Московский государственный университет имени М.В. Ломоносова

Ю.В. Прохоров, Л.С. Пономаренко

ЛЕКЦИИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

2-е издание, исправленное и дополненное

Допущено УМО по классическому университетскому образованию в качестве учебника для студентов высших учебных заведений, обучающихся по направлениям ВПО 010400 «Прикладная математика и информатика» и 010300 «Фундаментальная информатика и информационные технологии»

Издательство
Московского университета
2012

УДК 519.2
ББК 22.171; 22.172
П84

Публикуется по решению редакционно-издательского совета Московского государственного университета имени М. В. Ломоносова

Рецензенты:
доктор физико-математических наук, профессор В. В. Сенатов,
доктор физико-математических наук, профессор Д. М. Чибисов

Прохоров Ю.В., Пономаренко Л.С.
П84 Лекции по теории вероятностей и математической статистике: Учебник. — 2-е изд., испр. и доп. — М.: Издательство Московского университета, 2012. — 256 с. — (Классический университетский учебник).
ISBN 978-5-211-06234-4

Учебник основан на материале годового курса лекций по теории вероятностей и математической статистике, который много лет читался студентам второго курса факультета вычислительной математики и кибернетики МГУ.

Изложение учебного материала начинается со случая конечных вероятностных пространств, что дает возможность доказывать содержательные теоремы сравнительно простыми средствами. Далее излагаются общие основы теории вероятностей, рассматриваются предельные теоремы, сходимости последовательностей и рядов из случайных величин. Последние главы посвящены задачам математической статистики.

Особое внимание уделяется оценкам вероятностей в виде приближенных формул или в виде неравенств. Учебник содержит много примеров, иллюстрирующих основные понятия теории вероятностей и математической статистики.

Рекомендуется для студентов, обучающихся по специальностям «Прикладная математика и информатика», «Фундаментальная информатика и информационные технологии».

Ключевые слова: вероятность случайного события; независимость случайных величин и случайных событий; математическое ожидание и дисперсия; закон больших чисел; характеристические функции; предельные теоремы; доверительный интервал; несмещенные, состоятельные, эффективные оценки; проверка статистических гипотез.

УДК 519.2
ББК 22.171; 22.172

ISBN 978-5-211-06234-4

Предисловие

Уважаемый читатель!

Вы открыли одну из замечательных книг, изданных в серии «Классический университетский учебник», посвященной 250-летию Московского университета. Серия включает свыше 150 учебников и учебных пособий, рекомендованных к изданию Учеными советами факультетов, редакционным советом серии и издаваемых к юбилею по решению Ученого совета МГУ.

Московский университет всегда славился своими профессорами и преподавателями, воспитавшими не одно поколение студентов, впоследствии внесших заметный вклад в развитие нашей страны, составивших гордость отечественной и мировой науки, культуры и образования.

Высокий уровень образования, которое дает Московский университет, в первую очередь обеспечивается высоким уровнем написанных выдающимися учеными и педагогами учебников и учебных пособий, в которых сочетаются как глубина, так и доступность излагаемого материала. Аккумулируемый в этих книгах бесценный опыт методики и методологии преподавания становится достоянием не только Московского университета, но и других университетов России и всего мира.

Издание серии «Классический университетский учебник» наглядно демонстрирует тот вклад, который вносит Московский университет в классическое университетское образование в нашей стране, и, несомненно, служит его развитию.

Решение этой благородной задачи было бы невозможным без активной помощи со стороны издательств, принявших участие в выпуске книг серии «Классический университетский учебник». Мы расцениваем это как поддержку ими позиции, которую занимает Московский университет в вопросах науки и образования. Это служит также свидетельством того, что 250-летний юбилей Московского университета — выдающееся событие в жизни всей нашей страны, мирового образовательного сообщества.

Ректор Московского университета
академик РАН, профессор

В. Садовничий

Введение

Настоящее учебное пособие основано на материале лекций, которые много лет читались студентам второго курса факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова.

Вашему вниманию предлагается краткий вариант. В отдельные годы он дополнялся и другими вопросами, и представление о полной программе этого двухсеместрового курса можно получить из приведенного в приложении 2 списка экзаменационных вопросов.

Особое внимание обращается на оценки вероятностей либо в виде приближенных формул, либо в виде неравенств. Это вполне соответствует классической традиции, когда в названиях книг не было слов «теория вероятностей», а употреблялось выражение «исчисление вероятностей». В качестве примеров можно привести учебники Пуанкаре [35], Маркова [22].

Ряд теорем приводится без доказательства. Мы считаем, что студентам нужно в первую очередь уметь пользоваться этими теоремами. Источники, в которых при желании можно найти доказательства данных теорем, обычно указываются.

При изложении вопросов математической статистики произведен жесткий отбор материала. Часто из учебников большого объема математическая статистика представляется как цепочка определений, лемм и теорем без применения этих теорем к обработке статистического материала. Примером удачного соединения теории и практики может служить книга Г. Крамера «Математические методы статистики» [17].

Мы начинаем изложение со случая конечных вероятностных пространств, поскольку к началу третьего семестра студенты не имеют еще достаточных знаний по математическому анализу. Это позволяет иметь дело только с конечными суммами, все функции от элементарных исходов являются случайными величинами, математические ожидания и дисперсии существуют, и можно продвинуться с этими средствами достаточно далеко, например доказать закон больших чисел и даже центральную предельную теорему.

Ряд экзаменационных вопросов не освещен в этой книге (в частности, цепи Маркова). Мы исправим этот недостаток во втором томе данного курса лекций. Мы надеемся в будущем дополнить эту книгу новыми главами, которые могут быть полезны не только студентам, интересующимся теорией вероятностей, но и начинающим преподавателям этой дисциплины.

Авторы благодарят рецензентов профессора В. В. Сенатова и профессора Д. М. Чибисова за полезные замечания и поправки.

Мы будем благодарны всем читателям, которые пожелают сообщить нам свои замечания.

...Теория вероятностей есть в сущности не что иное, как здравый смысл, сведенный к исчислению.

Пьер-Симон Лаплас

Глава 1

ВЕРОЯТНОСТНОЕ ПРОСТРАНСТВО

«Теория вероятностей — математическая наука, позволяющая по *вероятностям* одних случайных событий находить *вероятности* других случайных событий, связанных каким-либо способом с первыми» — такое определение приводится в математической энциклопедии[1].

Поскольку теория вероятностей дает возможность количественно оценивать случайные события, случайные явления, то методы этой математической дисциплины находят широкое применение в различных областях науки и человеческой деятельности. К выводам теории вероятностей и возможности использовать их в практической деятельности надо относиться разумно.

«Утверждение о том, что какое-либо событие наступает с вероятностью, равной, например, 1/2, еще не представляет само по себе окончательной ценности, так как мы стремимся к *достоверному знанию*. Окончательную познавательную ценность имеют те результаты теории вероятностей, которые позволяют утверждать, что вероятность наступления какого-либо события A весьма близка к единице или (что то же самое) вероятность ненаступления события A весьма мала. В соответствии с принципом "пренебрежения достаточно малыми вероятностями" такое событие справедливо считают *практически достоверным*»[2].

О возникновении теории вероятностей и ее становлении как науки можно прочитать, например, в очерке А. Н. Ширяева[3]. Мы же отметим только, что в развитие теории вероятностей огромный вклад внесли российские и советские математики: П. Л. Чебышёв (1821–1894), А. А. Марков (1856–1922), А. М. Ляпунов (1857–1918), С. Н. Бернштейн (1880–1968), А. Н. Колмогоров (1903–1987) и др.

Знакомство с теорией вероятностей начнем с основного ее понятия — понятия вероятностного пространства.

---

[1] *Прохоров Ю.В., Севастьянов Б.А.* [23, т. 1, с. 655–665].

[2] Там же, с. 655.

[3] *Ширяев А.Н.* Математическая теория вероятностей. Очерк истории становления [14, с. 101–129].

Вероятностным пространством называется тройка элементов

$$(\Omega, \mathcal{A}, \mathbf{P}),$$

в которой Ω — некоторое множество, состоящее из точек ω, $\mathcal{A}$ — некоторый класс подмножеств Ω, называемых *случайными событиями*, $\mathbf{P}$ — *распределение вероятностей*, или *вероятность* случайных событий.

Разберем все эти понятия сначала для конечного вероятностного пространства.

1.1. Конечное вероятностное пространство

Пусть Ω — некоторое конечное множество, состоящее из s элементов $\omega_1, \ldots, \omega_s$, называемых элементарными исходами или элементарными событиями.

Обычно это множество Ω выбирается в соответствии с проводимым случайным экспериментом таким образом, чтобы всякий мыслимый неразложимый исход эксперимента описывался единственным элементарным исходом.

Событием в этом случае будем называть любое подмножество множества Ω. Таким образом, $\mathcal{A}$ будет состоять из 2^s событий (подмножеств), включая подмножество, совпадающее с Ω, и пустое подмножество $\varnothing$. Событие Ω называют достоверным, а $\varnothing$ — невозможным событием.

Распределение $\mathbf{P}$ для конечных вероятностных пространств задается следующим образом: каждому элементарному исходу ω_i ставится в соответствие число $\mathbf{P}(\omega_i) = p_i \geqslant 0$ так, чтобы $\sum\limits_{i=1}^{s} p_i = 1$.

Тогда вероятность $\mathbf{P}(A)$ произвольного случайного события $A = \{\omega_{i_1}, \ldots, \omega_{i_r}\}$ определяется следующим образом:

$$\mathbf{P}(A) = \sum_{k=1}^{r} p(\omega_{i_k}) = \sum_{\omega \in A} p(\omega), \quad \mathbf{P}(\varnothing) = 0.$$

Исходы $\omega \in A$ называют *благоприятствующими* для события A, так как наступление любого такого исхода повлечет наступление события A.

Подробнее рассмотрим частный случай конечного вероятностного пространства, которое будем называть классической вероятностной моделью.

1.2. Классическая вероятность

Пусть все s возможных исходов случайного эксперимента равновероятны, тогда в соответствии с изложенным выше $p(\omega_i) = 1/s$ для любого i и

$$\mathbf{P}(A) = \frac{r_A}{s},$$

где s — по-прежнему общее число возможных исходов, а r_A — число исходов, благоприятствующих A.

1.2.1. Генуэзская лотерея

Первое упоминание об этой лотерее в литературе относится к 1620 г. Условия лотереи таковы: проводится розыгрыш 5 выигрышных билетов из 90 имеющихся. Все билеты пронумерованы числами от 1 до 90. Участники розыгрыша могут делать ставки: на 1 номер (простая единица), на 2 номера (амбо), на 3 номера (терно), на 4 номера (кватерно), на 5 номеров (квинтерно).

При этом, если игрок угадывал, он получал свою ставку, умноженную на 15 для простой единицы, на 270 для амбо, на 5500 для терно, на 75 000 для кватерно, на 1 000 000 для квинтерно. Пользуясь моделью классической вероятности, найдем вероятности таких выигрышей.

Для этого введем элементарные исходы как упорядоченные наборы из 5 различных между собой чисел:

$$\omega = (i_1, i_2, i_3, i_4, i_5),$$

где i_j — номер числа, выпавшего в j-м розыгрыше, $1 \leqslant i_j \leqslant 90$.

Подсчет общего числа исходов s проводится следующим образом: для i_1 имеется всего 90 различных вариантов, для i_2 уже на единицу меньше, т. е. 89, для i_3 — 88, для i_4 — 87, для i_5 — 86. Перемножив эти числа, получим общее число исходов

$$s = 90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 = 5\,273\,912\,160.$$

Тогда вероятность любого элементарного исхода ω равна $\mathbf{P}(\omega) = 1/s$. Введем события A_i $(1 \leqslant i \leqslant 5)$ — выиграть при ставке на i номеров. Проведем для примера вычисление $\mathbf{P}(A_2)$.

Пусть 7 и 11 — два числа, на которые делается ставка амбо. Тогда каждый благоприятствующий исход для события A_2 будет иметь вид

$$(*, 7, *, 11, *),$$

где $*$ помечены различные между собой числа, не равные 7 и 11, причем 7 и 11 могут появиться в этой цепочке на любых местах. Таких благоприятствующих исходов будет

$$C_5^2 \cdot 2 \cdot 88 \cdot 87 \cdot 86,$$

где $C_5^2 \cdot 2$ — это число способов выбора двух мест, на которых можно в любом порядке расположить числа 7 и 11, 88 — возможные варианты выбора числа на первое из оставшихся мест, 87 — на второе оставшееся место, 86 — на третье свободное место. Получаем, что

$$\mathbf{P}(A_2) = \frac{C_5^2 \cdot 2 \cdot 88 \cdot 87 \cdot 86}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{2}{801}.$$

Аналогичным образом можно вычислить все вероятности $\mathbf{P}(A_i)$. Значения этих вероятностей приведены в таблице.

Таблица 1.1

Вероятности выигрышей в генуэзской лотерее

i	**Увеличение ставки в m_i раз**	$\mathbf{P}(A_i)$
1	15	$1/18 \approx 0.055555$
2	270	$2/801 \approx 0.002497$
3	5 500	$1/11\,748 \approx 0.0000851$
4	75 000	$1/511\,038 \approx 0.0000019$
5	1 000 000	$1/43\,949\,268 \approx 10^{-8} \cdot 2.275$

Замечание. Значения $\mathbf{P}(A_i)$ можно получить и другим способом, вводя в качестве элементарных исходов неупорядоченные наборы различных между собой 5 чисел. Тогда общее число исходов $s' = C_{90}^5$, а для события A_2 вероятность вычисляется следующим образом:

$$\mathbf{P}(A_2) = \frac{C_2^2 \cdot C_{88}^3}{C_{90}^5} = \frac{2}{801}.$$

Задача 1.1. Найдите вероятности всех событий A_i.

Задача 1.2. Какова вероятность того, что при розыгрыше лотереи номера выйдут в возрастающем или убывающем порядке? (Ответ: 1/60.)

Приведем имеющиеся статистические данные по розыгрышам этой лотереи в Праге. За 133 года, с 1754 по 1886 г., было проведено 2854 тиража, при этом номера выходили в возрастающем или убывающем порядке с частотой 0.01612, тогда как точное значение вероятности такого события 0.01667. Можно ли считать классическую модель, в рамках которой была вычислена эта вероятность, удовлетворительной? Ответ на этот вопрос мы сможем дать, когда познакомимся с основными понятиями теории проверки статистических гипотез во второй части нашего курса.

1.2.2. Игральные кости

Игроки бросают по 6 игральных костей. Ставка игрока составляет 10 копеек. Если сумма выпавших очков 6 или 36, то игрок получает премию 78 рублей, 7 или 35 очков — 24 рубля, 8 или 34 — 4 рубля, в остальных случаях не получает ничего[1].

Проведем вычисление вероятностей данных премий. Для этого занумеруем кости от 1 до 6 и введем элементарный исход как упорядоченный набор из шести чисел

$$\omega = (i_1, i_2, i_3, i_4, i_5, i_6),$$

где $1 \leqslant i_j \leqslant 6$ — результат бросания j-й кости. Всего элементарных исходов $s = 6^6 = 46\,656$.

Обозначим $S(\omega) = i_1 + i_2 + i_3 + i_4 + i_5 + i_6$ сумму выпавших очков. Тогда

$$\mathbf{P}\{S(\omega) = 6\} = \mathbf{P}\{S(\omega) = 36\} = \frac{1}{s},$$

так как единственными благоприятствующими исходами для этих событий являются $(1,1,1,1,1,1)$ и $(6,6,6,6,6,6)$ соответственно. Следовательно, вероятность выиграть 78 рублей равна

$$p_1 = \frac{2}{s} \approx 0.000042867.$$

Аналогичными рассуждениями можно получить, что

$$\mathbf{P}\{S(\omega) = 7\} = \mathbf{P}\{S(\omega) = 35\} = \frac{6}{s}.$$

[1] Игра относится ко второй половине XIX в. и описана П. Л. Чебышевым в лекциях по теории вероятностей, читавшихся им в С.-Петербургском университете в 1879–1880 гг. (см. [32]).

Действительно, для каждого из этих событий имеется уже по 6 благоприятствующих исходов. Например, событие $S(\omega) = 7$ произойдет в любом из 6 случаев

$$\begin{gathered}(2,1,1,1,1,1), \quad (1,2,1,1,1,1), \quad (1,1,2,1,1,1),\\ (1,1,1,2,1,1), \quad (1,1,1,1,2,1), \quad (1,1,1,1,1,2).\end{gathered}$$

Значит, вероятность выиграть 24 рубля равна

$$p_2 = \frac{2\cdot 6}{s} \approx 0.000257202.$$

Подобными рассуждениями, представив всеми возможными способами число $8 = 1+1+1+1+2+2 = \ldots = 1+1+1+1+1+3$, получим, что

$$\mathbf{P}\{S(\omega)=8\} = \mathbf{P}\{S(\omega)=34\} = \frac{C_6^2 + C_6^1}{6^6} = \frac{15+6}{6^6},$$

и вероятность выиграть 4 рубля равна

$$p_3 = \frac{2\cdot 21}{6^6} \approx 0.000900205.$$

Таким образом, вероятность выиграть хоть что-нибудь в этой игре равна

$$p_1 + p_2 + p_3 = 0.001200274,$$

а средний выигрыш равен

$$(78\cdot p_1 + 24\cdot p_2 + 4\cdot p_3)\times 100 = 1.3117284 \text{ (коп.)},$$

что составляет около 1 копейки и почти в 10 раз меньше ставки[1]. Как видим, условия этой игры весьма невыгодны для игроков, ставки должны быть повышены.

Подсчет вероятностей $\mathbf{P}\{S(\omega) = k\}$ был проведен в самых простых случаях, при приближении k к среднему значению 21 вычисления становятся все более громоздкими, и требуется уже некий алгоритм, заменяющий простой перебор.

Разберем способ, использующий бином Ньютона и разложения в ряды для простых функций, применимый при любом числе игральных костей.

[1] Чебышев [32, с. 213] пишет: «...Так что 9/10 ставки шло в пользу откупщика; поэтому такие лотереи во всех государствах упразднены».

Рассмотрим многочлен 36 степени

$$Q(x) = (x + x^2 + x^3 + x^4 + x^5 + x^6)^6 = a_{36}x^{36} + a_{35}x^{35} + \ldots + a_1 x + a_0,$$

$a_0 = 0$, $a_{36} = 1$. К примеру, коэффициент a_8 в этом разложении равен числу способов, которыми можно получить 8, т. е. 21. Это мы вычислили выше. Но можно этот коэффициент получить и по-другому. Используя формулу для суммы геометрической прогрессии, получим, что

$$Q(x) = \left(\frac{x(1-x^6)}{1-x}\right)^6 = \frac{x^6(1-x^6)^6}{(1-x)^6}.$$

Поскольку

$$\frac{1}{(1-x)^6} = (1-x)^{-6} = 1 + \frac{(-6)}{1}(-x) + \frac{(-6)(-7)}{1\cdot 2}(-x)^2 + \\ + \frac{(-6)(-7)(-8)}{1\cdot 2\cdot 3}(-x)^3 + \ldots + \frac{6\cdot 7\cdot \ldots \cdot (6+k-1)x^k}{k!} + \ldots,$$

то

$$Q(x) = x^6(1 - C_6^1 x^6 + C_6^2 x^{12} - \ldots + x^{36})(1 + 6x + 21x^2 + 56x^3 + \ldots) = \\ = x^6(1 - 6x^6 + 15x^{12} + \ldots)(1 + 6x + 21x^2 + \ldots).$$

Таким образом, подсчитывая коэффициенты при соответствующих степенях x, мы получим, в частности, что $a_8 = 21$, а $a_{15} = 1666$. Поскольку число a_k равно числу благоприятствующих исходов для события $\{S(\omega) = k\}$, то

$$\mathbf{P}\{S(\omega) = 8\} = \frac{21}{6^6}; \quad \mathbf{P}\{S(\omega) = 15\} = \frac{1666}{6^6}.$$

1.2.3. Случайные перестановки

Рассмотрим в качестве элементарных исходов ω различные перестановки чисел $1, 2, \ldots, n$. Общее число исходов $s = n!$, каждому исходу приписываем вероятность $1/n!$.

Пусть A_2 — событие, состоящее в том, что число 2 появится на втором месте, тогда

$$\mathbf{P}(A_2) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Какова вероятность того, что хотя бы одно число встретится на своем месте? Обозначим это событие A.

Для небольших n можно выписать все возможные исходы и подсчитать среди них число благоприятствующих. Так, для $n = 3$ имеем 6 исходов:

$$\begin{aligned} &\omega_1 = (1,2,3), \quad \omega_2 = (1,3,2), \quad \omega_3 = (2,1,3), \\ &\omega_4 = (2,3,1), \quad \omega_5 = (3,1,2), \quad \omega_6 = (3,2,1). \end{aligned}$$

Для события A благоприятствующими являются исходы ω_1, ω_2, ω_3, ω_6. Следовательно,

$$\mathbf{P}(A) = \frac{4}{6} = \frac{2}{3}.$$

Такой способ вычисления вероятности события A, конечно, не годится при больших n. В дальнейшем мы сможем показать, что $\mathbf{P}(A) \approx 1 - 1/e$, но для этого потребуется формула для вычисления вероятности объединения конечного числа случайных событий.

1.2.4. Игра в бридж

Рассмотрим колоду, состоящую из 52 карт. Четырем игрокам раздают по 13 карт. Какова вероятность того, что у одного из них, скажем у первого, на руках окажется полная масть?

Обозначим это событие B. Занумеруем карты следующим образом: номерами от 1 до 13 червы, от 14 до 26 — бубны, от 27 до 39 — трефы и от 40 до 52 — пики.

Здесь элементарный исход ω — это перестановка из 52 чисел, описывающая начальное расположение карт перед их раздачей. При этом первому игроку достаются первая, пятая, девятая и т. д. карты из колоды, второму — вторая, шестая и т. д.

Рассмотрим случай, когда первому игроку, например, достанутся все карты червовой масти. Это произойдет, если на местах $1, 5, 9, \ldots$ стоят числа $1, 2, \ldots, 13$ в том или ином порядке, а на остальных местах — все остальные числа также в произвольном порядке. Число таких исходов равно $13! \cdot 39!$. Поскольку этот подсчет верен для любой из четырех мастей, то всего благоприятствующих исходов — $4 \cdot 13! \cdot 39!$. Так как общее число элементарных исходов $s = 52!$, то

$$\mathbf{P}(B) = \frac{4 \cdot 13! \cdot 39!}{52!} \approx 6.3 \cdot 10^{-12}.$$

Хотя это событие и маловероятно, но возможно, что игроки в бридж могли наблюдать его[1].

1.2.5. Абсолютно случайные последовательности

Рассмотрим в качестве элементарных исходов цепочки длины n:

$$\omega = (\delta_1, \delta_2, \ldots, \delta_n),$$

где каждая δ_i $(1 \leqslant i \leqslant n)$ — это либо 0, либо 1, причем появление 0 или 1 на любом месте равновозможно. Всего таких исходов будет $s = 2^n$. Вероятность каждого отдельного исхода равна $\mathbf{P}(\omega) = 2^{-n}$. При больших n эти вероятности очень малы, так, при $n = 100$ вероятность отдельного элементарного исхода $\mathbf{P}(\omega) \approx 10^{-30}$, а при $n = 1000$ имеем $\mathbf{P}(\omega) \approx 10^{-300}$.

Такие последовательности из 0 и 1 называют *абсолютно случайными* последовательностями. Они играют важную роль в теории информации, в частности, используются при кодировании и защите информации.

Сформулируем ряд свойств таких последовательностей в виде задач.

Задача 1.3. Пусть $\omega = (\delta_1, \ldots, \delta_n)$ — абсолютно случайная последовательность длины n, и $1 \leqslant m_1 < m_2 < \ldots < m_k \leqslant n$ — некоторая последовательность длины k. Тогда новая подпоследовательность $\omega' = (\delta_{m_1}, \delta_{m_2}, \ldots, \delta_{m_k})$ — также абсолютно случайная последовательность, но длины k.

Поясним утверждение на примере. Пусть $n = 5$, $m_1 = 1$, $m_2 = 2$ и $\omega' = (0, 0)$. Тогда $\mathbf{P}(\omega') = 1/4$. Действительно, ω' может быть получена из любой из 8 последовательностей:

$$\begin{matrix} (00000), & (00001), & (00010), & (00011), \\ (00100), & (00101), & (00110), & (00111). \end{matrix}$$

Вероятность события A, для которого эти исходы являются благоприятствующими, равна

$$\mathbf{P}(A) = \frac{2^3}{2^5} = \frac{1}{2^2}.$$

Для реализации абсолютно случайных последовательностей

[1] Подробнее о вычислении вероятностей при игре в бридж можно прочитать в [22; 29].

используют либо физические датчики, либо последовательности так называемых псевдослучайных чисел, полученных по определенному вычислительному алгоритму и потому не являющихся случайными в прямом смысле, но обладающих схожими свойствами[1]. Существуют таблицы случайных чисел, самая значительная из них — «Миллион случайных чисел»[2].

Заметим, что если бросать монету и отмечать выпадение герба цифрой 1, а решки — цифрой 0, то получится последовательность нулей и единиц, но обычно полученные таким образом результаты не удовлетворяют критериям случайности. Это связано либо с тем, что монета не является идеально симметричной, либо с особенностями бросающего монету.

Задача 1.4. Пусть новая последовательность ω' длины $n-1$ получена из ω следующим образом: смены 0 и 1 соответствуют в новой последовательности 1, а отсутствие перемены — 0. Скажем, если $\omega = (01101)$, то $\omega' = (1011)$. Тогда ω' также абсолютно случайная последовательность длины $n-1$.

Задача 1.5. Рассмотрим фиксированную последовательность $\varepsilon_0 = (0,0,1,0,1)$ и случайную последовательность $\varepsilon_1 = (\delta_1,\ldots$ $\ldots,\delta_5)$. Сложив их почленно по $\operatorname{mod} 2$, получим последовательность $\varepsilon_2 = (\delta'_1,\ldots,\delta'_5)$. Тогда эта последовательность также абсолютно случайна. Докажите аналогичное утверждение для последовательностей произвольной длины n.

Пример 1.1. Рассмотрим буквы русского алфавита, отождествив предварительно Е с Ё. Их окажется 32. Закодируем их, ставя в соответствие каждой букве ее порядковый номер минус 1, записанный в двоичной системе и дополненный нулями, чтобы получилась цепочка длины 5: А(00000), Б(00001), В(00010), ..., Я(11111). Если взять произвольный текст, то из страницы текста получится некоторая последовательность ε_0. Сложим теперь ее по $\operatorname{mod} 2$ со случайной последовательностью ε_1 такой же длины. Тогда в силу утверждения предыдущей задачи полученная последовательность ε_2 будет абсолютно случайной. Расшифровать основной текст можно, сложив $\varepsilon_2 + \varepsilon_1 = \varepsilon_0 (\operatorname{mod} 2)$.

Это один из способов кодирования секретной информации[3].

[1] См. [10].

[2] A million random digits with 100 000 normal deviates. Glencoe, Illinois: Rand Corporation, 1955.

[3] Использовался в горячей линии Москва–Вашингтон.

Глава 2

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И СЛУЧАЙНЫЕ СОБЫТИЯ

В предыдущей главе мы познакомились с примерами, в которых вероятности вычислялись прямым подсчетом. Для дальнейшего развития теории нам понадобятся новые понятия и новые обозначения.

В теории вероятностей принята своя терминология, отличная от используемой, скажем, в математическом анализе[1].

2.1. Случайные величины

Рассмотрим конечное вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$. Напомним, что:

Ω — это конечное множество, состоящее из s элементарных исходов $\omega_1, \ldots, \omega_s$,

$\mathcal{A}$ — класс всех подмножеств Ω, называемых *случайными событиями*,

$\mathbf{P}$ — вероятность или распределение вероятностей на классе $\mathcal{A}$.

Определение 2.1. *Случайной величиной*, заданной на конечном вероятностном пространстве, называется любая функция, зависящая от элементарного исхода ω и принимающая числовые значения.

Обозначать случайные величины будем заглавными буквами конца латинского алфавита $X = X(\omega)$, $Y = Y(\omega)$, $Z = Z(\omega)$, $\ldots$, или греческими буквами $\xi = \xi(\omega)$, $\eta = \eta(\omega)$, $\zeta = \zeta(\omega)$, $\ldots$. Аргумент ω часто опускают.

Замечание. В более сложных случаях, когда пространство Ω состоит из бесконечного числа элементарных исходов, все значительно усложняется. Общее определение вероятностного пространства и случайной величины появится позже (см. главу 9).

[1] В учебнике А. Н. Ширяева «Вероятность» можно найти таблицу соответствия некоторых терминов [30, с. 149–150].

Определение 2.2. *Математическим ожиданием* случайной величины $X(\omega)$, определенной на конечном вероятностном пространстве, называется число

$$\mathbf{E}X = \sum_{\omega\in\Omega} X(\omega)\cdot p(\omega).$$

Часто мы будем использовать более короткую запись

$$\mathbf{E}X = \sum_{\omega} X(\omega)p(\omega).$$

Обозначение $\mathbf{E}X$ происходит от первой буквы английского слова expectation (ожидание). В отечественной литературе для обозначения математического ожидания также широко используется $\mathbf{M}X$.

Определение 2.3. *Индикатором* случайного события A называется случайная величина

$$I_A(\omega) = \begin{cases} 1, & \text{если} \quad \omega\in A, \\ 0, & \text{если} \quad \omega\notin A. \end{cases}$$

Для индикатора события A

$$\mathbf{E}I_A(\omega) = \sum_{\omega} I_A(\omega)p(\omega) = \sum_{\omega\in A} 1\cdot p(\omega) = \mathbf{P}(A).$$

Таким образом, *математическим ожиданием индикатора случайного события является его вероятность.*

Случайные события обычно обозначают первыми заглавными буквами латинского алфавита: $A, B, C, \ldots$. Наряду с отдельными случайными событиями мы будем рассматривать классы (множества) случайных событий, для обозначения которых будем использовать рукописные заглавные буквы: $\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$.

2.2. Операции над случайными событиями

На множестве случайных событий $\mathcal{A}$ определены нижеследующие операции.

1. *Объединение событий*:

$A\cup B = \{\omega\colon\ \omega\in A$ или $\omega\in B\}$ — это событие, состоящее из исходов, которые принадлежат или A, или B, или одновременно двум этим событиям.

2. *Пересечение событий*:

$A \cap B = \{\omega\colon\ \omega \in A \text{ и } \omega \in B\}$ — это событие, состоящее из тех элементарных исходов, которые одновременно принадлежат A и B. Эту операцию также обозначают как $A \cdot B = AB$.

3. *Разность событий*:

$A \setminus B = \{\omega\colon \omega \in A$, но $\omega \notin B\}$ — событие, состоящее из элементарных исходов, принадлежащих A, но не принадлежащих B.

4. *Отрицание*:

$\overline{A} = \Omega \setminus A$ — это событие состоит из всех элементарных исходов, которые не принадлежат A; $\overline{A}$ также называют дополнением к A или событием, противоположным A.

5. *Симметрическая разность*:

$A \triangle B = (A \cup B) \setminus (A \cap B)$ — событие, состоящее из исходов, которые содержатся или в A, или в B, но не принадлежат обоим событиям одновременно. Симметрическую разность двух событий можно записать по-другому: $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

События называются *несовместимыми*, если у них нет общих исходов, т. е. $A \cap B = \varnothing$.

Для несовместимых событий симметрическая разность этих событий совпадает с их объединением.

Введем также важное соотношение между случайными событиями.

6. *Включение событий*. Мы будем использовать одно из обозначений,

$$A \subset B \quad \text{или} \quad A \subseteq B,$$

если из условия $\omega \in A$ следует, что $\omega \in B$. В теории вероятностей в этом случае говорят, что событие A влечет наступление события B. Если одновременно выполняются включения $A \subset B$ и $B \subset A$, то $A = B$, то есть в этом случае события совпадают.

Задача 2.1. Докажите следующие свойства операций над случайными событиями:

1) коммутативность:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A;$$

2) ассоциативность:

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C,$$
$$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C;$$

3) дистрибутивность:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C);$$

4) принцип двойственности:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

2.3. Операции над индикаторами

Операциям над случайными событиями соответствуют операции над индикаторами этих событий.

Задача 2.2. Докажите, что
1) $I_{A \cap B}(\omega) = I_A(\omega) \cdot I_B(\omega)$,
2) $I_{\overline{A}}(\omega) = 1 - I_A(\omega)$,
3) $I_{A \cup B}(\omega) = I_A(\omega) + I_B(\omega) - I_{A \cap B}(\omega)$,
4) $I_{A \triangle B}(\omega) = |I_A(\omega) - I_B(\omega)| = (I_A(\omega) + I_B(\omega)) (\mathrm{mod}\, 2)$.

Приведем для примера доказательство свойства 3 индикаторов случайных событий:

$$I_{A \cup B}(\omega) = 1 - I_{\overline{A \cup B}}(\omega) = 1 - I_{\overline{A} \cap \overline{B}}(\omega) = 1 - I_{\overline{A}}(\omega) \cdot I_{\overline{B}}(\omega) =$$
$$= 1 - (1 - I_A(\omega)) \cdot (1 - I_B(\omega)) = I_A(\omega) + I_B(\omega) - I_{A \cap B}(\omega).$$

Свойство 3 для индикатора объединения событий можно сформулировать и для большего числа событий. Так, для трех событий

$$I_{A \cup B \cup C}(\omega) = I_A(\omega) + I_B(\omega) + I_C(\omega) -$$
$$- I_{A \cap B}(\omega) - I_{A \cap C}(\omega) - I_{B \cap C}(\omega) + I_{A \cap B \cap C}(\omega). \quad (2.1)$$

Задача 2.3. Докажите равенство (2.1), а также обобщение свойства 3 для n событий, а именно:

$$I_{\bigcup\limits_{i=1}^{n} A_i}(\omega) = \sum_{i=1}^{n} I_{A_i}(\omega) - \sum_{1 \leqslant i < j \leqslant n} I_{A_i A_j}(\omega) +$$
$$+ \sum_{1 \leqslant i < j < k \leqslant n} I_{A_i A_j A_k}(\omega) + \ldots + (-1)^{n+1} I_{A_1 A_2 \ldots A_n}(\omega).$$

В этом равенстве, как и в (2.1), с чередующимися знаками присутствуют сначала сумма индикаторов всех событий A_i, затем сумма индикаторов пересечений всех различных пар событий $A_i A_j$, сумма индикаторов пересечений всех различных троек событий $A_i A_j A_k$ и т. д., пока не последует индикатор пересечения всех событий $A_1 A_2 \ldots A_n$.

Задача 2.4. Докажите свойства симметрической разности:

1) $A \triangle (B \triangle C) = (A \triangle B) \triangle C$,
2) $A \cup B = (A \triangle B) \triangle (A \cap B)$.

Пусть $A_1, \ldots, A_n$ — произвольные n случайных событий.

Определение 2.4. Событие F *определяется по событиям* $A_1, \ldots, A_n$, если индикатор события F есть функция от индикаторов этих событий, то есть

$$I_F(\omega) = H(I_{A_1}(\omega), \ldots, I_{A_n}(\omega)),$$

где $H(\cdot)$ — функция, определенная в вершинах единичного n-мерного куба.

В этом случае, зная, какие из событий $A_1, \ldots, A_n$ наступили и какие не наступили, можно однозначно судить о том, наступило или нет событие F.

Глава 3

СВОЙСТВА ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ

3.1. Свойства вероятности

Из определения вероятности $\mathbf{P}(A)$, $A \in \mathcal{A}$, для конечных вероятностных пространств следует, что вероятность обладает следующими свойствами:

P1. $\mathbf{P}(A) \geqslant 0$;

P2. $\mathbf{P}(\Omega) = 1$, $\mathbf{P}(\varnothing) = 0$;

P3. Для попарно несовместимых случайных событий $A_1, \ldots$ $\ldots, A_n$

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i).$$

Последнее свойство называют свойством *конечной аддитивности* вероятности. Напомним, что попарная несовместимость событий означает, что $A_i \cap A_j = \varnothing$, если $i \neq j$.

Первые два свойства вероятности очевидны. Докажем свойство P3.

Доказательство проведем по индукции по числу событий n.

Пусть $n = 2$. Тогда

$$\mathbf{P}(A_1 \cup A_2) = \sum_{\omega \in A_1 \cup A_2} p(\omega) = \sum_{\omega \in A_1} p(\omega) + \sum_{\omega \in A_2} p(\omega) = \mathbf{P}(A_1) + \mathbf{P}(A_2).$$

Заметим, что возможность изменения порядка суммирования в данном случае сомнений не вызывает, поскольку все суммы содержат конечное число слагаемых.

Пусть свойство выполняется для $n-1$ события $A_1, \ldots, A_{n-1}$, тогда

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbf{P}\left\{\left(\bigcup_{i=1}^{n-1} A_i\right) \bigcup A_n\right\} = \mathbf{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbf{P}(A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i),$$

что и доказывает свойство P3.

Все остальные свойства, приведенные ниже, могут быть получены из свойств P1, P2, P3[1].

4. Если $A \subset B$, то $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.

Доказательство. Поскольку событие B можно представить как объединение несовместимых событий $B = A \cup (B \setminus A)$, то

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geqslant \mathbf{P}(A). \qquad \square$$

5. Для любого случайного события A выполняются неравенства $0 \leqslant \mathbf{P}(A) \leqslant 1$.

6. $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$.

Задача 3.1. Докажите свойства вероятности 5, 6.

7. Если $A_1, \ldots, A_n$ — произвольные случайные события, то

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathbf{P}(A_i). \tag{3.1}$$

Данное свойство называется свойством *полуаддитивности* вероятности.

Доказательство. Введем события

$$\begin{aligned}
&B_1 = A_1,\\
&B_2 = A_2 \setminus A_1 = A_2 \cap \overline{A_1},\\
&B_3 = A_3 \setminus (A_1 \cup A_2) = A_3 \cap \overline{A_1} \cap \overline{A_2},\\
&\vdots\\
&B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) = A_n \cap \overline{A_1} \cap \ldots \cap \overline{A_{n-1}}.
\end{aligned}$$

1. Заметим, что события $B_1, \ldots, B_n$ попарно несовместимы. Действительно, пусть $i < j$, тогда по построению $B_i \subset A_i$, $B_j \subset \overline{A_i}$, но это означает, что $B_i \cap B_j = \varnothing$.

2. Рассмотрим события

$$F = \bigcup_{i=1}^{n} B_i, \quad C = \bigcup_{i=1}^{n} A_i.$$

Покажем, что $F = C$.

[1] Это замечание выполняется также для общей теории, где свойства P1–P3 будут служить определением вероятности (с заменой конечной аддитивности на счетную).

Пусть $\omega \in C$, тогда существует такое i, что $\omega \in A_i$. Возьмем наименьшее такое i. Обозначим его i_0. Тогда $\omega \in A_{i_0}$, но $\omega \notin A_1, \ldots$ $\ldots, A_{i_0-1}$, а значит, $\omega \in B_{i_0}$, $\omega \in \bigcup\limits_{i=1}^{n} B_i = F$. Следовательно, $C \subset F$.

Докажем теперь обратное включение. Пусть $\omega' \in F$. Тогда найдется такое j_0, что $\omega' \in B_{j_0} \subset A_{j_0}$. Но это означает, что элементарный исход $\omega' \in \bigcup\limits_{j=1}^{n} A_j = C$. Следовательно, $F \subset C$ и, стало быть, $F = C$.

Из доказанного вытекает, что

$$\mathbf{P}\left(\bigcup_{j=1}^{n} A_j\right) = \mathbf{P}\left(\bigcup_{j=1}^{n} B_j\right) = \sum_{j=1}^{n} \mathbf{P}(B_j) \leqslant \sum_{j=1}^{n} \mathbf{P}(A_j).$$

$\square$

Замечание. В общей теории сходное доказательство будет применено и к бесконечным последовательностям случайных событий $A_1, A_2, \ldots$. В этом случае полуаддитивность вероятности (3.1) можно сформулировать следующим образом.

7*. Пусть $A_1, A_2, \ldots$ — произвольная последовательность случайных событий. Тогда

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

Задача 3.2. Докажите для произвольных событий $A_1, \ldots, A_n$, что

$$\mathbf{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1 \leqslant i < j \leqslant n} \mathbf{P}(A_i \cap A_j) + \\ + \sum_{1 \leqslant i < j < k \leqslant n} \mathbf{P}(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} \mathbf{P}(A_1 \cap \ldots \cap A_n).$$

3.2. Свойства математического ожидания случайных величин

В рассматриваемой схеме конечных вероятностных пространств случайной величиной называется произвольная числовая функция $X = X(\omega)$, зависящая от элементарного исхода.

Математическое ожидание мы определили как

$$\mathbf{E}X = \sum_{\omega} X(\omega)p(\omega). \tag{3.2}$$

Сформулируем основные свойства математического ожидания.

1. $\mathbf{E}I_A = \mathbf{P}(A)$.
2. $\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y$.
3. $\mathbf{E}(c \cdot X) = c \cdot \mathbf{E}X$, где c — произвольная константа.
4. Если $X(\omega) \geqslant 0$ для всех ω, то $\mathbf{E}X \geqslant 0$, причем из равенства $\mathbf{E}X = 0$ следует, что $\mathbf{P}\{X = 0\} = 1$.

Проведем для примера доказательство свойства 2.

Доказательство. Согласно определению

$$\mathbf{E}X = \sum_{\omega} X(\omega)p(\omega), \quad \mathbf{E}Y = \sum_{\omega} Y(\omega)p(\omega). \tag{3.3}$$

Складывая математические ожидания в (3.3), получаем

$$\mathbf{E}X + \mathbf{E}Y = \sum_{\omega} (X(\omega) + Y(\omega))p(\omega) = \mathbf{E}(X+Y).$$

$\square$

Замечание. Математическое ожидание можно рассматривать как *линейный функционал*, определенный на множестве случайных величин, заданных на одном вероятностном пространстве[1]. Линейный функционал, для которого выполняется свойство 4, называют *неотрицательным*. Можно привести следующий пример линейного неотрицательного функционала.

Рассмотрим множество всех непрерывных фукнкций на отрезке $[0,1]$. Пусть $X(\omega)$ — непрерывная функция на $[0,1]$. Тогда $\int_0^1 X(\omega)\,d\omega$ обладает всеми перечисленными свойствами математического ожидания.

В действительности связь между понятиями математического ожидания и интеграла более глубокая. Общее понятие математического ожидания в теории вероятностей соответствует понятию интеграла Лебега в функциональном анализе. Но пока

[1] В общей теории к этим свойствам будет добавлено некоторое свойство «непрерывности» математического ожидания.

мы ограничимся простым случаем конечного вероятностного пространства.

Для вычисления математического ожидания совсем не обязательно знать значения $X(\omega)$ для каждого $\omega \in \Omega$. Достаточно знать *распределение* этой случайной величины.

Для случайных величин, принимающих конечное число различных между собой значений (а мы имеем дело пока именно с такими случайными величинами), *распределением* называют набор значений $x_1, \ldots, x_n$ и соответствующих этим значениям вероятностей $p_1, \ldots, p_n$, где

$$p_i = \mathbf{P}\{X(\omega) = x_i\}, \quad i = 1, ..., n.$$

Иногда мы будем задавать распределение случайной величины X таблицей

$$\begin{array}{c|c|c|c} x_1 & x_2 & \ldots & x_n \\ \hline p_1 & p_2 & \ldots & p_n \end{array}\ .$$

Обратите внимание на то, что в этой таблице указываются все различные между собой значения случайной величины X, а сумма вероятностей всегда равна 1.

Приведем несколько утверждений, позволяющих вычислять $\mathbf{E}X$, используя распределение случайной величины X.

Определение 3.1. Класс $\mathcal{B} = \{B_1, \ldots, B_q\}$ называется *конечным разбиением* Ω, если $B_i \cap B_j = \varnothing$ при всех $i \neq j$ и $\Omega = \bigcup\limits_{i=1}^{q} B_i$.

Пусть есть два разбиения:

$$\mathcal{B} = \{B_1, \ldots, B_q\},$$
$$\mathcal{C} = \{C_1, \ldots, C_r\}.$$

Составим всевозможные пересечения $B_i \cap C_j$, $1 \leqslant i \leqslant q$, $1 \leqslant j \leqslant r$. Нетрудно убедиться в том, что получили новое разбиение $\mathcal{A}$, элементами которого являются события

$$A_{ij} = \{B_i \cap C_j\}.$$

Разбиение $\mathcal{A}$ называют *пересечением разбиений* $\mathcal{B}$ и $\mathcal{C}$ и обозначают

$$\mathcal{A} = \mathcal{B} \times \mathcal{C}.$$

Лемма 3.1. *Пусть $X(\omega)$ — случайная величина, заданная на конечном вероятностном пространстве, $\mathcal{B} = \{B_1, \ldots, B_q\}$ —*

некоторое разбиение Ω, *причем* $X(\omega)$ *постоянна на каждом из* B_j *и* $X(\omega) = x'_j$, *если* $\omega \in B_j$. *Тогда*

$$\mathbf{E}X = \sum_{j=1}^{q} x'_j \mathbf{P}(B_j).$$

Доказательство. Поскольку случайная величина $X(\omega)$ постоянна на каждом элементе B_j разбиения $\mathcal{B}$, то ее можно представить в виде

$$X(\omega) = \sum_{j=1}^{q} x'_j \cdot I_{B_j}(\omega).$$

Возьмем математические ожидания от обеих частей этого равенства и воспользуемся свойствами 1–3 математического ожидания. Получим

$$\mathbf{E}X(\omega) = \sum_{j=1}^{q} x'_j \cdot \mathbf{E}I_{B_j}(\omega) = \sum_{j=1}^{q} x'_j \cdot \mathbf{P}(B_j).$$

□

С каждой случайной величиной можно связать *естественное разбиение.* Это разбиение связано с распределением $X(\omega)$. Пусть это распределение задано таблицей

$$\begin{array}{c|c|c|c} x_1 & x_2 & \dots & x_n \\ \hline p_1 & p_2 & \dots & p_n \end{array}\,. \tag{3.4}$$

Обозначим $A_j = \{\omega\colon X(\omega) = x_j\}$, $p_j = \mathbf{P}(A_j)$, $j = 1, \dots, n$. События A_j образуют разбиение множества элементарных исходов Ω и являются аналогом линий уровня в математическом анализе, а разбиение $\mathcal{A} = \{A_1, \dots, A_n\}$ называют *естественным разбиением* Ω, *порожденным случайной величиной* X.

С помощью леммы 3.1 легко доказать следующую теорему.

Теорема 3.1. *Если распределение случайной величины* $X(\omega)$ *задается таблицей* (3.4), *а случайная величина* $Y(\omega) = h(X(\omega))$ *является функцией* $X(\omega)$, *то*

1) $\mathbf{E}X = \sum\limits_{j=1}^{n} x_j p_j$,

2) $\mathbf{E}Y = \sum\limits_{j=1}^{n} h(x_j) p_j$.

Данная теорема позволяет вычислять не только математическое ожидание случайной величины $X(\omega)$, используя распределение X, но также находить математические ожидания случайных величин, являющихся функциями от X.

3.3. Вероятность появления хотя бы одного из n событий

Пусть $A_1, \ldots, A_n$ — произвольные случайные события. Обозначим $N(\omega)$ случайную величину, равную числу наступающих событий среди $A_1, \ldots, A_n$. Для $0 \leqslant k \leqslant n$ событие $\{N(\omega) = k\}$ означает, что произойдет ровно k событий, т. е. что найдутся $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ такие, что $\omega \in A_{i_1}, A_{i_2}, \ldots, A_{i_k}$, но указать $k+1$ событие, которым одновременно принадлежит ω, уже нельзя.

Задача 3.3. Для всех k вычислите $\mathbf{P}\{\omega\colon N(\omega) = k\}$.

Рассмотрим событие $C = \bigcup\limits_{i=1}^{n} A_i$, состоящее в том, что наступит хотя бы одно из событий $A_1, \ldots, A_n$. Это событие также можно связать со случайной величиной $N(\omega)$:

$$C = \bigcup_{i=1}^{n} A_i = \{\omega : N(\omega) \geqslant 1\}.$$

Для вычисления вероятности события C воспользуемся свойствами индикаторов случайных событий и свойствами математических ожиданий.

Введем обозначения:

$$\begin{aligned}
S_1 &= \sum_{i=1}^{n} \mathbf{P}(A_i),\\
S_2 &= \sum_{1 \leqslant i < j \leqslant n} \mathbf{P}(A_i A_j),\\
S_3 &= \sum_{1 \leqslant i < j < k \leqslant n} \mathbf{P}(A_i A_j A_k),\\
&\vdots\\
S_n &= \mathbf{P}(A_1 A_2 \cdots A_n).
\end{aligned}$$

Вероятности событий, связанных со случайной величиной $N(\omega)$, выражаются через величины $S_1, \ldots, S_n$. В том числе справедлива следующая теорема.

Теорема 3.2.

$$\mathbf{P}(C) = S_1 - S_2 + S_3 - \ldots + (-1)^{n+1} S_n. \tag{3.5}$$

Доказательство. Для индикатора события C имеем

$$I_C(\omega) = 1 - I_{\overline{C}}(\omega) = 1 - I_{\overline{A_1} \cap \ldots \cap \overline{A_n}}(\omega) = 1 - \prod_{j=1}^{n} (1 - I_{A_j}(\omega)) =$$

$$= 1 - \left(1 - \sum_{j=1}^{n} I_{A_j}(\omega) + \sum_{1 \leqslant i < j \leqslant n} I_{A_i}(\omega) \cdot I_{A_j}(\omega) - \ldots + (-1)^n \prod_{j=1}^{n} I_{A_j}(\omega)\right).$$

Возьмем математическое ожидание от обеих частей этого равенства. Из свойств аддитивности математического ожидания следует

$$\mathbf{P}(C) = S_1 - S_2 + S_3 - \ldots + (-1)^{n+1} S_n.$$

□

Мы получили формулу для вероятности объединения n случайных событий.

Задача 3.4. Докажите, что, обрывая сумму в (3.5) на какомлибо члене, можно получить оценки сверху и снизу для вероятности события C:

$$S_1 - S_2 \leqslant \mathbf{P}(C) \leqslant S_1,$$
$$S_1 - S_2 + S_3 - S_4 \leqslant \mathbf{P}(C) \leqslant S_1 - S_2 + S_3$$

и т. д.

Замечание. Подробнее о вычислении вероятностей событий $\{N(\omega) = k\}$ можно прочитать в учебнике В. Феллера [29, т. 1][1].

Пример 3.1. Рассмотрим случайные перестановки. Всего существует $n!$ перестановок n различных чисел. Припишем каждой из них вероятность $1/n!$. Обозначим A_k — событие, состоящее в том, что число k в перестановке встречается на k-м месте, в этом случае будем говорить, что k — неподвижная точка.

[1] Эта книга выдержала несколько изданий, в том числе и на русском языке. Первый том посвящен дискретным вероятностным моделям, а второй — непрерывным.

Используя полученные формулы, вычислим вероятность того, что есть хотя бы одна неподвижная точка.

Поскольку

$$\mathbf{P}(A_i) = \frac{(n-1)!}{n!}, \quad 1 \leqslant i \leqslant n,$$
$$\mathbf{P}(A_iA_j) = \frac{(n-2)!}{n!}, \quad i \neq j, \quad \ldots,$$

то

$$S_1 = n \cdot \mathbf{P}(A_i) = 1, \quad S_2 = C_n^2 \cdot \mathbf{P}(A_iA_j) = \frac{1}{2!},$$
$$S_3 = C_n^3 \cdot \mathbf{P}(A_iA_jA_k) = \frac{1}{3!}, \quad \ldots .$$

Таким образом,

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1} \cdot \frac{1}{n!}. \tag{3.6}$$

Сравнивая это выражение с разложением для

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots,$$

получим, что выражение (3.6) — это сумма первых n слагаемых в разложении для $1 - e^{-1}$. Поскольку ряд сходится быстро, получим хорошее приближение

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \approx 1 - e^{-1} \approx 1 - 0.36788 = 0.63212.$$

Уже при $n = 7$ точное значение вероятности совпадает с приведенным приближенным значением в 4 знаках после запятой, если $n = 9$, то совпадение есть в 5 знаках.

В учебнике В.Феллера [29, гл. IV, § 4, с. 127] приводится способ вычисления вероятностей того, что имеет место в точности k совпадений, и показано, что

$$\mathbf{P}\{N(\omega) = k\} \approx \frac{e^{-1}}{k!}.$$

Так, для $n = 10$ результаты приведены в таблице:

$N=k$	$\mathbf{P}\{N=k\}$	$N=k$	$\mathbf{P}\{N=k\}$
0	0.36788	5	0.00307
1	0.36788	6	0.00051
2	0.18394	7	0.00007
3	0.06131	8	0.00001
4	0.01533	9	0.00000

Эти результаты можно использовать в задачах угадывания: пусть имеется два набора из 10 карт (Т, К, Д, В, 10, 9, 8, 7, 6, 5). Тогда вероятности угадать определенное число карт вычисляются по приведенным формулам и указаны в таблице.

Глава 4

НЕЗАВИСИМОСТЬ СЛУЧАЙНЫХ СОБЫТИЙ И СЛУЧАЙНЫХ ВЕЛИЧИН

4.1. Условная вероятность. Независимость двух случайных событий

Понятие независимости встречается во многих разделах математики: это линейная независимость векторов в линейной алгебре, функциональная независимость систем функций в математическом анализе и т. д. Чтобы подчеркнуть отличие понятия независимости в теории вероятностей, ее называют статистической или стохастической независимостью. Но поскольку в нашем курсе будет рассматриваться только такая независимость, то в дальнейшем мы будем говорить просто о *независимости* случайных величин или о *независимости* случайных событий.

С понятием независимости в теории вероятностей тесно связано понятие *условной вероятности*. Приведем для начала некоторые наглядные соображения, поясняющие эти понятия.

Пусть n раз подбрасывают 2 игральных кости, событие A происходит, если на первой кости выпало число 1, событие B — если на второй выпало число 5. Результаты опыта будем отмечать в таблице, ставя знак $+$, если в соответствующем опыте наблюдалось наступление события, и знак $-$, если это не так. К примеру, после 6 проведенных испытаний мы могли получить следующую таблицу:

Номер испытания	1	2	3	4	5	6
A	$+$	$+$	$+$	$-$	$-$	$+$
B	$+$	$-$	$-$	$+$	$+$	$-$

Пусть n — это число проведенных испытаний (в примере $n = 6$), n_A — число испытаний, в которых наблюдалось событие A ($n_A = 4$), n_B — число испытаний, в которых произошло событие B ($n_B = 3$), n_{AB} — число испытаний, в которых наступили оба события ($n_{AB} = 1$).

Тогда $\frac{n_{AB}}{n_A}$ — относительная частота наступления события B в ряду тех испытаний, в которых произошло событие A. Поскольку при большом числе испытаний относительная частота наступления события близка к вероятности этого события, то мы можем записать

$$\frac{n_{AB}}{n_A} = \frac{\frac{n_{AB}}{n}}{\frac{n_A}{n}} \approx \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

Определение 4.1. Отношение $\frac{\mathbf{P}(AB)}{\mathbf{P}(A)}$ называют *условной вероятностью* события B при условии, что наступило событие A, и обозначают $\mathbf{P}(B|A)$.

В случае, когда $\mathbf{P}(B) = \mathbf{P}(B|A)$, естественно говорить, что событие B не зависит от события A. Но в этом случае $\mathbf{P}(B) \times \times \mathbf{P}(A) = \mathbf{P}(AB)$, откуда следует, что $\mathbf{P}(A) = \mathbf{P}(A|B)$, если только $\mathbf{P}(B) \neq 0$. То есть в том случае, когда B не зависит от A, событие A также не зависит от B. Поэтому естественнее дать определение независимости двух событий в симметричной форме, распространяющееся и на события с нулевой вероятностью.

Определение 4.2. События A и B называются *независимыми*, если

$$\mathbf{P}(AB) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Задача 4.1. Опыт состоит в двукратном бросании игральной кости. Покажите, что события A и B, введенные в рассмотренном выше примере, независимы.

В математической статистике мы рассмотрим более содержательные примеры: независимость пола ребенка от возраста родителей, независимость пола второго ребенка от пола старшего ребенка в двухдетной семье, независимость длительности продолжения телефонного разговора от того, сколько он уже длится, и т. д.

4.2. Независимость случайных величин. Взаимная независимость нескольких случайных событий

Рассмотрим несколько случайных величин $X_1(\omega)$, $X_2(\omega)$, …, $X_N(\omega)$ с возможными значениями x_{k,j_k}, где первый индекс $1 \leqslant k \leqslant \leqslant N$ соответствует номеру случайной величины, а второй — номеру значения ($1 \leqslant j_k \leqslant m_k$), m_k обозначает число различных между собой значений случайной величины X_k.

Определение 4.3. Случайные величины $X_1, \ldots, X_N$ называются *взаимно независимыми* (или просто *независимыми*), если выполнены $m_1 \times m_2 \times \ldots \times m_N$ соотношений:

$$\mathbf{P}\{X_1(\omega) = x_{1,j_1}, \ldots, X_N(\omega) = x_{N,j_N}\} = \prod_{k=1}^{N} \mathbf{P}\{X_k(\omega) = x_{k,j_k}\},$$

где $1 \leqslant j_k \leqslant m_k, \quad k = 1, \ldots, N$.

Замечание. В практических задачах обычно на основании каких-либо физических или других соображений предполагается независимость некоторых величин, а из этого выводится независимость (или зависимость) других случайных величин.

Понятие *взаимной независимости* случайных событий можно ввести, исходя из определения независимости случайных величин.

Определение 4.4. События $A_1, \ldots, A_N$ *взаимно независимы*, если взаимно независимы индикаторы этих событий $I_{A_1}(\omega)$, …, $I_{A_N}(\omega)$.

Это означает, что

$$\mathbf{P}\{\omega : I_{A_1}(\omega) = \delta_1, \ldots, I_{A_N}(\omega) = \delta_N\} = \prod_{k=1}^{N} \mathbf{P}\{\omega : I_{A_k}(\omega) = \delta_k\},$$

где

$$\delta_k = \begin{cases} 1, \\ 0, \end{cases} \quad i = 1, \ldots, N.$$

Всего требуется выполнение 2^N таких соотношений.

Поясним определение взаимной независимости случайных событий для частных случаев $N = 2, 3$.

Пусть имеется два события A_1, A_2. Тогда необходимо выполнение 2^2 соотношений:

$$\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2), \qquad \mathbf{P}(\overline{A_1} A_2) = \mathbf{P}(\overline{A_1}) \cdot \mathbf{P}(A_2),$$
$$\mathbf{P}(A_1 \overline{A_2}) = \mathbf{P}(A_1) \cdot \mathbf{P}(\overline{A_2}), \qquad \mathbf{P}(\overline{A_1}\,\overline{A_2}) = \mathbf{P}(\overline{A_1})\mathbf{P}(\overline{A_2}).$$

Задача 4.2. Покажите, что в этом случае все соотношения вытекают из какого-либо одного, например из первого.

Из утверждения предыдущей задачи следует, что взаимная независимость двух событий эквивалентна определенной выше независимости событий.

Аналогичным образом можно выписать $2^3 = 8$ соотношений, необходимых для взаимной независимости событий A_1, A_2, A_3:

$$\begin{gathered} \mathbf{P}(A_1 A_2 A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3), \\ \vdots \\ \mathbf{P}(\overline{A_1}\,\overline{A_2}\,\overline{A_3}) = \mathbf{P}(\overline{A_1})\mathbf{P}(\overline{A_2})\mathbf{P}(\overline{A_3}). \end{gathered} \tag{4.1}$$

Задача 4.3. Докажите, что в этом случае выполнение условий (4.1) взаимной независимости трех событий эквивалентно выполнению следующих равенств:

$$\begin{aligned} &\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2), \\ &\mathbf{P}(A_1 A_3) = \mathbf{P}(A_1)\mathbf{P}(A_3), \\ &\mathbf{P}(A_2 A_3) = \mathbf{P}(A_2)\mathbf{P}(A_3), \\ &\mathbf{P}(A_1 A_2 A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3). \end{aligned} \tag{4.2}$$

Теорема 4.1. *Для взаимной независимости N случайных событий $A_1, \ldots, A_N$ необходимо и достаточно выполнение $2^N - N - 1$ условий:*

$$\mathbf{P}(A_{i_1} \ldots A_{i_k}) = \mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_k}) \tag{4.3}$$

для любого $2 \leqslant k \leqslant N$, для любых наборов различных между собой индексов $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$.

Задача 4.4. Проведите доказательство теоремы самостоятельно.

Иногда в качестве определения взаимной независимости случайных событий (ее также называют независимостью в совокупности) берется выполнение условий (4.3)[1].

Определение 4.5. События $A_1, \ldots, A_N$ называются *попарно независимыми*, если в (4.3) выполнены соотношения для $k = 2$.

Всего для попарной независимости N случайных событий требуется выполнение C_N^2 условий. В частности, три события A_1, A_2, A_3 попарно независимы, если в (4.2) выполнены первые три равенства.

Задача 4.5. Пусть события A, B, C, D взаимно независимы, $E = A \cup B$, $F = C \cup D$. Докажите, что события E и F независимы.

[1] См., например, [30, с. 89].

4.3. Свойства независимых случайных величин и взаимно независимых случайных событий

Сформулируем сначала общие свойства независимых случайных величин.

Теорема 4.2. *Пусть случайные величины* $X_1, \ldots, X_N$ *независимы. Тогда для любых наборов индексов* $1 \leqslant j_1 < j_2 < \ldots j_k \leqslant N$ *случайные величины* $X_{j_1}, X_{j_2}, \ldots, X_{j_k}$ *независимы.*

Доказательство. Проведем сначала доказательство для трех случайных величин. Чтобы показать, как это делается, и не загромождать записи индексами, рассмотрим независимые случайные величины U, V, W соответственно со значениями

$$\begin{aligned} U:&\ \ u_1, \ldots, u_l; \\ V:&\ \ v_1, \ldots, v_m; \\ W:&\ \ w_1, \ldots, w_n. \end{aligned}$$

Покажем, что из независимости трех случайных величин следует, что любые две, например U и V, также независимы. Для любых значений этих случайных величин из определения независимости случайных величин следует

$$\mathbf{P}\{U = u_i, V = v_j, W = w_k\} = \mathbf{P}\{U = u_i\}\mathbf{P}\{V = v_j\}\mathbf{P}\{W = w_k\}.$$

Просуммируем эти равенства по $1 \leqslant k \leqslant n$, получим

$$\sum_{k=1}^{n} \mathbf{P}\{U = u_i, V = v_j, W = w_k\} = \\ = \mathbf{P}\{U = u_i\}\mathbf{P}\{V = v_j\} \sum_{k=1}^{n} \mathbf{P}\{W = w_k\},$$

откуда получаем

$$\mathbf{P}\{U = u_i, V = v_j\} = \mathbf{P}\{U = u_i\}\mathbf{P}\{V = v_j\}.$$

Аналогичными рассуждениями в общем случае можно показать, что если независимы любые N случайных величин, то независимы и произвольные $N-1$ из этого набора случайных величин. А далее по индукции можно доказать независимость произвольного набора из $k \leqslant N$ случайных величин. $\square$

Приведем без доказательства еще одно свойство независимых случайных величин.

Теорема 4.3. *Пусть случайные величины $X_1,\ldots,X_N$ независимы и разбиты на непересекающиеся группы:*

$$\begin{array}{lll} X_{1,1}, & \ldots, & X_{1,n_1}, \\ X_{2,1}, & \ldots, & X_{2,n_2}, \\ \vdots & & \\ X_{k,1}, & \ldots, & X_{k,n_k}, \end{array}$$

$N = n_1 + n_2 + \ldots + n_k$, $k \geqslant 2$. *Определим новые случайные величины* $Y_1 = g_1(X_{1,1},\ldots,X_{1,n_1})$, $\ldots$, $Y_k = g_k(X_{k,1},\ldots,X_{k,n_k})$. *Тогда* Y_1, $\ldots$, Y_k *также независимы.*

Эти свойства вполне соответствуют нашим представлениям о независимости и аналогичны теоремам линейной алгебры о линейной независимости векторов[1].

Задача 4.6. Докажите утверждение теоремы 4.3 для случая $n_1 = n_2 = \ldots = n_k = 1$.

Для случайных событий из этих теорем вытекают следующие утверждения.

Теорема 4.4. *Если события $A_1,\ldots,A_N$ взаимно независимы, то и любые k из них также взаимно независимы.*

Для формулировки следующей теоремы напомним, что мы говорим, что событие A определяется по событиям $A_1,\ldots,A_l$, если

$$I_A(\omega) = f(I_{A_1}(\omega),\ldots,I_{A_l}(\omega)),$$

где $f(\cdot)$ — функция, определенная в вершинах единичного l-мерного куба и принимающая значения 0 и 1.

Определение 4.6. Событие A определяется по случайным величинам $X_1,\ldots,X_l$, если

$$I_A(\omega) = g(X_1(\omega),\ldots,X_l(\omega)),$$

т. е. если по известным значениям случайных величин можно установить, наступило событие A или нет.

Теорема 4.5. *Если взаимно независимые события $A_1,\ldots,A_N$*

[1] Применительно к векторам во второй теореме $g_1,\ldots,g_k$ — невырожденные линейные преобразования векторов.

разбиты на непересекающиеся группы событий

$$\begin{matrix} A_{1,1}, & \ldots, & A_{1,n_1}, \\ \vdots & & \\ A_{k,1}, & \ldots, & A_{k,n_k}, \end{matrix}$$

а события $B_1,\ldots,B_k$ определяются каждое по своей группе событий, то $B_1,\ldots,B_k$ также взаимно независимы.

4.4. Критерий независимости случайных величин

Теорема 4.6. *Для независимости случайных величин $X_1,\ldots$ $\ldots,X_N$ необходимо и достаточно, чтобы при любом выборе полуинтервалов $[a_1,b_1)$, $\ldots$, $[a_N,b_N)$ на числовой прямой выполнялись равенства*

$$\mathbf{P}\{\omega\colon a_1 \leqslant X_1(\omega) < b_1, \ \ldots, a_N \leqslant X_N(\omega) < b_N\} = \\ = \prod_{k=1}^{N} \mathbf{P}\{\omega\colon a_k \leqslant X_k < b_k\}. \quad (4.4)$$

Замечание. В дальнейшем, когда мы отойдем от конечного случая и будем рассматривать вероятностные пространства общего вида (см. главу 9), выполнение условий (4.4) берется в качестве определения независимости случайных величин.

Доказательство. Доказательство теоремы проведем для двух случайных величин U и V со значениями $u_1,\ldots,u_l$ и $v_1,\ldots,v_m$ соответственно.

1. Докажем сначала достаточность условия (4.4). Пусть u_j, v_k — произвольные значения этих случайных величин. Выберем интервалы $[a,b)$, $[c,d)$ таким образом, чтобы $u_j \in [a,b)$, $v_k \in [c,d)$, но других значений случайных величин в них не попадало. Тогда

$$\mathbf{P}\{U(\omega)=u_j,\ V(\omega)=v_k\} = \mathbf{P}\{a \leqslant U(\omega) < b,\ c \leqslant V(\omega) < d\} = \\ = \mathbf{P}\{a \leqslant U(\omega) < b\}\mathbf{P}\{c \leqslant V(\omega) < d\} = \mathbf{P}\{U(\omega)=u_j\}\mathbf{P}\{V(\omega)=v_k\}.$$

2. Для доказательства необходимости (4.4) рассмотрим событие

$$C = \{a \leqslant U(\omega) < b,\ c \leqslant V(\omega) < d\} =$$

$$= \bigcup_{u_j \in [a,b),\ v_k \in [c,d)} \{U(\omega) = u_j,\ V(\omega) = v_k\}.$$

Поскольку объединяемые события попарно не пересекаются, то

$$\mathbf{P}(C) = \sum_{u_j \in [a,b),\ v_k \in [c,d)} \mathbf{P}\{U(\omega) = u_j,\ V(\omega) = v_k\}.$$

Воспользовавшись независимостью случайных величин, получаем

$$\mathbf{P}(C) = \sum_{u_j \in [a,b),\ v_k \in [c,d)} \mathbf{P}\{U(\omega) = u_j\}\mathbf{P}\{V(\omega) = v_k\} =$$
$$= \left(\sum_{u_j \in [a,b)} \mathbf{P}\{U(\omega) = u_j\}\right) \cdot \left(\sum_{v_k \in [c,d)} \mathbf{P}\{V(\omega) = v_k\}\right) =$$
$$= \mathbf{P}\{a \leqslant U(\omega) < b\}\mathbf{P}\{c \leqslant V(\omega) < d\}.$$

С очевидными изменениями доказательство теоремы легко переносится на общий случай. □

4.5. Мультипликативное свойство математического ожидания независимых случайных величин

Теорема 4.7. *Пусть* $X_1, \ldots, X_N$ *— независимые случайные величины. Тогда*

$$\mathbf{E}\left(\prod_{i=1}^{N} X_i\right) = \prod_{i=1}^{N} \mathbf{E} X_i.$$

Доказательство. 1. Докажем утверждение для двух случайных величин U и V. Обозначим $A_j = \{U(\omega) = u_j\}$, $B_k = \{V(\omega) = v_k\}$.

События $A_j \cap B_k$ при $j = 1, \ldots, l$, $k = 1, \ldots, m$ образуют разбиение Ω, причем если $\omega \in A_j \cap B_k$, то $U(\omega) \cdot V(\omega) = u_j \cdot v_k$. Из леммы 3.1

$$\mathbf{E}(UV) = \sum_{j=1}^{l} \sum_{k=1}^{m} u_j v_k \mathbf{P}(A_j \cap B_k).$$

Используя независимость случайных величин, получим, что

$$\mathbf{E}(UV) = \sum_{j=1}^{l}\sum_{k=1}^{m} u_j v_k \mathbf{P}(A_j)\mathbf{P}(B_k) =$$
$$= \left(\sum_{j=1}^{l} u_j \mathbf{P}(A_j)\right) \cdot \left(\sum_{k=1}^{m} v_k \mathbf{P}(B_k)\right) = \mathbf{E}U \cdot \mathbf{E}V.$$

2. Пусть теперь $N \geqslant 3$ и для любого числа сомножителей, не превосходящего $N-1$, теорема верна. Представим

$$X_1 \cdots X_N = U \cdot V,$$

где $U = X_1 \cdots X_{N-1}$, $V = X_N$. В силу теоремы 4.2 случайные величины U и V независимы, поэтому

$$\mathbf{E}(UV) = \mathbf{E}U \cdot \mathbf{E}V = \mathbf{E}(X_1 \cdots X_{N-1}) \cdot \mathbf{E}X_N.$$

Но по предположению индукции первый множитель в правой части равен произведению математических ожиданий случайных величин, следовательно,

$$\mathbf{E}(X_1 \cdots X_N) = \mathbf{E}X_1 \cdots \mathbf{E}X_N.$$

$\square$

Глава 5

СУММИРОВАНИЕ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Начиная с 1713 г., когда впервые был опубликован закон больших чисел, открытый Я. Бернулли, на протяжении двух с половиной столетий основные теоремы теории вероятностей касались поведения сумм независимых случайных величин, играющих весьма важную роль как в теории вероятностей и математической статистике, так и в их приложениях. Около 60 лет назад вышла книга Б. В. Гнеденко и А. Н. Колмогорова «Предельные распределения для сумм независимых случайных величин», в которую были включены как результаты предшественников, так и результаты авторов, внесших значительный вклад в развитие теории суммирования случайных величин. Появление этой книги вызвало новый интерес к данному вопросу и послужило толчком в развитии методов исследования распределений сумм случайных величин.

Находить точные распределения сумм большого числа случайных величин, как правило, весьма сложно даже для простых распределений. Так, в первой главе мы искали точное распределение суммы очков, выпавших на 6 костях, которая представляет собой сумму $X_1+X_2+\ldots+X_6$ всего лишь шести слагаемых. Точные расчеты оказались весьма непростыми, хотя распределения самих X_i несложные: каждая из этих случайных величин принимает значения $1,2,\ldots,6$ с равными вероятностями: $\mathbf{P}\{X_i=k\}=1/6$.

Нам понадобится некоторый аналитический аппарат, связанный с математическими ожиданиями случайных величин.

5.1. Производящая функция целочисленной случайной величины

Пусть случайная величина X принимает целые неотрицательные значения $0,1,2,\ldots,m$ с вероятностями $p_0,p_1,\ldots,p_m$. Рассмотрим функцию действительной переменной z

$$G(z,X)=\mathbf{E}(z^X)=\sum_{k=0}^{m} z^k\cdot p_k. \tag{5.1}$$

Эта функция называется *производящей функцией* случайной величины X.

Замечание. Часть слагаемых в сумме (5.1) может равняться 0, если соответствующие вероятности $p_k = 0$.

Производящая функция может быть определена и для целочисленных случайных величин, принимающих бесконечное число значений:

$$G(z, X) = \sum_{k=0}^{\infty} z^k p_k,$$

но тогда степенной ряд может сходиться не при всех z. Однако при $|z| \leqslant 1$ такие ряды сходятся, причем равномерно, так как $\sum\limits_{k=1}^{\infty} p_k = 1$.

Всюду в дальнейшем для суммы случайных величин $X_1, \ldots$ $\ldots, X_N$ будем использовать обозначение

$$S_N = X_1 + \ldots + X_N.$$

Если случайные величины $X_1, \ldots, X_N$ независимы, то для производящей функции S_N выполняется свойство *мультипликативности*:

$$G(z, S_N) = \prod_{k=1}^{N} G(z, X_k).$$

Задача 5.1. Докажите свойство мультипликативности производящих функций.

По производящей функции распределение случайной величины восстанавливается однозначно, поскольку вероятности p_k есть не что иное, как коэффициенты разложения в ряд Тейлора для этой функции в окрестности 0:

$$p_0 = G(0, X), \quad p_1 = G'(0, X), \quad \ldots, \quad p_k = \frac{G^{(k)}(0, X)}{k!}, \quad \ldots .$$

Зная производящую функцию, можно найти математическое ожидание и другие числовые характеристики случайной величины.

Вычислим первые две производные производящей функции:

$$G'(z,X) = \sum_{k=0}^{m} kz^{k-1}p_k, \quad G''(z,X) = \sum_{k=0}^{m} k(k-1)z^{k-2}p_k.$$

Легко видеть, что

$$G'(1,X) = \sum_{k=0}^{m} kp_k = \mathbf{E}X, \quad G''(1,X) = \sum_{k=0}^{m} k(k-1)p_k = \mathbf{E}X(X-1).$$

Пример 5.1. Пусть производящая функция случайной величины X равна

$$G(z,X) = (1-p+pz)^2.$$

Найти распределение X.

Очевидно, что

$$G(z,X) = (1-p)^2 + 2p(1-p)z + p^2z^2.$$

Поскольку вероятности $p_k = \mathbf{P}\{X = k\}$ — это коэффициенты при z^k, то получаем

$$p_0 = (1-p)^2, \quad p_1 = 2p(1-p), \quad p_2 = p^2.$$

Таким образом, случайная величина X, имеющая заданную производящую функцию, принимает значения $0, 1, 2$ с вычисленными выше вероятностями.

Задача 5.2. Найдите производящую функцию случайной величины X, принимающей значения $0, 1, \ldots, n$ с вероятностями

$$p_k = C_n^k p^k (1-p)^{n-k}.$$

Такое распределение вероятностей называют *биномиальным* с параметрами $n \in \mathcal{N}$ и $0 < p < 1$. В этом случае будем писать $X \sim B(n,p)$.

Задача 5.3. Найдите производящую функцию случайной величины X, принимающей значения $0, 1, 2, \ldots$ с вероятностями

$$p_k = \frac{\lambda^k \cdot e^{-\lambda}}{k!},$$

где $\lambda > 0$ — некоторый параметр. Это распределение называют *распределением Пуассона* с параметром λ и обозначают $X \sim \Pi(\lambda)$.

Задача 5.4. Используя свойство мультипликативности производящих функций, покажите, что сумма независимых случайных величин X_1 и X_2, имеющих распределение Пуассона с параметрами λ_1 и λ_2 соответственно, также распределена по закону Пуассона, но с параметром $\lambda_1+\lambda_2$.

5.2. Производящая функция моментов

Рассмотрим случайную величину X, распределение которой задается таблицей

$$\begin{array}{c|c|c|c} x_1 & x_2 & \dots & x_m \\ \hline p_1 & p_2 & \dots & p_m \end{array}\ , \tag{5.2}$$

в которой указаны все различные между собой значения случайной величины и соответствующие им вероятности.

Производящей функцией моментов случайной величины X называется функция действительной переменной h, определенная следующим образом:

$$g(h,X)=\mathbf{E}\,e^{hX}=\sum_{k=1}^{m}e^{hx_k}p_k.$$

Замечание. Поясним происхождение названия этой функции. Рассмотрим разложение в ряд Тейлора по степеням h функции e^{hX}:

$$e^{hX}=1+\frac{hX}{1!}+\frac{h^2X^2}{2!}+\ldots+\frac{h^nX^n}{n!}+\ldots\ .$$

Если формально взять математические ожидания от обеих частей этого равенства, то получим

$$\mathbf{E}\,e^{hX}=1+\frac{\mathbf{E}X}{1!}h+\frac{\mathbf{E}X^2}{2!}h^2+\ldots+\frac{\mathbf{E}X^n}{n!}h^n+\ldots\ .$$

Таким образом, по функции $g(h,X)$, разложив ее в ряд Тейлора, можно найти $m_n=\mathbf{E}X^n$, называемые *моментами* случайной величины X.

Задача 5.5. Пусть $X_1,\ldots,X_N$ — независимые случайные величины,

$$S_N=X_1+X_2+\ldots+X_N$$

— сумма независимых случайных величин. Докажите свойство мультипликативности для производящей функции моментов суммы независимых случайных величин:

$$g(h,S_N) = g(h,X_1)\cdot g(h,X_2)\cdot \ldots \cdot g(h,X_N).$$

Сформулируем следующее свойство показательных функций.

Лемма 5.1. *Пусть* $-\infty < \alpha_1 < \alpha_2 < \alpha_3 < \ldots < \infty$ *— произвольные действительные числа. Если для всех* $-\infty < h < \infty$ *выполняется*

$$\lambda_1 e^{\alpha_1 h} + \lambda_2 e^{\alpha_2 h} + \ldots + \lambda_m e^{\alpha_m h} = 0,$$

то $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$.

Доказательство. Доказательство проведем для случая $m = 3$. Итак, пусть

$$\lambda_1 e^{\alpha_1 h} + \lambda_2 e^{\alpha_2 h} + \lambda_3 e^{\alpha_3 h} = 0 \tag{5.3}$$

для всех h. Продифференцируем (5.3) дважды по h:

$$\begin{aligned} \alpha_1\lambda_1 e^{\alpha_1 h} + \alpha_2\lambda_2 e^{\alpha_2 h} + \alpha_3\lambda_3 e^{\alpha_3 h} &= 0, \\ \alpha_1^2\lambda_1 e^{\alpha_1 h} + \alpha_2^2\lambda_2 e^{\alpha_2 h} + \alpha_3^2\lambda_3 e^{\alpha_3 h} &= 0. \end{aligned} \tag{5.4}$$

Подставляя $h = 0$ в (5.3), (5.4), получим систему линейных уравнений относительно переменных $\lambda_1, \lambda_2, \lambda_3$:

$$\begin{array}{rcrcrl} \lambda_1 & + & \lambda_2 & + & \lambda_3 & = 0, \\ \alpha_1\lambda_1 & + & \alpha_2\lambda_2 & + & \alpha_3\lambda_3 & = 0, \\ \alpha_1^2\lambda_1 & + & \alpha_2^2\lambda_2 & + & \alpha_3^2\lambda_3 & = 0. \end{array}$$

Определитель системы

$$\begin{vmatrix} 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 \end{vmatrix}$$

— это определитель Вандермонда[1]. Из курса линейной алгебры известно, что при $\alpha_1 < \alpha_2 < \alpha_3$ этот определитель не равен 0, значит, однородная система линейных уравнений имеет единственное решение

$$\lambda_1 = \lambda_2 = \lambda_3 = 0.$$

Это доказательство легко распространяется на случай произвольного m. □

[1] См., например, [23, т. 1, с. 578].

Замечание. Свойство показательных функций, сформулированное в лемме, в математическом анализе называется линейной независимостью этих функций.

Задача 5.6. Используя это свойство показательных функций, докажите, что если производящие функции моментов двух случайных величин X и Y совпадают при всех h, то эти случайные величины имеют одинаковые распределения, то есть принимают одинаковые значения с равными вероятностями.

Таким образом, производящая функция моментов случайной величины однозначно определяет распределение. Более того, соответствие между производящими функциями моментов (а также производящими функциями целочисленных случайных величин) и распределениями случайных величин оказывается непрерывным в том смысле, что если распределения близки между собой, то близки и производящие функции, и обратно[1].

5.3. Свойства числовых характеристик распределений сумм независимых случайных величин

Из свойств математического ожидания следует, что

$$\mathbf{E}S_N = \mathbf{E}X_1 + \mathbf{E}X_2 + \ldots + \mathbf{E}X_N.$$

Существуют ли другие числовые характеристики, которые для независимых случайных величин обладали бы столь же простым свойством?

Ответ на этот вопрос положителен, и одной из таких характеристик является *дисперсия*.

Определение 5.1. *Дисперсией* случайной величины X называется число

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2.$$

Эта числовая характеристика распределения случайной величины является мерой разброса (рассеивания) значений случайной величины от ее среднего значения $\mathbf{E}X$. Если распределение

[1] Теоремы непрерывности для производящих функций можно посмотреть, например, в [24, с. 51] или в [29, т. 1, с. 294].

задано таблицей (5.2), то

$$\mathbf{D}X = \sum_{k=1}^{m}(x_k - \mathbf{E}X)^2 p_k.$$

Иногда дисперсию удобно вычислять по-другому, используя тот факт, что

$$\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = m_2 - (m_1)^2,$$

где m_1, m_2 — обозначения для моментов первого и второго порядков.

Действительно, используя простые преобразования и свойства математического ожидания, имеем

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}\left(X^2 - 2X\,\mathbf{E}X + (\mathbf{E}X)^2\right) = \\ = \mathbf{E}X^2 - 2\,\mathbf{E}X \cdot \mathbf{E}X + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

Перечислим свойства дисперсии.

1. Для любой случайной величины $\mathbf{D}X \geqslant 0$. Отсюда, в частности, следует неравенство

$$\mathbf{E}X^2 \geqslant (\mathbf{E}X)^2.$$

2. Дисперсия не изменяется при добавлении к случайной величине произвольной константы c: $\mathbf{D}(X+c) = \mathbf{D}X$.

3. $\mathbf{D}(cX) = c^2\,\mathbf{D}X$ — при умножении случайной величины на константу c ее дисперсия изменяется в c^2 раз.

4. Если $X_1, \ldots, X_N$ — независимые случайные величины, то

$$\mathbf{D}S_N = \mathbf{D}X_1 + \ldots + \mathbf{D}X_N. \tag{5.5}$$

Доказательство. Доказательство этого свойства сначала проведем для двух случайных величин X_1 и X_2. Введем новые случайные величины

$$Y_1 = X_1 - \mathbf{E}X_1, \quad Y_2 = X_2 - \mathbf{E}X_2,$$

тогда $\mathbf{E}Y_1 = \mathbf{E}Y_2 = 0$. Применяя свойства математического ожидания, получим

$$\mathbf{D}(X_1 + X_2) = \mathbf{E}(X_1 + X_2 - (\mathbf{E}X_1 + \mathbf{E}X_2))^2 = \mathbf{E}(Y_1 + Y_2)^2 = \\ = \mathbf{E}(Y_1)^2 + 2\,\mathbf{E}(Y_1 Y_2) + \mathbf{E}(Y_2)^2 = \mathbf{D}X_1 + \mathbf{D}X_2 + 2\,\mathbf{E}Y_1 \cdot \mathbf{E}Y_2 = \\ = \mathbf{D}X_1 + \mathbf{D}X_2.$$

Далее по индукции можно получить, что утверждение верно для произвольного числа случайных величин N. □

Можно рассматривать и другие числовые характеристики распределений, обладающие свойством, подобным свойству (5.5) дисперсий. Их можно получить следующим образом.

Рассмотрим производящую функцию моментов

$$\mathbf{E}\, e^{hS_N} = \prod_{k=1}^{N} \mathbf{E}\, e^{hX_k}.$$

Ее логарифм равен

$$\ln(\mathbf{E}\, e^{hS_N}) = \sum_{k=1}^{N} \ln(\mathbf{E}\, e^{hX_k}),$$

а производные любого порядка от $\ln(\mathbf{E}\, e^{hS_N})$ равны сумме соответствующих производных от логарифмов производящих функций моментов отдельных слагаемых.

Задача 5.7. Проверить, что

$$\frac{d}{dh}\ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{E}\, X,$$
$$\frac{d^2}{dh^2}\ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{D}\, X,$$
$$\frac{d^3}{dh^3}\ln(\mathbf{E}\, e^{hX})|_{h=0} = \mathbf{E}(X - \mathbf{E}\, X)^3.$$

Замечание. Вычисленные при $h = 0$ производные $\ln(\mathbf{E}\, e^{hX})$ называют *семиинвариантами* случайной величины X. Обычно семиинвариант порядка k обозначают $\varkappa_k$. Так, $\varkappa_1 = \mathbf{E}\, X$, $\varkappa_2 = \mathbf{E}(X - \mathbf{E}\, X)^2 = \mathbf{D}\, X$, $\varkappa_3 = \mathbf{E}(X - \mathbf{E}\, X)^3$.

Задача 5.8. Выразите через моменты семиинварианты четвертого и пятого порядков.

Задача 5.9. Докажите следующее свойство семиинвариантов для независимых случайных величин:

$$\varkappa_k(S_N) = \varkappa_k(X_1) + \ldots + \varkappa_k(X_N).$$

Числовые характеристики $\mathbf{E}(X - \mathbf{E}\, X)^k$ называются *центральными моментами порядка k*.

Задача 5.10. Выразите центральные моменты до пятого порядка включительно через моменты и через семиинварианты.

Задача 5.11. Рассмотрим функцию $f(c) = \mathbf{E}(X - c)^2$. При каком значении c функция $f(c)$ достигает минимального значения?

Глава 6

НЕРАВЕНСТВА ЧЕБЫШЕВА. ОТКЛОНЕНИЯ СУММ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

6.1. Схемы Бернулли и Пуассона

Пусть $A_1, \ldots, A_N$ — независимые случайные события. Обозначим

$$p_j = \mathbf{P}(A_j), \quad q_j = \mathbf{P}(\overline{A_j}) = 1 - p_j.$$

Рассмотрим случайную величину $\mu_N(\omega)$ — число наступивших событий среди $A_1, \ldots, A_N$, если осуществился элементарный исход ω. Другими словами, $\mu_N(\omega)$ — это число событий A_k, которым одновременно принадлежит исход ω. Эту случайную величину удобно выразить через индикаторы случайных событий, а именно:

$$\mu_N(\omega) = I_1(\omega) + \ldots + I_N(\omega), \quad I_j(\omega) = I_{A_j}(\omega).$$

Из независимости событий следует независимость их индикаторов, поэтому, учитывая свойства математического ожидания и дисперсии для независимых случайных величин, получим

$$\begin{aligned} \mathbf{E}\,\mu_N &= \mathbf{E}\,I_1 + \ldots + \mathbf{E}\,I_N = p_1 + \ldots + p_N, \\ \mathbf{D}\,\mu_N &= \mathbf{D}\,I_1 + \ldots + \mathbf{D}\,I_N = p_1 q_1 + \ldots + p_N q_N. \end{aligned}$$

Производящая функция случайной величины μ_N равна

$$\mathbf{E}(z^{\mu_N}) = \prod_{j=1}^{N} \mathbf{E}\,z^{I_{A_j}} = \prod_{j=1}^{N} (q_j + z p_j).$$

Коэффициенты разложения производящей функции по степеням z — это вероятности $\mathbf{P}\{\mu_N(\omega) = k\}$.

В частности, если $p_0 = p_1 = \ldots = p_N = p$, $q = 1 - p$, то

$$\mathbf{E}(z^{\mu_N}) = (q + zp)^N = \sum_{k=0}^{N} C_N^k p^k q^{N-k} z^k.$$

Следовательно, в этом случае

$$\mathbf{P}\{\omega\colon \mu_N(\omega) = k\} = C_N^k p^k q^{N-k}, \quad 0 \leqslant k \leqslant N.$$

Такое распределение вероятностей называют *биномиальным* с параметрами (N, p), так как вероятности в этом случае определяются по формулам бинома Ньютона для $(p+q)^N$.

В общем случае, когда вероятности событий A_j не все равны между собой, формулы получаются громоздкими.

Описанная выше модель называется схемой Бернулли для равных вероятностей и схемой Пуассона в общем случае.

6.2. Неравенства Чебышева

В дальнейшем нам понадобятся две теоремы, носящие имя известного русского математика П. Л. Чебышёва (1821–1894), выпускника Московского университета, члена многих академий мира. П. Л. Чебышёв создал метод математических ожиданий, благодаря которому упростились и стали возможными доказательства многих теорем.

Лемма 6.1. *Пусть случайная величина* $Y \geqslant 0$, $\varepsilon > 0$ — *произвольное положительное число. Тогда*

$$\mathbf{P}\{\omega\colon Y(\omega) \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}Y}{\varepsilon}. \tag{6.1}$$

Доказательство. Обозначим $A = \{\omega\colon\ Y(\omega) \geqslant \varepsilon\}$. Воспользуемся тождеством

$$I_A(\omega) + I_{\overline{A}}(\omega) = 1.$$

Умножим обе части этого равенства на Y, получим

$$Y(\omega) = Y(\omega)I_A(\omega) + Y(\omega)I_{\overline{A}}(\omega).$$

Второе слагаемое неотрицательно, а первое можно оценить следующим образом:

$$Y(\omega)I_A(\omega) \geqslant \varepsilon I_A(\omega).$$

Но тогда

$$\mathbf{E}Y(\omega) \geqslant \mathbf{E}(Y(\omega)I_A(\omega)) \geqslant \varepsilon\mathbf{P}(A),$$

откуда и вытекает утверждение леммы. □

Теорема 6.1 (неравенство Чебышева). *Пусть* X — *случайная величина,* $\mathbf{E}X = a$, $\varepsilon > 0$. *Тогда*

$$\mathbf{P}\{\omega\colon |X(\omega) - a| \geqslant \varepsilon\} \leqslant \frac{\mathbf{D}X}{\varepsilon^2}.$$

Доказательство. Обозначим $Y = (X - a)^2 \geqslant 0$. Проводя простые преобразования и используя утверждение леммы, получим

$$\mathbf{P}\{|X - a| \geqslant \varepsilon\} = \mathbf{P}\{Y \geqslant \varepsilon^2\} \leqslant \frac{\mathbf{E}\,Y}{\varepsilon^2} = \frac{\mathbf{D}\,X}{\varepsilon^2}.$$

$\square$

Введем еще одну числовую характеристику распределения случайной величины X — *стандартное отклонение* $\sigma_X = \sqrt{\mathbf{D}X}$. Стандартное отклонение, так же как и дисперсия, является характеристикой рассеяния значений случайной величины, но имеет ту же размерность, что и сама случайная величина X. Рассмотрим $\varepsilon = t\sigma_X$, $t > 0$. Тогда из неравенства Чебышева следует, что вероятность отклонений[1], превышающих по абсолютной величине стандартное отклонение в t раз, не превосходит $1/t^2$:

$$\mathbf{P}\{|X - a| \geqslant t\sigma_X\} \leqslant \frac{\mathbf{D}X}{t^2\sigma_X^2} = \frac{1}{t^2},$$

а вероятность отклонений, в t раз меньших стандартного, наоборот, больше $1 - 1/t^2$:

$$\mathbf{P}\{|X - a| < t\sigma_X\} \geqslant 1 - \frac{1}{t^2}.$$

Теорема 6.2 (экспоненциальное неравенство Чебышева). *Для любого* $h > 0$

$$\mathbf{P}\{\omega\colon X(\omega) \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\,e^{hX}}{e^{h\varepsilon}}.$$

Доказательство. Действительно,

$$\mathbf{P}\{X(\omega) \geqslant \varepsilon\} = \mathbf{P}\{e^{hX} \geqslant e^{h\varepsilon}\}.$$

Применяя теперь лемму 6.1 к случайной величине $Y = e^{hX}$, получим нужное неравенство. $\square$

Замечание. Оба неравенства удобны для изучения распределений сумм независимых случайных величин, так как дисперсии в этом случае складываются, а производящие функции моментов перемножаются.

[1] Отклонением X (может быть, неудачный, но принятый термин) называют разность $X - \mathbf{E}X$.

6.3. Отклонения сумм независимых случайных величин

Применим неравенство Чебышева к суммам независимых случайных величин.

Пусть $X_1, \ldots, X_N$ — независимые случайные величины, для которых $\mathbf{E}\,X_k = a_k, \quad \mathbf{D}\,X_k = b_k^2$. Введем следующие обозначения:

$$S_N = X_1 + \ldots + X_N,$$
$$A_N = \mathbf{E}\,S_N = a_1 + \ldots + a_N,$$
$$B_N = \mathbf{D}\,S_N = b_1^2 + \ldots + b_N^2.$$

Тогда

$$\mathbf{P}\left\{\omega\colon |S_N - A_N| \geqslant t\sqrt{B_N}\right\} \leqslant \frac{1}{t^2}. \tag{6.2}$$

В действительности во многих случаях вероятность (6.2) при больших значениях t можно оценить гораздо точнее.

Пример 6.1. Рассмотрим независимые случайные величины $X_1, \ldots, X_N$, каждая из которых принимает значения 1 и -1 с вероятностями

$$\mathbf{P}\{X_k = 1\} = \mathbf{P}\{X_k = -1\} = \frac{1}{2}.$$

Для этих случайных величин $A_N = \mathbf{E}\,S_N = 0$, $B_N = \mathbf{D}\,S_N = N$. Но тогда

$$\mathbf{P}\{\omega\colon |S_N| \geqslant t\sqrt{N}\} \leqslant \frac{1}{t^2}.$$

В данном случае, как это будет показано ниже, вероятность можно оценить сверху величиной $2e^{-t^2/2}$, что при больших и даже не слишком больших t гораздо меньше. Так, для $t = 4$ имеем $\frac{1}{t^2} = 0.0625$, а $2e^{-t^2/2} = 0.000671$, при $t = 5$ соответственно $\frac{1}{t^2} = 0.04$, $e^{-t^2/2} = 7.453 \cdot 10^{-6}$.

Для получения более точной оценки понадобится экспоненциальное неравенство Чебышева. Заметим, что рассматриваемые в примере случайные величины имеют симметричное распределение.

Определение 6.1. Случайная величина X имеет *симметричное распределение*, если совпадают распределения случайных величин $Y = -X$ и X.

В силу симметричности распределений отдельных слагаемых X_k сумма S_N также имеет симметричное распределение[1]. Но тогда

$$\mathbf{P}\{\omega\colon |S_N| \geqslant k\} = 2\mathbf{P}\{\omega\colon S_N \geqslant k\}$$

для любого положительного числа k. Следовательно, при всех $h > 0$

$$\mathbf{P}\left\{\omega\colon |S_N| \geqslant t\sqrt{N}\right\} \leqslant 2\frac{\mathbf{E}\, e^{hS_N}}{e^{ht\sqrt{N}}}. \qquad (6.3)$$

Вычислим производящую функцию моментов:

$$\mathbf{E}\, e^{hS_N} = \prod_{j=1}^{N}\left(\frac{1}{2}e^{h} + \frac{1}{2}e^{-h}\right) = \prod_{j=1}^{N}\operatorname{ch} h = (\operatorname{ch} h)^N,$$

где $\operatorname{ch} h$ обозначает косинус гиперболический числа h. Перепишем (6.3) в виде

$$\mathbf{P}\left\{|S_N| \geqslant t\sqrt{N}\right\} \leqslant 2e^{-ht\sqrt{N}} \cdot (\operatorname{ch} h)^N = 2f(h).$$

Выберем теперь h таким образом, чтобы минимизировать $f(h)$. Поскольку минимумы функций $f(h)$ и $\ln f(h)$ достигаются одновременно, то достаточно это сделать для

$$g(h) = \ln f(h) = -ht\sqrt{N} + N\ln\operatorname{ch} h = N\left(-h\frac{t}{\sqrt{N}} + \ln\operatorname{ch} h\right).$$

Точное решение задачи легко выписать, но вид этого решения неудобен для дальнейшего анализа. Можно поступить следующим образом: оценить $g(h)$ сверху более простой функцией и выбрать число h, минимизирующее эту более простую функцию.

Обозначим $\varepsilon = \dfrac{t}{\sqrt{N}}$, тогда функцию $g(h)$ можно переписать следующим образом:

$$g(h) = N(-\varepsilon h + \ln\operatorname{ch} h).$$

Вычислим первые две производные функции $\ln\operatorname{ch} h = u(h)$:

$$u'(h) = \frac{\operatorname{sh} h}{\operatorname{ch} h}, \quad u''(h) = \frac{\operatorname{ch}^2 h - \operatorname{sh}^2 h}{\operatorname{ch}^2 h} = \frac{1}{\operatorname{ch}^2 h}.$$

---

[1] В нашем случае это почти очевидно. Доказательство данного факта для симметрично распределенных случайных величин общего вида содержится в лемме 8.3 главы 8.

Поскольку при всех h выполняется $\operatorname{ch} h \geqslant 1$, то

$$u(h) = u(0) + hu'(0) + \frac{h^2}{2}u''(\theta h) \leqslant \frac{h^2}{2} \quad (\theta \in (0,1)).$$

Следовательно,

$$g(h) \leqslant N\left(-\varepsilon h + \frac{h^2}{2}\right).$$

Выберем теперь $h = \varepsilon$, чтобы минимизировать правую часть неравенства. Получим

$$\mathbf{P}\left\{|S_N| \geqslant t\sqrt{N}\right\} \leqslant 2e^{-N\varepsilon^2/2}.$$

Вспоминая теперь, что $\varepsilon = \dfrac{t}{\sqrt{N}}$, получаем необходимое неравенство

$$\mathbf{P}\left\{|S_N| \geqslant t\sqrt{N}\right\} \leqslant 2e^{-t^2/2}.$$

Замечание. Доказательство последнего неравенства более полно использовало независимость случайных величин, а именно свойство мультипликативности производящей функции моментов, а также более сильное экспоненциальное неравенство Чебышева. Это позволило в конечном итоге получить гораздо более сильный результат, чем в неравенстве (6.2).

Пример 6.2. Рассмотрим генуэзскую лотерею с точки зрения игрока. Допустим, что игрок делает ставку на пару номеров (i,j). Выигрыш игрока в k-м тираже составляет

$$X_k = \begin{cases} 270M & \text{с вероятностью} \quad p = \dfrac{2}{801}, \\ 0 & \text{с вероятностью} \quad 1-p = \dfrac{799}{801}, \end{cases}$$

где M — ставка игрока. Обозначим $X'_k = X_k - M$ — изменение капитала игрока после k-го тиража, тогда суммарное изменение капитала после N тиражей запишется как

$$S'_N = X'_1 + X'_2 + \ldots + X'_N.$$

Поскольку

$$\begin{aligned} \mathbf{E}X'_k &= \mathbf{E}X_k - M = \frac{540}{801}M - M = -\frac{29}{89}M \approx -0.3258M, \\ \mathbf{D}X'_k &= \mathbf{D}X_k = 270^2 \cdot M^2 pq, \\ \sigma_{X'_k} &= 270 \cdot M\sqrt{pq} \approx 13.4747 \cdot M, \end{aligned}$$

то

$$\mathbf{E}S_{N'} \approx -0.3258 \cdot MN,$$
$$\sigma'_N = \sqrt{\mathbf{D}S'_N} \approx 13.4747 \cdot M\sqrt{N}.$$

Выберем $t > 0$. Тогда из неравенства Чебышева имеем

$$\mathbf{P}\{\omega\colon |S'_N - \mathbf{E}S'_N| < t\sigma'_N\} \geqslant 1 - \frac{1}{t^2}.$$

Раскрыв модуль, получаем, что для любого $t > 0$

$$\mathbf{P}\{\mathbf{E}S'_N - t\sigma'_N < S'_N < \mathbf{E}S'_N + t\sigma'_N\} \geqslant 1 - \frac{1}{t^2}.$$

Для игрока важен ответ на вопрос: при каком N вероятность отрицательных значений суммарного изменения капитала (то есть проигрыша) близка к 1? Например, если $t = 5$, то $1 - \frac{1}{t^2} = 0.96$. Но тогда с вероятностью, не меньшей 0.96, выполняется неравенство

$$-0.3258MN - 5 \cdot 13.4747 \cdot M\sqrt{N} < S'_N < -0.3258MN + 5 \cdot 13.4747M\sqrt{N}.$$

Невыгодность игры для отдельного игрока становится заметной не сразу, хотя ясно, что начиная с некоторого N ($N > (\frac{5 \cdot 13.4747}{0.3258})^2 = 42763.8$) правая часть неравенства становится отрицательной. На самом деле до разорения понадобится N значительно меньше. Для устроителей лотереи игра выгодна, поскольку участвуют в каждом розыгрыше много участников и в этом случае можно считать N изначально большим.

Изменение капитала игрока можно изобразить графически. Обозначим через S'_0 начальный капитал игрока. После каждого тиража, закончившегося проигрышем игрока, капитал игрока уменьшается на M единиц, после выигрыша возрастает сразу на $269M$ единиц.

На рис. 1 случайная ломаная показывает значение капитала игрока после каждого тиража. Первые десять тиражей, как это видно по рисунку, закончились проигрышем, зато в следующем тираже игрок выиграл. Если ломаная достигает в какой-либо момент оси абсцисс, то это означает, что в данный момент произошло разорение игрока.

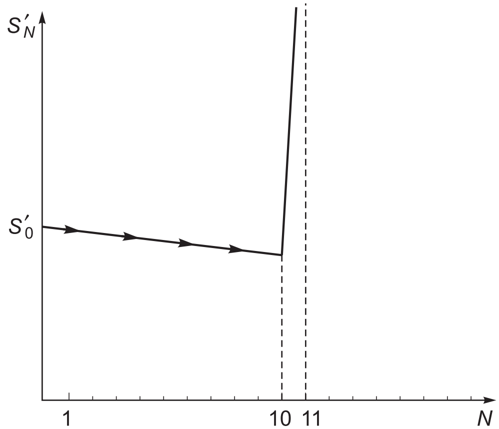

Рис. 1. График изменения капитала игрока с ростом числа тиражей

Аналогичный вид имеют траектории, описывающие деятельность страховых компаний. Несколько позже мы рассмотрим подробно задачу о разорении игрока при игре в орлянку, где также появятся похожего вида траектории.

Глава 7

ЗАКОН БОЛЬШИХ ЧИСЕЛ

В этой главе мы применим неравенство Чебышева к доказательству предельных теорем, касающихся поведения арифметического среднего случайных величин.

Функции от большого числа случайных переменных часто бывают «почти постоянными». В теории вероятностей утверждения о предельном постоянстве каких-либо случайных величин относят к законам больших чисел. В нашей книге мы познакомимся с несколькими такими утверждениями[1]. Теоремы подобного рода очень важны для приложений теории вероятностей, поскольку позволяют установить достаточные условия, при которых можно считать соответствующие случайные величины практически неизменными и на этом основании строить расчеты и прогнозы. С другой стороны, многие предельные теоремы позволяют глубже понять содержательную сторону теории вероятностей, объясняют смысл таких важных понятий, как вероятность случайного события, математическое ожидание случайной величины.

В дальнейшем мы будем рассматривать последовательности независимых случайных величин. Сразу же заметим, что такие объекты существуют. Всегда можно указать вероятностное пространство, на котором определены независимые случайные величины с заданными законами распределения[2].

7.1. Закон больших чисел в форме Чебышева

Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин.

Обозначим

$$\overline{S_N} = \frac{S_N}{N} = \frac{X_1 + \ldots + X_N}{N}$$

[1] См. главу 10.

[2] Этот факт следует из теоремы А.Н. Колмогорова [14, гл. III, § 4]. О последовательностях независимых случайных величин и событий см. также главу 10.

арифметическое среднее случайных величин. Тогда

$$\mathbf{E}\overline{S_N} = \frac{\mathbf{E}X_1 + \ldots + \mathbf{E}X_N}{N}, \quad \mathbf{D}\overline{S_N} = \frac{\mathbf{D}X_1 + \ldots + \mathbf{D}X_N}{N^2}.$$

Оказывается, что при достаточно общих предположениях о распределениях отдельных слагаемых X_k их арифметическое среднее $\overline{S_N}$ «почти постоянно» в том смысле, что если N достаточно велико, то значения $\overline{S_N}$ мало отличаются от константы $\mathbf{E}\overline{S_N}$.

Теорема 7.1 (Чебышев). *Пусть* $X_1, X_2, \ldots$ *— последовательность независимых случайных величин, дисперсии которых равномерно ограничены некоторой константой* $c > 0$: $\mathbf{D}X_k \leqslant c$ $(k = 1, 2, \ldots)$. *Тогда для любого* $\varepsilon > 0$ *при* $N \to \infty$

$$\mathbf{P}\left\{|\overline{S_N} - \mathbf{E}\overline{S_N}| \geqslant \varepsilon\right\} \to 0. \tag{7.1}$$

Доказательство. В силу неравенства Чебышева

$$\mathbf{P}\left\{|\overline{S_N} - \mathbf{E}\overline{S_N}| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{D}\overline{S_N}}{\varepsilon^2} = \frac{\mathbf{D}X_1 + \ldots + \mathbf{D}X_N}{\varepsilon^2 N^2} \leqslant \frac{cN}{\varepsilon^2 N^2} \to 0.$$

□

Утверждение (7.1) теоремы можно переписать иначе:

$$\mathbf{P}\left\{|\overline{S_N} - \mathbf{E}\overline{S_N}| \leqslant \varepsilon\right\} \to 1 \tag{7.2}$$

при $N \to \infty$.

Если для последовательности случайных величин выполнено (7.1) или (7.2), то говорят, что для нее выполняется *закон больших чисел.*

Замечание. Оба утверждения, (7.1) и (7.2), означают одно: разность между арифметическим средним случайных величин и арифметическим средним математических ожиданий этих случайных величин с ростом N стремится к 0. В теории вероятностей для такого вида сходимости используется специальное название: *сходимость по вероятности.*

Определение 7.1. Если $\forall \varepsilon > 0$ при $N \to \infty$

$$\mathbf{P}\{\omega\colon |X_N(\omega)| > \varepsilon\} \to 0,$$

то говорят, что X_N *сходятся к* 0 *по вероятности*, и в этом случае пишут

$$X_N(\omega) \xrightarrow{\mathbf{P}} 0.$$

Таким образом, утверждение теоремы можно переформулировать как

$$\overline{S_N} - \mathbf{E}\overline{S_N} \xrightarrow{\mathbf{P}} 0.$$

Следствие 7.1. *Пусть случайные величины $X_1, X_2, \ldots$ независимы, имеют одинаковое распределение с математическим ожиданием $\mathbf{E}X_k = a$ и $\mathbf{D}X_k = \sigma^2$. Тогда для всех $\varepsilon > 0$*

$$\mathbf{P}\{|\overline{S_N} - a| \leqslant \varepsilon\} \to 1.$$

Из последнего следствия вытекает, что арифметическое среднее значений случайной величины, наблюдаемых в независимых и проводимых в одинаковых условиях опытах, близко к математическому ожиданию. Это является способом практического оценивания неизвестного математического ожидания случайной величины.

Замечание. Утверждение теоремы и способ доказательства остаются справедливыми и для произвольных независимых случайных величин (не обязательно с конечным числом значений), если потребовать от них наличия конечных моментов первого и второго порядков. Из доказательства теоремы также видно, что условия на случайные величины можно несколько ослабить.

Введем числовые характеристики, отражающие взаимное влияние случайных величин друг на друга. Это *ковариация* и *коэффициент корреляции* случайных величин. Их обозначают соответственно $\operatorname{cov}(X,Y)$ и $\rho(X,Y)$.

Определение 7.2. *Ковариацией* двух случайных величин X, Y называется число $\operatorname{cov}(X,Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y)$.

Определение 7.3. *Коэффициентом корреляции* двух случайных величин X и Y с ненулевыми дисперсиями называется число

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Для независимых случайных величин ковариация и коэффициент корреляции равны 0. Если же $\operatorname{cov}(X,Y) = 0$, то такие случайные величины называют *некоррелированными*.

Задача 7.1. Приведите пример некоррелированных случайных величин, не являющихся независимыми.

Легко показать, что некоррелированность случайных величин означает, что $\mathbf{E}XY = \mathbf{E}X \cdot \mathbf{E}Y$.

Задача 7.2. Докажите, что если случайные величины $X_1, X_2, \ldots$ попарно некоррелированы[1], имеют равномерно ограниченные дисперсии, то для них также справедливо предельное соотношение (7.1).

Задача 7.3. Покажите, что утверждение теоремы 7.1 останется справедливым, если требование независимости заменить условием

$$\operatorname{cov}\left(X_i, X_j\right) \to 0$$

равномерно при $|i-j| \to \infty$.

7.2. Теорема Бернулли. Отклонение частоты наступления события от его вероятности

Частным случаем закона больших чисел является теорема Я. Бернулли, опубликованная в 1713 г., намного раньше теоремы Чебышева, и доказанная с помощью прямого подсчета вероятностей[2].

Теорема 7.2. *Пусть $A_1, A_2, \ldots$ — последовательность независимых событий с одинаковыми вероятностями $\mathbf{P}(A_k) = p$, $k = 1, 2, \ldots$, $\mu_N(\omega)$ — число наступивших среди первых N событий A_i. Тогда при любом $\varepsilon > 0$ и $N \to \infty$*

$$\mathbf{P}\left\{\omega\colon \left|\frac{\mu_N(\omega)}{N} - p\right| \geqslant \varepsilon\right\} \to 0.$$

Таким образом, если число испытаний велико, то относительная частота наступления случайных событий становится близка к вероятности этих событий. Этот факт позволяет в практических задачах использовать в качестве оценки неизвестной вероятности ее приближенное значение: $p \approx \mu_N/N$.

Доказательство. Поскольку μ_N представляет собой сумму независимых случайных величин $X_k = I_{A_k}$, для которых $\mathbf{E}X_k = p$, $\mathbf{D}X_k = pq$, то применима теорема 7.1, и утверждение теоремы Бернулли является следствием более общей теоремы Чебышева. □

[1] *Попарная некоррелированность* означает, что для любых двух случайных величин X_i, X_j $(i \neq j)$ их ковариация равна 0.

[2] В [1] воспроизводится часть сочинения Я. Бернулли «Искусство предположений», где впервые был изложен закон больших чисел.

Неравенство Чебышева позволяет также оценить отклонения относительной частоты наступления события от его вероятности, а именно:

$$\mathbf{P}\left\{\omega\colon \left|\frac{\mu_N}{N}-p\right| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{D}(\mu_N/N)}{\varepsilon^2} = \frac{Npq}{N^2\varepsilon^2} \leqslant \frac{1}{4N^2\varepsilon^2}. \qquad (7.3)$$

При выводе неравенства мы воспользовались тем, что $\mathbf{D}\mu_N = Npq$, а величина $pq = p(1-p)$ при $0 < p < 1$ достигает максимального значения в точке $p = 1/2$.

Замечание. В действительности правую часть неравенства (7.3) можно заметно усилить. Так, в учебнике [30, с. 81] приводится следующая оценка:

$$\mathbf{P}\left\{\omega\colon \left|\frac{\mu_N}{N}-p\right| \geqslant \varepsilon\right\} \leqslant 2e^{-2N\varepsilon^2}.$$

Как мы уже упоминали, Я. Бернулли доказал свою теорему задолго до появления теоремы П. Чебышева. Выписывая точные значения вероятностей

$$\mathbf{P}\{|\mu_N - Np| < N\varepsilon\} = \sum_{Np-N\varepsilon<k<Np+N\varepsilon} C_N^k p^k q^{N-k} \qquad (7.4)$$

при различных значениях p, он обнаружил, что биномиальные вероятности быстро убывают при удалении от Np. Так, при $p = q = 1/2$ величины биномиальных вероятностей равны $C_N^k/2^N$, поведение вероятностей определяется числами C_N^k.

Приведем для иллюстрации таблицу биномиальных коэффициентов до $N \leqslant 10$, известную как треугольник Паскаля (табл. на с. 64).

Мы видим, что поведение биномиальных коэффициентов всегда одинаково: сначала они возрастают, достигают максимального значения, а затем убывают, при этом средние коэффициенты значительно больше остальных. Еще более это заметно при больших N. Так, при $N = 20$

k	0	1	2	3	4	5	6	7	8	9	10
C_N^k	1	20	190	1140	4845	15 504	38 760	77 520	125 970	167 960	184 756

средние 5 коэффициентов (мы учли их симметричность) гораздо больше всех остальных. Таким образом, возвращаясь к вероятностям в распределении частоты успехов, замечаем, что значительная часть всей вероятности сосредоточена вокруг значения

Треугольник Паскаля

k	0	1	2	3	4	5	6	7	8	9	10
$N=$ 1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

$k=10$. То же самое верно и в общем случае. Заметив этот факт, Бернулли смог показать, что вероятность в (7.4) стремится к 1 при $N\to\infty$.

7.3. Вероятностное доказательство теоремы Вейерштрасса

Применим закон больших чисел для доказательства теоремы Вейерштрасса о равномерном приближении полиномами функции, непрерывной на отрезке.

Пусть $f(x)$ — непрерывная функция на $[0,1]$. Составим многочлены Бернштейна

$$B_n(x;f)=\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k} f\left(\frac{k}{n}\right).$$

Обозначим через μ_n число успехов в n испытаниях Бернулли с вероятностью успеха x в отдельном испытании. Тогда

$$B_n(x;f)=\mathbf{E}f\left(\frac{\mu_n}{n}\right).$$

Задача 7.4. Покажите, что:

1) если $f(x)=x$, то $B_n(X;f)=x$;

2) если $f(x)=x^2$, то $B_n(x;f)=\dfrac{(n-1)x^2}{n}+\dfrac{x}{n}$,

3) если $f(x)=e^x$, то $B_n(x;f)=(1+x(e^{1/n}-1))^n$.

Теорема 7.3. *Полиномы Бернштейна равномерно на* $[0,1]$ *сходятся к* $f(x)$ *при* $n\to\infty$.

Доказательство. Из свойств непрерывной на отрезке функции следует, что:

1) $f(x)$ ограничена и, значит, существует константа $M_f = \max\limits_{0\leqslant x\leqslant 1}|f(x)|$;

2) $f(x)$ равномерно непрерывна на $[0,1]$.

Рассмотрим величину

$$\omega(\delta;f)=\sup_{|x'-x''|<\delta}|f(x')-f(x'')|,$$

называемую *модулем непрерывности* функции $f(x)$. Равномерная непрерывность функции на отрезке $[0,1]$ эквивалентна тому, что $\omega(\delta;f)$ монотонно стремится к 0 при $\delta\to 0$.

Поскольку $\sum\limits_{k=0}^{n} C_n^k x^k(1-x)^{n-k}=1$ при всех $x\in[0,1]$, то

$$f(x)\equiv\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k}f(x).$$

Но тогда

$$f(x)-B_n(x;f)=\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k}(f(x)-f(k/n)).$$

Следовательно,

$$|f(x)-B_n(x;f)|\leqslant\Sigma_1+\Sigma_2,$$

где

$$\Sigma_1=\sum_{k:\ |k-nx|<n\delta} C_n^k x^k(1-x)^{n-k}|f(x)-f(k/n)|,$$

$$\Sigma_2=\sum_{k:\ |k-nx|\geqslant n\delta} C_n^k x^k(1-x)^{n-k}|f(x)-f(k/n)|.$$

Оценим каждую сумму по отдельности:

$$\Sigma_2\leqslant 2M_f\sum_{k:\ |k/n-x|\geqslant\delta} C_n^k x^k(1-x)^{n-k}=2M_f\mathbf{P}\left\{\left|\frac{\mu_n}{n}-x\right|\geqslant\delta\right\}\leqslant$$

$$\leqslant 2M_f\frac{\mathbf{D}\mu_n}{n^2\delta^2}\leqslant\frac{2M_f}{4n\delta^2},$$

$$\Sigma_1 \leqslant \left(\sum_{k:\ |k/n-x|<\delta} C_n^k x^k (1-x)^{n-k} \right) \omega(\delta; f) \leqslant \omega(\delta; f).$$

Таким образом,

$$|f(x) - B_n(x; f)| \leqslant \omega(\delta; f) + \frac{M_f}{2n\delta^2}.$$

Пусть $\varepsilon > 0$ — произвольное положительное число. Из равномерной непрерывности функции $f(x)$ следует, что существует δ_0 такое, что $\omega(\delta_0; f) < \frac{\varepsilon}{2}$. Выберем теперь n_0 так, чтобы при всех $n \geqslant n_0$ выполнялось неравенство $\frac{M_f}{2n\delta_0^2} < \frac{\varepsilon}{2}$. Но тогда, начиная с номера n_0, одновременно для всех $x \in [0, 1]$

$$|f(x) - B_n(x; f)| < \varepsilon,$$

то есть многочлены Бернштейна равномерно сходятся к функции $f(x)$. □

Замечание. Из этой теоремы можно получить вторую теорему Вейерштрасса о равномерном приближении непрерывной периодической функции тригонометрическими полиномами.

Глава 8

НЕРАВЕНСТВА ДЛЯ МАКСИМУМА СУММ НЕЗАВИСИМЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

До сих пор мы рассматривали вероятности, связанные с отдельными суммами случайных величин. Но иногда приходится иметь дело с последовательностями частичных сумм.

Пусть $X_1(\omega), X_2(\omega), \ldots$ — последовательность независимых случайных величин, заданных на некотором вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$. Введем новую последовательность случайных величин

$$\begin{aligned}
&S_0(\omega) = 0,\\
&S_1(\omega) = X_1(\omega),\\
&S_2(\omega) = X_1(\omega) + X_2(\omega),\\
&\ \vdots\\
&S_N(\omega) = X_1(\omega) + \ldots + X_N(\omega),\\
&\ldots\ldots\ldots\ldots,
\end{aligned}$$

называемых частичными суммами.

Последовательность $S_0(\omega), S_1(\omega), S_2(\omega), \ldots$ будем рассматривать как некий процесс, развивающийся во времени.

Замечание. В теории вероятностей всякое семейство случайных величин $X_t(\omega)$, заданных на одном вероятностном пространстве и зависящих от переменной $t \in T$, называют *случайным процессом*, а переменную t часто трактуют как время. Введенная выше последовательность случайных величин — это случайный процесс с дискретным временем.

Если зафиксировать элементарный исход ω, то мы получим числовую последовательность $S_0(\omega), S_1(\omega), S_2(\omega), \ldots,$ называемую траекторией случайного процесса. Поведение траектории (развитие процесса во времени) удобно изображать графически (рис. 2). Ломаная с вершинами (N, S_N) наглядно изображает изменения сумм с накоплением слагаемых. Практический интерес для многих приложений теории вероятностей представляют задачи о вероятностях нахождения сумм в заданных границах, а также о вероятностях выхода за какую-либо границу.

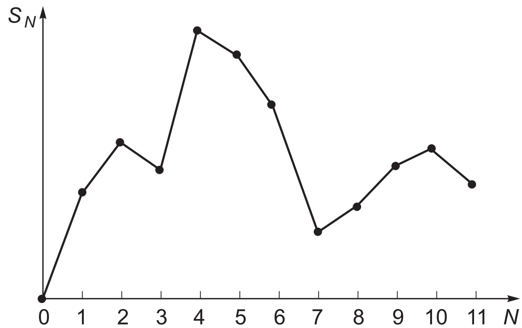

Рис. 2. Поведение частичных сумм

Вернемся к генуэзской лотерее. В примере 6.2 мы обозначили через S_0' начальный капитал игрока, который делает ставку M на два номера (амбо) и в случае выигрыша получает $270M$. На рис. 1 (с. 58) изображена одна из возможных траекторий процесса $S_o', S_1', S_2', \ldots$, показывающая изменение капитала игрока. Проигрыш игрока в каком-либо тираже означает уменьшение капитала на M единиц, а выигрыш — увеличение на $269M$ единиц. Если график достигает в какой-то момент оси абсцисс, происходит разорение игрока.

8.1. Неравенство А. Н. Колмогорова

Всюду в дальнейшем мы будем иметь дело со случайными величинами, определенными на одном вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$. Для краткости иногда будем опускать аргумент случайных величин ω.

Теорема 8.1 (неравенство Колмогорова). *Пусть X_1, X_2, ... — последовательность независимых случайных величин. Тогда*

$$\mathbf{P}\left\{\omega\colon \max_{1\leqslant k\leqslant n} |S_k - \mathbf{E}S_k| \geqslant \varepsilon\right\} \leqslant \frac{\mathbf{D}S_n}{\varepsilon^2}. \tag{8.1}$$

Доказательство. 1. Не ограничивая общности рассуждений, можно считать $\mathbf{E}X_k = 0$ для любого k. Если это не так, то можно ввести

$$X_k' = X_k - \mathbf{E}X_k, \quad S_k' = X_1' + \ldots + X_k',$$

для которых это условие выполнено. Тогда $\mathbf{D}X_k = \mathbf{D}X_k' = \mathbf{E}(X_k')^2$, и утверждение теоремы (8.1) примет вид

$$\mathbf{P}\left\{\omega\colon \max_{1\leqslant k\leqslant n}|S'_k|\geqslant\varepsilon\right\}\leqslant\frac{\mathbf{E}(S'_n)^2}{\varepsilon^2}. \tag{8.2}$$

2. Введем события

$$\begin{aligned}
&F_1=\{\omega\colon |S_1|\geqslant\varepsilon\},\\
&F_2=\{\omega\colon |S_1|<\varepsilon,\ \ |S_2|\geqslant\varepsilon\},\\
&\vdots\\
&F_n=\{\omega\colon |S_1|<\varepsilon,\ \ |S_2|<\varepsilon,\ \ \ldots,\ \ |S_{n-1}<\varepsilon,\ \ |S_n|\geqslant\varepsilon\},\\
&F_0=\{\omega\colon |S_1|<\varepsilon,\ \ \ldots,\ \ |S_n|<\varepsilon\}.
\end{aligned}$$

Эти события попарно не пересекаются, то есть $F_i\cap F_j=\varnothing$, если $i\neq j$. Кроме того, $\bigcup\limits_{k=0}^{n}F_k=\Omega$. Обозначим $I_k(\omega)=I_{F_k}(\omega)$, тогда $\sum\limits_{k=0}^{n}I_k(\omega)=1$, а для события

$$F=\bigcup_{k=1}^{n}F_k=\left\{\omega\colon \max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon\right\}$$

$I_F(\omega)=\sum\limits_{k=1}^{n}I_k(\omega)$.

Заметим, что $I_k(\omega)$ — это функция от частичных сумм $S_1,S_2,\ldots,S_k$, то есть это функция, зависящая от $X_1,X_2,\ldots,X_k$ и не зависящая от случайных величин $X_{k+1},\ldots,X_n$.

Лемма 8.1. *При* $1\leqslant k\leqslant n$ *и всех* ω

$$S_k^2(\omega)\cdot I_k(\omega)\geqslant\varepsilon^2\cdot I_k(\omega).$$

Доказательство. Утверждение почти очевидно. Действительно, если $\omega\in\overline{F_k}$, то это верно, поскольку $I_k(\omega)=0$. Если же $\omega\in F_k$, то $|S_k(\omega)|\geqslant\varepsilon$, $I_k(\omega)=1$, и неравенство также выполняется. □

Вернемся к доказательству теоремы. Нетрудно видеть, что

$$S_n^2=(S_k+(S_n-S_k))^2=S_k^2+2S_k(S_n-S_k)+(S_n-S_k)^2\geqslant$$
$$\geqslant S_k^2+2S_k(S_n-S_k).$$

Но тогда

$$S_n^2 = S_n^2 \cdot \sum_{k=0}^{n} I_k \geqslant \sum_{k=1}^{n-1} S_n^2 I_k + S_n^2 I_n \geqslant \sum_{k=1}^{n-1} S_k^2 I_k + S_n^2 I_n + 2\sum_{k=1}^{n-1} S_k I_k (S_n - S_k),$$

и в силу леммы

$$S_n^2 \geqslant \varepsilon^2 \sum_{k=1}^{n} I_k + 2\sum_{k=1}^{n-1} S_k I_k (S_n - S_k).$$

Переходя к математическим ожиданиям, получим

$$\mathbf{E}\, S_n^2 \geqslant \varepsilon^2 \sum_{k=1}^{n} \mathbf{E}\, I_k + 2\sum_{k=1}^{n} \mathbf{E}\big((S_k I_k)(S_n - S_k)\big).$$

Покажем, что вторая сумма равна 0. Поскольку случайная величина $S_k I_k$ зависит от $X_1, \ldots, X_k$, а $S_n - S_k$ — от $X_{k+1}, \ldots, X_n$, то эти случайные величины независимы и для всех k

$$\mathbf{E}\big((S_k I_k)(S_n - S_k)\big) = \mathbf{E}(S_k I_k)\,\mathbf{E}(S_n - S_k) = 0.$$

Но тогда

$$\mathbf{E}\, S_n^2 \geqslant \varepsilon^2 \mathbf{P}(F),$$

что и доказывает теорему. □

Замечание. В этом доказательстве существенно использовалась взаимная независимость случайных величин, что позволило получить результат намного сильнее, чем в неравенстве Чебышева. Сравнивая эти два неравенства, видим, что неравенство Чебышева можно получить как следствие из неравенства Колмогорова.

8.2. Неравенство Поля Леви

Для симметрично распределенных величин вместо неравенства Колмогорова можно использовать неравенство П. Леви.

Определение 8.1. Случайная величина X распределена симметрично, если случайные величины X и $-X$ имеют одинаковое распределение.

Пример 8.1. Рассмотрим случайную величину X, принимающую два значения 1 и -1 с вероятностью 1/2. Не вызывает сомнения, что X распределена симметрично.

Сформулируем два свойства симметрично распределенных случайных величин.

Лемма 8.2. *Если случайная величина X имеет симметричное распределение, то*

$$\mathbf{P}\{X \geqslant 0\} = \mathbf{P}\{X \leqslant 0\} \geqslant \frac{1}{2}.$$

Доказательство. Действительно, поскольку $\{X \geqslant 0\} \cup \cup \{X \leqslant 0\} = \Omega$, то

$$\mathbf{P}\{X \geqslant 0\} + \mathbf{P}\{X \leqslant 0\} \geqslant 1.$$

Учитывая, что слагаемые в этой сумме равны в силу симметричности распределения, получим утверждение леммы. □

Лемма 8.3. *Если X, Y — независимые симметрично распределенные случайные величины, то случайная величина $Z = X + Y$ также имеет симметричное распределение.*

Доказательство. Действительно,

$$\mathbf{P}\{Z = z\} = \sum_x \mathbf{P}\{X = x, Y = z - x\} = \sum_x \mathbf{P}\{X = x\}\mathbf{P}\{Y = z - x\} =$$
$$= \sum_x \mathbf{P}\{-X = x\}\mathbf{P}\{-Y = z - x\} = \mathbf{P}\{-X - Y = z\} = \mathbf{P}\{-Z = z\}.$$

□

Теорема 8.2 (неравенство Поля Леви). *Пусть $X_1, \ldots, X_n$ — независимые симметрично распределенные случайные величины, тогда для всех a выполняется неравенство*

$$\mathbf{P}\{S_n \geqslant a\} \leqslant \mathbf{P}\left\{\omega\colon \max_{1 \leqslant k \leqslant n} S_k \geqslant a\right\} \leqslant 2\mathbf{P}\{S_n \geqslant a\}. \tag{8.3}$$

Доказательство. В идейном плане доказательство неравенства Леви похоже на доказательство неравенства Колмогорова. Введем события

$$\begin{aligned}
&D_1 = \{\omega\colon S_1(\omega) \geqslant a\},\\
&D_2 = \{\omega\colon S_1(\omega) < a, \quad S_2(\omega) \geqslant a\},\\
&\vdots\\
&D_n = \{\omega\colon S_1(\omega) < a, \quad \ldots, \quad S_{n-1}(\omega) < a, \quad S_n(\omega) \geqslant a\},
\end{aligned}$$

$$D = \bigcup_{k=1}^{n} D_k = \left\{\omega\colon \max_{1 \leqslant k \leqslant n} S_k \geqslant a\right\}.$$

События $D_1, \ldots, D_n$ попарно не пересекаются. Введем также события

$$\begin{aligned}
&E_1 = \{\omega\colon S_n(\omega) - S_1(\omega) \geqslant 0\},\\
&E_2 = \{\omega\colon S_n(\omega) - S_2(\omega) \geqslant 0\},\\
&\vdots\\
&E_{n-1} = \{\omega\colon S_n(\omega) - S_{n-1}(\omega) \geqslant 0\},\\
&E_n = D_n.
\end{aligned}$$

Для произвольного $k \leqslant n-1$ события D_k определяются по случайным величинам $X_1, \ldots, X_k$, а события E_k — по случайным величинам $X_{k+1}, \ldots, X_n$. Следовательно, они независимы. Рассмотрим новые события

$$C_k = D_k \cap E_k \subset D_k, \quad 1 \leqslant k \leqslant n-1, \quad C_n = D_n \cap E_n = D_n.$$

Заметим, что, как и события D_k, события C_k попарно не пересекаются.

Если $\omega \in C_k\,(k \leqslant n-1)$, то $S_k(\omega) \geqslant a$, $S_n - S_k \geqslant 0$, но это означает, что $S_n(\omega) \geqslant a$, следовательно, $C_k \subset \{\omega\colon S_n(\omega) \geqslant a\}$. Но тогда

$$\bigcup_{k=1}^{n} C_k \subset \{\omega\colon S_n(\omega) \geqslant a\}.$$

Отсюда следует, что

$$\sum_{k=1}^{n} \mathbf{P}(C_k) \leqslant \mathbf{P}\{\omega\colon S_n(\omega) \geqslant a\}. \tag{8.4}$$

Рассмотрим левую часть неравенства (8.4). Из независимости D_k и E_k получим

$$\sum_{k=1}^{n} \mathbf{P}(C_k) = \mathbf{P}(C_n) + \sum_{k=1}^{n-1} \mathbf{P}(C_k) = \mathbf{P}(C_n) + \sum_{k=1}^{n-1} \mathbf{P}(D_k)\mathbf{P}(E_k).$$

Поскольку из лемм следует, что $\mathbf{P}(E_k) \geqslant \frac{1}{2}$, то

$$\sum_{k=1}^{n} \mathbf{P}(C_k) \geqslant \frac{1}{2}\left(\sum_{k=1}^{n-1} \mathbf{P}(D_k) + \mathbf{P}(D_n)\right) = \frac{1}{2}\mathbf{P}(D).$$

Учитывая это, из (8.4) получаем

$$\frac{1}{2}\mathbf{P}(D) \leqslant \mathbf{P}\{S_n \geqslant a\}.$$

Умножив полученное неравенство на 2, приходим к (8.3), что и доказывает теорему. □

Следствие 8.1. *Если $a \geqslant 0$, то выполняется неравенство*

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant a\right\} \leqslant 2\mathbf{P}\{|S_n| \geqslant a\}. \tag{8.5}$$

Доказательство. Поскольку

$$\left\{\omega\colon \max_{1\leqslant k\leqslant n} |S_k| \geqslant a\right\} = \left\{\omega\colon \max_k S_k \geqslant a\right\} \cup \left\{\omega\colon \min_k S_k \leqslant -a\right\},$$

то

$$\mathbf{P}\left\{\max_k |S_k| \geqslant a\right\} \leqslant \mathbf{P}\left\{\max_k S_k \geqslant a\right\} + \mathbf{P}\left\{\min_k S_k \leqslant -a\right\}.$$

К первому слагаемому этой суммы можно непосредственно применить утверждение теоремы, второе же придется оценивать.

Наряду со случайными величинами X_k рассмотрим также $X'_k = -X_k$ и их суммы $S'_k = -S_k$, $1 \leqslant k \leqslant n$. Тогда, используя симметричность распределений этих случайных величин, получим

$$\mathbf{P}\left\{\min_k S_k \leqslant -a\right\} = \mathbf{P}\left\{\max_k (-S_k) \geqslant a\right\} = \mathbf{P}\left\{\max_k S'_k \geqslant a\right\} \leqslant$$
$$\leqslant 2\mathbf{P}\{S'_n \geqslant a\} = 2\mathbf{P}\{S_n \leqslant -a\}.$$

Следовательно,

$$\mathbf{P}\left\{\max_k |S_k| \geqslant a\right\} \leqslant 2\mathbf{P}\{S_n \geqslant a\} + 2\mathbf{P}\{S_n \leqslant -a\} = 2\mathbf{P}\{|S_n| \geqslant a\}.$$

□

Замечание. Следует обратить внимание на то, что неравенство Леви (8.3) дает оценку, имеющую правильный порядок по a. В этом легко убедиться, сравнивая левые и правые части неравенства.

Применим неравенство Леви для оценки вероятности разорения одного из игроков в орлянку.

Пример 8.2. Пусть A и B — два игрока, начальные капиталы которых составляют a и b рублей соответственно. Пусть игрок A играет с бесконечно богатым противником, т. е. $b = \infty$. С вероятностью $p = 1/2$ игрок A в каждой игре может выиграть 1 рубль и с вероятностью $q = 1/2$ проиграть такую же сумму.

Обозначим $S_k = X_1 + \ldots + X_k$ — суммарный выигрыш игрока A после N игр.

Изменение выигрыша удобно изображать графически с помощью траекторий, которые выходят из 0 и в каждый момент времени могут смещаться на единицу вверх или вниз. Пример такой траектории приводится на рис. 3. Разорение игрока наступает в тот момент, когда траектория достигает уровня $-a$, т. е. когда впервые $S_k = -a$.

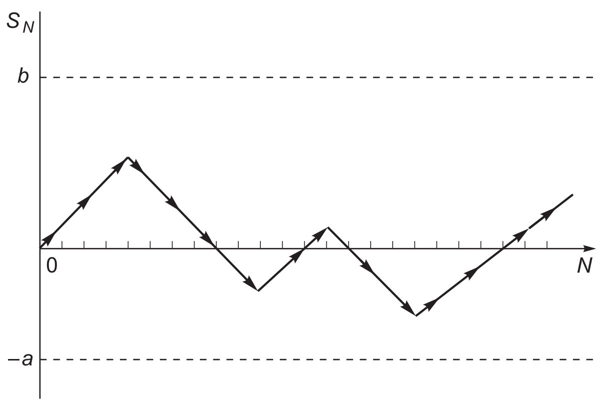

Рис. 3. К задаче о разорении игрока

Вероятность разорения до момента N есть $\mathbf{P}\left\{\min\limits_{1 \leqslant k \leqslant N} S_k \leqslant -a\right\}$. Поскольку

$$\mathbf{P}\left\{\min_{1 \leqslant k \leqslant N} S_k \leqslant -a\right\} = \mathbf{P}\left\{\max_{1 \leqslant k \leqslant N} S_k \geqslant a\right\},$$

то, применяя неравенство Леви, получим

$$\mathbf{P}\left\{\min_{1\leqslant k\leqslant N} S_k \leqslant -a\right\} \leqslant 2\mathbf{P}\{S_N \geqslant a\}.$$

Такие вероятности мы умеем оценивать (см. главу 6), а именно: при $a = t\sqrt{N}$

$$\mathbf{P}\left\{\min_{1\leqslant k\leqslant N} S_k \leqslant -t\sqrt{N}\right\} \leqslant 2\mathbf{P}\left\{S_N \geqslant t\sqrt{N}\right\} \leqslant 2e^{-t^2/2}.$$

Для иллюстрации возьмем конкретные числа: $t = 4$, $N = 625$. Тогда $2e^{-t^2/2} \approx 0.0007$, значит, вероятность разорения за 625 игр при начальном капитале всего лишь в 100 рублей очень невелика.

Ситуации, когда происходит игра с бесконечно богатым противником, встречаются довольно часто. Так в казино каждый из игроков играет против казино, капиталы которого много больше капитала отдельного игрока. С другой стороны, устроители казино играют с большим числом игроков, суммарный капитал которых тоже можно считать бесконечно большим. Аналогичное замечание касается также деятельности страховых компаний.

Глава 9

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

9.1. Общее определение вероятностного пространства

В главе 1 было введено понятие вероятностного пространства

$$(\Omega, \mathcal{A}, \mathbf{P})$$

как тройки элементов, в которой:

- Ω — некоторое множество элементарных исходов ω;
- $\mathcal{A}$ — класс подмножеств Ω, которые в дальнейшем назывались случайными событиями;
- $\mathbf{P}$ — вероятность или распределение вероятностей — неотрицательная функция, определенная на $\mathcal{A}$.

До сих пор мы рассматривали вероятностные пространства, в которых множество элементарных исходов конечно:

$$\Omega = \{\omega_1, \ldots, \omega_s\},$$

класс $\mathcal{A}$ — множество *всех подмножеств* Ω, а функция $\mathbf{P}$ определялась как

$$\mathbf{P}(A) = \sum_{\omega \in A} p(\omega),$$

где $p(\omega) \geqslant 0$ для всех ω и $\mathbf{P}(\Omega) = 1$, $\mathbf{P}(\varnothing) = 0$.

Для таких вероятностных пространств случайные величины — это любые функции, зависящие от элементарного исхода. Поскольку число различных значений любой случайной величины в этом случае конечно, то свойства математических ожиданий выводились без особого труда и в доказательствах этих свойств существенно использовалась конечность сумм. Доказательства и формулировки остальных теорем (неравенства Чебышева, Колмогорова, Леви, предельные теоремы) были приведены в форме, справедливой в общем случае.

Перейдем к общему определению вероятностного пространства. По-прежнему Ω — это множество элементарных исходов ω, но теперь уже произвольной мощности.

Подмножества Ω по-прежнему будем обозначать заглавными буквами начала латинского алфавита: $A, B, C, D, \ldots$, классы подмножеств — рукописными заглавными буквами $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \ldots$.

В общей теории нам встретятся следующие классы подмножеств Ω:

1) разбиения (конечные или счетные);
2) полукольца и полуалгебры;
3) алгебры;
4) σ-алгебры[1].

Определение 9.1. Непустые подмножества $A_1, A_2, \ldots$ образуют *разбиение* множества A, если они попарно не пересекаются ($A_i \cap A_j = \varnothing$, если $i \neq j$) и $A = \bigcup\limits_i A_i$.

Разбиение называется *конечным*, если оно состоит из конечного набора подмножеств $A_1, \ldots, A_n$, и *счетным*, если элементы этого разбиения $A_1, A_2, \ldots$ можно поставить во взаимно однозначное соответствие с натуральным рядом.

Определение 9.2. Класс $\mathcal{A}$ подмножеств Ω называется *алгеброй*, если выполнены следующие условия:

1) $\Omega \in \mathcal{A}$;
2) если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$;
3) если $A, B \in \mathcal{A}$, то $A \cup B \in \mathcal{A}$.

Другими словами, класс подмножеств множества Ω образует алгебру, если он содержит достоверное событие и если операции взятия дополнения и объединения событий не выводят из этого класса.

Задача 9.1. Докажите, что

1) невозможное событие $\varnothing \in \mathcal{A}$;
2) если $A, B \in \mathcal{A}$, то $A \cap B$, $A \triangle B$, $A \setminus B \in \mathcal{A}$;
3) если $A_1, \ldots, A_n \in \mathcal{A}$, то $\bigcup\limits_{i=1}^{n} A_i$, $\bigcap\limits_{i=1}^{n} A_i \in \mathcal{A}$ для любого натурального n.

Определение 9.3. Класс $\mathcal{A}$ называется *σ-алгеброй*, если выполняются свойства:

1) $\Omega \in \mathcal{A}$;

[1] Читается как «сигма-алгебры».

2) если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$;

3) для любой последовательности событий $A_1, A_2, \ldots \in \mathcal{A}$ выполнено $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A}$.

Задача 9.2. Докажите, что

1) всякая σ-алгебра является алгеброй,

2) σ-алгебра замкнута относительно конечных и счетных пересечений.

Определение 9.4. Класс $\mathcal{S}$ подмножеств Ω образует *полукольцо*, если выполнены свойства:

1) $\varnothing \in \mathcal{S}$;

2) если $A, B \in \mathcal{S}$, то $A \cap B \in \mathcal{S}$;

3) для любых $A, A_1 \in \mathcal{S}$ таких, что $A_1 \subset A$, существует конечное разбиение A, все элементы которого входят в $\mathcal{S}$ и одним из элементов которого является A_1.

Замечание. Последнее свойство в определении полукольца можно сформулировать по-другому: разность $A \setminus A_1$ любых двух элементов полукольца может быть представлена в виде объединения конечного числа попарно непересекающихся $A_i \in \mathcal{S}$. Заметим, однако, что сама эта разность полукольцу может и не принадлежать. Кроме того, это представление может оказаться не единственным.

Полукольцо подмножеств является, как мы это увидим ниже, очень полезным техническим средством при построении теории меры.

Рассмотрим несколько примеров наиболее часто используемых полуколец.

Пример 9.1. Класс $\mathcal{S}$ полуинтервалов $[a;b)$ $(a \leqslant b)$ на числовой прямой образует полукольцо.

Действительно, если $a = b$, то $[a;b) = \varnothing$ и, следовательно, невозможное событие является элементом этого класса. Легко видеть, что непустое пересечение двух полуинтервалов данного вида также является полуинтервалом из $\mathcal{S}$, а разность двух полуинтервалов, один из которых содержится в другом, представима в виде объединения не более чем двух непересекающихся полуинтервалов.

Пример 9.2. Класс $\mathcal{S}$ прямоугольников $[a;b) \times [c;d)$ на плоскости $(a \leqslant b,\ c \leqslant d)$ образует полукольцо (рис. 4).

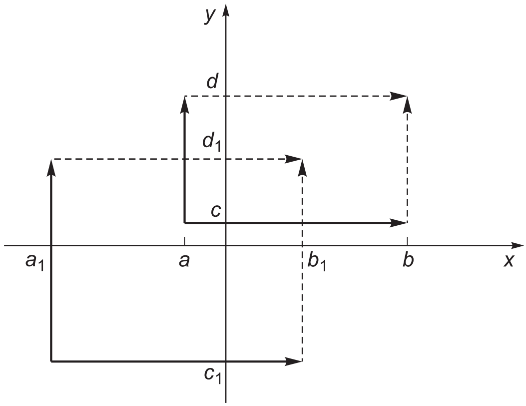

Рис. 4. Полукольцо прямоугольников на плоскости

Проверьте это самостоятельно.

Этот пример легко обобщается на случай произвольного евклидова пространства $\mathbb{R}^n$, где в качестве полукольца можно рассмотреть множество n-мерных параллелепипедов.

Понятие полуинтервалов можно распространить на случай, когда элементами Ω являются функции.

Пример 9.3. Полукольцо полуинтервалов в $\mathbb{C}[0;1]$ — множестве всех непрерывных на отрезке $[0;1]$ функций.

Выберем некоторое множество точек: $0 \leqslant t_1 < t_2 < \ldots < t_n \leqslant 1$, и столько же произвольных полуинтервалов: $[a_i;b_i)$ $(1 \leqslant i \leqslant n)$. Подмножества в $\mathbb{C}[0;1]$ вида

$$\{f(t)\colon\ a_1 \leqslant f(t_1) < b_1;\quad a_2 \leqslant f(t_2) < b_2;\quad \ldots,\quad a_n \leqslant f(t_n) < b_n\} \quad (9.1)$$

образуют полукольцо. Эти подмножества $\mathbb{C}[0;1]$ также называют полуинтервалами. На рис. 5 изображены несколько функций, попавших в полуинтервал.

Определение 9.5. Полукольцо $\mathcal{S}$ называется *полукольцом с единицей*, если $\Omega \in \mathcal{S}$.

Замечание. В примере 9.1, чтобы получить полукольцо с единицей, надо добавить полуинтервалы вида

$$(-\infty;b), \quad [a;+\infty), \quad (-\infty;+\infty),$$

то есть разрешить a, b принимать все значения, удовлетворяющие неравенству $-\infty \leqslant a \leqslant b \leqslant +\infty$.

Полукольцо с единицей также называют *полуалгеброй*.

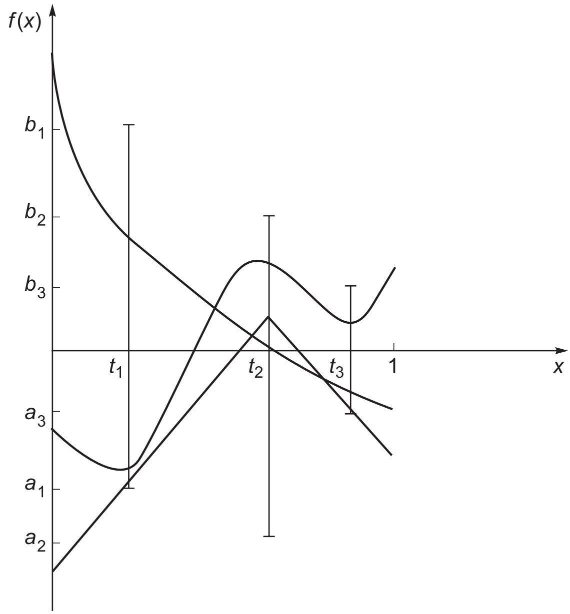

Рис. 5. Полуинтервалы функций

9.1.1. Порожденные алгебры и σ-алгебры

Лемма 9.1. *Для любого непустого класса $\mathcal{E}$ существует единственная алгебра $\mathcal{A}_0$ такая, что $\mathcal{E} \subset \mathcal{A}_0$ и для любой алгебры $\mathcal{A}_1$, содержащей $\mathcal{E}$, $\mathcal{A}_0 \subset \mathcal{A}_1$.*

Такую алгебру $\mathcal{A}_0$ будем обозначать $\mathcal{A}(\mathcal{E})$ и называть алгеброй, *порожденной* классом $\mathcal{E}$.

Сформулируем аналогичную лемму и для σ-алгебр.

Лемма 9.2. *Для любого непустого класса $\mathcal{E}$ существует единственная σ-алгебра $\mathcal{B}_0$ такая, что $\mathcal{E} \subset \mathcal{B}_0$ и для любой σ-алгебры $\mathcal{B}_1$, содержащей $\mathcal{E}$, выполняется $\mathcal{B}_0 \subset \mathcal{B}_1$.*

$\mathcal{B}_0$ называют σ-алгеброй, *порожденной классом* $\mathcal{E}$, это наименьшая σ-алгебра, содержащая класс $\mathcal{E}$.

Приведем доказательство первой леммы, вторая доказывается аналогично.

Доказательство. 1. Покажем сначала, что пересечение алгебр является алгеброй.

Пусть $\mathcal{A}_1, \mathcal{A}_2$ — две алгебры. Обозначим $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$.

Пусть $A, B \in \mathcal{A}$, но тогда $A, B \in \mathcal{A}_1$, $A, B \in \mathcal{A}_2$ и в силу определения алгебры события Ω, $\overline{A}$, $A \cup B$ принадлежат одновременно обеим алгебрам и, следовательно, принадлежат их пересечению $\mathcal{A}$. Значит, $\mathcal{A}$ является алгеброй.

Данное утверждение справедливо в более общем случае, когда пересечение $\bigcap\limits_{t \in T} \mathcal{A}_t$ берется по всем значениям t из непустого множества T, которое может содержать любое количество точек.

2. Определим теперь $\mathcal{A}_0$ как пересечение всех алгебр, содержащих $\mathcal{E}$. Поскольку для любого класса подмножеств $\mathcal{E}$ всегда найдется хотя бы одна алгебра, содержащая $\mathcal{E}$ (например, алгебра всех подмножеств Ω), то пересечение берется по непустой системе алгебр и, следовательно, $\mathcal{A}_0$ — минимальная алгебра, содержащая $\mathcal{E}$. $\square$

Приведенное доказательство ничего не говорит о конструкции алгебр, порожденных каким-либо классом подмножеств. Чтобы разобраться в этом вопросе, рассмотрим ряд примеров.

Пример 9.4. Алгебра, порожденная конечным разбиением Ω.

Пусть $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$, $A_i \cap A_j = \varnothing$ $(i \neq j)$. Тогда наименьшая алгебра, содержащая данное разбиение, должна содержать следующие подмножества Ω:

1) пустое подмножество $\varnothing$;
2) каждый из элементов разбиения: $A_1, A_2, \ldots, A_n$;
3) объединения любых двух элементов разбиения: $A_i \cup A_j$;
4) объединения любых трех элементов разбиения: $A_i \cup A_j \cup A_k$;

$\vdots$

5) объединение всех элементов разбиения $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Таким образом, всего элементов должно быть

$$1 + C_n^1 + C_n^2 + C_n^3 + \ldots + C_n^{n-1} + 1 = 2^n.$$

Так, например, разбиение, состоящее из четырех элементов, порождает алгебру из 16 элементов.

Замечание. Справедливо утверждение о том, что всякая конечная алгебра порождается некоторым разбиением, значит, число ее элементов может быть равным только степени числа 2.

Пример 9.5. Пусть A и B — произвольные два подмножества Ω, отличные от $\varnothing, \Omega$. Разберем, как устроена минимальная алгебра, содержащая A и B. Если $A=\overline{B}$, то минимальная алгебра содержит 4 элемента и равна

$$\mathcal{A}=\{\varnothing,\Omega,A,B\}.$$

Пусть теперь подмножества A и B таковы, что $A\cap B\neq\varnothing$, а $A\cup\cup B\neq\Omega$. С этими событиями свяжем следующее разбиение Ω:

$$\mathcal{E}=\left\{AB;\ A\overline{B};\ \overline{A}B;\ \overline{A}\overline{B}\right\},$$

каждый элемент которого должен принадлежать минимальной алгебре, содержащей A и B. Порожденная этим разбиением алгебра $\mathcal{A}(\mathcal{E})$ совпадает с минимальной алгеброй $\mathcal{A}$, содержащей A и B. Следовательно, $\mathcal{A}$ состоит из 16 элементов. В общем случае можно утверждать, что $\mathcal{A}$ состоит не более чем из $2^4 = 16$ элементов.

Задача 9.3. Приведите пример событий A и B таких, что минимальная алгебра $\mathcal{A}$, содержащая данные события, состоит из 8 элементов.

Столь же просто описываются алгебры, порожденные полукольцом с единицей.

Задача 9.4. Покажите, что алгебра, порожденная полукольцом с единицей, состоит из всех конечных объединений попарно непересекающихся элементов полукольца.

9.1.2. Борелевские σ-алгебры множеств

Пусть теперь множество элементарных исходов совпадает с s-мерным евклидовым пространством: $\Omega=\mathbb{R}^s$. В этом случае алгебры (σ-алгебры) событий связываются с наиболее типичными подмножествами $\mathbb{R}^s$, встречающимися в математическом анализе.

Пример 9.6. Рассмотрим на числовой прямой ($s = 1$) полукольцо полуинтервалов с ***1***.

Наименьшая σ-алгебра, содержащая все полуинтервалы вида $[a;b)$, где $-\infty\leqslant a\leqslant b\leqslant+\infty$, называется *борелевской* и обозначается $\mathcal{B}^1$ или $\mathcal{B}$. Нетрудно убедиться в том, что произвольные интервалы и отрезки числовой прямой принадлежат борелевской

σ-алгебре. Действительно,

$$(a;b) = \bigcup_{n=1}^{\infty} \left[a + \frac{1}{n}; b\right),$$

$$[a;b] = \bigcap_{n=1}^{\infty} \left[a; b + \frac{1}{n}\right).$$

Поскольку всякое открытое множество на прямой можно представить в виде объединения не более чем счетного множества непересекающихся интервалов, то все открытые множества также являются борелевскими, а значит, и их дополнения, замкнутые множества, тоже борелевские.

Аналогично определяется класс борелевских множеств $\mathcal{B}^2$: как наименьшая σ-алгебра, содержащая прямоугольники в $\mathbb{R}^2$.

Борелевские алгебры в $\mathbb{R}^s$ являются естественной областью определения вероятностных мер в евклидовых пространствах.

9.1.3. Вероятностные меры, или распределения вероятностей

Этот раздел посвящен третьей составляющей, входящей в определение вероятностного пространства, — вероятностной мере, или распределению вероятностей $\mathbf{P}$. Первое название связано с общим понятием меры в функциональном анализе. И тот и другой термин общеприняты в вероятностной литературе.

Определение 9.6. Функция множеств $\mathbf{P}(A)$, определенная на $\mathcal{A}$, называется *вероятностной мерой (вероятностью)*, если

- $\mathbf{P}(A) \geqslant 0$ $(\forall A \in \mathcal{A})$;
- $\mathbf{P}(\Omega) = 1$;
- выполнено свойство *счетной аддитивности*: для любой последовательности попарно несовместимых событий $A_1, A_2, \ldots$ $\left(\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A}\right)$

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_n).$$

Замечание. В этом определении класс $\mathcal{A}$ может быть полуалгеброй (полукольцом с единицей), алгеброй, σ-алгеброй.

Задача 9.5. Покажите, что

1) $\mathbf{P}(\varnothing) = 0$;

2) для вероятности выполняется свойство конечной аддитивности;

3) выполняются свойства вероятности 4–7, приведенные в главе 3.

Приведем без доказательства теорему, очень важную для построения всей теории вероятностей.

Теорема 9.1 (о продолжении вероятностной меры). *Пусть* $\mathbf{P}$ *— вероятностная мера, заданная на полукольце с единицей* $\mathcal{S}$. *Тогда существует, и притом только одна, вероятностная мера* $\mathbf{P}_1$, *определенная на* σ-*алгебре* $\mathcal{B}(\mathcal{S})$, *порожденной* $\mathcal{S}$, *и такая, что* $\forall A \in \mathcal{S}$ *выполняется равенство* $\mathbf{P}(A) = \mathbf{P}_1(A)$.

Другими словами, всякая счетно-аддитивная мера может быть единственным образом продолжена с полукольца с единицей на наименьшую σ-алгебру, содержащую это полукольцо[1].

Из аксиом в определении вероятности можно вывести другие важные свойства вероятностной меры. Для этого нам понадобятся некоторые вспомогательные утверждения.

Лемма 9.3. *Пусть* $\mathcal{S}$ *— полукольцо и* $A, A_1, \ldots, A_m \in \mathcal{S}$, *причем все* $A_i \subset A$, $A_i \cap A_j = \varnothing$, *если* $i \neq j$. *Тогда найдутся* $A_{m+1}, \ldots, A_n \in \mathcal{S}$ *такие, что события* $A_1, \ldots, A_m, A_{m+1}, \ldots, A_n$ *образуют конечное разбиение* A.

Это утверждение почти очевидным образом вытекает из определения полукольца.

Лемма 9.4. *Пусть* $\mathbf{P}$ *— вероятностная мера, определенная на полукольце* $\mathcal{S}$ *с единицей,* $A, A_1, \ldots, A_m$ *— события, удовлетворяющие условию предыдущей леммы, тогда*

$$\mathbf{P}(A) \geqslant \sum_{j=1}^{m} \mathbf{P}(A_j).$$

Доказательство. Дополним $A_1, \ldots, A_m$ до разбиения $\mathcal{A}$, а именно: $A_1, \ldots, A_m, A_{m+1}, \ldots, A_n$. Тогда из свойства счетной

[1] Подробнее об этом можно прочитать в учебниках [26; 30].

аддитивности следует, что

$$\mathbf{P}(A)=\sum_{j=1}^{n}\mathbf{P}(A_j)\geqslant\sum_{j=1}^{m}\mathbf{P}(A_j).$$

□

Лемма 9.5. *Пусть* $A\subset\bigcup\limits_{j=1}^{n}A_j$, $A, A_1,\ldots,A_n$ *— элементы полукольца с единицей* $\mathcal{S}$. *Тогда*

$$\mathbf{P}(A)\leqslant\sum_{j=1}^{n}\mathbf{P}(A_j).$$

Доказательство. Представим событие A в следующем виде:

$$A=\bigcup_{j=1}^{n}(A\cap A_j).$$

Из определения полукольца следует, что события $B_j=A\cap A_j\in\mathcal{S}$.

Рассмотрим для начала случай $n=2$:

$$A=B_1\cup B_2=B_1\cup(B_2\cap\overline{B_1}).$$

Так как $\mathcal{S}$ — полукольцо с единицей, то в силу леммы 9.3 для $\overline{B_1}=\Omega\setminus B_1$ существует конечное разбиение

$$\overline{B_1}=C_1\cup\ldots\cup C_l,$$

такое, что все $C_i\in\mathcal{S}$. Но тогда

$$B_2\cap\overline{B_1}=\bigcup_{i=1}^{l}(B_2\cap C_i).$$

Следовательно, для A существует конечное разбиение

$$A=B_1\cup(B_2\cap C_1)\cup\ldots(B_2\cap C_l),$$

все элементы которого принадлежат полукольцу $\mathcal{S}$. Но тогда в силу леммы 9.4

$$\mathbf{P}(A)=\mathbf{P}(B_1)+\sum_{i=1}^{l}\mathbf{P}(B_2\cap C_i)\leqslant\mathbf{P}(B_1)+\mathbf{P}(B_2).$$

По индукции можно доказать справедливость этого неравенства для произвольного n. □

Поскольку вероятностная мера, определенная на полукольце (алгебре), может быть единственным образом продолжена на порожденную этим полукольцом (алгеброй) σ-алгебру, то в дальнейшем будем считать, что **областью определения вероятности является σ-алгебра.**

Приведем несколько важных свойств вероятностной меры, определенной на σ-алгебре $\mathcal{A}$.

Свойство 1. Пусть $A_1 \subset A_2 \subset \ldots \subset A_n \subset A_{n+1} \subset \ldots \in \mathcal{A}$ — неубывающая последовательность случайных событий, $A = \bigcup\limits_{n=1}^{\infty} A_n$, тогда $\mathbf{P}(A) = \lim\limits_{n\to\infty} \mathbf{P}(A_n)$.

Свойство 2. Пусть $B_1 \supset B_2 \supset \ldots \supset B_n \supset B_{n+1} \supset \ldots \in \mathcal{A}$ — невозрастающая последовательность случайных событий, $B = \bigcap\limits_{n=1}^{\infty} B_n$. Тогда $\mathbf{P}(B) = \lim\limits_{n\to\infty} \mathbf{P}(B_n)$.

Свойство 3. Если $A = \bigcup\limits_{n=1}^{\infty} A_n$, $A_n \in \mathcal{A}$, то $\mathbf{P}(A) \leqslant \sum\limits_{n=1}^{\infty} \mathbf{P}(A_n)$.

Замечание. Эти свойства вытекают из свойства счетной аддитивности вероятностной меры, определенной на σ-алгебре. Первые два называются свойствами *непрерывности* по монотонным последовательностям событий, так как в этих случаях можно считать

$$A = \lim_{n\to\infty} A_n, \quad B = \lim_{n\to\infty} B_n$$

и сформулировать данные свойства в другом виде:

$$\mathbf{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbf{P}(A_n), \quad \mathbf{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to} \mathbf{P}(B_n).$$

Иногда в определении вероятности свойство счетной аддитивности заменяют двумя свойствами: конечной аддитивностью и непрерывностью вероятности по убывающей к $\varnothing$ последовательности событий[1].

Свойство 3 называют свойством *полуаддитивности* вероятности.

Докажем свойство 1.

[1] См., например, [28, с. 22–23].

Доказательство. Представим $A = \bigcup\limits_{n=1}^{\infty} A_n$ в виде объединения попарно непересекающихся событий B_n:

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus A_2, \quad \ldots, \quad B_n = A_n \setminus A_{n-1}, \quad \ldots .$$

Покажем, что $A = \bigcup\limits_{n=1}^{\infty} B_n$ и $A_N = \bigcup\limits_{n=1}^{N} B_n$. Действительно,

$$A_2 = B_1 \cup B_2, \quad A_3 = A_2 \cup B_3 = B_1 \cup B_2 \cup B_3, \quad \ldots .$$

Следовательно,

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbf{P}(B_n) = \lim_{N\to\infty} \mathbf{P}\left(\bigcup_{n=1}^{N} B_n\right) = \lim_{N\to\infty} \mathbf{P}(A_N).$$

$\square$

Свойство 2 выводится из свойства 1, если ввести неубывающую последовательность событий

$$A_1 = \overline{B_1}, \quad A_2 = \overline{B_2}, \quad \ldots, \quad A_n = \overline{B_n}, \quad \ldots .$$

Докажем свойство 3[1].

Доказательство. Снова представим событие A в виде объединения попарно несовместимых событий. Для этого введем

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \quad \ldots, \quad B_n = A_n \Big\backslash \bigcup_{i=1}^{n-1} A_i.$$

Покажем, что

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n. \tag{9.2}$$

Если $\omega_0 \in \bigcup\limits_{n=1}^{\infty} A_n$, то найдется номер n такой, что $\omega_0 \in A_n$. Среди таких номеров выберем наименьший — n_0. Тогда при всех $n < n_0$ $\omega \notin A_n$. Но тогда $\omega_{n_0} \in B_{n_0}$ (по определению B_{n_0}). Следовательно, $\omega_{n_0} \in \bigcup\limits_{n=1}^{\infty} B_n$.

[1] Сравните с доказательством свойства вероятности 7, приведенным в главе 3.

Покажем обратное включение. Пусть $\omega_1 \in \bigcup\limits_{n=1}^{\infty} B_n$. Тогда найдется номер n_1 такой, что $\omega_1 \in B_{n_1}$. Поскольку $B_{n_1} \subset A_{n_1}$ по построению, то $\omega_{n_1} \in A_{n_1} \subset \bigcup\limits_{n=1}^{\infty} A_n$, что и доказывает (9.2), следовательно,

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbf{P}(B_n) \leqslant \sum_{n=1}^{\infty} \mathbf{P}(A_n).$$

□

9.2. Вероятностные меры в евклидовых пространствах

Остановимся подробнее на этом случае, поскольку именно он чаще всего используется в приложениях теории вероятностей и в математической статистике.

Пусть $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}^d$ — наименьшая σ-алгебра, содержащая полуинтервалы (параллелепипеды).

9.2.1. Вероятностные распределения на прямой

Рассмотрим сначала задание вероятностной меры $\mathbf{P}$ на числовой прямой ($d = 1$).

Определение 9.7. *Функцией распределения*, соответствующей мере $\mathbf{P}$, называется функция $F_P(x)$ действительной переменной x, которая в каждой точке определяется равенством

$$F_P(x) = \mathbf{P}\{(-\infty; x)\}.$$

Всякая функция распределения обладает следующими свойствами:

1) $F_P(x)$ — неубывающая функция x;

2) в любой точке x функция $F_P(x)$ непрерывна слева, то есть $F_P(x_n) \to F_P(x)$, если $x_n \uparrow x$;

3) $0 \leqslant F_P(x) \leqslant 1$, причем $\lim\limits_{x\to-\infty} F_P(x) = 0$, $\lim\limits_{x\to+\infty} F_P(x) = 1$.

При доказательстве этих свойств функций распределения используются свойства вероятностных мер и равенство

$$F_P(x+h) - F_P(x) = \mathbf{P}([x; x+h)), \tag{9.3}$$

справедливое при всех $h > 0$. Приведем доказательство того, что $\lim\limits_{x\to+\infty} F_P(x) = 1$ (из свойства 3).

Доказательство. Пусть $x_n \uparrow +\infty$. Обозначим $A_n = (-\infty; x_n)$. Легко видеть, что

$$A_1 \subset A_2 \subset \ldots; \quad \bigcup_n A_n = (-\infty; +\infty).$$

Применив одно из свойств непрерывности вероятности, имеем

$$F_P(x_n) = \mathbf{P}(A_n) \to \mathbf{P}(-\infty; +\infty) = 1.$$

□

Задача 9.6. Проведите самостоятельно доказательство остальных свойств вероятности.

Замечание. Понятие функции распределения используется также в курсе функционального анализа для неубывающих функций с ограниченной вариацией, непрерывных слева.

Определение 9.8. *Функцией распределения* называют всякую функцию, обладающую свойствами 1–3 из определения 9.7.

Оказывается, что каждая функция распределения порождает вероятностную меру, и притом только одну.

Между классом функций распределения и множеством вероятностных мер на прямой существует взаимно однозначное соответствие.

Если $F(x)$ — функция распределения, то соответствующую ей меру $\mathbf{P}_F$ определим сначала на полукольце полуинтервалов $\mathcal{S}$ согласно равенствам

$$\begin{aligned} &\mathbf{P}_F([a;b)) = F(b) - F(a), \quad \mathbf{P}_F((-\infty;b)) = F(b), \\ &\mathbf{P}_F([a;+\infty)) = 1 - F(a), \quad \mathbf{P}_F((-\infty;+\infty)) = 1. \end{aligned} \tag{9.4}$$

Доказав счетную аддитивность меры $\mathbf{P}_F$ на полукольце с единицей $\mathcal{S}$, можно будет утверждать, что она допускает единственное продолжение на борелевскую σ-алгебру множеств на прямой.

Замечание. В некоторых учебниках (например, [26; 30]) функция распределения определяется как $F_P(x) = \mathbf{P}((-\infty; x])$. Различие в свойствах этих функций лишь в том, что функция, определенная вторым способом, оказывается непрерывной справа. Все остальные свойства остаются без изменений.

Функция $\mathbf{P}_F$, определяемая соотношениями (9.4), обладает также свойствами:

1) $\forall \triangle \in \mathcal{S}$ выполняется $\mathbf{P}_F(\triangle) \geqslant 0$;

2) $\mathbf{P}_F(\mathbb{R}^1) = 1$;

3) $\mathbf{P}_F$ счетно-аддитивна на $\mathcal{S}$.

Доказательство. Первые два свойства очевидным образом следуют из (9.4). Легко видеть также, что выполняется свойство конечной аддитивности. Докажем счетную аддитивность. Доказательство проведем для конечного полуинтервала $\triangle = [a; b) \in \mathcal{S}$. В остальных случаях рассуждения аналогичны, можете провести их самостоятельно.

Пусть полуинтервал $\triangle$ представлен в виде объединения попарно непересекающихся полуинтервалов $\triangle_n = [a_n; b_n) \in \mathcal{S}$:

$$\triangle = \bigcup_{n=1}^{\infty} \triangle_n.$$

Тогда для любого натурального N выполняется включение $\triangle \supset \bigcup\limits_{n=1}^{N} \triangle_n$. Согласно лемме 9.4 имеем

$$\mathbf{P}_F(\triangle) \geqslant \sum_{n=1}^{N} \mathbf{P}_F(\triangle_n), \tag{9.5}$$

$$\triangle_i \cap \triangle_j = \varnothing, \quad \triangle_1, \ldots, \triangle_N \in \mathcal{S}.$$

Переходя к пределу при $N \to \infty$, получим

$$\mathbf{P}_F(\triangle) \geqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n). \tag{9.6}$$

Покажем теперь, что также справедливо противоположное неравенство.

Пусть $\lambda, \lambda_1, \lambda_2, \ldots$ — пока произвольные достаточно малые положительные числа, для которых существуют определяемые ниже промежутки. Рассмотрим $\triangle' = [a; b - \lambda] \subset [a; b)$, $\triangle'_n = (a_n - \lambda_n; b_n) \supset [a_n; b_n)$. Система открытых интервалов $\triangle'_n$ образует покрытие отрезка $\triangle'$. По лемме Гейне–Бореля[1] из этой системы интервалов можно выделить конечную подсистему $\triangle'_{n_1}, \ldots, \triangle'_{n_k}$,

[1] Иногда эту лемму называют просто леммой Бореля. Доказательство этой леммы можно посмотреть, например, в [33, т. 1, с. 181].

также покрывающую $\triangle'$, то есть

$$\triangle' \subset \bigcup_{j=1}^{k} \triangle'_{n_j}.$$

В силу леммы 9.5 имеем

$$\mathbf{P}_F(\triangle') \leqslant \sum_{j=1}^{k} \mathbf{P}_F(\triangle'_{n_j}) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle'_n).$$

Из определения меры $\mathbf{P}_F$ на полуинтервалах получаем

$$F(b-\lambda) - F(a) \leqslant \sum_{n=1}^{\infty} (F(b_n) - F(a_n - \lambda_n)) =$$

$$= \sum_{n=1}^{\infty} (F(b_n) - F(a_n)) + \sum_{n=1}^{\infty} (F(a_n) - F(a_n - \lambda_n)).$$

Поскольку функция распределения $F(x)$ непрерывна слева во всех точках x, то $\forall \varepsilon > 0$ и $\forall n$ можно выбрать λ_n таким образом, чтобы

$$0 \leqslant F(a_n) - F(a_n - \lambda_n) < \frac{\varepsilon}{2^n}.$$

Но тогда

$$F(b-\lambda) - F(a) < \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n) + \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n) + \varepsilon.$$

Устремив $\varepsilon \to +0$, получим

$$F(b-\lambda) - F(a) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n).$$

Поскольку $\lambda > 0$ — произвольно, то можно также перейти к пределу при $\lambda \to +0$. Из непрерывности функции распределения слева в каждой точке следует, что

$$\mathbf{P}_F([a;b)) = F(b) - F(a) \leqslant \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n). \qquad (9.7)$$

Учитывая одновременное выполнение неравенств (9.6), (9.7), получим

$$\mathbf{P}_F(\triangle) = \sum_{n=1}^{\infty} \mathbf{P}_F(\triangle_n),$$

то есть мера $\mathbf{P}_F$, определенная на полукольце интервалов, счетно-аддитивна и, следовательно, допускает единственное продолжение на σ-алгебру $\mathcal{B}^1$, порожденную этим полукольцом. $\square$

Таким образом, доказана следующая теорема.

Теорема 9.2. *Пусть $F(x)$ — функция распределения. Тогда существует единственная вероятностная мера $\mathbf{P}_F$, определенная на борелевской σ-алгебре $\mathcal{B}^1$, для которой $F(x)$ является ее функцией распределения.*

9.2.2. Вероятностные распределения на плоскости и в пространстве

Пусть теперь $\Omega = \mathbb{R}^2$, σ-алгебра случайных событий — борелевские множества из $\mathcal{B}^2$, $\mathbf{P}$ — вероятностная мера на $\mathcal{B}^2$.

Определение 9.9. *Функцией распределения* вероятностной меры $\mathbf{P}$ называется функция $F_P(x,y)$ двух действительных переменных x, y, определяемая равенством

$$F_P(x,y) = \mathbf{P}\{(u,v)\colon\ u < x, \quad v < y\}.$$

Из свойств вероятностной меры вытекают следующие свойства функции распределения $F_P(x,y)$:

1) $F_P(x,y)$ не убывает по каждой из переменных;
2) $F_P(x,-\infty) = F_P(-\infty,y) = 0\ \forall\, x,y$;
3) $F_P(+\infty,+\infty) = 1$;
4) $\forall\, h,k > 0$, $\forall\, x,y$

$$F_P(x+h,y+k) - F_P(x+h,y) - F_P(x,y+k) + F_P(x,y) \geqslant 0; \quad (9.8)$$

5) свойство непрерывности слева по совокупности переменных: $F(x-h,y-k) \to F(x,y)$ при $h,k \to +0$.

Выражение в левой части (9.8) есть не что иное, как вероятность попадания в прямоугольник $\triangle = [x;x+h) \times [y;y+k)$ (рис. 6, где $a = x$; $b = x+h$; $c = y$; $d = y+k$).

Замечание. Если выражение в (9.8) поделить на hk, то получим разностный аналог второй смешанной производной функции

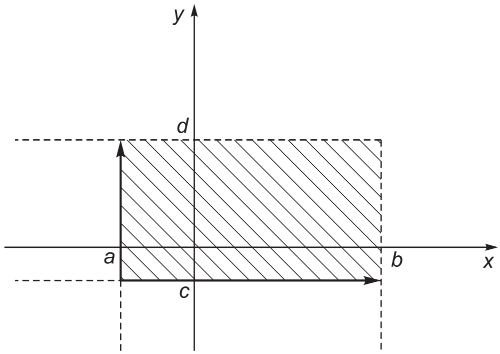

Рис. 6. Распределение вероятностей на плоскости

$F_P(x,y)$. Если перейти к пределу при $h, k \to +0$, то получим, что должно выполняться неравенство

$$\frac{\partial^2 F_P(x,y)}{\partial x \partial y} \geqslant 0$$

для тех точек, где эта производная существует.

Определение 9.10. Функция от двух переменных, для которой выполнены свойства 1–5 из определения 9.9, называется *двумерной функцией распределения*.

Теорема 9.3. *Для любой функции распределения $F(x,y)$ существует единственная вероятностная мера $\mathbf{P}_F$, определенная на $\mathcal{B}^2$ и такая, что $F(x,y)$ является ее функцией распределения.*

Доказательство аналогично тому, что проводилось в одномерном случае.

Похожим образом определяются вероятностные меры в евклидовых пространствах большей размерности: сначала задаются на прямоугольных параллелепипедах, а затем продолжаются на борелевские σ-алгебры. Отметим одно важное свойство вероятностных мер на борелевских множествах.

Пусть $(\mathbb{R}^d, \mathcal{B}^d, \mathbf{P})$ — вероятностное пространство, $\mathcal{G}$ — класс открытых подмножеств $\mathbb{R}^d$, $\mathcal{F}$ — класс замкнутых множеств. Если $B \in \mathcal{B}^d$,

$$\mathbf{P}(B) = \inf_{B \subset G \in \mathcal{G}} \mathbf{P}(G) = \sup_{B \supset F \in \mathcal{F}} \mathbf{P}(F). \tag{9.9}$$

Замечание. Вероятностную меру в произвольных метрических пространствах стремятся выбирать таким образом, чтобы выполнялось (9.9).

9.2.3. Два основных типа распределений в евклидовых пространствах

В большинстве приложений теории вероятностей мы имеем дело с распределениями одного из двух простых типов — дискретного и непрерывного.

1. Дискретный тип распределений.

Определение 9.11. Распределение вероятностей $\mathbf{P}$ на $(\mathbb{R}^1, \mathcal{B}^1)$ называется *дискретным*, если существует последовательность $x_1, x_2, \ldots$ различных между собой точек, таких, что

$$\mathbf{P}(x_k) = p_k \geqslant 0, \quad \sum_k p_k = 1.$$

В этом случае вся вероятность сосредоточена на счетном или конечном множестве точек. Вероятности других событий определяются как

$$\mathbf{P}(B) = \sum_{x_k \in B} p_k, \quad \mathbf{P}(\varnothing) = 0.$$

Примерами распределений дискретного типа могут служить биномиальное и пуассоновское распределения.

2. Непрерывный тип распределений[1]**.** Рассмотрим сначала распределения на прямой. Для распределения $\mathbf{P}$ *непрерывного типа* функция распределения $F_P(x)$ задается интегралом от некоторой неотрицательной функции $p(x)$, называемой *плотностью* распределения:

$$F_P(x) = \int\limits_{-\infty}^{x} p(u)\,du, \quad \int\limits_{-\infty}^{\infty} p(u)\,du = 1.$$

[1] Это название мы позаимствовали у Г. Крамера [17, с. 192]. Оно широко распространено у прикладников. В современной вероятностной литературе используется более точное название: *абсолютно непрерывные распределения по отношению к мере Лебега*, а часто — просто *абсолютно непрерывные распределения*.

Вероятность любого борелевского множества $B \in \mathcal{B}^1$ также можно вычислить с помощью интеграла от плотности:

$$\mathbf{P}(B) = \int_B p(x)\,dx.$$

Замечание. В известной монографии Г. Крамера [17] отмечается, что в приложениях ограничиваются случаями, когда плотность непрерывна всюду, кроме конечного числа точек, или когда множество точек разрыва бесконечно, но не имеет конечной предельной точки. Интеграл понимается в этих случаях как интеграл Римана.

Подобным же образом можно определить непрерывный тип распределения в двумерном случае:

$$F_P(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(u,v)\,du\,dv,$$

где функция $p(u,v) \geqslant 0$, и $\iint_{\mathbb{R}^2} p(u,v)\,du\,dv = 1$. Функцию $p(x,y)$ также называют плотностью распределения вероятностей. Как и в одномерном случае, $\forall\, B \in \mathcal{B}^2$

$$\mathbf{P}(B) = \iint_B p(x,y)\,dx\,dy.$$

9.3. Случайные величины

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — фиксированное вероятностное пространство.

Определение 9.12. *Случайной величиной*, заданной на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$, называется числовая функция $X(\omega)$ такая, что для любых a и b $(a \leqslant b)$

$$\{\omega\colon a \leqslant X(\omega) < b\} \in \mathcal{A}. \tag{9.10}$$

Задача 9.7. Покажите, что условие (9.10) в определении случайной величины эквивалентно каждому из следующих условий:

1) $\forall\, x \in \mathbb{R}^1$ $\{\omega\colon X(\omega) < x\} \in \mathcal{A}$;

2) $\forall B \in \mathcal{B}^1$ полный прообраз B при отображении X является элементом $\mathcal{A}$[1].

В теории функций понятие случайной величины эквивалентно понятию измеримой функции.

Пара $(\Omega, \mathcal{A})$ называется *измеримым пространством.* Рассмотрим также измеримое пространство $(\mathbb{R}^1, \mathcal{B}^1)$.

Определение 9.13. Функция $X\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^1, \mathcal{B}^1)$ называется $\mathcal{A}$-*измеримой*, если $\forall B \in \mathcal{B}^1$ выполняется $X^{-1}(B) \in \mathcal{A}$.

Теорема 9.4. *Пусть последовательность случайных величин* $X_1, X_2, \ldots$ *определена на* $(\Omega, \mathcal{A}, \mathbf{P})$ *и* $\forall \omega\ \exists \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$. *Тогда* $X(\omega)$ *также является случайной величиной.*

Другими словами, класс случайных величин, определенных на фиксированном вероятностном пространстве, замкнут относительно операции предельного перехода.

Доказательство. Покажем сначала, что $\forall c \in \mathbb{R}$ справедливо равенство

$$\{\omega\colon X(\omega) < c\} = \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n\geqslant m} \left\{\omega\colon X_n(\omega) < c - \tfrac{1}{k}\right\}. \tag{9.11}$$

1. Пусть $\omega_0 \in \{\omega\colon X(\omega) < c\}$. Тогда $X(\omega_0) < c$ и, следовательно, существует такое натуральное число k, что $X(\omega_0) < c - \dfrac{1}{k}$. Из определения предела вытекает, что найдется номер m, начиная с которого $\forall n \geqslant m\ X_n(\omega) < c - \dfrac{1}{k}$. Это означает, что

$$\omega_0 \in \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n\geqslant m} \left\{\omega\colon X_n(\omega) < c - \tfrac{1}{k}\right\},$$

то есть

$$\{\omega\colon X(\omega) < c\} \subset \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n\geqslant m} \left\{\omega\colon X_n(\omega) < c - \tfrac{1}{k}\right\}. \tag{9.12}$$

[1] Полным прообразом B при отображении X называют множество $X^{-1}(B) = \{\omega\colon X(\omega) \in B\}$.

2. Аналогичным образом доказывается обратное вложение

$$\{\omega\colon X(\omega) < c\} \supset \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n \geqslant m} \left\{\omega\colon X_n(\omega) < c - \frac{1}{k}\right\}. \qquad (9.13)$$

Из (9.12), (9.13) вытекает справедливость равенства (9.11). Поскольку из определения случайной величины для $X_n(\omega)$ следует выполнение

$$\left\{\omega\colon X_n(\omega) < c - \frac{1}{k}\right\} \in \mathcal{A}$$

для всех n, k, то событие в правой части (9.11) принадлежит σ-алгебре $\mathcal{A}$, а значит, и $\{\omega\colon X(\omega) < c\} \in \mathcal{A}$, то есть предел последовательности случайных величин является случайной величиной. □

Определение 9.14. Случайная величина $X(\omega)$ называется *элементарной*, если она принимает конечное или счетное число попарно различных значений $x_1, x_2, \ldots$.

Рассмотрим события

$$A_j = \{\omega\colon X(\omega) = x_j\} \in \mathcal{A}.$$

Элементарную случайную величину $X(\omega)$ можно представить в виде

$$X(\omega) = \sum_j x_j I_{A_j}(\omega). \qquad (9.14)$$

Поскольку события $A_1, A_2, \ldots$ образуют разбиение Ω, то в сумме (9.14) для любого элементарного исхода только одно слагаемое отлично от 0, и, следовательно, ряд сходится в каждой точке.

Лемма 9.6. *Пусть* $X_1, \ldots, X_n$ *— элементарные случайные величины на* $(\Omega, \mathcal{A}, \mathbf{P})$, *числовая функция* $f(x_1, \ldots, x_n)$ *определена на множестве значений случайных величин* $X_1, \ldots, X_n$. *Тогда* $Y(\omega) = f(X_1(\omega), \ldots, X_n(\omega))$ *также является элементарной случайной величиной на* $(\Omega, \mathcal{A}, \mathbf{P})$.

Доказательство. Доказательство достаточно провести для двух случайных величин. Пусть X и Y — две элементарные случайные величины, представленные в виде

$$X(\omega) = \sum_i x_i I_{A_i}(\omega), \quad Y(\omega) = \sum_j y_j I_{B_j}(\omega),$$

где $\mathcal{E}_1 = \{A_1, A_2, \ldots\}$ и $\mathcal{E}_2 = \{B_1, B_2, \ldots\}$ — два разбиения Ω, все элементы которых принадлежат σ-алгебре $\mathcal{A}$. Рассмотрим $Z(\omega) = f(X(\omega), Y(\omega))$ — произвольную функцию от случайных величин X и Y. Множества $C_{ij} = A_i \cap B_j \in \mathcal{A}$ при всех i, j и образуют новое разбиение Ω, а $Z(\omega)$ можно представить как

$$Z(\omega) = \sum_{i,j} f(x_i, y_j) I_{C_{i,i}}(\omega).$$

Следовательно, $Z(\omega)$ — случайная величина на $(\Omega, \mathcal{A}, \mathbf{P})$. □

Теорема 9.5. *Функция $X(\omega)$ является случайной величиной $\Leftrightarrow$ существует последовательность элементарных случайных величин $X_1(\omega), X_2(\omega), \ldots$, равномерно сходящаяся к $X(\omega)$.*

Доказательство. Достаточность следует из предыдущей теоремы, поэтому докажем только необходимость. Доказательство будет конструктивным. Приведем два примера таких последовательностей.

1. Пусть

$$X_n(\omega) = \sum_{k \in \mathbb{Z}} \frac{k}{n} \cdot I_{\{k/n \leqslant X < (k+1)/n\}}(\omega),$$

при этом полагаем, что $I_{\varnothing}(\omega) = 0$.

2.
$$X_n^*(\omega) = \sum_{l \in \mathbb{Z}} \frac{l}{2^n} I_{\{l/2^n \leqslant X < (l+1)/2^n\}}(\omega). \tag{9.15}$$

□

Задача 9.8. Покажите, что каждая из этих последовательностей равномерно сходится к случайной величине $X(\omega)$, причем последовательность $X_n^*(\omega)$ не убывает.

Следствие 9.1. *Для любой случайной величины X существует неубывающая последовательность элементарных случайных величин, равномерно сходящаяся к X.*

Теорема 9.6. *Пусть $f(x_1, \ldots, x_d)$ — непрерывная функция на $\mathbb{R}^d$ и $X_1(\omega), \ldots, X_d(\omega)$— случайные величины на $(\Omega, \mathcal{A}, \mathbf{P})$.*

Тогда $Y(\omega) = f(X_1(\omega), \ldots, X_d(\omega))$ *также является случайной величиной на данном вероятностном пространстве.*

Доказательство. Достаточно доказать, что существует последовательность случайных величин, в каждой точке ω сходящаяся к $Y(\omega)$. Мы укажем последовательность элементарных случайных величин, сходящуюся к $Y(\omega)$. Из предыдущей теоремы следует, что для каждой из $X_i(\omega)$ $(1 \leqslant i \leqslant d)$ существует последовательность элементарных случайных величин, которая сходится к ней равномерно на Ω. Пусть при $n \to \infty$

$$\begin{aligned}
&X_{1,n}(\omega) \ \text{ сходится равномерно к } \ X_1(\omega),\\
&X_{2,n}(\omega) \ \text{ сходится равномерно к } \ X_2(\omega),\\
&\vdots\\
&X_{d,n}(\omega) \ \text{ сходится равномерно к } \ X_d(\omega).
\end{aligned}$$

Покажем, что последовательность элементарных случайных величин

$$Y_n(\omega) = f(X_{1,n}(\omega), \ldots, X_{d,n}(\omega))$$

сходится к $Y(\omega)$. Выберем произвольное малое $\delta > 0$.

Поскольку равномерно сходящихся последовательностей $\{X_{i,n}(\omega)\}$ конечное число, то для любого $\delta > 0$ существует такое $N = N(\delta)$, что одновременно для всех i и ω выполняются неравенства

$$|X_{i,n}(\omega) - X_i(\omega)| < \frac{\delta}{\sqrt{d}},$$

если $n \geqslant N$. Для таких n

$$\sqrt{(X_{1,n}(\omega) - X_1(\omega))^2 + \ldots (X_{d,n}(\omega) - X_d(\omega))^2} < \delta.$$

Рассмотрим фиксированное $\omega_0 \in \Omega$. Функция $f(x_1, \ldots, x_d)$ непрерывна в точке $\boldsymbol{x}_0 = (X_1(\omega_0), \ldots, X_d(\omega_0))$. Следовательно, $\forall \varepsilon > 0$ $\exists \, \delta_0 = \delta(\varepsilon, \omega_0) > 0$ такое, что если

$$\sqrt{(X_{1,n}(\omega_0) - X_1(\omega_0))^2 + \ldots + (X_{d,n}(\omega_0) - X_d(\omega_0))^2} < \delta_0, \tag{9.16}$$

то

$$|f(X_1(\omega_0), \ldots, X_d(\omega_0)) - f(X_{1,n}(\omega_0), \ldots, X_{d,n}(\omega_0))| < \varepsilon.$$

Выберем теперь $N_0 = N(\delta_0)$, тогда при $n \geqslant N_0$ выполняется (9.16), а значит,

$$|Y_n(\omega_0) - Y(\omega_0)| < \varepsilon.$$

Но это и означает, что последовательность $Y_n(\omega_0)$ сходится к $Y(\omega_0)$. Поскольку сходимость выполняется для любого ω_0, то в силу теоремы 9.4 $Y(\omega) = f(X_1(\omega), \ldots, X_d(\omega))$ является случайной величиной. $\square$

Замечание. Заметим, что последовательность $Y_n(\omega)$ не обязательно сходится к $Y(\omega)$ равномерно. Приведите пример, когда это не так.

Как следствие из этой теоремы получаем, что сумма, разность, произведение и другие непрерывные функции от случайных величин являются случайными величинами. Предельным переходом, используя теорему 9.4, можно класс функций $f(x_1, \ldots$ $\ldots, x_d)$ значительно расширить.

Задача 9.9. Покажите, что $Y(\omega) = f(X_1, \ldots, X_d)$ является случайной величиной для любой борелевской функции от d переменных.

9.3.1. σ-Алгебра, порожденная случайной величиной

Пусть $X(\omega)$ — случайная величина на $(\Omega, \mathcal{A}, \mathbf{P})$. Остановимся подробнее на понятии полного прообраза некоторого борелевского множества B при отображении X.

Пример 9.7. Пусть $\Omega = \mathbb{R}^2$, $\omega = (x_1, x_2)$, $X(\omega) = \sqrt{x_1^2 + x_2^2}$. Тогда полными прообразами борелевских множеств

$$B_1 = (-\infty; b); \quad B_2 = [a; b), \quad B_3 = \{b\} \quad (0 < a < b)$$

будут соответственно круг, кольцо, окружность.

Пример 9.8. По-прежнему элементарные исходы $\omega = (x_1, x_2)$ — точки координатной плоскости, $X(\omega) = x_1$, тогда полными прообразами множеств из предыдущего примера служат полуплоскость, полоса, прямая.

Задача 9.10. Докажите справедливость следующих утверждений:

1) $X^{-1}(B_1 \cup B_2) = X^{-1}(B_1) \cup X^{-1}(B_2)$;
2) $X^{-1}(B_1 \cap B_2) = X^{-1}(B_1) \cap X^{-1}(B_2)$;
3) $X^{-1}(\overline{B}) = \overline{X^{-1}(B)}$.

Из этих соотношений следует, что класс подмножеств Ω, являющихся полными прообразами борелевских множеств при отображении $X(\omega)$, образует σ-алгебру. Будем обозначать ее $\mathcal{A}_X$.

σ-Алгебра $\mathcal{A}_X \subseteq \mathcal{A}$ и называется σ-алгеброй, *порожденной* случайной величиной X.

Задача 9.11. Выясните, какие σ-алгебры порождаются случайными величинами, принимающими конечное число значений, в том числе константами.

9.3.2. Распределения случайных величин

С каждой случайной величиной $X(\omega)$ можно связать вероятностное пространство $(\mathbb{R}, \mathcal{B}, \mathbf{P}_F)$, где вероятностная мера $\mathbf{P}_F$ — это мера, функция распределения которой равна

$$F(x) = \mathbf{P}\{\omega\colon X(\omega) < x\}. \tag{9.17}$$

Задача 9.12. Покажите, что функция, определенная в (9.17), действительно является функцией распределения.

Определение 9.15. Функция $F(x)$, определенная равенством (9.17), называется *функцией распределения* случайной величины $X(\omega)$.

В дальнейшем мы будем иметь дело только с двумя основными типами распределений: дискретным и непрерывным (абсолютно непрерывным) (см. раздел 9.2.3). Однако заметим, что в общем случае функция распределения — это смесь функций распределений трех разных типов: дискретного, непрерывного (абсолютно непрерывного) и сингулярного. Прочитать об этом можно, например, в учебнике по теории функций [16].

Примеры и свойства наиболее часто встречающихся распределений содержатся в приложении 1 в конце учебника.

9.4. Математические ожидания случайных величин (общий случай)

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — некоторое вероятностное пространство, $X(\omega)$ — элементарная случайная величина, принимающая значения $x_1, x_2, \ldots$ ($x_i \neq x_j$ $\forall\, i \neq j$). Обозначим $A_k = \{\omega\colon X(\omega) = x_k\}$.

Тогда $A_k \in \mathcal{A}$, образуют естественное разбиение Ω и

$$X(\omega) = \sum_k x_k I_{A_k}(\omega).$$

Определение 9.16. *Математическим ожиданием* элементарной случайной величины $X(\omega)$ называется сумма

$$\mathbf{E}X = \sum_k x_k \mathbf{P}(A_k), \tag{9.18}$$

причем для случайных величин, имеющих счетное число значений, ряд в (9.18) должен сходиться абсолютно.

Требование абсолютной сходимости в определении математического ожидания существенно. Напомним, что сумма абсолютно сходящегося ряда не зависит от порядка суммирования, тогда как условно сходящиеся ряды таким свойством не обладают. Поэтому, если в определении математического ожидания не потребовать абсолютную сходимость, эта числовая характеристика может оказаться зависящей от того, каким образом мы перенумеровали значения случайной величины.

Если ряд в (9.18) абсолютно сходится, то о случайной величине X будем говорить, что она имеет конечное математическое ожидание, или кратко, что у нее *существует математическое ожидание.*

Далее мы докажем лемму, аналогичную лемме 3.1.

Лемма 9.7. *Пусть $B_1, B_2, \ldots$ — конечное или счетное разбиение Ω, $B_l \in \mathcal{A}$ и случайная величина $X(\omega)$ постоянна на каждом из B_l, то есть ее можно представить в виде*

$$X(\omega) = \sum_l x'_l \cdot I_{B_l}(\omega).$$

Тогда

$$\mathbf{E}X = \sum_l x'_l \mathbf{P}(B_l), \tag{9.19}$$

если этот ряд сходится абсолютно.

Замечание. В определении математического ожидания (9.18) все значения случайной величины различны между собой, в (9.19) могут встречаться одинаковые. Это означает, что разбиение $B_1, B_2, \ldots$ более «мелкое», элементы естественного разбиения A_k можно представить в виде объединения событий B_l.

Доказательство. Пусть ряд (9.19) сходится абсолютно. Тогда его сумма не зависит от порядка суммирования и ряд можно переписать в другом виде:

$$\sum_l x'_l \mathbf{P}(B_l) = \sum_k \sum_{l:\, x'_l = x_k} x'_l \mathbf{P}(B_l) =$$

$$= \sum_k x_k \sum_{l:\, x'_l = x_k} \mathbf{P}(B_l) = \sum_k x_k \mathbf{P}(A_k) = \mathbf{E}X.$$

□

Замечание. Поскольку с абсолютно сходящимися рядами можно поступать так же, как с конечными суммами, то при доказательстве безразлично, с какой суммой (конечной или нет) мы имеем дело.

Основные свойства математического ожидания. Сформулируем основные свойства математического ожидания.

1. Математические ожидания $\mathbf{E}X$ и $\mathbf{E}|X|$ существуют или не существуют одновременно и $|\mathbf{E}X| \leqslant \mathbf{E}|X|$.

2. Если $X \geqslant 0$, то $\mathbf{E}X \geqslant 0$, причем в этом случае

$$\mathbf{E}X = 0 \Leftrightarrow \mathbf{P}\{\omega\colon X(\omega) = 0\} = 1.$$

3. Если существует $\mathbf{E}X$, то для любой константы c существует

$$\mathbf{E}(cX) = c \cdot \mathbf{E}X.$$

4. Если существуют $\mathbf{E}X, \mathbf{E}Y$, то существует математическое ожидание суммы этих случайных величин и

$$\mathbf{E}(X+Y) = \mathbf{E}X + \mathbf{E}Y.$$

5. Пусть $Y, X_1, X_2, \ldots$ — случайные величины с конечными математическими ожиданиями и такие, что

$$|X_n(\omega)| \leqslant Y(\omega), \quad X_n(\omega) \to X(\omega).$$

Тогда существует $\mathbf{E}X$ и

$$\lim_{n\to\infty} \mathbf{E}X_n = \mathbf{E}X.$$

Последнее утверждение носит название теоремы Лебега о мажорируемой сходимости.

Замечание. Математическое ожидание можно рассматривать как функцию, определенную на множестве случайных величин, заданных на фиксированном вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$. Такого рода функции называются функционалами. Данный функционал обладает свойствами линейности 3, 4, свойством неотрицательности 2. Аналогичными свойствами обладает интеграл.

Покажем справедливость свойства 4.

Доказательство. Пусть

$$X(\omega) = \sum_k x_k I_{A_k}(\omega), \quad Y(\omega) = \sum_l y_l I_{B_l}(\omega),$$

где $x_1, x_2, \ldots,\ y_1, y_2, \ldots$ — все различные между собой значения случайных величин X и Y. Рассмотрим события $C_{kl} = A_k \cap B_l$, на которых случайная величина $(X+Y)(\omega)$ постоянна и равна $x_k + y_l$, но тогда

$$X + Y = \sum_{k,l} (x_k + y_l) I_{C_{kl}}.$$

Следовательно, ее математическое ожидание равно сумме ряда

$$\mathbf{E}(X+Y) = \sum_{k,l} (x_k + y_l) \mathbf{P}(C_{kl}),$$

если этот ряд сходится абсолютно. С другой стороны, в силу утверждения леммы

$$\mathbf{E}X = \sum_{k,l} x_k \mathbf{P}(C_{kl}) \quad \mathbf{E}Y = \sum_{k,l} y_l (C_{kl}),$$

значит,

$$\mathbf{E}X + \mathbf{E}Y = \sum_{k,l} (x_k + y_l) \mathbf{P}(C_{kl}).$$

Из абсолютной сходимости первых двух рядов вытекает абсолютная сходимость последнего ряда. □

Задача 9.13. Проведите самостоятельно доказательство свойств 1–3.

Доказательство свойства 5 опирается на глубокие факты сходимостей случайных величин, приводить которые в данном курсе мы не будем. Желающие могут посмотреть доказательство этого свойства, например, в [30].

Введем обозначение $L_1 = L_1(\Omega, \mathcal{A}, \mathbf{P})$ для класса случайных величин, определенных на вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$ и обладающих свойством, что $\forall X \in L_1$ существуют элементарная случайная величина $X'(\omega)$ с конечным математическим ожиданием и положительная константа k такие, что при всех ω

$$|X(\omega) - X'(\omega)| \leqslant k.$$

В частности, в класс L_1 входят все элементарные случайные величины, имеющие конечное математическое ожидание, все ограниченные случайные величины (в этом случае можно взять $X'(\omega) \equiv 0$). Так же как в теореме о продолжении меры, можно доказать, что функционал $\mathbf{E}X$, определенный для элементарных случайных величин и обладающий свойствами 1–5, может быть продолжен на L_1.

Теорема 9.7. *На классе случайных величин $L_1(\Omega, \mathcal{A}, \mathbf{P})$ можно, и притом единственным образом, определить математическое ожидание $\mathbf{E}X$ так, что для элементарных случайных величин этот функционал совпадает с определенным ранее и для него выполнены свойства* 1–5.

Замечание. Обычно в учебниках описывают конструкцию, как строится этот функционал с той или иной степенью строгости. Наиболее полное изложение содержится в учебниках [26; 30]. Мы также приведем этапы построения этого продолжения.

1. Определение и доказательство свойств математического ожидания для элементарных случайных величин.

2. Согласно следствию из теоремы 9.5, для любой случайной величины $X(\omega)$ существует неубывающая последовательность элементарных случайных величин, равномерно сходящаяся к $X(\omega)$. В качестве таковой можно, к примеру, взять $X_n^*(\omega)$, определенную в (9.15):

$$X_n^*(\omega) = \sum_{l \in \mathbb{Z}} \frac{l}{2^n} I_{\{l/2^n \leqslant X < (l+1)/2^n\}}(\omega).$$

3. Поскольку $X_n^*(\omega) \leqslant X(\omega)$, то для того, чтобы выполнялось свойство 5 математического ожидания, необходимо опре-

делить $\mathbf{E}X$ следующим образом:

$$\mathbf{E}X = \lim_{n\to\infty} \mathbf{E}X_n^*(\omega).$$

4. Доказательство свойств математического ожидания для общего случая.

Из теорем о математическом ожидании приведем следующие.

Теорема 9.8. *Если существует* $\mathbf{E}Y(\omega)$ *и для всех* ω *выполняется неравенство* $|X(\omega)| \leqslant Y(\omega)$, *то существует и математическое ожидание случайной величины* $X(\omega)$ *и*

$$\mathbf{E}|X(\omega)| \leqslant \mathbf{E}Y(\omega).$$

Доказательство. Доказательство проведем для элементарных случайных величин. Поскольку

$$\sum_{k,l} |x_k| \mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} \leqslant \sum_{k,l} y_l \mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} = \\ = \sum_{l} y_l \mathbf{P}\{Y(\omega) = y_l\} = \mathbf{E}Y(\omega),$$

то ряд в левой части неравенства сходится, причем абсолютно, и можно изменять порядок суммирования. Но тогда

$$\sum_{k,l} |x_k| \mathbf{P}\{X(\omega) = x_k, Y(\omega) = y_l\} = \sum_{k} |x_k| \mathbf{P}\{X(\omega) = x_k\} = \mathbf{E}|X(\omega)|.$$

$\square$

Схема построения математического ожидания есть не что иное, как схема построения интеграла Лебега. Поэтому часто математическое ожидание записывают как интеграл Лебега:

$$\mathbf{E}X = \int_{\Omega} X(\omega) P(d\omega).$$

Если Ω — числовая прямая (плоскость, евклидово пространство $\mathbb{R}^d$), то в некоторых случаях этот интеграл может превратиться в обычный интеграл Римана.

Теорема 9.9. *Пусть распределение случайной величины* $X(\omega)$ *абсолютно непрерывно и* $p(x)$ — *плотность этого рас-*

пределения, тогда

$$\mathbf{E}X = \int\limits_{-\infty}^{+\infty} xp(x)\,dx, \tag{9.20}$$

при условии, что интеграл сходится абсолютно.

Доказательство. Введем последовательность элементарных случайных величин

$$X_n(\omega) = \sum_{k\in\mathbb{Z}} \frac{k}{n} \cdot I_{\{k/n\leqslant X<(k+1)/n\}}(\omega),$$

равномерно сходящуюся к $X(\omega)$. Тогда

$$\mathbf{E}X = \lim_{n\to\infty} \mathbf{E}X_n(\omega) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \int\limits_{k/n}^{(k+1)/n} p(x)\,dx =$$

$$= \lim_{n\to\infty}\left(\int\limits_{-\infty}^{\infty} xp(x)\,dx - \sum_{k=-\infty}^{\infty}\left(\int\limits_{k/n}^{(k+1)/n} xp(x)\,dx - \int\limits_{k/n}^{(k+1)/n} \frac{k}{n}p(x)\,dx\right)\right) =$$

$$= \int\limits_{-\infty}^{\infty} xp(x)\,dx - \lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\int\limits_{k/n}^{(k+1)/n}\left(x-\frac{k}{n}\right)p(x)\,dx.$$

Откуда получаем, что

$$\left|\mathbf{E}X - \int\limits_{-\infty}^{\infty} xp(x)\,dx\right| \leqslant \lim_{n\to\infty}\frac{1}{n}\sum_{k=-\infty}^{\infty}\int\limits_{k/n}^{(k+1)/n} p(x)\,dx = \lim_{n\to\infty}\frac{1}{n} = 0,$$

что и доказывает равенство (9.20). □

Аналогичным образом для произвольной измеримой функции $f(x)$ можно доказать, что

$$\mathbf{E}f(X) = \int\limits_{-\infty}^{+\infty} f(x)p(x)\,dx,$$

если интеграл сходится абсолютно.

9.5. Независимые случайные величины

Определение 9.17. Случайные величины $X_1,\ldots,X_n$ называются *независимыми (взаимно независимыми)*, если для любых полуинтервалов $\triangle_1 = [a_1,b_1)$, $\triangle_2 = [a_2,b_2)$, $\ldots$, $\triangle_n = [a_n,b_n)$ выполняется равенство

$$\mathbf{P}\{\omega\colon X_1(\omega)\in\triangle_1,\ X_2(\omega)\in\triangle_2,\ \ldots,\ X_n(\omega)\in\triangle_n\} = \\ = \mathbf{P}\{\omega\colon X_1(\omega)\in\triangle_1\}\mathbf{P}\{\omega\colon X_2(\omega)\in\triangle_2\}\cdots\mathbf{P}\{\omega\colon X_n(\omega)\in\triangle_n\}. \tag{9.21}$$

Определение независимости можно сформулировать и в терминах функций распределений, а для случайных величин с абсолютно непрерывными распределениями — в терминах плотностей.

Определение 9.18. Случайные величины $X_1,\ldots,X_n$ называются *независимыми*, если совместная функция распределения этих величин во всех точках равна произведению частных функций распределения, то есть

$$F(x_1,\ldots,x_n) = F_1(x_1)F_2(x_2)\cdots F_n(x_n),$$

где

$$F(x_1,\ldots,x_n) = \mathbf{P}\{\omega\colon X_1(\omega)<x_1,\ \ldots,\ X_n(\omega)<x_n\},$$
$$F_i(x_i) = \mathbf{P}\{\omega\colon X_i(\omega)<x_i\}.$$

Задача 9.14. Докажите эквивалентность двух определений независимости случайных величин.

Пусть совместное распределение случайных величин X и Y абсолютно непрерывно, то есть существует такая неотрицательная функция $p(x,y)$, что $\forall B\in\mathcal{B}^2$

$$\mathbf{P}\{(X,Y)\in B\} = \iint\limits_B p(x,y)\,dx\,dy.$$

Функция $p(x,y)$ называется совместной плотностью распределения. С функцией распределения $F(x,y)$ она связана соотношением

$$F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p(u,v)\,du\,dv,$$

выполняющимся $\forall x,y$.

Задача 9.15. Докажите, что абсолютно непрерывные случайные величины X и Y независимы $\Longleftrightarrow$ почти всюду по мере Лебега

$$p(x, y) = p_X(x)p_Y(y),$$

где $p_X(x), p_Y(y)$ — частные плотности распределений.

Для независимых случайных величин сохраняются все утверждения глав 4, 5, доказанные в предположении, что X и Y принимают конечное число значений. В частности, борелевские функции от независимых случайных величин также являются независимыми случайными величинами.

9.6. Мультипликативное свойство математического ожидания

Теорема 9.10. *Пусть X и Y — независимые случайные величины, для которых существуют математические ожидания. Тогда существует $\mathbf{E}(X \cdot Y)$ и*

$$\mathbf{E}XY = \mathbf{E}X\,\mathbf{E}Y.$$

Доказательство. При доказательстве этого свойства продемонстрируем прием перехода от элементарных случайных величин к произвольным.

1. Пусть X и Y — элементарные случайные величины:

$$X(\omega) = \sum_k x_k I_{A_k}(\omega), \quad Y(\omega) = \sum_l y_l I_{B_l}(\omega),$$

причем в этих представлениях $x_i \neq x_j$, $y_i \neq y_j$, если $i \neq j$,

$$A_k = \{\omega\colon X(\omega) = x_k\}, \quad B_l = \{\omega\colon Y(\omega) = y_l\}.$$

Обозначим $C_{kl} = A_k \cap B_l$. Тогда случайная величина $X(\omega)Y(\omega)$ постоянна на C_{kl} и

$$X(\omega)Y(\omega) = \sum_{k,l} x_k y_l I_{C_{kl}}.$$

По свойству математического ожидания, сформулированному в лемме 9.6,

$$\mathbf{E}X(\omega)Y(\omega) = \sum_{k,l} x_k y_l \mathbf{P}(C_{kl}), \tag{9.22}$$

если ряд (9.22) сходится абсолютно. Докажем абсолютную сходимость ряда. По условию теоремы существуют

$$\mathbf{E}X(\omega) = \sum_k x_k \mathbf{P}(A_k), \quad \mathbf{E}Y(\omega) = \sum_l y_l \mathbf{P}(B_l). \tag{9.23}$$

Абсолютно сходящиеся ряды можно перемножать и менять произвольным образом порядки суммирования, при этом будем получать также абсолютно сходящиеся ряды. Поэтому

$$\mathbf{E}X\,\mathbf{E}Y = \sum_k x_k \mathbf{P}(A_k) \sum_l y_l \mathbf{P}(B_l) = \sum_{k,l} x_k y_l \mathbf{P}(A_k)\mathbf{P}(B_l) =$$
$$= \sum_{k,l} x_k y_l \mathbf{P}(A_k \cap B_l) = \sum_{k,l} x_k y_l \mathbf{P}(C_{kl}).$$

Сравнивая последнее выражение с (9.22), видим, что математическое ожидание произведения случайных величин существует и $\mathbf{E}(XY) = \mathbf{E}X\,\mathbf{E}Y$, что доказывает утверждение теоремы для элементарных случайных величин.

2. Пусть теперь $X(\omega)$ и $Y(\omega)$ — произвольные случайные величины, имеющие конечные математические ожидания. Рассмотрим последовательности элементарных случайных величин

$$X_n^*(\omega) = \sum_k \frac{k}{n} I_{\{k/n \leqslant X(\omega) < k+1/n\}}(\omega), \qquad \mathbf{E}X(\omega) = \lim_{n\to\infty} \mathbf{E}X_n^*(\omega),$$
$$Y_n^*(\omega) = \sum_l \frac{l}{n} I_{\{l/n \leqslant Y(\omega) < l+1/n\}}(\omega), \qquad \mathbf{E}Y(\omega) = \lim_{n\to\infty} \mathbf{E}Y_n^*(\omega),$$

причем (см. задачу 9.7) построенные последовательности сходятся равномерно. По построению для всех ω

$$0 \leqslant X(\omega) - X_n^*(\omega) \leqslant \frac{1}{n}, \quad 0 \leqslant Y(\omega) - Y_n^*(\omega) \leqslant \frac{1}{n},$$

откуда следует, что

$$|X_n^*(\omega)| \leqslant |X(\omega)| + 1, \quad |Y_n^*(\omega)| \leqslant |Y(\omega)| + 1.$$

Из этих неравенств и конечности математических ожиданий для $X(\omega)$ и $Y(\omega)$ получаем, что для всех n существуют $\mathbf{E}X_n^*(\omega)$, $\mathbf{E}Y_n^*(\omega)$.

По условию теоремы случайные величины X и Y независимы, значит, для любых k и l выполнено равенство

$$\mathbf{P}\left\{\omega\colon\ \frac{k}{n}\leqslant X(\omega)<\frac{k+1}{n},\ \ \frac{l}{n}\leqslant Y(\omega)<\frac{l+1}{n}\right\}=$$
$$=\mathbf{P}\left\{\omega\colon \frac{k}{n}\leqslant X(\omega)<\frac{k+1}{n}\right\}\mathbf{P}\left\{\omega\colon \frac{l}{n}\leqslant Y(\omega)<\frac{l+1}{n}\right\}.$$

Но это означает, что элементарные случайные величины $X_n^*(\omega)$, $Y_n^*(\omega)$ являются независимыми. Тогда в силу доказанного выше имеем

$$\mathbf{E}(X_n^*(\omega)Y_n^*(\omega))=\mathbf{E}\,X_n^*(\omega)\,\mathbf{E}\,Y_n^*(\omega).$$

Осталось только перейти к пределу при $n\to\infty$. Выпишем несколько простых соотношений:

$$XY-X_n^*Y_n^*=(X-X_n^*)Y+X_n^*(Y-Y_n^*),$$
$$|XY-X_n^*Y_n^*|\leqslant|X-X_n^*|\cdot|Y|+|X_n^*|\cdot|Y-Y_n^*|\leqslant\frac{1}{n}(|X|+|Y|+1).$$

Но тогда

$$|\,\mathbf{E}(XY)-\mathbf{E}(X_n^*Y_n^*)|\leqslant\mathbf{E}\,|XY-X_n^*Y_n^*|\leqslant\frac{1}{n}\mathbf{E}(|X|+|Y|+1)\to 0$$

при $n\to\infty$. Следовательно,

$$\mathbf{E}(XY)=\lim_{n\to\infty}\mathbf{E}(X_n^*Y_n^*)=\lim_{n\to\infty}\mathbf{E}\,X_n^*\cdot\lim_{n\to\infty}\mathbf{E}\,Y_n^*=\mathbf{E}\,X\cdot\mathbf{E}\,Y.$$

□

Замечание. В других учебниках (см. [26; 30]) схема построения может быть несколько иной: сначала определяется математическое ожидание простых случайных величин, затем для неотрицательных, затем уже общий случай, но во всех этих схемах с помощью свойства 5 математического ожидания осуществляется предельный переход от простых случайных величин к случайным величинам, множество значений которых более чем счетно.

Глава 10

УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

10.1. Лемма Бореля–Кантелли

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — некоторое вероятностное пространство, $A_1, A_2, \ldots \in \mathcal{A}$ — последовательность случайных событий, $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, … — последовательность случайных величин, определенных на данном вероятностном пространстве.

Определение 10.1. События $A_1, A_2, \ldots$ *независимы*, если для всех целых $m \geqslant 2$, для любых наборов натуральных чисел $j_1 < < j_2 < \ldots < j_m$ события $A_{j_1}, A_{j_2}, \ldots, A_{j_m}$ независимы.

Определение 10.2. Случайные величины $X_1(\omega)$, $X_2(\omega)$, … называются *независимыми*, если для всех целых $m \geqslant 2$, для любых наборов натуральных чисел $j_1 < j_2 < \ldots < j_m$ независимы случайные величины $X_{j_1}(\omega)$, $X_{j_2}(\omega)$, …, $X_{j_m}(\omega)$.

Заметим, что независимость последовательности случайных событий эквивалентна независимости последовательности индикаторов этих событий: $I_{A_1}(\omega)$, $I_{A_2}(\omega)$, $I_{A_3}(\omega)$, … .

Введем следующие события:

$$C = \left\{\omega\colon \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\right\} \quad \text{и} \quad D = \left\{\omega\colon \sum_{n=1}^{\infty} I_{A_n}(\omega) = \infty\right\}.$$

Если $\omega \in C$, то сходится ряд $\sum\limits_{n=1}^{\infty} I_{A_n}(\omega)$ и, значит, элементарный исход ω может одновременно принадлежать лишь конечному числу событий A_n. В этом случае говорят, что происходит конечное число событий A_n.

Если $\omega \in D$, то ряд $\sum\limits_{n=1}^{\infty} I_{A_n}(\omega)$ расходится и, стало быть, найдется бесконечная подпоследовательность событий A_{n_k} такая, что ω принадлежит одновременно всем A_{n_k}. В этом случае говорят, что происходит одновременно бесконечно много событий A_n.

Задача 10.1. Покажите, что справедливы равенства:

$$C = \bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} \overline{A_l}, \quad D = \bigcap_{k=1}^{\infty} \bigcup_{l=k}^{\infty} A_l, \quad C = \overline{D}. \tag{10.1}$$

Лемма 10.1 (Бореля–Кантелли). 1. *Если последовательность событий $A_1, A_2, \ldots$ такова, что ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$, то* $\mathbf{P}(D) = 0$, $\mathbf{P}(C) = 1$.

2. *Если для последовательности независимых событий A_1, A_2, … ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$, то* $\mathbf{P}(D) = 1$, $\mathbf{P}(C) = 0$.

Доказательство. 1. Оценим вероятность события D сверху, для этого воспользуемся представлением (10.1):

$$\mathbf{P}(D) \leqslant \mathbf{P}\left\{\bigcup_{l=k}^{\infty} A_l\right\} \leqslant \sum_{l=k}^{\infty} \mathbf{P}(A_l), \tag{10.2}$$

— это неравенство выполняется для любого k. Из сходимости ряда $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n)$ следует, что правую часть неравенства (10.2) можно сделать сколь угодно малой, выбрав должным образом k. Следовательно,

$$\mathbf{P}(D) = 0, \quad \mathbf{P}(C) = 1 - \mathbf{P}(D) = 1.$$

2. Воспользовавшись представлением (10.1) для события C, получаем

$$\mathbf{P}(C) \leqslant \sum_{k=1}^{\infty} \mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\}.$$

Покажем, что каждое слагаемое в этой сумме равно 0. Действительно, в силу независимости событий A_l и справедливости почти очевидного неравенства $1 - z \leqslant e^{-z}$ имеем

$$\mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\} \leqslant \mathbf{P}\left\{\bigcap_{l=k}^{k+m} \overline{A_l}\right\} = \prod_{l=k}^{k+m} (1 - \mathbf{P}(A_l)) \leqslant \exp\left\{-\sum_{l=k}^{k+m} \mathbf{P}(A_l)\right\}.$$

Поскольку ряд $\sum\limits_{n=1}^{\infty} \mathbf{P}(A_n) = +\infty$, то, устремив $m \to \infty$, получим

$$\mathbf{P}\left\{\bigcap_{l=k}^{\infty} \overline{A_l}\right\} = 0$$

для всех k. Следовательно, $\mathbf{P}(C) = 0$. $\square$

10.2. Сходимость с вероятностью 1

Лемма Бореля–Кантелли является удобным инструментом для установления сходимости последовательностей случайных величин с вероятностью 1[1].

Определение 10.3. Последовательность случайных величин $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ... сходится к 0 *с вероятностью* 1, если

$$\mathbf{P}\left\{\omega\colon \lim_{n\to\infty} X_n(\omega) = 0\right\} = 1. \tag{10.3}$$

Равенство (10.3) означает, что сходимость к 0 имеет место для всех элементарных исходов за исключением, быть может, исходов, общая вероятность которых равна 0. Поэтому условие (10.3) можно переписать иначе:

$$\mathbf{P}\{\omega\colon X_n(\omega) \not\to 0\} = 0. \tag{10.4}$$

Сходимость с вероятностью 1 называют также сходимостью *почти всюду* или сходимостью *почти наверное* и обозначают

$$X_n(\omega) \xrightarrow{\text{п.в.}} 0 \quad \text{или} \quad X_n(\omega) \xrightarrow{\text{п.н.}} 0.$$

Лемма 10.2. *Для последовательности $X_1, X_2, \ldots$ случайных величин справедливо утверждение*

$$X_n(\omega) \xrightarrow{\text{п.н.}} 0 \Longleftrightarrow \forall\, \varepsilon > 0 \quad \sum_{n=1}^{\infty} \mathbf{P}\{\omega\colon |X_n(\omega)| \geqslant \varepsilon\} < \infty. \tag{10.5}$$

Доказательство леммы проведите самостоятельно, используя лемму Бореля–Кантелли.

---

[1] О различных видах сходимости для последовательностей случайных величин и соотношениях между ними более подробно можно прочитать в учебнике А. Н. Ширяева «Вероятность» (гл. 2, § 10).

Определение 10.4. Последовательность случайных величин $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ... сходится к случайной величине $X(\omega)$ *с вероятностью* 1 ($X_n(\omega) \xrightarrow{\text{п.н.}} X(\omega)$), если $X_n(\omega) - X(\omega) \xrightarrow{\text{п.н.}} 0$.

Приведем несколько утверждений о сходимости почти наверное.

Теорема 10.1.

$$X_n(\omega) \xrightarrow{\text{п.н.}} 0 \Longleftrightarrow \forall\, \varepsilon > 0 \quad \lim_{n\to\infty} \mathbf{P}\left\{\omega\colon \sup_{k\geqslant n} |X_k(\omega)| \geqslant \varepsilon\right\} = 0.$$

Доказательство. Введем обозначение для множества сходимости последовательности $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ..., а именно

$$A = \{\omega\colon X_n(\omega) \to 0\}.$$

Используя определение предела, мы можем записать

$$A = \bigcap_{\varepsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{k\geqslant n} \{\omega\colon |X_k(\omega)| < \varepsilon\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k\geqslant n} \left\{\omega\colon |X_k(\omega)| \leqslant \frac{1}{m}\right\}.$$

Тогда множество, где нет сходимости к 0, запишется как

$$\overline{A} = \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k\geqslant n} \left\{\omega\colon |X_k(\omega)| > \frac{1}{m}\right\}.$$

Рассмотрим события

$$B_{n,m} = \bigcup_{k\geqslant n} \left\{\omega\colon |X_k(\omega)| > \frac{1}{m}\right\} = \left\{\omega\colon \sup_{k\geqslant n} |X_k(\omega)| > \frac{1}{m}\right\},$$

образующие при любом m монотонно убывающую по n последовательность случайных событий. Проведем цепочку несложных рассуждений:

$$\mathbf{P}(A) = 1 \Longleftrightarrow \mathbf{P}(\overline{A}) = 0 \Longleftrightarrow \forall\, m \quad \mathbf{P}\left(\bigcap_{n=1}^{\infty} B_{n,m}\right) = 0 \Longleftrightarrow$$

$$\Longleftrightarrow \forall\, m \quad \lim_{n\to\infty} \mathbf{P}\left\{\omega\colon \sup_{k\geqslant n} |X_k(\omega)| > \frac{1}{m}\right\} = 0,$$

откуда следует утверждение теоремы. □

Определение 10.5. Последовательность случайных величин $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ... *фундаментальна с вероятностью* 1, если вероятность тех ω, для которых числовая последовательность $\{X_n(\omega)\}_{n=1}^{\infty}$ является фундаментальной, равна 1.

Теорема 10.2. *Последовательность случайных величин $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ... сходится с вероятностью 1 (фундаментальна с вероятностью 1)* $\Longleftrightarrow \forall\, \varepsilon > 0$

$$\mathbf{P}\{\sup_{k,\,l \geqslant n} |X_k(\omega) - X_l(\omega)| \geqslant \varepsilon\} \to 0$$

при $n \to \infty$.

Доказательство аналогично доказательству предыдущей теоремы.

10.3. Усиленный закон больших чисел

Теорема 10.3 (А. Н. Колмогоров). *Пусть* $X_1(\omega)$, $X_2(\omega)$, $X_3(\omega)$, ... — *последовательность независимых случайных величин, для которых существуют математические ожидания и дисперсии, причем* $\mathbf{E}X_k = m_k$, $\mathbf{D}X_k = \sigma_k^2$ *и сходится ряд* $\sum\limits_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$. *Тогда*

$$\mathbf{P}\left\{\omega\colon \frac{X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)}{n} - \frac{m_1 + m_2 + \ldots + m_n}{n} \to 0\right\} = 1. \quad (10.6)$$

Мы будем говорить, что для последовательности случайных величин выполняется *усиленный закон больших чисел*, если выполняется (10.6), другими словами, если

$$\frac{X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)}{n} - \frac{m_1 + m_2 + \ldots + m_n}{n} \xrightarrow{\text{п.н.}} 0. \quad (10.7)$$

Доказательство. Рассмотрим случайные величины $X_k'(\omega) = X_k(\omega) - m_k$. Для них $\mathbf{E}X_k' = 0$, $\mathbf{D}X_k' = \sigma_k^2$, а выполнение утверждения теоремы (10.6) или (10.7) означает одновременное выполнение усиленного закона больших чисел и для новых случайных величин. Поэтому в дальнейшем при доказательстве теоремы будем считать, что $m_k = 0$.

Обозначим $S_n = X_1 + \ldots + X_n$ и перепишем S_n, учитывая, что всегда найдется такое m, что $2^m \leqslant n < 2^{m+1}$, в следующем виде:

$$S_n = X_1 + (X_2 + X_3) + (X_4 + X_5 + X_6 + X_7) + \ldots + (X_{2^m} + \ldots + X_n) =$$

$$= \sum_{i=0}^{m-1} \left(\sum_{k=2^i}^{2^{i+1}-1} X_k \right) + \sum_{k=2^m}^{n} X_k.$$

Тогда

$$|S_n| \leqslant \sum_{i=0}^{m-1} \max_{2^i \leqslant j < 2^{i+1}} \left| \sum_{k=2^i}^{j} X_k \right| + \max_{2^m \leqslant j \leqslant n} \left| \sum_{k=2^m}^{j} X_k \right|.$$

Введем обозначения для случайных величин

$$\eta_i = \frac{1}{2^i} \left(\max_{2^i \leqslant j < 2^{i+1}} \left| \sum_{k=2^i}^{j} X_k \right| \right), \tag{10.8}$$

тогда

$$|S_n| \leqslant \eta_0 + 2\eta_1 + 2^2\eta_2 + \ldots + 2^m \eta_m,$$

и значит,

$$\frac{|S_n|}{n} \leqslant \frac{\eta_0 + 2\eta_1 + \ldots + 2^m \eta_m}{2^m}. \tag{10.9}$$

Лемма 10.3. *Пусть $\{a_n\}_0^\infty, \{b_n\}_0^\infty$ — две последовательности чисел, связанных соотношением*

$$b_n = \frac{a_0 + 2a_1 + \ldots + 2^n a_n}{2^n}.$$

Тогда

$$a_n \to 0 \Leftrightarrow b_n \to 0,$$

то есть последовательности чисел a_n и b_n стремятся (или не стремятся) к 0 одновременно.

Доказательство. 1. Пусть $b_n \to 0$. Тогда $2^n b_n = 2^{n-1} b_{n-1} + 2^n a_n$. Откуда следует

$$a_n = b_n - \frac{1}{2} b_{n-1} \to 0.$$

2. Пусть теперь $a_n \to 0$. Тогда $\forall \varepsilon > 0 \ \exists n_1$ такое, что при всех $n \geqslant n_1$ $|a_n| < \frac{\varepsilon}{4}$, а также существует константа $c > 0$ такая, что $|a_n| < c$ для всех n. Проведя несложные выкладки, получим

$$|b_n| \leqslant \frac{|a_0| + 2|a_1| + \ldots + 2^{n_1 - 1}|a_{n_1 - 1}|}{2^n} + \frac{2^{n_1}|a_{n_1}| + \ldots + 2^n |a_n|}{2^n} \leqslant$$

$$\leqslant \frac{c \cdot 2^{n_1}}{2^n} + \frac{\varepsilon(2^{n_1} + 2^{n_1+1} + \ldots + 2^n)}{4 \cdot 2^n} = \frac{c \cdot 2^{n_1}}{2^n} + \frac{\varepsilon \cdot 2^{n_1} \cdot (2^{n-n_1+1} - 1)}{4 \cdot 2^n}.$$

Последнее выражение можно сделать меньше ε, если выбрать n достаточно большим. Следовательно, $b_n \to 0$. Лемма доказана. □

Возвращаясь к (10.9), видим, что для доказательства теоремы осталось показать, что $\eta_n(\omega) \xrightarrow{\text{п.н.}} 0$.

Действительно, применяя неравенство Колмогорова для максимума сумм независимых случайных величин и проводя несложные выкладки, получим $\forall \varepsilon > 0$:

$$\mathbf{P}\{|\eta_n| \geqslant \varepsilon\} = \mathbf{P}\left\{\max_{2^n \leqslant j < 2^{n+1}} \left|\sum_{k=2^n}^{j} X_k\right| \geqslant \varepsilon \cdot 2^n\right\} \leqslant \frac{1}{2^{2n}\varepsilon^2} \sum_{k=2^n}^{2^{n+1}-1} \sigma_k^2 \leqslant$$

$$\leqslant \frac{2^{2(n+1)}}{\varepsilon^2 2^{2n}} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} = \frac{4}{\varepsilon^2} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} \to 0 \quad \text{при} \quad n \to \infty,$$

что следует из сходимости соответствующего ряда по условию теоремы. Но это лишь доказывает сходимость $\eta_n \to 0$ по вероятности[1]. Для доказательства $\eta_n(\omega) \xrightarrow{\text{п.н.}} 0$ применим лемму Бореля–Кантелли. Поскольку из условия теоремы и выше доказанного неравенства следует, что

$$\sum_{n=0}^{\infty} \mathbf{P}\{|\eta_n(\omega)| \geqslant \varepsilon\} \leqslant \frac{4}{\varepsilon^2} \sum_{n=0}^{\infty} \sum_{k=2^n}^{2^{n+1}-1} \frac{\sigma_k^2}{k^2} = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty,$$

то с вероятностью 1 осуществится лишь конечное число событий среди событий $\{|\eta_n(\omega)| \geqslant \varepsilon\}$, и значит, почти для всех ω, начиная с некоторого номера $N(\omega)$, выполняется неравенство $|\eta_n(\omega)| < \varepsilon$. Тем самым доказано, что

$$\mathbf{P}\{\omega\colon \eta_n(\omega) \to 0\} = 1.$$

□

Замечание. Достаточное для выполнения усиленного закона больших чисел условие

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

не является необходимым, но оно является неулучшаемым в том

[1] См. раздел «Сходимость по вероятности» главы 12.

смысле, что если ряд $\sum\limits_{k=1}^{\infty} \frac{\sigma_k^2}{k^2}$ расходится, то существует последовательность независимых случайных величин $\xi_1, \xi_2, \ldots$ таких, что

$$\mathbf{E}\,\xi_k = 0, \quad \mathbf{D}\,\xi_k = \sigma_k^2,$$

но усиленный закон больших чисел (10.6) не выполняется.

Пример 10.1. Пример также принадлежит А. Н. Колмогорову. Последовательность строится следующим образом. Для тех k, при которых $\frac{\sigma_k^2}{k^2} \leqslant 1$, выберем

$$\xi_k = \begin{cases} k, & \text{с вероятностью} \ \ \sigma_k^2/(2k^2), \\ k, & \text{с вероятностью} \ \ \sigma_k^2/(2k^2), \\ 0, & \text{с вероятностью} \ \ 1 - \sigma_k^2/k^2. \end{cases}$$

Если $\frac{\sigma_k^2}{k^2} > 1$, то для этих k

$$\xi_k = \begin{cases} \sigma_k, & \text{с вероятностью} \ \ 1/2, \\ -\sigma_k, & \text{с вероятностью} \ \ 1/2. \end{cases}$$

Предположим, что для $S_n = \sum\limits_{k=1}^{n} \xi_k$ выполняется неравенство

$$\mathbf{P}\left\{\frac{S_n}{n} \to 0\right\} > 0.$$

Но тогда из представления

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \cdot \frac{S_{n-1}}{n-1}$$

следует, что $\mathbf{P}\left\{\frac{\xi_n}{n} \to 0\right\} > 0$. С другой стороны,

$$\sum_{n=1}^{\infty} \mathbf{P}\left\{\left|\frac{\xi_n}{n}\right| \geqslant 1\right\} = \infty,$$

так как члены этого ряда или равны 1, или равны членам расходящегося ряда $\frac{\sigma_n^2}{n^2}$. В силу леммы Бореля–Кантелли с вероятностью 1 происходит бесконечно много событий $A_n = \left\{\left|\frac{\xi_n}{n}\right| \geqslant 1\right\}$, и значит, $\frac{\xi_n}{n}$ не может с положительной вероятностью сходиться к 0. Следовательно, для этой последовательности усиленный закон больших чисел не выполняется.

Приведем без доказательства еще одну теорему А. Н. Колмогорова.

Теорема 10.4. *Пусть $X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин. Тогда*

1) *условие $\mathbf{E}|X_i| < \infty$ является достаточным для выполнения усиленного закона больших чисел:*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{п.н.} \mathbf{E}X_1;$$

2) *если существует константа m такая, что $\frac{1}{n}\sum\limits_{i=1}^{n} X_i \xrightarrow{п.н.} m$, то $\mathbf{E}|X_i| < \infty$ и $m = \mathbf{E}X_1$.*

10.4. Сходимость рядов из независимых случайных величин. Закон «0 или 1»

Всюду в данном разделе будем полагать, что задано некоторое вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$, на котором в дальнейшем рассматриваются последовательности случайных величин и связанных с ними случайных событий.

Из леммы Бореля–Кантелли следовало, что ряды из индикаторов независимых случайных событий либо почти всюду сходятся, либо почти всюду расходятся. А как ведут себя другие ряды из независимых случайных величин?

Пример 10.2. Хорошо известно, что гармонический ряд $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ расходится, а знакочередующийся ряд $\sum\limits_{n=1}^{\infty} \frac{(-1)^n}{n}$ сходится. А что будет, если знаки расставлять случайным образом, например, если рассмотреть ряд $\sum\limits_{n=1}^{\infty} \frac{X_n}{n}$, где случайные величины $X_1, X_2, \ldots$ независимы, одинаково распределены и принимают два значения ± 1 с равными вероятностями $1/2$?

Теорема 10.5 (Колмогорова–Хинчина о двух рядах). 1. *Пусть $X_1, X_2, \ldots$ — последовательность независимых слу-*

чайных величин, для которых сходятся ряды

$$\sum_{n=1}^{\infty} \mathbf{E} X_n, \quad \sum_{n=1}^{\infty} \mathbf{D} X_n. \tag{10.10}$$

Тогда

$$\mathbf{P}\left\{\omega\colon \sum_{n=1}^{\infty} X_n(\omega) \text{ \textit{сходится}}\right\} = 1.$$

2. *Если дополнительно к условиям теоремы случайные величины равномерно ограничены некоторой константой $c > 0$, то условие сходимости рядов (10.10) будет также необходимым.*

Доказательство. Проведем только доказательство утверждения 1. Из условия теоремы следует, что множества сходимости рядов $\sum\limits_{n=1}^{\infty} X_n$ и $\sum\limits_{n=1}^{\infty} (X_n - \mathbf{E} X_n)$ совпадают. Поэтому можно в дальнейшем считать $\mathbf{E} X_n = 0$, $\mathbf{D} X_n = \sigma_n^2$.

Обозначим A множество сходимости ряда $\sum\limits_{n=1}^{\infty} X_n$. Используя критерий Коши для сходимости рядов, можно записать

$$A = \bigcap_{\varepsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n \geqslant N} \bigcap_{p=1}^{\infty} \left\{\omega\colon \left|\sum_{k=n}^{n+p} X_k(\omega)\right| < \varepsilon\right\}.$$

Тогда множество $\overline{A}$, где ряд расходится, равно

$$\overline{A} = \bigcup_{\varepsilon > 0} \bigcap_{N=1}^{\infty} \bigcup_{n \geqslant N} \bigcup_{p=1}^{\infty} \left\{\omega\colon \left|\sum_{k=n}^{n+p} X_k(\omega)\right| \geqslant \varepsilon\right\} =$$

$$= \bigcup_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n \geqslant N} \bigcup_{p=1}^{\infty} \left\{\omega\colon \left|\sum_{k=n}^{n+p} X_k(\omega)\right| > \frac{1}{m}\right\} =$$

$$= \bigcup_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \left\{\omega\colon \sup_{n \geqslant N, p \geqslant 1} \left|\sum_{k=n}^{n+p} X_k(\omega)\right| > \frac{1}{m}\right\}.$$

Покажем, что $\forall\, m$

$$\mathbf{P}\left\{\bigcap_{N=1}^{\infty} \left\{\omega\colon \sup_{n \geqslant N,\ p \geqslant 1} \left|\sum_{k=n}^{n+p} X_k(\omega)\right| > \frac{1}{m}\right\}\right\} = 0.$$

Для этого оценим вероятность события

$$B_N = \left\{ \omega\colon \sup_{n \geqslant N,\, p \geqslant 1} \left| \sum_{k=n}^{n+p} X_k(\omega) \right| > \frac{1}{m} \right\}.$$

Зафиксируем q и $n \geqslant N$ и с помощью неравенства Колмогорова оценим

$$\mathbf{P}\left\{ \omega\colon \max_{1 \leqslant p \leqslant q} \left| \sum_{k=n}^{n+p} X_k(\omega) \right| > \frac{1}{m} \right\} \leqslant m^2 \sum_{k=n}^{n+q} \sigma_k^2 \leqslant m^2 \sum_{k=n}^{\infty} \sigma_k^2 \leqslant m^2 \sum_{k=N}^{\infty} \sigma_k^2.$$

Поскольку полученная оценка не зависит ни от q, ни от n, ее можно использовать для оценки вероятности события B_N:

$$\mathbf{P}(B_N) \leqslant m^2 \sum_{k=N}^{\infty} \sigma_k^2 \to 0 \quad \text{при } N \to \infty.$$

Но тогда $\forall\, m$

$$\mathbf{P}\left\{ \bigcap_{N=1}^{\infty} B_N \right\} = \lim_{N \to \infty} \mathbf{P}(B_N) = 0,$$

и следовательно, $\mathbf{P}(\overline{A}) = 0$. □

Чтобы сформулировать утверждение следующей теоремы, введем обозначение

$$X_n^c = \begin{cases} X_n, & |X_n| \leqslant c, \\ 0, & |X_n| > c. \end{cases}$$

Теорема 10.6 (Колмогорова о трех рядах). *Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин. Тогда для сходимости с вероятностью 1 ряда $\sum\limits_{n=1}^{\infty} X_n$ необходимо и достаточно, чтобы при некотором $c > 0$ одновременно сходились три ряда*

$$\sum_{n=1}^{\infty} \mathbf{P}\{|X_n| > c\}, \quad \sum_{n=1}^{\infty} \mathbf{E} X_n^c, \quad \sum_{n=1}^{\infty} \mathbf{D} X_n^c. \tag{10.11}$$

Доказательство теоремы можно посмотреть в [14; 30].

Задача 10.2. Покажите, что из сходимости с вероятностью 1 ряда $\sum\limits_{n=1}^{\infty} X_n$ из независимых случайных величин следует, что для всякого $c > 0$ сходятся ряды (10.11).

Последняя теорема дает необходимые и достаточные условия для сходимости с вероятностью 1 рядов из независимых случайных величин, но не объясняет, что же будет, если сходимости с вероятностью 1 нет. Покажем, что так же, как и для индикаторов независимых событий, ряды с вероятностью 1 будут расходиться.

Пусть по-прежнему $X_1, X_2, \ldots$ — последовательность независимых случайных величин на $(\Omega, \mathcal{A}, \mathbf{P})$. Обозначим $\mathcal{A}_n$ — σ-алгебру, порожденную случайными величинами $X_1, \ldots, X_n$, то есть наименьшую σ-алгебру, содержащую все события вида

$$\{\omega\colon (X_1(\omega), \ldots, X_n(\omega)) \in B\}, \quad B \in \mathcal{B}^n,$$

где $\mathcal{B}^n$ обозначает борелевскую σ-алгебру в $\mathbb{R}^n$.

Аналогично пусть $\mathcal{A}_n^\infty$ — σ-алгебра, порожденная случайными величинами $X_n, X_{n+1}, \ldots$.

Рассмотрим произвольные события A и B такие, что $A \in \mathcal{A}_{n-1}$, а $B \in \mathcal{A}_n^\infty$. Поскольку событие A определяется случайными величинами X_i с номерами $1 \leqslant i \leqslant n-1$, а событие B — случайными величинами с номерами $i \geqslant n$, то из независимости случайных величин следует независимость событий A и B. Такие σ-алгебры называются *независимыми.*

Назовем *остаточной* для последовательности $X_1, X_2, \ldots$ σ-алгебру

$$\mathcal{E} = \bigcap_{n=1}^{\infty} \mathcal{A}_n^\infty,$$

а всякое событие $A \in \mathcal{E}$ — *остаточным событием.* Из определения следует, что всякое остаточное событие не зависит от любого конечного числа $X_1, \ldots, X_n$. Примерами остаточных событий могут служить события

$$\{\omega\colon \text{сходится } X_n(\omega)\}, \quad \{\omega\colon \text{сходится } \sum_{n=1}^{\infty} X_n(\omega)\}.$$

Эти события не изменятся, если откинуть любое конечное число первых членов последовательности.

Задача 10.3. Приведите другие примеры остаточных событий.

Теорема 10.7 (закон «0 или 1» Колмогорова). *Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин. Тогда для всякого остаточного события $A \in \mathcal{E}$ выполняется $\mathbf{P}(A) = 0$ или $\mathbf{P}(A) = 1$.*

Доказательство. Пусть $A \in \mathcal{E} = \bigcap\limits_{n=1}^{\infty} \mathcal{A}_n^{\infty}$, следовательно, $A \in \mathcal{A}_n^{\infty}$ для любого n и тем самым не зависит от $\mathcal{A}_{n-1}$ при всех n. Обозначим $\mathcal{A}_X$ наименьшую σ-алгебру, содержащую все $\mathcal{A}_n$. Тогда для любого $B \in \mathcal{A}_X$

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Так как $\mathcal{E} \subset \mathcal{A}_X$, то это равенство выполняется и для A:

$$\mathbf{P}(A \cap A) = \mathbf{P}(A) \cdot \mathbf{P}(A).$$

Следовательно, $\mathbf{P}(A)$ удовлетворяет уравнению

$$\mathbf{P}(A) = \mathbf{P}(A) \cdot \mathbf{P}(A),$$

и значит, $\mathbf{P}(A) = 0$ или $\mathbf{P}(A) = 1$. □

Поскольку множество сходимости ряда является остаточным событием, то получаем следующее следствие.

Следствие 10.1. *Ряд из независимых случайных величин или почти всюду сходится, или почти всюду расходится.*

Замечание. Ряды из функций, обладающие подобным свойством, встречаются и в математическом анализе. Так, можно рассмотреть на отрезке $[0,1]$ ряд $\sum\limits_{n=1}^{\infty} a_n \cos(2^n \cdot 2\pi x)$. Этот ряд или почти всюду на $[0,1]$ сходится, или почти всюду расходится, в зависимости от того, сходится или расходится ряд $\sum\limits_{n=1}^{\infty} a_n^2$.

Глава 11

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ И МЕТОД ХАРАКТЕРИСТИЧЕСКИХ ФУНКЦИЙ

11.1. Обозначения и формулировки предельных теорем

Определение 11.1. Случайная величина Z имеет *стандартное нормальное распределение*, если ее плотность распределения $\varphi(x)$ задается формулой

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

при всех значениях x. Этот факт будем обозначать $Z \sim N(0,1)$.

Если плотность распределения случайной величины Y равна

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)},$$

то распределение с такой плотностью называется *нормальным с параметрами* a и σ^2. В дальнейшем будем в этом случае писать $Y \sim N(a,\sigma^2)$. Заметим, что $\mathbf{E}Y = a$, $\mathbf{D}Y = \sigma^2$. Для стандартного нормального распределения $a = 0$, $\sigma = 1$.

Пусть $X_1, X_2, \ldots$ — независимые случайные величины, для которых существуют $\mathbf{E}X_k = a_k$, $\mathbf{D}X_k = b_k^2$, $\mathbf{E}|X_k - a_k|^3 = c_k$.

Обозначим

$$\begin{aligned}
S_n &= X_1 + X_2 + \ldots + X_n,\\
A_n &= \mathbf{E}S_n = a_1 + a_2 + \ldots + a_n,\\
B_n^2 &= \mathbf{D}S_n = b_1^2 + b_2^2 + \ldots + b_n^2,\\
C_n &= c_1 + c_2 + \ldots + c_n.
\end{aligned}$$

Введем новую случайную величину

$$S_n^* = \frac{S_n - A_n}{B_n},$$

которую называют *нормированной суммой*. Цель данной нормировки в том, чтобы сделать $\mathbf{E}S_n^* = 0$, $\mathbf{D}S_n^* = 1$.

Приведем без доказательства следующую теорему.

Теорема 11.1 (А. М. Ляпунов).

$$\Delta_n = \sup_{-\infty<c<d<+\infty} \left| \mathbf{P}\{c \leqslant S_n^* \leqslant d\} - \frac{1}{\sqrt{2\pi}} \int_c^d e^{-x^2/2}\, dx \right| \leqslant 2\frac{C_n}{B_n^3}. \quad (11.1)$$

Поясним содержание теоремы. Если величина, стоящая в правой части неравенства, мала, то это означает, что распределение S_n^* близко к стандартному нормальному распределению, и теорема Ляпунова позволяет оценить точность данного приближения[1].

Замечание. В теории вероятностей всякое утверждение о сходимости распределений сумм случайных величин к нормальному распределению относят к *центральной предельной теореме.* Сформулированная выше теорема — это центральная предельная теорема Ляпунова.

Приведем несколько примеров.

Пример 11.1. Пусть существует такая константа $L > 0$, что

$$|X_k - a_k| \leqslant L$$

при любом $k \geqslant 1$. Тогда $|X_k - a_k|^3 \leqslant L(X_k - a_k)^2$, а значит, $c_k \leqslant Lb_k^2$ и, следовательно, $C_n \leqslant LB_n^2$. В этом случае точность нормальной аппроксимации оценивается величиной

$$\Delta_n \leqslant \frac{2L}{B_n}.$$

Таким образом, если L мало по сравнению со стандартным отклонением всей суммы S_n, то можно считать распределение S_n^* «почти нормальным».

[1] А. М. Ляпунов, впервые применив метод характеристических функций, получил более грубую оценку вида $c\dfrac{\ln n}{\sqrt{n}}$, где c — некоторая абсолютная константа (см. [21, т. 1, с. 157–176] или [18, с. 97–100]). Позже с помощью неравенства Берри–Эссеена была получена оценка, имеющая такой же вид, как в неравенстве (11.1), но с большей константой (см., например, [29, т. 2, с. 608–611]). Значение абсолютной константы c после этого неоднократно улучшалось. В статье В. М. Золотарева [11] доказано, что $c \leqslant 1.8102$, и это не окончательный результат. Исследования в этой области продолжаются.

Пример 11.2. Пусть случайные величины X_k равномерно ограничены, то есть существует константа $L' > 0$ такая, что для всех k одновременно $|X_k| \leqslant L'$. Тогда $|X_k - a_k| \leqslant 2L'$, и получим, что $C_n \leqslant 2L'B_n^2$. Следовательно, если отдельные слагаемые малы по сравнению с B_n, то распределение нормированной суммы снова близко к нормальному.

Пример 11.3. Пусть $A_1, A_2, \ldots$ — последовательность независимых событий, причем для любого k

$$\mathbf{P}(A_k) = p, \quad 0 < p < 1, \quad q = 1 - p.$$

Обозначим $\mu_n(\omega) = \sum_{i=1}^{n} I_{A_i}(\omega)$ — число произошедших событий среди первых n. По теореме Бернулли $\frac{\mu_n}{n} \to p$ по вероятности. Так как случайная величина μ_n имеет биномиальное распределение с параметрами (n, p), то $\mathbf{E}\,\mu_n = np$, $\mathbf{D}\,\mu_n = npq$, и в этом случае

$$S_n^* = \frac{\mu_n - np}{\sqrt{npq}}.$$

Поскольку для индикатора любого случайного события A выполняется неравенство $|I_A(\omega) - p| \leqslant 1$, то в нашем случае можно применить вывод примера 11.1 и получить оценку

$$\Delta_n \leqslant \frac{2}{\sqrt{npq}}.$$

Нормальной аппроксимацией для биномиального распределения можно пользоваться, если значение величины $\frac{2}{\sqrt{npq}}$ мало.

Как следствие из теоремы Ляпунова получается следующая теорема.

Теорема 11.2 (интегральная теорема Муавра–Лапласа). *Если $n \to \infty$, то*

$$\mathbf{P}\{m_1 \leqslant \mu_n(\omega) \leqslant m_2\} \to \int\limits_c^d \varphi(x)\,dx,$$

где $c = \frac{m_1 - np}{\sqrt{npq}}$, $d = \frac{m_2 - np}{\sqrt{npq}}$.

Прямой вывод интегральной теоремы Муавра–Лапласа состоит в непосредственном подсчете вероятностей биномиального распределения

$$\mathbf{P}\{m_1 \leqslant \mu_n \leqslant m_2\} = \sum_{m=m_1}^{m_2} C_n^m p^m q^{n-m}$$

и применении формулы Стирлинга к отдельным слагаемым данной суммы.

Теорема Муавра–Лапласа позволяет также вычислять вероятности отклонений относительной частоты наступления события от его вероятности. Действительно,

$$\mathbf{P}\left\{c \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant d\right\} = \mathbf{P}\left\{c\frac{\sqrt{pq}}{\sqrt{n}} \leqslant \frac{\mu_n}{n} - p \leqslant d\frac{\sqrt{pq}}{\sqrt{n}}\right\}.$$

При этом видно, что эти отклонения имеют порядок $\sqrt{pq}/\sqrt{n}$.

Замечание. В практических задачах обычно заменяют

$$\mathbf{P}\{c \leqslant S_n^* \leqslant d\} \approx \int_c^d \varphi(x)\,dx,$$

не учитывая при этом, что в случаях, когда значения этого интеграла малы, применение теоремы неоправданно, так как погрешность аппроксимации может оказаться больше, чем значения интеграла.

Математики затратили немало усилий, чтобы справиться с этими трудностями. Соответствующие теоремы носят название теорем о больших уклонениях. Более подробно об этом можно прочитать в [29].

Попытки получить обобщение теоремы Муавра–Лапласа привели к созданию метода характеристических функций (А. М. Ляпунов, 1901).

11.2. Характеристические функции. Определение и свойства

Определение 11.2. *Характеристической функцией* случайной величины X называется функция $f(t;X)$ действительной переменной t, ($t \in (-\infty;+\infty)$), определяемая равенством

$$f(t;X) = \mathbf{E}\, e^{itX},$$

в котором i обозначает мнимую единицу.

Напомним, что по известной формуле Эйлера

$$e^{itX} = \cos tX + i \sin tX.$$

Математическое ожидание этой комплекснозначной случайной величины определим как

$$\mathbf{E}\, e^{itX} = \mathbf{E} \cos tX + i\, \mathbf{E} \sin tX.$$

Поскольку всякая ограниченная случайная величина имеет конечное математическое ожидание, то характеристическая функция существует для произвольной случайной величины. Иногда для краткости, когда понятно, о какой случайной величине идет речь, будем обозначать характеристическую функцию $f(t)$.

Рассмотрим два частных случая.

1. Пусть X — дискретная случайная величина и распределение X задано таблицей

x_1	x_2	$\ldots$	x_n	$\ldots$
p_1	p_2	$\ldots$	p_n	$\ldots$

Тогда характеристическая функция этой случайной величины равна

$$f(t;X) = \sum_{k=1}^{\infty} e^{itx_k} p_k,$$

причем если случайная величина X принимает конечное число значений, то и в сумме будет такое же число слагаемых. В случае бесконечного числа значений абсолютная сходимость ряда не вызывает сомнений.

2. Пусть теперь X имеет абсолютно непрерывное распределение с плотностью распределения вероятностей $p(x)$. Тогда ее характеристическая функция равна

$$f(t;X) = \int_{-\infty}^{+\infty} e^{itx} p(x)\, dx.$$

Замечание. Для читателей, знакомых с методами функций комплексного переменного, этот интеграл можно вычислять, не используя формулу Эйлера.

Пример 11.4. Рассмотрим в качестве примера вычисление характеристической функции индикатора случайного события A:

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \quad \text{с вероятностью} \quad p = P(A), \\ 0, & \omega \notin A, \quad \text{с вероятностью} \quad q = P(\overline{A}). \end{cases}$$

Характеристическая функция равна

$$f(t; I_A(\omega)) = f(t) = pe^{it} + q,$$

если $p = q = \frac{1}{2}$, то $f(t) = e^{it/2} \cos \frac{t}{2}$.

Пример 11.5. Вычислим характеристическую функцию случайной величины X с симметричным распределением:

$$X = \begin{cases} +1, & \text{с вероятностью} \quad \frac{1}{2}, \\ -1, & \text{с вероятностью} \quad \frac{1}{2}. \end{cases}$$

Нетрудно показать, что в этом случае характеристическая функция равна $f(t) = \cos t$.

Пример 11.6. Пусть $Z \sim N(0,1)$, тогда можно показать, что характеристическая функция равна

$$f(t) = e^{-t^2/2}.$$

Действительно, из симметричности распределения следует, что $\mathbf{E} \sin tZ = 0$ и характеристическая функция

$$f(t) = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos tx \, dx.$$

Вычислим производную этой функции, проведя дифференцирование под знаком интеграла и применив формулу интегрирования по частям:

$$f'(t) = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} (-\sin tx) x \, dx = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \sin tx \, d\left(e^{-x^2/2}\right) =$$

$$= -\int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos tx \cdot t \, dx = -tf(t).$$

Таким образом, характеристическая функция удовлетворяет дифференциальному уравнению

$$f'(t) = -tf(t), \quad f(0) = 1,$$

единственным решением которого является $f(t) = e^{-t^2/2}$.

Рассмотрим еще один пример вычисления характеристической функции симметрично распределенной случайной величины.

Пример 11.7. Пусть случайная X величина имеет равномерное на отрезке $[-a, a]$ распределение. Тогда ее плотность распределения равна

$$p(x) = \begin{cases} \dfrac{1}{2a}, & \text{если } \ |x| \leqslant a, \\ 0, & \text{если } \ |x| > a. \end{cases}$$

Характеристическая функция равна

$$f(t) = \int\limits_{-\infty}^{+\infty} e^{itx} p(x)\,dx = \int\limits_{-a}^{a} \cos tx \cdot \frac{dx}{2a} = \frac{\sin at}{at}.$$

Задача 11.1. Покажите, что для случайной величины X с плотностью распределения

$$p(x) = \begin{cases} \dfrac{1}{a}\left(1 - \dfrac{|x|}{a}\right), & \text{если } \ |x| \leqslant a, \\ 0, & \text{если } \ |x| > a, \end{cases}$$

характеристическая функция равна

$$f(t) = \frac{2(1 - \cos at)}{a^2 t^2}.$$

Это распределение по виду плотности называют *треугольным*.

Задача 11.2. Покажите, что характеристическая функция симметрично распределенной случайной величины является четной вещественной функцией.

Во многих учебниках по теории вероятностей имеются таблицы характеристических функций для наиболее часто встречающихся распределений. Существуют также более полные таблицы преобразований Фурье (а характеристическая функция есть не что иное, как преобразование Фурье плотности или функции распределения). В приложении 1 настоящего учебника приведены характеристические функции для некоторых часто используемых распределений.

Рассмотрим основные свойства характеристических функций[1].

1. *Мультипликативность*: если случайные величины X и Y независимы, то

$$f(t;X+Y)=f(t;X)\cdot f(t;Y)$$

— характеристическая функция суммы равна произведению характеристических функций слагаемых.

Доказательство. Пользуясь известными формулами тригонометрии и независимостью случайных величин, получаем

$$\begin{aligned} f(t;X+Y) &= \mathbf{E}\cos t(X+Y)+i\,\mathbf{E}\sin t(X+Y)= \\ &= \mathbf{E}\cos tX\,\mathbf{E}\cos tY-\mathbf{E}\sin tX\,\mathbf{E}\sin tY+ \\ &\quad + i\,\mathbf{E}\sin tX\,\mathbf{E}\cos tY+i\,\mathbf{E}\cos tX\,\mathbf{E}\sin tY=f(t;X)f(t;Y). \end{aligned}$$

□

2. *Для любой характеристической функции*

$$|f(t;X)|\leqslant 1,\quad f(0;X)=1.$$

Задача 11.3. Покажите, что если характеристическая функция

$$f(2\pi;X)=1,$$

то она является периодической с периодом 2π и соответствует целочисленной случайной величине.

3. *При линейном преобразовании случайной величины* X характеристическая функция изменяется следующим образом:

$$f(t;aX+b)=e^{itb}f(at;X).$$

Пример 11.8. Пусть $X\sim N(0,1)$, $Y=\sigma X+a$. Тогда $Y\sim$ $\sim N(a,\sigma^2)$, и характеристическая функция случайной величины Y равна

$$f(t;Y)=e^{ita}\,e^{-t^2\sigma^2/2}.$$

[1] Другие важные свойства, устанавливающие связь характеристических функций с моментами случайной величины, см. на с. 146.

4. *Взаимная однозначность соответствия между функциями распределения и характеристическими функциями.*

Если случайная величина X имеет функцию распределения $F(x)$, то

$$f(t;X)=\int\limits_{-\infty}^{+\infty} e^{itx}\mathbf{P}_F(dx),$$

где интеграл Лебега берется по мере $\mathbf{P}_F$, порожденной на числовой прямой функцией распределения $F(x)$. Следовательно, всякая функция распределения однозначно определяет характеристическую функцию. Но верно и обратное утверждение: функция распределения однозначно восстанавливается по характеристической функции. Доказывается это утверждение с помощью формул обращения.

11.3. Формулы обращения для характеристических функций

Прежде чем привести примеры различных формул обращения, вспомним некоторые свойства функций распределения. Мы определили ранее функцию распределения случайной величины X соотношением

$$F(x)=\mathbf{P}(X<x). \tag{11.2}$$

В учебной литературе можно встретить и другое определение[1], а именно

$$F(x)=\mathbf{P}(X\leqslant x). \tag{11.3}$$

По-разному определенные функции различаются только поведением в точках разрыва: первая из них непрерывна слева, а вторая справа. В точках непрерывности эти функции совпадают.

Всякая монотонно неубывающая ограниченная функция имеет не более чем счетное множество точек разрыва. Действительно, разрывов, по величине превосходящих 1/2, может быть не более одного, разрывов, превышающих 1/3, — не более двух, и т. д. Таким образом, все точки разрыва можно перенумеровать, и значит, их не более чем счетное число.

[1] Определение (11.2) приводится также в [4; 6; 28] и др., определение (11.3) используется, например, в [26; 29; 30].

Точек же непрерывности — континуум, причем множество точек непрерывности всякой функции распределения всюду плотно на числовой прямой.

Если $F_1(x)$ и $F_2(x)$ — две функции распределения, то множество общих точек непрерывности также всюду плотно.

Приведем без доказательства в качестве одной из формул обращения следующую теорему.

Теорема 11.3 (формула обращения[1]). *Если $F(x)$ и $f(t)$ — соответственно функция распределения и характеристическая функция случайной величины X, $x_1 < x_2$ — точки непрерывности $F(x)$, то*

$$F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi} \int\limits_{-T}^{T} \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t)\,dt.$$

Замечание. Если в этой формуле устремить $x_1 \to -\infty$ по точкам непрерывности, то получим $F(x_2)$. Поскольку x_2 — произвольная точка непрерывности, то функцию распределения можно однозначно восстановить по характеристической функции.

Существуют и другие варианты формул обращения. Так, если характеристическая функция абсолютно интегрируема, то функция распределения имеет плотность $p(x)$ и

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} f(t)\,dt. \tag{11.4}$$

Формулу 11.4 называют формулой обращения для плотности.

Задача 11.4. Докажите формулу обращения (11.4) для плотности, используя формулу обращения для функции распределения.

Замечание. Формулы

$$f(t) = \int\limits_{-\infty}^{\infty} e^{itx} p(x)\,dx,$$

[1] Доказательство этой теоремы можно прочитать, например, в [30].

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} f(t)\, dt$$

очень похожи, за исключением знака в показателях экспонент и множителя перед вторым интегралом. Первую из них называют преобразованием Фурье функции $p(x)$, вторую — обратным преобразованием Фурье. Подробно с теорией преобразований Фурье вы познакомитесь в курсе функционального анализа.

Приведем еще один вариант формулы обращения для целочисленной случайной величины X[1]:

$$\mathbf{P}\{\omega\colon X(\omega) = m\} = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} e^{-itm} f(t)\, dt.$$

Итак, по характеристической функции однозначно восстанавливается распределение случайной величины.

Таким образом, между функциями распределения и характеристическими функциями существует взаимно однозначное соответствие. В этом состояло свойство 4 характеристических функций.

11.4. Свойство непрерывности соответствия характеристических функций и функций распределения

Для того чтобы сформулировать точно это свойство, введем обозначения.

Пусть $X_1, X_2, \ldots$ — последовательность случайных величин. Обозначим $F_n(x)$ и $f(t; X_n)$ соответственно функцию распределения и характеристическую функцию случайной величины X_n $(n = 1, 2, \ldots)$.

Определение 11.3. Последовательность случайных величин X_n *сходится по распределению* к некоторой случайной величине X $(X_n \xrightarrow{d} X)$ с функцией распределения $F(x)$, если $F_n(x) \to F(x)$ в каждой точке непрерывности функции $F(x)$.

Определение 11.4. Сходимость функций распределения $F_n(x) \to F(x)$ в каждой точке непрерывности предельной

[1] См. лемму 11.7, в которой содержится вывод этой формулы.

функции $F(x)$ называют *слабой сходимостью* и обозначают $F_n(x) \Rightarrow F(x)$.

Замечание. Сходимость $X_n \xrightarrow{d} X$ по распределению равносильна сходимости

$$\mathbf{E}g(X_n) \to \mathbf{E}g(X), \quad n \to \infty,$$

для любой непрерывной и ограниченной функции $g(x)$[1].

Задача 11.5. Пусть $F(x)$ — непрерывная на всей прямой функция распределения и $F_n(x) \Rightarrow F(x)$. Покажите, что последовательность функций распределения $F_n(x)$ сходится к $F(x)$ равномерно.

Сформулируем теорему о непрерывном соответствии между характеристическими функциями и функциями распределения.

Теорема 11.4. *Для сходимости $F_n(x) \Rightarrow F(x)$ необходимо и достаточно, чтобы $f(t;X_n) \to f(t;X)$ равномерно по t в каждом конечном интервале.*

Определение 11.5. Если $f(t;X_n) \to f(t;X)$ равномерно по t в каждом конечном интервале, то такой вид сходимости характеристических функций также называют *слабой сходимостью* и обозначают

$$f(t;X_n) \Rightarrow f(t;X).$$

Задача 11.6. Приведите пример последовательности характеристических функций, которые сходятся равномерно на каждом конечном интервале, но не сходятся равномерно на всей прямой.

Мы не будем приводить доказательство теоремы 11.4 в общем случае, а ограничимся частным случаем, необходимым для доказательства центральной предельной теоремы. Пусть по-прежнему $Z \sim N(0,1)$. Докажем, что из слабой сходимости характеристических функций нормированных сумм S_n^* к характеристической функции Z

$$f(t;S_n^*) \Rightarrow e^{-t^2/2} \quad \left(\mathbf{E}\, e^{itS_n^*} \Rightarrow \mathbf{E}\, e^{itZ}\right) \tag{11.5}$$

следует сходимость функций распределения нормированных

[1] Подробнее об этой сходимости можно прочитать в [30].

сумм к функции распределения стандартного нормального закона.

Идея доказательства состоит в следующем:

1) сначала мы докажем, что из (11.5) следует сходимость математических ожиданий

$$\mathbf{E}H(S_n^*) \to \mathbf{E}H(Z) \tag{11.6}$$

для достаточно широкого класса функций $H(x)$;

2) затем с помощью (11.6) можно будет доказать, что для произвольных $-\infty < c < d < \infty$ выполняется

$$\mathbf{E}I_{[c,d]}(S_n^*) \to \mathbf{E}I_{[c,d]}(Z),$$

или, что то же самое,

$$\mathbf{P}\{c \leqslant S_n^* \leqslant d\} \to \mathbf{P}\{c \leqslant Z \leqslant d\}.$$

Но это и означает утверждение центральной предельной теоремы.

Пусть выполняется (11.5).

Рассмотрим для любого натурального m, произвольных комплексных чисел $c_1, c_2, \ldots c_m$ и действительных $t_1 < t_2 < \ldots < t_m$ функцию

$$Q(x) = \sum_{k=1}^{m} c_k e^{it_k x}. \tag{11.7}$$

Поскольку $\mathbf{E}Q(S_n^*) = \sum\limits_{k=1}^{m} c_k \mathbf{E}\, e^{it_k S_n^*}$, то из (11.5) следует, что

$$\mathbf{E}Q(S_n^*) \to \sum_{k=1}^{m} c_k \mathbf{E}\, e^{it_k Z} = \mathbf{E}Q(Z).$$

Возникает вопрос, насколько богат «запас» функций действительной переменной $H(x)$, которые можно приближать функциями вида (11.7) или функциями более общего вида

$$Q(x) = \sum_{k=1}^{m} (\alpha_k \cos t_k x + i\beta_k \sin t_k x)\,, \tag{11.8}$$

где α_k, β_k $(1 \leqslant k \leqslant m)$ — произвольные комплексные числа. Приведем несколько примеров таких функций $H(x)$.

Пример 11.9. Рассмотрим функцию

$$H(x) = \begin{cases} 1 - \dfrac{|x|}{a}, & |x| \leqslant a \\ 0, & |x| > a. \end{cases}$$

Ее можно представить (см. задачу 11.1) в виде интеграла

$$H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} \frac{1 - \cos at}{at^2}\, dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \cos tx \frac{1 - \cos at}{at^2}\, dt,$$

и значит, $H(x)$ можно приблизить функциями вида (11.7).

Пример 11.10. Рассмотрим периодическую с периодом 2 функцию

$$H(x) = \frac{1}{2} + \frac{4}{\pi^2} \left(\frac{\cos \pi x}{1^2} + \frac{\cos 3\pi x}{3^2} + \frac{\cos 5\pi x}{5^2} + \ldots \right),$$

которая при $x \in [-1, 1]$ совпадает с функцией предыдущего примера, если $a = 1$. Эта функция также может быть представлена в виде предела функций (11.8).

Пример 11.11. Следующая функция

$$H(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{itx} \frac{\sin lt}{2lt} \frac{\sin \delta t}{\delta t}\, dt$$

«похожа» на индикатор отрезка $[-l, l]$, если $0 < \delta < l$ мало (рис. 7). Ее также можно представить в виде предела функций (11.8).

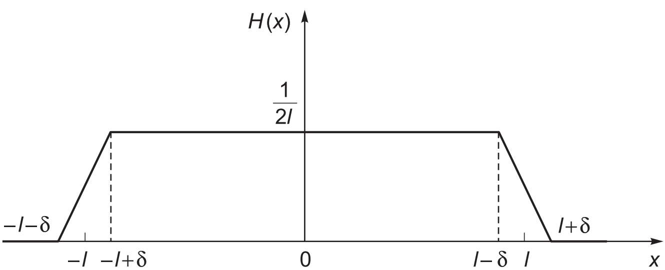

Рис. 7. Функция из примера 11.11

Задача 11.7. Покажите, что функция $H(x)$ предыдущего примера является плотностью суммы независимых случайных величин X, Y, одна из которых имеет равномерное распределение на $[-l, l]$, а другая — равномерное распределение на $[-\delta, \delta]$.

Пример 11.12 (распределение Коши). Рассмотрим преобразование Фурье плотности распределения Коши:

$$H(x) = e^{-\lambda|x|} = \int_{-\infty}^{\infty} e^{itx} \frac{\lambda}{\pi(\lambda^2 + t^2)} dt.$$

Эту функцию также можно приблизить функциями вида (11.8), рассмотрев соответствующие интегральные суммы.

Все рассмотренные в примерах функции представимы в виде

$$H(x) = \int_{-\infty}^{\infty} e^{itx} h(t)\, dt, \tag{11.9}$$

где $h(t)$ — непрерывная ограниченная функция, абсолютно интегрируемая на всей прямой. Обозначим $\mathcal{H}$ класс функций вида (11.9).

Нам для доказательства центральной предельной теоремы понадобится простое достаточное условие принадлежности функции $H(x)$ к классу $\mathcal{H}$.

Лемма 11.1. *Пусть функция $H(x)$ равна 0 вне конечного интервала, непрерывна и имеет две непрерывные производные на всей прямой. Тогда $H(x) \in \mathcal{H}$.*

Мы не будем приводить доказательство этой леммы, поскольку ее утверждение следует из известных теорем математического анализа о представлении функций интегралом Фурье[1].

Замечание. Функция примера 11.12 входит в класс $\mathcal{H}$, хотя и не удовлетворяет условиям леммы, то есть условие является достаточным, но не является необходимым.

Утверждения, которые последуют дальше, верны и для более широкого класса функций. Можно, например, отказаться от условия равенства 0 вне конечного интервала, но тогда для доказательства теорем понадобится теорема Фубини для интеграла Лебега, которая в настоящем курсе не рассматривалась.

[1] См., например, [9, с. 404].

Лемма 11.2. *Пусть $H(x) \in \mathcal{H}$. Тогда для произвольной случайной величины Y*

$$\mathbf{E}H(Y) = \int\limits_{-\infty}^{\infty} f(t;Y)h(t)\,dt.$$

Доказательство. Доказательство проведем в частных случаях: для дискретной случайной величины, принимающей конечное число значений, и для случайной величины со стандартным нормальным распределением.

1. Пусть случайная величина Y принимает s различных между собой значений $y_1, y_2, \ldots, y_s$ с вероятностями $p_1, p_2, \ldots, p_s$. Тогда

$$\mathbf{E}H(Y) = \sum_{j=1}^{s} H(y_j)p_j = \sum_{j=1}^{s} \left(\int\limits_{-\infty}^{\infty} e^{ity_j} h(t)\,dt \right) p_j =$$
$$= \int\limits_{-\infty}^{\infty} \left(\sum_{j=1}^{s} e^{ity_j} p_j \right) h(t)\,dt = \int\limits_{-\infty}^{\infty} f(t;Y)h(t)\,dt.$$

Заметим, что в рассмотренном случае правомерность замены порядков суммирования и интегрирования сомнений не вызывала.

2. Повторим рассуждения для случайной величины $Z \sim N(0,1)$ с плотностью $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$:

$$\mathbf{E}H(Z) = \int\limits_{-\infty}^{\infty} H(x)\varphi(x)\,dx = \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{\infty} e^{itx} h(t)dt \right) \varphi(x)\,dx =$$
$$= \int\limits_{-\infty}^{\infty} h(t) \left(\int\limits_{-\infty}^{\infty} e^{itx} \varphi(x)\,dx \right) dt = \int\limits_{-\infty}^{\infty} h(t) e^{-t^2/2}\,dt.$$

В этом случае замена порядков интегрирования хотя и верна, но уже не столь очевидна. Впрочем, для интегралов Римана теорема Фубини в курсе математического анализа рассматривалась, и вы можете убедиться в правильности проведенных выкладок. □

Теорема 11.5. *Пусть $H(x) \in \mathcal{H}$ и $Y_1, Y_2, \ldots$ — последовательность случайных величин, для которой характеристические* *функции сходятся к характеристической функции случайной величины* Y

$$f(t;Y_n) \Rightarrow f(t;Y)$$

равномерно в каждом конечном интервале. Тогда

$$\mathbf{E}H(Y_n) \to \mathbf{E}H(Y).$$

Доказательство. Рассмотрим разность

$$\mathbf{E}H(Y_n) - \mathbf{E}H(Y) = \int\limits_{-\infty}^{\infty} h(t)(f(t;Y_n) - f(t;Y))\,dt = I_1 + I_2 + I_3,$$

где интегралы I_1, I_2, I_3 берутся соответственно на промежутках $[-T, T]$, $(-\infty, -T]$, $[T, \infty)$. Оценим каждый из них по отдельности. Выбор константы T проведем позднее.

$$|I_1| \leqslant 2TA_h \sup_{|t| \leqslant T} |f(t;Y_n) - f(t;Y)|,$$

где $A_h = \sup\limits_{-\infty<t<+\infty} |h(t)|$,

$$|I_2| \leqslant 2 \int\limits_{-\infty}^{-T} |h(t)|\,dt, \quad |I_3| \leqslant 2 \int\limits_{T}^{\infty} |h(t)|\,dt.$$

Таким образом, получаем

$$|\mathbf{E}H(Y_n) - \mathbf{E}H(Z)| \leqslant |I_1| + |I_2| + |I_3|.$$

Поскольку $h(t)$ — абсолютно интегрируемая на всей прямой функция, то $\forall \varepsilon > 0$ существует $T = T(\varepsilon, h)$ такое, что

$$|I_2| + |I_3| < 2\varepsilon/3.$$

Оставшееся слагаемое

$$|I_1| \leqslant 2T(\varepsilon, h)A_h \sup_{|t| \leqslant T} |f(t;Y_n) - f(t;Y)|,$$

и его можно сделать меньше, чем $\varepsilon/3$, выбрав n достаточно большим, так как существует n_0 такое, что для всех $n \geqslant n_0$

$$\sup_{|t| \leqslant T} |f(t;Y_n) - f(t;Y)| < \frac{\varepsilon}{6TA_h}.$$

Следовательно, $\mathbf{E}H(Y_n) \to \mathbf{E}H(Y)$ для всех функций $H(x) \in \mathcal{H}$. $\square$

Как будет видно из доказательства следующей теоремы, аналогичное утверждение верно не только для функций из $\mathcal{H}$.

Теорема 11.6. *Если* $f(t; Y_n) \Rightarrow e^{-t^2/2}$, *то для любого отрезка* $[c, d]$

$$\lim_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\} = \mathbf{P}\{c \leqslant Z \leqslant d\}.$$

Доказательство. Рассмотрим функцию $I_{[c,d]}(x)$ — индикатор произвольного отрезка $[c, d]$. Такие функции можно приближать функциями из класса $\mathcal{H}$. А именно: для любого достаточно малого $\delta > 0$ (рис. 8) существует дважды непрерывно дифференцируемая функция $H_\delta^+(x)$ такая, что

$$H_\delta^+(x) \geqslant I_{[c,d]}(x), \quad H_\delta^+(x) = 0 \ \text{ для всех } \ x \notin [c - \delta, d + \delta].$$

В силу леммы 11.1 $H_\delta^+(x) \in \mathcal{H}$. Выпишем следующую цепочку соотношений:

$$\mathbf{P}\{c \leqslant Y_n \leqslant d\} = \mathbf{E}\, I_{[c,d]}(Y_n) \leqslant \mathbf{E}\, H_\delta^+(Y_n).$$

Переходя к пределу при $n \to \infty$, получим

$$\limsup_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\} \leqslant \limsup_{n\to\infty} \mathbf{E}\, H_\delta^+(Y_n) = \lim_{n\to\infty} \mathbf{E}\, H_\delta^+(Y_n) = \\ = \mathbf{E}\, H_\delta^+(Z) \leqslant \mathbf{E}\, I_{[c-\delta,d+\delta]}(Z).$$

Следовательно,

$$\limsup_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\} \leqslant \int\limits_{c-\delta}^{d+\delta} \varphi(x)\, dx \leqslant \\ \leqslant \int\limits_c^d \varphi(x)\, dx + \int\limits_{c-\delta}^{c} \varphi(x)\, dx + \int\limits_d^{d+\delta} \varphi(x)\, dx \leqslant \int\limits_c^d \varphi(x)\, dx + \frac{1}{\sqrt{2\pi}} 2\delta.$$

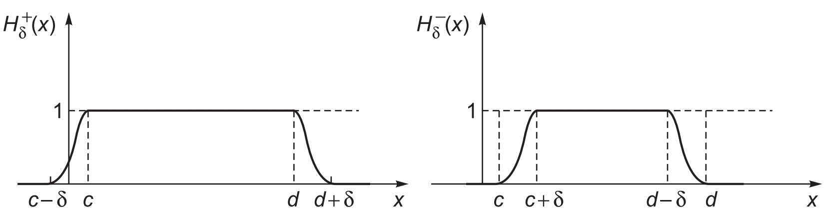

Рис. 8. Приближение индикаторов множеств функциями из $\mathcal{H}$

Устремив теперь $\delta \to +0$, получим

$$\limsup_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\} \leqslant \int_c^d \varphi(x)\,dx = \mathbf{P}\{c \leqslant Z \leqslant d\}. \qquad (11.10)$$

Аналогичным образом, выбирая $H_\delta^-(x) \in \mathcal{H}$ (см. рис. 8) такую, что

$$H_\delta^-(x) = 0 \quad \text{для всех} \quad x \notin [c,d], \qquad H_\delta^-(x) \leqslant I_{[c,d]}(x),$$

можно доказать, что

$$\mathbf{P}\{c \leqslant Z \leqslant d\} \leqslant \liminf_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\}. \qquad (11.11)$$

Сравнивая неравенства (11.10) и (11.11), получим, что

$$\lim_{n\to\infty} \mathbf{P}\{c \leqslant Y_n \leqslant d\} = \mathbf{P}\{c \leqslant Z \leqslant d\}.$$

□

11.5. Примеры слабой сходимости последовательностей характеристических функций

Рассмотрим сначала на примерах, как можно установить сходимость характеристических функций.

Пример 11.13. Пусть $X_1, X_2, \ldots$ — независимые случайные величины, принимающие значения 1 и -1 с вероятностями $1/2$, с характеристической функцией $f(t; X_k) = \cos t$, $\mathbf{E}\,X_k = 0$, $\mathbf{D}\,X_k = 1$. Тогда для нормированной суммы $S_n^* = \dfrac{S_n}{\sqrt{n}}$ имеем

$$f(t; S_n^*) = f\left(\frac{t}{\sqrt{n}}; S_n\right) = \cos^n\left(\frac{t}{\sqrt{n}}\right).$$

При оценке близости характеристических функций воспользуемся простой, но полезной леммой.

Лемма 11.3. *Если* $|\alpha| \leqslant 1$, $|\beta| \leqslant 1$, *то*

$$|\alpha^n - \beta^n| \leqslant |\alpha - \beta| \cdot n$$

для любого натурального n.

Воспользовавшись неравенством леммы 11.3, получим

$$\left|f(t;S_n^*)-e^{-t^2/2}\right| = \left|\cos^n\left(\frac{t}{\sqrt{n}}\right)-\left(e^{-t^2/(2n)}\right)^n\right| \leqslant$$
$$\leqslant n\left|\cos\left(\frac{t}{\sqrt{n}}\right)-e^{-t^2/(2n)}\right|.$$

Поскольку $|t| \leqslant T$, $n \to \infty$, то необходимо уметь оценивать эту разность лишь в малой окрестности 0. Воспользуемся конкретными свойствами функций:

$$\cos u = 1-\frac{u^2}{2!}+\frac{u^4}{4!}-\ldots+(-1)^n\frac{u^{2n}}{(2n)!}+R_n(u),$$

причем остаток $R_n(u)$ не превосходит по модулю первого отброшенного члена в разложении Тейлора и имеет знак первого отброшенного члена;

$$e^{-u} = 1-u+\frac{u^2}{2!}-\frac{u^3}{3!}+\ldots+(-1)^n\frac{u^n}{n!}+Q_n(u),$$

где остаток $Q_n(u)$ обладает аналогичным свойством для всех $u > 0$[1]. Следовательно,

$$0 \leqslant \cos u-\left(1-\frac{u^2}{2}\right) \leqslant \frac{u^4}{4!}, \quad 0 \leqslant e^{-u}-(1-u) \leqslant \frac{u^2}{2},$$

и значит,

$$0 \leqslant \cos\frac{t}{\sqrt{n}}-\left(1-\frac{t^2}{2n}\right) \leqslant \left(\frac{t^2}{n}\right)^2\frac{1}{24},$$
$$0 \leqslant e^{-t^2/(2n)}-\left(1-\frac{t^2}{2n}\right) \leqslant \left(\frac{t^2}{2n}\right)^2\frac{1}{2}.$$

Таким образом, для всех t имеем

$$\left|\cos\frac{t}{\sqrt{n}}-e^{-t^2/(2n)}\right| \leqslant \frac{t^4}{n^2}\left(\frac{1}{24}+\frac{1}{8}\right) = \frac{t^4}{6n^2}.$$

Следовательно, для всех $|t| \leqslant T$

$$|f(t;S_n^*)-e^{-t^2/2}| \leqslant n\left|\cos\frac{t}{\sqrt{n}}-e^{-t^2/(2n)}\right| \leqslant \frac{T^4}{6n} \to 0$$

равномерно в каждом конечном интервале.

[1] Такое свойство рядов Тейлора называют свойством «обертывания».

Пример 11.14. Пусть $A_1, A_2, \ldots$ — последовательность независимых событий, каждое из которых имеет вероятность p. Будем говорить, что в испытании с номером k произошел успех, если произошло событие A_k. Тогда $S_n = \sum_{k=1}^{n} I_{A_k}(\omega)$ — это число успехов в n испытаниях. Нетрудно убедиться в том, что

$$f(t; S_n^*) = \left(qe^{-itp/\sqrt{npq}} + pe^{itq/\sqrt{npq}}\right)^n.$$

Эта функция является комплекснозначной. Но и в этом случае, проводя непосредственые вычисления, можно убедиться в том, что

$$f(t; S_n^*) \Rightarrow e^{-t^2/2}.$$

11.6. Доказательство центральной предельной теоремы

Метод характеристических функций — основной при доказательстве *центральной предельной теоремы*. Ранее была сформулирована теорема Ляпунова. Приведем более простой вариант центральной предельной теоремы для одинаково распределенных случайных величин. Сформулируем центральную предельную теорему в форме Линдеберга–Леви.

Теорема 11.7. *Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины, имеющие конечные моменты первого и второго порядков: $a = \mathbf{E}X_k$, $b^2 = \mathbf{D}X_k$, $k = 1, 2, \ldots$. Тогда*

$$\sup_{-\infty \leqslant c < d \leqslant \infty} \left| \mathbf{P}\{c \leqslant S_n^* \leqslant d\} - \frac{1}{\sqrt{2\pi}} \int_c^d e^{-x^2/2}\, dx \right| \to 0 \qquad (11.12)$$

при $n \to \infty$.

Примеры предыдущего раздела показывали, каким образом можно установить, что

$$f(t; S_n^*) \Rightarrow e^{-t^2/2}.$$

Рассмотрим

$$S_n^* = \frac{S_n - an}{b\sqrt{n}} = \sum_{k=1}^{n} \frac{X_k - a}{b\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k^*,$$

где $X_k^* = \dfrac{X_k - a}{b}$, $\mathbf{E}X_k^* = 0$, $\mathbf{D}X_k^* = 1$.

Пользуясь свойствами характеристических функций, получаем

$$f(t;S_n^*)=f^n\left(t;\frac{X_1-a}{b\sqrt{n}}\right)=f^n\left(\frac{t}{\sqrt{n}};X_k^*\right).$$

Подобно тому, как мы поступали в примере 11.13 из предыдущего раздела, запишем

$$\begin{aligned}\left|f(t;S_n^*)-e^{-t^2/2}\right|=\left|f^n\left(\frac{t}{\sqrt{n}};X_1^*\right)-\left(e^{-t^2/(2n)}\right)^n\right|\leqslant\\ \leqslant n\left|f\left(\frac{t}{\sqrt{n}};X_1^*\right)-e^{-t^2/(2n)}\right|=\\ =n\left|f\left(\frac{t}{\sqrt{n}};X_1^*\right)-\left(1-\frac{t^2}{2n}\right)-\left(e^{-t^2/(2n)}-\left(1-\frac{t^2}{2n}\right)\right)\right|\leqslant\\ \leqslant n\left|f\left(\frac{t}{\sqrt{n}};X_1^*\right)-\left(1-\frac{t^2}{2n}\right)\right|+n\frac{t^4}{8n^2}.\end{aligned}$$

Если $|t|\leqslant T$, а $n\to\infty$, то $\frac{t}{\sqrt{n}}\to 0$, и значит, необходимо изучить поведение характеристических функций в окрестности 0.

Пример 11.15. Пусть случайная величина Y принимает конечное число различных значений $y_1,y_2,\ldots,y_s$ с вероятностями $q_1,q_2,\ldots,q_s$. В этом случае

$$f(t;Y)=\sum_{l=1}^{s}e^{ity_l}q_l.$$

У функции $f(t;Y)$ существуют производные всех порядков и равные

$$\frac{d^kf(t;Y)}{dt^k}=\sum_{l=1}^{s}(iy_l)^ke^{ity_l}q_l.$$

Если теперь возьмем эти производные в точке $t=0$, то получим

$$\left.\frac{d^kf(t;Y)}{dt^k}\right|_{t=0}=i^k\sum_{l=1}^{s}y_l^kq_l=i^k\,\mathbf{E}\,Y^k. \tag{11.13}$$

Таким образом, производные характеристической функции в точке $t=0$ связаны с моментами случайной величины.

Задача 11.8. Докажите в общем случае, что из существования у случайной величины конечного момента порядка l следует, что характеристическая функция l раз непрерывно дифференцируема и для ее производных ($k \leqslant l$) в точке $t = 0$ выполняется равенство (11.13).

Применим к характеристической функции из нашего примера разложение в ряд Тейлора в окрестности 0:

$$f(t;Y) = 1 + it\,\mathbf{E}\,Y + \frac{1}{2}t^2 i^2\,\mathbf{E}\,Y^2 + o(t^2).$$

Если $\mathbf{E}\,Y = 0$, $\mathbf{D}\,Y = 1$, то

$$f(t;Y) = 1 - \frac{t^2}{2} + o(t^2).$$

Если вместо t подставить $\frac{t}{\sqrt{n}}$, то получим

$$f\left(\frac{t}{\sqrt{n}};Y\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right).$$

Сформулируем общее утверждение.

Лемма 11.4. *Если у случайной величины Y существуют $\mathbf{E}\,Y, \mathbf{D}\,Y$, то в окрестности $t = 0$*

$$f(t;Y) = 1 + it\,\mathbf{E}\,Y - \frac{t^2}{2}\,\mathbf{E}\,Y^2 + o(t^2).$$

Приведем без доказательства[1] следующее утверждение.

Лемма 11.5. *При любом действительном z и любом натуральном n*

$$\left|e^{iz} - \left(1 + iz + \frac{(iz)^2}{2!} + \ldots + \frac{(iz)^n}{n!}\right)\right| \leqslant \frac{|z|^{n+1}}{(n+1)!}.$$

Лемма 11.6. *Пусть в дополнение к условиям леммы 11.4 случайная величина Y имеет третий конечный момент. Тогда*

$$\left|f(t;Y) - \left(1 + it\,\mathbf{E}\,Y - \tfrac{1}{2}t^2\,\mathbf{E}\,Y^2\right)\right| \leqslant \frac{1}{6}|t|^3\,\mathbf{E}\,|Y|^3.$$

Доказательство следует из леммы 11.5 при $n = 2$, если от обеих частей неравенства взять математические ожидания.

[1] Доказательство этой леммы можно провести самостоятельно или прочитать в учебниках [26; 30].

Возвращаясь к доказательству центральной предельной теоремы, получаем

$$\left|f(t;S_n^*) - e^{-t^2/2}\right| \leqslant n\left|f\left(\frac{t}{\sqrt{n}};X_1^*\right) - \left(1-\frac{t^2}{2n}\right)\right| + \frac{t^4}{8n} \leqslant$$
$$\leqslant \frac{n|t|^3\,\mathbf{E}|X_1^*|^3}{6(\sqrt{n})^3} + \frac{t^4}{8n}.$$

Поскольку $|t| \leqslant T$, то последнее выражение не превосходит

$$\frac{T^3\,\mathbf{E}\,|X_1-a|^3}{6\sqrt{n}\,b^3} + \frac{T^4}{8n} \to 0.$$

Следовательно, $f(t;S_n^*) \Rightarrow f(t;Z)$, причем скорость сходимости имеет порядок $O\left(\frac{1}{\sqrt{n}}\right)$.

Итак, доказана центральная предельная теорема в предположении существования конечного третьего момента. На самом деле для ее выполнения достаточно наличия первых двух моментов.

11.7. Теорема Пуассона

В теореме Пуассона устанавливается сходимость к другому распределению — распределению Пуассона. С. Пуассон — известный французский математик первой половины XIX в., доказавший, что в схеме испытаний Бернулли, когда $n \to \infty$, распределение числа успехов при определенных условиях сходится не к нормальному распределению, а к распределению Пуассона.

Рассмотрим схему серий испытаний Бернулли: в серии с номером n проводится n независимых испытаний, каждое из которых имеет вероятность успеха p_n, зависящую от номера серии. Обозначим $P_n(m)$ — вероятность того, что число успехов в n-й серии равно m.

Теорема 11.8. *Пусть* $n \to \infty$, $p_n \to 0$ *таким образом, что* $np_n \to \lambda > 0$. *Тогда*

$$P_n(m) = C_n^m p_n^m (1-p_n)^{n-m} \to \frac{e^{-\lambda}\cdot\lambda^m}{m!}.$$

Доказательство. Легко показать, что

$$P_n(0) = (1-p_n)^n = e^{n\ln(1-p_n)} = e^{-np_n\cdot\ln(1-p_n)/(-p_n)} \to e^{-\lambda},$$

$$\frac{P_n(k)}{P_n(k-1)} = \frac{n!p_n^k(1-p_n)^{n-k}}{k!(n-k)!} \cdot \frac{(k-1)!(n-k+1)!}{n!p_n^{k-1}(1-p_n)^{n-k+1}} =$$
$$= \frac{n-k+1}{k} \cdot \frac{p_n}{1-p_n} = \frac{np_n\left(1-\dfrac{k-1}{n}\right)}{k(1-p_n)} \to \frac{\lambda}{k}$$

для всех $k \geqslant 1$.

Тогда

$$P_n(m) = \frac{P_n(m)}{P_n(m-1)} \cdot \frac{P_n(m-1)}{P_n(m-2)} \cdots \frac{P_n(1)}{P_n(0)} \cdot P_n(0) \to$$
$$\to \frac{\lambda}{m} \cdot \frac{\lambda}{m-1} \cdots \frac{\lambda}{1} \cdot e^{-\lambda} = \frac{\lambda^m e^{-\lambda}}{m!}.$$

□

Определение 11.6. Распределение случайной величины X, принимающей значения $0, 1, 2, \ldots$ с вероятностями

$$\mathbf{P}\{X = m\} = \frac{\lambda^m e^{-\lambda}}{m!},$$

называется пуассоновским с параметром λ. В этом случае будем писать $X \sim \Pi(\lambda)$.

Мы доказали очень простым способом теорему Пуассона для схемы испытаний Бернулли. Эта теорема имеет обобщения, в том числе позволяющие получить оценку скорости сходимости к распределению Пуассона.

Рассмотрим следующую схему, которая называется *схемой Пуассона*.

Пусть $A_1, A_2, A_3, \ldots$ — независимые случайные события, для которых $\mathbf{P}(A_k) = p_k$, $k = 1, 2, \ldots$. Обозначим

$$\mu_n(\omega) = \sum_{k=1}^{n} I_{A_k}(\omega), \quad \lambda = \sum_{k=1}^{n} p_k,$$

$\mu_n(\omega)$— число наступивших событий A_k среди первых n событий нашей последовательности.

Замечание. Если $\mathbf{P}(A_k) = p$ для всех k, то получим как частный случай схему Бернулли, в которой случайную величину $\mu_n(\omega)$ называли числом успехов в n испытаниях.

Случайная величина $\mu_n(\omega)$ принимает целые значения от 0 до n, но выписать распределение в явном виде для различных между собой p_k и тем более работать с ним — задача достаточно сложная. Проиллюстрируем это на примере.

Пример 11.16. Рассмотрим схему Пуассона при $n=5$, $m=2$. Тогда вероятность

$$\mathbf{P}\{\mu_n=2\}=p_1p_2q_3q_4q_5+p_1q_2p_3q_4q_5+\ldots+q_1q_2q_3p_4p_5$$

выражается через сумму $C_5^2=10$ различных слагаемых.

Как видим, даже при малых n формулы получаются громоздкими, и вряд ли их можно использовать для доказательства предельных теорем. Поэтому воспользуемся методом характеристических функций.

Приведем несколько утверждений для характеристических функций целочисленных случайных величин.

Пусть случайная величина X принимает целые значения m ($m=0,\ \pm1,\ \pm2,\ \ldots$) с вероятностями p_m, $\sum\limits_m p_m=1$. Тогда характеристическая функция имеет вид

$$f(t;X)=\sum_{m=-\infty}^{\infty}e^{itm}p_m, \tag{11.14}$$

причем ряд сходится абсолютно и равномерно на всей прямой.

Ряд вида (11.14) называется *тригонометрическим*. Мы будем рассматривать тригонометрические ряды и более общего вида

$$\sum_{k=-\infty}^{\infty}c_ke^{itk},$$

где c_k $(-\infty<k<\infty)$ — комплексные числа, такие, что $\sum\limits_{k=-\infty}^{\infty}|c_k|<\infty$.

Функции e^{itk}, $k=0,\pm1,\pm2,\ldots$, обладают следующими свойствами:

$$\begin{gathered}\frac{1}{2\pi}\int\limits_{-\pi}^{\pi}e^{itk}\,dt=\begin{cases}1, & k=0,\\ 0, & k\neq0,\end{cases}\\ \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}e^{int}e^{-imt}\,dt=\begin{cases}1, & n=m,\\ 0, & n\neq m.\end{cases}\end{gathered} \tag{11.15}$$

Замечание. На множестве комплекснозначных функций, определенных на $[-\pi,\pi]$, для которых

$$\int_{-\pi}^{\pi} |f(t)|^2\,dt < \infty,$$

можно ввести скалярное произведение

$$(f,g) = \int_{-\pi}^{\pi} f(t)\overline{g}(t)\,dt.$$

Тогда свойство (11.15) означает ортогональность функций e^{itn} и e^{itm} при $m \neq n$.

Пусть

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{itk}, \quad \sum_{k-\infty}^{\infty} |c_k| < \infty. \tag{11.16}$$

Умножим (11.16) на $\frac{1}{2\pi}e^{-imt}$ и проинтегрируем по отрезку $[-\pi,\pi]$. Тогда, используя свойства абсолютно сходящихся рядов, получим

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-imt}f(t)\,dt = \sum_{k=-\infty}^{\infty} c_k \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-imt}e^{ikt}\,dt = c_m.$$

Следовательно, зная характеристическую функцию, мы можем однозначно восстановить распределение целочисленной случайной величины. Тем самым доказана следующая лемма.

Лемма 11.7. *Если $f(t;x)$ — характеристическая функция случайной величины X, принимающей значение m с вероятностью p_m $(m = 0, \pm 1, \ldots)$, то*

$$p_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itm}f(t;X)\,dt.$$

Это равенство представляет собой формулу обращения для целочисленных случайных величин.

Задача 11.9. Предложите аналог формулы обращения для дискретных случайных величин.

Лемма 11.8. *Пусть выполнены условия леммы* 11.7. *Тогда для произвольных целых чисел* $m_1 \leqslant m_2$

$$\mathbf{P}\{m_1 \leqslant X \leqslant m_2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{m_1,m_2}(t) f(t;X)\,dt, \tag{11.17}$$

где

$$S_{m_1,m_2} = \frac{e^{-it(m_2+1/2)} - e^{-it(m_1-1/2)}}{-2i\sin\frac{t}{2}}.$$

Доказательство. Вычислим $\mathbf{P}\{m_1 \leqslant X \leqslant m_2\}$:

$$\mathbf{P}\{m_1 \leqslant X \leqslant m_2\} = \sum_{m=m_1}^{m_2} p_m = \sum_{m=m_1}^{m_2} \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itm} f(t;X)\,dt =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t;X)\cdot \sum_{m=m_1}^{m_2} e^{-itm}\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t;X)\cdot \frac{e^{-itm_1} - e^{-it(m_2+1)}}{1-e^{-it}}\,dt.$$

Домножив числитель и знаменатель дроби, стоящей под интегралом, на $e^{it/2}$, получим

$$\mathbf{P}\{m_1 \leqslant X \leqslant m_2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{-it(m_1-1/2)} - e^{-it(m_2+1/2)}}{e^{it/2} - e^{-it/2}} f(t;X)\,dt =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{-it(m_1-1/2)} - e^{-it(m_2+1/2)}}{2i\sin(t/2)} f(t;X)\,dt,$$

откуда и следует утверждение леммы 11.8. □

Формула (11.17) является аналогом формулы обращения для функций распределения, приведенной в теореме 11.3.

Следствие 11.1. *Для всех* $m_1 \leqslant m_2$

$$|S_{m_1,m_2}(t)| \leqslant \frac{1}{|\sin t/2|}.$$

Лемма 11.9. *Если*

$$f(t) = \sum_k c_k e^{itk}, \quad \sum_k |c_k| < \infty,$$
$$g(t) = \sum_k d_k e^{itk}, \quad \sum_k |d_k| < \infty,$$

то

$$\sup_m |c_m - d_m| \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t) - g(t)|\, dt, \tag{11.18}$$

$$\sup_{m_1 \leqslant m_2} \left| \sum_{m=m_1}^{m_2} c_m - \sum_{m=m_1}^{m_2} d_m \right| \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|f(t) - g(t)|}{|\sin(t/2)|}\, dt. \tag{11.19}$$

Эта лемма позволяет оценить близость функций распределения, если известны соответствующие характеристические функции. Из леммы следует (для целочисленных случайных величин), что из сходимости характеристических функций $f_n(t) \Rightarrow g(t)$ вытекает сходимость функций распределения $F_n(x) \to F(x)$ в каждой точке непрерывности $F(x)$.

Доказательство. Из леммы 11.7 следует, что для всех целых m

$$c_m - d_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx} (f(t) - g(t))\, dt,$$

и значит,

$$\sup_m |c_m - d_m| \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t) - g(t)|\, dt.$$

Аналогично, применяя лемму 11.8 и равенства из доказательства этой леммы, для $m_1 \leqslant m_2$ получаем

$$\left| \sum_{m=m_1}^{m_2} c_m - \sum_{m=m_1}^{m_2} d_m \right| \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|f(t) - g(t)|}{|\sin(t/2)|}\, dt.$$

□

Вернемся к схеме Пуассона: $A_1, A_2, \ldots$ — независимые случайные события, $\mathbf{P}(A_k) = p_k$, и

$$\mu_n(\omega) = \sum_{k=1}^{n} I_{A_k}(\omega).$$

Характеристическая функция случайной величины $\mu_n(\omega)$ равна

$$f(t;\mu_n) = \prod_{k=1}^{n} f(t; I_{A_k}) = \prod_{k=1}^{n} (1 - p_k + p_k e^{it}) = \prod_{k=1}^{n} \left(1 + p_k(e^{it} - 1)\right).$$

Обозначим $\lambda = \mathbf{E}\,\mu_n = \sum\limits_{k=1}^{n} p_k$ и рассмотрим случайную величину Y, имеющую распределение Пуассона с параметром λ:

$$\mathbf{P}\{Y = m\} = \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0, 1, 2, \ldots\ .$$

Нетрудно показать, что

$$f(t;Y) = e^{\lambda(e^{it}-1)}, \quad \mathbf{E}\,Y = \lambda, \quad \mathbf{D}\,Y = \lambda.$$

Обозначим

$$\Delta_m = |\mathbf{P}\{\mu_n = m\} - \mathbf{P}\{Y = m\}|.$$

Теорема 11.9. *Для схемы Бернулли если $p_k = p \leqslant \frac{1}{2}$ для всех k и $\lambda = np$, то*

$$\Delta_m \leqslant \frac{3}{2} \cdot \frac{\lambda^2}{n}. \tag{11.20}$$

Доказательство. В силу леммы 11.9 и неравенства леммы 11.3 имеем

$$\Delta_m \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \left(1 + p(e^{it} - 1)\right)^n - \left(e^{(\lambda/n)(e^{it}-1)}\right)^n \right| dt \leqslant$$

$$\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} n \left| 1 + p(e^{it} - 1) - e^{p(e^{it}-1)} \right| dt.$$

Обозначим $z = e^{it} - 1$, тогда необходимо уметь оценивать сверху величину $|e^{pz} - (1 + pz)|$. Учитывая, что $p \to 0$ при $n \to \infty$ и фиксированном $\lambda > 0$, надо уметь это делать в малой окрестности 0.

Лемма 11.10. *Если z — комплексное число, такое, что $|z| \leqslant \leqslant 1$, то*

$$|e^z - (1+z)| \leqslant (e-2)|z|^2.$$

Доказательство. Действительно,

$$|e^z - (1+z)| = \left|\frac{z^2}{2!} + \frac{z^3}{3!} + \ldots\right| \leqslant |z|^2 \left(\frac{1}{2!} + \frac{1}{3!} + \ldots\right) = |z|^2 \cdot (e-2).$$

□

Применим эту лемму к доказательству теоремы. Так как

$$z = e^{it} - 1 = e^{it/2}\left(e^{it/2} - e^{-it/2}\right) = 2i \sin\frac{t}{2}\, e^{it/2},$$

то $|z| \leqslant 2|\sin\frac{t}{2}|$, и при $p \leqslant \frac{1}{2}$ имеем $|pz| \leqslant 2p \leqslant 1$. Следовательно, можно применить лемму 11.10:

$$|e^{pz} - (1+pz)| \leqslant |pz|^2(e-2) \leqslant p^2 \cdot 4\sin^2\frac{t}{2} \cdot (e-2).$$

Таким образом, получаем

$$\Delta_m \leqslant 2(e-2)p^2 n \cdot \frac{1}{2\pi}\int\limits_{-\pi}^{\pi} 2\sin^2\frac{t}{2}\, dt = 2(e-2)p^2 n \leqslant \frac{3}{2}\cdot\frac{\lambda^2}{n}.$$

Теорема 11.9 доказана. □

Рассмотрим теперь общий случай схемы Пуассона.

Теорема 11.10. *Пусть $A_1, A_2, \ldots$ — независимые случайные события, такие, что $p_k = \mathbf{P}(A_k) \leqslant \frac{1}{2}$. Тогда*

$$\Delta_m = |\mathbf{P}\{\mu_n = m\} - \mathbf{P}\{Y = m\}| \leqslant \frac{3}{2}\cdot\sum_{k=1}^{n} p_k^2.$$

Доказательство. Из леммы 11.9 следует, что

$$\Delta_m \leqslant \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\left|\prod_{k=1}^{n}\left(1 + p_k(e^{it}-1)\right) - \prod_{k=1}^{n} e^{p_k(e^{it}-1)}\right| dt. \qquad (11.21)$$

Для оценки подынтегральной функции воспользуемся следующей леммой.

Лемма 11.11. *Пусть* α_k, β_k $(k = 1, 2, \ldots, n)$ — *комплексные числа, такие, что* $|\alpha_k| \leqslant 1$, $|\beta_k| \leqslant 1$. *Тогда*

$$\left|\prod_{k=1}^{n} \alpha_k - \prod_{k=1}^{n} \beta_k\right| \leqslant \sum_{k=1}^{n} |\alpha_k - \beta_k|.$$

Доказательство. Доказательство леммы проведем по индукции. При $n = 1$ утверждение леммы выполняется тривиальным образом. Допустив верность утверждения для некоторого n, для $n+1$ получаем

$$\begin{aligned}
&|\alpha_1 \cdots \alpha_n \alpha_{n+1} - \beta_1 \cdots \beta_n \beta_{n+1}| \leqslant \\
&\quad \leqslant |\alpha_1 \cdots \alpha_n \alpha_{n+1} - \alpha_1 \cdots \alpha_n \beta_{n+1}| + |\alpha_1 \cdots \alpha_n \beta_{n+1} - \beta_1 \cdots \beta_n \beta_{n+1}| \leqslant \\
&\quad \leqslant |\alpha_{n+1} - \beta_{n+1}| \cdot \left|\prod_{k=1}^{n} \alpha_k\right| + |\beta_{n+1}| \cdot \left|\prod_{k=1}^{n} \alpha_k - \prod_{k=1}^{n} \beta_k\right| \leqslant \\
&\quad \leqslant |\alpha_{n+1} - \beta_{n+1}| + \sum_{k=1}^{n} |\alpha_k - \beta_k| = \sum_{k=1}^{n+1} |\alpha_k - \beta_k|.
\end{aligned}$$

□

Воспользовавшись этой леммой, из (11.21) получаем

$$\Delta_m \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{n} \left|1 + p_k(e^{it} - 1) - e^{p_k(e^{it}-1)}\right| dt.$$

Если все $p_k \leqslant 1/2$, то снова можно применять лемму 11.10. Получим

$$\begin{aligned}
\Delta_m &\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^{n} p_k^2 \cdot 4 \sin^2 \frac{t}{2} (e-2)\, dt \leqslant \\
&\leqslant 2(e-2) \sum_{k=1}^{n} p_k^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} 2 \sin^2 \frac{t}{2}\, dt \leqslant \frac{3}{2} \cdot \sum_{k=1}^{n} p_k^2.
\end{aligned}$$

□

Замечание. Пуассоновское приближение имеет смысл применять, когда величина $\sum\limits_{k=1}^{n} p_k^2$ мала.

Приведем несколько примеров.

Пример 11.17. Пусть в схеме Бернулли $n = 1000$, $p = 0.01$. Тогда $\lambda = \sum\limits_{k=1}^{n} p_k = 10$. Вычисляя по формулам биномиального распределения с точностью до пятого знака после запятой, получим

$$\mathbf{P}\{\mu_n = 10\} = 0.12574,$$

тогда как пуассоновская вероятность равна 0.12511, $\Delta_{10} = 0.00063$.

Неравенство теоремы 11.9 для этой величины дает оценку

$$\Delta_{10} \leqslant \frac{3}{2}\frac{\lambda^2}{n} = 0.15.$$

Мы видим, что оценка довольно грубая.

Пример 11.18. Пусть $n = 100$, $p = 0.01$, $\lambda = 1$. Тогда $\mathbf{P}\{\mu_n = 3\} = 0.060999$, а пуассоновская аппроксимация дает 0.061313. Здесь расхождение точного и приближенного значений составляет $\Delta_3 = 0.000314$, а оценка, указанная в теореме, дает более грубую границу $\Delta_m \leqslant 0.015$.

Как видим из примеров, неравенство теоремы 11.9 дает хотя и грубую, но гарантированную оценку скорости сходимости, правильно показывающую, что точность аппроксимации увеличивается с уменьшением $\sum\limits_{k=1}^{n} p_k^2$.

11.8. Сходимость к равномерному распределению

Одно из часто использующихся в приложениях теории вероятностей непрерывных распределений — это равномерное распределение на некотором измеримом множестве $B \in \mathbb{R}^m$, имеющем конечную ненулевую меру Лебега ($0 < \mathrm{mes}(B) < \infty$). Плотность равномерного на множестве B распределения задается формулой

$$p(x_1, \ldots, x_m) = \begin{cases} \dfrac{1}{\mathrm{mes}(B)}, & \text{если} \quad (x_1, \ldots, x_m) \in B, \\ 0, & \text{если} \quad (x_1, \ldots, x_m) \notin B. \end{cases}$$

Модели с таким распределением вероятностей часто называют геометрическими[1]. В данном разделе мы будем рассматривать частный случай равномерного распределения на отрезке $[0;1]$.

Обозначим через U случайную величину с равномерным на $[0;1]$ распределением. Для этой случайной величины плотность $p_U(x)$, функция распределения $F_U(x)$ и характеристическая функция $f(t,U)$ равны соответственно

$$p_U(x) = \begin{cases} 1, & \text{если} \quad x \in [0;1], \\ 0, & \text{если} \quad x \notin [0;1], \end{cases} \qquad F_U(x) = \begin{cases} 0, & \text{если} \quad x \leqslant 0, \\ x, & \text{если} \quad 0 < x \leqslant 1, \\ 1, & \text{если} \quad x > 1, \end{cases}$$

$$f(t,U) = \frac{1}{it}(e^{it} - 1).$$

Следующая теорема устанавливает необходимое и достаточное условие сходимости последовательности случайных величин по распределению к случайной величине U.

Теорема 11.11. *Пусть $X_1, X_2, \ldots$ — последовательность случайных величин таких, что $\mathbf{P}\{0 \leqslant X_j \leqslant 1\} = 1$ для всех X_j. Тогда для выполнения при $n \to \infty$ предельного соотношения*

$$\mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \to \mathbf{P}\{\alpha \leqslant U \leqslant \beta\} \tag{11.22}$$

для любых $0 < \alpha < \beta < 1$ необходимо и достаточно, чтобы для всех целых $k \neq 0$ при $n \to \infty$

$$\mathbf{E}\, e^{2\pi i k X_n} \to 0. \tag{11.23}$$

Доказательство. Докажем сначала необходимость выполнения условия (11.23). Выполнение предельного соотношения (11.22) равносильно слабой сходимости $f(t,X_n) \Rightarrow f(t,U)$ сответствующих характеристических функций. Учитывая, что $f(2\pi k,U) = 0$ при всех целых $k \neq 0$, условие (11.23) означает сходимость характеристических функций

$$f(2\pi k, X_n) \to f(2\pi k, U)$$

при $n \to \infty$ в точках определенного вида, когда $t = 2\pi k$ для любого целого $k \neq 0$. Следовательно, условие (11.23) является необходимым для сходимости распределений случайных величин X_n к равномерному на $[0;1]$ распределению.

[1] Прочитать о геометрических вероятностях можно в [6]. В этом учебнике разобрано много задач и приводятся примеры использования геометрических вероятностей.

Перейдем к доказательству достаточности условия (11.23). Из условия (11.23) теоремы следует, что для любого тригонометрического многочлена $T(x)$ вида

$$T(x) = \sum_{k=-k_0}^{k_0} a_k e^{2\pi ikx}, \tag{11.24}$$

где $a_{-k_0}, \ldots, a_{k_0}$ — произвольные комплексные числа, при $n \to \infty$ справедливо предельное соотношение

$$\mathbf{E}\,T(X_n) = \sum_{k=-k_0}^{k_0} a_k\,\mathbf{E}\,e^{2\pi ikX_n} \to \sum_{k=-k_0}^{k_0} a_k\,\mathbf{E}\,e^{2\pi ikU} = \mathbf{E}\,T(U). \tag{11.25}$$

Перепишем теперь утверждение теоремы в следующем виде:

$$\mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} = \mathbf{E}\,I_{[\alpha;\beta]}(X_n) \to \mathbf{E}\,I_{[\alpha;\beta]}(U) = \mathbf{P}\{\alpha \leqslant U \leqslant \beta\}$$

и сравним с (11.25). Понятно, что мы докажем утверждение теоремы, если сможем приблизить индикатор $I_{[\alpha;\beta]}(x)$ тригонометрическими многочленами.

Поступим так же, как и при доказательстве центральной предельной теоремы. Сначала приблизим функцию $I_{[\alpha;\beta]}(x)$ непрерывными функциями. Поскольку $0 < \alpha < \beta < 1$, то для достаточно малых $\delta > 0$ будут выполняться неравенства $0 < \alpha - \delta < \beta + \delta < 1$. Следовательно, существуют непрерывные, определенные на отрезке $[0;1]$ функции $0 \leqslant H_\delta^-(x) \leqslant 1$ и $0 \leqslant H_\delta^+(x) \leqslant 1$, обладающие свойствами

$$\begin{aligned}
&H_\delta^-(x) = 1, \quad \text{если} \quad x \in [\alpha+\delta;\beta-\delta],\\
&H_\delta^-(x) = 0, \quad \text{если} \quad x \notin [\alpha;\beta],\\
&H_\delta^+(x) = 1, \quad \text{если} \quad x \in [\alpha;\beta],\\
&H_\delta^+(x) = 0, \quad \text{если} \quad x \notin [\alpha-\delta;\beta+\delta],\\
&H_\delta^-(x) \leqslant I_{[\alpha;\beta]}(x) \leqslant H_\delta^+(x),
\end{aligned}$$

причем последнее неравенство выполняется для всех $0 \leqslant x \leqslant 1$. Но тогда и для всех случайных величин $0 \leqslant X_n \leqslant 1$ выполняется аналогичное неравенство

$$H_\delta^-(X_n) \leqslant I_{[\alpha;\beta]}(X_n) \leqslant H_\delta^+(X_n). \tag{11.26}$$

Воспользуемся теперь второй теоремой Вейерштрасса о равномерном приближении непрерывной периодической функции тригонометрическими многочленами[1]. Приведем формулировку этой теоремы в удобном для наших целей варианте.

Теорема 11.12. *Пусть $H(x)$ — непрерывная периодическая функция с периодом, равным* 1. *Тогда для любого $\varepsilon > 0$ существует тригонометрический многочлен $T_H(x)$ вида* (11.24) *такой, что*

$$\sup_{-\infty<x<+\infty} |H(x) - T_H(x)| < \varepsilon.$$

Замечание. Отметим, что многочлен $T_H(x)$ зависит от ε.

Так как функции $H_\delta^-(x)$ и $H_\delta^+(x)$ непрерывны на отрезке $[0;1]$ по построению и для них также выполнены равенства $H_\delta^-(0) = H_\delta^-(1) = 0$ и $H_\delta^+(0) = H_\delta^+(1) = 0$, то их можно периодически с периодом 1 продолжить на всю прямую. Выберем $\varepsilon > 0$. Согласно теореме Вейерштрасса, для функций $H_\delta^-(x)$ и $H_\delta^+(x)$ существуют тригонометрические многочлены $T_\varepsilon^-(x)$ и $T_\varepsilon^+(x)$ вида (11.24), для которых при всех x

$$|H_\delta^-(x) - T_\varepsilon^-(x)| < \varepsilon, \quad |H_\delta^+(x) - T_\varepsilon^+(x)| < \varepsilon. \tag{11.27}$$

В обозначениях данных тригонометрических многочленов верхний индекс указывает, для какой из двух функций выбран многочлен, а нижний индекс — для какого ε.

Тогда для всех $0 \leqslant x \leqslant 1$ можно выписать цепочку неравенств

$$T_\varepsilon^-(x) - \varepsilon < H_\delta^-(x) \leqslant I_{[\alpha;\beta]}(x) \leqslant H_\delta^+(x) < T_\varepsilon^+(x) + \varepsilon.$$

Следовательно, для случайных величин $0 \leqslant X_n \leqslant 1$ также выполнены неравенства

$$T_\varepsilon^-(X_n) - \varepsilon < H_\delta^-(X_n) \leqslant I_{[\alpha;\beta]}(X_n) \leqslant H_\delta^+(X_n) < T_\varepsilon^+(X_n) + \varepsilon. \tag{11.28}$$

Переходя в неравенствах (11.28) к математическим ожиданиям, получим

$$\mathbf{E}\, T_\varepsilon^-(X_n) - \varepsilon \leqslant \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \leqslant \mathbf{E}\, T_\varepsilon^+(X_n) + \varepsilon,$$

откуда, устремив $n \to \infty$, имеем в силу (11.25)

$$\mathbf{E}\, T_\varepsilon^-(U) - \varepsilon \leqslant \liminf_{n\to\infty} \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \leqslant \\ \leqslant \limsup_{n\to\infty} \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \leqslant \mathbf{E}\, T_\varepsilon^+(U) + \varepsilon. \tag{11.29}$$

[1] См., например, [33, т. 3, гл. XX, § 1].

Учитывая, что для функций $H_\delta^-(x)$ и $H_\delta^+(x)$ при всех $0 \leqslant x \leqslant 1$ также выполняются неравенства

$$I_{[\alpha+\delta;\beta-\delta]}(x) - \varepsilon \leqslant H_\delta^-(x) - \varepsilon \leqslant T_\varepsilon^-(x),$$
$$T_\varepsilon^+(x) \leqslant H_\delta^+(x) + \varepsilon \leqslant I_{[\alpha-\delta;\beta+\delta]}(x) + \varepsilon,$$

аналогичным образом можно получить, что

$$\mathbf{E}\, I_{[\alpha+\delta;\beta-\delta]}(U) - \varepsilon \leqslant \mathbf{E}\, T_\varepsilon^-(U) \tag{11.30}$$

и

$$\mathbf{E}\, T_\varepsilon^+(U) \leqslant \mathbf{E}\, I_{[\alpha-\delta;\beta+\delta]}(U) + \varepsilon. \tag{11.31}$$

Поскольку

$$\mathbf{E}\, I_{[\alpha+\delta;\beta-\delta]}(U) = \beta - \alpha - 2\delta, \quad \mathbf{E}\, I_{[\alpha-\delta;\beta+\delta]}(U) = \beta - \alpha + 2\delta,$$

из неравенств (11.29), учитывая (11.30) и (11.31), получим

$$\beta - \alpha - 2\delta - 2\varepsilon \leqslant \liminf_{n\to\infty} \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \leqslant$$
$$\leqslant \limsup_{n\to\infty} \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} \leqslant \beta - \alpha - 2\delta - 2\varepsilon.$$

Переходя в последнем неравенстве к пределу при $\delta \to 0$, $\varepsilon \to 0$, окончательно получаем, что

$$\lim_{n\to\infty} \mathbf{P}\{\alpha \leqslant X_n \leqslant \beta\} = \beta - \alpha.$$

□

Рассмотрим теперь пример сходимости к равномерному на отрезке $[0, 1]$ распределению.

Пример 11.19. Суммирование случайных величин по модулю 1. Пусть $Y_1, Y_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин. Введем обозначение для дробной части суммы первых n случайных величин:

$$X_n = \{Y_1 + \ldots + Y_n\}.$$

Применим основную теорему данного раздела для исследования условий сходимости к равномерному на отрезке $[0, 1]$ распределению.

Поскольку для произвольного числа x справедливо равенство $\{x\} = x - [x]$, в котором $[x]$ обозначает целую часть x, то при всех целых k выполнены равенства

$$e^{2\pi ikX_n} = e^{2\pi ik(Y_1+\ldots+Y_n)-2\pi ik[Y_1+\ldots+Y_n]} = e^{2\pi ik(Y_1+\ldots+Y_n)} = \prod_{j=1}^{n} e^{2\pi ikY_j}.$$

Так как случайные величины $Y_1, Y_2, \ldots$ независимы и одинаково распределены, то для математических ожиданий мы получим, что

$$\mathbf{E}\, e^{2\pi ikX_n} = \prod_{j=1}^{n} \mathbf{E}\, e^{2\pi ikY_j} = \left(\mathbf{E}\, e^{2\pi ikY_1}\right)^n.$$

Таким образом, условие

$$\left|\mathbf{E}\, e^{2\pi ikY_1}\right| < 1 \quad \text{при всех} \quad k \neq 0 \tag{11.32}$$

является необходимым и достаточным для того, чтобы

$$\{Y_1 + \ldots + Y_n\} \xrightarrow{d} U.$$

Разберем теперь ситуацию, когда найдется отличное от 0 целое число k_0, для которого

$$\left|\mathbf{E}\, e^{2\pi ik_0Y_1}\right| = 1.$$

В этом случае существует такое действительное число θ, что выполняется равенство

$$\mathbf{E}\, e^{2\pi ik_0Y_1} = e^{i\theta},$$

или, что то же самое,

$$\mathbf{E}\left(e^{2\pi ik_0Y_1 - i\theta}\right) = 1.$$

Последнее равенство равносильно тому, что

$$\mathbf{E}\cos(2\pi k_0Y_1 - \theta) = 1,$$

из которого следует, что с вероятностью 1

$$2\pi k_0Y_1 - \theta = 2\pi m, \quad \text{где} \quad m = 0, \pm1, \pm2, \ldots\ .$$

Следовательно, с вероятностью 1

$$Y_1 = \frac{m}{k_0} + \frac{\theta}{2\pi k_0}, \quad m = 0, \pm1, \pm2, \ldots\ . \tag{11.33}$$

Суммирование по модулю 1 для наглядности можно интерпретировать следующим образом.

Свернем числовую ось OX в окружность длины 1[1]. От точки $x = 0$ в нужном направлении, соответствующем знаку случайной величины Y_1, отложим Y_1. Получим точку X_1. Тогда $\{X_1\}$ — длина дуги OX_1, измеренная в положительном направлении. От полученной точки в нужном направлении отложим Y_2 и получим точку X_2 — и так далее. Если выполнено условие (11.32), то распределение полученных на окружности точек $X_1, X_2, \ldots$ «почти» равномерное, то есть для любых $0 < \alpha < \beta < 1$ выполняется при $n \to \infty$

$$\mathbf{P}\{\alpha \leqslant X \leqslant \beta\} \to \beta - \alpha.$$

Если же не выполняется (11.32), то выполнено условие (11.33), и все точки X_n попадают в этом случае в вершины правильного k_0-угольника. Сходимости к равномерному распределению в этом случае не будет.

[1] Подобным образом получают тригонометрический круг, сворачивая числовую ось в окружность радиуса 1.

Глава 12

ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. ОСНОВНЫЕ ПОНЯТИЯ

В качестве вступительного слова к новой части нашего курса — математической статистике — приведем цитату из статьи академика А. Н. Колмогорова в Математической энциклопедии: «Математическая статистика — раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов. При этом статистическими данными называют сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками»[1].

Ставший классическим учебник Г. Крамера «Методы математической статистики» содержит множество различных примеров статистических данных. Упомянем несколько из них:

- данные о возрасте родителей для проверки гипотезы о влиянии возраста родителей на пол ребенка[2];
- данные об уровнях воды в озере Вэнерн за 1807–1930 гг., которые использовались при проектировании плотины[3], и т. д.

В «Таблицах математической статистики» Л. Н. Большева и Н. В. Смирнова [3] половину объема всей книги составляет текст, в котором в рецептурной форме даются рекомендации по решению прикладных задач.

Теория вероятностей дает математический аппарат для решения задач математической статистики.

Результаты наблюдений в математической статистике трактуются как значения независимых и одинаково распределенных случайных величин $X_1, X_2, \ldots, X_n$. В некоторых случаях это могут быть и случайные векторы, и зависимые случайные величины с некоторым совместным распределением вероятностей.

Как правило, мы будем отдельно разбирать следующие случаи:

---

[1] [23, т. 3, с. 576–581].

[2] [17, с. 496].

[3] [17, с. 502].

- распределение случайных величин $X_1, X_2, \ldots, X_n$ — абсолютно непрерывное, имеет плотность распределения вероятностей;
- распределение случайных величин — дискретное.

Часто доказательство теорем мы будем проводить на примере одного из этих случаев, хотя утверждения теорем могут быть верны и в более общих ситуациях.

Что касается числа наблюдений n, то различают следующие случаи:

- большие выборки, когда число наблюдений n велико (порядка нескольких сотен), — тогда используют асимптотические теоремы теории вероятностей;
- умеренные выборки, когда $30 \leqslant n \leqslant 100$ (это наиболее сложный случай);
- малые выборки при $n \leqslant 30$ (в этом случае статистические выводы строятся на точных распределениях).

В математической статистике независимые одинаково распределенные случайные величины $X_1, X_2, \ldots, X_n$ с общей функцией распределения $F(x)$ называют случайной *выборкой* объема n из генеральной совокупности с распределением $F(x)$.

Приведем также классификацию основных задач математической статистики по характеру статистических выводов.

1. Проверка статистических гипотез. Во многих задачах формулируется некоторая гипотеза (например, гипотеза о влиянии возраста родителей на пол ребенка), которая может быть принята или отвергнута на основании статистических данных.

2. Оценивание неизвестных параметров распределения. Различают оценки точечные и интервальные.

3. Определение видов зависимостей между различными объектами.

В математической статистике вместо вероятностного пространства $(\Omega, \mathcal{A}, \mathbf{P})$ рассматривают семейства вероятностных пространств

$$\{(\Omega, \mathcal{A}, \mathbf{P}),\ \mathbf{P} \in \mathbb{P}\},$$

где $\mathbb{P}$ — семейство вероятностных распределений на $(\Omega, \mathcal{A})$. Это семейство распределений может быть параметрическим, зависящим от некоторого параметра θ. В этом случае мы будем писать

$$\mathbb{P} = \{\mathbf{P}_\theta,\ \theta \in \Theta\},$$

где Θ — множество допустимых значений параметра. Так, например, семейство нормальных распределений — это семейство распределений с двумя параметрами a и σ, $a \in (-\infty, +\infty)$, $\sigma \in \in (0, +\infty)$. Впрочем, параметризовать семейство можно было и по-другому, да и часто, когда значение одного из параметров известно, а другого — нет, это семейство рассматривается как однопараметрическое.

12.1. Сходимость по вероятности

Основным видом сходимости в математической статистике является сходимость по вероятности. Напомним определение.

Определение 12.1. Последовательность случайных величин $Y_1, Y_2, \ldots$ сходится к случайной величине Y по вероятности ($Y_n \xrightarrow{\mathbf{P}} Y$), если для любого $\varepsilon > 0$

$$\mathbf{P}\{\omega\colon |Y_n(\omega) - Y(\omega)| \geqslant \varepsilon\} \to 0 \qquad (12.1)$$

при $n \to \infty$.

Заметим, что выполнение (12.1) равносильно тому, что

$$\mathbf{P}\{\omega\colon |Y_n(\omega) - Y(\omega)| < \varepsilon\} \to 1. \qquad (12.2)$$

Определение 12.2. Последовательность случайных величин $W_1, W_2, \ldots$ называется *ограниченной по вероятности*, если для любого $\varepsilon > 0$ существует $K = K(\varepsilon)$ такое, что при всех n

$$\mathbf{P}\{\omega\colon |W_n(\omega)| \leqslant K\} \geqslant 1 - \varepsilon. \qquad (12.3)$$

В дальнейшем для краткости последовательность каких-либо чисел (случайных величин) $a_1, a_2, \ldots$ будем обозначать как $\{a_n\}$. В качестве самостоятельных упражнений докажите следующие свойства сходимости по вероятности.

Задача 12.1. Если последовательность случайных величин $\{a_n\}$ ограничена по вероятности, а $b_n \xrightarrow{\mathbf{P}} 0$, то $a_n b_n \xrightarrow{\mathbf{P}} 0$.

Задача 12.2. Если последовательности случайных величин $\{a_n\}$, $\{b_n\}$ ограничены по вероятности, то последовательность $\{c_n\}$ ($c_n = a_n b_n$) также является ограниченной по вероятности.

Сформулируем несколько утверждений, касающихся этого вида сходимости, которые нам понадобятся в дальнейшем.

Пусть $F_U(x)$ обозначает функцию распределения случайной величины U, а соотношение $F_{Y_n(x)} \Rightarrow F_Y(x)$ — слабую сходимость функций распределения.

Теорема 12.1. *Если* $Y_n \xrightarrow{\mathbf{P}} Y$, *то* $F_{Y_n}(x) \Rightarrow F_Y(x)$, *то есть из сходимости по вероятности следует слабая сходимость функций распределения.*

Теорема 12.2. *Пусть* $F_{W_n}(x) \Rightarrow F_W(x)$, $Y_n \xrightarrow{\mathbf{P}} 0$. *Тогда*

$$F_{W_n+Y_n}(x) \Rightarrow F_W(x).$$

Докажите эти утверждения самостоятельно.

Теорема 12.3. *Пусть* $h(x)$ *— всюду непрерывная функция, а* $Y_n \xrightarrow{\mathbf{P}} 0$. *Тогда* $h(a+Y_n) \xrightarrow{\mathbf{P}} h(a)$ *для всех* a.

Доказательство. В соответствии с определением сходимости по вероятности необходимо показать, что для любого $\varepsilon > 0$ при $n \to \infty$ выполняется

$$\mathbf{P}\{\omega\colon |h(a+Y_n) - h(a)| \geqslant \varepsilon\} \to 0. \tag{12.4}$$

Из непрерывности функции $h(x)$ в точке $x = a$ следует, что $\forall\, \varepsilon > 0$ существует $\lambda = \lambda(\varepsilon)$ такое, что если $|x - a| < \lambda$, то $|h(x) - h(a)| < \varepsilon$.

Представим вероятность в (12.4) следующим образом:

$$\mathbf{P}\{\omega\colon |h(a+Y_n) - h(a)| \geqslant \varepsilon\} = \mathbf{P}\{\omega\colon |h(a+Y_n) - h(a)| \geqslant \varepsilon,\ |Y_n| < \lambda\} + \\ + \mathbf{P}\{\omega\colon |h(a+Y_n) - h(a)| \geqslant \varepsilon,\ |Y_n| \geqslant \lambda\} = \mathbf{P}(A_n(\varepsilon,\lambda)) + \mathbf{P}(B_n(\varepsilon,\lambda)).$$

Поскольку $Y_n \xrightarrow{\mathbf{P}} 0$, то $\forall\, \lambda$, а значит, и для выбранного $\lambda = \lambda(\varepsilon)$ выполняется

$$\mathbf{P}\{\omega\colon |Y_n| \geqslant \lambda\} \to 0$$

при $n \to \infty$. Следовательно,

$$\mathbf{P}(B_n(\varepsilon,\lambda)) \leqslant \mathbf{P}\{|Y_n| \geqslant \lambda\} \to 0.$$

Рассмотрим другое слагаемое. Поскольку если $\omega \in A_n(\varepsilon,\lambda)$, то в этом случае $|Y_n(\omega)| < \lambda$, и значит, $|(Y_n + a) - a| < \lambda$. Но тогда в соответствии с выбором λ

$$|h(a+Y_n) - h(a)| < \varepsilon.$$

Следовательно, $\mathbf{P}(A_n(\varepsilon,\lambda)) = 0$, что и завершает доказательство теоремы. □

Теорема 12.4. *Если $Y_n \xrightarrow{\mathbf{P}} a$ при $n \to \infty$, то последовательность случайных величин $\{Y_n\}$ ограничена по вероятности.*

Доказательство. Пусть $Y_n \xrightarrow{\mathbf{P}} a$, тогда из (12.2) следует, что $\forall\, \varepsilon > 0$, $\forall\, \delta > 0$ $\exists\, N = N(\varepsilon, \delta)$ такое, что $\forall\, n > N$

$$\mathbf{P}\{\omega\colon\ |Y_n - a| \leqslant \delta\} > 1 - \frac{1}{2}\varepsilon.$$

Выберем $M^* = \max(|a+\delta|, |a-\delta|)$, тогда для всех $n > N$

$$\mathbf{P}\{\omega\colon\ |Y_n| \leqslant M^*\} > 1 - \frac{1}{2}\varepsilon.$$

Задача 12.3. Покажите, что для произвольной случайной величины X и любого $\varepsilon > 0$ существует $M_X = M_X(\varepsilon)$ такое, что

$$\mathbf{P}\{|X| \leqslant M_X\} > 1 - \varepsilon.$$

Пользуясь этим фактом, получим, что существуют константы $M_1, M_2, \ldots, M_N$ такие, что

$$\mathbf{P}\{|Y_1| \leqslant M_1\} > 1 - \frac{\varepsilon}{2N}, \ldots, \mathbf{P}\{|Y_N| \leqslant M_N\} > 1 - \frac{\varepsilon}{2N}.$$

Но тогда для $M_0 = \max(M_1, \ldots, M_N)$ выполняется

$$\mathbf{P}\{|Y_1| \leqslant M_0, \ldots, |Y_n| \leqslant M_0\} = 1 - \mathbf{P}\left\{\bigcup_{i=1}^{N}\{|Y_i| > M_0\}\right\} \geqslant$$

$$\geqslant 1 - \sum_{i=1}^{N}\mathbf{P}\{|Y_i| > M_0\} > 1 - \frac{\varepsilon}{2N}\cdot N = 1 - \frac{\varepsilon}{2}.$$

Следовательно, если выбрать $M = \max(M^*, M_0)$, то для всех n

$$\mathbf{P}\{|Y_n| \leqslant M\} > 1 - \varepsilon,$$

то есть последовательность $\{Y_n\}$ ограничена по вероятности. □

12.2. Асимптотическая нормальность

Определение 12.3. Последовательность случайных величин Y_n *асимптотически нормальна*[1] с математическим ожиданием

[1] В настоящей книге в качестве параметров асимптотической нормальности указываются математическое ожидание и стандартное отклонение (см. также [17; 26]). Не следует путать с параметрами нормального рас-

A_n и стандартным отклонением B_n, если для всех x при $n \to \infty$

$$\mathbf{P}\left\{\frac{Y_n - A_n}{B_n} < x\right\} \to \Phi(x). \qquad (12.5)$$

В определении предполагается, что $-\infty < A_n < \infty$, $B_n > 0$.

Пример 12.1. Пусть μ_n — число успехов в n испытаниях Бернулли с вероятностью успеха в одном испытании p, $(0 < p < 1)$. Из теоремы Муавра–Лапласа следует, что μ_n асимптотически нормальна с математическим ожиданием np и стандартным отклонением $\sqrt{npq}$, а относительная частота успеха $\frac{\mu_n}{n}$ асимптотически нормальна с параметрами p и $\sqrt{\frac{pq}{n}}$ соответственно.

Задача 12.4. Докажите, что из асимптотической нормальности последовательности случайных величин Y_n с математическим ожиданием A_n и стандартным отклонением B_n вытекает асимптотическая нормальность $U_n = \alpha Y_n + \beta$ с параметрами $\alpha A_n + \beta$ и αB_n.

Теорема 12.5. *Пусть:*

1) *последовательность случайных величин Y_n асимптотически нормальна с параметрами (a, b_n), причем $b_n \to 0$ при $n \to \infty$;*

2) *функция $g(x)$ дважды непрерывно дифференцируема на всей прямой и $g'(a) \neq 0$.*

Тогда последовательность случайных величин $W_n = g(Y_n)$ асимптотически нормальна с параметрами $(g(a); b_n|g'(a)|)$.

Доказательство. В общем случае доказательство теоремы довольно громоздко[1]. Поэтому мы разберем простой случай, когда случайные величины Y_n нормально распределены.

Пусть $Z_n \sim N(0,1)$, $Y_n = a + b_n Z_n$, $b_n > 0$, $b_n \to 0$. Покажем в этих дополнительных предположениях справедливость теоремы 12.5.

Для дважды непрерывно дифференцируемой функции $g(x)$ можно написать разложение по формуле Тейлора в окрестности

пределения в общепринятом обозначении $N(a, \sigma^2)$, где вторым параметром указана дисперсия.

[1] Доказательство в общем случае можно посмотреть в учебнике [26, с. 254–256].

точки a:

$$g(Y_n) = g(a) + b_n Z_n g'(a) + \frac{b_n^2 Z_n^2}{2} \cdot g''(a + \theta b_n Z_n),$$

где $|\theta| \leqslant 1$.

Для определенности будем считать, что $g'(a) > 0$. Тогда

$$\frac{g(Y_n) - g(a)}{b_n g'(a)} = Z_n + \frac{b_n Z_n^2}{2} \cdot \frac{1}{g'(a)} \cdot g''(a + \theta b_n Z_n). \tag{12.6}$$

Для того чтобы доказать утверждение теоремы, достаточно показать, что последнее слагаемое в (12.6) сходится по вероятности к 0.

Так как $F_{Z_n}(x) = \Phi(x)$ для всех n, то для любого $\varepsilon > 0$ $\exists M = M(\varepsilon)$ такое, что

$$\mathbf{P}\{|Z_n| \leqslant M\} > 1 - \varepsilon$$

для всех n одновременно. Значит, последовательность $\{Z_n\}$ ограничена по вероятности. Но тогда (утверждение задачи 12.1) $b_n Z_n \xrightarrow{\mathbf{P}} 0$ и, стало быть,

$$\frac{b_n Z_n^2}{2g'(a)} \xrightarrow{\mathbf{P}} 0.$$

Кроме того, из непрерывности второй производной функции $g(x)$ по теореме 12.3 следует сходимость

$$g''(a + \theta b_n Z_n) \xrightarrow{\mathbf{P}} g''(a).$$

Следовательно, по теореме 12.4 последовательность $g''(a + \theta b_n Z_n)$ ограничена по вероятности. Учитывая все, что было сказано выше, получим, что второе слагаемое в (12.6) сходится по вероятности к 0. Но тогда в силу утверждения теоремы 12.2

$$\frac{g(Y_n) - g(a)}{b_n g'(a)} \xrightarrow{d} Z,$$

или, что то же самое, $g(Y_n)$ асимптотически нормальна с параметрами $(g(a), b_n g'(a))$, что и требовалось доказать. □

12.3. Некоторые важные преобразования случайных величин

Пусть U — случайная величина, равномерно распределенная на $[0,1]$ с плотностью

$$p_U(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & x \notin [0,1], \end{cases}$$

и функцией распределения

$$F_U(x) = \begin{cases} 0, & x \leqslant 0, \\ x, & 0 < x \leqslant 1, \\ 1, & x > 1. \end{cases} \tag{12.7}$$

Лемма 12.1. *Пусть X — случайная величина с непрерывной функцией распределения $F(x)$, строго возрастающей в окрестности любой точки x_0, где $0 < F(x_0) < 1$, $F^{-1}(x)$ — функция, обратная к $F(x)$. Тогда:*

1) *случайная величина $Y = F(X)$ имеет равномерное распределение на $[0,1]$,*

2) *случайная величина $V = F^{-1}(U)$ имеет функцию распределения $F(x)$.*

Доказательство. 1. Поскольку $0 \leqslant F(x) \leqslant 1$, то $F_Y(x) = 0$ при $x \leqslant 0$ и $F_Y(x) = 1$ при $x > 1$. Для $0 < x < 1$ имеем

$$F_Y(x) = \mathbf{P}\{Y < x\} = \mathbf{P}\{F(X) < x\} = \mathbf{P}\{X < F^{-1}(x)\} = \\ = F\left(F^{-1}(x)\right) = x,$$

то есть функция распределения случайной величины Y совпадает с функцией распределения равномерного распределения на $[0,1]$.

2. Функция распределения случайной величины V равна

$$F_V(x) = \mathbf{P}\{V < x\} = \mathbf{P}\{F^{-1}(U) < x\} = \mathbf{P}\{U < F(x)\} = F(x),$$

что и требовалось доказать. □

Утверждение 2 леммы используется при моделировании случайных величин с заданной непрерывной функцией распределения $F(x)$. Для этого надо уметь моделировать равномерно распределенные случайные величины и к полученным значениям применить обратную функцию $F^{-1}(x)$.

Задача 12.5. Докажите, что утверждение 1 леммы верно для всякой непрерывной функции распределения $F(x)$.

Определение 12.4. Пусть $0 < \gamma < 1$ и $F(x)$ — непрерывная функция распределения. Решение уравнения $F(x_\gamma) = \gamma$ называется *квантилью уровня* γ распределения $F(x)$.

Замечание. Если уравнение $F(x_\gamma) = \gamma$ имеет несколько решений, то квантилью уровня γ называют наименьшее решение данного уравнения.

Если $\gamma = 1/2$, то $x_{1/2}$ называется *медианой* распределения, для нее

$$\mathbf{P}\{X \geqslant x_{1/2}\} = \mathbf{P}\{X \leqslant x_{1/2}\} = 1/2. \tag{12.8}$$

Если $\gamma = 1/4,\ 2/4,\ 3/4$, то $x_{1/4}$, $x_{2/4}$, $x_{3/4}$ называются *квартилями* распределения. Если $\gamma = k/10$ $(k = 1, 2, \ldots, 9)$, то $x_{k/10}$ называют *децилями* распределения.

Медиана является центром распределения в смысле (12.8), а квартили и децили служат характеристиками разброса относительно медианы; так, например,

$$\mathbf{P}\{x_{1/4} \leqslant X < x_{3/4}\} = 1/2.$$

Эти характеристики широко используются в экономико-статистических методах изучения свойств различных совокупностей. Скажем, если рассматривается распределение населения по уровню годового дохода, то нижняя дециль $x_{1/10}$ показывает, какая часть населения имеет наиболее низкий доход (доля бедных людей), а верхняя квантиль $x_{9/10}$ отделяет наиболее богатых. Сравнивая девятую и первую децили, измеряют соотношение уровней доходов 10% наиболее обеспеченного и 10% наименее обеспеченного населения.

На рис. 9 изображены распределения по доходам в двух различных совокупностях. В первой из них больше половины населения имеет крайне низкий доход, очень мало богатых. Во второй картина не такая удручающая: гораздо меньше доля очень бедных и больше богатых.

К наиболее часто используемым квантилям относятся квантили уровней

$$\begin{aligned} \gamma = 0.0001;\quad 0.005;\quad 0.01;\quad 0.025;\quad 0.05;\\ 0.9999;\quad 0.995;\quad 0.99;\quad 0.975;\quad 0.95. \end{aligned}$$

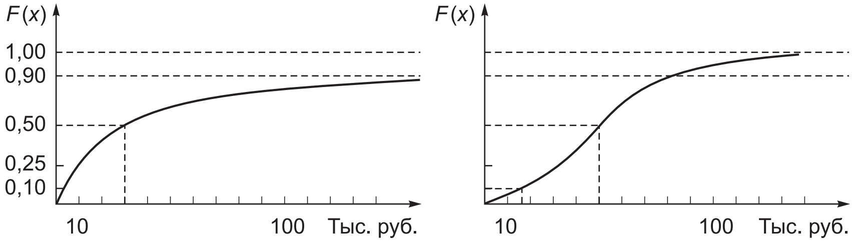

Рис. 9. Распределение населения по уровню доходов

Таблицы квантилей основных распределений можно найти в книге Л. Н. Большева и Н.В.Смирнова [3].

Лемма 12.2. *Если функция распределения $F(x)$ непрерывна во всех точках x, то она равномерно непрерывна на всей числовой прямой.*

Задача 12.6. Проведите самостоятельно доказательство этого утверждения.

Лемма 12.3. *Пусть случайные величины $X_1, X_2, \ldots, X_n$ — независимы, одинаково распределены с непрерывной функцией распределения $F(x)$. Тогда вероятность того, что какие-либо две случайные величины примут одинаковые значения, равна 0.*

Доказательство. Обозначим $B = \{\omega\colon \exists i \neq j$ такие, что $X_i(\omega) = X_j(\omega)\}$. Тогда

$$B \subseteq \bigcup_{1 \leqslant i<j \leqslant n} \{\omega : X_i(\omega) = X_j(\omega)\},$$

и, стало быть,

$$\mathbf{P}(B) \leqslant \sum_{1 \leqslant i<j \leqslant n} \mathbf{P}\{\omega : X_i(\omega) = X_j(\omega)\}.$$

Покажем, что каждое слагаемое в этой сумме равно 0. Пусть $h > 0$. Тогда

$$\{\omega\colon X_i(\omega) = X_j(\omega)\} \subseteq$$
$$\subseteq \bigcup_{k=-\infty}^{\infty} \{\omega\colon kh \leqslant X_i(\omega) < (k+1)h,\ kh \leqslant X_j(\omega) < (k+1)h\}.$$

Следовательно,

$$\mathbf{P}\{X_i = X_j\} \leqslant \sum_{k=-\infty}^{\infty} \mathbf{P}\{kh \leqslant X_i < (k+1)h\} \cdot \mathbf{P}\{kh \leqslant X_j < (k+1)h\} =$$

$$= \sum_{k=-\infty}^{\infty} (F(h(k+1)) - F(hk))^2 \leqslant \sup_k (F((k+1)h) - F(kh)) \cdot 1.$$

В силу леммы 12.2, выбирая должным образом h, мы можем сделать эту величину сколь угодно малой: $\forall \varepsilon > 0\ \exists h(\varepsilon)$ такое, что

$$\mathbf{P}\{X_i \neq X_j\} < \varepsilon.$$

Устремив $\varepsilon \to 0$, получим нужное утверждение. □

Из леммы 12.3 следует, что если мы наблюдаем независимые случайные величины $X_1, X_2, \ldots, X_n$ с одинаковой *непрерывной* функцией распределения $F(x)$, то с вероятностью 1 все наблюдения различны и их можно расположить в порядке возрастания. Полученную упорядоченную выборку будем обозначать

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}$$

и называть *вариационным рядом*, члены вариационного ряда $X_{(i)}$ — *вариантами* или *порядковыми статистиками*[1]. Заметим, что $X_{(1)} = \min\limits_{1 \leqslant i \leqslant n} X_i$, $X_{(n)} = \max\limits_{1 \leqslant i \leqslant n} X_{(i)}$.

Из леммы 12.1 вытекает, что $F(X_1)$, $F(X_2)$, …, $F(X_n)$ — независимые и равномерно распределенные на $[0, 1]$ случайные величины, причем с вероятностью 1 выполняется

$$F(X_{(1)}) < F(X_{(2)}) < \ldots F(X_{(n)}).$$

12.4. Эмпирическая функция распределения

Определение 12.5. *Эмпирической* (или *выборочной*) функцией распределения называется функция действительной переменной x

$$F_n(x; \omega) = F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i < x\}}(\omega),$$

равная числу наблюдений, меньших x, деленному на общее количество наблюдений.

[1] Здесь и далее при использовании вариационного ряда всюду предполагается, что функция распределения $F(x)$ непрерывна.

Эмпирическая функция распределения при любом фиксированном $x \in \mathbb{R}$ является случайной величиной, так как это среднее арифметическое случайных величин

$$I_{\{X_i<x\}}(\omega) = \begin{cases} 1, & \text{если} \quad X_i(\omega) < x, \\ 0, & \text{если} \quad X_i(\omega) \geqslant x. \end{cases}$$

В тех случаях, кода это надо подчеркнуть, мы будем также использовать запись $F_n(x) = F_n(x;\omega)$.

С другой стороны, если в результате наблюдений реализуется конкретный элементарный исход ω, случайные величины $X_1(\omega),\ldots,X_n(\omega)$ принимают конкретные значения

$$x_1, \ x_2, \ \ldots, \ x_n.$$

По этим значениям можно построить *вариационный ряд*

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)},$$

и тогда эмпирическая функция $F_n(x;\omega)$ представляет собой ступенчатую функцию, изменяющуюся скачками величины $1/n$ в точках $x_{(i)}$.

Эмпирическая функция распределения является примером случайной функции, или, как говорят еще по-другому, случайного процесса, когда элементарный исход ω определяет не число, а функцию от x.

Замечание. Поскольку обычно результаты наблюдений записываются с округлением до какого-то знака (часто определяемого точностью измерительного прибора), то в выборках встречаются одинаковые значения. В этом случае в соответствующей точке $x_{(i)}$ изменение эмпирической функции будет равно не $1/n$, а k/n, где k — кратность данного значения.

В дальнейшем нам часто придется в наших рассуждениях использовать индикаторы случайных событий, связанных со случайными величинами. Поэтому для обозначения индикаторов таких событий будем использовать более краткую запись:

$$I_{\{X_i<x\}}(\omega) = I\{X_i < x\}.$$

Для эмпирической функции распределения выполнены соотношения:

$$\mathbf{E}F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}I\{X_i < x\} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{P}\{X_i < x\} = F(x),$$

$$\mathbf{D}\, F_n(x) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{D}\, I\{X_i < x\} = \frac{1}{n}F(x)(1-F(x)),$$

и при любых x

$$F_n(x) \xrightarrow{\mathbf{P}} F(x).$$

Отметим, что перечисленные свойства справедливы для случайной выборки с любой функцией распределения $F(x)$, если же распределение наблюдений непрерывно, то можно сформулировать более сильные утверждения.

Теорема 12.6. *Пусть $X_1, X_2, \ldots, X_n$ — независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$, $F_n(x)$ — эмпирическая функция распределения. Обозначим*

$$D_n(\omega) = \sup_{-\infty<x<\infty} |F_n(x;\omega) - F(x)|.$$

Тогда распределение $D_n(\omega)$ не зависит от $F(x)$.

Доказательство. Доказательство теоремы 12.6 проведем при дополнительном предположении, что функция $F(x)$ строго возрастает в окрестности любой точки x_0, где $0 < F(x_0) < 1$.

Пусть $0 < y < 1$ и $F(x) = y$, тогда $x = F^{-1}(y)$. В этом случае

$$I\{X_j < x\} = I\{X_j < F^{-1}(y)\} = I\{F(X_j) < y\} = \begin{cases} 1, & \text{если } F(X_j) < y, \\ 0, & \text{если } F(X_j) \geqslant y. \end{cases}$$

Введем новые случайные величины $U_j = F(X_j)$, $j = 1, \ldots, n$. Эти случайные величины также независимы и одинаково распределены, но каждая из них имеет функцию распределения

$$G(y) = \begin{cases} 0, & y \leqslant 0, \\ y, & 0 < y \leqslant 1, \\ 1, & y > 1. \end{cases}$$

Чтобы подчеркнуть, что выборочная функция $F_n(x)$ является случайной величиной, зависящей от случайной выборки $X_1, \ldots$ $\ldots, X_n$ с распределением $F(x)$, будем использовать запись

$$F_n(x) = F_n(x; X_1, \ldots, X_n; F).$$

Тогда

$$F_n(x) = F_n(x; X_1, \ldots, X_n; F) = \frac{1}{n}\sum_{j=1}^{n} I\{X_j < x\} = \frac{1}{n}\sum_{j=1}^{n} I\{U_j < y\} = \\ = F_n(y; U_1, \ldots, U_n; G)$$

и, значит,

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = \sup_{-\infty<y<\infty} |F_n(y; U_1, \ldots, U_n; G) - G(y)| = \\ = \sup_{0\leqslant y\leqslant 1} |F_n(y; U_1, \ldots, U_n; G) - y|.$$

Но последняя величина определяется только значениями равномерно распределенных случайных величин U_j и не зависит от $F(x)$. Теорема 12.6 доказана. □

Введем специальное обозначение для функции распределения случайной величины $\sqrt{n}D_n(\omega)$. Пусть

$$K_n(\lambda) = \mathbf{P}\left\{\omega\colon \sqrt{n}D_n(\omega) < \lambda\right\}.$$

Из теоремы 12.6 следует, что функция $K_n(\lambda)$ при всех n и λ не зависит от $F(x)$.

Приведем без доказательства следующую теорему.

Теорема 12.7 (А. Н. Колмогоров). *Для любого* $\lambda > 0$ *при* $n \to \infty$

$$K_n(\lambda) \to K(\lambda) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2\lambda^2}.$$

Функция $K(\lambda)$ называется *функцией распределения Колмогорова.* Перепишем ее в несколько ином виде:

$$K(\lambda) = 1 + 2\sum_{j=1}^{\infty}(-1)^j e^{-2j^2\lambda^2}. \tag{12.9}$$

Значения этой функции приводятся в книге [3]. Там же имеются таблицы функции распределения $K_n(\lambda)$ статистики $\sqrt{n}D_n$ и при малых n. Функцию $K_n(\lambda)$ также называют функцией Колмогорова.

Отметим, что при $\lambda \geqslant 1$ знакопеременный ряд (12.9) сходится очень быстро, ошибка приближения конечной суммой для $K(\lambda)$ не превосходит по абсолютной величине первого отброшенного члена и совпадает с ним по знаку.

Например, при $\lambda \geqslant 1.5$

$$K(\lambda) \approx 1 - 2e^{-2\lambda^2},$$

и погрешность $0 < K(\lambda) - \left(1 - 2e^{-2\lambda^2}\right) \leqslant 2e^{-8\lambda^2} < 10^{-7}$.

Доказательство теоремы Колмогорова достаточно сложно, обычно в учебной литературе не приводится. Интересующиеся могут прочитать оригинальное доказательство А. Н. Колмогорова в сборнике [15, с. 134–141].

С помощью теоремы Колмогорова можно строить доверительные интервалы для неизвестной функции распределения случайной выборки, проверять гипотезы о виде распределения.

Глава 13

ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ

13.1. Критерий согласия А. Н. Колмогорова

Пусть по независимым наблюдениям $X_1, X_2, \ldots, X_n$ с одинаковой, но неизвестной непрерывной функцией распределения $F(x)$ проверяется гипотеза

$$H_0\colon\ F(x) = F_0(x),$$

согласно которой неизвестная функция распределения $F(x)$ совпадает с некоторой известной непрерывной функцией распределения $F_0(x)$.

В качестве меры расхождения наблюдаемых данных с выдвинутой гипотезой выберем статистику

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)|.$$

Поскольку $F_0(x)$ непрерывная функция, а $F_n(x)$ — ступенчатая, то максимальное различие между ними будет в одной из точек роста последней, то есть будет достигаться в одном (или нескольких) из наблюдаемых значений X_i.

Если проверяемая гипотеза верна, то вычисленное по выборке значение D_n не может быть слишком большим, в противном случае гипотезу H_0 следует отвергнуть или хотя бы подвергнуть сомнению.

Пример 13.1. Допустим, что вычисления по результатам наблюдений дали

$$\sqrt{n} D_n = 1.73.$$

Если гипотеза H_0 верна, то при достаточно больших n по теореме Колмогорова (обычно ее используют уже при $n \geqslant 30$)

$$\mathbf{P}_{H_0}\left\{\sqrt{n} D_n \geqslant \lambda\right\} \approx 1 - K(\lambda)$$

и, следовательно,

$$\mathbf{P}_{H_0}\left\{\sqrt{n} D_n \geqslant 1.73\right\} \approx 1 - K(1.73) = 1 - 0.995 = 0.005.$$

Мы видим, что вероятность наблюдать такие или бо́льшие расхождения с проверяемой гипотезой мала, следовательно, вряд ли мы могли наблюдать столь маловероятное событие, и тогда проверяемая гипотеза должна быть отвергнута. При этом, отвергая гипотезу, мы могли совершить ошибку, но вероятность этой ошибки невелика, она равна 0.005.

Аналогичным образом поступают и в общем случае, только обычно вероятность ошибки неправильного решения задают заранее. А именно: выбирают малое положительное число α и считают, что ошибка отвергнуть правильную гипотезу не должна превышать α. Число α называется *уровнем значимости критерия*.

Соображения, сколь малым должно быть это число, лежат вне теории вероятностей, они определяются теми потерями, которые могут произойти при неправильном решении отвергнуть верную гипотезу. В каждой конкретной области руководствуются своими допустимыми нормами вероятностей ошибок.

Итак, уровень значимости α выбран. По таблицам значений функции $K(\lambda)$ можно найти λ_α — решение уравнения

$$1 - K(\lambda_\alpha) = \alpha.$$

Если число наблюдений n достаточно велико ($n \geqslant 30$), то

$$\mathbf{P}\left\{\sqrt{n}D_n \geqslant \lambda_\alpha\right\} \approx \alpha,$$

и далее следует действовать как в примере: если в результате проведенных наблюдений осуществилось событие

$$S = \left\{\sqrt{n}D_n \geqslant \lambda_\alpha\right\}, \tag{13.1}$$

то гипотезу H_0 отвергают, в противном случае — принимают. При этом вероятность отвергнуть верную гипотезу равна α.

Множество (13.1) называют *критическим множеством*, а λ_α — *критическим значением уровня α.*

При небольшом числе наблюдений действуют аналогичным образом с той лишь разницей, что для отыскания критического значения используют таблицы значений функции $K_n(\lambda)$.

Сформулированный критерий проверки гипотезы

$$H_0\colon F(x) = F_0(x)$$

о виде распределения называют *критерием согласия Колмогорова.*

Существуют и другие критерии согласия.

13.2. Критерий согласия Пирсона хи-квадрат

Этот критерий был предложен Карлом Пирсоном.

Пусть по-прежнему есть случайная выборка $X_1, X_2, \ldots, X_n$ с неизвестной функцией распределения $F(x)$.

Разобьем числовую прямую точками

$$-\infty = z_0 < z_1 < z_2 < \ldots < z_{r-1} < z_r = +\infty$$

на r непересекающихся интервалов. Обозначим

ν_1 — число наблюдений X_i, попавших на $A_1 = (-\infty, z_1)$,

ν_2 — число наблюдений X_i, попавших на $A_2 = [z_1, z_2)$,

....

ν_r — число наблюдений X_i, попавших на $A_r = [z_{r-1}, \infty)$.

Тогда $\nu_1 + \nu_2 + \ldots + \nu_r = n$.

Пусть снова проверяется гипотеза

$$H_0\colon\ F(x) = F_0(x).$$

Заметим, что теперь $F_0(x)$ может быть любой функцией распределения, в том числе и дискретной.

Вычислим $p_j = \mathbf{P}_{H_0}\{X_i \in A_j\}$, $j = 1, 2, \ldots, r$. А именно

$$p_j = F_0(z_j) - F_0(z_{j-1}).$$

Если гипотеза H_0 верна, то для $1 \leqslant j \leqslant r$ справедливо $\nu_j \sim B(n; p_j)$. При этом случайные величины

$$\frac{\nu_j - np_j}{\sqrt{np_j(1-p_j)}}$$

асимптотически нормальны с параметрами $(0; 1)$, и $\frac{\nu_j}{n} \xrightarrow{\mathbf{P}} p_j$.

Основываясь на этих свойствах частот ν_j, Карл Пирсон предложил ввести следующую меру расхождения гипотезы H_0 с имеющимися данными:

$$\chi^2 = \sum_{j=1}^{r} \frac{(\nu_j - np_j)^2}{np_j} = \sum_{j=1}^{r} \left(\frac{\nu_j - np_j}{\sqrt{np_j q_j}} \right)^2 \cdot q_j, \tag{13.2}$$

где $q_j = 1 - p_j$.

Пирсон назвал эту статистику «хи-квадрат» по названию греческой буквы, которой он обозначил статистику критерия.

Теорема 13.1 (К. Пирсон). *Если верна гипотеза H_0, то для всех $x>0$ при $n\to\infty$*

$$\mathbf{P}\left\{\chi^2<x\right\}\to\int\limits_0^x p_{r-1}(y)\,dy,$$

где

$$p_{r-1}(y)=\frac{y^{(r-3)/2}\,e^{-y/2}}{2^{(r-1)/2}\Gamma\left(\frac{r-1}{2}\right)}. \tag{13.3}$$

Замечание. Отметим, что при конечных n распределение статистики χ^2 зависит от $F_0(x)$, но предельное распределение при $n\to\infty$ зависит только от r.

Рассмотрим частный случай $r=2$. Тогда

$$\chi^2=\frac{(\nu_1-np_1)^2}{np_1}+\frac{(n-\nu_1-n(1-p_1))^2}{n(1-p_1)}=\frac{(\nu_1-np_1)^2}{n}\cdot\left(\frac{1-p_1+p_1}{p_1(1-p_1)}\right)=$$
$$=\left(\frac{\nu_1-np_1}{\sqrt{np_1(1-p_1)}}\right)^2\xrightarrow{d}Z_1^2,$$

где Z_1 обозначает случайную величину со стандартным нормальным законом распределения. Плотность случайной величины Z_1^2 равна

$$p_1(y)=\begin{cases}\dfrac{y^{-1/2}\,e^{-y/2}}{\sqrt{2\pi}}, & \text{если}\quad y>0,\\ 0, & \text{если}\quad y\leqslant 0,\end{cases}$$

то есть плотность имеет вид (13.3) с $r=2$. Подобное обстоятельство имеет место и при $r>2$.

Определение 13.1. Пусть $Z_1,Z_2,\ldots,Z_k$ — независимые стандартные нормальные величины. Распределение случайной величины

$$\chi_k^2=Z_1^2+Z_2^2+\ldots+Z_k^2$$

называют *распределением хи-квадрат с k степенями свободы.*

Задача 13.1. Покажите, что плотность распределения χ_k^2 равна

$$p_k(x)=\begin{cases}\dfrac{x^{k/2-1}\,e^{-x/2}}{2^{k/2}\,\Gamma(k/2)}, & \text{если}\quad x>0,\\ 0, & \text{если}\quad x\leqslant 0.\end{cases}$$

Таким образом, теорема Пирсона утверждает, что если справедлива гипотеза H_0, то статистика χ^2, определенная в (13.2), сходится по распределению к χ^2_{r-1}.

Если число наблюдений n велико ($n \geqslant 30$), то теорема Пирсона позволяет проверять гипотезу

$$H_0\colon F(x) = F_0(x)$$

следующим образом:

1) по наблюдениям вычисляется значение статистики χ^2;

2) гипотезу H_0 отвергают, если расхождение χ^2 с гипотезой окажется велико, и принимают, если это расхождение невелико.

Критическое множество имеет вид

$$S = \{\chi^2 \geqslant x_\alpha\},$$

где критическое значение x_α находится по таблицам распределения χ^2_{r-1} из условия

$$\mathbf{P}\{\chi^2_{r-1} \geqslant x_\alpha\} = \alpha.$$

При таком выборе критического значения вероятность отклонить верную гипотезу равна

$$\mathbf{P}_{H_0}(S) = \mathbf{P}\{\chi^2_{r-1} \geqslant x_\alpha\} = \alpha.$$

Замечание. В случае $r = 2$ критическое множество

$$S = \left\{ \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} \right)^2 \geqslant x_\alpha \right\} = \left\{ \frac{|\nu_1 - np_1|}{\sqrt{np_1(1-p_1)}} \geqslant \sqrt{x_\alpha} \right\}.$$

Следовательно, мы должны принимать решение в зависимости от значения статистики

$$\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}},$$

которая при $n \to \infty$ асимптотически нормальна с параметрами $(0; 1)$.

Пример 13.2. В учебнике Г. Крамера [17] приведены данные о рождении мальчиков и девочек в 1935 г. в Швеции. Всего в октябре родилось 6903 ребенка, среди которых было 3512 мальчиков и 3391 девочка. Если считать рождение девочки успехом и обозначить вероятность этого события p, то для проверки гипотезы

$$H_0\colon p = 1/2$$

можно воспользоваться критерием χ^2. Вычисления дают

$$\frac{|\nu_1 - np_1|}{\sqrt{np_1(1-p_1)}} = \frac{|3391 - 0.5 \cdot 6903|}{\sqrt{6903 \cdot 0.5 \cdot 0.5}} = 1.456.$$

Если гипотеза H_0 верна, то вероятность наблюдать данное отклонение от гипотезы равна

$$\mathbf{P}\{|Z| \geqslant 1.456\} = 0.1453.$$

Мы видим, что это не такое уж редкое событие, критическая точка нормального распределения, соответствующая уровню значимости $\alpha = 0.05$, равна 1.96. Следовательно, гипотеза H_0 принимается. Оснований сомневаться в справедливости выдвинутой гипотезы нет.

С другой стороны, по данным Г. Крамера, в апреле из общего числа 7884 родившихся детей мальчиков было 4173, девочек — 3711. В этом случае вычисления дают

$$\frac{|3711 - 0.5 \cdot 7884|}{\sqrt{7884 \cdot 0.5 \cdot 0.5}} = 5.203,$$

вероятность таких и бóльших отклонений оказывается равной

$$\mathbf{P}\{|Z| \geqslant 5.203\} = 2 \cdot 10^{-7},$$

слишком малой, поэтому гипотеза H_0 должна быть отвергнута. К такому же выводу мы придем, если проанализируем данные за весь год: среди общего числа 88 273 родившихся детей мальчиков было 45 682, а девочек — 42 591. Значение статистики критерия в этом случае оказалось равным 10.4. Вероятность наблюдать такие отклонения равна $2 \cdot 10^{-25}$, что можно считать практически невозможным.

Таким образом, проанализировав большие группы новорожденных, статистики пришли к выводу, что вероятность рождения мальчика выше вероятности рождения девочки и приближенно равна 0.517.

Пример 13.3. Некий любознательный статистик, рассматривая витрины часовщиков, записывал время на часах, причем минуты не учитывались. Полученные данные приведены в таблице:

0	1	2	3	4	5	6	7	8	9	10	11
41	34	54	39	49	45	41	33	37	41	47	39

Можно ли считать, что верна гипотеза

$$H_0\colon\ p_0 = p_1 = \ldots = p_{12} = \frac{1}{12}$$

о равномерности распределения? Поскольку

$$\chi^2 = \sum_{j=1}^{12} \frac{(\nu_j - 500 \cdot 1/12)^2}{41.67} = 10.00,$$

то, вычислив по таблицам

$$\mathbf{P}\{\chi^2 \geqslant 10.00\} = 0.53039,$$

приходим к выводу, что если гипотеза верна, то это событие происходит в среднем в 53 случаях из 100. Сравнивая значение статистики критерия с 24.7 — критической точкой уровня 0.01 распределения хи-квадрат с одинадцатью степенями свободы, мы приходим к выводу, что данные не противоречат гипотезе о равномерном распределении. Гипотезу следует принять.

Задача 13.2. По данным розыгрышей генуэзской лотереи, приведенным в задаче 1.2, проверьте гипотезу о справедливости выбора классической модели.

Пример 13.4. Критерий хи-квадрат можно модернизировать и использовать для проверки гипотезы о независимости признаков. В учебнике [17] приведены данные на 927 570 новорожденных, а именно: в таблице на с. 496 приведено распределение по полу детей и по возрасту родителей. Мы составим упрощенную таблицу, в которой по горизонтали данные разделены на четыре группы:

A_1 — молодые отцы (до 35 лет), молодые матери (до 30 лет);
A_2 — молодые отцы, взрослые матери (старше 30 лет);
A_3 — взрослые отцы (старше 35 лет), молодые матери;
A_4 — взрослые отцы, взрослые матери:

	A_1	A_2	A_3	A_4	Σ
B_1	157 604	86 702	36 998	195 229	476 533
B_2	149 913	81 360	35 389	184 375	451 037
Σ	307 517	168 062	72 387	379 604	927 570

По вертикали данные разбиты на две группы по полу ребенка: B_1 — мальчики и B_2 — девочки. В клетке с номером i по горизонтали и с номером j по вертикали в этой таблице указано количество детей, одновременно обладающих признаками B_iA_j. Такая таблица называется таблицей сопряженности признаков.

Двойными линиями от основной таблицы отделены строка и столбец с суммарными значениями по столбцам и строкам соответственно.

Сформулируем проверяемую гипотезу

$$H_0\colon\ \mathbf{P}(A_jB_i) = \mathbf{P}(A_j)\cdot\mathbf{P}(B_i).$$

По имеющимся в таблице данным оценим вероятности попадания наблюдений в классы:

$$\begin{array}{lll} \widehat{\mathbf{P}}(A_1) = 0.332 & \widehat{\mathbf{P}}(A_3) = 0.078 & \widehat{\mathbf{P}}(B_1) = 0.513 \\ \widehat{\mathbf{P}}(A_2) = 0.181 & \widehat{\mathbf{P}}(A_4) = 0.409 & \widehat{\mathbf{P}}(B_2) = 0.487 \end{array}$$

В качестве оценок попадания в какой-либо класс были взяты соответствующие частоты. Статистика критерия вычисляется по формуле

$$\chi^2 = \sum_{j=1}^{k}\sum_{i=1}^{s} \frac{\left(\nu_{ji} - \nu\widehat{\mathbf{P}}(A_j)\widehat{\mathbf{P}}(B_i)\right)^2}{\nu\widehat{\mathbf{P}}(A_j)\widehat{\mathbf{P}}(B_i)},$$

в которой ν_{ji} — число наблюдений, попавших в класс $A_j \cap B_i$, $1 \leqslant j \leqslant k$, $1 \leqslant i \leqslant s$, а ν — общее число наблюдений. Сформулируем теорему, на которой основано принятие статистического решения.

Теорема 13.2. *Если гипотеза о независимости признаков верна, то при $n \to \infty$*

$$\chi^2 \xrightarrow{d} \chi^2_{(k-1)(s-1)}.$$

В нашем примере $k = 4$, $s = 2$, а наблюдаемое значение статистики оказалось равным $\chi^2 = 7.499$. Критическое значение, соответствующее уровню значимости $\alpha = 0.05$, найденное по таблицам распределения хи-квадрат с тремя степенями свободы, равно 7.81. Тем самым имеющиеся данные согласуются с гипотезой о независимости пола ребенка от возраста родителей.

Подробно о критерии хи-квадрат можно прочитать в учебнике [17].

Глава 14

ПРОВЕРКА ПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ. ФУНДАМЕНТАЛЬНАЯ ЛЕММА НЕЙМАНА–ПИРСОНА

Мы не случайно начали книгу с высказывания выдающегося математика П.-С. Лапласа. Как вы уже могли заметить, при принятии решений, выработке статистических критериев для проверки гипотез математики во многом руководствуются здравым смыслом. Нижеследующий материал станет тому дальнейшим подтверждением.

14.1. Квантили и процентные точки нормального распределения

Пусть $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,du$ — функция распределения стандартного нормального закона, соответствующая случайной величине Z.

Обозначим p_α, λ_α, u_α решения уравнений (14.1), (14.2), (14.3) соответственно:

$$\mathbf{P}\{Z \leqslant p_\alpha\} = \alpha \Longleftrightarrow \Phi(p_\alpha) = \alpha; \tag{14.1}$$

$$\mathbf{P}\{|Z| \geqslant \lambda_\alpha\} = \alpha \Longleftrightarrow 2(1 - \Phi(\lambda_\alpha)) = \alpha; \tag{14.2}$$

$$\mathbf{P}\{Z \geqslant u_\alpha\} = \alpha \Longleftrightarrow 1 - \Phi(u_\alpha) = \alpha. \tag{14.3}$$

Среди введенных величин p_α — это квантиль уровня α, u_α называют $\alpha \cdot 100$-процентной точкой распределения (или критической точкой уровня α), λ_α — двухсторонней критической точкой уровня α. Поскольку между p_α, λ_α, u_α выполняются равенства

$$p_\alpha = u_{1-\alpha}; \quad \lambda_\alpha = u_{\alpha/2},$$

то достаточно по значениям α уметь вычислять u_α.

Приведем таблицу некоторых значений u_α, которые используются наиболее часто:

α	0.0001	0.001	0.005	0.01	0.025	0.05
u_α	3.7190	3.0902	2.5758	2.3263	1.9600	1.6449

Из таблицы видно, как с ростом u_α значения α быстро убывают.

Для λ_α приведем лишь несколько значений. Если $\alpha = 0.5$, то $\lambda_{0.5} = 0.6745$. Эта точка называется *вероятным отклонением* стандартного нормального распределения, для нее

$$\mathbf{P}\{|Z| \leqslant \lambda_{0.5}\} = \mathbf{P}\{|Z| \geqslant \lambda_{0.5}\} = 0.5.$$

Если $\alpha = 0.0027$, то $\lambda_{0.0027} = 3$, то есть

$$\mathbf{P}\{|Z| \leqslant 3\} = 0.9973,$$

и почти все распределение сосредоточено на отрезке $[-3, 3]$. Для случайной величины $V \sim N(a, \sigma^2)$

$$\mathbf{P}\{|V - a| \geqslant 3\sigma\} = 0.0027.$$

Во многих приложениях теории вероятностей событие с такой малой вероятностью считают практически невозможным. Это служит обоснованием эмпирического правила «трех σ»: если наблюдаются отклонения от среднего значения, превышающие три стандартных отклонения, то гипотеза о нормальном распределении для такой случайной величины должна быть отклонена или хотя бы подвергнута сомнению.

Критерии Колмогорова и хи-квадрат К. Пирсона относятся к критериям согласия, когда по результатам наблюдений $X_1, X_2, \ldots, X_n$ с общей функцией распределения $F(x)$ проверяется гипотеза

$$H_0\colon\ F(x) = F_0(x),$$

где $F_0(x)$ — некоторая фиксированная функция распределения.

Пример 14.1. Пусть проводится только одно наблюдение $X \sim N(a, 1)$. По значению X требуется проверить гипотезу

$$H_0\colon\ F(x) = \Phi(x), \quad \text{или, что то же самое,} \quad a = 0.$$

Зададим $\alpha = 0.05$ — вероятность ошибки, которой можно пренебречь. Тогда $\lambda_{0.05} = 1.96$. Если гипотеза H_0 верна, то

$$\mathbf{P}\{|X| \geqslant 1.96\} = 0.05,$$

то есть вероятность наблюдать такие значения пренебрежимо мала. Следовательно, если мы получим в ходе наблюдений, что $|X| \geqslant 1.96$, то гипотеза должна быть отвергнута. Но в такой постановке задачи непонятно, что же делать дальше: какую гипотезу принимать?

14.2. Постановка задачи. Ошибки первого и второго рода

В 1930-е гг. Ю. Нейман и Э. Пирсон предложили другую постановку задач. Их идеи послужили основой современной теории проверки статистических гипотез.

Проиллюстрируем разницу в подходах на предыдущем примере. Сформулируем *основную гипотезу*

$$H_0\colon\ a = a_0 \quad (a_0 = 0)$$

и *альтернативную*, или *конкурирующую*, *гипотезу*

$$H_1\colon\ a = a_1 \quad (a_1 = 2.3).$$

Обозначим $p_0(x) = \varphi(x)$ — плотность распределения X при основной гипотезе H_0 и $p_1(x) = \varphi(x - a_1)$ — плотность распределения X при конкурирующей гипотезе H_1.

Выберем некоторую границу x_0 и будем поступать следующим образом:

если $X < x_0$, то будем принимать H_0, отвергая H_1;

если $X \geqslant x_0$, то отвергнем H_0 и примем H_1.

При этом мы можем совершить ошибку. Ошибки бывают двух типов:

1) *ошибка первого рода*, когда отвергаем верную гипотезу H_0;

2) *ошибка второго рода*, когда принимаем H_0, а верна H_1.

Этим ошибкам соответствуют вероятности

$$\mathbf{P}_{a_0}\{X \geqslant x_0\} = \int_{x_0}^{\infty} p_0(x)\,dx,$$

$$\mathbf{P}_{a_1}\{X < x_0\} = \int_{-\infty}^{x_0} p_1(x)\,dx.$$

На рис. 10 ошибке первого рода соответствует заштрихованная площадь справа от точки x_0, ошибке второго рода — закрашенная площадь слева от точки x_0. Из рисунка видно, что, уменьшая одну вероятность, мы увеличиваем другую.

При заданном числе наблюдений нельзя одновременно сделать обе ошибки сколь угодно малыми. Обычно задают величину вероятности ошибки первого рода α (ее называют *уровнем*

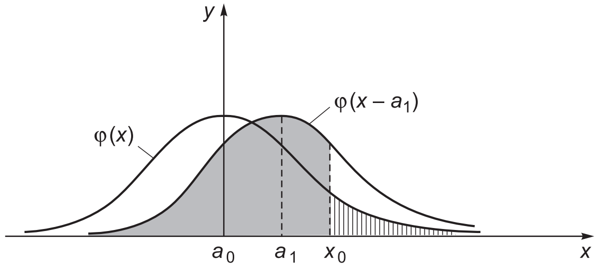

Рис. 10. Проверка гипотез о параметре сдвига нормального распределения

значимости критерия) и ищут такое x_0, чтобы минимизировать вероятность ошибки второго рода.

Выберем $x_0 = u_\alpha$, чтобы выполнялось равенство

$$\mathbf{P}_{a_0}\{X \geqslant x_0\} = \alpha.$$

Но тогда вероятность ошибки второго рода равна

$$\mathbf{P}_{a_1}\{X < u_\alpha\} = \mathbf{P}_{a_1}\{X - a_1 < u_\alpha - a_1\} = \Phi(u_\alpha - a_1).$$

Чтобы вероятность ошибки второго рода не превосходила заданную границу β, необходимо выполнение неравенства

$$\Phi(u_\alpha - a_1) \leqslant \beta = \Phi(-u_\beta).$$

В силу монотонности функции распределения $\Phi(x)$ получаем

$$u_\alpha - a_1 \leqslant -u_\beta \Longleftrightarrow a_1 - 0 \geqslant u_\alpha + u_\beta,$$

то есть вероятности ошибок, не превосходящие α и β, возможны лишь в том случае, когда гипотеза и альтернатива достаточно далеки друг от друга.

14.3. Лемма Неймана–Пирсона

Эта лемма может быть сформулирована несколькими способами. Приведем наиболее удобную с нашей точки зрения.

Мы начнем со случая конечного вероятностного пространства, когда $\Omega = \{\omega_1, \omega_2, \ldots, \omega_s\}$ состоит из конечного числа элементарных исходов, алгебра $\mathcal{A}$ содержит все подмножества и есть два возможных распределения вероятностей:

$$\mathbf{P}_0(A) = \sum_{\omega \in A} p_0(\omega),$$

$$\mathbf{P}_1(A) = \sum_{\omega \in A} p_1(\omega)$$

на измеримом пространстве $(\Omega, \mathcal{A})$. Предположим, что для всех ω выполнено $p_0(\omega) > 0$.

Для проверки основной гипотезы

$$H_0\colon \mathbf{P} = \mathbf{P}_0$$

против конкурирующей

$$H_1\colon \mathbf{P} = \mathbf{P}_1$$

строят критическое множество S и далее действуют по следующему правилу: если результат наблюдений $\omega \in S$, то гипотезу H_0 отвергают и принимают H_1, если же $\omega \notin S$, то принимают H_0 и отвергают H_1. Такой способ действий называется *критерием* S или S-*критерием*. Каждому выбору S соответствуют вероятности ошибок первого и второго рода:

$$\alpha(S) = \mathbf{P}_0(S) \quad \text{и} \quad \beta(S) = \mathbf{P}_1(\overline{S})$$

соответственно.

Теорема 14.1 (лемма Неймана–Пирсона). *Пусть* $0 < \alpha < 1$ *и существует* c_α *такое, что* $\mathbf{P}_0\left\{\omega\colon \frac{p_1(\omega)}{p_0(\omega)} \geqslant c_\alpha\right\} = \alpha$. *Выберем*

$$S^* = \left\{\omega\colon \frac{p_1(\omega)}{p_0(\omega)} \geqslant c_\alpha\right\}. \tag{14.4}$$

Тогда для любого критерия S *уровня* α ($\mathbf{P}_0(S) = \alpha$)

$$\mathbf{P}_1(\overline{S^*}) \leqslant \mathbf{P}_1(\overline{S}).$$

Доказательство. Пусть S — произвольный критерий уровня α, а S^* определяется в (14.4), тогда

$$\alpha = \mathbf{P}_0(S^*) = \mathbf{P}_0(S^* \cap \overline{S}) + \mathbf{P}_0(S^* \cap S),$$
$$\alpha = \mathbf{P}_0(S) = \mathbf{P}_0(\overline{S^*} \cap S) + \mathbf{P}_0(S^* \cap S).$$

Откуда получаем

$$\mathbf{P}_0(S^* \cap \overline{S}) = \mathbf{P}_0(\overline{S^*} \cap S).$$

Далее,

$$1 - \beta(S^*) = \mathbf{P}_1(S^*) = \mathbf{P}_1(S^* \cap \overline{S}) + \mathbf{P}_1(S^* \cap S),$$
$$1 - \beta(S) = \mathbf{P}_1(S) = \mathbf{P}_1(S \cap \overline{S^*}) + \mathbf{P}_1(S^* \cap S).$$

Мы желаем доказать, что $\beta(S^*) \leqslant \beta(S)$ или, что то же самое,

$$\mathbf{P}_1(S^* \cap \overline{S}) \geqslant \mathbf{P}_1(S \cap \overline{S^*}).$$

Действительно,

$$\mathbf{P}_1(S^* \cap \overline{S}) = \sum_{\omega \in S^* \cap \overline{S}} p_1(\omega) \geqslant c_\alpha \sum_{\omega \in S^* \cap \overline{S}} p_0(\omega) = c_\alpha \mathbf{P}_0(S^* \cap \overline{S}) =$$
$$= c_\alpha \mathbf{P}_0(\overline{S^*} \cap S) = c_\alpha \sum_{\omega \in \overline{S^*} \cap S} p_0(\omega) \geqslant \sum_{\omega \in \overline{S^*} \cap S} p_1(\omega) = \mathbf{P}_1(\overline{S^*} \cap S).$$

Теорема доказана. □

Определение 14.1. Число $W(S) = 1 - \beta(S)$ называют *мощностью критерия* S.

Критерий S^* в лемме Неймана–Пирсона минимизирует вероятность ошибки второго рода среди всех критериев уровня α. А это означает, что в данном классе критериев ($\mathbf{P}_0(S) = \alpha$) критерий S^* максимизирует мощность, поэтому его называют *наиболее мощным критерием уровня* α.

Задача 14.1. Покажите, что критерий S^* является наиболее мощным среди всех критериев S, удовлетворяющих условию

$$\mathbf{P}_0(S) \leqslant \alpha. \tag{14.5}$$

Определение 14.2. Всякий критерий, удовлетворяющий условию 14.5, также называют *критерием уровня* α (уровня значимости α).

Лемма Неймана–Пирсона показывает, как устроено наиболее мощное критическое множество S^*:

$$S^* = S^*(c_\alpha) = \left\{\omega\colon \frac{p_1(\omega)}{p_0(\omega)} \geqslant c_\alpha\right\}. \tag{14.6}$$

В критическое множество S^* включаются те точки, вероятность которых при H_1 существенно больше, чем при H_0. Ранее мы оговаривали, что $p_0(\omega) > 0$ для всех ω. От этого предположения можно отказаться, включив точки, в которых $p_0(\omega) = 0$, в критическое множество (формально в этом случае можно считать отношение вероятностей равным $+\infty$).

Замечание. Константа c_α определяется как решение уравнения

$$\mathbf{P}_0(S^*(c_\alpha)) = \alpha,$$

которое может иметь решение не для всех α. Чаще всего такое бывает для дискретных распределений. В таких случаях можно построить *рандомизированный критерий*, который определяется не критическим множеством, а критической функцией.

Определение 14.3. *Критической функцией* называют функцию $\varphi(\omega)$ такую, что $0 \leqslant \varphi(\omega) \leqslant 1$. Значение $\varphi(\omega)$ трактуют как вероятность отвергнуть основную гипотезу, если произойдет исход ω.

В случае, когда критическая функция принимает только два значения:

$$\varphi(\omega) = \begin{cases} 1, & \text{если } \omega \in S, \\ 0, & \text{если } \omega \in \overline{S}, \end{cases}$$

критерий называют *нерандомизированным* и выбор $\varphi(\omega)$ эквивалентен выбору критического множества S. Если существует элементарный исход ω, для которого $0 < \varphi(\omega) < 1$, то критерий называют *рандомизированным*[1], поскольку окончательное принятие решения в задаче проверки гипотез связано с реализацией некоторого эксперимента с двумя исходами, один из которых имеет вероятность, равную $\varphi(\omega)$.

Вероятность ошибки первого рода для рандомизированного критерия $\varphi(\omega)$ равна

$$\mathbf{E}_0\,\varphi = \sum_{\omega\in\Omega} \varphi(\omega)p_0(\omega),$$

а мощность критерия —

$$\mathbf{E}_1\,\varphi = \sum_{\omega\in\Omega} \varphi(\omega)p_1(\omega).$$

Приведем вариант леммы Неймана–Пирсона для рандомизированных критериев[2].

[1] Название происходит от английского random — случайный.

[2] Доказательство можно посмотреть, например, в [26] или провести самостоятельно.

Теорема 14.2. *Для любого* $0 < \alpha < 1$ *существует критерий*

$$\varphi^*(\omega) = \begin{cases} 1, & \text{если } \dfrac{p_1(\omega)}{p_0(\omega)} > c_\alpha, \\ 0, & \text{если } \dfrac{p_1(\omega)}{p_0(\omega)} < c_\alpha, \\ \gamma_\alpha, & \text{если } \dfrac{p_1(\omega)}{p_0(\omega)} = c_\alpha, \end{cases}$$

для которого $\mathbf{E}_0\,\varphi* = \alpha$. *Этот критерий является наиболее мощным среди всех критериев уровня* α, *то есть*

$$\mathbf{E}_1\,\varphi* \geqslant \mathbf{E}_1\,\varphi$$

для любого критерия, удовлетворяющего неравенству $\mathbf{E}_0\,\varphi \leqslant \alpha$.

Следует, однако, заметить, что в практических задачах стараются избегать рандомизированных критериев, немного изменяя в ту или иную сторону уровень значимости α.

Критерий $\varphi^*(\omega)$ называют также критерием отношения правдоподобия.

Лемма Неймана–Пирсона была доказана нами для дискретных случайных величин, но ее формулировка и доказательство легко переносятся на непрерывный случай.

В классических задачах математической статистики, как правило, множество элементарных исходов Ω совпадает с $\mathbb{R}^n$, результат наблюдения $\omega = (x_1, x_2, \ldots, x_n)$ — точка n-мерного пространства, а $X_1, \ldots, X_n$ — координатные случайные величины, для которых

$$X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n.$$

В качестве класса случайных событий $\mathcal{A}$ обычно выбирают борелевскую σ-алгебру в $\mathbb{R}^n$.

Пусть на $\mathbb{R}^n$ заданы два распределения вероятностей $\mathbf{P}_0$ и $\mathbf{P}_1$. Распределения будем рассматривать или дискретные, или абсолютно непрерывные.

Для $k = 0$ или $k = 1$ величина $\widehat{p}_k(x_1, x_2, \ldots, x_n)$ будет обозначать либо совместную вероятность $\mathbf{P}_k\{X_1 = x_1, \ldots, X_n = x_n\}$, либо совместную плотность распределения компонент случайного вектора $\boldsymbol{X} = (X_1, \ldots, X_n)$, вычисленные или при гипотезе H_0, или при гипотезе H_1 соответственно.

Если $X_1, \ldots, X_n$ — независимые, одинаково распределенные случайные величины, то

$$\widehat{p}_k(x_1, x_2, \ldots, x_n) = p_k(x_1) \cdot p_k(x_2) \cdots p_k(x_n),$$

где $p_k(x_i)$ обозначают или $\mathbf{P}_k\{X_i = x_i\}$, или плотность распределения случайной величины X_i при гипотезе H_k.

Тогда наиболее мощное критическое множество (14.6) перепишется в виде

$$S^* = S^*(c_\alpha) = \left\{ \frac{\widehat{p}_1(X_1, \ldots, X_n)}{\widehat{p}_0(X_1, \ldots, X_n)} \geqslant c_\alpha \right\}. \tag{14.7}$$

14.4. Проверка гипотез о параметрах нормального распределения

Применим лемму Неймана–Пирсона к задачам проверки гипотез о параметрах нормального распределения.

14.4.1. Проверка гипотез о математическом ожидании

Пусть $X_1, \ldots X_n$ — независимые случайные величины, распределенные одинаково по нормальному закону $N(a, \sigma^2)$. Рассмотрим задачу о различении двух простых гипотез:

$$\begin{aligned} &H_0\colon\ a = a_0, \\ &H_1\colon\ a = a_1 \quad (a_1 > a_0), \end{aligned}$$

при этом дисперсию σ^2 будем считать известной.

Отметим, что

$$a_0 = \mathbf{E}_0 X_i, \quad a_1 = \mathbf{E}_1 X_i, \quad \sigma^2 = \mathbf{D}_0 X_i = \mathbf{D}_1 X_i.$$

Индексы у математического ожидания и дисперсии показывают, по какому из распределений ($\mathbf{P}_0$ или $\mathbf{P}_1$) вычислялись эти числовые характеристики.

Согласно лемме Неймана–Пирсона, наиболее мощный критерий имеет вид

$$S_1^* = \left\{ \omega\colon\ \frac{p_1(X_1) \cdots p_1(X_n)}{p_0(X_1) \cdots p_0(X_n)} \geqslant c \right\} = \left\{ \omega\colon\ \sum_{j=1}^{n} \ln \frac{p_1(X_j)}{p_0(X_j)} \geqslant c' \right\},$$

где $c' = \ln c$.

Обозначим

$$W_j = \ln \frac{p_1(X_j)}{p_0(X_j)},$$

при этом случайные величины $W_1, \ldots, W_n$ независимы и одинаково распределены. Критическое множество можно переписать следующим образом:

$$S_1^* = \{W_1 + W_2 + \ldots + W_n \geqslant c'\}.$$

Вероятности ошибок первого и второго рода запишутся теперь как

$$\mathbf{P}_0\{W_1 + \ldots + W_n \geqslant c'\} = \alpha(S_1^*),$$
$$\mathbf{P}_1\{W_1 + \ldots + W_n < c'\} = \beta(S_1^*).$$

Поскольку в рассматриваемом случае

$$p_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a_0)^2/(2\sigma^2)},$$
$$p_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a_1)^2/(2\sigma^2)},$$

то

$$W_j = \ln \frac{p_1(X_j)}{p_0(X_j)} = \frac{-(X_j - a_1)^2 + (X_j - a_0)^2}{2\sigma^2} = \frac{a_1 - a_0}{\sigma^2} X_j - \frac{a_1^2 - a_0^2}{2\sigma^2}.$$

Но тогда

$$W_1 + \ldots + W_n = \frac{a_1 - a_0}{\sigma^2} \sum_{j=1}^{n} X_j - \frac{n(a_1^2 - a_0^2)}{2\sigma^2}.$$

Следовательно, критическое множество можно переписать в другом виде:

$$S_1^* = \left\{ \sum_{j=1}^{n} W_j \geqslant c' \right\} = \left\{ \sum_{j=1}^{n} X_j \geqslant c'' \right\},$$

где

$$c'' = \left(c' + \frac{n(a_1^2 - a_0^2)}{2\sigma^2} \right) \cdot \frac{\sigma^2}{a_1 - a_0}.$$

Воспользуемся тем, что $\sum\limits_{j=1}^{n} X_j \sim N(an; \sigma^2 n)$. Если верна гипотеза H_0, то

$$\mathbf{E}_0\left(\sum_{j=1}^{n} X_j\right) = a_0 n, \quad \mathbf{D}_0\left(\sum_{j=1}^{n} X_j\right) = n\sigma^2.$$

Если же верна H_1, то

$$\mathbf{E}_1\left(\sum_{j=1}^{n} X_j\right) = a_1 n, \quad \mathbf{D}_1\left(\sum_{j=1}^{n} X_j\right) = n\sigma^2.$$

Выберем $c'' = na_0 + u_\alpha \sigma\sqrt{n}$, тогда вероятность ошибки первого рода равна

$$\mathbf{P}_0\left\{\sum_{j=1}^{n} X_j \geqslant na_0 + u_\alpha\sigma\sqrt{n}\right\} = \mathbf{P}_0\left\{\frac{\sum_{j=1}^{n} X_j - na_0}{\sigma\sqrt{n}} \geqslant u_\alpha\right\} =$$
$$= 1 - \Phi(u_\alpha) = \alpha. \quad (14.8)$$

Вероятность ошибки второго рода при этом равна

$$\beta(S_1^*) = \mathbf{P}_1\left\{\sum_{j=1}^{n} X_j < na_0 + u_\alpha\sigma\sqrt{n}\right\} =$$
$$= \mathbf{P}_1\left\{\frac{\sum_{j=1}^{n} X_j - na_1}{\sigma\sqrt{n}} < u_\alpha - \frac{n(a_1 - a_0)}{\sigma\sqrt{n}}\right\} = \Phi\left(u_\alpha - \frac{(a_1 - a_0)\sqrt{n}}{\sigma}\right).$$

Воспользовавшись тем, что $\Phi(-x) = 1 - \Phi(x)$, получим

$$\beta(S_1^*) = 1 - \Phi\left(\sqrt{n}\,\frac{(a_1 - a_0)}{\sigma} - u_\alpha\right) \to 0$$

при $n \to \infty$.

Отметим также, что $\beta(S^*)$ зависит от величины $\frac{a_1 - a_0}{\sigma}$, которую называют информационным расстоянием между гипотезами. Это расстояние измеряется в стандартных отклонениях, оно тем больше, чем меньше стандартное отклонение. Чем меньше это расстояние, тем труднее различать гипотезы.

Полученные формулы позволяют оценить необходимое число наблюдений для достижения заранее установленных уровней ошибок первого и второго рода.

14.4.2. Необходимое число наблюдений для различения гипотез

Лемма 14.1. 1. *При* $x > 0$

$$1 - \Phi(x) \leqslant \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

2. *При* $x > 1$

$$1 - \Phi(x) \geqslant \frac{1}{x\sqrt{2\pi}} \left(1 - \frac{1}{x^2}\right) e^{-x^2/2}.$$

Доказательство.

$$1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_x^\infty e^{-u^2/2}\, du \leqslant \frac{1}{\sqrt{2\pi}} \int\limits_x^\infty \frac{u}{x} e^{-u^2/2}\, du = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

Доказательство второй части утверждения леммы можно посмотреть в [29, т. 1, с. 192]. □

Замечание. Воспользовавшись леммой, можно получить оценку для вероятности ошибки второго рода:

$$\beta(S_1^*) \leqslant ce^{-hn/2},$$

где c и h — некоторые положительные константы, зависящие от информационного расстояния и α и не зависящие от n. Таким образом, эта вероятность убывает с ростом числа наблюдений n, как геометрическая прогрессия и даже быстрее.

Выясним, сколько нужно иметь наблюдений, чтобы вероятность ошибки второго рода была меньше заданного числа β. Для этого решим неравенство

$$\begin{aligned}
&\Phi\left(-\sqrt{n}\,\frac{a_1 - a_0}{\sigma} + u_\alpha\right) \leqslant \beta \iff -\sqrt{n}\,\frac{a_1 - a_0}{\sigma} + u_\alpha \leqslant -u_\beta \iff \\
&\iff \sqrt{n}\,\frac{a_1 - a_0}{\sigma} \geqslant u_\alpha + u_\beta \iff \sqrt{n} \geqslant (u_\alpha + u_\beta)\frac{\sigma}{a_1 - a_0} \iff \\
&\iff n \geqslant \frac{\sigma^2 (u_\alpha + u_\beta)^2}{(a_1 - a_0)^2} = \frac{(u_\alpha + u_\beta)^2}{\left(\dfrac{a_1 - a_0}{\sigma}\right)^2}. \qquad (14.9)
\end{aligned}$$

Так, если $a_1 - a_0 = 0.01$, $\sigma = 0.01$, $\alpha = \beta = 0.05$, информационное расстояние между гипотезами равно 1 и необходимое число наблюдений $n \geqslant 11$.

Если изменить стандартное отклонение, увеличив его в 10 раз ($\sigma = 0.1$), то расстояние между гипотезами уменьшится в 10 раз, а число необходимых наблюдений увеличится почти в 100 раз, и мы получим $n \geqslant 1082$.

Чтобы различать близкие между собой гипотезы с малыми вероятностями ошибок первого и второго рода, необходимо проводить достаточно много наблюдений.

14.4.3. Проверка гипотез о дисперсии

Рассмотрим новую задачу. Пусть теперь случайная выборка взята из распределения $N(0, \sigma^2)$, о дисперсии которого выдвинуты две гипотезы:

$$H_0\colon \sigma^2 = \sigma_0^2,$$
$$H_1\colon \sigma^2 = \sigma_1^2 \quad (\sigma_1^2 > \sigma_0^2).$$

В этом случае

$$p_0(x) = \frac{1}{\sigma_0\sqrt{2\pi}}\, e^{-x^2/(2\sigma_0^2)},$$
$$p_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-x^2/(2\sigma_1^2)},$$

и

$$W_j = \ln\frac{p_1(X_j)}{p_0(X_j)} = \ln\frac{\sigma_0}{\sigma_1} + \frac{1}{2}X_j^2\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right).$$

Но тогда сумма

$$\sum_{j=1}^{n} W_j = n\ln\frac{\sigma_0}{\sigma_1} + \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum_{j=1}^{n} X_j^2$$

является возрастающей функцией от $\sum\limits_{j=1}^{n} X_j^2$ и наиболее мощный критерий можно переписать в более простом виде

$$S_2^* = \left\{\sum_{j=1}^{n} W_j \geqslant c'\right\} = \left\{\sum_{j=1}^{n} X_j^2 \geqslant c''\right\},$$

где новая константа c'' связана с константой c' соотношением

$$c'' = 2\left(c' - n(\ln\sigma_0 - \ln\sigma_1)\right)\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)^{-1}.$$

Константу c'' найдем из условия

$$\mathbf{P}_0\left\{\sum_{j=1}^{n} X_j^2 \geqslant c''\right\} = \alpha.$$

Если гипотеза H_0 верна, то случайная величина $\sum\limits_{j=1}^{n} X_j^2/\sigma_0^2$ имеет распределение χ_n^2. Обозначим $G_n(x) = \mathbf{P}\left\{\chi_n^2 < x\right\}$ функцию распределения χ_n^2, а α — квантиль $g(n;\alpha)$ этого распределения.

Выберем $c'' = \sigma_0^2 g(n;1-\alpha)$, тогда вероятность ошибки первого рода равна

$$\mathbf{P}_0(S_2^*) = \alpha,$$

а вероятность ошибки второго рода оказывается равной

$$\beta(S_2^*) = \mathbf{P}_1\left\{\sum_{j=1}^{n} X_j^2 < \sigma_0^2 g(n;1-\alpha)\right\} =$$
$$= \mathbf{P}_1\left\{\frac{\sum_{j=1}^{n} X_j^2}{\sigma_1^2} < \frac{\sigma_0^2}{\sigma_1^2} g(n;1-\alpha)\right\} = G_n\left(\frac{\sigma_0^2}{\sigma_1^2} g(n;1-\alpha)\right).$$

Последняя формула позволяет не только вычислить вероятность ошибки второго рода, но и оценить необходимое число наблюдений, при котором эта вероятность будет меньше заданной величины.

14.4.4. Сложные гипотезы. Равномерно наиболее мощные критерии

Наиболее мощные критерии, полученные в двух предыдущих задачах с помощью леммы Неймана–Пирсона:

$$S_1^* = \left\{\frac{\sum_{j=1}^{n} X_j - na_0}{\sigma\sqrt{n}} \geqslant u_\alpha\right\}$$

и

$$S_2^* = \left\{\frac{\sum_{j=1}^{n} X_j^2}{\sigma_0^2} \geqslant g(n;1-\alpha)\right\}$$

имеют одно общее важное свойство: оба они не зависят от альтернативы и могут быть использованы для решения более сложных задач.

Так, при проверке гипотез о математическом ожидании нормального распределения $N(a;\sigma^2)$ (σ^2 известно)

$$H_0\colon\ a = a_0,$$
$$H_1\colon\ a > a_0,$$

критерий S_1^* является наиболее мощным среди всех критериев уровня α при любой альтернативе $a_1 > a_0$.

Критерии, обладающие таким свойством, называют *равномерно наиболее мощными.*

Функцию мощности критерия S_1^* определяют равенством

$$W(a_1) = \mathbf{P}_{a_1}(S_1^*) = \Phi\left(\sqrt{n}\,\frac{(a_1 - a_0)}{\sigma} - u_\alpha\right), \quad a_1 > a_0.$$

Ее значение увеличивается с удалением альтернативы от гипотезы.

Этот критерий S_1^* можно использовать для проверки сложной гипотезы

$$H_0\colon\ a \leqslant a_0$$

против сложной альтернативы

$$H_1\colon\ a > a_0.$$

Критерий S_1^* является критерием уровня α и для этой задачи, поскольку

$$\mathbf{P}_a(S_1^*) = \Phi\left(\sqrt{n}\,\frac{(a - a_0)}{\sigma} - u_\alpha\right) \leqslant \alpha.$$

Более того, среди всех критериев уровня α критерий S_1^* является равномерно наиболее мощным. Этот критерий обладает еще одним важным свойством — *свойством несмещенности*: $\forall\, a \leqslant a_0$, $\forall\, a_1 > a_0$

$$\mathbf{P}_a(S_1^*) < \mathbf{P}_{a_1}(S_1^*),$$

то есть вероятность отвергуть основную гипотезу всегда меньше вероятности отвергнуть альтернативную гипотезу.

Аналогичными свойствами обладает критерий S_2^* при проверке гипотез, касающихся дисперсии нормальной выборки с математическим ожиданием, равным 0.

Задача 14.2. Покажите, что для нормальной выборки в условиях, когда дисперсия известна и равна σ^2, не существует равномерно наиболее мощного критерия в задаче проверки гипотезы

$$H_0\colon a = a_0$$

против двухсторонней альтернативы

$$H_1\colon a \neq a_0.$$

Замечание. Тем не менее в условиях предыдущей задачи существует равномерно наиболее мощный критерий среди всех несмещенных критериев. Задача построения равномерно наиболее мощного критерия становится еще более сложной, если оба параметра нормального распределения неизвестны. В этом случае гипотеза $H_0\colon a = a_0$ уже не является простой. Подробное изложение теории проверки параметрических гипотез содержится в известной монографии Э. Лемана [19].

14.5. Проверка гипотез о параметре биномиального распределения

Пусть $A_1, A_2, \ldots, A_n$ — взаимно независимые случайные события, имеющие одинаковую вероятность $p = \mathbf{P}(A_i)$. Рассмотрим следующую задачу проверки гипотез:

$$\begin{aligned} &H_0\colon p = p_0, \\ &H_1\colon p = p_1 \quad (p_0 < p_1). \end{aligned}$$

Обозначим $X_j(\omega) = I_{A_j}(\omega)$, $j = 1, \ldots, n$. Эти случайные величины независимы, одинаково распределены. Если верна основная гипотеза H_0, то

$$p_0(x) = \mathbf{P}_0\{X_j = x\} = {p_0}^x (1 - p_0)^{1-x}, \quad x = 0,\, 1.$$

Если же справедлива альтернатива H_1, то

$$p_1(x) = \mathbf{P}_1\left\{X_j = x\right\} = {p_1}^x (1 - p_1)^{1-x}, \quad x = 0,\, 1.$$

Поскольку распределения дискретные, то нерандомизированный критерий существует не при всех значениях α, поэтому будем строить *асимптотический критерий* при $n \to \infty$. По лемме Неймана–Пирсона, наиболее мощный критерий имеет вид

$$S^* = \left\{W_1 + W_2 + \ldots + W_n \geqslant c_\alpha\right\},$$

где

$$W_j = \ln \frac{p_1(X_j)}{p_0(X_j)} = \ln \frac{1 - p_1}{1 - p_0} + X_j \ln \left(\frac{p_1}{(1 - p_1)} \Big/ \frac{p_0}{(1 - p_0)} \right).$$

Простые преобразования позволяют переписать критическое множество в виде

$$S^* = \left\{ \sum_{j=1}^{n} X_j \geqslant c'_\alpha \right\}.$$

Учитывая асимптотическую нормальность $\sum\limits_{j=1}^{n} X_j$ с математическим ожиданием np и стандартным отклонением $\sqrt{npq}$, получим приближенное уравнение для нахождения c'_α

$$\alpha = \mathbf{P}_0 \left\{ \sum_{j=1}^{n} X_j \geqslant c'_\alpha \right\} \approx 1 - \Phi\left(\frac{c'_\alpha - np_0}{\sqrt{np_0(1-p_0)}} \right),$$

откуда находим

$$c'_\alpha = u_\alpha \sqrt{np_0 q_0} + np_0.$$

Вероятность ошибки второго рода тогда равна

$$\mathbf{P}_1 \left\{ \sum_{j=1}^{n} X_j < c'_\alpha \right\} = \mathbf{P}_1 \left\{ \sum_{j=1}^{n} X_j < u_\alpha \sqrt{np_0 q_0} + np_0 \right\}.$$

Чтобы эта вероятность при $n \to \infty$ равнялась β, необходимо выполнение равенства

$$\frac{u_\alpha \sqrt{np_0 q_0} - n(p_1 - p_0)}{\sqrt{np_1 q_1}} = -u_\beta.$$

Из последнего уравнения получим приближенную формулу для числа наблюдений n, при котором достигается заданный уровень ошибок:

$$n \approx \frac{(u_\alpha \sqrt{p_0 q_0} + u_\beta \sqrt{p_1 q_1})^2}{(p_1 - p_0)^2}. \tag{14.10}$$

Так, при $p_0 = 0.5$, $p_1 = 0.5 + 0.01$, $\alpha = 0.01$, $\beta = 0.05$ получим необходимое число наблюдений $n \approx 39\,420$.

Различать близкие между собой гипотезы не так просто. Оценки необходимого числа наблюдений в (14.9), (14.10) показывают, что числители дробей мало чувствительны к изменениям параметров, а порядок роста n определяют знаменатели этих дробей, то есть расстояния между гипотезами.

Глава 15

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

15.1. Постановка задачи и основные определения

Пусть случайные величины $X_1, X_2, \ldots, X_n$ — независимы и одинаково распределены с функцией распределения $F(x) = F(x,\theta)$, зависящей от неизвестного параметра $\theta \in \Theta$. Множество Θ — это множество допустимых значений параметра θ.

Важной задачей математической статистики, как это уже отмечалось ранее, является задача оценивания неизвестных параметров.

Всякую измеримую функцию от наблюдений $T(X_1, \ldots, X_n)$ называют *статистикой*, а статистику, которую используют в качестве приближенного значения неизвестного параметра θ, называют *оценкой* этого параметра. Оценку обычно обозначают той же буквой, что и сам параметр, только с каким-либо значком сверху, например: $\widehat{\theta}$, θ^* и т. д. При изучении асимптотических свойств оценок при $n \to \infty$, когда важно подчеркнуть зависимость оценки от n, будем использовать также обозначения T_n, $\widehat{\theta}_n$ и т. п.

Оценка неизвестного параметра — это функция от наблюдений. Если в эту функцию подставить полученные результаты наблюдений $x_1, \ldots, x_n$, то получим число, или точку в множестве допустимых значений параметра. Такой способ оценивания называют точечным оцениванием, а оценки — точечными оценками.

Определение 15.1. Оценка $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ называется *несмещенной* оценкой параметра θ, если для всех $\theta \in \Theta$ выполняется равенство

$$\mathbf{E}_\theta \widehat{\theta} = \theta.$$

Иногда в этом случае говорят, что оценка лишена систематической ошибки.

Определение 15.2. Оценка $\widehat{\theta}$ называется *состоятельной* для параметра θ, если $\forall \varepsilon > 0$

$$\mathbf{P}_\theta\{|\widehat{\theta} - \theta| > \varepsilon\} \to 0 \quad \text{при} \quad n \to \infty$$

для всех $\theta \in \Theta$.

Другими словами, оценка называется состоятельной для θ, если она *сходится по вероятности* к неизвестному параметру θ.

Более подробно о свойствах точечных оценок и методах их получения мы будем говорить в следующей главе, а пока остановимся на другом способе оценивания — интервальных оценках, или доверительных интервалах, когда по результатам наблюдений указывается не одно приближенное значение параметра, а целое множество приближенных значений.

Пусть $\mathbf{P}_\theta\{T_1(X_1,\ldots,X_n) < T_2(X_1,\ldots,X_n)\} = 1$ для всех $\theta \in \Theta$.

Определение 15.3. Случайный интервал

$$\big(T_1(X_1,\ldots,X_n), T_2(X_1,\ldots,X_n)\big)$$

называется *доверительным интервалом* для неизвестного параметра θ с надежностью $1-\alpha$, если для всех допустимых значений параметра

$$\mathbf{P}_\theta\{T_1(X_1,\ldots,X_n) < \theta < T_2(X_1,\ldots,X_n)\} = 1-\alpha. \tag{15.1}$$

Число $1-\alpha$ называют также *доверительной вероятностью* интервала, обычно его выбирают близким к 1. Таким образом, каково бы ни было значение параметра, интервал (T_1, T_2) накрывает параметр с заданной вероятностью.

Не всегда удается построить доверительный интервал так, чтобы в (15.1) выполнялось равенство[1]. Поэтому в таких случаях будем строить доверительные интервалы, удовлетворяющие неравенству

$$\mathbf{P}_\theta\{T_1(X_1,\ldots,X_n) < \theta < T_2(X_1,\ldots,X_n)\} \geqslant 1-\alpha, \tag{15.2}$$

добиваясь при этом, чтобы вероятность в (15.2) была как можно ближе к $1-\alpha$.

Существует несколько стандартных приемов построения доверительных интервалов. Продемонстрируем их при построении доверительных интервалов для параметров нормального и биномиального распределений.

[1] Это замечание относится прежде всего к дискретным распределениям.

15.2. Доверительный интервал для математического ожидания нормального распределения при известной дисперсии

Пусть случайные величины $X_1, \ldots, X_n$ независимы и имеют одинаковое нормальное распределение $N(a, \sigma_0^2)$. Математическое ожидание неизвестно и для него требуется построить доверительный интервал надежности $1-\alpha$, дисперсию считаем известной и равной σ_0^2.

Для построения доверительного интервала воспользуемся статистикой

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Известно, что $\overline{X} \sim N\left(a, \dfrac{\sigma_0^2}{n}\right)$. С помощью невырожденного линейного преобразования получим новую статистику

$$Z = \frac{\overline{X}-a}{\sigma_0\big/\sqrt{n}} = \frac{\overline{X}-a}{\sigma_0}\sqrt{n},$$

также имеющую нормальное распределение, но с параметрами $(0,1)$. Это распределение не зависит от неизвестного параметра. Поэтому для заданного числа $1-\alpha$ можно указать неслучайный интервал $(c_1(\alpha), c_2(\alpha))$ такой, что

$$\mathbf{P}_a\{c_1(\alpha) < Z < c_2(\alpha)\} = 1-\alpha$$

одновременно для всех $a \in (-\infty, \infty)$.

Такой интервал определен неоднозначно, и обычно при выборе чисел c_1, c_2 используют некоторые дополнительные соображения. Например, можно выбирать интервал наименьшей длины. Для стандартного нормального распределения это симметричный интервал, концы которого определяются равенствами

$$-c_1(\alpha) = c_2(\alpha) = u_{\alpha/2},$$

где $u_{\alpha/2}$ — это решение уравнения $\mathbf{P}\{|Z| \geqslant u_{\alpha/2}\} = \alpha$.

Поскольку

$$|Z| < u_{\alpha/2} \Longleftrightarrow \frac{|\overline{X}-a|}{\sigma_0}\sqrt{n} < u_{\alpha/2} \Longleftrightarrow |\overline{X}-a| < u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \Longleftrightarrow$$
$$\Longleftrightarrow \overline{X} - u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} < a < \overline{X} + u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \qquad (15.3)$$

то полученный интервал (15.3) является искомым доверительным интервалом.

Разобранный способ построения доверительного интервала использовал свойства статистики $Z = \frac{\overline{X} - a}{\sigma_0}\sqrt{n}$:

1) распределение статистики не зависит от неизвестных параметров;

2) статистика Z является монотонной функцией неизвестного параметра a, что позволило решить неравенство $|Z| < u_{\alpha/2}$ относительно неизвестной величины a.

Статистику, обладающую перечисленными свойствами, называют *центральной статистикой*, а описанный выше метод построения доверительного интервала — *методом, использующим центральную статистику.*

15.3. Построение доверительного интервала для дисперсии нормального распределения

Пусть случайные наблюдения $X_1, \ldots, X_n$ независимы, одинаково нормально распределены с параметрами (a_0, σ^2), но теперь неизвестным параметром является дисперсия σ^2, а математическое ожидание известно и равно a_0.

Для построения доверительного интервала воспользуемся статистикой

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a_0)^2,$$

которую можно использовать в качестве оценки для неизвестной дисперсии σ^2. Приведем некоторые свойства этой статистики. Ее можно представить в виде

$$S_0^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i - a_0}{\sigma}\right)^2 = \frac{\sigma^2}{n}\chi_n^2, \tag{15.4}$$

где χ_n^2 обозначает случайную величину, имеющую распределение хи-квадрат с n степенями свободы. Математическое ожидание и дисперсия соответственно равны

$$\mathbf{E}\,S_0^2 = \frac{\sigma^2}{n}\mathbf{E}\,\chi_n^2 = \frac{\sigma^2}{n}\cdot n = \sigma^2,$$

$$\mathbf{D}\,S_0^2 = \frac{\sigma^4}{n^2}\mathbf{D}\,\chi_n^2 = \frac{\sigma^4}{n^2}\cdot 2n = \frac{2\sigma^4}{n} \to 0,$$

и, следовательно, оценка S_0^2 является несмещенной, состоятельной оценкой для неизвестной дисперсии σ^2. Кроме того, из нее легко получить центральную статистику

$$G = \frac{nS_0^2}{\sigma^2} = \chi_n^2.$$

Поскольку эта статистика имеет распределение хи-квадрат с n степенями свободы, не зависящее от неизвестного параметра, то по таблицам данного распределения для заданной надежности $1-\alpha$ можно определить числа $c_1 = c_1(\alpha)$, $c_2 = c_2(\alpha)$ такие, что

$$\mathbf{P}_\sigma\{c_1 < G < c_2\} = 1 - \alpha.$$

Выбор чисел c_1 и c_2 неоднозначен. Можно, например, выбрать их из условий

$$\mathbf{P}\{\chi_n^2 \leqslant c_1\} = \alpha/2, \quad \mathbf{P}\{\chi_n^2 \geqslant c_2\} = \alpha/2.$$

Такой интервал называют центральным.

Обозначим $g(n;\alpha)$ — решение уравнения $\mathbf{P}\{\chi_n^2 \leqslant g(n;\alpha)\} = \alpha$. Такие числа, как известно, называют квантилями данного распределения. Тогда $c_1 = g(n;\alpha/2)$, $c_2 = g(n;1-\alpha/2)$. Решая неравенство относительно неизвестного параметра,

$$c_1 < G < c_2 \Longleftrightarrow c_1 < \frac{nS_0^2}{\sigma^2} < c_2,$$

находим искомый доверительный интервал:

$$\frac{nS_0^2}{g(n;1-\alpha/2)} < \sigma^2 < \frac{nS_0^2}{g(n;\alpha/2)}. \qquad (15.5)$$

Замечание. Если использовать, как и в предыдущем примере, соображения о кратчайшей длине интервала, то получим другие значения констант c_1, c_2.

Задача 15.1. Покажите, что в этом случае c_1, c_2 должны удовлетворять уравнению

$$c_1^2 p_n(c_1) = c_2^2 p_n(c_2),$$

где $p_n(x)$ — плотность распределения хи-квадрат с n степенями свободы.

15.3.1. Совместное распределение статистик $\overline{X}$ и S^2

Пусть по-прежнему $X_1,\ldots,X_n$ — независимые случайные величины с одинаковым нормальным распределением $N(a,\sigma^2)$, но теперь оба параметра этого распределения неизвестны.

Обозначим

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

выборочное среднее и выборочную дисперсию соответственно. Докажем следующую теорему о совместном распределении статистик $\overline{X}$ и S^2 в нормальных выборках.

Теорема 15.1. *Если $X_1,X_2,\ldots,X_n$ — независимая выборка из нормального распределения с параметрами (a,σ^2), то:*

1) *$\overline{X}$ и S^2 независимы;*

2) $\overline{X}\sim N\left(a,\dfrac{\sigma^2}{n}\right)$;

3) $S^2\overset{d}{=}\dfrac{\sigma^2}{n}\chi^2_{n-1}$

Замечание. Здесь и далее равенство $X\overset{d}{=}Y$ означает, что случайные величины X и Y имеют одинаковые распределения.

Доказательство теоремы использует свойства многомерного нормального распределения.

Определение 15.4. Случайный вектор $\boldsymbol{Y}=(Y_1,Y_2,\ldots,Y_n)^T$ имеет n-мерное стандартное нормальное распределение, если совместная плотность распределения равна

$$\widehat{p}_Y(y_1,\ldots,y_n)=\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}y_i^2\right\}.$$

Нетрудно проверить, что компоненты Y_i случайного вектора, имеющего стандартное нормальное распределение, независимы и $Y_i\sim N(0,1)$.

Лемма 15.1. *Пусть случайный вектор $\boldsymbol{Z}$ получен из вектора $\boldsymbol{Y}$ со стандартным нормальным распределением с помощью ортогонального преобразования $\boldsymbol{Z}=\boldsymbol{C}\boldsymbol{Y}$. Тогда $\boldsymbol{Z}$ также имеет стандартное нормальное распределение.*

Доказательство. Напомним, что строки матрицы ортогонального преобразования образуют ортонормированный базис в $\mathbb{R}^n$,

$$\boldsymbol{C}^{-1} = \boldsymbol{C}^T, \quad \det \boldsymbol{C} = 1$$

и что ортогональное преобразование сумму квадратов переменных переводит в сумму квадратов новых переменных.

Пусть A — событие из борелевской σ-алгебры в $\mathbb{R}^n$, а событие $B = \boldsymbol{C}^{-1}(A)$ — прообраз события A при преобразовании $\boldsymbol{C}$. Тогда

$$\mathbf{P}\{\boldsymbol{Z} \in A\} = \mathbf{P}\{\boldsymbol{C}\boldsymbol{Y} \in A\} = \int_B \widehat{p}_Y(y_1, \ldots, y_n)\, dy_1 \cdots dy_n = \\ = \int_B \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n y_i^2\right) dy_1 \cdots dy_n.$$

В последнем интеграле сделаем замену переменных $\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y}$, тогда

$$\mathbf{P}\{\boldsymbol{Z} \in A\} = \int_A \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n z_i^2\right) dz_1 \cdots dz_n.$$

Подынтегральная функция по определению является плотностью распределения случайного вектора $\boldsymbol{Z}$. Следовательно, распределение $\boldsymbol{Z}$ также является стандартным нормальным. □

Вернемся к доказательству теоремы.

Доказательство. Утверждение 2 теоремы почти очевидно, поэтому докажем только утверждения 1, 3. Наряду с исходными случайными величинами X_i рассмотрим

$$Y_i = \frac{X_i - a}{\sigma} \quad (i = 1, \ldots, n).$$

Отметим, что случайный вектор $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ имеет стандартное нормальное распределение. Рассмотрим ортогональную матрицу

$$\boldsymbol{C} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix},$$

у которой первая строка выбрана специальным образом, а остальные произвольным образом дополняют первую строку до ортогонального базиса в $\mathbb{R}^n$.

В силу леммы $\boldsymbol{Z} = \boldsymbol{CY}$ также имеет стандартное нормальное распределение, поэтому компоненты вектора $Z_i \sim N(0,1)$ и независимы. При этом

$$\overline{X} = a + \sigma\overline{Y} = a + \frac{\sigma}{\sqrt{n}} Z_1,$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2\right) =$$

$$= \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2 - \left(\frac{1}{\sqrt{n}}Z_1\right)^2\right) = \frac{\sigma^2}{n}\sum_{i=2}^{n} Z_i^2.$$

Мы видим, что выборочное среднее $\overline{X}$ зависит лишь от Z_1, а выборочная дисперсия S^2 — от не зависящих от Z_1 случайных величин $Z_2, \ldots, Z_n$. Следовательно, $\overline{X}$ и S^2 — независимые случайные величины.

Кроме того,

$$S^2 \stackrel{d}{=} \frac{\sigma^2}{n} \cdot \chi^2_{n-1},$$

так как по определению случайная величина, представленная в виде суммы квадратов независимых случайных величин со стандартным нормальным распределением, имеет распределение хи-квадрат с числом степеней свободы, равным числу независимых слагаемых. Теорема доказана. □

Воспользуемся доказанными свойствами статистик $\overline{X}$ и S^2 при построении доверительных интервалов для неизвестных математического ожидания a и дисперсии σ^2, когда оба параметра неизвестны.

Доверительный интервал для σ^2 строится с помощью центральной статистики

$$G = \frac{nS^2}{\sigma^2} \stackrel{d}{=} \chi^2_{n-1},$$

которая имеет распределение хи-квадрат с $n-1$ степенью свободы. Построение полностью повторяет рассуждения предыдущего раздела, поэтому сразу выпишем интервал с надежностью $1-\alpha$:

$$\frac{nS^2}{g(n-1;1-\alpha/2)} < \sigma^2 < \frac{nS^2}{g(n-1;\alpha/2)}, \tag{15.6}$$

где $g(n-1;\alpha)$ обозначает α-квантиль распределения хи-квадрат.

Замечание. Проводя простые вычисления, можно получить, что

$$\mathbf{E}S^2 = \frac{\sigma^2}{n} \cdot \mathbf{E}\chi^2_{n-1} = \frac{n-1}{n} \cdot \sigma^2,$$

то есть выборочная дисперсия не является несмещенной оценкой для неизвестной дисперсии. Используя эту оценку при небольшом количестве проведенных наблюдений n, мы будем получать в среднем числа, несколько меньшие действительного значения σ^2.

Но это легко исправить, если рассмотреть другую оценку

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

которую часто и называют исправленной выборочной дисперсией. Впрочем, асимптотические свойства этих двух оценок совпадают.

Задача 15.2. Докажите, что S^2 и s^2 являются состоятельными оценками для неизвестной дисперсии для любой независимой выборки из распределения с конечной дисперсией.

Доверительный интервал (15.6) можно переписать в виде

$$\frac{(n-1)s^2}{g(n-1;1-\alpha/2)} < \sigma^2 < \frac{(n-1)s^2}{g(n-1;\alpha/2)}.$$

15.3.2. Распределение Стьюдента

При построении доверительного интервала для a мы не можем, как ранее в разделе 15.2, использовать статистику $Z = \frac{(\overline{X}-a)\sqrt{n}}{\sigma}$, поскольку интервал будет зависеть от второго неизвестного параметра σ. Английский математик У. С. Госсет, публиковавший свои труды под псевдонимом Student, предложил центральную статистику

$$t_{n-1} = \frac{(\overline{X}-a)\sqrt{n}}{s}.$$

Покажем, что ее распределение не зависит от неизвестных параметров. Действительно,

$$t_{n-1} = \frac{\dfrac{(\overline{X}-a)\sqrt{n}}{\sigma}}{\dfrac{s}{\sigma}} = \frac{Z}{\sqrt{\dfrac{1}{n-1}\chi^2_{n-1}}}$$

представляется в виде дроби, числитель и знаменатель которой — независимые случайные величины, причем числитель имеет стандартное нормальное распределение, а знаменатель представим в виде $\sqrt{\dfrac{1}{n-1}\chi^2_{n-1}}$. Статистику t_{n-1} называют *статистикой Стьюдента*, а ее распределение — *распределением Стьюдента* с $n-1$ степенью свободы.

Замечание. Приведем формулу плотности для распределения Стьюдента с m степенями свободы:

$$p(x) = \frac{\Gamma((m+1)/2)}{\sqrt{\pi m}\,\Gamma(m/2)}(1+x^2/m)^{-(m+1)/2}.$$

В частности, для $m=1$ это распределение известно как распределение Коши.

Математическое ожидание существует при $m>1$ и равно 0, дисперсия существует при $m>2$ и равна $m/(m-2)$.

По таблицам распределения Стьюдента для заданной надежности $1-\alpha$ определим $t(n-1;\alpha/2)$, являющееся решением уравнения

$$\mathbf{P}\{|t_{n-1}| > t(n-1;\alpha/2)\} = \alpha, \quad \text{или} \quad \mathbf{P}\{t_{n-1} > t(n-1;\alpha/2)\} = \alpha/2.$$

Поскольку с вероятностью $1-\alpha$ выполняются соотношения

$$|t_{n-1}| < t(n-1;\alpha/2) \Longleftrightarrow \left|\frac{(\overline{X}-a)\sqrt{n}}{s}\right| < t(n-1;\alpha/2) \Longleftrightarrow$$
$$\Longleftrightarrow \overline{X} - \frac{t(n-1;\alpha/2)s}{\sqrt{n}} < a < \overline{X} + \frac{t(n-1;\alpha/2)s}{\sqrt{n}}, \quad (15.7)$$

то (15.7) представляет собой доверительный интервал кратчайшей длины для a с надежностью $1-\alpha$.

15.4. Асимптотический доверительный интервал для параметра p биномиального распределения

На следующем примере проиллюстрируем способ построения асимптотического доверительного интервала, который можно использовать только для достаточно больших выборок.

Пусть проведено n независимых испытаний Бернулли с вероятностью успеха p в отдельном испытании,

$$X_i = \begin{cases} 1, & \text{если испытание с номером } i \text{ прошло успешно,} \\ 0, & \text{если результат испытания } i \text{ — неуспех,} \end{cases}$$

$i = 1, 2, \ldots, n$. Построим асимптотический доверительный интервал для p.

Пусть $\mu_n = \sum\limits_{i=1}^{n} X_i$ — число успехов в n испытаниях. Известно, что относительная частота наступления успеха $\frac{\mu_n}{n}$ асимптотически нормальна с математическим ожиданием p и стандартным отклонением $\sqrt{\frac{p(1-p)}{n}}$. Это означает, что для всех x

$$\mathbf{P}\left\{ \frac{\frac{\mu_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} < x \right\} \to \Phi(x).$$

Существует несколько способов построения асимптотических доверительных интервалов.

Мы сначала рассмотрим способ, основанный на преобразовании, стабилизирующем дисперсию: ищется такое преобразование относительной частоты, в результате которого также получается асимптотически нормальная случайная величина, но с дисперсией, не зависящей от параметра p.

Для этого воспользуемся теоремой 12.5. Напомним ее содержание: если последовательность случайных величин Y_n асимптотически нормальна с параметрами (a, b_n), причем $b_n \to 0$, функция $g(x)$ дважды непрерывно дифференцируема, то последовательность $W_n = g(Y_n)$ также асимптотически нормальна, но с параметрами $(g(a), b_n|g'(a)|)$.

Рассмотрим функцию $g(x) = 2\arcsin\sqrt{x}$, $g'(x) = \dfrac{1}{\sqrt{x(1-x)}}$.

Применяя теорему, получим, что последовательность

$$\eta_n = 2\arcsin\sqrt{\frac{\mu_n}{n}}$$

асимптотически нормальна с параметрами $\left(2\arcsin\sqrt{p}; \frac{1}{\sqrt{n}}\right)$. Тогда при достаточно больших n (хотя бы несколько десятков) выполняется приближенное равенство

$$\mathbf{P}\left\{\left|2\arcsin\sqrt{\frac{\mu_n}{n}} - 2\arcsin\sqrt{p}\right| < \frac{u_{\alpha/2}}{\sqrt{n}}\right\} \approx 1-\alpha.$$

Решив неравенство в фигурных скобках относительно параметра p, получим доверительный интервал

$$\sin^2\left(\arcsin\sqrt{\frac{\mu_n}{n}} - \frac{u_{\alpha/2}}{2\sqrt{n}}\right) < p < \sin^2\left(\arcsin\sqrt{\frac{\mu_n}{n}} + \frac{u_{\alpha/2}}{2\sqrt{n}}\right). \quad (15.8)$$

Задача 15.3. Докажите, что последовательность случайных величин

$$V_n = \frac{(\mu_n/n - p)\sqrt{n}}{\sqrt{\mu_n/n(1-\mu_n/n)}}$$

асимптотически нормальна с параметрами $(0,1)$.

Используя этот факт, можно построить еще один асимптотический доверительный интервал с надежностью $1-\alpha$

$$\frac{\mu_n}{n} - u_{\alpha/2}\frac{\sqrt{\frac{\mu_n}{n}\left(1-\frac{\mu_n}{n}\right)}}{\sqrt{n}} < p < \frac{\mu_n}{n} + u_{\alpha/2}\frac{\sqrt{\frac{\mu_n}{n}\left(1-\frac{\mu_n}{n}\right)}}{\sqrt{n}}. \quad (15.9)$$

Доверительные интервалы (15.8), (15.9) при достаточно больших n не слишком сильно отличаются один от другого. Так, при $n=65$, $\mu_n=30$, $1-\alpha=0.95$ и тот и другой дают при подстановке данных результат

$$0.34 < p < 0.58,$$

тогда как точный доверительный интервал, вычисленный по таблицам Л. Н. Большева, Н. В. Смирнова [3]:

$$0.337 < p < 0.590.$$

Глава 16

ТОЧЕЧНЫЕ ОЦЕНКИ ДЛЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ

16.1. Сравнение свойств несмещенных оценок

Пример 16.1. По регистрационным номерам танков во время Второй мировой войны оценивался объем производства военной техники. Задача сводится к оценке параметра равномерного распределения.

Пусть $X_1, X_2, \ldots, X_n$ — независимые случайные величины с плотностью распределения

$$p(x;\theta) = \begin{cases} 1/\theta, & x \in [0,\theta], \\ 0, & x \notin [0,\theta]. \end{cases}$$

Введем новые случайные величины $X_j' = \dfrac{X_j}{\theta}$, которые распределены равномерно на отрезке $[0,1]$ и имеют плотность распределения

$$p_1(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases}$$

Для этих случайных величин

$$\mathbf{E}X_j' = \int_0^1 x\,dx = \frac{1}{2}; \quad \mathbf{E}(X_j')^2 = \int_0^1 x^2\,dx = \frac{1}{3}; \quad \mathbf{D}X_j' = \frac{1}{12}.$$

Пусть

$$Y = \max_{1 \leqslant j \leqslant n} X_j'.$$

Для случайной величины Y функция распределения и плотность при $x \in [0,1]$ соответственно равны

$$F_Y(x) = \mathbf{P}\{Y < x\} = \mathbf{P}^n(X_j' < x) = x^n, \quad p_Y(x) = nx^{n-1}.$$

Следовательно,

$$\mathbf{E}\,Y = \int_0^1 nx^n\,dx = \frac{n}{n+1},$$

$$\mathbf{D}\,Y = \mathbf{E}\,Y^2 - (\mathbf{E}\,Y)^2 = \frac{n}{(n+2)(n+1)^2}.$$

В качестве оценок для параметра θ равномерного распределения рассмотрим статистики

$$\theta_1^* = 2\overline{X}, \quad \theta_2^* = \frac{n+1}{n}\max_{1\leqslant j\leqslant n} X_j$$

и сравним их свойства.

Обе оценки являются несмещенными, так как для всех $\theta > 0$

$$\mathbf{E}_\theta\,\theta_1^* = 2\theta\,\mathbf{E}\,\frac{X_1' + \ldots + X_n'}{n} = \theta,$$

$$\mathbf{E}_\theta\,\theta_2^* = \theta\frac{n+1}{n}\,\mathbf{E}_\theta\,Y = \theta.$$

Обе оценки являются состоятельными, поскольку их дисперсии стремятся к 0 при $n\to\infty$. Действительно,

$$\mathbf{D}_\theta\,\theta_1^* = \frac{4\theta^2}{n^2}\cdot\frac{n}{12} = \frac{\theta^2}{3n},$$

$$\mathbf{D}_\theta\,\theta_2^* = \theta^2\,\mathbf{D}_\theta\left(\frac{n+1}{n}Y\right) = \frac{\theta^2}{n(n+2)}.$$

Однако при всех $n\geqslant 2$ и при всех θ одновременно дисперсия второй оценки меньше дисперсии первой оценки, при $n=1$ эти оценки просто совпадают. В таких ситуациях будем говорить, что θ_2^* имеет дисперсию *равномерно меньшую*, чем θ_1^*. А это означает, что вторая оценка предпочтительнее первой (особенно при больших n), так как меньше отклоняется от неизвестного параметра.

В общем случае поступают аналогичным образом: среди двух несмещенных оценок выбирают ту, дисперсия которой меньше. Если для параметра θ существует несмещенная оценка θ^*, имеющая равномерно наименьшую дисперсию среди всех несмещенных оценок, то такую оценку называют *оптимальной*.

В настоящей книге мы не будем рассматривать теорию и методы построения оптимальных оценок. Желающие могут прочитать об этом, например, в [12; 13].

16.2. Семейства распределений

Рассмотрим два основных типа параметрических семейств распределений.

I тип. Все распределения семейства получаются из некоторого фиксированного распределения путем преобразований, зависящих от неизвестного параметра θ:

а) θ — параметр сдвига, в этом случае

$$p(x;\theta) = p(x-\theta), \quad \theta \in \Theta = \mathbb{R};$$

б) θ — параметр масштаба, в этом случае

$$p(x;\theta) = \frac{1}{\theta} p\left(\frac{x}{\theta}\right), \quad \theta > 0;$$

в) присутствуют и параметр сдвига, и параметр масштаба, $\theta = (\theta_1, \theta_2)$:

$$p(x;\theta_1,\theta_2) = \frac{1}{\theta_2} p\left(\frac{x-\theta_1}{\theta_2}\right), \quad \theta_1 \in \mathbb{R}, \quad \theta_2 > 0.$$

Так получено семейство распределений примера 16.1, так получается семейство нормальных распределений на прямой.

II тип. Экспоненциальные семейства распределений. Плотность распределения (или совместная вероятность) имеет вид

$$p(x;\theta_1,\ldots,\theta_k) = h(x)\cdot c(\theta_1,\ldots,\theta_k)\cdot \exp\left(\sum_{i=1}^{k} \theta_i T_i(x)\right),$$

где $\theta = (\theta_1,\ldots,\theta_k)$ — k-мерный параметр.

Примерами экспоненциальных семейств распределений являются распределение Пуассона, биномиальное, нормальное, показательное распределения.

Пусть $X_1,\ldots,X_n$ — независимые одинаково распределенные случайные величины с плотностью распределения вероятностей $p(x;\theta)$. Совместная плотность распределения равна

$$\widehat{p}(x_1,\ldots,x_n;\theta) = p(x_1;\theta)\cdots p(x_n;\theta).$$

Для дискретных случайных величин $\widehat{p}(x_1,\ldots,x_n)$ будет обозначать совместную вероятность

$$\widehat{p}(x_1,\ldots,x_n;\theta) = \mathbf{P}_\theta\{X_1 = x_1,\ldots,X_n = x_n\}.$$

16.3. Метод максимального правдоподобия

Наиболее обоснованным и распространенным методом отыскания оценок является метод максимального правдоподобия.

Функция $\widehat{p}(x_1,\ldots,x_n;\theta)$, которую рассматривают как функцию параметра θ, считая при этом остальные переменные фиксированными, называется *функцией правдоподобия.* Чтобы подчеркнуть зависимость от параметра, часто функцию прадоподобия обозначают короче:

$$\widehat{p}(\theta)=\widehat{p}(x_1,\ldots,x_n;\theta).$$

Мы также будем рассматривать случайную величину, полученную из функции $\widehat{p}(x_1,\ldots,x_n;\theta)$ заменой ее аргументов $x_1,\ldots$ $\ldots,x_n$ на случайные величины $X_1,\ldots,X_n$. Обычно ее обозначают $L(\theta)$ или $L(\theta;X_1,\ldots,X_n)$:

$$L(\theta)=L(\theta;X_1,\ldots,X_n)=\widehat{p}(X_1,\ldots,X_n;\theta)$$

и также называют *функцией правдоподобия.* В дальнейшем мы будем рассматривать эту функцию и как функцию от θ (в том числе рассматривать производные функции правдоподобия по параметру θ), и как случайную величину (в частности, вычислять числовые характеристики распределения $L(\theta;X_1,\ldots$ $\ldots,X_n)$ при фиксированном значении параметра θ).

Определение 16.1. *Оценкой максимального правдоподобия* для неизвестного параметра θ называется точка максимума функции правдоподобия $L(\theta)$ по переменной θ.

Смысл оценки максимального правдоподобия состоит в следующем. Если в результате наблюдений $X_1=x_1,\ldots,X_n=x_n$, то в качестве оценки неизвестного параметра выбирается $\widehat{\theta}=$ $=\widehat{\theta}(x_1,\ldots,x_n)$, при котором

$$\widehat{p}(x_1,\ldots,x_n;\widehat{\theta})=\max_{\theta}\widehat{p}(x_1,\ldots,x_n;\theta),$$

то есть выбирается такое значение θ, при котором наблюдаемый результат наиболее вероятен.

Разберем несколько примеров построения оценок максимального правдоподобия.

Пример 16.2. По результатам n наблюдений над случайной величиной, распределенной по закону Пуассона, методом максимального правдоподобия оценим параметр λ этого распределения.

Пусть $X_1, \ldots, X_n$ — случайная выборка с распределением Пуассона,

$$\mathbf{P}\{X_i = k\} = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots .$$

Совместная вероятность этих величин равна

$$\mathbf{P}\{X_1{=}x_1, \ldots, X_n{=}x_n\} = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}\left(\prod_{i=1}^{n} x_i!\right)^{-1}.$$

Следовательно, функция правдоподобия равна

$$L(\lambda) = e^{-n\lambda}\lambda^{\sum_{i=1}^n X_i}\left(\prod_{i=1}^{n} X_i!\right)^{-1}.$$

Поскольку функции $L(\lambda)$ и $\ln L(\lambda)$ связаны одна с другой монотонным преобразованием, то точки максимума у них совпадают, но проще найти точку максимума для второй:

$$\ln L(\lambda) = -n\lambda + \ln\lambda \cdot \sum_{i=1}^{n} X_i + \ln\left(\prod_{i=1}^{n} X_i!\right)^{-1},$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i.$$

Поскольку первая производная убывает, то точка, в которой она обращается в 0, и будет точкой максимума. Таким образом, оценка максимального правдоподобия для параметра распределения Пуассона равна

$$\widehat{\lambda} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Пример 16.3. Пусть $X_1, \ldots, X_n$ — случайная выборка с показательным законом распределения, плотность которого равна

$$p(x;\theta) = \begin{cases} \theta e^{-\theta x} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x < 0. \end{cases}$$

Функция правдоподобия равна

$$L(\theta) = \prod_{i=1}^{n} \theta e^{-\theta X_i} = \theta^n e^{-\theta \sum_{i=1}^{n} X_i},$$

$$\ln L(\theta) = n \ln \theta - \theta \sum_{i=1}^{n} X_i,$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} X_i.$$

Приравняв производную к 0, найдем оценку максимального правдоподобия для параметра θ

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}.$$

Максимум функции правдоподобия не всегда достигается во внутренней точке области определения функции правдоподобия. Проиллюстрируем сказанное на следующем примере.

Пример 16.4. Пусть $X_1, \ldots, X_n$ — случайная выборка с равномерным на $[0, \theta]$ распределением, $\theta > 0$ — неизвестный параметр. Плотность равномерного распределения задается формулой

$$p(x; \theta) = \begin{cases} \dfrac{1}{\theta} & \text{при} \quad x \in [0, \theta], \\ 0 & \text{при} \quad x \notin [0, \theta]. \end{cases}$$

Выпишем совместную плотность распределения случайных величин $X_1, \ldots, X_n$:

$$\widehat{p}(x_1, \ldots, x_n; \theta) = \begin{cases} \left(\dfrac{1}{\theta}\right)^n, & \text{если } x_i \in [0, \theta] \text{ для всех } i, \\ 0, & \text{если иначе.} \end{cases}$$

Функция правдоподобия

$$L(\theta) = \begin{cases} \left(\dfrac{1}{\theta}\right)^n, & \text{если} \quad \theta \geqslant \max\limits_{1 \leqslant i \leqslant n} X_i, \\ 0, & \text{если} \quad \theta < \max\limits_{1 \leqslant i \leqslant n} X_i, \end{cases}$$

является монотонно убывающей по θ функцией на $\left[\max\limits_{1 \leqslant i \leqslant n} X_i, \infty\right)$, а в остальных точках равна 0. Значит, ее максимум достигается в точке $\theta = \max\limits_{1 \leqslant i \leqslant n} X_i$, и оценкой максимального правдоподобия является

$$\widehat{\theta} = \max_{1 \leqslant i \leqslant n} X_i.$$

Замечание. Отметим здесь, что несмещенная оценка θ_2^* примера 16.1 получена из оценки максимального правдоподобия

$$\theta_2^* = \frac{n+1}{n} \max_{1 \leqslant i \leqslant n} X_i$$

и уже не улучшаема в том смысле, что среди всех несмещенных оценок имеет наименьшую дисперсию одновременно при всех значениях параметра θ.

Оценка максимального правдоподобия не обязательно единственна.

Задача 16.1. Покажите, что для параметра θ равномерного распределения на отрезке $[\theta, \theta+1]$ существует бесконечное число оценок максимального правдоподобия. Выберите среди них несмещенную.

Приведем еще один пример построения оценки максимального правдоподобия в случае двух неизвестных параметров.

Пример 16.5. Пусть $X_1, \ldots, X_n$ — случайная выборка из нормального распределения с неизвестным математическим ожиданием a и неизвестной дисперсией σ^2. Функция правдоподобия в этом случае равна

$$L(a, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_i - a)^2}{2\sigma^2}\right] = \\ = \left(\sqrt{2\pi}\right)^{-n} \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - a)^2\right],$$

а ее логарифм

$$\ln L(a, \sigma) = \ln\left(\sqrt{2\pi}\right)^{-n} - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - a)^2.$$

Необходимым условием экстремума функции двух переменных является равенство 0 ее частных производных:

$$\frac{\partial \ln L(a, \sigma)}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - a) = 0,$$

$$\frac{\partial \ln L(a,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i - a)^2 = 0.$$

Решением этой системы уравнений максимального правдоподобия являются

$$\widehat{a} = \overline{X}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Чтобы убедиться, что $(\widehat{a}, \widehat{\sigma}^2)$ — точка максимума, надо вычислить вторые частные производные и показать, что матрица второго дифференциала в этой точке отрицательно определена. Действительно,

$$\frac{\partial^2 \ln L(a,\sigma)}{\partial a^2} = -\frac{n}{\sigma^2},$$

$$\frac{\partial^2 \ln L(a,\sigma)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(X_i - a)^2,$$

$$\frac{\partial^2 \ln L(a,\sigma)}{\partial a \partial \sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(X_i - a),$$

матрица второго дифференциала равна

$$\begin{pmatrix} -\frac{n}{\widehat{\sigma}^2} & 0 \\ 0 & -\frac{2n}{\widehat{\sigma}^2} \end{pmatrix}.$$

Эта матрица отрицательно определена, что и доказывает, что $(\widehat{a}, \widehat{\sigma}^2)$ — точка максимума функции правдоподобия.

Оценки максимального правдоподобия обладают хорошими свойствами, часто они оказываются состоятельными, асимптотически нормальными и асимптотически эффективными. Подробнее о свойствах оценок максимального правдоподобия можно прочитать в [12; 20]. Об эффективности и асимптотической эффективности речь пойдет в следующем разделе.

16.4. Неравенство Рао–Крамера

Будем рассматривать только случай одномерного параметра θ. Пусть по-прежнему $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ — n-мерный случайный вектор, а полученное в результате наблюдений значение этого вектора — $\boldsymbol{x} = (x_1, \ldots, x_n)^T$. Совместное распределение компонент вектора наблюдений имеет функцию правдоподобия $L(\theta) = \widehat{p}(\boldsymbol{x};\theta)$, и для этого семейства выполняются условия регулярности. Сформулируем их для абсолютно непрерывных распределений, когда $\widehat{p}(\boldsymbol{x};\theta)$ — совместная плотность распределения.

1. Множество $\{\boldsymbol{x}\colon \widehat{p}(\boldsymbol{x};\theta) > 0\}$ не зависит от параметра θ.

2. Можно дифференцировать по θ под знаком интеграла в равенстве

$$\int\limits_{\mathbb{R}^n} \widehat{p}(\boldsymbol{x};\theta)\, d\boldsymbol{x} = 1.$$

3. Можно дифференцировать по θ под знаком интеграла в равенстве

$$\int\limits_{\mathbb{R}^n} \theta^*(\boldsymbol{x})\widehat{p}(\boldsymbol{x};\theta)\, d\boldsymbol{x} = \theta,$$

где $\theta^*(\boldsymbol{x})$ — произвольная несмещенная оценка параметра θ.

4. Можно дважды продифференцировать по θ под знаком интеграла в равенстве

$$\int\limits_{\mathbb{R}^n} \widehat{p}(\boldsymbol{x};\theta)\, d\boldsymbol{x} = 1.$$

Для дискретных распределений условия регулярности сводятся к возможности провести почленное дифференцирование в соответствующих рядах.

Теорема 16.1 (неравенство Рао–Крамера). *Пусть для семейства распределений с функцией правдоподобия $\widehat{p}(\boldsymbol{x};\theta)$ выполнены условия регулярности* 1–3, *θ^* — несмещенная оценка с конечной дисперсией параметра θ. Тогда*

$$\mathbf{D}_\theta\, \theta^* \geqslant \frac{1}{\mathbf{E}_\theta \left(\dfrac{\partial \ln L(\theta;\boldsymbol{X})}{\partial\theta}\right)^2}. \qquad (16.1)$$

Неравенство (16.1) также называют *неравенством информации*, поскольку стоящую в знаменателе величину

$$\mathbf{E}_\theta \left(\frac{\partial \ln L(\theta;\boldsymbol{X})}{\partial\theta}\right)^2$$

называют содержащимся в выборке *количеством информации* о параметре θ. Будем обозначать ее $i_n(\theta)$. Тогда неравенство информации можно переписать в более простом виде

$$\mathbf{D}_\theta \theta^* \geqslant \frac{1}{i_n(\theta)}.$$

Согласно этому неравенству дисперсия несмещенной оценки параметра θ не может быть сколь угодно малой. В неравенстве Рао–Крамера указывается нижняя граница для дисперсии несмещенной оценки.

Определение 16.2. Если дисперсия несмещенной оценки параметра θ равна при всех θ нижней границе в неравенстве Рао–Крамера (16.1), то оценка называется *эффективной*.

Поясним понятие эффективности. Мы сравниваем точность несмещенных оценок по их дисперсиям. Пусть θ_1^* и θ_2^* — две несмещенные оценки. Поскольку их дисперсии являются функциями, зависящими от параметра, то может оказаться, что для одних значений параметра предпочтительнее одна оценка, а для других — другая. Такие две оценки будем считать несравнимыми.

Эффективная оценка, если существует, имеет дисперсию, которая одновременно для всех значений параметра меньше (или равна) дисперсии любой другой оценки. Следовательно, в этом смысле она предпочтительнее всех остальных[1].

Доказательство. Формальное доказательство неравенства информации достаточно просто, оно использует неравенство Коши–Буняковского, которое в дискретном случае сводится к тому, что скалярное произведение двух векторов по абсолютной величине не превосходит произведения длин этих векторов:

$$\left(\sum_{j=1}^{m} a_j b_j\right)^2 \leqslant \left(\sum_{j=1}^{m} a_j^2\right) \cdot \left(\sum_{j=1}^{n} b_j^2\right),$$

причем неравенство обращается в равенство тогда и только тогда, когда существуют числа λ, μ ($\lambda^2+\mu^2>0$) такие, что

$$\lambda a_j + \mu b_j = 0, \quad j = 1, 2, \ldots, m.$$

В непрерывном случае неравенство Коши–Буняковского при-

[1] Оценки можно сравнивать и по другим критериям.

нимает вид

$$\left(\int f(x)g(x)\,dx\right)^2 \leqslant \left(\int f^2(x)\,dx\right)\left(\int g^2(x)\,dx\right)$$

для любых функций $f(x), g(x) \in \boldsymbol{L}^2$. Для выполнения в этом неравенстве равенства необходимо и достаточно, чтобы существовали числа λ, μ ($\lambda^2+\mu^2>0$) такие, что

$$\lambda f(x)+\mu g(x)=0 \quad \text{почти всюду по } x.$$

Перейдем к непосредственному доказательству неравенства информации. Доказательство проведем для случая, когда случайный вектор $\boldsymbol{X}$ имеет абсолютно непрерывное распределение и $\widehat{p}(\boldsymbol{x};\theta)$ — плотность совместного распределения компонент этого вектора.

Поскольку тождественно по всем θ выполняются равенства

$$\int\limits_{\mathbb{R}^n} \widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}=1,$$

$$\int\limits_{\mathbb{R}^n} \theta^*(\boldsymbol{x})\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}=\theta,$$

то, продифференцировав их по θ, получим

$$\int\limits_{\mathbb{R}^n} \frac{\partial \ln \widehat{p}(\boldsymbol{x};\theta)}{\partial\theta}\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}=0, \tag{16.2}$$

$$\int\limits_{\mathbb{R}^n} \theta^*(\boldsymbol{x})\frac{\partial \ln \widehat{p}(\boldsymbol{x};\theta)}{\partial\theta}\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}=1. \tag{16.3}$$

При дифференцировании мы воспользовались тем, что $\frac{df}{dx} = \frac{d\ln f(x)}{dx}f(x)$.

Умножим (16.2) на θ и вычтем из (16.3), получим

$$\int\limits_{\mathbb{R}^n} (\theta^*(\boldsymbol{x})-\theta)\frac{\partial \ln \widehat{p}(\boldsymbol{x};\theta)}{\partial\theta}\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}=1. \tag{16.4}$$

Обозначим

$$f(\boldsymbol{x})=(\theta^*(\boldsymbol{x})-\theta)\sqrt{\widehat{p}(\boldsymbol{x};\theta)},$$

$$g(\boldsymbol{x}) = \frac{\partial \ln \widehat{p}(\boldsymbol{x};\theta)}{\partial \theta}\sqrt{\widehat{p}(\boldsymbol{x};\theta)}.$$

Тогда, применяя к (16.4) неравенство Коши–Буняковского, имеем

$$1 = \left(\int\limits_{\mathbb{R}^n} f(\boldsymbol{x})g(\boldsymbol{x})\,d\boldsymbol{x}\right)^2 \leqslant \int\limits_{\mathbb{R}^n} f^2(\boldsymbol{x})\,d\boldsymbol{x}\cdot\int\limits_{\mathbb{R}^n} g^2(\boldsymbol{x})\,d\boldsymbol{x} =$$
$$= \int\limits_{\mathbb{R}^n} (\theta^* - \theta)^2\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}\cdot\int\limits_{\mathbb{R}^n}\left(\frac{\partial \ln \widehat{p}(\boldsymbol{x};\theta)}{\partial \theta}\right)^2\widehat{p}(\boldsymbol{x};\theta)\,d\boldsymbol{x}.$$

Следовательно,

$$1 \leqslant \mathbf{D}_\theta\,\theta^*(\boldsymbol{X})\cdot\mathbf{E}_\theta\left(\frac{\partial \ln L(\theta;\boldsymbol{X})}{\partial \theta}\right)^2,$$

откуда получаем

$$\mathbf{D}_\theta\,\theta^* \geqslant \frac{1}{\mathbf{E}_\theta\left(\dfrac{\partial \ln L(\theta;\boldsymbol{X})}{\partial \theta}\right)^2}. \tag{16.5}$$

□

Следствие 16.1. *Если случайные величины $X_1,\ldots,X_n$ независимы и одинаково распределены с плотностью $p(x;\theta)$, то неравенство информации принимает вид*

$$\mathbf{D}_\theta\,\theta^* \geqslant \frac{1}{ni_1(\theta)}, \tag{16.6}$$

где $i_1(\theta) = \mathbf{E}_\theta\left(\dfrac{\partial \ln p(X_1;\theta)}{\partial \theta}\right)^2$ — количество информации о параметре, содержащееся в одном наблюдении.

Доказательство. Заметим, что в доказательстве теоремы нигде не использовались независимость и одинаковая распределенность наблюдений. Пусть теперь $\widehat{p}(\boldsymbol{x};\theta) = p(x_1;\theta)\cdots p(x_n;\theta)$. Тогда

$$\frac{\partial \ln L(\theta;\boldsymbol{X})}{\partial \theta} = \sum_{i=1}^n V_i,$$

где $V_i = \dfrac{\partial \ln p(X_i;\theta)}{\partial \theta}$ — независимые одинаково распределенные случайные величины, для которых из (16.2) следует, что

$\mathbf{E}_\theta V_i = 0$. Но тогда

$$\mathbf{E}_\theta \left(\frac{\partial \ln L(\theta; \boldsymbol{X})}{\partial \theta}\right)^2 = \mathbf{D}_\theta \sum_{i=1}^{n} V_i = n\,\mathbf{E}_\theta V_1^2 \tag{16.7}$$

и мы приходим к неравенству Рао–Крамера (16.6). □

Замечание. В примере 16.1 оценка θ_2^* имеет дисперсию, порядок убывания которой с ростом n равен $\frac{1}{n^2}$. Это пример «сверхэффективной» оценки. Дело в том, что для семейства равномерных распределений на $[0, \theta]$ не выполняются условия регулярности теоремы.

Из доказательства теоремы следует, что необходимым и достаточным условием достижения равенства в неравенстве информации является существование λ и μ, не зависящих от $\boldsymbol{x}$, но, вообще говоря, зависящих от θ, таких, что с вероятностью 1 для всех θ выполняется соотношение

$$\lambda(\theta^* - \theta) = \mu \frac{\partial \ln \widehat{p}(\boldsymbol{x}; \theta)}{\partial \theta}.$$

Отсюда получаем еще одно следствие.

Следствие 16.2. *Оценка $\theta^* = \theta^*(\boldsymbol{X})$ является эффективной для параметра $\theta \Longleftrightarrow$ выполняется*

$$\frac{\partial \ln L(\theta; \boldsymbol{X})}{\partial \theta} = k(\theta)(\theta^* - \theta), \tag{16.8}$$

где $k(\theta)$ не зависит от $\boldsymbol{X}$.

Задача 16.2. Пусть оценивается некоторая функция от параметра $\tau(\theta)$. Оценка $\widehat{\tau} = \widehat{\tau}(\boldsymbol{X})$ является эффективной оценкой $\tau(\theta) \Longleftrightarrow$

$$\Longleftrightarrow \frac{\partial \ln L(\theta; \boldsymbol{X})}{\partial \theta} = k(\theta)(\widehat{\tau} - \tau(\theta)), \tag{16.9}$$

где $k(\theta)$ не зависит от $\boldsymbol{X}$.

Задача 16.3. Покажите, что если для $\tau(\theta)$ существует эффективная оценка, то только функции вида $\alpha\tau(\theta) + \beta$ ($\alpha, \beta \in \mathbb{R}$ — произвольные константы) также допускают эффективное оценивание, то есть только единственная (с точностью до линейных преобразований) функция параметра может иметь эффективную оценку.

Основная теорема этого раздела также допускает обобщение на случай, когда оценивается некоторая функция параметра $\tau(\theta)$. А именно: при выполнений условий регулярности (условие 3 требуется соответствующим образом изменить) для дисперсии несмещенной оценки $\widehat{\tau}(\theta)$ справедливо неравенство

$$\mathbf{D}_\theta \widehat{\tau} \geqslant \frac{(\tau'(\theta))^2}{i_n(\theta)}.$$

Равенство (16.8) можно рассматривать как дифференциальное уравнение для плотности (совместной вероятности) распределения вектора наблюдений. Решая это уравнение, получим

$$\ln \widehat{p}(\boldsymbol{x};\theta) = \int k(\theta)(\theta^* - \theta)\, d\theta + c(\boldsymbol{x}),$$

или

$$\widehat{p}(\boldsymbol{x};\theta) = h(\boldsymbol{x}) \cdot \exp\left(\theta^* \int k(\theta)\, d\theta\right) \cdot u(\theta).$$

Окончательно перепишем плотность в следующем виде:

$$\widehat{p}(\boldsymbol{x};\theta) = h(\boldsymbol{x}) \cdot u(\theta) \cdot \exp(A(\theta)T(\boldsymbol{x})).$$

Следовательно, плотность принадлежит экспоненциальному семейству распределений. Таким образом, получено еще одно следствие.

Следствие 16.3. *Эффективная оценка может существовать только для некоторой функции параметра экспоненциального семейства распределений.*

Пример 16.6. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, распределенные по закону Пуассона с параметром $\theta > 0$. Тогда функция правдоподобия имеет вид

$$L(\theta; \boldsymbol{X}) = \left(\prod_{i=1}^{n} X_i!\right)^{-1} \theta^{\sum_{i=1}^{n} X_i} e^{-n\theta}.$$

Следовательно,

$$\frac{\partial \ln L(\theta; \boldsymbol{X})}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^{n} X_i - n = \frac{n}{\theta}(\overline{X} - \theta),$$

откуда получаем, что оценка максимального правдоподобия $\theta^* = \overline{X}$ является эффективной.

Данный пример является иллюстрацией общего утверждения: если существует эффективная оценка для параметра θ, то она совпадает с оценкой максимального правдоподобия, что еще раз говорит в пользу данного метода.

Задача 16.4. Докажите, что при выполнении условий регулярности 1, 2, 4 количество информации можно вычислять по формуле

$$i_n(\theta) = -\mathbf{E}_\theta \left(\frac{\partial^2 \ln L(\theta; \boldsymbol{X})}{\partial \theta^2} \right).$$

Замечание. Изложенная в данном разделе теория эффективных оценок дает хорошие результаты, когда мы точно знаем распределения наблюдаемых случайных величин. Но если реальное распределение хотя бы ненамного отличается от предполагаемого, то оценки утрачивают свои свойства и могут значительно отличаться от параметра.

Приведем следующий пример.

Пример 16.7. Пусть оценивается параметр сдвига a нормального распределения с известной дисперсией $\sigma^2 = 1$. Тогда $\overline{X}$ является эффективной оценкой для параметра a, ее эффективность достигает максимального значения

$$e(\overline{X}) = \frac{1/i_n(a)}{\mathbf{D}_a \overline{X}} = 1,$$

но эта оценка не обладает устойчивостью при изменениях распределения (скажем, если брать смеси нормального распределения с некоторым другим распределением). С другой стороны, выборочная медиана $\widehat{X}$, определяемая как средний член вариационного ряда, обладает свойством устойчивости, хотя несколько проигрывает первой оценке в эффективности. Асимптотическая эффективность выборочной медианы равна[1]

$$e(\widehat{X}) = \frac{1/i_n(a)}{\mathbf{D}_a \widehat{X}} \approx \frac{2}{\pi} \approx 0.64.$$

В математической статистике существует много различных подходов к построению точечных оценок. Это методы, осно-

[1] Об асимптотических свойствах выборочной медианы можно прочитать в [12]. Более полное сравнение свойств медианы и выборочного среднего содержится в [20].

ванные на порядковых статистиках [8], последовательные методы [5] и др. В заключение приведем простой метод, позволяющий находить состоятельные оценки.

16.5. Метод моментов

Этот метод достаточно простой и применим, когда у распределения существуют конечные моменты до порядка k включительно, причем число неизвестных параметров не превышает k.

Пусть $X_1, \ldots, X_n$ — случайная выборка из распределения $F(x; \theta_1, \ldots, \theta_k)$ и для этого распределения существуют

$$m_j = \mathbf{E} X_i^j = m_j(\theta_1, \ldots, \theta_k), \quad j = 1, \ldots, k.$$

Составим систему из k алгебраических уравнений с неизвестными $\theta_1, \ldots, \theta_k$:

$$\begin{cases} m_1(\theta_1, \ldots, \theta_k) = m_1, \\ \ldots\ldots\ldots\ldots \\ m_k(\theta_1, \ldots, \theta_k) = m_k. \end{cases} \qquad (16.10)$$

Пусть система 16.10 разрешима, имеет единственное решение

$$\begin{cases} \theta_1 = g_1(m_1, \ldots, m_k), \\ \ldots\ldots\ldots\ldots \\ \theta_k = g_k(m_1, \ldots, m_k), \end{cases}$$

и функции $g_1, \ldots, g_k$ непрерывны. Конечно, моменты распределения нам также неизвестны, но их легко оценить с помощью выборочных моментов.

Определение 16.3. *Выборочным моментом* порядка j называют статистику

$$M_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j.$$

Задача 16.5. Докажите, что выборочные моменты M_j являются несмещенными и состоятельными оценками для теоретических моментов m_j.

Оценки неизвестных параметров получим, если в решениях системы (16.10) теоретические моменты заменим на выборочные:

$$\widehat{\theta}_j = g_j(M_1, \ldots, M_k), \quad j = 1, \ldots, k.$$

Задача 16.6. Докажите, что полученные оценки являются состоятельными.

Пример 16.8. Построим методом моментов оценки неизвестных параметров a и b равномерного распределения на отрезке $[a, b]$. Поскольку число неизвестных параметров равно двум, то надо составить два уравнения. При этом отметим, что система получится проще, если во втором уравнении приравнивать дисперсии:

$$\begin{cases} m_1 = \dfrac{a+b}{2}, \\ \sigma^2 = \dfrac{(b-a)^2}{12} \end{cases} \iff \begin{cases} a = m_1 - \sqrt{3}\sigma, \\ b = m_1 + \sqrt{3}\sigma. \end{cases}$$

Заменим m_1 на $\overline{X}$, а σ^2 на выборочную дисперсию $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$. Получим искомые оценки

$$\widehat{a} = \overline{X} - \sqrt{3}S, \quad \widehat{b} = \overline{X} + \sqrt{3}S.$$

Приложение 1

ОСНОВНЫЕ РАСПРЕДЕЛЕНИЯ И ИХ СВОЙСТВА

Распределение Бернулли $B(1;p)$ с параметром $p \in (0,1)$.

Случайная величина $X \sim B(1;p)$, если

$$\mathbf{P}\{X=1\}=p, \quad \mathbf{P}\{X=0\}=1-p.$$
$$\mathbf{E}\,x=p, \quad \mathbf{D}\,x=p(1-p),$$

характеристическая функция $f(t)=1-p+pe^{it}$.

Примером случайной величины с таким распределением может служить индикатор случайного события A: $I_A(\omega)=1$, если $\omega \in A$, и $I_A(\omega)=0$, если $\omega \in \overline{A}$, $p=\mathbf{P}(A)$.

Биномиальное распределение $B(n;p)$ с параметрами $n \in \mathbb{N}$, $p \in (0;1)$.

Случайная величина $X \sim B(n;p)$, если

$$\mathbf{P}\{X=k\}=C_n^k p^k (1-p)^{n-k}, \quad k=0,1,\ldots,n.$$
$$\mathbf{E}X=np, \quad \mathbf{D}X=np(1-p), \quad f(t)=(1-p+pe^{it})^n.$$

Примером случайной величины с таким распределением может служить число успехов в n независимых испытаниях Бернулли с вероятностью успеха p в каждом из испытаний.

Распределение Пуассона $\Pi(\lambda)$ с параметром $\lambda>0$.

Случайная величина $X \sim \Pi(\lambda)$, если

$$\mathbf{P}\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,2,\ldots\ .$$
$$\mathbf{E}X=\lambda, \quad \mathbf{D}X=\lambda, \quad f(t)=\exp(\lambda(e^{it}-1)).$$

Распределение Пуассона применяется для описания «редких» событий, является предельным для биномиального в случае, когда $n\to\infty$, при этом $np_n \to \lambda>0$. Сформулируем одно из основных свойств этого распределения. Пусть случайные величины X_1 и X_2 — независимы, $X_1 \sim \Pi(\lambda_1)$, $X_2 \sim \Pi(\lambda_2)$. Тогда $X_1+X_2 \sim \sim \Pi(\lambda_1+\lambda_2)$.

Геометрическое распределение $G(p)$ с параметром $p \in (0,1)$.

Случайная величина $X \sim G(p)$, если

$$\mathbf{P}\{X = k\} = p(1-p)^k, \quad k = 0, 1, 2, \ldots .$$

$$\mathbf{E}X = \frac{1-p}{p}, \quad \mathbf{D}X = \frac{1-p}{p^2}, \quad f(t) = p(1-(1-p)e^{it})^{-1}.$$

Такое распределение имеет случайная величина, равная числу неудач, предшествующих первому успеху в испытаниях Бернулли. Важное свойство данного распределения — свойство отсутствия последействия: $\mathbf{P}\{X \geqslant m + t | X \geqslant m\} = \mathbf{P}\{X \geqslant t\}$. Если $X_1 \sim G(p_1)$, $X_2 \sim G(p_2)$ и эти случайные величины независимы, то $\min(X_1, X_2) \sim G(1-(1-p_1)(1-p_2))$.

Геометрическим также называют распределение случайной величины $Y = X + 1$, принимающей значения $k = 1, 2, \ldots$ с вероятностями $\mathbf{P}\{Y = k\} = p(1-p)^{k-1}$. Для Y свойства будут несколько иными.

Отрицательное биномиальное распределение $\mathrm{Bi}^-(m;p)$ с параметрами $m \in \mathbb{N}$, $p \in (0,1)$.

Случайная величина $X \sim \mathrm{Bi}^-(m;p)$, если

$$\mathbf{P}\{X = k\} = C_{k+m-1}^{k} p^m (1-p)^k, \quad k = 0, 1, 2, \ldots .$$

$$\mathbf{E}X = \frac{m(1-p)}{p}; \quad \mathbf{D}X = \frac{m(1-p)}{p^2}, \quad f(t) = p^m(1-(1-p)e^{it})^{-m}.$$

Примером случайной величины с отрицательным биномиальным распределением может служить число неудач, предшествующих наступлению m-го успеха в испытаниях Бернулли. Заметим, что $G(p) = \mathrm{Bi}^-(1;p)$. Если $X_1 \sim \mathrm{Bi}^-(m_1;p)$, $X_2 \sim \mathrm{Bi}^-(m_2;p))$ и эти случайные величины независимы, то $X_1 + X_2 \sim \mathrm{Bi}^-(m_1 + m_2;p)$.

Гипергеометрическое распределение $HG(N, M, n)$.

Случайная величина $X \sim HG(N, M, n)$, если

$$\mathbf{P}\{X = k\} = \frac{C_m^k C_{N-M}^{n-k}}{C_N^n},$$

где $\max(0; n - N + M) \leqslant k \leqslant \min(n; M)$, $k \in \mathbb{Z}$.

$$\mathbf{E}X = \frac{nM}{N}, \quad \mathbf{D}X = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

Примером случайной величины с таким распределением является случайная величина, равная числу красных шаров среди n

отобранных, если производится случайный выбор без возвращения из совокупности, содержащей M красных и $N-M$ белых шаров.

Полиномиальное распределение $B(n;p_1,\ldots,p_m)$ с параметрами $n\in\mathbb{N}$, $p_i>0$, $\sum\limits_{i=1}^{m}p_i=1$.

Случайный вектор $(X_1,\ldots,X_m)\sim B(n;p_1,\ldots,p_m)$, если

$$\mathbf{P}\{X_1=k_1,\ldots,X_m=k_m\}=\frac{n!}{k_1!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m},$$

где $k_1+\ldots+k_m=n$, k_m — целые неотрицательные числа. Среди свойств этого распределения отметим, что

$$X_i\sim B(n;p_i),\quad \operatorname{cov}((,X)_i,X_j)=-np_ip_j,\quad \text{если}\quad i\neq j.$$

Полиномиальное распределение применяется в независимых повторных испытаниях с m различными исходами для совместного распределения случайных величин $X_1,\ldots,X_m$, где X_i $(1\leqslant i\leqslant m)$ равна числу наступлений i-го исхода (p_i — вероятность i-го исхода в каждом из испытаний).

Нормальное распределение $N(a,\sigma^2)$ с параметрами $-\infty< <a<+\infty$, $\sigma>0$.

Случайная величина $X\sim N(a,\sigma^2)$, если ее плотность распределения имеет вид

$$p(x;a,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2/(2\sigma^2)}.$$

$$\mathbf{E}X=a,\quad \mathbf{D}X=\sigma^2,\quad f(t)=e^{iat-\sigma^2t^2/2}.$$

Для плотности и функции распределения стандартного нормального закона $N(0,1)$ приняты специальные обозначения:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\quad \Phi(x)=\int\limits_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du$$

соответственно. Для семейства нормальных распределений параметры a и σ являются параметрами сдвига и масштаба, поскольку

$$p(x;a,\sigma)=\frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right),\quad F(x;a,\sigma)=\Phi\left(\frac{x-a}{\sigma}\right),\quad \frac{X-a}{\sigma}\sim N(0,1),$$

где $F(x;a,\sigma)$ обозначает функцию распределения случайной величины X.

Любая нетривиальная линейная комбинация независимых нормально распределенных случайных величин снова имеет нормальное распределение, параметры которого легко вычислить, используя свойства математического ожидания и дисперсии.

Равномерное распределение $U(a,b)$ с параметрами $-\infty < a < b < +\infty$.

Случайная величина $X \sim U(a,b)$, если плотность распределения определяется формулой

$$p(x;a,b) = \begin{cases} \dfrac{1}{b-a} & \text{при} \quad x \in [a;b], \\ 0 & \text{при} \quad x \notin [a;b]. \end{cases}$$

$$\mathbf{E}X = \frac{a+b}{2}, \quad \mathbf{D}X = \frac{(b-a)^2}{12}, \quad f(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

Всякое невырожденное линейное преобразование $Y = cX + d$ снова имеет равномерное распределение, в том числе $X^* = \dfrac{X-a}{b-a} \sim U(0,1)$.

Показательное (экспоненциальное) распределение $E(\alpha,\beta)$ с параметрами $\alpha > 0$, $-\infty < \beta < +\infty$.

Случайная величина $X \sim E(\alpha,\beta)$, если ее плотность распределения равна

$$p(x;\alpha,\beta) = \begin{cases} \alpha e^{-\alpha(x-\beta)} & \text{при} \quad x > \beta, \\ 0 & \text{при} \quad x \leqslant \beta. \end{cases}$$

$$\mathbf{E}X = \beta + \frac{1}{\alpha}, \quad \mathbf{D}X = \frac{1}{\alpha^2}, \quad f(t) = \frac{\alpha e^{it\beta}}{\alpha - it}.$$

Параметры β и α^{-1} являются соответственно параметрами сдвига и масштаба, $Y = \alpha(X - \beta) \sim E(1,0)$. Для распределения $E(\alpha,0)$, так же как и для показательного, выполняется свойство отсутствия последействия. Если $X_1 \sim E(\alpha,0)$, $X_2 \sim E(\alpha,0)$ и эти случайные величины независимы, то $\min(X_1,X_2) \sim E(\alpha_1 + \alpha_2)$.

Гамма-распределение $\Gamma(\alpha,\lambda)$ с параметрами $\alpha > 0$, $\lambda > 0$.

Случайная величина $X \sim \Gamma(\alpha,\lambda)$, если она имеет плотность следующего вида:

$$p(x;\alpha,\lambda) = \begin{cases} \dfrac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x \leqslant 0. \end{cases}$$

Параметр λ называют параметром формы, а параметр $\dfrac{1}{\alpha}$ является параметром масштаба, поскольку $\alpha X \sim \Gamma(1,\lambda)$.

$$\mathbf{E}X = \frac{\lambda}{\alpha}, \quad \mathbf{D}X = \frac{\lambda}{\alpha^2}, \quad f(t) = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}.$$

Отметим, что распределение $\Gamma(\alpha, 1)$ совпадает с распределением $E(\alpha, 0)$.

Распределение хи-квадрат $\chi^2(k)$ с k степенями свободы, $k \in \mathbb{N}$.

Случайная величина $X \sim \chi^2(k)$, если ее плотность

$$p(x;k) = \begin{cases} \dfrac{x^{k/2-1}\, e^{-x/2}}{\Gamma(k/2)2^{k/2}} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x \leqslant 0. \end{cases}$$

$$\mathbf{E}X = k, \quad \mathbf{D}X = 2k, \quad \mathbf{E}(X-k)^3 = 8k, \quad f(t) = (1-2it)^{-k/2}.$$

Данное распределение имеет случайная величина

$$X = \sum_{j=1}^{k} \xi_j^2,$$

где $\xi_1, \ldots, \xi_k$ — независимые случайные величины с распределением $N(0,1)$. Это распределение совпадает с распределением $\Gamma\left(\frac{1}{2}, \frac{k}{2}\right)$.

Бета-распределение $\beta(r,s)$ с параметрами r, $s > 0$.

Случайная величина $X \sim \beta(r,s)$, если плотность распределения равна

$$p(x;r,s) = \begin{cases} \dfrac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1} & \text{при} \quad x \in (0,1), \\ 0 & \text{при} \quad x \leqslant 0, x \geqslant 1. \end{cases}$$

$$\mathbf{E}X = \frac{r}{r+s}, \quad \mathbf{D}X = \frac{rs}{(r+s)^2(r+s+1)}.$$

Отметим, что распределение $\beta(1,1)$ совпадает с $U(0,1)$.

Распределение Стьюдента $t(n)$ с n степенями свободы, $n \in \mathbb{N}$.

Случайная величина $X \sim t(n)$, если она по распределению совпадает со случайной величиной

$$\frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\xi_j^2}},$$

где $\xi_0, \xi_1, \ldots, \xi_n$ — независимые случайные величины с распределением $N(0;1)$. Плотность распределения X равна

$$p(x;n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-n+1/2}.$$

$$\mathbf{E}X = 0 \quad (n>1), \qquad \mathbf{D}X = \frac{n}{n-2} \quad (n>2).$$

Распределение Коши $K(a,b)$ с параметрами $-\infty < a < +\infty$, $b>0$.

Случайная величина $X \sim K(a,b)$, если плотность распределения задается формулой

$$p(x;a,b) = \frac{b}{\pi(b^2+(x-a)^2)}.$$

Математическое ожидание и дисперсия не существуют, $f(t) = e^{ita-b|t|}$. Параметры a и b являются параметрами сдвига и масштаба соответственно. Если $a=0$, $b=1$, то распределение Коши совпадает с распределением Стьюдента с одной степенью свободы.

Распределение Фишера $F(m,n)$ с параметрами n, $m \in \mathbb{N}$.

Случайная величина $X \sim F(m,n)$, если она совпадает по распределению с величиной

$$\frac{\frac{1}{m}\xi}{\frac{1}{n}\eta}, \quad \xi \sim \chi^2(m), \quad \eta \sim \chi^2(n),$$

причем ξ и η — независимые случайные величины.

Многомерное нормальное распределение $N(\boldsymbol{a}, \boldsymbol{\Sigma})$, $\boldsymbol{a} = (a_1, \ldots, a_n)^T$ — действительный n-мерный вектор, $\boldsymbol{\Sigma}$ — действительная неотрицательно определенная симметричная матрица порядка $n \times n$ с элементами σ_{ij}.

Случайный вектор $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ имеет невырожденное (собственное) n-мерное распределение $N(\boldsymbol{a}, \boldsymbol{\Sigma})$, если $\det \boldsymbol{\Sigma} > 0$

и плотность распределения имеет вид

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right).$$

Характеристическая функция этого распределения равна

$$f(\boldsymbol{t}) = \mathbf{E}\, e^{i(\boldsymbol{t},\boldsymbol{X})} = \exp\left(i(\boldsymbol{t},\boldsymbol{a}) - \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma}\boldsymbol{t}\right).$$

Эта формула имеет смысл и в том случае, когда плотность распределения не существует ($\det \boldsymbol{\Sigma} = 0$). Тогда распределение называют несобственным нормальным распределением. Несобственное нормальное распределение сосредоточено на некотором линейном подпространстве размерности $m = \operatorname{rank} \boldsymbol{\Sigma} < n$.

Пусть $\boldsymbol{Y}$ — проекция случайного вектора на это подпространство и $m > 0$. Тогда $\boldsymbol{Y}$ имеет невырожденное m-мерное нормальное распределение. Компоненты X_k случайного вектора $\boldsymbol{X}$ — это проекции на координатные оси,

$$X_k \sim N(a_k, \sigma_{kk}), \quad \operatorname{cov}(X_k, X_l) = \sigma_{kl}.$$

В частном случае при $n = 2$ и $\boldsymbol{a} = \mathbf{0}$:

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2\varrho\frac{x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right).$$

В этом случае пишут $(X_1, X_2) \sim N(a_1, a_2, \sigma_1^2, \sigma_2^2, \varrho)$, $\sigma_1 > 0$, $\sigma_2 > 0$, $|\varrho| < 1$.

$$\mathbf{E}\, X_k = a_k, \quad \mathbf{D}\, X_k = \sigma_k^2, \quad \varrho = cor(X_1, X_2).$$

Приложение 2

ЭКЗАМЕНАЦИОННЫЕ ВОПРОСЫ ПО КУРСУ «ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

1. Вероятностное пространство. Счетная аддитивность, монотонность вероятностной меры. Вероятность объединения событий. Лемма Бореля–Кантелли.

2. Независимость событий, случайных величин.

3. Случайные величины и их распределения вероятностей.

4. Биномиальное распределение. Пуассоновская аппроксимация (предельная теорема и неравенство).

5. Биномиальное распределение. Интегральная теорема Муавра–Лапласа (вывод ее из локальной теоремы Муавра–Лапласа или из центральной предельной теоремы).

6. Неравенство Чебышева и его уточнения. Закон больших чисел. Теорема Бернулли.

7. Вероятностное доказательство теоремы Вейерштрасса о равномерном приближении непрерывной на отрезке функции полиномами.

8. Классическое и геометрическое определение вероятности. Свойства вероятностей.

9. Формула композиции (свертки) распределений и ее применение. Распределение суммы независимых нормально распределенных случайных величин. Распределение суммы независимых случайных величин, равномерно распределенных на отрезке $[0;1]$.

10. Неравенства для распределений максимума сумм независимых случайных величин.

11. Неравенство Леви.

12. Задача о разорении игрока.

13. Математические ожидания и их свойства.

14. Характеристические функции: формула обращения, теорема единственности (план доказательства), теорема непрерывности.

15. Центральная предельная теорема для одинаково распределенных случайных величин. Теорема Ляпунова.

16. Задача выбора одной из двух простых гипотез. Оценка снизу необходимого числа независимых наблюдений с помощью неравенства Йенсена.

17. Задача выбора одной из двух простых гипотез. Лемма Неймана–Пирсона и ее применение к проверке гипотез о математическом ожидании и дисперсии нормального распределения.

18. Задача выбора одной из двух простых гипотез. Лемма Неймана–Пирсона и ее применение к проверке гипотез о вероятности успеха в схеме Бернулли.

19. Несмещенные оценки. Неравенство Рао–Крамера. Эффективные оценки.

20. Эффективные оценки: метод максимального правдоподобия, оценки с дисперсией, меньшей, чем граница Рао–Крамера.

21. Асимптотическое распределение выборочной медианы в выборке из нормального распределения. Выборочная медиана как оценка неизвестного математического ожидания в этом случае.

22. Определение доверительного интервала. Доверительные интервалы для неизвестного математического ожидания нормального распределения (при известной и неизвестной дисперсии).

23. Доверительные интервалы для вероятности успеха в схеме Бернулли.

24. Различные виды сходимости последовательностей случайных величин.

25. Теорема о предельном распределении статистики критерия согласия Пирсона хи-квадрат.

26. Усиленный закон больших чисел.

27. Сходимость рядов из независимых случайных величин.

28. Эргодическая теорема для однородных цепей Маркова с конечным числом состояний.

29. Сложение независимых случайных величин по $\bmod 1$. Сложение целочисленных случайных величин по $\bmod 1$. Сходимость к равномерному распределению на соответствующей группе.

Литература

1. *Бернулли Я.* О законе больших чисел. М.: Наука. Главная редакция физико-математической литературы, 1986.
2. *Бернштейн С.Н.* Теория вероятностей. 3-е изд. М.; Л.: ОНТИ, 1934.
3. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1983.
4. *Боровков А.А.* Теория вероятностей. 3-е изд. М.: Эдиториал УРСС, 1999.
5. *Вальд А.* Последовательный анализ. М.: Физматгиз, 1960.
6. *Гнеденко Б.В.* Курс теории вероятностей. 7-е изд. М.: Эдиториал УРСС, 2001.
7. *Гнеденко Б.В., Колмогоров А.Н.* Предельные распределения для сумм независимых случайных величин. М.; Л.: Гостехиздат, 1949.
8. *Дейвид Г.* Порядковые статистики. М.: Наука, 1979.
9. *Демидович Б.П.* Сборник задач и упражнений по математическому анализу. 10-е изд. М.: Наука, 1990.
10. *Ермаков С.М.* Метод Монте-Карло и смежные вопросы. М.: Наука, 1971.
11. *Золотарев В.М.* Абсолютная оценка остаточного члена в центральной предельной теореме // Теория вероятностей и ее применения. 1966. Т. 11, № 1. С. 108–119.
12. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. 2-е изд. М.: Высшая школа, 1992.
13. *Климов Г.П.* Теория вероятностей и математическая статистика. М.: Изд-во Московского университета, 1983.
14. *Колмогоров А.Н.* Основные понятия теории вероятностей. М.: Фазис, 1998.
15. *Колмогоров А.Н.* Теория вероятностей и математическая статистика: Сборник статей. М.: Наука, 1986.
16. *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. М.: Наука, 1968.
17. *Крамер Г.* Математические методы статистики. 2-е изд. М.: Мир, 1975.

18. *Крамер Г.* Случайные величины и распределения вероятностей. М.: Иностранная литература, 1947.
19. *Леман Э.* Проверка статистических гипотез. 2-е изд. М.: Наука, 1979.
20. *Леман Э.* Теория точечного оценивания. М.: Наука, 1991.
21. *Ляпунов А.М.* Собрание сочинений. М.: Изд-во АН СССР, 1954.
22. *Марков А.А.* Исчисление вероятностей. 4-е изд. М.: Госиздат, 1924.
23. Математическая Энциклопедия. В 5 т. / Под ред. И. М. Виноградова (гл. ред.) и др. М.: Советская энциклопедия, 1977–1985.
24. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей: Основные понятия. Предельные теоремы. Случайные процессы. 2-е изд. М.: Наука, 1973.
25. *Розанов Ю.А.* Лекции по теории вероятностей. 3-е изд. Долгопрудный: ИД «Интеллект», 2008.
26. *Севастьянов Б.А.* Курс теории вероятностей и математической статистики. М.: Институт компьютерных исследований, 2004.
27. Теория вероятностей и математическая статистика: Энциклопедия / Под ред. Ю. В. Прохорова (гл. ред.) и др. М.: Большая Российская энциклопедия, 1999.
28. *Чистяков В.П.* Курс теории вероятностей. 2-е изд. М.: Наука, 1982.
29. *Феллер В.* Введение в теорию вероятностей и ее приложения. В 2 т. М.: Мир, 1984.
30. *Ширяев А.Н.* Вероятность. М.: Наука, 1980.
31. *Ширяев А.Н.* Вероятность. В 2-х кн. 3-е изд. М.: МЦНМО, 2004.
32. *Чебышев П.Л.* Теория вероятностей: Лекции академика П. Л. Чебышева, читанные в 1879, 1880 гг. / Издано А. Крыловым по записи А. М. Ляпунова.— М.; Л., 1936.
33. *Фихтенгольц Г.М.* Курс дифференциального и интегрального исчисления. М.: Наука, 1969.
34. *Laplace P.S. de.* A philosophical essay on probabilities. 2nd edition. New York: Dover, 1951.
35. *Poincaré H.* Calcul des probabilités. Paris: G. Carré, 1896.

Предметный указатель

Moscow State University
Faculty of Computational Mathematics and Cybernetics

Lectures on Probability Theory and Mathematical Statistics

Yuri Prokhorov, Liubov Ponomarenko

SUMMARY

This textbook is based on the one-year lecture course on probability theory and mathematical statistics which has been read for many years for second-year students of the Computational Mathematics and Cybernetics Faculty of Moscow State University.

In this study course the main concepts and definitions are first introduced for the finite probability space case and illustrated by concrete examples. Originally the proofs of theorems are presented for the finite probability space but in the form that easily enables their extension to the general case. Such an approach makes it possible to prove comprehensive complex theories through application of relatively straightforward (simplified) mathematical tools. Further in the textbook the reader is introduced to the mathematical basis of the probability theory, the limit theorems, convergence of sequences and random variable series. The final chapters are devoted to the matters of mathematical statistics.

Particular attention is given to the probabilities estimations in the form of approximate formulae or inequalities. The textbook contains a number of specific and particularly useful examples illustrating the basic concepts of the probability theory and mathematical statistics.

Recommended for students specializing in: applied mathematics and informatics, information technologies.

KEYWORDS: probability of random event; independence of random variables and random events; mathematical expectation and variance; the law of large numbers; characteristic functions; limit theorems; confidence interval; unbiased, consistent, efficient estimators; testing statistical hypotheses

BOOK STRUCTURE: introduction, 16 chapters, 2 appendixes, list of references, index.

Moscow University Press, 2012.

Contents

Оглавление

Учебное издание

Прохоров Юрий Васильевич,
Пономаренко Любовь Степановна

ЛЕКЦИИ ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

2-е издание, исправленное и дополненное

Научный редактор *канд. физ.-мат. наук А. А. Куликов*
Редактор-корректор *Е. А. Босина*
Художник *В. А. Чернецов*
Художественный редактор *Ю. М. Добрянская*
Технический редактор *З. С. Кондрашова*
Оригинал-макет *П. Л. Поляков*

Подписано в печать 01.12.2011. Формат 60×90/16.
Бумага офсетная № 1. Гарнитура Антиква.
Усл. печ. л. 16,0. Уч.-изд. л. 11,6.
Тираж 1 000 экз. Изд. № 9419. Заказ № 4268.

Ордена «Знак Почета» Издательство Московского университета.
125009, Москва, ул. Б. Никитская, д. 5/7.

Тел.: (495) 629-50-91.
Факс: (495) 697-66-71, (495) 939-33-23 (отдел реализации).
E-mail: secretary-msu-press@yandex.ru.

Сайт издательства: http://www.msu.ru/depts/MSUPubl2005
Интернет-магазин: http://msupublishing.ru

Отпечатано в ППП «Типография «Наука»
121099, Москва, Шубинский пер., 6.