КЛАССИЧЕСКАЯ УЧЕБНАЯ ЛИТЕРАТУРА
ПО МАТЕМАТИКЕ

А. А. БОРОВКОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

•

УЧЕБНИК

Издание четвертое,
стереотипное

ЛАНЬ®
САНКТ-ПЕТЕРБУРГ · МОСКВА · КРАСНОДАР
2010

ББК 22.172я73
Б 83

Боровков А. А.

Б 83 Математическая статистика: Учебник. 4-е изд., стер. — СПб.: Издательство «Лань», 2010. — 704 с. — (Учебники для вузов. Специальная литература).

ISBN 978-5-8114-1013-2

В учебнике излагаются основания современной математической статистики, предельные теоремы для эмпирических распределений и основных типов статистик, основы теории оценок и теории проверки гипотез. Рассматриваются методы отыскания оптимальных и асимптотически оптимальных процедур. Значительное внимание уделено статистике разнораспределенных наблюдений и, в частности, задачам однородности, задачам регрессии и дискретного анализа, распознаванию образов и задаче о разладке. Излагается единый теоретико-игровой подход к задачам математической статистики. Изучаются статистические игры и основные принципы отыскания оптимальных и асимптотически оптимальных решающих правил. Многие результаты теории оценивания и теории проверки гипотез обобщаются на случай произвольной функции потерь.

Учебник рассчитан на студентов и аспирантов математических и физических специальностей вузов.

ББК 22.172я73

Обложка
А. Ю. ЛАПШИН

Охраняется законом РФ об авторском праве.
Воспроизведение всей книги или любой ее части
запрещается без письменного разрешения издателя.
Любые попытки нарушения закона
будут преследоваться в судебном порядке.

© Издательство «Лань», 2010
© А. А. Боровков, 2010
© Издательство «Лань»,
художественное оформление, 2010

ОГЛАВЛЕНИЕ

Предисловие ко второму изданию 11

Введение 18

Глава 1. **Выборка. Эмпирическое распределение. Асимптотические свойства статистик** 21

§ 1. Понятие выборки 21

§ 2. Эмпирическое распределение (одномерный случай) 24

§ 3. Выборочные характеристики. Основные типы статистик 28

1. Примеры выборочных характеристик (28). 2. Два типа статистик (29). 3. L-статистики (31). 4. M-статистики (32). 5. О других статистиках (33).

§ 4. Многомерные выборки 33

1. Эмпирические распределения (33). 2*. Более общие варианты теоремы Гливенко–Кантелли. Закон повторного логарифма (34). 3. Выборочные характеристики (35).

§ 5. Теоремы непрерывности 35

§ 6*. Эмпирическая функция распределения как случайный процесс. Сходимость к броуновскому мосту 40

1. Распределение процесса $nF_n^*(t)$ (40). 2. Предельное поведение процесса $w^n(t)$ (43).

§ 7. Предельное распределение для статистик первого типа 46

§ 8*. Предельное распределение для статистик второго типа 51

§ 9*. Замечания о непараметрических статистиках 62

§ 10*. Сглаженные эмпирические распределения. Эмпирические плотности 63

Глава 2. **Теория оценивания неизвестных параметров** 69

§ 11. Предварительные замечания 69

§ 12. Некоторые параметрические семейства распределений и их свойства 71

1. Нормальное распределение на прямой (71). 2. Многомерное нормальное распределение (71). 3. Гамма-распределение (72). 4. Распределение «хи-квадрат» $\mathbf{H}_k$ с k степенями свободы (73). 5. Экспоненциальное распределение (74). 6. Распределение Фишера с числом степеней свободы k_1, k_2 (74). 7. Распределение Стьюдента $\mathbf{T}_k$ с k степенями свободы (75). 8. Бэта-распределение (B-распределение) (77). 9. Равномерное распределение (78). 10. Распределение Коши $\mathbf{K}_{\alpha,\sigma}$ с параметрами (α, σ) (80). 11. Логнормальное распределение $\mathbf{L}_{\alpha,\sigma^2}$ (80). 12. Вырожденное распределение (81). 13. Распределение Бернулли $\mathbf{B}_p^n$ (81). 14. Распределение Пуассона $\mathbf{\Pi}_\lambda$ (81). 15. Полиномиальное распределение (81).

§ 13. Точечное оценивание. Основной метод получения оценок. Состоятельность, асимптотическая нормальность 82

1. Метод подстановки. Состоятельность (82). 2. Асимптотическая нормальность. Одномерный случай (85). 3. Асимптотическая нормальность. Случай многомерного параметра (86).

§ 14. Реализация метода подстановки в параметрическом случае. Метод моментов. M-оценки 87

1. Метод моментов. Одномерный случай (87). 2. Метод моментов. Многомерный случай (89). 3. M-оценивание как обобщенный метод моментов (90). 4. Состоятельность $\widehat{M}$-оценок (92). 5. Состоятельность M-оценок (95). 6. Асимптотическая нормальность M-оценок (96). 7. Замечания о многомерном случае (98).

§ 15*. Метод минимального расстояния 99

§ 16. Метод максимального правдоподобия. Оптимальность оценок максимального правдоподобия в классе M-оценок 101

1. Определения. Общие свойства (101). 2. Асимптотические свойства о. м. п. Состоятельность (109). 3. Асимптотическая нормальность о. м. п. Оптимальность в классе M-оценок (113).

§ 17. О сравнении оценок 115

1. Среднеквадратический подход. Одномерный случай (116). 2. Асимптотический подход. Одномерный случай (119). 3. Нижняя граница рассеивания для L-оценок (121). 4. Среднеквадратический и асимптотический подходы в многомерном случае (124). 5. Некоторые эвристические подходы к определению дисперсий оценок. Методы складного ножа и будстрэп (127).

§ 18. Сравнение оценок в параметрическом случае. Эффективные оценки 128

1. Одномерный случай. Среднеквадратический подход (129). 2. Асимптотический подход. Асимптотическая эффективность в классах M- и L-оценок (131). 3. Многомерный случай (138).

§ 19. Условные математические ожидания 140

1. Определение у. м. о. (140). 2. Свойства у. м. о. (143).

§ 20. Условные распределения 146

§ 21. Байесовский и минимаксный подходы к оцениванию параметров . . 150

§ 22. Достаточные статистики 157

§ 23*. Минимальные достаточные статистики 164

§ 24. Построение эффективных оценок с помощью достаточных статистик. Полные статистики 171

1. Одномерный случай (171). 2. Многомерный случай (173). 3. Полные статистики и эффективные оценки (173).

§ 25. Экспоненциальное семейство 178

§ 26. Неравенство Рао–Крамера и R-эффективные оценки 185

1. Неравенство Рао–Крамера и его следствия (185). 2. R-эффективные и асимптотически R-эффективные оценки (191). 3. Неравенство Рао–Крамера в многомерном случае (194). 4. Некоторые выводы (200).

§ 27*. Свойства информации Фишера 200

1. Одномерный случай (200). 2. Многомерный случай (203). 3. Матрица Фишера и замена параметра (206).

§ 28*. Оценки параметра сдвига и масштаба. Эффективные эквивариантные оценки 207

1. Оценки параметра сдвига и масштаба (207). 2. Эффективная оценка параметра сдвига в классе эквивариантных оценок (208). 3. Минимаксность оценки Питмена (211). 4. Об оптимальных оценках параметра масштаба (212).

§ 29*. Общая задача об эквивариантном оценивании 215

§ 30. Интегральное неравенство типа Рао–Крамера. Критерии асимптотической байесовости и минимаксности оценок 218

1. Эффективные и сверхэффективные оценки (218). 2. Основные неравенства (220). 3. Неравенства в случае, когда функция $q(\theta)/I(\theta)$ не дифференцируема (224). 4. Некоторые следствия. Критерии асимптотической байесовости и минимаксности (225). 5. Многомерный случай (228).

§ 31. Расстояния Кульбака–Лейблера, Хеллингера и χ^2. Их свойства . . . 228

1. Определения и основные свойства расстояний (228). 2. Связь расстояний Хеллингера и других с информацией Фишера (232). 3. Существование равномерных границ для $r(\Delta)/\Delta^2$ (233). 4. Многомерный случай (234). 5*. Связь рассматриваемых расстояний с оценками (236).

§ 32*. Свойства информации Фишера. Разностное неравенство типа Рао–Крамера 238

§ 33. Вспомогательные неравенства для отношения правдоподобия. Асимптотические свойства оценок максимального правдоподобия 243

1. Основные неравенства (244). 2. Оценки для распределения и моментов о. м. п. Состоятельность о. м. п. (246). 3. Асимптотическая нормальность (247). 4. Асимптотическая эффективность (249). 5. Асимптотическая байесовость о. м. п. (250).

§ 34. Асимптотические свойства отношения правдоподобия. Дальнейшие свойства оптимальности о. м. п. 251

§ 35*. Приближенное вычисление оценок максимального правдоподобия . 260

§ 36. Результаты § 33, 34 для случая многомерного параметра 269

1. Неравенства для отношения правдоподобия и свойства о. м. п. (результаты § 33) (269). 2. Асимптотические свойства отношения правдоподобия (результаты § 33, 34) (271). 3. Свойства о. м. п. (результаты § 33, 34) (275). 4. Приближенное вычисление о. м. п. (278). 5. Свойства о. м. п. при отсутствии условий регулярности (результаты § 14, 16) (278).

§ 37. Равномерность по θ асимптотических свойств отношения правдоподобия и оценок максимального правдоподобия 278

1. Равномерные закон больших чисел и центральная предельная теорема (278). 2. Равномерные варианты теорем об асимптотических свойствах отношения правдоподобия и оценок максимального правдоподобия (280). 3. Некоторые следствия (283).

§ 38*. О статистических задачах, связанных с выборками случайного объема. Последовательное оценивание 284

§ 39. Интервальное оценивание 285

1. Определения (285). 2. Построение доверительных интервалов в байесовском случае (286). 3. Построение доверительных интервалов в общем случае. Асимптотические доверительные интервалы (287). 4. Построение точного доверительного интервала с помощью заданной статистики (290). 5. Другие методы построения доверительных интервалов (293). 6. Многомерный случай (295).

§ 40. Точные выборочные распределения и доверительные интервалы для нормальных совокупностей 296

1. Точные распределения статистик $\overline{x}$, S_0^2 (296). 2. Построение точных доверительных интервалов для параметров нормального распределения (298).

Глава 3. **Теория проверки гипотез** 301

§ 41. Проверка конечного числа простых гипотез 301

1. Постановка задачи. Понятие статистического критерия. Наиболее мощный критерий (301). 2. Байесовский подход (303). 3. Минимаксный подход (308). 4. Наиболее мощные критерии (309).

§ 42. Проверка двух простых гипотез 311

§ 43*. Два асимптотических подхода к расчету критериев. Численное сравнение 315

1. Предварительные замечания (315). 2. Фиксированные гипотезы (316). 3. Близкие гипотезы (320). 4. Сравнение асимптотических подходов. Числовой пример (323). 5. Связь н. м. к. с асимптотической эффективностью о. м. п. (328).

§ 44. Проверка сложных гипотез. Классы оптимальных критериев 329

1. Постановка задачи и основные понятия (329). 2. Равномерно наиболее мощные критерии (331). 3. Байесовские критерии (332). 4. Минимаксные критерии (333).

§ 45. Равномерно наиболее мощные критерии 333

1. Односторонние альтернативы. Монотонное отношение правдоподобия (333). 2. Двусторонняя основная гипотеза. Экспоненциальное семейство (336). 3. Другой подход к рассматриваемым задачам (341). 4. Байесовский подход и наименее благоприятные априорные распределения при построении н. м. к. и р. н. м. к. (342).

§ 46*. Несмещенные критерии 344

1. Определения. Несмещенные р. н. м. к. (344). 2. Двусторонние альтернативы. Экспоненциальное семейство (347).

§ 47*. Инвариантные критерии 349

§ 48*. Связь с доверительными множествами 353

1. Связь статистических критериев и доверительных множеств. Связь свойств оптимальности (353). 2. Наиболее точные доверительные интервалы (356). 3. Несмещенные доверительные множества (359). 4. Инвариантные доверительные множества (360).

§ 49. Байесовский и минимаксный подходы к проверке сложных гипотез 363

1. Байесовские и минимаксные критерии (363). 2. Минимаксные критерии для параметра α нормальных распределений (366). 3. Вырожденные наименее благоприятные распределения для односторонних гипотез (373).

§ 50. Критерий отношения правдоподобия 375

§ 51*. Последовательный анализ 378

1. Вводные замечания (378). 2. Байесовский последовательный критерий (379). 3. Последовательный критерий, минимизирующий среднее число испытаний (383). 4. Вычисление параметров наилучшего последовательного критерия (385).

§ 52. Проверка сложных гипотез в общем случае 388

§ 53. Асимптотически оптимальные критерии. Критерий отношения правдоподобия как асимптотически байесовский критерий для проверки простой гипотезы против сложной 397

1. Асимптотические свойства к. о. п. и байесовского критерия (397). 2. Асимптотическая байесовость к. о. п. (399). 3. Асимптотическая несмещенность к. о. п. (403)

§ 54. Асимптотически оптимальные критерии для проверки близких сложных гипотез 404

1. Постановка задачи и определения (404). 2. Основные утверждения (407).

§ 55. Свойства асимптотической оптимальности критерия отношения правдоподобия, вытекающие из предельного признака оптимальности . . 412

1. А. р. н. м. к. для близких гипотез с односторонними альтернативами (412). 2. А. р. н. м. к. для двусторонних альтернатив (413). 3. Асимптотически минимаксный критерий для близких гипотез, касающихся многомерного параметра (415). 4. Асимптотически минимаксный критерий о принадлежности выборки параметрическому подсемейству (417).

§ 56. Критерий χ^2. Проверка гипотез по сгруппированным данным 423

1. Критерий χ^2. Свойства асимптотической оптимальности (423). 2. Применения критерия χ^2. Проверка гипотез по сгруппированным данным (427).

§ 57. Проверка гипотез о принадлежности выборки параметрическому семейству 430

1. Проверка гипотезы $X \in \mathbf{B}_{\theta(\alpha)}$. Группировка данных (430). 2. Общий случай (434).

§ 58. Устойчивость статистических решений (робастность) 438

1. Постановка задачи. Качественная и количественная характеризации робастности (438). 2. Оценка параметра сдвига (445). 3. Статистики Стьюдента и S_0^2 (448). 4. Критерий отношения правдоподобия (449).

Глава 4. **Статистические задачи с двумя и более выборками** 450

§ 59. Проверка гипотез об однородности (полной или частичной) в параметрическом случае 450

1. Рассматриваемый класс задач (450). 2. Асимптотически минимаксный критерий для проверки близких гипотез об обычной однородности (452). 3. Асимптотически минимаксные критерии для задачи об однородности при наличии мешающего параметра (458). 4. Асимптотически минимаксный критерий для задачи о частичной однородности (463). 5. Некоторые другие задачи (466).

§ 60. Задачи об однородности в общем случае 466

1. Постановка задачи (466). 2. Критерий Колмогорова–Смирнова (467). 3. Критерий знаков (469). 4. Критерий Вилкоксона (470). 5. Критерий χ^2 как асимптотически оптимальный критерий проверки однородности по сгруппированным данным (475).

§ 61. Задачи регрессии 476

1. Постановка задачи (476). 2. Оценка параметров (477). 3. Проверка гипотез относительно линейной регрессии (484). 4. Оценивание и проверка гипотез при наличии линейных связей (488).

§ 62. Дисперсионный анализ 491

1. Задачи дисперсионного анализа как задачи регрессии. Случай одного фактора (491). 2. Влияние двух факторов. Элементарный подход (494).

§ 63. Распознавание образов 496

1. Параметрический случай (497). 2. Общий случай (497).

Глава 5. **Статистика разнораспределенных наблюдений** 500

§ 64. Предварительные замечания. Примеры 500

§ 65. Основные методы построения оценок. M-оценки. Состоятельность и асимптотическая нормальность 507

1. Предварительные замечания и определения (507). 2. M-оценки (508). 3*. Состоятельность $\widehat{M}$-оценок (514). 4. Состоятельность M-оценок (519). 5. Асимптотическая нормальность M-оценок (521).

§ 66. Оценки максимального правдоподобия. Основные принципы сравнения оценок. Оптимальность о. м. п. в классе M-оценок 524

1. Оценки максимального правдоподобия (524). 2. Асимптотические свойства о. м. п. (526). 3. Основные принципы сравнения оценок. Асимптотическая эффективность о. м. п. в классе M-оценок (531).

§ 67. Достаточные статистики. Эффективные оценки. Экспоненциальные семейства 533

§ 68. Эффективные оценки в задаче оценивания «хвостов» распределений (пример 64.7). Асимптотические свойства оценок 535

1. Оценки максимального правдоподобия (535). 2. Асимптотическая нормальность $\widehat{\theta}^*$ в задаче B (537). 3*. Асимптотические нормальность и оптимальность в задаче A (540).

§ 69. Неравенство Рао–Крамера 545

§ 70. Неравенства для отношения правдоподобия и асимптотические свойства о. м. п. 547

1. Неравенства для отношения правдоподобия и состоятельность о. м. п. (547). 2. Асимптотическая нормальность о. м. п. (550). 3. Асимптотическая эффективность (552). 4. Замечания об о. м. п. для многомерного параметра (552).

§ 71. Замечания о проверке гипотез по неоднородным наблюдениям 553

§ 72. Задача о разладке (пример 64.6) 556

1. Задача о разладке как задача проверки гипотез (556). 2. Задача о разладке как задача оценки параметра (563). 3. Последовательные процедуры (566).

Глава 6. **Теоретико-игровой подход к задачам математической статистики** 575

§ 73. Предварительные замечания 575

§ 74. Основные понятия и теоремы, связанные с игрой двух лиц 576

1. Игра двух лиц (576). 2. Равномерно оптимальные стратегии в подклассах (577). 3. Байесовские стратегии (577). 4. Минимаксные стратегии (579). 5. Полный класс стратегий (587).

§ 75. Статистические игры 587

1. Описание статистических игр (587). 2. Классификация статистических игр (590). 3. Две фундаментальные теоремы теории статистических игр (591).

§ 76. Байесовский принцип. Полный класс решающих функций 593

§ 77. Достаточность, несмещенность, инвариантность 599

1. Достаточность (599). 2. Несмещенность (600). 3. Инвариантность (602).

§ 78. Асимптотически оптимальные оценки при произвольной функции потерь 605

§ 79. Оптимальные статистические критерии при произвольной функции потерь. Критерий отношения правдоподобия как асимптотически байесовское решение 616

1. Свойства оптимальности статистических критериев при произвольной функции потерь (616). 2. К. о. п. как асимптотически байесовский критерий (616).

§ 80. Асимптотически оптимальные решения при произвольной функции потерь в случае близких сложных гипотез 620

Таблицы 625

Приложение I 634

Приложение II 637

Приложение III 644

Приложение IV 647

Приложение V 658
Приложение VI 664
Приложение VII 669
Приложение VIII 679
Библиографические замечания 685
Список литературы 692
Список основных обозначений 698
Предметный указатель 701

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Предлагаемая книга является существенно переработанной и дополненной версией книги «Математическая статистика», вышедшей в 1984 г. в двух частях: первая часть имела подзаголовок «Оценка параметров, проверка гипотез», вторая — «Дополнительные главы». Позже эти две книги были опубликованы в виде монографии на французском (1987) и испанском (1988) языках. В 1997 г. дополненная и переработанная версия книги издана на английском языке.

Одним из главных нововведений в предлагаемой книге по сравнению с изданиями 1984 г. является появление новой главы, посвященной статистике разнораспределенных наблюдений.

В целом в основу предлагаемой книги положены лекции по математической статистике, которые автор в течение многих лет читал на математическом факультете Новосибирского университета. Со временем курс лекций неоднократно менялся в поисках варианта, который был бы по возможности более стройным, доступным и в то же время соответствовал современному состоянию этой науки. Были испробованы варианты, начиная с курса преимущественно рецептурного характера с изложением основных типов задач (построение оценок, критериев и изучение их свойств) и кончая курсом общего теоретико-игрового характера, в котором теория оценок и проверка гипотез представлялись как частные случаи единого подхода. Ограничения во времени не позволяли объединить эти дополняющие друг друга варианты, каждый из которых в отдельности обладал очевидными недостатками. В первом случае набор конкретных фактов мешал развитию общего взгляда на предмет. Второй вариант испытывал недостаток в простых конкретных результатах и оказывался загруженным многими новыми непростыми понятиями, трудными для усвоения. По-видимому, наиболее приемлемым является вариант, в котором изложение элементов теории оценок и теории проверки гипотез сочетается с последовательным проведением линии на отыскание оптимальных процедур.

Основу книги составляет объединение материала, читавшегося в разное время в лекциях, расширенное за счет разделов, присутствие которых диктуется самой логикой изложения. Главной целью являлось изложение современного состояния предмета в сочетании с максимально возможной доступностью и математической цельностью и стройностью.

Содержание книги распадается на 6 глав и приложения.

В главе 1 изучаются свойства (в основном асимптотические) эмпирических распределений, лежащие в основе математической статистики.

В главах 2, 3 излагаются соответственно теория оценок и теория проверки статистических гипотез. Первые части каждой главы посвящены описанию возможных подходов к решению поставленных задач и отысканию оптимальных процедур. Вторые части посвящены построению асимптотически оптимальных процедур.

Такую же структуру имеет глава 6, где излагается общий теоретико-игровой подход к задачам математической статистики.

Глава 4 посвящена задачам, связанным с двумя и более выборками. Статистические выводы для разнораспределенных наблюдений в более общей, чем в главе 4, постановке задачи рассматриваются в новой главе 5, которая уже упоминалась выше.

Книга содержит также 8 приложений. Они связаны с теми утверждениями в основном тексте, доказательство которых выходит за рамки основного изложения либо по своему характеру, либо по трудности.

Имеются еще библиографические замечания, не претендующие на полноту, но позволяющие проследить возникновение и развитие основных направлений математической статистики. При этом везде, где это возможно, оказывалось предпочтение ссылкам на монографии (как на более доступный вид литературы), а не на оригинальные статьи.

В настоящее время существует довольно много книг по математической статистике. Выделим из них следующие, в которых представлено большое количество материала, отражающее современное состояние предмета — это книги Г. Крамера [58], Э. Лемана [66, 67], Ш. Закса [48], И.А. Ибрагимова и Р.З. Хасьминского [49], L. Le Cam [63], G.R. Shorak, J.A. Wellner [139]. Из них наибольшее влияние на изложение настоящей книги оказали монографии [49] (некоторые идеи этой книги использовались в § 33, 34, 36, 37 главы 2) и [67] (изложение § 45–48 главы 3 близко по содержанию к соответствующим разделам [67]). Остальное изложение мало связано по строю с имеющимися книгами.

Существует много других книг, занимающих заметное место в статистической литературе (такие, как книги Блекуэлла и Гиршика [9], Кендалла и Стьюарта [50], Кокса и Хинкли [54], Фергюссона [112], Рао [96] и ряд других — все перечислить невозможно), но они по духу и по подбору материала существенно отличаются от предлагаемой монографии.

Наряду с известными результатами и подходами в настоящую книгу включены некоторые новые разделы, упрощающие изложение, ряд методологических улучшений, а также некоторые новые результаты и результаты, впервые публикуемые в монографической литературе.

Ниже приводится краткое описание методологической структуры книги (см. также оглавление и краткие предисловия к каждой из глав).

Глава 1 излагает основные результаты, связанные с понятиями выборки, эмпирического распределения и статистик.

В § 1, 2 вводятся понятия выборки, эмпирического распределения и устанавливается теорема Гливенко–Кантелли, которую можно рассматривать как фундаментальный факт, составляющий основу статистических выводов.

В § 3 вводится два типа статистик (статистики I и II типов), включающие в себя подавляющее большинство практически интересных статистик. Эти статистики определяются как значения $G(\mathbf{P}_n^*)$ функционалов G (удовлетворяющих некоторым условиям) от эмпирического распределения $\mathbf{P}_n^*$. Класс рассматриваемых статистик расширен включением в него также так называемых L-, M- и U-статистик. Позже, в § 7, 8, устанавливаются предельные теоремы о распределении этих статистик. Это разгружает последующее изложение и освобождает от необходимости для каждой конкретной статистики проводить очень схожие рассуждения, не относящиеся притом к самому предмету статистики.

В § 5 собраны вспомогательные теоремы (названные в книге теоремами непрерывности) о сходимости распределений статистик и о сходимости их моментов. Это также разгружает последующее изложение.

В § 6 (он не обязателен при первом чтении) устанавливается, что эмпирическая функция распределения $F_n^*(t)$ есть условный пуассоновский процесс, и приводится формулировка теоремы (доказанной в приложении I) о сходимости процесса $\sqrt{n}\left(F_n^*(t) - F(t)\right)$ к броуновскому мосту.

В § 10 вводятся сглаженные эмпирические распределения, позволяющие приближать не только само распределение, но и его плотность.

Глава 2 посвящена оценкам неизвестных параметров.

В § 13 вводится единый подход к построению оценок, названный *методом подстановки*. Он состоит в том, что оценку θ^* для параметра θ, представимого в виде функционала $\theta = G(\mathbf{P})$ от распределения $\mathbf{P}$ выборки, следует искать в виде $\theta^* = G(\mathbf{P}_n^*)$, где $\mathbf{P}_n^*$ — эмпирическое распределение. Все «разумные» оценки, используемые на практике, являются оценками подстановки. Оптимальность оценки достигается путем выбора подходящего функционала G. Если статистика $\theta^* = G(\mathbf{P}_n^*)$ является статистической I или II типов, то теоремы главы 1 позволяют немедленно установить состоятельность этих оценок и их асимптотическую нормальность.

В § 14, 15 этот подход иллюстрируется на оценках, полученных с помощью метода моментов и метода минимального расстояния. Там же доказаны состоятельность и асимптотическая нормальность M-оценок.

С тех же позиций в § 16 рассматриваются и оценки максимального правдоподобия. Кроме того, установлена нижняя граница рассеивания M-оценок и доказана асимптотическая оптимальность оценок максимального правдоподобия в классе M-оценок. Более полное изучение асимптотических свойств оценок максимального правдоподобия проводится позже, в § 34, 34.

В § 17 установлена нижняя граница рассеивания для L-оценок. Это дает возможность для параметра сдвига строить в явном виде асимптотически оптимальные L-оценки, асимптотически эквивалентные оценкам максимального правдоподобия.

Сравнение оценок в главе 2 происходит на основе двух подходов: среднеквадратического (сравниваются $\mathbf{E}_\theta(\theta^* - \theta)^2$) и асимптотического (сравниваются дисперсии предельного распределения $\sqrt{n}\,(\theta^* - \theta)$ в классе асимптотически нормальных оценок). В качестве иллюстрации с помощью таких подходов строятся асимптотически оптимальные оценки в классе L-оценок. В параметрическом случае такие подходы позволяют выделить три типа оптимальных оценок: эффективные оценки в классах K_b с фиксированным смещением b, байесовские и минимаксные оценки. На основе тех же принципов выделяются классы асимптотически оптимальных оценок при асимптотическом подходе. Для построения эффективных оценок используются следующие традиционные методы: первый носит качественный характер и связан с принципом достаточности (§ 22–24); второй метод основан на количественных соотношениях, вытекающих из неравенства Рао–Крамера (§ 26); третий связан с соображениями инвариантности (§ 28, 29), позволяющими сузить класс рассматриваемых оценок. Отысканию асимптотически оптимальных оценок, а также изучению асимптотических свойств функции правдоподобия посвящены § 30–38. Параграф 30 содержит интегральное не-

равенство типа Рао–Крамера, которое позволяет, в частности, получить простые критерии асимптотической байесовости и минимаксности оценок, а также обосновать выделение некоторого подкласса оценок $\widetilde{K}_0$, которым следует ограничиться в поисках асимптотически эффективных оценок. Это позволяет путем изучения асимптотических свойств оценок максимального правдоподобия в § 34 сразу установить асимптотическую байесовость и минимаксность названных оценок и их асимптотическую эффективность в $\widetilde{K}_0$. Параграфы 31–33 носят вспомогательный характер.

Интервальное оценивание параметров рассматривается в § 39, 40, 49 главы 3.

Глава 3 посвящена проверке гипотез.

В § 41, 42 рассматривается случай конечного числа простых гипотез. Выделяются (аналогично теории оценивания) три типа оптимальных критериев — наиболее мощные в подклассах, байесовские и минимаксные. Устанавливаются связи между этими критериями и находится их явный вид. При этом в основу рассмотрений положен байесовский принцип (а не лемма Неймана–Пирсона), что, на наш взгляд, упрощает изложение и делает его более прозрачным.

В § 43 излагаются два асимптотических подхода к расчету критериев для проверки двух простых гипотез и производится их сравнение.

В § 44 рассматривается общая постановка задачи о проверке двух сложных гипотез и определяются классы оптимальных критериев (равномерно наиболее мощные, байесовские, минимаксные).

Параграф 45 посвящен отысканию равномерно наиболее мощных критериев в тех случаях, когда это возможно.

В § 46, 47 решается та же задача, но в классах критериев, суженных на основании соображений несмещенности и инвариантности. При этом в основу рассмотрений, как и в § 41, 42, положен байесовский подход.

В § 48 с помощью полученных результатов строятся наиболее точные доверительные множества.

В § 49 рассматриваются байесовские и минимаксные критерии.

Параграфы 50, 53 посвящены критерию отношения правдоподобия. Он оказывается равномерно наиболее мощным во многих частных случаях и обладает свойством асимптотической байесовости при весьма широких предположениях.

Изучение свойств асимптотической оптимальности критерия отношения правдоподобия продолжено в § 55–57.

В § 51 установлена оптимальность этого критерия в задачах последовательного анализа.

Параграфы 54, 55 посвящены отысканию асимптотически оптимальных критериев для проверки близких гипотез и найден их простой явный вид для основных статистических задач.

Существенной особенностью первых трех глав является тот факт, что в них рассмотрены лишь статистические задачи, связанные с использованием одной выборки.

Глава 4 книги, как уже отмечалось, посвящена задачам с двумя и более выборками. К ним относятся прежде всего задачи об однородности (полной или частичной, § 59, 60), задачи регрессии (§ 61) и дисперсионного анализа

(§ 62). На основании результатов главы 3 для задач однородности в параметрическом случае построены асимптотически оптимальные критерии в предположении, что альтернативные гипотезы являются близкими к основной гипотезе об однородности. Для задач регрессии (как для линейной регрессии, так и для регрессии по произвольным функциям) с помощью результатов глав 2, 3 найдены эффективные оценки неизвестных параметров и построены критерии для проверки основных гипотез. Рассмотрены также так называемые задачи распознавания образов (§ 63), которые появляются в учебной литературе, по-видимому, впервые.

Новая глава 5 «Статистика разнораспределенных наблюдений» по своей структуре во многом повторяет главу 2. Ее появление в настоящем издании вызвано тем, что в приложениях стало появляться все больше задач, связанных с неоднородными наблюдениями, типичным примером которых могут служить задачи нелинейной регрессии (но не только они). В то же время общие методы исследования таких задач были разработаны лишь фрагментарно, отсутствовало их систематическое изложение. Не претендуя ни в какой мере на то, чтобы полностью заполнить этот пробел, автор имел своей целью перенести на неоднородный случай основные результаты и подходы, изложенные в исходной книге для статистики однородных наблюдений.

В § 64 приведен ряд типичных задач (некоторые из них представляют самостоятельный интерес), относящихся к статистике разнораспределенных наблюдений. Эти примеры в дальнейшем используются для иллюстраций и для более детального изучения.

В § 65 излагаются основные методы построения оценок. В основном это метод $\widehat{M}$- и M-оценивания. На эти оценки переносятся основные результаты главы 2 о состоятельности и асимптотической нормальности этих оценок.

В § 66 с помощью результатов § 65 устанавливаются асимптотические свойства оценок максимального правдоподобия (состоятельность, асимптотическая нормальность, асимптотическая оптимальность в классе M-оценок).

В § 67 содержатся комментарии об использовании результатов, связанных с достаточными статистиками и экспоненциальными семействами в случае разнораспределенных наблюдений.

В § 68 в качестве иллюстрации детально изучаются асимптотические свойства оценок параметров хвостов распределений.

В § 69 рассмотрено обобщение на случай разнородных наблюдений неравенства Рао–Крамера и близких задач; § 70 посвящен обобщению (на тот же случай) результатов главы 2 об асимптотических свойствах оценок максимального правдоподобия (§ 33, 34).

Перенос большинства результатов главы 3 на случай разнораспределенных наблюдений либо не требует подробных рассмотрений, либо основан на обобщениях результатов главы 2, упомянутых выше (комментарии по этому поводу см. в § 71).

В § 72 изучается задача о разладке. Для нее найдены асимптотически оптимальные процедуры.

Глава 6 посвящена общему теоретико-игровому подходу к задачам статистики. Он способствует выработке общего взгляда на предмет математической статистики и позволяет обобщить многие результаты глав 2, 3.

В § 74 излагаются основные понятия и результаты «обычной» теории игр (рассматриваются лишь игры двух лиц). В частности, устанавливаются

связи между основными типами оптимальных стратегий — байесовскими, минимаксными, равномерно наилучшими в подклассах.

В § 75 изучаются статистические игры.

В § 76 формулируется и доказывается так называемый байесовский принцип, который позволяет свести задачу отыскания байесовского статистического решения к значительно более простой задаче построения байесовской стратегии для обычной игры двух лиц.

В § 77 обсуждаются принципы достаточности, несмещенности и инвариантности для построения решений, равномерно наилучших в соответствующих подклассах.

Параграфы § 78–80 посвящены отысканию асимптотически оптимальных решающих правил. В § 78 изучаются асимптотически оптимальные оценки параметров при произвольной (не только квадратической) функции потерь. В этом случае удается установить результаты, близкие к результатам главы 2 об асимптотической оптимальности оценок максимального правдоподобия. В § 79, 80 изучаются асимптотически оптимальные критерии при произвольной функции потерь. В § 79 доказана асимптотическая байесовость критерия отношения правдоподобия; в § 80 установлен предельный признак оптимальности критериев для проверки близких гипотез (обобщение результатов § 54, 55 главы 3 на случай произвольной функции потерь).

Из приложений выделим приложение VII, в котором найдены асимптотически оптимальные процедуры в задаче о разладке, и приложение VIII, в котором доказаны две фундаментальные теоремы теории статистических игр. Чтение приложений требует более высокой математической подготовки.

Книга имеет многоцелевое назначение. В ее полном объеме она, конечно, ближе к программе кандидатского минимума по специальности «математическая статистика», чем к учебному пособию для студентов. Однако в книге предусмотрена система мер, упрощающих первое чтение, которая делает книгу доступной и для студентов. Параграфы повышенной трудности или «более продвинутые» по своему содержанию отмечены звездочкой, их следует при первом чтении опускать, как и текст, набранный петитом. Кроме того, изложение технически более сложного случая, относящегося к многомерному параметру, почти везде выделено в отдельные разделы и параграфы, которые также можно пропускать.

Преподаватели вузов, уже знакомые хотя бы частично с предметом, могут выбирать из книги совокупность параграфов (вариантов может быть много), используя которые (не обязательно полностью) можно составить семестровый курс математической статистики. Вот один из вариантов: § 1, 3, 5 главы 1; § 12–14, 16–22, 24, 26 (31, 33, 34), 40, 49 главы 2; § 41, 42, 44, 45, 52 (53, 56) главы 3. Параграфы, взятые в скобки, посвящены асимптотически оптимальным процедурам. Их следует в зависимости от подготовленности студентов либо максимально облегчить, либо опустить вовсе.

Чтение книги предполагает знакомство с курсом теории вероятностей, например, в объеме учебника А.А. Боровкова [17]. (Возможно, конечно, использование и других источников.) Ссылки на эту книгу, в отличие от других ссылок, появляются в тех местах, которые предполагаются читателю известными, и служат, главным образом, для напоминаний. При этом мы будем использовать сокращенную запись этих ссылок. Например, вместо § 3 главы 13 в [17] мы будем писать [17, § 13.3].

Нумерация параграфов в книге сквозная; нумерация теорем, лемм, примеров и т. д. в каждом параграфе самостоятельная. Для удобства чтения используются разные системы для ссылок на теоремы, леммы, примеры, формулы и т. д. в зависимости от того, в каком параграфе они находятся. Если делается ссылка на теорему 1 или формулу (12) читаемого параграфа, то ссылки на них будут выглядеть так: теорема 1, формула (12). Если делается ссылка на теорему 1 и формулу (12) одного из предыдущих параграфов (например, § 13), то ссылки будут иметь такой вид: теорема 13.1, формула (13.12).

Значок $\triangleleft$ означает окончание доказательства.

Разделы, помеченные звездочкой, при первом чтении можно опустить.

Для удобства читателя в конце книги помещены список основных обозначений и предметный указатель.

Написание этой книги было весьма трудным многоэтапным делом. Значительную помощь при подготовке первоначальных записок лекций к печати и по устранению недостатков в них мне оказал И.С. Борисов. Второй вариант рукописи прочел по моей просьбе К.А. Боровков, который представил мне длинный список полезных советов и замеченных им погрешностей. Он же существенно помог мне при «чистке» окончательного варианта рукописи. В поисках новой критики я обратился с просьбой ознакомиться с рукописью к А.И. Саханенко. Он также представил мне длинный список замечаний и предложений по улучшению изложения, многими из которых я воспользовался. Наиболее существенные изменения при этом коснулись доказательств в § 26, 31, 33, 37 главы 2, § 53–55 главы 3, приложениях II, IV, VIII (см. также библиографические замечания).

Много ценных замечаний, направленных на улучшение книги, было сделано Д.М. Чибисовым. Рукопись просматривали В.В. Юринский и А.А. Новиков и также сделали ряд полезных замечаний. Всем названным коллегам и всем тем, кто так или иначе помогал мне при работе над книгой, я хотел бы выразить здесь свою глубокую и искреннюю признательность и благодарность за их труд и содействие в написании книги.

А. А. Боровков

Памяти моего учителя
Андрея Николаевича Колмогорова
посвящается

ВВЕДЕНИЕ

Настоящая книга посвящена изложению основ раздела математики, который называется *математическая статистика.* Последнюю для краткости называют иногда просто *статистикой.* Однако следует иметь в виду, что такое сокращение возможно лишь при достаточно хорошем взаимопонимании, поскольку сам по себе термин «статистика» соответствует часто несколько иному понятию.

Что представляет собой предмет математической статистики? Можно приводить разные описательные «определения», которые в большей или меньшей степени отражают содержание этого раздела математики. Одно из самых простых и грубых основано на сравнении, связанном с понятием выборки из конечной генеральной совокупности и с задачей о гипергеометрическом распределении, которая обычно рассматривается в начале курса теории вероятностей. Зная состав генеральной совокупности, там изучают распределения для состава случайной выборки. Это типичная *прямая задача* теории вероятностей. Однако часто приходится решать и *обратные задачи*, когда известен состав выборки и по нему требуется определить, какой была генеральная совокупность (это делается, например, при социологических опросах общественного мнения). Такого рода обратные задачи и составляют, образно говоря, предмет математической статистики.

Несколько уточняя это сравнение, можно сказать так: в теории вероятностей мы, зная природу некоторого явления, выясняем, как будут вести себя (как распределены) те или иные изучаемые нами характеристики, которые можно наблюдать в экспериментах. В математической статистике наоборот — исходными являются экспериментальные данные (как правило, это наблюдения над случайными величинами), а требуется вынести то или иное суждение или решение о природе рассматриваемого явления. Таким образом, мы соприкасаемся здесь с одной из важнейших сторон человеческой деятельности — процессом познания. Тезис о том, что «критерий истины есть практика» имеет самое непосредственное отношение к математической статистике, поскольку именно эта наука изучает методы (в рамках точных математических моделей), которые позволяют отвечать на вопрос, соответствует ли практика, представленная в виде результатов эксперимента, данному гипотетическому представлению о природе явления или нет.

При этом необходимо подчеркнуть, что, как и в теории вероятностей, нас будут интересовать не те эксперименты, которые позволяют делать однозначные, детерминированные выводы о рассматриваемых в природе явлениях, а эксперименты, результатами которых являются случайные события. С развитием науки роль такого рода задач постоянно возрастает, поскольку с увеличением точности экспериментов становится все труднее избежать «случайного фактора», связанного с разного рода помехами и ограниченностью наших измерительных и вычислительных возможностей.

Математическая статистика является частью теории вероятностей в том смысле, что каждая задача математической статистики есть по существу задача (иногда весьма своеобразная) теории вероятностей. Однако сама по себе математическая статистика занимает и самостоятельное положение в табеле о науках. Математическая статистика может рассматриваться как наука о так называемом индуктивном поведении человека (и не только человека) в условиях, когда он должен на основании своего недетерминированного опыта принимать решения с наименьшими для него потерями*).

Математическую статистику называют также теорией статистических решений, поскольку ее можно характеризовать как науку об оптимальных решениях (последние два слова требуют пояснения), основанных на статистических (экспериментальных) данных. Точные постановки задач будут даны позже, в основном тексте книги. Здесь же мы ограничимся тем, что приведем три примера наиболее простых и типичных статистических задач.

Пример 1. Для многих изделий одним из основных параметров, которым характеризуется качество, является срок службы. Однако срок службы изделия (скажем, электролампы), как правило, случаен, и заранее определить его невозможно. Опыт показывает, что если процесс производства в известном смысле однороден, то сроки службы $\xi_1, \xi_2, \ldots$ соответственно 1-го, 2-го и т. д. изделий можно рассматривать как независимые одинаково распределенные величины. Интересующий нас параметр, определяющий срок службы, естественно отождествить с математическим ожиданием $\theta = \mathrm{E}\xi_i$. Одна из стандартных задач состоит в выяснении, чему равно θ. Для того чтобы определить это значение, берут n готовых изделий и проверяют их. Пусть $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ — сроки службы этих проверенных изделий. Мы знаем, что

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow[\text{п.н.}]{} \theta$$

при $n \to \infty$. Поэтому естественно ожидать, что число $\overline{\mathrm{x}} = \dfrac{1}{n}\sum_{i=1}^{n}\mathrm{x}_i$ при достаточно большом n окажется близким к θ и позволит в какой-то мере ответить на поставленные вопросы. При этом очевидно, что мы заинтересованы в том, чтобы требуемое число наблюдений n было по возможности наименьшим, а наша оценка числа θ по возможности более точной (завышение параметра θ, как и его занижение, приведут к материальным потерям).

Пример 2. Радиолокационное устройство в моменты времени $t_1, t_2, \ldots, t_n$ зондирует заданную часть воздушного пространства с целью обнаружения там некоторого объекта. Обозначим через $\mathrm{x}_1, \ldots, \mathrm{x}_n$ значения отраженных сигналов, принятых устройством. Если в заданной части пространства интересующий нас объект отсутствует, то значения x_i можно рассматривать как независимые случайные величины, распределенные так же, как некоторая случайная величина ξ, природа которой обусловлена характером различных помех. Если же в течение всего периода наблюдений объект находился в поле зрения, то x_i будут наряду с помехами содержать «полезный» сигнал a, и значения x_i будут распределены как $\xi + a$. Таким образом, если

*) Подробнее об этом см. в [72].

в первом случае наблюдения x_i имели функцию распределения $F(x)$, то во втором случае их функция распределения будет иметь вид $F(x-a)$. По выборке $\mathrm{x}_1, \ldots, \mathrm{x}_n$ требуется решить, какой из этих двух случаев имеет место, т. е. существует в заданном месте интересующий нас объект или нет.

В этой задаче окажется возможным указать в известном смысле «оптимальное решающее правило», которое будет решать поставленную задачу с минимальными ошибками. Сформулированная задача может быть усложнена следующим образом. Сначала объект отсутствует, а затем, начиная с наблюдения с неизвестным номером θ, появляется. Требуется по возможности более точно определить момент θ появления объекта. Это так называемая *задача о разладке*, имеющая и целый ряд других интерпретаций, важных для приложений.

Пример 3. Некоторый эксперимент производится сначала n_1 раз в условиях A и затем n_2 раз в условиях B. Обозначим через $\mathrm{x}_1, \ldots, \mathrm{x}_{n_1}$ и $\mathrm{y}_1, \ldots, \mathrm{y}_{n_2}$ результаты этих экспериментов соответственно в условиях A и B. Спрашивается, сказывается ли изменение условий эксперимента на его результатах? Иными словами, если обозначить через $\mathbf{P}_{\mathrm{A}}$ распределение x_i, $1 \leqslant i \leqslant n_1$, и через $\mathbf{P}_{\mathrm{B}}$ — распределение y_i, $1 \leqslant i \leqslant n_2$, то вопрос состоит в том, выполнено соотношение $\mathbf{P}_{\mathrm{A}} = \mathbf{P}_{\mathrm{B}}$ или нет.

Например, если нужно установить, влияет ли некоторый препарат на развитие, скажем, растений или животных, то параллельно ставятся две серии экспериментов (с препаратом и без), результаты которых необходимо уметь сравнивать.

Часто возникают и более сложные задачи, когда аналогичный вопрос ставится для многих серий наблюдений, проведенных в различных условиях. Если результаты наблюдений зависят от условий, то бывает необходимым проверить тот или иной характер этой зависимости (так называемая задача о регрессии).

Список примеров типичных статистических задач, разных по сложности и по своему существу, можно было бы продолжить. Однако общими для всех них будут следующие два обстоятельства.

1. Перед нами не было бы никаких проблем, если бы распределения результатов наблюдений, которые фигурируют в задачах, были нам известны.

2. В каждой из этих задач мы должны по результатам экспериментов принимать какое-то решение относительно распределения имеющихся наблюдений (отсюда и название «Теория статистических решений», упоминавшееся выше).

В связи с этими двумя замечаниями принципиальное значение для всего дальнейшего и, в частности, для решения приведенных в качестве примеров задач, приобретает следующий факт. Оказывается, по результатам наблюдений $\mathrm{x}_1, \ldots, \mathrm{x}_n$ над некоторой случайной величиной ξ можно при больших n сколь угодно точно восстановить неизвестное распределение $\mathbf{P}$ этой случайной величины. Аналогичное утверждение справедливо и для любого функционала $\theta = \theta(\mathbf{P})$ от этого неизвестного распределения.

Этот факт лежит в основе математической статистики. Ему и более точным постановкам задач будет посвящена глава 1.

Глава 1

ВЫБОРКА. ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ. АСИМПТОТИЧЕСКИЕ СВОЙСТВА СТАТИСТИК

В § 1–4 вводятся понятия выборки и эмпирического распределения и рассматриваются их простейшие свойства, главным образом, асимптотические, лежащие в основе математической статистики.

В § 5 излагаются так называемые теоремы непрерывности (о сходимости распределений функций от последовательностей случайных величин), используемые на протяжении всей книги.

Параграфы 6–10 посвящены более тонким асимптотическим свойствам эмпирических распределений и изучению предельных распределений для основных типов статистик.

§ 1. Понятие выборки

Исходным материалом любого статистического исследования является совокупность результатов наблюдений. В простейших случаях они представляют собой экспериментальные (полученные в результате опытов) значения некоторой случайной величины ξ. Мы уже отмечали, что в задачах статистики распределение $\mathbf{P}$ этой случайной величины хотя бы частично неизвестно.

Точнее, пусть G — эксперимент, связанный со случайной величиной ξ. Формально для этого эксперимента со случайной величиной ξ мы должны построить математическую модель, в которую входит вероятностное пространство $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}}, \mathbf{P})$, и задать на нем подходящим образом измеримую функцию, которая и называется случайной величиной ξ (см. [17]). Пространство $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}}, \mathbf{P})$, не ограничивая общности, можно считать «выборочным» (см. [17]), т. е. считать, что $\mathcal{X}$ есть пространство значений $\xi(x) = x$. В этом случае $\mathbf{P}$ можно называть распределением ξ. Если ξ — числовая случайная величина, то $\mathcal{X}$ есть числовая прямая $\mathbb{R}$; если ξ — вектор, то $\mathcal{X} = \mathbb{R}^m$, $m > 1$. В дальнейшем мы, как правило, и будем иметь в виду только эти два случая, т. е. под $\mathcal{X}$ понимать $\mathbb{R}$ (одномерный случай) или $\mathbb{R}^m$, $m > 1$ (многомерный случай). В качестве $\mathfrak{B}_{\mathcal{X}}$ чаще всего выбирают σ-алгебру борелевских множеств.

В то же время следует отметить, что многие результаты книги, особенно в главах 2–6, вообще никак не связаны с природой выборочного пространства $\mathcal{X}$, поскольку там рассматриваются не сами наблюдения, а функции от них со значениями в $\mathbb{R}$ или $\mathbb{R}^m$, $m > 1$.

Если заранее известно, что $\mathbf{P}$ сосредоточено на части $B \in \mathfrak{B}_{\mathcal{X}}$ пространства $\mathcal{X}$, то может оказаться удобным под $\mathcal{X}$ понимать B, а под $\mathfrak{B}_{\mathcal{X}}$ — сужение σ-алгебры $\mathfrak{B}_{\mathcal{X}}$ на B.

Рассмотрим n независимых повторений эксперимента G (см. [17, § 1.3]) и обозначим через $\mathrm{x}_1, \ldots, \mathrm{x}_n$ совокупность полученных наблюдений. Вектор

$$X_n = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$$

называется *выборкой объема n из совокупности с распределением* $\mathbf{P}$. Иногда используются укороченные или более полные варианты этого термина: *выборка из распределения* $\mathbf{P}$ или *простая выборка объема n из генеральной совокупности с распределением* $\mathbf{P}$.

Символически соотношение «X_n есть выборка из распределения $\mathbf{P}$» будет записываться с помощью знака $\Subset$ следующим образом:

$$X_n \Subset \mathbf{P}. \tag{1}$$

Такого рода запись будет использоваться нами и для других случайных величин. Например, соотношение

$$\xi \Subset \mathbf{P} \tag{2}$$

будет означать, что ξ имеет распределение $\mathbf{P}$. Такое использование символа $\Subset$ находится в соответствии с (1), поскольку последнее определено для любого n, в частности для $n = 1$.

Если ξ и η есть две случайные величины (вообще говоря, заданные на разных пространствах) с одинаковыми распределениями, то этот факт будем обозначать $\xi \underset{d}{=} \eta$, так что если X_n и Y_n есть две выборки одинакового объема из распределения $\mathbf{P}$, то можем писать $X_n \underset{d}{=} Y_n$.

В правых частях (1), (2) вместо распределения $\mathbf{P}$ иногда может стоять функция распределения, соответствующая $\mathbf{P}$. Так что если $F(x) = \mathbf{P}((-\infty, x))$, то запись

$$X_n \Subset F$$

будет тождественна с (1).

Само понятие «выборки из генеральной совокупности» встречается также при рассмотрении простейших вероятностных моделей, связанных с извлечением шаров из урны в классическом определении вероятности (см. [17, § 1.2]). Следует отметить, что введенное выше определение выборки находится в полном согласии с этим понятием, введенным ранее, и по существу совпадает с ним. Если x_i (или случайная величина ξ) могут принимать лишь s значений $a_1, \ldots, a_s$ и вероятности этих значений рациональны, т. е.

$$\mathbf{P}(\xi = a_j) = \frac{N_j}{N}, \quad \sum_{j=1}^{s} N_j = N,$$

то выборку X_n можно представить как результат «выборки с возвращением» (в смысле [17, § 3.2]) из урны с N шарами, из них на N_1 шарах написано a_1, на N_2 шарах написано a_2 и т. д.

Как математический объект выборка $X = X_n$ (индекс n мы часто будем опускать) есть не что иное, как случайная величина $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ со значениями в «n-мерном» пространстве $\mathcal{X}^n = \mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}$ и распределением,

определяемым для $B = B_1 \times B_2 \times \cdots \times B_n$, $B_j \in \mathfrak{B}_{\mathcal{X}}$ равенствами

$$\mathbf{P}(X \in B) = \mathbf{P}(\mathrm{x}_1 \in B_1, \ldots, \mathrm{x}_n \in B_n) = \prod_{i=1}^{n} \mathbf{P}(\mathrm{x}_i \in B_i). \tag{3}$$

Другими словами, распределение $\mathbf{P}$ на $\mathcal{X}^n$ есть n-кратное прямое произведение заданных «одномерных» распределений.

Относительно обозначений для распределения $\mathbf{P}$ и других будем придерживаться следующих соглашений, которые мы отчасти уже использовали в (3) и которые нигде не приведут нас к недоразумениям.

1. Мы будем использовать один и тот же символ (в частности, $\mathbf{P}$) для распределений в $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ и для прямого произведения этих распределений в $\left(\mathcal{X}^n, \mathfrak{B}_{\mathcal{X}}^n\right)$ (см. (3)), где $\mathfrak{B}_{\mathcal{X}}^n$ — σ-алгебра борелевских множеств в $\mathcal{X}^n$. Различие будет определяться лишь аргументом функции $\mathbf{P}$.

2. Вероятность попадания величины X в множество B, скажем, из $\mathfrak{B}_{\mathcal{X}}^n$, нам иногда будет удобно обозначать через $\mathbf{P}(B)$, а иногда через $\mathbf{P}(X \in B)$. Это одно и то же в силу того, что $\mathcal{X}^n$ есть выборочное пространство для X.

3. Наконец, символ $\mathbf{P}$ будем использовать для обозначения общего понятия вероятности (т. е. для вероятности, относящейся к каким-то другим случайным величинам без конкретизации вероятностного пространства).

В силу (3) можем рассматривать выборку X как элементарное событие в выборочном вероятностном пространстве $\left(\mathcal{X}, \mathfrak{B}_{\mathcal{X}}^n, \mathbf{P}\right)$ (см. [17, § 3.5]). Отметим, что относительно выборки X будем допускать двойственное толкование этого обозначения и объекта: как случайной величины и как вектора реальных числовых данных, полученных в фактически осуществленных экспериментах. Как показывает опыт, такое двойное толкование вполне приемлемо и не приводит к недоразумениям, хотя и допускает одновременное существование записей вида $\mathbf{P}(\mathrm{x}_1 < t) = F(t)$ и вида $\mathrm{x}_1 = 0{,}74$, $\mathrm{x}_2 = 0{,}83$ и т. д.

Отметим также, что координаты x_i выборки X мы обозначаем «прямыми» буквами x, оставляя «наклонные» буквы x для обозначения переменных величин. Вектор $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $x_i \in \mathcal{X}$, мы будем обозначать полужирными буквами $\boldsymbol{x} = (x_1, \ldots, x_n)$.

Выборка является основным исходным объектом в задачах математической статистики. Однако на практике ее элементы $\mathrm{x}_1, \mathrm{x}_2, \ldots$ далеко не всегда являются независимыми. В своих рассмотрениях мы также не будем исключать такой возможности. Чтобы при этом не делать дополнительных оговорок, мы в случае зависимых наблюдений будем считать, что имеем дело с выборкой объема $n = 1$, а наблюдения представляют собой координаты вектора x_1 (ведь природа пространства $\mathcal{X}$ произвольна).

В следующих рассмотрениях нам часто придется иметь дело с выборками X_n неограниченно возрастающего объема n. В таких случаях удобно предполагать, что задана выборка $X_\infty = (\mathrm{x}_1, \mathrm{x}_2, \ldots)$ бесконечного объема, а $X = X_n$ есть совокупность ее первых n координат. Под выборкой бесконечного объема X_∞ мы будем понимать элемент выборочного вероятностного пространства $\left(\mathcal{X}^\infty, \mathfrak{B}_{\mathcal{X}}^\infty, \mathbf{P}\right)$, где $\mathcal{X}^\infty$ — пространство последовательностей $(x_1, x_2, \ldots)$; σ-алгебра $\mathfrak{B}_{\mathcal{X}}^\infty$ порождена множествами $\bigcap\limits_{i \leqslant N} \{x_i \in B_i\}$, $B_i \in \mathfrak{B}_{\mathcal{X}}$, $N = 1, 2, \ldots$; распределение $\mathbf{P}$ обладает свойством (3). По теореме Колмогорова [17] такое распределение всегда существует. Стало быть, предположение

о существовании выборки X_∞ бесконечного объема никак не ограничивает общность.

Саму бесконечную последовательность (бесконечную выборку) X_∞ в рассмотрениях, носящих теоретико-вероятностный характер, можно трактовать как элементарное событие (ср. с [17]).

В тех случаях, когда нам нужно будет понимать X_n как подвектор X_∞, мы будем писать

$$X_n = [X_\infty]_n,$$

где $[\cdot]_n$ — оператор проектирования из $\mathcal{X}^\infty$ в $\mathcal{X}^n$, определенный очевидным образом. В соответствии с предыдущим запись

$$X_\infty \Subset \mathbf{P}$$

будет означать, что X_∞ есть выборка бесконечного объема из распределения $\mathbf{P}$.

Если возникает необходимость особо отметить тот факт, что речь идет о распределении в $\left(\mathcal{X}^\infty, \mathfrak{B}_\mathcal{X}^\infty\right)$ (или $\left(\mathcal{X}^n, \mathfrak{B}_\mathcal{X}^n\right)$ при $n<\infty$), а не в $(\mathcal{X}, \mathfrak{B}_\mathcal{X})$, то будем использовать также обозначение $\mathbf{P}^\infty$ ($\mathbf{P}^n$). Сохранение верхних индексов ∞ и n во всем тексте привело бы к громоздким обозначениям.

§ 2. Эмпирическое распределение (одномерный случай)

Пусть дана выборка $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n) \in \mathbf{P}$, $\mathrm{x}_i \in \mathcal{X} = \mathbb{R}$. Рассмотрим вещественную прямую $\mathbb{R}$ с σ-алгеброй борелевских множеств $\mathfrak{B}$ и дискретное распределение $\mathbf{P}_n^*$ на $(\mathbb{R}, \mathfrak{B})$, сосредоточенное в точках $\mathrm{x}_1, \ldots, \mathrm{x}_n$, для которого вероятность значения x_i полагается равной $1/n$. Другими словами, для любого $B \in \mathfrak{B}$ по определению

$$\mathbf{P}_n^*(B) = \frac{\nu(B)}{n}, \tag{1}$$

где $\nu(B)$ — число элементов выборки X, попавших во множество B. Распределение $\mathbf{P}_n^*$ называется *эмпирическим распределением*, построенным по выборке X (или соответствующим выборке X). Его можно представить также в следующем виде. Пусть $\mathbf{I}_x(B)$ — распределение, сосредоточенное в точке x:

$$\mathbf{I}_x(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B; \end{cases}$$

тогда, очевидно, $\nu(B) = \sum_{i=1}^n \mathbf{I}_{\mathrm{x}_i}(B)$,

$$\mathbf{P}_n^*(B) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\mathrm{x}_i}(B). \tag{2}$$

Ясно, что $\mathbf{P}_n^*(B)$ для любого борелевского B как функция от выборки есть случайная величина. Таким образом, мы имеем здесь дело со случайной функцией множеств, или со случайным распределением.

Предположим теперь, что $X_\infty \Subset \mathbf{P}$, $X_n = [X_\infty]_n$ и $n \to \infty$. Мы получим тогда последовательность эмпирических распределений $\mathbf{P}_n^*$. Замечательный

факт состоит в том, что эта последовательность неограниченно сближается с исходным распределением $\mathbf{P}$ наблюдаемой случайной величины. Этот факт имеет принципиальное значение для всего последующего изложения, поскольку он показывает, что неизвестное распределение $\mathbf{P}$ может быть сколь угодно точно восстановлено по выборке достаточно большого объема.

Теорема 1. *Пусть* $B \in \mathfrak{B}$ *и* $X_n = [X_\infty]_n \in \mathbf{P}$. *Тогда при* $n \to \infty$

$$\mathbf{P}_n^*(B) \underset{\text{п.н.}}{\longrightarrow} \mathbf{P}(B).$$

Сходимость с вероятностью 1 здесь имеется в виду относительно распределения $\mathbf{P} = \mathbf{P}^\infty$ в $(\mathbb{R}^\infty, \mathfrak{B}^\infty, \mathbf{P})$. Предположение $X_n = [X_\infty]$ нужно нам для того, чтобы случайные величины $\mathbf{P}_n^*(B)$ были заданы на одном вероятностном пространстве.

Доказательство. Обратимся к определению (2) и заметим, что $\mathbf{I}_{\mathrm{x}_i}(B)$ есть независимые одинаково распределенные случайные величины, $\mathbf{E}\mathbf{I}_{\mathrm{x}_i}(B) = \mathbf{P}(\mathbf{I}_{\mathrm{x}_i}(B) = 1) = \mathbf{P}(\mathrm{x}_i \in B) = \mathbf{P}(B)$. Так как $\mathbf{P}_n^*(B)$ есть среднее арифметическое этих величин, то остается воспользоваться усиленным законом больших чисел. $\triangleleft$

Теорема 1 устанавливает сходимость $\mathbf{P}_n^*(B)$ и $\mathbf{P}(B)$ в каждой «точке» B. Однако имеет место и более сильное утверждение о том, что такая сходимость в известном смысле равномерна по B.

Обозначим через $\mathfrak{I}$ совокупность множеств B, являющихся полуинтервалами вида $[a, b)$ с конечными или бесконечными концами, и предположим опять, что $X_n = [X_\infty]_n$.

Теорема 2 (Гливенко–Кантелли).

$$\sup_{B \in \mathfrak{I}} \left|\mathbf{P}_n^*(B) - \mathbf{P}(B)\right| \underset{\text{п.н.}}{\longrightarrow} 0.$$

Собственно, с именами Гливенко и Кантелли связано несколько иное утверждение, относящееся к важному понятию *эмпирической функции распределения*. По определению это есть функция распределения, соответствующая $\mathbf{P}_n^*$. Другими словами, *эмпирической функцией распределения* $F_n^*(x)$ называется функция

$$F_n^*(x) = \mathbf{P}_n^*((-\infty, x)).$$

Величина $nF_n^*(x)$ равна числу элементов выборки, которые меньше, чем x. В реальных условиях для построения $F_n^*(x)$ часто используют следующую процедуру. Элементы выборки $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ упорядочивают по возрастанию, т. е. образуют из нее последовательность

$$\mathrm{x}_{(1)} \leqslant \mathrm{x}_{(2)} \leqslant \ldots \leqslant \mathrm{x}_{(n)},$$

которая называется *вариационным рядом*. Тогда можем положить

$$F_n^*(x) = \frac{k}{n} \quad \text{при} \quad x \in (\mathrm{x}_{(k)}, \mathrm{x}_{(k+1)}],$$

где k пробегает значения от 0 до n, $\mathrm{x}_{(0)} = -\infty$, $\mathrm{x}_{(n+1)} = \infty$. Очевидно, $F_n^*(x)$ представляет из себя ступенчатую функцию, имеющую скачки величиной $1/n$ в точках x_i, если все x_i различны.

Пусть $F(x) = \mathbf{P}((-\infty, x))$ есть функция распределения ξ (или, что то же, x_1) и $X_n = [X_\infty]_n$. Теорема Гливенко–Кантелли состоит в следующем.

Теорема 2А. *При* $n \to \infty$

$$\sup_x |F_n^*(x) - F(x)| \xrightarrow[\text{п.н.}]{} 0.$$

Ниже индекс n у обозначений F_n^* мы будем опускать и писать просто F^*.

Доказательство теоремы 2А. Предположим сначала для простоты, что функция F непрерывна. Пусть $\varepsilon > 0$ — произвольно малое заданное число такое, что число $N = 1/\varepsilon$ целое. Так как F непрерывна, то можно указать числа $z_0 = -\infty$, $z_1, \ldots, z_{N-1}, z_N = \infty$ такие, что

$$F(z_0) = 0,\ F(z_1) = \varepsilon = \frac{1}{N},\ \ldots,\ F(z_k) = k\varepsilon = \frac{k}{N},\ \ldots,\ F(z_n) = 1.$$

При $z \in [z_k, z_{k+1})$ справедливы соотношения

$$\begin{aligned} F^*(z) - F(z) &\leqslant F^*(z_{k+1}) - F(z_k) = F^*(z_{k+1}) - F(z_{k+1}) + \varepsilon, \\ F^*(z) - F(z) &\geqslant F^*(z_k) - F(z_{k+1}) = F^*(z_k) - F(z_k) - \varepsilon. \end{aligned} \tag{3}$$

Обозначим через A_k множество элементарных событий $\omega = X_\infty$, на которых $F^*(z_k) \to F(z_k)$. По теореме 1 $\mathbf{P}(A_k) = 1$. Стало быть, для каждого $\omega \in A = \bigcup_{k=0}^{N} A_k$ найдется $n(\omega)$ такое, что при всех $n \geqslant n(\omega)$ будет выполняться

$$|F^*(z_k) - F(z_k)| < \varepsilon, \quad k = 0, 1, \ldots, N. \tag{4}$$

Но вместе с (3) эти неравенства влекут за собой

$$\sup_z |F^*(z) - F(z)| < 2\varepsilon. \tag{5}$$

Итак, это соотношение имеет место для произвольного $\varepsilon > 0$, всех $\omega \in A$ и всех достаточно больших $n \geqslant n(\omega)$. Так как $\mathbf{P}(A) = 1$, то теорема для непрерывной функции F доказана.

Для произвольной функции $F(x)$ доказательство происходит совершенно аналогично. Надо лишь воспользоваться следующим обстоятельством: для любой $F(x)$ существует конечное число точек $-\infty = z_0 < z_1 < \cdots < z_{N-1} < z_N = \infty$ таких, что

$$F(z_{k+1}) - F(z_k + 0) \leqslant \varepsilon, \quad k = 0, 1, \ldots, N-1 \tag{6}$$

(для определенности можно считать, что множество $\{z_j\}$ содержит все точки скачков F, по величине больших, чем, например, $\varepsilon/2$). Совершенно аналогично (3) получаем, что при $z \in (z_k, z_{k+1}]$

$$\begin{aligned} F^*(z) - F(z) &\leqslant F^*(z_{k+1}) - F(z_{k+1}) + \varepsilon, \\ F^*(z) - F(z) &\geqslant F^*(z_k + 0) - F(z_k + 0) - \varepsilon. \end{aligned} \tag{7}$$

К множествам A_k, которые определяются как прежде, добавим множества A_k^+, $k = 0, 1, \ldots, N$, на которых $F^*(z_k + 0) \to F(z_k + 0)$. Тогда по теореме 1 $\mathbf{P}(A_k) = \mathbf{P}(A_k^+) = 1$, и на множестве $A = \bigcap\limits_{k=0}^{N} A_k A_k^+$, $\mathbf{P}(A) = 1$, при достаточно больших $n \geqslant n(\omega)$ будет справедливо (4), а также неравенства

$$|F^*(z_k + 0) - F(z_k + 0)| < \varepsilon, \quad k = 0, 1, \ldots, N.$$

Вместе с (7) эти неравенства влекут за собой (5). $\triangleleft$

Теорема 2А является частным случаем теоремы 2, так как множества $(-\infty, x)$ принадлежат $\mathfrak{I}$; с другой стороны, теорему 2 нетрудно получить в качестве следствия теоремы 2А, поскольку для $B = [a, b)$

$$\bigl|\mathbf{P}_n^*(B) - \mathbf{P}(B)\bigr| \leqslant \bigl|F_n^*(b) - F(b)\bigr| + \bigl|F_n^*(a) - F(a)\bigr|,$$

и, стало быть,

$$\sup_{B \in \mathfrak{I}} \bigl|\mathbf{P}_n^*(B) - \mathbf{P}(B)\bigr| \leqslant \sup_{a,b} \bigl[\bigl|F_n^*(b) - F(b)\bigr| + \bigl|F_n^*(a) - F(a)\bigr|\bigr] \underset{\text{п.н.}}{\longrightarrow} 0.$$

Замечание 1. Нетрудно видеть, что такого же рода рассуждения позволяют нам в качестве совокупности множеств $\mathfrak{I}$ в теореме 2 брать совокупности интервалов (a, b), отрезков $[a, b]$ и их конечных (в числе не более некоторого N) объединений.

С другой стороны, если в качестве $\mathfrak{I}$ в теореме 2 взять достаточно богатый класс множеств, то утверждение теоремы перестает быть верным. Например, если $\mathfrak{I}$ содержит объединения любого конечного числа интервалов, то множество

$$B_n = \bigcup_{k=1}^{n} \left(\mathrm{x}_k - \frac{1}{n^2}, \mathrm{x}_k + \frac{1}{n^2}\right) \in \mathfrak{I}, \quad \mathbf{P}_n^*(B_n) = 1,$$

а для равномерного на $[0, 1]$ распределения $\mathbf{P}(B_n) \leqslant 2/n$, так что

$$\sup_{B \in \mathfrak{I}} \bigl|\mathbf{P}_n^*(B) - \mathbf{P}(B)\bigr| \geqslant \mathbf{P}_n^*(B_n) - \mathbf{P}(B_n) \to 1.$$

В заключение этого параграфа отметим, что представление (2) позволяет получить для $\mathbf{P}_n^*$ и более точные теоремы об асимптотическом поведении, чем теоремы типа Гливенко–Кантелли (эти результаты будут представлены в § 4, 6). Чтобы проиллюстрировать возможности, которые здесь есть, напомним, что $\sum\limits_{i=1}^{n} \mathbf{I}_{\mathrm{x}_i}(B)$ в (2) есть сумма независимых одинаково распределенных случайных величин в схеме Бернулли,

$$\mathbf{E}\mathbf{I}_{\mathrm{x}_i}(B) = \mathbf{P}(\mathbf{I}_{\mathrm{x}_i}(B) = 1) = \mathbf{P}(B),$$

$$\mathbf{E}\mathbf{I}_{\mathrm{x}_i}^2(B) = \mathbf{P}(B), \quad \mathbf{D}\mathbf{I}_{\mathrm{x}_i}(B) = \mathbf{P}(B)(1 - \mathbf{P}(B)).$$

Поэтому из центральной предельной теоремы немедленно вытекает следующее утверждение.

Теорема 3. $\mathbf{P}_n^*(B)$ *представимо в виде*

$$\mathbf{P}_n^*(B) = \mathbf{P}(B) + \frac{\zeta_n(B)}{\sqrt{n}}, \tag{8}$$

где распределение $\zeta_n(B) = \dfrac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} (\mathbf{I}_{\mathrm{x}_i}(B) - \mathbf{P}(B))$ *сходится к нормальному распределению с параметрами* $(0, \mathbf{P}(B)(1 - \mathbf{P}(B)))$.

Дальнейшее изучение $\mathbf{P}_n^*(B)$ в этом направлении содержится в § 6. Более точные теоремы о сходимости с вероятностью 1 см. в § 4.

§ 3. Выборочные характеристики. Основные типы статистик

1. Примеры выборочных характеристик. *Выборочными характеристиками* называют обычно различные измеримые функционалы от эмпирического распределения или, другими словами, функции от выборки, которые предполагаются измеримыми. Наиболее простые среди них — выборочные (или эмпирические) моменты. *Выборочным моментом порядка* k называется значение

$$a_k^* = a_k^*(X) = \int x^k \, dF_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{x}_i^k.$$

Выборочный центральный момент порядка k равен

$$a_k^{*\circ} = a_k^{*\circ}(X) = \int \left(x - a_1^*\right)^k dF_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} \left(\mathrm{x}_i - a_1^*\right)^k.$$

Для выборочных моментов a_1^* и $a_2^{*\circ}$ в литературе распространены специальные обозначения $\overline{\mathrm{x}}$ и S^2:

$$\overline{\mathrm{x}} = a_1^* = \frac{1}{n} \sum_{i=1}^{n} \mathrm{x}_i, \quad S^2 = a_2^{*\circ} = \frac{1}{n} \sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^2.$$

В статистических задачах используются самые различные выборочные характеристики. Например, *выборочная медиана* ζ^* — это среднее значение вариационного ряда, т. е. значение $\zeta^* = \mathrm{x}_{(m)}$, если $n = 2m - 1$ (нечетно), и $\zeta^* = (\mathrm{x}_{(m)} + \mathrm{x}_{(m+1)})/2$, если $n = 2m$ (четно). Напомним, что медианой ζ непрерывного распределения $\mathbf{P}$ называется решение уравнения $F(\zeta) = 1/2$.

Более общим понятием является понятие *квантили* ζ_p порядка p. Это число, для которого $F(\zeta_p) = p$. Так что медиана есть квантиль порядка $1/2$. Если F имеет точки разрыва (дискретную компоненту), то это определение теряет смысл. Поэтому в общем случае будем пользоваться следующим определением.

Квантилью ζ_p порядка p распределения $\mathbf{P}$ называется число

$$\zeta_p = \sup \{x : F(x) \leqslant p\}.$$

Как функция от p квантиль ζ_p есть не что иное, как функция $F^{-1}(p)$, обратная к $F(x)$.

Такое определение ζ_p (или $F^{-1}(p)$), в отличие от предыдущего, очевидно, имеет смысл для любых $F(x)$.

Ясно, что наряду с выборочными медианами можем рассматривать *выборочные квантили* ζ_p^* *порядка* p, равные по определению значению $\mathrm{x}_{(l)}$, где $l = [np] + 1$, $\mathrm{x}_{(k)}$ — члены вариационного ряда для выборки X, $k = 1, \ldots, n$. Для $p = 1/2$ мы сохраним определение $\zeta^* = \zeta_{1/2}^*$, данное выше (оно совпадает с приведенным лишь при нечетных n).

2. Два типа статистик. Пусть дана измеримая функция S от n аргументов. Выборочная характеристика $S(X) = S(\mathrm{x}_1, \ldots, \mathrm{x}_n)$ часто называется также *статистикой*. Из сказанного выше ясно, что любая статистика есть случайная величина. Ее распределение полностью определяется распределением $\mathbf{P}(B) = \mathbf{P}(\mathrm{x}_1 \in B)$ (напомним, что $S(X)$ можно рассматривать как случайную величину, заданную на $\left(\mathcal{X}^n, \mathfrak{B}_{\mathcal{X}}^n, \mathbf{P}\right)$, где $\mathbf{P}$ — n-кратное прямое произведение одномерных распределений x_1).

Мы выделим здесь два класса статистик, которые часто будут встречаться в дальнейшем. Они будут построены с помощью следующих двух видов функционалов $G(F)$ от функций распределения F.

I. Функционалы вида

$$G(F) = h\left(\int g(x)\, dF(x)\right),$$

где g — заданная борелевская функция, h — функция, непрерывная в точке $a = \int g(x)\, dF_0(x)$, где F_0 такова, что $X \Subset F_0$.

II. Функционалы $G(F)$, непрерывные в «точке» F_0 в равномерной метрике: $G(F^{(n)}) \to G(F_0)$, если $\sup\limits_x |F^{(n)}(x) - F_0(x)| \to 0$, носители*) распределений $F^{(n)}$ принадлежат носителю F_0. Здесь F_0 опять есть функция, для которой $X \Subset F_0$.

Соответствующие классы статистик мы определим с помощью равенства

$$S(X) = G\left(F_n^*\right),$$

где F_n^* — эмпирическая функция распределения. Мы получим тогда:

I. *Класс статистик* I *типа*, представимых в виде

$$S(X) = h\left(\int g(x)\, dF_n^*(x)\right) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i)\right).$$

Очевидно, все выборочные моменты имеют вид аддитивных статистик $\dfrac{1}{n}\sum\limits_{i=1}^{n} g(\mathrm{x}_i)$ и относятся к статистикам I типа.

*) Носитель N_F распределения $\mathbf{P}$ с функцией распределения F есть множество, для которого $\mathbf{P}(N_F) = 1$.

II. Класс статистик, которые будем называть *статистиками* II *типа* или *статистиками, непрерывными в точке* F_0.

Ясно, что, например, выборочная медиана будет непрерывной статистикой в точке F, если существует медиана ζ, $F(\zeta) = 1/2$, и F непрерывна и строго возрастает в точке ζ.

Принадлежность функционалов одному из названных классов, конечно, не является альтернативной. Функционал $G(F)$ может не принадлежать ни одному из этих классов или принадлежать обоим сразу. Например, если G есть функционал I типа, носитель F сосредоточен на отрезке $[a, b]$ ($F(a) = 0$, $F(b) = 1$), а функция g имеет ограниченную вариацию на $[a, b]$, то G будет одновременно функционалом II типа, так как в этом случае функционал

$$\int g(x)\,dF(x) = g(b) - \int_a^b F(x)\,dg(x)$$

непрерывен относительно F в равномерной метрике. Сказанное означает, что статистики I типа $\bar{\mathrm{x}}$ и S^2 будут также статистиками II типа, если $X \Subset \mathbf{P}$ и $\mathbf{P}$ сосредоточено на конечном интервале.

Статистики вида $G(F_n^*)$ иногда называют также *статистическими функциями*. Их систематическое исследование было начато фон Мизесом (см. [75, 76, 115]).

Теоремы 2.1 и 2.2 можем дополнить здесь следующим утверждением о п. н. сходимости выборочных характеристик.

Теорема 1. *Пусть, как и прежде,* $X_n = [X_\infty]_n \Subset F$. *Тогда если* $S(X) = G(F_n^*)$ *есть статистика* I *или* II *типа, то при* $n \to \infty$

$$G(F_n^*) \underset{\text{п.н.}}{\longrightarrow} G(F).$$

Здесь предполагается, конечно, что значение $G(F)$ существует.

Таким образом, выборки большого объема позволяют оценивать не только само распределение $\mathbf{P}$, но и функционалы от этого распределения — по крайней мере те, которые принадлежат одному из названных в теореме классов.

Доказательство утверждения для обоих классов статистик почти очевидно. Пусть, например, $G(F) = h\left(\int g(x)\,dF(x)\right)$. Тогда

$$S = S(X) = \int g(x)\,dF_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i)$$

есть сумма независимых случайных величин с математическим ожиданием

$$\mathbf{E}g(\mathrm{x}_1) = \int g(x)\,dF(x).$$

Поэтому по усиленному закону больших чисел $S \underset{\text{п.н.}}{\longrightarrow} \mathbf{E}g(\mathrm{x}_1)$. Пусть теперь $A = \{X_\infty : S(X) \to \mathbf{E}g(\mathrm{x}_1)\}$. Тогда $\mathbf{P}(A) = 1$, и если $X_\infty \in A$, то $S(X) \to \to \mathbf{E}g(\mathrm{x}_1)$, $h(S(X)) \to h(\mathbf{E}g(\mathrm{x}_1))$. Другими словами, на множестве A

$$G(F_n^*) \to G(F).$$

Утверждение теоремы для функционалов второго типа есть прямое следствие теоремы Гливенко–Кантелли. ◁

Из теоремы вытекает, что абсолютные и центральные выборочные моменты п. н. сходятся при $n \to \infty$ к соответствующим моментам распределения $\mathbf{P}$:

$$a_k^* = a_k^*(X) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i^k \underset{\text{п.н.}}{\longrightarrow} \mathbf{E}\mathrm{x}_1^k,$$

$$a_k^{*\circ} = a_k^{*\circ}(X) = \frac{1}{n}\sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^k \underset{\text{п.н.}}{\longrightarrow} \mathbf{E}(\mathrm{x}_1 - \mathbf{E}\mathrm{x}_1)^k.$$

В частности,

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^2 = \frac{1}{n}\sum \mathrm{x}_i^2 - \overline{\mathrm{x}}^2 \to \mathbf{D}\mathrm{x}_1.$$

Таким образом, мы установили важный факт, имеющий для нас принципиальное значение: эмпирическое распределение и широкий класс функционалов от него с ростом объема выборки неограниченно сближаются с соответствующими «теоретическими» значениями.

3. *L*-статистики. Так называются статистики, являющиеся линейными комбинациями членов вариационного ряда или функций от них:

$$S_n(X) = \frac{1}{n}\sum \varphi_{n,i}\, g(\mathrm{x}_{(i)}).$$

Здесь функция g и коэффициенты $\varphi_{n,i}$ заданы. Очевидно, что все выборочные квантили являются L-статистиками.

Если положить $\varphi_n(t) = \varphi_{n,i}$ при $t \in \left(\dfrac{i-1}{n}, \dfrac{i}{n}\right)$, то статистику $S_n(X)$ можно представить в виде

$$S_n(X) = \int_0^1 \varphi_n(t) g\big(F_n^{-1}(t)\big)\, dt, \tag{1}$$

где F_n^{-1} — обратная функция к эмпирической функции распределения F_n^*.

Если $\varphi_n(t)$ сближается при $n \to \infty$ в подходящем смысле с функцией $\varphi(t)$, то естественно ожидать, что $S_n(X)$ будет близко к статистике $G\big(F_n^*\big)$, где

$$G(F) = \int_0^1 \varphi(t) g(F^{-1}(t))\, dt = \int \varphi(F(u)) g(u)\, dF(u). \tag{2}$$

Очевидно, что $G\big(F_n^*\big)$ есть L-статистика при $\varphi_{n,i} = \varphi\left(\dfrac{i-1}{n}\right)$. Считая на время, что $F_n^*(u)$ непрерывна справа (это предположение ничего не меняет

в последующих доказательствах), получим несколько более удобную запись $\varphi_{n,i} = \varphi\left(\frac{i}{n}\right)$. Представление (2) доставляет весьма широкий «регулярный» выбор для коэффициентов $\varphi_{n,i}$ в определении $S_n(X)$. Поэтому, чтобы избежать технических осложнений, ограничимся в дальнейшем рассмотрением L-статистик вида $G\left(F_n^*\right)$, построенных на основании функционала G, определенного в (2).

Отметим, что выборочная квантиль порядка p и крайние значения вариационного ряда $\mathrm{x}_{(1)}$ и $\mathrm{x}_{(n)}$ являются L-статистиками вида (1), для которых $\varphi_n(t)$ сходятся при $n \to \infty$ к δ-функциям Дирака. В дальнейшем будем предполагать, как правило, что $\varphi(t)$ в (2) является «обычной» (не обобщенной) функцией.

Так как функции φ и g в представлении (2) определены неоднозначно (с точностью до постоянных множителей, произведение которых равно единице), то можно фиксировать значение $\varphi_0 = \int\limits_0^1 \varphi(t)\,dt$.

Запись (2) показывает, что статистики I типа при $h(x) = x$ являются L-статистиками, а L-статистики при $\varphi \equiv 1$ являются статистиками I типа.

4. M-статистики. Пусть $\psi(x,\theta)$ — некоторая функция $\mathcal{X} \times \mathbb{R}^k \to \mathbb{R}$. Рассмотрим функционал $G(F)$, который определяется как точка максимума по θ функции $\int \psi(x,\theta)\mathbf{P}(dx)$.

$\widehat{M}$-статистикой (или *$\widehat{M}$-оценкой*) называется значение $\theta^* = G\left(F_n^*\right)$ или, что то же, точка θ^*, в которой достигается

$$\max \sum \psi(\mathrm{x}_i,\theta) = \max n \int \psi(x,\theta)\,dF_n^*(x). \tag{3}$$

Если точек максимумов несколько, то любая из них является $\widehat{M}$-оценкой.

Часто используется и несколько иное (более регулярное в известном смысле) определение. *M-статистикой* (*M-оценкой*) называется решение уравнений

$$\sum g(\mathrm{x}_i,\theta) = 0, \tag{4}$$

где g, как и ψ, есть некоторая заданная функция на $\mathcal{X} \times \mathbb{R}^k$. Кроме того, термин M-статистика будем использовать иногда и как собирательный, включающий в себя оба вида упомянутых выше $\widehat{M}$- и M-статистик.

Если функция ψ в (3) дифференцируема по θ, то очевидно, что $\widehat{M}$-статистика (в смысле определения (3)) будет M-статистикой (в смысле определения (4) при замене в (4) g производной ψ' по θ). При этом определение (4) может добавлять «ненужные» значения θ^*, в которых достигаются локальные экстремумы.

Более подробно M-статистики будут изучены в главе 2. Они играют важную роль в теории асимптотически оптимальных оценок.

5. О других статистиках. Имеют распространение и некоторые другие «собирательные» классы статистик такие, например, как U-статистики, построенные на основе функционалов

$$G(F) = \int h(x_1, \ldots, x_m)\mathbf{P}(dx_1)\ldots\mathbf{P}(dx_m), \quad 1 \leqslant m \leqslant n,$$

где h — симметрическая функция, называемая иногда ядром функционала. *U-статистикой* называется статистика

$$G(F_n^*) = n^{-m}\sum h(\mathrm{x}_{i_1}, \ldots, \mathrm{x}_{i_m}),$$

где суммирование ведется по всем $i_1, \ldots, i_m$ между 1 и n.

Очевидно, что выборочные моменты являются U-статистиками.

Почти все встречающиеся в приложениях статистики принадлежат одному из названных выше классов. С другой стороны, асимптотическое поведение распределений статистик для каждого из этих классов может быть изучено весьма полно. Результаты в этом направлении будут изложены в § 7, 8 и в главе 2.

§ 4. Многомерные выборки

1. Эмпирические распределения. Совершенно аналогичным образом выглядят построения эмпирического распределения и выборочных характеристик в многомерном случае, когда наблюдаемая случайная величина ξ, а вместе с ней и выборочные значения $\mathrm{x}_1, \ldots, \mathrm{x}_n$ есть векторы размерности $m > 1$: $\mathrm{x}_k = (\mathrm{x}_{k,1}\ldots, \mathrm{x}_{k,m})$. Здесь $\mathbf{P}(B) = \mathbf{P}(\xi \in B)$ есть распределение в $\mathcal{X} = \mathbb{R}^m$, а выборочным пространством здесь будет $\left(\mathcal{X}^n, \mathfrak{B}_{\mathcal{X}}^n, \mathbf{P}\right)$, где $\mathbf{P}$ есть n-кратное прямое произведение распределений $\mathbf{P}$ в $\left(\mathbb{R}^m, \mathfrak{B}_{\mathcal{X}} = \mathfrak{B}_R^m\right)$. Обозначение $X \Subset \mathbf{P}$ полностью сохраняет свой смысл.

Эмпирическое распределение $\mathbf{P}_n^*$ по выборке X строится, как и раньше, как дискретное распределение с массами величины $1/n$ в точках $\mathrm{x}_1, \ldots, \mathrm{x}_n$, так что

$$\mathbf{P}_n^*(B) = \frac{\nu(B)}{n} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_{\mathrm{x}_i}(B),$$

где $\nu(B)$ — число точек, попавших в множество B, $\mathbf{I}_{\mathrm{x}_i}$ — распределение, сосредоточенное в точке x_i.

Утверждение теоремы 1 о сходимости $\mathbf{P}_n^*(B) \underset{\text{п.н.}}{\longrightarrow} \mathbf{P}(B)$ здесь, очевидно, сохранится.

Обобщение теоремы Гливенко–Кантелли на многомерный случай связано с появлением качественно новых вопросов. Один из них — обобщения понятия интервалов на многомерный случай. Таких обобщений может быть несколько — это могут быть, например, прямоугольники, выпуклые множества и др.

Простейший вариант обобщения теоремы Гливенко–Кантелли таков.

Пусть $y = (y_1, \ldots, y_m)$ есть точка $\mathbb{R}^m$ и B_t есть угол с вершиной в точке $t = (t_1, \ldots, t_m)$:

$$B_t = \{y \in \mathbb{R}^m : y_k < t_k,\ k = 1, \ldots, m\}.$$

Функция

$$F_n^*(t) = \mathbf{P}_n^*(B_t)$$

называется *эмпирической функцией распределения.*

Теорема 1. *Пусть* $X_n = [X_\infty]_n$, $X_\infty \in F$. *Тогда*

$$\sup_t \left|F_n^*(t) - F(t)\right| \underset{\text{п.н.}}{\longrightarrow} 0$$

при $n \to \infty$.

2*. Более общие варианты теоремы Гливенко–Кантелли. Закон повторного логарифма. Одно из возможных обобщений теоремы типа Гливенко–Кантелли состоит в следующем. Пусть $\mathfrak{C}$ — класс всех выпуклых множеств в $\mathbb{R}^m$.

Теорема 2. *Пусть* $X_n = [X_\infty]_n$, $X_\infty \in \mathbf{P}$, *распределение* $\mathbf{P}$ *абсолютно непрерывно относительно меры Лебега в* $\mathbb{R}^m$. *Тогда*

$$\sup_{B \in \mathfrak{C}} \left|\mathbf{P}_n^*(B) - \mathbf{P}(B)\right| \underset{\text{п.н.}}{\longrightarrow} 0. \tag{1}$$

Другие возможные обобщения теоремы 1 можно получить с помощью утверждений приложения I.

Замечание 1. Требование абсолютной непрерывности распределения $\mathbf{P}$ относительно меры Лебега в теореме 2 существенно. На это указывает следующий пример. Пусть $\mathbf{P}$ есть равномерное распределение на единичной окружности (т. е. на границе круга) в $\mathbb{R}^2$. Построим замкнутый многоугольник B_X с вершинами в точках $x_1, \ldots, x_n$, лежащих на окружности. Это выпуклое множество. Однако $\mathbf{P}(B_X) = 0$, $\mathbf{P}_n^*(B_X) = 1$, и, следовательно, соотношение (1), где $\mathfrak{C}$ — класс выпуклых множеств, не верно.

Утверждения теорем типа Гливенко–Кантелли можно существенно уточнить — по крайней мере для простейших классов множеств. Например, для эмпирических функций распределений $F_n^*(t)$ (см. теорему 1) можно указать такую детерминированную последовательность $b_n \to 0$ при $n \to \infty$, для которой с вероятностью 1 (для почти всех «точек» X_∞)

$$\limsup_{n\to\infty} b_n^{-1} \sup_t \left|F_n^*(t) - F(t)\right| = 1.$$

Оказывается, что порядок малости b_n такой же, как $\sqrt{\dfrac{\ln\ln n}{n}}$.

Теорема 3 (закон повторного логарифма). *Если* $F(t)$ *непрерывна, то*

$$\mathbf{P}\left(\limsup_{n\to\infty} \sqrt{\frac{2n}{\ln\ln n}} \sup_t \left|F_n^*(t) - F(t)\right| = 1\right) = 1.$$

Теорема 3 тесно связана с нормальным приближением для $F_n^*(t)$ вида (2.8), которое в многомерном случае, очевидно, также имеет место.

Доказательство теорем 1, 2 отнесено нами в приложение I. Доказательство теоремы 3 см. в [52].

3. Выборочные характеристики. В многомерном случае, как и в одномерном, они представляют различные измеримые функции от выборки. Простейшие из них — выборочные моменты. Например, выборочные моменты первого порядка равны

$$a^*_{1,j} = a^*_{1,j}(X) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{x}_{k,j}, \quad j = 1, \ldots, m,$$

моменты второго порядка (обычные и центральные) —

$$a^*_{2,ij} = a^*_{2,ij}(X) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{x}_{k,i}\mathrm{x}_{k,j}, \quad i, j = 1, \ldots, m,$$

$$a^{*\circ}_{2,ij} \equiv S_{ij} = \frac{1}{n}\sum_{k=1}^{n} \bigl(\mathrm{x}_{k,i} - a^*_{1,i}\bigr)\bigl(\mathrm{x}_{k,j} - a^*_{1,j}\bigr)$$

и т. д. Как и в одномерном случае, легко убедиться с помощью усиленного закона больших чисел, что эти характеристики с вероятностью 1 сходятся к соответствующим «теоретическим» моментам. В частности, $S_{ij} \underset{\text{п.н.}}{\longrightarrow} \mathbf{E}(\mathrm{x}_{1,i} - \mathbf{E}\mathrm{x}_{1,i})(\mathrm{x}_{1,j} - \mathbf{E}\mathrm{x}_{1,j})$. Нетрудно убедиться (подробнее об этом см. в следующем параграфе), что этим же свойством обладают и выборочные коэффициенты корреляции

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \underset{\text{п.н.}}{\longrightarrow} \rho(\mathrm{x}_{1,i}, \mathrm{x}_{1,j}) = \frac{\mathbf{E}(\mathrm{x}_{1,i} - \mathbf{E}\mathrm{x}_{1,i})(\mathrm{x}_{1,j} - \mathbf{E}\mathrm{x}_{1,j})}{\sqrt{\mathbf{D}\mathrm{x}_{1,i}\mathbf{D}\mathrm{x}_{1,j}}}.$$

Отметим, что все типы статистик, введенные в § 3, за исключением L-статистик, допускают естественное обобщение на многомерный случай и случай произвольного выборочного пространства $\mathcal{X}$.

Для получения более точных теорем о распределении выборочных характеристик нам будут полезны так называемые теоремы непрерывности.

§ 5. Теоремы непрерывности

В дальнейшем нам понадобятся некоторые вспомогательные утверждения, которые мы будем часто использовать и которые можно было бы назвать *теоремами непрерывности.* Для удобства выделяем их в отдельный параграф. Одну теорему такого типа мы уже использовали — это теорема 3.1. Первая теорема непрерывности будет очень близка к этой теореме.

Теорема 1 (первая теорема непрерывности). *Пусть* $X = [X_\infty]_n \in \mathbf{P}$. *Тогда если* $S_n = S_n(X)$ *есть последовательность скалярных или векторных статистик таких, что* $S_n \underset{\text{п.н.}}{\longrightarrow} S_0$, *и* $H(s)$ — *функция, непрерывная почти всюду относительно распределения случайной величины* S_0 (*т. е.* $H(s)$ *непрерывна в каждой точке множества* B, $\mathbf{P}(S_0 \in B) = 1$), *то* $H(S_n(X)) \underset{\text{п.н.}}{\longrightarrow} H(S_0)$.

Если S_n *сходится к* S_0 *по вероятности* ($S_n \underset{P}{\longrightarrow} S_0$), *то при тех же прочих условиях* $H(S_n) \underset{P}{\longrightarrow} H(S_0)$.

Доказательство теоремы почти очевидно. Так как вероятности событий $A = \{X_\infty : S_n(X_\infty) \to S_0(X_\infty)\}$ и $C = \{X_\infty : S_0(X_\infty) \in B\}$ равны 1, то в силу равенства $\mathbf{P}(A \cap C) = \mathbf{P}(A) + \mathbf{P}(C) - \mathbf{P}(A \cup C)$ вероятность события $A \cap C$ (на котором $H(S_n(X_\infty)) \to H(S_0(X_\infty))$ также равна 1.

Чтобы упростить доказательство о сходимости по вероятности, предположим дополнительно, что $S_0 = \text{const}$ (нам только этот случай и понадобится). Для заданного $\varepsilon > 0$ найдется $\delta > 0$ такое, что событие $A_n = \{X_\infty : |S_n - S_0| < \delta\}$ влечет за собой $|H(S_n) - H(S_0)| < \varepsilon$ и, кроме того, $\mathbf{P}(A_n) > 1 - \varepsilon$ при всех достаточно больших n. Следовательно, для таких n имеем $1 - \varepsilon < \mathbf{P}(A_n) \leqslant \mathbf{P}(|H(S_n) - H(S_0)| < \varepsilon)$. $\triangleleft$

Прежде чем формулировать последующие теоремы, введем некоторые обозначения, которые будут удобны в дальнейшем.

Пусть дана последовательность случайных векторов $\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(s)})$ (не обязательно на одном вероятностном пространстве). Если распределения η_n слабо сходятся при $n \to \infty$ к распределению некоторой случайной величины η, то этот факт будем обозначать символом

$$\eta_n \Rightarrow \eta. \tag{1}$$

Здесь мы используем для случайных величин знак $\Rightarrow$ слабой сходимости распределений. Этот знак будет употребляться по-прежнему и для самих распределений, так что соотношение (1) эквивалентно тому, что

$$\mathbf{Q}_n \Rightarrow \mathbf{Q},$$

где $\mathbf{Q}_n$ и $\mathbf{Q}$ суть соответственно распределения η_n и η. Такое соглашение удобно и не приводит к недоразумениям.

Ясно, что из $\eta_n \xrightarrow[P]{} \eta$ или $\eta_n \xrightarrow[\text{п.н.}]{} \eta$ следует $\eta_n \Rightarrow \eta$ (ср. с [17, § 6.2]).

Таким образом, если речь идет о соотношении (соответствующем слабой сходимости) между объектами одной природы (между случайными величинами или между распределениями), будем использовать символ $\Rightarrow$. Наряду с этим нам было бы удобно иметь также символ для выражения «*распределения η_n слабо сходятся при $n \to \infty$ к $\mathbf{Q}$*». Это соотношение будем записывать в виде

$$\eta_n \Subset \mathbf{Q}, \tag{2}$$

так что символ $\Subset$ выражает тот же факт, что и $\Rightarrow$, но соединяет объекты разной природы так же, как это делает символ $\in$ в соотношении $\eta \in \mathbf{Q}$ (слева в (2) стоят случайные величины, а справа — распределение).

Пусть η_n и η есть случайные векторы из $\mathbb{R}^s$.

Теорема 2 (вторая теорема непрерывности). *Если $\eta_n \Rightarrow \eta$ и $H(t)$, $t \in \mathbb{R}^s$, есть непрерывная функция из $\mathbb{R}^s$ в $\mathbb{R}^k$, то $H(\eta_n) \Rightarrow H(\eta)$.*

Отметим, что эта теорема на самом деле верна и в более общей форме*): *Если $\eta_n \Rightarrow \eta$ и $H(t)$ непрерывна в точках множества $A \in \mathfrak{B}^s$, $\mathbf{P}(\eta \in A) = 1$, то $H(\eta_n) \Rightarrow H(\eta)$.*

*) Подробнее об этом см. в [7].

Доказательство теоремы 2. Пусть $\mathbf{Q}_n$ и $\mathbf{Q}$ есть соответственно распределения η_n и η. Слабая сходимость $\mathbf{Q}_n \Rightarrow \mathbf{Q}$ означает по определению, что для любой непрерывной и ограниченной функции $f\colon \mathbb{R}^s \to \mathbb{R}$ выполняется

$$\int f(y)\mathbf{Q}_n(dy) \to \int f(y)\mathbf{Q}(dy),$$

или, что то же самое,

$$\mathbf{E}f(\eta_n) \to \mathbf{E}f(\eta). \tag{3}$$

Аналогичное соотношение мы должны получить и для распределений $H(\eta_n)$ и $H(\eta)$. То есть мы должны установить, что для любой непрерывной ограниченной функции $g\colon \mathbb{R}^k \to \mathbb{R}$ справедливо $\mathbf{E}g(H(\eta_n)) \to \mathbf{E}g(H(\eta))$. Но это очевидным образом следует из (3), так как суперпозиция $\widetilde{g} \equiv g \circ H : \mathbb{R}^s \to \mathbb{R}$ непрерывна и ограничена. $\triangleleft$

Теорема 3 (третья теорема непрерывности). *Пусть* $\eta_n \Rightarrow \eta \in \mathbb{R}$, $H(t)$, $t \in \mathbb{R}$, *есть функция, дифференцируемая в точке* a. *Тогда если* $b_n \to 0$ *есть числовая последовательность, то*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n} \Rightarrow \eta H'(a). \tag{4}$$

Доказательство. Рассмотрим функцию

$$h(x) = \begin{cases} \dfrac{H(a+x) - H(a)}{x}, & x \neq 0, \\ H'(a), & x = 0, \end{cases}$$

которая будет непрерывной в точке $x = 0$. Так как $b_n\eta_n \Rightarrow 0$, то в силу первой теоремы непрерывности $h(b_n\eta_n) \Rightarrow h(0) = H'(a)$. Пользуясь второй теоремой непрерывности, получим

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n} = h(b_n\eta_n)\eta_n \Rightarrow H'(a)\eta. \quad \triangleleft$$

Мы приведем теперь два последовательных обобщения теоремы 3 на многомерный случай, которые будут нам полезны.

Теорема 3А. *Пусть* $\eta_n \equiv (\eta_n^{(1)}, \ldots, \eta_n^{(s)}) \Rightarrow \eta \equiv (\eta^{(1)}, \ldots, \eta^{(s)})$, $H(t)$ — *скалярная функция вектора* $t = (t_1, \ldots, t_s)$ *такая, что существует производная* $H'(t) \equiv \left(\dfrac{\partial H}{\partial t_1}, \ldots, \dfrac{\partial H}{\partial t_s}\right)$ *в точке* a. *Тогда при* $b_n \to 0$

$$\frac{H(a + b_n\eta_n) - H(a)}{B_n} \Rightarrow \eta(H'(a))^{\top} = \sum_{j=1}^{s} \frac{\partial H(a)}{\partial t_j}\eta^{(j)}. \tag{5}$$

Здесь индекс $\top$ *соответствует транспонированию.*

Если $\eta(H'(a))^\top = 0$ *с вероятностью* 1 (*например*, $H'(a) = 0$), *а матрица* $H''(t)$ *производных* $\dfrac{\partial^2 H(t)}{\partial t_i \partial t_j}$ *существует в точке* a, *то*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n^2} \Rightarrow \frac{1}{2}\eta H''(a)\eta^\top = \frac{1}{2}\sum_{i,j=1}^{s} \frac{\partial^2 H(a)}{\partial t_i \partial t_j}\eta^{(i)}\eta^{(j)}. \tag{6}$$

Пусть теперь $H(t)$ — векторная функция. Тогда, очевидно, предельное распределение для каждой компоненты H_j будет описываться теоремой 3А, а относительно совместного распределения будет справедлива

Теорема 3В. *Пусть* $\eta_n \Rightarrow \eta \in \mathbb{R}^s$, $H(t) \in \mathbb{R}^k$ — *векторная функция такая, что производные* H'_j, $j = 1, \ldots, k$, *удовлетворяют условиям теоремы* 3А. *Тогда*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n} \Rightarrow \eta(H'(a))^\top.$$

Если $\eta(H'(a))^\top = 0$ *с вероятностью* 1, *а матрицы* H''_j, $j = 1, \ldots, k$, *существуют в точке* a, *то*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n^2} \Rightarrow \frac{1}{2}\left(\eta H''_1(a)\eta^\top, \ldots, \eta H''_k(a)\eta^\top\right).$$

Доказательства этих утверждений по существу ничем не отличаются от доказательства теоремы 3, и поэтому предоставляем их читателю в качестве упражнений. Кроме того, предлагаем убедиться, что символ $\Rightarrow$ в (4)–(6) можно заменить на $\underset{\text{п.н.}}{\longrightarrow}$ или $\underset{P}{\longrightarrow}$, если соответственно выполняется $\eta_n \underset{\text{п.н.}}{\longrightarrow} \eta$ или $\eta_n \underset{P}{\longrightarrow} \eta$.

Содержание теорем 1–3 можно резюмировать следующим образом. Пусть $\rightsquigarrow$ означает один из символов $\underset{\text{п.н.}}{\longrightarrow}$, $\underset{P}{\longrightarrow}$, $\Rightarrow$. Тогда если H непрерывна, то из $\eta_n \rightsquigarrow \eta$ следует $H(\eta_n) \rightsquigarrow H(\eta)$.

Если H дифференцируема в точке a, $\eta_n \rightsquigarrow \eta$, то для $b_n \to 0$

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n} \rightsquigarrow H'(a)\eta. \tag{7}$$

Замечание 1. Нетрудно видеть, что если $a = a_n$ зависит от n так, что $a_n \equiv a_0 + o(1)$, и производные в теоремах 3, 3А, 3В непрерывны, то соотношение (7) сохранится в форме

$$\frac{H(a_n + b_n\eta_n) - H(a_n)}{b_n} \rightsquigarrow H'(a_0)\eta. \tag{8}$$

Для доказательства достаточно заметить, что левая часть (8) представима в виде $H'(\alpha_n)\eta_n$, где $\alpha_n = \theta a_n + (1 - \theta)(a_n + b_n\eta_n) \rightsquigarrow a_0$, $0 \leqslant \theta \leqslant 1$, и воспользоваться второй теоремой непрерывности.

То же самое замечание справедливо для многомерных аналогов этого утверждения в теоремах 3A, 3B.

Сформулированные теоремы касались сходимости почти наверное и сходимости распределений. Четвертая теорема непрерывности относится к сходимости интегралов.

Теорема 4 (теорема непрерывности для моментов). *Пусть* $\{\eta_n\}$ — *последовательность числовых случайных величин и* $\eta_n \Rightarrow \eta$ *при* $n \to \infty$. *Тогда если выполнено хотя бы одно из следующих условий:*

1) $\limsup\limits_{n\to\infty} \int\limits_N^\infty \mathbf{P}(|\eta_n| > x)\,dx \to 0$ при $N \to \infty$,

2) $\mathbf{P}(|\eta_n| > x) \leqslant \varphi(x)$, $\int\limits_0^\infty \varphi(x)\,dx < \infty$,

3) $\mathbf{E}|\eta_n|^{1+\alpha} < c < \infty$ *для некоторого* $\alpha > 0$,

то $\lim\limits_{n\to\infty} \mathbf{E}\eta_n = \mathbf{E}\eta$.

Отметим, что условие 1) означает равномерную по n сходимость к нулю $\int\limits_N^\infty \mathbf{P}(|\eta_n| > x)\,dx$ при $N \to \infty$.

Доказательство. Из обобщенного неравенства Чебышева

$$\mathbf{P}(|\eta_n| > x) \leqslant \frac{\mathbf{E}|\eta_n|^{1+\alpha}}{x^{1+\alpha}}$$

следует, что условие 3) влечет за собой условие 2). В свою очередь, условие 2) влечет за собой условие 1).

Пусть выполнено условие 1). Для простоты рассуждений предположим сначала, что $\eta_n \geqslant 0$. Тогда, интегрируя по частям, получим

$$\mathbf{E}\eta_n = -\int\limits_0^\infty x\,d\mathbf{P}(\eta_n \geqslant x) = \int\limits_0^\infty \mathbf{P}(\eta_n \geqslant x)\,dx.$$

Из этого представления, из сходимости $\mathbf{P}(\eta_n \geqslant x) \to \mathbf{P}(\eta \geqslant x)$ для почти всех x и из равномерной по n сходимости интеграла $\int\limits_0^\infty \mathbf{P}(\eta_n \geqslant x)\,dx$ следует законность предельного перехода под знаком интеграла, в силу которого

$$\lim_{n\to\infty} \mathbf{E}\eta_n = \lim_{n\to\infty} \int\limits_0^\infty \mathbf{P}(\eta_n \geqslant x)\,dx = \int\limits_0^\infty \mathbf{P}(\eta \geqslant x)\,dx = \mathbf{E}\eta.$$

В общем случае следует воспользоваться представлением $\eta_n = \eta_n^+ - \eta_n^-$, где $\eta_n^+ = \max(\eta_n, 0)$, $\eta_n^- = \max(-\eta_n, 0)$. $\triangleleft$

Отметим, что условие 1) можно рассматривать так же, как условие равномерной интегрируемости η_n, из которой немедленно следует сходимость $\mathbf{E}\eta_n \to \mathbf{E}\eta$ (см., например, [17, гл. 6; 72, § 6.2)].

§ 6.* Эмпирическая функция распределения как случайный процесс. Сходимость к броуновскому мосту

В этом параграфе будем предполагать известным понятие *случайного процесса* (скажем, в объеме [17]) и, в частности, определения и простейшие свойства *винеровского* и *пуассоновского процессов*.

1. Распределение процесса $nF_n^*(t)$. Ограничимся рассмотрением случая $\mathcal{X} = \mathbb{R}$. Пусть, как и прежде, $F_n^*(t) = \mathbf{P}_n^*((-\infty, t))$ есть эмпирическая функция распределения, соответствующая выборке $X = X_n \in \mathbf{P}$.

Функция $F_n^*(t)$ есть функция двух переменных: t и X, или, что то же, случайная функция от t или *случайный процесс*.

Найдем *конечномерные распределения* этого процесса. Пусть $t_1 < t_2 < \ldots < t_m$ — произвольные m точек числовой оси. Положим $t_0 = -\infty$, $t_{m+1} = \infty$ и обозначим через

$$\Delta_j g = g(t_{j+1}) - g(t_j)$$

приращения функции $g(t)$ на полуинтервалах $\Delta_j = [t_j, t_{j+1})$, $j = 0, 1, \ldots, m$. Рассмотрим приращение $\Delta_j \pi_n$ процесса

$$\pi_n(t) = nF_n^*(t).$$

Очевидно, это есть число элементов выборки, попавших в Δ_j. Вероятность попадания одного элемента выборки (скажем, x_1) в Δ_j равна $p_j = \mathbf{P}(\Delta_j)$. Так как попадания в Δ_j, $j = 0, 1, \ldots, m$, представляют собой $m+1$ несовместных событий, то, очевидно, имеем здесь полиномиальное распределение (см. [17, § 5.2]) для вектора $(\Delta_0\pi_n, \ldots, \Delta_m\pi_n)$ с вероятностями $p_0, \ldots, p_m$, $\sum\limits_{j=0}^{m} p_j = 1$. Как известно,

$$\mathbf{P}(\Delta_0\pi_n = k_0, \ldots, \Delta_m\pi_n = k_m) = \frac{n!}{k_0! \ldots k_m!} p_0^{k_0} \ldots p_m^{k_m}, \tag{1}$$

где $\sum\limits_{j=0}^{m} k_j = n$.

Пусть теперь $\eta(u)$, $u \in [0, 1]$, есть непрерывный слева пуассоновский процесс (см. [17, § 16.4]) *с параметром* λ, $\eta(0) = 0$. Приращения этого процесса независимы,

$$\mathbf{P}(\eta(u) = k) = e^{-\lambda u} \frac{(\lambda u)^k}{k!}.$$

Если функция распределения $F(t) = \mathbf{P}((-\infty, t))$ непрерывна, то можем сделать непрерывную замену времени, положив $u = F(t)$, $-\infty < t < \infty$,

и определить тем самым процесс $\pi(t) = \eta(F(t))$ на всей оси. Рассмотрим приращения этого процесса

$$\Delta_j\pi = \pi(t_{j+1}) - \pi(t_j) = \eta(F(t_{j+1})) - \eta(F(t_j))$$

на интервалах Δ_j. Тогда

$$\mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m) = \prod_{j=0}^{m} e^{-\lambda p_j}\frac{(\lambda p_j)^{k_j}}{k_j!} = e^{-\lambda}\lambda^n \prod_{j=0}^{m}\frac{p_j^{k_j}}{k_j!},$$

а условная вероятность того же события при условии $\pi(\infty) = \sum_{j=0}^{m}\Delta_j\pi = n$ будет равна

$$\mathbf{P}\left(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m \,\middle|\, \sum_{j=0}^{m}\Delta_j\pi = n\right) = \frac{\mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m)}{\mathbf{P}(\pi(\infty) = n)} = \mathbf{P}(\Delta_0\pi = k_0, \ldots, \Delta_m\pi = k_m)\frac{e^\lambda n!}{\lambda^n} = n!\prod_{j=0}^{m}\frac{p_j^{k_j}}{k_j!}. \quad (2)$$

Мы получили при любом $\lambda > 0$ то же самое выражение, что и в правой части (1). Таким образом, мы доказали следующее утверждение.

Теорема 1. *Если $F(t)$ непрерывна, то распределение процесса $nF_n^*(t)$ совпадает с условным распределением процесса $\pi(t) = (\eta(F(t))$ при условии $\pi(\infty) = n$ $(\eta(1) = n)$.*

Теорема показывает, что отклонения $n\big(F_n^*(t) - F(t)\big)$ распределены так же, как $\eta(F(t)) - nF(t)$ при условии $\eta(1) = n$, и задача с точностью до замены времени $u = F(t)$ сводится к изучению отклонений $\eta(u) - nu$ для условного $(\eta(1) = n)$ пуассоновского процесса на отрезке $[0, 1]$ или, что то же, к изучению отклонений $n\big(F_n^*(t) - t\big)$, где $F_n^*(t)$ соответствует равномерному на $[0, 1]$ распределению.

Полезным бывает и другое представление для процесса $nF_n^*(t)$. Пусть $\zeta_1, \zeta_2, \ldots$ — точки скачков пуассоновского процесса $\eta(t)$, так что $\eta(\zeta_k + 0) = k$. Как известно [17], разности $\xi_k = \zeta_k - \zeta_{k-1}$ $(\zeta_0 = 0)$, $k = 1, 2, \ldots$, независимы и показательно распределены,

$$\mathbf{P}(\xi_k > x) = e^{-\lambda x},$$

ζ_k имеет Г-распределение с плотностью (см. также § 12)

$$\gamma_{\lambda,k}(x) = \frac{\lambda^k}{\Gamma(k)}e^{-\lambda x}x^{k-1}.$$

Чтобы упростить формулировки, предположим, что $F(t) = t$, $t \in [0,1]$, $t_0 = 0$, $t_{m+1} = 1$, так что $\eta(t) = \pi(t)$.

Теорема 2. *Распределение процесса $nF_n^*(t)$ совпадает при любом $v > 0$ с условным распределением процесса $\pi(tv)$, $0 < t < 1$, при условии $\zeta_{n+1} = v$.*

Другими словами, утверждение теоремы 1 сохранится, если условие $\pi(1) = n$ заменить значительно более узким условием $\pi(1) = n$, $\pi(1+0) = n+1$ (мы считаем, что траектории $\pi(t)$ непрерывны слева).

Поскольку вероятность этого нового условия равна 0, то, возможно, следует добавить (см. [17, §4.8 об условных математических ожиданиях, а также §19]), что под условным распределением мы понимаем здесь вероятности

$$\mathbf{P}(A/\zeta_{n+1} = v) = \frac{\mathbf{P}(A; \zeta_{n+1} \in dv)}{\mathbf{P}(\zeta_{n+1} \in dv)},$$

где $A = \{\Delta_0\pi(tv) = k_0, \ldots, \Delta_m\pi(tv) = k_m\}$, $\Delta_j\pi(tv) = \pi(t_{j+1}v) - \pi(t_jv)$.

Доказательство. Событие $\{\zeta_{n+1} \in dv\}$ представим в виде произведения двух событий

$$B = \{\pi(v) = n\}, \quad C = \{\pi(v+dv) - \pi(v) = 1\}.$$

События B и AB не зависят от C, так как события B и AB, с одной стороны, и событие C — с другой относятся к приращениям процесса π на непересекающихся интервалах времени. Поэтому

$$\mathbf{P}(A/\zeta_{n+1} = v) = \frac{\mathbf{P}(ABC)}{\mathbf{P}(BC)} = \frac{\mathbf{P}(AB)}{\mathbf{P}(B)} = \mathbf{P}(A/\pi(v) = n). \tag{3}$$

Точно так же, как в (2), убеждаемся, что это выражение от v (как и от λ) не зависит и совпадает с (1). ◁

Следствие 1. *Распределение процесса $nF_n^*(t)$ совпадает с распределением $\pi(t\zeta_{n+1})$, $0 \leqslant t \leqslant 1$.*

Это вытекает из того, что для

$$B = \{\Delta_0\pi(t\zeta_{n+1}) = k_0, \ldots, \Delta_m\pi(t\zeta_{n+1}) = k_m\}$$

имеем в силу (3)

$$\mathbf{P}(B) = \int \mathbf{P}(A/\zeta_{n+1} = v)\mathbf{P}(\zeta_{n+1} \in dv) = n! \prod_i \frac{\Delta_j^{k_j}}{k_j!}.$$

Из следствия 1 вытекает

Следствие 2. *Совместное распределение элементов вариационного ряда $\mathrm{x}_{(1)}, \ldots, \mathrm{x}_{(n)}$ выборки X из равномерного распределения совпадает с*

совместным распределением

$$\frac{\zeta_1}{\zeta_{n+1}}, \ldots, \frac{\zeta_n}{\zeta_{n+1}},$$

или, что то же, совместное распределение разностей $\mathrm{x}_{(1)}, \mathrm{x}_{(2)} - \mathrm{x}_{(1)}, \ldots$ $\ldots, \mathrm{x}_{(n)} - \mathrm{x}_{(n-1)}, 1 - \mathrm{x}_{(n)}$ *совпадает с совместным распределением*

$$\frac{\xi_1}{\zeta_{n+1}}, \ldots, \frac{\xi_{n+1}}{\zeta_{n+1}}.$$

В заключение этого пункта найдем моменты второго порядка для приращений процесса $n\bigl(F_n^*(t) - F(t)\bigr)$. Нам будет удобнее иметь дело с процессом

$$w^n(t) = \sqrt{n}\,\bigl(F_n^*(t) - F(t)\bigr).$$

Очевидно, что $\mathbf{E}\Delta_j w^n = 0$, $\mathbf{E}(\Delta_j w^n)^2 = \Delta_j F(1 - \Delta_j F)$. Для вычисления смешанных моментов заметим, что $(i \neq j)$

$$\mathbf{E}(\Delta_i w^n \cdot \Delta_j w^n) = \frac{1}{n} \sum_{k,l=1}^{n} \mathbf{E}(\mathbf{I}_{\mathrm{x}_k}(\Delta_i) - \mathbf{P}(\Delta_i))(\mathbf{I}_{\mathrm{x}_l}(\Delta_j) - \mathbf{P}(\Delta_j)) =$$

$$= \frac{1}{n} \sum_{k,l=1}^{n} \{\mathbf{E}\mathbf{I}_{\mathrm{x}_k}(\Delta_i)\mathbf{I}_{\mathrm{x}_l}(\Delta_j) - \mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j)\}.$$

Так как

$$\mathbf{E}\mathbf{I}_{\mathrm{x}_k}(\Delta_i)\mathbf{I}_{\mathrm{x}_l}(\Delta_j) = \begin{cases} \mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j) & \text{при } k \neq l, \\ 0 & \text{при } k = l, \end{cases}$$

то $\mathbf{E}(\Delta_i w^n \cdot \Delta_j w^n) = -\mathbf{P}(\Delta_i)\mathbf{P}(\Delta_j) = -\Delta_i F \cdot \Delta_j F$.

Таким образом, приращения процесса w^n отрицательно коррелированы.

2. Предельное поведение процесса $w^n(t)$. Мы будем считать, что $F(t)$ непрерывна. В предыдущем пункте мы видели, что задача отыскания распределения $nF_n^*(t)$ для произвольной непрерывной функции распределения F может быть редуцирована в известном смысле к той же задаче для функции распределения $U(t) = t$, $0 \leqslant t \leqslant 1$, соответствующей равномерному на $[0,1]$ распределению. Такой редукцией будем пользоваться и в дальнейшем. Сделаем в связи с этим несколько предварительных замечаний.

Пусть $x^{(F)} = \inf\{t\colon F(t) > 0\}$ — левая граница носителя распределения F. Положим $F^{-1}(0) = x^{(F)}$, $F^{-1}(u) = \sup\{v\colon F(v) < u\}$ при $u > 0$. Функция $F^{-1}(u)$ называется обратной к F, она так же, как и F, непрерывна слева. Если F непрерывна, то $F^{-1}(u)$ есть минимальное решение уравнения $F(v) = u$ (решение единственно, если $F(v)$ строго монотонна). Нетрудно видеть также, что если F непрерывна, то $F(F^{-1}(t)) \equiv t$, $t \in [0,1]$. В этом случае функции от наблюдений $\mathrm{y}_k = F(\mathrm{x}_k)$ представляют собой выборку Y из равномерного на $[0,1]$ распределения, так как по определению обратной функции

$$\mathbf{P}(\mathrm{y}_k < t) = \mathbf{P}(F(\mathrm{x}_k) < t) = \mathbf{P}(\mathrm{x}_k < F^{-1}(t)) = F(F^{-1}(t)) = t.$$

Так как, кроме того,

$$\mathbf{P}(F^{-1}(\mathrm{y}_k) < t) = \mathbf{P}(\mathrm{y}_k < F(t)) = F(t),$$

то

$$Y = (\mathrm{y}_1, \ldots, \mathrm{y}_k) \in U, \quad (F^{-1}(\mathrm{y}_1), \ldots, F^{-1}(\mathrm{y}_k)) \in F.$$

Если U_n^* — эмпирическая функция распределения, соответствующая выборке Y из равномерного на $[0,1]$ распределения, то из сказанного следует, что

$$F_n^*(t) = U_n^*(F(t)). \tag{4}$$

Действительно, левая и правая части этого соотношения равны k/n при $t \in (\mathrm{x}_{(k)}, \mathrm{x}_{(k+1)}]$ или, что то же, при

$$F(t) \in (F(\mathrm{x}_{(k)}, F(\mathrm{x}_{(k+1)})] = (\mathrm{y}_{(k)}, \mathrm{y}_{(k+1)}], \quad k = 1, \ldots, n.$$

Итак, если найдем способы приближать тем или иным образом распределение процесса $U_n^*(t)$ при больших n, то тем самым в силу (4) мы найдем подходы и к приближенному описанию распределения процесса $F_n^*(t)$ для произвольной непрерывной F.

В дальнейшем в этом пункте будем считать, что $F(t) = U(t)$ есть равномерное на $[0,1]$ распределение.

Обозначим через $w(t)$ *стандартный винеровский процесс*, т. е. процесс с независимыми приращениями, для которого $w(t)$ распределено нормально с параметрами $(0,t)$. Процесс

$$w^0(t) = w(t) - tw(1)$$

называют *броуновским мостом* (по той причине, что у него закреплены оба конца: $w^0(0) = w^0(1) = 0$). Распределение этого процесса совпадает с условным распределением процесса $w(t)$ при условии $w(1) = 0$ (более точно, надо взять условие $|w(1)| < \varepsilon$ и перейти к пределу при $\varepsilon \to 0$).

Оказывается, что конечномерные распределения процессов

$$w^n(t) = \sqrt{n}\,\big(F_n^*(t) - F(t)\big), \quad t \in [0,1],$$

сходятся при $n \to \infty$ к соответствующим распределениям броуновского моста $w^0(t)$.

Этот факт позволяет приближать процессы $w^n(t)$, которые называют иногда *эмпирическими процессами*, с помощью процесса $w^0(t)$. Именно, мы можем представлять себе, что при больших n имеет место приближенное равенство

$$\sqrt{n}\,\big(F_n^*(t) - F(t)\big) \approx w^0(t), \tag{5}$$

описывающее распределение уклонений $F_n^*(t)$ от $F(t)$ (напомним, что мы считаем здесь $F(t) = t$, $t \in [0,1]$).

Однако утверждение типа (5) нам понадобится в более сильной форме. Рассмотрим, например, статистику $S = \sqrt{n}\sup\limits_t\big(F_n^*(t) - (t)\big)$. Высказанное утверждение делает естественным предположение о том, что при больших n

случайная величина S приближенно распределена так же, как $\sup\limits_{0\leqslant t\leqslant 1} w^0(t)$. Однако из нашего утверждения это никак не следует, поскольку S нельзя представить как функцию от значений $w^n(t) = \sqrt{n}\left(F_n^*(t) - f(t)\right)$ в каком-либо конечном числе точек. Поэтому существенно более сильным является следующее утверждение.

Обозначим через $D(a,b)$ пространство функций на отрезке $[a,b]$, непрерывных слева (в точке a справа) и имеющих лишь конечное число скачков, и через $C(a,b)$ пространство всех непрерывных на $[a,b]$ функций. Через $D_0(a,b)$, $C_0(a,b)$ будем обозначать подпространства функций из $D(a,b)$, $C(a,b)$ соответственно, обращающихся в нуль в точках a и b. Очевидно, что траектории $w^n(t)$ принадлежат $D_0(0,1)$. Кроме того, известно (см. [17, § 16.2]), что траектории $w^0(t)$ принадлежат $C_0(0,1)$ с вероятностью 1. Для простоты можем считать, что все траектории $w(t)$ и, следовательно, $w^0(t)$ лежат в $C_0(0,1)$ (см. [17]). Поскольку $C_0(0,1) \subset D_0(0,1)$, то $(D_0(0,1), \sigma_D)$, где σ_D есть σ-алгебра подмножеств из $D_0(0,1)$, порожденная цилиндрическими множествами*), мы можем считать выборочным пространством**) процессов w^n и w^0.

Теорема 3 (функциональная предельная теорема для эмпирических процессов). *Пусть f — функционал, определенный на пространстве $D_0(0,1)$ и обладающий следующими свойствами:*

1) *$f(w_n)$ и $f(w^0)$ являются случайными величинами (т.е. $f(y)$ осуществляет измеримое отображение $(D_0(0,1), \sigma_D)$ в $(\mathbb{R}, \mathfrak{B})$);*

2) *$f(y)$ есть функционал, непрерывный в «точках» пространства $C_0(0,1)$ относительно равномерной метрики, т.е. $f(y_n) \to f(y)$ при $n \to \infty$, если $y \in C_0(0,1)$ и $\rho(y_n, y) = \sup\limits_{0\leqslant t\leqslant 1} |y_n(t) - y(t)| \to 0$.*

Если эти условия выполнены, то

$$f(w^n) \Rightarrow f(w^0).$$

Если функционал f непрерывен в равномерной метрике в любой точке $y \in D_0(0,1)$, то условие 1) выполняется автоматически.

Очевидно, что функционал S, рассмотренный выше, удовлетворяет условиям теоремы, так что при $n \to \infty$

$$S \Rightarrow \sup_{0\leqslant t\leqslant 1} w^0(t).$$

Поскольку распределение правой части в этом соотношении можно найти в явном виде (см., например, [7, 104])

$$\mathbf{P}\{\sup_{0\leqslant t\leqslant 1} w^0(t) > z\} = e^{-2z^2},$$

*) То есть множествами вида $\{y(t_1) \in B_1, \ldots, y(t_m) \in B_m\}$, где $B_1, \ldots, B_m$ — борелевские множества.

**) (D, σ) есть выборочное пространство процесса $\xi(t)$, если на нем задано распределение процесса ξ, так что траектории $\xi(t)$ лежат в D.

то мы получаем таким образом приближенное выражение для распределения S.

Использование теоремы 3 для вычисления предельного распределения других статистик рассмотрено в последующих параграфах.

Доказательство теоремы 3 отнесено нами в приложение II.

Возвращаясь к произвольным непрерывным функциям распределения F и пользуясь (4), можем записать

$$F_n^*(t) = U_n^*(F(t)) = F(t) + \frac{1}{\sqrt{n}} w^n(F(t)), \tag{6}$$

где поведение процесса $w^n(t)$ описано в теореме 3.

Отметим, что левая и правая части (6) равны и без предположения о непрерывности F.

§ 7. Предельное распределение для статистик первого типа

Напомним, что статистиками первого типа мы называем статистики $S_n(X) = G(F_n^*)$, где функционал G имеет вид $G(F) = h\left(\int g(x)\,dF(x)\right)$. Другими словами,

$$S_n(X) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i)\right).$$

Мы уже видели (теорема 3.1), что если $X \in F_0$ и h непрерывна в точке $a = \int g(x)\,dF_0(x)$, то $S_n \underset{\text{п.н.}}{\longrightarrow} h(a)$.

Теорема 1.

Если $X \in F_0$, h *дифференцируема в точке* a, $\int g^2(x)\,dF_0(x) < \infty$, *то*

$$\sqrt{n}\,(S_n(X) - h(a)) \Rightarrow h'(a)\xi,$$

где $\xi \in \Phi_{0,\sigma^2}$, $\sigma^2 = \int (g(x) - a)^2\,dF_0(x)$, Φ_{0,σ^2} — *нормальное распределение с параметрами* $(0, \sigma^2)$.

Доказательство. Статистику $S_n(X)$ представим в виде

$$h\left(a + \frac{1}{\sqrt{n}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(\mathrm{x}_i) - a)\right]\right),$$

где по центральной предельной теореме (см. [17], гл. 8 или приложение IV)

$$\eta_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(\mathrm{x}_i) - a) \Rightarrow \Phi_{0,\sigma^2},$$

$$\sigma^2 = \mathbf{E}(g(\mathrm{x}_1) - a)^2 = \int (g(x) - a)^2\,dF_0(x).$$

Отстается воспользоваться третьей теоремой непрерывности при $b_n = 1/\sqrt{n}$. $\triangleleft$

Иногда функционалы первого типа удобнее рассматривать в виде $G(F) = h\left(\int g(x)\,d(F-F_0)\right)$. Очевидно, все сказанное сохраняется и для них с той лишь разницей, что a следует положить равным нулю.

Приведем аналог теоремы 1 для случая, когда функция $g = (g_1, \ldots, g_s)$ есть вектор $\left(\text{т. е. } G(F) = h\left(\int g_1(x)\,dF(x), \ldots, \int g_s(x)\,dF(x)\right)\right)$.

Теорема 1А. *Пусть* $S_n(X) = G(F_n^*)$, $h(t)$ *дифференцируема в точке* $a = \int g(x)\,dF_0(x)$, *а матрица вторых моментов* $\sigma^2 = \|\sigma_{ij}\| = \mathbf{E}(g(\mathrm{x}_1) - a)^\top (g(\mathrm{x}_1) - a)$ *конечна. Тогда*

$$(S_n(X) - h(a))\sqrt{n} \Rightarrow \xi(h'(a))^\top = \sum_{j=1}^{s} \frac{\partial h(a)}{\partial t_j}\xi_j, \tag{1}$$

где $\xi = (\xi_1, \ldots, \xi_s) \in \Phi_{0,\sigma^2}$.

Если $\xi(h'(a))^\top = 0$ *с вероятностью* 1, *а матрица вторых производных* $h''(t) = \left\|\dfrac{\partial^2}{\partial t_i \partial t_j} h(t)\right\|$ *существует в точке* a, *то*

$$(S_n(X) - h(a))n \Rightarrow \frac{1}{2}\xi h''(a)\xi^\top = \frac{1}{2}\sum_{i,j=1}^{s} \left.\frac{\partial^2 h(t)}{\partial t_i \partial t_j}\right|_{t=a} \xi_i \xi_j.$$

Для доказательства теоремы 1А следует воспользоваться теоремой непрерывности 5.3А и многомерной центральной предельной теоремой, в силу которой $\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(\mathrm{x}_i) - a) \Rightarrow \xi$ (см. приложение IV).

Совершенно аналогично выглядит теорема о предельном распределении $S_n(X)$, когда функция h, а вместе с ней и статистика $S_n(X)$ являются векторами. Читатель без труда воспроизведет ее формулировку и доказательство с помощью теоремы 5.3В.

Пример 1. Пусть $X \in \mathbf{P}_0$ и $\mathbf{P}_0$ таково, что $\mathbf{E}\mathrm{x}_1 = \alpha > 0$, $\mathbf{D}\mathrm{x}_1 = d^2 < \infty$. Что из себя представляет в этих условиях предельное распределение статистики $S = \dfrac{1}{\overline{\mathrm{x}}}\left(\overline{\mathrm{x}} = \dfrac{1}{n}\sum_{i=1}^{n}\mathrm{x}_i\right)$? Условия теоремы 1 здесь, очевидно, выполнены при $h(t) = 1/t$, $g(x) = x$, при этом $a = \alpha$, $\sigma^2 = d^2$, $h(a) = 1/\alpha$, $h'(a) = -1/\alpha^2$. В силу теоремы 1

$$\left(S - \frac{1}{\alpha}\right)\sqrt{n} \Rightarrow -\frac{\xi}{\alpha^2}, \quad \xi \in \Phi_{0,d^2},$$

так что

$$\left(S - \frac{1}{\alpha}\right)\sqrt{n} \Subset \mathbf{\Phi}_{0,d^2/\alpha^4}.$$

Пример 2. Найдем предельное распределение статистики

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2,$$

если $\mathbf{E}\mathrm{x}_1 = \alpha$, $\mathbf{D}\mathrm{x}_1 = d^2$ и $\mathbf{E}\mathrm{x}_1^4 < \infty$. (Мы уже знаем, что в силу первой теоремы непрерывности $S^2 \underset{\text{п.н.}}{\longrightarrow} d^2$.) Найти требуемое предельное распределение нетрудно непосредственно, воспользовавшись представлениями

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i - \alpha)^2 - (\overline{\mathrm{x}} - \alpha)^2,$$

$$(S^2 - d^2)\sqrt{n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[(\mathrm{x}_i - \alpha)^2 - d^2] - \sqrt{n}(\overline{\mathrm{x}} - \alpha)^2.$$

Однако мы воспользуемся теоремой 1А. В условиях этой теоремы мы должны положить

$$G(F) = \int (x - \alpha)^2\, dF(x) - \left(\int x\, dF(x) - \alpha\right)^2,$$

так что $g_1(x) = (x - \alpha)^2$, $g_2(x) = x$, $h(t) = t_1 - (t_2 - \alpha)^2$. Поскольку в точке $a = (d^2, \alpha)$

$$\frac{\partial h(a)}{\partial t_1} = 1, \quad \frac{\partial h(a)}{\partial t_2} = 0,$$

то

$$(S^2 - d^2)\sqrt{n} \Rightarrow \xi, \quad \xi \Subset \mathbf{\Phi}_{0,v^2}, \quad v^2 = \mathbf{E}(\mathrm{x}_1 - \alpha)^4 - d^4.$$

Пример 3. *Статистика* χ^2. В заключение этого параграфа рассмотрим пример статистики, которую можно относить к статистикам как I, так и II типов.

Рассмотрим статистики, построенные с помощью функционалов вида

$$G(F) = h\left(\int g\, dF\right), \tag{2}$$

где g есть функция ограниченной вариации на отрезке $[a, b]$ таком, что $F(a) = 0$, $F(b) = 1$ (a и b могут быть бесконечными). Так как $\int g\, dF = g(b) - \int F\, dg$, то функционал $G(F)$ будет непрерывным в равномерной

метрике, если только непрерывна функция h. Нетрудно понять, что выделенный класс статистик представляет собой пересечение классов статистик I и II типов.

То же самое справедливо в случае, когда g есть векторнозначная функция с компонентами g_i, имеющими ограниченную вариацию.

Рассмотрим теперь разбиение вещественной оси (пространства $\mathcal{X}$) на непересекающиеся интервалы $\Delta_1, \ldots, \Delta_r$ и положим $\nu_i = n\mathbf{P}_n^*(\Delta_i)$, $p_i = \mathbf{P}_0(\Delta_i)$ ($\mathbf{P}_0$ есть распределение, соответствующее F_0, так что $X \in \mathbf{P}_0$). Статистикой «хи-квадрат» $\xi^2 = \chi^2(X)$ называют статистику

$$\chi^2(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}.$$

Очевидно, что это есть статистика II типа, так как она с точностью до множителя n соответствует функционалу

$$G(F) = G_1(\mathbf{P}) = \sum_{i=1}^{r} \frac{(\mathbf{P}(\Delta_i) - \mathbf{P}_0(\Delta_i))^2}{\mathbf{P}_0(\Delta_i)}.$$

Чтобы представить $\chi^2(X)$ как статистику I типа, рассмотрим функционал вида (2)

$$G(F) = h\left(\int g\, d(F - F_0)\right)$$

с функцией $h(u) = \sum_{j=1}^{r} u_j^2$ и векторной функцией g с координатами

$$g_j(x) = \begin{cases} \dfrac{1}{\sqrt{p_j}} & \text{при } x \in \Delta_j, \\ 0 & \text{при } x \notin \Delta_j. \end{cases}$$

Так как функция h дифференцируема, $\partial h(0)/\partial u_j = 0$, $\partial^2 h(0)/\partial u_i\, \partial u_j = 2\delta_{ij}$ (δ_{ij} — символ Кронекера), то, положив $S_n(X) = G\left(F_n^*\right)$, мы получим

$$nS_n(X) = n\sum_{j=1}^{r}\left[\left(\frac{\nu_j}{n} - p_j\right)\frac{1}{\sqrt{p_j}}\right]^2 = \chi^2(X).$$

При $X \in P_0$ в силу второй части теоремы 1А

$$\chi^2(X) \Rightarrow \sum_{j=1}^{r} \xi_j^2, \tag{3}$$

где $\xi = (\xi_1, \ldots, \xi_r)$ есть нормально распределенный вектор (предельный для $\left(\dfrac{\nu_1 - np_1}{\sqrt{np_1}}, \ldots, \dfrac{\nu_r - np_r}{\sqrt{np_r}}\right)$) с нулевым средним и с матрицей $\sigma^2 = \|\sigma_{ij}\|$ вторых моментов,

$$\sigma_{ij} = \mathbf{E}\xi_i\xi_j = \mathbf{E}(g_i(\mathrm{x}_1) - \sqrt{p_i})(g_j(\mathrm{x}_1) - \sqrt{p_j})$$

(из определения g_j следует, что $\mathbf{E}g_j(\mathrm{x}_1) = \sqrt{p_j}$). Так как $g_i(x)g_j(x) = 0$ при $i \neq j$ и $\mathbf{P}\big(g_j^2(\mathrm{x}_1) = 1/p_j\big) = p_j$, $\mathbf{P}\big(g_j^2(\mathrm{x}_1) = 0\big) = 1 - p_j$, то

$$\sigma_{ij} = \delta_{ij} - \sqrt{p_i p_j}.$$

Выясним теперь, что из себя представляет распределение правой части в (3) (т. е. предельное распределение $\chi^2(X)$).

Рассмотрим ортогональное преобразование в $\mathbb{R}^r$ с матрицей C и рассмотрим вектор

$$\eta = \xi C.$$

Вектор η так же, как и ξ, будет нормально распределен. Действительно, нормальность величины ξ означает, что ее характеристическая функция (х. ф.) равна (см. [17, § 7.6])

$$\mathbf{E}e^{it\xi^\top} = e^{-\frac{1}{2}t\sigma^2 t^\top},$$

где $\sigma^2 = \|\sigma_{ij}\|$ есть матрица вторых моментов. Но х. ф. для η

$$\mathbf{E}e^{it\eta^\top} = \mathbf{E}e^{itC^\top\xi^\top} = e^{-\frac{1}{2}tC^\top\sigma^2 Ct^\top}$$

имеет тот же вид и, следовательно, η есть нормальный вектор, но с матрицей вторых моментов $d^2 = C^\top\sigma^2 C = \|d_{ij}\|$, так что

$$d_{ij} = \mathbf{E}\eta_i\eta_j = \sum_{k,l} c_{li}\sigma_{lk}c_{kj} = \sum_{k,l} c_{li}(\delta_{lk} - \sqrt{p_l p_k})c_{kj} =$$

$$= \sum_l c_{li}c_{lj} - \Big(\sum_l c_{li}\sqrt{p_l}\Big)\Big(\sum_k c_{kj}\sqrt{p_k}\Big). \quad (4)$$

Выберем теперь матрицу C так, чтобы ее первый столбец имел координаты $c_{l1} = \sqrt{p_l}$ (это соответствует фиксированию первого вектора преобразованной системы координат и возможно, так как $\sum_{l=1}^r c_{l1}^2 = \sum p_l = 1$). Тогда, очевидно, второе слагаемое в (4) в силу ортогональности C равно единице лишь при $i = j = 1$ и нулю в противном случае. Это означает, что $d_{11} = \mathbf{E}\eta_1^2 = 0$, $d_{ij} = \mathbf{E}\eta_i\eta_j = \delta_{ij}$ при $i \geqslant 2$, и, следовательно, $\eta_1 = 0$ с вероятностью 1, а величины $\eta_2, \ldots, \eta_r$ независимы и нормально распределены с параметрами $(0, 1)$. На основании ортогональности C получаем

$$\sum_{j=1}^r \xi_j^2 = \sum_{j=1}^r \eta_j^2 = \sum_{j=2}^r \eta_j^2, \quad \chi^2(X) \Rightarrow \sum_{j=2}^r \eta_j^2. \quad (5)$$

Распределение правой части в этом равенстве называется *распределением* χ^2 («хи-квадрат») с $r - 1$ степенями свободы (см. [17, § 7.7], а также § 12). Мы будем встречать это распределение в последующем изложении неоднократно.

Еще одно доказательство (5) будет получено в следующем параграфе. Кроме того, (5) будет доказано в § 56 с помощью более общих соображений.

Некоторые другие примеры использования теорем 1, 1А будут содержаться в последующих главах.

§ 8.* Предельное распределение для статистик второго типа

В этом параграфе будем изучать предельное распределение $G(F_n^*)$ в случае $\mathcal{X} = \mathbb{R}$, где $G(F)$ есть функционал второго типа, осуществляющий измеримое отображение $(D(-\infty,\infty),\sigma_D)$ в $(\mathbb{R},\mathfrak{B})$ (или измеримое отображение $(V(-\infty,\infty),\sigma_V)$ в $(\mathbb{R},\mathfrak{B})$, где $V(-\infty,\infty)$ — пространство функций ограниченной вариации на $(-\infty,\infty)$, σ_V — σ-алгебра, порожденная цилиндрическими множествами).

Применительно к рассматриваемым задачам в дальнейшем под $C(-\infty,\infty)$ $(D(-\infty,\infty))$ можно понимать пространство функций $\widetilde{f}$ на $(-\infty,\infty)$, полученных из элементов $f \in C(0,1)$ $(f \in D(0,1))$ путем непрерывной замены «времени»: $\widetilde{f}(t) = f(F(t))$, где F — непрерывная функция распределения.

Отметим прежде всего, что для статистик II типа справедлива

Теорема 1. *Если* $X_n = [X_\infty]_n \in F_0$, *то* $G(F_n^*) \underset{\text{п.н.}}{\longrightarrow} G(F_0)$.

Это утверждение немедленно следует из теоремы 2.2А.

Для того чтобы можно было найти предельное распределение для функционала второго типа $G(F_n^*)$, необходимо, как и в предыдущем параграфе, наложить на функционалы $G(F)$ некоторые условия гладкости.

Положим для краткости $\|x\| = \sup\limits_{-\infty<t<\infty} |x(t)|$.

Определение 1. Функционал $G(F)$ называется *непрерывно дифференцируемым порядка k в точке F_0*, если существует функционал $G'(F_0,v)$, который для любой функции $v \in C(-\infty,\infty)$ и любой последовательности $v_h \in D(-\infty,\infty)$ такой, что $\|v_h - v\| \to 0$ при $h \to 0$, удовлетворяет соотношениям

$$\frac{G(F_0 + hv_h) - G(F_0)}{h^k} \to G'(F_0,v), \quad G'(F_0,v_h) \to G'(F_0,v). \tag{1}$$

Последнее соотношение означает, очевидно, непрерывность в равномерной метрике в точках из $C(-\infty,\infty)$ функционала $G'(F_0,v)$, который можно называть *производной порядка k от G по направлению v*.

Замечание 1. Введенное понятие производной шире понятий производных по Гато и Фреше, так как здесь не предполагается линейность функционала G' по v и равенство $k = 1$. Названные производные применительно к задачам робастности (см. § 58) называются также *функционалами влияния* (подробнее см., например, [124–126] и § 58). Значение $G'(F_0,v)$ совпадает, очевидно, с обычной производной функции $G(F_0 + hv)$ по h в точке $h = 0$.

Рассмотрим несколько примеров.

В дальнейшем будем обозначать через F_n^{-1} функцию, обратную F_n^* (см. § 6):

$$F_n^{-1}(0) = x_{(1)}, \quad F_n^{-1}(u) = \sup\{v\colon F_n^*(v) < u\} \quad \text{при} \quad u > 0.$$

Пример 1. Пусть $G(F)$ есть квантиль порядка p функции распределения F, т.е. $G(F) = F^{-1}(p)$. Это значит, что $G(F_n^*) = F_n^{-1}(p) = \zeta_p^*$ есть выборочная квантиль порядка p. Покажем, что в этом примере $G(F)$ дифференцируем первого порядка в точке F_0, если $F_0(t)$ дифференцируема в точке $\zeta_p = F_0^{-1}(p)$, $F'(\zeta_p) > 0$.

Действительно, по определению

$$G(F_0 + hv_h) = \max\{t\colon F_0(t) + hv_h(t) \leqslant p\}.$$

Так как этот функционал является непрерывным в равномерной метрике в точке F_0, то мы можем положить $G(F_0 + hv_h) = \zeta_p + \delta$, где $\delta = \delta(h) \to 0$ при $h \to 0$. Далее, из соотношения $\|v_h - v\| \to 0$, $v \in C(-\infty, \infty)$, следует

$$|v_h(p + \delta) - v_h(p)| \equiv r(h) \to 0$$

при $h \to 0$. Так как

$$F_0(\zeta_p + \delta) = p + \delta F_0'(\zeta_p) + o(\delta),$$

то для $t = G(F_0 + hv_h) = \zeta_p + \delta$ получаем

$$F_0(t) + hv_h(t) = p + \delta F_0'(\zeta_p) + o(\delta) + hv_h(\zeta_p + \delta) = \\ = p + \delta F_0'(\zeta_p) + o(\delta) + h(v_h(\zeta_p) + \tau r(h)) \leqslant p,$$

где $|\tau| \leqslant 1$. Аналогичное обратное неравенство можно написать, пользуясь тем, что $F_0(t + 0) + hv_h(t + 0) \geqslant p$. Отсюда следует, что

$$\delta + o(\delta) = -\frac{h(v_h(\zeta_p) + \tau_1 r(h))}{F_0'(\zeta_p)}, \quad |\tau_1| \leqslant 1,$$

так что

$$\frac{G(F_0 + hv_h) - G(F_0)}{h} = \frac{\delta}{h} \to -\frac{v(\zeta_p)}{F_0'(\zeta_p)}.$$

Таким образом, производная $G'(F_0, v)$ в этом примере равна

$$G'(F_0, v) = -\frac{v\left(F_0^{-1}(p)\right)}{F_0'\left(F_0^{-1}(p)\right)}. \tag{2}$$

Пример 2. Рассмотрим функционал

$$G(F) = \sup_t |F(t) - F_0(t)|.$$

Очевидно, что он дифференцируем первого порядка по любому направлению $v \in C(-\infty, \infty)$, так как $G(F_0) = 0$,

$$G'(F_0, v) = \frac{G(F_0 + hv)}{h} = \sup_t |v(t)|,$$

при этом $G'(F_0, v)$ от F_0 не зависит.

Пример 3. Функционал $G(F) = \int\limits_{-\infty}^{\infty} |F(t) - F_0(t)|^k \, dR(t)$ для любой функции ограниченной вариации $R(t)$ является непрерывно дифференцируемым порядка k по любому направлению $v \in C(-\infty, \infty)$, так как

$$G'(F_0, v) = \frac{G(F_0 + hv)}{h^k} = \int\limits_{-\infty}^{\infty} |v(t)|^k \, dR(t).$$

Пример 4. Рассмотрим функционал

$$G(F) = \sum_{j=1}^{r} \frac{(\Delta_j F - \Delta_j F_0)^2}{\Delta_j F_0},$$

где $\Delta_j F$ суть приращения функции F на интервалах $\Delta_j = [t_j, t_{j+1})$, образующих разбиение вещественной прямой. Очевидно, что $nG\left(F_n^*\right)$ есть не что иное, как статистика χ^2, рассмотренная в примере 7.3 в качестве статистики I типа. Рассматриваемый функционал будет непрерывно дифференцируемым второго порядка, так как для него

$$G'(F_0, v) = \frac{G(F_0 + hv)}{h^2} = \sum_{j=1}^{r} \frac{(\Delta_j v)^2}{\Delta_j F_0}.$$

Обобщением функционалов в примерах 2–4 являются функционалы вида $G(F) = G_1(F - F_0)$, где функционал G_1 является однородным в том смысле, что $G_1(hv) = h^k G_1(v)$. Очевидно, все такие функционалы будут дифференцируемыми.

Пример 5. Статистики I типа, введенные в § 3, соответствуют функционалам $G(F) = h\left(\int g(x)\, dF(x)\right)$.

Эти функционалы будут одновременно функционалами II типа, если функция h непрерывна в точке $\int g(x)\, dF_0(x)$, а g имеет ограниченную вариацию. В этом случае, считая, не ограничивая общности, $g(-\infty) = 0$, получим, что интеграл $\int g(x)\, dF(x) = \int (1 - F(x))\, dg(x)$ непрерывен относительно F в равномерной метрике. Если вариация g не ограничена, то такая непрерывность, вообще говоря, будет отсутствовать. Если, например, $h(x) \equiv x$, $g(x) \equiv x$, $F^{(n)} = \left(1 - \frac{1}{n}\right) F_0 + \frac{1}{n} F_1$, где $a = \int x\, dF_0(x) < \infty$, $\int x\, dF_1(x) = \infty$, то $\|F^{(n)} - F_0\| \to 0$, $G(F_0) = a$, $G(F^{(n)}) = \infty$ при всех n.

Если функция h дифференцируема в точке $a = \int g(x)\, dF_0(x)$, g имеет ограниченную вариацию, то функционал G дифференцируем первого по-

рядка в точке F_0, при этом ($\|v_\varepsilon - v\| \to 0$ при $\varepsilon \to 0$)

$$G'(F_0, v) = \lim_{\varepsilon \to 0} \varepsilon^{-1} \left(h\left(a - \varepsilon \int v_\varepsilon(x)\, dg(x)\right) - h(a)\right) = \\ = -h'(a) \left(\int v(x)\, dg(x)\right).$$

Пример 6. L-статистики были введены в § 3. Мы будем рассматривать L-статистики, соответствующие функционалу

$$G(F) = \int_0^1 \varphi(t) g(F^{-1}(t))\, dt, \tag{3}$$

где F^{-1} — функция, обратная к F; φ и g — заданные функции. Пусть

$$\int_0^1 \varphi(F_0(t)) g'(t)\, dt < \infty, \tag{4}$$

функции F_0 и g непрерывно дифференцируемы.

Опуская ввиду громоздкости обоснование предельного перехода под знаком интеграла, в силу примера 1 будем иметь ($F_0'(t) = f(t)$)

$$G(F_0 + hv_h) = \int_0^1 \varphi(t) g\left(F_0^{-1}(t) - \frac{hv\left(F_0^{-1}(t)\right)}{f\left(F_0^{-1}(t)\right)} + o(h)\right) dt = \\ = \int_0^1 v(t) \varphi\left(F_0^{-1}(t)\right) dt - h \int_0^1 \varphi(t) g'\left(F_0^{-1}(t)\right) \frac{v\left(F_0^{-1}(t)\right) dt}{f\left(F_0^{-1}(t)\right)} + o(h);$$

$$\lim_{n \to \infty} h^{-1} [G(F_0 + hv_h) - G(F_0)] = - \int_0^1 \varphi(t) g'\left(F_0^{-1}(t)\right) \frac{v\left(F_0^{-1}(t)\right) dt}{f\left(F_0^{-1}(t)\right)}.$$

Это значит, что функционал G, определенный в (3), дифференцируем первого порядка,

$$G'(F_0, v) = - \int_0^1 \varphi(t) g'\left(F_0^{-1}(t)\right) \frac{v\left(F_0^{-1}(t)\right) dt}{f\left(F_0^{-1}(t)\right)} = - \int_{-\infty}^{\infty} \varphi(F_0(t)) g'(t) v(t)\, dt. \tag{5}$$

Сформулируем теперь основную теорему о предельном распределении функционалов второго типа. Через fg мы будем обозначать функцию $fg(x) = f(g(x))$.

Теорема 2. *Пусть F_0 непрерывна, $X \in F_0$ и функционал $G(F)$ дифференцируем (порядка k в смысле определения* 1*). Тогда*

$$n^{k/2}\big(G(F_n^*) - G(F_0)\big) \Rightarrow G'(F_0, w^0 F_0), \tag{6}$$

где w^0 есть броуновский мост.

Доказательство. Воспользуемся теоремой 6.3. В ней доказана сходимость распределений функционалов от процесса $w^n(t) = \sqrt{n}\big(F_n^*(t) - F(t)\big)$, где $X \in F_0 = U$, U есть равномерное на $[0,1]$ распределение. В соответствии с этим функционалы в теореме 6.3 определены на пространстве $(D(0,1), \sigma_D)$, а не на $(D(-\infty,\infty)\sigma_D)$, как в общем случае. Поэтому предварительно мы должны произвести редукцию рассматриваемой задачи к условиям теоремы 6.3. Для этого воспользуемся соотношением $F_n^* = U_n^* F_0$ (см. (6.4)), где U_n^* — эмпирическая функция равномерного на $[0,1]$ распределения, F_0 — распределение, соответствующее выборке X. Это соотношение позволяет сводить изучение распределения $G(F_n^*)$ (G определен на $D(-\infty,\infty)$) к изучению распределения $G_0(U_n^*)$, где функционал G_0 определен на элементах H из $D(0,1)$ равенством $G_0(H) = G(HF_0)$, так что $G_0(\cdot)$ есть значение исходного функционала G, но от функции, у которой произведена замена «времени» t на $F_0(t)$. При таком определении G_0 имеем $G(F) = G\big(FF_0^{-1}F_0\big) = G_0\big(FF_0^{-1}\big)$.

Так как $F_n^* F_0^{-1} = U_n^*$ соответствует согласно (6.4) выборке из равномерного на $[0,1]$ распределения, то

$$G(F_n^*) = G_0(U_n^*).$$

Покажем теперь, что если функционал G дифференцируем, то G_0 также дифференцируем, $G_0'(U,v) = G'(F_0, vF_0)$.

Действительно, при $\|v_h - v\| \to 0$ имеем

$$G_0'(U,v) = \lim_{h\to 0} h^{-k}[G_0(U + hv_h) - G_0(U)] =$$

$$= \lim_{h\to 0} h^{-k}[G(F_0 + hv_hF_0) - G(F_0)] = G'(F_0, vF_0).$$

Итак, нам достаточно доказать теорему 2 в предположении $F_0 = U$. Рассмотрим величину

$$\big[G(F_n^*) - G(U)\big]n^{k/2} = G'(U, w^n) + H_n(w^n),$$

где $H_n(x) = [G(U + x/\sqrt{n}) - G(U)]n^{k/2} - G'(U,x)$. Так как в силу теоремы 6.3 и определения 1 $G'(U,w^n) \Rightarrow G'(U,w^0)$, нам достаточно убедиться, что

$$H_n(w^n) \underset{P}{\longrightarrow} 0. \tag{7}$$

Заметим, что для любого компакта $K \subset C(0,1)$ и любой последовательности $h_n \to 0$ при $n \to \infty$

$$\sup_{\substack{x \in D(0,1) \\ x \in (K)^{h_n}}} |H_n(x)| \to 0, \tag{8}$$

где $(K)^\varepsilon$ есть ε-окрестность множества K. Допустив противное, придем к существованию последовательности $x_n \in D(0,1)$ такой, что $\|x_n - x\| \to 0$, $x \in C(0,1)$, $\limsup\limits_{n\to\infty} |H_n(x_n)| > 0$, что противоречит дифференцируемости G.

Напомним теперь (см., например, [7]), что компакты в метрическом пространстве непрерывных функций $C(0,1)$ с равномерной метрикой описываются следующим образом. Каждой функции $\varphi(\Delta) > 0$, $\varphi(\Delta) \to 0$ при $\Delta \to 0$, и числу $N > 0$ соответствует компакт

$$K = K(\varphi, N) = \{y \in C(0,1) : \omega_\Delta(y) \leqslant \varphi(\Delta),\ |y(0)| \leqslant N\},$$

где $\omega_\Delta(y)$ есть модуль непрерывности y:

$$\omega_\Delta(y) = \sup_{|t-u| \leqslant \Delta} |y(t) - y(u)|.$$

Обозначим через K_h множество

$$K_h = \{y \in D(0,1) : \omega_\Delta(y) \leqslant \varphi(\Delta) \text{ для всех } \Delta \geqslant h;\ |y(0)| \leqslant N\}.$$

Множества K_h можно было бы назвать «предкомпактами» (этот термин используется в функциональном анализе в ином смысле), порожденными компактом K. Ясно, что $K_{h_1} \subset K_{h_2}$ при $h_1 \leqslant h_2$, $\bigcap\limits_{n=1}^{\infty} K_{1/n} = K$ и что $K_h \subset \subset (K)^{\varphi(h)}$.

Покажем теперь, что для заданного $\delta > 0$ существуют компакт K (и, стало быть, соответствующее ему семейство предкомпактов K_h) и последовательность $h_n \to 0$ при $n \to \infty$ такие, что

$$\limsup_{n\to\infty} \mathbf{P}(w^n \notin K_{h_n}) \leqslant \delta. \tag{9}$$

Действительно, по теореме 6.3 для любого функционала f, непрерывного в равномерной метрике, выполняется $f(w^n) \Rightarrow f(w^0)$, где $w^n(t) = \sqrt{n}\big(F_n^*(t) - t\big)$, $0 \leqslant t \leqslant 1$. Так как $\omega_\Delta(y)$ является таким функционалом, то $\omega_\Delta(w^n) \Rightarrow \omega_\Delta(w^0)$. Но $\omega_\Delta(w^0) \underset{\text{п.н.}}{\longrightarrow} 0$ при $\Delta \to 0$, так как траектории w^0 почти наверное непрерывны. Следовательно, для заданных ε и δ при достаточно малом Δ

$$\mathbf{P}(\omega_\Delta(w^0) > \varepsilon) \leqslant \delta.$$

Считая, не ограничивая общности, число ε точкой непрерывности распределения $\omega_\Delta(w^0)$, мы получим

$$\limsup_{n\to\infty} \mathbf{P}(\omega_\Delta(w^n) > \varepsilon) \leqslant \delta.$$

Пусть теперь $\varepsilon_k \downarrow 0$ — некоторая последовательность, а числа $\Delta_k \downarrow 0$ таковы, что

$$\limsup_{n\to\infty} \mathbf{P}(\omega_{\Delta_k}(w^n) > \varepsilon_k) \leqslant \frac{\delta}{2^{k+1}}.$$

Образуем функцию $\varphi(\Delta) = \varepsilon_k$ при $\Delta \in [\Delta_{k+1}, \Delta_k)$. Ясно, что $\varphi(\Delta) \to 0$ при $\Delta \to 0$, и мы можем рассмотреть предкомпакты K_h, построенные по функции φ. Тогда для любого $k < \infty$

$$\limsup_{n\to\infty} \mathbf{P}(w^n \notin K_{\Delta_k}) \leqslant \limsup_{n\to\infty} \sum_{j=1}^{k+1} (\omega_{\Delta_j}(w^n) > \varepsilon_j) \leqslant$$

$$\leqslant \sum_{j=1}^{k+1} \limsup_{n\to\infty} \mathbf{P}(\omega_{\Delta_j}(w^n) > \varepsilon_j) \leqslant \frac{\delta}{2}$$

(при $k = \infty$ это неравенство может быть неверным). Полученное соотношение означает, что при каждом δ существует последовательность $h_n \to 0$ при $n \to \infty$ такая, что выполнено (9).

На основании (8), (9) получаем

$$\mathbf{P}(|H_n(w^n)| > \varepsilon) \leqslant \mathbf{P}(|H_n(w^n)| > \varepsilon,\ w^n \in K_{h_n}) + \mathbf{P}(w^n \notin K_{h_n}),$$

$$\limsup_{n\to\infty} \mathbf{P}(|H_n(w^n)| > \varepsilon) \leqslant \delta.$$

Так как δ произвольно, то соотношение (7), а вместе с ним и утверждение теоремы доказаны. $\triangleleft$

Вернемся к рассмотрению примеров. В примере 1 для выборочной квантили $\zeta_p^* = F_p^{-1}(p)$ порядка p имеем согласно (2) и теоремы 2.

Следствие 1. *Если* $X_n \Subset F_0$, F_0 *непрерывно дифференцируема в точке* ζ_p, $F_0'(\zeta_p) > 0$, *то*

$$(\zeta_p^* - \zeta_p)\sqrt{n} \Rightarrow \frac{w^0(p)}{F_0'(\zeta_p)}.$$

Для доказательства надо заметить лишь, что условия следствия 1 означают непрерывную дифференцируемость F_0^{-1} в точке p:

$$\bigl(F_0^{-1}(p)\bigr)' = \frac{1}{F_0'\bigl(F_0^{-1}(p)\bigr)} = \frac{1}{F_0'(\zeta_p)}.$$

Так как $\mathbf{E}w^0(p) = 0$, $\mathbf{D}w^0(p) = \mathbf{E}(w(p) - pw(1))^2 = \mathbf{E}(w(p)(1-p) + p(w(1) - w(p)))^2 = p(1-p)^2 + p^2(1-p) = p(1-p)$, то утверждение след-

ствия 1 можно записать также в виде

$$\left(\zeta_p^* - \zeta_p\right)\sqrt{n} \Rightarrow \Phi_{0,\sigma^2}, \quad \sigma^2 = \frac{p(1-p)}{\left(F_0'(\zeta_p)\right)^2}. \quad \triangleleft$$

О распределении выборочных квантилей см. также п. 12.9.

В примере 2 функционал $G(F) = \sup\limits_{0\leqslant t\leqslant 1} |F(t) - F_0(t)|$ дифференцируем и, стало быть, по теореме 2

$$G\left(F_n^*\right)\sqrt{n} \Rightarrow \sup_t |w^0(F_0(t)| = \sup_{0\leqslant t\leqslant 1} |w^0(t)|.$$

Это соответствует тому, что для произвольной непрерывной функции распределения F_0 распределение статистики

$$D(X) = \sup_t \left|F_n^*(t) - F_0(t)\right|$$

остается тем же, что и для случая $F_0(t) = t$, $t \in [0,1]$.

Распределение $\eta = \sup\limits_{0\leqslant t\leqslant 1} |w^0(t)|$ найдено в явном виде [104]:

$$\mathbf{P}(\eta < z) = K(z) \equiv 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2z^2}.$$

Функция $K(z)$ называется *функцией Колмогорова.*

Мы получили, таким образом,

Следствие 2 (теорема Колмогорова). *Если* $X \in F_0$, F_0 *непрерывна, то*

$$\sqrt{n}D(X) \Rightarrow K.$$

Это означает, что максимальное уклонение $D(X)$ функции $F_n^*(t)$ от $F_0(t)$ имеет порядок $1/\sqrt{n}$ и может быть приближенно представлено в виде $D(X) \approx \approx \eta/\sqrt{n}$.

В примере 3 в силу теоремы 2 получаем

$$n^{k/2}\int \left|F_n^*(t) - F_0(t)\right|^k dR(t) \Rightarrow \int |w^0(F_0(t))|^k\, dR(t).$$

В частности, при $k = 2$ и $R(t) = F_0(t)$ мы получим статистику, обозначаемую ω^2, распределение которой инвариантно относительно F_0 и для которой справедливо

Следствие 3. *Если* $X \in F_0$, F_0 *непрерывна, то*

$$n\omega^2 \Rightarrow \int_0^1 [w^0(t)]^2\, dt.$$

Распределение $\int\limits_0^1 [w^0(t)]^2\, dt$ также найдено в явном виде и наряду с распределением $K(z)$ табулировано.

Применительно к примеру 4 теорема 2 дает

$$\chi^2 \Rightarrow \sum_{i=1}^{r} \frac{(\Delta_j w^0 F_0)^2}{\Delta_j F_0}.$$

Если обозначить через $\delta_j F$ приращение F на интервалах $\delta_j = [\tau_j, \tau_{j+1})$, $\tau_j = F_0(t_j)$ и обозначить той же буквой δ_j длину интервала δ_j, то мы получим

Следствие 4. *Если* $X \in F_0$, F_0 *непрерывна, то*

$$\chi^2 \Rightarrow \sum_{j=1}^{r} \frac{(\delta_j w^0)^2}{\delta_j}.$$

Если положить $\xi = (\xi_1, \ldots, \xi_r)$, $\xi_j = \delta_j w^0 / \sqrt{\delta_j}$, то, пользуясь тем, что $\delta_j w^0 = \delta_j w - w(1)\delta_j$, где w — стандартный винеровский процесс, получим

$$\chi^2 \Rightarrow \sum_{j=1}^{r} \xi_j^2, \quad \xi_j \in \Phi_{0,\sigma^2}.$$

Здесь матрица $\sigma^2 = \|\sigma_{ij}\|$ та же, что и в примере 7.3, так как

$$\delta_j w^0 = \delta_j w - \left(\sum_k \delta_k w\right)\delta_j = \sum_{k=1}^{r} a_{kj}\delta_k w,$$

$$a_{kj} = \delta_{kj} - \delta_j, \quad \mathbf{E}(\delta_k w)(\delta_l w) = \delta_{kl}\delta_k$$

(δ_{kl} — символ Кронекера),

$$\sigma_{ij} = \frac{\mathbf{E}(\delta_i w^0)(\delta_j w^0)}{\sqrt{\delta_i \delta_j}} = \frac{1}{\sqrt{\delta_i \delta_j}} \sum_{k=1}^{r} a_{ki} a_{kj} \delta_k =$$

$$= \frac{1}{\sqrt{\delta_i \delta_j}} (\delta_{ij}\delta_i - \delta_i \delta_j) = \delta_{ij} - \sqrt{\delta_i \delta_j}.$$

Повторяя рассуждения примера 7.3, получим, что $\sum_{j=1}^{r} \xi_j^2$ имеет распределение χ^2 с $r-1$ степенями свободы.

Для примера 5 справедливо

Следствие 5. *Если* $X \in F_0$, F_0 *непрерывна, функция* g *имеет ограниченную вариацию,* h *непрерывно дифференцируема в точке* $a = \int g(x)\, dF_0(x)$, *то*

$$\sqrt{n}\big(G(F_n^* - h(a)\big) \Rightarrow -h'(a) \int w^0(F_0(x))\, dg(x). \tag{10}$$

Правая часть здесь имеет нормальное распределение с параметрами

$$\left(0, (h'(a))^2 \int (g(x)-a)^2\, dF_0(x)\right). \tag{11}$$

Чтобы вычислить распределение правой части в (10), можно воспользоваться простейшими свойствами стохастических интегралов, в силу которых (интегрирование по частям)

$$-\int w^0(F_0(x))\, dg(x) = \int g(x)\, dw^0(F_0(x)) =$$
$$= \int g(x)\, dw(F_0(x)) - w(1)\int (g(x)\, dF_0(x)) = \int (g(x)-a)\, dw(F_0(x)).$$

Этот интеграл имеет нормальное распределение с параметрами

$$\left(0, \int (g(x)-a)^2\, dF_0(x)\right).$$

Если названные свойства стохастического интеграла (правило интегрирования по частям и нормальность $\int\limits_0^1 b(t)\, dw(t)$ читателю не известны, можно воспользоваться аналогичными свойствами сумм случайных величин, поскольку $\int b(t)\, dw(t)$ для непрерывных $b(\cdot)$ можно рассматривать как предел $\sum\limits_{k=1}^{n} \xi_{k,n} b(k/n)$, где $\xi_{k,n} = w(k/n) - w((k-1)/n)$ суть независимые, одинаково (нормально) распределенные случайные величины с параметрами $(0, 1/n)$. Аналогичную трактовку следует использовать для исходного интеграла вида

$$\int w^0(t)\, dg(t) \quad \left(w\left(\frac{k}{n}\right) = \sum_{i=1}^{k} \xi_{i,n},\ w^0\left(\frac{k}{n}\right) = w\left(\frac{k}{n}\right) - \frac{k}{n} w(1)\right).$$

Тогда, меняя порядок суммирования (аналог интегрирования по частям), мы придем к асимптотическому равенству допредельных сумм, а следовательно, и к равенству пределов.

Замечание 2. Второе утверждение следствия 5 (см. (11)), как показывает теорема 7.1, справедливо и при более общих предположениях, когда вместо ограниченности вариации функции g требуется лишь существование $\int g^2(x)\, dF_0(x)$.

Для *L-статистик* $G(F_n^*)$ в примере 6 (функционал G определен в (3)) в силу теоремы 2 справедливо

Следствие 6. *Если* $X \in F_0$, g *и* F_0 *непрерывно дифференцируемы и выполнено* (4), *то*

$$\sqrt{n}\big(G(F_n^*) - G(F_0)\big) \Rightarrow \int_0^1 a(t)w^0(t)\,dt, \tag{12}$$

где

$$a(t) = \varphi(t)g'\big(F_0^{-1}(t)\big)/f\big(F_0^{-1}(t)\big).$$

Правая часть в (12) *нормальна с параметрами*

$$\left(0, \int_0^1 A^2(t)\,dt - A^2\right), \tag{13}$$

где $A(t) = \int_t^1 a(t)\,dt$, $A = \int_0^1 A(t)\,dt$.

Чтобы убедиться в последнем утверждении, надо, как и в предыдущем следствии, воспользоваться интегрированием по частям:

$$\int_0^1 a(t)w(t)\,dt = \int_0^1 A(t)\,dw(t),$$

$$\int_0^1 a(t)w^0(t)\,dt = \int_0^1 A(t)\,dw(t) - w(1)\int_0^1 ta(t)\,dt = \int_0^1 (A(t) - A)\,dw(t),$$

где правая часть нормальна с параметрами

$$\left(0, \int_0^1 (A(t) - A)^2\,dt\right) = \left(0, \int_0^1 A^2(t)\,dt - A^2\right).$$

Замечание 3. Можно показать (см., например, [102]), что второе утверждение следствия 6 (о нормальном предельном распределении $\sqrt{n}\big(G(F_n^*) - G(F_0)\big)$ с параметрами (13)), как и в следствии 5 (см. замечание 2), остается справедливым и при более широких условиях, когда требуется лишь $\int_0^1 A^2(t)\,dt < \infty$.

Существует много других примеров, показывающих, что теорема 2 остается справедливой и при условиях более широких, чем те, что сфор-

мулированы в теореме 2. Если воспользоваться представлением $F_n^*(t) = F_0(t) + \frac{1}{\sqrt{n}} w^n(F_0(t))$ (см. (6.6)), где поведение w^n описано в теореме 6.3, то можно записать

$$\sqrt{n}\big(G(F_n^*) - G(F_0)\big) = \sqrt{n}\left(G\left(F_0 + \frac{1}{\sqrt{n}} w^n F_0\right) - G(F_0)\right) \approx G'(F_0, w^n F_0), \tag{14}$$

где производная

$$G'(F_0, v) = \lim_{h \to 0} h^{-1}(G(F_0 + hv) - G(F_0))$$

может быть определена на элементах v из какого-нибудь другого пространства (не обязательно $C(-\infty, \infty)$), а функционал $G'(F_0, v)$ не обязан быть непрерывным относительно v в непрерывной метрике. Тем не менее в этих условиях мы по-прежнему можем иметь требуемую сходимость $G'(F_0, w^n F_0) \Rightarrow G'(F_0, w^0 F_0)$ (как это имело место в примерах 5, 6), если, конечно, правая часть этого соотношения имеет смысл (см. также результаты § 14 о распределении M-оценок).

По своему существу теоремы 1 и 2 близки соответственно теоремам непрерывности 5.1 и 5.3, но они действуют в функциональных пространствах.

В заключение этого параграфа отметим, что далеко не все статистики, представляющие интерес, могут быть отнесены к статистикам I и II типа. Достаточно рассмотреть, например, U-статистики (скажем, $U(X) = \sum_{i,j=1}^{n} h(\mathrm{x}_i, \mathrm{x}_j)$; см. § 3) или статистики S, связанные с функционалами $G_n(F)$, где функционалы G_n «существенно» зависят от n (не только через выборку), такие, как, скажем, максимальный член вариационного ряда $S(X) = \mathrm{x}_{(n)} = \zeta^*_{1-1/n}$ и др.

§ 9.* Замечания о непараметрических статистиках

Есть одно свойство, в отношении которого статистика ζ_p^* в примере 8.1 существенно отличается от статистик в примерах 8.2–8.4. Это свойство состоит в том, что предельное распределение статистик в примерах 8.2–8.4 (см. следствия 8.2–8.4) никак с функцией распределения F_0 не связано. Этого нельзя сказать о статистике ζ_p^* и о статистиках в примерах 8.5, 8.6 (ср. со следствиями 8.1, 8.4, 8.5).

Определение 1. Статистика $S(X)$ называется *асимптотически непараметрической*, если $S(X) \Subset \mathbf{Q}$, когда $n \to \infty$, и $\mathbf{Q}$ не зависит от распределения X, т. е. не зависит от F_0, если $X \Subset F_0$.

Отметим, что сама функция S при этом может зависеть от F_0. Сам термин «непараметрическая» является не совсем удачным; однако он весьма распространен (он оправдан в случае, когда F_0 принадлежит некоторому параметрическому семейству — тогда распределение $\mathbf{Q}$ не зависит от параметра и в этом смысле является непараметрическим). Иногда используется другой термин — «свободный от распределения».

Мы видели в § 6–8, что статистики $\sqrt{n}\,D(X)$, $n\omega^2(X)$, $\chi^2(X)$ являются асимптотически непараметрическими.

Заметим теперь, что теорема 6.1 делает возможным введение и более узкого понятия. В теореме 6.1 установлено, что $nF_n^*(t)$ распределено так же, как $\eta(F_0(t))$, где $\eta(u)$ — условный пуассоновский процесс с произвольным параметром $\lambda > 0$ при условии $\eta(1) = n$ (см. § 6), т. е. процесс, от F_0 не зависящий. Таким образом, если статистика S построена как функционал $G(F_n^*)$ (или $G(F_n^* - F_0)$), который инвариантен относительно замены «времени» t у аргумента, то распределение S от F_0 зависеть не будет. Например,

$$D = \sup_t \left|F_n^*(t) - F_0(t)\right| \underset{d}{=} \frac{1}{n} \sup_t |\eta(F_0(t)) - nF_0(t)| = \frac{1}{n} \sup_{u\in[0,1]} |\eta(u) - un|. \tag{1}$$

Сказанное делает уместным

Определение 2. Статистика $S(X)$ называется *непараметрической*, если ее распределение не зависит от F_0 ($X \in F_0$).

Соотношения (1) означают, что статистика D является непараметрической.

Мы отмечали также (см. следствие 8.3), что статистика ω^2 наряду с D не зависит от F_0 и, стало быть, тоже является непараметрической.

Статистика χ^2, являясь асимптотически непараметрической, не будет обладать свойством непараметричности. В этом легко убедиться непосредственно на примере, положив $r = 2$, $n = 1$.

Другие примеры непараметрических статистик мы получим, если рассмотрим значения $F_n^*(\zeta_p)$, где ζ_p — квантиль порядка p, так что $nF_n^*(\zeta_p) \underset{d}{=} \eta(p)$ (см. § 6). Число r_j элементов выборки X, меньших, чем x_j, — так называемая ранговая статистика — также будет непараметрической статистикой.

Понятия непараметрической и асимптотически непараметрической статистик весьма полезны в теории проверки статистических гипотез (см. главу 3), так как распределение этих статистик, знать которое необходимо при построении критериев, достаточно вычислить лишь один раз (например, для равномерного распределения F_0) и оно будет годиться для любых других распределений выборки.

§ 10.* Сглаженные эмпирические распределения. Эмпирические плотности

В § 2 мы каждой выборке X поставили в соответствие распределение $\mathbf{P}_n^*$, которое назвали эмпирическим и которое представляет собой сумму n атомарных распределений, сосредоточенных в точках $x_1, \ldots, x_n$. Это распределение обладает рядом замечательных свойств, описанных в предыдущих параграфах. Однако использованное нами определение $\mathbf{P}_n^*$ далеко не единственно возможное, а в ряде случаев и не самое естественное. Существуют и другие подходы к определению $\mathbf{P}_n^*$, при которых изученные выше полезные свойства эмпирических распределений не только полностью сохраняются, но и дополняются рядом новых.

Мы ограничимся здесь обсуждением вопроса о природе распределений, которые мы помещаем в точки x_i. В использованном нами определении $\mathbf{P}_n^*$ это были вырожденные распределения $\mathbf{I}_{x_i}(B)$, так что

$$\mathbf{P}_n^*(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{x_i}(B). \tag{1}$$

В этом случае эмпирическое распределение оказывается сингулярным относительно меры Лебега и, стало быть, не имеет плотности. Это может оказаться неудобным в тех случаях, когда нам заранее известно, что исходное распределение $\mathbf{P}$ имеет плотность. При этом условии было бы желательно иметь такое гладкое эмпирическое распределение $\mathbf{P}_n^*$, для которого вместе со сходимостью $\mathbf{P}_n^* \to \mathbf{P}$ во всех установленных выше смыслах имела бы место также сходимость плотностей $f_n^* \to f$, где f_n^* и f есть плотности, соответствующие $\mathbf{P}_n^*$ и $\mathbf{P}$.

Этого нетрудно добиться следующим образом. Пусть $\mathbf{Q}$ — какое-нибудь распределение, *имеющее плотность*. Положим

$$\mathbf{P}_n^{**}(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Q}\left(\frac{B - x_i}{h_n}\right), \tag{2}$$

где $\dfrac{B-x}{h}$ есть множество точек $y \in \mathcal{X}$, для которых $x + yh \in B$; $h_n \to 0$ при $n \to \infty$.

Очевидно, что $\mathbf{P}_n^{**}(B)$ есть не что иное, как «средняя сумма» распределений $\mathbf{Q}$, сжатых до размеров h_n и «посаженных» в точки x_i. Определение (2) обобщает (1). Формула (1) получается из (2), если положить $\mathbf{Q} = \mathbf{I}_0$, так как $\mathbf{I}_{x_i}(B) = \mathbf{I}_0(B - x_i) = \mathbf{I}_0\left(\dfrac{B - x_i}{h_n}\right)$ при любой последовательности $\{h_n\}$.

Отметим следующие свойства распределения $\mathbf{P}_n^{**}$, которое будем называть *сглаженным эмпирическим распределением.*

1. Распределение $\mathbf{P}_n^{**}$ есть свертка распределений $\mathbf{P}_n^*$ и $\mathbf{Q}(B/h_n)$, а

$$\mathbf{P}_n(B) = \mathbf{E}\mathbf{P}_n^{**}(B) = \int \mathbf{Q}\left(\frac{B-y}{h_n}\right)\mathbf{P}(dy)$$

есть свертка распределений $\mathbf{P}$ и $\mathbf{Q}(B/h_n)$. Другими словами, $\mathbf{P}_n(B)$ есть распределение случайной величины $\xi + h_n\eta$, где $\xi \in \mathbf{P}$, $\eta \in \mathbf{Q}$. Из теорем непрерывности следует, что при $h_n \to 0$

$$\mathbf{P}_n \Rightarrow \mathbf{P}. \tag{3}$$

Напомним, что для распределения $\mathbf{P}_n^*$ мы имели точное равенство

$$\mathbf{E}\mathbf{P}_n^* = \mathbf{P}.$$

2. Распределение $\mathbf{P}_n^{**}$ будет удовлетворять теоремам типа теоремы Гливенко–Кантелли. Действительно, в этом случае сходимость (3) будет означать равномерную сходимость распределений по всем интервалам. Ограничиваясь для простоты одномерным случаем, будем иметь ($F_n^{**}(x)$, $F_n(x)$ и $Q(x)$ обозначают функции распределения, соответствующие $\mathbf{P}_n^{**}$, $\mathbf{P}_n$ и $\mathbf{Q}$)

$$F_n^{**} - F(x) = \int Q\left(\frac{x-y}{h_n}\right) dF_n^*(y) - F(x) =$$

$$= -\int F_n^*(y)\, d_y Q\left(\frac{x-y}{h_n}\right) - F(x) =$$

$$= F_n(x) - F(x) - \int \left(F_n^*(y) - F(y)\right) d_y Q\left(\frac{x-y}{h_n}\right).$$

Здесь, как уже отмечалось, разность $F_n(x) - F(x) \to 0$ равномерно по x, а интеграл, присутствующий в правой части, не превосходит

$$\sup_y \left|F_n^*(y) - F(y)\right| \underset{\text{п.н.}}{\longrightarrow} 0.$$

3. Преимущество $\mathbf{P}_n^{**}$ перед $\mathbf{P}_n^*$, ради которого это распределение и вводилось, состоит в том, что *это распределение имеет плотность*

$$f_n^*(x) = \frac{1}{nh_n}\sum_{i=1}^{n} q\left(\frac{x-\mathrm{x}_i}{h_n}\right) = \frac{1}{h_n}\int q\left(\frac{x-y}{h_n}\right) dF_n^*(y) \tag{4}$$

($q(x)$ есть плотность распределения $\mathbf{Q}$), которая при каждом x при $n \to \infty$, $h_n \to 0$ сближается с плотностью $f(x)$ распределения $\mathbf{P}$.

Прежде чем доказывать соответствующее утверждение, заметим, что для получения хороших результатов о сближении $f_n^*(x)$ с $f(x)$ следует пользоваться гладкими ограниченными плотностями q. При выборе, скажем, неограниченных q оценка $f_n^*(x)$ гладкой плотности $f(x)$ будет нарочито портиться. Так как выбор q находится в наших руках, то мы можем считать, что по крайней мере выполнено условие

$$d^2 = \int q^2(t)\, dt < \infty. \tag{5}$$

Теорема 1. *Если q удовлетворяет условию* (5), *$f(x)$ непрерывна и ограничена, $h_n \to 0$ при $n \to \infty$ так, что $nh_n \to \infty$, то*

$$f_n^*(x) = f_n(x) + \frac{\zeta_n(x)}{\sqrt{nh_n}}, \tag{6}$$

где $f_n(x)$ есть неслучайная функция

$$f_n(x) = \mathbf{E} f_n^*(x) = \mathbf{E} h_n^{-1} q\left(\frac{x-\mathrm{x}_1}{h_n}\right) =$$

$$= \frac{1}{h_n}\int q\left(\frac{x-t}{h_n}\right) f(t)\, dt = \int q(z) f(x - zh_n)\, dz \to f(x) \tag{7}$$

при $h_n \to 0$. *Случайные величины* $\zeta_n(x)$ *асимптотически нормальны,* $\zeta_n(x) \Subset \Phi_{0,\sigma^2(x)}$, $\sigma^2(x) = f(x)d^2$.

Доказательство. Сумма в (4) есть сумма независимых одинаково распределенных случайных величин в схеме серий, при этом $f_n(x) = \mathbf{E}f_n^*(x)$ представлено в (7). Положим

$$\xi_{k,n} = \frac{1}{\sqrt{nh_n}}\left[q\left(\frac{x-\mathrm{x}_k}{h_n}\right) - h_n f_n(x)\right].$$

Тогда

$$f_n^*(x) - f_n(x) = \frac{1}{\sqrt{nh_n}}\sum_{k=1}^{n}\xi_{k,n}, \quad \mathbf{E}\xi_{k,n} = 0,$$

$$\mathbf{E}\xi_{k,n}^2 = \frac{1}{n}\left[\mathbf{E}\frac{1}{h_n}q^2\left(\frac{x-\mathrm{x}_k}{h_n}\right) - h_n f_n^2(x)\right],$$

$$\mathbf{E}\frac{1}{h_n}q^2\left(\frac{x-\mathrm{x}_k}{h_n}\right) = \frac{1}{h_n}\int q^2\left(\frac{x-t}{h_n}\right)f(t)\,dt =$$

$$= \int q^2(z)f(x-zh_n)\,dz \to f(x)\int q^2(z)\,dz = f(x)d^2. \quad (8)$$

Таким образом, $\mathbf{E}\xi_{k,n}^2 \sim f(x)d^2/n$, если $f(x) > 0$. Условие Линдеберга в нашем случае имеет форму

$$n\mathbf{E}\left(\xi_{1,n}^2; |\xi_{1,n}| > \varepsilon\right) \to 0 \quad (9)$$

при $n \to \infty$ и любом $\varepsilon > 0$. Так как

$$h_n f_n^2(x) \to 0,$$

$$n\xi_{1,n}^2 \leqslant 2\left(q^2\left(\frac{x-\mathrm{x}_1}{h_n}\right) + h_n f_n^2(x)\right),$$

то для выполнения (9) достаточно, чтобы

$$\mathbf{E}\left(\frac{1}{h_n}q^2\left(\frac{x-\mathrm{x}_1}{h_n}\right);\quad q\left(\frac{x-\mathrm{x}_1}{h_n}\right) > \varepsilon\sqrt{nh_n}\right) \to 0.$$

Это соотношение имеет место, так как левая часть его равна (ср. с (8))

$$\int\limits_{q(z)>\varepsilon\sqrt{nh_n}} q^2(z)f(x-zh_n)\,dz \leqslant c \int\limits_{q(z)>\varepsilon\sqrt{nh_n}} q^2(z)\,dz \to 0.$$

Таким образом, к случайной величине $\zeta_n(x) = \sum_{k=1}^{n}\xi_{k,n}$ применима центральная предельная теорема. Это доказывает теорему 1. $\lhd$

В рассматриваемой задаче естественно возникает вопрос об оптимальном выборе h_n и функции $q(t)$. Однако ответ на него зависит от свойств гладкости $f(x)$. В самом деле, пусть, например, $f(x)$ положительна лишь на конечном интервале и дважды непрерывно дифференцируема с фиксированным значением $\varphi = \int (f''(x))^2\,dx$. Предположим также, что $\int zq(z)\,dz = 0$ (это всегда так для симметричных $q(z)$) и что $D^2 = \int z^2 q(z)\,dz < \infty$. Тогда

$$f_n(x) = \int q(z) f(x - zh_n)\,dz =$$

$$= \int q(z)\left[f(x) - zh_n f'(x) + \frac{z^2 h_n^2}{2} f''(x) + o\big(z^2 h_n^2\big)\right] dz =$$

$$= f(x) + \frac{h_n^2 f''(x)}{2}\int z^2 q(z)\,dz + o\big(h_n^2\big).$$

Мы видим, что

$$f_n^*(x) - f(x) = \frac{D^2 h_n^2 f''(x)}{2} + \frac{\zeta_n(x)}{\sqrt{nh_n}} + o\big(h_n^2\big),$$

$$\mathbf{E}\big[f_n^*(x) - f(x)\big]^2 = \left(\frac{D^2 h_n^2 f''(x)}{2}\right)^2 + \frac{d^2 f(x)}{nh_n} + o\big(h_n^4\big). \tag{10}$$

Минимизация этого выражения по h_n и q даст, в силу асимптотической нормальности $\zeta_n(x)$, наименьший возможный «разброс» $f_n^*(x)$ около значения $f(x)$. Однако минимизирующие значения h_n и q при этом будут зависеть от x через неизвестные значения $f(x)$ и $f''(x)$. Чтобы избавиться от этого эффекта и получить оптимальность «в среднем», естественно рассмотреть интеграл

$$\int \mathbf{E}\big[f_n^*(x) - f(x)\big]^2\,dx, \tag{11}$$

главная часть которого будет равна $(D^2 h_n^2/2)^2\varphi + d^2/nh_n$ (это получится, если в (10) убрать $o\big(h_n^4\big)$).

Минимум этого выражения достигается при $h_n = (d^2/nD^4\varphi)^{1/5}$. При таком выборе h_n интеграл (11) будет равен

$$\frac{5}{4}\varphi^{1/5}(Dd^2)^{4/5} n^{-4/5} + o(n^{-4/5}),$$

$$f_n^*(x) - f(x) = \left(\frac{Dd^2}{n\varphi}\right)^{2/5}\left(\frac{f''(x)}{2} + f(x)\sqrt{\varphi}\,\xi_n\right) + o(n^{-2/5}), \tag{12}$$

$$\xi_n \Rightarrow \mathbf{\Phi}_{0,1}.$$

Таким образом, скорость сходимости здесь составляет лишь $n^{-2/5}$ в отличие от скорости $n^{-1/2}$, которая имеет место для сходимости функций распределения. Это естественный факт, так как в оценке значения $f(x)$ принимает участие, грубо говоря, не вся выборка, а лишь те наблюдения, которые сосредоточились в некоторой убывающей окрестности точки x.

Выражение (12) позволяет оптимальным образом выбрать и функцию $q(z)$, т. е. функцию, для которой минимизируется Dd^2. Считая, не ограничивая общности, что $D = 1$, получим задачу минимизации $d^2 = \int q^2(z)\,dz$ при условиях $\int q(z)\,dz = \int z^2 q(z)\,dz = 1$, $\int zq(z)\,dz = 0$.

Отметим, что если f имеет непрерывные производные более высокого порядка $2m > 2$, то можно получать и более высокие скорости сходимости к нулю разности $f_n^*(x) - f(x)$. При этом, однако, надо использовать *обобщенные* распределения $\mathbf{Q}$, «плотность» которых q может принимать значения обоих знаков и позволяет удовлетворять условиям $\int z^{2m} q(z)\,dz = 1$, $\int z^j q(z)\,dz = 0$ при всех $1 \leqslant j \leqslant 2m-1$. В этом случае путем прежних рассуждений мы сможем получить скорость сходимости порядка $n^{-2m/(4m+1)} = n^{-1/2+1/2(4m+1)}$, которая будет тем лучше, чем больше m. Этот факт объясняется тем, что для более гладких $f(x)$ к оценке значения $f(x)$ привлекаются элементы выборки, лежащие во все более широких окрестностях точки x.

С другой стороны, выбирая гладкие функции $q(z)$, мы можем обеспечить возможность оценивать не только плотности $f(x)$, но и их производные. В этом также можно убедиться с помощью приведенных выше рассуждений.

Функции $f_n^*(x)$ вида (4) называют часто оценками Розенблата–Парзена плотности $f(x)$ или *ядерными оценками* $f(x)$. Функции $q(z)$ называют при этом *ядрами*. На практике часто используют «прямоугольные» ядра, т. е. полагают

$$q(z) = \begin{cases} 1 & \text{при } z \in \left[-\frac{1}{2}, \frac{1}{2}\right], \\ 0 & \text{при } z \notin \left[-\frac{1}{2}, \frac{1}{2}\right]. \end{cases}$$

Иногда поступают еще проще — разбивают вещественную прямую на маленькие интервалы Δ_j (длиной h_n) и полагают $f_n^*(x) = \nu_j/nh_n$ при $x \in \Delta_j$, где ν_j — число элементов выборки, попавших в Δ_j. Такая функция $f_n^*(x)$ называется *гистограммой* выборки. Нетрудно проверить, что если $f(x)$ непрерывна, то гистограмма $f_n^*(x)$ наряду с функцией, определенной в (4), также обладает свойством сходимости $f_n^*(x) \xrightarrow[P]{} f(x)$, если только $h_n \to 0$, $nh_n \to \infty$.

Глава 2

ТЕОРИЯ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ

Параграф 12 содержит описание наиболее распространенных параметрических семейств распределений и их основных свойств.

В § 13–16 излагаются основные методы получения точечных оценок.

В § 17, 18 обсуждаются подходы к сравнению оценок.

Параграфы 19–30 посвящены методам построения оптимальных (в том или ином смысле) оценок. Выделены следующие 4 направления:

1) (§ 19–21, 30) Байесовский и минимаксный подходы к построению оптимальных оценок. Параграфы 19, 20 носят вспомогательный характер и содержат определения и изложение основных свойств условных математических ожиданий и условных распределений.

2) (§ 22–25) Построение оптимальных (эффективных) оценок с помощью принципов достаточности и несмещенности.

3) (§ 26, 27, 30–32) Построение оптимальных (эффективных) оценок на основе неравенства Рао–Крамера.

4) (§ 28, 29) Использование соображений инвариантности.

В § 31, 33–37 изучаются асимптотические свойства отношения правдоподобия. На этой основе устанавливается асимптотическая оптимальность оценок максимального правдоподобия. Результаты этих параграфов составляют также основу теории асимптотически оптимальных критериев, развитой в главе 3.

Параграфы 39, 40 посвящены интервальному оцениванию.

§ 11. Предварительные замечания

Как мы уже отмечали в предыдущей главе, исходным объектом статистических исследований является выборка

$$X_n = (\mathrm{x}_1, \ldots, \mathrm{x}_n), \quad \mathrm{x}_i \in \mathcal{X},$$

из распределения $\mathbf{P}$, которое полностью или частично неизвестно. В математической статистике традиционно выделяют в качестве основных два следующих класса задач.

1. *Оценка неизвестных параметров.*

2. *Проверка статистических гипотез.*

Задачи первого класса возникают, когда по выборке $X = X_n$ нужно оценить какую-нибудь неизвестную числовую характеристику θ распределения $\mathbf{P}$ (оно ведь неизвестно). То есть для заданного функционала

$$\theta = \theta(\mathbf{P})$$

от распределения $\mathbf{P}$ мы должны указать функцию от выборки (или, что то же, статистику)

$$\theta^* = \theta_n^*(X_n),$$

предназначенную для использования вместо параметра θ в качестве его приближения. Мы видели в предыдущей главе, что предпосылки для этого существуют. Статистику θ^* называют *оценкой* параметра θ. Разумеется, оценок для параметра θ может быть очень много. Теорема 3.1 показывает, что, например, для оценки функционала $\theta = \theta(\mathbf{P})$ вида

$$\theta = \int g(x)\, dF(x)$$

естественно использовать статистику

$$\theta^* = \frac{1}{n} \sum_{i=1}^{n} g(\mathrm{x}_i).$$

Но можно, конечно, рассматривать и другие оценки, скажем L-статистики

$$\theta^* = \frac{1}{n - \nu_1 - \nu_2} \sum_{j=\nu_1+1}^{n-\nu_2} g(\mathrm{x}_{(j)})$$

(см. пример 8.5), где $\mathrm{x}_{(j)}$, $j = 1, \ldots, n$, — элементы вариационного ряда и т. д. В качестве θ^* можно брать и значения, не зависящие от выборки. Можно положить, например, $\theta^* \equiv 0$, хотя это не всегда разумно и совсем плохо, если множество возможных значений θ не содержит 0.

В связи с последним замечанием отметим, что часто в постановке задачи об оценивании содержится указание на то, каким является множество Θ возможных значений параметра θ. Например, если оценивается доля θ какого-нибудь минерала в руде, то ясно, что $\theta \in [0, 1]$.

Во многих случаях бывает заранее известно также, что распределение $\mathbf{P}$ выборки X не может быть произвольным, а принадлежит какому-то определенному семейству распределений $\mathcal{P}$.

К задачам оценки параметров относится пример 1 из введения.

Задачи второго класса имеют дело с проверкой того или иного предположения (гипотезы) о неизвестном распределении $\mathbf{P}$. Например, мы можем проверять гипотезу о том, что $\mathbf{P}$ имеет тот или иной заданный вид. К этому типу задач относится пример 2 из введения.

Позже мы увидим, что качественной разницы между задачами первого класса (теории оценок) и второго класса (проверка статистических гипотез) не существует.

В этой главе мы приведем постановки задач и подходы, которые тесно связаны с результатами предыдущей главы и которые можно назвать «чисто статистическими» в отличие от более общих теоретико-игровых подходов, изложенных в главе 6.

В известной мере чисто статистические подходы выражают существо методов математической статистики. Исторически они были осознаны значительно раньше, чем более общие методы. Что же касается их применения, то, по-видимому, человек пользовался ими, явно или неявно, на протяжении всего пути процесса познания.

Все это оправдывает отдельное изложение чисто статистических подходов несмотря на то, что некоторые моменты этого изложения в рамках более общих концепций можно рассматривать как частные случаи. Одновременно

мы обнаружим некоторую недостаточность чисто статистического подхода для более точных постановок задач. Это поможет нам понять целесообразность других точек зрения.

§ 12. Некоторые параметрические семейства распределений и их свойства

Рассмотрим некоторые семейства распределений, зависящих от параметров (или *параметрические семейства распределений*), которые часто возникают в приложениях и будут появляться в дальнейшем изложении как по существу его, так и в качестве иллюстраций.

1. Нормальное распределение на прямой. Символом $\mathbf{\Phi}_{\alpha,\sigma^2}$ мы будем обозначать нормальное распределение с параметрами (α, σ^2), т. е. распределение с плотностью

$$\varphi_{\alpha,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\alpha)^2/2\sigma^2},$$

так что

$$\mathbf{\Phi}_{\alpha,\sigma^2}(B) = \int\limits_B \varphi_{\alpha,\sigma^2}(x)\,dx.$$

Если $\xi \in \mathbf{\Phi}_{0,1}$ и $k \geqslant 0$ — целое число, то, очевидно,

$$\mathbf{E}\xi^{2k+1} = 0.$$

Для моментов четного порядка, пользуясь заменой $x = \sqrt{2u}$, находим

$$\mathbf{E}\xi^{2k} = \frac{2}{\sqrt{2\pi}} \int\limits_0^\infty x^{2k} e^{-x^2/2}\,dx = \frac{2^{k+1}}{\sqrt{2\pi}} \int\limits_0^\infty u^k e^{-u} \frac{du}{\sqrt{2u}} = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right),$$

где

$$\Gamma(\lambda) = \int\limits_0^\infty x^{\lambda-1} e^{-x}\,dx \tag{1}$$

есть Г-функция, $\Gamma(\lambda) = (\lambda - 1)\Gamma(\lambda - 1)$, $\Gamma(1/2) = \sqrt{\pi}$, так что

$$\mathbf{E}\xi^{2k} = (2k-1)!! = (2k-1)(2k-3)\ldots 1.$$

Такой же результат мы получили бы, дифференцируя $2k$ раз характеристическую функцию $e^{-t^2/2}$ в точке $t = 0$.

2. Многомерное нормальное распределение. В многомерном случае $\mathcal{X} = \mathbb{R}^m$ символ $\mathbf{\Phi}_{\alpha,\sigma^2}$ означает нормальное распределение в $\mathbb{R}^m$ с вектором математических ожиданий $\alpha = (\alpha_1, \ldots, \alpha_m)$ и матрицей центральных вторых моментов $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, \ldots, m$. Если A есть матрица, обратная

к σ^2 (в тех случаях, когда она существует), то плотность $\varphi_{\alpha,\sigma^2}(x)$ в $\mathbb{R}^m$ распределения $\mathbf{\Phi}_{\alpha,\sigma^2}$ будет иметь вид (см. [17, § 7.6])

$$\varphi_{\alpha,\sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}(x-\alpha)A(x-\alpha)^\top\right),$$

где $x^\top$ есть транспонированный вектор. Напомним также (мы этим фактом уже пользовались в § 7), что х. ф. величины $\xi \in \mathbf{\Phi}_{\alpha,\sigma^2}$ равна

$$\mathbf{E}e^{it\xi^\top} = \exp\left(it\alpha^\top - \frac{1}{2}t\sigma^2 t^\top\right),$$

где $t = (t_1, \ldots, t_m)$ есть вектор в $\mathbb{R}^m$.

3. Гамма-распределение. Символ $\mathbf{\Gamma}_{\alpha,\lambda}$ будет означать так называемое гамма-распределение (или Г-распределение) с параметрами (α, λ). Плотность $\gamma_{\alpha,\lambda}(x)$ этого распределения зависит от двух параметров $\alpha > 0$ и $\lambda > 0$ и равна (см. [17, § 7.7])

$$\gamma_{\alpha,\lambda}(x) = \begin{cases} \dfrac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x > 0, \\ 0, & x \leqslant 0, \end{cases} \tag{2}$$

где $\Gamma(\lambda)$ есть Г-функция, определенная в (1). Характеристическая функция Г-распределения имеет вид [17]

$$\int_0^\infty e^{itx}\gamma_{\alpha,\lambda}(x)\,dx = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}. \tag{3}$$

Если $\xi \in \mathbf{\Gamma}_{\alpha,\lambda}$, то

$$\mathbf{E}\xi^t = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty x^{\lambda+t-1}e^{-\alpha x}\,dx = \frac{\alpha^{-t}}{\Gamma(\lambda)} \int_0^\infty y^{\lambda+t-1}e^{-y}\,dy = \frac{\alpha^{-t}\Gamma(\lambda+t)}{\Gamma(\lambda)}. \tag{4}$$

Такой же результат при целых $t > 0$ можно было бы получить, дифференцируя характеристическую функцию. Полагая $t = 1, 2$, находим

$$\mathbf{E}\xi = \frac{\lambda}{\alpha}, \quad \mathbf{D}\xi = \frac{\lambda}{\alpha^2}. \tag{5}$$

Из формул (3), (4) видно, что параметр α играет роль масштаба, так что

$$\frac{\eta}{\alpha} \in \mathbf{\Gamma}_{\alpha,\lambda}, \quad \text{если} \quad \eta \in \mathbf{\Gamma}_{1,\lambda}.$$

В силу этого обстоятельства многие свойства Г-распределения достаточно изучать при каком-нибудь одном значении α, например при $\alpha = 1$ или $\alpha = 1/2$. Второе значение часто будет для нас более удобным, так как распределение $\mathbf{\Gamma}_{1/2,\lambda}$ играет важную самостоятельную роль в математической

статистике и называется распределением «хи-квадрат» (или χ^2-распределением).

4. Распределение «хи-квадрат» $\mathbf{H}_k$ с k степенями свободы. Так называется распределение $\mathbf{H}_k = \mathbf{\Gamma}_{1/2,k/2}$ при целых $k > 0$. Мы сохраним это название за распределением $\mathbf{H}_k$ и при произвольных $k > 0$. Характеристическая функция распределения $\mathbf{H}_k$ равна в силу (3)

$$(1 - 2it)^{-k/2}.$$

Отметим следующие три свойства распределения $\mathbf{H}_k$.

1) *Если* η_i *независимы,* $\eta_i \in \mathbf{H}_{k_i}$, $i = 1, \ldots, s$, *то*

$$\sum_{i=1}^{s} \eta_i \in \mathbf{H}_k, \quad k = \sum_{i=1}^{s} k_i.$$

Это свойство сразу вытекает из вида характеристической функции распределения $\mathbf{H}_k$.

2) *Если* $\xi \in \mathbf{\Phi}_{\alpha,\sigma^2}$, *где* $\mathbf{\Phi}_{\alpha,\sigma^2}$ *есть* k*-мерное нормальное распределение с невырожденной матрицей вторых моментов* σ^2, *то*

$$Q(\xi) = (\xi - \alpha)\sigma^{-2}(\xi - \alpha)^{\top} \in \mathbf{H}_k.$$

Действительно, характеристическая функция случайной величины $Q(\xi)$ равна

$$\mathbf{E}e^{itQ(\xi)} = \frac{\sqrt{|\sigma^{-2}|}}{(2\pi)^{k/2}} \int \exp\left(-\frac{1}{2}Q(x)(1 - 2it)\right) dx_1 \ldots dx_k.$$

Произведя замену переменных $x_j\sqrt{1 - 2it} = y_j$, мы получим выражение

$$(1 - 2it)^{-\frac{k}{2}} \frac{\sqrt{|\sigma^{-2}|}}{(2\pi)^{k/2}} \int e^{-\frac{1}{2}Q(y)}\, dy_1 \ldots dy_k = (1 - 2it)^{-\frac{k}{2}},$$

что и требовалось доказать. Независимость интеграла в левой части от изменения области интегрирования вытекает из аналитичности подынтегральной функции и ее быстрого убывания при $|y| \to \infty$.

Из сказанного вытекает, что распределение $\mathbf{H}_k$ имеет случайная величина

$$\chi^2 = \xi_1^2 + \cdots + \xi_k^2,$$

где ξ_j независимы, $\xi_j \in \mathbf{\Phi}_{0,1}$. Термин «число степеней свободы» связан именно с этим представлением.

3) Так как $\mathbf{E}\xi_1^2 = 1$, $\mathbf{E}\xi_1^4 = 3$, $\mathbf{D}\xi_1^2 = 2$ для $\xi_1 \in \mathbf{\Phi}_{0,1}$, то *в силу центральной предельной теоремы при* $k \to \infty$

$$\frac{\chi^2 - k}{\sqrt{2k}} \Rightarrow \mathbf{\Phi}_{0,1}. \tag{6}$$

Отсюда и из теорем непрерывности §5 следует, что наряду с (6)

$$\sqrt{2\chi^2} - \sqrt{2k-1} \Subset \mathbf{\Phi}_{0,1}.$$

Эта сходимость служит основой для приближенного (при больших k и x) равенства $\mathbf{H}_k((0,x)) \approx \Phi(\sqrt{2x} - \sqrt{2k-1})$, $\Phi(x) = \mathbf{\Phi}_{0,1}((-\infty, x))$, которое, как правило, оказывается более точным, чем приближение $\mathbf{H}_k((0,x)) \approx \approx \Phi\left(\dfrac{x-k}{\sqrt{2k}}\right)$, вытекающее из (6).

Отметим еще один частный случай Г-распределения, часто встречающийся в приложениях.

5. Экспоненциальное распределение. Это есть распределение $\mathbf{\Gamma}_{\alpha,1}$ с плотностью

$$\alpha e^{-\alpha x}, \quad x > 0.$$

Из формул (5) получаем для $\xi \Subset \mathbf{\Gamma}_{\alpha,1}$

$$\mathbf{E}\xi = \frac{1}{\alpha}, \quad \mathbf{D}\xi = \frac{1}{\alpha^2}.$$

Рассмотрим теперь некоторые распределения, связанные с нормальным и Г-распределениями и играющие важную роль в математической статистике. В отличие от предыдущих эти распределения нам прежде не встречались.

6. Распределение Фишера с числом степеней свободы k_1, k_2. Так называется распределение случайной величины

$$\varkappa = \frac{k_2\eta_1}{k_1\eta_2},$$

где η_j независимы, $\eta_j \Subset \mathbf{H}_{k_j}$, $j = 1, 2$. Нам будет удобнее изучать распределение $\mathbf{F}_{k_1,k_2}$ случайной величины

$$\zeta = \frac{\eta_1}{\eta_2},$$

которая, очевидно, связана с $\varkappa$ соотношением $\varkappa = \dfrac{k_2\zeta}{k_1}$. Причина введения в определении $\varkappa$ нормирующего множителя $\dfrac{k_2}{k_1}$ состоит в том, что при $k_1 \to \infty$, $k_2 \to \infty$ в силу закона больших чисел будет выполняться $\varkappa \underset{p}{\longrightarrow} 1$.

Из свойств Г-распределений следует, что распределение ζ останется тем же при $\eta_j \Subset \mathbf{\Gamma}_{\alpha,k_j/2}$ и любом $\alpha > 0$ и что ζ при целых k_j допускает представление

$$\zeta = \frac{\xi_1^2 + \cdots + \xi_{k_1}^2}{\zeta_1^2 + \ldots \zeta_{k_2}^2},$$

где случайные величины ξ_j, ζ_k независимы, $\xi_j \Subset \mathbf{\Phi}_{0,1}$, $\zeta_k \Subset \mathbf{\Phi}_{0,1}$.

Найдем плотность распределения $\mathbf{F}_{k_1,k_2}$. Имеем

$$\mathbf{P}(\zeta < x) = \iint\limits_{u/v<x} \mathbf{\Gamma}_{1,\lambda_1}(du)\mathbf{\Gamma}_{1,\lambda_2}(dv) = \int\limits_{v=0}^{\infty}\int\limits_{u=0}^{vx} \frac{u^{\lambda_1-1}v^{\lambda_2-1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)}e^{-u-v}\,du\,dv;$$

$$f_{(\zeta)}(x) = \frac{d\mathbf{P}(\zeta<x)}{dx} = \int\limits_0^{\infty} \frac{(vx)^{\lambda_1-1}v^{\lambda_2-1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)}e^{-v-vx}v\,dv =$$

$$= \frac{x^{\lambda_1-1}}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\int\limits_0^{\infty} v^{\lambda_1+\lambda_2-1}e^{-v(1+x)}\,dv = \frac{x^{\lambda_1-1}}{(1+x)^{\lambda_1+\lambda_2}}\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}. \quad (7)$$

Требуемая плотность получится, очевидно, если подставить сюда $\lambda_j = k_j/2$. Нетрудно найти моменты случайной величины ζ (если они существуют):

$$\mathbf{E}\zeta^l = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\int\limits_0^{\infty}\frac{x^{\lambda_1+l-1}}{(1+x)^{\lambda_1+\lambda_2}}\,dx = \frac{\Gamma(\lambda_1+l)\Gamma(\lambda_2-l)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}. \quad (8)$$

В частности, при $l = 1, 2$ находим

$$\mathbf{E}\zeta = \frac{\lambda_1}{\lambda_2-1}, \quad \mathbf{E}\zeta^2 = \frac{\lambda_1(\lambda_1+1)}{(\lambda_2-1)(\lambda_2-2)}.$$

Распределение Фишера называют иногда также распределением Снедекора. Это связано с тем, что Фишер предложил использовать и табулировал, собственно, не распределение $\varkappa$, а распределение случайной величины $\frac{1}{2}\ln\varkappa$. Распределение же $\varkappa$ было несколько позже подробно табулировано Снедекором.

В дальнейшем в формулировках результатов, связанных с распределением Фишера (см., например, § 61, 62), нам будет удобнее использовать распределение $\mathbf{F}_{k_1,k_2}$. При этом следует иметь в виду, что при отыскании квантилей распределения $\mathbf{F}_{k_1,k_2}$ с помощью таблиц распределения Фишера надо вводить поправочный множитель: если f_ε, $\varkappa_\varepsilon$ — квантили порядка $1-\varepsilon$ распределений $\mathbf{F}_{k_1,k_2}$ и Фишера, соответственно, то $f_\varepsilon = \dfrac{k_2\varkappa_\varepsilon}{k_1}$.

7. Распределение Стьюдента*) $\mathbf{T}_k$ с k степенями свободы. По определению это есть распределение случайной величины

$$t = \frac{\xi_0}{\sqrt{\frac{1}{k}\left(\xi_1^2+\cdots+\xi_k^2\right)}},$$

*) Стьюдент — псевдоним В. С. Госсета.

где ξ_j независимы, $\xi_j \in \mathbf{\Phi}_{0,1}$, $j = 0, \ldots, k$. Очевидно, что $-t$ имеет то же самое распределение и, стало быть, распределение Стьюдента симметрично относительно начала координат. Далее,

$$t^2 = \frac{k\xi_0^2}{\xi_1^2 + \cdots + \xi_k^2} = \frac{k\eta_1}{\eta_2},$$

где η_j независимы, $\eta_1 \in \mathbf{H}_1$, $\eta_2 \in \mathbf{H}_k$. Это значит, что t^2/k имеет распределение Фишера. Рассмотрим случайную величину $\tau = \sqrt{\zeta}$, $\zeta = \eta_1/\eta_2$, $\eta_j \in \mathbf{H}_{k_j}$. Так как $\mathbf{P}(\tau < x) = \mathbf{P}(\zeta < x^2)$, то плотность $f_{(\tau)}(x)$ случайной величины τ будет равна

$$f_{(\tau)}(x) = 2x f_{(\zeta)}(x^2) = 2x \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \frac{x^{2\lambda_1 - 2}}{(1 + x^2)^{\lambda_1 + \lambda_2}} =$$

$$= \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \frac{2x^{2\lambda_1 - 1}}{(1 + x^2)^{\lambda_1 + \lambda_2}}, \quad \lambda_j = \frac{k_j}{2}, \quad x > 0.$$

Отсюда при $\lambda_1 = 1/2$, $\lambda_2 = k/2$ очевидным образом можно получить плотность $|t|/\sqrt{k}$. Так как распределение t симметрично, то для плотности $f_{(t)}(x)$ случайной величины t окончательно получаем

$$f_{(t)}(x) = \frac{\Gamma((k+1)/2)}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}. \tag{9}$$

Ясно, что все моменты t нечетного порядка (если они существуют) равны нулю. Для моментов четного порядка $2l$ имеем в силу (8)

$$\mathbf{E}t^{2l} = k^l \mathbf{E}\zeta^l = k^l \frac{\Gamma(\lambda_1 + l)\Gamma(\lambda_2 - l)}{\Gamma(\lambda_1)\Gamma(\lambda_2)},$$

где следует положить $\lambda_1 = 1/2$, $\lambda_2 = k/2$, $2l < k$. При $l = 1$ получаем

$$\mathbf{E}t^2 = \frac{k}{k-2}.$$

По своей форме функция $f_{(t)}(x)$ напоминает плотность нормального закона. Более того, с ростом k

$$f_{(t)}(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

что означает сходимость $t \Rightarrow \mathbf{\Phi}_{0,1}$ при $k \to \infty$. Однако $f_{(t)}(x)$ имеет более «толстые хвосты», поскольку функция (9) убывает с ростом $|x|$ значительно медленнее, чем $e^{-x^2/2}$, так что при всех $b > 0$

$$\mathbf{T}_k((-b, b)) < \mathbf{\Phi}_{0,1}((-b, b)). \tag{10}$$

При этом разница между правой и левой частью в (10) при небольших k может быть существенной.

Сходимость $t = \sqrt{k\xi_0}/\sqrt{\eta_2}$ к нормальному закону читатель может доказать и другим путем с помощью теорем непрерывности. Например, достаточно заметить, что $\dfrac{\eta_2}{k} = \dfrac{1}{k}(\xi_1^2 + \cdots + \xi_k^2) \underset{\text{п.н.}}{\longrightarrow} 1$; и, стало быть, $t \underset{\text{п.н.}}{\longrightarrow} \xi_0$, $t \Rightarrow \xi_0$.

8. Бэта-распределение (B-распределение). Так называется распределение $\mathbf{B}_{\lambda_1,\lambda_2}$ с плотностью

$$f_{(\beta)}(x) = \begin{cases} \dfrac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} x^{\lambda_1-1}(1-x)^{\lambda_2-1}, & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases}$$

Своим названием оно обязано бета-функции

$$\mathrm{B}(\lambda_1,\lambda_2) = \int_0^1 x^{\lambda_1-1}(1-x)^{\lambda_2-1}\,dx = \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1+\lambda_2)}.$$

Бета-распределение связано с гамма-распределением и с распределением Фишера путем следующего утверждения.

Если η_j независимы, $\eta_j \in \mathbf{\Gamma}_{\alpha,\lambda_j}$ (или $\eta_j \in \mathbf{H}_{2\lambda_j}$), то

$$\beta = \frac{\eta_1}{\eta_1+\eta_2} = \frac{\zeta}{\zeta+1} \in \mathbf{B}_{\lambda_1,\lambda_2},$$

где $\zeta = \eta_1/\eta_2 \in \mathbf{F}_{2\lambda_1,2\lambda_2}$.

Доказательство этого утверждения весьма просто, так как в силу (7) $\mathbf{P}(\beta < x) = \mathbf{P}\left(\zeta < \dfrac{x}{1-x}\right)$,

$$f_{(\beta)}(x) = f_{(\zeta)}\left(\frac{x}{1-x}\right)\left(\frac{x}{1-x}\right)' = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)}\left(\frac{x}{1-x}\right)^{\lambda_1-1}(1-x)^{\lambda_2+\lambda_2-2} =$$
$$= \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} x^{\lambda_1-1}(1-x)^{\lambda_2-1}, \quad x \in [0,1].$$

Для моментов случайной величины β имеем

$$\mathbf{E}\beta^l = \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \int_0^1 x^{\lambda_1+l-1}(1-x)^{\lambda_2-1}\,dx = \frac{\Gamma(\lambda_1+\lambda_2)\Gamma(\lambda_1+l)}{\Gamma(\lambda_1)\Gamma(\lambda_1+\lambda_2+l)}.$$

При $l = 1, 2$ находим

$$\mathbf{E}\beta = \frac{\lambda_1}{\lambda_1+\lambda_2}, \quad \mathbf{E}\beta^2 = \frac{\lambda_1(\lambda_1+1)}{(\lambda_1+\lambda_2)(\lambda_1+\lambda_2+1)}.$$

9. Равномерное распределение. Частным случаем B-распределения является равномерное на $[0,1]$ распределение, которое получится, если положить $\lambda_1 = \lambda_2 = 1$.

Символом $\mathbf{U}_{a,b}$ мы будем обозначать равномерное распределение на отрезке $[a,b]$, так что $\mathbf{B}_{1,1} = \mathbf{U}_{0,1}$.

С помощью B-распределения можно описать распределение членов вариационного ряда $\mathrm{x}_{(k)}$ выборки X.

Теорема 1. *Если $X \in \mathbf{P}$ есть выборка из распределения $\mathbf{P}$ с непрерывной функцией распределения F, то*

$$\mathrm{y}_{(k)} = F(\mathrm{x}_{(k)}) \in \mathbf{B}_{k,n-k+1}.$$

Доказательство. Так как $\mathrm{y}_k = F(\mathrm{x}_k) \in \mathbf{U}_{0,1}$, то $\mathrm{y}_{(k)} = F(\mathrm{x}_{(k)})$ можно рассматривать как член вариационного ряда выборки $Y \in \mathbf{U}_{0,1}$. Найдем $\mathbf{P}(\mathrm{y}_{(k)} \in (u, u+du))$. Событие $\{\mathrm{y}_{(k)} \in (u, u+du)\}$ можно представить как объединение непересекающихся событий

$$A_j = \{\mathrm{y}_j \in (u, u+du),\ \mathrm{y}_j = \mathrm{y}_{(k)}\},$$

которые происходят, когда y_j попадает в $(u, u+du)$ (вероятность этого равна du), $k-1$ из оставшихся $n-1$ наблюдений попадает в область $(0,u)$, $n-k$ наблюдений — в область $(u,1)$. Следовательно,

$$\mathbf{P}(A_j) = C_{n-1}^{k-1} u^{k-1}(1-u)^{n-k}du,$$

$$\mathbf{P}(\mathrm{y}_{(k)} \in (u, u+du)) = nC_{n-1}^{k-1} u^{k-1}(1-u)^{n-k}du.$$

Это и значит, что плотность $\mathrm{y}_{(k)}$ существует и равна

$$\frac{n!}{(k-1)!(n-k)!}u^{k-1}(1-u)^{n-k} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}u^{k-1}(1-u)^{n-k}. \quad \triangleleft$$

На основании теоремы 1 можно без большого труда получить и *предельное* распределение членов вариационного ряда, когда объем выборки X неограниченно возрастает. Мы остановимся здесь лишь на одном результате, вытекающем из теорем непрерывности.

Теорема 2. *Если $a = \dfrac{k}{n+1} \to a_0 \in (0,1)$ при $n \to \infty$, то*

$$\mathrm{y}_{(k)} = a + \frac{\sqrt{a_0(1-a_0)}}{\sqrt{n}}\xi_n, \quad \xi_n \Rightarrow \mathbf{\Phi}_{0,1}.$$

Доказательство. В силу теоремы 1 $\mathrm{y}_{(k)} \in \mathbf{B}_{k,n-k+1}$ и, стало быть, в силу свойств B-распределения справедливо представление

$$\mathrm{y}_{(k)} \underset{d}{=} \frac{\eta_1}{\eta_1+\eta_2}, \quad \eta_j \in \mathbf{H}_{k_j}, \quad k_1 = 2k, \quad k_2 = 2(n-k+1).$$

Положим для удобства $a_1 = a$, $a_2 = 1 - a$ и предположим, что $a = a_0$ фиксировано. Тогда, очевидно, $k_j/(n+1) = 2a_j$, $j = 1, 2$, и в силу свойств распределения χ^2

$$\eta_j = k_j + \sqrt{2k_j}\,\xi_n^{(j)}, \quad \xi_n^{(j)} \Rightarrow \xi^{(j)} \in \mathbf{\Phi}_{0,1};$$

$$\mathrm{y}_{(k)} = \frac{a_1 + \sqrt{\dfrac{a_1}{n+1}}\,\xi_n^{(1)}}{a_1 + a_2 + \sqrt{\dfrac{a_1}{n+1}}\,\xi_n^{(1)} + \sqrt{\dfrac{a_2}{n+1}}\,\xi_n^{(2)}}.$$

Остается воспользоваться теоремой непрерывности 5.3А при

$$H(t) = \frac{t_1}{t_1 + t_2}, \quad b_n = \frac{1}{\sqrt{n+1}}, \quad \eta_n^{(j)} = \sqrt{a_j}\,\xi_n^{(j)}.$$

Так как η_j (а стало быть, и $\xi_n^{(j)}$) независимы и

$$\frac{\partial H}{\partial t_1} = \frac{t_2}{(t_1 + t_2)^2}, \quad \frac{\partial H}{\partial t_2} = -\frac{t_1}{(t_1 + t_2)^2},$$

то мы получаем

$$(\mathrm{y}_{(k)} - a_1)\sqrt{n+1} \Rightarrow a_2\sqrt{a_1}\,\xi^{(1)} - a_1\sqrt{a_2}\,\xi^{(2)} \underset{d}{=} \sqrt{a_1 a_2}\,\xi, \quad \xi \in \mathbf{\Phi}_{0,1}.$$

Если a зависит от n, то следует воспользоваться замечанием 5.1. $\triangleleft$

Следствие 1. *Если $a = k/(n+1) \to a_0 \in (0,1)$ и непрерывная функция F непрерывно дифференцируема в точке $\zeta_0 = F^{-1}(a_0)$ (квантиль порядка a_0), то*

$$\mathrm{x}_{(k)} = \zeta + \frac{\sqrt{a_0(1-a_0)}\,\xi_n}{f(\zeta_0)\sqrt{n}}, \quad \xi_n \Rightarrow \mathbf{\Phi}_{0,1}, \tag{11}$$

где $\zeta = F^{-1}(a)$ — квантиль порядка a, $f(x) = F'(x)$.

Это утверждение получается сразу из теоремы непрерывности 5.3 (с учетом замечания 5.1), если воспользоваться представлением

$$\mathrm{x}_{(k)} = F^{-1}(\mathrm{y}_{(k)}) = F^{-1}\left(a + \sqrt{\frac{a_0(1-a_0)}{n}}\,\xi_n\right)$$

и тем фактом, что $\dfrac{dF^{-1}(x)}{dx} = \dfrac{1}{f(F^{-1}(x))}$.

Замечание 1. Утверждение (11) несколько обобщает утверждение следствия 8.1. Его можно обобщить и в другом направлении. Пусть при $x \to \zeta$

$$|F(x) - F(\zeta)| \sim c|x - \zeta|^\gamma, \quad \gamma > 0.$$

Тогда нетрудно видеть, что при $y \to a$

$$|F^{-1}(y) - F^{-1}(a)| \sim \left|\frac{y-a}{c}\right|^{1/\gamma}$$

и что, стало быть,

$$(\mathrm{x}_{(k)} - \zeta)n^{1/2\gamma} \Rightarrow (a_0(1-a_0))^{1/2\gamma}\left|\frac{\xi}{c}\right|^{1/\gamma}\operatorname{sign}\xi, \quad \xi \in \mathbf{\Phi}_{0,1}. \tag{12}$$

При $\gamma = 1$, $c = f(\zeta)$ отсюда следует (11).

10. Распределение Коши $\mathbf{K}_{\alpha,\sigma}$ с параметрами (α, σ). Так называется распределение с плотностью

$$k_{\alpha,\sigma}(x) = \frac{\sigma}{\pi[\sigma^2 + (x-\alpha)^2]} = \frac{1}{\pi\sigma}\frac{1}{1 + \left(\dfrac{x-\alpha}{\sigma}\right)^2}.$$

Как и в случае нормального закона, параметры α и σ здесь являются соответственно параметрами сдвига и масштаба. Форма распределения $\mathbf{K}_{0,1}$ очень похожа на форму $\mathbf{\Phi}_{0,1}$, однако $k_{0,1}$, как и плотность распределения Стьюдента, имеет значительно более «толстые хвосты» (т. е. более медленное убывание при $|x| \to \infty$), так что распределение $\mathbf{K}_{0,1}$ не имеет даже конечного математического ожидания. В [17, гл. 8] отмечалось, что распределения $\mathbf{K}_{\alpha,\sigma}$, как и нормальные распределения, обладают свойством устойчивости. Характеристическая функция $\varkappa_{0,1}(t)$ распределения $\mathbf{K}_{0,1}$ равна

$$\varkappa_{0,1}(t) = \exp(-|t|),$$

поэтому

$$\varkappa_{\alpha,\sigma}(t) = \exp\{i\alpha t - \sigma|t|\},$$

$$\varkappa_{\alpha_1,\sigma_1}(t)\varkappa_{\alpha_2,\sigma_2}(t) = \exp\{i(\alpha_1+\alpha_2)t - (\sigma_1+\sigma_2)|t|\},$$

так что свертка $\mathbf{K}_{\alpha_1,\sigma_1}$ и $\mathbf{K}_{\alpha_2,\sigma_2}$ равна $\mathbf{K}_{\alpha_1+\alpha_2,\sigma_1+\sigma_2}$.

Нетрудно видеть, что $\mathbf{K}_{0,1} = \mathbf{T}_1$.

В приложениях часто приходится иметь дело с разного рода функциями от нормально распределенных случайных величин. Одна из них — экспоненциальная функция, с которой связано так называемое логнормальное распределение.

11. Логнормальное распределение $\mathbf{L}_{\alpha,\sigma^2}$. Мы будем говорить, что $\eta \in \mathbf{L}_{\alpha,\sigma^2}$, если $\ln\eta \in \mathbf{\Phi}_{\alpha,\sigma^2}$. Другими словами, $\eta = e^{\xi}$, где $\xi \in \mathbf{\Phi}_{\alpha,\sigma^2}$. Отсюда видно, что распределение $\mathbf{L}_{\alpha,\sigma^2}$ сосредоточено на положительной полуоси.

Плотность $\eta \in \mathbf{L}_{\alpha,\sigma^2}$ в силу формул для плотности функции от случайной величины (см. [17, § 3.2]) равна

$$\varphi_{\alpha,\sigma^2}(\ln x)x^{-1}.$$

Кроме того, находим

$$\mathbf{E}\eta = \int \exp y \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\alpha)^2}{2\sigma^2}\right] dy = \exp\frac{(\alpha+\sigma^2)^2-\alpha^2}{2\sigma^2} \times$$

$$\times \int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\alpha-\sigma^2)^2}{2\sigma^2}\right) dy = \exp\left(\frac{\alpha+\sigma^2}{2}\right),$$

$$\mathbf{E}\eta^2 = \int \exp(2y) \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\alpha)^2}{2\sigma^2}\right] dy = \exp(2\alpha+2\sigma^2).$$

12. Вырожденное распределение. Символ $\mathbf{I}_a$ (это обозначение нами уже использовалось в § 2) будет означать вырожденное распределение, сосредоточенное в точке a.

В общем случае, когда рассматривается *произвольное семейство распределений*, зависящих от параметра θ (скалярного или векторного), мы будем использовать обозначение $\mathbf{P}_\theta$. Само семейство будет обозначаться символом

$$\{\mathbf{P}_\theta\}_{\theta\in\Theta},$$

где Θ — множество возможных значений параметра θ. Такие же обозначения будут применяться к семействам распределений 1–12. Например, $\{\mathbf{\Phi}_{\alpha,1}\}_{\alpha\in\mathbb{R}}$ будет означать семейство всех нормальных распределений с единичной дисперсией.

Распределения 1–11 абсолютно непрерывны относительно меры Лебега. Введем теперь обозначения для трех хорошо известных *дискретных* распределений (абсолютно непрерывных относительно считающей меры $\mu(B)$: $\mu(B) = k$, если B содержит k целочисленных точек).

13. Распределение Бернулли $\mathbf{B}_p^n$. По определению $\xi \in \mathbf{B}_p^n$ (n целое, $p \in [0,1]$), если

$$\mathbf{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}, \quad 0 \leqslant k \leqslant n.$$

14. Распределение Пуассона $\mathbf{\Pi}_\lambda$. Это распределение определяется равенством

$$\mathbf{\Pi}_\lambda(B) = \sum_{\substack{k\in B,\\ k\geqslant 0}} \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0.$$

15. Полиномиальное распределение. Мы будем обозначать его $\mathbf{B}_p^n$, где $n > 0$ — целое число, $p = (p_1, \ldots, p_r)$, $p_j \geqslant 0$, $\sum_{j=1}^r p_j = 1$. Для целочисленного случайного вектора $\nu = (\nu_1, \ldots, \nu_r)$ мы будем писать $\nu \in \mathbf{B}_p^n$, если для $k = (k_1, \ldots, k_r)$, $k_j \geqslant 0$, $\sum_{j=1}^r k_j = n$ справедливо

$$\mathbf{P}(\nu = k) = \frac{n!}{k_1! \ldots k_r!} p_1^{k_1} \ldots p_r^{k_r}.$$

Распределение $\mathbf{B}_p^n$ соответствует последовательности n независимых испытаний, в каждом из которых наступает один из r несовместных возможных исходов $A_1, \ldots, A_r$; при этом вероятность появления исхода A_j в одном испытании равна p_j. Координаты ν_j вектора ν означают частоты появления событий A_j после n испытаний (см., например, [17, § 3.3]). Очевидно, что при каждом $j = 1, \ldots, r$

$$\nu_j \in \mathbf{B}_{p_j}^n.$$

Исход j-го испытания в описанном эксперименте можно описать r-координатным вектором x_j, у которого $r-1$ координат равны нулю и одна координата равна единице. Номер этой координаты есть номер события, которое произошло при j-м испытании. Очевидно, что $\nu = \sum_{j=1}^{n} \mathrm{x}_j$. Относительно выборки X, составленной из $\mathrm{x}_1, \ldots, \mathrm{x}_n$, нам будет удобнее писать

$$X \in \mathbf{B}_p,$$

где $\mathbf{B}_p = \mathbf{B}_p^1$. Пространство $\mathcal{X}$ для такой выборки, очевидно, конечно и состоит из r точек. Если $p = (p_1, p_2)$, $p_1 + p_2 = 1$, то мы получим схему Бернулли, для которой мы будем использовать те же обозначения, отождествляя $\mathbf{B}_{(p_1,p_2)}$ с $\mathbf{B}_{p_1} = \mathbf{B}_{p_1}^1$ (см. п. 13). В общем случае распределение $\mathbf{B}_p$ зависит на самом деле лишь от $(r-1)$-мерного параметра $(p_1, \ldots, p_{r-1})$, так что вместо индекса p можно было бы писать $(p_1, \ldots, p_{r-1})$.

Многие из рассмотренных выше распределений, например распределения $\mathbf{\Phi}_{0,1}$, $\mathbf{H}_k$, $\mathbf{F}_{k_1,k_2}$, $\mathbf{T}_k$, $\mathbf{\Pi}_\lambda$, табулированы в руководствах по математической статистике и в специальных таблицах (см., например, [10]).

§ 13. Точечное оценивание. Основной метод получения оценок. Состоятельность, асимптотическая нормальность

1. Метод подстановки. Состоятельность. Понятие оценки было введено нами в § 11. Формально оценка — это то же самое, что и статистика, т. е. любая измеримая функция θ^* от выборки. Неформально смысл, вкладываемый в этот новый термин, состоит в том, что оценками θ^* мы называем лишь статистики, предназначенные для использования вместо неизвестного параметра θ. Другими словами, θ^* есть некоторое приближение для θ, основанное на выборке. Величину θ^* называют также *точечной* оценкой для θ в отличие от *интервальных* оценок, которые будут рассмотрены позже.

Задание оценки обычно предполагает задание функций (от выборок X_n), определенных при всех возможных значениях n. Поэтому в дальнейшем термин «оценка», будет означать семейство статистик $\theta^* = \theta_n^*(X_n)$, определенных при всех $n = 1, 2, \ldots$, где θ^* есть функция на $\mathcal{X}^n$, или, что то же, одну функцию $\theta^* = \theta^*(n, X_\infty)$, определенную на произведении множества целых чисел и $\mathcal{X}^\infty$.

В соответствии с § 11 мы будем считать, что в постановке задачи об оценивании определено множество Θ возможных значений параметра θ и семейство $\mathcal{P}$ возможных распределений $\mathbf{P}$ выборки X (это могут быть, скажем, лишь нормальные распределения $\mathbf{\Phi}_{\alpha,1}$ или распределения Пуассона $\mathbf{\Pi}_\lambda$, для которых требуется оценить неизвестные параметры α, λ). Если какие-либо

ограничения на θ (на $\mathbf{P}$) отсутствуют, то можно считать, что Θ ($\mathcal{P}$) совпадает с евклидовым пространством соответствующей размерности (с множеством всех распределений).

Если для обозначения параметра вместо θ используется какая-нибудь другая буква, например λ, то оценки этого параметра будут обозначаться тем же способом: добавлением к λ верхнего индекса «звездочка». Например, для параметра α нормального закона естественно рассматривать оценку

$$\alpha^* = \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i.$$

Выборочные моменты, используемые для оценки

$$\mathbf{E}\mathrm{x}_1 = \int x = \mathbf{P}(dx), \quad \mathbf{D}\mathrm{x}_1 = \int (x - \mathbf{E}\mathrm{x}_1)^2 \mathbf{P}(dx),$$

имеют свои специальные традиционные обозначения

$$\overline{\mathrm{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^2.$$

Мы уже отмечали, что для данного параметра можно указать сколь угодно много разных оценок, и прежде чем обсуждать, каким образом в каждой конкретной ситуации следует сравнивать их достоинства, мы остановимся на некоторых общих «регулярных» методах их построения.

Эти методы объединяют в себе наиболее разумные подходы к проблеме оценивания и позволят нам получать в дальнейшем наилучшие в том или ином смысле оценки.

В основе почти всех приемов оценивания лежит следующий основной метод, который можно было бы назвать *методом подстановки эмпирического распределения* (или, для краткости, *методом подстановки.*)

Пусть $X_n \in \mathbf{P}$ и неизвестный параметр θ представим в виде некоторого функционала G от распределения $\mathbf{P}$:

$$\theta = G(\mathbf{P}).$$

Пусть, далее, $\mathbf{P}_n^*$, как и прежде, означает эмпирическое распределение. Тогда метод подстановки предписывает в качестве оценки θ^* взять функцию

$$\theta^* = G\big(\mathbf{P}_n^*\big).$$

Такие оценки мы будем называть *оценками по методу подстановки* или просто *оценками подстановки.*

Функционал G иногда бывает задан в неявном виде как решение некоторого уравнения $H(\theta, \mathbf{P}) = 0$, разрешимого относительно θ. В этом случае оценками подстановки в соответствии с основным определением мы будем называть любое решение уравнения $H\big(\theta, \mathbf{P}_n^*\big) = 0$.

Если известно, что множество возможных значений параметра $\theta \in \mathbb{R}^k$ ограничено областью Θ из $\mathbb{R}^k$, не совпадающей с $\mathbb{R}^k$, то эту информацию можно учесть при построении оценок подстановки. Допустим, что область Θ замкнута, и пусть $\mathcal{P}$ есть множество возможных распределений выборки X,

$\Theta = \{G(\mathbf{P})\}_{\mathbf{P}\in\mathcal{P}}$. Определим функционал $G_1(\mathbf{P})$ для произвольного $\mathbf{P}$ как значение $t \in \Theta$, для которого достигается

$$\min_{t\in\Theta} |t - G(\mathbf{P})| = |G_1(\mathbf{P}) - G(\mathbf{P})|, \tag{1}$$

так что $G_1(\mathbf{P})$ есть точка из Θ, ближайшая к $G(\mathbf{P})$. Так как $G_1(\mathbf{P}) = G(\mathbf{P}) = \theta$, если $\mathbf{P} \in \mathcal{P}$, то оценка

$$\theta^* = G_1\big(\mathbf{P}_n^*\big) \tag{2}$$

вместе с $G\big(\mathbf{P}_n^*\big)$ будет оценкой подстановки, при этом множество возможных значений θ^* будет принадлежать Θ.

Про оценки (1), (2) мы будем говорить, что они получены *сужением метода подстановки.*

Пусть, например, оценивается параметр α нормального распределения $\mathbf{\Phi}_{\alpha,1}$, и нам заранее известно, что $\alpha \in [0,1]$. Тогда может оказаться, что оценка $\alpha^* = \overline{\mathrm{x}} \notin [0,1]$ (очевидно, что $\overline{\mathrm{x}} = \int t\, dF_n^*(t)$ есть оценка подстановки). Сужение метода подстановки рекомендует в качестве оценки взять точку из $[0,1]$, ближайшую к $\overline{\mathrm{x}}$.

Отметим теперь, что в сформулированном виде метод подстановки не всегда имеет смысл. Дело в том, что исходный функционал G может оказаться не определенным на множестве эмпирических распределений. Пусть, например, заранее известно, что распределение $\mathbf{P}$ принадлежит классу $\mathcal{P}$ абсолютно непрерывных относительно меры Лебега распределений, так что каждое $\mathbf{P} \in \mathcal{P}$ имеет плотность f. Нас интересует значение

$$\theta = G(\mathbf{P}) = \int f^2(x)\, dx = \int \left(\frac{d\mathbf{P}}{dx}\right)^2 dx.$$

Ясно, что в этом случае $G\big(\mathbf{P}_n^*\big)$ не имеет смысла, так как $\mathbf{P}_n^*$ есть дискретное распределение. В таких случаях метод подстановки всегда можно естественным образом модифицировать так, чтобы он сохранил свое существо. В приведенном примере, где $G(\mathbf{P})$ есть функционал от плотности f, следует в качестве θ^* в соответствии с методом подстановки рассмотреть значение $G\big(\mathbf{P}_n^{**}\big)$, где $\mathbf{P}_n^{**}$ — сглаженное эмпирическое распределение (см. § 10), обеспечивающее сходимость эмпирической плотности к $f(x)$.

Может оказаться также, что в некоторых случаях $G\big(\mathbf{P}_n^*\big)$ будет иметь смысл не для всех X_n, а лишь для $X_n \in A_n$, где $\mathbf{P}(X_n \in A_n) \to 1$ при $n \to \infty$. Это обстоятельство для существа дальнейшего изложения роли играть не будет, и для определенности можно положить $G\big(\mathbf{P}_n^*\big) = 0$ для $X_n \notin A_n$.

В этом параграфе мы для простоты будем предполагать, что $G\big(\mathbf{P}_n^*\big)$ имеет смысл при всех $X_n \in \mathcal{X}^n$ и что θ^* есть случайный вектор, т. е. функция $G\big(\mathbf{P}_n^*\big)$ осуществляет измеримое отображение $\mathcal{X}^n$ в $\mathbb{R}^k$, где k есть размерность θ.

Принцип подстановки представляет собой весьма естественный подход к задаче, поскольку, как мы уже знаем, распределение $\mathbf{P}_n^*$ неограниченно сближается с $\mathbf{P}$ с ростом n.

Пусть $X_n = [X_\infty]_n$.

Определение 1. Оценка $\theta^* = \theta_n^*(X_n)$ (или последовательность $\theta_n^*(X)$) называется *состоятельной*, если

$$\theta^* \underset{P}{\longrightarrow} \theta$$

при $n \to \infty$.

Оценка θ^* называется *сильно состоятельной*, если при $n \to \infty$

$$\theta^* \underset{\text{п.н.}}{\longrightarrow} \theta.$$

Пусть F, как обычно, есть функция распределения, соответствующая $\mathbf{P}$.

Теорема 1. *Пусть $\theta = G(\mathbf{P})$ и функционал G принадлежит одному из двух классов: либо он представим в виде*

$$G(\mathbf{P}) = h\left(\int g(x)\,dF(x)\right), \tag{I}$$

где h — функция, непрерывная в точке $a = \int g(x)\,dF_0(x)$ (функционал I *типа), либо в виде*

$$G(\mathbf{P}) = G_1(F), \tag{II}$$

где функционал G_1 непрерывен в точке F_0 в равномерной метрике (функционал II *типа). Тогда если $X \in F_0$, то $\theta^* = G(\mathbf{P}_n^*)$ есть сильно состоятельная оценка:*

$$\theta^* \underset{\text{п.н.}}{\longrightarrow} \theta.$$

Утверждение этой теоремы немедленно следует из теоремы 4.1.

2. Асимптотическая нормальность. Одномерный случай

Определение 2. Оценка θ^* параметра θ называется *асимптотически нормальной* (а. н.) *с коэффициентом* $\sigma^2 > 0$, если $(\theta^* - \theta)\sqrt{n} \Rightarrow \Phi_{0,\sigma^2}$.

Последнее соотношение может читаться также следующим образом: оценка θ^* а. н. с параметрами $(\theta, \sigma^2/n)$. Коэффициент σ^2 мы будем иногда называть также *коэффициентом рассеивания* (или *разброса*).

Пусть θ^* есть оценка подстановки параметра $\theta = G(\mathbf{P})$ и выполнено (I), т. е.

$$\theta^* = h\left(\frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i)\right) \tag{3}$$

есть статистика I типа. Тогда из результатов § 7 вытекает следующее утверждение. Предположим, что θ — скалярный параметр, а g — скалярная функция.

Теорема 2. *Пусть $X \in F_0$, h дифференцируема в точке $a = \int g(x)\,dF_0(x)$, $0 < |h'(a)| < \infty$, $\int g^2(x)\,dF_0(x) < \infty$. Тогда θ^* — а. н. оценка с коэффициентом*

$$\sigma^2 = [h'(a)]^2 \int (g(x) - a)^2\,dF_0(x).$$

Примеры, рассмотренные в § 7, можно взять в качестве иллюстраций и к этой теореме, так как статистики, рассмотренные в них, используются в качестве оценок.

Аналогичным образом мы могли бы, пользуясь результатами § 8, получить условия асимптотической нормальности оценок, являющихся статистиками II типа. Читатель может получить нужные утверждения, если воспользуется теоремой 8.1 без каких-либо ее изменений, потребовав, однако, чтобы в ее формулировке выполнялось $k = 1$, а производная g была такой, что $g(F_0, w^0) \Subset \mathbf{\Phi}_{0,\sigma^2}$.

3. Асимптотическая нормальность. Случай многомерного параметра

Определение 2А. Оценка $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ называется *а. н. оценкой* $\theta = (\theta_1, \ldots, \theta_k)$ *с матрицей* σ^2, если

$$(\theta^* - \theta)\sqrt{n} \Rightarrow \mathbf{\Phi}_{0,\sigma^2}, \tag{4}$$

где $\mathbf{\Phi}_{0,\sigma^2}$ есть k-мерное нормальное распределение с нулевым вектором математических ожиданий и матрицей вторых моментов $\sigma^2 = \|\sigma_{ij}\|$. Плотность этого распределения равна (см. § 12)

$$\varphi_{0,\sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{k/2}} e^{-\frac{1}{2}xAx^\top},$$

где A — матрица, обратная к σ^2, $\mathbf{x} = (x_1, \ldots, x_k)$. Матрицу σ^2 мы будем называть также *матрицей рассеивания* (или *разброса*).

Если θ^* есть оценка подстановки и она является статистикой I типа (т. е. представима в виде (3), где g, вообще говоря, вместе с θ^* и h есть векторная функция), то для выяснения условий асимптотической нормальности можно воспользоваться теоремой 7.1А и замечанием к ней. Мы получим тогда следующее утверждение.

Теорема 2А. *Пусть* $\theta^* \in \mathbb{R}^k$ *определяется равенством* (I), *где* $g = (g_1, \ldots, g_s) \in \mathbb{R}^s$, *а вектор-функция* $h(t) = (h_1(t), \ldots, h_k(t))$, $t = (t_1, \ldots, t_s)$, *имеет в точке* $a = (a_1, \ldots, a_s)$, $a_j = \int g_j(x)\, dF_0(x)$, *частные производные* $\dfrac{\partial h_l}{\partial t_j}(a)$, $l = 1, \ldots, k$, $j = 1, \ldots, s$. *Тогда если* $X \Subset F_0$, *то*

$$(\theta^* - \theta)\sqrt{n} \Rightarrow \xi H^\top,$$

где $\xi = (\xi_1, \ldots, \xi_s) \Subset \mathbf{\Phi}_{0,d^2}$ *есть нормально распределенный вектор с нулевым средним и матрицей вторых моментов* $d^2 = \|d_{ij}\|$, $d_{ij} = \mathbf{E}(g_i(\mathrm{x}_1) - a_i)(g_j(\mathrm{x}_1) - a_j)$, $i, j = 1, \ldots, s$; $H = \|h_{ij}\|$ *есть матрица размера* $k \times s$ *с элементами* $h_{ij} = \dfrac{\partial h_i}{\partial t_j}(a)$, $i = 1, \ldots, k$; $j = 1, \ldots, s$.

Это означает, в свою очередь, что θ^* при выполнении условий теоремы 2А является а. н. оценкой с матрицей $\sigma^2 = Hd^2H^\top = \mathbf{E}H\xi^\top\xi H^\top$. Отметим, что матрицы σ^2 и d^2 здесь имеют, вообще говоря, разные размерности (k и s).

§ 14. Реализация метода подстановки в параметрическом случае. Метод моментов. M-оценки

Пусть $X \in \mathbf{P}_\theta$, где $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$ есть известное нам семейство распределений $\mathbf{P}_\theta$, зависящих от параметра θ. Неизвестный параметр θ из множества Θ может быть в наших рассмотрениях как скалярным, так и векторным. Например, если $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$, то $\theta = (\alpha, \sigma^2)$ двумерно, а множество Θ может быть как полуплоскостью $\{-\infty < \alpha < \infty, \sigma \geqslant 0\}$, так и какой-нибудь ее частью.

Математическое ожидание и дисперсию статистики $S = S(X)$ по распределению $\mathbf{P}_\theta$ будем обозначать соответственно $\mathbf{E}_\theta S$, $\mathbf{D}_\theta S$.

Мы рассмотрим ниже несколько методов оценивания, каждый из которых можно трактовать как реализацию принципа подстановки эмпирического распределения.

1. Метод моментов. Одномерный случай. Выберем $g(x)$ так, чтобы функция

$$m(\theta) = \mathbf{E}_\theta g(\mathrm{x}_1) = \int g(x)\mathbf{P}_\theta(dx) \tag{1}$$

была монотонной и непрерывной. Область $m(\Theta)$ значений $m(\theta)$, $\theta \in \Theta$, имеет ту же «природу», что и Θ. Если, например, Θ есть отрезок вещественной оси, то $m(\Theta)$ также будет отрезком.

Очевидно, что уравнение $m(\theta) = t$ однозначно и непрерывно разрешимо в области $m(\Theta)$ относительно θ: $\theta = m^{-1}(t)$, и что (1) эквивалентным образом можно записать в виде

$$\theta = m^{-1}\left(\int g(x)\mathbf{P}_\theta(dx)\right). \tag{2}$$

Предположим для простоты, что

$$\overline{g} \equiv \int g(x)\, d\mathbf{P}_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i) \in m(\Theta)$$

при всех $X \in \mathcal{X}^n$.

Определение 1. Оценкой по *методу моментов* называется оценка

$$\theta^* = m^{-1}(\overline{g}). \tag{3}$$

Если $\overline{g} \notin m(\Theta)$, то можно положить согласно (13.1), (13.2)

$$\theta^* = m^{-1}(\overline{g}_0),$$

где $\overline{g}_0 \in m(\Theta)$ есть ближайшая к $\overline{g}$ точка из $m(\Theta)$.

Нетрудно видеть, что это есть оценка по принципу подстановки. Выбор функции $m(\theta)$ позволил нам выразить θ в виде функционала (2). Ясно также, что оценка (3) является статистикой I типа, так что в силу теоремы 13.1 оценки по методу моментов будут сильно состоятельны. Если к тому же

функция m дифференцируема в точке θ, $\int g^2(x)\mathbf{P}_\theta(dx) < \infty$, то по теореме 13.2 оценка по методу моментов будет а. н. с коэффициентом рассеивания $(m'(\theta))^{-2}\mathbf{D}_\theta g(\mathrm{x}_1)$.

Метод моментов был предложен К. Пирсоном (в несколько более частном виде) и исторически является первым регулярным методом построения оценок.

Свое название «метод моментов» получил потому, что его существо состоит в приравнивании друг другу «теоретических» и эмпирических моментов (математических ожиданий) величины $g(\mathrm{x}_1)$: ведь оценка (3) есть не что иное, как решение уравнения

$$m(\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\mathrm{x}_i). \tag{4}$$

Можно добавить также, что в качестве $g(x)$ часто выбирают функцию $g(x) = x$ или $g(x) = x^k$, $k > 1$, так что наше уравнение превращается в уравнение для обычных моментов.

Равенство (4) можно рассматривать так же, как результат приравнивания среднего значения величины $g(\mathrm{x}_1)$ «по пространству» ее среднему значению «по времени».

Неоднозначность метода моментов, как и всего принципа подстановки, здесь видна особенно хорошо: ведь выбор функции $g(x)$ почти ничем не ограничен.

Пример 1. Пусть $X \Subset \mathbf{\Gamma}_{\alpha,1}$ и α не известно. Построим оценки по методу моментов с двумя простейшими функциями $g(x)$: $g_1(x) = x$ и $g_2(x) = x^2$. Справедливы следующие равенства (см. п. 12.2):

$$m_1(\alpha) = \mathbf{E}_\alpha g_1(\mathrm{x}_1) = \int_0^\infty x\mathbf{\Gamma}_{\alpha,1}(dx) = \frac{1}{\alpha},$$

$$m_2(\alpha) = \mathbf{E}_\alpha g_2(\mathrm{x}_1) = \int_0^\infty x^2\mathbf{\Gamma}_{\alpha,1}(dx) = \frac{2}{\alpha^2}.$$

Решая уравнения $m_1(\alpha) = \overline{\mathrm{x}}$, $m_2(\alpha) = \dfrac{1}{n}\sum_{i=1}^{n}\mathrm{x}_i^2$, мы получим оценки по методу моментов

$$\alpha^* = (\overline{\mathrm{x}})^{-1}, \quad \alpha^{**} = \left(\frac{1}{2n}\sum_{i=1}^{n}\mathrm{x}_i^2\right)^{-1/2}. \tag{5}$$

Обе эти оценки являются статистиками I типа, и мы можем описать их асимптотические свойства. На основании равенств (12.4) получаем

$$\mathbf{D}_\alpha g_1(\mathrm{x}_1) = \mathbf{D}_\alpha \mathrm{x}_1 = \frac{1}{\alpha^2}, \quad \mathbf{D}_\alpha g_2(\mathrm{x}_1) = \mathbf{D}_\alpha \mathrm{x}_1^2 = \frac{20}{\alpha^4}.$$

Так как для первой оценки $m_1'(\alpha) = -1/\alpha^2$, а для второй $m_2'(\alpha) = -4/\alpha^2$, то на основании теорем 13.1, 13.2 мы получаем, что обе оценки α^* и α^{**} являются сильно состоятельными и а. н. с коэффициентами рассеивания соответственно

$$\frac{1}{\alpha^2}\alpha^4 = \alpha^2, \quad \frac{20}{\alpha^4}\cdot\frac{\alpha^6}{16} = \frac{5}{4}\alpha^2.$$

Видимо, оценку α^* следует предпочесть, так как ее «разброс» при больших n вокруг истинного значения α, измеряемый дисперсией предельного распределения, меньше, чем «разброс» для α^{**}.

2. Метод моментов. Многомерный случай. Совершенно аналогичным образом рассматривается случай, когда θ — многомерный параметр.

Пусть, как и прежде, k есть размерность θ. Выберем вектор-функцию $g(x) = (g_1(x), \ldots, g_k(x))$ так, чтобы уравнение

$$m(\theta) = t,$$

где

$$t = (t_1, \ldots, t_k), \quad m(\theta) = (m_1(\theta), \ldots, m_k(\theta)),$$

$$m_j(\theta) = \mathbf{E}_\theta g_j(\mathrm{x}_1) = \int g_j(x)\mathbf{P}_\theta(dx),$$

было однозначно и непрерывно разрешимо относительно $\theta = m^{-1}(t)$ в области $m(\Theta)$ значений $m(\theta)$, $\theta \in \Theta$. Допустим для простоты, что вектор

$$\overline{g} = \left(\frac{1}{n}\sum_{i=1}^n g_1(\mathrm{x}_i), \ldots, \frac{1}{n}\sum_{i=1}^n g_k(\mathrm{x}_i)\right)$$

принадлежит области $m(\Theta)$ при всех $X \in \mathcal{X}^n$.

Определение 1А. Оценка $\theta^* = m^{-1}(\overline{g})$ называется *оценкой по методу моментов*.

Как и прежде, из теоремы 13.1 следует, что такие оценки θ^* будут сильно состоятельными.

Для того чтобы имела место а. н. θ^*, нужно потребовать дополнительно, чтобы функция m была дифференцируемой, $\int g_j^2(x)\mathbf{P}_\theta(dx) < \infty$. Утверждение о предельном распределении θ^* нетрудно получить с помощью теоремы 13.2А.

Пример 2. Рассмотрим в качестве $\{\mathbf{P}_\theta\}$ семейство нормальных распределений $\mathbf{\Phi}_{\alpha,\sigma^2}$. Полагая $g_1(x) = x$, $g_2(x) = x^2$, получим следующие уравнения для метода моментов:

$$\alpha = \overline{\mathrm{x}}, \quad \sigma^2 + \alpha^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{x}_i^2,$$

решение которых есть

$$\alpha^* = \overline{\mathrm{x}}, \quad (\sigma^2)^* = \frac{1}{n}\sum_{i=1}^n \mathrm{x}_i^2 - (\overline{\mathrm{x}})^2 = S^2.$$

Мы предлагаем читателю в качестве упражнения найти оценки по методу моментов для всех параметрических семейств, приведенных в § 12.

3. M-оценивание как обобщенный метод моментов. Возможно следующее обобщение метода моментов, которое существенно расширяет рассмотренный выше класс оценок.

Пусть θ_0 — истинное значение параметра: $X \in \mathbf{P}_{\theta_0}$. Оценка по методу моментов была основана на том, что уравнение относительно θ

$$m(\theta) \equiv \int g(x)\mathbf{P}_\theta(dx) = \int g(x)\mathbf{P}(dx) \tag{6}$$

имело единственное решение $\theta = G(\mathbf{P}) = m^{-1}\left(\int g(x)\mathbf{P}(dx)\right)$, совпадающее при $\mathbf{P} = \mathbf{P}_{\theta_0}$ с θ_0, так что оценка по методу моментов $\theta^* = G(\mathbf{P}_n^*)$, как мы уже отмечали, оказывается оценкой подстановки.

Первое соотношение в (6) можно записать в виде

$$\int (g(x) - m(\theta))\mathbf{P}_\theta(dx) \equiv 0, \tag{7}$$

а уравнение (6) в виде

$$\int (g(x) - m(\theta))\mathbf{P}(dx) = 0.$$

Основываясь на этих двух соотношениях, мы можем рассмотреть теперь более общее уравнение. Пусть функция $g(x,\theta) \not\equiv 0$ такова, что

$$\int g(x,\theta)\mathbf{P}_\theta(dx) = 0 \tag{8}$$

($g(x,\theta) = g(x) - m(\theta)$ в (7)). Функцию $g(x,\theta)$, удовлетворяющую (8), можно получить из любой интегрируемой функции $g_1(x,\theta)$, положив $g(x,\theta) = g_1(x,\theta) - \int g_1(x,\theta)\mathbf{P}_\theta(dx)$. Если уравнение

$$\int g(x,\theta)\mathbf{P}(dx) = 0 \tag{9}$$

имеет единственное решение $\theta = G(\mathbf{P})$, то в силу (8) выполняется $G(\mathbf{P}_{\theta_0}) = \theta_0$. Стало быть, метод моментов допускает следующее естественное обобщение.

Определение 2. Значение θ^*, являющееся решением уравнения

$$\sum g(\mathrm{x}_i,\theta) = 0 \tag{10}$$

(ср. с (9) при $\mathbf{P} = \mathbf{P}_n^*$), называется оценкой решения θ уравнения (9) по *обобщенному методу моментов* или *M-оценкой*.

Очевидно, что M-оценка является M-статистикой (см. п. 3.4).

Здесь, как и прежде, $\theta^* = G(\mathbf{P}_n^*)$, где $G(\mathbf{P})$ есть решение (9) относительно θ. Нетрудно видеть, что определение 2 M-оценки *допускает выход за рамки параметрического случая и сохраняется в более общем случае, когда речь идет об оценке по выборке* $X \Subset \mathbf{P}$ *фукционала* $\theta = G(\mathbf{P})$, который определяется как решения уравнения (9) при некоторой функции $g(x, \theta)$; при этом $\mathbf{P}$ может быть произвольным распределением, не обязательно принадлежащим параметрическому семейству. Равенство (8) в параметрическом случае нужно лишь для справедливости равенства $G(\mathbf{P}_{\theta_0}) = \theta_0$.

Если $\theta \in \mathbb{R}^k$ — *многомерный параметр*, то в качестве g в (9), (10) следует брать вектор-функцию той же размерности, что и θ, так, чтобы уравнения (9), (10) представляли собой систему из k независимых уравнений.

Иногда M-оценки определяют иначе. Пусть $\psi(x, \theta)$ — заданная скалярная измеримая функция.

Определение 2А. *$\widehat{M}$-оценкой* θ^* называется значение θ, при котором достигается

$$\max_{\theta} \sum_{i=1}^{n} \psi(\mathrm{x}_i, \theta). \tag{11}$$

Если ψ для п. в. x дифференцируема по θ, то, полагая $g(x, \theta) = \psi'_\theta(x, \theta)$, мы получим, что $\widehat{M}$-оценка будет M-оценкой в смысле определения 2. Обратное, вообще говоря, не верно, так как M-оценки в смысле определения 2 могут быть точками локальных экстремумов $\sum \psi(\mathrm{x}_i, \theta)$. Если решение θ^* уравнения (10) единственно и в точке θ^* достигается максимум, то, очевидно, θ^* будет M- и $\widehat{M}$-оценкой.

Использование $\widehat{M}$-оценок естественно в тех случаях, когда существует функция $\psi(x, \theta)$ такая, что неизвестный параметр θ_0 максимизирует $\int \psi(x, \theta)\mathbf{P}(dx)$.

Пример 3. Среднее θ распределения $\mathbf{P}$ минимизирует $\int (x - \theta)^2 \mathbf{P}(dx)$. Поэтому $\widehat{M}$-оценкой параметра θ будет значение θ^*, минимизирующее $\sum(\mathrm{x}_i - \theta)^2$, или, что то же, решение $\theta^* = \bar{\mathrm{x}}$ уравнения $\sum(\mathrm{x}_i - \theta) = 0$, являющееся M-оценкой. Здесь $\psi(x, t) = -(x - \theta)^2$, $g(x, t) = 2(x - \theta)$.

Так как $\theta = \int x\mathbf{P}(dx)$, то сказанное согласуется, очевидно, с тем, что θ^* является также «непосредственной» оценкой подстановки $\theta^* = \int x\mathbf{P}_n^*(dx) = \bar{\mathrm{x}}$.

Аналогичное замечание справедливо и относительно медианы выборки X. Предварительно заметим, что квантиль ζ_p порядка p распределения $\mathbf{P}$ с функцией распределения F может быть определена как решение уравнения

$$\int g(x - \theta)\, dF(x) = 0,$$

где

$$g(x) = \begin{cases} -1 & \text{при } x < 0, \\ 0 & \text{при } x = 0, \\ \dfrac{p}{1-p} & \text{при } x > 0. \end{cases}$$

Поэтому выборочная квантиль ζ_p^* является M-оценкой в смысле определения 2. Она определяется как решение уравнения

$$-\sum I(\mathrm{x}_i < \theta) + \frac{p}{1-p}\sum I(\mathrm{x}_i > \theta) = 0.$$

Медиану $\zeta_{1/2}$ распределения $\mathbf{P}$ с непрерывной в точке $\zeta_{1/2}$ функцией распределения F можно рассматривать также как значение, минимизирующее $\int |x-\theta|\, dF(x)$, т. е. $\widehat{M}$-оценку. Действительно, производная этого интеграла п. в. существует и равна $\mathbf{P}((\theta,\infty)) - \mathbf{P}((-\infty,\theta))$. Она монотонно убывает с ростом θ и обращается в нуль при $\theta = \zeta_{1/2}$. Мы получили, что $\zeta_{1/2}^*$ является одновременно M- и $\widehat{M}$-оценкой. Она является решением уравнения

$$\sum I(\mathrm{x}_i > \theta) - \sum I(\mathrm{x}_i < \theta) = 0.$$

(Если n нечетно, то $\zeta_{1/2}^* = \mathrm{x}_{((n+1)/2)}$. Если n четно, то левая часть этого уравнения обращается в нуль при любом $\theta \in (\mathrm{x}_{(n/2)}, \mathrm{x}_{(n/2+1)})$; для определенности можно положить $\theta^* = \dfrac{\mathrm{x}_{(n/2+1)} + \mathrm{x}_{(n/2)}}{2} = \zeta_{1/2}^*$.)

Более содержательные примеры M-оценок рассмотрены в § 16, 33, 34. Там же можно найти примеры $\widehat{M}$-оценок, не являющихся M-оценками (см. пример 16.5).

Легко построить также примеры M-оценок, не являющихся $\widehat{M}$-оценками. Таким образом, классы M- и $\widehat{M}$-оценок близки, но не совпадают. В известном смысле $\widehat{M}$-оценки имеют более широкую область определения: $\widehat{M}$-оценки определены и в тех случаях, когда функция ψ не дифференцируема, и уравнение (10) для $g = \psi'_\theta$ не определено. Если же ψ'_θ существует, то $\widehat{M}$-оценки являются M-оценками.

M-оценки были введены П. Хьюбером в [127] как обобщение оценок максимального правдоподобия (см. § 16; об M-оценках см. также [129, 67, 102]). Основное внимание в настоящей книге уделяется регулярному случаю, когда функция ψ является гладкой. Изучение асимптотических свойств M-оценок (а стало быть, и $\widehat{M}$-оценок) посвящены п. 4–6. Для полноты картины мы приводим также условия состоятельности $\widehat{M}$-оценок в «нерегулярном» случае, когда не предполагается гладкость ψ.

4. Состоятельность $\widehat{M}$-оценок. Мы приведем здесь условия сильной состоятельности $\widehat{M}$-оценок при весьма слабых ограничениях на функцию $\psi(x,\theta)$ в (11). Введем в рассмотрение следующие условия.

(A_c) *Множество* Θ *компактно.*

Это условие вводится для упрощения. С точки зрения приложений всегда можно указать достаточно большой компакт $K = [t_0, t_1]$ такой, что искомое θ_0 наверняка принадлежит внутренности K и, следовательно, можно ограничиться поисками точки θ^* максимума (11) лишь в компакте K. При этом под θ^* можно понимать любую точку максимума (11), принадлежащую K. Если же максимум достигается вне K, скажем в точке $\theta > t_1$, то можно положить θ^* равной t_1 или равной точке максимума на K. Содержание последующих утверждений от этого не зависит.

[A_0] *Функция* $\psi(x,t)$ *такова, что*

$$\int \psi(x,t)\mathbf{P}(dx)$$

достигает своего максимума в единственной точке $t = \theta$, *так что*

$$\Psi(t,\theta,\mathbf{P}) \equiv \int (\psi(x,t) - \psi(x,\theta))\mathbf{P}(dx) < 0$$

при всех $t \neq \theta$, $t \in \Theta$.

Эти два условия мы будем предполагать выполненными на протяжении всего этого параграфа.

Далее, по выборке $X \Subset \mathbf{P}$ строится $\widehat{M}$-оценка θ^*, являющаяся точкой максимума $\sum \psi(\mathrm{x}_i, t)$, или, что то же, $\Psi\left(t,\theta,\mathbf{P}_n^*\right)$.

Нам понадобится также следующий равномерный вариант условия [A_0].

($\overline{\mathrm{A}}_0$) *Для любого* $\delta > 0$

$$\inf_{t:|t-\theta|\geqslant\delta} \Psi(t,\theta,\mathbf{P}) < -\varepsilon$$

при некотором $\varepsilon = \varepsilon(\delta) > 0$.

Очевидно, что условие ($\overline{\mathrm{A}}_0$) будет следствием (A_c), [A_0] и непрерывности $\Psi(t,\theta,\mathbf{P})$ по t. Обозначим

$$\psi^{\Delta}(x,t) = \sup_{|u|\leqslant\Delta} \psi(x, t+u)$$

и наряду с [A_0], ($\overline{\mathrm{A}}_0$) рассмотрим также условие

(A_0^{Δ}) *Для любого* $\delta > 0$ *существует* $\Delta = \Delta(\delta)$ *такое, что при всех* t, $|t-\theta| > \delta$

$$\int (\psi^{\Delta}(x,t) - \psi(x,\theta))\mathbf{P}(dx) < -\varepsilon \tag{12}$$

при некотором $\varepsilon > 0$.

Теорема 1. *Если выполнено* (A_0^{Δ}), *то* $\widehat{M}$*-оценка* θ^* *сильно состоятельна.*

Как будет показано в § 16, одного условия $(\overline{\mathrm{A}}_0)$ для состоятельности θ^* недостаточно.

Доказательство. Так как

$$\Psi(\theta^*, \theta, \mathbf{P}_n^*) \geqslant \Psi(\theta, \theta, \mathbf{P}_n^*) = 0,$$

то для доказательства теоремы нам достаточно убедиться, что с вероятностью 1

$$\limsup_{n\to\infty} \sup_{|t-\theta|\geqslant\delta} \Psi(t, \theta, \mathbf{P}_n^*) < -\varepsilon$$

при некотором $\varepsilon > 0$. (Это и будет означать, что для п. в. $X_\infty \in \mathbf{P}$, начиная с некоторого $n = n(X_\infty) < \varepsilon$, выполняется $|\theta^* - \theta| < \delta$.)

Пусть δ фиксировано, Δ удовлетворяет (12), $(\theta)^\delta$ есть δ-окрестность точки θ. Покроем множество $\Theta \setminus (\theta)^\delta$ окрестностями $\Delta_k = \{t : |t - t_k| \leqslant \Delta\}$, $k = 1, \ldots, N < \infty$, где $t_k \in \Theta$, $t_k \notin (\theta)^\delta$. Тогда по усиленному закону больших чисел в силу (12)

$$\sup_{|t-\theta|\geqslant\delta} \Psi(t, \theta, \mathbf{P}_n^*) \leqslant \max_k \sup_{t\in\Delta_k} \Psi(t, \theta, \mathbf{P}_n^*) \leqslant$$

$$\leqslant \max_k \frac{1}{n} \sum_{i=1}^{n} \sup_{\Delta_k} [\psi(\mathrm{x}_i, t) - \psi(\mathrm{x}_i, \theta)] \underset{\text{п.н.}}{\longrightarrow}$$

$$\underset{\text{п.н.}}{\longrightarrow} \max_k \mathbf{E}[\psi^\Delta(\mathrm{x}_1, t_k) - \psi(\mathrm{x}_1, t_k)] < -\varepsilon. \quad \triangleleft$$

Условие (A_0^Δ) можно эквивалентным образом представить в несколько иной форме. Обозначим $\psi^0(x, t) = \limsup_{u\to t} \psi(x, u)$.

Теорема 2. *Условие* (A_0^Δ) *эквивалентно одновременному выполнению следующих двух условий:*

(A_0^0) *для всех* $t \neq \theta$

$$\int [\psi^0(x, t) - \psi(x, \theta)]\, \mathbf{P}(dx) < 0;$$

(J) *при всех* t *и некотором* $\Delta > 0$

$$\int [\psi^\Delta(x, t) - \psi(x, \theta)] \mathbf{P}(dx) < \infty.$$

Условие (J), как, впрочем, и (A_0^0), (A_0^Δ), означает интегрируемость положительных частей подынтегральных функций. Такие функции естественно называть *интегрируемыми сверху.*

Условие (J) в силу (A_c) по существу эквивалентно ограниченности сверху интеграла

$$\int [\psi(x, \Theta) - \psi(x, \theta)] \mathbf{P}(dx) < \infty,$$

где $\psi(x, \Theta) = \sup_{t\in\Theta} \psi(x, t)$.

Доказательство теоремы 2. Тот факт, что из (A_0^{Δ}) следует (A_0^0) и (J), очевиден. Пусть теперь выполнены (A_0^0) и (J). Если допустить, что (A_0^{Δ}) не имеет места, то найдутся последовательности $t_k \to t \in \Theta$, $\Delta_k \to 0$, $\varepsilon_k \to 0$ такие, что

$$\int [\psi^{\Delta_k}(x, t_k) - \psi(x, \theta)]\mathbf{P}(dx) > -\varepsilon_k.$$

Здесь подынтегральная функция мажорируется в силу условия (J) интегрируемой сверху функцией, поэтому в силу леммы Фату

$$\limsup_{k\to\infty} \int [\psi^{\Delta_k}(x, t_k) - \psi(x, \theta)]\mathbf{P}(dx) \leqslant \int [\psi^0(x, t) - \psi(x, \theta)]\mathbf{P}(dx) < 0.$$

Мы получили противоречие, доказывающее теорему. $\triangleleft$

Более простые достаточные условия, обеспечивающие выполнение (A_0^0) и (J) для конкретных функций $\psi(x, t)$, приведены в § 16.

Асимптотическая нормальность $\widehat{M}$-оценок требует выполнения условий гладкости по t функции $\psi(x, t)$. Поэтому ее естественно изучать в рамках п. 6, посвященного асимптотической нормальности M-оценок.

5. Состоятельность M-оценок. Сначала мы ограничимся рассмотрением одномерного случая $\Theta \subset \mathbb{R}$ и вновь для простоты будем предполагать везде, что выполнено условие (A_c) о том, что $\Theta \subset [t_0, t_1]$ — компакт. Положим

$$s(\theta) = \mathbf{E}g(\mathrm{x}_1, \theta) = \int g(x, \theta)\mathbf{P}(dx). \tag{13}$$

Нам понадобятся следующие условия.

(M_1) *Пусть*

$$\sup_{\theta\in\Theta} |g(x, \theta)| \leqslant \gamma(x), \quad \int \gamma(x)\mathbf{P}(dx) < \infty.$$

(M_2) *Функция $g(x, \theta)$ для п. в. x непрерывна по θ.*

(M_3) *Уравнение $s(\theta) = 0$ имеет внутри Θ единственное решение θ_0.*

Теорема 3. *При выполнении условий (M_1)–(M_3) M-оценка θ^* сильно состоятельна.*

Для доказательства теоремы нам понадобится следующее утверждение. Положим

$$s_n^*(t) = \frac{1}{n}\sum g(\mathrm{x}_i, t).$$

Лемма 1. *При выполнении (M_1), (M_2) и $n \to \infty$*

$$\sup_{t\in\Theta} \left|s_n^*(t) - s(t)\right| \underset{\text{п.н.}}{\longrightarrow} 0,$$

функция $s(t)$ непрерывна.

Доказательство. Чтобы упростить рассуждения, везде в дальнейшем мы будем считать $g(x,t)$ непрерывной по t при всех x (этого можно добиться, выбросив из $\mathcal{X}$ множество нулевой $\mathbf{P}$-вероятности). Обозначим через

$$\omega_\Delta(x) = \sup_{t\in\Theta,\, t+u\in\Theta,\, |u|\leqslant\Delta} |g(x,t+u) - g(x,t)|$$

модуль непрерывности функции $s(t)$ на Θ. Тогда

$$\omega_\Delta \equiv \mathbf{E}\omega_\Delta(\mathrm{x}_1) = \int \omega_\Delta(x)\mathbf{P}(dx) \to 0$$

при $\Delta \to \infty$. Это следует из теоремы о мажорируемой сходимости, так как в силу (M_1), (M_2) имеем $\omega_\Delta(x) \to 0$ при $\Delta \to 0$, $\omega_\Delta(x) < 2\gamma(x)$. Из этих же соображений вытекает непрерывность $s(t)$.

По усиленному закону больших чисел

$$\omega^*_\Delta(X) \equiv \sup_{t\in\Theta,\, t+u\in\Theta,\, |u|\leqslant\Delta} \left|s^*_n(t+u) - s^*_n(t)\right| \leqslant \frac{1}{n}\sum_{i=1}^{n} \omega_\Delta(\mathrm{x}_i) \underset{\text{п.н.}}{\longrightarrow} \omega_\Delta.$$

Далее, пусть $\tau_k = t_0 + k\Delta$, $k = 0,1,\ldots,m$ $(\tau_0 = t_0,\ \tau_m = t_1)$, — точки разбиения отрезка $[t_0,t_1] \supset \Theta$. Тогда $s^*_n(\tau_i) \underset{\text{п.н.}}{\longrightarrow} s(\tau_i)$ при всех i. Остается заметить, что

$$s_n \equiv \sup_{t\in\Theta} \left|s^*_n(t) - s(t)\right| \leqslant \max_i \left|s^*_n(\tau_i) - s(\tau_i)\right| + \omega^*_\Delta(X) + \omega_\Delta \underset{\text{п.н.}}{\longrightarrow} 2\omega_\Delta,$$

где левая часть от Δ не зависит, а правая выбором Δ может быть сделана сколь угодно малой. Это доказывает утверждение леммы.

Доказательство теоремы 3. Пусть δ таково, что $(\theta_0-\delta,\theta_0+\delta) \subset \Theta$ и для определенности $s(t) > 0$ при $t > \theta_0$ и $s(t) < 0$ при $t < \theta_0$. Ясно, что для любого $\delta > 0$ найдется $\delta_1 = \delta_1(\delta)$ такое, что

$$\inf_{t>\theta_0+\delta} s(t) > 2\delta_1, \qquad \sup_{t<\theta_0-\delta} s(t) < -2\delta_1. \tag{14}$$

В силу леммы 1 для п. в. $X \in \mathbf{P}$ найдется $n(X)$ такое, что при всех $n \geqslant n(X)$ будет выполняться $\inf\limits_{t>\theta_0+\delta} s^*_n(t) > \delta_1$, $\sup\limits_{t<\theta_0-\delta} s^*_n(t) < -\delta_1$. Так как функция $s^*_n(t)$ непрерывна, то при $n \geqslant n(X)$ существует $\theta^* \in (\theta_0-\delta,\theta_0+\delta)$ такое, что $s^*_n(\theta^*) = 0$. Итак, мы получили, что для заданного $\delta > 0$ существует $n(X)$ такое, что при всех $n \geqslant n(X)$ выполняется $|\theta^* - \theta_0| < \delta$. $\triangleleft$

6. Асимптотическая нормальность M-оценок. Здесь в дополнение к (M_1), (M_2) нам понадобятся следующие условия.

(M'_1) *Функция $g(x,\theta)$ для п. в. x непрерывно дифференцируема по θ,*

$$\sup_{\iota\in\Theta} |g'(x,t)| < \gamma_1(x), \qquad \int \gamma_1(x)\mathbf{P}(dx) < \infty.$$

Из этого условия следует, что в (13) законно дифференцирование по θ:

$$s'(\theta) = \mathbf{E}g'(\mathrm{x}_1, \theta).$$

(M_2') $s'(\theta_0) \neq 0$.
(M_4) $\sigma^2 \equiv \mathbf{E}g^2(\mathrm{x}_1, \theta_0) < \infty$.

В этих условиях θ_0 — решение уравнения (9).

Теорема 4. *При выполнении условий* (M_1'), (M_2'), (M_3), (M_4) *M-оценка θ^* асимптотически нормальна с коэффициентом $\sigma^2[s'(\theta_0)]^{-2}$.*

Доказательство. Так как (M_1'), (M_2') влекут за собой (M_1), (M_2), то условия теоремы 3 выполнены и $\theta^* \underset{\text{п.н.}}{\longrightarrow} \theta_0$.

Так как $s_n^*(t)$ непрерывно дифференцируема, то (полагаем для краткости $s_n^*(t) = s_n(t)$)

$$s_n(\theta^*) = 0 = s_n(\theta_0) + (\theta^* - \theta_0)s_n'(\widetilde{\theta}), \quad \widetilde{\theta} \in [\theta_0, \theta^*], \tag{15}$$

где $\widetilde{\theta} \underset{\text{п.н.}}{\longrightarrow} \theta_0$.

Заметим теперь, что в силу (M_1'), (M_2') функция $s_n'(t)$ удовлетворяет условиям леммы 1 и, следовательно, $s'(t)$ непрерывна,

$$\sup_{t\in\Theta}\left|s_n'(t) - s'(t)\right| \underset{\text{п.н.}}{\longrightarrow} 0.$$

Это значит, что $s_n'(\widetilde{\theta}) - s'(\widetilde{\theta}) \underset{\text{п.н.}}{\longrightarrow} 0$, и так как $s'(\widetilde{\theta}) \underset{\text{п.н.}}{\longrightarrow} s'(\theta_0)$, то $s_n'(\widetilde{\theta}) \underset{\text{п.н.}}{\longrightarrow} \underset{\text{п.н.}}{\longrightarrow} s'(\theta_0)$.

Итак, в силу (15)

$$\sqrt{n}\,(\theta^* - \theta_0) = -\sqrt{n}\,s_n(\theta_0)\big(s_n'(\widetilde{\theta})\big)^{-1},$$

где $s_n'(\widetilde{\theta}) \underset{\text{п.н.}}{\longrightarrow} s'(\theta_0) \neq 0$ и по центральной предельной теореме $-\sqrt{n}\,s_n(\theta_0) \Subset \Subset \mathbf{\Phi}_{0,\sigma^2}$. Нам остается воспользоваться второй теоремой непрерывности. $\triangleleft$

Из сказанного в доказательствах теорем 1, 2 легко извлечь также, что для любой последовательности $\delta_n \to 0$ случайный процесс $y(u) = \sqrt{n}\,s_n \times$ $\times(\theta_0 + u/\sqrt{n})$ допускает в области $|u| < \delta_n\sqrt{n}$ представление

$$y(u) = \xi_n + us'(\theta_0)(1 + \varepsilon_n(\theta_0, X, u)),$$

где $\xi_n \Subset N_{0,\sigma^2}$, $|\varepsilon_n(\theta_0, X, u)| \leqslant \varepsilon_n(\theta_0, X) \underset{P}{\longrightarrow} 0$ при $n \to \infty$.

Замечание 1. Предельное распределение для M-оценок можно получать и пользуясь подходами § 8, где изучалось асимптотическое поведение статистик второго типа (см. замечание в конце § 8). Так как M-оценка параметра θ есть значение $G(F_n^*)$, где функционал $G(F)$ является решением

относительно θ уравнения $\int g(x,\theta)\,dF(x)=0$, то мы имеем тождество

$$\int g(x,G(F))\,dF(x)=0, \tag{16}$$

пользуясь которым можно найти значение производной $G'(F_0,v)$ (см. § 8). Пусть $v=F_1-F_0$, где F_1, как и F_0, есть функция распределения. Тогда F_0+hv — снова функция распределения, и, подставляя в (16) F_0+hv вместо F, получим

$$\int [g(x,G(F_0+hv))]\,d(F_0(x)+hv(x))=0.$$

Предположим, что функция $g(x,\theta)$ для всех x непрерывно дифференцируема по θ. Так как $G'(F_0,v)$ является обычной производной функции $G(F_0+hv)$ по h в точке $h=0$ (см. замечание 8.1), то, пользуясь процедурой отыскания производной от неявной функции, получим

$$\int g'(x,G(F_0))G'(F_0,v)\,dF_0(x)+\int g(x,G(F_0))\,dv(x)=0,$$

$$G'(F_0,v)=-\int g(x,G(F_0))\,dv(x)\left(\int g'(x,G(F_0))\,dF_0\right)^{-1}.$$

Стало быть, в силу (8.10), (16)

$$\sqrt{n}\left(G\left(F_n^*\right)-G(F_0)\right)\approx$$

$$\approx-\int g(x,G(F_0))\,dw^n(F_0(x))\left(\int g'(x,G(F_0))\,dF_0(x)\right)^{-1}\Rightarrow$$

$$\Rightarrow-(s'(\theta_0))^{-1}\int g(x,\theta_0)\,dw^0(F_0(x))=-(s'(\theta_0))^{-1}\int g(x,\theta_0)\,dw(F_0(x)).$$

Это значит, что $\sqrt{n}\left(G\left(F_n^*\right)-G(F_0)\right)$ асимптотически нормально с параметрами, указанными в теореме 2.

7. Замечания о многомерном случае. Если неизвестный параметр $\theta=(\theta_1,\ldots,\theta_k)\in\mathbb{R}^k$ многомерен ($k>1$), то исходным при построении M-оценок является предположение о том, что θ является решением системы уравнений $s(\theta)=0$, где $s(\theta)=(s_1(\theta),\ldots,s_k(\theta))$,

$$s_i(\theta)=\int g_i(x,\theta)\mathbf{P}(dx).$$

В соответствии с этим M-оценка определяется как решение системы уравнений $s_n^*(t)=0$, где $s_n^*(t)=\left(s_{n,1}^*(t),\ldots,s_{n,k}^*(t)\right)$,

$$s_{n,i}^*(t)=\frac{1}{n}\sum_{r=1}^{n}g_i(\mathrm{x}_r,t).$$

Утверждения, относящиеся к состоятельности и асимптотической нормальности M-оценок, сохранятся при естественной модификации условий. Основным для доказательства асимптотической нормальности является следующий аналог уравнения (15):

$$s^*_{n,i}(\theta^*) = 0 = s^*_{n,i}(\theta) + ((\theta^* - \theta, \operatorname{grad} s^*_{n,i}(\theta)) + o(|\theta^* - \theta|), \quad i = 1, \ldots, k. \tag{17}$$

Здесь матрица $\left\| \dfrac{\partial s^*_{n,i}(t)}{\partial t_j} \right\|_{t=\theta}$ сходится п. н. к матрице $\boldsymbol{s}'(\theta) = \|s^{(ij)}\|$, $s^{(ij)} = \dfrac{\partial s_j(\theta)}{\partial \theta_i}$, $i, j = 1, \ldots, k$. Поэтому уравнение (17) в векторной форме можно записать в виде

$$s^*_n(\theta) + (\theta^* - \theta)\boldsymbol{s}'(\theta) + o(|\theta^* - \theta|) = 0,$$

которое в предположении $|\boldsymbol{s}'(\theta)| \neq 0$ дает соотношение

$$(\theta^* - \theta) \approx -s^*_n(\theta)(\boldsymbol{s}'(\theta))^{-1},$$

означающее асимптотическую нормальность θ^*:

$$(\theta^* - \theta)\sqrt{n} \Subset \boldsymbol{\Phi}_{0,M},$$

$$M = [(\boldsymbol{s}'(\theta))^{-1}]^\top D(\boldsymbol{s}'(\theta))^{-1}, \quad D = \|g^{(ij)}\|, \quad g^{(ij)} = \mathbf{E} g_i(\mathrm{x}_1, t) g_j(\mathrm{x}_1, \theta).$$

§ 15.* Метод минимального расстояния

Метод, указанный в заголовке, как и методы § 14, является реализацией принципа подстановки и состоит в следующем. Рассмотрим какой-нибудь функционал от двух распределений $d(\mathbf{P}, \mathbf{Q})$, который обладает тем свойством, что как функция от $\mathbf{Q}$ он достигает своего минимума при $\mathbf{Q} = \mathbf{P}$ и $d(\mathbf{P}, \mathbf{Q}) > d(\mathbf{P}, \mathbf{P})$ при $\mathbf{Q} \neq \mathbf{P}$. Мы будем рассматривать величину $d(\mathbf{P}, \mathbf{Q})$ (или $d(\mathbf{P}, \mathbf{Q}) - d(\mathbf{P}, \mathbf{P})$ как «расстояние» между $\mathbf{Q}$ и $\mathbf{P}$, так что $\mathbf{P}$ можно определить как значение $\mathbf{Q}$, при котором $d(\mathbf{P}, \mathbf{Q})$ достигает своего наименьшего значения.

Пусть теперь $X \Subset \mathbf{P}$, $\mathbf{P}$ неизвестно и принадлежит семейству $\mathcal{P}$. Обозначим через $(\mathbf{Q})_{\mathcal{P}}$ распределение из $\mathcal{P}$, ближайшее к распределению $\mathbf{Q}$ в смысле расстояния d, и допустим, что оно существует:

$$d((\mathbf{Q})_{\mathcal{P}}, \mathbf{Q}) = \min_{\Pi \in \mathcal{P}} d(\mathbf{\Pi}, \mathbf{Q}),$$

так что $(\mathbf{Q})_{\mathcal{P}} = \mathbf{Q}$, если $\mathbf{Q} \in \mathcal{P}$.

Определение 1. *Оценкой распределения* $\mathbf{P}$ *по минимуму расстояния* d называется распределение $\mathbf{P}^* = (\mathbf{P}^*_n)_{\mathcal{P}} \in \mathcal{P}$, где $\mathbf{P}^*_n$, как и прежде, эмпирическое распределение.

Таким образом, при $\mathbf{\Pi} = \mathbf{P}^*! = (\mathbf{P}^*_n)_{\mathcal{P}}$ минимизируется $d(\mathbf{\Pi}, \mathbf{P}^*_n)$. Если $\mathcal{P}$ совпадает с множеством всех функций распределения, то, очевидно, $\mathbf{P}^* = \mathbf{P}^*_n$.

Пусть теперь $\mathcal{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}$ есть параметрическое семейство, удовлетворяющее следующему условию.

(A_0) *Пусть*

$$\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2} \quad \textit{при} \quad \theta_1 \neq \theta_2.$$

В этом случае отображение $\theta \to \mathbf{P}_\theta$ взаимно однозначно, так что распределение $\mathbf{P} \in \mathcal{P}$ позволяет единственным образом восстановить параметр θ, для которого $\mathbf{P} = \mathbf{P}_\theta$. Этот факт можно выразить и иначе: существует функционал G, определенный на $\mathcal{P}$, такой, что $\theta = G(\mathbf{P}_\theta)$.

Введем в рассмотрение функционал $G_1(\mathbf{Q}) = G((\mathbf{Q})_\mathcal{P})$. Это есть, очевидно, значение $\theta \in \Theta$, при котором $\mathbf{P}_\theta$ будет ближайшим к $\mathbf{Q}$ распределением в смысле расстояния d, так что

$$G_1(\mathbf{P}_\theta) = G(\mathbf{P}_\theta) = \theta. \tag{1}$$

Определение 2. Оценка $\theta^* = G_1(\mathbf{P}_n^*)$ называется *оценкой параметра θ по минимуму расстояния d.*

Другими словами, θ^* есть значение из Θ, для которого

$$d(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*) = \inf_{\theta\in\Theta} d(\mathbf{P}_\theta, \mathbf{P}_n^*).$$

Очевидно, что здесь мы снова имеем дело с принципом подстановки. Это вытекает из определений и (1). Разумеется, расстояние d и семейство $\mathcal{P} = \{\mathbf{P}_\theta\}$ должны обладать свойствами, обеспечивающими измеримость отображения $\mathcal{X}^n$ в $\mathbb{R}^k$, осуществляемого функционалом $G_1(\mathbf{P}_n^*)$ так, чтобы θ^* была случайной величиной.

Отметим теперь, что *в параметрическом случае при выполнении условия* (A_0) *сужение метода подстановки* (см. (13.1), (13.2)) *и метод минимального расстояния дают один и тот же класс оценок.*

Действительно, мы уже знаем, что оценки минимального расстояния θ^* являются оценками по методу подстановки, при этом $\theta^* \in \Theta$. Пусть теперь θ^* есть оценка по методу подстановки: $\theta^* = G(\mathbf{P}_n^*)$, где $G(\mathbf{P}_\theta) = \theta$, $\theta^* \in \Theta$. Определим расстояние $d(\mathbf{P}, \mathbf{Q}) = |G(\mathbf{P}) - G(\mathbf{Q})|$. Тогда, очевидно, при $\theta = \theta^*$ достигается

$$\inf_{\theta\in\Theta} d(\mathbf{P}_\theta, \mathbf{P}_n^*) = \inf_{\theta\in\Theta} \left|G(\mathbf{P}_\theta) - G(\mathbf{P}_n^*)\right| = \inf_{\theta\in\Theta} \left|\theta - G(\mathbf{P}_n^*)\right| = 0.$$

Ясно, что можно указать много «разумных» расстояний d, которые можно использовать для построения оценок. Мы могли бы в качестве d взять расстояние

$$d(\mathbf{P}, \mathbf{Q}) = \sup_x |F_P(x) - F_Q(x)|$$

или

$$d(\mathbf{P}, \mathbf{Q}) = \int (F_P(x) - F_Q(x))^2 \, dF_Q(x),$$

где $F_P(x)$ — функция распределения, соответствующая распределению $\mathbf{P}$. Оценками θ^* по минимальному расстоянию здесь будут значения θ, для

которых достигается соответственно

$$\inf_\theta \sup_x \left|F_{P_\theta}(x) - F_n^*(x)\right|,$$

$$\inf_\theta \int \left(F_{P_\theta}(x) - F_n^*(x)\right)^2 dF_n^*(x) = \inf_\theta \frac{1}{n}\sum_{k=1}^{n}\left(F_{P_\theta}(\mathrm{x}_{(k)}) - \frac{k-1}{n}\right)^2. \tag{2}$$

В некоторых задачах (ср. с [58]) используются так называемые оценки по минимуму χ^2 (хи-квадрат). Это есть оценки по минимуму расстояния

$$d(\mathbf{P},\mathbf{Q}) = \sum_{i=1}^{r}\frac{(\mathbf{P}(\Delta_i) - \mathbf{Q}(\Delta_i))^2}{\mathbf{P}(\Delta_i)},$$

где $\Delta_i,\ldots,\Delta_r$ — разбиение $\mathbb{R}$ (или $\mathbb{R}^m$, если x_j m-мерны) на $r<\infty$ интервалов, так что $\bigcup_{i=1}^{r}\Delta_i = \mathbb{R}$. Таким образом, оценка θ^* по минимуму χ^2 есть значение θ, при котором минимизируется

$$n\sum_{i=1}^{r}\frac{(\mathbf{P}_\theta(\Delta_i) - \nu_i/n)^2}{\mathbf{P}_\theta(\Delta_i)} = \sum_{i=1}^{r}\frac{(n\mathbf{P}_\theta(\Delta_i) - \nu_i)^2}{n\mathbf{P}_\theta(\Delta_i)}. \tag{3}$$

Здесь $\nu_i = n\mathbf{P}_n^*(\Delta_i)$ есть число наблюдений x_j, попавших в интервал Δ_i. Статистика в правой части (3) есть уже известная нам статистика χ^2, отсюда и происходит название оценки.

Ниже мы увидим, что существует такой функционал G, $\theta = G(\mathbf{P}_\theta)$, для которого оценки по принципу подстановки, называемые оценками максимального правдоподобия, будут в известном смысле наилучшими. Эти оценки будут являться также $\widehat{M}$-оценками, M-оценками (при некоторых условиях регулярности) и оценками минимального расстояния. Расстояние, минимизируемое этими оценками, будет близко к расстоянию Кульбака–Лейблера между распределениями, которое будет введено в § 31.

§ 16. Метод максимального правдоподобия. Оптимальность оценок максимального правдоподобия в классе M-оценок

1. Определения. Общие свойства. Пусть снова $\mathcal{P}$ есть параметрическое семейство $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$. Относительно этого семейства будем предполагать в дальнейшем везде, где это потребуется, выполненным условие

(A_0) *Пусть*

$$\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2} \quad \textit{при} \quad \theta_1 \neq \theta_2,$$

а также условие

(A_μ) *Существует σ-конечная мера μ в фазовом пространстве $(\mathcal{X},\mathfrak{B}_\mathcal{X})$ такая, что все распределения $\mathbf{P}_\theta \in \mathcal{P}$ имеют относительно*

этой меры плотность $f_\theta(x) = \dfrac{d\mathbf{P}_\theta}{d\mu}(x)$, *так что*

$$\mathbf{P}_\theta(B) = \int\limits_B f_\theta(x)\mu(dx).$$

В этом случае говорят, что мера μ *доминирует* распределения $\mathbf{P}_\theta$.

Все семейства распределений, рассмотренные в § 12, очевидно, удовлетворяют условиям (A_0), (A_μ). В качестве μ для одних распределений надо взять меру Лебега (абсолютно непрерывные распределения), для других — считающую меру (дискретные распределения). Считающая мера μ определяется так: $\mu(B) = k$, где k есть число точек с целочисленными координатами, принадлежащих B.

К первым относятся нормальное $\mathbf{\Phi}_{\alpha,\sigma^2}$, логнормальное $\mathbf{L}_{\alpha,\sigma^2}$ распределения, Г- и В-распределения, равномерное распределение, распределение Коши, распределения Стьюдента и Фишера. Ко вторым — распределения Бернулли, Пуассона, вырожденное в нуле и полиномиальные распределения. Вид плотностей $f_\theta(x)$ этих распределений приведен в § 12. В дискретном случае (когда μ есть считающая мера) плотность $f_\theta(x)$ совпадает с вероятностью $\mathbf{P}_\theta(\{x\})$ события $\{\mathrm{x}_1 = x\}$; здесь $\{x\}$ означает множество, состоящее из одной точки x. Можно отметить также, что, например, нормальное распределение $\mathbf{\Phi}_{\alpha,\sigma^2}$ и распределение Пуассона взаимно сингулярны. Вместо меры Лебега и считающей меры мы могли бы брать и другие меры — например, нормальное распределение $\mathbf{\Phi}_{0,1}$ и распределения Пуассона $\mathbf{\Pi}_1$ соответственно. Однако в этом случаи плотности $f_\theta(x)$ будут, очевидно, другими. Мы предлагаем читателю найти их. Приведенные выше примеры относились к случаю $\mathcal{X} = \mathbb{R}$ или $\mathcal{X} = \mathbb{R}^m$, $m > 1$. В произвольном фазовом пространстве $(\mathcal{X}, \mathfrak{B}_\mathcal{X})$ природа меры μ может быть более сложной.

Введение условия (A_μ) удобно прежде всего тем, что оно позволит нам в дальнейшем с единой точки зрения рассматривать два наиболее важных в приложениях типа распределений — абсолютно непрерывные и дискретные. С точки зрения условия (A_μ) качественная разница между ними отсутствует. Кроме того, становится несущественной размерность фазового пространства $\mathcal{X}$.

Условимся писать

$$f(x) = g(x) \quad \text{п. в.} \quad [\mu],$$

если существует множество A, $\mu(A) = 0$ такое, что $f(x) = g(x)$ для всех $x \notin A$. Очевидно, $f(x) = g(x)$ п. в. $[\mu]$ тогда и только тогда, когда

$$\int (f(x) - g(x))^2 \mu(dx) = 0.$$

Лемма 1. *Пусть f и g — две плотности вероятности относительно меры μ. Тогда*

$$\int f(x) \ln f(x) \mu(dx) \geqslant \int f(x) \ln g(x) \mu(dx), \tag{1}$$

если оба эти интеграла конечны. Знак равенства возможен лишь в случае $f = g$ п. в. $[\mu]$.

Здесь принято соглашение, что интегралы в (1) по множеству A, на котором $f(x) = 0$, равны нулю при любой $g(x)$.

Доказательство. Нам надо доказать, что

$$\int f(x) \ln \frac{g(x)}{f(x)} \mu(dx) \leqslant 0.$$

Так как $\ln(1+x) \leqslant x$ при всех $x \geqslant -1$ и знак равенства возможен лишь при $x = 0$, то

$$\ln \frac{g(x)}{f(x)} = \ln \left(1 + \left(\frac{g(x)}{f(x)} - 1\right)\right) \leqslant \frac{g(x)}{f(x)} - 1,$$

и знак равенства здесь возможен лишь при $f(x) = g(x)$. Поэтому

$$\int f(x) \ln \frac{g(x)}{f(x)} \mu(dx) \leqslant \int f(x) \left(\frac{g(x)}{f(x)} - 1\right) \mu(dx) = \\ = \int g(x)\mu(dx) - \int f(x)\mu(dx) = 0. \quad (2)$$

Если соотношение $f = g$ п. в. $[\mu]$ не имеет места, то, очевидно, знак неравенства в (2) будет строгим. $\triangleleft$

Рассмотрим теперь семейство $\mathcal{P} = \{\mathbf{P}_\theta\}_{\theta \in \Theta}$, удовлетворяющее условиям (A_0), (A_μ), и «расстояние» $d(\mathbf{P}_\theta, \mathbf{Q})$ между произвольным распределением $\mathbf{Q}$ и распределением $\mathbf{P}_\theta \in \mathcal{P}$

$$d(\mathbf{P}_\theta, \mathbf{Q}) = -\int \ln f_\theta(x) \mathbf{Q}(dx). \quad (3)$$

Определим функционал $G(\mathbf{Q})$ как значение θ, при котором достигается

$$\min_\theta d(\mathbf{P}_\theta, \mathbf{Q}) = d(\mathbf{P}_{G(Q)}, \mathbf{Q}).$$

Из леммы 1 и условия (A_0) следует, что

$$-\int f_{\theta_0} \ln f_\theta \mu(dx) > -\int f_{\theta_0} \ln f_{\theta_0} \mu(dx),$$

$$d(\mathbf{P}_\theta, \mathbf{P}_{\theta_0}) > d(\mathbf{P}_{\theta_0}, \mathbf{P}_{\theta_0})$$

при $\theta \neq \theta_0$. Это означает, что

$$G(\mathbf{P}_{\theta_0}) = \theta_0. \quad (4)$$

Определение 1. *Оценкой максимального правдоподобия* (о. м. п.) называется значение $\widehat{\theta}^* = G(\mathbf{P}_n^*)$, т. е. значение θ, при котором достигается

$$\max_\theta \int \ln f_\theta(x) \mathbf{P}_n^*(dx) = \max_\theta \frac{1}{n} \sum_{i=1}^n \ln f_\theta(\mathrm{x}_i). \quad (5)$$

Символ «$\widehat{\ }$» над обозначением оценки везде в дальнейшем будет соответствовать о. м. п.

Из определения и (4) следует, что *о. м. п. является оценкой подстановки.* Ее можно рассматривать также как $\widehat{M}$-оценку и оценку по минимуму расстояния (3). Это расстояние, как уже отмечалось, тесно связано с расстоянием Кульбака–Лейблера между распределениями, которое играет в математической статистике особую роль и будет нами рассмотрено позже.

В определении 1 семейство $\{\mathbf{P}_\theta\}$ предполагается таким, чтобы $\widehat{\theta}^*$ была случайной величиной*).

Поскольку максимум некоторой функции может достигаться в нескольких точках, то о. м. п., вообще говоря, не единственна. Соответствующий пример мы приведем несколько позже.

Название этой оценки связано со следующей важной интерпретацией выражения

$$\sum_{i=1}^{n} \ln f_\theta(\mathrm{x}_i) = \ln \prod_{i=1}^{n} f_\theta(\mathrm{x}_i),$$

присутствующего в (5). Для простоты рассмотрим сначала дискретный случай, когда μ — считающая мера. Тогда $\prod_{i=1}^{n} f_\theta(\mathrm{x}_i)$ есть вероятность появления исхода $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Стало быть, мы выбираем в качестве $\widehat{\theta}^*$ значение параметра, которое *максимизирует эту вероятность* (ведь функции $\varphi(\theta) > 0$ и $\ln \varphi(\theta)$ достигают экстремумов в одних и тех же точках).

Аналогичное толкование имеет место и в общем случае. В силу независимости x_i имеем для множеств $B = B_1 \times \cdots \times B_n$, $B_i \in \mathfrak{B}_\mathcal{X}$

$$\mathbf{P}_\theta(X \in B) = \int\limits_{B_1} f_\theta(x_1)\mu(dx_1) \cdots \int\limits_{B_n} f_\theta(x_n)\mu(dx_n). \tag{6}$$

Напомним, что x_i в отличие от элементов выборки x_i обозначают переменные величины, вектор $(x_1, \ldots, x_n)$ обозначается через $\boldsymbol{x}$. Пусть μ^n есть n-кратное прямое произведение мер μ, так что $\mu^n(d\boldsymbol{x}) = \prod_{i=1}^{n} \mu(dx_i)$. Тогда (6) означает, что

$$\mathbf{P}_\theta(X \in B) = \int\limits_B \left(\prod_{i=1}^{n} f_\theta(x_i) \right) \mu^n(d\boldsymbol{x})$$

и, стало быть, функция $f_\theta(\boldsymbol{x}) = \prod_{i=1}^{n} f_\theta(x_i)$ есть плотность распределения случайного вектора X в $\mathcal{X}^n$ относительно меры μ^n,

$$\int\limits_{\mathcal{X}^n} f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = 1.$$

*) То есть $\widehat{\theta}^*$ осуществляет измеримое отображение $(\mathcal{X}^n, \mathfrak{B}^n_\mathcal{X})$ в $(\mathbb{R}^k, \mathfrak{B}^k)$.

Таким образом, $\prod_{i=1}^{n} f_\theta(x_i)\mu^n(dx)$ можно интерпретировать (аналогично дискретному случаю) как вероятность попадания выборки в параллелепипед, образованный пересечением «полос» $(x_i, x_i + dx_i)$, и оценка максимального правдоподобия максимизирует по θ эту вероятность.

Функция

$$f_\theta(X) = \prod_{i=1}^{n} f_\theta(\mathrm{x}_i)$$

как функция от θ называется *функцией правдоподобия*, а функция

$$L(X, \theta) = \ln f_\theta(X) = \sum_{i=1}^{n} l(\mathrm{x}_i, \theta),$$

где $l(x, \theta) = \ln f_\theta(x)$, — *логарифмической функцией правдоподобия*.

Такое же название за функциями f и L мы будем сохранять и в том случае, когда в качестве аргумента вместо X стоит переменный вектор $\boldsymbol{x}$. Таким образом, функция правдоподобия $f_\theta(\boldsymbol{x})$ есть функция на $\mathcal{X}^n \times \Theta$, являющаяся при каждом $\theta \in \Theta$ плотностью вероятности относительно меры μ^n, так что плотность $f_\theta(x_1)$ в $\mathcal{X}$ также есть функция правдоподобия для случая $n = 1$.

С другой стороны, $f_\theta(X)$, например, в случае $\mathcal{X} = \mathbb{R}$ можно рассматривать как функцию правдоподобия выборки объема 1 в многомерном случае, когда $\mathcal{X} = \mathbb{R}^m = \mathbb{R}^n$.

Важно отметить, что о. м. п. от выбора меры μ никак не зависит, так как при замене μ на какую-нибудь эквивалентную меру μ_1 функция правдоподобия $f_\theta(x)$ изменится лишь на множитель $\dfrac{d\mu^n}{d\mu_1^n}(x)$, от θ не зависящий.

В соответствии с определением 14.2А о. м. п. является $\widehat{M}$-оценкой для функции $\psi(x, \theta) = l(x, \theta)$. Это соответствует тому, что значение $t = \theta$ максимизирует

$$\int l(x, t)\mathbf{P}_\theta(dx) = \int \ln f_t(x) f_\theta(x)\mu(dx). \tag{7}$$

Если $l(x, \theta)$ имеет непрерывные производные $l'(x, \theta)$ по θ, то точка $\widehat{\theta}^*$, в которой достигается максимум (5), будет с необходимостью удовлетворять уравнению

$$L'(X, \theta) = \sum_{i=1}^{n} l'(\mathrm{x}_i, \theta) = 0. \tag{8}$$

В соответствии с определением 14.2 и замечаниями к нему это означает, что в этом случае о. м. п. $\widehat{\theta}^*$ является M-оценкой (надо положить в определении 14.2 $g(x, \theta) = l'(x, t)$). Это соответствует тому, что значение $t = \theta$ в предположении законности дифференцирования интеграла (7) удовлетворяет уравнению

$$\int l'(x, t)\mathbf{P}_\theta(dx) = 0. \tag{9}$$

Тот факт, что о. м. п. является M-оценкой, можно получить и из других соображений, не связанных с максимизацией (как и тот факт, что $t = \theta$ является решением уравнения (9)). Для этого в рамках параметрического случая надо предполагать, что законно дифференцирование по θ равенства

$$\int f_\theta(x)\mu(dx) = 1,$$

что дает нам равенство вида (14.8) при $g = l'$

$$\int l'(x,\theta)\mathbf{P}_\theta(dx) = \int l'(x,\theta)f_\theta(x)\mu(dx) = \int f'_\theta(x)\mu(dx) = 0. \qquad (10)$$

В этом случае согласно определению 14.2 решение уравнения (8) (или уравнения (14.10)) является M-оценкой параметра θ.

Обратное утверждение о том, что M-оценка, соответствующая функции $g(x,\theta) = l'(x,\theta)$, является о. м. п., вообще говоря, не верно. Функция $L(X,\theta)$ может иметь несколько изолированных точек локальных максимумов, в которых будет удовлетворяться (8), однако $\max L(X,\theta)$ достигаться не будет. Это будет иметь место в тех случаях, когда $J(t) \equiv \int l(x,t)\mathbf{P}_\theta(dx)$ имеет несколько точек максимума по t и, следовательно, решение уравнения (9) не единственно. Соответствующие примеры легко построить, если взять в качестве θ параметр сдвига: $f_\theta(x) = f(x-\theta)$ ($\mathcal{X} = \mathbb{R}$, $\Theta = \mathbb{R}$), а в качестве $f(x)$ симметричную «двухгорбную» функцию (например, $f(x) = \frac{1}{2}(\varphi(x+1) + \varphi(x-1))$, где φ — плотность нормального распределения с параметрами $(0,1)$). Тогда интеграл $J(\theta)$ будет иметь три точки $\theta = 0$, $\theta = \pm\theta_1$, $\theta_1 > 0$, в которых $J(\theta)$ достигает локальных максимумов, $J'(0) = J'(\pm\theta_1) = 0$, при этом $J(0) > J(\pm\theta_1)$. Примерно такая же картина при больших n будет иметь место и для функции $L(X,\theta)$.

Найдем функции правдоподобия и о. м. п. для некоторых распределений из § 12. Для гладких функций правдоподобия максимум проще всего искать путем приравнивания нулю первых производных.

Пример 1. *Нормальное распределение* $\mathbf{\Phi}_{\alpha,\sigma^2}$ и $\mathcal{X} = \mathbb{R}$ имеет плотность

$$\varphi_{\alpha,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\alpha)^2}{2\sigma^2}\right], \quad -\infty < \alpha < \infty, \quad \sigma > 0.$$

Полагая в этом случае $\theta = (\alpha,\sigma^2)$, получим

$$f_\theta(x) = (2\pi)^{-n/2}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathrm{x}_i-\alpha)^2\right\},$$

$$L(X,\theta) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathrm{x}_i-\alpha)^2.$$

Так как ln есть монотонная функция, то, как уже отмечалось, f и L достигают максимума при одних и тех же θ. Имеем

$$\frac{\partial L}{\partial \alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\mathrm{x}_i - \alpha), \quad \frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(\mathrm{x}_i - \alpha)^2.$$

Решая для точки максимума систему уравнений

$$\frac{\partial L}{\partial \alpha} = 0, \quad \frac{\partial L}{\partial \sigma} = 0,$$

получаем

$$\widehat{\alpha}^* = \overline{\mathrm{x}}, \quad (\widehat{\sigma}^2)^* = S^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2.$$

Нетрудно проверить, что в этой точке действительно достигается максимум L.

Пример 2. Рассмотрим Г-*распределение* с плотностью

$$\gamma_\alpha(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)}x^{\lambda-1}e^{-\alpha x}, \quad x \geqslant 0, \quad \alpha > 0,$$

в случае, когда параметр λ известен. Имеем

$$L(X,\alpha) = \lambda n \ln\alpha - n\ln\Gamma(\lambda) + (\lambda - 1)\sum_{i=1}^{n}\ln \mathrm{x}_i - \alpha\sum_{i=1}^{n}x_i,$$

$$\frac{\partial L}{\partial \alpha} = \frac{\lambda n}{\alpha} - \overline{\mathrm{x}}n, \quad \widehat{\alpha}^* = \frac{\lambda}{\overline{\mathrm{x}}}.$$

Пример 3. *Биномиальное распределение* $\mathbf{B}_p$. Здесь для $X \in \mathbf{B}_p$ имеем $\mathbf{P}(\mathrm{x}_i = 1) = p$, $\mathbf{P}(\mathrm{x}_i = 0) = 1 - p$,

$$f_p(X) = p^\nu(1-p)^{n-\nu},$$

где ν — число появлений 1 среди элементов $\mathrm{x}_1, \ldots, \mathrm{x}_n$. Стало быть,

$$L(X,p) = \nu\ln p + (n-\nu)\ln(1-p),$$

$$\frac{\partial L}{\partial p} = \frac{\nu}{p} - \frac{n-\nu}{1-p}, \quad \widehat{p}^* = \frac{\nu}{n}.$$

Мы предлагаем читателю в виде упражнения попытаться найти о. м. п. для всех параметрических семейств, приведенных в § 12, и сравнить их с оценками по методу моментов.

Приведем теперь два примера несколько иного типа, когда функция f_θ не является гладкой по θ и методы отыскания о. м. п., связанные с дифференцированием, не действуют.

Пример 4. Пусть $X \in \mathbf{U}_{\theta,1+\theta}$ (*равномерное распределение на* $[\theta, 1+\theta]$). Здесь

$$f_\theta(x) = \begin{cases} 1, & x \in [\theta, 1+\theta], \\ 0, & x \notin [\theta, 1+\theta], \end{cases}$$

$$f_\theta(X) = \begin{cases} 1, & \theta \leqslant \mathrm{x}_{(1)} \leqslant \mathrm{x}_{(n)} \leqslant 1+\theta, \\ 0 & \text{в противном случае,} \end{cases}$$

где $\mathrm{x}_{(1)} \leqslant \ldots \leqslant \mathrm{x}_{(n)}$ — вариационный ряд. Оценка м. п. в этом примере не единственна. Действительно, $f_\theta(X) = 1$ (т. е. максимальному значению) при любых θ, удовлетворяющих соотношениям $\mathrm{x}_{(n)} - 1 \leqslant \theta \leqslant \mathrm{x}_{(1)}$. Так как $\mathrm{x}_{(n)} - \mathrm{x}_{(1)} \leqslant 1$, то такие θ всегда существуют. Мы можем взять, в частности, $\widehat{\theta}_1^* = \mathrm{x}_{(1)}$ или $\widehat{\theta}_2^* = \mathrm{x}_{(n)} - 1$. Так как $\widehat{\theta}_1^* > \theta$, $\widehat{\theta}_2^* < \theta$, то лучше, по-видимому, взять

$$\widehat{\theta}^* = \frac{1}{2}(\widehat{\theta}_1^* + \widehat{\theta}_2^*) = \frac{1}{2}(\mathrm{x}_{(1)} + \mathrm{x}_{(n)} - 1).$$

Пример 5. Пусть $X \in \mathbf{U}_{0,\theta}$. Здесь

$$f_\theta(x) = \begin{cases} \theta^{-1}, & x \in [0, \theta], \\ 0, & x \notin [0, \theta], \end{cases}$$

$$f_\theta(X) = \begin{cases} \theta^{-n}, & \text{если } \mathrm{x}_i \in [0,\theta] \text{ при всех } i = 1, 2, \ldots, n, \\ 0 & \text{в противном случае.} \end{cases}$$

Чтобы получить более простую запись функции $f_\theta(X)$ как функции от θ, перепишем условие $\mathrm{x}_i \in [0, \theta]$, $i = 1, \ldots, n$, в эквивалентной форме $\theta \geqslant \max \mathrm{x}_i = \mathrm{x}_{(n)}$. Таким образом, $f_\theta(X) = 0$ при $\theta \in [0, \mathrm{x}_{(n)})$ и $f_\theta(X) = \theta^{-n}$ при $\theta \in [\mathrm{x}_{(n)}, \infty)$. График этой функции представлен на рис. 1. Здесь, как и в предыдущем примере, функция f_θ разрывна. Максимум f_θ достигается в точке $\widehat{\theta}^* = \mathrm{x}_{(n)}$.

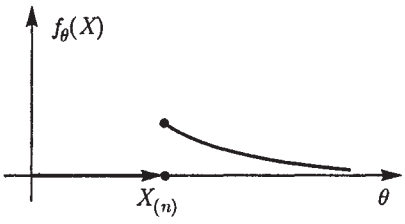

Рис. 1

Аналогичным образом читатель может найти о. м. п. для двумерного неизвестного параметра (α, β), когда $X \in \mathbf{U}_{\alpha,\beta}$.

Если $f_\theta(x)$ не ограничена и точки x_θ, в которых $f_\theta(x_\theta) = \infty$, зависят от θ, то метод максимального правдоподобия в значительной мере теряет свой смысл (мы приняли здесь соглашение $f_\theta(x_\theta) = \infty$, если $f_\theta(x) \to \infty$ при $x \downarrow x_\theta$ или $x \uparrow x_\theta$). Проще всего это понять на примере параметра сдвига, когда $f_\theta(x) = f(x - \theta)$, $f(x) > 0$, $f(0) = \infty$. Тогда $f_\theta(X) = \infty$ при $\theta = \mathrm{x}_1, \ldots$, $\theta = \mathrm{x}_n$ и, стало быть, $\widehat{\theta}^*$ принимает по крайней мере n значений, совпадающих с элементами выборки. Существо этого эффекта в том, что в этом случае «всплески» $f_\theta(X)$ не дают возможности судить о положении точки «истинного» максимума $f_\theta(X)$, обусловленного влиянием всей выборки (ср. с § 34). Чтобы получить таковой, следовало бы каким-то образом «сгладить» выбросы $f_\theta(X)$.

Оценки м. п. обладают следующим важным свойством инвариантности относительно замены параметра.

Теорема 1. *Пусть $\beta(\theta)$ есть функция, осуществляющая взаимно однозначное отображение множества Θ на множество B. Тогда если $\widehat{\theta}^*$ есть о. м. п. по выборке X параметра θ, то $\widehat{\beta}^* = \beta(\widehat{\theta}^*)$ будет о. м. п. по выборке X параметра $\beta = \beta(\theta)$ для параметрического семейства $\{\mathbf{Q}_\beta = \mathbf{P}_{\theta(\beta)}\}_{\beta\in B}$, где $\theta(\beta)$ есть функция, обратная к $\beta(\theta)$.*

Доказательство теоремы мы опускаем ввиду его очевидности.

Отметим, что неявно мы уже пользовались теоремой 1 в примере 1, где в поисках о. м. п. для σ^2 искали максимум L по σ и затем взяли $(\widehat{\sigma}^2)^* = (\widehat{\sigma}^*)^2$.

Другой пример использования теоремы 1 — отыскание о. м. п. в случае $X \in \mathbf{L}_{\alpha,\sigma^2}$, т. е. в случае, когда распределение x_i логнормально: $\ln \mathrm{x}_i \in \in \mathbf{\Phi}_{\alpha,\sigma^2}$. Для таких x_i среднее a и дисперсия d^2 равны соответственно (см. § 12)

$$a = \exp\left\{\alpha + \frac{\sigma^2}{2}\right\}, \quad d^2 = a^2(\exp(\sigma^2) - 1).$$

Если обозначить $\widehat{a}^*$ и $(\widehat{d^2})^*$ о. м. п. для a и d^2, то в силу свойства инвариантности получим для функции $(a, d^2) = \beta(\alpha, \sigma^2)$ (см. пример 1)

$$\widehat{a}^* = \exp\left\{\overline{\mathrm{y}} + \frac{S_Y^2}{2}\right\}, \quad (\widehat{d^2})^* = (\widehat{a}^*)^2(\exp(S_Y^2) - 1),$$

где $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_n)$, $\mathrm{y}_i = \ln \mathrm{x}_i$, $\overline{\mathrm{y}} = \dfrac{1}{n}\sum\limits_{i=1}^{n} \mathrm{y}_i$, $S_Y^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(\mathrm{y}_i - \overline{\mathrm{y}})^2$.

Приближенному вычислению о. м. п. в более сложных ситуациях посвящен § 35.

2. Асимптотические свойства о. м. п. Состоятельность. Как уже отмечалось, о. м. п. являются $\widehat{M}$-оценками и при выполнении некоторых условий регулярности — M-оценками. Если сравнивать условия состоятельности, которые можно получить для о. м. п., используя соответствующие результаты для $\widehat{M}$- и M-оценок, то окажется, что M-оценки являются более общими уже потому, что не требуют дифференцируемости $l(x, \theta)$ по θ. Поэтому для получения условий состоятельности о. м. п. мы воспользуемся результатами § 14 о состоятельности $\widehat{M}$-оценок.

Следуя п. 14.1, мы будем предполагать везде, что выполнено условие (A_c) о компактности Θ (замечания по этому поводу см. в п. 14.1). Условие $[\mathrm{A}_0]$ и его аналоги в § 14 мы будем использовать в новой редакции применительно к функциям $\psi(x, \theta) = l(x, \theta)$ и $\mathbf{P}(A) = \mathbf{P}_\theta(A)$. Эта новая редакция $[\mathrm{A}_0]$ следует из леммы 1 и приведена в начале параграфа в виде условия (A_0).

Она означает, что расстояние Кульбака–Лейблера (см. § 31)

$$\rho(\theta,t)=\int \ln\frac{f_\theta(x)}{f_t(x)}\cdot f_\theta(x)\mu(dx)>0$$

при $t\neq\theta$. Условия $(\overline{\mathrm{A}}_0)$, (A_0^Δ) будут иметь соответственно вид:

$(\overline{\mathrm{A}}_0)$ *Для любого* $\delta>0$ $\inf\limits_{t:|t-\theta|\geqslant\delta}\rho(\theta,t)>\varepsilon$ *при некотором* $\varepsilon=\varepsilon(\delta)>0$.

(A_0^Δ) *Для любого* $\delta>0$ *существует* $\Delta=\Delta(\delta)>0$ *такое, что при всех* t, $|t-\theta|>\delta$,

$$\int \ln\frac{f_\theta(x)}{f_t^\Delta(x)}\cdot f_\theta(x)\mu(dx)>\varepsilon \tag{11}$$

при некотором $\varepsilon>0$, *где*

$$f_t^\Delta(x)=\sup_{|u|\leqslant\Delta} f_{t+u}(x).$$

Из теоремы 14.1 вытекает

Следствие 1. *Если выполнено* (A_0^Δ), *то о. м. п.* $\widehat{\theta}^*$ *сильно состоятельна.*

Пусть $f_t^0(x)=\limsup\limits_{u\to t} f_u(x)$. Из теоремы 14.2 вытекает

Следствие 2. *Условие* (A_0^Δ) *эквивалентно одновременному выполнению следующих двух условий.*

(A_0^0) *Для всех* $t\neq\theta$

$$\int \ln\frac{f_t^0(x)}{f_\theta(x)}\cdot f_\theta(x)\mu(dx)<0.$$

(J) *При всех* t *и некотором* $\Delta>0$

$$\int \ln\frac{f_t^\Delta(x)}{f_\theta(x)}\cdot f_\theta(x)\mu(dx)<\infty.$$

Условие (J), означающее интегрируемость положительных частей подынтегральных функций, эквивалентно по существу ограниченности

$$\int \ln\frac{f^\Theta(x)}{f_\theta(x)}\cdot f_\theta(x)\mu(dx)<\infty,$$

где

$$f^\Theta(x)=\sup_{t\in\Theta} f_t(x).$$

Приведем теперь более простые достаточные условия, обеспечивающие выполнение (A_0^0) и (J) и, стало быть, сильную состоятельность о. м. п.

Определение 2. Будем говорить, что $f_t(x)$ принадлежит классу D_0, если для каждого $t\in\Theta$ существует множество $C_t\in\mathfrak{B}_\mathcal{X}$, $\mathbf{P}_\theta(C_t)=1$, на котором $f_t(x)$ непрерывна по t $f_{t_k}(x)\to f_t(x)$ при $t_k\to t$, $x\in C_t$.

Классу D_0 помимо непрерывных (по t) $f_t(x)$ на не зависящем от t множестве C, $\mathbf{P}_\theta(C) = 1$, принадлежат, конечно, и многие другие функции такие, например, для которых $f_t(x)$ в плоскости (t, x) имеет изолированные линии разрыва, не имеющие участков, параллельных оси x. Так будет, в частности, если $f_t(x)$ как функция от x имеет изолированные разрывы в точках $x_t^{(1)}, x_t^{(2)}, \ldots$, которые непрерывно зависят от t.

Теорема 2. *Если* $f_t(x) \in D_0$ *и выполнено* (J), *то выполнено условие* (A_0^0), *и, следовательно о. м. п.* $\widehat{\theta}^*$ *сильно состоятельна.*

Доказательство. Если $f_t(x) \in D_0$, то $f_t^0(x) = f_t(x)$ при $x \in C_t$ и, стало быть,

$$\int \ln \frac{f_t^0(x)}{f_\theta(x)} \cdot f_\theta(x)\mu(dx) = -\rho(\theta, t) < 0. \quad \triangleleft$$

Следствие 3. *Если* $f_t(x) \in D_0$ *ограничена, а интеграл*

$$\int f_\theta(x) \ln f_\theta(x)\mu(dx) \tag{12}$$

конечен, то о. м. п. сильно состоятельна.

Утверждение следствия 3 немедленно вытекает из теоремы 2, так как ограниченность $f_t(x)$ и конечность интеграла (12) влекут за собой условие (J).

Следствие 4. *Если*

$$\varphi(\Delta) = \int \sup_{|u| \leqslant \Delta} |f_{t+u}(x) - f_t(x)|\mu(dx) \to 0 \tag{13}$$

при $\Delta \to 0$, *то о. м. п. сильно состоятельна.*

Доказательство. Воспользуемся теоремой 2. Принадлежность $f_t(x) \in D_0$ очевидна, так как (13) может выполняться лишь в случае, когда $f_{t+u}(x) \to f_t(x)$ при $u \to 0$ для п. в. $[\mu]$ значений x. Далее,

$$\int f_t^\Delta(x)\mu(dx) \leqslant \varphi(\Delta) + \int f_t(x)\mu(dx) = \varphi(\Delta) + 1,$$

и условие (13) означает также интегрируемость $f_t^\Delta(x)$. Так как $\ln \dfrac{f_t^\Delta(x)}{f_\theta(x)} \leqslant$ $\leqslant \dfrac{f_t^\Delta(x)}{f_\theta(x)} - 1$, то отсюда получаем, что интеграл в условии (J) не превосходит

$$\int f_t^\Delta(x)\mu(dx) - 1 \leqslant \varphi(\Delta). \quad \triangleleft$$

Вместо (13) мы могли бы требовать сходимости к нулю величины

$$\varphi_1(\Delta) = \int \sup_{|u| \leqslant \Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)})^2 \mu(dx),$$

так как $\varphi(\Delta)$ можно оценить с помощью $\varphi_1(\Delta)$ неравенством

$$\varphi(\Delta) \leqslant \int \sup_{|u|\leqslant\Delta} |\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)}| \sup_{|u|\leqslant\Delta} |\sqrt{f_{t+u}(x)} + \sqrt{f_t(x)}|\mu(dx) \leqslant$$

$$\leqslant \varphi_1^{1/2}(\Delta)\left[\int \sup_{|u|\leqslant\Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)} + 2\sqrt{f_t(x)})^2\mu(dx)\right]^{1/2} \leqslant$$

$$\leqslant [2\varphi_1(\Delta)(\varphi_1(\Delta)+4)]^{1/2}.$$

Следствие 5. *Если $f_t(x)$ дифференцируема по t для п. в. $[\mu]$ значений x и*

$$\int |f_t'(x)|\mu(dx) < c < \infty, \tag{14}$$

то о. м. п. $\widehat{\theta}^*$ сильно состоятельна.

Доказательство. Принадлежность $f_t(x) \in D_0$ очевидна. Для выполнения условия (J) достаточно, как мы видели в доказательстве следствия 4, интегрируемости $f_t^{\Delta}(x)$. Но

$$\int f_t^{\Delta}(x)\mu(dx) \leqslant \int \left[f_t(x) + \int_{-\Delta}^{\Delta} |f_{t+u}'(x)|\, du\right]\mu(dx) =$$

$$= 1 + \int_{-\Delta}^{\Delta}\left[\int |f_{t+u}'(x)|\mu(dx)\right] du \leqslant 1 + 2\Delta c.$$

Остается воспользоваться теоремой 2.

Следствие 6. *Пусть θ есть параметр сдвига семейства $f_\theta(x) = f(x-\theta)$, $\int f(x)\ln f(x)\,dx > -\infty$. Если функция $f(x)$ ограничена (иначе метод максимального правдоподобия теряет свой смысл; см. замечание перед теоремой 1) и имеет множество B точек разрыва, лебегова мера замыкания которого $\mu(B^c)$ равна нулю, то о. м. п. $\widehat{\theta}^*$ сильно состоятельна.*

Доказательство. Проверим выполнение условий теоремы 2. Условие (J) выполнено очевидным образом. Принадлежность $f_t(x) \in D_0$ вытекает из определения D_0, в котором надо положить $C_t = \overline{B}^c - t$ (это есть сдвиг множества $\overline{B}^c$ на t, $\overline{B}^c$ есть дополнение к замыканию множества B). Так как множество $\overline{B}^c$ открыто, то $x - t \in \overline{B}^c - t$ влечет за собой $x - t_k \in \overline{B}^c - t$ при достаточно малых $|t_k - t|$. Это означает, что $f(x-t_k) \to f(x-t)$. Следствие доказано. $\triangleleft$

Отметим, что в условиях следствия 6 предполагать выполненным условие (A_0) излишне — оно выполняется автоматически. Если допустить, что

(A_0) не имеет места, то мы придем к периодичности функции $f(x)$, что невозможно.

Относительно условий следствия 6 заметим, что условие на «непрерывность» $f(x)$, сформулированное в нем, является весьма слабым. Но и оно, по-видимому, не существенно. На это указывает в какой-то мере следующий пример.

Пример 6. Пусть $f(x)$ — произвольная функция, имеющая ограниченный носитель $(a, b) = \{x : f(x) > 0\}$. Тогда

$$\mathbf{P}_\theta(|\widehat{\theta}^* - \theta| > \delta) \leqslant (1 - F_0(a+\delta))^n + F_0^n(b-\delta), \tag{15}$$

где $F_0(x) = \int\limits_{-\infty}^{x} f(y)\,dy$. Неравенство (15) означает сильную состоятельность $\widehat{\theta}^*$. Оно следует из соотношений вида

$$\{\widehat{\theta}^* - \theta > \delta\} \subset \left\{\prod_{i=1}^n f_{\theta+\delta}(\mathrm{x}_i) > 0\right\} \subset \bigcap_{i=1}^n \{\mathrm{x}_i \geqslant a + \theta + \delta\},$$

$$\mathbf{P}_\theta(\widehat{\theta}^* - \theta > \delta) \leqslant [1 - F_\theta(a+\theta+\delta)]^n = [1 - F_0(a+\delta)]^n.$$

Условие на конечность интеграла $\int f(x) \ln f(x)\,dx$ в следствии 6 в известном смысле тоже не существенно: легко построить пример, когда этот интеграл обращается в $-\infty$, а условие (J) будет выполнено. Все сказанное в следствии 6 и после него полностью сохраняет свою силу для параметра масштаба.

Если бы для отыскания условий состоятельности о. м. п. мы стали применять теорему 14.3 о состоятельности M-оценок, то мы столкнулись бы среди других с требованием дифференцируемости $f_t(x)$ по t, которое само по себе есть намного более сильное условие (при выполнении (A_0)), чем, скажем, (11).

Замечание 1. В целом применительно к о. м. п. имеет место в известном смысле парадоксальная ситуация (мы будем убеждаться в этом на примерах): для доказательства состоятельности (и особенно асимптотической нормальности; см. ниже) мы предполагаем выполнение некоторых свойств гладкости $f_t(x)$. На самом же деле, чем менее регулярна $f_\theta(x)$, тем, как правило, лучше оказываются свойства о. м. п. Иллюстрации этого факта можно найти в § 30, 32.

3. Асимптотическая нормальность о. м. п. Оптимальность в классе M-оценок. Обратимся теперь к свойству асимптотической нормальности о. м. п. Она будет вытекать из следующего утверждения, являющегося следствием теоремы 14.4. Это утверждение устанавливает также нижнюю границу для разброса M-оценок и оптимальность о. м. п. в классе M.

Теорема 3. *Предположим, что функция $g(x, \theta)$ удовлетворяет условиям теоремы* 14.4 *и законно дифференцирование под знаком интеграла*

в равенстве

$$\int g(x,\theta) f_\theta(x)\mu(dx) = 0.$$

Пусть, кроме того, выполнены следующие условия:

a) $l(x,\theta)$ *дважды непрерывно дифференцируема по* θ;

$$\sup_{\theta\in\Theta} |l''(x,\theta)| < l(x), \quad \int l(x) f_\theta(x)\mu(dx) < \infty; \tag{16}$$

b) *пусть*

$$0 < I(\theta) \equiv \mathbf{E}_\theta (l'(\mathrm{x}_1,\theta))^2 = \int (l'(x,\theta))^2 f_\theta(x)\mu(dx) < \infty; \tag{17}$$

c) *законно дифференцирование под знаком интеграла в равенствах*

$$\int f_\theta(x)\mu(dx) = 1, \quad \int l'(x,\theta) f_\theta(x)\mu(dx) = 0;$$

d) *уравнение относительно* t

$$\int l'(x,t) f_\theta(x)\mu(dx) = 0$$

имеет единственное решение $t = \theta$.

(В c) второе равенство получается дифференцированием первого; из c) вытекает, что $\mathbf{E}_\theta l''(\mathrm{x}_1,\theta) = -I(\theta)$.)

При выполнении этих условий M*-оценка* θ^**, соответствующая функции* g*, и о. м. п.* $\widehat{\theta}^*$ *асимптотически нормальны. Коэффициент разброса* θ^* (см. теорему 14.4) *удовлетворяет неравенству*

$$\mathbf{E}_\theta g^2(\mathrm{x}_1,\theta)(\mathbf{E}_\theta g'(\mathrm{x}_1,\theta))^{-2} \geqslant I^{-1}(\theta) \tag{18}$$

и достигает своего наименьшего значения, равного $I^{-1}(\theta)$ *при* $g(x,\theta) = l'(x,\theta)$ (*т. е. коэффициент разброса* θ^* *всегда больше или равен коэффициенту разброса* $\widehat{\theta}^*$).

Доказательство. Заметим прежде всего, что условия a)–d) достаточны для того, чтобы функция $l'(x,\theta)$ наряду с $g(x,\theta)$ удовлетворяла условиям теоремы 14.4. Поэтому нам надо доказать лишь последнюю часть теоремы, связанную с неравенством (18). Имеем (аргументы для краткости опускаем)

$$\int g'f + \int gf' = 0 = \int g'f + \int gl'f,$$

$$\mathbf{E}_\theta g'(\mathrm{x}_1,\theta) = \int g'f = -\int gl'f = -\mathbf{E}_\theta g(\mathrm{x}_1,\theta) l'(\mathrm{x}_1,\theta).$$

По неравенству Коши–Буняковского отсюда находим

$$(\mathbf{E}_\theta g'(\mathrm{x}_1,\theta))^2 = (\mathbf{E}_\theta g(\mathrm{x}_1,\theta) l'(\mathrm{x}_1,\theta))^2 \leqslant \mathbf{E}_\theta g^2(\mathrm{x}_1,\theta)\mathbf{E}_\theta (l'(\mathrm{x}_1,\theta))^2. \qquad (19)$$

Это доказывает неравенство (18).

В силу a)–d) все сказанное применимо и к функции $g = l'$, для которой в (19) достигается равенство. Теорема доказана. $\triangleleft$

Замечание 2. Условие c) о законности дифференцирования будет выполнено, если эти интегралы после дифференцирования сходятся равномерно по θ (в окрестности точки θ). Это условие, в свою очередь, будет выполнено, если в (16) и в условии (M_1') из § 14 потребовать, чтобы оценочные интегралы вида $\int l(x) f_\theta(x)\mu(dx)$ сходились равномерно по θ (в окрестности точки θ); см. об этом также в § 34 и в приложении V.

Значение $I(\theta) = \mathbf{E}_\theta (l'(\mathrm{x}_1,\theta))^2$, присутствующее в (17), (18), играет значительную роль в теории оценивания параметров. Оно будет характеризовать наименьший возможный разброс оценки около истинного значения параметра θ не только в классе M-оценок, но и в классе вообще всех оценок.

Асимптотические свойства о.м.п. будут более подробно изучены в § 33–37.

В примерах 1–5 мы имели дело с распределениями, для которых о. м. п. без труда могла быть найдена в явном виде. В общем случае поиски о. м. п. сопряжены обычно со значительными вычислительными трудностями. Например, для параметра сдвига θ распределения Коши с плотностью $[\pi(1 + (x-\theta)^2)]^{-1}$ о. м. п. $\widehat{\theta}^*$ является решением уравнения

$$\sum_{i=1}^{n} \frac{\mathrm{x}_i - \theta}{1 + (\mathrm{x}_i - \theta)^2} = 0$$

степени $2n-1$, так что при больших n отыскание $\widehat{\theta}^*$ становится весьма сложным. Все условия теорем 2, 3 при этом выполнены, так что $\widehat{\theta}^*$ будет а. н. оценкой с параметрами $(0, 2)$ $\left(\text{в этом случае } l'(x,\theta) = \dfrac{2(x-\theta)}{1+(x-\theta)^2},\right.$ $\left.\mathbf{E}_\theta (l'(\mathrm{x}_1,\theta))^2 = \dfrac{4}{\pi}\int \dfrac{x^2}{(1+x^2)^3}\,dx = \dfrac{1}{2}\right)$.

Процедура приближенного вычисления $\widehat{\theta}^*$ в этом примере приведена в § 35 (см. также § 18 об асимптотической эквивалентности $\widehat{\theta}^*$ L-оценке). Там же содержатся другие примеры, удовлетворяющие условиям теорем 2, 3 и иллюстрирующие процедуру вычисления о. м. п.

§ 17. О сравнении оценок

Мы видели, что существует много весьма естественных подходов к получению оценок. Возникает вопрос — как сравнивать между собой разные оценки и какие оценки следует предпочитать другим? В этой главе мы выделим два подхода к сравнению оценок — среднеквадратический и асимптотический. Более общий подход, основанный на использовании элементов

теории игр и произвольных (не только квадратических) функций потерь, изложен в главе 6.

Первый из двух названных подходов основан на сравнении среднеквадратичных уклонений. Второй подход применим лишь к выборкам большого объема, так как он основан на сравнении «рассеиваний» распределений для $(\theta^*-\theta)\sqrt{n}$ при больших n. Основанием для такого сравнения служит обычно вид предельных распределений для $(\theta^*-\theta)\sqrt{n}$ при $n\to\infty$ (если таковые существуют). Соответствующие предельные теоремы дают нам условия, при которых распределение $(\theta^*-\theta)\sqrt{n}$ при больших n можно приближать с помощью упомянутых предельных распределений.

В этом параграфе предполагается, что оценки сравниваются при каком-то одном неизвестном, но фиксированном распределении выборки $\mathbf{P}$.

1. Среднеквадратический подход. Одномерный случай. Этот подход используется при рассмотрении оценок по выборке X любого фиксированного (не обязательно большого) объема. Он состоит в сравнении среднеквадратичных уклонений $\mathbf{E}(\theta^*-\theta)^2$.

Правило 1. Оценку θ_1^*, в соответствии со среднеквадратическим подходом, будем считать *лучшей*, чем θ_2^*, если

$$\mathbf{E}\big(\theta_1^*-\theta\big)^2<\mathbf{E}\big(\theta_2^*-\theta\big)^2. \tag{1}$$

Широко распространено представление о том, что среднеквадратичная погрешность является наиболее подходящей числовой характеристикой точности оценки, хотя это обстоятельство со многих точек зрения является спорным — ведь можно сравнивать, скажем, величины $\mathbf{E}|\theta^*-\theta|$, также описывающие средние значения уклонений θ^* от θ.

Несомненное преимущество характеристик $\mathbf{E}(\theta^*-\theta)^2$ составляет тот факт, что $(\theta^*-\theta)^2$ есть аналитическая функция разности $\theta^*-\theta$. Это делает многие рассмотрения более удобными, а также позволяет приближать, как мы увидим позже, значения $\mathbf{E}f(\theta^*-\theta)$ для гладких функций f.

Наряду со среднеквадратичным уклонением для описания свойств оценок часто используется также величина смещения.

Определение 1. *Смещением* оценки θ^* называется величина

$$b=\mathbf{E}\theta^*-\theta.$$

Оценка θ^*, для которой $b=0$, называется *несмещенной*.

Среднеквадратичное уклонение связано со смещением и дисперсией оценки равенством

$$\mathbf{E}(\theta^*-\theta)^2=\mathbf{D}\theta^*+b^2,$$

так что для несмещенных оценок среднеквадратичное уклонение совпадает с дисперсией.

Само по себе свойство несмещенности представляется, очевидно, желательным свойством оценок, поскольку оно означает, что в данной последовательности оцениваний среднее значение оценок будет совпадать с истинным значением параметра. Если этого свойства нет, то оценка называется *смещенной*.

Пример 1. Рассмотрим следующие три оценки для среднего значения $\theta = \mathbf{E}\mathrm{x}_1$ распределения $\mathbf{P}$

$$\theta_1^* = \overline{\mathrm{x}}, \quad \theta_2^* = \zeta^*, \quad \theta_3^* = \frac{\mathrm{x}_{(1)} + \mathrm{x}_{(n)}}{2}, \tag{2}$$

где ζ^* — выборочная медиана, $\mathrm{x}_{(k)}$, $k = 1, \ldots, n$, — значения вариационного ряда, так что $\zeta^* = \mathrm{x}_{((n+1)/2)}$, если n нечетно, и $\zeta^* = \frac{1}{2}(\mathrm{x}_{(n/2)} + \mathrm{x}_{(n/2+1)})$, если n четно (при $n = 1, 2$ все три оценки совпадают). Все оценки являются несмещенными, если распределение $\mathbf{P}$, из которого извлечена выборка, симметрично относительно θ ($\mathbf{P}((-\infty, \theta - t)) = \mathbf{P}((\theta + t, \infty))$ при любом $t \geqslant 0$). Это следует из того, что распределение всех трех оценок также будет симметричным относительно θ. Для $\overline{\mathrm{x}}$ утверждение о несмещенности $\mathbf{E}\overline{\mathrm{x}} = \theta$ очевидно и без предположения о симметрии.

Вычислим среднеквадратичные уклонения оценок (2). Для простоты мы ограничимся случаем $\mathbf{P} = \mathbf{U}_{0,1}$, $n = 3$, для которого оценки (2) перейдут в

$$\theta_1^* = \overline{\mathrm{x}}, \quad \theta_2^* = \mathrm{x}_{(2)}, \quad \theta_3^* = \frac{\mathrm{x}_{(1)} + \mathrm{x}_{(3)}}{2}.$$

Имеем

$$\mathbf{D}\mathrm{x}_1 = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12}, \quad \mathbf{E}\left(\theta_1^* - \theta\right)^2 = \mathbf{D}\overline{\mathrm{x}} = \frac{\mathbf{D}\mathrm{x}_1}{3} = \frac{1}{36}.$$

Далее, в силу определения медианы (n нечетно) $\{\zeta^* < x\} = \{F_n^*(x) > 1/2\}$ и, стало быть,

$$\mathbf{P}(\zeta^* < x) = \mathbf{P}\left(F_n^*(x) > \frac{1}{2}\right) = \sum_{k=(n+1)/2}^{n} \mathbf{P}\left(nF_n^*(x) = k\right). \tag{3}$$

Для $n = 3$

$$\mathbf{P}\left(F_3^*(x) = 1\right) = \mathbf{P}\left(\bigcap_{i=1}^{3} \{\mathrm{x}_i < x\}\right) = F^3(x),$$

$$\mathbf{P}\left(3F_3^*(x) = 2\right) = 3F^2(x)(1 - F(x)).$$

Вероятность $\mathbf{P}(\zeta^* \in (u, u+du))$ складывается из вероятностей событий вида $\{\mathrm{x}_1 \in (u, u + du)\}\{\mathrm{x}_2 < u\}\{\mathrm{x}_3 > u\}$. Так как всего возможно шесть таких комбинаций, то $\mathbf{P}(\zeta^* \in (u, u + du)) = 6f(u)F(u)(1 - F(u))\,du$, и, следовательно, ζ^* имеет плотность, равную (это вытекает и из (3))

$$6f(u)F(u)(1 - F(u)),$$

где $F(u) = \int\limits_{-\infty}^{u} f(t)\,dt = \mathbf{P}(\mathrm{x}_1 < u)$. В случае $\mathbf{P} = \mathbf{U}_{0,1}$ эта плотность будет равна $6x(1-x)$ при $x \in [0,1]$, так что

$$\mathbf{E}(\zeta^*)^2 = \int\limits_0^1 6x^3(1-x)\,dx = 6\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{3}{10},$$

$$\mathbf{D}\zeta^* = \mathbf{E}(\zeta^*)^2 - (\mathbf{E}\zeta^*)^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

Нам осталось найти дисперсию оценки

$$\theta_3^* = \frac{\mathrm{x}_{(1)} + \mathrm{x}_{(3)}}{2}.$$

Рассуждая аналогично предыдущему, нетрудно убедиться, что вероятность $\mathbf{P}(\mathrm{x}_{(1)} \in (u, u+du),\ \mathrm{x}_{(3)} \in (v, v+dv))$ при $u < v$ равна $6f(u)f(v) \times$ $\times\,(F(v) - F(u))\,du\,dv$. Поэтому для $\mathbf{P} = \mathbf{U}_{0,1}$

$$\mathbf{E}\big(\theta_3^*\big)^2 = \int\limits_0^1 \int\limits_0^v \left(\frac{u+v}{2}\right)^2 6(v-u)\,du\,dv.$$

Значение этого интеграла равно $11/40$ (вычисления читатель может провести самостоятельно), так что

$$\mathbf{D}\theta_3^* = \mathbf{E}\big(\theta_3^*\big)^2 - \big(\mathbf{E}\theta_3^*\big)^2 = \frac{11}{40} - \frac{1}{4} = \frac{1}{40}.$$

Итак, оценка θ_3^* оказалась наилучшей. Для других значений n и других распределений $\mathbf{P}$ положение может оказаться другим. Мы увидим, например, что для $\mathbf{P} = \mathbf{\Phi}_{\alpha,\sigma^2}$ наилучшей оценкой для α будет $\theta_1^* = \overline{\mathrm{x}}$.

Пример 2. *Несмещенные оценки дисперсии.* Рассмотрим оценку для дисперсии

$$S^2 = \frac{1}{n}\sum(\mathrm{x}_i - \overline{\mathrm{x}})^2 = \frac{1}{n}\sum \mathrm{x}_i^2 - (\overline{\mathrm{x}})^2,$$

а также оценку

$$S_1^2 = \frac{1}{n}\sum(\mathrm{x}_i - \mathbf{E}\mathrm{x}_i)^2 = \frac{1}{n}\sum \mathrm{x}_i^2 + (\mathbf{E}\mathrm{x}_1)^2 - 2\overline{\mathrm{x}}\mathbf{E}\mathrm{x}_1$$

(обе по принципу подстановки) в том случае, когда $\mathbf{E}\mathrm{x}_1$ известно. Оценка S_1^2, очевидно, является несмещенной. В то же время

$$S^2 = \frac{1}{n}\sum(\mathrm{x}_i - \overline{\mathrm{x}})^2 = \frac{1}{n}\sum(\mathrm{x}_i - \overline{\mathrm{x}} \pm \mathbf{E}\mathrm{x}_1)^2 =$$

$$= \frac{1}{n}\sum(\mathrm{x}_i - \mathbf{E}\mathrm{x}_1)^2 - (\overline{\mathrm{x}} - \mathbf{E}\mathrm{x}_1)^2 = S_1^2 - (\overline{\mathrm{x}} - \mathbf{E}\mathrm{x}_1)^2 < S_1^2.$$

Таким образом, оценка S^2 является смещенной,

$$\mathbf{E}S^2 = \mathbf{D}\mathrm{x}_1 - \mathbf{D}\overline{\mathrm{x}} = \left(1 - \frac{1}{n}\right)\mathbf{D}\mathrm{x}_1.$$

Это соотношение показывает, что мы можем рассматривать и в случае неизвестного $\mathbf{E}\mathrm{x}_1$ несмещенную оценку дисперсии, равную

$$S_0^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum(\mathrm{x}_i - \overline{\mathrm{x}})^2, \quad \mathbf{E}S_0^2 = \mathbf{D}\mathrm{x}_1.$$

Перейдем теперь к асимптотическому подходу к проблеме сравнивания оценок. В этом случае правило для предпочтения оценок выбирается однозначно.

2. Асимптотический подход. Одномерный случай. Пусть даны две оценки θ_1^* и θ_2^* такие, что

$$\frac{(\theta_1^* - \theta)\sqrt{n}}{\sigma_1} \Rightarrow \mathbf{Q}, \quad \frac{(\theta_2^* - \theta)}{\sigma_2} \Rightarrow \mathbf{Q}, \tag{4}$$

где $\mathbf{Q}$ — некоторый предельный закон распределения, один и тот же для θ_1^* и θ_2^*, а $\sigma_2 > \sigma_1$. Тогда при больших n распределения $(\theta_i^* - \theta)\sqrt{n}/\sigma_i$, $i = 1, 2$, будут близки к $\mathbf{Q}$, и несомненно, что «рассеивание» θ_2^* вокруг θ будет больше «рассеивания» θ_1^*, и нам следует предпочесть θ_1^*. Числа σ_i естественно называть *коэффициентами рассеивания*.

Таким образом, существо асимптотического подхода состоит в сравнении предельных распределений оценок.

Мы уже видели и будем убеждаться в этом и в дальнейшем, что многие возникающие естественным образом оценки, в том числе и оптимальные (об этом ниже), являются асимптотически нормальными, т. е. для них справедливо (4) при $\mathbf{Q} = \mathbf{\Phi}_{0,1}$. Это позволяет нам сформулировать следующее естественное правило сравнения а. н. оценок.

Пусть даны две а. н. оценки θ_1^* и θ_2^* с коэффициентами рассеивания σ_1^2 и σ_2^2 соответственно.

Правило 2. Оценку θ_1^* следует признать *лучшей*, чем θ_2^*, если $\sigma_1^2 < \sigma_2^2$.

Мы использовали это правило для выделения оптимальных оценок в п. 16.3.

В дальнейшем, при использовании этих и других правил, будем наряду с термином «лучше» использовать, где это требуется, также слова «не хуже», «хуже», «не лучше». Они будут соответствовать знакам неравенства $\leqslant$, $>$, $\geqslant$ между σ_1^2 и σ_2^2 (или между $\mathbf{E}(\theta_1^* - \theta)^2$ и $\mathbf{E}(\theta_2^* - \theta)^2$ в (1)). Если $\sigma_1^2 = \sigma_2^2$, то оценки мы будем называть асимптотически эквивалентными. Предлагаемое соглашение естественно, и в дальнейших определениях мы оговаривать его каждый раз не будем, а будем ограничиваться лишь определением отношения «лучше» или аналогичных ему.

Отметим, что в классе а. н. оценок минимальность рассеивания θ^* означает, что величина

$$\lim_{n\to\infty} \mathbf{P}(|\theta^* - \theta| < u\sqrt{n})$$

будет максимальной при каждом u. Это обстоятельство делает указанное правило для сравнения а. н. оценок бесспорным.

Асимптотический подход при всей своей естественности обладает существенным недостатком: *он применим лишь к выборкам большого объема и лишь в классе а. н. оценок.*

Отмеченные два подхода в известном смысле близки друг к другу — в обоих случаях дело сводится к сравнению дисперсий или величин, близких к ним. Конечно, величина σ_i^2/n в (4) при $\mathbf{Q} = \mathbf{\Phi}_{0,1}$ может, вообще говоря, существенно отличаться от $\mathbf{E}(\theta^* - \theta)^2$. Однако примеры, иллюстрирующие этот факт (мы предлагаем читателю построить их), носят, как правило, искусственный характер.

Дальнейшее изложение этой главы во многом связано с построением оценок, оптимальных для каждого из двух введенных подходов.

Пример 3. Пусть $X \in \mathbf{\Gamma}_{\alpha,1}$. В примере 14.1 было показано, что обе оценки

$$\alpha_1^* = (\overline{\mathrm{x}})^{-1}, \quad \alpha_2^* = \left(\frac{1}{2n}\sum \mathrm{x}_i^2\right)^{-1/2}$$

являются оценками по методу моментов. Кроме того, α_1^* является также о. м. п. Было установлено, далее, что обе оценки а. н. с коэффициентами соответственно α^2 и $\frac{5}{4}\alpha^2$, так что оценка α_1^* лучше, чем α_2^*, с точки зрения асимптотического подхода. Тот же результат при $n \geqslant 2$ получим для среднеквадратического подхода.

Приведем теперь пример, показывающий, что в зависимости от свойств распределения одна и та же оценка может быть лучше или хуже, чем некоторая другая фиксированная оценка.

Пример 4. Рассмотрим задачу об оценке $\theta = \mathbf{E}\mathrm{x}_1$, если известно, что $X \in \mathbf{P}$ и $\mathbf{P}$ — симметричное относительно точки θ распределение (ср. с примером 1). В этом случае медиана распределения ζ совпадает с θ. Мы рассмотрим две оценки для θ (обе по принципу подстановки): среднее $\theta_1^* = \overline{\mathrm{x}}$ и выборочную медиану $\theta_2^* = \zeta^*$. Пусть, для определенности, n нечетно. Из следствия 12.1 при $k = (n+1)/2$ следует, что если функция распределения F непрерывно дифференцируема в точке $\theta = \zeta$, то

$$(\zeta^* - \zeta)\sqrt{n} \Rightarrow \frac{\xi}{2f(\theta)}, \quad \xi \in \mathbf{\Phi}_{0,1}, \quad f(x) = F'(x).$$

Иными словами, в этом случае ζ^* является а. н. оценкой с коэффициентом $\sigma_2^2 = 1/(4f^2(\zeta))$.

С другой стороны, а. н. оценка $\overline{\mathrm{x}}$ имеет коэффициент $\sigma_1^2 = \mathbf{D}\mathrm{x}_1$. Таким образом, если

$$\int (x-\zeta)^2\, dF(x) < \frac{1}{4f^2(\zeta)},$$

то мы должны предпочесть оценку $\overline{\mathrm{x}}$. Если знак неравенства будет обратным, то ζ^*. Следует отметить, что числа $\int (x-\zeta)^2\, dF(x)$ и $f(\zeta)$ представляют очень мало связанные между собой характеристики распределения.

Рассмотрим важный частный случай, когда мы оцениваем параметр α по выборке $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$. В этом случае $f(\alpha) = f(\zeta) = 1/\sigma\sqrt{2\pi}$, так что

$$\sigma_2^2 = \frac{\pi}{2}\sigma^2 > \sigma^2 = \sigma_1^2.$$

Это значит, что статистика $\overline{x}$, являющаяся в данной ситуации о. м. п. для a, лучше, чем ζ^*. Это обстоятельство тесно связано с общим фактом, который будет установлен нами несколько позже, что в широких классах а. н. оценок θ^* о. м. п. $\widehat{\theta}^*$ является асимптотически наилучшей: ее коэффициент рассеивания является наименьшим в этих классах.

Возвращаясь к примеру 4 о сравнении выборочных медианы и среднего, еще раз отметим, что, как было показано, существуют распределения, для которых выборочная медиана будет предпочтительнее $\overline{x}$.

Пример с медианой весьма поучителен и в другом отношении. Он показывает, что скорость убывания степени рассеивания $\zeta^* - \zeta$ с ростом n может быть любой. Чтобы убедиться в этом, достаточно обратиться к замечанию 12.1. В условиях этого замечания нормирующим множителем, который обеспечивает сходимость распределения $\zeta^* - \zeta$ к предельному распределению, является величина $n^{1/(2\gamma)}$, где γ может быть любым неотрицательным числом (см. (12.12)). Множитель $\sqrt{n}$ соответствует лишь гладким распределениям.

Приведем теперь один численный эксперимент с выборкой X объема $n = 101$ из нормальной совокупности $\mathbf{\Phi}_{0,1}$ и посмотрим*), как значения $\overline{x}$ и ζ^* приближают 0 при $n = 11, 21, 51, 101$. Полученные данные приведены в следующей таблице:

n	11	21	51	101
$\overline{x}$	−0,283	−0,254	−0,148	−0,072
ξ^*	−0,291	−0,292	−0,078	−0,044

В этом примере оценка ζ^* при $n = 51, 101$ ведет себя лучше, что есть результат случайного отклонения. Чтобы почувствовать преимущества $\overline{x}$, надо было бы провести много таких экспериментов.

Приведем теперь некоторые общие результаты, связанные с рассмотренными примерами.

3. Нижняя граница рассеивания для L-оценок. В связи со сформулированным выше правилом сравнения а. н. оценок возникает естественный вопрос, насколько малым может быть сделан коэффициент рассеивания этих оценок. Для некоторых классов оценок нижняя граница возможных значений рассеиваний может быть найдена в явном виде. Для M-оценок такие границы были найдены в теореме 16.3. Здесь мы остановимся на этой задаче для L-оценок, т. е. для оценок, являющихся L-статистиками, рассмотренными нами в § 3, 8. Это есть класс оценок (или статистик), представимых в виде линейной комбинации членов вариационного ряда (мы предполагаем,

*) Выборка X построена с помощью случайных чисел, взятых из таблиц в [10] (взяты первые 101 чисел на с. 434).

что $\mathrm{x}_i \in \mathbb{R}$), или функций от них. Как и в главе 1, ограничимся L-оценками не самого общего вида:

$$\theta^* = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\frac{i}{n}\right) g(\mathrm{x}_{(i)}) = \int_{-\infty}^{\infty} \varphi\big(F_n^*(u)\big) g(u)\, dF_n^*(u) = G\big(F_n^*\big), \tag{5}$$

или в несколько иной, близкой, записи:

$$\theta^* = \int_0^1 \varphi(t) g\big(F_n^{-1}(t)\big)\, dt, \tag{6}$$

где φ, g — заданные функции, F_n^{-1} — функция, обратная к F_n^*. В § 8 было установлено (см. следствие 8.6, замечание 8.3), что если $X \in F$, φ и F непрерывно дифференцируемы, то L-оценки являются а. н. оценками:

$$\begin{gathered} \sqrt{n}\,\big(G\big(F_n^*\big) - G(F)\big) \Rightarrow \mathbf{\Phi}_{0,\sigma^2}, \\ \sigma^2 = \int_0^1 A^2(t)\, dt - A^2, \quad A = \int_0^1 A(t)\, dt, \\ A(t) = \int_t^1 a(t)\, dt, \quad a(t) = \frac{\varphi(t) g'(F^{-1}(t))}{F'(F^{-1}(t))}. \end{gathered} \tag{7}$$

Так как σ^2 инвариантно относительно замены $A(t)$ на $\pm A(t) + c$, то можно сказать также, что в (7) $A(t)$ — любая функция с производной $A'(t)$, равной $-a(t)$, и считать ее выбранной так, что $A = 0$. Простоты ради положим $g(x) \equiv x$. Это не очень ограничивает общность, так как для монотонной g значения $g(\mathrm{x}_{(i)})$ вместе с $\mathrm{x}_{(i)}$ образуют вариационный ряд (см. (5)). В этом случае все разнообразие L-оценок доставляется функцией φ. Фиксируем значение

$$\int_0^1 \varphi(t)\, dt = \mathrm{const} = \varphi_0, \tag{8}$$

где φ_0 может выбираться, например, так, чтобы оценка θ^* была несмещенной.

Итак, если $X \in F$, F имеет плотность f, то в сделанных предположениях мы можем записать

$$\sigma^2 = \int_0^1 A^2(t)\, dt, \quad A'(t) = -a(t) = -\frac{\varphi(t)}{f(F^{-1}(t))}, \tag{9}$$

где φ удовлетворяет (8). Через $\dot f$, $\dot l$, $\ddot f$, $\ddot l$ будем обозначать производные $f(x)$ и $l(x) = \ln f(x)$ по x, сохраняя обозначения f'_θ, f''_θ, l', l'' для производных $f_\theta(x)$ и $l(x,\theta) = \ln f_\theta(x)$ по параметру θ.

Теорема 1. *Если f имеет производную $\dot f$, то при всех φ, удовлетворяющих* (8),

$$\sigma^2 \geqslant \sigma_0^2 = \varphi_0^2 I^{-1}, \tag{10}$$

$$I = \int \frac{(\dot f(u))^2}{f(u)}\,du = \int (\dot l(u))^2 f(u)\,du. \tag{11}$$

Если $I < \infty$, f имеет вторую производную $\ddot f$ и существует $\int \ddot f(u)\,du$, то $I = -\int \ddot l f(u)\,du$ и равенство в (10) *достигается при*

$$\varphi(t) = \varphi_0 I^{-1} \ddot l(F^{-1}(t)), \tag{12}$$

где φ_0 из условия (8).

Здесь мы опять сталкиваемся с тем, что интеграл I, аналогичный (16.17), определяет нижнюю границу возможных значений коэффициентов разброса.

Доказательство. Так как $f(u) \to 0$ при $u \to \pm\infty$, то, интегрируя по частям, получим в силу (9)

$$\varphi_0 = \int_0^1 \varphi(t)\,dt = \int_0^1 a(t) f(F^{-1}(t))\,dt = \int_0^1 A(t) \frac{\dot f(F^{-1}(t))}{f(F^{-1}(t))}\,dt.$$

По неравенству Коши–Буняковского

$$\varphi_0^2 \leqslant \int_0^1 A^2(t)\,dt \int_0^1 \left[\frac{\dot f(F^{-1}(t))}{f(F^{-1}(t))}\right]^2 dt, \tag{13}$$

где второй интеграл после замены $t = F(u)$ превращается в I. Отсюда следует (10). Равенство (11) следует из соотношений

$$\dot f = \dot l f, \quad \ddot f = \ddot l f + (\dot l)^2 f, \quad \int \ddot f(u)\,du = 0,$$

последнее из которых получается, если два раза продифференцировать по θ в точке $\theta = 0$ равенство

$$\int f(u+\theta)\,du = 1.$$

Знак равенства в (13) достигается при

$$A(t) = c\frac{\dot f(F^{-1}(t))}{f(F^{-1}(t))} = c\dot l(F^{-1}(t)). \tag{14}$$

Отрицательная производная этой функции равна

$$a(t) = -c\frac{\ddot l(F^{-1}(t))}{f(F^{-1}(t))}.$$

Сравнивая это с (9), получаем

$$\varphi(t) = -c\ddot{l}(F^{-1}(t)).$$

Для этой функции $A(t) = c\dot{l}(F^{-1}(t)) + c_1$. Условие $A = 0$ и соотношение (14) будут выполнены при $c_1 = 0$. Условие (8) будет выполнено при $c = -\varphi_0 I^{-1}$. Теорема доказана. $\triangleleft$

Заметим, что свойство $\ddot{l}(u) = \text{const}$ реализуется лишь для нормального закона. С другой стороны, из (6) видно, что оценка θ^* будет статистикой I типа (см. § 3) лишь при $\varphi(u) = \text{const}$. Это означает, что нижняя граница возможных значений σ^2 может достигаться на статистиках I типа (в классе L-статистик) лишь для выборок из нормальных совокупностей.

Посмотрим теперь, как выглядят два подхода к сравнению оценок, сформулированные в п. 1, 2 в многомерном случае, когда θ есть вектор $(\theta_1, \ldots, \theta_k)$.

4. Среднеквадратический и асимптотический подходы в многомерном случае. Асимптотический подход, как и прежде, мы будем использовать лишь в классе а. н. оценок. В этом случае все дело сводится к сравнению многомерных нормальных распределений (предельных для $(\theta^* - \theta)\sqrt{n}$), которые полностью описываются матрицей вторых моментов σ^2 (см., например, теорему 13.2А).

Если рассматривать среднеквадратический подход к сравнению точных распределений θ^*, то дело также сведется к возможности сравнивать два распределения в $\mathbb{R}^k$ на основании знания моментов $(\theta^* - \theta)$ второго порядка. Таким образом, в обоих случаях мы должны уметь сравнивать по «степени рассеивания» матрицы моментов второго порядка.

Рассмотрим наиболее естественные способы сравнения. Пусть $\mathbf{Q}_1$ и $\mathbf{Q}_2$ — произвольные распределения в $\mathbb{R}^k$. Обозначим через ξ_1 и ξ_2 какие-нибудь случайные векторы, обладающие этими распределениями: $\xi_i \Subset \mathbf{Q}_i$.

Определение 2. Мы будем говорить, что среднеквадратичное рассеивание распределения $\mathbf{Q}_1$ около точки $\alpha \in \mathbb{R}^k$ не больше, чем рассеивание $\mathbf{Q}_2$, если для любого вектора $a = (a_1, \ldots, a_k)$

$$\mathbf{E}(\xi_1 - \alpha, a)^2 \leqslant \mathbf{E}(\xi_2 - \alpha, a)^2, \tag{15}$$

где $(x, a) = \sum_{i=1}^{k} x_i a_i$ есть скалярное произведение.

Будем говорить, что рассеивание для $\mathbf{Q}_1$ меньше, чем для $\mathbf{Q}_2$, если в (15) имеет место знак строгого неравенства хотя бы для одного a.

Если $\alpha = \mathbf{E}\xi_1 = \mathbf{E}\xi_2$, то неравенство (15) означает, что по любому направлению a дисперсия распределения $\mathbf{Q}_1$ (т. е. дисперсия проекции ξ_1 на a) не превосходит такой же величины для $\mathbf{Q}_2$.

Если $d_l^2 = \|d_{ij}^{(l)}\|$ — матрица вторых моментов $\mathbf{Q}_l$, $l = 1, 2$, то, раскрывая скобки в (5) при $\alpha = 0$, получим для любых $a_1, \ldots, a_k$

$$\sum_{i,j=1}^{k} d_{ij}^{(1)} a_i a_j \leqslant \sum_{i,j=1}^{k} d_{ij}^{(2)} a_i a_j. \tag{16}$$

Это соотношение на языке матриц мы будем обозначать

$$d_1^2 \leqslant d_2^2. \tag{17}$$

Оно означает неотрицательную определенность матрицы $d_2^2 - d_1^1$.

Таким образом, среднеквадратичное рассеивание $\mathbf{Q}_1$ около нуля не превосходит такового для $\mathbf{Q}_2$ тогда и только тогда, когда для матриц моментов второго порядка имеют место неравенства (16), (17).

Правила предпочтения оценок в многомерном случае можно сформулировать следующим образом.

Среднеквадратический подход: оценка θ_1^* лучше, чем θ_2^*, если среднеквадратичное рассеивание θ_1^* около точки θ меньше, чем такая же величина для θ_2^*. Если d_l^2 есть матрица вторых моментов $\theta_l^* - \theta$, то утверждение «оценка θ_1^* лучше, чем θ_2^*» означает, что $d_1^2 < d_2^2$.

Асимптотический подход: оценка θ_1^* лучше, чем θ_2^*, если среднеквадратичное рассеивание около нуля предельного распределения для $(\theta_1^* - \theta)\sqrt{n}$ меньше, чем такая же величина для $(\theta_2^* - \theta)\sqrt{n}$.

Другими словами, если $(\theta_l^* - \theta)\sqrt{n} \Rightarrow \Phi_{\alpha,\sigma_l^2}$, то утверждение «оценка θ_1^* лучше, чем θ_2^*» означает, что $\sigma_1^2 < \sigma_2^2$.

Можно показать, что если θ_1^* и θ_2^* — две а. н. оценки и θ_1^* лучше, чем θ_2^*, то

$$\lim_{n\to\infty} \big((\theta_1^* - \theta)\sqrt{n} \in B\big) > \lim_{n\to\infty} \mathbf{P}\big((\theta_2^* - \theta)\sqrt{n} \in B\big) \tag{18}$$

для любого центрального эллипсоида *) B.

Мы видим, что в обоих случаях сравнение оценок сводится к установлению неравенств для матриц моментов второго порядка. Некоторое отличие состоит в том, что в первом случае моменты не обязательно центральные.

Установим теперь некоторые соотношения, эквивалентные (16), (17). Положим

$$v(\theta^*) = \mathbf{E}(\theta^* - \theta)V(\theta^* - \theta)^\top$$

и обозначим через $\mathfrak{B}_+$ множество всех неотрицательно определенных матриц $V = \|v_{ij}\|$. Если $\|d_{ij}\|$ есть матрица вторых моментов $\theta^* - \theta$, то, очевидно, $v(\theta^*) = \sum v_{ij}d_{ij}$.

Лемма 1. *$d_1^2 \leqslant d_2^2$ тогда и только тогда, когда $v(\theta_1^*) \leqslant v(\theta_2^*)$ для любых $V \in \mathfrak{B}_+$.*

Доказательство. В одну сторону утверждение очевидно, так как матрица $V_a = \|a_i a_j\| \in \mathfrak{B}_+$, и для такой матрицы

$$v_a(\theta_l^*) = \mathbf{E}(\theta_l^* - \theta)V_a(\theta_l^* - \theta)^\top = \sum a_i a_j d_{ij}^{(l)}$$

(см. (16)).

---

*) Ради краткости условимся область $\sum\limits_{i,j=1}^{k} d_{ij}x_i x_j \leqslant c$ называть *эллипсоидом* в $\mathbb{R}^k$, а поверхность $\sum\limits_{i,j=1}^{k} d_{ij}x_i x_j = c$ — *эллипсом*.

Чтобы доказать утверждение в обратную сторону, заметим, что частичный порядок, основанный на неравенствах (15), инвариантен относительно вращения осей координат. Именно, если C — матрица ортогонального преобразования и θ_1^* лучше θ_2^* для параметра θ, то $\theta_1^* C$ лучше $\theta_2^* C$ для параметра θC. Это следует из равенств

$$\big(\theta_l^* C - \theta C, a\big) = \big((\theta_l^* - \theta)C, a\big) = \big(\theta_l^* - \theta, aC^\top\big)$$

и определения 2.

Пусть теперь $d_1^2 < d_2^2$, т. е.

$$\sum d_{ij}^{(1)} a_i a_j < \sum d_{ij}^{(2)} a_i a_j. \tag{19}$$

Это означает, что $v\big(\theta_1^*\big) < v\big(\theta_2^*\big)$ для матриц V вида $V_a = \|a_i a_j\|$, а стало быть, и для диагональных матриц $V_{\text{diag}} \in \mathfrak{B}_+$, поскольку последние представимы в виде суммы k матриц вида V_a. Пусть теперь V — произвольная матрица из $\mathfrak{B}_+$, C есть ортогональное преобразование такое, что $C^\top V C = V_{\text{diag}}$. Тогда

$$v\big(\theta_1^*\big) = \mathbf{E}\big(\theta_1^* - \theta\big)V\big(\theta_1^* - \theta\big)^\top = \mathbf{E}\big(\theta_1^* - \theta\big)CV_{\text{diag}}C^\top\big(\theta_1^* - \theta\big)^\top.$$

Из сделанных выше двух замечаний и из (19) следует, что правая часть этого равенства меньше, чем

$$\mathbf{E}\big(\theta_2^* - \theta\big)CV_{\text{diag}}C^\top\big(\theta_2^* - \theta\big)^\top = \mathbf{E}\big(\theta_2^* - \theta\big)V\big(\theta_2^* - \theta\big)^\top = v\big(\theta_2^*\big). \quad \lhd$$

Существует и другой способ сравнения рассеивания (см. [58]), который предполагает, однако, что оба распределения $\mathbf{Q}_1$ и $\mathbf{Q}_2$ не вырождены в $\mathbb{R}^k$ и имеют нулевое среднее. В этом случае матрицы центральных вторых моментов d_l^2 будут положительно определены, и для них существуют обратные $A_l = \big(d_l^2\big)^{-1}$.

Пусть d^2 есть матрица вторых моментов распределения $\mathbf{Q}$ и $A = (d^2)^{-1}$.

Определение 3. *Эллипсоидом рассеивания распределения* $\mathbf{Q}$ называется эллипсоид

$$tAt^\top \leqslant k + 2,$$

который однозначно выделяется среди всех эллипсоидов следующим своим свойством: если рассмотреть равномерное распределение $\mathbf{U}$ на этом эллипсоиде (т. е. распределение в $\mathbb{R}^k$ с постоянной плотностью внутри эллипсоида и нулевой плотностью вне его), то первые и вторые моменты $\mathbf{Q}$ и $\mathbf{U}$ совпадают (см. [58, с. 333]).

Лемма 2. *Пусть матрицы* d_l^2, $l = 1, 2$, *невырождены. Среднеквадратичное рассеивание* $\mathbf{Q}_1$ *около нуля не больше рассеивания* $\mathbf{Q}_2$ *тогда и только тогда, когда эллипсоид рассеивания для* $\mathbf{Q}_1$ *лежит в эллипсоиде для* $\mathbf{Q}_2$.

Доказательство. Пусть эллипс $tA_1t^\top = 1$ лежит внутри $tA_2t^\top = 1$. Как известно, существует невырожденное линейное преобразование

$t = uL$, которое переводит эллипс $tA_1t^\top = 1$ в единичную сферу S_1, а эллипс $tA_2t^\top = 1$ — в эллипс S_2 с главными осями по направлениям осей координат. Это означает, что $\widetilde{A}_1 \equiv LA_1L^\top = E$ (единичная матрица), $\widetilde{A}_2 \equiv LA_2L^\top = \operatorname{diag}(\lambda_1^2, \ldots, \lambda_k^2)$, $0 < \lambda_j^2 \leqslant 1$, $j = 1, \ldots, k$. Так как $\widetilde{A}_1^{-1} = E$, $\widetilde{A}_2^{-1} = \operatorname{diag}(\lambda_1^{-2}, \ldots, \lambda_k^{-2})$, то эллипс $t\widetilde{A}_2^{-1}t^\top = 1$ будет инверсией относительно единичной сферы S_1 эллипса S_2, и, следовательно, он будет располагаться в S_1. Так как $\widetilde{A}_2^{-1} = (L^\top)^{-1}A_2L^{-1}$, то, делая «обратное» преобразование $u = tL^\top$, получим, что эллипс $tA_1^{-1}t^\top = td_1^2t^\top = 1$ лежит снаружи от $tA_2^{-1}t^\top = td_2^2t^\top = 1$. Очевидно, что такое же соотношение справедливо для эллипсов $td_1^2t^\top = c$ и $td_2^2t^\top = c$. Но это означает, что равенство $td_1^2t^\top = c$ влечет за собой $td_1^2t^\top = c \leqslant td_2^2t^\top$. Утверждение в обратную сторону доказывается точно так же. $\triangleleft$

Важно заметить теперь, что в отличие от одномерного случая сравнение рассеиваний распределений с помощью матриц вторых моментов устанавливает лишь частичный порядок на множестве всех распределений. Например, матрицы $d_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ и $d_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ не лучше и не хуже одна другой, так как для вектора $a = (1, 0)$ неравенство (16) справедливо, а для вектора $a = (0, 1)$ оно будет обратным. Это существенное неудобство введенного порядка, хотя сам по себе он сомнений не вызывает.

Мы можем сделать множество оценок (или множество распределений) вполне упорядоченным, если будем сравнивать, скажем, $\mathbf{E}|\theta^* - \theta|^2$, где $|\cdot|$ есть евклидова норма в $\mathbb{R}^k$, так что

$$\mathbf{E}|\theta^* - \theta|^2 = \mathbf{E}\sum_{i=1}^{k}(\theta_i^* - \theta_i)^2. \tag{20}$$

Такой способ упорядочивания является уже спорным, так как в разных обстоятельствах точность по различным направлениям может цениться поразному. Чтобы как-то учесть это обстоятельство, можно в качестве обобщения рассматривать меру точности

$$v(\theta^*) = \mathbf{E}(\theta^* - \theta)V(\theta^* - \theta)^\top,$$

где V — неотрицательно определенная матрица (случай (20) соответствует $V = E$).

Из леммы 1 следует, что если рассеивание θ_1^* около θ меньше, чем рассеивание θ_2^*, то $v(\theta_1^*) < v(\theta_2^*)$. Обратное, вообще говоря, неверно — выполнение неравенства $v(\theta_1^*) < v(\theta_2^*)$ для какой-нибудь одной матрицы V (предложенный выше полный порядок основан на фиксированной матрице) не означает еще, что рассеивание θ_1^* около θ меньше, чем рассеивание θ_2^*.

5. Некоторые эвристические подходы к определению дисперсий оценок. Методы складного ножа и будстрэп. Мы видели, что важнейшую роль при сравнении оценок играют величины дисперсий этих оценок. Однако вычисление этих величин может представлять значительные трудности. В связи

с этим был предложен ряд эвристических процедур (иногда они могут быть теоретически обоснованными) для определения дисперсий оценок по самой выборке. Мы ограничимся рассмотрением лишь одномерного случая $\theta \in \mathbb{R}$.

1. *Метод складного ножа* (jacknife; см. [78, 107]). Для определения дисперсии заданной оценки $\theta^* = \theta_n^*(X)$, построенной по выборке X объема n (т. е. заданной функции от X и n) предлагается рассмотреть n «подвыборок» $X^{(1)}, \ldots, X^{(n)}$ объема $n-1$, где $X^{(i)}$ получается из X выбрасыванием i-го наблюдения x_i. По каждой из этих подвыборок строятся оценки $\theta_{n-1}^*(X^{(i)})$ и $\overline{\theta}^*(X) = \dfrac{1}{n}\sum\limits_{i=1}^{n} \theta_{n-1}^*(X^{(i)})$ (последняя оценка, как правило, имеет меньшее смещение по сравнению с $\theta^*(X)$). В качестве оценки дисперсии $\sigma^2 = \mathbf{D}\theta^*$ предлагается использовать значение

$$(\sigma^2)^* = \frac{n-1}{n}\sum_{i=1}^{n}\bigl(\theta_{n-1}^*(X^{(i)}) - \overline{\theta}^*(X)\bigr)^2.$$

2. *Будстрэп* (bootstrap; см. [47]). Здесь для оценки $\sigma^2 = \mathbf{D}\theta^* = \sigma^2(F)$, $X \Subset F$ предлагается использовать значение $\sigma^2\bigl(F_n^*\bigr)$, а для оценки $\sigma^2\bigl(F_n^*\bigr)$ — следующую процедуру типа Монте-Карло. По выборке $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ строятся новые выборки $X^{(1)}, \ldots, X^{(N)}$, каждая из которых получается независимым образом как выборка из распределения F_n^*. Иначе говоря, элементами $X^{(i)} = \bigl(\mathrm{x}_1^{(i)}, \ldots, \mathrm{x}_n^{(i)}\bigr)$ являются $\mathrm{x}_1, \ldots, \mathrm{x}_n$, при этом $\mathrm{x}_1^{(i)}$ есть результат случайного равновероятного выбора с возвращением из совокупности $(\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Точно также независимо от $\mathrm{x}_1^{(i)}$ определяется $\mathrm{x}_2^{(i)}$ и т. д. Оценкой будстрэпа для θ, построенной по $X^{(i)}$, называется значение $\theta_n^*(X^{(i)})$. В качестве оценки значения $\sigma^2\bigl(F_n^*\bigr)$ принимается величина

$$\bigl(\sigma^2(F_n^*)\bigr)^* = \frac{1}{N}\sum_{i=1}^{N}\bigl(\theta_n^*(X^{(i)}) - \overline{\theta}^*\bigr)^2, \quad \overline{\theta}^* = \frac{1}{N}\sum_{i=1}^{N}\theta_n^*(X^{(i)}),$$

которая выбором N может быть сделана сколь угодно точной.

Вернемся к общей задаче оценивания и перейдем к рассмотрению важного *параметрического* случая, когда оцениваются неизвестные параметры распределений из параметрических семейств.

§ 18. Сравнение оценок в параметрическом случае. Эффективные оценки

В предыдущем параграфе мы выделили два подхода (среднеквадратический и асимптотический) к сравнению качества оценок. Введем теперь некоторые понятия, связанные с этими подходами в параметрическом случае, когда распределение выборки X принадлежит некоторому семейству $\mathcal{P} = \{\mathbf{P}_\theta\}$. Символами $\mathbf{E}_\theta$ и $\mathbf{D}_\theta$ обозначаем, как и прежде, математическое ожидание и дисперсию по распределению $\mathbf{P}_\theta$.

1. Одномерный случай. Среднеквадратический подход. Напомним, что в соответствии со среднеквадратическим подходом мы должны говорить, что θ_1^* лучше, чем θ_2^*, если

$$d_1^2(\theta) = \mathbf{E}_\theta(\theta_1^* - \theta)^2 < \mathbf{E}_\theta(\theta_2^* - \theta)^2 = d_2^2(\theta). \tag{1}$$

Но в параметрическом случае $d_l^2(\theta)$, $l = 1, 2$, суть функции от θ, и мы должны говорить «θ_1^* лучше θ_2^* *в точке* θ», если $d_1(\theta) < d_2(\theta)$.

Аналогичным образом дело обстоит при использовании асимптотического подхода, когда сравниваются а. н. оценки при больших объемах выборки n путем сравнения их предельных распределений. Оценка θ_1^* считается лучше, чем θ_2^*, в *точке* θ, если в соотношениях

$$(\theta_l^* - \theta) = \sqrt{n} \Subset \mathbf{\Phi}_{0,\sigma_l^2(\theta)}, \quad l = 1, 2, \tag{2}$$

справедливо $\sigma_1(\theta) < \sigma_2(\theta)$ *).

Таким образом, в обоих случаях задача сравнения оценок приводит к вопросу *о сравнении функций*, скажем, $d_l(\theta)$, $\theta \in \Theta$. Это множество неупорядоченное, и в классе всех оценок можно ввести частичный порядок следующим образом.

Правило 1. *Оценка* θ_1^* *лучше, чем* θ_2^*, если $d_1(\theta) \leqslant d_2(\theta)$ (или соответственно $\sigma_1(\theta) \leqslant \sigma^2(\theta)$) при всех $\theta \in \Theta$ и хотя бы для одного θ выполняется строгое неравенство $d_1(\theta) < d_2(\theta)$.

Если оценка θ^* такова, что для нее существует оценка θ_1^*, которая лучше, чем θ^*, то в таких случаях говорят, что θ^* — *недопустимая оценка.*

Остановимся сначала на среднеквадратическом подходе в одномерном случае и рассмотрим существующие тут возможности сравнивать оценки. Прежде всего следует отметить, что наилучшей оценки в смысле приведенного определения, вообще говоря, не существует. То есть не существует оценки θ^* такой, что для любой другой оценки θ_1^* справедливо неравенство $d(\theta) \leqslant d_1(\theta)$, где $d_1(\theta)$ определено в (1), $d(\theta)$ соответствует θ^*.

Действительно, если взять оценку $\theta_1^* = \theta_1 = \text{const} \in \Theta$, то $d_1^2(\theta) = \mathbf{E}_\theta(\theta_1^* - \theta)^2 = 0$ при $\theta = \theta_1$, и для наилучшей оценки θ^* (если бы такая существовала) будет выполняться $d^2(\theta_1) = \mathbf{E}_{\theta_1}(\theta^* - \theta_1)^2 = 0$. Поскольку θ_1 произвольно, то $d^2(\theta) \equiv 0$. Но это возможно лишь в «вырожденном» случае, когда наблюдения однозначно определяют значение параметра θ. Например, когда $X \in \mathbf{I}_\theta$ или $X \in \mathbf{U}_{\theta,\theta+1}$ и $\Theta = \{1, 2, \ldots\}$.

Таким образом, нижняя огибающая всех функций $d^2(\theta)$ равна нулю, но в «невырожденных» случаях ни для какой оценки θ^* эта функция не реализуется.

Задачу можно сделать более содержательной, если искать наилучшие оценки θ^* в тех или иных *подклассах* оценок, которые выбираются достаточно разумным образом. Одним из возможных способов выделения таких подклассов является фиксирование смещения $b(\theta)$.

*) Мы уже отмечали, что в широком классе случаев $d_l^2(\theta) = n^{-1}\sigma_l^2(\theta) + o(n^{-1})$. Однако из определений чисел $d_l^2(\theta)$ и $\sigma_l^2(\theta)$ это не следует.

Определение 1. Оценка $\theta_0^* \in K$ называется *эффективной в классе* K, если для любой другой оценки $\theta^* \in K$ $\mathbf{E}_\theta(\theta_0^* - \theta)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta)^2$ при всех $\theta \in \Theta$.

Особую роль играет класс K_0 несмещенных оценок — это класс оценок, для которых $b(\theta) \equiv 0$.

Эффективные оценки в классе $K_0 = \{\theta^* : \mathbf{E}_\theta\theta^* = \theta\}$ несмещенных оценок называются просто *эффективными*. Так что эффективные оценки — это несмещенные оценки с минимальной дисперсией.

Само по себе свойство несмещенности, как уже отмечалось, является несомненно желательным, так как оно означает отсутствие систематической погрешности при использовании оценки.

Вопрос о существовании оценок с данным смещением $b(\theta)$ (в частности, несмещенных оценок) сводится к разрешимости интегрального уравнения относительно $g(x)$

$$\int g(\boldsymbol{x})\mathbf{P}_\theta(X \in d\boldsymbol{x}) = \theta + b(\theta), \tag{3}$$

где $g(X) = \theta^*$; левая часть этого уравнения представляет собой $\mathbf{E}_\theta\theta^*$.

Если выполнено условие (A_μ) и $f_\theta(\boldsymbol{x}) = \prod_{i=1}^{n} f_\theta(x_i)$ есть функция правдоподобия, то уравнение принимает вид

$$\int g(\boldsymbol{x})f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \theta + b(\theta). \tag{4}$$

Следует отметить, что решение (4) для заданной $b(\theta)$ существует далеко не всегда, и, в частности, *не для всех семейств* $\{\mathbf{P}_\theta\}$ *существуют несмещенные оценки параметра* θ. Рассмотрим, например, схему Бернулли с неизвестным параметром p (вероятностью исхода $\{\mathrm{x}_1 = 1\}$) и допустим, что нам дано оценить параметр $\theta = \varphi(p)$, где φ — заданная функция. Тогда уравнение (4) для несмещенной оценки будет иметь вид

$$\sum_{\boldsymbol{x}} g(\boldsymbol{x})f_\theta(\boldsymbol{x}) = \theta$$

или, что то же,

$$\sum_{k=0}^{n} G(k)p^k(1-p)^{n-k} = \varphi(p), \tag{5}$$

где $G(k) = \sum_{\boldsymbol{x}\in A_k} g(\boldsymbol{x})$, A_k есть множество точек x, у которых k координат равны единице. Но левая часть (5) есть полином от p степени n. Это означает, что уравнение (5) может быть разрешимо, если только $\varphi(p)$ есть полином степени не выше n.

Рассмотрим теперь класс K_b оценок с фиксированным смещением $b(\theta)$ и допустим, что существует эффективная в K_b оценка.

Теорема 1. *Эффективная в K_b оценка единственна с точностью до значений на множестве $A \subset \mathcal{X}^n$, для которого $\mathbf{P}_\theta(A) = 0$ при всех $\theta \in \Theta$.*

Доказательство. Пусть θ_0^*, θ_1^* — две эффективные в K_b оценки. Обозначим

$$D = \mathbf{D}_\theta \theta_l^*, \quad \Delta_l = \theta_l^* - \theta, \quad \theta^* = \frac{\theta_0^* + \theta_1^*}{2}, \quad l = 0, 1.$$

Так как

$$\left(\frac{\Delta_0 + \Delta_1}{2}\right)^2 + \left(\frac{\Delta_0 - \Delta_1}{2}\right)^2 = \frac{\Delta_0^2 + \Delta_1^2}{2},$$
$$\frac{\Delta_0 + \Delta_1}{2} = \theta^* - \theta, \quad \Delta_0 - \Delta_1 = \theta_0^* - \theta_1^*, \tag{6}$$

то

$$\mathbf{E}_\theta(\theta^* - \theta)^2 + \frac{1}{4}\mathbf{E}_\theta\left(\theta_0^* - \theta_1^*\right)^2 = D + b^2(\theta). \tag{7}$$

Но $\theta^* \in K_b$, и, стало быть, $\mathbf{E}_\theta(\theta^* - \theta)^2 \geqslant D + b^2(\theta)$. Из (7) вытекает тогда, что

$$\mathbf{E}_\theta\left(\theta_0^* - \theta_1^*\right)^2 \leqslant 0,$$

$\theta_1^* = \theta_0^*$ п. в. *) $\triangleleft$

Проведенное обсуждение проблемы сравнения оценок относилось к среднеквадратическому подходу. К нему же относится по существу и следующее

Определение 2. Оценка $\theta_1^* \in K$ называется *асимптотически эффективной* (а. э.) в классе K, если при $n \to \infty$ для любой другой оценки θ^* из K и каждого $\theta \in \Theta$

$$\limsup_{n\to\infty} \frac{\mathbf{E}_\theta\left(\theta_1^* - \theta\right)^2}{\mathbf{E}_\theta(\theta^* - \theta)^2} \leqslant 1. \tag{8}$$

2. Асимптотический подход. Асимптотическая эффективность в классах M- и L-оценок. Мы перейдем теперь к *асимптотическому подходу*, с которым определение 2 тесно связано. Задача здесь, как и прежде, состоит в сравнении функций $\sigma(\theta)$, характеризующих предельное нормальное распределение, но проблема в целом несколько упрощается. Дело прежде всего в том, что сравнение происходит лишь в классе а. н. оценок, который мы в дальнейшем будем обозначать через K_Φ.

---

*) Справедливо следующее утверждение, обобщающее в известном смысле теорему 1. *Если θ_0^* — эффективная в K_b, а θ^* — произвольная в K_b оценки, так что $h = \mathbf{D}_\theta\theta_0^*/\mathbf{D}_\theta\theta^* \leqslant 1$, то коэффициент корреляции $\rho(\theta_1^*, \theta^*)$ между оценками θ_0^* и θ^* равен $\rho(\theta_0^*, \theta^*) = \sqrt{h}$.*

Доказательство читатель может провести самостоятельно, убедившись, что при $\rho(\theta_0^*, \theta^*) \neq \sqrt{h}$ и соответствующем выборе α оценка $\theta_1^* = (1-\alpha)\theta_0^* + \alpha\theta^* \in K_b$ будет удовлетворять неравенству $\mathbf{D}_\theta\theta_1^* < \mathbf{D}_\theta\theta_0^*$, противоречащему эффективности θ_0^*.

Пусть K — некоторый класс оценок из K_{Φ}. Временно вместо (8) мы будем использовать несколько иное определение асимптотически эффективной оценки.

Определение 3. Оценка $\theta_1^* \in K$ называется *асимптотически эффективной в* K, если для любой другой оценки $\theta^* \in K$

$$\sigma_1^2(\theta) \leqslant \sigma^2(\theta)$$

при всех $\theta \in \Theta$, где $\sigma^2(\theta)$ и $\sigma_1^2(\theta)$ — коэффициенты рассеивания θ^* и θ_1^* соответственно.

Позже мы покажем, что при широких предположениях определения 2 и 3 эквивалентны.

Если в качестве K взять класс M-оценок K_M, построенных по функциям g, удовлетворяющим условия теоремы 14.4 (об асимптотической нормальности), то из теоремы 16.3 немедленно вытекает следующее утверждение.

Следствие 1. *При выполнении условий регулярности, приведенных в теореме* 16.3, *о. м. п.* $\widehat{\theta}^*$ *является а. э. оценкой в* K_M *с коэффициентом рассеивания*

$$\sigma^2(\theta) = I^{-1}(\theta); \quad I(\theta) = \int \big(l'(\mathrm{x}_1, \theta)\big)^2 f_\theta(x)\mu(dx), \tag{9}$$

где $f_\theta(x)$ — *плотность наблюдений* x_i.

Класс K_M является весьма широким (в него входят, например, все оценки, являющиеся статистиками I типа; достаточно положить $g(x, \theta) = g(x) - h^{-1}(\theta)$, где g и h — функции из определения статистик I типа в § 3). Однако остается неясным, содержит ли K_M все а. н. оценки, т. е. все оценки из класса K_{Φ}, в котором нам хотелось бы построить а. э. оценки. Чтобы двигаться в этом направлении, мы произведем несколько позже иное сужение класса K_{Φ}, по-видимому, менее существенное.

Обратимся теперь к L-оценкам применительно к задачам оценивания параметров (изначально определение L-статистик с параметрическими семействами не связаны).

Использование результатов теоремы 17.1 в параметрическом случае будет эффективно, если функция $\ddot{l}\big(F_\theta^{-1}(t), \theta\big)$ (см. (17.12)) от θ не зависит. По-видимому, это возможно лишь для параметров сдвига и масштаба.

Рассмотрим сначала задачу об оценке параметра сдвига $\theta = \alpha$ для параметрического семейства $\mathbf{P}_\alpha(A) = \mathbf{P}(A - \alpha)$ $(A - \alpha = \{x : x + \alpha \in A\}$, $X = \mathbb{R}$, $\Theta = \mathbb{R})$. В этом случае $F_\alpha(x) = F(x - \alpha)$, $f_\alpha(x) = f(x - \alpha)$, где функции F и $f = F'$ известны, $F_\alpha^{-1}(t) = \alpha + F^{-1}(t)$; $f_\alpha\big(F_\alpha^{-1}(t)\big) = f(F^{-1}(t))$ от α не зависит.

Заметим, что

$$G(F_\alpha) \equiv \int_0^1 \varphi(t) F_\alpha^{-1}(t)\, dt = \int_0^1 \varphi(t)(\alpha + F^{-1}(t))\, dt = \alpha\varphi_0 + G(F_0) = \alpha + G(F)$$

при любой функции φ, удовлетворяющей условию

$$\varphi_0 = \int \varphi(t)\,dt = 1. \tag{10}$$

Поэтому оценка подстановки $\alpha^* = G(F_n^*) - G(F)$ является с точностью до постоянной $G(F)$ (зависящей от φ!) L-статистикой, зависящей от φ, и естественно поставить вопрос об оптимальном выборе φ. Оценки, являющиеся с точностью до постоянных слагаемых L-статистиками, также будем называть L-оценками. Пусть f имеет вторую производную $\ddot{f}$ и φ удовлетворяет (10). Обозначим через K_L класс L-оценок вида

$$\alpha^* = G(F_n^*) - G(F), \quad \alpha^* = \int \varphi(F_n^*(u))u\,dF_n^*(u) - G(F).$$

Тогда согласно следствию 8.6 и замечанию 8.3 $K_L \subset K_{\Phi}$.

Несомненное преимущество L-оценок перед M-оценками состоит в том, что первые выписываются в явном виде. Из теоремы 17.1 вытекает

Следствие 2. *Пусть* $l(x) = \ln f(x)$,

$$0 < I = \int (\dot{l}(x))^2 f(x)\,dx < \infty. \tag{11}$$

Тогда L-оценка $\alpha^* \in K_L$ при $\varphi(x) = I^{-1}\ddot{l}(F^{-1}(x))$ является а. э. в классе K_L и имеет коэффициент рассеивания I^{-1}.

Очевидно, что для параметра сдвига $\theta = \alpha$ значение $I(\alpha)$ в (9) от α не зависит и совпадает с I. Это означает, что оценка

$$\alpha^* = I^{-1}\int \ddot{l}(F^{-1}(F_n^*(u)))u\,dF_n^*(u) - G(F) =$$

$$= (nI)^{-1}\sum \ddot{l}\left(F^{-1}\left(\frac{i}{n}\right)\right)\mathrm{x}_{(i)} - G(F) \tag{12}$$

(мы считаем здесь в соответствии с договоренностью в § 3, что $F_n^*(u)$ непрерывна справа, так что $F_n^*(\mathrm{x}_{(i)}) = i/n$) асимптотически эквивалентна о. м. п. (имеет то же предельное распределение) и в *этом смысле* оценку (12) можно рассматривать также как M-оценку, соответствующую решению уравнения $\sum \dot{l}(\mathrm{x}_l - \alpha) = 0$: она будет обладать всеми свойствами асимптотической оптимальности, каким обладает о. м. п. $\widehat{\alpha}^*$.

Замечание 1. Так как $\ddot{l}(u) = \mathrm{const}$ лишь для нормального распределения, то представление (12) дает следующую характеризацию этого распределения: а. э. оценка параметра сдвига (или о. м. п.) реализуется как статистика вида $\frac{1}{n}\sum \varphi\left(\frac{i}{n}\right)\mathrm{x}_i$ (статистика I типа) лишь для нормального закона $F = \Phi$.

Аналогично обстоит дело с оценкой параметра масштаба $\theta = \sigma$ для параметрического семейства $\mathbf{P}_\sigma(A) = \mathbf{P}\left(\dfrac{A}{\sigma}\right)$.

В этом случае

$$F_\sigma(x) = F\left(\frac{x}{\sigma}\right), \quad f_\sigma(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right), \tag{13}$$

где функции F и $f = F'$ известны, $F_\sigma^{-1}(t) = \sigma F^{-1}(t)$. Имеем

$$G(F_\sigma) \equiv \int_0^1 \varphi(t) F_\sigma^{-1}(t)\, dt = \sigma \int_0^1 \varphi(t) F_1^{-1}(t)\, dt = \sigma G(F), \tag{14}$$

так что при любой φ такой, что $G(F) = G(F, \varphi) \neq 0$

$$\sigma = \frac{G(F_\sigma)}{G(F)}. \tag{15}$$

Это делает естественным рассматривать в качестве оценки σ^* параметра σ значение

$$\sigma^* = \frac{G(F_n^*)}{G(F)},$$

являющееся L-оценкой. Для этой оценки функция φ в «исходном» представлении L-оценок (как $\int \varphi(t) F_n^{-1}(t)\, dt$) удовлетворяет условию

$$\int \varphi(t) F^{-1}(t)\, dt = 1. \tag{16}$$

(Это равенство дает возможность аналогично (17.10) получать с помощью неравенства Коши–Буняковского другие нижние границы для разброса L-оценок.)

Следствие 3. *Если выполнено* (11) *и*

$$I_1 = \int \ddot{l}(u) u f(u)\, du \neq 0, \tag{17}$$

то L-оценка

$$\sigma^* = \int \ddot{l}(t) F_n^{-1}(t)\, dt = \frac{1}{n} \sum \ddot{l}\left(\frac{i}{n}\right) \mathrm{x}_{(i)},$$

соответствующая функции $\varphi(t) = \ddot{l}(F^{-1}(t))$, является а. э. в классе L-оценок, удовлетворяющих (16), *и имеет коэффициент рассеивания $\sigma^2 I_1^{-2} I$.*

Доказательство. Воспользуемся теоремой 17.1, в которой роль F, I, $f(F^{-1}(t))$, $\ddot{l}(F^{-1}(t))$ булут играть соответственно F_σ, $\sigma^{-2} I$, $\dfrac{1}{\sigma} f(F^{-1}(t))$,

$\frac{1}{\sigma^2}\ddot{l}(F^{-1}(t))$, где F и f из (13) от σ не зависят. Согласно этой теореме в качестве оптимальной φ надо взять $\varphi(t) = \varphi_0 I^{-1}\ddot{l}(F^{-1}(t))$, где φ_0 определяется из (16):

$$\varphi_0 I^{-1}\int \ddot{l}(F^{-1}(t))F^{-1}(t)\,dt = 1; \quad \varphi_0 = I_1^{-1}I.$$

Следствие доказано. $\triangleleft$

Для весьма распространенного класса *симметричных распределений* F ($f(-x) = f(x)$, $\ddot{l}(-x) = \ddot{l}(x)$) условие (17) не выполнено. В этом случае можно перейти к наблюдениям $\mathrm{y}_i = |\mathrm{x}_i|$ (в § 22 увидим, что $\{|\mathrm{x}_i|\}$ образуют здесь достаточную статистику и потери информации при переходе от x_i к $|\mathrm{x}_i|$ не происходит). После этого можно перейти к наблюдениям $\mathrm{z}_i = \ln \mathrm{y}_i$, для которых

$$\mathbf{P}(\mathrm{z}_i < t) = \mathbf{P}(|\mathrm{x}_i| < e^t) = 2F_\sigma(e^t) - 1 = 2F(e^{t-\ln\sigma}) - 1 = \widetilde{F}(t-a),$$

где $\alpha = \ln\sigma$, $\widetilde{F}(u) = 2F(e^u) - 1$. Мы свели задачу об оценке параметра масштаба σ к оценке параметра сдвига для параметрического семейства с плотностью $\widetilde{f}(u-\alpha)$, $\widetilde{f}(u) = 2f(e^u)e^u$, $u \geqslant 0$. Остается воспользоваться следствием 2. При этом в новой задаче при очевидном соглашении относительно обозначений

$$\widetilde{l}(u) = l(e^u) + \ln 2 + u, \quad \dot{\widetilde{l}}(u) = \dot{l}(e^u)e^u + 1, \quad \ddot{\widetilde{l}} = \ddot{l}(e^u)e^{2u} + \dot{l}(e^u)e^u,$$

так что

$$\widetilde{I} = -2\int_0^\infty f(t)\,dt[\ddot{l}(t)t^2 + \dot{l}(t)t]. \tag{18}$$

Итак, из следствия 2 вытекает

Следствие 4. *Если распределение F симметрично, то оценка*

$$\alpha^* = (n\widetilde{I})^{-1}\sum \ddot{\widetilde{l}}\left(\widetilde{F}^{-1}\left(\frac{i}{n}\right)\right)\mathrm{z}_{(i)} - G(\widetilde{F}),$$
$$G(\widetilde{F}) = \widetilde{I}^{-1}\int \ddot{\widetilde{l}}(\widetilde{F}^{-1}(u))\widetilde{F}^{-1}(u)\,du = \widetilde{I}^{-1}\int \ddot{\widetilde{l}}(u)u\widetilde{f}(u)\,du, \tag{19}$$

будет а. э. в классе оценок L и будет иметь коэффициент разброса $\widetilde{I}^{-1}$. Оценка $\sigma^* = e^{\alpha^*}$ будет а. н. с коэффициентом разброса $\sigma^2\widetilde{I}^{-1}$.

Для стандартного нормального распределения $\dot{l}(u) = -u$, $\ddot{l}(u) = -1$, $\widetilde{I} = 2$ (см. (18), а также § 26, 27) и оценка $\sigma^* = e^{\alpha^*}$ будет иметь разброс $\sigma^2/2$. Такой же разброс будет иметь о. м. п.

$$\widehat{\sigma}^* = \left[\frac{1}{n}\sum \mathrm{x}_i^2\right]^{1/2}. \tag{20}$$

Это вытекает из следствия 1 или непосредственно из соотношений

$$n\sigma^{-2}(\widehat{\sigma}^*)^2 \in \mathbf{H}_n, \quad \xi_n = \frac{n\sigma^{-2}(\widehat{\sigma}^*)^2 - n}{\sqrt{2n}} \Rightarrow \mathbf{\Phi}_{0,1}$$

(см. 12.4), так что

$$\widehat{\sigma}^* = \sigma\left(1 + \xi_n\sqrt{\frac{2}{n}}\right)^{1/2} = \sigma\left(1 + \frac{1}{\sqrt{2n}}\xi_n + \varepsilon_n\right),$$

где $\varepsilon_n \underset{P}{\longrightarrow} 0$. Таким образом, для нормальных распределений оценка $\sigma^* = e^{\alpha^*}$ оказалась асимптотически эквивалентной а. э. оценке $\widehat{\sigma}^*$ в классе K_M.

Отметим еще раз, что несомненным преимуществом оценки $\sigma^* = e^{\alpha^*}$ является наличие ее явного вида.

Сравним оценки (19), (20) с естественной несмещенной L-оценкой параметра σ вида

$$\sigma^* = \frac{1}{na}\sum |\mathrm{x}_i|, \tag{21}$$

где $a = 2\int\limits_0^\infty x\,dF(x)$, так что $\mathbf{E}|\mathrm{x}_i| = a\sigma$ ($X \in F_\sigma$, распределение F симметрично). Если обозначить $d^2 = \int\limits_{-\infty}^{\infty} x^2\,dF(x) < \infty$, то мы получим, что σ^* является а. н. оценкой с коэффициентом разброса

$$\mathbf{E}\left(\frac{|\mathrm{x}_i|}{a} - \sigma\right)^2 = \sigma^2(d^2a^{-2} - 1) \equiv D^2\sigma^2,$$

$$(\sigma^* - \sigma)\sqrt{n} = \frac{1}{\sqrt{n}}\sum\left(\frac{|\mathrm{x}_i|}{a} - \sigma\right) \Rightarrow \mathbf{\Phi}_{0,D^2\sigma^2}.$$

Для нормального закона $a^2 = \dfrac{2}{\pi}$, $d^2 = 1$, $D^2 = \sigma^2\left(\dfrac{\pi}{2} - 1\right)$. Так как $\dfrac{\pi}{2} - 1 > \dfrac{1}{2}$, то коэффициент разброса для оценки (21) больше, чем для оценок (19), (20).

Вернемся теперь к общим определениям 2, 3 асимптотической эффективности и введем в рассмотрение несколько иное сужение $K_{\mathbf{\Phi}}$, упоминавшееся выше, которое, по существу, обеднять $K_{\mathbf{\Phi}}$ не будет. Именно, будем рассматривать класс $K_{\mathbf{\Phi},2} \subset K_{\mathbf{\Phi}}$ а. н. оценок θ^*, обладающих тем свойством, что для них сходимость

$$(\theta^* - \theta)\sqrt{n} \Rightarrow \mathbf{\Phi}_{0,\sigma^2(\theta)}$$

происходит вместе с двумя первыми моментами:

$$\mathbf{E}_\theta(\theta^* - \theta)\sqrt{n} \to 0, \quad \mathbf{E}_\theta(\theta^* - \theta)^2 n \to \sigma^2(\theta). \tag{22}$$

Отметим, что первое из этих двух соотношений нетрудно получить из второго с помощью теоремы непрерывности для моментов (см. § 5).

Сужение K_Φ до класса $K_{\Phi,2}$ мало обедняет первый из этих классов по двум причинам. Во-первых, а. н. оценки, в которых (22) нарушено, практически не встречаются (мы отмечали, что для этого обычно нужны искусственные построения). Во-вторых, для $\theta^* \in K_\Phi$ по лемме Фату

$$\liminf_{n\to\infty} \mathbf{E}_\theta n(\theta^* - \theta)^2 \geqslant \sigma^2(\theta)$$

(мы имеем дело с интегралами от неотрицательных функций), так что $\mathbf{E}_\theta n(\theta^* - \theta)^2$ при больших n может отличаться от $\sigma^2(\theta)$ лишь в сторону бóльших значений. Но оценки с таким свойством вряд ли могут конкурировать с оценками, для которых (22) выполнено.

Таким образом, при асимптотическом подходе мы можем в качестве класса а. н. оценок, в котором производится сравнение, рассматривать класс $K_{\Phi,2}$. Он будет для нас удобнее.

Если K — класс оценок такой, что $K \subset K_{\Phi,2}$, то определение 3 *асимптотической эффективности эквивалентно определению* 2.

Эквивалентность определений вытекает из того, что для $\theta^* \in K_{\Phi,2}$

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \frac{\sigma^2(\theta)}{n}(1 + r_n(\theta)), \quad r_n(\theta) \to 0 \quad \text{при} \quad n \to \infty.$$

В этом случае соотношение (8), означающее, что

$$\mathbf{E}_\theta\big(\theta_1^* - \theta\big)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta)^2\big(1 + r_n'(\theta)\big), \quad r_n'(\theta) \to 0,$$

для любой $\theta^* \in K$, очевидно, эквивалентно неравенству в определении 3. $\triangleleft$

Упомянутое ранее некоторое упрощение проблемы сравнения при асимптотическом подходе состоит в том, что здесь мы сравниваем лишь дисперсии предельных законов. Роль смещения $b(\theta)$ оценок здесь исчезает, поскольку в классе $K_{\Phi,2}$ в силу (22) выполняется соотношение $b(\theta) = o(1/\sqrt{n})$, означающее «почти несмещенность» оценок или «асимптотическую пренебрегаемость» смещения с точки зрения соотношений (2).

Аналогично теореме 1 может быть получена

Теорема 2. *Пусть $K \subset K_{\Phi,2}$. Тогда если θ_1^* и θ_2^* — две а. э. оценки в K такие, что $\frac{1}{2}\big(\theta_1^* + \theta_2^*\big) \in K$, то они асимптотически совпадают, т. е.*

$$\sqrt{n}\,\big(\theta_1^* - \theta_2^*\big) \underset{P}{\longrightarrow} 0, \quad \mathbf{E}_\theta\big[\sqrt{n}\,\big(\theta_1^* - \theta_2^*\big)\big]^2 \to 0.$$

Доказательство. Нам достаточно установить второе соотношение, так как первое из него следует. Пусть

$$M_{l,n} = \mathbf{E}_\theta n\big(\theta_l^* - \theta\big)^2, \quad \Delta_l = \theta_l^* - \theta, \quad \theta^* = \frac{\theta_l^* + \theta_2^*}{2}, \quad l = 1, 2.$$

Тогда на основании (6) получаем

$$\mathbf{E}_\theta n(\theta^* - \theta)^2 + \frac{1}{4}\mathbf{E}_\theta n\big(\theta_1^* - \theta_2^*\big)^2 = \frac{M_{1,n} + M_{2,n}}{2}. \tag{23}$$

Но $\theta^* \in K$ и, стало быть, после перехода к пределу в последнем равенстве получим в силу асимптотической эффективности θ_l^*

$$\lim_{n\to\infty} \mathbf{E}_\theta n\big(\theta_1^* - \theta_2^*\big)^2 \leqslant 0. \quad \triangleleft$$

Приведенные выше рассмотрения содержали лишь один из возможных путей выделения оценок (в нашем случае — эффективных оценок), которые по ряду естественных соображений следует предпочесть другим. Однако возможны, конечно, и иные подходы (напомним, что нам приходится сравнивать неупорядоченные элементы — функции $d(\theta)$ или $\sigma(\theta)$). Поскольку оценок с наименьшим при каждом θ возможным значением $d(\theta)$, вообще говоря, не существует, можно сравнивать, скажем, средние значения $\int d(t)q(t)\,dt$, где $q(t) \geqslant 0$, $\int q(t)\,dt = 1$, или максимальные значения $\max\limits_{\theta\in\Theta} d(\theta)$. Это есть способы упорядочивания множества *всех* оценок.

Первый из этих двух способов мы назовем позже *байесовским*, второй — *минимаксным*. Байесовские и мимимаксные оптимальные оценки будут рассмотрены нами в § 21, эффективные оценки — в более поздних параграфах.

Более подробно проблема выбора оценок будет рассматриваться в главе 6.

3. Многомерный случай. Рассмотрим теперь случай, когда θ и θ^* есть векторы из $\mathbb{R}^k$. Здесь задача сравнения оценок становится труднее. Дело в том, что в многомерном случае нам пришлось ввести *частичный* порядок уже для сравнения оценок при *фиксированном* θ. Для сравнения оценок *на всем множестве* Θ, как и в одномерном случае, также приходится вводить частичный порядок, но уже в «другом направлении» (ведь в основе сравнения лежит среднеквадратичное уклонение, которое есть функция двух переменных: θ и вектора a, на который проектируется уклонение $\theta^* - \theta$).

Наилучшие оценки в «обоих направлениях» и составляют предмет следующих определений.

Определение 4. Оценка θ_0^* является *эффективной в классе* K, если для любой оценки θ^* из K среднеквадратичное рассеивание θ^* около θ при всех $\theta \in \Theta$ не меньше, чем рассеивание θ_0^*.

Это определение эквивалентно следующему.

Векторная оценка θ_0^* параметра θ эффективна в K, если для любого вектора a оценка $\alpha_0^* = (\theta_0^*, a)$ является эффективной оценкой скалярного параметра $\alpha = (\theta, a)$ в классе оценок $\alpha^* = (\theta^*, a)$, $\theta^* \in K$, т. е. если при всех $\theta \in \Theta$, $a \in \mathbb{R}^k$, $\theta^* \in K$

$$\mathbf{E}_\theta(\theta_0^* - \theta, a)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta, a)^2. \tag{24}$$

Как мы уже видели, это неравенство эквивалентным образом записывается в виде $d_0^2(\theta) \leqslant d^2(\theta)$ или

$$\sum_{i,j} d_{ij}^{(0)}(\theta) a_i a_j \leqslant \sum_{i,j} d_{ij} a_i a_j$$

при всех $\theta \in \Theta$, $a \in \mathbb{R}^k$, где $d^2(\theta) = \|d_{ij}(\theta)\|$ и $d_0^2(\theta) = \left\|d_{ij}^{(0)}(\theta)\right\|$ — матрицы вторых моментов $\theta^* - \theta$ и $\theta_0^* - \theta$ соответственно.

Эффективные оценки в классе K_0 несмещенных оценок называются просто *эффективными.*

Поскольку определение (24) эффективности строится на использовании одномерного случая, из теоремы 1 нетрудно установить, что эффективная оценка в классе K_b оценок с фиксированным смещением $b(\theta) = \mathbf{E}\theta^* - \theta$ единственна.

Определение а. э. оценок в многомерном случае аналогично определениям 2, 3.

Определение 5. Векторная оценка θ_1^* параметра θ *асимптотически эффективна в* K, если для любого вектора a оценка $\left(\theta_1^*, a\right)$ является а. э. оценкой скалярного параметра $\alpha = (\theta, a)$ в классе оценок $\alpha^* = (\theta^*, a) \in K$.

Другими словами (см. § 17), среднеквадратичное рассеивание предельного распределения $\left(\theta_1^* - \theta\right)\sqrt{n}$ для а. э. оценки является наименьшим. Это, в свою очередь, означает что для любых $\theta^* \in K$, $a \in \mathbb{R}^k$, $\theta \in \Theta$ выполняется $\sigma_1^2(\theta) \leqslant \sigma^2(\theta)$, или

$$\sum_{i,j} \sigma_{ij}^{(1)}(\theta) a_i a_j \leqslant \sum_{i,j} \sigma_{ij}(\theta) a_i a_j,$$

где $\sigma^2(\theta) = \|\sigma_{ij}(\theta)\|$, $\sigma_1^2(\theta) = \left\|\sigma_{ij}^{(1)}(\theta)\right\|$ суть соответственно матрицы вторых моментов предельных распределений $(\theta^* - \theta)\sqrt{n}$ и $\left(\theta_1^* - \theta\right)\sqrt{n}$.

Из рассмотрений предыдущего параграфа можно заключить, что множество оценок в многомерном случае при фиксированном θ можно сделать упорядоченным, если качество оценки измерять количеством (при среднеквадратическом подходе)

$$v(\theta^*) = \mathbf{E}_\theta (\theta^* - \theta) V (\theta^* - \theta)^\top \equiv v(\theta^*, \theta), \tag{25}$$

где V — неотрицательно определенная матрица. Аналогичное количество, связанное с матрицей вторых моментов предельного нормального распределения, можно рассматривать и при асимптотическом подходе в классе $K_{\Phi,2}$.

Двигаясь дальше по этому пути, можно сделать вполне упорядоченным множество всех оценок и на всем множестве Θ. Именно, можно проводить сравнение средних

$$\int v(\theta^*, t) q(t)\, dt, \quad q(t) \geqslant 0, \quad \int q(t)\, dt = 1,$$

или максимальных $\max\limits_{t \in \Theta} v(\theta^*, t)$ значений количеств $v(\theta^*, \theta)$, определенных в (25).

Если окажется, что оценка, наилучшая при таком подходе, будет оставаться наилучшей при любой неотрицательно определенной матрице V, это будет означать в силу леммы 17.1, что эта оценка будет наилучшей и в смысле частичного порядка, установленного в § 17 (т. е. усредненное среднеквадратичное уклонение будет наименьшим в любом направлении).

Чтобы строить оценки, оптимальные в смысле определений, рассмотренных в этом параграфе, нам понадобятся понятия и свойства условных математических ожиданий и достаточных статистик.

§ 19. Условные математические ожидания

В этом параграфе напомним определение и основные свойства условных математических ожиданий (у. м. о). Более полное изложение см. в приложении III, а также в [17, 35, 46, 72, 137].

1. Определение у. м. о. Пусть ξ и η есть две случайные величины, заданные на вероятностном пространстве $(\Omega, \mathfrak{F}, \mathbf{P})$.

Условное математическое ожидание $\mathbf{E}(\xi/B)$ случайной величины ξ относительно события B, $\mathbf{P}(B) > 0$, определяется равенством

$$\mathbf{E}(\xi/B) = \frac{\mathbf{E}(\xi; B)}{\mathbf{P}(B)}, \tag{1}$$

где $\mathbf{E}(\xi; B) = \int\limits_B \xi \, d\mathbf{P} = \mathbf{E}(\xi I_B)$, $I_B = I_B(\omega)$ есть случайная величина, равная индикатору множества B.

Допустим, что ξ и η независимы, $B = \{\eta = x\}$ и $\mathbf{P}(B) > 0$. Тогда для любой измеримой функции $\varphi(x, y)$ согласно (1)

$$\mathbf{E}[\varphi(\xi, \eta)/\eta = x] = \frac{\mathbf{E}\varphi(\xi, \eta) I_{\{\eta = x\}}}{\mathbf{P}(\eta = x)} = \frac{\mathbf{E}\varphi(\xi, x) I_{\{\eta = x\}}}{\mathbf{P}(\eta = x)} = \mathbf{E}\varphi(\xi, x). \tag{2}$$

Последнее равенство справедливо, так как случайные величины $\varphi(\xi, x)$ и $I_{\{\eta = x\}}$ как функции соответственно от ξ и η независимы и, стало быть,

$$\mathbf{E}\varphi(\xi, x) I_{\{\eta = x\}} = \mathbf{E}\varphi(\xi, x) \mathbf{E} I_{\{\eta = x\}} = \mathbf{E}\varphi(\xi, x) \mathbf{P}(\eta = x).$$

Соотношения (2) показывают, что понятие у. м. о. может сохранять свой смысл и в случае, когда вероятность условия равна нулю — ведь само по себе равенство

$$\mathbf{E}[\varphi(\xi, \eta)/\eta = x] = \mathbf{E}\varphi(\xi, x)$$

для независимых ξ и η представляется естественным и с предположением $\mathbf{P}(\eta = x) > 0$ никак не связано.

Пусть $\mathfrak{A}$ есть σ-подалгебра $\mathfrak{F}$. Определим теперь понятие у. м. о. случайной величины ξ относительно $\mathfrak{A}$, которое будем обозначать через $\mathbf{E}(\xi/\mathfrak{A})$. Определение дадим сначала для «дискретного» случая, но так, чтобы оно легко обобщалось.

«Дискретным» мы называем случай, когда σ-алгебра $\mathfrak{A}$ образована (порождена) не более чем счетной последовательностью непересекающихся событий $A_1, A_2, \ldots$; $\cup A_i = \Omega$, $\mathbf{P}(A_i) > 0$. Этот факт мы записываем в виде $\mathfrak{A} = \sigma(A_1, A_2, \ldots)$; это означает, что элементами $\mathfrak{A}$ являются всевозможные объединения множеств $A_1, A_2, \ldots$

С помощью случайной величины ξ и системы событий $(A_1, A_2, \ldots)$ построим новую случайную величину $\widehat{\xi} = \widehat{\xi}(\omega)$ следующим образом:

$$\widehat{\xi} = y_k \equiv \mathbf{E}(\xi/A_k) = \frac{\mathbf{E}(\xi; A_k)}{\mathbf{P}(A_k)} \quad \text{при} \quad \omega \in A_k,\ k = 1, 2, \ldots$$

Другими словами,

$$\widehat{\xi} = \sum_k \frac{\mathbf{E}(\xi; A_k)}{\mathbf{P}(A_k)} I_{A_k},$$

где I_A есть индикатор множества A.

Определение 1. Случайная величина $\widehat{\xi}$ называется *у. м. о. ξ относительно σ-алгебры* $\mathfrak{A}$ и обозначается $\mathbf{E}(\xi/\mathfrak{A})$.

Таким образом, в отличие от обычных математических ожиданий у. м. о $\mathbf{E}(\xi/\mathfrak{A})$ есть *случайная величина*. В нашем случае она постоянна на множествах A_k и равна на этих множествах усреднению ξ по A_k. Если ξ и $\mathfrak{A}$ независимы (т. е. $\mathbf{P}(\xi \in B; A_k) = \mathbf{P}(\xi \in B)\mathbf{P}(A_k)$), то, очевидно, $\mathbf{E}(\xi; A_k) = \mathbf{E}\xi\mathbf{P}(A_k)$ и $\widehat{\xi} = \mathbf{E}\xi$.

Если же $\mathfrak{A} = \mathfrak{F}$, то $\mathfrak{F}$ тоже «дискретна», ξ постоянна на множествах A_k и, стало быть, $\widehat{\xi} = \xi$.

Отметим следующие два основных свойства у. м. о.:

1) *$\widehat{\xi}$ измерима относительно* $\mathfrak{A}$;

2) *для любого события* $A \in \mathfrak{A}$

$$\mathbf{E}(\widehat{\xi}; A) = \mathbf{E}(\xi; A).$$

Первое свойство очевидно. Второе следует из того, что любое событие $A \in \mathfrak{A}$ представимо в виде $A = \bigcup_k A_{j_k}$ и, стало быть,

$$\mathbf{E}(\widehat{\xi}; A) = \sum_k \mathbf{E}(\widehat{\xi}; A_{j_k}) = \sum_k y_{j_k}\mathbf{P}(A_{j_k}) = \sum_k \mathbf{E}(\xi; A_{j_k}) = \mathbf{E}(\xi; A).$$

Это свойство довольно ясно: усреднение по множеству A величины ξ дает тот же результат, что усреднение уже усредненной по A_{j_k} величины $\widehat{\xi}$.

Лемма 1. *Свойства* 1), 2) *однозначно определяют у. м. о. и эквивалентны определению* 1.

Доказательство. В одну сторону утверждение леммы уже доказано. Пусть теперь выполнены условия 1), 2). Измеримость $\widehat{\xi}$ относительно $\mathfrak{A}$ означает, что $\widehat{\xi}$ постоянна на множествах A_k. Обозначим значение $\widehat{\xi}$ на A_k через y_k. Так как $A_k \in \mathfrak{A}$, то из свойства 2) следует, что

$$\mathbf{E}(\widehat{\xi}; A_k) = y_k\mathbf{P}(A_k) = \mathbf{E}(\xi; A_k)$$

и, стало быть, при $\omega \in A_k$

$$\widehat{\xi} = y_k = \frac{\mathbf{E}(\xi; A_k)}{\mathbf{P}(A_k)}. \quad \triangleleft$$

Теперь мы может дать общее определение у. м. о.

Определение 2. Пусть ξ есть случайная величина на вероятностном пространстве $(\Omega, \mathfrak{F}, \mathbf{P})$ и $\mathfrak{A} \subset \mathfrak{F}$ есть σ-подалгебра $\mathfrak{F}$. *Условным математическим ожиданием ξ относительно* $\mathfrak{A}$ называется случайная величина $\widehat{\xi}$, обозначаемая $\mathbf{E}(\xi/\mathfrak{A})$, которая обладает следующими свойствами:

1) *$\widehat{\xi}$ измерима относительно* $\mathfrak{A}$;

2) *для всякого $A \in \mathfrak{A}$ справедливо* $\mathbf{E}(\widehat{\xi}; A) = \mathbf{E}(\xi; A)$.

В этом определении случайная величина ξ может быть как скалярной, так и векторной.

Сразу возникают вопросы: существует ли такая величина $\widehat{\xi}$ и единственна ли она? В «дискретном» случае мы видели, что ответ на эти вопросы положителен. В общем случае справедлива

Теорема 1. *Если $\mathbf{E}|\xi|$ конечно, то функция $\widehat{\xi} = \mathbf{E}(\xi/\mathfrak{A})$ в определении 2 всегда существует и единственна с точностью до значений на множестве вероятности нуль.*

Доказательство. Предположим сначала, что ξ скалярна, $\xi \geqslant 0$. Тогда функция множества

$$\mathbf{Q}(A) = \int_A \xi \, d\mathbf{P} = \mathbf{E}(\xi; A), \quad A \in \mathfrak{A},$$

будет мерой на $(\Omega, \mathfrak{A})$, которая абсолютно непрерывна относительно $\mathbf{P}$, поскольку $\mathbf{P}(A) = 0$ влечет за собой $\mathbf{Q}(A) = 0$. Следовательно, по теореме Радона–Никодима (см. [17, приложение 3]). существует $\mathfrak{A}$-измеримая функция $\widehat{\xi} = \mathbf{E}(\xi/\mathfrak{A})$, единственная с точностью до значений на множестве меры нуль такая, что

$$\mathbf{Q}(A) = \int_A \widehat{\xi} \, d\mathbf{P}.$$

В общем случае положим $\xi = \xi^+ - \xi^-$, $\xi^+ = \max(0, \xi) \geqslant 0$, $\xi^- = \max(0, -\xi) \geqslant 0$,

$$\widehat{\xi} = \widehat{\xi}^+ - \widehat{\xi}^-,$$

где $\widehat{\xi}^{\pm}$ — у. м. о. для $\xi^{\pm}$. Этим доказано существование у. м. о., так как $\widehat{\xi}$ будет удовлетворять условиям 1), 2) определения 2. Отсюда же следует и единственность, так как предположение о неединственности $\widehat{\xi}$ будет означать неединственность $\widehat{\xi}^+$ или $\widehat{\xi}^-$. Доказательство для векторных ξ сводится к одномерному случаю, так как свойствами 1), 2) будут обладать координаты $\widehat{\xi}$, существование и единственность которых нами доказана. $\triangleleft$

Существо проведенного доказательства достаточно прозрачно: ведь по условию 2) для всякого $A \in \mathfrak{A}$ задано $\mathbf{E}(\widehat{\xi}; A) = \int\limits_A \widehat{\xi}\, d\mathbf{P}$, т. е. заданы значения интегралов от $\widehat{\xi}$ по всем множествам $A \in \mathfrak{A}$. Ясно, что это должно определять $\mathfrak{A}$-измеримую функцию $\widehat{\xi}$ однозначно с точностью до значений на множестве меры нуль.

Смысл $\mathbf{E}(\xi/\mathfrak{A})$ остается прежним — грубо говоря, это есть усреднение ξ по «неделимым» элементам $\mathfrak{A}$.

Если $\mathfrak{A} = \mathfrak{F}$, то, очевидно, $\widehat{\xi} = \xi$ удовлетворяет свойствам 1), 2) и, стало быть, $\mathbf{E}(\xi/\mathfrak{F}) = \xi$.

Определение 3. Если $\xi = I_C$ есть индикатор множества $C \in \mathfrak{F}$, то $\mathbf{E}(I_C/\mathfrak{A})$ будет называться *условной вероятностью* $\mathbf{P}(C/\mathfrak{A})$ события C относительно $\mathfrak{A}$. Функция множеств $\mathbf{P}(\xi \in C/\mathfrak{A})$ называется *условным распределением* ξ *относительно* $\mathfrak{A}$.

Рассмотрим теперь важный специальный случай, когда σ-алгебра $\mathfrak{A}$ порождена некоторой случайной величиной η.

Определение 4. Пусть ξ и η — случайные величины на $(\Omega, \mathfrak{F}, \mathbf{P})$, $\mathfrak{A} = \sigma(\eta)$ есть σ-алгебра, порожденная случайной величиной η. Тогда $\mathbf{E}(\xi/\mathfrak{A})$ называется также *условным математическим ожиданием величины* ξ *относительно* η.

Иногда ради упрощения записи будем вместо $\mathbf{E}(\xi/\sigma(\eta))$ писать $\mathbf{E}(\xi/\eta)$. Это не приводит к недоразумениям.

Так как $\mathbf{E}(\xi/\eta)$ по определению есть $\sigma(\eta)$-измеримая случайная величина, это означает (см. [17, § 3.4]), что существует измеримая функция $g(x)$, для которой

$$\mathbf{E}(\xi/\eta) = g(\eta). \tag{3}$$

По аналогии с дискретным случаем мы можем истолковывать здесь величину $g(x)$ как результат усреднения ξ по множеству $\{\eta = x\}$. (Напомним, что в дискретном случае $g(x) = \mathbf{E}(\xi/\eta = x)$.)

Аналогично определению 3 в случае $\mathfrak{A} = \sigma(\eta)$ будем называть функцию множеств $\mathbf{P}(\xi \in C/\eta)$ *условным распределением* ξ *относительно* η (см. также § 20).

2. Свойства у. м. о.

1. *У. м. о. обладает свойствами обычных математических ожиданий* (см. [17, § 4.9]) с той лишь разницей, что они выполняются почти наверное (с вероятностью 1):

1a) $\mathbf{E}(c\xi/\mathfrak{A}) = c\mathbf{E}(\xi/\mathfrak{A})$, *если* $c = \text{const}$,

1b) $\mathbf{E}(\xi_1 + \xi_2/\mathfrak{A}) = \mathbf{E}(\xi_1/\mathfrak{A}) + \mathbf{E}(\xi_2/\mathfrak{A})$,

1c) *если* $\xi_1 \leqslant \xi_2$ *п. н., то* $\mathbf{E}(\xi_1/\mathfrak{A}) \leqslant \mathbf{E}(\xi_2/\mathfrak{A})$.

2. *Справедливо неравенство типа Чебышева*: если ξ вещественна, $\xi \geqslant 0$, то для любого $x > 0$

$$\mathbf{P}(\xi \geqslant x/\mathfrak{A}) \leqslant \frac{\mathbf{E}(\xi/\mathfrak{A})}{x}.$$

Это соотношение между у. м. о., как и равенства п. 1, выполняется п. н. Такое же соглашение будет справедливо в дальнейшем для всех соотношений между у. м. о.

3. *Если* σ*-алгебры* $\mathfrak{A}$ *и* $\sigma(\xi)$ *независимы, то* $\mathbf{E}(\xi/\mathfrak{A}) = \mathbf{E}\xi$.

Отсюда следует, в частности, что если ξ и η независимы, то $\mathbf{E}(\xi/\eta) = \mathbf{E}\xi$. Если σ-алгебра $\mathfrak{A}$ тривиальна, очевидно, также получим $\mathbf{E}(\xi/\mathfrak{A}) = \mathbf{E}\xi$.

4. *Для у. м. о. верны теоремы сходимости, справедливые для обычных математических ожиданий*, например теорема о монотонной сходимости: если $\xi_n \uparrow \xi$, $\xi_n \geqslant 0$, то $\mathbf{E}(\xi_n/\mathfrak{A}) \uparrow \mathbf{E}(\xi/\mathfrak{A})$ п. н.

5. *Если* η *скалярна и измерима относительно* $\mathfrak{A}$, $\mathbf{E}|\xi| < \infty$, $\mathbf{E}|\xi_n| < \infty$, *то*

$$\mathbf{E}(\eta\xi/\mathfrak{A}) = \eta\mathbf{E}(\xi/\mathfrak{A}).$$

Другими словами, $\mathfrak{A}$-измеримые случайные величины ведут себя относительно операции у. м. о. как постоянные (ср. со свойством 1а)).

6. *Для у. м. о. остаются справедливыми все основные неравенства для обычных математических ожиданий*, в частности неравенство Коши–Буняковского

$$\mathbf{E}(|\xi_1\xi_2|/\mathfrak{A}) \leqslant \left[\mathbf{E}\left(\xi_1^2/\mathfrak{A}\right)\mathbf{E}\left(\xi_2^2/\mathfrak{A}\right)\right]^{1/2}$$

и неравенство Иенсена: если $\mathbf{E}|\xi| < \infty$, то для любой выпуклой вниз функции $g(x)$

$$g(\mathbf{E}(\xi/\mathfrak{A})) \leqslant \mathbf{E}(g(\xi)/\mathfrak{A}).$$

7. *Формула полной вероятности* (свойство 2 определения 2 при $A = \Omega$)

$$\mathbf{E}\xi = \mathbf{E}\mathbf{E}(\xi/\mathfrak{A}).$$

8. *Последовательное усреднение* (обобщение свойства 7): если $\mathfrak{A} \subset \mathfrak{A}_1 \subset \mathfrak{F}$, то

$$\mathbf{E}(\xi/\mathfrak{A}) = \mathbf{E}(\mathbf{E}(\xi/\mathfrak{A}_1)/\mathfrak{A}).$$

Доказательство этих свойств можно найти в приложении III.

Очевидно, что свойства 1, 3–5, 7, 8 справедливы как для скалярных, так и для векторных случайных величин ξ.

Отметим, что если рассматривать у. м. о. $\widehat{\xi} = \mathbf{E}(\xi/\eta)$ относительно случайной величины η, то их, пользуясь свойствами 5, 7, иногда удобно определять эквивалентным (по отношению к определению 2) образом так:

Определение 2А. Случайная величина $\widehat{\xi}$ является у. м. о. относительно η, если она измерима относительно $\sigma(\eta)$, и для любой измеримой функции m

$$\mathbf{E}(\widehat{\xi}m(\eta)) = \mathbf{E}(\xi m(\eta)). \tag{4}$$

Действительно, если $\widehat{\xi} = \mathbf{E}(\xi/\eta)$, то согласно свойствам 5, 7

$$\mathbf{E}(\xi m(\eta)) = \mathbf{E}\mathbf{E}(\xi m(\eta)/\eta) = \mathbf{E}m(\eta)\mathbf{E}(\xi/\eta) = \mathbf{E}(\widehat{\xi}m(\eta)).$$

Обратно, если выполнено (4), то, полагая $m(y) = I_C$, получим

$$\mathbf{E}(\widehat{\xi}; \eta \in C) = \mathbf{E}(\widehat{\xi m}(\eta)) = \mathbf{E}(\xi; \eta \in C),$$

что означает выполнение условий определения 2.

Вернемся к свойствам у. м. о. в общем случае. Следующее свойство проясняет «геометрическую» сущность у. м. о.

9. Известно, что функция $\varphi(a) = \mathbf{E}(\xi - a)^2$ достигает своего минимума при $a = \mathbf{E}\xi$ (см., например, [17, § 4.5]). Аналогичное свойство справедливо и для у. м. о.: *при* $a(\omega) = \mathbf{E}(\xi/\mathfrak{A})$ *достигается минимум* $\mathbf{E}(\xi - a(\omega))^2$ *среди всех* $\mathfrak{A}$*-измеримых функций* $a(\omega)$.

Действительно, $\mathbf{E}(\xi - a(\omega))^2 = \mathbf{E}\mathbf{E}((\xi - (\omega))^2/\mathfrak{A})$, но $a(\omega)$ ведет себя как постоянная относительно операции $\mathbf{E}(\cdot/\mathfrak{A})$ (см. свойство 5), так что

$$\mathbf{E}((\xi - a(\omega))^2/\mathfrak{A}) = \mathbf{E}((\xi - \mathbf{E}(\xi/\mathfrak{A}))^2/\mathfrak{A}) + \mathbf{E}((\mathbf{E}(\xi/\mathfrak{A}) - a(\omega))^2/\mathfrak{A}),$$

и минимум этого выражения достигается при $a(\omega) = \mathbf{E}(\xi/\mathfrak{A})$. Это свойство можно рассматривать как определение у. м. о., эквивалентное определению 2. Согласно нему $\mathbf{E}(\xi/\mathfrak{A})$ можно трактовать как «проекцию» ξ на $\mathfrak{A}$.

Свойство 9 допускает обобщение на многомерный случай, когда $\xi = (\xi_1, \ldots, \xi_s)$ есть случайный вектор в $\mathbb{R}^s$.

9A. *Пусть* $V = \|v_{ij}\|$ — *произвольная неотрицательно определенная матрица размера* $s \times s$, $a \in \mathbb{R}^s$,

$$\zeta(a) = (\xi - a)V(\xi - a)^\top$$

(в частности, при $V = E$ получим $\zeta(a) = |\xi - a|^2$). *Тогда на функции* $a(\omega) = \mathbf{E}(\xi/\mathfrak{A})$ *достигается минимум* $\min\limits_{a \in A} \mathbf{E}\zeta(a)$ *по классу* A *всех* $\mathfrak{A}$*-измеримых функций.*

Доказательство этого факта протекает так же, как и в одномерном случае. Положим $\alpha = \mathbf{E}(\xi/\mathfrak{A})$. Тогда $\mathbf{E}\zeta(a) = \mathbf{E}\mathbf{E}(\zeta(a)/\mathfrak{A})$,

$$\begin{aligned}\mathbf{E}(\zeta(a)/\mathfrak{A}) &= \mathbf{E}((\xi - a)V(\xi - a)^\top/\mathfrak{A}) = \\ &= \mathbf{E}((\xi - \alpha)V(\xi - \alpha)^\top/\mathfrak{A}) + \mathbf{E}((\alpha - a)V(\xi - \alpha)^\top/\mathfrak{A}) + \\ &\quad + \mathbf{E}((\xi - \alpha)V(\alpha - a)^\top/\mathfrak{A}) + \mathbf{E}((\alpha - a)V(\alpha - a)^\top/\mathfrak{A}). \quad (5)\end{aligned}$$

Так как $\alpha - a$ есть $\mathfrak{A}$-измеримый вектор, то по свойству 5

$$\mathbf{E}((\alpha - a)V(\xi - \alpha)^\top/\mathfrak{A}) = (\alpha - a)V\mathbf{E}((\xi - \alpha)^\top/\mathfrak{A}) = 0,$$
$$\mathbf{E}((\xi - \alpha)V(\alpha - a)^\top/\mathfrak{A}) = [\mathbf{E}((\xi - \alpha)/\mathfrak{A})]V(\alpha - a)^\top = 0.$$

Поскольку последнее слагаемое в (4) неотрицательно и равно нулю при $a = \alpha$, утверждение доказано. $\triangleleft$

§ 20. Условные распределения

Наряду с у. м. о. можно рассматривать условные распределения относительно σ-подалгебр и относительно случайных величин. В этом параграфе остановимся лишь на последних.

Пусть ξ и η — две случайные величины на $(\Omega, \mathfrak{F}, \mathbf{P})$ со значениями соответственно в $\mathbb{R}^s$ и $\mathbb{R}^k$ и $\mathfrak{B}^s$ — σ-алгебра борелевских множеств из $\mathbb{R}^s$.

Определение 1. Функция $P(B/y)$ двух переменных $y \in \mathbb{R}^k$, $B \in \mathfrak{B}^s$ называется *условным распределением ξ при условии $\eta = y$*, если:

1) при каждом $B \in \mathfrak{B}^s$ $P(B/\eta)$ есть условная вероятность (условное распределение) $\mathbf{P}(\xi \in B/\eta)$ события $\{\xi \in B\}$ относительно η, т. е. $P(B/y)$ есть борелевская функция от y такая, что для любого $A \in \mathfrak{B}^k$

$$\mathbf{E}(P(B/\eta); \eta \in A) \equiv \int\limits_A P(B/y)\mathbf{P}(\eta \in dy) = \mathbf{P}(\xi \in B, \eta \in A);$$

2) при каждом y $P(B/y)$ есть распределение вероятностей по B.

Функцию $P(B/y)$ мы будем иногда записывать в более «расшифрованной форме» в виде

$$\mathbf{P}(B/y) = \mathbf{P}(\xi \in B/\eta = y).$$

Мы знаем, что для каждого $B \in \mathfrak{B}^s$ существует борелевская функция $g_B(y)$ такая, что $g_B(\eta) = \mathbf{P}(\xi \in B/\eta)$. Таким образом, положив $P(B/y) = g_B(y)$, мы удовлетворим условию 1) определения 1. Однако условие 2) при этом из свойств у. м. о. никак не следует и выполняться совсем не обязано — ведь условная вероятность $\mathbf{P}(\xi \in B/\eta)$ определена при каждом B с точностью до значений на множестве N_B меры нуль (так что существует много вариантов у. м. о.), и это множество может быть своим для каждого B. Поэтому если объединение $N = \bigcup\limits_{B \in \mathfrak{B}^s} N_B$ имеет не нулевую вероятность, то может оказаться, что, например, равенства

$$\mathbf{P}(\xi \in B_1 \cup B_2/\eta) = \mathbf{P}(\xi \in B_1/\eta) + \mathbf{P}(\xi \in B_2/\eta)$$

(аддитивность вероятности) сразу для всех непересекающихся B_1, B_2 из $\mathfrak{B}^s$ не выполняются ни для одного ω из N, т. е. на ω-множестве N положительной вероятности функция $g_B(y)$ не будет распределением как функция B.

Однако в нашем случае, *когда ξ есть случайная величина со значениями в $\mathbb{R}^s$ с σ-алгеброй борелевских множеств $\mathfrak{B}^s$, $g_B(\eta) = \mathbf{P}(\xi \in B/\eta)$ всегда можно выбрать таким образом, что $g_B(y)$ будет условным распределением* (см. [35, 46]).

Как и следовало ожидать, условные распределения обладают тем естественным свойством, что у. м. о. выражаются в виде интегралов по условным распределениям.

Теорема 1. *Для любой измеримой функции $g(x)$, отображающей $\mathbb{R}^s$ в $\mathbb{R}$, такой, что $\mathbf{E}|g(\xi)| < \infty$, справедливо равенство*

$$\mathbf{E}(g(\xi)/\eta) = \int g(x)P(dx/\eta). \tag{1}$$

Доказательство. Достаточно рассмотреть случай, когда $g(x) \geqslant 0$. Если $g(x) = I_A(x)$ есть индикатор множества A, то формула (1), очевидно, верна. Стало быть, она верна для любой простой функции $g_n(x)$ (т. е. для функции, принимающей конечное число значений). Остается взять последовательность $g_n \uparrow g$ и воспользоваться монотонностью обеих частей в (1) и свойством 4 в § 19. $\triangleleft$

В реальных задачах для подсчета условных распределений часто можно пользоваться следующим простым правилом, которое для наглядности мы запишем в форме

$$\mathbf{P}(\xi \in B/\eta = y) = \frac{\mathbf{P}(\xi \in B.\ \eta \in dy)}{\mathbf{P}(\eta \in dy)}. \tag{2}$$

Очевидно, что формально оба условия определения 1 будут удовлетворены.

Если ξ и η имеют плотности распределения, этому равенству будет придан точный смысл.

Определение 2. Пусть условное распределение $P(B/y)$ при каждом y абсолютно непрерывно относительно некоторой меры μ в $\mathbb{R}^s$

$$\mathbf{P}(\xi \in B/\eta = y) = \int\limits_B f(x/y)\mu(dx).$$

Тогда плотность $f(x/y)$ называется *условной плотностью* ξ (*относительно меры* μ) *при условии* $\eta = y$.

Другими словами, измеримая по паре переменных x, y функция $f(x/y)$ есть условная плотность ξ при условии $\eta = y$, если

1) для любых борелевских множеств $A \subset \mathbb{R}^k$, $B \subset \mathbb{R}^s$

$$\int\limits_{y\in A}\int\limits_{x\in B} f(x/y)\mu(dx)\mathbf{P}(\eta \in dy) = \mathbf{P}(\xi \in B,\ \eta \in A); \tag{3}$$

2) при каждом y функция $f(x/y)$ есть плотность распределения вероятностей.

Из теоремы 1 вытекает, что если существует условная плотность, то

$$\mathbf{E}(g(\xi)/\eta) = \int g(x)f(x/\eta)\mu(dx).$$

Если предположить дополнительно, что распределение η имеет плотность $q(y)$ относительно некоторой меры λ в $\mathbb{R}^k$, то (3) можно записать в виде

$$\int\limits_{y\in A}\int\limits_{x\in B} f(x/y)q(y)\mu(dx)\lambda(dy) = \mathbf{P}(\xi \in B,\ \eta \in A). \tag{4}$$

Рассмотрим теперь прямое произведение пространств $\mathbb{R}^s$ и $\mathbb{R}^k$ и на нем прямое произведение мер $\mu \times \lambda$ (если $C = B \times A$, $B \subset \mathbb{R}^s$, $A \subset \mathbb{R}^k$, то $\mu \times \lambda(C) = \mu(B)\lambda(A)$). В этом пространстве соотношение (4) означает, очевидно, что совместное распределение ξ и η в $\mathbb{R}^s \times \mathbb{R}^k$ имеет плотность относительно $\mu \times \lambda$, равную

$$f(x, y) = f(x/y)q(y).$$

Справедливо и обратное утверждение.

Теорема 2. *Если совместное распределение ξ и η в $\mathbb{R}^s \times \mathbb{R}^k$ имеет плотность $f(x, y)$ относительно $\mu \times \lambda$, то функция*

$$f(x/y) = \frac{f(x, y)}{q(y)}, \quad \text{где} \quad q(y) = \int f(x, y)\mu(dx),$$

есть условная плотность ξ при условии $\eta = y$, а функция $q(y)$ есть плотность η относительно меры λ.

Доказательство. Утверждение теоремы относительно $q(y)$ очевидно, так как $\int\limits_A q(y)\lambda(dy) = \mathbf{P}(\eta \in A)$. Остается заметить, что $f(x/y) = f(x, y)/q(y)$ удовлетворяет всем условиям в определении 2 условной плотности (равенство (4), эквивалентное (3), выполнено очевидным образом). $\triangleleft$

Замечание 1. Случайные величины ξ и η в теореме 2 можно поменять местами. Тогда получим, что наряду с $f(x/y)$ существует условная плотность

$$q(y/x) = \frac{f(x, y)}{f(x)}, \quad f(x) = \int f(x, y)\lambda(dy),$$

случайной величины η при условии $\xi = x$. Это простое следствие теоремы 2 будет играть существенную роль в последующем изложении. Применительно к задачам статистики оно позволит нам в следующем параграфе получить формулу Байеса, которая будет затем часто использоваться на протяжении всего курса.

Пример 1. Пусть Φ_{α,σ^2} есть двумерное нормальное распределение величин ξ_1 и ξ_2, где $\alpha = (\alpha_1, \alpha_2)$, $\alpha_i = \mathbf{E}\xi_i$, $\sigma^2 = \|\sigma_{ij}\|$, $\sigma_{ij} = \mathbf{E}(\xi_i - \alpha_i)(\xi_j - \alpha_j)$, $i, j = 1, 2$. Определитель матрицы вторых моментов равен

$$|\sigma^2| = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - \rho^2),$$

где ρ есть коэффициент корреляции между ξ_1 и ξ_2. Таким образом, если $|\rho| \neq 1$, то матрица вторых моментов не вырождена и для нее существует обратная матрица

$$A = (\sigma^2)^{-1} = \frac{1}{|\sigma^2|}\begin{Vmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{Vmatrix} = \frac{1}{1 - \rho^2}\begin{Vmatrix} \dfrac{1}{\sigma_{11}} & -\dfrac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}} \\ -\dfrac{\rho}{\sqrt{\sigma_{11}\sigma_{22}}} & \dfrac{1}{\sigma_{22}} \end{Vmatrix}.$$

Стало быть, совместная плотность ξ_1 и ξ_2 (относительно меры Лебега) равна (см. § 12)

$$f(x,y) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \times$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\alpha_1)^2}{\sigma_{11}} - \frac{2\rho(x-\alpha_1)(y-\alpha_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(y-\alpha_2)^2}{\sigma_{22}}\right]\right\}.$$

Одномерные плотности ξ_1 и ξ_2 соответственно равны

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_{11}}}e^{-(x-\alpha_1)^2/2\sigma_{11}}, \quad q(y) = \frac{1}{\sqrt{2\pi\sigma_{22}}}e^{-(y-\alpha_2)^2/2\sigma_{22}}.$$

Поэтому условная плотность ξ_1 при условии $\xi_2 = y$ равна

$$f(x/y) = \frac{f(x,y)}{q(y)} = \frac{1}{\sqrt{2\pi\sigma_{11}(1-\rho^2)}} \times$$
$$\times \exp\left\{-\frac{1}{2\sigma_{11}(1-\rho^2)}\left(x - \alpha_1 - \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}(y-\alpha_2)\right)^2\right\};$$

это есть плотность нормального распределения со средним $\alpha_1 + \rho\sqrt{\sigma_{11}/\sigma_{22}} \times$ $\times (y-\alpha_2)$ и дисперсией $\sigma_{11}(1-\rho^2)$. Отсюда следует, в частности, что у. м. о. ξ_1 относительно ξ_2 равно

$$\mathbf{E}(\xi_1/\xi_2) = \alpha_1 + \rho\sqrt{\frac{\sigma_{11}}{\sigma_{22}}}(\xi_2 - \alpha_2).$$

Прямую $x = \alpha_1 + \rho\sqrt{\sigma_{11}/\sigma_{22}}\,(y-\alpha_2)$ называют линией регрессии ξ_1 на ξ_2. Она дает наилучшие среднеквадратичные приближения величины ξ_1 при данной $\xi_2 = y$.

Пример 2. Рассмотрим задачу о вычислении плотности случайной величины $\xi = \varphi(\zeta,\eta)$, где ζ и η независимы. Из формулы (3) при $A = \mathbb{R}^k$ вытекает, что плотность $f(x)$ распределения ξ выражается через условную плотность $f(x/y)$ равенством

$$f(x) = \int f(x/y)\mathbf{P}(\eta \in dy). \tag{5}$$

Применительно к рассматриваемой задаче под $f(x/y)$ надо понимать плотность случайной величины $\varphi(\zeta,y)$, поскольку $\mathbf{P}(\xi \in B/\eta = y) = \mathbf{P}(\varphi(\zeta,y) \in B)$.

Формула (5) бывает очень полезной при вычислении распределений различных статистик. Например, в п. 12.6 мы могли бы формулу (12.7) для для плотности распределения Фишера выписывать сразу, а не выводить ее из вида функции распределения.

§ 21. Байесовский и минимаксный подходы к оцениванию параметров

Суть байесовского подхода состоит в том, что неизвестный параметр θ рассматривается как *случайная величина* с некоторой (известной или неизвестной) плотностью распределения $q(t)$, $t \in \Theta$, относительно меры λ, которая, как и мера μ в условии (A_μ), чаще всего будет представлять из себя либо меру Лебега, либо считающую меру. Плотность $q(t)$ называется *априорной*, т. е. данной *до эксперимента*. Байесовский подход предполагает, что неизвестный параметр θ выбран случайным образом из распределения с плотностью $q(t)$.

Пусть, далее, $f_t(x)$, $t \in \Theta$, $\boldsymbol{x} \in \mathcal{X}^n$, есть функция правдоподобия, введенная нами в § 16. Как мы отмечали, $f_t(\boldsymbol{x})$ при каждом t есть плотность распределения в $\mathcal{X}^n$. Поэтому функция

$$f(\boldsymbol{x}, t) = f_t(\boldsymbol{x}) q(t)$$

есть плотность некоторого распределения в $\mathcal{X}^n \times \Theta$ относительно меры $\mu^n \times \lambda$, которую можно интерпретировать как *плотность совместного распределения* X и θ. При таком подходе в силу теоремы 20.2 функция $f_t(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}^n$, есть условная плотность X при условии $\theta = t$

$$f_t(\boldsymbol{x}) = f(\boldsymbol{x}/t), \quad \mathbf{E}_\theta g(X) = \mathbf{E}(g(X)/\theta).$$

В этих рассмотрениях формальная сторона дела требует, чтобы $f_t(\boldsymbol{x})$ была измеримой по t и $\boldsymbol{x}$ функцией. В дальнейшем везде, где это потребуется, будем предполагать, что это свойство имеет место.

В последующем параметр, как случайную величину, везде будем обозначать через θ, в то время как для фиксированных значений параметра часто будем использовать обозначения t, u и др., так что

$$\mathbf{E}_t g(X) = \mathbf{E}(g(X)/\theta = t).$$

Наряду с $f(\boldsymbol{x}/t)$ мы можем выписать *условную плотность* $q(t/\boldsymbol{x})$ *величины* θ *при условии* $X = \boldsymbol{x}$:

$$q(t/\boldsymbol{x}) = \frac{f_t(\boldsymbol{x}) q(t)}{f(\boldsymbol{x})}, \quad f(\boldsymbol{x}) = \int f_t(\boldsymbol{x}) q(t) \lambda(dt). \tag{1}$$

Эта плотность определяет так называемое *апостериорное* (т. е. *после эксперимента*) распределение θ, которое мы будем обозначать через $\mathbf{Q}_{\boldsymbol{x}}$. Равенство (1) называется *формулой Байеса* для плотности апостериорного распределения. В дальнейшем эта формула будет играть существенную роль.

Свойство 9 у. м. о. применительно к байесовскому случаю означает следующее: среди всех функций $\theta^* = \varphi(X)$ наилучшей оценкой для θ (в смысле минимизации $\mathbf{E}(\theta - \varphi(X))^2$) является функция

$$\theta^*_Q = \mathbf{E}(\theta/X) = \int t q(t/X) \lambda(dt) = \int t \mathbf{Q}_X(dt). \tag{2}$$

Определение 1. Оценка θ^*_Q, определенная формулами (2), (1), называется *байесовской, соответствующей априорному распределению* $\mathbf{Q}$ *с плотностью* $q(t)$.

Отметим еще раз, что для байесовской оценки безусловное среднеквадратичное уклонение

$$\mathbf{E}(\theta^* - \theta)^2 = \mathbf{E}\mathbf{E}((\theta^* - \theta)^2/\theta) = \mathbf{E}\mathbf{E}_\theta(\theta^* - \theta)^2 = \int \mathbf{E}_t(\theta^* - t)^2 q(t)\lambda(dt) \quad (3)$$

принимает наименьшее возможное значение. Соотношение (3) показывает, что байесовская оценка минимизирует среднее значение (с данной весовой функцией $q(t)\lambda(dt)$) величины $\mathbf{E}_t(\theta^* - t)^2$.

Другими словами, если θ выбирается случайно, с плотностью $q(t)$, то *байесовская оценка является наилучшей в смысле среднеквадратического подхода.* Среднеквадратичное уклонение (3) байесовской оценки можно представить в виде (см. (1))

$$\mathbf{E}\big(\theta_Q^* - \theta\big)^2 = \int \mathbf{E}_t\big(\theta_Q^* - t\big)^2 q(t)\lambda(dt) =$$

$$= \iint \big(t - \theta_Q^*\big)^2 f_t(\boldsymbol{x}) q(t)\lambda(dt)\mu^n(d\boldsymbol{x}) = \int \sigma^2_{Q_{\boldsymbol{x}}} f(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \mathbf{E}\sigma^2_{Q_X},$$

где $\sigma^2_{Q_X}$ есть дисперсия апостериорного распределения $\mathbf{Q}_X$:

$$\sigma^2_{Q_X} = \int \big(t - \theta_Q^*\big)^2 q(t/X)\lambda(dt) = \int (t - \mathbf{E}(\theta/X))^2 \mathbf{Q}_X(dt). \quad (4)$$

Это был *байесовский подход* к отысканию наилучших оценок.

Другой подход к сравнению оценок, который отмечался нами в § 18, основан на сравнении $\sup\limits_{t\in\Gamma} \mathbf{E}_t(\theta^* - t)^2$, где $\Gamma \subset \Theta$ — заданное подмножество Θ (Γ либо совпадает с Θ, либо равно той его части, относительно которой удалось установить, что $\theta \in \Gamma$).

Определение 2. Оценка $\overline{\theta}^*$ называется *минимаксной*, если для любой другой оценки θ^*

$$\sup_{t\in\Gamma} \mathbf{E}_t(\overline{\theta}^* - t)^2 \leqslant \sup_{t\in\Gamma} \mathbf{E}_t(\theta^* - t)^2.$$

Другими словами, для минимаксной оценки достигается

$$\inf_{\theta^*} \sup_{t\in\Gamma} E_t(\theta^* - t)^2 = \sup_{t\in\Gamma} \mathbf{E}_t(\overline{\theta}^* - t)^2. \quad (5)$$

Установим некоторые полезные связи между байесовскими и минимаксными оценками.

Теорема 1. *Обозначим через θ_Q^* байесовскую оценку для априорного распределения $\mathbf{Q}$ с плотностью q. Если существуют оценка θ_1^* и распределение $\mathbf{Q}$ такие, что при всех t*

$$\mathbf{E}_t\big(\theta_1^* - t\big)^2 \leqslant \int \mathbf{E}_u\big(\theta_Q^* - u\big)^2 q(u)\lambda(du), \quad (6)$$

то оценка θ_1^* минимаксна.

Доказательство. Пусть θ^* — любая другая оценка. Тогда

$$\sup_t \mathbf{E}_t(\theta^* - t)^2 \geqslant \int \mathbf{E}_t(\theta^* - t)^2 q(t)\lambda(dt) \geqslant$$

$$\geqslant \int \mathbf{E}_t\big(\theta^*_Q - t\big)^2 q(t)\lambda(dt) \geqslant \mathbf{E}_t\big(\theta^*_1 - t\big)^2. \quad \triangleleft$$

Заметим, что почти для всех t, принадлежащих носителю $N_Q = \{t : q(t) > 0\}$ распределения $\mathbf{Q}$, в неравенстве (6) с необходимостью должно выполняться равенство, так как в противном случае мы получили бы

$$\int \mathbf{E}_t\big(\theta^*_1 - t\big)^2 q(t)\lambda(dt) < \int \mathbf{E}_t\big(\theta^*_Q - t\big)^2 q(t)\lambda(dt),$$

что противоречит определению байесовской оценки.

Это замечание позволяет нам сформулировать следующий критерий минимаксной оценки, эквивалентный теореме 1.

Теорема 2. *Если оценка* θ^*

1) *является байесовской для некоторого распределения* $\mathbf{Q}$,

2) $\mathbf{E}_t(\theta^* - t)^2 = c = \text{const}$ *при* $t \in N_Q$,

3) $\mathbf{E}_t(\theta^* - t)^2 \leqslant c$ *для остальных* t,

то θ^* *является минимаксной оценкой.*

Если $\theta^* = \theta^*_Q = \overline{\theta}^*$ удовлетворяет этому критерию, то, очевидно,

$$\sup_t \mathbf{E}_t(\overline{\theta}^* - t)^2 = \int \mathbf{E}_t(\overline{\theta}^* - t)^2 q(t)\lambda(dt). \tag{7}$$

Таким образом, минимаксная оценка — это есть байесовская оценка, которая «выравнивает» погрешности $\mathbf{E}_t(\overline{\theta}^* - t)^2$ при различных t. Это значит, что априорное распределение $\overline{\mathbf{Q}}$, соответствующее этой оценке, заставляет быть одинаково внимательным ко всем возможным значениям θ, а не ориентироваться, как это делают байесовские оценки θ^*_Q, соответствующие другим априорным распределениям $\mathbf{Q} \neq \overline{\mathbf{Q}}$, на какие-то выделенные (более вероятные) значения θ. Поскольку в последнем случае мы используем дополнительную информацию о θ, то естественно, что при $\mathbf{Q} \neq \overline{\mathbf{Q}}$ оценки θ^*_Q будут обладать меньшими значениями безусловных среднеквадратичных уклонений

$$\int \mathbf{E}_t\big(\theta^*_Q - t\big)^2 \mathbf{Q}(dt) \leqslant \int \mathbf{E}_t\big(\theta^*_{\overline{Q}} - t\big)^2 \overline{\mathbf{Q}}(dt).$$

В связи с этим распределение $\overline{\mathbf{Q}}$ в теореме 2, соответствующее минимаксной оценке $\overline{\theta}^*$, часто называют *наихудшим.*

Так как такое наихудшее распределение $\overline{\mathbf{Q}}$ не всегда существует (это бывает обычно в случаях, когда Θ — неограниченное множество), то можно предложить следующий модифицированный критерий для отыскания минимаксной оценки.

Теорема 3. *Если существует оценка* θ_1^* *и последовательность распределений* $\mathbf{Q}^{(k)}$ *с плотностями* $q^{(k)}$ *такие, что при всех* t

$$\mathbf{E}_t\left(\theta_1^* - t\right)^2 \leqslant \limsup_{k\to\infty} \int \mathbf{E}_t\left(\theta_{Q^{(k)}}^* - t\right)^2 q^{(k)}(t)\lambda(dt),$$

то оценка θ_1^* *минимаксная.*

Доказательство этой теоремы столь же просто. Для любой оценки θ^* справедливо

$$\sup_t \mathbf{E}_t(\theta^* - t)^2 \geqslant \int \mathbf{E}_t(\theta^* - t)^2 q^{(k)}(t)\lambda(dt) \geqslant \int \mathbf{E}_t\left(\theta_{Q^{(k)}}^* - t\right)^2 q^{(k)}(t)\lambda(dt).$$

Отсюда следует, что

$$\sup_t \mathbf{E}_t(\theta^* - t)^2 \geqslant \limsup_{k\to\infty} \int \mathbf{E}_t\left(\theta_{Q^{(k)}}^* - t\right)^2 q^{(k)}(t)\lambda(dt) \geqslant \mathbf{E}_t\left(\theta_1^* - t\right)^2. \quad \triangleleft$$

Пример 1. Пусть $X \in \mathbf{\Phi}_{\alpha,1}$. Выясним, что из себя представляет байесовская оценка $\alpha_{Q^{(k)}}^*$ параметра α с априорным нормальным распределением $\mathbf{Q}^{(k)} = \mathbf{\Phi}_{0,k}$. В этом случае мы должны положить $\lambda(dt) = dt$,

$$q^{(k)}(t) = \frac{1}{\sqrt{2\pi k}} e^{-t^2/2k}.$$

Апостериорное распределение $\mathbf{Q}_X^{(k)}$ будет иметь плотность $q^{(k)}(t/X)$, пропорциональную (как функция от t) $q^{(k)}(t)f_t(X)$ или, что то же, пропорциональную

$$\exp\left\{-\frac{t^2}{2k} - \frac{1}{2}\sum(\mathrm{x}_i - t)^2\right\}.$$

Из равенства

$$-\frac{t^2}{2}\left(\frac{1}{k} + n\right) + \bar{\mathrm{x}}nt = -\frac{1}{2}\left(\frac{1}{k} + n\right)\left(t - \frac{\bar{\mathrm{x}}n}{\frac{1}{k} + n}\right)^2 + \frac{(\bar{\mathrm{x}}n)^2}{2\left(\frac{1}{k} + n\right)}$$

следует, что

$$\mathbf{Q}_X^{(k)} = \mathbf{\Phi}_{\bar{\mathrm{x}}nk/(1+nk),\, k/(1+nk)}.$$

Так как байесовская оценка $\alpha_{Q^{(k)}}^*$ параметра α равна математическому ожиданию апостериорного распределения, то отсюда получаем

$$\alpha_{Q^{(k)}}^* = \frac{\bar{\mathrm{x}}nk}{1+nk} = \frac{\bar{\mathrm{x}}}{1 + \frac{1}{nk}}.$$

Дисперсия апостериорного распределения $\sigma^2_{Q_X^{(k)}} = \dfrac{k}{1+nk}$ от X не зависит. Стало быть, в силу (4) среднеквадратичная ошибка байесовской оценки равна

$$\frac{k}{1+nk} \to \frac{1}{n}$$

при $k \to \infty$. Поэтому для оценки $\alpha^* = \overline{\mathrm{x}}$ имеем

$$\mathbf{E}_t(\overline{\mathrm{x}} - t)^2 = \frac{1}{n} = \lim_{k\to\infty} \int \mathbf{E}_t\bigl(\alpha^*_{Q^{(k)}} - t\bigr)^2 q^{(k)}(t)\,dt,$$

и, стало быть, по теореме 3 оценка $\alpha^* = \overline{\mathrm{x}}$ минимаксна. «Наихудшим» распределением здесь было бы равномерное распределение на всей прямой («предельное» для $\mathbf{\Phi}_{0,k}$), если бы таковое существовало*).

В следующем примере множество Θ компактно и «наихудшее» распределение существует.

Пример 2. Пусть $X \in \mathbf{B}_p$, т. е. x_j, $j = 1, \ldots, n$, принимают значения 1 и 0 соответственно с вероятностями p и $1-p$, $p \in \Theta = [0,1]$. Как мы знаем, в этом случае для оценки $p^* = \overline{\mathrm{x}}$ справедливо

$$\mathbf{E}_p(\overline{\mathrm{x}} - p)^2 = \frac{p(1-p)}{n},$$

так что критерий теоремы 2 не выполнен. Рассмотрим оценку

$$p^* = \frac{\overline{\mathrm{x}} + 1/2\sqrt{n}}{1 + 1/\sqrt{n}}. \tag{8}$$

Для нее погрешность

$$\mathbf{E}_p(p^* - p)^2 = \left(1 + \frac{1}{\sqrt{n}}\right)^{-2} \mathbf{E}_p\left(\overline{\mathrm{x}} - p + \frac{1}{2\sqrt{n}} - \frac{p}{\sqrt{n}}\right)^2 =$$

$$= \frac{n}{(1+\sqrt{n})^2}\left(\frac{p(1-p)}{n} + \frac{(1-2p)^2}{4n}\right) = \frac{1}{4(1+\sqrt{n})^2}$$

не зависит от p. Если мы убедимся теперь, что оценка (8) является байесовской, мы установим тем самым ее минимаксность. Рассмотрим априорное распределение $\mathbf{Q} = \mathbf{B}_{N+1,N+1}$, где $\mathbf{B}_{\lambda_1,\lambda_2}$ есть бэта-распределение с плотностью (см. п. 12.8)

$$\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} t^{\lambda_1-1}(1-t)^{\lambda_2-1}.$$

Тогда, поскольку

$$f_t(X) = t^{\overline{\mathrm{x}}n}(1-t)^{n(1-\overline{\mathrm{x}})}, \quad q(t) = \frac{\Gamma(2N+2)}{\Gamma^2(N+1)} t^N(1-t)^N,$$

*) Интересно отметить, что оценка $\alpha^* = \overline{\mathrm{x}}$ перестает обладать отмеченным свойством, если X есть выборка из многомерного нормального распределения размерности, большей двух ($\mathrm{x}_j \in \mathbb{R}^k$, $\alpha \in \mathbb{R}^k$, $k \geqslant 3$). Подробнее об этом см. в [49].

то апостериорное распределение будет иметь плотность $q(t/X)$, которая как функция от t будет пропорциональна $f_t(X)q(t)$ или, что то же, пропорциональна

$$t^{N+\bar{x}n}(1-t)^{N+(1-\bar{x})n}.$$

Это означает, что апостериорное распределение совпадает с $\mathbf{B}_{N+\bar{x}n+1,\,N+n(1-\bar{x})+1}$. Так как среднее значение распределения $\mathbf{B}_{\lambda_1,\lambda_2}$ равно $\lambda_1/(\lambda_1+\lambda_2)$ (см. п. 12.8), то байесовская оценка p^*_Q, соответствующая $\mathbf{Q}$, будет равна

$$p^*_Q = \frac{N+n\bar{x}+1}{2N+n+2} = \frac{\bar{x}+(N+1)/n}{1+2(N+1)/n}.$$

При $N+1=\sqrt{n}/2$ эта оценка будет совпадать с оценкой p^*, определенной в (8), и в силу теоремы 2 будет минимаксной. Распределение $\mathbf{Q}$ будет наихудшим. Оно концентрируется с ростом n около «наихудшего» значения параметра $p=1/2$, при котором дисперсия оценки $\bar{x}$, равная $p(1-p)/n=1/(4n)$, будет максимальна. Сама оценка $\bar{x}$ не является минимаксной, так как

$$\sup_p \frac{p(1-p)}{n} = \frac{1}{4n} > \frac{1}{4(1+\sqrt{n})^2}.$$

В то же время ясно, что для всех значений p, лежащих вне узкой окрестности точки $p=1/2$, оценка $\bar{x}$ все же будет лучше, чем p^*_Q, — это будет иметь место для всех p, для которых

$$p(1-p) < \frac{1}{4(1+1/\sqrt{n})^2}.$$

В общем случае отыскание точных выражений (явных функций от X) для байесовских и минимаксных оценок не всегда возможно. Поэтому естественно использовать также *асимптотический подход.*

Прежде чем вводить соответствующие определения, напомним, что байесовские и минимаксные оценки θ^*_Q и $\overline{\theta}^*$ определялись неравенствами

$$\begin{aligned} &\mathbf{E}(\theta^*_Q-\theta)^2 - \mathbf{E}(\theta^*-\theta)^2 \leqslant 0,\\ &\sup_{t\in\Gamma}\mathbf{E}_t(\overline{\theta}^*-t)^2 - \sup_{t\in\Gamma}\mathbf{E}_t(\theta^*-t)^2 \leqslant 0 \end{aligned} \tag{9}$$

для любой оценки θ^*. Было бы неразумно определять асимптотическую байесовость и минимаксность оценок, просто добавляя к левым частям знак предельного перехода $\lim\limits_{n\to\infty}$, так как обычно для а. н. оценок $\mathbf{E}_\theta(\theta^*-\theta)^2 \sim$ $\sim \sigma^2(\theta)/n$ и левые части в (9) будут сходиться к нулю. Поэтому естественно рассматривать, скажем, отношение слагаемых в (9). Учитывая, что в дальнейшем мы будем иметь дело главным образом с оценками, для которых $\mathbf{E}_\theta(\theta^*-\theta)^2$ имеет порядок малости $1/n$, можно эквивалентным образом использовать следующее определение.

Определение 3. Оценка θ_1^* называется *асимптотически байесовской* или *асимптотически минимаксной*, если для любой другой оценки θ^* выполняется соответственно

$$\limsup_{n\to\infty}\left[\mathbf{E}n\left(\theta_1^*-\theta\right)^2-\mathbf{E}n(\theta^*-\theta)^2\right]\leqslant 0,$$

$$\limsup_{n\to\infty}\left[\sup_{t\in\Gamma}\mathbf{E}_t n\left(\theta_1^*-t\right)^2-\sup_{t\in\Gamma}\mathbf{E}_t n(\theta^*-t)^2\right]\leqslant 0.$$

Отыскание асимптотически байесовских и асимптотически минимаксных оценок возможно, как мы увидим, при весьма широких предположениях.

В *многомерном случае* (когда $\theta\in\mathbb{R}^k$ есть вектор) свойство 9 у. м. о., как мы видели, сохраняется, и оценка

$$\theta_Q^*=\mathbf{E}(\theta/X)$$

будет минимизировать

$$v(\theta^*)=\mathbf{E}(\theta^*-\theta)V(\theta^*-\theta)^\top=\mathbf{E}\mathbf{E}_\theta(\theta^*-\theta)V(\theta^*-\theta)^\top=$$
$$=\int\mathbf{E}_t(\theta^*-t)V(\theta^*-t)^\top q(t)\lambda(dt)$$

для любой неотрицательно определенной матрицы V или, что то же (см. § 18), минимизировать усредненное (с весом $q(t)$) среднеквадратичное уклонение $\theta^*-\theta$ по любому направлению $a\in\mathbb{R}^k$.

Определение 4. Оценка θ_Q^* называется *байесовской*, если для любой другой оценки θ^* и для любой неотрицательно определенной матрицы V

$$v\left(\theta_Q^*\right)\leqslant v(\theta^*).$$

Оценка θ_1^* называется *асимптотически байесовской*, если

$$\limsup_{n\to\infty}\left[nv\left(\theta_1^*\right)-nv\left(\theta_Q^*\right)\right]\leqslant 0.$$

Определение 5. Оценка $\overline{\theta}^*$ называется *минимаксной*, если для любой другой оценки θ^* и для любой неотрицательной определенной матрицы V

$$\sup_{t\in\Gamma}\mathbf{E}_t(\overline{\theta}^*-t)V(\overline{\theta}^*-t)^\top-\sup_{t\in\Gamma}\mathbf{E}_t(\theta^*-t)V(\theta^*-t)^\top\leqslant 0.$$

Оценка θ_1^* называется *асимптотически минимаксной*, если

$$\limsup_{n\to\infty}\left[\sup_{t\in\Gamma}\mathbf{E}_t n\left(\theta_1^*-t\right)V\left(\theta_1^*-t\right)^\top-\sup_{t\in\Gamma}\mathbf{E}_t n\left(\overline{\theta}^*-t\right)V\left(\overline{\theta}^*-t\right)^\top\right]\leqslant 0.$$

В заключение этого параграфа еще раз отметим, что обозначения $\mathbf{E}_\theta S$, $\mathbf{P}_\theta(A)$, $f_\theta(x)$ в байесовском случае можем рассматривать, если потребуется, с новой точки зрения — как условные относительно θ математические ожидания, вероятности и плотности $\mathbf{E}(S/\theta)$, $\mathbf{P}(A/\theta)$, $f(x/\theta)$.

§ 22. Достаточные статистики

В предыдущем параграфе мы рассматривали вопрос о построении двух типов оптимальных оценок — байесовских и минимаксных. В этом параграфе введем понятие достаточной статистики, которое позволит нам строить эффективные оценки — другой тип оптимальных оценок, выделенный нами в § 18.

Понятие достаточной статистики играет важную роль в математической статистике вообще и в теории оценок в частности. Условимся обозначать статистики, т. е. произвольные измеримые (скалярные и векторные) функции от X символом $S = S(X)$.

Пусть $X \in \mathbf{P}_\theta$, $\mathbf{P}_\theta \in \mathcal{P} = \{\mathbf{P}_\theta\}$. Рассмотрим условное относительно случайной величины S распределение $\mathbf{P}_\theta(X \in B/S)$, $B \in \mathfrak{B}^n_{\mathcal{X}}$, порожденное распределением $\mathbf{P}_\theta$ в $\mathcal{X}^n$.

Определение 1. Статистика $S = S(X)$ называется *достаточной для параметра* θ, если существует вариант условного распределения $\mathbf{P}_\theta(X \in B/S)$, не зависящий от θ.

Очевидно, что это определение эквивалентно следующему (ср. с определением 19.2А и теоремой 20.1): *статистика S является достаточной*, если для любой статистики $T = T(\mathbf{X})$ существует вариант у. м. о. $\mathbf{E}_\theta(T/S)$, не зависящий от θ.

Мы знаем, что $\mathbf{P}_\theta(X \in B/S)$ при каждом B есть у. м. о. и, следовательно, существует функция $P(B/s)$, борелевская по s при каждом B такая, что

$$\mathbf{P}_\theta(X \in B/S) = P(B/S).$$

Мы можем считать (см. § 20), что $P(B/s)$ как функция от B есть *условное при условии $S = s$ распределение вероятностей*. Это распределение можно интерпретировать как *распределение X на поверхности $S(x) = s$*.

Но если S — достаточная статистика, то это распределение *не зависит от θ*! Это означает, что знание того, где находится выборочная точка X на поверхности $S(x) = s$, не сообщает нам никакой дополнительной информации о параметре θ. (Ведь ясно, что никто не станет определять неизвестный параметр в примере 1 введения с помощью подбрасывания монеты по той причине, что распределение числа «гербов» или «решек» при таком подбрасывании от θ никак не зависит.)

Это важное обстоятельство означает, в свою очередь, что вся информация о параметре θ содержится в значении статистики S. Отсюда и происходит ее название — достаточная статистика; грубо говоря, знания $S(X)$ *достаточно* для построения оценки параметра θ; остальные данные, содержащиеся в выборке X, бесполезны.

Пример 1. Пусть $X \in \mathbf{\Pi}_\lambda$. Покажем, что статистика $S = n\overline{\mathrm{x}} = \sum_{i=1}^{n} \mathrm{x}_i$ является достаточной для параметра закона Пуассона λ. Нам надо убедиться, что распределение положения точки X на поверхности $\sum_{i=1}^{n} x_i = s$ (s целое)

от λ не зависит. Так как $\mathbf{P}\left(X = \boldsymbol{x}, \sum \mathrm{x}_i = s\right) = \mathbf{P}(X = \boldsymbol{x})$ при $\sum_{i=1}^{n} x_i = s$, то

$$\mathbf{P}(X = \boldsymbol{x}/n\bar{\mathrm{x}} = s) = \begin{cases} \dfrac{\mathbf{P}(\mathrm{x}_1 = x_1, \ldots, \mathrm{x}_n = x_n)}{\mathbf{P}(n\bar{\mathrm{x}} = s)}, & \text{если } \sum_{i=1}^{n} x_i = s, \\ 0, & \text{если } \sum_{i=1}^{n} x_i \neq s. \end{cases} \tag{1}$$

Поскольку x_i независимы, $\sum_{i=1}^{n} \mathrm{x}_i \in \mathbf{\Pi}_{n\lambda}$, то правая часть (1) равна

$$\left(e^{-n\lambda}\frac{(n\lambda)^s}{s!}\right)^{-1} \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \frac{s!}{n^s \prod_{i=1}^{n} x_i!}.$$

Таким образом, условное при условии $S = s$ распределение X совпадает с полиномиальным распределением $\mathbf{B}_p^s$ (см. § 12) с n равновероятными исходами (т. е. с вектором вероятностей $p = (1/n, \ldots, 1/n)$) и с s независимыми испытаниями. Очевидно, это распределение от λ не зависит, так что $S = n\bar{\mathrm{x}}$ является достаточной статистикой для λ.

Понятие достаточной статистики было введено Фишером в 1922 г. Следующая теорема Неймана–Фишера носит название факторизационной теоремы, она устанавливает простой критерий существования достаточной статистики.

Предположим, что выполнено условие (A_μ) о существовании плотности $f_\theta(x) = \dfrac{d\mathbf{P}_\theta}{d\mu}(x)$.

Теорема 1. *Для того чтобы S была достаточной статистикой для θ, необходимо и достаточно, чтобы функция правдоподобия $f_\theta(\boldsymbol{x}) = \prod_{i=1}^{n} f_\theta(x_i)$ была представима в виде*

$$f_\theta(\boldsymbol{x}) = \psi(S(\boldsymbol{x}), \theta)h(\boldsymbol{x}) \quad \textit{п. в.} \quad [\mu^n], \tag{2}$$

где каждая из функций $\psi \geqslant 0$ и $h \geqslant 0$ зависит только от своих аргументов, $\psi(s, \theta)$ измерима по s, $h(\boldsymbol{x})$ измерима по $\boldsymbol{x}$.

Понятно, что представление (2) неоднозначно. Компоненты его определены с точностью до произвольной положительной функции от $S(x)$.

В рассмотренном нами выше примере с распределением Пуассона

$$f_\lambda(\boldsymbol{x}) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{n\bar{\mathrm{x}}} \prod_{i=1}^{n} \frac{1}{x_i!}, \quad n\bar{\mathrm{x}} = \sum_{i=1}^{n} x_i,$$

так что мы можем положить при $S = n\bar{\mathrm{x}}$

$$\psi(S, \lambda) = e^{-n\lambda}\lambda^S, \quad h(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{x_i!}.$$

Отсюда в силу теоремы 1 будет следовать, что $S = n\bar{\mathrm{x}}$ есть достаточная статистика.

Из методических соображений мы приведем два варианта доказательства теоремы 1. Первый — для наиболее простого специального случая, когда наблюдения x_i дискретны. В этом случае доказательство оказывается весьма простым. Второй, несколько более сложный вариант доказательства теоремы Неймана–Фишера, будет относиться к некоторому обобщению теоремы 1, в котором условие (A_μ) не предполагается.

Доказательство теоремы 1 в дискретном случае.

В *дискретном случае* μ есть считающая мера на счетном множестве $\mathcal{X}$ возможных значений x_1, и, стало быть, $f_\theta(x) = \mathbf{P}_\theta(\mathrm{x}_1 = x)$, $x \in \mathcal{X}$. Пусть сначала выполнено (2). Тогда для фиксированной точки $\boldsymbol{x} \in \mathcal{X}^n$

$$\mathbf{P}_\theta(X = \boldsymbol{x}/S(X) = S(\boldsymbol{x})) = \frac{\mathbf{P}_\theta(X = \boldsymbol{x}, S(X) = S(\boldsymbol{x}))}{\mathbf{P}_\theta(S(X) = S(\boldsymbol{x}))}. \tag{3}$$

Так как $\{X = \boldsymbol{x},\ S(X) = S(\boldsymbol{x})\} = \{X = \boldsymbol{x}\}$, то правая часть (3) равна

$$\frac{\mathbf{P}_\theta(X = \boldsymbol{x})}{\mathbf{P}_\theta(S(X) = S(\boldsymbol{x}))} = \frac{f_\theta(\boldsymbol{x})}{\sum\limits_{\boldsymbol{y}:S(\boldsymbol{y})=S(\boldsymbol{x})} f_\theta(\boldsymbol{y})} =$$

$$= \frac{\psi(S(\boldsymbol{x}), \theta)h(\boldsymbol{x})}{\sum\limits_{\boldsymbol{y}:S(\boldsymbol{y})=S(\boldsymbol{x})} \psi(S(\boldsymbol{x}), \theta)h(\boldsymbol{y})} = \frac{h(\boldsymbol{x})}{\sum\limits_{\boldsymbol{y}:S(\boldsymbol{y})=S(\boldsymbol{x})} h(\boldsymbol{y})}.$$

Таким образом, $\mathbf{P}_\theta(X = \boldsymbol{x}/S(X) = S(\boldsymbol{x}))$ от θ не зависит.

Наоборот, если левая часть в (3) от θ не зависит, то, обозначая ее через $h(\boldsymbol{x})$, из (3) находим

$$\mathbf{P}_\theta(X = \boldsymbol{x}) = f_\theta(\boldsymbol{x}) = \mathbf{P}_\theta(X = \boldsymbol{x};\ S(X) = S(\boldsymbol{x})) = h(\boldsymbol{x})\mathbf{P}_\theta(S(X) = S(\boldsymbol{x})),$$

где $\mathbf{P}_\theta(S(X) = S(\boldsymbol{x})) = \psi(S(\boldsymbol{x}), \theta)$ зависит только от $S(\boldsymbol{x})$ и θ. $\triangleleft$

Лишь немногим сложнее доказывается теорема 1 и в другом важном частном случае — «гладком», когда μ есть мера Лебега в $\mathbb{R}$, а статистика $S(X)$ предполагается гладкой функцией от X.

Чтобы доказать теорему 1 в общем случае, сначала несколько изменим ее формулировку. Заметим прежде всего, что для простоты мы можем, не ограничивая общности, считать $n = 1$, так как выборку $\mathbf{X}$ мы можем рассматривать как выборку объема 1 из совокупности с плотностью распределения $f_\theta(\boldsymbol{x}) = \prod\limits_{i=1}^{n} f_\theta(x_i)$. Тогда теорема 1 будет эквивалентна следующему утверждению.

Теорема 1А. *Для того чтобы S была достаточной статистикой, необходимо и достаточно, чтобы при каждых θ, θ' отношение $\dfrac{f_\theta(\boldsymbol{x})}{f_{\theta'}(\boldsymbol{x})}$ было функцией от $S(\boldsymbol{x})$, т. е. чтобы существовала измеримая по s функция $r_{\theta,\theta'}(s)$ такая, что*

$$\frac{f_\theta(\boldsymbol{x})}{f_{\theta'}(\boldsymbol{x})} = r_{\theta,\theta'}(S(\boldsymbol{x})). \tag{4}$$

Действительно, если верно (2), то (4) очевидно. Если верно (4), то, фиксируя θ', получим равенство

$$f_\theta(\boldsymbol{x}) = r_{\theta,\theta'}(S(\boldsymbol{x}))f_{\theta'}(\boldsymbol{x}),$$

которое влечет за собой (2) при $h(\boldsymbol{x}) = f_{\theta'}(\boldsymbol{x})$, $\psi(\theta, s) = r_{\theta,\theta'}(s)$ (θ' фиксировано). $\triangleleft$

Характеризация достаточности с помощью теоремы 1А обладает двумя преимуществами.

1. В ней отсутствует проблема неоднозначности представления (2), связанного с неопределенным множителем, зависящим от S. (Поэтому в ряде случаев для отыскания достаточной статистики S удобнее пользоваться теоремой 1А, а не теоремой 1.)

2. Она позволяет получать критерии достаточности, не связанные с условием (A_μ).

Чтобы убедиться в последнем, заметим, что теорему 1А можно переформулировать следующим образом.

Статистика S является достаточной тогда и только тогда, когда при каждых θ, θ' отношение

$$\frac{f_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x}) + f_{\theta'}(\boldsymbol{x})} \equiv R_{\theta,\theta'}(\boldsymbol{x}) = \frac{r_{\theta,\theta'}(\boldsymbol{x})}{1 + r_{\theta,\theta'}(\boldsymbol{x})} \tag{5}$$

является функцией от $S(\boldsymbol{x})$.

Рассмотрим теперь усредненное распределение

$$\overline{\mathbf{P}} = \mathbf{P}_{\theta,\theta'} = \frac{1}{2}(\mathbf{P}_\theta + \mathbf{P}_{\theta'}).$$

Очевидно, что распределение $\mathbf{P}_\theta$ абсолютно непрерывно относительно $\mathbf{P}_{\theta,\theta'}$ и, стало быть, существует плотность

$$f_{\theta,\theta'}(\boldsymbol{x}) = \frac{d\mathbf{P}_\theta}{d\mathbf{P}_{\theta,\theta'}}(\boldsymbol{x}) \tag{6}$$

(она всегда ограничена, так как $f_{\theta,\theta'}(\boldsymbol{x}) + f_{\theta',\theta}(\boldsymbol{x}) = 2$). Если выполнено условие (A_μ), то

$$f_{\theta,\theta'}(\boldsymbol{x}) = 2R_{\theta,\theta'}(\boldsymbol{x}).$$

Таким образом, функция R, присутствующая в (5), допускает определение, никак не связанное с условием (A_μ) (этого нельзя сказать о функции r в

теореме 1А, так как $\mathbf{P}_\theta$ не всегда абсолютно непрерывно относительно $\mathbf{P}_{\theta'}$, как, например, в случае, когда θ есть параметр сдвига: $\mathbf{P}_\theta(B) = \mathbf{P}(B - \theta)$, $\mathbf{P}_{\theta_0}([0,1]) = 1$ при некотором θ_0).

Теперь мы можем сформулировать характеризационную теорему в виде, не связанном с условием (A_μ).

Теорема 1В. *Статистика S является достаточной тогда и только тогда, когда при каждых θ, θ' плотность $f_{\theta,\theta'}(\boldsymbol{x})$ есть функция от S, т. е. существует измеримая по s функция $g_{\theta,\theta'}(s)$ такая, что $f_{\theta,\theta'}(\boldsymbol{x}) = g_{\theta,\theta'}(S(\boldsymbol{x}))$.*

Из сделанных выше замечаний ясно, что при выполнении (A_μ) это утверждение эквивалентно теореме 1А и, следовательно, теореме 1. Доказав теорему 1В, тем самым докажем теорему 1.

Доказательство теоремы 1В. Необходимость. Пусть статистика S является достаточной. Тогда для любой статистики $T = T(\mathbf{X})$

$$\mathbf{E}_\theta(T/S) = \mathbf{E}_{\theta'}(T/S) \equiv H(S(\boldsymbol{x})) \tag{7}$$

от θ и θ' не зависит. Наряду с у. м. о. $\mathbf{E}_\theta$, построенным по распределению $\mathbf{P}_\theta$, можем строить также у. м. о. $\overline{\mathbf{E}}$ по усредненному распределению $\overline{\mathbf{P}} = \mathbf{P}_{\theta,\theta'}$. Покажем, что $\overline{\mathbf{E}}(T/S) = \mathbf{E}_\theta(T/S)$ и, следовательно, также не зависит от θ, θ'. Действительно, для любой измеримой функции m в силу (7) и свойств у. м. о.

$$\overline{\mathbf{E}}(Tm(S)) = \frac{1}{2}[\mathbf{E}_\theta(Tm(S)) + \mathbf{E}_{\theta'}(Tm(S))] =$$

$$= \frac{1}{2}[\mathbf{E}_\theta(H(S)m(S)) + \mathbf{E}_{\theta'}(H(S)m(S))] = \overline{\mathbf{E}}H(S)m(S).$$

Отсюда следует, что (см. определение 19.2А)

$$\overline{\mathbf{E}}(T/S) = H(S) \tag{8}$$

от θ и θ' не зависит (т. е. S является достаточной статистикой и для семейства $\overline{\mathbf{P}} = \mathbf{P}_{\theta,\theta'}$).

Нам надо доказать, что $f = f_{\theta,\theta'}(\boldsymbol{x})$ является функцией от $S(\boldsymbol{x})$. Для этого покажем, что

$$f_{\theta,\theta'}(\mathbf{X}) = \overline{\mathbf{E}}(f_{\theta,\theta'}(\mathbf{X})/S) \quad \text{п. в.}$$

или, что то же,

$$d \equiv \overline{\mathbf{E}}[f - \overline{\mathbf{E}}(f/S)]^2 = \overline{\mathbf{E}}f^2 - 2\overline{\mathbf{E}}f\overline{\mathbf{E}}(f/S) + \overline{\mathbf{E}}(\overline{\mathbf{E}}(f/S))^2 = 0.$$

Но в этом соотношении в силу свойств у. м. о.

$$\overline{\mathbf{E}}f\overline{\mathbf{E}}(f/S) = \overline{\mathbf{E}}(\overline{\mathbf{E}}(f/S)\overline{\mathbf{E}}(f/S)) = \overline{\mathbf{E}}(\overline{\mathbf{E}}(f/S))^2$$

и, стало быть,

$$d = \overline{\mathbf{E}}f^2 - \overline{\mathbf{E}}f\overline{\mathbf{E}}(f/S),$$

где в силу равенства (6)

$$\overline{\mathbf{E}}f^2 = \int f \cdot f\mathbf{P}_{\theta,\theta'}(d\boldsymbol{x}) = \int f\mathbf{P}_\theta(d\boldsymbol{x}) = \mathbf{E}_\theta f,$$

$$\overline{\mathbf{E}}f\overline{\mathbf{E}}(f/S) = \int \overline{\mathbf{E}}(f/S)f\mathbf{P}_{\theta,\theta'}(d\boldsymbol{x}) = \int \overline{\mathbf{E}}(f/S)\mathbf{P}_\theta(d\boldsymbol{x}) = \mathbf{E}_\theta\overline{\mathbf{E}}(f/S). \tag{9}$$

Но согласно (7), (8) $\overline{\mathbf{E}}(f/S) = \mathbf{E}_\theta(f/S)$. Следовательно, правая часть (9) будет равна $\mathbf{E}_\theta\mathbf{E}_\theta(f/S) = \mathbf{E}_\theta f$, $d = 0$.

Достаточность. Пусть теперь $f_{\theta,\theta'}(\boldsymbol{x}) = g_{\theta,\theta'}(S(\boldsymbol{x}))$ является функцией от $S(\boldsymbol{x})$. Покажем, что тогда для произвольной статистики T и произвольных θ, θ'

$$\mathbf{E}_\theta(T/S) = \mathbf{E}_{\theta'}(T/S), \tag{10}$$

что означает независимость $\mathbf{E}_\theta(T/S)$ от θ и достаточность S. Положим $H(S) = \overline{\mathbf{E}}(T/S)$. Тогда для любой измеримой функции m

$$\mathbf{E}_\theta m(S)H(S) = \int m(S(\boldsymbol{x}))H(S(\boldsymbol{x}))\mathbf{P}_\theta(d\boldsymbol{x}) =$$

$$= \int m(S(\boldsymbol{x}))H(S(\boldsymbol{x}))f_{\theta,\theta'}(\boldsymbol{x})\overline{\mathbf{P}}(d\boldsymbol{x}) = \int m(s)H(s)g_{\theta,\theta'}(s)\overline{\mathbf{P}}(S(\mathbf{X}) \in ds) =$$

$$= \overline{\mathbf{E}}m(S)H(S)g_{\theta,\theta'}(S) = \overline{\mathbf{E}}m(S)Tg_{\theta,\theta'}(S).$$

Те же рассуждения в обратном направлении позволяют продолжить эти равенства следующим образом:

$$= \int m(S(\boldsymbol{x}))T(\boldsymbol{x})g_{\theta,\theta'}(S(\boldsymbol{x}))\overline{\mathbf{P}}(d\boldsymbol{x}) = \int m(S(\boldsymbol{x}))T(\boldsymbol{x})\mathbf{P}_\theta(d\boldsymbol{x}) = \mathbf{E}_\theta m(S)T.$$

Отсюда следует

$$H(S) = \mathbf{E}_\theta(T/S).$$

Поменяем теперь местами θ и θ'. Функция $f_{\theta',\theta}(\boldsymbol{x})$ также будет функцией от $S(\boldsymbol{x})$. Поэтому совершенно аналогично получим

$$H(S) = \mathbf{E}_{\theta'}(T/S),$$

что доказывает (10). Теорема доказана. $\triangleleft$

Вернемся теперь к теореме 1 и примерам достаточных статистик.

Пример 2. Пусть $X \in \boldsymbol{\Phi}_{\alpha,\sigma^2}$. Здесь параметр $\theta = (\alpha, \sigma^2)$ двумерный. Имеем

$$f_\theta(X) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\mathrm{x}_i - \alpha)^2}{2\sigma^2}\right\} = \sigma^{-n}(2\pi)^{-n/2} \exp\left\{-\frac{\sum(\mathrm{x}_i - \alpha)^2}{2\sigma^2}\right\} =$$

$$= \sigma^{-n} \exp\left\{-\frac{\sum \mathrm{x}_i^2 - 2\alpha n\overline{\mathrm{x}} + n\alpha^2}{2\sigma^2}\right\}(2\pi)^{-n/2}.$$

Полагая $S = (S_1, S_2)$, $S_1 = n\bar{x}$, $S_2 = \sum_{i=1}^{n} x_i^2$, получим представление (2), где

$$\psi(S, \theta) = \sigma^{-n} \exp\left\{-\frac{S_2 - 2\alpha S_1 + \alpha^2}{2\sigma^2}\right\}, \quad h(X) = (2\pi)^{-n/2}.$$

Здесь мы могли бы, конечно, множитель $(2\pi)^{-n/2}$ отнести и к функции ψ, положив $h(X) = 1$.

Таким образом, мы получили, что статистика (S_1, S_2) является достаточной векторной статистикой для (α, σ^2). Из всей информации, содержащейся в выборке, нам достаточно знать $\bar{x}$ и $\sum x_i^2$.

Предлагаем читателю найти достаточные статистики для всех семейств распределений, приведенных в § 12.

Остановимся подробнее лишь на одном их этих семейств.

Пример 3. Пусть $X \in \mathbf{U}_{0,\theta}$. Здесь условие (A_μ) выполнено относительно меры Лебега и

$$f_\theta(X) = \begin{cases} \theta^{-n}, & \text{если } 0 \leqslant x_i \leqslant \theta \text{ при всех } i = 1, \ldots, n, \\ 0 & \text{в противном случае.} \end{cases}$$

Пусть $x_{(1)} = \min x_i$, $x_{(n)} = \max x_i$. Тогда, как мы видели в примере 16.5, функцию $f_\theta(X)$ можно записать в виде $f_\theta(X) = \psi(x_{(n)}, \theta)h(X)$, где

$$h(X) = \begin{cases} 1, & \text{если } x_{(1)} \geqslant 0, \\ 0 & \text{в противном случае,} \end{cases}$$

$$\psi(s, \theta) = \begin{cases} \theta^{-n} & \text{при } s \leqslant \theta, \\ 0 & \text{в противном случае.} \end{cases}$$

Это означает, что $S(X) = x_{(n)}$ есть достаточная статистика для θ.

Аналогичным образом читатель может убедиться, что для выборки $X \in \mathbf{U}_{\theta,1+\theta}$ достаточной статистикой для параметра θ является двумерная статистика $S(X) = (x_{(1)}, x_{(n)})$. Такой же будет достаточная статистика и для двумерного параметра $\theta = (a, b)$, когда выборка извлечена из распределения $\mathbf{U}_{a,b}$.

Приведем некоторые следствия теоремы 1.

Следствие 1. *Если S — достаточная статистика для θ, то оценка максимального правдоподобия зависит лишь от S.*

Точнее, о. м. п. $\widehat{\theta}^*$ не зависит от X при фиксированном $S(X)$.

Это следствие очевидно, так как о. м. п. есть значение θ, при котором достигается максимум $f_\theta(X) = \psi(S(X), \theta)h(X)$ или, что то же, максимум $\psi(S(X), \theta)$.

Следствие 2. *Если S — достаточная статистика и функция ψ такова, что отображение $u = \varphi(v)$ взаимно однозначно и измеримо в обе стороны, то $S_1 = \varphi(S)$ также будет достаточной статистикой.*

Это следствие также очевидно, поскольку $\psi(S,\theta)$ в (2) можно записать в виде $\psi(\varphi^{-1}(S_1),\theta)=\psi_1(S_1,\theta)$.

Следствие 3. *Пусть нам надо оценить параметр* $\gamma=\varphi(\theta)$, *где* φ *осуществляет взаимно однозначное отображение, так что определена функция* $\theta=\varphi^{-1}(\gamma)$. *Пусть, далее,* S *— достаточная статистика для* θ. *Тогда* S *будет также достаточной статистикой для* γ.

Это следствие вытекает из тех же соображений, что и в следствии 2. Действительно, функция правдоподобия для γ будет иметь вид $\psi(S(X),\varphi^{-1}(\gamma))\times$ $\times\, h(X)$, где первый множитель можно записать в виде $\psi_1(S(X),\gamma)$.

Справедлив еще один критерий достаточности статистики S, который можно рассматривать как некоторую модификацию теоремы 1В.

Теорема 2. *Статистика* S *является достаточной для* θ *тогда и только тогда, когда для любого априорного распределения* $\mathbf{Q}$ *параметра* θ *апостериорное распределение* $\mathbf{Q}_X$ *зависит от* X *только через* $S(X)$ (*т. е. остается неизменным на поверхности* $S(X)=s$).

Доказательство. Пусть S — достаточная статистика, $q(t)$ — плотность $\mathbf{Q}$ относительно какой-нибудь меры λ. Тогда апостериорная плотность $q(t/X)$ относительно этой меры будет по формуле Байеса равна

$$q(t/X)=\frac{f_t(X)q(t)}{\int f_u(X)q(u)\lambda(du)}=\frac{\psi(S(X),t)q(t)}{\int\psi(S(X),u)q(u)\lambda(du)}.$$

Докажем теперь обратное утверждение теоремы. Выберем априорное распределение так, чтобы $q(t)>0$ всюду на Θ и при всех t

$$f_t(X)=\frac{q(t/X)f(X)}{q(t)},\quad f(X)=\int f_u(X)q(u)\lambda(du).$$

Если $q(t/X)=g(t,S(X))$, то, полагая $\psi(s,t)=g(t,s)/q(t)$, $h(X)=f(X)$, получим представление (2). $\triangleleft$

Следствие 4. *Если* S *— достаточная статистика, то все байесовские оценки и минимаксные оценки, определенные с помощью теоремы* 21.2, *зависят лишь от* S.

В дальнейшем получим много других подтверждений того, что достаточная статистика S содержит в себе исчерпывающую информацию о θ.

§ 23.* Минимальные достаточные статистики

Рассмотрим теперь вопрос о выборе достаточных статистик. Ясно, что их может быть очень много. Например, статистика $S(X)\equiv X$, очевидно, всегда является достаточной. Эта достаточная статистика называется *тривиальной* достаточной статистикой. Однако мы заинтересованы (и в дальнейшем станет ясно, почему) в более «экономных» статистиках. Оказывается, далеко не всегда можно построить достаточные статистики, которые были бы существенно «экономнее» тривиальной достаточной статистики. К этому вопросу мы вернемся после того, как определим более точно понятия, связанные с

«экономностью» достаточных статистик. Для этого введем на множестве всех достаточных статистик (для некоторого параметра θ) частичный порядок.

Определение 1. Будем говорить, что статистика S_1 *подчинена* S_2, если S_1 есть измеримая функция от S_2: $S_1 = \varphi(S_2)$.

Это соотношение и означает, что S_1 «экономнее» S_2.

Определение 2. Если S_1 подчинена S_2 и S_2 подчинена S_1, то статистики S_1 и S_2 называются *эквивалентными*.

Очевидно, S_1 эквивалентна S_2 тогда и только тогда, когда $S_1 = \varphi(S_2)$ и φ — взаимно однозначное измеримое в обе стороны отображение.

Определение 3. Достаточная статистика S_θ называется *минимальной*, если она подчинена любой другой достаточной статистике S.

Минимальная достаточная статистика является самой экономной. Если мы построили минимальную достаточную статистику S, то при сохранении свойства достаточности дальнейшее сокращение данных по сравнению с S невозможно. Остальные данные, содержащиеся в выборке, можно рассматривать как порожденные некоторым случайным механизмом, не зависящим от θ. Они не несут в себе никакой информации о θ.

Введенные понятия, как и исходное понятие достаточной статистики, *можно в слегка обобщенной форме излагать на языке σ-алгебр*, который в ряде случаев оказывается более удобным и наглядным. В самом начале — в определении 22.1 предыдущего параграфа — условное распределение $\mathbf{P}_\theta(X \in B/S)$ можно заменить на условное распределение $\mathbf{P}_\theta(X \in B/\mathfrak{A})$ относительно σ-подалгебры $\mathfrak{A} \subset \mathfrak{B}^n_{\mathcal{X}}$ и называть $\mathfrak{A}$ *достаточной σ-алгеброй*, если существует вариант $\mathbf{P}_\theta(X \in B/\mathfrak{A})$, не зависящий от θ.

Факторизационная теорема при таком подходе сохранится, если функцию $\psi(S(X), \theta)$ заменить на $\mathfrak{A}$-измеримую по X функцию $\psi(X, \theta)$. Доказательство этой теоремы по существу останется прежним.

Достаточную статистику теперь можно определить как статистику S, для которой порожденная ею σ-алгебра $\sigma(S)$ будет достаточной.

Подчиненность достаточных статистик (см. определение 1) на языке σ-алгебр не требует введения дополнительных понятий и совпадает просто с вложением σ-алгебр: S_1 подчинена S_2, если $\sigma(S_1) \subset \sigma(S_2)$. Таким образом, S_1 экономнее S_2, если σ-алгебра $\sigma(S_1)$ беднее (грубее), чем $\sigma(S_2)$. Эквивалентность S_1 и S_2 означает, что $\sigma(S_1) = \sigma(S_2)$.

Минимальная достаточная σ-алгебра $\mathfrak{A}_0$ определяется как σ-алгебра, которая вкладывается в любую достаточную σ-алгебру.

Минимальная достаточная σ-алгебра всегда существует. Прежде чем сформулировать соответствующее утверждение, заметим, что существует распределение $\mathbf{Q}$ на Θ такое, что все $\mathbf{P}_\theta$ абсолютно непрерывны относительно распределения $\mathbf{P}_Q = \int \mathbf{P}_t \mathbf{Q}(dt)$. Действительно, рассмотрим байесовскую постановку задачи, когда θ есть случайная величина с априорным распределением $\mathbf{Q}$. Пусть $q(t) > 0$ есть плотность распределения относительно подходящей меры λ на Θ.

Допустим для простоты, что для п. в. x функция $f_t(x)$ в каждой точке $t \in \Theta$ непрерывна справа и слева. Тогда все распределения $\mathbf{P}_t$ будут абсолютно непрерывны относительно $\mathbf{P}_Q = \int \mathbf{P}_t q(t)\lambda(dt)$, так как $\mathbf{P}_Q(A) = 0$ влечет за собой $\mathbf{P}_t(A) = 0$ для всех t.

Это означает, что либо $f_Q(X) = \int f_t(X)\mathbf{Q}(dt) > 0$ при всех X, либо из равенства $f_Q(X) = 0$ следует $f_\theta(X) = 0$ при всех θ. В этом случае говорят, что $\mathbf{P}_Q$ *доминирует* семейство $\{\mathbf{P}_\theta\}$, так что мы могли бы $\mathbf{P}_Q$ взять в качестве меры μ. Плотность распределения $\mathbf{P}_\theta$ относительно этой меры равна

$$\frac{d\mathbf{P}_\theta}{d\mathbf{P}_Q}(\boldsymbol{x}) = \frac{f_\theta(\boldsymbol{x})}{f_Q(\boldsymbol{x})} \equiv r(\boldsymbol{x}, \theta).$$

Ясно (ср. с теоремой 22.2), что если S — достаточная статистика, то $r(\boldsymbol{x}, \theta)$ зависит от x лишь через $S(\boldsymbol{x})$.

Теорема 1. *σ-алгебра $\mathfrak{A}_0 = \sigma(r(X, \theta);\ \theta \in \Theta)$, порожденная случайными величинами $r(X, \theta) = f_\theta(X)/f_Q(X)$ при различных $\theta \in \Theta$, является минимальной достаточной σ-алгеброй.*

Доказательство теоремы весьма просто. Достаточность $\mathfrak{A}_0$ следует из факторизационной теоремы и того, что

$$f_\theta(X) = r(X, \theta) f_Q(X), \tag{1}$$

где $f_Q(X)$ не зависит от θ, а $r(X, \theta)$ измерима относительно $\mathfrak{A}_0$.

Пусть теперь $\mathfrak{A}$ есть любая достаточная σ-алгебра. Тогда $f_\theta(X) = \psi(X, \theta)h(X)$, где функция $\psi(X, \theta)$ $\mathfrak{A}$-измерима. Рассмотрим σ-алгебру $\mathfrak{A}_\psi \equiv \sigma(\psi(X, \theta),\ \theta \in \Theta) \subset \mathfrak{A}$. Из определения $r(X, \theta)$ следует, что

$$r(X, \theta) = \frac{\psi(X, \theta)}{\int \psi(X, t)\mathbf{Q}(dt)},$$

и, стало быть, $\mathfrak{A}_0 \subset \mathfrak{A}_\psi \subset \mathfrak{A}$. $\triangleleft$

С этой теоремой и теоремой 22.2 тесно связано другое полезное замечание. Апостериорная плотность, соответствующая $q(t)$, равна

$$q(t/X) = \frac{f_t(X)q(t)}{f_Q(X)} = r(X, t)q(t),$$

и, стало быть, минимальную достаточную σ-алгебру $\mathfrak{A}_0$ можно рассматривать как порожденную апостериорным распределением:

$$\mathfrak{A}_0 = \sigma(q(t/X);\ t \in \Theta).$$

Отыскание распределений $\mathbf{Q}$ и $\mathbf{P}_Q$, фигурирующих в теореме 1, обычно труда не составляет. Например, если носитель $N_{\mathbf{P}_\theta}$ распределения $\mathbf{P}_\theta$ от θ не зависит, что имеет место для большинства распределений, перечисленных в § 12, то можно взять $\mathbf{P}_Q = \mathbf{P}_{\theta_0}$ при любом $\theta_0 \in \Theta$.

Таким образом, мы располагаем теоремой существования и конструктивным способом построения минимальных достаточных σ-алгебр *).

Однако в большинстве случаев нам все же будет удобнее иметь дело со статистиками. Основная задача этого параграфа — отыскание минимальных достаточных статистик.

Прежде всего, каким образом можно проверить, что данная достаточная статистика S_0 является минимальной?

Одна из возможностей состоит в использовании теоремы 1. Если $\sigma(S_0)$ совпадает с σ-алгеброй, порожденной $f_\theta(X)/f_Q(X)$, то S_0 есть минимальная достаточная статистика.

Пример 1. Мы видели, что статистика $S = n\bar{x}$ является достаточной для параметра λ распределения Пуассона $\mathbf{\Pi}_\lambda$. Это будет минимальная достаточная статистика, так как $\sigma(S)$ совпадает, очевидно, с σ-алгеброй, порожденной $f_\lambda(X)/f_{\lambda_1}(X) = e^{n(\lambda_1-\lambda)}(\lambda/\lambda_1)^S$ (мы взяли здесь распределение $\mathbf{Q}$, сосредоточенное в точке λ_1).

Пример 2. Пусть $X \in \mathbf{U}_{0,\theta}$. Тогда статистика $S = x_{(n)} = \max x_i$ является минимальной достаточной статистикой. В самом деле, возьмем в качестве $\mathbf{Q}$ какое-нибудь распределение на $[0,\infty)$ с плотностью $q(t) > 0$ при всех $t > 0$. Тогда

$$f_\theta(X) = \begin{cases} \theta^{-n}, & \theta \geqslant S, \\ 0, & \theta < S, \end{cases}$$

$$f_Q(X) = \int_0^\infty f_t(X) q(t)\, dt = \int_S^\infty t^{-n} q(t)\, dt > 0$$

при всех X. При этом $S = \sup\{\theta : f_\theta(X)/f_Q(X) = 0\}$. Это означает, что S измерима относительно минимальной σ-алгебры $\mathfrak{A}_0$, $\sigma(S) \subset \mathfrak{A}_0$ и что, стало быть, S — минимальная достаточная статистика.

Можно указать другой способ отыскания минимальных достаточных статистик, также связанный с функцией правдоподобия. В самом деле, любая статистика, и в частности достаточная статистика S, порождает разбиение выборочного пространства на классы эквивалентности — т. е. на подмножества точек $\boldsymbol{x}$ с одинаковым значением $S(\boldsymbol{x})$.

Если S_1 подчинена S_2, т. е. $S_1 = \varphi(S_2)$, то, очевидно, разбиение для S_1 будет более крупным, так как классы эквивалентности для S_2 содержатся в классах эквивалентности для S_1. Таким образом, минимальной достаточной статистике соответствует самое «крупное» среди разбиений, порожденных достаточными статистиками.

Можно рассматривать просто разбиения пространства на классы эквивалентности, не связывая их непосредственно со статистиками. Класс эквивалентности, содержащий точку $\boldsymbol{x}$, обозначим через $D(\boldsymbol{x})$. Каждый класс однозначно определяется какой-нибудь одной своей точкой. *Разбиение на классы D будем называть достаточным, если*

$$f_\theta(\boldsymbol{x}) = \varphi(\boldsymbol{x}, \theta) h(\boldsymbol{x}), \tag{2}$$

*) Существование минимальной достаточной σ-алгебры $\mathfrak{A}_0$ можно установить и иным путем, доказав, что $\mathfrak{A}_0$ есть пересечение всех пополненных достаточных σ-алгебр.

где $\varphi(\boldsymbol{x},\theta) = \varphi(\boldsymbol{x}_0,\theta)$ *постоянно при* $\boldsymbol{x} \in D(\boldsymbol{x}_0)$ (т.е. $\varphi(\boldsymbol{x},\theta) = \text{const}$ внутри класса эквивалентности). Если классы $D(\boldsymbol{x})$ определяются соотношениями $S(\boldsymbol{x}) = s$, то из теоремы 21.1 сразу вытекает, что статистика $S(\boldsymbol{x})$ достаточна тогда и только тогда, когда разбиение на классы D достаточно.

Рассмотрим теперь разбиение, построенное следующим образом: возьмем точку $\boldsymbol{x}_0$ и объявим, что $\boldsymbol{x}$ принадлежит классу $D(\boldsymbol{x}_0)$, если отношение

$$\frac{f_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{x}_0)} = h(\boldsymbol{x},\boldsymbol{x}_0) \tag{3}$$

не зависят от θ. Очевидно, что при таком построении $D(\boldsymbol{x}_1) = D(\boldsymbol{x}_2) = D(\boldsymbol{x}_0)$, если $\boldsymbol{x}_1 \in D(\boldsymbol{x}_0)$, $\boldsymbol{x}_2 \in D(\boldsymbol{x}_0)$, так что правило (3) порождает разбиение всего пространства на непересекающиеся классы.

Это разбиение соответствует разбиению, порожденному минимальной достаточной статистикой S.

В самом деле, пусть S — минимальная достаточная статистика. Возьмем произвольную точку $\boldsymbol{x}_0$. Тогда на поверхности $S(\boldsymbol{x}) = S(\boldsymbol{x}_0)$ отношение $f_\theta(\boldsymbol{x})/f_\theta(\boldsymbol{x}_0)$ равно $h(\boldsymbol{x})/h(\boldsymbol{x}_0)$ и, стало быть, от θ не зависит. Таким образом, разбиение на кассы D не менее крупное, чем разбиение для S.

С другой стороны, это разбиение достаточное. Действительно, каждой поверхности D мы можем поставить в соответствие одну какую-нибудь ее точку $\boldsymbol{x}_D$, по которой эта поверхность будет однозначно определяться. Рассмотрим функцию $\boldsymbol{x}_0(\boldsymbol{x})$, которая определяется соотношением $\boldsymbol{x}_0(\boldsymbol{x}) = \boldsymbol{x}_D$, если $\boldsymbol{x} \in D$. Тогда в силу (3) при $\boldsymbol{x} \in D$

$$f_\theta(\boldsymbol{x}) = f_\theta(\boldsymbol{x}_D)h(\boldsymbol{x},\boldsymbol{x}_D) = f_\theta(\boldsymbol{x}_0(\boldsymbol{x}))h(\boldsymbol{x},\boldsymbol{x}_0(\boldsymbol{x})), \tag{4}$$

что означает выполнение (2).

Проведенные рассмотрения были не совсем строгими, так как мы не связывали их с вопросом об измеримости функций, входящих в (4).

Сказанное можно резюмировать следующим образом. *Пусть дана статистика* $S(X)$ *такая, что* $S(\boldsymbol{x}) = S(\boldsymbol{x}_0)$ *тогда и только тогда, когда отношение* (3) *не зависит от* θ. *В этом случае* S *есть минимальная достаточная статистика.*

В отличие от подходов, связанных с теоремой 1, где рассматривались отношения $f_\theta(\boldsymbol{x})/f_Q(\boldsymbol{x})$ или $f_\theta(\boldsymbol{x})/f_{\theta_1}(\boldsymbol{x})$ при разных θ и θ_1 (называемые часто отношениями правдоподобия), правило, сформулированное выше, использует отношение $f_\theta(\boldsymbol{x})/f_\theta(\boldsymbol{x}_0)$ при одинаковых значениях параметра θ. В примере 1, например, отношение

$$\frac{f_\lambda(\boldsymbol{x})}{f_\lambda(\boldsymbol{x}_0)} = \prod \frac{\lambda^{x_i - x_{i0}} x_{i0}!}{x_i!} = \lambda^{n(\bar{\boldsymbol{x}} - \bar{\boldsymbol{x}}_0)} \prod \frac{x_{i0}!}{x_i!}$$

не будет зависеть от λ тогда и только тогда, когда $\overline{\boldsymbol{x}} = \overline{\boldsymbol{x}}_0 = \dfrac{1}{n}\sum_{i=1}^{n} x_{i0}$, где x_{i0} — координаты вектора $\boldsymbol{x}_0$. Этого достаточно, чтобы заключить, что $S(\boldsymbol{x}) = \overline{\boldsymbol{x}}$ есть минимальная достаточная статистика.

Рассмотрим теперь с помощью предложенного правила пример, когда «экономных» достаточных статистик не существует. Прежде заметим, что вариационный ряд $S_V = (\mathrm{x}_{(1)}, \mathrm{x}_{(2)}, \ldots, \mathrm{x}_{(n)})$, построенный по выборке X, очевидно, всегда является достаточной статистикой, так как $f_\theta(X) = \prod_{i=1}^{n} f_\theta(\mathrm{x}_i) = \prod_{k=1}^{n} f_\theta(\mathrm{x}_{(k)})$. Эта статистика «чуть-чуть экономнее» самой выборки X. Отсюда, в частности, следует, что любая минимальная достаточная статистика инвариантна относительно перестановки координат x_i в выборке X.

Если плотность $f_\theta(x)$ симметрична, т. е. $f_\theta(-x) = f_\theta(x)$ при всех θ, то, очевидно, будет существовать еще немного более «экономная» достаточная статистика, представляющая из себя упорядоченную по возрастанию совокупность $(\mathrm{x}_1^2, \ldots, \mathrm{x}_n^2)$, которую мы обозначим через S_{V^2}.

Пример 3. *Если* $X \in \mathbf{K}_{0,\sigma}$, *т. е. если* x_i *имеют плотность распределения Коши с параметром* $\theta = \sigma$

$$k_{0,\sigma}(x) = \frac{\sigma}{\pi(x^2 + \sigma^2)},$$

то статистика S_{V^2} *является минимальной достаточной статистикой.*

Действительно, в этом случае

$$f_\sigma(x) = \left(\frac{\sigma}{\pi}\right)^n \prod_{i=1}^n (x_i^2 + \sigma^2)^{-1},$$

так что

$$\frac{f_\sigma(x)}{f_\sigma(x_0)} = \prod_{i=1}^n \frac{x_{i0}^2 + \sigma^2}{x_i^2 + \sigma^2} \tag{5}$$

есть отношение двух полиномов от σ^2. Оно не зависит от σ тогда и только тогда, когда коэффициенты при соответствующих степенях σ^2 в числителе и знаменателе совпадают. Это, в свою очередь, имеет место тогда и только тогда, когда множества «нулей» $\{-x_{i0}^2\}$ и $\{-x_i^2\}$ совпадают. Другими словами, для независимости (5) от σ необходимо и достаточно, чтобы точка $x^2 = (x_1^2, \ldots, x_n^2)$ имела координаты, которые отличаются от координат x_0^2 разве лишь перестановкой их мест. Это и означает, что S_{V^2} есть минимальная достаточная статистика.

Совершенно аналогично можно показать, что S_V является минимальной достаточной статистикой для параметра α и, стало быть, для параметра $\theta = (\alpha, \sigma)$ распределения $\mathbf{K}_{\alpha,\sigma}$.

Другой пример, в котором S_V будет минимальной достаточной статистикой, получим, если рассмотрим семейство

$$\mathbf{P}_{\alpha,\theta_1,\theta_2} = \alpha \mathbf{P}_{\theta_1} + (1 - \alpha)\mathbf{P}_{\theta_2}, \quad \alpha \in [0, 1],$$

где $\{\mathbf{P}_\theta\}$ — экспоненциальное семейство (см. § 25; в качестве $\mathbf{P}_\theta$ можно взять нормальное распределение или распределение Пуассона) и где хотя бы один из параметров α, θ_1, θ_2 неизвестен.

Докажем теперь теорему, указывающую простой «конструктивный» способ отыскания минимальных достаточных статистик.

Для простоты изложения рассмотрим случай одномерного параметра θ.

Теорема 2. *Пусть функция правдоподобия* $f_\theta(x)$ *при всех* x *как функция от* θ *непрерывна справа (или слева). Тогда если оценка м. п.* $\widehat{\theta}^*$ *единственна и является достаточной статистикой, то* $\widehat{\theta}^*$ *есть минимальная достаточная статистика.*

Доказательство. Пусть S — произвольная достаточная статистика. Мы докажем теорему, если покажем, что $\widehat{\theta}^*$ измерима относительно $\sigma(S)$ и, стало быть, $\widehat{\theta}^*$ подчинена S.

В силу факторизационной теоремы

$$f_\theta(\boldsymbol{x}) = \psi(S(\boldsymbol{x}), \theta)h(\boldsymbol{x}) \quad \text{п. в.} \quad [\mu^n], \tag{6}$$

где $h(\boldsymbol{x})$ — измеримая функция по $\boldsymbol{x}$, $\psi(s,t)$ непрерывна (справа или слева) по t и измерима по s. Так как $\mathbf{P}_\theta$ не изменится, если плотность $f_\theta(\boldsymbol{x})$ изменить на множестве μ^n-меры 0, мы можем считать, что (6) справедливо при всех $\boldsymbol{x}$.

В силу (6) точка абсолютного максимума $f_\theta(\boldsymbol{x})$ является точкой абсолютного максимума и для $\psi(S(\boldsymbol{x}), \theta)$. Поэтому в силу единственности $\widehat{\theta}^*$

$$\{\widehat{\theta}^* < t\} = \{\sup_{\theta<t} \psi(S(X), \theta) > \sup_{\theta \geqslant t} \psi(S(X), \theta)\}.$$

Так как $\psi(S(X), \theta)$ при каждом $S(X)$ непрерывна по θ справа (или слева), то существует счетное всюду плотное множество $\Theta_c = \{\theta_j\}_{j=1}^{\infty} \subset \Theta$ (одно и то же для всех $S(X)$) такое, что

$$\sup_{\theta<t} \psi(S(X), \theta) = \sup_{\substack{\theta_j<t \\ \theta_j \in \Theta_c}} \psi(S(X), \theta_j). \tag{7}$$

Такое же соотношение будет справедливо для области $\theta \geqslant t$. Так как $\psi(S(X), \theta_j)$ измеримы относительно $\sigma(S)$, то в силу (7) значения $\sup\limits_{\theta<t} \psi(S, \theta)$ и $\sup\limits_{\theta \in t} \psi(S, \theta)$ будут случайными величинами, также измеримыми относительно $\sigma(S)$. Следовательно, $\{\widehat{\theta}^* < t\} \in \sigma(S)$, и теорема доказана. $\triangleleft$

Условие достаточности о. м. п. $\widehat{\theta}^*$ в условии приведенного утверждения существенно, поскольку оценка м. п. $\widehat{\theta}^*$ сама по себе не обязана быть достаточной статистикой. Соответствующий пример легко получить, рассмотрев любое семейство распределений $\{\mathbf{P}_\theta\}$ со скалярным параметром θ и векторной (размерности, большей единицы) минимальной достаточной статистикой S. В этом случае оценка м. п. $\widehat{\theta}^*$ будет также скалярной, так что σ-алгебра $\sigma(S)$ будет более богатой, чем $\sigma(\widehat{\theta}^*)$, и, стало быть, включение $\sigma(S) \subset \sigma(\widehat{\theta}^*)$, вытекающее из минимальности S и достаточности $\widehat{\theta}^*$, невозможно.

Пример 4. Пусть $X \Subset \mathbf{U}_{\theta,1+\theta}$, $\Theta = \mathbb{R}$. Тогда, как мы видели в примере 16.4,

$$f_\theta(X) = \begin{cases} 1 & \text{при } \theta \leqslant \mathrm{x}_{(1)} \leqslant \mathrm{x}_{(n)} \leqslant 1+\theta, \\ 0 & \text{в противном случае,} \end{cases}$$

так что $f_\theta(X)$ зависит от X только через $\mathrm{x}_{(1)}$ и $\mathrm{x}_{(n)}$. Это означает, что $S = (\mathrm{x}_{(1)}, \mathrm{x}_{(n)})$ есть достаточная статистика. Ни одна из величин $\mathrm{x}_{(1)}$, $\mathrm{x}_{(n)}$ в отдельности не является достаточной статистикой, на что указывают следующие соотношения:

$$\mathbf{P}(\mathrm{x}_{(1)} \geqslant u, \mathrm{x}_{(n)} < v) = \prod_{i=1}^{n} \mathbf{P}(\mathrm{x}_i \in [u, v)) = (v-u)^n$$

при $u \geqslant \theta$, $v \leqslant 1+\theta$, $v > u$.

Поэтому совместная плотность распределения $(\mathrm{x}_{(1)}, \mathrm{x}_{(n)})$ будет равна

$$g(u,v) = \begin{cases} n(n-1)(v-u)^{n-2} & \text{при } u \geqslant \theta,\ v \leqslant 1+\theta,\ v > u, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Далее, $\mathbf{P}(\mathrm{x}_{(1)} \geqslant u) = (1+\theta-u)^n$ при $\theta \leqslant u \leqslant 1+\theta$, так что плотность $\mathrm{x}_{(1)}$ равна

$$g(u) = n(1+\theta-u)^{n-1} \quad \text{при} \quad \theta \leqslant u \leqslant 1+\theta.$$

Отсюда уже легко получить, что условная плотность $g(v/u)$ величины $\mathrm{x}_{(n)}$ при условии $\mathrm{x}_{(1)} = u$ (а стало быть, и соответствующее условное распределение) будет зависеть от θ. Это означает, что $\mathrm{x}_{(1)}$ (как и $\mathrm{x}_{(n)}$) в отдельности не являются достаточными статистиками. Так как в качестве о. м. п. $\widehat{\theta}^*$ мы можем взять $\widehat{\theta}^* = \mathrm{x}_{(1)}$ (см. пример 16.4), то мы доказали, что для семейства $\mathbf{U}_{\theta,1+\theta}$ о. м. п. $\widehat{\theta}^*$ не является достаточной статистикой.

Предлагаем читателю с помощью теоремы 1 убедиться, что $S{=}(\mathrm{x}_{(1)}, \mathrm{x}_{(n)})$ является минимальной достаточной статистикой для $\mathbf{U}_{\theta,1+\theta}$.

Условие о достаточности $\widehat{\theta}^*$ в теореме 2 будет автоматически выполнено, если предположить, что существуют скалярная (для одномерного θ) достаточная статистика S_0, для которой функция φ в равенстве $\widehat{\theta}^* = \varphi(S_0)$ будет взаимно однозначной (т. е. $\widehat{\theta}^*$ и S_0 будут эквивалентными).

§ 24. Построение эффективных оценок с помощью достаточных статистик. Полные статистики

Определение 1. Оценка θ^* называется *достаточной*, если она является достаточной статистикой.

1. Одномерный случай. Будем предполагать здесь, что θ — скалярный параметр. Пусть K_b есть класс всех оценок θ^* со смещением $b(\theta)$, так что $\theta^* \in K_b$, если $a(\theta) = \mathbf{E}_\theta \theta^* = \theta + b(\theta)$. Для $\theta^* \in K_b$ имеем

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \mathbf{E}_\theta(\theta^* - a(\theta))^2 + (a(\theta) - \theta)^2 = \mathbf{D}_\theta \theta^* + b^2(\theta).$$

Индекс θ у символов $\mathbf{E}_\theta$, $\mathbf{D}_\theta$ будем в этом параграфе иногда опускать.

Следующее утверждение было получено независимо Блекуэлом, Рао и Колмогоровым.

Теорема 1. *Пусть S — достаточная статистика, $\theta^* \in K_b$. Тогда функция $\theta^*_S = \mathbf{E}_\theta(\theta^*/S)$ является оценкой, обладающей свойствами:*

1) $\theta^*_S \in K_b$,

2) *θ^*_S зависит от выборки лишь через $S(X)$,*

3) $\mathbf{E}_\theta(\theta^*_S - \theta)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta)^2$ *при всех* θ.

Последнее неравенство превращается в равенство, если только $\theta^* = \theta^*_S$ п. в. относительно $\mathbf{P}_\theta$.

Другими словами, в классе K_b применение к θ^* операции $\mathbf{E}_\theta(\cdot/S)$ равномерно улучшает оценку θ^*.

Доказательство. Тот факт, что θ_S^* является оценкой, означает, что θ_S^* не зависит от θ и является измеримой функцией от X. Независимость от θ вытекает из свойств достаточных статистик, так как распределение X при фиксированном S от θ не зависит ($\mathbf{E}_\theta(\theta^*/S)$ для произвольной статистики S, вообще говоря, от θ зависит). В то же время в силу свойств у. м. о. θ_S^* есть измеримая функция от S, а стало быть, и от X. Таким образом, θ_S^* является оценкой, удовлетворяющей свойству 2 теоремы.

Равенство

$$\mathbf{E}_\theta\theta_S^* = \mathbf{E}_\theta\mathbf{E}_\theta(\theta^*/S) = \mathbf{E}_\theta\theta^*,$$

доказывающее, что $\theta_S^* \in K_b$, также следует непосредственно из свойств у. м. о. Далее,

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \mathbf{E}_\theta\big(\theta^* - \theta \pm \theta_S^*\big)^2 =$$
$$= \mathbf{E}_\theta\big(\theta_S^* - \theta\big)^2 + \mathbf{E}_\theta\big(\theta^* - \theta_S^*\big)^2 + 2\mathbf{E}_\theta\big(\theta_S^* - \theta\big)\big(\theta^* - \theta_S^*\big).$$

Используя опять свойства у. м. о., получим

$$\mathbf{E}_\theta\big(\theta_S^* - \theta\big)\big(\theta^* - \theta_S^*\big) = \mathbf{E}_\theta\mathbf{E}_\theta\big[\big(\theta_S^* - \theta\big)\big(\theta^* - \theta_s^*\big)/S\big] =$$
$$= \mathbf{E}_\theta\big[\big(\theta_S^* - \theta\big)\mathbf{E}_\theta\big(\theta^* - \theta_S^*/S\big)\big] = 0,$$

и, следовательно,

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \mathbf{E}_\theta\big(\theta_S^* - \theta\big)^2 + \mathbf{E}_\theta\big(\theta^* - \theta_S^*\big)^2. \quad \triangleleft$$

Собственно, неравенство 3) теоремы 1 можно получить непосредственно из свойства у. м. о. $(\mathbf{E}(\xi/S))^2 \leqslant \mathbf{E}(\xi^2/S)$, так как тогда

$$\big(\theta_S^* - \theta\big)^2 = [\mathbf{E}_\theta((\theta^* - \theta)/S)]^2 \leqslant \mathbf{E}_\theta[(\theta^* - \theta)^2/S],$$
$$\mathbf{E}_\theta\big(\theta_S^* - \theta\big)^2 \leqslant \mathbf{E}(\theta^* - \theta)^2.$$

Факт, изложенный в теореме 1, можно интерпретировать следующим образом. Пусть S и T — две достаточные статистики, $\theta^* = \varphi(T)$ и S подчинена T, тогда $\mathbf{E}_\theta\big(\theta_S^* - \theta\big)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta)^2$.

Другими словами, чем «экономнее» достаточная статистика S (или чем беднее соответствующая σ-алгебра), тем оценки θ_S^* лучше. Таким образом, для построения оптимальных оценок мы должны искать *минимальные* достаточные статистики (или минимальные достаточные σ-алгебры). При этом в качестве исходных оценок θ^* могут фигурировать и «плохие» оценки, не обладающие, например, даже свойством состоятельности. Поучительным в этом отношении является следующий

Пример 1. Пусть $X \Subset \mathbf{\Pi}_\lambda$. Оценка $\lambda^* = \mathrm{x}_1$, очевидно, является несмещенной $\mathbf{E}\lambda^* = \mathbf{E}\mathrm{x}_1 = \lambda$ ($b(\lambda) = 0$), но не состоятельной, так как она от n не зависит. Минимальной достаточной статистикой для λ является статистика $S = n\overline{\mathrm{x}} = \sum \mathrm{x}_i$. Из примера 22.1 следует, что условное относительно S

распределение x_1 есть распределение $\mathbf{B}^S_{1/n}$ в схеме Бернулли с вероятностью успеха $1/n$

$$\mathbf{P}(\mathrm{x}_1 = k/S = s) = C_s^k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{s-k}.$$

Следовательно,

$$\lambda_S^* = \mathbf{E}(\mathrm{x}_1/S) = \sum_{k=1}^{S} k C_S^k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{S-k} = \frac{S}{n} = \overline{\mathrm{x}}.$$

В одном из последующих примеров покажем, что $\overline{\mathrm{x}}$ — эффективная оценка.

2. Многомерный случай. Получим аналоги теоремы 1 для многомерного случая, когда θ и θ^* есть векторы из $\mathbb{R}^k$.

Как и в одномерном случае, вектор $b(\theta) = \mathbf{E}_\theta \theta^* - \theta$ будем называть *смещением оценки* θ^* и обозначать через K_b класс всех оценок со смещением b.

Теорема 1А. *Пусть S — достаточная статистика и $\theta^* \in K_b$. Тогда оценка $\theta_S^* = \mathbf{E}_\theta(\theta^*/S)$ обладает свойствами:*

1) $\theta_S^* \in K_b$;

2) *θ_S^* зависит лишь от $S(X)$;*

3) *среднеквадратичное рассеивание θ_S^* не превосходит среднеквадратичного рассеивания θ^*, или, что то же, для любого вектора $a \in \mathbb{R}^k$*

$$\mathbf{E}_\theta\big(\theta_S^* - \theta, a\big)^2 \leqslant \mathbf{E}_\theta(\theta^* - \theta, a)^2. \tag{1}$$

Равенство (при всех a) здесь возможно лишь в случае $\theta^* = \theta_S^*$ п. в. относительно $\mathbf{P}_\theta$.

Доказательство. Первые два утверждения очевидны. Неравенства (1) следуют из теоремы 1, поскольку все сводится к рассмотрению одномерных оценок (θ^*, a) и $\mathbf{E}_\theta[(\theta^*, a)/S] = \big(\theta_S^*, a\big)$ параметра (θ, a). Если в (1) при всех a справедливо равенство, то при каждом a будем иметь $\big(\theta_S^*, a\big) = (\theta^*, a)$ п. в. Это и означает, что $\theta_S^* = \theta^*$ п. в. $\triangleleft$

Таким образом, в многомерном случае роль достаточных статистик та же: квадратичная форма $\sum \sigma_{ij} a_i a_j$, где $\sigma^2 = \|\sigma_{ij}\|$ есть матрица вторых моментов для $\theta_S^* - \theta$, будет тем меньше, чем меньше σ-алгебра $\sigma(S)$, порожденная S.

3. Полные статистики и эффективные оценки. Приведем теперь весьма простой критерий неулучшаемости оценок, основанный на понятии полноты статистики S. Размерность статистики S будем обозначать через l. Обычно она больше или равна размерности k параметра θ.

Для двух измеримых функций $f_1(s)$ и $f_2(s)$: $\mathbb{R}^l \to \mathbb{R}^k$ будем писать $f_1(s) = f_2(s)$ п. в. $[\mathcal{P}]$, где $\mathcal{P}$ — семейство распределений в $(\mathbb{R}^l, \mathfrak{B}^l)$, если $f_1(s) = f_2(s)$ всюду, за исключением множества N такого, что $\mathbf{P}(N) = 0$ для всех $\mathbf{P} \in \mathcal{P}$.

Определение 2. Семейство $\mathcal{G} = \{\mathbf{G}_\theta\}$ распределений в $(\mathbb{R}^l, \mathfrak{B}^l)$, зависящих от k-мерного параметра $\theta \in \Theta \subset \mathbb{R}^k$, называется *полным*, если равенство

$$\int y(s)\mathbf{G}_\theta(ds) = 0 \quad \text{при всех} \quad \theta \in \Theta \tag{2}$$

влечет за собой $y(s) = 0$ п. в. $[\mathcal{G}]$. Уравнение (2) рассматривается в классе функций $y\colon \mathbb{R}^l \to \mathbb{R}^k$, для которых интеграл (2) существует.

Определение 3. Статистика S называется *полной*, если семейство $\mathcal{G}$ ее распределений $\mathbf{G}_\theta$, индуцированных распределением $\mathbf{P}_\theta$ в $\left(\mathcal{X}^n, \mathfrak{B}^n_{\mathcal{X}}\right)$, является полным.

Уравнение (2) для статистик может быть написано в виде $\mathbf{E}_\theta y(S) = 0$ при всех $\theta \in \Theta \subset \mathbb{R}^k$.

Теорема 2. *Статистика S является полной тогда и только тогда, когда при каком-нибудь $b_0(\theta)$ $\sigma(S)$-измеримая*) оценка θ^* единственна в классе $K^S_{b_0}$ всех $\sigma(S)$-измеримых оценок из K_{b_0}.*

Если $\sigma(S)$-измеримая оценка единственна в $K^S_{b_0}$, то $\sigma(S)$-измеримые оценки будут обладать свойством единственности и в любом другом классе K^S_b.

Доказательство этого утверждения почти очевидно, так как наличие в K_{b_0} двух $\sigma(S)$-измеримых оценок $\theta^*_1 = \varphi_1(S)$, $\theta^*_2 = \varphi_2(S)$ означает, что $\int \varphi_i(s)\mathbf{G}_\theta(ds) = b_0(\theta)$, $i = 1, 2$,

$$\int [\varphi_1(s) - \varphi_2(s)]\mathbf{G}_\theta(ds) = 0 \quad \text{при всех} \quad \theta \in \Theta,$$

так что полнота S влечет за собой $\varphi_1(s) = \varphi_2(s)$ п. в. $[\mathcal{G}]$. Наоборот, пусть $\int y(s)\mathbf{G}_\theta(ds) = 0$ для всех $\theta \in \Theta$, $\theta^*_1 = \varphi_1(s) \in K_b$. Тогда $\theta^*_2 = \varphi_1(s) + y(s) \in K_b$, единственность $\sigma(S)$-измеримой оценки означает, что $y(s) = 0$ п. в. $[\mathcal{G}]$. $\triangleleft$

Теорема 3. *Если достаточная статистика S является полной и $\theta^* \in K_b$, то оценка $\theta^*_S = \mathbf{E}_\theta(\theta^*/S)$ является единственной эффективной оценкой в K_b.*

Эта теорема дает нам достаточно простые критерии эффективности оценок.

Доказательство. В силу теоремы 2 $\sigma(S)$-измеримая оценка в классе K^S_b единственна.

Пусть θ^{**} — любая другая оценка из K_b. Тогда $\theta^{**}_S = \mathbf{E}_\theta(\theta^{**}/S) \in K^S_b$ и, стало быть, $\theta^{**}_S = \theta^*_S$ п. в. $[\mathcal{G}]$. Отсюда и из теоремы 1 следует, что

$$\mathbf{E}_\theta\left(\theta^*_S - \theta\right) = \mathbf{E}_\theta\left(\theta^{**}_S - \theta\right)^2 \leqslant \mathbf{E}_\theta(\theta^{**} - \theta)^2,$$

и равенство возможно лишь при $\theta^{**} = \theta^*_S$ п. н. $\triangleleft$

---

*) То есть измеримая относительно σ-алгебры $\sigma(S)$, порожденной S, и, стало быть, представимая в виде $\varphi(S)$, где φ — борелевская функция.

Следствие 1. *Если S — достаточная полная статистика, а θ^* — несмещенная оценка, то θ^*_S есть эффективная оценка и она единственна.*

Пример 2. В примере 1 с распределениями Пуассона мы получили, что для $\lambda^* = \mathrm{x}_1$

$$\lambda^*_S = \mathbf{E}_\lambda(\mathrm{x}_1/S) = \overline{\mathrm{x}},$$

где $S = n\overline{\mathrm{x}}$. Покажем, что S — полная статистика и, стало быть, $\overline{\mathrm{x}}$ — эффективная оценка. Уравнение (2) для статистики S имеет вид

$$\sum_{k=0}^{\infty} y(k) e^{-n\lambda} \frac{(n\lambda)^k}{k!} = 0 \quad \text{при всех} \quad \lambda \geqslant 0$$

или, что то же,

$$v(z) = \sum y(k) \frac{z^k}{k!} = 0 \quad \text{при всех} \quad z \geqslant 0. \tag{3}$$

Это, очевидно, влечет за собой $y(k) = 0$, так как из сходимости ряда (3), скажем, при $z = 1$ следует, что $v(z)$ аналитична при $|z| < 1$ и тождественно равна нулю. Стало быть, коэффициенты $y(k)$ ее разложения в ряд равны нулю.

Пример 3. Пусть $X \in \mathbf{U}_{0,\theta}$. Покажем, что статистика $S = \mathrm{x}_{(n)} = \max\limits_{i \leqslant n} \mathrm{x}_i$ является полной. Достаточность (и минимальность) S была установлена в примере 23.2. Распределение S определяется равенством

$$\mathbf{P}(S < s) = (s/\theta)^n, \quad 0 \leqslant s \leqslant \theta,$$

так что S имеет плотность, равную $ns^{n-1}\theta^{-n}$ при $s \in [0, \theta]$. Уравнение (2) в этом случае имеет вид

$$\int_0^\theta y(s) \frac{ns^{n-1}}{\theta^n}\, ds = 0 \quad \text{при} \quad \theta \in (0, \infty).$$

Из равенства $\int\limits_0^\theta y(s) s^{n-1}\, ds = 0$ при всех θ следует, очевидно, что $y(s)s^{n-1} = 0$, $y(s) = 0$ п. в.

Мы предлагаем читателю проверить, являются ли полными достаточные статистики, найденные методами § 22, для других параметрических семейств. В частности, установить, что $\alpha^* = \dfrac{1}{\overline{\mathrm{x}}}\left(1 - \dfrac{1}{n}\right)$ есть единственная эффективная оценка параметра α семейства $\mathbf{\Gamma}_{\alpha,1}$ (см. § 12).

Замечание 1. Во всех приведенных выше утверждениях и примерах молчаливо предполагалось, что множество возможных значений оценок θ^* принадлежит Θ. Если это не так, то в изложение надо внести некоторые коррективы. Проиллюстрируем сказанное на примере 2, в котором мы показали, что $\lambda^* = \overline{\mathrm{x}}$ является эффективной оценкой параметра λ для распределения Пуассона $\mathbf{\Pi}_\lambda$. Рассмотрим теперь простейшую модификацию этого

примера. Пусть по-прежнему надо оценить параметр λ распределения Пуассона, но нам даны в качестве наблюдений не значения $z_i \in \Pi_\lambda$, а «зашумленные» значения $x_i = z_i + y_i$, где $y_i \in \Pi_1$, y_i не зависят от z_i. Это значит, что мы должны в этом случае оценить λ по выборке $X \in \Pi_{\lambda+1}$. Как и в примере 2, здесь полной достаточной статистикой является $\bar{x}$. Однако при этом мы не можем воспользоваться конструкцией примера 2. Действительно, там в качестве основы для построения эффективной оценки параметра $\theta = \lambda + 1$ мы использовали несмещенную оценку $\theta^* = x_1$. В нашем случае аналогичная оценка могла бы иметь вид $\lambda^* = x_1 - 1$. Но это невозможно, так как λ^* с вероятностью $e^{-\lambda-1}$ принимает «недопустимое» значение -1, которое естественно заменить на ближайшее допустимое значение, равное нулю, так что «подправленная» оценка имеет вид $\lambda^* = \max(0, x_1 - 1)$. Эта оценка имеет положительное смещение, равное $e^{-\lambda-1}$ и не годится для построения эффективной оценки. Более естественно подправлять не исходную оценку, а эффективную оценку в примере 2, взяв в качестве оценки $\lambda^* = \max(0, \bar{x} - 1)$. Эта оценка также не является несмещенной, но смещение ее при больших n очень мало *).

К тому же ее качество по сравнению с «недопустимой» оценкой $\bar{x} - 1$ может только улучшиться, так как

$$\mathbf{E}_\lambda(\lambda^* - \lambda)^2 = \mathbf{E}_\lambda((\bar{x} - 1 - \lambda)^2;\ \bar{x} \geqslant 1) + \mathbf{E}_\lambda(\lambda^2;\ \bar{x} < 1) \leqslant$$

$$\leqslant \mathbf{E}_\lambda(\bar{x} - 1 - \lambda)^2 = \frac{\lambda + 1}{n}.$$

Смещение оценки λ^* равно

$$\sum_{k=1}^{n} \mathbf{P}_\lambda(\bar{x} - 1 = -k/n)\frac{k}{n} < \mathbf{P}(n\bar{x} < n) =$$

$$= \mathbf{P}(-n(\bar{x} - \lambda - 1) > \lambda n) < e^{-n\lambda}(\lambda + 1)^n.$$

Последнее неравенство получается, если воспользоваться неравенством Чебышева для сумм $S_n = -n(\bar{x} - \lambda - 1)$ центрированных случайных величин и тем, что $\mathbf{E}_\lambda e^{-tn\bar{x}} = e^{n(\lambda+1)(e^{-t}-1)}$ (см. также [17, § 8.8]). Таким образом, смещение λ^* экспоненциально убывает при $n \to \infty$ и пренебрежимо мало по сравнению с «разбросом» оценки, имеющим порядок $1/\sqrt{n}$.

---

*) Несмещенных оценок в этом случае вообще не существует. Пусть для простоты $n = 1$. Существование несмещенной оценки для λ означало бы существование функции $\varphi \geqslant 0$ такой, что

$$\lambda^* = \varphi(x_1), \quad \sum_{k=0}^{\infty} \varphi(k) e^{-(\lambda+1)} \frac{(\lambda+1)^k}{k!} = \lambda,$$

при всех $\lambda > 0$. Полагая $\lambda + 1 = \theta$, последнее соотношение запишем в виде

$$\sum_{k=0}^{\infty} \varphi(k)\frac{\theta^k}{k!} = e^\theta(\theta - 1) = -1 + \sum_{k=1}^{\infty} \theta^k \left(\frac{1}{(k-1)!} - \frac{1}{k!}\right).$$

Из этого равенства двух степенных рядов следует с необходимостью $\varphi(0) = -1$, что противоречит предположению о существовании несмещенной оценки $\lambda^* = \varphi(x_1) \geqslant 0$.

Аналогичным образом дело будет обстоять и в других примерах, поскольку замена значений θ^*, вышедших за пределы допустимой области Θ, на подходящие значения, лежащие на границе этой области, будут приводить лишь к уменьшению разброса исходной «недопустимой» оценки θ^* около истинного значения. Величины смещений при этом будут экспоненциально малы.

Отметим теперь, что теорема 3 указывает на существование связей между понятиями полноты и минимальности. На этот счет справедливо следующее утверждение, дающее вместе с теоремами § 23 критерий минимальности достаточных статистик.

Теорема 4. *Всякая полная достаточная статистика S является минимальной достаточной статистикой.*

Доказательство. Пусть $\mathfrak{A}_0$ — минимальная достаточная σ-алгебра (по теореме 23.1 такая существует). Предположим, что $\mathbf{E}_\theta S$ существует, и рассмотрим функцию $\psi = S - \mathbf{E}_\theta(S/\mathfrak{A}_0)$. Так как $\mathfrak{A}_0 \subset \sigma(S)$, то ψ будет $\sigma(S)$-измеримой, так что $\psi = \psi(S)$. Обозначим через $\mathbf{G}_\theta$ распределение S. Тогда при всех θ, очевидно, $\mathbf{E}_\theta\psi(S) = 0$ или, что то же,

$$\int \psi(s)\mathbf{G}_\theta(ds) = 0 \quad \text{при всех} \quad \theta \in \Theta.$$

Отсюда в силу полноты S следует, что $\psi(s) = 0$ п. в. $[\mathcal{G}]$, $\mathcal{G} = \{\mathbf{G}_\theta\}$. Это означает, что $S = \mathbf{E}_\theta(S/\mathfrak{A}_0)$ п. в. $[\mathcal{G}]$ и, стало быть, S измерима относительно*) $\mathfrak{A}_0$, $\sigma(S) = \mathfrak{A}_0$.

Если $\mathbf{E}_\theta S$ не существует, надо вместо S рассмотреть статистику $\operatorname{arctg} S$, которая, очевидно, эквивалентна S относительно свойств достаточности, полноты и минимальности. ◁

Отметим, что обратное утверждение не верно: *минимальная достаточная статистика не обязана быть полной.* Соответствующие примеры легко получаются в случаях, когда размерность l статистики больше, чем размерность k параметра θ. Например, в § 23 мы видели, что совместная плотность минимальной достаточной статистики $S = (\mathrm{x}_{(1)}, \mathrm{x}_{(n)})$ для семейства $\mathbf{U}_{\theta,1+\theta}$ равна

$$g_\theta(u,v) = \begin{cases} n(n-1)(v-u)^{n-2} & \text{при } u \geqslant \theta,\ v \leqslant 1+\theta,\ v > u, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Если взять функцию $y(u,v) = \varphi(v-u)$ и сделать ортогональное преобразование $(v-u)/\sqrt{2} = t$, $(v+u)/\sqrt{2} = z$, то интеграл в (2) (по треугольнику $u \geqslant \theta$, $v \leqslant 1+\theta$, $v > u$) будет равен

$$\int y(u,v)g_\theta(u,v)\,du\,dv = n(n-1)\int_0^1 \varphi(x)x^{n-2}(1-x)\,dx.$$

Очевидно, что интеграл в правой части от θ не зависит и легко подобрать функцию $\varphi(x) \not\equiv 0$, обращающую его в нуль.

*) Под $\mathfrak{A}_0$ здесь надо понимать σ-алгебру, пополненную множествами N, для которых $\mathbf{P}_\theta(N) = 0$ при всех θ

§ 25. Экспоненциальное семейство

Пусть $\theta = (\theta_1, \ldots, \theta_k)$ — k-мерный параметр и плотность $f_\theta(x)$ представима в виде

$$f_\theta(x) = h(x) \exp\left\{\sum_{j=1}^{k} a_j(\theta) U_j(x) + V(\theta)\right\}, \tag{1}$$

где все функции, входящие в правую часть, конечны и измеримы.

Определение 1. Семейства распределений $\{\mathbf{P}_\theta\}$ с плотностью такого вида будем называть *экспоненциальными семействами* и обозначать их символом $\mathcal{E}$.

Чтобы сделать представление (1) по возможности более однозначным, предположим, что *функции* $a_0(\theta) \equiv 1$, $a_1(\theta), \ldots, a_k(\theta)$ *линейно независимы на* Θ.

Как мы увидим, экспоненциальные семейства занимают особое место среди параметрических семейств распределений, так как для них многие общие конструкции математической статистики можно реализовать в явном виде.

Иногда экспоненциальными семействами называют семейства распределений более частного вида*), когда $a_j(\theta) = \theta_j$.

К эскпоненциальным семействам относятся, например, семейства распределений $\{\mathbf{\Phi}_{\alpha,\sigma^2}\}$, $\{\mathbf{\Pi}_\lambda\}$, $\{\mathbf{B}_p\}$, $\{\mathbf{\Gamma}_{\alpha,\lambda}\}$ и ряд других.

Пример 1. Рассмотрим распределение $\mathbf{\Gamma}_{\alpha,\lambda}$. Плотность $\gamma_{\alpha,\lambda}(x)$ этого распределения можно представить в виде

$$\gamma_{\alpha,\lambda}(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} \exp(-\alpha x) = x^{-1} \exp\left\{\lambda \ln x - \alpha x + \ln \frac{\alpha^\lambda}{\Gamma(\lambda)}\right\}, \quad x > 0,$$

так что здесь можно положить

$$h(x) = \begin{cases} x^{-1}, & x > 0, \\ 0, & x \leqslant 0, \end{cases}$$

$$U_1(x) = \ln x, \quad U_2(x) = x, \quad V(\alpha, \lambda) = \ln \frac{\alpha^\lambda}{\Gamma(\lambda)},$$

$$a_1(\alpha, \lambda) = \lambda, \quad a_2(\alpha, \lambda) = -\alpha. \quad \triangleleft$$

Функция правдоподобия для $X \in \mathbf{P} \in \mathcal{E}$ равна

$$f_\theta(X) = \exp\{(a(\theta), S) + nV(\theta)\} \prod_{i=1}^{n} h(\mathrm{x}_i),$$

*) По существу это то же самое: мы придем к частному виду, если сделаем взаимно однозначное преобразование $\gamma = \gamma(\theta)$, $\gamma = (\gamma_1, \ldots, \gamma_k)$ над параметром θ, положив $\gamma_j = a_j(\theta)$.

где

$$a(\theta) = (a_1(\theta), \ldots, a_k(\theta)), \quad S = (S_1, \ldots, S_k),$$

$$S_j = S_j(X) = \sum_{i=1}^{n} U_j(\mathrm{x}_i),$$

(a, S) есть скалярное произведение. Отсюда и из теоремы 22.1 следует, что S есть достаточная статистика для θ. Мы докажем, что S является минимальной достаточной статистикой.

Так как функции $a_j(\theta)$, $U_j(x)$, $V(\theta)$ конечны, экспонента в (1) всегда положительна. Это означает, что в качестве распределения $\mathbf{Q}$ в теореме 23.1 (при котором все $\mathbf{P}_\theta$ абсолютно непрерывны относительно $\mathbf{P}_Q = \int \mathbf{P}_t \mathbf{Q}(dt)$) можно взять распределение, сосредоточенное в любой фиксированной точке θ^0. Поэтому из теоремы 23.1 вытекает, что σ-алгебра $\mathfrak{A}_0$, порожденная функцией

$$r(X, \theta) = \frac{f_\theta(X)}{f_{\theta^0}(X)} = \exp\{(a(\theta) - a(\theta^0), S) + n(V(\theta) - V(\theta^0))\},$$

является минимальной достаточной σ-алгеброй.

Теорема 1. *Статистика S является минимальной достаточной статистикой.*

Доказательство. Из линейной независимости функций $1, a_1(\theta), \ldots, a_k(\theta)$ на Θ следует линейная независимость $a_1(\theta) - a_1(\theta^0), \ldots, a_k(\theta) - a_k(\theta^0)$. Это значит, что в Θ найдутся k точек $\theta^1, \ldots, \theta^k$ таких, что значения $a_{ij} = a_i(\theta^j) - a_i(\theta^0)$ образуют матрицу A с определителем, отличным от нуля. Это означает, в свою очередь, что уравнения

$$(a(\theta^j) - a(\theta^0), S) = \ln r(X, \theta^j) - n(V(\theta^j) - V(\theta^0)), \quad j = 1, \ldots, k,$$

однозначно разрешимы относительно S и, стало быть, $\sigma(S) \subset \sigma(r(X, \theta_j); j = 1, \ldots, k) \subset \mathfrak{A}_0$. $\triangleleft$

В примере 1 мы рассмотрели Г-распределение и установили, что для него справедливо представление (1) при $\theta = (\alpha, \lambda)$ с функциями

$$U_1(x) = \ln x, \quad U_2(x) = x,$$
$$a_1(\alpha, \lambda) = \lambda, \quad a_2(\alpha, \lambda) = -\alpha.$$

Очевидно, что условия теоремы 1 выполнены, и статистика $S = \left(\sum \ln \mathrm{x}_i, \sum \mathrm{x}_i\right)$ или, что то же, статистика $\left(\prod \mathrm{x}_i, \sum \mathrm{x}_i\right)$ является минимальной достаточной статистикой.

Если условия теоремы 1 несколько усилить, то статистика S будет полной достаточной статистикой (в этом случае минимальность S можно было бы получить как следствие полноты).

Теорема 2. *Пусть $X \in \mathbf{P} \in \mathcal{E}$. Если функция a и множество Θ таковы, что $a(\theta)$ зачеркивает k-мерный параллелепипед, когда θ пробегает Θ, то S — полная достаточная статистика.*

Условия теоремы относительно параллелепипеда, очевидно, будут выполнены, если множество Θ «телесно», т. е. содержит внутренние точки (а вместе с ними и сферы в $\mathbb{R}^k$ достаточно малого радиуса), а функции $a_j(\theta)$ в окрестности какой-нибудь «телесной» точки θ^0 являются линейно независимыми и гладкими. Тогда преобразование $a = a(\theta)$ переводит окрестность точки θ^0 в телесное множество.

Очевидно, что пример 1 с Г-распределением удовлетворяет условиям теоремы 2, так что статистика $\left(\prod \mathrm{x}_i, \sum \mathrm{x}_i\right)$ является полной.

Столь же просто читатель может проверить, что для нормального распределения Φ_{α,σ^2} статистика $\left(\sum \mathrm{x}_i, \sum \mathrm{x}_i^2\right)$ является также полной достаточной статистикой.

Доказательство теоремы 2. В нашем случае функции $\psi(s,\theta)$ и $h(\boldsymbol{x})$ в факторизационной теореме Неймана–Фишера равны

$$\psi(s,\theta) = \exp\{(a(\theta), s) + nV(\theta)\},$$

$$h(\boldsymbol{x}) = \prod_{i=1}^{n} h(x_i).$$

Рассмотрим на $(\mathbb{R}^k, \mathfrak{B}^k)$ не зависящую от θ меру

$$\nu(B) = \int\limits_{S^{-1}(B)} h(\boldsymbol{x})\mu^n(d\boldsymbol{x}),$$

где $S^{-1}(B)$ есть множество всех x, для которых $S(\boldsymbol{x}) \in B$.

Выделим в виде лемм следующие два вспомогательных утверждения.

Лемма 1. *Распределение $\mathbf{G}_\theta(B) = \mathbf{P}_\theta(S(X) \in B)$ статистики S абсолютно непрерывно относительно ν и имеет плотность в точке s, равную $\psi(s,\theta)$.*

Доказательство следует из равенства

$$\mathbf{G}_\theta(B) = \int\limits_{S(\boldsymbol{x})\in B} \psi(S(\boldsymbol{x}),\theta)h(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \int\limits_{s\in B} \psi(s,\theta)\nu(ds),$$

которое является следствием замены переменных. $\triangleleft$

Лемма 2. *Пусть $\mathbf{G}_1$ и $\mathbf{G}_2$ — две σ-конечные меры в $(\mathbb{R}^k, \mathfrak{B}^k)$. Тогда если $\int e^{(a,u)}\mathbf{G}_1(du) = \int e^{(a,u)}\mathbf{G}_2(du)$ существуют для всех a из некоторого параллелепипеда I в $\mathbb{R}^k$, то $\mathbf{G}_1 = \mathbf{G}_2$.*

Доказательство. Чтобы упростить рассуждения, рассмотрим одномерный случай $k = 1$ и будем считать, что $I = \{x : |x| \leqslant \alpha\}$. Тогда

$$h_j(a) = \int e^{au}\mathbf{G}_j(du), \quad j = 1, 2,$$

есть аналитические функции при $|a| < \alpha$. Кроме того, при всех $b \in \mathbb{R}$ определены функции $h_j(z) = \int e^{(a+ib)u}\mathbf{G}_j(du)$ комплексного переменного $z = a + ib$. Ясно, что $h_j(z)$ будут аналитическими в полосе $|a| < \alpha$, $-\infty < b < \infty$. Так как $h_1(z) = h_2(z)$ на отрезке прямой $b = 0$, $|a| < \alpha$, то $h_1(z) = h_2(z)$ при всех z из указанной полосы. Стало быть,

$$\int e^{ibu}\mathbf{G}_1(du) = \int e^{ibu}\mathbf{G}_2(du). \tag{2}$$

Заметим, что так как $h_j(0) = \int \mathbf{G}_j(du) < \infty$, то $\mathbf{G}_j$ можно считать вероятностными мерами. Из теоремы о взаимно однозначном соответствии между характеристическими функциями и распределениями (см. [17, § 7.2]) и из (2) следует, что $\mathbf{G}_1 = \mathbf{G}_2$.

Если параллелепипед I имеет вид $\{x : |x - \alpha_0| \leqslant \alpha\}$, то следует перейти к мерам $\mathbf{G}_j^*(du) = e^{\alpha_0 u}\mathbf{G}_j(du)$.

В многомерном случае $k > 1$ доказательство проходит точно так же. ◁

Мы можем перейти теперь непосредственно к доказательству теоремы 2.

Нам надо доказать, что если φ — измеримая функция в $(\mathbb{R}^k, \mathfrak{B}^k)$ и существует

$$\int \varphi(s)\mathbf{G}_\theta(ds) = 0 \quad \text{при всех} \quad \theta \in \Theta, \tag{3}$$

то $\varphi(s) = 0$ п. в. $[\mathcal{G}]$, $\mathcal{G} = \{\mathbf{G}_\theta\}_{\theta\in\Theta}$. Пусть $\varphi = \varphi^+ - \varphi^-$, где $\varphi^\pm \geqslant 0$. Тогда из (3) следует $\int \varphi^+(s)\mathbf{G}_\theta(ds) = \int \varphi^-(s)\mathbf{G}_\theta(ds)$ или в силу леммы 1

$$\int \varphi^+(s)\psi(s,\theta)\nu(ds) = \int \varphi^-(s)\psi(s,\theta)\nu(ds),$$

$$\int \varphi^+(s)e^{(s,a(\theta))}\nu(ds) = \int \varphi^-(s)e^{(s,a(\theta))}\nu(ds).$$

Если образовать σ-конечные меры $\nu^\pm(ds) = \varphi^\pm(s)\nu(ds)$, получим

$$\int e^{(s,a)}\nu^+(ds) = \int e^{(s,a)}\nu^-(ds)$$

при всех a из некоторого параллелепипеда в $\mathbb{R}^k$. Остается воспользоваться леммой 2. ◁

Следствие 1. *Если* $X \Subset \mathbf{P} \in \mathcal{E}$, $\theta^* \in K_b$, *и выполнены условия теоремы* 2, *то оценка* $\theta_S^* = \mathbf{E}(\theta^*/S)$ *является эффективной оценкой в* K_b.

Предположим теперь, что нам надо оценить параметр $\gamma = \varphi(\theta)$, где φ осуществляет взаимно однозначное непрерывное отображение Θ в Γ, множество Γ телесно в $\mathbb{R}^k$. В этом случае определена непрерывная функция $\theta = \varphi^{-1}(\gamma)$, переводящая Γ в Θ, и согласно теореме 2 статистика $S(X) = \sum_{i=1}^{n} U(\mathrm{x}_i)$ будет полной достаточной статистикой и для параметра γ (см. также следствие 22.3). Следовательно, для γ сохраняется и утверждение следствия 1: если $\gamma^* \in K_{b_1}$ — оценка γ, то $\gamma^*_S = \mathbf{E}(\gamma^* \mid S)$ будет эффективной оценкой в классе K_{b_1}, $b_1(\gamma) = \mathbf{E}_\theta \gamma^* - \gamma = \mathbf{E}_{\varphi^{-1}(\gamma)}\gamma^* - \gamma$.

Если φ — линейная функция и θ^* — эффективная оценка, то $\gamma^* = \varphi(\theta^*)$ будет эффективной оценкой для γ.

Пример 2. Задача восстановления изображения. Рассмотрим следующую простейшую задачу восстановления «дискретного» изображения, представляющую существенный интерес в томографии, астрофизике (см., например, [106]) и в других приложениях. Имеется источник излучения, которому на плоской фотопластине (будем считать ее квадратной) соответствует изображение, описываемое функцией (полем) яркости $\theta(u) = \theta(u_1, u_2)$, $u \in [0,1]^2$. Дискретизируем это изображение, разбивая пластинку на маленькие квадраты $\Delta_{ij} = \left\{u : u_1 \in \left[\frac{i-1}{m}, \frac{i}{m}\right), u_2 \in \left[\frac{j-1}{m}, \frac{j}{m}\right)\right\}$ и введя для описания изображения усредненные яркости $\theta_{i,j} = \int\limits_{\Delta_{ij}} \theta(u)\, du$, $i, j = 1, \ldots, m$. Значения полей $\theta(u)$ и, стало быть, $\{\theta_{ij}\}$ неизвестны. Наблюдателю доступно лишь «зашумленное» и «размытое» («рассеянное») изображение источника. Более точно, будем предполагать, что имеем фотографическое изображение суммарного воздействия двух пуассоновских источников частиц. Первый соответствует изучаемому источнику. Для него «в идеале» каждая частица попадает в квадрат Δ_{ij} с вероятностью $\theta_{ij}/\overline{\theta}$ ($\overline{\theta} = \int \theta(u)\, du$ есть интенсивность источника). Однако из-за присутствия эффекта рассеивания этого не происходит. Реальная интенсивность γ_{ij} попадания частиц в Δ_{ij} определяется не только «идеальной» интенсивностью θ_{ij}, но и интенсивностями соседних квадратов, так что γ_{ij} есть взвешенное среднее

$$\gamma_{ij} = \sum_{k,l} \theta_{k,l} h^{(i,j)}_{k,l}, \quad \sum_{i,j} h^{(i,j)}_{k,l} = 1, \quad i, j = 1\ldots, m. \tag{4}$$

Здесь суммирование в первой сумме ведется по $1 \leqslant k \leqslant m$, $1 \leqslant l \leqslant m$, так что (4) исключает влияние «квадратов», расположенных вне фотопластины. Во многих случаях рассеивание можно считать инвариантным относительно сдвига, т. е. считать, что

$$h^{(i+a,\, j+b)}_{k+a,\, l+b} = h^{(i,j)}_{k,l}$$

в области, не затрагивающей границу.

Второй пуассоновский поток частиц представляет собой «шум»: число частиц второго потока, попавших в Δ_{ij} за единицу времени имеет распределение Пуассона с параметром α_{ij}. Если поток однороден, то $\alpha_{ij} = \alpha$ не зависят от i, j.

Значения α_{ij}, $h_{k,l}^{(i,j)}$ нам известны. В приведенном выше описании мы имеем дело с «расщеплением» и сложением пуассоновских потоков. Если пуассоновский поток частиц с параметром θ $\left(\text{вероятность появления } k \text{ частиц за единицу времени равна } e^{-\theta}\frac{\theta^k}{k!}\right)$, то под «расщеплением» будем понимать разделение этого потока на r «подпотоков» следующим образом: каждая частица независимо от других с вероятностью p_k, $k = 1, \ldots, r$, $\sum p_k = 1$, направляется в k-й подпоток. Полученные подпотоки снова будут пуассоновскими (с интенсивностями θp_k) и независимыми. Действительно, вероятность иметь набор $k_1, \ldots, k_r$, $\sum k_i = k$, частиц соответственно в первом, $\ldots, r$-м подпотоке равна

$$e^{-\theta}\frac{\theta^k}{k!} \cdot \frac{k!}{k_1! \ldots k_r!} p_1^{k_1} \ldots p_r^{k_r} = \prod_{i=1}^{r} e^{-\theta p_i} \frac{(\theta p_i)^{k_i}}{k_i!}.$$

При сложении независимых пуассоновских потоков мы, очевидно, также будем получать в результате пуассоновский поток с интенсивностью, равной сумме интенсивностей слагаемых.

В силу сказанного рассматриваемое фотографическое отображение можно описывать как результат воздействия m^2 независимых пуассоновских потоков частиц, соответствующих квадратам Δ_{ij}, $i, j = 1, \ldots, m$. Интенсивности этих потоков равны соответственно $\lambda_{ij} = \gamma_{ij} + \alpha_{ij}$. Стало быть, количества ν_{ij} частиц, попавших в Δ_{ij} за время n, независимы и будут иметь распределение Пуассона с параметрами $n\lambda_{ij}$,

$$\lambda_{ij} = \gamma_{ij} + \alpha_{ij}.$$

Набор частот $\nu = \{\nu_{ij}\}$ можно рассматривать как достаточную статистику для параметра $\lambda = \{\lambda_{ij}\}$ (построенную по выборке $X = \{\nu_{ij}^{(1)}\}, \ldots, \{\nu_{ij}^{(n)}\}$, где $\nu_{ij}^{(k)}$ — число частиц, попавших в Δ_{ij} за k-ю единицу времени). Функция правдоподобия в этом случае имеет вид

$$f_\lambda(X) = \prod_{k=1}^{n} \prod_{ij} e^{-\lambda_{ij}} \frac{\lambda_{ij}^{\nu_{ij}^{(k)}}}{\nu_{ij}^{(k)}!} = \prod_{ij} e^{-n\lambda_{ij}} \lambda_{ij}^{\nu_{ij}} \prod_{k} \frac{1}{\nu_{ij}^{(k)}!}.$$

Очевидно, что мы имеем дело с экспоненциальным семейством и ν является полной достаточной статистикой для λ при этом $\lambda^* = \nu/n$ будет эффективной оценкой λ.

Если матрица $H = \left\{h_{k,l}^{(ij)}\right\}$ имеет обратную, отображение (4), которое можно записать в виде

$$\lambda = \theta H + \alpha,$$

обратимо: $\theta = (\lambda - \alpha)H^{-1}$ и, стало быть, ν будет полной достаточной статистикой и для θ (см. замечания, предшествующие примеру 2). При этом было бы естественно положить $\theta^* = (\lambda^* - \alpha)H^{-1}$. Однако здесь возникают ограничения, связанные с областью возможных значений $\theta \geqslant 0$. Эти ограничения были рассмотрены на простом примере в замечании 24.1. В соответствии со сказанным в этом замечании мы должны положить

$$\theta^* = [(\lambda^* - \alpha)^+ H^{-1}]^+,$$

где $[\theta]^+$ для $\theta = \{\theta_{ij}\}$ означает «вектор» $\{\theta_{ij}^+\}$, $\theta_{ij}^+ = \max(0, \theta_{ij})$. При этом качество оценки от введения операции $[]^+$ только улучшится (см. замечание 24.1).

С технической стороны наибольшие трудности во всем этом может представлять вычисление обратной матрицы H^{-1}. Если изображение сосредоточено в квадрате $\left[\dfrac{s}{m}, \dfrac{m-s}{m}\right]^2$ ($\theta_{ij} = 0$ при $i \notin [s+1,\ m-s]$ или $j \notin$ $\notin [s+1,\ m-s]$, $s \geqslant 0$) и $h_{k+a,l+b}^{(i+a,j+b)} = h_{k,l}^{i,j} = 0$ при $|i-k| > s$, $|j-l| > s$ (ограниченное однородное рассеивание), то построение обратного преобразования можно осуществить с помощью производящих функций (отметим, что при этом будем иметь дело с $(m-2s)^2$ независимыми параметрами θ_{ij}, $i, j = s+1, \ldots, m-s$, и таким же числом независимых γ_{ij}, $i, j = 1, \ldots, m$). Действительно, положим $h_{k,l}^{(k+a,\,l+b)} = h_{a,b}$,

$$h(y,z) = \sum_{a,b=-s}^{s} h_{a,b} z^a z^b, \quad \theta(y,z) = \sum_{i,j=s+1}^{m-s} \theta_{ij} y^i z^i, \quad \gamma(y,z) = \sum_{i,j}^{m} \gamma_{ij} y^i z^j.$$

Тогда уравнение (4) можно записать в виде свертки

$$\gamma_{ij} = \sum_{k,l=-s}^{s} \theta_{i-k,j-l} h_{k,l},$$

что при переходе к производящим функциям дает

$$\gamma(y,z) = \sum_{i,j=1}^{m} y^i z^i \sum_{k,l=-s}^{s} \theta_{i-k,j-l} h_{k,l} =$$

$$= \sum_{k,l=-s}^{s} h_{k,l} y^k z^l \sum_{i,j=1}^{m} y^{i-k} z^{j-l} \theta_{i-k,j-l} = h(y,z)\theta(y,z), \quad (5)$$

$$\theta(y,z) = \gamma(y,z) h^{-1}(y,z).$$

Разлагая $h^{-1}(y,z)$ в ряд по положительным и отрицательным степеням y и z (в подходящей области изменения y и z)

$$h^{-1}(y,z) = \sum g_{k,l} y^k z^l$$

и приравнивая коэффициенты при одинаковых степенях в (5), получим обратное линейное преобразование

$$\theta_{ij} = \sum_{k,l=1}^{m} \gamma_{k,l} g_{i-k,j-l}, \quad i,j = s+1,\ldots,m-s.$$

Если, например, $h_{-s,-s} > 0$, то $h(y,z)y^s z^s$ образует полином, отличный от нуля в точке $y = z = 0$, так что при достаточно малых y, z

$$h^{-1}(y,z) = y^s z^s [y^s z^s h(y,z)]^{-1} = \sum_{k,l=s}^{\infty} g_{k,l} y^k z^l.$$

Если теперь воспользоваться теми же, что и прежде, правилами построения оценок, получим

$$\theta^*_{ij} = \left[\sum_{k,l=1}^{m} \left(\lambda^*_{k,l} - \alpha_{k,l}\right)^+ g_{i-k,j-l} \right]^+, \quad \lambda^*_{k,l} = \frac{\nu_{k,l}}{n}.$$

Ясно, что во всех этих рассмотрениях для получения достаточно хороших оценок нужно, чтобы частоты ν_{kl} были велики или, что то же, чтобы время наблюдения n было велико по сравнению с размерностью m^2 (или $(m-2s)^2$) неизвестного параметра $\theta = \{\theta_{ij}\}$ (считаем, что $\bar{\theta}$ сравнимо с единицей).

Точность оценки поля $\theta = \{\theta_{ij}\}$ можно значительно увеличить, если уменьшить размерность оцениваемого параметра. Это можно сделать, если нам заранее известна возможная форма поля θ_{ij} (или $\theta(u)$). Если, например, θ_{ij} есть «одновершинное» поле вида

$$\theta_{ij} = c \exp\{b_1(i-a_1)^2 + b_2(j-a_2)^2\},$$

то дело сводится к оценке всего лишь пяти (а не m^2) неизвестных параметров c, a_1, a_2, b_1, b_2.

§ 26. Неравенство Рао–Крамера и R-эффективные оценки

1. Неравенство Рао–Крамера и его следствия. Результаты предыдущих параграфов дали нам ряд критериев эффективности оценок. Однако эти критерии носили в известном смысле качественный характер. В этом параграфе мы продолжим изучение вопроса об эффективных оценках, но с несколько иной точки зрения. Выясним прежде всего, какова минимальная величина среднеквадратической погрешности, которую можно получить.

Рассмотрим сначала одномерный случай, когда θ — скалярный параметр. Относительно множества Θ будем предполагать для определенности, что это есть конечный или бесконечный интервал, замкнутый или открытый.

Для того чтобы ответить на поставленный вопрос, нам понадобятся условия регулярности на $f_\theta(x)$. Пусть, как и прежде,

$$l(x,\theta) = \ln f_\theta(x), \quad L(X,\theta) = \sum_{i=1}^{n} l(\mathrm{x}_i,\theta), \quad a(\theta) = \mathbf{E}_\theta \theta^* = \theta + b(\theta).$$

Предположим, что выполнено условие

(R) *Функции* $\sqrt{f_\theta(x)}$ *для п. в.* $[\mu]$ *значений* x *непрерывно дифференцируемы по* $\theta \in \Theta$, *интеграл*

$$I(\theta) \equiv \int \frac{\left(f'_\theta(x)\right)^2}{f_\theta(x)} \mu(dx) = \mathbf{E}_\theta [l'(\mathrm{x}_1, \theta)]^2 \tag{1}$$

существует, положителен и непрерывен по θ. (Здесь и в дальнейшем штрих означает дифференцирование по θ.)

Относительно интеграла (1) необходимо отметить следующее. Если x вместе со своей окрестностью не принадлежит носителю $N_{P_\theta} = \{x : f_\theta(x) > 0\}$ распределения $\mathbf{P}_\theta$, то подынтегральная функция $\left(f'_\theta(x)\right)^2 / f_\theta(x)$ превращается в неопределенность типа $0/0$. Условимся считать это отношение равным нулю. Такого же правила будем придерживаться относительно производной $l'(x, \theta) = f'_\theta(x)/f_\theta(x)$ при ее интегрировании. Этих оговорок можно не делать, если с самого начала рассматривать интегралы вида $\mathbf{E}_\theta \varphi(\mathrm{x}_1, \theta)$ лишь по области N_{P_θ}.

Функция $I(\theta)$ носит название *информации Фишера*. Мы уже встречали ее в § 16. Она играет важную роль в математической статистике и будет неоднократно встречаться в дальнейшем. Некоторые свойства функции $I(\theta)$ обсуждаются в § 27.

Если множество Θ компактно, то *непрерывность* $I(\theta)$ *в условиях* (R) *эквивалентна условию*

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta ([l'(\mathrm{x}_1, \theta)]^2; |l'(\mathrm{x}_1, \theta)| > N) \to 0$$

при $N \to \infty$, которое можно назвать равномерной сходимостью интеграла $I(\theta)$ (см. приложение VI).

Имеет место следующее неравенство для дисперсии оценок θ^* со смещением b.

Теорема 1 (неравенство Рао–Крамера). *Если* $\theta^* \in K_b$, *выполнено условие* (R) *и* $\mathbf{E}_\theta (\theta^*)^2 < c < \infty$, *то*

$$\mathbf{D}_\theta \theta^* \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)}. \tag{2}$$

Если в этом неравенстве достигается равенство на некотором отрезке $\theta \in [\theta_1, \theta_2] \subset \Theta$ *и* $\mathbf{D}_\theta \theta^* > 0$ *на этом отрезке, то функция правдоподобия* $f_\theta(X)$ *при* $\theta \in [\theta_1, \theta_2]$ *представима в виде*

$$f_\theta(X) = \exp\{\theta^* A(\theta) + B(\theta)\} h(X), \tag{3}$$

где $A(\theta)$, $B(\theta)$ *от* X *не зависят.*

Обратно, если $\theta^* = \mathrm{const}$ *или если справедливо представление* (3), *то в неравенстве* (2) *достигается равенство.*

Условие (3), очевидно, означает принадлежность распределения в $\mathcal{X}^n$ с плотностью $f_\theta(\boldsymbol{x})$ к экспоненциальному семейству $\mathcal{E}$.

Следствие 1. *При выполнении условий теоремы* 1

$$\mathbf{E}_\theta(\theta^* - \theta)^2 \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)} + b^2(\theta).$$

Для любой несмещенной оценки θ^*

$$\mathbf{E}_\theta(\theta^* - \theta)^2 \geqslant \frac{1}{nI(\theta)}.$$

Таким образом, в классах K_b наименьшее возможное значение среднеквадратичных уклонений отлично от нуля и определяется правыми частями выписанных неравенств.

Замечание 1. По поводу условия $\mathbf{E}_\theta(\theta^*)^2 < c < \infty$ можно заметить, что при $\mathbf{E}_\theta(\theta^*)^2 = \infty$ выполняется $\mathbf{D}_\theta\theta^* = \infty$ и неравенство (2) становится тривиальным. Условие $\mathbf{D}_\theta\theta^* > 0$ в силу (2) можно заменить на $(1+b'(\theta))^2 > 0$.

Замечание 2. Наряду с условием (R) можно указать несколько других условий, обеспечивающих утверждение теоремы 1, которые будут отличаться друг от друга очень мало. Мы остановились на том из них, которое нам будет удобнее в последующих параграфах. Условия несколько иного вида будут приведены в § 32.

Нам понадобится одно вспомогательное утверждение.

Лемма 1. *Пусть выполнено условие* (R) *и* $S = S(X)$ *есть любая статистика, для которой* $\mathbf{E}_\theta S^2 < c < \infty$ *при* $\theta \in \Theta$. *Тогда функция*

$$a_S(\theta) = \mathbf{E}_\theta S = \int S(\boldsymbol{x}) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) \tag{4}$$

непрерывно дифференцируема по θ, *причем*

$$a'_S(\theta) = \int S(\boldsymbol{x}) f'_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \mathbf{E}_\theta S L'(X, \theta). \tag{5}$$

Это утверждение носит технический характер, а его доказательство существенно отяготило бы рассмотрения. Поэтому доказательство леммы 1 отнесено нами в приложение V.

Доказательство теоремы 1. Полагая в (5) $S \equiv 1$, получим $a_S(\theta) \equiv 1$,

$$\mathbf{E}_\theta L' = 0, \quad \mathbf{E}_\theta a(\theta) L' = 0. \tag{6}$$

Используя снова (5) при $S = \theta^*$ и (6), получаем

$$\mathbf{E}_\theta \theta^* L' = a'(\theta), \quad \mathbf{E}_\theta(\theta^* - a(\theta)) L' = a'(\theta). \tag{7}$$

По неравенству Коши–Буняковского

$$(a'(\theta))^2 \leqslant \mathbf{E}_\theta(\theta^* - a(\theta))^2 \mathbf{E}_\theta(L')^2 \tag{8}$$

или, что то же,

$$\mathbf{D}_\theta\theta^* \geqslant \frac{(1 + b'(\theta))^2}{\mathbf{E}_\theta(L')^2}. \tag{9}$$

Так как случайные величины $l_j = l'(\mathrm{x}_j, \theta)$ независимы, одинаково распределены и имеют в силу (6) ненулевое математическое ожидание $\mathbf{E}_\theta l_j = 0$, то $\mathbf{E}_\theta l_i l_j = 0$ при $i \neq j$, $\mathbf{E}_\theta (L')^2 = \mathbf{E}_\theta \Big(\sum\limits_j l_j \Big)^2 = \sum\limits_{i,j} \mathbf{E}_\theta l_i l_j = n\mathbf{E}_\theta l_1^2 =$ $= nI(\theta)$. Вместе с (9) это доказывает неравенство (2).

Докажем второе утверждение теоремы. Для простоты будем считать, что Θ совпадает с $[\theta_1, \theta_2]$ и что $\mathcal{X}$ совпадает с объединением носителей $\mathcal{X} = \bigcup\limits_{t \in \Theta} N_{\mathbf{P}_t}$.

Итак, пусть в (2) (или в (8)) достигается знак равенства. Тогда функция $\mathbf{D}_\theta \theta^* = \dfrac{(a'(\theta))^2}{nI(\theta)}$ вместе с $a'(\theta)$ и $I(\theta)$ непрерывна по θ и

$$(\mathbf{E}_\theta (\theta^* - a(\theta)) L')^2 = \mathbf{E}_\theta (\theta^* - a(\theta))^2 \mathbf{E}_\theta (L')^2.$$

Последнее равенство вместе с (7) при $c(\theta) = \sqrt{\dfrac{nI(\theta)}{\mathbf{D}_\theta \theta^*}}$ дает

$$\mathbf{E}_\theta (c(\theta)(\theta^* - a(\theta)) - L')^2 = nI(\theta) - 2\sqrt{\frac{nI(\theta)}{\mathbf{D}_\theta \theta^*}} \sqrt{\mathbf{D}_\theta \theta^* \cdot nI(\theta)} + nI(\theta) = 0$$

или, что то же,

$$\int (c(\theta)(\theta^* - a(\theta)) - L'(\boldsymbol{x}, \theta))^2 f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = 0, \tag{10}$$

где функции $a(\theta)$, $c(\theta)$ непрерывны.

Пусть $M(t)$ — множество точек $\boldsymbol{x} \in \mathcal{X}^n$, для которых

$$[c(t)(\theta^* - a(t)) - L'(\boldsymbol{x}, t)] \sqrt{f_t(\boldsymbol{x})} = 0. \tag{11}$$

Из (10) следует, что $\mu^n(\overline{M}(t)) = 0$, где $\overline{M}(t)$ — дополнение к $M(t)$. Обозначим $M = \bigcap\limits_{t \in T} M(t)$, где T — счетное всюду плотное на Θ множество. Тогда, очевидно, $\mu^n(\overline{M}) = 0$. Фиксируем точку $\boldsymbol{x} \in M$. Так как $f_t(\boldsymbol{x})$ непрерывна по t и $\mathcal{X}$ совпадает с объединением носителей, найдется интервал $(t_1, t_2) \subset \Theta$, $t_2 > t_1$, на котором $f_t(\boldsymbol{x}) > 0$ и, стало быть (см. (11)),

$$L'(\boldsymbol{x}, t) = c(t)(\theta^* - a(t)) \quad \text{для} \quad t \in T \cap [t_1, t_2]. \tag{12}$$

Обе части этого равенства непрерывны на $[t_1, t_2]$ и, следовательно, это равенство справедливо всюду на (t_1, t_2). Кроме того, правая часть равномерно ограничена на (t_1, t_2). Но это означает, что $L(\boldsymbol{x}, t)$ также равномерно ограничена, а $f_t(\boldsymbol{x})$ равномерно положительно всюду на (t_1, t_2). Это противоречит допущению о существовании интервала (t_1, t_2), на котором $f_t(\boldsymbol{x}) > 0$, но $\min(f_{t_1}(\boldsymbol{x}), f_{t_2}(\boldsymbol{x})) = 0$. Поэтому $f_t(\boldsymbol{x})$ положительно всюду на $\Theta = [\theta_1, \theta_2]$, так что (12) справедливо при всех t. Интегрируя (12) в пределах от θ_1 до θ,

получим

$$L(x,\theta)=\theta^*\int\limits_{\theta_1}^{\theta}c(t)\,dt-\int\limits_{\theta_1}^{\theta}c(t)a(t)\,dt+L(x,\theta_1),$$

что эквивалентно (3) для $[\mu^n]$ п. в. x. Так как изменение $f_\theta(x)$ на множестве μ^n-меры нуль роли не играет, то (3) доказано.

Обратимся к последнему утверждению теоремы. Если $\theta^*=\text{const}$, то $b'(\theta)=-1$, и обе части неравенства (2) обращаются в нуль. Пусть теперь выполнено (3). Тогда, дифференцируя по θ функцию $L(X,\theta)$, получим

$$L'(X,\theta)=\theta^*A'(\theta)+B'(\theta).$$

Из (7) следует, что $a(\theta)A(\theta)+B(\theta)=0$. Поэтому

$$L'(X,\theta)=A'(\theta)(\theta^*-a(\theta))$$

и, стало быть (см. (12)), в (2) достигается равенство. $\triangleleft$

Замечание 3. Так как в представлении (3) $A(\theta)=\int\limits_{\theta_1}^{\theta}c(t)\,dt>0$, то (3) означает, что носитель $N_{\mathbf{P}_\theta}=\{\boldsymbol{x}:h(\boldsymbol{x})>0,\ \theta^*>-\infty\}$ от θ не зависит, и можно считать $N_{\mathbf{P}_\theta}=\mathcal{X}$. Таким образом, если $N_{\mathbf{P}_\theta}$ зависит от θ (как, например, в задаче оценки параметра сдвига, когда $f_\theta(x)=f(x-\theta)$ и $f(x)>0$ лишь для $x\in[0,a]$, $0<a\leqslant\infty$), то не существует оценок θ^*, для которых в (2) достигалось бы равенство.

Если доказывать второе утверждение теоремы 2 лишь для случая $N_{\mathbf{P}_\theta}=\mathcal{X}$, то доказательство можно несколько упростить, сделав ненужными рассмотрения, связанные с (11), (12) и непрерывностью функций $c(\theta)$, $a(\theta)$. Чтобы получить требуемое представление для $L(x,\theta)$, достаточно заметить, что при $f_\theta(\boldsymbol{x})>0$, $\boldsymbol{x}\in\mathcal{X}^n$, наряду с (10) справедливо

$$\int[c(t)(\theta^*-a(t))-L'(\boldsymbol{x},t)]\varphi(\boldsymbol{x})\mu^n(d\boldsymbol{x})=0$$

для любой ограниченной измеримой функции φ. Надо теперь проинтегрировать это равенство по t в пределах от θ_1 до θ, поменять порядок интегрирования и воспользоваться произвольностью φ.

В дальнейшем тривиальный случай $\theta^*=\text{const}$ будем исключать из рассмотрений и предполагать, что $\mathbf{D}_\theta\theta^*>0$ всюду на Θ. Тогда справедливо

Следствие 2. *При выполнении условий* (R) *для достижения нижней границы в неравенстве Рао–Крамера необходимо и достаточно, чтобы оценка θ^* была достаточной и функция $\psi(\theta^*,\theta)$ в факторизационном равенстве имела вид*

$$\psi(\theta^*,\theta)=\exp\{\theta^*A(\theta)+B(\theta)\},$$

где $A(\theta)$ и $B(\theta)$ — дифференцируемые функции.

Следствие 3. *Если выполнены условия* (R), $\theta^* \in K_b$, *и в неравенстве Рао–Крамера достигается равенство, то* θ^* — *эффективная в* K_b *оценка.*

Это утверждение следует из представления

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \mathbf{D}_\theta\theta^* + b^2(\theta).$$

Заметим, что обратное, вообще говоря, не верно: оценка может быть эффективной в K_b, но нижняя граница $\dfrac{(1+b'(\theta))^2}{nI(\theta)}$ для дисперсии может не достигаться.

Пример 1. Пусть $X \in \mathbf{\Gamma}_{\alpha,1}$. Здесь $f_\alpha(X) = \alpha^n e^{-\alpha n\bar{\mathrm{x}}}$. Условия (R) в области $\Theta \subseteq \{\alpha \geqslant \delta > 0\}$ выполнены. Очевидно, $S = n\bar{\mathrm{x}}$ есть полная достаточная статистика. Поэтому оценка $\alpha^* = \bar{\mathrm{x}}^{-1} = \mathbf{E}_\alpha(\bar{\mathrm{x}}^{-1}/S)$ является эффективной в классе K_b со смещением $b(\alpha) = \mathbf{E}_\alpha\bar{\mathrm{x}}^{-1} - \alpha$.

Заметим теперь, что $S \in \mathbf{\Gamma}_{\alpha,n}$, так что при $n > 1$ (см. § 12) $\mathbf{E}_\alpha\bar{\mathrm{x}}^{-1} = n\mathbf{E}_\alpha S^{-1} = \dfrac{n}{n-1}\alpha$.

Таким образом, оценка $\alpha^{**} = \dfrac{n-1}{n\bar{\mathrm{x}}} = \alpha^*\left(1 - \dfrac{1}{n}\right)$ будет при $n > 1$ несмещенной.

Аналогично при $n > 2$ находим (см. § 12, а также пример 14.1)

$$\mathbf{E}_\alpha(\alpha^{**})^2 = (n-1)^2\mathbf{E}_\alpha S^{-2} = \frac{n-1}{n-2}\alpha^2,$$

$$\mathbf{D}_\alpha\alpha^{**} = \alpha^2\left[\frac{n-1}{n-2} - 1\right] = \frac{\alpha^2}{n-2}.$$

Итак, при $n > 2$ оценка α^{**} является эффективной. Однако критерий (3) не выполнен, так как

$$f_\alpha(X) = \alpha^{-n}e^{-\alpha(n-1)/\alpha^{**}}.$$

Следовательно, в неравенстве Рао–Крамера нижняя граница не достигается. В этом можно убедиться и непосредственно. В самом деле, здесь $l(x,\alpha) = \ln\alpha - \alpha x$, $l'(x,\alpha) = 1/\alpha - x$ и

$$I(\alpha) = \mathbf{E}_\alpha[l'(\mathrm{x}_1,\alpha)]^2 = \mathbf{E}_\alpha\left(\frac{1}{\alpha} - \mathrm{x}_1\right)^2 = \frac{1}{\alpha^2}.$$

Стало быть, при $n > 2$

$$\frac{1}{nI(\alpha)} = \frac{\alpha^2}{n} < \frac{\alpha^2}{n-2} = \mathbf{D}_\alpha\alpha^{**}.$$

Таким образом, достижение нижней границы в (2) есть более жесткое требование, чем эффективность.

2. *R*-эффективные и асимптотически *R*-эффективные оценки. Пусть выполнены условия (R). В этом случае достижение нижней границы (точное или асимптотическое) для дисперсии в неравенстве Рао–Крамера может служить важным показателем качества оценок, тесно связанным с понятием эффективности.

Определение 1. Оценка $\theta^* \in K_b$, для которой

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta),$$

называется *R-эффективной* (или *регулярно эффективной*) *в классе* K_b.

R-эффективная оценка в классе K_0 несмещенных оценок называется просто *R-эффективной.*

Оценка θ^* называется *асимптотически R-эффективной* (а. R-э.), если

$$\mathbf{E}_\theta(\theta^* - \theta)^2 = \frac{1 + o(1)}{nI(\theta)}.$$

Мы видим, что в отличие от определений § 18, которые носили более качественный характер, определения R-эффективности основаны на сравнении с известными числовыми значениями, связанными, главным образом, с информацией Фишера, точнее, с количеством $(nI(\theta))^{-1}$.

Для R-эффективности оценки θ^* необходимо и достаточно выполнение (3).

Из сказанного выше следует, что *R-эффективные оценки являются эффективными, но не наоборот*; R-эффективные оценки просто реже существуют, что является недостатком нижней границы в неравенстве Рао–Крамера, а не оценок.

В существующей литературе по математической статистике R-эффективные оценки называются просто эффективными. Однако нам представляется более естественным сохранить термин «эффективность» для наилучших оценок в более широком понимании (см. определение 18.1).

Теорема 2. *Если выполнены условия* (R) *и R-эффективная оценка существует, то она совпадает с оценкой м. п.*

Доказательство. Мы уже видели, что выполнение (3) влечет за собой равенство (см. (11))

$$L'(X, \theta) = (\theta^* - \theta)c(\theta).$$

Кроме того, так как $b(\theta) = 0$, из (12) следует

$$c(\theta) = \frac{1}{\mathbf{D}_\theta \theta^*} = nI(\theta) > 0$$

для любых $\theta \in \Theta$. Это означает, что $L'(X, \theta) < 0$ при $\theta > \theta^*$ и $L'(X, \theta) > 0$ при $\theta < \theta^*$. Следовательно, при $\theta = \theta^*$ достигается максимум $L(X, \theta)$. $\triangleleft$

Приведенный выше пример 1 показывает, что эффективные оценки в отличие от R-эффективных могут не совпадать с о. м. п. В этом примере

о. м. п. есть $(\overline{\mathrm{x}})^{-1}$, в то время как эффективная оценка равна $\dfrac{n-1}{n}(\overline{\mathrm{x}})^{-1}$. Обе эти оценки являются, очевидно, а. R-э. оценками.

Рассмотрим класс $\widetilde{K}_0$ оценок θ^*, для которых

$$|b(\theta)| \leqslant \frac{\varepsilon(\theta,n)}{\sqrt{n}}, \quad |b'(\theta)| \leqslant \varepsilon(\theta,n), \quad \mathbf{E}_\theta(\theta^*)^2 < c < \infty$$

при некоторой функции $\varepsilon(\theta,n) = o(1)$ при $n \to \infty$ и каждом $\theta \in \Theta$.

Каждый такой класс замечателен тем, что для него нижняя граница в неравенстве Рао–Крамера имеет вид $(1+o(1))/[nI(\theta)]$. В § 30 мы увидим, что в ряде случаев при отыскании асимптотически оптимальных оценок можно ограничиться изучением оценок θ^* из таких классов.

Теорема 3. *Пусть выполнены условия* (R). *Тогда всякая а. R-э. оценка из $\widetilde{K}_0$ является а. э. в $\widetilde{K}_0$ оценкой.*

Доказательство теоремы очевидно: если θ_1^* есть а. R-э. оценка, то

$$\mathbf{E}_\theta\big(\theta_1^* - \theta\big)^2 = \frac{1+o(1)}{nI(\theta)}.$$

Кроме того, как уже отмечалось, по неравенству Рао–Крамера для всех $\theta^* \in \widetilde{K}_0$

$$\liminf_{n\to\infty} \mathbf{E}_\theta n(\theta^* - \theta)^2 \geqslant I^{-1}(\theta) = \lim_{n\to\infty} \mathbf{E}_\theta n\big(\theta_1^* - \theta\big)^2. \quad \triangleleft$$

Ясно также, что если а. R-э. оценка существует, любая а. э. оценка в $\widetilde{K}_0$ будет а. R-э. оценкой.

Позже мы увидим (см. § 34), что при некоторых дополнительных предположениях а. R-э. оценки существуют всегда. Ими оказываются о. м. п. $\widehat{\theta}^*$. Этот факт тесно связан с теоремой 16.3, в силу которой $\sqrt{n}\,(\widehat{\theta}^* - \theta) \Subset \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)}$. Чтобы установить, что $\widehat{\theta}^*$ является а. R-э. оценкой, нужны дополнительные рассмотрения, которые будут проведены в § 33. Сказанное означает, что утверждение теоремы 3 справедливо и в обратную сторону: а. э. оценка в $\widetilde{K}_0$ является а. R-э. оценкой, т. е. для нее $\mathbf{E}_\theta(\theta^* - \theta)^2 \sim [nI(\theta)]^{-1}$.

Теорема 4. *Пусть выполнены условия* (R). *Если θ_1^*, θ_2^* принадлежат $\widetilde{K}_0$ и являются а. R-э. оценками, то они асимптотически эквивалентны в следующем смысле:*

$$\sqrt{n}\,\big(\theta_1^* - \theta_2^*\big) \underset{P}{\longrightarrow} 0.$$

Доказательство этого утверждения проходит точно так же, как в теореме 18.2. Так как $\theta^* = \big(\theta_1^* + \theta_2^*\big)/2 \in \widetilde{K}_0$, на основании (18.11) и неравенства Рао–Крамера получаем

$$\limsup_{n\to\infty} \mathbf{E}_\theta n\big(\theta_1^* - \theta_2^*\big)^2 \leqslant 0. \quad \triangleleft$$

Пример 2. Оценка $\alpha^* = \overline{\mathrm{x}}$ среднего значения α нормальной совокупности $\mathbf{\Phi}_{\alpha,\sigma^2}$ при известном σ^2 является R-эффективной оценкой. В этом легко убедиться, проверив, например, условие (3). Другая возможность состоит в сопоставлении $\mathbf{D}_\alpha\alpha^* = \sigma^2/n$ с наименьшим возможным значением $(nI(\alpha))^{-1}$ дисперсий несмещенных оценок. В нашем случае

$$l(x,\alpha) = -\ln\sqrt{2\pi}\,\sigma - \frac{(x-\alpha)^2}{2\sigma^2},$$

$$l'(x,\alpha) = \frac{x-\alpha}{\sigma^2},$$

$$I(\alpha) = \mathbf{E}_\alpha[l'(\mathrm{x}_1,\alpha)]^2 = \frac{\mathbf{E}_\alpha(\mathrm{x}_1-\alpha)^2}{\sigma^4} = \frac{1}{\sigma^2},$$

так что $\mathbf{D}_\alpha\alpha^* = (nI(\alpha))^{-1} = \sigma^2/n$.

Пример 3. Рассмотрим оценку $\theta^* = S_1^2 = \dfrac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i-\alpha)^2$ параметра $\theta = \sigma^2$ нормальной совокупности с известным α. Нетрудно подсчитать, что $\mathbf{D}_\theta\theta^* = \mathbf{E}_\theta(\theta^*-\sigma^2)^2 = 2\sigma^4/n$. С другой стороны, здесь

$$l'(\mathrm{x}_1,\theta) = -\frac{1}{2\theta} + \frac{(\mathrm{x}_1-\alpha)^2}{2\theta^2},$$

$$I(\theta) = \mathbf{E}_\theta[l'(\mathrm{x}_1,\theta)]^2 = \frac{1}{4\theta^4}\mathbf{E}_\theta[(\mathrm{x}_1-\alpha)^2-\theta]^2 = \frac{1}{2\theta^2} = \frac{1}{2\sigma^4}.$$

Таким образом, здесь также $\mathbf{D}_\theta\theta^* = (nI(\theta))^{-1}$, и оценка $\theta^* = S_1^2$ является R-эффективной.

Дисперсия несмещенной оценки $S_0^2 = \dfrac{1}{n-1}\sum(\mathrm{x}_i-\overline{\mathrm{x}})^2$ равна $\dfrac{2\sigma^4}{n-1}$, так что она не является R-эффективной или просто эффективной оценкой σ^2. В то же время очевидно, что S_0^2 есть а. R-э. оценка.

Если мы будем в качестве неизвестного параметра оценивать не σ^2, а $\theta = \sigma$, то R-эффективной оценки мы не получим. Несмещенной оценкой σ будет оценка

$$\sigma^* = \sqrt{\frac{n}{2}}\,\frac{\Gamma\left(\dfrac{n}{2}\right)}{\Gamma\left(\dfrac{n+1}{2}\right)}S_1,$$

так как

$$\mathbf{E}_\sigma S_1 = \frac{\sigma}{\sqrt{n}}\mathbf{E}_\sigma\sqrt{\frac{1}{\sigma^2}\sum(\mathrm{x}_i-\alpha)^2},$$

где $\dfrac{nS_1^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum(\mathrm{x}_i-\alpha)^2$ имеет распределение $\mathbf{H}_n = \mathbf{\Gamma}_{1/2,n/2}$, поэтому (см. § 12)

$$\mathbf{E}_\sigma S_1 = \frac{\sigma\sqrt{2}}{\sqrt{n}}\,\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)}, \quad \mathbf{E}_\sigma\sigma^* = \sigma.$$

Так как S_1 есть минимальная и полная достаточная статистика, то σ^* является эффективной оценкой. С помощью формулы Стирлинга нетрудно убедиться, что $\sigma^* = S_1(1 + O(1/n))$.

Сравним теперь величину $\mathbf{D}_\sigma\sigma^*$ с нижней границей $(nI(\sigma))^{-1}$. Имеем

$$\mathbf{D}_\sigma\sigma^* = \mathbf{E}_\sigma\left(\sqrt{\frac{n}{2}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}S_1 - \sigma\right)^2 = \sigma^2\left[\frac{n}{2}\frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1\right]. \qquad (13)$$

С другой стороны, здесь

$$l'(x,\sigma) = -\frac{1}{\sigma} + \frac{(x-\alpha)^2}{\sigma^3},$$

$$I(\sigma) = \mathbf{E}_\sigma[l'(\mathrm{x}_1,\sigma)]^2 = \frac{1}{\sigma^6}\mathbf{E}_\sigma[(\mathrm{x}_1-\alpha)^2 - \sigma^2]^2 = \frac{2}{\sigma^2},$$

так что $(nI(\sigma))^{-1} = \sigma^2/(2n)$. Но это значение отличается от (13). Их отношение, например, для $n = 3$ равно 0,936. Таким образом, R-эффективных оценок здесь не существует. При $n \to \infty$ коэффициент при σ^2 в (13) ведет себя асимптотически как $\frac{1}{2n} + O\left(\frac{1}{n^2}\right)$, так что σ^* есть а. R-э. оценка.

3. Неравенство Рао–Крамера в многомерном случае. В этом пункте $\theta = (\theta_1,\ldots,\theta_k)$ есть k-мерный вектор, как и оценка $\theta^* = (\theta_1^*,\ldots,\theta_k^*)$. Как и прежде, положим

$$a(\theta) = \mathbf{E}_\theta\theta^* = \theta + b(\theta), \quad b(\theta) = (b_1(\theta),\ldots,b_k(\theta))$$

и рассмотрим классы K_b оценок с фиксированным смещением $b(\theta)$.

Обобщение условий (R) на многомерный случай будет выглядеть следующим образом. Обозначим

$$l(x,\theta) = \log f_\theta(x), \quad l_i'(x,\theta) = \frac{\partial}{\partial\theta_i}l(x,\theta),$$

$$I_{ij}(\theta) = \mathbf{E}_\theta l_i'(\mathrm{x}_1,\theta)l_j'(\mathrm{x}_1,\theta)$$

и предположим, что выполнено условие

(R) *Функции* $\sqrt{f_\theta(x)}$ *непрерывно дифференцируемы по* θ_j *для п. в.* $[\mu]$ *значений* x. *Матрица*

$$I(\theta) = \|I_{ij}(\theta)\|, \quad I_{ij}(\theta) = \int l_i'(x;\theta)l_j'(x;\theta)f_\theta(x)\mu(dx)$$

непрерывна по θ*), ее определитель* $|I(\theta)|$ *отличен от нуля.*

*) Для этого достаточно требовать равномерной сходимости интегралов $I_{ij}(\theta)$ (см. приложение V).

Так как $I(\theta)$ есть матрица вторых моментов $\mathbf{E}_\theta l_i l_j$ случайных величин $l_i = l_i'(\mathrm{x}_1, \theta)$, то она будет положительно определенной матрицей, так как для любого вектора $\alpha = (\alpha_1, \ldots, \alpha_k) \neq 0$ выполняется

$$\sum \alpha_i \alpha_j \mathbf{E}_\theta l_i l_j = \mathbf{E}_\theta \left(\sum \alpha_i l_i \right)^2 \geqslant 0,$$

где равенство нулю исключается условием $|I(\theta)| \neq 0$.

Неравенство между матрицами $\sigma_1^2 \geqslant \sigma_2^2$, как и раньше, будем понимать как неравенство $\alpha \sigma_1^2 \alpha^\top \geqslant \alpha \sigma_2^2 \alpha^\top$ для любой вектор-строки $\alpha = (\alpha_1, \ldots, \alpha_k) \neq 0$. Это эквивалентно, очевидно, тому, что матрица $\sigma_1^2 - \sigma_2^2$ неотрицательно определена. Строгое неравенство будет соответствовать положительной определенности, так что, например, $I(\theta) > 0$.

Теорема 1А. *Если $\theta^* \in K_b$ и выполнено условие* (R), *то для матрицы вторых моментов $\sigma^2 = \|\sigma_{ij}\| = \mathbf{E}_\theta (\theta^* - a(\theta))^\top (\theta^* - a(\theta))$ любой оценки θ^* вектор-строки θ справедливо неравенство*

$$\sigma^2 \geqslant \frac{1}{n} (E + D(\theta)) I^{-1}(\theta) (E + D(\theta))^\top, \tag{14}$$

где E есть единичная матрица, $D(\theta) = \|b_{ij}(\theta)\|$, $b_{ij}(\theta) = \dfrac{\partial b_i(\theta)}{\partial \theta_j}$.

Пусть $|\sigma^2| > 0$ (или $|E + D(\theta)| > 0$) при всех θ. В этом случае знак равенства в (14) *достигается тогда и только тогда, когда распределение выборки принадлежит экспоненциальному семейству специального вида, т. е. когда для некоторых скалярных функций $B(\theta)$ и $h(X)$ выполняется*

$$f_\theta(X) = \exp\{(\theta^*, A(\theta)) + B(\theta)\} h(X), \tag{15}$$

где вектор $A(\theta) = (A_1(\theta), \ldots, A_k(\theta))$ имеет матрицу производных, равную

$$\|A_{ij}\| \equiv \left\| \frac{\partial A_i(\theta)}{\partial \theta_j} \right\| = n[(E + D(\theta))^{-1}]^\top I(\theta).$$

Для несмещенных оценок θ^*, очевидно,

$$\sigma^2 \geqslant (nI(\theta))^{-1},$$

и равенство возможно лишь при выполнении (15), *где $\|A_{ij}\| = nI(\theta)$.*

Таким образом, если нам удастся найти несмещенную оценку θ^* с матрицей вторых моментов $[nI(\theta)]^{-1}$, то она будет эффективной оценкой.

Отметим, что

$$\mathbf{E}_\theta (\theta^* - \theta)^\top (\theta^* - \theta) = \sigma^2 + b^\top(\theta) b(\theta).$$

Доказательство теоремы 1А. Положим

$$L_j' = L_j'(X, \theta) = \sum_{i=1}^n l_j'(\mathrm{x}_i, \theta), \quad L' = L'(X, \theta) = \left(L_1', \ldots, L_k'\right).$$

Тогда совершенно аналогично одномерному случаю устанавливаем, что справедливы равенства

$$\mathbf{E}_\theta l'_j(\mathrm{x}_1, \theta) = 0, \quad \mathbf{E}_\theta \theta_i^* L'_j(X, \theta) = 1 + b_{ij}(\theta),$$

в которых $b_{ij}(\theta)$ непрерывны, или, что то же, равенства

$$\mathbf{E}_\theta L' = 0, \tag{16}$$

$$\mathbf{E}_\theta (\theta^*)^\top L' = E + D(\theta), \tag{17}$$

в которых матрица $D(\theta)$ непрерывна. Отсюда получаем

$$\mathbf{E}_\theta (\theta^* - a(\theta))^\top L' = E + D(\theta). \tag{18}$$

Докажем теперь следующее неравенство (матричный вариант неравенства Коши–Буняковского).

Лемма 2. *Пусть ξ и η — матрицы одинакового размера (не обязательно квадратные) со случайными элементами, матрица $\mathbf{E}\eta\eta^\top$ имеет обратную. Тогда*

$$\mathbf{E}\xi\xi^\top \geqslant \mathbf{E}\xi\eta^\top (\mathbf{E}\eta\eta^\top)^{-1} \mathbf{E}\eta\xi^\top. \tag{19}$$

При этом равенство возможно лишь в случае $\xi = z\eta$, $z = \mathbf{E}\xi\eta^\top (\mathbf{E}\eta\eta^\top)^{-1}$.

Доказательство. Так как для любой матрицы A справедливо неравенство $AA^\top \geqslant 0$ ($AA^\top$ неотрицательно определена), то

$$0 \leqslant \mathbf{E}(\xi - z\eta)(\xi - z\eta)^\top = \mathbf{E}\xi\xi^\top - z\mathbf{E}\eta\xi^\top - \mathbf{E}\xi\eta^\top z^\top + z\mathbf{E}\eta\eta^\top z^\top.$$

Полагая $z = \mathbf{E}\xi\eta^\top (\mathbf{E}\eta\eta^\top)^{-1}$, получим требуемое неравенство.

Утверждение относительно условий равенства в (19) очевидно. $\triangleleft$

Вернемся к доказательству теоремы 1А. Положим в (19) $\xi = (\theta^* - a(\theta))^\top$, $\eta = (L')^\top$. Тогда

$$\mathbf{E}_\theta \xi\xi^\top = \mathbf{E}_\theta (\theta^* - a(\theta))^\top (\theta^* - a(\theta)) = \sigma^2.$$

Из (16) и независимости x_i получаем

$$\mathbf{E}_\theta \eta\eta^\top = \mathbf{E}_\theta (L')^\top L' = nI(\theta).$$

Наконец, из (18) находим

$$\mathbf{E}_\theta \xi\eta^\top = \mathbf{E}_\theta (\theta^* - a(\theta))^\top L' = E + D(\theta).$$

Неравенство (14) доказано.

Равенство в (14) возможно в силу леммы 2, если только для точек $(\boldsymbol{x}, \theta)$ таких, что $f_\theta(\boldsymbol{x}) > 0$, справедливо

$$(\theta^* - a(\theta))^\top = (E + D(\theta))(nI(\theta))^{-1} (L')^\top$$

или, что то же,

$$L' = (\theta^* - a(\theta)) n [(E + D(\theta))^{-1}]^\top I(\theta). \tag{20}$$

Заметим теперь, что из равенства в (14) следует

$$|E + D(\theta)|^2 = n|\sigma^2| \cdot |I(\theta)|,$$

и отделенность от нуля определителя $|\sigma^2|$ означает то же самое для $|E + D(\theta)|$ и означает существование равномерно ограниченной обратной матрицы $(E + D(\theta))^{-1}$. Поэтому производная L' в (20) будет ограниченной, $f_\theta(\boldsymbol{x}) > 0$ всюду на Θ и само равенство (20) будет справедливо всюду на Θ. Если теперь s есть любой путь, соединяющий точки θ_0 и θ в области Θ, то

$$L(X, \theta) = \int_s (L', ds) + L(X, \theta_0),$$

где ds означает векторный элемент пути s, $((L', ds) = (L', s'(l))\, dl$ есть приращение $L(X, \theta)$ на этом пути, l — «длина» пройденного пути). Стало быть, в силу (20)

$$L(X, \theta) = \theta^* A(\theta) + B(\theta) + H(X), \tag{21}$$

где $B(\theta)$ и $H(X)$ — скалярные функции, $A(\theta) = (A_1(\theta), \ldots, A_k(\theta))$ — вектор, зависящий только от своих аргументов. Это означает справедливость (15).

Если выполнено (21), то

$$L' = \theta^* \|A_{ij}\| + B'(\theta),$$

где в силу равенства $\mathbf{E}_\theta L' = 0$ справедливо

$$B'(\theta) = -a(\theta)\|A_{ij}\|.$$

Умножая обе части равенства $L' = (\theta^* - a(\theta))\|A_{ij}\|$ слева на $(\theta^* - a(\theta))^\top$, получим в силу (18), что для выполнения условия (20), означающего равенство в (14), должно выполняться

$$\|A_{ij}\| = n[(E + D(\theta))^{-1}]^\top I(\theta). \quad \triangleleft$$

В многомерном случае сохраняют свою силу все замечания, сделанные к одномерному неравенству Рао–Крамера, а также определение R-эффективности, в которые следует внести лишь очевидные изменения, связанные с размерностью θ.

В частности, а. R-э. оценками будем называть оценки θ^*, для которых

$$\mathbf{E}_\theta(\theta^* - \theta)^\top(\theta^* - \theta) = \sigma^2 + b^\top(\theta)b(\theta) = (nI(\theta))^{-1} + o(1/n).$$

Аналог теоремы 2 здесь будет выглядеть следующим образом.

Теорема 2А. *Пусть выполнены условия* (R). *Если θ^* есть R-эффективная оценка, то это есть оценка максимального правдоподобия.*

Доказательство. Чтобы доказать, что R-эффективная оценка является единственной точкой максимума, достаточно убедиться, что $L'(X, \theta^*) = 0$ и что при $\theta = \theta^* + u$, $u \neq 0$

$$(\operatorname{grad} L(X, \theta), u) = (L'(X, \theta),\ \theta - \theta^*) < 0.$$

Но в случае существования R-эффективной оценки выполняется (см. (20))

$$L'(X,\theta) = (\theta^* - \theta)nI(\theta),$$

откуда немедленно вытекают оба требуемые соотношения. Второе следует из того, что

$$(L', u) = -unI(\theta)u^\top,$$

где $uI(\theta)u^\top$ есть положительно определенная квадратичная форма. $\triangleleft$

Пример 4. Рассмотрим двупараметрическое семейство нормальных распределений Φ_{α,σ^2}. Оно принадлежит экспоненциальному семейству, так как (здесь $\theta = (\theta_1, \theta_2)$, $\theta_1 = \alpha$, $\theta_2 = \sigma^2$)

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(x-\alpha)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2} + \frac{x\alpha}{\sigma^2} - \frac{\alpha^2}{2\sigma^2} - \ln\sigma\right\}.$$

Оценка $\theta^* = \left(\theta_1^*, \theta_2^*\right)$, где $\theta_1^* = \overline{\mathrm{x}}$, $\theta_2^* = S_0^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2 =$

$= \dfrac{1}{n-1}\left(\sum \mathrm{x}_i^2 - n\overline{\mathrm{x}}^2\right)$, является эффективной, поскольку она принадлежит K_0 и статистика $\left(\sum \mathrm{x}_i, \sum \mathrm{x}_i^2\right)$, как мы видели в § 25, является полной достаточной статистикой (см. теорему 24.4).

Оценка м. п. $\left(\overline{\mathrm{x}}, \dfrac{1}{n}\sum(\mathrm{x}_i - \overline{\mathrm{x}})^2\right)$ отличается от θ^* лишь множителем $\dfrac{n-1}{n}$ при второй координате, что делает ее смещенной. Для выбранной оценки θ^* специальное экспоненциальное представление (15) функции $f_\theta(X)$ реализовываться не будет, так как

$$f_\theta(X) = (2\pi)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum \mathrm{x}_i^2 + \frac{\alpha}{\sigma^2}\sum \mathrm{x}_i - \frac{n\alpha^2}{2\sigma^2} - n\ln\sigma\right\} =$$

$$= (2\pi)^{-n/2}\exp\left\{\frac{\alpha n}{\sigma^2}\theta_1^* - \frac{n-1}{2\sigma^2}\theta_2^* - \frac{n}{2\sigma^2}\left(\theta_1^*\right)^2 - \frac{n\alpha^2}{2\sigma^2} - n\ln\sigma\right\}.$$

Это значит, что нижняя граница в многомерном неравенстве Рао–Крамера достигаться не будет.

Наименьший эллипсоид рассеивания, определяемый по теореме 1А матрицей $I(\theta)$ (или $I^{-1}(\theta)$), будет достигаться лишь асимптотически при $n \to \infty$, так что оценка θ^*, не являясь R-эффективной, будет а. R-э. оценкой. Убедимся в этом непосредственно.

Вычислим сначала матрицу $I(\theta)$. Имеем

$$l_1'(x,\theta) = \frac{(x-\alpha)}{\sigma^2}, \quad l_2'(x,\theta) = \frac{(x-\alpha)^2}{2\sigma^4} - \frac{1}{2\sigma^2}$$

(напомним, что l_2' есть производная по σ^2, а не по σ; ср. с примером 3). Поэтому

$$I_{11}(\theta) = \mathbf{E}_\theta \frac{(\mathrm{x}_1 - \alpha)^2}{\sigma^4} = \frac{1}{\sigma^2},$$

$$I_{12}(\theta) = I_{21}(\theta) = \mathbf{E}_\theta \left[\frac{(\mathrm{x}_1 - \alpha)^3}{2\sigma^6} - \frac{\mathrm{x}_1 - \alpha}{2\sigma^4} \right] = 0,$$

$$I_{22}(\theta) = \frac{1}{4\sigma^8} \mathbf{E}_\theta [(\mathrm{x}_1 - \alpha)^2 - \sigma^2]^2 = \frac{1}{2\sigma^4}.$$

Отсюда находим

$$(nI(\theta))^{-1} = \begin{bmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{bmatrix}. \tag{22}$$

Вычислим теперь для сравнения матрицу вторых центральных моментов оценки θ^*.

Имеем

$$\mathbf{E}_\theta \big(\theta_1^* - \theta_1\big)^2 = \mathbf{E}_\theta (\overline{\mathrm{x}} - \alpha)^2 = \frac{\sigma^2}{n},$$

$$\mathbf{E}_\theta \big(\theta_2^* - \theta_2\big)^2 = \mathbf{E}_\theta \big(S_0^2 - \sigma^2\big)^2 = \frac{2\sigma^4}{n-1},$$

$$\mathbf{E}_\theta \big(\theta_1^* - \theta_1\big)\big(\theta_2^* - \theta_2\big) = 0.$$

Последние два равенства устанавливаются непосредственным подсчетом Рассмотрим, например, второе из них. Нам достаточно убедиться, что

$$\mathbf{E}_\theta (\overline{\mathrm{x}} - \alpha) S_0^2 = 0. \tag{23}$$

Но

$$S_0^2 = \frac{1}{n-1} \left[\sum (\mathrm{x}_i - \alpha)^2 - (\overline{\mathrm{x}} - \alpha)^2 \right],$$

$$(\overline{\mathrm{x}} - \alpha) S_0^2 = \frac{1}{n(n-1)} \left[\sum (\mathrm{x}_i - \alpha) \right] \left[\sum (\mathrm{x}_i - \alpha)^2 \right] - \frac{1}{n(n-1)} (\overline{\mathrm{x}} - \alpha)^3.$$

Поскольку

$$\mathbf{E}_\theta (\overline{\mathrm{x}} - \alpha)^3 = \mathbf{E}_\theta (\mathrm{x}_i - \alpha)^3 = \mathbf{E}_\theta (\mathrm{x}_i - \alpha)(\mathrm{x}_j - \alpha)^2 = 0,$$

то (23) доказано.

Таким образом, матрица вторых моментов $\theta^* - \theta$ равна

$$\begin{bmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n-1} \end{bmatrix}.$$

Ясно, что различие между этой матрицей и матрицей $(nI(\theta))^{-1}$ может быть существенным лишь при небольших значениях n.

4. Некоторые выводы. В заключение этого параграфа подведем некоторые итоги исследований, проведенных в последних шести параграфах. Главная их цель состояла в отыскании способов построения оптимальных (в том или ином смысле) оценок и установлении нижних границ для их среднеквадратичных уклонений. В результате можно указать следующие четыре основные направления, по которым можно идти в поисках наилучших оценок.

1. Построение байесовских (если есть априорная информация о θ) и минимаксных оценок.

2. Отыскание полных (или минимальных) достаточных статистик S. Тогда оценка $\theta^*_S = \mathbf{E}_\theta(\theta^*/S)$ будет эффективной в том классе K_b, которому принадлежат θ^*.

3. Использование о. м. п. в тех случаях, когда выполнен критерий (3) теоремы 1 (или критерий (15) теоремы 1А). При этом также получим эффективные (и даже R-эффективные) оценки в классах с фиксированным смещением.

4. Количественный подход, основанный на сравнении среднеквадратичного уклонения $\mathbf{E}_\theta(\theta^* - \theta)^2$ оценки θ^*, которую хотим использовать, с нижней границей R, определяемой неравенством Рао–Крамера. Если отношение $\mathbf{E}_\theta(\theta^* - \theta)^2/R$ будет близко к единице, то оценка θ^* может быть рекомендована к использованию. На этом пути мы получим в дальнейшем весьма общие результаты, связанные с построением асимптотически эффективных, асимптотически байесовских и асимптотически минимаксных оценок.

Сделаем также следующее замечание. Во всех отмеченных выше направлениях существенную роль играет форма зависимости распределения выборки $\mathbf{P}_\theta$ от оцениваемого параметра θ. На практике, однако, нередко возникают задачи оценки не самого θ, а некоторой функции от него $\varphi(\theta)$. При этом нетрудно видеть (см. пример со схемой Бернулли в (18.4), (18.5)), что оценка $\varphi^* = \varphi(\theta^*)$ далеко не всегда будет обладать теми же свойствами, которыми обладала оценка θ^* (несмещенность, эффективность и т. д. Сохранятся лишь свойства асимптотической эффективности, если φ — гладкая функция). С этой точки зрения представляется естественным с самого начала рассматривать задачу оценивания *функций* $\varphi(\theta)$ от исходного параметра θ. Мы отказались от такого подхода, так как на этом пути многие основные результаты, полученные нами, существенно усложнились бы. С другой стороны, если φ осуществляет взаимно однозначное отображение, то задача об оценке $\varphi(\theta)$ сводится к задаче, рассматриваемой нами путем «перепараметризации», т. е. введения нового параметра $\gamma = \varphi(\theta)$, которому будет соответствовать семейство распределений $\mathbf{G}_\gamma = \mathbf{P}_{\varphi^{-1}(\gamma)}$.

§ 27.* Свойства информации Фишера

Мы уже видели и будем убеждаться в дальнейшем, что информация Фишера играет важную роль в математической статистике. В связи с этим выясним некоторые ее полезные свойства.

1. Одномерный случай. Информация Фишера

$$I(\theta) = \int \frac{\left(f'_\theta(x)\right)^2}{f_\theta(x)}\mu(dx) = \mathbf{E}_\theta(l'(\mathrm{x}_1, \theta))^2$$

появилась в рассмотрениях § 16, 26. Величину

$$I^X(\theta) = \mathbf{E}_\theta[L'(X,\theta)]^2$$

принято рассматривать как *меру количества информации, содержащейся в выборке X относительно параметра θ*. В теореме 26.1 мы доказали *аддитивность* информации: $I^X(\theta) = nI(\theta)$, т. е. что $I^X(\theta)$ равна сумме информаций $I^{\mathrm{x}_i}(\theta) = \mathbf{E}_\theta[l'(\mathrm{x}_i,\theta)]^2 = I(\theta)$, содержащихся в независимых наблюдениях $\mathrm{x}_1,\ldots,\mathrm{x}_n$.

Докажем еще одно свойство информации Фишера. Пусть $S = S(X)$ есть некоторая статистика со значениями в $\mathbb{R}^l$ и $g_\theta(s)$ есть плотность ее распределения, индуцированного распределением $\mathbf{P}_\theta$ в $\left(\mathcal{X}^n, \mathfrak{B}^n_{\mathcal{X}}\right)$ относительно некоторой меры λ в $(\mathbb{R}^l, \mathfrak{B}^l)$. В соответствии с предыдущими обозначениями величину

$$I^S(\theta) = \mathbf{E}_\theta[(\log g_\theta(S))']^2$$

будем называть *информацией, содержащейся в статистике S относительно параметра θ*.

Отметим, что значение $I^S(\theta)$ не зависит от выбора меры λ. Действительно, пусть $\widetilde{\lambda}$ — какая-нибудь другая мера и $\nu = \lambda + \widetilde{\lambda}$. Тогда λ и $\widetilde{\lambda}$ абсолютно непрерывны относительно ν, а плотность $g^\nu_\theta(s)$ распределения S относительно меры ν равна

$$g^\nu_\theta(s) = g_\theta(s)\frac{d\lambda}{d\nu} = \widetilde{g}_\theta(s)\frac{d\widetilde{\lambda}}{d\nu},$$

где $\widetilde{g}_\theta$ — плотность относительно $\widetilde{\lambda}$. Так как $\dfrac{d\lambda}{d\nu}$ и $\dfrac{d\widetilde{\lambda}}{d\nu}$ от θ не зависят, то производные от логарифмов всех трех выражений будут совпадать.

Теорема 1. *Пусть плотности $f_\theta(x)$ и $g_\theta(s)$ удовлетворяют условиям* (R). *Тогда*

$$I^S(\theta) \leqslant I^X(\theta). \tag{1}$$

Равенство здесь достигается тогда и только тогда, когда S есть достаточная статистика.

Доказательство. Для любого $B \in \mathfrak{B}^l$ обозначим через $S^{-1}(B) \in \mathfrak{B}^n_{\mathcal{X}}$ множество $\boldsymbol{x} \in \mathcal{X}^n$, для которых $S(x) \in B$. Тогда по определению у. м. о.

$$\int\limits_{S^{-1}(B)} L'(\boldsymbol{x},\theta)\mathbf{P}_\theta(d\boldsymbol{x}) = \mathbf{E}_\theta[L'(X,\theta);\, X \in S^{-1}(B)] = \\ = \mathbf{E}_\theta[\mathbf{E}_\theta(L'(X,\theta)/S);\, S \in B]. \tag{2}$$

С другой стороны,

$$\int\limits_{S^{-1}(B)} L'(\boldsymbol{x},\theta)\mathbf{P}_\theta(d\boldsymbol{x}) = \frac{\partial}{\partial\theta}\int\limits_{S^{-1}(B)} f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \frac{\partial}{\partial\theta}\int\limits_B g_\theta(s)\lambda(ds) = \\ = \int\limits_B \frac{\partial}{\partial\theta} g_\theta(s)\lambda(ds) = \mathbf{E}_\theta[(\log g_\theta(S))';\, S \in B]. \tag{3}$$

Сравнивая (2) и (3), видим, что п. в. $[\mathbf{P}_\theta]$

$$\mathbf{E}_\theta(L'(X,\theta)/S) = (\log g_\theta(S))'. \tag{4}$$

Далее, имеем

$$0 \leqslant \mathbf{E}_\theta[L'(X,\theta) - (\log g_\theta(S))']^2 = I^X(\theta) + I^S(\theta) - 2\mathbf{E}_\theta L'(X,\theta)(\log g_\theta(S))',$$

где в силу (4)

$$\mathbf{E}_\theta L'(X,\theta)(\log g_\theta(S))' =$$
$$= \mathbf{E}_\theta[(\log g_\theta(S))'\mathbf{E}_\theta(L'(X,\theta)/S)] = \mathbf{E}_\theta[(\log g_\theta(S))']^2 = I^S(\theta).$$

Это доказывает неравенство (1).

Пусть теперь S — достаточная статистика для θ. Тогда

$$f_\theta(X) = \psi(S,\theta)h(X). \tag{5}$$

Возьмем в качестве λ меру

$$\lambda(B) = \int\limits_{S^{-1}(B)} h(x)\mu^n(dx).$$

Тогда, как показано в лемме 25.1, распределение S будет абсолютно непрерывно относительно λ и будет иметь плотность $g_\theta(s)$, равную $g_\theta(s) = \psi(s,\theta)$. Отсюда в силу (5) получаем

$$I^X(\theta) = \mathbf{E}_\theta[L'(X,\theta)]^2 = \mathbf{E}_\theta[(\log\psi(S,\theta))']^2 = I^S(\theta).$$

Покажем теперь, что из равенств $I^X(\theta) = I^S(\theta)$ при всех θ следует, что S — достаточная статистика. Действительно, $I^X(\theta)$ есть дисперсия $L'(X,\theta)$, так что

$$I^X(\theta) = \mathbf{E}_\theta[L'(X,\theta) - \mathbf{E}_\theta(L'(X,\theta)/S)]^2 + \mathbf{E}_\theta[\mathbf{E}_\theta(L'(X,\theta)/S)]^2. \tag{6}$$

Но в силу (4) последнее слагаемое равно

$$\mathbf{E}_\theta[(\log g_\theta(S))']^2 = I^S(\theta).$$

Так как $I^X(\theta) = I^S(\theta)$, то в (6) п. в. $[\mathbf{P}_\theta]$ при всех θ

$$L'(X,\theta) - \mathbf{E}_\theta(L'(X,\theta)/S) = 0.$$

То есть $L'(X,\theta)$ измерима относительно $\sigma(S)$ и, стало быть, существует измеримая функция $\varphi(S,\theta)$ такая, что

$$L'(X,\theta) = \varphi(S,\theta), \quad L(X,\theta) = \Phi(S,\theta) + h_1(X),$$
$$f_\theta(X) = \exp[\Phi(S,\theta) + h_1(X)]. \quad \triangleleft$$

Мы уже отмечали, что достаточные статистики являются тем единственным типом статистик, которые сокращают выборочные данные без потери информации о параметре θ. Теорема 1 придает этому утверждению точный смысл применительно к информации Фишера.

Пример 1. Пусть $X \in \mathbf{B}_p$. Здесь

$$f_p(x) = p^x(1-p)^{1-x},$$

где x равно нулю или единице, $f_p(x)$ есть плотность относительно считающей меры. Поэтому

$$l(x,p) = x\ln p + (1-x)\ln(1-p), \quad l'(x,p) = \frac{x}{p} - \frac{1-x}{1-p},$$

$$I(p) = \mathbf{E}_p[l'(\mathrm{x}_1,p)]^2 = p\left(\frac{1}{p}\right)^2 + (1-p)\left(\frac{1}{1-p}\right)^2 = \frac{1}{p(1-p)}.$$

Таким образом, информация одного наблюдения в схеме Бернулли равна $(p(1-p))^{-1}$ и достигает своего наименьшего значения при $p = 1/2$.

Информация всей выборки равна $n/(p(1-p))$. Обозначим теперь через ν число «успехов» в выборке X (число единичных исходов) и найдем информацию этого наблюдения. Плотности (опять относительно считающей меры) для ν будут равны

$$g_p(x) = C_n^x p^x (1-p)^{n-x}, \quad x = 0,1,\ldots,n,$$

так что $\log g_p(x) = x\log p + (n-x)\log(1-p) + \log C_n^x$,

$$I^\nu(p) = \mathbf{E}_p[(\log g_p(\nu))']^2 =$$

$$= \sum_{x=0}^{n} C_n^x p^x (1-p)^{n-x}\left(\frac{x}{p} - \frac{n-x}{1-p}\right)^2 = \sum_{x=0}^{n} C_n^x p^x (1-p)^{n-x}\frac{(x-np)^2}{(p(1-p))^2} =$$

$$= \frac{1}{(p(1-p))^2}\mathbf{D}\nu = \frac{n}{p(1-p)}.$$

Это равенство полностью согласуется с теоремой 1.

Мы предлагаем читателю найти в виде упражнений информации наблюдений для выборок из распределений, зависящих от одномерного параметра и приведенных в § 12.

2. Многомерный случай. Пусть теперь $\theta \in \mathbb{R}^k$, $k > 1$. В этом случае будем иметь дело с информационной матрицей Фишера наблюдения x_1

$$I(\theta) = \|I_{ij}(\theta)\|, \quad I_{ij}(\theta) = \mathbf{E}_\theta \frac{\partial}{\partial\theta_i} l(\mathrm{x}_1,\theta)\frac{\partial}{\partial\theta_j} l(\mathrm{x}_1,\theta),$$

где предполагается, конечно, что функция $f_\theta(x)$ дифференцируема.

Если положить

$$\varphi(x,\theta) = (\varphi_1(x,\theta),\ldots,\varphi_k(x,\theta)) =$$

$$= 2(\sqrt{f_\theta(x)})' = \frac{1}{\sqrt{f_\theta(x)}}\left(\frac{\partial f_\theta(x)}{\partial\theta_1},\ldots,\frac{\partial f_\theta(x)}{\partial\theta_k}\right),$$

то матрицу $I(\theta)$ можно записать также в виде

$$I(\theta) = \int_{\mathcal{X}} \varphi^{\top}(x,\theta)\varphi(x,\theta)\mu(dx).$$

Мы уже установили в § 26, что, как и в одномерном случае, информация Фишера аддитивна, т. е. информационная матрица Фишера выборки X равна сумме информационных матриц отдельных наблюдений. Если обозначить

$$I^X(\theta) = \left\|I^X_{ij}(\theta)\right\|, \quad I^X_{ij}(\theta) = \mathbf{E}_\theta \frac{\partial}{\partial\theta_i} L(X,\theta) \frac{\partial}{\partial\theta_j} L(X,\theta),$$

то $I^X(\theta) = nI(\theta)$.

Теорема 1 также сохраняется полностью. Пусть $g_\theta(s)$ — плотность некоторой статистики $S = S(X)$ со значениями в $\mathbb{R}^l$ относительно некоторой меры λ. Положим

$$I^S(\theta) = \left\|I^S_{ij}(\theta)\right\|, \quad I^S_{ij}(\theta) = \mathbf{E}_\theta \frac{\partial}{\partial\theta_i} \log g_\theta(S) \frac{\partial}{\partial\theta_j} \log g_\theta(S).$$

Это есть информационная матрица наблюдения S.

Теорема 1А. *Если плотности $f_\theta(x)$ и $g_\theta(s)$ удовлетворяют условиям* (R) § 26, *то*

$$I^S(\theta) \leqslant I^X(\theta), \tag{7}$$

т. е. матрица $I^X(\theta) - I^S(\theta)$ является неотрицательно определенной. Равенство в (7) *имеет место тогда и только тогда, когда S — достаточная статистика.*

Доказательство этой теоремы вполне аналогично доказательству теоремы 1, и для сокращения текста мы его опускаем. Это доказательство можно найти, например, в [48, 49].

Пример 2. В § 26 мы уже вычисляли информационную матрицу для нормального распределения. Вычислим теперь информационную матрицу для двупараметрического семейства распределений

$$f_\theta(x) = \frac{1}{\sigma} f\left(\frac{x-\alpha}{\sigma}\right),$$

где $\theta = (\alpha,\sigma)$, f — заданная дифференцируемая функция, для которой существуют интегралы

$$I_i = \int x^i \frac{(f'(x))^2}{f(x)}\,dx = \mathbf{E}_{(0,1)} \mathrm{x}_1^i (l'(\mathrm{x}_1))^2, \quad i = 0,1,2.$$

Здесь $l(x) = \log f(x)$, штрих означает обычное дифференцирование, параметры α и σ суть параметры *сдвига* и *масштаба* распределения с плотностью $f(x)$. Таким образом, нам известен вид распределения, но лишь с точностью до линейного преобразования аргумента. Параметры α и σ нормального

распределения Φ_{α,σ^2}, очевидно, являются параметрами сдвига и масштаба. Параметр α Г-распределения при фиксированном λ является параметром масштаба, как и параметр θ в распределении $\mathbf{U}_{0,\theta}$.

Имеем

$$l(x,\theta) = \log f_\theta(x) = -\log\sigma + l\left(\frac{x-\alpha}{\sigma}\right),$$

$$\frac{\partial l(x,\theta)}{\partial\alpha} = -\frac{1}{\sigma}l'\left(\frac{x-\alpha}{\sigma}\right),$$

$$\frac{\partial l(x,\theta)}{\partial\sigma} = -\frac{1}{\sigma} - \frac{(x-\alpha)}{\sigma^2}l'\left(\frac{x-\alpha}{\sigma}\right).$$

Отсюда находим

$$I_{11}(\theta) = \frac{1}{\sigma^2}\mathbf{E}_\theta\left[l'\left(\frac{\mathrm{x}_1-\alpha}{\sigma}\right)\right]^2 = \frac{1}{\sigma^2}\int\frac{\left[f'\left(\frac{x-\alpha}{\sigma}\right)\right]^2}{\sigma f\left(\frac{x-\alpha}{\sigma}\right)}dx = \frac{1}{\sigma^2}I_0,$$

$$I_{12}(\theta) = \frac{1}{\sigma^2}\mathbf{E}_\theta l'\left(\frac{\mathrm{x}_1-\alpha}{\sigma}\right)\left[1+\frac{\mathrm{x}_1-\alpha}{\sigma}l'\left(\frac{\mathrm{x}_1-\alpha}{\sigma}\right)\right] = \frac{1}{\sigma^2}I_1,$$

$$I_{22}(\theta) = \frac{1}{\sigma^2}\mathbf{E}_\theta\left[1+\frac{\mathrm{x}_1-\alpha}{\sigma}l'\left(\frac{\mathrm{x}_1-\alpha}{\sigma}\right)\right]^2 = \frac{1}{\sigma^2}[I_2-1],$$

так как $2\int\frac{x-\alpha}{\sigma}f'\left(\frac{x-\alpha}{\sigma}\right)\frac{dx}{\sigma} = -2\int f(x)\,dx = -2$. Таким образом,

$$I(\theta) = \frac{1}{\sigma^2}\begin{Vmatrix} I_0 & I_1 \\ I_1 & I_2-1 \end{Vmatrix}.$$

Если f — симметричная функция, то, очевидно, $I_1 = 0$.

Вырождение матрицы $I(\theta)$ означает, что ее определитель обращается в нуль или, что то же,

$$[\mathbf{E}_{(0,1)}l'(\mathrm{x}_1)(1+\mathrm{x}_1l'(\mathrm{x}_1))]^2 = \mathbf{E}_{(0,1)}(l'(\mathrm{x}_1))^2\mathbf{E}_{(0,1)}(1+\mathrm{x}_1l'(\mathrm{x}_1))^2.$$

Это возможно лишь в случае, когда либо $1+xl'(x) = cl'(x)$ при каком-нибудь c, либо $l'(x) = 0$. Из первого равенства следует, что

$$l(x) = -\ln(x-c) + c_1, \quad f(x) = \frac{e^{c_1}}{x-c}.$$

Ясно, что такая функция $f(x)$ не может быть плотностью распределения. Аналогично рассматривается возможность $l'(x) = 0$. Стало быть, $I(\theta)$ положительно определена.

В частности, для нормального семейства $[\Phi_{\alpha,\sigma^2}]$ при $\theta = (\alpha, \sigma)$

$$I(\theta) = \frac{1}{\sigma^2}\begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix},$$

так как в этом случае $l(x) = -x^2/2 - \ln\sqrt{2\pi}$, $l'(x) = -x$, $I_0 = \mathbf{E}_{(0,1)}\mathrm{x}_1^2 = 1$, $I_1 = \mathbf{E}_{(0,1)}\mathrm{x}_1^3 = \theta$, $I_2 = \mathbf{E}_{(0,1)}\mathrm{x}_1^4 = 3$. Тот же самый результат мы могли бы получить с помощью примера 26.4, используя п. 3 этого параграфа, где выяснено поведение информационной матрицы при замене параметра (в примере 26.4 $\theta = (\alpha, \sigma^2)$, а не (α, σ)). Предлагаем читателю убедиться теперь, что в соответствии с теоремой 1А статистика $\left(\overline{\mathrm{x}}, \sum \mathrm{x}_i^2\right)$ имеет информационную матрицу

$$I^S(\theta) = \frac{n}{\sigma^2}\begin{Vmatrix} 1 & 0 \\ 0 & 2 \end{Vmatrix} = nI(\theta).$$

3. Матрица Фишера и замена параметра. Рассмотрим вопрос о том, как ведет себя информационная матрица при замене параметра. Положим $\theta = v(\beta)$, $\beta \in \mathbb{R}^k$, где v — дифференцируемая вектор-функция, и рассмотрим параметрическое семейство $\mathbf{P}_\beta^{(1)} = \mathbf{P}_{v(\beta)}$. Чтобы найти информационную матрицу $J(\beta)$ для этого семейства, мы должны найти производные

$$\frac{\partial}{\partial\beta_j} l(\mathrm{x}_1, v(\beta)) = \sum_{i=1}^{k} \frac{\partial}{\partial\theta_i} l(\mathrm{x}_1, v(\beta)) \frac{\partial v_i(\beta)}{\partial\beta_j}. \tag{8}$$

Если обозначить $V = \left\| \dfrac{\partial v_i(\beta)}{\partial\beta_j} \right\|$, $i, j = 1, \ldots, k$, то мы получим, что вектор производных в (8) $l'_\beta(\mathrm{x}_1, v(\beta))$ представим в виде $l'_\theta(\mathrm{x}_1, v(\beta))V$, так что

$$J(\beta) = \mathbf{E}_\beta \big(l'_\theta(\mathrm{x}_1, v(\beta))V\big)^\top \big(l'_\theta(\mathrm{x}_1, v(\beta))V\big) = V^\top I(v(\beta))V.$$

В частности, если $\theta = \beta C$, $C = \|c_{ij}\|$, $i, j = 1, \ldots, k$, то $V = C^\top$ и

$$J(\beta) = CI(\theta)C^\top. \tag{9}$$

Отметим, что если мы рассмотрим в параметрическом пространстве эллипсоид

$$(\theta - \theta_1)I(\theta)(\theta - \theta_1)^\top < c, \tag{10}$$

то запись (10) этого множества инвариантна относительно линейного обратимого преобразования C над параметром θ. Именно, если положим $\theta = \beta C$, то в новых переменных множество (10) будет иметь вид

$$(\beta - \beta_1)J(\beta)(\beta - \beta_1)^\top < c,$$

где $\beta_1 = \theta_1 C^{-1}$. Это немедленно получится, если подставить $\theta = \beta C$ в (10) и воспользоваться (9).

§ 28.* Оценки параметра сдвига и масштаба. Эффективные эквивариантные оценки

Мы видели в § 22–26 и будем убеждаться в дальнейшем, насколько полезным является понятие достаточной статистики вообще и при построении эффективных оценок в частности. Круг идей, связанный с использованием достаточных статистик, можно было бы назвать *принципом достаточности.*

Принцип достаточности мы сочетали при построении эффективных оценок с другим — *принципом несмещенности.* Он состоит в выделении классов оценок с фиксированным и, в частности, с нулевым смещением. Без фиксации смещения выделение эффективных оценок было бы невозможно.

В этом и следующем параграфах (а также в главе 3) мы рассмотрим еще один важный принцип математической статистики — *принцип инвариантности.*

Смысл введения всех названных принципов один и тот же — они позволяют *естественным образом* сужать класс рассматриваемых оценок так, что в полученных сужениях оказывается возможным отыскание эффективных оценок.

1. Оценки параметра сдвига и масштаба. Задачей об оценке параметра сдвига называется задача оценки параметра α в семействе распределений $\{\mathbf{P}_\alpha\}$, обладающих свойством

$$\mathbf{P}_\alpha(A) = \mathbf{P}(A - \alpha).$$

Здесь $\mathbf{P}$ — некоторое фиксированное распределение, $A-\alpha = \{x : x+\alpha \in A\}$, и предполагается, что параметрическое множество Θ имеет ту же природу, что и $\mathcal{X}$. В случае $\mathcal{X} = \mathbb{R}^m$ можно, конечно, рассматривать сдвиги θ и «меньшей размерности», например скалярные, но тогда надо фиксировать направление (вектор $e \in \mathcal{X}$) сдвига и рассматривать $\mathbf{P}_\alpha(A) = \mathbf{P}(A - \alpha e)$. Будем рассматривать для определенности лишь первую возможность и считать, что $\Theta = \mathcal{X} = \mathbb{R}^m$.

Заметим, что $\mathbf{P}_\alpha$-распределение $\mathrm{x}_i + c$ ($c \in \mathbb{R}^m$) совпадает с распределением $\mathbf{P}_{\alpha+c}$ величины x_i, т. е. сдвиг всех наблюдений на c приводит к выборке из распределения $\mathbf{P}_{\alpha+c}$. Это делает естественным рассматривать лишь те оценки $\alpha^* = \alpha^*(X)$ параметра α, которые обладают свойством

$$\alpha^*(X + c) = \alpha^*(X) + c. \tag{1}$$

Здесь и в дальнейшем $X + c$ означает вектор с координатами $(\mathrm{x}_1 + c, \ldots$ $\ldots, \mathrm{x}_n + c)$. Нарушение этого равенства означало бы, что оценка α^* зависит от начала отсчета, т. е. от выбора начала координат в пространстве $\mathcal{X} = \mathbb{R}^m$.

Аналогичный подход возникает при оценке параметра масштаба, когда оценивается параметр σ в семействе $\{\mathbf{P}_\sigma\}$ со свойством $\mathbf{P}_\sigma(A) = \mathbf{P}(A/\sigma)$, $\sigma \in (0, \infty)$. Мы считаем здесь σ скалярным, хотя можно рассматривать и матричный случай. В этом случае $\mathbf{P}_\sigma$-распределение значений $\mathrm{x}_i c$ совпадает с распределением $\mathbf{P}_{\sigma c}$ величин x_i, т. е. умножение наблюдений на c приводит к выборке из $\mathbf{P}_{\sigma c}$. Стало быть, в этом случае естественно ограничиться рассмотрением оценок, обладающих свойством

$$\sigma^*(Xc) = c\sigma^*(X), \tag{2}$$

где $Xc = (\mathrm{x}_1 c, \ldots, \mathrm{x}_n c)$, поскольку при изменении в c раз масштаба наблюдений во столько же раз меняется и параметр.

Следующие утверждения читатель может легко получить самостоятельно.

Если семейство $\mathbf{P}_\theta$ *удовлетворяет условию* (A_μ), *то параметр* θ *будет параметром сдвига (масштаба) тогда и только тогда, когда*

$$f_\theta(x) = f(x-\theta) \quad \left(f_\theta(x) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right)\right).$$

Если $\mathcal{X} = \mathbb{R} = \Theta$, $X \in \mathbf{P}_\alpha$ и α есть параметр сдвига, то $Y = e^X = (e^{\mathrm{x}_1}, \ldots, e^{\mathrm{x}_n}) \in \mathbf{Q}_\sigma$, где для распределений $\mathbf{Q}_\sigma$ параметр $\sigma = e^\alpha$ есть параметр масштаба. Это сразу следует из того, что плотность $\mathrm{y}_1 = e^{\mathrm{x}_1}$ равна (см. [17, § 3.2])

$$\frac{1}{y} f(\ln y - \alpha) = \frac{1}{\sigma}\left[\frac{\sigma}{y} f\left(\ln \frac{y}{\sigma}\right)\right].$$

Наоборот, если $\mathcal{X} = (0, \infty) = \Theta$, $X \in \mathbf{P}_\sigma$ и σ есть параметр масштаба, то $Y = \ln X = (\ln \mathrm{x}_1, \ldots, \ln \mathrm{x}_n) \in \mathbf{Q}_\alpha$, где $\alpha = \ln \sigma$ есть параметр сдвига распределений $\mathbf{Q}_\alpha$. Таким образом, задачи оценивания параметров сдвига и масштаба могут быть сведены одна к другой.

Можно рассматривать задачу и об одновременной оценке неизвестных параметров α и σ в случае, когда $\mathbf{P}_{\alpha,\sigma}(A) = \mathbf{P}\left(\dfrac{A-\alpha}{\sigma}\right)$. В этих условиях в качестве оценки σ естественно рассматривать функции, обладающие свойством

$$\alpha^*(X + c) = \alpha^*(X), \quad \sigma^*(Xc) = c\sigma^*(X). \tag{3}$$

Оценки, удовлетворяющие в рассмотренных примерах условиям (1)–(3), называются *эквивариантными* (общее определение см. в § 29). Смысл введения таких оценок состоит в сужении класса всех рассматриваемых оценок, что упрощает задачу поиска оптимальных оценок. Так, в § 18 нами было установлено, что найти равномерно (т. е. при всех θ) наилучшие оценки в классе всех оценок невозможно. Оказывается, что в классе эквивариантных оценок такие равномерно наилучшие оценки уже существуют и могут быть в ряде случаев найдены в явном виде. На примере оценок параметров сдвига и масштаба мы этот факт и проиллюстрируем.

2. Эффективная оценка параметра сдвига в классе эквивариантных оценок. Будем предполагать здесь, что выполнено условие (A_μ) и, стало быть, $f_\alpha(x) = f(x-\alpha)$ и что μ есть мера Лебега.

Обозначим через S_0 статистику

$$S_0 = S_0(X) = (\mathrm{x}_2 - \mathrm{x}_1, \ldots, \mathrm{x}_n - \mathrm{x}_1),$$

очевидно, инвариантную относительно сдвига: $S_0(X+c) = S_0(X)$. Через K_E обозначим класс всех эквивариантных оценок α^*, т. е. оценок, удовлетворяющих (1), и через $|\alpha|^2$ — квадрат евклидовой нормы $\alpha \in \mathbb{R}^m$.

Теорема 1. *Пусть* $\alpha^* = \alpha^*(X)$ — *любая эквивариантная оценка с конечным значением* $\mathbf{E}_0\alpha^*$. *Тогда оценка*

$$\alpha_0^* = \alpha^* - \mathbf{E}_0(\alpha^*/S_0) \tag{4}$$

от выбора α^* *не зависит и является единственной эффективной в* K_E *оценкой, т. е.* $\mathbf{E}_\alpha|\alpha_0^* - \alpha|^2 = \min\limits_{\alpha^* \in K_E} \mathbf{E}_\alpha|\alpha^* - \alpha|^2$ *при всех* α *и* $\mathbf{E}_\alpha|\alpha^* - \alpha|^2 = \mathbf{E}_\alpha|\alpha_0^* - \alpha|^2$, *если только* $\mathbf{E}_0(\alpha^*/S_0) = 0$ *п. в. Оценка* α_0^* *может быть представлена в форме*

$$\alpha_0^* = \frac{\int u f_u(X)\,du}{\int f_u(X)\,du} = \frac{\int u f(X-u)\,du}{\int f(X-u)\,du}. \tag{5}$$

Оценка α_0^* называется *оценкой Питмена.* Из (4) нетрудно видеть, что она является эквивариантной и несмещенной. Эквивариантность следует из эквивариантности α^* и инвариантности относительно сдвига функции $V(S_0) = \mathbf{E}_0(\alpha^*/S_0)$, зависящей только от S_0. Несмещенность вытекает из равенств

$$\mathbf{E}_\alpha\alpha_0^* = \alpha + \mathbf{E}_\alpha\alpha^*(X-\alpha) - \mathbf{E}_\alpha V(S_0), \tag{6}$$

где $\mathbf{E}_\alpha V(S_0) = \mathbf{E}_0 V(S_0)$, $\mathbf{E}_\alpha\alpha^*(X-\alpha) = \mathbf{E}_0\alpha^*(X)$. Последнее соотношение вытекает из того, что $X - \alpha \in \mathbf{P}_0$, если $X \in \mathbf{P}_\alpha$. Поэтому сумма двух последних слагаемых в (6) равна

$$\mathbf{E}_0\alpha^* - \mathbf{E}_0[\mathbf{E}_0(a^*/S_0)] = 0, \quad \mathbf{E}_\alpha\alpha_0^* = \alpha.$$

Доказательству теоремы предпошлем следующее вспомогательное утверждение.

Лемма 1. *Пусть* $X \in \mathbf{P}_0$. *Для любой статистики* $S = S(X)$ *с конечным математическим ожиданием* $\mathbf{E}_0|S| < \infty$ *у. м. о.* S *относительно* S_0 *равно*

$$\mathbf{E}_0(S/S_0) = S_1(X) \equiv \frac{\int S(X-u) f_u(X)\,du}{\int f_u(X)\,du}. \tag{7}$$

Доказательство. Все функции под знаками интегралов в (7) есть функции от $X-u$. Стало быть, после замены $x_1 - u = v$ они будут функциями от $(v, x_2 - x_1 + v, \ldots, x_n - x_1 + v)$. Это означает, что правая часть (7) зависит лишь от S_0. В силу свойств у. м. о. нам для доказательства леммы достаточно убедиться, что для любого $A \in \sigma(S_0)$

$$\mathbf{E}_0(S_1; A) = \mathbf{E}_0(S; A). \tag{8}$$

Пусть $Z = Z(S_0)$ — любая ограниченная $\sigma(S_0)$-измеримая статистика, тогда

$$\mathbf{E}_0 Z S_1 = \int\limits_{\mathcal{X}^n} \frac{Z(S_0)\int\limits_\Theta S(\boldsymbol{x}-u)f_u(\boldsymbol{x})\,du}{\int\limits_\Theta f_u(\boldsymbol{x})\,du} f(\boldsymbol{x})\,d\boldsymbol{x} =$$

$$= \int\limits_\Theta \int\limits_{\mathcal{X}^n} \frac{Z(S_0)S(\boldsymbol{x}-u)f(\boldsymbol{x}-u)f(\boldsymbol{x})}{\int\limits_\Theta f(\boldsymbol{x}-v)\,dv}\,d\boldsymbol{x}\,du.$$

После замены $\boldsymbol{x}-u \to \boldsymbol{x}$ во внутреннем интеграле получим ($S_0(x)$ при этом перейдет в себя)

$$\int\limits_\Theta \int\limits_{\mathcal{X}^n} \frac{Z(S_0)S(\boldsymbol{x})f(\boldsymbol{x})f(\boldsymbol{x}+\boldsymbol{u})}{\int\limits_\Theta f(\boldsymbol{x}+u-v)\,dv}\,d\boldsymbol{x}du = \int\limits_{\mathcal{X}^n} Z(S_0)S(\boldsymbol{x})f(\boldsymbol{x})\,d\boldsymbol{x} = \mathbf{E}_0 Z S.$$

Это доказывает (8). Изменение порядка интегрирования, которое мы дважды производили, законно в силу абсолютной интегрируемости S и ограниченности Z. $\triangleleft$

Доказательство теоремы 1. Заметим прежде всего, что для эквивариантной оценки $\mathbf{E}_\alpha|\alpha^* - \alpha|^2$ не зависит от α. Действительно,

$$\mathbf{E}_\alpha|\alpha^*(X) - \alpha|^2 = \mathbf{E}_\alpha|\alpha^*(X-\alpha)|^2 = \mathbf{E}_0|\alpha^*(X)|^2.$$

Таким образом, для отыскания равномерно оптимальной эквивариантной оценки надо найти оценку α^*, минимизирующую $\mathbf{E}_0|\alpha^*|^2$.

Пусть α^* — любая эквивариантная оценка α. В силу свойств у. м. о.

$$\mathbf{E}_0|\alpha^*|^2 = \mathbf{E}_0|\alpha^* - \mathbf{E}_0(\alpha^*/S_0)|^2 + \mathbf{E}_0|\mathbf{E}_0(\alpha^*/S_0)|^2 \geqslant \mathbf{E}_0|\alpha^* - \mathbf{E}_0(\alpha^*/S_0)|^2. \quad (9)$$

Остается заметить, что в силу леммы 1 оценка $\alpha_0^* = \alpha^* - \mathbf{E}_0(\alpha^*/S_0)$ равна (5) и от выбора α^* не зависит. Равенство в (9), очевидно, возможно тогда и только тогда, когда $\mathbf{E}_0(\alpha^*/S_0) = 0$ п. в. $\triangleleft$

Из доказательства теоремы видно, что особую роль в построении оптимальной эквивариантной оценки играет статистика $S_0 = (x_2 - x_1, \ldots, x_n - x_1)$, инвариантная относительно преобразования сдвига. Инвариантность статистики есть качество, в известном смысле противоположное достаточности, а конструкция оценки $\theta_0^* = \theta^* - \mathbf{E}_0(\theta^*/S_0)$ на основе произвольной оценки θ^* представляет собой подход к улучшению оценки θ^*, также в некотором смысле противоположный подходу, при котором для улучшения оценки θ^* с помощью достаточной статистики S рассматривалась оценка $\theta_S^* = \mathbf{E}_\theta(\theta^*/S)$. Противоположность состоит в следующем. Достаточная статистика содержит в себе *всю* информацию о параметре θ; инвариантная статистика — *никакой*. В поисках наилучших оценок мы искали минимальные достаточные статистики; здесь, как мы увидим, нам нужны максимальные инвариантные статистики (таковой является статистика S_0). Оценка θ_S^* есть

«проекция» θ^* на S, в то время как оценка θ_0^* получается вычитанием из θ^* ее «проекции» на S_0.

В конечном счете результаты, полученные этими двумя путями, часто совпадают, как это будет видно из следующих двух примеров.

Пример 1. Пусть $\mathcal{X} = \mathbb{R}$, $X \in \mathbf{\Phi}_{\alpha,1}$. Тогда

$$f_\alpha(X) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum(\mathrm{x}_i - \alpha)^2\right\} =$$

$$= \frac{1}{\sqrt{n}\,(2\pi)^{(n-1)/2}} \exp\left\{\frac{1}{2}\sum(\mathrm{x}_i - \overline{\mathrm{x}})^2\right\} \cdot \sqrt{\frac{n}{2\pi}} \exp\left[-\frac{n}{2}(\alpha - \overline{\mathrm{x}})^2\right].$$

Здесь второй множитель как функция α есть функция плотности нормального закона с параметрами $(\overline{\mathrm{x}}, 1/n)$. Так как первый множитель от α не зависит, то он сократится в (5), и оценка Питмена будет равна $\alpha^* = \overline{\mathrm{x}}$. Таким же будет результат в многомерном случае.

Пример 2. Пусть $\mathcal{X} = \mathbb{R}$, $X \in \mathbf{U}_{\theta,1+\theta}$. Тогда

$$f_\theta(X) = \begin{cases} 1 & \text{при } \mathrm{x}_{(n)} - 1 \leqslant \theta \leqslant \mathrm{x}_{(1)}, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Поэтому в силу (5)

$$\theta^* = \int\limits_{\mathrm{x}_{(n)}-1}^{\mathrm{x}_{(1)}} \frac{u\,du}{\mathrm{x}_{(1)} - \mathrm{x}_{(n)} + 1} = \frac{1}{2}(\mathrm{x}_{(1)} + \mathrm{x}_{(n)} - 1).$$

Мы видим, таким образом, что в классе K_E эквивариантных оценок можно в явном виде строить эффективные оценки, при этом не требуется никаких условий на гладкость $f_\theta(x)$, и сама эффективность носит точный (не асимптотический) характер.

3. Минимаксность оценки Питмена. Обратим теперь внимание на форму оценки Питмена. Грубо говоря, это есть байесовская оценка для «равномерного на всей оси» априорного распределения. Так как указанного распределения не существует, то мы сформулируем высказанное утверждение более точно. Пусть $\mathcal{X} = \mathbb{R}$ и $\mathbf{Q}^{(N)}$ есть равномерное распределение на $[-N, N]$, т. е. распределение с плотностью

$$q^{(N)}(t) = \begin{cases} (2N)^{-1}, & |t| \leqslant N, \\ 0, & |t| > N. \end{cases}$$

Байесовская оценка, соответствующая $\mathbf{Q}^{(N)}$, будет равна

$$\alpha^*_{Q^{(N)}} = \frac{\int u q^{(N)}(u) f_u(X)\,du}{\int q^{(N)}(u) f_u(X)\,du} = \frac{\int\limits_{-N}^{N} u f_u(X)\,du}{\int\limits_{-N}^{N} f_u(X)\,du}.$$

Очевидно, что для всех X оценка Питмена α_0^* есть предел $\alpha_0^* = \lim\limits_{N\to\infty} \alpha^*_{Q^{(N)}}$. Это обстоятельство наводит на мысль, что одновременно будут сходиться и моменты второго порядка:

$$\mathbf{E}_\alpha\bigl(\alpha^*_{Q^{(N)}} - \alpha\bigr)^2 \to \mathbf{E}_\alpha\bigl(\alpha_0^* - \alpha\bigr)^2.$$

Оказывается, что в области $|\alpha| \leqslant N - \sqrt{N}$ это так и есть, причем сходимость будет равномерной по α в указанном интервале значений α. (Доказательство связано с оценкой $\mathbf{E}_\alpha(\alpha_0^* - \alpha^*_{Q^{(N)}})^2$, носит в основном технический характер, и мы его опускаем.)

Но в таком случае мы можем воспользоваться критерием минимаксности оценок в теореме 21.3: *если оценка α^* такова, что при всех α*

$$\mathbf{E}_\alpha(\alpha^* - \alpha)^2 \leqslant \limsup_{N\to\infty} \int \mathbf{E}_t\bigl(\alpha^*_{Q^{(N)}} - t\bigr)^2 \mathbf{Q}^{(N)}(dt) \tag{10}$$

для некоторой последовательности априорных распределений $\mathbf{Q}^{(N)}$ (не обязательно равномерных) и соответствующих им байесовских оценок $\alpha^*_{Q^{(N)}}$, то α^* — минимаксная оценка.

В нашем случае $m = \mathbf{E}_\alpha\bigl(\alpha_0^* - \alpha\bigr)^2$ от α не зависит. Поэтому в силу отмеченных выше свойств сходимости вторых моментов

$$\limsup_{N\to\infty} \int \mathbf{E}_t\bigl(\alpha^*_{Q^{(N)}} - t\bigr)^2 \mathbf{Q}^{(N)}(dt) \geqslant$$

$$\geqslant \limsup_{N\to\infty} \frac{1}{2N} \int\limits_{|t|<N-\sqrt{N}} \mathbf{E}_t\bigl(\alpha^*_{Q^{(N)}} - t\bigr)^2 dt \geqslant$$

$$\geqslant \limsup_{N\to\infty} \frac{1}{2N} 2(N - \sqrt{N})(m - \varepsilon) = m - \varepsilon$$

при любом $\varepsilon > 0$. Это означает, что свойство (10) выполнено.

Таким образом, *оценка Питмена является минимаксной в классе всех оценок параметра сдвига* (то, что она минимаксна в классе инвариантных оценок, очевидно, следует из эффективности).

Сказанное можно интерпретировать также следующим образом: «наихудшим» априорным распределением (см. § 21) для параметра сдвига является «равномерное на всей оси» распределение.

Указанием на минимаксность оценки Питмена могла бы служить также отмеченная выше независимость $\mathbf{E}_\alpha\bigl(\alpha_0^* - \alpha\bigr)^2$ от α (ср. с теоремой 21.2).

4. Об оптимальных оценках параметра масштаба. Как мы уже отмечали, проблема оценки параметра масштаба σ может быть сведена в известном смысле к проблеме оценки параметра сдвига. Пусть для простоты $\mathcal{X} = (0, \infty) = \Theta$. Тогда если $X \in \mathbf{P}_\sigma$, $\mathbf{P}_\sigma(A) = \mathbf{P}(A/\sigma)$, то $Y = \ln X = (\ln \mathrm{x}_1, \ldots, \ln \mathrm{x}_n) \in \mathbf{P}_\alpha^{(1)}$, где $\alpha = \ln \sigma$, а распределение $\mathbf{P}_\alpha^{(1)}$ имеет плотность

величины $y_1 = \ln x_1$ в точке y (условие (A_μ) выполнено, $\dfrac{d\mathbf{P}_1(x)}{d\mu} = f(x)$), равную (см. [17, § 3.2])

$$f\left(\frac{e^y}{\sigma}\right)\frac{e^y}{\sigma} = f(e^{y-\alpha})e^{y-\alpha} = f^{(1)}(y-\alpha),$$

$$f^{(1)}(y) = f(e^y)e^y.$$

Таким образом, мы можем оценить наилучшим образом параметр α с помощью оценки Питмена $\alpha^* = \alpha^*(Y)$, а затем положить $\sigma^*(X) = e^{\alpha^*(Y)}$. Нетрудно видеть, что $\sigma^*(X)$ будет эквивариантной, так как

$$\sigma^*(cX) = e^{\alpha^*(Y+\ln c)} = e^{\alpha^*(Y)+\ln c} = c\sigma^*(X).$$

Однако здесь важно отметить, что оценка Питмена минимизирует $\mathbf{E}_\alpha(\alpha^* - \alpha)^2$. Стало быть, полученная оценка σ^* будет минимизировать

$$\mathbf{E}_\sigma\left(\ln\frac{\sigma^*}{\sigma}\right)^2, \tag{11}$$

а не величину $\mathbf{E}_\sigma(\sigma^* - \sigma)^2$, с которой мы обычно имели дело. Но в задаче об эквивариантной оценке параметра σ и неразумно было рассматривать среднеквадратичную погрешность, поскольку она, в отличие от (11), зависит от преобразования сжатия, примененного одновременно к σ^* и σ. Аналогом инвариантной статистики S_0 здесь будет статистика $(x_2/x_1, \ldots, x_n/x_1)$. Наряду с (11) возможно, конечно, рассмотрение и других погрешностей. Если, например, мы будем минимизировать величину

$$\mathbf{E}_\sigma\left(\frac{\sigma^*}{\sigma} - 1\right)^2,$$

наилучшей эквивариантной оценкой будет

$$\sigma^* = \frac{\int \sigma^{-n-2} f(X/\sigma)\, d\sigma}{\int \sigma^{-n-3} f(X/\sigma)\, d\sigma} \tag{12}$$

(см. [112, с. 191]).

Пример 3. Обнаружение источника излучения. Приведем пример реальной физической задачи, которая связана с оценками параметров сдвига и масштаба.

Предположим, что в некоторой неизвестной точке z трехмерного пространства находится источник γ-излучения. Задача состоит в том, чтобы, используя плоский детектор (пусть он совпадает с одной из координатных плоскостей) и фиксируя на нем следы излучения, т. е. следы взаимодействия γ-квантов, испускаемых точкой z, с чувствительной поверхностью детектора, определить координаты точки z.

Эта задача намного упростилась бы, если бы мы имели источник излучения заряженных частиц высокой энергии. Тогда можно было бы поставить один за другим два параллельных плоских детектора и фиксировать на них точки прохождения, т. е. взаимодействия с поверхностью экрана, всего лишь двух частиц. Это дало бы нам направления полета этих частиц и вместе с ними координаты точки z как точки пересечения этих направлений. Однако для слабого γ-излучения, используемого в рентгеноскопии, это неосуществимо, и возможно введение лишь одного детектора.

Направление, в котором распространяются излучаемые γ-кванты, случайно и имеет равномерное распределение на поверхности сферы (если направление определять точкой на сфере с центром в точке z).

Чтобы упростить задачу, рассмотрим ее двумерный вариант. Пусть источник находится на плоскости переменных (x, y) в неизвестной точке $z = (\alpha, \sigma)$, $\sigma > 0$. Угол направления излучения, составленный с осью Oy имеет равномерное распределение на $[0, 2\pi]$. Чувствительный детектор совпадает с осью абсцисс. Результатами наблюдений будут точки $\mathrm{x}_1, \mathrm{x}_2, \ldots$, в которых было зафиксировано взаимодействие γ-квантов с детектором (осью абсцисс).

Некоторое своеобразие этой задачи состоит в том, что объем n выборки, полученной за фиксированное время t будет случайным: число γ-квантов, испускаемых источником за время t, имеет распределение Пуассона, число γ-квантов, достигших детектора, также распределено по закону Пуассона, так как каждый квант достигает оси абсцисс с вероятностью $1/2$. Однако в нашем случае n и наблюдения $\mathrm{x}_1, \mathrm{x}_2, \ldots$ независимы. Поэтому мы можем рассматривать то число n наблюдений, которое получилось, и считать его фиксированным (для каждого такого фиксированного n распределение x_i будет одним и тем же).

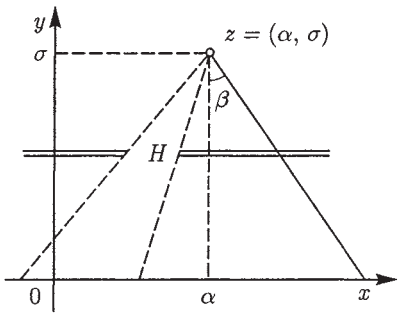

Рис. 2

Итак, пусть даны наблюдения $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Наша задача состоит в оценке координат (α, σ). Покажем, что $X \in \mathbf{K}_{\alpha,\sigma}$, т. е. x_i имеют распределение Коши с параметрами сдвига α и масштаба σ.

Действительно, условное распределение угла β направления движения γ-кванта с осью $(0, -y)$ при условии, что квант достиг детектора (ось абсцисс), будет равномерным на отрезке $[-\pi/2, \pi/2]$. Так как $(x - \alpha)/\sigma = \operatorname{tg}\beta$ (см. рис. 2), то

$$\mathbf{P}_{\alpha,\sigma}(\mathrm{x}_1 < x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg}\frac{x-\alpha}{\sigma}.$$

Стало быть, плотность распределения x_1 будет равна плотности распределения Коши (см. § 12)

$$k_{\alpha,\sigma}(x) = \frac{1}{\pi\sigma}\frac{1}{(1 + ((x-\alpha)/\sigma)^2)} = \frac{\sigma}{\pi(\sigma^2 + (x-\alpha)^2)}.$$

Предположим теперь, что σ известно, пусть, например, $\sigma = 1$. Тогда наилучшей инвариантной оценкой параметра сдвига α будет оценка Питмена,

которая получается как среднее значение $\alpha^* = \int u\varphi(u)\,du$ распределения с плотностью

$$\varphi(u) = \varphi(u, X) = \frac{k_u(X)}{\int k_v(X)\,dv}, \quad k_u(X) = \prod_{i=1}^{n} k_u(\mathrm{x}_i),$$

$$k_u(\mathrm{x}_i) = k_{u,1}(\mathrm{x}_i) = \frac{1}{\pi(1 + (u - \mathrm{x}_i)^2)}.$$

О. м. п. $\widehat{\alpha}^*$ будет представлять из себя точку, в которой достигается $\max\varphi(u)$. Позже мы покажем (см. § 34), что α^* и $\widehat{\alpha}^*$ асимптотически эквивалентны и имеют а. н. распределение с коэффициентом $1/I = 2$ (в рассматриваемом случае $I = \int (k_0')^2/k_0\,dx = 4\pi^{-1}\int x^2(1 + x^2)^{-3}\,dx = 1/2$). Из сказанного следует, что погрешность оценок α^* и $\widehat{\alpha}^*$ при больших n имеет порядок малости $1/\sqrt{n}$.

Интересно отметить, что в рассматриваемой задаче можно добиться путем вмешательства в эксперимент более высокой степени точности. Это можно сделать, поставив между точкой $z = (\alpha, 1)$ и детектором экран, параллельный оси абсцисс, с отверстием H, через которое лишь и могут проходить γ-кванты. Положение экрана и отверстия выбираются экспериментатором и, стало быть, нам известны.

В этом случае распределение наблюдений на экране будет разрывным и при небольших отверстиях H будет близко к $\mathbf{U}_{a\alpha, a\alpha+b}$ при некоторых известных нам постоянных a и b. Вид эффективной эквивариантной оценки α_H^* для такого распределения был найден в примере 2. Оценка α_H^* определяется крайними значениями выборки и имеет точность порядка $1/n_H$, где $n_H \leqslant n$ — число элементов выборки, которые соответствуют квантам, прошедшим через щель (n_H, как и n, на самом деле случайно и распределено по закону Пуассона). Так как в среднем n_H пропорционально n, то при достаточно больших n получим $1/n_H \ll 1/\sqrt{n}$.

§ 29.* Общая задача об эквивариантном оценивании

Рассмотрим группу G измеримых преобразований g пространства $\mathcal{X}^n$ в себя, обладающих следующими свойствами.

1. Каждое g отображает $\mathcal{X}^n$ на все пространство $\mathcal{X}^n$, т. е. для любого $x_2 \in \mathcal{X}^n$ найдется $x_1 \in \mathcal{X}^n$ такое, что $x_2 = gx_1$.

2. Отображения g взаимно однозначны.

Измеримость g нужна для того, чтобы gX была случайной величиной. Групповое свойство означает, что $g_2g_1 \in G$, если $g_1 \in G$, $g_2 \in G$; тождественное преобразование e и обратное преобразование g^{-1} принадлежат G (так что $g^{-1}g = e$).

Определение 1. Семейство распределений $\{\mathbf{P}_\theta\}$ называется *инвариантным относительно группы преобразований* G (или, для краткости, просто инвариантным), если для каждых $g \in G$ и $\theta \in \Theta$ существует единственное $\theta_g \in \Theta$ такое, что соотношение $X \in \mathbf{P}_\theta$ влечет за собой $gX \in \mathbf{P}_{\theta_g}$.

Значение θ_g, однозначно определяемое θ и g, обозначим через $\theta_g = \overline{g}\theta$. Тогда определение означает, что

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\overline{g}\theta}(X \in A).$$

Так как в силу определения 1 условие (A_0) выполнено, то множество $\overline{G}$ всех преобразований $\overline{g}$ пространства Θ в себя образует группу. Действительно, распределение $g_2 g_1 X$ дается одновременно распределениями $\mathbf{P}_{\overline{g_2 g_1}\theta}$ и $\mathbf{P}_{\overline{g}_2\overline{g}_1\theta}$. Из условия (A_0) следует, что $\overline{g_2 g_1} = \overline{g}_2\overline{g}_1$ и что $\overline{g}_1^{-1} \in \overline{G}$ (достаточно положить $g_2 = g_1^{-1}$). Преобразования $\overline{g}$ из $\overline{G}$ автоматически взаимно однозначны. Однако изоморфизма между G и $\overline{G}$ может и не быть. Пусть, например, $X \in \mathbf{\Phi}_{0,\sigma^2}$, $\sigma \in (0,\infty)$. В этом случае плотность $f_{0,\sigma^2}(X)$ (функция правдоподобия) зависит лишь от $\sum \mathrm{x}_i^2$. Стало быть, если в качестве G рассмотреть группу вращений (ортогональных преобразований $\mathcal{X}^n$), то условия определения 1 будут выполнены, но $\overline{g} = \overline{e}$, и группа $\overline{G}$ состоит из единственного элемента $\overline{e}$ — тождественного преобразования $\Theta = (0,\infty)$ в себя.

Мы предоставляем читателю возможность в качестве упражнения проверить, что если $\{\mathbf{P}_\theta\}$ инвариантно относительно группы G и G_1 есть подгруппа G, то $\{\mathbf{P}_\theta\}$ инвариантно относительно G_1.

При рассмотрении общей задачи эквивариантного оценивания нам необходимо иметь несколько более общую постановку задачи относительно сравнения оценок. До сих пор мы делали это с помощью среднеквадратичных уклонений, меряя погрешность оценки величиной $(\theta^* - \theta)^2$. Будем теперь предполагать, что измерение погрешности θ^* происходит с помощью функции $w(\theta^*, \theta)$ и что эта функция обладает свойством «однородности» *):

$$w(\overline{g}\theta, \overline{g}\theta^*) = w(\theta, \theta^*) \quad \text{при всех} \quad \theta. \tag{1}$$

Именно этим свойством обладают функции $w(\theta,\theta^*) = (\theta-\theta^*)^2$ для параметра сдвига (преобразование сдвига) и $w(\theta,\theta^*) = \left(\ln\dfrac{\theta}{\theta^*}\right)^2$ или $\left(\dfrac{\theta}{\theta^*} - 1\right)^2$ для параметра масштаба (преобразование сжатия).

Мы видели в п. 28.4, что задача отыскания наилучшей эквивариантной оценки может быть весьма чувствительной к выбору меры погрешности $w(\theta,\theta^*)$ оценки θ^*.

Обратимся теперь к проблеме оценивания для инвариантных семейств $\{\mathbf{P}_\theta\}$. Пусть мы имеем выборку X и по ней построили оценку $\theta^* = \theta^*(X)$ параметра θ. Если мы рассмотрим выборку $Y = gX \in \mathbf{P}_{\overline{g}\theta}$, то $\theta^*(Y)$ будет оценкой для $\overline{g}\theta$. При этом естественно предполагать, что оценки $\theta^*(X)$ и $\theta^*(Y)$ связаны между собой так же, как оцениваемые параметры θ и $\overline{g}\theta$, т. е. преобразованием $\overline{g}$

$$\theta^*(Y) = \overline{g}\theta^*(X). \tag{2}$$

*) Это свойство не обязательно в теории эквивариантного оценивания. Можно лишь требовать существования $\overline{g}\theta^*$ такого, что $w(\overline{g}\theta, \overline{g}\theta^*) = w(\theta,\theta^*)$ при всех θ (см. [112]).

В силу (1) оценка $\theta^*(Y)$ параметра $\overline{g}\theta$ дает ту же самую ошибку, что и оценка $\theta^*(X)$ параметра θ. Таким образом, имеем две «одинаковые» проблемы оценивания. Произведенные преобразования gX и $\overline{g}\theta$ можно интерпретировать как замены систем координат. Тогда (2) означает, что оценка θ^* не зависит от выбора системы координат и удовлетворяет соотношению

$$\theta^*(X) = \overline{g}^{-1}\theta^*(gX). \tag{3}$$

Иными словами, если выбрана θ^*, удовлетворяющая (2), то безразлично, какую из двух проблем оценивания, указанных выше, решать, поскольку выводы о $\overline{g}\theta$ во второй проблеме можно превратить в выводы о θ в первой с помощью равенства (3).

Определение 2. Оценка θ^* параметра θ инвариантного семейства $\mathbf{P}_\theta$, удовлетворяющая (3), называется *эквивариантной**).

Рассмотрим какую-нибудь точку $\theta_0 \in \Theta$ и множество «эквивалентных» точек $\theta = \overline{g}\theta_0$, $\overline{g} \in \overline{G}$. Такое выделение классов «эквивалентных» точек делит все пространство Θ на подмножества, называемые *орбитами*.

Теорема 1. *Значение* $\mathbf{E}_\theta w(\theta, \theta^*)$ *для эквивариантной оценки* θ^* *постоянно на орбите, т. е.*

$$\mathbf{E}_\theta w(\theta, \theta^*) = \mathbf{E}_{\overline{g}\theta} w(\overline{g}\theta, \theta^*)$$

при любых $\theta \in \Theta$ *и* $\overline{g} \in \overline{G}$.

Доказательство. Имеем

$$\mathbf{E}_\theta w(\theta, \theta^*(X)) = \mathbf{E}_\theta w(\overline{g}\theta, \overline{g}\theta^*(X)) = \mathbf{E}_\theta w(\overline{g}\theta, \theta^*(gX)) = \mathbf{E}_{\overline{g}\theta} w(\overline{g}\theta, \theta^*(X)). \quad \triangleleft$$

Если орбита $\{\theta : \theta = \overline{g}\theta_0,\ \overline{g} \in \overline{G}\}$ совпадает с Θ (как это имело место для параметров сдвига и масштаба), то $\mathbf{E}_\theta w(\theta, \theta^*) = \text{const}$ на Θ. Выполнение этого равенства есть характерный признак минимаксности θ^* (ср. с теоремой 21.2), так что наилучшие эквивариантные оценки часто оказываются минимаксными в классе всех оценок (подробнее см. [112]).

Из теорем § 21 вытекает, например,

Теорема 2. *Если* Θ *является орбитой и эквивариантная оценка* θ^* *оказалась байесовской (или пределом байесовских оценок* θ_N^* *со сходимостью* $\mathbf{E}_\theta w(\theta, \theta^*) = \lim\limits_{N\to\infty} \mathbf{E}_\theta w\big(\theta, \theta_N^*\big)$*), то* θ^* *— минимаксная оценка.*

Отметим также следующее важное свойство эквивариантных оценок. Нам будет удобно через $\nu(g\,d\boldsymbol{x})/\nu(d\boldsymbol{x})$ обозначить плотность меры ν_g, $\nu_g(B) = \nu(gB)$ относительно меры ν в точке $\boldsymbol{x} \in \mathcal{X}^n$.

Теорема 3. *Пусть выполнено условие* (A_μ) *и* $\mu^n(g\,d\boldsymbol{x})/\mu^n(d\boldsymbol{x})$ *конечно и положительно при каждом* $g \in G$ *и п. в.* $[\mu^n]$ *значений* $\boldsymbol{x}$. *Пусть,*

*) Такие оценки называют иногда инвариантными. Однако этот термин менее точен. Его лучше оставить для оценок, обладающих свойством $\theta^*(gX) = \theta^*(X)$ (т. е. для случая, когда $\overline{g} = \overline{e}$ для любого g).

кроме того, о. м. п. $\widehat{\theta}^*$ *единственна при каждом* X. *Тогда если семейство* $\mathbf{P}_\theta$ *инвариантно, то* $\widehat{\theta}^*$ *является эквивариантной оценкой.*

Доказательство. Имеем

$$f_{\widehat{\theta}^*}(X) = \frac{\mathbf{P}_{\widehat{\theta}^*(X)}(d\boldsymbol{x})}{\mu^n(d\boldsymbol{x})} = \max_\theta \frac{\mathbf{P}_\theta(d\boldsymbol{x})}{\mu^n(d\boldsymbol{x})} \tag{4}$$

в точке $\boldsymbol{x} = X$. Полагая $Y = gX$, мы можем записать также

$$f_{\widehat{\theta}^*(Y)}(Y) = \frac{\mathbf{P}_{\widehat{\theta}^*(Y)}(g\,d\boldsymbol{x})}{\mu^n(g\,d\boldsymbol{x})} = \max_\theta \frac{\mathbf{P}_\theta(g\,d\boldsymbol{x})}{\mu^n(g\,d\boldsymbol{x})}.$$

В силу инвариантности $\mathbf{P}_\theta$ и конечности $\mu^n(g\,d\boldsymbol{x})/\mu^n(d\boldsymbol{x}) > 0$ это эквивалентно тому, что

$$\frac{\mathbf{P}_{\overline{g}^{-1}\widehat{\theta}^*(Y)}(d\boldsymbol{x})}{\mu^n(d\boldsymbol{x})} = \max_\theta \frac{\mathbf{P}_{\overline{g}^{-1}\theta}(d\boldsymbol{x})}{\mu^n(d\boldsymbol{x})} = \max_\theta \frac{\mathbf{P}_\theta(d\boldsymbol{x})}{\mu^n(d\boldsymbol{x})}.$$

Сравнивая с (4) и пользуясь единственностью $\widehat{\theta}^*(X)$, получаем $\overline{g}^{-1}\widehat{\theta}^*(gX) = \widehat{\theta}^*(X)$. $\triangleleft$

§ 30. Интегральное неравенство типа Рао–Крамера. Критерии асимптотической байесовости и минимаксности оценок

Этот параграф можно было бы назвать также «Неравенство для среднеквадратичного уклонения в байесовском случае». В значительной своей части он относится к асимптотической теории оценивания.

Вопросы, связанные с асимптотическим подходом к сравнению оценок, затрагивались нами и раньше. Теперь, и главным образом в §§ 33–37, они станут основным объектом изучения.

1. Эффективные и сверхэффективные оценки. В § 26, посвященном неравенству Рао–Крамера, остался невыясненным следующий важный вопрос. Пусть выполнено условие (R). Тогда для несмещенных оценок

$$\mathbf{E}_\theta(\theta^* - \theta)^2 \geqslant \frac{1}{nI(\theta)}.$$

Правую часть этого неравенства называют иногда *границей Рао–Крамера*. Она достигается для R-эффективных оценок. Вопрос состоит в том, можно ли за счет выбора смещения сколько-нибудь существенно улучшить R-эффективные или а. R-э. оценки? Это есть вопрос о существенности границы Рао–Крамера и о роли смещения.

Тот факт, что в отдельной фиксированной точке θ_0 значение $\mathbf{E}_\theta(\theta^* - \theta)^2$ может быть сделано значительно меньше, чем граница Рао–Крамера, нами уже отчасти обсуждался. Достаточно взять $\theta^* = \theta_0$. Однако при этом в других точках эта оценка будет очень плохой.

Можно привести другой, менее тривиальный пример, где улучшение достигается не за счет других точек. Пусть $X \in \mathbf{\Phi}_{\alpha,1}$, $\alpha \in \Theta = [0, \infty)$. Тогда оценка $\alpha^* = \overline{\mathrm{x}}$ является эффективной и даже R-эффективной. Однако оценка

$\alpha^{**} = \max(0, \overline{\mathrm{x}})$ в нашем случае, когда $\Theta = [0, \infty)$, будет, очевидно, лучше, поскольку она уменьшает среднеквадратичные уклонения, заменяя недопустимые отрицательные значения на нуль. Оценка α^{**}, очевидно, будет уже смещенной: $\mathbf{E}_\alpha \alpha^{**} > \alpha$, но в точке $\alpha = 0$ имеем $I(\alpha) = 1$, $\mathbf{E}_0(\alpha^*)^2 = \dfrac{1}{n}$, $\mathbf{E}_0(\alpha^{**})^2 = \dfrac{1}{2n} < \dfrac{1}{nI(0)}$.

В этом примере улучшение связано с тем, что мы сузили область значений оценки α^* до множества Θ. Приведем еще один пример (принадлежащий Ходжесу), в котором улучшение α^* происходит не за счет ограничения Θ.

Пусть опять $X \in \mathbf{\Phi}_{\alpha,1}$, $\alpha \in \Theta = (-\infty, \infty)$. Наряду с эффективной оценкой $\alpha^* = \overline{\mathrm{x}}$ рассмотрим при $\beta < 1$ оценку

$$\alpha^{**} = \begin{cases} \overline{\mathrm{x}}, & \text{если} \quad |\overline{\mathrm{x}}| \geqslant n^{-1/4}, \\ \beta\overline{\mathrm{x}}, & \text{если} \quad |\overline{\mathrm{x}}| < n^{-1/4}. \end{cases}$$

Нетрудно видеть, что при $\alpha > 0$ по центральной предельной теореме

$$\mathbf{P}_\alpha(|\overline{\mathrm{x}}| < n^{-1/4}) \leqslant \mathbf{P}_\alpha((\overline{\mathrm{x}} - \alpha)\sqrt{n} < n^{1/4} - \alpha\sqrt{n}) \to 0$$

при $n \to \infty$. Аналогичное утверждение верно при $\alpha < 0$. Поэтому α^{**} при $\alpha \neq 0$ совпадает с $\overline{\mathrm{x}}$ на множестве вероятности, сходящейся к единице, и, стало быть, по теореме непрерывности при $\alpha \neq 0$

$$(\alpha^{**} - \alpha)\sqrt{n} \Rightarrow \mathbf{\Phi}_{0,1}.$$

При $\alpha = 0$

$$\mathbf{P}_0(|\overline{\mathrm{x}}| < n^{-1/4}) = \mathbf{P}_0(|\overline{\mathrm{x}}\sqrt{n}| < n^{1/4}) \to 1,$$

и α^{**} на множестве вероятности, сходящейся к единице, совпадает с $\beta\overline{\mathrm{x}}$, так что $(\alpha^{**} - \alpha)\sqrt{n} \Rightarrow \mathbf{\Phi}_{0,\beta^2}$. Таким образом, при всех α оценка α^{**} является а. н., $(\alpha^{**} - \alpha)\sqrt{n} \Rightarrow \mathbf{\Phi}_{0,\sigma^2(\alpha)}$, где

$$\sigma^2(\alpha) = \begin{cases} 1 & \text{при} \quad \alpha \neq 0, \\ \beta^2 < 1 & \text{при} \quad \alpha = 0. \end{cases}$$

Таким образом, в точке $\alpha = 0$ коэффициент рассеивания $\sigma^2(0)$ оказался меньше нижней границы Рао–Крамера, равной единице.

Асимптотически нормальные оценки в приведенных примерах, когда коэффициент рассеивания для них $\sigma^2(\theta) \leqslant I^{-1}(\theta)$ оказывается при некоторых θ строго меньше $I^{-1}(\theta)$, называют иногда *суперэффективными*.

Однако оказалось, что эти примеры все же мало меняют правильную в целом картину о предпочтительности эффективных оценок. Именно, Ле Камом было установлено, что добиться проиллюстрированного выше улучшения оценок можно, грубо говоря, лишь в малом числе точек.

Здесь мы покажем, что наряду с соотношением $\inf\limits_{\theta^*} \mathbf{E}_t(\theta^* - t)^2 = 0$, справедливом при каждом t, для *интеграла* от $\mathbf{E}_t(\theta^* - t)^2$ уже существует *положительная нижняя граница, не зависящая от* θ^* и тесно связанная

с аналогичным интегралом от функции $(nI(t))^{-1}$. Именно, мы получим в одномерном случае $\theta \in \mathbb{R}$ неравенство для

$$\inf_{\theta^*} \int \mathbf{E}_t(\theta^* - t)^2 q(t)\, dt \tag{1}$$

при любой весовой функции $q(t) \geqslant 0$, $\int q(t)\, dt = 1$, правая часть которого не зависит от θ^* (в том числе и от смещения $b(t)$, присутствующего в неравенстве Рао–Крамера) и близка к значению J/n, где

$$J = \int \frac{q(t)}{I(t)}\, dt. \tag{2}$$

2. Основные неравенства. Прежде чем сформулировать соответствующие теоремы, заметим, что интеграл в (1) можно рассматривать как безусловное математическое ожидание $\mathbf{E}(\theta^* - \theta)^2$ в байесовском случае, когда θ имеет априорное распределение с плотностью $q(t)$ относительно меры Лебега. В этом случае $J = \mathbf{E}I^{-1}(\theta)$.

Обозначим через $f(\boldsymbol{x}, t) = f_t(\boldsymbol{x})q(t)$ плотность совместного распределения X и θ; $f_t'(\boldsymbol{x})$, как и прежде, будет означать производную $f_t(\boldsymbol{x})$ по t.

Пусть, далее, $N_h \subset \Theta$ есть носитель функции h, определенной на Θ; $N_h = \{t : h(t) \neq 0\}$, и N есть носитель $f(\boldsymbol{x}, t)$ в $\mathcal{X}^n \times \Theta$.

Теорема 1. *Предположим, что $f_t(x)$ дифференцируема по t, а функция $\sqrt{I(t)}$ интегрируема на любом конечном интервале. Тогда для любой финитной (т. е. равной нулю вне конечного интервала) дифференцируемой функции $h(t)$ такой, что $N_h \subset N_q$, справедливо неравенство*

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{[\mathbf{E}(h(\theta)/q(\theta))]^2}{n\mathbf{E}(I(\theta)[h(\theta)/q(\theta)]^2) + \mathbf{E}[h'(\theta)/q(\theta)]^2} = \\ = \frac{\left[\int h(t)\, dt\right]^2}{n\int I(t)h^2(t)/q(t)\, dt + \int h'(t)^2/q(t)\, dt}. \tag{3}$$

Доказательство. В силу финитности $h(t)$ имеем

$$\int (f_t(\boldsymbol{x})h(t))'\, dt = \int d(f_t(\boldsymbol{x})h(t)) = 0,$$

$$\int t(f_t(\boldsymbol{x})h(t))'\, dt = -\int f_t(\boldsymbol{x})h(t)\, dt.$$

Стало быть, для любой θ^*

$$\int\limits_{\mathcal{X}^n}\int\limits_{\Theta} (\theta^* - t)(f_t(\boldsymbol{x})h(t))'\, dt\mu^n(d\boldsymbol{x}) = \int\limits_{\mathcal{X}^n}\int\limits_{\Theta} f_t(\boldsymbol{x})h(t)\, dt\mu^n(d\boldsymbol{x}) = \int\limits_{\Theta} h(t)\, dt. \tag{4}$$

Эти интегралы можно рассматривать в силу условия $N_h \subset N_q$ как интегралы по N. Стало быть, мы можем умножить и разделить подынтегральное выражение в (4) на $f(x,t)$. Тогда получим

$$\mathbf{E}\left[(\theta^*-\theta)\frac{(f_\theta(X)h(\theta))'}{f(X,\theta)}\right]=\int\limits_{N_q}h(t)\,dt=\mathbf{E}\frac{h(\theta)}{q(\theta)}.$$

В силу неравенства Коши–Буняковского отсюда следует

$$\mathbf{E}(\theta^*-\theta)^2\geqslant\frac{[\mathbf{E}(h(\theta)/q(\theta))]^2}{\mathbf{E}[(f_\theta(X)h(\theta))'/(f_\theta(X)q(\theta))]^2}. \tag{5}$$

Остается привести это неравенство к виду (3). Заметим предварительно, что

$$\mathbf{E}_t|L'(X,t)|\leqslant n\sqrt{I(t)}$$

и что для почти всех *) t

$$\mathbf{E}_tL'(X,t)=0. \tag{6}$$

Первое из этих утверждений следует из соотношений

$$\mathbf{E}_t|L'(X,t)|\leqslant n\mathbf{E}_t|l'(\mathrm{x}_1,t)|\leqslant n\{\mathbf{E}_t[l'(\mathrm{x}_1,t)]^2\}^{1/2}=n\sqrt{I(t)},$$

вытекающих из неравенства Коши–Буняковского. Чтобы доказать второе утверждение, возьмем произвольную финитную функцию $g(t)$, имеющую всюду непрерывную производную $g'(t)$. Тогда

$$\int g(t)f_t'(X)\,dt=-\int g'(t)f_t(X)\,dt.$$

Кроме того,

$$\int|g(t)|\mathbf{E}_t|L'(X,t)|\,dt\leqslant n\int|g(t)|\sqrt{I(t)}\,dt<\infty.$$

Отсюда следует, что можно менять порядок интегрирования в следующем выражении:

$$\int g(t)\mathbf{E}_tL'(X,t)\,dt=\int\limits_{\mathcal{X}^n}\int\limits_{\Theta}g(t)f_t'(x)\,dt\mu^n(dx)=$$

$$=-\int\limits_{\mathcal{X}^n}\int\limits_{\Theta}g'(t)f_t(x)\,dt\mu^n(dx)=-\int\limits_{\Theta}g'(t)\,dt=-\int\limits_{\Theta}dg(t)=0.$$

Выполнение этого равенства для всех g и означает справедливость (6).

*) В § 26 мы доказали, что это равенство при выполнении условий (R) имеет место *при всех* t. Здесь нам будет достаточно его выполнения для почти всех t.

Теперь можем преобразовать правую часть (5). Опуская для краткости аргументы функций, получим

$$\mathbf{E}\left[\frac{(f_\theta(X)h(\theta))'}{f_\theta(X)q(\theta)}\right]^2 = \mathbf{E}\left[L'\frac{h}{q}+\frac{h'}{q}\right]^2 = \mathbf{E}\left[\left(\frac{h}{q}\right)^2\mathbf{E}_\theta(L')^2\right] +$$

$$+2\mathbf{E}\left[\frac{h'h}{q^2}\mathbf{E}_\theta L'\right]+\mathbf{E}\left(\frac{h'}{q}\right)^2 = n\mathbf{E}\left[\left(\frac{h}{q}\right)^2 I\right]+\mathbf{E}\left(\frac{h'}{q}\right)^2.$$

Здесь мы воспользовались тем, что в силу (6)

$$\mathbf{E}\left[\frac{h'h}{q^2}\mathbf{E}_\theta L'\right] = \int\limits_{N_q}\frac{h'h}{q}\mathbf{E}_t L'\,dt = 0$$

и что (см. § 26) $\mathbf{E}_\theta(L')^2 = nI(\theta)$. ◁

В последующих утверждениях везде будем предполагать, что $f_t(x)$ удовлетворяет условиям теоремы 1.

Теорема 2. *Если функция* $h(t) = h_0(t) \equiv q(t)/I(t)$ *является финитной и дифференцируемой, то*

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{J}{n}\left(1+\frac{H}{nJ}\right)^{-1} \geqslant \frac{J}{n}-\frac{H}{n^2}, \tag{7}$$

где $H = \int\left[\left(\frac{q(t)}{I(t)}\right)'\right]^2\frac{dt}{q(t)}$.

Замечание 1. Неравенства в теоремах 1, 2 являются *интегральными* в том смысле, что они относятся к интегралам от $\mathbf{E}_t(\theta^*-\theta)^2$. В этом смысле неравенства § 26 можно называть *локальными*.

Доказательство. Это утверждение немедленно вытекает из теоремы 1, так как правая часть в (3) при $h = q/I$ преобразуется к виду $J^2/(nJ+H)$. ◁

Мы видим, таким образом, что нижняя граница возможных значений $\mathbf{E}(\theta^*-\theta)^2$ при больших n лишь немногим отличается от границы $\dfrac{J}{n} = \int\dfrac{q(t)\,dt}{nI(t)}$, равной значению $\mathbf{E}(\theta_0^*-\theta)^2$ для R-эффективной оценки θ_0^*. Это указывает на целесообразность использования эффективных оценок, поскольку для них при любой функции q почти достигается экстремальное значение $\mathbf{E}(\theta^*-\theta)^2$.

Оценка (7) неулучшаема, на что указывает

Пример 1. Пусть $X \in \Phi_{\alpha,1}$. Как мы знаем, в этом случае $I(\alpha) = 1$. Пусть, далее, параметр α выбирается случайно с гладкой плотностью $q(t)$, $t \in (-\infty,\infty)$. Тогда правая часть (7) превращается в $(n+H)^{-1}$, где

$$H = \int\frac{(q')^2}{q}\,dt = \mathbf{E}[(\ln q(\alpha))']^2.$$

В нашем случае байесовская оценка α_Q^*, соответствующая априорному распределению $\mathbf{Q}$ с плотностью q и минимизирующая $\mathbf{E}(\alpha^* - \alpha)^2$, равна (см. § 20)

$$\alpha_Q^* = \frac{\int tq(t)f_t(X)\,dt}{\int q(t)f_t(X)\,dt} = \\ = \frac{\int tq(t)\exp(n\bar{\mathrm{x}}t - t^2n/2)\,dt}{\int q(t)\exp(n\bar{\mathrm{x}}t - t^2n/2)\,dt} = \frac{\int tq(t)\exp(-n(\bar{\mathrm{x}} - t)^2/2)\,dt}{\int q(t)\exp(-n(\bar{\mathrm{x}} - t)^2/2)\,dt}. \quad (8)$$

Нетрудно найти асимптотическое представление этого отношения и показать, что

$$\alpha_Q^* = \bar{\mathrm{x}} + \frac{q'(\bar{\mathrm{x}})}{nq(\bar{\mathrm{x}})} + O\left(\frac{1}{n^2}\right), \quad \mathbf{E}\left(\alpha_Q^* - \alpha\right)^2 = \frac{1}{n} - \frac{H}{n^2} + O\left(\frac{1}{n^3}\right).$$

Однако мы поступим проще, предположив, что $q(t) = \dfrac{1}{\sqrt{2\pi}}e^{-t^2/2}$.

Тогда, очевидно, $H = 1$, и правая часть в (7) превращается в $1/(n+1)$. Но в примере 21.1 было установлено, что

$$\mathbf{E}\left(\alpha_Q^* - \alpha\right)^2 = \frac{1}{n+1}.$$

Тем самым неулучшаемость неравенств (7) и (3) доказана.

Теорема 3. *Если интервал* $(a - \varepsilon, a + \varepsilon)$ *содержится в* Θ, *то для любой оценки* θ^*

$$\max_{t\in(a-\varepsilon,a+\varepsilon)} \mathbf{E}_t(\theta^* - t)^2 \geqslant \frac{1}{n \max\limits_{t\in(a-\varepsilon,a+\varepsilon)} I(t) + \pi^2\varepsilon^{-2}}.$$

Доказательство. Воспользуемся неравенством

$$\max_{t\in(a-\varepsilon,a+\varepsilon)} \mathbf{E}_t(\theta^* - t)^2 \geqslant \int\limits_{a-\varepsilon}^{a+\varepsilon} \mathbf{E}_t(\theta^* - t)^2 q(t)\,dt,$$

справедливым для любой плотности $q(t)$, равной нулю вне $(a - \varepsilon, a + \varepsilon)$. Требуемое утверждение следует из теоремы 1, если положить в ней

$$h(t) = q(t) = \frac{1}{\varepsilon}\cos^2\frac{\pi(t-a)}{2\varepsilon}, \quad |t - a| \leqslant \varepsilon.$$

Тогда

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{1}{n\int I(t)q(t)\,dt + \int (q'(t))^2/q(t)\,dt},$$

где

$$\int \frac{(q'(t))^2}{q(t)}\,dt = \int\limits_{-\varepsilon}^{\varepsilon} \frac{\left(\dfrac{\pi}{2\varepsilon^2}\cdot 2\cos\dfrac{\pi t}{2\varepsilon}\sin\dfrac{\pi t}{2\varepsilon}\right)^2 \varepsilon}{\cos^2\dfrac{\pi t}{2\varepsilon}}\,dt = \frac{1}{\varepsilon^2}\int\limits_{-1}^{1} \pi^2\sin^2\frac{\pi t}{2}\,dt = \frac{\pi^2}{\varepsilon^2}. \quad \triangleleft$$

Можно отметить, что на функции $q(t) = \cos^2(\pi t/2)$ достигается минимум функционала $\int_{-1}^{1} (q'(t))^2/q(t)\,dt$ в классе всех дифференцируемых плотностей $q(t)$.

Из теоремы 3 вытекает, в частности, что интервал значений θ, при которых оценка θ^* является сверхэффективной, не может иметь длину, большую чем $O(1/\sqrt{n})$.

3. Неравенства в случае, когда функция $q(\theta)/I(\theta)$ не дифференцируема. Если функция $h_0 = q/I$ не удовлетворяет условиям теоремы 1, справедливо следующее полезное утверждение, позволяющее оценивать асимптотику $\mathbf{E}(\theta^* - \theta)^2$ в общем случае.

Теорема 4. *Пусть последовательность функций $h_\varepsilon(t)$, зависящих от параметра $\varepsilon > 0$, такова, что каждая функция h_ε удовлетворяет условиям теоремы* 1 *и*

1) $h_\varepsilon(t) \leqslant h_0(t)$,

2) $H(\varepsilon) = \int \dfrac{(h'_\varepsilon(t))^2}{q(t)}\,dt < \infty$.

Тогда для любого $\varepsilon > 0$

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{\left(\int h_\varepsilon(t)\,dt\right)^2}{nJ + H(\varepsilon)}.$$

Доказательство немедленно вытекает из теоремы 1, если положить $h = h_\varepsilon$.

Из теоремы 4 получаем следующее важное следствие.

Теорема 5. *Если функция q интегрируема по Риману, $J < \infty$, то*

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{J}{n}(1 + \delta_n),$$

где $\delta_n = o(1)$ при $n \to \infty$.

Доказательство. Положим $\widehat{q}_\varepsilon(t) = \min\limits_{|u| \leqslant \varepsilon} q(t+u)$,

$$q_\varepsilon(t) = \begin{cases} \widehat{q}_\varepsilon(t), & \text{если} \quad \widehat{q}_\varepsilon(t) \geqslant \varepsilon, \\ 0, & \text{если} \quad \widehat{q}_\varepsilon(t) < \varepsilon, \end{cases}$$

$$I_\varepsilon(t) = \max(\varepsilon, I(t)),$$

$$h_\varepsilon(t) = \frac{1}{2\varepsilon} \int_{t-\varepsilon}^{t+\varepsilon} \frac{q_\varepsilon(v)}{I_\varepsilon(v)}\,dv \leqslant h_0(t).$$

Очевидно, что функция h_ε финитна и дифференцируема при любом $\varepsilon > 0$.

Из того, что $q(t)$ интегрируема по Риману, следует, что $q_\varepsilon(t) \uparrow q(t)$ почти всюду при $\varepsilon \to 0$. Чтобы доказать это, убедимся, что

$$\int_a^b [q(t) - q_\varepsilon(t)]\, dt \downarrow 0. \tag{9}$$

Из интегрируемости $q(t)$ по Риману вытекает сходимость

$$\sum_k q_\delta(2k\delta)2\delta \uparrow \int q(t)\, dt, \quad \sum_k q_\delta((2k+1)\delta)2\delta \uparrow \int q(t)\, dt$$

при $\delta \to 0$. Поэтому

$$\int_a^b q_\varepsilon(t)\, dt \geqslant \sum_k q_{2\varepsilon}(2k\varepsilon)2\varepsilon =$$

$$= \frac{1}{2}\left(\sum q_{2\varepsilon}(4k\varepsilon)4\varepsilon + \sum q_{2\varepsilon}((4k+2)\varepsilon)4\varepsilon\right) \to \int q(t)\, dt.$$

Соотношение (9) и вместе с ним сходимость $q_\varepsilon(t) \uparrow q(t)$ доказаны.

Пользуясь теперь этой сходимостью, получаем $\dfrac{q_\varepsilon(t)}{I_\varepsilon(t)} \uparrow h_0(t)$,

$$\int h_\varepsilon(t)\, dt = \int \frac{dt}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \frac{q_\varepsilon(t+v)}{I_\varepsilon(t+v)}\, dv = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} dv \int \frac{q_\varepsilon(t)}{I_\varepsilon(t)}\, dt = \int \frac{q_\varepsilon(t)}{I_\varepsilon(t)}\, dt \uparrow J.$$

Кроме того,

$$\left|h'_\varepsilon(t)\right| = \frac{1}{2\varepsilon}\left|\frac{q_\varepsilon(t+\varepsilon)}{I_\varepsilon(t+\varepsilon)} - \frac{q_\varepsilon(t-\varepsilon)}{I_\varepsilon(t-\varepsilon)}\right| \leqslant \frac{q(t)}{\varepsilon^2},$$

$$H(\varepsilon) \leqslant \int \left(\frac{q(t)}{\varepsilon^2}\right)^2 q^{-1}(t)\, dt = \frac{1}{\varepsilon^4}.$$

Мы можем воспользоваться теперь теоремой 3. Полагая $\varepsilon = \varepsilon(n) = n^{-1/5}$, $n \to \infty$, получим $\varepsilon(n) \to 0$,

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{\left(\int h_\varepsilon(t)\, dt\right)^2}{nJ + n^{4/5}} = \frac{J}{n}(1 + o(1)). \quad \triangleleft$$

4. Некоторые следствия. Критерии асимптотической байесовости и минимаксности. Один из основных выводов, которые можно сделать из результатов этого параграфа, состоит, грубо говоря, в следующем. Если существует а. R-э. оценка, то какую бы другую оценку мы не взяли, мы не получим «в целом» (или «в среднем») асимптотически лучшего результата. Этим фактом

мы воспользуемся позже в § 33. Здесь же мы приведем критерии асимптотической байесовости и асимптотической минимаксности оценок, вытекающие из теорем 2, 5.

Определение 1. Оценка θ_1^*, обладающая свойством

$$\mathbf{E}n\big(\theta_1^* - \theta\big)^2 = J + o(1) \tag{10}$$

при $n \to \infty$, называется *асимптотически R-байесовской.*

Это есть оценки, для которых асимптотически достигается нижняя граница для среднеквадратичных уклонений, определенная в теоремах 2, 5. Их можно было бы назвать также *асимптотически R-эффективными «в целом»* (или «в среднем»).

Напомним (см. § 21), что оценка θ_1^* называется *асимптотически байесовской* (относительно распределения $\mathbf{Q}$), если для всякой другой оценки θ^*

$$\limsup_{n\to\infty}\big[\mathbf{E}n\big(\theta_1^* - \theta\big)^2 - \mathbf{E}n(\theta^* - \theta)^2\big] \leqslant 0. \tag{11}$$

Следствие 1. *Пусть выполнены условия теоремы* 1 *и функция* $q(t)$ *интегрируема по Риману. Тогда асимптотически R-байесовская оценка является асимптотически байесовской.*

Доказательство. Пусть θ_1^* — асимптотически R-байесовская оценка. В силу теоремы 5 для любой оценки θ^*

$$\liminf_{n\to\infty} \mathbf{E}n(\theta^* - \theta)^2 \geqslant J.$$

Отсюда и из (10) следует (11). $\triangleleft$

Ясно также, что если асимптотически R-байесовская оценка существует, то любая асимптотически байесовская оценка будет асимптотически R-байесовской (ср. с замечаниями к теореме 26.3).

Из теоремы 5 вытекает также

Следствие 2. *Пусть выполнены условия теоремы* 1 *и функция* $q(t)$ *интегрируема по Риману. Если* θ_1^* *и* θ_2^* *— две асимптотически R-байесовские оценки, они асимптотически эквивалентны в следующем смысле:*

$$\mathbf{E}n\big(\theta_1^* - \theta_2^*\big)^2 \to 0, \quad \big(\theta_1^* - \theta_2^*\big)\sqrt{n} \underset{P}{\longrightarrow} 0,$$

где сходимость по вероятности понимается относительно совместного распределения X и θ в $\mathcal{X}^n \times \Theta$.

Доказательство совершенно аналогично доказательствам теорем 18.2, 26.4. Исходным является равенство (18.23), которое в силу теоремы 5 дает

$$\limsup_{n\to\infty} \mathbf{E}n\big(\theta_1^* - \theta_2^*\big)^2 \leqslant 0. \ \triangleleft$$

В § 18, 21 мы отмечали, что для сравнения оценок наряду со *средними* значениями $\int q(t)\mathbf{E}_t(\theta^* - t)^2\,dt$ можно рассматривать *максимальные* значения

$$\sup_{t\in\Gamma} \mathbf{E}_t(\theta^* - t)^2, \quad \Gamma \subset \Theta.$$

В качестве Γ берется либо все множество Θ, либо та его часть, которая по предварительным данным содержит неизвестное значение θ. Напомним, что оценка $\overline{\theta}^*$ называется минимаксной, если для любой оценки θ^*

$$\sup_{t\in\Gamma}\mathbf{E}_t(\overline{\theta}^*-t)^2\leqslant\sup_{t\in\Gamma}\mathbf{E}_t(\theta^*-t)^2.$$

Оценка θ_1^* называется *асимптотически минимаксной*, если для любой оценки θ^*

$$\limsup_{n\to\infty}\sup_{t\in\Gamma}\mathbf{E}_t\left[\sqrt{n}\,(\theta_1^*-t)\right]^2\leqslant\liminf_{n\to\infty}\sup_{t\in\Gamma}\mathbf{E}_t\left[\sqrt{n}\,(\theta^*-t)\right]^2.$$

Следствие 3. *Пусть информация Фишера $I(\theta)$ существует и непрерывна. Тогда если для любого отрезка $\Gamma\subset\Theta$*

$$\limsup_{n\to\infty}\sup_{t\in\Gamma}\mathbf{E}_t\left[\sqrt{n}\,(\theta_1^*-t)\right]^2\leqslant\sup_{t\in\Gamma}I^{-1}(t),\tag{12}$$

то оценка θ_1^* асимптотически минимаксна.

Доказательство. Нам достаточно убедиться, что для любой оценки θ^*

$$\liminf_{n\to\infty}\sup_{t\in\Gamma}\mathbf{E}_t[\sqrt{n}\,(\theta^*-t)]^2\geqslant\sup_{t\in\Gamma}I^{-1}(t).\tag{13}$$

Для любого распределения $\mathbf{Q}$ на Γ с гладкой плоскостью $q(t)$ относительно меры Лебега

$$\sup_{t\in\Gamma}\mathbf{E}_t[\sqrt{n}\,(\theta^*-t)]^2\geqslant\int\mathbf{E}_t[\sqrt{n}\,(\theta^*-t)]^2q(t)\,dt.$$

По теореме 2 интеграл в правой части для любой оценки θ^* не меньше, чем $J-H/n$. Поэтому левая часть (13) больше или равна

$$J=\int I^{-1}(t)q(t)\,dt.$$

Но q — произвольная гладкая плотность, и для заданного $\varepsilon>0$ ее всегда можно выбрать в силу непрерывности $I^{-1}(t)$ так, чтобы

$$J\geqslant\sup_{t\in\Gamma}I^{-1}(t)-\varepsilon.$$

Так как ε произвольно, (13) доказано. $\triangleleft$

В заключение этого пункта сделаем одно важное замечание. Оно состоит в том, что при отыскании асимптотически оптимальных оценок можно ограничиться классом $\widetilde{K}_0$ асимптотически несмещенных оценок, введенным нами в § 26. Это вытекает из следующих соображений.

Мы уже отмечали, что правая часть неравенства теоремы 5, равная $J/n+o(1/n)$, от смещения $b(\theta)$ никак не зависит. В то же время, если для построения нижней границы для $\mathbf{E}(\theta^*-\theta)^2$ мы будем исходить из неравенства Рао–Крамера в § 26, то мы получим

$$\mathbf{E}(\theta^*-\theta)^2\geqslant\min_b\int q(t)\left[\frac{(1+b'(t))^2}{nI(t)}+b^2(t)\right]dt.$$

Можно показать (ср. с [131]), что этот минимум по всем смещениям $b(\theta)$ имеет (при некоторых предположениях на гладкость $q(t)$ и $I(t)$) тот же вид $J/n + o(1/n)$ и, что для нас существенно, достигается на смещении $b(\theta)$, обладающем при $n \to \infty$ свойствами

$$b'(t) = o(1), \quad b(t) = o\left(\frac{1}{\sqrt{n}}\right).$$

Класс оценок θ^* с такими смещениями и есть $\widetilde{K}_0$ (см. § 26). Выход θ^* из класса $\widetilde{K}_0$ делает границу $J/n + o(1/n)$ недостижимой. Таким образом, при асимптотическом подходе, когда а. н. оценки сравниваются с помощью значений $\mathbf{E}(\theta^* - \theta)^2$ при гладких $q(t)$ и $I(t)$, можно ограничиться рассмотрением оценок из класса $K = K_{\Phi,2} \cap \widetilde{K}_0$ (класс $K_{\Phi,2}$ обсуждался нами в § 18), поскольку оценки вне класса $\widetilde{K}_0$ являются «недопустимыми» в указанном выше смысле.

5. Многомерный случай. В случае $\theta \in \mathbb{R}^k$ можно получить аналоги для всех теорем этого параграфа и сделать те же выводы, которые нами были получены для одномерного случая.

В частности, утверждение теоремы 5, одной из основных в этом параграфе, будет иметь вид

$$d^2 \geqslant \frac{J}{n} + o\left(\frac{1}{n}\right),$$

где $d^2 = \|d_{ij}\|$, $d_{ij} = \mathbf{E}\big(\theta_i^* - \theta_i\big)\big(\theta_j^* - \theta_j\big)$, $J = \mathbf{E}I^{-1}(\theta)$.

Рассуждения, связанные с байесовскими и минимаксными оценками, также сохранятся, если в качестве погрешности оценки рассматривать значение

$$v(\theta^*) = \mathbf{E}_\theta(\theta^* - \theta)V(\theta^* - \theta)^\top,$$

где V — неотрицательно определенная матрица. Байесовскими и минимаксными оценками (или асимптотически байесовскими и минимаксными) следует называть оценки, погрешности которых удовлетворяют соответствующим неравенствам для любой неотрицательно определенной матрицы V.

§ 31. Расстояния Кульбака–Лейблера, Хеллингера и χ^2. Их свойства

Результаты этого параграфа будут существенны для получения основных результатов асимптотической теории оценивания, а также для результатов главы 3.

1. Определения и основные свойства расстояний. Пусть $\mathbf{P}$ и $\mathbf{G}$ — два распределения на $(\mathcal{X}, \mathfrak{B}_\mathcal{X})$, абсолютно непрерывных относительно меры μ. Обозначим

$$\frac{d\mathbf{P}}{d\mu} = p, \quad \frac{d\mathbf{G}}{d\mu} = g,$$

N_p — носитель распределения $\mathbf{P}$: $N_P = \{x : p(x) > 0\}$.

Определение 1. *Расстоянием Кульбака–Лейблера* между распределениями $\mathbf{P}$ и $\mathbf{G}$ называется величина

$$\rho_1(\mathbf{P},\mathbf{G}) = \int\limits_{N_P} \ln\frac{p(x)}{g(x)}\mathbf{P}(dx) = \int\limits_{N_p} \ln\frac{p(x)}{g(x)}p(x)\mu(dx).$$

На самом деле $\rho_1(\mathbf{P},\mathbf{G})$ не есть, конечно, расстояние или метрика в общепринятом смысле, поскольку $\rho_1(\mathbf{P},\mathbf{G})$ не есть симметричная функция от $\mathbf{P}$ и $\mathbf{G}$. Тем не менее мы увидим, что $\rho_1(\mathbf{P},\mathbf{G})$ очень по существу (со статистической точки зрения) характеризует отклонение $\mathbf{G}$ от $\mathbf{P}$.

Из неравенства $\ln(1+v) - v \leqslant 0$ и представления

$$\rho_1(\mathbf{P},\mathbf{G}) = -\int\left[\ln\frac{g}{p} - \left(\frac{g}{p}-1\right)\right]p\mu(dx)$$

следует, что всегда $\rho_1(\mathbf{P},\mathbf{G}) \geqslant 0$. В лемме 16.1 было установлено, что равенство $\rho_1(\mathbf{P},\mathbf{G}) = 0$ возможно, лишь если $\mathbf{P} = \mathbf{G}$.

Определение 2. *Расстоянием* χ^2 между распределениями $\mathbf{P}$ и $\mathbf{G}$ будем называть величину

$$\rho_2(\mathbf{P},\mathbf{G}) = \int\limits_{N_P\cup N_G} \frac{(p(x)-g(x))^2}{p(x)}\mu(dx).$$

К этому расстоянию относятся почти все замечания, сделанные к определению 1. Название χ^2 объясняется причинами, которые станут ясны позже.

Определение 3. *Расстоянием Хеллингера* между распределениями $\mathbf{P}$ и $\mathbf{G}$ называется величина

$$\rho_3(\mathbf{P},\mathbf{G}) = \int\limits_{N_P\cup N_G} (\sqrt{p(x)} - \sqrt{g(x)})^2\mu(dx).$$

Расстояние Хеллингера является уже симметричной функцией $\mathbf{P}$ и $\mathbf{G}$, а значение $\sqrt{\rho_3(\mathbf{P},\mathbf{G})}$ обладает всеми свойствами метрики (между функциями $\sqrt{p(x)}$ и $\sqrt{g(x)}$ в метрическом пространстве $L_2(\mathcal{X},\mu)$). Нетрудно видеть, что

$$\rho_3(\mathbf{P},\mathbf{G}) = 2\left(1 - \int\sqrt{pg}\mu(dx)\right) \leqslant 2. \tag{1}$$

Все три введенных нами расстояния играют существенную роль в различных задачах математической статистики. В какой-то мере мы в этом убедимся.

Если с помощью этих расстояний характеризовать степень близости распределений, когда отношение p/g близко к единице, то окажется, что все они асимптотически ведут себя одинаково с точностью до постоянных множителей. Действительно, пользуясь разложением

$$\ln\frac{g}{p} = \ln\left(1+\left(\frac{g}{p}-1\right)\right) = \left(\frac{g}{p}-1\right) - \frac{1}{2}\left(\frac{g}{p}-1\right)^2 + O\left(\left|\frac{g}{p}-1\right|^3\right),$$

получаем

$$\rho_1(\mathbf{P},\mathbf{G}) = -\int \ln\frac{g}{p}\cdot p\mu(dx) \approx \frac{1}{2}\int\left(\frac{g}{p}-1\right)^2 p\mu(dx) = \frac{1}{2}\rho_2(\mathbf{P},\mathbf{G}),$$

$$\rho_2(\mathbf{P},\mathbf{G}) = \int\frac{(p-g)^2}{p}\mu(dx) = \int(\sqrt{p}-\sqrt{g})^2\left(1+\sqrt{\frac{g}{p}}\right)^2\mu(dx) \approx 4\rho_3(\mathbf{P},\mathbf{G}).$$

Из последнего равенства следует также, что $\rho_2(\mathbf{P},\mathbf{G}) \geqslant \rho_3(\mathbf{P},\mathbf{G})$. Кроме того, $\rho_1(\mathbf{P},\mathbf{G}) \geqslant \rho_3(\mathbf{P},\mathbf{G})$. Действительно, так как $\ln(1+x) \leqslant x$, то

$$\ln\frac{g}{p} = 2\ln\left(1+\left(\sqrt{\frac{g}{p}}-1\right)\right) \leqslant 2\left(\sqrt{\frac{g}{p}}-1\right),$$

$$\rho_1(\mathbf{P},\mathbf{G}) = -\int\ln\frac{g}{p}p\mu(dx) \geqslant -2\left(\int\sqrt{pg}\mu(dx)-1\right) = \rho_3(\mathbf{P},\mathbf{G}).$$

В дальнейшем будем рассматривать параметрический случай и предполагать, что выполнено условие (A_μ). Нас будут интересовать расстояния ρ_i, $i=1,2,3$, между распределениями $\mathbf{P}=\mathbf{P}_{\theta_1}$ и $\mathbf{G}=\mathbf{P}_{\theta_2}$ в $(\mathcal{X},\mathfrak{B}_\mathcal{X})$, а также между соответствующими выборочными распределениями (обозначим их здесь через $\mathbf{P}^n_{\theta_1}$, $\mathbf{P}^n_{\theta_2}$) в $\left(\mathcal{X}^n,\mathfrak{B}^n_\mathcal{X}\right)$. (Отметим, что расстояния имеют смысл для произвольных распределений и с природой пространств никак не связаны.) Если $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$, то мы можем записать

$$\rho_1(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2}) = \int\ln\frac{f_{\theta_1}}{f_{\theta_2}}f_{\theta_1}\mu(dx) = \mathbf{E}_{\theta_1}\ln\frac{f_{\theta_1}(\mathrm{x}_1)}{f_{\theta_2}(\mathrm{x}_1)},$$

$$\rho_2(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2}) = \int\frac{(f_{\theta_1}-f_{\theta_2})^2}{f_{\theta_1}}\mu(dx) = \mathbf{E}_{\theta_1}\left(\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}-1\right)^2, \tag{2}$$

$$\rho_3(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2}) = \int(\sqrt{f_{\theta_1}}-\sqrt{f_{\theta_2}})^2\mu(dx) = \mathbf{E}_{\theta_1}\left(\sqrt{\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}}-1\right)^2.$$

Если условие $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$ не выполнено, то $\rho_2(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2})$, $\rho_3(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2})$ будут больше соответствующих математических ожиданий в (2).

Отметим, что наряду с (2) имеет место следующее полезное равенство, вытекающее из (1):

$$\mathbf{E}_{\theta_1}\sqrt{f_{\theta_2}(\mathrm{x}_1)/f_{\theta_1}(\mathrm{x}_1)} = \int\sqrt{f_{\theta_2}(x)f_{\theta_1}(x)}\,\mu(dx) = 1-\frac{1}{2}\rho_3(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2}). \tag{3}$$

Связь между расстояниями $\rho_i(\mathbf{P}_{\theta_1},\mathbf{P}_{\theta_2})$ и $\rho_i\left(\mathbf{P}^n_{\theta_1},\mathbf{P}^n_{\theta_2}\right)$ устанавливается следующим утверждением.

Теорема 1. *Справедливы равенства*

$$\begin{aligned} \rho_1\bigl(\mathbf{P}_{\theta_1}^n, \mathbf{P}_{\theta_2}^n\bigr) &= n\rho_1(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}), \\ 1 + \rho_2\bigl(\mathbf{P}_{\theta_1}^n, \mathbf{P}_{\theta_2}^n\bigr) &= (1 + \rho_2(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}))^n, \\ 1 - \frac{1}{2}\rho_3\bigl(\mathbf{P}_{\theta_1}^n, \mathbf{P}_{\theta_2}^n\bigr) &= \left(1 - \frac{1}{2}\rho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})\right)^n. \end{aligned} \tag{4}$$

Доказательство почти очевидно, если предположить для простоты, что $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$ (в общем случае выкладки по существу сохранятся, но станут чуть более громоздкими). Действительно, в этом случае мы можем воспользоваться равенствами (2). Тогда первое из соотношений (4) немедленно следует из того, что

$$\ln \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} = \sum_{i=1}^n \ln \frac{f_{\theta_1}(\mathrm{x}_i)}{f_{\theta_2}(\mathrm{x}_i)}.$$

Далее, в силу (2), (3)

$$\begin{aligned} 1 + \rho_2(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}) &= \mathbf{E}_{\theta_1}\left(\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}\right)^2, \\ 1 - \frac{1}{2}\rho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}) &= \mathbf{E}_{\theta_1}\sqrt{\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}}, \end{aligned}$$

и такие же соотношения справедливы для расстояний между $\mathbf{P}_{\theta_1}^n$ и $\mathbf{P}_{\theta_2}^n$ (с заменой в правых частях x_1 на X). Так как

$$\mathbf{E}_{\theta_1}\left(\frac{f_{\theta_2}(X)}{f_{\theta_1}(X)}\right)^\alpha = \mathbf{E}_{\theta_1}\prod_{i=1}^n\left(\frac{f_{\theta_2}(\mathrm{x}_i)}{f_{\theta_1}(\mathrm{x}_i)}\right)^\alpha = \left[\mathbf{E}_{\theta_1}\left(\frac{f_{\theta_2}(\mathrm{x}_1)}{f_{\theta_1}(\mathrm{x}_1)}\right)^\alpha\right]^n,$$

то отсюда при $\alpha = 2$ и $\alpha = 1/2$ получаем (4).

Доказательство теоремы в общем случае (т. е. если условие $N_{P_{\theta_2}} \subset N_{P_{\theta_1}}$ не выполнено) мы предоставляем читателю. $\triangleleft$

Из теоремы 1 вытекает

Следствие 1. *Справедливо неравенство*

$$\rho_3\bigl(\mathbf{P}_{\theta_1}^n, \mathbf{P}_{\theta_2}^n\bigr) \leqslant n\rho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}).$$

Действительно, $1 - \beta^n \leqslant (1 - \beta)n$ при любом $\beta \geqslant 0$. Полагая $\beta = 1 - \frac{1}{2}\rho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})$, получим из (4)

$$\rho_3\bigl(\mathbf{P}_{\theta_1}^n, \mathbf{P}_{\theta_2}^n\bigr) = 2(1 - \beta^n) \leqslant 2(1 - \beta)n = n\rho_3(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}). \quad \triangleleft$$

2. Связь расстояний Хеллингера и других с информацией Фишера. Среди трех введенных в предыдущем пункте расстояний наибольший интерес для нас в дальнейшем будет представлять расстояние Хеллингера. В то же время характер главных утверждений, приводимых ниже (теоремы 2, 3), и характер доказательств будет один и тот же для всех трех расстояний. Поэтому для краткости изложения ограничимся в этом пункте изучением расстояния Хеллингера, которое будем обозначать теперь (опуская индекс у символа ρ_3)

$$\rho(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2}) = \int (\sqrt{f_{\theta_1}} - \sqrt{f_{\theta_2}})^2 \mu(dx).$$

Положим $r(\theta_1, \theta_2) = \rho(\mathbf{P}_{\theta_1}, \mathbf{P}_{\theta_2})$.

Лемма 1. *Если $f_\theta(x)$ при п. в. $[\mu]$ значениях x непрерывна по θ, $\theta_1 \neq \neq \theta_2$, то*

$$\liminf_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2}. \tag{5}$$

Если функция $\sqrt{f_\theta(x)}$ при п. в. $[\mu]$ значениях x дифференцируема по θ, то

$$\liminf_{\substack{\theta' \to \theta \\ \theta'' \to \theta}} \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{I(\theta)}{4}. \tag{6}$$

Кроме того,

$$\frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4} \int_0^1 I(\theta_1 + (\theta_2 - \theta_1)y)\, dy. \tag{7}$$

Здесь предполагается, конечно, что значения θ', θ'', θ_1, θ_2, θ принадлежат Θ.

Доказательство. Чтобы проверить (5), достаточно воспользоваться леммой Фату и непрерывностью $f_\theta(x)$ в соотношении

$$\liminf_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \int \liminf_{\substack{\theta' \to \theta_1 \\ \theta'' \to \theta_2}} \left(\frac{\sqrt{f_{\theta'}} - \sqrt{f_{\theta''}}}{\theta' - \theta''} \right)^2 \mu(dx).$$

Так как при $\theta_1 = \theta_2 = \theta$ подынтегральное выражение в последнем интеграле равно $(f'_\theta)^2/(4f_\theta)$, то мы получаем (6).

Чтобы доказать (7), положим $a = \theta_2 - \theta_1$ и представим приращение $\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}}$ в виде

$$\frac{1}{2} \int_{\theta_1}^{\theta_2} \frac{f'_t}{\sqrt{f_t}}\, dt = \frac{a}{2} \int_0^1 \frac{f'_{\theta_1 + ay}}{\sqrt{f_{\theta_1 + ay}}}\, dy.$$

В силу неравенства Коши–Буняковского

$$(\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}})^2 = \frac{a^2}{4}\left[\int_0^1 \frac{f'_{\theta_1+ay}}{\sqrt{f_{\theta_1+ay}}}\,dy\right]^2 \leqslant \frac{a^2}{4}\int_0^1 \frac{(f'_{\theta_1+ay})^2}{f_{\theta_1+ay}}\,dy.$$

Пользуясь неотрицательностью подынтегральной функции, мы можем менять порядок интегрирования в следующих соотношениях:

$$\frac{r(\theta_1,\theta_2)}{a^2} \leqslant \frac{1}{4}\int_{\mathcal{X}}\left(\int_0^1 \frac{(f'_{\theta_1+ay})^2}{f_{\theta_1+ay}}\,dy\right)\mu(dx) = \frac{1}{4}\int_0^1 I(\theta_1+ay)\,dy.$$

Неравенство (7) доказано. $\triangleleft$

Положим $r(\Delta) = r(\theta, \theta+\Delta)$. Из леммы 1 непосредственно вытекает

Теорема 2. *Если функция* $\sqrt{f_\theta(x)}$ *при п. в.* $[\mu]$ *значениях* x *дифференцируема по* θ *и* $I(\theta)$ *непрерывна, то существует*

$$\lim_{\Delta\to 0}\frac{r(\Delta)}{\Delta^2} = \frac{I(\theta)}{4}. \tag{8}$$

Замечание 1. Это утверждение остается справедливым и для расстояний ρ_1, ρ_2, если положить

$$r(\Delta) = \frac{1}{4}\rho_2(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta}), \quad r(\Delta) = \frac{1}{2}\rho_1(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta}).$$

Соотношение (6) доказывается при этом точно так же, как в лемме 1. Доказательство же (8) может потребовать привлечения дополнительных условий регулярности (близких к условиям (R)), обеспечивающих законность предельного перехода под знаком интеграла.

Итак, $\rho_i(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta})$, $i = 1, 2, 3$, ведут себя асимптотически одинаково, и $I(\theta)$ характеризует скорость их стремления к нулю при $\Delta \to 0$ (ведь $\frac{1}{4}I(\theta)$ пропорционально второй производной $r(v)$ в точке $v = 0$).

Если положить $r^{(n)}(\Delta) = \rho(\mathbf{P}^n_{\theta+\Delta}, \mathbf{P}^n_\theta)$, то из теорем 1, 2 будет следовать

$$\lim_{\Delta\to 0}\frac{r^{(n)}(\Delta)}{\Delta^2} = \frac{nI(\theta)}{4}.$$

Такие же соотношения сохранятся для расстояний ρ_1, ρ_2.

3. Существование равномерных границ для $r(\Delta)/\Delta^2$. Наличие таких границ позволит нам в дальнейшем получить весьма полезные оценки для моментов отношения правдоподобия.

Для того чтобы упростить изложение и избежать введения других более громоздких условий, часто будем предполагать в последующих рассмотрениях выполненным условие

(A_c) *Множество* Θ *компактно.*

С точки зрения приложений это условие, означающее ограниченность и замкнутость параметрического множества, как правило, не представляется ограничительным.

Ниже мы будем использовать также условие (A_0), которое было введено нами в § 14, 16 и которое означает, что $f_{\theta_1} \neq f_{\theta_2}$ при $\theta_1 \neq \theta_2$. При этом условии $r(\theta_1, \theta_2) > 0$ при $\theta_1 \neq \theta_2$.

Теорема 3. *Если выполнены условия* (A_0), (A_c) *и* $0 < I(\theta) \leqslant 4h < \infty$ *при всех* $\theta \in \Theta$, *то существует постоянная* $g > 0$ *такая, что для всех* $\theta_1, \theta_2 \in \Theta$

$$g \leqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant h. \tag{9}$$

Доказательство. Оценка сверху следует непосредственно из (7). Покажем теперь, что

$$\inf_{\theta_1, \theta_2} \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \geqslant g > 0. \tag{10}$$

Допустим, что (10) неверно, тогда найдется последовательность $\left(\theta_1^{(n)}, \theta_2^{(n)}\right)$ такая, что

$$\frac{r\left(\theta_1^{(n)}, \theta_2^{(n)}\right)}{\left|\theta_1^{(n)} - \theta_2^{(n)}\right|^2} \to 0 \tag{11}$$

при $n \to \infty$. В силу условия (A_c) можем считать, не ограничивая общности, что $\theta_1^{(n)} \to \theta_1 \in \Theta$, $\theta_2^{(n)} \to \theta_2 \in \Theta$. Если $\theta_1 \neq \theta_2$, то (11) противоречит (5), так как в силу условия (A_0) $r(\theta_1, \theta_2) > 0$. Если же $\theta_1 = \theta_2 = \theta$, то (11) противоречит (6), так как $I(\theta) > 0$. $\triangleleft$

4. Многомерный случай. В этом пункте получим аналоги утверждений п. 2, 3 для многомерного параметра (содержание п. 1 с размерностью θ не связано). Обозначим через $\varphi(x, \theta)$ вектор-функцию с координатами

$$\varphi_i(x, \theta) = \frac{1}{\sqrt{f_\theta(x)}} \frac{\partial f_\theta(x)}{\partial \theta_i}.$$

Тогда производная функция $\sqrt{f_\theta(x)}$ по направлению единичного вектора $\omega = (\omega_1, \ldots, \omega_k)$ равна $((\sqrt{f_\theta(x)})', \omega) = (\operatorname{grad} \sqrt{f_\theta(x)}, \omega) = \frac{1}{2}(\varphi(x, \theta), \omega)$. Матрица Фишера $I(\theta)$ равна в этих обозначениях

$$I(\theta) = \int \varphi^\top(x, \theta) \varphi(x, \theta) \mu(dx).$$

Пусть $|u|$ означает евклидову норму $u = (u_1, \ldots, u_k)$.

В многомерном случае имеет место следующее обобщение леммы 1.

Лемма 1А. *Первое утверждение леммы* 1 (см. (5)) *при* $k > 1$ *полностью сохраняется.*

Если функция $\sqrt{f_\theta(x)}$ *при п. в.* $[\mu]$ *значениях* x *дифференцируема по* θ, $\theta' \to \theta$, $\theta'' = \theta' + \omega''\delta$, $\omega'' \to \omega$, $|\omega''| = |\omega| = 1$, $\delta \to 0$, *то*

$$\liminf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \frac{1}{4}\omega I(\theta)\omega^\top. \tag{12}$$

Кроме того, если ω, $|\omega| = 1$, *есть вектор, коллинеарный* $\theta_2 - \theta_1$, *так что* $\theta_2 = \theta_1 + a\omega$, $a = |\theta_2 - \theta_1|$, *то*

$$\frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4}\int_0^1 \omega I(\theta_1 + a\omega y)\omega^\top \, dy. \tag{13}$$

Доказательство. Первое утверждение леммы 1 с размерностью не связано. Второе вытекает из леммы Фату и соотношений

$$\liminf \frac{r(\theta', \theta'')}{|\theta' - \theta''|^2} \geqslant \int \liminf \frac{(\sqrt{f_{\theta'}} - \sqrt{f_{\theta''}})^2}{|\theta' - \theta''|^2}\mu(dx) =$$

$$= \frac{1}{4}\int (\varphi(x, \theta), \omega)^2 \mu(dx) = \frac{1}{4}\omega I(\theta)\omega^\top.$$

Чтобы доказать (13), заметим, что

$$\sqrt{f_{\theta_2}} - \sqrt{f_{\theta_1}} = \frac{1}{2}\int_0^a (\varphi(x, \theta_1 + y\omega), \omega)\, dy = \frac{a}{2}\int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega)\, dy,$$

$$r(\theta_1, \theta_2) = \frac{a^2}{4}\int_{\mathcal{X}} \left[\int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega)\, dy\right]^2 \mu(dx) \leqslant$$

$$\leqslant \frac{a^2}{4}\int_{\mathcal{X}}\int_0^1 (\varphi(x, \theta_1 + ay\omega), \omega)^2 \, dy\mu(dx) =$$

$$= \frac{a^2}{4}\int_0^1\int_{\mathcal{X}} (\varphi(x, \theta_1 + ay\omega), \omega)^2 \mu(dx)\, dy = \frac{a^2}{4}\int_0^1 \omega I(\theta_1 + ay\omega)\omega^\top \, dy. \quad \triangleleft$$

Положим, как и прежде, $r(\Delta) = r(\theta, \theta + \Delta)$. Из леммы 1А вытекает

Теорема 2А. *Если функция* $\sqrt{f_\theta(x)}$ *дифференцируема при п. в.* $[\mu]$ *значениях* x *и матрица* $I(\theta)$ *непрерывна, то для любого вектора* ω *единичной длины существует*

$$\lim_{\delta \to 0} \frac{r(\delta\omega)}{\delta^2} = \frac{1}{4}\omega I(\theta)\omega^\top.$$

Как и в одномерном случае, из леммы 1А мы можем получить также следующее следствие. Обозначим через $\operatorname{Sp} I(\theta)$ след матрицы $I(\theta)$.

Теорема 3А. *Если выполнены условия* (A_0), (A_c), *матрица* $I(\theta)$ *является положительно определенной в* Θ, $4h \equiv \sup\limits_{\theta} \operatorname{Sp} I(\theta) < \infty$, *то существует постоянная* $g > 0$ *такая, что при всех* $\theta_1, \theta_2 \in \Theta$

$$g \leqslant \frac{r(\theta_1, \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant h. \tag{14}$$

Доказательство. Обозначим через $\Lambda_1(\theta)$ и $\Lambda_k(\theta)$ соответственно минимальное и максимальное собственные числа матрицы $I(\theta)$, так что при $|\omega| = 1$

$$\Lambda_1(\theta) \leqslant \omega I(\theta) \omega^\top \leqslant \Lambda_k(\theta). \tag{15}$$

По условию теоремы $\Lambda_1(\theta) > 0$ всюду на Θ. Так как $(\varphi, \omega)^2 \leqslant |\varphi|^2 = \sum\limits_{j=1}^{k} \varphi_j^2$, то

$$\int\limits_{\mathcal{X}} (\varphi, \omega)^2 \mu(dx) = \omega I(\theta) \omega^\top \leqslant \operatorname{Sp} I(\theta)$$

и, стало быть, $\Lambda_k(\theta) \leqslant \operatorname{Sp} I(\theta) \leqslant 4h$. Из неравенства (13) получаем

$$\frac{r(\theta_1 - \theta_2)}{|\theta_1 - \theta_2|^2} \leqslant \frac{1}{4} \int\limits_0^1 \Lambda_k(\theta_1 + ay\omega)\, dy \leqslant h.$$

Докажем теперь второе неравенство в (14). Предположим, что оно не верно. Тогда, как и в теореме 3, найдется последовательность $(\theta_1^{(n)}, \theta_2^{(n)})$, $\theta_1^{(n)} \to \theta_1 \in \Theta$, $\theta_2^{(n)} \to \theta_2 \in \Theta$, для которой будет справедливо (11). Если $\theta_1 \neq \theta_2$, то это будет противоречить (5). Если $\theta_1 = \theta_2 = \theta$, то в силу компактности сферы $|\omega| = 1$ можно считать, не ограничивая общности, что $\theta_2^{(n)} = \theta_1^{(n)} + \delta\omega^{(n)}$, $\omega^{(n)} \to \omega$, $|\omega^{(n)}| = |\omega| = 1$. Но в этом случае (11) будет противоречить (12), (15). $\triangleleft$

5*. Связь рассматриваемых расстояний с оценками. Рассмотрим расстояние Кульбака–Лейблера между распределениями $\mathbf{P}_\theta$ и не зависящим от θ распределением $\mathbf{G}$

$$\rho_1(\mathbf{G}, \mathbf{P}_\theta) = \int \ln \frac{d\mathbf{G}}{d\mu} \mathbf{G}(dx) - \int \ln f_\theta(x) \mathbf{G}(dx).$$

Здесь от θ зависит лишь второе слагаемое

$$d(\mathbf{P}_\theta, \mathbf{G}) = -\int \ln f_\theta(x) \mathbf{G}(dx).$$

С другой стороны, напомним, что о. м. п. определялась нами в § 16 как значение θ, при котором минимизируется $d(\mathbf{P}_\theta, \mathbf{P}_n^*)$. Если распределение x_1 дискретно, а μ — считающая мера, то выражение

$$d(\mathbf{P}_n^*, \mathbf{P}_n^*) = -\int \ln \frac{d\mathbf{P}_n^*}{d\mu} \mathbf{P}_n^*(dx)$$

имеет смысл, $\rho_1(\mathbf{P}_n^*, \mathbf{P}_\theta) = d(\mathbf{P}_\theta, \mathbf{P}_n^*) - d(\mathbf{P}_n^*, \mathbf{P}_n^*)$ и, следовательно, мы можем считать, что о. м. п. минимизирует расстояние Кульбака–Лейблера $\rho_1(\mathbf{P}_n^*, \mathbf{P}_\theta)$ между $\mathbf{P}_\theta$ и $\mathbf{P}_n^*$. В общем случае такое толкование можно принимать лишь условно.

Основанием для такого толкования может служить приближенное описание распределения $\mathbf{P}_\theta$ с помощью дискретного распределения $\mathbf{P}_\theta^N$. Для этого надо предположить, например, что существуют множества $\Delta_1, \ldots, \Delta_N$ такие, что при больших N (малых Δ_i) $f_\theta(x)$ при $x \in \Delta_i$ хорошо приближается (в подходящем смысле) значениями $f_\theta(x_i)$, $y_i \in \Delta_i$, $i = 1, \ldots, N$. Тогда можно считать, что $f_\theta(x)$ постоянна на Δ_i и равна $f_\theta(y_i)$, а наблюдения, попавшие в Δ_i, отождествляются (объединяются в группы). Полученное новое сгруппированное эмпирическое распределение обозначим через $\mathbf{P}_n^{*N}$. Соответствующие вероятности попадания в i-ю группу равны

$$\mathbf{P}_\theta^N(\Delta_i) = \int_{\Delta_i} f_\theta(x)\mu(dx) \sim f_\theta(y_i)\mu(\Delta_i).$$

В этом случае будем иметь дело с выборкой из распределения $\mathbf{P}_\theta^N$ дискретного типа. При этом о. м. п. $\widehat{\theta}^*$ будет близка значению $\widehat{\theta}_N^*$, максимизирующему

$$\sum_i \ln f_\theta(y_i)\mathbf{P}_n^*(\Delta_i) \approx \int \ln f_\theta(x)\mathbf{P}_n^*(dx).$$

Другими словами, $\widehat{\theta}_N^*$ будут сближаться с ростом N с $\widehat{\theta}^*$, где $\widehat{\theta}_N^*$ суть значения, минимизирующие расстояния Кульбака–Лейблера $\rho_1(\mathbf{P}_n^{*N}, \mathbf{P}_\theta^N)$. Более содержательный «предельный переход» здесь построить трудно, так как значения $\rho_1(\mathbf{P}_n^{*N}, \mathbf{P}_\theta^N)$, как и $d(\mathbf{P}_n^{*N}, \mathbf{P}_\theta^N)$, неограниченно возрастают при $N \to \infty$ ($\mu(\Delta_i) \to 0$).

Для дискретных распределений x_1 можно рассматривать также расстояния $\rho_i(\mathbf{P}_\theta, \mathbf{P}_n^*)$ при $i = 2, 3$ и оценки, минимизирующие эти расстояния. Например, при $i = 2$ получим

$$\rho_2(\mathbf{P}_\theta, \mathbf{P}_n^*) = \sum_i \frac{(\nu_i/n - f_\theta(a_i))^2}{f_\theta(a_i)},$$

где ν_i — число элементов выборки, попавших в точку a_i, для которой $f_\theta(a_i) = \mathbf{P}_\theta(\{a_i\}) > 0$. Это есть статистика χ^2 (см. § 7, 8), в связи с чем это наименование присвоено нами и расстоянию ρ_2.

Так как расстояния ρ_i обладают близкими асимптотическими свойствами, то оценки, минимизирующие эти расстояния, как это выяснится позже, будут асимптотически совпадать.

§ 32.* Свойства информации Фишера. Разностное неравенство типа Рао–Крамера

Этот параграф стоит несколько в стороне от основного изложения. В нем мы попытаемся хотя бы частично ответить на вопрос о том, что происходит с нижней допустимой границей для $\mathbf{E}_\theta(\theta^* - \theta)^2$ в нерегулярном случае, т. е. в случае, когда функция $f_\theta(x)$ не дифференцируема по θ или когда $I(\theta) = \infty$.

Начнем с примера, который показывает, что при этих условиях поведение среднеквадратичных уклонений оценок (или их дисперсий) может быть совсем иным, чем поведение правой части неравенства Рао–Крамера.

Пример 1. Пусть $X \in \mathbf{U}_{0,\theta}$. Условие (R) здесь не выполнено, так как функция $f_\theta(x)$ разрывна. Как мы знаем, для этого семейства статистика $S = \max \mathrm{x}_i$ является полной и достаточной (см. пример 24.3). Возьмем несмещенную оценку $\theta^* = 2\mathrm{x}_1$. Тогда в силу результатов § 24 оценка $\theta^*_S = 2\mathbf{E}_\theta(\mathrm{x}_1/S)$ будет эффективной. Вычислим значение $\mathbf{E}_\theta(\mathrm{x}_1/S)$. Так как $\mathbf{P}_\theta(S < z) = (z/\theta)^n$, $z \in [0, \theta]$, то S имеет плотность, равную nz^{n-1}/θ^n на $[0, \theta]$ и нулю вне этого интервала. Для отыскания условного распределения $P(B/s) = \mathbf{P}_\theta(\mathrm{x}_1 \in B/S = s)$ величины x_1 при условии $S = s$ воспользуемся правилом (20.2):

$$P(dy/s) = \mathbf{P}_\theta(\mathrm{x}_1 \in dy/S = s) = \frac{\mathbf{P}_\theta(\mathrm{x}_1 \in dy,\ S \in ds)}{\mathbf{P}_\theta(S \in ds)}.$$

Здесь числитель равен

$$\mathbf{P}_\theta(\mathrm{x}_1 \in dy,\ S \in ds) = \begin{cases} \dfrac{dy}{\theta}\dfrac{(n-1)s^{n-2}ds}{\theta^{n-1}} & \text{при} \quad y < s, \\ \dfrac{ds}{\theta}\dfrac{s^{n-1}}{\theta^{n-1}} & \text{при} \quad y = s, \\ 0 & \text{при} \quad y > s. \end{cases}$$

Отсюда следует, что $P(dy/s) = \dfrac{(n-1)\,dy}{ns}$ при $0 \leqslant y < s$, $P(\{s\}/s) = \dfrac{1}{n}$. Таким образом,

$$\mathbf{E}_\theta(\mathrm{x}_1/S) = \int_0^S y\frac{n-1}{nS}\,dy + \frac{S}{n} = \frac{S(n-1)}{2n} + \frac{S}{n} = \frac{n+1}{2n}S,$$

$$\theta^*_S = S\left(1 + \frac{1}{n}\right).$$

Имеем

$$\mathbf{D}_\theta\theta^*_S = \mathbf{E}_\theta\left(\theta^*_S\right)^2 - \theta^2 = \int_0^\theta s^2\left(1 + \frac{1}{n}\right)^2 \frac{ns^{n-1}}{\theta^n}\,ds - \theta^2 =$$

$$= \left(\frac{(n+1)^2}{n(n+2)} - 1\right)\theta^2 = \frac{\theta^2}{n(n+2)}. \quad (1)$$

Так как θ_S^* эффективна, для любой несмещенной оценки θ^* получаем

$$\mathbf{D}_\theta\theta^* \geqslant \frac{\theta^2}{n(n+2)}. \tag{2}$$

Таким образом, при больших n среднеквадратичное уклонение $\mathbf{E}_\theta\big(\theta_S^* - \theta\big)^2$ будет иметь порядок малости $1/n^2$. С точки зрения нижней границы неравенства Рао–Крамера, имеющей порядок $1/n$, это ненормально высокая точность *). Можно показать, что это есть точность, с которой по выборке определяются любые точки скачков $f_\theta(x)$ (запрещаемые условием (R)). Мы видели в примере 17.4 об оценке медианы, что точки, где плотность $f_\theta(x)$ бесконечна, можно определить с еще большей точностью, так что, грубо говоря, чем больше нарушение регулярности в точке, тем точнее эта точка оценивается по выборке. Скажем, если $X \in \mathbf{P}_\theta$, где $\mathbf{P}_\theta = \frac{1}{2}\mathbf{U}_{0,\theta} + \frac{1}{2}\mathbf{I}_\theta$, $\mathbf{I}_\theta$ есть распределение, сосредоточенное в точке θ, то $\mathbf{P}_\theta(S \neq \theta) = 2^{-n}$ $(S = \max \mathrm{x}_i)$, так что дисперсия $\theta^* - \theta$ при $\theta^* = S$ будет убывать с ростом n экспоненциальным образом.

Можно ли в этих условиях указать нижнюю границу для дисперсии оценок? Ниже мы получим неравенство, аналогичное неравенству Рао–Крамера, с помощью которого такие границы можно строить при более слабых условиях регулярности, чем условие (R).

Будем предполагать лишь, что выполнено условие (A_μ), хотя оно и не очень существенно (см. замечание в конце параграфа).

Обозначим через $\Delta\varphi(\theta)$ приращение функции $\varphi(\theta)$ на интервале $(\theta, \theta + \Delta)$, через $N_{P_\theta}^n$ — носитель в $\mathcal{X}^n$ распределения выборки: $N_{P_\theta}^n = \{\boldsymbol{x} : f_\theta(\boldsymbol{x}) \neq 0\}$ и положим $N^n = N_{P_\theta}^n \cup N_{P_{\theta+\Delta}}^n$.

Теорема 1 (Неравенство Чепмена–Роббинса). *Пусть* $\theta \in \Theta$, $\theta + \Delta \in \Theta$, $a(\theta) = \mathbf{E}_\theta\theta^*$. *Тогда при любом* $\Delta \neq 0$

$$\mathbf{D}_\theta\theta^* \geqslant \frac{(\Delta a(\theta))^2}{\int[\Delta f_\theta(\boldsymbol{x})]^2/f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x})} = \frac{(\Delta a(\theta))^2}{\rho_2\big(\mathbf{P}_\theta^n, \mathbf{P}_{\theta+\Delta}^n\big)}, \tag{3}$$

где ρ_2 *есть расстояние* χ^2, *рассматривавшееся в* § 31. *Для несмещенных оценок числитель здесь следует заменить на* Δ^2.

В силу теоремы 31.1 знаменатель в (3) имеет вид $\rho_2\big(\mathbf{P}_\theta^n, \mathbf{P}_{\theta+\Delta}^n\big) = \big(1 + r_2(\Delta)\big)^n - 1$, где

$$r_2(\Delta) = \rho_2(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta}) = \int \frac{[\Delta f_\theta(x)]^2}{f_\theta(x)}\mu(dx). \tag{4}$$

Таким образом, чем больше расстояние $\rho_2(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta})$ между $\mathbf{P}_{\theta+\Delta}$ и $\mathbf{P}_\theta$ (при фиксированном Δ), тем меньше нижняя граница для $\mathbf{D}\theta^*$.

*) Для параметра θ существуют также оценки, дисперсия которых имеет порядок $1/n$. Например, для оценки $\theta^{**} = 2\bar{\mathrm{x}}$ имеем $\mathbf{E}\theta^{**} = \theta$, $\mathbf{D}\theta^{**} = 4\mathbf{D}\mathrm{x}_1/n = \theta^2/3n$.

Если $\mathbf{P}_{\theta+\Delta}$ абсолютно непрерывно относительно $\mathbf{P}_\theta$, то $N^n_{P_{\theta+\Delta}} \subset N^n_{P_\theta} = N^n$ и $\rho_2(\mathbf{P}^n_\theta, \mathbf{P}^n_{\theta+\Delta})$ можно записать в виде (см. (31.2))

$$\rho_2(\mathbf{P}^n_\theta, \mathbf{P}^n_{\theta+\Delta}) = \mathbf{E}_\theta \left[\frac{\Delta f_\theta(X)}{f_\theta(X)}\right]^2;$$

аналогично $r_2(\Delta) = \mathbf{E}_\theta \left[\dfrac{\Delta f_\theta(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)}\right]^2$.

Если же распределение $\mathbf{P}_{\theta+\Delta}$ не абсолютно непрерывно относительно $\mathbf{P}_\theta$, то существует подмножество из $N_{P_{\theta+\Delta}}$ положительной $\mathbf{P}_{\theta+\Delta}$-меры, на котором $f_\theta(x) = 0$, так что интеграл в (4) обращается в бесконечность, а само неравенство (3) становится тривиальным. Следует вновь отметить, что выражение $\mathbf{E}_\theta[\Delta f_\theta(X)/f_\theta(X)]^2$, понимаемое как интеграл по N_{P_θ}, может при этом оставаться конечным.

Доказательство теоремы 1. Из сказанного выше следует, что, не ограничивая общности, можем считать, что $\mathbf{P}_{\theta+\Delta}$ абсолютно непрерывно относительно $\mathbf{P}_\theta$, так что $N^n_{P_{\theta+\Delta}} \subset N^n_{P_\theta} = N^n$. Так как $f_\theta(\boldsymbol{x})$ и $f_{\theta+\Delta}(\boldsymbol{x})$ есть плотности в $\mathcal{X}^n$, то

$$\int \Delta f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = 0.$$

Кроме того,

$$\int \theta^* \Delta f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \Delta a(\theta).$$

Отсюда следует, что

$$\int\limits_{N^n} (\theta^* - a(\theta))\Delta f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \Delta a(\theta). \tag{5}$$

На множестве N^n можем подынтегральную функцию в (5) представить в виде произведения

$$(\theta^* - a(\theta))\sqrt{f_\theta(x)}\,\frac{\Delta f_\theta(x)}{\sqrt{f_\theta(x)}}.$$

Применяя затем неравенство Коши–Буняковского, получим

$$(\Delta a(\theta))^2 \leqslant \int\limits_{N^n} (\theta^* - a(\theta))^2 f_\theta(\boldsymbol{x})\mu^n(d\boldsymbol{x}) \int\limits_{N^n} \frac{(\Delta f_\theta(\boldsymbol{x}))^2}{f_\theta(\boldsymbol{x})}\mu^n(d\boldsymbol{x}). \quad \triangleleft$$

В дальнейшем, в соответствии со сделанным выше замечанием, как и в доказательстве теоремы 1, ограничимся случаем, когда $\mathbf{P}_{\theta+\Delta}$ абсолютно непрерывно относительно $\mathbf{P}_\theta$ (ведь иначе неравенство (3) становится тривиальным).

Следствие 1. *Если выполнены условия регулярности, обеспечивающие существование* (см. замечание 31.1 к теореме 31.2) $\lim\limits_{\Delta\to 0} r_2(\Delta)/\Delta^2 = I(\theta)$, *то*

$$\mathbf{D}_\theta\theta^* \geqslant \frac{(a'_+(\theta))^2}{nI(\theta)}, \tag{6}$$

где $a'_+(\theta) = \limsup\limits_{\Delta\to 0} \dfrac{\Delta a(\theta)}{\Delta}$.

Чтобы получить (6) из теоремы 1, надо лишь заметить, что мы можем выбрать последовательность $\Delta \to 0$ так, чтобы $\dfrac{\Delta a(\theta)}{\Delta} \to a'_+(\theta)$. $\triangleleft$

Неравенство (6) является по форме некоторым обобщением неравенства Рао–Крамера (обобщением, скорее всего, фиктивным, так как упомянутые условия регулярности, по-видимому, влекут за собой существование $a'(\theta)$).

Неравенство (3) естественно называть *разностным* в отличие от неравенства (6), которое можно было бы назвать *дифференциальным*.

Таким образом, если $r_2(\Delta) \sim I(\theta)\Delta^2$ (это соответствует тому, что f_θ дифференцируема), то из *разностного неравенства Чепмена–Роббинса следует дифференциальное неравенство Рао–Крамера.*

Однако если функция f_θ не дифференцируема, поведение $r_2(\Delta)$ при уменьшении Δ будет иным.

Если, скажем, f_θ дифференцируема всюду за исключением конечного числа точек разрыва $\theta = \theta(x)$, зависящих от x, то мы будем иметь

$$r_2(\Delta) \sim c|\Delta|. \tag{7}$$

Проще всего это прояснить на довольно типичном примере, рассмотренном в начале параграфа.

Пусть $X \in \mathbf{U}_{0,\theta}$. Чтобы было выполнено условие абсолютной непрерывности $\mathbf{P}_{\theta+\Delta}$ относительно $\mathbf{P}_\theta$, будем считать в случае $\mathbf{P}_\theta = \mathbf{U}_{0,\theta}$, что $\Delta < 0$, $|\Delta| < \theta$. Тогда

$$\Delta f_\theta(x) = \begin{cases} \dfrac{1}{\theta+\Delta} - \dfrac{1}{\theta} & \text{при} \quad x \in [0, \theta+\Delta], \\ -\dfrac{1}{\theta} & \text{при} \quad x \in [\theta+\Delta, \theta], \\ 0 & \text{при} \quad x \notin [0,\theta], \end{cases}$$

$$r_2(\Delta) = \int\limits_0^\theta \frac{(\Delta f_\theta(x))^2}{f_\theta(x)}\,dx =$$

$$= \int\limits_0^{\theta+\Delta} \left[\frac{\Delta}{\theta(\theta+\Delta)}\right]^2 \theta\,dx + \int\limits_{\theta+\Delta}^{\theta} \frac{1}{\theta^2}\theta\,dx = \frac{\Delta^2}{\theta(\theta+\Delta)} + \frac{|\Delta|}{\theta}.$$

Существенным здесь является наличие интервала длины, сравнимой с $|\Delta|$, на котором $|\Delta f_\theta(x)| > c > 0$, где c от Δ не зависит. Это и обеспечивает порядок малости (7) для $r_2(\Delta)$.

Возвращаясь к нашему примеру, видим, что для несмещенных оценок параметра θ

$$\mathbf{D}\theta^* \geqslant \max_{\Delta} \frac{\Delta^2}{\left(1 + \dfrac{|\Delta|}{\theta} + \dfrac{\Delta^2}{\theta(\theta+\Delta)}\right)^n - 1}.$$

Каков порядок малости правой части этого неравенства при $n \to \infty$? Полагая $|\Delta| = y\theta/n$, получим

$$\mathbf{D}\theta^* \geqslant \frac{\theta^2}{n^2} \max_{y} \frac{y^2}{\left(1 + \dfrac{y}{n} + \dfrac{y^2}{n(n-y)}\right)^n - 1}.$$

Ясно, что выражение со знаком max асимптотически эквивалентно $h = \max\limits_y y^2/(e^y - 1) \approx 0{,}65$, так что

$$\mathbf{D}\theta^* \geqslant \frac{\theta^2}{n^2}(h + o(1)).$$

Это неравенство имеет ту же по порядку малости правую часть, что и неулучшаемое неравенство (2), однако постоянный множитель при θ^2/n^2 в (2) «лучше» и равен единице.

Наряду с (7) могут возникать и другие скорости сходимости $r_2(\Delta)$ к нулю при $\Delta \to 0$. Мы можем получить, например, и $r_2(\Delta) \sim c\Delta^\alpha$, $\alpha < 1$, если $f_\theta(x)$ имеет линии $\theta = \theta(x) \neq \text{const}$, при приближении к которым $f_\theta(x) \to \infty$; и $r_2(\Delta) \sim c\Delta^\alpha$, $2 > \alpha > 1$, если f_θ непрерывна по θ, но не дифференцируема, а удовлетворяет лишь условию Гёльдера в окрестности некоторой линии $\theta = \theta(x) \neq \text{const}$. Нетрудно видеть, что порядок малости

$$\max_{\Delta} \frac{\Delta^2}{(1 + c\Delta^\alpha)^n - 1}$$

при $\alpha < 2$ будет определяться значением $\Delta = (y/cn)^{1/\alpha}$, так что

$$\mathbf{D}\theta^* \geqslant \frac{1}{(cn)^{2/\alpha}} \max_{y} \frac{y^{2/\alpha}}{e^y - 1}(1 + o(1)).$$

В «регулярном» случае $\alpha = 2$ максимум по y достигается в предельной точке $y = 0$ ($\Delta = 0$).

В заключение этого параграфа отметим, что оценки для $\mathbf{D}\theta^*$ аналогичным образом могут быть получены и для взаимно не абсолютно непрерывных $\mathbf{P}_\theta$ и $\mathbf{P}_{\theta+\Delta}$. Для этого в (5) подынтегральную функцию следует умножать и делить на $\sqrt{f_\theta(x) + f_{\theta+\Delta}(x)}$, а не на $\sqrt{f_\theta(x)}$. Условие (A_μ) также не столь существенно, так как меры $\mathbf{P}_\theta$ и $\mathbf{P}_{\theta+\Delta}$ всегда абсолютно непрерывны относительно $\dfrac{1}{2}(\mathbf{P}_\theta + \mathbf{P}_{\theta+\Delta})$.

§ 33. Вспомогательные неравенства для отношения правдоподобия. Асимптотические свойства оценок максимального правдоподобия

В § 22–26 мы изучали вопросы, связанные с существованием и отысканием в явном виде эффективных и R-эффективных оценок. Мы видели, что существуют они далеко не всегда, а найти их удается лишь когда функция правдоподобия имеет специальный вид или когда мы знаем в явном виде полную достаточную статистику (первое из этих условий часто влечет за собой второе (см. § 25)).

Перейдем теперь к построению *асимптотически* оптимальных оценок. Здесь условия их существования будут значительно более широкими. Соответствующие результаты будут опираться прежде всего на асимптотические свойства функции

$$Z(u) = \frac{f_{\theta+u}(X)}{f_\theta(X)} = \exp\{L(X, \theta+u) - L(X,\theta)\}, \tag{1}$$

где, как и раньше, $L(X,\theta) = \sum_{i=1}^{n} l(\mathrm{x}_i, \theta)$. Число θ в (1), как правило, будет считаться фиксированным и представлять собой истинное значение параметра, т. е. такое, что $X \in \mathbf{P}_\theta$. В этом случае $Z(u)$ есть функция двух переменных u и X и, стало быть, вместе с функцией правдоподобия $f_{\theta+u}(X)$ будет *случайной функцией* переменного u. Функцию $Z(u)$ будем называть *отношением правдоподобия*, она играет важную роль в математической статистике. Основная задача этого и следующего параграфов — изучить свойства $Z(u)$.

Будет установлено, что $Z(u)$ близко к нулю вне окрестности точки $u = 0$. В окрестности этой точки $Z(u)$ сближается в известном смысле с дельта-функцией, точнее, $Z(v/\sqrt{n}\,)$ сближается асимптотически при $n \to \infty$ с функцией плотности нормального закона.

В § 33, 34 будем рассматривать лишь *одномерный* случай. Случай многомерного параметра будет рассмотрен отдельно в § 36.

Важную роль в последующих оценках будет играть расстояние Хеллингера

$$r(u) = \rho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta) = \int (\sqrt{f_{\theta+u}(x)} - \sqrt{f_\theta(x)}\,)^2 \mu(dx)$$

между распределениями $\mathbf{P}_{\theta+u}$ и $\mathbf{P}_\theta$. Оно рассматривалось нами в § 31. Напомним, что

$$0 \leqslant r(u) = 2\left(1 - \int \sqrt{f_{\theta+u}(x) f_\theta(x)}\,\mu(dx)\right) \leqslant 2,$$

так что

$$\mathbf{E}_\theta \sqrt{\frac{f_{\theta+u}(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)}} = \int \sqrt{f_{\theta+u}(x) f_\theta(x)}\,\mu(dx) = 1 - \frac{r(u)}{2}, \tag{2}$$

$$\mathbf{E}_\theta \mathbf{Z}^{1/2}(u) = \left(1 - \frac{r(u)}{2}\right)^n. \tag{3}$$

Относительно параметрического семейства $\{\mathbf{P}_\theta\}$ будем предполагать в этом и последующих параграфах, что *наряду с* (A_μ) *выполнены условия* (A_0) $(f_{\theta_1}(x) \neq f_{\theta_2}(x)$ *при* $\theta_1 \neq \theta_2)$ *и* (A_c) (Θ *есть компакт*). Несущественность с точки зрения приложений последнего условия мы уже отмечали. Это связано с тем, что в реальных задачах из априорных соображений обычно бывает возможно указать границы возможных значений θ. Для простоты там, где это потребуется, будем предполагать также, что Θ выпукло (в одномерном случае это означает, что $\Theta = [a, b]$, $-\infty < a < b < \infty$).

Кроме того, в этом параграфе будем предполагать, что функция $\sqrt{f_\theta}$ *дифференцируема для п. в.* $[\mu]$ *значений* x, *а информация Фишера*

$$I(\theta) = \int \frac{\left(f'_\theta(x)\right)^2}{f_\theta(x)}\mu(dx) = \mathbf{E}_\theta\left(\frac{f'_\theta(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)}\right)^2$$

строго положительна и ограничена в Θ (это есть условие (R) без свойства непрерывности). При этих условиях в теореме 31.3 было доказано, что при всех допустимых θ и $\theta + u$ (т. е. таких, что $\theta \in \Theta$, $\theta + u \in \Theta$) для величины $r(u) = \rho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta)$ справедливо неравенство

$$\inf_{\theta,u} \frac{r(u)}{u^2} \geqslant g > 0. \tag{4}$$

1. Основные неравенства. Обозначим для краткости $p(u) = Z^{3/4}(u)$ и будем считать, что все перечисленные выше условия выполнены.

Теорема 1. *Справедливы неравенства*

$$\begin{gathered}\mathbf{E}_\theta Z^{1/2}(u) \leqslant e^{-ngu^2/2}, \quad \mathbf{E}_\theta p(u) \leqslant e^{-ngu^2/4}, \\ \mathbf{E}_\theta|p'(u)| \leqslant \frac{3}{4}\sqrt{nI(\theta+u)}\, e^{-ngu^2/4}.\end{gathered} \tag{5}$$

Из рассмотрений § 31 следует, что для значений $u = o(1)$ вместо g в этих неравенствах можно брать значения, сколь угодно близкие к $I(\theta)/4$.

Доказательство. Имеем в силу (3), (4)

$$\mathbf{E}_\theta Z^{1/2}(u) = (1 - r(u)/2)^n \leqslant e^{-nr(u)/2} \leqslant e^{-ngu^2/2}.$$

Далее, в силу неравенства Коши–Буняковского

$$\mathbf{E}_\theta p(u) \leqslant [\mathbf{E}_\theta Z^{1/2}(u) \cdot \mathbf{E}_\theta Z(u)]^{1/2} = [\mathbf{E}_\theta Z^{1/2}(u)]^{1/2} \leqslant e^{-u^2 ng/4}.$$

Используя снова неравенство Коши–Буняковского и соотношение

$$p'(u) = \frac{3}{4}L'(X, \theta+u)Z^{3/4}(u),$$

находим

$$\begin{aligned}\mathbf{E}_\theta|p'(u)| &= \frac{3}{4}\mathbf{E}_\theta|L'(X,\theta+u)|Z^{1/2}(u)Z^{1/4}(u) \leqslant \\ &\leqslant \frac{3}{4}[\mathbf{E}_\theta[L'(X,\theta+u)]^2 Z(u) \cdot \mathbf{E}_\theta Z^{1/2}(u)]^{1/2} \leqslant \\ &\leqslant \frac{3}{4}[\mathbf{E}_{\theta+u}[L'(X,\theta+u)]^2]^{1/2} e^{-u^2 ng/4}. \quad \triangleleft\end{aligned}$$

Теорема 2. *При всех* z, $n \geqslant 1$

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant u} Z\left(\frac{v}{\sqrt{n}}\right) > e^z\right) \leqslant ce^{-3z/4}e^{-u^2 g/4},$$

где $c = 1 + 3\sqrt{I_0/g}$, $I_0 = \sup\limits_{\theta\in\Theta} I(\theta)$ *от* θ *не зависят.*

Для доказательства теоремы нам понадобится

Лемма 1. *При всех* $x \geqslant 0$

$$\int_x^\infty e^{-v^2/2}\,dv \leqslant \sqrt{2\pi}\,e^{-x^2/2}.$$

Доказательство*). Характеристическая функция случайной величины $\xi \in \Phi_{0,1}$ равна $\mathbf{E}e^{it\xi} = e^{-t^2/2}$ и определена во всей плоскости. Полагая $t = -ix$, получим $\mathbf{E}e^{x\xi} = e^{x^2/2}$. Отсюда с помощью неравенства Чебышева получаем

$$\mathbf{P}(\xi > x) = \mathbf{P}(e^{\xi x} > e^{x^2}) \leqslant e^{-x^2}\mathbf{E}e^{\xi x} = e^{-x^2/2}. \quad \triangleleft$$

Доказательство теоремы 2. Оценим функцию

$$H(\delta) = \mathbf{E}_\theta \sup_{|v|>\delta} p(v).$$

Если $w \in [\theta + \delta, b]$, то

$$p(w-\theta) = p(\delta) + \int_\delta^{w-\theta} p'(u)\,du \leqslant p(\delta) + \int_\delta^{b-\theta} |p'(u)|\,du.$$

Так как правая часть здесь от w не зависит, то

$$\sup_{u\geqslant\delta} p(u) \leqslant p(\delta) + \int_{u\geqslant\delta} |p'(u)|\,du,$$

$$H_+(\delta) \equiv \mathbf{E}_\theta \sup_{u\geqslant\delta} p(u) \leqslant \mathbf{E}_\theta p(\delta) + \int_{u\geqslant\delta} \mathbf{E}_\theta |p'(u)|\,du.$$

*) При больших x более точными являются следующие неравенства:

$$\frac{1}{x+1}e^{-x^2/2} < \int_x^\infty e^{-v^2/2}\,dv < \frac{1}{x}e^{-x^2/2},$$

которые читатель без труда может получить, сравнив производные рассматриваемых функций (значения самих функций при $x = \infty$ совпадают).

В силу теоремы 1 отсюда получаем

$$H_+(\delta) \leqslant e^{-ng\delta^2/4} + \frac{3}{4}\sqrt{n} \int\limits_{u \geqslant \delta} \sqrt{I(\theta+u)}\, e^{-ngu^2/4}\, du.$$

На основании леммы 1

$$H_+(\delta) \leqslant e^{-ng\delta^2/4} + \frac{3}{4}\sqrt{nI_0} \int\limits_{|u| \geqslant \delta} e^{-ngu^2/4}\, du \leqslant$$

$$\leqslant e^{-ng\delta^2/4} + \frac{3}{4}\sqrt{2I_0/g} \int\limits_{v \geqslant \delta\sqrt{ng/2}} e^{-v^2/2}\, dv \leqslant e^{-ng\delta^2/4}(1 + 3\sqrt{I_0/g}).$$

Ясно, что точно такая же оценка будет справедлива для функции

$$H_-(\delta) = \sup_{u \leqslant -\delta} p(u).$$

Поэтому

$$H(\delta) \leqslant \max\left(H_+(\delta), H_-(\delta)\right) \leqslant \left(1 + 3\sqrt{\frac{I_0}{g}}\right) e^{-ng\delta^2/4}.$$

Остается воспользоваться неравенством Чебышева

$$\mathbf{P}_\theta(\sup_{|t| \geqslant \delta} Z(t) > e^z) = \mathbf{P}_\theta(\sup_{|t| \geqslant \delta} p(t) > e^{3z/4}) \leqslant H(\delta) e^{-3z/4}. \quad \triangleleft$$

2. Оценки для распределения и моментов о. м. п. Состоятельность о. м. п.

Теорема 3. *Существуют значения $c < \infty$, $g > 0$ такие, что*

$$\mathbf{P}_\theta(\sqrt{n}\,|\widehat{\theta}^* - \theta| \geqslant v) \leqslant ce^{-gv^2/4} \tag{6}$$

для всех v и $n \geqslant 1$.

Доказательство. Из теоремы 2 вытекает, что

$$\mathbf{P}_\theta(\sup_{|t| \geqslant v/\sqrt{n}} Z(t) > 1) \leqslant ce^{-gv^2/4}.$$

Остается воспользоваться соотношением

$$\{|\widehat{\theta}^* - \theta| \geqslant \delta\} = \{\sup_{|t| \geqslant \delta} Z(t) \geqslant \sup_{|t| \leqslant \delta} Z(t)\} \subset \{\sup_{|t| \geqslant \delta} Z(t) \geqslant Z(0) = 1\} \tag{7}$$

при $\delta = v/\sqrt{n}$. $\triangleleft$

Следствие 1. *Пусть $u_n \to \infty$ — любая неограниченно возрастающая последовательность. Тогда*

$$(\widehat{\theta}^* - \theta)\frac{\sqrt{n}}{u_n} \xrightarrow[P]{} 0. \tag{8}$$

Если же u_n *таковы, что при любом* $\alpha > 0$

$$\sum e^{-\alpha u_n^2} < \infty, \tag{9}$$

то

$$(\widehat{\theta}^* - \theta)\frac{\sqrt{n}}{u_n} \underset{\text{п.н.}}{\longrightarrow} 0. \tag{10}$$

Эти соотношения являются, очевидно, усилениями соответственно состоятельности ($\widehat{\theta}^* - \theta \underset{P}{\longrightarrow} 0$) и сильной состоятельности ($\widehat{\theta}^* - \theta \underset{\text{п.н.}}{\longrightarrow} 0$) о. м. п.

Доказательство. Соотношение (8) следует непосредственно из (6), если в последнем положить $v = \delta u_n$. Соотношение (10) также вытекает из (6), так как сумма правых частей в (6) при выполнении (9) будет образовывать сходящийся ряд. ◁

Например, даже такая медленно растущая последовательность, как $u_n = \ln n$, удовлетворяет условию (9), так что*)

$$(\widehat{\theta}^* - \theta)\frac{\sqrt{n}}{\ln n} \underset{\text{п.н.}}{\longrightarrow} 0.$$

Следствие 2. *Существует значение* $c_1 < \infty$, *не зависящее от* n, θ, *такое, что при любом* $\alpha \leqslant g/5$

$$\mathbf{E}_\theta \exp\{\alpha(u^*)^2\} < c_1, \tag{11}$$

где $u^* = \sqrt{n}\,(\widehat{\theta}^* - \theta)$.

Доказательство. Интегрируя по частям, получаем

$$\mathbf{E}e^{\alpha\xi^2} = -\int_0^\infty e^{\alpha v^2}\, d\mathbf{P}(|\xi| \geqslant v) = 1 + 2\alpha \int_0^\infty v e^{\alpha v^2} \mathbf{P}(|\xi| \geqslant v)\, dv.$$

Поэтому в силу теоремы 3

$$\mathbf{E}_\theta e^{\alpha(u^*)^2} \leqslant 1 + \frac{2gc}{5}\int_0^\infty v e^{-gv^2/20}\, dv \equiv c_1 < \infty. \quad \triangleleft$$

3. Асимптотическая нормальность. Доказательство асимптотической нормальности о. м. п., как мы видели в § 14, 16, требует более сильных, чем (R), условий регулярности. Обозначим эти условия символом (RR), чтобы отметить таким способом, что это есть условия регулярности и что они усиливают условия (R).

Условия (RR).

1. *Выполнены условия* (A_0), (A_c), (R).

*) Из замечания 34.2 будет следовать, что (10) справедливо и для еще более медленно растущих u_n.

2. *Функция $l(x,\theta)$ для п. в. $[\mu]$ значений x дважды непрерывно дифференцируема по θ. Функция $|l''(x,t)|$ мажорируется функцией $l(x)$, не зависящей от t: $|l''(x,t)| < l(x)$, для которой*

$$\mathbf{E}_t l(\mathrm{x}_1) = \int l(x) f_t(x) \mu(dx) < \infty. \tag{12}$$

3. *Законно дифференцирование по θ под знаком интеграла в равенствах*

$$\int f_\theta(x)\mu(dx) = 1, \quad \int l'(x,\theta) f_\theta(x)\mu(dx) = 0.$$

Свойство 3 вместе с равенствами (аргументы для краткости опускаем)

$$l' = \frac{f'}{f}, \quad l'' = \frac{f''}{f} - \left(\frac{f'}{f}\right)^2$$

означает, что

$$\mathbf{E}_\theta l'(\mathrm{x}_1,\theta) = 0, \quad \mathbf{E}_\theta l''(\mathrm{x}_1,\theta) = -\mathbf{E}_\theta (l'(\mathrm{x}_1,\theta))^2 = -I(\theta).$$

Условие о единственности решения $t = \theta$ уравнения (относительно t)

$$\int l'(x,t) f_\theta(x)\mu(dx) = 0$$

здесь не требуется, так как состоятельность в силу следствия 1 уже установлена (см. условие d) в теореме 16.3).

Как увидим ниже, условие 3 является излишним, если потребовать, чтобы интеграл (12) сходился равномерно (см. § 34 и приложение V).

Условия (RR), по-видимому, завышены для последующих утверждений; см. об этом ниже замечание 34.3. Напомним основной результат п. 16.3.

Теорема 4. *При выполнении условий* (RR)

$$(\widehat{\theta}^* - \theta)\sqrt{n} = \xi_n I^{-1}(\theta)(1 + \varepsilon_n(X,\theta)),$$

где

$$\xi_n = \frac{L'(X,\theta)}{\sqrt{n}} \Subset \mathbf{\Phi}_{0,I(\theta)}, \quad \varepsilon_n(X,\theta) \underset{P}{\longrightarrow} 0.$$

Это утверждение можно дополнить следующим образом.

Теорема 5. *Сходимость*

$$u^* \equiv (\widehat{\theta}^* - \theta)\sqrt{n} \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)} \tag{13}$$

имеет место вместе с моментами любого порядка, т. е. вместе с (13) *при любом $k > 0$ выполняется*

$$\mathbf{E}_\theta (u^*)^k \to \mathbf{E}\eta^k, \quad \eta \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)}. \tag{14}$$

Более того, для любой непрерывной функции $w(t)$ такой, что $|w(t)| < e^{gt^2/6}$ (см. 4)),

$$\mathbf{E}_\theta w(u^*) \to \mathbf{E}w(\eta), \quad \eta \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)}. \tag{15}$$

Доказательство. Соотношения (14), (15) получаются из (13) и теоремы непрерывности для моментов (см. § 5), так как в силу следствия 2

$$\mathbf{E}_\theta w^{6/5}(u^*) \leqslant \mathbf{E}_\theta \exp\left\{\frac{(u^*)^2 g}{5}\right\} < c < \infty. \quad \triangleleft$$

Замечание 1. Из (13), (14) вытекает, что $\widehat{\theta}^*$ принадлежит классу оценок $K_{\mathbf{\Phi},2}$, в котором сходимость $(\widehat{\theta}^* - \theta)\sqrt{n} \Subset \mathbf{\Phi}_{0,\sigma^2(\theta)}$ имеет место вместе со сходимостью $\mathbf{E}_\theta(\theta^* - \theta)^2 \to \sigma^2(\theta)$ двух первых моментов. Как уже отмечалось в § 18, в этом классе асимптотический подход к сравнению оценок совпадает по существу со среднеквадратическим.

Установим теперь один из центральных результатов этой главы, состоящий в том, что о. м. п. при выполнении условий (RR) обладает всеми возможными свойствами асимптотической оптимальности, которые мы рассматривали выше, т. е. является а. э., асимптотически байесовской (для любого априорного распределения, имеющего плотность) и асимптотически минимаксной оценкой одновременно.

Асимптотическая эффективность M-оценок в классе K_M оценок была доказана в § 16.

4. Асимптотическая эффективность. В § 26 мы ввели в рассмотрение класс $\widetilde{K}_0$ асимптотически несмещенных оценок, т. е. оценок θ^*, смещение которых $b(\theta) = \mathbf{E}_\theta\theta^* - \theta$ обладает свойствами

$$b(\theta) = o\left(\frac{1}{\sqrt{n}}\right), \quad b'(\theta) = o(1). \tag{16}$$

В § 30 были приведены соображения, по которым в поисках асимптотически эффективных «в целом» оценок можно ограничиться классом $\widetilde{K}_0$.

Установим теперь следующий факт.

Следствие 3. *Справедливо* $\widehat{\theta}^* \in \widetilde{K}_0$.

Доказательство. Первое из соотношений (16) вытекает из (14) при $k = 1$. Чтобы доказать второе, заметим, что (см. § 26)

$$1 + b'(\theta) = \mathbf{E}_\theta\widehat{\theta}^* L'(X,\theta) = \mathbf{E}_\theta(\widehat{\theta}^* - \theta)L'(X,\theta) =$$

$$= \mathbf{E}_\theta((\widehat{\theta}^* - \theta)\sqrt{n}\,\xi_n) = \mathbf{E}_\theta\frac{\xi_n^2}{I(\theta)}(1 + \varepsilon_n(X,\theta)),$$

$$\varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0.$$

Если здесь верна теорема непрерывности для моментов, получим отсюда требуемое соотношение $1 + b'(\theta) \to 1$ или, что то же, $b'(\theta) \to 0$. Чтобы установить справедливость этой теоремы в нашем случае, достаточно убедиться (см. § 5), что

$$\mathbf{E}_\theta|(\widehat{\theta}^* - \theta)\sqrt{n}\,\xi_n|^{3/2} < c < \infty, \tag{17}$$

где c от n не зависит. Воспользуемся неравенством Гёльдера

$$\mathbf{E}|\xi\eta|^r \leqslant (\mathbf{E}|\xi|^{pr})^{1/p}(\mathbf{E}|\eta|^{qr})^{1/q}, \quad p>0, \quad q>0, \quad \frac{1}{p}+\frac{1}{q}=1,$$

при $r=3/2$, $p=4$, $q=4/3$. Получим тогда для левой части (17) оценку $(\mathbf{E}_\theta[(\widehat{\theta}^*-\theta)\sqrt{n}\,]^6)^{1/4}(\mathbf{E}\xi_n^2)^{3/4}$, которая в силу (13) дает нам требуемое неравенство. $\triangleleft$

Следующее следствие ввиду его важности сформулируем в виде теоремы.

Теорема 6. *О. м. п.* $\widehat{\theta}^*$ *является а. R-э. оценкой. Кроме того,* $\widehat{\theta}^*$ *является а. э. в* $\widetilde{K}_0$.

Доказательство. Тот факт, что $\widehat{\theta}^*$ есть а. R-э. оценка, следует непосредственно из определения 26.1 и того, что

$$\mathbf{E}_\theta(\widehat{\theta}^*-\theta)^2=\frac{1+o(1)}{nI(\theta)}.$$

Асимптотическая эффективность в $\widetilde{K}_0$ следует из теоремы 26.3. $\triangleleft$

Теорема 2 вместе с замечаниями к теореме 26.3 означает, что при выполнении условий (RR) любая а. э. оценка в $\widetilde{K}_0$ будет а. R-э. оценкой.

Отметим, что сужение множества рассматриваемых оценок до $\widetilde{K}_0$ — далеко не единственное сужение, при котором $\widehat{\theta}^*$ становится а. э.

Укажем другое сужение, связанное на этот раз со свойством θ быть асимптотической медианой распределения а. н. оценок, т. е. со свойством

$$\mathbf{P}_\theta(\widehat{\theta}^*>\theta)\to\frac{1}{2} \tag{18}$$

при $n\to\infty$.

Обозначим через $\widetilde{K}^0$ класс оценок θ^*, для которых (18) выполняется равномерно по θ. Класс $\widetilde{K}^0$ можно было бы назвать классом *асимптотически центральных оценок.*

Теорема 7. *О. м. п.* $\widehat{\theta}^*\in\widetilde{K}^0$ *и является а. э. оценкой в классе* $\widetilde{K}^0$.

Доказательство этой теоремы мы отложим до п. 37.3 и 43.5.

5. Асимптотическая байесовость о. м. п. В этом пункте всюду, где будет предполагаться существование плотности $q(t)$ априорного распределения $\mathbf{Q}$ относительно меры Лебега на Θ, будем предполагать также, не оговаривая это особо, что плотность интегрируема по Риману, так что будут удовлетворяться условия теоремы 30.5.

Теорема 8. *О. м. п.* $\widehat{\theta}^*$ *является асимптотически R-байесовской оценкой. Если* $\mathbf{Q}$ *есть произвольное априорное распределение, имеющее плотность* $q(t)$ *относительно меры Лебега, то* $\widehat{\theta}^*$ *является также и асимптотически байесовской оценкой, соответствующей распределению* $\mathbf{Q}$.

Доказательство. Асимптотическая R-байесовость о. м. п. следует из соотношений

$$\lim_{n\to\infty} \mathbf{E}[\sqrt{n}\,(\widehat{\theta}^* - \theta)]^2 = \lim_{n\to\infty} \mathbf{E}\mathbf{E}_\theta[\sqrt{n}\,(\widehat{\theta}^* - \theta)]^2 =$$

$$= \mathbf{E} \lim_{n\to\infty} \mathbf{E}_\theta[\sqrt{n}\,(\widehat{\theta}^* - \theta)]^2 = \mathbf{E}I^{-1}(\theta) = J.$$

Предельный переход под знаком математического ожидания здесь законен по теореме о мажорируемой сходимости, так как в силу следствия 2 значение $\mathbf{E}_\theta[\sqrt{n}\,(\widehat{\theta}^* - \theta)]^2$ равномерно ограничено постоянной, не зависящей от n и θ.

Асимптотическая байесовость вытекает из следствия 30.1. $\triangleleft$

Из замечаний к следствию 30.1 и теоремы 8 следует, что любая асимптотически байесовская оценка является асимптотически R-байесовской.

Утверждение теоремы 8 можно усилить. Оказывается, что о. м. п. и байесовская оценка для любой априорной плотности q «почти» совпадают.

Теорема 9. *Справедливы соотношения*

$$\mathbf{E}n\big(\widehat{\theta}^* - \theta_Q^*\big)^2 \to 0, \quad \big(\theta_Q^* - \widehat{\theta}^*\big)\sqrt{n} \underset{P}{\longrightarrow} 0,$$

где θ_Q^* есть байесовская оценка, соответствующая распределению $\mathbf{Q}$; *сходимость по вероятности понимается относительно совместного распределения X и θ на $\mathcal{X} \times \Theta$.*

Теорема 9 немедленно вытекает из следствия 30.2. Ее утверждение эквивалентно тому, что для почти всех t

$$\mathbf{E}_t n\big(\widehat{\theta}^* - \theta_Q^*\big)^2 \to 0.$$

§ 34. Асимптотические свойства отношения правдоподобия. Дальнейшие свойства оптимальности о. м. п.

В предыдущем параграфе нами были изучены свойства отношения правдоподобия $Z(u)$ вне окрестности точки $u = 0$. Теперь рассмотрим поведение $Z(u)$ в окрестности этой точки. Как и в п. 3 предыдущего параграфа, будем предполагать, что выполнены условия (RR). Из полученных ниже результатов будет следовать, в частности, уже доказанная нами асимптотическая нормальность о. м. п. $\widehat{\theta}^*$. Мы приведем доказательство этого факта в таком виде, который позволяет без труда переносить его на многомерный случай.

Некоторые из последующих утверждений повторяют отчасти рассуждения § 14. Нам понадобятся соотношения

$$\mathbf{E}_\theta l'(\mathrm{x}_1, \theta) = 0, \tag{1}$$

$$\mathbf{E}_\theta l''(\mathrm{x}_1, \theta) = -\mathbf{E}_\theta(l'(\mathrm{x}_1, \theta))^2 = -I(\theta), \tag{2}$$

вытекающие из условий (RR), и следующая

Лемма 1. *Если выполнены условия* (RR), *то имеет место непрерывность* $l''(x,\theta)$ *«в среднем» в следующем смысле*:

$$\mathbf{E}_\theta \omega''_\Delta(\mathrm{x}_1) = \int \omega''_\Delta(x) f_\theta(x)\mu(dx) \to 0 \tag{3}$$

при $\Delta \to 0$, *где* $\omega''_\Delta(x)$ — *модуль непрерывности функции* $l''(x,\theta)$:

$$\omega''_\Delta(x) = \sup_{\substack{\theta\in\Theta,\ \theta+u\in\Theta \\ |u|\leqslant\Delta}} |l''(x,\theta+u) - l''(x,\theta)|. \tag{4}$$

Доказательство. В силу теоремы о мажорируемой сходимости соотношение (3) будет следствием обычной непрерывности l'', так как в этом случае $\omega''_\Delta(x) \to 0$ для п. в. $[\mu]$ значений x при $\Delta \to 0$ и, кроме того, $\left|\omega''_\Delta(x)\right| \leqslant 2l(x)$. $\triangleleft$

Положим

$$\gamma_n(\Delta,\theta) = \sup_{|v|\leqslant\Delta} \left| \frac{L'(X,\theta+v) - L'(X,\theta)}{nv} + I(\theta) \right|.$$

Лемма 2. *Пусть выполнены условия* (RR), $\delta_n > 0$, $n = 1,2,\ldots$, *есть любая последовательность, сходящаяся к нулю. Тогда для любого* $\theta \in \Theta$ *и для* $X \in \mathbf{P}_\theta$

$$\gamma_n(\delta_n,\theta) \underset{\text{п.н.}}{\longrightarrow} 0, \quad \gamma_n(\delta_n,\widehat{\theta}^*) \underset{\text{п.н.}}{\longrightarrow} 0.$$

В этих соотношениях $I(\theta)$ *можно заменить на* $I(\widehat{\theta}^*)$ *и наоборот.*

Доказательство. Докажем сначала первое утверждение. Так как $\mathbf{E}_\theta l''(\mathrm{x}_1,\theta) = -I(\theta)$, $L''(X,\theta)/n \underset{\text{п.н.}}{\longrightarrow} -I(\theta)$, то нам достаточно убедиться, что $\gamma_n(\delta_n) \to 0$, где

$$\gamma_n(\Delta) = \sup_{|v|\leqslant\Delta} \left| \frac{L'(X,\theta+v) - L'(X,\theta)}{nv} - \frac{L''(X,\theta)}{n} \right|.$$

Но

$$\gamma_n(\delta_n) \leqslant \sup_{|v|\leqslant\delta_n} \frac{1}{n}|L''(X,\theta+v) - L''(X,\theta)| \leqslant \frac{1}{n}\sum_{i=1}^n \omega''_{\delta_n}(\mathrm{x}_i) \equiv \overline{\omega}''_{\delta_n}(X),$$

где $\omega''_\Delta(x)$ означает модуль непрерывности $l''(x,\theta)$, определенный в (4). Очевидно, что для любого фиксированного $\Delta > 0$ при достаточно больших n

$$\overline{\omega}''_{\delta_n}(X) \leqslant \overline{\omega}''_\Delta(X).$$

Кроме того, по усиленному закону больших чисел

$$\overline{\omega}''_\Delta(X) \underset{\text{п.н.}}{\longrightarrow} \mathbf{E}_\theta \omega''_\Delta(\mathrm{x}_1) \equiv \omega''_\Delta.$$

В силу леммы 1 $\omega''_\Delta \to 0$ при $\Delta \to 0$. Отсюда следует, что

$$\overline{\omega}''_{\delta_n}(X) \underset{\text{п.н.}}{\longrightarrow} 0. \tag{5}$$

Первое утверждение доказано. Из (5) и определения п. н. сходимости следует, что наряду с (5)

$$\overline{\omega}''_{\delta_n+\eta_n}(X) \xrightarrow[\text{п.н.}]{} 0$$

для любой последовательности случайных величин $\eta_n \xrightarrow[\text{п.н.}]{} 0$. Нам остается заметить, что

$$\sup_{|v|<\delta_n} \left| \frac{L'(X,\widehat{\theta}^*+v) - L'(X,\widehat{\theta}^*)}{nv} - \frac{L''(X,\theta)}{n} \right| \leqslant \overline{\omega}''_{\delta_n+|\widehat{\theta}^*-\theta|}(X). \tag{6}$$

и воспользоваться следствием 33.1. Возможность замены $I(\theta)$ на $I(\widehat{\theta}^*)$ также вытекает из следствия 33.1 (и непрерывности $I(\theta)$). $\lhd$

Мы можем теперь сформулировать основные утверждения об асимптотическом поведении отношения правдоподобия $Z(t)$. Положим

$$Y(u) = \ln Z\left(\frac{u}{\sqrt{n}}\right) = L\left(X, \theta + \frac{u}{\sqrt{n}}\right) - L(X,\theta)$$

и условимся обозначать через $\varepsilon_n(X,\theta)$ (иногда с дополнительными индексами) различные последовательности случайных величин, п. н. сходящиеся к нулю относительно распределения $\mathbf{P}_\theta$ выборки X_∞.

Теорема 1. *Пусть выполнены условия* (RR), $\delta_n > 0$ *есть произвольная последовательность, сходящаяся к нулю. Тогда при* $|u/\sqrt{n}| < \delta_n$

$$Y(u) = u\xi_n - \frac{u^2}{2} I(\theta)(1 + \varepsilon_n(X,\theta,u)), \tag{7}$$

где

$$|\varepsilon_n(X,\theta,u)| \leqslant \varepsilon_n(X,\theta) \xrightarrow[\text{п.н.}]{} 0, \quad \xi_n = \frac{1}{\sqrt{n}} L'(X,\theta) \Rightarrow \mathbf{\Phi}_{0,I(\theta)}.$$

Точка $u^* = (\widehat{\theta}^* - \theta)\sqrt{n}$, *в которой* $Y(u)$ *достигает максимального значения, обладает свойствами*

$$u^* = \frac{\xi_n}{I(\theta)}(1 + \varepsilon_n(X,\theta)), \tag{8}$$

$$2Y(u^*) = 2\ln Z(\widehat{\theta}^* - \theta) = \frac{\xi_n^2}{I(\theta)}(1 + \varepsilon_n(X,\theta)) \Rightarrow \mathbf{H}_1. \tag{9}$$

Наряду с (7) *справедливо представление*

$$\begin{gathered} Y(u) = Y(u^*) - \frac{(u-u^*)^2}{2} I(\theta)(1 + \varepsilon_n(X,\theta,u)), \\ |\varepsilon_n(X,\theta,u)| < \varepsilon_n(X,\theta). \end{gathered} \tag{10}$$

Во всех приведенных утверждениях $I(\theta)$ *можно заменить на* $I(\widehat{\theta}^*)$.

В этой теореме, как и в лемме 2, предполагается, что $\theta + u\sqrt{n} \in \Theta$. Это соотношение будет автоматически выполнено при достаточно больших n, если θ — внутренняя точка Θ.

Замечание 1. Важно отметить, что в (7) случайные величины ξ_n и $\varepsilon_n(X, \theta)$ от u не зависят. Поэтому первое утверждение теоремы можно записать в виде

$$\sup_{|u| \leqslant \delta_n \sqrt{n}} \left| \frac{Y(u) - u\xi_n + \frac{u^2}{2} I(\theta)}{u^2} \right| \underset{\text{п.н.}}{\longrightarrow} 0.$$

Если δ_n таково, что

$$\sum e^{-ng\delta_n^2/4} < \infty, \tag{11}$$

из теоремы 33.2 следует, что в дополнительной области $|u| > \delta_n \sqrt{n}$

$$\sup_{|u| > \delta_n \sqrt{n}} Y(u) \underset{\text{п.н.}}{\longrightarrow} -\infty.$$

Доказательство теоремы 1. Из леммы 2 при $|v| \leqslant \delta_n$ получаем

$$L'(X, \theta + v) = L'(X, \theta) - nvI(\theta)(1 + \varepsilon_n(X, \theta, v)),$$

$$|\varepsilon_n(X, \theta, v)| \leqslant \varepsilon_n(X, \theta).$$

Интегрируя это равенство по v в пределах от нуля до $u/\sqrt{n}$, находим

$$L\left(X, \theta + \frac{u}{\sqrt{n}}\right) - L(X, \theta) = \frac{1}{\sqrt{n}} uL'(X, \theta) - \frac{u^2}{2} I(\theta)(1 + \varepsilon_n(X, \theta, u)), \tag{12}$$

$$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n(X, \theta).$$

Это есть, очевидно, разложение в ряд Тейлора, в котором $L''(X, \theta)/n$ заменено на $-I(\theta)$, а остаточный член допускает равномерную оценку. Так как

$$\xi_n \equiv \frac{1}{\sqrt{n}} L'(X, \theta) = \frac{1}{\sqrt{n}} \sum l'(\mathrm{x}_i, \theta)$$

есть сумма независимых одинаково распределенных случайных величин с нулевым средним и дисперсией $I(\theta)$ (см. (2)), то по центральной предельной теореме $\xi_n \Rightarrow \Phi_{0, I(\theta)}$. Представление (7) доказано. Чтобы доказать (8), вернемся к лемме 2. Она означает, что существует множество A, $\mathbf{P}_\theta(A) = 1$, такое, что для $X_\infty \in A$, $n \to \infty$

$$\sup_{|v| < \delta_n} \left| \frac{L'(X, \theta + v) - L'(X, \theta)}{nv} + I(\theta) \right| \to 0. \tag{13}$$

Кроме того, в силу следствия 33.1 существуют последовательность $u_n \to \infty$, $u_n/\sqrt{n} \equiv \gamma_n \to 0$ (u_n должна удовлетворять (33.9)) и множество B, $\mathbf{P}_\theta(B) = 1$, такие, что для $X_\infty \in B$, $n \to \infty$

$$v^* = (\widehat{\theta}^* - \theta) = o(\gamma_n). \tag{14}$$

Так как последовательность $\delta_n \to 0$ в (13) произвольна, то при $X_\infty \in A \cap B$, $\mathbf{P}_\theta(A \cap B) = 1$, в силу (14) величина под знаком sup в (13) в точке $v = v^*$ будет сходиться к нулю. Вспоминая, что $L'(X, \theta + v^*) = L'(X, \widehat{\theta}^*) = 0$, получим для $X_\infty \in A \cap B$

$$\left| I(\theta) - \frac{L'(X,\theta)}{n(\widehat{\theta}^* - \theta)} \right| \to 0.$$

Это означает, что $\xi_n - I(\theta)u^* = u^*\varepsilon_n(X,\theta)$, и доказывает (8).

Пользуясь теми же аргументами, можно подставить

$$u = u^* = v^*\sqrt{n} = (\widehat{\theta}^* - \theta)\sqrt{n} = \frac{\xi_n}{I(\theta)}(1 + \varepsilon_n(X,\theta))$$

в (12). Это дает

$$L(X,\widehat{\theta}^*) - L(X,\theta) = \frac{\xi_n^2}{I(\theta)}(1 + \varepsilon_n(X,\theta))$$

и доказывает первую часть соотношения (9). Сходимость распределений $\xi_n^2/I(\theta)$ к распределению χ^2 с одной степенью свободы следует из теорем непрерывности, так как $\xi_n/\sqrt{I(\theta)} \Rightarrow \mathbf{\Phi}_{0,1}$.

Соотношение (10) доказывается совершенно аналогично (7), если воспользоваться вторым утверждением леммы 2 и найти на его основе представление для $L(X, \theta + u\sqrt{n}) - L(X, \widehat{\theta}^*)$. $\triangleleft$

Замечание 2. Соотношение (8) позволяет также точно описать «максимальные уклонения» $(\widehat{\theta}^* - \theta)\sqrt{n}$ при $n \to \infty$. Именно, известно, (см. [17, 72, 137]), что нормированные суммы ξ_n независимых одинаково распределенных величин с нулевым средним и дисперсией $I(\theta)$ удовлетворяют закону повторного логарифма в силу которого

$$\mathbf{P}\left(\limsup_{n\to\infty} \frac{|\xi_n|}{\sqrt{2I(\theta)\ln\ln n}} = 1 \right) = 1.$$

Так как в (4) $\limsup\limits_{n\to\infty} \varepsilon_n(X,\theta) = 0$ п. н., то получаем, что

$$\mathbf{P}_\theta\left(\limsup_{n\to\infty} \frac{|\widehat{\theta}^* - \theta|\sqrt{nI(\theta)}}{\sqrt{2\ln\ln n}} = 1 \right) = 1.$$

Замечание 3. На языке распределений первое утверждение теоремы 1 можно сформулировать следующим образом:

$$Y(u) \Rightarrow \mathbf{\Phi}_{-u^2 I(\theta)/2,\, u^2 I(\theta)}. \tag{15}$$

Мы отмечали выше, что второе условие (RR) (о существовании $l''(x,\theta)$) не всегда существенно для доказываемых утверждений. В несущественности этого условия для сходимости (15) можно убедиться путем следующих рассуждений. Величина

$$Y(u) = L\left(X, \theta + \frac{u}{\sqrt{n}}\right) - L(X,\theta) = \sum_{i=1}^{n} \left[l\left(\mathrm{x}_i, \theta + \frac{u}{\sqrt{n}}\right) - l(\mathrm{x}_i,\theta) \right]$$

есть сумма независимых одинаково распределенных величин. Поэтому по центральной предельной теореме для схемы серий (слагаемые зависят от n, проверку условий Линдеберга опускаем)

$$Y(u) \Rightarrow \Phi_{\alpha(u),\sigma^2(u)},$$

где

$$\alpha(u) = \lim_{n\to\infty} n\mathbf{E}_\theta \left[l\left(\mathrm{x}_1, \theta + \frac{u}{\sqrt{n}}\right) - l(\mathrm{x}_1, \theta)\right] =$$

$$= \lim_{n\to\infty} n\mathbf{E}_\theta \ln \frac{f_{\theta+u/\sqrt{n}}(\mathrm{x}_1)}{f_\theta(\mathrm{x}_1)} = -u^2 \lim_{\Delta\to 0} \frac{\rho_1(\mathbf{P}_\theta, \mathbf{P}_{\theta+\Delta})}{\Delta^2} = -\frac{u^2 I(\theta)}{2}$$

(см. теорему 31.2 и замечание 31.1). Далее

$$\sigma^2(u) = \lim_{n\to\infty} n\mathbf{E}_\theta \left[l\left(\mathrm{x}_1, \theta + \frac{u}{\sqrt{n}}\right) - l(\mathrm{x}_1, \theta)\right]^2 =$$

$$= u^2 \lim_{\Delta\to 0} \int \left[\frac{l(x, \theta+\Delta) - l(x,\theta)}{\Delta}\right]^2 f_\theta(x)\mu(dx) =$$

$$= u^2 \int (l'(x,\theta))^2 f_\theta(x)\mu(dx) = u^2 I(\theta).$$

Если бы при вычислении $\alpha(u)$ и $\sigma^2(u)$ мы воспользовались разложением $l(x, \theta + u/\sqrt{n})$ в ряд с двумя производными, то получили бы тот же результат. Однако мы убедились, что это делать не обязательно.

В заключение этого параграфа получим из теоремы 1 еще одно полезное следствие, которое потребуется в дальнейшем и которое касается поведения интегралов от отношения правдоподобия.

Теорема 2. *Пусть выполнены условия* (RR), *функция* $w(t)$ *удовлетворяет условию*

$$|w(t)| \leqslant ce^{\alpha|t|^2}, \quad c < \infty, \quad \alpha = \frac{g}{10}$$

($g > 0$ определено в § 31), *а функция* $q(t)$ *непрерывна в точке* $t = \theta$ *и ограничена. Пусть, кроме того,* $\mathbf{\Pi}$ *— любая мера на* $(\mathbb{R}, \mathfrak{B})$ *такая, что* $\int e^{-\alpha|u|^2/4}\mathbf{\Pi}(du) < \infty$. *Тогда если* θ *— внутренняя точка* Θ *и* $X \in \mathbf{P}_\theta$, *то*

$$J \equiv \int w(u^* - u) q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) \mathbf{\Pi}(du) =$$

$$= e^{Y(u^*)} q(\theta) \left[\int w(u^* - u) e^{-(1/2)(u-u^*)^2 I(\theta)} \mathbf{\Pi}(du) + \varepsilon_n(X, \theta)\right]. \quad (16)$$

В частности, если $\mathbf{\Pi}$ *есть мера Лебега,* $\mathbf{\Pi}(du) = du$, *то*

$$J = \sqrt{\frac{2\pi}{I(\theta)}} e^{Y(u^*)} q(\theta)(\mathbf{E}w(\eta) + \varepsilon_n(X, \theta)),$$

где $\varepsilon_n(X,\theta) \xrightarrow[\text{п.н.}]{} 0$, $\eta \in \Phi_{0, I^{-1}(\theta)}$.

Утверждение (16) весьма естественно, так как множитель $q(\theta + u/\sqrt{n})$ «почти постоянен», а функция $Z(u/\sqrt{n}) = e^{Y(u)}$ сближается с точностью до постоянного множителя по теореме 1 с плотностью нормального распределения.

Доказательство. Чтобы упростить запись, ограничимся рассмотрением случая, когда $\mathbf{\Pi}$ есть мера Лебега. Переход к общему случаю никаких затруднений не вызывает.

Оценим сначала часть интеграла (16) по области $|u| > r$. Обозначим ее через $J(r)$. Так как $f_\theta(X)/f_{\widehat{\theta}^*}(X) \leqslant 1$, то, полагая для краткости $Z = Z(u^*/\sqrt{n}) = e^{Y(u^*)}$, $t = \theta + u/\sqrt{n}$, получим

$$Z^{-1}Z\left(\frac{u}{\sqrt{n}}\right) = \frac{f_t(X)}{f_{\widehat{\theta}^*}(X)} \leqslant \left(\frac{f_t(X)}{f_{\widehat{\theta}^*}(X)}\right)^{3/4} \leqslant Z^{3/4}\left(\frac{u}{\sqrt{n}}\right).$$

Поэтому в силу неравенства Коши–Буняковского, теоремы 33.1 и следствия 33.2

$$\mathbf{E}_\theta w(u^*-u)Z^{-1}Z\left(\frac{u}{\sqrt{n}}\right) \leqslant \left[\mathbf{E}_\theta w^2(\sqrt{n}\,(\widehat{\theta}^* - t))\mathbf{E}_\theta Z^{1/2}\left(\frac{u}{\sqrt{n}}\right)\right]^{1/2} \leqslant ce^{-u^2 g/4}.$$

Так как $\max q(t) < \infty$, то отсюда и из леммы 33.1 находим

$$\mathbf{E}_\theta Z^{-1}J(r) \leqslant ce^{-gr^2/4}.$$

Пользуясь неравенством Чебышева, получим оценки того же порядка и для $\mathbf{P}_\theta(Z^{-1}J(r) > \delta)$. Поэтому если $r = r_n \to \infty$ так, что

$$\sum e^{-r_n^2 g/4} < \infty, \tag{17}$$

то при $y \geqslant r_n$

$$Z^{-1}J(y) \underset{\text{п.н.}}{\longrightarrow} 0. \tag{18}$$

Выберем $r_n = o(\sqrt{n})$ и рассмотрим оставшуюся часть $V(y) = J - J(y)$ интеграла J при $y = 2r_n$. По теореме 1

$$Z^{-1}V(2r) = Z^{-1}\int\limits_{|u|<2r_n} q\left(\theta + \frac{u}{\sqrt{n}}\right) w(u^*-u)Z\left(\frac{u}{\sqrt{n}}\right) du =$$

$$= \int\limits_{|u|<2r_n} (q(\theta) + \varepsilon_n(u))w(u^*-u)\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)(1+\varepsilon_n(X,\theta,u))\right\} du,$$

где $|\varepsilon_n(u)| < \varepsilon_n \to 0$, $|\varepsilon_n(X,\theta,u)| \leqslant \varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0$ при $n \to \infty$. Поэтому в силу (18) для доказательства теоремы нам достаточно убедиться в близости

интегралов

$$\int\limits_{|u|<2r_n} w(u^*-u)\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)(1+\varepsilon_n(X,\theta,u))\right\}du,$$

$$\sqrt{\frac{2\pi}{I(\theta)}}\mathbf{E}w(\eta)=\int w(u^*-u)\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)\right\}du.$$

В силу (17) и следствия 33.1 существует множество A, $\mathbf{P}_\theta(A)=1$, такое, что $|u^*|\leqslant r_n$ для $X_\infty\in A$ при всех достаточно больших $n=n(X_\infty)$. Так как $I(\theta)\geqslant g$, $|u-u^*|^2>u^2/2$ при $|u|>2r_n$, $|u^*|<r_n$, то на множестве A (см. лемму 33.1)

$$\int\limits_{|u|\geqslant 2r_n} w(u^*-u)\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)\right\}du<ce^{-gr_n^2}\to 0.$$

Поэтому нам остается оценить

$$\int\limits_{|u|<2r_n} w(u^*-u)\left|\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)(1+\varepsilon_n(X,\theta,u))\right\}-\right.$$

$$\left.-\exp\left\{-\frac{1}{2}(u-u^*)^2 I(\theta)\right\}\right|du\leqslant$$

$$\leqslant\int w(v)\left|\exp\left\{-\frac{1}{2}v^2 I(\theta)(1+\varepsilon_n(X,\theta,v+u^*))\right\}-\exp\left\{-\frac{1}{2}v^2 I(\theta)\right\}\right|dv.$$

Но этот интеграл сходится на множестве AB к нулю, где $B=\{X_\infty:\ \varepsilon_n(X,\theta)\to 0\}$, $\mathbf{P}_\theta(B)=1$. Это следует из сходимости к нулю при каждом v подынтегральной функции и того, что она мажорируется интегрируемой функцией. $\triangleleft$

Мы можем усилить теперь утверждение теоремы 33.9.

Теорема 3. *Пусть θ — произвольная внутренняя точка Θ, $X\Subset\mathbf{P}_\theta$. Пусть, далее, $q(t)$ — произвольная непрерывная и положительная внутри Θ плотность априорного распределения. Тогда* $\sqrt{n}\,(\widehat{\theta}^*-\theta_Q^*)\underset{\text{п.н.}}{\longrightarrow}0$.

Доказательство следует из теоремы 2. Действительно,

$$\theta_Q^*-\widehat{\theta}^*=\frac{\int(t-\widehat{\theta}^*)q(t)f_t(X)\,dt}{\int q(t)f_t(X)\,dt}.$$

Делая замену переменных $t=\theta+u/\sqrt{n}$ и деля числитель и знаменатель в этом выражении на $f_\theta(X)$, получим

$$\theta_Q^*-\widehat{\theta}^*=\frac{\int(u-u^*)q(\theta+u/\sqrt{n})Z(u/\sqrt{n})\,du}{\sqrt{n}\int q(\theta+u/\sqrt{n})Z(u/\sqrt{n})\,du}.$$

Надо воспользоваться теперь теоремой 2 при $w(t) = t$ и $w(t) = 1$. Так как в первом случае $\mathbf{E}w(\eta) = \mathbf{E}\eta = 0$, то мы получаем

$$\theta_Q^* - \widehat{\theta}^* = \frac{\varepsilon_n(X,\theta)}{\sqrt{n}}, \quad \varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0. \quad \triangleleft$$

Имеет место также асимптотическая минимаксность о. м. п. Чтобы установить этот факт, нам будет удобнее пользоваться несколько иной формой условий (RR) (см. § 33). В новой версии (RR_u) этих условий будет отсутствовать свойство 3, а интеграл (33.12) будет предполагаться сходящимся равномерно по θ. Приведем полностью эту новую версию условий (RR).

Условия (RR_u).

1. *Выполнены условия* (A_0), (A_c), (R).

2. *Функция* $l(x,\theta)$ *для п. в.* $[\mu]$ *значений* x *дважды непрерывно дифференцируема по* θ. *Функция* $|l''(x,t)|$ *мажорируется функцией* $l(x)$, *не зависящей от* t: $|l''(x,t)| < l(x)$, *для которой интеграл*

$$\mathbf{E}_t l(\mathrm{x}_1) = \int l(x) f_t(x) \mu(dx)$$

сходится равномерно по $t \in \Theta$ *).

Под равномерной сходимостью интеграла мы понимаем сходимость **)

$$\sup_\theta \int\limits_{x:|l(x)|>N} l(x) f_\theta(x) \mu(dx) \to 0$$

при $N \to \infty$.

В дальнейшем нам понадобятся следующие два свойства, вытекающие из (RR_u).

---

*) Все последующее изложение полностью сохранится, если условие о существовании мажоранты ослабить следующим образом: область Θ может быть покрыта конечным числом областей $\Theta_1, \ldots, \Theta_s$ так, что при $\theta \in \Theta_j$ функция $|l''(x,\theta)|$ мажорируется функцией $l_{(j)}(x)$, не зависящей от θ: $|l''(x,\theta)| < l_{(j)}(x)$, для которой интеграл

$$\mathbf{E}_\theta l_{(j)}(\mathrm{x}_1) = \int l_{(j)}(x) f_\theta(x) \mu(dx)$$

сходится равномерно по $\theta \in \Theta_j$, $j = 1, \ldots, s$.

**) Такое понимание равномерной сходимости находится в согласии с равномерной сходимостью, использовавшейся в теореме 5.4 — там оно относилось к функции $l(x) \equiv x$. В то же время это не есть равномерная сходимость $\int \varphi(x,\theta)\mu(dx)$ при $\varphi(x,\theta) = l(x) f_\theta(x)$, когда предполагается, что при $N \to \infty$

$$\sup_\theta \int\limits_{x:|\varphi(x,\theta)|>N} \varphi(x,\theta)\mu(dx) \to 0.$$

1. *Законность двойного дифференцирования по параметру под знаком интеграла в равенстве*

$$\int f_\theta(x)\mu(dx) = 1,$$

означающая справедливость соотношений

$$\int f'_\theta(x)\mu(dx) = 0, \quad \int f''_\theta(x)\mu(dx) = 0.$$

2. *Равномерная сходимость интеграла*

$$I(\theta) = \int (l'(x,\theta))^2 f_\theta(x)\mu(dx).$$

(Это свойство вытекает из (R) и понадобится в § 37.)

Чтобы разгрузить основное изложение, доказательство этих следствий условий (RR_u) отнесено в приложение V.

Для большинства утверждений этого параграфа безразлично, какой версией условий (RR) пользоваться: они будут справедливы при выполнении хотя бы одного из этих условий, при этом доказательства останутся прежними. Поэтому под (RR) мы можем понимать любую версию этих условий. Однако в утверждениях, связанных со свойством минимаксности, нам нужно будет условие (RR_u).

Теорема 4. *При выполнении* (RR_u) *о. м. п.* $\widehat{\theta}^*$ *является асимптотически минимаксной оценкой.*

Эта теорема выводится непосредственно из следствия 30.3 и следующего утверждения.

Лемма 3. *Справедливо равенство*

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} I^{-1}(\theta),$$

где Γ — *любой отрезок, лежащий внутри* Θ.

Лемма 3 вытекает из равномерной по θ сходимости (33.14). Равномерность будет доказана в § 37 (см. п. 37.3).

В дальнейшем под (RR) будем понимать любой из двух приведенных выше вариантов этих условий. (В теореме 4 приложения V будет доказано, что условия (RR_u) влекут за собой (RR).) В утверждениях, связанных с минимаксностью о. м. п., под (RR) будем понимать (RR_u).

§ 35.* Приближенное вычисление оценок максимального правдоподобия

Мы видели, что в задачах оценивания параметров наибольший интерес представляют эффективные и а. э. оценки, каковыми часто являются о. м. п. Возникает вопрос о практическом отыскании таких оценок. Отыскание точного значения о. м. п. $\widehat{\theta}^*$ в реальных задачах может представлять значительные трудности. Особенно это относится к распределениям, не имеющим сравнительно простых достаточных статистик.

С другой стороны, отыскание какой-нибудь а. н. оценки θ^* обычно трудностей не вызывает.

Приведем здесь способ приближенного отыскания M-оценок и, в частности, о. м. п., используя в качестве исходной какую-нибудь состоятельную или а. н. оценку θ^*. С помощью θ^* будет построена оценка θ_1^*, асимптотически эквивалентная о. м. п. и, стало быть, а. э. В основе этого способа лежит известный метод Ньютона (метод касательных) приближенного решения уравнения $s(t) = 0$, где s — некоторая непрерывно дифференцируемая функция. Напомним предварительно его суть. Пусть θ — решение уравнения $s(t) = 0$, и пусть в каждой точке t мы можем вычислить значения $s(t)$ и $s'(t)$. Будем предполагать также, что $s'(\theta) \neq 0$ и что «исходное» для приближений значение t_0 близко к θ и, стало быть, $s'(t_0) \neq 0$. Тогда, обозначив $\Delta = t_0 - \theta$, будем иметь

$$s(\theta) = 0 = s(t_0 - \Delta) = s(t_0) - \Delta s'(t_0) + o(\Delta),$$

$$\frac{s(t_0)}{s'(t_0)} = \Delta + o(\Delta).$$

Поэтому если положить

$$t_1 = U(t_0) \coloneqq t_0 - \frac{s(t_0)}{s'(t_0)}, \tag{1}$$

то получим

$$t_1 - \theta = U(t_0) - \theta = \Delta - \frac{s(t_0)}{s'(t_0)} = o(\Delta).$$

Таким образом, применение к t_0 преобразования U дает значение t_1, которое «на порядок» ближе к θ, чем t_0.

Перейдем теперь к статистическим оценкам. Положим

$$\begin{aligned} U(t) &= t - L'(X,t)(L''(X,t))^{-1}, \quad t \in \Theta, \\ U_1(t) &= t + L'(X,t)(nI(t))^{-1}, \quad t \in \Theta, \end{aligned} \tag{2}$$

где $L(X,t)$, как и прежде, логарифмическая функция правдоподобия.

Теорема 1. *Пусть выполнены условия* (RR), $X \in \mathbf{P}_\theta$ *и* θ^* — *любая а. н. оценка:*

$$(\theta^* - \theta)\sqrt{n} \Rightarrow \Phi_{0,\sigma^2(\theta)}.$$

Тогда оценка $\theta_1^* = U(\theta^*)$ (*или* $\theta_1^* = U_1(\theta^*)$) *будет асимптотически эквивалентной* $\widehat{\theta}^*$, *т. е.*

$$\left(\theta_1^* - \widehat{\theta}\right)\sqrt{n} \underset{P_\theta}{\longrightarrow} 0.$$

Доказательство теоремы будет опираться на следующую лемму.

Лемма 1. *Пусть выполнены условия* (RR), $X \in \mathbf{P}_\theta$, $\delta_n > 0$ *есть произвольная последовательность, сходящаяся к нулю. Тогда, если* θ_n

таково, что $|\theta_n - \theta| \leqslant \delta_n$, *то*

$$U(\theta_n) - \widehat{\theta}^* = (\theta_n - \widehat{\theta}^*)\varepsilon_n(\theta_n, \theta, X),$$

где $\bar{\varepsilon}_n \equiv \max\limits_{\theta_n : |\theta_n - \theta| \leqslant \delta_n} |\varepsilon_n(\theta_n, \theta, X)| \underset{P_\theta}{\longrightarrow} 0.$

То же самое утверждение будет справедливо, если вместо U *использовать функцию* U_1.

Другими словами, если пользоваться методом последовательных приближений к $\widehat{\theta}^*$ и положить $\theta_0^* = \theta_n$, $\theta_1^* = U(\theta_0^*)$ (или $\theta_1^* = U_1(\theta_0^*)$), то $\theta_1^* - \widehat{\theta}^* = o(\theta_0^* - \widehat{\theta}^*)$, так что приближение θ_1^* существенно лучше, чем θ_0^*.

Доказательство. Из рассмотрений § 34 и непрерывности L'' вытекает (см, например, лемму 34.1), что

$$L'(X, \theta_n) = (\theta_n - \widehat{\theta}^*)L''(X, \widetilde{\theta}), \quad L''(X, \widetilde{\theta}) = -n(I(\theta) + \varepsilon_n'(\theta_n, \theta, X)),$$

где $\widetilde{\theta} \in [\theta_n, \widehat{\theta}^*]$, $\max\limits_{\theta_n : |\theta_n - \theta| \leqslant \delta_n} \varepsilon_n'(\theta_n, \theta, X) \underset{P_\theta}{\longrightarrow} 0$ при любой последовательности $\delta_n \to 0$. Далее,

$$L''(X, \theta_n) = -n\big(I(\theta) + \varepsilon_n''\big),$$

$$\big(I(\theta) + \varepsilon_n'\big)\big(I(\theta) + \varepsilon_n''\big)^{-1} = 1 + \varepsilon_n,$$

где ε_n'', ε_n обладают тем же свойством, что и ε_n'. Следовательно,

$$U(\theta_n) - \widehat{\theta}^* = \theta_n - \widehat{\theta}^* - L'(X, \theta_n)(L''(X, \theta_n))^{-1} =$$
$$= \theta_n - \widehat{\theta}^* - (\theta_n - \widehat{\theta}^*)(1 + \varepsilon_n) = (\theta_n - \widehat{\theta}^*)\varepsilon_n.$$

Доказательство для функции U_1 проводится точно так же. $\triangleleft$

Доказательство теоремы 1. Выберем какое-нибудь $\delta_n \to 0$ такое, что $\delta_n\sqrt{n} \to \infty$, и представим $\big(\theta_1^* - \widehat{\theta}^*\big)\sqrt{n}$ в виде

$$(U(\theta^*) - \widehat{\theta}^*)\sqrt{n} = \sqrt{n}\,(\theta^* - \widehat{\theta}^*)\varepsilon_n(\theta^*, \theta, X)I_{(|\theta^* - \theta| \leqslant \delta_n)} + r_n,$$

где $r_n \neq 0$ лишь на множестве $B_n = \{X : |\theta^* - \theta| > \delta_n\}$, и в силу леммы 1

$$\bar{\varepsilon}_n = \max_{|t - \theta| \leqslant \delta_n} \varepsilon_n(t, \theta, X) \underset{P_\theta}{\longrightarrow} 0.$$

Так как, кроме того, $\mathbf{P}_\theta(B_n) \to 0$, то отсюда следует, что

$$\big|\theta_1^* - \widehat{\theta}^*\big|\sqrt{n} \leqslant \sqrt{n}\,|\theta^* - \theta|\bar{\varepsilon}_n + \sqrt{n}\,|\widehat{\theta}^* - \theta|\bar{\varepsilon}_n + r_n \underset{P_\theta}{\longrightarrow} 0. \quad \triangleleft$$

Теорема 1 показывает, что метод последовательных приближений, отправляясь от любой а. н. оценки, за один шаг приводит нас в точку $\widehat{\theta}^*$ с точностью до значений $o(1/\sqrt{n}\,)$.

Однако при выполнении некоторых условий последовательные приближения можно начинать с произвольного значения t, принадлежащего некоторому множеству. Тогда можно за «небольшое» число итераций преобразования $U(t)$, определенного в (2), получить оценку, асимптотически эквивалентную о. м. п. Обратимся сначала к неслучайной функции $s(t)$. Справедливо следующее вспомогательное предложение, которое поясняет суть дела. Пусть $U(t)$ — преобразование, определенное в (1).

Лемма 2. *Пусть отрезок* $K = [u_1, u_2]$ *таков, что решение* θ *уравнения* $s(t) = 0$ *принадлежит* K *и при всех* $t \in K$ *выполняется* $s'(t) \neq 0$ *и*

$$r_1 \leqslant \frac{s(t)}{(t-\theta)s'(t)} \leqslant r_2 \tag{3}$$

при некоторых $r_1 \in (0,1)$, $r_2 \in (1,2)$. *Тогда*

$$|U(t) - t| < p|t-\theta|, \quad p = \max(r_2 - 1,\, 1 - r_1) < 1. \tag{4}$$

Значение θ неизвестно, и поэтому проверять условие (3), по-видимому, удобнее с помощью следующего более сильного достаточного условия: *при всех* t, $u \in K$

$$r_1 \leqslant \frac{s(t) - s(u)}{(t-u)s'(t)} \leqslant r_2. \tag{5}$$

Ясно также, что при достаточно малом отрезке K условие (3) всегда выполнено.

Из леммы 2 следует, что если $U^k(t)$ есть k-я итерация преобразования U и все $U^{(k)} \in K$, то

$$|U^{(k)}(t) - \theta| < p^k|t-\theta|.$$

Принадлежность $U^{(k)}(t) \in K$ будет всегда иметь место по крайней мере для одного из концов u_1, u_2 отрезка K (для значения u_i, ближайшего к θ; это проверяется по ходу вычислений).

Доказательство леммы 2. Пусть для определенности $t > \theta$, $\Delta = t - \theta$. Тогда

$$U(t) - \theta = \Delta - \frac{s(t)}{s'(t)} \geqslant \Delta - r_2\Delta = -(r_2 - 1)\Delta,$$

$$U(t) - \theta \leqslant \Delta - r_1\Delta = (1 - r_1)\Delta.$$

Отсюда следует (4). Случай $t < \theta$ рассматривается аналогично. ◁

Из леммы 2 вытекает следующая рекомендация. Пусть с помощью оценки θ^* или априорных сведений нам удалось выделить отрезок $K \subset \Theta$ такой, что выполнено (5) для $s(t) = L'(X,t)$. Тогда для одного из значений $t = u_1$ или $t = u_2$ все $U^{(k)}(t)$ будут принадлежать K и будет иметь место

экспоненциальная сходимость $U^{(k)}(t)$ к θ (на самом деле сходимость будет еще более быстрой, так как при уменьшении K числа r_1 и r_2 сближаются с единицей). Признаком, по которому следует прекратить итерации, является, очевидно, первое совпадение (в рамках принятой точности) значений $U^{(k)}(t)$ и $U^{(k+1)}(t)$.

Описанную процедуру можно применять и без предварительной проверки условия (5). Критерием того, что она привела к значению $\widehat{\theta}^* = U^{(k_0)}(t)$ является совпадение $U^{(k_0)}(t)$ при всех k, начиная с некоторого k_0 ($L'(X, U^{(k)}(t)) = 0$; t — нулевое приближение), и неравенство $L''(X, U^{(k_0)}(t)) < 0$.

В многомерном случае $\theta \in \mathbb{R}^k$, $k > 1$, дело обстоит аналогичным образом. Как и прежде, рассмотрим сначала соответствующую детерминированную задачу: найти приближенное решение θ системы уравнений

$$s_i(t) = 0, \quad i = 1, \ldots, k,$$

для непрерывно дифференцируемой функции $s(t) = (s_1(t), \ldots, s_k(t))$. Здесь, полагая $\Delta = (\Delta_1, \ldots, \Delta_k) = t - \theta$, будем иметь

$$s_i(\theta) = 0 = s_i(t - \Delta) = s_i(t) - (\Delta, \operatorname{grad} s_i(t)) + o(|\Delta|).$$

Предположим, что $\det s'(\theta) \neq 0$, где

$$s'(t) = \|s_{ij}(t)\|, \quad s_{ij}(t) = \frac{\partial s_j(t)}{\partial t_i}.$$

Тогда при малых Δ будем иметь также $\det s'(t) \neq 0$ и можно записать в векторной форме

$$s(t) = \Delta s'(t) + o(|\Delta|), \quad s(t)(s'(t))^{-1} = \Delta + o(\Delta).$$

Полагая аналогично предыдущему

$$U(t) = t - s(t)(s'(t))^{-1},$$

получим

$$U(t) - \theta = \Delta - s(t)(s'(t))^{-1} = o(|\Delta|).$$

Это дает возможность перенести на многомерный случай все основные результаты, полученные выше. В частности, сохраняет свою силу теорема 1, в которой в определениях (2) под $L''(X, t)$ и $I(t)$ следует понимать соответствующие матрицы.

Пример 1. Классификация частиц. Рассмотрим источник, который излучает частицы двух типов: с вероятностью p — частицы типа A и с вероятностью $1 - p$ — частицы типа B. Энергия частиц случайна и имеет плотность $f_1(x)$ для частиц типа A и $f_2(x)$ для частиц типа B. Функции $f_i(x)$ известны. Зарегистрировано n частиц с энергиями $\mathrm{x}_1, \ldots, \mathrm{x}_n$. Чему равна вероятность p? Функция правдоподобия здесь равна

$$f_p(X) = \prod_{i=1}^{n} (p f_1(\mathrm{x}_i) + (1 - p) f_2(\mathrm{x}_i)),$$

так что

$$L'(X,p)=\sum_{i=1}^{n}\frac{f_1(\mathrm{x}_i)-f_2(\mathrm{x}_i)}{pf_1(\mathrm{x}_i)+(1-p)f_2(\mathrm{x}_i)}. \tag{6}$$

Мы видим, что отыскание о. м. п. $\widehat{p}^*$ приводит к уравнению $L'=0$ степени $n-1$ относительно p, решать которое при больших n весьма сложно. Воспользуемся теоремой 1. Для этого нам нужна какая-нибудь а. н. оценка p^*. Предположим, что

$$\int (F_1-F_2)^2\,dx<\infty,$$

где $F_i(x)=\int\limits_{-\infty}^{x} f_i(t)\,dt$, и рассмотрим следующий естественный подход. Определим p^* как значение, минимизирующее

$$\int \bigl(F_n^*(x)-F(x)\bigr)^2\,dx,\quad F(x)=pF_1(x)+(1-p)F_2(x). \tag{7}$$

Приравнивая нулю производную от (7), получим $\int \bigl(F_n^*-F\bigr)(F_1-F_2)\,dx=0$,

$$p^*=\frac{\int \bigl(F_n^*-F_2\bigr)(F_1-F_2)\,dx}{\int (F_1-F_2)^2\,dx}.$$

Оценка p^* является, очевидно, M-оценкой. Нетрудно видеть, что $\mathbf{E}p^*=p$ и что

$$(p^*-p)\sqrt{n}=\frac{\int \bigl(F_n^*-F\bigr)\sqrt{n}\,(F_1-F_2)\,dx}{\int (F_1-F_2)^2\,dx}. \tag{8}$$

Из результатов § 6–8 следует, что p^* есть а. н. оценка, а предельное распределение (8) совпадает с распределением

$$\frac{\int w^0(F(x))(F_1-F_2)\,dx}{\int (F_1-F_2)^2\,dx}.$$

Следовательно, в силу теоремы 1 оценка

$$p_1^*=p^*-L(X,p^*)(L''(X,p^*))^{-1},$$

где L' определено в (6),

$$L''=-\sum\frac{(f_1(\mathrm{x}_i)-f_2(\mathrm{x}_i))^2}{(pf_1(\mathrm{x}_i)+(1-p)f_2(\mathrm{x}_i))^2},$$

будет асимптотически эквивалентной о. м. п. $\widehat{p}^*$. Коэффициент рассеивания p_1^* будет определяться информацией

$$I(p) = \int \frac{(f_1(x) - f_2(x))^2}{pf_1(x) + (1-p)f_2(x)}\, dx$$

и будет меньше, чем коэффициент рассеивания p^*.

Пример 2. Предлагаем читателю аналогичным образом найти приближение для о. м. п. параметра α распределения Коши $\mathbf{K}_{\alpha,1}$ с плотностью

$$k_{\alpha,1}(x) = \frac{1}{\pi(1 + (x-\alpha)^2)}.$$

В качестве «предварительной» а. н. оценки можно взять выборочную медиану ζ^* (см. § 12 или § 3, 8; оценку $\alpha^* = \bar{\mathrm{x}}$ здесь брать нельзя, так как $\mathbf{E}_\alpha \alpha^*$ не существует). Оценка

$$\alpha_1^* = \zeta^* - L'(X, \zeta^*)(L''(X, \zeta^*))^{-1},$$

где

$$L'(X,\alpha) = -2\sum \frac{\mathrm{x}_i - \alpha}{1 + (\mathrm{x}_i - \alpha)^2}, \quad L''(X,\alpha) = 2\sum \frac{1 - (\mathrm{x}_i - \alpha)^2}{(1 + (\mathrm{x}_i - \alpha)^2)^2},$$

будет асимптотически эквивалентной о. м. п. $\widehat{\alpha}^*$. Так как

$$I(\alpha) = \int \frac{\left(k'_{\alpha,1}(x)\right)^2}{k_{\alpha,1}(x)}\, dx = \frac{4}{\pi}\int \frac{x^2}{(1+x^2)^3}\, dx = \frac{1}{2},$$

то коэффициенты рассеивания ζ^* и α_1^* будут равны соответственно (см. § 12)

$$\left(\frac{1}{2k_{\alpha,1}(\alpha)}\right)^2 = \left(\frac{\pi}{2}\right)^2, \quad I^{-1}(\alpha) = 2, \quad \left(\frac{\pi}{2}\right)^2 > 2.$$

Предлагаем читателю с помощью следствия 18.2 построить в явном виде другую оценку, асимптотически эквивалентную о. м. п. $\widehat{\alpha}^*$ и являющуюся L-оценкой. Здесь в обозначениях следствия 18.2 будем иметь

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x, \quad F^{-1}(t) = \operatorname{tg}\left[\pi\left(t - \frac{1}{2}\right)\right],$$

$$\dot{l}(x) = -\frac{2x}{1+x^2}, \quad \ddot{l}(x) = \frac{-2(1-x^2)}{(1+x^2)^2}.$$

Пример 3. Каждый человек имеет кровь, принадлежащую одной из четырех групп, которые обозначим через 0 (ноль), A, B, AB. Наследование групп крови управляется тремя генами: A, B, 0, при этом ген 0 «подавляется» генами A и B. Поэтому если p, q и $r = 1 - p - q$ обозначают вероятности появления генов A, B и 0 у родителей, то вероятности появления групп крови у потомка будут равны следующим величинам:

Таблица 1

i(номер группы)	Группа	Комбинации генов, дающие эту группу	Вероятности
1	0	00	r^2
2	A	AA, A0	p^2+2pr
3	B	BB, B0	q^2+2qr
4	AB	AB	$2pq$

Пусть ν_1, ν_2, ν_3, ν_4 — частоты появления соответствующих групп крови в обследуемой популяции общей численностью n человек. Как найти о. м. п. для p и q? В нашем случае вероятности $p_i(\theta)$, $\theta=(p,q)$, появления i-й группы крови и их частные производные по p и q представлены в табл. 2.

Таблица 2

	i			
	1	2	3	4
$p_i(\theta)$	r^2	$p(p+2r)$	$q(q+2r)$	$2pq$
$\frac{\partial p_i(\theta)}{\partial p}$	$-2r$	$2r$	$-2q$	$2q$
$\frac{\partial p_i(\theta)}{\partial q}$	$-2r$	$-2p$	$2r$	$2p$

Поэтому для логарифмической функции правдоподобия $L(X,\theta)=\sum_{i=1}^{4}\nu_i\ln p_i(\theta)$ получим

$$\frac{\partial L}{\partial p}=\sum\frac{\nu_i}{p_i}\frac{\partial p_i}{\partial p}=-\frac{2\nu_1}{r}+\frac{2r\nu_2}{p(p+2r)}-\frac{2\nu_3}{q+2r}+\frac{\nu_4}{p},$$
$$\frac{\partial L}{\partial q}=\sum\frac{\nu_i}{p_i}\frac{\partial p_i}{\partial q}=-\frac{2\nu_1}{r}-\frac{2\nu_2}{p+2r}+\frac{2r\nu_3}{q(q+2r)}+\frac{\nu_4}{q}. \tag{9}$$

Приравнивая к нулю эти производные, мы придем к системе из двух уравнений для θ^* четвертого порядка. Решение такой системы связано с техническими трудностями. Поэтому проще воспользоваться теоремой 1. Для этого заметим, что справедливы равенства

$$p_1=r^2,\quad p_1+p_2=(p+r)^2,\quad p_1+p_3=(q+r)^2. \tag{10}$$

Эффективные оценки для p_i равны $p_i^*=\nu_i/n$. Подставляя в (10) эти оценки и решая полученные уравнения, получим

$$p^*=\sqrt{p_1^*+p_2^*}-\sqrt{p_1^*},\quad q^*=\sqrt{p_1^*+p_3^*}-\sqrt{p_1^*}.$$

Так как p_i^* есть а. н. оценка p_i (т. е. $(p_i^*-p_i)\sqrt{n}\Rightarrow\Phi_{0,p_i(1-p_i)}$), то p^*, q^* в силу теоремы § 5 также будут а. н. оценками для p, q.

Чтобы воспользоваться теоремой 1, остается вычислить матрицу $(L''(X,\theta^*))^{-1}$ или матрицу $(nI(\theta^*))^{-1}$, $\theta^* = (p^*, q^*)$.

Приведем пример реальной выборки X, полученной в результате обследования $n = 353$ человека.

Распределение людей по группам крови приведено в табл. 3.

Таблица 3

	0	A	B	AB	Всего
ν_i	121	120	79	33	353
p_i^*	0,343	0,340	0,224	0,093	1

Из этой таблицы находим $p^* = 0{,}241$, $q^* = 0{,}167$, $r^* = 1 - p^* - q^* = 0{,}592$. С помощью табл. 2 для элементов матрицы $I(\theta)$ при $\theta = \theta^*$ получаем

$$\sum \left(\frac{\partial p_i(\theta)}{\partial p}\right)^2 \frac{1}{p_i(\theta)} = 4 + \frac{4r^2}{p(p+2r)} + \frac{4q}{q+2r} + \frac{2q}{p} = 9{,}970,$$

$$\sum \left(\frac{\partial p_i(\theta)}{\partial q}\right)^2 \frac{1}{p_i(\theta)} = 4 + \frac{4p}{p+2r} + \frac{4r^2}{q(q+2r)} + \frac{2p}{q} = 13{,}761,$$

$$\sum \frac{\partial p_i(\theta)}{\partial p}\frac{\partial p_i(\theta)}{\partial q}\frac{1}{p_i(\theta)} = 4 - \frac{4r}{p+2r} - \frac{4r}{q+2r} + 2 = 2{,}585.$$

Отсюда находим $|I(\theta^*)| = 130{,}512$,

$$I^{-1}(\theta^*) = \begin{vmatrix} 0{,}105 & -0{,}020 \\ -0{,}020 & 0{,}076 \end{vmatrix}.$$

Из формул для $\dfrac{\partial L}{\partial p}$ и $\dfrac{\partial L}{\partial q}$ (см. (9)) получаем

$$L'(\theta^*, X) = (25{,}443;\ 34{,}161), \tag{11}$$

так что для второго приближения θ_1^* имеем

$$\theta_1^* = \theta^* + \frac{1}{n} L'(\theta^*, X) I^{-1}(\theta^*) = (0{,}246;\ 0{,}173). \tag{12}$$

Это дает нам в дополнение к табл. 3 оценки, приведенные в табл. 3А.

Таблица 3А

	0	A	B	AB
$p_i(\theta^*)$	0,351	0,343	0,226	0,080
$p_i(\theta_1^*)$	0,337	0,347	0,231	0,085

Применение еще одной итерации вида (12) оценку θ_1^* уже не меняет (в рамках используемой нами точности), так как

$$L'(\theta_1^*, X) = (-0{,}076;\ -0{,}167)$$

(ср. с (11)), так что третье приближение для $\widehat{\theta}^*$ и все последующие будут совпадать с θ_1^*.

§ 36. Результаты § 33, 34 для случая многомерного параметра

В этом параграфе мы перенесем на многомерный случай все основные результаты § 33, 34. Излагаться эти результаты будут в той же последовательности, что и в указанных параграфах, при этом мы будем останавливаться лишь на тех моментах, где многомерность либо меняет формулировку результата, либо требует модификации рассуждений.

Итак, пусть $\theta \in \Theta \subset \mathbb{R}^k$, $k > 1$. Формулировки условий (A_μ), (A_c), (A_0), как и определения отношения правдоподобия

$$Z(u) = \frac{f_{\theta+u}(X)}{f_\theta(X)}$$

и расстояния Хеллингера

$$r(u) = \rho(\mathbf{P}_{\theta+u}, \mathbf{P}_\theta) = \int (\sqrt{f_{\theta+u}(x)} - \sqrt{f_\theta(x)})^2 \mu(dx),$$

с размерностью никак не связаны.

1. Неравенства для отношения правдоподобия и свойства о. м. п. (результаты § 33). Для изучения поведения функции $Z(u)$ в окрестности нуля нам понадобится следующее условие: *функция* $\sqrt{f_\theta(x)}$ *дифференцируема по* θ, *информационная матрица Фишера*

$$I(\theta) = \|I_{ij}(\theta)\| = \left\| \mathbf{E}_\theta \frac{\partial}{\partial\theta_i} l(\mathrm{x}_1, \theta) \frac{\partial}{\partial\theta_j} l(\mathrm{x}_1, \theta) \right\| \tag{1}$$

при всех $\theta \in \Theta$ *ограничена и положительно определена.*

При этом условии из теоремы 31.3А следует, что при всех θ

$$0 < g \leqslant \frac{r(u)}{|u|^2} \leqslant h = \frac{1}{4} \sup_\theta \mathrm{Sp}\, I(\theta) < \infty. \tag{2}$$

Здесь и в дальнейшем $|u|$ означает евклидову норму $|u| = \sqrt{u_1^2 + \ldots + u_k^2}$ вектора $u = (u_1, \ldots, u_k)$.

Первое утверждение теоремы 33.1 и его доказательство переносится на многомерный случай без каких-либо изменений, так как они с размерностью по существу не связаны.

Теорема 1. *Если выполнено* (2), *то*

$$\mathbf{E}_\theta Z^{1/2}(u) \leqslant e^{-ng|u|^2/2}.$$

Для обобщения теоремы 33.2 нам понадобится дополнительное условие, которое состоит в том, что

$$\gamma \equiv \sup_\theta \mathbf{E}_\theta |l'(\mathrm{x}_1, \theta)|^s < \infty \tag{3}$$

при некотором $s > k$.

Теорема 2 (аналог теоремы 33.2). *Если выполнены условия* (2), (3), *то при любых* z, $n \geqslant 1$

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant u} Z\left(\frac{v}{\sqrt{n}}\right) > e^z\right) \leqslant c\gamma(e^{-z} + e^{-z/2})e^{-\beta u^2}, \tag{4}$$

где $c < \infty$, $\beta > 0$ *зависят лишь от* k, g *и* s.

Для доказательства этого утверждения в одномерном случае мы использовали возможность оценивать $\sup\limits_{u\in(0,1)} p(u)$ через значения $p(0)$ и $\int\limits_0^1 |p'(u)|\,du$. Такой же подход в многомерном случае наталкивается на трудности, поскольку максимальное значение $p(u)$ в некоторой области $D \subset \mathbb{R}^k$, $k > 1$, не удается оценить, вообще говоря, через значения $p(u_0)$, $u_0 \in D$ и интеграл от $|p'(u)|$ ($p'(u) = \operatorname{grad} p(u)$) по какой-нибудь одной фиксированной кривой из D. Существует по крайней мере три пути преодоления этой трудности.

Первый является полной аналогией одномерному подходу и состоит в использовании оценки вида (в этой формуле мы для простоты ограничиваемся двумерным случаем $k = 2$)

$$\sup_{u\in K_{0,1}} p(u) \leqslant |p(0)| + \int_0^1 \left|\frac{\partial p((0,u_2))}{\partial u_2}\right| du_2 + \int_0^1 \left|\frac{\partial p((u_1,0))}{\partial u_1}\right| du_1 + \int_0^1\int_0^1 \left|\frac{\partial^2 p(u)}{\partial u_1 \partial u_2}\right| du_1 du_2,$$

где $u = (u_1, u_2)$, $K_{0,1}$ — единичный куб: $K_{0,1} = \{u : 0 \leqslant u_j \leqslant 1, j = 1, \ldots, k\}$. Однако для использования этого подхода мы должны предполагать существование производных k-го порядка функции $l_\theta(x)$ ($f_\theta(x)$) (см. определение функции p в § 33) и уметь оценивать нужные нам средние значения производных функции p l-го порядка, $l \leqslant k$.

Второй путь использует неравенства, восходящие к С. Л. Соболеву, об оценке $\sup\limits_{u\in K_{0,1}} p(u)$ через значения $p(0)$ и

$$\int_{K_{0,1}} |p'(u)|^s\,du \quad (p'(u) = \operatorname{grad} p(u),\ u = (u_1, \ldots, u_k))$$

при некотором $s > k$ (при $s = k$ оценка невозможна). При этом, разумеется, мы должны располагать оценками для $\mathbf{E}_\theta|p'(u)|^s$ при $s > k$ *).

В этой книге мы остановимся на третьем варианте, использующем некоторые идеи А. Н. Колмогорова об оценке распределения максимума случайного процесса. Ввиду того, что доказательство теоремы 2 занимает довольно много места, мы поместили его в приложение VI.

*) Этот вариант доказательства приведен в испанском издании [19] книги.

Доказательства утверждений о состоятельности о. м. п. и об оценках для моментов п. 33.2 с размерностью не связаны. Сами утверждения сохранятся в следующем виде.

Теорема 3 (аналог теоремы 33.3). *Если выполнены условия* (2), (3), *то при любых* v, $n \geqslant 1$ *справедливо* (33.6) *с заменой числа* $g/4$ *на* β (см. теорему 2).

Утверждения следствий 33.1, 33.2 сохраняются полностью с той же заменой $g/4$ на β.

2. Асимптотические свойства отношения правдоподобия (результаты § 33, 34). Под условиями (RR) в многомерном случае будем понимать совокупность следующих условий.

1. *Условия* (A_0), (A_c), (R).

2. *Непрерывная дифференцируемость второго порядка по* θ *внутри* Θ *функции* $l(x,t)$ *для п. в.* $[\mu]$ *значений* x. *При этом предполагается, что производные*

$$l''_{ij}(x,t) = \frac{\partial^2 l(x,t)}{\partial t_i \partial t_j}$$

допускают мажоранту $l(x)$, *не зависящую от* t: $\left|l''_{ij}(x,t)\right| \leqslant l(x)$, *для которой интеграл*

$$\mathbf{E}_t l(\mathrm{x}_1) = \int l(x) f_t(x) \mu(dx)$$

*сходится равномерно**) *по* $t \in \Theta$.

3. *Кроме того, будем предполагать, когда это потребуется, что выполнено условие* (3).

Как и в одномерном случае, нам понадобятся следующие два свойства, вытекающие из (RR).

1. *Возможность дважды дифференцировать по* θ *под знаком интеграла в равенстве*

$$\int f_\theta(x) \mu(dx) = 1,$$

означающая справедливость соотношений

$$\int \frac{\partial}{\partial \theta_i} f_\theta(x) \mu(dx) = 0, \quad \int \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta(x) \mu(dx) = 0. \tag{5}$$

2. *Равномерная сходимость интеграла* $I(\theta)$:

$$\sup_\theta \mathbf{E}_\theta [(l'_i(\mathrm{x}_1,\theta))^2;\ |l'_i(\mathrm{x}_1,\theta)| > N] \to 0 \tag{6}$$

при $N \to \infty$, *где* $l'_i(x,\theta) = \dfrac{\partial l(x,\theta)}{\partial \theta_i}$.

Доказательство этих свойств отнесено в приложение V. Для упрощения изложения они могут быть введены в условия (RR).

*) См. сноску о равномерной сходимости в § 34.

В силу равенств

$$l'_i(x,\theta) = \frac{1}{f_\theta(x)}\frac{\partial f_\theta(x)}{\partial \theta_i},$$

$$l''_{ij}(x,\theta) = \frac{1}{f_\theta(x)}\frac{\partial^2 f_\theta(x)}{\partial\theta_i\,\partial\theta_j} - \frac{1}{f^2_\theta(x)}\frac{\partial f_\theta(x)}{\partial\theta_i}\frac{\partial f_\theta(x)}{\partial\theta_j},$$

из соотношений (5) следует, что

$$\mathbf{E}_\theta l'_i(\mathrm{x}_1,\theta) = 0,$$

$$\mathbf{E}_\theta l''_{ij}(\mathrm{x}_1,\theta) = -\mathbf{E}_\theta l'_i(\mathrm{x}_1,\theta) l'_j(\mathrm{x}_1,\theta) = -I_{ij}(\theta).$$

Как и в одномерном случае, условия (RR) означают, что будут иметь место утверждения теорем 2, 3 об оценках для $\sup\limits_{|v|\geqslant u} Z(v/\sqrt{n})$ и для $\sqrt{n}\,(\widehat{\theta}^* - \theta)$.

При выполнении условий (RR) справедливы также следующие аналоги лемм 34.1, 34.2.

Лемма 1. *Функции $l''_{ij}(x,\theta)$ непрерывны «в среднем»*:

$$\mathbf{E}_\theta \omega''_\Delta(\mathrm{x}_1) \to 0$$

равномерно по θ при $\Delta \to 0$, где $\omega''_\Delta(x) = \max\limits_{i,j} \sup\limits_{\theta,|u|<\Delta} \bigl|l''_{ij}(x,\theta+u) - l''_{ij}(x,\theta)\bigr|$.

Доказательство в точности повторяет рассуждения леммы 34.1. $\triangleleft$

Положим

$$\gamma_n(\delta,\theta) = \sup_{\substack{\Delta\leqslant\delta\\ |\omega|=1}} \left|\frac{(L'(X,\theta+\omega\Delta),\omega) - (L'(X,\theta),\omega)}{n\Delta} + \omega I(\theta)\omega^\top\right|.$$

Лемма 2 (аналог леммы 34.2). *Пусть выполнены условия* (RR), *$\delta_n > 0$ — любая последовательность, сходящаяся к нулю. Тогда для $X \in \mathbf{P}_\theta$*

$$\gamma_n(\delta_n,\theta) \underset{\text{п.н.}}{\longrightarrow} 0, \quad \gamma_n(\delta_n,\widehat{\theta}^*) \underset{\text{п.н.}}{\longrightarrow} 0.$$

В этих соотношениях значения $I(\theta)$ и $I(\widehat{\theta}^*)$ можно заменить одно другим.

Доказательство. Как и в одномерном случае, нам достаточно убедиться, что $\gamma_n(\delta_n) \to 0$, где

$$\gamma_n(\delta) = \sup_{\substack{\Delta\leqslant\delta\\ |\omega|=1}} \left|\frac{(L'(X,\theta+\omega\Delta),\omega) - (L'(X,\theta),\omega)}{n\Delta} - \frac{\omega L''(X,\theta)\omega^\top}{n}\right|.$$

Но $\gamma_n(\delta_n) \leqslant \dfrac{1}{n}\sum_i \sum_{k,j} \omega''_{\delta_n}(\mathrm{x}_i)|\omega_k\omega_j|$, где $\omega''_\delta(x)$ есть максимальный модуль непрерывности функций $l''_{ij}(x,\theta)$. Так как

$$\sum_{k,j} |\omega_k\omega_j| \leqslant k|\omega|^2 = k,$$

то

$$\gamma_n(\delta_n) \leqslant \frac{k}{n}\sum_i \omega''_{\delta_n}(\mathrm{x}_i). \tag{7}$$

Дальнейшее доказательство основано на лемме 1 и в точности повторяет рассуждения леммы 34.2. $\triangleleft$

Обобщением теоремы 34.1 на многомерный случай здесь является

Теорема 4. *Пусть выполнены условия* (RR), $\delta_n > 0$, $n = 1, 2, \ldots$, — *любая последовательность, сходящаяся к нулю. Тогда если* $X \in \mathbf{P}_\theta$, *то для и таких, что* $|u/\sqrt{n}| \leqslant \delta_n$,

$$Y(u) \equiv \ln Z\left(\frac{u}{\sqrt{n}}\right) = (\xi_n, u) - \frac{1}{2}uI(\theta)u^\top(1 + \varepsilon_n(X,\theta,u)), \tag{8}$$

где

$$|\varepsilon_n(X,\theta,u)| \leqslant \varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0,$$

$$\xi_n = \frac{1}{\sqrt{n}}\operatorname{grad} L(X,\theta) = \frac{1}{\sqrt{n}}L'(X,\theta) \Subset \mathbf{\Phi}_{0,I(\theta)}.$$

Значение $u^* = \sqrt{n}\,(\widehat{\theta}^* - \theta)$, *при котором* $Y(u)$ *достигает максимума, представимо в виде*

$$u^* = \xi_n I^{-1}(\theta)(E + \varepsilon_n(X,\theta)), \quad \varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0, \tag{9}$$

где E *есть единичная матрица. Кроме того,*

$$2Y(u^*) = \xi_n I^{-1}(\theta)\xi_n^\top(1 + \varepsilon_n(X,\theta)) \Subset \xi I^{-1}(\theta)\xi^\top \in \mathbf{H}_k, \quad \xi \in \mathbf{\Phi}_{0,I(\theta)}. \tag{10}$$

Наряду с (8) *справедливо представление*

$$Y(u) - Y(u^*) = \frac{1}{2}(u - u^*)I(\theta)(u - u^*)^\top(1 + \varepsilon_n(X,\theta,u)),$$

$$|\varepsilon_n(X,\theta,u)| \leqslant \varepsilon_n(X,\theta).$$

Во всех приведенных утверждениях $I(\theta)$ *можно заменить на* $I(\widehat{\theta}^*)$.

В этом параграфе мы, как и в § 34, понимаем под $\varepsilon_n(X,\theta)$ различные последовательности, обладающие свойством $\varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0$ относительно P_θ.

Отметим также, что главную часть в (8) можно записать в виде

$$\xi_n u^\top - \frac{1}{2}uI(\theta)u^\top = -\frac{1}{2}(u - \xi_n I^{-1}(\theta))I(\theta)(u - \xi_n I^{-1}(\theta))^\top + \frac{1}{2}\xi_n I^{-1}(\theta)\xi_n^\top.$$

Это соответствует плотности многомерного нормального распределения со средним $\xi_n I^{-1}(\theta)$ и матрицей вторых моментов $I^{-1}(\theta)$.

Доказательство теоремы 4 вполне аналогично доказательству теоремы 34.1. Из леммы 2 при $\Delta \leqslant \delta_n$ получаем

$$(L'(X,\theta+\Delta\omega),\omega) = (L'(X,\theta),\omega) - n\Delta\omega I(\theta)\omega^{\top}(1+\varepsilon_n(X,\theta,\Delta\omega)),$$

$$|\varepsilon_n(X,\theta,\Delta\omega)| \leqslant \varepsilon_n(X,\theta).$$

Интегрируя это равенство по Δ от нуля до $|u|/\sqrt{n}$ и полагая $\omega = u/|u|$, получим

$$L\left(X,\theta+\frac{u}{\sqrt{n}}\right) - L(X,\theta) = \int\limits_0^{|u|/\sqrt{n}} (L'(X,\theta+\Delta u),u)\,d\Delta =$$

$$= \frac{|u|}{\sqrt{n}}(L'(X,\theta),\omega) - \frac{|u|^2}{2}\omega I(\theta)\omega^{\top}(1+\varepsilon_n(X,\theta,u)) =$$

$$= (\xi_n,u) - \frac{1}{2}uI(\theta)u^{\top}(1+\varepsilon_n(X,\theta,u)), \quad |\varepsilon_n(X,\theta,u)| \leqslant \varepsilon_n(X,\theta).$$

Здесь по многомерной центральной предельной теореме (см. приложение IV)

$$\xi_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l'(\mathrm{x}_i,\theta) \Subset \mathbf{\Phi}_{0,I(\theta)}.$$

Представление (8) доказано. Остальные утверждения теоремы доказываются точно так же, как в теореме 34.1, с учетом продемонстрированных изменений, связанных с многомерностью. Соотношение

$$\xi I^{-1}(\theta)\xi^{\top} \Subset \mathbf{H}_k$$

в (10) следует из свойств нормального распределения (см. п. 12.4). $\lhd$

В связи с соотношением (10) полезно также следующее

Замечание 1. Матрица $I^{-1}(\theta)$ вместе с $I(\theta)$ является положительно определенной, и существует матрица $I^{-1/2}(\theta)$, являющаяся корнем квадратным из $I^{-1}(\theta)$, т. е. удовлетворяющая соотношению

$$I^{-1/2}(\theta)I^{-1/2}(\theta) = I^{-1}(\theta).$$

Действительно, если некоторая матрица $M > 0$ (положительно определена), то существует ортогональная матрица C, для которой $CMC^{\top} = \operatorname{diag}(\lambda_1,\ldots,\lambda_k)$ есть диагональная матрица с положительными элементами $\lambda_i > 0$ на диагонали. Если положить теперь $M^{1/2} = C^{\top}\operatorname{diag}(\lambda_1^{1/2},\ldots,\lambda_k^{1/2})C$, то получим, очевидно, корень квадратный из M.

Пользуясь этим и симметричностью матрицы $I^{-1}(\theta)$, мы можем (10) записать в виде

$$(\xi_n I^{-1/2}(\theta))(\xi_n I^{-1/2}(\theta))^{\top}.$$

Здесь вектор $\eta_n = \xi_n I^{-1/2}(\theta)$ есть, очевидно, нормированная сумма n независимых одинаково распределенных случайных векторов с нулевым средним и матрицей вторых моментов

$$\mathbf{E}_\theta (\xi_n I^{-1/2}(\theta))^\top (\xi_n I^{-1/2}(\theta)) = \mathbf{E}_\theta I^{-1/2}(\theta) \xi_n^\top \xi_n I^{-1/2}(\theta) = E,$$

поскольку

$$\mathbf{E}_\theta \xi_n^\top \xi_n = \mathbf{E}_\theta (l'(\mathrm{x}_1, \theta))^\top (l'(\mathrm{x}_1, \theta)) = I(\theta).$$

Это означает, что по многомерной центральной предельной теореме $\xi_n I^{-1/2}(\theta) \Subset \Subset \mathbf{\Phi}_{0,E}$.

Теорема 5 (аналог теоремы 34.2). *Пусть выполнены условия теоремы* 34.2 *для многомерного* $\theta \in \mathbb{R}^k$ *и при* $\alpha = \beta/2$ (β *определено в теореме* 2). *Тогда*

$$J \equiv \int w(u^* - u) q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) \mathbf{\Pi}(du) =$$
$$= e^{Y(u^*)} q(\theta) \left[\int w(u^* - u) \exp\left\{ -\frac{1}{2}(u - u^*) I(\theta) (u - u^*)^\top \right\} \times \right.$$
$$\left. \times \mathbf{\Pi}(du) + \varepsilon_n(X, \theta) \right]. \quad (11)$$

Если $\mathbf{\Pi}$ *есть мера Лебега,* $\mathbf{\Pi}(du) = du$, *то*

$$J = \frac{(2\pi)^{k/2}}{\sqrt{|I(\theta)|}} e^{Y(u^*)} q(\theta) (\mathbf{E} w(\eta) + \varepsilon_n(X, \theta)), \quad (12)$$

где $\varepsilon_n(X, \theta) \underset{\text{п.н.}}{\longrightarrow} 0$, $\eta \Subset \mathbf{\Phi}_{0, I^{-1}(\theta)}$ (*последовательность* $\varepsilon_n(X, \theta)$ *является векторной, если* $w(t)$ *есть вектор-функция*).

Доказательство теоремы 5 проходит так же, как доказательство теоремы 34.2, поскольку последнее с размерностью не связано.

3. Свойства о. м. п. (результаты § 33, 34). Здесь мы будем предполагать везде выполненными условия (RR).

Аналог теоремы 33.5 будет иметь следующий вид.

Теорема 6. *О. м. п.* $\widehat{\theta}^*$ *является асимптотически нормальной оценкой, при этом сходимость*

$$u^* = (\widehat{\theta}^* - \theta)\sqrt{n} \Subset \mathbf{\Phi}_{0, I^{-1}(\theta)}$$

имеет место вместе со сходимостью моментов любого порядка. В частности,

$$\mathbf{E}_\theta n (\widehat{\theta}^* - \theta)^\top (\widehat{\theta}^* - \theta) \to I^{-1}(\theta). \quad (13)$$

Более того, для любой непрерывной функции $w(t)$ *такой, что* $|w(t)| < e^{\beta |t|^2/2}$ (*число* β *определено в теореме* 2),

$$\mathbf{E}_\theta w(u^*) \to \mathbf{E} w(\eta), \quad \eta \Subset \mathbf{\Phi}_{0, I^{-1}(\theta)}.$$

Соотношение (13) означает, что $\widehat{\theta}^* \in K_{\Phi,2}$.

Утверждение теоремы 6 следует из теоремы 4 (см. (9)) и многомерного аналога следствия 33.2, вытекающего из теоремы 3 (ср. с доказательством теоремы 33.5). $\triangleleft$

Определим класс $\widetilde{K}_0$ как совокупность оценок θ^*, для которых смещение $b(\theta) = (b_1(\theta), \ldots, b_k(\theta)) = \mathbf{E}_\theta \theta^* - \theta$ обладает свойствами

$$|b(\theta)| = o\left(\frac{1}{\sqrt{n}}\right), \quad b_{ij}(\theta) = \frac{\partial b_i(\theta)}{\partial \theta_j} \to 0$$

при $n \to \infty$.

Аналог теорем 33.6, 33.7 имеет здесь ту же самую форму.

Теорема 7. *$\widehat{\theta}^*$ является а. R-э. оценкой. Кроме того, $\widehat{\theta}^* \in \widetilde{K}_0$ является асимптотически эффективной в $\widetilde{K}_0$.*

Асимптотическая R-эффективность $\widehat{\theta}^*$, эквивалентная (13), очевидно, имеет место. Доказательство принадлежности $\widehat{\theta}^* \in \widetilde{K}_0$ и асимптотической эффективности в $\widetilde{K}_0$ происходит совершенно аналогично одномерному случаю.

Перейдем теперь к свойству асимптотической байесовости. Асимптотическая R-байесовость оценки θ^* означает по определению, что (ср. с § 30)

$$\mathbf{E}(\theta^* - \theta)^\top (\theta^* - \theta) = \frac{J}{n} + o\left(\frac{1}{n}\right), \quad J = \int I^{-1}(t)\mathbf{Q}(dt). \tag{14}$$

Асимптотическая байесовость θ^* означает

$$\limsup_{n\to\infty} \left[nv(\theta^*) - nv\left(\theta_Q^*\right)\right] \leqslant 0, \tag{15}$$

θ_Q^* — байесовская оценка, минимизирующая $v(\theta^*) = \mathbf{E}(\theta^* - \theta)V(\theta^* - \theta)^\top$ для любой неотрицательно определенной матрицы V.

Теорема 8 (аналог теоремы 33.8). *$\widehat{\theta}^*$ является асимптотически R-байесовской оценкой. Если априорное распределение $\mathbf{Q}$ имеет плотность относительно меры Лебега на Θ, то $\widehat{\theta}^*$ — асимптотически байесовская оценка.*

Доказательство вполне аналогично теореме 33.8. Соотношение (14) для $\theta^* = \widehat{\theta}^*$ следует из того, что

$$\lim_{n\to\infty} \mathbf{E} n(\widehat{\theta}^* - \theta)^\top (\widehat{\theta}^* - \theta) = \mathbf{E} \lim_{n\to\infty} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^\top (\widehat{\theta}^* - \theta) = \mathbf{E} I^{-1}(\theta) = J.$$

Предельный переход под знаком математического ожидания (т. е. интеграла) здесь законен, так как величина $\mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^\top (\widehat{\theta}^* - \theta)$ ограничена постоянной, не зависящей от n и θ (ср. со следствием 33.2).

Чтобы доказать (15), заметим, что в соответствии с § 30 интегральное неравенство Рао–Крамера в случае, когда $\mathbf{Q}$ имеет плотность, имеет вид

$$\mathbf{E} n(\theta^* - \theta)^\top (\theta^* - \theta) \geqslant J + o(1).$$

Это означает, что

$$nv(\theta_Q^*) \geqslant \sum v_{ij} J_{ij} + o(1),$$

где $\|J_{ij}\| = J$, $\|v_{ij}\| = V$. С другой стороны, в силу (14) при $\theta^* = \widehat{\theta}^*$

$$nv(\widehat{\theta}^*) = \sum v_{ij} J_{ij} + o(1).$$

Из этих соотношений, очевидно, вытекает (15) при $\theta^* = \widehat{\theta}^*$. $\triangleleft$

Аналоги теорем 33.9, 34.3 также будут иметь место. Например, из теоремы 5 вытекает

Теорема 9 (аналог теоремы 34.3). *Пусть* $X \in \mathbf{P}_\theta$ *и* θ *есть произвольная внутренняя точка* Θ. *Если* $q(t)$ *— произвольная непрерывная и положительная внутри* Θ *плотность априорного распределения* $\mathbf{Q}$, *то*

$$\sqrt{n}\,(\widehat{\theta}^* - \theta_Q^*) \underset{\text{п.н.}}{\longrightarrow} 0,$$

где θ_Q^* *есть байесовская оценка, соответствующая* $q(t)$.

Асимптотическая минимаксность $\widehat{\theta}^*$ может быть установлена аналогично теореме 34.3 с помощью многомерного аналога критерия асимптотической минимаксности в следствии 30.3:

$$\lim_{n\to\infty} \sup_{t\in\Gamma} \mathbf{E}_t n(\widehat{\theta}^* - \theta) V (\widehat{\theta}^* - \theta)^\top = \sup_{t\in\Gamma} \sum I_{ij}^{-1}(\theta)_{ij},$$

$$\left\| I_{ij}^{-1}(\theta) \right\| = I^{-1}(\theta),$$

и равномерного характера сходимости в (13), которая будет вытекать из результатов следующего параграфа.

Свойства асимптотической оптимальности $\widehat{\theta}^*$ в случае многомерного параметра θ, когда его размерность k велика, следует использовать с осторожностью. Нужно следить за тем, чтобы отношение n/k было велико (число наблюдений на один скалярный параметр). В противном случае выводы могут оказаться ошибочными.

Пример 1. В лаборатории проверяется концентрация n растворов. Каждая из неизвестных n концентраций $\mu_1, \ldots, \mu_n$ проверяется дважды. Предполагается, что дисперсия σ^2 всех n наблюдений $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_n, \mathrm{y}_n)$ одна и та же, а сами наблюдения независимы и нормально распределены, так что

$$f_\theta(X) = \frac{1}{\sigma^{2n}(2\pi)^n} \exp\left\{ -\frac{1}{2\sigma^2} \sum [(\mathrm{x}_i - \mu_i)^2 + (\mathrm{y}_i - \mu_i)^2] \right\},$$

где

$$\theta = (\mu_1, \ldots, \mu_n, \sigma^2).$$

О. м. п. для μ_i здесь равны

$$\widehat{\mu}_i^* = \frac{1}{2}(\mathrm{x}_i + \mathrm{y}_i).$$

Очевидно, эти оценки являются несмещенными, но не состоятельными. О. м. п. для σ^2 равна

$$(\widehat{\sigma}^2)^* = \frac{1}{4n}\sum(x_i - y_i)^2 \xrightarrow[P]{} \frac{\sigma^2}{2} \quad \text{при} \quad n \to \infty.$$

Эта оценка с большой надежностью дает ложное значение для параметра σ^2 (вдвое меньшее).

4. Приближенное вычисление о. м. п. Как мы видели в § 35, содержание этого параграфа в многомерном случае полностью сохранится.

5. Свойства о. м. п. при отсутствии условий регулярности (результаты § 14, 16). Условия состоятельности $\widehat{\theta}^*$, сформулированные в теоремах 14.1, 14.2 и в п. 16.2, с размерностью по существу не связаны. Доказательство этих теорем полностью сохраняется с точностью до очевидных изменений, связанных с тем, что множество Θ надо покрывать теперь (в силу условия (A_c)) не конечным числом интервалов, а конечным числом шаров. То же самое можно сказать и по поводу теоремы 16.2 и следствий 16.1–16.6.

§ 37. Равномерность по θ асимптотических свойств отношения правдоподобия и оценок максимального правдоподобия

В последующих рассмотрениях, главным образом в § 53–55 следующей главы, будут полезны утверждения § 33, 34 и 36 в их равномерной по θ форме. Большинство этих утверждений (скажем, утверждения о предельном $\mathbf{P}_\theta$-распределении $(\widehat{\theta}^* - \theta)\sqrt{n}$) были получены в предположении, что θ — фиксированная точка из Θ. Нас будет интересовать теперь, что будет происходить, если θ не фиксировано, а меняется вместе с n. Ясно, что при этом вместе с n будут меняться и распределения $\mathbf{P}_\theta$, так что каждая выборка X_n будет иметь «свое» распределение при $n = 1, 2, \ldots$

Мы приходим, таким образом, к *схеме серий* (см. [17, гл. 8]), для которой формулировки основных предельных теорем будут уже несколько иными. В частности, усиленный закон больших чисел, вообще говоря, теряет свой смысл, так как рассматриваемые случайные величины перестают быть заданными (при разных n) на одном вероятностном пространстве.

1. Равномерные закон больших чисел и центральная предельная теорема. Пусть $X \in \mathbf{P}_\theta$, $\eta_{n,\theta} = \eta_n(X, \theta)$.

Определение 1. Мы будем говорить, что последовательность $\eta_{n,\theta}$ *равномерно сходится по вероятности к постоянной* $a(\theta)$, если для любого $\varepsilon > 0$ при $n \to \infty$

$$\sup_{\theta \in \Theta} \mathbf{P}_\theta(|\eta_{n,\theta} - a(\theta)| > \varepsilon) \to 0.$$

Это соотношение будем записывать в форме «$\eta_{n,\theta} \xrightarrow[P_\theta]{} a(\theta)$ *равномерно по* θ».

Определение 2. Мы будем говорить, что $\eta_{n,\theta}$ *сходится по распределению к случайной величине* η_θ *равномерно по* θ, если для любой непрерывной и ограниченной функции φ при $n \to \infty$

$$\sup_{\theta} |\mathbf{E}_\theta \varphi(\eta_n, \theta) - \mathbf{E}\varphi(\eta_\theta)| \to 0. \tag{1}$$

Это соотношение будем записывать в виде «$\eta_{n,\theta} \Rightarrow \eta_\theta$ *равномерно по* θ». Такой же смысл будем придавать соотношению «$\eta_{n,\theta} \Subset \mathbf{G}_\theta$ равномерно по θ», где $\mathbf{G}_\theta$ означает распределение η_θ.

Мы предоставляем читателю самостоятельно убедиться, что если функции распределения η_θ непрерывны равномерно по θ, то соотношение (1) эквивалентно

$$\sup_{\theta,x} |\mathbf{P}_\theta(\eta_{n,\theta} < x) - \mathbf{P}(\eta_\theta < x)| \to 0.$$

Отметим, что равномерная сходимость $\eta_{n,\theta} \xrightarrow[P_\theta]{} a(\theta)$ и равномерная сходимость по распределению $\eta_{n,\theta} \Rightarrow a(\theta)$ к вырожденной случайной величине $a(\theta)$ эквивалентны.

Отметим также, что для равномерной сходимости сохранятся все основные теоремы непрерывности. Например, если H есть непрерывная функция, то из равномерной сходимости $\eta_{n,\theta} \Rightarrow \eta_\theta$ следует равномерная сходимость

$$H(\eta_{n,\theta}) \Rightarrow H(\eta_\theta). \tag{2}$$

Эти утверждения вытекают непосредственно из определений.

В приложении IV доказаны следующие «равномерные» предельные теоремы.

Пусть $X \Subset \mathbf{P}_\theta$ и $a(x,\theta)$ — заданная измеримая вектор-функция: $\mathcal{X} \times \Theta \to \mathbb{R}^l$. Рассмотрим суммы

$$s_n(\theta) = \sum a(\mathrm{x}_i, \theta)$$

независимых случайных векторов, зависящих от параметра $\theta \in \Theta$ как непосредственно через функцию $a(x,\theta)$, так и через распределение $\mathrm{x}_i \Subset \mathbf{P}_\theta$.

Напомним, что интеграл $\int \psi(x,\theta)\mathbf{P}_\theta(dx)$ мы называем *сходящимся равномерно по* θ *в области* Θ, если

$$\sup_{\theta\in\Theta} \int\limits_{|\psi(x,\theta)|>N} |\psi(x,\theta)|\mathbf{P}_\theta(dx) \to 0$$

при $N \to \infty$.

Теорема 1 (равномерный закон больших чисел). *Если интеграл* $a(\theta) = \int a(x,\theta)\mathbf{P}_\theta(dx)$ *сходится равномерно по* $\theta \in \Theta$, *то при* $n \to \infty$

$$\frac{s_n(\theta)}{n} \xrightarrow[P_\theta]{} a(\theta)$$

равномерно по θ.

Следствие 1. *Если последовательность* $\{\theta_n\} \in \Theta$, *то в условиях теоремы* 1

$$\mathbf{P}_{\theta_n}\left(\left|\frac{s_n(\theta_n)}{n} - a(\theta_n)\right| > \varepsilon\right) \to 0.$$

Этот факт будем обозначать

$$\frac{s_n(\theta_n)}{n} - a(\theta_n) \underset{P_{\theta_n}}{\longrightarrow} 0.$$

При рассмотрении центральной предельной теоремы для сумм $s_n(\theta)$ нам будет удобнее считать $a(\theta)$ равным нулю. (Это не есть ограничение общности, так как мы можем рассмотреть новые слагаемые $a^1(\mathrm{x}_1,\theta) = a(\mathrm{x}_1,\theta) - a(\theta)$.) Положим $\sigma^2(\theta) = \mathbf{E}_\theta(a^\top(\mathrm{x}_1,\theta)a(\mathrm{x}_1,\theta))$ и обозначим через $a_j(\mathrm{x}_1,\theta)$, $j = 1,2,\ldots,l$, координаты векторов $a(\mathrm{x}_1,\theta)$.

Теорема 2 (равномерная центральная предельная теорема). *Пусть интегралы* $\int (a_j^2(x,\theta))\mathbf{P}_\theta(dx)$, $j = 1,\ldots,l$, *сходятся равномерно в* Θ. *Тогда*

$$\eta_{n,\theta} = \frac{s_n(\theta)}{\sqrt{n}} \Rightarrow \eta_\theta \Subset \mathbf{\Phi}_{0,\sigma^2(\theta)}$$

равномерно по θ.

2. Равномерные варианты теорем об асимптотических свойствах отношения правдоподобия и оценок максимального правдоподобия. Заметим предварительно, что при выполнении условий (RR) результаты § 33 являются равномерными по θ по самой своей форме, так как правые части неравенств в теоремах 33.1–33.3 (и в теоремах 36.1–36.3) от θ не зависят. Перейдем к результатам § 34, 36 об асимптотическом поведении $Z(u/\sqrt{n}\,)$.

При этом под условиями (RR) в одномерном случае будем понимать условия (RR_u) (см. § 34).

Утверждения лемм 34.1, 36.1, 34.2, 36.2 можно сделать равномерными по θ.

Лемма 1. *При* $\Delta \to 0$

$$\sup_\theta \mathbf{E}_\theta \omega''_\Delta(\mathrm{x}_1) \to 0, \tag{3}$$

где $\omega''_\Delta(\mathrm{x}_1)$ *есть максимальный модуль непрерывности функций* $l''_{ij}(x,\theta)$.

Доказательство. Справедливость (3) при фиксированном θ доказана в лемме 36.1. Если допустить при этом отсутствие равномерности по θ, то мы придем к существованию $\varepsilon > 0$ и последовательностей $\theta_n \to \theta \in \Theta$, $\Delta_n \to 0$ таких, что

$$\mathbf{E}_{\theta_n}\omega''_{\Delta_n}(\mathrm{x}_1) > \varepsilon. \tag{4}$$

Полагая для краткости $\omega''_{\Delta_n}(\mathrm{x}_1) = \omega''$, получим

$$\mathbf{E}_{\theta_n}\omega'' = \mathbf{E}_{\theta_n}(\omega''; f_{\theta_n}(\mathrm{x}_1) \leqslant 2f_\theta(\mathrm{x}_1)) + \mathbf{E}_{\theta_n}(\omega''; f_{\theta_n}(\mathrm{x}_1) > 2f_\theta(\mathrm{x}_1), l(\mathrm{x}_1) \leqslant N) + \mathbf{E}_{\theta_n}(\omega''; f_{\theta_n}(\mathrm{x}_1) > 2f_\theta(\mathrm{x}_1), l(\mathrm{x}_1) > N).$$

Здесь первое слагаемое не превосходит $2\mathbf{E}_\theta\omega''$ и сходится к нулю в силу леммы 36.1, второе — не превосходит $2NJ_n$, где

$$J_n = \int\limits_{f_{\theta_n}(x)>2f_\theta(x)} f_{\theta_n}(x)\mu(dx) = 1 - \int\limits_{f_{\theta_n}(x)\leqslant 2f_\theta(x)} f_{\theta_n}(x)\mu(dx) \to 0$$

по теореме о мажорируемой сходимости. Наконец, последнее слагаемое не превосходит $\mathbf{E}_{\theta_n}(2l(\mathrm{x}_1); l(\mathrm{x}_1) > N)$ и в силу (RR) может быть сделано выбором N сколь угодно малым. Мы получили противоречие с (4), доказывающее лемму. $\triangleleft$

Лемма 2. *Утверждение леммы* 36.2 *сохранится, если в нем п. н. сходимость заменить сходимостью* $\gamma_n(\delta_n, \theta) \underset{P_\theta}{\longrightarrow} 0$, $\gamma_n(\delta_n, \widehat{\theta}^*) \underset{P_\theta}{\longrightarrow} 0$, *равномерной по* θ.

Доказательство. Будем следовать доказательству леммы 26.2. Заметим предварительно, что в силу теоремы 1 и равномерной сходимости интеграла в (RR)

$$\frac{L''(X, \theta)}{n} \underset{P_\theta}{\longrightarrow} -I(\theta)$$

равномерно по θ (сходимость матриц понимается поэлементно). Кроме того, из теорем 33.3, 36.3 вытекает, что $\widehat{\theta}^* \underset{P_\theta}{\longrightarrow} \theta$ равномерно по θ. Отсюда следует, что в соотношении $\gamma_n(\delta_n, \theta) \underset{P_\theta}{\longrightarrow} 0$ (см. лемму 36.2) можем $I(\theta)$ заменить на $L''(\theta)/n$ и на $I(\widehat{\theta}^*)$.

В силу неравенства (36.7) задача оценки $\gamma_n(\delta_n, \theta)$ сводится к оценке

$$\overline{\omega}''_{\delta_n}(X) = \frac{1}{n} \sum_{i=1}^{n} \omega''_{\delta_n}(\mathrm{x}_i, \theta),$$

где $\omega''_\Delta(x, \theta)$ — максимальный модуль непрерывности функций $l''_{ij}(x, \theta)$. Из неравенства Чебышева получаем

$$\sup_\theta \mathbf{P}_\theta\left(\overline{\omega}''_{\delta_n}(X) > \varepsilon\right) \leqslant \frac{1}{\varepsilon} \sup_\theta \mathbf{E}_\theta \omega''_{\delta_n}(\mathrm{x}_1, \theta).$$

Но в силу леммы 1 $\sup\limits_\theta \mathbf{E}_\theta \omega''_\Delta(\mathrm{x}_1, \theta) \to 0$ при $\Delta \to 0$. Это доказывает, что

$$\overline{\omega}''_{\delta_n}(X) \underset{P_\theta}{\longrightarrow} 0, \quad \gamma_n(\delta_n, \theta) \underset{P_\theta}{\longrightarrow} 0 \tag{5}$$

равномерно по θ.

Далее, из неравенств (34.6) следует, что задача оценки $\gamma_n(\delta_n, \widehat{\theta}^*)$ сводится к оценке $\overline{\omega}''_{\delta_n + |\widehat{\theta}^* - \theta|}(X)$. Так как $\widehat{\theta}^* - \theta \underset{P_\theta}{\longrightarrow} 0$ равномерно по θ, то из (5) получаем, что

$$\overline{\omega}''_{\delta_n + |\widehat{\theta}^* - \theta|}(X) \underset{P_\theta}{\longrightarrow} 0, \quad \gamma_n(\delta_n, \widehat{\theta}^*) \underset{P_\theta}{\longrightarrow} 0$$

равномерно по θ. $\triangleleft$

Теорема 3 (аналог теоремы 36.4). *При выполнении условий* (RR) *утверждения теоремы* 36.4 *сохранятся при следующих изменениях:* $\varepsilon_n(X, \theta) \underset{P_\theta}{\longrightarrow} 0$ *равномерно по* θ, $\xi_n \Subset \mathbf{\Phi}_{0, I(\theta)}$, $2Y(u^*) \Subset \mathbf{H}_k$ *равномерно по* θ.

Доказательство теоремы полностью основано на лемме 2 так же, как доказательство теоремы 36.4 — на лемме 36.2. Поэтому требуемое доказательство получается путем внесения очевидных изменений в доказательство теоремы 36.4, связанных с заменой (вытекающей из леммы 36.2) сходимости $\varepsilon_n(X,\theta) \underset{\text{п.н.}}{\longrightarrow} 0$ на равномерную сходимость $\varepsilon_n(X,\theta) \underset{P_\theta}{\longrightarrow} 0$. Кроме того, надо добавить, что

$$\xi_n = \frac{1}{n}\sum_{i=1}^{n} l'(\mathrm{x}_i,\theta) \Subset \mathbf{\Phi}_{0,I(\theta)}$$

равномерно по θ в силу теоремы 2 и равномерной сходимости (36.6) интеграла $I(\theta)$ (это есть матрица вторых моментов для $l'(\mathrm{x}_1,\theta)$), которая вытекает из условий (RR_u) (см. приложение V). Отсюда и из замечаний, касающихся (2), получаем равномерную сходимость

$$2Y(u^*) \Subset \mathbf{H}_k. \quad \triangleleft$$

Те же изменения, что и в теореме 3 (по сравнению с теоремой 36.4), могут быть внесены в теоремы 36.5, 36.6.

Приведем здесь следующие два следствия теоремы 3.

Теорема 4.

$$u^* = \sqrt{n}\,(\widehat{\theta}^* - \theta) \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)} \tag{6}$$

равномерно по θ. *При этом для любой функции* $w(x)$, *непрерывной п. в. относительно меры Лебега и такой, что* $|w(x)| < Ce^{\beta|x|^2/2}$ (*значение* $\beta > 0$ *определено в теореме* 36.2), *выполняется*

$$\sup_\theta |\mathbf{E}_\theta w(u^*) - \mathbf{E}w(\eta_\theta)| \to 0, \tag{7}$$

где $\eta_\theta \in \mathbf{\Phi}_{0,I^{-1}(\theta)}$.

Доказательство. Первое утверждение вытекает из соотношений

$$u^* = \xi_n I^{-1}(\theta)(E + \varepsilon_n(X,\theta)),$$

$$|\varepsilon_n(X,\theta)| \underset{P_\theta}{\longrightarrow} 0, \quad \xi_n \Subset \mathbf{\Phi}_{0,I(\theta)},$$

равномерных по θ и содержащихся в теореме 3.

Чтобы доказать второе утверждение, допустим, что (7) не верно. Тогда найдутся $\delta > 0$ и последовательность $\theta_n \to \theta \in \Theta$ такие, что

$$|\mathbf{E}_{\theta_n} w(u^*) - \mathbf{E}w(\eta_{\theta_n})| > \delta \tag{8}$$

при всех n.

Но $\mathbf{\Phi}_{0,I^{-1}(\theta_n)} \Rightarrow \mathbf{\Phi}_{0,I^{-1}(\theta)}$ и, стало быть, в силу (6) $\mathbf{P}_{\theta_n}$-распределение u^* $(w(u^*))$ слабо сходится к распределению η_θ $(w(\eta_\theta))$. Кроме того, по следствию 33.2 (см. также § 36)

$$\sup_\theta \mathbf{E}_\theta w^{3/2}(u^*) \leqslant \sup_\theta \mathbf{E}_\theta \exp\left|\frac{3(u^*)^2\beta}{4}\right| < c_1 < \infty.$$

Отсюда и из теорем непрерывности для моментов следует, что

$$\mathbf{E}_{\theta_n} w(u^*) \to \mathbf{E} w(\eta_\theta).$$

Так как $\mathbf{E} w(\eta_{\theta_n}) \to \mathbf{E} w(\eta_\theta)$, то полученное соотношение противоречит (8). $\triangleleft$

Пусть $A_n \subset \mathcal{X}^n$.

Теорема 5. *Если* $\mathbf{P}_\theta(A_n) \to 0$, *то при любом фиксированном* N

$$\sup_{|u| \leqslant N} \mathbf{P}_{\theta + u/\sqrt{n}}(A_n) \to 0.$$

Это свойство последовательностей распределений $\mathbf{P}_{\theta+u/\sqrt{n}}$ при $n \to \infty$ называется *контигуальностью* (см. [100]). Мы будем использовать его в главе 3.

Доказательство. Имеем

$$\mathbf{P}_{\theta+u/\sqrt{n}}(A_n) = \mathbf{E}_\theta \left\{ Z\left(\frac{u}{\sqrt{n}}\right); A_n \right\} \leqslant$$

$$\leqslant \mathbf{E}_\theta \left(Z\left(\frac{u}{\sqrt{n}}\right); A_n \cap \{Y(u) \leqslant c\} \right) + \mathbf{P}_{\theta+u/\sqrt{n}}(Y(u) > c) \leqslant$$

$$\leqslant e^c \mathbf{P}_\theta(A_n) + \mathbf{P}_{\theta+u/\sqrt{n}}(Y(u) > c).$$

Так как $\mathbf{P}_\theta(A_n) \to 0$, то для доказательства теоремы нам надо рассмотреть лишь $\sup\limits_{|u| \leqslant N} \mathbf{P}_{\theta+u/\sqrt{n}}(Y(u) > c)$. Согласно теореме 3 равномерно по u

$$Y(u) = (\xi_n, u) - \frac{1}{2} u I(\theta) u^\top \left(1 + \varepsilon_n\left(X, \theta + \frac{u}{\sqrt{n}}\right)\right) \Rightarrow \Phi_{-\frac{1}{2}\sigma^2, \sigma^2}, \qquad (9)$$

где $\sigma^2 = u I(\theta) u^\top \leqslant N^2 \Lambda_k(\theta)$ при $|u| \leqslant N$, $\Lambda_k(\theta)$ — максимальное собственное число матрицы $I(\theta)$. Так как $\Phi_{-\frac{1}{2}\sigma^2,\sigma^2}((c,\infty)) \leqslant \Phi_{0,\sigma^2}((c,\theta))$, то в силу равномерности в (9)

$$\lim_{n\to\infty} \sup_{|u|\leqslant N} \mathbf{P}_{\theta+u/\sqrt{n}}(Y(u) > c) \leqslant \sup_{|u|\leqslant N} \Phi_{0,\sigma^2}((c,\infty)) = \Phi_{0,N^2\Lambda_k(\theta)}((c,\infty)).$$

Это значение выбором c может быть сделано сколь угодно малым. $\triangleleft$

3. Некоторые следствия.

1) В § 33 была сформулирована теорема 33.7, в которой утверждалось, в частности, что $\widehat{\theta}^* \in \widetilde{K}^0$, где $\widetilde{K}^0$ есть класс асимптотически центральных оценок, определяемый соотношением (рассматривается одномерный случай)

$$\mathbf{P}_\theta(\widehat{\theta}^* > \theta) \to \frac{1}{2}$$

равномерно по θ. Из теоремы 4 вытекает, что упомянутая часть теоремы верна, так как

$$\mathbf{P}_\theta(\widehat{\theta}^* > \theta) = \mathbf{P}_\theta(\sqrt{n}\,(\widehat{\theta}^* - \theta)I^{-1/2}(\theta) > 0) \to \mathbf{\Phi}_{0,1}((0,\infty)) = \frac{1}{2}$$

равномерно по θ. $\triangleleft$

2) В § 34 была сформулирована теорема 34.4 об асимптотической минимаксности $\widehat{\theta}^*$. Для доказательства этой теоремы нужно установить справедливость леммы 34.3 о том, что

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} I^{-1}(\theta), \tag{10}$$

где Γ — любой отрезок из Θ. Но это утверждение есть прямое следствие равномерной по $\theta \in \Theta$ сходимости $\mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^2 \to I^{-1}(\theta)$, которая делает законным предельный переход под знаком $\sup\limits_{\theta\in\Gamma}$

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} \lim_{n\to\infty} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)^2 = \sup_{\theta\in\Gamma} I^{-1}(\theta). \quad \triangleleft$$

Утверждение, аналогичное (10) и обеспечивающее асимптотическую минимаксность $\widehat{\theta}^*$, будет иметь место, очевидно, и в многомерном случае:

$$\lim_{n\to\infty} \sup_{\theta\in\Gamma} \mathbf{E}_\theta n(\widehat{\theta}^* - \theta)V(\widehat{\theta}^* - \theta)^\top = \sup_{\theta\in\Gamma} \sum v_{ij} I_{ij}^{-1}(\theta),$$

$$\left\| I_{ij}^{-1}(\theta) \right\| = I^{-1}(\theta),$$

для любой матрицы V.

§ 38.* О статистических задачах, связанных с выборками случайного объема. Последовательное оценивание

Тот факт, что выборки случайного объема возникают на практике и являются естественными, подтверждают примеры 25.2, 28.3. Еще один пример связан с так называемым последовательным оцениванием. Оно применяется в тех случаях, когда мы можем производить наблюдения последовательно — одно за другим, и когда мы заинтересованы в минимизации числа этих наблюдений, скажем, в силу их высокой стоимости. При этом наряду с правилом оценивания (т. е. конструкцией оценки θ^*) мы должны задать правило остановки эксперимента. Эти правила могут быть различными: например, можем суммировать заданные стоимости $c(\mathrm{x}_i)$ наблюдений x_i до тех пор, пока не исчерпаем некоторое допустимое количество t. В этом случае момент ν остановки (номер последнего наблюдения или объем выборки) будет определяться как

$$\nu = \min\left\{ k : \sum_{i=1}^{k} c(\mathrm{x}_i) \geqslant t \right\}$$

— это есть «время первого прохождения уровня t» в блуждании со скачками $c(\mathrm{x}_i)$ (см. [17, гл. 9]). Можно суммировать «информации» $I(\mathrm{x}_i, \theta) = (l'(\mathrm{x}_i, \theta))^2$ и прекращать наблюдения, когда будет достигнут заданный уровень, и т. д.

В этих примерах ν есть марковский момент, т. е. $\{\nu > n\} \in \sigma(\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Это есть одно из основных предположений, при которых рассматриваются задачи последовательного оценивания. При этом предположении и выполнении ряда дополнительных, менее существенных условий сохраняется неравенство Рао–Крамера в виде

$$\mathbf{D}_\theta \theta^* \geqslant \frac{1}{I(\theta)\mathbf{E}\nu},$$

где $\theta^* = \theta^*(\mathrm{x}_1, \ldots, \mathrm{x}_\nu)$ — несмещенная оценка θ, $I(\theta)$ — информация Фишера. Доказательство этого неравенства вполне аналогично доказательствам § 26; надо лишь для вычисления информации Фишера, содержащейся в выборке $(\mathrm{x}_1, \ldots, \mathrm{x}_\nu)$, воспользоваться тождеством Вальда (см. [17, § 14.2]).

Если ν зависит от некоторого параметра t, как это было в примере 28.3, так что $\nu \to \infty$ п. н. при $t \to \infty$, то возможно построение асимптотически оптимальных оценок со среднеквадратичной погрешностью, асимптотически эквивалентной $(I(\theta)\mathbf{E}\nu)^{-1}$.

§ 39. Интервальное оценивание

1. Определения. До сих пор мы изучали свойства и способы отыскания наилучших *точечных* оценок неизвестного параметра θ, который определяет распределение $\mathbf{P}_\theta$ из семейства $\mathcal{P} = \{\mathbf{P}_\theta\}$, соответствующее выборке X. Точечные оценки используются в тех случаях, когда мы должны назвать некоторое число θ^*, предназначенное для использования вместо неизвестного θ.

Однако весьма распространен и другой подход к проблеме.

Будем считать θ скалярным параметром (многомерный случай будет рассмотрен в п. 6). Как мы знаем, точное определение θ по данной выборке невозможно. Но мы могли бы попытаться указать такой интервал (θ^-, θ^+), что он с заданной достаточно высокой вероятностью будет накрывать неизвестное значение θ. При этом несомненно, что чем у́же этот интервал, тем лучше. Во многих задачах заранее требуется, скажем, путем увеличения объема выборки построить такой интервал (θ^-, θ^+), чтобы его ширина не превосходила заданных размеров.

Определение 1. Пусть для заданного $\varepsilon > 0$ существуют случайные величины $\theta^\pm = \theta^\pm(\varepsilon, X)$ такие, что

$$\mathbf{P}_\theta(\theta^-(\varepsilon, X) < \theta, \quad \theta^+(\varepsilon, X) > \theta) \geqslant 1 - \varepsilon. \tag{1}$$

Тогда интервал (θ^-, θ^+) называется *доверительным интервалом для* θ *уровня* $1 - \varepsilon$.

Очевидно, что (1) можно записать в виде

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) \geqslant 1 - \varepsilon.$$

Событие, стоящее здесь под знаком вероятности, состоит в том, что *случайный интервал* (θ^-, θ^+) *накрыл неизвестное значение* θ. Читать это событие как «попадание θ в интервал (θ^-, θ^+)» было бы несколько менее аккуратно, так как θ, вообще говоря, не случайно.

Значения $\theta^{\pm}$ называются *границами доверительных интервалов*, число $1-\varepsilon$ — *коэффициентом* или *уровнем доверия*.

Итак, отличие интервального оценивания от точечного состоит в следующем.

1. Доверительный интервал как оценка менее «точен», так как указывается целое множество возможных значений θ.

2. С другой стороны, утверждение «$\theta \in (\theta^-, \theta^+)$ с вероятностью $\geqslant 1-\varepsilon$» является истинным, в то время как событие $\theta = \theta^*$, как правило, имеет вероятность, равную нулю.

В качестве ε выбирается обычно малое число. По нему строятся $\theta^{\pm}(\varepsilon, X)$, и затем на основании выборки объявляется, что $\theta \in (\theta^-(\varepsilon, X), \theta^+(\varepsilon, X))$. Поступая таким образом, мы будем ошибаться в длинном ряду экспериментов примерно в $100\varepsilon\,\%$ всех случаев. Например, если $\varepsilon = 0{,}001$, то ошибка будет происходить примерно 1 раз на 1000 случаев.

Объявляя верным соотношение $\theta \in (\theta^-, \theta^+)$, мы используем тот факт, что если некоторое событие имеет вероятность ε и это ε мало, то практически невозможно, чтобы это событие произошло при единичном испытании. Авиапассажир, садясь в самолет, интуитивно твердо верит в это. Ему достаточно знать, что вероятность того, что полет закончится благополучно, довольно высока (ведь ему известно, что она не равна 1). Именно такой подход и лежит в основе построения многих статистических процедур.

Мы выделим сначала один случай, когда построение доверительных интервалов особенно естественно и может быть произведено без существенных затруднений. Это так называемый байесовский случай, который уже рассматривался нами в § 20, 21, 30.

2. Построение доверительных интервалов в байесовском случае. Мы предполагаем здесь, что параметр θ выбирается *случайно* с известной нам априорной плотностью распределения $q(t)$ относительно некоторой меры λ на Θ. После этого производится выборка $X \in \mathbf{P}_\theta$, и нам надо построить доверительный интервал для выбранного значения θ.

Если выполнено условие (A_μ), то, как мы знаем из § 20, в этом случае существует апостериорное распределение θ (условное относительно X) с плотностью

$$q(t/X) = \frac{f_t(X)q(t)}{\int f_u(X)q(u)\lambda(du)}$$

относительно меры λ. Это означает, что в качестве $\theta^{\pm}(\varepsilon, X)$ достаточно взять любые два числа $\theta^{\pm}$, для которых

$$\int_{\theta^-}^{\theta^+} q(u/X)\lambda(du) = 1-\varepsilon$$

(или $\geqslant 1-\varepsilon$, если $\int_{-\infty}^{t} q(u/X)\lambda(du)$ меняется при изменении t дискретно).

Другими словами, в качестве θ^- и θ^+ следует взять квантили апостериорного

распределения соответственно порядков $1-\varepsilon_2$ и ε_1 при любых ε_1 и ε_2 таких, что $\varepsilon_1+\varepsilon_2=\varepsilon$.

Здесь в соотношении $\theta^- \leqslant \theta \leqslant \theta^+$ в отличие от небайесовского случая случайными являются все три элемента: границы интервала $\theta^{\pm}$ и сама величина θ.

Нетрудно видеть, что в описанной процедуре существует некоторый произвол, связанный с выбором чисел ε_1 и ε_2. Иногда этот произвол устраняется самой постановкой задачи — например, в случаях, когда нам важно установить лишь верхнюю или нижнюю доверительную границу. В этом случае одно из чисел ε_1, ε_2 следует положить равным нулю, а соответствующую границу сделать бесконечной. Если же границы играют симметричную роль, то естественно следует выбирать ε_i так, чтобы сделать интервал (θ^-, θ^+) по возможности более коротким. Для распределений $q(t/X)$, близких к симметричным, это достигается при $\varepsilon_1=\varepsilon_2=\varepsilon/2$.

3. Построение доверительных интервалов в общем случае. Асимптотические доверительные интервалы. Основные методы построения доверительных интервалов базируются на использовании точечных оценок. Рассмотрим сначала асимптотический подход к построению доверительных интервалов.

Определение 2. Пусть $X=[X_\infty]_n \in \mathbf{P}_\theta$, и для заданного $\varepsilon>0$ существуют случайные величины $\theta^{\pm}(\varepsilon, X)$ такие, что

$$\liminf_{n\to\infty} \mathbf{P}_\theta(\theta^-(\varepsilon,X) < \theta < \theta^+(\varepsilon,X)) \geqslant 1-\varepsilon. \tag{2}$$

Тогда интервал (θ^-, θ^+) называется *асимптотическим доверительным интервалом уровня* $1-\varepsilon$.

В этом определении необходимо подчеркнуть, что речь идет на самом деле о *последовательности* интервалов $\left(\theta_n^-, \theta_n^+\right)$, определенных для каждого n. Формально понятие асимптотического доверительного интервала применительно к выборке фиксированного объема является малосодержательным. Тем не менее соотношением (2) пользуются при больших n — так же, как пользуются центральной предельной теоремой для приближенного вычисления распределений сумм конечного числа случайных величин.

Мы видели в предыдущих пунктах, что большинство рассматривавшихся нами точечных оценок асимптотически нормальны. Ниже мы приводим конструкцию асимптотических доверительных интервалов, основанных на таких оценках.

Пусть θ^* — а. н. оценка:

$$(\theta^*-\theta)\sqrt{n} \Rightarrow \Phi_{0,\sigma^2(\theta)}, \tag{3}$$

и $\sigma(\theta)$ есть непрерывная функция. Так как $\theta^* \xrightarrow[P]{} \theta$, то последнее условие означает, что $\sigma(\theta^*) \xrightarrow[P]{} \sigma(\theta)$. Отсюда и из (3) по второй теореме непрерывности следует, что

$$\frac{(\theta^*-\theta)\sqrt{n}}{\sigma(\theta^*)} \Rightarrow \Phi_{0,1}. \tag{4}$$

Обозначим через λ_δ квантиль нормального распределения порядка $1-\delta$, т. е. такое число, что $\Phi_{0,1}((-\infty, \lambda_\delta)) = 1-\delta$, или $\mathbf{P}(|\xi|<\lambda_\delta)=1-2\delta$, если

$\xi \in \mathbf{\Phi}_{0,1}$. Для $\lambda_{\varepsilon/2}$ при заданном фиксированном $\varepsilon > 0$ введем временно более короткое обозначение, положив

$$\lambda_{\varepsilon/2} = \beta.$$

Тогда из (4) следует

$$\lim_{n\to\infty} \mathbf{P}_\theta \left(\left| \frac{(\theta^* - \theta)\sqrt{n}}{\sigma(\theta^*)} \right| < \beta \right) = 1 - \varepsilon.$$

Но это соотношение можно переписать в виде

$$\lim_{n\to\infty} \mathbf{P}_\theta \left(\theta^* - \frac{\beta\sigma(\theta^*)}{\sqrt{n}} < \theta < \theta^* + \frac{\beta\sigma(\theta^*)}{\sqrt{n}} \right) = 1 - \varepsilon.$$

Таким образом, числа

$$\theta^{\pm} = \theta^* \pm \frac{\beta\sigma(\theta^*)}{\sqrt{n}} \tag{5}$$

удовлетворяют определению 2 и, стало быть, являются границами асимптотического доверительного интервала уровня $1 - \varepsilon$.

Если теперь для заданной фиксированной выборки X объема n построим интервал (5), его истинный уровень будет, вообще говоря, отличаться от ε, но будет мало отличаться, если n достаточно велико. Поэтому обращаться с асимптотическими доверительными интервалами надо с известной осторожностью, выяснив предварительно, начиная с каких n вероятность события $\{\theta \in (\theta^-, \theta^+)\}$ достаточно точно приближается к предельным значениям. Обычно чем меньше ε, тем выше требование к объему выборки n. Необходимый объем зависит также от распределения $\mathbf{P}_\theta$ и от статистики θ^*.

Пример 1. Пусть $X \in \mathbf{\Gamma}_{\alpha,1}$, и мы пользуемся эффективной оценкой $\alpha^* = \dfrac{n-1}{n\bar{x}}$. В примерах 14.1, 26.1 было установлено, что

$$\mathbf{E}_\alpha \alpha^* = \alpha, \quad \mathbf{D}_\alpha \alpha^* = \frac{\alpha^2}{n-2},$$

так что здесь $\sigma^2(\alpha) = \alpha^2$. Соотношение (5) дает нам

$$\alpha^{\pm} = \frac{n-1}{n\bar{x}} \left(1 \pm \frac{\beta}{\sqrt{n}} \right). \tag{6}$$

Чему на самом деле равен уровень этого интервала?

Нам надо найти $\mathbf{\Gamma}_{\alpha,1}$ — вероятность неравенства

$$\frac{n-1}{n\bar{x}} \left(1 - \frac{\beta}{\sqrt{n}} \right) < \alpha < \frac{n-1}{n\bar{x}} \left(1 + \frac{\beta}{\sqrt{n}} \right)$$

или, что то же, неравенства

$$1 - \frac{\beta}{\sqrt{n}} < \frac{n\alpha\bar{x}}{n-1} < 1 + \frac{\beta}{\sqrt{n}},$$

где $n\alpha\bar{x} \in \Gamma_{1,n}$. Так как α есть параметр масштаба, то $2n\alpha\bar{x} \in \Gamma_{1/2,n} = H_{2n}$. Таким образом, точный уровень интервала (6) равен

$$\int_{2(n-1)(1-\beta/\sqrt{n})}^{2(n-1)(1+\beta/\sqrt{n})} \gamma_{1/2,n}(x)\,dx, \tag{7}$$

где $\gamma_{1/2,n}$ определено в § 12 *).

При $\varepsilon = 0{,}05$, $n = 30$ имеем $\beta = 1{,}96$, $(n-1)(1-\beta/\sqrt{n})/n \approx 0{,}6201$, $(n-1)(1+\beta/\sqrt{n})/n \approx 1{,}3126$.

Таким образом, асимптотический доверительный интервал уровня $1-\varepsilon = 0{,}95$ применительно к случаю $n = 30$ представляет собой интервал $(0{,}620/\bar{x},\ 1{,}313/\bar{x})$.

Если воспользоваться таблицами распределения χ^2 с 60 степенями свободы, то мы в силу (7) обнаружим, что точный уровень значимости этого доверительного интервала равен (с точностью до трех знаков) $0{,}937 = 1-0{,}063$. При этом «вклады» левого и правого конца доверительного интервала далеко не равнозначны (ср. с нормальным приближением) и составляют соответственно 0,010 и 0,053.

Для $n = 50$ асимптотический доверительный интервал уровня $0{,}95$ будет иметь вид $(0{,}708/\bar{x},\ 1{,}252/\bar{x})$. Истинный уровень его значимости будет равен $0{,}942 = 1 - 0{,}058$ (вклады равны соответственно 0,014 и 0,044). Ясно, что если продолжать увеличивать n, то эти вклады будут сближаться с 0,025.

Вернемся к доверительному интервалу (5), построенному нами с помощью а. н. оценки θ^*. В отличие от байесовского случая здесь имеется произвол, связанный с выбором оценки θ^*. Форма границ интервала показывает, что достичь заданных размеров интервала можно как за счет увеличения объема выборки n (что по разным причинам не всегда осуществимо), так и за счет возможного уменьшения $\sigma(\theta^*)$. Здесь мы приходим к важному выводу, что при равных объемах выборки лучший доверительный интервал будет давать оценка с меньшим рассеиванием $\sigma(\theta)$. Таким образом, *наилучшие асимптотические доверительные интервалы будут получены при использовании асимптотически эффективных оценок.*

При выполнении условий (RR) и принадлежности θ^* классу $\widetilde{K}_0 \cap K_{\Phi,2}$ (см. § 18, 26) наилучший асимптотический доверительный интервал имеет границы

$$\theta^{\pm} = \theta^* \pm \frac{\beta}{\sqrt{nI(\theta^*)}},$$

где θ^* — любая а. э. оценка, например о. м. п.

Некоторые другие методы построения асимптотических доверительных интервалов будут рассмотрены в п. 6.

*) Замечание о том, что $\Gamma_{1/2,n} = H_{2n}$, полезно, так как оно позволяет для вычисления $\Gamma_{\alpha,\lambda}$ (если 2λ целое) использовать таблицы распределения χ^2, приведенные в приложении, а также во многих других руководствах по математической статистике.

4. Построение точного доверительного интервала с помощью заданной статистики. Допустим, что в качестве статистики мы выбрали оценку θ^*. Тогда симметричный доверительный интервал уровня $1-\varepsilon$ с помощью этой оценки было бы естественно искать в форме $\theta^* \pm \Delta(\varepsilon, X)$ или в форме $\theta^*(1 \pm \Delta(\varepsilon, X))$, как это было в рассмотренном выше примере. Однако если мы попытаемся реализовать такой план, то окажется, что дело обстоит не так просто, так как в общем случае границы $\pm\Delta(\varepsilon, X)$ будут зависеть от неизвестного параметра θ — ведь $\Delta(\varepsilon, X)$ надо выбрать из условия

$$\mathbf{P}_\theta(\theta^* - \Delta(\varepsilon, X) < \theta < \theta^* + \Delta(\varepsilon, X)) \geqslant 1 - \varepsilon,$$

в которое θ входит существенным и весьма сложным образом, прежде всего через само распределение $\mathbf{P}_\theta$.

Поэтому требуется некоторая специальная конструкция для построения доверительных интервалов с помощью данной оценки θ^*.

В приводимом ниже построении наряду с оценкой θ^* может участвовать любая статистика S. Обозначим символом $\mathbf{G}_\theta$ распределение S и положим $G_\theta(x) = \mathbf{G}_\theta((-\infty, x))$.

Определение 3. Будем говорить, что статистика S *по распределению монотонно зависит от* θ, если при любых x, $\theta_1 < \theta_2$

$$\mathbf{G}_{\theta_1}((x, \infty)) \leqslant \mathbf{G}_{\theta_2}((x, \infty))$$

или, что то же,

$$G_{\theta_1}(x) \geqslant G_{\theta_2}(x). \tag{8}$$

Все разумные оценки θ^* обычно этим свойством обладают.

Если дополнительно монотонная зависимость $G_\theta(x)$ от θ является непрерывной, то уравнение

$$G_\theta(x) = \gamma$$

всегда разрешимо относительно θ при каждом $\gamma \in (0, 1)$. Решение этого уравнения обозначим через $b(x, \gamma)$.

Теорема 1. *Если* $\varepsilon_1 + \varepsilon_2 = \varepsilon$, *статистика* S *по распределению монотонно зависит от* θ *и функция* $G_\theta(x)$ *непрерывна по* θ *и* x, *то значения*

$$\theta^- = b(S, 1 - \varepsilon_2), \quad \theta^+ = b(S, \varepsilon_1)$$

образуют доверительный интервал уровня $1 - \varepsilon$.

Доказательство теоремы почти очевидно. Воспользуемся тем фактом, что если функция распределения $F(x)$ непрерывна и $\xi \Subset F$, то $F(\xi) \Subset \Subset \mathbf{U}_{0,1}$ ($\mathbf{P}(F(\xi) < x) = \mathbf{P}(\xi < F^{-1}(x)) = F(F^{-1}(x)) \equiv x$). В силу этого замечания $G_\theta(S) \Subset \mathbf{U}_{0,1}$ и, стало быть,

$$\mathbf{P}_\theta(\varepsilon_1 < G_\theta(S) < 1 - \varepsilon_2) = 1 - \varepsilon,$$

$$\mathbf{P}_\theta(b(S, 1 - \varepsilon_2) < \theta < b(S, \varepsilon_1)) = 1 - \varepsilon. \quad \triangleleft$$

Часто «обращение» функции $G_\theta(S)$, используемое в теореме, удобно проводить в два этапа. Сначала $G_\theta(x)$ обращается относительно x, т. е. определяются квантили $G_\theta^{-1}(\gamma)$ как решения уравнений $G_\theta(x) = \gamma$, а затем решаются относительно θ уравнения

$$G_\theta^{-1}(\varepsilon_1) = S, \quad G_\theta^{-1}(1-\varepsilon_2) = S.$$

Такие решения всегда будут существовать, так как по условию теоремы $G_\theta^{-1}(\gamma)$ монотонно и непрерывно зависит от θ.

На рис. 3 представлены кривые $y = G_\theta^{-1}(\varepsilon_1)$ и $y = G_\theta^{-1}(1-\varepsilon_2)$, определяющие при каждом θ область значений y, вероятность попадания в которую для некоторой оценки $S = \theta^*$ равна $1-\varepsilon$. Как уже отмечалось, процедура построения доверительного интервала есть обращение функций

$$y = G_\theta^{-1}(\varepsilon_1), \quad y = G_\theta^{-1}(1-\varepsilon_1),$$

т. е. отыскание точек пересечения соответствующих им кривых с уровнем $y = S$. Полученные точки пересечения и дают требуемый интервал (θ^-, θ^+).

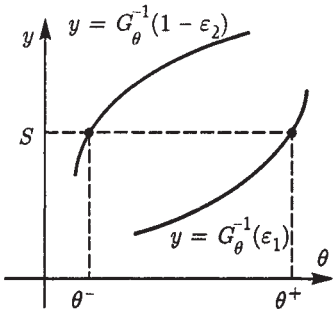

Рис. 3

Если условие о непрерывности $G_\theta(x)$ не выполнено, что будет иметь место для дискретных случайных величин S, то в целом изложенная процедура и утверждение теоремы 1 сохранятся с той лишь разницей, что при соответствующем определении квантилей $G_\theta^{-1}(\gamma)$ следует удовлетворить *неравенству* $\mathbf{G}_\theta\big((G_\theta^{-1}(\varepsilon_1), G_\theta^{-1}(1-\varepsilon_2))\big) \geqslant 1-\varepsilon$, на месте которого мы прежде имели точное равенство. В соответствии с этим утверждение теоремы 1 в этом случае будет иметь вид

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) \geqslant 1-\varepsilon,$$

где $\theta^\pm$ есть решения уравнений $G_\theta^{-1}(\varepsilon_1) = S$, $G_\theta^{-1}(1-\varepsilon_2) = S$. При этом мы будем, как и прежде, называть интервал (θ^-, θ^+) доверительным интервалом уровня $1-\varepsilon$.

Если мы строим доверительный интервал (θ^-, θ^+) с помощью оценки θ^*, то из рис. 3 видно, что этот интервал будет тем у́же, чем у́же интервал $\big(G_\theta^{-1}(\varepsilon_1), G_\theta^{-1}(1-\varepsilon_2)\big)$ или, что то же, чем более сконцентрировано распределение θ^* около θ. Таким образом, мы приходим здесь к той же проблеме, что в теории точечных оценок — к отысканию оценок θ^*, наиболее точно оценивающих θ.

Проблема построения наилучших доверительных интервалов более подробно будет рассмотрена в § 48.

Так как вид функций распределения $G_\theta(x)$ даже для простых семейств распределений, приведенных в § 12, оказывается обычно достаточно сложным, то указанная процедура обращения $G_\theta(x)$ становится на практике довольно трудным делом. Поэтому вычисление доверительных границ во многом табулировано. В следующем примере, где мы проиллюстрируем по-

строение доверительных интервалов по схеме, описанной в теореме 1, мы для упрощений будем пользоваться нормальным приближением.

Пример 2. Пусть $X \in \mathbf{B}_p$. В качестве оценки для p возьмем эффективную оценку $p^* = \nu/n$, где ν — число успехов в n испытаниях (числом ν может быть, например, число бракованных изделий, обнаруженных при контрольной проверке n образцов). Требуется построить доверительный интервал для доли брака p.

Имеем $(q = 1 - p)$

$$G_p(x) = \mathbf{P}_p(p^* < x) = \mathbf{P}_p\left(\frac{\nu - np}{\sqrt{npq}} < \frac{xn - np}{\sqrt{npq}}\right).$$

В соответствии с теоремой 1 нам надо решить уравнение

$$G_p(p^*) = \gamma \tag{9}$$

для значений γ, равных $\varepsilon/2$ и $1 - \varepsilon/2$. При больших n в силу центральной предельной теоремы $G_p(x) \approx \Phi((x-p)n/\sqrt{npq})$, где $\Phi(y) = \mathbf{\Phi}_{0,1}((-\infty, y))$, и, стало быть, уравнение (9) можно заменить его приближением

$$\Phi\left(\frac{(p^* - p)n}{\sqrt{npq}}\right) = \gamma, \quad \gamma = \frac{\varepsilon}{2}, 1 - \frac{\varepsilon}{2},$$

или, что то же, $|(p^* - p)n/\sqrt{npq}| = \lambda_{\varepsilon/2} = \beta$,

$$(p^* - p)^2 = \frac{\beta^2 p(1-p)}{n}.$$

Это уравнение для границ $p^{\pm}$ доверительного интервала представляет собой уравнение эллипса, вытянутого при больших n вдоль биссектрисы $p^* - p = 0$. Решая это уравнение относительно p, получим

$$p^{\pm} \approx p^* \pm \beta\sqrt{\frac{p^*(1-p^*)}{n}}.$$

Нетрудно проверить, что такой же результат мы получили бы, пользуясь асимптотическим подходом, изложенным в п. 3.

Если n невелико, то приходится $G_p(x)$ вычислять по точной формуле

$$G_p(x) = \sum_{k<nx} C_n^k p^k (1-p)^{n-k},$$

применяя затем процедуру теоремы 1.

Пусть, например, из $n = 10$ изделий $\nu = 2$ оказались бракованными. Тогда при $\varepsilon = 0{,}05$ точные границы доверительного интервала оказываются равными $p^- = 0{,}037$, $p^+ = 0{,}507$. Большая ширина интервала объясняется незначительной информацией, которой мы располагаем.

Если же $n = 100$, $\nu = 20$, то при $\varepsilon = 0{,}05$

$$p^- = 0{,}137, \quad p^+ = 0{,}277.$$

Эти цифры заимствованы из специальных таблиц, дающих численное решение задачи о доверительных интервалах для числа p при разных n и ν (см. [10]).

5. Другие методы построения доверительных интервалов. В этом пункте мы рассмотрим некоторые обобщения предложенной ранее процедуры построения доверительных интервалов.

Теорема 2. *Допустим, что на* $\Theta \times \mathcal{X}^n$ *существует такая функция* $G(\theta, \boldsymbol{x})$, *что распределение* $\mathbf{H}(B) = \mathbf{P}_\theta(G(\theta, X) \in B)$ *не зависит от* θ. *Предположим также, что* $G(\theta, \boldsymbol{x})$ *при каждом* $\boldsymbol{x}$ *непрерывна и монотонна по* θ.

Пусть, далее, y^-, y^+ *удовлетворяют соотношению* $\mathbf{H}((y^-, y^+)) = 1 - \varepsilon$. *Тогда статистики*

$$\theta^- = G^{-1}(y^-, X), \quad \theta^+ = G^{-1}(y^+, X), \quad \text{если} \quad G(\theta, \cdot) \uparrow,$$
$$\theta^- = G^{-1}(y^+, X), \quad \theta^+ = G^{-1}(y^-, X), \quad \text{если} \quad G(\theta, \cdot) \downarrow,$$

являются границами доверительного интервала уровня $1 - \varepsilon$. *Здесь* $G^{-1}(y, X)$ *есть решение уравнения* $G(\theta, X) = y$.

Доказательство. В силу монотонности $G(\theta, \boldsymbol{x})$ (пусть для определенности $G(\theta, \boldsymbol{x})$ возрастает по θ) событие $\{G^{-1}(y^-, X) < \theta < G^{-1}(y^+, X)\}$ совпадает с событием $A = \{y^- < G(\theta, X) < y^+\}$.

По определению $\mathbf{H}(\cdot)$ и $y^\pm$ имеем

$$\mathbf{P}_\theta(\theta^- < \theta < \theta^+) = \mathbf{P}_\theta(G^{-1}(y^-, X) < \theta < G^{-1}(y^+, X)) =$$
$$= \mathbf{P}_\theta(A) = \mathbf{H}((y^-, y^+)) = 1 - \varepsilon. \quad \triangleleft$$

Замечание 1. В теореме 1 в качестве $G(\theta, X)$ мы рассматривали функцию $G_\theta(S)$. При этом выполнялось $\mathbf{H} = \mathbf{U}_{0,1}$.

Замечание 2. Можно рассмотреть асимптотический аналог теоремы 2, допустив существование последовательности функций $\{G_n(\theta, x)\}$, непрерывных и монотонных по θ и таких, что при $n \to \infty$

$$\mathbf{P}_\theta(G_n(\theta, X) \in B) \to \mathbf{H}(B),$$

где $\mathbf{H}(\cdot)$ не зависит от θ. Мы получим тогда способ построения асимптотических доверительных интервалов, обобщающий изложенный в п. 3 метод построения асимптотических доверительных интервалов с помощью а. н. оценок.

Предложим теперь еще один способ (наряду с теоремой 1) выбора функции $G(\theta, \boldsymbol{x})$, фигурирующей в теореме 2.

Теорема 3. *Пусть* $F_\theta(x) = \mathbf{P}_\theta(\mathrm{x}_1 < x)$, *причем*
1) $F_\theta(x)$ *непрерывна по* x *для всех* $\theta \in \Theta$,

2) $F_\theta(x)$ *непрерывна и монотонна по* θ *для любого фиксированного* x. *Тогда функция*

$$G(\theta, \boldsymbol{x}) = -\sum_{i=1}^{n} \ln\left(F_\theta(x_i)\right)$$

удовлетворяет условиям теоремы 2.

Если числа $y^\pm$ *таковы, что*

$$\frac{1}{\Gamma(n)} \int_{y^-}^{y^+} x^{n-1} e^{-x}\, dx = 1 - \varepsilon, \tag{10}$$

то $y^\pm = G^{-1}(y^\pm, X)$ *образуют границы доверительного интервала уровня* $1 - \varepsilon$.

Доказательство. Проверим выполнение условий теоремы 2. Поскольку по условию 1 $F_\theta(\mathrm{x}_i)$ распределена равномерно на $[0,1]$, то $-\ln F_\theta(\mathrm{x}_i) \in \in \boldsymbol{\Gamma}_{1,1}$ и $G(\theta, X) \in \boldsymbol{\Gamma}_{1,n}$. Другими словами, $\mathbf{P}_\theta(G(\theta, X) \in B) = \boldsymbol{\Gamma}_{1,n}(B)$ и $\mathbf{H} = \boldsymbol{\Gamma}_{1,n}$ на зависит от θ. Монотонность и непрерывность $G(\theta, \boldsymbol{x})$ при каждом $\boldsymbol{x}$ следуют из условия 2. Кроме того, в силу (10)

$$\mathbf{H}((y^-, y^+)) = \boldsymbol{\Gamma}_{1,n}((y^-, y^+)) = 1 - \varepsilon. \quad \triangleleft$$

Можно указать и некоторые другие конструкции доверительных интервалов. При этом так же, как в теории точечного оценивания, сразу же возникает вопрос о том, какой доверительный интервал, если их получено несколько, считать лучшим. Подходов, которые здесь существуют, мы коснемся в § 48. Однако из предыдущего изложения ясно, что по существу задача поиска оптимального доверительного интервала во многом очень близка к задаче оптимального точечного оценивания. Ясно также, что если мы строим доверительные интервалы с помощью точечных оценок, то предпочтение следует отдавать доверительным интервалам, построенным с помощью лучших оценок.

Близость задач оптимизации точечного и интервального оценивания можно проиллюстрировать на примере следующего утверждения.

Теорема 4. *Рассмотрим асимптотический доверительный интервал* (θ^-, θ^+) *уровня* $1 - \varepsilon$ *и предположим, что случайная величина* $\theta^* = (\theta^+, \theta^-)/2$ *есть асимптотически центральная* (см. п. 33.4) *а. н. оценка, а величина* $\Delta = (\theta^+ - \theta^-)/2$ *такова, что* $\delta = \liminf\limits_{n\to\infty} \sqrt{n}\,\Delta$ *от* X *не зависит. Тогда* $\delta \geqslant \beta/\sqrt{I(\theta)}$.

Это значит, что ширина доверительного интервала (θ^-, θ^+) не может быть существенно меньше, чем $2\beta/\sqrt{nI(\theta)}$, т. е. чем ширина интервала $1 - \varepsilon$, построенного с помощью о. м. п. $\widehat{\theta}^*$.

Доказательство. Допустим противное. Тогда найдется подпоследовательность чисел $\{n'\}$, для которых $\Delta\sqrt{n'} \to c\beta/\sqrt{I(\theta)}$, $c < 1$. Так как

$\theta^{\pm} = \theta^* \pm \Delta$, то

$$1 - \varepsilon = \lim_{n' \to \infty} \mathbf{P}_\theta(\theta^- < \theta < \theta^+) = \lim_{n' \to \infty} \mathbf{P}_\theta(|\theta^* - \theta| < \Delta) =$$

$$= \lim_{n' \to \infty} \mathbf{P}_\theta\left(|\theta^* - \theta|\sqrt{n'} < \frac{c\beta}{\sqrt{I(\theta)}}\right) \leqslant$$

$$\leqslant \lim_{n \to \infty} \mathbf{P}_\theta\left(|\widehat{\theta}^* - \theta|\sqrt{n} < \frac{c\beta}{\sqrt{I(\theta)}}\right). \quad (11)$$

Последнее неравенство следует из того, что о. м. п. $\widehat{\theta}^*$ является а. э. в классе $\widetilde{K}^0$ асимптотически центральных оценок (см. теорему 33.7). Так как правая часть в (11) строго меньше $1 - \varepsilon$, мы получили противоречие, доказывающее теорему. $\triangleleft$

6. Многомерный случай. Понятие доверительного интервала обобщается в случае многомерного параметра $\theta \in \mathbb{R}^k$ в понятие доверительной области или доверительного множества.

Определение 4. Случайное*) подмножество $\Theta^* = \Theta(\varepsilon, X)$ параметрического пространства Θ называется *доверительным множеством уровня* $1 - \varepsilon$, если

$$\mathbf{P}_\theta(\Theta^* \ni \theta) \geqslant 1 - \varepsilon. \quad (12)$$

Другими словами, доверительное множество уровня $1 - \varepsilon$ накрывает неизвестное истинное значение θ с вероятностью, не меньшей $1 - \varepsilon$.

Определение 5. Если $X = [X_\infty]_n \in \mathbf{P}_\theta$, и случайное множество Θ^* удовлетворяет соотношению

$$\liminf_{n \to \infty} \mathbf{P}_\theta(\Theta^* \ni \theta) \geqslant 1 - \varepsilon,$$

то Θ^* называется *асимптотическим доверительным множеством уровня* $1 - \varepsilon$.

Изучению «точных» доверительных множеств, в том числе оптимальных, посвящен § 48 следующей главы.

Что касается асимптотических доверительных множеств, то принцип их построения остается прежним. Учитывая теорему 4, рассмотрим сразу доверительные множества, построенные с помощью о. м. п. $\widehat{\theta}^*$. Как мы знаем, при выполнении условий (RR), $X \in \mathbf{P}_\theta$

$$(\widehat{\theta}^* - \theta)\sqrt{n}\, I^{1/2}(\theta) \Rightarrow \mathbf{\Phi}_{0,E}.$$

Отсюда следует, что

$$n(\widehat{\theta}^* - \theta)I(\theta)(\widehat{\theta}^* - \theta)^\top \Rightarrow \mathbf{H}_k,$$

$$n(\widehat{\theta}^* - \theta)I(\widehat{\theta}^*)(\widehat{\theta}^* - \theta)^\top \Rightarrow \mathbf{H}_k.$$

*) В этом контексте множество $\Theta^*(\varepsilon, X)$ будем называть случайным, если при каждом t множество $\{X : t \in \Theta^*(\varepsilon, X)\}$ измеримо и, стало быть, определена вероятность (12) (ср. с § 48).

Другими словами, если h_ε означает квантиль порядка $1-\varepsilon$ распределения χ^2 с k степенями свободы, то

$$\lim_{n\to\infty} \mathbf{P}_\theta(n(\theta-\widehat{\theta}^*)I(\widehat{\theta}^*)(\theta-\widehat{\theta}^*)^\top < h_\varepsilon) = 1-\varepsilon. \tag{13}$$

Мы построили асимптотическое доверительное множество Θ^* уровня $1-\varepsilon$, которое представляет собой эллипсоид с центром в точке $\widehat{\theta}^*$ и с осями, определяемыми матрицей $nI(\widehat{\theta}^*)/h_\varepsilon$. Вычислять матрицу $I(\theta)$ для построения Θ^* при этом не обязательно. Как мы знаем, при выполнении условий (RR), $X \in \mathbf{P}_\theta$

$$L(X,\theta) - L(X,\widehat{\theta}^*) \approx -\frac{n}{2}(\theta-\widehat{\theta}^*)I(\widehat{\theta}^*)(\theta-\widehat{\theta}^*)^\top.$$

Поэтому эллипсоид Θ^*, определенный в (13), можно представить как совокупность значений θ, для которых

$$L(X,\theta) - L(X,\widehat{\theta}^*) \geqslant -\frac{h_\varepsilon}{2}.$$

В § 36 было установлено, что предел $\mathbf{P}_\theta$-вероятности этого неравенства (см. замечание 36.2) равен $1-\varepsilon$.

Отсюда следует, в частности, что в одномерном случае границы $\theta^\pm$ асимптотического доверительного интервала уровня $1-\varepsilon$ можно определять как решения уравнения

$$L(X,\theta) - L(X,\widehat{\theta}^*) = -\frac{h_\varepsilon}{2} = -\frac{\beta^2}{2}.$$

§ 40. Точные выборочные распределения и доверительные интервалы для нормальных совокупностей

Среди всех распределений, перечисленных в § 12, нормальное распределение наиболее распространено в приложениях. Поэтому в этом параграфе мы остановимся отдельно на построении точных доверительных интервалов для параметров α и σ^2 распределения $\mathbf{\Phi}_{\alpha,\sigma^2}$.

1. Точные распределения статистик $\overline{\mathrm{x}}$, S_0^2. Пусть $X \in \mathbf{\Phi}_{0,1}$, и $C = \|c_{ij}\|$ $(i,j=1,2,\ldots,n)$ есть ортогональная матрица.

Изучим распределение n-мерного вектора

$$Y = XC, \quad Y = (\mathrm{y}_1,\ldots,\mathrm{y}_n), \quad \mathrm{y}_i = \sum_{j=1}^n \mathrm{x}_j c_{ji}.$$

Лемма 1. *Если C — ортогональная матрица, то $Y \in \mathbf{\Phi}_{0,1}$, т. е. координаты $\mathrm{y}_1,\ldots,\mathrm{y}_n$ есть независимые случайные величины, $\mathrm{y}_i \in \mathbf{\Phi}_{0,1}$, $i=1,2,\ldots,n$.*

Доказательство. Пусть t есть вектор $(t_1,\ldots,t_n)$. Нормальность распределения X означает, что его характеристическая функция равна

$$\mathbf{E}e^{itX^T} = e^{-\frac{1}{2}tmt^\top},$$

где $m = \|m_{ij}\|$ есть матрица вторых моментов, равная в нашем случае единичной матрице E, для которой $tEt^\top = \sum_{j=1}^{n} t_j^2$,

$$\mathbf{E}e^{itX^\top} = e^{-\frac{1}{2}\sum_{j=1}^{n} t_j^2}.$$

Характеристическая функция совместного распределения $\mathrm{y}_1, \ldots, \mathrm{y}_n$ (или распределения вектора Y) имеет вид

$$f(t) = \mathbf{E}e^{itY^\top} = \mathbf{E}e^{itC^\top X^\top}.$$

Делая замену переменных $t = uC$ и замечая, что $CC^\top = E$, получим

$$f(t) = \mathbf{E}e^{iuCY^\top} = \mathbf{E}e^{iuX^\top} = e^{-\frac{1}{2}\sum_{i=1}^{n} u_i^2} = e^{-\frac{1}{2}\sum_{i=1}^{n} t_i^2}.$$

Это означает, что Y имеет ту же характеристическую функцию и, стало быть, то же распределение, что и X. ◁

Теперь докажем важное для дальнейшего утверждение, которое носит название леммы Фишера.

Лемма 2. *Пусть по-прежнему* $X \in \mathbf{\Phi}_{0,1}$, C — *ортогональная матрица*, $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_n) = XC$. *Тогда квадратичная форма*

$$T(X) = \sum_{i=1}^{n} \mathrm{x}_i^2 - \mathrm{y}_1^2 - \ldots - \mathrm{y}_r^2$$

не зависит от случайных величин $\mathrm{y}_1, \ldots, \mathrm{y}_r$ *и имеет распределение* χ^2 *с* $n - r$ *степенями свободы:* $T(X) \in \mathbf{H}_{n-r}$.

Доказательство почти очевидно, так как после применения ортогонального преобразования C получим

$$T(X) = \sum_{i=1}^{n} y_i^2 - y_1^2 - \cdots - y_r^2 = y_{r+1}^2 + \cdots + y_n^2.$$

Остается лишь воспользоваться леммой 1. ◁

Перейдем теперь к изучению совместного распределения статистик $\overline{\mathrm{x}}$ и $S_0^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2$.

Теорема 1. *Пусть* $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$. *Тогда*

1) $(\overline{\mathrm{x}} - \alpha)\sqrt{n}/\sigma \in \mathbf{\Phi}_{0,1}$,

2) $(n-1)S_0^2/\sigma^2 \in \mathbf{H}_{n-1}$,

3) *случайные величины* $\overline{\mathrm{x}}$ *и* S_0^2 *независимы.*

Доказательство. Утверждение 1 очевидно. Далее, ясно, что, не ограничивая общности, можем считать $\alpha = 0$, $\sigma = 1$. Имеем

$$(n-1)S_0^2 = \sum_{i=1}^{n} \mathrm{x}_i^2 - n\overline{\mathrm{x}}^2.$$

Заметим, что

$$\sqrt{n}\,\overline{\mathrm{x}} = \frac{1}{\sqrt{n}}\mathrm{x}_1 + \ldots + \frac{1}{\sqrt{n}}\mathrm{x}_n$$

и что n-мерный вектор-столбец $\begin{pmatrix} 1/\sqrt{n} \\ \ldots \\ 1/\sqrt{n} \end{pmatrix}$ (его норма равна единице) всегда можно дополнить до какой-нибудь ортогональной матрицы C. Тогда $\mathrm{y}_1 = \sqrt{n}\,\overline{\mathrm{x}}$ есть первая координата $Y = XC$, и в силу леммы 2 получаем, что

$$(n-1)S_0^2 = \sum_{i=1}^{n} \mathrm{x}_i^2 - \mathrm{y}_1^2 \in \mathbf{H}_{n-1}$$

и что случайные величины $(n-1)S_0^2$ и $\mathrm{y}_1 = \sqrt{n}\,\overline{\mathrm{x}}$ независимы. ◁

Следствие 1. *Пусть* $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$. *Тогда случайная величина* $t = (\overline{\mathrm{x}} - \alpha)\sqrt{n}/S_0 \in \mathbf{T}_{n-1}$, *т. е. имеет распределение Стьюдента с* $n-1$ *степенями свободы.*

Это следует из теоремы 1 и представления

$$t = \frac{(\overline{\mathrm{x}} - \alpha)\sqrt{n}}{\sigma} \frac{1}{\sqrt{\dfrac{1}{n-1}\dfrac{(n-1)S_0^2}{\sigma^2}}}. \quad \triangleleft$$

Утверждения теоремы 1 о независимости S_0^2 и $\overline{\mathrm{x}}$ можно немного усилить. Пусть $e = (1, \ldots, 1)$ есть вектор с единичными координатами. Тогда $\overline{\mathrm{x}}$ не зависит не только от S_0^2, но и от вектора $\mathbf{X} - \overline{\mathrm{x}}e$ (т. е. от слагаемых суммы, определяющей S_0^2). Действительно, из леммы 1 и рассуждений в доказательстве теоремы 1 следует, что ортогональные векторы $\mathbf{X} - \overline{\mathrm{x}}e$ и $\overline{\mathrm{x}}e$ независимы, что означает, очевидно, независимость $\overline{x}$ и $\mathbf{X} - \overline{\mathrm{x}}e$.

2. Построение точных доверительных интервалов для параметров нормального распределения. Разберем сначала две простейшие ситуации.

а) Пусть $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$ и σ^2 известно. Требуется построить доверительный интервал уровня $1 - \varepsilon$ для параметра α. Вид доверительного интервала в этом случае очевидным образом вытекает из равенств

$$\mathbf{P}\left(\left|\frac{(\overline{\mathrm{x}} - \alpha)\sqrt{n}}{\sigma}\right| < \beta\right) = \mathbf{P}\left(-\frac{\sigma\beta}{\sqrt{n}} < \overline{\mathrm{x}} - \alpha < \frac{\sigma\beta}{\sqrt{n}}\right) = 1 - \varepsilon,$$

где, как и прежде, $\beta = \lambda_{\varepsilon/2}$, $\mathbf{\Phi}_{0,1}((-\infty, \lambda_\delta)) = 1 - \delta$, так что

$$\alpha^{\pm}(\varepsilon, X) = \overline{\mathrm{x}} \pm \frac{\sigma\beta}{\sqrt{n}}.$$

Читателю в виде упражнения предлагается воспользоваться несколько более формальной процедурой, изложенной в теореме 39.2, с использованием функции $G(\alpha, X) = (\overline{\mathrm{x}} - \alpha)\sqrt{n}/\sigma \in \mathbf{\Phi}_{0,1}$.

б) Пусть теперь известно α. Требуется построить доверительный интервал уровня $1 - \varepsilon$ для σ^2.

Положим

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i - \alpha)^2.$$

Тогда, очевидно, $nS_1^2/\sigma^2 \in \mathbf{H}_n$ и, стало быть,

$$\mathbf{P}\left(y_n^- < \frac{nS_1^2}{\sigma^2} < y_n^+\right) = \mathbf{H}_n\big((y_n^-, y_n^+)\big) = \mathbf{P}\left(\frac{nS_1^2}{y_n^+} < \sigma^2 < \frac{nS_1^2}{y_n^-}\right).$$

Таким образом, границы доверительного интервала уровня $1-\varepsilon$ будут иметь вид

$$(\sigma^2)^{\pm} = \frac{nS_1^2}{y_n^{\mp}}$$

при любых $y_n^{\pm}$ таких, что $\mathbf{H}_n\big((y_n^-, y_n^+)\big) = 1 - \varepsilon$.

Если использовать процедуру теоремы 39.2, то следует положить $G(\sigma, X) = nS_1^2/\sigma^2 \in \mathbf{H}_n$.

Перейдем теперь к случаю, когда оба параметра α и σ^2 неизвестны.

в) Для построения доверительного интервала для σ^2 воспользуемся статистикой $G_1(\sigma, X) = (n-1)S_0^2/\sigma^2$. На основании теоремы 1 $G_1(\sigma, X) \in \in \mathbf{H}_{n-1}$. Далее поступаем так же, как и в случае б). Границы доверительного интервала для σ^2 будут иметь вид

$$(\sigma^2)^{\pm} = (n-1)S_0^2/y_{n-1}^{\mp}.$$

Легко видеть, что статистики $G(\sigma, X)$ и $G_1(\sigma, X)$ в случаях б) и в) имеют одинаковое распределение и, стало быть, дают одинаковые доверительные интервалы для σ^2, если только в случае в) мы имеем на одно наблюдение больше, чем в б). Образно говоря, в случае в) мы «теряем» одно наблюдение ввиду наличия дополнительной неопределенности — неизвестного параметра α. Это наблюдение как бы предназначается для устранения «мешающего» параметра*) α.

---

*) Интересно отметить, что вопреки первоначальным интуитивным представлениям по одному наблюдению $\mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,\sigma^2}$ *возможно* построить доверительный интервал для σ^2 при неизвестном α. Следующие рассуждения, показывающие это, были сообщены нам Л. Н. Большевым.

Выберем u так, чтобы $\Phi(1/u) - \Phi(-1/u) = \varepsilon$, где $\Phi(x) = \Phi_{0,1}((-\infty, x))$. Тогда

$$\mathbf{P}(\sigma > u|\mathrm{x}_1|) = \mathbf{P}\left(-\frac{\sigma}{u} < \mathrm{x}_1 < \frac{\sigma}{u}\right) = \mathbf{P}\left(-\frac{1}{u} - \frac{\alpha}{\sigma} < \frac{(\mathrm{x}_1 - \alpha)}{\sigma} < \frac{1}{u} - \frac{\alpha}{\sigma}\right) =$$

$$= \Phi\left(\frac{1}{u} - \frac{\alpha}{\sigma}\right) - \Phi\left(-\frac{1}{u} - \frac{\alpha}{\sigma}\right) \leqslant \Phi\left(\frac{1}{u}\right) - \Phi\left(-\frac{1}{u}\right) = \varepsilon. \quad \triangleleft$$

г) Построим теперь доверительный интервал для α. Воспользуемся статистикой $G_1(\alpha, X) = (\overline{\mathrm{x}} - \alpha)\sqrt{n}/S_0$. В силу следствия теоремы 1

$$G_1(\alpha, X) \Subset \mathbf{T}_{n-1}.$$

Так как функция $G_1(\alpha, X)$ удовлетворяет условиям теоремы 39.2, то дальнейшие рассуждения в точности повторяют соответствующие рассуждения в случаях а)–в). Границы доверительного интервала имеют вид (для простоты возьмем симметричный интервал)

$$\alpha^{\pm} = \overline{\mathrm{x}} \pm \tau_\varepsilon S_0/\sqrt{n},$$

где τ_ε определяется из равенства

$$\mathbf{P}(|t_{n-1}| < \tau_\varepsilon) = \mathbf{T}_{n-1}((-\tau_\varepsilon, \tau_\varepsilon)) = 1 - \varepsilon.$$

Заметим, что если значение S_0 близко к σ, то полученный доверительный интервал будет шире, чем интервал в а), так как $\tau_\varepsilon > \beta$ (см. замечание в § 12). Это опять же объясняется наличием «мешающего» параметра σ который в а) известен.

Числа $y^{\pm}$, для которых в приведенных рассмотрениях выполнилось соотношение

$$\mathbf{P}(G(\theta, X) \Subset (y^-, y^+)) = 1 - \varepsilon,$$

на практике обычно определяются с помощью таблиц математической статистики.

В § 48 мы покажем, что доверительные интервалы, построенные в этом параграфе, в известном смысле являются наилучшими.

Глава 3
ТЕОРИЯ ПРОВЕРКИ ГИПОТЕЗ

В § 41–43, 51 излагается теория проверки конечного числа (в частности, двух) простых гипотез.

Параграфы 44–52 посвящены методам построения оптимальных критериев для проверки двух сложных гипотез. В частности, рассмотрены байесовские и минимаксные критерии (§ 44, 49) и использованы принципы достаточности, несмещенности и инвариантности для построения равномерно наиболее мощных критериев.

В § 53–57 изучаются методы построения асимптотически оптимальных критериев.

Параграф 58 посвящен устойчивости (робастности) статистических решений.

§ 41. Проверка конечного числа простых гипотез

1. Постановка задачи. Понятие статистического критерия. Наиболее мощный критерий. В этой главе речь будет идти о проверке каких-либо предположений (гипотез) относительно распределения $\mathbf{P}$, из которого извлечена выборка X. Здесь, как и в теории оценок, проблема отсутствовала бы, если бы распределение $\mathbf{P}$, из которого извлечена выборка X, было известно.

Решение о том, справедлива данная гипотеза H или нет, должно основываться лишь на знании выборки $X \in \mathbf{P}$, которая нам дана, и, возможно, также на знании априорной информации относительно $\mathbf{P}$, если таковой мы располагаем.

Таким образом, чтобы определить процедуру принятия решения по выборке X, мы должны задать в той или иной форме отображение выборочного пространства $\mathcal{X}^n$ на множество рассматриваемых гипотез. Такое отображение называют обычно статистическим критерием. Точные определения для различных конкретных ситуаций будут даны ниже.

Мы начнем с наиболее простой задачи — проверки конечного числа простых гипотез.

Определение 1. *Простой гипотезой* будем называть любое предположение, однозначно определяющее распределение выборки X.

Пусть даны r распределений $\mathbf{P}_1, \ldots, \mathbf{P}_r$ и пусть нам известно, что X есть выборка из одного из этих распределений. Задача состоит в том, чтобы определить, к какому именно $\mathbf{P}_j$, $j = 1, 2, \ldots, r$, относится X. Каждая из r гипотез

$$H_j = \{X \in \mathbf{P}_j\} \tag{1}$$

будет простой, и, стало быть, речь идет о проверке r простых гипотез.

В этой главе, как и в главе 2, мы часто будем рассматривать параметрический случай, когда выборка X извлечена из распределения $\mathbf{P}_\theta \in \mathcal{P} = \{\mathbf{P}_\theta\}_{\theta \in \Theta}$. В этом случае при выполнении условия (A_0) простые гипотезы

будут записываться в виде $H_j = \{X \Subset \mathbf{P}_{\theta_j}\}$, где $\theta_1, \ldots, \theta_r$ — фиксированные точки из Θ. Случай (1) тоже можно рассматривать как параметрический с конечным множеством $\Theta = \{1, \ldots, r\}$.

Эти рассуждения показывают, что принципиальной разницы между задачами оценки параметров и проверки гипотез нет — и там, и там определяем неизвестное значение θ. Некоторое различие все же имеется, и состоит оно в том, что в рассматриваемой задаче проверки гипотез возможные значения θ дискретны и подходы, связанные со сравнением, скажем, среднеквадратичных уклонений, развитые нами в главе 2, тут не всегда применимы. Здесь нами будут выбраны другие критерии для сравнения правил принятия тех или иных гипотез по выборке X.

С дискретностью множества возможных значений θ связано и другое новое качество, которое здесь появляется, — мы можем теперь с ненулевой вероятностью точно указать неизвестное значение θ_i (или распределение $\mathbf{P}_{\theta_i}$), в то время как в задачах оценки параметров вероятность такого события, как правило, равна нулю.

Отметим, что существуют задачи, связанные с дискретными параметрами, в которых естественным образом применимы оба подхода: с позиций теории оценивания и с позиций теории проверки гипотез. Одной из таких задач является задача о разладке, упомянутая в примере 2 введения. Эта задача (она имеет дело с параметром *случайного процесса*) будет детально рассмотрена в § 72 главы 5.

Вернемся к общей задаче о проверке конечного числа простых гипотез.

Определение 2. *Статистическим критерием* для проверки r гипотез $H_1, \ldots, H_r$ называется любое измеримое отображение δ: $\mathcal{X}^n \to \{H_1, \ldots, H_r\}$.

Другими словами, $\delta(X)$ есть случайная «величина», принимающая значения $H_1, H_2, \ldots, H_r$: если $\delta(X) = H_k$, то принимаем гипотезу H_k (т. е. считаем, что $\theta = \theta_k$ в параметрическом случае).

Отображение $\delta(\cdot)$ иногда называют *решающим правилом* или *решающей функцией*. Понятно, что задание решающего правила эквивалентно разбиению пространства $\mathcal{X}^n$ на r непересекающихся борелевских множеств $\Omega_1, \Omega_2, \ldots, \Omega_r$, на которых принимаются соответственно гипотезы H_1, $H_2, \ldots, H_r$.

Качество критерия чаще всего характеризуется набором вероятностей ошибочных решений

$$\alpha_i = \alpha_i(\delta) = \mathbf{P}_i(X \notin \Omega_i) = \mathbf{P}_i(\delta(X) \neq H_i).$$

Число α_i есть вероятность отвергнуть гипотезу H_i, когда она на самом деле верна. Это число называют *вероятностью ошибки i-го рода критерия δ*.

Если нам удалось выбрать критерий δ так, что все числа α_i малы, то в соответствии с нашим основным принципом, упомянутым в § 39, будем считать, что в одном испытании ошибка практически невозможна, и будем объявлять, что верна гипотеза H_k, если $\delta(X) = H_k$. При этом будем ошибаться примерно в доле случаев $\alpha_i = \mathbf{P}_i(\delta(X) \neq H_i)$, если на самом деле верна H_i.

Желательно, конечно, провести проверку гипотез таким образом, чтобы свести к минимуму вероятности всех ошибок. Однако если объем выборки X

задан, то мы не можем управлять всеми вероятностями ошибок одновременно. Можно лишь, фиксируя некоторые из вероятностей ошибок, пытаться минимизировать остальные.

Мы приходим здесь к вопросу о том, как сравнивать между собой различные критерии. Введем на множестве всех критериев для проверки гипотез $H_1, \ldots, H_r$ частичный порядок.

Определение 3. *Критерий* δ_1 *лучше критерия* δ_2, если для всех $i = 1, 2, \ldots, r$

$$\alpha_i(\delta_1) \leqslant \alpha_i(\delta_2)$$

и хотя бы для одного i имеет место строгое неравенство.

Однако далеко не всегда критерии δ_1 и δ_2 могут быть сравнимы в этом смысле. Точно так же, как могут быть не сравнимы две оценки θ_1^* и θ_2^* с точки зрения среднеквадратического подхода, когда в качестве критерия качества мы берем $\mathbf{E}_\theta(\theta^* - \theta)^2$. Чтобы иметь возможность сравнивать критерии, надо сузить множество рассматриваемых решающих правил. Для этого введем в рассмотрение классы

$$K_{\alpha_1, \ldots, \alpha_{r-1}} = \{\delta : \alpha_j(\delta) = \alpha_j;\ j = 1, 2, \ldots, r-1\}.$$

В классах $K_{\alpha_1, \ldots, \alpha_{r-1}}$ уже можно установить отношение порядка между критериями по величине α_r: чем меньше $\alpha_r(\delta)$, тем лучше критерий.

Определение 4. Критерий $\delta_0 \in K_{\alpha_1, \ldots, \alpha_{r-1}}$ называется *наиболее мощным критерием* (н. м. к.) *в классе* $K_{\alpha_1, \ldots, \alpha_{r-1}}$, если для любого $\delta \in K_{\alpha_1, \ldots, \alpha_{r-1}}$

$$\alpha_r(\delta_0) \leqslant \alpha_r(\delta).$$

Напомним, что нечто подобное мы делали в главе 2, когда занимались сравнением оценок. Там мы выделяли, например, классы K_b оценок с фиксированным смещением.

Наряду с только что введенным в теории проверки гипотез так же, как и в теории оценок, существуют два других подхода, которые позволяют упорядочить множество всех решающих правил с помощью одной числовой характеристики, — это байесовский и минимаксный подходы.

Прежде чем изучать методы построения наиболее мощных критериев в классах $K_{\alpha_1, \ldots, \alpha_{r-1}}$, мы рассмотрим эти два подхода.

2. Байесовский подход. Этот подход предполагает, что распределение $\mathbf{P}_j$, из которого извлечена выборка X, было выбрано случайно. В этом случае гипотезы $H_j = \{X \in \mathbf{P}_j\}$, $j = 1, \ldots, r$, становятся случайными событиями, и мы обозначим вероятности этих событий через

$$\mathbf{Q}(H_j) = q(j),$$

так что $\mathbf{Q}$ есть априорное распределение на множестве гипотез $\{H_1, \ldots, H_r\}$, а $q(j)$ есть априорные вероятности этих гипотез (ср. с § 21). В этом случае сравнивать критерии становится проще, поскольку мы можем определить

здесь среднюю вероятность $\alpha_Q(\delta)$ ошибки критерия δ:

$$\alpha_Q(\delta) = \sum_{j=1}^{r} \mathbf{Q}(H_j)\mathbf{P}_j(\delta(X) \neq H_j) = \sum_{j=1}^{r} q(j)\alpha_j(\delta) \tag{2}$$

и тем самым полностью упорядочить множество критериев по величине $\alpha_Q(\delta)$.

Определение 5. Критерий $\delta = \delta_Q$, который минимизирует вероятность ошибки $\alpha_Q(\delta)$, называется *байесовским критерием, соответствующим априорному распределению* $\mathbf{Q}$.

Пусть выполнено условие (A_μ), т. е. распределения $\mathbf{P}_j$ имеют плотности $f_j(x)$ относительно некоторой σ-конечной меры μ. Функцию $f_j(X) = \prod_{i=1}^{n} f_j(\mathrm{x}_j)$, как и прежде, будем называть *функцией правдоподобия*.

Функция $f(\boldsymbol{x}) = \sum q(j)f_j(\boldsymbol{x})$ есть безусловная плотность распределения X относительно меры μ^n, а $q(j)f_j(\boldsymbol{x})$ есть плотность совместного распределения пары (θ, X), в которой номер θ гипотезы выбирается случайно.

Таким образом, если дана выборка X, то в байесовском случае можно построить *апостериорное распределение* $\mathbf{Q}_X$ гипотез H_j (мера λ, фигурирующая в § 21, здесь — считающая мера), которое определяется по формуле Байеса

$$\mathbf{Q}_X(H_k) = q(k/X) = \frac{q(k)f_k(X)}{f(X)}. \tag{3}$$

Это есть *условное относительно* X *распределение* θ.

Через $\mathbf{E}$ будем обозначать безусловное математическое ожидание, соответствующее распределению $\mathbf{P}$ пары (θ, X).

Теорема 1. 1) *Вероятность ошибки* $\alpha_Q(\delta)$ *любого критерия* δ *удовлетворяет неравенству*

$$\alpha_Q(\delta) \geqslant 1 - \mathbf{E}\max_j q(j/X). \tag{4}$$

2) *Для того чтобы критерий* $\delta = \delta_Q$ *был байесовским для априорного распределения* $\mathbf{Q}$, *необходимо и достаточно, чтобы для* $\mathbf{P}$*-почти всех значений* X *он удовлетворял соотношениям*

$$\delta(X) = H_k, \quad \text{если} \quad q(k/X) = \max_j q(j/X). \tag{5}$$

При $\delta = \delta_Q$ *в неравенстве* (4) *достигается равенство.*

Отметим, что правая часть в (4) от δ не зависит.

Доказательство. Пусть дан критерий δ. Рассмотрим событие D_δ, состоящее в том, что критерий δ приводит к ошибочному решению:

$$D_\delta = \bigcup_{j=1}^{r} \{\theta = j,\ \delta(X) \neq H_j\}.$$

Тогда очевидно, что $\alpha_Q(\delta) = \mathbf{P}(D_\delta)$ и запись (2) есть результат последовательного усреднения: сначала по X при фиксированном $\theta = j$, потом по θ. Но мы можем $\alpha_Q(\delta)$ записать и иначе: сначала провести усреднение по θ при фиксированном X, а потом по X:

$$\alpha_Q(\delta) = \int \mathbf{P}(D_\delta/X = \boldsymbol{x}) f(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) =$$

$$= \mathbf{E}\mathbf{P}(D_\delta/X) = \mathbf{E}\sum_{j=1}^{r} \mathbf{P}(\theta = j,\ \delta(X) \neq H_j/X).$$

Но $\delta(X)$ измеримо относительно X, поэтому

$$\mathbf{P}(\theta = j,\ \delta(X) \neq H_j/X) = I_{\{\delta(X)\neq H_j\}}\mathbf{P}(\theta = j/X) = (1 - I_{\{\delta(X)=H_j\}})q(j/X).$$

Отсюда получаем

$$\alpha_Q(\delta) = 1 - \mathbf{E}\sum_{j=1}^{r} q(j/X) I_{\{\delta(X)=H_j\}} \geqslant 1 - \mathbf{E}\max_j q(j/X).$$

Первое утверждение теоремы доказано.

Достаточность второго утверждения теоремы с очевидностью вытекает из первого утверждения, так как установленная нижняя граница для $\alpha_Q(\delta)$ достигается для критерия δ_Q, определенного в (5). Изменение $\delta_Q(X)$ на множестве нулевой $\mathbf{P}$-вероятности, очевидно, $\alpha_Q(\delta_Q)$ не меняет.

Необходимость второго утверждения доказывается столь же просто. Действительно, пусть $\delta = \delta_Q$ есть байесовский критерий и пусть $\delta(X) = H_k$, $q(k/X) < q(l/X) = \max\limits_j q(j/X)$ при $X \in A$, $\mathbf{P}(A) > 0$. Тогда для критерия $\delta_1(X)$, который отличается от $\delta(X)$ лишь на множестве A: $\delta_i(X) = H_l$ при $X \in A$, получим

$$\mathbf{P}(D_{\delta_1}; A) = \mathbf{P}(A) - \mathbf{E}\left[\sum_j q(j/X) I_{\{\delta_1(X)=H_j\}}; A\right] =$$

$$= \mathbf{P}(A) - \mathbf{E}[q(l/X); A] < \mathbf{P}(A) - \mathbf{E}[q(k/X); A] = \mathbf{P}(D_\delta; A);$$

$$\mathbf{P}(D_{\delta_1}) < \mathbf{P}(D_\delta) = \mathbf{P}(D_{\delta_Q}).$$

Мы получили противоречие. $\triangleleft$

Отметим теперь, что запись (5) еще не полностью определяет критерий δ_Q: из нее не ясно, какую гипотезу следует принимать, если максимальными оказались два или более значений $q(j/x)$. Речь идет, очевидно, об определении функции $\delta_Q(X)$ на границах

$$\Gamma_k = \{\boldsymbol{x} \in \mathcal{X}^n : q(k) f_k(\boldsymbol{x}) = \max_{j\neq k} q(j) f_j(\boldsymbol{x})\}$$

множеств

$$\widetilde{\Omega}_k^Q = \{\boldsymbol{x} \in \mathcal{X}^n : q(k) f_k(\boldsymbol{x}) > \max_{j\neq k} q(j) f_j(\boldsymbol{x})\}, \tag{6}$$

на которых согласно (5) критерием δ_Q принимается гипотеза H_k.

Стало быть, $\widetilde{\Omega}_k^Q$ есть «внутренность» области

$$\Omega_k^Q = \{\boldsymbol{x} \in \mathcal{X}^n : \delta_Q(\boldsymbol{x}) = H_k\}$$

приема гипотезы H_k, и нам надо в дополнение к (6) определить лишь, какие точки границы Γ_k относятся к Ω_k^Q и какие нет. Но эта задача, как видно из приведенных рассуждений, может быть решена весьма просто: мы можем присоединять точки Γ_k к любой из «прилегающих» областей $\widetilde{\Omega}_j^Q$; при этом мы получим то же значение $\alpha_Q(\delta)$, поскольку (5) будет выполнено. Точнее, если $A \subset \Gamma_{k_1} \cap \ldots \cap \Gamma_{k_l}$, то при $X \in A$ согласно байесовскому критерию безразлично, какую из гипотез $H_{k_1}, \ldots, H_{k_l}$ принимать. Мы можем принимать решение случайно — с вероятностью p_{k_i} выбирать гипотезу H_{k_i}, $i = 1, \ldots, l$, $\sum_{i=1}^{l} p_{k_i} = 1$. Значение $\alpha_Q(\delta)$ при этом не изменится.

Мы приходим здесь к более общему понятию *рандомизированного* статистического критерия (от английского слова *random*), которое оказывается весьма полезным.

Определение 6. *Рандомизированным статистическим критерием* для проверки гипотез $H_1, \ldots, H_r$ называется любое измеримое отображение π: $\mathcal{X}^n \to \mathbb{R}^{(r)}$, где $\mathbb{R}^{(r)}$ есть множество векторов $(\pi_1, \ldots, \pi_r)$, $\pi_i \geqslant 0$, $\sum_{i=1}^{r} \pi_i = 1$.

Рандомизированный критерий каждому $\boldsymbol{x} \in \mathcal{X}^n$ ставит в соответствие распределение вероятностей $\pi(\boldsymbol{x}) = (\pi_1(\boldsymbol{x}), \ldots, \pi_r(\boldsymbol{x}))$ на множестве $\{H_1, \ldots, H_r\}$, и окончательное решение о приеме гипотезы «разыгрывается» случайным образом с этим распределением (уже независимо от X, после того как $\pi_i(X)$ определены).

Обычный статистический критерий, очевидно, есть частный случай рандомизированного, когда все π_i равны нулю и лишь одно — единице. Такие критерии будем называть *нерандомизированными.*

Вероятность ошибки i-го рода $\alpha_i(\pi)$ для рандомизированного критерия определяется аналогично:

$$\alpha_i(\pi) = \mathbf{P}_i \text{ (не принять } H_i) = 1 - \mathbf{E}_i \pi_i(X).$$

В байесовском случае задача минимизации

$$\alpha_Q(\pi) = \sum_{j=1}^{r} q(j)\alpha_j(\pi)$$

рассматривается совершенно аналогичным образом. Если через θ, как и раньше, обозначить номер случайно выбранной гипотезы с априорным распределением $\mathbf{Q}$, так что $\mathbf{Q}(\theta = j) = q(j)$, и через $\mathbf{E}$, как и прежде, символ

безусловного математического ожидания, то

$$\alpha_Q(\pi) = 1 - \sum_{j=1}^{r} q(j)\mathbf{E}_i\pi_i(X) = 1 - \mathbf{E}\pi_\theta(X) = 1 - \mathbf{E}\mathbf{E}(\pi_\theta(X)/X) =$$

$$= 1 - \mathbf{E}\sum_{j=1}^{r} q(j/X)\pi_j(X) \geqslant 1 - \mathbf{E}\max_j q(j/X).$$

Итак, мы получили ту же нижнюю границу для $\alpha_Q(\pi)$, что и для нерандомизированных критериев. Это значит, что, расширяя класс критериев, мы не можем в нашем случае улучшить значение $\alpha_Q(\delta)$. Более того, наименьшее значение достигается на нерандомизированном критерии δ_Q. Однако при этом число байесовских *рандомизированных* критериев π^Q, т. е. критериев, для которых $\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q)$, будет значительно больше, чем нерандомизированных, так как на множестве

$$\Gamma_{k_1,\dots,k_l} = \bigcap_{i=1}^{l} \Gamma_{k_i} \bigcap_{j\neq k_1,\dots,k_l} \overline{\Gamma}_j,$$

где $\overline{\Gamma} = \mathcal{X}^n \setminus \Gamma$, мы можем взять в качестве $\pi^Q(\boldsymbol{x})$ любой вектор из подмножества $R_{k_1,\dots,k_l} \subset \mathbb{R}^{(r)}$, состоящего из векторов π, у которых отличны от нуля лишь координаты с номерами $k_1,\dots,k_l$. Очевидно, что R_k состоит из единственного вектора e_k, у которого k-я координата равна единице, а остальные — нулю, и мы должны положить

$$\pi^Q(\boldsymbol{x}) = e_k \quad \text{при} \quad \boldsymbol{x} \in \widetilde{\Omega}_k^Q.$$

Поскольку приведенные соотношения с точностью до значения $\pi^Q(\boldsymbol{x})$ на множестве $\mathbf{P}$-меры 0 необходимы и достаточны для того, чтобы

$$\alpha_Q(\pi^Q) = \alpha_Q(\delta_Q) = 1 - \mathbf{E}\max_j q(j/X),$$

то мы можем наряду с теоремой 1 сформулировать следующее утверждение.

Теорема 1А. 1) *Для любого рандомизированного критерия*

$$\alpha_Q(\pi) \geqslant 1 - \mathbf{E}\max_j q(j/X).$$

2) *Для того чтобы критерий π^Q был байесовским, необходимо и достаточно выполнения соотношений*

$$\begin{aligned} \pi^Q(\boldsymbol{x}) &= e_k \qquad \textit{при} \quad \boldsymbol{x} \in \widetilde{\Omega}_k^Q, \\ \pi^Q(\boldsymbol{x}) &\in R_{k_1,\dots,k_l} \qquad \textit{при} \quad \boldsymbol{x} \in \Gamma_{k_1,\dots,k_l} \end{aligned} \tag{7}$$

для $\mathbf{P}$-почти всех значений $\boldsymbol{x}$.

3) *Для любых* $g_j \geqslant 0$, $j = 1, \ldots, r$; $\sum_{j=1}^{r} g_j = 1$ *справедливо неравенство*

$$\alpha_Q(\pi^Q) = \sum_{j=1}^{r} q(j)\alpha_j(\pi^Q) \leqslant \sum_{j=1}^{r} q(j)(1 - g_j). \tag{8}$$

Если $\min\limits_j g_j > 0$ *и не все* $f_i(\boldsymbol{x})$ *совпадают, т. е. если существуют значения* k, j *и множество* A, $\mathbf{P}(A) > 0$, *на котором* $f_k(\boldsymbol{x}) \neq f_j(\boldsymbol{x})$, *то знак в неравенстве* (8) *будет строгим.*

Замечание 1. Из (8) следует, что

$$\alpha_Q(\pi^Q) \leqslant 1 - \max_j q(j). \tag{9}$$

В правой части здесь стоит вероятность ошибки критерия, который выбирает H_k, если $q(k) = \max\limits_j q(j)$ (это есть байесовский критерий среди всех критериев, не зависящих от выборки X).

Доказательство теоремы 1А. Первые два утверждения нами уже доказаны. Чтобы доказать последнее утверждение, достаточно сравнить байесовский критерий π^Q с критерием $\pi^0(X) = g = (g_1, \ldots, g_r)$, не зависящим от X, для которого, очевидно, $\alpha_j(\pi^0) = 1 - g_j$,

$$\alpha_Q(\pi^0) = \sum_{j=1}^{r} q(j)(1 - g_j) \geqslant \alpha_Q(\pi^Q).$$

Если в (8) имеет место равенство, то критерий $\pi^0(X) = g = \text{const}$ будет байесовским. По второму утверждению теоремы это возможно лишь в случае, когда $q(1/X) = \ldots = q(r/X)$ $\mathbf{P}$-почти всюду. Это, в свою очередь, возможно, лишь когда $f_1(X) = \ldots = f_r(X)$ $\mathbf{P}$-почти всюду, $q(1) = \ldots = q(r)$. $\triangleleft$

Итак, введение рандомизированных критериев не позволяет уменьшить вероятность ошибки α_Q, но увеличивает само разнообразие критериев и, в частности, число байесовских критериев π^Q. Это обстоятельство иногда оказывается полезным.

В дальнейшем под статистическим критерием мы, как правило, будем понимать рандомизированный критерий π.

3. Минимаксный подход. Если в байесовском случае мы измеряли качество критерия средней величиной $\alpha_Q(\pi) = \sum q(j)\alpha_j(\pi)$, то теперь мы будем сравнивать максимальные значения

$$\alpha(\pi) = \max_j \alpha_j(\pi) = \max_Q \alpha_Q(\pi).$$

Очевидно, это тоже позволяет упорядочить множество всех критериев.

Определение 7. Критерий $\pi = \overline{\pi}$, для которого

$$\alpha(\overline{\pi}) = \min_\pi \alpha(\pi),$$

называется *минимаксным.*

Следующее утверждение является полным аналогом теоремы 21.2.

Теорема 2. *Пусть существует байесовский критерий* $\overline{\pi}$ (*соответствующий некоторому априорному распределению* $\overline{\mathbf{Q}}$), *для которого*

$$\alpha_1(\overline{\pi}) = \ldots = \alpha_r(\overline{\pi}). \tag{10}$$

Тогда $\overline{\pi}$ — *минимаксный критерий.*

Доказательство. Обозначим через $\overline{q}(j)$ априорные вероятности, соответствующие $\overline{\mathbf{Q}}$. Тогда для любого критерия π имеем

$$\alpha(\pi) \geqslant \sum_{j=1}^{r} \overline{q}(j)\alpha_j(\pi) \geqslant \sum_{j=1}^{r} \overline{q}(j)\alpha_j(\overline{\pi}) = \max_j \alpha_j(\overline{\pi}) = \alpha(\overline{\pi}). \quad \triangleleft$$

Распределение $\overline{\mathbf{Q}} = \{\overline{q}(j)\}$, соответствующее критерию $\overline{\pi}$, называют *наихудшим* (или *наименее благоприятным*, ср. с § 21). Это связано с тем, что при $\mathbf{Q} = \overline{\mathbf{Q}}$ достигается

$$\max_{\mathbf{Q}} \alpha_Q(\pi^Q) = \max_{\mathbf{Q}} \min_{\pi} \alpha_Q(\pi),$$

так что минимаксный критерий (10) есть байесовский критерий с самой большой вероятностью ошибиться. Доказательство этого факта можно найти в § 75. Там же будет показано, что наихудшее распределение и минимаксный критерий всегда существуют.

Следует, однако, отметить, что в отличие от байесовских критериев минимаксные *нерандомизированные* критерии существуют далеко не всегда. Дело в том, что разделяющие границы Γ_k множеств $\widetilde{\Omega}_k^Q$ (см. (6)) могут иметь ненулевую вероятность $\mathbf{P}_k(X \in \Gamma_k) > 0$ и, стало быть, значения $\alpha_k(\delta_Q)$ при непрерывном изменении $\mathbf{Q}$ могут меняться скачками. Это означает, в свою очередь, что $r-1$ уравнения $\alpha_1(\delta_Q) = \ldots = \alpha_r(\delta_Q)$ для $r-1$ неизвестного $q(1), \ldots, q(r-1)$ $\left(q(r) = 1 - \sum_{j=1}^{r-1} q(j)\right)$ могут не иметь решения. Однако в классе рандомизированных байесовских критериев минимаксный критерий существует всегда. В качестве иллюстрации подробно рассмотрим этот вопрос для случая $r = 2$ в следующем параграфе.

Итак, мы нашли явный вид байесовских критериев и установили, что с их помощью можно строить минимаксные критерии. Оказывается, что аналогичным образом можно строить и наиболее мощные критерии в классах $K_{\alpha_1,\ldots,\alpha_{r-1}}$, введенных в п. 1.

4. Наиболее мощные критерии. Определение нерандомизированного н. м. к. было дано в п. 1. Нам будет удобно здесь распространить это определение на класс рандомизированных критериев. Пусть аналогично п. 1 $K_{\alpha_1,\ldots,\alpha_{r-1}}$ означает класс *рандомизированных* критериев с фиксированными значениями вероятностей ошибок j-го рода, $j = 1, \ldots, r-1$,

$$K_{\alpha_1,\ldots,\alpha_{r-1}} = \{\pi : \alpha_j(\pi) = \alpha_j;\ j = 1, \ldots, r-1\}.$$

Определение 8. Критерий $\pi_0 \in K_{\alpha_1,\ldots,\alpha_{r-1}}$ называется *н. м. к.* в $K_{\alpha_1,\ldots,\alpha_{r-1}}$, если для любого $\pi \in K_{\alpha_1,\ldots,\alpha_{r-1}}$

$$\alpha_r(\pi_0) \leqslant \alpha_r(\pi).$$

Теорема 3. *Предположим, что существует распределение* $\mathbf{Q} = \{q(1),\ldots,q(r)\}$ *такое, что*

$$\alpha_j(\pi^Q) \equiv 1 - \mathbf{E}_j \pi_j^Q(X) = \alpha_j, \quad j = 1,\ldots,r-1 \tag{11}$$

(по существу, мы имеем здесь $r-1$ уравнение для неизвестных $q(1),\ldots,q(r-1)$). *Тогда байесовский критерий* π^Q, *определенный в* (6), (7), *будет наиболее мощным в классе* $K_{\alpha_1,\ldots,\alpha_{r-1}}$.

Доказательство. По определению байесовского критерия

$$\alpha_Q(\pi^Q) \leqslant \alpha_Q(\pi).$$

Это значит, что для $\pi \in K_{\alpha_1,\ldots,\alpha_{r-1}}$ будем иметь

$$\sum_{j=1}^{r} q(j)\alpha_j(\pi^Q) \leqslant \sum_{j=1}^{r-1} q(j)\alpha_j + q(r)\alpha_r(\pi).$$

Но $\alpha_j(\pi^Q) = \alpha_j$ при $j \leqslant r-1$ и, стало быть, $\alpha_r(\pi^Q) \leqslant \alpha_r(\pi)$. $\triangleleft$

Здесь по той же причине, что и при отыскании минимаксных критериев, уравнения (11) в классе нерандомизированных критериев δ не всегда разрешимы. В классе рандомизированных критериев положение существенно меняется. Это обстоятельство будет проиллюстрировано в следующем параграфе.

Приведем теперь пример довольно распространенной реальной задачи о проверке конечного числа простых гипотез.

Пример 1. Пусть гипотеза H_1 означает, что пациент, пришедший на осмотр к врачу, здоров, а H_k — болен некоторой определенной болезнью A_k, $k \geqslant 2$. Задача врача состоит в том, чтобы на основании наблюдений (которые можно записать в виде вектора $\mathrm{x}_1 = (\mathrm{x}_{11}, \mathrm{x}_{12}, \ldots, \mathrm{x}_{1s})$, представляющего из себя многомерную выборку X единичного объема) принять одну из гипотез H_j. Болезни A_k фиксируются нами для того, чтобы гипотезы H_k были простыми и тем самым полностью определяли распределение выборки X. Если врач принимает гипотезу H_k, $k \geqslant 2$, в то время как на самом деле верна H_1, он совершает ошибку одного типа. Если же наоборот — он больного (H_k) признает здоровым (H_1), — то совершает ошибку другого рода. Нетрудно понять, что «эффекты» от ошибок этих двух типов могут быть существенно различными.

Из результатов, изложенных выше, мы видим, что для построения наилучшего решающего правила мы должны знать распределение вектора наблюдаемых величин $(\mathrm{x}_{11},\ldots,\mathrm{x}_{1s})$ для здоровых людей и людей, больных болезнью A_k (для этого нужны обширные статистические данные медицинских обследований). Разумеется, значительная часть проблемы состоит в самом выборе s и наблюдений $(\mathrm{x}_{11}, \mathrm{x}_{12}, \ldots, \mathrm{x}_{1s})$. Во многом именно здесь проявляется искусство и опыт врачей.

Если же вектор $(x_{11}, \ldots, x_{1s})$ выбран достаточно обоснованным образом, то теоремы 1, 2 указывают нам прямой путь для алгоритмизации задач диагностики заболеваний.

§ 42. Проверка двух простых гипотез

В этом параграфе мы несколько подробнее остановимся на частном случае, когда проверяются $r = 2$ простые гипотезы.

В задачах проверки гипотез последние часто играют несимметричную роль, как это было, скажем, в примере 41.1. Поэтому одну из гипотез, например H_1, принято называть *основной* или *нулевой*, а остальные — *альтернативами*. Вероятность ошибки первого рода $\alpha_1(\delta)$ критерия δ в этом случае называют также *размером*, а число $1 - \alpha_1(\delta)$ — *уровнем* критерия. Число $\beta(\delta) = 1 - \alpha_2(\delta)$ называется *мощностью* критерия.

Область $\Omega_2 \subset \mathcal{X}^n$ приема гипотезы H_2 нерандомизированным критерием δ в случае $r = 2$ называют *критической областью*. Вероятность $\mathbf{P}_2(X \in \Omega_2)$ попадания в эту область, когда верна H_2, будет равна мощности критерия $\beta(\delta)$. Отсюда происходит название «наиболее мощный критерий» для критерия δ, на котором $\beta(\delta)$ достигает своего максимума при фиксированном уровне критерия δ.

Отметим теперь, что в случае $r = 2$ любой критерий, в том числе и рандомизированный, можно характеризовать одной числовой функцией. Действительно, произвольный рандомизированный критерий $\pi(\boldsymbol{x})$ полностью определяется значением его r координат $(\pi_1(\boldsymbol{x}), \ldots, \pi_r(\boldsymbol{x}))$. Но так как $\sum \pi_j(\boldsymbol{x}) = 1$, то в случае $r = 2$ достаточно задать одну функцию, скажем $\pi_2(\boldsymbol{x})$. Эта функция определяет вероятность принятия альтернативы H_2. Мы будем обозначать ее через $\pi(\boldsymbol{x})$ и называть *критической функцией* критерия π, который будем обозначать той же буквой π. Очевидно, что для нерандомизированных критериев $\pi(\boldsymbol{x})$ принимает лишь значения нуль и единица; в общем случае $0 \leqslant \pi(\boldsymbol{x}) \leqslant 1$.

Размер $\alpha_1(\pi)$ критерия π (или δ) и его мощность $\beta(\pi)$ выражаются через $\pi(\boldsymbol{x})$ следующим образом:

$$\alpha_1(\pi) = \mathbf{E}_1\pi(X), \quad \beta(\pi) = 1 - \alpha_2(\pi) = \mathbf{E}_2\pi(X).$$

Обозначим через Z отношение правдоподобия

$$Z = Z(\boldsymbol{x}) = \frac{f_2(\boldsymbol{x})}{f_1(\boldsymbol{x})},$$

которое будем рассматривать лишь для значений $\boldsymbol{x}$, при которых оно определено, т. е. для $\boldsymbol{x}$, при которых $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) > 0$.

Теорема 1. 1) *Пусть* $c = q(1)/q(2)$, *где* $\mathbf{Q} = (q(1), q(2))$, $q(2) = 1 - q(1)$ *есть заданное априорное распределение. Тогда критерий* $\pi_{c,p}$ *с критической функцией*

$$\pi_{c,p}(x) = \begin{cases} 1, & \text{если} \quad Z(\boldsymbol{x}) > c, \\ p(\boldsymbol{x}), & \text{если} \quad Z(\boldsymbol{x}) = c, \\ 0, & \text{если} \quad Z(\boldsymbol{x}) < c, \end{cases} \tag{1}$$

при любой измеримой функции $p(\boldsymbol{x})$, $0 \leqslant p(\boldsymbol{x}) \leqslant 1$, *является байесовским для распределения* $\mathbf{Q}$: $\pi_{c,p} = \pi^{\mathbf{Q}}$.

Параметры $\alpha_1(\pi_{c,p})$, $\alpha_2(\pi_{c,p})$ *критерия* $\pi_{c,p}$ *удовлетворяют неравенству*

$$\sum_{j=1}^{2} q(j)\alpha_j(\pi_{c,p}) \leqslant \sum_{j=1}^{2} q(j)(1-g_j) \tag{2}$$

при любых $g_j \geqslant 0$, $g_1 + g_2 = 1$.

2) *Для заданного* $\varepsilon > 0$ *такого, что* $\mathbf{P}_1(Z > 0) \geqslant \varepsilon$, *существуют* $c > 0$ *и* $p(\boldsymbol{x}) \equiv p = \mathrm{const}$ *такие, что* $\pi_{c,p} \in K_\varepsilon = \{\pi : \alpha_1(\pi) = \varepsilon\}$ *и* $\pi_{c,p}$ *является н. м. к. в* K_ε. *Числа* c *и* p *определяются как решение уравнения*

$$\alpha_1(\pi_{c,p}) \equiv \mathbf{E}_1\pi_{c,p}(X) \equiv \mathbf{P}_1(Z(X) > c) + p\mathbf{P}_1(Z(X) = c) = \varepsilon. \tag{3}$$

При этом мощность критерия $\beta(\pi_{c,p}) = 1 - \alpha_2(\pi_{c,p})$ *удовлетворяет неравенству*

$$\beta(\pi_{c,p}) \geqslant \varepsilon. \tag{4}$$

Если соотношение $f_2(\boldsymbol{x}) = f_1(\boldsymbol{x})$ *п. в.* $[\mu]$ *не выполнено (аналог условия* (A_0) *главы 2), то неравенства* (4) *и* (2) *при* $0 < q_1 < 1$ *являются строгими.*

Критерий $\pi_{c,p}$ *минимизирует вероятность ошибки первого рода* $\alpha_1(\pi)$ *в классе* K *всех критериев* π *с фиксированной вероятностью ошибки второго рода:* $K = \{\pi : \alpha_2(\pi) = \alpha_2(\pi_{c,p})\}$.

3) *Существуют* $c > 0$ *и* $p(\boldsymbol{x}) \equiv p = \mathrm{const}$ *такие, что критерий* $\pi_{c,p}$ *будем минимаксным. Числа* c *и* p *определяются из уравнения* $\alpha_1(\pi_{c,p}) = \alpha_2(\pi_{c,p})$ *или, что то же, из уравнения*

$$\mathbf{P}_1(Z(X) > c) + \mathbf{P}_2(Z(X) > c) + p[\mathbf{P}_1(Z(X) = c) + \mathbf{P}_2(Z(X) = c)] = 1. \tag{5}$$

Очевидно, что если $\mathbf{P}_1$-распределение $Z(X)$ непрерывно, т. е. $\mathbf{P}_1(Z(X) = c) = 0$ при всех $c \geqslant 0$, то в последних двух утверждениях теоремы можем положить $p \equiv 1$ или $p \equiv 0$.

Заметим также, что

$$\mathbf{P}_1(Z(X) = c) = \int\limits_{Z(\boldsymbol{x})=c} f_1(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \int\limits_{Z(\boldsymbol{x})=c} \frac{f_2(\boldsymbol{x})}{c}\mu^n(d\boldsymbol{x}) = \frac{1}{c}\mathbf{P}_2(Z(X) = c),$$

так что непрерывность на $(0, \infty)$ $\mathbf{P}_1$-распределения Z влечет за собой непрерывность $\mathbf{P}_2$-распределения Z.

Критерий $\pi_{c,p}$, основанный на отношении правдоподобия Z, называют *критерием отношения правдоподобия.*

Теорема 1 показывает, что *все оптимальные критерии являются критериями отношения правдоподобия.*

Второе утверждение теоремы 1 носит название *леммы Неймана–Пирсона.* Если в нем условие $\mathbf{P}_1(Z > 0) \geqslant \varepsilon$ не выполнено, т. е. если $\mathbf{P}_1(Z = 0) = 1 - \gamma$, $\gamma < \varepsilon$, то н. м. к. $\pi(\boldsymbol{x}) = I_{\{Z(\boldsymbol{x})>0\}}$ будет иметь мощность 1 и размер

$\gamma < \varepsilon$. Например, если носители распределений $\mathbf{P}_1$ и $\mathbf{P}_2$ не пересекаются, то $Z = 0$ на множестве, где $f_1(\boldsymbol{x}) > 0$ и, стало быть, $\mathbf{P}_1(Z > 0) = 0$. В этом случае гипотезы H_1 и H_2 различаются по одному наблюдению с вероятностями ошибок, равными нулю, т. е. различаются детерминированным образом.

Доказательство теоремы 1. Первое утверждение теоремы является прямым следствием теоремы 41.1А.

Для доказательства второго утверждения можно воспользоваться теоремой 41.3. Покажем сначала, что уравнение (3) всегда разрешимо относительно c и p. Очевидно, что функция $\varphi(c) = \mathbf{P}_1(Z > c)$ не возрастает на $[0, \infty)$. Случайная величина Z относительно распределения $\mathbf{P}_1$ является собственной, т. е.

$$\varphi(c) = \mathbf{P}_1(Z > c) = \int\limits_{Z(\boldsymbol{x})>c} f_1(\boldsymbol{x})\mu^n(d\boldsymbol{x}) < \frac{1}{c}\int\limits_{Z(\boldsymbol{x})>c} f_2(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \frac{1}{c}\mathbf{P}_2(Z > c) \to 0$$

при $c \to \infty$. Так как по условию $\varphi(0) \geqslant \varepsilon$, то найдется $c_\varepsilon \in (0, \infty)$ такое, что ($\varphi(c)$ непрерывна справа)

$$\varphi(c_\varepsilon - 0) \geqslant \varepsilon, \quad \varphi(c_\varepsilon) \leqslant \varepsilon. \tag{6}$$

Если мы положим в (3) $c = c_\varepsilon$ и обозначим $\Delta_\varepsilon = \varphi(c_\varepsilon - 0) - \varphi(c_\varepsilon)$, то получим

$$\alpha_1(\pi_{c_\varepsilon, p}) = \varphi(c_\varepsilon) + p\Delta_\varepsilon.$$

В силу (6) здесь, очевидно, всегда можно выбрать $p \in [0, 1]$ так, чтобы*) $\varphi(c_\varepsilon) + p\Delta_\varepsilon = \varepsilon$.

Теперь мы можем поступить так же, как в доказательстве теоремы 41.3. Положим $q(1) = q_\varepsilon = c_\varepsilon/(c_\varepsilon + 1)$ и фиксируем выбранное нами p. Тогда критерий $\pi_{c_\varepsilon, p}$ будет байесовским, соответствующим распределению $\mathbf{Q}_\varepsilon = (q_\varepsilon, 1 - q_\varepsilon)$, и в то же время $\alpha_1(\pi_{c_\varepsilon, p}) = \varepsilon$. Это означает в силу теоремы 41.3, что $\pi_{c_\varepsilon, p}$ является н. м. к. в K_ε.

Если взять критерий $\pi(\boldsymbol{x}) \equiv \varepsilon$, получим

$$\pi \in K_\varepsilon, \quad \alpha_2(\pi_{c_\varepsilon, p}) \leqslant \alpha_2(\pi) \equiv 1 - \varepsilon, \quad \beta(\pi_{c_\varepsilon, p}) \geqslant \varepsilon.$$

Это есть не что иное, как неравенство (2) (ср. (41.8)) при $g_2 = \varepsilon$. Стало быть, если соотношение $f_2(x) = f_1(x)$ п. в. $[\mu]$ не выполнено, то эти неравенства будут строгими. Утверждение теоремы о минимизации $\alpha_1(\pi)$ на критерии $\pi_{c,p}$ в классе $K = \{\pi : \alpha_2(\pi) = \alpha_2(\pi_{c,p})\}$ следует из приведенных выше рассуждений и симметрии относительно гипотез H_1 и H_2 постановки задачи в первом утверждении теоремы.

Для доказательства третьего утверждения теоремы 1 следует воспользоваться теоремой 41.2. Для этого нам надо проверить лишь, что уравнение

*) Ясно, что если $\varphi(c)$ непрерывна в точке c_ε, то проблема решения (3) сводится к отысканию квантиля распределения Z порядка $1 - \varepsilon$.

$\alpha_1(\pi_{c,p}) = \alpha_2(\pi_{c,p})$ разрешимо относительно c и p. Это уравнение можно записать в виде

$$\mathbf{E}_1\pi_{c,p}(X) = 1 - \mathbf{E}_2\pi_{c,p}(X)$$

или, что то же, в виде (5). Разрешимость его устанавливается точно так же, как разрешимость уравнения (3). Надо заметить лишь, что всегда $\mathbf{P}_1(Z > 0) + \mathbf{P}_2(Z > 0) \geqslant 1$, так как $\mathbf{P}_2(Z > 0) = \int\limits_{f_2(\boldsymbol{x})>0} f_2(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = 1$. $\triangleleft$

Мы видели еще раз, что цель введения рандомизированных байесовских критериев состоит в обеспечении «непрерывного» изменения параметров этих критериев (возможные значения размеров критериев $\pi_{c,p}$ заполняют весь интервал $(0, 1)$). Отсутствие такого непрерывного изменения параметров, связанное с тем, что на множестве положительной $\mathbf{P}_1$-вероятности возможно равенство $f_1(x) = cf_2(x)$, составляет основное препятствие при отыскании критериев заданного уровня или минимаксных критериев в классе нерандомизированных критериев. Эта картина полностью сохраняется и в случае проверки бóльшего числа гипотез.

Важно отметить, что два вида оптимальных критериев — наиболее мощные и минимаксные — оказываются байесовскими при тех или иных априорных распределениях. Нетрудно видеть также, что класс всех наиболее мощных критериев совпадает в известном смысле с классом всех байесовских критериев. Такое положение, при котором в качестве основы для выбора оптимальных критериев можно использовать байесовский подход, во многом сохранится и в дальнейшем.

Пример 1. Рассмотрим пример 2, приведенный во введении. В этом примере гипотезы H_1 и H_2 имеют вид $H_1 = \{\mathrm{x}_i \in F(x)\}$, $H_2 = \{\mathrm{x}_i \in \in F(x - a)\}$, где $F(x)$ — заданная функция распределения, a — данное число. Предположим, что $F(x)$ имеет плотность $f(x)$ и что случайная величина $f(\mathrm{x}_1 - a)/f(\mathrm{x}_1)$ имеет непрерывное распределение. Тогда по лемме Неймана–Пирсона (п. 2 теоремы 1) среди всех критериев уровня $1 - \varepsilon$ критерий

$$\prod_{i=1}^{n} \frac{f(\mathrm{x}_i - a)}{f(\mathrm{x}_i)} \geqslant c_\varepsilon$$

будет наиболее мощным в рассматриваемой задаче о проверке гипотезы H_1 (объект отсутствует) против гипотезы H_2 (объект присутствует). Число c_ε определяется из условия

$$\mathbf{P}_1\left(\sum_{i=1}^{n} \ln \frac{f(\mathrm{x}_i - a)}{f(\mathrm{x}_i)} > \ln c_\varepsilon\right) = \varepsilon.$$

При больших n для вычисления этой вероятности мы можем пользоваться, очевидно, центральной предельной теоремой.

§ 43.* Два асимптотических подхода к расчету критериев. Численное сравнение

1. Предварительные замечания. В § 41, 42 мы нашли форму оптимальных критериев для проверки простых гипотез. Термин «расчет критериев», который мы употребили в заголовке, будет означать вычисление параметров, характеризующих критерий. В задаче о н. м. к. в случае $r = 2$ это есть отыскание величин c_ε и p для заданного $\varepsilon > 0$ и определение вероятности ошибки второго рода $\alpha_2(\pi_{c_\varepsilon,p})$ или, что то же, мощности критерия $\beta(\pi_{c_\varepsilon,p}) = 1 - \alpha_2(\pi_{c_\varepsilon,p})$. Можно поставить вопрос и несколько иначе. Мы видели, что все оптимальные критерии в случае $r = 2$ имеют вид функций $\pi_{c,p}$, представленных в (42.1). Пусть дан критерий $\pi_{c,p}$. Как определить для него вероятности ошибок $\alpha_i(\pi_{c,p})$?

Этот же вопрос возникает, конечно, и в общем случае $r > 2$ для критерия (41.7), но мы в этом параграфе ограничимся для простоты лишь случаем двух простых гипотез.

Ниже рассматриваются асимптотические подходы, позволяющие *приближенно* (при больших n) решать эти задачи. Такие же подходы можно использовать и для расчета критериев, которые будут рассматриваться в дальнейшем.

Итак, пусть дан критерий (42.1) и пусть для простоты распределение $Z(X)$ непрерывно, так что можем положить $p \equiv 1$. Тогда критерий (42.1) станет нерандомизированным (обозначим его через δ_c), и нам требуется найти значения

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\frac{f_2(X)}{f_1(X)} \geqslant c\right), \quad \alpha_2(\delta_c) = \mathbf{P}_2\left(\frac{f_2(X)}{f_1(X)} < c\right). \tag{1}$$

Так как $f_j(X) = \prod_{i=1}^{n} f_j(\mathrm{x}_i)$, то событие, стоящее под знаком вероятности в (1), можно записать в виде

$$\sum_{i=1}^{n} \ln \frac{f_2(\mathrm{x}_i)}{f_1(\mathrm{x}_i)} \geqslant \ln c,$$

где слагаемые

$$\eta_i = \ln \frac{f_2(\mathrm{x}_i)}{f_1(\mathrm{x}_i)},$$

очевидно, есть независимые, одинаково распределенные в каждом из случаев $X \in \mathbf{P}_j$, $j = 1, 2$, случайные величины.

Таким образом, дело сводится к изучению распределений сумм $\sum_{i=1}^{n} \eta_i$ случайных величин η_i.

В дальнейшем мы будем предполагать, что объем n выборки X неограниченно возрастает. При этом под критерием мы будем понимать на самом деле *последовательность* критериев, определенных при каждом n (такое же соглашение мы использовали для оценок в главе 2).

2. Фиксированные гипотезы. В этом пункте мы будем предполагать, что распределения $\mathbf{P}_i$ фиксированы, т. е. от объема $n \to \infty$ выборки $X_n = [X_\infty]_n$ не зависят. Рассмотрим задачу расчета н. м. к. фиксированного уровня $1 - \varepsilon$. Имеем

$$\mathbf{E}_1\eta_i \equiv -a = \int f_1(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = -\rho_1(\mathbf{P}_1, \mathbf{P}_2) < 0,$$

$$\mathbf{E}_2\eta_i \equiv b = \int f_2(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx) = \rho_1(\mathbf{P}_2, \mathbf{P}_1) > 0,$$

где ρ_1 есть расстояние Кульбака–Лейблера (см. § 31).

Это означает в силу закона больших чисел, что $\mathbf{P}_1$-распределение $\dfrac{1}{n}\sum_{i=1}^{n} \eta_i$ будет сосредоточено в окрестности точки $-a$, а $\mathbf{P}_2$-распределение — в окрестности точки b. И это «разделение» распределений будет в смысле леммы Неймана–Пирсона наилучшим. Положим $\sigma_j^2 = \mathbf{D}_j\eta_1$ и предположим, что $\sigma_j^2 < \infty$. Тогда

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\sum_{i=1}^{n} \eta_i \geqslant \ln c\right) = \mathbf{P}_1\left(\frac{1}{\sigma_1\sqrt{n}} \sum_{i=1}^{n} (\eta_i + a) \geqslant \frac{\ln c + an}{\sigma_1\sqrt{n}}\right). \quad (2)$$

Выберем в качестве $c = c(n)$ любую последовательность, для которой

$$\frac{\ln c + an}{\sigma_1\sqrt{n}} \to \lambda_\varepsilon,$$

где λ_ε, как и прежде, квантиль нормального распределения уровня $1 - \varepsilon$. Тогда из (2) и центральной предельной теоремы следует

$$\alpha_1(\delta_c) \sim 1 - \Phi\left(\frac{\ln c + an}{\sigma_1\sqrt{n}}\right) \to \varepsilon. \quad (3)$$

Определение 1. Критерий π, удовлетворяющий соотношению

$$\lim_{n\to\infty} \alpha_1(\pi) = \lim_{n\to\infty} \mathbf{E}_1\pi(X) = \varepsilon,$$

называется *критерием асимптотического уровня* $1 - \varepsilon$ (или асимптотического размера ε).

Стало быть, при

$$\ln c = -an + \lambda_\varepsilon\sigma_1\sqrt{n} + o(\sqrt{n}) \quad (4)$$

критерий δ_c будет иметь асимптотический уровень $1 - \varepsilon$.

Соотношение (4) можно рассматривать как приближенное решение уравнения для числа c_ε, для которого $\alpha_1(\delta_{c_\varepsilon}) = \varepsilon$.

Положим для определенности $\ln c = -an + \lambda_\varepsilon \sigma_1 \sqrt{n}$ и найдем для выбранного c асимптотическое поведение вероятности ошибки второго рода

$$\alpha_2(\delta_c) = \mathbf{P}_2\left(\sum_{i=1}^n \eta_i < \ln c\right) = \mathbf{P}_2\left(\sum_{i=1}^n \eta_i < -an + \lambda_\varepsilon \sigma_1 \sqrt{n}\right) =$$

$$= \mathbf{P}_2\left(\frac{1}{\sigma_2\sqrt{n}} \sum_{i=1}^n (\eta_i - b) < -\frac{(a+b)\sqrt{n}}{\sigma_2} + \frac{\lambda_\varepsilon \sigma_1}{\sigma_2}\right). \quad (5)$$

Так как $-(a+b)\sqrt{n}/\sigma^2 + \lambda_\varepsilon \sigma_1/\sigma_2 \to -\infty$ при $n \to \infty$, то применение здесь центральной предельной теоремы дает нам лишь, что $\alpha_2(\delta_c) \to 0$.

Задача вычисления точного асимптотического поведения правой части в (5) *приводит нас к задаче о вероятностях больших уклонений для сумм случайных величин* η_j.

Приведем здесь результаты о вероятностях больших уклонений, изложенные в [17, § 8.8]. Пусть нам надо вычислить асимптотическое поведение $\mathbf{P}\left(\sum_{i=1}^n \xi_i > x\right)$ при $n \to \infty$, $x \to \infty$, где ξ_i независимы и одинаково распределены. Предположим, что распределение ξ_i имеет абсолютно непрерывную компоненту и что

$$\psi(\lambda) = \mathbf{E}e^{\lambda \xi_1} < \infty$$

при некоторых $\lambda > 0$. Пусть, далее,

$$\lambda_+ = \sup\{\lambda : \psi(\lambda) < \infty\},$$
$$\Lambda(\alpha) = \inf_\lambda \{-\alpha\lambda + \ln \psi(\lambda)\}, \quad (6)$$

и $\lambda(\alpha)$ есть значение λ, при котором этот $\inf\{\cdot\}$ достигается.

Тогда справедливо следующее утверждение. (См. [17, § 8.8, теоремы 12, 13, теоремы 9, 10]. Условия $\mathbf{D}\xi_i = 1$, $\mathbf{E}\xi_i = 0$, фигурирующие в этих теоремах, не играют роли.)

Теорема 1. *Пусть* $\dfrac{x - n\mathbf{E}\xi_1}{\sqrt{n}} \to \infty$ *так, что*

$$\limsup_{n\to\infty} \frac{x}{n} < \alpha_+ \equiv \frac{\psi'(\lambda_+)}{\psi(\lambda_+)}.$$

Тогда уравнение

$$\alpha\psi(\lambda) = \psi'(\lambda) \quad (7)$$

для точки $\lambda(\alpha)$ *имеет при* $\alpha < \alpha_+$ *единственное решение,*

$$\mathbf{P}\left(\sum_{i=1}^n \xi_i > x\right) = \frac{1}{\sigma(\alpha)|\lambda(\alpha)|\sqrt{2\pi n}} \exp\{-n\Lambda(\alpha)\}(1 + o(1)), \quad (8)$$

где

$$\alpha = \frac{x}{n}, \quad \sigma^2(\alpha) = \frac{\psi''(\lambda(\alpha))}{\psi(\lambda(\alpha))} - \alpha^2.$$

Кроме того, справедливы соотношения

$$\Lambda(\mathbf{E}\xi_1) = 0, \quad \Lambda'(\alpha) = \lambda(\alpha),$$

$$\Lambda''(\alpha) = \lambda'(\alpha) = \frac{\psi(\lambda(\alpha))}{\psi''(\lambda(\alpha)) - \alpha^2\psi(\lambda(\alpha))}.$$

Вернемся теперь к вычислению асимптотического поведения величины $\alpha_2(\delta_c)$, определенной в (5) и равной

$$\mathbf{P}_2\left(-\sum_{i=1}^{n}\eta_i > an - y\sqrt{n}\right) = \mathbf{P}_2\left(\sum_{i=1}^{n}(-\eta_i + b) > (a+b)n - y\sqrt{n}\right)$$

при $y = \lambda_\varepsilon\sigma_1$. Чтобы воспользоваться приведенной теоремой, надо положить

$$\xi_i = -\eta_i = \ln\frac{f_1(\mathrm{x}_i)}{f_2(\mathrm{x}_i)}, \quad x = an - y\sqrt{n}.$$

Тогда получим при $0 \leqslant \lambda \leqslant 1$

$$\psi(\lambda) = \mathbf{E}_2 e^{-\lambda\eta_1} = \int f_2(x)\left(\frac{f_1(x)}{f_2(x)}\right)^\lambda \mu(dx) =$$

$$= \int f_1^\lambda(x) f_2^{1-\lambda}(x)\mu(dx) \leqslant \left(\int f_1(x)\mu(dx)\right)^\lambda \left(\int f_2(x)\mu(dx)\right)^{1-\lambda} = 1.$$

Отсюда следует также, что $\psi(\lambda)$ будет конечным и в некоторой окрестности точки $\lambda = 1$, если

$$\int f_1(x)\left(\frac{f_1(x)}{f_2(x)}\right)^\gamma \mu(dx) < \infty \tag{9}$$

при каком-нибудь $\gamma > 0$. Далее, уравнение для точки $\lambda(\alpha)$ будет иметь вид

$$-\alpha + \frac{\psi'(\lambda)}{\psi(\lambda)} = 0$$

или

$$\psi'(\lambda) \equiv \int f_2(x)\left(\frac{f_1(x)}{f_2(x)}\right)^\lambda \ln\frac{f_1(x)}{f_2(x)}\mu(dx) = \alpha\int f_2(x)\left(\frac{f_1(x)}{f_2(x)}\right)^\lambda \mu(dx). \tag{10}$$

Если $\alpha = a = \rho_1(\mathbf{P}_1, \mathbf{P}_2) = \int f_1(x)\ln\frac{f_1(x)}{f_2(x)}\mu(dx)$, то (10) будет удовлетворено при $\lambda = 1$. Это означает, что

$$\lambda(a) = 1, \quad \psi(\lambda(a)) = \psi(1) = 1.$$

Отсюда следует, что

$$\Lambda(a) = a\lambda(a) - \ln\psi(\lambda(a)) = a,$$

$$\psi''(\lambda(a)) = \psi''(1) = \int f_1(x)\left(\ln\frac{f_1(x)}{f_2(x)}\right)^2\mu(dx),$$

$$\sigma^2(a) = \psi''(1) - a^2 = \sigma_1^2,$$

$$\Lambda'(a) = \lambda(a) = 1, \quad \Lambda''(a) = \sigma_1^{-2}.$$

Условия приведенной теоремы будут выполнены, если:

1) $\mathbf{P}_2$-распределение $\ln\dfrac{f_1(\mathrm{x}_1)}{f_2(\mathrm{x}_1)}$ имеет абсолютно непрерывную компоненту;

2) $\int f_1(x)(f_1(x)/f_2(x))^\gamma\mu(dx) < \infty$ при каком-нибудь $\gamma > 0$.

Учитывая, что в нашем случае функции $\sigma(\alpha)$, $\lambda(\alpha)$, $\Lambda''(\alpha)$ непрерывны в окрестности точки $\alpha = a$ и что $\alpha = x/n = a - y/\sqrt{n}$, мы получаем $\Lambda(\alpha) = a - y/\sqrt{n} + y^2/2\sigma_1^2 n + o(1/n)$.

Итак, мы можем сформулировать теперь следующее следствие цитированной теоремы.

Следствие 1. *Пусть выполнено условие* (9) *и* $\mathbf{P}_2$*-распределение* $\ln\dfrac{f_1(\mathrm{x}_1)}{f_2(\mathrm{x}_1)}$ *имеет абсолютно непрерывную компоненту. Тогда при* $n \to \infty$

$$\alpha_2(\delta_c) = \mathbf{P}_2\left(-\sum_{i=1}^n \eta_i > an - y\sqrt{n}\right) \sim \frac{1}{\sigma_1\sqrt{2\pi n}}\exp\left\{-na + y\sqrt{n} - \frac{y^2}{2\sigma_1^2}\right\} =$$

$$= \frac{1}{\sigma_1\sqrt{2\pi n}}\exp\left\{-n\rho_1(\mathbf{P}_1,\mathbf{P}_2) + \lambda_\varepsilon\sigma_1\sqrt{n} - \frac{\lambda_\varepsilon^2}{2}\right\}. \quad (11)$$

Таким образом, $\alpha_2(\delta_c)$ убывает при $n \to \infty$ экспоненциально*).

*) Мы получили заодно возможность дать еще одно определение (интерпретацию) расстояния Кульбака–Лейблера:

$$\rho_1(\mathbf{P}_1,\mathbf{P}_2) = -\lim_{n\to\infty}\frac{1}{n}\ln\alpha_2(\delta_c) = -\lim_{n\to\infty}\frac{1}{n}\inf_{\delta\in K_\varepsilon}\ln\alpha_2(\delta).$$

В связи с этим отметим, что тот же порядок малости $\exp\{-n\rho_1(\mathbf{P}_1,\mathbf{P}_2)\}$ имеет $\mathbf{P}_2$-вероятность того, что эмпирическая функция распределения F_n^* попадет в окрестность функции распределения F_1, соответствующей $\mathbf{P}_1$. Точнее, если $\gamma = \gamma(n) \to 0$ достаточно медленно, то

$$-\lim_{n\to\infty}\frac{1}{n}\ln\mathbf{P}_2\left(\sup_x|F_n^*(x) - F_1(x)| < \gamma\right) = \rho_1(\mathbf{P}_1,\mathbf{P}_2) \quad (12)$$

(теорема Санова). Таким образом, расстояние $\rho_1(\mathbf{P}_1,\mathbf{P}_2)$ имеет глубокий вероятностный смысл. Доказательство соотношения (12) читатель может получить, преодолев незначительные трудности, из теоремы 6 в [17, § 5.2].

Нетрудно видеть, что если мы возьмем фиксированное c в (1), то обе вероятности $\alpha_1(\delta_c)$ и $\alpha_2(\delta_c)$ будут убывать экспоненциально, как и значение $\alpha_Q(\delta_Q)$ при любом фиксированном $\mathbf{Q}$.

Так как

$$\mathbf{E}_1 e^{\lambda\eta_1} = \int f_1(x)\left(\frac{f_2(x)}{f_1(x)}\right)^{\lambda}\mu(dx) = \psi(1-\lambda),$$

$$\min_\lambda \psi(\lambda) = \min_\lambda \psi(1-\lambda),$$

то $\alpha_1(\delta_c)$, $\alpha_2(\delta_c)$ будут при этом убывать с одинаковой скоростью (зависимость от n будет одна и та же). Это означает, что минимаксный критерий будет соответствовать некоторому фиксированному c, приближенное значение которого нетрудно найти, решая уравнение $\alpha_1(\delta_c) = \alpha_2(\delta_c)$ и пользуясь асимптотическим анализом правой части (8) при $\alpha = c/n$, $n \to \infty$.

Экспоненциальное приближение (11) действует достаточно хорошо при больших n, если только нормированное уклонение

$$\frac{x + n\mathbf{E}_2\eta}{\sigma_2\sqrt{n}} = \frac{\sqrt{n}}{\sigma_2}(\rho_1(\mathbf{P}_1,\mathbf{P}_2) + \rho_1(\mathbf{P}_1,\mathbf{P}_2)) - \frac{\lambda_\varepsilon\sigma_1}{\sigma_2} \tag{13}$$

также велико (см. формулировку теоремы).

В тех прикладных задачах, где число n ограничено значениями порядка 100, это условие выполняется довольно редко, и значение (13) часто оказывается сравнимым с единицей. Это делает затруднительным использование описанного подхода к вычислению $\alpha_2(\delta_c)$ и соответствует той ситуации, когда значение $\alpha_2(\delta_c)$ вместе с $\alpha_1(\delta_c)$ не очень мало (имеет величину, сравнимую, скажем, с $0,1$). В то же время значения n порядка 100 вполне достаточны для удовлетворительного применения центральной предельной теоремы в зоне «нормальных уклонений».

Таким образом, вопрос, интересующий нас, состоит в том, когда мы можем пользоваться нормальными приближениями

$$\begin{aligned}
\alpha_1(\delta_c) &= \mathbf{P}_1\left(\sum_{i=1}^n \eta_i \geqslant \ln c\right) \approx 1 - \Phi\left(\frac{\ln c - n\mathbf{E}_1\eta_1}{\sigma_1\sqrt{n}}\right), \\
\alpha_2(\delta_c) &= \mathbf{P}_2\left(\sum_{i=1}^n \eta_i < \ln c\right) \approx \Phi\left(\frac{\ln c - n\mathbf{E}_2\eta_1}{\sigma_2\sqrt{n}}\right)
\end{aligned} \tag{14}$$

для вычисления обоих значений $\alpha_1(\delta_c)$ и $\alpha_2(\delta_c)$.

Для обоснования формул (14) возникает другой подход, основанный на предположении, что гипотезы H_1 и H_2 являются *близкими*.

3. Близкие гипотезы. Мы будем здесь рассматривать выборку X в схеме серий и предполагать, что распределения $\mathbf{P}_1$, $\mathbf{P}_2$ зависят от n так, что

$$\rho_1(\mathbf{P}_1,\mathbf{P}_2) + \rho_1(\mathbf{P}_2,\mathbf{P}_1) \to 0 \tag{15}$$

при $n \to \infty$, а последовательность (13) сходится к конечному положительному пределу.

Чтобы облегчить рассуждения и сделать их полезными для дальнейшего, мы ограничимся здесь параметрическим случаем, когда $X \in \mathbf{P}_\theta$,

$$H_1 = \{\theta = \theta_1\}, \quad H_2 = \{\theta = \theta_2\},$$

а семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям регулярности (RR) (см. § 33).

Сделаем сначала несколько неформальных замечаний, поясняющих существо дела. Мы рассматриваем близкие гипотезы, т. е. предполагаем, что $\theta_2 = \theta_1 + \gamma$, где γ мало. Тогда логарифм отношения правдоподобия, на котором построен н. м. к., представим в виде *)

$$\ln \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} \sim \gamma L'(X, \theta_1). \tag{16}$$

Статистику $U = L'(X, \theta_1)$ — главную часть в (16) — называют иногда *эффективным вкладом*. Если верна гипотеза H_1, то

$$\mathbf{E}_{\theta_1} U = 0, \quad \mathbf{D}_{\theta_1} U = nI(\theta_1).$$

Так как

$$L'(X, \theta_1) - L'(X, \theta_2) \sim \gamma L''(X, \theta_2), \quad \mathbf{E}_{\theta_2} L''(X, \theta_2) = -nI(\theta_2),$$

то

$$\mathbf{E}_{\theta_2} U \sim \gamma n I(\theta_2) \sim \gamma n I(\theta_1), \quad \mathbf{D}_{\theta_2} U \sim nI(\theta_2) \sim nI(\theta_1).$$

Это означает, что распределения U при гипотезах H_1, H_2 и при больших n будут различимы, если только величина $\mathbf{E}_{\theta_2} U - \mathbf{E}_{\theta_1} U \sim \gamma n I(\theta_1)$ будет значительно больше или сравнима с $\sqrt{\mathbf{D}_{\theta_i} U} \sim \sqrt{nI(\theta_1)}$. Другими словами, должно выполняться равенство $\gamma n = v\sqrt{n}$, $v \neq 0$, или, что то же, $\gamma = v/\sqrt{n}$.

Итак, переходя к более точному изложению, мы будем предполагать, что

$$\theta_2 = \theta_1 + \frac{v}{\sqrt{n}}, \tag{17}$$

где величины θ_1 и v будем считать фиксированными.

Положим, следуя обозначениям главы 2,

$$Z_i(t) = \frac{f_{\theta_i + t}(X)}{f_{\theta_i}(X)}, \quad Y_i(v) = \ln Z_i\left(\frac{v}{\sqrt{n}}\right).$$

Тогда

$$\sum_{i=1}^{n} \eta_i = \ln \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = Y_1(v) = -Y_2(-v). \tag{18}$$

В силу теоремы 37.3 при $X \in \mathbf{P}_{\theta_1}$ имеем

$$Y_1(v) = \xi_n v - \frac{1}{2} v^2 (I(\theta_1) + \varepsilon_n), \tag{19}$$

*) Знак $\sim$, используемый здесь, означает асимптотическую эквивалентность при $\gamma \to 0$.

где $\varepsilon_n \xrightarrow[P_{\theta_1}]{} 0$, $\xi_n I^{-1/2}(\theta_1) \Subset \Phi_{0,1}$. Аналогично при $X \in \mathbf{P}_{\theta_2}$

$$-Y_2(-v) = \xi_n v + \frac{1}{2}v^2(I(\theta_2) + \varepsilon_n),$$

где $\varepsilon_n \xrightarrow[P_{\theta_2}]{} 0$, $\xi_n I^{-1/2}(\theta_2) \Subset \Phi_{0,1}$.

Так как $I(\theta_2) \to I(\theta_1)$, получаем, что при гипотезе H_j, $j = 1, 2$,

$$\sum_{i=1}^{n} \eta_i \Rightarrow \xi|v|\sqrt{I(\theta_1)} + (-1)^j \frac{v^2}{2} I(\theta_1), \quad \xi \in \Phi_{0,1}.$$

Это означает, что из теоремы 37.3 вытекает

Следствие 2. *Пусть выполнены условия* (RR), (17). *Тогда при любом фиксированном* c *справедливы формулы* (14) *или, более точно,*

$$\begin{aligned} \alpha_1(\delta_c) &= \mathbf{P}_{\theta_1}\left(\sum_{i=1}^{n} \eta_i \geqslant \ln c\right) \to 1 - \Phi\left(\frac{\frac{v^2}{2} I(\theta_1) + \ln c}{|v|\sqrt{I(\theta_1)}}\right), \\ \alpha_2(\delta_c) &= \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n} \eta_i < \ln c\right) \to \Phi\left(\frac{-\frac{v^2}{2} I(\theta_1) + \ln c}{|v|\sqrt{I(\theta_1)}}\right). \end{aligned} \tag{20}$$

Определение 2. Критерии π_1 и π_2 называются *асимптотически эквивалентными*, если

$$\limsup_{n\to\infty} |\alpha_j(\pi_1) - \alpha_j(\pi_2)| = 0, \quad j = 1, 2.$$

Критерий π называется *асимптотически наиболее мощным критерием* (а. н. м. к.), если он асимптотически эквивалентен н. м. к.

Так как в представлениях (18), (19) $\xi_n = L'(X, \theta_1)n^{-1/2}$, то из этих представлений следует, что критерий δ с критической областью

$$\frac{vL'(X, \theta_1)}{\sqrt{nI(\theta_1)}} > vd, \quad d = \frac{v^2 I(\theta) + 2\ln c}{2|v|\sqrt{I(\theta_1)}}$$

(здесь играет роль знак v) будет иметь те же предельные значения $\alpha_i(\delta)$, что и критерий δ_c, и, следовательно, является а. н. м. к.

Кроме того, в силу результатов § 37

$$\xi_n = \frac{L'(X, \theta_1)}{\sqrt{n}} = (\widehat{\theta}^* - \theta_1)\sqrt{n}\, I(\theta_1)(1 + \varepsilon_n(X, \theta_1)),$$

$\varepsilon_n(X, \theta_1) \xrightarrow[P_{\theta_i}]{} 0$. Отсюда следует, что критерий с критической областью

$$v(\widehat{\theta}^* - \theta_1)\sqrt{nI(\theta_1)} > vd \tag{21}$$

также будет а. н. м. к.

Чтобы получить н. м. к. δ_c асимптотического уровня $1 - \varepsilon$, достаточно в (20) положить $d = \lambda_\varepsilon$. Вероятность ошибки второго рода $\alpha_2(\delta_c)$ будет сходиться к $\Phi(-v\sqrt{I(\theta_1)} + \lambda_\varepsilon)$.

При $c = 1$ оба предела в (20) будут иметь одно и то же значение

$$\lim_{n\to\infty} \alpha_j(\delta_c) = \Phi\left(-\frac{v\sqrt{I(\theta_1)}}{2}\right).$$

В этом случае критерий δ_c (ср. с теоремой 41.2) естественно называть *асимптотически минимаксным.*

4. Сравнение асимптотических подходов. Числовой пример. В п. 2, 3 мы рассмотрели два асимптотических подхода, каждый из которых оправдан в определенных условиях и которые позволяют указывать приближенные значения вероятностей ошибок первого и второго рода н. м. к. *). В случае *фиксированных гипотез* эти формулы даны в (3), (5), (11), в случае *близких гипотез* — в (14), (20). Формулы (11), (20) являются «вторичным» приближением по сравнению с (8), (14). Поэтому при возможности следует отдавать предпочтение последним (т. е. (8), (14)).

Мы уже отмечали, что для малых значений $\alpha_1(\delta)$, $\alpha_2(\delta)$ (скажем, порядка 0,01 и менее) целесообразнее использовать подход, связанный с фиксированными гипотезами. Здесь нам важно иметь достаточно хорошую *относительную* точность приближения, которая обеспечивается формулами (8) и не гарантируется центральной предельной теоремой. Если же $\alpha_1(\delta)$ и $\alpha_2(\delta)$ сравнимы с 0,1 (скажем, $\geqslant 0{,}1$), то можно рекомендовать второй подход, рассматривая данную вторую гипотезу $H_2 = \{\theta = \theta_2\}$ как элемент последовательности близких гипотез $H_{2,n} = \{\theta = \theta_1 + v/\sqrt{n}\}$, где, очевидно, надо при данных θ_1, θ_2 положить $v = \sqrt{n}\,(\theta_2 - \theta_1)$. Так как ожидаемые значения $\alpha_1(\delta)$, $\alpha_2(\delta)$ не являются очень малыми, то абсолютная величина $v/\sqrt{I(\theta_1)}$ не должна быть большой.

Пример 1. Приведем теперь числовой пример, который иллюстрирует в какой-то мере соотношение между двумя предложенными выше способами аппроксимации.

Пусть $X \in \mathbf{\Gamma}_{\theta,1}$, т. е. x_i имеют плотность

$$f_\theta(x) = \theta e^{-\theta x}, \quad x \geqslant 0,$$

*) Отметим, что наряду с предложенными двумя подходами можно рассматривать целый спектр промежуточных случаев, которые на параметрическом языке можно представить в форме (ср. с (17)) $\theta_2 = \theta_1 + zn^{-\gamma}$, $0 \leqslant \gamma \leqslant 1/2$. Такого рода близкие гипотезы представляют интерес при подборе приближенных формул, наиболее точно отражающих данную конкретную ситуацию.

и пусть основная гипотеза H_1 имеет вид $H_1 = \{\theta = 1\}$. В качестве альтернатив рассмотрим простые гипотезы $H_2^{(1)} = \{\theta = 0{,}5\}$, $H_2^{(2)} = \{\theta = 0{,}8\}$, $H_2^{(3)} = \{\theta = 0{,}9\}$.

По выборке X гипотеза H_1 будет проверяться против одной из гипотез $H_2^{(j)}$, $j = 1, 2, 3$. Таким образом, здесь $\theta_1 = 1$, а для θ_2 имеются три варианта: $\theta_2 = 0{,}5$, $\theta_2 = 0{,}8$, $\theta_2 = 0{,}9$, два последних из которых мы попытаемся рассматривать как соответствующие гипотезам, «близким» к H_1. Мы произведем расчет критериев для выборок объемов $n = 30, 100, 300, 1000$.

В нашем случае

$$\eta_i = \ln \frac{f_{\theta_2}(\mathrm{x}_i)}{f_{\theta_1}(\mathrm{x}_i)} = \ln\theta_2 - (\theta_2 - 1)\mathrm{x}_i, \tag{22}$$

$$l'(\mathrm{x}_i, \theta_1) = 1 - \mathrm{x}_i, \tag{23}$$

$$\widehat{\theta}^* = \frac{1}{\overline{\mathrm{x}}}.$$

Отсюда следует, что н. м. к. δ_c, а также оба а. н. м. к., рассмотренные выше (с критическими областями вида $\sum l'(\mathrm{x}_i, \theta_1) < d_1$ и $\widehat{\theta}^* - \theta_1 < d_1/(nI(\theta_1))$, $d_1 = d\sqrt{nI(\theta_1)}$) будут иметь форму $\delta_c(X) = H_2^{(j)}$, если

$$\sum_{i=1}^{n} (\mathrm{x}_i - 1) > d_1. \tag{24}$$

Если $X \in \mathbf{\Gamma}_{1,1}$ (гипотеза H_1), то

$$\mathbf{E}_1\mathrm{x}_1 = 1, \quad \mathbf{D}_1\mathrm{x}_1 = 1 = I(1) = \mathbf{E}_1[l'(\mathrm{x}_1, 1)]^2.$$

Стало быть, если положим $d_1 = 2\sqrt{n}$, то (ср. с (14))

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\sum_{i=1}^{n}(\mathrm{x}_i - 1) > d_1\right) =$$

$$= \mathbf{P}_1\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathrm{x}_i - 1) > 2\right) \to 1 - \Phi(2) \approx 0{,}023 \tag{25}$$

при $n \to \infty$. Так как в нашем случае $\sum_{i=1}^{n} \eta_i = n\ln\theta_2 + (1 - \theta_2)\sum_{i=1}^{n} \mathrm{x}_i$, то $\ln c$ в (14) (или в (20)) связан с d_1 соотношением

$$\ln c = n(\ln\theta_2 + 1 - \theta_2) + (1 - \theta_2)d_1.$$

Ниже мы приводим три таблицы. Во всех трех d_1 предполагается выбранным так, что выполнено (25) (т. е. $d_1 = 2\sqrt{n}$). В табл. 1 сравниваются истинные значения $\alpha_1(\delta_c)$ с приближением (25). В табл. 2 приведены

истинные значения вероятности ошибки второго рода $\alpha_2(\delta_c)$ и приближения для $\alpha_2(\delta_c)$, полученные по формулам больших уклонений (8). В табл. 3 сравниваются истинные значения $\alpha_2(\delta_c)$ с приближениями, полученными по формулам близких гипотез (14). Отметим, что мы пользуемся здесь приближениями (8), (14), а не вторичными приближениями (11), (20), которые содержат дополнительные погрешности. Все необходимые вычисления приведены ниже.

Таблица 1

Значения $\alpha_1(\delta_c)$. Верхний ряд — истинные значения, нижний — приближения (14), (25)

Объем выборки			
30	100	300	1000
0,031 0,023	0,028 0,023	0,026 0,023	0,024 0,023

Таблица 2

Значения $\alpha_2(\delta_c)$. Верхний ряд — истинные значения, нижний — приближения (8) или (26) (большие уклонения)

θ_2	Объем выборки			
	30	100	300	1000
0,5	0,028 0,033	$15 \cdot 10^{-7}$ $15 \cdot 10^{-7}$	$19 \cdot 10^{-19}$ $19 \cdot 10^{-19}$	$18 \cdot 10^{-72}$ $18 \cdot 10^{-72}$
0,8	0,71 —	0,35 —	0,028 —	$33 \cdot 10^{-8}$ $34 \cdot 10^{-8}$
0,9	0,89 —	0,79 —	0,53 —	0,085 0,11

Таблица 3

Значения $\alpha_2(\delta_c)$. Верхний ряд — истинные значения, нижний — приближения (14) (близкие гипотезы)

θ_2	Объем выборки			
	30	100	300	1000
0,5	0,028 0,041	$15 \cdot 10^{-7}$ $31 \cdot 10^{-6}$	$19 \cdot 10^{-19}$ —	$18 \cdot 10^{-72}$ —
0,8	0,71 0,69	0,35 0,35	0,028 0,031	$33 \cdot 10^{-8}$ $12 \cdot 10^{-7}$
0,9	0,89 0,89	0,79 0,79	0,53 0,52	0,085 0,086

Числа в табл. 1–3 приведены с точностью до двух значений цифр после запятой.

Сравнение табл. 2, 3 показывает, что в соответствии со сделанными выше замечаниями приближение, основанное на больших уклонениях, лучше действует в верхней правой части таблиц (где $(\theta_1 - \theta_2)\sqrt{n} = (1 - \theta_2)\sqrt{n} > 3$),

а приближение, основанное на близких гипотезах — в нижней левой части таблиц (где $(1-\theta_2)\sqrt{n} < 3$). Прочерки в таблицах проставлены там, где применение рассматриваемого подхода бессмысленно (в табл. 2, например, приближение (8) не применяется во всех случаях, когда $\alpha_2(\delta_c) > 0{,}1$). Вычисление $\alpha_2(\delta_c)$, когда это значение меньше, скажем, 10^{-6}, редко имеет практический смысл. Очень малые значения $\alpha_2(\delta_c)$ в табл. 2 при $\theta_2 = 0{,}5$, $n = 300, 1000$ подсчитаны нами для сравнения результатов вычислений.

Для завершения комментариев к таблицам нам осталось пояснить, как вычислялись истинные значения $\alpha_i(\delta_c)$, $i = 1, 2$, и во что превращаются приближения (8), (14) в нашем конкретном случае.

Значение $\alpha_2(\delta_c)$ равно

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n}(\mathrm{x}_i - 1) < 2\sqrt{n}\right).$$

Так как $\mathbf{E}_{\theta_2}\mathrm{x}_i = 1/\theta_2$, $\mathbf{D}_{\theta_2}\mathrm{x}_i = 1/\theta_2^2$, то нормальное приближение (14) для $\alpha_2(\delta_c)$ будет иметь вид

$$\Phi\left(\frac{\theta_2}{\sqrt{n}}\left[\left(1-\frac{1}{\theta_2}\right)n + 2\sqrt{n}\right]\right) = \Phi((\theta_2 - 1)\sqrt{n} + 2\theta_2).$$

Рассмотрим теперь формулу (8), в которой в нашем случае надо положить $\xi_i = \mathrm{x}_i$, $x = -n - 2\sqrt{n}$. Здесь условие теоремы 1

$$\frac{x - n\mathbf{E}\xi_1}{\sqrt{n}} = \frac{-n - 2\sqrt{n} + n/\theta_2}{\sqrt{n}} = \sqrt{n}\left(\frac{1-\theta_2}{\theta_2}\right) - 2 \to \infty$$

выполнено. Далее,

$$\psi(\lambda) = \mathbf{E}_{\theta_2}e^{-\lambda \mathrm{x}_i} = \theta_2\int_0^\infty e^{-\lambda x - \theta_2 x}\,dx = \frac{\theta_2}{\lambda + \theta_2},$$

$$\lambda_+ = \infty, \quad \alpha_+ = \lim_{\lambda\to\infty}\frac{\psi'(\lambda)}{\psi(\lambda)} = 0, \quad \alpha = \frac{x}{n} = -1 - \frac{2}{\sqrt{n}}.$$

Так как $\lim\limits_{n\to\infty}\alpha = -1 < 0$, то условие $\limsup\limits_{n\to\infty}\dfrac{x}{n} < \alpha_+$ также выполнено. Уравнение (7) имеет в нашем случае вид

$$\frac{\alpha\theta_2}{\lambda + \theta_2} = -\frac{\theta_2}{(\lambda + \theta_2)^2},$$

и его решение есть $\lambda(\alpha) = -1/\alpha - \theta_2$. Отсюда находим

$$\Lambda(\alpha) = -\ln(-\alpha\theta_2) - 1 - \alpha\theta_2, \quad \sigma^2(\alpha) = \frac{1}{\lambda'(\alpha)} = \alpha^2.$$

Таким образом, в силу (8) получаем

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n}\xi_i > x\right) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n}(\mathrm{x}_i - 1) < 2\sqrt{n}\right) \sim$$

$$\sim \frac{1}{(1+\alpha\theta_2)\sqrt{2\pi n}}\exp\{n[\ln(-\alpha\theta_2) + 1 + \alpha\theta_2]\}. \quad (26)$$

Полагая здесь $\alpha = -1 - 2/\sqrt{n}$, мы получим формулы, по которым вычислялись значения $\alpha_2(\delta_c)$ в табл. 2 (нижний ряд).

Отметим для сравнения, что правая часть (11) превращается в нашем случае в выражение

$$\frac{1}{(1-\theta_2)\sqrt{2\pi n}} \exp\{n[\ln\theta_2 + 1 - \theta_2] + 2(1-\theta_2)\sqrt{n} - 2\}, \tag{27}$$

которое можно получить из (26), подставляя туда $\alpha = -1 - 2\sqrt{n}$ и отбрасывая после разложений в ряд члены порядка $1/\sqrt{n}$ и выше.

Первый множитель в знаменателе (26), равный $\sigma(\alpha)|\lambda(\alpha)| = 1 + \alpha\theta_2 = 1 - \theta_2 - 2\theta_2/\sqrt{n}$, заменяется в (27) на $\sigma_1 = 1 - \theta_2$. Если θ_2 близко к единице, то относительная погрешность, связанная с поправочным слагаемым $-2\theta_2/\sqrt{n}$, может оказаться значительной. Например, для $\theta_2 = 0{,}8$, $n = 100$ получаем $2\theta_2/\sqrt{n} = 0{,}16$, $\sigma_1 = 1 - \theta_2 = 0{,}2$, $\sigma(\alpha)|\lambda(a)| = 0{,}2 - 0{,}16 = 0{,}04$, так что первый множитель в (27) в пять раз (!) больше, чем в (26). Этот пример показывает, что в случае близких гипотез, когда множитель σ_1 в (11) мал, приближения (11) (или (27)) надо использовать с большой осторожностью.

Для вычисления истинных значений $\alpha_i(\delta_c)$ использовался следующий факт. Пусть $\eta(t)$ есть процесс восстановления (см. [17, гл. 9]) для блуждания со скачками $\mathrm{x}_1, \mathrm{x}_2, \ldots$, т. е.

$$\eta(t) = \min\left\{k : \sum_{i=1}^{k} \mathrm{x}_i \geqslant t\right\}.$$

Тогда если $\mathrm{x}_i \in \Gamma_{\theta,1}$, то, как показано в [17, § 16.4], процесс $\xi(t) = \eta(t) - 1$ представляет собой при $t > 0$ процесс Пуассона с параметром θ, т. е.

$$\mathbf{P}(\eta(t) - 1 = k) = e^{-\theta t} = \frac{(\theta t)^k}{k!}.$$

Заметим теперь, что $\left\{\sum_{i=1}^{n} \mathrm{x}_i \geqslant t\right\} = \{\eta(t) \leqslant n\}$, и, стало быть,

$$\mathbf{P}_\theta\left(\sum_{i=1}^{n} \mathrm{x}_i \geqslant t\right) = \sum_{k=0}^{n-1} e^{-\theta t}\frac{(\theta t)^k}{k!}. \tag{28}$$

Поэтому при $t = n + 2\sqrt{n}$

$$\alpha_1(\delta_c) = \mathbf{P}_1\left(\sum_{i=1}^{n} \mathrm{x}_i \geqslant t\right) = \sum_{k=0}^{n-1} e^{-t}\frac{t^k}{k!},$$

$$\alpha_2(\delta_c) = \mathbf{P}_{\theta_2}\left(\sum_{i=1}^{n} \mathrm{x}_i < t\right) = 1 - \sum_{k=0}^{n-1} e^{-\theta_2 t}\frac{(\theta_2 t)^k}{k!}.$$

Эти равенства и использовались для вычислений точных значений $\alpha_i(\delta_c)$, $i = 1, 2$.

Отметим, что наряду с (28) можно выписать и другие формулы для распределения $\sum_{i=1}^{n} \mathrm{x}_i$, основанные на том, что $\sum_{i=1}^{n} \mathrm{x}_i \in \Gamma_{\theta,n}$.

5. Связь н. м. к. с асимптотической эффективностью о. м. п. Используя проведенные вычисления и результаты § 41, 42, мы можем доказать теперь теорему 33.7 об асимптотической эффективности о. м. п. $\widehat{\theta}^*$ в классе $\widetilde{K}^0$ асимптотически центральных оценок (принадлежность $\widehat{\theta}^* \in \widetilde{K}^0$ была установлена в п. 37.3).

Доказательство теоремы 33.7. Допустим противное, что существует а. н. оценка θ^* такая, что при каком-нибудь θ_1

$$\lim_{n\to\infty} \mathbf{E}_{\theta_1} n(\theta^* - \theta_1)^2 = \sigma^2(\theta_1) < I^{-1}(\theta_1) = \lim_{n\to\infty} \mathbf{E}_{\theta_1} n(\widehat{\theta}^* - \theta_1)^2.$$

Рассмотрим задачу проверки гипотезы $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$ против $H_2 = \{X \in \mathbf{P}_\theta,\ \theta = \theta_1 + vn^{-1/2}\}$ и построим для этого критерий δ следующего вида:

$$\delta(X) = \begin{cases} H_1, & \text{если} \quad \theta^* \leqslant \theta_1 + vn^{-1/2}, \\ H_2, & \text{если} \quad \theta^* > \theta_1 + vn^{-1/2}, \end{cases}$$

где мы приняли для определенности, что $v > 0$. Тогда

$$\alpha_1(\delta) = \mathbf{P}_{\theta_1}(\theta^* > \theta_1 + vn^{-1/2}) = \\ = \mathbf{P}_{\theta_1}\left(\frac{(\theta^* - \theta_1)\sqrt{n}}{\sigma(\theta_1)} > \frac{v}{\sigma(\theta_1)}\right) \to 1 - \Phi\left(\frac{v}{\sigma(\theta_1)}\right).$$

Далее, принадлежность $\theta^* \in \widetilde{K}^0$ означает, что

$$\alpha_2(\delta) = \mathbf{P}_\theta(\theta^* \leqslant \theta_1 + vn^{-1/2}) = \mathbf{P}_\theta(\theta^* \leqslant \theta) \to \frac{1}{2}.$$

Рассмотрим теперь другой критерий $\delta_0(X)$ с критической областью

$$\widehat{\theta}^* - \theta_1 > \frac{v + \gamma}{\sqrt{n}}, \quad \gamma > 0,$$

который, как мы установили, будет а. н. м. к. (см. (21)). Так как при достаточно малом $\gamma > 0$

$$(v + \gamma)\sqrt{I(\theta_1)} < \frac{v}{\sigma(\theta_1)},$$

то для этого критерия

$$\lim_{n\to\infty} \alpha_1(\delta_0) = 1 - \Phi((v + \gamma)\sqrt{I(\theta_1)}\,) > 1 - \Phi\left(\frac{v}{\sigma(\theta_1)}\right),$$

$$\lim_{n\to\infty} \alpha_2(\delta_0) = \lim_{n\to\infty} \mathbf{P}_\theta\left(\widehat{\theta}^* \leqslant \theta + \frac{\gamma}{\sqrt{n}}\right) > \frac{1}{2}.$$

Это означает, что начиная с некоторого n критерий δ будет лучше, чем н. м. к. Полученное противоречие доказывает теорему. $\triangleleft$

§ 44. Проверка сложных гипотез. Классы оптимальных критериев

1. Постановка задачи и основные понятия. В § 41, 42 мы рассмотрели наименее сложные задачи проверки гипотез, когда эти гипотезы являются простыми. Однако часто проверяемые гипотезы имеют более сложную природу. В параметрическом случае, например, гипотеза может иметь вид $\{X \in \mathbf{P}_\theta;\ \theta \in \Theta_1\}$, где Θ_1 — данное подмножество Θ. Такая гипотеза, очевидно, уже не определяет однозначно распределение выборки.

Всякую гипотезу H, не являющуюся простой, мы будем называть *сложной*.

Например, гипотезы $\{X \in \mathbf{\Phi}_{0,\sigma^2};\ \sigma \geqslant 1\}$, $\{X \in \mathbf{\Phi}_{\alpha,1};\ \alpha \geqslant 0\}$ являются сложными.

Везде в дальнейшем в этой главе будем рассматривать задачи о проверке *двух* гипотез, которые будем обозначать через H_1 и H_2. При этом в ближайших параграфах ограничимся изучением *параметрического случая* $X \in \mathbf{P}_\theta$, $\theta \in \Theta$. Гипотезы H_i в этом случае можно записать в виде

$$H_i = \{X \in \mathbf{P}_\theta;\ \theta \in \Theta_i\}, \quad \Theta_i \subset \Theta, \quad \Theta_1 \cap \Theta_2 = \varnothing.$$

Так как остальные значения θ, не принадлежащие $\Theta_1 \cup \Theta_2$, вообще не рассматриваются, то, не ограничивая общности, можем считать, что $\Theta = \Theta_1 \cup \Theta_2$ и что H_2 есть гипотеза, *дополнительная* (или противоположная) к H_1, так что гипотезу H_2 можно записать также в виде $H_2 = \{H_1$ *не верна*$\}$. Аналогично тому, как это делалось в § 42, одну из гипотез — в нашем рассмотрении H_1 — будем называть *основной* или *нулевой*. Простые гипотезы $H_\theta = \{X \in \mathbf{P}_\theta\}$, $\theta \in \Theta_2$, будем называть *альтернативами*.

Выделение среди гипотез основной часто отражает отношение исследователя к изучаемому объекту. Основная гипотеза обычно соответствует некоторой концепции, а альтернатива — отклонениям от нее, наличие которых надо либо доказывать, либо отвергать. Обычно имеется одна или небольшое количество основных гипотез и огромное количество мыслимых альтернатив.

Процедура приема гипотез основывается на статистическом критерии. Так как мы рассматриваем всего две гипотезы, то, как и в § 42, любой (рандомизированный) критерий π однозначно определяется измеримой функцией $\pi(\boldsymbol{x})$, $0 \leqslant \pi(\boldsymbol{x}) \leqslant 1$, которая определяет вероятность принятия $\pi(X)$ гипотезы H_2 для каждой выборки X (реализация случайного выбора с вероятностью $\pi(X)$ должна производиться с помощью независимого дополнительного устройства). Функция $\pi(\boldsymbol{x})$ называется, как и в § 42, *критической*. Для нерандомизированного критерия δ функция $\pi(\boldsymbol{x}) = \delta(\boldsymbol{x})$ принимает лишь два значения 0 и 1; область Ω_2 пространства $\mathcal{X}^n$, в которой $\delta(\boldsymbol{x}) = 1$ (область приема H_2), называется в этом случае *критической*, ее часто отождествляют с критерием δ.

Определение 1. *Размером* или *вероятностью ошибки первого рода* критерия π называется число

$$\alpha_1(\pi) = \sup_{\theta \in \Theta_1} \mathbf{E}_\theta \pi(X).$$

Для нерандомизированных критериев очевидно

$$\alpha_1(\delta) = \sup_{\theta\in\Theta_1} \mathbf{P}_\theta(X \in \Omega_2).$$

Это есть максимальная (по $\theta \in \Theta_1$) вероятность отвергнуть гипотезу H_1, когда она на самом деле верна. Обычно, чтобы облегчить поиски оптимальных критериев, рассматривают критерии π, удовлетворяющие условию

$$\alpha_1(\pi) = \varepsilon \quad (\text{или } \alpha_1(\pi) \leqslant \varepsilon).$$

Класс таких критериев обозначим через K_ε.

Число $1 - \alpha_1(\pi) = 1 - \varepsilon$ будем называть *уровнем (значимости)**) критерия π.

Статистически использование критерия $\delta \in K_\varepsilon$ означает, что в длинном ряду экспериментов по проверке гипотезы H_1 с помощью критерия $\delta \in K_\varepsilon$ будем ошибаться не чаще, чем в доле случаев ε, если на самом деле была верна гипотеза H_1.

Выбор уровня значимости критерия в значительной степени произволен. Часто в качестве ε выбирают одно из стандартных значений таких, как 0,005; 0,01; 0,05; 0,1. Эта стандартизация имеет то преимущество, что позволяет сократить объем таблиц, используемых статистиком в своей работе. Никакой другой специальной причины для выбора именно этих значений нет. Выбирая уровень значимости критерия π, нужно обращать внимание на *мощность* критерия

$$\beta_\pi(\theta) = \mathbf{E}_\theta \pi(X), \quad \theta \in \Theta_2.$$

Если она окажется слишком малой, то, возможно, следует заменить уровень $1 - \varepsilon$ на меньшее значение.

Важным обстоятельством, которое может влиять на выбор уровня значимости, является наше отношение к гипотезе до проведения эксперимента. Если мы твердо верим в истинность гипотезы (априорная вероятность $\mathbf{Q}(H_1)$ в байесовской постановке задачи велика), то потребуются убедительные свидетельства против нее для того, чтобы мы отказались от своей уверенности. В таких условиях нужны критерии высокого уровня, и ε выбирается очень малым (тогда попадание в Ω_2 будет крайне неправдоподобным, если верна H_1).

Здесь используется та же концепция, которую мы излагали при построении доверительных интервалов. Она состоит в следующем: если вероятность ε некоторого события A мала, то мы считаем практически невозможным, что это событие произойдет при единичном испытании.

Среди ряда специалистов по математической статистике существует и другая точка зрения. Она состоит в том, что нет надобности задавать фиксированный уровень значимости и что для его предварительного выбора

*) Часто уровнем значимости называют число ε, а не $1 - \varepsilon$. Но это несколько противоестественно: ведь естественнее считать, что чем выше уровень значимости, тем «значимее» критерий. Именно из этих соображений мы определяли уровень значимости (или доверия) для доверительных интервалов. Так как между статистическими критериями и доверительными интервалами существует прямая связь (см. § 48), то было бы неразумно менять терминологию при переходе к критериям на противоположную.

нет разумного правила. Они рассматривают проверку гипотез не как процедуру, ведущую обязательно к приему одной из двух гипотез, а как некоторый процесс в сознании исследователя, определяющий его отношение к гипотезам. С этой точки зрения фиксированному уровню значимости можно предпочесть «фактически достигаемый» уровень, который определяется следующим образом. Рассмотрим семейство нерандомизированных критериев δ уровня $1-\varepsilon$, когда ε пробегает значения из интервала $(0,1)$, и обозначим через $\Omega_{2,\varepsilon}$ критическую область δ, предположив, что $\Omega_{2,\varepsilon_2} \subset \Omega_{2,\varepsilon_1}$ при $\varepsilon_2 < \varepsilon_1$.

Определение 2. *Фактически достигаемым уровнем* семейства критериев δ на выборке X называется случайная величина $1-\varepsilon(X)$, где

$$\varepsilon(X) = \inf\{\varepsilon : X \in \Omega_{2,\varepsilon}\}.$$

Чем больше $1-\varepsilon(X)$, тем сильнее свидетельствует выборка против гипотезы H_1.

Значение $\varepsilon(X)$ дает возможность принимать или отвергать гипотезу при любом заранее данном уровне $1-\varepsilon$ путем простого сравнения $\varepsilon(X)$ с ε.

Пример 1. В предыдущем параграфе мы построили н. м. к. для проверки гипотезы $H_1 = \{X \in \mathbf{\Gamma}_{1,1}\}$ против гипотезы $H_2 = \{X \in \mathbf{\Gamma}_{1/2,1}\}$. Этот критерий имеет критическую область

$$\Omega_2 = \left\{ \boldsymbol{x} \in \mathcal{X}^n : \sum_{i=1}^{n} (\mathrm{x}_i - 1) > d_1 \right\}.$$

Допустим, что для выборки X объема $n = 10$ оказалось $\sum\limits_{i=1}^{10} \mathrm{x}_i = 18$. Так как при гипотезе H_1 $\sum\limits_{i=1}^{n} \mathrm{x}_i \in \mathbf{\Gamma}_{1,n}$ и $\mathbf{\Gamma}_{1,n}((a,b)) = \mathbf{H}_{2n}((2a,2b))$, то $\mathbf{\Gamma}_{1,10}((18,\infty)) = \mathbf{H}_{20}(36,\infty)) = 0{,}0154$ (см. табл. III или [10]) и фактически достигаемый уровень в этом случае будет равен $1-\varepsilon(X) = 1-0{,}0154 = 0{,}9846$, так что гипотеза H_1 будет отвергаться н. м. к. с уровнем $1-\varepsilon = 0{,}98$ и не будет отвергаться н. м. к. с уровнем $1-\varepsilon = 0{,}99$.

2. Равномерно наиболее мощные критерии. Вернемся к произвольным рандомизированным критериям π, которые мы договорились задавать критической функцией $\pi(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}^n$. (Функцию $\pi(\boldsymbol{x})$ можно называть также статистической (рандомизированной) решающей функцией.)

Если существует достаточная статистика $S(X)$, то можно ограничиться критериями $\pi(X)$, которые зависят от X только через достаточную статистику $S(X)$, т. е. критериями, представимыми в форме $\pi(X) = \varphi(S(X))$. Ведь мы знаем, что вся информация о неизвестном параметре сосредоточена в S и привлечение других статистик (другой информации о выборке X) не имеет смысла.

Как мы уже отмечали, для отыскания оптимальных критериев обычно сужают множество рассматриваемых критериев до класса K_ε критериев фиксированного уровня. Среди них можно пытаться отыскать такой критерий, для которого мощность

$$\beta_\pi(\theta) = \mathbf{E}_\theta \pi(X)$$

в области Θ_2 максимальна (т. е. вероятность ошибки второго рода $1-\beta_\pi(\theta)$ минимальна). Другими словами, максимальной должна быть вероятность принять гипотезу H_2, когда она верна.

Функция $\beta_\pi(\theta)=\mathbf{E}_\theta\pi(X)$ часто называется также *функцией мощности* критерия π.

Определение 3. Критерий $\pi^0\in K_\varepsilon$ называется *равномерно наиболее мощным критерием* (р. н. м. к.) в K_ε, если для любого $\pi\in K_\varepsilon$

$$\beta_{\pi^0}(\theta)\geqslant\beta_\pi(\theta)\quad\text{при всех}\quad\theta\in\Theta_2. \tag{1}$$

Конечно, р. н. м. к. существуют далеко не всегда. Если такой критерий π^0 существует, то для него функция мощности $\beta_{\pi^0}(\theta)$ располагается на графике выше любой другой функции $\beta_\pi(\theta)$ в области Θ_2 при условии, что обе они не превосходят значения ε в области Θ_1 (ведь $\alpha_1(\pi)=\sup\limits_{\theta\in\Theta_1}\beta_\pi(\theta)$), так что $\beta_{\pi^0}(\theta)$ есть огибающая семейства $\{\beta_\pi(\theta)\}$ в области Θ_2.

Допустим, что $\Theta_1=\{\theta_1\}$, $\mathbf{E}_{\theta_1}\pi^0(X)=\varepsilon$. Тогда р. н. м. к. π^0 будет, очевидно, н. м. к. уровня $1-\varepsilon$ для проверки гипотезы $\{\theta=\theta_1\}$ против альтернативы $\{\theta=\theta_2\}$ при любом $\theta_2\in\Theta_2$. Так как форма н. м. к. известна, то отсюда возникает следующий естественный путь отыскания р. н. м. к.: мы найдем его, если окажется, что н. м. к. в поставленной выше задаче о проверке гипотез $\{\theta=\theta_1\}$ и $\{\theta=\theta_2\}$ не зависит от θ_2.

Верно и обратное: если н. м. к. из K_ε для проверки гипотезы $\{\theta=\theta_1\}$ против $\{\theta=\theta_2\}$, $\theta_2\in\Theta_2$, существенно зависит от θ_2, это будет означать, что р. н. м. к. для проверки $\{\theta=\theta_1\}$ против $\{\theta\in\Theta_2\}$ не существует.

Если гипотеза H_2 простая (Θ_2 состоит из одной точки θ_2), то понятие р. н. м. к. частично теряет свой смысл и превращается в понятие обычного н. м. к., т. е. критерия, для которого в классе K_ε максимизируется $\mathbf{E}_{\theta_2}\pi(X)$.

Определим теперь байесовские и минимаксные критерии для проверки сложных гипотез.

3. Байесовские критерии. При проверке сложных гипотез будем различать два байесовских подхода.

1. Полный байесовский подход. Он состоит в предположении, что гипотезы $H_\theta=\{X\Subset\mathbf{P}_\theta\}$, $\theta\in\Theta$, выбираются случайно с априорным распределением $\mathbf{Q}$. Другими словами, на $\Theta=\Theta_1\cup\Theta_2$ фиксируется некоторая σ-алгебра подмножеств $\mathfrak{S}$, $\Theta_1\in\mathfrak{S}$, $\Theta_2\in\mathfrak{S}$, и θ рассматривается как случайная величина на выборочном пространстве $(\Theta,\mathfrak{S},\mathbf{Q})$.

Распределение $\mathbf{Q}$ индуцирует $\mathbf{Q}_i$ на Θ_i, $i=1,2$, и вероятности $q_i=\mathbf{Q}(\theta\in\Theta_i)$, так что $\mathbf{Q}=q_1\mathbf{Q}_1+q_2\mathbf{Q}_2$. Гипотезу о том, что $\theta\in\Theta_i$ выбирается случайным образом с распределением $\mathbf{Q}_i$, обозначим через H_{Q_i}.

Определение 4. Критерий π_Q называется *байесовским*, если он является байесовским критерием, соответствующим априорному распределению (q_1,q_2) для проверки двух простых гипотез H_{Q_1} и H_{Q_2} (см. § 41).

2. Частично байесовский подход. Здесь предполагаются заданными априорные распределения $\mathbf{Q}_i$ на Θ_i, но отсутствуют априорные веро-

ятности q_1, q_2. В этом случае имеем дело с проверкой двух простых гипотез H_{Q_1} и H_{Q_2}.

Положим, как и прежде,

$$K_\varepsilon = \{\pi : \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi(X) \leqslant \varepsilon\}, \quad K_\varepsilon^{Q_1} = \{\pi : \mathbf{E}_{Q_1}\pi(X) \leqslant \varepsilon\},$$

где $\mathbf{E}_{Q_i}$ обозначает безусловное математическое ожидание по распределению на $\Theta_i \times \mathcal{X}^n$, порожденному $\mathbf{Q}_i$ и $\mathbf{P}_\theta$.

Определение 5. Критерий π_{Q_1,Q_2} называется *байесовским в* $K_\varepsilon^{Q_1}$ если это есть н. м. к. уровня $1-\varepsilon$ для проверки двух простых гипотез H_{Q_1} и H_{Q_2}.

Если одна из гипотез H_i вырождается в простую (Θ_1 или Θ_2 одноточечно), то будет вырождаться и соответствующее распределение. В этом случае индекс у обозначения π_{Q_1,Q_2} будем укорачивать и писать, например, π_{Q_1} вместо π_{Q_1,Q_2}, если $\Theta_2 = \{\theta_2\}$ одноточечно.

Построение критериев π_{Q_1,Q_2} трудностей не вызывает. Они будут использоваться нами как вспомогательное средство для построения р. н. м. к. и минимаксных критериев.

4. Минимаксные критерии

Определение 6. Критерий $\overline{\pi}$ для проверки $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$ называется *минимаксным* в K_ε (в $K_\varepsilon^{Q_1}$), если $\overline{\pi} \in K_\varepsilon$ ($\overline{\pi} \in K_\varepsilon^{Q_1}$) и для него максимизируется

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi(X) = \inf_{\theta\in\Theta_2} \beta_\pi(\theta).$$

Такой критерий правильнее было бы называть *максиминным* (максимизируется минимум). Однако будем использовать все же единый термин «минимаксный», тем более что он сохраняет свой смысл, если иметь дело не с мощностью, а с вероятностями ошибок второго рода.

Более подробно байесовские и минимаксные критерии будут изучаться в § 49. Ближайшие же параграфы будут посвящены выяснению условий, при которых возможно построение р. н. м. к.

§ 45. Равномерно наиболее мощные критерии

В этом параграфе мы рассмотрим два важных частных случая, относящихся к одномерному параметру θ, когда р. н. м. к. удается построить. Мы получим также один полезный результат, касающийся построения н. м. к.

1. Односторонние альтернативы. Монотонное отношение правдоподобия. Пусть основная гипотеза H_1 состоит в том, что $\theta \leqslant \theta_1$, а альтернативная гипотеза H_2 в том, что $\theta > \theta_1$. Такую гипотезу H_2 будем называть *односторонней* в отличие, скажем, от гипотезы $H_2 = \{\theta \neq \theta_1\}$ (дополнительной к $H_1 = \{\theta = \theta_1\}$), которая является *двусторонней*, так как допускает отклонения от θ_1 в обе стороны.

Другое наше предположение будет состоять в следующем. Пусть выполнено условие (A_0) и существует функция $T(\boldsymbol{x})$ такая, что при всех θ, θ_0, $\theta > \theta_0$, отношение правдоподобия

$$\frac{f_\theta(\boldsymbol{x})}{f_{\theta_0}(\boldsymbol{x})} \tag{1}$$

является неубывающей (или невозрастающей) функцией от $T(\boldsymbol{x})$. В этом случае говорят, что семейство $\{\mathbf{P}_\theta\}$ имеет *монотонное отношение правдоподобия.*

Так как T — достаточная статистика, то $f_\theta(\boldsymbol{x}) = \psi(T(\boldsymbol{x}), \theta)h(\boldsymbol{x})$, и сформулированное условие будет относиться к отношению $\psi(T,\theta)/\psi(T,\theta_0)$. Это условие означает, что при всех $\theta > \theta_0$ и любом $d > 0$ неравенство $f_\theta(\boldsymbol{x})/f_{\theta_0}(\boldsymbol{x}) \geqslant d$ будет разрешимо в виде $T(\boldsymbol{x}) \geqslant c_n(\theta, \theta_0, d)$ (или $T(\boldsymbol{x}) \leqslant \leqslant c_n(\theta, \theta_0, d)$).

Например, семейства $\{\mathbf{\Phi}_{\alpha,1}\}$ и $\{\mathbf{\Phi}_{0,\sigma^2}\}$ имеют монотонное отношение правдоподобия, так как

$$\frac{f_{\alpha,1}(X)}{f_{\alpha_0,1}(X)} = \exp\left\{(\alpha - \alpha_0)n\bar{\mathrm{x}} - \frac{n}{2}(\alpha^2 - \alpha_0^2)\right\},$$

$$\frac{f_{0,\sigma^2}(X)}{f_{0,\sigma_0^2}(X)} = \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^n \mathrm{x}_i^2\right\},$$

и соответствующие неравенства будут иметь вид $(\alpha > \alpha_0,\ \sigma > \sigma_0)$

$$\bar{\mathrm{x}} \geqslant c_n(\alpha, \alpha_0, d) = \frac{1}{2}(\alpha + \alpha_0) + \frac{\ln d}{n(\alpha - \alpha_0)} \qquad (T(X) = \bar{\mathrm{x}}),$$

$$\sum_{i=1}^n \mathrm{x}_i^2 \geqslant c_n(\sigma, \sigma_0, d) = \frac{2(\sigma\sigma_0)^2}{\sigma^2 - \sigma_0^2}\ln d \qquad \left(T(X) = \sum_{i=1}^n \mathrm{x}_i^2\right).$$

Многие параметрические семейства § 12 также имеют монотонное отношение правдоподобия. В дальнейшем для определенности будем считать, что (1) есть *неубывающая* функция $T(\boldsymbol{x})$.

Теорема 1. *Пусть θ — одномерный параметр, а семейство $\{\mathbf{P}_\theta\}$ имеет монотонное отношение правдоподобия. Тогда:*

1) *в K_ε существует р. н. м. к. для проверки гипотезы $H_1 = \{\theta \leqslant \theta_1\}$ против альтернативы $H_2 = \{\theta > \theta_1\}$, который имеет вид*

$$\pi^0(X) = \begin{cases} 1, & \text{если} \quad T(X) > c, \\ p, & \text{если} \quad T(X) = c, \\ 0, & \text{если} \quad T(X) < c, \end{cases} \tag{2}$$

где c и p определяются из условия

$$\mathbf{E}_{\theta_1}\pi^0(X) = \mathbf{P}_{\theta_1}(T(X) > c) + p\mathbf{P}_{\theta_1}(T(X) = c) = \varepsilon; \tag{3}$$

2) *функция мощности* $\beta^0(\theta) = \mathbf{E}_\theta \pi^0(X)$ *строго возрастает по* θ *при всех* θ, *для которых* $\beta^0(\theta) < 1$;

3) *при всех* θ_0 *критерий* (2) *является р. н. м. к. в классе* $K_{\beta^0(\theta_0)}$ *для проверки гипотезы* $H_1^0 = \{\theta \leqslant \theta_0\}$ *против* $H_2^0 = \{\theta > \theta_0\}$;

4) *для любого* $\theta < \theta_1$ *критерий* (2) *минимизирует* $\beta(\theta) = \mathbf{E}_\theta \pi(X)$ *в классе* K_ε.

Доказательство. Рассмотрим сначала простые гипотезы $\{\theta = \theta_1\}$ и $\{\theta = \theta_2\}$, $\theta_2 > \theta_1$. Н. м. к. для проверки этих гипотез в классе критериев π, для которых $\mathbf{E}_{\theta_1} \pi(X) = \varepsilon$, имеет согласно теореме 42.1 вид (2), так как неравенство $Z(X) > d$ для отношения правдоподобия эквивалентно неравенству $T(X) > c$ (при надлежащем соответствии между c и d), где постоянные c и p определяются из (3) (ср. с (42.3)). Так как числа c и p из уравнения вида (3) определяются единственным образом, то мы получаем также, что критерий (2) будет н. м. к. в $K_{\beta^0(\theta_0)}$ для проверки гипотезы $\{\theta = \theta_0\}$ против $\{\theta = \theta_2\}$, $\theta_2 > \theta_0$. Отсюда и из теоремы 42.1 (см. (42.4)) следует, что $\beta^0(\theta_2) > \beta^0(\theta_0)$.

Так как $\beta^0(\theta)$ не убывает, то

$$\mathbf{E}_\theta \pi^0(X) \leqslant \varepsilon \quad \text{при} \quad \theta \leqslant \theta_1. \tag{4}$$

Класс K_ε критериев π, удовлетворяющих (4), содержится в классе $\{\pi\colon \mathbf{E}_{\theta_1} \pi(X) = \varepsilon\}$. Так как критерий (2) максимизирует $\beta(\theta_2)$ в этом последнем классе, он будет максимизировать $\beta(\theta_2)$ и в K_ε. Остается заметить, что критерий (2) никак не зависит от θ_2 и, стало быть, сделанные выводы справедливы при любом $\theta_2 > \theta_1$. Тем самым первые три утверждения теоремы доказаны.

Четвертое утверждение теоремы следует из первых трех, если их применить к задаче о проверке гипотезы $H_1' = \{\theta \geqslant \theta_1\}$ против $H_2' = \{\theta < \theta_1\}$, для которой р. н. м. к. в классе $\{\Pi(X) : \mathbf{E}_\theta \Pi(X) \leqslant 1-\varepsilon,\ \theta \geqslant \theta_1\}$ будет иметь вид $\Pi^0(X) = 1 - \pi^0(X)$, а функция $1 - \beta^0(\theta) = \mathbf{E}_\theta \Pi^0(X)$ станет максимальной при $\theta < \theta_1$ функцией мощности. $\triangleleft$

Важный класс семейств распределений, допускающих монотонное отношение правдоподобия, образует однопараметрическое экспоненциальное семейство (см. § 25), когда плотность $f_\theta(x)$ представима в виде

$$f_\theta(x) = h(x) \exp\{a(\theta) U(x) + V(\theta)\}. \tag{5}$$

Действительно, в этом случае

$$\frac{f_\theta(\boldsymbol{x})}{f_{\theta_0}(\boldsymbol{x})} = \exp\left\{(a(\theta) - a(\theta_0)) \sum_{i=1}^n U(x_i) + n(V(\theta) - V(\theta_0))\right\},$$

и отношение правдоподобия будет монотонно зависеть от $T(\boldsymbol{x}) = \sum_{i=1}^n U(x_i)$, если $a(\theta) - a(\theta_0)$ сохраняет знак при всех θ, θ_0, $\theta > \theta_0$.

Следствие 1. *Пусть $f_\theta(x)$ имеет вид* (5), *где $a(\theta)$ есть монотонная функция. Тогда существует р. н. м. к. π^0 в классе K_ε для проверки гипотезы $H_1 = \{\theta \leqslant \theta_1\}$ против $H_2 = \{\theta > \theta_1\}$. Если $a(\theta)$ возрастает, этот критерий имеет вид* (2), (3). *Если $a(\theta)$ убывает, неравенства в* (2), (3) *заменяются на противоположные.*

Отметим, что если проверяется двусторонняя альтернатива, например гипотеза $H_1 = \{\theta = \theta_1\}$ против $H_2 = \{\theta \neq \theta_1\}$, то р. н. м. к. для экспоненциального семейства (5) уже не существует. Действительно, допустим для простоты, что $a(\theta)$ возрастает и что $\mathbf{P}_\theta$-распределение $T(X) = \sum_{i=1}^{n} U(\mathrm{x}_i)$ при всех θ абсолютно непрерывно. Тогда согласно теореме 42.1 н. м. к. для проверки $\{\theta = \theta_1\}$ против $\{\theta = \theta_2\}$ будет нерандомизированным и будет иметь критическую область $T(X) \geqslant c$, если $\theta_2 > \theta_1$. Если же $\theta_2 < \theta_1$, то критическая область будет иметь вид $T(X) < c$. Мы видим, что максимальная мощность в точке θ_2 будет достигаться на существенно разных критериях в зависимости от знака разности $\theta_2 - \theta_1$. Из теоремы 1 видно, что если мы возьмем какой-нибудь из этих двух критериев, например тот, для которого $\pi(X) = 1$ при $T(X) \geqslant c$, то он будет р. н. м. к. для всех $\theta_2 > \theta_1$ и заведомо не будет таковым для $\theta_2 < \theta_1$.

Уже отмечалось, что положение двух простых гипотез в теореме 42.1 о н. м. к. в известном смысле симметрично (н. м. к. минимизирует вероятность ошибки второго рода $\alpha_2(\pi)$, если фиксировано значение $\alpha_1(\pi)$, и наоборот — минимизирует $\alpha_1(\pi)$, если фиксировано $\alpha_2(\pi)$). Такой симметрии в постановке задачи о проверке сложных гипотез нет. С этим обстоятельством связан следующий интересный факт. Мы только что видели, что для экспоненциального семейства не существует р. н. м. к. для проверки гипотезы $H_1 = \{\theta = \theta_1\}$ против $H_2 = \{\theta \neq \theta_1\}$. Из приведенных рассмотрений легко понять, что не существует также р. н. м. к. для проверки гипотезы $\{\theta_1 < \theta < \theta_2\}$ против альтернативы $\{\theta \notin (\theta_1, \theta_2)\}$. Однако если рассмотрим теперь в качестве основной гипотезы H_1 гипотезу $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, а в качестве альтернативы гипотезу $H_2 = \{\theta \in (\theta_1, \theta_2)\}$, то р. н. м. к. в K_ε будет уже существовать. Итак, мы рассмотрим теперь вторую возможность, когда удается построить р. н. м. к.

2. Двусторонняя основная гипотеза. Экспоненциальное семейство.

Теорема 2. *Пусть $f_\theta(x)$ определяется равенством* (5) *и проверяется гипотеза $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, $\theta_1 < \theta_2$, против альтернативы $H_2 = \{\theta \in (\theta_1, \theta_2)\}$. Тогда если функция $a(\theta)$ монотонна, то*:

1) *В классе $K_\varepsilon = \{\pi : \sup\limits_{\theta \notin (\theta_1, \theta_2)} \mathbf{E}_\theta \pi(X) \leqslant \varepsilon\}$ существует р. н. м. к. π^0, который имеет вид*

$$\pi^0(\boldsymbol{x}) = \begin{cases} 1, & \text{если} \quad c_1 < T(\boldsymbol{x}) < c_2, \\ p_i, & \text{если} \quad T(\boldsymbol{x}) = c_i, \ i = 1, 2, \\ 0, & \text{если} \quad T(\boldsymbol{x}) \notin [c_1, c_2], \end{cases} \tag{6}$$

где $T(\boldsymbol{x}) = \sum_{i=1}^{n} U(x_i)$, *а постоянные* c_i, p_i *определяются из условий*

$$\mathbf{E}_{\theta_1}\pi^0(X) = \mathbf{E}_{\theta_2}\pi^0(X) = \varepsilon. \tag{7}$$

2) *Этот критерий максимизирует функцию мощности* $\beta(\theta) = \mathbf{E}_\theta\pi(X)$ *при условии* (7) *внутри интервала* (θ_1, θ_2) *и минимизирует ее вне этого интервала* (см. рис. 4).

3) *При* $0 < \varepsilon < 1$ *функция* $\beta^0(\theta)$ *имеет максимум в некоторой точке* $\theta_0 \in (\theta_1, \theta_2)$ *и строго убывает при удалении* θ *от* θ_0 *вправо или влево. При этом мы исключаем тот случай, когда распределение* $T(X)$ *сосредоточено в двух точках, т. е. когда существуют такие* t_1, t_2, *что*

$$\mathbf{P}_\theta(T(X) = t_1) + \mathbf{P}_\theta(T(X) = t_2) = 1 \qquad \textit{при всех} \quad \theta. \tag{8}$$

В приводимых рассмотрениях полезно также следующее утверждение.

Лемма 1. *Уравнения* (7) *при* $0 < \varepsilon < 1$ *всегда разрешимы относительно* c_i, p_i, $i = 1, 2$.

Доказательство этой леммы мы приведем позже.

Доказательство теоремы 2. Запишем функцию правдоподобия в виде

$$f_\theta(\boldsymbol{x}) = c(\theta)e^{a(\theta)T(\boldsymbol{x})}h(\boldsymbol{x}), \tag{9}$$

где будем считать для определенности, что $a(\theta)$ строго возрастает.

Рассмотрим следующую байесовскую постановку задачи. Пусть проверяется основная «смешанная» гипотеза H, которая состоит в том, что $\{\theta = \theta_1\}$ с вероятностью q и $\{\theta = \theta_2\}$ с вероятностью $1-q$, против альтернативы $H_0 = \{\theta = \theta_0\}$, $\theta_0 \in (\theta_1, \theta_2)$. Пусть, далее, априорные вероятности гипотез H и H_0 равны соответственно r и $1-r$. Так как гипотезы H и H_0 полностью определяют распределение выборки, то их можно считать простыми и воспользоваться результатами § 42. Байесовский критерий (обозначим его через π^0) будет иметь в этом случае вид

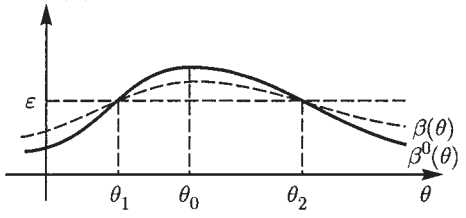

Рис. 4. Вид функции мощности $\beta^0(\theta) = \mathbf{E}_\theta\pi^0(X)$ и функции $\beta(\theta) = \mathbf{E}_\theta\pi(x)$ для произвольного критерия $\pi \in K_\varepsilon$

$$\pi^0(X) = \begin{cases} 1, & \text{если} \quad R(X) \equiv \dfrac{f_{\theta_0}(X)}{qf_{\theta_1}(X) + (1-q)f_{\theta_2}(X)} > \dfrac{r}{1-r}, \\ p(X), & \text{если} \quad R(X) = \dfrac{r}{1-r}, \\ 0, & \text{если} \quad R(X) < \dfrac{r}{1-r}. \end{cases} \tag{10}$$

Неравенство $R(X) > r/(1-r)$ в силу (9) эквивалентно неравенству

$$q\frac{c(\theta_1)}{c(\theta_0)}e^{(a(\theta_1)-a(\theta_0))T} + (1-q)\frac{c(\theta_2)}{c(\theta_0)}e^{(a(\theta_2)-a(\theta_0))T} < \frac{1-r}{r}. \tag{11}$$

Так как $a(\theta_1) - a(\theta_0) < 0$, $a(\theta_2) - a(\theta_0) > 0$, то левая часть здесь есть выпуклая функция от T. Это означает, что (11) можно записать в виде

$$c_1 < T < c_2,$$

где $c_i = c_i(q, r)$; числа $c_1 < c_2$ пробегают при изменении q и r все возможные значения. Функцию $p(X)$ в (10) положим равной p_1, если $T(X) = c_1$, и p_2, если $T(X) = c_2$.

Согласно лемме 1 найдутся c_i, $i = 1, 2$ (или, что то же, q и r), и p_i такие, что (7) будет выполнено. Покажем теперь, что функция $\pi^0(X)$, определенная в (10) или, что то же, в (6), будет обладать всеми сформулированными в теореме 2 свойствами. Сказанное означает, что мы стали теперь рассматривать π^0 одновременно как решающую функцию для проверки H_1 против H_2. Так как критерий π^0 байесовский (для проверки H против H_0), то для любого другого критерия π

$$r[q\mathbf{E}_{\theta_1}\pi^0 + (1-q)\mathbf{E}_{\theta_2}\pi^0] + (1-r)\mathbf{E}_{\theta_0}(1-\pi^0) \leqslant \\ \leqslant r[q\mathbf{E}_{\theta_1}\pi + (1-q)\mathbf{E}_{\theta_2}\pi] + (1-r)\mathbf{E}_{\theta_0}(1-\pi). \tag{12}$$

Стало быть, если критерий π наряду с π^0 будет удовлетворять (7), то

$$\mathbf{E}_{\theta_0}\pi^0 \geqslant \mathbf{E}_{\theta_0}\pi.$$

Это значит, что в каждой внутренней точке $\theta_0 \in (\theta_1, \theta_2)$ критерий π^0 максимизирует функцию мощности $\beta(\theta) = \mathbf{E}_\theta\pi$ в классе критериев π, удовлетворяющих (7). Но условия (7) выделяют класс критериев, более широкий, чем K_ε. Следовательно, π^0 будет максимизировать $\beta(\theta)$ и в K_ε. Так как критерий π^0 не зависит от θ_0, то он будет р. н. м. к. в K_ε.

Отметим еще, что в силу теоремы 42.1

$$\beta^0(\theta_0) = \mathbf{E}_{\theta_0}\pi^0 \geqslant \varepsilon$$

и равенство здесь возможно лишь в случае, когда

$$qf_{\theta_1}(\boldsymbol{x}) + (1-q)f_{\theta_2}(\boldsymbol{x}) = f_{\theta_0}(\boldsymbol{x}) \tag{13}$$

μ^n-почти всюду.

Совершенно аналогично с помощью (12) можно убедиться, что π^0 будет минимизировать $\mathbf{E}_{\theta_1}\pi$ при фиксированных $\mathbf{E}_{\theta_0}\pi$, $\mathbf{E}_{\theta_2}\pi$ (мы используем здесь те же соображения, что и в доказательстве теорем § 41).

Покажем теперь, что π^0 минимизирует $\beta(\theta)$ вне (θ_1, θ_2). Пусть $\theta^0 < \theta_1$. Заменим в предыдущих рассмотрениях тройку точек $(\theta_1, \theta_0, \theta_2)$ на тройку $(\theta^0, \theta_1, \theta_2)$ и заметим, что для новой задачи критерий π^0 снова будет байесовским (ведь его форма от выбора точек θ_i, $i = 0, 1, 2$, не зависит) в классе критериев π, для которых $\mathbf{E}_{\theta^0}\pi = \beta^0(\theta^0)$, $\mathbf{E}_{\theta_2}\pi = \varepsilon$. Но согласно сделанному

выше замечанию π^0 будет минимизировать $\mathbf{E}_{\theta^0}\pi$ при фиксированных $\mathbf{E}_{\theta_1}\pi$ и $\mathbf{E}_{\theta_2}\pi$. Первые два утверждения теоремы доказаны.

Докажем третье утверждение. Заметим предварительно, что, пользуясь заменой переменных интегрирования, мы можем записать

$$\mathbf{P}_\theta(T \in A) = c(\theta) \int\limits_{\{\boldsymbol{x}:T(\boldsymbol{x})\in A\}} e^{a(\theta)T(\boldsymbol{x})}h(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = c(\theta)\int\limits_{t\in A} e^{a(\theta)t}\nu(dt),$$

где мера ν определяется соотношением

$$\nu(A) = \int\limits_{\{\boldsymbol{x}:T(\boldsymbol{x})\in A\}} h(\boldsymbol{x})\mu^n(d\boldsymbol{x}).$$

Это означает, что распределение T относительно меры ν имеет плотность (см. также лемму 25.1) $g_\theta(t) = c(\theta)e^{a(\theta)t}$ и, стало быть, также принадлежит экспоненциальному семейству. Далее, в силу монотонности $a(\theta)$ можно ввести новый параметр $\gamma = a(\theta)$, совершенно не изменив при этом задачи и ее условий. Следовательно, мы можем считать, не ограничивая общности, что $a(\theta) = \theta$. В этом случае функции $c(\theta) = \left[\int e^{\theta t}\nu(dt)\right]^{-1}$ и $\beta^0(\theta) = \mathbf{E}_\theta\pi^0(X)$ будут, очевидно, непрерывными. Допустим теперь, что утверждение теоремы о характере поведения $\beta^0(\theta)$ неверно. Тогда найдутся три точки $\theta' < \theta'' < \theta'''$, для которых

$$\beta^0(\theta') = \beta^0(\theta'') = \beta^0(\theta''') = \alpha \in (0,1). \tag{14}$$

Мы видели, что π^0 максимизирует $\beta(\theta'')$ при условиях $\beta(\theta') = \beta(\theta'') = \alpha$, при этом если не выполнено условие вида (13), то $\beta^0(\theta'') > \alpha$. Но равенство (13) в нашем случае означает, что

$$q\frac{f_{\theta'}}{f_{\theta''}} + (1-q)\frac{f_{\theta'''}}{f_{\theta''}} \equiv q\frac{c(\theta')}{c(\theta'')}e^{(\theta'-\theta'')T} + (1-q)\frac{c(\theta''')}{c(\theta'')}e^{(\theta'''-\theta'')T} = 1$$

ν-почти всюду. В силу выпуклости левой части по T это равенство возможно не более чем для двух значений T. Стало быть, если (8) исключено, то $\beta^0(\theta'') > \beta^0(\theta') = \alpha$, и (14) невозможно. $\triangleleft$

Доказательство леммы 1. Проведем доказательство при упрощающем предположении, что распределение $T(X)$ непрерывно, т. е. $\mathbf{P}_\theta(T = c) = 0$ при всех θ и c. Это освободит нас от несущественных усложнений. В этом случае в силу замечаний, сделанных в конце доказательства теоремы 2, мы можем записать

$$\mathbf{E}_\theta\pi^0(X) = \mathbf{P}_\theta(T \in (c_1, c_2)) = \int\limits_{c_1}^{c_2} g_\theta(t)\nu(dt) = c(\theta)\int\limits_{c_1}^{c_2} e^{\theta t}\nu(dt).$$

Это будет непрерывная функция от θ, c_1, c_2.

Обозначим через c_+ значение c, для которого $\mathbf{P}_{\theta_1}(T \leqslant c_+) = 1-\varepsilon$. Тогда на $(-\infty, c_+)$ определена функция $d(c)$ такая, что

$$\mathbf{P}_{\theta_1}(T \in (c, d(c))) = \int_c^{d(c)} g_{\theta_1}(t)\nu(dt) = \varepsilon.$$

Ясно, что $d(c)$ есть непрерывная возрастающая функция.

Докажем требуемое утверждение, если убедимся, что функция

$$\psi(c) = \mathbf{P}_{\theta_2}(T \in (c, d(c))) = \int_c^{d(c)} g_{\theta_2}(t)\nu(dt)$$

непрерывно возрастает, $\psi(-\infty) < \varepsilon$, $\psi(c_+) > \varepsilon$. В этом случае найдется c_0 такое, что $\psi(c_0) = \varepsilon$, и, стало быть, $\mathbf{P}_{\theta_i}(c_0, d(c_0)) = \varepsilon$, $i = 1, 2$.

Непрерывность $\psi(c)$ очевидна. Докажем монотонность. Запишем $\psi(c)$ в виде

$$\psi(c) = \int_c^{d(c)} g_{\theta_1}(t) r(t)\nu(dt), \tag{15}$$

где $r(t)$ есть плотность $\mathbf{P}_{\theta_2}$-распределения T относительно $\mathbf{P}_{\theta_1}$-распределения:

$$r(t) = \frac{c(\theta_2)}{c(\theta_1)} e^{(\theta_2-\theta_1)t}.$$

Пусть Δ для определенности таково, что $c + \Delta < d(c)$. Тогда так как

$$\int_c^{c+\Delta} g_{\theta_1}(t)\nu(dt) = \int_{d(c)}^{d(c+\Delta)} g_{\theta_1}(t)\nu(dt), \tag{16}$$

то

$$\psi(c+\Delta) - \psi(c) = \int_{d(c)}^{d(c+\Delta)} g_{\theta_1}(t) r(t)\nu(dt) - \int_c^{c+\Delta} g_{\theta_1}(t) r(t)\nu(dt) \geqslant$$

$$\geqslant [r(d(c)) - r(c+\Delta)]\lambda \geqslant 0,$$

где λ есть общее значение интеграла (16).

Убедимся теперь, что $\psi(-\infty) < \varepsilon$. Обозначим через t_0 решение уравнения $r(t) = 1$. Если $d(-\infty) \leqslant t_0$, то $r(t) < 1$ на интервале $(-\infty, d(-\infty))$, и требуемое неравенство в силу (15) очевидно. Если же $d(-\infty) > t_0$, то аналогично находим

$$\psi(-\infty) = 1 - \mathbf{P}_{\theta_2}(T \in (d(-\infty), \infty)) <$$

$$< 1 - \mathbf{P}_{\theta_1}(T \in (d(-\infty), \infty)) = \mathbf{P}_{\theta_1}(T \in (-\infty, d(-\infty))) = \varepsilon.$$

Точно так же устанавливается, что $\psi(c_+) > \varepsilon$. $\triangleleft$

Замечание 1. Мы предоставляем читателю самому убедиться в том, что при $\theta_1 < \theta_2$ утверждение теоремы 2 и все проведенные рассмотрения сохранятся, если мы интервал (θ_1, θ_2) заменим на отрезок $[\theta_1, \theta_2]$, т. е. будем проверять гипотезу $H_1 = \{\theta \notin [\theta_1, \theta_2]\}$ против $H_2 = \{\theta \in [\theta_1, \theta_2]\}$.

Замечание 2. Требование экспоненциальности семейства $\{\mathbf{P}_\theta\}$, как это видно из доказательства теоремы, может быть ослаблено до условия выпуклости отношения

$$q\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} + (1-q)\frac{f_{\theta_2}(X)}{f_{\theta_0}(X)}$$

относительно некоторой статистики T (ср. с (10), (11)).

Замечание 3. Обратим внимание еще раз на то, что если бы основной гипотезой была $H_2 = \{\theta \in (\theta_1, \theta_2)\}$, а альтернативой $H_1 = \{\theta \notin (\theta_1, \theta_2)\}$, то р. н. м. к. не существовало бы, так как в этом случае «односторонние» критерии вида $T > c$ или $T < c$ соответственно для альтернатив $\theta > \theta_2$ и $\theta < \theta_1$ оказались бы более мощными, чем критерий вида $T \notin (c_1, c_2)$. Например, для альтернатив $\theta > \theta_2$ будет существовать р. н. м. к. вида $T > c$, а условие $\pi \in K_\varepsilon$ приведет к единственному ограничению $\mathbf{E}_{\theta_2}\pi \leqslant \varepsilon$ (см. замечания в конце п. 4).

Тем не менее окажется, что если класс K_ε дополнительно несколько сузить вполне естественным образом (см. § 46, 47), то р. н. м. к. будет существовать и в этой задаче.

3. Другой подход к рассматриваемым задачам. Математическое существо главного утверждения теоремы 2, а также теорем § 41, 42 весьма просто и заслуживает того, чтобы его выделить. Например, в теореме 2 оно состоит в следующей вариационной задаче. В классе функций π, удовлетворяющих условиям

$$\int \pi(\boldsymbol{x}) f_{\theta_i}(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \varepsilon, \quad i = 1, 2,$$

мы ищем элемент π^0, для которого максимизируется

$$\int \pi(\boldsymbol{x}) f_{\theta_0}(\boldsymbol{x}) \mu^n(d\boldsymbol{x}).$$

Следующее утверждение называется обычно *обобщением фундаментальной леммы Неймана–Пирсона.*

Лемма 2. *Пусть $f_1, \ldots, f_{m+1}$ — действительные функции, определенные на $\mathcal{X}^n$ и интегрируемые относительно меры μ^n. Пусть критические функции π таковы, что*

$$\int \pi(\boldsymbol{x}) f_i(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \varepsilon_i, \quad i = 1, \ldots, m. \tag{17}$$

Тогда элемент π^0, *на котором* $\int \pi(\boldsymbol{x}) f_{m+1}(\boldsymbol{x}) \mu^n(d\boldsymbol{x})$ *достигает максимума, имеет вид*

$$\pi^0(\boldsymbol{x}) = \begin{cases} 1, & \text{если} \quad f_{m+1}(\boldsymbol{x}) > \sum_{i=1}^{m} k_i f_i(\boldsymbol{x}), \\ 0, & \text{если} \quad f_{m+1}(\boldsymbol{x}) < \sum_{i=1}^{m} k_i f_i(\boldsymbol{x}), \end{cases}$$

где $k_1, \ldots, k_m$ *определяются из условий* (17).

Доказательство. Положим $F_i(\pi) = \int \pi(\boldsymbol{x}) f_i(\boldsymbol{x}) \mu^n(d\boldsymbol{x})$, $i = 1, \ldots, m+1$. Элемент π, удовлетворяющий условиям $F_i(\pi) = \varepsilon_i$, $i = 1, \ldots, m$, максимизирует $F_{m+1}(\pi)$ тогда и только тогда, кода он максимизирует $F_{m+1}(\pi) - \sum_{i=1}^{m} k_i F_i(\pi)$ при каких-нибудь $k_1, \ldots, k_m$ (ведь значение суммы здесь фиксировано). Стало быть, нам достаточно, чтобы π максимизировало

$$\int \left(f_{m+1}(\boldsymbol{x}) - \sum_{i=1}^{m} k_i f_i(\boldsymbol{x}) \right) \pi(\boldsymbol{x}) \mu^n(d\boldsymbol{x}).$$

Но это выражение делается максимальным, если положить $\pi(\boldsymbol{x}) = 1$ там, где $f_{m+1}(\boldsymbol{x}) - \sum_{i=1}^{m} k_i f_i(\boldsymbol{x}) > 0$, и $\pi(\boldsymbol{x}) = 0$ там, где это выражение отрицательно. Постоянные k_i, от которых зависит это π, а также «свободные» значения π на множестве $\left\{ f_{m+1}(\boldsymbol{x}) = \sum_{i=1}^{m} k_i f_i(\boldsymbol{x}) \right\}$ следует подобрать так, чтобы выполнялось (17). $\triangleleft$

4. Байесовский подход и наименее благоприятные априорные распределения при построении н. м. к. и р. н. м. к. Лемма 2 выясняет математическое существо построений, которые проводились нами в этом параграфе. В этом пункте речь будет идти также о существе этих рассмотрений, но с несколько иной точки зрения. Дело в том, что при доказательстве теоремы 2 мы неявно использовали подход, связанный с построением минимаксных критериев с помощью байесовских (ср. с теоремой 41.2). Этот подход позже будет рассматриваться более подробно. Здесь же мы получим одно общее утверждение, полезное при построении р. н. м. к. в общем случае, и поясним его связь с минимаксным подходом.

Пусть проверяется основная гипотеза $H_1 = \{\theta \in \Theta_1\}$ против простой альтернативы $H_2 = \{\theta = \theta_2\}$, $\theta_2 \notin \Theta_1$. В качестве H_2 здесь можно взять и произвольную альтернативу $\{X \in \mathbf{G}\}$, где $\mathbf{G}$ имеет плотность g относительно μ и с семейством $\{\mathbf{P}_\theta\}$ никак не связано. Задача состоит в отыскании

н. м. к. π уровня $1-\varepsilon$ для проверки H_1 против H_2. Другими словами, надо найти функцию π из K_ε,

$$K_\varepsilon = \{\pi : \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi(X) \leqslant \varepsilon\}, \tag{18}$$

которая минимизирует $\beta(\theta_2) = \mathbf{E}_{\theta_2}\pi(X)$. В предыдущих рассмотрениях мы не раз наблюдали известную «дуальность» в постановке задачи: максимизация мощности при фиксированной вероятности ошибки первого рода эквивалентна минимизации последней при фиксированной мощности. Но при таком обращении мы приходим в нашей задаче к проблеме минимизации (18), а это и есть проблема построения минимаксного критерия (более подробно эта проблема будет обсуждаться в § 50). Это объясняет в какой-то мере сходство доказываемого ниже утверждения с теоремой 41.2.

Итак, рассмотрим частично байесовскую постановку задачи, в силу которой параметр θ на множестве Θ_1 выбирается случайным образом с распределением $\mathbf{Q}_1$. Сложная гипотеза H_1 заменяется при этом простой гипотезой H_{Q_1}, при которой плотность X определяется как усредненное по мере $\mathbf{Q}_1$ значение

$$f_{Q_1}(\boldsymbol{x}) = \int_{\Theta_1} f_\theta(\boldsymbol{x})\mathbf{Q}_1(d\theta).$$

Для проверки H_{Q_1} против H_2 в классе $K_\varepsilon^{Q_1} = \{\pi : \mathbf{E}_{Q_1}\pi(X) \leqslant \varepsilon\}$ критериев уровня $1-\varepsilon$ существует н. м. к. π_{Q_1}, имеющий вид (π_{Q_1} есть критерий $\pi_{Q_1Q_2}$ в обозначениях § 44, где $\mathbf{Q}_2$ — вырожденное в точке θ_2 распределение):

$$\pi_{Q_1}(\boldsymbol{x}) = \begin{cases} 1, & \text{если} \quad g(\boldsymbol{x}) > cf_{Q_1}(\boldsymbol{x}), \\ 0, & \text{если} \quad g(\boldsymbol{x}) < cf_{Q_1}(\boldsymbol{x}) \end{cases} \tag{19}$$

(здесь $g(\boldsymbol{x}) = f_{\theta_2}(\boldsymbol{x})$ в параметрическом случае).

Теорема 3. *Допустим, что существует такое распределение* $\mathbf{Q}_1$, *сосредоточенное на подмножестве* $\Theta_1^0 \subset \Theta_1 \big(\mathbf{Q}_1(\Theta_1^0) = 1\big)$, *для которого:*

1) *имеем*

$$\pi_{Q_1} \in K_\varepsilon^{Q_1}, \tag{20}$$

2) *справедливы равенства*

$$\mathbf{E}_\theta \pi_{Q_1}(X) = \text{const} = \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_{Q_1}(X) \tag{21}$$

для всех $\theta \in \Theta_1^0$.

Тогда критерий $\pi_{Q_1} \in K_\varepsilon$ *является н. м. к. для проверки* H_1 *против* H_2.

Доказательство. Проверим сначала принадлежность $\pi_{Q_1} \in K_\varepsilon$. В силу условий теоремы

$$\sup_{\theta\in\Theta_1^0} \mathbf{E}_\theta \pi_{Q_1}(X) = \int_{\Theta_1^0} \mathbf{E}_\theta \pi_{Q_1}(X)\mathbf{Q}_1(d\theta) = \mathbf{E}_{Q_1}\pi_{Q_1}(X) \leqslant \varepsilon. \tag{22}$$

Пусть теперь π — любой другой критерий из K_ε, т. е. критерий уровня $1-\varepsilon$ для проверки H_1 против H_2. Тогда

$$\mathbf{E}_{Q_1}\pi(X) = \int \pi(\boldsymbol{x}) f_{Q_1}(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \int\limits_{\Theta_1} \mathbf{E}_\theta \pi(X)\mathbf{Q}_1(d\theta) \leqslant \varepsilon,$$

и, стало быть, $\pi \in K_\varepsilon^{Q_1}$. Но тогда в силу определения π_{Q_1}

$$\mathbf{E}_{\theta_2}\pi_{Q_1}(X) \geqslant \mathbf{E}_{\theta_2}\pi(X),$$

что и требовалось доказать. $\triangleleft$

Распределение $\mathbf{Q}_1$, фигурирующее в теореме, называют *наименее благоприятным*. Это связано со следующим обстоятельством. Величина $\beta_{Q_1}(\theta_2) = \mathbf{E}_{\theta_2}\pi_{Q_1}(X)$ есть наибольшее значение мощности, которое можно достичь в $K_\varepsilon^{Q_1}$ при «априорном» распределении $\mathbf{Q}_1$ на Θ_1. Если мы возьмем теперь любое другое распределение $\mathbf{Q}'$ на Θ_1, то получим

$$\beta_{Q'}(\theta_2) \geqslant \beta_{Q_1}(\theta_2), \quad \beta_{Q_1}(\theta_2) = \inf_{\mathbf{Q}'} \beta_{Q'}(\theta_2)$$

(это и есть смысл термина «наихудшее распределение»). В самом деле, π_{Q_1} в силу (22) принадлежит K_ε и, стало быть, $K_\varepsilon^{Q'}$. Это означает, что его мощность $\beta_{Q_1}(\theta_2) = \mathbf{E}_{\theta_2}\pi_{Q_1}(X)$ не будет превосходить мощности н. м. к. в $K_\varepsilon^{Q'}$, равной по определению $\beta_{Q'}(\theta_2)$.

Мы могли бы теперь с помощью теоремы 3 доказывать теоремы 1 и 2. Множество Θ_1^0, на котором сосредоточено наименее благоприятное распределение в теоремах 1 и 2 , состоит соответственно из одной $\{\theta_1\}$ и двух $\{\theta_1, \theta_2\}$ точек. Условия (20), (21) превращаются соответственно в условия (3) и (7).

Аналогичным образом следует использовать теорему 3 для построения р. н. м. к. и в других случаях: если построенный критерий π_{Q_1} не зависит от $\theta_2 \in \Theta_2$, то он будет р. н. м. к. для проверки $H_1 = \{\theta \in \Theta_1\}$ в классе K_ε.

Наименее благоприятное распределение $\mathbf{Q}_1$, удовлетворяющее условиям теоремы 3, существует при весьма широких предположениях, которые в реальных задачах обычно выполнены. Достаточно требовать компактность Θ_1 и непрерывность $f_\theta(x)$ по θ для п. в. x (подробнее об этом см. [65] и главу 6).

Дальнейшее обсуждение связей между байесовским и минимаксным подходами см. в § 49.

§ 46.* Несмещенные критерии

В этом и следующем параграфах мы будем использовать принципы несмещенности и инвариантности для естественного сужения класса рассматриваемых критериев. Цель такого сужения — отыскание оптимальных критериев в построенных подклассах.

1. Определения. Несмещенные р. н. м. к. Пусть, как и в предыдущем параграфе, мы рассматриваем проверку сложной гипотезы $H_1 = \{\theta \in \Theta_1\}$

против $H_2 = \{\theta \in \Theta_2\}$ по выборке $X \Subset \mathbf{P}_\theta$, $\theta \in \Theta = \Theta_1 \cup \Theta_2$. Рассмотрим критерии π из класса $K_\varepsilon = \{\pi : \sup\limits_{\theta\in\Theta_1} \mathbf{E}_\theta\pi \leqslant \varepsilon\}$.

Если, например, Θ_1 состоит из одной точки θ_1, $\mathbf{E}_{\theta_1}\pi = \varepsilon$, то ε есть вероятность отклонения H_1, когда H_1 верна. Естественное требование к критерию π состоит в том, чтобы *вероятность отклонения H_1, когда H_1 неверна, должна быть больше ε*. Если это не так, то найдутся альтернативы, при которых принятие H_1 более вероятно, чем в случаях, когда H_1 верна. Такое положение является нежелательным. Мы приходим к необходимости выделить следующий важный класс критериев.

Определение 1. Критерий π называется *несмещенным*, если для него

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta\pi(X) \geqslant \sup_{\theta\in\Theta_1} \mathbf{E}_\theta\pi(X). \tag{1}$$

Таким образом, критерий $\pi \in K_\varepsilon$, для которого $\sup\limits_{\theta\in\Theta_1} \mathbf{E}_\theta\pi = \varepsilon$, будет несмещенным, если $\beta_\pi(\theta) \geqslant \varepsilon$ при $\theta \in \Theta_2$. Класс несмещенных критериев уровня $1 - \varepsilon$ мы обозначим через $\check{K}_\varepsilon$.

Односторонний критерий π с критической областью $T > c$ (или $T < c$) для экспоненциальных семейств, упоминавшийся в предыдущем параграфе, не может быть несмещенным для проверки $H_1 = \{X \Subset \mathbf{P}_{\theta_1}\}$ против $H_2 = \{X \Subset \mathbf{P}_{\theta_1}, \theta \neq \theta_1\}$, так как здесь $\Theta_2 = \{\theta : \theta \neq \theta_1\}$, $\mathbf{E}_\theta\pi < \varepsilon$ при $\theta < \theta_1$, если $\mathbf{E}_{\theta_1}\pi = \varepsilon$ (см. теорему 45.1).

Наоборот, р. н. м. к., если они существуют, с необходимостью являются несмещенными, так как для них мощность $\beta(\theta)$ при $\theta \in \Theta_2$ не может быть меньше мощности тривиального критерия $\pi(X) \equiv \varepsilon$.

Принцип несмещенности*) представляет самостоятельный интерес, поскольку он дает возможность естественного сужения класса критериев. Это позволяет нам строить р. н. м. к. в классе $\check{K}_\varepsilon$ в тех случаях, когда р. н. м. к. в классе K_ε не существуют.

Как мы увидим, это относится, в частности, к задаче о проверке гипотезы $H_1 = \{\theta \in [\theta_1, \theta_2]\}$, $\theta_1 \leqslant \theta_2$, против двусторонней альтернативы $H_2 = \{\theta \notin [\theta_1, \theta_2]\}$ (ср. с п. 45.2).

Отыскание р. н. м. несмещенных критериев во многом может быть сведено к использованию уже применявшихся приемов, существо которых изложено в лемме 45.2. При этом может быть полезным следующее утверждение.

Допустим, что существует общая непустая граница Γ множеств Θ_1 и Θ_2 из $\mathbb{R}^k$:

$$\Gamma = \partial\Theta_1 \cap \partial\Theta_2$$

($\partial\Theta_i$ обозначает границу Θ_i), т. е. множество точек, предельных одновременно для Θ_1 и Θ_2. Допустим, кроме того, что для всех $\pi \in \check{K}_\varepsilon$

$$\beta_\pi(\theta) = \mathbf{E}_\theta\pi(X) = \varepsilon \quad \text{при всех} \quad \theta \in \Gamma. \tag{2}$$

*) Термин «несмещенность» использовался и применительно к оценкам. Свойство несмещенности оценки в известном смысле аналогично свойству несмещенности критерия: если оценка θ^* является смещенной, то $\mathbf{E}_{\theta_0}\theta^* \neq \theta_0$ и найдутся другие значения параметра $\theta \neq \theta_0$, при которых среднее значение $\mathbf{E}_\theta\theta^*$ будет равно θ_0.

Это свойство будет, очевидно, всегда выполнено, если $\beta_n(\theta)$ непрерывно зависит от θ для любого критерия π из $\check{K}_\varepsilon$.

Так как

$$\beta_\pi(\theta) = \int \pi(\boldsymbol{x}) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}), \quad 0 \leqslant \pi(\boldsymbol{x}) \leqslant 1,$$

то непрерывность $\beta_\pi(\theta)$ будет иметь место, если непрерывной по θ для μ^n-п. в. $\boldsymbol{x}$ будет функция $f_\theta(\boldsymbol{x})$. Это вытекает из следствия 1 приложения V.

Обозначим через $\overline{K}_\varepsilon$ класс всех критериев π, удовлетворяющих (2).

Лемма 1. *Пусть $\check{K}_\varepsilon \subset \overline{K}_\varepsilon$ (т. е. выполнено (2)). Тогда если $\check{\pi}$ есть р. н. м. к. в $\overline{K}_\varepsilon \cap K_\varepsilon$, то $\check{\pi}$ — р. н. м. к. в $\check{K}_\varepsilon$.*

Доказательство. Нам достаточно убедиться, что $\check{\pi} \in \check{K}_\varepsilon$ и что $\check{K}_\varepsilon \subset \overline{K}_\varepsilon \cap K_\varepsilon$. Второе из этих соотношений вытекает из предположения $\check{K}_\varepsilon \subset \overline{K}_\varepsilon$. Первое следует из того, что критерий $\pi \equiv \varepsilon$ принадлежит $\overline{K}_\varepsilon \cap K_\varepsilon$ и, стало быть, $\inf\limits_{\theta \in \Theta_2} \mathbf{E}_\theta \check{\pi}(X) \geqslant \inf\limits_{\theta \in \Theta_2} \mathbf{E}_\theta \pi = \varepsilon$. $\triangleleft$

Таким образом, лемма 1 позволяет свести поиски р. н. м. несмещенных критериев к поискам обычных р. н. м. к., но при наличии граничных условий (2). Если число точек границы Γ конечно, то мы окажемся в условиях леммы 45.2, где нам останется проверить независимость полученной оптимальной критической функции π от значения $\theta \in \Theta_2$, для которого максимизировался функционал $\mathbf{E}_\theta \pi(X)$. Это и будет означать равномерную наибольшую мощность.

Отметим теперь следующее обстоятельство, связанное с вырождением условий (2), которое проще всего пояснить для одномерного случая. Если $\Theta_1 = [\theta_1, \theta_2]$, а Θ_2 есть дополнение к Θ_1, то условия (2) будут представлять собой два уравнения $\mathbf{E}_{\theta_i} \pi(X) = \varepsilon$, $i = 1, 2$. Однако в предельном случае $\theta_1 = \theta_2$ эти уравнения превратятся в одно. Но в силу несмещенности критерия π его мощность $\beta_\pi(\theta)$ должна будет в точке θ_1 достигать своего минимума (см. (1)). Стало быть, если $\beta_\pi(\theta)$ дифференцируема, то роль уравнений (2) в случае $\theta_1 = \theta_2$ будут играть равенства

$$\beta_\pi(\theta_1) = \varepsilon, \quad \beta'_\pi(\theta_1) = 0. \tag{3}$$

Условия дифференцируемости $\int f_\theta(x) \mu(dx)$, а тем самым и $\beta_\pi(\theta) = \mathbf{E}_\theta \pi(X)$, выяснены в приложении V. Если эти условия выполнены, то

$$\beta'_\pi(\theta) = \int \pi(\boldsymbol{x}) f'_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \int \pi(\boldsymbol{x}) L'(\boldsymbol{x}, \theta) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \mathbf{E}_\theta \pi(X) L'(X, \theta).$$

Это означает, что условия (3) вновь можно записать в терминах интегралов:

$$\mathbf{E}_{\theta_1} \pi(X) = \varepsilon, \quad \mathbf{E}_{\theta_1} \pi(X) L'(X, \theta_1) = 0. \tag{4}$$

Например, для экспоненциального семейства (45.9)

$$L'(\boldsymbol{x}, \theta) = \frac{c'(\theta)}{c(\theta)} + a'(\theta) T(\boldsymbol{x}).$$

Так как $\mathbf{E}_\theta L'(X,\theta)=0$, то $c'(\theta)/c(\theta)=-a'(\theta)\mathbf{E}_\theta T(X)$,

$$\mathbf{E}_\theta\pi(X)L'(X,\theta)=-a'(\theta)\mathbf{E}_\theta T(X)\cdot\mathbf{E}_\theta\pi(X)+a'(\theta)\mathbf{E}_\theta\pi(X)T(X),$$

и уравнения (4) примут вид

$$\mathbf{E}_{\theta_1}(\pi(X)-\varepsilon)=0,\quad \mathbf{E}_{\theta_1}(\pi(X)-\varepsilon)T(X)=0.$$

В качестве иллюстрации рассмотрим случай, к разбору которого, по существу, уже все подготовлено.

2. Двусторонние альтернативы. Экспоненциальное семейство.

Теорема 1. *Пусть $f_\theta(x)$ определяется равенством* 45.9 *и проверяется гипотеза $H_1=\{\theta\in[\theta_1,\theta_2]\}$, $\theta_1\leqslant\theta_2$, против альтернативы $H_2=\{\theta\notin[\theta_1,\theta_2]\}$. Тогда если функция $a(\theta)$ монотонна, то:*

1) *в классе $\check{K}_\varepsilon$ несмещенных критериев уровня $1-\varepsilon$ существует р. н. м. к. $\check{\pi}$, который имеет вид*

$$\check{\pi}(x)=\begin{cases}0, & \text{если} \quad c_1<T(\boldsymbol{x})<c_2,\\ p_i, & \text{если} \quad T(\boldsymbol{x})=c_i,\ i=1,2,\\ 1, & \text{если} \quad T(\boldsymbol{x})\notin[c_1,c_2],\end{cases}\tag{5}$$

где $T(\boldsymbol{x})=\sum\limits_{i=1}^{n}U(x_i)$, а постоянные c_i, p_i, $i=1,2$, определяются из условий

$$\mathbf{E}_{\theta_i}\check{\pi}(X)=\varepsilon,\quad i=1,2,\tag{6}$$

если $\theta_1<\theta_2$, и из условий

$$\mathbf{E}_{\theta_1}\check{\pi}(X)=\varepsilon,\quad \mathbf{E}_{\theta_1}(\check{\pi}(X)-\varepsilon)T(X)=0,\tag{7}$$

если $\theta_1=\theta_2$;

2) *критерий $\check{\pi}$ минимизирует функцию $\beta_\pi(\theta)=\mathbf{E}_\theta\pi(X)$ при условиях* (6) *внутри отрезка $[\theta_1,\theta_2]$ и максимизирует ее вне $[\theta_1,\theta_2]$ при условиях* (6) *или* (7) (*последнее при $\theta_1=\theta_2$*);

3) *при $0<\varepsilon<1$, $\theta_1<\theta_2$ функция $\check{\beta}(\theta)=\mathbf{E}_\theta\check{\pi}(X)$ имеет минимум в некоторой точке $\theta_0\in(\theta_1,\theta_2)$ и строго возрастает при удалении θ от θ_0 вправо или влево. При этом исключаем случай* (45.8).

Нетрудно видеть, что формулировка этой теоремы почти повторяет утверждение теоремы 45.2 с той, однако, разницей, что сами утверждения иногда носят «противоположный» характер и не исключается равенство $\theta_1=\theta_2$.

Доказательство. В случае $\theta_1<\theta_2$ оно полностью следует доказательству теоремы 45.2. В замечании 1 к упомянутой теореме сказано, что при $\theta_1<\theta_2$ все рассуждения теоремы полностью сохранятся в случае, когда проверяется основная гипотеза $\{\theta\notin[\theta_1,\theta_2]\}$ против $\{\theta\in[\theta_1,\theta_2]\}$, т. е. в обозначениях этого параграфа — гипотеза H_2 против H_1. Положим $\check{\pi}(x)=1-\pi^0(x)$, где π^0 — функция, определенная в (45.6) при условиях $\mathbf{E}_{\theta_i}\pi^0(X)=1-\varepsilon$, $i=1,2$, вместо (45.7). Тогда, очевидно, утверждения 2), 3) будут прямыми следствиями соответствующих утверждений теоремы 45.2.

Первое утверждение теоремы следует из второго, так как класс критериев π, удовлетворяющих (6), шире, чем $\check{K}_\varepsilon$, и, стало быть, $\check{\pi}$ будет максимизировать $\mathbf{E}_\theta\pi(X)$ в классе $\check{K}_\varepsilon$ в любой точке θ вне $[\theta_1,\theta_2]$. Это означает, что $\check{\pi}$ есть р. н. м. несмещенный критерий.

Нам осталось рассмотреть случай $\theta_1=\theta_2$. Здесь, по-видимому, проще воспользоваться леммой 45.2. Возьмем любое $\theta\neq\theta_1$ и рассмотрим задачу максимизации $\mathbf{E}_\theta\pi(X)$ при условиях

$$\mathbf{E}_{\theta_1}\pi(X)=\varepsilon,\quad \mathbf{E}_{\theta_1}\pi(X)T(X)=\varepsilon\mathbf{E}_{\theta_1}T(X). \tag{8}$$

Очевидно, мы будем находиться в условиях леммы 45.2, если положим $m=2$, $f_1=f_{\theta_1}$, $f_2=Tf_{\theta_1}$, $f_3=f_\theta$, $\varepsilon_1=\varepsilon$, $\varepsilon_2=\varepsilon\mathbf{E}_\theta T(X)$. Согласно этой лемме максимум $\mathbf{E}_\theta\pi$ будет достигаться на функции

$$\check{\pi}(\boldsymbol{x})=\begin{cases}1, & \text{если}\quad f_\theta(\boldsymbol{x})>k_1f_{\theta_1}(\boldsymbol{x})+k_2T(\boldsymbol{x})f_{\theta_1}(\boldsymbol{x}),\\ 0, & \text{если}\quad f_\theta(\boldsymbol{x})<k_1f_{\theta_1}(\boldsymbol{x})+k_2T(\boldsymbol{x})f_{\theta_1}(\boldsymbol{x}).\end{cases}$$

Рассмотрим последнее неравенство, которое можно записать в форме

$$\frac{c(\theta)}{c(\theta_1)}e^{(a(\theta)-a(\theta_1))T(\boldsymbol{x})}<k_1+k_2T(\boldsymbol{x}).$$

Ясно, что для любых $c_1<c_2$ можно всегда выбрать k_1, k_2 так, чтобы это неравенство было эквивалентно

$$c_1<T<c_2.$$

Это доказывает, что критерий вида (5) максимизирует $\mathbf{E}_\theta\pi(X)$ при условиях (7), если только c_i, p_i, $i=1,2$, в (5) можно выбрать так, чтобы удовлетворялось (7) (или (8)). Этот критерий будет, очевидно, р. н. м. несмещенным критерием, так как класс критериев π, удовлетворяющих (8), шире, чем $\check{K}_\varepsilon$, и, стало быть, $\check{\pi}$ будет максимизировать $\mathbf{E}_\theta\pi(X)$ и в $\check{K}_\varepsilon$. Итак, для доказательства теоремы нам осталось убедиться, что справедлива

Лемма 2. *Уравнение* (7) *при* $0<\varepsilon<1$ *разрешимо относительно* c_i, p_i, $i=1,2$.

Доказательство этой леммы мы приведем, как и доказательство леммы 45.1, при упрощающем предположении, что $\mathbf{P}_{\theta_1}$-распределение $T(X)$ непрерывно, т. е. $\mathbf{P}_{\theta_1}(T(X)=c)=0$ при всех c.

Напомним, что плотность распределения T относительно некоторой меры ν можем считать равной (см. § 45) $g_\theta(t)=c(\theta)e^{\theta t}$. Тогда уравнения (7), (8) будут эквивалентны соотношениям

$$\begin{aligned}&\mathbf{E}_{\theta_1}(1-\pi(X))=c(\theta_1)\int\limits_{c_1}^{c_2}e^{\theta_1 t}\nu(dt)=1-\varepsilon,\\ &\mathbf{E}_{\theta_1}(1-\pi(X))T(X)=c(\theta_1)\int\limits_{c_1}^{c_2}te^{\theta_1 t}\nu(dt)=(1-\varepsilon)c(\theta_1)\int\limits_{c_1}^{c_2}te^{\theta_1 t}\nu(dt).\end{aligned} \tag{9}$$

Обозначая $r(t) = t$, $m = \mathbf{E}_{\theta_1} T(X) = c(\theta_1) \int te^{\theta_1 t} \nu(dt)$, мы можем уравнения (9) записать в виде

$$c(\theta_1) \int_{c_1}^{c_2} e^{\theta_1 t} \nu(dt) = 1 - \varepsilon, \quad c(\theta_1) \int_{c_1}^{c_2} r(t) e^{\theta_1 t} \nu(dt) = (1 - \varepsilon) m. \tag{10}$$

Мы пришли к задаче, совпадающей с задачей, рассматривавшейся в лемме 45.1, с той лишь разницей, что распределение с плотностью $r(t) g_{\theta_1}(t)$ может быть обобщенным (т. е. принимать и отрицательные значения). В этих новых условиях сохраняются рассуждения леммы 45.1. $\triangleleft$

§ 47.* Инвариантные критерии

В этом параграфе рассмотрим еще один способ сужения класса всех критериев, основанный на этот раз на соображениях инвариантности.

Предположим, что $X \in \{\mathbf{P}_\theta\}$ и $\{\mathbf{P}_\theta\}$ есть инвариантное семейство. Напомним необходимые обозначения и понятия (см. § 29). Пусть дана группа G измеримых преобразований g пространства $\mathcal{X}^n$ в себя. Семейство $\{\mathbf{P}_\theta\}$ будет *инвариантным относительно* G, если для каждого $g \in G$ и $\theta \in \Theta$ найдется элемент $\theta_g \in \Theta$ такой, что

$$\mathbf{P}_{\theta_g}(X \in A) = \mathbf{P}_\theta(gX \in A)$$

для любого $A \in \mathfrak{B}^n_{\mathcal{X}}$.

Преобразования $\overline{g}$ пространства Θ, определяемые равенством $\overline{g}\theta = \theta_g$, при выполнении условий (A_0) образуют группу $\overline{G}$ (см. § 29).

Определение 1. Будем говорить, что задача проверки гипотезы $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$, $\Theta_1 \cup \Theta_2 = \Theta$, является *инвариантной*, если выполнены следующие два условия:

1) семейство $\{\mathbf{P}_\theta\}$ является инвариантным относительно G;

2) множества Θ_1 и Θ_2 являются инвариантными относительно $\overline{g} \in \overline{G}$, т. е. $\overline{g}\Theta_i = \Theta_i$, $i = 1, 2$.

Если задача проверки гипотез является инвариантной, то для ее решения естественно воспользоваться инвариантным критерием.

Определение 2. Критерий π называется *инвариантным*, если $\pi(X)$ есть инвариантная относительно g статистика*):

$$\pi(g\boldsymbol{x}) = \pi(\boldsymbol{x}) \quad \text{для всех} \quad \boldsymbol{x} \in \mathcal{X}^n, \quad g \in G.$$

Если π есть нерандомизированный критерий и Ω_j есть область приема гипотезы H_j, то инвариантность π означает, что $g\Omega_j = \Omega_j$, $j = 1, 2$.

Естественность использования инвариантных критериев, по-видимому, проще всего понять на примерах. Общее обсуждение, связанное с трактовкой g как замены координат и нечувствительности к этой замене соответствующих статистик, содержится в § 29.

*) См. сноску в § 29.

Пример 1. Наиболее простые примеры относятся к тому случаю, когда группа $\overline{G}$ тривиальна, т. е. $\overline{g}$ для любого g есть тождественное преобразование $\overline{e}$ пространства Θ.

Пусть $X \in \mathbf{\Phi}_{0,\sigma^2}$ и проверяется гипотеза $H_1 = \{\sigma_1 \leqslant \sigma \leqslant \sigma_2\}$ против дополнительной альтернативы H_2. В этом случае

$$f_{\sigma^2}(X) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} \mathrm{x}_i^2\right\}.$$

Очевидно, что семейство $\mathbf{\Phi}_{0,\sigma^2}$ инвариантно относительно группы G ортогональных преобразований g (вращений) пространства $\mathcal{X}^n$, при этом $\overline{g} = \overline{e}$ для любого g. Поэтому естественно рассматривать критерии π, зависящие от статистики $T(X) = \sum_{i=1}^{n} \mathrm{x}_i^2$. Так как $\sigma^{-2}T(X) \in \mathbf{\Gamma}_{1/2,n/2} = \mathbf{H}_n$, то $T(X) \in \mathbf{\Gamma}_{\alpha,n/2}$ при $\alpha = 1/(2\sigma^2)$, и мы приходим к задаче о проверке гипотезы $H_1 = \{\alpha_1 \leqslant \alpha \leqslant \alpha_2\}$, $\alpha_1 = 1/(2\sigma_2^2)$, $\alpha_2 = 1/(2\sigma_1^2)$, по наблюдению $T(X)$, имеющему распределение $\mathbf{\Gamma}_{\alpha,n/2}$ из экспоненциального семейства. С помощью результатов предыдущих параграфов мы можем построить р. н. м. несмещенный критерий уровня $1-\varepsilon$, принимающий H_1 при

$$c_1 \leqslant T(X) \leqslant c_2, \tag{1}$$

где c_i выбираются так, чтобы $\mathbf{\Gamma}_{\alpha_1,n/2}(\mathbb{R} \setminus [c_1, c_2]) = \mathbf{\Gamma}_{\alpha_2,n/2}(\mathbb{R} \setminus [c_1, c_2]) = \varepsilon$.

Отметим, что критерий вида (1) в этом примере мы могли бы построить и из других соображений, основываясь на принципе достаточности, так как статистика T является достаточной. Ведь мы знаем, что вся информация о параметре σ^2 сосредоточена в T и привлечение других статистик (т. е. другой информации о выборке) не имеет смысла.

В дальнейшем там, где это возможно, мы будем сразу редуцировать задачу к задаче о достаточных статистиках.

Пример 2. Пусть $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$, $H_1 = \{\sigma_1 \leqslant \sigma \leqslant \sigma_2\}$. В этом случае $\theta = (\alpha, \sigma^2)$, и преобразование сдвига $gX = X + c = (\mathrm{x}_1 + c, \ldots, \mathrm{x}_n + c)$ индуцирует преобразование $\overline{g}\alpha = \alpha + c$, которое оставляет гипотезу H_1 неизменной. Если ограничиться рассмотрением достаточных статистик

$$T_1 = \overline{\mathrm{x}}, \quad T_2 = \sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^2,$$

то преобразование g дает

$$T_1(gX) = \overline{\mathrm{x}} + c, \quad T_2(gX) = T_2(X).$$

Таким образом, статистика T_2 является инвариантной относительно G. То есть инвариантный критерий π, основанный на достаточных статистиках, должен быть функцией от T_2. (Ниже мы увидим, что *любой* инвариантный критерий π должен быть функцией от T_2.) В силу § 40 $\sigma^{-2}T_2 \in \mathbf{\Gamma}_{1/2,(n-1)/2}$, и мы приходим к задаче, разобранной в предыдущем примере. Р. н. м. инвариантный несмещенный критерий будет иметь вид $c_1 \leqslant T_2 \leqslant c_2$.

Пример 3. Оба рассмотренные выше примера касались нормального распределения. Применительно к распределению выборки X это было многомерное нормальное распределение с диагональной матрицей вторых моментов. Для дальнейшего полезно отметить, что семейство произвольных многомерных нормальных распределений $\mathbf{\Phi}_{\alpha,\sigma^2}$, $\alpha \in \mathbb{R}^m$, $\sigma^2 = \|\sigma_{ij}\|$, $i,j = 1,\ldots,m$, является инвариантным относительно группы G линейных невырожденных преобразований

$$gx = (x - a)C,$$

где C — обратимая матрица. В самом деле, мы должны убедиться, что при некотором преобразовании $\overline{g}$ выполняется $\mathbf{P}_{\overline{g}\theta}(A) = \mathbf{P}_\theta(g^{-1}A)$, где $\mathbf{P}_\theta \equiv \mathbf{\Phi}_{\alpha,\sigma^2}$, $\theta = (\alpha,\sigma^2)$, $g^{-1}A$, как обычно, означает множество $g^{-1}A = \{x \in \mathbb{R}^m : gx \in A\}$. Имеем ($\sigma = \sqrt{|\sigma^2|}$)

$$\mathbf{\Phi}_{\alpha,\sigma^2}(g^{-1}A) = \frac{1}{(2\pi)^{m/2}\sigma} \int\limits_{g^{-1}A} \exp\left\{-\frac{1}{2}(x-\alpha)\sigma^{-2}(x-\alpha)^\top\right\} dx.$$

После замены $y = gx$ получим

$$\mathbf{\Phi}_{\alpha,\sigma^2}(g^{-1}A) = \frac{1}{(2\pi)^{m/2}\sigma|C|} \int\limits_{A} \exp\left\{-\frac{1}{2}(g^{-1}y-\alpha)\sigma^{-2}(g^{-1}y-\alpha)^\top\right\} dy.$$

Учитывая, что $g^{-1}y = yC^{-1} + a$, мы можем показатель у экспоненты в последнем интеграле записать в виде

$$(y - (\alpha - a)C)C^{-1}\sigma^{-2}(C^{-1})^\top(y - (\alpha - a)C)^\top.$$

Стало быть, если положить

$$\overline{g}\theta = \overline{g}(\alpha,\sigma^2) = (g\alpha, C^\top\sigma^2 C) = ((\alpha - a)C, C^\top\sigma^2 C), \tag{2}$$

то мы получим

$$\mathbf{\Phi}_{\alpha,\sigma^2}(g^{-1}A) = \mathbf{\Phi}_{\overline{g}(\alpha,\sigma^2)}(A). \tag{3}$$

Пример 4. Пусть гипотезы H_j имеют следующий вид: $H_j = \{X \in \in \mathbf{P}_{j,\alpha}\}$, $\alpha \in \mathcal{X}$, $j = 1,2$, где $\mathbf{P}_{j,\alpha}$ есть распределения с плотностями $f_j(x - \alpha)$, $j = 1,2$. Другими словами, нас интересует, какому из двух типов распределений, с точностью до сдвига, принадлежит выборка X. Здесь следует положить $\theta = (\nu,\alpha)$, $\nu = 1,2$, $\alpha \in \mathcal{X}$, и рассмотреть преобразование $gX = X + c$, которое в параметрическом пространстве индуцирует преобразование $\overline{g}\theta = (\nu, \alpha + c)$. Ясно, что гипотезы $H_j = \{\nu = j\}$, $j = 1,2$, инвариантны относительно $\overline{g}$ и, стало быть, задача проверки этих гипотез является инвариантной. Статистика

$$Y = \{\mathrm{x}_1 - \mathrm{x}_n, \ldots, \mathrm{x}_{n-1} - \mathrm{x}_n\}$$

будет инвариантом относительно g (ср. с § 28). Ее распределение в точке $\boldsymbol{y} = (y_1, \ldots, y_{n-1})$ при гипотезе H_j имеет плотность

$$f_j^Y(\boldsymbol{y}) = \int \prod_{i=1}^{n-1} f_j(y_i + z) f_j(z) \mu(dz). \tag{4}$$

Отсюда видно, что для наблюдения Y гипотезы H_j превращаются в простые гипотезы, в соответствии с которыми плотности f_j^Y для Y имеют вид (4). В этих условиях мы можем воспользоваться леммой Неймана–Пирсона и построить н. м. к. π, который принимает H_2, если

$$\frac{f_2^Y(Y)}{f_1^Y(Y)} > c. \tag{5}$$

Так как этот критерий от α не зависит, он будет р. н. м. к. для проверки H_1 против H_2 среди всех инвариантных критериев, основанных на статистике Y.

В связи с рассмотренными примерами желательно иметь уверенность в том, что остальные инвариантные критерии в этих задачах также являются функциями выбранных нами инвариантных статистик. Особенно это относится к последнему примеру, так как в двух предыдущих наш выбор критериев определялся также соображениями достаточности.

Для того чтобы выяснить взаимоотношения между инвариантами, введем некоторые понятия. Две точки $\boldsymbol{x}$ и $\boldsymbol{x}'$ из $\mathcal{X}^n$ будем называть *эквивалентными* относительно группы G, если найдется $g \in G$ такое, что $\boldsymbol{x}' = g\boldsymbol{x}$. Так как G есть группа, все пространство $\mathcal{X}^n$ разбивается на непересекающиеся классы эквивалентности, которые в § 29 мы называли *орбитами*. Чтобы получить некоторую орбиту, достаточно взять какую-нибудь одну ее точку $\boldsymbol{x}_0$ и применить к ней все преобразования g из G. Например, для ортогональных преобразований в примере 1 орбиты образуют сферы с центром в начале координат.

Инвариантность статистики T относительно G равнозначна тому, что T постоянно на каждой орбите.

Определение 3. Статистика T называется *максимальным инвариантом*, если она инвариантна и из $T(\boldsymbol{x}') = T(\boldsymbol{x})$ следует $\boldsymbol{x}' = g\boldsymbol{x}$ при некотором $g \in G$.

Это означает, что максимальный инвариант принимает разные значения на разных орбитах.

Теорема 1. *Пусть T — максимальный инвариант. Статистика S инвариантна тогда и только тогда, когда S зависит от X через T, т. е. когда существует функция φ такая, что $S(X) = \varphi(T(X))$.*

Для упрощения изложения мы не затрагиваем здесь важный вопрос об измеримости φ. Отметим лишь, что в рассматриваемых в этом параграфе примерах такая измеримость будет иметь место*).

*) Подробнее см., например, [48, 65].

Доказательство. Если $S(\boldsymbol{x}) = \varphi(T(\boldsymbol{x}))$, то $S(g\boldsymbol{x}) = \varphi(T(g\boldsymbol{x})) = \varphi(T(\boldsymbol{x})) = S(\boldsymbol{x})$, и, стало быть, S инвариантна. Чтобы доказать обратное утверждение, нам надо убедиться, что из $T(\boldsymbol{x}) = T(\boldsymbol{x}')$ следует $S(\boldsymbol{x}) = S(\boldsymbol{x}')$. Но это так в силу того, что $T(\boldsymbol{x}) = T(\boldsymbol{x}')$ влечет за собой существование g такого, что $\boldsymbol{x}' = g\boldsymbol{x}$. Так как S — инвариант, то $S(\boldsymbol{x}) = S(\boldsymbol{x}')$. $\triangleleft$

Рассмотрим в качестве примера группу G сдвигов

$$g(\boldsymbol{x}) = \boldsymbol{x} + c = (x_1 + c, \ldots, x_n + c).$$

Как мы уже отмечали, статистика $Y(\boldsymbol{x}) = (x_1 - x_n, \ldots, x_{n-1} - x_n)$ является инвариантом. Покажем, что это максимальный инвариант. Действительно, из $Y(\boldsymbol{x}) = Y(\boldsymbol{x}') \equiv (x_1' - x_n', \ldots, x_{n-1}' - x_n')$ следует, что $x_i - x_n = x_i' - x_n'$ при всех $i = 1, \ldots, n-1$. Полагая $x_n' - x_n = c$, получим $x_i' = x_i + c$, $i = 1, \ldots, n$, $\boldsymbol{x}' = \boldsymbol{x} + c = g\boldsymbol{x}$, что и означает требуемую эквивалентность $\boldsymbol{x}'$ и $\boldsymbol{x}$.

Мы можем вернуться теперь к примеру 3 и утверждать, что критерий (5) является р. н. м. среди всех инвариантных критериев, поскольку все инвариантные критерии являются по теореме 1 функциями от Y и, стало быть, предположение о существовании более мощного инвариантного критерия, чем (5), будет противоречивым.

Аналогично предыдущему читатель может убедиться, что статистика $\sum_{i=1}^{n} \mathrm{x}_i^2$ в примере 1 также является максимальным инвариантом.

Если существуют достаточные статистики, то исходную задачу бывает удобно редуцировать сначала к задаче относительно распределения достаточных статистик, а затем уже применять соображения инвариантности, как это было в примере 2, где статистика $T_2 = \sum_{i=1}^{n} (\mathrm{x}_i - \overline{\mathrm{x}})^2$ является, очевидно, максимальным инвариантом в наблюдении $(\overline{\mathrm{x}}, T_2)$.

В заключение этого параграфа отметим еще раз, что существо подхода, связанного с инвариантностью, состоит в редукции рассматриваемых задач о проверке гипотез к более простым задачам, относящимся к распределению максимальных инвариантов. В этих новых более простых условиях в ряде случаев оказывается возможным построить н. м. или р. н. м. критерии. В этом отношении «принцип инвариантности» близок к «принципам» достаточности и несмещенности, в соответствии с которыми исходная задача редуцируется к задаче в терминах достаточных или несмещенных статистик.

§ 48.* Связь с доверительными множествами

1. Связь статистических критериев и доверительных множеств. Связь свойств оптимальности. Понятия доверительного множества и статистического критерия тесно связаны между собой. Определение доверительного множества было дано в § 39. Напомним его.

Пусть $X \in \mathbf{P}_\theta$, $\theta \in \Theta$.

Определение 1. Случайное подмножество $\Theta^* = \Theta^*(X, \varepsilon)$ параметрического пространства Θ называется *доверительным множеством уровня* $1 - \varepsilon$, если

$$\mathbf{P}_\theta(\Theta^*(X, \varepsilon) \ni \theta) \geqslant 1 - \varepsilon \tag{1}$$

для всех $\theta \in \Theta$.

Очевидно, доверительный интервал есть частный случай доверительного множества. Последнее сохраняет тот же смысл: с вероятностью $\geqslant 1 - \varepsilon$ оно накрывает истинное значение параметра.

Положим

$$\Omega(\theta, \varepsilon) = \{\boldsymbol{x} \in \mathcal{X}^n : \theta \in \Theta^*(\boldsymbol{x}, \varepsilon)\}. \tag{2}$$

Тогда соотношения

$$\theta \in \Theta^*(\boldsymbol{x}, \varepsilon) \quad \text{и} \quad \boldsymbol{x} \in \Omega(\theta, \varepsilon) \tag{3}$$

будут эквивалентными.

Определение доверительного множества предполагает, что множество $\Omega(\theta, \varepsilon)$ в (2) является измеримым, так что вероятность в (1) имеет смысл и равна $\mathbf{P}_\theta(X \in \Omega(\theta, \varepsilon))$.

Доверительные множества и статистические критерии для проверки гипотезы $H_1 = \{\theta = \theta_1\}$ против дополнительной альтернативы $H_2 = \{\theta \in \Theta_2\}$, $\theta_1 \notin \Theta_2$, связаны между собой следующим образом. Пусть для каждого θ_1 определено свое множество $\Theta_2 = \Theta_2(\theta_1) \not\ni \theta_1$.

Теорема 1. 1) *Рассмотрим для каждого* θ_1 *нерандомизированный критерий* $\pi = \delta$ *уровня* $1 - \varepsilon$ *для проверки гипотезы* H_1 *против* H_2 *и обозначим через* $\Omega(\theta_1, \varepsilon)$ *его область принятия гипотезы* H_1. *Тогда множество*

$$\Theta^*(X, \varepsilon) = \{\theta \in \Theta : X \in \Omega(\theta, \varepsilon)\}$$

будет доверительным множеством уровня $1 - \varepsilon$.

Наоборот, если $\Theta^*(X, \varepsilon)$ *— доверительное множество уровня* $1 - \varepsilon$, *то множество* $\Omega(\theta_1, \varepsilon) \subset \mathcal{X}^n$, *определенное в* (2) *и взятое в качестве области приема* H_1, *будет определять критерий для проверки* $H_1 = \{\theta = \theta_1\}$ *против* $H_2 = \{\theta \in \Theta_2(\theta_1)\}$ *уровня* $1 - \varepsilon$ *для любого* $\Theta_2(\theta_1)$, $\theta_1 \notin \Theta_2(\theta_1)$.

2) *Если критерий* π *с областью приема* $\Omega(\theta_1, \varepsilon)$ *гипотезы* H_1 *является р. н. м. к., то соответствующее множество* $\Theta^*(X, \varepsilon)$ *минимизирует вероятность*

$$\mathbf{P}_\theta(\theta' \in \Theta^*(X, \varepsilon)) \quad \textit{для всех } \theta, \theta', \quad \theta \in \Theta_2(\theta'), \tag{4}$$

в классе всех доверительных множеств уровня $1 - \varepsilon$.

Верно и обратное утверждение: минимальность (4) *означает, что соответствующее множество* $\Omega(\theta, \varepsilon)$ *будет порождать р. н. м. к.*

Для одномерного параметра наиболее употребительными являются случаи $\Theta_2(\theta') = \{\theta : \theta \neq \theta'\}$ и $\Theta_2(\theta') = \{\theta : \theta > \theta'\}$ (или $\{\theta : \theta < \theta'\}$). В первом из них в (4) будет иметь место минимизация при всех $\theta' \neq \theta$, во втором — при всех $\theta' < \theta$.

Итак, теорема утверждает в (4), что для Θ^* минимизируется вероятность $\mathbf{P}_\theta$ того, что в доверительное множество попадет любое другое значение $\theta' \neq \theta$ такое, что $\theta \in \Theta_2(\theta')$. Это есть один из способов выделения оптимальных доверительных интервалов.

Определение 2. Доверительные множества, для которых минимизируется (4) при условии (1), называются *наиболее точными доверительными множествами* (уровня $1-\varepsilon$) *относительно альтернатив* θ' *таких, что* $\theta \in \Theta_2(\theta')$.

Ниже будет приведено некоторое дополнительное обоснование для такого понимания оптимальности доверительного интервала.

Таким образом, теорема 1 устанавливает, что «обращение» множества $\Omega(\theta_1, \varepsilon)$ для р. н. м. к. дает наиболее точное доверительное множество. При этом важно отметить, что рассматриваемая процедура построения доверительных множеств с размерностью θ никак не связана. Можно даже рассматривать бесконечномерные параметры θ или отождествлять θ с самим распределением $\mathbf{P}$ выборки X. Тогда соотношения эквивалентности (3), где $\Omega(\theta, \varepsilon) = \Omega(\mathbf{P}, \varepsilon)$ есть область приема гипотезы $\{X \in \mathbf{P}\}$ при альтернативе $\{X \in \mathbf{P}_1 \neq \mathbf{P}\}$, позволяют построить доверительное множество для $\mathbf{P}$. Например, мы видели в § 6, что распределение статистики $D_n = \sqrt{n} \sup\limits_t \left|F_n^*(t) - F(t)\right|$ при условии $X \in \mathbf{P}$, где F есть непрерывная функция распределения, соответствующая $\mathbf{P}$, от F не зависит и может быть найдено. Стало быть, можем найти такое $d = d(\varepsilon)$, что $\mathbf{P}(D_n \leqslant d(\varepsilon)) = 1 - \varepsilon$. Таким образом, неравенство

$$\sqrt{n} \sup_t \left|F_n^*(t) - F(t)\right| \leqslant d$$

определяет область приема гипотезы $\{X \in \mathbf{P}\}$ для критерия уровня $1 - \varepsilon$.

Но это же неравенство определяет и доверительное множество для F — просто ввиду симметрии этого неравенства относительно F и F_n^* специальной процедуры «обращения» здесь не требуется.

Доказательство теоремы 1 почти очевидно. В основе его лежит эквивалентность (3), в силу которой

$$\mathbf{P}_\theta(\theta \in \Theta^*(X, \varepsilon)) = \mathbf{P}_\theta(X \in \Omega(\theta, \varepsilon)) \geqslant 1 - \varepsilon.$$

Это доказывает первое утверждение. Чтобы доказать второе, рассмотрим какое-нибудь другое доверительное множество $\widetilde{\Theta}^*(X, \varepsilon)$, и пусть $\widetilde{\Omega}(\theta, \varepsilon)$ — соответствующее ему подмножество в $\mathcal{X}^n$. Тогда

$$\mathbf{P}_\theta(X \in \widetilde{\Omega}(\theta, \varepsilon)) = \mathbf{P}_\theta(\theta \in \widetilde{\Theta}^*(X, \varepsilon)) \geqslant 1 - \varepsilon,$$

$$\mathbf{P}_\theta(X \in \widetilde{\Omega}(\theta_1, \varepsilon)) \geqslant \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon))$$

при $\theta \in \Theta_2(\theta_1)$ и, стало быть,

$$\mathbf{P}_\theta(\theta_1 \in \widetilde{\Theta}^*(X, \varepsilon)) \geqslant \mathbf{P}_\theta(\theta_1 \in \Theta^*(X, \varepsilon)). \quad \triangleleft$$

Рассмотрим теперь важный частный случай, связанный с одномерным θ.

2. Наиболее точные доверительные интервалы

Теорема 2. *Пусть множество $\Omega(\theta,\varepsilon)$ р. н. м. к., рассмотренного в теореме* 1, *имеет вид*

$$c_1(\theta,\varepsilon) \leqslant T(\boldsymbol{x}) \leqslant c_2(\theta,\varepsilon),$$

где $c_1(\theta,\varepsilon)$ монотонно и непрерывно) зависят от θ. Пусть для определенности $c_i(\theta,\varepsilon)$ возрастают. Тогда наиболее доверительное множество (уровня $1-\varepsilon$) относительно альтернатив θ' таких, что $\theta \in \Theta_2(\theta')$, будет иметь форму интервала*

$$c_2^{-1}(T,\varepsilon) \leqslant \theta \leqslant c_2^{-1}(T,\varepsilon),$$

где $T = T(X)$, $c_i^{-1}(t,\varepsilon)$ есть решения относительно θ уравнений $c_i(\theta,\varepsilon) = t$.

Мы видим, таким образом, что процедура построения доверительного интервала здесь, по существу, та же, что и в § 39, — с той лишь особенностью, что в качестве статистики S здесь используется статистика T р. н. м. критерия.

Доказательство теоремы вполне очевидно, и оно предоставляется читателю.

Рассмотрим теперь более подробно *односторонние доверительные интервалы для скалярного θ*. Они используются там, где наибольший интерес представляет лишь одна граница для оцениваемого параметра. Такие ситуации возникают, когда оценивается вероятность какого-нибудь нежелательного события или, скажем, оценивается величина разрывающего усилия для нового сплава.

Ввиду симметрии можно ограничиться рассмотрением нижней доверительной границы $\theta^-(X,\varepsilon)$, для которой

$$\mathbf{P}_\theta(\theta^-(X,\varepsilon) \leqslant \theta) \geqslant 1-\varepsilon. \tag{5}$$

Определение 3. Граница $\theta^- = \theta^-(X,\varepsilon)$, для которой $\mathbf{P}_\theta(\theta^- \leqslant \theta')$ минимальна при всех $\theta' < \theta$, называется *наиболее точной нижней доверительной границей уровня* $1-\varepsilon$.

Допустим, что $w(\theta^-,\theta)$ есть какая-нибудь мера потерь, возникающих от «недооценки» θ: $w(\theta^-,\theta) = 0$ при $\theta^- \geqslant 0$ и $w(\theta^-,\theta) > 0$ при $\theta^- < \theta$; при этом $w(\theta^-,\theta)$ непрерывно возрастает при удалении θ^- от θ, $\mathbf{E}_\theta w(\theta^-,\theta) < \infty$.

Следующее утверждение в какой-то мере поясняет смысл определения 3.

Лемма 1. *Наиболее точная нижняя граница θ^- минимизирует значение $\mathbf{E}_\theta w(\theta^-,\theta)$ при условии* (5) *и при любой функции w, обладающей сформулированными выше свойствами.*

*) Свойства монотонности и непрерывности $c_i(\theta,\varepsilon)$ обычно вытекают из таких же свойств функции распределения $\mathbf{P}_\theta(T(X) < c)$. В обозначениях § 39 $c_1(\theta,\varepsilon) = G_\theta^{-1}(\varepsilon_1)$, $c_2(\theta,\varepsilon) = G_\theta^{-1}(1-\varepsilon_2)$, где G_θ — функция распределения $T(X)$, $\varepsilon_1+\varepsilon_2 = \varepsilon$.

Доказательство. Пусть $\widetilde{\theta}^-$ — другая нижняя граница. Тогда, так как приращения $d_u w(u,\theta)$ по u в области $u<\theta$ отрицательны, то

$$\mathbf{E}_\theta w(\theta^-,\theta)=\int\limits_{-\infty}^{\theta} w(u,\theta)\,d_u\mathbf{P}_\theta(\theta^-<u)=-\int\limits_{-\infty}^{\theta}\mathbf{P}_\theta(\theta^-<u)\,d_u w(u,\theta)\leqslant$$

$$\leqslant-\int\limits_{-\infty}^{\theta}\mathbf{P}_\theta(\widetilde{\theta}^-<u)\,d_u w(u,\theta)=\mathbf{E}_\theta w(\widetilde{\theta}^-,\theta).\ \triangleleft$$

Мы видим, таким образом, что подход к определению наиболее точных доверительных множеств в случае односторонних множеств является вполне естественным. Можно теперь с помощью теорем 1, 2 и результатов § 45 построить в явном виде односторонние доверительные интервалы для случая, когда отношение правдоподобия монотонно.

Теорема 3. *Пусть $X\in\mathbf{P}_\theta$, семейство $\{\mathbf{P}_\theta\}$ имеет монотонное отношение правдоподобия относительно статистики $T(X)$, $\mathbf{P}_\theta$-распределение которой $G_\theta(t)=\mathbf{P}_\theta(T(X)<t)$ непрерывно по θ и t. Тогда статистика T по распределению монотонно и непрерывно зависит от θ, т. е. $G_\theta(t)$ непрерывно убывает с ростом θ* (см. определение 39.3). *Если $b(t,\gamma)$ есть решение относительно θ уравнения $G_\theta(t)=\gamma$, то наиболее точная нижняя граница $\theta^-(X,\varepsilon)$ уровня $1-\varepsilon$ равна*

$$\theta^-(X,\varepsilon)=b(T(X),1-\varepsilon).$$

Другими словами, в утверждении теоремы 39.1 мы получим наиболее точную нижнюю доверительную границу, если будем использовать в качестве S статистику T.

Доказательство. В нашем случае в условиях теорем 1, 2 надо положить $\Theta_2(\theta)=\{t:t>\theta\}$. В силу теоремы 45.1 существует нерандомизированный р. н. м. к. для проверки $H_1=\{\theta=\theta_1\}$ против $H_2=\{\theta>\theta_1\}$ с областью $\Omega(\theta_1,\varepsilon)=\{X:T(X)<c\}$ принятия H_1, где $c=c(\theta_1,1-\varepsilon)=G_{\theta_1}^{-1}(1-\varepsilon)$ определяется из условия

$$\mathbf{P}_{\theta_1}(T(X)<c(\theta_1,1-\varepsilon))=1-\varepsilon.$$

При этом

$$\mathbf{P}_\theta(T(X)\geqslant c)>\varepsilon=\mathbf{P}_{\theta_1}(T(X)\geqslant c)$$

при $\theta>\theta_1$. Последнее означает, что $c(\theta_1,1-\varepsilon)<c(\theta,1-\varepsilon)$ при $\theta_1<\theta$, т. е. функция $c(\theta,1-\varepsilon)$ возрастает по θ. Непрерывность $c(\theta,1-\varepsilon)=G_\theta^{-1}(1-\varepsilon)$ по θ следует из непрерывности G_θ.

Мы видим, что условия теорем 1, 2 полностью выполнены при $c_2(\theta,\varepsilon)=$ $=c(\theta,1-\varepsilon)$ и, стало быть, наиболее точное доверительное множество имеет форму полуинтервала $(c^{-1}(T(X),1-\varepsilon),\infty)$, где, как мы видели в теореме 39.1, $c^{-1}(T,1-\varepsilon)=b(T,1-\varepsilon)$. $\triangleleft$

Точно так же можно построить *наиболее точную верхнюю границу* $\theta^+(X,\varepsilon)$.

Пусть теперь $\theta^-(X,\varepsilon_1) < \theta^+(X,\varepsilon_2)$ обозначают верхнюю и нижнюю доверительные границы уровней $1-\varepsilon_1$ и $1-\varepsilon_2$ соответственно. Так как события $\{\theta^-(X,\varepsilon_1) > \theta\}$ и $\{\theta^+(X,\varepsilon_2) < \theta\}$ не пересекаются, то

$$\mathbf{P}_\theta(\theta^-(X,\varepsilon_1) < \theta < \theta^+(X,\varepsilon_2)) = 1-\varepsilon_1-\varepsilon_2,$$

и $(\theta^-(X,\varepsilon_1),\theta^+(X,\varepsilon_2))$ представляет собой доверительный интервал уровня $1-\varepsilon_1-\varepsilon_2$.

Пусть $w_1(\theta^-,\theta)$ и $w_2(\theta^+,\theta)$ — функции потерь для границ $\theta^\mp$, обладающие свойствами, описанными в формулировке леммы 1.

Лемма 2. *Пусть* $w(\theta^-,\theta^+,\theta) = w_1(\theta^-,\theta) + w_2(\theta^+,\theta)$. *Тогда доверительный интервал* (θ^-,θ^+), *образованный наиболее точными верхними и нижними границами, минимизирует* $\mathbf{E}_\theta w(\theta^-,\theta^+,\theta)$ *при условиях*

$$\mathbf{P}_\theta(\theta^- > \theta) \leqslant \varepsilon_1, \quad \mathbf{P}_\theta(\theta^+ < \theta) \leqslant \varepsilon_2.$$

Эта лемма является очевидным следствием леммы 1. Она показывает, что доверительный интервал, построенный с помощью точной нижней и точной верхней границ, также будет обладать свойством оптимальности.

Теорема 3 дает возможность в явном виде строить такие интервалы для параметрических семейств с монотонным отношением правдоподобия.

Предлагаем читателю убедиться с помощью сделанных замечаний, что доверительные интервалы, построенные в § 40 для среднего и дисперсии нормального распределения, будут иметь наиболее точные верхние и нижние границы.

В теореме 1 и последующих рассмотрениях фигурировало условие, что р. н. м. к. является нерандомизированным. Однако это ограничение не является существенным. Любой рандомизированный критерий π можно представить как нерандомизированный, если ввести в рассмотрение дополнительное наблюдение Y, независимое от X и равномерно распределенное на $[0,1]$. Действительно, рассмотрим для новой выборки (X,Y) критическую область

$$\Omega = \{(\boldsymbol{x},y) : \pi(\boldsymbol{x}) \geqslant y\},$$

т. е. положим $\delta(X,Y) = 1$, если $(X,Y) \in \Omega$, и $\delta(X,Y) = 0$ в противном случае. Тогда при любом распределении X

$$\mathbf{P}(\delta(X,Y)=1) = \mathbf{P}(\pi(X) \geqslant Y) = \int_0^1 \mathbf{P}(\pi(X) \geqslant y)\,dy = \mathbf{E}\pi(X)$$

и, стало быть, критерий δ по своим параметрам эквивалентен π.

Как воспользоваться этим обстоятельством для построения доверительных интервалов в условиях теоремы 3? Допустим для простоты, что статистика $T(X)$ целочисленна (как мы видели, отсутствие нерандомизированных р. н. м. к. может быть вызвано лишь дискретностью распределения T). Тогда наблюдение $S(X,Y) = T(X) + Y$, $Y \in \mathbf{U}_{0,1}$, сохраняет всю информацию, содержащуюся в $T(X)$, так как $T(X)$ есть целая часть $S(X,Y)$. Р. н. м. к.

уровня $1-\varepsilon$ можно придать выбором нецелого $c(\theta,\varepsilon)$ следующую форму: принимается гипотеза H_1, если

$$S(X,Y)\leqslant c(\theta_1,1-\varepsilon).$$

Тем самым мы построили требуемые множества $\Omega(\theta,\varepsilon)$ и остается лишь «обратить» их тем же способом, что и прежде. Получим нижнюю границу

$$\theta^-(X,Y,\varepsilon)=c^{-1}(T(X)+Y,1-\varepsilon),$$

где c^{-1} — функция, обратная к c по первому аргументу. Здесь из самой формы записи видно, что для определения θ^- нужно произвести дополнительное наблюдение Y.

Пример 1. Пусть $X\in\mathbf{B}_p$, и нас интересует верхняя доверительная граница p^+ уровня $1-\varepsilon$ для вероятности $p=\mathbf{P}(\mathrm{x}_i=1)=1-\mathbf{P}(\mathrm{x}_i=0)$. Семейство распределений $\{\mathbf{B}_p\}$ экспоненциально и удовлетворяет условиям теоремы 3, где следует положить $T(X)=\sum_{i=1}^{n}\mathrm{x}_i$. Рассмотрим наблюдение

$$S=\sum_{i=1}^{n}\mathrm{x}_i+Y,\quad Y\in\mathbf{U}_{0,1}.$$

Оно имеет в точке t, $0\leqslant t\leqslant n+1$, плотность $C_n^{[t]}p^{[t]}(1-p)^{n-[t]}$. Обозначим через $G_p(t)$ функцию распределения с этой плотностью. Тогда p^+ будет решением уравнения $G_p(t)=\varepsilon$.

3. Несмещенные доверительные множества. Вернемся к вопросу о наиболее точных доверительных множествах. С помощью теоремы 3 можем строить наиболее точные верхние и нижние границы, основываясь на том, что для односторонних альтернатив $\{\theta>\theta_1\}$, $\{\theta<\theta_1\}$ гипотез $\{\theta=\theta_1\}$ в ряде случаев существует р. н. м. к. Если же попытаться использовать теоремы 1, 2 непосредственно при построении наиболее точных доверительных интервалов, то нам потребуется существование р. н. м. к. для проверки гипотезы $\{\theta=\theta_1\}$ против $\{\theta\neq\theta_1\}$, что бывает весьма редко. Выход из этого положения состоит в естественном сужении класса рассматриваемых доверительных интервалов на той же основе, на которой мы сужали классы рассматриваемых критериев в § 47, 48. Именно, введем понятия несмещенных и инвариантных доверительных множеств.

Пусть, как и прежде, каждому θ соответствует множество $\Theta_2(\theta)$, $\theta\notin\Theta_2(\theta)$.

Определение 4. Доверительное множество $\Theta^*(X,\varepsilon)$ для θ уровня $1-\varepsilon$ называется *несмещенным относительно альтернатив θ' таких, что* $\theta\in\Theta_2(\theta')$, если

$$\mathbf{P}_\theta(\theta'\in\Theta^*(X,\varepsilon))\leqslant 1-\varepsilon\quad\text{при всех }\theta,\theta',\quad\theta\in\Theta_2(\theta'). \tag{6}$$

Множество $\Theta^*(X,\varepsilon)$ называется просто *несмещенным*, если (6) справедливо для всех $\theta'\neq\theta$.

Несмещенность доверительного множества означает, что *вероятность покрытия им ложного значения θ' не больше, чем истинного.*

Определение 5. Доверительные множества, для которых минимизируется (4) при условиях (1), (6), называются *наиболее точными несмещенными доверительными множествами* (уровня $1-\varepsilon$) относительно альтернатив, для которых $\theta \in \Theta_2(\theta')$.

Теорема 4. 1) *Несмещенные нерандомизированные критерии порождают в силу эквивалентности* (3) *несмещенные доверительные множества, и наоборот.*

2) *Если $\Omega(\theta_1, \varepsilon)$ при каждом $\theta_1 \in \Theta$ есть область принятия гипотезы $\{\theta = \theta_1\}$ нерандомизированного р. н. м. несмещенного критерия при альтернативе $\{\theta \in \Theta_2(\theta_1)\}$, то соответствующее множество $\Theta^*(X, \varepsilon)$ будет наиболее точным несмещенным доверительным множеством, и наоборот.*

Доказательство теоремы полностью повторяет рассуждения теоремы 1, к которым надо добавить лишь, что свойство несмещенности при переходе от критериев к доверительным множествам и обратно сохраняется. Действительно, соотношения (1), (6) эквивалентны

$$\sup_{\theta \in \Theta_2(\theta_1)} \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)) \leqslant 1 - \varepsilon \leqslant \mathbf{P}_{\theta_1}(X \in \Omega(\theta_1, \varepsilon)).$$

Если $\pi(X)$ есть критическая функция нерандомизированных критериев, фигурирующих в теореме ($\pi(X) = 0$ при $X \in \Omega(\theta_1, \varepsilon)$), то мы получаем

$$\mathbf{E}_\theta \pi(X) = 1 - \mathbf{P}_\theta(X \in \Omega(\theta_1, \varepsilon)), \qquad \inf_{\theta \in \Theta_2(\theta_1)} \mathbf{E}_\theta \pi(X) \geqslant \varepsilon \geqslant \mathbf{E}_{\theta_1} \pi(X).$$

Это и есть, очевидно, свойство несмещенности, эквивалентное (6). $\triangleleft$

Если воспользоваться результатами § 46 и построить несмещенное наиболее доверительное множество для параметра θ экспоненциального семейства, то мы получим тот же доверительный интервал (θ^-, θ^+), который мы строили, используя монотонность отношения правдоподобия, т. е. интервал, у которого θ^- и θ^+ есть соответственно наиболее точные нижняя и верхняя границы уровней $1 - \varepsilon/2$.

4. Инвариантные доверительные множества. Следующее определение использует обозначения и понятия предыдущего параграфа. Пусть $\{\mathbf{P}_\theta\}$ — инвариантное семейство относительно G.

Определение 6. Доверительное множество $\Theta^*(X, \varepsilon)$ называется *инвариантным**) *относительно группы G*, если

$$\Theta^*(gX, \varepsilon) = \overline{g}\Theta^*(X, \varepsilon) \tag{7}$$

для всех $g \in G$.

*) Если следовать замечанию, сделанному в сноске в § 29, то доверительное множество со свойством (7) было бы естественнее называть *эквивариантным*.

Смысл этого понятия аналогичен смыслу эквивариантной оценки (§ 29). Если преобразования g и $\overline{g}$ интерпретировать как замену системы координат, сохраняющую распределение, то (7) означает, что доверительное множество не зависит от координатной системы, в которой выражены исходные данные.

Определение 7. Доверительное множество $\Theta^*(X,\varepsilon)$ называется *наиболее точным инвариантным доверительным множеством* уровня $1-\varepsilon$, если на нем минимизируется $\mathbf{P}_\theta(\theta' \in \Theta^*(X,\varepsilon))$ при всех $\theta' \neq \theta$ в классе всех множеств Θ^*, удовлетворяющих (7) и условию $\mathbf{P}_\theta(\theta \in \Theta^*(X,\varepsilon)) = 1-\varepsilon$.

Пусть $\Omega(\theta_1,\varepsilon)$ есть область приема гипотезы $H_1 = \{\theta = \theta_1\}$ при альтернативе $\{\theta \neq \theta_1\}$ для инвариантного критерия уровня $1-\varepsilon$. Отметим, что есть существенное различие в определениях инвариантного критерия и инвариантного доверительного множества (его не было бы, если бы требовалось выполнение равенства $g\Omega(\theta,\varepsilon) = \Omega(\overline{g}\theta,\varepsilon)$, а не $g\Omega(\theta,\varepsilon) = \Omega(\theta,\varepsilon)$). С этим фактом связано то обстоятельство, что соответствие между р. н. м. инвариантными критериями и наиболее точными инвариантными доверительными интервалами выглядит несколько сложнее, чем в предыдущих теоремах.

Рассмотрим группу преобразований G и предположим, что для каждого θ_1 в ней есть подгруппа $G[\theta_1]$, оставляющая инвариантной задачу проверки гипотезы $H_1 = \{\theta = \theta_1\}$. Другими словами, $\overline{g}\theta_1 = \theta_1$ при $g \in G[\theta_1]$.

Теорема 5. *Пусть* $\Theta^*(X,\varepsilon)$ — *инвариантное относительно* G *доверительное множество уровня* $1-\varepsilon$. *Тогда*:

1) *область* $\Omega(\theta,\varepsilon) = \{\boldsymbol{x} : \theta \in \Theta^*(\boldsymbol{x},\varepsilon)\}$ *будет инвариантной относительно* $G[\theta]$ *при каждом* θ;

2) *если область* $\Omega(\theta_1,\varepsilon)$, *соответствующая* $\Theta^*(X,\varepsilon)$, *есть область приема* H_1 *при альтернативе* $\{\theta \neq \theta_1\}$ *для р. н. м. инвариантного критерия уровня* $1-\varepsilon$, *то* $\Theta^*(X,\varepsilon)$ *будет наиболее точным доверительным инвариантным множеством.*

Доказательство. 1) Пусть $g \in G[\theta]$. Тогда $\overline{g}\theta = \theta$,

$$g\Omega(\theta,\varepsilon) = \{g\boldsymbol{x} : \theta \in \Theta^*(\boldsymbol{x},\varepsilon)\} =$$
$$= \{\boldsymbol{x} : \theta \in \Theta^*(g^{-1}\boldsymbol{x},\varepsilon)\} = \{\boldsymbol{x} : \theta \in \overline{g}^{-1}\Theta^*(\boldsymbol{x},\varepsilon)\} =$$
$$= \{\boldsymbol{x} : \overline{g}\theta \in \Theta^*(\boldsymbol{x},\varepsilon)\} = \{\boldsymbol{x} : \theta \in \Theta^*(\boldsymbol{x},\varepsilon)\} = \Omega(\theta,\varepsilon).$$

2) Пусть $\widetilde{\Theta}^*$ — любое другое инвариантное доверительное множество уровня $1-\varepsilon$. Согласно первому утверждению ему соответствует инвариантный критерий уровня $1-\varepsilon$ с областью $\widetilde{\Omega}(\theta_1,\varepsilon)$ приема H_1.

Так как по предположению

$$\mathbf{P}_\theta(X \in \Omega(\theta_1,\varepsilon)) \geqslant \mathbf{P}_\theta(X \in \widetilde{\Omega}(\theta_1,\varepsilon)),$$

то

$$\mathbf{P}_\theta(\theta_1 \in \Theta^*(X,\varepsilon)) \geqslant \mathbf{P}_\theta(\theta_1 \in \widetilde{\Theta}^*(X,\varepsilon))$$

при $\theta_1 \neq \theta$, что и требовалось доказать. $\triangleleft$

Пример 2. Пусть $X \in \Phi_{\alpha,\sigma^2}$. Требуется построить наиболее точное доверительное множество для параметра σ^2 при неизвестном α. Мы видели в примере 2 предыдущего параграфа, что семейство Φ_{α,σ^2} инвариантно относительно преобразований сдвига $gX = X + c$, если положить $\overline{g}(\alpha,\sigma^2) = (\alpha + c, \sigma^2)$. Статистика $S_0^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2$ является максимальным инвариантом, построенным по достаточной статистике. Кроме того, гипотеза $H_1 = \{\sigma = \sigma_1\}$ инвариантна относительно G. Согласно примеру 47.2 р. н. м. инвариантный несмещенный критерий для проверки H_1 имеет вид

$$h_{1,\varepsilon}\sigma_1^2 < (n-1)S_0^2 < h_{2,\varepsilon}\sigma_1^2, \tag{8}$$

где $h_{i,\varepsilon}$ определяются из условий (см. условие (46.7) теоремы 46.1)

$$\mathbf{P}\big(h_{1,\varepsilon} < \chi_{n-1}^2 < h_{2,\varepsilon}\big) = 1 - \varepsilon,$$

$$\mathbf{E}\big(\chi_{n-1}^2; h_{1,\varepsilon} < \chi_{n-1}^2 < h_{2,\varepsilon}\big) = (1-\varepsilon)\mathbf{E}\chi_{n-1}^2,$$

$$\chi_{n-1}^2 \in \mathbf{H}_{n-1}.$$

Соответствующее (8) доверительное множество $\Theta^*(X,\varepsilon)$ имеет вид интервала

$$\frac{(n-1)S_0^2}{h_{2,\varepsilon}} < \sigma^2 < \frac{(n-1)S_0^2}{h_{1,\varepsilon}}. \tag{9}$$

Этот интервал, очевидно, инвариантен относительно g, как и критерий (8) (в этом примере $G[\sigma_1] \equiv G$ для любого σ_1). Следовательно, в силу вторых утверждений теорем 4, 5 интервал (9) есть наиболее точное инвариантное несмещенное доверительное множество уровня $1 - \varepsilon$.

Пример 3. Пусть $X \in \Phi_{\alpha,\sigma^2}$. Требуется построить наиболее точное доверительное множество для параметра α при неизвестном σ. Здесь

$$f_{\alpha,\sigma^2}(X) = \frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mathrm{x}_i - \alpha)^2\right\}.$$

Семейство Φ_{α,σ^2} будет инвариантно относительно группы G линейных преобразований $gX = aX + b$, если положить $\overline{g}(\alpha,\sigma) = (a\alpha + b, a\sigma)$. Пара наблюдений $\big(\overline{\mathrm{x}}, S_0^2\big)$ образует достаточную статистику. Нетрудно видеть, что с ее помощью нельзя построить статистику, инвариантную относительно G. Однако для каждого α_1 можно выделить подгруппу $G[\alpha_1]$ преобразований $gX = a(X - \alpha_1) + \alpha_1$, относительно которой статистика $(\overline{\mathrm{x}} - \alpha_1)/S_0$ будет максимальным инвариантом. Гипотеза $H_1 = \{\alpha = \alpha_1\}$ остается инвариантной относительно $G[\alpha_1]$. Исследуя плотность $(\overline{\mathrm{x}} - \alpha_1)/S_0$, можно показать с помощью методов § 47 (ввиду сложности этих рассмотрений мы их опускаем*)), что при каждом σ р. н. м. несмещенный инвариантный критерий для проверки гипотезы H_1 против $\{\alpha \neq \alpha_1\}$ существует и имеет область

*) Подробнее об этом см. [65, с. 312].

приема H_1 вида

$$\frac{\sqrt{n}\,|\overline{\mathrm{x}} - \alpha_1|}{S_0} < \tau_\varepsilon, \tag{10}$$

где τ_ε определяется из условия $\mathbf{P}(|t_{n-1}| \geqslant \tau_\varepsilon) = \varepsilon$, $t_{n-1} \in \mathbf{T}_{n-1}$.

Соответствующее доверительное множество Θ^* имеет вид

$$\overline{\mathrm{x}} - \frac{\tau_\varepsilon S_0}{\sqrt{n}} < \alpha < \overline{\mathrm{x}} + \frac{\tau_\varepsilon S_0}{\sqrt{n}}. \tag{11}$$

Легко видеть, что этот доверительный интервал является инвариантным ($\Theta^*(gX, \varepsilon) = \overline{g}\Theta^*(X, \varepsilon)$). В соответствии с первым утверждением теоремы 5 критерий (10) будет инвариантным относительно $G[\alpha_1]$. Согласно второму утверждению доверительный интервал (11) будет наиболее точным (равномерно по σ) инвариантным несмещенным доверительным интервалом уровня $1 - \varepsilon$.

Таким образом, в этом параграфе мы установили, что все доверительные интервалы, построенные нами в § 40, *являются* в известном смысле оптимальными.

§ 49. Байесовский и минимаксный подходы к проверке сложных гипотез

1. Байесовские и минимаксные критерии. Байесовский и минимаксный подходы были описаны в § 44. Там же были даны соответствующие определения, которые мы по ходу изложения напомним.

Пусть, как и прежде, проверяется гипотеза $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$ по выборке $X \in \mathbf{P}_\theta$.

Полный байесовский подход предполагает, что θ выбирается случайно с априорным распределением $\mathbf{Q}$ на $\Theta = \Theta_1 \cup \Theta_2$. Распределение $\mathbf{Q}$ индуцирует распределения $\mathbf{Q}_i$ на Θ_i, $i = 1, 2$, и вероятности $q(i) = \mathbf{Q}(\theta \in \Theta_i)$, так что $\mathbf{Q} = q(1)\mathbf{Q}_1 + q(2)\mathbf{Q}_2$. Гипотезу о том, что $\theta \in \Theta_i$ выбирается случайным образом с распределением $\mathbf{Q}_i$, обозначим через H_{Q_i}. В соответствии с этой гипотезой X имеет плотность

$$f_{Q_i}(\boldsymbol{x}) = \int f_\theta(\boldsymbol{x})\mathbf{Q}_i(d\theta).$$

Подразумевается, конечно (см. § 44), что на Θ_i определены σ-алгебры $\mathfrak{S}_i$, на которых заданы $\mathbf{Q}_i$, и что $f_\theta(\boldsymbol{x})$ измерима относительно $\mathfrak{S}_1 \times \mathfrak{B}^n_{\mathcal{X}}$.

Из результатов § 41, 42 следует, что байесовский критерий π_Q для проверки H_{Q_1} против H_{Q_2} в описанной выше задаче будет иметь вид

$$\pi_Q(X) = \begin{cases} 1, & \text{если} \quad f_{Q_2}(X) > c f_{Q_1}(X), \\ p, & \text{если} \quad f_{Q_2}(X) = c f_{Q_1}(X), \\ 0, & \text{если} \quad f_{Q_2}(X) < c f_{Q_1}(X), \end{cases} \tag{1}$$

где $c = q(1)/q(2)$, $p \in [0, 1]$ произвольно.

Частично байесовский подход связан с проверкой гипотезы H_{Q_1} против H_{Q_2} в том случае, когда априорное распределение между H_{Q_1} и H_{Q_2} (определяемое вероятностями $q(1)$, $q(2)$) отсутствует. Положим

$$K_\varepsilon^{Q_1} = \{\pi : \mathbf{E}_{Q_1}\pi(X) \leqslant \varepsilon\}.$$

Тогда критерий $\pi_{Q_1Q_2}$ мы называем *байесовским* в $K_\varepsilon^{Q_1}$, если это есть н. м. к. уровня $1-\varepsilon$ для проверки H_{Q_1} против H_{Q_2}. Критерий $\pi_{Q_1Q_2}$ будет иметь тот же вид (1), где c, p выбираются из условия $\mathbf{E}_{Q_1}\pi_{Q_1Q_2}(X) = \varepsilon$.

Вместо $\pi_{Q_1Q_2}$ мы будем писать π_{Q_1} или π_{Q_2}, если одно из множеств Θ_1 или Θ_2 вырождается в одноточечное множество $\{\theta_1\}$ или $\{\theta_2\}$.

В приложениях не часто встречаются задачи, в которых распределения $\mathbf{Q}_i$ полностью известны. Однако, как мы уже не раз видели, польза байесовского подхода не ограничивается лишь возможностью его непосредственного применения. Он позволяет строить р. н. м. к., а также минимаксные критерии (ср. с § 41, 45, 46). Позже мы будем использовать байесовский подход также для построения асимптотически оптимальных критериев. Пусть, как и раньше,

$$K_\varepsilon = \{\pi : \sup_{\theta\in\Theta_1} \mathbf{E}_\theta\pi(X) \leqslant \varepsilon\}. \tag{2}$$

Тогда критерий $\overline{\pi}$ мы называем *минимаксным* в K_ε (в $K_\varepsilon^{Q_1}$), если $\overline{\pi} \in K_\varepsilon \left(\overline{\pi} \in K_\varepsilon^{Q_1}\right)$ и для него максимизируется

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta\pi(X) = \inf_{\theta\in\Theta_2} \beta(\theta). \tag{3}$$

Следует отметить, что если функции мощности $\beta(\theta) = \mathbf{E}_\theta\pi(X)$ непрерывны и множества Θ_1 и Θ_2 соприкасаются, то

$$\beta = \sup_{\pi\in K_\varepsilon} \inf_{\theta\in\Theta_2} \beta(\theta) \leqslant \varepsilon \tag{4}$$

и неравенство $\beta > \varepsilon$ выполняться не может. Поэтому если желать, чтобы гарантированная мощность (3) была достаточно большой (во всяком случае больше ε), то следует рассматривать «разделенные» множества Θ_1 и Θ_2. Другими словами, зону значений θ, где $\beta(\theta)$ близка к ε, как зону «безразличия» критериев следует изъять и рассматривать в качестве Θ_2 множество, не соприкасающееся с Θ_1.

Если все же множества соприкасаются, то *любой несмещенный критерий в K_ε будет минимаксным.* Действительно, для несмещенных критериев $\beta(\theta) = \mathbf{E}_\theta\pi(X) \geqslant \varepsilon$, $\theta \in \Theta_2$, и, стало быть, $\beta = \inf\limits_{\theta\in\Theta_2} \beta(\theta) \geqslant \varepsilon$ достигает в силу (4) своего максимального значения.

Обратное утверждение верно в общем случае: *минимаксный критерий, если он существует, является несмещенным.* Это следует из того, что

$$\beta = \sup_{\pi\in K_\varepsilon} \inf_{\theta\in\Theta_2} \beta(\theta) \geqslant \varepsilon$$

(мы можем взять $\pi(X) \equiv \varepsilon$), и того, что для минимаксного критерия

$$\inf_{\theta\in\Theta_2} \beta(\theta) = \beta.$$

Р. н. м. несмещенный критерий $\check{\pi}$ *в классе* $\check{K}_\varepsilon$ *всех несмещенных критериев является минимаксным в* K_ε. Действительно, пусть $\check{\beta}(\theta)$ — функция мощности критерия $\check{\pi}$. Тогда для любых $\pi \in \check{K}_\varepsilon$, $\theta \in \Theta_2$

$$\check{\beta}(\theta) \geqslant \beta(\theta), \quad \inf_{\theta\in\Theta_2} \check{\beta}(\theta) \geqslant \inf_{\theta\in\Theta_2} \beta(\theta),$$
$$\inf_{\theta\in\Theta_2} \check{\beta}(\theta) = \sup_{\pi\in\check{K}_\varepsilon} \inf_{\theta\in\Theta_2} \beta(\theta) = \sup_{\pi\in K_\varepsilon} \inf_{\theta\in\Theta 2} \beta(\theta). \tag{5}$$

Последнее равенство объясняется тем, что добавление к $\check{K}_\varepsilon$ критериев из K_ε, для которых $\inf\limits_{\theta\in\Theta_2} \beta(\theta) < \varepsilon$, величины $\sup\limits_{\pi\in\check{K}_\varepsilon}$ в (5) не меняет. $\triangleleft$

В теореме 45.3 мы использовали байесовские критерии для отыскания р. н. м. к. Следующее утверждение есть некоторое «развитие» теоремы 45.3. Оно является также аналогом теорем 41.2, 21.2 и устанавливает, что минимаксные критерии следует искать в классе критериев (1), явный вид которых нам известен.

Теорема 1. *Пусть существуют распределения* $\mathbf{Q}_i$, *сосредоточенные соответственно на подмножествах* $\Theta_i^0 \subset \Theta_i$, $i = 1, 2$, *и постоянные* c *и* r *такие, что критерий* $\pi_{Q_1Q_2}$, *определенный в* (1), *обладает свойствами:*

1) $$\pi_{Q_1Q_2} \in K_\varepsilon^{Q_1},$$

2) $$\mathbf{E}_\theta \pi_{Q_1Q_2}(X) = \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \tag{6}$$

при всех $\theta \in \Theta_1^0$,

3) $$\mathbf{E}_\theta \pi_{Q_1Q_2}(X) = \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \tag{7}$$

при всех $\theta \in \Theta_2^0$.

Тогда $\pi_{Q_1Q_2} \in K_\varepsilon$ *и является минимаксным критерием в* K_ε *для проверки* H_1 *против* H_2.

Пара распределений $\mathbf{Q}_1$, $\mathbf{Q}_2$, *обладающая свойствами* 2), 3), *является наименее благоприятной в том смысле, что для любых двух других распределений* $\mathbf{Q}_1'$, $\mathbf{Q}_2'$

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2} \leqslant \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1'Q_2'},$$

где $\pi_{Q_1'Q_2'}$ — *критерий вида* (1) *из* K_ε.

Последнее утверждение означает, что среди всех байесовских критериев (1) критерий $\pi_{Q_1Q_2}$ обладает наименьшей гарантированной мощностью.

Доказательство. Так как

$$\sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) = \int_{\Theta_1^0} \mathbf{E}_\theta \pi_{Q_1Q_2} \mathbf{Q}_1(d\theta) = \mathbf{E}_{Q_1} \pi_{Q_1Q_2} = \varepsilon,$$

то $\pi_{Q_1Q_2} \in K_\varepsilon$. Гарантированная мощность $\pi_{Q_1Q_2}$ равна (см. (7))

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) = \int\limits_{\Theta_2^0} \mathbf{E}_\theta \pi_{Q_1Q_2} \mathbf{Q}_2(d\theta) = \mathbf{E}_{Q_2} \pi_{Q_1Q_2} \equiv \beta_{Q_1Q_2}. \tag{8}$$

Пусть теперь π — любой другой критерий из K_ε для проверки H_1 против H_2. Тогда π будет одновременно критерием из $K_\varepsilon^{Q_1}$ для проверки H_{Q_1} против H_{Q_2}, так как

$$\mathbf{E}_{Q_1}\pi(X) = \int\limits_{\Theta_1^0} \mathbf{E}_\theta \pi(X) \mathbf{Q}_1(d\theta) \leqslant \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi(X) \leqslant \varepsilon. \tag{9}$$

Но критерий $\pi_{Q_1Q_2}$ является н. м. к. в $K_\varepsilon^{Q_1}$ для проверки H_{Q_1} против H_{Q_2}. Стало быть, в силу (8)

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) = \beta_{Q_1Q_2} \geqslant \mathbf{E}_{Q_2}\pi(X) \geqslant \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi(X). \tag{10}$$

Первое утверждение теоремы доказано. Пусть теперь $\mathbf{Q}_1'$, $\mathbf{Q}_2'$ — два любых других распределения на Θ_1, Θ_2 соответственно. Критерий $\pi_{Q_1Q_2}$, как и $\pi_{Q_1'Q_2'}$, будет критерием из $K_\varepsilon^{Q_1'}$ для проверки $H_{Q_1'}$ против $H_{Q_2'}$, так как

$$\mathbf{E}_{Q_1'} \pi_{Q_1Q_2}(X) = \int\limits_{\Theta_1} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \mathbf{Q}_1'(d\theta) \leqslant \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \leqslant \varepsilon.$$

Но критерий $\pi_{Q_1'Q_2'}$ является н. м. к. для этих гипотез, поэтому в силу (8)

$$\beta_{Q_1'Q_2'} \equiv \mathbf{E}_{Q_2'} \pi_{Q_1'Q_2'}(X) \geqslant \mathbf{E}_{Q_2'} \pi_{Q_1Q_2}(X) =$$
$$= \int\limits_{\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \mathbf{Q}_2'(d\theta) \geqslant \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_{Q_1Q_2}(X) \equiv \beta_{Q_1Q_2}. \quad \triangleleft$$

Основная трудность в применении теоремы 1 к реальным задачам состоит в отыскании (или угадывании) наименее благоприятных распределений $\mathbf{Q}_1$ и $\mathbf{Q}_2$. Иногда при этом могут оказаться полезными соображения инвариантности, как это имеет место в примерах следующего пункта. Эти примеры представляют самостоятельный интерес и будут использоваться в дальнейшем.

2. Минимаксные критерии для параметра α нормальных распределений

Пример 1. Пусть $X = \mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,E}$ есть выборка объема $n = 1$ из m-мерного нормального распределения со средним $\alpha = (\alpha_1, \ldots, \alpha_m)$ и единичной матрицей вторых моментов. Положим $|\alpha|^2 = \sum\limits_{i=1}^m \alpha_i^2$ и рассмотрим задачу о проверке гипотезы $H_1 = \{|\alpha| \leqslant a\}$ против $H_2 = \{|\alpha| \geqslant b\}$, $b > a$ (здесь имеется «разделяющая» зона $a < |\alpha| < b$).

Если, например, X определяет в канале связи амплитуды вектор-сигнала, состоящего из «шума» $X_0 \in \Phi_{0,E}$ и полезного сигнала α, $|\alpha| \geqslant b$, то гипотезы H_i при $a = 0$ можно рассматривать как гипотезы о наличии полезного сигнала.

Так как рассматриваемый пример будет неоднократно использоваться нами в дальнейшем, то утверждение, относящееся к виду минимаксного критерия, мы сформулируем в виде теоремы.

Теорема 2. *Минимаксный критерий* $\overline{\pi} \in K_\varepsilon$ *для проверки* $H_1 = \{|\alpha| \leqslant a\}$ *против* $H_2 = \{|\alpha| \geqslant b\}$, $a < b$, *по наблюдению* $X \in \Phi_{\alpha,E}$ *имеет вид*

$$\pi(X) = \begin{cases} 1, & \text{если} \quad |X| > c_\varepsilon, \\ 0, & \text{если} \quad |X| \leqslant c_\varepsilon, \end{cases}$$

где c_ε *выбирается из условия* $p_c(a) = \varepsilon$, *гарантированная мощность* $\overline{\pi}$ *равна* $p_{c_\varepsilon}(b)$,

$$p_c(t) = \mathbf{P}\big((\xi_1 - t)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2\big),$$

$\xi_i \in \Phi_{0,1}$ *и независимы.*

Доказательство. Начнем с наводящих соображений. В нашем случае для $x = (x^{(1)}, \ldots, x^{(m)})$ имеем

$$f_\alpha(x) = \frac{1}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}(x - \alpha)(x - \alpha)^\top\right\},$$

где $x^\top$ — вектор-столбец. Отсюда видно, что рассматриваемое семейство распределений инвариантно относительно ортогонального преобразования $gx = xC$, где C есть матрица ортогонального преобразования в $\mathbb{R}^m$. При этом надо положить $\overline{g}\alpha = \alpha C$. Гипотезы H_i будут инвариантными относительно $\overline{g}$.

Пусть для простоты $a = 0$. Если бы распределение $\mathbf{Q}_2$ на $\Theta_2 = \{\alpha : |\alpha| \geqslant b\}$ не обнаруживало инвариантности относительно $\overline{g}$ (так будет, например, если оно сосредоточено в окрестности какой-нибудь точки α_0), эту несимметрию можно было бы как-то использовать при решении задачи (при сделанном только что допущении мы были бы близки к задаче проверки двух простых гипотез $\{\alpha = 0\}$ и $\{\alpha = \alpha_0\}$ и получили бы при этом критерий с большой мощностью). Стало быть, такое распределение не может быть наименее благоприятным. Им должно быть распределение $\mathbf{Q}_2$, инвариантное относительно $\overline{g}$. Кроме того, ясно, что наихудший вариант мы получим, если все распределение будет сосредоточено на границе Θ_2 (чем ближе гипотезы, тем труднее их различать). Аналогичные наводящие рассуждения можно проводить относительно $\mathbf{Q}_1$, если $a \neq 0$.

Итак, представляется естественным, что наименее благоприятные распределения $\mathbf{Q}_1$ и $\mathbf{Q}_2$ в нашем примере есть равномерные распределения на сферах $\Theta_1^0 = \{\alpha : |\alpha| = a\}$ и $\Theta_2^0 = \{\alpha : |\alpha| = b\}$. В этом случае согласно теореме 1 минимаксный критерий $\overline{\pi}$ будет иметь вид $\overline{\pi}(x) = \pi_{Q_1 Q_2}(x)$,

где $\pi_{Q_1Q_2}(x) = 1$, если

$$\int\limits_{\Theta_2^0} \exp\left\{-\frac{1}{2}(x-v)(x-v)^\top\right\}\frac{dV(v)}{V_2} > c\int\limits_{\Theta_1^0}\exp\left\{-\frac{1}{2}(x-v)(x-v)^\top\right\}\frac{dV(v)}{V_1} \tag{11}$$

и $\pi_{Q_1Q_2}(x) = 0$ в противном случае. Здесь $dV(v)$ означает площадь элемента соответствующей сферы, $V_i = \operatorname{mes}\Theta_i^0$, $i = 1, 2$.

Рассмотрим любой из этих интегралов, например правый, и заметим, что его можно записать в виде

$$\exp\left\{-\frac{1}{2}xx^\top - a^2\right\}\int\limits_{\Theta_1^0}\exp\{xv^\top\}\frac{dV(v)}{V_1}.$$

Интеграл здесь равен

$$\int\limits_{\Theta^0}\exp\{|x|ae_xv^\top\}\frac{dV(v)}{V}, \quad V = \operatorname{mes}\Theta^0,$$

где Θ^0 — поверхность единичного шара, $e_x = x/|x|$. Стало быть, если обозначить

$$\psi(t) = \int\limits_{\Theta^0}\exp\{te_xv^\top\}\,dV(v), \tag{12}$$

то область (11) приема H_2 примет форму

$$\psi(|x|b) > c\psi(|x|a) \tag{13}$$

(через c обозначаем здесь постоянные, не обязательно совпадающие со значением в (11)). Но $\psi(t)$, очевидно, от x не зависит, поскольку значение интеграла (12) не зависит от направления единичного вектора e_x. Поэтому

$$\psi(t) = \int\limits_{\Theta^0}\exp\{tv_1\}\,dV(v),$$

где v_1 есть первая координата вектора v.

Так как $\psi'(0) = 0$, $\psi''(t) > 0$ при $t > 0$, то $\psi(t)$ есть выпуклая возрастающая функция на $[0, \infty)$. Отсюда следует, что неравенство (13) (или (11)) эквивалентно

$$|x| > c. \tag{14}$$

Это есть, очевидно, инвариантный критерий. Проверим для него выполнение условий 1)–3) теоремы 1 и установим тем самым, что это минимаксный критерий.

Имеем

$$\mathbf{E}_\alpha\pi_{Q_1Q_2}(X) = \mathbf{P}_\alpha(|X| > c) = \mathbf{\Phi}_{0,E}(\{x : |x-\alpha| > c\}).$$

Ясно, что перемещение точки α по сфере $|\alpha| = \text{const}$ этой вероятности не меняет. Стало быть, она зависит лишь от $|\alpha|$ и, следовательно,

$$\mathbf{E}_\alpha \pi_{Q_1Q_2}(X) = \mathbf{P}(|\xi - \alpha|^2 > c^2) =$$
$$= \mathbf{P}\left(\sum_{i=1}^{m}(\xi_i - \alpha_i)^2 > c^2\right) = \mathbf{P}\big((\xi_1 - |\alpha|)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2\big),$$

где $\xi_i \in \mathbf{\Phi}_{0,1}$ есть независимые координаты вектора ξ.

Лемма 1. *Функция* $p_c(t) = \mathbf{P}\big((\xi_1 - t)^2 + \xi_2^2 + \ldots + \xi_m^2 > c^2\big)$ *при каждом* c *является возрастающей функцией* $|t|$.

Из этой леммы вытекает, что

$$\mathbf{E}_\alpha \pi_{Q_1Q_2}(X) = p_c(|\alpha|) \leqslant p_c(a) \quad \text{при} \quad |\alpha| \leqslant a,$$
$$\mathbf{E}_\alpha \pi_{Q_1Q_2}(X) = p_c(|\alpha|) \geqslant p_c(b) \quad \text{при} \quad |\alpha| \geqslant b,$$

Эти соотношения эквивалентны условиям 2), 3) теоремы 1. Чтобы критерий $\pi_{Q_1Q_2}$ был критерием уровня $1 - \varepsilon$, мы должны положить c равным решению c_ε уравнения $p_c(a) = \varepsilon$. Таким образом, $\pi_{Q_1Q_2}$ есть минимаксный критерий уровня $1 - \varepsilon$, и его гарантированная мощность равна $p_{c_\varepsilon}(b)$. $\triangleleft$

Доказательство леммы 1. Так как $p_c(t) = p_c(-t)$, то мы можем ограничиться рассмотрением значений $t \geqslant 0$.

Рассмотрим сначала случай $m = 1$. Обозначим функцию $p_c(t)$ в этом случае через $p(t)$. Имеем

$$p(t) = \mathbf{P}(|\xi_1 - t|^2 > c^2) = \Phi(t - c) + 1 - \Phi(t + c).$$

Стало быть, производная по t равна

$$p'(t) = \frac{1}{\sqrt{2\pi}}[e^{-(t-c)^2/2} - e^{-(c+t)^2/2}] = \frac{1}{\sqrt{2\pi}} e^{-(c^2+t^2)^2/2}[e^{ct} - e^{-ct}] \geqslant 0,$$

и функция $p(t)$ возрастает при $t \geqslant 0$.

При $m > 1$ функция $p_c(t)$ есть свертка функции $p(t) = p(t, c^2)$ с распределением χ^2 с $m - 1$ степенями свободы:

$$p_c(t) = \int_0^\infty p(t, c^2 - u)\, d\mathbf{H}_{m-1}(u).$$

Очевидно, это также есть возрастающая функция t при $t \geqslant 0$. $\triangleleft$

По поводу теоремы 2 можно заметить следующее. Пусть для простоты $a = 0$. Тогда гипотеза $H_1 = \{\alpha = 0\}$ станет простой. Если мы построим н. м. к. для каждой альтернативы $\alpha \in \Theta_2$, получим критерий вида

$$x\alpha^\top > c.$$

Это значит, что каждое направление $\alpha = \alpha_0 t$, $\alpha_0 \in \Theta_2^0$, $t \geqslant 1$, будет иметь свой н. м. к. уровня $1 - \varepsilon$

$$x\alpha_0^\top > c_\varepsilon, \tag{15}$$

где c_ε зависит только от ε, так как $\mathbf{E}_0(X\alpha_0^\top) = 0$, $\mathbf{D}_0(X\alpha_0^\top) = |\alpha_0|^2 = b$. Но критическая область минимаксного (инвариантного) критерия должна быть одинаково чувствительна по всем альтернативам. В соответствии с этим она имеет форму объединения полупространств (15), представляющего собой внешность сферы.

Пример 2. Пусть теперь $X = \mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,\sigma^2}$, где $\sigma^2 = \|\sigma_{ij}\|$ есть произвольная положительно определенная матрица вторых моментов. Рассмотрим задачу о проверке гипотезы $H_1 = \{\alpha\sigma^{-2}\alpha^\top \leqslant a^2\} = \{|\alpha\sigma^{-1}| \leqslant a\}$ против $H_2 = \{\alpha\sigma^{-2}\alpha^\top \geqslant b^2\} = \{|\alpha\sigma^{-1}| \geqslant b\}$, $a < b$. Из теоремы 2 вытекает

Теорема 2А. *Критическое множество минимаксного критерия уровня* $1 - \varepsilon$ *для проверки* H_1 *против* H_2 *имеет вид*

$$x\sigma^{-2}x^\top > c_\varepsilon^2$$

и гарантированную мощность $p_{c_\varepsilon}(b)$, *где* c_ε, *как и прежде, есть решение уравнения* $p_c(a) = \varepsilon$.

Доказательство. Положим $gx = x\sigma$ и заметим, что в силу (47.3)

$$\mathbf{\Phi}_{\alpha,E}(A) = \mathbf{\Phi}_{\overline{g}(\alpha,E)}(gA),$$

где $\overline{g}(\alpha, E) = (\alpha\sigma, \sigma^2)$. Для шара $A = \{x : |x| < c\}$ будем иметь

$$\begin{gathered} gA = \{y = x\sigma : xx^\top < c^2\} = \{y : y\sigma^{-2}y^\top < c^2\}, \\ \mathbf{\Phi}_{\alpha,E}(A) = \mathbf{\Phi}_{\alpha\sigma,\sigma^2}(\{x : x\sigma^{-2}x^\top < c^2\}). \end{gathered} \tag{16}$$

Множество $\{\alpha : |\alpha| \leqslant a\}$ переходит после преобразования $\overline{g}$ в множество $\{\beta = \alpha\sigma : \alpha\alpha^\top \leqslant a^2\} = \{\beta : \beta\sigma^{-2}\beta^\top \leqslant a^2\}$.

Таким образом, все соотношения, установленные нами в примере 1 для $\mathbf{\Phi}_{\alpha,E}(A)$ при $|\alpha| \leqslant a$ или $|\alpha| \geqslant b$, будут справедливы для $\mathbf{\Phi}_{\beta,\sigma^2}(\{x : x\sigma^{-2}x^\top < c^2\})$ при $|\beta\sigma^{-1}| \leqslant a$ или $|\beta\sigma^{-1}| \geqslant b$ соответственно. Это доказывает теорему 2А. $\triangleleft$

Пример 3. Рассмотрим снова выборку из нормального распределения $\mathbf{\Phi}_{\alpha,E}$ с единичной матрицей вторых моментов. Однако в отличие от примера 1 проверяемые гипотезы H_i будут касаться лишь части координат вектора α. Представим α в виде набора двух векторов $\alpha = (\alpha', \alpha'')$, где $\alpha' = (\alpha_1, \ldots, \alpha_l)$, $\alpha'' = (\alpha_{l+1}, \ldots, \alpha_m)$, и рассмотрим задачу о проверке гипотезы $H_1 = \{|\alpha''| \leqslant a\}$ против $H_2 = \{|\alpha''| \geqslant b\}$ по выборке $X = \mathrm{x}_1 = (\mathrm{x}_{1,1}, \ldots, \mathrm{x}_{1,m})$ объема $n = 1$. Величина α' при каждой из гипотез одна и та же и может принимать произвольное значение. Поступим точно так

же, как в примере 1, но выберем в качестве $\mathbf{Q}_1$, $\mathbf{Q}_2$ равномерные распределения на «сферах» $\Theta_1^0 = \{\alpha : |\alpha''| = a, \alpha' = \alpha_0'\}$, $\Theta_2^0 = \{\alpha : |\alpha''| = b, \alpha' = \alpha_0'\}$, где α_0' — какая-нибудь фиксированная точка. Если обозначить $\mathrm{x}_1' = (\mathrm{x}_{1,1}, \ldots, \mathrm{x}_{1,l})$, $\mathrm{x}_1'' = (\mathrm{x}_{1,l+1}, \ldots, \mathrm{x}_{1,m})$, то мы получим в результате минимаксный критерий

$$|\mathrm{x}_1''| > c_\varepsilon,$$

где c_ε есть решение уравнения

$$\mathbf{P}\big((\xi_1 - a)^2 + \xi_2^2 + \ldots + \xi_{m-l}^2 > c^2\big) = \varepsilon \tag{17}$$

(множители $\exp\left\{-\frac{1}{2}(\mathrm{x}_1' - \alpha_0')(\mathrm{x}_1' - \alpha_0')^\top\right\}$ в неравенстве $f_{Q_2}(X)/f_{Q_1}(X) > c$ сократятся, и оно перейдет в неравенство типа (11)). Этот результат является вполне естественным, так как координаты $\mathrm{x}_{1,j}$ в нашем случае независимы и, стало быть, подвектор x_1' не несет в себе никакой информации относительно α''. Поэтому из всей выборки $X = \mathrm{x}_1$ достаточно рассматривать лишь подвектор x_1'', а в этом случае задача сводится к примеру 1.

Проверка гипотез в примере 3 относится к классу задач при наличии так называемого «мешающего» параметра. В нашем случае мешающим параметром служил вектор α'. В силу отмеченных выше причин он на самом деле практически не мешал построению минимаксного критерия, который автоматически получился не зависящим от α'.

Несколько иначе дело обстоит в следующем более общем примере, когда координаты x_{1j} зависимы.

Пример 4. Пусть $X = \mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,\sigma^2}$. Рассмотрим задачу о проверке гипотезы

$$H_1 = \{\alpha d^{-2}\alpha^\top \leqslant a^2\} \quad \text{против} \quad H_2 = \{\alpha d^{-2}\alpha^\top \geqslant b^2\}, \tag{18}$$

где d^{-2} — неотрицательно определенная матрица ранга $m - l < m$, полученная из σ^{-2} заменой элементов каких-нибудь l строк и l столбцов (с теми же номерами) на нули. Для определенности можно считать, что для положительно определенной матрицы σ_2^{-2} порядка $(m-l) \times (m-l)$, обратной к матрице

$$\sigma_2^2 = \mathbf{E}_{\alpha,\sigma^2}(\mathrm{x}_1'' - \alpha'')^\top(\mathrm{x}_1'' - \alpha''),$$

образованной последними $m - l$ столбцами и строками матрицы $\sigma^2 = \|\sigma_{ij}\|$, проверяется гипотеза $H_1 = \{\alpha''\sigma_2^{-2}\alpha''^\top \leqslant a^2\}$ против $H_2 = \{\alpha''\sigma_2^{-2}\alpha''^\top \geqslant b^2\}$, где x_1'', α'' означают те же подвекторы векторов x_1 и α, что и в предыдущем примере. Мешающий параметр α' может быть в каждой из гипотез H_i произвольным.

В этом примере распределение x_1', вообще говоря, зависит от α''. Сделаем следующее преобразование по превращению x_1 в вектор с «ортонормированными» координатами. Положим

$$\mathrm{y} = \mathrm{x}_1\Lambda, \tag{19}$$

где $\Lambda = \|a_{ij}\|$ — треугольная матрица с элементами $a_{ij} = 0$ при $j > i$. Оставшиеся элементы выбираются из условия $\mathrm{y} \in \mathbf{\Phi}_{\beta,E}$, где $\beta = (\beta_1, \ldots, \beta_m) = \alpha\Lambda$. Это всегда можно сделать, так как из (19) получаем

$$\mathrm{y}_m = \mathrm{x}_{1,m} a_{m,m},$$
$$\mathrm{y}_{m-1} = \mathrm{x}_{1,m} a_{m,m-1} + \mathrm{x}_{1,m-1} a_{m-1,m-1},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Отсюда и из условий

$$\mathbf{E}_{\alpha,\sigma^2}(\mathrm{y}_i - \beta_i)^2 = 1,$$
$$\mathbf{E}_{\alpha,\sigma^2}(\mathrm{y}_i - \beta_i)(\mathrm{y}_j - \beta_j) = 0, \quad i \neq j,$$

поочередно определяются значения

$$a_{m,m}^2 = \frac{1}{\sigma_{m,m}},$$
$$\sigma_{m,m} a_{m,m-1} + \sigma_{m-1,m} a_{m-1,m-1} = 0,$$
$$\sigma_{m,m} a_{m,m-1}^2 + 2\sigma_{m,m-1} a_{m,m-1} a_{m-1,m-1} + \sigma_{m-1,m-1} a_{m-1,m-1}^2 = 1,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Таким образом, треугольная матрица Λ такова, что

$$\mathbf{E}_{\alpha,\sigma^2}(\mathrm{y} - \beta)^\top (\mathrm{y} - \beta) = \mathbf{E}_{\alpha,\sigma^2} \Lambda^\top (\mathrm{x}_1 - \alpha)^\top (\mathrm{x}_1 - \alpha)\Lambda = \Lambda^\top \sigma^2 \Lambda = E.$$

Из треугольности Λ следует, что вектор $\beta'' = (\beta_{l+1}, \ldots, \beta_m)$ зависит лишь от α'' и наоборот. Если обозначить через Λ_2 треугольную матрицу порядка $(m-l) \times (m-l)$, полученную из последних $m-l$ строк и столбцов матрицы Λ, то мы получим, очевидно, $\beta'' = \alpha''\Lambda_2$, $\Lambda_2^\top \sigma_2^2 \Lambda_2 = E$. Множество $\Theta_1 = \{\alpha : \alpha'' \sigma_2^{-2} \alpha''^\top \leqslant a^2\}$ перейдет в множество

$$\{\beta : \beta = \alpha\Lambda,\ \alpha'' \sigma_2^{-2} \alpha''^\top \leqslant a^2\} = \{\beta : \beta'' \Lambda_2^{-1} \sigma_2^{-2} \Lambda_2^{-1\top} \beta''^\top \leqslant a^2\} =$$
$$= \{\beta : \beta'' \beta''^\top \leqslant a^2\} = \{\beta : |\beta''| \leqslant a\}.$$

«Подпараметр» β' может быть произвольным, если произвольно α'.

Мы пришли к задаче примера 3. Минимаксный критерий уровня $1-\varepsilon$ для проверки H_1 против H_2 имеет, следовательно, вид $\mathrm{y}''\mathrm{y}''^\top > c_\varepsilon$ или $(\Lambda_2 \Lambda_2^\top = \sigma_2^{-2})$

$$\mathrm{x}_1'' \sigma_2^{-2} \mathrm{x}_1''^\top > c_\varepsilon,$$

где c_ε — решение уравнения (17).

Последний пример является наиболее общим из примеров 1–4. Он резюмирует содержание этих примеров следующим образом.

Теорема 2В. *Если по выборке* $X = \mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,\sigma^2}$ *проверяются гипотезы* (18), *связанные со значением* $\alpha d^{-2} \alpha^\top$, *то минимаксный критерий уровня* $1-\varepsilon$ *имеет вид*

$$\mathrm{x}_1 d^{-2} \mathrm{x}_1^\top > c_\varepsilon, \tag{20}$$

где c_ε *определяется в* (17), $m-l$ *есть ранг* d^{-2}.

Гарантированная мощность критерия (20) *равна*

$$\mathbf{P}\big((\xi_1 - b)^2 + \xi_2 + \ldots + \xi_{m-l} > c_\varepsilon^2\big), \quad \xi_i \in \mathbf{\Phi}_{0,1}.$$

Если выборка X имеет объем n, то $\overline{\mathrm{x}} \in \mathbf{\Phi}_{\alpha,\sigma^2/n}$ будет достаточной статистикой, а минимаксный критерий будет иметь вид

$$\overline{\mathrm{x}} d^{-2} \overline{\mathrm{x}}^\top > \frac{c_\varepsilon}{n}.$$

Следующий пример будет носить несколько иной характер.

Пример 5. Пусть, как и в примере 1, $X = \mathrm{x}_1 \in \mathbf{\Phi}_{\alpha,E}$ есть выборка объема $n = 1$ из m-мерного нормального распределения со средним $\alpha = (\alpha_1, \ldots, \alpha_m)$. Пусть $H_1 = \{\alpha = 0\}$, а гипотеза H_2 состоит в том, что α принадлежит некоторому множеству Θ_2, не содержащему точки $\alpha = 0$. Обозначим через $\overline{\Theta}_2$ выпуклое замыкание множества Θ_2 (наименьшее выпуклое замкнутое множество, содержащее Θ_2), и пусть β есть точка из $\overline{\Theta}_2$, ближайшая к началу координат. Тогда если $\beta \in \Theta_2$, то распределение $\mathbf{Q}_2$, сосредоточенное в точке β, будет наименее благоприятным, и минимаксный критерий $\overline{\pi}$ будет иметь вид $\overline{\pi}(X) = 1$, если

$$(X - \beta)(X - \beta)^\top < XX^\top + c_1$$

или, что то же, если

$$\frac{X\beta^\top}{|\beta|} > c_2,$$

где c_2 выбирается из условия $\overline{\pi} \in K_\varepsilon$.

В самом деле, нам достаточно проверить условие (7). Имеем

$$\mathbf{E}_\alpha \overline{\pi}(X) = \mathbf{P}_\alpha \left(\frac{X\beta^\top}{|\beta|} > c_2 \right),$$

где $X\beta^\top/|\beta| \in \mathbf{\Phi}_{\alpha\beta^\top/|\beta|,1}$, так что

$$\mathbf{E}_\alpha \overline{\pi}(X) = 1 - \Phi\left(c_2 - \frac{\alpha\beta^\top}{|\beta|} \right).$$

Это значит, что минимум $\mathbf{E}_\alpha \overline{\pi}(X)$, $\alpha \in \Theta_2$, достигается при α, минимизирующем функцию $\alpha\beta^\top/|\beta|$. Но очевидно, что $\alpha\beta^\top \geqslant \beta\beta^\top = |\beta|^2$ при всех $\alpha \in \Theta_2$, так что

$$\mathbf{E}_\beta \overline{\pi}(X) = \inf_{\alpha \in \Theta_2} \mathbf{E}_\alpha \overline{\pi}(X). \quad \triangleleft$$

Мы предлагаем читателю построить минимаксный критерий в той же задаче, когда $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$, σ^2 есть произвольная матрица вторых моментов.

3. Вырожденные наименее благоприятные распределения для односторонних гипотез. Пусть $X \in \mathbf{P}_\theta$, где θ и элементы x_i выборки X вещественны.

Предположим, что мы проверяем одностороннюю гипотезу $H_1 = \{\theta \leqslant \theta_1\}$ против $H_2 = \{\theta \geqslant \theta_2\}$ при наличии непустой «зоны безразличия» $\theta_1 < \theta < \theta_2$. При каких условиях наименее благоприятные распределения будут сосредоточены в точках θ_1 и θ_2? Ведь в этом случае минимаксный критерий $\overline{\pi}$ уровня $1-\varepsilon$ выглядел бы весьма просто:

$$\overline{\pi}(X) = \begin{cases} 1, & \text{если} \quad f_{\theta_2}(X) > cf_{\theta_1}(X), \\ p, & \text{если} \quad f_{\theta_2}(X) = cf_{\theta_1}(X), \\ 0, & \text{если} \quad f_{\theta_2}(X) < cf_{\theta_1}(X), \end{cases} \tag{21}$$

где p и c определяются равенством $\mathbf{E}_{\theta_1}\overline{\pi}(X) = \varepsilon$.

Мы уже знаем, что если отношение правдоподобия монотонно, то такой критерий будет р. н. м., а стало быть, и минимаксным. Следующее утверждение дает другое достаточное условие для того, чтобы критерий был минимаксным.

Теорема 3. *Пусть плотность $f_\theta(x)$ обладает тем свойством, что отношение $f_{\theta'}(x)/f_\theta(x)$ не убывает по x при любых $\theta' > \theta$. Тогда наименее благоприятные распределения $\mathbf{Q}_1$ и $\mathbf{Q}_2$ будут сосредоточены соответственно в точках θ_1 и θ_2 и, стало быть, критерий* (21) *будет минимаксным.*

Доказательство. Пусть сначала $n = 1$. По условию теоремы найдутся такие $a \leqslant b$, что $f_{\theta'}(x)/f_\theta(x) \leqslant 1$ при $x \in (-\infty, a]$, $f_{\theta'}(x)/f_\theta(x) = 1$ при $x \in (a, b)$, $f_{\theta'}(x)/f_\theta(x) \geqslant 1$ при $x \in [b, \infty)$. Так как $\overline{\pi}(x)$ не убывает, то $\overline{\pi}(b) \geqslant \overline{\pi}(a)$ и

$$\mathbf{E}_{\theta'}\overline{\pi}(X) - \mathbf{E}_\theta\overline{\pi}(X) \geqslant$$

$$\geqslant \overline{\pi}(a)\int_{-\infty}^{a}(f_{\theta'}(x) - f_\theta(x))\mu(dx) + \overline{\pi}(b)\int_{b}^{\infty}(f_{\theta'}(x) - f_\theta(x))\mu(dx) =$$

$$= (\overline{\pi}(b) - \overline{\pi}(a))\int_{b}^{\infty}(f_{\theta'}(x) - f_\theta(x))\mu(dx) \geqslant 0.$$

Если $n > 1$, то для получения такого же неравенства нужно воспользоваться последовательным интегрированием (сначала по x_1, затем по x_2 и т. д.) и тем, что $\overline{\pi}(X)$ не убывает по каждому из своих аргументов.

Таким образом, мы установили, что мощность $\beta(\theta) = \mathbf{E}_\theta\overline{\pi}(X)$ является неубывающей функцией.

Отсюда следует, что уровень $\overline{\pi}$ равен $1-\varepsilon$ и $\beta(\theta_1) = \sup\limits_{\theta \leqslant \theta_1}\beta(\theta)$, $\beta(\theta_2) = \inf\limits_{\theta \geqslant \theta_2}\beta(\theta)$. Это означает, что выполнены все условия теоремы 1. Теорема 3 доказана. ◁

Если θ — параметр сдвига: $f_\theta(x) = f(x-\theta)$, то можно показать, что $f_{\theta'}(x)/f_\theta(x)$ будет монотонно по x тогда и только тогда, когда функция $-\ln f(x)$ является выпуклой (см. [65]).

§ 50. Критерий отношения правдоподобия

В предыдущих параграфах мы получили ряд результатов, касающихся построения разного рода оптимальных критериев. Важный вывод, который можно сделать из приведенных рассмотрений, состоит в том, что эти оптимальные критерии существуют лишь при весьма ограниченных условиях. В теории оценивания мы имели примерно ту же ситуацию: эффективные оценки существуют также лишь при ограничительных условиях. Однако мы видели в главе 2, что если рассматривать не точное, а асимптотическое свойство эффективности, то оценки, обладающие этим свойством, существуют уже довольно часто, при весьма широких условиях, связанных главным образом с регулярностью семейства $\{\mathbf{P}_\theta\}$. Такими оценками являются о. м. п.

Другое выражение асимптотической оптимальности о. м. п., как мы видели, состоит в том, что о. м. п. асимптотически эквивалентны байесовским оценкам для любого фиксированного гладкого априорного распределения.

В теории проверки гипотез некоторым аналогом о. м. п. является так называемый *критерий отношения правдоподобия* (к. о. п.). При широких предположениях он совпадает с оптимальными критериями, если таковые существуют, и оказывается асимптотически эквивалентным байесовскому критерию в случае $\Theta_1 = \{\theta_1\}$ при любом фиксированном гладком априорном распределении $\mathbf{Q}_2$ на Θ_2. Это свойство и ряд других асимптотических свойств к. о. п. будут установлены в ближайших параграфах.

Дадим определение к. о. п. Пусть в параметрическом случае, когда $X \in \mathbf{P}_\theta$, проверяется гипотеза $H_1 = \{\theta \in \Theta_1\}$ против гипотезы $H_2 = \{\theta \in \Theta_2\}$.

Определение 1. Критерий $\widehat{\pi}(X)$ с критической областью

$$R(X) \equiv \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c \tag{1}$$

называется *критерием отношения правдоподобия* (к. о. п.) для проверки гипотезы H_1 против H_2.

Постоянная c обычно выбирается из условия

$$\sup_{\theta\in\Theta_1} \mathbf{P}_\theta(R(X) > c) = \varepsilon, \tag{2}$$

при котором к. о. п. будет иметь уровень $1 - \varepsilon$.

Наряду с критерием (1) часто рассматривают, по существу, эквивалентный ему критерий (который также называется к. о. п.) вида

$$R_1(X) \equiv \frac{\sup\limits_{\theta\in\Theta} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} = \frac{f_{\widehat{\theta}^*}(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c. \tag{3}$$

Близость этих критериев вытекает из того, что при $\Theta = \Theta_1 \cup \Theta_2$

$$f_{\widehat{\theta}^*}(X) = \max\{\sup_{\theta\in\Theta_1} f_\theta(X),\ \sup_{\theta\in\Theta_2} f_\theta(X)\},$$

и, стало быть, $R^1(X) = \max\{1, R(X)\}$.

Если гипотеза H_1 простая: $\Theta_1 = \{\theta_1\}$, $H_2 = \{\theta \neq \theta_1\}$, так что $\Theta_2 = \Theta \setminus \{\theta_1\}$, то для $f_\theta(\boldsymbol{x})$, непрерывных по θ, будем иметь

$$R(X) = R_1(X) = \frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)}.$$

По своей форме критерий (1) естественным образом обобщает н. м. к. для проверки простых гипотез в лемме Неймана–Пирсона. И хотя *точных* свойств оптимальности этот критерий в общем случае, по-видимому, не имеет, *асимптотически* он часто оказывается наилучшим (см. § 53–56).

Многие несмещенные инвариантные и минимаксные критерии, рассмотренные выше, являются к. о. п. Рассмотрим для иллюстрации примеры 49.1–49.4, в которых строились минимаксные критерии для параметра α нормальных совокупностей. *Во всех этих примерах минимаксные критерии являются к. о. п.* Докажем это. Задачи примеров 49.2, 49.4 с точностью до линейных преобразований параметра сводились к задачам примеров 49.1, 49.3. Так как отношение правдоподобия (1) от таких замен (при соответствующем изменении областей Θ_i) не зависит, нам достаточно рассмотреть лишь примеры 49.1, 49.3.

В примере 49.1 по выборке $X \in \boldsymbol{\Phi}_{\alpha,E}$ единичного объема из многомерной нормальной совокупности с единичной матрицей E вторых моментов проверялась гипотеза $H_1 = \{|\alpha| \leqslant a\}$ против $H_2 = \{|\alpha| \geqslant b\}$, $a < b$. Оказалось, что минимаксный критерий имеет форму

$$|X| > c. \tag{4}$$

В нашем случае $\sup\limits_{\theta\in\Theta_i} f_\theta(X)$ определяется значением

$$\inf_{\alpha\in\Theta_i} (X-\alpha)(X-\alpha)^\top = \inf_{a\in\Theta_i} |X-\alpha|^2,$$

так что для статистики $R(X)$ в (1) будем иметь

$$\ln R(X) = \begin{cases} -\dfrac{1}{2}(|X|-b)^2, & \text{если} \quad |X| \leqslant a, \\ -\dfrac{1}{2}(|X|-b)^2 + \dfrac{1}{2}(|X|-a)^2, & \text{если} \quad a < |X| < b, \\ -\dfrac{1}{2}(|X|-a)^2, & \text{если} \quad |X| \geqslant b. \end{cases}$$

Это есть непрерывная возрастающая функция от $|X|$. Поэтому области (1) и (4) при подходящих значениях c совпадают.

Предлагаем читателю убедиться, что критерий (3) в этом примере также имеет вид (4).

В примере 49.3 по выборке $x \in \mathbf{\Phi}_{\alpha,E}$ единичного объема проверялась гипотеза $H_1 = \{|\alpha''| \leqslant a\}$ против $H_2 = \{|\alpha''| \geqslant b\}$, где $\alpha'' = (\alpha_{l+1}, \ldots, \alpha_m)$ есть подвектор вектора α, составленный из его последних $m - l$ координат. Минимаксный критерий имеет вид

$$|X''| > c, \tag{5}$$

где X'' составлен из $m-l$ последних координат вектора X. Но в этом случае

$$\inf_{\alpha\in\Theta_1} (X-\alpha)(X-\alpha)^\top = \inf_{\alpha'':|\alpha''|\leqslant a} (X''-\alpha'')(X''-\alpha'')^\top.$$

Аналогичное неравенство справедливо для Θ_2. Поэтому все сводится к рассмотрениям, проведенным для примера 49.1, и к. о. п. (1) и (3) будут совпадать с (5).

В условиях § 45 построенные там р. н. м. к. для экспоненциальных семейств

$$f_\theta(\boldsymbol{x}) = c(\theta)e^{\theta T(\boldsymbol{x})}h(\boldsymbol{x}) \tag{6}$$

также будут совпадать с к. о. п. Читатель может проверить это самостоятельно, заметив, что функция

$$\varphi(\theta) = \ln(\theta) = -\ln\left(\int e^{\theta T(\boldsymbol{x})}h(\boldsymbol{x})\mu^n(d\boldsymbol{x})\right)$$

является выпуклой, поскольку $\varphi'(\theta) = -\mathbf{E}_\theta T$, $\varphi''(\theta) = -\mathbf{D}_\theta T < 0$. Из выпуклости φ вытекает однозначная разрешимость уравнения

$$\varphi'(\theta) + T(X) = 0$$

для о. м. п. $\widehat{\theta}^* = \psi(T)$ и монотонность функции ψ. При этом один из $\sup_{\theta\in\Theta_i} f_\theta(X)$ будет достигаться в точке $\widehat{\theta}^*$, а другой — в точках θ_1 или θ_2.

Проверка сделанного утверждения для нормальных семейств $\mathbf{\Phi}_{\alpha,E}$, являющихся частным случаем (6), содержится в § 55.

Несколько иначе обстоит дело в примере 49.5, в котором по выборке $X \in \mathbf{\Phi}_{\alpha,E}$ проверялась гипотеза $H_1 = \{\alpha = 0\}$ против $H_2 = \{\alpha \in \Theta_2\}$. Предполагается, что множество Θ_2 и его выпуклое замыкание $\overline{\Theta}_2$ не содержит точки $\alpha = 0$. Если ближайшая к началу координат точка β множества $\overline{\Theta}_2$ принадлежит Θ_2, то минимаксный критерий существует и имеет вид

$$X\beta^\top > c. \tag{7}$$

Этот критерий не является инвариантным относительно какой-либо группы преобразований. Предлагаем читателю убедиться, что в этом случае к. о. п. будет отличным от (7) и будет иметь вид

$$\rho^2(X,\Theta_2) - \rho^2(X,0) < c,$$

где $\rho(X,\Theta_2) = \inf_{\alpha\in\Theta_2} |X-\alpha|$, $\rho(X,0) = |X|$.

Установим теперь, что при выполнении некоторых предположений критерий отношения правдоподобия обладает свойствами инвариантности. Пусть

G — любая группа преобразований в $\mathcal{X}^n$, относительно которой задача проверки гипотез H_1 и H_2 инвариантна, и пусть $\overline{G}$ — соответствующая группа преобразований $\overline{g}$ на Θ.

Теорема 1. *Если $f_\theta(\boldsymbol{x})$ обладает свойством*

$$f_\theta(g\boldsymbol{x}) = c(g,\boldsymbol{x}) f_{\overline{g}\theta}(\boldsymbol{x}), \tag{8}$$

критерий отношения правдоподобия инвариантен относительно G.

По поводу условия (8) отметим, что оно всегда выполнено, если μ есть мера Лебега и g — преобразование, сохраняющее меру (сдвиг, поворот). В этом случае $c(g,x) \equiv 1$. Для преобразований сжатия $c(g,x) \equiv \text{const}$.

Доказательство теоремы 1. В силу того, что $\overline{g}\Theta_i = \Theta_i$, $i = 1,2$, будем иметь

$$R(g\boldsymbol{x}) = \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(g\boldsymbol{x})}{\sup\limits_{\theta\in\Theta_1} f_\theta(g\boldsymbol{x})} = \frac{\sup\limits_{\theta\in\Theta_2} c(g,\boldsymbol{x}) f_{\overline{g}\theta}(\boldsymbol{x})}{\sup\limits_{\theta\in\Theta_1} c(g,\boldsymbol{x}) f_{\overline{g}\theta}(\boldsymbol{x})} = \frac{\sup\limits_{\theta\in\overline{g}\theta_2} f_\theta(\boldsymbol{x})}{\sup\limits_{\theta\in\overline{g}\Theta_1} f_\theta(\boldsymbol{x})} = R(\boldsymbol{x}). \quad \triangleleft$$

Другие свойства к. о. п. см. в § 51, 53–56.

§ 51.* Последовательный анализ

1. Вводные замечания. Во всех предыдущих рассмотрениях объем n выборки $X = X_n$, которой мы располагаем, был фиксированным. При этом условии мы искали критерии, обладающие теми или иными свойствами оптимальности. Например, в самом простом случае, когда проверяются две простые гипотезы $H_i = \{X \in \mathbf{P}_i\}$, $i = 1,2$, оказалось, что существует н. м. к. π уровня $1-\varepsilon$, который имеет вид (см. теорему 42.1)

$$\pi(X) = \begin{cases} 1, & \text{если} \quad f_2(X) > cf_1(X), \\ p, & \text{если} \quad f_2(X) = cf_1(X), \\ 0, & \text{если} \quad f_2(X) < cf_1(X). \end{cases}$$

Здесь c и p определяются из условия $\mathbf{E}_1\pi(X) = \varepsilon$, $f_i(x)$ суть плотности распределений $\mathbf{P}_i$, $i = 1,2$, относительно некоторой меры μ.

Возможно ли дальнейшее улучшение этой статистической процедуры? В сформулированных условиях, конечно, нет. Но если отказаться от фиксации объема выборки, т. е. сделать число наблюдений n случайной величиной, зависящей от уже проведенных наблюдений, то улучшения возможны. Имеется в виду сокращение числа наблюдений, необходимых для построения критерия с заданными параметрами. Это обстоятельство существенно для тех экспериментов, где проведение опытов сопряжено с затратами.

Принципиальную возможность такого улучшения критериев можно пояснить на следующем примере. Допустим, что распределения $\mathbf{P}_1$ и $\mathbf{P}_2$ не являются взаимно абсолютно непрерывными, и пусть существуют множества B_1 и B_2 из $\mathfrak{B}_\mathcal{X}$ такие, что $f_1(x) > 0$, $f_2(x) = 0$ при $x \in B_1$ и $f_1(x) = 0$, $f_2(x) > 0$ при $x \in B_2$. Тогда ясно, что если $\mathrm{x}_1 \in B_1$ ($\mathrm{x}_1 \in B_2$), то мы безошибочно можем утверждать, что имеет место гипотеза H_1 (H_2). При этом нет никакой необходимости делать дальнейшие наблюдения.

Таким образом, если проводить эксперименты не сразу в количестве n штук, а последовательно, рассматривая результат каждой новой серии полученных наблюдений, то можно объем наблюдений сократить.

Введение последовательной процедуры представляется очень естественным и с точки зрения байесовского подхода. В самом деле, байесовский критерий, рассмотренный нами в § 42, предписывает принимать гипотезу H_2, если апостериорная вероятность $q(2/X)$ этой гипотезы не меньше $1/2$. При этом в критическом множестве среди других будут находиться как выборки X, для которых $q(2/X)$ близка к единице (для таких X прием H_2 представляется целесообразным), так и выборки X, для которых $q(2/X)$ близко к $1/2$. Эти последние было бы естественно рассматривать как «недостаточные» для приема решения и требующие дополнительных экспериментов. Кроме того, как и в приведенном выше примере, апостериорная вероятность $q(2/X)$ может оказаться большой уже после первых опытов, и тогда решение можно было бы принимать без последующих испытаний (в упомянутом примере $q(2/X) = 1$ при $X = \mathrm{x}_1 \in B_2$ для любого априорного распределения $(q(1), q(2)), q(2) > 0$).

Ниже мы рассмотрим последовательную процедуру для проверки двух простых гипотез, на которой будет достигаться максимально возможное сокращение числа наблюдений.

2. Байесовский последовательный критерий. Рассмотрим сначала байесовскую постановку задачи и обозначим через $q(1) = q$, $q(2) = 1 - q$ априорные вероятности гипотез H_1, H_2. Тогда апостериорная вероятность гипотезы H_i после наблюдений $X = X_n$ будет равна

$$q(i/X_n) = \frac{q(i)f_i(X_n)}{q(1)f_1(X_n) + q(2)f_2(X_n)}. \tag{1}$$

Будем проводить наблюдения последовательно и при каждом n будем вычислять значения $q(2/X_n)$, $n = 1, 2, \ldots$ (или $q(1/X_n)$). В плоскости переменных (n, y) рассмотрим *случайную траекторию* апостериорных вероятностей (случайную ломаную), выходящую при $n = 0$ из точки $y = q(2)$ и принимающую в точках $n = 1, 2, \ldots$ значения $y = q(2/X_n)$. С помощью этой траектории можно построить следующий критерий для проверки гипотезы H_1 против H_2: рассмотрим в плоскости (n, y) две прямолинейные границы $y = \gamma_i$, $i = 1, 2$; $0 < \gamma_1 < \gamma_2 < 1$, для переменной $q(2/X_n)$. Принимается гипотеза H_2, если траектория $q(2/X_n)$, $n = 0, 1, \ldots$, впервые выйдет из полосы (γ_1, γ_2) через верхнюю границу γ_2. Если траектория $q(2/X_n)$, $n = 0, 1, \ldots$, выйдет из этой полосы через нижнюю границу γ_1, принимается H_1. Позже мы увидим, что $\mathbf{P}_i$-вероятность $(i = 1, 2)$ того, что $q(2/X_n)$ никогда не выйдет из полосы (γ_1, γ_2), т. е. вероятность события

$$\{\gamma_1 < q(2/X_n) < \gamma_2,\ n = 0, 1, \ldots\} \tag{2}$$

равна нулю.

Число испытаний ν, которое потребуется для того, чтобы принять одну из гипотез (т. е. чтобы нарушить неравенства (2)), очевидно, является *марковской случайной величиной* (моментом остановки) относительно последовательности $\mathrm{x}_1, \mathrm{x}_2, \ldots$ для каждого из распределений $\mathbf{P}_1$, $\mathbf{P}_2$. В этом смысле приведенное правило принятия гипотез является *последовательным*. Оно

довольно хорошо согласуется с правилами, которыми руководствуется человек в своей практической деятельности, — принимать то или иное решение только после того, как наблюдения позволят уменьшить в достаточной степени неопределенность, имеющуюся в отношении изучаемого объекта.

Построенный критерий зависит от $q = q(1)$ и вектора $\gamma = (\gamma_1, \gamma_2)$. Поэтому обозначим его через $\delta_{q,\gamma}$. Установим теперь, что критерий $\delta_{q,\gamma}$ является оптимальным. Для этого введем сначала общее понятие последовательного критерия, существенными характеристиками которого наряду с вероятностями ошибок первого и второго рода становятся средние значения $\mathbf{E}_1\nu$ и $\mathbf{E}_2\nu$ для числа наблюдений ν, необходимых для принятия решения.

Пусть на $(\mathcal{X}^\infty, \mathfrak{B}_{\mathcal{X}}^\infty)$ дана произвольная целочисленная случайная величина $\nu \geqslant 0$, являющаяся марковской относительно последовательности $\mathrm{x}_1, \mathrm{x}_2, \ldots$ $(\{\nu \geqslant n\} \in \sigma(\mathrm{x}_1, \ldots, \mathrm{x}_n) = \mathfrak{B}_{\mathcal{X}}^n)$. Обозначим через $\mathcal{X}^\nu$ пространство векторов (n, X_n) таких, что $\nu(X_\infty) = n$, $X_n = [X_\infty]_n$. На $\mathcal{X}^\nu$ введем σ-алгебру $\mathfrak{B}^\nu$, порожденную событиями $\{\nu = n,\ X_n \in B^n\}$, $B^n \in \mathfrak{B}_{\mathcal{X}}^n$, $n = 0, 1, \ldots$ Ясно, что любое распределение на $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ (или на $(\mathcal{X}^\infty, \mathfrak{B}_{\mathcal{X}}^\infty)$) индуцирует соответствующее распределение на $(\mathcal{X}^\nu, \mathfrak{B}^\nu)$.

Определение 1. *Последовательным критерием* δ для проверки H_1 против H_2 называется пара (ν, Ω), где $\Omega \in \mathfrak{B}^\nu$ представляет собой область приема H_2 (критическую область), а случайная величина ν предполагается собственной относительно обоих распределений $\mathbf{P}_1$, $\mathbf{P}_2$ $(\mathbf{P}_i(\nu < \infty) = 1$, $i = 1, 2)$.

В тех случаях, когда необходимо будет отметить, что ν и Ω относятся к критерию δ, будем писать $\nu(\delta)$ и $\Omega(\delta)$.

Ясно, что эквивалентным образом последовательный критерий можно задать с помощью двузначной измеримой функции на $\mathcal{X}^\nu$. Ясно также, что последовательный критерий δ можно задавать и с помощью построения критической области (обозначим ее снова через Ω) во всем пространстве $\mathcal{X}^\infty$. Однако при таком отображении области Ω и $\mathcal{X}^\nu \setminus \Omega$ приема гипотез H_2 и H_1 не обязательно покрывают все элементы из $\mathcal{X}^\infty$: на тех из них, для которых $\nu(X_\infty) = \infty$, не принимается никакая гипотеза. Но в соответствии с определением, $\mathbf{P}_i$-вероятности множества таких X_∞ равны нулю.

Обычный нерандомизированный критерий δ является частным случаем последовательного, когда $\nu(\delta) \equiv n$ постоянно (если $\nu(\delta) \equiv 0$, то решение принимается без проведения испытаний).

Последовательный критерий δ, как и любой обычный критерий для проверки простых гипотез, характеризуется вероятностями $\alpha_i(\delta)$ ошибок i-го рода $(i = 1, 2)$:

$$\alpha_i(\delta) = \mathbf{P}_i((\nu, X_\nu) \notin \Omega_i),$$

где $\Omega_2 = \Omega$, $\Omega_1 = \mathcal{X}^\nu \setminus \Omega_2$. Кроме того, как уже отмечалось, мы будем характеризовать последовательный критерий средними значениями $\mathbf{E}_i\nu$, $i = 1, 2$. Очевидно, что для обычного критерия δ, построенного по выборке X_n, выполняется $\mathbf{E}_i\nu(\delta) \equiv n$.

Чтобы учесть появление этих новых факторов в постановке задачи (т. е. характеристик, связанных с величиной ν), мы будем считать, что проведение каждого наблюдения требует затрат величиной a. Нам будет удобно также характеризовать потери, возникающие при неправильном решении,

различными значениями w_1 и w_2. Именно, будем считать, что потери i-го рода, возникающие при ошибочном решении, когда верна H_i, равны w_i, $i = 1, 2$.

При этих соглашениях математическое ожидание $R(q, \delta)$ потерь, которые возникают при использовании критерия δ, равно

$$R(q, \delta) = q[\alpha_1(\delta)w_1 + a\mathbf{E}_1\nu(\delta)] + (1 - q)[\alpha_2(\delta)w_2 + a\mathbf{E}_2\nu(\delta)]. \qquad (3)$$

Это выражение называют *байесовским риском* в рассматриваемой задаче. Если положить здесь $a = 0$, $w_1 = w_2 = 1$, то получим выражение для вероятности ошибочного решения критерия δ, которое нами уже неоднократно использовалось в § 41, 42.

Определение 2. Последовательный критерий δ, минимизирующий байесовский риск (3), называется *байесовским*.

Следующее утверждение устанавливает оптимальность (байесовость) критерия $\delta_{q,\gamma}$, построенного в начале этого пункта.

Теорема 1. *Для заданных* a, w_1, w_2 *существуют* γ_1, γ_2 *такие, что критерий* $\delta_{q,\gamma}$ *является байесовским.*

Доказательство. Обозначим через δ_i критерий, который принимает гипотезу H_i без проведения испытаний, так что $\nu(\delta_i) = 0$, $\alpha_i(\delta_i) = 0$. Выясним сначала, в каких случаях критерий δ, минимизирующий $R(q, \delta)$, совпадает с δ_1 или δ_2. Очевидно, что

$$R(q, \delta_1) = (1 - q)w_2, \quad R(q, \delta_2) = qw_1.$$

Пусть K — класс критериев $\{\delta = \delta(X)\}$, которые зависят хотя бы от одного наблюдения, т. е. класс критериев δ, для которых $\nu(\delta) \geqslant 1$. Очевидно, что $R(q, \delta) \geqslant a$ для $\delta \in K$. Положим

$$R(q) = \inf_{\delta \in K} R(q, \delta).$$

Так как критерий δ, основанный на одном испытании ($\nu(\delta) \equiv 1$), принадлежит K, то $R(q) < \infty$.

Для любого $p \in (0, 1)$ имеем в силу линейности $R(q, \delta)$ как функции от q:

$$R(pq_1 + (1 - p)q_2) = \inf_{\delta \in K}[pR(q_1, \delta) + (1 - p)R(q_2, \delta)] \geqslant pR(q_1) + (1 - p)R(q_2).$$

Это означает, что $R(q)$ — вогнутая функция. Так как $a < R(q) < \infty$, то отсюда следует, что $R(q)$ является также непрерывной на $[0, 1]$ функцией. Сравним теперь риски критериев δ_i и $\delta \in K$ в зависимости от q (см. рис. 5).

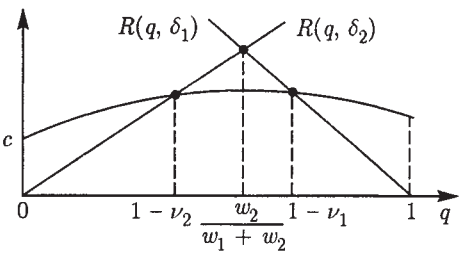

Рис. 5

Одно из двух: либо $R(q) \geqslant \min_i R(q, \delta_i)$ при всех q (это соответствует тому, что $R\left(\frac{w_2}{w_1 + w_2}\right) \geqslant \frac{w_1 w_2}{w_1 + w_2}$), либо существуют решения,

которые обозначим $1-\gamma_1$, $1-\gamma_2$, $1-\gamma_1 > 1-\gamma_2$, соответственно уравнений $R(q,\delta_1) = R(q)$, $R(q,\delta_2) = R(q)$. Очевидно, что $R(q) < \min\limits_i R(q,\delta_i)$ внутри интервала $(1-\gamma_2, 1-\gamma_1)$. Для первой из перечисленных двух возможностей положим

$$1-\gamma_1 = 1-\gamma_2 = \frac{w_1}{w_1+w_2},$$

так что

$$R(1-\gamma_1,\delta_1) = R(1-\gamma_1,\delta_2).$$

Из приведенных рассуждений и рис. 5 вытекает следующее оптимальное правило действий. По данным a, w_1, w_2 вычисляем $1-\gamma_1$, $1-\gamma_2$. Если $q \leqslant 1-\gamma_2$ или, что то же, $1-q \geqslant \gamma_2$, то наименьший риск среди всех критериев дает δ_2 (т. е. надо немедленно принимать H_2). Если $q \geqslant 1-\gamma$ $(1-q \leqslant \gamma_1)$, то наименьший риск дает δ_1 (надо принимать H_1). И лишь в случае $1-\gamma_2 < 1-\gamma_1$, $q \in (1-\gamma_2, 1-\gamma_1)$ (или $1-q \in (\gamma_1,\gamma_2)$) следует воспользоваться критерием из K, т. е. необходимо проводить эксперимент.

Воспользуемся теперь индукцией. Пусть проведено n наблюдений и мы имеем выборку X_n. Перед $(n+1)$-м наблюдением мы имеем ту же альтернативу: либо больше не проводить наблюдений и принять одну из гипотез H_i тут же, либо продолжить наблюдения. Тот факт, что мы уже понесли потери an, роли не играет, так как они ничем уже устранены быть не могут. Существенные изменения касаются лишь априорного распределения. Теперь роль вероятностей $q(1) = q$ и $q(2) = 1-q$ должны играть апостериорные вероятности $q(1/X_n)$, $q(2/X_n)$. Применительно к этой новой ситуации уже выработанное нами оптимальное правило дает, что нужно принимать H_2, если $q(2/X_n) \geqslant \gamma_2$, и H_1, если $q(2/X_n) \leqslant \gamma_1$. Если же $q(2/X_n) \in (\gamma_1,\gamma_2)$, то следует продолжать наблюдения. Но полученное правило есть не что иное, как критерий $\delta_{q,\gamma}$. Таким образом, мы нашли $\gamma_i = \gamma_i(a,w_1,w_2)$, обладающие тем свойством, что критерий $\delta_{q,\gamma}$ минимизирует риск $R(q,\delta)$. $\triangleleft$

Отметим, что числа $\gamma_i(a,w_1,w_2)$ остаются неизменными при умножении a, w_1, w_2 на одно и то же число; это очевидно из их определения, так как такая операция приводит лишь к умножению на то же число всех рисков $R(q,\delta)$. Таким образом, на самом деле γ_i есть функция лишь двух переменных, например a и w_1, если считать, что $w_2 = 1-w_1$.

Что собой представляет байесовский критерий $\delta_{q,\gamma}$? Он предписывает не проводить наблюдений в двух случаях: если $\gamma_1 = \gamma_2$ (это бывает, когда a велико по сравнению с w_1, w_2) либо если $q(2) \leqslant \gamma_1$ или $q(2) \geqslant \gamma_2$. В остальных случаях следует проводить эксперименты до первого нарушения неравенств

$$\gamma_1 < q(2/X_n) < \gamma_2$$

или, что то же (см. (1)), при первом нарушении неравенств

$$\frac{\gamma_1 q(1)}{(1-\gamma_1)q(2)} < \frac{f_2(X_n)}{f_1(X_n)} < \frac{\gamma_2 q(1)}{(1-\gamma_2)q(2)}. \tag{4}$$

При этом принимается гипотеза H_2, если впервые нарушается правое неравенство, и гипотеза H_1 — если левое. В таком виде «переменная» часть

критерия $\delta_{q,\gamma}$ с байесовской постановкой задачи уже не связана, и мы можем, обозначив через Γ_1, Γ_2 левую и правую границы в (4), рассматривать последовательный критерий δ_Γ, $\Gamma = (\Gamma_1, \Gamma_2)$, который называют *последовательным критерием отношения правдоподобия.* Он был впервые введен Вальдом.

3. Последовательный критерий, минимизирующий среднее число испытаний

Теорема 2. *Пусть* $\Gamma_1 < 1 < \Gamma_2$. *Обозначим через* α_1, α_2 *вероятности ошибок первого и второго рода критерия* δ_Γ. *Тогда среди всех последовательных критериев* δ, *для которых* $\alpha_1(\delta) \leqslant \alpha_1$, $\alpha_2(\delta) \leqslant \alpha_2$, *критерий* δ_Γ *имеет наименьшие значения* $\mathbf{E}_1\nu(\delta)$ *и* $\mathbf{E}_2\nu(\delta)$.

Эта теорема означает, в частности, что если δ есть критерий, построенный по выборке X_n фиксированного объема, для которого $\alpha_1(\delta) \leqslant \alpha_1$, $\alpha_2(\delta) \leqslant \alpha_2$, то

$$\mathbf{E}_i\nu(\delta_\Gamma) \leqslant n, \quad i = 1, 2.$$

Доказательство. Байесовский критерий $\delta_{q,\gamma}$, рассматривавшийся в теореме 1, определяется набором чисел (q, a, w_1, w_2). Но, как уже отмечалось, умножение a, w_1, w_2 на одно и то же число не меняет границ γ_i, так что $\delta_{q,\gamma}$ на самом деле определяется тремя параметрами, например (q, a, w), если принять $w_1 = w$, $w_2 = 1 - w$.

Если исходить из этого соглашения, то в теореме 1 мы по заданным (a, w) построили числа $\gamma_i = \gamma_i(a, w)$, при которых критерий $\delta_{q,\gamma}$ является байесовским. Теперь нам потребуется в известном смысле обратное утверждение о том, что для данных γ_1, γ_2 существуют a, w такие, что $\gamma_i(a, w) = \gamma_i$, т. е. такие a, w, для которых критерий $\delta_{q,\gamma}$ будет байесовским в задаче, соответствующей набору (q, a, w). Это утверждение носит технический характер и доказывается довольно сложно (см. [65]). Поэтому мы примем его как допущение*).

Итак, рассмотрим критерий δ_Γ и при данном q найдем γ_i из уравнений

$$\frac{\gamma_i q}{(1-\gamma_i)(1-q)} = \Gamma_i.$$

Для полученных значений $\gamma_i = \Gamma_i(1-q)/(\Gamma_i(1-q)+q)$ найдем a, w, при которых критерий $\delta_{q,\gamma}$ будет байесовским в задаче, соответствующей набору (q, a, w). Так как $\Gamma_1 < 1 < \Gamma_2$, то $\gamma_1 < 1-q < \gamma_2$ и $\nu(\delta_{q,\gamma}) \geqslant 1$. Это означает, что $\delta_{q,\gamma} = \delta_\Gamma$.

Пусть теперь δ — любой другой критерий, для которого $\alpha_i(\delta) \leqslant \alpha_i$. Так как критерий $\delta_{q,\gamma} = \delta_\Gamma$ минимизирует байесовский риск, то

$$q[\alpha_1 w + a\mathbf{E}_1\nu(\delta_\Gamma)] + (1-q)[\alpha_2(1-w) + a\mathbf{E}_2\nu(\delta_\Gamma)] \leqslant$$
$$\leqslant q[\alpha_1(\delta)w + a\mathbf{E}_1\nu(\delta)] + (1-q)[\alpha_2(\delta)(1-w) + a\mathbf{E}_2\nu(\delta)].$$

*) Мы не доказываем здесь и другое полезное утверждение о том, что для непрерывных $\mathbf{P}_i$-распределений величины $f_2(X)/f_1(X)$ и любых заданных α_1, α_2 найдутся Γ_1, Γ_2 такие, что $\alpha_1(\delta_\Gamma) = \alpha_1$, $\alpha_2(\delta_\Gamma) = \alpha_2$. По своему существу это утверждение близко к леммам 46.1 и 47.1, но доказывается труднее.

Отсюда следует, что

$$q\mathbf{E}_1\nu(\delta_\Gamma) + (1-q)\mathbf{E}_2\nu(\delta_\Gamma) \leqslant q\mathbf{E}_1\nu(\delta) + (1-q)\mathbf{E}_2\nu(\delta).$$

Так как число $q \in (0,1)$ здесь произвольно, то $\mathbf{E}_1\nu(\delta_\Gamma) \leqslant \mathbf{E}_1\nu(\delta)$, $\mathbf{E}_2\nu(\delta_\Gamma) \leqslant \mathbf{E}_2\nu(\delta)$. $\triangleleft$

Мы использовали здесь для доказательства тот же прием сравнения с байесовскими критериями, который применяли в § 41, 42, 45.

Рассмотрим некоторые *свойства критерия* δ_Γ. Обозначим через Ω_i^n подмножества из $\mathcal{X}^\infty$, которые определяются следующим образом ($X_k = [X_\infty]_k$):

$$\Omega_1^n = \left\{X_\infty : \Gamma_1 < \frac{f_2(X_k)}{f_1(X_k)} < \Gamma_2,\ k = 1,\ldots,n-1,\ \frac{f_2(X_n)}{f_1(X_n)} \leqslant \Gamma_1\right\}.$$

Множество Ω_2^n определяется так же, но последнее неравенство надо заменить на $f_2(X_n)/f_1(X_n) \geqslant \Gamma_2$. Очевидно, что Ω_i^n не пересекаются, $\Omega_i = \bigcup\limits_{n=1}^{\infty} \Omega_i^n$ есть область приема H_i, $\nu(\delta_\Gamma) = n$ в области $\{\boldsymbol{x} \in \mathcal{X}^\infty : \boldsymbol{x} \in \Omega_i^n\}$,

$$\alpha_1(\delta_\Gamma) = \sum_{n=1}^{\infty} \mathbf{P}_1\{\Omega_2^n\} = \sum_{n=1}^{\infty} \int\limits_{\mathcal{X}^n \cap \Omega_2^n} f_1(\boldsymbol{x})\mu^n(d\boldsymbol{x}) \leqslant$$

$$\leqslant \sum_{n=1}^{\infty} \int\limits_{\mathcal{X}^n \cap \Omega_2^n} f_2(\boldsymbol{x})\Gamma_2^{-1}\mu^n(d\boldsymbol{x}) = \frac{1-\alpha_2(\delta_\Gamma)}{\Gamma_2}. \quad (5)$$

Аналогично устанавливается, что

$$\alpha_2(\delta_\Gamma) \leqslant \Gamma_1(1-\alpha_1(\delta_\Gamma)). \quad (6)$$

Положим для краткости $\alpha_i(\delta_\Gamma) = \alpha_i$. Степень точности полученных неравенств

$$\Gamma_2 \leqslant \frac{1-\alpha_2}{\alpha_1}, \quad \Gamma_1 \geqslant \frac{\alpha_2}{1-\alpha_1} \quad (7)$$

обсудим позже. А сейчас выясним свойства критерия, который получим, если воспользуемся соотношениями (7) в качестве основы для определения Γ_i по заданным α_i. Если положить

$$\Gamma_1' = \frac{\alpha_2}{1-\alpha_1}, \quad \Gamma_2' = \frac{1-\alpha_2}{\alpha_1}, \quad \alpha_i' = \alpha_i(\delta_{\Gamma'}),$$

для полученного критерия $\delta_{\Gamma'}$ будем иметь в силу (7)

$$\frac{\alpha_2}{1-\alpha_1} \geqslant \frac{\alpha_2'}{1-\alpha_1'}, \quad \frac{1-\alpha_2}{\alpha_1} \leqslant \frac{1-\alpha_2'}{\alpha_1'}. \quad (8)$$

Отсюда следует, что

$$\alpha_1' \leqslant \frac{\alpha_1(1-\alpha_2')}{1-\alpha_2} \leqslant \frac{\alpha_1}{1-\alpha_2}, \quad \alpha_2' \leqslant \frac{\alpha_2(1-\alpha_1')}{1-\alpha_1} \leqslant \frac{\alpha_2}{1-\alpha_1}.$$

Приводя неравенства (8) к общему знаменателю и складывая их, получим также

$$\alpha_1' + \alpha_2' \leqslant \alpha_1 + \alpha_2.$$

Таким образом, если α_i малы, то критерий $\delta_{\Gamma'}$ будет иметь значения α_i', которые в сумме не превышают $\alpha_1 + \alpha_2$, и каждое из α_i' может превышать α_i лишь незначительно и в известных нам пределах.

Пример 1. Пусть x_i имеет биномиальное распределение с вероятностью успеха p (распределение Бернулли $\mathbf{B}_p^1$). Задача состоит в проверке гипотезы $H_1 = \{p = p_1\}$ против $H_2 = \{p = p_2\}$, $p_1 < p_2$. В этом случае

$$\frac{f_2(X)}{f_1(X)} = \frac{p_2^{\eta_n}(1-p_2)^{n-\eta_n}}{p_1^{\eta_n}(1-p_1)^{n-\eta_n}} = \left(\frac{p_2(1-p_1)}{p_1(1-p_2)}\right)^{\eta_n}\left(\frac{1-p_2}{1-p_1}\right)^n,$$

где η_n есть число успехов в n испытаниях. Для значений $p_1 = 0{,}05$, $p_2 = 0{,}17$, $\alpha_1 = 0{,}05$, $\alpha_2 = 0{,}10$ получаем*) $\Gamma_1' = 0{,}105$, $\Gamma_2' = 18$, $\alpha_1' = 0{,}031$, $\alpha_2' = 0{,}099$,

$$\mathbf{E}_1\nu(\delta_{\Gamma'}) = 31{,}4, \quad \mathbf{E}_2\nu(\delta_{\Gamma'}) = 30{,}0.$$

С другой стороны, процедура с фиксированным объемом выборки и вероятностями ошибок первого и второго рода 0,05 и 0,10 требует $n = 57$ наблюдений. Таким образом, последовательная процедура в этом примере почти вдвое сокращает среднее число наблюдений.

4. Вычисление параметров наилучшего последовательного критерия. Соотношения (7), (8) дают возможность устанавливать некоторое соответствие между границей Γ и вероятностями ошибок $\alpha_i(\delta_\Gamma)$. Рассмотрим теперь задачу расчета критерия δ_Γ более подробно.

а) Точные формулы. Положим

$$z_k = \ln\frac{f_2(\mathrm{x}_k)}{f_1(\mathrm{x}_k)}, \quad k = 1, 2, \ldots, \quad A_i = \ln\Gamma_i, \quad i = 1, 2.$$

Тогда критерию δ_Γ можно придать следующую форму: если $A_1 < 0 < A_2$, то последовательно проводятся эксперименты и суммируются независимые одинаково распределенные значения z_k до тех пор, пока сумма $Z_n = \sum_{k=1}^{n} z_k$ не коснется впервые одной из границ A_i. Если верна гипотеза H_2, то описанное блуждание будет направлено в среднем вверх, так как

$$\mathbf{E}_2 z_1 = \int \ln\frac{f_2(x)}{f_1(x)} f_2(x)\mu(dx) = \rho_1(\mathbf{P}_2, \mathbf{P}_1) > 0$$

(см. лемму 16.1). Аналогично устанавливается, что $\mathbf{E}_1 z_1 = -\rho_1(\mathbf{P}_1, \mathbf{P}_2) < 0$.

Если границы A_i удаляются от начала координат, то это соответствует (ср. с (5), (6)) уменьшению вероятностей ошибок первого и второго рода.

*) Числовые данные заимствованы из [65, с. 143].

Множества Ω_2^n в терминах блуждания $\{Z_k\}$ будут иметь вид

$$\Omega_2^n = \{A_1 < Z_k < A_2,\ k = 1, \ldots, n-1,\ Z_n \geqslant A_2\}.$$

Аналогичную форму будут иметь множества Ω_1^n.

Обозначим через $\eta(t)$ случайную величину, равную времени первого выхода случайного блуждания $Z_0 = 0, Z_1, Z_2, \ldots$ за границу t:

$$\eta(t) = \begin{cases} \min\{k : Z_k \geqslant t\} & \text{при} \quad t > 0, \\ \min\{k : Z_k \leqslant t\} & \text{при} \quad t < 0. \end{cases}$$

Это есть процесс восстановления, соответствующий последовательности $\{Z_k\}$ (см. [17, гл. 9]). Разности $\chi(A_i) = Z_{\eta(A_i)} - A_i$ будут представлять собой величины эксцессов (перескоков) через уровни A_i в блуждании $\{Z_k\}$ (см. [17]).

Для вероятности ошибки первого рода можем записать

$$\alpha_1(\delta_\Gamma) = \sum_{n=1}^{\infty} \mathbf{P}_1(\Omega_2^n) = \sum_{n=1}^{\infty} \int\limits_{\mathcal{X}^n \cap \Omega_2^n} \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} f_2(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) =$$

$$= \sum_{n=1}^{\infty} \mathbf{E}_2(e^{-Z_n}; \Omega_2^n) = \Gamma_2^{-1} \mathbf{E}_2(e^{-\chi(A_2)}; \Omega_2), \quad (9)$$

где $\Omega_2 = \bigcup\limits_{n=1}^{\infty} \Omega_2^n$ есть область приема H_2. Аналогично

$$\alpha_2(\delta_\Gamma) = \Gamma_1 \mathbf{E}_1(e^{\chi(A_1)}; \Omega_1), \quad \Omega_1 = \bigcup_{n=1}^{\infty} \Omega_1^n. \quad (10)$$

Далее, для значений $\mathbf{E}_i \nu$, $i = 1, 2$; $\nu = \nu(\delta_\Gamma)$, в силу тождества Вальда находим

$$\mathbf{E}_i(Z_\nu) = \mathbf{E}_i z_1 \mathbf{E}_i \nu, \quad i = 1, 2.$$

Так как $Z_\nu = A_2 + \chi(A_2)$ на множестве Ω_2 и $Z_\nu = A_1 + \chi(A_1)$ на множестве Ω_1, то

$$\begin{aligned} \mathbf{E}_1 \nu &= \frac{1}{\mathbf{E}_1 z_1} [\alpha_1 A_2 + \mathbf{E}_1(\chi(A_2); \Omega_2) + (1 - \alpha_1) A_1 + \mathbf{E}_1(\chi(A_1); \Omega_1)], \\ \mathbf{E}_2 \nu &= \frac{1}{\mathbf{E}_2 z_1} [(1 - \alpha_2) A_2 + \mathbf{E}_2(\chi(A_2); \Omega_2) + \alpha_2 A_1 + \mathbf{E}_2(\chi(A_1); \Omega_1)]. \end{aligned} \quad (11)$$

В ряде случаев правые части в формулах (9)–(11) могут быть найдены в явном виде. Эти формулы оказываются весьма полезными и в приближенных вычислениях.

б) Приближенные формулы (при больших A_1, A_2) и неравенства. Мы уже отмечали, что большие значения $|A_i|$, $i = 1, 2$, соответствуют малым вероятностям ошибок $\alpha_i(\delta_\Gamma)$. Рассмотрим значение

$$\alpha_1(\delta_\Gamma) = \mathbf{P}_1(\sup_{k \leqslant \eta(A_1)} Z_k \geqslant A_2) =$$

$$= \mathbf{P}_1(\sup_{k \geqslant 0} Z_k \geqslant A_2) - \mathbf{P}_1(\sup_{k \leqslant \eta(A_1)} Z_k < A_2, \sup_{k > \eta(A_1)} Z_k \geqslant A_2). \quad (12)$$

Последнее слагаемое здесь не превосходит в силу марковости случайной величины $\eta(t)$ значений

$$\mathbf{P}_1(\sup_{k>\eta(A_1)}(Z_k - Z_{\eta(A_1)}) \geqslant A_2 - Z_{\eta(A_1)}) \leqslant \mathbf{P}_1(\sup_{k\geqslant 0} Z_k \geqslant A_2 - A_1).$$

Так как почти во всех практически интересных случаях вероятность $u(A) = \mathbf{P}_1(\sup_{k\geqslant 0} Z_k \geqslant A)$ убывает экспоненциально с ростом A (см., например, [111, т. 2]; это же можно извлечь из главы 11 в [17], где изложены методы вычислений $u(A)$ *)), то при больших $|A_i|$ значение $u(A_2 - A_1)$ будет иметь более высокий порядок малости, чем $u(A_2)$. Это означает в силу (12), что

$$\alpha_1(\delta_\Gamma) \approx \mathbf{P}_1(\sup_{k\geqslant 0} Z_k \geqslant A_2) = u(A_2), \tag{13}$$

так что второй границей при больших A_1, A_2 в (12) можно пренебречь. Точно так же получается приближение

$$\alpha_2(\delta_\Gamma) \approx \mathbf{P}_2(\inf_{k\geqslant 0} Z_k \leqslant A_1). \tag{14}$$

Если $|A_i|$ велики, α_i малы, то главные части в (11) дают

$$\mathbf{E}_1\nu \approx \frac{A_1}{\mathbf{E}_1 z_1}, \quad \mathbf{E}_2\nu \approx \frac{A_2}{\mathbf{E}_2 z_1}. \tag{15}$$

В основе этих формул также лежит пренебрежение второй границей (их можно получить и с помощью приближений $\mathbf{E}_i\nu \approx \mathbf{E}_i\eta(A_i) \approx A_i/\mathbf{E}_i z_1$. Последнее соотношение имеет место в силу теоремы восстановления [17]).

Учет следующих по порядку малости членов в (11) дает

$$\begin{aligned} \mathbf{E}_1\nu &= \frac{1}{\mathbf{E}_1 z_1}(A_1 + \alpha_1(A_2 - A_1) + \mathbf{E}_1\chi_1), \\ \mathbf{E}_2\nu &= \frac{1}{\mathbf{E}_2 z_1}(A_2 - \alpha_2(A_2 - A_1) + \mathbf{E}_2\chi_2), \end{aligned} \tag{16}$$

где α_i определяются приближениями (12), (13), значения $\mathbf{E}_i\chi_i = \lim_{|A_i|\to\infty} \mathbf{E}_i\chi(A_i)$ могут быть найдены методами главы 11 в [17].

Рассмотрим теперь неравенства (8). Так как $\chi(A_1) \leqslant 0$, $\chi(A_2) \geqslant 0$, то эти неравенства следуют из (9), (10), если $\chi(A_i)$ заменить на нуль. Стало быть, точность этих неравенств определяется погрешностью, вызываемой такой заменой.

Если случайные величины z_i ограничены, $b_1 \leqslant z_i \leqslant b_2$, то, очевидно, $\chi(A_2) \leqslant b_2$, $\chi(A_1) \geqslant b_1$, и наряду с (5), (6) могут быть написаны обратные неравенства. Именно,

$$\begin{aligned} \alpha_1(\delta_\Gamma) = \Gamma_2^{-1}\mathbf{E}_2(e^{-\chi(A_2)}; \Omega_2) &\geqslant \Gamma_2^{-1}e^{-b_2}(1 - \alpha_2), \\ \alpha_2(\delta_\Gamma) &\geqslant \Gamma_1 e^{b_1}(1 - \alpha_1). \end{aligned} \tag{17}$$

*) Подробнее об этом см. [12].

Чтобы проиллюстрировать полученные соотношения, обратимся к примеру 1. Для него

$$Z_n = \eta_n \ln \frac{p_2(1-p_1)}{p_1(1-p_2)} + n \ln \frac{1-p_2}{1-p_1},$$

где η_n есть число успехов в n испытаниях. Это значит, что z_1 для $\mathbf{P}_i$-распределения принимает значение $b_2 = \ln(p_2/p_1) \approx 1{,}224$ с вероятностью p_i и значение $b_1 = \ln \dfrac{1-p_2}{1-p_1} \approx -0{,}135$ с вероятностью $1-p_i$, $i = 1, 2$. Отсюда находим

$$\mathbf{E}_1 z_1 = -0{,}067, \quad \mathbf{E}_2 z_1 = 0{,}096, \quad e^{b_2} = 3{,}400, \quad e^{b_1} = 0{,}874.$$

Из двух последних значений близко к единице лишь второе, так что сравнительно точным будет лишь второе из неравенств (17). Используя это неравенство и (7) для критерия $\delta_{\Gamma'}$, получим

$$0{,}102 = \frac{\alpha_2'}{1-\alpha_1'} \leqslant \Gamma_1' \leqslant \frac{\alpha_2'}{(1-\alpha_1')e^{b_1}} = 0{,}117.$$

Это дает довольно точные границы для значения $\Gamma_1' = 0{,}105$. В нашем случае

$$A_1' = \ln \Gamma_1' \approx -2{,}254, \quad A_2' = \ln \Gamma_2' \approx 2{,}890.$$

Отсюда, используя приближенные формулы (15), получим для $\mathbf{E}_i \nu'$, $i = 1, 2$, значения

$$\frac{A_1'}{\mathbf{E}_1 z_1} = 33{,}639, \quad \frac{A_2'}{\mathbf{E}_2 z_1} = 30{,}108.$$

Мы видим, что даже довольно грубые приближения такие, как (15), дают правильное представление о величинах $\mathbf{E}_i \nu'$. Результаты будут значительно более точными, если воспользоваться формулами (16).

§ 52. Проверка сложных гипотез в общем случае

В этом параграфе мы не будем предполагать, что выборка относится к какому-либо параметрическому семейству.

Задача проверки двух гипотез в общем случае выглядит следующим образом. Пусть $\mathcal{P}_1$ и $\mathcal{P}_2$ — два семейства распределений такие, что распределение $\mathbf{P}$ выборки X принадлежит $\mathcal{P}_1 \cup \mathcal{P}_2$. Проверяется гипотеза $H_1 = \{X \in \in \mathbf{P}, \mathbf{P} \in \mathcal{P}_1\}$ против $H_2 = \{X \in \mathbf{P}, \mathbf{P} \in \mathcal{P}_2\}$. Общий принцип построения (нерандомизированного*)) критерия $\pi(X) = \delta(X)$ здесь остается прежним — таким же, каким он был описан в § 44 для параметрического случая. Именно, строится критическое множество $\Omega \subset \mathcal{X}^n$ (часто отождествляемое с понятием критерия) такое, что мы принимаем H_2, если $X \in \Omega$, и принимаем H_1 в противном случае. Число

$$1 - \varepsilon = \inf_{\mathbf{P} \in \mathcal{P}_1} \mathbf{P}(X \notin \Omega)$$

*) В дальнейшем для единообразия обозначений мы будем статистические критерии обозначать символом π, хотя в пределах этой главы это будут, как правило, нерандомизированные критерии.

называется уровнем значимости критерия. Величина

$$\beta_\pi(\mathbf{P}) = \mathbf{P}(X \in \Omega), \quad \mathbf{P} \in \mathcal{P}_2,$$

есть значение мощности критерия π в «точке» $\mathbf{P} \in \mathcal{P}_2$.

Сравнивать мощности $\beta_\pi(\mathbf{P})$ критериев π в случае, когда множество $\mathcal{P}_2$ альтернатив $\mathbf{P}$ очень богато, и строить оптимальные критерии в этих условиях весьма затруднительно или просто невозможно. Минимальные требования к критериям в этом случае состоят обычно в том, чтобы для каждого фиксированного $\mathbf{P} \in \mathcal{P}_2$ выполнялось

$$\lim_{n\to\infty} \beta_\pi(\mathbf{P}) = 1.$$

Определение 1. Критерий π, обладающий этим свойством, называется *состоятельным.*

Существо рассматриваемых критериев, как и всех статистических критериев, соответствует основному принципу математической статистики, о котором уже говорилось в § 4 и в § 39. Если ε мало, то при выполнении гипотезы H_1 и при многократном использовании построенного критерия уровня $1-\varepsilon$ мы будем ошибаться (т. е. попадать в критическую область) в среднем лишь в $100\varepsilon\,\%$ всех испытаний. Поэтому попадание при выполнении гипотезы H_1 в эту область в единичном испытании мы считаем практически невозможным. Так что если мы все-таки туда попали, значит, сделанное предположение неверно, и мы объявляем, что гипотеза H_1 неверна. В этом случае говорят, что *результаты эксперимента не согласуются с гипотезой H_1 с точки зрения критерия уровня $1-\varepsilon$.*

Очень распространенными являются критерии проверки простой гипотезы $H_1 = \{X \Subset \mathbf{P}_1\}$ против сложной альтернативы $H_2 = \{X \Subset \mathbf{P} \neq \mathbf{P}_1\}$; гипотеза H_2 означает, что X есть выборка из произвольного распределения $\mathbf{P} \neq \mathbf{P}_1$.

В основе построения критериев для проверки простой гипотезы $H_1 = \{X \Subset \mathbf{P}_1\}$ обычно лежит «удаленность» эмпирического распределения $\mathbf{P}_n^*$ от распределения $\mathbf{P}_1$ в смысле некоторого «расстояния» $d(\mathbf{P}, \mathbf{Q})$. Желательным свойством этого расстояния является обращение $d(\mathbf{P}, \mathbf{Q})$ в нуль *лишь* при $\mathbf{Q} = \mathbf{P}$, а также непрерывность $d(\mathbf{P}, \mathbf{Q})$ в «окрестности» точки $\mathbf{Q} = \mathbf{P}$, например в равномерной метрике (иначе малые отклонения $\mathbf{Q}$ от $\mathbf{P}$ могут приводить к большим значениям расстояния d). Напомним, что в параметрическом случае аналогичные соображения мы использовали при построении оценок неизвестного параметра по методу минимального расстояния.

Итак, пусть $d(\mathbf{P}, \mathbf{Q})$ есть некоторое расстояние (не обязательно метрика) в пространстве распределений. Предположим, что по заданному $\varepsilon > 0$ можно найти такое $c > 0$, для которого

$$\mathbf{P}_1\big(d(\mathbf{P}_1, \mathbf{P}_n^*) > c\big) = \varepsilon. \tag{1}$$

Тогда критерий строится следующим образом:

$$\pi(X) = \begin{cases} 0, & \text{если} \quad d\big(\mathbf{P}_1, \mathbf{P}_n^*\big) \leqslant c, \\ 1 & \text{в противном случае.} \end{cases}$$

Очевидно, что π есть критерий уровня $1-\varepsilon$.

Так же, как и в § 43, можно ввести понятие критерия *асимптотического уровня* $1-\varepsilon$, для которого

$$\lim_{n\to\infty} \mathbf{P}_1\big(d(\mathbf{P}_1, \mathbf{P}_n^*) > c\big) = \varepsilon. \tag{2}$$

Описанные критерии часто называют *критериями согласия* (с предположением $\{X \in \mathbf{P}_1\}$). Их конструкцию можно эквивалентным образом представить и в несколько иной форме. Пусть дан функционал $G(\mathbf{P})$ (или последовательность функционалов $G_n(\mathbf{P})$) такой, что $G(\mathbf{P}) \neq G(\mathbf{P}_1)$ при $\mathbf{P} \neq \mathbf{P}_1$. Тогда можем положить $\pi(X) = 1$, если $\big|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)\big| > c$, и $\pi(X) = 0$ в противном случае, где c выбирается из тех же соображений, что и в (1), (2). Нетрудно проверить, что этот второй подход эквивалентен первому, поскольку по функционалу G можно построить расстояние $d(\mathbf{P}, \mathbf{P}_1) = |G(\mathbf{P}) - G(\mathbf{P}_1)|$ (ср. с принципом подстановки в теории оценивания), и наоборот по расстоянию $d(\mathbf{P}, \mathbf{P}_1)$ можно построить функционал $G(\mathbf{P}) = d(\mathbf{P}, \mathbf{P}_1)$ $(G(\mathbf{P}_1) = 0)$, удовлетворяющий требуемым свойствам.

Если функционал G в описанной конструкции критерия обладает к тому же свойством $G(\mathbf{P}_n^*) \underset{P}{\longrightarrow} G(\mathbf{P})$ при $X \in \mathbf{P}$ (это всегда так, если G является функционалом первого или второго типа; см. § 3), то *построенный критерий будет состоятельным.* Действительно, в этом случае число $c = c(n)$, обеспечивающее равенство (2), должно сходиться к нулю (так как $\mathbf{P}_1\big(\big|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)\big| > \varepsilon\big) \to 0$ при любом $\varepsilon > 0$) и, следовательно, будем иметь $G(\mathbf{P}_n^*) \underset{P}{\longrightarrow} G(\mathbf{P})$, $\mathbf{P}\big(\big|G(\mathbf{P}_n^*) - G(\mathbf{P}_1)\big| > c(n)\big) \to 1$ при каждом фиксированном $\mathbf{P} \neq \mathbf{P}_1$.

Рассмотрим теперь некоторые хорошо известные критерии согласия, которые являются реализацией описанного выше подхода.

1. Критерий Колмогорова. Рассмотрим статистику (расстояние)

$$D(\mathbf{P}_1, \mathbf{P}_n^*) = \sup_t \big|F_n^*(t) - F(t)\big|,$$

где $F_n^*(t)$ и $F(t)$ — функции распределения, соответствующие мерам $\mathbf{P}_n^*$ и $\mathbf{P}_1$. В § 8 нами было установлено, что если $F(t)$ непрерывна, $X \in \mathbf{P}_1$, то

$$d_K(\mathbf{P}_1, \mathbf{P}_n^*) \equiv \sqrt{n}\, D(\mathbf{P}_1, \mathbf{P}_n^*) \Rightarrow \sup_{0\leqslant t\leqslant 1} |w^0(t)|,$$

где $w^0(t)$ — броуновский мост. Отсюда следует

Теорема 1 (А. Н. Колмогоров). *Если $F(t)$ непрерывна, то существует*

$$\lim_{n\to\infty} \mathbf{P}_1\big(d_K(\mathbf{P}_1, \mathbf{P}_n^*) < x\big) = K(x) \equiv \mathbf{P}\Big(\sup_{0\leqslant t\leqslant 1} |w^0(t)| < x\Big).$$

Функцию $K(x)$ можно найти в явном виде. Она равна

$$K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}.$$

С помощью этой теоремы можно конструировать критерии асимптотического уровня $1-\varepsilon$. Функция $K(x)$ табулирована во многих руководствах по математической статистике. Поэтому для заданного ε можем с помощью таблиц найти постоянную $c = c_\varepsilon$, для которой $K(c) = 1-\varepsilon$. Положив $\pi(X) = 1$ при $d_K(\mathbf{P}_1, \mathbf{P}_n^*) > c_\varepsilon$, мы получим критерий согласия асимптотического уровня $1-\varepsilon$. Нетрудно видеть, что полученный критерий состоятелен, так как функционал $G(\mathbf{P}) = \sup\limits_t |F_P(t) - F(t)|$ (здесь $F_P(t) = \mathbf{P}((-\infty, t))$), с помощью которого построен критерий Колмогорова, непрерывен относительно F_P в равномерной метрике и, стало быть, является функционалом II типа (см. главу 1), для которого $G(\mathbf{P}_n^*) \underset{\text{п.н.}}{\longrightarrow} G(\mathbf{P})$ при $X \in \mathbf{P}$. Остается воспользоваться сделанными выше замечаниями об условиях состоятельности критериев согласия.

С помощью результатов главы 1 мы можем найти асимптотическое поведение мощности критерия Колмогорова относительно близких альтернатив (см. § 43). Допустим, что $X \in \mathbf{P}$, где распределение $\mathbf{P}$ имеет функцию распределения

$$F_P(x) = F(x) + p(x)n^{-1/2}. \tag{3}$$

Будем предполагать для простоты, что $p(x)$ непрерывна, а $F(x)$ непрерывна и строго монотонна. Мощность $\beta(\mathbf{P})$ критерия Колмогорова в «точке» $\mathbf{P}$ будет равна

$$\beta(\mathbf{P}) = \mathbf{P}\big(d_K(\mathbf{P}_1, \mathbf{P}_n^*) > c\big) = \mathbf{P}\big(\sup_t |F(t) - F_n^*(t)|\sqrt{n} > c\big) =$$
$$= \mathbf{P}\big(\sup_t |F_P(t) - p(t)n^{-1/2} - F_n^*(t)|\sqrt{n} > c\big).$$

Если сделать замену $t = F_P^{-1}(u)$, где F_P^{-1} есть функция, обратная к F_P, то мы получим выражение

$$\mathbf{P}\left(\sup_{0\leqslant u\leqslant 1} \left|u - p\big(F_P^{-1}(u)\big)n^{-1/2} - F_n^*\big(F_P^{-1}(u)\big)\right|\sqrt{n} > c\right). \tag{4}$$

Здесь $U_n^*(u) = F_n^*\big(F_P^{-1}(u)\big)$ есть эмпирическая функция, соответствующая равномерному на $[0,1]$ распределению $U_{0,1}$, так что (4) равно

$$\mathbf{P}\left(\sup_{0\leqslant u\leqslant 1} \left|u - U_n^*(u) - p\big(F_P^{-1}(u)\big)n^{-1/2}\right|\sqrt{n} > c\right).$$

Кроме того, $F_P^{-1}(u) \to F^{-1}(u)$ в силу строгой монотонности F. Отсюда и из непрерывности p следует, что

$$\lim_{n\to\infty} \beta(\mathbf{P}) = \mathbf{P}\left(\sup_{0\leqslant t\leqslant 1} |w^0(t) - a(t)| > c\right), \tag{5}$$

где $a(t) = p(F^{-1}(t))$.

Можно показать, что это выражение минимально при $a(t) \equiv 0$ ($p \equiv 0$). В этом смысле критерий Колмогорова является асимптотически несмещенным.

2. Критерий Мизеса–Смирнова (критерий ω^2). Рассмотрим в качестве расстояния между $\mathbf{P}_1$ и $\mathbf{P}_n^*$ статистику

$$\omega_n^2 = d_{\omega^2}(\mathbf{P}_1, \mathbf{P}_n^*) = n \int \left(F(x) - F_n^*(x)\right)^2 dF(x),$$

с помощью которой также можно построить критерий согласия заданного уровня. В главе 1 доказано, что здесь, как и в предыдущем случае, справедлива

Теорема 2. *Существует предельное распределение*

$$\lim_{n\to\infty} \mathbf{P}_1(\omega_n^2 < x) = \Omega(x) \equiv \mathbf{P}\left(\int_0^1 (w^0(t))^2\, dt < x\right).$$

Функция $\Omega(x)$ имеет весьма сложный вид (см. [10]), и мы ее приводить здесь не будем.

Так как функционал

$$G(\mathbf{P}) = \int (F(t) - F_P(t))^2\, dF(t)$$

является функционалом II типа (§ 3), то по тем же соображениям, что и в случае критерия Колмогорова, критерий ω^2 является состоятельным.

Следуя рассуждениям предыдущего пункта, можно найти и асимптотическое поведение мощности $\beta(\mathbf{P})$ критерия ω^2 для близких альтернатив $\mathbf{P}$ вида (3). Получим совершенно аналогичным образом, что

$$\beta(\mathbf{P}) = \mathbf{P}(\omega_n^2 > c) \to \mathbf{P}\left(\int (w^0(t) - a(t))^2\, dt > c\right),$$

где $a(t)$ определено в (5). Полученное предельное значение так же, как и (5), минимально при $a(t) \equiv 0$, так что критерий ω^2 тоже является асимптотически несмещенным.

Рассмотренные два критерия, как и другие критерии согласия с гипотезой $H_1 = \{X \in \mathbf{P}_1\}$, построенные с помощью расстояний $d(\mathbf{P}, \mathbf{Q})$, позволяют сразу получать *доверительные множества* для неизвестной функции распределения $F(x)$ или для неизвестного распределения $\mathbf{P}_1$ выборки X. Действительно, соотношение (1) (или (2)) можно трактовать и так: вероятность того, что c-окрестность «точки» $\mathbf{P}_n^*$ (в смысле расстояния d) накроет «точку» $\mathbf{P}_1$, равна $1 - \varepsilon$. (Для (2) мы получим асимптотический вариант этого утверждения.) Это означает (см. § 48), что c-окрестность точки $\mathbf{P}_n^*$ представляет собой доверительное множество уровня $1 - \varepsilon$ для неизвестного распределения $\mathbf{P}_1$, $X \in \mathbf{P}_1$. Критерий Колмогорова, например, определяет эту окрестность в терминах функций распределения: это множество всех $F(x)$, для которых

$$\sup_t \left|F(t) - F_n^*(t)\right| \leqslant \frac{c_\varepsilon}{\sqrt{n}},$$

где c_ε определяется из (1).

Вернемся к критериям. Мы уже отмечали, что асимптотическим уровням значимости можно доверять лишь при больших n. Если же объем выборки невелик, то при построении критерия (точнее, при нахождении $c = c_\varepsilon$) необходимо использовать точные формулы для распределения $d(\mathbf{P}_1, \mathbf{P}_n^*)$. Однако их получение связано, как правило, с большими трудностями. В этой связи важную роль играют так называемые *непараметрические* критерии, основанные на статистиках, распределение которых не зависит от истинного распределения $\mathbf{P}_1$ (или не зависит от параметра θ, если $X \in \mathbf{P}_\theta$).

В этом случае вероятности $\mathbf{P}_1\big(d(\mathbf{P}_1, \mathbf{P}_n^*) < x\big)$ от $\mathbf{P}_1$ не зависят, и, стало быть, можно один раз произвести вычисления, составить таблицы и затем использовать их при любых $\mathbf{P}_1$.

Критерий Колмогорова и критерий ω^2 являются непараметрическими критериями. Этот факт установлен в § 6.

Непараметрические критерии возникают и при проверке *двух сложных* гипотез.

3. Критерий знаков. Пусть $F(x)$ есть функция распределения для $\mathbf{P}_1$ и гипотеза H_1 состоит в том, что $F(a) = p$ для заданной точки a. Очевидно, это есть сложная гипотеза. Гипотеза H_2 является дополнительной: $H_2 = \{X \in \mathbf{P}, F_P(a) \neq p\}$. В этом случае естественно воспользоваться следующей статистикой: обозначим через $\nu(X)$ число наблюдений x_i, для которых знак разности $\mathrm{x}_i - a$ отрицателен. В качестве критического множества Ω рассмотрим все выборки X, для которых

$$\nu(X) \notin (c_1, c_2)$$

при некоторых $c_1 < c_2$.

Если верна гипотеза H_1, то

$$\mathbf{P}_1(\nu(X) = k) = C_n^k p^k (1-p)^{n-k}.$$

Итак, распределение $\nu(X)$ при гипотезе H_1 от $\mathbf{P}_1$ не зависит, так что наш критерий является непараметрическим. Числа c_i следует выбрать так, чтобы

$$\mathbf{P}(\nu(X) \in (c_1, c_2)) \geqslant 1 - \varepsilon$$

(знак равенства из-за дискретности $\nu(X)$ здесь может не достигаться). Неоднозначность в выборе c_i можно устранить требованием несмещенности относительно изменений p. В целом эта задача эквивалентна проверке гипотезы о том, что вероятность успеха в схеме Бернулли равна p. Аналогичным образом можно строить «односторонние» критерии для проверки гипотез $F(a) \leqslant p$.

Если в качестве обобщения рассмотренной задачи мы будем проверять гипотезу $F(a_i) = p_i$, $i = 1, \ldots, r$, для заданных значений a_i, p_i, то придем к критерию χ^2, который подробно рассматривается в § 56.

4. Критерий Морана. Так называется следующий критерий для проверки гипотезы о том, что $X \in \mathbf{P}_1$. Пусть $\mathrm{x}_{(1)}, \ldots, \mathrm{x}_{(n)}$ — вариационный ряд, построенный по выборке X. Предположим, что $\mathbf{P}_1$ имеет непрерывную

функцию распределения F, и составим статистику

$$M_n = \sum_{k=0}^{n} [F(\mathrm{x}_{(k+1)}) - F(\mathrm{x}_{(k)})]^2, \tag{6}$$

где принято $F(\mathrm{x}_{(0)}) = 0$, $F(\mathrm{x}_{(n+1)}) = 1$. Критерий Морана отвергает гипотезу $\{X \in \mathbf{P}_1\}$, если $M_n > c$.

Очевидно, этот критерий является непараметрическим, так как $F(\mathrm{x}_k) \in \in \mathbf{U}_{0,1}$. Стало быть, достаточно рассматривать критерий $M_n > c$, основанный на статистике

$$M_n = \sum_{k=0}^{n} (\mathrm{x}_{(k+1)} - \mathrm{x}_{(k)})^2$$

и предназначенный для проверки равномерности распределения X. Использование статистики M_n в этом случае представляется естественным, так как величина $\sum_{i=0}^{n} y_i^2$ достигает своего минимума при условии $\sum_{i=0}^{n} y_i = 1$ в точке $y_0 = \ldots = y_n = \dfrac{1}{n+1}$.

Для вычисления асимптотического уровня критерия Морана может служить следующая

Теорема 3. *Если* $X \in \mathbf{P}_1$, *то*

$$\sqrt{n}\left(\frac{nM_n}{2} - 1\right) \Rightarrow \mathbf{\Phi}_{0,1}.$$

Доказательство. Пусть ξ_j независимы и $\xi_j \in \mathbf{\Gamma}_{\alpha,1}$, $j = 1, 2, \ldots$ Тогда $\zeta_k = \sum_{j=1}^{k} \xi_j \in \mathbf{\Gamma}_{\alpha,k}$, и в силу следствия 6.2 совместное распределение разностей

$$\mathrm{x}_{(1)}, \mathrm{x}_{(2)} - \mathrm{x}_{(1)}, \ldots, \mathrm{x}_{(n)} - \mathrm{x}_{(n-1)}, 1 - \mathrm{x}_{(n)}$$

совпадает с совместным распределением

$$\frac{\xi_1}{\zeta_{n+1}}, \frac{\xi_2}{\zeta_{n+1}}, \ldots, \frac{\xi_{n+1}}{\zeta_{n+1}},$$

так что *)

$$M_n \underset{d}{=} \zeta_{n+1}^{-2} \sum_{j=1}^{n+1} \xi_j^2.$$

*) Знак $\underset{d}{=}$ означает совпадение распределений.

Распределение M_n от α не зависит, и можно положить $\alpha = 1$. Тогда (см. § 12)

$$\mathbf{E}\xi_j^k = \Gamma(k+1) = k!, \quad \mathbf{D}\xi_j = 1, \quad \mathbf{D}\xi_j^2 = 20,$$

$$\rho_n \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\xi_j - 1) \Rightarrow \mathbf{\Phi}_{0,1}, \quad \eta_n \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\xi_j^2 - 2\right) \Rightarrow \mathbf{\Phi}_{0,20}.$$

Имеем

$$nM_{n-1} \underset{d}{=} \frac{n\left[2n + \sum\limits_{j=1}^{n}\left(\xi_j^2 - 2\right)\right]}{\left(n + \sum\limits_{j=1}^{n}(\xi_j - 1)\right)^2} = \frac{n(2n + \eta_n\sqrt{n})}{(n + \rho_n\sqrt{n})^2} = \frac{2 + \eta_n n^{-1/2}}{(1 + \rho_n n^{-1/2})^2}, \tag{7}$$

$$(nM_{n-1} - 2)\sqrt{n} = \frac{\eta_n - 4\rho_n - 2\rho_n^2 n^{-1/2}}{(1 + \rho_n n^{-1/2})^2}.$$

Здесь

$$\eta_n - 4\rho_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\xi_j' + 2), \quad \xi_j' = \xi_j^2 - 4\xi_j,$$

$$\mathbf{E}\xi_j' = -2, \quad \mathbf{D}\xi_j' = \mathbf{E}\left(\xi_j^4 - 8\xi_j^3 + 16\xi_j^2\right) - 4 = 4.$$

Стало быть, $\eta_n - 4\rho_n \Rightarrow \mathbf{\Phi}_{0,4}$, так что в силу теорем непрерывности из (7) получаем

$$\sqrt{n}\left(\frac{nM_{n-1}}{2} - 1\right) \Rightarrow \mathbf{\Phi}_{0,1}.$$

Это доказывает утверждение теоремы. $\triangleleft$

Приведем теперь соображения, показывающие что критерий Морана является состоятельным. Рассмотрим статистику (6) для $X \in \mathbf{P}$, где $\mathbf{P}$ отлично от $\mathbf{P}_1$. Одно из распределений $\mathbf{P}_1$ или $\mathbf{P}$ мы можем, не ограничивая общности, считать равномерным. Пусть это будет $\mathbf{P}$. Относительно F предположим для простоты, что существует непрерывная плотность $f(t) = F'(t)$, сосредоточенная на $[0,1]$. Тогда для $X \in \mathbf{U}_{0,1}$ главная часть nM_n будет равна

$$n\sum_{k=0}^{n}[f(\mathrm{x}_{(k+1)})(\mathrm{x}_{(k+1)} - \mathrm{x}_{(k)})]^2 \underset{d}{=} n\sum_{k=1}^{n+1}[f(\zeta_k/\zeta_{n+1})\xi_k/\zeta_{n+1}]^2. \tag{8}$$

По усиленному закону больших чисел $k^{-1}\zeta_k \underset{\text{п.н.}}{\longrightarrow} 1$ при $k \to \infty$. Поэтому главная часть (8) в свою очередь будет равна

$$\sum_{k=1}^{n} f^2(k/n)\xi_k^2/n. \tag{9}$$

Используя снова закон больших чисел (или неравенство Чебышева), получим, что это выражение сходится по вероятности к

$$2\int_0^1 f^2(t)\,dt \geqslant 2\left(\int_0^1 f(t)\,dt\right)^2 = 2.$$

Здесь знак неравенства строгий, если $f(t) \not\equiv 1$. Это значит, что при $X \in \mathbf{P} = \mathbf{U}_{0,1} \neq \mathbf{P}_1$ и при $n \to \infty$

$$\sqrt{n}\left(\frac{nM_n}{2} - 1\right) \xrightarrow[P]{} \infty,$$

что влечет за собой в силу теоремы 3 состоятельность критерия Морана любого фиксированного уровня $1 - \varepsilon$. $\triangleleft$

Будучи состоятельным, критерий Морана *не различает*, однако, *близких гипотез*. Предположим, что $X \in \mathbf{P} = \mathbf{U}_{0,1}$,

$$\begin{gathered} F(t) = t + p(t)n^{-1/2}, \quad t \in [0,1], \\ p(0) = p(1) = 0 \end{gathered} \tag{10}$$

и что функция $p(t)$ непрерывно дифференцируема. Тогда

$$n^{3/2}M_n = n^{3/2}\sum_{k=0}^{n}(\mathrm{x}_{(k+1)} - \mathrm{x}_{(k)})^2 +$$

$$+ 2n\sum_{k=0}^{n}(\mathrm{x}_{(k+1)} - \mathrm{x}_{(k)})(p(\mathrm{x}_{(k+1)}) - p(\mathrm{x}_{(k)})) + \sqrt{n}\sum_{k=0}^{n}(p(\mathrm{x}_{(k+1)} - p(\mathrm{x}_{(k)}))^2. \tag{11}$$

Главная часть второй суммы здесь равна $2n\sum_{k=0}^{n} p'(\mathrm{x}_{(k)})(\mathrm{x}_{(k+1)} - \mathrm{x}_{(k)})^2$ или в силу тех же соображений, что и в (9),

$$2\sum_{k=1}^{n} p'(k/n)\xi_k^2/n \xrightarrow[P]{} 4\int_0^1 p'(t)\,dt = 0.$$

Последнее слагаемое в (11) также сходится по вероятности к нулю, так как имеет главную часть, совпадающую по распределению с

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[p'(k/n)]^2\xi_k^2/n$$

или с $\dfrac{2}{\sqrt{n}}\displaystyle\int_0^1 [p'(t)]^2\,dt \to 0$. Изложенное означает, что для функции F вида (10) статистика $n^{3/2}M_n/2 - \sqrt{n}$ будет иметь то же предельное распределение $\Phi_{0,1}$, что и для $F(t) = t$. $\triangleleft$

Следует отметить, что из этого факта нельзя делать поспешных выводов о том, что критерий Морана плох. Дело в том, что, не различая близких гипотез вида (10), критерий Морана различает гипотезы (также в известном смысле близкие),

которые другие критерии, рассмотренные в этом параграфе, различить не могут. Речь идет о *гипотезах для плотностей.*

Рассмотрим гипотезу $H_2 = \{X \in \mathbf{P}\}$, где распределение $\mathbf{P}$ имеет плотность

$$f(t) = \begin{cases} 2 & \text{при} \quad 2k\Delta_n \leqslant t < (2k+1)\Delta_n, \\ 0 & \text{при} \quad (2k+1)\Delta_n \leqslant t < (2k+2)\Delta_n, \end{cases}$$

$k = 0, 1, \ldots, N-1$, где $\Delta_n = 1/2N$, $N = N_n > 0$ целое. Тогда при $\Delta_n = o(n^{-1/2})$ функция распределения $F_P(t)$, соответствующая распределению $\mathbf{P}$, будет обладать свойством

$$\sup_t |F_P(t) - t| = o(n^{-1/2}).$$

Это значит, что гипотеза H_2 как *гипотеза для функции распределения* будет настолько близкой к $H_1 = \{X \in \mathbf{U}_{0,1}\}$, что критерии Колмогорова и ω^2 асимптотически их различать не будут (предельное значение мощности в точке $\mathbf{P}$ будет совпадать с предельным уровнем критерия). Однако как гипотезы *для плотностей* гипотезы H_1 и H_2 существенно различны, так как $\sup|f(t) - 1| = 1$. Поскольку $x_{(0)} = 0$, $x_{(n+1)} = 1$, то для $X \in \mathbf{P}$ статистика M_n будет превосходить величину $\Delta_n^2 N = \Delta_n/2$. Стало быть, при $n/N = 2n\Delta_n \to \infty$ с $\mathbf{P}$-вероятностью, равной 1, будем иметь

$$nM_n \to \infty.$$

Фиксировав критическое множество $\Omega_2 = \{nM_n > 3\}$, получим $\mathbf{P}_1(\Omega_2) \to 0$. Это означает, что при $\Delta_n = o(n^{-1/2})$, $\Delta_n n \to \infty$ критерий Морана с вероятностью, близкой к 1, будет различать гипотезы H_1 и H_2. Другими словами, статистика M_n чувствительна к отклонениям в плотности, а сам критерий Морана можно рекомендовать как критерий для проверки гипотез, относящихся к *плотностям.* С другой стороны, мы знаем из § 10, что скорость сближения эмпирических плотностей с истинной медленнее, чем $n^{-1/2}$. Поэтому «неразличимость» гипотез о плотностях, отличающихся друг от друга на величину порядка $n^{-1/2}$ (см. (10)), не является удивительной.

В связи с критерием Морана и некоторыми другими критериями, рассмотренными ранее, можно сделать одно общее замечание. Если сравнивать два критерия одного фиксированного уровня, первый из которых предназначен для отбраковки бóльшего числа альтернатив, чем второй, то мощность первого критерия для каждой фиксированной альтернативы (отвергаемой обоими критериями) будет, как правило, меньше, чем мощность второго. Наиболее простой пример, иллюстрирующий это обстоятельство, читатель может получить, рассмотрев критерии $|x_1| > \lambda_{\varepsilon/2}$ и $x_1 > \lambda_\varepsilon$, предназначенные соответственно для проверки гипотез $\{\alpha \neq 0\}$ и $\{\alpha > 0\}$ против $\{\alpha = 0\}$ по наблюдению $x_1 \in \Phi_{\alpha,1}$. Здесь λ_ε есть квантиль распределения $\Phi_{0,1}$ порядка $1 - \varepsilon$. Мощности в точке $\alpha > 0$ будут равны соответственно

$$1 - \Phi_{0,1}(-\lambda_{\varepsilon/2} - \alpha, \lambda_{\varepsilon/2} - \alpha) < 1 - \Phi(\lambda_\varepsilon - \alpha).$$

§ 53. Асимптотически оптимальные критерии. Критерий отношения правдоподобия как асимптотически байесовский критерий для проверки простой гипотезы против сложной

1. Асимптотические свойства к. о. п. и байесовского критерия. Рассмотрим задачу проверки простой гипотезы $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$ против сложной альтернативы $H_2 = \{X \in \mathbf{P}_\theta;\ \theta \neq \theta_1,\ \theta \in \Theta\}$ в параметрическом случае.

В предыдущих параграфах мы видели на примерах, что р. н. м. к. в этом случае, как правило, не существует.

Будем рассматривать «частично байесовскую» постановку задачи, которая была описана в § 44, 49. Она состоит в предположении, что θ выбирается на $\Theta_2 = \Theta \setminus \{\theta_1\}$ случайно с распределением $\mathbf{Q}_2 = \mathbf{Q}$. Можно считать, что $\mathbf{Q}$ задано на Θ, $\mathbf{Q}(\{\theta_1\}) = 0$. В этом случае распределение выборки X будет определяться «усредненной» плотностью

$$f_Q(\boldsymbol{x}) = \int f_t(\boldsymbol{x})\mathbf{Q}(dt). \tag{1}$$

Таким образом, если $\mathbf{Q}$ известно, то гипотезу $H_{Q_2} = H_Q$, в силу которой X имеет распределение с плотностью (1), вместе с H_1 можно рассматривать как простую гипотезу и пользоваться леммой Неймана–Пирсона для построения наиболее мощного критерия.

Оказывается, что в этом случае для «почти всех» гладких $\mathbf{Q}$ наиболее мощные критерии будут асимптотически совпадать с критерием отношения правдоподобия

$$R(X) = \frac{\sup\limits_{\theta\in\Theta} f_\theta(X)}{f_{\theta_1}(X)} = \frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > c \tag{2}$$

и, стало быть, не зависеть от $\mathbf{Q}$. Этот факт позволяет считать найденный критерий асимптотически оптимальным по крайней мере в тех случаях, когда можно предполагать, что θ в Θ_2 выбирается случайно, но его распределение $\mathbf{Q}$ нам неизвестно.

Прежде чем формулировать соответствующую теорему, напомним некоторые нужные нам результаты и докажем одно вспомогательное утверждение. Главную роль в нем будут играть уже известные нам асимптотические свойства отношения правдоподобия. Будем сразу рассматривать случай многомерного параметра; все необходимое для этого имеется в § 36, 37.

Итак, пусть $\theta \in \Theta \subset \mathbb{R}^k$, $k \geqslant 1$, и выполнены условия регулярности (RR), формулировка которых дана в § 36. Пусть, кроме того, $\mathbf{Q}$ имеет плотность $q(t)$ относительно меры Лебега $\lambda(dt) = dt$.

В соответствии с леммой Неймана–Пирсона наиболее мощный нерандомизированный критерий $\pi_{Q_2} = \pi_Q$ для проверки H_1 против H_Q будет иметь следующий вид: $\pi_Q(X) = 1$, если

$$X \in \Omega(c) = \left\{\boldsymbol{x} : \frac{f_Q(\boldsymbol{x})}{f_{\theta_1}(\boldsymbol{x})} > c\right\}, \quad f_Q(\boldsymbol{x}) = \int q(u) f_u(\boldsymbol{x})\, du, \tag{3}$$

где $c = c_n$ выберем позже по заданному уровню критерия.

Такой же вид будут иметь и байесовские критерии для проверки H_1 против H_Q.

Вероятности ошибок первого и второго рода равны соответственно

$$\begin{aligned} \alpha_1(\pi_Q) &= \mathbf{P}_{\theta_1}\left(\frac{f_Q(X)}{f_{\theta_1}(X)} > c\right), \\ 1 - \beta(\pi_Q) &= \int q(t)\mathbf{P}_t\left(\frac{f_Q(X)}{f_{\theta_1}(X)} \leqslant c\right) dt, \end{aligned} \tag{4}$$

где $\beta(\pi_Q) = \int\limits_{\{f_Q(x)\leqslant cf_{\theta_1}(x)\}} f_Q(x)\mu^n(dx)$ есть мощность наиболее мощного критерия.

Аналогичные выражения сможем записать для к. о. п. $\widehat{\pi}$, который принимает H_Q при выполнении (2):

$$\alpha_1(\widehat{\pi}) = \mathbf{P}_{\theta_1}\left(\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > c\right),$$

$$\alpha_2(\widehat{\pi}) = \int q(t)\mathbf{P}_t\left(\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} \leqslant c\right)dt = \int\limits_{\{f_{\widehat{\theta}_*}(x)<cf_{\theta_1}(x)\}} f_Q(x)\mu^n(dx). \tag{5}$$

Положим $I = I(\theta_1)$ (значение информационной матрицы Фишера в точке θ_1),

$$\frac{f_Q(X)}{f_{\theta_1}(X)} \equiv \left(\frac{2\pi}{n}\right)^{k/2}\frac{q(\theta_1)}{\sqrt{|I|}}e^{T(X)}, \quad \frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} \equiv e^{\widehat{T}(X)}. \tag{6}$$

Тогда критические области критериев π_Q и $\widehat{\pi}$ (см. (3), (2)) можно записать соответственно в виде

$$T(X) > c_Q, \quad \widehat{T}(X) > \widehat{c}. \tag{7}$$

Лемма 1. *Пусть выполнены условия* (RR) § 36, $X \in \mathbf{P}_{\theta_1}$, θ_1 — *внутренняя точка* Θ. *Тогда*

$$2T(X) = 2\widehat{T}(X)(1+\varepsilon_n(X)) \Subset \mathbf{H}_k, \quad \varepsilon_n(X) \underset{\text{п.н.}}{\longrightarrow} 0.$$

Доказательство. Утверждение леммы есть очевидное следствие теорем 36.4, 36.5. Надо лишь заметить, что $\widehat{T}(X)$ в обозначениях теоремы 36.4 есть не что иное, как $Y(u^*)$ (при $\theta = \theta_1$). $\triangleleft$

2. Асимптотическая байесовость к. о. п. Перейдем к формулировке основного утверждения. Напомним, что когда мы изучаем асимптотические свойства критериев, на самом деле имеем в виду не один, а целую последовательность критериев $\pi = \pi_n$, где π_n есть критерий, основанный на выборке X_n. Такую же ситуацию мы имели, рассматривая асимптотические свойства оценок. Таким образом, здесь и в дальнейшем — везде, где это потребуется — под критерием π мы будем понимать последовательность функций $\pi_n(X_n)$, определенных при каждом n и $X_n = [X_\infty]_n$.

Определение 1. *Критерий* π для проверки гипотезы $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$ *принадлежит классу* $\widetilde{K}_\varepsilon$ *критериев асимптотического уровня* $1-\varepsilon$, если

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta_1}\mathbf{E}_\theta\pi(X) \leqslant \varepsilon. \tag{8}$$

В нашем случае, когда гипотеза H_1 является простой и $\Theta_1 = \{\theta_1\}$, соотношение (8) превращается в неравенство

$$\limsup_{n\to\infty} \mathbf{E}_{\theta_1}\pi(X) \leqslant \varepsilon.$$

Пусть h_ε есть квантиль порядка $1-\varepsilon$ распределения χ^2 с k степенями свободы ($\mathbf{H}_k((h_\varepsilon,\infty)) = \varepsilon$). Тогда из леммы 1 вытекает, что $\pi_Q \in \widetilde{K}_\varepsilon$, $\widehat{\pi} \in \widetilde{K}_\varepsilon$, если $c_Q = \widehat{c} = h_\varepsilon/2$.

Определение 2. Положим $c_Q = h_\varepsilon/2$, так что $\pi_Q \in \widetilde{K}_\varepsilon$. Критерий $\pi \in \widetilde{K}_\varepsilon$ называется *асимптотически байесовским критерием* (а. б. к.) в $\widetilde{K}_\varepsilon$ для проверки гипотезы $H_1 = \{\theta = \theta_1\}$ против H_Q, если для вероятностей ошибок второго рода, вычисленных при гипотезе H_Q, справедливо соотношение

$$\limsup_{n\to\infty} \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = \limsup_{n\to\infty} \frac{1-\beta(\pi)}{1-\beta(\pi_Q)} = \limsup_{n\to\infty} \frac{\mathbf{E}_Q(1-\pi(X))}{\mathbf{E}_Q(1-\pi_Q(X))} = 1.$$

Мы использовали в этом определении отношение (а не разность) вероятностей ошибок второго рода, так как $\alpha_2(\pi_Q) \to 0$ при $n \to \infty$.

Теорема 1. *Пусть выполнены условия* (RR), *точка* θ_1 *является внутренней точкой* Θ. *Тогда критерий отношения правдоподобия* $\widehat{\pi}$ (см. (2), (7)) *при* $\widehat{c} = h_\varepsilon/2$ *принадлежит* $\widetilde{K}_\varepsilon$ *и является а. б. к. в* $\widetilde{K}_\varepsilon$ *для проверки* H_1 *против* H_Q *при любом распределении* $\mathbf{Q}$, *плотность* $q(t)$ *которого непрерывна и положительна в* Θ. *При этом*

$$\alpha_2(\widehat{\pi}) \sim \alpha_2(\pi_Q) \sim \frac{q(\theta_1)}{n^{k/2}\sqrt{|I|}} V_k h_\varepsilon^{k/2},$$

где $I = I(\theta_1)$, V_k — *объем единичного шара в* $\mathbb{R}^k$.

Доказательство. Принадлежность $\widehat{\pi} \in \widetilde{K}_\varepsilon$ при $\widehat{c} = h_\varepsilon/2$ нами уже доказана. Рассмотрим теперь ошибки второго рода. В силу (4), (7) имеем

$$\alpha_2(\pi_Q) = \int\limits_{\{T(\boldsymbol{x})\leqslant c_Q\}} f_Q(\boldsymbol{x})\mu^n(d\boldsymbol{x}) = \mathbf{E}_{\theta_1}\left\{\frac{f_Q(X)}{f_{\theta_1}(X)};\ 2T(X) \leqslant h_\varepsilon\right\} =$$

$$= \left(\frac{2\pi}{n}\right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}} \mathbf{E}_{\theta_1}\{e^{T(X)};\ 2T(X) \leqslant h_\varepsilon\}.$$

Здесь под знаком математического ожидания стоит ограниченная функция от $2T$, непрерывная п. в. относительно предельного распределения $\mathbf{H}_k$. По-

этому при $n \to \infty$, $\chi_k^2 \Subset \mathbf{H}_k$

$$\mathbf{E}_{\theta_1}(e^{T(X)}; 2T(X) \leqslant h_\varepsilon) \to \mathbf{E}\{e^{\frac{1}{2}\chi_k^2}; \chi_k^2 \leqslant h_\varepsilon\} =$$
$$= (2\pi)^{-k/2} \int\limits_{\{|y|^2 \leqslant h_\varepsilon\}} e^{\frac{1}{2}|y|^2 - \frac{1}{2}|y|^2}\, dy_1 \ldots dy_k = (2\pi)^{-k/2} h_\varepsilon^{k/2} V_k.$$

Найдем теперь асимптотическое поведение $\alpha_2(\widehat{\pi})$. Положим $A_n = \{X : \pi_Q \neq \neq \widehat{\pi}\}$. В силу леммы 1 $\mathbf{P}_{\theta_1}(A_n) \to 0$. Поэтому из теоремы 37.5 вытекает, что при любом фиксированном N

$$\sup_{|u| \leqslant N} \mathbf{P}_{\theta + u/\sqrt{n}}(A_n) \to 0. \tag{9}$$

Воспользуемся представлением (см. (5))

$$\alpha_2(\widehat{\pi}) = \int q(t)\mathbf{P}_t(\widehat{T}(X) \leqslant \widehat{c})\, dt =$$
$$= \int\limits_{|t-\theta_1| \leqslant N/\sqrt{n}} + \int\limits_{|t-\theta_1| > N/\sqrt{n}} \leqslant \int q(t)\mathbf{P}_t(T(X) \leqslant \widehat{c})\, dt +$$
$$+ \int\limits_{|t-\theta_1| \leqslant N/\sqrt{n}} q(t)\mathbf{P}_t(A_n)\, dt + \int\limits_{|t-\theta_1| > N/\sqrt{n}} q(t)\mathbf{P}_t(\widehat{T}(X) \leqslant c)\, dt.$$

В силу (9) мы получаем

$$\limsup_{n\to\infty} n^{k/2}\alpha_2(\widehat{\pi}) \leqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q) +$$
$$+ \max_t q(t) \limsup_{n\to\infty} \int\limits_{|t-\theta_1| > N/\sqrt{n}} \mathbf{P}_t\left(\frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} \leqslant e^{\widehat{c}}\right) dt.$$

Но вероятность под знаком интеграла не превосходит

$$\mathbf{P}_t\left(\frac{f_{\theta_1}(X)}{f_t(X)} \geqslant e^{-\widehat{c}}\right) \leqslant \exp\left\{\frac{\widehat{c}}{2} - \frac{|t-\theta_1|^2 ng}{2}\right\}. \tag{10}$$

Здесь мы использовали теорему 36.1. Следовательно, сам интеграл не превосходит

$$\frac{1}{\sqrt{n}} e^{\widehat{c}/2} \int\limits_{|u|>N} e^{-|u|^2 g/2}\, du \to 0$$

при $N \to \infty$. Отсюда следует, что

$$\limsup_{n\to\infty} n^{k/2}\alpha_2(\widehat{\pi}) \leqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q). \tag{11}$$

Очевидно, это равнозначно тому, что $\widehat{\pi}$ есть а. б. к.

Нам осталось установить только, что $\alpha_2(\widehat{\pi}) \sim \alpha_2(\pi_Q)$ или, что то же в силу (11),

$$\liminf_{n\to\infty} n^{k/2}\alpha_2(\widehat{\pi}) \geqslant \lim_{n\to\infty} n^{k/2}\alpha_2(\pi_Q). \tag{12}$$

Для этого заметим, что построенный нами критерий π_Q является байесовским, соответствующим априорной вероятности q_1 гипотезы H_1, определяемой из уравнения (ср. (3), (6))

$$\frac{q_1}{1-q_1} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{q(\theta_1)}{\sqrt{|I|}} e^{\widehat{c}}.$$

Это означает, что вероятность ошибки π_Q будет вести себя асимптотически как

$$\varepsilon q_1 + (1-q_1)\alpha_2(\pi_Q) \sim \varepsilon q_1 + \alpha_2(\pi_Q).$$

Если допустить, что (12) неверно, то мы получим критерий $\widehat{\pi}$, для которого вероятность ошибки будет меньше. Так как это невозможно, (12) доказано. Теорема полностью доказана. $\triangleleft$

Из приведенных рассуждений видно, что основной вклад в вероятности ошибок второго рода вносят случайные значения θ, попадающие в $n^{-1/2}$-окрестность точки θ_1 (этим обуславливается порядок малости $n^{-k/2}$ этих вероятностей).

Незначительные изменения в рассуждениях доказательства теоремы 1 позволяют получить также следующее утверждение.

Теорема 2. *Критерии π' и π'' с критическими областями*

$$\begin{aligned} \Omega' &= \{\boldsymbol{x} \in \mathcal{X}^n : n(\widehat{\theta}^* - \theta_1)I(\theta_1)(\widehat{\theta}^* - \theta_1)^\top > h_\varepsilon\}, \\ \Omega'' &= \{\boldsymbol{x} \in \mathcal{X}^n : L'(\boldsymbol{x},\theta_1)I^{-1}(\theta_1)(L'(\boldsymbol{x},\theta_1))^\top > h_\varepsilon\} \end{aligned} \tag{13}$$

являются наряду с $\widehat{\pi}$ а. б. к. в $\widetilde{K}_\varepsilon$. Это свойство сохранится, если в (13) *$I(\theta_1)$ заменить на $I(\widehat{\theta}^*)$.*

Критерии (13) получаются, если воспользоваться разложением

$$\ln \frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} = L(X,\widehat{\theta}^*) - L(X,\theta_1)$$

в ряд около точки $\widehat{\theta}^*$ (см. теорему 36.4). Форма критерия $\widehat{\pi}$ в известном смысле более удобна, так как она с размерностью не связана.

Доказательство теоремы 2 мы предоставляем читателю.

В одномерном случае критическое множество Ω' (при замене $I(\theta_1)$ на $I(\widehat{\theta}^*)$) имеет вид

$$\Omega' = \left\{ |\widehat{\theta}^* - \theta_1| > \left[\frac{h_\varepsilon}{n|I(\widehat{\theta}^*)|}\right]^{1/2} \right\}, \tag{14}$$

где, очевидно, $h_\varepsilon = \lambda^2_{\varepsilon/2}$, $\Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$. Мы видим, что соответствующий (14) критерий π', асимптотически эквивалентный $\widehat{\pi}$, можно интерпретировать так: $\pi'(X) = 1$, *если* θ_1 *не попало в доверительный интервал асимптотического уровня* $1 - \varepsilon$ *для параметра* θ, *построенный с помощью о. м. п.* $\widehat{\theta}^*$.

Такая же интерпретация сохранится, очевидно, и в многомерном случае; доверительные множества при этом будут иметь форму эллипсоидов:

$$(\widehat{\theta}^* - \theta) I(\widehat{\theta}^*) (\widehat{\theta}^* - \theta)^\top \leqslant n^{-1} h_\varepsilon.$$

Мы видим, таким образом, что о. м. п. тесно связаны с а. б. к.

Пример 1. Пусть $X \in \mathbf{\Pi}_\lambda$ и проверяется гипотеза $H_1 = \{\lambda = \lambda_1\}$ против $H_2 = \{\lambda \neq \lambda_1\}$. В этом случае $\widehat{\lambda}^* = \overline{\mathrm{x}}$, $I(\lambda) = \lambda^{-1}$, и а. б. к. будет иметь вид

$$(\overline{\mathrm{x}} - \lambda_1)^2 > \frac{h_\varepsilon \lambda_1}{n},$$

где $\mathbf{H}_1((h_\varepsilon, \infty)) = \varepsilon$.

Пример 2. Пусть $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$ и проверяется гипотеза $H_1 = \{(\alpha, \sigma^2) = (\alpha_1^2, \sigma_1^2)\}$ против дополнительной альтернативы. Здесь $\widehat{\alpha}^* = \overline{\mathrm{x}}$, $\widehat{\sigma}^{2*} = S^2 = \frac{1}{n}\sum_{i=1}^n (\mathrm{x}_i - \overline{\mathrm{x}})^2$, $I(\alpha, \sigma^2) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4}/2 \end{pmatrix}$ (см. § 26). Поэтому а. б. к. имеет вид

$$\frac{(\overline{\mathrm{x}} - \alpha_1)^2}{\sigma_1^2} + \frac{(S^2 - \sigma_1^2)}{2\sigma_1^4} > \frac{h_\varepsilon}{n},$$

где $\mathbf{H}_2((h_\varepsilon, \infty)) = \varepsilon$.

3. Асимптотическая несмещенность к. о. п. В заключение этого параграфа установим, что к. о. п. (2) является асимптотически несмещенным. Напомним предварительно, что критерий π для проверки $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$ называется несмещенным, если

$$\inf_{\theta \in \Theta_2} \mathbf{E}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{E}_\theta \pi \geqslant 0.$$

Определение 3. Критерий π называется *асимптотически несмещенным*, если

$$\liminf_{n \to \infty} \left(\inf_{\theta \in \Theta_2} \mathbf{E}_\theta \pi - \sup_{\theta \in \Theta_1} \mathbf{E}_\theta \pi \right) \geqslant 0.$$

Теорема 3. *К. о. п.* $\widehat{\pi}$ (см. (2), (6), (7)) *для проверки* $H_1 = \{\theta = \theta_1\}$ *против* $H_2 = \{\theta \neq \theta_1\}$ *является асимптотически несмещенным.*

Доказательство. Так как в нашем случае $\Theta_1 = \{\theta_1\}$ и $\lim\limits_{n \to \infty} \mathbf{E}_{\theta_1} \widehat{\pi} = \varepsilon$, то достаточно убедиться, что

$$\liminf_{n \to \infty} \inf_{t \in \Theta} \mathbf{E}_t \widehat{\pi} = \liminf_{n \to \infty} \inf_{t \in \Theta} \mathbf{P}_t \left(\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{\widehat{c}} \right) \geqslant \varepsilon, \tag{15}$$

где $\widehat{c} = h_\varepsilon/2$.

Из оценки (10) следует, что существует $N > 0$ такое, что

$$\inf_{|t-\theta_1|\geqslant N/\sqrt{n}} \mathbf{P}_t\left(\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{\widehat{c}}\right) > \varepsilon.$$

Остается доказать, что $\inf\limits_{|t-\theta_1|\leqslant N/\sqrt{n}} \mathbf{E}_t\widehat{\pi} \to \varepsilon$.

Но в силу теорем 36.4, 37.3 при $X \in \mathbf{P}_t$ равномерно по u, $|u| \leqslant N$, $u = \sqrt{n}\,(t-\theta_1)$,

$$\widehat{T}(X) \Rightarrow \frac{1}{2}(\xi-u)I(\xi-u)^{\top}, \quad \xi \in \mathbf{\Phi}_{0,I^{-1}},$$

$$\mathbf{E}_t\widehat{\pi} \to \mathbf{P}_t\left(\frac{1}{2}(\xi-u)I(\xi-u)^{\top} > \hat{c} = \frac{h_\varepsilon}{2}\right).$$

Правая часть здесь достигает своего наименьшего значения при $u = 0$. Это значение равно $\mathbf{P}(\xi I\xi^{\top} > h_\varepsilon) = \varepsilon$. $\triangleleft$

§ 54. Асимптотически оптимальные критерии для проверки близких сложных гипотез

1. Постановка задачи и определения. В § 43 мы обсуждали два асимптотических подхода в задаче о проверке двух простых гипотез H_1 и H_2. Если эти гипотезы считать фиксированными, т. е. неизменными при растущем объеме n выборки X_n, то мы приходим при вычислении вероятностей ошибок к задаче о вероятностях больших уклонений, так что по крайней мере вероятность одной из ошибок сходится к нулю. Согласно другому подходу гипотезы H_1 и H_2 рассматриваются как элементы последовательности «сближающихся» гипотез, при этом скорость сближения подбирается так, чтобы вероятности ошибок первого и второго рода сходились к собственным (отличным от нуля и единицы) пределам. Мы видели, что в параметрическом случае значения параметра θ_1 и θ_2, соответствующие близким гипотезам H_1 и H_2, должны отличаться на величину порядка $n^{-1/2}$. Каждый из этих подходов может быть оправдан в зависимости от конкретных условий.

В предыдущем параграфе мы рассматривали не зависящее от n распределение $\mathbf{Q}$ для альтернативного значения θ и, как это естественно было ожидать, получили, что вероятность ошибки второго рода сходится к нулю как $n^{-k/2}$. Это обусловлено тем, что основной вклад в эту вероятность вносят близкие гипотезы, для которых θ удалено от θ_1 на расстояние порядка $n^{-1/2}$ (объем области, содержащей такие θ, и будет иметь порядок малости $n^{-k/2}$).

В этом параграфе мы рассмотрим задачу о проверке *близких сложных гипотез*, когда альтернативные значения параметра сближаются при $n \to \infty$. Оказывается, что в этом случае задачу проверки гипотез можно в известном смысле редуцировать к значительно более простой задаче для нормального распределения.

Перейдем к более точным формулировкам. Пусть по выборке $X \in \mathbf{P}_\theta$ проверяется гипотеза $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$. Фиксируем какую-нибудь внутреннюю точку θ_1 множества Θ и положим

$$\theta = \theta_1 + \gamma n^{-1/2}. \tag{1}$$

Предположим теперь, что множества Θ_i имеют вид

$$\Theta_i = \theta_1 + \Gamma_i n^{-1/2}, \tag{2}$$

где Γ_i от n не зависят. Запись (2) означает, что $\theta \in \Theta_i$ тогда и только тогда, когда в (1) $\gamma \in \Gamma_i$. Гипотезы $H_i = \{\theta \in \Theta_i\}$ при условии (1) мы будем, как и в § 43, называть *близкими* (на самом деле это *последовательность* гипотез — своих для каждого n).

Задачу проверки близких гипотез H_i по выборке $X \in \mathbf{P}_\theta$ будем называть *задачей* А.

Рассмотрим теперь другую задачу. Пусть $Y \in \mathbf{\Phi}_{\gamma, I^{-1}}$ есть выборка единичного объема из нормальной совокупности $\mathbf{\Phi}_{\gamma, I^{-1}}$ с вектором средних значений γ и матрицей вторых моментов $I^{-1} = I^{-1}(\theta_1)$, где $I(\theta_1)$ есть информационная матрица Фишера для задачи А в точке θ_1. Обозначим через $\mathfrak{h}_i$ гипотезы $\{\gamma \in \Gamma_i\}$. Задачу проверки гипотез $\mathfrak{h}_i$ по одному наблюдению $Y \in \mathbf{\Phi}_{\gamma, I^{-1}}$ будем называть *задачей* В.

Замечательный факт, позволяющий делать упомянутую выше редукцию, состоит, грубо говоря, в следующем. Пусть $\pi(Y)$ есть оптимальный в том или ином смысле критерий (р. н. м. к., байесовский, минимаксный) для проверки $\mathfrak{h}_1$ против $\mathfrak{h}_2$ в задаче В. И пусть $\widehat{\theta}^*$, как обычно, есть о. м. п. в задаче А, $\gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n}$. Тогда критерий $\pi(\gamma^*)$ для проверки H_1 против H_2 в задаче А будет *асимптотически* обладать теми же свойствами оптимальности, что и критерий $\pi(Y)$ в задаче В.

Таким образом, чтобы найти асимптотически оптимальный критерий в задаче А, мы должны рассмотреть более простую задачу В и найти в ней (если это возможно) критерий π, обладающий нужным свойством оптимальности. Если теперь взять в качестве наблюдения Y значение γ^* и подставить его в π, то мы получим искомый критерий в задаче А.

Этот факт можно было бы назвать *предельным признаком оптимальности*. Смысл его довольно прост. Ведь мы знаем из результатов главы 2, что при $X \in \mathbf{P}_\theta$

$$\sqrt{n}\,(\widehat{\theta}^* - \theta) I^{1/2}(\theta) \Subset \mathbf{\Phi}_{0,E}$$

равномерно по θ. Стало быть, для $\theta = \theta_1 + \gamma n^{-1/2}$

$$\sqrt{n}\,(\widehat{\theta}^* - \theta_1) - \gamma \Subset \mathbf{\Phi}_{0, I^{-1}(\theta_1)}$$

или, что то же,

$$\gamma^* \Subset \mathbf{\Phi}_{\gamma, I^{-1}}.$$

Таким образом, $\mathbf{\Phi}_{\gamma, I^{-1}}$ — распределение, присутствующее в задаче В, есть не что иное, как предельное распределение для γ^*. Поэтому предельный признак оптимальности является весьма естественным: он сводит задачу проверки гипотез к «предельной» задаче. Замечательным во всем этом является тот факт, что при такой редукции никакой существенной потери информации относительно θ не происходит: критерий, оптимальный в задаче В, сохраняет эту оптимальность и для задачи А.

Чтобы придать изложенному точный смысл, введем теперь основные понятия асимптотической оптимальности критериев для проверки близких гипотез в задаче А.

Определение класса $\widetilde{K}_\varepsilon$ критериев π асимптотического уровня $1-\varepsilon$ дано в предыдущем параграфе (определение 53.1). Для $\pi \in \widetilde{K}_\varepsilon$ справедливо

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi(X) \leqslant \varepsilon.$$

Определение 1. Критерий $\pi_1 \in \widetilde{K}_\varepsilon$ называется *асимптотически р. н. м. к.* (а. р. н. м. к.) *в* $\widetilde{K}_\varepsilon$, если для любого $\gamma \in \Gamma_2$ и любого $\pi \in \widetilde{K}_\varepsilon$

$$\liminf_{n\to\infty}(\mathbf{E}_\theta \pi_1(X) - \mathbf{E}_\theta \pi(X)) \geqslant 0,$$

где $\theta = \theta_1 + \gamma n^{-1/2} \in \Theta_2$ при $\gamma \in \Gamma_2$.

Пусть на Γ_i заданы распределения $\mathbf{\Pi}_i$. Эти распределения индуцируют на Θ_i некоторые другие распределения (сосредоточенные в $n^{-1/2}$-окрестности точки θ_1), которые мы обозначим через $\mathbf{Q}_i$, $i = 1, 2$. Гипотезы о том, что θ выбирается случайно с распределением $\mathbf{Q}_i$, как и прежде, обозначим через H_{Q_i}.

Через $\widetilde{K}_\varepsilon^{Q_1}$ будем обозначать класс критериев π, для которых

$$\limsup_{n\to\infty} \mathbf{E}_{Q_1} \pi(X) \leqslant \varepsilon,$$

где $\mathbf{E}_{Q_i}$ означает безусловное математическое ожидание по совместному распределению θ и X, $\theta \in \mathbf{Q}_i$, $X \in \mathbf{P}_\theta$. Очевидно, что $\widetilde{K}_\varepsilon \subset \widetilde{K}_\varepsilon^{Q}$ при любом $\mathbf{Q}$.

Определение 2. Критерий $\pi_1 \in \widetilde{K}_\varepsilon^{Q_1}$ для проверки H_{Q_1} против H_{Q_2} называется *асимптотически байесовским критерием* (а. б. к.) *в* $\widetilde{K}_\varepsilon^{Q_1}$, если для любого другого критерия $\pi \in \widetilde{K}_\varepsilon^{Q_1}$

$$\liminf_{n\to\infty}(\mathbf{E}_{Q_2} \pi_1(X) - \mathbf{E}_{Q_2} \pi(X)) \geqslant 0. \tag{3}$$

Можно дать эквивалентное определение асимптотической байесовости, в котором вместо (3) требуется, чтобы

$$\liminf_{n\to\infty}(\mathbf{E}_{Q_2} \pi_1(X) - \mathbf{E}_{Q_2} \pi_{Q_1Q_2}(X)) \geqslant 0, \tag{4}$$

где $\pi_{Q_1Q_2}$ — байесовский критерий из $\widetilde{K}_\varepsilon^{Q_1}$ для проверки гипотез H_{Q_1} против H_{Q_2} (или, что то же, н. м. к. критерий для проверки H_{Q_1} против H_{Q_2} асимптотического уровня $1-\varepsilon$).

Следует отметить, что определение 2 несколько отличается от определения а. б. к., которое было дано в предыдущем параграфе (см. определение 43.2; там фигурирует отношение вероятностей ошибок, а не разность). С точки зрения последующего изложения эти определения эквивалентны, но последнее из них нам будет удобнее.

Определение 3. Критерий $\pi_1 \in \widetilde{K}_\varepsilon$ называется *асимптотически минимаксным в* $\widetilde{K}_\varepsilon$ критерием для проверки H_1 против H_2, если для любого другого критерия $\pi \in \widetilde{K}_\varepsilon$ выполняется

$$\liminf_{n\to\infty}(\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_1(X) - \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi(X)) \geqslant 0. \tag{5}$$

Как и при рассмотрении обычных минимаксных критериев (см. § 49), множества Θ_1 и Θ_2 во избежание малосодержательных рассмотрений разумно разделить некоторой промежуточной зоной так, чтобы они не соприкасались. В противном случае оба нижних предела в (5) могут оказаться равными ε для любого асимптотически несмещенного критерия π.

Из приведенных определений видно, что свойство той или иной асимптотической оптимальности отличается от обычного свойства той же оптимальности лишь тем, что перед соответствующей разностью появляется знак $\liminf$ вместо $\inf$.

Наряду с асимптотически байесовскими и минимаксными критериями в *классах* $\widetilde{K}_\varepsilon$ и $\widetilde{K}_\varepsilon^{Q_1}$ можно изучать *обычные* асимптотически байесовские и минимаксные критерии. Пусть на $\Theta = \Theta_1 \cup \Theta_2$ задано распределение $\mathbf{Q} = q(1)\mathbf{Q}_1 + q(2)\mathbf{Q}_2$, $q(1) + q(2) = 1$. Тогда критерий π_1 называется *асимптотически байесовским для априорного распределения* $\mathbf{Q}$, если для любого другого критерия π

$$\liminf_{n\to\infty} [q(1)\mathbf{E}_{Q_1}\pi_1(X) + q(2)\mathbf{E}_{Q_2}(1-\pi_1(X)) - \\ - q(1)\mathbf{E}_{Q_1}\pi(X) - q(2)\mathbf{E}_{Q_2}(1-\pi(X))] \leqslant 0. \quad (6)$$

Усредненная по $\mathbf{Q}$ вероятность ошибки критерия π, присутствующая в этом неравенстве, может быть записана с помощью вероятности $\alpha(\pi,\theta)$ ошибки в точке θ в виде $\mathbf{E}_Q\alpha(\pi,\theta)$, где

$$\alpha(\pi,\theta) = \begin{cases} \mathbf{E}_\theta\pi(X) & \text{при} \quad \theta \in \Theta_1, \\ \mathbf{E}_\theta(1-\pi(X)) & \text{при} \quad \theta \in \Theta_2. \end{cases}$$

Тогда неравенство (6) примет вид

$$\liminf_{n\to\infty} \mathbf{E}_Q[\alpha(\pi_1(X),\theta) - \alpha(\pi(X),\theta)] \leqslant 0.$$

Критерий π_1 будет *асимптотически минимаксным*, если

$$\liminf_{n\to\infty} [\sup_{\theta\in\Theta} \alpha(\pi_1,\theta) - \sup_{\theta\in\Theta} \alpha(\pi,\theta)] \leqslant 0$$

для любого другого критерия π.

Изучение асимптотических байесовских (в $\widetilde{K}_\varepsilon^{Q_1}$) и асимптотически минимаксных (в $\widetilde{K}_\varepsilon$) критериев и просто асимптотически байесовских и минимаксных критериев, по существу, представляет собой одно и то же. Например, байесовский критерий в $\widetilde{K}_\varepsilon^{Q_1}$ есть обычный байесовский критерий при соответствующем $q(1)$. В этом параграфе мы будем изучать критерии из классов $\widetilde{K}_\varepsilon$, $\widetilde{K}_\varepsilon^{Q_1}$; обычные асимптотически байесовские и минимаксные критерии будут рассмотрены в главе 6 при исследовании более общей постановки задачи.

2. Основные утверждения. Чтобы максимально упростить дальнейшее изложение, введем одно предположение, которое с существом дела никак не связано и которое при желании может быть убрано — все необходимые для этого результаты имеются. Именно, будем предполагать, что множества Γ_i ограничены, т. е. существует $N > 0$ такое, что $\Gamma_i \subset \{\gamma : |\gamma| \leqslant N\}$.

Определение 4. Критерии π_1, π_2 для проверки близких гипотез $H_1 = \{\theta \in \Theta_1\}$, $H_2 = \{\theta \in \Theta_2\}$ по выборке X называются *асимптотически эквивалентными*, если

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_1\cup\Theta_2} |\mathbf{E}_\theta \pi_1(X) - \mathbf{E}_\theta \pi_2(X)| = 0. \tag{7}$$

При сделанном предположении можем под знаком sup в (7) поставить область $|\theta - \theta_1| \leqslant N/\sqrt{n}$.

Асимптотически эквивалентные критерии π_1, π_2 обладают следующими свойствами:

1) *если* $\pi_1 \in \widetilde{K}_\varepsilon$ (*или* $\widetilde{K}_\varepsilon^{Q_1}$), *то* $\pi_2 \in \widetilde{K}_\varepsilon$ ($\widetilde{K}_\varepsilon^{Q_1}$);

2) *если* π_1 *обладает одним из свойств асимптотической оптимальности в определениях* 1–3, *то тем же свойством будет обладать критерий* π_2.

Первое утверждение вытекает из (7) и неравенства

$$\sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_2(X) \leqslant \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_1(X) + \sup_{\theta\in\Theta_1} |\mathbf{E}_\theta(\pi_2 - \pi_1)|.$$

Второе утверждение доказывается аналогично. Если, например, π_1 является асимптотически минимаксным, то асимптотическая минимаксность π_2 будет следствием (7) и неравенства

$$\inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_2(X) \geqslant \inf_{\theta\in\Theta_2} \mathbf{E}_\theta \pi_1(X) - \sup_{\theta\in\Theta_2} |\mathbf{E}_\theta(\pi_2 - \pi_1)|. \quad \triangleleft$$

Условия асимптотической эквивалентности критериев устанавливает

Лемма 1. *Пусть в окрестности точки* θ_1 *выполнены условия* (RR), $\pi_i(X) = I_{\{T_n(X)+\varepsilon_{ni}(X)>c\}}$, $i = 1, 2$, *где для* $X \in \mathbf{P}_{\theta_1}$ *имеют место соотношения* $\varepsilon_{ni}(X) \underset{P_{\theta_1}}{\longrightarrow} 0$, $T_n(X) \Rightarrow \mathbf{G}$, *распределение* $\mathbf{G}$ *непрерывно. Тогда критерии* π_1, π_2 *асимптотически эквивалентны.*

Доказательство. Имеем $|\mathbf{E}_t\pi_1(X) - \mathbf{E}_t\pi_2(X)| \leqslant \mathbf{P}_t(A_n)$, где для события $A_n = \{\pi_1(X) \neq \pi_2(X)\}$ выполняется $\mathbf{P}_{\theta_1}(A_n) = \mathbf{P}_{\theta_1}(T_n(X) + \varepsilon_{n1}(X) > c,\ T_n(X) + \varepsilon_{n2}(X) \leqslant c) + \mathbf{P}_{\theta_1}(T_n(X) + \varepsilon_{n1}(X) \leqslant c,\ T_n(X) + \varepsilon_{n2}(X) > c) \to 0$ при $n \to \infty$, так как предельное распределение T_n непрерывно. Стало быть, в силу теоремы 37.5

$$\sup_{|t-\theta_1|\leqslant N/\sqrt{n}} \mathbf{P}_t(A_n) \to 0. \quad \triangleleft$$

Байесовский критерий уровня $1 - \varepsilon$ в задаче B для проверки гипотез $\mathfrak{h}_{\Pi_i}$ о том, что γ выбирается случайно с распределением $\mathbf{\Pi}_i$ на Γ_i, $i = 1, 2$, обозначим через $\pi_{\Pi_1\Pi_2}(Y)$. Он имеет вид

$$r(Y) \equiv \frac{\int \exp\left\{-\frac{1}{2}(Y-u)I(Y-u)^\top\right\} \mathbf{\Pi}_2(du)}{\int \exp\left\{-\frac{1}{2}(Y-u)I(Y-u)^\top\right\} \mathbf{\Pi}_1(du)} > c, \tag{8}$$

где $c = c_\varepsilon$ выбирается из условия

$$\int \varphi(\gamma, c)\mathbf{\Pi}_1(d\gamma) = \varepsilon, \quad \varphi(\gamma, c) = \mathbf{P}(r(Y) > c), \quad Y \in \mathbf{\Phi}_{\gamma, I^{-1}}. \tag{9}$$

Это соотношения означают, очевидно, что $\mathbf{E}_{\mathbf{\Pi}_1}\pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y) = \varepsilon$.

Заметим, что $r(y)$ есть аналитическая функция от y. В силу аналитичности эта функция не может принимать постоянное значение на множестве положительной меры Лебега или меры $\mathbf{\Phi}_{\gamma, I^{-1}}$ (в противном случае она была бы постоянной всюду, что возможно лишь при $\mathbf{\Pi}_1 = \mathbf{\Pi}_2$). Значит, $\mathbf{P}(r(Y) = c) = 0$ при любом c и распределение $r(Y)$ непрерывно.

Пусть, как и прежде, $\pi_{Q_1Q_2}(X)$ обозначает байесовский критерий асимптотического уровня $1 - \varepsilon$ в задаче A.

Теорема 1. *Пусть условия* (RR) *выполнены в окрестности точки* θ_1. *Тогда критерий* $\pi(X) = \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(\gamma^*)$, $\gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n}$, *асимптотически эквивалентен критерию* $\pi_{Q_1Q_2}$ *и, стало быть, является асимптотически байесовским.*

Кроме того,

$$\sup_{|\gamma| \leqslant N} |\mathbf{E}_{\theta_1 + \gamma/\sqrt{n}}\pi(X) - \varphi(\gamma, c)| \to 0 \tag{10}$$

при $n \to \infty$, *где* $\varphi(\gamma, c) = \mathbf{E}_\gamma \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y)$ *определено в* (9).

Доказательство. Рассмотрим байесовский критерий $\pi_{Q_1Q_2}$ в задаче A. Он имеет вид

$$T(X) = \frac{\int f_{\theta_1 + u/\sqrt{n}}(X)\mathbf{\Pi}_2(du)}{\int f_{\theta_1 + u/\sqrt{n}}(X)\mathbf{\Pi}_1(du)} > c.$$

Если $X \in \mathbf{P}_{\theta_1}$, то в силу теоремы 36.5

$$T(X) = r(\gamma^*)(1 + \varepsilon(X, \theta_1))$$

($\gamma^* = u^*$ при $\theta = \theta_1$). Так как распределение $r(Y)$ непрерывно, $\gamma^* \Rightarrow Y \in \mathbf{\Phi}_{0, I^{-1}}$ и так как критерий π имеет вид $r(\gamma^*) > c$, то в силу леммы 1 первое утверждение теоремы доказано.

Соотношение (10) следует из представления

$$\mathbf{E}_{\theta_1 + \gamma/\sqrt{n}}\pi(X) = \mathbf{E}_{\theta_1 + \gamma/\sqrt{n}} I_{\{r(\gamma^*) > c\}} \to \mathbf{P}(r(Y) > c),$$

$Y \in \mathbf{\Phi}_{\gamma, I^{-1}}$, и теоремы 37.4. $\triangleleft$

Теорема 2. *Пусть в окрестности точки* θ_1 *выполнены условия* (RR), $\gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n}$.

Пусть, далее, существует минимаксный критерий $\pi_1(Y)$ *уровня* $1 - \varepsilon$ *для проверки* $\mathfrak{h}_1$ *против* $\mathfrak{h}_2$ *в задаче* B *и этот критерий является байесовским*

$$\pi_1(Y) = \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y) \tag{11}$$

при априорных распределениях $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$, *удовлетворяющих условиям*

$$\mathbf{E}_{\mathbf{\Pi}_1}\pi_1(Y) = \sup_{\gamma\in\Gamma_1}\mathbf{E}_\gamma\pi(Y),$$
$$\mathbf{E}_{\mathbf{\Pi}_2}\pi_1(Y) = \inf_{\gamma\in\Gamma_2}\mathbf{E}_\gamma\pi(Y), \quad Y\in\mathbf{\Phi}_{\gamma,I^{-1}} \tag{12}$$

(ср. с условиями теоремы 49.1). *Тогда критерий* $\pi(X) = \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(\gamma^*)$ *будет асимптотически минимаксным в классе* $\widetilde{K}_\varepsilon$ *критериев для проверки* H_1 *против* H_2 *в исходной задаче* A.

Доказательство. Так как π_1 есть критерий уровня $1-\varepsilon$, то

$$\sup_{\gamma\in\Gamma_1}\mathbf{E}_\gamma\pi_1(Y) = \mathbf{E}_{\mathbf{\Pi}_1}\pi(Y) = \varepsilon.$$

В силу (10), (12) отсюда получаем

$$\lim_{n\to\infty}\sup_{\gamma\in\Gamma_1}\mathbf{E}_{\theta_1+\gamma/\sqrt{n}}\pi_{Q_1Q_2}(X) = \lim_{n\to\infty}\mathbf{E}_{Q_1}\pi_{Q_1Q_2}(X) = \varepsilon.$$

Это означает, что $\pi_{Q_1Q_2}\in\widetilde{K}_\varepsilon$, $\pi_{Q_1Q_2}\in\widetilde{K}_\varepsilon^{Q_1}$.

Надо доказать теперь, что для любого критерия $\pi^*\in\widetilde{K}_\varepsilon$

$$\liminf_{n\to\infty}\Big[\inf_{\theta\in\Theta_2}\mathbf{E}_\theta\pi(X) - \inf_{\theta\in\Theta_2}\mathbf{E}_\theta\pi^*(X)\Big]) \geqslant 0.$$

Имеем

$$\limsup_{n\to\infty}\inf_{\theta\in\Theta_2}\mathbf{E}_\theta\pi^*(X) \leqslant \limsup_{n\to\infty}\mathbf{E}_\theta\pi^*(X) \leqslant \limsup_{n\to\infty}\mathbf{E}_{Q_2}\pi_{Q_1Q_2}(X). \tag{13}$$

Последнее неравенство справедливо в силу байесовости $\pi_{Q_1Q_2}$ (т. е. минимальности $q_1\mathbf{E}_{Q_1}\pi_{Q_1Q_2} + (1-q_1)\mathbf{E}_{Q_2}(1-\pi_{Q_1Q_2})$ при соответствующем q_1) и того, что $\limsup\mathbf{E}_{Q_1}\pi^*(X) \leqslant \varepsilon$, $\lim\mathbf{E}_{Q_1}\pi_{Q_1Q_2} = \varepsilon$.

Далее, в силу (10), (12) и теоремы 1 правая часть в (13) равна

$$\lim_{n\to\infty}\mathbf{E}_{Q_2}\pi_1(\gamma^*) = \mathbf{E}_{\mathbf{\Pi}_2}\pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y) =$$
$$= \inf_{\gamma\in\Gamma_2}\mathbf{E}_\gamma\pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y) = \lim_{n\to\infty}\inf_{\gamma\in\Gamma}\mathbf{E}_{\theta_1+\gamma/\sqrt{n}}\pi_{Q_1Q_2}(X). \quad \triangleleft$$

Теорема 3. *Пусть существует р. н. м. к.* $\pi_1(Y)$ *уровня* $1-\varepsilon$ *для проверки* $\mathfrak{h}_1$ *против* $\mathfrak{h}_2$ *в задаче* B. *Предположим, кроме того, что для любого* $\gamma_2\in\Gamma_2$ *существует распределение* $\mathbf{\Pi}_1$ *на* Γ_1 *такое, что*

$$\pi_1(Y) = \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y) \tag{14}$$

есть байесовский критерий для проверки $\mathfrak{h}_{\mathbf{\Pi}_1}$ *против* $\mathfrak{h}_{\mathbf{\Pi}_2}$ (*здесь* $\mathbf{\Pi}_2$ *сосредоточено в точке* γ_2). *Тогда критерий* $\pi(X) = \pi_1(\gamma^*)$ *является а. р. н. м. к.* (*асимптотического уровня* $1-\varepsilon$) *для проверки* H_1 *против* H_2 *в исходной задаче* A.

Отметим, что для задач § 45–47 условие (14) всегда выполнено. Это следует из самого построения р. н. м. к. в этих параграфах.

Доказательство теоремы 3. Принадлежность $\pi_1(\gamma^*) \in \widetilde{K}_\varepsilon$ следует из теоремы 1, так как

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi_1(\gamma^*) = \sup_{\theta\in\Theta_1} \lim_{n\to\infty} \mathbf{E}_\theta \pi_1(\gamma^*) = \sup_{\gamma\in\Gamma_1} \varphi(\gamma, c) \leqslant \varepsilon.$$

Пусть теперь π^* — любой другой критерий из $\widetilde{K}_\varepsilon$. Тогда

$$\limsup_{n\to\infty} \mathbf{E}_{Q_1} \pi^*(X) \leqslant \limsup_{n\to\infty} \sup_{\theta\in\Theta_1} \mathbf{E}_\theta \pi^*(X) \leqslant \varepsilon,$$

и, стало быть, π^* можно рассматривать так же, как критерий из $\widetilde{K}_\varepsilon^{Q_1}$ для проверки H_{Q_1} против H_{Q_2}, где $\mathbf{Q}_1$ индуцировано распределением $\mathbf{\Pi}_1$ (см. формулировку теоремы), а $\mathbf{Q}_2$ сосредоточено в точке $\theta_2 = \theta_1 + \gamma_2 n^{-1/2}$. Если $\pi_{Q_1Q_2}$ — байесовский критерий асимптотического уровня $1 - \varepsilon$ для этих распределений, то

$$\lim_{n\to\infty} \mathbf{E}_{\theta_2} \pi_{Q_1Q_2}(X) \geqslant \limsup_{n\to\infty} \mathbf{E}_{\theta_2} \pi^*(X).$$

Но левая часть этого неравенства совпадает в силу теоремы 1 со значением

$$\lim_{n\to\infty} \mathbf{E}_{\theta_2} \pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(\gamma^*) = \lim_{n\to\infty} \mathbf{E}_{\theta_2} \pi_1(\gamma^*). \quad \triangleleft$$

Аналогичным образом можно искать асимптотически р. н. м. к. в классе асимптотически несмещенных критериев.

Замечание 1. Если распределения $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$ сосредоточены соответственно в точках γ_1 и γ_2, то

$$r(Y) = \frac{\exp\left\{-\frac{1}{2}(Y-\gamma_2)I(Y-\gamma_2)^\top\right\}}{\exp\left\{-\frac{1}{2}(Y-\gamma_1)I(Y-\gamma_1)^\top\right\}}.$$

Стало быть, критическая область $\pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}(Y)$ будет иметь вид

$$YI(\gamma_2-\gamma_1)^\top = (YI, (\gamma_2-\gamma_1)) > c.$$

В одномерном случае мы получаем отсюда а. н. м. к. (43.21), изучавшийся нами в § 43.

Замечание 2. Если распределение $\mathbf{\Pi}_1$ сосредоточено в точке $u = 0$, а распределение Π_2 является равномерным в шаре $|u| \leqslant N$, то знаменатель функции $r(Y)$ будет равен $\exp\left\{-\frac{1}{2}YIY^\top\right\}$, а числитель при больших N и $|\gamma| < N - \sqrt{N}$ будет близок к $\sqrt{|I|}\,(2\pi)^{k/2}$. Стало быть, критическая область для $\pi_{\mathbf{\Pi}_1\mathbf{\Pi}_2}$ при таких $\mathbf{\Pi}_1$, $\mathbf{\Pi}_2$ будет близка к внешности эллипсоида

$$YIY^\top > c,$$

а критическая область асимптотически байесовского критерия $\pi_{\Pi_1\Pi_2}(\gamma^*)$ будет близка к

$$\gamma^* I \gamma^{*\top} > c.$$

Это есть не что иное, как асимптотическая форма к. о. п., изучавшегося в предыдущем параграфе (ср. с теоремой 43.2).

Замечание 3. В теоремах 2, 3 присутствуют условия о том, что минимаксный критерий (теорема 2) или р. н. м. к. (теорема 3) для задачи В являются байесовскими при некоторых распределениях Π_i на Γ_i. Мы увидим в главе 6, что эти условия излишни — класс всех байесовских критериев содержит в себе все «неулучшаемые» критерии, в том числе р. н. м. к. и минимаксные.

§ 55. Свойства асимптотической оптимальности критерия отношения правдоподобия, вытекающие из предельного признака оптимальности

В этом параграфе мы рассмотрим некоторые следствия результатов § 54, связанные с критерием отношения правдоподобия. Мы установим, в частности, асимптотические равномерную наибольшую мощность и минимаксность к. о. п. для некоторых важных конкретных задач, связанных с проверкой близких гипотез.

Везде в дальнейшем предполагается, что в окрестности точки θ_1 выполнены условия (RR). Для сокращения выкладок нам будет удобно, как и в предыдущем параграфе, считать, где это потребуется, что множества Γ_i ограничены.

1. А. р. н. м. к. для близких гипотез с односторонними альтернативами. Пусть параметр θ одномерный и проверяется односторонняя гипотеза $H_1 = \{\theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$ против гипотезы $H_2 = \{\theta > \theta_2 = \theta_1 + \gamma_2 n^{-1/2}\}$, $\gamma_1 \leqslant \gamma_2$.

Теорема 1. *Критерий отношения правдоподобия $\widehat{\pi}(X)$ с критической областью*

$$R(X) \equiv \frac{\sup\limits_{\theta\in\Theta_2} f_\theta(X)}{\sup\limits_{\theta\in\Theta_1} f_\theta(X)} > c \tag{1}$$

при $\Theta_1 = \{\theta : \theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$, $\Theta_2 = \{\theta : \theta > \theta_1 + \gamma_2 n^{-1/2}\}$ и подходящем с асимптотически эквивалентен критерию

$$\gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n} > c_\varepsilon = \lambda_\varepsilon I^{-1/2} + \gamma_1, \quad \Phi_{0,1}(\lambda_\varepsilon) = 1 - \varepsilon \tag{2}$$

и является а. р. н. м. к. асимптотического уровня $1 - \varepsilon$ для проверки $H_1 = \{\theta \leqslant \theta_1 + \gamma_1 n^{-1/2}\}$ против $H_2 = \{\theta > \theta_1 + \gamma_2 n^{-1/2}\}$. В формулах (2) I означает информацию Фишера $I(\theta_1)$ в точке θ_1 для семейства f_θ.

Доказательство. Из § 45 следует, что для выборки $Y \in \Phi_{\gamma, I^{-1}}$ единичного объема из нормальной совокупности с известной дисперсией I^{-1} существует р. н. м. к. для проверки гипотезы $\mathfrak{h}_1 = \{\gamma \leqslant \gamma_1\}$ против $\mathfrak{h}_2 = \{\gamma > \gamma_2\}$ вида $Y > c_\varepsilon$, где c_ε определено в (2). Так же будет выглядеть,

очевидно, байесовский критерий для вырожденных распределений, сосредоточенных в точках γ_1 и γ_2 (или в точках γ_1 и $\gamma > \gamma_1$, если $\gamma_1 = \gamma_2$). На основании этого из теоремы 54.3 следует, что асимптотически р. н. м. к. асимптотического уровня $1 - \varepsilon$ для проверки H_1 против H_2 существует и имеет вид (2).

Нам осталось доказать, что критерии (1) и (2) асимптотически эквивалентны. По теореме 36.4, полагая $Z_1(t) = \dfrac{f_{\theta_1+t}(X)}{f_{\theta_1}(X)}$, будем иметь при $X \in \mathbf{P}_{\theta_1}$

$$R(X) = \frac{\sup\limits_{u>\gamma_2} Z_1(un^{-1/2})}{\sup\limits_{u\leqslant\gamma_1} Z_1(un^{-1/2})} = \frac{\sup\limits_{u>\gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X)\right\}}{\sup\limits_{u\leqslant\gamma_1} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X)\right\}} = T_n(X) + \varepsilon_n^{(3)}(X),$$

где $\varepsilon_n^{(i)}(X) \xrightarrow[P_{\theta_1}]{} 0$, $i = 1, 2, 3$,

$$T_n(X) = r(\gamma^*) \equiv \frac{\sup\limits_{u>\gamma_2} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}}{\sup\limits_{u\leqslant\gamma_1} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}} = \begin{cases} \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{при} \quad \gamma^* \leqslant \gamma_1, \\ \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I + \frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{при} \quad \gamma_1 < \gamma^* < \gamma_2, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{при} \quad \gamma^* \geqslant \gamma_2. \end{cases}$$

Это есть непрерывная, монотонно возрастающая функция от γ^*. Стало быть, неравенство $T_n(X) > c$ эквивалентно неравенству $\gamma^* > c'$ при некотором c'. Так как, кроме того, $\gamma^* \Rightarrow Y \in \mathbf{\Phi}_{0,I^{-1}}$, то распределение $r(Y)$ абсолютно непрерывно. Условия леммы 54.1 для критериев (1), (2) выполнены. $\triangleleft$

2. А. р. н. м. к. для двусторонних альтернатив. Пусть параметр θ по-прежнему одномерный, а задача А состоит в проверке гипотезы $H_1 = \{(\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$ против $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$, $\gamma_2 > \gamma_1$. Положим

$$\overline{\gamma} = \frac{\gamma_1 + \gamma_2}{2}, \quad \Delta = \frac{\gamma_2 - \gamma_1}{2}.$$

Теорема 2. *Критерий отношения правдоподобия* $\widehat{\pi}(X)$, *определенный в* (1) *при соответствующем* c *и* $\Theta_1 = \{\theta : (\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$, $\Theta_2 = \{\theta : (\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$, *как и критерий*

$$|\gamma^* - \overline{\gamma}| = |(\widehat{\theta}^* - \theta_1)\sqrt{n} - \overline{\gamma}| < c_\varepsilon, \tag{3}$$

где c_ε *определяется из уравнения* $\Phi_{0,I^{-1}}(-c - \Delta, c - \Delta) = \varepsilon$, *являются а. р. н. м. к. асимптотического уровня* $1 - \varepsilon$ *для проверки* $H_1 = \{(\theta - \theta_1)\sqrt{n} \notin (\gamma_1, \gamma_2)\}$ *против* $H_2 = \{(\theta - \theta_1)\sqrt{n} \in (\gamma_1, \gamma_2)\}$.

Доказательство этой теоремы вполне аналогично доказательству предыдущей. Из § 45 следует, что для задачи B, состоящей в проверке по наблюдению $Y \in \Phi_{\gamma,I^{-1}}$ гипотезы $\mathfrak{h}_1 = \{\gamma \notin (\gamma_1, \gamma_2)\}$ против $\mathfrak{h}_2 = \{\gamma \in (\gamma_1, \gamma_2)\}$, существует р. н. м. к. вида $c' < Y < c''$, где c', c'' подбираются так, чтобы

$$\Phi_{\gamma_1,I^{-1}}((c', c'')) = \Phi_{\gamma_2,I^{-1}}((c', c'')) = \varepsilon.$$

Нетрудно видеть, что мы удовлетворим этим соотношениям, если положим $c' = \overline{\gamma} - c_\varepsilon$, $c'' = \overline{\gamma} + c_\varepsilon$, так как

$$\Phi_{\gamma_1,I^{-1}}((\overline{\gamma} - c_\varepsilon, \overline{\gamma} + c_\varepsilon)) = \Phi_{0,I^{-1}}((-c_\varepsilon + \Delta, c_\varepsilon + \Delta)) = \varepsilon,$$
$$\Phi_{\gamma_2,I^{-1}}((\overline{\gamma} - c_\varepsilon, \overline{\gamma} + c_\varepsilon)) = \Phi_{0,I^{-1}}((-c_\varepsilon - \Delta, c_\varepsilon - \Delta)) = \varepsilon.$$

Кроме того, мы видели в § 45, что для любого $\gamma_0 \in (\gamma_1, \gamma_2)$ существует $q \in (0, 1)$ такое, что байесовский критерий $\pi_{\Pi_1\Pi_2}$ для проверки гипотезы $\mathfrak{h}_{\Pi_1}$ для распределения Π_1: $\Pi_1(\{\gamma_1\}) = q$, $\Pi_1(\{\gamma_2\}) = 1 - q$, против гипотезы $\mathfrak{h}_{\Pi_2} = \{\gamma = \gamma_0\}$ будет иметь вид

$$c' < Y < c''.$$

Это означает, что условия теоремы 54.3 будут выполнены и критерий (3) будет а. р. н. м. к. для проверки H_1 против H_2.

Рассмотрим теперь к. о. п. (1) для областей Θ_i, определенных в теореме, и покажем, что он асимптотически эквивалентен (3). Как и в доказательстве теоремы 1, из теоремы 36.4 получаем, что при $X \in \mathbf{P}_{\theta_1}$

$$\frac{\sup\limits_{u \in (\gamma_1,\gamma_2)} Z_1(un^{-1/2})}{\sup\limits_{u \notin (\gamma_1,\gamma_2)} Z_1(un^{-1/2})} = \frac{\sup\limits_{u \in (\gamma_1,\gamma_2)} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(1)}(X)\right\}}{\sup\limits_{u \notin (\gamma_1,\gamma_2)} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I + \varepsilon_n^{(2)}(X)\right\}} =$$
$$= T_n(X) + \varepsilon_n^{(3)}(X),$$

где $\varepsilon_n^{(i)}(X) \xrightarrow[P_{\theta_1}]{} 0$, $i = 1, 2, 3$,

$$T(X) = r(\gamma^*) = \frac{\sup\limits_{u \in (\delta_1, \gamma_2)} \exp\left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}}{\sup\limits_{u \notin (\gamma_1, \gamma_2)} \left\{-\frac{1}{2}(\gamma^* - u)^2 I\right\}} =$$

$$= \begin{cases} \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{при} \quad \gamma^* \leqslant \gamma_1, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_1)^2 I\right\} & \text{при} \quad \gamma_1 < \gamma^* \leqslant \overline{\gamma}_1, \\ \exp\left\{\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{при} \quad \overline{\gamma} < \gamma^* \leqslant \gamma_2, \\ \exp\left\{-\frac{1}{2}(\gamma^* - \gamma_2)^2 I\right\} & \text{при} \quad \gamma_2 < \gamma^*. \end{cases}$$

Из этих равенств видно, что $r(\gamma^*)$ есть непрерывная монотонно убывающая функция от $|\gamma^* - \overline{\gamma}|$ (она симметрична относительно точки $\gamma^* = \overline{\gamma}$). Поэтому неравенство $r(\gamma^*) > c$ эквивалентно неравенству $|\gamma^* - \overline{\gamma}| < c'$. Так как $\gamma^* \Rightarrow$ $\Rightarrow Y \in \mathbf{\Phi}_{0, I^{-1}}$, то условия леммы 54.1 выполнены. $\triangleleft$

3. Асимптотически минимаксный критерий для близких гипотез, касающихся многомерного параметра. Рассмотрим теперь многомерный параметр θ. В этом случае а. р. н. м. к. для проверки гипотезы $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$, как правило, не существует, и мы рассмотрим задачу построения асимптотически минимаксных критериев.

Сначала сделаем одно общее замечание, которое упростит дальнейшие рассуждения. Оно состоит в том, что рассматриваемую задачу проверки гипотез всегда можно «перепараметризировать» (т. е. ввести новый параметр) так, чтобы информационная матрица $I = I(\theta_1)$ в точке θ_1 стала единичной. Для этого достаточно сделать линейное преобразование и ввести новый параметр β равенством

$$\theta = \beta I^{-1/2}.$$

Тогда информационная матрица Фишера $J(\beta)$ для параметрического семейства $\mathbf{P}_{\beta I^{-1/2}}$ будет равна в точке $\beta_1 = \theta_1 I^{1/2}$

$$J(\beta_1) = I^{-1/2} I I^{-1/2} = E.$$

В этом пункте иногда нам будет проще иметь дело с параметром β. Вернуться к исходному параметру всегда можно с помощью обратного линейного преобразования.

Итак, пусть $I = I(\theta_1) = E$. Рассмотрим задачу А, состоящую в проверке по выборке $X \in \mathbf{P}_\theta$ гипотезы

$$H_1 = \{|\theta - \theta_1| \leqslant a n^{-1/2}\} \quad \text{против} \quad H_2 = \{|\theta - \theta_1| \geqslant b n^{-1/2}\}, \ a < b. \qquad (4)$$

Теорема 3. *Критерий отношения правдоподобия* $\widehat{\pi}$, *определенный в* (1) *при соответствующем* c *и* $\Theta_1 = \{\theta : |\theta - \theta_1| \leqslant an^{-1/2}\}$, $\Theta_2 = \{\theta : |\theta - \theta_1| \geqslant bn^{-1/2}\}$, *асимптотически эквивалентен при любых* $0 \leqslant a < b < \infty$ *критериям*

$$\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > c, \tag{5}$$

$$|\gamma^*| \equiv |(\widehat{\theta}^* - \theta_1)\sqrt{n}\,| > c_\varepsilon, \tag{6}$$

где c_ε^2 *есть решение относительно* c *уравнения*

$$p_c(a) \equiv \mathbf{P}\big((\xi_1 + a)^2 + \xi_2^2 + \ldots + \xi_k^2 > c^2\big) = \varepsilon, \tag{7}$$

и является асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ *для проверки гипотез* H_1 *и* H_2, *определенных в* (4). *Случайные величины* ξ_i *в* (7) *независимы,* $\xi_i \in \mathbf{\Phi}_{0,1}$. *Гарантированная предельная мощность критериев* $\widehat{\pi}$, (5), (6) *равна* $p_{c_\varepsilon}(b)$.

Доказательство. Задача B здесь будет состоять в проверке по наблюдению $Y \in \mathbf{\Phi}_{\gamma,E}$ гипотезы $\mathfrak{h}_1 = \{|\gamma| \leqslant a\}$ против $\mathfrak{h}_2 = \{|\gamma| \geqslant b\}$. В примере 49.1 мы видели, что в этой задаче существует минимаксный критерий уровня $1 - \varepsilon$, который имеет вид

$$|Y| > c_\varepsilon.$$

При построении этого критерия мы пользовались теоремой 49.1. Это означает, что условия теоремы 54.2 выполнены. Следовательно, критерий

$$|\gamma^*| > c_\varepsilon$$

будет асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ для задачи A.

Критерий отношения правдоподобия (1) здесь будет иметь вид

$$R(X) = \frac{\sup\limits_{|u| \geqslant b} Z_1(un^{-1/2})}{\sup\limits_{|u| \leqslant a} Z_1(un^{-1/2})} > c. \tag{8}$$

Следуя в точности рассуждениям в доказательствах теорем 1, 2, мы получим, что $R(X) = T_n(X) + \varepsilon_n(X)$, $\varepsilon_n(X) \underset{P_{\theta_1}}{\longrightarrow} 0$, где

$$T_n(X) = r(\gamma^*) = \frac{\sup\limits_{|u| \geqslant b} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}{\sup\limits_{|u| \leqslant a} \exp\left\{-\frac{1}{2}|\gamma^* - u|^2\right\}}.$$

Как и прежде, отсюда следует абсолютная непрерывность распределения $r(Y)$ и асимптотическая эквивалентность критериев $R(X) > c$ и $T_n(X) > c$. Последний эквивалентен критерию

$$|\gamma^*| > c',$$

который при $c' = c_\varepsilon$ будет критерием уровня $1 - \varepsilon$. По теореме 54.1 (см. (54.10)) он будет иметь гарантированную предельную мощность, равную $p_{c_\varepsilon}(b)$ (см. теорему 49.2). $\triangleleft$

Замечание 1. Если вернуться к первоначальному параметру (до параметризации, переводящей $I(\theta_1)$ в единичную матрицу), то получим, что утверждение теоремы будет справедливо относительно гипотез $H_i = \{\theta \in \in \Theta_i\}$, где (ср. с примером 49.2 при $\sigma^2 = I^{-1}$)

$$\Theta_1 = \{\theta : (\theta - \theta_1) I(\theta_1) (\theta - \theta_1)^\top \leqslant a^2 n^{-1}\},$$

$$\Theta_2 = \{\theta : (\theta - \theta_1) I(\theta_1) (\theta - \theta_1)^\top \geqslant b^2 n^{-1}\}.$$

Критерий (6) примет форму

$$(\widehat{\theta}^* - \theta_1) I(\theta_1) (\widehat{\theta}^* - \theta_1)^\top n > c_\varepsilon^2$$

или (см. теорему 53.2)

$$L'(X, \theta_1) I^{-1}(\theta_1) (L'(X, \theta_1))^\top > c_\varepsilon^2. \tag{9}$$

Критерий отношения правдоподобия, очевидно, не изменится, так как величина максимума $f_\theta(X)$ в области Θ_i не зависит от замены переменных (при соответствующем преобразовании областей Θ_i).

Отметим также, что форма критерия (9) иногда более удобна, чем (5), (6), так как не связана с вычислениями $\widehat{\theta}^*$. Аналогичные замены можно сделать для критериев (2), (3) в теоремах 1, 2. Мы предоставляем это читателю.

Замечание 2. Совершенно аналогично теореме 3 можно построить асимптотически минимаксный критерий для таких задач A, которые могут быть редуцированы к задаче B, рассмотренной в примере 49.5.

Замечание 3. В § 53 мы построили а. б. к. для проверки гипотезы $\{\theta = \theta_1\}$ против $\{\theta \neq \theta_1\}$, который имеет вид к. о. п.

$$\frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > c.$$

Таким образом, этот критерий, являясь а. б. к., обладает также свойством асимптотической минимаксности при проверке гипотезы $\{\theta = \theta_1\}$ против $\{(\theta - \theta_1) I(\theta_1) (\theta - \theta_1)^\top \geqslant b^2 n^{-1}\}$ при любом $b > 0$.

4. Асимптотически минимаксный критерий о принадлежности выборки параметрическому подсемейству. Рассмотрим теперь к. о. п. в несколько более сложной задаче о проверке гипотезы $H_1 = \{\theta \in \Theta_1\}$ против $H_2 = \{\theta \in \Theta_2\}$, когда размерность l множества Θ_1 положительна, но меньше $k > 1$. Именно, пусть дана гладкая функция $\theta = g(\alpha)$ от l-мерного ($l < k$) параметра $\alpha \in A_1 \subset \mathbb{R}^l$. Образ множества A_1 в Θ, порожденный отображением g, обозначим через Θ_1. Задача состоит в проверке гипотезы $H_1 = \{\theta \in \Theta_1\}$ о том, что параметр θ принадлежит «кривой» Θ_1 (или о том, что $X \in \mathbf{P}_{g(\alpha)}$ при некотором $\alpha \in A_1$), против дополнительной альтернативы $\{X \in \mathbf{P}_\theta;\ \theta \notin \Theta_1\}$, так что в этом случае $\Theta_2 = \Theta \setminus \Theta_1$. Другими словами,

это есть задача о проверке принадлежности выборки X параметрическому подсемейству распределений $\{\mathbf{P}_{g(\alpha)}; \alpha \in A_1\}$.

К этому классу задач относятся, например, уже известные нам задачи о проверке гипотезы $\{X \in \mathbf{\Phi}_{\alpha_0,\sigma^2}\}$ против $\{X \in \mathbf{\Phi}_{\alpha,\sigma^2};\ \alpha \neq \alpha_0\}$ при данном α_0 и неизвестном σ^2 или о проверке гипотезы $\{X \in \mathbf{\Phi}_{\alpha,\sigma_0^2}\}$ против $\{X \in \mathbf{\Phi}_{\alpha,\sigma^2};\ \sigma \neq \sigma_0\}$ при данном σ_0 и неизвестном α и другие.

Относительно кривой $\theta = g(\alpha)$ в Θ мы будем предполагать, что она является дважды непрерывно дифференцируемой и что матрица $G = \|\partial g_i(\alpha)/\partial \alpha_j\|$ $(i = 1,\ldots,k;\ j = 1,\ldots,l;\ g_i(\alpha),\ \alpha_i$ суть координаты $g(\alpha)$ и α соответственно) имеет ранг l. Это значит, что мы можем произвести дифференцируемую взаимно однозначную замену параметра (перепараметризацию задачи) так, чтобы первые l координат (можно положить их, не ограничивая общности, равными $\alpha = (\alpha_1,\ldots,\alpha_l)$) определяли положение точки θ на кривой Θ_1, а остальные (обозначим их $\beta = (\beta_1,\ldots,\beta_{k-l})$) определяли положение θ в «плоскости» (подпространстве), скажем, ортогональной (это не обязательно) к «кривой» $g(\alpha)$ в точке α. Тогда задача сводится к проверке гипотезы $\{\beta = 0\}$ против $\{\beta \neq 0\}$ при наличии неизвестного «мешающего» подпараметра α.

При этом будем рассматривать близкие гипотезы, положив $\beta = \gamma'' n^{-1/2}$, и проверять гипотезу $\{\gamma'' = 0\}$ против $\{\gamma'' \neq 0\}$ или против

$$\{\gamma'' M_2(\alpha) \gamma''^{\top} \geqslant b^2\} \tag{10}$$

при $b > 0$ и некоторой положительно определенной матрице $M_2(\alpha)$.

В исходных координатах последняя задача будет соответствовать проверке гипотезы $H_1 = \{\theta \in \Theta_1\}$ против близких альтернатив, когда параметр θ располагается в $n^{-1/2}$-окрестности кривой Θ_1 и находится вне некоторой «трубки», содержащей Θ_1 и соответствующей множеству (10). Возможен и другой вариант постановки задачи о проверке близких гипотез, исходящий из того, что параметр θ «локализован», и мы знаем, что он располагается в окрестности некоторой точки $\theta_0 = g(\alpha^0)$, $\alpha^0 \in A_1$. Тогда новый параметр $\tau = (\beta, \alpha - \alpha^0)$ будет локализован около точки $\tau_0 = (0,0)$. Положим $\alpha - \alpha^0 = \gamma' n^{-1/2}$, $\beta = \gamma'' n^{-1/2}$ и будем проверять гипотезу $\{\gamma'' = 0\}$ против $\{\gamma'' \neq 0\}$ или против $\{\gamma'' M_2(\alpha^0) \gamma''^{\top} \geqslant b^2\}$ при наличии мешающего параметра γ'.

Интересующие нас результаты в этих двух постановках задач практически совпадают. Но нам удобнее будет иметь дело со второй постановкой, поскольку в этом случае мы располагаем всеми необходимыми предварительными результатами. Предположение о локализации параметра θ носит условный характер, и форма утверждений, полученных ниже, от θ_0 зависеть не будет.

Итак, будем считать, что новый параметр $\tau = (\alpha - \alpha^0, \beta)$ имеет вид

$$\tau = \gamma n^{-1/2}, \quad \gamma = (\gamma', \gamma''),$$

и будем проверять гипотезу $H_1 = \{\gamma'' = 0\}$ против $H_2 = \{\gamma'' M_2 \gamma''^{\top} \geqslant b^2\}$, где в качестве $M_2 = M_2(\alpha^0)$ мы возьмем информационную матрицу Фишера для

параметрического семейства $\{\mathbf{P}_{\theta(0,\beta)}\}$ в точке $\beta = 0$, где $\theta(\tau) = \theta((\alpha-\alpha^0),\beta))$ есть функция, восстанавливающая θ по значению $\tau = (\tau', \tau'')$.

Теорема 4. *Пусть $\theta_0 = g(\alpha^0)$ является внутренней точкой Θ и в окрестности этой точки выполнены условия* (RR). *Пусть, далее, функция $g(\alpha)$ дважды непрерывно дифференцируема в точке α^0, а матрица $G = \|\partial g_i(\alpha)/\partial\alpha_j\|_{\alpha=\alpha^0}$ имеет ранг l. Тогда к. о. п.* (1) *при определенных выше Θ_1, Θ_2 и при соответствующем с асимптотически эквивалентен критериям*

$$R_1(X) \equiv \frac{f_{\widehat{\theta}^*}(X)}{f_{g(\widehat{\alpha}^*)}(X)} > e^{h_\varepsilon/2}, \tag{11}$$

$$\begin{gathered}(\widehat{\theta}^* - g(\widehat{\alpha}^*))I(g(\widehat{\alpha}^*))(\widehat{\theta}^* - g(\widehat{\alpha}^*))^\top > h_\varepsilon n^{-1},\\ (\widehat{\theta}^* - g(\widehat{\alpha}^*))I(\widehat{\theta}^*)(\widehat{\theta}^* - g(\hat{\alpha}^*))^\top > h_\varepsilon n^{-1}\end{gathered} \tag{12}$$

и является асимптотически минимаксным критерием асимптотического уровня $1-\varepsilon$ для проверки гипотезы $H_1 = \{\theta \in \Theta_1\} = \{\gamma'' = 0\}$ против $H_2 = \{\gamma'' M_2 \gamma''^\top \geqslant b^2\}$.

Распределение статистики $2\ln R_1(X)$ при $X \Subset \mathbf{P}_{g(\alpha^0)}$ (т. е. при гипотезе H_1) сходится при $n \to \infty$ к распределению χ^2 с $k-l$ степенями свободы (и, стало быть, от f_θ и α^0 не зависит). В соответствии с этим h_ε в (11), (12) *означает квантиль порядка $1-\varepsilon$ распределения $\mathbf{H}_{k-l}$.*

Гарантированная асимптотическая мощность к. о. п. равна

$$\mathbf{P}\big((\xi_1 + b)^2 + \xi_2^2 + \ldots + \xi_{k-l}^2 > h_\varepsilon\big),$$

где $\xi_i \Subset \mathbf{\Phi}_{0,1}$ и независимы.

Мы видим, что асимптотически минимаксные критерии (11), (12) со значениями α^0 никак не связаны.

Замечание 4. Гипотезу H_2 в терминах исходного параметра θ можно записать в виде

$$H_2 = \{\inf_{\gamma'}(\theta - g(\alpha^0 + \gamma' n^{-1/2}))I(g(\alpha^0))(\theta - g(\alpha^0 + \gamma' n^{-1/2}))^\top \geqslant b^2 n^{-1}\}.$$

Напомним, что мы считаем множества Γ_i ограниченными, так что здесь $(\theta - \theta_0) \leqslant N n^{-1/2}$, $|\gamma'| \leqslant N$ при некотором $N > 0$.

Замечание 5. Как будет видно из доказательства, утверждение теоремы полностью сохранится, если гипотезу $H_1 = \{\gamma'' = 0\}$ заменить на $H_1 = \{\gamma'' M_2 \gamma''^\top \leqslant a^2\}$, $a < b$, при соответствующей замене множества Θ_1.

Доказательство теоремы 4. Здесь в качестве «основного» мы будем рассматривать критерий (11), эквивалентный (1) и более удобный по своей форме. Установим его асимптотическую эквивалентность асимптотически минимаксному критерию, а затем его асимптотическую эквивалентность (12).

Рассмотрим распределения $\mathbf{P}_\theta$ и $\mathbf{P}_{g(\alpha)}$ как зависящие от параметров $\tau = \{\tau', \tau''\}$ и $\alpha = \tau' + \alpha^0$ соответственно. Положим $\tau = \gamma n^{-1/2}$, $\gamma = (\gamma', \gamma'')$, так что $\tau' = \gamma' n^{-1/2}$, $\tau'' = \gamma'' n^{-1/2}$, и будем проверять гипотезу $H_1 = \{\gamma'' = 0\}$ против

$H_2 = \{\gamma'' M_2 \gamma''^{\top} \geqslant b^2\}$, где M_2 — информационная матрица Фишера для семейства $\mathbf{P}_{g(\alpha)}$ в точке α^0. Сделаем теперь еще одно преобразование над параметром, подобное тому, которое проводилось нами в примере 49.4 и которое превращает информационные матрицы в единичные. Именно, положим $\rho = \tau\Lambda$ и соответственно $\eta = \gamma\Lambda$ ($\rho = \eta n^{-1/2}$), где Λ — треугольная матрица, подробно описанная в примере 49.4 и обладающая свойствами

$$J^{-1} = \Lambda^{\top} M^{-1} \Lambda = E, \quad J_2^{-1} = \Lambda_2^{\top} M_2^{-1} \Lambda_2 = E,$$

где J, M, J_2, M_2 есть информационные матрицы в точке θ_0 соответственно для ρ, τ, ρ'', τ'' (верхние штрихи у обозначений имеют тот же смысл, что у τ', τ'', γ', γ''), Λ_2 — матрица порядка $(k-l)\times(k-l)$, образованная последними $k-l$ строками и столбцами матрицы Λ, так что $\rho'' = \tau''\Lambda_2$, $\eta'' = \gamma''\Lambda_2$.

В новых параметрах гипотезы H_1 и H_2 запишутся в виде

$$H_1 = \{\eta'' = 0\}, \quad H_2 = \{|\eta''| \geqslant b\}.$$

Из свойств произведенных преобразований ясно, что $\theta = \theta(\rho)$ есть взаимно однозначная функция от ρ и что все рассматриваемые нами параметрические семейства (в том числе с параметрами ρ, ρ'') удовлетворяют условиям (RR). Положим $\rho_0 = \theta^{-1}(\theta_0)$ (это есть решение уравнения $\theta(\rho) = \theta_0$),

$$Z_0(t) = \frac{f_{\theta(\rho_0+t)}(X)}{f_{\theta_0}(X)}, \quad Y_0(u) = \ln Z_0(un^{-1/2}).$$

Воспользуемся теоремой 37.3. Мы получим при $|u| \leqslant \eta_n\sqrt{n}$, $\rho = \rho_0 + \eta n^{-1/2}$, $X \in \mathbf{P}_{\theta(\rho)}$, что

$$Y_0(u) = (\xi_n + \eta, u) - \frac{1}{2}(u,u) + (|u|^2 + |\eta|^2)\varepsilon_n(X,u,\eta), \tag{13}$$

где $|\varepsilon_n(X,u,\eta)| \leqslant \varepsilon_n(X) \underset{P_{\theta(\rho)}}{\longrightarrow} 0$ равномерно по η при $|\eta| \leqslant \eta_n\sqrt{n}$, η_n — произвольная последовательность, сходящаяся к нулю. В этих равенствах мы использовали тот факт, что информационная матрица для параметра ρ является единичной. Вектор ξ_n есть вектор производных функции $n^{-1/2}L(X,\theta(\rho))$ по ρ_j в точке $\rho = \rho_0 + \eta n^{-1/2}$, так что $\xi_n \Subset \Phi_{0,E}$ равномерно по ρ (по η) при $|\eta| \leqslant \eta_n\sqrt{n}$. (Ввиду предположения об ограниченности $(\theta - \theta_0)\sqrt{n}$ равномерность сходимости здесь и в дальнейшем достаточно устанавливать для $|\eta| \leqslant N$ при произвольном фиксированном N. Однако ничто не мешает нам устанавливать требуемую равномерность и в более широкой области $|\eta| \leqslant \eta_n\sqrt{n} \to \infty$.)

Положим теперь в (13) $u = (u', u'')$, $u'' = 0$. Мы сможем записать тогда при прежнем соглашении относительно обозначений со штрихами

$$Y_0((u',0)) = (\xi_n' + \eta', u') - \frac{1}{2}(u',u') + (|u'|^2 + |\eta|^2)\varepsilon_n(X,u',\eta). \tag{14}$$

Из (13), (14) следует, что максимальные значения $Y_0(u)$ и $Y_0((u',0))$ достигаются соответственно при

$$u = (\xi_n + \eta)(E + \varepsilon_n(X,\eta)), \quad u' = (\xi_n' + \eta')(E + \varepsilon_n^{(1)}(X,\eta)), \tag{15}$$

где $\varepsilon_n(X,\eta) \underset{P_{\theta(\rho)}}{\longrightarrow} 0$, $\varepsilon_n^{(1)}(X,\eta) \underset{P_{\theta(\rho)}}{\longrightarrow} 0$ равномерно по η, $|\eta| < \eta_n\sqrt{n}/2$. Надо лишь заметить, что вероятность больших значений $|\xi_n + \eta|$ равномерно мала, так как $\xi_n + \eta \Subset \mathbf{\Phi}_{\eta,E}$ равномерно по η, $|\eta| < \eta_n\sqrt{n}$, и $\mathbf{P}_\theta(|\xi_n + \eta| > \eta_n\sqrt{n}) \to 0$ равномерно по η, $|\eta| < \eta_n\sqrt{n}/2$.

Обратимся теперь к к. о. п. Имеем при $\theta = \theta(\rho)$, $X \in \mathbf{P}_\theta$, $\rho = \rho_0 + \eta n^{-1/2}$,

$$R_1(X) = \frac{\sup\limits_{\theta} f_\theta(X)}{\sup\limits_{\alpha} f_{g(\alpha)}(X)} = \frac{\sup\limits_{u} e^{Y_0(u)}}{\sup\limits_{u'} e^{Y_0((u',0))}} =$$

$$= \frac{\exp\left\{\frac{1}{2}|\xi_n + \eta|^2 + \widetilde{\varepsilon}_n(X,\eta)\right\}}{\exp\left\{\frac{1}{2}|\xi_n' + \eta'|^2 + \widetilde{\varepsilon}_n^{(1)}(X,\eta)\right\}} = \exp\left\{\frac{1}{2}|\xi_n'' + \eta''|^2 + \varepsilon_n''(X,\eta)\right\}, \quad (16)$$

где функции ε_n с различными индексами сходятся к нулю по $\mathbf{P}_\theta$-вероятности равномерно при $|\eta| < \eta_n\sqrt{n}$;

$$2\ln R_1(X) \Rightarrow |Y'' + \eta''|^2, \quad Y \in \mathbf{\Phi}_{0,E}, \quad (17)$$

равномерно по η.

Так как при $\theta = g(\alpha)$ с необходимостью $\eta'' = 0$, то отсюда следует утверждение теоремы относительно статистики $2\ln R_1(X)$.

Вспомним теперь, что (см. теорему 36.4) $\xi_n = u^*(E + \varepsilon_n(X,\eta))$, где $u^* = (\widehat{\rho}^* - \rho)\sqrt{n}$, $\widehat{\rho}^*$ — о. м. п. для параметра ρ. Отсюда и из равенства $\rho_0'' = 0$, полагая $\eta^* = (\widehat{\rho}^* - \rho_0)\sqrt{n}$, получим

$$\xi_n + \eta = \sqrt{n}\,(\widehat{\rho}^* - \rho + \rho - \rho_0) + u^*\varepsilon_n(X,\eta) = \sqrt{n}\,(\widehat{\rho}^* - \rho_0) + u^*\varepsilon_n(X,\eta) =$$
$$= \eta^* + u^*\varepsilon_n(X,\eta) \Subset \mathbf{\Phi}_{\eta,E},$$

$$\xi_n'' + \eta'' = (\eta^*)'' + (u^*\varepsilon_n(X,\eta))''.$$

Стало быть, правая часть в (16) может быть записана также в виде $\exp\left\{\frac{1}{2}|(\eta^*)''|^2 + \right.$ $\left. + \varepsilon_n'''(X,\eta)\right\}$, $\varepsilon_n'''(X,\eta) \underset{P_\theta}{\longrightarrow} 0$. Это означает, что критерий

$$|(\eta^*)''|^2 > h_\varepsilon \quad (18)$$

и к. о. п. асимптотически эквивалентны, т. е.

$$\lim_{n\to\infty} \sup_{\alpha} \mathbf{P}_{g(\alpha)}(R_1(X) > e^{h_\varepsilon/2}) = \lim_{n\to\infty} \sup_{\alpha} \mathbf{P}_{g(\alpha)}(|(\eta^*)''| > h_\varepsilon) = \varepsilon,$$

$$\lim_{n\to\infty} \sup_{\theta\in\Theta_2} \mathbf{P}_\theta(R_1(X) > e^{h_\varepsilon/2}) = \lim_{n\to\infty} \sup_{\theta\in\Theta_2} \mathbf{P}_\theta(|(\eta^*)''|^2 > h_\varepsilon) =$$
$$= \sup_{|\eta''|>b^2} \mathbf{P}(|Y'' + \delta''|^2 > h_\varepsilon) = \mathbf{P}\big((\mathrm{y}_1 + b)^2 + \mathrm{y}_2^2 + \ldots + \mathrm{y}_{k-l}^2 > h_\varepsilon\big),$$

где $\mathrm{y}_i \in \mathbf{\Phi}_{0,1}$ независимы.

Докажем теперь, что критерий (18) является асимптотически минимаксным критерием асимптотического уровня $1-\varepsilon$. Воспользуемся теоремой 54.2. В нашем случае $\eta^* = (\widehat{\rho}^* - \rho_0)\sqrt{n} \Leftarrow \Phi_{\gamma,E}$. Задача В для $Y \in \Phi_{\gamma,E}$ рассматривалась нами в примерах 49.3, 49.4. Мы установили там, что критерий

$$|Y''|^2 > h_\varepsilon$$

является минимаксным уровня $1-\varepsilon$. Следовательно, по теореме 54.2 критерий (18) является асимптотически минимаксным.

Чтобы завершить доказательство, нам осталось установить асимптотическую эквивалентность (11) и (12). Эта эквивалентность легко следует из результатов § 36 и леммы 54.1. $\triangleleft$

Пример 1. Пусть $X \in \Phi_{\lambda,\sigma^2}$, где λ, и σ^2 — скалярные параметры. (Мы используем здесь обозначение λ вместо традиционного α, чтобы не возникло путаницы с аргументом функции $g(\alpha)$.) Требуется проверить гипотезу $\{\lambda = \lambda_0\}$ против $\{\lambda \neq \lambda_0\}$ или против $\{|\lambda - \lambda_0| \geqslant bn^{-1/2}\}$, $b > 0$, когда σ неизвестно. Мы знаем, что о. м. п. в этом случае имеют следующий вид. Если обе компоненты λ и σ^2 вектора $\theta = (\lambda, \sigma^2)$ неизвестны, то о. м. п. для θ есть

$$\widehat{\theta}^* = (\lambda, \sigma^2)^* = (\overline{\mathrm{x}}, S^2), \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2.$$

Если $\lambda = \lambda_0$, то о. м. п. для σ^2 имеет вид $(\sigma^2)^* = S_1^2 = \dfrac{1}{n}\sum(\mathrm{x}_i - \lambda_0)^2$, так что $g(\widehat{\alpha}^*) = \left(\lambda_0, S_1^2\right)$. Так как

$$f_\theta(X) = (\sqrt{2\pi}\,\sigma)^{-n} \exp\left\{-(2\sigma^2)^{-1}\sum(\mathrm{x}_i - \lambda)^2\right\},$$

то критерий отношения правдоподобия (11) будет иметь вид

$$\frac{S_1^2}{S^2} > c.$$

В силу равенства $S_1^2 = S^2 + (\overline{\mathrm{x}} - \lambda_0)^2$ этот критерий эквивалентен критерию

$$\frac{|\overline{\mathrm{x}} - \lambda_0|}{S} > c_1. \tag{19}$$

Но это есть хорошо известный критерий Стьюдента, рассматривавшийся нами ранее (об оптимальных свойствах этого критерия см. (48.10) и § 47, 48).

Нетрудно проверить, что такую же форму будет иметь критерий (12). Действительно, в § 26 мы видели, что матрица $I(\theta)$ для семейства Φ_{λ,σ^2} имеет вид

$$I(\theta) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & (2\sigma^4)^{-1} \end{pmatrix}.$$

В нашем случае $\widehat{\theta}^* - g(\widehat{\alpha}^*) = \left(\overline{\mathrm{x}} - \lambda_0, S^2 - S_1^2\right) = (\overline{\mathrm{x}} - \lambda_0, -(\overline{\mathrm{x}} - \lambda_0)^2)$,

$$I^{1/2}(\widehat{\theta}^*) = \begin{pmatrix} S^{-1} & 0 \\ 0 & (\sqrt{2}\,S^2)^{-1} \end{pmatrix}.$$

Так как в левой части (12) стоит квадрат нормы $|(g(\widehat{\alpha}^*) - \widehat{\theta}^*)I^{1/2}(\widehat{\theta}^*)|^2$, то критерий (12) будет иметь вид

$$\frac{(\overline{\mathrm{x}} - \lambda_0)^2}{S^2} + \frac{(\overline{\mathrm{x}} - \lambda_0)^4}{2S^4} > c_2,$$

который, очевидно, эквивалентен (19).

Если вместо $I(\widehat{\theta}^*)$ здесь использовать $I(g(\widehat{\alpha}^*))$, то мы получим асимптотически эквивалентный критерий

$$\frac{|\overline{\mathrm{x}} - \lambda_0|}{S_1} > c_1.$$

Пример 2. Пусть $X \in \mathbf{\Phi}_{\lambda,\sigma^2}$. Требуется проверить гипотезу $\{\sigma = \sigma_0\}$ против $\{|\sigma^2 - \sigma_0^2| \geqslant bn^{-1/2}\}$, когда λ неизвестно. Здесь, очевидно, о. м. п. $\widehat{\theta}^*$ для $\theta = (\lambda, \sigma^2)$ будет той же, что и в предыдущем примере. Если $\sigma = \sigma_0$, то $\widehat{\lambda}^* = \overline{\mathrm{x}}$, так что $g(\widehat{\alpha}^*) = (\overline{\mathrm{x}}, \sigma_0^2)$, $\widehat{\theta}^* - g(\widehat{\alpha}^*) = (0, S^2 - \sigma_0^2)$.

Критерий (11) (или, что то же, критерий отношения правдоподобия) будет иметь вид

$$\frac{(S^2 - \sigma_0^2)^2}{\sigma_0^4} > 2h_\varepsilon n^{-1},$$

эквивалентный, очевидно,

$$\left|\frac{S^2}{\sigma_0^2} - 1\right| > \sqrt{2h_\varepsilon n^{-1}},$$

где $\mathbf{\Phi}_{0,1}\big((h_\varepsilon^{1/2}, \infty)\big) = \varepsilon/2$. Этот критерий также уже рассматривался в § 47.

Дальнейшие примеры использования теоремы 4 см. в § 57.

§ 56. Критерий χ^2. Проверка гипотез по сгруппированным данным

1. Критерий χ^2. Свойства асимптотической оптимальности. Сам по себе «исходный» критерий χ^2 предназначается для проверки по выборке X из полиномиального распределения $\mathbf{B}_\theta$, $\theta = (\theta_1, \ldots, \theta_r)$, $\sum_{i=1}^{r} \theta_i = 1$, простой гипотезы $H_1 = \{\theta = p\}$ против дополнительной альтернативы $H_2 = \{\theta \neq p\}$, $p = (p_1, \ldots, p_r)$. Полиномиальное распределение $\mathbf{B}_\theta$ описывается вероятностями $\theta_i = \mathbf{P}(A_i)$, $i = 1, \ldots, r$, появления в отдельном испытании одного из непересекающихся r событий $A_1, \ldots, A_r$. Элемент x_j выборки X из этого распределения можно представлять себе как один из векторов $e_1, \ldots, e_r$ с r координатами. У вектора e_k $(r-1)$ координата равна нулю, а координата с номером k равна единице; при этом $\mathrm{x}_j = e_k$, если произошло событие A_k.

Обозначим через ν_k число наступлений события A_k в n независимых испытаниях. Тогда $\nu = (\nu_1, \ldots, \nu_r) = \sum_{i=1}^{n} \mathrm{x}_i$ есть достаточная статистика для θ, так как функция правдоподобия $f_\theta(X)$ имеет вид

$$f_\theta(X) = \prod_{i=1}^{r} \theta_i^{\nu_i}. \tag{1}$$

Статистика χ^2 есть по определению

$$\chi^2(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i},$$

а критическое множество критерия χ^2 (область приема H_2) имеет вид

$$\chi^2(X) \geqslant c,$$

где c выбирается по заданному уровню значимости.

Рассмотрим более внимательно сформулированную выше задачу о проверке гипотезы $H_1 = \{\theta = p\}$ против $H_2 = \{\theta \neq p\}$.

Ясно, что распределения $\{\mathbf{B}_\theta\}$ образуют параметрическое семейство, зависящее от $k = (r-1)$-мерного параметра $(\theta_1, \ldots, \theta_{r-1})$; значение θ_r определяется равенством $\theta_r = 1 - \sum_{i=1}^{r-1} \theta_i$. Вектор $(\theta_1, \ldots, \theta_{r-1})$ так же, как и $(\theta_1, \ldots, \theta_r)$, будем обозначать буквой θ. Недоразумений от этого не возникнет. Область Θ представляет собой симплекс $\theta_i \geqslant 0$, $i = 1, \ldots, r-1$, $\sum_{i=1}^{r-1} \theta_i \leqslant 1$.

Логарифмическая функция правдоподобия $L(X, \theta)$ равна

$$L(X, \theta) = \sum_{k=1}^{r} \nu_k \ln \theta_k = \sum_{i=1}^{n} l(\mathrm{x}_i, \theta). \tag{2}$$

Семейство $\{\mathbf{B}_\theta\}$ удовлетворяет условиям (A_0), (A_μ), (A_c), а также условиям регулярности (RR) в любой внутренней точке из Θ, т. е. в любой точке θ, для которой все $\theta_i > 0$. Действительно, в нашем случае

$$l(\mathrm{x}_1, \theta) = \ln \theta_j \quad \text{при} \quad \mathrm{x}_1 = e_j;$$

$$\frac{\partial l(\mathrm{x}_1, \theta)}{\partial \theta_j} = \begin{cases} \theta_j^{-1} & \text{при} \quad \mathrm{x}_1 = e_j, \\ -\theta_r^{-1} & \text{при} \quad \mathrm{x}_1 = e_r, \\ 0 & \text{при} \quad \mathrm{x}_1 \neq e_j,\ \mathrm{x}_1 \neq e_r, \end{cases} \tag{3}$$

$$\frac{\partial^2 l(\mathrm{x}_1, \theta)}{\partial \theta_i \partial \theta_j} = \begin{cases} -\dfrac{\delta_{ij}}{\theta_i \theta_j} & \text{при} \quad \mathrm{x}_1 = e_j, \\ -\theta_r^{-2} & \text{при} \quad \mathrm{x}_1 = e_r, \\ 0 & \text{при} \quad \mathrm{x}_1 \neq e_j,\ \mathrm{x}_1 \neq e_r, \end{cases} \tag{4}$$

где δ_{ij} — символ Кронекера. Из этих формул видно, что

$$\frac{\partial^2 l(\mathrm{x}_1,\theta)}{\partial\theta_i\partial\theta_j} = -\frac{\partial l(\mathrm{x}_1,\theta)}{\partial\theta_i}\frac{\partial l(\mathrm{x}_1,\theta)}{\partial\theta_j}, \quad i,j \leqslant r-1.$$

Часть условий (RR), связанных с существованием математических ожиданий, здесь выполняется очевидным образом, так как множество $\mathcal{X}$ в нашем случае конечно.

Из (3) или (4) находим

$$I(\theta) = \|I_{ij}(\theta)\| = -\left\|\mathbf{E}_\theta \frac{\partial^2 l(\mathrm{x}_1,\theta)}{\partial\theta_i\partial\theta_j}\right\| = \left\|\frac{\delta_{ij}}{\theta_i} + \frac{1}{\theta_r}\right\|, \quad i,j = 1,\ldots,r-1. \tag{5}$$

Если в этой матрице первую строку вычесть из всех остальных, а затем воспользоваться разложением по элементам первой строки, то мы получим

$$|I(\theta)| = \left(1 + \sum_{j=1}^{r-1}\frac{\theta_j}{\theta_r}\right)\prod_{j=1}^{r-1}\theta_j^{-1} = \left(\prod_{j=1}^{r}\theta_j\right)^{-1}.$$

Таким образом, $0 < |I(\theta)| < \infty$, если $\prod\limits_{k=1}^{r}\theta_k > 0$, т. е. если точка θ является внутренней точкой симплекса Θ.

Итак, мы видим, что вправе воспользоваться результатами § 53, 54 об асимптотически оптимальных критериях. Из этих результатов вытекает, что для проверки гипотезы $H_1 = \{\theta = p\}$ против $H_2 = \{\theta \neq p\}$ существует а. б. к., совпадающий с критерием отношения правдоподобия

$$\frac{f_{\widehat{\theta}^*}(X)}{f_p(X)} > c. \tag{6}$$

Этот же критерий будет асимптотически минимаксным для проверки H_1 против гипотезы $\{(\theta-p)I(\theta)(\theta-p)^\top > b^2 n^{-1}\}$ (см. теорему 55.3).

Чтобы найти в более удобной форме критическую область (6), вычислим значение $f_{\widehat{\theta}^*}(X)$. Дифференцируя (2) по $\theta_1,\ldots,\theta_{r-1}$, получим

$$\frac{\partial L(X,\theta)}{\partial\theta_i} = \frac{\nu_i}{\theta_i} - \frac{\nu_r}{\theta_r}, \quad i = 1,\ldots,r-1.$$

Приравнивая эти производные к нулю, получим, что о. м. п. равна

$$\widehat{\theta}^* = n^{-1}\nu,$$

так что $\widehat{\theta}_i^* = n^{-1}\nu_i$.

Таким образом, переходя к логарифмам, критерий (6) можно записать в виде

$$\psi^2(X) \equiv \sum_{i=1}^{r}\nu_i \ln\frac{\nu_i}{np_i} > c_1. \tag{7}$$

Согласно теореме 53.1 (см. также лемму 53.1) статистика $2\psi^2(X)$ при гипотезе H_1 имеет предельное распределение χ^2 с $r-1$ степенями свободы. Поэтому мы получим критерий асимптотического уровня $1-\varepsilon$, если положим $c_1 = h_\varepsilon/2$, где h_ε — квантиль распределения $\mathbf{H}_{r-1}$ порядка $1-\varepsilon$.

Что же собой представляет в наших условиях асимптотически эквивалентный (6) критерий π', полученный в теореме 53.2 и имеющий вид

$$n(\widehat{\theta}^* - p)I(p)(\widehat{\theta}^* - p)^\top > h_\varepsilon? \tag{8}$$

При $t = (t_1, \ldots, t_{r-1})$, $s = \sum_{i=1}^{r-1} t_i$, получим

$$\begin{gathered} tI(p) = \left(\frac{t_1}{p_1} + \frac{s}{p_r}, \ldots, \frac{t_{r-1}}{p_{r-1}} + \frac{s}{p_r}\right), \\ tI(p)t^\top = \sum_{i=1}^{r-1} \frac{t_i^2}{p_i} + \frac{s^2}{p_r} = \sum_{i=1}^{r} \frac{t_i^2}{p_i}, \end{gathered} \tag{9}$$

где

$$t_r = -s, \quad \sum_{i=1}^{r} t_i = 0. \tag{10}$$

Полагая $t = \widehat{\theta}^* - p$ и замечая, что условие (10) выполнено, мы получим в качестве (8) критерий

$$\sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} > h_\varepsilon. \tag{11}$$

Это есть не что иное, как критерий χ^2. Из цитированных утверждений вытекает, что $\chi^2(X) \Subset \mathbf{H}_{r-1}$.

Критерий π'' в теореме 53.2, асимптотически эквивалентный (7) и (11), будет иметь вид

$$\sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{\nu_i} > h_\varepsilon. \tag{12}$$

Учитывая также теорему 55.3 и замечание 55.1, мы можем резюмировать изложенное в виде следующего утверждения.

Теорема 1. *Критерий* (7) *при* $c_1 = h_\varepsilon/2$, *а также критерий* χ^2 (11) *и критерий* (12) *имеют асимптотический уровень* $1-\varepsilon$ *и являются а. б. к. для проверки по выборке* $X \in \mathbf{B}_\theta$ *гипотезы* $\{\theta = p\}$ *против* $\{\theta \neq p\}$. *Они же являются асимптотически минимаксными критериями для проверки гипотезы* $\{\theta = p\}$ *против близкой альтернативы* $\left\{\sum_{i=1}^{r}(\theta_i - p_i)^2/p_i > b^2/n\right\}$ *при любом* $b > 0$.

Асимптотическую эквивалентность критериев (7), (11), (12) можно было бы установить и непосредственно, пользуясь разложением в ряд $\ln \dfrac{\nu_i}{np_i} = \ln\left(1 + \dfrac{\nu_i - np_i}{np_i}\right)$ в (7).

Эти критерии являются асимптотически непараметрическими, так как предельное распределение используемых в них статистик является «абсолютным», т. е. с природой исходного распределения никак не связано.

2. Применения критерия χ^2. Проверка гипотез по сгруппированным данным. Критерий χ^2 является очень распространенным, и значение его выходит за пределы задачи, рассмотренной в предыдущем пункте.

Обратимся к общей задаче о проверке гипотезы $H_1 = \{X \in \mathbf{P}_1\}$ против $H_2 = \{X \in \mathbf{P}, \mathbf{P} \neq \mathbf{P}_1\}$, изучавшейся в § 52. Так как сколько-нибудь развитая теория об оптимальных критериях существует лишь в параметрическом случае, то естественно пытаться как-то «параметризовать» эту задачу *).

Наиболее простым и естественным способом сделать это в общем случае является *группировка данных.* Она состоит в следующем. Область возможных значений наблюдаемых величин (т. е. пространство $\mathcal{X}$) разбивается на r непересекающихся областей $\Delta_1, \ldots, \Delta_r$ и вместо наблюдения x_j указывается лишь тот интервал Δ_k, в который это наблюдение попало.

Другими словами, мы огрубляем наблюдения, и те x_j, которые попали в Δ_k, можем заменить каким-нибудь одним значением $z_k \in \Delta_k$. Ясно, что, выбирая достаточно богатое разбиение, мы можем приблизить наблюдение x_j с помощью z_k сколь угодно точно.

Итак, группировка приводит к тому, что наблюдение x_j заменяется вектором e_k, если произошло событие $A_k = \{\mathrm{x}_j \in \Delta_k\}$ (векторы e_k были определены в начале предыдущего пункта). Но новая выборка, полученная в результате такой операции, очевидно, есть не что иное, как выборка из $\mathbf{B}_\theta$, $\theta_k = \mathbf{P}(\mathrm{x}_j \in \Delta_k)$. Мы уже знаем, что в этом случае вектор $\nu = (\nu_1, \ldots, \nu_r)$ частот попаданий в интервалы $\Delta_i, \ldots, \Delta_r$ будет достаточной статистикой.

Произведенная редукция выборки X к вектору ν и называется *группировкой данных.*

Ясно, что группировка связана с некоторым «обеднением» выборки X и частичной потерей информации.

На произведенную параметризацию возможен и несколько иной взгляд. Допустим для наглядности, что $\mathcal{X} = \mathbb{R}$, и все рассматриваемые распределения сосредоточены на конечном интервале и имеют плотность, т. е. удовлетворяют условию (A_μ), где μ — мера Лебега. При данном разбиении $\Delta, \ldots, \Delta_r$ наряду с плотностью $f(x)$ рассмотрим кусочно-постоянную плотность

$$f_\theta(x) = \frac{\mathbf{P}_i(\Delta_i)}{\Delta_i} = \frac{1}{\Delta_i} \int\limits_{\Delta_i} f(x)\,dx = \frac{\theta_i}{\Delta_i} \quad \text{при} \quad x \in \Delta_i. \tag{13}$$

*) Имеется в виду конечномерный параметр. Любую задачу можно считать параметрической, если допускать бесконечномерный параметр, так как его можно отождествить с распределением $\mathbf{P}$, $X \in \mathbf{P}$.

Здесь Δ_i обозначает также длину интервала Δ_i. Это есть параметрическое семейство распределений $\mathbf{P}_\theta$, $\mathbf{P}_\theta(B) = \int\limits_B f_\theta(x)\,dx$.

Мы получим выборку Y из $\mathbf{P}_\theta$, если для каждого k соберем все наблюдения из $X \in \mathbf{P}$, попавшие в Δ_k, и затем случайно и равномерно «разбросаем» их по Δ_k. По существу, это то же самое, что мы делали раньше, так как сведения о том, в какой точке интервала Δ_k находится наблюдение y_i, никакой информации в себе относительно параметра θ не содержит: функция правдоподобия $f_\theta(Y)$ при «перемещении» наблюдений в пределах своих интервалов не меняется. Таким образом, достаточно лишь знать количества $\nu_1, \ldots, \nu_r$ наблюдений, попавших в $\Delta_1, \ldots, \Delta_r$.

Ясно, что если $f(x)$ — гладкая функция, то $f_\theta(x)$ будет хорошо приближать $f(x)$ при достаточно «мелком» разбиении $\{\Delta_1, \ldots, \Delta_r\}$.

Соотношения (13) означают другой способ параметризации, эквивалентный первому. Эквивалентность следует из совпадения с точностью до множителя, не зависящего от параметра, функций правдоподобия. Для распределения (13) она равна

$$f_\theta(Y) = \prod_{i=1}^{r} \theta_i^{\nu_i} \prod_{i=1}^{r} \Delta_i^{-\nu_i},$$

где первый множитель представляет собой функцию правдоподобия выборки из $\mathbf{B}_\theta$ (см. (1)).

Следует отметить, что группировка наблюдений возникает часто и сама по себе, не для целей параметризации, а просто как удобный и часто более экономный способ записи информации, содержащейся в выборке. Если, например, $n = 10^4$, а точность измерений, наблюдаемых значений на $[0, 1]$ сравнима с $0,1$, то ясно, что практически бесполезно знать все 10^4 наблюдений, а достаточно указать 10 частот $\nu_1, \ldots, \nu_{10}$ попадания в интервалы $\Delta_i = ((i-1)/10, i/10)$, $i = 1, \ldots, 10$, т. е. знать лишь *гистограмму выборки.*

Вернемся к задаче проверки гипотезы $H_1 = \{X \in \mathbf{P}_1\}$ против $H_2 = \{X \in \mathbf{P} \neq \mathbf{P}_1\}$. Мы будем предполагать произведенную группировку наблюдений такой, что значимое для нас отклонение распределения $\mathbf{P}$ выборки X от $\mathbf{P}_1$ обязательно скажется на распределении сгруппированных данных. Тогда нашу задачу можно рассматривать как задачу проверки гипотезы $\{\theta = p\}$, где $p_i = \mathbf{P}_1(\Delta_i)$, против $\{\theta \neq p\}$ для параметрических семейств $\mathbf{B}_\theta$ или (13). Как мы уже знаем, *критерий* χ^2 (*как и критерии* (7), (12)) *будет асимптотически оптимальным в этой задаче в смысле, сформулированном в теореме* 1.

Кроме того, критерий χ^2 является *асимптотически непараметрическим*, так как предельное распределение статистики $\chi^2(X)$ при гипотезе H_1 от исходного распределения выборки X не зависит.

Отметим при этом, что проверка гипотезы $\{\theta = p\}$ для семейств (13) или $\mathbf{B}_\theta$ все же не равнозначна проверке гипотезы $\{X \in \mathbf{P}_1\}$, хотя при некотором разбиении $\{\Delta_1, \ldots, \Delta_r\}$ может быть и близка к ней. На самом деле, для выборки X проверяется гипотеза $X \in \mathbf{P}$, $\mathbf{P}(\Delta_i) = p_i = \mathbf{P}_1(\Delta_i)$. Это делает

критерий χ^2 *несостоятельным* относительно альтернатив $\mathbf{P} \neq \mathbf{P}_1$, для которых $\theta_i = \mathbf{P}(\Delta_i) = \mathbf{P}_1(\Delta_i) = p_i$. Поэтому еще раз отметим, что критерий χ^2 есть критерий, обладающий рядом свойств асимптотической оптимальности, но он действует лишь против альтернатив, меняющих вектор θ, т. е. против альтернатив, для которых $\{\mathbf{P}(\Delta_i)\} \neq \{\mathbf{P}_1(\Delta_i)\} = \{p_i\}$.

Сделаем несколько замечаний, касающихся применений критериев χ^2, (7), (12). Мы будем говорить при этом в основном лишь о критерии χ^2, так как, с одной стороны, названные критерии близки друг к другу, а с другой — критерий χ^2 исторически (отчасти ввиду своей наглядности) получил значительно более широкое распространение.

Уровень значимости критерия $\chi^2(X) > h_\varepsilon$ равен $1-\varepsilon$ лишь «в пределе». Опыт показывает, что для $\varepsilon \geqslant 0{,}01$ истинный уровень значимости этого критерия удовлетворительно приближается значением $1-\varepsilon$ лишь при $np_i \geqslant 8$, $i = 1, \ldots, r$.

При большом числе групп r, скажем, при $n > r > 30$, можно пользоваться нормальным приближением как для распределения $\dfrac{1}{\sqrt{2r}}(\chi_r^2 - r)$, $\chi_r^2 \in \mathbf{H}_r$ (см. § 12), так и для распределения при гипотезе H_1 статистики $\chi^2(X)$, нормированной моментами

$$\mathbf{E}\chi^2(X) = r - 1,$$

$$\mathbf{D}\chi^2(X) = 2(r-1) + \frac{1}{n}\left(\sum_{i=1}^{r} p_i^{-1} - r^2 - 2r + 2\right).$$

Часто используют также нормальное приближение $\mathbf{\Phi}_{0,1}$ для распределения случайной величины (см. § 12) $\sqrt{2\chi_r^2} - \sqrt{2r-1}$, $\chi_r^2 \in \mathbf{H}_r$.

Отметим также, что при увеличении числа групп улучшается приближение плотности $f(x)$ с помощью ступенчатой функции, построенной по значениям $\mathbf{P}_1(\Delta_i) = \int\limits_{\Delta_i} f(x)\,dx$. Это значит, что возрастает число альтернатив, не согласующихся с H_1, и критерий χ^2 все более становится критерием для проверки *гипотез о плотности*. В связи с этим мощность критериев χ^2 фиксированного уровня при увеличении числа групп будет убывать (ср. с замечаниями предыдущего параграфа о критерии Морана; подробнее об этом см. [14, 135]).

К недостаткам критерия χ^2 следует отнести тот факт, что в ряде случаев разбиение $\{\Delta_1, \ldots, \Delta_r\}$ приходится устанавливать статистику. Здесь следует проявлять осторожность, так как при этом вносится элемент субъективизма в «обеднение» выборки X. Кроме того, иногда это разбиение выбирают в зависимости от выборки X, что, вообще говоря, не всегда допустимо, так как при этом Δ_i становятся случайными (подробнее об этом см. [50], § 30, 31).

Пример 1*). Некто в городе N обследовал показания 500 часов, выставленных в витринах различных часовых мастерских. Результаты на-

*) Пример заимствован из [58].

блюдений были разбиты на 12 групп (в зависимости от положения часовой стрелки на циферблате). Приведем полученную таблицу наблюдений:

Промежутки на циферблате	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11	11–12
Число наблюдений	41	34	54	39	49	45	41	33	37	41	47	39

Проверяется простая гипотеза $H_1 = \{$*распределение положения стрелки по часовым группам на циферблате равномерно*$\}$ против сложной дополнительной альтернативы.

В этом примере $n = 500$, $p_i = 1/12$, $i = 1, \ldots, 12$; $np_i \approx 41{,}67$. На основании теоремы 1 можно приближенно считать, что $\chi^2(X) \in \mathbf{H}_{11}$. В нашем примере непосредственным подсчетом убеждаемся, что $\chi^2(X) \approx 10$ и фактически достигаемый уровень критерия χ^2 приближенно равен $1 - \mathbf{H}_{11}((10, \infty)) \approx 0{,}47$ (см. табл. III). Это означает, что результаты эксперимента согласуются с гипотезой H_1 с точки зрения критерия χ^2 любого уровня $1 - \varepsilon$, заключенного между 0,47 и 1.

Мы уже отмечали, что критерий χ^2 имеет широкое распространение. При этом сфера его приложений не ограничивается проверкой лишь простых гипотез. Один из таких примеров будет рассмотрен в следующем параграфе.

§ 57. Проверка гипотез о принадлежности выборки параметрическому семейству

Рассмотрим задачу о проверке сложной гипотезы $H_1 = \{X \in \mathbf{P}_\alpha, \alpha \in A\}$ о том, что распределение выборки принадлежит параметрическому семейству $\{\mathbf{P}_\alpha\}_{\alpha \in A}$ против дополнительной альтернативы $H_2 = \{X \in \mathbf{P}, \mathbf{P} \notin \{\mathbf{P}_\alpha\}_{\alpha \in A}\}$. Примером такого рода гипотез могут служить утверждения о том, что X есть выборка из какой-нибудь нормальной совокупности (гипотеза H_1), и утверждение, дополнительное к этому (гипотеза H_2).

Другим примером может служить проверка гипотезы о том, что $X \in \mathbf{B}_{\theta(\alpha)}$, где размерность α меньше размерности θ. Эту задачу можно, конечно, трактовать и как задачу о проверке гипотезы о принадлежности X параметрическому подсемейству (см. § 55), однако верной будет и первая трактовка, поскольку в том случае, когда в результате эксперимента наступает лишь конечное число возможных исходов (см. определение $\mathbf{B}_\theta$ в § 12), семейство $\mathbf{B}_\theta$ содержит в себе *все* мыслимые распределения выборки.

В следующем пункте мы рассмотрим задачу о проверке гипотезы $X \in \mathbf{B}_{\theta(\alpha)}$ и покажем, что общая задача о принадлежности параметрическому семейству может быть редуцирована к этой задаче путем группировки данных.

1. Проверка гипотезы $X \in \mathbf{B}_{\theta(\alpha)}$. Группировка данных. Рассмотрим сначала общую задачу, сформулированную в начале параграфа, в случае произвольного пространства $\mathcal{X}$. Возьмем разбиение пространства $\mathcal{X}$ на области («интервалы») $\{\Delta_1, \ldots, \Delta_r\}$ такое, что число «интервалов» r больше $l + 1$,

где l есть размерность параметра α. Произведем группировку наблюдений по этим интервалам. Если гипотеза $H_1 = \{X \in \mathbf{P}_\alpha\}$ верна, то вероятности попадания наблюдений в интервалы Δ_i будут равны

$$p_i(\alpha) = \mathbf{P}_\alpha(\Delta_i).$$

Это означает, что в этом случае вектор $\theta = (\theta_1, \ldots, \theta_r)$ вероятностей попадания наблюдения в Δ_i должен лежать на кривой $\theta = p(\alpha) = (p_1(\alpha), \ldots, p_r(\alpha))$.

Таким образом, по полученной при группировке выборке $Y \in \mathbf{B}_\theta$ мы должны проверить гипотезу H_1 о принадлежности Y параметрическому подсемейству $\mathbf{B}_{p(\alpha)}$ против альтернативы $\{Y \in \mathbf{B}_\theta\}$, где θ не располагается на кривой $\theta = p(\alpha)$, $\alpha \in A$. Эта задача рассматривалась нами в § 55, где был найден асимптотически минимаксный критерий для проверки H_1 против близкой альтернативы

$$H_2 = \{Y \in \mathbf{B}_\theta;\ \inf_\gamma |\theta - p(\alpha_0 + \gamma n^{-1/2}) I^{1/2}(p(\alpha_0 + \gamma n^{-1/2}))^\top| > b n^{-1/2}\} \quad (1)$$

(см. замечание 55.3 к теореме 55.4; точка α_0 означает «локализованное» значение параметра такое, что альтернативы располагаются в окрестности точки $\theta_0 = p(\alpha_0)$). Критерий отношения правдоподобия (55.11) в нашем случае имеет вид

$$\ln R_1(X) = \max_\theta \sum_{i=1}^r \nu_i \ln \theta_i - \max_\alpha \sum_{i=1}^r \nu_i \ln p_i(\alpha) > \frac{h_\varepsilon}{2}$$

или, что то же,

$$\sum_{i=1}^r \nu_i \ln \frac{\nu_i}{n p_i(\widehat{\alpha}^*)} > \frac{h_\varepsilon}{2},$$

где $\widehat{\alpha}^*$ есть о. м. п. параметра α по выборке Y (или по вектору $\nu = (\nu_1, \ldots, \nu_r)$). Этот критерий асимптотически эквивалентен (см. теорему 55.4) критерию

$$(p(\widehat{\alpha}^*) - \nu n^{-1}) I(p(\widehat{\alpha}^*))(P(\widehat{\alpha}^*) - \nu n^{-1})^\top > h_\varepsilon.$$

Так как вид матрицы $I(\theta)$ нам известен (см. (56.5)), то, пользуясь (56.9), мы получим из теоремы 55.4

Следствие 1. *Если $r - 1 > l$ и функция $p(\alpha)$ удовлетворяет условиям теоремы* 55.4, *то критерий отношения правдоподобия асимптотического уровня $1 - \varepsilon$ для проверки по сгруппированным данным гипотезы $H_1 = \{X \in \mathbf{P}_\alpha, \mathbf{P}_\alpha \in \{\mathbf{P}_\alpha\}_{\alpha \in A}\}$ против дополнительной альтернативы H_2 является асимптотически минимаксным (для проверки гипотезы H_1 против* (1)) *и имеет вид*

$$\sum_{i=1}^r \nu_i \ln \frac{\nu_i}{n p_i(\widehat{\alpha}^*)} > \frac{h_\varepsilon}{2}, \quad (2)$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с $r-l-1$ степенями свободы. Этот критерий асимптотически эквивалентен критерию

$$\widehat{\chi}^2(X)=\sum_{i=1}^{r}\frac{\bigl(\nu_i-np_i(\widehat{\alpha}^*)\bigr)^2}{np_i(\widehat{\alpha}^*)}>h_\varepsilon. \tag{3}$$

Последний критерий называют также *критерием* χ^2, но в случае, когда неизвестные «мешающие» параметры оцениваются по выборке. Распределение статистики $\widehat{\chi}^2(X)$, как это вытекает из следствия 1, сходится при гипотезе H_1 к распределению χ^2 с $r-l-1$ степенями свободы (число степеней свободы $r-1$ в предельном распределении статистики $\chi^2(X)$ уменьшилось на число скалярных параметров $\alpha_1,\dots,\alpha_l$, оцениваемых по выборке).

Пример 1. В примере 35.3 был описан механизм наследования групп крови 0 (ноль), A, B, AB. Он управляется генами трех типов: A, B, 0. Вероятности появления этих генов в данной популяции обозначим через p, q, $r=1-p-q$. Вероятности $p_i(\alpha)$, $\alpha=(p,q)$, того, что человек будет иметь i-ю группу крови, найдены в примере 35.3 и приведены в табл. 1 из § 35.

Нам дана выборка X с частотами ν_i, $i=1,2,3,4$ (см. табл. 1), появления i-й группы крови, полученная в результате обследования $n=353$ человек. В примере 35.3 для этой выборки были найдены приближенные значения о. м. п. $\widehat{\alpha}^*=(p^*,q^*)=(0{,}246;\ 0{,}173)$. Это дает нам значения $p_i(\widehat{\alpha}^*)$, приведенные в табл. 1.

Таблица 1

Распределение людей по группам крови

	0	A	B	AB	Всего
ν_i	121	120	79	33	353
$p_i^*=\nu_i n^{-1}$	0,343	0,340	0,224	0,093	1
$p_i(\widehat{\alpha}^*)$	0,337	0,347	0,231	0,085	1

Мы получаем возможность воспользоваться следствием 1 для проверки гипотезы о том, что описанный выше механизм наследования крови имеет место. С помощью табл. 1 находим, что статистика $\widehat{\chi}^2(X)$ (см. (3)) в нашем случае равна примерно 0,44. Это дает хорошее согласование с гипотезой, так как критическое значение h_ε, соответствующее распределению χ^2 с одной степенью свободы и значению $\varepsilon=0{,}2$, равно $h_{0,2}\approx 1{,}64$.

Пример 2. Задача о сопряженных признаках. Допустим, что выборка X есть результат обследования некоторых объектов, каждый из которых характеризуется признаками A и B. Первый признак может принимать значения $A_1,\dots,A_s$, второй — $B_1,\dots,B_t$. Спрашивается, зависимы ли эти признаки? Например, мы можем проводить некоторый эксперимент G с исходами $B_1,\dots,B_t$ в разных условиях $A_1,\dots,A_s$. Задача состоит в том, чтобы выяснить, зависят ли результаты эксперимента G от условий, в которых он производится.

Эту задачу можно рассматривать также как задачу о проверке независимости двух случайных величин ξ и η по сгруппированным наблюдениям над парой (ξ,η).

Результаты экспериментов в нашем примере представляют собой матрицу значений $\|\nu_{ij}\|$, где ν_{ij} есть число появлений исходов с признаками A_i и B_j в выборке X объема n (каждый элемент выборки есть пара признаков обследуемого объекта).

Положим $p_{ij} = \mathbf{P}(A_iB_j)$, $p_{i\cdot} = \sum_{j=1}^{t} p_{ij}$, $p_{\cdot j} = \sum_{i=1}^{s} p_{ij}$. Тогда гипотеза H_1 о независимости признаков будет иметь вид $H_1 = \{p_{ij} = p_{i\cdot}p_{\cdot j}\}$. Нетрудно видеть, что это есть гипотеза о принадлежности распределения выборки параметрическому подсемейству, где роль параметра α играет $(s+t-2)$-мерный вектор $\alpha = (p_{1\cdot}, \ldots, p_{s-1,\cdot}; p_{\cdot 1}, \ldots, p_{\cdot,t-1})$ (значения $p_{s\cdot}$ и $p_{\cdot t}$ определяются из равенств $p_{s\cdot} = 1 - \sum_{i=1}^{s-1} p_{i\cdot}$, $p_{\cdot t} = 1 - \sum_{j=1}^{t-1} p_{\cdot j}$).

Функция правдоподобия выборки X при гипотезе H_1 равна

$$\prod_{i,j} p_{ij}^{\nu_{ij}} = \prod_i p_{i\cdot}^{\nu_{i\cdot}} \prod_j p_{\cdot j}^{\nu_{\cdot j}}, \quad \nu_{i\cdot} = \sum_{j=1}^{t} \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=1}^{s} \nu_{ij}.$$

Из результатов § 56 (ср. с (56.1)) следует, что о. м. п. $\widehat{\alpha}^*$ для такой функции правдоподобия имеет вид

$$\widehat{p}_{i\cdot}^* = \frac{\nu_{i\cdot}}{n}, \quad \widehat{p}_{\cdot j}^* = \frac{\nu_{\cdot j}}{n}.$$

Итак, критерий χ^2 в нашем случае принимает вид

$$\widehat{\chi}^2(X) = \sum_{i,j} \frac{(\nu_{ij} - n\widehat{p}_{i\cdot}^*\widehat{p}_{\cdot j}^*)^2}{n\widehat{p}_{i\cdot}^*\widehat{p}_{\cdot j}^*} = n \sum_{i,j} \frac{(\nu_{ij} - n^{-1}\nu_{i\cdot}\nu_{\cdot j})^2}{\nu_{i\cdot}\nu_{\cdot j}} > h_\varepsilon,$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с числом степеней свободы $st - 1 - (s+t-2) = (s-1)(t-1)$.

Можно указать много прикладных задач, где используется построенный нами критерий сопряженности признаков. Рассмотрим в качестве иллюстрации одну из них — задачу социологического обследования о связи между доходом семей и количеством детей в них (см. [58, с. 481]).

Пример 2А. Пусть признак A означает количество детей и принимает значения 0, 1, 2, 3, $\geqslant 4$. Признак B указывает, какому диапазону (0–1), (1–2), (2–3), ($\geqslant 3$) (за единицу принято 1000 шведских крон) принадлежит зарплата. По результатам $n = 25\,263$ обследований были получены следующие данные, приведенные в табл. 2.

Таблица 2

Признак A	Признак B				
	0–1	1–2	2–3	$\geqslant 3$	$\nu_{i\cdot}$
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
$\geqslant 4$	39	98	31	14	182
$\nu_{\cdot j}$	6116	10928	5173	3016	25263

В этом примере $\widehat{\chi}^2(X) = 568{,}5$, что значимо больше критического значения h_ε для распределения χ^2 с $(5-1)(4-1) = 12$ степенями свободы даже при достаточно малых ε. Так что мы вынуждены признать несостоятельность гипотезы $H_1 = \{A \text{ и } B \text{ независимы (несопряжены)}\}$.

Отметим, однако, что более тонкий анализ указывает на *очень слабую* зависимость признаков A и B.

2. Общий случай. Критерий χ^2, примененный в задаче этого параграфа, обладает теми же недостатками, которые отмечались нами применительно к задачам предыдущего параграфа.

Задача проверки гипотезы $\{X \in \mathbf{P}_\theta\}$ о принадлежности X параметрическому семейству $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$ допускает, конечно, и более широкий подход, аналогичный тому, который излагается в § 52. Выберем некоторое расстояние $d(\mathbf{P}, \mathbf{Q})$ в пространстве распределений. Далее, найдем точку $\mathbf{P}_{\theta^*}$ из $\{\mathbf{P}_\theta\}$, ближайшую к $\mathbf{P}_n^*$ в смысле расстояния d. В качестве $\mathbf{P}_{\theta^*}$ можно брать также $\mathbf{P}_{\widehat{\theta}^*}$, где $\widehat{\theta}^*$ есть о. м. п. (см. § 15) или какая-нибудь другая разумная оценка. Если гипотеза H_1 верна, то $d\big(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*\big)$ не должно быть большим, и наоборот, если верна H_2, то $d\big(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*\big)$ будет значимо. Это соображение дает нам следующую конструкцию критерия: мы отвергаем гипотезу H_1, если $d\big(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*\big) > c$, и принимаем в противном случае.

Число c следует выбрать так, чтобы

$$\sup_{\theta\in\Theta} \mathbf{P}_\theta\big(d\big(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*\big) > c\big) \leqslant \varepsilon$$

или чтобы это соотношение выполнялось асимптотически. Следствие 1 предлагает в качестве расстояния $d\big(\mathbf{P}_{\theta^*}, \mathbf{P}_n^*\big)$ взять статистики в (2), (3). Среди других они обладают еще и тем достоинством, что являются асимптотически непараметрическими: при гипотезе $H_1 = \{X \in \mathbf{P}_\theta\}$ предельное распределение $\widehat{\chi}^2(X)$, например, от θ не зависит.

Рассмотрим реализацию изложенного выше общего подхода в двух важных частных случаях, когда параметрические семейства образованы параметрами сдвига и масштаба.

1. Пусть мы проверяем гипотезу $X \in \mathbf{P}_\theta$, $\theta \in \mathbb{R}$, где $\mathbf{P}_\theta(A) = \mathbf{P}(A-\theta)$, $A \subset \mathbb{R}$. Обозначим через $F(x)$ функцию распределения, соответствующую $\mathbf{P}$, и положим $F_\theta(x) = F(x-\theta)$. В качестве d возьмем расстояние, использованное нами в критерии Колмогорова.

Теорема 1. *Пусть $X \in \mathbf{P}_\theta$, $F_\theta(x) = F(x-\theta)$, функция $F(x)$ имеет ограниченную равномерно непрерывную плотность $f(x) = F'(x)$, $\int x^2 f(x)\,dx < \infty$. Если обозначить $\int x f(x)\,dx = a$, $\theta^* = \overline{\mathrm{x}} - a$, то при любом θ*

$$\lim_{n\to\infty} \mathbf{P}_\theta\big(\sup_x \sqrt{n}\,\big|F_n^*(x) - F_{\theta^*}(x)\big| > c\big) =$$

$$= \mathbf{P}\left(\sup_x \left|w^0(F(x)) + f(x)\int w^0(F(t))\,dt\right| > c\right),$$

где w^0 — стандартный броуновский мост.

Правая часть в этом соотношении от θ не зависит. Вычисляя ее для заданной F и выбирая $c = c_\varepsilon$ так, чтобы она равнялась ε, мы получим критерий

$$D_n \equiv \sup_x \sqrt{n}\,\bigl|F_n^*(x) - F(x-\theta^*)\bigr| > c_\varepsilon$$

асимптотического уровня $1-\varepsilon$ для проверки гипотезы H_1 о принадлежности выборки X параметрическому семейству $\{\mathbf{P}_\theta\}$, где θ — параметр сдвига.

Доказательство теоремы 1. Рассмотрим процесс

$$W_n(x) = \sqrt{n}\,\bigl(F_n^*(x) - F_{\theta^*}(x)\bigr) = w_n(x) - \sqrt{n}\,(F_{\theta^*}(x) - F_\theta(x)),$$

где $w_n(x) = \sqrt{n}\,\bigl(F_n^*(x) - F_\theta(x)\bigr)$. Имеем при $t \to \theta$

$$F_t(x) - F_\theta(x) = -(t-\theta)(f(x-\theta) + \varepsilon(t,\theta,x)),$$
$$|\varepsilon(t,\theta,x)| \leqslant \omega_{|t-\theta|},$$

где ω_Δ — не зависящий от x модуль непрерывности функции f, $\omega_\Delta \to 0$ при $\Delta \to 0$. Так как $\theta^* \xrightarrow[P_\theta]{} \theta$, то, полагая $t = \theta^*$ и принимая, не ограничивая общности, $a = 0$, получим

$$\sqrt{n}\,(F_{\theta^*}(x) - F_\theta(x)) = -f(x-\theta)\int t d\bigl[\sqrt{n}\,\bigl(F_n^*(t) - F_\theta(t)\bigr)\bigr] + \varepsilon(\theta^*,\theta,x) =$$

$$= -f(x-\theta)\int t\,dw_n(t) + \varepsilon(\theta^*,\theta,x),$$

$$|\varepsilon(\theta^*,\theta,x)| \leqslant \omega(\theta^*-\theta) \equiv \sqrt{n}\,|\theta^*-\theta|\omega_{|\theta^*-\theta|} \xrightarrow[P_\theta]{} 0.$$

Далее, функционал

$$H_N(w_n) = \sup_x \left| w_n(x) - f(x-\theta)\int_{-N}^{N} w_n(t)\,dt \right|$$

при любом $N > 0$ является непрерывным в равномерной метрике. Замена переменной x на $F_\theta^{-1}(y) = \theta + F^{-1}(y)$, которую естественно сделать для применения теоремы 6.3, этого факта не меняет. Поэтому в силу названной теоремы

$$H_N(w_n) \Rightarrow \sup_x \left| w^0(F(x-\theta)) + f(x-\theta)\int_{-N}^{N} w^0(F(t-\theta))\,dt \right|.$$

Чтобы доказать требуемое соотношение

$$D_n \Rightarrow \sup_x \left| w^0(F(x-\theta)) + f(x-\theta)\int w^0(F(t-\theta))\,dt \right|$$

(сдвиг на θ аргумента в правой части значение последней не меняет), нам в силу соотношений

$$|D_n - H_N(w_n)| \leqslant \omega(\theta^* - \theta) + c \int\limits_{|t|\geqslant N} w_n(t)\,dt, \qquad \omega(\theta^* - \theta) \underset{P_\theta}{\longrightarrow} 0, \tag{4}$$

остается убедиться, что интеграл в (4) вместе с интегралом $\int\limits_{|t|\geqslant N} w^0(F(t))\,dt$ (положим для простоты $\theta = 0$) сходятся по вероятности к нулю при $n \to \infty$, $N \to \infty$. Оценки обоих названных интегралов проще всего, по-видимому, провести, доказав малость их дисперсий и воспользовавшись неравенством Чебышева. Так как первые два момента подынтегральных выражений в обоих интегралах ведут себя одинаково, то мы можем ограничиться оценкой лишь одного из этих интегралов. Рассмотрим, например,

$$\int\limits_{-\infty}^{-N} w^0(F(t))\,dt.$$

В силу соотношений $\mathbf{E}w^0(s)w^0(u) = \min(s,u) + su \leqslant 2\min(s,u)$ при $s \leqslant 1$, $u \leqslant 1$ имеем

$$\mathbf{E}\left(\int\limits_{-\infty}^{-N} w^0(F(t))\,dt\right)^2 \leqslant 2\int\limits_{-\infty}^{-N}\int\limits_{-\infty}^{-N} \min(F(t),F(s))\,dt\,ds = 4\int\limits_{-\infty}^{-N} (-t-N)F(t)\,dt \leqslant -8\int\limits_{-\infty}^{-N} tF(t)\,dt \to 0$$

при $N \to \infty$, так как $\int t^2\,dF(t) < \infty$. Аналогично рассматриваются остальные интегралы. $\triangleleft$

2. Пусть теперь проверяется гипотеза $X \in \mathbf{P}_\theta$, $\theta \in \mathbb{R}$, $\theta > 0$, где $\mathbf{P}_\theta(A) = \mathbf{P}(A/\theta)$, $A \subset \mathbb{R}$. Обозначим снова через F функцию распределения, соответствующую $\mathbf{P}$, и положим

$$\sigma^2 = \mathbf{E}_1 \mathrm{x}_1^2 = \int x^2 \mathbf{P}(dx), \quad \theta^* = \frac{1}{\sigma}\sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathrm{x}_i^2}.$$

Теорема 2. *Пусть $X \in \mathbf{P}_\theta$, $F_\theta(x) = F(x/\theta)$, существует ограниченная непрерывная плотность $f(x) = F'(x)$ такая, что*

$$\int x^4 f(x)\,dx < \infty, \quad xf(x) \to 0 \tag{5}$$

при $x \to 0$. *Тогда при любом* θ

$$\lim_{n\to\infty} \mathbf{P}_\theta\left(\sup_x \sqrt{n}\left|F_n^*(x) - F(x/\theta^*)\right| > c\right) =$$

$$= \mathbf{P}\left(\sup_x \left|w^0(F(x)) + xf(x)\int tw^0(F(t))\,dt\right| > c\right).$$

Доказательство этой теоремы совершенно аналогично доказательству теоремы 1. Имеем

$$W_n(x) \equiv \sqrt{n}\left(F_n^*(x) - F\left(\frac{x}{\theta^*}\right)\right) = w_n(x) - \sqrt{n}\left(F\left(\frac{x}{\theta^*}\right) - F\left(\frac{x}{\theta}\right)\right),$$

$$w_n(x) = \sqrt{n}\left(F_n^*(x) - F\left(\frac{x}{\theta}\right)\right).$$

При $t \to \infty$

$$F_t(x) - F_\theta(x) = \left(\frac{x}{t} - \frac{x}{\theta}\right)\left(f\left(\frac{x}{\theta}\right) + \varepsilon(t,\theta,x)\right),$$

где в силу соотношения $f(x) = o(1/|x|)$ и равномерной непрерывности f на любом конечном интервале выполняется $\sup\limits_x |x\varepsilon(t,\theta,x)| \leqslant \omega_{|t-\theta|} \to 0$. Полагая $t = \theta^* \underset{P_\theta}{\longrightarrow} \theta$, получим

$$\sqrt{n}\left(F\left(\frac{x}{\theta^*}\right) - F\left(\frac{x}{\theta}\right)\right) = \sqrt{n}\left(\frac{x}{\theta^*} - \frac{x}{\theta}\right) f\left(\frac{x}{\theta}\right) - \frac{\sqrt{n}\,(\theta^* - \theta)}{\theta^*}\frac{x}{\theta}\varepsilon(\theta^*,\theta,x),$$

где $\sup\limits_x$ второго слагаемого сходится по $\mathbf{P}_\theta$-вероятности к нулю. Нам остается воспользоваться рассуждениями предыдущей теоремы (малость интегралов $\int\limits_{|t|>N} tw^0(F(t))\,dt$ и $\int\limits_{|t|>N} tw_n(t)\,dt$ обеспечивается условием (5)) и заметить, что главная часть $W_n(x)$ равна (примем, не ограничивая общности, $\sigma^2 = 1$)

$$w_n(x) - \frac{\sqrt{n}\,x(\theta^{*2} - \theta^2)}{\theta\theta^*(\theta+\theta^*)} f\left(\frac{x}{\theta}\right) =$$

$$= w_n(x) - \frac{xf(x/\theta)}{\theta\theta^*(\theta+\theta^*)}\int t^2 dw_n(t) w_n(x) = w_n(x)\frac{2xf(x/\theta)}{\theta\theta^*(\theta+\theta^*)}\int tw_n(t)\,dt,$$

где $\theta^*(\theta+\theta^*) \underset{P_\theta}{\longrightarrow} 2\theta^2$. Поэтому

$$\sup_x |W_n(x)| \Rightarrow \sup_x \left|w^0\left(F\left(\frac{x}{\theta}\right)\right) + \theta^{-3}xf\left(\frac{x}{\theta}\right)\int tw^0\left(F\left(\frac{t}{\theta}\right)\right)dt\right| =$$

$$= \sup_x \left|w^0\left(F\left(\frac{x}{\theta}\right)\right) + \frac{x}{\theta}f\left(\frac{x}{\theta}\right)\int tw^0(F(t))\,dt\right|.$$

Так как преобразование сжатия по x под знаком $\sup\limits_x$ ничего не меняет, то теорема 2 доказана.

Аналогичные результаты читатель может получить и для статистик

$$\int \left(F_n^*(x) - F_{\theta^*}(x)\right) dF_{\theta^*}(x).$$

§ 58. Устойчивость статистических решений (робастность)

1. Постановка задачи. Качественная и количественная характеризации робастности. При построении различных статистических процедур в предыдущих параграфах — в задачах оценивания или проверки гипотез — мы исходили каждый раз из некоторой совокупности условий. Эти условия касались, в частности, независимости наблюдений и их одинаковой распределенности, предположений о характере распределения $\mathbf{P}$ элементов выборки. Невыполнение этих условий делало бы соответствующие утверждения (например, о характере предельного распределения или об оптимальности той или иной статистики), вообще говоря, неверными.

С другой стороны, на практике обсуждаемые условия есть, как правило, результат аппроксимации и неизбежной идеализации. Стало быть, в точном виде эти условия обычно не выполнены, и возникают опасения, связанные с правомочностью рекомендаций, сделанных с помощью той или иной выбранной статистической процедуры.

Следовательно, как и в любой другой области математики, касающейся приложений, здесь необходимо на последней стадии перед применениями разработанных методов выяснить, как велики должны быть отклонения от принятых условий, чтобы это заставило нас изменить сформулированные выводы.

С математической точки зрения это есть задача очень близкая к проблеме устойчивости. В англоязычной литературе для этого круга задач принят термин «robustness» *). В связи с этим в русской литературе наряду с «устойчивостью» используется также слово «робастность».

Робастность статистических процедур в широком понимании есть малая чувствительность этих процедур к малым отклонениям от исходных предположений.

Наиболее распространенные отклонения от упомянутых выше условий состоят в следующем.

1. Распределение x_i не в точности равно $\mathbf{P}$, а лишь приближенно.

2. В серии наблюдений X присутствует малая доля «выбросов», т. е. наблюдений, вызванных грубыми ошибками измерений или записи или порожденных каким-то другим «мешающим» механизмом, отличным от исследуемой системы. Отделить эти наблюдения, как правило, невозможно. Вместо этого ищут процедуры, которые были бы мало чувствительны к таким «засорениям» выборки.

3. Элементы выборки X не являются независимыми, а лишь слабо зависимы.

Задача состоит в том, чтобы построить решающие правила для основных задач математической статистики, которые были бы близки по своей эффективности к оптимальным и в то же время были бы нечувствительны к перечисленным отклонениям от принятых условий или хотя бы к тем из них, которые для нас существенны. Эту задачу, которая является весьма трудной и не всегда точно поставленной, нельзя считать изученной в полном объеме. Результаты здесь носят пока еще разрозненный характер.

Остановимся несколько подробнее на названных видах отклонений от исходных предположений. В задачах из практики эти отклонения могут

*) Robust — крепкий, здоровый.

присутствовать, конечно, все одновременно. Здесь мы рассмотрим их влияние по отдельности.

1. Пусть на самом деле выборка X принадлежит не распределению F_0, а распределению $F_\varepsilon = (1-\varepsilon)F_0 + \varepsilon F_1$, где ε мало. Мы будем рассматривать оценки подстановки неизвестного параметра θ (мы видели в главе 2, что все разумные оценки являются таковыми). Другими словами, если θ есть значение функционала $G(F_0)$ от распределения F_0, то мы будем пользоваться в качестве оценки значением $\theta^* = G(F_n^*)$, где F_n^* — эмпирическая функция распределения, построенная по выборке X.

Возможны как *качественная*, так и *количественная* характеризации робастности таких оценок.

а) Качественная робастность. Оценку $\theta^* = G(F_n^*)$ мы будем называть *робастной* (относительно замены F_0 на близкое распределение F_ε), если функционал $G(F)$ непрерывен в точке F_0 относительно слабой сходимости распределений (слабая сходимость, по-видимому, лучше других описывает интуитивное понимание близости распределений). Если считать для простоты, что функция распределения F_0 непрерывна, то слабая сходимость будет эквивалентна равномерной сходимости. *Все статистики (оценки) второго типа, рассмотренные в* § 3, 8, *являются робастными оценками.* Если $\theta^* = G(F_n^*)$ есть статистика второго типа и $X \Subset F_\varepsilon = (1-\varepsilon)F_0 + \varepsilon F_1$, то $G(F_n^*)$ будет близко при больших n к значению $G((1-\varepsilon)F_0 + \varepsilon F_1)$, которое, в свою очередь, будет близко при малых ε и любом F_1 к $G(F_0) = \theta$.

Статистики первого типа (см. § 3) не являются, вообще говоря, робастными оценками. Пусть, например, $\theta = G(F_0) = \int g(x)\,dF_0(x) < \infty$, $\int g(x)\,dF_1(x) = \infty$, $X \Subset F_\varepsilon = (1-\varepsilon)F_0 + \varepsilon F_1$. Тогда $\theta^* = G(F_n^*)$ будет близко при больших n (см. § 7) к $(1-\varepsilon)\int g(x)\,dF_0(x) + \varepsilon \int g(x)\,dF_1(x) = \infty$.

b) Количественная характеризация робастности. Степень близости $G(F_n^*)$ для $X \Subset F_\varepsilon$ и $\theta = G(F_0)$ можно характеризовать и количественно. Для этого нам придется предполагать, что функционал G является достаточно «гладким». Именно, мы будем считать, что функционал G дифференцируем (см. определение 8.1 и замечания в конце § 8), что означает в нашем случае существование

$$\lim_{\varepsilon\to 0} \frac{G(F_\varepsilon) - G(F_0)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{G(F_0 + \varepsilon(F_1 - F_0)) - G(F_0)}{\varepsilon} = G'(F_0, F_1 - F_0).$$

Тогда

$$G(F_n^*) - G(F_0) = \big(G(F_n^*) - G(F_\varepsilon)\big) + \big(G(F_\varepsilon) - G(F_0)\big). \tag{1}$$

Первое слагаемое здесь согласно § 8 можно представить приближенно при больших n в виде

$$\frac{1}{\sqrt{n}} G'(F_\varepsilon, w^n F_\varepsilon)$$

(это в свою очередь близко к $n^{-1/2}G'(F_0, w^0F_0)$ при наличии свойств непрерывности G'). Второе слагаемое в (1) при малых ε близко к

$$\varepsilon G'(F_0, F_1 - F_0). \tag{2}$$

Из сказанного видно, что суммарная погрешность в (1) имеет тот же порядок малости, что и

$$n^{-1/2}G'(F_0, w^0F_0) + \varepsilon G'(F_0, F_1 - F_0).$$

Здесь к задаче робастности имеет отношение лишь второе слагаемое, которое и характеризует робастность оценки количественно. Если оно мало по сравнению с $n^{-1/2}$, то влиянием отклонений в предположениях о распределении X можно пренебречь. Заметим, что малость (2) может быть обусловлена и близостью F_1 к F_0, так как $G'(F_0, F_1 - F_0) = 0$ при $F_1 = F_0$. Это имеет место, например, когда некоторые элементы выборки известны лишь приближенно (огрублены) и интервалы приближения малы. Если произведение (2) не мало по сравнению с $n^{-1/2}$, а F_1 и ε известны, то значение $\varepsilon G'\bigl(F_0^*, F_1 - F_n^*\bigr)$ можно использовать как «поправку» к $\theta^* = G\bigl(F_n^*\bigr)$, уменьшающую смещение оценки.

При оценке параметра θ по выборке X из параметрического семейства $\{\mathbf{P}_\theta\}$ следует проявлять осторожность в связи с тем, что значение $G(F_\varepsilon)$ при $F_\varepsilon \notin \{\mathbf{P}_\theta\}$ не обязано иметь смысл параметра этого семейства.

c) Устойчивость статистик. Возможны и несколько иные по сравнению с п. a), b) подходы к задачам робастности, не связанные с методом подстановки и анализом функции $G\bigl(F_n^*\bigr)$, где $G(F_0) = \theta$ есть оцениваемый параметр. Пусть в некоторой статистической процедуре используется статистика $S(X)$. Тогда анализ робастности этой процедуры можно проводить, используя свойства регулярности функции S. Будем предполагать здесь, что:

(C_1) Статистика $S(X)$, $X \Subset F_0$, асимптотически нормальна:

$$\sigma^{-1}\sqrt{n}\,(S(X) - \theta) \Rightarrow \Phi_{0,1}$$

(в правой части здесь может стоять и любое другое распределение).

(C_2) Для любого $\varepsilon > 0$ найдется множество Ω_ε и постоянная L_ε такие, что

$$\mathbf{P}_0(\Omega_\varepsilon) > 1 - \varepsilon, \quad |S(X) - S(Z)| \leqslant \frac{L_\varepsilon}{n} \sum |\mathrm{x}_i - \mathrm{z}_i|,$$

где $Z = (\mathrm{z}_1, \ldots, \mathrm{z}_n)$, $\mathrm{z}_i \Subset F_i$.

Предположим теперь, что F_i и F_0 обладают свойствами:

(C_3) Функции F_i непрерывны,

$$\bigl|F_0^{-1}(F_0(t)(1 + \Delta)) - t\bigr| < c|\Delta|^\beta, \quad \beta > 0.$$

(C_4) Справедливо

$$F_i(x) = F_0(x)(1 + \varepsilon_i(x)),$$

где $\varepsilon_i(x) = \varepsilon_{i,n}(x) \to 0$ при $n \to \infty$ таким образом, что

$$n^{-1/2} \sum_{i=1}^{n} |\varepsilon_{i,n}(z_i)|^\beta \xrightarrow[P]{} 0.$$

Таким образом, мы имеем дело здесь со схемой серий, когда $F_i = F_{i,n}$ зависят от n.

Лемма 1. *Пусть* $x_i \in F_0$, $z_i \in F_i$ *и выполнены условия* (C_1)–(C_4). *Тогда выборки* X *и* Z *можно так задать на одном вероятностном пространстве, что* $(S(X) - S(Z))\sqrt{n} \xrightarrow[P]{} 0$. *Это означает, что наряду с условием* (C_1) *справедливо также*

$$\sigma^{-1}\sqrt{n}\,(S(Z) - \theta) \Rightarrow \Phi_{0,1}. \tag{3}$$

Сходимость (3) означает, что статистические выводы, основанные на статистике S в предположении $x_i \in F_0$, сохраняются и в случае $x_i \in F_i$.

Применение этого подхода проиллюстрировано в задаче об оценке хвостов распределения в § 68.

Отметим, что все статистики I типа $S(X) = h\left(\frac{1}{n} g(x_i)\right)$ удовлетворяют условиям (C_1), (C_2) при подходящих условиях на h, g. Например, если $\mathbf{P}(\overline{g}(X) \in [a,b]) \to 1$, $\mathbf{P}(\overline{g}(Z) \in [a,b]) \to 1$ при $n \to \infty$, где $\overline{g}(X) = \frac{1}{n}\sum g(x_i)$, и если h непрерывно дифференцируема на $[a,b]$, g удовлетворяет условию Липшица, то на множестве $\Omega = \{\overline{g}(X) \in [a,b], \overline{g}(Z) \in [a,b]\}$ получаем

$$|S(X) - S(Z)| = |h(\overline{g}(X)) - h(\overline{g}(Z))| = |h'(\widetilde{g})(\overline{g}(X) - \overline{g}(Z))| \leqslant$$
$$\leqslant |h'(\widetilde{g})| n^{-1} \sum |g(x_i) - g(z_i)| \leqslant |h'(\tilde{g})| n^{-1} c \sum |x_i - z_i|,$$

где $\widetilde{g} \in [a,b]$.

Доказательство леммы 1. Пусть $z_i \in F_i$. Так как F_i непрерывны, то

$$x_i = F_0^{-1}(F_i(z_i)) \in F_0.$$

Далее,

$$|x_i - z_i| = \left|F_0^{-1}(F_i(z_i)) - z_i\right| = \left|F_0^{-1}(F_0(z_i))(1 + \varepsilon_i(z_i)) - z_i\right| \leqslant c|\varepsilon_i(z_i)|^\beta.$$

Для любых $\varepsilon > 0$, $\gamma > 0$ имеем

$$\mathbf{P}(\sqrt{n}\,|S(X) - S(Z)| > \gamma) \leqslant \varepsilon + \mathbf{P}(\sqrt{n}\,|S(X) - S(Z)| > \gamma, X \in \Omega_\varepsilon) \leqslant$$
$$\leqslant \varepsilon + \mathbf{P}\left(L_\varepsilon n^{-1/2} \sum |x_i - z_i| > \gamma\right) \leqslant \varepsilon + \mathbf{P}\left(n^{-1/2} \sum |\varepsilon_i(z_i)|^\beta > \frac{\gamma}{cL_\varepsilon}\right).$$

Правая часть здесь выбором ε и n может быть сделана сколь угодно малой. Лемма доказана.

2. Рассмотрим теперь второй тип отклонений от предположения $X \in F_0$, когда известно, что доля $\varepsilon = \varepsilon_n = m/n > 0$ элементов выборки $X \in F_0$ заменена на значения $y_1, \ldots, y_m$, $m = \varepsilon n$, относительно которых ничего не известно. Другими словами, мы располагаем выборкой $\widetilde{X}_n$, состоящей из выборки X_{n-m} и выборки $Y = (y_1, \ldots, y_m)$, которую выделить из $\widetilde{X}_n$ невозможно. Эмпирическую функцию распределения $\widetilde{F}_n^*$ для выборки $\widetilde{X}_n$ можно представить в виде $\widetilde{F}_n^* = (1-\varepsilon)F_{n-m}^* + \varepsilon F_{(Y)}^*$ при очевидном толковании обозначения $F_{(Y)}^*$. Наша задача по-прежнему состоит в оценке $\theta = G(F_0)$. Имеем согласно (6.6)

$$
\begin{aligned}
G\big(\widetilde{F}_n^*\big) &= G\big((1-\varepsilon)F_{n-m}^* + \varepsilon F_{(Y)}^*\big) = \\
&= G\left((1-\varepsilon)F_0 + \frac{1-\varepsilon}{\sqrt{n-m}} w^{n-m}(F_0) + \varepsilon F_{(Y)}^*\right) = \\
&= G\left(F_0 + \frac{1-\varepsilon}{\sqrt{n-m}} w^{n-m}(F_0) + \varepsilon\big(F_{(Y)}^* - F_0\big)\right) \approx \\
&\approx G(F_0) + b_n G'\left(F_0, \frac{1-\varepsilon}{\sqrt{n-m}\, b_n} w^{n-m}(F_0) + \frac{\varepsilon}{b_n}\big(F_{(Y)}^* - F_0\big)\right),
\end{aligned}
$$

где предполагается, что

$$
b_n = \max\left(\varepsilon, \frac{1-\varepsilon}{\sqrt{n-m}}\right) = \frac{1}{n}\max\big(m, \sqrt{n-m}\,\big) \to 0
$$

при $n \to \infty$. Здесь опять важно сравнение величин ε и $(n-m)^{-1/2}$, но не только это, поскольку в рассматриваемой ситуации нам ничего не известно относительно $F_{(Y)}^*$.

Пусть, например, $G'(F_0, v)$ есть линейный функционал относительно v. Тогда при $b_n = \sqrt{n-m}/n$ (при $m < \sqrt{n-m}$) имеем

$$
\begin{aligned}
G'\left(F_0, \frac{1-\varepsilon}{\sqrt{n-m}\, b_n} w^{n-m}(F_0) + \frac{\varepsilon}{b_n}\big(F_{(Y)}^* - F_0\big)\right) &= \\
= G_1(F_0, w^{n-m}(F_0)) + \frac{1}{\sqrt{n-m}} \sum_{k=1}^{m} G'\big(F_0, F_{(y_k)}^* - F_0\big), \quad (4)
\end{aligned}
$$

где $F_{(y)}^*(x) = 0$ при $x \leqslant y$, $F_{(y)}^*(x) = 1$ при $x > y$. Влияние второго слагаемого в (4) будет мало, если $G'\big(F_0, F_{(y)}^* - F_0\big)$ равномерно ограничено по y. Однако это не всегда имеет место. Например, если

$$
G'(F_0, v) = \int l(x)\, dv(x),
$$

то

$$
G'\big(F_0, F_{(y)}^* - F_0\big) = l(y) - \int l(x)\, dF_0(x). \quad (5)
$$

Это значение может быть сколь угодно большим, если функция $l(y)$ не ограничена (считаем, что $\int l(x)\,dF_0(x) < \infty$).

Отсюда видно, что функция от y

$$I(F_0, y) = G'\bigl(F_0, F^*_{(y)} - F_0\bigr)$$

играет важную роль для характеризации робастности функционала G. Ее называют *функцией влияния* (влияния наблюдения y на изменение функционала $G\bigl(\widetilde{F}^*_n\bigr)$; подробнее об этом см., например, в [122, 127–130]).

Итак, второе слагаемое в (4) будет иметь порядок $m(n-m)^{-1/2}$ при всех Y, если $\sup\limits_y I(F_0, y) < \infty$ и, следовательно, при $m = o(\sqrt{n}\,)$ будет пренебрежимо малым по сравнению с первым слагаемым, которое описывает «естественный» разброс значений $G\bigl(F^*_{n-m}\bigr)$ около значения θ. В этом случае оценка $G\bigl(\widetilde{F}^*_n\bigr)$ будет робастной.

3. Рассмотрим теперь третий тип отклонений от предположения $X \in F_0$, состоящий в том, что наблюдения $\mathrm{x}_i \in F_0$ не являются независимыми. Чтобы рассматривать задачу устойчивости относительно этого допущения, мы должны предполагать, что зависимость между наблюдениями x_i и x_{i+k} ослабевает с ростом k или равномерно мала при всех k. Остановимся на первом варианте и будем характеризовать ослабление зависимости весьма грубо через убывание коэффициентов корреляции. В отличие от п. 1, 2 здесь робастными оказываются статистики обоих типов, рассмотренных в § 3. Как мы увидим, это связано с тем, что закон больших чисел и центральная предельная теорема устойчивы относительно появления небольшой зависимости между наблюдениями.

Пусть $\theta^* = G\bigl(F^*_n\bigr)$, G есть функционал I типа $G(F) = h\left(\int g(x)\,dF(x)\right)$, где функция h непрерывна в точке $g_0 = \mathbf{E}g(\mathrm{x}_i) = \int g(x)\,dF_0(x)$. Положим

$$\Delta_i = g(\mathrm{x}_i) - g_0, \quad \rho(k) = \mathbf{E}\Delta_i\Delta_{i+k},$$

так что $\rho(0) = \mathbf{D}g(\mathrm{x}_i)$, и будем считать, что $\rho(k)$ от i не зависят.

Теорема 1. *Если*

$$0 < \rho(0) + 2\sum_{k=1}^{\infty}\rho(k) = \sigma^2 < \infty, \tag{6}$$

то для выборки X из зависимых наблюдений $\mathrm{x}_i \in F_0$ выполняется $G\bigl(F^*_n\bigr) \underset{P}{\longrightarrow} h(g_0) = G(F_0)$.

Доказательство. Имеем

$$\mathbf{E}\left(\sum_{i=1}^{n}\Delta_i\right)^2 = n\rho(0) + 2\sum_{k=1}^{n-1}\sum_{i=1}^{n-k}\mathbf{E}\Delta_i\Delta_{i+k} =$$

$$= n\rho(0) + 2\sum_{k=1}^{n-1}(n-k)\rho(k) = n\sigma^2 - 2\sum_{k=1}^{n-1}k\rho(k) + r_n,$$

где $r_n = o(1)$,

$$\sum_{k=1}^{n-1}k\rho(k) = \sum_{k=1}^{n-1}\sum_{j=k}^{n-1}\rho(j) = o(n)$$

при $n \to \infty$. Поэтому для эмпирической функции распределения F_n^*, построенной по выборке из зависимых наблюдений, в силу неравенства Чебышева будет выполняться

$$\mathbf{P}\left(\left|\int g(x)\,dF_n^*(x) - g_0\right| > \varepsilon\right) \leqslant \varepsilon^{-2}n^{-2}\mathbf{E}\left(\sum_{i=1}^{n}\Delta_i\right)^2 =$$

$$= \varepsilon^{-2}n^{-1}(\sigma^2 + o(1)) \to 0$$

при $n \to \infty$. Остается воспользоваться теоремой непрерывности.

Аналогично обстоит дело с функционалами II типа. Обозначим через $I_A(x)$ индикатор множества A, $g(x) = I_{(-\infty,z)}(x)$.

Теорема 2. *Пусть для введенной функции g выполнено* (6) *при любом z. Тогда если функционал G непрерывен в равномерной метрике в точке F_0, то* $G\big(F_n^*\big) \underset{P}{\longrightarrow} G(F_0)$.

Доказательство. Нам достаточно доказать, что для выборки X из зависимых наблюдений справедливо

$$\sup_x\big|F_n^*(x) - F_0(x)\big| \underset{P}{\longrightarrow} 0.$$

Это следует из того, что в силу теоремы 1 $F_n^*(z) \underset{P}{\longrightarrow} F_0(z)$ при любом z и из рассуждений, аналогичных доказательству теоремы 2.2. $\triangleleft$

Нетрудно провести и количественную характеризацию робастности в условиях этого параграфа. Для статистик I типа предельное распределение $\sqrt{n}\left(\int g(x)\,dF_n^*(x) - g_0\right)$ в широких предположениях будет нормальным Φ_{0,σ^2}, так что влияние зависимости будет осуществляться лишь через параметр σ^2, определенный в (6). Таким образом, появление зависимости наблюдений сказывается лишь на степени разброса θ^* около θ (для независимых x_i выполняется $\sigma^2 = \rho(0)$). Для статистики II типа эта картина в целом сохраняется, хотя изучение предельного распределения $G\big(F_n^*\big)$ для

функционалов второго типа и зависимых наблюдений выглядит сложнее (качественно все результаты § 8 в этом случае сохраняются).

Рассмотрим теперь несколько примеров. Первый из них будет связан главным образом с анализом влияния выбросов (второй тип отклонений от предположения $X \in F_0$).

2. Оценка параметра сдвига. Пусть по выборке $X \in \mathbf{P}_\alpha$, $\mathbf{P}_\alpha(A) = \mathbf{P}(A - \alpha)$ оценивается параметр сдвига α. Мы рассмотрим здесь два типа оценок: M-оценки (как мы видели в § 14; они ведут себя в известном смысле как аддитивные статистики) и L-оценки (основанные на порядковых статистиках; см. § 8).

При рассмотрении M-оценок наиболее интересными являются два случая, когда функция $g(\mathrm{x}_i, \alpha)$, участвующая в построении M-оценок (см. § 14), имеет вид $g(x, \alpha) = g(x)$ (метод моментов) и $g(x, \alpha) = g(x - \alpha)$. В первом случае истинное значение α_0 параметра α определяется как решение уравнения

$$a(\alpha) \equiv \int g(x)\mathbf{P}_\alpha(dx) = \int g(x)\mathbf{P}_{\alpha_0}(dx),$$

так что M-оценка α^* в этом случае (являясь оценкой подстановки) имеет вид

$$\alpha^* = a^{-1}\left(\int g(x)\, dF_n^*(x)\right). \tag{7}$$

Во втором случае для $g(x, \alpha) = g(x - \alpha)$ получим уравнение (считаем для простоты, что $0 \in \Theta$)

$$\int g(x - \alpha)\mathbf{P}_{\alpha_0}(dx) = \int g(x - \theta)\mathbf{P}_\alpha(dx) \equiv$$

$$\equiv \int g(x - \alpha)\mathbf{P}(dx - \alpha) \equiv \int g(x)\mathbf{P}(dx) = a(0).$$

Таким образом, если решение уравнения

$$\int g(x - \alpha)\, dF(x) = a(0)$$

обозначить через $G(F)$, то M-оценка α^* в этом случае будет иметь вид $\alpha^* = G(F_n^*)$. К этому второму случаю относятся, очевидно, и о. м. п. Достаточно положить $g(x - \alpha) = l'(x, \alpha)$, где $l(x, \alpha) = \ln f_\alpha(x) = \ln f(x - \alpha)$.

Если $g(x) \equiv x$, то $a(\alpha) = a(0) + \alpha$ и оценки α^* в обоих рассматриваемых случаях будут совпадать:

$$\alpha^* = \overline{\mathrm{x}} - a(0).$$

Очевидно, эта оценка является а. н. оценкой с параметрами $(0, \sigma^2)$, $\sigma^2 = \int x^2\mathbf{P}(dx)$.

Для $g(x) = x$, как и для произвольной неограниченной функции g, в обоих случаях M-оценки не являются робастными относительно присутствия выбросов. В первом случае это видно непосредственно из формы

оценки (см. (7)), во втором — из того, что производная функционала G (см. § 14) равна

$$G'(F_{\alpha_0}, v) = -b^{-1}\int g(x-\alpha_0)\,dv(x), \quad b = \int g'(x)\mathbf{P}(dx),$$

так что в этом последнем случае в обозначениях (5)

$$l(y) = -b^{-1}g(x-\alpha_0).$$

Чтобы сделать рассматриваемые M-оценки робастными, можно вместо g использовать ограниченную функцию

$$g_N(x) = \min[\max(g(x), -N), N].$$

Следует отметить, что при этом вычисления функции $a(\theta)$ и функционала $G(F)$ могут усложниться, а коэффициент разброса оценок увеличиться (напомним, что для ряда задач линейная функция $g(x) = x$ дает эффективные оценки). Число N следует выбирать исходя из специфики задачи, а также в зависимости от n и от величин выбросов (они не превосходят крайних значений выборки). В связи со сказанным отметим также, что сам класс M-оценок был введен в [127] как такое расширение класса о. м. п., в котором можно искать робастные оценки, не слишком теряя при этом в эффективности.

Рассмотрим теперь L-оценки (см. § 8), связанные с функционалом (см. (8.2))

$$G(F) = \int_0^1 \varphi(t)g(F^{-1}(t))\,dt - b, \tag{8}$$

при $g(x) \equiv x$, $\varphi(t) = \dfrac{1}{1-2\gamma}I_{(\gamma,1-\gamma)}(t)$, где $\gamma \in (0, 1/2)$, $I_A(t)$ — индикатор множества A, постоянную b мы выберем так, чтобы $G(F_0) = \theta$. Имеем $F_\theta^{-1}(t) = F_0^{-1}(t) + \theta$ (будем считать для простоты, что $0 \in \Theta$),

$$G(F_\theta) + b = \frac{1}{1-2\gamma}\int_\gamma^{1-\gamma} F_\theta^{-1}(t)\,dt = \theta + \frac{1}{1-2\gamma}\int_\gamma^{1-\gamma} F_0^{-1}(t)\,dt.$$

Поэтому нам следует положить

$$b = \frac{1}{1-2\gamma}\int_\gamma^{1-\gamma} F_0^{-1}(t)\,dt.$$

При таком b мы можем пользоваться оценкой подстановки (L-оценкой)

$$\theta^* = \frac{1}{1-2\gamma}\sum_{k=[\gamma n]+1}^{[(1-\gamma)n]} \mathrm{x}_{(k)} - b. \tag{9}$$

Предположим, что функция распределения $F_\theta(x)$ имеет плотность $F'_\theta(x) = f(x-\theta)$. Тогда согласно § 8 оценка θ^* является а. н. с параметрами $\left(0, \int\limits_0^1 A^2(t)\,dt - A^2\right)$, где

$$A(t) = \int\limits_t^1 a(t)\,dt, \quad A = \int\limits_0^1 A(t)\,dt, \quad a(t) = \frac{\varphi(t)}{f\left(F_{\theta_0}^{-1}(t)\right)}.$$

Производная функционала G здесь равна (см. § 8)

$$G'(F_{\theta_0}, v) = \frac{-}{1} 1 - 2\gamma \int\limits_\gamma^{1-\gamma} \frac{v(F_{\theta_0}(t))}{f\left(F_{\theta_0}^{-1}(t)\right)}\,dt.$$

При $v = F^*_{(y)} - F_{\theta_0}$ получим функцию влияния

$$I(F_{\theta_0}, y) = G'\left(F_{\theta_0}, F^*_{(y)} - F_{\theta_0}\right) =$$

$$= -\frac{1}{1-2\gamma} \int\limits_{\max(\gamma, F_{\theta_0}(y))}^{1-\gamma} \frac{dt}{f\left(F_{\theta_0}^{-1}(t)\right)} + \frac{1}{1-2\gamma} \int\limits_\gamma^{1-\gamma} \frac{t\,dt}{f\left(F_{\theta_0}^{-1}(t)\right)}.$$

Очевидно, что она ограничена по y и, стало быть, оценка (9) является робастной (относительно выбросов; см. пп. 2 из п. 1). Об оптимальности (в известном смысле) выбора функции φ в (8) см. в [66, 102] и § 17.

Если γ сближаются с $1/2$, то оценка (9) сближается со «сдвинутой» выборочной медианой $\zeta^* - b$ $\left(\text{или с оценкой } \frac{1}{2l} \sum\limits_{k=-l}^{l} \mathrm{x}_{([n/2-k])} - b\right)$, которой в (8) соответствует обобщенная δ-функция с единичной массой в точке $t = 1/2$. В нечувствительности ζ^* к появлению выбросов $\mathrm{y}_1, \ldots, \mathrm{y}_m$ (см. пп. 2 из п. 1) можно убедиться и непосредственно. Действительно, ζ^* будет располагаться между значениями $\mathrm{x}_{(k_1)}$ и $\mathrm{x}_{(k_2)}$, где

$$k_1 = \left[\frac{n+1}{2} - m\right], \quad k_2 = \left[\frac{n+1}{2} + m\right],$$

$\mathrm{x}_{(k)}$, $k = 1, \ldots, n-m$, образуют вариационный ряд выборки $X_{n-m} \Subset F_{\theta_0}$. Но асимптотические свойства $\mathrm{x}_{(k_1)}$, $\mathrm{x}_{(k_2)}$ одинаковы и совпадают со свойствами выборочной медианы (см. § 8, 12).

Если $\gamma \to 0$, то оценка (9) сближается с $\overline{\mathrm{x}} - b$. Таким образом, при $\gamma \in (0, 1/2)$ оценки (9) заполняют весь «промежуточный спектр» оценок, лежащий между $\zeta^* - b$ и $\overline{\mathrm{x}} - b$. Если оценка $\overline{\mathrm{x}} - b$ оптимальна, то выбирая $\gamma > 0$ достаточно малым, мы получим робастную оценку, близкую к эффективной. В связи со сказанным отметим следующее. Свойства оценки $\theta^* = \overline{\mathrm{x}}$

мало зависят от изменений $\mathbf{P}$, сохраняющих дисперсию $\sigma_1^2 = \int t^2 f(t)\,dt$ и, в частности, от локальных изменений $f(t)$ в точке $t = 0$. В этом смысле она устойчива. Но *свойство оптимальности* этой оценки, имеющее место для $\mathbf{P} = \mathbf{\Phi}_{\alpha,\sigma_1^2}$, *неустойчиво*. Действительно, пусть при малом $\varepsilon > 0$

$$\mathbf{P} = (1-\varepsilon)\mathbf{\Phi}_{\theta,1} + \varepsilon U_{\theta-\varepsilon,\theta+\varepsilon}.$$

Тогда $f(0) = (1-\varepsilon)/\sqrt{2\pi} + 1/2 > 1/2$ и, как показывают соотношения

$$(\overline{\mathrm{x}} - \theta)\sqrt{n} \Subset \mathbf{\Phi}_{0,\sigma_1^2}, \quad \sigma_1^2 = \int t^2 f(t)\,dt = (1-\varepsilon) + \frac{\varepsilon^3}{6},$$

$$(\zeta^* - \theta)\sqrt{n} \Subset \mathbf{\Phi}_{0,\sigma_2^2}, \quad \sigma_2^2 = \frac{1}{4f^2(0)} = \left(1 + \frac{2(1-\varepsilon)}{\sqrt{2\pi}}\right)^{-2},$$

оценка ζ^* будет лучше, чем $\overline{\mathrm{x}}$ ($\sigma_2 < \sigma_1$), для $\varepsilon < \varepsilon_0$, где $\varepsilon_0 > 0$ легко вычислить.

Сказанное означает также, что распределение статистик ζ^* неустойчиво относительно малых изменений F_{θ_0} в окрестности точки $F_{\theta_0}^{-1}(1/2)$ (значение $f(0)$ в приведенных выше рассмотрениях может при этом сильно меняться).

Соображения об устойчивости оценок параметра сдвига, приведенные в этом пункте, нетрудно переформулировать применительно к статистическим критериям, относящимся к проверке гипотез о параметре сдвига, например к несмещенным р. н. м. к. $|\overline{\mathrm{x}} - \alpha_0| > c$ для проверки по выборке $X \Subset \mathbf{\Phi}_{\alpha,1}$ гипотезы $H_1 = \{\alpha = \alpha_0\}$ против $H_2 = \{|\alpha - \alpha_0| > d > 0\}$.

3. Статистики Стьюдента и S_0^2. Рассмотрим теперь вопрос относительно устойчивости статистических процедур (оценивания и проверки гипотез), связанных со статистиками

$$t = \frac{(\overline{\mathrm{x}} - \alpha)\sqrt{n}}{S_0}, \quad S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{x}_i - \overline{\mathrm{x}})^2.$$

На этих статистиках, как мы знаем (см. § 47, 48), основаны оптимальные критерии соответственно для проверки гипотез относительно среднего α и дисперсии σ^2 нормальных совокупностей в случае, когда второй параметр (σ^2 или α) распределения $\mathbf{\Phi}_{\alpha,\sigma^2}$ неизвестен.

Статистики t и S_0^2 ведут себя по-разному относительно нарушений условия $X \Subset \mathbf{\Phi}_{\alpha,\sigma^2}$. Допустим, что n велико, а $X \Subset \mathbf{P}$, где $\mathbf{P}$ — любое распределение со средним α и конечной дисперсией. Тогда распределение t так же, как и в случае $X \Subset \mathbf{\Phi}_{\alpha,\sigma^2}$, будет сближаться с нормальным распределением $\mathbf{\Phi}_{0,1}$. Это следует из теорем непрерывности (§ 5) и того, что

$$\frac{(\overline{\mathrm{x}} - \alpha)\sqrt{n}}{\sqrt{\mathbf{D}\mathrm{x}_1}} \Subset \mathbf{\Phi}_{0,1}, \quad S_0^2 \xrightarrow[P]{} \mathbf{D}\mathrm{x}_1.$$

Сказанное означает, что размер критерия Стьюдента при больших n будет мало отличаться от заданного, если даже распределение $\mathbf{P}$ выборки X существенно отличается от нормального.

Этого нельзя сказать относительно критериев, построенных по статистике S_0^2. Это обстоятельство связано с тем, что предельное распределение S_0^2 зависит от значения $\mathbf{E}x_i^4$. Действительно, из рассмотрений главы 1 следует, что

$$(S_0^2 - \sigma^2)\sqrt{n} \Rightarrow \Phi_{0,d^2}, \quad d^2 = \mathbf{E}(x_1^2 - \sigma^2)^2 = \mathbf{D}x_1^2.$$

Стало быть, размер критерия, построенного по статистике S_0^2 для нормальной совокупности, может существенно отличаться от заданного, если $X \in \mathbf{P}$ и $\mathbf{P}$ отличается от Φ_{α,σ^2} (разницы не будет при совпадении четвертых моментов $\mathbf{P}$ и Φ_{α,σ^2}).

Обе статистики t и S_0^2 чувствительны к отказу от предположения о независимости наблюдений в выборке X. Если, например, все наблюдения в выборке коррелированы друг с другом и коэффициент корреляции равен ρ, то, принимая, не ограничивая общности, $\alpha = 0$, получим

$$\mathbf{E}S_0^2 = \frac{1}{n-1}\mathbf{E}\left[\sum_{i=1}^n x_i^2 - n(\bar{x})^2\right] = \frac{1}{n-1}\left[n\sigma^2 - \frac{1}{n}\mathbf{E}\left(\sum_{i=1}^n x_i\right)^2\right] =$$

$$= \frac{1}{n-1}[n\sigma^2 - \sigma^2(1-\rho) - n\sigma^2\rho] = \sigma^2(1-\rho).$$

Таким образом, здесь нарушается даже свойство несмещенности S_0^2, хотя при малых ρ отклонение будет небольшим. Отыскание распределений t и S_0^2 при появлении зависимости обычно представляет собой значительные трудности.

4. Критерий отношения правдоподобия. Этот критерий обычно весьма чувствителен к наличию выбросов и даже к малым отклонениям в предположениях о распределении X. Пусть, например, проверяются две простые гипотезы $H_1 = \{X \in \Phi_{0,1}\}$ и $H_2 = \{X \in \mathbf{U}_{-1,1}\}$. Ясно, что при использовании н. м. к. Неймана–Пирсона появление хотя бы одного наблюдения x вне отрезка $[-1, 1]$ при идеальном соответствии остальных наблюдений равномерному распределению $\mathbf{U}_{-1,1}$ заставит нас (с нулевой вероятностью ошибки!) принять гипотезу H_1. Это значит, что наличие хотя бы одного выброса или появление даже малых отклонений от распределения $\mathbf{U}_{-1,1}$ могут вынудить нас принимать ложное решение.

В этом отношении критерий Колмогорова, например, является значительно более устойчивым (хотя и менее мощным относительно H_2). Вообще, непараметрические критерии, как того и следовало ожидать, в целом значительно более устойчивы, чем «индивидуальные» критерии, обладающие свойством оптимальности в той или иной конкретной задаче.

Что касается рассматриваемой задачи о проверке нормальности (H_1) против равномерности (H_2) выборки X, то отыскание мощных и в то же время устойчивых относительно выбросов критериев можно произвести, используя по-прежнему отношение правдоподобия, но для «усеченных» выборок (ср. с (9)). Можно также идти по пути подбора какого-нибудь другого критерия. В этом отношении наличие достаточно большого запаса различных критериев и оценок представляется весьма полезным. Это связано зачастую не только с соображениями устойчивости, но и с вопросами удобства вычислений.

Глава 4

СТАТИСТИЧЕСКИЕ ЗАДАЧИ С ДВУМЯ И БОЛЕЕ ВЫБОРКАМИ

В § 59, 60 рассмотрены задачи об однородности двух выборок.
В § 61 рассмотрены задачи регрессии.
В § 62 излагаются основные результаты дисперсионного анализа.
В § 63 рассмотрены задачи распознавания образов.

§ 59. Проверка гипотез об однородности (полной или частичной) в параметрическом случае

1. Рассматриваемый класс задач. В предыдущих главах основным объектом всех рассмотрений была выборка X объема n из полностью или частично неизвестного распределения $\mathbf{P}$. Мы переходим теперь к статистическим задачам, где фигурирует не одна, а *две* или *более* выборок.

Один из основных классов задач, которые при этом рассматриваются, составляют задачи о проверке *однородности* (полной или частичной) двух выборок.

Сюда относятся следующие основные три типа задач.

I. Проверка «обычной» однородности. Задача здесь состоит в проверке гипотезы о том, что две выборки X и Y извлечены из одного и того же неизвестного распределения. Такие задачи возникают, например, при сравнении двух способов обработки в каком-нибудь технологическом процессе или в сельском хозяйстве. Основой для сравнения обычно служат числовые характеристики конечного продукта (выборки), которые носят случайный характер. Такого же рода задачи возникнут, если мы по состоянию больных будем проверять воздействие нового лекарства, сравнивая подопытную группу пациентов с контрольной.

К задачам об однородности относится пример 3 из введения.

В этом параграфе мы будем рассматривать *параметрический случай.* Пусть дано параметрическое семейство распределений $\{\mathbf{P}_\theta\}_{\theta\in\Theta}$ и даны две независимые выборки $X = (\mathrm{x}_1, \ldots, \mathrm{x}_{n_1})$ и $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_{n_2})$ объемов n_1 и n_2 соответственно, относительно которых заранее известно, что они относятся к семейству $\{\mathbf{P}_\theta\}$:

$$X \Subset \mathbf{P}_{\theta_1}, \quad Y \Subset \mathbf{P}_{\theta_2}, \tag{1}$$

при некоторых θ_1, θ_2. Обычная задача об однородности состоит в проверке гипотезы $H_1 = \{\theta_1 = \theta_2\}$ против дополнительной альтернативы $H_2 = \{\theta_1 \neq \theta_2\}$. Очевидно, что здесь обе гипотезы H_1 и H_2 являются сложными.

II. Проверка однородности при наличии мешающего параметра. Здесь предполагается, что размерность k параметра θ больше

единицы. Запишем вектор θ в виде набора $\theta = (u, v)$ двух «подвекторов» u и v и обозначим через u_j, v_j подвекторы векторов θ_j в (1), $j = 1, 2$.

Пусть нам заранее известно, что «подпараметр» v у обеих выборок хотя и неизвестен, но является общим: $v_1 = v_2 = v$. Проверяется гипотеза $H_1 = \{u_1 = u_2\}$ против $H_2 = \{u_1 \neq u_2\}$.

Это и есть задача об однородности при наличии мешающего параметра v. Она отличается от обычной задачи об однородности тем, что альтернатива для гипотезы $H_1 = \{\theta_1 = \theta_2\}$ имеет вид $H_2 = \{u_1 \neq u_2,\ v_1 = v_2\}$.

Можно привести следующий пример возникновения такого рода задач. Допустим, что нас интересует состояние некоторого объекта, которое характеризуется вектором a. Но измерить a непосредственно нельзя. Нам доступны лишь измерения, при которых на a накладывается случайный шум. Природа шума при разных наблюдениях остается неизменной. Надо проверить гипотезу о неизменности a в двух сериях наблюдений X и Y.

Если, скажем, измерения имеют вид $\mathrm{x}_i = a_1 + \xi_i$, где $\xi_i \in \mathbf{\Phi}_{\lambda,\sigma^2}$ определяют роль шума, и такой же характер имеют наблюдения y_i с заменой a_1 на a_2, то мы можем записать $X \in \mathbf{\Phi}_{a_1+\lambda,\sigma^2}$, $Y \in \mathbf{\Phi}_{a_2+\lambda,\sigma^2}$. Мы пришли к задаче о проверке равенства средних $\{\alpha_1 = \alpha_2\}$ двух нормальных совокупностей $\mathbf{\Phi}_{\alpha_1,\sigma^2}$ и $\mathbf{\Phi}_{\alpha_2,\sigma^2}$ при общем неизвестном значении σ^2.

III. Проверка частичной однородности. Здесь проверяется гипотеза H_1 лишь о «частичном» совпадении θ_1 и θ_2. Именно, проверяется гипотеза $H_1 = \{u_1 = u_2\}$ (в обозначениях предыдущего пункта) против $H_2 = \{u_1 \neq u_2\}$. Значения v_1, v_2 для каждой из выборок X, Y могут быть свои.

Допустим, например, что в лаборатории оценивается результат воздействия нового способа обработки на урожайность какого-нибудь злака. Наблюдения представляют собой суммарный вес зерен в отдельных колосках. Предположим, что $\mathrm{x}_i \in \mathbf{\Phi}_{\alpha_1,\sigma_1^2}$, $i = 1, \ldots, n_1$, для опытной партии колосков и $\mathrm{y}_i \in \mathbf{\Phi}_{\alpha_2,\sigma_2^2}$ для контрольной партии. Естественно допустить, что «разброс» σ^2 в результате изменения обработки мог измениться. Для нас же существенно, изменился ли главный показатель α, определяющий урожайность культуры. Мы приходим к задаче о проверке гипотезы $H_1 = \{\alpha_1 = \alpha_2\}$ против $H_2 = \{\alpha_1 \neq \alpha_2\}$ для нормальных совокупностей, дисперсии которых могут быть разными. Это задача, хорошо известная в литературе под названием *проблемы Беренса–Фишера**).

В этом параграфе мы произведем редукцию всех трех типов задач для произвольных параметрических семейств к задаче, рассмотренной в § 55, о принадлежности одной выборки параметрическому подсемейству и найдем вид асимптотически минимаксных критериев в предположении близо-

*) Поиску ее оптимальных решений посвящена довольно обширная литература (см. библиографию, например, в [65, 66]). Значительный вклад в исследование проблемы Беренса–Фишера, оказавшейся весьма трудной, принадлежит Ю. В. Линнику и его ученикам. Эти исследования требуют введения новых понятий и использования весьма сложного математического аппарата. Это делает невозможным цитирование и доказательство (в рамках настоящей книги) полученных результатов. Положение в задачах об обычной однородности и об однородности при наличии мешающего параметра для нормальных совокупностей несколько лучше (в ряде задач удается найти равномерно наиболее мощные (р. н. м.) инвариантные несмещенные критерии, однако необходимые для этого построения также оказываются довольно сложными; подробнее об этом см. [65, 66]).

сти проверяемых гипотез. Это будут критерии отношения правдоподобия, которые для нормальных совокупностей будут совпадать с критериями, построенными в поисках той или иной *точной* оптимальности (если таковые имеются; ср. с [65, 66]).

Статистический критерий π для проверки H_1 против H_2 в нашем случае будет функцией $\pi = \pi(X, Y)$ двух выборок X и Y и будет, как и ранее, означать вероятность приема H_2 при данной объединенной выборке (X, Y) (см. гл. 3). Определения асимптотического уровня и асимптотической оптимальности критерия π здесь остаются теми же, что и в § 43.

Определение 1. Будем говорить, что критерий π *имеет асимптотический уровень* $1 - \varepsilon$ (принадлежит классу $\widetilde{K}_\varepsilon$), если

$$\limsup_{\substack{n_1\to\infty\\ n_2\to\infty}} \sup_{(\theta_1,\theta_2)\in\overline{\Theta}_1} \mathbf{E}_{\theta_1\theta_2}\pi(X,Y) \leqslant \varepsilon,$$

где $\mathbf{E}_{\theta_1\theta_2}$ означает математическое ожидание по распределению $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$, $\overline{\Theta}_1$ есть множество значений (θ_1, θ_2), при которых выполнена гипотеза H_1 (например, множество всех точек (θ_1, θ_2), лежащих на «биссектрисе» $\theta_1 = \theta_2$ в задаче об обычной однородности).

Определение 2. Критерий $\pi_1 \in \widetilde{K}_\varepsilon$ называется *асимптотически минимаксным в* $\widetilde{K}_\varepsilon$ для проверки H_1 против H_2, если для любого критерия $\pi \in \widetilde{K}_\varepsilon$ выполняется

$$\liminf_{\substack{n_1\to\infty\\ n_2\to\infty}} \Bigl(\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_2} \mathbf{E}_{\theta_1\theta_2}\pi_1(X,Y) - \inf_{(\theta_1,\theta_2)\in\overline{\Theta}_2} \mathbf{E}_{\theta_1\theta_2}\pi(X,Y) \Bigr) \geqslant 0,$$

где $\overline{\Theta}_2$ есть множество значений (θ_1, θ_2), соответствующих альтернативам из H_2.

2. Асимптотически минимаксный критерий для проверки близких гипотез об обычной однородности. Введем новый параметр $\overline{\theta} = (\theta_1, \theta_2)$, характеризующий «объединенную» выборку (X, Y). Функция правдоподобия этой выборки равна $f_{\overline{\theta}}(X, Y) = f_{\theta_1}(X) f_{\theta_2}(Y)$.

Допустим для простоты, что объемы выборок совпадают: $n_1 = n_2 = n$. Тогда выборку (X, Y) можно представлять себе как выборку объема n, образованную наблюдениями $(x_1, y_1), \ldots, (x_n, y_n)$ из распределения $\mathbf{P}_{\overline{\theta}} = \mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2}$, имеющего плотность $f_{\theta_1}(x) f_{\theta_2}(y)$. Мы приходим к задаче, рассмотренной в § 55, о проверке по выборке (X, Y) гипотезы H_1 о том, что параметр $\overline{\theta}$ располагается на «кривой» $\theta_1 = \theta_2$. В обозначениях § 55 гипотеза H_1 в нашем случае имеет вид $\overline{\theta} = g(\alpha)$, где $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha, \alpha)$. Очевидно, что матрица $G = \left\| \dfrac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots, 2k$; $j = 1, \ldots, k$, имеет вид $\begin{pmatrix} E \\ E \end{pmatrix}$, где E — единичная матрица k-го порядка, так что ранг G равен k.

Будем считать параметр $\overline{\theta}$ локализованным, т. е. будем считать, что значения θ_1 и θ_2 близки и, стало быть, возможные значения $\overline{\theta}$ располагаются в окрестности точки $\overline{\theta}_0 = (\theta_0, \theta_0)$ при некотором фиксированном θ_0.

Если следовать § 55, то нам будет удобнее ввести новый параметр $\tau = (\tau', \tau'') = (\gamma'/\sqrt{n}, \gamma''/\sqrt{n}) = \gamma/\sqrt{n}$, где $\tau' = \theta_1 - \theta_0$, $\tau'' = \theta_2 - \theta_1$, так что отображение $\overline{\theta} = \overline{\theta}(\tau)$ взаимно однозначно: $\theta_1 = \tau' + \theta_0$, $\theta_2 = \tau'' + \tau' + \theta_0$. В терминах параметров τ и γ гипотеза H_1 об однородности примет вид $H_1 = \{\tau'' = 0\} = \{\gamma'' = 0\}$. В качестве альтернативы мы рассмотрим «отделенную» гипотезу

$$H_2^b = \{\gamma'' I \gamma''^{\top} \geqslant b^2\}, \quad b > 0, \tag{2}$$

где $I = I(\theta_0)$ есть матрица Фишера для семейства $\{\mathbf{P}_\theta\}$ в точке θ_0.

Теорема 1. *Пусть в окрестности точки θ_0 семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям* (RR) (см. § 36). *Тогда критерий отношения правдоподобия*

$$R_1(X,Y) = \frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{3}$$

является асимптотически минимаксным критерием асимптотического уровня $1-\varepsilon$ для проверки $H_1 = \{\theta_1 = \theta_2\}$ против $H_2^b = \{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^{\top} \geqslant b^2/n\}$ при любом $b > 0$, где h_ε есть квантиль порядка $1-\varepsilon$ распределения χ^2 с k степенями свободы (такое предельное распределение при гипотезе H_1 имеет статистика $2 \ln R_1(X,Y)$).

Пусть $\widehat{\theta}_X^*$, $\widehat{\theta}_Y^*$, $\widehat{\theta}^*$ есть о. м. п. параметра $\theta = \theta_1 = \theta_2$ соответственно по выборкам X, Y, (X,Y). Тогда критерий

$$(\widehat{\theta}_X^* - \widehat{\theta}^*) I(\widehat{\theta}^*) (\widehat{\theta}_X^* - \widehat{\theta}^*)^{\top} + (\widehat{\theta}_Y^* - \widehat{\theta}^*) I(\widehat{\theta}^*) (\widehat{\theta}_Y^* - \widehat{\theta}^*)^{\top} > \frac{h_\varepsilon}{n} \tag{4}$$

будет асимптотически эквивалентным критерию (3).

Доказательство. Приведенное утверждение является прямым следствием теоремы 55.4. Нам надо лишь выяснить, что собой представляет матрица Фишера $\overline{I}(\overline{\theta}_0) = \overline{I}(\theta_0, \theta_0)$ для «объединенного» параметра $\overline{\theta} = (\theta_1, \theta_2)$ и матрица M_2 для параметрического семейства $\{\mathbf{P}_{(\theta_0, \theta_0+\beta)}\}$ в точке $\beta = 0$. Имеем

$$\ln f_{\theta_1}(x) f_{\theta_2}(y) = l(x, \theta_1) + l(y, \theta_2).$$

Обозначим через t_i, $i = 1, \ldots, 2k$, координаты вектора $\overline{\theta}$. Тогда если через $\mathbf{E}_{\overline{\theta}}$ обозначить математическое ожидание по распределению $\mathbf{P}_{\overline{\theta}}$, то элементы $\overline{I}_{ij}(\overline{\theta})$ матрицы $\overline{I}(\overline{\theta})$ будут равны

$$\overline{I}_{ij}(\overline{\theta}) = \mathbf{E}_{\overline{\theta}} \left(\frac{\partial l(\mathrm{x}_1, \theta_1)}{\partial t_i} + \frac{\partial l(\mathrm{y}_1, \theta_2)}{\partial t_i} \right) \left(\frac{\partial l(\mathrm{x}_1, \theta_1)}{\partial t_j} + \frac{\partial l(\mathrm{y}_1, \theta_2)}{\partial t_j} \right).$$

Отсюда в силу независимости x_1 и y_1 получаем

$$\overline{I}(\overline{\theta}) = \begin{pmatrix} I(\theta_1) & 0 \\ 0 & I(\theta_2) \end{pmatrix}.$$

Поэтому критерий (4) есть не что иное, как критерий (55.12) в теореме 55.4.

Аналогичные вычисления показывают, что $M_2 = I(\theta_0)$, так как при $\beta = (\beta_1, \ldots, \beta_k) = 0$

$$\frac{\partial l(\mathrm{x}_1, \theta_0)}{\partial \beta_i} + \frac{\partial l(\mathrm{y}_1, \theta_0 + \beta)}{\partial \beta_i} = \frac{\partial l(\mathrm{y}_1, \theta_0)}{\partial t_i}, \quad i = 1, \ldots, k. \ \triangleleft$$

Замечание 1. Утверждение теоремы 1 мы получили в предположении, что $n_1 = n_2$. Однако это ограничение совершенно несущественно. Рассмотрим, например, случай, когда $n_1 \to \infty$, $n_2 \to \infty$ так, что отношение n_1/n_2 равно рациональному числу r_1/r_2 (r_1, r_2 — произвольные целые фиксированные числа, $n_i = nr_i$, $n \to \infty$). Мы введем опять новый параметр $\overline{\theta} = (\theta_1, \theta_2)$ и будем рассматривать объединенную выборку (X, Y) как выборку объема n с наблюдениями $(\mathrm{x}_1, \ldots, \mathrm{x}_{r_1}; \mathrm{y}_1, \ldots, \mathrm{y}_{r_2})$, $(\mathrm{x}_{r_1+1}, \ldots, \mathrm{x}_{2r_1}; \mathrm{y}_{r_2+1}, \ldots, \mathrm{y}_{2r_2}), \ldots$ из распределения

$$\mathbf{P}_{\overline{\theta}} = \underbrace{\mathbf{P}_{\theta_1} \times \cdots \times \mathbf{P}_{\theta_1}}_{r_1 \text{ раз}} \times \underbrace{\mathbf{P}_{\theta_2} \times \cdots \times \mathbf{P}_{\theta_2}}_{r_2 \text{ раз}},$$

зависящего от параметра $\overline{\theta}$. Функция правдоподобия здесь вновь будет иметь вид

$$f_{\overline{\theta}}(X, Y) = f_{\theta_1}(X) f_{\theta_2}(Y).$$

Если ввести, как и прежде, новый параметр $\tau = (\tau', \tau'') = (\theta_1 - \theta_0,\ \theta_2 - \theta_1)$ и положить $\tau = \gamma/\sqrt{n} = (\gamma'/\sqrt{n},\ \gamma''/\sqrt{n})$, то рассматриваемая задача состоит в проверке $H_1 = \{\gamma'' = 0\}$ против $H_2^b = \{\gamma'' M_2 \gamma''^{\top} \geqslant b^2\}$, где M_2 — матрица Фишера для $\mathbf{P}_{(\theta_0, \theta_0+\beta)}$ в точке $\beta = 0$. Легко видеть, что в нашем случае $M_2 = r_2 I(\theta_0)$, так что множество альтернатив сохраняет свою форму (2):

$$H_2^b = \left\{ \gamma'' I \gamma''^{\top} > \frac{b^2}{r^2} \right\}.$$

Матрица Фишера $\overline{I}(\overline{\theta})$ будет иметь вид

$$\begin{pmatrix} r_1 I(\theta_1) & 0 \\ 0 & r_2 I(\theta_2) \end{pmatrix}.$$

Остается воспользоваться теоремой 55.4. Получим тогда утверждение теоремы 1, в которой критерий (4) следует заменить на

$$n_1 \big(\widehat{\theta}_X^* - \theta^*\big) I(\theta^*) \big(\widehat{\theta}_X^* - \theta^*\big)^{\top} + n_2 \big(\widehat{\theta}_Y^* - \theta^*\big) I(\theta^*) \big(\widehat{\theta}_Y^* - \theta^*\big) > h_\varepsilon. \qquad (5)$$

С помощью теоремы 55.4 можно указать также гарантированную асимптотическую мощность критериев (3)–(5).

Утверждение теоремы остается справедливым и в общем случае, когда $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to c$, где c — произвольное число из $(0,1)$. Однако доказательство этого факта требует дополнительных рассмотрений.

Замечание 2. Утверждение теоремы 1 сохранится, если гипотезу $H_1 = \{\theta_1 = \theta_2\}$ заменить гипотезой (см. § 55)

$$H_1^a = \left\{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^\top \leqslant \frac{a^2}{n}\right\}, \quad 0 < a < b.$$

Замечание 3. Форма асимптотически минимаксных критериев в теореме 1 от θ_0 не зависит. Значение θ_0 входит только в определение гипотезы H_2^b через $I = I(\theta_0)$ (см. (2)), хотя и этого появления θ_0 можно было бы избежать, заменив I в (2) на $(\theta_1 + \theta_2)/2$. Это дало бы нам «асимптотически эквивалентную» H_2^b гипотезу $\widetilde{H}_2^b$, для которой утверждение теоремы 3 полностью сохраняется. Появление в (2) значения θ_0 есть следствие использования наиболее простого пути редукции рассматриваемой задачи к результатам § 55.

Пример 1. Пусть X и Y есть выборки объемов n_1, n_2 из полиномиальных распределений $X \in \mathbf{B}_{\theta_1}$, $Y \in \mathbf{B}_{\theta_2}$, $\theta_i \in \mathbb{R}^k$, $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})$, $i = 1, 2$. Векторы частот $\nu = (\nu_1, \ldots, \nu_k)$ и $\mu = (\mu_1, \ldots, \mu_k)$ появлений событий $A_1, \ldots, A_k$ (см. § 12) образуют достаточные статистики,

$$f_{\theta_1}(X) = \prod_{i=1}^k \theta_{1i}^{\nu_i}, \quad f_{\theta_2}(Y) = \prod_{i=1}^k \theta_{2i}^{\mu_i}.$$

О. м. п. имеют вид $\widehat{\theta}_X^* = \nu/n_1$, $\widehat{\theta}_Y^* = \mu/n_2$, $\widehat{\theta}^* = (\nu + \mu)/(n_1 + n_2)$. Матрица $I(\theta)$ определена в (56.5), так что (см. (56.9))

$$tI(\theta_0)t^\top = \sum_{i=1}^k \frac{t_i^2}{\theta_{0i}}.$$

Итак, в силу теоремы 1 и замечания 1 асимптотически минимаксный критерий асимптотического уровня $1 - \varepsilon$ для проверки $H_1 = \{\theta_1 = \theta_2\}$ против

$$H_2^b = \left\{\sum_{i=1}^k \frac{(\theta_{1i} - \theta_{2i})^2}{\theta_{0i}} \geqslant \frac{b^2}{n_2}\right\}$$

имеет вид

$$\ln R_1(X, Y) = \sum_{i=1}^k \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^k \mu_i \ln \frac{\mu_i}{n_2} - \sum (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2},$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с $k-1$ степенями свободы. Согласно (4), (5) асимптотически эквивалентным будет критерий

$$n_1\sum_{i=1}^{k}\left(\frac{\nu_i}{n_1}-\frac{\nu_i+\mu_i}{n_1+n_2}\right)^2\frac{n_1+n_2}{\nu_i+\mu_i}+n_2\sum_{i=1}^{k}\left(\frac{\mu_i}{n_2}-\frac{\nu_i+\mu_i}{n_1+n_2}\right)^2\frac{n_1+n_2}{\nu_i+\mu_i}=$$

$$=\sum_{i=1}^{k}\left(\frac{\nu_i}{n_1}-\frac{\mu_i}{n_2}\right)^2\frac{n_1n_2}{\nu_i+\mu_i}>h_\varepsilon. \quad (6)$$

Пример 1А. В примере 35.3 был описан механизм наследования групп крови, которые были обозначены как 0 (ноль), А, В, АВ. Он управляется с помощью генов трех типов А, В, 0. Вероятности появления этих генов в данной популяции обозначим соответственно через p, q, $r=1-p-q$. Вероятности $p_i(\alpha)$, $\alpha=(p,q)$, того, что человек будет иметь i-ю группу крови, выражаются через α по формулам, приведенным в § 35, табл. 1.

Нам даны две выборки X и Y с частотами ν_i и μ_i, $i=1,\ldots,4$, появления i-й группы крови, полученные в результате обследования $n_1=353$ из общины I и $n_2=364$ человек из общины II. Распределение людей по группам крови приведено в табл. 1.

Таблица 1

	0	А	В	АВ	Всего
Община I	121	120	79	33	353
Община II	118	95	121	30	364
Всего	239	215	200	63	717

Надо проверить гипотезу о принадлежности обследуемых общин одной популяции, т. е. гипотезу о равенстве вероятностей p и q для этих групп или, что то же, о равенстве вероятностей $p_i(\alpha)$. Это есть, очевидно, задача об однородности, рассмотренная в примере 1.

Если проверять совпадение вероятностей четырех групп крови, то статистике (см. (6))

$$\chi_1^2=\sum_{i=1}^{4}\left(\frac{\nu_i}{n_1}-\frac{\mu_i}{n_2}\right)^2\frac{n_1n_2}{\nu_i+\mu_i}$$

будет соответствовать распределение χ^2 с тремя степенями свободы. В нашем случае значение χ_1^2 составляет 11,74. Фактически достигаемый (см. § 44) уровень полученного уклонения превосходит 0,99. Это значит, что гипотезу об однородности следует забраковать с точки зрения критерия $\chi_1^2>h_{0,01}$ уровня 0,99.

Отметим, что примененный критерий не полностью отвечает природе рассмотренного явления, так как мы должны проверять совпадение вероятностей p и q, а не вероятностей p_i появлений групп крови. Если следовать в точности теореме 1, то мы должны методами § 35 вычислить о. м. п. $\widehat{\alpha}_X^*$, $\widehat{\alpha}_Y^*$ и $\widehat{\alpha}^*$ параметра $\alpha=(p,q)$ соответственно по выборкам X, Y, (X,Y) и

воспользоваться статистикой

$$\chi_2^2 = 2\big[L\big(\widehat{\alpha}_X^*, X\big) + L\big(\widehat{\alpha}_Y^*, Y\big) - L(\widehat{\alpha}^*, (X,Y))\big] =$$
$$= 2\left[\sum_{i=1}^{4} \nu_i \ln p_i\big(\widehat{\alpha}_X^*\big) + \sum_{i=1}^{4} \mu_i \ln p_i\big(\widehat{\alpha}_Y^*\big) - \sum_{i=1}^{4} (\nu_i + \mu_i) \ln p_i(\widehat{\alpha}^*)\right],$$

имеющей при больших n распределение, близкое к распределению χ^2 с двумя степенями свободы. Если провести необходимые вычисления (см. пример 35.3), то мы получим $\chi_2^2 \approx 11{,}04$, что для двух степеней свободы дает более значимое уклонение, чем 11,74 для трех.

О проверке самой гипотезы о принадлежности X и Y параметрическим подсемействам $\mathbf{B}_{p(\alpha)}$, где $p(\alpha) = (p_1(\alpha), \ldots, p_4(\alpha))$, смотри пример 57.1. Обе выборки хорошо согласуются с этой гипотезой.

Пример 2. Пусть $X \in \mathbf{\Phi}_{\alpha_1,\sigma_1^2}$, $Y \in \mathbf{\Phi}_{\alpha_2,\sigma_2^2}$, где точки $\theta_i = \big(\alpha_i, \sigma_i^2\big)$ располагаются в окрестности точки $\theta_0 = \big(\alpha_0, \sigma_0^2\big)$. Здесь

$$I(\theta_0) = \begin{pmatrix} \sigma_0^{-2} & 0 \\ 0 & \dfrac{1}{2}\sigma_0^{-4} \end{pmatrix}$$

(см. § 26), и мы будем рассматривать задачу о проверке гипотезы $H_1 = \{\theta_1 = \theta_2\}$ против

$$H_2^b = \left\{ \frac{(\alpha_1 - \alpha_2)^2}{\sigma_0^2} + \frac{\big(\sigma_1^2 - \sigma_2^2\big)^2}{2\sigma_0^4} \geqslant \frac{b^2}{n_2} \right\}, \quad n = n_1 + n_2.$$

Имеем $\widehat{\theta}_X^* = \big(\overline{\mathrm{x}}, S_X^2\big)$, $S_X^2 = \dfrac{1}{n_1}\sum\limits_{i=1}^{n_1} (\mathrm{x}_i - \overline{\mathrm{x}})^2$, $f_{\widehat{\theta}_X^*}(X) = \big(2\pi e S_X^2\big)^{-n_1/2}$. Аналогичные формулы справедливы для выборки Y. Далее,

$$\widehat{\theta}^* = \big(\overline{\mathrm{z}}, S_{X,Y}^2\big), \quad \overline{\mathrm{z}} = \frac{\left(\sum\limits_{i=1}^{n_1} \mathrm{x}_i + \sum\limits_{i=1}^{n_2} \mathrm{y}_i\right)}{n_1 + n_2} = a\overline{\mathrm{x}} + (1-a)\overline{\mathrm{y}},$$

$$S_{X,Y}^2 = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1} (\mathrm{x}_i - \overline{\mathrm{z}})^2 + \sum_{i=1}^{n_2} (\mathrm{y}_i - \overline{\mathrm{z}})^2\right] =$$
$$= aS_X^2 + (1-a)S_Y^2 + (1-a)a(\overline{\mathrm{x}} - \overline{\mathrm{y}})^2, \quad (7)$$

где $a = n_1/(n_1 + n_2)$, $f_{\widehat{\theta}^*}(X) f_{\widehat{\theta}^*}(Y) = \big(2\pi e S_{X,Y}^2\big)^{-(n_1+n_2)/2}$.

Таким образом, асимптотически минимаксным для проверки H_1 против H_2^b будет критерий

$$\frac{S_{X,Y}^2}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)},$$

где h_ε — квантиль распределения χ^2 с двумя степенями свободы. Мы предлагаем читателю в виде упражнения найти асимптотически эквивалентный критерий вида (5).

3. Асимптотически минимаксные критерии для задачи об однородности при наличии мешающего параметра. В этом и последующих пунктах мы будем считать для простоты, что объемы выборок X и Y совпадают: $n_1 = n_2$. Это ограничение несущественно. В случае $n_1/n_2 = r_1/r_2$ (r_1, r_2 целые) читатель может избавиться от него самостоятельно так же, как это было сделано в замечании 1 к теореме 1.

Итак, пусть даны две выборки $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$, объемов $n_1 = n_2 = n$. Проверяется гипотеза $\{u_1 = u_2\}$ против $\{u_1 \neq u_2\}$ в предположении, что $v_1 = v_2 = v$ и v нам неизвестно. Размерность u_i обозначим через r, $r < k$.

Введем новый параметр $\overline{\theta} = (u_1, u_2, v)$. Объединенную выборку (X, Y) представим как выборку объема n с наблюдениями $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_n, \mathrm{y}_n)$, плотность распределения которых равна $f_{\overline{\theta}}(x, y) = f_{(u_1,v)}(x) f_{(u_2,v)}(y)$. Для этого параметрического семейства рассматриваемая нами задача эквивалентна задаче о проверке гипотезы H_1, состоящей в том, что значение $\overline{\theta}$ лежит на «кривой» $\overline{\theta} = g(\theta_1) = (u_1, u_1, v)$, против дополнительной альтернативы. Матрица $G = \left\| \dfrac{\partial g_i}{\partial \theta_{1j}} \right\|$, $i = 1, \ldots, k+r$; $j = 1, \ldots, k$, имеет форму $\begin{pmatrix} E_r & 0 \\ & E_k \end{pmatrix}$, где сверху стоит единичная матрица порядка r, а внизу — единичная матрица порядка k, так что ранг G равен k.

Как и в предыдущем пункте, будем считать параметр θ локализованным около точки $\theta_0 = (u_0, v_0)$. Введем параметр $\tau = \tau(\overline{\theta}) = (\tau', \tau'', \tau''') = (u_1 - u_0, u_2 - u_1, v - v_0)$. Обратное отображение $\overline{\theta} = \overline{\theta}(\tau)$ всегда существует и имеет координаты $u_1 = \tau' + u_0$, $u_2 = \tau'' + \tau' + u_0$, $v = \tau''' + v_0$. Положим $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''')$.

Для нового параметра τ (или γ) гипотеза об однородности имеет вид $H_1 = \{\gamma'' = 0\}$. В качестве альтернативы рассмотрим «отделенную» гипотезу $H_2^b = \{\gamma'' I_1(\theta_0) \gamma''^{\top} \geqslant b^2\}$, где $I_1(\theta)$ есть подматрица исходной информационной матрицы Фишера $I(\theta)$, образованная ее первыми r строками и столбцами.

Теорема 2. *Пусть в окрестности точки θ_0 семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям* (RR). *Тогда критерий отношения правдоподобия*

$$R_1(X, Y) = \frac{\sup\limits_{(u_1,u_2,v)} f_{(u_1,v)}(X) f_{(u_2,v)}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{8}$$

является асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ для проверки $H_1 = \{u_1 = u_2\}$ против

$$H_2^b = \left\{ (u_1 - u_2) I_1(\theta_0) (u_1 - u_2)^{\top} \geqslant \frac{b^2}{n} \right\} \tag{9}$$

при общем значении $v_1 = v_2 = v$ *и при любом* $b > 0$. *Здесь* h_ε — *квантиль порядка* $1 - \varepsilon$ *распределения* χ^2 *с* r *степенями свободы.* (*Таким будет предельное распределение* $2 \ln R_1(X, Y)$ *при гипотезе* H_1.)

Обозначим через $\overline{\theta}^*$ *значение параметра* $\overline{\theta}$, *при котором достигается максимум числителя в* (8), *и* $\theta^* = (u^*, v^*)$ — *значение* θ, *при котором достигается максимум знаменателя. Представим матрицу* $I(\theta)$ *в виде*

$$I(\theta) = \begin{pmatrix} I_1(\theta) & I_{21}(\theta) \\ I_{12}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

Тогда критерий

$$(\overline{\theta}^* - (u^*, u^*, v^*))\overline{I}(\overline{\theta}^*)(\overline{\theta}^* - (u^*, u^*, v^*))^\top > \frac{h_\varepsilon}{n}, \tag{10}$$

где

$$\overline{I}(\overline{\theta}) = \begin{pmatrix} I_1(\theta_1) & 0 & I_{21}(\theta_1) \\ 0 & I_1(\theta_2) & I_{21}(\theta_2) \\ I_{12}(\theta_1) & I_{12}(\theta_2) & I_{22}(\theta_1) + I_{22}(\theta_2) \end{pmatrix}, \tag{11}$$

будет асимптотически эквивалентным (8).

Доказательство. Эта теорема также является прямым следствием теоремы 55.4. Нам надо лишь выяснить структуру матрицы $\overline{I}(\overline{\theta})$ для выборки (X, Y) и «объединенного» параметра $\overline{\theta}$ и матрицы M_2. Имеем

$$l \equiv \ln f_{\overline{\theta}}(x, y) = l(x, (u_1, v)) + l(y, (u_2, v)).$$

Обозначим через t_i, $i = 1, \ldots, k + r$, координаты вектора $\overline{\theta}$. Тогда

$$\frac{\partial l}{\partial t_i} = \begin{cases} \dfrac{\partial l(x, (u_1, v))}{\partial t_i}, & 0 < i \leqslant r, \\ \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & r < i \leqslant 2r, \\ \dfrac{\partial l(x, (u_1, v))}{\partial t_i} + \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & 2r < i \leqslant k + r; \end{cases}$$

отсюда без труда получается (11).

Матрица M_2 для параметрического семейства $\mathbf{P}_{\overline{\theta}(0,\beta,0)} = \mathbf{P}_{(u_0, u_0+\beta, v_0)}$ в точке $\beta = 0$ вычисляется аналогично, равна $I_1(\theta_0)$ и соответствует средней подматрице матрицы $\overline{I}(\overline{\theta}_0)$. $\triangleleft$

В примерах объемы выборок n_1 и n_2 мы будем считать произвольными.

Пример 3. Пусть $X \in \mathbf{\Phi}_{\alpha_1, \sigma^2}$, $Y \in \mathbf{\Phi}_{\alpha_2, \sigma^2}$. Надо проверить гипотезу $H_1 = \{\alpha_1 = \alpha_2\}$, когда σ^2 неизвестно. Для отыскания с помощью теоремы 2 асимптотически минимаксных критериев нам надо найти статистику $R_1(X, Y)$ в (8), где в нашем случае $u_i = \alpha_i$, $v = \sigma^2$, $\overline{\theta} = (\alpha_1, \alpha_2, \sigma^2)$.

Имеем

$$\ln f_{(\alpha_1,\sigma^2)}(X) f_{(\alpha_2,\sigma^2)}(Y) = \\ = -\frac{1}{2}(n_1+n_2)\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(\mathrm{x}_i-\alpha_1)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n_2}(\mathrm{y}_i-\alpha_2)^2.$$

Обращая в нуль производные этой функции по α_1, α_2 и σ^2 и решая полученные уравнения, находим (в обозначениях примера 2)

$$\overline{\theta}^* = \left(\overline{\mathrm{x}}, \overline{\mathrm{y}}, aS_X^2 + (1-a)S_Y^2\right), \quad a = \frac{n_1}{n_1+n_2}, \\ f_{\overline{\theta}^*}(X,Y) = \left[2\pi e\left(aS_X^2 + (1-a)S_Y^2\right)\right]^{-(n_1+n_2)/2}. \tag{12}$$

Поступая таким же образом с функцией

$$\ln f_\theta(X) f_\theta(Y) = \ln f_{(\alpha,\sigma^2)}(X) f_{(\alpha,\sigma^2)}(Y),$$

получим (см. пример 2)

$$\theta^* = \left(\overline{z}, S_{X,Y}^2\right), \\ f_{\theta^*}(X) f_{\theta^*}(Y) = \left(2\pi e S_{X,Y}^2\right)^{-(n_1+n_2)/2}. \tag{13}$$

Таким образом, асимптотически оптимальный критерий имеет вид

$$\frac{S_{X,Y}^2}{aS_X^2 + (1-a)S_Y^2} > e^{h_\varepsilon/(n_1+n_2)}$$

или (см. (7))

$$\frac{\sqrt{a(1-a)}\,|\overline{\mathrm{x}} - \overline{\mathrm{y}}|}{\sqrt{aS_X^2 + (1-a)S_Y^2}} > \sqrt{\frac{h_\varepsilon}{n_1+n_2}},$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с одной степенью свободы, так что $\sqrt{h_\varepsilon}$ можно заменить значением $\lambda_{\varepsilon/2}$, для которого $\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1-\varepsilon$. Нетрудно видеть, что левая часть неравенства

$$\frac{\sqrt{a(1-a)(n_1+n_2)}\,|\overline{\mathrm{x}} - \overline{\mathrm{y}}|}{\sqrt{aS_X^2 + (1-a)S_Y^2}} > \lambda_{\varepsilon/2}, \tag{14}$$

определяющего асимптотически минимаксный критерий, после замены $|\overline{\mathrm{x}} - \overline{\mathrm{y}}|$ на $\overline{\mathrm{x}} - \overline{\mathrm{y}}$ будет асимптотически нормальной с параметрами $(0,1)$ случайной величиной.

Но этот критерий можно сделать точным (т. е. имеющим в точности наперед заданный уровень). Действительно, в силу результатов § 40 при

гипотезе H_1

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\frac{\overline{\mathrm{x}}-\overline{\mathrm{y}}}{\sigma} \in \mathbf{\Phi}_{0,1},$$

$$\frac{(n_1+n_2)aS_X^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(\mathrm{x}_i-\overline{\mathrm{x}})^2 \in \mathbf{H}_{n_1-1},$$

$$\frac{(n_1+n_2)(1-a)S_Y^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n_2}(\mathrm{y}_i-\overline{\mathrm{y}})^2 \in \mathbf{H}_{n_2-1}.$$

Так как все три случайные величины независимы, то отношение

$$|\overline{\mathrm{x}}-\overline{\mathrm{y}}|\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\frac{n_1+n_2}{n_1+n_2-2}\left(aS_X^2+(1-a)S_Y^2\right)\right]^{-1/2} =$$

$$= \frac{(\overline{\mathrm{x}}-\overline{\mathrm{y}})\sqrt{a(1-a)(n_1+n_2-2)}}{\sqrt{aS_X^2+(1-a)S_Y^2}} \in \mathbf{T}_{n_1+n_2-2}$$

имеет распределение Стьюдента с n_1+n_2-2 степенями свободы. Таким образом, критерий (ср. с (14))

$$\frac{(\overline{\mathrm{x}}-\overline{\mathrm{y}})\sqrt{a(1-a)(n_1+n_2-2)}}{\sqrt{aS_X^2+(1-a)S_Y^2}} > \tau_\varepsilon,$$

где τ_ε такое, что $\mathbf{T}_{n_1+n_2-2}(-\tau_\varepsilon,\tau_\varepsilon) = 1-\varepsilon$, будет иметь уровень значимости, в точности равный $1-\varepsilon$, и его можно использовать при любых (а не только больших) значениях n_1, n_2. Этот критерий называется критерием Стьюдента. Он обладает также некоторыми свойствами точной (а не только асимптотической) оптимальности (см. [65, 66]).

Пример 4. Пусть $X \in \mathbf{\Phi}_{\alpha,\sigma_1^2}$, $Y \in \mathbf{\Phi}_{\alpha,\sigma_2^2}$. Проверяется гипотеза $\{\sigma_1{=}\sigma_2\}$, когда α неизвестно. Поступая аналогично предыдущему, придем к значению R_1 в (8), у которого знаменатель тот же, что и в предыдущем примере, а числитель равен

$$\sup_{(\alpha,\sigma_1,\sigma_2)} f_{(\alpha,\sigma_1^2)}(X) f_{(\alpha,\sigma_2^2)}(Y). \tag{15}$$

Выписывая уравнения для точки максимума, получим

$$\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(\mathrm{x}_i-\alpha)^2 = S_X^2+(\overline{\mathrm{x}}-\alpha)^2,$$

$$\sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(\mathrm{y}_i-\alpha)^2 = S_Y^2+(\overline{\mathrm{y}}-\alpha)^2,$$

$$\frac{n_1}{\sigma_1^2}(\overline{\mathrm{x}}-\alpha)+\frac{n_2}{\sigma_2^2}(\overline{\mathrm{y}}-\alpha) = 0.$$

Отсюда, полагая

$$p = \frac{a}{\sigma_1^2} \cdot \frac{1}{a/\sigma_1^2 + (1-a)/\sigma_2^2} \in (0,1), \tag{16}$$

находим

$$\alpha = p\overline{x} + (1-p)\overline{y},$$

$$\sigma_1^2 = S_X^2 + (1-p)^2\Delta^2, \quad \sigma_2^2 = S_Y^2 + p^2\Delta^2,$$

где мы для краткости положили $\Delta = \overline{x} - \overline{y}$; p можно рассматривать как решение уравнения (16) или

$$p = \frac{a\left(S_Y^2 + p^2\Delta^2\right)}{a\left(S_Y^2 + p^2\Delta^2\right) + (1-a)\left(S_X^2 + (1-p)^2\Delta^2\right)}.$$

Так как максимум в (15) равен

$$(2\pi e)^{-(n_1+n_2)/2}\left(S_X^2 + (1-p)^2\Delta^2\right)^{-n_1/2}\left(S_Y^2 + p^2\Delta^2\right)^{-n_2/2}, \tag{17}$$

то, сравнивая это с (13), (7), мы получим асимптотически минимаксный критерий

$$\frac{aS_X^2 + (1-a)S_Y^2 + a(1-a)\Delta^2}{\left(S_X^2 + (1-p)^2\Delta^2\right)^a\left(S_Y^2 + p^2\Delta^2\right)^{1-a}} > e^{h_\varepsilon/(n_1+n_2)} \tag{18}$$

или

$$\frac{aS_X^2 + (1-a)S_Y^2}{S_X^{2a}S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}A^{-1}, \tag{19}$$

где

$$A = \frac{1 + \dfrac{a(1-a)\Delta^2}{aS_X^2 + (1-a)S_Y^2}}{\left(1 + (1-p)^2\Delta^2/S_X^2\right)^a\left(1 + p^2\Delta^2/S_Y^2\right)^{1-a}},$$

h_ε есть квантиль распределения χ^2 с одной степенью свободы. Здесь $\Delta^2 = \left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)\xi^2$, $\xi \Subset \Phi_{0,1}$, $S_X^2/\sigma_1^2 \xrightarrow[P]{} 1$, $S_Y^2/\sigma_2^2 \xrightarrow[P]{} 1$, $\sigma_1^2/\sigma_2^2 \to 1$, $p \to a$ (можно считать для простоты, что $a = n_1/(n_1+n_2)$ фиксировано), $\ln A \xrightarrow[P]{} 0$ для каждой из рассматриваемых близких гипотез. Стало быть, правая часть в (19) имеет вид

$$1 + \frac{h_\varepsilon + \gamma_n}{n_1+n_2}, \quad \gamma_n \xrightarrow[P]{} 0.$$

Левая часть (19) представляет собой отношение среднего арифметического к среднему геометрическому величин S_X^2 и S_Y^2. Если обозначить $S_X^2/S_Y^2 = Z^2$, то неравенство, обратное к (19), можно записать в виде

$$\frac{aZ^2 + (1-a)}{Z^{2a}} - 1 \leqslant \frac{h_\varepsilon + \gamma_n}{n_1+n_2}. \tag{20}$$

В левой части здесь стоит выпуклая вниз функция от Z (можно для определенности считать $a \leqslant 1/2$), имеющая в точке $Z = 1$ кратный нуль. Так как правая часть этого неравенства мала, то решение удобно искать в виде $Z^2 = 1 + \zeta$ при малом ζ. Пользуясь разложением в ряд по степеням ζ и отбрасывая члены третьего и более высоких порядков малости, мы получим для границ ζ_1, ζ_2 интервала, в котором справедливо (20), значения

$$\zeta_1 = \sqrt{\frac{2(h_\varepsilon + \gamma_n')}{a(1-a)(n_1+n_2)}}, \quad \zeta_2 = \sqrt{\frac{2(h_\varepsilon + \gamma_n'')}{a(1-a)(n_1+n_2)}},$$

$$\gamma_n' \underset{P}{\longrightarrow} 0, \quad \gamma_n'' \underset{P}{\longrightarrow} 0.$$

Это означает, если вернуться к исходным переменным, что область

$$\sqrt{\frac{1}{2}a(1-a)(n_1+n_2)}\left|\frac{S_X^2}{S_Y^2} - 1\right| > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{21}$$

(λ_ε определено в примере 3) определяет критерий, асимптотически эквивалентный (18) и, стало быть, асимптотически минимаксный.

Здесь, как и в примере 3, мы можем сделать полученный критерий точным, поскольку нам известно точное распределение статистики S_X^2/S_Y^2. Действительно,

$$\frac{n_1 S_X^2}{\sigma_1^2} \in \mathbf{H}_{n_1-1}, \quad \frac{n_2 S_Y^2}{\sigma_2^2} \in \mathbf{H}_{n_2-1}$$

и при гипотезе $H_1 = \{\sigma_1 = \sigma_2\}$

$$\frac{n_1 S_X^2}{n_2 S_Y^2} \in \mathbf{F}_{n_1-1,n_2-1},$$

где $\mathbf{F}_{n_1-1,n_2-1}$ есть распределение, введенное нами в § 12 и табулированное в руководствах по математической статистике. Напомним (см. § 12, п. 6), что если пользоваться таблицами распределения Фишера, то надо для отыскания квантилей f_ε распределения $\mathbf{F}_{k_1,k_2}$ порядка $1-\varepsilon$ вводить поправочный множитель. Если $\varkappa_\varepsilon$ — квантиль порядка $1-\varepsilon$ распределения Фишера с (k_1, k_2) степенями свободы, то $f_\varepsilon = \dfrac{k_2}{k_1}\varkappa_\varepsilon$.

Сказанное означает, что можно вычислить точный уровень значимости критерия (21) и применять его при любых n_1 и n_2 (о свойствах точной оптимальности этого критерия см. [65, 66]). Если n_1 и n_2 велики, то левая часть в (21) (без знака абсолютной величины) асимптотически нормальна с параметрами $(0, 1)$.

4. Асимптотически минимаксный критерий для задачи о частичной однородности. Пусть $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$. Проверяется гипотеза $\{u_1 = u_2\}$ против $\{u_1 \neq u_2\}$, когда значения v_1 и v_2 в выборках X и Y могут быть любыми. Размерность u_i, как и прежде, обозначим через r, $r < k$.

Введем новый параметр $\overline{\theta} = (\theta_1, \theta_2) = (u_1, v_1, u_2, v_2)$ размерности $2k$. Как и раньше, выборку (X, Y) (при $n_1 = n_2 = n$) представим как выборку с наблюдениями $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_n, \mathrm{y}_n)$, имеющими плотность

$$f_{\overline{\theta}}(x, y) = f_{(u_1, v_1)}(x) f_{(u_2, v_2)}(y).$$

Для этого семейства исходная задача о частичной однородности эквивалентна задаче о проверке гипотезы H_1, состоящей в том, что $\overline{\theta}$ лежит на «кривой» $\theta = g(\alpha) = (u_1, v_1, u_1, v_2)$, где $\alpha = (u_1, v_1, v_2)$ есть «подпараметр» размерности $2k - r$. Мы предлагаем читателю, следуя рассуждениям двух предыдущих разделов, выписать матрицу $G = \left\| \dfrac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots, 2k$; $j = 1, \ldots, 2k - r$. Ее ранг будет равен $2k - r$.

Как и в п. 2, 3, мы будем считать задачу «локализованной» около точки $\theta_0 = (u_0, v_0)$.

Введем наряду с $\overline{\theta}$ параметр $\tau = \tau(\overline{\theta}) = (\tau', \tau'', \tau''', \tau^{\mathrm{IV}}) = (u_1 - u_0, v_1 - v_0, u_2 - u_1, v_2 - v_0)$. Обратное преобразование $\overline{\theta} = \overline{\theta}(\tau)$ имеет координаты

$$u_1 = \tau' + u_0, \quad v_1 = \tau'' + v_0, \quad u_2 = \tau''' + \tau' + u_0, \quad v_2 = \tau^{\mathrm{IV}} + v_0.$$

Если положить $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''', \gamma^{\mathrm{IV}})$, то гипотеза H_1 будет иметь вид $H_1 = \{\gamma''' = 0\}$. В качестве альтернативы будем рассматривать «отделенную» гипотезу $H_2^b = \{\gamma''' I_1(\theta_0) \gamma'''^{\top} \geqslant b^2\}$, где $I_1(\theta)$ имеет тот же смысл, что и в теореме 2.

Теорема 3. *Пусть в окрестности точки θ_0 семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям* (RR). *Тогда критерий отношения правдоподобия*

$$R_1(X, Y) = \frac{\sup\limits_{(\theta_1, \theta_2)} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{(u, v_1, v_2)} f_{(u, v_1)}(X) f_{(u, v_2)}(Y)} > e^{h_\varepsilon/2} \tag{22}$$

является асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ для проверки H_1 против гипотезы H_2^b, определенной в (9), *при произвольных значениях v_1 и v_2. Значение h_ε здесь то же, что и в теореме* 2.

Доказательство этой теоремы повторяет рассуждения предыдущих разделов и тоже полностью основано на теореме 55.4. Отыскание информационной матрицы Фишера $\overline{I}(\overline{\theta})$ для параметра $\overline{\theta}$ и матрицы M_2 для семейства с плотностью $f_{\overline{\theta}(0,0,\beta,0)} = f_{(u_0, v_0, u_0 + \beta, v_0)}$ в точке $\beta = 0$ мы предоставляем читателю.

С помощью матрицы $\overline{I}\big((\widehat{\theta}_X^*, \widehat{\theta}_Y^*)\big)$ и вектора $(\widehat{\theta}_X^*, \widehat{\theta}_Y^*) - (u^*, v_1^*, u^*, v_2^*)$, где $(\widehat{\theta}_X^*, \widehat{\theta}_Y^*)$ и (u^*, v_1^*, v_2^*) суть векторы, в которых достигаются максимумы числителя и знаменателя в (22), мы можем, как и прежде, с помощью теоремы 55.4 (см. (55.12)) построить асимптотически эквивалентный критерий, использующий квадратичную форму от введенных оценок. $\triangleleft$

Пример 5. Сравнение дисперсий нормальных совокупностей. Пусть $X \in \mathbf{\Phi}_{\alpha_1,\sigma_1^2}$, $Y \in \mathbf{\Phi}_{\alpha_2,\sigma_2^2}$, $H_1 = \{\sigma_1 = \sigma_2\}$. Здесь вычисления значительно проще, чем в примере 4, поскольку значение числителя в (22) (как и вектор $(\widehat{\theta}_X^*, \widehat{\theta}_Y^*) = (\overline{\mathrm{x}}, S_X^2, \overline{\mathrm{y}}, S_Y^2)$) нам известно, а значение знаменателя найдено в примере 3 (см. (12)). Неравенство (22) здесь будет иметь вид

$$\frac{aS_X^2 + (1-a)S_Y^2}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}.$$

Сравнивая это с (19) и с последующими рассмотрениями, мы придем к тем же критериям и выводам, что и в примере 4.

Пример 6. Проблема Беренса–Фишера о сравнении двух нормальных совокупностей. Пусть

$$X \in \mathbf{\Phi}_{\alpha_1,\sigma_1^2}, \quad Y \in \mathbf{\Phi}_{\alpha_2,\sigma_2^2}, \quad H_1 = \{\alpha_1 = \alpha_2\},$$

значения σ_1 и σ_2 произвольны. Для этого примера числитель в (22) тот же, что и в предыдущем примере, а знаменатель найден в примере 4 (см. (17); там это был числитель для (8)).

Асимптотически минимаксный критерий, стало быть, будет иметь вид

$$\left(\frac{S_X^2 + (1-p)^2\Delta^2}{S_X^2}\right)^a \left(\frac{S_Y^2 + p^2\Delta^2}{S_Y^2}\right)^{1-a} > e^{h_\varepsilon/(n_1+n_2)}; \tag{23}$$

здесь $\Delta = \overline{\mathrm{x}} - \overline{\mathrm{y}}$ представимо в виде

$$\Delta = (\alpha_1 - \alpha_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,\xi, \quad \xi \in \mathbf{\Phi}_{0,1},$$

$$\frac{S_X^2}{\sigma_1^2} \underset{P}{\longrightarrow} 1, \quad \frac{S_Y^2}{\sigma_2^2} \underset{P}{\longrightarrow} 1,$$

так что $\Delta \underset{P}{\longrightarrow} 0$ при гипотезе H_1. Это соотношение сохранится, очевидно, и для каждой из близких альтернатив. Чтобы найти более простой по форме критерий, асимптотически эквивалентный (23), выделим в обеих частях неравенства (23) главные части. Получим

$$\frac{a(1-p)^2\Delta^2}{S_X^2} + \frac{(1-a)p^2\Delta^2}{S_Y^2} + \Delta^4\rho_n > \frac{h_\varepsilon}{n_1+n_2} + O\left(\frac{1}{(n_1+n_2)^2}\right),$$

где $\rho_n \underset{P}{\longrightarrow} \rho = \mathrm{const}$. Учитывая, что

$$p = \frac{aS_Y^2}{aS_Y^2 + (1-a)S_X^2} + \Delta^2\rho_n', \quad \rho_n' \underset{P}{\longrightarrow} \rho' = \mathrm{const},$$

мы получим

$$\frac{a(1-a)^2 S_X^2\Delta^2(n_1+n_2) + a^2(1-a)S_Y^2\Delta^2(n_1+n_2)}{\left(aS_Y^2 + (1-a)S_X^2\right)^2} + \\ + \Delta^4(n_1+n_2)\rho_n'' > h_\varepsilon + O\left(\frac{1}{n_1+n_2}\right),$$

где $\rho_n'' \to \rho'' = \text{const}$, $\Delta^4(n_1+n_2) \xrightarrow[P]{} 0$. Это неравенство можно эквивалентным образом записать в форме

$$\frac{\Delta^2(n_1+n_2)}{S_X^2/a + S_Y^2/(1-a)} > h_\varepsilon + \gamma_n, \quad \gamma_n \xrightarrow[P]{} 0.$$

Отсюда вытекает, что критерий

$$\frac{|\overline{\mathrm{x}} - \overline{\mathrm{y}}|\sqrt{n_1+n_2}}{\sqrt{S_X^2/a + S_Y^2/(1-a)}} > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{24}$$

является асимптотически эквивалентным (23) и, стало быть, асимптотически минимаксным для проблемы Беренса–Фишера. Здесь $\lambda_{\varepsilon/2}$ имеет тот же смысл, что и в примере 4. В отличие от примеров 2–4 здесь допредельное распределение статистики в левой части (24) зависит при гипотезе H_1 от неизвестных параметров σ_1^2, σ_2^2.

5. Некоторые другие задачи. Мы отметим здесь еще два класса задач, асимптотическое решение которых может быть найдено с помощью теоремы 55.4.

1. Первый класс задач составляют задачи, обобщающие задачи п. 2–4 на случай, когда проверяются гипотезы вида $\{\theta_1 = f(\theta_2)\}$ (например, $\{\theta_1 = a + b\theta_2\}$) в условиях п. 2 и вида $\{u_1 = f(u_2)\}$ в условиях п. 3, 4. Нетрудно видеть, что проведенные в п. 2–4 рассмотрения переносятся на этот более общий случай.

2. Второй класс задач составляют задачи с тремя и более выборками. Рассмотрим, например, задачу об однородности для трех выборок. Пусть $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $X \in \mathbf{P}_{\theta_3}$. Проверяется гипотеза $H_1 = \{\theta_1 = \theta_2 = \theta_3\}$ против дополнительной альтернативы. Предположим для простоты, что объемы выборок n_1, n_2, n_3 равны: $n_1 = n_2 = n_3 = n$. Рассмотрим объединенную выборку (X, Y, Z) как выборку объема n с наблюдениями $(\mathrm{x}_1, \mathrm{y}_1, \mathrm{z}_1), \ldots$ $\ldots, (\mathrm{x}_n, \mathrm{y}_n, \mathrm{z}_n)$, имеющими плотность $f_{\overline{\theta}}(x,y,z) = f_{\theta_1}(x) f_{\theta_2}(y) f_{\theta_3}(z)$, где $\overline{\theta} = (\theta_1, \theta_2, \theta_3)$. Гипотеза H_1 тогда будет эквивалентна тому, что $\overline{\theta}$ лежит на «кривой» $\overline{\theta} = g(\alpha)$, $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha, \alpha, \alpha)$. Мы видим, что проблема вновь сводится к задаче, рассмотренной в теореме 55.4.

§ 60. Задачи об однородности в общем случае

1. Постановка задачи. В этом параграфе мы будем рассматривать две выборки X и Y объемов n_1 и n_2 соответственно, не предполагая, что они относятся к какому-либо параметрическому семейству.

Задача об однородности выборок X и Y в общем случае выглядит следующим образом. Обозначим через $\mathbf{P}_1$ и $\mathbf{P}_2$ распределения выборок X и Y: $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$. Проверяется гипотеза $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\}$ против $H_2 = \{\mathbf{P}_1 \neq \mathbf{P}_2\}$. Обе гипотезы, очевидно, являются сложными. Распределения $\mathbf{P}_1$, $\mathbf{P}_2$ могут выбираться из заданного семейства $\mathcal{P}$ или быть произвольными. Общий принцип построения статистического критерия для проверки H_1 против H_2 остается тем же, что и в главе. 3. Разница, как и в § 59, состоит лишь в том, что здесь он основан на объединенной выборке

(X, Y), так что $\pi = \pi(X, Y)$ есть вероятность принять H_2 при данной выборке (X, Y). В нерандомизированном случае (π равно нулю или единице) критерий π определяется критической областью $\Omega \subset \mathcal{X}^{n_1+n_2}$ такой, что при $(X, Y) \in \Omega$ принимается H_2. Число

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathcal{P}} \mathbf{P}_1 \times \mathbf{P}_1((X, Y) \notin \Omega)$$

называется *уровнем значимости*, а значение

$$\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2((X, Y) \in \Omega), \quad \mathbf{P}_1 \in \mathcal{P}, \quad \mathbf{P}_2 \in \mathcal{P},$$

— мощностью критерия π в «точке» $(\mathbf{P}_1, \mathbf{P}_2)$.

Критерий π называется *состоятельным*, если $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ при $n_1 \to \infty$, $n_2 \to \infty$ и при любых $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathcal{P}$, $\mathbf{P}_2 \in \mathcal{P}$.

Мы уже знаем, что с ростом n_1, n_2 эмпирические распределения $\mathbf{P}_X^*, \mathbf{P}_Y^*$, соответствующие выборкам X и Y, неограниченно сближаются с $\mathbf{P}_1$, $\mathbf{P}_2$ соответственно. Поэтому естественной основой для построения критериев однородности является использование разного рода «расстояний» $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$ между $\mathbf{P}_X^*$ и $\mathbf{P}_Y^*$, где d удовлетворяет тем же общим условиям, которые описаны в § 52. При этом особый интерес представляют непараметрические и асимптотически непараметрические критерии, которые определяются следующим образом.

Пусть $d(\mathbf{P}, \mathbf{Q})$ — некоторое расстояние (не обязательно метрика) в пространстве распределений. Если вероятность

$$\mathbf{P}_1 \times \mathbf{P}_1\big(d(\mathbf{P}_X^*, \mathbf{P}_Y^*) > c\big) = \varepsilon \tag{1}$$

не зависит от выбора $\mathbf{P}_1$, то критерий π, определенный равенствами

$$\pi(X, Y) = \begin{cases} 0, & \text{если} \quad d(\mathbf{P}_X^*, \mathbf{P}_Y^*) \leqslant c, \\ 1 & \text{в противном случае,} \end{cases} \tag{2}$$

называется *непараметрическим*. Очевидно, построенный непараметрический критерий будет иметь уровень $1 - \varepsilon$.

Аналогичным образом определяются асимптотически непараметрические критерии, когда (1) сохраняется асимптотически, при добавлении в левой части операции $\lim\limits_{n_1 \to \infty, n_2 \to \infty}$. В этом случае критерий (2) будет иметь *асимптотический уровень* $1 - \varepsilon$. При отсутствии непараметричности (точной или асимптотической) строить критерии проверки однородности заданного уровня весьма трудно.

Рассмотрим некоторые основные критерии проверки однородности.

2. Критерий Колмогорова–Смирнова. Пусть $\mathbf{P}_1$, $\mathbf{P}_2$ принадлежат классу $\mathcal{P}$ всех непрерывных распределений на прямой, и пусть F_X^* и F_Y^* — эмпирические функции распределения, соответствующие $\mathbf{P}_X^*$ и $\mathbf{P}_Y^*$. Критерий Колмогорова–Смирнова в качестве расстояния $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$ рассматривает статистику

$$D_{n_1, n_2} = \sup_t \big|F_X^*(t) - F_Y^*(t)\big|.$$

Критерий $D_{n_1,n_2} > c$, построенный с помощью статистики D_{n_1,n_2}, является непараметрическим. Действительно, пусть верна гипотеза H_1 и $F(t)$ есть общая функция распределения для X и Y. Статистику D_{n_1,n_2} можно записать в виде

$$D_{n_1,n_2} = \sup_t \left|G_X^*(F(t)) - G_Y^*(F(t))\right|, \tag{3}$$

где $G_X^*(u) = F_X^*(F^{-1}(u))$ есть эмпирическая функция распределения, соответствующая равномерному на $[0,1]$ распределению (см. § 6, 52). Но в силу (3) $D_{n_1,n_2} = \sup_u \left|G_X^*(u) - G_Y^*(u)\right|$, так что распределение D_{n_1,n_2} от F никак не зависит.

Можно найти точное распределение статистики D_{n_1,n_2}. Например, при $n_1 = n_2 = n$

$$\mathbf{P}(nD_{n,n} \geqslant k) = 2\left(C_{2n}^n\right)^{-1} \sum_{j=1}^{[n/k]} (-1)^{j+1} C_{2n}^{n-jk}, \tag{4}$$

$k = 1, 2, \ldots, n$. Этот факт был установлен Гнеденко и Королюком сведением задачи к простой проблеме для случайных блужданий (см. [111]).

В § 6 мы видели, что распределение $n_1 G_X^*(u)$ совпадает с распределением пуассоновского процесса $\zeta_1(u)$ при условии $\zeta_1(1) = n_1$. Так как $G_X^*(u)$ и $G_Y^*(u)$ независимы, то распределение $G_X^*(u) - G_Y^*(u)$, $u \in [0,1]$, совпадает с распределением сложного пуассоновского процесса $\zeta(u)$, в котором с интенсивностью n_1 происходят скачки величиной $1/n_1$ и с интенсивностью n_2 — скачки величиной $-1/n_2$; распределение надо брать при условии, что произошло $n_1 + n_2$ скачков и что $\zeta(1) = 0$. Поэтому

$$\mathbf{P}(D_{n_1,n_2} < x) = \mathbf{P}\Big(\sup_{u \leqslant 1} |\zeta(u)| < \frac{x}{\zeta(1)} = 0;\ \text{произошло } n_1 + n_2 \text{ скачков}\Big).$$

На основании этого факта в приложении II наряду с теоремой 6.2 о сходимости процесса $w_n(u) = \sqrt{n_1}\left(G_X^*(u) - u\right)$ к броуновскому мосту $w^0(u)$ доказано также утверждение о сходимости к броуновскому мосту и процесса

$$w_{n_1,n_2}(u) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left(G_X^*(u) - G_Y^*(u)\right).$$

Точнее, для любого измеримого и непрерывного в равномерной метрике функционала f распределение $f(w_{n_1,n_2})$ сходится к распределению $f(w^0)$. Отсюда сразу вытекает следующее утверждение, называемое теоремой Смирнова.

Теорема 1.

$$\lim_{\substack{n_1 \to \infty \\ n_2 \to \infty}} \mathbf{P}\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D_{n_1,n_2} < x\right) = \mathbf{P}\Big(\sup_{u \leqslant 1} |w^0(u)| < x\Big) = K(x),$$

где $K(x)$ — *функция Колмогорова* (см. § 8, 52)).

Так как функция $K(x)$ табулирована, то теорема 1 дает удобное средство для приближенного вычисления уровня значимости критерия Колмогорова–Смирнова.

Предоставляем читателю убедиться, что критерий Колмогорова–Смирнова является состоятельным.

3. Критерий знаков. Пусть $n_1 = n_2 = n$. Тогда из наблюдений выборок X и Y можно составить n разностей

$$\mathrm{x}_1 - \mathrm{y}_1, \ldots, \mathrm{x}_n - \mathrm{y}_n. \tag{5}$$

Если верна гипотеза H_1 и $\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 = 0) = 0$ при всех $\mathbf{P}_1 \in \mathcal{P}$ (это, очевидно, всегда так, если $\mathcal{P}$ — множество непрерывных распределений), то

$$\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 > 0) = \mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 < 0) = \frac{1}{2}.$$

Статистика ν критерия знаков представляет собой число положительных разностей в (5)*). Сам критерий можно построить, взяв в качестве критического множества

$$\Omega = \left\{(X,Y)\colon \left|\nu - \frac{n}{2}\right| > c\right\}.$$

Так как распределение ν от $\mathbf{P}_1$ не зависит,

$$\mathbf{P}_1 \times \mathbf{P}_1(\nu = k) = C_n^k 2^{-n},$$

то этот критерий является непараметрическим.

Число c по заданному уровню $1-\varepsilon$ критерия выбирается из соотношения

$$\sum_{k\colon |2k-n| \leqslant 2c} C_n^k 2^{-n} \geqslant 1 - \varepsilon. \tag{6}$$

Так как левая часть здесь возрастает с ростом c дискретно, то в качестве решения следует взять наименьшее c, при котором левая часть в (6) превосходит $1 - \varepsilon$.

Мы видим, что здесь используется критерий для проверки гипотезы о том, что вероятность успеха в схеме Бернулли равна $1/2$. С точки зрения исходной задачи проверяется не гипотеза об однородности, а более широкая гипотеза о том, что

$$\mathbf{P}_1 \times \mathbf{P}_2(\mathrm{x}_1 - \mathrm{y}_1 < 0) = \int F_1(t)\, dF_2(t) = \frac{1}{2}, \tag{7}$$

где F_i соответствуют $\mathbf{P}_i$, $i = 1, 2$. Соотношение (7) означает, что медиана распределения $\mathrm{x}_1 - \mathrm{y}_1$ равна нулю.

Критерий знаков асимптотического уровня $1 - \varepsilon$ будет иметь вид

$$\begin{gathered} \pi(X,Y) = 1, \quad \text{если} \quad \frac{2|\nu - n/2|}{\sqrt{n}} > \lambda_{\varepsilon/2}, \\ \Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1 - \varepsilon. \end{gathered} \tag{8}$$

*) Если в выборках X, Y из-за огрубления данных окажется, что некоторые разности $\mathrm{x}_i - \mathrm{y}_i = 0$, то их следует просто выбросить, а в качестве n взять число разностей, отличных от нуля.

Этот критерий не является состоятельным, так как для $\mathbf{P}_1 \neq \mathbf{P}_2$, удовлетворяющих (7), $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to \varepsilon < 1$ при $n_1 \to \infty$, $n_2 \to \infty$.

4. Критерий Вилкоксона. Этот критерий весьма распространен при проверке гипотез об однородности.

Объединим выборки X и Y в одну выборку (X, Y) и построим из нее вариационный ряд, т. е. расположим все наблюдения в порядке возрастания. Получим последовательность вида

$$\mathrm{y}^{(1)}, \mathrm{y}^{(2)}, \mathrm{x}^{(3)}, \mathrm{y}^{(4)}, \mathrm{x}^{(5)}, \ldots, \tag{9}$$

где верхний индекс означает номер наблюдения в *общем вариационном ряде*, а буква указывает на принадлежность к выборке. Пусть $r_1, r_2, \ldots, r_{n_1}$ означают номера элементов выборки X в вариационном ряде (9). Для выписанной в (9) последовательности $r_1 = 3$, $r_2 = 5$. *Статистикой Вилкоксона* называется функция

$$U = U(X, Y) = \sum_{i=1}^{n_1} (r_i - i),$$

где $r_i - i$ есть число элементов выборки Y, меньших чем $\mathrm{x}_{(i)}$.

Так как порядок наблюдений в (9) инвариантен относительно монотонных преобразований над переменными (порядок для $F_X^*(t)$, $F_Y^*(t)$ будет тем же, что и для $F_X^*(F^{-1}(t))$, $F_Y^*(F^{-1}(t))$, где F — функция распределения), то критерий, построенный по статистике U, будет непараметрическим.

Теорема 2. *Пусть $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$, $F_i \in \mathcal{F}$ суть функции распределения, соответствующие $\mathbf{P}_i$, $i = 1, 2$; $\mathcal{F}$ есть класс всех непрерывных функций распределения. Предположим также, что $a = n_1/(n_1+n_2) \to a_0$ при $n_1 \to \infty$, $n_2 \to \infty$. Тогда*

$$\frac{U - n_1 n_2 \mathbf{E} F_2(\mathrm{x}_1)}{\sqrt{n_1 n_2 (n_1 + n_2)}} \Subset \mathbf{\Phi}_{0,\sigma^2}, \tag{10}$$

где $\sigma^2 = (1 - a_0)\mathbf{D}F_2(\mathrm{x}_1) + a_0 \mathbf{D} F_1(\mathrm{y}_1)$.

Если $F_1 = F_2 = F$, то $F_2(\mathrm{x}_1) \Subset \mathbf{U}_{0,1}$, $F_1(\mathrm{y}_1) \Subset \mathbf{U}_{0,1}$ и, стало быть, $\mathbf{E}F_2(\mathrm{x}_1) = 1/2$, $\mathbf{D}F_2(\mathrm{x}_1) = \mathbf{D}F_1(\mathrm{y}_1) = 1/12$.

Следовательно, критерий Вилкоксона асимптотического уровня $1 - \varepsilon$ будет иметь вид

$$\left| U - \frac{n_1 n_2}{2} \right| > \frac{\lambda_{\varepsilon/2} \sqrt{n_1 n_2 (n_1 + n_2)}}{2\sqrt{3}}, \tag{11}$$

$$\mathbf{\Phi}_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1 - \varepsilon.$$

Из (10) видно, что этот критерий нацелен главным образом на проверку гипотезы (ср. с (7))

$$\int F_2(t)\, dF_1(t) = \frac{1}{2} \quad \text{или} \quad \int (F_2(t) - F_1(t))\, dF_1(t) = 0. \tag{12}$$

Если принять, не ограничивая общности, что $F_1(t) \equiv t$, $t \in [0,1]$, и предположить, что $F_2(0) = 0$, $F_2(1) = 1$, то в силу равенства

$$\int_0^1 (1 - F_2(t))\, dt = \mathbf{E}\mathrm{y}_1$$

проверяемая гипотеза примет вид $\mathbf{E}\mathrm{y}_1 = 1/2$.

Это значит, что критерий Вилкоксона, как и критерий знаков, чувствителен главным образом к сдвигам распределений одно относительно другого. Для таких сдвинутых альтернатив мощность их бывает достаточно большой (см. пример 1). Если же $F_2 \neq F_1$ и выполнено (12), то согласно критерию Вилкоксона гипотеза $\{F_2 = F_1\}$ будет приниматься с вероятностью, близкой к $\mathbf{\Phi}_{0,1}\left(-\dfrac{\lambda_{\varepsilon/2}}{2\sqrt{3}\,\sigma}, \dfrac{\lambda_{\varepsilon/2}}{2\sqrt{3}\,\sigma}\right)$. Это означает, что критерий Вилкоксона будет несостоятельным.

Доказательство теоремы 2. Статистику U можно записать в виде

$$U = \sum_{i=1}^{n_1} n_2 F_Y^*(\mathrm{x}_i) = n_1 n_2 \int F_Y^*(t)\, dF_X^*(t).$$

Положим

$$w_X(t) = \sqrt{n_1}\,\big(F_X^*(t) - F_1(t)\big), \quad w_Y(t) = \sqrt{n_2}\,\big(F_Y^*(t) - F_2(t)\big).$$

Тогда очевидно

$$U = n_1 n_2 \int F_2(t)\, dF_1(t) + \\ + \sqrt{n_1 n_2 (n_1 + n_2)}\left[\sqrt{a}\int w_Y(t)\, dF_1(t) + \sqrt{1-a}\int F_2(t)\, dw_X(t)\right] + \\ + \sqrt{n_1 n_2}\int w_Y(t)\, dw_X(t). \quad (13)$$

Так как здесь $\int F_2(t)\, dw_X(t) = \int w_X(t)\, dF_2(t)$ и, стало быть, второй и третий интегралы в (13) имеют одну и ту же форму и независимы, то для доказательства теоремы нам достаточно убедиться, что

$$\int w_Y(t)\, dF_1(t) \Subset \mathbf{\Phi}_{0,\sigma_1^2}, \quad \sigma_1^2 = \mathbf{D}F_1(\mathrm{y}_1), \quad (14)$$

и что

$$\frac{1}{\sqrt{n_1 + n_2}}\int w_Y(t)\, dw_X(t) \underset{P}{\longrightarrow} 0. \quad (15)$$

В силу теоремы 6.2

$$\int w_Y(t)\, dF_1(t) \Subset \int w^0(F_2(t))\, dF_1(t), \quad (16)$$

где $w^0(u)$ есть броуновский мост. Чтобы найти распределение последнего интеграла, заметим, что траектории винеровского процесса $w(u)$ с вероятностью 1 непрерывны (см. [17, § 16.2]) $w^0(u) = w(u) - uw(1)$ и что, следовательно, интеграл (16) есть по определению результат почти наверное сходимости при $N \to \infty$ сумм

$$\sum_{i=1}^{N} w(F_2(t_i))\Delta_i F_1 - m_1 w(1), \tag{17}$$

где $m_1 = \int F_2(t)\,dF_1(t)$, $\{t_i\}_{i=0}^{N}$ образуют разбиение вещественной оси, $\Delta_i g = g(t_i) - g(t_{i-1})$, $w(F_2(t_i)) = \sum_{l=1}^{i} \Delta_l w(F_2)$, $w(1) = \sum_{l=1}^{N} \Delta_l w(F_2)$.

В силу преобразования Абеля

$$\sum_{i=1}^{N} \left(\sum_{l=1}^{i} d_l \right) b_i = \sum_{i=1}^{N} \left(\sum_{l=i}^{N} b_l \right) d_i.$$

Поэтому (17) равно

$$\sum_{i=1}^{N} (1 - F_1(t_{i-1}) - m_1)\Delta_i w(F_2). \tag{18}$$

Здесь $1 - m_1 = \int F_1(t)\,dF_2(t) \equiv m_2$, а $\Delta_i w(F_2)$ суть независимые нормально распределенные случайные величины с параметрами $(0, \Delta_i F_2)$. Поэтому распределение (17), (18) будет нормальным с нулевым средним и дисперсией

$$\sum_{i=1}^{N} (m_2 - F_1(t_{i-1}))^2 \Delta_i F_2 \to \int (m_2 - F_1(t))^2\,dF_2(t) = \mathbf{D}F_1(\mathrm{y}_1).$$

Соотношение (14) доказано.

Чтобы доказать (15)*), проще всего оценить дисперсию интеграла в (15). Приближая интеграл снова с помощью конечной суммы, можно убедиться, что дисперсия

$$\mathbf{D}_{X,Y} = \mathbf{E}\left(\int w_Y(t)\,dw_X(t) \right)^2$$

ограничена при $n_1 \to \infty$, $n_2 \to \infty$. Отсюда и из неравенства Чебышева следует (15). Доказательство ограниченности $\mathbf{D}_{X,Y}$ мы опускаем ввиду громоздкости и рутинного характера вычислений. $\triangleleft$

По поводу критериев знаков и Вилкоксона см. также [34].

Пример 1. Мы отмечали, что критерии знаков и Вилкоксона наиболее чувствительны к сдвигам. Поэтому интересно сравнить их мощности с мощностью оптимального критерия в задаче, где однородность проверяется

*) Интеграл в (15) сходится по распределению к $\int w^0(F_2(t))\,dw^0(F_1(t))$.

для семейства $\mathcal{P}$ распределений, отличающихся лишь сдвигами. Именно, пусть

$$\mathcal{P} = \{\mathbf{\Phi}_{\alpha,1}\}, \quad \mathbf{P}_1 = \mathbf{\Phi}_{\alpha_1,1}, \quad \mathbf{P}_2 = \mathbf{\Phi}_{\alpha_2,1}, \quad n_1 = n_2 = n.$$

В этом случае согласно теореме 59.1 для проверки гипотезы $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\} = \{\alpha_1 = \alpha_2\}$ против $H_2^b = \{|\alpha_1 - \alpha_2| \geqslant b/\sqrt{n}\}$ существует асимптотически минимаксный критерий π_0 уровня $1 - \varepsilon$, имеющий вид

$$|\overline{\mathrm{x}} - \overline{\mathrm{y}}| > \lambda_{\varepsilon/2}\sqrt{\frac{2}{n}}, \quad \mathbf{\Phi}_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1 - \varepsilon$$

(то, что это неравенство в нашем примере эквивалентно (59.3), (59.4), читатель может проверить сам). Примем этот критерий за эталон для сравнения с другими критериями и рассмотрим альтернативу $(\mathbf{P}_1, \mathbf{P}_2)$, где $\alpha_2 = \alpha_1 + c/\sqrt{n}$ (мы рассматриваем близкие альтернативы, чтобы не иметь дела с задачей о больших уклонениях). Очевидно, что в этом случае $(\overline{\mathrm{x}} - \overline{\mathrm{y}}) \Subset \mathbf{\Phi}_{-c/\sqrt{n},2/n}$. Стало быть,

$$\beta_{\pi_0}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(|\overline{\mathrm{x}} - \overline{\mathrm{y}}| > \lambda_{\varepsilon/2}\sqrt{\frac{2}{n}}\right) = 1 - \mathbf{\Phi}_{-c/\sqrt{2},1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) =$$

$$= 1 - \mathbf{\Phi}_{0,1}\left(-\lambda_{\varepsilon/2} + \frac{c}{\sqrt{2}}, \lambda_{\varepsilon/2} + \frac{c}{\sqrt{2}}\right) \equiv \beta_0(c). \quad (19)$$

Рассмотрим теперь критерий знаков (8) и обозначим его через π_1. Пользуясь разложением в ряд по степеням $c/\sqrt{n}$, находим ($\mathbf{\Phi}_{\alpha,\sigma^2}(x) = \mathbf{\Phi}_{\alpha,\sigma^2}((-\infty, x))$)

$$\mathbf{P}_1 \times \mathbf{P}_2(\mathrm{x}_1 - \mathrm{y}_1 < 0) = \mathbf{\Phi}_{0,2}\left(\frac{c}{\sqrt{n}}\right) = \frac{1}{2} + \frac{c}{\sqrt{n}}\frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Поэтому в точке $(\mathbf{P}_1, \mathbf{P}_2)$

$$\frac{2}{\sqrt{n}}\left(\nu - \frac{n}{2} + \frac{c\sqrt{n}}{2\sqrt{\pi}}\right) \Subset \mathbf{\Phi}_{0,1}.$$

Стало быть, для критерия знаков π_1 асимптотического уровня $1 - \varepsilon$

$$\beta_{\pi_1}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(2\left|\nu - \frac{n}{2}\right| > \lambda_{\varepsilon/2}\sqrt{n}\right) \to$$

$$\to 1 - \mathbf{\Phi}_{0,1}\left(-\lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}, \lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}\right).$$

Обратимся, наконец, к критерию π_2 Вилкоксона (см. (11)), имеющему в нашем случае вид

$$\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2} n^{3/2}}{\sqrt{6}}.$$

Очевидно, что статистика U инвариантна относительно преобразования сдвига элементов выборок X и Y. Поэтому можно считать, что $\mathbf{P}_1 = \mathbf{\Phi}_{0,1}$, $\mathbf{P}_2 = \mathbf{\Phi}_{c/\sqrt{n},1}$ и, стало быть,

$$\mathbf{E}F_2(\mathrm{x}_1) = \int F_2(t)\,dF_1(t) = \int \mathbf{\Phi}_{0,1}\left(t - \frac{c}{\sqrt{n}}\right) d\mathbf{\Phi}_{0,1}(t) = \\ = \mathbf{\Phi}_{0,2}\left(-\frac{c}{\sqrt{n}}\right) = \frac{1}{2} - \frac{c}{\sqrt{n}}\frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Так как $\mathbf{D}F_2(\mathrm{x}_1) \to \mathbf{D}F_1(\mathrm{x}_1) = 1/12$, $\mathbf{D}F_1(\mathrm{y}_1) \to \mathbf{D}F_1(\mathrm{x}_1) = 1/12$, то по теореме 2

$$\beta_{\pi_2}(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2\left(\left|U - \frac{n^2}{2}\right| > \frac{\lambda_{\varepsilon/2} n^{3/2}}{\sqrt{6}}\right) = \\ = 1 - \mathbf{P}_1 \times \mathbf{P}_2\left(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}} \leqslant \sqrt{6}\, n^{-3/2}\left(U - \frac{n^2}{2} + \frac{n^{3/2}c}{2\sqrt{\pi}}\right) \leqslant \right. \\ \left. \leqslant \lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}\right) \to 1 - \mathbf{\Phi}_{0,1}\left(-\lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}},\ \lambda_{\varepsilon/2} + c\sqrt{\frac{3}{2\pi}}\right).$$

Заметим теперь, что $\beta_0(c)$ (см. (19)) есть монотонно возрастающая функция от c и что при больших n

$$\beta_{\pi_1}(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{2}{\pi}}\, c\right), \quad \beta_{\pi_2}(\mathbf{P}_1, \mathbf{P}_2) \approx \beta_0\left(\sqrt{\frac{3}{\pi}}\, c\right).$$

Таким образом, при каждом $c > 0$ наиболее мощным среди π_0, π_1, π_2 оказывается, как этого и следовало ожидать, критерий π_0. Затем идут критерий Вилкоксона и критерий знаков; при этом критерий Вилкоксона уступает критерию π_0 совсем немного, так как $\sqrt{3/\pi} \approx 0{,}977$.

Если мы при том же смещении $\alpha_2 - \alpha_1 = c/\sqrt{n}$ рассмотрим выборки X' и Y' объема $n' > n$, то для того чтобы с помощью проведенных вычислений получить мощность в точке $(\mathbf{P}_1, \mathbf{P}_2)$ критериев $\pi_j(X', Y')$, мы должны рассмотреть прежнюю задачу при новом значении c, равном $c' = c\sqrt{n'}/\sqrt{n}$ (тогда $\alpha_2 - \alpha_1$ можно будет записать в виде $c'/\sqrt{n'}$). Стало быть, мощности $\pi_1(X', Y')$ и $\pi_2(X', Y')$ в той же точке $(\mathbf{P}_1, \mathbf{P}_2)$ будут равны приближенно

$$\beta_0\left(\sqrt{\frac{2}{\pi}}\, c'\right) = \beta_0\left(\sqrt{\frac{2n'}{\pi n}}\, c\right), \quad \beta_0\left(\sqrt{\frac{3}{\pi}}\, c'\right) = \beta_0\left(\frac{3n'}{\pi n}\, c\right).$$

Приравнивая $\dfrac{2n'}{\pi n} = 1$, $\dfrac{3n'}{\pi n} = 1$, мы получим значения $n' = \dfrac{\pi}{2}n$, $n' = \dfrac{\pi}{3}n$ (они от c не зависят) для числа наблюдений, которые надо проделать, чтобы соответственно с помощью критериев π_1 и π_2 получить ту же мощность, что и для критерия π_0 с n наблюдениями. Например, для $n = 100$ наблюдений

с критерием π_0 нам потребуется для получения тех же результатов $n' \approx 105$ наблюдений с критерием π_2 и $n' \approx 157$ наблюдений с критерием π_1.

Совсем другие результаты мы получили бы, если бы проверяли однородность для семейства $\mathcal{P} = \{\mathbf{\Phi}_{0,\sigma^2}\}$. В этом случае критерии знаков и Вилкоксона оказались бы несостоятельными. Более того, критерий знаков уровня $1-\varepsilon$ был бы, по существу, равнозначен критерию $\pi \equiv \varepsilon$, не зависящему от выборок, так как $\mathbf{E}(\mathrm{x}_1 - \mathrm{y}_1) = 0$ и $\mathbf{P}_1 \times \mathbf{P}_2(\mathrm{x}_1 - \mathrm{y}_1 > 0) = 1/2$ для любой пары распределений $\mathbf{P}_1$ и $\mathbf{P}_2$ из $\mathcal{P}$. Для этой задачи можно было бы рассмотреть другие непараметрические критерии, использующие статистики r_i, например критерий $\sum_{i=0}^{n_1} (r_{i+1} - r_i)^2$, $r_0 = 0$, $r_{n_1+1} = n_2$, напоминающий по своим свойствам критерий Морана (§ 52).

5. Критерий χ^2 как асимптотически оптимальный критерий проверки однородности по сгруппированным данным. В этом пункте мы будем предполагать, что данные в обеих выборках X и Y объемов соответственно n_1 и n_2 сгруппированы (см. § 56). Тогда вместо выборок X, Y можно использовать векторы $\nu = (\nu_1, \ldots, \nu_r)$ и $\mu = (\mu_1, \ldots, \mu_r)$ частот наблюдений соответственно из выборок X и Y, попавших в интервалы $\Delta_1, \ldots, \Delta_r$, определяющие группировку.

Обозначим через $\theta_i = (\theta_{i1}, \ldots, \theta_{ir})$, $i = 1, 2$, векторы вероятностей попадания наблюдений соответственно из первой и второй выборок в интервалы $\Delta_1, \ldots, \Delta_r$, так что $\theta_{1l} = \mathbf{P}(\mathrm{x}_j \in \Delta_l)$, $\theta_{2l} = \mathbf{P}(\mathrm{y}_j \in \Delta_l)$. Огрубленные выборки X и Y можно рассматривать тогда как выборки из параметрических семейств $\mathbf{B}_{\theta_1}$ и $\mathbf{B}_{\theta_2}$ соответственно.

Таким образом, задача становится параметрической и мы можем использовать результаты, изложенные в примере 1 предыдущего параграфа. Из этого примера следует, что если мы проверяем гипотезу об однородности $H_1 = \{\theta_1 = \theta_2\}$ в случае, когда параметр θ локализован, т. е. значения θ_1, θ_2 располагаются в окрестности точки $\theta_0 = (\theta_{01}, \ldots, \theta_{0r})$, то асимптотически минимаксный критерий асимптотического уровня $1-\varepsilon$ для проверки H_1 против

$$H_2^b = \left\{ \sum_{i=1}^{r} \frac{(\theta_{1i} - \theta_{2i})^2}{\theta_{0i}} \geqslant \frac{b^2}{n_2} \right\}$$

имеет вид

$$\sum_{i=1}^{r} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2} \right)^2 \frac{n_1 n_2}{\nu_i + \mu_i} \geqslant h_\varepsilon,$$

где h_ε есть квантиль порядка $1-\varepsilon$ распределения χ^2 с $r-1$ степенями свободы. Это и есть критерий χ^2 для проверки однородности по сгруппированным данным.

В качестве асимптотически эквивалентного можно рассматривать критерий

$$\sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^{r} \mu_i \ln \frac{\mu_i}{n_2} - \sum_{i=1}^{r} (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2}.$$

§ 61. Задачи регрессии

1. Постановка задачи. В приложениях часто возникают задачи, относящиеся к наблюдениям, распределение которых меняется в разных экспериментах при изменении некоторых параметров, характеризующих эти эксперименты. Набор значений указанных параметров в i-м эксперименте, $i = 1, \ldots, n$, обозначим через

$$\mathrm{x}_i = (\mathrm{x}_{i,1}, \ldots, \mathrm{x}_{i,r})$$

(так что r есть размерность векторов x_i). Значения $\mathrm{x}_{i,k}$ определяются либо экспериментатором, либо природой изучаемого явления. Вектор $(\mathrm{x}_{1,k}, \ldots, \mathrm{x}_{n,k})$ мы обозначим через X_k, а матрицу $\begin{pmatrix} X_1 \\ \ldots \\ X_r \end{pmatrix} = (\mathrm{x}_1^\top, \ldots, \mathrm{x}_n^\top)$ — через X. Таким образом, здесь в отличие от предыдущего X является матрицей порядка $r \times n$ и может быть произвольным неслучайным набором чисел, природа которых нас интересовать не будет. Вектор наблюдений обозначим через $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_n)$.

Задачи регрессии связаны с предположением, что наблюдения y_i в зависимости от набора параметров $\mathrm{x}_i = (\mathrm{x}_{i,1}, \ldots, \mathrm{x}_{i,r})$ имеют вид

$$\mathrm{y}_i = \alpha_1 \mathrm{x}_{i,1} + \cdots + \alpha_r \mathrm{x}_{i,r} + \xi_i, \quad i = 1, \ldots, n, \tag{1}$$

где $\alpha = (\alpha_1, \ldots, \alpha_r)$ — неизвестные нам постоянные, $\xi_i \in \Phi_{0,\sigma^2}$ и независимы.

Постоянная α_1 часто играет особую роль, так как в ряде случаев она выделяет в представлении (1) постоянное слагаемое, что соответствует тому, что в матрице X заранее полагается $X_1 = (1, \ldots, 1)$ $(\mathrm{x}_{i,1} \equiv 1)$. Мы этим предположением пользоваться не будем. Случайные величины ξ_i обуславливаются шумами, флуктуациями или ошибками измерений.

В матричной форме соотношения (1) можно записать в виде

$$Y = \alpha X + \xi. \tag{2}$$

Регрессия вида (1), (2) называется линейной (как по α, так и по X). В качестве задач регрессии могут рассматриваться как задача оценки неизвестных параметров α, σ^2, если известно, что справедливо (1), (2), так и задача проверки самой гипотезы о том, что представление (1), (2) имеет место. Исходными данными в обоих случаях служит «выборка» (X, Y). Термин «выборка» здесь используем уже в более широком, чем прежде, смысле, обозначая им набор результатов наблюдений, не обязательно имеющих одинаковую природу. Кроме того, напомним, что первая из двух «выборок» X и Y может быть неслучайной. Матрицу X называют иногда *регрессором*, а вектор Y — *откликом*.

Модель регрессии (1), (2) является весьма общей, если иметь в виду зависимость y_i от набора параметров. Полагая, например, $\mathrm{x}_{i,k} = \psi_k(\mathrm{z}_i)$, где $\psi_1, \ldots, \psi_r$ — заданный набор функций, а z_i — значения одномерного

параметра, мы получим модель

$$y_i = \alpha_1\psi_1(z_i) + \ldots + \alpha_r\psi_r(z_i) + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

регрессии по произвольным функциям $\psi_1, \ldots, \psi_r$ (и по-прежнему линейной по α). Если $\psi_1(z) \equiv 1$, $\psi_2(z) \equiv z$, $r = 2$, то мы получим модель *простейшей* (одномерной) *линейной* регрессии (рис. 6).

В отличие от простейшей, общую модель (1), (2) иногда называют *множественной* регрессией. В целом задачи регрессии, как мы видим, связаны с изучением (с наличием) функциональной зависимости $y = \varphi(x)$ для заданного класса функций φ в тех случаях, когда наблюдения над переменной y при данном x «зашумлены» случайными отклонениями.

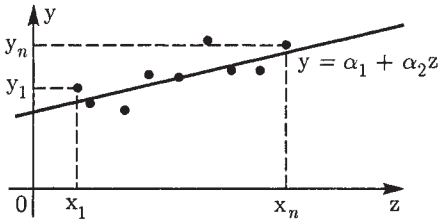

Рис. 6

Строки $X_1, \ldots, X_r$ матрицы X в (2) обычно выбирают таким образом, чтобы они были *линейно независимы* (иначе координаты α нам не удастся оценить). Мы тоже будем следовать этому соглашению, означающему, что ранг матрицы X равен r.

Иногда бывает удобнее иметь дело с ортогональными векторами $X_1, \ldots, X_r$, т. е. удовлетворяющими условию $(X_i, X_j) = 0$, $i \neq j$, где (a, b) означает скалярное произведение. Если исходный набор линейно независимых векторов $\{X_k\}$ таким свойством не обладает, то его можно ортогонализировать, введя новые векторы

$$\begin{aligned} X_1' &= X_1, \\ X_2' &= X_2 + a_{2,1}X_1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_r' &= X_r + a_{r,r-1}X_{r-1} + \cdots + a_{r,1}X_1. \end{aligned} \tag{4}$$

Коэффициенты $a_{k,j}$ легко определяются из условий ортогональности $X_k' \perp X_j'$, $k \neq j$, так что, например, $a_{2,1} = -(X_1, X_2)/(X_1, X_1)$. Соотношения (4) можно записать в виде $X' = AX$, где A — треугольная обратимая матрица (с единицами по главной диагонали). Отсюда получаем $X = A^{-1}X'$, $Y = \alpha A^{-1}X' + \xi$. Мы пришли к задаче регрессии с коэффициентами $\beta = \alpha A^{-1}$. Вектор α очевидным образом восстанавливается по β с помощью равенства $\alpha = \beta A$.

Для простейшей линейной регрессии предположение об ортогональности $X_1 = (1, \ldots, 1)$ и $X_2 = (z_1, \ldots, z_n)$ означает предположение $\sum z_i = 0$, которое очевидным образом может быть удовлетворено путем изменения начала отсчета переменной z.

2. Оценка параметров. Везде в дальнейшем будем предполагать, что $r < n$ и что векторы X_k, $k = 1, \ldots, r$, линейно независимы. Функция прав-

доподобия наблюдения Y при данном X для регрессии (1), (2) равна

$$f_{\alpha,\sigma^2}(Y) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\mathrm{y}_i - \sum_{k=1}^{r}\alpha_k \mathrm{x}_{i,k}\right)^2\right\} =$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{|Y-\alpha X|^2}{2\sigma^2}\right\}. \quad (5)$$

Функция (5) зависит от параметра $\theta = (\alpha, \sigma^2)$. Отметим, что если (5) рассматривать как функцию правдоподобия не одного наблюдения Y (или (X, Y)), а n наблюдений $\mathrm{y}_1, \ldots, \mathrm{y}_n$, то она не будет соответствовать выборке из какого-нибудь одного параметрического семейства. Наблюдения y_i относятся к разным распределениям $\mathbf{\Phi}_{\gamma_i,\sigma^2}$, $\gamma_i = \sum_{k=1}^{r}\alpha_k \mathrm{x}_{i,k}$, зависящим от x_i. Поэтому те рассмотрения глав 1–3, в которых использовалась одинаковая распределенность элементов выборки, здесь непосредственно неприменимы. (О статистике неодинаково распределенных наблюдений см. гл. 5.)

Итак, будем рассматривать (5) как функцию правдоподобия наблюдения (X, Y). Воспользуемся методом максимального правдоподобия. Непосредственно из (5) видно, что оценка максимального правдоподобия $\alpha^* = \widehat{\alpha}^*$, максимизирующая $f_\theta(Y)$ по α, есть оценка, минимизирующая $|Y - \alpha X|^2$. Поэтому в нашем случае метод максимального правдоподобия совпадает с «методом наименьших квадратов».

Обозначим через $\mathcal{L}[X]$ подпространство, натянутое на векторы $X_1, \ldots, X_r$. Это есть совокупность точек вида αX, когда α пробегает значения из $\mathbb{R}^r$. Это пространство имеет размерность r и содержит единственную точку $\beta = \alpha^* X$,

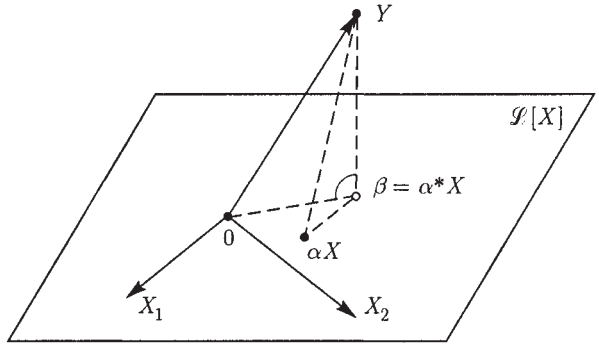

Рис. 7

наименее удаленную от Y (рис. 7). Значение β однозначно определяется условием ортогональности $Y - \beta$ и $\mathcal{L}[X]$ или, что то же, условиями

$$(Y - \alpha^* X, X_k) = (Y - \alpha^* X)X_k^\top = 0, \quad k = 1, \ldots, r.$$

В матричной форме эти условия можно записать в виде $(Y - \alpha^* X)X^\top = 0$. Отсюда находим

$$\alpha^* = YX^\top (XX^\top)^{-1}. \quad (6)$$

Здесь обратная матрица $(XX^\top)^{-1}$ (порядка $r \times r$) существует, так как матрица $D = XX^\top$ является положительно определенной. Действительно, мы видели, что существует невырожденная матрица A такая, что строки матрицы $X' = AX$ ортогональны. Стало быть, матрицу D можно записать в виде

$$XX^\top = A^{-1}X'(X')^\top(A^{-1})^\top = A^{-1}B(A^{-1})^\top,$$

где $B = X'(X')^\top$ — диагональная матрица с элементами

$$(X'_i, X'_j) = \begin{cases} |X'_i|^2 > 0 & \text{при} \quad i = j, \\ 0 & \text{при} \quad i \neq j. \end{cases}$$

Следовательно, B положительно определена, $aBa^\top > 0$ для любого $a \in \mathbb{R}^r$, $a \neq 0$. Полагая $b = aA$, получим $bDb^\top = aAXX^\top A^\top a^\top = aBa^\top > 0$ при любом $b \in \mathbb{R}^r$, $b \neq 0$, что и требовалось доказать.

Если X_k ортогональны, то из (6) находим $\alpha_k^* = (Y, X_k)/(X_k, X_k)$.

Результат (6) мы могли бы получить и другим путем — дифференцируя (5) по α_k и приравнивая производные нулю.

Разность $Y - \alpha^* X$ называют иногда *остатком*. Эта разность ортогональна $\mathcal{L}[X]$, а вместе с этим и любому вектору $\gamma X \in \mathcal{L}[X]$, $\gamma \in \mathbb{R}^r$. Если взять $\gamma = \alpha^* - \alpha$, то из равенства $Y - \alpha X = Y - \alpha^* X + (\alpha^* - \alpha)X$ будет следовать

$$|Y - \alpha X|^2 = |Y - \alpha^* X|^2 + |(\alpha^* - \alpha)X|^2. \tag{7}$$

Найдем теперь о. м. п. для σ^2. Из (5) видно, что это будет та же оценка, что и для нормального семейства (можно заново продифференцировать (5) по σ и приравнять производную нулю), так что

$$(\widehat{\sigma}^2)^* = \frac{1}{n}|Y - \alpha^* X|^2. \tag{8}$$

Положим

$$(\sigma^2)^* = \frac{1}{n-r}|Y - \alpha^* X|^2 = \frac{n}{n-r}(\widehat{\sigma}^2)^*. \tag{9}$$

В дальнейшем E_l будет означать единичную матрицу порядка l, $\sigma^* = \sqrt{(\sigma^2)^*}$.

Теорема 1. *Оценки* (6) *и* (9) *являются независимыми несмещенными эффективными оценками параметров* α *и* σ^2. *Кроме того,*

$$(\alpha^* - \alpha)D^{1/2} \in \mathbf{\Phi}_{0,\sigma^2 E_r}, \quad D = XX^\top, \tag{10}$$

$$\frac{(n-r)(\sigma^2)^*}{\sigma^2} = \frac{|Y - \alpha^* X|^2}{\sigma^2} \in \mathbf{H}_{n-r}. \tag{11}$$

Если X_k *ортогональны, то* α_k^* *независимы,*

$$(\alpha_k^* - \alpha_k)|X_k| \in \mathbf{\Phi}_{0,\sigma^2}. \tag{12}$$

Следствие 1. *Из* (10), (11) *вытекает, что*

$$\frac{(\alpha^*-\alpha)D(\alpha^*-\alpha)^\top}{(n-r)(\sigma^2)^*}=\frac{|(\alpha^*-\alpha)X|^2}{|Y-\alpha^*X|^2}\in\mathbf{F}_{r,n-r}. \tag{13}$$

Пусть $\overline{\alpha}$, $\overline{\alpha}^*$ — *«подвекторы» размерности* $l\leqslant r$ *векторов* α *и* α^*, *составленные из координат с фиксированными номерами* $k_1,\ldots,k_l$, *и пусть* $\overline{X}$ — *матрица, составленная из строк* $X_{k_1},\ldots,X_{k_l}$. *Тогда если* X_k, $k=1,\ldots,r$, *ортогональны, то*

$$(\overline{\alpha}^*-\overline{\alpha})(\overline{X}\overline{X}^\top)^{1/2}\in\mathbf{\Phi}_{0,\sigma^2E_l},\quad\frac{(\alpha_k^*-\alpha_k)|X_k|}{\sigma^*}\in\mathbf{T}_{n-r}. \tag{14}$$

Доказательство теоремы 1. Так как $YX^\top=\alpha XX^\top+\xi X^\top$, то

$$\alpha=(YX^\top-\xi X^\top)D^{-1},\quad\alpha^*-\alpha=\xi X^\top D^{-1}. \tag{15}$$

Матрица вторых моментов вектора $(\alpha^*-\alpha)D^{1/2}$ равна

$$\mathbf{E}D^{1/2}(\alpha^*-\alpha)^\top(\alpha^*-\alpha)D^{1/2}=D^{1/2}D^{-1}X(\mathbf{E}\xi^\top\xi)X^\top D^{-1}D^{1/2}=\sigma^2E_r.$$

Так как компоненты этого вектора нормальны, то они независимы и $\frac{1}{\sigma^2}|(\alpha^*-\alpha)D^{1/2}|^2\in\mathbf{H}_r$. Далее, в силу (7), (9)

$$(n-r)(\sigma^2)^*=|Y-\alpha^*X|^2=|\xi|^2-|(\alpha^*-\alpha)X|^2.$$

Убедимся теперь, что векторы α^* и $Y-\alpha^*X$ (а стало быть, α^* и σ^*) независимы. В силу их нормальности достаточно проверить, что коэффициенты корреляции между их компонентами равны нулю или, что то же, что матрица центральных вторых моментов $\mathbf{E}(\alpha^*-\alpha)^\top(Y-\alpha^*X)$ равна нулю. Заметим, что в силу (6)

$$\alpha^*X=YX^\top(XX^\top)^{-1}X=YX^\top D^{-1}X$$

и вектор α^*X получается из Y путем проектирования Y на $\mathcal{L}[X]$. Оператор проектирования, определяемый матрицей $\Pi=X^\top D^{-1}X$, обладает очевидными свойствами: $\Pi^2=\Pi$, $BX\Pi=BX$ для любой матрицы B, имеющей r столбцов. Поэтому в силу (15)

$$\mathbf{E}(\alpha^*-\alpha)^\top(Y-\alpha^*X)=\mathbf{E}D^{-1}X\xi^\top(\xi-\xi X^\top D^{-1}X)=D^{-1}X\sigma^2(E_n-\Pi)=0.$$

Докажем теперь (11). В силу (7)

$$|Y-\alpha^*X|^2=|\xi|^2-|(\alpha^*-\alpha)X|^2=|\xi|^2-|(\alpha^*-\alpha)D^{1/2}|^2,$$

где $\frac{1}{\sigma^2}|\xi|^2\in\mathbf{H}_n$, $\frac{1}{\sigma^2}|(\alpha^*-\alpha)D^{1/2}|^2\in\mathbf{H}_r$ (см. (10)). Утверждение (11) будет следствием этих соотношений и леммы 1.

Лемма 1. *Если* $\eta = \eta_1 + \eta_2$, *где* η_1 *и* η_2 *независимы,* $\eta \in \mathbf{H}_n$, $\eta_1 \in \mathbf{H}_r$, *то* $\eta_2 \in \mathbf{H}_{n-r}$.

Доказательство. Если обозначить через $\varphi(t)$ характеристическую функцию распределения $\mathbf{H}_1$: $\varphi(t) = (1+2it)^{-1/2}$, то

$$\mathbf{E}e^{i\eta t} = \varphi(t)^n = \varphi(t)^r \cdot \mathbf{E}e^{i\eta_2 t}.$$

Так как $\varphi(t) \neq 0$ на вещественной оси, то $\mathbf{E}e^{i\eta_2 t} = \varphi(t)^{n-r}$. Лемма доказана.

Несмещенность оценок α^*, $(\sigma^2)^*$ с очевидностью следует из (10), (11) ($\mathbf{E}\eta = l$, если $\eta \in \mathbf{H}_l$).

Нам осталось доказать эффективность оценки $\theta^* = (\alpha^*, (\sigma^2)^*)$. Для этого заметим, что параметрическое семейство (5) относится к экспоненциальному типу, так как (5) представимо в форме (см. (25.1))

$$f_\theta(Y) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(|Y|^2 - 2(Y, \alpha X) + |\alpha X|^2)\right\} =$$

$$= h(Y)\exp\left\{\sum_{k=1}^{r+1} a_k(\theta)U_k(Y) + V(\theta)\right\},$$

где

$$h(Y) = (2\pi)^{-n/2}, \quad V(\theta) = -n\ln\sigma - \frac{1}{2\sigma^2}|\alpha X|^2,$$

$$a_k(\theta) = \frac{\alpha_k}{\sigma^2}, \quad U_k(Y) = (Y, X_k), \quad k = 1, \ldots, r,$$

$$a_{r+1}(\theta) = -\frac{1}{2\sigma^2}, \quad U_{r+1}(Y) = |Y|^2.$$

Так как условия теорем 25.1, 25.2 здесь выполнены, то статистика $U = (U_1(X), \ldots, U_{r+1}(X))$ (а вместе с ней и θ^*) является минимальной полной достаточной статистикой). Отсюда следует (см. следствие 25.1) эффективность θ^*.

Утверждение (12) с очевидностью следует из (10), так как для ортогональных X_k матрица $D^{1/2}$ диагональна с элементами $|X_k|$ по диагонали. Теорема доказана.

Замечание 1. Хотеллинг (см. [101]) доказал, что $\mathbf{D}\alpha_k^* \geqslant \sigma^2/|X_k|^2$ и равенство достигается только в том случае, когда X_k ортогональны. Таким образом, при планировании эксперимента при заданных значениях $|X_k|$ оптимальный выбор регрессора X состоит в том, чтобы сделать X_k ортогональными.

Замечание 2. Интересно сравнить матрицу вторых моментов оценки θ^* с нижней границей для несмещенных оценок, определяемой в силу многомерного неравенства Рао–Крамера матрицей $I^{-1}(\theta)$, где $I(\theta)$ — информационная матрица Фишера:

$$I(\theta) = \|I_{ij}(\theta)\|, \quad I_{ij}(\theta) = \mathbf{E}_\theta \frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}, \quad L = L(Y;\theta) = \ln f_\theta(Y).$$

Здесь мы приняли $\theta_k = \alpha_k$, $k = 1, \ldots, r$; $\theta_{r+1} = \sigma^2$. Пусть для простоты X_k ортогональны. Из независимости θ_k^* следует, что матрица $\mathbf{E}_\theta(\theta^* - \theta)^\top(\theta^* - \theta)$ будет диагональной с диагональными элементами

$$\mathbf{E}_\theta\big(\alpha_k^* - \alpha\big)^2 = \frac{\sigma^2}{|X_k|^2}, \quad k = 1, \ldots, r,$$

$$\mathbf{E}_\theta((\sigma^2)^* - \sigma^2)^2 = \mathbf{E}\left(\frac{\sigma^2\chi^2_{n-r}}{n-r} - \sigma^2\right)^2 = \frac{\sigma^4}{n-r}\mathbf{E}\big(\chi_1^2 - 1\big)^2 = \frac{2\sigma^4}{n-r},$$

где $\chi_l^2 \in \mathbf{H}_l$.

С другой стороны, для матрицы $I(\theta)$ в силу того, что

$$\frac{\partial L}{\partial \alpha_k} = \frac{1}{\sigma^2}\sum_{i=1}^n \left(\mathrm{y}_i - \sum_{j=1}^r \alpha_j \mathrm{x}_{ij}\right)\mathrm{x}_{ik} = \frac{1}{\sigma^2}(Y - \alpha X)X_k^\top,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n \left(\mathrm{y}_i - \sum_{j=1}^r \alpha_j \mathrm{x}_{ij}\right)^2 = \frac{1}{2\sigma^2}\left(\frac{|Y - \alpha X|^2}{\sigma^2} - n\right),$$

находим при $k = 1, \ldots, r$

$$I_{kk}(\theta) = \mathbf{E}_\theta \frac{1}{\sigma^4} X_k(Y - \alpha X)^\top (Y - \alpha X)X_k^\top =$$

$$= \frac{1}{\sigma^4}\mathbf{E}X_k\xi^\top \xi X_k^\top = \frac{1}{\sigma^4}\mathbf{E}(\xi, X_k)^2 = \frac{|X_k|^2\mathbf{E}|\xi|^2}{\sigma^4} = \frac{|X_k|^2}{\sigma^2},$$

$$I_{r+1,r+1}(\theta) = \frac{1}{4\sigma^4}\mathbf{E}\left[\sum_{i=1}^n \left(\frac{\xi_1^2}{\sigma^2} - 1\right)\right]^2 = \frac{n}{2\sigma^4}, \quad I_{ij}(\theta) = 0$$

при $i \neq j$. Так что

$$I^{-1}(\theta) = \begin{pmatrix} \dfrac{\sigma^2}{|X_1|^2} & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & \dfrac{\sigma^2}{|X_r|^2} & 0 \\ 0 & \cdots & 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}.$$

Таким образом, в неравенстве Рао–Крамера

$$\mathbf{E}_\theta(\theta^* - \theta)^\top(\theta^* - \theta) \geqslant I^{-1}(\theta) \tag{16}$$

для первых r компонент θ^* достигается равенство. Для $(r+1)$-й оно достигаться не может (хотя асимптотически при $n \to \infty$ обе части (16) ведут

себя одинаково), так как необходимое и достаточное условие теоремы 26.1А здесь не выполнено.

Замечание 3. Предположение о нормальности ξ_i становится малосущественным для утверждений (10)–(12), если n велико (в (11) лучше произвести нормировку и утверждать близость к нормальному закону).

Замечание 4. Сам термин «регрессия» относится к совместному распределению двух случайных величин ξ и η и означает кривую

$$g(x) = \mathbf{E}(\eta/\xi = x),$$

которую называют также регрессией η на ξ. Например, если $(\xi, \eta) \in \Phi_{\gamma,\sigma^2}$, $\gamma = (\gamma_1, \gamma_2)$, $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, 2$, то, как мы видели в § 20, $g(x) = \gamma_2 + \dfrac{\sigma_{12}}{\sigma_{22}}(x - \gamma_1)$. Это есть простейшая линейная регрессия.

Замечание 5. Предположение $\xi_i \in \Phi_{0,\sigma^2}$ об одинаковой распределенности ξ_i при известном σ^2 может быть ослаблено. Мы можем считать, что $\xi_i \in \Phi_{0,\sigma_i^2}$, если σ_i различны и известны. В этом случае, обозначив через σ диагональную матрицу $\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$ и введя новые переменные $\xi' = \xi\sigma^{-1}$, $X' = X\sigma^{-1}$, $Y' = Y\sigma^{-1}$ (так что $\xi'_i = \xi_i/\sigma_i$, $\mathrm{x}'_i = \mathrm{x}_i/\sigma_i$, $\mathrm{y}'_i = \mathrm{y}_i/\sigma_i$), мы придем к задаче регрессии

$$Y' = \alpha X' + \xi',$$

в которой вектор наблюдений Y' и регрессор X' нам известны, $\xi' \in \Phi_{0,E_n}$. Легко проверить (мы предоставляем это читателю), что в этих условиях справедлив следующий аналог теоремы 1.

Теорема 2. *Оценка*

$$\alpha^* = Y\sigma^{-2}X^\top(D')^{-1}, \quad D' = X\sigma^{-2}X^\top,$$

является несмещенной эффективной оценкой α,

$$(\alpha^* - \alpha)(D')^{1/2} \in \Phi_{0,E_r},$$

$$|Y' - \alpha^* X'|^2 = \sum_{i=1}^{n} \frac{\left(\mathrm{y}_i - \sum\limits_{k=1}^{r} \alpha_k^* \mathrm{x}_{ik}\right)^2}{\sigma_i^2} \in \mathbf{H}_{n-r}.$$

Обратимся снова к теореме 1. Соотношения (10)–(12), установленные в ней, дают возможность строить доверительные множества как для отдельных координат θ, так и для вектора θ в целом. Например,

$$\mathbf{P}_\theta\left(\frac{(n-r)(\sigma^2)^*}{h^{(2)}_{\varepsilon/2}} < \sigma^2 < \frac{(n-r)(\sigma^2)^*}{h^{(1)}_{\varepsilon/2}}\right) = 1 - \varepsilon, \tag{17}$$

и если X_k ортогональны, то

$$\mathbf{P}_\theta\left(|\alpha_k - \alpha_k^*| < \frac{t_{\varepsilon/2}\sigma^*}{|X_k|}\right) = 1 - \varepsilon, \tag{18}$$

где $\mathbf{T}_{n-r}((-t_{\varepsilon/2}, t_{\varepsilon/2})) = 1 - \varepsilon$, $\mathbf{H}_{n-r}\big((h^{(1)}_{\varepsilon/2}, h^{(2)}_{\varepsilon/2})\big) = 1 - \varepsilon$.

Пусть X_k ортогональны. Обозначим через $\overline{\alpha}$ «подвектор» вектора α, определенный в следствии 1. В силу теоремы 1 доверительное множество для $\overline{\alpha}$ естественно строить с помощью соотношения

$$\frac{|(\overline{\alpha} - \overline{\alpha}^*)\overline{X}|^2}{(n-r)(\sigma^2)^*} < f_\varepsilon. \tag{19}$$

Значение f_ε, соответствующее заданному уровню $1-\varepsilon$, определяется известным образом (см. гл. 3) с помощью распределения $\mathbf{F}_{l,n-r}$ с числом степеней свободы $l, n-r$.

Если σ^2 известно, то доверительное множество будет определяться соотношением

$$|(\overline{\alpha} - \overline{\alpha}^*)\overline{X}|^2 < \sigma^2 h_\varepsilon, \tag{20}$$

где h_ε соответствует распределению $\mathbf{H}_l$.

В задачах регрессии может понадобиться также оценить значение поверхности регрессии $\mathrm{y} = \alpha \mathrm{z}^\top$ в новой, наперед заданной точке $\mathrm{z} = (\mathrm{z}_1, \ldots, \mathrm{z}_r) \in \mathbb{R}^r$. Положим $\mathrm{y}^* = \alpha^* \mathrm{z}^\top$. Тогда, как и раньше, находим

$$\mathrm{y}^* - \mathrm{y} = (\alpha^* - \alpha)\mathrm{z}^\top = \xi X^\top D^{-1} \mathrm{z}^\top \in \mathbf{\Phi}_{0,d^2},$$

$$d^2 = \sigma^2 \mathrm{z} D^{-1} \mathrm{z}^\top, \quad \frac{\mathrm{y}^* - \mathrm{y}}{d\sigma^*} \in \mathbf{T}_{n-r}.$$

Это дает возможность строить доверительные интервалы для y.

Следует отметить, что отыскание доверительной области для поверхности регрессии «в целом» есть задача более трудоемкая (ср. с [101]). Совокупность поверхностей, входящих в доверительное множество, будет определяться доверительным множеством для θ, построенным, например, с помощью (10), (11) (см. § 48). Подробнее об этом см. [101].

3. Проверка гипотез относительно линейной регрессии. Коснемся здесь двух типов задач.

1. Пусть известно, что представление (1), (2) имеет место. Требуется проверить гипотезу о том, что θ равно заданному значению θ' или что набор координат $\theta_{k_1}, \ldots, \theta_{k_l}$ равен набору $\theta'_{k_1}, \ldots, \theta'_{k_l}$, в то время как остальные координаты неизвестны.

Критерии для проверки таких гипотез удобно строить с помощью доверительных множеств (17)–(20) (см. § 48). Пусть, например, требуется проверить гипотезу H_1 о независимости Y от X для простейшей линейной регрессии, т. е. гипотезу $H_1 = \{\alpha_2 = 0\}$. Тогда из (18) (или из (14)) получим критерий уровня $1-\varepsilon$, отвергающий H_1, если

$$|\alpha_2^*| \geqslant \frac{t_{\varepsilon/2}\sigma^*}{|X_2|}. \tag{21}$$

В общем случае регрессии (1) с ортогональными X_k гипотеза о независимости Y от X будет иметь вид $H_1 = \{\overline{\alpha} = 0\}$, где $\overline{\alpha} = (\alpha_2, \ldots, \alpha_r)$, $x_{i1} \equiv 1$, и для ее проверки можно воспользоваться критерием

$$\frac{|\overline{\alpha}^* \overline{X}|^2}{(n-r)(\sigma^2)^*} \geqslant f_\varepsilon, \tag{22}$$

где $\overline{X}$ и f_ε определены в (19) при $l = r - 1$.

Можно использовать также подходы § 55, где рассматривалась проверка принадлежности выборки параметрическому подсемейству. Мы придем тогда к критерию отношения правдоподобия, который будет в известном смысле близок к (22). Если σ^2 известно, то к. о. п. для проверки $H_1 = \{\overline{\alpha} = 0\}$ будет иметь вид

$$\sigma^{-2}|\overline{\alpha}^* \overline{X}|^2 > h_\varepsilon,$$

где h_ε есть квантиль $\mathbf{H}_{r-1}$ порядка $1-\varepsilon$. Этот критерий будет минимаксным (см. § 49, 50) для соответствующим образом отделенных альтернатив.

2. Проверка гипотезы о наличии в выборке (X, Y) самой регрессии (1), (2). Под этим мы понимаем гипотезу о том, что при каких-нибудь α, σ имеет место представление (1), (2), т. е. что при каких-нибудь α, σ справедливо $\sigma^{-1}(Y - \alpha X) \in \mathbf{\Phi}_{0,E_n}$. Это есть задача о принадлежности Y параметрическому семейству. Но, как уже отмечалось, наблюдения в Y неодинаково распределены. Чтобы свести задачу к случаю одинаково распределенных наблюдений (см. § 57), мы воспользуемся следующим утверждением, дополняющим теорему 1. Будем считать X_k ортогональными.

Теорема 3. *Пусть C — любая ортогональная матрица порядка $n \times n$, содержащая в качестве первых r столбцов столбцы матрицы $X^\top D^{-1/2}$. Тогда вектор $\gamma = (Y - \alpha^* X)C$ имеет независимые координаты, обладающие свойством $\gamma_1 = \ldots = \gamma_r = 0$, $\gamma_i \in \mathbf{\Phi}_{0,\sigma^2}$, $i = r+1, \ldots, n$.*

Таким образом, задача сводится к проверке гипотезы о принадлежности выборки $\gamma_{r+1}, \ldots, \gamma_n$ объема $n - r$ семейству $\mathbf{\Phi}_{0,\sigma^2}$ (r наблюдений ушло, грубо говоря, на оценки для α). Эта задача рассматривалась в § 57. Чтобы получить значение γ_i, надо по выборкам X, Y вычислить последовательно значения α^*, $Y - \alpha^* X$ и применить к $Y - \alpha^* X$ любое преобразование C, обладающее отмеченными в теореме 3 свойствами.

Если σ известно, то мы придем к задаче о проверке простой гипотезы о принадлежности $\mathbf{\Phi}_{0,\sigma^2}$. Однако в этом случае для проверки интересующей нас гипотезы можно использовать и теорему 1, в силу которой

$$\frac{(n-r)(\sigma^2)^*}{\sigma^2} \in \mathbf{H}_{n-r}.$$

Доказательство теоремы 3. Если $Z \perp \mathcal{L}[X]$, то первые r координат вектора ZC образуют вектор $ZX^\top D^{-1/2} = 0$. Так как $(Y - \alpha^* X) \perp \mathcal{L}[X]$ и $\gamma = (Y - \alpha^* X)C$, то отсюда следует, что $\gamma_1 = \ldots = \gamma_r = 0$. Далее,

$$\gamma = (Y - \alpha X)C - (\alpha^* - \alpha)XC = \eta - \overline{\eta} D^{-1/2} XC,$$

где $\eta = \xi C$, $\overline{\eta} = (\eta_1, \ldots, \eta_r) = (\alpha^* - \alpha)D^{1/2} = \xi X^\top D^{-1/2}$ и, стало быть, γ есть результат линейного преобразования над η,

$$|\gamma|^2 = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha - \alpha^*)X|^2 = \sum_{i=1}^{n} \eta_i^2 - |\overline{\eta}|^2 = \sum_{i=r+1}^{n} \eta_i^2,$$

так что $\sum\limits_{i=r+1}^{n} \gamma_i^2 = \sum\limits_{i=r+1}^{n} \eta_i^2$. Это возможно только в том случае, когда $(\gamma_{r+1}, \ldots, \gamma_n)$ есть результат поворота вектора $(\eta_{r+1}, \ldots, \eta_n)$ или, что то же, результат ортогонального преобразования над $(\eta_{r+1}, \ldots, \eta_n)$. Так как $\sigma^{-1}\eta \in \mathbf{\Phi}_{0,E_n}$, то теорема доказана. $\triangleleft$

Пример 1. В этом примере мы опишем математическую сторону физического эксперимента, с помощью которого был обнаружен эффект распада φ-мезона на два π-мезона (см. [103]). Полученный результат носит статистический характер, и в нем, по существу, использовалась модель регрессии.

Исследование относится к изучению взаимодействия электронов (e^-) и позитронов (e^+) на встречных пучках. Если суммарная энергия этих частиц $2E$ находится в окрестности точки $2E_0 = 1019{,}6$ МэВ (рис. 8), то при

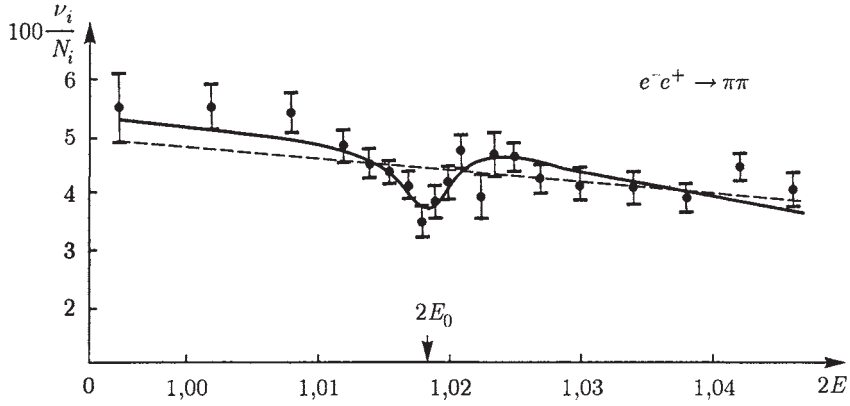

Рис. 8. Кривые изображают оценки линий регрессии при гипотезах H_1 и H_2

их «столкновении» в результате взаимодействия образуются (наряду с другими) частицы двух типов: φ-мезоны и пары π-мезонов. Вероятность возникновения пары π-мезонов при взаимодействии e^+, e^- в зависимости от энергии E описывается с большой точностью линейной функцией, которую представим в виде (гипотеза H_1)

$$p_1^{\pi\pi}(E) = a_0 + a_1 x, \quad x = E - E_0, \tag{23}$$

где a_0, a_1 неизвестны.

Было выдвинуто предположение (гипотеза H_2) о том, что при распаде родившихся φ-мезонов также могут появляться пары π-мезонов. Обнаружить непосредственно этот эффект практически невозможно, так как было установлено, что это явление, если оно происходит, то происходит очень редко,

не чаще чем один раз на 10^4 распадов φ-мезонов. Однако благодаря эффекту интерференции этого дополнительного канала рождения π-мезонов с основным каналом вероятность возникновения этих частиц будет равна не (23), а

$$p_2^{\pi\pi}(E) = [a_0 + a_1 x]\left[1 + \frac{b_0 + b_1 x + b_2 x^2}{x^2 + d^2}\right] \tag{24}$$

(как и (23), это есть весьма точное приближение несколько более сложной формулы, основанное на том, что рассматриваемый диапазон изменения $x = E - E_0$ мал по сравнению с E_0). В этом равенстве коэффициенты b_i, как и a_i, неизвестны, d известно.

Чтобы установить, какое из двух соотношений (23) или (24) на самом деле имеет место, было проведено $n = 20$ экспериментов при различных значениях энергии $E_1, \ldots, E_{20}$.

Таблица 1

Таблица экспериментальных данных

Номер эксперимента i	E_i, МэВ	N_i	ν_i	Номер эксперимента i	E_i, МэВ	N_i	ν_i
1	497,75	6960	384	11	510,40	14322	716
2	500,65	7908	435	12	510,92	13470	568
3	503,65	8102	432	13	511,39	12008	569
4	505,40	22259	1080	14	512,17	23951	1117
5	506,62	16938	765	15	513,20	27796	1185
6	507,66	21728	951	16	514,62	37771	1539
7	508,40	14014	603	17	516,58	25902	1036
8	508,90	13793	545	18	518,64	27857	1057
9	509,40	14075	615	19	520,61	23228	989
10	509,90	14867	691	20	522,88	26482	1066

Результатами экспериментов (см. табл. 1, рис. 8) являются количества N_i, $i = 1, \ldots, 20$, взаимодействий e^+, e^- и количества ν_i родившихся пар π-мезонов при энергии E_i. В каждом из проведенных экспериментов числа N_i и ν_i достаточно велики (N_i имеют порядок 10^4). Так как при фиксированном N_i число пар π-мезонов ν_i имеет распределение Бернулли $\mathbf{B}_{p_i}^{N_i}$ ($p_i = p_1^{\pi\pi}(E_i)$ при гипотезе H_1 и $p_i = p_2^{\pi\pi}(E_i)$ при гипотезе H_2), то, пользуясь нормальным приближением, мы вправе считать, что имеет место представление

$$\mathrm{y}_i \equiv \frac{\nu_i}{N_i} = p_i + \xi_i, \quad \xi_i \Subset \mathbf{\Phi}_{0,\sigma_i^2}$$

(в слагаемое ξ_i входят также случайные помехи (фон)). В силу (23), (24) мы будем иметь два возможных варианта регрессии:

$$p_i = \sum_{k=0}^{1} \alpha_k \psi_k(\mathrm{x}_i), \quad \psi_k(x) = x^k, \quad k = 0, 1 \tag{25}$$

(гипотеза H_1) и

$$p_i = \sum_{k=0}^{3} \alpha_k \psi_k(\mathrm{x}_i), \quad \psi_k(x) = \frac{x^k}{x^2 + d^2}, \quad k = 0, 1, 2, 3 \tag{26}$$

(гипотеза H_2).

Значения σ_i^2 при изменении гипотез меняются очень мало; они могут быть весьма точно оценены, и их можно считать известными. Тогда на основании теоремы 2 статистика

$$\chi^2 = \left| Y' - \sum_k \alpha_k^* \psi_k \right|^2 = \sum_{i=1}^{n} \frac{\left(y_i - \sum_k \alpha_k^* \psi_k(x_i) \right)^2}{\sigma_i^2} \tag{27}$$

будет иметь распределение $\mathbf{H}_{n-r}$, где r — число оцениваемых параметров α_k ($r = 2$ при гипотезе H_1 и $r = 4$ при гипотезе H_2).

Проводя необходимые вычисления в соответствии с рекомендациями теоремы 2, мы получим для статистики (27) значения: в первом случае ($r = 2$) $\chi_1^2 = 36{,}8$, во втором ($r = 4$) $\chi_2^2 = 19{,}0$. Фактически достигаемые уровни значимости (см. § 44) критерия $\chi^2 > c$ для проверки (в качестве основных) гипотез H_1 и H_2 составят соответственно $\mathbf{H}_{18}((0,\ 36{,}8)) = 0{,}9944$, $\mathbf{H}_{16}((0,\ 19{,}0)) = 0{,}731$.

Другими словами, предположение об отсутствии дополнительного канала рождения пар π-мезонов отвергается критерием, основанным на статистике χ^2, с уровнем значимости, равным, например, 0,99. В то же время предположение о наличии этого канала хорошо согласуется с экспериментальными данными.

Если быть более точным, то в этой задаче мы должны были бы проверять две сложные параметрические гипотезы, соответствующие предположениям (25), (26) для значений вероятностей появления пар π-мезонов. Если использовать критерий отношения правдоподобия, то он, как нетрудно проверить, будет основан на разности статистик χ^2, соответствующих моделям (25), (26), и, стало быть, результаты его будут примерно теми же.

4. Оценивание и проверка гипотез при наличии линейных связей. Рассмотрим, как и прежде, линейную регрессию (1), (2), но в предположении, что координаты вектора α связаны $s < r$ линейными соотношениями

$$\sum_{k=1}^{r} \alpha_k a_{kl} = c_l, \quad l = 1, \ldots, s.$$

В матричной форме эти соотношения можно записать в виде

$$\alpha A = c, \tag{28}$$

где A — матрица порядка $r \times s$. Мы будем предполагать, что A имеет ранг s. В этом случае мы могли бы выразить s переменных (скажем, $\alpha_{r-s+1}, \ldots, \alpha_r$) через остальные (т. е. через $\alpha_1, \ldots, \alpha_{r-s}$), подставить полученные значения в (1), (2) и получить снова стандартную задачу линейной регрессии (но с измененным регрессором).

Однако для дальнейшего нам будет удобнее подойти к решению этой задачи несколько иначе. Обратимся к доказательству теоремы 1. Подпространство $\mathcal{A}$ значений α, определяемое соотношениями (28), выделяет в $\mathcal{L}[X]$ подпространство размерности $r - s$ значений αX, которое обозначим через $\mathcal{L}_A[X]$. Очевидно, что оценивание $\alpha \in \mathcal{A}$ можно вести теперь теми же

приемами, которые мы использовали в теореме 1. Нужная оценка $\alpha_A^* \in \mathcal{A}$ будет определяться, как и в теореме 1, с помощью проекции $\alpha_A^* X$ вектора Y на $\mathcal{L}_A[X]$. Таким образом, наряду с соотношением $(Y - \alpha^* X) \perp \mathcal{L}[X]$ будем иметь соотношение $(Y - \alpha_A^* X) \perp \mathcal{L}_A[X]$, которое однозначно определяет α_A^*. Для получения самого значения α_A^* удобнее воспользоваться аналитическим подходом — применить метод неопределенных множителей Лагранжа для отыскания $\min\limits_{\alpha} |Y - \alpha X|^2$ при условии $\alpha A = c$. Для этого мы должны решить уравнения

$$\alpha A = c, \quad \frac{\partial}{\partial \alpha}[|Y - \alpha X|^2 + \lambda(\alpha A - c)^\top] = 0 \tag{29}$$

(мы используем множители $\lambda_1, \ldots, \lambda_s$, образующие вектор λ и соответствующие условиям (28)). Так как $|Y - \alpha X|^2 = (Y - \alpha X)(Y - \alpha X)^\top$, то второе из уравнений (29) примет вид

$$-2YX^\top - 2\alpha XX^\top + \lambda A^\top = 0.$$

Отсюда находим

$$\alpha_A^* = YX^\top D^{-1} - \frac{1}{2}\lambda A^\top D^{-1} = \alpha^* - \frac{1}{2}\lambda A^\top D^{-1}.$$

В силу (29) $c = \alpha_A^* A = \alpha^* A - \frac{1}{2}\lambda A^\top D^{-1} A$. Так как матрица D положительно определена, A имеет ранг s, то $B = D^{-1/2}A$ также будет иметь ранг s, а матрица $B^\top B = A^\top D^{-1} A$ будет также положительно определенной (см. п. 1). Следовательно,

$$-\frac{1}{2}\lambda = (c - \alpha^* A) D_A, \quad \alpha_A^* = \alpha^* + (c - \alpha^* A) D_A A^\top D^{-1}, \tag{30}$$

где мы положили для краткости $D_A = [A^\top D^{-1} A]^{-1}$.

Читатель может проверить, что это есть о. м. п. параметра α при условии $\alpha A = c$. Тот же результат (30) можно получить и из геометрических соображений, используя соотношения $\alpha_A^* X \in \mathcal{L}_A[X]$ и ортогональность

$$\begin{gathered} (Y - \alpha_A^* X) \perp \mathcal{L}_A[X], \\ (\alpha_A^* - \alpha^*)X = [(Y - \alpha^* X) - (Y - \alpha_A^* X)] \perp \mathcal{L}_A[X]. \end{gathered} \tag{31}$$

Обратимся теперь к задаче *проверки линейных гипотез*. Гипотезу H_1 относительно параметра α будем называть *линейной*, если она имеет вид $H_1 = \{\alpha A = c\}$, где матрицы A и c определены выше.

Сразу же можно отметить, что, введя новый параметр $\beta = \alpha A_c$, где A_c — любая невырожденная матрица, у которой первые s столбцов совпадают с A, мы сведем задачу к регрессии

$$Y = \beta X' + \xi, \quad X' = A_c^{-1} X, \tag{32}$$

и к проверке гипотезы $\overline{\beta} = c$, $\overline{\beta} = (\beta_1, \ldots, \beta_s)$ (см. п. 2).

Естественно также исходить из следующих соображений. Чем больше αA отличается от c, тем αX удаленнее от $\mathcal{L}_A[X]$ и тем больше точки αX, $\alpha^* X$ будут отличаться от $\alpha_A^* X \in \mathcal{L}_A[X]$. Поэтому в основу критерия для проверки H_1 естественно положить расстояние $\alpha_A^* X$ от $\alpha^* X$. Если гипотеза H_1 верна, то в силу (31)

$$\left|(\alpha_A^* - \alpha^*)X\right|^2 = \left|Y - \alpha_A^* X\right|^2 - |Y - \alpha^* X|^2. \tag{33}$$

В силу (30) (с заменой c на αA) $\alpha_A^* - \alpha^*$ есть результат линейного преобразования над $\alpha - \alpha^*$. Поэтому $(\alpha_A^* - \alpha^*)X$ не зависит от $Y - \alpha^* X$ (см. теорему 1).

Далее, в силу (30)

$$\begin{aligned}\left|(\alpha_A^* - \alpha^*)X\right|^2 &= (\alpha_A^* - \alpha^*)XX^\top(\alpha_A^* - \alpha^*)^\top = \\ &= (c - \alpha^* A)D_A(c - \alpha^* A) = (\alpha^* - \alpha)AD_A A^\top(\alpha^* - \alpha)^\top.\end{aligned} \tag{34}$$

Так как

$$(\alpha^* - \alpha)A = \xi X^\top D^{-1} A \Subset \mathbf{\Phi}_{0,\sigma^2 A^\top D^{-1} A} = \mathbf{\Phi}_{0,\sigma^2 D_A^{-1}},$$

то в силу (34) и § 12 (п. 4)

$$\frac{1}{\sigma^2}\left|(\alpha_A^* - \alpha^*)X\right|^2 \Subset \mathbf{H}_s. \tag{35}$$

Из сказанного и из теоремы 1 следует, что

$$\frac{\left|(\alpha_A^* - \alpha^*)X\right|^2}{|Y - \alpha^* X|^2} = \frac{\left|Y - \alpha_A^* X\right|^2}{|Y - \alpha^* X|^2} - 1 \Subset \mathbf{F}_{s,n-r}. \tag{36}$$

Соотношения (35), (36) дают нам возможность строить критерии (основанные на использовании удаленности $\alpha^* X$ от $\alpha_A^* X$) для проверки гипотезы H_1 соответственно в случаях, когда σ^2 известно и неизвестно (см. гл. 3).

Важно отметить, что гипотеза H_1 есть гипотеза о принадлежности α параметрическому подсемейству (при наличии мешающего параметра σ^2, если σ^2 неизвестно), а статистики (35), (36) есть не что иное, как статистики отношения правдоподобия (см. § 50, 55). Действительно, пусть, например, σ^2 неизвестно. Тогда (см. (5), (8))

$$\begin{aligned}\sup_{\alpha,\sigma} f_\theta(Y) &= \sup_{\alpha,\sigma}(\sqrt{2\pi}\,\sigma)^{-n}\exp\left\{-\frac{|Y - \alpha X|^2}{2\sigma^2}\right\} = \\ &= (\sqrt{2\pi}\,\widehat{\sigma}^*)^{-n}\exp\left\{-\frac{|Y - \alpha^* X|^2}{2(\sigma^2)^*}\right\} = \left(\sqrt{2\pi}\frac{|Y - \alpha^* X|}{n}\right)^{-n} e^{-n/2}.\end{aligned}$$

Значение $\sup\limits_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)$ вычисляется точно так же. Надо лишь заметить, что о. м. п. для α в случае $\alpha \in \mathcal{A}$ будет равна α_A^*, а о. м. п. для σ^2 будет равна

аналогично (8) $\frac{1}{n}|Y-\alpha_A^*X|^2$. Поэтому

$$\sup_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)=\left(\sqrt{2\pi}\frac{|Y-\alpha_A^*X|}{n}\right)^{-n}e^{-n/2},$$

$$\frac{\sup\limits_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)}{\sup\limits_{\alpha,\sigma} f_\theta(Y)}=\frac{|Y-\alpha^*X|^n}{|Y-\alpha_A^*X|^n},$$

и, стало быть, статистика критерия отношения правдоподобия эквивалентна (36).

Если σ^2 известно, то в основу критерия для проверки H_1 можно положить соотношение (35). Читатель может аналогично предыдущему убедиться, что это тоже есть критерий отношения правдоподобия. Так как этот критерий инвариантен относительно замены параметра (см. § 50), то в силу замечания и утверждений § 49, 50 можно утверждать, что к. о. п.

$$|(\alpha_A^*-\alpha^*)X|^2>\sigma^2h_\varepsilon,$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения $\mathbf{H}_s$, будет минимаксным критерием уровня $1-\varepsilon$ для проверки H_1 против отделенных соответствующим образом альтернатив.

Изложенное выше и результаты глав 2, 3 (см., в частности, § 55) дают основание считать, что критерий (36), как и оценка (30), также будут обладать свойствами оптимальности. Подробнее на этом мы останавливаться не будем. Весьма полное изложение задач регрессии можно найти в [101].

§ 62. Дисперсионный анализ

Задачи дисперсионного анализа, излагаемые ниже, по своему существу относятся к задачам регрессии. В последних изучалась зависимость наблюдений от *числового* фактора x, который мог принимать любые наперед заданные значения $x_1,\dots,x_n$ и каждому из них соответствовало одно наблюдение. В задачах дисперсионного анализа обычно изучается воздействие лишь *дискретных* факторов (одного, двух или более), которые могут принимать лишь конечное число значений. При каждом из этих значений мы располагаем набором наблюдений (выборкой). Дисперсионный анализ объединяет группу статистических приемов, основанных на анализе среднеквадратических уклонений и предназначенных для проверки различных гипотез и оценки параметров, связанных с воздействием факторов. Основы дисперсионного анализа были заложены Фишером.

1. Задачи дисперсионного анализа как задачи регрессии. Случай одного фактора. Пусть дано r независимых выборок

$$Y_1=(y_{11},\dots,y_{1n_1}),\dots,Y_r=(y_{r1},\dots,y_{rn_r})$$

объемов $n_1,\dots,n_r$ из нормальных совокупностей $Y_k\in\Phi_{\alpha_k,\sigma^2}$. Предполагается, что наблюдения Y_k, $k=1,\dots,r$, проводились при разных значениях

некоторого фактора, роль которого нас интересует, и что влияние этого фактора сказывается на значении среднего α_k. Значение дисперсии σ^2 предполагается одним и тем же для всех выборок и, как правило, неизвестным. Задачи дисперсионного анализа включают в себя проверку гипотез, относящихся к значениям $\alpha_1, \ldots, \alpha_r$, и, в частности, гипотезы об однородности $\alpha_1 = \ldots = \alpha_r = \alpha$ (эта последняя задача рассматривалась нами в § 59), а также оценки параметров α_k и их изменчивости.

Дисперсионный анализ, как и задачи регрессии вообще, имеет чрезвычайно широкое поле применений, особенно в социологии, сельском хозяйстве, биологии, медицине. В качестве весьма типичной задачи для применения методов дисперсионного анализа можно назвать, например, проблему о выяснении зависимости содержания холестерина в крови человека от его профессии.

Сформулированные выше задачи дисперсионного анализа есть частные случаи задач линейной регрессии. Действительно, наблюдения y_{ki} можно представить в форме

$$y_{ki} = \alpha_k + \xi_{ki}, \quad \xi_{ki} \in \Phi_{0,\sigma^2}, \quad k = 1, \ldots, r, \quad i = 1, \ldots, n_k. \tag{1}$$

Образуем вектор

$$Y = (y_{11}, \ldots, y_{1n_1}; y_{21}, \ldots, y_{2n_2}; \ldots; y_{r1}, \ldots, y_{rn_r})$$

и вектор ξ по тому же правилу. Тогда соотношения (1) можно записать в матричной форме $Y = \alpha X + \xi$, где X есть матрица размера $r \times n$, $n = n_1 + \ldots + n_r$, вида

$$X = \left(\begin{array}{cccc|cccc|c|cccc} 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 & & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 & & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & & 1 & 1 & \ldots & 1 \end{array}\right).$$

Очевидно, что строки этой матрицы (векторы X_j) ортогональны. Гипотезу $H_1 = \{\alpha_1 = \alpha_2 = \ldots = \alpha_r\}$ можно записать в виде

$$\alpha A = 0, \tag{2}$$

где A есть матрица размера $r \times (r-1)$

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \\ -1 & -1 & \ldots & -1 \end{pmatrix}.$$

Очевидно, что A имеет ранг $r-1$.

Мы видим, что проверка основной гипотезы H_1 дисперсионного анализа есть не что иное, как задача о проверке линейной гипотезы для регрессии.

Выясним, что собой представляют эффективные оценки для α и σ^2, найденные в теореме 61.1. В нашем случае $|X_k|^2 = n_k$, матрица $D = XX^\top$

порядка $r \times r$ имеет вид

$$D = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_r \end{pmatrix},$$

$$\alpha_k^* = \frac{(Y, X_k)}{(X_k, X_k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ki} \equiv \overline{y}_{k\cdot}, \tag{3}$$

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y}_{k\cdot})^2 \equiv Q_2(Y).$$

При этом $\alpha_1^*, \dots, \alpha_r^*$, $(\sigma^2)^*$ независимы. Доверительные интервалы для параметров α, σ^2 и функций от них строятся так же, как и в § 61.

Для проверки линейной гипотезы (2) нам нужно вычислить также о. м. п. α_A^* при наличии условия (2) (см. п. 4 предыдущего параграфа). Простейший путь здесь — использовать подход, изложенный в начале п. 61.3, и выразить $\alpha_1, \dots, \alpha_r$ через независимые переменные. В нашем случае независимая переменная лишь одна — пусть это будет $\alpha_r = \mu$, и $\alpha_A^* = (\mu^*, \dots, \mu^*)$, где μ^* минимизирует

$$|Y - (\mu, \dots, \mu)X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \mu)^2.$$

Очевидно, что

$$\mu^* = \overline{y} \equiv \frac{1}{n} \sum_{k=1}^{r} \sum_{i=1}^{n_k} y_{ki},$$

$$\left|Y - \alpha_A^* X\right|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y})^2 \equiv Q(Y) =$$

$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y} + \overline{y}_{k\cdot} - \overline{y}_{k\cdot})^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \overline{y}_{k\cdot})^2 + \sum_{k=1}^{r} n_k (\overline{y}_{k\cdot} - \overline{y})^2$$

(сумма смешанных произведений равна нулю, так как $\sum_{i=1} (y_{ki} - \overline{y}_{k\cdot}) = 0$).

Если гипотеза H_1 верна, то в силу (61.33), (3) и только что полученного равенства

$$\left|(\alpha_A^* - \alpha^*)X\right|^2 = Q(Y) - Q_2(Y) = \sum_{k=1}^{r} n_k (\overline{y}_{k\cdot} - \overline{y})^2 \equiv Q_1(Y).$$

На основании (61.36) при выполнении H_1 получаем $Q_1(Y)/Q_2(Y) \in \mathbf{F}_{r-1,n-r}$, что дает нам возможность строить критерий $Q_1(Y)/Q_2(Y) > f_\varepsilon$ (f_ε — квантиль $\mathbf{F}_{r-1,n-r}$ порядка $1-\varepsilon$) для проверки H_1, который будет к. о. п. Если

σ^2 известно, к. о. п. будет иметь вид

$$Q_1(Y) > \sigma^2 h_\varepsilon$$

(h_ε — квантиль $\mathbf{H}_{r-1}$) и будет минимаксным критерием при соответствующим образом отделенных альтернативах (см. § 49).

2. Влияние двух факторов. Элементарный подход. В задачах этого раздела исследуется влияние на результаты эксперимента факторов двух видов. Применительно, скажем, к сельскому хозяйству это может быть изучение влияния на урожайность состава почвы (фактор A принимает r значений) и способа обработки (фактор B принимает s значений).

Здесь наблюдения можно представить в форме

$$\begin{gathered} \mathrm{y}_{kli} = \alpha_{kl} + \xi_{kli}, \quad \xi_{kli} \in \mathbf{\Phi}_{0,\sigma^2}, \\ k = 1, \ldots, r, \quad l = 1, \ldots, s, \quad i = 1, \ldots, n_{kl}, \end{gathered} \tag{4}$$

и рассматриваемая модель, по существу, ничем не будет отличаться от модели (1), рассмотренной в п. 1. Стало быть, здесь также применимы все результаты § 61, но прямое их применение становится более громоздким. Громоздким является уже само присутствие тройных индексов. Чтобы несколько упростить задачу, положим $n_{kl} \equiv 1$, это избавит нас от одного индекса (индекса i в (4)). Кроме того, мы предложим в этом пункте несколько иной, элементарный подход, который независимо от теорем § 61 позволит получить утверждения, необходимые для проверки основных гипотез.

Итак, будем рассматривать выборки $Y_{kl} = \mathrm{y}_{kl}$ единичного объема, так что набор экспериментальных данных Y здесь будет представлять собой матрицу $r \times s$ чисел y_{kl}, определяющих результат эксперимента под воздействием k-го значения фактора A и l-го значения фактора B. Эту матрицу можно трактовать как r выборок (строк) объема s, соответствующих разным значениям фактора A, или как s выборок (столбцов) объема r, соответствующих разным значениям фактора B. В соответствии с этим в дальнейшем и будет происходить группировка наблюдений. Положим

$$\overline{\mathrm{y}}_{k\cdot} = \frac{1}{s}\sum_{l=1}^{s} \mathrm{y}_{kl}, \quad \overline{\mathrm{y}}_{\cdot l} = \frac{1}{r}\sum_{k=1}^{r} \mathrm{y}_{kl}, \quad \overline{\mathrm{y}} = \frac{1}{rs}\sum_{k,l} \mathrm{y}_{kl}.$$

Справедливо тождество

$$Q(Y) \equiv \sum_{k,l} (\mathrm{y}_{kl} - \overline{\mathrm{y}})^2 = Q_1(Y) + Q_2(Y) + Q_3(Y), \tag{5}$$

где

$$Q_1(Y) = s\sum_{k} (\overline{\mathrm{y}}_{k\cdot} - \overline{\mathrm{y}})^2, \quad Q_2(Y) = r\sum_{l} (\mathrm{y}_{\cdot l} - \overline{\mathrm{y}})^2,$$

$$Q_3(Y) = \sum_{k,l} (\mathrm{y}_{kl} - \overline{\mathrm{y}}_{k\cdot} - \overline{\mathrm{y}}_{\cdot l} + \overline{\mathrm{y}})^2.$$

Предположим, что влияние факторов аддитивно, т. е. существуют a_k, b_l такие, что

$$\alpha_{kl} = a_k + b_l, \quad k = 1, \ldots, r, \ l = 1, \ldots, s. \tag{6}$$

Очевидно, что Q_1 определяет изменчивость значений a_k (т. е. связано с фактором A), Q_2 определяет изменчивость b_l (фактор B), а Q_3 есть сумма, создаваемая всецело случайностью. Очевидно также, что

$$Q_i(Y+a) = Q_i(Y), \quad i = 1, 2, 3. \tag{7}$$

Теорема 1. 1) *Справедливо*

$$\frac{Q_3(Y)}{\sigma^2} \in \mathbf{H}_{(r-1)(s-1)}. \tag{8}$$

2) *Если верна гипотеза* $H_A = \{a_1 = \ldots = a_r = a\}$, *то* $Q_1(Y)$ *не зависит от* $Q_2(Y)$ *и* $Q_3(Y)$, $Q_1(Y)/\sigma^2 \in \mathbf{H}_{r-1}$. *Аналогичное утверждение имеет место относительно* Q_2 *и гипотезы* $H_B = \{b_1 = \ldots = b_s = b\}$.

3) *Если справедлива гипотеза* $H_1 = \{\alpha_{kl} = \alpha\}$, *то все три квадратичные формы* Q_1, Q_2, Q_3 *независимы.*

Доказательство. Положим, не ограничивая общности, $\sigma^2 = 1$. Тогда

$$\mathbf{E}y_{kl}y_{ij} = \begin{cases} \alpha_{kl}\alpha_{ij}, & \text{если} \quad (i,j) \neq (k,l), \\ \alpha_{kl}^2 + 1, & \text{если} \quad (i,j) = (k,l). \end{cases}$$

Отсюда следует, что

$$\mathbf{E}\Big(\sum_{\mathrm{I}} y_{kl}\Big)\Big(\sum_{\mathrm{II}} y_{kl}\Big) = \Big(\sum_{\mathrm{I}} \alpha_{kl}\Big)\Big(\sum_{\mathrm{II}} \alpha_{kl}\Big) + m,$$

где m — число одинаковых слагаемых в суммах $\sum\limits_{\mathrm{I}}$ и $\sum\limits_{\mathrm{II}}$. Пользуясь этим равенством, легко получить теперь, что

$$\mathbf{E}(\overline{y}_{k\cdot} - \overline{y})(\overline{y}_{\cdot l} - \overline{y}) = (\alpha_{k\cdot} - \overline{\alpha})(\alpha_{\cdot l} - \overline{\alpha}) = (a_k - \overline{a})(b_l - \overline{b}) \tag{9}$$

при естественных соглашениях относительно обозначений $\alpha_{k\cdot}$, $\alpha_{\cdot l}$, $\overline{\alpha}$, $\overline{a}$, $\overline{b}$. Если верна гипотеза $H_A = \{\alpha_1 = \ldots = a_r = a\}$, то математическое ожидание в (9) равно нулю. Так как при этом $\mathbf{E}(\overline{y}_{k\cdot} - \overline{y}) = \alpha_{k\cdot} - \overline{\alpha} = 0$, то установленный факт означает, что набор случайных величин $\{\overline{y}_{k\cdot} - \overline{y}\}$ не зависит от $\{\overline{y}_{\cdot l} - \overline{y}\}$.

Аналогично устанавливаем, что при любых k, l, i

$$\mathbf{E}(y_{kl} - \overline{y}_{k\cdot})(\overline{y}_{i\cdot} - \overline{y}) = 0.$$

Это значит, что совокупность $\{\overline{y}_{k\cdot} - \overline{y}\}$ не зависит также от $\{y_{kl} - \overline{y}_{k\cdot} - \overline{y}_{\cdot l} + \overline{y}\}$. Это означает в свою очередь, что $Q_1(Y)$ при выполнении H_A не зависит от $Q_2(Y)$ и $Q_3(Y)$. Тот факт, что $Q_1(Y) \in \mathbf{H}_{r-1}$, вытекает из леммы Фишера (§ 40).

Аналогично обстоит дело, если выполнена гипотеза H_B. Если же справедлива гипотеза H_1 (т. е. справедливы H_A и H_B), то, очевидно, все три названных выше набора случайных величин будут независимы. Это означает независимость $Q_1(Y)$, $Q_2(Y)$, $Q_3(Y)$.

Нам осталось найти распределение $Q_3(Y)$. Так как это распределение от a_k, b_l не зависит, то мы можем считать, что $a_k = b_l = 0$ для всех k, l

и что, стало быть, выполнена H_1. Тогда из определения $Q(Y)$ следует, что $Q(Y) \in \mathbf{H}_{rs-1}$. Кроме того, справедливо (5), где $Q_1(Y) \in \mathbf{H}_{r-1}$, $Q_2(Y) \in \in \mathbf{H}_{s-1}$. Остается воспользоваться независимостью $Q_i(Y)$ и леммой 61.1. Теорема доказана. $\triangleleft$

Аналогичный подход может быть применен и к задачам п. 1.

Из теоремы 1 вытекает возможность построения следующих статистических процедур.

1. Оценка параметров $a_k - a_i$, $b_l - b_j$, σ^2 (числа a_k, b_l в (6) определены с точностью до последнего слагаемого) с помощью оценок $\overline{y}_{k\cdot} - \overline{y}_{i\cdot}$, $\overline{y}_{\cdot l} - \overline{y}_{\cdot j}$, $(\sigma^2)^* = Q_3(Y)/(r-1)(s-1)$. Так как, по существу, проведенные выше рассмотрения совпадают с тем, что мы делали в § 61 и в п. 1 настоящего параграфа, то названные оценки будут эффективными. Доверительные интервалы для σ^2, $a_k - a_i$ можно строить с помощью соотношений (8),

$$\overline{y}_{k\cdot} - \overline{y}_{i\cdot} - (a_k - a_i) \in \mathbf{\Phi}_{0,2\sigma^2/s},$$

$$\frac{\overline{y}_{k\cdot} - \overline{y}_{i\cdot} - (a_k - a_i)}{\sqrt{\dfrac{2Q_3(Y)}{s(r-1)(s-1)}}} \in \mathbf{T}_{(r-1)(s-1)}$$

(для $b_l - b_j$ все происходит аналогично).

2. Проверка гипотезы H_A с помощью критерия $Q_1/Q_3 > f_\varepsilon$. Критерий будет иметь уровень $1-\varepsilon$, если f_ε есть квантиль порядка $1-\varepsilon$ распределения $\mathbf{F}_{r-1,(r-1)(s-1)}$.

Аналогичный вид будет иметь критерий для проверки H_B: $Q_2/Q_3 > f_\varepsilon$, где f_ε — квантиль порядка $1-\varepsilon$ распределения $\mathbf{F}_{s-1,(r-1)(s-1)}$. $\triangleleft$

3. Проверка гипотезы H_1 с помощью критерия

$$\frac{Q_1 + Q_2}{Q_3} > f_\varepsilon$$

уровня $1-\varepsilon$, где f_ε — квантиль порядка $1-\varepsilon$ распределения $\mathbf{F}_{r+s-2,(r-1)(s-1)}$.

Более подробно задачи дисперсионного анализа рассмотрены в [101, 136].

§ 63. Распознавание образов

В этом параграфе мы коснемся кратко круга задач, для обозначения которого наряду с названием «распознавание образов» иногда используются также термины «классификация» и «дискриминантный анализ» *).

В § 41 мы рассматривали следующую задачу о проверке r простых гипотез. Даны распределения $\mathbf{P}_1, \ldots, \mathbf{P}_r$ и выборка X объема n. Требуется определить, какая из гипотез

$$H_j = \{X \in \mathbf{P}_j\} \tag{1}$$

верна.

*) Следует отметить, что последние два термина употребляются также и для обозначения иных задач, например задач, в которых распределения $\mathbf{P}_i$ в (1) известны.

Однако в практических задачах распределения $\mathbf{P}_j$ часто неизвестны, и судить о них мы можем опять-таки лишь по выборкам.

Итак, пусть дано r выборок $X_i = (\mathrm{x}_{i1}, \ldots, \mathrm{x}_{in_i})$, $i = 1, \ldots, r$, объемов $n_1, \ldots, n_r$ соответственно, отвечающих r различным *неизвестным* нам распределениям $\mathbf{P}_1, \ldots, \mathbf{P}_r$, и пусть, кроме того, дана выборка X. Надо решить все ту же задачу: определить, какая из гипотез (1) верна. Другими словами, надо установить, продолжением какой из выборок $X_1, \ldots, X_r$ является выборка X. Это и есть задача распознавания образов.

Для простоты мы ограничимся рассмотрением случая $r = 2$.

1. Параметрический случай. Предположим сначала, что $\mathbf{P}_i$ принадлежат некоторому семейству $\{\mathbf{P}_\theta\}$, удовлетворяющему условию (A_μ), т. е. $X_1 \in \mathbf{P}_{\theta_1}$, $X_2 \in \mathbf{P}_{\theta_2}$, $X \in \mathbf{P}_\theta$ при некоторых $\theta_1 \neq \theta_2$ и $\theta = \theta_1$ или $\theta = \theta_2$. Первое из этих равенств соответствует гипотезе $H_1 = \{X \in \mathbf{P}_{\theta_1}\}$, второе — гипотезе $H_2 = \{X \in \mathbf{P}_{\theta_2}\}$.

Предположим, далее, опять-таки для простоты, что объемы выборок n_1, n_2, n равны: $n_1 = n_2 = n$.

Рассмотрим объединенную выборку (X_1, X_2, X) и представим ее как выборку объема n, образованную наблюдениями $(\mathrm{x}_{1i}, \mathrm{x}_{2i}, \mathrm{x}_i)$ и относящуюся к распределению $\mathbf{P}_{\theta_1} \times \mathbf{P}_{\theta_2} \times \mathbf{P}_\theta$, имеющему плотность $f_{\theta_1}(x_1) f_{\theta_2}(x_2) f_\theta(x)$, зависящую от параметра $\overline{\theta} = (\theta_1, \theta_2, \theta)$. Очевидно, что функция правдоподобия выборки (X_1, X_2, X) будет равна

$$f_{\overline{\theta}}(X_1, X_2, X) = f_{\theta_1}(X_1) f_{\theta_2}(X_2) f_\theta(X).$$

Мы приходим к задаче о проверке гипотезы H_1 о том, что параметр $\overline{\theta}$ лежит на «кривой» $\theta = \theta_1$, против гипотезы H_2 о том, что $\overline{\theta}$ располагается на другой «кривой» $\theta = \theta_2$. Это есть задача проверки гипотезы о принадлежности параметрическому подсемейству (см. § 55), однако в том случае, когда конкурирующая гипотеза означает принадлежность другому параметрическому подсемейству. Рассмотрение этой задачи происходит аналогично § 55, но по своей технической трудности выходит за рамки настоящей книги. Мы ограничимся здесь тем, что для случая одномерного параметра θ опишем вкратце существо результата. Оно вполне аналогично содержанию § 55: если параметр θ локализован, т. е. точки θ_1, θ_2 располагаются в окрестности некоторой точки θ_0, $|\theta_1 - \theta_2| > b/\sqrt{n}$, и если семейство $\{\mathbf{P}_\theta\}$ удовлетворяет в точке θ_0 условиям регулярности (RR), то критерий отношения правдоподобия

$$\frac{\sup\limits_{\theta_1, \theta_2} f_{\theta_1}(X_1) f_{\theta_2}(X_2) f_{\theta_2}(X)}{\sup\limits_{\theta_1, \theta_2} f_{\theta_1}(X_1) f_{\theta_2}(X_2) f_{\theta_1}(X)} > c \tag{2}$$

будет при $n \to \infty$ асимптотически минимаксным для проверки H_1 против H_2.

Ограничение $n_1 = n_2 = n$ несущественно. Оно устраняется так же, как и в рассмотрениях § 59.

2. Общий случай. В общем случае, когда у нас нет оснований предполагать, что X_i связаны с параметрическим семейством, возможен общий подход, в основе которого лежат те же соображения, которые использовались

нами при построении в § 60 критериев однородности. В этом случае критерий π для проверки H_1 против H_2 будет функцией трех выборок, так что $\pi = \pi(X_1, X_2, X)$ есть вероятность принять H_2 при данных (X_1, X_2, X). Как и прежде, нерандомизированный критерий определяется критической областью $\Omega \subset \mathcal{X}^{n_1+n_2+n}$ в пространстве значений (X_1, X_2, X). Уровнем значимости критерия естественно называть число

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathcal{P}, \mathbf{P}_2 \in \mathcal{P}} \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_1((X_1, X_2, X) \notin \Omega),$$

где $\mathcal{P}$ есть класс допустимых распределений. Значение

$$\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2 \times \mathbf{P}_2((X_1, X_2, X) \in \Omega),$$
$$\mathbf{P}_1 \in \mathcal{P}, \quad \mathbf{P}_2 \in \mathcal{P},$$

есть мощность критерия в точке $(\mathbf{P}_1, \mathbf{P}_2)$.

Критерий π называется состоятельным, если $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ при $n_1 \to \infty$, $n_2 \to \infty$, $n \to \infty$ при любых $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathcal{P}$, $\mathbf{P}_2 \in \mathcal{P}$.

В качестве основы для построения состоятельных критериев можно использовать хорошо известный нам факт о сближении эмпирических распределений $\mathbf{P}^*_{X_1}$ и $\mathbf{P}^*_{X_2}$ для выборок X_1 и X_2 с $\mathbf{P}_1$ и $\mathbf{P}_2$ соответственно. Если $d(\mathbf{P}, \mathbf{Q})$ есть некоторое расстояние между распределениями, то при гипотезе H_2 расстояние $d\big(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X\big)$ должно быть меньше, чем $d\big(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X\big)$. Поэтому в качестве критерия можно использовать неравенство

$$d\big(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X\big) - d\big(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X\big) < c,$$

при выполнении которого принимается H_2. Расчет такого рода критериев (вычисление их уровней значимости и мощности) связан обычно с большими трудностями (ср. с более простыми задачами, изложенными в § 60).

Используя *группировку наблюдений*, мы можем в общем случае применять асимптотически оптимальный критерий (2). Предположим, что группировка по областям $\Delta_1, \ldots, \Delta_m$ произведена и $(\nu_{i1}, \ldots, \nu_{im})$ и $(\nu_1, \ldots, \nu_m)$ есть частоты попаданий в эти области наблюдений выборок X_i, $i = 1, 2$, и X соответственно. Пусть, кроме того, $\theta_i = (\theta_{i1}, \ldots, \theta_{im})$ есть вероятности $(\mathbf{P}_i(\Delta_1), \ldots, \mathbf{P}_i(\Delta_m))$ попаданий в области $\Delta_1, \ldots, \Delta_m$ для распределений $\mathbf{P}_i$, $i = 1, 2$. Так как для сгруппированной выборки X_i, $i = 1, 2$, функция правдоподобия $f_{\theta_i}(X_i)$ равна $f_{\theta_i}(X_i) = \prod_{k=1}^{m} \theta_{ik}^{\nu_{ik}}$, то критерий (2) будет иметь вид

$$\sup_{\theta_2} \sum_{k=1}^{m} (\nu_{2k} + \nu_k) \ln \theta_{2k} + \sup_{\theta_1} \sum_{k=1}^{m} \nu_{1k} \ln \theta_{1k} -$$
$$- \sup_{\theta_1} \sum_{k=1}^{m} (\nu_{1k} + \nu_k) \ln \theta_{1k} - \sup_{\theta_2} \sum_{k=1}^{m} \nu_{2k} \ln \theta_{2k} > \ln c$$

или

$$\sum_{k=1}^{m}(\nu_{2k}+\nu_k)\ln\frac{\nu_{2k}+\nu_k}{n_2+n}+\sum_{k=1}^{m}\nu_{1k}\ln\frac{\nu_{1k}}{n_1}>$$

$$>\ln c+\sum_{k=1}^{m}(\nu_{1k}+\nu_k)\ln\frac{\nu_{1k}+\nu_k}{n_1+n}+\sum_{k=1}^{m}\nu_{2k}\ln\frac{\nu_{2k}}{n_2}. \quad (3)$$

Аналогичные рассмотрения могут быть проведены и для $r>2$.

Глава 5

СТАТИСТИКА РАЗНОРАСПРЕДЕЛЕННЫХ НАБЛЮДЕНИЙ

В § 64 приведены примеры задач, которые помогают сформулировать общую постановку проблемы.

В § 65, 66 излагаются основные методы построения оценок в случае разнораспределенных наблюдений. Изучаются асимптотические свойства M-оценок и устанавливается асимптотическая оптимальность оценок максимального правдоподобия в классе M-оценок.

Параграф 67 посвящен достаточным статистикам и построению эффективных оценок на их основе.

В § 68 рассмотрена задача об оценке «хвостов» правильно меняющихся распределений.

В § 69 установлены аналоги неравенства Рао–Крамера для случая разнораспределенных наблюдений. Найдены критерии эффективности оценок, основанные на этом неравенстве.

Параграф 70 посвящен анализу отношения правдоподобия и уточнению асимптотических свойств оценок максимального правдоподобия.

В § 71 рассмотрены задачи проверки гипотез для разнораспределенных наблюдений.

Параграф 72 посвящен задаче о разладке.

§ 64. Предварительные замечания. Примеры

В этой главе мы будем рассматривать задачи математической статистики, для которых экспериментальные данные представлены в виде выборки $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, состоящей, в отличие от глав 1–3, из наблюдений x_i разной природы. В какой-то мере мы уже сталкивались с такими задачами в гл. 4, например в задачах линейной регрессии об оценке параметров $(\alpha_1, \ldots, \alpha_r)$ по выборке Y, состоящей из наблюдений $\mathrm{y}_i = \alpha_1 \mathrm{x}_{i,1} + \ldots$ $\ldots + \alpha_r \mathrm{x}_{i,r} + \xi_i$, $i = 1, \ldots, n$, где векторы $\mathrm{x}_i = (\mathrm{x}_{i,1}, \ldots, \mathrm{x}_{i,r})$ различны при разных i, $\xi_i \in \mathbf{\Phi}_{0,\sigma^2}$. Эти задачи оказалось возможным в значительной мере редуцировать к традиционным подходам, изложенным в гл. 1–3. Такая редукция возможна и в ряде других задач, внешне формулируемых в терминах разнораспределенных (неоднородных) наблюдений, но далеко не всегда.

Рассмотрим несколько примеров, которые с разных сторон характеризуют рассматриваемые в этой главе задачи.

Пример 1. Пусть надо оценить параметры (α, σ^2) по выборке $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, где $\mathrm{x}_i \in \mathbf{\Phi}_{\alpha t_i, \sigma^2 u_i}$, t_i, u_i известны, $i = 1, \ldots, n$. Если α известно и надо оценить лишь σ^2, то можно рассмотреть преобразованные наблюдения $\dfrac{\mathrm{x}_i - \alpha t_i}{\sqrt{u_i}} \in \mathbf{\Phi}_{0,\sigma^2}$, которые позволяют все свести к стандартной

задаче об оценке σ^2 по однородным наблюдениям. Если σ^2 неизвестно и надо оценить α, то такого сведéния, вообще говоря, осуществить не удается, так как $\mathrm{x}_i/\sigma\sqrt{u_i}$ при $\alpha \neq 0$ и $u_i \neq t_i^2$ имеют разные распределения. Если $u_i = t_i^2$, $i = 1, \ldots, n$, то $\mathrm{x}_i/\sqrt{u_i} \in \mathbf{\Phi}_{\alpha,\sigma^2}$ и сведéние вновь возможно.

Для пуассоновских наблюдений $\mathrm{x}_i \in \mathbf{\Pi}_{\lambda t_i}$, где t_i известны, использование линейных преобразований не позволяет получить требуемую редукцию к случаю однородных наблюдений, так как x_i/t_i $\left(\mathbf{E}\dfrac{\mathrm{x}_i}{t_i} = \lambda\right)$ не имеют распределения Пуассона (эти значения даже не обязаны быть целочисленными).

Пример 2. Нелинейная регрессия. Так называются задачи об оценке неизвестного параметра θ (или о проверке соответствующих параметрических гипотез), когда наблюдения x_i имеют вид

$$\mathrm{x}_i = \varphi(\theta, \mathrm{z}_i) + \xi_i \tag{1}$$

или

$$\mathrm{x}_i = \varphi(\theta, \mathrm{z}_i)(1 + \xi_i), \tag{2}$$

где $\varphi(\theta, z)$ — известная функция; $z_1, z_2, \ldots$ — суть значения «регрессора», при которых проводятся наблюдения; распределения ξ_i известны или $\xi_i \in \mathbf{P}_\theta$. В простейшем случае в задаче (1) предполагается, что $\xi_i \in \mathbf{\Phi}_{0,\sigma^2}$ (σ^2 может быть неизвестным). Если $\mathrm{x}_i > 0$, $\varphi(\theta, \mathrm{z}_i) > 0$, то путем логарифмирования задача (2) может быть сведена к (1). Линейная регрессия, рассмотренная в § 61, является, очевидно, частным случаем задачи (1). Вместо $\varphi(\theta, \mathrm{z}_i)$ в (1), (2) можно писать также $\varphi_i(\theta)$, где функции φ_i известны.

Пусть, например, в пространстве движутся прямолинейно и равномерно два тела A и B, сближаясь друг с другом. Тело A может в моменты времени $t_i = id$, $i = 1, \ldots, n$, зондировать с помощью локатора расстояния до тела B, которые получаются «зашумленными» из-за наличия ошибок ξ_i. По данным локации надо оценить расстояние h, на котором тело A пройдет мимо тела B. Не ограничивая общности, можно считать, что тело B неподвижно; тело A движется со скоростью v, которая также может быть неизвестной (см. рис. 9). Тогда наблюдение x_i будет иметь вид

$$\mathrm{x}_i = \sqrt{(\lambda - ivd)^2 + h^2} + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

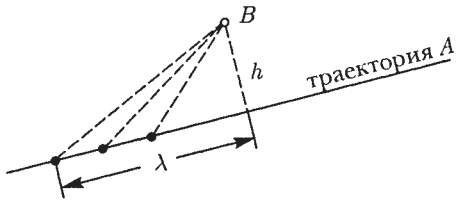

Рис. 9

где неизвестными параметрами могут быть в лучшем случае h, λ, в худшем — h, λ, v и $\sigma^2 = \mathbf{D}\xi_i$ (можно считать, например, что $\xi_i \in \mathbf{\Phi}_{0,\sigma^2}$, λ определяет расстояние, на котором начинается зондирование).

Пример 3. Усеченные и зашумленные наблюдения. Пусть $Y = (\mathrm{y}_1, \ldots, \mathrm{y}_n)$ есть выборка из распределения $\mathbf{P}_\theta$ с плотностью $f_\theta(y)$ относительно некоторой меры μ. Наблюдателю доступны лишь значения $\mathrm{x}_i = \min(\mathrm{y}_i, s_i)$, где $s_1, \ldots, s_n$ — заданная последовательность чисел (возможно, случайных). Требуется, как и в исходной задаче, оценить параметр θ.

Отображение $x = \min(y, s)$ не имеет обратного, и восстановить по x_i значения y_i невозможно. Чтобы охарактеризовать распределение x_i, введем в рассмотрение меры μ_i', сосредоточенные в точках s_i:

$$\mu_i'(B) = \begin{cases} 1, & \text{если} \quad s_i \in B, \\ 0, & \text{если} \quad s_i \notin B, \end{cases}$$

и положим $\mu_i = \mu + \mu_i'$. Тогда очевидно, что распределение x_i имеет следующую плотность $f_{\theta,i}(x)$ относительно μ_i:

$$f_{\theta,i}(x) = \begin{cases} f_\theta(x) & \text{при} \quad x < s_i, \\ (1 + \mu(s_i))^{-1}\mathbf{Q}(\theta, s_i) & \text{при} \quad x = s_i, \\ 0 & \text{при} \quad x > s_i, \end{cases}$$

где $\mathbf{Q}(\theta, s) = \mathbf{P}_\theta(y_1 \geqslant s) = \int\limits_s^\infty f_\theta(x)\mu(dx)$.

Возможны ситуации, когда при больших n значения s_i таковы, что $\mathbf{P}_\theta(\mathrm{y}_i < s_i)$ малы. В этом случае следует рассматривать схему серий, когда для каждого n имеем свой набор значений $s_i = s_{i,n}$ такой, что $\mathbf{P}_\theta(\mathrm{y}_i < s_{i,n}) \to$ $\to 0$ при $n \to \infty$.

Встречаются также ситуации, когда наблюдателю доступны лишь значения $\mathrm{x}_i = \mathrm{y}_i + s_i$, где s_i — случайные помехи (при тех же, что и прежде, y_i). Если s_i имеет распределение G_i, то плотность x_i относительно μ будет равна

$$f_{\theta,i}(x) = \int dG_i(u) f_\theta(x - u).$$

Пример 4. Статистика случайных процессов. Пусть $\xi(t)$ — случайный процесс, заданный на $[0, T]$, распределение которого зависит от неизвестного параметра θ. Будем различать три вида статистических данных.

(a) Рассматриваются n независимых реализаций $\xi_1(t), \ldots, \xi_n(t)$ этого процесса, относительно которых нам известны значения $\mathrm{x}_i = \xi_i(u_i)$ в известные моменты времени $u_0 = 0, u_1, \ldots, u_n$.

(b) Мы рассматриваем одну реализацию $\xi(t)$ этого процесса, и нам известны значения $\mathrm{x}_i = \xi(u_i)$ этой реализации в моменты $u_0 = 0, u_1, \ldots, u_n$.

(c) Нам полностью известна одна реализация процесса ξ на $[0, T]$.

Конечно, возможны и другие виды статистических данных. Для процессов с независимыми приращениями при $\xi(0) = 0$ виды данных п. (a), (b) в известном смысле совпадают, так как реализации $\xi(u_i + t) - \xi(u_i)$, $t \in (0, u_{i+1} - u_i)$, $i \geqslant 0$, независимы при разных i.

Если обозначить через $\mathbf{P}_{\theta,t}$ распределение $\xi(t)$, то в условиях п. (a) мы будем иметь выборку $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, состоящую из независимых наблюдений $\mathrm{x}_i \in \mathbf{P}_{\theta,u_i}$. В условиях п. (b) мы также будем иметь $\mathrm{x}_i \in \mathbf{P}_{\theta,u_i}$, но значения x_i будут, вообще говоря, *зависимы*. Для процесса с независимыми приращениями из выборки X можно получить выборку из независимых наблюдений $\mathrm{y}_i = \mathrm{x}_i - \mathrm{x}_{i-1} \in \mathbf{P}_{\theta,u_i - u_{i-1}}$, $i \geqslant 1$.

В условиях п. (c) нам доступны данные, предусмотренные п. (b) *при любых* n и $u_1, \ldots, u_n$. Однако к этому обстоятельству, предъявляющему повышенные требования к точности измерений $\xi(u_i)$ при малых $u_i - u_{i-1}$, надо подходить с осторожностью (об этом см. ниже) в связи с тем, что точность наблюдений в экспериментах всегда ограничена и, стало быть, абсолютно точные значения $\xi(t)$ в реальных задачах знать невозможно. Одна из возможностей построения математических моделей, более точно описывающих эксперимент, состоит в предположении, что нам даны не «точные» значения $\xi(t)$, а значения $\xi(t) + \Delta(t)$, где $\Delta(t)$ — процесс «ошибок» с независимыми значениями, $\Delta(t) \in U_{-\Delta,\Delta}$ или $\Delta(t) \in \mathbf{\Phi}_{0,\Delta}$, где Δ определяется условиями эксперимента.

Если $\xi(t)$ — диффузионный процесс с коэффициентами сноса и диффузии (α, σ^2) (см. [17, § 18.4]), то в условиях п. (a), (b) мы будем иметь выборку $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, $\mathrm{x}_i \in \mathbf{\Phi}_{\alpha u_i, \sigma^2 u_i}$, рассмотренную в примере 1.

Если $\xi(t)$ — пуассоновский процесс со сносом α и параметром λ и если u_1 мы можем выбирать произвольно малым, то с высокой вероятностью значение α определяется по одному наблюдению: при малых u_1 с вероятностью $e^{-\lambda u_1}$ имеем $\mathrm{x}_i = \alpha u_1$, $\alpha = \mathrm{x}_1/u_1$. Если известно, что αu_i не целые, то можно использовать также арифметические соображения. В силу этого будем считать, что α известно. Тогда в условиях п. (a), (b) мы будем иметь выборку со значениями $\mathrm{x}_i = \xi_i(u_i) - \alpha u_i \in \mathbf{\Pi}_{\lambda u_i}$ (см. пример 1).

Если $\xi(t)$ — диффузионный процесс с неизвестными *переменными* коэффициентами сноса и диффузии $\alpha(x)$, $\sigma^2(x)$ (см. [17, § 18.4]) и нам надо оценить по траекториям процесса эти функции, то мы придем к задаче об оценке «бесконечномерных» параметров $(\alpha(x), \sigma^2(x))$. При оценке таких параметров приходится сталкиваться с качественно новыми явлениями, которые при оценке конечномерных параметров, по сути, не возникают, но о существовании которых можно догадываться. Например, мы встречались с фактом, что если число оцениваемых параметров велико (сравнимо с объемом выборки или превосходит его), то нельзя надеяться на существование «хороших» оценок. То же происходит с бесконечномерным параметром: когда параметрическое множество очень богато, то все известные нам регулярные методы оценивания, в том числе метод максимального правдоподобия, как правило, утрачивают свою эффективность. Это можно проиллюстрировать на примере оценки функции плотности $f(x)$ (ее можно отождествить с неизвестным бесконечномерным параметром) по выборке X из распределения с этой плотностью (см. § 10). Функция правдоподобия в этом случае равна $f(X) = \prod_{i=1}^{n} f(\mathrm{x}_i)$. Если не накладывать никаких ограничений на множество «допустимых» значений параметра $f \geqslant 0$, $\int f(x)\mu(dx) = 1$, то максимум $f(X)$ будет равен ∞ и будет достигаться на всех f, для которых $f(\mathrm{x}_i) = \infty$ хотя бы при одном x_i, $f(\mathrm{x}_j) \neq 0$ при остальных x_j, $j \neq i$. Чтобы получать здесь более содержательные результаты, надо вводить ограничения на множество допустимых f (например, существование вторых производных, см. § 10). Случай, когда f принадлежит параметрическому множеству, нами во многом изучен в гл. 2.

При переходе к изучению статистики случайных процессов появляется и другое важное обстоятельство. При рассмотрении в качестве наблюдений не конечной выборки, а всей траектории процесса на $[0, T]$, $T < \infty$ (см. п. (c)), может оказаться, что относительно некоторых параметров (скалярных или конечномерных) эта траектория несет в себе «бесконечное количество информации», что позволяет определить этот параметр точно (детерминированно). Это будет проиллюстрировано ниже на примере оценки параметра диффузии $\sigma^2 = \text{const}$ (но не сдвига!) по конечному отрезку траектории диффузионного процесса.

В целом задачи об *оценке параметров* случайных процессов часто существенно отличаются от аналогичных задач, изложенных в гл. 2. В то же время подходы к задаче *проверки гипотез* (те, что не связаны с размерностями; см. также элементы общей теории статистических решений в гл. 6) во многом остаются прежними, а некоторые результаты, не связанные с параметрическими гипотезами, полностью сохраняют свою силу. Например, в задаче о проверке двух простых гипотез оптимальный критерий (байесовский, минимаксный, н. м. к) по-прежнему будет основан на статистике отношения правдоподобия — ведь наблюдаемую траекторию процесса можно рассматривать как выборку объема 1 в соответствующем пространстве $\mathcal{X}$. Наиболее трудным элементом этой задачи становится отыскание в явном виде плотностей распределения (если они существуют) этих наблюдений относительно какой-нибудь меры.

Разработка общей теории статистических выводов для случайных процессов является делом технически весьма сложным и выходит за рамки настоящей книги. Представление о некоторых основных результатах этой теории можно получить, например, по книгам [6, 8, 61, 70]. Отдельные простые примеры, упомянутые выше и относящиеся к случайным процессам (см. также пример 6), будут рассмотрены позже в качестве иллюстраций.

До сих пор в примерах шла речь в основном об оценке параметров. Рассмотрим теперь примеры задач о проверке гипотез по разнораспределенным наблюдениям. К ним можно отнести, конечно, все задачи о проверке параметрических гипотез в примерах 1–3. Приведем здесь также пример иного рода.

Пример 5. Проверка двух простых гипотез. Пусть нам известно, что возможна лишь одна из двух гипотез H_1 или H_2 относительно природы наблюдений $\mathrm{x}_1, \ldots, \mathrm{x}_n$. В соответствии с гипотезой H_1 наблюдение x_i имеет распределение с плотностью $f_i^{(1)}(x)$ относительно некоторой меры μ_i; по гипотезе H_2 распределения x_i имеют плотности $f_i^{(2)}(x)$ относительно той же меры μ_i. Задача состоит в построении критериев (желательно оптимальных) для проверки гипотез H_1, H_2 по выборке X и в вычислении вероятностей ошибок этих критериев.

Как было установлено в § 42, оптимальные критерии (в том числе наиболее мощный критерий) для проверки H_i, $i = 1, 2$, имеют критическую область вида

$$Z(X) = \prod_{i=1}^{n} \frac{f_i^{(1)}(\mathrm{x}_i)}{f_i^{(2)}(\mathrm{x}_i)} > c, \tag{4}$$

построенную с помощью отношения плотностей (относительно меры $\mu_1 \times \ldots \times \mu_n$) наблюдения X при гипотезах H_i. Как и в § 43, мы можем здесь рассматривать два асимптотических подхода к исследованию критерия (4):

a) считать гипотезы H_i фиксированными, т. е. плотности $f_i^{(j)}$ фиксированы;

b) считать гипотезы H_i близкими, т. е. $f_i^{(j)} = f_{i,n}^{(j)}$ зависят от n и сближаются при $n \to \infty$. Если ограничиться рассмотрением параметрических гипотез, т. е. считать, что $f_i^{(j)}(x) = f_i(x, \theta_j)$, $\theta_1 = \theta_2$, то предположение о близости будет означать предположение о близости θ_1 и θ_2: $\theta_2 = \theta_{2,n} = \theta_1 + \varepsilon(n)$, $\varepsilon(n) \to 0$ при $n \to \infty$. В этих случаях мы будем иметь дело со *схемой серий* для разнораспределенных наблюдений. Более подробно об этом см. § 71.

Приведем теперь пример, относящийся к статистике случайных процессов, в котором естественным образом применимы оба подхода: с позиций теории оценивания и с позиций теории проверки гипотез. Речь идет о задаче о разладке, упомянутой в примере 2 введения.

Пример 6. Задача о разладке. Пусть дана совокупность наблюдений $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ объема n, $n \leqslant \infty$, устроенная следующим образом. Первые $\theta - 1$ наблюдений, $1 \leqslant \theta \leqslant n$, независимы и имеют плотность распределения $g_1(x)$ относительно некоторой меры μ. Последующие $n - \theta$ наблюдений также независимы, но имеют плотность распределения $g_2(x) \neq g_1(x)$. Требуется найти или оценить момент θ, называемый *моментом разладки*. Это есть задача *оценки дискретного параметра θ случайного процесса X*. Одновременно ее можно рассматривать как задачу о *проверке n простых гипотез $H_j = \{\theta = j\}$* по единственному наблюдению X.

Вернемся к задачам, связанным с оценкой параметра скалярных наблюдений. Следующий пример иллюстрирует редукцию задачи о разнораспределенных наблюдениях при наличии мешающего «бесконечномерного» параметра к однородным наблюдениям в схеме серий.

Пример 7. Оценка «хвостов» распределений. Эта задача представляет интерес в теории страхования (см., например, [59, 87]).

Пусть $\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_n$ — наблюдения над случайными величинами с плотностями (относительно меры Лебега)

$$f_{\alpha,i}(y) = \gamma_i(\alpha, y) y^{-\alpha-1}, \quad y > 0,$$

где $\alpha > 0$. Относительно функций $\gamma_i(\alpha, y)$ известно лишь, что $\gamma_i(\alpha, y) \to \gamma$ при $y \to \infty$ при всех i. Требуется оценить параметр α. Предел γ может быть любым. Поэтому можно рассматривать в известном смысле эквивалентную задачу об оценке параметра $\theta = (\alpha, \gamma)$ по выборке из распределения с плотностью

$$f_{\theta,i}(y) = f_{(\alpha,\gamma),i}(y) = \gamma(1 + \varepsilon_i(\alpha, y)) y^{-\alpha-1}, \quad \varepsilon_i(\alpha, y) \to 0, \tag{5}$$

при $y \to \infty$.

Сделаем сначала упрощающее предположение, что при достаточно большом N и всех $y \geqslant N$ слагаемые $\varepsilon_i(\alpha, y)$ пренебрежимо малы (т. е. можно считать, что $\varepsilon_i(\alpha, y) = 0$ при $y > N$). Введем в рассмотрение наблюдения

$$\mathrm{x}_i = \max(N, \mathrm{y}_i) \tag{6}$$

и положим $\mu_N = \mu + \mu'_N$, где μ — мера Лебега, μ'_N — мера, сосредоточенная в точке N (ср. с примером 3). Тогда x_i будут иметь распределения с плотностью $f_\theta(x)$ относительно μ_N:

$$f_\theta(x) = \begin{cases} \gamma x^{-\alpha-1} & \text{при} \quad x > N, \\ 1-p & \text{при} \quad x = N, \\ 0 & \text{при} \quad x < N, \end{cases} \tag{7}$$

где

$$p = p(\theta, N) = \int_N^\infty \gamma^{-\alpha-1}\, dx = \frac{\gamma N^{-\alpha}}{\alpha}.$$

Мы получили вполне «регулярную» задачу для однородных наблюдений x_i, которая согласно § 36 при $n \to \infty$ допускает а. н. оценки (α^*, γ^*) (например, о. м. п. $\widehat{\alpha}^*$, $\widehat{\gamma}^*$), имеющие точность при фиксированном N порядка $n^{-1/2}$. Это значит, что при больших n основная погрешность при оценке α может быть связана не с разбросом оценок α^*, γ^*, а со сделанным выше упрощающим предположении о замене $\varepsilon_i(\alpha, y)$ при $y > N$ на нуль. Чтобы уменьшить эту вторую погрешность, естественно рассматривать значения $N = N(n)$, растущие вместе с n. Мы вновь получим задачу в схеме серий. С ростом N доля $p(\theta, N)$ «нетривиальных» наблюдений x_i (отличных от N) будет уменьшаться. Остается выбрать скорость роста $N(n)$ так, чтобы слагаемыми $\varepsilon_i(\alpha, y)$ при $y > N$ действительно можно было пренебречь, что возможно, если известны оценки для скорости убывания $\varepsilon_i(\alpha, y) < \varepsilon(y)$. Подробнее об этом см. в § 68.

Аналогичным образом может быть рассмотрена задача об оценке «экспоненциальных хвостов», в которой предполагается, что плотность наблюдений y_i имеет вид

$$f_{\alpha,i}(y) = \gamma_i(\alpha, y) e^{-\alpha y}, \quad y > 0,$$

где функции γ_i обладают теми же свойствами, что и прежде. Если вместо наблюдений y_i рассмотреть наблюдения $\mathrm{z}_i = e^{\mathrm{y}_i}$, то плотность этих новых наблюдений будет иметь вид

$$\widetilde{f}_{\alpha,i}(z) = f_{\alpha,i}(\ln z) z^{-1} = \gamma_i(\alpha, \ln z) z^{-\alpha-1}, \quad z \geqslant 1,$$

так что задача сводится к предыдущей.

В последующем изложении в качестве основного объекта в этом примере будем рассматривать задачи об оценке параметра $\theta = (\alpha, \gamma)$ по выборке $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ с плотностями (5) (задача A) и по выборке X с плотностью наблюдений (7) в схеме серий при $N \to \infty$ (задача B).

Приведенные примеры будут более подробно рассмотрены в последующих параграфах либо в качестве иллюстраций, либо в качестве объектов изучения, представляющих самостоятельный интерес. Эти примеры позволяют заметить следующее.

1. В каждом из них мы могли выборку $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ рассматривать как одно n-мерное наблюдение и применять к нему качественные результаты глав 1–4, 6. Это позволяет нам перенести на случай разнораспределенных

наблюдений результаты, связанные с принципами сравнения оценок и критериев, с понятиями эффективных оценок и оценок максимального правдоподобия, оптимальных критериев, достаточных статистик, с неравенствами Рао–Крамера и др.

2. При этом, как правило, оказывается невозможным использовать основные асимптотические результаты глав 1–3, поскольку для разнораспределенных наблюдений отсутствует сам исходный объект этих глав — выборка X *из одного и того же распределения* $\mathbf{P}$. Такие фундаментальные факты, как теорема Гливенко–Кантелли и другие теоремы об асимптотическом поведении эмпирических распределений и статистик, остаются в стороне от рассматриваемых задач, а вместе с ними и такой фундаментальный принцип построения оценок, как метод подстановки.

3. В связи с этим желательно выработать более общие подходы, позволяющие сохранить аналоги тех асимптотических результатов глав 2, 3, которые допускают обобщение на случай разнораспределенных наблюдений.

В последующих разделах мы реализуем в какой-то мере программу, намеченную в п. 1, 3. При этом мы будем придерживаться той же последовательности изложения, которая была принята в гл. 2, 3.

§ 65. Основные методы построения оценок. M-оценки. Состоятельность и асимптотическая нормальность

1. Предварительные замечания и определения. Как уже отмечалось, в условиях этой главы отсутствует, вообще говоря, какое-то одно распределение, характеризующее выборку, и задача оценки этого распределения или функционалов от этого распределения не возникает. Понятие оценки здесь будет относиться только к параметрическому случаю, когда известно, что $x_i \in \mathbf{P}_{\theta,i}$, $i = 1, \ldots, n$, где $\mathbf{P}_{\theta,1}, \mathbf{P}_{\theta,2}, \ldots$ — данная последовательность распределений (разного типа, вообще говоря), зависящих от параметра θ. Из всех подходов к построению оценок, изложенных в гл. 2, естественным образом переносятся на случай разнораспределенных наблюдений лишь метод M-оценивания (см. § 14) и, в частности, метод моментов и метод максимального правдоподобия. Оценки подстановки как собирательный класс оценок, выражающий некоторый общий подход к их получению, здесь во многом теряют свой смысл.

Как и в гл. 2, нам понадобятся понятия состоятельности и асимптотической нормальности оценок. В дальнейшем мы будем писать $X \in \mathbf{P}_\theta$, если $x_i \in \mathbf{P}_{\theta,i}$, понимая под $\mathbf{P}_\theta$ набор распределений $\{\mathbf{P}_{\theta,i}\}$. При этом часто мы будем иметь в виду *схему серий*, т. е. будем считать, что распределения $\mathbf{P}_{\theta,i}$, $i = 1, 2, \ldots$, вообще говоря, зависят и от n.

Свойство *состоятельности* оценки θ^* параметра θ определяется, как и в § 13 (см. определение 13.1), соотношением

$$\mathbf{P}(|\theta^* - \theta| > \varepsilon) \to 0$$

при $n \to \infty$ для $X \in \mathbf{P}_\theta$ и для любого $\varepsilon > 0$, которое можно записать также в виде

$$\theta^* \xrightarrow[P]{} \theta$$

при $X \in \mathbf{P}_\theta$, $n \to \infty$. (Сходимость θ^* по вероятности к постоянной θ не требует задания θ^* при разных n на одном вероятностном пространстве.)

Если мы не имеем дело со схемой серий и выборка X образована первыми n элементами данной бесконечной последовательности X_∞, $X_i \in \mathbf{P}_{\theta,i}$, то можно рассматривать также *сильно состоятельные* оценки

$$\theta^* \underset{\text{п.н.}}{\longrightarrow} \theta \quad \text{при} \quad n \to \infty.$$

Понятие а. н. оценок здесь несколько иное, чем в гл. 2.

Определение 1. Оценка $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ k-мерного параметра $\theta = (\theta_1, \ldots, \theta_k)$ называется *асимптотически нормальной с разбросом* $\sigma^2(\theta, n)$, где $\sigma^2(\theta, n)$ — положительно определенная матрица, если при $X \in \mathbf{P}_\theta$, $n \to \infty$,

$$(\theta^* - \theta)\sigma^{-1}(\theta, n) \Rightarrow \mathbf{\Phi}_{0,E},$$

где E — единичная матрица. Часто $\sigma^2(\theta, n)$ удается представить в виде $\sigma^2(\theta)b^{-1}(n)$, где $\sigma^2(\theta)$ — матрица вторых моментов. При этом здесь $b(n)$ в отличие от определений § 13, вообще говоря, не равны $\sqrt{n}$. Разброс $\sigma^2(\theta, n)$ часто совпадает с матрицей вторых моментов $(\theta^* - \theta)$.

Для пуассоновских наблюдений $\mathrm{x}_i \in \mathbf{\Pi}_{\lambda t_i}$ оценка

$$\lambda^* = b^{-1}(n) \sum \mathrm{x}_i, \quad b(n) = \sum_{i=1}^n t_i$$

является а. н. с разбросом $\lambda b^{-1}(n)$, так как $\lambda^* b(n) = \sum_{i=1}^n \mathrm{x}_i \in \mathbf{\Pi}_{\lambda b(n)}$ и, стало быть,

$$\frac{(\lambda^* - \lambda)b(n)}{\sqrt{\lambda b(n)}} = (\lambda^* - \lambda)b(n)^{1/2}\lambda^{-1/2} \Rightarrow \mathbf{\Phi}_{0,1}.$$

2. M-оценки. Как уже отмечалось, в случае разнораспределенных наблюдений основной общий метод получения оценок связан с M-оценками. Базой для построения M-оценок являются функции $g_i(x, \theta)$, обладающие свойством

$$\int g_i(x, \theta)\mathbf{P}_{\theta,i}(dx) = 0. \tag{1}$$

Везде ниже в этом параграфе будет предполагаться, что функции $g_i(x, \theta)$ этим свойством обладают. Получить функцию $g_i(x, \theta)$ с таким свойством можно из любой интегрируемой функции $\widetilde{g}_i(x, \theta)$, положив

$$g_i(x, \theta) = \widetilde{g}_i(x, \theta) - \int \widetilde{g}_i(x, \theta)\mathbf{P}_{\theta,i}(dx).$$

Фиксируем теперь некоторое $\theta_0 \in \Theta$ и положим

$$s_i(\theta) = \int g_i(x, \theta)\mathbf{P}_{\theta_0,i}(dx). \tag{2}$$

Тогда θ_0 можно рассматривать как решение уравнения $s_i(\theta) = 0$. Эмпирическим аналогом правой части в (2) в случае $X \in \mathbf{P}_{\theta_0}$ является значение $g_i(\mathrm{x}_i, \theta)$. Эти факты и лежат в основе следующего определения.

Определение 2. Любое решение уравнения

$$\sum_{i=1}^{n} g_i(\mathrm{x}_i, \theta) = 0 \tag{3}$$

называется *M-оценкой* параметра θ, соответствующего выборке X.

Если θ — k-мерный параметр, то в качестве g_i следует брать k-мерную вектор-функцию с линейно независимыми координатами, так что (3) будет представлять собой систему из k уравнений.

Если

$$g_i(x, \theta) = g_i(x) - m_i(\theta), \quad m_i(\theta) = \int g_i(x)\mathbf{P}_{\theta,i}(dx),$$

то M-оценки по аналогии со случаем однородных наблюдений можно называть *оценками по методу моментов*, хотя прозрачность этого подхода здесь значительно меньше, чем в однородном случае.

Если считать, что θ — скалярный параметр, $m(\theta) = \sum_{i=1}^{n} m_i(\theta)$ — монотонная функция, то оценка по методу моментов единственна и определяется равенством

$$\theta^* = m^{-1}\left(\sum g_i(\mathrm{x}_i)\right)$$

(здесь и ниже, если не указаны пределы суммирования, имеется в виду суммирование от единицы до n). Подобрать функции $g_i(x)$ так, чтобы $m_i(\theta)$ (а следовательно, и $m(\theta)$) были монотонными, возможно, если исходить из «физического» смысла параметра θ. Например, если θ — параметр сдвига, то $g_i(x)$ следует брать линейными, если θ — параметр масштаба, то квадратичными.

Пример 1. Пусть $\mathbf{P}_i$ — известные распределения, $\theta = (\alpha, \sigma^2)$, $\mathrm{x}_i \in \mathbf{P}_{\theta,i}$, $\mathbf{P}_{\theta,i}(B) = \mathbf{P}_i\left(\dfrac{B - \alpha t_i}{\sigma\sqrt{u_i}}\right)$. Предполагается, что $\mathbf{P}_i$ имеют конечные моменты второго порядка,

$$a_i = \int x\mathbf{P}_i(dx) = 0, \quad b_i = \int x^2\mathbf{P}_i(dx) = 1.$$

(Случай произвольных a_i, b_i также может быть рассмотрен с той лишь разницей, что уравнения (3) для α и σ^2, рассматриваемые ниже, в этом случае будут немного сложнее.) В примере 64.1 в качестве $\mathbf{P}_i$ фигурировали нормальные распределения $\mathbf{\Phi}_{0,1}$.

Положим $g_i(x) = (g_{i,1}(x), g_{i,2}(x))$, $i = 1, 2$; $g_{i,1}(x) = v_i x$, $g_{i,2}(x) = w_i x^2$, где v_i, w_i выберем позже. Тогда

$$m_{i,1}(\theta) = \mathbf{E}_\theta g_{i,1}(\mathrm{x}_i) = \alpha t_i v_i,$$

$$m_{i,2}(\theta) = \mathbf{E}_\theta g_{i,2}(\mathrm{x}) = w_i\left[\mathbf{E}(\mathrm{x}_i - \alpha t_i)^2 + \alpha^2 t_i^2\right] = w_i\left(\sigma^2 u_i + \alpha^2 t_i^2\right).$$

Оценки по методу моментов будут определяться уравнениями (см. (3))

$$\sum v_i(\mathrm{x}_i - \alpha t_i) = 0, \quad \sum w_i\left(\mathrm{x}_i^2 - u_i\sigma^2 - \alpha^2 t_i^2\right) = 0,$$

которые дают

$$\alpha^* = \frac{\sum \mathrm{x}_i v_i}{\sum t_i v_i}, \quad (\sigma^2)^* = \frac{\sum w_i\left(\mathrm{x}_i^2 - (\alpha^*)t_i^2\right)}{\sum u_i w_i}. \tag{4}$$

Рассмотрим сначала оценку α^*. Имеем

$$\alpha^* - \alpha = \frac{\sum v_i(\mathrm{x}_i - \alpha t_i)}{\sum t_i v_i}, \quad \mathbf{E}(\alpha^* - \alpha)^2 = \frac{\sigma^2 \sum v_i^2 u_i}{\left(\sum t_i v_i\right)^2}.$$

Стало быть, если

$$b_1(n) = \frac{\left(\sum t_i v_i\right)^2}{\sum v_i^2 u_i} \to \infty, \tag{5}$$

то α^* — состоятельная оценка. В силу центральной предельной теоремы (см. приложение IV) α^* будет а. н. с разбросом $\sigma^2(\alpha, n) = \sigma^2 b_1^{-1}(n)$, если суммы $\sum v_i(\mathrm{x}_i - \alpha t_i)$ удовлетворяют условию Линдеберга или условию Ляпунова:

$$\left(\sum v_i^2 u_i\right)^{-3/2} \sum \mathbf{E}\left|\frac{\mathrm{x}_i - \alpha t_i}{\sqrt{u_i}}\right|^3 \to 0 \quad \text{при} \quad n \to \infty.$$

Отметим, что выбор v_i в (5) может влиять не только на постоянный множитель, но и на саму скорость роста $b_i(n)$. Естественно выбрать v_i так, чтобы максимизировать $b_1(n)$. Так как $\left(\sum a_i b_i\right)^2 \leqslant \sum a_i^2 \sum b_i^2$, то

$$\left(\sum t_i v_i\right)^2 \leqslant \left(\sum \frac{t_i^2}{u_i}\right)\left(\sum v_i^2 u_i\right), \quad b_1(n) \leqslant \sum \frac{t_i^2}{u_i},$$

где правая часть от v_i не зависит. Эта верхняя граница для $b_1(n)$, очевидно, достигается при $v_i = t_i/u_i$. Для таких значений v_i получаем «оптимальную» в классе (4) оценку

$$\alpha^* = \frac{\sum \dfrac{\mathrm{x}_i t_i}{u_i}}{\sum \dfrac{t_i^2}{u_i}},$$

для которой

$$b_1(n) = \sum \frac{t_i^2}{u_i}.$$

Как уже отмечалось, оценка α^* будет состоятельной, если $b_1(n) \to \infty$. Это же условие необходимо для асимптотической нормальности α^*. В примере 64.4

об оценке параметров диффузионного процесса $t_i = u_i$ и условия $b_1(n) \to \infty$ эквивалентно $\sum t_i \to \infty$.

Аналогичным образом может быть рассмотрена оценка (4) для σ^2. Здесь оптимальными будут значения $w_i = u_i^{-1}$. Читатель без труда может убедиться в этом аналогично предыдущему. Мы опускаем здесь эти рассмотрения еще и потому, что оптимальность выбора $w_i = u_i^{-1}$ будет независимым образом подтверждена ниже (как, впрочем, и выбора $v_i = t_i/u_i$). В результате будем иметь дело с оценкой

$$(\sigma^2)^* = \frac{1}{n}\sum \frac{\mathrm{x}_i^2 - (\alpha^*)t_i^2}{u_i} = \frac{1}{n}\sum \frac{(\mathrm{x}_i - \alpha^* t_i)^2}{u_i} = \\ = \frac{1}{n}\left[\sum \frac{(\mathrm{x}_i - \alpha^* t_i)^2}{u_i} - (\alpha^* - \alpha)^2 b_1(n)\right], \quad (6)$$

так что

$$((\sigma^2)^* - \sigma^2)\sqrt{n} = \frac{1}{\sqrt{n}}\sum \left[\frac{(\mathrm{x}_i - \alpha t_i)^2}{u_i} - \sigma^2\right] - \frac{1}{\sqrt{n}}(\alpha^* - \alpha)^2 b_1(n).$$

Так как $\mathbf{E}(\alpha^* - \alpha)^2 b_1(n) = \sigma^2$, то последнее слагаемое в правой части сходится по вероятности к нулю (если α известно, то это слагаемое будет отсутствовать). Первое слагаемое можно представить в виде $\dfrac{\sigma^2}{\sqrt{n}}\sum (\xi_i^2 - 1)$, $\xi_i = \dfrac{\mathrm{x}_i - \alpha t_i}{\sigma\sqrt{u_i}}$, где $\mathbf{E}\xi_i = 0$, $\mathbf{D}\xi_i = 1$. Оно будет асимптотически нормально с параметрами $(0, \sigma^4 d^4)$, $d^4 = \dfrac{1}{n}\sum \mathbf{E}\left(\xi_i^2 - 1\right)^2 = \dfrac{1}{n}\sum \mathbf{E}\xi_i^4 - 1$ в предположении, что выполнено условие Линдеберга или условие Ляпунова:

$$n^{-1-\gamma/2}\sum \mathbf{E}|\xi_i|^{4+2\gamma} \to 0, \quad \gamma > 0. \quad (7)$$

Итак, при выполнении (7) и $d^4 = \mathrm{const}$ оценка $(\sigma^2)^*$ будет а. н. с разбросом $\sigma^4 d^4 b_2^{-1}(n)$, $b_2(n) = n$, который, очевидно, не зависит от последовательностей $\{t_i, u_i\}$. Кроме того, нетрудно видеть, что, как и в однородном случае, α^* и $(\sigma^2)^*$ будут асимптотически независимыми (некоррелированными), т. е. оценка $\theta^* = (\alpha^*, (\sigma^2)^*)$ для $\theta = (\alpha, \sigma^2)$ будет а. н. с разбросом $\sigma^2(\theta, n) = \sigma^2(\theta) b^{-1}(n)$, где

$$\sigma^2(\theta) = \|\sigma_{ij}(\theta)\|, \quad b(n) = \|b_{ij}(n)\|, \quad \sigma_{ij}(\theta) = b_{ij}(n) = 0 \quad \text{при} \quad i \neq j,$$

$$\sigma_{11}(\theta) = \sigma^2, \quad \sigma_{22}(\theta) = \sigma^4 d^4, \quad b_{11}(n) = b_1(n), \quad b_{22}(n) = b_2(n) = n.$$

Отметим, что здесь соотношение между $b_1(n) = \sum \dfrac{t_i^2}{u_i}$ и $b_2(n) = n$ может быть любым.

Отметим также, что из асимптотической нормальности $(\sigma^2)^*$ с разбросом $\sigma^4 d^4 n^{-1}$ следует состоятельность

$$(\sigma^2)^* \xrightarrow[P]{} \sigma^2$$

при $n \to \infty$, хотя условия состоятельности могут быть, конечно, и более слабыми (см. п. 4). Далее, названные выше условия (см. (7)) выполнены в примерах 64.1 и 64.4 об оценках параметров диффузионного процесса. В последнем примере в условиях п. (c) мы располагаем реализацией процесса $\xi(t)$ на $[0, T]$ и, взяв разности наблюдений в точках $u_i = iT/n$, мы можем по ней построить выборку $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, $\mathrm{x}_i = \xi(i\Delta) - \xi((i-1)\Delta) \in \mathbf{\Phi}_{(\alpha\Delta, \sigma^2\Delta)}$ любого объема n, где $\Delta = T/n$ (здесь $t_i = u_i = \Delta$). В соответствии с предыдущим мы будем иметь здесь оценки

$$\alpha^* = \frac{\sum \mathrm{x}_i}{T} = \frac{\xi(T)}{T},$$

$$(\sigma^*)^2 = \frac{1}{T}\sum \left(\mathrm{x}_i^2 - (\alpha^*)^2\Delta^2\right) = \frac{1}{T}\sum \mathrm{x}_i^2 - \frac{(\sigma^*)^2 T}{n}.$$

Здесь $(\alpha^* - \alpha) \in \mathbf{\Phi}_{0, \sigma^2/T}$ и оценка α^* при фиксированном T и растущем n несостоятельна; при этом лучших оценок не существует (см. § 67 или § 25). Для оценки $(\sigma^*)^2$ имеем

$$(\sigma^*)^2 = \frac{1}{T}\sum (\mathrm{x}_i - \alpha\Delta)^2 - (\alpha - \alpha^*)^2\Delta,$$

где $(\alpha - \alpha^*)^2\Delta \xrightarrow[P]{} 0$ при $n \to \infty$, $T \leqslant \mathrm{const}$. Так как по закону больших чисел

$$\frac{1}{T}\sum (\mathrm{x}_i - \alpha\Delta)^2 = \frac{1}{n}\sum \left(\frac{\mathrm{x}_i - \alpha\Delta}{\sqrt{\Delta}}\right)^2 \xrightarrow[P]{} \sigma^2,$$

то $(\sigma^*)^2 \xrightarrow[P]{} \sigma^2$. Это означает, что, пользуясь траекторией $\xi(T)$ лишь на малом отрезке $[0, T]$, мы можем выбором n восстановить любой заданной точностью параметр σ^2. Как уже отмечалось в примере 64.4, процедура построения оценок $(\sigma^*)^2$ при этом требует знания траектории $\xi(T)$ с неограниченной точностью (в связи с этим к сделанному выводу надо относиться с осторожностью). Чтобы избежать затруднений, связанных с этим обстоятельством, можно предположить, например, что мы имеем в качестве наблюдений в моменты $i\Delta$ не $\xi(i\Delta)$, а значения $\xi(i\Delta) + \eta_i$, где ошибки наблюдений η_i независимы и одинаково распределены, $\mathbf{E}\eta_i = 0$, $\mathbf{E}\eta_i^2 = \varepsilon/2$, $i \geqslant 1$. Число ε будем считать малым. В этом случае мы будем иметь дело с выборкой

$$\mathrm{x}_i = \xi(i\Delta) - \xi((i-1)\Delta) + \eta_i - \eta_{i-1}, \quad \xi(0) = \eta_0 = 0.$$

Метод моментов приводит к оценке

$$\alpha^* = \frac{1}{T}\sum \mathrm{x}_i = \frac{\xi(T) + \eta_n}{T},$$

которая, как и прежде, несостоятельна при фиксированном T, $\mathbf{D}\alpha^* = \dfrac{2\sigma^2 T + \varepsilon^2}{2T^2}$, и оценке

$$(\sigma^*)^2 = \frac{1}{T}\sum \mathrm{x}_i^2 - \frac{\varepsilon^2}{\Delta} - (\alpha^*)^2\Delta = \frac{1}{T}\sum (\mathrm{x}_i - \alpha\Delta)^2 - \frac{\varepsilon^2}{\Delta} - (\alpha - \alpha^*)^2\Delta. \quad (8)$$

Считая для простоты $\alpha = 0$ известным, получим

$$(\sigma^*)^2 - \sigma^2 = \frac{1}{n}\sum \left[\left(\frac{\mathrm{x}_i}{\sqrt{\Delta}}\right)^2 - \frac{\varepsilon^2}{\Delta} - \sigma^2\right].$$

При каких Δ разброс $\mathbf{E}((\sigma^*)^2 - \sigma^2)^2$ будет минимальным? Предположим, что $\mathbf{E}\left(\dfrac{\eta_i}{\varepsilon}\right)^4 < \text{const}$, $T \leqslant \text{const}$. Тогда получим

$$\mathbf{E}((\sigma^*)^2 - \sigma^2)^2 \leqslant \frac{1}{n}\mathbf{E}\left[\left(\frac{\mathrm{x}_i}{\sqrt{\Delta}}\right)^2 - \frac{\varepsilon^2}{\Delta} - \sigma^2\right]^2 \leqslant c\left[\Delta + \frac{\varepsilon^4}{\Delta}\right],$$

$c = \text{const}$. Наименьшее значение правой части достигается при $\Delta = \varepsilon^2$, так как разброс (наименьший) оценки $(\sigma^*)^2$ будет иметь порядок ε^2. Очевидно, это сохранится и в общем случае (8). Точный подсчет наименьшего значения $\mathbf{E}((\sigma^*)^2 - \sigma^2)^2$ предоставляем читателю.

Пример 2. Пусть $\mathrm{x}_i \Subset \mathbf{\Pi}_{\lambda t_i}$. Здесь параметр λ не является ни параметром сдвига, ни параметром масштаба. Тем не менее ясно, что при $g_i(x) = v_i x$ мы получим монотонную линейную функцию $m(\lambda)$ и, стало быть, можем, как и в примере 1, найти оценку по методу моментов в явном виде. Действительно, здесь

$$m_i(\lambda) = \mathbf{E}_\lambda v_i \mathrm{x}_i = \lambda t_i v_i, \quad m(\lambda) = \lambda \sum t_i v_i, \quad \lambda^* = \frac{\sum v_i \mathrm{x}_i}{\sum v_i t_i}.$$

Так как

$$\lambda^* - \lambda = \frac{\sum v_i(\mathrm{x}_i - \lambda t_i)}{\sum v_i t_i},$$

то разброс λ^* равен $\mathbf{E}(\lambda^* - \lambda)^2 = \lambda b^{-1}(n)$, $b(n) = \dfrac{\sum (v_i t_i)^2}{\sum v_i^2 t_i}$. Здесь мы имеем ту же задачу минимизации, что и в примере 1, но при $u_i = t_i$. Поэтому оптимальные коэффициенты равны $v_i \equiv 1$,

$$\lambda^* = \frac{\sum \mathrm{x}_i}{\sum t_i}, \qquad b(n) = \sum t_i.$$

Очевидно, что эта оценка будет а. н. (с разбросом $\lambda b^{-1}(n)$), если $\sum_{i=1}^{n} t_i \to \infty$.

Вернемся к общим определениям. Как и в § 14, здесь можно рассматривать несколько иной подход к определению M-оценок. Пусть $\psi_i(x, \theta)$ обозначают заданные функции на $\mathcal{X} \times \Theta$.

Определение 2А. *$\widehat{M}$-оценкой* параметра θ по выборке $X \in \mathbf{P}_\theta$ называется значение θ, при котором достигается максимум по t (или минимум)

$$\sum_{i=1}^{n} \psi_i(\mathrm{x}_k, t). \tag{9}$$

В качестве $\psi_i(x,t)$ естественно брать такие функции, для которых $\int \psi_i(x,t)\mathbf{P}_{\theta,i}(dx)$ достигает своего максимума (соответственно минимума) при $t=\theta$.

Например, если $\mathbf{P}_{\theta,i}(B) = \mathbf{P}_i(B-\theta u_i)$, где $\mathbf{P}_i$ известны, $\int x\mathbf{P}_i(dx) = a_i$ (см. пример 1), то среднее значение для $\mathbf{P}_{\theta,i}$ равно $a_i - \theta u_i$ и при $t=\theta$ достигается $\min\limits_t \int (x-a_i+tu_i)^2\mathbf{P}_{\theta,i}(dx)$, так что $\widehat{M}$-оценкой θ будет значение, минимизирующее (9) при $\psi_i(x,t) = (x-a_i+tu_i)^2$. Таким образом,

Оценки по методу наименьших квадратов есть $\widehat{M}$-оценки.

В примере 64.2 о нелинейной регрессии

$$\mathrm{x}_i = \varphi_1(\theta) + \xi_i, \quad \xi_i \in \mathbf{P}_i, \quad \mathbf{E}\xi_i = 0.$$

Оценка θ^* по методу наименьших квадратов минимизирует $\sum(\mathrm{x}_i - \varphi_i(\theta))^2$ и будет $\widehat{M}$-оценкой. Другим важным примером $\widehat{M}$-оценок является о. м. н., о которых речь пойдет ниже. Связь $\widehat{M}$-оценок и M-оценок здесь та же, что и в случае однородных наблюдений. Она обсуждалась в § 14.

3*. Состоятельность $\widehat{M}$-оценок. Как и в однородном случае, мы будем предполагать здесь выполненным условие

(A_c) *Множество Θ компактно.*

Комментарии по поводу этого условия см. в § 14. Естественно предполагать также выполнение условия

(A_0) *Функции $\psi_i(x,\theta)$ таковы, что $\int \psi_i(x,t)\mathbf{P}_{\theta,i}(dx)$ достигает максимума в единственной точке $t=\theta$, так что при $t\neq\theta$*

$$\int (\psi_i(x,t) - \psi_i(x,\theta))\mathbf{P}_{\theta,i}(dx) \equiv \mathbf{E}_{\theta,i}\eta_i(t) < 0,$$

где

$$\eta_i(t) = \psi_i(\mathrm{x}_i, t) - \psi_i(\mathrm{x}_i, \theta) \tag{10}$$

($\mathbf{E}_{\theta,i}$ есть среднее относительно $\mathbf{P}_{\theta,i}$).

В дальнейшем мы будем допускать, что рассматривается схема серий, т. е. что функции ψ_i, η_i и распределения $\mathbf{P}_{\theta,i}$ могут зависеть также от n.

Нам понадобятся дополнительные условия. Обозначим

$$\psi_i^\Delta(x,t) = \sup_{|u|\leqslant\Delta} \psi_i(x, t+u), \quad \eta_i^\Delta(t) = \psi_i^\Delta(x,t) - \psi_i(x,\theta).$$

Наряду с семейством $\{\mathbf{P}_{\theta,i}\}$ через $\mathbf{P}_\theta$ будем обозначать также распределение $\mathbf{P}_{\theta,1} \times \ldots \times \mathbf{P}_{\theta,n}$ на $\mathcal{X}^n$ и через $\mathbf{E}_\theta$ — математическое ожидание относительно этого распределения, так что $\mathbf{E}_{\theta,i}\eta_i(t)$ можно записывать и в виде $\mathbf{E}_\theta\eta_i(t)$. Через $(\theta)^\delta$ мы будем обозначать δ-окрестность точки θ.

Введем в рассмотрения условия

(A_1^Δ) *Для любых* $\delta > 0$ *и* $t \notin (\theta)^\delta$, $t \in \Theta$, *существуют* $\Delta = \Delta(\delta)$ *и* $N > 0$ *такие, что при некотором* $\varepsilon = \varepsilon(\delta, t, N) > 0$

$$\limsup_{n\to\infty} \frac{1}{b_n(t)} \sum_{i=1}^n \mathbf{E}_\theta\big(\eta_i^\Delta(t); \eta_i^\Delta(\theta) > -N\big) < -\varepsilon,$$

где

$$b_n(t) = \sum \mathbf{E}_\theta\big(|\eta_i^\Delta(t)|; \eta_i^\Delta(t) > -N\big) \to \infty. \tag{11}$$

(A_2^Δ) *Для любых* $\tau > 0$, $t \in \Theta$ *выполняется*

$$\limsup_{n\to\infty} \frac{1}{b_n(t)} \sum_{i=1}^n \mathbf{E}_\theta\big(\eta_i^\Delta(t); \eta_i^\Delta(\theta) > \tau b_n(t)\big) = 0.$$

Из (A_2^Δ) следует, очевидно, что

$$\sum \mathbf{P}\big(\eta_k^\Delta(t) > \tau b_n(t)\big) \to 0 \tag{12}$$

при $n \to \infty$.

В качестве $b_n = b_n(t)$ можно брать также $b_n = -\sum \mathbf{E}_\theta\eta_i^\Delta(t)$ или $b_n = -\sum \mathbf{E}_\theta\eta_i(t)$, если эти суммы имеют «тот же порядок», что и (11). Это возможно, если они не содержат бесконечно больших отрицательных слагаемых (больших положительных слагаемых они содержать не могут в силу (A_2^Δ)). Об условиях, достаточных для выполнения (A_1^Δ), (A_2^Δ), см. ниже.

Теорема 1. *При выполнении* (A_c), (A_1^Δ), (A_2^Δ) $\widehat{M}$*-оценка состоятельна.*

Доказательство. Обозначим

$$\Psi^*(t) = \sum_{i=1}^n (\psi_i(\mathrm{x}_i, t) - \psi_i(\mathrm{x}_i, \theta)).$$

Если θ^* — $\widehat{M}$-оценка, то

$$\mathbf{P}_\theta(|\theta^* - \theta| > \delta) = \mathbf{P}_\theta\Big(\sup_{|t-\theta|>\delta} \Psi^*(t) > \sup_{|t-\theta|\leqslant\delta} \Psi^*(t)\Big) \leqslant \mathbf{P}_\theta\Big(\sup_{|t-\theta|>\delta} \Psi^*(t) > 0\Big).$$

Пусть δ фиксировано, Δ и N удовлетворяют (A_1^Δ). Покроем множество $\Theta \setminus (\theta)^\delta$ окрестностями $\Delta_j = (t_j)^\Delta$, $j = 1, \ldots, m$, $t_i \in \Theta \setminus (\theta)^\delta$. Тогда

$$\sup_{|t-\theta|>\delta} \Psi^*(t) \leqslant \max_j \sup_{t\in\Delta_j} \Psi^*(t) \leqslant$$

$$\leqslant \max_j \sum_{i=1}^n \big(\psi_i^\Delta(\mathrm{x}_i, t_j) - \psi_i(\mathrm{x}_i, \theta)\big) = \max_j \sum_{i=1}^n \eta_i^\Delta(t_j).$$

Отсюда следует, что

$$\mathbf{P}_\theta(|\theta^* - \theta| > \delta) \leqslant \sum_{j=1}^{m} \mathbf{P}_\theta\left(\sum_{i=1}^{n} \eta_i^{\Delta}(t_j) > 0\right),$$

и нам достаточно доказать, что для $t \notin (\theta)^\delta$

$$\mathbf{P}_\theta\left(\sum_{i=1}^{n} \eta_i^{\Delta}(t) > 0\right) \to 0.$$

Обозначим

$$\eta_i^{(N)} = \eta_i^{\Delta}(t) I\big(\eta_i^{\Delta}(t) > -N\big),$$

где $I(A)$ — индикатор A. Тогда

$$\mathbf{P}_\theta\left(\sum \eta_i^{\Delta}(t) > 0\right) \leqslant \mathbf{P}_\theta\left(\sum \eta_i^{(N)} > 0\right) =$$

$$= \mathbf{P}_\theta\left(\sum \big(\eta_i^{(N)} - a_i(-N)\big) > -\sum a_i(-N)\right),$$

где

$$a_i(-N) \equiv \mathbf{E}_\theta \eta_i^{(N)} = \mathbf{E}\big(\eta_i^{\Delta}(t); \eta_i^{\Delta}(t) > -N\big).$$

По условию (A_1^{Δ})

$$-\frac{1}{b_n}\sum a_i(-N) > \varepsilon.$$

Поэтому нам достаточно убедиться, что

$$\mathbf{P}_\theta\left(\frac{1}{b_n}\sum \big(\eta_i^{(N)} - a_i(-N)\big) > \varepsilon\right) \to 0.$$

Воспользуемся следующим законом больших чисел (см. приложение IV).

Теорема А. *Пусть* $\xi_{i,n}$, $i = 1, \ldots, n$, *независимы и удовлетворяют условиям*

$$\mathbf{E}\xi_{i,n} = 0, \quad \sum \mathbf{E}|\xi_{i,n}| < c, \quad \sum \mathbf{E}(|\xi_{i,n}|; |\xi_{i,n}| > \tau) \to 0$$

при любом $\tau > 0$ *и* $n \to \infty$. *Тогда* $\mathbf{P}\left(\left|\sum \xi_{i,n}\right| > \varepsilon\right) \to 0$ *при* $n \to \infty$ *и любом* $\varepsilon > 0$.

Нам надо проверить, что величины $\xi_{i,n} = \dfrac{\eta_i^{(N)} - a_i(-N)}{b_n}$ удовлетворяют условиям теоремы А. Первое и второе условия выполнены очевидным образом. Далее, имеем $-N \leqslant a_i(-N) \leqslant \dfrac{\tau b_n}{3} + a_i\left(\dfrac{\tau b_n}{3}\right)$, где по условию (A_2^{Δ}) $a_i\left(\dfrac{\tau b_n}{3}\right) = o(b_n)$. Поэтому при достаточно больших n

$$-N \leqslant a_i(-N) < \frac{\tau b_n}{2}, \quad \eta_i^{(N)} - a_i(-N) > -N - \frac{\tau b_n}{2} > -\tau b_n$$

и событие $\{\eta_i^{(N)} - a_i(-N) < -\tau b_n\}$ невозможно. Следовательно, при достаточно больших n

$$\mathbf{E}\big(|\eta_i^{(N)} - a_i(-N)|; |\eta_i^{(N)} - a_i(-N)| > \tau b_n\big) =$$
$$= \mathbf{E}\big(\eta_i^{(N)} - a_i(-N); \eta_i^{(N)} - a_i(-N) > \tau b_n\big) \leqslant$$
$$\leqslant N\mathbf{P}\left(\eta_i^{(N)} > \frac{\tau b_n}{2}\right) + \mathbf{E}\left(\eta_i^{(N)}; \eta_i^{(N)} > \frac{\tau b_n}{2}\right);$$

$$\frac{1}{b_n}\sum_{i=1}^{n}\mathbf{E}\big(|\eta_i^{(N)} - a_i(-N)|; |\eta_i^{(N)} - a_i(-N)| > \tau b_n\big) \leqslant$$
$$\leqslant \frac{N}{b_n}\sum_{i=1}^{n}\mathbf{P}\left(\eta_i^{(N)} > \frac{\tau b_n}{2}\right) + \frac{1}{b_n}\sum_{i=1}^{n}\mathbf{E}\left(\eta_i^{(N)}; \eta_i^{(N)} > \frac{\tau b_n}{2}\right) \to 0$$

при $n \to \infty$ в силу (A_2^{Δ}) и (12). Теорема доказана. $\triangleleft$

Перейдем к выяснению условий, достаточных для выполнения (A_1^{Δ}), (A_2^{Δ}). Обозначим

$$\xi_i = \frac{\eta_i^{\Delta}(t)}{b_n(t)}, \quad \xi^+ = \max(0, \xi)$$

и рассмотрим условие

(L_2) *При каком-нибудь s выполняется*

$$\limsup_{n\to\infty}\sum_{i=1}^{n}\mathbf{E}_\theta \min\left(\xi_i^+, \left(\xi_i^+\right)^s\right) = 0.$$

Так как $\mathbf{E}\min\left(\xi_i^+, \left(\xi_i^+\right)^s\right) \leqslant \mathbf{E}\left(\xi_i^+\right)^s$, то это условие можно рассматривать как ослабление условия типа Ляпунова.

Условие (L_2) *достаточно для выполнения* (A_2^{Δ}).

Действительно $(\tau \leqslant 1)$,

$$\mathbf{E}_\theta\left(\frac{\eta_i^{\Delta}(t)}{b_n}; \frac{\eta_i^{\Delta}(t)}{b_n} > \tau\right) \leqslant \mathbf{E}_\theta\big(\xi_i^+, \xi_i^+ \geqslant 1\big) +$$
$$+ \tau^{1-s}\mathbf{E}_\theta\big(\left(\xi_i^+\right)^s; \tau < \xi_i^+ < 1\big) \leqslant \tau^{1-s}\mathbf{E}_\theta \min\left(\xi_i^+, \left(\xi_i^+\right)^s\right).$$

Из условий (A_1^{Δ}), (A_2^{Δ}) следует, что они не предполагают ограничений в области отрицательных значений $\eta_i^{\Delta}(t)$ того же типа, которые предполагаются условием (A_2^{Δ}) в области положительных значений $\eta_i^{\Delta}(t)$. Однако если такие ограничения ввести, то условия, достаточные для выполнения (A_1^{Δ}), (A_2^{Δ}), можно сделать более симметричными.

Обозначим в ближайших рассмотрениях

$$b_n(t) = \sum_{i=1}^{n} \mathbf{E}_\theta \left|\eta_i^\Delta(t)\right| \tag{13}$$

(это соответствует $N = \infty$ в (11)) и введем условия

(L_0) *При* $t \notin (\theta)^\delta$ *выполняется*

$$b_n(t) \to \infty, \quad \limsup_{n\to\infty} \sum_{i=1}^{n} \mathbf{E}_\theta \xi_i < -\varepsilon < 0.$$

(L_1) *При* $s > 1$ *выполняется*

$$\limsup_{n\to\infty} \sum_{i=1}^{n} \mathbf{E}_\theta \min\left(|\xi_i|, |\xi_i|^s\right) = 0.$$

Условие (L_1) эквивалентно условию (см. приложение IV): при любом $\tau > 0$

$$\limsup \sum \mathbf{E}_\theta(|\xi_i|, |\xi_i| > \tau) = 0. \tag{14}$$

Условия (L_0), (L_1) *влекут за собой* (A_1^Δ), (A_2^Δ).

Действительно, условие (L_1) влечет (L_2) и, следовательно, (A_2^Δ). Условие (A_1^Δ) при $N = \infty$ вытекает из (L_0).

Неудобство рассмотренных выше условий состоит в том, что они выражены в терминах величин $\eta_i^\Delta(t)$, $(\psi_i^\Delta(\mathrm{x}_i, t))$, а не $\eta_i(t)$ $(\psi(\mathrm{x}_i, t))$. Этого неудобства можно избежать, если потребовать, чтобы функции $\psi_i(\mathrm{x}_i, t)$ были в известном смысле непрерывны по t. Обозначим

$$\omega_i^\Delta(x,t) = \sup_{|u|\leqslant\Delta} |\psi_i(x, t+u) - \psi_i(x,t)|, \quad \omega_i^\Delta(t) = \mathbf{E}_\theta \omega_i^\Delta(\mathrm{x}_i, t),$$

$$\beta_n(t) = \sum_{i=1}^{n} \mathbf{E}_\theta \eta_i(t), \quad b_n(t) = \sum_{i=1}^{n} \mathbf{E}_\theta |\eta_i(t)|.$$

Может быть полезной

Лемма 1. *Пусть при* $t \notin (\theta)^\delta$, $c < \infty$

$$-\beta_n(t) \geqslant c b_n(t) \to \infty, \quad \sum \omega_i^\Delta(t) = o(b_n(t)) \quad \textit{при} \quad \Delta \to 0.$$

Тогда выполнено условие (A_1^Δ) *при* $N = \infty$.

Доказательство леммы 1 носит чисто технический характер и основано на неравенстве $\left|\eta_i^\Delta(t) - \eta_i(t)\right| \leqslant \omega_i^\Delta(\mathrm{x}_i, t)$. Мы предоставляем его читателю.

Отметим, что наиболее важным «качественным» элементом рассмотренных выше условий, обеспечивающих состоятельность M-оценок, является сходимость при $t \neq \theta$

$$\beta_n(t) = \sum_{i=1}^{n} \mathbf{E}_\theta \eta_i(t) \to -\infty. \tag{15}$$

Асимптотическая нормальность $\widehat{M}$-оценок требует выполнения условий дифференцируемости $\psi_i(x,t)$ по t. Поэтому ее естественно изучать в рамках п. 5, посвященного асимптотической нормальности M-оценок.

4. Состоятельность M-оценок. Чтобы упростить изложение, ограничимся здесь в основном рассмотрением одномерного случая, когда θ — скалярный параметр. Комментарии по поводу многомерного случая см. в конце параграфа. Как и в предыдущем пункте, мы будем предполагать, что выполнено условие компактности (A_c). Кроме того, нам понадобится ряд условий на исходные функции $g_i(x,\theta)$ (см. определение 2), которые в схеме серий могут зависеть от n.

(M_1) *Функции* $g_i(x,\theta)$ *интегрируемы и удовлетворяют условию типа Линдеберга в следующем смысле*:

$$\sup_{t\in\Theta}|g_i(x,t)| \equiv \gamma_i(x) < \infty, \quad \gamma_i \equiv \int \gamma_i(x)\mathbf{P}_{\theta,i}(dx) < \infty,$$

$$\Gamma^{-1}(n)\sum_{i=1}^{n}\int\limits_{\gamma_i(x)>\tau\Gamma(n)} \gamma_i(x)\mathbf{P}_{\theta,i}(dx) \to 0, \quad \Gamma(n) = \sum_{i=1}^{n}\gamma_i,$$

при $n\to\infty$, *где* θ — *«истинное» значение параметра, соответствующее выборке* X, $\tau > 0$ — *любое фиксированное число.*

(M_2) *Функции* $g_i(x,\theta)$ *непрерывны по* θ *«в среднем» в следующем смысле. Пусть*

$$\omega_{\Delta,i}(x) = \sup_{t\in\Theta,t+u\in\Theta,|u|\leqslant\Delta} |g_i(x,t+u) - g_i(x,t)|$$

есть модуль непрерывности g_i *на* Θ, $\omega_{\Delta,i} = \mathbf{E}_\theta\omega_{\Delta,i}(\mathrm{x}_i) = \int \omega_{\Delta,i}(x)\mathbf{P}_{\theta,i}(dx)$.

Мы будем предполагать, что

$$\Gamma^{-1}(n)\sum\omega_{\Delta,i} \leqslant \omega_\Delta \to 0$$

при $\Delta\to 0$, *где* ω_Δ *от* n *не зависит* (в лемме 14.1 было показано, что $\omega_{\Delta,i}\to 0$ при $\Delta\to 0$).

(M_3) $\theta\in\Theta$ *является единственным решением уравнения* $S(t) = 0$, *где* $S(t) = \sum s_i(t)$, $s_i(t) = \int g_i(x,t)\mathbf{P}_{t,i}(dx)$ (см. (2)). *Кроме того, функция* $s(t) = \Gamma^{-1}(n)S(t)$ *удовлетворяет условию: для любого достаточно малого* $\varepsilon > 0$ *найдется* $\delta > 0$ *такое, что*

$$\inf_{t\in\Theta,|\theta-t|>\delta}|s(t)| > \varepsilon \tag{16}$$

при всех n (ε, δ *от* n *не зависят*).

Теорема 2. *При выполнении условий* (M_1)–(M_3) *M-оценка* θ^* *состоятельна.*

Положим

$$s_n^*(t) = \Gamma^{-1}(n) \sum g(\mathrm{x}_i, t).$$

Для доказательства теоремы 2 нам понадобится

Лемма 2. *Если выполнены условия* (M_1), (M_2), *то**)

$$\sup_{t\in\Theta} \left| s_n^*(t) - s(t) \right| \xrightarrow[P]{} 0$$

при $n \to \infty$, *функция* $s(t)$ *непрерывна.*

Доказательство. Имеем

$$\omega_\Delta(X) \equiv \sup_{t\in\Theta, t+u\in\Theta, |u|\leqslant\Delta} \left| s_n^*(t+u) - s_n^*(t) \right| \leqslant \Gamma^{-1}(n) \sum_{i=1}^n \omega_{\Delta,i}(\mathrm{x}_i) =$$

$$= \Gamma^{-1}(n) \sum_{i=1}^n \omega_{\Delta,i} + \Gamma^{-1}(n) \sum_{i=1}^n (\omega_{\Delta,i}(\mathrm{x}_i) - \omega_{\Delta,i}). \quad (17)$$

Покажем, что второе слагаемое здесь сходится по вероятности к нулю. Это следует из закона больших чисел (теорема А) и следующего почти очевидного утверждения, которое означает, что слагаемые $\xi_{i,n} = \Gamma^{-1}(n) \times$ $\times(\omega_{\Delta,i}(\mathrm{x}_i) - \omega_{\Delta,i})$ удовлетворяют условиям теоремы А.

Лемма 3.

$$\mathbf{E}\xi_{i,n} = 0, \quad \sum \mathbf{E}|\xi_{i,n}| < \mathrm{const}, \quad \sum_{i=1}^n \mathbf{E}[(\xi_{i,n}); |\xi_{i,n}| > \tau) \to 0 \quad (18)$$

при $n \to \infty$ *и любом* $\tau > 0$.

Из этой леммы и из (17) следует, что $\omega_\Delta(X) \leqslant \omega_\Delta + \varepsilon_n(X)$, $\varepsilon_n(X) \xrightarrow[P]{} 0$.

Дальнейшее доказательство леммы 2 ничем не отличается от доказательства леммы 14.1 (надо лишь сходимость п. н. заменить на сходимость по вероятности).

Непрерывность $s(t)$ очевидна, так как

$$s(t+\Delta) - s(t) \leqslant \Gamma^{-1}(n) \sum (s_i(t+\Delta) - s_i(t)) \leqslant \Gamma^{-1}(n) \sum \omega_{\Delta,i} \leqslant \omega_\Delta. \quad (19)$$

Нам осталось доказать лемму 3. Имеем

$$\sum_{i=1}^n \mathbf{E}|\xi_{i,n}| = \Gamma^{-1}(n) \sum \mathbf{E}|\omega_{\Delta,i}(\mathrm{x}_i) - \omega_{\Delta,i}| \leqslant 2\Gamma^{-1}(n) \sum \omega_{\Delta,i} \leqslant 2\omega_\Delta.$$

Это доказывает неравенство в (18). Чтобы доказать сходимость в (18), заметим, что из (M_1) и комментариев к теореме 1 приложения IV следует

---

*) Здесь и в дальнейшем для краткости мы используем запись $\zeta_n \xrightarrow[P]{} 0$ вместо $\mathbf{P}(|\zeta_n| > \varepsilon) \to 0$ для любого $\varepsilon > 0$.

(см. там (6), (S_1)), что

$$\Gamma^{-1}(n)\max\gamma_i \to 0, \quad \sum \mathbf{P}(\gamma_i(\mathrm{x}_i) > \varepsilon\Gamma(n)) \to 0 \tag{20}$$

при $n \to \infty$. Отсюда получаем

$$\max_i \omega_{\Delta,i} \leqslant 2\max\gamma_i = o(\Gamma(n))$$

и следующие вложения для области интегрирования $B_{i,n} = \{|\xi_{i,n}| > \tau\}$ в (18) при достаточно больших n:

$$B_{i,n} \equiv \{|\omega_{\Delta,i}(\mathrm{x}_i) - \omega_{\Delta,i}| > \tau\Gamma(n)\} \subset$$
$$\subset \left\{\omega_{\Delta,i}(\mathrm{x}_i) > \frac{2\tau}{3}\Gamma(n)\right\} \subset \left\{\gamma_i(\mathrm{x}_i) > \frac{\tau\Gamma(n)}{3}\right\} \equiv A_{i,n}.$$

Следовательно, последняя сумма в (18) не превосходит

$$\sum \mathbf{E}(|\xi_{i,n}|; A_{i,n}) \leqslant \Gamma^{-1}(n)\sum \omega_{i,\Delta}\mathbf{P}(A_{i,n}) +$$
$$+ 2\Gamma^{-1}(n)\sum \mathbf{E}\left(|\gamma_i(\mathrm{x}_i)|; \gamma_i(\mathrm{x}_i) > \frac{\tau\Gamma(n)}{3}\right).$$

Здесь правая часть в силу (20), (M_1) и неравенств $\omega_{\Delta,i} \subseteq \omega_\Delta\Gamma(n)$ сходится к нулю при $n \to \infty$. Лемма 3, а вместе с ней и лемма 2 доказаны.◁

Доказательство теоремы 2 почти ничем не отличается от доказательства теоремы 14.3. Надо лишь сходимость п. н. заменить на сходимость по вероятности. Неравенства (14.14), используемые в доказательстве теоремы 14.3, вытекают из условия (M_3) (см. (16)).

5. Асимптотическая нормальность *M*-оценок. Здесь нам потребуются условия (M_1)–(M_3) и аналоги условий (M_1), (M_2) для производных $g_i'(x,\theta)$ по θ.

(M_1') *Функции $g_i(x,t)$ непрерывно дифференцируемы по θ,*

$$\sup_{t\in\Theta}|g_i'(x,t)| \equiv \gamma_i'(x) < \infty, \quad \gamma_i' \equiv \int \gamma_i'(x)\mathbf{P}_{\theta,i}(dx) < \infty,$$

$$\Gamma_1^{-1}(n)\sum_{i=1}^{n} \int\limits_{\gamma_i'(x)>\tau\Gamma_1(n)} \gamma_i'(x)\mathbf{P}_{\theta,i}(dx) \to 0$$

при $n \to \infty$, где $\Gamma_1(n) = \sum\limits_{i=1}^{n}\gamma_i'$, θ — истинное значение параметра, $\tau > 0$ — любое фиксированное число.

Положим

$$\omega_{\Delta,i}'(x) = \sup_{t\in\Theta,t+u\in\Theta,|u|\leqslant\Delta} |g_i'(x,t+u) - g_i'(x,t)|,$$

$$\omega_{\Delta,i}' = \mathbf{E}_\theta\omega_{\Delta,i}'(\mathrm{x}_i) = \int \omega_{\Delta,i}'(x)\mathbf{P}_{\theta,i}(dx).$$

(M_2') $\Gamma_1^{-1}(n)\sum \omega'_{\Delta,i} \leqslant \omega'_\Delta \to 0$ *при* $\Delta \to \infty$, *где* ω'_Δ *от* n *не зависит*;

$$|s'(\theta)| > \delta > 0.$$

Заметим, что в силу (M_2') законно дифференцирование интегралов $s_i(t)$ по параметру, так что

$$s'(t) = \Gamma_1^{-1}(n)\sum_{i=1}^{n} \mathbf{E}_\theta g_i'(\mathrm{x}_i, t).$$

Кроме того, мы будем предполагать выполненным условие Линдеберга

(M_4) $$D^{-1}(n)\sum \mathbf{E}_\theta\big(g_i^2(\mathrm{x}_i,\theta); |g_i(\mathrm{x}_i,\theta)| > \tau D(n)\big) \to 0,$$

где $D(n) = \sum \mathbf{E}_\theta g_i^2(\mathrm{x}_i,\theta)$.

Условия (M_1), (M_2) (они тоже будут нам нужны), как и в § 14, во многом являются следствиями условий (M_1'), (M_2'). Однако для краткости изложения не будем выделять минимальные требования, а *будем предполагать* (M_1), (M_2) *выполненными наряду с* (M_1'), (M_2'). Если состоятельность M-оценки θ^* будет установлена каким-либо другим способом (см. § 68), то нужда в условиях (M_1)–(M_3) в следующей теореме отпадает.

Теорема 3. *При выполнении условий* (M_1)–(M_4), (M_1'), (M_2') *M-оценка* θ^* *асимптотически нормальна с разбросом*

$$\sigma^2(\theta,n) = D(n)(s'(\theta))^{-2}\Gamma_1^{-2}(n) = D(n)\Big(\sum \mathbf{E}_\theta g'(\mathrm{x}_i,\theta)\Big)^{-2}.$$

Доказательство теоремы 3 аналогично доказательству теоремы 14.4. Условия теоремы 3 влекут за собой условия теоремы 2 и, следовательно, $\theta^* \xrightarrow[P]{} \theta$. Функция

$$s_n^*(\theta) = \Gamma_1^{-1}(n)\sum g_i(\mathrm{x}_i,\theta)$$

непрерывно дифференцируема и допускает представление

$$0 = s_n^*(\theta^*) = s_n^*(\theta) + (\theta^* - \theta)\big(s_n^*(\widetilde{\theta})\big)', \quad \widetilde{\theta} \in [\theta,\theta^*],$$

где $\widetilde{\theta} \xrightarrow[P]{} \theta$.

В силу условий (M_1'), (M_2') функция $\big(s_n^*(t)\big)$ удовлетворяет условиям леммы 2 и, следовательно,

$$\sup_{t\in\Theta}\big|\big(s_n^*(t)\big)' - s'(t)\big| \xrightarrow[P]{} 0.$$

Кроме того, аналогично предыдущему (см. (19)) получаем

$$\sup_{|u|\leqslant\Delta}|s'(t+u) - s'(t)| \leqslant \omega'_\Delta \to 0$$

при $\Delta \to 0$. Из этих соотношений и (M_2') следует, что

$$\bigl(s_n^*(\widetilde{\theta})\bigr)' - s'(\widetilde{\theta}) \underset{P}{\longrightarrow} 0, \quad s'(\widetilde{\theta}) - s'(\theta) \underset{P}{\longrightarrow} 0,$$

$$\bigl(s_n^*(\widetilde{\theta})\bigr)' = s'(\theta)(1+\varepsilon_n), \quad \varepsilon_n \underset{P}{\longrightarrow} 0.$$

Это дает нам

$$(\theta^* - \theta) = -s_n^*(\theta)(s'(\theta))^{-1}(1+\varepsilon_n)^{-1},$$

где в силу центральной предельной теоремы (см. приложение IV) и условия (M_4) сумма $\Gamma_1(n)s_n^*(\theta)$ асимптотически нормальна с параметрами $\bigl(0,\, D(n)\Gamma_1^{-2}(n)(s'(\theta))^{-2}\bigr)$. Остается воспользоваться теоремой непрерывности. Теорема 3 доказана. $\triangleleft$

Изучение асимптотических свойств M-оценок в *многомерном случае* происходит аналогичным образом, но технически сложнее. Поэтому, как и в § 14, ограничимся здесь краткими комментариями по поводу асимптотической нормальности. В многомерном случае функции g_i являются k-мерными векторами $g_i(x,\theta) = \bigl(g_i^{(1)}, \ldots, g_i^{(k)}\bigr)$ и M-оценки суть решения системы из k уравнений (в векторной форме)

$$\sum g_i(\mathrm{x}_i, \theta) = 0.$$

Характер условий на функцию g_i остается прежним. Роль величины $D(n)$ в условии (M_4) может играть теперь минимальное собственное значение матрицы

$$D(\theta, n) = \sum_{i=1}^{n} G_i, \quad G_i = \bigl\|\mathbf{E}_\theta g_i^\top(\mathrm{x}_i, \theta) g_i(\mathrm{x}_i, \theta)\bigr\|.$$

Роль матрицы $n\boldsymbol{s}'(\theta)$ в § 14 будет играть сумма

$$\boldsymbol{S}'(\theta, n) = \sum_{i=1}^{n} \boldsymbol{s}_i'(\theta),$$

где $\boldsymbol{s}_i'(t)$ — матрица, составленная из столбцов $\bigl(\operatorname{grad} s_i^{(1)}(t)\bigr)^\top, \ldots$ $\ldots, \bigl(\operatorname{grad} s_i^{(k)}(t)\bigr)^\top$, $s_i^{(r)}(\theta) = \mathbf{E}_\theta g_i^{(r)}(\mathrm{x}_i, \theta)$. Основное асимптотическое соотношение для $\theta^* - \theta$ остается прежним:

$$(\theta^* - \theta) \approx -S_n^*(\theta)(\boldsymbol{S}'(\theta, n))^{-1}, \quad S_n^*(\theta) = \sum g_i(\mathrm{x}_i, \theta).$$

Отсюда при выполнении соответствующих условий можно получить, что θ^* будет асимптотически нормальна с разбросом

$$[(\boldsymbol{S}'(\theta, n))^{-1}]^\top D(\theta, n)(\boldsymbol{S}'(\theta, n))^{-1}. \tag{21}$$

§ 66. Оценки максимального правдоподобия. Основные принципы сравнения оценок. Оптимальность о. м. п. в классе M-оценок

1. Оценки максимального правдоподобия. С методами M-оценивания тесно связан общий метод построения оценок, называемый методом максимального правдоподобия (ср. с гл. 2).

Предположим, что $X \in \mathbf{P}_\theta$, распределения $\mathbf{P}_{\theta,i}$ имеют плотности $f_{\theta,i}(x)$ относительно некоторой меры μ_i, не зависящей от θ.

Определение 1. *Функцией правдоподобия* (относительно меры $\mu_1 \times \ldots \times \mu_n$) выборки $X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ называется функция

$$f_\theta(X) = \prod_{i=1}^{n} f_{\theta,i}(\mathrm{x}_i).$$

Логарифмической функцией правдоподобия называется функция

$$L(X,\theta) = \ln f_\theta(X) = \sum l_i(\mathrm{x}_i, \theta),$$

где $l_i(x,\theta) = \ln f_{\theta,i}(x)$. *Оценкой максимального правдоподобия* $\widehat{\theta}^*$ параметра θ называется значение θ, при котором достигается $\max f_\theta(X)$ по θ.

Очевидно, что о. м. п. $\widehat{\theta}^*$ является $\widehat{M}$-оценкой при $\psi_i(x,\theta) = l_i(x,\theta)$.

Читатель без труда может построить функции правдоподобия в примерах 64.1–64.7. Например, в условиях примера 64.1 $\theta = (\alpha, \sigma)$,

$$f_\theta(X) = (2\pi)^{-n/2}\sigma^{-n}\left(\prod u_i\right)^{-1/2} \exp\left\{-\sum \frac{(\mathrm{x}_i - \alpha t_i)^2}{2\sigma^2 u_i}\right\}. \qquad (1)$$

Логарифмическая функция правдоподобия $L(X,\theta)$ здесь имеет вид

$$L(X,\theta) = c_n - n\ln\sigma - \sum \frac{(\mathrm{x}_i - \alpha t_i)^2}{2\sigma^2 u_i},$$

где c_n от θ и X не зависит.

В более общей постановке, относящейся к схеме серий (см. примеры 64.3–64.7), мы должны предполагать, что распределения $\mathbf{P}_{\theta,i}$ наблюдений x_i зависят также и от n: $\mathbf{P}_{\theta,i} = \mathbf{P}^{(n)}_{\theta,i}$. Зависеть от n могут также меры $\mu_1^{(n)}, \ldots, \mu_n^{(n)}$ и, несомненно, плотности $f^{(n)}_{\theta,i}(x)$ относительно меры $\mu_i^{(n)}$.

В дальнейшем верхний индекс (n) для краткости мы будем опускать. Если рассмотрения исключают схему серий, это будет оговариваться особо.

Если функции $f_{t,i}(x)$ (или $l_i(x,t)$) непрерывно дифференцируемы по $t = (t_1, \ldots, t_k)$, то точка максимума $L(X,t)$ удовлетворяет уравнениям

$$\frac{\partial L(X,t)}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial}{\partial t_j} l_i(\mathrm{x}_i, t) = 0, \quad j = 1, \ldots, k.$$

Так как, кроме того, $\int l_i(x,t)f_{\theta,i}(x)\mu_i(dx)$ достигает максимума при $t=\theta$ (см. § 16), то в случае законности дифференцирования этих интегралов по t получаем

$$\int \frac{\partial}{\partial t_j} l_i(x,\theta) f_{\theta,i}(x)\mu_i(dx) = 0. \tag{2}$$

Это означает, что при выполнении названных условий *о. м. п. является М-оценкой.* Равенство (2) также можно получить, дифференцируя тождество $\int f_{\theta,i}(x)\mu_i(dx) = 1$, если законно дифференцирование под знаком интеграла.

Для выборки X из нормальных распределений в примере 64.1 эти условия выполнены, и о. м. п. является решением уравнений $\dfrac{\partial L}{\partial \alpha} = 0$, $\dfrac{\partial L}{\partial \sigma} = 0$ или, что то же,

$$\sum \frac{(\mathrm{x}_i - \alpha t_i)t_i}{u_i} = 0, \quad \sum \frac{(\mathrm{x}_i - \alpha t_i)^2}{\sigma^3 u_i} = \frac{n}{\sigma},$$

так что

$$\widehat{\alpha}^* = \frac{\sum \mathrm{x}_i t_i/u_i}{\sum t_i^2/u_i}, \tag{3}$$

$$(\widehat{\sigma}^2)^* = \frac{1}{n}\sum \frac{(\mathrm{x}_i - \widehat{\alpha}^* t_i)^2}{u_i} = \frac{1}{n}\left(\sum \frac{\mathrm{x}_i^2 - (\widehat{\alpha}^*)^2 t_i^2}{u_i}\right).$$

Для пуассоновских наблюдений $\mathrm{x}_i \in \mathbf{\Pi}_{\lambda t_i}$ функция правдоподобия и о. м. п. $\widehat{\lambda}^*$ соответственно имеют вид

$$f_\lambda(X) = \prod_{i=1}^n e^{-\lambda t_i}\frac{(\lambda t_i)^{\mathrm{x}_i}}{\mathrm{x}_i!}, \quad \widehat{\lambda}^* = \frac{\sum \mathrm{x}_i}{\sum t_i}.$$

Мы сразу получили здесь оценки, оптимальные в классе оценок по методу моментов и найденные нами в п. 65.2.

Отметим, что других общих «регулярных» методов построения оценок для разнораспределенных наблюдений, помимо M-оценивания и метода максимального правдоподобия как важнейшего частного случая, по существу, нет. Поэтому большое значение приобретает метод последовательных приближений при отыскании о. м. п., изложенный в § 35.

Среди немногих других методов можно указать подходы к построению оценок, связанные с некоторыми частными предположениями. Допустим, например, что при каждом i можно построить несмещенную оценку θ_i^* параметра θ, построенную по одному наблюдению x_i: $\mathbf{E}_\theta \theta_i^* = \theta$ (имеется в виду математическое ожидание по распределению $\mathbf{P}_{\theta,i}$). Тогда оценка $\theta^* = \sum p_i \theta_i^*$, $\sum_{i=1}^n p_i = 1$ также будет несмещенной и может обладать хорошими асимптотическими свойствами.

2. Асимптотические свойства о. м. п. Рассмотрим сначала одномерный случай $\Theta \subset \mathbb{R}$. Как уже отмечалось, при выполнении некоторых условий регулярности о. м. п. являются M-оценками. Это дает возможность воспользоваться результатами предыдущего параграфа (теоремы 65.1–65.3), в которых следует положить

$$\psi_i(x,\theta) = l_i(x,\theta), \quad g_i(\mathrm{x}_i,\theta) = l_i'(\mathrm{x}_i,\theta).$$

Из теоремы 65.1 вытекает

Теорема 1. *Пусть функции $\psi_i(x,\theta) = l_i(x,\theta)$ удовлетворяют условиям теоремы* 65.1. *Тогда о. м. п. $\widehat{\theta}^*$ состоятельна.*

Условие (65.15) для о. м. п. превращается в условие неограниченного роста при $t \neq 0$ сумм расстояний Кульбака–Лейблера

$$\sum_{i=1}^{n} \int f_{\theta,i}(x) \ln \frac{f_{\theta,i}(x)}{f_{t,i}(x)} \mu(dx) \to \infty. \tag{4}$$

Пример 1. Рассмотрим пример 64.2, в котором $\mathrm{x}_i = \varphi_i(\theta) + \xi_i$ (мы положили здесь $\varphi(\theta, \mathrm{z}_i) = \varphi_i(\theta)$), и предположим, что ξ_i имеют всюду положительную плотность f относительно меры Лебега. В этом случае

$$f_{\theta,i}(x) = f(x - \varphi_i(\theta)), \quad \psi_i(x,\theta) = l_i(x,\theta) \equiv \ln f(x - \varphi_i(\theta)).$$

Пусть $\rho_i(\theta,t)$ — расстояние Кульбака–Лейблера между $f_{\theta,i}$ и $f_{t,i}$:

$$\rho_i(\theta,t) = \int f(x - \varphi_i(\theta)) \ln \frac{f(x - \varphi_i(\theta))}{f(x - \varphi_i(t))}\, dx = \\ = \int f(x) \ln \frac{f(x)}{f(x - \varphi_i(t) + \varphi_i(\theta))}\, dx.$$

Предположим, что для $t \notin (\theta)^\delta$ при любом фиксированном $\delta > 0$ и $n \to \infty$

$$b_n(t) \equiv \sum_{i=1}^{n} \rho_i(\theta,t) \to \infty. \tag{5}$$

Пусть, кроме того, при $n \to \infty$ и некотором $\gamma > 0$

$$\sum_{i=1}^{n} \int f(x) \left| \ln \frac{f(x)}{f(x - \varphi_i(t) + \varphi_i(\theta))} \right|^{1+\gamma} dx = o\big(b_n^{1+\gamma}\big), \tag{6}$$

$$\sum_{i=1}^{n} \int f(x) r_i(x,t,\Delta)^{1+\gamma}\, dx = o\big(b_n^{1+\gamma}\big) \tag{7}$$

и для любого $\varepsilon > 0$ найдется $\Delta = \Delta(\varepsilon)$ такое, что

$$\sum_{i=1}^{n} \int f(x) r_i(x,t,\Delta)\, dx \leqslant \varepsilon b_n, \tag{8}$$

где

$$r_i(x,t,\Delta)=\sup_{|u|\leqslant\Delta}\left|\ln\frac{f(x-\varphi_i(t+u)+\varphi_i(\theta))}{f(x-\varphi_i(t)+\varphi_i(\theta))}\right|,$$
$$b_n=\inf_{t\notin(\theta)^\delta} b_n(t)\to\infty. \tag{9}$$

Тогда условия теоремы 65.1 выполнены. Чтобы убедиться в этом, проверим выполнение условий (L_0), (L_1) из § 65.

Из (5), (8), (9) следует (L_0) при любом $\varepsilon\in(0,1)$ (определения $b_n(t)$ в (5) и (L_0) близки, но не совпадают). Условие (L_1) следует из (7). Стало быть, о. м. п. $\widehat{\theta}^*$ при выполнении (6)–(9) состоятельна.

Асимптотическая постановка задачи в той конкретизации примера 64.2, которая дана в § 65, приводит к схеме серий, когда в (64.3) вместо $\varphi_i(\lambda,h)=\sqrt{(\lambda-ivd)^2+h^2}$ (здесь $\theta=(\lambda,h)$) следует рассматривать $\varphi_{i,n}(\lambda,h)=\sqrt{\lambda^2(1-i\beta/n)^2+h^2}$, $\beta<1$ (интервал d между наблюдениями мал и равен $\lambda\beta/vn$). Если, кроме того, в этом примере $\xi_i\Subset\mathbf{\Phi}_{0,1}$, то мы получим (см. (9))

$$r_i(x,t,\theta)=\sup_{|u|\leqslant\Delta}\Big|x(\varphi_i(t+u)-\varphi_i(t))-$$
$$-\frac{1}{2}(\varphi_i(t+u)-\varphi_i(\theta))^2+\frac{1}{2}(\varphi_i(t)-\varphi_i(\theta))^2\Big|\leqslant c_1|x|\Delta+c_2$$

при всех i. Отсюда очевидно выполнение (6)–(9) при $b_n=cn$.

Перейдем теперь к *асимптотической нормальности*. Сначала опять ограничимся рассмотрением одномерного случая $\Theta\subset\mathbb{R}$. Аргументы у функций $g_i(x,\theta)$, $f_{\theta,i}(x)$, $l_i(x,\theta)$ иногда для краткости будем опускать.

Теорема 2. *Пусть некоторый набор функций $\{g_i\}$ и функции l_i' удовлетворяют условиям теоремы* 65.3. *Пусть, кроме того, законно дифференцирование под знаком интеграла в равенствах*

$$\int f_{\theta,i}=0,\quad \int g_i f_{\theta,i}=0,\quad \int l_i' f_{\theta,i}=0$$

(*последнее получено дифференцированием первого*). *При выполнении этих условий М-оценка θ^*, соответствующая функциям $\{g_i\}$, и о. м. п. $\widehat{\theta}^*$ асимптотически нормальны. Разброс $\sigma^2(\theta,n)$ оценки θ^* удовлетворяет неравенству*

$$\sigma^2(\theta,n)=\Big(\sum\mathbf{E}_\theta g_i^2(\mathrm{x}_i,\theta)\Big)\Big(\sum\mathbf{E}_\theta g_i'(\mathrm{x}_i,\theta)\Big)^{-2}\geqslant\Big(\sum I_i(\theta)\Big)^{-1}, \tag{10}$$

где $I_i(\theta)$ — информация Фишера наблюдения x_i:

$$I_i(\theta)=\mathbf{E}_\theta\big(l_i'(\mathrm{x}_i,\theta)\big)^2,$$

и достигает своего нормального значения, равного $\left(\sum I_i(\theta)\right)^{-1}$ *при* $g_i(x,\theta) = l_i'(x,\theta)$.

В этом утверждении нуждается в доказательстве лишь та его часть, которая связана с неравенством (10). Так как

$$\left|\mathbf{E}_\theta g_i'(\mathrm{x}_i,\theta)\right| = \left|\int g_i(x,t) f_{\theta,i}'(x)\mu_i(dx)\right| \leqslant \left[\mathbf{E}_\theta g_i^2(\mathrm{x}_i,\theta) I_i(\theta)\right]^{1/2},$$

то

$$\left(\sum \mathbf{E}_\theta g_i'(\mathrm{x}_i,\theta)\right)^2 \leqslant \left[\sum \left(\mathbf{E}_\theta g_i^2(\mathrm{x}_i,\theta)\right)^{1/2} I_i^{1/2}(\theta)\right]^2 \leqslant$$
$$\leqslant \left(\sum \mathbf{E}_\theta g_i^2(\mathrm{x}_i,\theta)\right)\left(\sum I_i(\theta)\right).$$

В силу условий теоремы все сказанное применимо и к функциям $g_i = l_i'$, для которых в (10) достигается равенство. Теорема 2 доказана.

В *многомерном случае* при выполнении соответствующих условий регулярности на вектор-функции $g_i = \left(g_i^{(1)},\dots,g_i^{(k)}\right)$ и $l_i' = \left(\dfrac{\partial l_i}{\partial\theta_1},\dots,\dfrac{\partial l_i}{\partial\theta_k}\right)$ M-оценка θ^* и о. м. п. $\widehat{\theta}^*$ будут а. н., при этом разброс M-оценки θ^* (см. (65.21)) будет удовлетворять неравенству

$$[(\boldsymbol{S}'(\theta,n))^{-1}]^{\top} D(\theta,n)(\boldsymbol{S}'(\theta,n))^{-1} \geqslant I^{-1}(\theta,n),$$

где $I(\theta,n) = \sum I_i(\theta)$, $I_i(\theta)$ — информационная матрица i-го эксперимента

$$I_i(\theta) = \left\|\mathbf{E}_\theta \frac{\partial l_i}{\partial\theta_r}\frac{\partial l_i}{\partial\theta_j}\right\|, \quad l_i = l_i(x,\theta) = \ln f_{\theta,i}(x).$$

При $g_i = l_i'$ в этом неравенстве $D(\theta,n) = I(\theta,n)$, $\boldsymbol{S}'(\theta,n) = I(\theta,n)$ и оно превращается в равенство.

В примере 64.2 при гладких функциях $\varphi_i(\theta)$ все требуемые условия будут выполнены, так что о. м. п. $\widehat{\theta}^*$ будет асимптотически оптимальной (см. ниже) в классе M-оценок и иметь разброс $I^{-1}(\theta,n)$. При $\xi_i \in \boldsymbol{\Phi}_{0,1}$ будем иметь

$$l_i(x,\theta) = -\ln\sqrt{2\pi} - \frac{(x-\varphi_i(\theta))^2}{2},$$

$$\frac{\partial l_i}{\partial\theta_r} = (x-\varphi_i(\theta))\frac{\partial\varphi_i}{\partial\theta_r}, \quad r = 1,\dots,k.$$

Поэтому здесь

$$I_i(\theta) = \left\|I_i^{(rs)}(\theta)\right\|, \quad I_i^{(rs)}(\theta) = \frac{\partial\varphi_i}{\partial\theta_r}\frac{\partial\varphi_i}{\partial\theta_s}.$$

О приближенном вычислении о. м. п. в многомерном случае см. § 35.

Отметим, что иногда приближенное представление функции правдоподобия может подсказать удобную явную форму оценок, а асимптотические свойства этих оценок проще устанавливать непосредственно, исходя из их явного вида (а не через весьма громоздкие условия теорем 1, 2). Проиллюстрируем это на следующем примере.

Пример 2. Рассмотрим пример 64.3 в случае, когда y_i неотрицательны; s_i малы ($s_i \to 0$ при $n \to \infty$), μ — мера Лебега. В этом случае плотность $f_{\theta,i}(x)$ относительно меры μ_i, введенной в примере 64.3, будет иметь вид

$$f_{\theta,i}(x) = \begin{cases} f_\theta(x) & \text{при} \quad x < s_i, \\ 1 - Q(\theta, s_i) & \text{при} \quad x = s_i, \end{cases}$$

$$Q(\theta, s) = \int_0^s f_\theta(x)\, dx.$$

Будем считать, что $f_\theta(x)$ непрерывна по x в точке $x = 0$, и рассмотрим два случая:

a) $f_\theta(0) = a_0(\theta)$ монотонно зависит от θ;

b) $a_0(\theta) = a_0 > 0$ не зависит от θ, $f_\theta(x)$ дифференцируема по x в точке $x = 0$:

$$f_\theta(x) = a_0 + x a_1(\theta) + o(x),$$

где $a_1(\theta) = \dfrac{\partial f_\theta(x)}{\partial x}\Big|_{x=0}$ монотонно зависит от θ.

В первом случае для простоты отождествим $a_0(\theta)$ и θ. Тогда $Q(\theta, s) = \theta s + o(s)$ и логарифмическая функция правдоподобия будет приближенно равна

$$L(X, \theta) = \sum_{i=1}^n I(\mathrm{x}_i < s_i) \ln \theta + \sum_{i=1}^n I(\mathrm{x}_i = s_i) \ln (1 - s_i\theta).$$

Приравнивая к нулю производную

$$L'(X, \theta) = \frac{1}{\theta} \sum I(\mathrm{x}_i < s_i) - \sum \frac{s_i I(\mathrm{x}_i = s_i)}{1 - s_i\theta}$$

и решая приближенно (при малых s_i) полученное уравнение, получим значение

$$\frac{\sum I(\mathrm{x}_i < s_i)}{\sum s_i I(\mathrm{x}_i = s_i)}, \tag{11}$$

которое дает нам основание в качестве оценки θ^* рассмотреть значение $\theta^* = \dfrac{\sum I(\mathrm{x}_i < s_i)}{\sum s_i}$, близкое к (11). Покажем, что *если*

$$S_n \equiv \sum_{i=1}^n s_i \to \infty \quad \textit{при} \quad n \to \infty, \tag{12}$$

то оценка θ^* *состоятельна* (условие (12) соответствует (65.15)). Действительно, в этом случае

$$S_n^{-1}\sum_{i=1}^{n}\mathbf{P}_\theta(\mathrm{x}_i < s_i) = S_n^{-1}\sum_{i=1}^{n}(\theta s_i + o(s_i)) = \theta + o(1)$$

и, стало быть,

$$\theta^* - \theta = \sum_{i=1}^{n}\xi_{i,n} + o(1),$$

где

$$\xi_{i,n} = S_n^{-1}(J_i - \mathbf{E}_\theta J_i), \quad J_i = J(\mathrm{x}_i < s_i).$$

В силу закона больших чисел (см. приложение IV) $\sum_{i=1}^{n}\xi_{i,n} \xrightarrow[P]{} 0$, что доказывает состоятельность θ^*.

Чтобы установить, что θ^* является а. н. оценкой, предположим дополнительно, что функции $f_\theta(x)$ дифференцируемы по x в точке $x = 0$, $\sum s_i^2 = o(\sqrt{S_n})$, $\theta \neq 0$. Тогда

$$\mathbf{E}_\theta\sum J_i = \sum\left[\theta s_i + \frac{a_1(\theta)s_i^2}{2} + o(s_i^2)\right] = \theta S_n + O\Big(\sum s_i^2\Big) = \theta S_n + o(\sqrt{S_n}),$$

$$\mathbf{D}_\theta\sum J_i = \theta S_n + o(S_n)$$

и, стало быть,

$$(\theta^* - \theta)\sqrt{S_n} = \frac{\sum(J_i - \mathbf{E}_\theta J_i)}{\sqrt{S_n}} + o(1),$$

где распределение первого слагаемого в правой части сходится при $n \to \infty$ к $\mathbf{\Phi}_{0,\theta}$. Итак, θ^* — а. н. оценка с разбросом θS_n^{-1}.

Рассмотрим теперь случай b), в котором мы положим, не ограничивая общности (с точки зрения последующих выводов), $a_0 = 1$ и отождествим $a_1(\theta)$ с θ. Тогда $Q(\theta, s) = s + \dfrac{\theta s^2}{2} + o(s^2)$ и логарифмическая функция правдоподобия с точностью до величин более высокого порядка малости будет иметь вид

$$L(X,\theta) = \sum J_i \ln(1 + \theta\mathrm{x}_i) + \sum I(\mathrm{x}_i = s_i)\ln\left(1 - s_i - \frac{s_i^2}{2}\theta\right).$$

В этом случае приближенное решение уравнения $L'(X,\theta) = 0$ показывает, что естественно использовать оценку

$$\theta^* = 3\frac{\sum(\mathrm{x}_i J_i - s_i^2/2)}{\sum s_i^3}.$$

Нетрудно видеть, что если $\sum_{i=1}^{n} s_i^3 \to \infty$ при $n \to \infty$, то из закона больших чисел будет следовать состоятельность θ^*. Мы предлагаем читателю найти условия, при которых эта оценка будет асимптотически нормальна.

О приближенном вычислении о. м. п. в общем случае см. § 35.

3. Основные принципы сравнения оценок. Асимптотическая эффективность о. м. п. в классе M-оценок. Принципы сравнения оценок для разнораспределенных наблюдений остаются, по существу, теми же, что были рассмотрены в гл. 2: это среднеквадратический и асимптотический подходы (более общие подходы см. в гл. 6).

Понятие *эффективной оценки в классе* K_b оценок с фиксированным смещением $b(\theta)$ (и, в частности, в классе K_0 несмещенных оценок) сохраняется полностью. Сохраняется и утверждение теоремы о единственности эффективных оценок, так как его доказательство с объемом выборки никак не связано и, стало быть, применимо к выборке, состоящей из одного наблюдения X.

Асимптотический подход к сравнению оценок естествен в классе K_Φ всех а. н. оценок. Определение а. н. оценок было дано в начале § 65. Главной характеристикой а. н. оценки является ее разброс $\sigma^2(\theta, n)$. Класс K_Φ а. н. оценок здесь в известном смысле шире, чем в однородном случае, так как он включает в себя а. н. оценки с разными скоростями убывания $\sigma^2(\theta, n)$ (а не только $\sigma^2(n)(\theta, n) = \sigma^2 n^{-1}$, как это было в гл. 2).

Если θ_k^*, $k = 1, 2$, суть две а. н. скалярные оценки с разбросами $\sigma_k^2(\theta, n)$, $k = 1, 2$, соответственно, то, очевидно, оценка θ_1^* будет асимптотически предпочтительней, если $\sigma_1(\theta, n) \leqslant \sigma^2(\theta, n)$ при всех достаточно больших n и всех $\theta \in \Theta$.

Определение 2. Оценка $\theta_1^* \in K$ называется *асимптотически эффективной* (а. э.) в K, если для любой другой оценки $\theta^* \in K$

$$\limsup \frac{\sigma_1^2(\theta, n)}{\sigma^2(\theta, n)} \leqslant 1 \quad \text{при всех} \quad \theta \in \Theta, \tag{13}$$

где $\sigma_1^2(\theta, n)$ и $\sigma^2(\theta, n)$ суть разбросы оценок θ_1^* и θ^* соответственно.

Мы можем в качестве K взять, например, класс K_M M-оценок, построенных по функциям g_i, удовлетворяющим условиям теоремы 65.3. Тогда из теоремы 2 будет вытекать следующее утверждение.

Следствие 1. *При выполнении условий регулярности, приведенных в теореме* 2, *о. м. п.* $\widehat{\theta}^*$ *является а. э. оценкой в* K_M.

Роль этого утверждения для разнораспределенных наблюдений представляется более значительной, чем в однородном случае, так как здесь M-оценивание, как уже отмечалось, по сути, исчерпывает все общие «регулярные» методы получения оценок. Тем не менее представляет интерес и другой подход к отысканию асимптотически оптимальных оценок, изложенный в § 18 и состоящий в построении а. э. оценок во всем классе а. н. оценок K_Φ, лишь

незначительно суженном до класса $K_{\Phi,2}$ оценок, для которых сходимость $(\theta^* - \theta)\sigma^{-1}(\theta, n) \Rightarrow \Phi_{0,1}$ имеет место вместе с двумя первыми моментами:

$$\mathbf{E}_\theta \frac{(\theta^* - \theta)}{\sigma(\theta, n)} \to 0, \quad \mathbf{E}_\theta \frac{(\theta^* - \theta)^2}{\sigma^2(\theta, n)} \to 1$$

(здесь $\mathbf{E}_\theta$ означает интегрирование по распределению $\mathbf{P}_{\theta,1} \times \ldots \times \mathbf{P}_{\theta,n}$ выборки X).

Аналогично § 18 асимптотическая эффективность оценки $\theta_1^* \in K \subset K_{\Phi,2}$ в классе K (согласно определению 18.2) будет эквивалентна неравенству

$$\limsup \frac{\mathbf{E}_\theta (\theta_1^* - \theta)^2}{\mathbf{E}_\theta (\theta^* - \theta)^2} \leqslant 1$$

для любой другой оценки $\theta^* \in K$ и при всех $\theta \in \Theta$.

Сохраняет свою силу аналог теоремы об асимптотической единственности а. э. оценок: *если* θ_1^* *и* θ_2^* *— две а. э. оценки в* $K \subset K_{\Phi,2}$ *такие, что* $\frac{1}{2}(\theta_1^* + \theta_2^*) \in K$, *то они асимптотически совпадают*, т. е.

$$\frac{\theta_1^* - \theta_2^*}{\sigma(\theta, n)} \underset{P}{\longrightarrow} 0, \quad \mathbf{E}_\theta \frac{(\theta_1^* - \theta_2^*)^2}{\sigma^2(\theta, n)} \to 0.$$

Доказательство повторяет рассуждения теоремы 18.2.

Переход к *многомерному случаю* выглядит так же, как в § 18. Однако здесь возникают и некоторые усложнения, связанные в тем, как понимать соотношение (13) определения 2 в многомерном случае. Наиболее естественный аналог этого определения в многомерном случае выглядит следующим образом. Оценка $\widetilde{\theta}^*$ называется *асимптотически эффективной* в $K \in K_{\Phi,2}$, если для любой другой оценки $\theta^* \in K$ и для любого $a = (a_1, \ldots, a_k)$

$$\limsup \frac{\sum \widetilde{\sigma}_{ij}(\theta, n) a_i a_j}{\sigma_{ij}(\theta, n) a_i a_j} \leqslant 1, \tag{14}$$

где $\widetilde{\sigma}^2 = (\widetilde{\sigma}_{ij})$ и $\sigma^2 = (\sigma_{ij})$ — матрицы разброса соответственно оценок $\widetilde{\theta}^*$ и θ^*).

В классе $K_{\Phi,2}$ неравенство (14) эквивалентно неравенству

$$\limsup \frac{\mathbf{E}(\widetilde{\theta}^* - \theta, a)^2}{\mathbf{E}(\theta^* - \theta, a)^2} \leqslant 1.$$

Понятия байесовской и минимаксной оценок с объемом выборки никак не связаны (см. также гл. 6) и потому в условиях этой главы полностью сохраняются. Сохраняются также все основные определения и утверждения § 21 о байесовских и минимаксных оценках.

§ 67. Достаточные статистики. Эффективные оценки. Экспоненциальные семейства

Определение достаточной статистики, утверждение основной факторизационной теоремы 22.1 и остальные утверждения § 22 для разнораспределенных наблюдений полностью сохраняются (они не связаны с объемом выборки).

В примере 64.1 функция правдоподобия имеет вид 66.1, и, стало быть, достаточная статистика равна $\left(\sum \frac{\mathrm{x}_i^2}{u_i}, \sum \frac{\mathrm{x}_i t_i}{u_i}\right)$. Для пуассоновских наблюдений $\mathrm{x}_i \in \mathbf{\Pi}_{\lambda t_i}$ достаточная статистика остается той же, что и в однородном случае, и равна $\sum \mathrm{x}_i$.

В примере 64.2 для $\xi_i \in \mathbf{\Phi}_{0,\sigma^2}$ нетривиальная достаточная статистика отсутствует. В задаче В примера 64.7 функция правдоподобия имеет вид

$$f_\theta(X) = (1-p)^{n-\nu} \prod_{i:\mathrm{x}_i>N} \gamma \mathrm{x}_i^{-\alpha-1} = (1-p)^n \left(\frac{\gamma}{1-p}\right)^\nu \prod_{i:\mathrm{x}_i>N} \mathrm{x}_i^{-\alpha-1} =$$

$$= [(1-p)N^{\alpha+1}]^n \left(\frac{\gamma}{(1-p)N^{\alpha+1}}\right)^\nu \prod \mathrm{x}_i^{-\alpha-1}, \quad (1)$$

где $p = p(\theta, N) = \dfrac{\gamma N^{-\alpha}}{\alpha}$, ν — число наблюдений x_i, отличных от N. Стало быть, достаточная статистика здесь равна $\left(\nu, \prod\limits_{i:\mathrm{x}_i>N} \mathrm{x}_i\right)$ или $\left(\nu, \prod \mathrm{x}_i\right)$.

Процедура построения эффективных оценок с помощью достаточных статистик остается без изменений.

Определения, связанные с экспоненциальным семейством, здесь также выглядят аналогично. Пусть $\theta = (\theta_1, \ldots, \theta_k)$ — k-мерный параметр. Будем говорить, что *выборка X соответствует экспоненциальному семейству*, если ее функция правдоподобия (относительно некоторой меры) представима в виде

$$f_\theta(X) = \exp\{(a(\theta), S(X)) + V(\theta)\} h(X), \quad (2)$$

где $a(\theta) = (a_1(\theta), \ldots, a_k(\theta))$, $S(\boldsymbol{x}) = (S_1(\boldsymbol{x}), \ldots, S_k(\boldsymbol{x}))$, под $\boldsymbol{x}$ понимается вектор $(x_1, \ldots, x_n)$, функции $a_0(\theta) \equiv 1$, $a_1(\theta), \ldots, a_k(\theta)$ линейно независимы.

В примере 64.1 $\theta = (\alpha, \sigma^2)$, функция правдоподобия определена в (66.1), так что в этом случае

$$h(X) = \prod (2\pi u_i)^{-1/2}, \quad a_1(\theta) = -\frac{1}{2}\sigma^{-2}, \quad a_2(\theta) = \alpha\sigma^{-2},$$

$$V(\theta) = \alpha^2 \sigma^{-2} \sum \frac{t_i^2}{u_i}, \quad S_1(x) = \sum \frac{\mathrm{x}_i^2}{u_i}, \quad S_2(x) = \sum \frac{\mathrm{x}_i t_i}{u_i}. \quad (3)$$

Для распределения Пуассона $\theta = \lambda$,

$$f_\lambda(X) = \prod e^{-\lambda t_i}\frac{(\lambda t_i)^{\mathrm{x}_i}}{\mathrm{x}_i!}, \quad h(X) = \prod \frac{t_i^{\mathrm{x}_i}}{\mathrm{x}_i!},$$
$$a_1(\lambda) = \ln\lambda, \quad V(\lambda) = -\lambda\sum t_i, \quad S_1(X) = \sum \mathrm{x}_i. \tag{4}$$

В примере 64.7 $\theta = (\alpha, \gamma)$, функция правдоподобия определена в (1). Здесь в представлении (2)

$$h(X) = 1, \quad V(\theta) = [n\ln(1 - p(\alpha,\gamma,N))], \quad a_1(\theta) = \ln\frac{\gamma}{(1 - Q(\alpha,\gamma,N))},$$
$$a_2(\theta) = -\alpha - 1, \quad S_1(X) = \nu = \sum I(\mathrm{x}_i > N),$$
$$S_2(x) = \sum_{i:\mathrm{x}_i > N} \ln \mathrm{x}_i = \sum \ln \mathrm{x}_i I(\mathrm{x}_i > N).$$

Так как результаты § 25 с объемом выборки никак не связаны, то они для разнораспределенных наблюдений полностью сохраняются. В частности, справедливы следующие утверждения.

Теорема 1. 1. *Если* $a_0(\theta) \equiv 1$, $a_1(\theta), \ldots, a_k(\theta)$ *линейно независимы, то статистика* S *в представлении* (2) *является минимальной достаточной статистикой.*

2. *Если, кроме того,* $a(\theta)$ *зачеркивает* k*-мерный параллелепипед в* $\mathbb{R}^k$*, когда* θ *пробегает* Θ*, то* S *— полная достаточная статистика.*

Отсюда вытекает

Следствие 1. *Если* X *соответствует экспоненциальному семейству, выполнены условия п.* 2 *теоремы* 1 *и* $\theta^* \in K_b$*, то оценка* $\theta^*_S = \mathbf{E}(\theta^* \mid S)$ *является эффективной в* K_b*.*

Если выполнены условия п. 2 *теоремы* 1*, то о. м. п.* $\widehat{\theta}^*$ *является эффективной в классах с соответствующим им смещением.*

Последнее утверждение вытекает из того, что о. м. п. измеримы относительно S.

Вернемся к примерам § 64. В примере 64.1 несмещенная оценка

$$\alpha^* = \left(\sum \frac{\mathrm{x}_i t_i}{u_i}\right)\left(\sum \frac{t_i^2}{u_i}\right)^{-1}$$

в силу следствия 1 является эффективной. Напомним, что такую же оценку мы получили в п. 65.2 путем варьирования коэффициентов v_i в оценках по методу моментов. Найдем теперь несмещенную оценку для σ^2. Для этого вычислим смещение оценки

$$(\sigma^2)^* = \frac{1}{n}\sum \frac{(\mathrm{x}_i - \alpha^* t_i)^2}{u_i} = \frac{1}{n}\sum \frac{\left(\mathrm{x}_i^2 - (\alpha^*)^2 t_i^2\right)}{u_i} =$$
$$= \frac{1}{n}\left[\sum \frac{(\mathrm{x}_i - \alpha t_i)^2}{u_i} - (\alpha^* - \alpha)^2 b_1(n)\right], \quad b_1(n) = \sum \frac{t_i^2}{u_i}$$

(см. (65.6)). Так как $\mathbf{E}(\alpha^* - \alpha)^2 b_1(n) = \sigma^2$ (см. (65.5)), то мы получаем

$$\mathbf{E}(\sigma^2)^* = \frac{1}{n}\left[\sum \sigma^2 - \sigma^2\right] = \frac{n-1}{n}\sigma^2.$$

Отсюда, из (3) и следствия 1 вытекает, что несмещенная оценка

$$(\sigma^2)_1^* = \frac{n}{n-1}(\sigma^2)^* = \frac{1}{n-1}\sum \frac{(\mathrm{x}_i - \alpha^* t_i)^2}{u_i} \tag{5}$$

является эффективной для оценки σ^2. Таким образом, относительно смещения и эффективности оценки $(\sigma^2)_1^*$ здесь сохраняется полная аналогия с однородным случаем (ср. с § 17). Оценка (5) отличается постоянным множителем $\dfrac{n}{n-1}$ от оценки, найденной в п. 65.2.

Для пуассоновских наблюдений оценка $\lambda^* = \left(\sum \mathrm{x}_i\right)\left(\sum t_i\right)^{-1}$ будет, очевидно, также эффективной.

Отметим, что оценки α^*, $(\sigma^2)^*$, λ^* в этих примерах являются также о. м. п.

Примерам 64.6, 64.7 мы посвятим специальные параграфы, которые служат иллюстрацией того, как решать нестандартные задачи оценки параметров и проверки гипотез, как иметь дело, например, с бесконечномерными мешающими параметрами, как изучать сложные статистики и использовать схему серий.

§ 68. Эффективные оценки в задаче оценивания «хвостов» распределений (пример 64.7). Асимптотические свойства оценок

В этом параграфе мы рассмотрим применение результатов предыдущих разделов к «идеализированной» задаче об оценке параметра $\theta = (\alpha, \gamma)$ в примере 64.7 с функцией правдоподобия (67.1) (задача В). Будут найдены о. м. п., которые, будучи измеримыми относительно достаточных статистик от выборки экспоненциального типа, будут эффективными оценками в классах с соответствующими смещениями. Далее будут найдены асимптотические свойства этих оценок — доказана их асимптотическая нормальность (это можно установить и с помощью общих результатов § 66; для одномерных неизвестных параметров α или γ это следует из теоремы 66.1). Будет выяснено, при каких условиях мы можем пользоваться оценками, найденными в задаче В, для оценивания параметров в исходной задаче А с плотностями наблюдений x_i (64.5) при $\mathrm{x}_i > N$. При этом рассматриваются два подхода. Первый основан на изучении условий, при которых задачи А и В оказываются асимптотически эквивалентными. Второй использует построение наблюдений и нужных нам статистик в задачах А и В на одном вероятностном пространстве. Условия, при которых нужные нам статистики становятся асимптотически эквивалентными, оказываются более широкими, чем при первом подходе.

1. Оценки максимального правдоподобия. Функция правдоподобия выборки X в задаче В имеет вид (см. (67.1))

$$f_\theta(X) = (1-p)^{n-\nu}\gamma^\nu \Pi^{-\alpha-1},$$

где

$$\nu = \sum I(\mathrm{x}_i > N), \quad \ln \Pi = \sum I(\mathrm{x}_i > N) \ln \mathrm{x}_i, \quad p = \frac{\gamma N^{-\alpha}}{\alpha}.$$

Логарифм $L = L(X, \theta)$ этой функции правдоподобия примет вид

$$L = (n - \nu) \ln (1 - p) + \nu \ln \gamma - (\alpha + 1) \ln \Pi. \tag{1}$$

Для отыскания о. м. п. нам понадобятся частные производные выражений, присутствующих в (1). При очевидных соглашениях относительно обозначений имеем

$$p'_\alpha = -p \left(\ln N + \frac{1}{\alpha} \right), \quad p'_\gamma = \frac{p}{\gamma},$$

$$L'_\alpha = (n - \nu) \frac{p(\ln N + 1/\alpha)}{1 - p} - \ln \Pi, \quad L'_\gamma = -(n - \nu) \frac{p}{\gamma(1 - p)} + \frac{\nu}{\gamma}.$$

Уравнение $L'_\gamma = 0$ дает

$$p = \frac{\nu}{n}, \quad \gamma = \frac{\nu \alpha N^\alpha}{n}. \tag{2}$$

Из уравнения $L'_\alpha = 0$ получаем равенство $p(n - \nu) \ln N + \dfrac{(n - \nu)p}{\alpha} = (1 - p) \ln \Pi$, которое при подстановке $p = \nu/n$ дает

$$\widehat{\alpha}^* = \frac{\nu}{\ln \Pi - \nu \ln N} = \frac{\sum \xi_i}{\sum \eta_i}, \tag{3}$$

где

$$\xi_i = I(\mathrm{x}_i > N), \quad \eta_i = I(\mathrm{x}_i > N) \ln \frac{\mathrm{x}_i}{N}.$$

Отметим, что оценка $\widehat{\alpha}^*$ определена лишь на множестве $\{\nu > 0\}$. Если $\nu = 0$, то $\widehat{\alpha}^*$ можно положить равной любому числу из множества возможных значений. В силу (2) имеем также

$$\widehat{\gamma}^* = \frac{\nu \widehat{\alpha}^*}{n} N^{\widehat{\alpha}^*}. \tag{4}$$

По следствию 67.1 оценка $\widehat{\theta}^* = (\widehat{\alpha}^*, \widehat{\gamma}^*)$ эффективна в классах оценок с соответствующим смещением. Если γ известно, то о. м. п. $\widehat{\alpha}^*$ будет решением весьма сложного уравнения. Проще пользоваться оценкой (3), которая, как мы увидим ниже, будет асимптотически эффективной. Если α известно, то о. м. п. для γ имеет вид $\widehat{\gamma}^* = \alpha \nu N^\alpha / n$. Эта оценка в соответствии с теоремой 67.1 будет а. н. и а. э. в классе M-оценок. Мы будем рассматривать лишь случай двух неизвестных параметров.

2. Асимптотическая нормальность $\widehat{\theta}^*$ в задаче B. Начнем с оценки $\widehat{\alpha}^*$. Для представления (3) имеем

$$\begin{gathered}\mathbf{E}\xi_i = p, \quad \mathbf{E}\xi_i^2 = p, \quad \mathbf{D}\xi_i = p(1-p),\\ \mathbf{E}\eta_i = \int\limits_N^\infty \gamma x^{-\alpha-1} \ln\frac{x}{N}\, dx = \frac{p}{\alpha},\\ \mathbf{E}\eta_i^2 = \int \gamma x^{-\alpha-1}\left(\ln\frac{x}{N}\right)^2 dx = \frac{2p}{\alpha^2},\\ \mathbf{D}\eta_i = \frac{2p}{\alpha^2} - \frac{p^2}{\alpha^2} = \frac{p(2-p)}{\alpha^2}\end{gathered} \tag{5}$$

(для вычисления последних двух интегралов следует воспользоваться интегрированием по частям). Поэтому

$$\mathbf{E}(\xi_i - \alpha\eta_i) = 0, \quad \mathbf{E}(\xi_i - \alpha\eta_i)^2 = \mathbf{E}\xi_i^2 - 2\alpha\mathbf{E}\eta_i + \alpha^2\mathbf{E}\eta_i^2 = p.$$

Мы будем рассматривать схему серий, когда $N = N(n) \to \infty$ так, что $np \to \infty$. Воспользуемся представлениями (см. (5))

$$\begin{gathered}\nu = np\left(1 + \sqrt{\frac{1-p}{np}}\zeta_1\right),\\ \widehat{\alpha}^* - \alpha = \frac{\sum(\xi_i - \alpha\eta_i)}{\sum\eta_i} = \frac{\sqrt{np}\,\zeta_3}{\dfrac{np}{\alpha}\left(1 + \zeta_2\sqrt{\dfrac{2-p}{np}}\right)},\end{gathered} \tag{6}$$

где

$$\begin{aligned}\zeta_1 &= \frac{1}{\sqrt{np-(1-p)}}\sum(\xi_i - p) \Rightarrow \mathbf{\Phi}_{0,1},\\ \zeta_2 &= \frac{\alpha}{\sqrt{np(2-p)}}\sum\left(\eta_i - \frac{p}{\alpha}\right) \Rightarrow \mathbf{\Phi}_{0,1},\\ \zeta_3 &= \frac{1}{\sqrt{np}}\sum(\xi_i - \alpha\eta_i) \Rightarrow \mathbf{\Phi}_{0,1}.\end{aligned} \tag{7}$$

Сходимость к нормальному распределению здесь имеет место, так как условие Линдеберга очевидным образом выполнено.

Отсюда и из теоремы следует, что

$$(\widehat{\alpha}^* - \alpha)\frac{\sqrt{n}\,p}{\alpha} = \frac{\zeta_3}{(1 + \zeta_2\sqrt{(2-p)/np})} = \zeta_3\left(1 + O\left(\frac{|\zeta_2|}{\sqrt{np}}\right)\right) \tag{8}$$

и, стало быть, $\widehat{\alpha}^*$ а. н. с разбросом α^2/np. Далее,

$$\widehat{\alpha}^* = \alpha\left[1 + \frac{\zeta_3}{\sqrt{np}}\left(1 + O\left(\frac{|\zeta_2|}{\sqrt{np}}\right)\right)\right], \quad \widehat{\gamma}^* = \frac{\nu}{n}\widehat{\alpha}^* N^{\widehat{\alpha}^*},$$

$$\ln\widehat{\gamma}^* = \ln\nu - \ln n + \ln\widehat{\alpha}^* + \widehat{\alpha}^*\ln N =$$

$$= \ln np + \ln\left(1+\sqrt{\frac{1-p}{n-p}}\zeta_1\right) - \ln n + \ln\alpha +$$

$$+\ln\left(1+\frac{1}{\sqrt{np}}\zeta_3\left(1+O\left(\frac{|\zeta_2|}{\sqrt{np}}\right)\right)\right)+\alpha\ln N+\frac{\alpha\ln N}{\sqrt{np}}\zeta_3\left(1+O\left(\frac{|\zeta_2|}{\sqrt{np}}\right)\right) =$$

$$= \ln\gamma + \zeta_1\sqrt{\frac{1-p}{np}} + \frac{\zeta_3}{\sqrt{np}} + \frac{\alpha\zeta_3\ln N}{\sqrt{np}} + O\left(\frac{\zeta\ln N}{np}\right), \quad (9)$$

где ζ есть случайная величина (не всегда одна и та же в последующих формулах), распределение которой слабо сходится к некоторому собственному распределению. В итоге в силу (7) получаем

$$\ln\widehat{\gamma}^* = \ln\gamma + \frac{1}{np}\sum(\xi_i(1+A) - \alpha A\eta_i - p) + O\left(\frac{\zeta\ln N}{np}\right), \quad (10)$$

где $A = 1+\alpha\ln N$, слагаемые под знаком $\sum$ независимы, одинаково распределены и имеют нулевое среднее. Кроме того, непосредственный подсчет с помощью равенств $\xi_i^2 = \xi_i$, $\eta_i\xi_i = \eta_i$ и (5) дает

$$\mathbf{E}(\xi_i(1+A) - \alpha\eta_i A - p)^2 = p(A^2+1-p).$$

Поэтому

$$\widehat{\gamma}^* = \gamma\left(1+\sqrt{\frac{A^2+1-p}{np}}\zeta_4 + O\left(\frac{\zeta\ln N}{np}\right)\right), \quad (11)$$

$$\zeta_4 = [np(A^2+1-p)]^{-1/2}\sum(\xi_i(1+A) - \alpha A\eta_i - p) \Rightarrow \mathbf{\Phi}_{0,1}.$$

Это означает, что $\widehat{\gamma}^*$ также является а. н. оценкой с разбросом $A^2(A^2+1-p)/np$.

Покажем, что и $\widehat{\theta}^* = (\widehat{\alpha}^*, \widehat{\gamma}^*)$ является а. н. оценкой. Предположим на время, что N, а вместе с ним и A — фиксированные числа. Введем в рассмотрение вектор $(\widetilde{\alpha}, \widetilde{\gamma})'$, где $\widetilde{\alpha}$, $\widetilde{\gamma}$ — главные части оценок $\widehat{\alpha}^*$, $\widehat{\gamma}^*$ в представлениях (8), (11), являющиеся линейными функциями от ζ_3, ζ_4. Но совместное распределение ζ_3, ζ_4 сходится, очевидно, к нормальному с коэффициентом корреляции

$$\mathbf{E}\zeta_3\zeta_4 = (p\sqrt{A^2+1-p})^{-1/2}\mathbf{E}(\xi_i - \alpha\eta_i)(\xi_i(A+1) - \alpha A\eta_i - p) =$$

$$= \frac{A}{\sqrt{A^2+1-p}}$$

(он стремится к единице при $A\to\infty$). Отсюда следует, что $\mathbf{E}(\widetilde{\alpha}-\alpha)(\widetilde{\gamma}-\gamma) = \alpha\gamma A/np$ и, стало быть, матрица вторых моментов $((\widetilde{\alpha}-\alpha),(\widetilde{\gamma}-\gamma))$ имеет вид

$$\sigma^2(\theta, n) = \frac{1}{np}\begin{pmatrix}\alpha^2 & \alpha\gamma A\\ \alpha\gamma A & A^2(A^2+1-p)\end{pmatrix} \equiv \frac{M}{np}. \quad (12)$$

(Если показать, что $\widehat{\alpha}^* \in K_{\Phi,2}$, $\widehat{\gamma}^* \in K_{\Phi,2}$, то это будет также главная часть матрицы вторых моментов $(\widehat{\alpha}^*-\alpha)(\widehat{\gamma}^*-\gamma)$.) Эта матрица имеет определитель

$$\operatorname{Det}\sigma^2(\theta,n) = (np)^{-2}\alpha^2\gamma^2(1-p) > 0$$

и обратную матрицу

$$\sigma^{-2}(\theta,n) = \frac{np}{1-p}\begin{pmatrix} \dfrac{A^2+1-p}{\alpha^2} & -\dfrac{A}{\alpha\gamma} \\ -\dfrac{A}{\alpha\gamma} & \dfrac{1}{\gamma^2} \end{pmatrix}.$$

Это означает, что матрица вторых моментов вектора $((\widetilde{\alpha}-\alpha),(\widetilde{\gamma}-\gamma))\sigma^{-1}(\theta,n)$ равна единичной матрице E, $((\widetilde{\alpha}-\alpha),(\widetilde{\gamma}-\gamma))\sigma^{-1}(\theta,n) \Subset \mathbf{\Phi}_{0,E}$. Эти соотношения сохранятся, очевидно, и в случае $N \to \infty$, а вместе с ними и сходимость

$$((\widehat{\alpha}^*-\alpha),(\widehat{\gamma}^*-\gamma))\sigma^{-1}(\theta,n) \Subset \mathbf{\Phi}_{0,E}. \tag{13}$$

Как уже отмечалось, асимптотическая нормальность $(\widehat{\alpha}^*,\widehat{\gamma}^*)$ может быть установлена также с помощью общих теорем об асимптотической нормальности о. м. п. (ср. с п. 70.4 или с § 36, модифицированным на случай схемы серий). В соответствии с этими результатами оценка $(\widehat{\alpha},\widehat{\gamma}^*)$ будет а. н. с разбросом $(nI(\theta))^{-1}$, где $I(\theta)$ — информационная матрица Фишера; при этом сходимость

$$(\widehat{\alpha}^*-\alpha,\widehat{\gamma}^*-\gamma)(nI(\theta))^{1/2} \Subset \mathbf{\Phi}_{0,E} \tag{14}$$

имеет место вместе с моментами любого порядка. Матрицу $I(\theta)$ можно подсчитать непосредственно исходя из определения, при этом в соответствии с (13) окажется, что $nI(\theta) = \sigma^{-2}(\theta,n)$. Действительно, в нашем случае функция $l = l(x,\theta)$ $(\theta = (\alpha,\gamma))$ равна

$$l = \begin{cases} \ln\gamma - (\alpha+1)\ln x & \text{при} \quad x > N, \\ \ln(1-p) & \text{при} \quad x = N,\ p = \dfrac{\gamma N^{-\alpha}}{\alpha}, \end{cases}$$

так что

$$l'_\alpha = \begin{cases} -\ln x, & x > N, \\ \dfrac{pA}{(1-p)\alpha}, & x = N, \end{cases} \qquad l'_\gamma = \begin{cases} \dfrac{1}{\gamma}, & x > 0, \\ -\dfrac{p}{(1-p)\gamma}, & x = N. \end{cases}$$

Отсюда находим элементы матрицы $I(\theta)$:

$$\mathbf{E}_\theta\big(l'_\alpha\big)^2 = -\mathbf{E}_\theta l''_{\alpha\alpha} = \frac{p}{1-p}\left(\frac{A^2+1-p)}{\alpha^2}\right),$$

$$\mathbf{E}_\theta\big(l'_\gamma\big)^2 = -\mathbf{E}_\theta l''_{\gamma\gamma} = \frac{p}{(1-p)\gamma^2}, \quad \mathbf{E}_\theta l'_\alpha l'_\gamma = -\mathbf{E}l''_{\alpha\gamma} = -\frac{p}{1-p}\cdot\frac{A}{\alpha\gamma}.$$

Дальнейшей нашей целью является установление асимптотической нормальности $\widehat{\theta}^*$ в задаче A и отыскание N, при котором задачи A и B будут в некотором смысле асимптотически эквивалентными.

3*. Асимптотические нормальность и оптимальность в задаче A. Рассмотрим теперь вопрос о том, в каком соотношении находятся «идеализированная» задача B и исходная задача A. Этот вопрос можно рассматривать как проблему робастности (ср. с § 58) в задаче B относительно малых возмущений функции плотности наблюдений (ср. (64.5) и (64.7)) в области $x > N$. Переход от наблюдений y_i к x_i (см. (64.6)) означает отбрасывание «неинформативных» наблюдений $\mathrm{y}_i \leqslant N$, поскольку функция плотности этих наблюдений начинает больше зависеть от неизвестных функций $\gamma_i(\alpha, y)$ и, стало быть, содержит меньше информации о θ. Задача здесь состоит в оптимальном выборе N, который возможен, очевидно, лишь в случае, когда «мешающие» функции $\varepsilon_i(\alpha, y)$ в (64.5) допускают оценку скорости убывания, например

$$|\varepsilon_i(\alpha, y)| \leqslant \varepsilon(y) y^{-\rho}, \quad \rho > 0, \tag{15}$$

где $\varepsilon(y) \to 0$ при $y \to \infty$, или

$$|\varepsilon_i(\alpha, y)| \leqslant \varepsilon(y) e^{-\rho y}, \quad \rho > 0. \tag{16}$$

Мы должны выбрать возможно меньшее N, которое позволяло бы статистические выводы в задаче B переносить на исходную задачу A.

Рассмотрим два подхода к задаче о выборе N.

(a) Первый подход связан с таким выбором N, при котором задачи A и B оказываются в известном смысле асимптотически эквивалентными.

Теорема 1. *Если выполнено* (15) *и* $N = n^{1/(\alpha+\rho)}$, *то любая а. н. оценка* (α^*, γ^*) *для задачи* B *имеет то же самое предельное распределение в задаче* A (*исходной*), *и наоборот. В частности, оценка* $(\widehat{\alpha}^*, \widehat{\gamma}^*)$, *определенная в* (3), (4), *является а. н. в задаче* A *с разбросом* $\sigma^2(\theta, n)$, *определенным в* (12). *Такое же утверждение справедливо при* $N = \dfrac{1}{\rho} \ln n$, *если выполнено* (16).

Если $N = n^{1/(\rho+\alpha)}$, то $np = \dfrac{\gamma}{\alpha} n^{\rho/(\rho+\alpha)}$ и разброс $\sigma^2(\theta, n)$ будет иметь вид (см. (12))

$$\sigma^2(\theta, n) = \frac{\alpha}{\gamma} n^{-\rho/(\rho+\alpha)} M, \tag{17}$$

при этом $A = 1 + \alpha \ln N = 1 + \dfrac{\alpha}{\alpha+\rho} \ln n$. Аналогичным образом выписывается матрица разброса в случае $N = \dfrac{1}{\rho} \ln n$.

Замечание 1. Утверждение теоремы означает, что оценка $\widehat{\theta}^*$, будучи асимптотически оптимальной (в известном смысле, см. § 66) в задаче B, будет таковой и в задаче A. Более точно, пусть класс оценок $K \in K_\Phi$ таков, что принадлежность θ^* классу $K \subset K_\Phi$ в задаче A означает принадлежность

$\theta^* \in K$ в задаче B. Тогда если $\widetilde{\theta}^*$ — а. э. оценка в K в задаче B, то эта оценка будет а. э. в K в задаче A.

Действительно, если $\widetilde{\theta}^*$ — а. э. оценка в K в задаче B, то для любой другой оценки $\theta^* \in K$ выполнено (66.12). Так как матрицы разброса $\widetilde{\sigma}^2$ и σ^2 этих оценок в задаче A сохраняются (с точностью до асимптотической эквивалентности), то $\widetilde{\theta}^*$ будет а. э. в K и в задаче A.

(b) Второй подход связан с таким построением наблюдений x_i в задачах A, B на одном вероятностном пространстве, что имеет место асимптотическая эквивалентность оценок $\widehat{\theta}^*$ в этих задачах. Статистики θ^* и θ^{**} мы называем здесь асимптотически эквивалентными, если они а. н. с разбросом $\sigma^2(\theta, n)$ и $(\theta^* - \theta^{**})\sigma^{-1}(\theta, n) \underset{P}{\longrightarrow} 0$ (или $(\theta^* - \theta^{**})\sigma^{-2}(\theta, n)(\theta^* - \theta^{**})^\top \underset{P}{\longrightarrow} 0$).

Теорема 2. *Пусть выполнено* (15) *и* $N = n^{1/(2\rho+\alpha)}$. *Тогда оценка* $(\widehat{\alpha}^*, \widehat{\gamma}^*)$, *определенная формулами* (3), (4), *является а. н. в задаче* A *с разбросом* $\sigma^2(\theta, n)$, *определенным в* (12). *Такое же утверждение справедливо при* $N = \dfrac{\ln n}{2\rho}$, *если выполнено* (16).

Отметим, что значение N в теореме 2 меньше, чем значение N в теореме 1. При выполнении (15) и $N = n^{1/(2\rho+\alpha)}$ разброс $\sigma^2(\theta, n)$ будет иметь вид

$$\sigma^2(\theta, n) = \frac{\alpha}{\gamma} n^{-2\rho/(2\rho+\alpha)} M.$$

Ясно, что этот разброс меньше, чем (17), и достигается это за счет возможности выбора меньшего N.

Доказательство теоремы 1. Напомним, что функция правдоподобия в задаче A имеет вид (ср. с (64.5))

$$f_\theta^A(X) = \prod_{i:\mathrm{x}_i \leqslant N} (1 - p_i) \prod_{i:\mathrm{x}_i > N} \gamma \mathrm{x}_i^{-\alpha-1}(1 + \varepsilon_i(\alpha, \mathrm{x}_i)),$$
$$p_i = p_i(\theta, N) = \int_N^\infty \gamma(1 + \varepsilon_i(\alpha, y)) y^{-\alpha-1}\, dy. \tag{18}$$

Функция правдоподобия $f_\theta^B(X)$ в задаче B приведена в (67.1). Обозначим через P_θ^A, P_θ^B соответственно распределения выборок в задачах A, B (это суть распределения с плотностями f_θ^A и f_θ^B). Допустим, что мы можем выбрать $N = N(n)$ так, что $f_\theta^A(x)/f_\theta^B(x) \to 1$ равномерно на множестве Ω_n при каждом $\theta \in \Theta$, где

$$\mathbf{P}^B(\Omega_n) \to 1, \quad \mathbf{P}^A(\Omega_n) \to 1$$

при $n \to \infty$. *Тогда все утверждения о предельном распределении статистик* (α^*, γ^*) *в условиях задач* A, B *будут совпадать.* Действительно, в

этом случае

$$P_\theta^A(X \in C) =$$

$$= \int\limits_{C\Omega_n} f_\theta^A(\boldsymbol{x})\mu(d\boldsymbol{x}) + P_\theta^A(C\overline{\Omega}_n) = \int\limits_{C\Omega_n} f_\theta^B(\boldsymbol{x})(1+\Delta(\theta,\boldsymbol{x}))\mu(d\boldsymbol{x}) + o(1) =$$

$$= P_\theta^B(X \in C\Omega_n)(1+r\Delta(\theta)) + o(1) = \mathbf{P}_\theta^B(X \in C) + o(1), \quad (19)$$

где $r \leqslant 1$, $\Delta(\theta) = \sup\limits_{x\in\Omega_n} \Delta(\theta,x) \to 0$ при $n \to \infty$. Это означает, в частности, что асимптотическая нормальность (α^*,γ^*) в задаче B влечет за собой асимптотическую нормальность (α^*,γ^*) с теми же параметрами в задаче A:

$$\mathbf{P}_\theta^A((\alpha^*,\gamma^*)\sigma^{-1}(\theta,n) \in C_2) =$$

$$= \mathbf{P}_\theta^B((\alpha^*,\gamma^*)\sigma^{-1}(\theta,n) \in C_2) + o(1) \to \Phi_{0,E}(C_2), \quad (20)$$

где $C_2 \subset \mathbb{R}^2$. (Вместе с тем (19), (20), вообще говоря, не влекут за собой сходимость моментов (α^*,γ^*).)

Нам осталось выяснить, при каких N возможно (19). Положим

$$\Omega_n = \{\nu < 2pn\}. \quad (21)$$

Согласно результатам п. 1, 2 $\mathbf{P}_\theta^B(\Omega_n) \to 1$ при $n \to \infty$. Аналогично в задаче A имеем $\nu \leqslant \nu_1$, $\nu_1 \in B_{p_1}^n$, $p_1 = p(1+\Delta)$ при достаточно малом $\Delta > 0$. Отсюда при $\Delta < 1$ следует $\mathbf{P}_\theta^A(\Omega_n) \to 1$. (Нетрудно видеть, что на самом деле справедливы экспоненциальные оценки, например $1 - \mathbf{P}_\theta^B(\Omega_n) < \exp(-pn)$ (см. [17, § 5.5]).) Утверждение теоремы 1 будет вытекать из следующего предложения.

Лемма 1. *Если* $N = n^{1/(\alpha+\rho)}$ *и* Ω_n *определены в* (21), *то выполнено* (19). *В случае* (16) *надо положить* $N = \dfrac{1}{\rho}\ln n$.

Доказательство. Сравнивая (18) и (67.1), получаем

$$\ln f_\theta^A(X) - \ln f_\theta^B(X) = \sum_{i:\mathrm{x}_i>N} \ln(1+\varepsilon_i(\alpha,\mathrm{x}_i)) + \sum_{i:\mathrm{x}_i\leqslant N} \ln\frac{1-p_i}{1-p}. \quad (22)$$

Будем считать, не ограничивая общности, что $|\varepsilon_i(\alpha,x)| \downarrow 0$. Тогда при выполнении (15) первая сумма в (22) на множестве Ω_n не превосходит

$$\sum_{i:\mathrm{x}_i>N} \varepsilon_i(\alpha,\mathrm{x}_i) \leqslant \nu\varepsilon(N)N^{-\rho} \leqslant 2\rho n\varepsilon(N)N^{-\rho}. \quad (23)$$

Во второй сумме при $p \to 0$, $p_i \to 0$

$$p_i = p + \gamma \int_N^\infty \varepsilon_i(\alpha, y) y^{-\alpha-1}\, dy,$$

$$\left| \int_N^\infty \varepsilon_i(\alpha, y) y^{-\alpha-1}\, dy \right| \leqslant \int_N^\infty \varepsilon(y) y^{-\alpha-\rho-1}\, dy \leqslant \varepsilon(N) \frac{N^{-\rho-a}}{\rho+\alpha} = \frac{\varepsilon(N)\alpha}{\gamma(\alpha+\rho)} p N^{-\rho},$$

$$\ln \frac{1-p_i}{1-p} = p - p_i + O\bigl(|p^2 - p_i^2|\bigr), \quad \left| \ln \frac{1-p_i}{1-p} \right| \leqslant \frac{\varepsilon(N)\alpha}{\alpha+\rho} p N^{-\rho}(1+o(1)).$$

Поэтому

$$\left| \sum_{i:\mathrm{x}_i \leqslant N} \ln \frac{1-p_i}{1-p} \right| \leqslant \left(\frac{\varepsilon(N)\alpha p}{\alpha+\rho} \right) N^{-\rho}(1+o(1))n \leqslant cpnN^{-\rho}\varepsilon(N). \tag{24}$$

Эта оценка вместе с (23) означает, что разность (22) будет сходиться к нулю на множестве Ω_n, если в качестве N взять значение $N = n^{1/\alpha+\rho}$, при котором $npN^{-\rho} = \alpha/\gamma$ и правые части в (23), (24) сходятся к нулю. Аналогичные выводы можно сделать в случае (16) при $N = \ln n/\rho$. Лемма 1 доказана.

Обратимся теперь ко второму подходу, связанному с построением наблюдений и статистик в задачах A, B на одном вероятностном пространстве.

Доказательство теоремы 2. Обозначим

$$p_B(t) = \min\left(1, \frac{\gamma}{\alpha} t^{-\alpha-1}\right), \quad P_{A,i}(t) = \mathbf{P}_\theta^A(\mathrm{y}_i > t). \tag{25}$$

Очевидно, что $p_B(t) = \mathbf{P}_\theta^B(\mathrm{x}_i > t)$ при $t \geqslant N$,

$$P_{a,i}(t) = p_B(t)(1 + \Delta_i(t)),$$

где

$$p_B(T)|\Delta_i(t)| = \left| \gamma \int_t^\infty \varepsilon_i(\alpha, y)^{-\alpha-1}\, dy \right| \leqslant \frac{\gamma|\varepsilon_i(\alpha,t)|t^{-\alpha}}{\alpha} = p_B(t)|\varepsilon_i(\alpha,t)|.$$

Будем считать для простоты, что функции $P_{A,i}(t)$ непрерывны. Зададим случайные величины x_i в задачах A, B на одном вероятностном пространстве следующим образом. В качестве исходного возьмем вероятностное пространство $\left(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, \mathbf{P}_\theta^A\right)$ в задаче A и рассмотрим наряду с наблюдениями y_i (с плотностью (64.5)) наблюдения $\mathrm{z}_i = p_B^{-1}(P_{A,i}(\mathrm{y}_i))$.

Очевидно, что z_i имеют распределение p_B и, следовательно, значения $\mathrm{x}_i^A = \max(N, \mathrm{z}_i)$ можно рассматривать как наблюдения x_i в задаче B.

Рассмотрим теперь статистики, на которых строятся основные части оценок $\widehat{\alpha}^*$, $\widehat{\gamma}^*\widehat{\alpha}^*$, и выясним, чем они отличаются в задачах A и B. Начнем со статистики

$$S(Y) = \sqrt{np}\,\zeta_3 = \sum(\xi_i - \alpha\eta_i), \quad \xi_i = I(\mathrm{y}_i > N), \quad \eta_i = I(\mathrm{y}_i > N)\ln\frac{\mathrm{y}_i}{N}.$$

В задаче B аналогичная статистика на том же вероятностном пространстве имеет вид

$$S_B(Y) = \sum\left(\xi_i^B - \alpha\eta_i^B\right), \quad \xi_i^B = I(\mathrm{z}_i > N), \quad \eta_i^B = I(\mathrm{z}_i > N)\ln\frac{\mathrm{z}_i}{N},$$

так что

$$S(Y) - S_B(Y) = \sum\left(\xi_i - \xi_i^B\right) - \alpha\sum\left(\eta_i - \eta_i^B\right); \tag{26}$$

$$\sum\left(\xi_i - \xi_i^B\right) = \sum\left(I(\mathrm{y}_i > N) - I(\mathrm{z}_i > N)\right) = \\ = \sum I(\mathrm{y}_i > N,\, \mathrm{z}_i \leqslant N) - \sum I(\mathrm{y}_i \leqslant N,\, \mathrm{z}_i > n). \tag{27}$$

Рассмотрим сначала первую сумму в правой части (27) и предположим для определенности, что $\Delta_i(N) > 0$. Событие $\{\mathrm{y}_i > N,\, \mathrm{z}_i \leqslant N\}$ возможно, если $\mathrm{y}_i = N + \Delta$, $\Delta > 0$, $\mathrm{z}_i = p_B^{-1}(P_{A,i}(\mathrm{y}_i)) \leqslant N$ или, что то же (см. (24)), $p_B^{-1}(u) = (\alpha u/\gamma)^{-1/\alpha}$,

$$\mathrm{z}_i = \left(\frac{\gamma}{\alpha}\right)^{1/\alpha}\left(\frac{\gamma}{\alpha}\mathrm{y}_i^{-\alpha}(1 + \Delta_i(\mathrm{y}_i))\right)^{-1/\alpha} = \\ = \mathrm{y}_i(1 + \Delta_i(\mathrm{y}_i))^{-1/\alpha} = (N + \Delta)(1 + \Delta_i(N + \Delta))^{-1/\alpha} \leqslant N. \tag{28}$$

Это, в свою очередь, возможно, если ($|\Delta_i(t)|$ и $|\varepsilon_i(\alpha, t)|$ монотонно убывают)

$$N\Delta_i(N) \geqslant \alpha\Delta(1 + o(1)),$$

$$\Delta \leqslant \frac{N\Delta_i(N)}{\alpha}(1 + o(1)) \leqslant \frac{N\varepsilon_i(\alpha, N)}{\alpha}(1 + o(1)).$$

Изложенное означает, что первая сумма в (27) не превосходит

$$\sum I\left(\mathrm{y}_i \in \left(N, N\left(1 + \frac{\varepsilon_i(\alpha, N)}{\alpha}(1 + o(1))\right)\right)\right). \tag{29}$$

Среднее значение этой суммы при достаточно больших N не превосходит

$$2nf_\theta(N)\frac{N\varepsilon_i(\alpha, N)}{\alpha} \leqslant 3n\gamma N^{-\alpha}\frac{\varepsilon_i(\alpha, N)}{\alpha} \leqslant \frac{3\gamma n}{\alpha}\varepsilon(n)N^{-\rho-\alpha}. \tag{30}$$

Аналогично происходит оценка второй суммы в (27) в случае $\Delta < 0$.

Оценим теперь вторую сумму в (26):

$$\sum\left(\eta_i - \eta_i^B\right) = \sum\eta_i\left(1 - \xi_i^B\right) - \sum\eta_i^B(1 - \xi_i) + \sum\left(\eta_i - \eta_i^B\right)\xi_i\xi_i^B. \tag{31}$$

Здесь первые две суммы оцениваются так же, как суммы в (27), с той лишь разницей, что оценку (29) надо еще домножить на малый множитель, не превосходящий $\ln\dfrac{N+\Delta}{N} \leqslant \dfrac{\Delta}{N}$.

Последняя сумма в (31) согласно (28) по абсолютной величине не превосходит

$$\sum\left|\ln\frac{\mathrm{y}_i}{\mathrm{z}_i}\right| = \frac{1}{\alpha}\sum|\ln(1+\Delta_i(\mathrm{y}_i))| \leqslant \frac{1}{\alpha}\sum|\Delta_i(\mathrm{y}_i)| \leqslant$$

$$\leqslant \frac{1}{\alpha}\sum|\varepsilon_i(\alpha,\mathrm{y}_i)| \leqslant \frac{1}{\alpha}\sum\varepsilon(\mathrm{y}_i)\mathrm{y}_i^{-\rho}.$$

Среднее значение этой суммы не превосходит

$$\frac{\gamma}{\alpha}\sum\int_N^\infty \varepsilon(y)y^{-\rho}y^{-\alpha-1}(1+\varepsilon_i(\alpha,y))\,dy \leqslant \frac{2\gamma n\varepsilon(N)N^{-\rho-\alpha}}{\alpha(\alpha+\rho)}.$$

Мы получили здесь ту же оценку, что и в (30). Это означает в силу неравенства Чебышева, что с вероятностью, сходящейся к единице,

$$|S(Y)-S_B(Y)| \leqslant cn\varepsilon^{1/2}(N)N^{-\rho-\alpha},$$

$$|\widehat{\alpha}^*-\widehat{\alpha}_B^*| = \frac{\alpha|S(Y)-S_B(Y)|}{np}\left(1+O\left(\frac{\zeta}{\sqrt{np}}\right)\right) \leqslant c\varepsilon^{1/2}(N)N^{-\rho} \tag{32}$$

(см. (6)); постоянные c здесь различны; $\widehat{\alpha}_B^*$ есть о. м. п. для α в задаче B, построенная на наблюдениях y_i.

Если выбором N мы сделаем квадрат правой части неравенства (32) меньше разброса оценки α_B^*, равного $(np)^{-1}$, то оценки $\widehat{\alpha}^*$ и $\widehat{\alpha}_B^*$ будут асимптотически эквивалентными (иметь то же самое предельное распределение). Требуемое неравенство будет выполнено, если мы положим $N^{-2\rho} = \dfrac{\gamma}{\alpha}(np)^{-1}$ или, что то же, $N = n^{1/(2\rho+\alpha)}$.

Сравнение в задачах A и B статистик вида $\sum(\xi_i(1+A)-\alpha A\eta_i-p)$, определяющих оценку $\widehat{\gamma}^*$, проводится аналогично, как и анализ оценок $\widehat{\theta}^*$ при выполнении (16). Остальные члены для $(\widehat{\alpha}^*-\alpha)$ и $(\widehat{\gamma}^*-\gamma)$ в представлениях (8), (9), (11) имеют в задаче A то же асимптотическое поведение, что и в задаче B (надо использовать центральную предельную теорему в приложении IV). Теорема 2 доказана.

§ 69. Неравенство Рао–Крамера

Вернемся к изучению свойств оценок в общем случае. Пусть, как и прежде, $X \in \mathbf{P}_\theta = \{\mathbf{P}_{\theta,1},\ldots,\mathbf{P}_{\theta_n}\}$, $f_{\theta,i}(x)$ — плотность распределения $\mathbf{P}_{\theta,i}$ относительно меры μ_i, K_b — класс оценок со смещением $b(\theta)$ (см. § 66). Мы будем предполагать, что выполнено условие

(R) *Функции* $\sqrt{f_{\theta,i}(x)}$ *для п. в.* $[\mu_i]$ *значений* x *непрерывно дифференцируемы по* θ. *Информации Фишера* $I_i(\theta) = \mathbf{E}_\theta\big[l_i'(\mathrm{x}_i,\theta)\big]^2$ *существуют, положительны и непрерывны.*

Если Θ — компакт (выполнено (A_c)), то непрерывность $I_i(\theta)$ в условии (R) эквивалентна условию

$$\sup_\Theta \mathbf{E}_\theta\big(\big[l_i'(\mathrm{x}_i,\theta)\big]^2\big|l_i'(\mathrm{x}_i,\theta)\big| > N\big) \to 0$$

при $N \to \infty$ (см. приложение V, теорема 3).

Неравенство Рао–Крамера для неоднородных наблюдений естественным образом обобщает неравенства для однородного случая, полученные в § 26.

Теорема 1. *Если* $\theta^* \in K_b$, *выполнено условие* (R) *и* $\mathbf{E}_\theta(\theta^*)^2 < c < \infty$, *то*

$$\mathbf{D}_\theta\theta^* \geqslant \frac{1+b'(t)}{I(\theta,n)}, \quad I(\theta,n) = \sum_{i=1}^n I_i(\theta). \tag{1}$$

Если в этом неравенстве достигается равенство на некотором отрезке $\theta \in [\theta_1,\theta_2] \subset \Theta$ *и* $\mathbf{D}_\theta\theta^* > 0$ *на этом отрезке, то функция правдоподобия* $f_\theta(X)$ *при* $\theta \in [\theta_1,\theta_2]$ *представима в виде*

$$f_\theta(X) = \exp\{\theta^* A(\theta) + B(\theta)\}h(X), \tag{2}$$

где $A(\theta)$, $B(\theta)$ *от* X *не зависят.*

Обратно, если $\theta^* = \mathrm{const}$ *или если справедливо представление* (2), *то в неравенстве* (1) *достигается равенство.*

Доказательство теоремы 1 почти полностью повторяет рассуждения в доказательстве теоремы 26.1 с той лишь разницей, что применительно к новым условиям в (26.9) надо подставить

$$\mathbf{E}_\theta(L')^2 = \mathbf{E}_\theta\left(\sum_{i=1}^n l_i'(\mathrm{x}_i,\theta)\right)^2 = \sum_{i=1}^n \mathbf{E}_\theta\big(l_i'(\mathrm{x}_i,\theta)\big)^2 = I(\theta,n).$$

Здесь полностью сохраняет свою силу следствие 26.2 о том, что *для достижения нижней границы в неравенстве Рао–Крамера необходимо и достаточно, чтобы оценка* θ^* *была достаточной статистикой, а функция* $\psi(\theta^*,\theta)$ *в факторизационном равенстве имела вид*

$$\psi(\theta^*,\theta) = \exp\{\theta^* A(\theta) + B(\theta)\}.$$

Сохраняется также следствие 26.3 о том, что *если* $\theta^* \in K_b$ *и в неравенстве Рао–Крамера достигается равенство, то* θ^* *— эффективная в* K_b *оценка.*

Сохраняют свою силу и другие утверждения и определения § 26 (с заменой $I(\theta)$ на $I(\theta,n)$; многие из них с объемом выборки не связаны).

В *многомерном случае* изменения аналогичны. Под (R) следует понимать условие

(R) *Функции* $\sqrt{f_{\theta,i}(x)}$, $i = 1, \ldots, n$, *непрерывно дифференцируемы по* θ_j, $j = 1, \ldots, k$, *для п. в.* $[\mu_i]$ *значений* x. *Матрицы Фишера* $I_i(\theta) = \left\|I_i^{rs}(\theta)\right\|$,

$$I_i^{rs}(\theta) = \mathbf{E}\frac{\partial l_i(\mathrm{x}_i,\theta)}{\partial\theta_r}\frac{\partial l_i(\mathrm{x}_i,\theta)}{\partial\theta_s}$$

непрерывны по θ, *определитель* $|I(\theta,n)|$, $I(\theta,n) = \sum_{i=1}^{n} I_i(\theta)$, *отличен от нуля.*

В неравенстве (16.14) теоремы 26.1А вместо $(nI(\theta))^{-1}$ следует поставить матрицу $I(\theta,n)^{-1}$.

Интегральное неравенство Рао–Крамера также сохраняется при естественной модификации формулировок и доказательств. Утверждение теоремы 30.1 для скалярного параметра θ в условиях этого параграфа будет иметь вид

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{\left[\mathbf{E}\dfrac{h(\theta)}{q(\theta)}\right]^2}{\mathbf{E}I(\theta,n)\left(\dfrac{h(\theta)}{q(\theta)}\right)^2 + \mathbf{E}\left(\dfrac{h'(\theta)}{q(\theta)}\right)^2}.$$

При $h(\theta) = q(\theta)/I(\theta,n)$ в предположении, что эта функция является финитной и дифференцируемой, получаем

$$\mathbf{E}(\theta^* - \theta)^2 \geqslant \frac{J_n^2}{J_n + \mathbf{E}[(q(\theta)I^{-1}(\theta,n))'q^{-1}(\theta)]^2}, \quad J_n = \mathbf{E}I^{-1}(\theta,n).$$

Это значит, что для гладких $q(\theta)$ и $I(\theta,n)$ средний разброс $\mathbf{E}(\theta^* - \theta)^2$ (усреднение $\mathbf{E}$ берется по совместному распределению θ и выборки X с плотностью $q(t)f_t(\boldsymbol{x})$) имеет тот же порядок малости, что и J_n. Если $I_i(t)$ при разных i соизмеримы и $I(t,n) \to \infty$, то J_n будет сходиться к нулю.

Важным следствием интегральных неравенств Рао–Крамера является возможность устанавливать тот факт, что а. R-э. оценки являются асимптотически байесовскими (см. § 30). Это означает, что «в среднем» никакие другие оценки не дают выигрыша по сравнению с а. R-э. оценками.

§ 70. Неравенства для отношения правдоподобия и асимптотические свойства о. м. п.

Этот параграф является аналогом § 33, 34 и содержит результаты о свойствах асимптотической оптимальности о. м. п.

1. Неравенства для отношения правдоподобия и состоятельность о. м. п. Как и прежде, отношением правдоподобия мы называем функцию (предполагаем выполненным условие (A_μ))

$$Z(u) = \frac{f_{\theta+u}(X)}{f_\theta(X)} = \exp\{L(X,\theta+u) - L(X,\theta)\}, \quad X \in \mathbf{P}_\theta.$$

Рассмотрим сначала одномерный случай (многомерный случай будет рассмотрен в п. 4). Обозначим через $r_i(u)$ расстояние Хеллингера между $\mathbf{P}_{\theta+u,i}$ и $\mathbf{P}_{\theta,i}$:

$$r_i(u) = \rho(\mathbf{P}_{\theta+u,i}, \mathbf{P}_{\theta,i}) = \int \left(\sqrt{f_{\theta+u,i}(x)} - \sqrt{f_{\theta,i}(x)}\right)^2 \mu_i(dx).$$

В этом и последующих разделах будем предполагать, что выполнены условия (A_0) $(f_{t,i}(x) \neq f_{\theta,i}(x)$ при $t \neq \theta)$ и (A_c) (множество Θ — компакт). Кроме того, будем предполагать, что *функции* $\sqrt{f_{\theta,i}(x)}$ *дифференцируемы для п. в.* $[\mu_i]$ *значений* x; *информация Фишера* $I_i(\theta) = \mathbf{E}_\theta \left(\dfrac{f'_{\theta,i}(\mathrm{x}_i)}{f_{\theta,i}(\mathrm{x}_i)}\right)^2$ *положительна и ограничена на* Θ. (Это есть условие (R) без свойств непрерывности.) При выполнении этих условий в § 31 было установлено, что

$$\inf_{\theta,u} \frac{r_i(u)}{u^2} \geqslant g_i > 0, \tag{1}$$

где g_i при $|u| \leqslant u_0$ и малых u_0 может быть выбрано сколь угодно близким к $\dfrac{1}{4} I_i(\theta)$. Нам будет удобнее использовать свойство (1) в его равномерном по отношению к частному $g_i/I_i(\theta)$ виде. Именно, мы будем предполагать, что выполнено условие

(A_u) *Существуют* $g > 0$, $c_1 < \infty$ *такие, что при всех* θ, $\theta + u \in \Theta$

$$\frac{r_i(u)}{u^2} \geqslant g I_i(\theta), \quad I_i(\theta+u) < c_1 I_i(\theta).$$

Обозначим, как и в § 33, $p(u) = Z^{3/4}(u)$.

Теорема 1. *При выполнении сформулированных выше условий*

$$\mathbf{E}_\theta Z^{1/2}(u) \leqslant \exp\left\{-\frac{gu^2}{2} I(\theta, n)\right\}, \quad \mathbf{E}_\theta p(u) \leqslant \exp\left\{-\frac{gu^2}{4} I(\theta, n)\right\},$$

$$\mathbf{E}_\theta |p'(u)| \leqslant \frac{3}{4} I^{1/2}(\theta+u, n) \exp\left\{-\frac{gu^2}{4} I(\theta, n)\right\}.$$

Доказательство аналогично доказательству теоремы 33.1. Из результатов § 31 следует, что

$$\mathbf{E}_\theta Z^{1/2}(u) = \prod_{i=1}^n \left(1 - \frac{r_i(u)}{2}\right).$$

Правая часть здесь не превосходит в силу (A_u)

$$\exp\left\{-\frac{1}{2}\sum_{i=1}^n r_i(u)\right\} \leqslant \exp\left\{-\frac{gu^2}{2} I(\theta, n)\right\}.$$

Далее, в силу неравенств Коши–Буняковского (ср. с § 33)

$$\mathbf{E}_\theta p(u) \leqslant [\mathbf{E}_\theta Z^{1/2}(u)\mathbf{E}_\theta Z(u)]^{1/2} = [\mathbf{E}_\theta Z^{1/2}(u)]^{1/2} \leqslant \exp\left\{-\frac{gu^2}{4}I(\theta,n)\right\},$$

$$\mathbf{E}_\theta|p'(u)| = \frac{3}{4}\mathbf{E}_\theta|L'(X,\theta+u)|Z^{1/2}(u)Z^{1/4}(u) \leqslant$$

$$\leqslant \frac{3}{4}\left\{\mathbf{E}_\theta\left[L'_\theta(x,\theta+u)\right]^2 Z(u)\mathbf{E}_\theta Z^{1/2}(u)\right\}^{1/2} \leqslant$$

$$\leqslant \frac{3}{4}\left\{\mathbf{E}_{\theta+u}\left[L'_\theta(X,\theta+u)\right]^2\right\}^{1/2}\exp\left\{-\frac{gu^2}{4}I(\theta,n)\right\}.$$

Теорема 1 доказана.

Теорема 2. *При всех* $z, n \geqslant 1$

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant u} Z(vI^{-1/2}(\theta,n)) > \exp z\right) \leqslant c\exp\left\{-\frac{3z}{4}\right\}\exp\left\{-\frac{gu^2}{4}\right\},$$

где $c = 1 + 3\sqrt{c_1/g}$, c_1 — *постоянная из условия* (A_u).

Доказательство. Оценим функцию $H(\varepsilon) = \mathbf{E}_\theta \sup\limits_{|v|\geqslant\varepsilon} p(v)$. Будем считать, что $\Theta = [a,b]$. Тогда для $w \in [\theta+\varepsilon, b]$

$$p(w-\theta) = p(\varepsilon) + \int\limits_{\varepsilon}^{w-\theta} p'(u)\,du \leqslant p(\varepsilon) + \int\limits_{\varepsilon}^{w-\theta} |p'(u)|\,du,$$

$$H_+(\varepsilon) = \sup_{u\geqslant\varepsilon} p(u) \leqslant p(\varepsilon) + \int\limits_{u\geqslant\varepsilon} |p'(u)|\,du.$$

Пользуясь теоремой 1 и условием (A_u), получаем (ср. с § 33)

$$H_+(\varepsilon) \leqslant \exp\left\{-\frac{g\varepsilon^2}{4}I(\theta,n)\right\} + \frac{3}{4}\int\limits_{u\geqslant\varepsilon} I^{1/2}(\theta+u,n)\exp\left(-\frac{gu^2}{4}I(\theta,n)\right)du \leqslant$$

$$\leqslant \exp\left(-\frac{g\varepsilon^2}{4}I(\theta,n)\right) + \frac{3c_1^{1/2}}{4}I^{1/2}(\theta,n)\int\limits_{u\geqslant\varepsilon}\exp\left(-\frac{gu^2}{4}I(\theta,n)\right)du \leqslant$$

$$\leqslant \exp\left\{-\frac{g\varepsilon^2}{4}I(\theta,n)\right\}\left(1 + \frac{3}{2}\sqrt{\frac{\pi c_1}{g}}\right).$$

Вместе с аналогичной оценкой для $H_-(\varepsilon) = \sup\limits_{u\leqslant-\varepsilon} p(u)$ это дает такое же неравенство для $H(\varepsilon) = \max(H_-(\varepsilon), H_+(\varepsilon))$. Остается, как и в теореме 33.2,

воспользоваться неравенством Чебышева для оценки

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant u} Z(vI^{-1/2}(\theta,n)) > \exp z\right) = \mathbf{P}_\theta\left(\sup_{|v|\geqslant uI^{-1/2}(\theta,n)} p(v) > \exp\left\{\frac{3z}{4}\right\}\right) \leqslant$$

$$\leqslant \exp\left\{-\frac{3z}{4}\right\} H(uI^{-1/2}(\theta,n)) \leqslant \left(1+3\sqrt{\frac{c_1}{g}}\right)\exp\left\{-\frac{3z}{4}\right\}\exp\left\{-\frac{gu^2}{4}\right\}.$$

Теорема 2 доказана. $\triangleleft$

Из теоремы 2 вытекает

Теорема 3. *Существуют значения* $c<\infty$, $g>0$ *такие, что*

$$\mathbf{P}_\theta(|\widehat{\theta}^* - \theta| I^{1/2}(\theta,n) \geqslant v) \leqslant c\exp\left\{-\frac{gv^2}{4}\right\}.$$

Доказательство. Как и в § 33, следует воспользоваться включением $\{|\widehat{\theta}^* - \theta| \leqslant \varepsilon\} \supseteq \left\{\sup_{|t|\geqslant\varepsilon} Z(t) \geqslant 1\right\}$, в силу которого и теоремы 2 получаем утверждение теоремы. $\triangleleft$

Следствие 1. *Если* $I(\theta,n)\to\infty$ *при* $n\to\infty$, *то оценка* $\widehat{\theta}^*$ *состоятельна.*

Более того, теорема 3 дает оценки вероятностей уклонений $|\widehat{\theta}^* - \theta| \geqslant \varepsilon$. Она позволяет также утверждать, что справедливо

Следствие 2. *При любом* $\alpha \leqslant g/5$ *существует* $c_2<\infty$, *не зависящее от* n *и* θ, *такое, что*

$$\mathbf{E}_\theta \exp\{\alpha(u^*)^2\} < c_2,$$

где

$$u^* = I^{1/2}(\theta,n)(\widehat{\theta}^* - \theta)$$

(ср. с § 33).

Теорема 3 и следствие 2 представляют собой значительно более сильные утверждения, чем утверждение о состоятельности о. м. п., приведенное в § 66. Они означают, в частности, существование второго момента $\widehat{\theta}^* - \theta$, убывающего, как $I^{-1}(\theta,n)$. Утверждения § 66 такого вывода делать не позволяли.

2. Асимптотическая нормальность о. м. п. Как и в однородном случае, доказательство асимптотической нормальности требует более ограничительных условий на гладкость функций $l_i(x,\theta)$. Эти условия можно найти в теоремах 65.3, 66.2. Воспроизведем здесь эти условия.

(RR) 1. *Выполнены условия* (A_0), (A_c), (A_u), (R).

2. *Функции* $l_i(x,\theta)$ *дважды непрерывно дифференцируемы по* θ, *при этом*

$$\sup_{\theta\in\Theta}|l_i''(x,\theta)| \leqslant l_i(x), \quad l_i \equiv \int l_i(x) f_{\theta,i}(x)\mu_i(dx) < \infty.$$

Сходимость интегралов здесь предполагается равномерной в следующем смысле: при $n \to \infty$ *и при каждом* $\tau > 0$

$$\Gamma^{-1}(n) \sum_{l_i(x) > \tau\Gamma(n)} \int l_i(x) f_{\theta,i}(x) \mu_i(dx) \to 0, \quad \Gamma(n) = \sum_{i=1}^{n} l_i.$$

Обозначим

$$\omega_{\Delta,i}(x) = \sup_{t\in\Theta, t+u\in\Theta, |u|\leqslant\Delta} \left| l_i''(x, t+u) - l_i''(x,t) \right|,$$

$$\omega_{\Delta,i} = \mathbf{E}_\theta \omega_{\Delta,i}(\mathrm{x}_i).$$

3. *Функции* $l_i''(x,\theta)$ *непрерывны в среднем:*

$$\Gamma^{-1}(n) \sum \omega_{\Delta,i} \leqslant \omega_\Delta \to 0$$

при $n \to \infty$, *где* ω_Δ *от* n *не зависит.*

4. *При* $n \to \infty$

$$0 < I(\theta, n) = \sum_{i=1}^{n} I_i(\theta) \to \infty,$$

$$I^{-1}(\theta, n) \sum \mathbf{E}_\theta \left[\left(l_i'(\mathrm{x}_i, \theta) \right)^2; \left| l_i'(\mathrm{x}_i, \theta) \right| > \tau I(\theta, n) \right] \to 0.$$

5. *Законно дифференцирование по* θ *под знаком интеграла в равенствах*

$$\int f_{\theta,i} = 1, \quad \int l_i' f_{\theta,i} = 0.$$

В формулировке теоремы 66.2 присутствуют также условия (M_1)–(M_3), нужные для доказательства состоятельности $\widehat{\theta}^*$. Так как состоятельность в силу следствия 1 уже доказана, то необходимость в этих условиях здесь отпадает.

Из теорем 65.3, 66.2 вытекает

Теорема 4. *При выполнении условий* (RR)

$$(\widehat{\theta}^* - \theta) I^{1/2}(\theta, n) = \xi_n (1 + \varepsilon_n(X, \theta)), \tag{2}$$

где

$$\varepsilon_n(X,\theta) \underset{P}{\longrightarrow} 0, \quad \xi_n = I^{-1/2}(\theta, n) \sum_{i=1}^{n} l_i'(\mathrm{x}_i, \theta) \Rightarrow \Phi_{0,1}.$$

Это утверждение в силу п. 1 можно дополнить следующим образом.

Теорема 5. *Сходимость*

$$u^* \equiv (\widehat{\theta}^* - \theta) I^{1/2}(\theta, n) \Rightarrow \Phi_{0,1}, \tag{3}$$

вытекающая из (2), *имеет место вместе со сходимостью моментов любого порядка, т. е. вместе с* (3) *при любом* $k > 0$ *выполняется*

$$\mathbf{E}_\theta (u^*)^k \to \mathbf{E}\eta^k, \quad \eta \in \mathbf{\Phi}_{0,1}.$$

Более того, для любой непрерывной функции $w(t)$ *такой, что* $|w(t)| < e^{gt^2/6}$ (см. следствие 2), $\mathbf{E}_\theta w(u^*) \to \mathbf{E}w(\eta)$, $\eta \in \mathbf{\Phi}_{0,1}$.

Доказательство этого утверждения повторяет доказательство теоремы 33.5.

3. Асимптотическая эффективность. Асимптотическая эффективность о. м. п. в классе K_M M-оценок была доказана в § 66. Совершенно аналогично однородному случаю можно установить теперь, что о. м. п. обладает всеми другими возможными свойствами асимптотической оптимальности из тех, что были введены нами в рассмотрение в п. 66.3 и в § 18, 21.

Предварительно следует заметить, что

1. О. м. п. $\widehat{\theta}^*$ принадлежит классу $K_{\Phi,2}$ оценок θ^*, для которых имеет место сходимость со вторыми моментами (см. п. 66.3):

$$(\theta^* - \theta)\sigma^{-1}(\theta, n) \Rightarrow \mathbf{\Phi}_{0,1}, \quad \mathbf{E}_\theta (\theta^* - \theta)^2 \sigma^{-2}(\theta, n) \to 1.$$

В этом классе среднеквадратический и асимптотический подходы к сравнению оценок совпадают.

2. О. м. п. $\widehat{\theta}^*$ принадлежит классу $\widetilde{K}_0$ асимптотически несмещенных оценок, для которых $b(\theta) = o(\sigma(\theta, n))$, $b'(\theta) = o(1)$.

Это доказывается точно так же в § 33, 34. Аналогично § 33, 34 доказываются и следующие утверждения.

Теорема 6. *Если выполнены условия* (RR), *то*

1) *о. м. п.* $\widehat{\theta}^*$ *является а.R-э. оценкой*;

2) *о. м. п.* $\widehat{\theta}^*$ *является а. э. оценкой в* $\widetilde{K}_0$;

3) *о. м. п.* $\widehat{\theta}^*$ *является асимптотически R-байесовской и асимптотически байесовской оценкой для любого априорного распределения* Q, *имеющего плотность относительно меры Лебега* (см. теоремы 33.8, 30.5).

Все изменения, которые надо внести в доказательства § 33, 34 при переходе к неоднородным наблюдениям, сводятся к замене $nI(\theta)$ на $I(\theta, n)$, а также к некоторым другим не менее очевидным модификациям.

Утверждение об асимптотической минимаксности о. м. п. требует введения элементов равномерности по θ в условиях (RR) (ср. с условиями (RR_u) в § 34). Установление свойств минимаксности при таких равномерных условиях безусловно возможно, но требует предварительных рассмотрений, аналогичных тем, что были проведены в § 37. Читателю, хорошо усвоившему результаты и доказательства § 33, 34, это под силу.

4. Замечания об о. м. п. для многомерного параметра. В многомерном случае $\Theta \subset \mathbb{R}^k$, $k > 1$, наиболее широкие условия *состоятельности* о. м. п. можно получить из теоремы 66.1 (условия этой теоремы с размерностью θ, по существу, не связаны).

Доказательство *асимптотической нормальности* о. м. п. в многомерном случае возможны путем использования как подходов, близких к тем, что изложены в § 65 при изучении M-оценок, так и подходов, изложенных в § 34 и 70. При этом доказательство и сама формулировка условий регулярности (условий типа (RR)), обеспечивающих эту нормальность, оказываются более громоздкими и трудоемкими. Изменится вид тех компонент условий (RR) (ср. с теоремами 67.5 и 36.6), которые содержат отношения сходимости и неравенства (например, условия (RR_u), (36.3) и др.). Вместо условия $I(\theta,n)\to\infty$ в одномерном случае здесь надо предполагать $\Lambda_1(\theta,n)\to\infty$, где $\Lambda_1(\theta,n)$ — минимальное собственное значение матрицы $I(\theta,n)=\sum\limits_{i=1}^{n} I_i(\theta)$. При этом минимальное и максимальное собственные значения $I(\theta,n)$ могут сходиться к ∞, вообще говоря, с разной скоростью (см., например, значение $I(\theta,n)$ в § 68). «Качественный» вид основного утверждения остается прежним:

$$(\widehat{\theta}^*-\theta)I^{1/2}(\theta,n)\Rightarrow\mathbf{\Phi}_{0,E},$$

где E — единичная матрица; при этом сходимость имеет место и для моментов любого порядка; о. м. п. $\widehat{\theta}^*$ является асимптотически эффективной в том же смысле, что и в теореме 6.

§ 71. Замечания о проверке гипотез по неоднородным наблюдениям

Все основные принципы построения критериев для проверки статистических гипотез, изложенные в гл. 3, при переходе к статистике неоднородных наблюдений полностью сохраняются.

Пусть дана выборка $X=(\mathrm{x}_1,\ldots,\mathrm{x}_n)\in\mathbf{P}=\mathbf{P}_1\times\ldots\times\mathbf{P}_n$ и по ней проверяется некоторая гипотеза H относительно распределения $\mathbf{P}$ этой выборки. В этом случае наблюдение X можно рассматривать как выборку объема 1 из распределения $\mathbf{P}$ и применять все выводы и результаты гл. 3. (Отметим, что это замечание справедливо и для зависимых наблюдений $\mathrm{x}_1,\ldots,\mathrm{x}_n$; ср. также с гл. 6.)

Рассмотрим простейший пример (см. пример 64.5), когда проверяются две простые гипотезы $H_j=\{X\in\mathbf{P}^{(j)}\}$, $j=1,2$, при выполнении условия (A_μ) о том, что распределение элемента выборки x_i имеет плотность относительно некоторой меры μ_i. Обозначим через $f_i^{(j)}(x)$ эту плотность при гипотезе H_j, $j=1,2$, и через $f^{(j)}(X)$ — функцию правдоподобия $f^{(j)}(X)=\prod\limits_{i=1}^{n} f_i^{(j)}(\mathrm{x}_i)$, так что $f^{(j)}(X)$ есть плотность распределения $\mathbf{P}^{(j)}$ выборки X при гипотезе H_j относительно меры $\mu_1\times\ldots\times\mu_n$. Пусть, далее, $Z(X)$ есть отношение правдоподобия $Z(X)=f^{(2)}(X)/f^{(1)}(X)$.

Тогда согласно теореме 42.1 все оптимальные критерии (байесовские, минимаксные, н. м. к.) будут основаны на статистике $Z(X)$, и в том случае, когда распределение $Z(X)$ при обеих гипотезах непрерывно, будут иметь вид

$\delta_c(X) = I(Z(X) > c)$ при надлежащем выборе c. Если $\mathbf{P}^{(j)}(Z(X) = c) > 0$, то необходимо рандомизировать критерий (обозначим его через $\pi_c(X)$), положив $\pi_c(X) = p(X) > 0$ на множестве $\{Z(X) = c\}$; $\pi_c(X) = I(Z(X) > c)$ при других X. Пусть для простоты $\mathbf{P}^{(j)}(Z(X) = c) = 0$ при всех c. Тогда о. м. к. $\delta_c(X)$ будет иметь вероятности ошибок первого и второго рода равными соответственно

$$\alpha_1(\delta_c) = \mathbf{P}^{(1)}(Z(X) \geqslant c), \quad \alpha_2(\delta_c) = \mathbf{P}^{(2)}(Z(X) < c). \tag{1}$$

Стоящее под знаком вероятности в (1) событие можно записать в виде

$$\sum_{i=1}^{n} \eta_i \geqslant \ln c, \quad \eta_i = \ln \frac{f_i^{(2)}(\mathrm{x}_i)}{f_i^{(1)}(\mathrm{x}_i)},$$

где случайные величины η_i независимы. Если рассматривать н. м. к. уровня ε, то надо выбрать c так, чтобы $\mathbf{P}^{(1)}\left(\sum_{i=1}^{n} \eta_i \geqslant \ln c\right) = \varepsilon$.

Положим

$$a_i = \mathbf{E}_{P^{(1)}}\eta_i = \int f_i^{(1)}(x) \ln \frac{f_i^{(2)}(x)}{f_i^{(1)}(x)} \mu_i(dx) = -\rho_1\left(\mathbf{P}_i^{(1)}, \mathbf{P}_i^{(2)}\right),$$

$$b_i = \mathbf{E}_{P^{(2)}}\eta_i = \int f_i^{(2)}(x) \ln \frac{f_i^{(2)}(x)}{f_i^{(1)}(x)} \mu_i(dx) = \rho_1\left(\mathbf{P}_i^{(2)}, \mathbf{P}_i^{(1)}\right),$$

где распределения $\mathbf{P}_i^{(j)}$ соответствуют плотностям $f_i^{(j)}$, $\rho_1(\mathbf{P}, \mathbf{Q})$ — расстояние Кульбака–Лейблера между $\mathbf{P}$ и $\mathbf{Q}$ (см. замечание 31). Положим далее

$$\left(\sigma_i^{(j)}\right)^2 = D^{(j)}\eta_i, \quad \left(B_n^{(j)}\right)^2 = \sum_{i=1}^{n} \left(\sigma_i^{(j)}\right)^2, \quad j = 1, 2.$$

В дальнейшем мы не будем исключать схему серий, когда распределения $\mathbf{P}_i^{(j)}$ (плотности $f_i^{(1)}$) зависят от n. Мы можем теперь во введенных обозначениях записать $\alpha_1(\delta_c)$ в виде

$$\alpha_1(\delta_c) = \mathbf{P}^{(1)}\left(\sum \eta_i \geqslant \ln c\right) = \mathbf{P}^{(1)}\left(\frac{\sum(\eta_i - a_i)}{B_n^{(1)}} \geqslant \frac{\ln c - \sum a_i}{B_n^{(1)}}\right).$$

При выполнении условия Линдеберга

$$\sum_{i=1}^{n} \mathbf{E}_{P^{(1)}}\left(\eta_i^2; |\eta_i| > \tau B_n^{(1)}\right) \to 0, \quad \tau > 0,$$

и при c таком, что $\left(\ln c - \sum a_i\right) \Big/ B_n^{(1)} = \lambda_\varepsilon$, где λ_ε, как и прежде, — квантиль нормального распределения порядка $1 - \varepsilon$, мы будем иметь в силу центральной предельной теоремы

$$\alpha_1(\delta_c) = \mathbf{P}^{(1)}\left(\frac{\sum(\eta_i - a_i)}{B_n^{(1)}} \geqslant \lambda_\varepsilon\right) \to \varepsilon,$$

так что при выбранном c критерий δ_c будет иметь асимптотический уровень $1-\varepsilon$. Этот критерий имеет следующую вероятность ошибки второго рода:

$$\alpha_2(\delta_c) = P^{(2)}\left(\frac{\sum(\eta_i - b_i)}{B_n^{(2)}} < \frac{\ln c - \sum b_i}{B_n^{(2)}}\right) = \\ = \mathbf{P}^{(2)}\left(\frac{\sum(\eta_i - b_i)}{B_n^{(2)}} < -\frac{\sum(b_i - a_i)}{B_n^{(2)}} + \lambda_\varepsilon \frac{B_n^{(1)}}{B_n^{(2)}}\right).$$

Здесь

$$\sum(b_i - a_i) = \sum\left(\rho_1\left(\mathbf{P}_i^{(2)}, \mathbf{P}_i^{(1)}\right) + \rho_1\left(\mathbf{P}_i^{(1)}, \mathbf{P}_i^{(2)}\right)\right) = \\ = \rho_1(\mathbf{P}^{(2)}, \mathbf{P}^{(1)}) + \rho_1(\mathbf{P}^{(1)}, \mathbf{P}^{(2)}) = \rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$$

является «симметризированным» расстоянием Кульбака–Лейблера $\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$ между $\mathbf{P}^{(1)}$ и $\mathbf{P}^{(2)}$. Если выполнено условие Линдеберга

$$\sum \mathbf{E}_{P^{(2)}}\left(\eta_i^2; |\eta_i| > \tau B_n^{(2)}\right) \to 0, \quad \tau > 0,$$

то здесь также применима центральная предельная теорема и соотношение $\alpha_2(\delta_c) \to \beta \geqslant 0$ эквивалентно тому, что

$$-\frac{\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})}{B_n^{(2)}} + \lambda_\varepsilon \frac{B_n^{(1)}}{B_n^{(2)}} \to \lambda_{1-\beta}. \tag{2}$$

Поэтому малые значения β (или $\beta = 0$) возможны лишь в случае, когда при больших n велико значение $\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})/B_n^{(2)}$.

Будем говорить, что гипотезы H_j *фиксированы*, если $f_i^{(j)}$ (а следовательно, и распределение η_i) от n не зависят. В этом случае необходимым условием состоятельности критерия δ_c (не ограничивая общности, можно считать, что $\sigma_1^{(2)} > 0$) является неограниченный рост вместе с n расстояния $\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$. Если $\left(B_n^{(2)}\right)^2$ растет с той же скоростью, что и $\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)})$, как это имеет место в случае однородных наблюдений, то условие $\rho(\mathbf{P}^{(1)}, \mathbf{P}^{(2)}) \to \\ \to \infty$ является и достаточным для состоятельности δ_c.

Будем называть гипотезы H_j *близкими* (но различимыми), если $f_i^{(j)}$ зависят от n, $\rho\left(\mathbf{P}_i^{(1)}, \mathbf{P}_i^{(2)}\right) \to 0$ и имеет место (2) при $\lambda_{1-\beta} \neq -\infty$, $\beta < \\ < 1-\varepsilon$. Сюда будет относиться, в частности, случай параметрических гипотез $H_j = \{X \in P_{\theta_j}\}$, $j = 1, 2$, когда $\mathbf{P}_\theta = \mathbf{P}_{\theta,1} \times \ldots \times \mathbf{P}_{\theta,n}$, $\mathbf{P}_{\theta,i}$ от n не зависят, $\theta_2 = \theta_1 + vI^{-1/2}(\theta_1, n)$, информационное количество Фишера $I(\theta_1, n) = \sum_{i=1}^{n} I_i(\theta_1)$ обладает свойством $I(\theta_1, n) \to \infty$ при $n \to \infty$ (параметр θ для простоты считаем скалярным). Если выполнены условия регулярности типа (RR), то, пользуясь результатами § 70, читатель без труда

сможет получить приближенные значения $\alpha_j(\delta_c)$ (ср. с (43.20))

$$\alpha_1(\delta_c) \approx 1 - \Phi\left(\frac{v^2/2 + \ln c}{|v|}\right), \quad \alpha_2(\delta_c) \approx \left(\frac{-v^2/2 + \ln c}{|v|}\right).$$

Изучение асимптотического поведения $\alpha_j(\delta_c)$ в случае фиксированных гипотез оказывается более трудной и менее исследованной задачей (за исключением случая однородных наблюдений; ср. с § 43).

Используя аналогичные модификации, можно перенести на случай неоднородных наблюдений и результаты § 53–55 об асимптотически оптимальных критериях для проверки близких сложных гипотез. При этом при необходимости следует использовать результаты об асимптотическом представлении отношения правдоподобия, приведенные в § 70.

§ 72. Задача о разладке (пример 64.6)

Обратимся теперь к примеру 64.6 и несколько уточним постановку задачи. Пусть дана последовательность X независимых наблюдений $X = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_{\theta-1}, \mathrm{x}_\theta, \ldots)$, первые $\theta - 1$ из которых имеют распределение G_1, остальные — распределение $G_2 \neq G_1$. Не ограничивая общности, можно считать, что G_1 и G_2 имеют плотности (абсолютно непрерывны) относительно некоторой меры μ. В качестве μ можно взять, например, $\mu = G_1 + G_2$. Параметр $\theta < \infty$ неизвестен, и задача, которую мы будем рассматривать, состоит в том, чтобы по возможности точно «угадать» или оценить этот параметр по наблюдению X. Двойственный характер этой задачи состоит в том, что, с одной стороны, ее можно рассматривать как задачу о проверке конечного или счетного числа гипотез $H_j = \{\theta = j\}$, $j = 1, 2, \ldots, n$; $n \leqslant \infty$, и тогда естественно рассматривать задачу об отыскании оптимального критерия $\delta(X)$, который минимизировал бы вероятность ошибки или, что то же, максимизировал бы вероятности $\mathbf{P}_t(\delta(X) = H_t)$ «угадывания» неизвестного параметра. Через $\mathbf{P}_t$ мы здесь обозначаем аналогично предыдущему распределение в пространстве последовательностей X в случае $\theta = t$.

С другой стороны, параметр θ носит количественный характер и, стало быть, естественно рассматривать задачу об оценке параметра θ и об отыскании такой оценки $\theta^* = \theta^*(X)$, которая минимизировала бы среднеквадратические уклонения $\mathbf{E}_\theta(\theta^* - \theta)^2$. Ниже мы рассмотрим оба этих подхода в их асимптотической форме, предполагая параметр θ большим.

Задаче о разладке посвящена обширная литература (см., например, [41, 53, 71, 79, 90, 138]); некоторые из известных результатов мы будем использовать.

1. Задача о разладке как задача проверки гипотез. Мы уже отмечали, что θ является количественным параметром. Поэтому в этом разделе наряду с термином «критерий $\delta(X)$», принимающим значения в множестве $\{H_1, H_2, \ldots\}$, нам будет удобно использовать также термин «оценка $\theta^* = \theta^*(X)$», где θ^* принимает значения $1, 2, \ldots$ Функцию $\theta^*(X)$, как и $\delta(X)$, в рамках этого раздела можно также называть критерием.

Обозначим

$$r_i = \frac{g_1(\mathrm{x}_i)}{g_2(\mathrm{x}_i)}, \quad R_k = R(k, X) = \prod_{i=1}^{k-1} r_i. \tag{1}$$

Определение 1. *Критерием* (оценкой) максимального правдоподобия называется такое наименьшее число $\widehat{\theta}^* = \widehat{\theta}^*(X)$, что

$$R(\widehat{\theta}^*, X) \geqslant R(t, X)$$

при всех $t \leqslant n+1$, где $n \leqslant \infty$ — число наблюдений.

Если $n < \infty$, то $\theta = \widehat{\theta}^*$ максимизирует функцию правдоподобия

$$f_\theta(X) = \prod_{i=1}^{\theta-1} g_1(\mathrm{x}_i) \prod_{i=\theta}^{n} g_2(\mathrm{x}_i). \tag{2}$$

Определение 2. Критерий $\theta^* = \theta^*(X)$ называется *асимптотически однородным*, если

$$\mathbf{P}_t(\theta^* = t) \to p \geqslant 0 \tag{3}$$

при $t \to \infty$.

Если мы имеем дело с ограниченной последовательностью $n < \infty$ наблюдений, то в (3) предполагается также, что $n \to \infty$, $n - t \to \infty$.

Покажем, что о. м. п. $\widehat{\theta}^*$ является асимптотически однородной и найдем соответствующее ей значение $p = \widehat{p}$. Нам понадобятся следующие обозначения. Рассмотрим две независимые друг от друга последовательности $\{\xi_i\}$ и $\{\zeta_i\}$ независимых случайных величин, определенных равенствами

$$\xi_i = -\ln \frac{g_1(\mathrm{x}_i)}{g_2(\mathrm{x}_i)}, \quad \mathrm{x}_i \in G_1;$$

$$\zeta_i = \ln \frac{g_1(\mathrm{x}_i)}{g_2(\mathrm{x}_i)}, \quad \mathrm{x}_i \in G_2.$$

Очевидно, что

$$\mathbf{E}\xi_i = -\int g_1(x) \ln \frac{g_1(x)}{g_2(x)} \mu(dx) = -\rho_1(G_1, G_2) < 0,$$

$$\mathbf{E}\zeta_i = \int g_2(x) \ln \frac{g_1(x)}{g_2(x)} \mu(dx) = -\rho_1(G_2, G_1) < 0,$$

где ρ_1 — расстояние Кульбака–Лейблера, введенное в § 31. Оба расстояния $\rho_1(G_1, G_2)$ и $\rho(G_2, G_1)$ мы будем предполагать конечными. Обозначим далее

$$S_n = \sum_{i=1}^{n} \xi_i, \quad Z_n = \sum_{i=1}^{n} \zeta_i,$$

$$S = \sup_{k \geqslant 1} S_k, \quad Z = \sup_{k \geqslant 1} Z_k.$$

Так как $\mathbf{E}\xi_i$ и $\mathbf{E}\zeta_i$ отрицательны, то в силу усиленного закона больших чисел S и Z — собственные случайные величины,

$$\mathbf{P}(S<0)>0, \quad \mathbf{P}(Z<0)>0$$

(см., например, [12]; [17, § 11.3]).

Распределения S и Z могут быть найдены в явном виде в терминах компонент факторизации функций $1-z\varphi(\lambda)$ и $1-z\psi(\lambda)$ соответственно, где $\varphi(\lambda)=\mathbf{E}e^{\lambda\xi_1}$, $\psi(\lambda)=\mathbf{E}e^{\lambda\zeta_1}$ (см. [12]; [17, § 11.3]). Отметим, что $\psi(\lambda)=\varphi(1-\lambda)$ и $\varphi(\lambda)\leqslant 1$ при $\lambda\in[0,1]$. Действительно,

$$\varphi(\lambda)=\int g_1(x)\left(\frac{g_1(x)}{g_2(x)}\right)^{-\lambda}\mu(dx)=$$

$$=\int g_1^{1-\lambda}(x)g_2^{\lambda}(x)\mu(du)\leqslant\left(\int g_1\,d\mu\right)^{1/(1-\lambda)}\left(\int g_2\,d\mu\right)^{1/\lambda}=1.$$

Такое же неравенство справедливо для $\psi(\lambda)$. Кроме того, производные $\varphi'(\lambda)=$ $=-\psi'(1-\lambda)$ в силу сделанных предположений конечны на всем отрезке $[0,1]$, $\varphi'(0)<0$, $\varphi'(1)>0$.

Теорема 1. *Оценка максимального правдоподобия $\widehat{\theta}^*$ является асимптотически однородной,*

$$\mathbf{P}_t(\widehat{\theta}^*=t)\downarrow\widehat{p}=\mathbf{P}(S<0)\mathbf{P}(Z\leqslant 0) \tag{4}$$

при $t\uparrow\infty$ (и $n-t\uparrow\infty$, если $n<\infty$). Кроме того,

$$\begin{aligned}\mathbf{P}_t(\widehat{\theta}^*>t)&\to\mathbf{P}(Z>S;\ Z>0),\\ \mathbf{P}_t(\widehat{\theta}^*<t)&\to\mathbf{P}(Z\leqslant S;\ S\geqslant 0),\end{aligned} \tag{5}$$

$$\mathbf{P}_t(\widehat{\theta}^*>t+k)\leqslant 2\varphi^k, \quad \mathbf{P}_t(\widehat{\theta}^*<t-k)\leqslant 2\varphi^k, \tag{6}$$

где

$$\varphi=\min_{\lambda}\mathbf{E}e^{\lambda\xi}<1.$$

Более полные сведения о предельном $\mathbf{P}_t$-распределении разности $\widehat{\theta}^*-t$ содержатся в приложении VII.

Доказательство. Имеем

$$\{\widehat{\theta}^*=t\}=\left\{\prod_{i=1}^{t-1}r_i>\prod_{i=1}^{k-1}r_i \quad \text{при всех} \quad k,\ 1\leqslant k<t;\right.$$

$$\left.\prod_{i=1}^{t-1}r_i\geqslant\prod_{i=1}^{k-1}r_i \quad \text{при всех} \quad k,\ n+1\geqslant k>t\right\}.$$

Поэтому в силу независимости x_i

$$\mathbf{P}_t(\widehat{\theta}^* = t) = \mathbf{P}(-\xi_k - \ldots - \xi_{t-1} > 0 \quad \text{при всех} \quad k,\ 1 \leqslant k < t) \times$$
$$\times\, \mathbf{P}(\zeta_t + \ldots + \zeta_{k-1} \leqslant 0 \quad \text{при всех} \quad k,\ t < k \leqslant n+1) \uparrow$$
$$\uparrow \mathbf{P}(S < 0)\mathbf{P}(Z \leqslant 0)$$

при $t \uparrow \infty$, $n - t \uparrow \infty$. Аналогично находим

$$\{\widehat{\theta}^* > t\} = \left\{ \max_{k>t} \prod_{i=1}^{k-1} r_i > \max_{s \leqslant t} \prod_{i=1}^{s-1} r_i \right\};$$

$$\mathbf{P}_t(\widehat{\theta}^* > t) = \mathbf{P}\left(\max_{t<k\leqslant n+1} (\zeta_t + \ldots + \zeta_{k-1}) > \max_{s \leqslant t} (\xi_s + \ldots + \xi_{t-1}) \right) =$$
$$= \mathbf{P}\left(\max_{1 \leqslant l \leqslant n-t} Z_l > \max_{0 \leqslant l \leqslant t-1} S_l \right) \to \mathbf{P}(Z > S;\ Z > 0).$$

Далее,

$$\{\widehat{\theta}^* > t + k\} = \left\{ \max_{j>t+k} \prod_{i=1}^{j-1} r_i > \max_{j>t+k} \prod_{i=1}^{j-1} r_i \right\};$$

$$\mathbf{P}_t(\widehat{\theta}^* > t + k) = \mathbf{P}\Big(S_{t-1} + Z_k + \max_{1 \leqslant l \leqslant n-t-k} Z_l^* >$$
$$> \max(S_1, \ldots, S_{t-1}, S_{t-1} + \zeta_1, \ldots, S_{t-1} + Z_k) \Big), \quad (7)$$

где последовательность $\{Z_l^*\}$ имеет то же распределение, что и последовательность $\{Z_l\}$, и не зависит от нее. Так как последний max под знаком вероятности в (7) больше или равен S_{t-1}, то вероятность (7) не превосходит

$$\mathbf{P}(Z_k + Z^* > 0) \leqslant \mathbf{P}(Z_k > 0) + \int_{-\infty}^{\infty} \mathbf{P}(Z_k \in dv)\mathbf{P}(Z^* > -v).$$

Так как $\psi(\lambda) = \mathbf{E}e^{\lambda\zeta_1} = 1$ при $\lambda = 1$ (см. ниже), то

$$\mathbf{P}(Z^* > u) \leqslant e^{-u}, \quad \mathbf{P}(Z_k > 0) \leqslant \psi^k(\lambda)$$

при любом λ, $1 \geqslant \lambda \geqslant 0$ (см., например, [12, 17]). Поэтому

$$\mathbf{P}_t(\widehat{\theta}^* > t + k) \leqslant \psi^k(\lambda) + \int_{-\infty}^{\infty} e^{v\lambda}\mathbf{P}(Z_k \in dv) \leqslant \psi^k(\lambda) + \psi^k(\lambda) = 2\varphi^k$$

при $\lambda = \lambda_0$, $\varphi = \psi(\lambda_0) = \min\limits_{(\lambda)} \psi(\lambda) < 1$. Второе неравенство в (6) доказывается аналогично. Теорема доказана. $\triangleleft$

Класс $K_{(0)}$ асимптотически однородных критериев, введенный в определении 2, следует считать весьма естественным, поскольку он состоит из критериев (оценок), ведущих себя «симметричным образом» по отношению к разным большим значениям параметра θ. Аналогичное свойство симметричности в теории оценивания и в теории проверки гипотез вводилось, например, понятиями несмещенности или инвариантности. Если мы не имеем дело с байесовской постановкой задачи, то у нас нет оснований выделять те или иные значения θ, так что для разумных оценок θ^* значения $\mathbf{P}_t(\theta^* = t)$ должны быть примерно одинаковыми при всех больших значениях t (и $n - t$, если $n < \infty$). Теорема 1 показывает, что класс $K_{(0)}$ не пуст.

Покажем теперь, что о. м. п. $\widehat{\theta}^*$ обладает свойством оптимальности в $K_{(0)}$.

Теорема 2. *Пусть θ^* — любая оценка из $K_{(0)}$:*

$$\mathbf{P}_t(\theta^* = t) \to p \geqslant 0$$

при $t \to \infty$ (и $n - t \to \infty$, если $n < \infty$). Тогда $p \leqslant \widehat{p}$.

Доказательство. Рассмотрим байесовскую постановку задачи, когда на множестве $\{1, \ldots, n\}$ задано априорное распределение Q с вероятностями $q(k) = 1/n$. (Точнее, мы будем рассматривать *последовательность* $Q = Q^{(n)}$ распределений, зависящих от $n \to \infty$.) Байесовский критерий θ_Q^* выбирает в качестве θ значения k, при котором максимизируется апостериорная вероятность

$$q(k/X) = \frac{q(k)f_k(X)}{\sum\limits_{i=1}^{n} q(i)f_i(X)} = \frac{f_k(X)}{\sum\limits_{i=1}^{n} f_i(X)}.$$

Это означает, что $\theta_Q^* = \widehat{\theta}^*$. Далее, для вероятности ошибки $\alpha_Q(\theta^*) = \sum\limits_{k=1}^{n} q(k)\mathbf{P}_k(\theta^* \neq k)$ асимптотически однородного критерия θ^* будем иметь $\alpha_Q(\theta_\theta^*) \leqslant \alpha_Q(\theta^*)$,

$$p = \lim_{n\to\infty} \frac{1}{n} \sum_{n=1}^{n} \mathbf{P}_k(\theta^* = k) = \lim_{n\to\infty} (1 - \alpha_Q(\theta^*)) \leqslant$$

$$\leqslant \lim_{n\to\infty} (1 - \alpha_Q(\widehat{\theta}^*)) = \lim_{n\to\infty} \frac{1}{n} \sum_{n=1}^{n} \mathbf{P}_k(\widehat{\theta}^* = k) = \widehat{p}.$$

Теорема доказана. $\triangleleft$

Можно выделить еще один класс критериев (оценок), в котором о. м. п. будет асимптотически оптимальной. Это класс байесовских критериев с «плавно меняющимися» априорными вероятностями $q(k)$.

Чтобы ввести этот класс, рассмотрим произвольную последовательность $\delta \to 0$. Если $n < \infty$, то мы будем считать, что $n = n(\delta) \to \infty$ при $\delta \to 0$ (можно отождествить, например, δ с $1/n$).

Обозначим через $\mathcal{Q}$ класс распределений Q на $\{1,\dots,n\}$, $n \leqslant \infty$, зависящих от параметра δ и обладающих следующим свойством: для любых $k_1 = k_1(\delta) < k_2 = k_2(\delta)$ таких, что $k_1 \to \infty$, $n - k_2 \to \infty$ при $\delta \to 0$, выполняется

$$\begin{aligned} &\frac{q(k+1)}{q(k)} < e^{\varepsilon} \quad \text{при} \quad k_1 \leqslant k < n, \\ &\frac{q(k-1)}{q(k)} < e^{\varepsilon} \quad \text{при} \quad 2 \leqslant k \leqslant k_2, \end{aligned} \tag{8}$$

где $\varepsilon = \varepsilon(k_1, k_2) \to 0$ при $\delta \to 0$.

Очевидно, что для любого фиксированного l

$$\frac{q(k+l)}{q(k)} \to 1$$

при $\delta \to 0$, $k \to \infty$, $n - k \to \infty$.

Равномерное распределение $q(k) = 1/n$ на множестве $\{1,\dots,n\}$ принадлежит $\mathcal{Q}$ при $\varepsilon = 0$, $\delta = 1/n$. Геометрическое распределение $q(k) = p(1-p)^{k-1}$, $k = 1, 2, \dots$, также принадлежит $\mathcal{Q}$ при $\delta = p \to 0$, $\varepsilon = -\ln(1-p)$.

В обоих этих примерах соотношения (11) выполняются и без условия роста k_1 и $n - k_2$.

Названные два примера можно обобщить следующим образом. Пусть, как и прежде, δ — произвольная последовательность, сходящаяся к нулю, и $p(t)$ — плотность, непрерывная и положительная внутри отрезка $[0, a]$ и равная нулю вне его, $a \leqslant \infty$. Тогда при широких условиях на поведение $p(t)$ на концах интервала $(0, a)$ последовательность

$$q(k) = \int\limits_{(k-1)\delta}^{k\delta} p(u)\,du, \quad k = 1, \dots, \frac{a}{\delta}$$

(считаем для простоты, что δ нацело делит a), будет принадлежать $\mathcal{Q}$. Можно взять, например,

$$p(u) = \begin{cases} c(\alpha,\beta)u^{\alpha}(1-u)^{\beta}, & u \in [0,1], \\ 0, & u \notin [0,1], \end{cases}$$

при любых $\alpha \geqslant 0$, $\beta \geqslant 0$, где $c(\alpha,\beta)$ — нормирующая постоянная.

Для определенности будем называть байесовским критерием $\theta_Q^* = \theta_Q^*(X)$ функцию, которая выбирает *минимальное* значение k, для которого $q(k/X) = \max\limits_j q(j/X)$.

Теорема 3. *Если* $Q \in \mathcal{Q}$, *то*

$$\mathbf{P}_t\big(\theta_Q^* - t > k\big) \leqslant 2(\varphi e^{\varepsilon\lambda_0})^k, \quad \mathbf{P}_t\big(\theta_Q^* - t < -k\big) \leqslant 2(\varphi e^{\varepsilon\lambda_0})^k, \tag{9}$$

где, как и прежде, λ_0 *— точка, в которой достигается* $\min\psi(\lambda)$, $\varphi = \psi(\lambda_0)$. *Если* $Q \in \mathcal{Q}$, $\mathbf{P}(S_j = 0) = \mathbf{P}(Z_j = 0) = 0$ *при всех* j, *то*

$$\mathbf{P}_t\big(\theta_Q^* \neq \widehat{\theta}^*\big) \to 0 \tag{10}$$

при $t \to \infty$, $n - t \to \infty$.

Очевидно, что наряду с (10) имеет место также сходимость $\mathbf{P}\big(\theta_Q^* \neq \widehat{\theta}^*\big) \to 0$ относительно распределения $\mathbf{P}$ пары (Q, X), $\theta \in Q \in \mathcal{Q}$.

Вместе с теоремой 2 утверждение (10) означает, что все байесовские критерии, соответствующие $Q \in \mathcal{Q}$, асимптотически эквивалентны о. м. п. $\widehat{\theta}^*$ и имеют в пределе одну и ту же вероятность ошибки, равную $1 - \widehat{p}$.

Так как о. м. п. $\widehat{\theta}^*$ является байесовским критерием (для равномерного априорного распределения), а $\mathbf{P}_t(\widehat{\theta}^* = t)$ асимптотически не зависит от t, то нетрудно установить также, что $\widehat{\theta}^*$ является *асимптотически минимаксным критерием* при соответствующем естественном определении этого понятия (ср. с § 53).

Доказательство теоремы 3. Доказательство неравенств (9) вполне аналогично доказательству неравенств (6) в теореме 1. Имеем

$$\big\{\theta_Q^* > t + k\big\} = \left\{\max_{j>t+k} q(j)\prod_{i=1}^{j-1} r_i > \max_{j\leqslant t+k} q(j)\prod_{i=1}^{j-1} r_i\right\}.$$

Совершенно аналогично с (7) имеем

$$\begin{aligned}\mathbf{P}_t\big(\theta_Q^* > t + k\big) &= \mathbf{P}\Big(S_{t-1} + Z_k + \max_{l\geqslant 1}\big(Z_l^* + \ln q(t+k+l)\big) > \\ &> \max\big(S_1 + \ln q(2), \ldots, S_{t-1} + \ln q(t);\\ &S_{t-1} + Z_1 + \ln q(t+1), \ldots, S_{t-1} + Z_k + \ln q(t+k)\big)\Big) \leqslant \\ &\leqslant \mathbf{P}\Big(S_{t-1} + Z_k + \varepsilon k + \max_{l\geqslant 1}\big(Z_l^* + \varepsilon l\big) > S_{t-1}\Big) \leqslant \\ &\leqslant \mathbf{P}(Z_k + \varepsilon k > 0) + \int_{-\infty}^{0} \mathbf{P}(Z_k + \varepsilon k \in dv)\mathbf{P}\Big(\max_{l\geqslant 1}\big(Z_l^* + \varepsilon l\big) > -v\Big).\end{aligned}$$

Пусть $s > 0$ — решение уравнения $\mathbf{E}e^{\lambda\zeta+1+\varepsilon\lambda} = \psi(\lambda)e^{\varepsilon\lambda} = 1$. Очевидно, что $s < 1$, $s \to 1$ при $\varepsilon \to 0$. Поэтому при любом фиксированном $\lambda \in (0, 1)$

$$\mathbf{P}_t\big(\theta_Q^* > t + k\big) < (\psi(\lambda)e^{\varepsilon\lambda})^k + \int_{-\infty}^{0} e^{sv}\mathbf{P}(Z_k + \varepsilon k \in dv) < 2(\psi(\lambda)e^{\varepsilon\lambda})^k.$$

Отсюда следует первое утверждение теоремы. Второе неравенство в (9) получается аналогично.

Докажем (10). Событие $\theta_Q^* \neq \widehat{\theta}^*$ наступает в том случае, когда максимум последовательности $q(k)\prod_{i=1}^{k-1} r_i$ достигается в точке k, отличной от точки максимума $\prod_{i=1}^{k-1} r_i$. В силу неравенств (6), (9) мы можем ограничиться рассмотрением значений k, $|t-k| \leqslant N$, при достаточно большом, но фиксированном N. Но в этих пределах

$$m_1 = \min_{|k-t|\leqslant N} \frac{q(k)}{q(t)} \to 1, \quad m_2 = \max_{|k-t|\leqslant N} \frac{q(k)}{q(t)} \to 1$$

при $t \to \infty$, $n-t \to \infty$. Стало быть, $\mathbf{P}_t(\theta_Q^* \neq \widehat{\theta}^*)$ оценивается $\mathbf{P}_t$-вероятностью того, что S_r и Z_s при некоторых $r \leqslant N$, $s \leqslant N$ попадут в интервал в окрестности нуля, шириной не превышающей $\ln m_2 - \ln m_1 \to 0$. Так как распределения S_r и Z_s непрерывны в точке 0, то $\mathbf{P}_t$-вероятности этих событий сходятся к нулю. Теорема доказана. $\triangleleft$

2. Задача о разладке как задача оценки параметра. Если число $\widehat{p}$ в теореме 1 невелико, то шансы точно угадать параметр θ тоже невелики. Поэтому подходы предыдущего раздела, максимизирующие вероятность точного угадывания, не всегда являются наиболее естественными. Часто целью является не точное угадывание θ, а определение значения θ^*, которое было бы по возможности близко к θ, например в среднеквадратичном смысле. Отметим при этом, что мы оцениваем *целочисленный* параметр *случайного процесса* $\mathrm{x}_1, \mathrm{x}_2, \ldots$ по *одному наблюдению* X (в гл. 2 мы оценивали параметры *распределений* x_i и, как правило, по многим наблюдениям). Так как мы имеем лишь одно наблюдение X, то невозможно рассчитывать получить оценки, которые обладали бы свойством состоятельности (это видно и из содержания предыдущего раздела). Мы опять будем рассматривать асимптотическую постановку задачи.

Определение 3. Оценка θ^* называется *асимптотически однородной* (в среднеквадратичном), если

$$\mathbf{E}_t(\theta^* - t)^2 \to \sigma^2 \leqslant \infty \tag{11}$$

при $t \to \infty$ (и $n-t \to \infty$, если $n < \infty$). Класс таких оценок обозначим через $K_{(2)}$. Из теоремы 1 следует, что о. м. п. $\widehat{\theta}^* \in K_{(2)}$, при этом $\sigma^2 = \widehat{\sigma}^2 = \mathbf{E}D^2$, где D — случайная величина, определенная в приложении VII.

Определение 4. Оценка $\widetilde{\theta}^*$, определенная равенством

$$\widetilde{\theta}^* = \frac{\sum\limits_{k\geqslant 1} kR(k)}{\sum\limits_{k\geqslant 1} R(k)}, \tag{12}$$

называется *оценкой среднего правдоподобия*.

Ясно, что эта оценка не обязательно целочисленна. Наряду с $\widetilde{\theta}^*$ можно рассматривать эквивалентную ей в известном смысле рандомизированную целочисленную оценку $\widetilde{\theta}^{**}$, принимающую значение $[\widetilde{\theta}^*]$ с вероятностью $1-\{\widetilde{\theta}^*\}$ и значение $[\widetilde{\theta}^*]+1$ с вероятностью $\{\widetilde{\theta}^*\}$, где $[x]$ и $\{x\}$ — целая и дробная части x соответственно.

Введем в рассмотрение случайные величины

$$E_1=\sum_{k=1}^{\infty}e^{S_k},\quad E_2=\sum_{k=1}^{\infty}e^{Z_k},\quad E_1'=\sum_{k=1}^{\infty}ke^{S_k},\quad E_2'=\sum_{k=1}^{\infty}ke^{Z_k}.$$

Так как по усиленному закону больших чисел $S_k<-ck<0$, $Z_k<-ck<0$ с вероятностью 1 при некотором $c>0$ ($c<-\mathbf{E}\xi_1$, $c<-\mathbf{E}\zeta_1$) и при всех k, начиная с некоторого, то рассматриваемые ряды сходятся, а введенные случайные величины являются собственными (заметим, что при этом $\mathbf{E}e^{S_k}=\mathbf{E}e^{Z_k}=1$, $\mathbf{E}E_i=\infty$). Более детальные сведения о распределении E_j см. в приложении VII.

Теорема 4. *$\mathbf{P}_\theta$-распределение $\widetilde{\theta}^*-\theta$ слабо сходится при $\theta\to\infty$ ($и$ $n-\theta\to\infty$, если $n<\infty$, $n\to\infty$) к распределению* $\dfrac{E_2'-E_1'}{1+E_1+E_2}$. *Эта сходимость имеет место вместе со сходимостью моментов любого порядка,* $\widetilde{\sigma}^2=\mathbf{E}\left(\dfrac{E_2'-E_1'}{1+E_1+E_2}\right)^2<\infty$. *Оценка среднего правдоподобия $\widetilde{\theta}^*$ является асимптотически оптимальной в $K_{(2)}$: для любой оценки $\theta^*\in K_{(2)}$ с предельным разбросом σ^2* (см. (11)) *выполняется $\sigma^2\geqslant\widetilde{\sigma}^2$. Кроме того, для любого $Q\in\mathcal{Q}$ байесовская оценка θ_Q^* и оценка $\widetilde{\theta}^*$ асимптотически эквивалентны:*

$$\mathbf{P}_t\bigl(|\theta_Q^*-\widetilde{\theta}^*|>\varepsilon\bigr)\to 0 \tag{13}$$

при $t\to\infty$, $n-t\to\infty$.

Во избежание недоразумений напомним, что байесовский критерий θ_Q^*, фигурирующий в предыдущем разделе, и обозначаемая тем же символом байесовская оценка θ_Q^* определяются по-разному.

Доказательство. Докажем сначала, что $\widetilde{\theta}^*\in K_{(2)}$. Имеем

$$R(k)=\begin{cases}e^{-S_{k-1}} & \text{при}\quad k\leqslant\theta,\\ e^{-S_{\theta-1}+Z_{k-\theta}} & \text{при}\quad k\geqslant\theta.\end{cases}$$

Умножим числитель и знаменатель в (12) на $e^{S_{\theta-1}}$. Тогда

$$\widetilde{\theta}^* - \theta = \left(\sum_{k\leqslant\theta}(k-\theta)e^{S_{\theta-1}-S_{k-1}} + \sum_{k>\theta}(k-\theta)e^{Z_{k-\theta}}\right) \times$$

$$\times \left(\sum_{k\leqslant\theta} e^{S_{\theta-1}-S_{k-1}} + \sum_{k>\theta} e^{Z_{k-\theta}}\right)^{-1} \underset{d}{=} \frac{E'_{2,n-\theta} - E'_{1,\theta}}{1 + E_{2,n-\theta} + E_{1,\theta}},$$

где

$$E_{1,s} = \sum_{k=1}^{s-1} e^{S_k}, \quad E_{2,s} = \sum_{k=1}^{s+1} e^{Z_k}, \quad E'_{1,s} = \sum_{k=1}^{s-1} ke^{S_k}, \quad E'_{2,s} = \sum_{k=1}^{s+1} ke^{Z_k}.$$

Ясно, что $E'_{i,s} \to E'_i$, $i = 1, 2$. Очевидно далее, что

$$\left|\frac{E'_{1,\theta}}{1 + E_{1,\theta} + E_{2,n-\theta}} - \frac{E'_1}{1 + E_1 + E_2}\right| \xrightarrow[P]{} 0$$

при $\theta \to \infty$, $n - \theta \to \infty$. Кроме того, согласно лемме 1 из приложения VII момент любого фиксированного порядка величин

$$\frac{E'_{1,\theta}}{E_{1,\theta}} \geqslant \frac{E'_{1,\theta}}{1 + E_{1,\theta} + E_{2,n-\theta}}$$

равномерно по θ ограничен. Отсюда следует сходимость распределений $\dfrac{E'_{1,\theta}}{1 + E_{1,\theta} + E_{2,n-\theta}}$ с моментами любого порядка. Такое же утверждение справедливо для $\dfrac{E'_{2,n-\theta}}{1 + E_{1,\theta} + E_{2,n-\theta}}$. Это доказывает сходимость $\mathbf{P}_\theta$-распределения $\theta^* - \theta$ к распределению $\dfrac{E'_2 - E_1}{1 + E_1 + E_2}$ вместе с моментами любого порядка.

Рассмотрим теперь байесовскую постановку задачи и равномерное на $\{1, \ldots, n\}$ априорное распределение Q. Тогда, очевидно, что $\widetilde{\theta}^*$ совпадает с байесовской оценкой θ^*_Q. Обозначая через $\mathbf{E} = \mathbf{E}_Q$ математическое ожидание относительно распределения (θ, X), $\theta \in Q$, получим, что для любой оценки $\theta^* \in K_{(2)}$ с предельным разбросом σ^2 (см. (11)) справедливо

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}_t(\theta^* - t)^2 = \lim_{n\to\infty} \mathbf{E}(\theta^* - \theta)^2 \geqslant$$

$$\geqslant \lim_{n\to\infty} \mathbf{E}\big(\theta^*_Q - \theta\big)^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{E}_t(\widetilde{\theta}^2 - t)^2 = \widetilde{\sigma}^2.$$

Нам осталось убедиться, что байесовская оценка $\theta_Q^* = \mathbf{E}(\theta/X)$ для любого априорного распределения $Q \in \mathcal{Q}$ асимптотически эквивалентна $\widetilde{\theta}^*$, т. е. что выполнено (13). Имеем

$$\theta_Q^* - \widetilde{\theta}^* = \frac{\sum\limits_k (k-\theta)R(k)q(k)}{\sum\limits_k q(k)R(k)} - \frac{\sum\limits_k (k-\theta)R(k)}{\sum\limits_k R(k)}. \tag{14}$$

Из доказательства первой части теоремы мы видели, что второе слагаемое в (14) будет сколь угодно близким к $\dfrac{E'_{2,N} - E'_{1,N}}{1 + E_{1,N} + E_{2,N}}$ при достаточно большом N, т. е. близким к слагаемому того же вида, в котором, однако, суммирование в числителе и знаменателе ведется лишь по значениям k, $|k-\theta| \leqslant N$.

То же самое справедливо, очевидно, и для первого слагаемого в (14), и для их разности. Но для $Q \in \mathcal{Q}$ выполняется $\dfrac{q(k)}{q(\theta)} \to 1$ при $|k-\theta| \leqslant N$, $\theta \to \infty$, $n-\theta \to \infty$. Поэтому разность (14), в которой суммирование ведется по k, $|k-\theta| \leqslant N$, сходится к нулю при $\theta \to \infty$, $n-\theta \to \infty$. Другими словами, при $\theta = t$ $\mathbf{P}_t\bigl(|\theta_Q^* - \widetilde{\theta}^*| > \varepsilon\bigr) \to 0$ при $t \to \infty$, $n-t \to \infty$. Теорема доказана. $\triangleleft$

Утверждение, аналогичное теореме 4, можно получить и для целочисленной рандомизированной оценки $\widetilde{\theta}^{**}$. При этом под классом $K_{(2)}$ надо будет понимать класс асимптотически однородных рандомизированных целочисленных оценок,

$$\mathbf{E}_t(\widetilde{\theta}^{**} - \theta)^2 = \mathbf{E}_t(\widetilde{\theta}^* - \theta)^2 + \mathbf{E}_t(\widetilde{\theta}^* - \widetilde{\theta}^{**})^2$$

будет отличаться от $\mathbf{E}_t(\widetilde{\theta}^* - \theta)^2$ на величину

$$\mathbf{E}_t(\widetilde{\theta}^* - \widetilde{\theta}^{**})^2 = 2\mathbf{E}_t\{\widetilde{\theta}^*\}(1 - \{\widetilde{\theta}^*\}) \leqslant \frac{1}{2}.$$

Отметим также, что оценка среднего правдоподобия, являясь асимптотически байесовской и однородной ($\mathbf{E}_t(\widetilde{\theta}^* - t)^2$ асимптотически от t не зависит), будет также *асимптотически минимаксной оценкой* при соответствующем естественном определении этого понятия (ср. с § 21).

3. Последовательные процедуры. Последовательные процедуры представляют интерес в тех случаях, когда существует необходимость принять решение о наличии разладки и ее локализации в возможно короткие сроки (по минимальному числу наблюдений). Сокращение числа наблюдений естественно проводить с помощью следующей процедуры. Обозначим через $\widehat{\theta}_m^*$ наименьшее значение k, при котором $R(k) \geqslant R(j)$ при всех $j \leqslant m+1$, так что $\widehat{\theta}_n^* = \widehat{\theta}^*$, если мы располагаем n наблюдениями. В дальнейшем для определенности будем считать $n = \infty$. В качестве момента ν остановки наблюдений естественно выбирать такое m, при котором разность

$\ln R(\widehat{\theta}_m^*) - \ln R(m)$ достигнет достаточно большого значения, которое мы обозначим через $N + b(m)$. (Нетрудно видеть, что если $b(m) = o(m)$, то при $m - \theta \to \infty$ это произойдет с вероятностью 1. Если же $m < \theta$, то $\widehat{\theta}_m^*$ будет расти вместе с m и рассматриваемая разность будет оставаться «ограниченной».) Итак, мы положим

$$\nu = \min\{m\colon \ln R(\widehat{\theta}_m^*) - \ln R(m) \geqslant N + b(m)\}. \tag{15}$$

Технически, чтобы упростить вычисления и не запоминать все наблюдения, можно воспользоваться следующей эквивалентной процедурой. Рассмотрим последовательность $\{w_n\}$, определенную рекурсивными соотношениями

$$\begin{gathered} w_k = (w_{k-1} + \gamma_k)^+, \quad k \geqslant 1, \\ x^+ = \max(0, x), \quad \gamma_n = -\ln r_n = -\ln \frac{g_1(\mathrm{x}_n)}{g_2(\mathrm{x}_n)}. \end{gathered} \tag{16}$$

Основные свойства последовательности w_n мы сформулируем в виде леммы.

Лемма 1. *Уравнение* (16) *имеет решение*

$$w_n = \max\Big(w_0 + \Gamma_n, \max_{k \leqslant n}(\Gamma_n - \Gamma_k)\Big) = \Gamma_n + \max\big(w_0, \overline{\Gamma}_n^-\big),$$

где

$$\Gamma_n = \sum_{k=1}^{n} \gamma_k, \quad \overline{\Gamma}_n^- = \max_{k \leqslant n}(-\Gamma_k).$$

Если γ_k независимы и одинаково распределены, как γ, $\mathbf{E}\gamma < 0$, то распределение цепи Маркова w_n сходится к стационарному распределению величины $w = \overline{\Gamma}_\infty^- = \sup\limits_{k \geqslant 0}(-\Gamma_k)$.

Если уравнение $\varphi_\gamma(\lambda) = \mathbf{E}e^{\lambda\gamma} = 1$ имеет решение $\mu > 0$, то

$$\mathbf{P}(w > x) < e^{-\mu x}.$$

Если дополнительно $\varphi'_\gamma(\mu) < \infty$, то

$$\mathbf{P}(w > x) \sim ce^{-\mu x}, \quad 0 < c < 1.$$

Утверждение леммы вытекает из результатов глав 1, 4 в [12].

Из леммы при $w_0 = 0$ получаем, что $w_m = \ln R(\widehat{\theta}_m^*) - \ln R(m)$, а $\widehat{\theta}_m^*$ совпадает со значением $m - s(m)$, где $s(m)$ — наименьшее значение s, при котором $w_{m-s} = 0$.

Итак, наряду с (15) мы можем положить также

$$\nu = \nu(N) = \min\{m\colon w_m \geqslant N + b(m)\}. \tag{17}$$

Такой момент остановки называют моментом, полученным методом кумулятивных сумм (CUSUM). Этот метод в случае $b(m) \equiv 0$ был предложен в [71, 90]. Ясно, что $\mathbf{P}(\nu < \infty) = 1$, так как при $m > \theta$ в силу усиленного

закона больших чисел $w_{m+k} > ck$, $0 < c < -\mathbf{E}\zeta_1$, при всех достаточно больших k.

Если $k < \theta$, то распределение w_k совпадает с распределением $\overline{S}_{k-1} = \max\limits_{0\leqslant j\leqslant k-1} S_j$, которое монотонно сходится к распределению $S^+ = \sup\limits_{j\geqslant 0} S_j$.

Поэтому если $N + b(m) = \text{const}$, а θ велико, то событие $\{w_m \geqslant N + b(m)\}$ рано или поздно произойдет, так что вероятность $\mathbf{P}_\theta(\nu < \theta)$ близка к 1 и момент остановки ν с высокой вероятностью будет «ложным». Стало быть, для того чтобы избежать этого, мы должны выбирать растущие последовательности $b(m) \uparrow \infty$ (по возможности медленно растущие, чтобы $\nu - \theta$ было не очень велико). Пусть $b(m) \uparrow \infty$ так, что

$$b = \sum_{k=1}^{\infty} e^{-b(k)} \leqslant \infty.$$

(Можно взять, например, $b(k) = \ln k + (1+\beta)\ln\ln k$, $\beta > 0$ при $k \geqslant 3$; $b(1) = b(2) = 0$. Любая такая последовательность будет верхней для $\{w_k\}$ при $\theta = \infty$, а последовательность $b(k) = \ln k + (1-\beta)\ln\ln k$, $\beta > 0$, — нижней последовательностью.)

Свойства момента остановки $\nu = \nu(N)$ $(\{\nu \leqslant n\} \in \sigma(\mathrm{x}_1, \ldots, \mathrm{x}_n))$, определенного в (17), описываются следующим утверждением. Чтобы упростить рассуждения, мы будем предполагать, что распределение $-\xi_i$ непрерывно в области $(0, \infty)$. Каким образом обойти технические трудности, возникающие при нарушении этого условия, отражено в замечании 2 к теореме 8.

Обозначим $P(t) = P(t, N) = \mathbf{P}_t(\nu(N) < t)$.

Теорема 5. *Пусть* $\beta > 0$,

$$b(k) = \ln k + (1+\beta)\ln\ln(k), \quad b = 2 + \sum_{3}^{\infty} k^{-1}(\ln(k))^{-(1+\beta)}.$$

Тогда при всех t

$$P(t) \leqslant be^{-N}. \tag{18}$$

Если распределение $-\xi_i$ *непрерывно в* $(0, \infty)$, *то* $P(t)$ *и* $P(\infty)$ *непрерывно зависят от* N *и для любого* $\varepsilon < \varepsilon_0$ *существует* $N = N(\varepsilon)$, *являющееся решением уравнения* $P(\infty, N) = \varepsilon$, *так что* $P(t) \uparrow P(\infty) = \varepsilon$ *при* $t \to \infty$, *где* $\varepsilon_0 = \mathbf{P}_\infty\left(\bigcup\limits_{k=1}^{\infty}\{w_k \geqslant b(k)\}\right)$. *Кроме того, для нерешетчатых* ζ_i

$$\mathbf{E}_t\left(\nu - \frac{t}{\nu} \geqslant t\right) = \frac{\mathbf{E}_t(\nu - t)^+}{1 - P(t)} = \frac{1}{\rho}\big(b(t) + N - \mathbf{E}\max(S^+, Z^+) + \mathbf{E}_\chi\big) + o(1), \tag{19}$$

где $\rho = \rho - 1(G_2, G_1)$, S^+ *и* Z^+ *независимы,* χ *есть величина перескока через бесконечно удаленный барьер последовательности сумм* $-Z_k$. *Если* ζ_i *решетчаты, то слагаемые* $\dfrac{\mathbf{E}_\chi}{\rho} + o(1)$ *в правой части* (19) *следует заменить функцией, ограниченной известной постоянной.*

Доказательство. Оценка (18) вытекает из очевидного соотношения

$$P(t) = \mathbf{P}_t(\nu < t) = \mathbf{P}_t\left(\bigcup_{k=1}^{t-1}\{w_k \geqslant N + b(k)\}\right) \leqslant \sum_{k=1}^{t-1}\mathbf{P}(S^+ \geqslant N + b(k)) \leqslant$$

$$\leqslant e^{-N}\sum_{k=1}^{\infty} e^{-b(k)} = be^{-N}. \quad (20)$$

Далее, если распределение $-\xi_i$ непрерывно на $(0,\infty)$, то непрерывными будут и распределения w_k на $(0,\infty)$. Стало быть, $P(t)$ и $P(\infty)$ в силу (20) будут непрерывно зависеть от N, так как вероятность скачка в точке N оценивается суммой $\sum_{k=1}^{\infty}\mathbf{P}(w_k = N + b(k))$. Далее,

$$P(t) = \mathbf{P}_\infty\left(\bigcup_{k=1}^{t-1}\{w_k \geqslant N + b(k)\}\right) \uparrow P(\infty).$$

Таким образом, функция $P(\infty) = P(\infty, N)$ с ростом $N > 0$ монотонно и непрерывно убывает от ε_0 до 0. Положив $N(\varepsilon)$ равным решению уравнения $P(\infty, N) = \varepsilon$, мы получим утверждение теоремы, касающееся $N(\varepsilon)$.

Докажем теперь (19). В силу (16), (17), (20) и леммы 1 мы имеем представление

$$\mathbf{E}_t\left(\nu - \frac{t}{\nu} > t\right) = \mathbf{E}\eta,$$

где η есть марковский момент

$$\begin{aligned} \eta &= \min\{k\colon -Z_k \geqslant B(t,k)\}, \\ B(t,k) &= b(t+k) + N - \max(w_t, \overline{Z}_k), \end{aligned} \quad (21)$$

$\overline{Z}_k = \max\limits_{0<j<k} Z_j$, $w_t \underset{d}{=} \overline{S}_{t-1}$ и $\{Z_k\}$ независимы.

Пусть для простоты $\beta \leqslant 1$. Тогда при $t > \max\left(3, \frac{6}{\rho}\right)$ выполняется $b(t+k) \leqslant 3\ln(t+k) \leqslant 3\ln t + 3\frac{k}{t} \leqslant 3\ln t + \frac{k\rho}{2}$. Поэтому

$$\eta \leqslant \min\{k\colon -Z_k \geqslant b(t+k) + N\} \leqslant \eta_0 =$$

$$= \min\left\{k\colon -Z_k \geqslant 3\ln t + N + \frac{k\rho}{2}\right\}, \quad (22)$$

$$\mathbf{E}\eta \leqslant \mathbf{E}\eta_0 \leqslant \frac{6\ln t}{\rho} + \text{const}.$$

Кроме того, $\mathbf{E}|\zeta_k| < \infty$. Поэтому к мартингалу $-Z_k^0 = -Z_k - \rho k$ применима теорема Вальда–Дуба (см., например, [17, § 14.2]), в силу которой

$$\begin{gathered}\mathbf{E}(-Z_\eta - \rho\eta) = 0,\\ \mathbf{E}\eta = \frac{1}{\rho}\mathbf{E}[b(t+\eta) + N - \max(w_t, \overline{Z}_\eta) + \chi_t],\end{gathered} \tag{23}$$

где $\chi_t = -Z_\eta - B(t,\eta)$. Мы опускаем здесь доказательство того, что $\sup\limits_t \mathbf{E}\chi_t < \infty$ (для ограниченных ζ_i это очевидно), и того, что для нерешетчатых ζ_i выполняется $\mathbf{E}\chi_t \to \mathbf{E}\chi$. Доказательство этих утверждений можно найти в [23]. Далее, распределения w_t и $\overline{Z}_\eta$ сходятся вместе с экспоненциальными моментами соответственно к распределению S^+ и Z^+. Кроме того, в силу (22)

$$\mathbf{E}b(t+\eta) = \ln t + O\left(\frac{\ln t}{t}\right) = \ln t + O(1).$$

Изложенное вместе с (23) доказывает (19). Теорема доказана. $\triangleleft$

Если вернуться к исходной задаче оценивания θ с помощью последовательной процедуры, то из сказанного выше следует, что в качестве приближений для $\widehat{\theta}^*$ естественно взять значение

$$\widehat{\theta}^{**} = \nu - s(\nu),$$

а в качестве приближения для $\widetilde{\theta}^*$ — значение

$$\widetilde{\theta}^{**} = \frac{\sum\limits_{k=\nu-2s(\nu)}^{\nu} kR(k)}{\sum\limits_{k=\nu-2s(\nu)}^{\nu} R(k)}.$$

Теорема 6.

$$\begin{gathered}\widehat{\theta}^{**} \leqslant \widehat{\theta}^*,\\ \mathbf{P}_t(\widehat{\theta}^* > \widehat{\theta}^{**}) \leqslant (b + e^{-b(t)})e^{-N}.\end{gathered} \tag{24}$$

Кроме того, для любого $\varepsilon > 0$

$$\mathbf{P}_t(|\widehat{\theta}^* - \widehat{\theta}^{**}| > \varepsilon) \to 0 \tag{25}$$

при $N \to \infty$, $t \to \infty$.

Доказательство. Первое утверждение теоремы очевидно. Далее,

$$\mathbf{P}_t(\widehat{\theta}^{**} < \widehat{\theta}^*) = \mathbf{P}_t(\widehat{\theta}^{**} < \widehat{\theta}^*;\, \nu \geqslant t) + \mathbf{P}_t(\widehat{\theta}^{**} < \widehat{\theta}^*;\, \nu < t). \tag{26}$$

Событие $\{\widehat{\theta}^{**} < \widehat{\theta}^*;\, \nu \geqslant t\}$ означает, что в области $k \geqslant t$ траектория $\ln R(k)$ будет иметь по крайней мере еще один максимум, превосходящий $\widehat{\theta}^{**}$. Стало

быть, $\mathbf{P}_t$-вероятность этого события не превосходит $\mathbf{P}(Z > N + b(t)) \leqslant \leqslant e^{-N-b(t)}$.

Вторая вероятность в (26) не превосходит (см. (20))

$$\mathbf{P}_t(\nu < t) \leqslant be^{-N}.$$

Доказательство последнего утверждения теоремы (25) мы предоставляем читателю. Все необходимое для этого уже подготовлено. $\triangleleft$

Вернемся к теореме 5. Из нее следует, что при фиксированной вероятности ложной тревоги $\mathbf{P}_t(\nu < t) \approx \varepsilon$ среднее значение «задержки» с объявлением тревоги $\mathbf{E}_t(\nu - \theta/\nu \geqslant \theta)$ имеет при больших θ порядок $\dfrac{\ln t}{\rho}$. Оказывается, что в широких предположениях улучшить (уменьшить) величину этой задержки в классе произвольных моментов остановки τ ($\{\tau \leqslant n\} \in \in \sigma(\mathrm{x}, \ldots, \mathrm{x}_n)$) с той же вероятностью ложной тревоги $\mathbf{P}_t(\tau < t)$ невозможно.

Чтобы сформулировать более точное утверждение, введем в рассмотрение классы асимптотически «однородных» (ср. с определениями 2, 3) марковских моментов.

Определение 5. Момент остановки τ называется *асимптотически однородным*, если существует

$$\lim_{t\to\infty} \frac{\mathbf{E}_t(\tau - t/\tau \geqslant t)}{\ln t} = v, \quad 0 \leqslant v \leqslant \infty. \tag{27}$$

Класс таких моментов, для которых $\mathbf{P}_t(\nu < t) \to \varepsilon$ при $t \to \infty$, обозначим через $K_{(\varepsilon)}$.

Теорема 7. *Для любого момента остановки $\tau \in K_{(\varepsilon)}$ выполняется $v \geqslant 1/\rho$.*

Если не требовать существования предела (27), то из доказательства этой теоремы, приведенного в приложении VII, будет следовать, что

$$\limsup_{t\to\infty} \frac{\mathbf{E}_t(\tau - t/\tau \geqslant t)}{\ln t} \geqslant \frac{1}{\rho}.$$

Итак, из теорем 5 и 7 вытекает, что для момента остановки ν, определенного в (17), нижняя граница $1/\rho$ возможных значений v достигается. В этом смысле момент ν является *асимптотически оптимальным*. В приложении VII будет построен еще один асимптотически оптимальный момент остановки, который в ряде случаев может быть более предпочтительным.

Приведем теперь утверждение об оптимальном байесовском моменте остановки в задаче о разладке, на которое мы будем опираться при доказательстве теоремы 7 в приложении VII и которое представляет самостоятельный интерес.

Итак, пусть задано априорное распределение $Q = \{q(n) = \mathbf{P}(\theta = n)\}$. Обозначим $X_n = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$. Тогда определено апостериорное распределе-

ние

$$q(k/X_n) = \frac{q(k)f_k(X_n)}{\sum\limits_{i=1}^{\infty} q(i)f_i(X_n)} = \frac{q(k)R(k)}{\sum\limits_{i=1}^{n} q(i)R(i) + R(n+1)Q_n}, \tag{28}$$

где $f_k(X_n)$ и $R(k)$ определяются так же, как в (1), (2), $Q_n = \sum\limits_{i=n+1}^{\infty} q(i)$. Положим

$$Q(n/X_n) = \mathbf{P}(\theta \leqslant n/X_n) = \sum_{k=1}^{n} q(k/X_n) = \left(1 + \frac{R(n+1)Q_n}{\sum\limits_{i=1}^{n} q(i)R(i)}\right)^{-1}$$

и определим момент остановки

$$\tau(A) = \min\{n \geqslant 1\colon Q(n/X_n) \geqslant A\} = \min\left\{n \geqslant 1\colon \frac{\sum\limits_{i=1}^{n} q(i)R(i)}{R(n+1)Q_n} \geqslant \frac{A}{1-A}\right\}.$$

Вероятность ложной тревоги $\mathbf{P}_Q(\tau(A) < \theta)$ равна

$$\alpha(A) = \mathbf{P}_Q(\tau(A) < \theta) = \sum_{n=1}^{\infty} q(n)\mathbf{P}_n(\tau(A) < n). \tag{29}$$

Обозначим через $K_{(\varepsilon,Q)}$ класс всех моментов остановки τ, для которых $\mathbf{P}_Q(\tau < \theta) \leqslant \varepsilon$.

Теорема 8. *Пусть A_ε — решение уравнения $\alpha(A) = \varepsilon$. Тогда момент остановки $\tau(A_\varepsilon) \in K_{(\varepsilon,Q)}$ является оптимальным в $K_{(\varepsilon,Q)}$ в следующем смысле: для любого $\tau \in K_{(\varepsilon,Q)}$ выполняется*

$$\mathbf{E}_Q(\tau - \theta)^+ \geqslant \mathbf{E}_Q(\tau(A_\varepsilon) - \theta)^+. \tag{30}$$

Если $\mathbf{P}_Q(\tau < \theta) = \varepsilon$, то (30) эквивалентно неравенству

$$\mathbf{E}_Q\left(\tau - \frac{\theta}{\tau} \geqslant \theta\right) \geqslant \mathbf{E}_Q\left(\tau(A_\varepsilon) - \frac{\theta}{\tau(A_\varepsilon)} \geqslant \theta\right).$$

Доказательство этой теоремы для геометрического априорного распределения

$$q(n) = (1-p)^{n-1}p, \quad n = 1, 2, \ldots, \tag{31}$$

содержится в [138]. В этом случае мы и будем пользоваться этой теоремой для доказательства теоремы 7 в приложении VII, хотя ее утверждение остается справедливым для любого априорного распределения.

Замечание 1. Нетрудно видеть (ср. с теоремой 5), что

$$\mathbf{P}_t(\tau(A) < t) = \mathbf{P}_t\left(\bigcup_{k=1}^{t-1}\{Q(k/X_k) \geqslant A\}\right)$$

будет непрерывно зависеть от A, если распределения ξ_i (и, стало быть, $R(i)$) непрерывны. В этом случае $\alpha(A)$ в (29) также непрерывно зависит от A. Так как, кроме того, $\alpha(A)$ монотонно убывает по A,

$$\alpha(A) = \mathbf{P}(\tau(A) < \theta) = 1 - \mathbf{P}(\theta \leqslant \tau(A)) = \\ = 1 - \mathbf{E}\mathbf{P}\left(\theta \leqslant \frac{\tau(A)}{X_{\tau(A)}}\right) \leqslant 1 - A, \quad (32)$$

то для любого $\varepsilon \in (0,1)$ существует единственное решение A_ε уравнения $\alpha(A) = \varepsilon$. Отметим также, что если все $q(n)$ малы, то «перескок» через уровень A будет тоже мал и наряду с (32) будет справедливо приближенное равенство $\alpha(A) \approx 1 - A$. Например, если в (31) $p \to 0$, то $\alpha(A) \to 1 - A$, $A_\varepsilon \to 1 - \varepsilon$.

Замечание 2. Предположение о непрерывности распределения ξ_i (в области $(-\infty, 0)$) делалось нами также в теореме 5, чтобы обеспечить непрерывность функции $P(\infty, N)$. Эти предположения о непрерывности ξ_i в теоремах 5, 8 не носят принципиального характера и могут быть обойдены путем следующей рандомизации процедуры принятия решения. Пусть, например, мы находимся в условиях теоремы 5 и монотонно убывающая по N функция

$$P(\infty, N) = \mathbf{P}_\infty(\nu(N) < \infty) = \mathbf{P}\left(\bigcup_k \{w_k \geqslant b(k) + N\}\right)$$

такова, что уравнение $P(\infty, N) = \varepsilon$ для заданного фиксированного $\varepsilon > 0$ не имеет решения. Это означает, что найдется точка $N(\varepsilon)$, в которой $P(\infty, N)$ имеет скачок,

$$P(\infty, N(\varepsilon) - 0) = \varepsilon_1 > \varepsilon, \quad P(\infty, N(\varepsilon) + 0) = \varepsilon_2 < \varepsilon.$$

Введем случайную величину $\nu(N(\varepsilon), p)$, которая при первом попадании w_k в область $(b(k) + N(\varepsilon), \infty)$ ведет себя как $\nu(N)$. Если же $w_k = b(k) + N(\varepsilon)$, то «тревога» о разладке объявляется с вероятностью p независимо от X; с дополнительной вероятностью $1 - p$ испытания продолжаются по той же схеме: либо до первого попадания w_k в $(b(k) + N(\varepsilon), \infty)$ (и тогда объявляется тревога), либо до следующего попадания w_k в точку $b(k) + N(\varepsilon)$, и тогда повторяется процедура «случайного розыгрыша» решения, описанная выше. Очевидно, $\mathbf{P}_\infty(\nu(N(\varepsilon), 1) < \infty) = \mathbf{P}_\infty(\nu(N(\varepsilon)) < \infty) = \varepsilon_1 > \varepsilon$, $\mathbf{P}_\infty(\nu(N(\varepsilon), 0) < \infty) = \mathbf{P}_\infty(\nu(N(\varepsilon) + 0) < \infty) = \varepsilon_2 > \varepsilon$. Кроме того, $\mathbf{P}_\infty(\nu(N(\varepsilon), p) < \infty)$ непрерывным и монотонным образом зависит от p. Поэтому найдется $p(\varepsilon) \in (0,1)$ такое, что $\mathbf{P}_\infty(\nu(N(\varepsilon), p(\varepsilon)) < \infty) = \varepsilon$.

Далее, $\nu(N, p)$ есть вновь момент остановки, но уже относительно последовательности $(\mathrm{x}_1, \omega_1), \ldots, (\mathrm{x}_n, \omega_n)$, где ω_i принимают значения 1 с вероятностью p и 0 с вероятностью $1 - p$ независимо от X и других значений ω_k.

Нетрудно проверить, что все остальные рассмотрения в доказательстве теоремы 5 сохранят свою силу применительно к $\nu(N, p)$.

Аналогичные модификации можно осуществить и относительно момента остановки $\tau(A)$ в теореме 8, после которых уравнение $\alpha(A) = \varepsilon$ всегда будет иметь решение.

Таким образом, мы можем считать, что предположения о непрерывности ξ_i в теоремах 5, 8 не ограничивают общности.

Теоремы 2–8 (см. также теорему 2 из приложения VII) дают нам вид асимптотически оптимальных статистик, которые следует использовать в задачах о разладке при больших θ. Однако теорема 5 оставляет существенный произвол в выборе статистической процедуры. Речь идет о выборе взаимосвязанных компонент N и $b(k)$ криволинейной границы $b(k) + N$. Если θ очень велико, то, видимо, предпочтение следует отдавать «пологим» границам с функцией $b(k)$, рассмотренной в теореме 5. Если же θ не очень велико, то для таких границ слагаемое $N(\varepsilon)$ в оценке (19) может оказаться существенно больше $b(\theta)$. Поэтому для уменьшения $N(\varepsilon)$, возможно, следует выбирать более быстро растущую функцию $b(t)$, например $(1 + \beta) \ln t$ или даже βt. Отыскание оптимального соотношения между $b(t)$ и N представляет собой трудную задачу, так как явный вид зависимости $N(\varepsilon)$ от ε и функции $b(t)$ нам неизвестен.

Глава 6

ТЕОРЕТИКО-ИГРОВОЙ ПОДХОД К ЗАДАЧАМ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В § 73–75 вводятся понятия обычной и статистической игр, выделяются основные классы оптимальных стратегий.

В § 76, 77 рассмотрены методы отыскания оптимальных статистических решений.

Материал § 78–80 посвящен построению асимптотически оптимальных решающих правил.

§ 73. Предварительные замечания

В предыдущих главах было рассмотрено большое количество разнообразных статистических задач. Всех их объединяет то обстоятельство, что статистик на основании экспериментальных данных должен принять некоторое решение. В теории оценок эти решения могут иметь форму точечных оценок θ^*, которые мы должны принять в качестве неизвестного параметра θ, в теории проверки статистических гипотез — форму утверждений, в которых говорится, какие предположения относительно природы исследуемого объекта являются верными, а какие — нет. Эти решения, если они ошибочны, сопряжены с последующими потерями. Например, ошибка в лабораторной оценке (производимой с помощью выборки) содержания различных компонент в руде может привести к нарушению оптимального режима плавки и ухудшению качества выплавляемого металла. Это означает, что мы понесем материальные потери, которые будут зависеть от величины ошибки. Ошибка, касающаяся эффективности действия медицинского препарата, который проверяется на выборочной группе больных, очевидно, тоже может привести к потерям, которые мы для единообразия подхода также будем считать возможным исчислять в некоторых единицах. Такое же соглашение мы примем и относительно других задач статистики, в которых потери не носят ясно выраженный материальный характер.

Изложенное позволяет нам в задачах математической статистики выделить следующие четыре общих элемента. Эти элементы, собственно, и определяют сущность каждой конкретной задачи. Для простоты в последующем изложении мы будем снова, как и в гл. 1–3, иметь в виду лишь *задачи с одной выборкой X фиксированного объема n.* (Это не мешает успешно применять полученные результаты к выборкам объема 1 в задачах гл. 4, 5.) Эти элементы таковы.

1. *Множество Θ, элементы которого $\theta \in \Theta$ определяют состояние исследуемого объекта.* Если θ известно, то необходимость построения статистического решения отсутствует. Множество Θ называют также множеством параметров, хотя θ могут допускать и более широкое толкование (на-

пример, множество Θ может быть очень богатым и совпадать с множеством всех распределений на некотором пространстве $\mathcal{X}$).

2. Чтобы получить информацию о неизвестном θ, статистик ставит эксперимент и производит наблюдения над некоторой случайной величиной, распределение которой зависит от θ. Другими словами, статистик располагает выборкой X из распределения $\mathbf{P}_\theta$. Как мы уже знаем, из выборки X можно извлечь информацию о $\mathbf{P}_\theta$ и, стало быть, о θ. Можно считать, что выполнено условие (A_0) (см. § 16) о взаимно однозначном соответствии между θ и $\mathbf{P}_\theta$.

3. В задачах статистики всегда определено *множество* $D = \{\delta\}$ *решений*, которые может принимать статистик. В теории оценивания множество D обычно совпадает с Θ, в задачах проверки гипотез множество D конечно, а число его элементов равно числу проверяемых гипотез. Если θ известно, то решение $\delta = \varphi(\theta)$ определяется однозначно. Если θ неизвестно, то решение δ желательно выбрать в известном смысле оптимально. Но оптимизация решений требует, чтобы мы имели возможность их сравнивать. Для этого будем предполагать, что задана функция потерь, определяющая количественно последствия принятия решений.

4. *Функция потерь* $w(\delta, \theta)$ определена на $D \times \Theta$ и указывает, какие потери мы будем нести, если примем решение δ, а исследуемый объект, к которому относится решение, находится в состоянии θ. Будем считать, что $w(\delta, \theta) > 0$ при $\delta \neq \varphi(\theta)$, $w(\varphi(\theta), \theta) = 0$.

Если из перечисленных четырех элементов убрать п. 2 об экспериментальных данных, то мы получим объект, представляющий собой *обычную игру двух лиц*, игру, в которой роль первого игрока играет статистик (исследователь), а второго — природа.

§ 74. Основные понятия и теоремы, связанные с игрой двух лиц

1. Игра двух лиц

Определение 1. *Игрой двух лиц* называется тройка (D, Θ, w), составленная из множеств D и Θ и функции w, отображающей $D \times \Theta$ в полупрямую $[0, \infty)$. Элементы δ множества D называются *стратегиями* (действиями) I *игрока*, элементы $\theta \in \Theta$ — *стратегиями* II игрока, а w есть *функция потерь* I игрока (или функция выигрыша второго игрока), определяющая потери $w(\delta, \theta)$, которые будет нести I игрок, если он выберет стратегию δ, а II игрок — стратегию θ.

Основная задача теории игр двух лиц состоит в выборе оптимальной стратегии I игрока, которого мы часто будем отождествлять с собой. Для этого множество стратегий нужно как-то упорядочить. Сделать это непросто, так как потери $w(\delta, \theta)$, с помощью которых мы должны производить упорядочение, зависят от двух аргументов, так что для каждого θ стратегия δ, минимизирующая $w(\delta, \theta)$, будет, вообще говоря, своя.

Определение 2. Будем говорить, что стратегия δ_1 *лучше, чем* δ_2, если

$$w(\delta_1, \theta) \leqslant w(\delta_2, \theta) \quad \text{при всех} \quad \theta \in \Theta \tag{1}$$

и существует по крайней мере одно значение $\theta_1 \in \Theta$, при котором $w(\delta_1, \theta_1) < w(\delta_2, \theta_1)$.

Если выполнено лишь (1), то мы будем говорить, что стратегия δ_1 *не хуже, чем* δ_2.

Стратегию δ_0, для которой

$$w(\delta_0, \theta) \leqslant w(\delta, \theta) \quad \text{при всех} \quad \delta \text{ и } \theta,$$

будем называть *равномерно оптимальной* (или равномерно наилучшей).

Равномерно наилучшая стратегия обеспечивает наименьшие потери при всех θ. Однако, как правило, такие стратегии не существуют.

Отметим следующие три подхода к исследованию оптимальных стратегий I игрока:

— отыскание равномерно оптимальных стратегий в подклассах;

— отыскание байесовских и минимаксных стратегий;

— изучение совокупности всех неулучшаемых стратегий (так называемого полного класса стратегий).

2. Равномерно оптимальные стратегии в подклассах. Применительно к задачам математической статистики часто используется следующий прием (см. § 77). Из некоторых соображений, не связанных непосредственно с потерями (соображения симметрии, естественность процедуры, простота вычислений и др.), иногда оказывается возможным сузить класс рассматриваемых стратегий. Если это сужение оказывается таким, что после него существует равномерно оптимальная стратегия, то тем самым проблема выбора стратегии решается. Этот подход необходимо сопровождать исследованиями вопроса о том, не утратили ли мы путем сужения класса стратегий возможности получения существенно лучшего результата. Примеры использования такого подхода, относящиеся, правда, к более сложному объекту — статистическим играм, будут приведены в последующих двух параграфах. Читатель уже знает о них по гл. 2, 3, где рассматривались наилучшие (эффективные) оценки в подклассе несмещенных оценок и равномерно наиболее мощные критерии в подклассах всех инвариантных или несмещенных критериев.

3. Байесовские стратегии. Они возникают в тех случаях, когда второй игрок выбирает свою стратегию случайным образом, с некоторым распределением (известным или неизвестным) на Θ.

Чтобы иметь возможность рассматривать в дальнейшем «случайные» стратегии, мы будем предполагать, что на Θ и D выделены некоторые естественные σ-алгебры подмножеств $\mathfrak{F}_\Theta$ и $\mathfrak{F}_D$. Тогда на $(\Theta, \mathfrak{F}_\Theta)$, $(D, \mathfrak{F}_D)$ можно определить распределения $\mathbf{Q}$ и π соответственно, так что $(\Theta, \mathfrak{F}_\Theta, \mathbf{Q})$, $(D, \mathfrak{F}_D, \pi)$ будет вероятностными пространствами.

Задание распределений π и $\mathbf{Q}$ индуцирует вероятностное пространство $(D \times \Theta, \mathfrak{F}_{D\times\Theta}, \pi \times \mathbf{Q})$, где $\mathfrak{F}_{D\times\Theta}$ есть σ-алгебра, порожденная прямыми произведениями множеств из $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$. Выбор σ-алгебр $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$ должен быть таким, чтобы выполнялись следующие два условия.

1. $\mathfrak{F}_D$ *и* $\mathfrak{F}_\Theta$ *содержат одноточечные множества* $\{\delta\}$ *и* $\{\theta\}$.

2. *Функция потерь* $w(\delta, \theta)$ *измерима относительно* $\mathfrak{F}_{D\times\Theta}$.

Определение 3. Распределения π на $(D, \mathfrak{F}_D)$ и $\mathbf{Q}$ на $(\Theta, \mathfrak{F}_\Theta)$ мы будем называть *смешанными* или *рандомизированными стратегиями* соответственно I и II игроков.

Распределение $\mathbf{Q}$ мы часто будем называть *априорным*. Смысл этого термина должен быть ясен из гл. 2, 3. Он будет пояснен дополнительно в следующем параграфе.

Множества всех смешанных стратегий I и II игроков (т. е. множества всех распределений на $(D, \mathfrak{F}_D)$ и $(\Theta, \mathfrak{F}_\Theta)$) обозначим через $\widetilde{D}$ и $\widetilde{\Theta}$. Так как $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$ содержат одноточечные множества, то $\widetilde{D}$ и $\widetilde{\Theta}$ будут содержать распределения, сосредоточенные в одной точке, и, стало быть, мы можем считать, что $\widetilde{D}$ и $\widetilde{\Theta}$ содержат в себе стратегии δ и θ, которые, чтобы иметь возможность их выделить, мы будем называть *чистыми стратегиями*. Соглашение, по которому мы будем распределения из $\widetilde{D}$ и $\widetilde{\Theta}$, сосредоточенные в одной точке δ или θ, обозначать соответственно теми же символами δ и θ, нигде к недоразумениям не приведет.

Определим теперь потери $\widetilde{w}(\pi, \mathbf{Q})$ при использовании смешанных стратегий равенством

$$\widetilde{w}(\pi, \mathbf{Q}) = \mathbf{M}_{\pi \times Q} w(\delta, \theta) = \int w(u, t)\pi(du)\mathbf{Q}(dt). \tag{2}$$

Таким образом, наряду с исходной игрой мы можем рассматривать игру $(\widetilde{D}, \widetilde{\Theta}, \widetilde{w})$ с функцией потерь (2), которая называется *усреднением* или *рандомизацией игры* (D, Θ, w).

В соответствии с принятым соглашением мы будем писать

$$\widetilde{w}(\pi_{(\delta)}, \mathbf{Q}) = \widetilde{w}(\delta, \mathbf{Q}), \quad \widetilde{w}(\pi, \mathbf{Q}_{(\theta)}) = \widetilde{w}(\pi, \theta),$$
$$\widetilde{w}(\delta, \theta) = w(\delta, \theta),$$

если $\pi_{(\delta)}$ и $\mathbf{Q}_{(\theta)}$ есть распределения, сосредоточенные соответственно в точках δ и θ.

Очевидно, что рандомизация игры (D, Θ, w) означает переход к игре с более богатыми множествами стратегий, по отношению к которой исходная пара является «вложенной» — она получается, если рассматривать лишь чистые стратегии обоих игроков. Задачи упорядочения стратегий в играх (D, Θ, w) и $(\widetilde{D}, \widetilde{\Theta}, \widetilde{w})$, как мы увидим, тесно связаны.

Определение 4. Стратегия $\pi = \pi_Q$, для которой

$$\widetilde{w}(\pi_Q, \mathbf{Q}) = \inf_\pi \widetilde{w}(\pi, \mathbf{Q}),$$

называется *байесовской, соответствующей априорному распределению* $\mathbf{Q}$.

Таким образом, байесовская стратегия есть не что иное, как наилучшая стратегия π при данном $\mathbf{Q}$ в усредненной игре.

Стратегия $\delta_Q \in D$, для которой $\widetilde{w}(\delta_Q, \mathbf{Q}) = \inf_\pi \widetilde{w}(\pi, \mathbf{Q})$, называется *чистой байесовской*.

Теорема 1. *Если для данного* $\mathbf{Q}$ *существует смешанная байесовская стратегия* π_Q, *то существует и чистая байесовская стратегия* δ_Q *такая, что*

$$\widetilde{w}(\delta_Q, \mathbf{Q}) = \widetilde{w}(\pi_Q, \mathbf{Q}).$$

Доказательство почти очевидно. Обозначим $a = \widetilde{w}(\pi_Q, \mathbf{Q})$. Ясно, что

$$\widetilde{w}(\delta, \mathbf{Q}) \geqslant \inf_{\delta} \widetilde{w}(\gamma, \mathbf{Q}) \geqslant a.$$

Если допустить, что $\widetilde{w}(\delta, \mathbf{Q}) > a$ при всех δ, то, производя усреднение по δ с помощью π_Q, получим

$$a = \int \widetilde{w}(u, \mathbf{Q}) \pi_Q(du) > a.$$

Это противоречие доказывает теорему. $\triangleleft$

Таким образом, если $\inf\limits_{\pi} \widetilde{w}(\pi, \mathbf{Q})$ достигается, то он достигается и на чистых стратегиях.

Если $\inf\limits_{\delta} \widetilde{w}(\delta, \mathbf{Q})$ не достигается, то байесовские стратегии не существуют. В этом случае полезным оказывается понятие *ε-байесовской стратегии*, которая существует всегда и определяется как стратегия δ_Q, для которой

$$\widetilde{w}(\delta_Q, \mathbf{Q}) \leqslant \inf_{\delta} \widetilde{w}(\delta, \mathbf{Q}) + \varepsilon \tag{3}$$

для заданного $\varepsilon > 0$. Однако в дальнейшем для упрощения изложения мы будем обычно ограничиваться теми задачами, в которых байесовские стратегии существуют. В связи с этим везде, где это понадобится, мы будем считать существование байесовских стратегий обеспеченным. Это имеет место, например, во всех случаях, когда множество D конечно или когда D есть метрический компакт, а функция $\widetilde{w}(g, Q)$ непрерывна «снизу» по δ: $\liminf\limits_{\delta_k \to \delta_0} \widetilde{w}(\delta_k, Q) = \widetilde{w}(\delta_0, Q)$. Последнее свойство означает, что если $\delta_k \to \delta_0$ и $\widetilde{w}(\delta_k, Q) \downarrow$, то $\widetilde{w}(\delta_k, Q) \downarrow \widetilde{w}(\delta_0, Q)$. В нашем случае всегда найдется последовательность стратегий δ_k' такая, что $\widetilde{w}(\delta_k', Q) \downarrow \inf\limits_{\delta} \widetilde{w}(\delta, Q)$. Так как D — компакт, то из нее можно выделить подпоследовательность δ_k такую, что $\delta_k \to \delta_0 \in D$. В силу изложенного выше получаем $\widetilde{w}(\delta_k, Q) \downarrow$ $\downarrow \widetilde{w}(\delta_0, Q) = \inf\limits_{\delta} (\delta, Q)$. Это означает, что существует байесовская стратегия $\delta_Q = \delta_0$. Можно приводить также и другие условия, обеспечивающие существование байесовских стратегий, связанные, например, с выпуклостью функции $\widetilde{w}(\delta, Q)$.

Вопрос о практическом использовании байесовских стратегий является довольно тонким. Если наличие априорного распределения обусловлено некоторым реальным физическим механизмом, то этот подход бесспорен. Но байесовский подход можно оправдывать и в тех случаях, когда его связывают с наличием некоторых, быть может, даже субъективных и не всегда достаточно полных представлений, которые тем не менее сбрасывать со счета нельзя. Более подробное обсуждение вопроса об использовании байесовского подхода см. в п. 4.

4. Минимаксные стратегии. Если априорная информация относительно θ отсутствует, то при упорядочении стратегий можно ориентироваться на «наихудшую» стратегию противника. Если мы выберем стратегию δ, то макси-

мальные потери составят

$$\sup_{\theta} w(\delta, \theta) \equiv w(\delta, \uparrow). \tag{4}$$

Это количество зависит лишь от δ и так же, как и значения $w(\delta, \mathbf{Q})$, позволяет упорядочить δ.

Определение 5. Стратегия $\overline{\delta}$ называется *минимаксной*, если

$$w(\overline{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow) \equiv w^*. \tag{5}$$

Термин «минимаксный» образован из соединения наименований операций в правой части соотношения

$$w(\overline{\delta}, \uparrow) = \min_{\delta} \max_{\theta} w(\delta, \theta).$$

Очевидно, что минимаксные стратегии так же, как и байесовские, могут, вообще говоря, не существовать. В этом случае аналогично (3) можно ввести понятие *ε-минимаксной стратегии*. В последующих рассмотрениях будем исходить из того, что sup и inf в (4), (5) достигаются.

Так как при любом θ

$$w(\overline{\delta}, \theta) \leqslant w(\overline{\delta}, \uparrow) = w^*,$$

то минимаксная стратегия $\overline{\delta}$ характеризуется тем, что обеспечивает потери I игрока в размере, не большем чем w^*.

Определение 6. Значения

$$w^* = \inf_{\delta} w(\delta, \uparrow) \quad (w(\delta, \uparrow) = \sup_{\theta} w(\delta, \theta)),$$

$$w_* = \sup_{\Theta} w(\downarrow, \Theta) \quad (w(\downarrow, \theta) = \inf_{\delta} w(\delta, \theta))$$

называются соответственно *верхней* и *нижней ценой игры*. Если $w^* = w_*$, то говорят, что *существует цена игры*, равная общему значению w^* и w_*.

Из сказанного выше и соображений симметрии ясно, что II игрок, действуя аналогично первому и выбирая свою стратегию $\overline{\theta}$ из тех же минимаксных соображений, может всегда обеспечить себе выигрыш не меньше, чем w_*. (Такую стратегию $\overline{\theta}$ было бы правильнее называть максиминной, но мы будем использовать для нее тот же термин: минимаксная стратегия.) Стало быть, *если существует цена игры*, то, выбирая минимаксную стратегию $\overline{\delta}$, мы обеспечим себе *неулучшаемый результат* в том смысле, что если противник выберет $\overline{\theta}$, то никакая другая стратегия не даст нам потери, меньшие, чем $w_* = w^*$. Очевидно, что $w(\overline{\delta}, \overline{\theta}) = w^* = w_*$.

В общем случае всегда $w^* \geqslant w_*$, так как при всех δ и θ $w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant$ $\geqslant w(\downarrow, \theta)$ и, следовательно,

$$w^* = \inf_{\delta} w(\delta, \uparrow) \geqslant \sup_{\theta} w(\downarrow, \theta) = w_*. \tag{6}$$

Если $w^* > w_*$, то минимаксную стратегию $\overline{\delta}$ *можно улучшить*, вводя смешанные стратегии. В этом состоит одно из главных назначений последних.

Минимаксные стратегии для усредненной игры (если они существуют) обозначим соответственно символами $\overline{\pi}$ и $\overline{\mathbf{Q}}$ и положим

$$\widetilde{w}^* = \inf_{\pi} \sup_{\mathbf{Q}} \widetilde{w}(\pi, \mathbf{Q}), \quad \widetilde{w}_* = \sup_{\mathbf{Q}} \inf_{\pi} \widetilde{w}(\pi, \mathbf{Q}).$$

Покажем сначала, что при усреднении игры верхняя и нижняя цены игры сближаются.

Теорема 2. *Всегда справедливо*

$$w^* \geqslant \widetilde{w}^* \geqslant \widetilde{w}_* \geqslant w_*.$$

Доказательство этой теоремы, как и теоремы 1, очень просто. Так как усреднение игры можно производить в два этапа: сначала по множеству D, а затем по Θ, то для доказательства достаточно рассмотреть лишь частичное усреднение $(\widetilde{D}, \Theta, \widetilde{w})$ игры (D, Θ, w). Имеем

$$\widetilde{w}^* = \inf_{\pi} \sup_{\theta} \widetilde{w}(\pi, \theta) \leqslant \inf_{\delta} \sup_{\theta} w(\delta, \theta) = w^*.$$

Так как при всех π

$$\widetilde{w}(\pi, \theta) = \int w(u, \theta)\pi(du) \geqslant \inf_{\delta} w(\delta, \theta) = w(\downarrow, \theta),$$

то $\inf\limits_{\pi} \widetilde{w}(\pi, \theta) \geqslant w(\downarrow, \theta)$,

$$\widetilde{w}_* = \sup_{\theta} \inf_{\pi} \widetilde{w}(\pi, \theta) \geqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

Неравенство $\widetilde{w}^* \geqslant \widetilde{w}_*$ доказано в (6). $\triangleleft$

Фундаментальным фактом теории игр является так называемая теорема о минимаксе, которая утверждает, что при весьма широких предположениях *усредненные игры имеют цену* $\widetilde{w}^* = \widetilde{w}_*$ *и для них существуют минимаксные стратегии.*

Более точно это утверждение будет сформулировано в следующем параграфе в более общей ситуации применительно к статистическим играм.

Исходная игра (D, Θ, w), особенно в случае, когда D и Θ конечны, цены, как правило, не имеет.

Пример 1. Рассмотрим простейшую игру, когда множества D и Θ двухточечные, $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$. Значения функции потерь $w(\delta, \theta)$ определяются матрицей $\|w(\delta_i, \theta_j)\|$, $i, j = 1, 2$, которую положим равной $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$. Это соответствует, например, игре с угадыванием, когда I игрок угадывает, в какой руке II игрок спрятал монету. Угадывание означает нулевой

проигрыш ($w(\delta_1, \theta_1) = w(\delta_2, \theta_2) = 0$), ошибка — проигрыш в размере 1 рубля ($w(\delta_1, \theta_2) = w(\delta_2, \theta_1) = 1$). Очевидно, что здесь $w(\delta_i, \uparrow) = 1$, $w^* = 1$, $w(\downarrow, \theta_i) = 0$, $w_* = 0$, так что игра не имеет цены, а I игрок не может себе гарантировать проигрыш меньше чем 1 рубль. Само понятие минимаксной стратегии здесь бесполезно.

Рассмотрим теперь усреднение этой игры. Здесь классы стратегий $\widetilde{D}$ и $\widetilde{\Theta}$ представляют собой совокупность всех распределений на двухточечном множестве. Очевидно, каждое из распределений на D и Θ описывается одной вероятностью p и q соответственно выбора стратегий δ_1 и θ_1. Поэтому можно считать, что $\widetilde{D} = [0, 1]$, $\widetilde{\Theta} = [0, 1]$. Потери I игрока в этой игре равны

$$\widetilde{w}(p, q) = p(1 - q) + q(1 - p) = p + q - 2pq,$$

$$\widetilde{w}(p, \uparrow) = \begin{cases} p + 1 - 2p = 1 - p & \text{при} \quad 2p < 1, \\ p & \text{при} \quad 2p \geqslant 1, \end{cases}$$

$$\widetilde{w}^* = 1/2.$$

Аналогично находим, что $\widetilde{w}_* = 1/2$. Таким образом, усредненная игра уже имеет цену, и первый игрок, выбирая δ_1 и δ_2 с вероятностью $p = 1 - p = 1/2$, может гарантировать себе проигрыш не больше $1/2$. Улучшить эту стратегию нельзя, так как такой же выигрыш может гарантировать себе II игрок, выбирая $q = 1/2$. Таким образом, стратегия $p = 1/2$ будет минимаксной, соответствующей стратегии $q = 1/2$ второго игрока.

Если все же окажется, что усредненная игра не имеет цены (это может быть только в играх специальной сложной конструкции), то повторное усреднение не даст никаких результатов, поскольку это повторное усреднение по существу будет совпадать с одинарным.

Байесовский и минимаксный подходы к решению игровых проблем весьма распространены в повседневной человеческой деятельности. Байесовский подход ориентирован на наличие некоторых представлений, хотя бы приближенных, о поведении второго игрока. Минимаксный подход оправдан в тех случаях, когда необходимо гарантировать себя от большого проигрыша.

Пример 2. Студент готовится к экзамену. Будем считать, что это не идеальный студент и что у него не оказалось достаточно времени для того, чтобы хорошо подготовить к сдаче весь материал. Кроме того, целью этого студента является получение возможно более высокой оценки.

В описанных условиях студент может выучить на отлично лишь часть материала. Поэтому для него возможны по крайней мере два пути: 1) выучить на отлично лишь те разделы, которые по имеющимся сведениям чаще всего спрашивает экзаменатор; 2) выучить все понемногу, чтобы гарантировать себе хорошую или удовлетворительную оценку. Первый вариант будет соответствовать байесовскому, второй — минимаксному подходу.

Ясно, что равномерно оптимальной стратегией здесь было бы выучить весь материал, но по условию задачи она невозможна.

Минимаксные стратегии в конкретных ситуациях выглядят не всегда разумными.

Пример 3. Пусть $\Theta = [0,1]$, а множество $D = \{\delta_1, \delta_2\}$ состоит из двух элементов. Функция потерь определяется соотношениями (см. рис. 10)

$$w(\delta_1, \theta) \equiv 1,$$

$$w(\delta_2, \theta) = 4(1+\varepsilon)\theta(1-\theta).$$

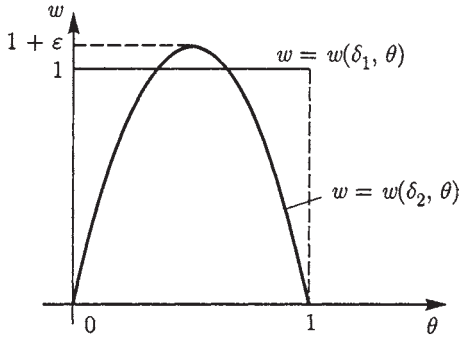

Рис. 10

Здесь $w(\delta_1, \uparrow) = 1$, $w(\delta_2, \uparrow) = 1+\varepsilon$, $w^* = 1$, и минимаксной стратегией будет δ_1, хотя при малых $\varepsilon > 0$ для «большинства» значений θ стратегия δ_2 будет лучше: $w(\delta_2, \theta) < 1$ при θ из области $\left|\theta - \frac{1}{2}\right| > \frac{1}{2}\sqrt{\frac{\varepsilon}{1+\varepsilon}}$. Для «большинства» распределений $\mathbf{Q}$ на $\Theta = [0,1]$ (чья масса не сосредоточена в окрестности точки $\theta = 1/2$) байесовские стратегии также будут совпадать с δ_2.

Понятия байесовской и минимаксной стратегии тесно связаны между собой. Следующее утверждение дает способ отыскания минимаксных стратегий с помощью байесовских.

Определение 7. Стратегия $\overline{\pi}$ называется *уравнивающей* на множестве $\Theta_0 \subset \Theta$, если

1) $\widetilde{w}(\overline{\pi}, \theta) = c = \text{const}$ при $\theta \in \Theta_0$,

2) $\widetilde{w}(\overline{\pi}, \theta) \leqslant c$ при всех θ.

Теорема 3. *Пусть существуют априорное распределение $\overline{\mathbf{Q}}$ и соответствующая ему байесовская стратегия $\pi_{\overline{\mathbf{Q}}}$, которая является уравнивающей на носителе $N_{\overline{\mathbf{Q}}}$ распределения $\overline{\mathbf{Q}}$. Тогда $\overline{\pi} = \pi_{\overline{\mathbf{Q}}}$ является минимаксной стратегией.*

Если $N_{\overline{\mathbf{Q}}} = \Theta$, то уравнивающая стратегия $\overline{\pi}$ делает игру второго игрока «безразличной» — от него не зависящей (ср. с примером 1).

Доказательство теоремы 3. Обозначим

$$\sup_{\theta} \widetilde{w}(\pi, \theta) = \sup_{\mathbf{Q}} \widetilde{w}(\pi, \mathbf{Q}) = \widetilde{w}(\pi, \uparrow),$$

$$\inf_{\delta} \widetilde{w}(\delta, \mathbf{Q}) = \inf_{\pi} \widetilde{w}(\pi, \mathbf{Q}) = \widetilde{w}(\downarrow, \mathbf{Q}).$$

Нам надо убедиться, что

$$\widetilde{w}(\pi_{\overline{\mathbf{Q}}}, \uparrow) = \inf_{\pi} \widetilde{w}(\pi, \uparrow).$$

Это вытекает из следующих неравенств, справедливых для любого π:

$$\widetilde{w}(\pi, \uparrow) \geqslant \widetilde{w}(\pi, \overline{\mathbf{Q}}) \geqslant \widetilde{w}(\pi_{\overline{\mathbf{Q}}}, \overline{\mathbf{Q}}) = \int \widetilde{w}(\pi_{\overline{\mathbf{Q}}}, t)\overline{\mathbf{Q}}(dt) = c \geqslant \widetilde{w}(\pi_{\overline{\mathbf{Q}}}, \uparrow). \quad \triangleleft$$

Иногда бывает полезным следующее небольшое обобщение теоремы 3.

Теорема 3А. *Пусть существуют последовательности* $\mathbf{Q}_n$, $\pi_{\overline{\mathbf{Q}}_n}$ *такие, что* $\widetilde{w}(\pi_{\overline{\mathbf{Q}}_n}, \mathbf{Q}_n) \to c$. *Пусть, кроме того, существует стратегия* $\overline{\pi}$, *обладающая свойством* $\widetilde{w}(\overline{\pi}, \theta) \leqslant c$ *при всех* θ. *Тогда* $\overline{\pi}$ — *минимаксная стратегия.*

Доказательство столь же просто:

$$\widetilde{w}(\pi, \uparrow) \geqslant \widetilde{w}(\pi, \mathbf{Q}_n) \geqslant \widetilde{w}(\pi_{\mathbf{Q}_n}, \mathbf{Q}_n) \to c.$$

Это может быть только тогда, когда $\inf\limits_{\pi} \widetilde{w}(\pi, \uparrow) \geqslant c$. Так как $c \geqslant \widetilde{w}(\overline{\pi}, \uparrow)$, то теорема доказана. $\triangleleft$

Распределение $\overline{\mathbf{Q}}$ в теореме 3, определяющее байесовскую минимаксную стратегию $\pi_{\overline{\mathbf{Q}}}$, обладает одним замечательным свойством: оно будет наихудшим в том смысле, что для него байесовские потери $\widetilde{w}(\pi_{\mathbf{Q}}, \mathbf{Q})$ будут максимальными.

Определение 8. Распределение $\overline{\mathbf{Q}}$ называется *наименее благоприятным* или *наихудшим*, если

$$\widetilde{w}(\pi_{\overline{\mathbf{Q}}}, \overline{\mathbf{Q}}) = \sup_{\mathbf{Q}} \widetilde{w}(\pi_{\mathbf{Q}}, \mathbf{Q})$$

или, другими словами, $\widetilde{w}(\downarrow, \overline{\mathbf{Q}}) = \sup\limits_{\mathbf{Q}} \widetilde{w}(\downarrow, \mathbf{Q})$.

Последнее соотношение означает, что определение наихудшего распределения в наших условиях совпадает с определением минимаксной (максиминной) стратегии $\overline{\mathbf{Q}}$.

Теорема 4. *Пусть игра* $(\widetilde{D}, \widetilde{\Theta}, \widetilde{w})$ *имеет цену, а оба игрока имеют минимаксные стратегии* $\overline{\pi}$ *и* $\overline{\mathbf{Q}}$. *Тогда распределение* $\overline{\mathbf{Q}}$ *является наихудшим, а* $\overline{\pi}$ *является байесовской стратегией* $\overline{\pi} = \pi_{\overline{\mathbf{Q}}}$, *отвечающей* $\overline{\mathbf{Q}}$.

Замечание 1. Из того, что в силу теоремы 1 наряду с $\pi_{\overline{\mathbf{Q}}}$ существует чистая байесовская стратегия $\delta_{\overline{\mathbf{Q}}}$, вовсе не следует, что последняя также будет минимаксной.

Замечание 2. В силу фундаментальной теоремы о минимаксах условие теоремы 4 о существовании цены усредненной игры и минимаксных стратегий не следует рассматривать как существенное ограничение.

Замечание 3. Из соображений симметрии и теоремы 4 следует, что распределение $\overline{\pi}$ также является «наихудшим», а $\overline{\mathbf{Q}}$ является байесовской стратегией, отвечающей $\overline{\pi}$.

Нам понадобится следующее вспомогательное утверждение, которое мы сформулируем в терминах исходной (не усредненной) игры.

Лемма 1. *Пусть игра* (D, Θ, w) *имеет цену и минимаксные стратегии* $\overline{\delta}$, $\overline{\theta}$ *обоих игроков*

$$w(\overline{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow), \quad w(\downarrow, \overline{\theta}) = \sup_{\theta} w(\downarrow, \theta).$$

Тогда

$$w(\overline{\delta},\uparrow) = w(\overline{\delta},\overline{\theta}) = w(\downarrow,\overline{\theta}), \tag{7}$$

$$w^* = w(\overline{\delta},\overline{\theta}) = w_*. \tag{8}$$

Обратно, если при некоторых $\overline{\delta}$, $\overline{\theta}$ *выполнено* (7), *то справедливо* (8), *а* $\overline{\delta}$, $\overline{\theta}$ *являются минимаксными стратегиями.*

Доказательство. При всех δ и θ имеем

$$w(\delta,\uparrow) \geqslant w(\delta,\theta) \geqslant w(\uparrow,\theta).$$

Отсюда следует, что

$$w^* = w(\overline{\delta},\uparrow) \geqslant w(\overline{\delta},\overline{\theta}) \geqslant w(\downarrow,\overline{\theta}) = w_*. \tag{9}$$

Так как по условию $w^* = w_*$, то в (9) все знаки неравенства надо заменить равенствами. Это доказывает (7), (8).

Обратно, если справедливо (7), то

$$w^* = \inf_{\delta} w(\delta,\uparrow) \leqslant w(\overline{\delta},\uparrow) = w(\downarrow,\overline{\theta}) \leqslant \sup_{\theta} w(\downarrow,\theta) = w_*.$$

Так как всегда $w^* \geqslant w_*$, то приведенные неравенства означают, что $w^* = w_*$, а стратегии $\overline{\delta}$, $\overline{\theta}$ являются минимаксными. Лемма доказана. ◁

Точка $(\overline{\delta},\overline{\theta})$, обладающая свойством (7), называется *седловой точкой*, а лемма 1 — критерием наличия седловой точки неулучшаемых минимаксных стратегий.

Доказательство теоремы 4. Тот факт, что $\overline{\mathbf{Q}}$ является наихудшим распределением, следует непосредственно из определения 8. Применим теперь лемму 1 к усредненной игре $(\widetilde{D},\widetilde{\Theta},\widetilde{w})$. Получим тогда, что

$$\widetilde{w}(\overline{\pi},\overline{\mathbf{Q}}) = \widetilde{w}(\downarrow,\overline{\mathbf{Q}}).$$

Отсюда следует, что $\overline{\pi}$ есть байесовская стратегия, соответствующая $\overline{\mathbf{Q}}$. Теорема доказана. ◁

Из леммы 1 вытекает также следующий критерий минимаксности стратегий в игре $(\widetilde{D},\widetilde{\Theta},\widetilde{w})$.

Теорема 5. *Пусть стратегии* $\overline{\pi}$, $\overline{\mathbf{Q}}$ *таковы, что* $\overline{\pi}$ *является байесовской относительно* $\overline{\mathbf{Q}}$: $\overline{\pi} = \pi_{\overline{\mathbf{Q}}}$; $\overline{\mathbf{Q}}$ *является оптимальной относительно* $\overline{\pi}$: $\overline{\mathbf{Q}} = \mathbf{Q}_{\overline{\pi}}$. *Тогда* $\overline{\pi}$ *и* $\overline{\mathbf{Q}}$ — *минимаксные стратегии,*

$$\widetilde{w}_* = \widetilde{w}(\overline{\pi},\overline{\mathbf{Q}}) = \widetilde{w}^*.$$

Доказательство немедленно следует из леммы 1, так как $\overline{\pi}$, $\overline{\mathbf{Q}}$ удовлетворяют соотношениям

$$\widetilde{w}(\overline{\pi},\overline{\mathbf{Q}}) = \widetilde{w}(\pi_{\overline{\mathbf{Q}}},\overline{\mathbf{Q}}) = \widetilde{w}(\downarrow,\overline{\mathbf{Q}}), \quad \widetilde{w}(\overline{\pi},\overline{\mathbf{Q}}) = \widetilde{w}(\overline{\pi},\mathbf{Q}_{\overline{\pi}}) = \widetilde{w}(\overline{\pi},\uparrow),$$

которые влекут за собой (7). ◁

Содержание приведенных утверждений можно резюмировать (с некоторыми дополнениями) в виде следующего предложения, которое весьма полно описывает связь минимаксных и байесовских стратегий. Введем в рассмотрение следующие условия.

А. *Игра* $(\widetilde{D}, \widetilde{\Theta}, \widetilde{w})$ *имеет цену и минимаксные стратегии.*

1. *Стратегия* $\overline{\pi}$ *является минимаксной.*

2. *Стратегия* $\overline{\pi}$ *является байесовской и уравнивающей.*

3. *Стратегия* $\overline{\pi}$ *является байесовской и соответствует наихудшему распределению* $\overline{\mathbf{Q}}$: $\overline{\pi} = \pi_{\overline{\mathbf{Q}}}$. *Стратегия* $\overline{\mathbf{Q}}$ *является байесовской по отношению к* $\overline{\pi}$.

Теорема 6. *Справедливы следующие импликации:*

$$1, \mathrm{A} \Rightarrow 3; \quad 3 \Rightarrow 2, \mathrm{A}; \quad 3 \Rightarrow 1, \mathrm{A}, \quad 2 \Rightarrow 1.$$

Отсюда следует, что при выполнении А *условия* 1–3 *эквивалентны.*

Доказательство. Первое, третье и четвертое соотношения доказаны соответственно в теоремах 4, 5 и 3 (см. также замечание 3). Оставшаяся вторая импликация доказана в теореме 5 лишь частично. Чтобы показать, что справедливо $3 \Rightarrow 2$, надо доказать, что байесовская стратегия, соответствующая наихудшему распределению, является уравнивающей. Так как $\overline{\pi}$ в условии 3 в силу теоремы 5 является минимаксной стратегией, то мы будем иметь

$$\widetilde{w}_* = \widetilde{w}(\overline{\pi}, \overline{\mathbf{Q}}) = \int \widetilde{w}(\overline{\pi}, t)\overline{\mathbf{Q}}(dt) \leqslant \sup_t \widetilde{w}(\overline{\pi}, t) = \widetilde{w}(\overline{\pi}, \uparrow) = \inf \widetilde{w}(\pi, \uparrow) = \widetilde{w}^*.$$

Это означает, что $\int \widetilde{w}(\overline{\pi}, t)\overline{\mathbf{Q}}(dt) = \sup_t \widetilde{w}(\overline{\pi}, t)$ и, стало быть,

$$\widetilde{w}(\overline{\pi}, t) = \widetilde{w}(\overline{\pi}, \uparrow) \quad \text{п. в.} \quad [\overline{\mathbf{Q}}].$$

Так как, кроме того, всегда $\widetilde{w}(\overline{\pi}, t) \leqslant \widetilde{w}(\overline{\pi}, \uparrow)$, то $\overline{\pi}$ является уравнивающей стратегией. Теорема доказана. $\triangleleft$

Вернемся к вопросу о применении рассматриваемых классов стратегий. Допустим, что мы не можем выделить удовлетворяющий нас подкласс стратегий, среди которых существовала бы равномерно наилучшая. Допустим, далее, что мы располагаем некоторыми представлениями о поведении второго игрока (т. е. об ожидаемых значениях θ), недостаточными, однако, для применения байесовского подхода в его чистом виде. Минимаксный подход в этих условиях будет означать пренебрежение имеющейся у нас информацией. В такой ситуации можно использовать промежуточный подход, который выглядит следующим образом.

1. Сначала нужно оградить себя от высоких потерь, т. е. рассматривать лишь те стратегии δ, для которых $w(\delta, \theta) \leqslant w^* + a$ при подходящем значении $a > 0$ и при всех θ. Множество стратегий, удовлетворяющих этому неравенству, обозначим символом D_a.

2. В этом подмножестве (т. е. в игре (D_a, Θ, w)) уже можно применять байесовский подход с использованием доступных нам приближений для априорного распределения $\mathbf{Q}$.

Такой смешанный подход также постоянно используется в повседневной человеческой деятельности. В условиях примера 2 этот подход будет означать, что студент совсем немного (чтобы избежать неудовлетворительной оценки) выучит весь материал, а затем получше выучит то, что чаще спрашивают.

Математически использование смешанного подхода должно сопровождаться исследованиями устойчивости байесовских потерь в игре (D_a, Θ, w) при допустимых изменениях $\mathbf{Q}$.

5. Полный класс стратегий. Если все описанные выше подходы не дают возможности однозначно выбрать стратегию, то решение задачи ограничивают описанием так называемого полного класса стратегий.

Определение 9. Класс стратегий $D^c \subset \widetilde{D}$ называется *полным*, если для любого $\pi \notin D^c$ существует стратегия $\pi_0 \in D^c$, которая лучше, чем π.

Класс D_0^c называется *минимальным полным классом*, если D_0^c есть полный класс, но никакой его собственный подкласс полным классом не является.

Другими словами, минимальный полный класс состоит из одних неулучшаемых стратегий.

Польза построения минимального полного класса или полного класса, который существенно меньше, чем D, очевидна. Это позволяет редуцировать игру $(\widetilde{D}, \widetilde{\Theta}, \widetilde{w})$ к игре $(D^c, \widetilde{\Theta}, \widetilde{w})$, которая может иметь более простую структуру.

Вторая фундаментальная теорема теории игр состоит в том, что при широких предположениях *класс всех байесовских стратегий* $\{\pi_{\mathbf{Q}}\}$, $\mathbf{Q} \in \widetilde{\Theta}$, *является полным классом*. Точная формулировка этой теоремы будет приведена в следующем параграфе. В некоторых случаях полные классы можно строить и непосредственно, используя структуру игры. Допустим, например, что существует разбиение пространства D на подмножества D_b, $D = \bigcup\limits_{b \in B} D_b$, $D_{b_1} \neq D_{b_2}$ при $b_1 \neq b_2$ такое, что в каждом из этих подмножеств (т. е. для игр (D_b, Θ, w)) существует равномерно оптимальная стратегия $\delta_b \in D_b$. Ясно, что в этом случае класс $D^c = \{\delta_b\}_{b \in B}$ будет полным. Такой подход к построению полного класса будет проиллюстрирован в § 75.

§ 75. Статистические игры

1. Описание статистических игр. Основные элементы статистической игры образуются той же тройкой (D, Θ, w), которая рассматривалась нами в предыдущем параграфе. Однако к ним добавляется следующее.

1. В статистических играх роль I игрока играет *статистик* (исследователь), роль II игрока — *природа* (точнее, природа того явления, которое исследуется). Последняя выбирает (или «загадывает») параметр (стратегию) θ, который нам неизвестен и который определяет состояние исследуемого объекта. Большинство задач математической статистики так или иначе связано с принятием таких решений δ, которое как можно более точно «угадывали» бы это неизвестное θ. При этом следует иметь в виду, что природа как игрок не стремится к максимальному выигрышу (т. е. не стремится причинить

нам максимальные потери) и в этом смысле является игроком, «беспристрастным» к выбору своих стратегий.

2. В статистических играх мы имеем возможность «разведывать» стратегию природы с помощью экспериментов, которые дают нам в виде выборки $X \in \mathbf{P}_\theta$ «наводящие» указания на то, каким может быть значение θ. Итак, элементом статистической игры является выборка X объема*) n из распределения $\mathbf{P}_\theta$, зависящего от θ.

В этих условиях наше решение δ мы должны выбирать, очевидно, в зависимости от X. Стало быть, стратегиями статистика становятся теперь все функции $\delta(X)$, отображающие $\mathcal{X}^n$ в D. Эти функции $\delta(X)$ называются *решающими функциями* или *решающими правилами*, или просто *решениями*. Мы ограничимся рассмотрением лишь функций $\delta(X)$, осуществляющих *измеримое* отображение $(\mathcal{X}^n, \mathfrak{B}_\mathcal{X}^n)$ в $(D, \mathfrak{F}_D)$. Множество всех таких функций обозначим через $\mathcal{D}$.

Множество стратегий II игрока (природы) Θ остается прежним.

Если мы воспользуемся решением $\delta(X)$, а природа выберет θ, то наши потери составят $w(\delta(X), \theta)$. Это есть *случайная величина*. Чтобы избежать этого неудобства, естественно в качестве потерь для стратегий $\delta = \delta(\cdot) \in \mathcal{D}$ и $\theta \in \Theta$ принять значение математического ожидания

$$W(\delta(\cdot), \theta) = \mathbf{E}_\theta w(\delta(X), \theta) = \int w(\delta(\boldsymbol{x}), \theta) \mathbf{P}_\theta(d\boldsymbol{x}), \tag{1}$$

которое называется *функцией риска* (появление слова «риск» здесь естественно, так как применение $\delta(\cdot)$ дает случайный результат). Если выполнено условие (A_μ) о наличии плотности $f_\theta(x)$ распределения $\mathbf{P}_\theta$ относительно некоторой σ-конечной меры μ, то функцию риска можно записать в виде

$$W(\delta(\cdot), \theta) = \int w(\delta(\boldsymbol{x}), \theta) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}).$$

Мы можем дать теперь следующее

Определение 1. *Статистической игрой* называется тройка $(\mathcal{D}, \Theta, W)$, где Θ есть множество стратегий природы, $\mathcal{D}$ — множество всех измеримых отображений пространства $\mathcal{X}^n$ в множество D, W определено в (1). Для более полной характеризации статистической игры вместе с тройкой $(\mathcal{D}, \Theta, W)$ можно считать также заданной пару $(X, \mathbf{P}_\theta)$, где $X \in \mathbf{P}_\theta$.

Пример 1. Пусть $\theta \in [0, 1]$ определяет содержание некоторой химической компоненты руды, приготовленной для плавки. Если мы примем решение, что доля этой компоненты равна $\delta \neq \theta$, и в соответствии с этим решением будет организован весь процесс плавки, то в результате качество выплавленного металла будет хуже, чем при $\delta = \theta$, а расходы энергии выше. Другими словами, мы будем нести потери $w(\delta, \theta)$, которые будут тем больше,

*) В построениях этого параграфа мы могли бы, не ограничивая общности, считать, что $n = 1$. Однако мы сохраним понятие выборки объема n с тем, чтобы сохранить простые связи с предыдущими результатами и с последующими рассмотрениями (§ 78–80).

Более общая и более сложная концепция статистической игры имеет дело с неограниченной выборкой $X_\infty = (\mathrm{x}_1, \mathrm{x}_2, \ldots)$, использование элемента x_n которой сопряжено с затратами $c_n \geqslant 0$ (см. [3, 112]).

чем больше δ отличается от θ. Предположим для простоты, что $w(\delta,\theta)$ пропорциональна квадрату отклонения δ от θ:

$$w(\delta,\theta)=c(\delta-\theta)^2.$$

(Если функция $w(\delta,\theta)$ гладкая и если рассматривать окрестность прямой $\delta=\theta$, то упрощающим предположением здесь будет лишь независимость c от θ.) В результате получим игру (D,Θ,w), у которой $D=[0,1]$, $\Theta=[0,1]$,

$$w(\delta,\uparrow)=\sup_{\theta} w(\delta,\theta)=\begin{cases} c\delta^2 & \text{при} \quad \delta>\dfrac{1}{2},\\ c(1-\delta)^2 & \text{при} \quad \delta\leqslant\dfrac{1}{2},\end{cases}$$

$$w^*=\inf_{\delta} w(\delta,\uparrow)=w\left(\frac{1}{2},\uparrow\right)=\frac{c}{4}.$$

Таким образом, стратегия $\delta=1/2$ является минимаксной и гарантирует потери не больше $c/4$. Так как $w_*=0$, то эта игра цены не имеет. Рандомизация игры минимаксную стратегию $\delta=1/2$ не улучшает (она дает $\widetilde{w}_*=c/4$). Предоставляем читателю убедиться, что байесовская стратегия δ_Q здесь имеет вид $\delta_Q=\mathbf{E}_Q\theta=\int t\mathbf{Q}(dt)$ (это вытекает из равенств $\widetilde{w}(\delta,\mathbf{Q})=c\mathbf{E}_Q(\delta-\theta)^2=c\mathbf{E}_Q(\theta-\mathbf{E}_Q\theta)^2+c\mathbf{E}_Q(\delta-\mathbf{E}_Q\theta)^2$) и что наихудшее распределение $\overline{\mathbf{Q}}$ будет иметь вид $\overline{\mathbf{Q}}(\{0\})=\overline{\mathbf{Q}}(\{1\})=1/2$. Очевидно, что соответствующая ему байесовская стратегия есть $\delta_{\overline{\mathbf{Q}}}=1/2$.

Предположим теперь, что руда неоднородна и что мы имеем возможность брать n проб руды. Эти пробы организованы так, что результаты лабораторных анализов для содержания упомянутой компоненты в пробах будут случайны и дадут нам независимые значения $(\mathrm{x}_1,\ldots,\mathrm{x}_n)=X$, относительно которых известно, что $\mathbf{E}\mathrm{x}_i=\theta$, $\mathbf{D}\mathrm{x}_i=b(\theta)$. В этом случае решениями $\delta(X)$ будут всевозможные оценки $\theta^*=\delta(X)$ параметра θ по выборке X. Риск решающей функции $\delta(X)$ будет равен

$$W(\delta,\theta)=c\mathbf{E}_\theta(\delta(X)-\theta)^2,$$

и мы приходим к задаче отыскания оценки $\theta^*=\delta(X)$, минимизирующей в том или ином смысле этот риск. Если положить, например, $\delta_1(X)=\overline{\mathrm{x}}$, то получим

$$W(\delta_1,\theta)=\frac{cb(\theta)}{n}. \tag{2}$$

Наибольшее значение $b(\theta)$ равно $\theta(1-\theta)$ и достигается на распределении x_1, сосредоточенном в точках 0 и 1. Так как такую возможность можно исключить, то

$$b(\theta)<\theta(1-\theta)\leqslant\frac{1}{4},\quad W(\delta_1,\theta)<\frac{c}{4n}.$$

Таким образом, даже при $n=1$ и при использовании, возможно, не наилучшей стратегии мы получим результат, лучший чем для минимаксной

стратегии в игре без выборки. Соотношение (2) показывает также, что риск сходится к нулю при $n \to \infty$ и будет тем меньше, чем меньше $b(\theta)$. $\triangleleft$

Из приведенного выше определения статистической игры видно, что последняя обладает значительно более богатым множеством стратегий $\mathcal{D}$ и по сравнению с исходной игрой (D, Θ, w).

Как и в § 74, наряду с игрой $(\mathcal{D}, \Theta, W)$, стратегии которой будем называть *чистыми*, можно рассматривать *рандомизированные* или *смешанные игры* $(\widetilde{\mathcal{D}}, \widetilde{\Theta}, \widetilde{W})$. Здесь множество $\widetilde{\mathcal{D}}$ является множеством отображений $\pi(X)\colon \mathcal{X}^n \to \widetilde{D}$. Эти отображения должны быть такими, чтобы значения

$$\widetilde{w}(\pi(X), \theta) = \int\limits_D w(u, \theta)\pi(X, du)$$

были случайными величинами ($\pi(X, A)$ есть вероятность множества $A \subset D$ в соответствии с решающим правилом π в точке X). Тогда по определению полагаем

$$\widetilde{W}(\pi(\cdot), \mathbf{Q}) = \int\limits_\Theta \int\limits_{\mathcal{X}^n} \int\limits_D w(u, t)\pi(\boldsymbol{x}, du)\mathbf{P}_t(d\boldsymbol{x})\mathbf{Q}(dt).$$

Стратегия $\pi(X)$ называется *рандомизированным решающим правилом.*

Отношения частичного порядка между стратегиями, равномерно наилучшие стратегии, байесовские и минимаксные стратегии, полные классы стратегий для статистических игр определяются точно так же, как для игр обычных (с заменой множества D на $\mathcal{D}$, а функций w, $\widetilde{w}$ — на W, $\widetilde{W}$).

На статистические игры полностью переносятся утверждения теорем 73.1–73.6, так как последние с природой множества D никак не связаны.

2. Классификация статистических игр. С природой множеств D и Θ связана следующая классификация, выделяющая основные виды статистических игр.

1. Если $\Theta = A$, $D = A$, где A есть «телесное» подмножество в $\mathbb{R}^k$ (например, параллелепипед), $w(t, t) = 0$, $w(t, u) > 0$ при $t \neq u$, то мы получим задачи теории *точечного оценивания* неизвестного параметра θ.

2. Если множества $\Theta = \{\theta_1, \ldots, \theta_r\}$, $D = \{\delta_1, \ldots, \delta_r\}$ конечны и содержат одинаковое число элементов, $w(\delta_i, \theta_i) = 0$, $w(\delta_i, \theta_j) > 0$ при $i \neq j$, то мы получим задачи *проверки конечного числа простых гипотез.*

3. Если Θ есть «телесная» область в $\mathbb{R}^k$, $D = \{\delta_1, \delta_2\}$ состоит из двух элементов, $w(\delta_1, \theta) = 0$ при $\theta \in \Theta_1$, $w(\delta_2, \theta) = 0$ при $\theta \in \Theta_2$ ($\Theta_1 \cup \Theta_2$ пусто) и $w(\delta_i, \theta) > 0$ в остальных случаях, то мы приходим к задаче *проверки гипотез* $\{\theta \in \Theta_1\}$ и $\{\theta \in \Theta_2\}$.

Возможны, конечно, и другие классы задач. Мы выделили эти три типа, поскольку они рассматривались нами в гл. 2, 3. Мы рассматривали эти задачи с чисто «статистических» позиций, что соответствует специальному выбору функций $w(\delta, \theta)$: в первой группе задач потери определялись среднеквадратичным уклонением, что соответствует функции потерь $w(\delta, \theta) = (\delta - \theta)^2$;

во второй группе задач потери определялись вероятностью ошибиться, что соответствует функции

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

То же относится и к третьей группе задач, в которой мы использовали функцию потерь

$$w(\delta_1, \theta) = \begin{cases} 0 & \text{при} \quad \theta \in \Theta_1, \\ 1 & \text{при} \quad \theta \in \Theta_2, \end{cases} \qquad w(\delta_2, \theta) = \begin{cases} 1 & \text{при} \quad \theta \in \Theta_1, \\ 0 & \text{при} \quad \theta \in \Theta_2. \end{cases}$$

Эти функции потерь, соответствующие чисто статистическому подходу к проблемам, будем называть *статистическими*.

Приведенная классификация показывает, что никакой принципиальной разницы между задачами теории оценивания и проверкой статистических гипотез нет. Все дело в природе множеств Θ и D и виде функций потерь.

На примере этой классификации можно отметить еще одно своеобразие статистических игр (в дополнение к п. 1, 2 в начале этого параграфа); оно состоит в том, что множество D в статистических играх либо совпадает с Θ, либо представляет собой множество, более бедное, чем Θ.

3. Две фундаментальные теоремы теории статистических игр. Сформулируем теперь основные результаты теории статистических игр. Мы уже отмечали, что утверждения теорем 73.1–73.6 для статистических игр полностью сохраняются, так как они с природой игр не связаны. Чтобы получить две фундаментальные теоремы, упоминавшиеся в § 74, введем некоторые предположения. Это далеко не самые общие предположения (иначе формулировки и доказательства чрезвычайно усложнились бы), но они достаточно широкие, чтобы охватить наиболее интересный и содержательный круг задач и, в частности, задачи, рассмотренные в гл. 2, 3.

Условие (A). *Каждое из множеств Θ и D либо конечно, либо представляет собой компактное множество в $\mathbb{R}^k$.*

Как уже отмечалось, случай, когда Θ конечно, а $D \subset \mathbb{R}^k$, можно не рассматривать. В остальных трех случаях мы будем предполагать, что функция потерь $w(\delta, \theta)$ удовлетворяет следующему условию.

Условие (B). 1. *Если $D \subset \mathbb{R}^k$, $\Theta \subset \mathbb{R}^k$, то функция $w(\delta, \theta)$ непрерывна на $D \times \Theta$.*

2. *Если $\Theta \subset \mathbb{R}^k$, а $D = \{\delta_1, \ldots, \delta_r\}$ конечно, то каждая из r функций $w(\delta_i, \theta)$, $i = 1, \ldots, r$, непрерывна на Θ.*

Если $\Theta = \{\theta_1, \ldots, \theta_r\}$ и $D = \{\delta_1, \ldots, \delta_r\}$ конечны, то значения $w(\delta_i, \theta_j)$, $i, j = 1, \ldots, r$, могут быть произвольными.

Кроме того, мы потребуем, чтобы выполнялось

Условие (C). *Мы располагаем выборкой $X \in \mathbf{P}_\theta$ из распределения $\mathbf{P}_\theta$, абсолютно непрерывного при всех θ относительно некоторой*

σ-конечной меры μ. Если $\Theta \subset \mathbb{R}^k$, то плотность $\dfrac{d\mathbf{P}_\theta}{d\mu}(x) = f_\theta(x)$ *непрерывна в* $L_1(\mathcal{X}, \mathfrak{B}_\mathcal{X}, \mu)$ *относительно* θ, *т. е. при* $\theta_m \to \theta$

$$\int |f_{\theta_m}(x) - f_\theta(x)|\mu(dx) \to 0. \tag{3}$$

Нетрудно проверить, что обычная непрерывность $f_\theta(x)$ по θ для $[\mu]$ п. в. влечет за собой непрерывность (3).

Следующие две теоремы носят название фундаментальных теорем теории статистических игр.

Теорема 1. *Если выполнены условия* (A), (B), (C), *то усредненная игра* $(\widetilde{\mathcal{D}}, \widetilde{\Theta}, \widetilde{W})$ *имеет цену и минимаксные стратегии* $\overline{\pi}(X)$ *и* $\overline{\mathbf{Q}}$:

$$\widetilde{W}(\overline{\pi}(\cdot), \uparrow) = \inf_\pi \widetilde{W}(\pi, (\cdot), \uparrow), \quad \widetilde{W}(\downarrow, \overline{\mathbf{Q}}) = \sup_\mathbf{Q} \widetilde{W}(\downarrow, \mathbf{Q}).$$

Из теорем 73.4–73.6 предыдущего параграфа мы знаем, что $\overline{\mathbf{Q}}$ есть наихудшее распределение,

$$\widetilde{W}(\pi_{\overline{\mathbf{Q}}}(\cdot), \overline{\mathbf{Q}}) = \sup_\mathbf{Q} \widetilde{W}(\pi_\mathbf{Q}(\cdot), \mathbf{Q}) = \sup_\mathbf{Q} \widetilde{W}(\downarrow, \mathbf{Q}),$$

и $\overline{\pi}(X) = \pi_{\overline{\mathbf{Q}}}(X)$ является байесовской стратегией, соответствующей $\overline{\mathbf{Q}}$.

Мы знаем также (см. теоремы 74.3–74.5), что для того чтобы стратегия $\overline{\pi}(X)$ была минимаксной, необходимо и достаточно, чтобы она была байесовской: $\overline{\pi}(X) = \pi_\mathbf{Q}(X)$ для некоторого априорного распределения $\mathbf{Q}$ и

$$\widetilde{W}(\overline{\pi}(\cdot), \overline{\theta}) = c = \text{const} \quad \text{п. в.} \quad [\mathbf{Q}], \quad \widetilde{W}(\overline{\pi}(\cdot), \overline{\theta}) \leqslant c.$$

Этот последний критерий минимаксности мы уже неоднократно использовали в различных частных ситуациях (см. § 21, 41, 45, 49).

Теорема 2. *При выполнении условий* (A), (B), (C) *класс всех байесовских стратегий является полным.*

В приложении VIII мы приводим доказательства теорем 1, 2 в их более общей форме, когда D и Θ — произвольные компактные метрические пространства (условие (A)); функция $w(\delta, \theta)$: $D \times \Theta \to \mathbb{R}$ непрерывна по δ и θ в соответствующих метриках (условие (B)); распределение $\mathbf{P}_\theta$ непрерывно относительно θ по вариации (условие (C)).

Доказательства теорем 1, 2 при некоторых дополнительных предположениях можно извлечь из [30]. Доказательства для случая конечных D и Θ можно найти в [9, 109]. В этих же монографиях можно найти сравнительно полное изложение элементов общей теории статистических игр (и, в частности, обсуждение для некоторых случаев конструкции минимального полного класса; см. [109]).

Теоремы 1, 2 показывают, насколько важной является задача описания класса всех байесовских решающих правил. Этой задаче посвящен следующий параграф.

§ 76. Байесовский принцип. Полный класс решающих функций

Мы видели, что по своей конструкции статистическая игра является объектом более сложным, чем исходная игра (D, Θ, w). Для последней, особенно в случае простых множеств D и Θ (например, конечных), отыскание байесовских и минимаксных стратегий может оказаться делом сравнительно простым. В то же время даже простейшие статистические игры имеют весьма сложную природу множества $\mathcal{D}$, и это заметно усложняет их изучение, если подходить к ним как к обычным играм.

Пример 1. Пусть множества $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$ двухточечные, $w(\delta_i, \theta_j) = w_{ij}$, $w_{ii} = 0$, $i, j = 1, 2$. Пусть $\mathbf{Q} = (q, 1-q)$ есть априорное распределение на Θ. Тогда

$$\widetilde{w}(\delta_i, \mathbf{Q}) = qw_{i1} + (1-q)w_{i2}.$$

Стало быть, байесовская стратегия $\pi_{\mathbf{Q}}$ имеет вид

$$\pi_{\mathbf{Q}}(\delta_2) = \begin{cases} 0, & \text{если } \widetilde{w}(\delta_1, \mathbf{Q}) < \widetilde{w}(\delta_2, \mathbf{Q}) \quad (qw_{21} > (1-q)w_{12}), \\ 1, & \text{если } \widetilde{w}(\delta_2, \mathbf{Q}) < \widetilde{w}(\delta_1, \mathbf{Q}) \quad (qw_{21} < (1-q)w_{12}) \end{cases} \tag{1}$$

($\pi_{\mathbf{Q}}(\delta_i)$ есть вероятность принятия δ_i).

Если

$$\widetilde{w}(\delta_1, \mathbf{Q}) = \widetilde{w}(\delta_2, \mathbf{Q}) \tag{2}$$

или, что то же, $q = \overline{q} = w_{12}/(w_{12} + w_{21})$, то в качестве $\pi_{\mathbf{Q}}$ можно взять любое распределение π на множестве (δ_1, δ_2). Точно так же всегда можно найти распределение $\pi = (p, 1-p)$ такое, что

$$\widetilde{w}(\pi, \theta_1) = \widetilde{w}(\pi, \theta_2) \quad \text{или} \quad pw_{12} = (1-p)w_{21}.$$

Решение этого уравнения $\overline{p} = w_{21}/(w_{21} + w_{12})$, очевидно, отвечает байесовской уравнивающей стратегии $\pi_{\overline{\mathbf{Q}}}$, $\overline{\mathbf{Q}} = (\overline{q}, 1-\overline{q})$, которая в силу теорем 73.2, 73.5, 73.6 будет минимаксной. Распределение $\overline{\mathbf{Q}}$ (минимаксная стратегия «природы») будет наихудшим.

Мы видим, что «решение» этой игры осуществляется довольно просто. Если же перейти к статистической игре, то даже в простейшем случае $w_{12} = w_{21} = 1$ мы получим задачу о байесовских и минимаксных критериях, для рассмотрения которой нам потребовались § 41, 42.

Замечательный факт, которому посвящен настоящий параграф, состоит в том, что задача отыскания байесовских стратегий (а стало быть, полного класса и минимаксных стратегий) для статистических игр может быть редуцирована в известном смысле к той же задаче для исходных игр (D, Θ, w). В основе этой редукции лежит следующее утверждение, которое назовем *байесовским принципом*. Пусть, как и прежде,

$$f_\theta(X) = \prod_{i=1}^{n} f_\theta(\mathrm{x}_i)$$

есть функция правдоподобия выборки X; она же — плотность X в $\mathcal{X}^n$ относительно μ^n. Пусть, кроме того, априорное распределение $\mathbf{Q}$ на $(\Theta, \mathfrak{F}_\Theta)$ имеет

плотность $q(t)$ относительно некоторой меры λ (очевидно, это не есть ограничение). Тогда согласно § 21 функция $f(\boldsymbol{x}, t) = q(t) f_t(\boldsymbol{x})$ будет плотностью совместного распределения (X, θ) в $\mathcal{X}^n \times \Theta$. Это означает, что функция

$$q(t/\boldsymbol{x}) = \frac{q(t) f_t(\boldsymbol{x})}{f(\boldsymbol{x})}, \quad f(\boldsymbol{x}) = \int q(t) f_t(\boldsymbol{x}) \lambda(dt), \tag{3}$$

определяет условную плотность распределения θ при условии $X = \boldsymbol{x}$. Эта плотность соответствует *апостериорному распределению* $\mathbf{Q}_{\boldsymbol{x}}$ случайной величины θ при условии $X = \boldsymbol{x}$. Соотношение (3) носит название *формулы Байеса* (см. § 20, 21).

Теорема 1 (байесовский принцип). *Пусть выполнено условие* (A_μ), *априорное распределение на* Θ *имеет плотность* $q(t)$ *и* $\mathbf{Q}_X$ *означает апостериорное распределение с плотностью* (3), *соответствующее априорному распределению* $\mathbf{Q}$. *Пусть, далее, исходная игра* (D, Θ, w) *для любого априорного распределения* $\mathbf{Q}$ *имеет байесовскую стратегию* $\pi_{\mathbf{Q}}$. *Тогда статистическая игра* $(\mathcal{D}, \Theta, W)$ *имеет байесовскую стратегию* $\pi_{\mathbf{Q}}(X)$, *соответствующую распределению* $\mathbf{Q}$, *которая совпадает с* $\pi_{\mathbf{Q}_X}$ *— байесовской стратегией исходной игры, соответствующей апостериорному распределению* $\mathbf{Q}_X$.

Утверждение этой теоремы можно выразить одним равенством

$$\pi_{\mathbf{Q}}(X) = \pi_{\mathbf{Q}_X}.$$

Оно сводит поставленную задачу к нахождению апостериорного распределения $\mathbf{Q}_X$ и к отысканию байесовских стратегий для исходной игры.

Теорема 1 очень существенна для понимания механизма влияния информации, полученной от выборки, на выбор оптимальной стратегии. Априорная информация, представленная распределением $\mathbf{Q}$ на Θ, под влиянием экспериментальных данных постоянно меняется. Оптимальной стратегией будет та, которая учитывает эти изменения следующим образом: надо взять оптимальную стратегию в исходной игре, но стратегию, соответствующую уже не $\mathbf{Q}$, а $\mathbf{Q}_X$.

Доказательство теоремы 1. Имеем

$$\widetilde{W}(\pi(\cdot), \mathbf{Q}) = \int\limits_{\Theta} \int\limits_{\mathcal{X}^n} \widetilde{w}(\pi(\boldsymbol{x}), t) f_t(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) q(t) \lambda(dt) =$$

$$= \int\limits_{\mathcal{X}^n} f(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) \int\limits_{\Theta} \widetilde{w}(\pi(\boldsymbol{x}), t) q(t/\boldsymbol{x}) \lambda(dt). \tag{4}$$

Здесь мы воспользовались (3). Изменение порядка интегрирования законно в силу неотрицательности подынтегральной функции. Второй интеграл в правой части (4) есть не что иное, как $\widetilde{w}(\pi(\boldsymbol{x}), \mathbf{Q}_{\boldsymbol{x}})$. Но при любом $\boldsymbol{x}$

$$\widetilde{w}(\pi(\boldsymbol{x}), \mathbf{Q}_{\boldsymbol{x}}) \geqslant \widetilde{w}(\pi_{\mathbf{Q}_{\boldsymbol{x}}}, \mathbf{Q}_{\boldsymbol{x}}) = \int\limits_{\Theta} \widetilde{w}(\pi_{\mathbf{Q}_{\boldsymbol{x}}}, t) q(t/\boldsymbol{x}) \lambda(dt).$$

Подставляя это неравенство в (4) и возвращаясь к исходному порядку интегрирования, получим

$$\widetilde{W}(\pi(\cdot),\mathbf{Q}) \geqslant \int\limits_{\mathcal{X}^n} f(\boldsymbol{x})\mu^n(d\boldsymbol{x}) \int\limits_{\Theta} \widetilde{w}(\pi_{\mathbf{Q}_{\boldsymbol{x}}},t)q\,(t/\boldsymbol{x})\,\lambda(dt) = \widetilde{W}(\pi_{\mathbf{Q}.},\mathbf{Q}).$$

Так как $\pi(\boldsymbol{x})$ здесь произвольна, то это означает, что

$$\pi_{\mathbf{Q}}(\boldsymbol{x}) = \pi_{\mathbf{Q}_{\boldsymbol{x}}}. \quad \triangleleft$$

Замечание 1. Аккуратности ради мы должны в проведенных рассмотрениях оговаривать измеримость относительно $\mathfrak{B}^n \times \mathfrak{F}_\Theta$ функции $\widetilde{w}(\pi_{\mathbf{Q}_{\boldsymbol{x}}},t)$. Эти оговорки мы опускаем, так как они носят чисто технический характер, а при выполнении условий (A), (B), (C) из § 75 они заведомо излишни. Последнее утверждение читатель может проверить самостоятельно, пользуясь тем, что для дискретных D и Θ эта измеримость устанавливается очевидным образом, и тем, что произвольная игра при условиях (A), (B) может быть сколь угодно точно «приближена» к дискретной.

Возвращаясь к примеру 1, мы можем теперь на основании теоремы 1 сразу указать вид байесовских стратегий для соответствующей статистической игры. Именно, из (1) получаем

$$\pi_{\mathbf{Q}_X}(\delta_2) = \begin{cases} 0, & \text{если} \quad q_X \equiv \dfrac{qf_{\theta_1}(X)}{qf_{\theta_1}(X)+(1-q)f_{\theta_2}(X)} > \dfrac{w_{12}}{w_{12}+w_{21}}, \\ 1, & \text{если} \quad q_X < \dfrac{w_{12}}{w_{12}+w_{21}}. \end{cases} \tag{5}$$

Если

$$q_X = \frac{w_{12}}{w_{12}+w_{21}}, \tag{6}$$

то в качестве $\pi_{\mathbf{Q}_X}$ можно брать любое распределение на (δ_1,δ_2). Неравенство (5) можно переписать в виде

$$\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}, \quad a = \frac{w_{12}}{w_{12}+w_{21}}. \tag{7}$$

Это знакомый нам критерий отношения правдоподобия.

Далее,

$$\widetilde{W}(\pi_{\mathbf{Q}.},\theta_j) = w_{1i}\mathbf{E}_{\theta_j}\pi_{\mathbf{Q}_X}(\delta_1) + w_{2j}\mathbf{E}_{\theta_j}\pi_{\mathbf{Q}_X}(\delta_2), \quad j=1,2.$$

Допустим для простоты, что равенство (6) имеет место с $\mathbf{P}_{\theta_j}$-вероятностью 0, так что байесовская стратегия с $\mathbf{P}_{\theta_j}$-вероятностью 1 будет чистой, $j=1,2$. Тогда

$$\mathbf{E}_{\theta_j}\pi_{\mathbf{Q}_X}(\delta_1) = \mathbf{P}_{\theta_j}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right),$$

$$\widetilde{W}(\pi_{\mathbf{Q}.},\theta_1) = w_{21}\mathbf{P}_{\theta_1}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} < \frac{a(1-q)}{q(1-a)}\right),$$

$$\widetilde{W}(\pi_{\mathbf{Q}.},\theta_2) = w_{12}\mathbf{P}_{\theta_2}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

Отсюда уже нетрудно найти значение $\overline{q}$, соответствующее наихудшему распределению $\overline{\mathbf{Q}}$, при котором $\pi_{\overline{\mathbf{Q}}_X}$ будет уравнивающей стратегией, т. е. стратегией, при которой

$$\widetilde{W}(\pi_{\overline{\mathbf{Q}}.}, \theta_1) = \widetilde{W}(\pi_{\overline{\mathbf{Q}}.}, \theta_2).$$

Эта стратегия согласно теоремам 74.4, 74.6 будет минимаксной.

Распространение описанной процедуры отыскания минимаксной стратегии на общий случай, когда $\mathbf{P}_{\theta_1}$- или $\mathbf{P}_{\theta_2}$-распределения $f_{\theta_1}(X)/f_{\theta_2}(X)$ содержат дискретную компоненту, мы предоставляем читателю.

Пользуясь теоремой 1, мы можем аналогичным образом получить обобщение результатов § 41, 42 на случай произвольных конечных D и Θ и произвольной функции потерь $w(\delta_i, \theta_j) = w_{ij}$, которую в этом случае можно называть также матрицей потерь $\|w(\delta_i, \theta_j)\|$. (В § 41, 42 мы рассматривали частный случай $w_{ij} = 1$ при $i \neq j$.) При произвольных w_{ij} байесовское решающее правило будет иметь следующий вид. Пусть $\mathbf{Q} = (q(\theta_1), \ldots, q(\theta_r))$, $\mathbf{Q}_X = (q_X(\theta_1), \ldots, q_X(\theta_r))$,

$$q_X(\theta_j) = \frac{q(\theta_j) f_{\theta_j}(X)}{\sum\limits_i q(\theta_i) f_{\theta_i}(X)}.$$

Тогда $\widetilde{w}(\delta_i, \mathbf{Q}_X) = \sum\limits_{j=1}^{r} w_{ij} q_X(\theta_j)$ и, стало быть,

$$\pi_{\mathbf{Q}_X}(\delta_k) = 1, \quad \text{если} \quad \widetilde{w}(\delta_k, \mathbf{Q}_X) \leqslant \widetilde{w}(\delta_i, \mathbf{Q}_X) \quad \text{при всех } i$$

или, что то же, если

$$\sum_{j=1}^{r} w_{kj} f_{\theta_j}(X) q(\theta_j) \leqslant \sum_{j=1}^{r} w_{ij} f_{\theta_j}(X) q(\theta_j).$$

Если существует несколько значений k, обладающих этим свойством (обозначим их через $k_1, \ldots, k_s$), то байесовской стратегией $\pi_{\mathbf{Q}_X}$ будет также любое распределение на $\delta_{k_1}, \ldots, \delta_{k_s}$.

Отыскание минимаксной стратегии происходит следующим образом. Предположим опять для простоты, что $\mathbf{P}_{\theta_j}$-распределения $\widetilde{w}(\delta_i, \mathbf{Q}_X)$ не имеют дискретных компонент. Тогда

$$\widetilde{W}(\pi_{\mathbf{Q}.}, \theta_j) = \sum_{i \neq j} w_{ij} \mathbf{P}_{\theta_j}(\widetilde{w}(\delta_i, \mathbf{Q}_X) < \min_{l \neq i} \widetilde{w}(\delta_l, \mathbf{Q}_X)).$$

В силу теоремы 75.1 существует $\overline{\mathbf{Q}} = (\overline{q}(\theta_1), \ldots, \overline{q}(\theta_r))$, при котором стратегия $\pi_{\overline{\mathbf{Q}}_X}$ будет уравнивать значения $\widetilde{W}(\pi_{\overline{\mathbf{Q}}_X}, \theta_j)$ при всех j. Эта стратегия и будет минимаксной.

Из проведенных рассмотрений и теоремы 75.2 нетрудно получить также вид полного класса стратегий статистической игры $(\mathcal{D}, \Theta, \widetilde{W})$ в случае конечных D и Θ.

Рассмотрим стратегии $\pi_{\mathbf{Q}_X}$, представляющие собой произвольное распределение на тех $\delta_{k_1}, \ldots, \delta_{k_s}$, для которых

$$\min_i \left(\sum_{j=1}^{r} (w_{k_i,j} - w_{ij}) f_{\theta_j}(X) q(\theta_j) \right) = 0.$$

Класс таких стратегий (байесовских), который получится, если $q(\theta_1), \ldots, q(\theta_r)$ будут пробегать все возможные значения, и будет полным классом. Мы видели, что в случае $r = 2$ этот класс оказывается весьма простым и узким (см. (7)): он состоит из решающих функций $\pi(X) = (\pi(X, \delta_1), \pi(X, \delta_2))$, где $\pi(X, \delta_i)$ суть вероятности принятия решения δ_i,

$$\pi(X, \delta_1) = \begin{cases} 1, & \text{если} \quad R(X) > c, \\ p \in [0,1], & \text{если} \quad R(X) = c, \\ 0, & \text{если} \quad R(X) < c. \end{cases} \tag{8}$$

$$R(X) = \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}, \quad 0 \leqslant c \leqslant \infty.$$

В играх с *континуальными* множествами D и Θ для некоторых важных конкретных функций потерь также можно найти в явном виде форму байесовских решений. Пусть, например, D и Θ — области из $\mathbb{R}^k$, а функция потерь является квадратичной:

$$w(\delta, \theta) = c|\delta - \theta|^2 = c \sum_{i=1}^{k} (\delta_i - \theta_i)^2, \tag{9}$$

где δ_i, θ_i — координаты δ и θ. Тогда

$$\widetilde{w}(\delta, \mathbf{Q}) = c \int |\delta - t|^2 \mathbf{Q}(dt) = c\mathbf{E}_{\mathbf{Q}} |\delta - \theta|^2.$$

Мы знаем, что минимум этого выражения достигается при $\delta = \mathbf{E}_{\mathbf{Q}}\theta = \int t\mathbf{Q}(dt)$. Это и есть, очевидно, байесовская стратегия $\delta_{\mathbf{Q}} = \mathbf{E}_{\mathbf{Q}}\theta$. Отсюда и из байесовского принципа следует, что байесовская стратегия $\delta_{\mathbf{Q}}(X) = \theta^*_{\mathbf{Q}}$ в статистической игре будет иметь вид

$$\theta^*_{\mathbf{Q}} = \delta_{\mathbf{Q}_X} = \int_{R^k} t\mathbf{Q}_X(dt) = \int_{R^k} tq\,(t/X)\,\lambda(dt). \tag{10}$$

Этот результат нам уже известен по гл. 2.

Риск стратегии $\theta^*_{\mathbf{Q}}$ равен $W\left(\theta^*_{\mathbf{Q}}, \theta\right) = c\mathbf{E}_\theta \left|\theta^*_{\mathbf{Q}} - \theta\right|^2$. Априорное распределение $\overline{\mathbf{Q}}$, для которого $\mathbf{E}_\theta \left|\theta^*_{\mathbf{Q}} - \theta\right|^2 = \text{const}$, дает нам минимаксную оценку

$\overline{\theta}^* = \delta_{\overline{\mathbf{Q}}}(X)$. Примеры построения минимаксных оценок на этом пути содержатся в § 21.

Класс оценок (10), где $\mathbf{Q}$ пробегает значения в классе всех распределений на Θ, представляет собой полный класс.

Рассмотрим теперь другой частный случай функции потерь

$$w(\delta, \theta) = c|\delta - \theta| \tag{11}$$

и предположим, что $\Theta = \mathbb{R}$, $D = \mathbb{R}$. Тогда

$$\widetilde{w}(\delta, \mathbf{Q}) = c\mathbf{E}_{\mathbf{Q}}|\delta - \theta| = c\int |\delta - t|\mathbf{Q}(dt) = c\int\limits_{-\infty}^{\delta} (\delta - t)\mathbf{Q}(dt) + c\int\limits_{\delta}^{\infty} (t - \delta)\mathbf{Q}(dt).$$

Пользуясь интегрированием по частям и положив $F(t) = \mathbf{Q}((-\infty, t))$, находим

$$\widetilde{w}(\delta, \mathbf{Q}) = c\int\limits_{-\infty}^{\delta} (\delta - t)\, dF(t) - c\int\limits_{\delta}^{\infty} (t - \delta) d(1 - F(t)) =$$

$$= c\left[\int\limits_{-\infty}^{\delta} F(t)\, dt + \int\limits_{\delta}^{\infty} (1 - F(t))\, dt\right].$$

Производная этого выражения по δ существует п. в. и равна $c[2F(\delta) - 1]$. Эта функция монотонно возрастает и меняет знак в точке $\overline{\delta}$, равной медиане распределения F: $F(\overline{\delta} - 0) \leqslant 1/2$, $F(\overline{\delta} + 0) \geqslant 1/2$. Отсюда следует, что $\widetilde{w}(\delta, \mathbf{Q})$ будет выпуклой по δ и в точке $\overline{\delta}$ будет иметь минимум.

В силу байесовского принципа это означает, что *байесовской оценкой* $\theta^*_{\mathbf{Q}} = \delta_{\mathbf{Q}}(X)$ *для априорного распределения* $\mathbf{Q}$ *и функции потерь* (11) *будет медиана апостериорного распределения* $\mathbf{Q}_X$. Это дает возможность, как и в случае (9), найти минимаксную решающую функцию и полный класс.

Аналогичным образом можно рассмотреть случай

$$w(\delta, \theta) = c|\delta - \theta|^{\alpha}, \quad \alpha > 0.$$

В заключение этого параграфа отметим, что квадратичная функция потерь (9) при $c = 1$ для континуальных множеств D и Θ и функция потерь

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j, \end{cases} \tag{12}$$

для конечных D и Θ играют в теории статистических игр особую роль. Для них функции риска превращаются соответственно в сумму дисперсии и квадрата смещения оценки для континуальных D и Θ и в вероятность ошибиться — для конечных D и Θ. Эти характеристики, являющиеся естественными сами по себе, служили нам основой для выбора оптимальных правил в гл. 2–5. Если статистическая задача не содержит прямых указаний на форму функции $w(\delta, \theta)$, то чаще всего в качестве $w(\delta, \theta)$ выбираются именно эти две функции (9) или (12). Мы условились называть их *статистическими функциями потерь*.

§ 77. Достаточность, несмещенность, инвариантность

Принципы достаточности, несмещенности, инвариантности служат для сужения класса решающих правил. Они состоят в том, чтобы использовать в качестве решающих функций лишь соответственно достаточные, несмещенные и инвариантные решающие правила. Использование одного из этих принципов, двух из них или всех трех (если это оказывается возможным) позволяет в ряде случаев настолько сузить класс рассматриваемых стратегий, что его пересечение с полным классом оказывается состоящим из одной-единственной решающей функции. Это означает, что в выделенном подклассе существует равномерно наилучшая стратегия (ср. с п. 73.1), и это решает проблему выбора решения.

Все три принципа представляются довольно естественными и уже обсуждались нами в разных конкретных задачах в гл. 2–5.

Наиболее бесспорным из них является принцип достаточности, который часто представляет собой не что иное, как способ описания полного класса.

1. Достаточность. Предположим, что выполнено условие (A_μ) и что существует достаточная статистика S, т. е. (см. § 22)

$$f_\theta(X) = \psi(\theta, S) \cdot h(X).$$

Пусть, далее, априорное распределение $\mathbf{Q}$ имеет плотность $q(t)$ относительно некоторой меры λ. Тогда в силу байесовского принципа байесовская стратегия будет полностью определяться апостериорной плотностью

$$q(t/X) = \frac{q(t)f_t(X)}{\int q(u)f_u(X)\lambda(du)} = \frac{q(t)\psi(t,S)}{\int q(u)\psi(u,S)\lambda(du)}, \tag{1}$$

зависящей лишь от S. Так как любое распределение $\mathbf{Q}$ имеет плотность относительно подобранной соответствующим образом меры λ (можно положить, например, $\lambda = \mathbf{Q}$, $q(t) \equiv 1$), то сказанное означает, что все байесовские решающие правила $\pi_{\mathbf{Q}}(X)$ будут функциями лишь от S:

$$\pi_{\mathbf{Q}}(X) = p_{\mathbf{Q}}(S).$$

Другими словами, любая байесовская стратегия $\pi_{\mathbf{Q}}(X)$ не зависит от X при фиксированной S.

Пусть теперь выполнены условия (A)–(C) из § 75. Тогда высказанное утверждение будет относиться и к минимаксным стратегиям. Оно будет означать также, что все решающие правила, построенные как функции лишь от S (т. е. все измеримые отображения $\mathcal{S} \to \widetilde{D}$, где $\mathcal{S}$ есть пространство, в котором лежат значения S), образуют полный класс $\mathcal{D}_S$. Это следует из того, что $\mathcal{D}_S$ содержит в себе все байесовские стратегии, образующие, как мы знаем, полный класс. Очевидно, класс $\mathcal{D}_S$ будет наименьшим для минимальной достаточной статистики S.

Ясно, что в минимальный полный класс входят далеко не все функции от S (со значениями в $\widetilde{D}$), а лишь узкая их часть. Это подтверждает формула (1), из которой следует, например, что для двухточечных множеств D и Θ (см. (76.8)) полный класс образуют функции $\pi(X)$, у которых вероятность $\pi(X, \delta_1)$ принятия решения δ_1 имеет форму индикатора множества $\{R(X) > c\}$, где $R(X) = \psi(\theta_1, S)/\psi(\theta_2, S)$ (точнее см. (76.8)).

Если $D \subset \mathbb{R}^k$, $\Theta \subset \mathbb{R}^k$, а функция потерь $w(\delta, \theta)$ имеет вид $w(\delta, \theta) = w(\delta - \theta)$, где $w(u)$ — выпуклая функция в $\mathbb{R}^k$, то принципу достаточности можно придать очень конструктивную форму, позволяющую эффективно характеризовать полный класс. Именно, имеет место следующее обобщение теоремы 24.1.

Теорема 1 (Блекуэлл). *Для любой решающей функции (оценки)* $\theta^* = \delta(X)$ *существует оценка*

$$\theta_S^* = \mathbf{E}_\theta\left(\frac{\theta^*}{S}\right)$$

(θ_S^* *от* θ *не зависит, так как* S *— достаточная статистика*), *которая не хуже, чем* θ^*. *Именно, для всех* $\theta \in \Theta$

$$\mathbf{E}_\theta w(\theta_S^* - \theta) \leqslant \mathbf{E}_\theta w(\theta^* - \theta).$$

Доказательство. Имеет место следующее неравенство Иенсена (см. § 19): если g — выпуклая функция в $\mathbb{R}^k$, ξ — случайная величина со значениями в $\mathbb{R}^k$, а $\mathfrak{F}$ — какая-нибудь σ-подалгебра основной σ-алгебры, то

$$\mathbf{E}\left(\frac{g(\xi)}{\mathfrak{F}}\right) \geqslant g\left(\mathbf{E}\left(\frac{\xi}{\mathfrak{F}}\right)\right).$$

В соответствии с этим неравенством

$$\mathbf{E}_\theta w(\theta^* - \theta) = \mathbf{E}_\theta\left\{\mathbf{E}_\theta\left(\frac{w(\theta^* - \theta)}{S}\right)\right\} \geqslant$$

$$\geqslant \mathbf{E}_\theta w\left(\mathbf{E}_\theta\left(\theta^* - \frac{\theta}{S}\right)\right) = \mathbf{E}_\theta w(\theta_S^* - \theta). \quad \triangleleft$$

Если достаточная статистика S является полной, то теорема 1 вместе с принципом несмещенности позволяет однозначно определить равномерно наилучшую оценку. Действительно, рассмотрим класс K_0 всех несмещенных оценок $\theta^* = \delta(X)$:

$$\mathbf{E}_\theta \theta^* = \theta \quad \text{для} \quad \theta^* \in K_0.$$

Тогда, следуя в точности рассуждениям § 24 (теорема 24.3), убеждаемся, что $\theta_S^* = \mathbf{E}_\theta(\theta^*/S)$ для всех $\theta^* \in K_0$ совпадают и, стало быть, пересечение K_0 и полного класса состоит из одной-единственной оценки $\varphi(S)$, которую естественно называть *эффективной*.

Из сказанного видно, что *эффективные оценки, если они существуют, будут одни и те же для произвольной выпуклой функции потерь* $w(\delta - \theta)$. Это позволяет использовать для любой такой функции потерь все утверждения соответствующих теорем гл. 2, полученных для $w(u) = u^2$.

Приведенные рассуждения иллюстрируют совместное применение принципов достаточности и несмещенности.

2. Несмещенность. Мы только что видели, какую роль может играть принцип несмещенности в теории оценок. В § 46 было установлено, что

аналогичный эффект (существование равномерно наиболее мощных несмещенных критериев) может быть получен при использовании несмещенных критериев в теории проверки статистических гипотез.

В общем случае определение несмещенности выглядит следующим образом. Допустим, что проблема статистического решения состоит в «определении» неизвестного значения θ и что, следовательно, множества D и Θ совпадают. Функция потерь $w(\delta, \theta)$ может быть произвольной.

Определение 1. Решающая функция $\delta(X)$ называется *несмещенной*, если

$$\mathbf{E}_\theta w(\delta(X), \theta) \leqslant \mathbf{E}_\theta w(\delta(X), \theta')$$

при любых $\theta, \theta' \neq \theta$.

Другими словами, при $v = \theta$ достигается $\min\limits_v \mathbf{E}_\theta w(\delta(X), v)$. Это означает, что $\delta(X)$ в среднем находится ближе всего к неизвестному θ, чем к какой-нибудь другой точке.

Легко видеть, что данное ранее определение несмещенных оценок является частным случаем этого определения.

Если проверяются две сложные гипотезы $H_1 = \{\theta \in \Theta_1\}$ и $H_2 = \{\theta \in \Theta_2\}$, то множество $D = \{\delta_1, \delta_2\}$ может существенно отличаться от Θ. В этом случае определение несмещенности будет формально несколько иным, хотя смысл его остается тем же. Именно, определение 1 можно модифицировать так (см. [65, 66]), что оно перейдет в следующее определение.

Определение 1А. Решающая функция $\delta(X)$ называется *несмещенной*, если

$$\mathbf{E}_\theta w(\delta(X), \theta) \leqslant \mathbf{E}_\theta w(\delta(X), \theta')$$

при любых $\theta \in \Theta_1$, $\theta' \in \Theta_2$ или $\theta \in \Theta_2$, $\theta' \in \Theta_1$.

Пусть для простоты $w(\delta_1, \theta) = w_1 = \text{const}$ при $\theta \in \Theta_2$; $w(\delta_2, \theta) = w_2 = \text{const}$ при $\theta \in \Theta_1$; $\delta_1 = 0$, $\delta_2 = 1$ и $\delta(X)$ означает вероятность (1 или 0) принятия H_2. Тогда

$$\mathbf{E}_\theta w(\delta(X), \theta) = \begin{cases} w_2 \mathbf{P}_\theta(\delta(X) = 1) & \text{при} \quad \theta \in \Theta_1, \\ w_1 \mathbf{P}_\theta(\delta(X) = 0) & \text{при} \quad \theta \in \Theta_2, \end{cases}$$

$$\mathbf{E}_\theta w(\delta(X), \theta') = \begin{cases} w_1 \mathbf{P}_\theta(\delta(X) = 0) & \text{при} \quad \theta \in \Theta_1, \ \theta' \in \Theta_2, \\ w_2 \mathbf{P}_\theta(\delta(X) = 1) & \text{при} \quad \theta \in \Theta_2, \ \theta' \in \Theta_1, \end{cases}$$

и неравенства в определении 1А означают, что

$$w_2 \mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant w_1 \mathbf{P}_{\theta_1}(\delta(X) = 0) \quad \text{при} \quad \theta_1 \in \Theta_1,$$
$$w_1 \mathbf{P}_{\theta_2}(\delta(X) = 0) \leqslant w_2 \mathbf{P}_{\theta_2}(\delta(X) = 1) \quad \text{при} \quad \theta_2 \in \Theta_2$$

или, что то же,

$$\mathbf{P}_{\theta_1}(\delta(X) = 1) \leqslant \frac{w_1}{w_1 + w_2}, \quad \mathbf{P}_{\theta_2}(\delta(X) = 1) \geqslant \frac{w_1}{w_1 + w_2}.$$

Отсюда следует, что

$$\sup_{\theta\in\Theta_1} \mathbf{E}_\theta\delta(X) \leqslant \inf_{\theta\in\Theta_2} \mathbf{E}_\theta\delta(X)$$

и что, стало быть, критерий δ будет несмещенным в смысле определений § 46. Наоборот, если справедливо последнее неравенство, то критерий δ будет несмещенным в смысле определения 1А при подходящем выборе функции потерь $w(\delta,\theta)$, например при $w_1/(w_1+w_2) = \sup\limits_{\theta\in\Theta_1} \mathbf{E}_\theta\delta(X)$.

Дополнительные примеры использования принципа несмещенности (помимо результатов § 46) можно найти в [65, 66].

3. Инвариантность. Мы видели, что пересечение полного класса, порожденного «достаточными» решениями, с классом несмещенных решений может состоять из единственной стратегии. Еще одним естественным классом стратегий, в котором может оказаться единственное неулучшаемое решение, является класс инвариантных решающих правил (ср. с § 28, 29, 47).

Определение инвариантной проблемы статистического решения связано с группами преобразований во всех трех пространствах, которые участвуют в определении статистической игры: в пространствах D, Θ и выборочном пространстве $\mathcal{X}^n$. В основе определения лежат измеримые взаимно однозначные преобразования g пространства $\mathcal{X}^n$ в себя, образующие некоторую группу G с групповой операцией, определенной как композиция: если $g_1 \in G$, $g_2 \in G$, то g_2g_1 определяется как преобразование $\boldsymbol{x} \to g_2(g_1\boldsymbol{x})$, которое снова обязано принадлежать G. Тождественное преобразование обозначим через e. Обратное к g преобразование g^{-1} определяется как преобразование, для которого $g^{-1}g = e$. Измеримость $g \in G$ означает, что gX вместе с X будет случайной величиной в $\mathcal{X}^n$.

С введенной группой G тесно связано понятие инвариантности семейства $\mathbf{P}_\theta$, которое было определено нами в § 29 и 47. Оно означает, что для каждых $g \in G$ и $\theta \in \Theta$ найдется элемент $\theta_g \in \Theta$ такой, что

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\theta_g}(X \in A). \tag{2}$$

Преобразования $\overline{g}$ пространства Θ в себя, определенные равенством $\overline{g}\theta = \theta_g$, при выполнении условия (A_0) образуют группу $\overline{G}$ (см. § 29).

В терминах математических ожиданий условие (2) означает, что для любой интегрируемой функции φ

$$\mathbf{E}_\theta\varphi(gX) = \mathbf{E}_{\overline{g}\theta}\varphi(X). \tag{3}$$

Определение 2. Проблема статистического решения, связанная со статистической игрой $(\mathcal{D},\Theta,w)$, $(X,\mathbf{P}_\theta)$, называется *инвариантной относительно группы* G, если семейство $\mathbf{P}_\theta$ инвариантно относительно G и функция потерь w инвариантна относительно G в следующем смысле: для любых $\delta \in D$, $g \in G$ найдется единственное $\delta' \in D$ такое, что

$$w(\delta,\theta) = w(\delta',\overline{g}\theta) \quad \text{при всех} \quad \theta \in \Theta. \tag{4}$$

Значение δ', однозначно определенное по g, обозначим через $g'\delta$.

Лемма 1. *Преобразования g' пространства D в себя, порожденные группой G, образуют группу G'.*

Доказательство. Покажем, что совокупность G' всех преобразований g' замкнута относительно композиции, причем справедливо равенство $g_2'g_1' = (g_2g_1)'$.

Действительно,

$$w(\delta,\theta) = w(g_1'\delta, \overline{g}_1\theta) = w(g_2'g_1'\delta, \overline{g}_2\overline{g}_1\theta) = w((g_2g_1)'\delta, (\overline{g_2g_1})\theta).$$

Так как $(\overline{g_2g_1}) = \overline{g}_2\overline{g}_1$, то в силу единственности $(g_2g_1)' = g_2'g_1'$. Лемма доказана. $\triangleleft$

Итак, с основной группой G преобразований g пространства $\mathcal{X}^n$ в себя связаны еще две группы $\overline{G}$ и G' преобразований пространств Θ и D в себя. Применение одновременно всех трех преобразований g, $\overline{g}$ и g' оставляет проблему решения неизменной (инвариантной). Поэтому естественно выбирать такие решающие правила, которые не менялись бы при переходе от одной эквивалентной проблемы решения к другой. Естественность такого подхода нами обсуждалась весьма детально в § 28, 29, 47.

Определение 3. Решающая функция $\delta(X)$ инвариантной проблемы решения называется *инвариантной*, если

$$\delta(gX) = g'\delta(X).$$

Рандомизированное инвариантное правило $\pi(X)$ определяется как любое распределение, сосредоточенное на инвариантных решающих правилах.

Примеры использования принципа инвариантности содержатся в уже цитированных нами § 28, 29, 47, где рассматривались эквивариантные оценки и инвариантные критерии. Отметим некоторое своеобразие с точки зрения общего подхода этих двух частных случаев.

В *проблеме оценивания* группа преобразований G' не вводилась вовсе. В этом случае множества D и Θ совпадают, и с самого начала предполагалось, что $g'\delta = \overline{g}\delta$. Поэтому эквивариантные оценки определялись с помощью равенства $\theta^*(gX) = \overline{g}\theta^*(X)$.

В *теории проверки гипотез* преобразование g' полагалось равным тождественному $g' = e$, так что инвариантный критерий π определялся соотношением $\pi(gX) = \pi(X)$.

В этом случае для инвариантности проблемы проверки двух гипотез $\{\theta \in \Theta_1\}$ и $\{\theta \in \Theta_2\}$ нужно предполагать также (см. (4)), что $g\Theta_i = \Theta_i$.

Наличием некоторой разницы в этих двух подходах и объясняется в какой-то мере использование двух разных терминов: «эквивариантность» (для оценок) и «инвариантность» (для проверки гипотез) для обозначения инвариантных решающих правил. Дополнительно к примерам инвариантных проблем решения, рассмотренным в гл. 2 и 3, приведем еще один.

Пример 1. Пусть $X \in \mathbf{\Phi}_{\alpha,\sigma^2}$. Здесь Θ есть полуплоскость $\{\theta = (\alpha,\sigma)\colon \sigma \geqslant 0\}$. Пусть D есть вещественная прямая $\mathbb{R}$ и пусть $w(\delta,\theta) = (\delta-\alpha)^2/\sigma^2$.

Рассмотрим группу G преобразований $g_{a,b}X = a + bX = (a + bx_1, \ldots, a + bx_n)$, где $b \neq 0$. Случайную величину $g_{a,b}X$ в $\mathcal{X}^n$ можно рассматривать, очевидно, как выборку из $\mathbf{\Phi}_{a+b\alpha,b^2\sigma^2}$. Следовательно, семейство $\mathbf{\Phi}_{\alpha,\sigma^2}$

инвариантно относительно G, если положить $\overline{g}_{a,b}\theta = (a + b\alpha, |b|\sigma)$. Функция потерь будет инвариантной, если мы положим $g'_{a,b}\delta = a + b\delta$, поскольку

$$w\bigl(g'_{a,b}\delta, \overline{g}_{a,b}\theta\bigr) = \frac{(a + b\delta - a - b\alpha)^2}{b^2\sigma^2} = w(\delta, \theta).$$

Таким образом, имеем инвариантную относительно G проблему решения. Инвариантные решающие функции $\delta(X)$: $\mathcal{X}^n \to \mathbb{R}$ должны обладать свойством

$$\delta(a + bX) = \delta(g_{a,b}X) = g'_{a,b}\delta(X) = a + b\delta(X). \tag{5}$$

Далее, нетрудно установить, что рассматриваемая проблема решения является инвариантной и относительно группы F всех перестановок f координат вектора X; при этом $\overline{f}$ и f' будут тождественными преобразованиями. Поэтому если потребовать, чтобы функция $\delta(X)$ была инвариантным решением и относительно F, то должно выполняться также

$$\delta(fX) = \delta(X). \tag{6}$$

Отметим, что класс функций, удовлетворяющих (5), (6), все еще достаточно широк: в него входят, например, все линейные формы

$$\delta(X) = \sum_{k=1}^{n} a_k \mathrm{x}_{(k)}, \quad \sum_{k=1}^{n} a_k = 1,$$

где $\mathrm{x}_{(1)}, \ldots, \mathrm{x}_{(n)}$ — вариационный ряд выборки X. Если использовать принцип несмещенности, то получим еще одно условие на коэффициенты a_k:

$$\sum_{k=1}^{n} a_k \mathbf{E}_\theta(\mathrm{x}_{(k)} - \alpha) = 0. \quad \triangleleft$$

При построении оптимальных инвариантных решений в теории оценивания и теории проверки статистических гипотез важную роль играют понятия, в известном смысле близкие друг другу: понятие *орбиты* в теории оценок и понятие *инварианта* в теории проверки гипотез. Напомним, что орбитой в пространстве Θ называется множество $\{\overline{g}\theta_0, \overline{g} \in \overline{G}\}$, где θ_0 есть некоторая точка из Θ. Другими словами, θ_1 и θ_2 принадлежат одной орбите, если существует $\overline{g} \in \overline{G}$ такое, что $\theta_1 = \overline{g}\theta_2$.

Аналогично можно определить орбиты в $\mathcal{X}^n$. Тогда инварианты по определению есть статистики, постоянные на орбитах в $\mathcal{X}^n$.

Понятие орбиты сохраняет свое значение и в общем случае.

Лемма 2. *Функция риска инвариантной проблемы решения для инвариантного решающего правила постоянна на орбите*:

$$W(\delta(\cdot), \theta) = W(\delta(\cdot), \overline{g}\theta)$$

при всех $\theta \in \Theta$, $\overline{g} \in \overline{G}$.

Доказательство. В силу инвариантности соответственно функции потерь, решающего правила и семейства $\mathbf{P}_\theta$ (см. (3), (4)) имеем

$$W(\delta(\cdot),\theta) = \mathbf{E}_\theta w(\delta(X),\theta) = \mathbf{E}_\theta w(g'\delta(X),\overline{g}\theta) =$$
$$= \mathbf{E}_\theta w(\delta(gX),\overline{g}\theta) = \mathbf{E}_{\overline{g}\theta} w(\delta(X),\overline{g}\theta) = W(\delta(\cdot),\overline{\theta}). \quad \triangleleft$$

Постоянство на орбите риска для рандомизированных инвариантных решающих правил следует из их определения и леммы 2.

Из леммы 2 вытекает, что в том случае, когда все пространство Θ является орбитой (т. е. $\Theta = \{\overline{g}\theta_0, \overline{g} \in \overline{G}\}$ при каком-нибудь θ_0; это имеет место, например, для преобразований сдвига), инвариантное решающее правило становится уравнивающим. Поэтому из леммы 2 и теорем 74.3–74.6 получаем немедленно следующее утверждение, устанавливающее важную связь между инвариантностью и минимаксностью.

Теорема 2. *Пусть пространство Θ является орбитой и существует априорное распределение $\mathbf{Q}$, для которого байесовская стратегия $\pi_{\mathbf{Q}}(X)$ является инвариантной. Тогда $\pi_{\mathbf{Q}}(X)$ — минимаксная стратегия.*

Из теоремы 75.3 вытекает также, что имеет место следующее обобщение теоремы 2.

Теорема 2А. *Пусть существует априорное распределение $\mathbf{Q}$, сосредоточенное на одной из орбит Θ_0, такое, что байесовская стратегия $\pi_{\mathbf{Q}}(X)$ является инвариантной. Тогда если при всех θ*

$$W(\pi_{\mathbf{Q}}(\cdot),\theta) \leqslant W(\pi_{\mathbf{Q}}(\cdot),\theta_0), \quad \theta_0 \in \Theta_0,$$

то $\pi_{\mathbf{Q}}(X)$ является минимаксной.

Этим критерием мы пользовались в § 49.

§ 78. Асимптотически оптимальные оценки при произвольной функции потерь

Многие из результатов гл. 2 об асимптотически оптимальных оценках и гл. 3 об асимптотически оптимальных критериях допускают обобщения на функции потерь весьма общего вида.

В этом параграфе остановимся на задачах теории оценивания и предположим, что $w(\delta,\theta) = w(\delta-\theta)$.

Сделаем сначала одно общее замечание. В гл. 2 мы видели, что в регулярном случае ($X \Subset \mathbf{P}_\theta$, $\mathbf{P}_\theta$ удовлетворяет условиям (RR); см. § 34, 36) все разумные оценки $\theta^* = \delta(X)$ параметра θ «сосредоточены» в $1/\sqrt{n}$-окрестности точки θ. Так, например, для асимптотически нормальных оценок $(\theta^*-\theta)\sqrt{n} \Subset \mathbf{\Phi}_{0,\sigma^2(\theta)}$. Отсюда следует, что при широких предположениях относительно функции $w(t)$ асимптотическое поведение риска $\mathbf{E}_\theta w(\theta^*-\theta)$ будет определяться свойствами функции $w(t)$ в окрестности точки $t=0$. Если $w(t)$ дважды непрерывно дифференцируема в нуле, $w'' > 0$, то при $t \to 0$

$$w(t) = \frac{w''(0)}{2}t^2 + o(t^2). \tag{1}$$

Это значит, что в интересующей нас области значений t (порядка $1/\sqrt{n}$) функция $w(t)$ будет вести себя так же, как квадратичная функция потерь $w_0(t) = ct^2$ при $c = w''(0)/2$, для которой были установлены результаты гл. 2. Если к тому же $w(t) < e^{\alpha|t|^2}$ при достаточно малом $\alpha > 0$ (см. теорему 36.6), то все эти результаты останутся в силе — перенесение их на случай функции $w(t)$ вида (1) — дело несложной техники, вполне доступной читателю.

В этом параграфе мы рассмотрим значительно более содержательное обобщение. Мы предположим, что функция потерь $w(\delta, \theta)$ зависит от n и представима в виде

$$w(\delta, \theta) = w_n(\delta - \theta) = w(\sqrt{n}\,(\delta - \theta)), \tag{2}$$

где функция $w(t) \geqslant 0$ определена во всем пространстве $\mathbb{R}^k$. Очевидно, что в этом случае будут существенными значения $w(t)$ во всей области значений t. Предположение (2) означает, что величина потерь чувствительна даже к небольшим отклонениям δ от θ. Мы будем предполагать, что функция w в (2) удовлетворяет следующим условиям.

1. $w(t) \leqslant e^{c|t|}$ *при некотором* $c > 0$.

Такая форма условия 1 несколько упрощает выкладки. На самом деле все результаты сохранятся, если потребовать, чтобы $w(t) \leqslant c_1 e^{\alpha|t|^2}$ при достаточно малом $\alpha > 0$.

В дальнейшем существенную роль будет играть функция

$$V_{\sigma^2}(s) = \int w(s-u)e^{-\frac{1}{2}u\sigma^2 u^\top}\,du,$$

где σ^2 — некоторая положительно определенная матрица вторых моментов. Функцию $V_{\sigma^2}(s)$ можно интерпретировать как

$$V_{\sigma^2}(s) = \frac{(2\pi)^{k/2}}{\sqrt{|\sigma^2|}}\mathbf{E}w(s-\xi), \quad \xi \Subset \mathbf{\Phi}_{0,\sigma^{-2}}.$$

Так как

$$V_{\sigma^2}(s) = \int w(v)e^{-\frac{1}{2}(s-v)\sigma^2(s-v)^\top}\,dv,$$

то эта функция будет аналитической функцией переменных s и σ^2.

Нам понадобятся следующие условия.

2. *Функция* $V_{\sigma^2}(s)$ *достигает своего минимального значения по* s *в единственной точке, которую мы обозначим через* b_w.

3. $b_w = 0$.

4. *Функция* $w(t)$ *непрерывна.*

Условие 2 будет заведомо выполнено, если $w(s) \neq \text{const}$ есть выпуклая вниз функция. В этом случае, очевидно, $V_{\sigma^2}(s)$ также будет выпуклой и не будет содержать «линейных» участков (т. е. матрица вторых производных будет всюду положительно определенной).

Условие 3 будет выполнено, если

$$V'_{\sigma^2}(0) = -\int uw(u)e^{-\frac{1}{2}u\sigma^2u^\top}\,du = 0,$$

что всегда имеет место для симметричных функций $w(u) = w(-u)$.

Значение b_w можно было бы назвать *смещением* функции потерь w. Оно удовлетворяет уравнению $V'_{\sigma^2}(b_w) = 0$. Условие 3 о том, что $b_w = 0$, не является существенным и лишь упрощает изложение, которое читатель может без труда перенести и на случай $b_w \neq 0$. Изменения в формулировках теорем, которые при этом произойдут, будут проиллюстрированы в замечании 2 к теореме 1.

Напомним теперь, во что перейдут здесь определения оптимальных стратегий § 74, 75. Оценка $\theta^*_{\mathbf{Q}}$ будет *байесовской* относительно априорного распределения $\mathbf{Q}$ с плотностью q относительно меры Лебега (и функции потерь w_n), если

$$\int W\left(\theta^*_{\mathbf{Q}}, t\right)q(t)\,dt = \min_{\theta^*}\int W(\theta^*, t)q(t)\,dt, \tag{3}$$

где $W(\theta^*, t) = \mathbf{E}_t w_n(\theta^* - t)$. Здесь интеграл в правой части (3) можно записать в виде безусловного математического ожидания $\mathbf{E}w_n(\theta^* - \theta)$, где усреднение берется по распределению с плотностью $f_t(\boldsymbol{x})q(t)$.

Оценка $\overline{\theta}^*$ будет *минимаксной*, если для любой другой оценки θ^*

$$\sup_t W(\overline{\theta}^*, t) \leqslant \sup_t W(\theta^*, t).$$

Изложенное делает естественными следующие определения, совершенно аналогичные определениям § 21.

Определение 1. Оценку θ^* будем называть *асимптотически байесовской*, если

$$\limsup_{n\to\infty}\left[\mathbf{E}w_n(\theta^* - \theta) - \mathbf{E}w_n\left(\theta^*_{\mathbf{Q}} - \theta\right)\right] \leqslant 0, \tag{4}$$

где $\theta^*_{\mathbf{Q}}$ — байесовская оценка.

Определение 2. Оценку θ^*_1 будем называть *асимптотически минимаксной*, если для любой другой оценки θ^*

$$\limsup_{n\to\infty}\left[\sup_{t\in\Theta_0} W\left(\theta^*_1, t\right) - \sup_{t\in\Theta_0} W(\theta^*, t)\right] \leqslant 0, \tag{5}$$

где Θ_0 — любое замкнутое подмножество, лежащее внутри Θ.

При изучении асимптотически оптимальных оценок в этом параграфе мы будем оперировать лишь понятиями, введенными в определениях 1 и 2. Это составляет некоторое отличие от гл. 2, где присутствовали также асимптотически эффективные оценки. Их отсутствие здесь объясняется тем, что для произвольных функций потерь w мы не располагаем неравенствами типа Рао–Крамера для $\inf\limits_{\theta^*\in K_0} W(\theta^*, \theta)$ (K_0 есть класс несмещенных оценок), с

помощью которого можно было по значению $W(\theta^*, \theta)$ судить о качестве θ^* и выделять, в частности, эффективные (и асимптотически эффективные) оценки, т. е. оценки, равномерно наилучшие в классе K_0.

Следующие утверждения устанавливают, что оценка максимального правдоподобия является, как и в условиях гл. 2, асимптотически байесовской и асимптотически минимаксной. Кроме того, мы получим *асимптотическую* нижнюю границу для функции риска при произвольной функции потерь w (неравенство Рао–Крамера дает *точную* нижнюю границу). Во всех последующих трех теоремах мы предполагаем, что выполнено условие (RR).

Теорема 1. *Пусть* $X \in \mathbf{P}_\theta$, $\widehat{\theta}^*$ *есть о. м. п., а* $\theta^*_{\mathbf{Q}}$ *есть байесовская оценка, соответствующая функции потерь* w (см. (2)), *удовлетворяющей условиям* 1–3, *и априорному распределению* $\mathbf{Q}$ *с ограниченной плотностью* q *относительно меры Лебега. Тогда*

$$\left|\theta^*_{\mathbf{Q}} - \widehat{\theta}^*\right|\sqrt{n} \underset{P_\theta}{\longrightarrow} 0, \tag{6}$$

$$\left(\theta^*_{\mathbf{Q}} - \theta\right)\sqrt{n} \in \mathbf{\Phi}_{0, I^{-1}(\theta)} \tag{7}$$

равномерно по $\theta \in \Theta_0$, Θ_0 — *любое замкнутое подмножество, лежащее внутри* Θ, *на котором* $q(\theta) > q_0 > 0$ *непрерывна.*

Если, кроме того, функция w *удовлетворяет условию* 4, *то*

$$\mathbf{E}w_n\left(\theta^*_{\mathbf{Q}} - \theta\right) = \mathbf{E}w\left(\sqrt{n}\left(\theta^*_{\mathbf{Q}} - \theta\right)\right) \to \mathbf{E}w(\eta_\theta) = \mathbf{E}\frac{\sqrt{I(\theta)}}{(2\pi)^{k/2}} V_{I(\theta)}(0), \tag{8}$$

где $\eta_\theta \in \mathbf{\Phi}_{0, I^{-1}(\theta)}$, $\theta \in \mathbf{Q}$; $\mathbf{E}$, *как и прежде, означает безусловное математическое ожидание с плотностью* $f_t(\boldsymbol{x})q(t)$ $(X \in \mathbf{P}_\theta, \theta \in \mathbf{Q})$.

Замечание 1. Наряду со сходимостью (6) можно установить также сходимость почти наверное относительно $\mathbf{P}_\theta$.

Замечание 2. Если w таково, что смещение $b_w \neq 0$, то утверждение теоремы 1 полностью сохранится, если $\theta^*_{\mathbf{Q}}$ в (6)–(8) заменить на $\theta^*_{\mathbf{Q}} - b_w/\sqrt{n}$. Таким образом, b_w имеет смысл *асимптотического смещения величины* $\left(\theta^*_{\mathbf{Q}} - \theta\right)\sqrt{n}$.

Теорема 2. *Пусть функция* w *удовлетворяет условиям* 1–4. *Тогда для любой оценки* θ^*

$$\liminf_{n\to\infty} \sup_{t\in\Theta_0} \mathbf{E}_t w_n(\theta^* - t) \geqslant \sup_{t\in\Theta_0} \mathbf{E}w(\eta_t), \qquad \eta_t \in \mathbf{\Phi}_{0, I^{-1}(t)}. \tag{9}$$

Любая оценка θ^*, *для которой*

$$\mathbf{E}_t w_n(\theta^* - t) \to \mathbf{E}w(\eta_t) \tag{10}$$

равномерно по t, *является асимптотически минимаксной.*

Теорема 3. *Пусть* $X \in \mathbf{P}_\theta$ *и функция* w *удовлетворяет условиям* 1–4. *Тогда оценка максимального правдоподобия* $\widehat{\theta}^*$ *является асимптотически минимаксной и асимптотически байесовской для любого априорного распределения* $\mathbf{Q}$ *с непрерывной положительной в точке* θ *плотностью* q.

Все эти утверждения вполне аналогичны соответствующим утверждениям гл. 2. Они делают правдоподобным предположение, что и для произвольной функции потерь w, удовлетворяющей условиям 1–4, о. м. п. является асимптотически равномерно наилучшей в классе асимптотически несмещенных оценок (ср. с § 33, 34, 36).

Доказательство теоремы 1. Байесовская оценка определяется в силу байесовского принципа как значение $\theta^*_{\mathbf{Q}}$, обладающее свойством

$$\int w_n(\theta^*_{\mathbf{Q}} - t) q(t/X)\, dt = \min_{u \in \Theta} \int w_n(u-t) q(t/X)\, dt = \\ = \min_{u \in \Theta} \int w(\sqrt{n}(u-\theta) - \sqrt{n}(t-\theta)) \frac{q(t) f_t(X)}{\int q(v) f_v(X)\, dv}\, dt.$$

Это значит, что в качестве $(\theta^*_{\mathbf{Q}} - \theta)\sqrt{n} \equiv u^*_{\mathbf{Q}}$ можно взять любое значение s, при котором достигается $\min\limits_s U(s)$,

$$U(s) = \int w(s-v) q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv, \tag{11}$$

где, как и прежде, $Z(t) = f_{\theta+t}(X)/f_\theta(X)$.

Нам понадобятся утверждения об асимптотическом поведении $U(s)$. В § 36, 37 было установлено (теорема 36.5), что при выполнении условий (RR)

$$U(u^*) = e^{Y(u^*)} q(\widehat{\theta}^*)(V_{I(\theta)}(0) + \varepsilon_n(X, \theta)), \tag{12}$$

где $\varepsilon_n(X, \theta) \underset{P_\theta}{\longrightarrow} 0$ равномерно по θ (мы заменили здесь $\dfrac{(2\pi)^{k/2}}{\sqrt{I(\theta)}} \mathbf{E} w(\eta)$ на $V_{I(\theta)}(0)$; $q(\theta)$ на $q(\widehat{\theta}^*)$)

Заметим теперь, что

$$\mathbf{P}\left(\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \widehat{\theta}^*| \geqslant \varepsilon\right) = \mathbf{P}\left(|u^*_{\mathbf{Q}} - u^*| \geqslant \varepsilon\right) \leqslant \mathbf{P}\left(\min_{|s-u^*| \geqslant \varepsilon} U(s) \leqslant U(u^*)\right). \tag{13}$$

Поскольку асимптотическое представление для $U(u^*)$ мы имеем, то нам надо оценить здесь значение $U(s)$. Из теорем 36.4, 37.3 следует, что для произвольной последовательности $\varepsilon_n \to 0$ при $|v| < \varepsilon_n \sqrt{n}$ выполняется

$$\ln Z\left(v\sqrt{n}\right) = Y(u^*) - \frac{1}{2}(v-u^*) I(\theta) (v-u^*)^{\mathsf{T}} (1 + \varepsilon_n(X, \theta, u)),$$

$|\varepsilon_n(X, \theta, u)| \leqslant \varepsilon_n^{(1)}(X, \theta) \underset{P_\theta}{\longrightarrow} 0$ равномерно по θ. Но

$$U(s) \geqslant U_n(s) \equiv \int\limits_{|v-u^*| \leqslant \varepsilon_n \sqrt{n}} w(s-v) q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Рассмотрим множество

$$A_n = \left\{ \varepsilon_n^{(1)}(X,\theta) < \rho,\ \inf_{|v-u^*|\leqslant \varepsilon_n\sqrt{n}} q\left(\theta + \frac{v}{\sqrt{n}}\right) > q(\widehat{\theta}^*)(1-\rho) \right\},$$
$$\rho > 0,$$

обладающее очевидным свойством

$$\mathbf{P}_\theta(A_n') \to 1. \tag{14}$$

Равномерно по θ на этом множестве

$$U_n(s) \geqslant (1-\rho) q(\widehat{\theta}^*) e^{Y(u^*)} \times$$
$$\times \int\limits_{|v-u^*|\leqslant \varepsilon_n\sqrt{n}} w(s-v) \exp\left\{ -\frac{1}{2}(v-u^*) I(\theta)(v-u^*)^\top (1+\rho) \right\} dv =$$
$$= (1-\rho) q(\widehat{\theta}^*) e^{Y(u^*)} [V_{I(\theta)(1+\rho)}(s-u^*) - r_n(s)], \tag{15}$$

где по условию 1

$$r_n(s) = \int\limits_{|v-u^*|>\varepsilon_n\sqrt{n}} w(s-v) \exp\left\{ -\frac{1}{2}(v-u^*) I(\theta)(v-u^*)^\top (1+\rho) \right\} dv \leqslant$$
$$\leqslant e^{c\sqrt{n}d} \frac{(2\pi)^{k/2}}{\sqrt{I(\theta)|c}} \mathbf{P}(|\eta| > \varepsilon_n\sqrt{n}), \quad \eta \Subset \mathbf{\Phi}_{0,I(\theta)(1+\rho)},$$

d — диаметр области Θ. Аналогично лемме 33.1 нетрудно убедиться, что

$$\mathbf{P}(|\eta| > \varepsilon_n\sqrt{n}) \leqslant e^{-\alpha n \varepsilon_n^2}, \quad \alpha > 0.$$

Выбирая $\varepsilon_n = n^{-1/9}$, мы получим, что при всех s и достаточно больших n

$$r_n(s) \leqslant e^{-n^{3/4}}. \tag{16}$$

Воспользуемся теперь условиями 2, 3, в силу которых

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)}(s-u^*) \geqslant V_{I(\theta)}(0) + 4\tau, \quad \tau = \tau(\varepsilon) > 0.$$

В силу аналитических свойств $V_{\sigma^2}(s)$ мы получим, что при достаточно малых ρ

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)(1+\rho)}(s-u^*) \geqslant V_{I(\theta)}(0) + 3\tau,$$

и в силу (15), (16) при $X \in A_n$ и достаточно больших n

$$\min_{|s-u^*|\geqslant\varepsilon} U_n(s) \geqslant (1-\rho) q(\widehat{\theta}^*) e^{Y(u^*)} [V_{I(\theta)}(0) + 2\tau].$$

Используя (12), (13), окончательно находим

$$\mathbf{P}_\theta\big(\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \widehat{\theta}^*| \geqslant \varepsilon\big) \leqslant \mathbf{P}_\theta\left(\min_{|s-u^*|\geqslant\varepsilon} U_n(s) \leqslant U(u^*)\right) \leqslant$$

$$\leqslant \mathbf{P}_\theta(X \notin A_n) + \mathbf{P}_\theta((1-\rho)[V_{I(\theta)}(0) + 2\tau] \leqslant V_{I(\theta)} + \varepsilon_n(X,\theta)).$$

Выбирая дополнительно ρ столь малым, чтобы выполнялось $(1-\rho)2\tau - \rho V_{I(\theta)}(0) \geqslant \tau$, получим

$$\mathbf{P}\big(\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \widehat{\theta}^*| \geqslant \varepsilon\big) \leqslant \mathbf{P}_\theta(X \notin A_n) + \mathbf{P}_\theta(\varepsilon_n(X,\theta) > \tau) \to 0$$

при $n \to \infty$. В силу (12), (14) утверждение (6) доказано.

Из (6) и теорем § 37 вытекает (7). Докажем теперь соотношение (8). В силу (7) и свойства 4

$$w\big(\sqrt{n}\,(\theta^*_{\mathbf{Q}} - \theta)\big) \Rightarrow w(\eta_\theta), \quad \eta_\theta \Subset \mathbf{\Phi}_{0,I^{-1}(\theta)}.$$

По лемме Фату

$$\liminf_{n\to\infty} \mathbf{E}_t w\big(\sqrt{n}\,(\theta^*_{\mathbf{Q}} - t)\big) \geqslant \mathbf{E}w(\eta_t),$$

$$\liminf_{n\to\infty} \mathbf{E}w\big(\sqrt{n}\,(\theta^*_{\mathbf{Q}} - \theta)\big) \geqslant \int q(t)\mathbf{E}w(\eta_t)\,dt = \mathbf{E}w(\eta_\theta).$$

С другой стороны, по определению $\theta^*_{\mathbf{Q}}$

$$\mathbf{E}w\big(\sqrt{n}\,(\theta^*_{\mathbf{Q}} - (\theta)\big) \leqslant \mathbf{E}w(\sqrt{n}\,(\widehat{\theta}^* - \theta)) \to \mathbf{E}w(\eta_\theta).$$

Последнее соотношение вытекает из равномерной сходимости $\mathbf{E}_t w \times$ $\times(\sqrt{n}(\widehat{\theta}^* - t)) \to \mathbf{E}w(\eta_t)$, доказанной в § 37. Теорема доказана. $\triangleleft$

Доказательство теоремы 2. Возьмем распределение $\mathbf{Q}$, сосредоточенное на Θ_0, с ограниченной плотностью $q(t) > 0$ при $t \in \Theta_0$, и пусть $\theta^*_{\mathbf{Q}}$ — байесовская оценка, отвечающая $\mathbf{Q}$. Тогда для любой оценки θ^*

$$\sup_{t\in\Theta_0} \mathbf{E}_\theta w_n(\theta^* - t) \geqslant \int\limits_{\Theta_0} \mathbf{E}_t w_n(\theta^* - t) q(t)\,dt \geqslant$$

$$\geqslant \int\limits_{\Theta_0} \mathbf{E}_t w_n\big(\theta^*_{\mathbf{Q}} - t\big) q(t)\,dt = \mathbf{E}w_n\big(\theta^*_{\mathbf{Q}} - \theta\big).$$

По лемме Фату в силу (8)

$$\liminf_{n\to\infty} \sup_{t\in\Theta_0} \mathbf{E}_t w_n(\theta^* - t) \geqslant \liminf_{n\to\infty} \mathbf{E}w_n\big(\theta^*_{\mathbf{Q}} - \theta\big) \geqslant \mathbf{E}w(\eta_\theta) = \int\limits_{\Theta_0} \mathbf{E}w(\eta_t) q(t)\,dt.$$

Так как функция $\mathbf{E}w(\eta_t) = \dfrac{\sqrt{I(t)}}{(2\pi)^{k/2}} V_{I(t)}(0)$ непрерывна относительно t, то выбором $q(t)$ интеграл

$$\int\limits_{\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0) q(t)\, dt$$

может быть сделан сколь угодно близким к $\sup\limits_{t\in\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0) = \sup\limits_{t\in\Theta_0} \mathbf{E}w(\eta_t)$. Это доказывает (9).

Пусть теперь оценка θ_1^* обладает свойством (10) и θ^* — любая другая оценка. Тогда в силу (9) и равномерности сходимости (10)

$$\limsup_{n\to\infty} \bigl[\sup_{t\in\Theta_0} \mathbf{E}_t w_n\bigl(\theta_1^* - t\bigr) - \sup_{t\in\Theta_0} \mathbf{E}_t w_n(\theta^* - t)\bigr] \leqslant$$

$$\leqslant \sup_{t\in\Theta_0} \lim_{n\to\infty} \mathbf{E}_t w_n\bigl(\theta_1^* - t\bigr) - \sup_{t\in\Theta_0} \mathbf{E}w(\eta_t) = 0.$$

Неравенство (5) определения асимптотической минимаксности, а вместе с ним и теорема 2 доказаны. $\triangleleft$

Доказательство теоремы 3. Асимптотическая минимаксность $\widehat{\theta}^*$ следует из того, что для о. м. п. $\widehat{\theta}^*$ по теореме 37.4 справедливо (10).

Асимптотическая байесовость $\widehat{\theta}^*$ следует из того, что при $\theta^* = \widehat{\theta}^*$ выполнено (4), ибо для $\widehat{\theta}^*$ имеет место равномерная сходимость (10) и, стало быть,

$$\lim_{n\to\infty} \mathbf{E}w_n\bigl(\widehat{\theta}^* - \theta\bigr) = \lim_{n\to\infty} \int \mathbf{E}_t w_n(\widehat{\theta}^* - \theta) q(t)\, dt =$$

$$= \mathbf{E}w(\eta_\theta) = \lim_{n\to\infty} \mathbf{E}w_n\bigl(\theta_{\mathbf{Q}}^* - \theta\bigr).$$

Последнее неравенство вытекает из (8). Теорема доказана. $\triangleleft$

Утверждение теоремы 1 можно усилить, если потребовать дополнительно, чтобы функция $w(t)$ росла достаточно быстро. Именно, положим $w_N = \min\limits_{|t|>N} w(t)$, $W_M = \max\limits_{|t|\leqslant M} w(t)$ и рассмотрим условие

5. *Существует* $\gamma < 1$ *такое, что* $w_N > 2W_{\gamma N}$ *при всех достаточно больших* N.

Если $w(t)$ растет при $|t| \to \infty$ как степенная или экспоненциальная функция, то условие 5 выполнено.

Теорема 4. *Если выполнены условия* 1, 5, $q(t) > q_0 > 0$ *на замкнутом множестве* Θ_0 *и* $q(t) \leqslant q_m < \infty$, *то при некоторых* $c < \infty$, $\alpha > 0$, *не зависящих от* t,

$$\mathbf{P}_t\bigl(\sqrt{n}\,(\theta_{\mathbf{Q}}^* - t) > N\bigr) \leqslant ce^{-\alpha N^2}, \quad t \in \Theta_0.$$

Отсюда и из теоремы 1 следует, что для любой непрерывной функции $v(t)$ такой, что $|v(t)| \leqslant e^{-\alpha N^2/2}$, справедливо

$$\mathbf{E}_t v\left(\sqrt{n}\left(\theta^*_{\mathbf{Q}} - t\right)\right) \to \mathbf{E}v(\eta_t), \quad t \in \Theta_0.$$

Доказательство теоремы 4. Обозначим

$$u(r) = \int\limits_{|v|\geqslant r} w(-v) q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv$$

(это есть часть интеграла $U(0)$ по области $|v| \geqslant r$). Для доказательства теоремы 4 нам понадобится

Лемма 1. *Если $w(t)$ удовлетворяет условию* 1, $q_m = \max\limits_u q(u) < \infty$, *то при некоторых $\beta > 0$, $a < \infty$, не зависящих от θ, и всех $0 < \varepsilon < 1$*

$$\mathbf{P}_\theta(u(r) > \varepsilon) \leqslant \frac{a}{\varepsilon} e^{-\beta r^2}.$$

Это неравенство сохраняется при $w(t) \equiv 1$.

Доказательство. Имеем

$$\mathbf{P}_\theta(u(r) > \varepsilon) \leqslant \mathbf{P}_\theta\left(\sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) > 1\right) + \mathbf{P}_\theta\left(u(r) > \varepsilon,\ \sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \leqslant 1\right).$$

Оценка для первого слагаемого дана в теореме 33.2, в силу которой это слагаемое не превосходит $c_1 e^{-r^2\beta}$, $\beta > 0$. Второе слагаемое не превосходит

$$\mathbf{P}_\theta\left(\int\limits_{|v|\geqslant r} w(-v) q\left(\theta + \frac{v}{\sqrt{n}}\right) Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) dv > \varepsilon\right). \tag{17}$$

Так как в силу теоремы 33.1

$$\mathbf{E}_\theta Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) \leqslant e^{-2|v|^2\beta}, \quad \beta > 0,$$

то математическое ожидание интеграла в (17) не будет превосходить (см. лемму 33.1)

$$q_m \int\limits_{|v|\geqslant r} e^{c|v|} e^{-2|v|^2\beta}\, dv \leqslant c_2 e^{r^2\beta}.$$

Поэтому в силу неравенства Чебышева вероятность (17) не превосходит $c_2 e^{-r^2\beta}/\varepsilon$. Лемма доказана. $\triangleleft$

Обозначим через $u_1(r)$ значение интеграла $u(r)$ при $w(t) \equiv 1$:

$$u_1(r) = \sum_{|v|\geqslant r} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Лемма 2. *Если $q(\theta) > 0$ на замкнутом множестве Θ_0, то при некотором $b < \infty$, не зависящем от θ, любом $\varepsilon > 0$ и всех достаточно больших n*

$$\mathbf{P}_\theta(u_1(0) < \varepsilon) \leqslant b\varepsilon^2, \quad \theta \in \Theta_0.$$

Доказательство. При всех достаточно больших n имеем

$$u_1(0) \geqslant \int\limits_{|v|\leqslant 1} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv \geqslant$$

$$\geqslant q_0 \int\limits_{|v|\leqslant 1} \exp\left\{L\left(X, \theta + \frac{v}{\sqrt{n}}\right) - L(X,\theta)\right\} dv = q_0 \int\limits_{|v|\leqslant 1} \exp\left\{(v, \zeta_n) + \frac{1}{2} v\gamma_n v^\top\right\} dv,$$

где

$$q_0 = \min_{t\in\Theta_0} q(t) > 0, \quad \zeta_n = \frac{1}{\sqrt{n}} L'(X,t), \quad \gamma_n = \frac{1}{n}\left\|L''_{ij}(X,\widetilde{\theta})\right\|,$$

$\widetilde{\theta} = \theta + \rho v n^{-1/2}$, $|\rho| \leqslant 1$. (Здесь L' есть вектор производных логарифмической функции правдоподобия, L''_{ij} — частные производные второго порядка.) Так как $|(v,\zeta_n)| \leqslant |v||\zeta_n|$ и так как в силу условий (RR)

$$|v\gamma_n v^\top| \leqslant \frac{1}{n}\sum_{i=1}^n l(\mathrm{x}_i) \sum_{l,j=1}^k |v_l v_j| \leqslant \frac{k|v|^2}{n} L_n,$$

где $L_n = \sum\limits_{i=1}^n l(\mathrm{x}_i)$, то на множестве $A = \{|\zeta_n| \leqslant 1/\varepsilon,\ L_n \leqslant n/\varepsilon^2 k\}$ справедливо

$$u_1(0) \geqslant q_0 \int\limits_{|v|\leqslant 1} \exp\left\{-\frac{|v|}{\varepsilon} - \frac{|v|^2}{2\varepsilon^2}\right\} dv \geqslant q_0\varepsilon \int\limits_{|s|\leqslant\varepsilon^{-1}} \exp\left\{-|s| - \frac{|s|^2}{2}\right\} ds \geqslant c_1\varepsilon.$$

Это означает, что имеет место вложение $\{u_1(0) < c_1\varepsilon\} \subset \overline{A}$. Так как

$$\mathbf{P}_\theta(\overline{A}) \leqslant \mathbf{P}_\theta(|\zeta_n| \geqslant \varepsilon^{-1}) + \mathbf{P}_\theta\left(L_n > \frac{n}{\varepsilon^2 k}\right) \leqslant \varepsilon^2 \mathbf{E}_\theta|\zeta_n|^2 + \frac{\varepsilon^2 k}{n}\mathbf{E}_\theta L_n,$$

$$\mathbf{E}_\theta|\zeta_n|^2 = \sum_{i=1}^k I_{ii}(\theta), \quad \mathbf{E}_\theta L_n = n\mathbf{E}_\theta l(\mathrm{x}_1),$$

то

$$\mathbf{P}_\theta(\overline{A}) \leqslant c_2\varepsilon^2.$$

Лемма доказана. ◁

Доказательство теоремы 4. Обозначим через M_n множество точек s, в которых достигается $\min U(s)$ (т. е. множество точек $\left(\theta^*_{\mathbf{Q}} - \theta\right)\sqrt{n}$; см. (11)) *). Тогда

$$\{M_n \subset D\} = \left\{\min_{s\in D} U(s) < \min_{s\notin D} U(s)\right\}. \tag{18}$$

*) Вместо M_n можно было бы рассматривать, например, наименьшую по норме точку, в которой достигается $\min U(s)$.

Стало быть,

$$\{\sqrt{n}|\theta^*_{\mathbf{Q}} - \theta| > 2N\} = \left\{\min_{|s|>2N} U(s) < \min_{|s|\leqslant 2N} U(s)\right\} \subset \min_{|s|\geqslant 2N} U(s) < U(0)\}.$$

Здесь

$$\min_{|s|\geqslant 2N} U(s) \geqslant w_n \int\limits_{|u|<N} q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) du = w_N(u_1(0) - u_1(N)),$$

$$w_N \equiv \min_{\substack{|s|>2N \\ |u|<N}} w(s-u) = \min_{|t|>N} w(t).$$

Далее,

$$U(0) = \int w(-u) q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) du \leqslant (u_1(0) - u_1(M))W_M + u(M),$$

где $W_M = \max\limits_{|t|\leqslant M} w(t)$. Отсюда получаем

$$\{\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \theta| > 2N\} \subset \{w_N(u_1(0) - u_1(N)) < W_M(u_1(0) - u_1(M)) + u(M)\} \subset$$
$$\subset \left\{\left(\frac{w_n}{W_M} - 1\right) u_1(0) < \frac{u(M)}{W_M} + \frac{u_1(N)w_N}{W_M} + u_1(M)\right\}.$$

В силу условия 5 выберем $M = \gamma N$, $\gamma < 1$, так, чтобы $w_N > 2W_M$ при всех достаточно больших N. Кроме того, воспользуемся неравенствами $W_M > 2$ (при достаточно большом M), $w_N < w(N) < e^{cN}$. Тогда, очевидно,

$$\{\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \theta| > 2N\} \subset \{u_1(0) < u(\gamma N) + u_1(N)e^{cN}\}. \tag{19}$$

В силу леммы 1 находим

$$\mathbf{P}_\theta\left(u(\gamma N) > \frac{1}{2}e^{-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2\gamma^2 + \alpha N^2},$$

$$\mathbf{P}_\theta\left(u_1(N) > \frac{1}{2}e^{-cN-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2 + \alpha N^2 + cN}.$$

Выбирая $\alpha < \frac{1}{2}\beta\gamma^2$, мы получим, что при достаточно больших N из (19) следует

$$\mathbf{P}_\theta(\sqrt{n}\,|\theta^*_{\mathbf{Q}} - \theta| > 2N) \leqslant 4ae^{-\alpha N^2} + \mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}).$$

Нам остается лишь воспользоваться леммой 2, в силу которой

$$\mathbf{P}_\theta(u_1(0) < e^{-\alpha N^2}) \leqslant be^{-2\alpha N^2}.$$

Теорема доказана. $\triangleleft$

§ 79. Оптимальные статистические критерии при произвольной функции потерь. Критерий отношения правдоподобия как асимптотически байесовское решение

1. Свойства оптимальности статистических критериев при произвольной функции потерь. Мы видели в предыдущих двух параграфах, что многие главные результаты теории оценивания качественно сохраняются при переходе к более общим задачам статистического решения с потерями $w(\delta, \theta)$, $\delta \in D \subset \mathbb{R}^k$, $\theta \in \Theta \subset \mathbb{R}^k$, отличными от квадратичных.

Такая же картина имеет место и в теории проверки гипотез. Мы видели в § 76, что оптимальные решающие правила для игр с конечными множествами D и Θ и произвольной функцией потерь имеют ту же форму, что и оптимальные критерии для проверки конечного числа простых гипотез, рассмотренные в § 41. Результаты § 45, 47, 49, 51, 53–55 также в основном сохранятся. В частности, теоремы о р. н. м. к. в § 45–47 перейдут в утверждения о равномерно наилучших стратегиях в соответствующих статистических играх ($\Theta \subset \mathbb{R}^k$, $D = \{\delta_1, \delta_2\}$ двухточечно), в которых, однако, функция потерь $w(\delta_i, \theta) = w_i(\theta)$, $w_i(\theta) = 0$ при $\theta \in \Theta_i$, $i = 1, 2$, уже будет не обязательно статистической ($w_i(\theta) = 1$ при $\theta \notin \Theta_i$), а будет лишь удовлетворять некоторым весьма общим условиям (например, свойствам монотонного возрастания $w_i(\theta)$ при удалении θ от Θ_i). Роль классов K_ε, в которых мы искали р. н. м. к., будут играть теперь классы решающих функций $\pi(X)$ с фиксированным максимальным значением ε «потерь первого рода»:

$$\varepsilon = \sup_{\theta \in \Theta_1} W(\pi(\cdot), \theta) = \sup_{\theta \in \Theta_1} w_2(\theta) \mathbf{E}_\theta \pi(X, \delta_2). \tag{1}$$

Минимизироваться будет значение «потерь второго рода»:

$$W(\pi(\cdot), \theta) = w_1(\theta) \mathbf{E}_\theta \pi(X, \delta_1) \quad \text{при} \quad \theta \in \Theta_2. \tag{2}$$

Здесь $\pi(X, \delta_i)$ означает вероятность принятия решения δ_i критерием π. Для сокращения записи положим, следуя гл. 3, $\pi(X, \delta_2) = \pi(X)$, так что $\pi(X, \delta_1) = 1 - \pi(X)$. Обозначение одним символом $\pi(X)$ критерия и числа $\pi(X, \delta_2)$ удобно и, как мы видели в гл. 3, к недоразумениям не приводит.

В (1), (2) ищутся экстремумы выражений, которые отличаются от соответствующих выражений для статистических функций потерь лишь не зависящими от $\pi(X)$ множителями. Если эти множители обладают естественным свойством монотонности, то изложение § 45–47, 49, 51 при переходе к задаче, определяемой (1), (2), существенно не изменится.

По существу мало изменятся и результаты асимптотического характера в § 53–55. В этом параграфе мы остановимся более подробно на обобщении на случай произвольной функции потерь результатов § 53 и убедимся, что это обобщение действительно не требует никаких дополнительных усилий.

2. К. о. п. как асимптотически байесовский критерий. Рассмотрим статистическую игру $(\mathcal{D}, \Theta, W)$, в которой множество Θ континуально и есть выпуклое компактное множество в $\mathbb{R}^k$, а множество D стратегий статистика

двухточечно: $D = \{\delta_1, \delta_2\}$. Функция потерь $w(\delta, \theta)$ имеет вид

$$w(\delta_1, \theta) = \begin{cases} 0, & \theta = \theta_1, \\ w_1(\theta), & \theta \neq \theta_1, \end{cases} \quad w(\delta_2, \theta) = \begin{cases} w_2, & \theta = \theta_1, \\ 0, & \theta \neq \theta_1, \end{cases}$$

где θ_1 — фиксированная внутренняя точка Θ. При $w_2 = w_1(\theta) \equiv 1$ это соответствует задаче проверки простой гипотезы $H_1 = \{\theta = \theta_1\}$ против дополнительной альтернативы $H_2 = \{\theta \neq \theta_1\}$.

Чтобы, пользуясь байесовским принципом, найти вид байесовского решения, рассмотрим обычную игру (D, Θ, w) и предположим, что на Θ задано распределение $\mathbf{Q}$ такое, что $q = \mathbf{Q}(\{\theta_1\}) > 0$ (полная байесовская постановка задачи). Обозначим $\mathbf{Q}_2 = \dfrac{\mathbf{Q} - q\mathbf{I}_{\theta_1}}{1 - q}$, где $\mathbf{I}_\theta$ есть вырожденное распределение, сосредоточенное в точке θ. Тогда

$$\widetilde{w}(\delta_1, \mathbf{Q}) = (1 - q) \int w_1(t) \mathbf{Q}_2(dt), \quad \widetilde{w}(\delta_2, \mathbf{Q}) = q w_2.$$

Это означает, что байесовская стратегия $\pi_{\mathbf{Q}}(\delta_2) = 1$, если

$$(1 - q) \int w_1(t) \mathbf{Q}_2(dt) > q w_2, \tag{3}$$

и $\pi_{\mathbf{Q}}(\delta_1) = 1$, если имеет место обратное неравенство. Соотношение 3 можно записать в виде

$$\int w(t) \mathbf{Q}(dt) > 0,$$

где

$$w(t) = \begin{cases} w_1(t) & \text{при} \quad t \neq \theta_1, \\ -w_2 & \text{при} \quad t = \theta_1. \end{cases}$$

В силу байесовского принципа байесовское решающее правило $\pi_{\mathbf{Q}}(X)$ имеет вид $\pi_{\mathbf{Q}}(X) = 1$, если

$$\int w(t) \mathbf{Q}_X(dt) > 0,$$

где $\mathbf{Q}_X$ — апостериорное распределение. Пусть $\lambda(dt) = dt$ при $t \neq \theta_1$, $\lambda(\{\theta_1\}) = 1$ и распределение $\mathbf{Q}_2$ имеет плотность $q_2(t)$ относительно меры Лебега. Тогда априорное распределение $\mathbf{Q}$ имеет плотность $q(t)$ относительно λ, равную $(1 - q)q_2(t)$ при $t \neq \theta_1$ и $q(t) = q$ при $t = \theta_1$. Это означает, что апостериорная плотность относительно меры λ будет равна

$$q(t/X) = \frac{f_t(X) q(t)}{f(X)}, \quad f(X) = \int f_u(X) q(u) \lambda(du).$$

Следовательно, байесовское решающее правило $\pi_{\mathbf{Q}}(X)$ имеет вид $\pi_{\mathbf{Q}}(X) = 1$, если

$$(1 - q) \int w_1(t) f_t(X) q_2(t)\, dt > w_2 q f_{\theta_1}(X). \tag{4}$$

Риск этого правила равен

$$\widetilde{W}(\pi_{\mathbf{Q}}(\cdot),\mathbf{Q}) = qw_2\mathbf{P}_{\theta_1}(\pi_{\mathbf{Q}}(X)=1)+(1-q)\int w_1(u)q_2(u)\mathbf{P}_u(\pi_{\mathbf{Q}}(X)=0)\,du.$$

Сравнивая эти соотношения с содержанием § 53, мы видим, что область (4) приема решения δ_2 имеет здесь тот же вид, что и область $\Omega(c)$ в (53.3) при $c = w_2q/(1-q)$ и при замене функции $q(t)$ в (53.3) на $w_1(t)q_2(t)$. Другими словами,

$$\pi_{\mathbf{Q}}(X) = \begin{cases} 1, & \text{если} \quad r_{\mathbf{Q}_2}(X) > c, \\ \gamma, & \text{если} \quad r_{\mathbf{Q}_2}(X) = c, \\ 0, & \text{если} \quad r_{\mathbf{Q}_2}(X) < c, \end{cases} \tag{5}$$

где

$$r_{\mathbf{Q}_2}(X) = \frac{\int w_1(t)q_2(t)f_t(X)\,dt}{f_{\theta_1}(X)}, \quad c = \frac{w_2q}{1-q}.$$

Далее, мы можем, следуя рассуждениям § 53, поступить следующим образом. Из совокупности байесовских правил (5) путем изменения числа q выбрать такое решение $\pi_{\mathbf{Q}}(X)$, которое имело бы фиксированное значение «потерь первого рода»:

$$w_2[\mathbf{P}_{\theta_1}(\pi_{\mathbf{Q}}(X)=1)+\gamma\mathbf{P}_{\theta_1}(\pi_{\mathbf{Q}}(X)=\gamma)] = \alpha.$$

Тогда среди всех правил $\pi(X)$, для которых

$$\alpha_1(\pi) = w_2\mathbf{E}_{\theta_1}\pi(X) \leqslant \alpha, \tag{6}$$

решение $\pi_{\mathbf{Q}}(X)$ будет минимизировать «потери второго рода», равные

$$\alpha_2(\pi) = \int w_1(u)q_2(u)\mathbf{E}_u(1-\pi(X))\,du. \tag{7}$$

Это есть прямое следствие байесовости решения $\pi_{\mathbf{Q}}$. Сравнение значений (6), (7) с величинами вероятностей ошибок первого и второго рода (53.4) показывает, что мы снова имеем дело с несущественными отличиями, главное из которых состоит в замене функции $q(u)$ в (53.4) на функцию $w_1(u)q_2(u)$. Числа c и γ в (5) определяются по α.

Изложенное позволяет нам, в точности следуя рассуждениям § 53, сформулировать следующие определения и утверждения.

Определение 1. Решающее правило $\pi(X)$ принадлежит классу $\widetilde{K}_\varepsilon$ (имеет асимптотический уровень $1-\varepsilon$), если

$$\limsup_{n\to\infty}\mathbf{E}_{\theta_1}\pi(X) \leqslant \varepsilon.$$

Это определение, по существу, ничем не отличается от определения 53.1.

Покажем теперь, что выбором q мы можем добиться того, чтобы $\pi_{\mathbf{Q}} \in \widetilde{K}_\varepsilon$. Положим

$$r_{\mathbf{Q}_2}(X) \equiv \frac{\int w_1(t)q_2(t)f_t(X)\,dt}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}} e^{T(X)},$$

где $I = I(\theta_1)$ есть информационная матрица Фишера в точке θ_1. Предположим, далее, что выполнены условия (RR) и что θ_1 есть внутренняя точка Θ, функция $w_1(t)q_2(t)$ непрерывна в точке θ_1 и положительна,

$$c = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)q_2(\theta_1)}{\sqrt{|I|}} e^z. \tag{8}$$

Тогда в силу леммы 53.1 для функции $p_t(c) = \mathbf{P}_t(r_{\mathbf{Q}_2}(X) > c)$ получаем

$$p_{\theta_1}(c) = \mathbf{P}_{\theta_1}(T(X) > z) \to \mathbf{H}_k((2z, \infty)).$$

Следовательно, полагая $q = c/(c + w_2)$, где c определено в (8), $z = h_\varepsilon/2$, h_ε есть квантиль порядка $1 - \varepsilon$ распределения χ^2 с k степенями свободы, мы получим

$$\lim_{n\to\infty} p_{\theta_1}\left(\frac{w_2 q}{1-q}\right) = \varepsilon,$$

и, стало быть, $\pi_{\mathbf{Q}}(X) \in \widetilde{K}_\varepsilon$.

Определение 2. Для данного априорного распределения $\mathbf{Q}$ решающее правило $\pi(X)$ называется *асимптотически байесовским в* $\widetilde{K}_\varepsilon$, если $\pi_{\mathbf{Q}} \in \widetilde{K}_\varepsilon$,

$$\limsup_{n\to\infty} \frac{\alpha_2(\pi)}{\alpha_2(\pi_{\mathbf{Q}})} = 1.$$

Теорема 1. *Пусть выполнены условия* (RR), *точка* θ_1 *является внутренней точкой* Θ. *Тогда в* $\widetilde{K}_\varepsilon$ *существует асимптотически байесовское решающее правило* $\widehat{\pi}(X)$ — *одно и то же для любых распределений* $\mathbf{Q}_2$ *и любых функций* $w_1(t)$ *таких, что функция* $w_1(t)q_2(t)$ *непрерывна и положительна в точке* θ_1 *и ограничена на* Θ. *Критерий* π *определяется соотношением*

$$\widehat{\pi}(X) = 1, \quad \textit{если} \quad \frac{f_{\widehat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{h_\varepsilon/2}. \tag{9}$$

Теорема доказывается точно так же, как теорема 53.1, с точностью до замены функции $q(t)$ на $w_1(t)q_2(t)$. Теорема 53.1 позволяет также найти значение «потерь второго рода» (см (7)) критерия $\widehat{\pi}$.

Критерий (9) есть не что иное, как критерий отношения правдоподобия.

§ 80. Асимптотически оптимальные решения при произвольной функции потерь в случае близких сложных гипотез

В этом параграфе мы рассмотрим обобщение на случай произвольной функции потерь результатов § 54. Это обобщение будет более содержательным, чем в предыдущем параграфе, так как функции потерь будут зависящими от n (ср. с § 78).

Пусть $(\mathcal{D}, \Theta, W)$ — статистическая игра, в которой $\Theta \subset \mathbb{R}^k$, множество $D = \{\delta_1, \delta_2\}$ двухточечно, $w(\delta_i, \theta) = w_i(\theta)$, где $w_i(\theta) = 0$ при $\theta \in \Theta_i$, $i = 1, 2$, пересечение $\Theta_1 \cap \Theta_2$ пусто.

Если $w_i(\theta) = 1$ при $\theta \notin \Theta_i$, то мы получим задачу о проверке гипотез $H_i = \{\theta \in \Theta_i\}$, $i = 1, 2$, рассмотренную в § 54, 55.

Найдем байесовскую стратегию для игры (D, Θ, w). Пусть $\mathbf{Q}_i$ — распределения на Θ_i,

$$\mathbf{Q} = q_1\mathbf{Q}_1 + q_2\mathbf{Q}_2, \quad q_1 + q_2 = 1.$$

Тогда, очевидно, $\widetilde{w}(\delta_i, \mathbf{Q}) = \int w_i(t)\mathbf{Q}(dt)$ и $\pi_{\mathbf{Q}}(\delta_2) = 1$, если

$$\int w_2(t)\mathbf{Q}(dt) < \int w_1(t)\mathbf{Q}(dt),$$

или

$$q_1 \int w_2(t)\mathbf{Q}_1(dt) < q_2 \int w_1(t)\mathbf{Q}_2(dt).$$

Следовательно, в силу байесовского принципа байесовское решающее правило будет иметь вид $\pi_{\mathbf{Q}}(X) = 1$, если

$$\int w_2(t)\mathbf{Q}_X(dt) < \int w_1(t)\mathbf{Q}_X(dt). \tag{1}$$

Допустим, что распределения $\mathbf{Q}_i$ имеют плотности $q_i(t)$, $i = 1, 2$, относительно меры λ. Тогда $\mathbf{Q}$ и апостериорное распределение $\mathbf{Q}_X$ будут иметь плотности соответственно $q(t) = q_1q_1(t) + q_2q_2(t)$ и

$$q(t/X) = \frac{q(t)f_t(X)}{f(X)}, \quad f(X) = \int q(u)f_u(X)\lambda(du).$$

Это значит, что соотношение (1) можно переписать в виде

$$q_1 \int_{\Theta_1} w_2(t)q_1(t)f_t(X)\lambda(dt) < q_2 \int_{\Theta_2} w_1(t)q_2(t)f_t(X)\lambda(dt). \tag{2}$$

Риск байесовского правила $\pi_{\mathbf{Q}}(X)$ равен

$$W(\pi_{\mathbf{Q}}(\cdot), \theta) = w_1(\theta)\mathbf{E}_\theta\pi_{\mathbf{Q}}(X) + w_2(\theta)(1 - \mathbf{E}_\theta\pi_{\mathbf{Q}}(X)),$$

$$\widetilde{W}(\pi_{\mathbf{Q}}(\cdot)\mathbf{Q}) = \int W(\pi_{\mathbf{Q}}(\cdot), t)q(t)\lambda(dt).$$

Перейдем теперь к рассмотрению *близких альтернатив*. Пусть θ_1 — какое-нибудь фиксированное значение параметра θ. Как и в § 54, мы будем предполагать, что множества Θ_i имеют вид

$$\Theta_i = \theta_i + \frac{\Gamma_i}{\sqrt{n}}, \tag{3}$$

где Γ_i от n не зависят. Относительно $\mathbf{Q}_i$ будем предполагать, что они индуцированы некоторыми распределениями $\mathbf{\Pi}_i$, сосредоточенными на Γ_i и от n не зависящими. Если множества Γ_i ограничены, то стратегии природы θ располагаются в $b/\sqrt{n}$-окрестности точки θ_1, $b < \infty$. Поэтому если $w_1(t)$, $w_2(t)$ непрерывны, $w_i(t) > c > 0$, $i = 1, 2$, соответственно на множествах Θ_2, Θ_1, то статистическая игра $(\mathcal{D}, \Theta, W)$ для такой функции потерь практически не будет отличаться по своим свойствам от игры со статистической функцией потерь $w_i(t) = 1$ при $t \notin \Theta_i$, рассмотренной нами в § 54, 55.

Рассмотрим здесь более содержательное обобщение, аналогичное тому, которое проводилось нами в § 78. Будем считать, что функция потерь $w(\delta_i, \theta) = w_i(\theta)$ зависит от n так, что

$$w_i(\theta) = w_{i,n}(\theta) = v_i(\sqrt{n}\,(\theta - \theta_1)), \tag{4}$$

где $v_i(t)$ — ограниченные измеримые функции, не зависящие от n. Как и в § 78, предположение (4) означает, что функции потерь очень чувствительны к отклонениям θ от θ_1.

Следуя § 54, мы будем задачу отыскания по выборке $X \in \mathbf{P}_\theta$ решения игры $(\mathcal{D}, \Theta, W)$, описанной выше, называть *задачей* A. Если выполнено (3), (4), то будем говорить о задаче A для близких гипотез с функциями потерь $v_i(t)$.

Рассмотрим теперь другую статистическую игру $(\mathcal{D}_B, \Gamma, V)$, относящуюся к выборке $Y \in \mathbf{\Phi}_{\gamma, I^{-1}}$ единичного объема, где $I = I(\theta_1)$ есть информационная матрица Фишера для семейства $\mathbf{P}_\theta$ в точке θ_1. Эта игра имеет двухточечное множество решений $D_B = (d_1, d_2)$ и множество стратегий природы (параметрическое множество) $\Gamma = \Gamma_1 \cup \Gamma_2$. Функция потерь $v(d, \gamma)$: $D_B \times \Gamma \to \mathbb{R}$ определяется соотношениями

$$v(d_i, \gamma) = v_i(\gamma), \quad v_i(\gamma) = 0 \quad \text{при} \quad \gamma \in \Gamma_i.$$

Таким образом, в этой игре $\mathcal{D}_B$ есть класс всех решений $d(Y)$: $\mathbb{R}^k \to D_B$, функция риска равна

$$V(d(\cdot), \gamma) = v_1(\gamma)\mathbf{\Phi}_{\gamma, I^{-1}}(d(Y) = d_1) + v_2(\gamma)\mathbf{\Phi}_{\gamma, I^{-1}}(d(Y) = d_2)$$

(одно из слагаемых в правой части равно нулю). Аналогично выписываются потери для рандомизированных стратегий $\pi(Y)$ в терминах $\mathbf{E}\pi(Y)$, $Y \in \mathbf{\Phi}_{\gamma, I^{-1}}$. Сформулированную задачу будем называть *задачей* B.

Между задачами A и B существует то же соотношение, которое было установлено между этими задачами в § 55. Пусть $\pi(Y)$ есть решение задачи B, оптимальное в том или ином смысле (байесовское, минимаксное). И пусть $\widehat{\theta}^*$ есть о. м. ц. в задаче A, $\gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n}$. Тогда *решение* $\pi(\gamma^*)$ *будет асимптотически оптимальным* (*в том же смысле*) *решением задачи* A.

Сформулированный «предельный признак оптимальности» позволяет редуцировать задачу A к более простой задаче B.

Чтобы придать изложенному точный смысл, дадим следующие определения. Пусть на Γ_i заданы распределения $\mathbf{\Pi}_i$. Положим $\mathbf{\Pi} = q_1\mathbf{\Pi}_1 + q_2\mathbf{\Pi}_2$, $q_1 + q_2 = 1$, и обозначим через $\mathbf{Q}$ распределение на Θ, индуцированное распределением $\mathbf{\Pi}$ и преобразованием $\theta = \theta_1 + \gamma/\sqrt{n}$.

Определение 1. Решение $\pi_1(X)$ называется *асимптотически байесовским*, если

$$\limsup_{n\to\infty} [\widetilde{W}(\pi_1(\cdot), \mathbf{Q}) - \widetilde{W}(\pi_{\mathbf{Q}}(\cdot), \mathbf{Q})] \leqslant 0.$$

Здесь, как и прежде,

$$\widetilde{W}(\pi(\cdot), \theta) = w_1(\theta)\mathbf{E}_\theta\pi(X) + w_2(\theta)(1 - \mathbf{E}_\theta\pi(X)),$$
$$\widetilde{W}(\pi(\cdot), \Theta) = \int \widetilde{W}(\pi(\cdot), t)\mathbf{Q}(dt),$$

$\pi_{\mathbf{Q}}$ — байесовское решающее правило.

Определение 2. Решение $\pi_1(X)$ называется *асимптотически минимаксным*, если для любого другого решения $\pi(X)$

$$\limsup_{n\to\infty} \left[\sup_{\theta\in\Theta} \widetilde{W}(\pi_1(\cdot), \theta) - \sup_{\theta\in\Theta} \widetilde{W}(\pi(\cdot), \theta)\right] \leqslant 0.$$

Если следовать определению 1, то мы могли бы здесь сравнивать π_1 с минимаксным правилом $\overline{\pi}$.

Аналогично § 54 мы могли бы рассматривать также асимптотически байесовские и минимаксные решения в классе $\widetilde{K}_\varepsilon$ решений с фиксированными асимптотическими «потерями первого рода»:

$$\varepsilon = \limsup_{n\to\infty} \sup_{\theta\in\Theta_1} w_1(\theta)\mathbf{E}_\theta\pi(X).$$

Чтобы получить соответствующие результаты, достаточно будет сравнить содержание этого параграфа с § 54.

Обозначим через $\pi_{\mathbf{\Pi}}(Y)$ байесовское решение игры $(\mathcal{D}_B, \Gamma, V)$ (т. е. задачи B), соответствующее априорному распределению $\mathbf{\Pi}$, и предположим для простоты, что множества Γ_i ограничены.

Теорема 1. *Пусть в окрестности точки θ_1 выполнены условия* (RR), *а функции v_i и распределение $\mathbf{\Pi}_i$ таковы, что*

$$0 < \int v_1(u)\mathbf{\Pi}_2(du) < \infty, \quad 0 < \int v_2(u)\mathbf{\Pi}_1(du) < \infty.$$

Тогда во введенных обозначениях критерий

$$\pi_1(X) = \pi_{\mathbf{\Pi}}(\gamma^*), \quad \gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n},$$

будет асимптотически байесовским решением игры $(\mathcal{D}, \Theta, W)$ (т. е. задачи A), соответствующим априорному распределению $\mathbf{Q}$.

Теорема 2. *Пусть в окрестности точки θ_1 выполнены условия* (RR) *и в задаче* B *существуют минимаксное решение $\overline{\pi}(Y)$ и соответствующее ему наихудшее распределение $\overline{\mathbf{\Pi}}$. Тогда критерий $\pi_1(X) = \overline{\pi}(\gamma^*)$ будет асимптотически минимаксным решением задачи* A.

Замечание 1. Условия теоремы о существовании $\overline{\pi}$ и $\overline{\mathbf{\Pi}}$ в силу теорем § 75 будут выполнены, если v_i — непрерывные функции.

Доказательство теоремы 1 вполне аналогично доказательству теоремы 54.1. Из (2) следует, что байесовское решающее правило $\pi_{\mathbf{Q}}$ имеет вид $\pi_{\mathbf{Q}}(X) = 1$, если

$$\frac{\int w_1(t)q_2(t)f_t(X)\lambda(dt)}{\int w_2(t)q_1(t)f_t(X)\lambda(dt)} > \frac{q_1}{q_2}. \tag{5}$$

Полагая $Z_1(t) = \dfrac{f_{\theta_1+t}(X)}{f_{\theta_1}(X)}$ и учитывая, что

$$q_i(t)\lambda(dt) = \mathbf{Q}_i(dt), \quad \mathbf{Q}_i\left(\theta_1 + \frac{du}{\sqrt{n}}\right) = \mathbf{\Pi}_i(du),$$
$$w_1\left(\theta_1 + \frac{u}{\sqrt{n}}\right) - v_i(u),$$

мы можем неравенство (5) с помощью замены $t = \theta_1 + u/\sqrt{n}$ преобразовать к виду

$$\frac{\int v_1(u)Z_1\left(u/\sqrt{n}\right)\mathbf{\Pi}_2(du)}{\int v_2(u)Z_1\left(u/\sqrt{n}\right)\mathbf{\Pi}_1(du)} = \frac{\int Z_1\left(u/\sqrt{n}\right)\mathbf{\Pi}_2'(du)}{\int Z_1\left(u/\sqrt{n}\right)\mathbf{\Pi}_1'(du)} > c, \quad c = \frac{q_1}{q_2}, \tag{6}$$

где обобщенные распределения $\mathbf{\Pi}_i'(A) = \int\limits_A v_{i+1}(u)\mathbf{\Pi}_i(du)$ $(v_3(u) \equiv v_1(u),$ $i = 1, 2)$ можно превратить в вероятностные перенормировкой, введя распределения $\mathbf{\Pi}_2''(A) = \mathbf{\Pi}_i'(A)/\mathbf{\Pi}_i'(\Gamma_i)$ (по условию $0 < \mathbf{\Pi}_i'(\Gamma_i) < \infty$). Получим тогда в качестве (5) неравенство точно того же вида, как в § 54.

Последующие рассуждения доказательства отличаются от соответствующих рассуждений в § 54 лишь упрощениями. Мы предоставляем провести их читателю. В основе этих рассуждений лежит равномерная по γ сходимость $(\theta = \theta_1 + \gamma/\sqrt{n})$

$$\widetilde{W}(\pi_{\mathbf{Q}}(\cdot), \theta) \to \widetilde{V}(\pi_{\mathbf{\Pi}}(\cdot), \gamma), \quad \widetilde{W}(\pi_1(\cdot), \theta) \to \widetilde{V}(\pi_{\mathbf{\Pi}}(\cdot), \gamma), \tag{7}$$

где $\pi_1(X) = \pi_{\mathbf{\Pi}}(\gamma^*)$. $\triangleleft$

Для доказательства теоремы 2 нам понадобится

Лемма 1. *Пусть $\mathbf{Q}$ — априорное распределение и π_1 — соответствующее ему асимптотически байесовское решение такое, что*

$$\limsup_{n\to\infty} \widetilde{W}(\pi_1(\cdot), \mathbf{Q}) = c_1, \quad \limsup_{n\to\infty}\sup_{\theta\in\Theta} \widetilde{W}(\pi_1(\cdot), \theta) \leqslant c. \tag{8}$$

Тогда π_1 — асимптотически минимаксное решение.

Доказательство. Обозначим, как и прежде, через $\pi_{\mathbf{Q}}$ байесовское решение. Тогда для любого решения π имеем

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta}\widetilde{W}(\pi,\theta)\geqslant\limsup_{n\to\infty}\widetilde{W}(\pi,\mathbf{Q})\geqslant\limsup_{n\to\infty}\widetilde{W}(\pi_{\mathbf{Q}},\mathbf{Q})\geqslant$$

$$\geqslant\limsup_{n\to\infty}\widetilde{W}(\pi_1,\mathbf{Q})=c\geqslant\limsup_{n\to\infty}\sup_{\theta\in\Theta}\widetilde{W}(\pi_1,\theta).\quad\triangleleft$$

Доказательство теоремы 2. Пусть $\overline{\mathbf{\Pi}}$ — наихудшее распределение на Γ, так что $\overline{\pi}(Y)=\pi_{\overline{\mathbf{\Pi}}}(Y)$ есть минимаксное решающее правило в игре $(\mathcal{D}_B,\Gamma,V)$. Тогда по теореме 1 $\pi_1(X)=\pi_{\overline{\mathbf{\Pi}}}(\gamma^*)$ будет асимптотически байесовским решением для распределения $\overline{\mathbf{Q}}$, соответствующего $\overline{\mathbf{\Pi}}$, и для доказательства теоремы нам достаточно убедиться, что $\overline{\mathbf{Q}}$ и π_1 удовлетворяют условиям леммы 1.

Обозначим через $N_{\overline{\mathbf{\Pi}}}$ носитель распределения $\overline{\mathbf{\Pi}}$. Тогда в силу теорем § 75

$$\widetilde{V}(\pi_{\overline{\mathbf{\Pi}}}(\cdot),\gamma)=c\quad\text{при}\quad\gamma\in N_{\overline{\mathbf{\Pi}}},\quad\sup_{\gamma\in\Gamma}\widetilde{V}(\pi_{\overline{\mathbf{\Pi}}}(\cdot),\gamma)\leqslant c.\tag{9}$$

Но для $\theta=\theta_1+\gamma/\sqrt{n}$ имеет место (см. (7)) равномерная по γ сходимость $\widetilde{W}(\pi_1(\cdot),\theta)\to\widetilde{V}(\pi_{\overline{\mathbf{\Pi}}}(\cdot),\gamma)$. Отсюда и из (9) следует (8). Теорема доказана. $\triangleleft$

Таблица I. Нормальное распределение $\Phi_{0,1}$

В таблице приведены значения

$$\overline{\Phi}(x) = \Phi_{0,1}((x, \infty)) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt.$$

Таблица I

x	0	1	2	3	4
0,0	0,5000	0,4960	0,4920	0,4880	0,4840
0,1	,4602	,4562	,4522	,4483	,4443
0,2	,4207	,4168	,4129	,4090	,4052
0,3	,3821	,3783	,3745	,3707	,3669
0,4	,3446	,3409	,3372	,3336	,3300
0,5	,3085	,3050	,3015	,2981	,2946
0,6	,2743	,2709	,2676	,2643	,2611
0,7	,2420	,2389	,2358	,2327	,2297
0,8	,2119	,2090	,2061	,2033	,2005
0,9	,1841	,1814	,1788	,1762	,1736
1,0	,1587	,1562	,1539	,1515	,1492
1,1	,1357	,1335	,1314	,1292	,1271
1,2	,1151	,1131	,1112	,1093	,1075
1,3	,0968	,0951	,0934	,0918	,0901
1,4	,0808	,0793	,0778	,0764	,0749
1,5	,0668	,0655	,0643	,0630	,0618
1,6	,0548	,0537	,0526	,0516	,0505
1,7	,0446	,0436	,0427	,0418	,0409
1,8	,0359	,0351	,0344	,0336	,0329
1,9	,0288	,0281	,0274	,0268	,0262
2,0	,0228	,0222	,0217	,0212	,0207
2,1	,0179	,0174	,0170	,0166	,0162
2,2	,0139	,0136	,0132	,0129	,0125
2,3	,0107	,0104	,0102	,0099	,0096
2,4	,0082	,0080	,0078	,0075	,0073
2,5	,0062	,0060	,0059	,0057	,0055
2,6	,0047	,0045	,0044	,0043	,0041
2,7	,0035	,0034	,0033	,0032	,0031
2,8	,0026	,0025	,0024	,0023	,0023
2,9	,0019	,0018	,0018	,0017	,0016
$x =$	3,0	3,1	3,2	3,3	3,4
$\overline{\Phi}(x) =$	0,0013	0,0010	0,0007	0,0005	0,0003

Таблица I (окончание)

x	5	6	7	8	9
0,0	0,4810	0,4761	0,4721	0,4681	0,4641
0,1	,4404	,4364	,4325	,4286	,4247
0,2	,4013	,3974	,3936	,3897	,3859
0,3	,3632	,3594	,3557	,3520	,3483
0,4	,3264	,3228	,3192	,3156	,3121
0,5	,2912	,2877	,2843	,2810	,2776
0,6	,2578	,2546	,2514	,2483	,2451
0,7	,2266	,2236	,2206	,2177	,2148
0,8	,1977	,1949	,1922	,1894	,1867
0,9	,1711	,1685	,1660	,1635	,1611
1,0	,1469	,1446	,1423	,1401	,1379
1,1	,1251	,1230	,1210	,1190	,1170
1,2	,1056	,1038	,1020	,1003	,0985
1,3	,0885	,0869	,0853	,0838	,0823
1,4	,0735	,0721	,0708	,0694	,0681
1,5	,0606	,0594	,0582	,0571	,0559
1,6	,0495	,0485	,0475	,0465	,0455
1,7	,0401	,0392	,0384	,0375	,0367
1,8	,0322	,0314	,0307	,0301	,0294
1,9	,0256	,0250	,0244	,0239	,0233
2,0	,0202	,0197	,0192	,0188	,0183
2,1	,0158	,0154	,0150	,0146	,0143
2,2	,0122	,0119	,0116	,0113	,0110
2,3	,0094	,0091	,0089	,0087	,0084
2,4	,0071	,0069	,0068	,0066	,0064
2,5	,0054	,0052	,0051	,0049	,0048
2,6	,0040	,0039	,0038	,0037	,0036
2,7	,0030	,0029	,0028	,0027	,0026
2,8	,0022	,0021	,0021	,0020	,0019
2,9	,0016	,0015	,0015	,0014	,0014
$x=$	3,5	3,6	3,7	3,8	3,9
$\overline{\Phi}(x)=$	0,0002	0,0002	0,0001	0,0001	0,0000

Таблица II. Квантили нормального распределения

В таблице приведены значения λ_ε такие, что

$$\overline{\Phi}(\lambda_\varepsilon) = \Phi_{0,1}((\lambda_\varepsilon, \infty)) = \varepsilon.$$

Таблица II

100ε	λ_ε	100ε	λ_ε	100ε	λ_ε
50	0,0000	20	0,8416	0,5	2,5758
45	0,1257	15	1,0364	0,1	3,0902
40	0,2533	10	1,2816	0,05	3,2905
35	0,3853	5	1,6449	0,01	3,7190
30	0,5244	2,5	1,9600	0,005	3,8906
25	0,6745	1	2,3263		

Таблица III. Распределение хи-квадрат $\mathbf{H}_k$

В таблице приведены значения (см. § 2.2)

$$H_k(x) = \mathbf{H}_k((x,\infty)) = \frac{1}{2^{k/2}\Gamma(k/2)} \int_x^\infty t^{k/2-1} e^{-t/2}\, dt$$

при $1 \leqslant k \leqslant 20$. При бо́льших k можно пользоваться приближением (см. § 12, табл. I)

$$H_k(x) \approx \overline{\Phi}(\sqrt{2x} - \sqrt{2k-1}) \equiv \widetilde{H}_k(x). \tag{1}$$

Последний столбец таблицы содержит значения $\widetilde{H}_k(x)$ при $k = 20$. Путем сравнения с предыдущим столбцом можно оценить степень точности приближения (1). С ростом k погрешность уменьшается.

Таблица III

x	Объем выборки k				
	1	2	3	4	5
0,1	0,7518	0,9512	0,9918	0,9988	0,9998
,2	,6547	,9048	,9776	,9953	,9991
,4	,5271	,8187	,9402	,9825	,9953
,6	,4386	,7408	,8964	,9631	,9880
,8	,3711	,6703	,8495	,9385	,9770
1,0	,3173	,6065	,8013	,9098	,9626
,5	,2207	,4724	,6823	,8266	,9131
2	,1573	,3679	,5725	,7358	,8492
3	,0833	,2231	,3916	,5578	,7000
4	,0455	,1353	,2615	,4060	,5494
5	,0254	,0821	,1718	,2873	,4159
6	,0143	,0498	,1116	,1992	,3062
7	,0082	,0302	,0719	,1359	,2206
8	,0047	,0183	,0460	,0916	,1562
9	,0027	,0111	,0293	,0611	,1091
10	,0016	,0067	,0186	,0404	,0752
11	,0009	,0041	,0117	,0266	,0514
12	,0005	,0025	,0074	,0174	,0348
13	,0003	,0015	,0046	,0113	,0234
14	,0002	,0009	,0029	,0073	,0156
15	,0001	,0006	,0018	,0047	,0104
16	,0001	,0003	,0011	,0030	,0068
17		,0002	,0007	,0019	,0045
18		,0001	,0004	,0012	,0030
19		,0001	,0003	,0008	,0019
20		,0001	,0002	,0001	,0013
21			,0001	,0003	,0008
22			,0001	,0002	,0005
23				,0001	,0003
24				,0001	,0002
25				,0001	,0001

Таблица III (продолжение)

x	Объем выборки k				
	6	7	8	9	10
0,5	0,9978	0,9995	0,9999	1,0000	1,0000
1,0	,9856	,9948	,9983	,9994	,9998
,5	,9595	,9823	,9927	,9972	,9989
2,0	,9197	,9598	,9810	,9915	,9963
,5	,8685	,9271	,9617	,9809	,9909
3	,8089	,8850	,9344	,9643	,9814
4	,6767	,7798	,8571	,9114	,9474
5	,5438	,6600	,7576	,8343	,8912
6	,4232	,5398	,6472	,7399	,8153
7	,3204	,4284	,5366	,6371	,7254
8	,2381	,3326	,4335	,5342	,6288
9	,1736	,2527	,3423	,4373	,5321
10	,1246	,1886	,2650	,3505	,4405
11	,0884	,1386	,2017	,2757	,3575
12	,0620	,1006	,1512	,2133	,2851
13	,0430	,0721	,1119	,1626	,2237
14	,0296	,0512	,0818	,1223	,1730
15	,0203	,0360	,0592	,0909	,1321
16	,0138	,0251	,0424	,0669	,0996
17	,0093	,0174	,0301	,0487	,0744
18	,0062	,0120	,0212	,0352	,0550
19	,0042	,0082	,0149	,0252	,0403
20	,0028	,0056	,0103	,0179	,0193
21	,0018	,0038	,0072	,0127	,0211
22	,0012	,0025	,0049	,0084	,0151
23	,0008	,0017	,0034	,0062	,0108
24	,0005	,0011	,0023	,0043	,0076
25	,0003	,0008	,0016	,0030	,0054
26	,0002	,0005	,0011	,0020	,0037
27	,0002	,0003	,0007	,0014	,0026
28	,0001	,0002	,0004	,0010	,0018
29	,0001	,0002	,0003	,0007	,0013
30		,0001	,0002	,0004	,0009

Таблица III (продолжение)

x	Объем выборки k				
	11	12	13	14	15
2	0,9985	0,9994	0,9998	0,9999	1,0000
3	,9907	,9955	,9979	,9991	0,9996
4	,9699	,9834	,9912	,9955	,9977
5	,9312	,9580	,9752	,9858	,9921
6	,8734	,9161	,9462	,9665	,9798
7	,7991	,8576	,9022	,9347	,9577
8	,7133	,7852	,8436	,8893	,9238
9	,6219	,7029	,7729	,8311	,8775
10	,5304	,6160	,6939	,7622	,8197
12	,3636	,4457	,5276	,6063	,6790
14	,2330	,3007	,3738	,4497	,5255
16	,1411	,1912	,2491	,3134	,3821
18	,0816	,1157	,1575	,2068	,2627
20	,0453	,0671	,0952	,1301	,1719
21	,0334	,0504	,0729	,1016	,1368
22	,0244	,0375	,0554	,0786	,1078
23	,0177	,0277	,0417	,0603	,0841
24	,0127	,0203	,0311	,0458	,0651
25	,0091	,0148	,0231	,0346	,0499
26	,0065	,0170	,0107	,0259	,0380
27	,0046	,0077	,0124	,0193	,0287
28	,0032	,0055	,0091	,0142	,0216
29	,0023	,0039	,0066	,0105	,0161
30	,0016	,0028	,0047	,0076	,0119
31	,0011	,0020	,0034	,0055	,0088
32	,0008	,0014	,0024	,0040	,0064
33	,0005	,0010	,0017	,0029	,0047
34	,0004	,0007	,0012	,0021	,0034
35	,0003	,0005	,0009	,0015	,0025
36	,0002	,0003	,0006	,0010	,0018
37	,0001	,0002	,0004	,0007	,0013
38	,0001	,0002	,0003	,0005	,0009
39	,0001	,0001	,0002	,0004	,0006
40		,0001	,0001	,0003	,0005

Таблица III (окончание)

x	Объем выборки k					$\widetilde{H}_{20}(x)$
	16	17	18	19	20	
4	0,9989	0,9995	0,9998	0,9999	1,0000	0,9997
5	,9958	,9978	,9989	,9994	0,9997	,9990
6	,9881	,9932	,9962	,9979	,9989	,9973
7	,9733	,9836	,9901	,9942	,9967	,9938
8	,9489	,9666	,9786	,9867	,9919	,9876
9	,9134	,9403	,9597	,9735	,9829	,9774
10	,8666	,9036	,9319	,9530	,9682	,9619
12	,7440	,8001	,8472	,8856	,9161	,9109
14	,5987	,6671	,7291	,7837	,8305	,8298
16	,4530	,5238	,5926	,6573	,7166	,7218
18	,3239	,3888	,4557	,5224	,5874	,5968
20	,2202	,2742	,3328	,3946	,4579	,4683
22	,1432	,1847	,2320	,2843	,3405	,3489
24	,0895	,1194	,1550	,1962	,2424	,2472
26	,0540	,0745	,0998	,1302	,1658	,1670
28	,0316	,0449	,0621	,0834	,1094	,1078
30	,0180	,0264	,0375	,0518	,0699	,0667
31	,0135	,0200	,0288	,0404	,0552	,0517
32	,0100	,0151	,0220	,0313	,0433	,0396
33	,0074	,0113	,0167	,0240	,0337	,0301
34	,0054	,0084	,0126	,0184	,0261	,0227
35	,0040	,0062	,0095	,0140	,0201	,0169
36	,0029	,0046	,0071	,0106	,0154	,0125
37	,0021	,0034	,0052	,0080	,0117	,0092
38	,0015	,0025	,0039	,0059	,0089	,0067
39	,0011	,0018	,0029	,0044	,0067	,0048
40	,0008	,0013	,0021	,0033	,0050	,0035
41	,0006	,0009	,0015	,0024	,0037	,0025
42	,0004	,0007	,0011	,0018	,0028	,0017
43	,0003	,0005	,0008	,0013	,0020	,0012
44	,0002	,0003	,0006	,0009	,0015	,0010
45	,0001	,0002	,0004	,0007	,0011	,0006

Таблица IV. Распределение Стьюдента T_k

В таблице приведены значения

$$T_k(x) = \mathbf{T}_k((x,\infty)) = \frac{\Gamma((k+1)/2)}{\sqrt{k\pi}\,\Gamma(k/2)} \int\limits_x^\infty \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2} dt$$

при $1 \leqslant k \leqslant 20$. При бо́льших значениях k можно пользоваться приближением (см. § 12, табл. I)

$$T_k(x) \approx \overline{\Phi}(x) = \Phi_{0,1}((x,\infty)). \tag{2}$$

Точность приближения (2) при $k = 20$ можно оценить путем сравнения последнего столбца таблицы с табл. I.

Таблица IV

x	Объем выборки k				
	1	2	3	4	5
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3524	,3333	,3257	,3217	,3191
1,0	,2500	,2113	,1955	,1869	,1816
2	,2211	,1765	,1581	,1482	,1419
4	,1974	,1482	,1280	,1170	,1102
6	,1778	,1253	,1039	,0924	,0852
8	,1614	,1068	,0848	,0731	,0659
2,0	,1476	,0917	,0697	,0581	,0510
2	,1358	,0794	,0576	,0463	,0395
4	,1257	,0692	,0479	,0372	,0308
6	,1169	,0679	,0402	,0300	,0241
8	,1092	,0537	,0339	,0244	,0190
3,0	,1024	,0477	,0282	,0200	,0150
2	,0964	,0427	,0247	,0165	,0120
4	,0910	,0383	,0212	,0136	,0096
6	,0862	,0346	,0184	,0114	,0078
8	,0819	,0314	,0160	,0095	,0063
4,0	,0780	,0286	,0140	,0081	,0052
2	,0744	,0261	,0123	,0068	,0045

Таблица IV (продолжение)

x	Объем выборки k				
	1	2	3	4	5
4	0,0711	0,0240	0,0109	0,0058	0,0035
6	,0681	,0221	,0097	,0050	,0029
8	,0654	,0204	,0086	,0043	,0024
5,0	,0628	,0199	,0077	,0037	,0020
2	,0605	,0175	,0069	,0033	,0017
4	,0583	,0163	,0062	,0028	,0015
6	,0562	,0152	,0056	,0025	,0012
8	,0543	,0142	,0051	,0022	,0011
6,0	,0526	,0133	,0046	,0019	,0009
2	,0509	,0125	,0042	,0017	,0008
4	,0493	,0118	,0039	,0015	,0007
6	,0479	,0111	,0035	,0014	,0006
8	,0465	,0105	,0033	,0012	,0005
7,0	,0452	,0099	,0030	,0011	,0005
2	,0439	,0094	,0028	,0010	,0004
4	,0428	,0089	,0025	,0009	,0004
6	,0416	,0086	,0024	,0008	,0003
8	,0406	,0080	,0022	,0007	,0003
8,0	,0396	,0076	,0020	,0007	,0002

Таблица IV (продолжение)

x	Объем выборки k				
	6	7	8	9	10
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3174	,3162	,3153	,3145	,3139
1,0	,1780	,1753	,1733	,1717	,1704
2	,1377	,1346	,1322	,1304	,1289
4	,1055	,1021	,0995	,0975	,0959
6	,0804	,0768	,0741	,0720	,0703
8	,0610	,0574	,0548	,0527	,0510
2,0	,0462	,0428	,0403	,0383	,0367
2	,0350	,0319	,0295	,0277	,0262
4	,0266	,0237	,0216	,0199	,0186
6	,0203	,0177	,0158	,0144	,0132
8	,0156	,0132	,0116	,0104	,0094
3,0	,0120	,0100	,0085	,0075	,0067
2	,0093	,0075	,0063	,0054	,0047
4	,0072	,0057	,0047	,0039	,0034
6	,0057	,0044	,0035	,0029	,0024
8	,0045	,0034	,0026	,0022	,0017
4,0	,0035	,0026	,0020	,0015	,0013
2	,0028	,0020	,0015	,0012	,0009
4	,0023	,0016	,0011	,0009	,0007
6	,0018	,0012	,0009	,0006	,0005
8	,0015	,0010	,0007	,0005	,0004
5,0	,0012	,0008	,0005	,0004	,0003

Таблица IV (продолжение)

x	Объем выборки k				
	11	12	13	14	15
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3135	,3131	,3127	,3124	,3112
1,0	,1694	,1685	,1678	,1671	,1666
2	,1277	,1266	,1258	,1250	,1244
4	,0945	,0934	,0925	,0916	,0909
6	,0689	,0678	,0668	,0660	,0652
8	,0496	,0485	,0475	,0467	,0460
2,0	,0354	,0343	,0334	,0326	,0320
2	,0250	,0241	,0232	,0225	,0219
4	,0176	,0168	,0160	,0154	,0149
6	,0123	,0116	,0110	,0105	,0100
8	,0086	,0080	,0075	,0071	,0067
3,0	,0060	,0055	,0051	,0048	,0045
2	,0042	,0038	,0035	,0032	,0030
4	,0030	,0026	,0024	,0022	,0020
6	,0021	,0018	,0016	,0014	,0013
8	,0015	,0013	,0011	,0010	,0009
4,0	,0010	,0009	,0008	,0007	,0006

Таблица IV (окончание)

x	Объем выборки k				
	16	17	18	19	20
0,0	0,5000	0,5000	0,5000	0,5000	0,5000
5	,3119	,3117	,3116	,3114	,3113
1,0	,1661	,1657	,1653	,1649	,1646
2	,1238	,1233	,1228	,1224	,1221
4	,0903	,0898	,0893	,0888	,0884
6	,0646	,0640	,0635	,0630	,0626
8	,0454	,0448	,0443	,0439	,0435
2,0	,0314	,0309	,0304	,0300	,0296
2	,0214	,0210	,0205	,0202	,0199
4	,0145	,0141	,0137	,0134	,0131
6	,0097	,0093	,0090	,0082	,0086
8	,0064	,0061	,0059	,0057	,0055
3,0	,0042	,0040	,0038	,0037	,0035
2	,0028	,0026	,0025	,0024	,0022
4	,0018	,0017	,0016	,0015	,0014
6	,0012	,0011	,0010	,0009	,0009
8	,0008	,0007	,0007	,0006	,0006
4,0	,0005	,0005	,0004	,0004	,0003

Приложение I

Теоремы типа Гливенко–Кантелли

В этом приложении будут доказаны утверждения, из которых будут вытекать теоремы 4.1, 4.2. Обозначения § 4, в котором эти теоремы сформулированы, мы будем использовать без пояснений. Докажем сначала один общий вспомогательный вариант теоремы Гливенко–Кантелли.

Определение. Класс $\mathfrak{K}$ множеств из $\mathfrak{B}_{\mathcal{X}} = \mathfrak{B}^m$ будем называть *конечно аппроксимируемым* (относительно распределения $\mathbf{P}$), если, каково бы ни было $\varepsilon > 0$, для него существует другой класс множеств $\mathfrak{S}(\varepsilon)$, состоящий из конечного числа $N = N(\varepsilon)$ элементов $S_1, \ldots, S_N$, $S_i \in \mathfrak{B}^m$, такой, что для любого $B \in \mathfrak{K}$ найдутся множества A_1 и A_2 из $\mathfrak{S}(\varepsilon)$ со свойствами

$$\begin{aligned} &A_1 \subset B \subset A_2, \\ &\mathbf{P}(A_2 - A_1) < \varepsilon. \end{aligned} \tag{1}$$

Определим над классами множеств операции *сложения*, *умножения* и *дополнения*. Классами $\mathfrak{K}_1 + \mathfrak{K}_2$ и $\mathfrak{K}_1\mathfrak{K}_2$ будем называть соответственно классы множеств вида $A \cup B$ и $A \cap B$, где $A \in \mathfrak{K}_1$, $B \in \mathfrak{K}_2$. Дополнением $\overline{\mathfrak{K}}$ будем называть класс множеств, составленный из дополнений $\overline{A}$, $A \in \mathfrak{K}$.

Теорема 1. 1) *Пусть* $X_n = [X_\infty]_n$, $X_\infty \in \mathbf{P}$, *а класс* $\mathfrak{K}$ *является конечно аппроксимируемым. Тогда*

$$\sup_{B \in \mathfrak{K}} \left| \mathbf{P}_n^*(B) - \mathbf{P}(B) \right| \underset{\text{п.н.}}{\longrightarrow} 0. \tag{2}$$

2) *Совокупность конечно аппроксимируемых классов замкнута относительно введенных операций.*

Доказательство. Первое утверждение получается с помощью тех же соображений, которые применялись нами в одномерном случае в теореме 2.2. Для заданных $B \in \mathfrak{K}$ и $\varepsilon > 0$ найдутся $N = N(\varepsilon)$ и множества A_1, A_2, обладающие свойством (1). Для них имеем

$$\begin{aligned} \mathbf{P}_n^*(B) - \mathbf{P}(B) &\leqslant \mathbf{P}_n^*(A_2) - \mathbf{P}(A_1) < \mathbf{P}_n^*(A_2) - \mathbf{P}(A_2) + \varepsilon, \\ \mathbf{P}_n^*(B) - \mathbf{P}(B) &\geqslant \mathbf{P}_n^*(A_1) - \mathbf{P}(A_2) > \mathbf{P}_n^*(A_1) - \mathbf{P}(A_1) - \varepsilon. \end{aligned}$$

Поэтому

$$\bigcap_{k=1}^{N} \left\{ \left| \mathbf{P}_n^*(S_k) - \mathbf{P}(S_k) \right| < \varepsilon \right\} \subset \left\{ \sup_{B \in \mathfrak{K}} \left| \mathbf{P}_n^*(B) - \mathbf{P}(B) \right| < 2\varepsilon \right\},$$

где $S_1, \ldots, S_n$ — элементы $\mathfrak{S}(\varepsilon)$. Так как $\mathbf{P}_n^*(S_k) \underset{\text{п.н.}}{\longrightarrow} \mathbf{P}(S_k)$, то отсюда уже без труда получаем (2) (ср. с доказательством теоремы 2.2А).

Второе утверждение теоремы 1 почти очевидно. Пусть дано $\varepsilon > 0$ и пусть $\mathfrak{S}_1(\varepsilon_1)$ и $\mathfrak{S}_2(\varepsilon_2)$ — аппроксимирующие классы для $\mathfrak{K}_1$ и $\mathfrak{K}_2$ соответственно. Пусть, далее, A и B — любые множества из $\mathfrak{K}_1$, $\mathfrak{K}_2$. Из соотношений $\varepsilon_1 + \varepsilon_2 = \varepsilon$,

$$A_1 \subset A \subset A_2, \quad \mathbf{P}(A_2 - A_1) < \varepsilon_1 \quad (A_i \in \mathfrak{S}_1(\varepsilon_1)),$$

$$B_1 \subset B \subset B_2, \quad \mathbf{P}(B_2 - B_1) < \varepsilon_2 \quad (B_i \in \mathfrak{S}_2(\varepsilon_2))$$

получаем

$$A_1 B_1 \subset AB \subset A_2 B_2,$$

$$A_2 B_2 - A_1 B_1 \subset (A_2 - A_1) \cup (B_2 - B_1),$$

$$\mathbf{P}(A_2 B_2 - A_1 B_1) \leqslant \varepsilon.$$

Таким образом, класс $\mathfrak{K}_1 \mathfrak{K}_2$ конечно аппроксимируем. Сумма $\mathfrak{K}_1 + \mathfrak{K}_2$ и дополнение $\overline{\mathfrak{K}}$ рассматриваются аналогично. $\triangleleft$

Следствие 1. *Пусть* $\mathcal{X} = \mathbb{R}^m$, $X_n = [X_\infty]_n \Subset F$. *Тогда*

$$\sup_t \left| F_n^*(t) - F(t) \right| \underset{\text{п.н.}}{\longrightarrow} 0$$

при $n \to \infty$, *где* $F_n^*(t)$ — *эмпирическая функция распределения.*

Доказательство. Из доказательства теоремы 2.2А видно, что классы множеств $\mathfrak{K}_j = \{\boldsymbol{y} \in \mathbb{R}^m : y_j < t_j\}$, $-\infty < t_j < \infty$, при каждом $j = 1, \ldots, m$ являются конечно аппроксимируемыми. В качестве системы $\mathfrak{S}(\varepsilon)$ достаточно взять полупространство $\{y_j < z_k\}$ и $\{y_j \leqslant z_k\}$, $k = 1, \ldots, N$, где z_k определены в (2.6).

По второму утверждению теоремы 1 класс углов $\mathfrak{K} = \mathfrak{K}_1 \mathfrak{K}_2 \ldots \mathfrak{K}_m$ также будет конечно аппроксимируемым. Нам остается воспользоваться первым утверждением теоремы 1. $\triangleleft$

Следствие 1 есть не что иное, как теорема 4.1.

Рассмотрим теперь классы множеств $\mathfrak{K}$, которые удовлетворяют следующему условию (Г). Пусть K_M есть куб

$$K_M = \left\{ \boldsymbol{y} = (y_1, \ldots, y_m) : \max_{1 \leqslant k \leqslant m} |y_k| \leqslant M \right\}.$$

(Г) *Все множества* $B \in \mathfrak{K}$ *обладают свойством: окрестность* Γ_B^ε *границы* $\Gamma_B = \partial(B \cap K_M)$ *имеет меру Лебега (объем)* $\mu(\Gamma_B^\varepsilon) \leqslant \varphi(\varepsilon, M)$, *где* φ *зависит только от своих аргументов и при любом* M $\varphi(\varepsilon, M) \to 0$ *при* $\varepsilon \to 0$.

Теорема 2. *Пусть* $\mathcal{X} = \mathbb{R}^m$, $X \Subset \mathbf{P}$ *и распределение* $\mathbf{P}$ *абсолютно непрерывно относительно меры Лебега. Тогда класс* $\mathfrak{K}$, *удовлетворяющий условию* (Г), *конечно аппроксимируем и, следовательно, для него справедливо* (2).

Доказательство. Заметим прежде всего, что задачу с пространства $\mathbb{R}^m$ можно редуцировать к кубу K_M в следующем смысле. Допустим, что при любом фиксированном M найдется класс $\mathfrak{S}$ подмножеств из K_M такой, что для любого $B' \in \mathfrak{K}$ и $B = B' \cap K_M$ выполнено (1). Тогда $\mathfrak{K}$ будет конечно аппроксимируем. Действительно, для $\varepsilon > 0$, выбранного в (1), найдем $M = M(\varepsilon)$ такое, что $\mathbf{P}(K_M) \geqslant \geqslant 1 - \varepsilon$, и положим $A_1' = A_1$, $A_2' = A_2 \cup \overline{K}_M$, где A_i — множества из (1), $\overline{K}_M$ — дополнение до K_M. Тогда, очевидно,

$$A_1' \subset B' \subset A_2', \quad \mathbf{P}(A_2' - A_1') \leqslant 2\varepsilon.$$

Итак, мы можем считать, что $\mathbf{P}(K_M) = 1$, $\mathfrak{K}$ составлено из подмножеств K_M.

Рассмотрим в качестве $\mathfrak{S}$ фигуры A_j, образованные различными объединениями замкнутых кубов с ребрами длиной δ и с вершинами в точках

$$(j_1\delta, \ldots, j_m\delta), \quad -\frac{M}{\delta} < j_k < \frac{M}{\delta}, \quad k = 1, \ldots, m$$

(можно считать для простоты, что δ нацело делит M). Определим множества A_1, A_2 соответственно как объединения всех кубов, принадлежащих и задевающих B. Очевидно, что

$$A_1 \subset B \subset A_2,$$

$$\mu(A_2 - A_1) \leqslant \mu\left(\Gamma_B^{2\delta\sqrt{m}}\right) \leqslant \varphi(2\delta\sqrt{m}, M).$$

Выбором δ правая часть этого неравенства может быть сделана сколь угодно малой.

Далее, $\mathbf{P}$ абсолютно непрерывна относительно μ. Поэтому для заданного ε можно найти $\gamma = \gamma(\varepsilon)$ такое, что $\sup\limits_{\mu(A)<\gamma} \mathbf{P}(A) < \varepsilon$. Если теперь δ выбрать так, чтобы $\varphi\left(2\delta\sqrt{m}, M\right) < \gamma$, то получим

$$\mathbf{P}(A_2 - A_1) < \varepsilon. \quad \triangleleft$$

Следствие 2. *Класс $\mathfrak{C}$ всех выпуклых множеств конечно аппроксимируем и, следовательно, для абсолютно непрерывных $\mathbf{P}$*

$$\sup_{B\in\mathfrak{C}} \left|\mathbf{P}_n^*(B) - \mathbf{P}(B)\right| \underset{\text{п.н.}}{\longrightarrow} 0.$$

Действительно, максимальная «площадь» поверхности выпуклого множества в K_M составляет $2m(2M)^{m-1}$, что есть «площадь» поверхности K_M, а максимальный объем $\mu((\partial K_M)^\varepsilon)$ ε-окрестности ∂K_M не превосходит $2\varepsilon \cdot 2m(2M)^{m-1}$. Это означает, что условие (Г) выполнено. $\triangleleft$

Следствие 2 совпадает с теоремой 4.2. Замечание о существенности условия об абсолютной непрерывности $\mathbf{P}$ содержится в § 4.

Нетрудно видеть, что условие (Г) будет выполнено и для классов невыпуклых множеств с достаточно гладкими границами.

Приложение II

Функциональная предельная теорема для эмпирических процессов

Мы докажем здесь следующее утверждение (теорему 6.3). Пусть

$$w^n(t) = \sqrt{n}\,(F_n^*(t) - t)$$

есть эмпирический процесс, определенный в § 6, и $w^0(t)$ — броуновский мост.

Теорема 1. *Если f есть измеримый функционал $D(0,1) \to \mathbb{R}$, непрерывный в точках пространства $C(0,1)$ в равномерной метрике, то при $n \to \infty$*

$$f(w^n) \Rightarrow f(w^0).$$

Для доказательства теоремы нам понадобятся две леммы.

Лемма 1. *Конечномерные распределения процессов w^n слабо сходятся при $n \to \infty$ к соответствующим распределениям процесса w^0.*

Доказательство. Рассмотрим случайные $(m+1)$-мерные векторы

$$\boldsymbol{w}^n = (\Delta_0 w^n, \ldots, \Delta_m w^n),$$

где $\Delta_j g$, как и в § 6, обозначает разности

$$\Delta_j g = g(t_{j+1}) - g(t_j), \quad t_{j+1} > t_j, \quad j = 0, \ldots, m, \quad t_0 = 0, \quad t_{m+1} = 1.$$

Через $\boldsymbol{w}^0$ обозначим аналогичный вектор для процесса $w^0(t)$. В силу второй теоремы непрерывности для доказательства леммы нам достаточно показать, что $\boldsymbol{w}^n \Rightarrow \boldsymbol{w}^0$.

Найдем характеристические функции $\boldsymbol{w}^n$ и $\boldsymbol{w}^0$. Для вектора $\boldsymbol{u} = (u_0, \ldots, u_m)$ имеем

$$\mathbf{E} e^{i\boldsymbol{w}^0 \boldsymbol{u}^\top} = \mathbf{E} \exp\left\{ i \sum_{j=0}^{m} u_j \Delta_j w^0 \right\} = \mathbf{E} \exp\left\{ i \sum_{j=0}^{m} u_j (\Delta_j w - w(1)\Delta_j) \right\},$$

где $\Delta_j = t_{j+1} - t_j$, $j = 0, \ldots, m$, $w(t)$ — стандартный винеровский процесс.

Представим показатель у экспоненты как сумму независимых величин. Обозначив для краткости $\sum_{j=0}^{m} u_j \Delta_j = U$, получим

$$\sum_{j=0}^{m} u_j (\Delta_j w - w(1)\Delta_j) = \sum_{j=0}^{m} (u_j - U)\Delta_j w.$$

Так как $\mathbf{E}\exp\{iuw(s)\} = \exp\{-u^2 s/2\}$, то

$$\mathbf{E}\exp\{iw^0 u^\top\} = \\ = \exp\left\{-\frac{1}{2}\sum_{j=0}^{m}(u_j - U)^2\Delta_j\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m}u_j^2\Delta_j - U^2\right)\right\}. \quad (1)$$

Рассмотрим теперь $\mathbf{E}\exp\{iw^n u^\top\}$. Пусть, как и прежде (см. § 6),

$$\pi_n(t) = nF_n^*(t).$$

Тогда, как мы уже знаем (см. (6.1)),

$$\mathbf{P}(\Delta_0\pi_n = k_0, \ldots, \Delta_m\pi_n = k_m) = \frac{n!}{k_0!\ldots k_m!}\Delta_0^{k_0}\ldots\Delta_m^{k_m}.$$

В правой части стоят члены разложения полинома $(\Delta_0 + \ldots + \Delta_m)^n$. Используя этот довод, получаем

$$\mathbf{E}\exp\left\{i\sum_{j=0}^{m}u_j\Delta_j\pi_n\right\} = \sum\frac{n!}{k_0!\ldots k_m!}(\exp[iu_0]\Delta_0)^{k_0}\ldots(\exp[iu_m]\Delta_m)^{k_m} = \\ = (\exp[iu_0]\Delta_0 + \ldots + \exp[iu_m]\Delta_m)^n.$$

Так как

$$\Delta_j w^n = \sqrt{n}\,(F_n^*(t_{j+1}) - F_n^*(t_j) - \Delta_j) = \frac{\Delta_j\pi_n - n\Delta_j}{\sqrt{n}},$$

то

$$\mathbf{E}\exp\{iw^n u^\top\} = \exp\left\{-i\sum_{j=0}^{m}u_j\sqrt{n}\,\Delta_j\right\}\mathbf{E}\exp\left\{\frac{i}{\sqrt{n}}\sum_{j=0}^{m}u_j\Delta_j\pi_n\right\} = \\ = \exp\{i\sqrt{n}U\}\left(\sum_{j=0}^{m}\exp\left\{\frac{iu_j}{\sqrt{n}}\right\}\Delta_j\right)^n.$$

Отсюда для фиксированного u, пользуясь равенствами

$$\exp\alpha = 1 + \alpha + \frac{\alpha^2}{2} + O(\alpha^3), \quad \ln(1+\alpha) = \alpha - \frac{\alpha^2}{2} + O(\alpha^3)$$

при $\alpha = o(1)$, находим

$$\ln \mathbf{E} \exp\{iw^n u^\top\} = -iU\sqrt{n} + n \ln \left[1 - \sum_{j=0}^{m} \left(1 - \exp\left\{\frac{iu_j}{\sqrt{n}}\right\}\right)\Delta_j\right] =$$

$$= -iU\sqrt{n} + n \ln \left[1 + \sum_{j=0}^{m} \left(i\frac{u_j}{\sqrt{n}} - \frac{u_j^2}{2n} + O(n^{-3/2})\right)\Delta_j\right] =$$

$$= -iU\sqrt{n} + n \left[\frac{iU}{\sqrt{n}} - \frac{1}{2n}\sum_{j=0}^{m} u_j^2 \Delta_j + \frac{U^2}{2n} + O(n^{-3/2})\right] =$$

$$= \frac{1}{2}\left[-\sum_{j=0}^{m} u_j^2 \Delta_j + U^2\right] + O(n^{-1/2}).$$

Сравнивая с (1), мы видим, что при $n \to \infty$

$$\mathbf{E} \exp\{iw^n u^\top\} \to \mathbf{E} \exp\{iw^0 u^\top\}. \tag{2}$$

Остается воспользоваться теоремой непрерывности для характеристических функций многомерных распределений (см. [17, § 7.6]). $\lhd$

Лемма 2. *Для любого* $\varepsilon > 0$

$$\limsup_{n\to\infty} \mathbf{P}(\omega_\Delta(w^n) > \varepsilon) \to 0 \tag{3}$$

при $\Delta \to 0$, *где* $\omega_\Delta(y)$ *есть модуль непрерывности функции* $y \in D(0,1)$:

$$\omega_\Delta(y) = \sup_{\substack{0 \leqslant t_1 < t_2 \leqslant 1 \\ |t_1 - t_2| \leqslant \Delta}} |y(t_1) - y(t_2)|.$$

Доказательство. Не ограничивая общности, мы можем рассматривать лишь двоично-рациональные $\Delta = 2^{-l}$. Имеем при $m > l$

$$\omega_\Delta(w^n) \leqslant \omega_\Delta^{[m]} + 2 \max_{k \leqslant 2^m} \omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right),$$

где

$$\omega_\Delta^{[m]} = \max_{|(k-j)/2^m| \leqslant \Delta} \left|w^n\left(\frac{k}{2^m}\right) - w^n\left(\frac{j}{2^m}\right)\right|,$$

$$\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) = \sup_{(k-1)/2^m \leqslant u < k/2^m} \left|w^n\left(\frac{k}{2^m}\right) - w^n(u)\right|.$$

Чтобы доказать (3), рассмотрим

$$\mathbf{P}(\omega_\Delta(w^n) \geqslant 3\varepsilon) \leqslant \mathbf{P}(\omega_\Delta^{[m]} \geqslant \varepsilon) + \mathbf{P}\left(\bigcup_{k=1}^{2^m} \left\{\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon\right\}\right). \tag{4}$$

Возьмем здесь первое слагаемое. Нетрудно видеть, что при $l > 3$ событие

$$\bigcap_{r=l}^{m}\bigcap_{k=1}^{2^r}\left\{\left|w^n\left(\frac{k}{2^r}\right)-w^n\left(\frac{k-1}{2^r}\right)\right|<\frac{\varepsilon}{r^2}\right\}$$

влечет за собой $\{\omega_\Delta^{[m]}<\varepsilon\}$. Поскольку для дополнительных событий имеет место обратное включение, то

$$\mathbf{P}\left(\omega_\Delta^{[m]}\geqslant\varepsilon\right)\leqslant\mathbf{P}\left(\bigcup_{r=l}^{m}\bigcup_{k=1}^{2^r}\left\{\left|w^n\left(\frac{k}{2^r}\right)-w^n\left(\frac{k-1}{2^r}\right)\right|\geqslant\frac{\varepsilon}{r^2}\right\}\right). \tag{5}$$

Но

$$w^n\left(\frac{k}{2^r}\right)-w^n\left(\frac{k-1}{2^r}\right)=\sqrt{n}\left(F_n^*\left(\frac{k}{2^r}\right)-F_n^*\left(\frac{k-1}{2^r}\right)-\frac{1}{2^r}\right),$$

где

$$n\left(F_n^*\left(\frac{k}{2^r}\right)-F_n^*\left(\frac{k-1}{2^r}\right)\right)$$

есть частота попадания элементов выборки в интервал длины 2^{-r}. Другими словами, это есть сумма S_n случайных величин в схеме Бернулли с n испытаниями и вероятностью исхода 1, равной $p=2^{-r}$. Так как (см. [17, § 5.1])

$$\mathbf{E}(S_n-np)^4=n(p(1-p)^4+(1-p)p^4)+3n(n-1)p^2(1-p)^2\leqslant np+3n^2p^2,$$

то по неравенству типа Чебышева

$$\mathbf{P}\left(\left|w^n\left(\frac{k}{2^r}\right)-w^n\left(\frac{k-1}{2^r}\right)\right|\geqslant\frac{\varepsilon}{r^2}\right)=$$

$$=\mathbf{P}\left(|S_n-np|\geqslant\frac{\varepsilon\sqrt{n}}{r^2}\right)\leqslant\frac{(np+3n^2p^2)r^8}{\varepsilon^4n^2}=\frac{r^8}{\varepsilon^4 2^r n}+\frac{3r^8}{\varepsilon^4 2^{2r}}.$$

Следовательно, правая часть в (5) не превосходит

$$\sum_{r=l}^{m}\left[\frac{r^8}{\varepsilon^4 n}+\frac{3r^8}{\varepsilon^4 2^r}\right]\leqslant c\left(\frac{m^9}{\varepsilon^4 n}+\frac{l^8}{\varepsilon^4 2^l}\right),$$

где c — некоторая абсолютная постоянная $\left(\sum_{r=l}^{m}r^8\sim m^9/9\text{ при }m\to\infty,\ \sum_{r=l}^{\infty}r^8 2^{-r}\sim\right.$ $\left.\sim 2l^8 2^{-l}\text{ при }l\to\infty\right)$. Полагая $m=3\log_2 n$, получим, что

$$\limsup_{n\to\infty}\mathbf{P}\left(\omega_\Delta^{[m]}\geqslant\varepsilon\right)\leqslant c\frac{l^8}{\varepsilon^4 2^l}.$$

Это выражение может быть выбором l (или Δ) сделано сколь угодно малым.

Оценим теперь второе слагаемое в (4), не превосходящее

$$2^m \mathbf{P}\left(\omega\left(\frac{k-1}{2^m}, \frac{k}{2^m}\right) \geqslant \varepsilon\right). \tag{6}$$

Событие, стоящее здесь под знаком вероятности, означает при выбранном m, что на интервале $((k-1)/n^3,\ k/n^3)$ шириной n^{-3} отклонение $n\big(F_n^*(u)-u\big)$ от $n\big(F_n^*(k/n^3) - k/n^3\big)$ превзойдет $\sqrt{n\varepsilon}$. Так как $\sqrt{n\varepsilon} \geqslant 3$ при достаточно большом n, то для этого в интервал $((k-1)/n^3,\ k/n^3)$ должно попасть по крайней мере два элемента выборки X, т. е. должно произойти событие $\{S_n \geqslant 2\}$, если воспользоваться снова обозначениями для схемы Бернулли при $p = n^{-3}$. Но так как

$$1 = (1-p+p)^n = (1-p)^n + np(1-p)^{n-1} + O(n^2p^2),$$

то

$$\mathbf{P}(S_n \geqslant 2) = 1 - (1-p)^n - np(1-p)^{n-1} = O(n^2p^2).$$

Таким образом, (6) не превосходит $n^3 O(n^{-4}) = O(n^{-1}) = o(1)$. Это заканчивает доказательство леммы. $\triangleleft$

Доказательство теоремы 1. Для любого $x \in D(0,1)$ положим

$$\|x\| = \sup_{0\leqslant t\leqslant 1} |x(t)|, \quad f_\varepsilon^+(x) = \sup_{\|y-x\|\leqslant\varepsilon} f(y), \quad f_\varepsilon^-(x) = \inf_{\|y-x\|\leqslant\varepsilon} f(y)$$

и обозначим через x_Δ непрерывную ломаную с узлами в точках $(k\Delta, x(k\Delta) = x_\Delta(k\Delta))$, $k = 0, \ldots, 1/\Delta$, где Δ нацело делит 1. Заметим, что

$$\|x - x_\Delta\| \leqslant \omega_\Delta(x) \tag{7}$$

и что $f_\varepsilon^\pm(x_\Delta)$ суть непрерывные функции от вектора $(x(0), x(\Delta), x(2\Delta), \ldots, x(1))$. В силу леммы 1 и второй теоремы непрерывности

$$f_\varepsilon^\pm(w_\Delta^n) \Rightarrow f_\varepsilon^\pm(w_\Delta^0). \tag{8}$$

при $n \to \infty$. Кроме того, из непрерывности w^0 и функционала f следует, что

$$\|w_\Delta^0 - w^0\| \leqslant \omega_\Delta(w^0) \underset{P}{\longrightarrow} 0 \quad \text{при} \quad \Delta \to 0, \tag{9}$$

$$f_\varepsilon^\pm(w^0) \underset{P}{\longrightarrow} f(w^0) \quad \text{при} \quad \varepsilon \to 0. \tag{10}$$

Из определения f_ε^- следует, что $f_\varepsilon^-(y) \leqslant f(x)$ на множестве $\|y-x\| \leqslant \varepsilon$. Поэтому

$$\mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}\big(f_\varepsilon^-(w_\Delta^n) \leqslant t;\ \|w_\Delta^n - w^n\| \leqslant \varepsilon\big) + \\ + \mathbf{P}\big(\|w_\Delta^n - w^n\| > \varepsilon\big) \leqslant \mathbf{P}\big(f_\varepsilon^-(w_\Delta^n) \leqslant t\big) + \mathbf{P}(\omega_\Delta(w^n) > \varepsilon).$$

Переходя здесь к пределу при $n \to \infty$ и используя (8), (9), получим

$$\limsup_{n\to\infty} \mathbf{P}(f(w^n) \leqslant t) \leqslant \mathbf{P}\big(f_\varepsilon^-(w_\Delta^0) \leqslant t\big) + \limsup_{n\to\infty} \mathbf{P}(\omega_\Delta(w^n) > \varepsilon). \tag{11}$$

Аналогично находим

$$\mathbf{P}\left(f_{\varepsilon}^{-}(w_{\Delta}^{0}) \leqslant t\right) \leqslant \mathbf{P}\left(f_{2\varepsilon}^{-}(w^{0}) \leqslant t\right) + \mathbf{P}(\omega_{\Delta}(w^{0}) > \varepsilon).$$

Подставим теперь последнее выражение в (11) и перейдем к пределу при $\Delta \to \infty$. Из (9) и леммы 2 получим тогда, что

$$\limsup_{n\to\infty} \mathbf{P}(f(w^{n}) \leqslant t) \leqslant \mathbf{P}\left(f_{2\varepsilon}^{-}(w^{0}) \leqslant t\right).$$

Отсюда и из (10) вытекает, что

$$\limsup_{n\to\infty} \mathbf{P}(f(w^{n}) \leqslant t) \leqslant \mathbf{P}(f(w^{0}) \leqslant t).$$

Аналогично устанавливается обратное неравенство

$$\limsup_{n\to\infty} \mathbf{P}(f(w^{n}) < t) \geqslant \mathbf{P}(f(w^{0}) < t).$$

Полученные неравенства означают, очевидно, что $f(w^{n}) \Rightarrow f(w^{0})$. $\triangleleft$

Рассмотрим еще одну функциональную предельную теорему для эмпирических процессов, весьма близкую к теореме 1.

Пусть наряду с выборкой X объема n_1 нам дана независимая от нее выборка Y объема n_2 из того же равномерного на $[0,1]$ распределения. В условиях этого раздела нам удобнее будет эмпирические функции распределения выборок X и Y обозначать соответственно через $F_{X}^{*}(t)$ и $F_{Y}^{*}(t)$. Положим

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left(F_{X}^{*}(t) - F_{Y}^{*}(t)\right).$$

Теорема 2. *Если функционал f удовлетворяет условиям теоремы* 1, *то при* $n_1 \to \infty$, $n_2 \to \infty$

$$f(w_{X,Y}) \Rightarrow f(w^{0}).$$

Доказательство. Докажем эту теорему при упрощающем предположении, что

$$a \equiv \frac{n_2}{n_1+n_2} \to a_0 \in [0,1]$$

при $n = \min(n_1, n_2) \to \infty$. Имеем

$$w_{X,Y}(t) = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\left(F_{X}^{*}(t) - t\right) - \left(F_{Y}^{*}(t) - t\right)\right] = \sqrt{a}\, w_{X}(t) + \sqrt{1-a}\, w_{Y}(t), \quad (12)$$

где $w_X(t)$ и $w_Y(t)$ — эмпирические процессы, соответствующие выборкам X и Y.

Так как $\omega_{\Delta}(x+y) \leqslant \omega_{\Delta}(x) + \omega_{\Delta}(y)$, то из (12) и леммы 2 сразу вытекает аналог леммы 2 для процесса $w_{X,Y}(t)$: для любого $\varepsilon > 0$

$$\limsup_{n\to\infty} \mathbf{P}(\omega_{\Delta}(w_{X,Y} > \varepsilon) \to 0.$$

Сходимость конечномерных распределений $w_{X,Y}$ и w^0 также вытекает из (12). Действительно, обозначим через $\boldsymbol{w}_{X,Y}$, $\boldsymbol{w}_X$, $\boldsymbol{w}_Y$ векторы, построенные по процессам $w_{X,Y}(t)$, $w_X(t)$, $w_Y(t)$ точно так же, как строился вектор $\boldsymbol{w}^n$ по процессу $w^n(t)$. Тогда, используя независимость X и Y и доказательство леммы 1, получим

$$\mathbf{E}\exp\{i\boldsymbol{w}_{X,Y}\boldsymbol{u}^\top\} = \mathbf{E}\exp\{i\sqrt{a}\boldsymbol{w}_X\boldsymbol{u}^\top\}\mathbf{E}\exp\{\sqrt{1-a}\boldsymbol{w}_Y\boldsymbol{u}^\top\} \to$$

$$\to \exp\left\{-\frac{a_0+(1-a_0)}{2}\left(\sum_{j=0}^{m}u_j^2\Delta_j - U^2\right)\right\} = \exp\left\{-\frac{1}{2}\left(\sum_{j=0}^{m}u_j^2\Delta_j - U^2\right)\right\} =$$

$$= \mathbf{E}\exp\{iw^0\boldsymbol{u}^\top\}.$$

В остальном доказательство теоремы 2 ничем не отличается от доказательства теоремы 1. $\triangleleft$

Приложение III

Свойства условных математических ожиданий

В § 19 перечислены основные свойства у. м. о. Ниже приводятся доказательства этих свойств, расположенных в том же порядке, что и в § 19.

1a. $\mathbf{E}(c\xi/\mathfrak{A}) = c\mathbf{E}(\xi/\mathfrak{A})$.

1b. $\mathbf{E}(\xi_1 + \xi_2/\mathfrak{A}) = \mathbf{E}(\xi_1/\mathfrak{A}) + \mathbf{E}(\xi_2/\mathfrak{A})$.

1c. *Если* $\xi_1 \leqslant \xi_2$ *п. н., то* $\mathbf{E}(\xi_1/\mathfrak{A}) \leqslant \mathbf{E}(\xi_2/\mathfrak{A})$ *п. н.*

Чтобы доказать свойство 1a, надо убедиться согласно определению 19.2, что

1) $c\mathbf{E}(\xi/\mathfrak{A})$ есть $\mathfrak{A}$-измеримая функция,

2) $\mathbf{E}(c\mathbf{E}(\xi/\mathfrak{A}); A) = \mathbf{E}(c\xi; A)$ для любого $A \in \mathfrak{A}$.

Выполнение первого свойства очевидно. Второе свойство вытекает из свойств линейности обычного математического ожидания (или обычного интеграла):

$$\mathbf{E}(c\mathbf{E}(\xi/\mathfrak{A}); A) = c\mathbf{E}(\mathbf{E}(\xi/\mathfrak{A}); A) = c\mathbf{E}(\xi; A) = \mathbf{E}(c\xi; A).$$

Свойство 1b доказывается точно так же.

Чтобы доказать свойство 1c, положим для краткости $\widehat{\xi}_j = \mathbf{E}(\xi_j/\mathfrak{A})$. Тогда для любого $A \in \mathfrak{A}$

$$\int\limits_A \widehat{\xi}_1 \, d\mathbf{P} = \mathbf{E}(\widehat{\xi}_1; A) = \mathbf{E}(\xi_1; A) \leqslant \mathbf{E}(\xi_2; A) = \int\limits_A \widehat{\xi}_2 \, d\mathbf{P}, \quad \int\limits_A (\widehat{\xi}_2 - \widehat{\xi}_1) \, d\mathbf{P} \geqslant 0.$$

Отсюда вытекает, что $\widehat{\xi}_2 - \widehat{\xi}_1 \geqslant 0$ п. н.

2. Неравенство Чебышева. *Если* $\xi \geqslant 0$, $x > 0$, *то*

$$\mathbf{P}(\xi \geqslant x/\mathfrak{A}) \leqslant \frac{\mathbf{E}(\xi/\mathfrak{A})}{x}.$$

Это свойство вытекает из 1c, так как $\mathbf{P}(\xi \geqslant x/\mathfrak{A}) = \mathbf{E}(I_{\{\xi \geqslant x\}}/\mathfrak{A})$, где I_A — индикатор события A и справедливо неравенство $I_{\{\xi \geqslant x\}} \leqslant \xi/x$.

3. *Если* $\mathfrak{A}$ *и* $\sigma(\xi)$ *независимы, то* $\mathbf{E}(\xi/\mathfrak{A}) = \mathbf{E}\xi$.

Так как $\widehat{\xi} = \mathbf{E}\xi$ является $\mathfrak{A}$-измеримой функцией, то нам остается проверить лишь второе условие определения 19.2: для любого $A \in \mathfrak{A}$

$$\mathbf{E}(\widehat{\xi}; A) = \mathbf{E}(\xi; A).$$

Справедливость этого равенства вытекает из независимости случайных величин I_A и ξ и соотношений

$$\mathbf{E}(\xi; A) = \mathbf{E}(\xi I_A) = \mathbf{E}\xi \cdot \mathbf{E}I_A = \widehat{\xi}\mathbf{P}(A) = \mathbf{E}(\widehat{\xi}; A).$$

4. Теорема о монотонной сходимости. *Если* $0 \leqslant \xi_n \uparrow \xi$ *п. н., то* $\mathbf{E}(\xi_n/\mathfrak{A}) \uparrow \mathbf{E}(\xi/\mathfrak{A})$ *п. н.*

Действительно, из $\xi_{n+1} \geqslant \xi_n$ п. н. следует $\widehat{\xi}_{n+1} \geqslant \widehat{\xi}_n$ п. н., где $\widehat{\xi}_n = \mathbf{E}(\xi_n/\mathfrak{A})$. Поэтому существует $\mathfrak{A}$-измеримая $\widehat{\xi}$ такая, что $\widehat{\xi}_n \uparrow \widehat{\xi}$ п. н. В силу обычной теоремы о монотонной сходимости для любого $A \in \mathfrak{A}$

$$\int\limits_A \widehat{\xi}_n \, d\mathbf{P} \to \int\limits_A \widehat{\xi} \, d\mathbf{P}, \quad \int\limits_A \xi_n \, d\mathbf{P} \to \int\limits_A \xi \, d\mathbf{P}.$$

Так как левые части этих соотношений совпадают, то совпадают и правые. Это и означает, что $\widehat{\xi} = \mathbf{E}(\xi/\mathfrak{A})$.

5. *Если* η *вещественна и* $\mathfrak{A}$*-измерима, то*

$$\mathbf{E}(\eta\xi/\mathfrak{A}) = \eta\mathbf{E}(\xi/\mathfrak{A}). \tag{1}$$

Если $\eta = I_B$ (индикатор множества $B \in \mathfrak{A}$), то утверждение справедливо, так как при любом $A \in \mathfrak{A}$

$$\int\limits_A \mathbf{E}(I_B\xi/\mathfrak{A}) \, d\mathbf{P} = \int\limits_A I_B\xi \, d\mathbf{P} = \int\limits_{AB} \xi \, d\mathbf{P} = \int\limits_{AB} \mathbf{E}(\xi/\mathfrak{A}) \, d\mathbf{P} = \int\limits_A I_B\mathbf{E}(\xi/\mathfrak{A}) \, d\mathbf{P}.$$

Отсюда и из линейности у. м. о. следует, что утверждение справедливо и для любых простых функций η.

Если $\xi \geqslant 0$, $\eta \geqslant 0$, то, взяв последовательность простых функций $0 \leqslant \eta_n \uparrow \eta$ и пользуясь теоремой о монотонной сходимости в равенстве

$$\mathbf{E}(\eta_n\xi/\mathfrak{A}) = \eta_n\mathbf{E}(\xi/\mathfrak{A}),$$

получим (1). Переход к случаю произвольных ξ и η осуществляется обычным образом — путем рассмотрения положительных и отрицательных частей случайных величин ξ и η. При этом для того чтобы полученные разности и суммы имели смысл, нужно требовать существования $\mathbf{E}|\xi| < \infty$, $\mathbf{E}|\xi\eta| < \infty$.

6. Неравенство Коши–Буняковского

$$\mathbf{E}(\xi_1\xi_2/\mathfrak{A}) \leqslant \left[\mathbf{E}(\xi_1^2/\mathfrak{A})\mathbf{E}(\xi_2^2/\mathfrak{A})\right]^{1/2}$$

доказывается точно так же, как для обычных математических ожиданий (см., например, [17, § 4.7, 4.8]), поскольку доказательство не использует других свойств математических ожиданий, кроме линейности.

Неравенство Йенсена

$$g(\mathbf{E}(\xi/\mathfrak{A})) \leqslant \mathbf{E}(g(\xi)/\mathfrak{A}) \tag{2}$$

для любой выпуклой вниз функции g вытекает из следующих соотношений (ср. с [17]). В силу выпуклости $g(x)$ для каждого y найдется число $g_1(y)$ такое, что

$$g(x) \leqslant g(y) + (x - y)g_1(y).$$

Положим здесь $x = \xi$, $y = \widehat{\xi} = \mathbf{E}(\xi/\mathfrak{A})$ и возьмем у.м.о. обеих частей этого неравенства. Так как в силу свойства 5

$$\mathbf{E}[(\xi - \widehat{\xi})g_1(\widehat{\xi})/\mathfrak{A}] = g_1(\widehat{\xi})\mathbf{E}[\xi - \widehat{\xi}/\mathfrak{A}] = 0,$$

то мы получим (2).

7. Формула полной вероятности следует из свойства 8, если в качестве $\mathfrak{A}$ взять тривиальную σ-алгебру.

8. *Если* $\mathfrak{A} \subset \mathfrak{A}_1 \subset \mathfrak{F}$, *то справедлива формула «последовательного усреднения»*

$$\mathbf{E}(\xi/\mathfrak{A}) = \mathbf{E}(\mathbf{E}(\xi/\mathfrak{A}_1)\mathfrak{A}).$$

Действительно, для любого $A \in \mathfrak{A}$ в силу того, что $A \in \mathfrak{A}_1$

$$\int\limits_A \mathbf{E}(\mathbf{E}(\xi/\mathfrak{A}_1)/\mathfrak{A})\,d\mathbf{P} = \int\limits_A \mathbf{E}(\xi/\mathfrak{A}_1)\,d\mathbf{P} = \int\limits_A \xi\,d\mathbf{P} = \int\limits_A \mathbf{E}(\xi/\mathfrak{A})\,d\mathbf{P}.$$

В заключение отметим, что свойство 5 при широких предположениях допускает следующее обобщение.

5А. *Если* η *измерима относительно* $\mathfrak{A}$, $\varphi(\omega, \eta)$ *есть измеримая функция переменных* $\omega \in \Omega$ *и* $\eta \in \mathbb{R}^k$, *то*

$$\mathbf{E}(\varphi(\omega, \eta)/\mathfrak{A}) = \psi(\omega, \eta), \tag{3}$$

где $\psi(\omega, y) = \mathbf{E}(\varphi(\omega, y)/\mathfrak{A})$.

Докажем это свойство в предположении, что существует последовательность простых функций η_n такая, что $\varphi(\omega, \eta_n) \uparrow \varphi(\omega, \eta)$, $\psi(\omega, \eta_n) \uparrow \psi(\omega, \eta)$ п.н. Действительно, пусть $\eta_n = y_k$ при $\omega \in A_k \subset \mathfrak{A}$. Тогда

$$\varphi(\omega, \eta_n) = \sum_k \varphi(\omega, y_k) I_{A_k}.$$

В силу свойства 5 отсюда следует выполнение (3) для функций η_n. Остается воспользоваться теоремой о монотонной сходимости (свойство 4) в равенстве

$$\mathbf{E}(\varphi(\omega, \eta_n)/\mathfrak{A}) = \psi(\omega, \eta_n).$$

Приложение IV

Закон больших чисел и центральная предельная теорема. Равномерные варианты

1. Закон больших чисел в схеме серий. Рассмотрим последовательность произвольных серий $\xi_{1,n},\dots,\xi_{n,n}$, $n=1,2,\dots$, независимых случайных величин, где распределения $\xi_{k,n}$ могут зависеть от n.

Обозначим $\zeta_n = \sum_{k=1}^{n} \xi_{k,n}$. С точки зрения последующих результатов можно считать, не ограничивая общности, что

$$\mathbf{E}\xi_{k,n} = 0. \tag{1}$$

Предположим, что выполнено условие

$$D_1 \equiv \sum_{k=1}^{n} \mathbf{E} \min\left(|\xi_{k,n}|, |\xi_{k,n}|^2\right) \to 0 \tag{D_1}$$

при $n \to \infty$.

Теорема 1 (закон больших чисел). *При выполнении условий* (1), (D_1) *для любого* $\varepsilon > 0$

$$\mathbf{P}(|\zeta_n| > \varepsilon) \to 0$$

или, что то же, $\zeta_n \xrightarrow[P]{} 0$.

Рассмотрим внимательнее условие (D_1). Очевидно, что его можно записать также в форме

$$D_1 = \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > 1) + \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^2; |\xi_{k,n}| \leqslant 1) \to 0.$$

Если

$$M_1 \equiv \sum_{k=1}^{n} \mathbf{E}|\xi_{k,n}| \leqslant c < \infty \tag{2}$$

при $n \to \infty$, то для выполнения (D_1) необходимо и достаточно выполнение условия

$$M_1(\tau) \equiv \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > \tau) \to 0 \tag{M_1}$$

при любом $\tau > 0$ и $n \to \infty$, которое можно назвать условием типа Линдеберга (условие Линдеберга (M_2) будет введено в следующем разделе. Действительно

$(\tau \leqslant 1,\ g_1(x) = \min(|x|, |x|^2))$,

$$D_1 = \sum_{k=1}^{n} \mathbf{E} g_1(\xi_{k,n}) \leqslant \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > \tau) + \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^2; |\xi_{k,n}| \leqslant \tau) \leqslant$$

$$\leqslant M_1(\tau) + \tau \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| \leqslant \tau) \leqslant M_1(\tau) + \tau M_1(0). \quad (3)$$

Так как $M_1(0) = M_1 \leqslant c$, а τ может быть произвольно малым, то $D_1 \to 0$ при $n \to \infty$.

Обратно (здесь условие (2) не требуется), при $\tau \leqslant 1$

$$M_1(\tau) \leqslant \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > 1) + \tau^{-1} \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^2; \tau < |\xi_{k,n}| \leqslant 1) \leqslant \tau^{-1} D_1 \to 0 \quad (4)$$

при $n \to \infty$ и при любом $\tau > 0$.

Каждое из условий (D_1), (M_1) влечет за собой равномерную малость $\mathbf{E}|\xi_{k,n}|$:

$$\max_{1 \leqslant k \leqslant n} \mathbf{E}|\xi_{k,n}| \to 0 \quad \text{при} \quad n \to \infty. \quad (5)$$

Действительно, (M_1) означает, что существует достаточно медленно убывающая последовательность $\tau_n \to 0$ такая, что $M_1(\tau_n) \to 0$. Поэтому

$$\max_{k \leqslant n} \mathbf{E}|\xi_{k,n}| \leqslant \max_{k \leqslant n} [\tau_n + \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > \tau_n)] \leqslant \tau_n + M_1(\tau_n) \to 0. \quad (6)$$

Из (5) следует, в частности, свойство пренебрежимой малости слагаемых $\xi_{k,n}$.

Будем говорить, что $\xi_{k,n}$ *пренебрежимо малы* или, что то же, *обладают свойством* (S), если при любом $\varepsilon > 0$

$$\max_{k \leqslant n} \mathbf{P}(|\xi_{k,n}| > \varepsilon) \to 0. \quad (\mathrm{S})$$

Свойство (S) можно было бы назвать также *равномерной сходимостью $\xi_{k,n}$ по вероятности к нулю*. Свойство (S) сразу следует из (5) и неравенства Чебышева. Оно следует также из более сильных соотношений, вытекающих из (M_1):

$$\mathbf{P}\left(\max_{k \leqslant n} |\xi_{k,n}| > \varepsilon\right) = \mathbf{P}(\cup\{|\xi_{k,n}| > \varepsilon\}) \leqslant$$

$$\leqslant \sum_{k \leqslant n} \mathbf{P}(|\xi_{k,n}| > \varepsilon) \leqslant \varepsilon^{-1} \sum_{k \leqslant n} \mathbf{E}(|\xi_{k,n}|; |\xi_{k,n}| > \varepsilon) \to 0. \quad (\mathrm{S}_1)$$

Доказательство теоремы 1. Положим

$$\varphi_{k,n}(t) = \mathbf{E} e^{it\xi_{k,n}}, \quad \Delta_k(t) = \varphi_{k,n}(t) - 1.$$

Надо доказать, что при каждом t и при $n \to \infty$

$$\varphi_{\zeta_n}(t) = \mathbf{E} e^{it\zeta_n} = \prod_{k=1}^{n} \varphi_{k,n}(t) \to 1.$$

Нам понадобятся следующие неравенства (см., например, [17, § 7.5, леммы 2, 3]). *Если* $\operatorname{Re}\beta \leqslant 0$, *то*

$$|e^{\beta}-1| \leqslant |\beta|, \quad |e^{\beta}-1-\beta| \leqslant \frac{|\beta|^2}{2}, \quad \left|e^{\beta}-1-\beta-\frac{\beta^2}{2}\right| \leqslant \frac{|\beta|^3}{6}. \tag{7}$$

Если $|a_k| \leqslant 1$, $|b_k| \leqslant 1$, $k=1,\ldots,n$, *то*

$$\left|\prod_1^n a_k - \prod_1^n b_k\right| \leqslant \sum_1^n |a_k - b_k|. \tag{8}$$

В силу (8)

$$|\varphi_{\zeta_n}(t)-1| = \left|\prod_{k=1}^n \varphi_{k,n}(t) - \prod_{k=1}^n 1\right| \leqslant \sum_{k=1}^n |\Delta_k(t)| =$$

$$= \sum_{k=1}^n |\mathbf{E}(e^{it\xi_{k,n}}-1)| = \sum_{k=1}^n |\mathbf{E}(e^{it\xi_{k,n}}-1-it\xi_{k.n})|.$$

В силу (7) ($g_1(x) = \min(|x|, x^2)$)

$$|e^{itx}-1-itx| \leqslant \min\left(2|tx|, \frac{t^2x^2}{2}\right) \leqslant 2g_1(tx) \leqslant 2h(t)g_1(x),$$

где $h(t) = \max(|t|, |t|^2)$. Поэтому

$$|\varphi_{\zeta_n}(t)-1| \leqslant 2h(t)\sum_{k=1}^n \mathbf{E}g_1(\xi_{k,n}) = 2h(t)D_1 \to 0. \quad \triangleleft$$

Замечание 1. Если $\{\xi_k\}$ есть заданная фиксированная (не зависящая от n) последовательность случайных величин, $S_n = \sum_{k=1}^n \xi_k$, $\mathbf{E}\xi_k = a_k$, то для нее мы будем изучать поведение сумм

$$\zeta_n = \sum_{k=1}^n \xi_{k,n} = \frac{S_n - \sum_{k=1}^n a_k}{b_n},$$

где $\xi_{k,n} = (\xi_k - a_k)/b_n$ удовлетворяют (1), $b_n \to \infty$. Если положить, не ограничивая общности, $a_k = 0$, то условия (D_1), (M_1) будут в этом случае иметь вид

$$D_1 = \frac{1}{b_n}\sum \mathbf{E}\min\left(|\xi_k|, \frac{|\xi_k|^2}{b_n}\right) \to 0,$$

$$M_1(\tau) = \frac{1}{b_n}\sum \mathbf{E}(|\xi_k|; |\xi_k| > \tau b_n) \to 0.$$

Если ξ_k одинаково распределены, то из теоремы 1 вытекает в качестве частного случая «обычный» закон больших чисел: если $\mathbf{E}|\xi_k| < \infty$ (в этом случае выполнено $M_1(\tau)$ при $b_n = n$), то $\zeta_n \xrightarrow[P]{} 0$.

Для выполнения (D_1) достаточно также выполнения следующего условия: *при некотором* s, $2 \geqslant s > 1$,

$$\sum_{k=1}^{n} \mathbf{E}|\xi_{k,n}|^s \to 0. \tag{L_s}$$

Это утверждение очевидно, так как $g_1(x) \leqslant |x|^s$ при $2 \geqslant s > 1$. Условие (L_s) можно было бы назвать условием типа Ляпунова (ср. с условием Ляпунова (L_s) в следующем разделе.

Замечание 2. Теорему 1 можно доказывать и иначе, используя не характеристические функции, а так называемые срезки случайных величин и оценки их дисперсий. Такое доказательство приведено в первом издании книги.

2. Центральная предельная теорема для сумм произвольных независимых случайных величин. Как и в п. 1, мы будем рассматривать серии случайных величин

$$\xi_{1,n}, \ldots, \xi_{n,n}$$

и их суммы

$$\zeta_n = \sum_{k=1}^{n} \xi_{k,n}. \tag{9}$$

Будем предполагать, что $\xi_{k,n}$ имеют конечные вторые моменты $\sigma_{k,n}^2 = \mathbf{D}\xi_{k,n} < \infty$, и считать, не ограничивая общности, что

$$\mathbf{E}\xi_{k,n} = 0, \quad \sum_{k=1}^{n} \sigma_{k,n}^2 = \mathbf{D}\zeta_n = 1. \tag{10}$$

Введем в рассмотрение условие: *при некотором* $s > 2$

$$D_2 \equiv \sum_{k=1}^{n} \mathbf{E} \min\left(\xi_{k,n}^2, |\xi_{k,n}|^s\right) \to 0, \tag{D_2}$$

которое будет играть важную роль в дальнейшем. При этом многие рассмотрения, связанные с условием (D_2) и с условиями (M_2), (L_s), вводимыми ниже, будут вполне аналогичны рассмотрениям, связанным с условиями (D_1), (M_1), (L_s) в п. 1.

При выполнении (10) *необходимым и достаточным для выполнения* (D_2) *является условие Линдеберга*

$$M_2(\tau) \equiv \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^2; |\xi_{k,n}| > \tau) \to 0 \tag{M_2}$$

при любом $\tau > 0$ и $n \to \infty$. Действительно, положим

$$g_2(x) = \min(x^2, |x|^s).$$

Тогда (ср. с (3), (4); $\tau \leqslant 1$)

$$D_2 = \sum_{k=1}^{n} \mathbf{E} g_2(\xi_{k,n}) \leqslant \sum_{k=1}^{n} \mathbf{E}\left(\xi_{k,n}^2; |\xi_{k,n}| > \tau\right) + \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^s; |\xi_{k,n}| \leqslant \tau) \leqslant$$
$$\leqslant M_2(\tau) + \tau^{s-2} M_2(0) = M_2(\tau) + \tau^{s-2}.$$

Так как τ произвольно, то $D_2 \to 0$ при $n \to \infty$.

Обратно, при $\tau \leqslant 1$ (условие 10 здесь не требуется)

$$M_2(\tau) \leqslant \sum_{k=1}^{n} \mathbf{E}\left(\xi_{k,n}^2; |\xi_{k,n}| > 1\right) + \frac{1}{\tau^{s-2}} \sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^s;\ \tau < |\xi_{k,n}| \leqslant 1) \leqslant \frac{1}{\tau^{s-2}} D_2 \to 0$$

при любом $\tau > 0$ и $n \to \infty$. $\triangleleft$

Итак, *при выполнении* (10) *условие* (D_2) *инвариантно по* $s > 2$,

Условию (D_2) можно придать более общую форму:

$$\sum_{k=1}^{n} \mathbf{E}\xi_{k,n}^2 h(\xi_{k,n}) \to 0,$$

где $h(x)$ — любая функция, обладающая свойствами $h(x) > 0$ при $x > 0$, $h(x) \uparrow$, $h(x) \to 0$ при $x \to 0$, $h(x) \to c < \infty$ при $x \to \infty$. При этом все основные свойства условия (D_2) сохранятся. Условие Линдеберга выясняет существо условия (D_2) с несколько иной точки зрения. В нем $h(x) = I_{(\tau,\infty)}$, $\tau \in (0,1)$.

Аналогично тому, как это было сделано в п. 1 относительно условия (M_1), легко проверить, что условие (M_2) влечет за собой сходимость (см. (6))

$$\max_{k \leqslant n} \mathbf{D}\xi_{k,n} \to 0 \tag{11}$$

и пренебрежимую малость $\xi_{k,n}$ (свойство (S)). Очевидно также неравенство

$$M_1(\tau) \leqslant \frac{1}{\tau} M_2(\tau).$$

Если $\{\xi_k\}$ есть заданная фиксированная (не зависящая от n) последовательность независимых случайных величин, $S_n = \sum_{k=1}^{n} \xi_k$, $\mathbf{E}\xi_k = a_k$, $\mathbf{D}\xi_k = \sigma_k^2$, то для нее мы будем изучать поведение нормированных сумм

$$\zeta_n = \frac{S_n - \sum_{k=1}^{n} a_k}{B_n}, \quad B_n^2 = \sum_{k=1}^{n} \sigma_k^2,$$

которые, очевидно, также имеют вид (9), где $\xi_{k,n} = (\xi_k - a_k)/B_n$.

Условия (D_2), (M_2) для ξ_k будут иметь вид

$$D_2 = \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbf{E} \min\left((\xi_k - a_k)^2, \frac{|\xi_k - a_k|^s}{B_n^{s-2}}\right) \to 0,$$
$$M_2(\tau) = \frac{1}{B_n^2} \sum_{k=1}^{n} \mathbf{E}((\xi_k - a_k)^2;\ |\xi_k - a_k| > \tau B_n) \to 0. \tag{12}$$

Теорема 2 (центральная предельная теорема). *Если последовательности независимых случайных величин* $\{\xi_{k,n}\}_{k=1}^{n}$, $n = 1, 2, \ldots$, *удовлетворяют условиям* (10), (D_2), *то* $\mathbf{P}(\zeta_n < x) \to \Phi(x)$ *равномерно по* x.

Доказательство. Достаточно убедиться, что

$$\varphi_{\zeta_n}(t) = \prod_{k=1}^{n} \varphi_{k,n}(t) \to e^{-t^2/2}.$$

В силу неравенств (8)

$$|\varphi_{\zeta_n}(t) - e^{-t^2/2}| = \left|\prod_{k=1}^{n} \varphi_{k,n}(t) - \prod_{k=1}^{n} e^{-t^2\sigma_{k,n}^2/2}\right| \leqslant \sum_{k=1}^{n} |\varphi_{k,n}(t) - e^{-t^2\sigma_{k,n}^2/2}| \leqslant$$

$$\leqslant \sum_{k=1}^{n} \left|\varphi_{k,n}(t) - 1 + \frac{1}{2}t^2\sigma_{k,n}^2\right| + \sum \left|e^{-t^2\sigma_{k,n}^2/2} - 1 + \frac{1}{2}t^2\sigma_{k,n}^2\right|. \quad (13)$$

Так как в силу (7)

$$\left|e^{ix} - 1 - ix + \frac{x^2}{2}\right| \leqslant \min\left(x^2, \frac{|x^3|}{6}\right) \leqslant g_2(x),$$

(при $s \leqslant 3$), то первая сумма в (13) не превосходит

$$\sum_{k=1}^{n} \left|\mathbf{E}\left(e^{it\xi_{k,n}} - 1 - it\xi_{k,n} + \frac{t^2\xi_{k,n}^2}{2}\right)\right| \leqslant$$

$$\leqslant \sum_{k=1}^{n} \mathbf{E}g_2(|t\xi_{k,n}|) \leqslant h(t) \sum_{k=1}^{n} \mathbf{E}g_2(|\xi_{k,n}|) \leqslant h(t)\mathbf{D}_2 \to 0,$$

где $h(t) = \max(t^2, |t|^s)$. Вторая сумма в (13) (снова в силу 7) не превосходит (см. (10), (11))

$$\frac{t^4}{8} \sum_{k=1}^{n} \sigma_{k,n}^4 \leqslant \frac{t^4}{8} \max_k \sigma_{k,n}^2 \to 0 \quad \text{при} \quad n \to \infty. \ \triangleleft$$

Иногда утверждение теоремы 2 бывает удобным в следующей форме.

Следствие 1. *Пусть* $\mathbf{E}\xi_{k,n} = 0$, $\mathbf{D}\zeta_n \to \sigma^2 > 0$ *и выполнено условие* (D_2) *или* (M_2), *тогда* $\zeta_n \Rightarrow \Phi_{0,\sigma^2}$.

Замечание 3. Утверждение об асимптотической нормальности $\zeta_n = \dfrac{S_n - an}{\sigma\sqrt{n}}$, когда ξ_k одинаково распределены, следует из теоремы 2. Условие Линдеберга выполнено, так как в этом случае $B_n^2 = n\sigma^2$, и (12) примет вид

$$M_2(\tau) = \frac{1}{\sigma^2}\mathbf{E}\big((\xi_k - a)^2; |\xi_k - a| > \tau\sigma\sqrt{n}\big) \to 0 \quad \text{при} \quad n \to \infty$$

для любого $\tau > 0$.

Замечание 4. Достаточным условием для выполнения (D_2), (M_2) является более ограничительное *условие Ляпунова*. Проверять его в ряде случаев легче. Предположим, что выполнено (10). Величину

$$L_s = \sum_{k=1}^{n} \mathbf{E}|\xi_{k,n}|^s$$

при некотором $s > 2$ называют *отношением Ляпунова порядка* s. Условие

$$L_s \to 0 \qquad (L_s)$$

называют *условием Ляпунова*. Величина L_s получила название *отношения* потому, что для $\xi_{k,n} = (\xi_k - a_k)/B_n$ ($a_k = \mathbf{E}\xi_k$, $B_n^2 = \mathbf{D}\zeta_n$, ξ_k от n не зависят) L_s имеет вид

$$L_s = \frac{\sum_{k=1}^{n} \mathbf{E}|\xi_k - a_k|^s}{B_n^s}.$$

Если ξ_k одинаково распределены, $a_k = a$, $\mathbf{D}\xi_k = \sigma^2$, $\mathbf{E}|\xi_k - a|^s = \mu < \infty$, то

$$L_s = \frac{\mu}{\sigma^s n^{(s-2)/2}} \to 0.$$

Достаточность условия Ляпунова вытекает из очевидных неравенств $g_2(x) \leqslant |x|^s$ при любом s, $D_2 \leqslant L_s$.

3. Закон больших чисел и центральная предельная теорема в многомерном случае. В этом пункте мы будем предполагать, что $\xi_{1,n}, \ldots, \xi_{n,n}$ есть случайные векторы в схеме серий, $\mathbf{E}\xi_{k,n} = 0$, $\zeta_n = \sum_{k=1}^{n} \xi_{k,n}$.

Закон больших чисел $\zeta_n \xrightarrow[P]{} 0$ получается немедленно из теоремы 1, если предположить, что условиям теоремы 1 удовлетворяют координаты $\xi_{k,n}$. Поэтому ничего не мешает считать, что теорема 1 сформулирована и доказана для векторов.

Чуть сложнее обстоит дело с центральной предельной теоремой. Здесь мы будем предполагать, что $\mathbf{E}|\xi_{k,n}|^2 < \infty$, где $|x|^2 = (x, x)$ есть квадрат нормы x. Обозначим

$$\sigma_{k,n}^2 = \mathbf{E}\xi_{k,n}^{\top}\xi_{k,n}, \quad \sigma_n^2 = \sum_{k=1}^{n} \sigma_{k,n}^2$$

(индекс $\top$ означает транспонирование, так что $\xi_{k,n}^{\top}$ есть вектор-столбец).

Введем в рассмотрение условие

$$\sum_{k=1}^{n} \mathbf{E} \min\left(|\xi_{k,n}|^2, |\xi_{k,n}|^s\right) \to 0, \quad s > 2, \qquad (D_2)$$

и условие Линдеберга

$$\sum_{k=1}^{n} \mathbf{E}(|\xi_{k,n}|^2; |\xi_{k,n}| > \tau) \to 0 \qquad (M_2)$$

при $n \to \infty$ и любом $\tau > 0$. Как и в одномерном случае, легко убедиться, что условия (M_2) и (D_2) при $\operatorname{tr}\sigma_n^2 < c < \infty$ равносильны.

Теорема 3. *Если* $\sigma_n^2 \to \sigma^2$, *где* σ^2 — *положительно определенная матрица, и выполнено условие* (D_2) (*или* (M_2)), *то*

$$\zeta_n \Rightarrow \Phi_{0,\sigma^2}.$$

Следствие 2 («обычная» центральная предельная теорема). *Если* $\xi_1, \xi_2, \ldots$ — *последовательность независимых одинаково распределенных векторов,* $\mathbf{E}\xi_k = 0$, $\sigma^2 = \mathbf{E}\xi_k^\top \xi_k$, $S_n = \sum_1^n \xi_k$, *то при* $n \to \infty$

$$\frac{S_n}{\sqrt{n}} \Rightarrow \Phi_{0,\sigma^2}.$$

Это утверждение является следствием теоремы 3, так как случайные величины $\xi_{k,n} = \xi_k/\sqrt{n}$ удовлетворяют ее условиям.

Доказательство теоремы 3. Рассмотрим характеристические функции

$$\varphi_{k,n}(t) = \mathbf{E}e^{i(t,\xi_{k,n})}, \quad \varphi_n(t) = \mathbf{E}e^{i(t,\zeta_n)} = \prod_{k=1}^n \varphi_{k,n}(t).$$

Для доказательства надо убедиться, что при любом t и при $n \to \infty$

$$\varphi_n(t) \to \exp\left\{-\frac{1}{2}t\sigma^2 t^\top\right\}.$$

Воспользуемся теоремой 2. Функции $\varphi_{k,n}(t)$ и $\varphi_n(t)$ можно рассматривать как характеристические функции

$$\varphi_{k,n}^r(v) = \mathbf{E}\exp\left(iv\xi_{k,n}^r\right), \quad \varphi_n^r(v) = \mathbf{E}\exp\left(iv\zeta_n^r\right)$$

случайных величин $\xi_{k,n}^r = (\xi_{k,n}, r)$, $\zeta_n^r = (\zeta_n, r)$, где $r = t/|t|$, $v = |t|$. Покажем, что скалярные случайные величины $\xi_{k,n}^r$ удовлетворяют условиям теоремы 2 (или следствия 1) для одномерного случая. Очевидно,

$$\mathbf{E}\xi_{k,n}^r = 0,$$

$$\sum_{k=1}^n \mathbf{E}\left(\xi_{k,n}^r\right)^2 = \sum_{k=1}^n \mathbf{E}(\xi_{k,n}, r)^2 = r\sigma_n^2 r^\top \to r\sigma^2 r^\top > 0.$$

Выполнение условия (D_2) следует из очевидных неравенств

$$(\xi_{k,n}, r)^2 = \left|\xi_{k,n}^r\right|^2 \leqslant |\xi_{k,n}|^2,$$

$$\sum_{k=1}^n \mathbf{E}g_2\left(\xi_{k,n}^r\right) \leqslant \sum_{k=1}^n \mathbf{E}g_2(|\xi_{k,n}|),$$

где $g_2(x) = \min(x^2, |x|^s)$, $s > 2$. Таким образом, для любых v и r (т.е. для любых t) в силу следствия 1 теоремы 2

$$\varphi_n(t) = \mathbf{E}\exp\{iv\zeta_n^r\} \to \exp\left\{-\frac{1}{2}v^2 r\sigma^2 r^\top\right\} = \exp\left\{-\frac{1}{2}t\sigma^2 t^\top\right\}. \quad \triangleleft$$

В теореме 3 остается вне поля зрения случай, когда элементы матрицы σ_n^2 неограниченно растут или ведут себя так, что ранг предельной матрицы σ^2 становится меньше размерности векторов $\xi_{k,n}$. Такое может происходить, когда дисперсии разных координат $\xi_{k,n}$ имеют разные порядки убывания (или роста). В этом случае вместо ζ_n следует рассматривать преобразованные суммы $\zeta_n' = \zeta_n\sigma_n^{-1}$. По существу теорема 3 является следствием следующего более общего утверждения, которое в свою очередь вытекает из теоремы 3.

Теорема 3А. *Если случайные величины* $\xi_{k,n}' = \xi_{k,n}\sigma_n^{-1}$ *удовлетворяют условию* (D_2) *(или* (M_2)*), то* $\zeta_n' \Subset \mathbf{\Phi}_{0,E}$, *где* E — *единичная матрица.*

4. Равномерные предельные теоремы для сумм случайных величин, зависящих от параметра. Мы докажем в этом пункте теоремы 37.1, 37.2.

Пусть $X \Subset \mathbf{P}_\theta$, $a(x,\theta)$ — заданная измеримая функция $\mathcal{X}\times\Theta \to \mathbb{R}^l$,

$$s_n(\theta) = \sum_{j=1}^n a(\mathrm{x}_j, \theta).$$

Будем говорить, что интеграл $a(\theta) = \int a(x,\theta)\mathbf{P}_\theta(dx)$ сходится равномерно по θ в области $\Theta_0 \subset \Theta$, если

$$\sup_{\theta\in\Theta_0} \int\limits_{|a(x,\theta)|>N} |a(x,\theta)|\mathbf{P}_\theta(dx) \to 0$$

при $N \to \infty$.

Теорема 4 (равномерный закон больших чисел). *Если интеграл* $a(\theta) = \int a(x,\theta)\mathbf{P}_\theta(dx)$ *сходится равномерно по* θ *в области* $\Theta_0 \subset \Theta$, *то*

$$\zeta_n(\theta) = \frac{s_n(\theta)}{n} - a(\theta) \underset{P_\theta}{\longrightarrow} 0 \tag{14}$$

равномерно по $\theta \in \Theta_0 \to 0$.

Доказательство. Допустим, что (14) не имеет места. Тогда найдутся $\varepsilon > 0$, $\delta > 0$ и последовательность $\theta_n \in \Theta_0$ такие, что

$$\mathbf{P}_{\theta_n}\left(\left|\frac{\zeta_n(\theta_n)}{n}\right| > \varepsilon\right) > \delta \tag{15}$$

при всех n.

Рассмотрим случайные величины

$$\xi_{j,n} = \frac{a(\mathrm{x}_j, \theta_n) - a(\theta_n)}{n}.$$

Нетрудно видеть, что они удовлетворяют условиям теоремы 1. Действительно, положим $A_n = \{x\colon |a(x,\theta_n) - a(\theta_n)| > \tau n\}$. Тогда

$$n\mathbf{E}_{\theta_n}|\xi_{j,n}| \leqslant 2a \equiv 2 \sup_{\theta\in\Theta_0} \int |a(x,\theta)|\mathbf{P}_\theta(dx) < \infty,$$

$$n\mathbf{E}_{\theta_n}(|\xi_{j,n}|; |\xi_{j,n}| > \tau) = \int\limits_{A_n} |a(x,\theta_n) - a(\theta_n)|\mathbf{P}_{\theta_n}(dx) \to 0.$$

Последнее соотношение следует из равномерной сходимости интеграла $a(\theta)$ и неравенства Чебышева

$$\mathbf{P}_{\theta_n}(A_n) \leqslant \frac{\mathbf{E}_{\theta_n}|\xi_{1,n}|}{\tau} \leqslant \frac{2a}{\tau n} \to 0.$$

Сказанное означает, что последовательность $\xi_{j,n}$ удовлетворяет закону больших чисел

$$\mathbf{P}_{\theta_n}\left(\left|\sum_{j=1}^{n} \xi_{j,n}\right| > \varepsilon\right) \to 0$$

для любого $\varepsilon > 0$. Это противоречит (15) и доказывает теорему. $\triangleleft$

Перейдем к *центральной предельной теореме*. Пусть $\mathbf{E}_\theta a(\mathrm{x}_1, \theta) = 0$.

Положим $\sigma^2(\theta) = \|\sigma_{ij}(\theta)\| = \mathbf{E}_\theta a^\top(\mathrm{x}_1,\theta)a(\mathrm{x}_1,\theta)$ и обозначим через $a_j(x,\theta)$, $j = 1,\ldots,l$, координаты векторов $a(x,\theta)$.

Теорема 5 (равномерная центральная предельная теорема). *Пусть интегралы $\sigma_{jj}(\theta) = \mathbf{E}_\theta a_j^2(\mathrm{x}_1,\theta)$ сходятся равномерно в $\Theta_0 \subset \Theta$, т. е.*

$$\begin{gathered} \sup_{\Theta_0} \sigma_{jj}(\theta) < \infty, \\ \sup_{\Theta_0} \mathbf{E}_\theta\left(a_j^2(\mathrm{x}_1,\theta); |a_j(\mathrm{x}_1,\theta)| > N\right) \to 0 \end{gathered} \tag{16}$$

при $N \to \infty$. Тогда

$$\frac{s_n(\theta)}{\sqrt{n}} \Rightarrow \mathbf{\Phi}_{0,\sigma^2(\theta)} \tag{17}$$

при $n \to \infty$ равномерно по $\theta \in \Theta_0$.

Доказательство. Невыполнение (17) будет означать существование последовательности $\theta_n \in \Theta_0$, для которой суммы случайных величин $\xi_{j,n} = a(\mathrm{x}_j,\theta_n)/\sqrt{n}$ не будут по распределению сближаться с $\mathbf{\Phi}_{0,\sigma^2(\theta_n)}$.

В силу компактности замыкания $\{\sigma^2(\theta), \theta \in \Theta_0\}$ последовательность θ_n можно считать выбранной так, что для некоторой матрицы σ^2

$$\sigma^2(\theta_n) = n\mathbf{E}_{\theta_n}\xi_{1,n}^\top\xi_{1,n} \to \sigma^2. \tag{18}$$

Тогда наше допущение о невыполнении (17) будет означать, что $\sum_{j=1}^{n} \xi_{j,n}$ не будет сближаться по распределению с Φ_{0,σ^2}. Но это невозможно в силу теоремы 3, так как $\xi_{j,n}$ удовлетворяют условиям этой теоремы. Действительно, в силу (18) достаточно проверить условие Линдеберга. Для множеств $A_{i,n} = \{|a_i(\mathrm{x}_1, \theta_n)| > \tau\sqrt{n/l}\,\}$ находим

$$\sup_{\theta\in\Theta_0} \mathbf{P}_\theta(A_{i,n}) \leqslant \sup_{\theta\in\Theta_0} \frac{l\sigma_{ii}(\theta)}{n\tau^2} \to 0$$

при $n \to \infty$. Используя тот факт, что $\{|\xi_{1,n}| > \tau\} \subset \bigcup_{i=1}^{l} A_{i,n}$, получаем

$$n\mathbf{E}_{\theta_n}(|\xi_{1,n}|^2; |\xi_{1,n}| > \tau) \leqslant \sum_{i,k=1}^{l} \mathbf{E}_{\theta_n}\big(a_i^2(\mathrm{x}_1, \theta_n); A_{k,n}\big). \tag{19}$$

Здесь $\mathbf{E}_{\theta_n}\big(a_i^2(\mathrm{x}_1, \theta_n); A_{i,n}\big) \to 0$ в силу равномерной сходимости интеграла $\sigma_{ii}(\theta)$. Если $i \neq k$, то, полагая $B_{i,N} = \{|a_i(\mathrm{x}_1, \theta_n)| > N\}$, получим

$$\mathbf{E}_{\theta_n}\big(a_i^2; A_{k,n}\big) = \mathbf{E}_{\theta_n}\big(a_i^2;\ A_{k,n}B_{i,N}\big) + \mathbf{E}_{\theta_n}\big(a_i^2;\ A_{k,n}\overline{B}_{i,N}\big).$$

Здесь для заданного $\varepsilon > 0$ можно выбрать N так, что первое слагаемое в силу (16) будет меньше ε. Второе слагаемое не превосходит $N^2\mathbf{P}_{\theta_n}(A_{k,n}) \to 0$ при $n \to \infty$. Это означает, что (19) сходится к нулю при $n \to \infty$. $\triangleleft$

Приложение V

Некоторые утверждения, относящиеся к интегралам, зависящим от параметра

1. Теоремы о сходимости интегралов, зависящих от параметра. Пусть $\{\psi(t,y)\}$ — семейство измеримых функций, заданных на измеримом пространстве $(\mathcal{Y},\mathfrak{B}_{\mathcal{Y}})$ с мерой ν на нем. Нас будут интересовать условия, при которых

$$\int \psi(t,y)\nu(dy) \to \int \psi(\theta,y)\nu(dy) \quad \text{при} \quad t \to \theta. \tag{1}$$

Пусть $\{A(t) = A(t,\theta), t \in \Theta\}$ есть некоторое семейство множеств из $\mathfrak{B}_{\mathcal{Y}}$. Обозначим через $I_{A(t)}(x)$ индикатор $A(t)$ и через $\overline{A}(t)$ — дополнение к $A(t)$.

Следующее утверждение является некоторым обобщением известной теоремы Лебега.

Теорема 1. *Пусть семейство $\{A(t)\}$ таково, что*

1) $\psi(t,y)I_{A(t)}(y) \to \psi(\theta,y)$ *при* $t \to \theta$ *для п. в.* $[\nu]$ *значений* y, *для которых* $\psi(\theta,y) \neq 0$;

2) $\sup\limits_t |\psi(t,y)I_{A(t)}(y)| \leqslant \psi(y)$, *где* ψ — *интегрируемая функция*

$$\int \psi(y)\nu(dy) < \infty.$$

Тогда для выполнения (1) *необходимо и достаточно, чтобы*

$$\int \psi(t,y)I_{\overline{A}(t)}(y)\nu(dy) \to 0 \quad \text{при} \quad t \to \infty. \tag{2}$$

Доказательство. В силу теоремы Лебега

$$\int \psi(t,y)I_{A(t)}(y)\nu(dy) \to \int \psi(\theta,y)\nu(dy).$$

Так как

$$\int \psi = \int \psi I_A + \int \psi I_{\overline{A}},$$

то (1) эквивалентно (2). $\triangleleft$

Если $\int \psi(\theta,y)\nu(dy)$ существует, то в качестве множества $A(t)$ для п. в. $[\nu]$ непрерывных $\psi(t,y)$ можно использовать множество

$$A(t) = \{y\colon |\psi(t,y)| \leqslant 2|\psi(\theta,y)|\},$$

как это делается, например, в следующем утверждении.

Следствие 1. *Пусть $\pi(x)$ — любая измеримая ограниченная функция $\mathcal{X}^n \to \mathbb{R}$, $f_\theta(\boldsymbol{x})$ непрерывна по θ для п. в. $[\mu^n]$ значений $\boldsymbol{x} \in \mathcal{X}^n$. Тогда функция*

$$\mathbf{E}_\theta \pi(X) = \int \pi(\boldsymbol{x}) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x})$$

непрерывна по θ.

Доказательство. Используем теорему 1 при $\mathcal{Y} = \mathcal{X}^n$, $y = \boldsymbol{x}$, $\nu = \mu^n$, $\psi(t, \boldsymbol{x}) = \pi(\boldsymbol{x}) f_t(\boldsymbol{x})$, $A(t) = \{\boldsymbol{x}\colon f_t(\boldsymbol{x}) \leqslant 2 f_\theta(\boldsymbol{x})\}$. Очевидно, что условия 1), 2) теоремы 1 выполнены. Так как для $\pi(\boldsymbol{x}) \equiv 1$ функция $\mathbf{E}_\theta \pi(X) \equiv 1$ непрерывна, то выполнено (см. (2))

$$\int\limits_{\boldsymbol{x} \notin A(t)} f_t(x) \mu^n(d\boldsymbol{x}) \to 0$$

при $t \to \theta$. Но по теореме 1 отсюда следует непрерывность $\mathbf{E}_\theta \pi(X)$ для любой ограниченной функции π. $\triangleleft$

Если речь вести лишь о достаточном условии для сходимости (1) в случае $\psi(t, y) \to \psi(\theta, y)$ п. в. при $t \to \theta$, то в качестве такового можно использовать равномерную сходимость интегралов в (1). Последнюю можно определить как существование конечной меры λ такой, что неравенство $\lambda(A) < \delta = \delta(\varepsilon)$ влечет за собой $\sup\limits_t \int\limits_A |\psi(t, y)| \nu(dy) < \varepsilon$ для заданного $\varepsilon > 0$.

Если существует интегрируемая мажоранта $\psi(y) = \sup\limits_t \psi(t, y)$, то такая мера λ всегда существует: достаточно положить $\lambda(A) = \int\limits_A \psi(y) \nu(dy)$.

2. Следствия условий (R). Докажем здесь лемму 26.1 и равномерную сходимость интеграла $I(\theta)$:

$$\sup_\theta \mathbf{E}_\theta(|l'(\mathrm{x}_1, \theta)|^2; |l'(\mathrm{x}_1, \theta)| > N) \to 0 \tag{3}$$

при $N \to \infty$ (именно такая равномерность имеется в виду в §§ 33, 34, 36, 37). Так как рассмотрения для одномерного и многомерного параметра практически не отличаются, то мы в этом и следующем пункте ограничимся изучением одномерного случая.

Теорема 2 (лемма 26.1). *Пусть выполнено условие* (R) *и $S = S(X)$ есть любая статистика, для которой $\mathbf{E}_\theta S^2 < c < \infty$ при $\theta \in \Theta$. Тогда в равенстве*

$$a_S(\theta) = \mathbf{E}_\theta(S) = \int S(\boldsymbol{x}) f_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x})$$

возможно дифференцирование под знаком интеграла:

$$a'_S(\theta) = \int S(\boldsymbol{x}) f'_\theta(\boldsymbol{x}) \mu^n(d\boldsymbol{x}) = \mathbf{E}_\theta S L'(X, \theta), \tag{4}$$

при этом функция $a'_S(\theta)$ непрерывна.

Доказательство. Заметим предварительно, что из (4) при $S(x) \equiv 1$ и $n = 1$ следует, что

$$\int f'_\theta(x)\mu(dx) = 0. \tag{5}$$

Так как в этом случае $L'(X,\theta) = \sum_{i=1}^{n} l'(\mathrm{x}_i,\theta)$ есть сумма независимых случайных величин с нулевым средним (см. (5)), то

$$\mathbf{D}_\theta L'(X,\theta) = \mathbf{E}_\theta(L'(X,\theta))^2 = n\mathbf{E}_\theta(l'(\mathrm{x}_1,\theta))^2 = nI(\theta). \tag{6}$$

Предположим теперь, что функция

$$I_n(\theta) = \mathbf{E}_\theta(L'(X,\theta))^2 = 4\int \left(\sqrt{f_\theta(x)}'\right)^2 \mu^n(dx)$$

непрерывна по θ (мы еще не можем пользоваться (6)) и докажем непрерывность интеграла в (4). Воспользуемся теоремой 1 при $\mathcal{Y} = \mathcal{X}^n$, $\nu = \mu^n$, $\psi(t,x) = \left(\sqrt{f_t(x)}'\right)^2$, $\tau = t - \theta$,

$$A(t) = A_1(\tau) \equiv$$
$$\equiv \left\{x : \sup_{v:|\theta-v|<|\tau|} \sqrt{f_v(x)} < 2\sqrt{f_\theta(x)}, \sup_{v:|\theta-v|<|\tau|} \left|\sqrt{f_v(x)}'\right| \leqslant 2\left|\sqrt{f_\theta(x)}'\right|\right\}. \tag{7}$$

Условия 1), 2) теоремы 1 при $\psi(x) = 2\psi(\theta,x)$ в силу непрерывности функций $\sqrt{f_\theta}$ и $\sqrt{f_\theta}'$ выполнены. Поэтому из сходимости $I_n(t)$ к $I_n(\theta)$ при $t \to \infty$ получаем (см. (2)), что при $t \to \theta$

$$\varepsilon(t) = \int\limits_{x \notin A_1(\tau)} \left(\sqrt{f_t(x)}'\right)^2 \mu^n(dx) \to 0. \tag{8}$$

Аналогично тому, как это делалось в следствии 1, мы получаем отсюда непрерывность $\int S(x)f'_\theta(x)\mu^n(dx)$. Чтобы убедиться в этом, надо воспользоваться теоремой 1 «в обратную сторону» при тех же множествах $A(t)$ и $\psi(t,x) = S(x)f'_t(x)$. Условия 1), 2) теоремы 1, очевидно, будут выполнены $\Big(\psi(x) = 2\big|S(x)f'_\theta(x)\big|$, $\int \psi(x)\mu^n(dx) \leqslant 4\mathbf{E}_\theta S^2 \int \left(\sqrt{f_\theta(x)}'\right)^2 \mu^n(dx)\Big)$. Выполнение (2) обеспечивается (8) и только что приведенным неравенством, в котором интегрирование надо вести по множеству $x \notin A_1(\tau)$.

Обратимся теперь непосредственно к доказательству (4). Заметим, что

$$\frac{1}{\tau}\left(\int Sf_{\theta+\tau}\mu^n - \int Sf_\theta\mu^n\right) = \int\int_0^1 Sf_{\theta+u\tau}\,du\mu^n = \int\int_0^1 2S\sqrt{f_{\theta+u\tau}}\left(\sqrt{f_{\theta+u\tau}}\right)'du\mu^n.$$

Воспользуемся снова теоремой 1 при $\mathcal{Y} = \mathbb{R}\times\mathcal{X}^n$, $y = (u,x)$, $\nu = \lambda\times\mu^n$ (λ — мера Лебега), $\psi(\tau,y) = S(x)f'_{\theta+u\tau}(x)$, $\tau \to 0$, $A(\tau) = A_1(\tau)$, где $A_1(\tau)$ определено

в (7). Опять из непрерывности $\sqrt{f_\theta(x)}$ и $\sqrt{f_\theta(x)}'$ следует выполнение условий 1), 2) теоремы 1

$$\psi(\tau, y) I_{A(\tau)}(x) \to S(x) f'_\theta(x) = \psi(0, y) \quad \text{при} \quad \tau \to 0,$$
$$\sup_\tau |\psi(\tau, y) I_{A(\tau)}(x) S(x)| \leqslant 4S(x) |f'_\theta(x)|,$$

где в силу неравенства Коши–Буняковского

$$\int 4S|f'_\theta|\mu^n \leqslant 4\left[\int S^2 f_\theta \mu^n \int \frac{(f'_\theta)^2}{f_\theta}\mu^n\right]^{1/2} < \infty.$$

Таким образом, для доказательства (4) нам надо проверить условие (2). Оно вытекает из неравенства Коши–Буняковского и соотношения (8):

$$\left|\int\limits_{x \notin A_1(\tau)} \int\limits_0^1 S\sqrt{f_{\theta+u\tau}}\left(\sqrt{f_{\theta+u\tau}}\right)' du\mu^n\right| \leqslant$$
$$\leqslant \left[\int\int\limits_0^1 S^2 f_{\theta+u\tau}\, du\mu^n\right]^{1/2} \left[\int\limits_{x \notin A_1(\tau)} \int\limits_0^1 \left(\sqrt{f_{\theta+u\tau}}'\right)^2 du\mu^n\right]^{1/2} \leqslant$$
$$\leqslant c^{1/2}\left[\int\limits_0^1 \varepsilon(\theta + u\tau)\, du\right]^{1/2} \to 0$$

при $\tau \to 0$.

Итак, мы доказали (4) в предположении, что $I_n(\theta)$ непрерывно. Но $I_n(\theta) = I(\theta)$ при $n = 1$, и это предположение выполнено в силу условий (R). Стало быть, (4) верно при $n = 1$ и, стало быть, верно (5). Но из (5) следует соотношение (6), означающее непрерывность $I_n(\theta)$. Теорема доказана.

Теорема 3. *Если множество Θ компактно, функция $\sqrt{f_\theta(x)}$ для $[\mu]$ п. в. значений x непрерывно дифференцируема по θ, то непрерывность $I(\theta)$ имеет место тогда и только тогда, когда выполнено* (3).

Теорема означает, что непрерывность $I(\theta)$ в условии (R) можно заменить условием (3).

Доказательство. Допустим, что $I(\theta)$ непрерывна, а (3) не выполнено. Тогда существует $\gamma > 0$ и последовательности $t \to \theta \in \Theta$, $N_t \to \infty$, такие, что

$$m(t) \equiv \mathbf{E}_t[|l'(x_1, t)|^2; l'(x_1, t) > N_t] > \gamma \tag{9}$$

при всех t из выбранной последовательности.

Воспользуемся теоремой 1 при $\mathcal{Y} = \mathcal{X}$, $\nu = \mu$, $\psi(t, x) = \left(\sqrt{f_t(x)}'\right)^2 = \frac{1}{4}(l'(x,t))^2 f_t(x)$, $A(t) = \{x: \left|\sqrt{f_t(x)}'\right| \leqslant 2\left|\sqrt{f_\theta(x)}'\right|\}$. В силу непрерывности

$\sqrt{f_\theta(x)}'$ условия 1), 2) теоремы 1 выполнены и, следовательно, из непрерывности $I(t)$ будет вытекать, что

$$m_1(t) = \int\limits_{x \notin A(t)} \left|\sqrt{f_t(x)}'\right|^2 \mu(dx) \to 0$$

при $t \to \theta$. Но $m(t) \leqslant m_1(t) + m_2(t)$, где

$$m_2(t) = \int\limits_{B(t) \cap A(t)} \left(\sqrt{f_t}'\right)^2 \mu, \quad B(t) = \left\{x : 2\left|\sqrt{f_t(x)}'\right| > N_t \sqrt{f_t(x)}\right\}.$$

Из вида множества $A(t)$ следует, что

$$m_2(t) \leqslant 4 \int\limits_{B(t)} \left|\sqrt{f_\theta}'\right|^2 \mu.$$

Используя опять сходимость $\left(\sqrt{f_t(x)}\right)' \to \left(\sqrt{f_\theta(x)}\right)'$, $\sqrt{f_t(x)} \to \sqrt{f_\theta(x)}$ при $t \to \theta$, мы получаем, что $B(t)$ сходится к множеству μ-меры нуль. Это значит, что $\mu(B(t)) \to 0$, $m_2(t) \to 0$, $m(t) \to 0$ при $t \to \infty$. Мы получили противоречие с (9). Соотношение 3 доказано.

Путь теперь выполнено (3). В силу теоремы 1 для доказательства непрерывности $I(t)$ нам достаточно убедиться, что при том же, что и выше, множестве $A(t)$ выполняется $m_1(t) \to 0$ при $t \to \infty$. Но

$$m_1(t) \leqslant \int\limits_{|l'|>N} |l'|^2 f_t \mu + N^2 \int\limits_{x \notin A(t)} f_t \mu,$$

где первый интеграл выбором N может быть сделан в силу (3) сколь угодно малым. Для оценки второго интеграла заметим, что $\mu(A(t)) \to 0$ и что при $C(t) = \{x: f_t(x) \leqslant 2f_\theta(x)\}$ выполняется $\int\limits_{x \notin C(t)} f_t \mu \to 0$ при $t \to \theta$ (см. доказательство следствия 1). Поэтому

$$\int\limits_{x \notin A(t)} f_t \mu \leqslant 2 \int\limits_{x \notin A(t)} f_\theta \mu + \int\limits_{x \notin C(t)} f_t \mu \to 0$$

при $t \to \theta$. $\triangleleft$

3. Следствия условий ($\mathbf{RR_u}$). Условия (RR_u) были введены в § 34.

Теорема 4. *Если выполнены условия* (RR_u), *то* $\int f''_\theta(x)\mu(dx) = 0$.

Вместе с теоремой 2 это обеспечивает выполнение нужных нам в § 33, 34 условий (RR), 3.

Доказательство. В силу теоремы 2 при всех $\theta \in \Theta$

$$\int f'_\theta(x)\mu(dx) = 0$$

и нам достаточно доказать, что при $t \to 0$

$$J(t) \equiv \frac{1}{t-\theta}\left[\int f_t'\mu - \int f_\theta'\mu\right] \to \int f_\theta''\mu.$$

Заметим, что $\dfrac{1}{t-\theta}(f_t' - f_\theta') = \varphi_t f_t + \dfrac{f_\theta'}{f_\theta}\dfrac{f_t - f_\theta}{t-\theta}$, где $\varphi_t = \dfrac{1}{t-\theta}\left(\dfrac{f_t'}{f_t} - \dfrac{f_\theta'}{f_\theta}\right)$. Пользуясь этим равенством, мы можем $J(t)$ представить в виде суммы четырех слагаемых $J(t) = J_1 + J_2 + J_3 + J_4$, где

$$J_1 = \int \varphi_t f_\theta \mu, \quad J_2 = \int\limits_{l \leqslant N} \varphi_t(f_t - f_\theta)\mu,$$

$$J_3 = \int\limits_{l \geqslant N} \varphi_t(f_t - f_\theta)\mu, \quad J_4 = \int \frac{f_\theta'}{f_\theta}\frac{f_t - f_\theta}{t-\theta}\mu,$$

$l = l(x)$ есть мажоранта для $l''(x,t)$ в условиях (RR_u). В силу теоремы 2 при $n = 1$, $S(x) = l'(x,\theta)$ получаем

$$J_4 = \frac{1}{t-\theta}(\mathbf{E}_t l'(\mathrm{x}_1,\theta) - \mathbf{E}_\theta l'(\mathrm{x}_1,\theta)) \to \mathbf{E}_\theta(l'(\mathrm{x}_1,\theta))^2 = I(\theta). \tag{10}$$

Далее,

$$|\varphi_t| < l \tag{11}$$

и, стало быть, по теореме Лебега

$$\lim_{t\to\theta} J_1 = \int \lim_{t\to\theta} \varphi_t f_\theta \mu = \int l'' f_\theta \mu = \int f_\theta''\mu - I(\theta). \tag{12}$$

Используя опять (11), получаем в силу условий (RR)

$$|J_3| \leqslant \left|\int\limits_{l\geqslant N} l f_t \mu\right| + \left|\int\limits_{l\geqslant N} l f_\theta \mu\right| \to 0$$

при $N \to \infty$. Наконец, в силу неравенства Коши–Буняковского

$$|J_2| \leqslant N\int |f_t - f_\theta|\mu \leqslant N\int\int\limits_\theta^t |f_u'|\,du\mu \leqslant N\int\limits_\theta^t \sqrt{I(u)}\,du \to 0 \tag{13}$$

при $t \to \theta$. Сопоставляя (10)–(13), мы получим, что $0 \equiv J(t) \to \int f_\theta''\mu$. $\triangleleft$

Приложение VI

Неравенства для распределения отношения правдоподобия в многомерном случае

В этом приложении будет доказана следующая теорема (теорема 36.2; обозначения см. в § 31, 33, 36).

Теорема 1. *Пусть выполнены следующие условия:*

$$\inf_u \frac{r(u)}{u^2} \geqslant g(\theta) > 0, \tag{1}$$

$$\mathbf{E}_\theta l'(\mathrm{x}_1, \theta) = 0, \tag{2}$$

$$\gamma = \sup_\theta \mathbf{E}_\theta |l'(\mathrm{x}_1, \theta)|^s < \infty \tag{3}$$

при некотором $s > k$. Тогда при любых $z, n \geqslant 1$

$$\mathbf{P}_\theta \left(\sup_{|v| \geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z \right) \leqslant c\gamma (e^{-z/2} + e^{-z}) e^{-\beta g(\theta) r^2},$$

где $\beta > 0$ зависит лишь от k и s, $c < \infty$ зависит от k, s и $g(\theta)$ и может быть выбрана не зависящей от θ, если $g(\theta) > g > 0$ при всех θ.

Нам понадобятся несколько вспомогательных утверждений *). Через c_s, $c_{k,s}$ мы будем обозначать различные постоянные, зависящие лишь от своих индексов.

Лемма 1. *Пусть ξ_k, $k = 1, 2, \ldots$, независимы и одинаково распределены, $\mathbf{E}\xi_1 = 0$, $\mathbf{E}|\xi_1|^s < \gamma < \infty$, $s \geqslant 2$. Тогда*

$$\mathbf{E}\left|\sum_{k=1}^n \xi_k\right|^s \leqslant c_s \gamma n^{s/2}.$$

Доказательство. Для упрощения рассуждений ограничимся рассмотрением случая, когда $s = 2m$ есть целое четное число **). В этом случае достаточно рассмотреть скалярные величины ξ_k. Имеем

$$\mathbf{E}\left|\sum_{k=1}^n \xi_k\right|^s = \sum_{k_1, \ldots, k_n} \mathbf{E}\xi_1^{k_1} \ldots \mathbf{E}\xi_n^{k_n}, \tag{4}$$

*) Другой вариант доказательства теоремы 1 предложен в испанском издании этой книги [19].

**) Доказательство в общем случае см., например, в [49, с. 255].

где суммирование ведется по всем целым $k_1, \ldots, k_n$ таким, что $\sum_i k_i = s$, $k_i \neq 1$ ($k_i = 1$ исключаются, так как $\mathbf{E}\xi_k = 0$). По неравенству Гёльдера

$$|\mathbf{E}\xi_j^{k_j}| \leqslant (\mathbf{E}|\xi_j|^s)^{k_j/s} = \gamma^{k_j/s}$$

и, стало быть,

$$\prod_{j=1}^{n} |\mathbf{E}\xi_j^{k_j}| \leqslant \prod_{j=1}^{n} \gamma^{k_j/s} = \gamma.$$

Нам осталось оценить $\sum_{k_1,\ldots,k_n} 1$. Обозначим через $(l_1, \ldots, l_p)$ ненулевые элементы ($l_i \geqslant 2$) набора $(k_1, \ldots, k_n)$ $\left(\sum_{j=1}^{p} l_j = s\right)$. Тогда оцениваемая сумма будет равна $\sum_{(l_1,\ldots,l_p)} A_p$, где A_p есть число расположений элементов $l_1, \ldots, l_p$ по n местам. Очевидно, что $A_p \leqslant n(n-1)\ldots(n-p+1)$. Самое большое возможное значение p равно $m = s/2$ (оно соответствует набору $(2, 2, \ldots, 2)$), так что $A_p \leqslant A_m \leqslant n^m$. Но число различных наборов $(l_1, \ldots, l_p)$ зависит только от s. Следовательно, оцениваемая сумма не превосходит $c_s n^m$. $\triangleleft$

Положим $p(u) = Z^{1/s}(u)$.

Лемма 2. *При выполнении условий* (2), (3)

$$\mathbf{E}_\theta |p'(u)|^s \leqslant c_s \gamma n^{s/2}.$$

Доказательство. Имеем

$$\mathbf{E}_\theta |p'(u)|^s = \mathbf{E}_\theta \left|\frac{1}{s} L'(X, \theta+u) Z^{1/s}(u)\right|^s =$$

$$= s^{-s} \mathbf{E}_\theta |L'(X, \theta+u)|^s Z(u) = s^{-s} \mathbf{E}_{\theta+u} |L'(X, \theta+u)|^s.$$

Остается воспользоваться леммой 1, применив ее к случайным величинам $\xi_k = l'(\mathrm{x}_k, \theta+u)$. $\triangleleft$

Лемма 3. *При выполнении условий теоремы* 1

$$\mathbf{E}_\theta |p(u+v) - p(u)|^s \leqslant |v|^s c_s \gamma n^{s/2},$$

где c_s *то же, что и в лемме* 2.

Доказательство. В силу неравенства Гёльдера и леммы 2

$$\mathbf{E}_\theta |p(u+v) - p(u)|^s = \mathbf{E}_\theta \left| \int_0^{|v|} \left(p'\left(u + \frac{tv}{|v|}\right), \frac{v}{|v|} \right) dt \right|^s =$$

$$= |v|^s \mathbf{E}_\theta \left| \int_0^1 \left(p'(u+hv), \frac{v}{|v|} \right) dh \right|^s \leqslant |v|^s \int_0^1 \mathbf{E}_\theta |p'(u+hv)|^s \, dh \leqslant |v|^s c_s \gamma n^{s/2}. \quad \triangleleft$$

Обозначим через $K_{u,\Delta}$ куб в $\mathbb{R}^k$ со стороной длины Δ и с вершиной в точке $u = (u_1, \ldots, u_k)$:

$$K_{u,\Delta} = \{v \in \mathbb{R}^k : u_i \leqslant v_i \leqslant u_i + \Delta, \quad i = 1, \ldots, k\}$$

и положим $\Delta = e^{-|u|^2 g(\theta)/(4k)}$.

Лемма 4. *При выполнении условий теоремы* 1

$$\mathbf{P}_\theta \left(\sup_{v \in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right) \geqslant e^z \right) \leqslant c\gamma \Delta^k (e^{-z/2} + e^{-z}) e^{-|u|^2 \beta g(\theta)},$$

где постоянные $c < \infty$ и $\beta > 0$ зависят только от k и s.

Такая же оценка будет верна для любого куба со стороной длины Δ, содержащего точку u.

Доказательство. Точку $v \in K_{u,\Delta}$ представим в виде $v = u + t\Delta$, где $t \in K_{0,1}$. Для координат t_i вектора t воспользуемся двоичным разложением

$$t_i = \sum_{r=1}^{\infty} \frac{\delta_{ir}}{2^r},$$

где все δ_{ir} равны либо единице, либо нулю. Положим

$$t_i^m = \sum_{r=1}^{m} \frac{\delta_{ir}}{2^r}, \quad t^m = (t_1^m, \ldots, t_k^m), \quad t^0 = 0, \tag{5}$$

так что t^m есть двоичное приближение t: $|t - t^m| < 2^{-m}\sqrt{k}$.

Положим

$$\varphi(t) = p\left(\frac{v}{\sqrt{n}}\right) = p\left(\frac{u + t\Delta}{\sqrt{n}}\right).$$

Тогда

$$\varphi(t) = \varphi(0) + \sum_{m=1}^{\infty} (\varphi(t^m) - \varphi(t^{m-1})).$$

Точки $t_{(1)}^m$ и $t_{(2)}^m$ назовем соседними, если их представления (5) отличаются лишь одним из чисел $\delta_{1m}, \ldots, \delta_{km}$. Ясно, что если для любых двух соседних точек

$$|\varphi(t_{(1)}^m) - \varphi(t_{(2)}^m)| < \frac{c_m}{\sqrt{k}}, \tag{6}$$

то $|\varphi(t^m) - \varphi(t^{m-1})| < c_m$. Стало быть, если (6) выполнено для всех соседних точек при всех m и при $c_m = a(1-q)q^m$, $q < 1$, и, кроме того,

$$\varphi(0) < a(1-q), \tag{7}$$

то

$$\varphi(t) < \sum_{m=0}^{\infty} a(1-q)q^m = a,$$

$$\sup_{t \in K_{0,1}} \varphi(t) < a. \tag{8}$$

Обратимся теперь непосредственно к утверждению леммы. Нам надо оценить

$$\mathbf{P}_\theta\left(\sup_{v\in K_{u,\Delta}} Z\left(\frac{v}{\sqrt{n}}\right)\geqslant e^z\right)=\mathbf{P}_\theta\left(\sup_{t\in K_{0,1}}\varphi(t)\geqslant a\right) \tag{9}$$

при $a=e^{z/s}$. Неравенство, стоящее под знаком вероятности, означает нарушение (8) и, стало быть, влечет за собой нарушение одного из неравенств (6), (7). Следовательно, вероятность (9) оценивается суммой вероятностей

$$\mathbf{P}_\theta(\varphi(0)\geqslant a(1-q))+\sum_{m=1}^{\infty}\sum\mathbf{P}_\theta\left(\left|\varphi\left(t_{(1)}^m\right)-\varphi\left(t_{(2)}^m\right)\right|\geqslant\frac{a(1-q)q^m}{\sqrt{k}}\right), \tag{10}$$

где последняя сумма образована $k(2^m)^k$ слагаемыми, соответствующими всем возможным парам соседних точек $t_{(1)}^m$ и $t_{(2)}^m$. Так как $\left|t_{(1)}^m-t_{(2)}^m\right|=2^{-m}$, то для каждой такой пары в силу леммы 3 и неравенства Чебышева

$$\mathbf{P}_\theta\left(\left|\varphi\left(t_{(1)}^m\right)-\varphi\left(t_{(2)}^m\right)\right|\geqslant\frac{a(1-q)q^m}{\sqrt{k}}\right)\leqslant\left(\frac{\Delta}{\sqrt{n}\,2^m}\right)^s c_s\gamma n^{s/2}\left(\frac{a(1-q)q^m}{\sqrt{k}}\right)^{-s}.$$

Следовательно, двойная сумма в (10) будет оцениваться значением

$$\Delta^s c_s\gamma[a(1-q)]^{-s}k^{1-s/2}\sum_{m=1}^{\infty}2^{-m(s-k)}q^{-ms}. \tag{11}$$

Выберем $q=2^{(k-s)/2s}<1$. Тогда ряд в (11) будет сходиться и все выражение (11) будет иметь вид $c_{k,s}\gamma\Delta^s a^{-s}$. Первое слагаемое в (10) оценивается в силу теоремы 36.1 и неравенства Чебышева:

$$\mathbf{P}_\theta(\varphi(0)\geqslant a(1-q))=\mathbf{P}_\theta\left(Z^{1/2}\left(\frac{u}{\sqrt{n}}\right)\geqslant(a(1-q))^{s/2}\right)\leqslant(a(1-q))^{-s/2}e^{-|u|^2g(\theta)/2}.$$

Полагая $\Delta=e^{-|u|^2g(\theta)/(4k)}$, учитывая, что $a^{-s}=e^{-z}$, и считая, не ограничивая общности, что $s\leqslant 2k$, мы получим для (9) оценку

$$c_{k,s}\gamma\Delta^k e^{-|u|^2g(\theta)(s-k)/(4k)}(e^{-z}+e^{-z/2}).$$

При $\beta=(s-k)/4k$ получим первое утверждение леммы.

Справедливость второго утверждения очевидна, так как в доказательстве вместо $\varphi(0)$ мы смогли бы взять значение $\varphi(t_0)$ в любой фиксированной точке $t_0\in K_{0,1}$ (первая сумма в (10) соответствует значению в точке, а вторая — полному возможному изменению $\varphi(t)$ в $K_{0,1}$). Лемма доказана.

Доказательство теоремы 1. Покроем все пространство $\mathbb{R}^k$ системой кубиков $K_{u,\Delta}$, у которых координаты точек u кратны Δ. Число таких кубиков, пересекающихся со слоем $S_r=\{v\in\mathbb{R}^k: r\leqslant|v|\leqslant r+1\}$, ограничено количеством

$c_k r^{k-1}\Delta^{-k}$. Следовательно,

$$\mathbf{P}_\theta\left(\sup_{v\in S_r} Z\left(\frac{v}{\sqrt{n}}\right)\geqslant e^z\right)\leqslant c_k r^{k-1} c_{k,s}\gamma(e^{-z/2}+e^{-z})e^{-r^2\beta g(\theta)},$$

$$\mathbf{P}_\theta\left(\sup_{|v|\geqslant r} Z\left(\frac{v}{\sqrt{n}}\right)> e^z\right)\leqslant c_k c_{k,s}\gamma(e^{-z/2}+e^{-z})\sum_{j=0}^{\infty}(r+j)^{k-1}e^{-(r+j)^2\beta g(\theta)}.$$

Так как $\sup\limits_{r,j}(r+j)^{k-1}e^{-\beta g(\theta)(r+j)^2/2}$ и $\sum\limits_j e^{-\beta g(\theta)[(r+j)^2-r^2]/2} < \sum\limits_j e^{-\beta g(\theta)j^2/2}$ ограничены постоянной $c(\beta g(\theta))$, зависящей от $\beta g(\theta)$, то полученное выражение не превосходит $c_{k,s}\gamma c(\beta g(\theta))e^{-r^2\beta g(\theta)/2}(e^{-z/2}+e^{-z})$, где $c(\beta g(\theta)) < c(\beta g) < \infty$, если $g(\theta)\geqslant g > 0$ при всех θ. $\triangleleft$

Приложение VII

Задача о разладке. Вспомогательные предложения и теоремы об асимптотически оптимальных моментах остановки

Везде ниже мы будем использовать обозначения § 72, как правило, без напоминаний.

1. Теорема о предельном распределении оценки максимального правдоподобия. В этом пункте мы получим утверждение, обобщающее теорему 72.1 о предельном $\mathbf{P}_t$-распределении разности $\widehat{\theta}^* - t$ при $t \to \infty$, где $\widehat{\theta}^*$ — оценка максимального правдоподобия. Обозначим через $n(S)$ момент *последнего* достижения неотрицательного максимума S, так что $S_{n(S)} \equiv S$, если $S \geqslant 0$, и через $n(Z)$ — момент *первого* достижения $Z \geqslant 0$, $Z_{n(Z)} \equiv Z$. Если $S < 0$, $Z < 0$, то полагаем $n(S) = 0$, $n(Z) = 0$ соответственно. Определим на одном вероятностном пространстве две независимые пары $(S, n(S))$ и $(Z, n(Z))$ и случайную величину

$$D = \begin{cases} -n(S) < 0 & \text{на множестве} \quad Z \leqslant S,\ S \geqslant 0; \\ 0 & \text{на множестве} \quad S < 0,\ Z \leqslant 0; \\ n(Z) > 0 & \text{на множестве} \quad Z > S,\ Z > 0. \end{cases}$$

Если $\mathbf{P}(S_k = 0) = 0$ при всех k, то моменты первого и последнего достижения максима S совпадают с вероятностью 1 (см. [12]). То же справедливо для Z, если $\mathbf{P}(Z_k = 0) = 0$ при всех k. В этом случае распределение D совпадает с распределением

$$D_0 = \begin{cases} -n(S) < 0 & \text{на множестве} \quad Z^+ \leqslant S^+,\ S^+ > 0; \\ 0 & \text{на множестве} \quad S^+ = Z^+ = 0; \\ n(Z) > 0 & \text{на множестве} \quad Z^+ > S^+,\ Z^+ > 0. \end{cases}$$

Совместные распределения пар $(S^+, n(S))$ и $(Z^+, n(Z))$ (точнее, двойные преобразования $\mathbf{E}z^{n(S)}e^{\lambda S^+}$ и $\mathbf{E}z^{n(Z)}e^{\lambda Z^+}$) могут быть найдены в явном виде в терминах компонент факторизации функций $1 - z\varphi(\lambda)$ и $1 - z\psi(\lambda)$ соответственно (см. [12]). Если $\varpi_{z+}(v)$ — положительная компонента канонической факторизации функции $1 - z\varphi(iv)$, т. е. функция, аналитическая в верхней полуплоскости, непрерывная, включая границу $\operatorname{Im} v = 0$, и обладающая свойствами

$$\inf_{\operatorname{Im} v > 0} |\varpi_{z+}(v)| > 0, \quad \varpi_{z+}(i\infty) = 1,$$

(подробнее см. § 14, 15 в [12]; гл. 11 в [17]), то

$$\mathbf{E}z^{n(S)}e^{ivS^+} = \frac{\varpi_{1+}(0)}{\varpi_{z+}(v)}.$$

В широких предположениях справедливо асимптотическое представление

$$\mathbf{P}(n(S)=k) \sim \frac{c\varphi^k}{k^{3/2}}, \quad c=\text{const}>0$$

(см. [12, § 21]). Аналогичные соотношения верны для $\mathbf{E}z^{n(Z)}e^{ivZ^+}$ и $\mathbf{P}(n(Z)=k)$.

Теорема 1. *$\mathbf{P}_t$-распределение $\widehat{\theta}^* - t$ сходится к распределению D:*

$$\mathbf{P}_t(\theta^* - t = k) \to \mathbf{P}(D=k)$$

при $t \to \infty$.

В силу оценок (72.6) эта сходимость будет иметь место вместе со сходимостью моментов любого порядка.

Доказательство теоремы следует доказательству теоремы 72.1 и отличается лишь большей детализацией. Поэтому мы предоставляем его читателю.

2. Оценки распределений, связанных со случайными величинами E_i. В этом пункте мы получим оценки для «хвостов» распределений случайных величин $\dfrac{E_i'}{E_i}$ и $\ln E_i$, $i=1,2$, где E_i и E_i' определены в теореме 72.4. Эти оценки использованы в доказательствах теоремы 72.4 и теоремы 2 этого приложения.

Так как величины ξ_i и ζ_i, определенные в п. 1 § 72, обладают одними и теми же свойствами ($\mathbf{E}\xi_i < 0$, $\mathbf{E}\zeta_i < 0$; $\varphi(1) = \mathbf{E}e^{\xi_i} = 1$, $\psi(1) = \mathbf{E}e^{\zeta_i} = 1$), то можно ограничиться рассмотрением лишь случайных величин, ассоциированных с ξ_i (т. е. с E_1). При этом мы несколько расширим наши допущения и будем считать, что ξ_i — произвольные независимые случайные величины, распределенные как ξ, $\mathbf{E}\xi < 0$,

$$\varphi(\lambda) = \mathbf{E}e^{\lambda\xi} < \infty \quad \text{при некотором} \quad \lambda > 0.$$

Как и прежде, обозначим

$$E_{1,s} = \sum_{k=1}^{s} e^{S_k}, \quad E_{1,s}' = \sum_{k=1}^{s} k e^{S_k},$$

так что $E_1 = E_{1,\infty}$, $E_1' = E_{1,\infty}'$. Обозначим $b^{-1} = 3 + \sum\limits_{k=1}^{\infty} k^{-1}\ln^{-2} k$.

Лемма 1. *При всех достаточно больших x справедливо неравенство*

$$\mathbf{P}\left(\frac{E_1'}{E_1} > x\right) \leqslant \frac{2(bx)^\lambda}{1-\varphi} \exp\left(\frac{\sqrt{bx}}{\ln\sqrt{bx}} \ln\varphi\right),$$

где $\lambda > 0$ есть значение аргумента, при котором достигается $\min\varphi(\lambda) = \varphi < 1$.

Такое же неравенство справедливо при всех $s > 3$ для $\mathbf{P}\left(\dfrac{E_{1,s}'}{E_{1,s}} > x\right)$.

Таким образом, случайная величина $\dfrac{E_1'}{E_1}$ имеет моменты любого порядка. Из доказательства леммы будет видно, что полученная оценка может быть улучшена.

Однако истинная асимптотика $\mathbf{P}\left(\dfrac{E_1'}{E_1} > x\right)$ при $x \to \infty$ остается неясной. Возможно, что она имеет тот же вид, что и $\mathbf{P}(n(S) > x)$ (см. выше).

Доказательство. Обозначим через η значение k, при котором достигается максимум последовательности S_k^*, $S_k^* = S_k + 2\ln k + 2\ln\ln k$, $k \geqslant 3$; $S_0^* = 0$, $S_1^* = S_1$, $S_2^* = S_2$. Обозначим $A_{n,dv} = \{\eta = n, S_\eta^* \in dv\}$, $B = \left\{\dfrac{E_1'}{E_1} > x\right\}$. Тогда

$$\mathbf{P}(B) = \sum_{n=0}^{\infty} \int_0^{\infty} \mathbf{P}(BA_{ndv}). \tag{1}$$

На событии $A_{n,dv}$ выполняется

$$E_1' = \sum_{k=1}^{\infty} ke^{S_k} \leqslant 3e^v + \sum_{k=3}^{\infty} ke^{v-2\ln k-2\ln\ln k} \leqslant e^v\left(3 + \sum_{k=3}^{\infty} k^{-1}\ln^{-2} k\right) = b^{-1}e^v;$$

$$E_1 \leqslant e^{v-2\ln n-2\ln\ln n} = e^v(n\ln n)^{-2}; \quad \frac{E_1'}{E_1} \leqslant b^{-1}(n\ln n)^2.$$

Поэтому произведение $BA_{n,dv}$ не пусто лишь при n таких, что $n\ln n \geqslant \sqrt{bx}$,

$$BA_{n,dv} \subset A_{n,dv}\, I\left\{n \geqslant \frac{\sqrt{bx}}{\ln\sqrt{bx}}\right\}.$$

Следовательно, в силу (1)

$$\mathbf{P}(B) \leqslant \mathbf{P}\left(\eta \geqslant \frac{\sqrt{bx}}{\ln\sqrt{bx}}\right). \tag{2}$$

Но $\{\eta = k\} \subset \{S_k + 2\ln k + 2\ln\ln k \geqslant 0\}$,

$$\mathbf{P}(\eta = k) \leqslant \mathbf{P}(S_k \geqslant -2\ln k - 2\ln\ln k) \leqslant \varphi^k(\lambda)e^{\lambda(2\ln k+2\ln\ln k)} = \varphi^k(k\ln k)^{2\lambda},$$

где под $\lambda > 0$ мы понимаем значение, при котором достигается $\min\varphi(\lambda) = \varphi < 1$. Следовательно, при достаточно больших k

$$\mathbf{P}(\eta \geqslant k) \leqslant \frac{2\varphi^k}{1-\varphi}(k\ln k)^{2\lambda},$$

и в силу (2)

$$\mathbf{P}(B) \leqslant \frac{2}{1-\varphi}\exp\left(\frac{\sqrt{bx}}{\ln\sqrt{bx}}\ln\varphi\right)\left(\frac{\sqrt{bx}}{\ln\sqrt{bx}}\ln\frac{\sqrt{bx}}{\ln\sqrt{bx}}\right)^{2\lambda}.$$

Очевидно, что при $bx \geqslant e$ последний множитель не превосходит $(\sqrt{bx})^{2\lambda}$. Это доказывает первое утверждение теоремы.

Второе утверждение доказывается точно так же. Надо лишь вместо η рассмотреть значение $\eta_{(s)} = \min(s, \eta)$, при котором достигается максимум S_k^*, $k \leqslant s$. Теорема доказана. $\triangleleft$

В следующем утверждении изучается асимптотика $\mathbf{P}(\ln E_1 > x)$ при $x \to \infty$.

Лемма 2. *Пусть* $s = \sup\{\lambda : \varphi(\lambda) \leqslant 1\}$. *Тогда*

$$\ln \mathbf{P}(\ln E_1 > x) = -sx(1 + o(1)). \tag{3}$$

Если

$$\varphi(s) = 1, \quad \varphi'(s) < \infty, \tag{4}$$

то

$$\mathbf{P}(\ln E_1 > x) = c_1 e^{-sx}(1 + o(1)), \quad o < c_1 < \infty. \tag{5}$$

Доказательство. Рассмотрим последовательность

$$W_n = \ln(e^{\xi_n} + e^{\xi_n + \xi_{n-1}} + \ldots + e^{\xi_n + \ldots + \xi_1}). \tag{6}$$

Очевидно, что она удовлетворяет рекурсивным соотношениям

$$\begin{aligned} e^{W_n} &= e^{\xi_n} + e^{\xi_n} e^{W_{n-1}}, \\ W_n &= W_{n-1} + \xi_n + \ln(1 + e^{-W_{n-1}}) \end{aligned} \tag{7}$$

и является цепью Маркова. Распределение W_n при любом начальном значении W_1 (не обязательно $W_1 = \xi_1$) сходится к распределению $W = \ln E_1$, которое будет стационарным распределением цепи, так что цепь $\{W_n\}$ является эргодической. Приращение этой цепи $\xi(x) = W_n - W_{n-1}$ при условии $W_{n-1} = x$ имеет вид $\xi(x) = \xi_n + \ln(1 + e^{-x})$. Это означает, что цепь $\{W_n\}$ является асимптотически однородной (в пространстве; см. [20, 21]) и к ней применима теорема 3 из [21], в силу которой выполнено (3). Чтобы применить теорему 4 из [21] о точной асимптотике $\mathbf{P}(W > x)$ (из которой будет следовать (5) при $0 \leqslant c_1 < \infty$), наряду с условиями (4) требуется также выполнение условий

$$\int_{-\infty}^{\infty} e^{st} |\mathbf{P}(\xi + \ln(1 + e^{-y}) > t) - \mathbf{P}(\xi > t)| \, dt \leqslant l(y), \tag{8}$$

где $l(y)$ — правильно меняющаяся функция такая, что

$$\int_0^{\infty} l(y) \, dy < \infty. \tag{9}$$

Разность под знаком интеграла (8) неотрицательна, и поэтому нам достаточно оценить разность интегралов

$$\int_{-\infty}^{\infty} e^{st} \mathbf{P}(\xi > t) \, dt = \frac{1}{s} \int_{-\infty}^{\infty} e^{st} \, d\mathbf{P}(\xi < t) = \frac{\varphi(s)}{s}$$

и

$$\int\limits_{-\infty}^{\infty} e^{st}\mathbf{P}(\xi + h(y) > t)\,dt = \frac{\varphi(s)e^{sh(y)}}{s}, \quad h(y) = \ln(1+e^{-y}),$$

равную

$$\frac{1}{s}\varphi(s)(1-e^{sh(y)}) \leqslant \frac{1}{s}\varphi(s)(1-(1+e^{-y})^{s}) \leqslant \varphi(s)e^{-y}.$$

Это гарантирует выполнение условий (8), (9). Положительность c_1 в (5) следует из неравенства $\ln E_1 \geqslant S$ и асимптотических представлений в лемме 72.1. Лемма доказана. $\triangleleft$

Доказательство теоремы 72.7. Чтобы разгрузить само доказательство от технических деталей, мы начнем со вспомогательных утверждений.

Лемма 3. *Если* $p > 0$, $p \to 0$, *то*

$$p\sum_{k=1}^{\infty}(1-p)^{k-1}\ln k = -\ln p + c + o(1),$$

где $|c| < \infty$, *с от* p *не зависит.*

Доказательство. Положим $\Delta = -\ln(1-p) = p + O(p^2)$ и воспользуемся представлением

$$\int\limits_{k}^{k+1} e^{-x\Delta}\ln x\,dx = e^{-k\Delta}\ln k \int\limits_{k}^{k+1}(1+O(\Delta))\left(1+O\left(\frac{1}{x}\right)\right)dx =$$

$$= e^{-k\Delta}\ln k\left(1+O(\Delta)+O\left(\frac{1}{k}\right)\right).$$

Из него следует, что

$$p\sum_{k=1}^{\infty}(1-p)^{k-1}\ln k = p(1-p)^{-1}\int\limits_{1}^{\infty} e^{-\Delta x}\ln x\left(1+O(\Delta)+O\left(\frac{1}{x}\right)\right)dx =$$

$$= \frac{p}{\Delta}(1-p)^{-1}\int\limits_{\Delta}^{\infty} e^{-y}\ln\frac{y}{\Delta}\left(1+O(\Delta)+O\left(\frac{\Delta}{y}\right)\right)dy =$$

$$= (1+o(1))\left(\int\limits_{0}^{\infty} e^{-y}\ln y\,dy - \ln\Delta + o(1)\right).$$

Лемма доказана. $\triangleleft$

Рассмотрим теперь последовательность геометрических априорных распределений Q (72.31) при $p \to 0$. Рассмотрим байесовский момент остановки $\tau(A)$, определенный в теореме 72.8, и для заданного $\varepsilon > 0$ выберем A так, чтобы $\mathbf{P}(\tau(A) < \theta) = \mathbf{P}_Q(\tau(A) < \theta) = \varepsilon$. Это будет означать, что $\tau(A) \in K_{(\varepsilon,Q)}$ (см. п. 3 из § 72). Как уже отмечалось в п. 3 из § 72, при малых $q(n) = p(1-p)^{n-1}$ выполняется $A \approx 1-\varepsilon$.

Нам понадобится асимптотика $\mathbf{E}(\tau(A)-\theta)^*$ при $p\to 0$.

Лемма 4.

$$E(\tau(A)-\theta)^+ = \frac{(1-\varepsilon)(-\ln p - 1 + V)}{\rho} + u_p + o(1), \tag{10}$$

где

$$V = \ln\frac{A}{1-A} - \mathbf{E}\ln(1+E_1+E_2), \quad \sup_p u_p < c < \infty.$$

Если распределение ζ_i нерешетчато, то $u_p \to (1-\varepsilon)\mathbf{E}\chi < \infty$, где χ — перескок через бесконечно удаленный барьер в блуждании $\{-Z_k\}$.

Доказательство. Имеем

$$E(\tau(A)-\theta)^+ = p\sum_{t=1}^{\infty}(1-p)^{t-1}\mathbf{E}_t(\tau(A)-t)^+ =$$

$$= p\sum(1-p)^{t-1}\mathbf{P}_t(\tau(A)\geqslant t)\mathbf{E}_t\left(\tau(A)-\frac{t}{\tau(A)}\geqslant t\right). \tag{11}$$

Наша ближайшая цель — найти асимптотику $\mathbf{E}_t(\tau(A)-t/\tau(A)\geqslant t)$ при $t\to\infty$. Рассмотрим более внимательно момент остановки

$$\tau(A) = \min\left\{n\geqslant 1\colon \frac{\sum_{k=1}^{n}(1-p)^{k-1}R(k)}{(1-p)^n R(n+1)} \geqslant \frac{A}{p(1-A)}\right\} = \min\{n\geqslant 1\colon M_n\geqslant e^h\}, \tag{12}$$

где

$$h = \ln\frac{A}{p(1-A)}, \tag{13}$$

$$M_n = (1-p)r_n^{-1} + (1-p)^2(r_n r_{n-1})^{-1} + \ldots + (1-p)^n(r_n\ldots r_1)^{-1}.$$

В нашем случае

$$r_k^{-1} = e^{\xi_k} \quad \text{при} \quad k<t,$$
$$r_k^{-1} = e^{-\zeta_k} \quad \text{при} \quad k\geqslant t.$$

Поэтому, полагая $\Delta = -\ln(1-p) = p + O(p^2)$, получим

$$M_n = e^{-\zeta_n-\Delta} + e^{-\zeta_n-\zeta_{n-1}-2\Delta} + \ldots + e^{-\zeta_n-\ldots-\zeta_t-(n-t+1)\Delta} +$$
$$+ e^{-\zeta_n-\ldots-\zeta_t-(n-t+1)\Delta+\xi_{t-1}-\Delta} + \ldots + e^{-\zeta_n-\ldots-\zeta_t-(n-t+1)\Delta+\xi_{t-1}+\ldots+\xi_1-(t-1)\Delta} =$$
$$= e^{-\zeta_n-\ldots-\zeta_t-(n-t+1)\Delta}[1 + e^{\zeta_t+\Delta} + \ldots + e^{\zeta_t+\ldots+\zeta_{n-1}+(n-t)\Delta} +$$
$$+ e^{\xi_{t-1}-\Delta} + \ldots + e^{\xi_{t-1}+\ldots+\xi_1-(t-1)\Delta}].$$

Пользуясь независимостью $\{\xi_i\}$ и $\{\zeta_i\}$ и сдвигая индексы у $\zeta_\bullet$ на $t-1$, получим

$$\tau(A)-t+1 \underset{d}{=} \tau \equiv \min\left\{k\geqslant 1\colon -Z_k^{\Delta} + \ln\left[1+E_{1,t-1}^{\Delta}+E_{2,k-1}^{\Delta}\right]\geqslant h\right\},$$

где

$$Z_k^\Delta = Z_k + k\Delta, \quad S_k^\Delta = S_k - k\Delta,$$

$$E_{1,n}^\Delta = \sum_{k=1}^n e^{S_k^\Delta}, \quad E_{2,n}^\Delta = \sum_{k=1}^n e^{Z_k^\Delta}.$$

Таким образом, величина $\tau(A) - t + 1$ при условии $\tau(A) \geqslant t$ имеет такое же распределение, как момент остановки τ для мартингала $\{-Z_k^\Delta - k(\rho - \Delta)\}$. Нетрудно видеть, что здесь, как и для величины η в доказательстве теоремы 72.5, применимо тождество Вальда–Дуба, в силу которого $-\mathbf{E}Z_\tau^\Delta = \mathbf{E}\tau(\rho - \Delta)$. Отсюда вытекает

$$\mathbf{E}\tau = -\frac{\mathbf{E}Z_\tau^\Delta}{\rho - \Delta},$$

где

$$-Z_\tau^\Delta = h - \ln\left[1 + E_{1,t}^\Delta + E_{2,\tau-1}^\Delta\right] + \chi_t.$$

Здесь «перескок» χ_t имеет тот же смысл, что в доказательстве теоремы 72.5, и для $\mathbf{E}\chi_t$ справедливо то же утверждение, что в теореме 72.5. Кроме того, $E_{i,k}^\Delta \uparrow E_{i,\infty}^\Delta < \infty$ и в силу леммы 2 $\mathbf{E}\ln\left[1 + E_{1,\infty}^\Delta + E_{2,\infty}^\Delta\right] < \infty$ при достаточно малых Δ;

$$E_{i,\infty}^\Delta \to E_i \quad \text{при} \quad \Delta \to 0, \quad i = 1, 2.$$

Поэтому

$$\mathbf{E}\tau = \frac{h - \mathbf{E}\ln[1 + E_1 + E_2] + \mathbf{E}\chi_t}{\rho} + o(1),$$

где $\sup\limits_{t,\Delta} \mathbf{E}\chi_t < c < \infty$ и для нерешетчатых ζ_k выполняется $\mathbf{E}\chi_t = \mathbf{E}\chi + o(1)$ при $t \to \infty$, $\Delta \to 0$ (см. [23]). Остается заметить, что в силу (11)

$$\begin{aligned}\mathbf{E}(\tau(A) - \theta)^+ &= p\sum_{t=1}^\infty (1-p)^{t-1}\mathbf{P}_t(\tau(A) \geqslant t)\mathbf{E}\tau = \\ &= \frac{h - \mathbf{E}\ln[1 + E_1 + E_2]}{\rho}\mathbf{P}_t(\tau(A) \geqslant t) + u_p + o(1) = \\ &= \frac{-\ln p + \ln\dfrac{A}{1-A} - \mathbf{E}\ln[1 + E_1 + E_2]}{\rho}(1 - \varepsilon) + u_p + o(1),\end{aligned}$$

где $\sup\limits_p u_p < \infty$, $u_p = (1 - \varepsilon)\mathbf{E}\chi + o(1)$ для нерешетчатых ζ_k. Лемма доказана. $\triangleleft$

Мы можем обратиться теперь непосредственно к доказательству теоремы 72.7.

Возьмем произвольный асимптотически однородный момент остановки $\tau \in \in K_{(\varepsilon_1)}$, $\varepsilon_1 < \varepsilon$ (см. п. 3 из § 72). Тогда $\mathbf{P}_t(\tau < t) \to \varepsilon_1$ при $t \to \infty$ и, стало быть, $\mathbf{P}_Q(\tau < \theta) \to \varepsilon_1$ при $p \to 0$. Это означает, что $\tau \in K_{(\varepsilon,Q)}$ при всех достаточно

малых p. Имеем в силу леммы 3

$$\mathbf{E}(\tau - Q)^+ = p\sum_{t=1}^{\infty}(1-p)^{t-1}\mathbf{E}_t(\tau - t)^+ =$$

$$= p\sum_{t=1}^{\infty}(1-p)^{t-1}\mathbf{P}_t(\tau \geqslant t)v\ln t(1+o(1)) = -v(\ln p)(1-\varepsilon_1)(1+o(1)).$$

Следовательно, в силу теоремы 72.8 при $A = A_\varepsilon$ и леммы 4

$$v(1-\varepsilon_1) = \lim_{p\to 0}\frac{\mathbf{E}_Q(\tau-\theta)^+}{-\ln p} \geqslant \lim_{p\to 0}\frac{\mathbf{E}_Q(\tau(A)-\theta)^+}{-\ln p} = \frac{(1-\varepsilon)}{\rho},$$

$$v \geqslant \frac{(1-\varepsilon)}{(1-\varepsilon_1)\rho}.$$

Так как ε_1 — произвольное число, меньшее ε, то мы получаем $v \geqslant 1/\rho$. Теорема доказана. $\triangleleft$

Из теоремы 72.5 видно, что полученная нижняя граница $1/\rho$ достигается для момента остановки ν, построенного в (72.17). Сказанное выше означает также, что ν асимптотически эквивалентно (в смысле приведенных утверждений) байесовскому моменту $\tau(A)$ относительно априорного распределения Q (72.31) при $p \to 0$. Нетрудно убедиться, что это останется справедливым и для других распределений Q из класса $\mathcal{Q}$, описанного в п. 2 из § 72.

4. Еще одна конструкция асимптотически оптимального момента остановки. Эта конструкция возникает естественным образом из приведенных в доказательстве теоремы 72.7 (см. также лемму 4) рассмотрений, из которых видно, что оптимальное правило строится на последовательности статистик M_n (см. (12), (13)), которые асимптотически близки при $p \to 0$ к последовательности

$$M_n^0 = r_n^{-1} + (r_n r_{n-1})^{-1} + \ldots + (r_n \ldots r_1)^{-1}.$$

Из доказательства леммы 2 видно, что статистики $W_n = \ln M_n^0$, допускающие рекурсивное представление

$$W_{n+1} = -\ln r_{n+1} + \ln(1+e^{W_n}) = W_n + \gamma_{n+1} + \ln(1+e^{-W_n}), \tag{14}$$

очень близки по своим свойствам к последовательности $\{w_n\}$, определенной в (72.16). Аналогично тому, как это делалось в п. 3 из § 72 мы можем определить момент остановки

$$\nu_1 = \nu_1(N) = \min\{k \geqslant 1 : W_k \geqslant b(k) + N\}, \tag{15}$$

где функция $b(k)$ имеет тот же вид, что и в теореме 72.5.

Момент остановки ν_1 при $b(k) \equiv 0$ был предложен в [97]. Относительно него в [92, 93] установлено, что он является асимптотически оптимальным в следующем смысле: он минимизирует главную часть асимптотики $\sup_t \mathbf{E}_t(\nu_1 - t/\nu_1 \geqslant t)$ при $B \to \infty$ в классе моментов τ, для которых $E_\infty \tau \geqslant B$.

Для ν_1 справедлив следующий аналог теоремы 72.5.

Теорема 2. *При всех* t

$$\mathbf{P}_t(\nu_1 < t) \leqslant c_1 b e^{-N}, \quad c_1 = \text{const} < \infty. \tag{16}$$

Если распределения ξ_k *непрерывны, то* $P_1(t) = P_1(t, N) = \mathbf{P}_t(\nu_1 < t)$ *и* $P_1(\infty)$ *непрерывно зависят от* N *и для любого* $\varepsilon < \varepsilon_0 = \mathbf{P}\left(\bigcup_{k=1}^{\infty} \{W_k \geqslant b(k)\}\right)$ *существует* $N_1 = N_1(\varepsilon)$ *такое, что* $P_1(t) \uparrow P_1(\infty) = \varepsilon$ *при* $t \to \infty$. *Кроме того,*

$$\mathbf{E}_t\left(\nu_1 - \frac{t}{\nu_1} \geqslant t\right) = \frac{b(t) + N - \mathbf{E}\ln(1 + E_1 + E_2) + \mathbf{E}\chi}{\rho} + o(1), \tag{17}$$

где χ *для нерешетчатых* ζ_k *имеет тот же смысл, что и в теореме* 72.5. *Для нерешетчатых* ζ_k *значение* $\mathbf{E}\chi$ *следует заменить функцией, ограниченной известной постоянной (той же, что и в теореме* 72.5).

Как и в теоремах 72.5, 72.8, условие о непрерывности распределения ξ_k здесь не является существенным.

Доказательство теоремы 2. Если $\theta = \infty$, то $\gamma_k = \xi_k$ и последовательность W_n образует эргодическую цепь Маркова, рассмотренную в доказательстве леммы 2. Распределение W_n монотонно сходится к распределению $W = \ln M_\infty^0 \underset{d}{=} \ln E_1$.

Очевидно, что в нашем случае $\gamma = \xi$ выполнены условия второй части леммы 2 при $s = 1$ и, стало быть,

$$P_1(t, N) = \mathbf{P}_t(\nu_1 < t) = \mathbf{P}\left(\bigcup_{k=1}^{t} \{W_k \geqslant b(k) + N\}\right) \leqslant$$

$$\leqslant \sum_{k=1}^{\infty} \mathbf{P}(W_k \geqslant b(k) + N) \leqslant \sum_{k=1}^{\infty} c_1 e^{-b(k)-N}(1 + o(1)) = c_1 b e^N (1 + o(1)).$$

Это доказывает (16) при, возможно, несколько ином значении постоянной c_1. Непрерывная зависимость $P_1(t, N)$ от N доказывается так же, как в теореме 72.5.

Если $\theta = t$, то $\nu_1 - t$ при условии $\nu_1 \geqslant t$ распределено так же, как время τ первого прохождения уровня $b(t + k) + N$ блужданием W_k при $\gamma_j = -\zeta_j$ и с начальным значением W_0, распределение которого монотонно сходится при $t \to \infty$ к распределению $\ln E_1$. В этом случае приращения $\xi(x) = W_{k+1} - W_k$ превосходят $-\zeta_{k+1}$, а в области $W_k \geqslant M$ допускают мажоранту $\xi(x) \leqslant -\zeta_{k+1} + e^{-M}$. Отсюда и из рассмотрений [23] нетрудно получить существование предельного значения $\mathbf{E}\chi$ в нерешетчатом случае и ограниченность соответствующей функции в общем случае. Далее,

$$e^{W_n} = e^{Z_n + W_0} + e^{Z_n} + e^{Z_n - Z_1} + \ldots + e^{Z_n - Z_{n-1}} = e^{Z_n}(1 + E_1 + E_2 + \varepsilon_n),$$

где $\varepsilon_n \underset{P}{\to} 0$. Таким образом, W_n ведет себя асимптотически как $Z_n + \ln(1 + E_1 + E_2)$, где E_1 и E_2 независимы. С помощью тождества Дуба–Вальда находим, как и прежде,

$$\mathbf{E}\tau = \frac{\mathbf{E}Z_\tau}{\rho} = \rho^{-1}\mathbf{E}[b(t + \tau) + N_1 - \ln(1 + E_1 + E_2) + \mathbf{E}\chi_t] =$$

$$= \rho^{-1}\left[b(t) + N_1 - \mathbf{E}\ln(1 + E_1 + E_2) + \mathbf{E}\chi_t + o(1)\right],$$

где χ_t имеет прежний смысл величины перескока, $\mathbf{E}\chi_t \to \mathbf{E}\chi$ в нерешетчатом случае. Теорема доказана. $\triangleleft$

Представляет интерес сравнение средних значений задержек моментов остановки ν и ν_1 в условиях теорем 72.5 и 2 (см. правые части в (72.19) и (17)) при фиксированной вероятности ложной тревоги ε. Однако сделать это довольно трудно, так как вид функций $N(\varepsilon)$ и $N_1(\varepsilon)$ в этих теоремах нам неизвестен. Мы может лишь высказать некоторые соображения, которые позволят хотя бы приближенно понять природу этих функций. Заметим, что для функции $P(\infty, N)$ в теореме 72.5 вместо (72.18) при больших N можно использовать в силу леммы 72.1 более точную оценку

$$P(\infty, N) \approx cbe^{-N}, \tag{18}$$

где постоянная c из леммы 72.1. Слабая зависимость событий $A_n = \{w_n < b(n) + N\}$ и A_{n+k} при больших k и сравнение с некоторыми близкими результатами о распределении $\max\limits_{j \leqslant n} w_j$ позволяют сделать предположение, что асимптотическое поведение $P(\infty, N)$ отличается от (18) лишь постоянным множителем K, так что

$$P(\infty, N) \approx Kcbe^{-N}. \tag{19}$$

Такое же замечание справедливо относительно функции

$$P_1(\infty, N) \approx K_1 c_1 be^{-N}, \tag{20}$$

где c_1 из леммы 2, при этом в силу большого сходства в устройстве последовательностей $\{w_k\}$ и $\{W_k\}$ значения множителей K и K_1 близки друг другу. Из соотношений (19), (20) следует, что при малых ε выполняется

$$N_1(\varepsilon) - N(\varepsilon) \approx \ln c_1 - \ln c.$$

Взяв разность между правыми частями в (72.19) и (17), мы получим, что «выигрыш» при использовании ν_1 по сравнению с ν будет измеряться величиной

$$\mathbf{E} \ln (1 + E_1 + E_2) - \mathbf{E} \max (S^+, Z^+) - \ln \frac{c_1}{c}. \tag{21}$$

Если эта величина окажется значимой положительной, то момент остановки ν_1 следует считать более предпочтительным. Относительно разности (21) заметим, что в ней, очевидно,

$$\ln (1 + E_1) > S^+, \quad \ln (1 + E_1 + E_2) > \max (S^+, Z^+), \quad c_1 > c.$$

Приложение VIII

Доказательство двух фундаментальных теорем теории статистических игр

Мы будем предполагать здесь, что выполнены следующие условия.

Условие (A). *Множество решений D и множество параметров (чистых стратегий природы) Θ представляют собой компактные метрические пространства соответственно с метриками ρ_D и ρ_Θ.*

Условие (B). *Функция потерь $w(\delta,\theta)\colon D\times\Theta\to\mathbb{R}$ непрерывна по δ и θ в метриках ρ_D и ρ_Θ соответственно.*

Свойство $w(\delta,\theta)\geqslant 0$ нам не потребуется, и мы не будем предполагать, что оно имеет место.

Далее, мы располагаем выборкой $X\in\mathbf{P}_\theta$ из распределения $\mathbf{P}_\theta$. Ее объем n, не ограничивая общности, можно считать равным единице.

Условие (C). *Распределения $\mathbf{P}_\theta$ непрерывны по вариации относительно θ, т. е.*

$$\sup_{B\in\mathfrak{B}_\mathcal{X}}|\mathbf{P}_{\theta_m}(B)-\mathbf{P}_\theta(B)|\to 0,$$

если $\rho_\Theta(\theta_m,\theta)\to 0$ при $m\to\infty$.

Если выполнено условие (A_μ), т. е. $\mathbf{P}_\theta$ имеет плотность $f_\theta(x)$ относительно некоторой σ-конечной меры μ на $(\mathcal{X},\mathfrak{B}_\mathcal{X})$:

$$f_\theta(x)=\frac{d\mathbf{P}_\theta}{d\mu}(x),$$

то условие (C) будет эквивалентно непрерывности $f_\theta(x)$ в $L_1(\mathcal{X},\mathfrak{B}_\mathcal{X},\mu)$:

$$\int|f_{\theta_m}(x)-f_\theta(x)|\mu(dx)\to 0,$$

если $\rho_\Theta(\theta_m,\theta)\to 0$ при $m\to\infty$.

Условия (A)–(C) допускают, конечно, возможность множествам D или Θ быть конечными.

Если D конечно и состоит из точек $\delta_1,\ldots,\delta_r$, то условие (A) относительно D будет выполненным (выбор ρ_D не играет роли), а условие (B) будет означать непрерывность относительно ρ_Θ функций $w(\delta_1,\theta),\ldots,w(\delta_r,\theta)$.

Если оба множества D и Θ конечны, то условия (A)–(C) автоматически выполнены.

Обозначим через σ_D, σ_Θ соответственно σ-алгебры борелевских множеств из D и из Θ. Следуя § 75, обозначим через $(\widetilde{\mathcal{D}},\widetilde{\Theta},\widetilde{W})$ усредненную статистическую игру, где элементами $\widetilde{\Theta}$ являются распределения $\mathbf{Q}$ на (Θ,σ_Θ), а элементами $\widetilde{\mathcal{D}}$ являются распределения $\pi(x)=\pi(x,\cdot)$ на (D,σ_D) (для каждого $x\in\mathcal{X}$), где $\pi(x,A)$ для каждого $A\in\sigma_D$ есть функция, измеримая по x.

Функция риска $\widetilde{W}(\pi,\mathbf{Q})$ определяется равенством

$$\widetilde{W}(\pi,\mathbf{Q})=\int\limits_\Theta\int\limits_\mathcal{X}\int\limits_D w(u,t)\pi(x,du)\,f_t(x)\mu(dx)\mathbf{Q}(dt).$$

Если вместо аргумента $\mathbf{Q}$ подставить θ, то $\widetilde{W}(\pi,\theta)$ будет означать $\widetilde{W}(\pi,\mathbf{I}_\theta)$, где $\mathbf{I}_\theta$ есть распределение, сосредоточенное в точке θ. Такое же соглашение будет действовать относительно замены $\pi\in\widetilde{D}$ на $\delta\in\mathcal{D}$. Нам будет удобнее также вместо $\widetilde{W}$ писать W. Это нигде не приведет к недоразумениям.

Лемма 1. *Если выполнены условия* (A)–(C), *то функция* $W(\pi,\theta)$ *непрерывна по* θ *для любой стратегии* $\pi(x)$.

Доказательство. Имеем при $\theta_n\to\theta$

$$|W(\pi,\theta_n)-W(\pi,\theta)|\leqslant|\mathbf{E}_\theta\mathbf{E}[w(\pi(X),\theta)-w(\pi(X),\theta_n)/X]|+$$
$$+|\mathbf{E}_\theta\mathbf{E}[w(\pi(X),\theta_n)/X]-\mathbf{E}_{\theta_n}\mathbf{E}[w(\pi(X),\theta_n)/X]|\leqslant$$
$$\leqslant\int|w(\pi(x),\theta)-w(\pi(x),\theta_n)|\mathbf{P}_\theta(dx)+\sup_{\delta,\theta}|w(\delta,\theta)|\int|\mathbf{P}_{\theta_n}(dx)-\mathbf{P}_\theta(dx)|. \quad (1)$$

Первый интеграл здесь сходится к нулю в силу непрерывности функции w по θ. Сходимость к нулю второго интеграла вытекает из условия (C). Действительно, пусть $f_{\theta_n}(x)$ есть плотность $\mathbf{P}_{\theta_n}$ относительно меры

$$\mu=\mathbf{P}_\theta+\sum_{j=1}^{\infty}2^{-j}\mathbf{P}_{\theta_j},$$

и пусть $B_n=\{x\colon f_{\theta_n}(x)\geqslant f_\theta(x)\}$. Тогда второй интеграл в (1) равен

$$\int|f_{\theta_n}(x)-f_\theta(x)|\mu(dx)=2\int_{B_n}(f_{\theta_n}(x)-f_\theta(x))\mu(dx)=2(\mathbf{P}_{\theta_n}(B_n)-\mathbf{P}_\theta(B_n))\to 0.$$

Лемма доказана. $\triangleleft$

Теорема 1 (первая фундаментальная теорема). *Если выполнены условия* (A)–(C), *то игра* $(\widetilde{\mathcal{D}},\widetilde{\Theta},W)$ *имеет цену и минимаксные стратегии обоих игроков. Другими словами, существуют наименее благоприятное распределение* $\overline{\mathbf{Q}}$ *и минимаксное решающее правило* $\overline{\pi}(x)$:

$$W_*=\sup_{\mathbf{Q}}\inf_{\pi}W(\pi,\mathbf{Q})=W(\overline{\pi},\overline{\mathbf{Q}})=\inf_{\pi}\sup_{\mathbf{Q}}W(\pi,\mathbf{Q})=W^*. \quad (2)$$

В силу леммы 74.1 утверждение (2) эквивалентно тому, что

$$W(\overline{\pi},\uparrow)\equiv\sup_{\mathbf{Q}}W(\overline{\pi},\mathbf{Q})=W(\overline{\pi},\overline{\mathbf{Q}})=\inf_{\pi}(\pi,\overline{\mathbf{Q}})\equiv W(\downarrow,\overline{\mathbf{Q}}). \quad (3)$$

Теорема 2 (вторая фундаментальная теорема). *Если выполнены условия* (A)–(C), *то байесовские решения* $\pi_Q(x)$ *образуют полный класс. Другими словами, для любого* $\pi_0\in\widetilde{\mathcal{D}}$ *найдутся* $\mathbf{Q}\in\widetilde{\Theta}$, $\pi_Q\in\widetilde{\mathcal{D}}$, *такие, что*

1) $W(\pi_Q,\mathbf{Q})=W(\downarrow,\mathbf{Q})$,

2) $W(\pi_Q,\theta)\leqslant W(\pi_0,\theta)$ *при всех* θ.

Доказательство теоремы 2. Вторая фундаментальная теорема является следствием первой. Действительно, рассмотрим произвольную стратегию $\pi_0\in\widetilde{\mathcal{D}}$ и

игру $(\widetilde{\mathcal{D}}, \widetilde{\Theta}, W_0)$, где W_0 построена по функции $w_0(\delta,\theta) = w(\delta,\theta) - W(\pi_0,\theta)$, так что

$$W_0(\pi,\theta) = W(\pi,\theta) - W(\pi_0,\theta). \tag{4}$$

Функция $v(\theta) = W(\pi_0,\theta)$ в силу леммы 1 непрерывна по θ и, стало быть, функция потерь $w_0(\delta,\theta) = w(\delta,\theta) - v(\theta)$ вместе с $w(\delta,\theta)$ удовлетворяет условию (B). Это означает, что к игре $(\widetilde{\mathcal{D}}, \widetilde{\Theta}, W_0)$ применима теорема 1. Так как $W_0(\pi_0,\uparrow) = 0$ (см. (4)), то верхняя цена этой границы удовлетворяет неравенству $W_0^* \leqslant 0$. Из (2), (3) следует тогда, что существуют $\overline{\pi}$, $\overline{\mathbf{Q}}$ такие, что

$$\sup_{\mathbf{P}} W_0(\overline{\pi},\mathbf{P}) = \sup_{\theta} W_0(\overline{\pi},\theta) \leqslant 0, \quad \overline{\pi} = \pi_{\overline{Q}}.$$

Эти два соотношения эквивалентны утверждениям 2, 1 теоремы 2, если положить $\overline{\mathbf{Q}} = \mathbf{Q}$, $\overline{\pi} = \pi_Q$. Теорема доказана. $\triangleleft$

Доказательство теоремы 1 будет вытекать из следующих двух лемм.

Лемма 2. *При выполнении условий* (A)–(C) *существует распределение* $\overline{\mathbf{Q}}$ *такое, что* $W(\downarrow,\overline{\mathbf{Q}}) \geqslant \inf\limits_{\pi} W(\pi,\uparrow) \equiv W^*$.

Лемма 3. *При выполнении условий* (A)–(C) *существует стратегия* $\overline{\pi}$ *такая, что* $W(\overline{\pi},\uparrow) \leqslant W^*$.

Из неравенств лемм 2, 3 следует соотношение

$$W^* \geqslant W(\overline{\pi},\uparrow) \geqslant W(\overline{\pi},\overline{\mathbf{Q}}) \geqslant W(\downarrow,\overline{\mathbf{Q}}) \geqslant W^*,$$

эквивалентное (3) и, стало быть, (2). Это доказывает теорему 1. $\triangleleft$

Леммы 2, 3 разделяют доказательство теоремы 1 на две части. Первая из них (лемма 2) очень мало связана с тем, что игра является статистической. Эта часть доказательства проводится примерно так же, как для обычных игр (ср. с [73]).

Доказательство леммы 2. Пусть V есть множество функций $\Theta \to \mathbb{R}$, представимых в виде $v(\theta) = W(\pi,\theta)$, $\pi \in \widetilde{\mathcal{D}}$. В силу леммы 1 все функции из V непрерывны, так что $V \subset C(\Theta)$, где $C(\Theta)$ — пространство всех непрерывных функций на Θ. Далее, пусть $v_1(\theta) = W(\pi_1,\theta)$, $v_2(\theta) = W(\pi_2,\theta)$. Так как при $p \in [0,1]$

$$v(\theta) = pv_1(\theta) + (1-p)v_2(\theta) = W(p\pi_1 + (1-p)\pi_2,\theta),$$

$$\pi = p\pi_1 + (1-p)\pi_2 \in \widetilde{\mathcal{D}},$$

то $v \in V$ и, стало быть, множество V выпукло.

Заметим теперь, что $W^* = \inf\limits_{\pi} W(\pi,\uparrow) = \inf\limits_{v\in V} \sup\limits_{\theta} v(\theta)$. Нам будет удобнее вместо исходной функции $w(\delta,\theta)$ рассматривать функцию $\dfrac{w(\delta,\theta) - v_0 + 1}{W^* - v_0 + 1}$, $v_0 = \inf\limits_{v\in V} \inf\limits_{\theta} v(\theta)$. Обозначая новую функцию снова через $w(\delta,\theta)$ (задача при этом остается неизменной), мы получим, что для нее

$$W^* = 1, \quad v_0 > 0. \tag{5}$$

Пусть теперь U есть множество непрерывных функций $v(\theta)\colon \Theta \to \mathbb{R}$ таких, что $\sup\limits_{\theta} v(\theta) < 1$. Очевидно, что U — выпуклое открытое множество из $C(\Theta)$. Кроме

того, из (5) следует, что пересечение $V \cap U$ пусто. Поэтому в силу теоремы Хана–Банаха (см., например, [140, с. 171, 200–206]) существует линейный функционал $L(v)\colon C(\Theta) \to \mathbb{R}$ такой, что

$$L(v) < 1 \quad \text{при} \quad v \in U, \quad L(v) \geqslant 1 \quad \text{при} \quad v \in V. \tag{6}$$

Этот функционал с необходимостью обладает свойством $L(v) \geqslant 0$, если $v(\downarrow) \equiv \equiv \inf\limits_{\theta} v(\theta) \geqslant 0$. Действительно, допустив существование элемента $v_0 \in C(\Theta)$, $v_0(\downarrow) \geqslant 0$, для которого $L(v_0) < 0$, мы получим, что $v_s = -sv_0 \in U$ при любом $s > 0$, $L(v_s) = -sL(v_0) > 1$ при достаточно большом s. Это приводит к противоречию с (6).

Но неотрицательный функционал L в силу теоремы Рисса (см. [120, с. 240]) допускает представление в виде интеграла

$$L(v) = \int\limits_{\Theta} v(\theta)\lambda(d\theta),$$

где λ — конечная мера. Так как $1 \geqslant \sup\limits_{v \in U} L(v) = \lambda(\Theta)$, то, полагая $\overline{\mathbf{Q}}(A) = \lambda(A)/\lambda(\Theta)$, мы получим при $v \in V$

$$L(v) = \int W(\pi, \theta)\lambda(d\theta) = \lambda(\Theta)W(\pi, \overline{\mathbf{Q}}),$$

$$W(\downarrow, \overline{\mathbf{Q}}) = \frac{1}{\lambda(\Theta)} \inf_{v \in V} L(v) \geqslant 1 = W^*.$$

Лемма доказана. ◁

Доказательство леммы 3. Так как функция $W(\pi, \theta)$ при каждом $\pi \in \widetilde{\mathcal{D}}$ непрерывна по θ (см. лемму 1), то нам достаточно построить стратегию $\overline{\pi}$, для которой при всех $k = 1, 2, \ldots$

$$W(\overline{\pi}, \theta_k) \leqslant W^*, \tag{7}$$

где θ_k суть точки некоторого счетного всюду плотного в D множества $T = \{\theta_1, \theta_2, \ldots\}$. По определению верхней цены W^* существует последовательность стратегий $\pi_n = \pi_n(x, \cdot)$ такая, что

$$W(\pi_n, \theta_k) < W^* + \frac{1}{n} \tag{8}$$

при всех k.

Построим теперь с помощью распределений π_n последовательность специально подобранных случайных элементов ζ_n и выделим из нее сходящуюся подпоследовательность. Для этого обозначим через $f_{\theta_k}(x)$ плотность распределения $\mathbf{P}_{\theta_k}$ относительно вероятностной меры $\mu = \sum\limits_{j=1}^{\infty} 2^{-k}\mathbf{P}_{\theta_k}$, так что

$$W(\pi_n, \theta_k) = \iint w(u, \theta_k)\pi_n(x, du)f_{\theta_k}(x)\mu(dx). \tag{9}$$

Рассмотрим пространство $D \times \mathbb{R}^\top$, где $\mathbb{R}^\top$ есть пространство значений элементов $f(x) = \{f_{\theta_1}(x), f_{\theta_2}(x), \ldots\}$ с σ-алгеброй $\mathfrak{B}^\top$, порожденной цилиндрическими множествами. Каждой стратегии π поставим в соответствие вероятностное пространство $(D \times \mathcal{X}, \sigma_D \times \mathfrak{B}_\mathcal{X}, \mathbf{P})$, где распределение $\mathbf{P}$ определяется равенством

$$\mathbf{P}(\delta \in A, X \in B) = \int_B \mu(dx)\pi(x, A), \quad A \in \sigma_D, \quad B \in \mathfrak{B}_X. \tag{10}$$

На этом пространстве определим случайные элементы

$$\zeta = \zeta(\delta; X) = (\delta; f_{\theta_1}(X), f_{\theta_2}(X), \ldots) = (\delta; f(X))$$

и обозначим через ζ_n элементы, соответствующие π_n, так что ζ_n есть случайные величины на выборочном вероятностном пространстве $(D \times \mathbb{R}^\top, \sigma_D \times \mathfrak{B}^\top, \mathbf{\Pi}_n)$, а распределение $\mathbf{\Pi}_n$ порождено π_n, формулой (10) и отображением $\xi(\delta, x)\colon D \times \mathcal{X} \to \to D \times \mathbb{R}^\top$.

Обозначим через $\mathbf{\Pi}_n^{(k)}$ сужения распределения $\mathbf{\Pi}_{n_s}$ на $D \times \mathbb{R}^k$ (это есть совместное распределение $(\delta; f_{\theta_1}(X), \ldots, f_{\theta_k}(X))$ и через λ — распределение $f(X)$ на $(\mathcal{X}, \mathfrak{B}_\mathcal{X}, \mu)$. Нам понадобится

Лемма 4. *Существуют такое распределение* $\overline{\mathbf{\Pi}}$ *на измеримом пространстве* $(D \times \mathbb{R}^\top, \sigma_D \times \mathfrak{B}^\top)$ *и такая подпоследовательность* $\{\pi_{n^*}\} \subset \{\pi_n\}$, *что*

$$\mathbf{\Pi}_{n^*}^{(k)} \Rightarrow \overline{\mathbf{\Pi}}^{(k)} \tag{11}$$

при любом k ($\overline{\mathbf{\Pi}}^{(k)}$ *суть сужения* $\overline{\mathbf{\Pi}}$),

$$\overline{\mathbf{\Pi}}(D \times C) = \lambda(C), \quad C \in \mathfrak{B}^\top. \tag{12}$$

Доказательство леммы 4 проведем позже.

Обозначим через $\overline{\zeta} = (\overline{\delta}, \overline{f})$ случайный элемент с распределением $\overline{\mathbf{\Pi}}$. Соотношение (12) означает, что распределение $\overline{f}$ совпадает с λ (вторая «координата» ζ_n при изменении n распределения не меняет). Так как пространство D есть метрический компакт, то оно сепарабельно и, стало быть (см. [35, с. 191]), существует условное (регулярное) распределение $\overline{\delta}$ относительно $\overline{f}(X)$, которое мы обозначим через $\mathbf{\Pi}(\bullet/\overline{f}(x))$.

Рассмотрим стратегию $\overline{\pi}(x, A) = \mathbf{\Pi}(\overline{\delta} \in A/\overline{f}(X))$ и докажем, что для нее выполняется (7).

Заметим предварительно, что

$$\mathbf{E}w(\overline{\delta}, \theta_k)\overline{f}_{\theta_k} = \mathbf{E}\overline{f}_{\theta_k}\mathbf{E}\left(\frac{w(\overline{\delta}, \theta_k)}{X}\right) = \int f_{\theta_k}(x) \int w(u, \theta_k)\overline{\pi}(u, dx)\mu(dx) = W(\overline{\pi}, \theta_k). \tag{13}$$

Далее, в силу леммы 4 распределение $(\delta_{n^*}, f_{\theta_k}(X))$ слабо сходится к распределению $(\overline{\delta}, \overline{f}_{\theta_k}(X))$. Поскольку функция w непрерывна, то совместное распределение $(w(\delta_{n^*}, \theta_k), f_{\theta_k}(X))$ слабо сходится к распределению $(w(\overline{\delta}, \theta_k), \overline{f}_{\theta_k}(X))$. Но функция $g(u, v) = w(u, \theta_k)v$ непрерывна по u и v и мажорируется функцией $g(v) = cv$,

$c = \max\limits_u w(u, \theta_k)$, такой, что $\mathbf{E}g(f_{\theta_k}(X)) = c\int f_{\theta_k}(x)\mu(dx) = c$. Поэтому по теореме непрерывности для моментов (см. теорему 5.4)

$$\lim_{n^*\to\infty} \mathbf{E}g(\delta_{n^*}, f_{\theta_k}(X)) = \mathbf{E}g(\overline{\delta}, \overline{f}_{\theta_k}(X)),$$

или, что то же,

$$\lim_{n^*\to\infty} \mathbf{E}w(\delta_{n^*}, \theta_k) f_{\theta_k}(X) = \mathbf{E}w(\overline{\delta}, \theta_k)\overline{f}_{\theta_k}(X).$$

В силу (9), (13) это дает нам сходимость

$$\lim_{n^*\to\infty} W(\pi_{n^*}, \theta_k) = W(\overline{\pi}, \theta_k).$$

Так как левая часть этого равенства (см. (8)) не превосходит W^*, то лемма 3 доказана. $\triangleleft$

Доказательство леммы 4. Фиксируем какое-нибудь $k \geqslant 1$ и рассмотрим $D \times \mathbb{R}^k$ как полное метрическое сепарабельное пространство относительно метрики, порожденной евклидовой метрикой в $\mathbb{R}^k$ и метрикой ρ_D. Для любого $\varepsilon > 0$ в $\mathbb{R}^k$ найдется компакт K_ε такой, что $\mathbf{P}((f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon$. Так как $D \times K_\varepsilon$ есть компакт в $D \times \mathbb{R}^k$ и так как

$$\mathbf{P}(\delta_n \in D, (f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon) \geqslant 1 - \varepsilon,$$

то последовательность распределений $\mathbf{\Pi}_n^{(k)}$ является плотной (см. [7]). Следовательно, по теореме Прохорова [7] существуют распределение $\overline{\mathbf{\Pi}}^{(k)}$ и подпоследовательность $n^{(k)} = (n_1^{(k)}, n_2^{(k)}, \ldots)$ такие, что $\mathbf{\Pi}_{n^{(k)}}^{(k)} \Rightarrow \overline{\mathbf{\Pi}}^{(k)}$. Но распределения $\overline{\mathbf{\Pi}}^{(k)}$, очевидно, согласованы и, следовательно, по теореме Колмогорова на $(D\times\mathbb{R}^\top, \sigma_D \times \mathfrak{B}^\top)$ существует распределение $\overline{\mathbf{\Pi}}$, для которого $\overline{\mathbf{\Pi}}^{(k)}$ являются сужениями на $(D \times \mathbb{R}^k, \sigma_D \times \mathfrak{B}^k)$.

С другой стороны, можно считать, что $n^{(k+1)} \subset n^{(k)}$. Положив $n^* = (n_1^{(1)}, n_2^{(2)}, n_3^{(3)}, \ldots)$, получим подпоследовательность, для которой $\mathbf{\Pi}_{n^*}^{(k)} \Rightarrow \overline{\mathbf{\Pi}}^{(k)}$ при всех k.

Докажем теперь (12). Пусть $C \in \mathfrak{B}^\top$ есть цилиндрическое множество такое, что $\overline{\mathbf{\Pi}}$-мера его границы равна нулю. Обозначим через $C^{(k)} = C \cap \mathbb{R}^k \in \mathfrak{B}^k$ множество из $\mathbb{R}^k$, образованное первыми k координатами точек из C, и положим $\overline{C}^{(k)} = C^{(k)} \times \mathbb{R}^{T-k} \in \mathfrak{B}^\top$. Тогда $\lambda(\overline{C}^{(k)}) = \mathbf{\Pi}_{n^*}^{(k)}(D \times C^{(k)}) \to \overline{\mathbf{\Pi}}^{(k)}(D \times C^{(k)})$. Так как $\overline{C}^{(k+1)} \subset \overline{C}^{(k)}$, $C = \bigcap\limits_{k=1}^{\infty} \overline{C}^{(k)}$, то

$$\lambda(C) = \lim_{k\to\infty} \lambda(\overline{C}^{(k)}) = \lim_{k\to\infty} \overline{\mathbf{\Pi}}^{(k)}(D \times C^{(k)}) = \lim_{k\to\infty} \overline{\mathbf{\Pi}}(D \times \overline{C}^{(k)}) = \overline{\mathbf{\Pi}}(D \times C).$$

Лемма 4 доказана. $\triangleleft$

БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ

Ниже приводятся библиографические комментарии, в которых делается попытка проследить историю появления основных идей и результатов, излагаемых в книге. Эти комментарии не претендуют на полноту и часто будут содержать ссылки не на оригинальные труднодоступные статьи, а на учебники, монографии или обзорные статьи, в которых нужные результаты найти проще. Более обширные библиографические указания и исторические справки содержатся, например, в [48, 65–67].

Некоторые основные понятия математической статистики возникли еще в начале прошлого века и связаны с именами Лапласа и Гаусса. В конце прошлого века работы Пирсона положили начало периоду интенсивного развития этой науки. Оно было обусловлено фундаментальными работами Р. Фишера, Ю. Неймана, А. Н. Колмогорова, А. Вальда. В нашей стране развитие математической статистики связано прежде всего с именами А. Н. Колмогорова и Н. И. Смирнова.

Глава 1

§ 2–4. Теорема Гливенко–Кантелли была установлена в 1933 г. (Гливенко принадлежит доказательство для непрерывного распределения, Кантелли — для общего случая).

Доказательство теоремы 2.2 близко к доказательству, приведенному в [72, с. 28] и представляет собой частный случай использования более общего подхода, основанного на «конечной аппроксимируемости» рассматриваемого класса множеств. В полном виде этот подход изложен в приложении I, где доказана теорема 4.2. Независимо аналогичный подход был рассмотрен в [43]. Закон повторного логарифма (теорема 4.3) установлен в [52].

Статистики вида $G(F_n^*)$ иногда называют также статистическими функциями. Их систематическое исследование было начато фон Мизесом (см. [75, 76, 115]).

M-статистики были введены Хьюбером в [127] (см. также [128, 129, 122, 102, 67]) как обобщение оценок максимального правдоподобия. Об L-статистиках см. в [67, 102, 139].

§ 5. Теоремы непрерывности, изложенные в § 5, встречаются в качестве вспомогательных утверждений во многих работах; о некоторых из них см., например, [102].

§ 6. Теоремы 6.1, 6.2 о распределении $nF_n^*(t)$ содержатся в книге Феллера [111, т. 2, гл. III, § 3]. Теорема 6.3 о сходимости процесса $\sqrt{n}\,(F_n^*(t) - F(t))$ к броуновскому мосту, доказанная в приложении II, была установлена Донскером в [44]. Несколько иное по сравнению с приложением II доказательство теоремы 6.3 можно найти в книге Биллингсли [7].

§ 7. Утверждение примера 7.3 о предельном распределении статистики $\chi^2(X)$ (хи-квадрат) впервые получено Пирсоном (см. [58, с. 454]).

§ 8. Исследование предельного распределения статистик второго типа или «статистических функций» по Мизесу было начато в [75, 76]. Ближе других к теореме 2 находятся результаты Филиппповой [115], которая изучала статистические функции специального вида. При доказательстве теорем 8.2 и 6.3 реализован так называемый аппроксимативный метод изучения сходимости случайных процессов, изложенный в [11]. Утверждение следствия 8.2 составляет содержание теоремы Колмогорова, а следствия 8.3 — теоремы Смирнова. Последняя включает в себя также явный вид распределения $\int_0^1 [w^0(t)]^2\,dt$, который ввиду его сложности мы не приводим (см. [105]).

О предельном распределении L-статистик см. [139, 102].

§ 10. Оценки плотности, рассматриваемые в этом параграфе, были введены Парзеном [89] и Розенблаттом [98]. Обзор результатов в этом направлении и библиографию см. в обзорной работе Розенблатта [99] и в § 25 книги Ченцова [131].

Глава 2

§ 12. Некоторые другие параметрические семейства описаны в книге Уилкса [109]. Весьма полное исследование распределений членов вариационного ряда выполнено Гнеденко. Полное изложение результатов и обширную библиографию по этому поводу можно найти в книге Дейвида [42].

§ 14. Метод моментов исторически является первым регулярным методом построения оценок. Он был предложен Пирсоном в 1894 г. M-оценки были введены Хьюбером (см. [127–129]). Об асимптотической нормальности M-оценок см. также [67, 102].

§ 15. Метод минимума χ^2 был предложен Фишером в 1922 г.

§ 16. Метод максимального правдоподобия в частных случаях применялся еще Гауссом. Как общий метод получения оценок он был предложен Фишером в краткой заметке в 1912 г. Позднее, в 1925 г. в классической работе [117] им же были изучены асимптотические свойства о. м. п.

Исследование состоятельности о. м. п. было начато в 30–40-х годах прошлого столетия в работах Дуба [45], Вальда [28], Вольфовица [33], Крамера [58]. Основные условия состоятельности в [28] включают в себя (помимо условий (A_μ), (A_c), (A_0)) принадлежность $f_t(x)$ классу D_0 и интегрируемость

$$\int \ln f_t^{\Delta}(x) f_\theta(x) \mu(dx).$$

В монографии [49] получены условия состоятельности, использующие сходимость

$$\int \sup_{|u| \leqslant \Delta} (\sqrt{f_{t+u}(x)} - \sqrt{f_t(x)})^2 \mu(dx) \to 0 \quad \text{при} \quad \Delta \to 0.$$

Результаты теорем 14.1, 14.3 и их следствий 16.1–16.5 являются более общими. Метод доказательства близок к [28]. Достаточность условий (A_0^0) и (7) была замечена Саханенко.

Асимптотическая нормальность оценок максимального правдоподобия изучалась во многих работах (см., например, [49, 67, 102, 117]). Условия асимптотической нормальности в теореме 16.3 близки к минимальным (ср. также с результатами в § 34). Оптимальность оценок максимального правдоподобия в классе M-оценок отмечена в [102].

§ 17, 18. Излагаемые подходы к сравнению оценок являются общепринятыми. Доказательство леммы 17.3 заимствовано из [58]. Понятие эффективной оценки было введено Фишером в 1922 г. в [116].

В существующей статистической литературе один и тот же термин «эффективность» используется в двух разных смыслах. Эффективными оценками называют: (a) оценки с минимальными дисперсиями и (b) оценки, для которых достигается нижняя граница в неравенстве Рао–Крамера. Классы оценок (a) и (b), вообще говоря, не совпадают. В настоящей книге для оценок (b) используется термин «R-эффективность» (см. § 26).

§ 19, 20. Фундаментальное понятие условного математического ожидания было введено Колмогоровым в 1933 г. в классической работе [56]. Свойства условных распределений детально изучены в [35, 46, 137].

§ 21. Байесовский подход широко использовался еще в прошлом веке Лапласом. Этот подход подвергался критике со стороны Фишера, и в 20–30-х годах прошлого столетия центр тяжести исследований переместился на эффективные и асимптотически эффективные оценки. Затем по мере того, как стала осознаваться фундаментальная роль байесовского подхода, интерес к нему вновь стал возрастать.

Понятие минимаксной оценки вошло в математическую статистику вместе с теоретико-игровым подходом, развитым в работах Бореля (1921) и Дж. Неймана (1928), теоремы 21.1–21.3 были получены Ходжесом и Леманом [124].

§ 22. Фундаментальное понятие достаточной статистики было введено Фишером в [116] в 1922 г. Фишер [116] и позже Ю. Нейман [82] предложили простой критерий, обнаруживающий существование и вид достаточной статистики. Этот критерий носит название факторизационной теоремы Неймана–Фишера и представлен в теореме 22.1. Строгое теоретико-множественное доказательство теоремы Неймана–Фишера было получено лишь в 1949 г. Халмошем и Сэвиджем [121].

Более простое доказательство этой теоремы, приведенное в книге и основанное на теореме 22.1В, автору предложил В. В. Юринский.

§ 23. Понятие достаточной σ-алгебры шире, чем понятие достаточной статистики. Необходимые и достаточные условия для их совпадения приведены в [48]. Конструкция достаточных разбиений и теорема 23.1 связаны с работой Лемана и Шеффе [68], посвященной выяснению условий существования и построению минимальных достаточных статистик. Краткое изложение этой статьи можно найти в [48]. Доказательство теоремы 23.2 было предложено И. С. Борисовым.

§ 24. Теорема 24.1 была получена независимо Блекуэллом [8] (1947), Рао [95] (1945), [96] (1949) и Колмогоровым [55] (1950). Теорема 24.3 принадлежит Рао [96] (1949) и Блекуэллу [8] (1947).

§ 25. Экспоненциальное семейство упоминалось еще Фишером в [116]. Теоретическое значение этого семейства было осознано в 30-х годах в работах Питмена, Купмена, Дармуа. Именами последних иногда и называют экспоненциальное семейство. Теорема 25.2 доказана Леманом [65, с. 183; 66].

§ 26, 27. Неравенство Рао–Крамера иногда называют также неравенством информации. Оно принадлежит, по существу, Фишеру [119], хотя в приводимой форме оно было получено независимо Фреше [119] в 1943 г., Рао [94] в 1945 г. и Крамером [57] в 1946 г.

Условия регулярности, нужные для выполнения неравенства, в руководствах по математической статистике трактуются не всегда аккуратно. Мы имеем в виду условия, обеспечивающие законность дифференцирования по параметру под знаком интеграла. Доказательство этой законности часто содержит пробелы (см., например, [48]) либо отсутствует вовсе (например, в [31]). В ряде случаев она оговаривается в виде условия [31], что неудобно для проверки в реальных задачах.

Условия регулярности, принятые в книге, весьма просты, хотя, по-видимому, не являются наиболее общими (ср. с [49]). Тот факт, что при этих условиях можно дифференцировать под знаком интеграла, доказан в приложении V, написанном на основе результатов А. И. Саханенко.

Различные обобщения неравенства Рао–Крамера см. в [48, 131]. Понятие информации (Фишера) было введено в [117]. При доказательстве теорем 26.1А и 27.1 мы следуем книгам [48, 49].

§ 28, 29. Идея использования инвариантных соображений принадлежит Хотеллингу и Питмену. Значительный вклад в разработку теории внес Стейн. Основное содержание теоремы 28.1 принадлежит Питмену. При ее доказательстве мы использовали изложение в [48, 49]. Минимаксность оценки Питмена была установлена Гиршиком и Сэвиджем.

§ 30. Результаты этого параграфа были получены автором совместно с Саханенко [24]. Некоторые неравенства при более жестких ограничениях можно получить также из работ [39, 130].

§ 31. Расстояние Кульбака–Лейблера в параметрическом случае называется также функцией информации Кульбака–Лейблера. К этому расстоянию независимым образом пришел И. Н. Санов при описании вероятностей больших уклонений эмпирического распределения. Идея широкого использования расстояния Хеллингера для изучения свойств отношения правдоподобия нами заимствована из книги Ибрагимова и Хасьминского [49]. На результатах этой книги базируются и доказательства основных теорем § 33. Доказательство теоремы 31.3 было существенно упрощено Саханенко.

§ 32. Теорема 32.1 была установлена Чепменом и Роббинсом в 1951 г. в [132] и Кифером в 1952 г. в [51].

§ 33, 34. Излагается материал лекций, значительно усовершенствованный после появления книги Ибрагимова и Хасьминского [49]. Основные усовершенствования связаны с систематическим использованием расстояния Хеллингера для оценки $\mathbf{E}_\theta Z^{1/2}(u)$. Предложение использовать $\int \mathbf{E}_\theta |(Z^{3/4}(u))'|\,du$ для оценки $\sup\limits_u Z(u)$ (см. теоремы 33.1, 33.2) принадлежит А. И. Саханенко. Асимптотическая нормальность и асимптотическая эффективность о. м. п. была установлена еще Фишером [117]. Весьма общие условия асимптотической нормальности о. м. п. получены в [49].

Асимптотическая нормальность апостериорной плотности (или отношения правдоподобия) была обнаружена С. Н. Бернштейном в 1927 г. Теорема 33.8 принадлежит Бахадуру [2]. Асимптотическая байесовость и минимаксность о. м. п. легко получаются благодаря результатам § 30. Ранее асимптотическая байесовость о. м. п. устанавливалась при более жестких ограничениях на плотность априорного распределения.

При доказательстве теорем 34.1, 34.2 мы использовали некоторые усовершенствования, предложенные А. И. Саханенко.

§ 35. Излагается один из вариантов численного метода Ньютона–Рафсона отыскания экстремума функции. Более подробное изложение см. в [48]. Пример 3 заимствован из книги Рао [96].

§ 36, 37. См. комментарии к § 33, 34. Пример 36.3 заимствован из книги Ван-дер-Вардена [31]. В изложении параграфов внесен ряд усовершенствований, предложенных А. И. Саханенко (в частности, добавлена теорема 37.5). Эти изменения позволили упростить текст в § 43–45 гл. 3.

§ 38. Подробнее о последовательном оценивании см., например, в [48].

§ 39, 40. Впервые доверительные интервалы встречаются, по-видимому, у Лапласа. Еще в 1812 г. он показал, что можно обратить относительно p утверждение о степени расхождения наблюдаемой частоты и биномиальной вероятности p с тем, чтобы найти интервал для возможных значений p. Правильная интерпретация доверительных интервалов (не предполагающая случайность параметра) дана в 1927 г. Вилсоном.

Общий метод отыскания точных доверительных интервалов для вещественного параметра был предложен Фишером в 1930 г. в [118]. В 1937–38 г. Ю. Нейманом была развита общая теория доверительных утверждений и установлена их связь с теорией проверки гипотез. Весьма полное современное изложение вопроса можно найти в книге Лемана [65]. Это изложение использовано нами в § 47.

Теорема 40.1 и лемма 40.2 принадлежат Фишеру.

Глава 3

Первые отдельные применения статистических критериев восходят к Лапласу (конец XVIII в.). Систематическое использование критериев для проверки гипотез начинается с работ Пирсона, предложившего в 1900 г. критерий χ^2. Фундаментальные понятия ошибок первого и второго рода были введены Нейманом и Пирсоном в 1928 г. в [84]. Этими же авторами впервые была осознана роль альтернатив для рационального выбора критерия. В итоговой работе Неймана и Пирсона [85] развита теория р. н. м. к.

Систематическое изложение теории проверки гипотез содержится в книге Лемана [65].

§ 41–43. Фундаментальная лемма Неймана–Пирсона получена в [85]. Теоремы 41.1, 41.2 можно извлечь из книги Блекуэлла и Гиршика [9]. Теорема 42.1 содержится в книге Лемана [65]. Теорема 43.1 о больших уклонениях принадлежит Крамеру (см. [17]). Оценка качества критериев, связанная с вероятностями больших уклонений, легла в основу понятия эффективности критерия по Бахадуру. Итоги исследований в этом направлении изложены в [4, 22].

Роль статистики эффективного вклада была отмечена еще в 1925 г. в работе Фишера [117]. В дальнейшем подход, связанный с рассмотрением близких гипотез, интенсивно развивался в работах Ле Кама, Русаса, Чибисова (см. также комментарии к § 44, 45).

§ 44. Излагаемая общая концепция статистических критериев является общепринятой (см. [58, 65]). Понятие р. н. м. к. было введено Нейманом и Пирсом в [85]. Байесовский подход использовался еще Лапласом в XIX в.

§ 45–48. Основные результаты этих параграфов взяты из книги Лемана [65]. Изложение также близко к этой книге, но отличается тем, что основано на байесовском подходе, а не на обобщенной лемме Неймана–Пирсона (лемма 45.2, см. также [65]). Это упрощает изложение и делает его более цельным.

По поводу доверительных множеств см. комментарии к § 39, 40.

О возможности перенесения основных результатов на случайные процессы см. книгу Гренандера [37].

§ 49. Теорема 49.1 принадлежит Ходжесу и Леману [124].

§ 50. Фундаментальная роль отношения правдоподобия в математической статистике была выяснена в работах Неймана и Пирсона [84, 85]. К. о. п. посвящена обширная литература. Попытки установить те или иные свойства асимптотической оптимальности этого критерия содержатся в работах [3, 28, 88, 109, 123].

§ 51. Основной вклад в развитие теории последовательного анализа принадлежит Вальду [29]. Наиболее компактное изложение главных результатов, которому мы следуем в книге, можно найти в [65].

§ 52. По поводу критериев Колмогорова и ω^2 см. § 8 и комментарии к нему. О некоторых модификациях критерия Колмогорова, дающих наибольшую возможную мощность, см. [26]. Критерий Морана был предложен в [80]. Его мощность для близких альтернатив изучалась в [32, 133].

§ 53. Асимптотическая байесовость к. о. п. установлена в работе автора [13]. Результаты о предельном распределении отношения правдоподобия при основной гипотезе принадлежат Уилксу [108] и Вальду [27] (см. также книгу Уилкса [109]). Идея замены сложной гипотезы усредненной использовалась Вальдом. Асимптотический вид байесовских критериев содержится в [69]. См. также комментарии к § 36, 37 гл. 2.

§ 54, 55. Основные идеи, связанные с отысканием асимптотически оптимальных тестов для близких гипотез, изложены в работах Вальда [27], Ле Кама, Русаса (см. книгу Русаса [100]), Чибисова [134]. О возможности перенесения основных результатов на случай бесконечномерного параметра (на случайные процессы) см. [25]. Форма изложения § 54, 55 с цитированными работами связана мало. Редукция исходной задачи A к задаче B для параметра нормального распределения при отыскании оптимальных критериев для основных типов задач, рассмотренных в § 54, содержится в работе Вальда [27]. Утверждение теоремы 55.4 о распределении статистики $2 \ln R_1(X)$ при гипотезе H_1 можно найти в [109]. См. также комментарии к § 36, 37 гл. 2.

§ 56, 57. Критерий χ^2 был предложен Пирсоном в 1900 г. Ему посвящена обширная литература (см., например, специальную монографию Ланкастера [62]). Обсуждение различных свойств оптимальности см. в [27, 88, 109, 123] и др. О поведении мощности критерия χ^2 при увеличении числа групп см., например, [14, 135]. Примеры 56.1, 57.2 заимствованы из книги Крамера [58], пример 57.1 — из книги Рао [96].

§ 58. Начальный этап в изучении устойчивости статистических решений проследить трудно. Более поздние исследования имеют в своей основе работы Тьюки, Ходжеса, Лемана. Обстоятельный обзор этого направления представлен в работе Хьюбера [128].

Глава 4

§ 59. Критерий χ^2 в задаче примера 59.1, критерий Стьюдента в задаче примера 59.3 и критерий Фишера в задачах примеров 59.4, 59.5 используются очень часто. О других свойствах оптимальности этих критериев см. книгу Лемана [65]. Пример 59.1A заимствован из [96]. По проблеме Беренса–Фишера (пример 59.6) имеется обширная литература (см. [65]).

§ 60. Точное распределение статистики $D_{n,n}$ найдено Гнеденко и Королюком (см. [111]); предельное распределение статистики D_{n_1,n_2} — Смирновым. Теорема 60.2 доказана впервые в [74] методом моментов. О критериях знаков и Вилкоксона см. также [34].

§ 61, 62. Подробнее задачи регрессии и дисперсионного анализа изложены в специальных монографиях Себера [101] и Шеффе [136]. См. также [58, 65, 96].

§ 63. Замечание об асимптотической оптимальности критерия (63.3) содержится в [13].

Глава 5

§ 64. Изучению задач регрессии (пример 2) посвящена обширная литература (см. комментарии к § 61, 62 гл. 4, а также [101, 136]). Теория статистических выводов для случайных процессов (пример 4) является технически весьма сложной и, по-видимому, окончательно еще не сформировалась. Некоторые важные элементы этой теории можно найти в монографиях [1, 38, 61, 70]. Пример 7 индуцирован работами, ссылки на которые содержатся в [59, 87].

§ 65, 66. Асимптотические свойства оценок максимального правдоподобия в неоднородном случае изучались в [114, 113, 125].

§ 68. Основные результаты и библиографию, относящиеся к задаче об оценке хвостов распределений, см., например, в [59, 87]. Некоторые результаты § 68 близки результатам этих работ.

§ 70. См. комментарии к § 65, 66.

§ 72. Задаче о разладке посвящена обширная литература (см., например, [41, 53, 71, 79, 90, 138]). Использование последовательности $\{w_n\}$, определенной в (72.16), было предложено в [71, 90]. Теорема 72.7 об оптимальном байесовском моменте остановки получена в [138]. Основные результаты § 72 и приложения VII об асимптотически однородных оптимальных процедурах были получены автором специально для предлагаемой книги.

Глава 6

В математике теоретико-игровое направление возникло с появлением работ Бореля 1921 г. и фон Неймана 1928 г. В математической статистике исходной работой, подготовившей использование теории игр, можно считать классическую работу Неймана и Пирсона [86], в которой были высказаны многие основные идеи теории статистических решений. Фундаментальный вклад в развитие общей теории статистических решений принадлежит Вальду. Основные положения этой теории изложены в его завершающей книге [30]. Общематематическая теория игр получила свое полное развитие в книге фон Неймана и Моргенштерна [81]. Доступное изложение основ теории статистических игр можно найти в книгах Блекуэлла и Гиршика [9] и Фергюссона [112].

§ 74. Сравнительно полным введением в обычную теорию игр является книга Мак Кинси [73].

§ 75, 76. Более полное описание основ теории статистических игр см. в [9, 112]. В этих книгах две фундаментальные теоремы теории статистических игр доказаны лишь в частном случае для дискретных множеств D и Θ. Объясняется это большой сложностью изложения в общем случае (см. [30]). В приложении VIII приведено наиболее простое из известных нам доказательств этих теорем. Оно было предложено А. И. Саханенко.

Роль байесовского подхода в разное время оценивалась по-разному. Он широко использовался в прошлом веке Лапласом. Затем байесовский подход был подвергнут критике со стороны Фишера, и в 20–30-х годах прошлого столетия центр тяжести переместился на эффективные оценки и наиболее мощные критерии. Затем по мере того, как стала осознаваться фундаментальная теория байесовского подхода, интерес к нему стал вновь возрастать. Эта фундаментальная роль выясняется в теоремах 75.1, 75.2.

§ 77. Фундаментальное понятие достаточной статистики было введено Фишером в [116] в 1922 г. Фишер [116] и позже Ю. Нейман [82] предложили простой критерий, обнаруживающий существование и вид достаточной статистики. Этот критерий носит название факторизационной теоремы Неймана–Фишера и представлен в теореме 22.1. Строгое теоретико-множественное доказательство теоремы Неймана–Фишера было получено лишь в 1949 г. Халмошем и Сэвиджем [121].

Понятие достаточной σ-алгебры шире, чем понятие достаточной статистики. Необходимые и достаточные условия для их совпадения приведены в [48]. Теорема 77.1 (сначала для квадратической функции потерь) была получена независимо Блекуэллом [8] (1947), Рао [94] (1945), [95] (1949) и Колмогоровым [55] (1950). Обобщения на случай произвольной функции потерь связаны с именами Лемана и Шеффе [68].

Идея использования инвариантных соображений принадлежит Хотеллингу и Питмену. Значительный вклад в разработку теории внес Стейн (см. [48, 49]).

О несмещенности подробнее см. [48].

§ 78. Результаты, близкие к теоремам этого параграфа, можно найти в книге Ибрагимова и Хасьминского [49].

§ 79. Асимптотическая байесовость к. о. п. установлена в работе автора [13]. Результаты о предельном распределении отношения правдоподобия при основной гипотезе принадлежат Уилксу [108] и Вальду [27] (см. также книгу Уилкса [109]). Идея замены сложной гипотезы усредненной использовалась Вальдом. Асимптотический вид байесовских критериев содержится в [69].

§ 80. Основные идеи, связанные с отысканием асимптотически оптимальных тестов для близких гипотез, изложены в работах Вальда [27], Ле Кама, Русаса (см. книгу Русаса [100]), Чибисова [134]. О возможности перенесения основных результатов на случай бесконечномерного параметра (на случайные процессы) см. [25]. Форма изложения § 80 и § 54, 55 с цитированными работами связана мало. Редукция исходной задачи A к задаче B (для параметра нормального распределения) при отыскании оптимальных критериев для основных типов задач содержится в работе Вальда [27].

Приложения

Доказательство двух фундаментальных теорем теории статистических игр из приложения VIII можно найти в [30] (см. также [9, 112]). Это доказательство технически очень сложно и сопровождается ограничительными условиями (предполагается, например, что распределение наблюдений либо решетчато, либо имеет плотность относительно меры Лебега

в $\mathbb{R}^k$). Другой значительно более общий подход был предложен Гошем [36] и Ле Камом [63, 77, 110]. Однако в этих работах требуемые доказательства также труднодоступны, так как их приходится извлекать из громоздких и весьма общих утверждений (включающих в себя критерии последовательного анализа, к тому же в предположении неограниченности функции потерь; см. [63, 110]).

В настоящей книге излагается более короткое и обозримое доказательство, занимающее менее шести страниц в детальном изложении. Оно подготовлено Саханенко и близко по смыслу идеям Ле Кама. Центральной его частью являются леммы 2, 3. Лемма 2, по существу, не связана со статистическим характером игры, основана на теоремах Хана–Банаха и Рисса и близка по своей идее рассуждениям, использованным, например, в [140]. Доказательство леммы 3 основано на теоремах Колмогорова [56] и Прохорова [7].

При составлении таблиц I–IV использовалась книга Большева и Смирнова [10].

СПИСОК ЛИТЕРАТУРЫ

1. Басава И., Пракаса Рао Б. (Basawa I.V., Prakasa Rao B.) Statistic inference for stochastic processes. — N.Y.: Acad. Press, 1980.
2. Бахадур Р. (Bahadur R.R.) On Fisher's bound for asymptotic variances // Ann. Math. Statist. 1964. V. 35, № 4. P. 1545–1552.
3. Бахадур Р. (Bahadur R.R.) An optimal property of the likelihood ratio statistic // Proc. 5-th Berkeley sympos. math. statist. prob. Berkeley; Los Angeles, 1965. V. 1. P. 27–40.
4. Бахадур Р. (Bahadur R.R.) Some limit theorems in statistics. — Philadelphia: SIAM, 1971.
5. Бахадур Р., Леман Э. (Bahadur R.R., Lehman E.L.) Two comments on «Sufficiency and statistical decision functions» // Ann. Math. Statist. 1955. V. 26. P. 139–141.
6. Биллингсли П. (Billingsley P.) Statistical inference for Markov processes. — Chicago: Univ. Chicago Press, 1961.
7. Биллингсли П. Сходимость вероятностных мер. — М.: Наука, 1977.
8. Блекуэлл Д. (Blackwell D.H.) Conditional expectation and unbiased sequential estimation // Ann. Math. Statist. 1947. V. 18. P. 105–110.
9. Блекуэлл Д., Гиршик М.А. Теория игр и статистических решений. — М.: ИЛ, 1958.
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука, 1965.
11. Боровков А.А. Сходимость распределений функцибналов от случайных процессов // Успехи мат. наук. 1972. Т. 27, № 1. С. 3–41.
12. Боровков А.А. Вероятностные процессы в теории массового обслуживания. — М.: Наука, 1972.
13. Боровков А.А. Асимптотически оптимальные тесты для проверки сложных гипотез // Теория вероятностей и ее применения. 1975. Т. 20, № 3. С. 463–487.
14. Боровков А.А. О мощности критерия χ^2 при увеличении числа групп // Теория вероятностей и ее применения. 1977. Т. 22, № 2. С. 375–379.
15. Боровков А.А. Математическая статистика. Оценка параметров. Проверка гипотез. — М.: Наука, 1984.
16. Боровков А.А. Математическая статистика. Дополнительные главы. — М.: Наука, 1984.
17. Боровков А.А. Теория вероятностей. 3-е изд. перераб. и доп. — М.: Эдиториал УРСС, 1999.
18. Боровков А.А. (Borovkov A.A.) Statistique mathematique. — Moscow: Mir, 1987.
19. Боровков А.А. (Borovkov A.A.) Estadistica matematica. — Moscow: Mir, 1988.
20. Боровков А.А. Эргодичность и устойчивость случайных процессов. — М.: Эдиториал УРСС, 1999.
21. Боровков А.А., Коршунов Д.А. Вероятности больших уклонений одномерных цепей Маркова. I // Теория вероятностей и ее применения. 1996. Т. 41, вып. 1. С. 3–30.
22. Боровков А.А., Могульский А.А. Большие уклонения и проверка статистических гипотез. — Новосибирск: Наука, 1992.

23. Боровков А.А., Фосс С.Г. Оценки для перескока случайного блуждания через произвольную границу и их применения // Теория вероятности и ее применение. 1999. Т. 44, № 2. С. 245–277.

24. Боровков А.А., Саханенко А.И. Неравенства типа Рао–Крамера для байесовского риска // Теория вероятностей и ее применения. 1980. Т. 25, № 1. С. 207–209.

25. Боровков А.А., Саханенко А.И. Об асимптотически оптимальных тестах для проверки сложных близких гипотез // Тр. Ин-та математики им. С.Л. Соболева СО АН СССР. 1982. Т. 1. С. 79–89.

26. Боровков А.А., Сычева Н.М. О некоторых асимптотически оптимальных непараметрических критериях // Теория вероятностей и ее применения. 1968. Т. 13, № 3. С. 385–418.

27. Вальд А. (Wald A.) Tests of statistical hypotheses concerning several parameters when the number of observations is large // Trans. Amer. Math. Soc. 1943. V. 54, № 3. P. 426–482.

28. Вальд А. (Wald A.) Note on the consistency of the maximum likelihood estimate // Ann. Math. Statist. 1949. V. 20. P. 595–601.

29. Вальд А. Последовательный анализ. — М.: Физматгиз, 1960.

30. Вальд А. Статистические решающие функции // Позиционные игры. — М.: Наука, 1967. С. 300–522.

31. Ван-дер-Варден Б. Математическая статистика. — М.: ИЛ, 1960.

32. Вейс Л. (Weiss L.) The asymptotic power of certain tests of fit based on sample spacings // Ann. Math. Statist. 1957. V. 28, № 3. P. 783–786.

33. Вольфовиц Дж. (Wolfowitz J.) On Wald's proof of the consistency of the maximum likelihood estimate // Ann. Math. Statist. 1949. V. 20. P. 601–602.

34. Гаек Я., Шидак З. Теория ранговых критериев. — М.: Наука, 1971.

35. Гихман И.И., Скороход А.В. Введение в теорию случайных процессов. — М.: Наука, 1977.

36. Гош М. (Ghosh M.N.) An extension on Wald's decision theory to unbounded weight functions // Sankhyā. 1952. V. 12. P. 8–26.

37. Гренандер У. Случайные процессы и статистические выводы. — М.: ИЛ, 1961.

38. Гренандер У. (Grenander U.) Abstract inference. — N.Y.: J. Wiley, 1981.

39. Гусев С.И. Асимптотические разложения, связанные с некоторыми статистическими оценками в гладком случае. II // Теория вероятностей и ее применения. 1976. Т. 21, № 1. С. 16–33.

40. Дакуна-Кастелли Д., Дюфло М. (Dacunha-Castelle D., Duflo M.) Probabilités et statistique. Т. 1, 2. — Paris: Masson, 1982, 1983.

41. Дарховский Б.С., Бродский Б.Е. Непараметрический метод скорейшего обнаружения изменений среднего случайной последовательностью // Теория вероятностей и ее применение. 1987. Т. 32, № 4. С. 703–711.

42. Дейвид Г. Порядковые статистики. — М.: Наука, 1979.

43. Де Хардт Дж. (De Hardt J.) Generalizations of the Glivenko–Cantelli theorem // Ann. Math. Statist. 1971. V. 42. P. 2050–2055.

44. Донскер М. (Donsker M.) Justifications and extensions of Doob's heuristic approach to the Kolmogorov–Smirnov theorems // Ann. Math. Statist. 1952. V. 23. P. 277–281.

45. Дуб Дж. (Doob J.L.) Probability and statistics // Trans. Amer. Math. Soc. 1934. V. 36, № 4. P. 759–775.

46. Дуб Дж. Вероятностные процессы. — М.: ИЛ, 1956.

47. Ефрон Б. (Efron B.) Bootstrap methods: another look at the jacknife // Ann. Statist. 1979. V. 7, № 1. P. 1–26.

48. Закс Ш. Теория статистических выводов. — М.: Мир, 1975.

49. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. — М.: Наука, 1979.
50. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.
51. Кифер Дж. (Kiefer J.) On minimum variance estimators // Ann. Math. Statist. 1952. V. 23. P. 627–629.
52. Кифер Дж. (Kiefer J.) On large deviations of the empiric D. C. of vector chance variables and a law of the iterated logarithm // Pacific J. Math. 1961. V. 11, № 2. P. 649–660.
53. Кличене Н., Телькснис Л. Методы обнаружения момента изменения свойств случайных последовательностей // Автоматика и телемеханика. 1983. № 10. С. 5–56.
54. Кокс Д.Р., Хинкли Д.В. Теоретическая статистика. — М.: Мир, 1978.
55. Колмогоров А.Н. Несмещенные оценки // Изв. АН СССР. Сер. мат. 1950. Вып. 14, № 4. С. 303–326.
56. Колмогоров А.Н. Основные понятия теории вероятностей. — М.: Наука, 1974.
57. Крамер Г. (Cramer H.) A contribution to the theory of statistical estimation // Aktuariestidskrift. 1946. V. 29. P. 458–463.
58. Крамер Г. Математические методы статистики. — М.: Мир, 1975.
59. Ксерго С., Дехувельс П., Масон Д. (Csörgõ S., Deheuvels P., Mason D.M.) Kernel estimates of the tail index of a distribution // Ann. Statist. 1985. V. 13, № 3. P. 1050–1077.
60. Кульбак С., Лейблер Р. (Kullback S., Leibler R.A.) On information and sufficiency // Ann. Math. Statist. 1951. V. 22. P. 79–86.
61. Кутоянц Ю.А. Оценивание параметров случайных процессов. — Ереван: Изд-во АН Арм. ССР, 1980.
62. Ланкастер Г. (Lancaster H.) The chi-squared distribution. — N.Y.: J. Wiley, 1969.
63. Ле Кам Л. (Le Cam L.) An extension of Wald's theory of statistical decision functions // Ann. Math. Statist. 1955. V. 26. P. 69–81.
64. Ле Кам Л. (Le Cam L.) Asymptotic methods in statistical decision theory. — N.Y.; Berlin: Springer-Verl., 1986.
65. Леман Э.Проверка статистических гипотез. — М.: Наука, 1979.
66. Леман Э. (Lehmann E. L.) Testing statistical hypotheses. Second Edition. — N.Y.: J. Wiley, 1986.
67. Леман Э. Теория точечного оценивания. — М.: Наука, 1991.
68. Леман Э., Шеффе Г. (Lehmann E.L., Scheffe H.) Completeness, similar regions and unbiased estimation. I // Sankhyā. Ser A. 1950. V. 10. P. 305–340.
69. Линдли Д. (Lindley D.) The use of prior probability disributions in statistical inference and decision // Proc. 4-th Berkley sympos. math. statist. prob., Berkeley; Los Angeles, 1960. V. 1. P. 453–468.
70. Липцер Р.Ш., Ширяев А.Н. Статистика случайных процессов. — М.: Наука, 1974.
71. Лорден Г. (Lorden G.) Procedures for reacting to a change in distribution // Ann. Math. Stat. 1971. V. 42, № 6. P. 1897–1908.
72. Лоэв М. Теория вероятностей. — М.: ИЛ, 1962.
73. Мак Кинси Дж. Введение в теорию игр. — М.: Физматгиз, 1960.
74. Манн Г., Уитни Д. (Mann H.B., Whitney D.R.) On a test whether one of two random variables is stochastically larger than the other // Ann. Math. Statist. 1947. V. 18. P. 50–60.
75. Мизес Р. (Mises R. von) On the asymptotic distribution of differentiable statistical functions // Ann. Math. Statist. 1947. V. 18. P. 309–348.
76. Мизес Р. (Mises R. von) Mathematical theory of probability and statistics. — N.Y.: Acad. Press, 1964.

77. Миллар П. (Millar P.) The minimax principle in asymptotic statistical theory // Lecture Notes in Math. 1981. V. 976. P. 76–267.
78. Миллер Р. (Miller R.) The jacknife: a reviev // Biometrica. 1974. V. 61. P. 1–16.
79. Монстакидес Г. (Monstakides G.V.) Optimal stopping times for detecting changes in distributions // Ann. Stat. 1986. V. 14, № 4. P. 1379–1387.
80. Моран П. (Moran P.A.P.) The random division of an interval // J. Roy. Statist. Soc. Ser. A. 1947. V. 9. P. 92–98.
81. Нейман Дж. фон, Моргенштерн О. Теория игр и экономическое поведение. — М.: Наука, 1970.
82. Нейман Ю. (Neyman J.) Sur un theorems concerente le cosidette statistische sufficienti // Inst. Ital. Atti. Giorn. 1935. V. 6. P. 320–334.
83. Нейман Ю. Вводный курс теории вероятностей и математической статистики. — М.: Наука, 1968.
84. Нейман Ю., Пирсон Э. (Neyman J., Pearson E.S.) On the use and interpretation of certain test criteria // Biometrika. Ser. A. 1928. V. 20. P. 175–240, 263–294.
85. Нейман Ю., Пирсон Э. (Neyman J., Pearson E.S.) On the problem of the most efficient tests of statistical hypotheses // Phil. Trans. Roy. Soc. Ser. A. 1933. V. 231. P. 289–337.
86. Нейман Ю., Пирсон Э. (Neyman J., Pearson E.S.) The testing of statistical hypothess in relation to probabilities a priori // Proc. Camb. Phil. Soc. 1933. V. 34. P. 492–510.
87. Новак С.Ю., Утев С.А. Об асимптотике распределения отношения сумм случайных величин // Сиб. мат. журн. 1990. Т. 31, № 5. С. 92–101.
88. Оостерхов Дж., Цвет В. ван (Oosterhoff J.W.R., van Zwet W.R.) The likelihood ratio test for rhe multinomial distribution // Proc. 6-th Berkley sympos. math. statist. prob., Berkeley; Los Angeles, 1970. V. 1. P. 31–50.
89. Парзен Е. (Parzen E.) On estimation of a probability density function and mode // Ann. Math. Statist. 1962. V. 33, № 3. P. 1065–1076.
90. Пейдж Е. (Page E.S.) Continuous inspection schemes // Biometrica. 1954. V. 41, № 1. P. 100–115.
91. Питмен Е. (Pitman E.J.G.) The estimation of the location and scale parameters of a continuous population of any given form // Biometrika. 1938. V. 30. P. 391–421.
92. Поллак М. (Pollak M.) Optimal defection of a change in distribution // Ann. Statist. 1985. V. 13, № 1. P. 206–227.
93. Поллак М. (Pollak M.) Average run length of an optimal method of defecting a change in distribution // Ann. Statist. 1987. V. 15, № 2. P. 749–779.
94. Рао С. (Rao C.R.) Information and accuracy attainable in estimation of statistical parameters // Bull. Calcutta Math. Soc. 1945. V. 37. P. 81–91.
95. Рао С. (Rao C.R.) Sufficient statistics and minimum variance estimates // Proc. Camb. Phil. Soc. 1949. V. 45. P. 213–218.
96. Рао С. Линейные статистические методы и их применения. — М.: Наука, 1968.
97. Робертс С. (Roberts S.W.) A comparison of control chart, procedures // Technometrics. 1966. V. 8. P. 411–430.
98. Розенблатт М. (Rosenblatt M.) Remarks on some nonparametric estimates of a density function // Ann. Math. Statist. 1956. V. 27, № 3. P. 832–837.
99. Розенблатт М. (Rosenblatt M.) Curve estimation // Ann. Math. Statist. 1971. V. 42, № 6. P. 1815–1842.
100. Русас Дж. Контигуальность вероятностных мер. — М.: Мир, 1975.
101. Себер Дж. Линейный регрессионный анализ. — М.: Мир, 1980.
102. Серфлинг Р. (Serfling R.J.) Approximation theorems of mathematical statistics. — N.Y.: J. Wiley, 1980.

103. Сидоров В.А. (Sidorov V. A.) Measurement of the $\varphi \to \pi^+\pi^-$ branching ration // Phys. Lett. B. 1981. V. 99. P. 62–65.

104. Скороход А.В. Случайные процессы с независимыми приращениями. — М.: Наука, 1964.

105. Смирнов Н.В. О распределении ω^2 Мизеса // Смирнов Н.В. Теория вероятностей и математическая статистика. Избранные труды. — М.: Наука, 1970.

106. Теребиж В.Ю. Максимально правдоподобное восстановление изображений. I. Основные соотношения // Астрофизика. 1990. Т. 32, № 2. С. 327–339.

107. Тьюки Дж. (Tukey J.W.) Bias and confidence in non-quite large samples // Ann. Math. Statist. 1958. V. 29, № 2. P. 614.

108. Уилкс С. (Wilks S.S.) The large sample distribution of the likelihood ratio for testing composite hypotheses // Ann. Math. Statist. 1938. V. 9. P. 60–62.

109. Уилкс С. Математическая статистика. — М.: Наука, 1967.

110. Фаррел Р. (Farrel R.H.) Weak limits of sequences of Bayes procedures in estimation theory // Proc. of V-th Berkeley sympos. math. statist. prob., Berkeley, 1965. V. 1. P. 83–111.

111. Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, 1967. Т. 1, 2.

112. Фергюсон Т. (Ferguson Th.S.) Mathematical statistics: A decision theoretic approach. — N.Y.; London: Acad. Press, 1968.

113. Филипп А., Руссас Дж. (Philippou A.N., Roussas G.G.) Asymptotic distribution of the likelihood function in the independent not identically distributed case // Ann. Statist. 1973. V. 1, № 3. P. 454–471.

114. Филипп А., Руссас Дж. (Philippou A.N., Roussas G.G.) Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case // Ann. Inst. Statist. Math. 1975. V. 27, № 1. P. 45–55.

115. Филиппова А.А. Теорема Мизеса о предельном поведении функционалов от эмпирических функций распределения и ее статистические применения // Теория вероятностей и ее применения. 1962. Т. 7, № 1. С. 26–60.

116. Фишер Р. (Fischer R.A.) On the mathematical foundations of theoretical statistics // Phil. Trans. Roy. Soc. Ser. A. 1922. V. 222. P. 309–368.

117. Фишер Р. (Fischer R.A.) Theory of statistical estimation // Proc. Camb. Phil. Soc. 1925. V. 22. P. 700–725.

118. Фишер Р. (Fischer R.A.) Inverse probability // Proc. Camb. Phil. Soc. 1930. V. 26. P. 528–535.

119. Фреше М. (Fréchet M.) Sur l'extension de certains evalutions statistiques au cas de petits echantillons // Rev. Inst. Internat. Statist. 1943. V. 11. P. 182–205.

120. Халмош П. Теория меры. — М.: ИЛ, 1953.

121. Халмош П., Сэвидж Л. (Halmos P.R., Savage L.J.) Application of the Radon–Nikodym theorem to the theory of sufficient statistics // Ann. Math. Statist. 1949. V. 20. P. 225–241.

122. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. — М.: Мир, 1989.

123. Хеффдинг У. (Hoeffding W.) Asymptotically optimal tests for multinomial distributions // Ann. Math. Statist. 1965. V. 36, № 2. P. 369–401.

124. Ходжес Дж., Леман Э. (Hodges J., Lehman E.) Some problems in minimax estimation // Ann. Math. Statist. 1955. V. 21, № 2. P. 182–197.

125. Ходли Б. (Hoadley B.) Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case // Ann. Math. Statist. 1971. V. 42, № 6. P. 1977–1991.

126. Хотеллинг Х. (Hotelling H.) The generalization of student's ratio // Ann. Math. Statist. 1931. V. 2. P. 360–378.

127. Хьюбер П. (Huber P.J.) Robust estimation of a lokation parameter // Ann. Math. Statist. 1964. V. 35. P. 73–101.

128. Хьюбер П. (Huber P.J.) Robust statistics: a review // Ann. Math. Statist. 1972. V. 43. P. 1041–1067.

129. Хьюбер П. Робастность в статистике. — М.: Мир, 1984.

130. Ченцов Н.Н. Об оценке неизвестного среднего многомерного нормального распределения // Теория вероятностей и ее применения. 1967. Т. 12, № 4. С. 619–633.

131. Ченцов Н.Н. Статистические решающие правила и оптимальные выводы. — М.: Наука, 1972.

132. Чепмен Д., Роббинс Г. (Chapman D.G., Robbins H.E.) Minimum variance estimation without regularity assumptions // Ann. Math. Statist. 1951. V. 22. P. 581–586.

133. Чибисов Д.М. О критериях согласия, основанных на выборочных промежутках // Теория вероятностей и ее применения. 1961. Т. 6, № 1. С. 354–358.

134. Чибисов Д.М. (Chibisov D.M.) Transition to the limiting process for deriving asymptotically optimal tests // Sankhyā. Ser. A. 1969. V. 31, № 3. P. 241–258.

135. Чибисов Д.М., Гванцеладзе Л.Г. О критериях согласия, основанных на сгруппированных данных // III Советско-японский симпозиум по теории вероятностей. — Ташкент: Фан, 1975. С. 183–185.

136. Шеффе Г. Дисперсионный анализ. — М.: Физматгиз, 1963.

137. Ширяев А.Н. Вероятность. — М.: Наука, 1980.

138. Ширяев А.Н. Статистический последовательный анализ. — М.: Наука, 1976.

139. Шорак Дж., Веллнер Дж. (Shorack G.R., Wellner J.A.) Empirical processes with applicatons to statistics. — N.Y.: J. Wiley, 1986.

140. Эдвардс Р. Функциональный анализ. — М.: Мир, 1969.

СПИСОК ОСНОВНЫХ ОБОЗНАЧЕНИЙ*)

а. б. к. — асимптотически байесовский критерий, 400, 406

а. н. оценка — асимптотически нормальная оценка, 85

а. р. н. м. к. — асимптотически равномерно наиболее мощный критерий, 406

а. э. оценка — асимптотически эффективная оценка, 131

а.R-э. оценка — асимптотически R-эффективная оценка, 191

к. о. п. — критерий отношения правдоподобия, 375

н. м. к — наиболее мощный критерий, 303, 310

о. м. п. — оценка максимального правдоподобия, 103

р. н. м. к. — равномерно наиболее мощный критерий, 332

у. м. о. — условное математическое ожидание, 140

(A_0) — условие взаимнооднозначного соответствия между параметрическим множеством Θ и семейством распределений $\mathcal{P} = \{\mathbf{P}_\theta\}_{\theta\in\Theta}$ ($\mathbf{P}_{\theta_1} \neq \mathbf{P}_{\theta_2}$, если $\theta_1 \neq \theta_2$), 100

(A_c) — условие, состоящее в том, что параметрическое множество Θ компактно, 233

(A_μ) — условие, в силу которого все распределения семейства $\mathcal{P} = \{\mathbf{P}_\theta\}$ доминируются мерой μ (существует плотность $f_\theta = d\mathbf{P}_\theta/d\mu$), 101, 102

b, $b(\theta)$ — смещение, 116, 129

$\mathbf{B}_p$ — полиномиальное распределение (в том числе распределение Бернулли), 82

$\mathfrak{B}$ — σ-алгебра борелевских множеств на прямой $\mathbb{R}$, 24

$\mathfrak{B}_\mathcal{X}$ — σ-алгебра в фазовом пространстве $\mathcal{X}$ (борелевских множеств, если $\mathcal{X} = \mathbb{R}^m$), 21

$C(a,b)$ — пространство непрерывных на $[a,b]$ функций, 45

$D(a,b)$ — пространство функций на $[a,b]$, непрерывных слева (в точке a справа) и имеющих лишь конечное число скачков, 45

$\mathbf{D}_\theta$ — дисперсия по распределению $\mathbf{P}_\theta$, 128

E — единичная матрица, 195

$\mathcal{E}$ — экспоненциальное семейство распределений, 178

$\mathbf{E}_\theta$ — математические ожидание по распределению $\mathbf{P}_\theta$, 128

$\mathbf{E}(\xi/\mathfrak{A})$ — условное математическое ожидание ξ относительно σ-алгебры $\mathfrak{A}$, 141

$\mathbf{E}(\xi/\eta)$ — условное математическое ожидание ξ относительно случайной величины η, 143

$f_\theta(x)$ — плотность распределения $\mathbf{P}_\theta$ относительно меры μ, 102

$f_\theta(X)$ — функция правдоподобия, равная по определению $\prod_{i=1}^{n} f_\theta(\mathrm{x}_i)$, 105

$F(x)$ — как правило, функция, распределения, соответствующая распределению $\mathbf{P}$, 22

$F_n^*(x)$ — эмпирическая функция распределения, 25, 34

$F_X^*(x)$ — эмпирическая функция распределения, соответствующая выборке X, 467

$\mathbf{F}_{k_1,k_2}$ — распределение, ассоциированное с распределением Фишера, 74

G — группа преобразований $\mathcal{X}$ в себя, соответствующая инвариантному семейству, 215

h_ε — квантиль распределения χ^2, 296, 400

H_i — гипотеза, 301

$\mathbf{H}_k$ — распределение χ^2, 73

$I(\theta)$ — информация Фишера, 186, 200, 201

I_A — индикатор множества A, 141

$I(\theta) = \|I_{ij}(\theta)\|$, $I_{ij}(\theta) = \mathbf{E}_\theta \frac{\partial}{\partial\theta_i} l(\mathrm{x}_1,\theta) \frac{\partial}{\partial\theta_j} \times$ $\times\, l(\mathrm{x}_1,\theta)$ — информационная матрица Фишера, 204

$\mathbf{I}_x$ — распределение, сосредоточенное в точке x, 24

K_b — класс оценок со смещением $b = b(\theta)$, 131

K_0 — класс несмещенных оценок, 130, 139

$\widetilde{K}_0$ — класс асимптотически несмещенных оценок, 192

$\widetilde{K}^0$ — класс асимптотически центральных оценок, 250

$K_{\Phi,2}$ — класс асимптотически нормальных оценок θ^*, для которых $\mathbf{E}_\theta n(\theta^* - \theta)^2 \to$ $\to\, \sigma^2(\theta)$, где $\sigma^2(\theta)$ — дисперсия предельного для $\sqrt{n}\,(\theta^* - \theta)$ нормального распределения, 137

---

*) Обозначения расположены в алфавитном порядке: сначала русский алфавит, затем латинский, готический и греческий. В конце помещены математические символы.

K_ε — (в гл. 3) класс критериев размера ε (уровня $1-\varepsilon$), 312, 330

$\check{K}_\varepsilon$ — класс несмещенных критериев размера ε, 345

$\widetilde{K}_\varepsilon$ — класс критериев асимптотического уровня $(1-\varepsilon)$, 400

$K_\varepsilon^{Q_1}$ — класс критериев размера ε для частично байесовского подхода, 364

$\widetilde{K}_\varepsilon^{Q_1}$ — класс критериев асимптотического размера ε для частично байесовского подхода, 406

$K_{\alpha_1,\dots,\alpha_{r-1}}$ — класс критериев с фиксированными значениями α_i вероятностей ошибок i-го рода, $i=1,\dots,r-1$, 303

$\mathbf{K}_{\alpha,\sigma}$ — распределение Коши, 80

$l(x,\theta)=\ln f_\theta(x)$, 105

$L(X,\theta)=\ln f_\theta(X)$ — логарифмическая функция правдоподобия, 105

$\mathbf{L}_{\alpha,\sigma^2}$ — логнормальное распределение, 80

n — объем выборки, 22

N_P, N_F — носитель распределения $\mathbf{P}$ с функцией распределения F, 29, 228

$\mathcal{P}$ — семейство распределений, 70

$\mathbf{P}$ — символ распределения, употребляемый в разных смыслах, указанных на с. 22, 23

$\mathbf{P}_n^*$ — эмпирическое распределение, 24

$\mathbf{P}_X^*$ — эмпирическое распределение, соответствующее выборке X, 467

$\mathbf{P}_\theta$ — распределение, зависящее от параметра θ, 81

$P(B/y)$ — условное распределение, 146

$\mathbf{Q}$ — рандомизированная стратегия [[природы]] (априорное распределение θ), 578

$\mathbf{Q}_X$ — апостериорное распределение, 150, 304

$\overline{\mathbf{Q}}$ — наихудшее распределение θ (минимаксная стратегия «природы»), 581

$q(t/X)$ — плотность апостериорного распределения, 150

$\mathbb{R}$ — вещественная прямая, 21

$\mathbb{R}^m$ — m-мерное евклидово пространство, 21

R-э. оценка — R-эффективная (регулярно-эффективная) оценка, 191

(R) — условие регулярности параметрического семейства, в силу которого функция $\sqrt{f_\theta(x)}$ непрерывно дифференцируема по θ, а информация Фишера положительна и непрерывна, 186, 194

(RR) — условия регулярности параметрического семейства, требующие выполнения условий (A_0), (A_c), (R), а также непрерывной дифференцируемости второго порядка функции $l(x,\theta)$ и существования мажоранты $l(x)\geqslant|l''(x,t)|$, для которой интеграл $\mathbf{E}_\theta l(\mathrm{x}_1)$ сходится равномерно в Θ, 247, 271

$S=S(X)$ — статистика, 29

S^2 — эмпирическая дисперсия, 28

$S_0^2=\dfrac{n}{n-1}S^2$, 119

$\mathbf{T}_k$ — распределение Стьюдента, 75

$u^*=\sqrt{n}\,(\widehat{\theta}^*-\theta)$ — нормированная оценка максимального правдоподобия, 253, 282

$\mathbf{U}_{a,b}$ — равномерное распределение на $[a,b]$, 78

$w(t)$ — (не везде) винеровский процесс, 44

$w^0(t)$ — броуновский мост, 44

$w^n(t)$ — эмпирический процесс, 44

x_i — элемент выборки, 22

$\mathrm{x}_{(i)}$ — i-й элемент вариационного ряда, 25

$\boldsymbol{x}=(x_1,\dots,x_n)$ — элемент $\mathcal{X}^n$, 23

$\overline{\mathrm{x}}$ — эмпирическое среднее, 28

$X=X_n=(\mathrm{x}_1,\dots,\mathrm{x}_n)$ — выборка объема n, 22

$[X_\infty]_n=X_n$ — часть бесконечной выборки, состоящая из первых n ее элементов, 24

$\mathcal{X}$ — пространство, которому принадлежат наблюдения (фазовое пространство выборки), 21

$(\mathcal{X},\mathfrak{B}_\mathcal{X},\mathbf{P})$ — выборочное вероятностное пространство, соответствующее одному наблюдению, 21

$(\mathcal{X}^n,\mathfrak{B}_\mathcal{X}^n,\mathbf{P})$ — выборочное вероятностное пространство, соответствующее выборке объема n, 23

$\alpha_i(\pi)$ — вероятность ошибки i-го рода критерия π, 306

$\beta(\delta)$ — мощность критерия δ, 311

$\beta_\pi(\theta)$ — функция мощности критерия π, 330

$\mathbf{B}_{\lambda_1,\lambda_2}$ — бэта-распределение, 77

$\mathbf{\Gamma}_{\alpha,\lambda}$ — гамма-распределение, 72

δ — стратегия I игрока, 576

$\delta=\delta(X)$ — (в гл. 3) решающее правило (критерий), 302

$\delta=\delta(X)$ — решающая функция, 588

ζ_p — квантиль порядка p, 28

ζ_p^* — выборочная квантиль порядка p, 29

θ — параметр (стратегия «природы»), 69, 81, 576

θ^* — оценка параметра θ, 70

θ_Q^* — байесовская оценка параметра θ, соответствующая априорному распределению $\mathbf{Q}$, 150

$\theta^\pm$ — границы доверительного интервала для параметра θ, 285

$\overline{\theta}^*$ — минимаксная оценка параметра θ, 151

$\widehat{\theta}^*$ — оценка максимального правдоподобия параметра θ, 103

Θ — множество возможных значений параметра θ (пространство стратегий «природы»), 70, 81, 576

Θ^* — доверительное множество, 295

λ_ε — квантиль нормального распределения, 469

$\pi = \pi(X)$ — (в гл. 3) рандомизированный критерий, 306, 329

$\pi = \pi(X)$ — рандомизированное решающее правило (критерий), 590

π^0 — равномерно наиболее мощный критерий, 332

π_Q — байесовский критерий, соответствующий априорному распределению $\mathbf{Q}$, 332

$\pi_{Q_1Q_2}$ — байесовский критерий для частично байесовского подхода, 333

$\overline{\pi}$ — минимаксный критерий, 308, 333

$\widehat{\pi}$ — критерий отношения правдоподобия, 375

$\mathbf{\Pi}_\lambda$ — распределение Пуассона, 81

$\mathbf{\Phi}_{\alpha,\sigma^2}$ — нормальное распределение, 71

$\Phi(x)$ — функция стандартного нормального распределения, 74

$\underset{d}{=}$ — знак, означающий совпадение распределений выборок или случайных величин, 22

$\underset{\text{п.н.}}{\longrightarrow}$ — знак сходимости почти наверное (с вероятностью 1), 25

$\underset{P}{\longrightarrow}$ — знак сходимости по вероятности, 35

$\Rightarrow$ — знак слабой сходимости распределений (употребляется как между случайными величинами, так и между распределениями, 36

$\Subset$ — знак, употребляемый между обозначениями выборки (или случайной величины) и распределения и означающий, что выборка извлечена из данного распределения (случайная величина имеет данное распределение), 22

$\Subset\!\!\Rightarrow$ — знак слабой сходимости. Соотношение $\xi_n \Subset\!\!\Rightarrow \mathbf{P}$ означает, что распределение ξ_n слабо сходится при $n \to \infty$ к $\mathbf{P}$, 36

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Асимптотический доверительный интервал, 287
— подход к сравнению оценок, 119, 125, 131
Асимптотическое доверительное множество, 295

Байесовский подход полный, 332, 363
— — частичный, 332, 364
— принцип, 593
Беренса–Фишера проблема, 465
Близкие гипотезы, 320, 404, 555
Броуновский мост, 44
Бэта-распределение, 77

Вариационный ряд, 25
Вероятность ошибки i-го рода, 302, 306
— — первого рода, 329
Винеровский процесс, 44
Выборка, 22
Выборочная квантиль, 29
— медиана, 28
Выборочные моменты, 28, 35
— характеристики, 28, 35
Выборочный коэффициент корреляции, 35

Гамма-распределение, 72
Гипотеза основная, 311, 329
— простая, 301
— сложная, 329
Границы доверительных интервалов, 287
Группировка данных, 427

Доверительное множество, 295, 354
— — инвариантное, 360
— — наиболее точное, 355
— — несмещенное, 359
Доверительный интервал, 285
Достаточная оценка, 172
— статистика, 157
— σ-алгебра, 165

Закон повторного логарифма для эмпирических распределений, 34

Игра двух лиц, 576
— рандомизированная (смешанная), 577
— статистическая, 588
Инвариантное семейство, 349
Интервальное оценивание, 285
Информация Фишера, 186, 200
Источник излучения, 213

Квантиль, 28
Классификация частиц, 264
Критерий, 302
— асимптотически байесовский, 400, 407
— — минимаксный, 407
— — наиболее мощный, 322
— — несмещенный, 403
— — равномерно наиболее мощный, 406
— — эквивалентный, 322, 408
— асимптотического уровня, 316, 399
— байесовский, 304, 332, 364
— Вилкоксона, 470
— знаков, 393, 469
— инвариантный, 349
— Колмогорова, 390
— Колмогорова–Смирнова, 467
— минимаксный, 308, 333
— Морана, 393
— наиболее мощный, 303, 310
— непараметрический, 393
— несмещенный, 345
— отношения правдоподобия, 375
— проверки однородности, 451, 467
— — — асимптотически минимаксный, 452
— — — непараметрический, 468
— — — состоятельный, 469
— — — χ^2, 475
— равномерно наиболее мощный, 332, 333
— состоятельный, 389
— χ^2, 423
— ω^2 (Мизеса–Смирнова), 392
Критическая область, 311

Лемма Неймана–Пирсона, 312
— — — обобщенная, 341
Логарифмическая функция правдоподобия, 105

Максимальный инвариант, 352
Медиана, 28
Метод максимального правдоподобия, 101
— минимального расстояния, 99
— моментов, 87
— подстановки, 83
Минимальная достаточная σ-алгебра, 165
Монотонное отношение правдоподобия, 333
Мощность критерия, 311, 330

Неравенство Иенсена для у. м. о., 144
— Коши–Буняковского для матриц, 196
— — — — у. м. о., 144
— Рао–Крамера, 186
— — — интегральное, 218
— — — разностное, 238
Нормальность асимптотическая, 519, 527
Носитель распределения, 29

Орбита, 217
Отношение правдоподобия, 243, 269
Оценка, 69, 82
— асимптотически байесовская, 156, 226, 607
— — минимаксная, 156, 227, 607
— — нормальная, 85
— — эффективная, 131, 531
— — R-байесовская, 226
— — R-эффективная, 291
— байесовская, 150, 156
— достаточная, 171
— максимального правдоподобия, 103, 529
— минимаксная, 151, 156
— недопустимая, 129
— несмещенная, 116
— параметра масштаба, 207
— — сдвига, 207
— параметров, 504
— Питмена, 209
— плотности, 65
— подстановки, 83
— сильно состоятельная, 85
— состоятельная, 85
— суперэффективная, 219
— эквивариантная, 208, 217
— эффективная, 130, 138, 531
— R-эффективная, 191
Оценки по методу моментов, 509

Плотность апостериорная, 150
— априорная, 150
— условная, 147
Полное семейство распределений, 174
Полный класс стратегий, 587
Последовательное оценивание, 284
Последовательный анализ, 378
— критерий, 379
Приближенное вычисление о.м.п., 261
Принцип достаточности, 207
— инвариантности, 207
— несмещенности, 207, 345
Проверка однородности, 450
Пуассоновский процесс, 40

Равномерная сходимость интеграла, 260
— — по вероятности, 278
— — распределений, 279
— центральная предельная теорема, 280
Равномерный закон больших чисел, 278
Размер критерия, 311, 329
Распределение, 21
Распределение апостериорное, 150, 304
— априорное, 150, 304
— Бернулли, 81
— вырожденное, 81
— Коши, 80
— логнормальное, 80
— наименее благоприятное (наихудшее), 309, 344, 365, 584
— наихудшее, 152
— нормальное, 71
— полиномиальное, 81
— Пуассона, 81
— равномерное, 78
— Стьюдента, 75
— Фишера, 74
— экспоненциальное, 74
— χ^2 (хи-квадрат), 73
Расстояние Кульбака–Лейблера, 229, 316
— Хеллингера, 229
— χ^2, 229
Регрессия, 483
— линейная, 476
Регрессор, 476
Решающая функция (решающее правило, решение), 588
— — асимптотически байесовская, 619, 622
— — асимптотически минимаксная, 622
— — инвариантная, 603
— — несмещенная, 601
— — рандомизированная, 590
Риск (функция риска), 588
Робастность, 439

Смещение, 116
Среднеквадратический подход к сравнению оценок, 116, 125
Статистика, 29
— достаточная, 157
— минимальная достаточная, 165
— непараметрическая, 63
— полная, 174
— тривиальная достаточная, 164
— χ^2, 49, 53
Статистики I и II типов, 29, 30
Стратегия, 576
— байесовская, 578
— минимаксная, 580
— равномерно оптимальная, 577
— рандомизированная (смешанная), 577
— уравнивающая, 583
— чистая, 578, 590
Сужение метода подстановки, 84

Теорема Гливенко–Кантелли, 25, 34
— Неймана–Фишера (факторизационная), 158
Теоремы непрерывности, 35

Уровень доверия, 286
— значимости критерия, 330
Условие (A_0), 100, 101
— (A_c), 233
— (A_μ), 101
— (R), 186, 194
— (RR), 247, 271
Условная вероятность, 143
— плотность, 147
Условное математическое ожидание, 142
— распределение, 146

Устойчивость статистических решений, 438

Фактически достигаемый уровень, 331
Формула Байеса, 150
— полной вероятности, 144
Функционал непрерывно дифференцируемый, 51
Функционалы I и II типов, 29
Функциональная предельная теорема для эмпирических процессов, 45
Функция логарифмическая, 105
— правдоподобия, 105, 524
— — логарифмическая, 524

Цена игры, 580

Эквивариантное оценивание, 215
Экспоненциальное семейство распределений, 178
Эллипсоид рассеивания, 126
Эмпирическая функция распределения, 25, 34
Эмпирический процесс, 44
Эмпирическое распределение, 24, 33
— — сглаженное, 64

Ядерные оценки плотности, 68

Александр Алексеевич БОРОВКОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебник

Издание четвертое, стереотипное

ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.07.953.П.004173.04.07
от 26.04.2007 г., выдан ЦГСЭН в СПб

Издательство «ЛАНЬ»
lan@lpbl.spb.ru; www.lanbook.com
192029, Санкт-Петербург, Общественный пер., 5.
Тел./факс: (812)412-29-35, 412-05-97, 412-92-72.
Бесплатный звонок по России: 8-800-700-40-71

ГДЕ КУПИТЬ

ДЛЯ ОРГАНИЗАЦИЙ:

Для того, чтобы заказать необходимые Вам книги, достаточно обратиться в любую из торговых компаний Издательского Дома «ЛАНЬ» :

по России и зарубежью
«ЛАНЬ-ТРЕЙД». 192029, Санкт-Петербург, ул. Крупской, 13
тел.: (812) 412-85-78, 412-14-45, 412-85-82; тел./факс: (812) 412-54-93
e-mail: trade@lanpbl.spb.ru; ICQ: 446-869-967
www.lanpbl.spb.ru/price.htm

в Москве и в Московской области
«ЛАНЬ-ПРЕСС». 109263, Москва, 7-ая ул. Текстильщиков, д. 6/19
тел.: (499) 178-65-85; e-mail: lanpress@ultimanet.ru

в Краснодаре и в Краснодарском крае
«ЛАНЬ-ЮГ». 350072, Краснодар, ул. Жлобы, д. 1/1
тел.: (8612) 74-10-35; e-mail:lankrd98@mail.ru

ДЛЯ РОЗНИЧНЫХ ПОКУПАТЕЛЕЙ:

интернет-магазины:
«Сова»: http://www.symplex.ru; «Ozon.ru»: http://www.ozon.ru
«Библион»: http://www.biblion.ru
также Вы можете отправить заявку на покупку книги
по адресу: 192029, Санкт-Петербург, ул. Крупской, 13

Подписано в печать 20.10.09.
Бумага офсетная. Гарнитура Школьная. Формат 70 ×100 $^{1}/_{16}$.
Печать офсетная. Усл. п. л. 57,20. Тираж 1500 экз.

Заказ № .

Отпечатано в полном соответствии с качеством
предоставленных материалов в ОАО «Дом печати — ВЯТКА»
610033, г. Киров, ул. Московская, 122