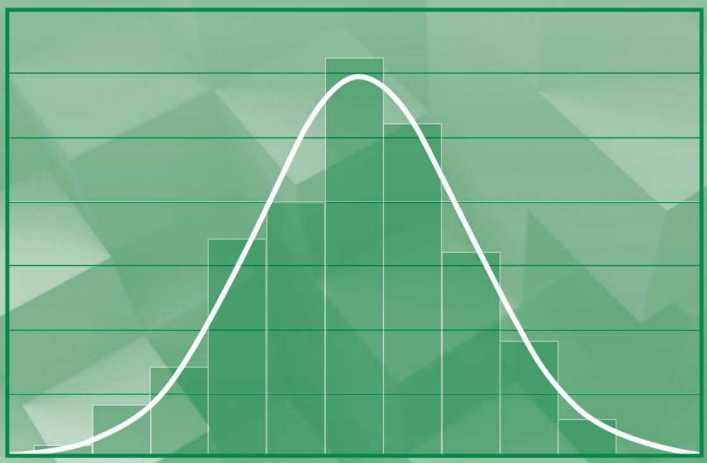
Ю. В. Прохоров
А. В. Прохоров

Курс лекций по теории вероятностей и математической статистике



УДК 519.21(075.8)
ББК 22.171
П84

Прохоров Ю. В., Прохоров А. В.
Курс лекций по теории вероятностей и математической статистике.
М.: МЦНМО, 2020.
144 с.
ISBN 978-5-4439-3392-4

Предлагаемая к изданию книга содержит курс теории вероятностей с элементами математической статистики, рассчитанный на семестр университетского учебного плана.

Впервые этот курс был прочитан в Московском инженерно-физическом институте, затем в переработанном и дополненном виде на факультете вычислительной математики и кибернетики и на инженерном отделении механико-математического факультета МГУ. Изложение основ теории вероятностей дополнено нестандартным материалом. Так, например, в курс включены простейшие примеры марковских случайных процессов, закон больших чисел для зависимых случайных величин, в частности цепей Маркова, элементы теории массового обслуживания. Кроме того, дано полное оригинальное доказательство основных теорем классической теории вероятностей: теоремы Пуассона и центральной предельной теоремы.

Книга может служить учебным пособием для студентов университетов, в образовательной программе которых в соответствии с учебным планом теория вероятностей и математическая статистика занимают важное место.

Подготовлено на основе книги:
Прохоров Ю. В., Прохоров А. В. Курс лекций по теории вероятностей и математической статистике. — М.: МЦНМО, 2019. — 144 с. — ISBN 978-5-4439-1392-6

Учебное издание для вузов

6+

Издательство Московского центра
непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11,
тел. (499)241-08-04.
http://www.mccme.ru

ISBN 978-5-4439-3392-4

© Прасолов В. В., 2020.
© МЦНМО, 2020.

Оглавление

Предисловие 5
Введение 7

Глава 1. Классическая вероятностная модель: урновая схема 8

Глава 2. Основные понятия математической теории вероятностей . . 14
§ 1. Конечное вероятностное пространство 14
§ 2. Условная вероятность. Формула полной вероятности 17
§ 3. Независимость случайных событий 19
§ 4. Случайные величины 23
§ 5. Совместное распределение вероятностей и независимость случайных величин 30
§ 6. Неравенство Чебышёва. Закон больших чисел 39
§ 7. Вероятностные модели с бесконечным числом элементарных событий 44

Глава 3. Суммирование независимых случайных величин. Предельные теоремы теории вероятностей 57
§ 1. Производящие и характеристические функции 58
§ 2. Предельная теорема Пуассона 63
§ 3. Центральная предельная теорема для независимых случайных величин 65
§ 4. Замечания к центральной предельной теореме 72
§ 5. Применение центральной предельной теоремы к статистическим задачам 75

Глава 4. Последовательности зависимых случайных величин. Цепи Маркова. Простейшие марковские процессы в задачах массового обслуживания 80
§ 1. Вариант закона больших чисел для зависимых случайных величин. Теорема Бернштейна 80
§ 2. Цепи Маркова 83
§ 3. Эргодические теоремы для цепей Маркова 90
§ 4. Закон больших чисел для цепей Маркова 94
§ 5. Понятие о марковском случайном процессе. Пуассоновский процесс 97
§ 6. Пример системы обслуживания с ожиданием 101
§ 7. Пример системы обслуживания с отказами 103

Глава 5. Элементы математической статистики 109
§ 1. Постановка задач в математической статистике 109

§ 2. Статистическая оценка параметра биномиального распределения 111
§ 3. Общие проблемы теории статистических оценок 115
§ 4. Методы получения точечных статистических оценок 121
§ 5. Интервальные статистические оценки 125
§ 6. Проверка статистических гипотез 129
§ 7. Фундаментальная лемма Неймана—Пирсона 136
Литература 141

Предисловие

Предлагаемая к изданию книга содержит курс теории вероятностей с элементами математической статистики, рассчитанный на семестр университетского учебного плана.

Впервые этот курс был прочитан в Московском инженерно-физическом институте, затем в переработанном и дополненном виде на факультете вычислительной математики и кибернетики и на инженерном отделении механико-математического факультета МГУ. Первоначально курс был задуман как введение в теорию вероятностей для студентов, обучающихся на отделениях прикладной математики, студентов инженерных и физических факультетов, но опыт показал, что он может быть полезен и студентам чисто математической специализации. Изложение основ теории вероятностей дополнено нестандартным материалом. Так, например, в курс включены простейшие примеры марковских случайных процессов, закон больших чисел для зависимых случайных величин, в частности цепей Маркова, элементы теории массового обслуживания. Кроме того, дано полное оригинальное доказательство основных теорем классической теории вероятностей: теоремы Пуассона и центральной предельной теоремы.

Авторы курса лекций — академик Ю. В. Прохоров (1929—2013), возглавлявший кафедру математической статистики факультета ВМК МГУ им. М. В. Ломоносова и отдел теории вероятностей Математического института им. В. А. Стеклова РАН, и А. В. Прохоров — доцент кафедры математической статистики механико-математического факультела МГУ, заслуженный преподаватель МГУ. Оба автора являются непосредственными учениками А. Н. Колмогорова.

Книга может служить учебным пособием для студентов университетов, в образовательной программе которых в соответствии с учебным планом теория вероятностей и математическая статистика занимают важное место.

Другие учебные пособия авторов:

Прохоров Ю. В., *Розанов Ю. А.* Теория вероятностей. 3-е изд. М.: Наука, 1987.

Прохоров Ю. В., *Пономаренко Л. С.* Лекции по теории вероятностей и математической статистике. 2-е изд. М.: Изд-во МГУ, 2012.

Прохоров А. В., *Ушаков В. Г.*, *Ушаков Н. Г.* Задачи по теории вероятностей. 2-е изд. М.: Изд-во «Книжный дом Университета», 2009.

Козлов М. В., *Прохоров А. В.* Введение в математическую статистику. М.: Изд-во МГУ, 1987.

Колмогоров А. Н., *Журбенко И. Г.*, *Прохоров А. В*. Введение в теорию вероятностей. 3-е изд. М.: МЦНМО, 2015.

Введение

Теория вероятностей изучает случайные явления. К категории случайных явлений относят те, для которых учет всех факторов, влияющих на их появление, принципиально невозможен. Например, такой простой опыт, как подбрасывание монеты, имеет два взаимно исключающих исхода — выпадение «орла» и «решки». Результат является случайным событием, так как заранее исход предопределить практически невозможно. В подобных ситуациях при рассмотрении результатов отдельных опытов бывает очень трудно обнаружить какие-либо закономерности. Вероятностные закономерности обнаруживаются в массовых явлениях. Существуют ситуации, когда одно и то же испытание или наблюдение, результат которого случаен, может быть повторено большое число раз в одинаковых условиях, при этом в полученной последовательности испытаний средние результаты обнаруживают устойчивость. Если проведено n опытов с подбрасыванием симметричной монеты и в результате выпало m «орлов», то $\frac{m}{n}$ приближается к $\frac{1}{2}$ при $n \to \infty$, например, при $n = 4040$ бросаниях $\frac{m}{n} = 0{,}508$ (Бюффон, XVIII в.), при $n = 24\,000$ бросаниях $\frac{m}{n} = 0{,}5005$ (К. Пирсон, XX в.). Исходя из подобных результатов разумно число $\frac{1}{2}$ принять за вероятность выпадения «орла» при одном подбрасывании симметричной монеты.

Центральное понятие теории вероятностей — вероятность как числовая характеристика степени возможности осуществления какого-либо события при определенных условиях. С самой общей точки зрения теория вероятностей может быть названа «исчислением вероятностей», так как она дает возможность вычислять вероятности одних событий по вероятностям других событий. В свое время один из основоположников теории вероятностей П. Лаплас писал: «Теория вероятностей есть в сущности не что иное, как здравый смысл, сведенный к исчислению». В этом плане теория вероятностей является такой же частью математики, как математический анализ, теория дифференциальных уравнений и др., однако методы теории вероятностей предназначены для обнаружения закономерностей в тех явлениях, в наступлении которых существенную роль играет случайность.

Глава 1

Классическая вероятностная модель: урновая схема

Приведем пример подсчета вероятностей с помощью простейших комбинаторных рассуждений в так называемой «урновой» схеме. Занумеруем шары номерами от 1 до N. Предположим также, что шары в урне хорошо перемешаны. Шары один за другим извлекаются из урны «наудачу». Возможны две процедуры выбора:

а) выбор с возвращением,

б) выбор без возвращения.

Если общее число вынутых из урны шаров равно n (объем «выборки» равен n) и если j_k — номер шара при k-м извлечении, $1 \leqslant j \leqslant N$, то в каждом из указанных случаев элементарные исходы (события) опыта можно символически описать следующим образом:

а) $(j_1, j_2, \ldots, j_n)$, некоторые из j_k могут быть равными,

б) $(j_1, j_2, \ldots, j_n)$, все j_k различны.

Общее число возможных элементарных исходов в каждом из этих случаев равно соответственно

а) $\underbrace{N \cdot N \cdot \ldots \cdot N}_{n \text{ раз}} = N^n$,

б) $N(N-1)\ldots(N-n+1) = A_N^n$, где $A_N^n = \dfrac{N!}{(N-n)!}$ называют числом размещений из N по n.

Вычислим в схеме выбора с возвращением вероятность того, что среди n извлеченных шаров нет ни одной пары с одинаковыми номерами. Для этого подсчитаем число элементарных исходов опыта, благоприятствующих данному событию, в предположении, что элементарные исходы взаимно исключают друг друга и равновозможны, а затем найдем отношение числа благоприятных исходов к общему их числу. Это отношение и будет искомой вероятностью в соответствии с классическим определением.

Пример 1 (парадокс дней рождения)**.** Предположим, что в группе n студентов и каждый из них мог быть рожден с одинаковой вероятностью в любой день года из $N = 365$ (но, возможно, в разные годы). Тогда вероятность того, что все студенты группы рождены в разные дни года, равна $\dfrac{A_N^n}{N^n}$, а соответственно вероятность того,

что хотя бы у двух студентов совпадут дни рождения, будет равна $p = 1 - \frac{A_N^n}{N^n}$. Зависимость значений этой вероятности от размеров группы отражена в следующей табличке.

n	22	23	30	55	65
p	0,476	0,507	0,706	0,990	0,999

При $n = 366$ (мы приняли число дней в году за 365) вероятность интересующего нас события равна 1, и это дает основание считать результаты этих подсчетов парадоксальными. □

Предположим теперь, что среди N шаров в урне ровно M белых и $N - M$ черных шаров. Занумеруем эти шары так, что все белые шары имеют номера от 1 до M, а все черные шары — соответственно номера от $M + 1$ до N.

Вычислим вероятность того, что среди n извлеченных шаров ровно m белых. Для этого подсчитаем в каждом случае (при выборе с возвращением и без возвращения) число элементарных исходов опыта, благоприятствующих данному событию.

Число способов размещения m белых шаров на n местах без учета их номеров и порядка равно $C_n^m = \frac{A_N^n}{n!}$. Если места белых шаров фиксированы, то число размещений любых m из M шаров на этих местах в сочетании с размещением любых $n - m$ из $N - M$ шаров на оставшихся местах равно в каждом из рассматриваемых случаев соответственно

а) $M^m (N - M)^{n-m}$,

б) $A_M^m A_{N-M}^{n-m}$.

Искомая вероятность при выборе с возвращением равна

$$C_n^m \frac{M^m (N - M)^{n-m}}{N^n} = C_n^m p^m (1 - p)^{n-m},$$

где $p = \frac{M}{N}$ — доля белых шаров в урне (очевидно, что абсолютные значения M и N здесь несущественны). При данных n и p совокупность чисел $C_n^m p^m (1 - p)^{n-m}$, $m = 0, 1, \ldots, n$, образует *биномиальное* распределение вероятностей. Эти вероятности вычисляются либо по специальным таблицам биномиального распределения, либо с помощью таблиц $\lg n!$ или таблиц $\lg C_n^m$.

При выборе без возвращения искомая вероятность равна

$$C_n^m \frac{A_M^m A_{N-M}^{n-m}}{A_N^n} = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Для данных значений N, M, n $(n \leqslant N)$ числа $C_n^m \dfrac{A_M^m A_{N-M}^{n-m}}{A_N^n}$ при $m = 0, 1, \ldots, \min(n, M)$ образуют *гипергеометрическое* распределение вероятностей.

Пример 2 (биномиальное распределение)**.** Идеальный опыт, заключающийся в 100 подбрасываниях симметричной (шансы на выпадение «орла» и «решки» при отдельном бросании одинаковы) монеты, включается в урновую схему с $n = 100$, $p = q = \dfrac{1}{2}$.

Пример 3 (гипергеометрическое распределение)**.** Обратимся к знаменитой генуэзской или французской лотерее, существовавшей в Европе с XVII до начала XX века. Суть ее в том, что, имея 90 номеров, можно было делать ставку на любое число номеров $k = 1, 2, 3, 4, 5$. Если ставка сделана на k номеров и все эти k номеров после розыгрыша имеются среди 5 номеров, «вышедших в тираж», то выигрыши таковы:

$$\begin{aligned} 15 &\text{ ставок при } k = 1; \\ 270 &\text{ ставок при } k = 2; \\ 5500 &\text{ ставок при } k = 3; \\ 75\,000 &\text{ ставок при } k = 4; \\ 1\,000\,000 &\text{ ставок при } k = 5. \end{aligned}$$

Подсчитаем вероятности выигрыша, пользуясь схемой выбора без возвращения при $N = 90$, $n = 5$, $M = k$, $m = k$. Имеем

$$C_5^k \frac{A_k^k A_{90-k}^{5-k}}{A_{90}^5} = \frac{C_{90-k}^{5-k}}{C_{90}^5}.$$

Для значений $k = 1, 2, 3, 4, 5$ эти вероятности равны соответственно

$$\frac{1}{18}, \ \frac{2}{801}, \ \frac{1}{11748}, \ \frac{1}{511033}, \ \frac{1}{43949268}.$$

События, которые обычно люди считают маловероятными, имеют вероятность порядка 10^{-5} (вероятность того, что при раздаче карт четырем игрокам хотя бы один получит полную масть, имеет порядок 10^{-11}). Подсчет вероятностей устраняет иллюзии в оценке шансов на выигрыш, например, при ставке на пять номеров в генуэзской лотерее.

Рассмотрим еще два поучительных примера подсчета вероятностей.

Пример 4 (задача Галилея)**.** Какой исход вероятнее при бросании трех игральных костей: 9 или 10 очков?

Некто заметил, что 9 очков при большом числе испытаний появляется реже, чем 10 очков, хотя он считал, что эти суммы очков на костях должны выпадать одинаково часто. Этот человек, как говорят, обратился за разъяснениями к Галилею, и тот показал, что выводы были основаны на неверном подсчете числа различных равновозможных исходов. При бросании трех игральных костей суммарное число очков на них может принимать значения от 3 до 18. В нижеследующей таблице помещены результаты неверных подсчетов: 6 случаев, соответствующих сумме 9 очков (левый столбец), и 6 случаев, соответствующих сумме 10 очков (правый столбец), которые не являются равновозможными элементарными исходами опыта. В круглых скобках справа указано правильное число равновозможных элементарных исходов, соответствующих каждому случаю.

$$
\begin{array}{llll}
9 = 1+2+6 & (6), & 10 = 1+3+6 & (6),\\
1+3+5 & (6), & 1+4+5 & (6),\\
1+4+4 & (3), & 2+2+6 & (3),\\
2+2+5 & (3), & 2+3+5 & (6),\\
2+3+4 & (6), & 2+4+4 & (3),\\
3+3+3 & (1), & 3+3+4 & (3).
\end{array}
$$

Итак, вывод о равновероятности двух событий был основан на неправильном подсчете (6 исходов и 6 исходов), тогда как выписанные выше исходы не являются элементарными и поэтому не равновероятны. На самом деле при бросании трех игральных костей каждое элементарное событие представляется тройкой чисел (i, j, k), где i, j, k — числа очков на первой, второй, третьей костях соответственно. Первое из выписанных событий {сумма очков равно 9} включает следующие элементарные исходы: $(1, 2, 6)$, $(1, 6, 2)$, $(2, 1, 6)$, $(2, 6, 1)$, $(6, 1, 2)$, $(6, 2, 1)$. Всего элементарных исходов $6^3 = 216$. Событию {сумма очков равна 9} благоприятствуют 25 элементарных исходов, а событию {сумма очков равна 10} — 27 исходов. Поэтому вероятность события {сумма очков 9} равна $\frac{25}{216} = 0{,}1157$, а вероятность события {сумма очков 10} равна $\frac{27}{216} = 0{,}1250$ и разность этих вероятностей равна 0,0092. Возникает вопрос: сколько раз нужно осуществить данный опыт, чтобы на практике испытать эту разницу?

Пример 5 («парадокс» де Мере)**.** По легенде известный игрок де Мере задал Паскалю вопрос: сколько раз нужно бросить одну правильную кость, чтобы вероятность события {хотя бы раз выпадет 6} превосходила $\frac{1}{2}$, и сколько раз нужно бросить две правильные ко-

сти, чтобы вероятность события {хотя бы раз на обеих костях выпадет 6} превосходила $\frac{1}{2}$? Речь и в том и в другом случае идет о наименьшем числе бросаний. Полагаясь на ошибочные рассуждения, де Мере пришел к выводу, что в первом случае достаточно 4, а во втором — 24 бросаний. Будучи уверенным, что это так, де Мере неоднократно держал пари на то, что при 24 бросаниях двух костей хотя бы раз выпадет одновременно 6 очков на обеих костях, но со временем заметил, что в среднем он проигрывает. Как показал Паскаль, в первом случае правильный ответ де Мере был «угадан»: вероятность того, что при 4 бросаниях кости хотя бы раз выпадет 6 очков, немного больше $\frac{1}{2}$, а именно равна $1-\frac{5^4}{6^4}=0{,}51775$. Вероятность того, что при 24 бросаниях двух костей хотя бы раз получится результат (6,6), равна $1-\frac{35^{24}}{36^{24}}=0{,}49140$, т. е. немного меньше $\frac{1}{2}$. Можно вновь задать вопрос о том, сколько наблюдений необходимо сделать, чтобы различить эти вероятности. □

В связи с поставленными проблемами уместно сделать замечание о соотношении прямой и обратной задач теории вероятностей, не выходя пока за пределы урновой схемы.

Прямая задача (задача вероятностного прогноза) заключается в том, чтобы по известному составу урны вычислить вероятность того или иного соотношения шаров в выборке.

Обратная задача состоит в том, чтобы определить неизвестный состав урны по соотношению шаров в выборке из этой урны. Это задача математической статистики, которой посвящена гл. 5.

Поставленные выше вопросы о различении вероятностей и о числе наблюдений, необходимых для этого, соответствуют специально сформулированной обратной задаче: пусть состав урны с белыми и черными шарами неизвестен, но имеется предположение о том, что доля белых шаров в урне либо p_1, либо p_2, причем p_1 и p_2 — заданные числа между 0 и 1. Безошибочный ответ на вопрос, какова доля белых шаров в урне — p_1 или p_2, принципиально невозможен, однако, произведя выбор с возвращением фиксированного объема n, можно установить такое значение n, которое необходимо для различения указанных гипотез с заданной точностью. Точность определяется малой вероятностью ошибки. Рассмотрим пример.

Пример 6 (проверка гипотезы о составе урны)**.** Пусть имеются две гипотезы о доле белых шаров в урне — гипотеза H_1: $p=p_1=0{,}9$ и гипотеза H_2: $p=p_2=0{,}1$. Будем считать, что шары в урне хо-

рошо перемешаны. Наудачу вынем из урны один шар и примем следующее правило: если вынут белый шар, то верна гипотеза H_1, т. е. $p = p_1$; если вынут черный шар, то будем считать, что верна гипотеза H_2, т. е. $p = p_2$. Очевидно, что, принимая каждое из решений, можно ошибиться. Легко подсчитать вероятности ошибок. Пусть, например, верна гипотеза H_2, но вынут белый шар и мы приняли гипотезу H_1. Таким образом, вероятность ошибки равна вероятности того, что шар, извлеченный из урны с долей белых шаров $p_2 = 0{,}1$, окажется белым, т. е. равна 0,1. Вероятность другой возможной ошибки в рассматриваемом конкретном случае также равна 0,1. Это означает, что при следовании принятому правилу, называемому критерием проверки двух гипотез, мы будем ошибаться в среднем в одном случае из 10. Вероятности ошибок можно уменьшить, вынимая из урны не один шар, а три шара с возвращением шаров каждый раз в урну. В этом случае естественно следующее правило поведения: если два или три шара белые, то принимается гипотеза H_1, а если соотношение шаров в выборке обратное, то принимается гипотеза H_2. Вероятность первой из этих ошибок совпадает с вероятностью вытащить два или три белых шара из урны, в которой доля белых шаров равна $p_2 = 0{,}1$. Эта вероятность равна $p_2^3 + 3p_2^2(1 - p_2) = 0{,}028$ и совпадает с вероятностью другой возможной ошибки. Увеличивая число шаров в выборке, можно значительно уменьшить вероятность ошибок, как показывает следующая табличка.

Объем выборки	Вероятность ошибки
5	0,00856
7	0,00273
11	0,00030

Если заранее задать вероятности ошибок, то можно решить задачу об объеме выборки, необходимом для различения двух гипотез о доле белых шаров. Подробнее об этом пойдет речь в гл. 5, посвященной задачам математической статистики.

Глава 2

Основные понятия математической теории вероятностей

§1. Конечное вероятностное пространство

В первую очередь к основным понятиям относятся: пространство элементарных событий, случайное событие, вероятность, случайная величина, математическое ожидание. О первых трех мы уже получили наглядное представление, теперь перейдем к строгим определениям.

Мы рассматриваем эксперименты, каждый из которых может закончиться одним и только одним элементарным исходом из общего числа s. Элементарные исходы опыта условимся обозначать $\omega_1, \omega_2, \ldots, \omega_s$ и называть *элементарными событиями*. Основное свойство элементарных событий — они неразложимы и взаимно исключают друг друга. Совокупность всех элементарных событий $\Omega = \{\omega_1, \ldots, \omega_s\}$ будем называть *пространством элементарных событий*. Предположим, что каждому элементарному событию ω_i поставлено в соответствие такое число p_i, что

1) $p_i \geqslant 0$, $i = 1, \ldots, s$,

2) $\sum_{i=1}^{s} p_i = 1$.

Числа p_i называются *вероятностями* элементарных событий ω_i (иногда мы будем опускать индексы у элементарных событий и употреблять обозначение $p(\omega)$ для всех ω). Выбор математической модели равнозначен перечислению элементарных событий и указанию соответствующих им вероятностей. Числа p_i можно выбирать по-разному, мы считаем p_i уже выбранными.

Всякое событие A, связанное с данным экспериментом, задается как совокупность элементарных событий, с каждым из которых оно осуществляется, и поэтому является подмножеством пространства Ω: $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$, $1 \leqslant k \leqslant s$. К числу всех событий причисляют Ω и пустое множество $\varnothing$, не содержащее элементарных событий.

Вероятностью события A называется сумма вероятностей входящих в него элементарных событий: если $A=\{\omega_{i_1}, \ldots, \omega_{i_k}\}$, то

$$\mathrm{P}\{A\} = p_{i_1} + \ldots + p_{i_k}, \quad \text{или} \quad \mathrm{P}\{A\} = \sum_{\omega \in A} p(\omega).$$

Функция $\mathrm{P}\{A\}$ — функция множеств, область определения которой в рассматриваемом конечном случае совпадает с классом всех подмножеств пространства Ω. Эта функция называется распределением вероятностей на Ω. Если все элементарные события равновероятны, т. е.

$$p_1 = p_2 = \ldots = p_s = \frac{1}{s},$$

то мы приходим к классическому, исторически самому первому определению вероятности:

$$\mathrm{P}\{A\} = \frac{k}{s} = \frac{\text{число элементарных событий, благоприятствующих } A}{\text{число всех элементарных событий}}.$$

Пространство элементарных событий Ω вместе с введенным на нем распределением вероятностей составляют вероятностную модель (Ω, P).

Замечание. Пространство Ω может состоять из счетного числа элементарных исходов, например, пространство, соответствующее эксперименту, состоящему в извлечении шаров из урны до тех пор, пока не появится шар с номером 1. В этом случае в определении вероятностей конечные суммы заменяются рядами. □

Перед тем как указать свойства вероятности, остановимся на операциях над событиями, которые соответствуют обычным операциям над множествами.

Объединение двух событий $A_1 \cup A_2$ — совокупность всех элементарных событий, входящих в A_1 или в A_2.

Пересечение или *произведение* двух событий $A_1 \cap A_2$ — совокупность всех элементарных событий, входящих одновременно в A_1 и в A_2. Два события A_1 и A_2, для которых $A_1 \cap A_2 = \varnothing$, называются *несовместными*.

Разность двух событий $A_1 \setminus A_2$ — совокупность тех элементарных событий, которые входят в A_1, но не входят в A_2.

Противоположное событие, дополнение к событию A — разность $\Omega \setminus A = \overline{A}$.

Операции объединения и пересечения событий определяются аналогично для любого числа событий $A_1, \ldots, A_n$, принадлежащих

одному и тому же пространству элементарных событий Ω. Особенно часто используются следующие соотношения «двойственности»:

• для двух событий:

$$\overline{A_1 \cup A_2} = \bar{A}_1 \cap \bar{A}_2, \quad \overline{A_1 \cap A_2} = \bar{A}_1 \cup \bar{A}_2,$$

• для любого конечного числа событий $A_1, \ldots, A_s$:

$$\overline{\bigcup_{k=1}^{s} A_k} = \bigcap_{k=1}^{s} \bar{A}_k, \quad \overline{\bigcap_{k=1}^{s} A_k} = \bigcup_{k=1}^{s} \bar{A}_k.$$

Отметим следующие главные свойства вероятности:

1) *неотрицательность*: $\mathrm{P}\{A\} \geqslant 0$;

2) *нормированность*: $\mathrm{P}(\Omega) = 1$;

3) *аддитивность*: если $A_1 \cap A_2 = \varnothing$, то $\mathrm{P}\{A_1 \cup A_2\} = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2\}$.

Доказательство свойства 3 следует непосредственно из определения вероятности:

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

По индукции это свойство доказывается для любого конечного числа попарно несовместных событий. □

Отметим, кроме того, еще свойства, которые являются следствиями уже перечисленных:

4) $\mathrm{P}\{\bar{A}\} = 1 - \mathrm{P}\{A\}$;

5) $\mathrm{P}\{\varnothing\} = 0$;

6) $\mathrm{P}\{A_1 \setminus A_2\} = \mathrm{P}\{A_1 \cap \bar{A}_2\}$;

7) для любых A_1 и A_2 имеет место формула

$$\mathrm{P}\{A_1 \cup A_2\} = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2\} - \mathrm{P}\{A_1 \cap A_2\}.$$

Доказательство свойства 7 просто:

$$\mathrm{P}\{A_1 \cup A_2\} = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2 \setminus A_1 \cap A_2\},$$
$$\mathrm{P}\{A_2 \setminus A_1 \cap A_2\} = \mathrm{P}\{A_2\} - \mathrm{P}\{A_1 \cap A_2\}.$$ □

Свойство 7 можно обобщить для любого числа событий $A_1, \ldots, A_n$:

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

Последняя формула доказывается по индукции, но мы ее докажем иным способом несколько позже (см. с. 28).

Отметим, что события Ω и $\varnothing$, которым соответствуют вероятности 1 и 0, называются *достоверным* и *невозможным* событиями (эти понятия имеют смысл только для данного пространства элементарных событий).

Рассмотрим совокупность $\mathrm{D}\{A_1, \ldots, A_n\}$ событий из Ω, которые обладают свойствами $\bigcup_{k=1}^{n} A_k = \Omega$ и $A_k \cap A_j = \varnothing$ при $k \neq j$. Тогда D называют *конечным разбиением пространства* Ω, а $A_1, \ldots, A_n$ — *полной группой* событий.

§ 2. Условная вероятность. Формула полной вероятности

Определение. Пусть $A, B \subset \Omega$ и $\mathrm{P}\{B\} > 0$. Тогда *условной вероятностью* события A при условии, что произошло событие B (или кратко: условной вероятностью события A при условии B), называется отношение

$$\mathrm{P}\{A|B\} = \frac{\mathrm{P}\{A \cap B\}}{\mathrm{P}\{B\}}.$$

Вероятность $\mathrm{P}\{A\}$ называют по отношению к условной вероятности $\mathrm{P}\{A|B\}$ *безусловной* вероятностью события A. Условная вероятность $\mathrm{P}\{A|B\}$ в сравнении с безусловной вероятностью $\mathrm{P}\{A\}$ показывает, как изменяются шансы события A, если поступает дополнительная информация о том, что произошло событие B. В случае, когда $\mathrm{P}\{B\} = 0$, полагают $\mathrm{P}\{A|B\} = 0$.

Пример. Производится трехкратное бросание симметричной монеты. Событие «орел» обозначаем 1, а событие «решка» обозначаем 0. Пространство Ω насчитывает $2^3 = 8$ элементарных событий:

$$\begin{aligned} \omega_1 &= (0,0,0), \quad \omega_5 = (1,1,0), \\ \omega_2 &= (1,0,0), \quad \omega_6 = (1,0,1), \\ \omega_3 &= (0,1,0), \quad \omega_7 = (0,1,1), \\ \omega_4 &= (0,0,1), \quad \omega_8 = (1,1,1). \end{aligned}$$

Рассмотрим события $A = \{$«орел» ровно 1 раз$\}$ и $B = \{$«орел» нечетное число раз$\}$. Очевидно, что $\mathrm{P}\{A\} = \frac{3}{8}$, $\mathrm{P}\{B\} = \frac{4}{8} = \frac{1}{2}$ и $\mathrm{P}\{A \,|\, B\} = \frac{3}{4}$, так как $A = \{\omega_2, \omega_3, \omega_4\}$, $B = \{\omega_2, \omega_3, \omega_4, \omega_8\}$.

Следующие две теоремы называются *теоремами умножения вероятностей* (ср. с теоремами сложения на предыдущей странице).

Теорема 1 (формула умножения вероятностей для двух событий). *Если* $A, B \subset \Omega$, *то*

$$\mathrm{P}\{A \cap B\} = \mathrm{P}\{A|B\}\mathrm{P}\{B\} = \mathrm{P}\{A\}\mathrm{P}\{B|A\}.$$

Теорема 2 (формула умножения вероятностей для n событий). *Если* $A_1, \ldots, A_n \subset \Omega$, *то*

$$\mathrm{P}\{A_1 \cap \ldots \cap A_n\} = \mathrm{P}\{A_1\}\mathrm{P}\{A_2|A_1\}\ldots\mathrm{P}\{A_k|A_1 \cap \ldots \cap A_{n-1}\}.$$

Другие варианты этой формулы получаются перестановкой $A_1, \ldots, A_n$. Доказательство теоремы 2 проводится по индукции.

Связь между безусловной вероятностью события A и условными вероятностями устанавливает формула «полной вероятности».

Теорема 3 (формула полной вероятности). *Пусть события* $B_1, \ldots, B_n$ *образуют конечное разбиение пространства* Ω, *и пусть заданы* $\mathrm{P}\{B_k\}$ *и* $\mathrm{P}\{A|B_k\}$, $k = 1, \ldots, n$. *Тогда*

$$\mathrm{P}\{A\} = \sum_{k=1}^{n} \mathrm{P}\{B_k\}\mathrm{P}\{A|B_k\}.$$

Доказательство. Очевидно, что $A = \bigcup\limits_{k=1}^{n} \{A \cap B_k\}$, а все события $A \cap B_1, \ldots, A \cap B_n$ попарно несовместны. Используя формулы сложения и формулы умножения, получаем

$$\mathrm{P}\{A\} = \mathrm{P}\left\{\sum_{k=1}^{n} A \cap B_k\right\} = \sum_{k=1}^{n} \mathrm{P}\{A \cap B_k\} = \sum_{k=1}^{n} \mathrm{P}\{A|B_k\}\mathrm{P}\{B_k\}. \qquad \square$$

В паре с формулой полной вероятности часто рассматривают так называемую «формулу Байеса».

Теорема 4 (формула Байеса). *Справедливо равенство*

$$\mathrm{P}\{B_k|A\} = \frac{\mathrm{P}\{A|B_k\}\mathrm{P}\{B_k\}}{\sum\limits_{j=1}^{n} \mathrm{P}\{A|B_j\}\mathrm{P}\{B_j\}}.$$

Доказательство. Из определения условной вероятности $\mathrm{P}\{B_k|A\}$, используя формулу умножения $\mathrm{P}\{A \cap B_k\} = \mathrm{P}\{B_k|A\}\mathrm{P}\{A\}$ и формулу полной вероятности для $\mathrm{P}\{A\}$, выводим формулу Байеса. $\square$

Замечание. Формула Байеса выглядела бы тривиальной, как обычная формула условной вероятности, если бы не ее интерпретация вместе с формулой полной вероятности в одном важном контексте. С помощью формулы полной вероятности можно предсказать появление события A в условиях, когда известны все предшествующие событию A события $B_1, \ldots, B_n$, образующие полную группу событий. Влияние событий B_k на возможность наступления события A

выражено в условных вероятностях $\mathrm{P}\{A|B_k\}$. Эти события $B_1, \ldots, B_n$ стали называть «гипотезами» о предпосылках появления события A, а их вероятности — «априорными» вероятностями гипотез. Но тогда условные вероятности $\mathrm{P}\{B_k|A\}$ можно рассматривать как вероятности гипотез $B_1, \ldots, B_n$ в случае, когда событие A произошло, и называть их поэтому «апостериорными» вероятностями гипотез. Формула Байеса явилась исторически первой попыткой решения так называемой «обратной задачи» теории вероятностей, которую в описанной простой ситуации можно было бы сформулировать так: после завершения некоторого случайного эксперимента по его исходу восстановить вероятностную модель, в рамках которой произошло событие A. Решение же «прямой задачи» теории вероятностей основано на использовании формулы полной вероятности, которая дает вероятностный прогноз события A по предшествующим ему событиям. Примеры применения формулы Байеса можно найти в книге [11].

§ 3. Независимость случайных событий

Понятие независимости — одно из центральных понятий теории вероятностей. Многие важнейшие модели теории вероятностей включают гипотезу о независимости тех или иных событий. Существует наглядное житейское представление о статистической независимости — обычно независимыми считаются события, которые кажутся несвязанными. Таковы, например, событие, заключающееся в том, что некий человек выиграл в лотерее, и событие, заключающееся в том, что у его соседа трое детей. На практике независимость достигается с помощью специальных процедур. Например, перемешивание шаров в урне или тасование карт в колоде преследуют цель сделать результат очередного испытания независимым от результатов предыдущих испытаний. Таким образом, существуют понятия независимости в интуитивном понимании и независимости в узком, но точно определенном смысле, которое в теории вероятностей вводится как допущение. При этом последнее формальное определение основывается на убеждении, что оно применимо во многих реальных ситуациях.

Определение 1. События $A_1, \ldots, A_n$ *взаимно независимы*, если выполняются 2^n соотношений:

$$\mathrm{P}\{A_1^\delta \cap \ldots \cap A_n^\delta\} = \mathrm{P}\{A_1^\delta\} \ldots \mathrm{P}\{A_n^\delta\},$$

где A_k^δ — это или A_k, или $\overline{A}_k$.

Однако некоторые из этих условий являются излишними. В действительности достаточно $2^n - n - 1$ условий.

Определение 2. События $A_1, \ldots, A_n$ *взаимно независимы*, если для любого натурального $s \geqslant 2$ и любых $i_1, \ldots, i_s$, $1 \leqslant i_1 < i_2 < \ldots < i_s \leqslant n$, выполняются соотношения

$$\mathrm{P}\left\{\bigcap_{k=1}^{s} A_{i_k}\right\} = \prod_{k=1}^{s} \mathrm{P}\{A_{i_k}\}.$$

Можно показать, что определения 1 и 2 взаимной независимости событий $A_1, \ldots, A_n$ эквивалентны. Например, при $n = 2$ события A_1 и A_2 независимы в соответствии с определением 1, если

$$\begin{aligned}
\mathrm{P}\{A_1 \cap A_2\} &= \mathrm{P}\{A_1\} \cdot \mathrm{P}\{A_2\}, \\
\mathrm{P}\{\overline{A}_1 \cap A_2\} &= \mathrm{P}\{\overline{A}_1\} \cdot \mathrm{P}\{A_2\}, \\
\mathrm{P}\{A_1 \cap \overline{A}_2\} &= \mathrm{P}\{A_1\} \cdot \mathrm{P}\{\overline{A}_2\}, \\
\mathrm{P}\{\overline{A}_1 \cap \overline{A}_2\} &= \mathrm{P}\{\overline{A}_1\} \cdot \mathrm{P}\{\overline{A}_2\}.
\end{aligned}$$

Но на самом деле независимость A_1 и A_2 определяется одним условием

$$\mathrm{P}\{A_1 \cap A_2\} = \mathrm{P}\{A_1\} \cdot \mathrm{P}\{A_2\},$$

а остальные три условия являются его следствиями: например,

$$\begin{aligned}
\mathrm{P}\{\overline{A}_1 \cap A_2\} &= \mathrm{P}\{A_2 \backslash A_1 \cap A_2\} = \mathrm{P}\{A_2\} - \mathrm{P}\{A_1 \cap A_2\} = \\
&= \mathrm{P}\{A_2\} - \mathrm{P}\{A_1\} \cdot \mathrm{P}\{A_2\} = \mathrm{P}\{A_2\}(1 - \mathrm{P}\{A_1\}) = \mathrm{P}\{A_2\} \cdot \mathrm{P}\{\overline{A}_1\}.
\end{aligned}$$

Добавим, что эти равенства, определяющие независимость событий, представляют собой формулы умножения вероятностей для независимых событий. Но в случае двух событий A_1 и A_2 определение независимости эквивалентно любому из двух условий

$$\mathrm{P}\{A_1 \,|\, A_2\} = \mathrm{P}\{A_1\}, \quad \mathrm{P}\{A_2 \,|\, A_1\} = \mathrm{P}\{A_2\},$$

если $\mathrm{P}\{A_1\} > 0$ и $\mathrm{P}\{A_2\} > 0$.

Часто совершают ошибку, полагая, что взаимная независимость трех событий равносильна условию

$$\mathrm{P}\{A_1 \cap A_2 \cap A_3\} = \mathrm{P}\{A_1\}\mathrm{P}\{A_2\}\mathrm{P}\{A_3\}.$$

Это противоречит разумному пониманию независимости, так как из независимости трех событий в таком определении не следовала бы независимость любых двух из этих трех событий. Поэтому для

взаимной независимости событий A_1, A_2, A_3 необходимо, чтобы выполнялись равенства

$$\begin{aligned} \mathrm{P}\{A_1 \cap A_2 \cap A_3\} &= \mathrm{P}\{A_1\}\mathrm{P}\{A_2\}\mathrm{P}\{A_3\}, \\ \mathrm{P}\{A_1 \cap A_2\} &= \mathrm{P}\{A_1\}\mathrm{P}\{A_2\}, \\ \mathrm{P}\{A_1 \cap A_3\} &= \mathrm{P}\{A_1\}\mathrm{P}\{A_3\}, \\ \mathrm{P}\{A_2 \cap A_3\} &= \mathrm{P}\{A_2\}\mathrm{P}\{A_3\}. \end{aligned}$$

Три последних соотношения определяют важное свойство *попарной независимости* событий A_1, A_2, A_3. В общем случае попарную независимость n событий $A_1, \ldots, A_n$ определяют C_n^2 условий $\mathrm{P}\{A_i \cap A_j\} = \mathrm{P}\{A_i\} \cdot \mathrm{P}\{A_j\}$ для каждой пары событий A_i и A_j, $i \neq j$, из $A_1, \ldots, A_n$. Известен пример С. Н. Бернштейна (см. [11, т. 1, с. 114]), который показывает, что из попарной независимости событий не следует их взаимная независимость.

Пример 1. Три грани тетраэдра окрашены соответственно в белый, красный и синий цвета, а на четвертую грань нанесены три полоски перечисленных цветов. Пусть при подбрасывании тетраэдра события A_1, A_2, A_3 означают выпадение грани с любым из трех цветов. Тогда

$$\mathrm{P}\{A_1\} = \mathrm{P}\{A_2\} = \mathrm{P}\{A_3\} = \frac{1}{2},$$

$$\mathrm{P}\{A_1 \cap A_2\} = \mathrm{P}\{A_1 \cap A_3\} = \mathrm{P}\{A_2 \cap A_3\} = \frac{1}{4},$$

$$\mathrm{P}\{A_1 \cap A_2 \cap A_3\} = \frac{1}{4}$$

и можно сделать вывод о том, что события A_1, A_2, A_3 попарно независимы, но зависимы в совокупности, так как

$$\mathrm{P}\{A_1 \cap A_2 \cap A_3\} \neq \frac{1}{8}.$$

Рассмотрим примеры независимости и зависимости событий в классических вариантах урновой модели.

Пример 2 (случайный выбор с возвращением)**.** Пусть из урны, содержащей N шаров, среди которых M белых шаров и $N - M$ черных шаров, наудачу извлечены n шаров с возвращением. Если при k-м извлечении событие A_k соответствует белому шару, а событие $\overline{A}_k$ — черному шару и $A_k^\delta = A_k$ или $A_k^\delta = \overline{A}_k$, то общий результат выбора шаров определяется событием

$$\bigcap_{k=1}^{n} A_k^\delta.$$

Если в этом произведении ровно m $(0 \leqslant m \geqslant n)$ сомножителей A_k и $n-m$ сомножителей $\overline{A}_k$, то вероятность $\mathrm{P}\Big\{\bigcap\limits_{k=1}^{n} A_k^{\delta}\Big\}$ равна $p^m(1-p)^{n-m}$ при фиксированном n и $p=\dfrac{M}{N}$. Например, при $n=m=2$ имеем

$$\mathrm{P}\{A_1 \cap A_2\} = \frac{M^2}{N^2} = \frac{M}{N}\cdot\frac{M}{N} = \mathrm{P}\{A_1\}\mathrm{P}\{A_2\},$$

что говорит о том, что события A_1 и A_2 независимы. Этот вывод справедлив для любого набора событий A_1^{δ}, ..., A_n^{δ}:

$$\mathrm{P}\{A_1^{\delta} \cap \ldots A_n^{\delta}\} = \mathrm{P}\{A_1\}\ldots\mathrm{P}\{A_n\},$$

и поэтому события A_1, ..., A_n взаимно независимы.

Пример 3 (случайный выбор без возвращения)**.** Как показано в гл. 1 на с. 9, вероятность получить ровно m белых шаров среди n шаров, извлеченных из урны по схеме выбора без возвращения, равна

$$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{C_n^m C_{N-n}^{M-m}}{C_N^M}.$$

Эти вероятности, которые зависят от числа всех шаров в урне N и числа белых шаров в урне M, задают распределение, называемое *гипергеометрическим*. Вероятность события $\bigcap\limits_{k=1}^{n} A_k^{\delta}$ может быть вычислена по общей формуле умножения

$$\mathrm{P}\Big\{\bigcap_{k=1}^{n} A_k^{\delta}\Big\} = \mathrm{P}\{A_1^{\delta}\}\mathrm{P}\{A_2^{\delta}\,|\,A_1^{\delta}\}\ldots\mathrm{P}\{A_1^{\delta}\}\mathrm{P}\{A_n^{\delta}\,|\,A_1^{\delta}\ldots|\,A_{n-1}^{\delta}\}$$

и оказывается равной $\dfrac{C_{N-n}^{M-m}}{C_N^M}$ независимо от порядка извлечения шаров по цвету. Например, $\mathrm{P}\{A_2 \cap A_1\} = \mathrm{P}\{A_1\}\mathrm{P}\{A_2\,|\,A_1\} = \dfrac{M}{N}\cdot\dfrac{M-1}{N-1}$, и потому по формуле полной вероятности

$$\mathrm{P}\{A_2\} = \mathrm{P}\{A_2\,|\,A_1\}\mathrm{P}\{A_1\} + \mathrm{P}\{A_2\,|\,\overline{A}_1\}\mathrm{P}\{\overline{A}_2\} = \\ = \frac{M}{N}\cdot\frac{M-1}{N-1} + \frac{N=M}{N}\cdot\frac{M}{N-1} = \frac{M}{N}.$$

В результате справедливо равенство $\mathrm{P}\{A_1\} = \mathrm{P}\{A_2\} = \dfrac{M}{N}$, но события A_1 и A_2 зависимы, так как $\mathrm{P}\{A_2\,|\,A_1\} = \dfrac{M-1}{N-1} \neq \dfrac{M}{N} = \mathrm{P}\{A_2\}$. Аналогично можно показать, что все события A_1, ..., A_n зависимы, хотя

при любом значении n и при любых значениях $k \neq l$ выполнены равенства $\mathrm{P}\{A_k\} = \mathrm{P}\{A_l\} = \dfrac{M}{N}$. Указанное выше свойство цепочки зависимых случайных событий называют свойством «стационарности», присущим данной вероятностной модели: вероятность некоторого события в ряду других событий не зависит от момента его наступления. Добавим, что для этой модели справедливо свойство симметрии: при любых $i \neq j$ выполнено равенство

$$\mathrm{P}\{A_i \,|\, A_j\} = \mathrm{P}\{A_j \,|\, A_i\}.$$

§ 4. Случайные величины

Случайная величина — это любая функция, заданная на пространстве элементарных событий[1] Ω: $X(\omega)$, $\omega \in \Omega$. Ее значениями обычно являются действительные числа, но могут быть векторы или какие-нибудь другие объекты.

Со всяким событием $A \in \Omega$ можно связать случайную величину, которая определяется только этим событием A. Это *индикатор* множества A:

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

Операции над индикаторами соответствуют операциям над событиями:

1) $I_{\bar{A}}(\omega) = 1 - I_A(\omega)$,

2) $I_{A_1 \cap A_2}(\omega) = I_{A_1}(\omega) \cdot I_{A_2}(\omega)$,

2′) $I_{\bigcap\limits_{k=1}^{n} A_k}(\omega) = \prod\limits_{k=1}^{n} I_{A_k}(\omega)$,

3) $I_{A_1 \cup A_2}(\omega) = I_{A_1}(\omega) + I_{A_2}(\omega) - I_{A_1 \cap A_2}(\omega)$,

3′) $I_{\bigcup\limits_{k=1}^{n} A_k}(\omega) = \sum\limits_{k=1}^{n} I_{A_k}(\omega) - \sum\limits_{(i,j)\,:\, i<j} I_{A_i \cap A_j}(\omega) + \sum\limits_{(i,j,l)\,:\, i<j<l} I_{A_i \cap A_j \cap A_l}(\omega) - \ldots$

$$\ldots + (-1)^{n-1} I_{\bigcap\limits_{k=1}^{n} A_k}(\omega).$$

При доказательстве этих свойств используются определение индикатора, формулы двойственности и метод математической индукции (для свойств 2′ и 3′).

Пусть $\mathrm{D}\{A_1, \ldots, A_s\}$ — конечное разбиение пространства Ω и $x_1, \ldots, x_s$ — различные действительные числа. Рассмотрим случайную величину

$$X(\omega) = x_1 I_{A_1}(\omega) + \ldots + x_s I_{A_s}(\omega).$$

[1] Это определение корректно для конечного пространства Ω.

Тогда если $\omega \in A_k$, $k = 1, \ldots, s$, то $X(\omega) = x_k$. Таким образом, случайная величина $X(\omega)$ принимает значения $x_1, \ldots, x_s$ на множествах $A_1, \ldots, A_s$ соответственно с вероятностями

$$\mathrm{P}\{\omega\colon X(\omega) = x_k\} = \mathrm{P}\{\omega\colon I_{A_k}(\omega) = 1\} = \mathrm{P}\{A_k\}.$$

Такие случайные величины $X(\omega)$ обычно называются *простыми*. Представление случайных величин через индикаторы событий является полезным для вывода многих свойств. Но можно к определению простых случайных величин подойти несколько иначе: пусть функция $X(\omega)$ принимает значения $x_1, \ldots, x_s$, и пусть A_k — множество всех тех ω, для которых $X(\omega) = x_k$, т. е. $A_k = \{\omega\colon X(\omega) = x_k\}$, тогда

$$\mathrm{P}\{A_k\} = \sum_{\omega \in A_k} \mathrm{P}(\omega) = \mathrm{P}\{\omega\colon X(\omega) = x_k\},$$

$$\mathrm{P}\{A_k\} = \sum_{\omega \in A_k} \mathrm{P}(\omega) = p_k.$$

Составим таблицу значений $X(\omega)$ и соответствующих им вероятностей.

$X(\omega)$	x_1	x_2	…	x_s
$\mathrm{P}\{X = x_i\}$	p_1	p_2	…	p_s

Совокупность различных возможных значений случайной величины $X(\omega)$ и соответствующих им вероятностей называется *распределением вероятностей* случайной величины $X(\omega)$. Распределение вероятностей случайной величины можно задать и формулой. В дальнейшем мы увидим, что в большинстве задач теории вероятностей не обязательно знать точную зависимость $X(\omega)$ от ω, а достаточно знать только распределение вероятностей случайной величины $X(\omega)$.

Математическое ожидание случайной величины и его свойства

Определение 1. *Математическое ожидание индикатора* $I_A(\omega)$ случайного события A равно

$$\mathrm{E}I_A(\omega) = \sum_{\omega \in A} 1 \cdot \mathrm{P}(\omega) + \sum_{\omega \in \bar{A}} 0 \cdot \mathrm{P}(\omega) = \mathrm{P}\{A\}.$$

Очевидны свойства математического ожидания индикатора:

$$\mathrm{E}(cI_A(\omega)) = c\mathrm{E}I_A(\omega) = c\mathrm{P}\{A\}$$

для любой постоянной c,

$$\mathrm{E}(I_{A_1}(\omega) + I_{A_2}(\omega)) = \mathrm{E}I_{A_1}(\omega) + \mathrm{E}I_{A_2}(\omega) = \mathrm{P}\{A_1\} + \mathrm{P}\{A_2\}.$$

Определение 2. *Математическим ожиданием* или *средним значением* простой случайной величины $X(\omega) = \sum_{k=1}^{s} x_k I_{A_k}(\omega)$ со значениями $x_1, \ldots, x_s$ называется

$$\mathrm{E}X(\omega) = \sum_{k=1}^{s} x_k \mathrm{E} I_{A_k}(\omega).$$

Вычислим математическое ожидание простой случайной величины, используя свойства математических ожиданий индикаторов:

$$\mathrm{E}X(\omega) = \sum_{k=1}^{s} x_k \mathrm{P}\{A_k\}.$$

Определение 3. *Математическим ожиданием* случайной величины $X(\omega)$ называется

$$\mathrm{E}X(\omega) = \sum_{\omega \in \Omega} X(\omega)\mathrm{P}(\omega).$$

Покажем, что для вычисления математического ожидания случайной величины и любой функции от случайной величины достаточно знать распределение вероятностей самой случайной величины.

Лемма. *Пусть случайная величина $Y(\omega)$ принимает постоянные значения $y_1, \ldots, y_m$ (среди которых могут быть и одинаковые) на множествах $A_1, \ldots, A_m$. Тогда*

$$\mathrm{E}Y(\omega) = \sum_{i=1}^{m} y_i \mathrm{P}\{A_i\}.$$

Доказательство. Так как $A_i = \{\omega \colon Y(\omega) = y_i\}$, из определения математического ожидания получаем

$$\mathrm{E}Y(\omega) = \sum_{\omega \in \Omega} Y(\omega)p(\omega) = \sum_{k=1}^{m} \sum_{\omega \in A_k} Y(\omega)p(\omega) =$$

$$= \sum_{k=1}^{m} y_k \sum_{\omega \in A_k} p(\omega) = \sum_{k=1}^{m} y_k \mathrm{P}\{A_k\}. \qquad \square$$

Теорема 1. *Если $X(\omega)$ принимает различные значения $x_1, \ldots, x_s$ с вероятностями $p_1, \ldots, p_s$, то*

$$\mathrm{E}X(\omega) = \sum_{k=1}^{s} x_k p_k, \quad p_k = \mathrm{P}\{X = x_k\}.$$

Очевидно, это то же самое выражение, что и в определении 2.

Доказательство. Введем разбиение $\{A_1, \ldots, A_s\}$, где $A_k = \{\omega \colon X(\omega) = x_k\}$, и применим лемму. $\square$

Теорема 2. *Если $X(\omega)$ принимает различные значения $x_1, \ldots, x_s$ с вероятностями $p_1, \ldots, p_s$ и $Y(\omega) = f(X(\omega))$, где функция $f(\cdot)$ взаимно однозначна, то*

$$\mathrm{E}f(X(\omega)) = \sum_{k=1}^{s} f(x_k)p_k.$$

Доказательство. Введем множества разбиения $A_k = \{\omega\colon X(\omega) = x_k\}$, на которых $Y(\omega)$ принимает постоянные значения $y_k = f(x_k)$, $k = 1, \ldots, s$, и применим лемму. □

Теоремы 1 и 2 дают формулы для вычисления математического ожидания случайной величины и любой функции от нее по ее распределению, без обращения к пространству элементарных событий, на котором случайная величина определена.

Укажем простейшие свойства математического ожидания:

1) $\mathrm{E}(cX) = c\mathrm{E}X$ при любом c;

2) $\mathrm{E}(c_1X + c_2Y) = c_1\mathrm{E}X + c_2\mathrm{E}Y$.

Доказательство свойства 2. Из определения следует, что

$$\sum_{\omega}[c_1X(\omega) + c_2 2Y(\omega)]\mathrm{P}(\omega) = \\ = c_1\sum_{\omega}X(\omega)\mathrm{P}(\omega) + c_2\sum_{\omega}Y(\omega)\mathrm{P}(\omega) = c_1\mathrm{E}X + c_2\mathrm{E}Y.$$

В сочетании свойства 1 и 2 дают свойство линейности математического ожидания. □

Это свойство распространяется на любое конечное число случайных величин:

3) если $X(\omega) \geqslant Y(\omega)$ при всех ω, то $\mathrm{E}X(\omega) \geqslant \mathrm{E}Y(\omega)$.

В частности,

3′) если $X(\omega) \geqslant 0$ при всех ω, то $\mathrm{E}X(\omega) \geqslant 0$. □

Рассмотрим несколько примеров вычисления математического ожидания.

1. Пусть случайная величина $S_n(\omega)$ имеет биномиальное распределение, т. е. $\mathrm{P}\{S_n = m\} = C_n^m p^m (1-p)^{n-m}$. Тогда

$$\mathrm{E}S_n(\omega) = \sum_{m=0}^{n} mC_n^m p^m (1-p)^{n-m} = np\sum_{m=0}^{n-1} C_{n-1}^m p^m (1-p)^{n-1-m} = np.$$

Укажем другой способ вычисления математического ожидания биномиального распределения. Пусть в урновой модели с возвращением событие A_k, состоит в том, что k-й вынутый шар белый, $\overline{A}_k$ —

событие, противоположное A_k, $k=1, 2, \ldots, n$. Введем индикатор

$$I_{A_k}(\omega) = \begin{cases} 1, & \text{если } \omega \in A_k, \\ 0, & \text{если } \omega \notin A_k, \end{cases}$$

при этом $\mathrm{E}I_{A_k}(\omega) = \mathrm{P}\{A_k\}$. Рассмотрим случайную величину $S_n(\omega) = I_{A_1}(\omega) + \ldots + I_{A_n}(\omega)$. Очевидно, что $S_n(\omega)$ — это число белых шаров среди извлеченных n шаров и $S_n(\omega)$ имеет биномиальное распределение. По свойству математического ожидания индикаторов

$$\mathrm{E}S_n(\omega) = \mathrm{E}I_{A_1}(\omega) + \ldots + \mathrm{E}I_{A_n}(\omega) = \mathrm{P}\{A_1\} + \ldots + \mathrm{P}\{A_n\} = np,$$

так как при выборе с возвращением $\mathrm{P}\{A_1\} = \ldots = \mathrm{P}\{A_n\} = p$.

2. Для того чтобы освоить вероятностный смысл математического ожидания в примере 1, найдем наиболее вероятное значение случайной величины с биномиальным распределением, т. е. такое значение m^*, для которого вероятность $\mathrm{P}\{S_n = m\}$ максимальна при фиксированных значениях n и p. Решим для этого систему неравенств

$$\begin{cases} \dfrac{C_n^m p^m (1-p)^{n-m}}{C_n^{m+1} p^{m+1} (1-p)^{n-m-1}} \geqslant 1, \\ \dfrac{C_n^m p^m (1-p)^{n-m}}{C_n^{m-1} p^{m-1} (1-p)^{n-m+1}} \geqslant 1. \end{cases}$$

Получим

$$\begin{cases} \dfrac{(n-m+1)p}{m(1-p)} \geqslant 1, \\ \dfrac{(m+1)(1-p)}{(n-m)p} \geqslant 1, \end{cases} \quad \text{или} \quad \begin{cases} (n+1)p \geqslant m, \\ m \geqslant np - (1-p). \end{cases}$$

Окончательно получим неравенство для m^*:

$$np - (1-p) \leqslant m^* \leqslant np + p \quad \text{или} \quad (n+1)p - 1 \leqslant m^* \leqslant (n+1)p.$$

Теперь можно сделать вывод: если $(n+1)p$ — целое число, то $m_1^* = (n+1)p - 1$ и $m_2^* = (n+1)p$; в ином случае $(n+1)p - 1 < m^* < (n+1)p$, или, что то же самое, $m^* = [(n+1)p]$, где $[a]$ — целая часть числа a. Получаем оценку $|m^* - np| \leqslant \min(p, 1-p)$, что указывает на близость математического ожидания и наиболее вероятного значения случайной величины с биномиальным распределением.

3. Рассмотрим урновую модель случайного выбора без возвращения. Как и в предыдущем случае, событие A_k состоит в том, что при

k-м извлечении появляется белый шар, а $\bar{A}_k$ — противоположное событие при $k=1,\ldots,n$, и пусть $I_{A_k}(\omega)$ — индикатор события A_k, так что $\mathrm{E}I_{A_k}(\omega)=\mathrm{P}\{A_k\}$. Как доказано в примере 3 на с. 22, эти вероятности $\mathrm{P}\{A_k\}$ при любом значении k одинаковы и равны $p=\dfrac{M}{N}$, где N — общее число шаров в урне, а M — число белых шаров, т. е. при любом $k=1,\ldots,n$ выполнено равенство

$$\mathrm{E}I_{A_k}(\omega)=p.$$

В рассматриваемой модели случайная величина $S'_n(\omega)=I_{A_1}(\omega)+\ldots$ $\ldots+I_{A_n}(\omega)$ имеет гипергеометрическое распределение, и поэтому при фиксированных значениях N, M, n и $m=0,\ldots,\min(n,m)$ имеем

$$\mathrm{E}S'_n(\omega)=\sum_m m\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Но это математическое ожидание легче вычислить по формуле

$$\mathrm{E}S'_n(\omega)=\mathrm{E}I_{A_1}(\omega)+\ldots+\mathrm{E}I_{A_n}(\omega)$$

и получить $\mathrm{E}S'_n(\omega)=\dfrac{M}{N}+\ldots+\dfrac{M}{N}=n\dfrac{M}{N}=np$. Таким образом, математическое ожидание случайной величины S'_n совпадает с математическим ожиданием величины S_n в модели выбора с возвращением.

4. Генуэзская лотерея (см. пример 3 гл. 1 на с. 10). Обозначим ставку игрока в лотерее через a. В случае выигрыша игрок получает

$15a$	с вероятностью	$p_1=1/18$,
$270a$	с вероятностью	$p_2=2/201$,
$5500a$	с вероятностью	$p_3=1/11\,748$,
$75000a$	с вероятностью	$p_4=1/511\,038$,
$1\,000\,000a$	с вероятностью	$p_5=1/43\,949\,268$.

Подсчитаем среднее значение «чистого» выигрыша E_k при ставке на k номеров:

$$\mathrm{E}_1=[15a\cdot p_1+0\cdot(1-p_1)]-a=-\frac{1}{6}a<0,$$

$$\mathrm{E}_2=[270a\cdot p_2+0\cdot(1-p_2)]-a=-\frac{29}{89}a<0$$

и т. д. Оказывается, все $\mathrm{E}_k<0$. Таким образом, в среднем при большом числе игр в выигрыше оказываются устроители лотереи.

5. Как было обещано, выведем общую формулу для вероятности $\mathrm{P}\{A_1\cup\ldots\cup A_n\}$. Обозначим $A=A_1\cup\ldots\cup A_n$ и, пользуясь свойствами

индикаторов событий, получим

$$I_{A_i}(\omega) = 1 - \prod_{i=1}^{n} I_{\overline{A}_i}(\omega) = 1 - \prod_{i=1}^{n}(1 - I_{A_i}(\omega)) =$$

$$= 1 - \Bigg\{1 - \sum_{i=1}^{n} I_{A_i}(\omega) + \sum_{1 \leqslant i < j \leqslant n} I_{A_i}(\omega) I_{A_j}(\omega) -$$

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

Раскроем скобки и перейдем к математическим ожиданиям в левой и правой частях. В силу того что $\mathrm{E} I_A(\omega) = \mathrm{P}\{A\}$, получим

$$\mathrm{P}\{A_1 \cup \ldots \cup A_n\} = \sum_{i=1}^{n} \mathrm{P}\{A_i\} - \sum_{1 \leqslant i < j \leqslant n} \mathrm{P}\{A_i \cap A_j\} +$$

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

(см. [11, т. 1, гл. IV]).

Можно также найти вероятность события {наступает ровно k событий из числа всех событий $A_1, \ldots, A_n$}, $1 \leqslant k \leqslant n$.

Моменты случайной величины

Определение 4. *Моментом порядка* $r \in N$ случайной величины $X(\omega)$ называется математическое ожидание $\mathrm{E}X^r$; для любого действительного числа a математическое ожидание $\mathrm{E}(X-a)^r$ называется *моментом порядка* r *относительно* a; при $a = \mathrm{E}X$ математическое ожидание $\mathrm{E}(X - \mathrm{E}X)^r$ называется *центральным моментом* порядка r.

Математическое ожидание случайной величины $\mathrm{E}X$ — это момент порядка $r = 1$. Кроме математического ожидания особую роль играет центральный момент второго порядка, или дисперсия случайной величины.

Определение 5. *Дисперсия* случайной величины $X(\omega)$ — это математическое ожидание

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

Дисперсия определяет меру отклонения значений случайной величины от ее среднего значения. Другими словами, дисперсия дает представление в среднем о степени концентрации распределения вблизи математического ожидания. Математическое ожидание и дисперсия служат самыми важными числовыми характеристиками распределения вероятностей случайных величин.

Во многих случаях вместо определения можно для вычисления дисперсии применять удобную формулу. Используя свойства математического ожидания, получаем

$$\mathrm{D}X(\omega) = \mathrm{E}(X - \mathrm{E}X)^2 = \mathrm{E}[X^2 - 2X \cdot \mathrm{E}X + (\mathrm{E}X)^2] = \\ = \mathrm{E}X^2 - 2(\mathrm{E}X)^2 + (\mathrm{E}X)^2 = \mathrm{E}X^2 - (\mathrm{E}X)^2.$$

Для применения этой формулы необходимо вычислить $\mathrm{E}X^2$ — момент порядка 2 случайной величины $X(\omega)$. Например, дисперсию индикатора случайного события A легко вычислить по полученной формуле:

$$\mathrm{E}I_A = \mathrm{E}I_A^2 - (\mathrm{E}I_A)^2 = \mathrm{P}\{A\} - \mathrm{P}^2\{A\} = \mathrm{P}\{A\}\mathrm{P}\{\overline{A}\},$$

поскольку $\{I_A^2 = 1\} = \{I_A = 1\}$ и $\mathrm{P}\{I_A^2 = 1\} = \mathrm{P}\{I_A = 1\} = \mathrm{P}\{A\}$.

Перечислим основные свойства дисперсии $\mathrm{D}X(\omega)$:

1) $\mathrm{D}X(\omega) \geqslant 0$;

2) $\mathrm{D}X(\omega) = 0$ тогда и только тогда, когда $\mathrm{P}\{X(\omega) = c\} = 1$ (доказательство основано на свойстве математического ожидания: если $X(\omega) \geqslant 0$ и $\mathrm{E}X(\omega) = 0$, то $X(\omega) = 0$ с вероятностью 1);

3) $\mathrm{D}(cX(\omega)) = c^2\mathrm{D}X(\omega)$, где c — постоянная;

4) $\mathrm{D}(X(\omega) + c) = \mathrm{D}X(\omega)$, где c — постоянная;

5) $\mathrm{D}X(\omega) = \min\limits_{a \in \mathbb{R}} \mathrm{E}(X(\omega) - a)^2$.

Наряду с дисперсией для измерения отклонения значений случайной величины от математического ожидания используется так называемое *стандартное отклонение*:

$$\sigma_X = +\sqrt{\mathrm{D}X(\omega)}.$$

Во многих случаях для вычисления дисперсий различных распределений нужно знать, как вычислять дисперсию суммы или линейной комбинации случайных величин, см. теорему 5 на с. 38.

§ 5. Совместное распределение вероятностей и независимость случайных величин

Пусть случайные величины $X_1(\omega), \ldots, X_n(\omega)$ заданы на одном пространстве элементарных событий, и пусть $\mathscr{X}_1, \ldots, \mathscr{X}_n$ — конечные множества их значений соответственно.

Определение 1. *Совместным распределением* вероятностей случайных величин называют совокупность вероятностей

$$\mathrm{P}\{\omega\colon X_1(\omega) = x^{(1)}, \ldots, X_n(\omega) = x^{(n)}\}$$

для всех возможных значений $x^{(1)} \in \mathscr{X}_1, \ldots, x^{(n)} \in \mathscr{X}_n$ этих величин, если сумма всех этих вероятностей равна 1.

Совместное распределение случайных величин позволяет рассматривать различные случайные величины в рамках одной вероятностной модели. Совместное распределение обладает максимальной информацией в отношении распределения каждой из этих случайных величин, числовых характеристик этих случайных величин и связей между этими величинами.

Рассмотрим для примера совместное распределение двух случайных величин

$$p_{ij} = \mathrm{P}\{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}\}, \quad i = 1, \ldots, s, \quad j = 1, \ldots, r,$$

где $x_1^{(1)}, \ldots, x_s^{(1)}$ — значения величины X_1, $x_1^{(2)}, \ldots, x_r^{(2)}$ — значения величины X_2, при этом $\sum\limits_{i,j} p_{ij} = 1$. Найдем распределение каждой из этих случайных величин. Обозначим $p_{i\cdot} = \mathrm{P}\{X_1 = x_i^{(1)}\}$ и $p_{\cdot j} = \mathrm{P}\{X_2 = x_j^{(2)}\}$. Поскольку для любого значения i имеет место равенство

$$p_{i\cdot} = \mathrm{P}\{X_1 = x_i^{(1)}\} = \sum_{j=1}^{r} \mathrm{P}\{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}\} = \sum_{j=1}^{r} p_{ij},$$

эта формула задает распределение случайной величины X_1. Аналогично можно определить распределение величины X_2. По отношению к совместному распределению случайных величин X_1 и X_2 полученные распределения называют *частными* или *маргинальными* распределениями. Все сказанное справедливо для совместного распределения любого конечного числа случайных величин. Рассмотрим еще один пример.

Пример 1. Определим для случайных величин $X_1, \ldots, X_n$ совместное распределение следующей формулой:

$$\mathrm{P}\{X_1 = x_1, \ldots, X_n = x_n\} = p^{x_1 + \ldots + x_n}(1-p)^{n - (x_1 + \ldots + x_n)},$$

где $x_k = 1$ или $x_k = 0$, а $p \in [0, 1]$. Найдем частное распределение величины X_1. При фиксированных n и x_1 вычислим

$$\mathrm{P}\{X_1 = x_1\} = \\ = \sum_{x_1, \ldots, x_n \,:\, x_2 + \ldots + x_n = n - x_1} p^{x_1 + \ldots + x_n}(1-p)^{n - (x_1 + \ldots + x_n)} = p^{x_1}(1-p)^{1 - x_1}$$

и аналогично $\mathrm{P}\{X_k = x_k\} = p^{x_k}(1-p)^{1-x_k}$ для любого k.

Таким образом, каждая случайная величина имеет одинаковое распределение: $\mathrm{P}\{X_k = 1\} = p$ и $\mathrm{P}\{X_k = 0\} = 1 - p$, а последовательность $X_1, \ldots, X_n$ есть не что иное, как схема испытаний Бернулли с вероятностью «удачи» p.

Важно отметить, что по распределению каждой из случайных величин в общем случае получить совместное распределение невозможно.

Независимость случайных величин

Понятие независимости или стохастической независимости случайных величин является одной из наиболее существенных характеристик вероятностных моделей.

Определение 2. Рассмотрим случайные величины $X_1, \ldots, X_n$ и обозначим через $\{x^{(k)}\}$ множество возможных значений случайной величины X_k, $k = 1, \ldots, n$. Величины $X_1, \ldots, X_n$ называются *взаимно независимыми*, если при любом наборе $x^{(1)}, \ldots, x^{(n)}$ выполняется равенство

$$\mathrm{P}\{\omega\colon X_1(\omega) = x^{(1)}, \ldots, X_n(\omega) = x^{(n)}\} = \\ = \mathrm{P}\{\omega\colon X_1(\omega) = x^{(1)}\} \cdot \ldots \cdot \mathrm{P}\{\omega\colon X_n(\omega) = x^{(n)}\}.$$

Легко установить связь между независимостью событий и случайных величин. Рассмотрим случайные величины $I_{A_1}, \ldots, I_{A_n}$, являющиеся индикаторами независимых событий $A_1, \ldots, A_n$. Тогда

$$\mathrm{P}\{I_{A_1} = 1, \ldots, I_{A_n} = 1\} = \mathrm{P}\{A_1 \cap \ldots \cap A_n\} = \\ = \mathrm{P}\{A_1\} \cdot \ldots \cdot \mathrm{P}\{A_n\} = \mathrm{P}\{I_{A_1} = 1\} \cdot \ldots \cdot \mathrm{P}\{I_{A_n} = 1\}.$$

Аналогично выводятся и остальные $2^n - 1$ соотношений.

Пример 2. Пусть в урне 10 шаров с номерами 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Пусть при выборе с возвращением двух шаров: номер первого шара i, номер второго шара j. Очевидно, что $\mathrm{P}(\omega) = \frac{1}{100}$. Рассмотрим значения двух случайных величин $X_1(\omega) = i$ и $X_2(\omega) = j$. Покажем, что X_1 и X_2 независимы в смысле данного определения, т. е. что

$$\mathrm{P}\{X_1 = i, X_2 = j\} = \mathrm{P}\{X_1 = i\}\mathrm{P}\{X_2 = j\}.$$

Поскольку i и j принимают значения от 0 до 9, получаем, что $\mathrm{P}\{X_1 = i\} = \frac{1}{10}$ и $\mathrm{P}\{X_2 = j\} = \frac{1}{10}$ для любых i и j. Следовательно,

$$\mathrm{P}\{X_1 = i\}\mathrm{P}\{X_2 = j\} = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100}$$

и

$$\mathrm{P}\{X_1 = i, X_2 = j\} = \mathrm{P}\{\omega\colon X_1 = i, X_2 = j\} = \mathrm{P}(\omega) = \frac{1}{100},$$

что указывает на независимость X_1 и X_2.

Пример 3. При выборе без возвращения двух шаров всего имеется $10 \cdot 9 = 90$ пар (i, j), $i \neq j$. Рассмотрим событие $\{X_1 = i\}$. Очевидно, $\mathrm{P}\{X_1 = i\} = \frac{1}{10}$ для любого j. Несложные подсчеты приводят к заключению, что $\mathrm{P}\{X_2 = j\} = \frac{1}{10}$ для любого j. Ясно, что X_1 и X_2 не будут независимыми, так как номер 2-го шара определяется номером 1-го шара. Проверим это:

$$\mathrm{P}\{X_1 = i\}\mathrm{P}\{X_2 = j\} = \frac{1}{10} \cdot \frac{1}{10} = \frac{1}{100},$$

$$\mathrm{P}\{X_1 = i, X_2 = j\} = \begin{cases} 0 & \text{при } i = j, \\ \frac{1}{9} & \text{при } i \neq j. \end{cases}$$

Таким образом, $X_1(\omega)$ и $X_2(\omega)$ не являются независимыми в смысле нашего определения.

Мы видим, что независимость случайных величин выводится из независимости случайных событий, определяющих данные случайные величины.

Пример 4. Было показано (см. пример 1 на с. 31), что если совместное распределение случайных величин $X_1, \ldots, X_n$ выражается формулой

$$\mathrm{P}\{X_1 = x_1, \ldots, X_n = x_n\} = p^{x_1 + \ldots + x_n}(1-p)^{n - (x_1 + \ldots + x_n)},$$

где $x_k = 1$ или $x_k = 0$, то распределение любой случайной величины X_k определяется равенством $p^{x_k}(1-p)^{1-x_k}$. Очевидно соотношение, выражающее совместное распределение величин $X_1, \ldots, X_n$ через частные распределения:

$$\mathrm{P}\{X_1 = x_1, \ldots, X_n = x_n\} = p^{x_1 + \ldots + x_n}(1-p)^{n - (x_1 + \ldots + x_n)} =$$
$$= p^{x_1}(1-p)^{1-x_1} \cdot \ldots \cdot p^{x_n}(1-p)^{1-x_n} = \mathrm{P}\{X_1 = x_1\} \ldots \mathrm{P}\{X_n = x_n\}.$$

Поскольку это равенство выполняется для любых значений $x_1, \ldots, x_n$, можно сделать вывод о том, что случайные величины $X_1, \ldots, X_n$

независимы. Совокупность этих случайных величин задает известную вероятностную модель — схему Бернулли независимых испытаний. Свойство независимости случайных величин предоставляет нам единственную возможность найти совместное распределение случайных величин по распределению каждой из них.

Пример 5. Рассмотрим модель случайного выбора без возвращения. Пусть событие $\{X_k = 1\}$ состоит в том, что k-й извлеченный шар белый, а событие $\{X_k = 0\}$ — в том, что k-й извлеченный шар черный. Тогда, как выведено на с. 28, распределение величин X_k определяется формулами

$$\mathrm{P}\{X_k = 1\} = \frac{M}{N}, \quad \mathrm{P}\{X_k = 0\} = \frac{N-M}{N},$$

где N — общее число шаров, а M — число белых шаров. Как следует из примера на предыдущей странице, в этой модели совместное распределение любых двух случайных величин X_k и X_l, $k \neq l$, вычисляется по формулам

$$\begin{aligned}
\mathrm{P}\{X_k = 1, X_l = 1\} &= \frac{M}{N} \cdot \frac{M-1}{N-1}, \\
\mathrm{P}\{X_k = 1, X_l = 0\} &= \frac{M}{N} \cdot \frac{N-M}{N-1}, \\
\mathrm{P}\{X_k = 0, X_l = 1\} &= \frac{N-M}{N} \cdot \frac{M}{N-1}, \\
\mathrm{P}\{X_k = 0, X_l = 0\} &= \frac{N-M}{N} \cdot \frac{N-M-1}{N-1}.
\end{aligned}$$

Очевидно, что вероятность получить при двух испытаниях белый и черный шары

$$\mathrm{P}\{X_k = 1, X_l = 0\} = \mathrm{P}\{X_k = 0, X_l = 1\} = \frac{M}{N} \cdot \frac{N-M}{N-1}$$

не зависит от порядка их извлечения и не совпадает с произведением вероятностей

$$\mathrm{P}\{X_k = 1\}\mathrm{P}\{X_l = 1\} = \frac{M}{N} \cdot \frac{N-M}{N},$$

что говорит нам о зависимости величин X_k и X_l при $k \neq l$. При случайном выборе n шаров совместное распределение величин $X_1, \ldots, X_n$ определяется лишь числом извлеченных белых и черных шаров и не зависит от порядка их извлечения. Поскольку любые две величины X_k и X_l зависимы, в совокупности случайные величины тоже $X_1, \ldots, X_n$ оказываются зависимыми.

Свойства независимых случайных величин

Пусть $X_1, \ldots, X_n$ — взаимно независимые случайные величины.

1. Любые m $(m < n)$ из них $X_{i_1}, \ldots, X_{i_m}$ также взаимно независимы.

2. Пусть $f_1(x), \ldots, f_n(x)$ — некоторые действительные функции. Рассмотрим новые случайные величины $Y_1 = f_1(X_1), \ldots, Y_n = f_n(X_n)$. Тогда случайные величины $Y_1, \ldots, Y_n$ также независимы.

3. Разобьем набор случайных величин $X_1, \ldots, X_n$ на две группы: $X_1, \ldots, X_m$ и $X_{m+1}, \ldots, X_n$ при любом $m < n$. Рассмотрим две функции $Z_1 = f(X_1, \ldots, X_m)$ и $Z_2 = g(X_{m+1}, \ldots, X_n)$. Тогда Z_1 и Z_2 снова являются независимыми случайными величинами.

Доказательство этих утверждений несложно и заключается в аккуратной проверке необходимого количества соотношений. Общие для всех утверждений принципы доказательства видны на следующем примере.

Пусть X_1, X_2, X_3 взаимно независимы. Покажем, что X_1 и X_2 также независимы. В силу дистрибутивности операции объединения

$$\bigcup_k \{A \cap B_k\} = A \cap \Big\{\bigcup_k B_k\Big\}$$

имеем

$$\bigcup_k \{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}, X_3 = x_k^{(3)}\} = \{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}\}.$$

Отсюда следует, что

$$\sum_k \mathrm{P}\{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}, X_3 = x_k^{(3)}\} = \mathrm{P}\{X_1 = x_i^{(1)}, X_2 = x_j^{(2)}\}.$$

С другой стороны, в силу независимости X_1, X_2, X_3 имеем

$$\begin{aligned}\sum_k \mathrm{P}\{X_1 = x_i^{(1)}\}\mathrm{P}\{X_2 = x_j^{(2)}\}\mathrm{P}\{X_3 = x_k^{(3)}\} &= \\ = \mathrm{P}\{X_1 = x_i^{(1)}\}\mathrm{P}\{X_2 = x_j^{(2)}\} \cdot \sum_k \mathrm{P}\{X_3 = x_k^{(3)}\} &= \\ = \mathrm{P}\{X_1 = x_i^{(1)}\}\mathrm{P}\{X_2 = x_j^{(2)}\}, &\end{aligned}$$

откуда следует, что X_1 и X_2 независимы.

Свойства математического ожидания и дисперсии независимых случайных величин

Теорема 1. *Пусть X и Y — независимые случайные величины, и пусть* $\mathrm{E}X$, $\mathrm{E}Y$ *и* $\mathrm{E}XY$ *— соответствующие математические ожидания. Тогда*

$$\mathrm{E}XY = \mathrm{E}X \cdot \mathrm{E}Y.$$

Доказательство. Докажем теорему для случая, когда число значений случайных величин X и Y конечно. Пусть x_i и y_j $(i=1,\ldots,n;\ j=1,\ldots,m)$ — возможные значения величин X и Y. Введем множества

$$A_i=\{\omega\colon X(\omega)=x_i\},\quad B_j=\{\omega\colon Y(\omega)=y_j\}.$$

Очевидно, что на множествах $A_i\cap B_j$ выполнены равенства $X(\omega)=x_i$ и $Y(\omega)=y_j$ и поэтому $XY(\omega)=x_iy_j$. В силу независимости X и Y получаем

$$\mathrm{P}\{A_i\cap B_j\}=\mathrm{P}\{X=x_i,\,Y=y_j\}=\mathrm{P}\{X=x_i\}\mathrm{P}\{Y=y_j\}=\mathrm{P}\{A_i\}\mathrm{P}\{B_j\},$$

т. е. A_i и B_i независимы. Имеем

$$\begin{aligned}\mathrm{E}XY&=\sum_{\omega}X(\omega)Y(\omega)\mathrm{P}\{\omega\}=\sum_i\sum_j\Big[\sum_{\omega\in A_i\cap B_j}X(\omega)Y(\omega)\mathrm{P}\{\omega\}\Big]=\\&=\sum_i\sum_j x_iy_j\Big[\sum_{\omega\in A_i\cap B_j}\mathrm{P}\{\omega\}\Big]=\sum_i\sum_j x_iy_j\mathrm{P}\{A_i\cap B_j\}=\\&=\sum_i\sum_j x_iy_j\mathrm{P}\{A_i\}\mathrm{P}\{B_j\}=\Big[\sum_i x_i\mathrm{P}\{A_i\}\Big]\Big[\sum_j y_j\mathrm{P}\{B_j\}\Big]=\mathrm{E}X\cdot\mathrm{E}Y.\ \square\end{aligned}$$

Замечание. Теорема распространяется на любое число взаимно независимых случайных величин:

$$\mathrm{E}(X_1\ldots X_n)=\mathrm{E}X_1\ldots\mathrm{E}X_n.$$

Попарной независимости для справедливости этого равенства недостаточно.

Теорема 2. *Если величины X и Y независимы и имеют дисперсии* $\mathrm{D}X$ *и* $\mathrm{D}Y$, *то*

$$\mathrm{D}(X+Y)=\mathrm{D}X+\mathrm{D}Y.$$

Доказательство. Используя теорему 1, получаем

$$\begin{aligned}\mathrm{D}(X+Y)&=\mathrm{E}(X+Y-\mathrm{E}(X+Y))^2=\mathrm{E}[(X-\mathrm{E}X)+(Y-\mathrm{E}Y)]^2=\\&=\mathrm{E}(X-\mathrm{E}X)^2+2\mathrm{E}\{(X-\mathrm{E}X)(Y-\mathrm{E}Y)\}+\mathrm{E}(Y-\mathrm{E}Y)^2=\\&=\mathrm{D}X+2\mathrm{E}(X-\mathrm{E}X)\cdot\mathrm{E}(Y-\mathrm{E}Y)+\mathrm{D}Y=\\&=\mathrm{D}X+\mathrm{D}Y+2(\mathrm{E}X-\mathrm{E}X)(\mathrm{E}Y-\mathrm{E}Y)=\mathrm{D}X+\mathrm{D}Y,\end{aligned}$$

что и требовалось доказать. $\square$

Замечание. Здесь было использовано лишь следствие независимости, а именно то, что по теореме 1 выполнено равенство

$$\mathrm{E}\{(X-\mathrm{E}X)(Y-\mathrm{E}Y)\}=\mathrm{E}(X-\mathrm{E}X)\cdot\mathrm{E}(Y-\mathrm{E}Y).$$

Определение 3. Выражение $\mathrm{E}\{(X-\mathrm{E}X)(Y-\mathrm{E}Y)\}$ называется *ковариацией* двух случайных величин X и Y и обозначается $\operatorname{cov}(X,Y)$.

Легко показать, что $\operatorname{cov}(X,Y)=\mathrm{E}XY-\mathrm{E}X\cdot\mathrm{E}Y$. Пользуясь теоремой 1, можно утверждать, что для независимых случайных величин ковариация равна нулю.

Определение 4. Отношение $r(X,Y)=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}X\cdot\mathrm{D}Y}}$ носит название *коэффициента корреляции*[1] случайных величин X и Y.

Определение 5. Случайные величины X_1 и X_2, для которых ковариация или, что равнозначно, коэффициент корреляции равны нулю, называются *некоррелированными*.

Из предыдущего следует, что независимые величины некоррелированы. Однако обратное утверждение в общем случае неверно. Рассмотрим два примера.

Пример 6. Пусть случайная величина X распределена по закону

$$\mathrm{P}\{X=-1\}=\mathrm{P}\{X=1\}=\frac{1}{4},\quad \mathrm{P}\{X=0\}=\frac{1}{2}$$

с $\mathrm{E}X=0$. Тогда случайная величина X^2 имеет распределение

$$\mathrm{P}\{X^2=0\}=\mathrm{P}\{X^2=1\}=\frac{1}{2}$$

с $\mathrm{E}X^2=\dfrac{1}{2}$. Очевидно, что $\mathrm{E}(X\cdot X^2)=\mathrm{E}X^3=0$ и $\mathrm{E}X\cdot\mathrm{E}X^2=0$, так что величины X и X^2 некоррелированы, но связаны функциональной зависимостью.

Пример 7. Пусть I_1 и I_2 — индикаторы двух случайных событий A_1 и A_2. Пусть величины I_1 и I_2 некоррелированы, т. е.

$$\operatorname{cov}(I_1,I_2)=\mathrm{E}(I_1\cdot I_2)-\mathrm{E}I_1\cdot\mathrm{E}I_2=0,$$

что равносильно равенству $\mathrm{E}I_1I_2=\mathrm{E}I_1\cdot\mathrm{E}I_2$. Так как $\mathrm{E}I_1I_2=\mathrm{P}\{A_1\cap A_2\}$, $\mathrm{E}I_1=\mathrm{P}\{A_1\}$ и $\mathrm{E}I_2=\mathrm{P}\{A_2\}$, это равенство означает, что события A_1 и A_2 независимы, стало быть, независимы случайные величины I_1 и I_2:

$$\mathrm{P}\{I_1=1,I_2=1\}=\mathrm{P}\{A_1\cap A_2\}=\mathrm{E}I_1I_2=\mathrm{E}I_1\cdot\mathrm{E}I_2=$$
$$=\mathrm{P}\{A_1\}\mathrm{P}\{A_2\}=\mathrm{P}\{I_1=1\}\mathrm{P}\{I_2=1\}.$$

Теорема 3. *В общем случае*

$$\mathrm{D}(X_1+X_2)=\mathrm{D}X_1+\mathrm{D}X_2+2\operatorname{cov}(X_1,X_2),$$

[1] Точнее, коэффициента корреляции Пирсона.

или

$$\mathrm{D}(X_1 + X_2) = \mathrm{D}X_1 + \mathrm{D}X_2 + 2\sqrt{\mathrm{D}X_1} \cdot \sqrt{\mathrm{D}X_2} \cdot r(X_1, X_2),$$

где $r(X_1, X_2)$ *— коэффициент корреляции между* X_1 *и* X_2.

Теорема 4. *Если случайные величины* X_1 *и* X_2 *некоррелированы, то*

$$\mathrm{D}(X_1 + X_2) = \mathrm{D}X_1 + \mathrm{D}X_2.$$

Доказательство теорем 3 и 4 следует из теоремы 2 и подтверждает, что в теореме 2 было использовано лишь следствие независимости. □

Следует отметить, что ковариация $\mathrm{cov}(X_1, X_2)$ является моментом совместного распределения случайных величин X_1 и X_2, а именно центральным «смешанным» моментом порядка 2, т. е. моментом того же порядка, что и дисперсия, и, более того, $\mathrm{cov}(X_1, X_1) = \mathrm{D}X_1$. Также следует иметь в виду, что и ковариация, и коэффициент корреляции являются характеристиками двух случайных величин, поэтому при рассмотрении множества случайных величин, связанных совместным распределением, нужно говорить о попарной некоррелированности или, наоборот, о попарной корреляционной зависимости.

Перейдем к обобщению рассмотренных теорем.

Теорема 5. *Пусть случайные величины* $X_1, \ldots, X_n$ *распределены с дисперсиями* $\mathrm{D}X_1, \ldots, \mathrm{D}X_n$ *и попарными ковариациями* $\mathrm{cov}(X_i, X_j)$. *Тогда*

$$\mathrm{D}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \mathrm{D}X_k + 2 \sum_{(i,j):\, i<j} \mathrm{cov}(X_i, X_j).$$

Если $r(X_i, X_j)$ *— коэффициенты корреляции между* X_i *и* X_j, *то*

$$\mathrm{D}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \mathrm{D}X_k + 2 \sum_{(i,j):\, i<j} \sqrt{\mathrm{D}X_i} \cdot \sqrt{\mathrm{D}X_j} \cdot r(X_i, X_j).$$

Если величины $X_1, \ldots, X_n$ *некоррелированы, то*

$$\mathrm{D}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \mathrm{D}X_k.$$

Отметим, что коэффициент корреляции представляет собой нормированное значение ковариации. Рассмотрим те преимущества, которые мы получаем при переходе от ковариации к коэффициенту корреляции. В следующей теореме установим свойства коэффициента корреляции.

Теорема 6. *Пусть $r(X_1, X_2)$ — коэффициент корреляции случайных величин X_1 и X_2. Тогда*

1) $-1 \leqslant r(X_1, X_2) \leqslant 1$ *для любых случайных величин X_1 и X_2,*

2) *если случайные величины X_1 и X_2 независимы, то $r(X_1, X_2) = 0$,*

3) *если $|r(X_1, X_2)| = 1$, то случайные величины X_1 и X_2 связаны линейной зависимостью,*

4) *при линейном преобразовании случайных величин коэффициент корреляции остается неизменным; в частности, если $X_1^* = \dfrac{X_1 - \mathrm{E}X_1}{\sqrt{\mathrm{D}X_1}}$ и $X_2^* = \dfrac{X_2 - \mathrm{E}X_2}{\sqrt{\mathrm{D}X_2}}$, то*

$$r(X_1^*, X_2^*) = r(X_1, X_2).$$

Доказательство. 1. По неравенству Коши—Буняковского

$$|\operatorname{cov}(X_1, X_2)| \leqslant \sqrt{\mathrm{D}X_1} \cdot \sqrt{\mathrm{D}X_2},$$

а это означает, что $|r(X_1, X_2)| \leqslant 1$.

2. Это свойство доказано на с. 37.

3. Поскольку неравенство Коши—Буняковского обращается в равенство тогда и только тогда, когда $c_1X_1 + c_2X_2 = C$ при c_1 и c_2, не равных нулю одновременно, X_1 и X_2 связаны линейной зависимостью, например, $X_2 = kX_1 + b$, $k \neq 0$. Если $r(X_1, X_2) = -1$, то $k < 0$; если $r(X_1, X_2) = 1$, то $k > 0$.

4. Так как $\mathrm{E}X_1^* = 0$, $\mathrm{E}X_2^* = 0$, $\mathrm{D}X_1^* = 1$, $\mathrm{D}X_2^* = 1$, получаем, что

$$r(X_1^*, X_2^*) = \frac{\mathrm{E}X_1^*X_2^* - \mathrm{E}X_1^*\mathrm{E}X_2^*}{\sqrt{\mathrm{D}X_1^* \cdot \mathrm{D}X_2^*}} = \mathrm{E}X_1^*X_2^* = \frac{\operatorname{cov}(X_1, X_2)}{\sqrt{\mathrm{D}X_1^* \cdot \mathrm{D}X_2^*}} = r(X_1, X_2). \quad \square$$

§ 6. Неравенство Чебышёва. Закон больших чисел

Вероятностный смысл взаимосвязанности двух главных числовых характеристик случайной величины — математического ожидания и дисперсии — определяется при выводе одного важного в теории вероятностей неравенства.

Лемма (неравенство Маркова)**.** *Пусть случайная величина $Y(\omega) \geqslant 0$ имеет математическое ожидание* $\mathrm{E}Y$. *Тогда для любого $\varepsilon > 0$ выполнено неравенство*

$$\mathrm{P}\{Y \geqslant \varepsilon\} \leqslant \frac{\mathrm{E}Y}{\varepsilon}.$$

Доказательство. Имеем

$$\mathrm{E}Y(\omega) = \sum_{\omega} Y(\omega)\mathrm{P}\{\omega\} \geqslant \sum_{\omega:\, Y(\omega)\geqslant\varepsilon} Y(\omega)\mathrm{P}\{\omega\} \geqslant$$

$$\geqslant \varepsilon \sum_{\omega:\, Y(\omega)\geqslant\varepsilon} \mathrm{P}\{\omega\} = \varepsilon\mathrm{P}\{Y(\omega) \geqslant \varepsilon\}.$$

Разделив крайние части на ε, получим искомое неравенство. □

Теорема 1 (неравенство Чебышёва)**.** *Пусть случайная величина* $X(\omega)$ *имеет математическое ожидание* $\mathrm{E}X$ *и дисперсию* $\mathrm{D}X$. *Тогда при любом* $\varepsilon > 0$ *имеет место неравенство*

$$\mathrm{P}\{|X - \mathrm{E}X| \geqslant \varepsilon\} \leqslant \frac{\mathrm{D}X}{\varepsilon^2}.$$

Доказательство. Обозначая $Y = (X - \mathrm{E}X)^2 \geqslant 0$ и применяя к Y лемму, получаем

$$\mathrm{P}\{|X - \mathrm{E}X| \geqslant \varepsilon\} = \mathrm{P}\{Y \geqslant \varepsilon^2\} \leqslant \frac{\mathrm{E}(X - \mathrm{E}X)^2}{\varepsilon^2} = \frac{\mathrm{D}X}{\varepsilon^2}.$$ □

Неравенство Чебышёва имеет в теории вероятностей принципиальное значение, оно выражает роль дисперсии при оценивании отклонений значений случайной величины от ее среднего значения. Неравенство Чебышёва дает во многих случаях достаточно грубую, но гарантированную оценку. При этом сами отклонения от среднего значения могут быть измерены в масштабе стандартных отклонений: обозначим $\mathrm{D}Y = \sigma^2$, $\sigma = +\sqrt{\sigma^2}$ и $t = \dfrac{\varepsilon}{\sigma} > 0$, тогда $\varepsilon = t\sigma$ и неравенство Чебышёва приобретает вид

$$\mathrm{P}\{|X - \mathrm{E}X| \geqslant t\sigma\} \leqslant \frac{1}{t^2}.$$

Пример. Рассмотрим случайную величину S_n — число «орлов» при n бросаниях монеты, и пусть p — вероятность выпадения «орла». Математическое ожидание случайной величины S_n равно np, а дисперсия равна $np(1-p)$. Вычисляя эти характеристики распределения при значениях параметров $n = 100$ и $p = 1/2$ и используя неравенство Чебышёва, получим

$$\mathrm{P}\{|S_n - 50| > 15\} \leqslant \frac{25}{225} = \frac{1}{9}$$

(ср. с вычисленным по таблицам биномиального распределения значением этой вероятности $\mathrm{P}\{|S_n - np| > 15\} = 0{,}00178$).

Неравенство Чебышёва весьма полезно при выводе так называемого закона больших чисел.

Закон больших чисел

Законом больших чисел (ЗБЧ) называется, по предложению А. А. Маркова, группа теорем, характеризующих поведение средних арифметических значений в последовательности случайных величин. В самых общих чертах содержание закона больших чисел — в устойчивости поведения средних арифметических неограниченно большого числа случайных величин, а в наиболее простом случае — в том, что это среднее арифметическое почти постоянно. Различие форм закона больших чисел объяснимо разнообразием необходимых и достаточных условий.

Теорема 2 (Я. Бернулли, исторически первый вариант закона больших чисел, 1713)**.** *Пусть последовательность случайных величин $X_1, \dots, X_n$ соответствует схеме испытаний Бернулли: $X_1, \dots, X_n$ взаимно независимы и $\mathrm{P}\{X_n = 1\} = p$, $\mathrm{P}\{X_n = 0\} = 1 - p$ для любых значений n и p, $0 < p < 1$. Пусть $S_n = X_1 + \dots + X_n$. Тогда для любого $\varepsilon > 0$ при $n \to 0$ выполнено соотношение*[1]

$$\mathrm{P}\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} \to 0.$$

Это есть простое следствие неравенства Чебышёва. Я. Бернулли сравнивал эти вероятности с геометрической прогрессией.

Отношение $\dfrac{S_n}{n}$ есть не что иное, как частота события $\{X_k = 1\}$, происходящего в каждом отдельном испытании с вероятностью p. Данное предельное соотношение разъясняет смысл утверждения о том, что частота события с увеличением числа испытаний стремится к его вероятности.

Закон больших чисел для схемы Бернулли можно записать в форме, лишенной явного вероятностного содержания: при любом $\varepsilon > 0$, $p \in [0, 1]$ и $n \to \infty$ выполнено соотношение

$$\sum_{m:\left|\frac{m}{n} - p\right| > \varepsilon} C_n^m p^m (1-p)^{n-m} \to 0.$$

Докажем теорему, которая является непосредственным следствием теоремы Бернулли, но имеет общематематический смысл.

[1] Это предельное соотношение соответствует типу сходимости случайных величин, которая называется сходимостью по вероятности и в данном случае обозначается $\dfrac{S_n}{n} \xrightarrow{\mathrm{P}} p$ при $n \to \infty$.

Теорема 3. *Пусть функция $f(x)$ непрерывна на отрезке $[0,1]$ и*

$$B_n(x)=\sum_{m=0}^{n} f\Big(\frac{m}{n}\Big)C_n^m x^m(1-x)^{n-m}$$

есть многочлен степени $n\in\mathbb{N}$. Тогда при $n\to\infty$ выполнено соотношение

$$\sup_{x\in[0,1]}|f(x)-B_n(x)|\to 0.$$

В этом утверждении нетрудно узнать теорему Вейерштрасса о приближении непрерывных функций многочленами. Многочлен $B_n(x)=B_n(x;f)$ называют полиномом Бернштейна функции $f(x)$, а следующее ниже «вероятностное» доказательство теоремы Вейерштрасса принадлежит С. Н. Бернштейну (см. [2, т. 1]).

Доказательство. Функция $f(x)$ непрерывна на $[0,1]$, поэтому, во-первых, она ограничена при $x\in[0,1]$, т. е. существует такая постоянная $C>0$, что $|f(x)|\leqslant C$, и, во-вторых, она равномерно непрерывна на $[0,1]$, т. е. для любого $\varepsilon>0$ существует такое значение $\delta>0$, что при любых $x_1,x_2\in[0,1]$ из условия $|x_1-x_2|<\delta$ следует неравенство $|f(x_1)-f(x_2)|<\varepsilon$. Имеем

$$\begin{aligned}|f(x)-B_n(x)|&=\Big|\sum_{m=0}^{n}\Big(f(x)-f\Big(\frac{m}{n}\Big)\Big)C_n^m x^m(1-x)^{n-m}\Big|\leqslant\\&\leqslant\sum_{m:\,\left|\frac{m}{n}-x\right|\leqslant\delta}|f(x)-f\Big(\frac{m}{n}\Big)|C_n^m x^m(1-x)^{n-m}+\\&\quad+\sum_{m:\,\left|\frac{m}{n}-x\right|>\delta}\Big|f(x)-f\Big(\frac{m}{n}\Big)\Big|C_n^m x^m(1-x)^{n-m}.\end{aligned}$$

Первую из этих сумм оцениваем числом $\frac{\varepsilon}{2}$ исходя из того, что

$$\sup_{x\in[0,1]}\Big|f(x)-f\Big(\frac{m}{n}\Big)\Big|<\frac{\varepsilon}{2}$$

и

$$\sum_{m:\,\left|\frac{m}{n}-x\right|\leqslant\delta}C_n^m x^m(1-x)^{n-m}\leqslant\sum_{m=0}^{n}C_n^m x^m(1-x)^{n-m}=1.$$

Вторую сумму оцениваем, используя соображения, положенные в основу теоремы Бернулли: так как

$$\mathrm{P}\{S_n=m\}=C_n^m p^m(1-p)^{n-m}\quad\text{и}\quad B_n\Big(\frac{m}{n}\Big)=\mathrm{E}f\Big(\frac{S_n}{n}\Big),$$

получаем, что

$$\left|f(x)-f\Big(\frac{m}{n}\Big)\right|\leqslant 2C \quad \text{и} \quad \sum_{m:\,\left|\frac{m}{n}-x\right|>\delta} C_n^m x^m(1-x)^{n-m}\leqslant \frac{1}{4n\delta^2}.$$

Затем выбираем подходящие значения n_0 и $\delta>0$ так, чтобы выполнялось неравенство $\dfrac{C}{2n\delta^2}<\dfrac{\varepsilon}{2}$ при $n\geqslant n_0$.

Используя эти две оценки, получаем требуемый результат: при $n\geqslant n_0$ имеем

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

равномерно относительно всех x из $[0,1]$. Таким образом, многочлен $B_n(x;f)$ сходится при $n\to\infty$ к функции $f(x)$ равномерно на $[0,1]$. □

Теорема 4 (Пуассон, 1837). *Пусть последовательность случайных величин $X_1,\ldots,X_n$ соответствует схеме испытаний Пуассона: $X_1,\ldots,X_n$ независимы и имеют распределение*

$$\mathrm{P}\{X_k=1\}=p_k,\quad \mathrm{P}\{X_k=0\}=1-p_k,\quad 0<p_k<1,\ k=1,\ldots,n,\ n\in\mathbb{N}.$$

Пусть $S_n=\sum\limits_{k=1}^{n}X_k$. Тогда для любого $\varepsilon>0$ при $n\to\infty$ выполнено соотношение

$$\mathrm{P}\left\{\left|\frac{S_n}{n}-\frac{p_1+\ldots+p_n}{n}\right|>\varepsilon\right\}\to 0.$$

Именно эту теорему, полученную из варианта центральной предельной теоремы (см. с. 65), Пуассон назвал «законом больших чисел».

Оказалось, что результаты Бернулли и Пуассона являются частными случаями закона, который справедлив в более общей ситуации. Этот факт был обнаружен П. Л. Чебышёвым.

Определение. Говорят, что последовательность случайных величин $X_1,\ldots,X_n$ с математическими ожиданиями $\mathrm{E}X_k$ удовлетворяет закону больших чисел, если для любого $\varepsilon>0$ при $n\to\infty$ выполнено соотношение

$$\mathrm{P}\left\{\left|\frac{X_1+\ldots+X_n}{n}-\frac{\mathrm{E}X_1+\ldots+\mathrm{E}X_n}{n}\right|>\varepsilon\right\}\to 0.$$

Теорема 5 (П. Л. Чебышёв). *Пусть $X_1,\ldots,X_n$ — взаимно независимые случайные величины с математическими ожиданиями $\mathrm{E}X_k$ и дисперсиями $\mathrm{D}X_k$. Для выполнения закона больших чисел достаточно, чтобы существовала такая постоянная $C>0$, что $\mathrm{D}X_k\leqslant C$ для любого $k=1,\ldots$*

Очевидно, что случайные величины, рассматриваемые в теоремах Бернулли и Пуассона, удовлетворяют достаточным условиям теоремы Чебышёва.

Доказательство. Используем неравенство Чебышёва и свойство дисперсий для независимых случайных величин. Выпишем для случайной величины $\frac{X_1+\ldots+X_n}{n}$ неравенство Чебышёва:

$$\mathrm{P}\left\{\left|\frac{X_1+\ldots+X_n}{n}-\frac{\mathrm{E}X_1+\ldots+\mathrm{E}X_n}{n}\right|>\varepsilon\right\}\leqslant\frac{\mathrm{D}X_1+\ldots+\mathrm{D}X_n}{n^2\varepsilon^2}.$$

В силу того что $\mathrm{D}X_n<C$, правая часть неравенства не превосходит величины $\frac{C}{n\varepsilon^2}$, которая стремится к 0 при $n\to\infty$, если C и ε фиксированы. □

В доказательстве теоремы на самом деле используется не взаимная независимость, а некоррелированность случайных величин.

Теорема 6. *Пусть случайные величины $X_1,\ldots,X_n$ попарно некоррелированы, а их дисперсии равномерно ограничены*: $\mathrm{D}X_k\leqslant C$, $C>0$. *Тогда для случайных величин $X_1,\ldots,X_n$ выполняется закон больших чисел.*

В доказательстве используется тот факт, что для некоррелированных случайных величин сохраняется равенство

$$\mathrm{D}(X_1+\ldots+X_n)=\mathrm{D}X_1+\ldots+\mathrm{D}X_n.$$

О применимости закона больших чисел к последовательностям зависимых случайных величин говорится в гл. 4.

§ 7. Вероятностные модели с бесконечным числом элементарных событий

Об общих вероятностных пространствах

Выше рассматривались конечные пространства элементарных событий. В этом случае для всех подмножеств пространства Ω были определены вероятности и все функции от элементарных событий были случайными величинами. При переходе к пространствам Ω с бесконечным числом элементарных событий возникают трудности. Уже при рассмотрении счетных множеств элементарных событий неясно, все ли подмножества пространства Ω имеют вероятность. Так, например, в игре «орел-решка» мы имеем дело с неограниченной последовательностью бросаний монеты. При этом отдельный исход опыта будет изображаться бесконечной последовательностью единиц и нулей (единица обозначает появление «орла»,

а нуль — появление «решки»), например, $\omega = (0100011011\ldots)$. Пространство элементарных событий будет совпадать с пространством всех таких последовательностей. Неясно, как задавать вероятности в таком пространстве. Если, однако, считать, что игра заканчивается при первом появлении «орла», то в качестве элементарных событий естественно рассматривать последовательности

$$\omega_1 = (1),$$
$$\omega_2 = (01),$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\omega_n = (000\ldots01),$$
$$\ldots\ldots\ldots\ldots\ldots$$

По соображениям независимости нулей и единиц в схеме испытаний Бернулли вероятность ω_n следует задать так:

$$\mathrm{P}\{\omega_n\} = \frac{1}{2^n}.$$

Тогда

$$\sum_{n=1}^{\infty} \mathrm{P}\{\omega_n\} = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

Таким образом, в рамках этой модели событие, заключающееся в том, что в бесконечной последовательности испытаний ни разу не выпадет «орел», имеет вероятность, равную нулю. Некоторая искусственность этого связана с невозможностью вместить рассматриваемую ситуацию в рамки конечного пространства элементарных событий.

Таким образом, добавляя модель со счетным числом элементарных исходов, можно расширить наше представление о вероятностных моделях и перейти к так называемому *дискретному* вероятностному пространству.

Введенные ранее конечные распределения благополучно вписываются в эту конструкцию: конечному или счетному множеству точек пространства Ω ставятся в соответствие положительные числа, сумма которых равна 1.

Кроме свойств неотрицательности, нормированности и аддитивности, которыми должна обладать вероятность $\mathrm{P}\{A\}$, мы должны в рассматриваемом случае потребовать выполнения свойства счетной аддитивности: если $A_1, \ldots, A_n, \ldots$ таковы, что $A_i \cap A_j = \varnothing$,

$i \neq j$, то

$$\mathrm{P}\left\{\bigcup_1^\infty A\right\} = \sum_1^\infty \mathrm{P}\{A\},$$

где все вероятности определены. В общей дискретной вероятностной модели класс $\mathscr{A}$ всех событий, т. е. подмножеств пространства Ω, для которых определена вероятность, уже не будет совпадать с классом всех подмножеств пространства Ω, как было в конечном случае. Класс событий должен обладать следующими свойствами: если A_1 и A_2 — элементы класса $\mathscr{A}$, то $\bar{A}_1$, $A_1 \cup A_2$, $A_1 \cap A_2$, $A_1 \backslash A_2$ снова входят в класс $\mathscr{A}$, и, кроме того, класс $\mathscr{A}$ должен быть замкнут относительно операций счетного объединения и пересечения. Такие классы событий называются *алгеброй* и *сигма-алгеброй* множеств соответственно.

Точками пространства Ω могут быть сколь угодно сложные объекты. Укажем два примера.

1. Пусть частица совершает случайные перемещения на плоскости. Введем на плоскости декартову систему координат (t, x) и рассмотрим изменение координаты x на отрезке времени $[0, T]$. Элементарным событием здесь будет непрерывная функция $x(t)$ — координата частицы, определенная на отрезке $[0, T]$, а пространство Ω будет совпадать с пространством $C[0, T]$ непрерывных функций на $[0, T]$.

2. В интервале $[0, 1]$ на прямой выбирается «наудачу» точка. Элементарные события — все точки интервала. Точный смысл термина «наудачу» состоит в том, что мы приписываем каждому подынтервалу интервала $[0, 1]$ с концами a и b $(a < b)$ вероятность, равную его длине $b - a$. Однако при этом неясно, как определить вероятность того, что точка, выбранная наудачу на отрезке, имеет рациональную координату. Ясно, что прежний способ задания вероятностей не годится.

При ответе на вопрос, для каких подмножеств элементарных событий можно определить вероятность, возникают те же трудности, что и при определении площади и объема. Дело в том, что существуют множества, для которых определить вероятность нельзя. Но существование неизмеримых множеств — это проблема общей теории меры. Поэтому строгое математическое изучение общих пространств элементарных событий невозможно без обращения к теории меры и интеграла Лебега. В этом курсе мы по возможности

будем стараться не затрагивать проблему теоретического обоснования и будем заниматься частными, но имеющими практическую ценность случаями.

Перейдем к общему определению вероятностного пространства $(\Omega, \mathscr{A}, \mathrm{P})$. Здесь пространство элементарных событий Ω — множество любой природы и произвольной мощности, а класс $\mathscr{A}$ подмножеств Ω является алгеброй, т. е. обладает следующими свойствами:

1) $A \in \mathscr{A} \Rightarrow \overline{A} = \Omega \setminus A \in \mathscr{A}$;

2) $A_1, A_2 \in \mathscr{A} \Rightarrow A_1 \cap A_2 \in \mathscr{A}$, $A_1 \cup A_2 \in \mathscr{A}$,

3) $\Omega \in \mathscr{A}$,

или, более общим образом, сигма-алгеброй, т. е. алгеброй, которая замкнута относительного счетных пересечений и объединений своих элементов:

$$A_1, \ldots, A_n, \ldots \in \mathscr{A} \Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathscr{A}, \quad \bigcup_{k=1}^{\infty} A_k \in \mathscr{A}.$$

Наконец, функция $\mathrm{P}\{A\}$, определенная на сигма-алгебре $\mathscr{A}$, называется *распределением вероятностей* или *вероятностной мерой*, если она удовлетворяет трем аксиомам Колмогорова:

1) $0 \leqslant \mathrm{P}\{A\} \leqslant 1$,

2) $\mathrm{P}\{\Omega\} = 1$,

3) $A_1, \ldots, A_n, \ldots \in \mathscr{A}$, $A_i \cap A_j = \varnothing$, $i \neq j \Rightarrow \mathrm{P}\left\{\bigcup_{k=1}^{\infty} A_k\right\} = \sum_{k=1}^{\infty} \mathrm{P}\{A_k\}$.

Последняя аксиома есть условие счетной аддитивности. Для конечных моделей свойства 1—3 выполняются автоматически, но в общем случае благодаря аксиомам, предложенным А. Н. Колмогоровым в 1933 г., теория вероятностей получила свое математическое обоснование.

Пусть пространство элементарных событий Ω совпадает с n-мерным евклидовым пространством R^n: $\omega = (x_1, \ldots, x_n)$, $x_k \in \mathrm{R}^1$. В этом случае сигма-алгебра событий, называемая борелевской алгеброй, содержит все «типичные» подмножества пространства R^n. Борелевская алгебра в евклидовых пространствах является естественной областью определения вероятностного распределения. Рассмотрим, например, евклидову прямую $\mathrm{R} = \mathrm{R}^1$. Соответствующая борелевская алгебра включает произвольные интервалы и полуинтервалы, отрезки, полупрямые, все открытые и замкнутые множества. С точки зрения обоснования теории вероятностей существует конструктивная схема построения алгебры событий на основе некоторого подходящего класса подмножеств, например, на основе конечно-

го разбиения пространства элементарных событий. На элементах выбранного класса подмножеств легко задать вероятность, затем найти вероятности для элементов алгебры, и наконец доопределить («продолжить») вероятность на элементы сигма-алгебры на основании общей теоремы Каратеодори о продолжении меры в измеримых пространствах.

Рассмотрим вероятностное пространство $(\mathrm{R}, \mathscr{B}, \mathrm{P})$ и элементы борелевской алгебры $\mathscr{B}$ вида $(-\infty, x]$. Назовем функцией распределения вероятностного распределения следующую функцию действительного аргумента $x \in \mathrm{R}$:

$$\mathscr{F}(x) = \mathrm{P}\{(-\infty, x]\}.$$

Эта функция обладает следующими свойствами:

1) $\mathscr{F}(x)$ — неубывающая функция переменной x;

2) $\mathscr{F}(x)$ — непрерывная справа в любой точке функция переменной x: $\mathscr{F}_n(x_n) \to \mathscr{F}(x)$ при $x_n \downarrow x$;

3) $\lim\limits_{x \to -\infty} \mathscr{F}(x) = 0$, $\lim\limits_{x \to \infty} \mathscr{F}(x) = 1$.

Важная роль этой функции в теории вероятностей определяется тем, что каждой такой функции, заданной этими тремя свойствами, соответствует ровно одно распределение вероятностей. При этом для любых действительных a и b, $a < b$, выполнено равенство

$$\mathrm{P}\{a, b\} = \mathscr{F}(b) - \mathscr{F}(a).$$

Функция распределения может быть найдена для любого распределения, в том числе и для дискретного, но исключительную роль она играет для распределений, которые мы называем непрерывными.

Непрерывные распределения

В качестве Ω рассмотрим n-мерное евклидово пространство R^n, т. е. элементарными событиями будут точки $\omega = (x_1, \ldots, x_n)$, где $x_k \in \mathrm{R}$, а в качестве $\mathscr{A}$ — соответствующую борелевскую сигма-алгебру. Предположим, что в R^n задана функция $p(\omega)$, обладающая двумя свойствами:

1) $p(\omega) \geqslant 0$,

2) $\int\limits_{\mathrm{R}^n} p(\omega)\, d(\omega) = 1$.

Положим тогда $\mathrm{P}\{A\} = \int\limits_A p(\omega)\, d(\omega)$ для таких множеств $A \in \mathscr{A}$, для которых интеграл существует. Так определенная функция множеств называется *непрерывным распределением вероятностей* на $(\Omega, \mathscr{A})$, а функция $p(\omega)$ называется *плотностью* этого распределения.

Действительная функция $X(\omega)$, $\omega \in \Omega$, называется случайной величиной, если для любых действительных a и b $(-\infty < a < b < \infty)$ определена вероятность $\mathrm{P}\{\omega\colon a \leqslant X(\omega) \leqslant b\}$.

Пусть $X(\omega)$ — действительная случайная величина, заданная на вероятностном пространстве $(\Omega, \mathscr{A}, \mathrm{P})$. Рассмотрим все множества B на прямой, для которых $\{\omega\colon X(\omega) \in B\} \in \mathscr{A}$, где $\mathscr{A}$ есть сигма-алгебра в пространстве $(\Omega, \mathscr{A}, \mathrm{P})$. Тогда для любого такого множества B определена вероятность $\mathrm{P}\{\omega\colon X(\omega) \in B\}$ и можно положить $\mathrm{P}_X\{B\} = \mathrm{P}\{\omega\colon X(\omega) \in B\}$. Объединим все такие множества B в класс $\mathscr{B}$. Этот класс будет борелевской алгеброй подмножеств прямой R. Функция $\mathrm{P}_X\{B\}$, $B \in \mathscr{B}$, называется *распределением вероятностей случайной величины* X. Если для любого множества $B \in \mathscr{B}$ допустимо представление

$$\mathrm{P}_X\{B\} = \int\limits_B p_X(x)\,dx, \quad \text{где} \quad p_X(x) \geqslant 0 \quad \text{и} \quad \int\limits_{\mathrm{R}^1} p(x)\,dx = 1,$$

то $p_X(x)$ называется *плотностью распределения вероятностей случайной величины* X, и в этом случае говорят, что $X(\omega)$ имеет непрерывное распределение.

Пример 1 (равномерное распределение на отрезке)**.** Пусть на прямой $\Omega = \mathrm{R}^1$ задано распределение вероятностей с плотностью

$$p(x) = \begin{cases} C > 0, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases}$$

где $a < b$ — фиксированные действительные числа. Постоянная C находится из условия $\int\limits_{-\infty}^{\infty} p(x)\,dx = 1$:

$$c = \frac{1}{b-a}.$$

Распределение с такой плотностью называется *равномерным распределением* на отрезке $[a, b]$; при этом вероятность попасть в любой подынтервал $[\alpha, \beta]$, $a \leqslant \alpha < \beta \leqslant b$, пропорциональна его длине:

$$\mathrm{P}\{X \in [\alpha, \beta]\} = \frac{\beta - \alpha}{b-a}.$$

Функция распределения случайной величины X с равномерным распределением на отрезке $[a, b]$ вычисляется по формулам

$$\mathscr{F}(x) = 0,\ x \leqslant a; \quad \mathscr{F}(x) = \frac{x-a}{b-a},\ a < x \leqslant b; \quad \mathscr{F}(x) = 1,\ x > b.$$

Непрерывные распределения в R^n

В евклидовом пространстве R^n особую роль играют координатные случайные величины $X_1(\omega)=x_1, \ldots, X_n(\omega)=x_n$, принимающие значения на координатных осях. Пусть $n=2$. Рассмотрим распределение на плоскости с плотностью $p(\omega)=p(x_1, x_2)$:

$$\int\limits_{\mathrm{R}^2} p(x_1, x_2)\, dx_1\, dx_2 = 1$$

и $p(x_1, x_2) \geqslant 0$ при $(x_1, x_2) \in \mathrm{R}^2$, и пусть $X_1(\omega)=x_1$, $X_2(\omega)=x_2$ — координатные случайные величины. Покажем, что по распределению в R^2 можно найти распределения различных функций от координатных случайных величин.

1. Найдем вероятность события $\{\omega\colon a \leqslant X_1(\omega) \leqslant b\}$, где a и b — произвольные действительные числа. Имеем

$$\mathrm{P}\{\omega\colon a \leqslant X_1(\omega) \leqslant b\} = \iint\limits_{\substack{a\leqslant x_1\leqslant b,\\ -\infty<x_2<\infty}} p(x_1, x_2)\, dx_1\, dx_2 =$$

$$= \int\limits_a^b \left[\int\limits_{-\infty}^{\infty} p(x_1, x_2)\, dx_2\right] dx_1 = \int\limits_a^b p(x_1)\, dx_1.$$

Согласно общему определению $p(x_1)$ — плотность распределения случайной величины $X_1(\omega)$. Аналогично отыскивается плотность $X_2(\omega)$. Таким образом, если распределение на плоскости имеет плотность, то каждая из координатных случайных величин имеет плотность.

2. Найдем плотность распределения случайной величины $X = X_1 + X_2$. Для этого вычислим вероятность события $\{a \leqslant X_1 + X_2 \leqslant b\}$ для любых действительных a и b. Обозначим

$$\mathscr{D} = \{\omega\colon a \leqslant X_1 + X_2 \leqslant b\}.$$

Имеем

$$\mathrm{P}\{\mathscr{D}\} = \iint\limits_{\mathscr{D}} p(x_1, x_2)\, dx_1\, dx_2 = \int\limits_{-\infty}^{\infty} dx_1 \left[\int\limits_{a-x_1}^{b-x_1} p(x_1, x_2)\, dx_2\right] =$$

$$= \int\limits_{-\infty}^{\infty} dx_1 \left(\int\limits_a^b p(x_1, y-x_1)\, dy\right) = \int\limits_a^b dy \left[\int\limits_{-\infty}^{\infty} p(x_1, y-x_1)\, dx_1\right].$$

Итак, $\mathrm{P}\{a \leqslant X_1 + X_2 \leqslant b\} = \int\limits_a^b q(y)\, dy$, где $q(y) = \int\limits_{-\infty}^{\infty} p(x, y-x)\, dx$ есть по определению плотность случайной величины $X_1 + X_2$.

Таким образом, по плотности двумерного распределения можно найти плотность функции от координатных случайных величин. Однако часто интегралы, через которые эти плотности выражаются, не вычисляются в элементарных функциях и их приходится табулировать в целях практического использования.

Из предыдущего ясно, что распределения на плоскости можно рассматривать как совместные распределения координатных случайных величин. Перейдем теперь к совместным распределениям произвольных случайных величин. Рассмотрим две произвольные случайные величины $X_1(\omega)$ и $X_2(\omega)$ и вектор (X_1, X_2). Пусть $\mathscr{D}$ — любое такое множество на плоскости, что для события $\{\omega\colon (X_1, X_2)\in\mathscr{D}\}$ определена вероятность. Тогда

$$\mathrm{P}\{\omega\colon (X_1(\omega), X_2(\omega))\in\mathscr{D}\} = \mathrm{P}\{\mathscr{D}\}$$

определяет *совместное распределение* вероятностей случайных величин X_1 и X_2. Очевидно, что класс множеств $\mathscr{D}$ «порождает» борелевскую алгебру на плоскости R^2. Если $\mathscr{D}$ имеет вид $\{X_1\leqslant x_1, X_2\leqslant x_2\}$, то $\mathscr{F}_{X_1,X_2}(x_1, x_2) = \mathrm{P}\{X_1\leqslant x_1, X_2\leqslant x_2\}$ называется функцией распределения соответствующего совместного распределения величин X_1 и X_2. Если для любого борелевского множества $\mathscr{D}$ эту вероятность можно представить в виде

$$\mathrm{P}\{\omega\colon (X_1, X_2)\in\mathscr{D}\} = \int\limits_{\mathscr{D}} p(x_1, x_2)\,dx_1\,dx_2,$$

то такая функция $p(x_1, x_2)$, что $p(x_1, x_2)\geqslant 0$ и $\iint\limits_{\mathrm{R}^2} p(x_1, x_2)\,dx_1\,dx_2 = 1$, называется *плотностью совместного распределения* величин X_1 и X_2. Плотность $p_{X_1}(x_1)$ случайной величины X_1 и плотность $p_{X_2}(x_2)$ случайной величины X_2 называются *частными* или *маргинальными* плотностями по отношению к совместному распределению величин X_1 и X_2.

Пример 2. Рассмотрим сферу единичного радиуса в R^3 и равномерное распределение на сфере. Введем в R^3 географические координаты точки — расстояние от центра ρ, долготу θ и широту φ ($\rho\geqslant 0$, $0\leqslant\theta\leqslant 2\pi$, $-\frac{\pi}{2}\leqslant\varphi\leqslant\frac{\pi}{2}$), т. е. перейдем от координатных случайных величин X_1, X_2, X_3 к случайным величинам R, Θ, Φ:

$$X_1 = R\cos\Theta\cos\Phi,\quad X_2 = R\sin\Theta\cos\Phi,\quad X_3 = R\sin\Phi.$$

Тогда совместная плотность величин R, Θ, Φ равна

$$g(\rho, \theta, \varphi) = p(\rho\cos\theta\cos\varphi, \rho\sin\theta\cos\varphi, \rho\sin\varphi)\rho^2\cos\varphi.$$

Сфера единичного радиуса в R^3 описывается с помощью θ и φ следующим образом:

$$x_1 = \cos\theta\cos\varphi;$$
$$x_2 = \sin\theta\cos\varphi, \quad 0 \leqslant \theta \leqslant 2\pi;$$
$$x_3 = \sin\varphi, \qquad -\frac{\pi}{2} \leqslant \varphi \leqslant \frac{\pi}{2}.$$

Точке ω на сфере соответствует единственная точка (θ, φ) на плоскости (θ, φ), области $\mathscr{D}$ на сфере соответствует область $\mathscr{D}^*$ на (θ, φ)-плоскости ($\mathscr{D}^*$ — образ области $\mathscr{D}$). При этом

$$\text{площадь } \mathscr{D} = \iint\limits_{\mathscr{D}} \cos\varphi \, d\theta \, d\varphi.$$

Так как распределение на сфере равномерное:

$$\mathrm{P}\{\mathscr{D}\} = \frac{\text{площадь } \mathscr{D}}{4\pi},$$

тем самым на (θ, φ)-плоскости задается плотность совместного распределения долготы Θ и широты Φ:

$$p(\theta, \varphi) = \frac{1}{4\pi}\cos\varphi, \quad 0 \leqslant \theta \leqslant 2\pi; -\frac{\pi}{2} \leqslant \varphi \leqslant \frac{\pi}{2}.$$

При любом выборе $\varphi_1 < \varphi_2$ и $\theta_1 < \theta_2$ имеем

$$\mathrm{P}\{\mathscr{D}\} = \frac{(\sin\varphi_2 - \sin\varphi_1)(\theta_2 - \theta_1)}{4\pi}.$$

Легко показать, что

$$\mathrm{P}\{\theta_1 \leqslant \Theta \leqslant \theta_2\} = \frac{\theta_2 - \theta_1}{2\pi},$$

т. е. распределение Θ является равномерным на $[0, 2\pi]$ с плотностью $p_\Theta(\theta) = \frac{1}{2\pi}$. Далее, имеем

$$\mathrm{P}\{\varphi_1 \leqslant \Phi \leqslant \varphi_2\} = \int\limits_{\varphi_1}^{\varphi_2} \frac{1}{2}\cos\varphi \, d\varphi = \frac{\sin\varphi_2 - \sin\varphi_1}{2}.$$

Поэтому $p_\Phi(\varphi) = \frac{1}{2}\cos\varphi$, $|\varphi| \leqslant \frac{\pi}{2}$, и, таким образом, плотность Φ не является равномерной. Итак, получено соотношение

$$\mathrm{P}\{\theta_1 \leqslant \Theta \leqslant \theta_2, \varphi_1 \leqslant \Phi \leqslant \varphi_2\} = \mathrm{P}\{\theta_1 \leqslant \Theta \leqslant \theta_2\}\mathrm{P}\{\varphi_1 \leqslant \Phi \leqslant \varphi_2\}.$$

Случайные величины, плотности которых удовлетворяют выписанному соотношению, называются независимыми. Таким образом, показано, что географические долгота и широта на сфере независимы:

$$p_{\Theta,\Phi}(\theta, \varphi) = p_\Theta(\theta)p_\Phi(\varphi) = \frac{\cos\varphi}{4\pi}.$$

Независимые случайные величины

Определение. Две действительные случайные величины X_1 и X_2 называются *независимыми*, если при любом наборе действительных чисел $a_1 < b_1$, $a_2 < b_2$ выполняются соотношения

$$\mathrm{P}\{\omega\colon a_1 \leqslant X_1 \leqslant b_1, a_2 \leqslant X_2 \leqslant b_2\} = \\ = \mathrm{P}\{\omega\colon a_1 \leqslant X_1 \leqslant b_1\} \cdot \mathrm{P}\{\omega\colon a_2 \leqslant X_2 \leqslant b_2\}.$$

Если $\mathscr{F}_{X_1,X_2}(x_1, x_2)$ — функция распределения совместного распределения величин X_1 и X_2, а $\mathscr{F}_{X_1}(x_1)$ и $\mathscr{F}_{X_2}(x_2)$ — функции распределения величин X_1 и X_2 соответственно, то случайные величины X_1 и X_2 независимы тогда и только тогда, когда при любых действительных x_1 и x_2 выполняется равенство

$$\mathscr{F}_{X_1,X_2}(x_1, x_2) = \mathscr{F}_{X_1}(x_1) \cdot \mathscr{F}_{X_2}(x_2).$$

Можно показать, что если совместное распределение (X_1, X_2) имеет плотность $p_{X_1,X_2}(x_1, x_2)$, то необходимым и достаточным условием независимости является условие

$$p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2),$$

т. е. плотность совместного распределения равна произведению частных плотностей.

Если случайные величины X_1 и X_2 независимы, то плотность распределения суммы $X_1 + X_2$ определяется по формуле «свертки»:

$$p_{X_1+X_2}(x) = \int\limits_{-\infty}^{\infty} p_{X_1}(y) p_{X_2}(x - y)\, dy.$$

В общем случае плотность распределения суммы двух случайных величин вычисляется на с. 50.

Пример 3. Если случайная величина (X_1, X_2) распределена равномерно в квадрате со стороной 1, то координатные случайные величины X_1 и X_2 независимы.

Пример 4. Если случайная величина (X_1, X_2) распределена равномерно в круге единичного радиуса с центром в начале координат, то координатные случайные величины X_1 и X_2 не являются независимыми.

Условие независимости является весьма сильным в том смысле, что в предположении независимости случайных величин можно получить значительную информацию о распределении этих случай-

ных величин. Приведем пример, когда распределение определяется предположением о независимости.

Теорема. *Пусть X_1 и X_2 — координатные случайные величины на декартовой плоскости. Предположим, что*

1) *существует плотность $p(x_1, x_2)$ совместного распределения величин X_1 и X_2;*

2) $p(x_1, x_2) = f\left(\sqrt{x_1^2 + x_2^2}\right)$, *т. е. плотность распределения (X_1, X_2) зависит лишь от $r = \sqrt{x_1^2 + x_2^2}$;*

3) *случайные величины X_1 и X_2 независимы.*

Тогда при дополнительном предположении о дифференцируемости некоторых функций от плотностей случайные величины X_1 и X_2 распределены с плотностями $p(x_i) = \dfrac{1}{\sqrt{2\pi\sigma_i^2}} e^{-x_i^2/2\sigma_i^2}$, $i = 1, 2$, где постоянные σ_1 и σ_2 являются параметрами распределения.

Доказательство. Обозначим через $p(x_1)$ и $p(x_2)$ частные плотности распределений X_1 и X_2 соответственно. В силу независимости X_1 и X_2, а также условия 2 теоремы имеем $p(x_1, x_2) = p(x_1)p(x_2) = f(r)$. Легко показать, что ни в одной из точек ни $p(x_1)$, ни $p(x_2)$ не обращаются в нуль, т. е. $p(x_1) > 0$, $p(x_2) > 0$, $f(r) > 0$ (в ином случае $p(x_1)$ и $p(x_2)$ тождественно равны нулю). Перейдем в последнем равенстве к логарифмам: $\ln p(x_1) + \ln p(x_2) = \ln f(r)$. Предполагая дифференцируемость функций $p(x_1)$, $p(x_1)$ и $f(r)$ по соответствующим переменным, продифференцируем отдельно по x_1 и по x_2:

$$\frac{1}{p(x_1)} \cdot \frac{dp(x_1)}{dx_1} = \frac{1}{f(r)} \cdot \frac{df(r)}{dr} \cdot \frac{x_1}{r},$$
$$\frac{1}{p(x_2)} \cdot \frac{dp(x_2)}{dx_2} = \frac{1}{f(r)} \cdot \frac{df(r)}{dr} \cdot \frac{x_2}{r}.$$

Отсюда получаем

$$\frac{1}{x_1 p(x_1)} \cdot \frac{dp(x_1)}{dx_1} = \frac{1}{x_2 p(x_2)} \cdot \frac{dp(x_2)}{dx_2} = \frac{1}{rf(r)} \cdot \frac{df(r)}{\partial r}.$$

Последнее равенство возможно в том и только в том случае, когда

$$\frac{1}{x_1 p(x_1)} \cdot \frac{dp(x_1)}{dx_1} = \frac{1}{x_2 p(x_2)} \cdot \frac{dp(x_2)}{dx_2} = 0.$$

Тогда, в частности, $\dfrac{d \ln p(x_1)}{dx_1} = cx_1$. Решая дифференциальное уравнение, получим $p(x_1) = e^{cx_1^2/2 + c'}$, где c и c' — постоянные. Аналогичное равенство верно для $p(x_2)$. □

Полученные плотности определяют нормальное распределение.

Нормальное распределение

Определение. О случайной величине X говорят, что она имеет *нормальное распределение* с параметрами a и σ^2 ($a \in \mathrm{R}$, $\sigma^2 > 0$), если ее плотность распределения имеет вид

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/2\sigma^2}.$$

Функция

$$F_{a,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-(u-a)^2/2\sigma^2}\, du$$

действительного аргумента x является функцией распределения случайной величины X, или *функцией нормального распределения*.

Особо выделяется нормальное распределение с параметрами $a = 0$ и $\sigma^2 = 1$, называемое *стандартным* нормальным распределением. Плотность и функция распределения стандартного нормального распределения, обычно обозначаемые $\varphi(x)$ и $\Phi(x)$, табулированы (см., например, [3]). Легко показать, что при любых a и $\sigma^2 > 0$ плотность и функция нормального распределения с параметрами a и σ^2 выражаются через плотность и функцию стандартного нормального распределения следующим образом:

$$F_{(a,\sigma)}(x) = \Phi\Big(\frac{x-a}{\sigma}\Big), \quad p_{(a,\sigma)}(x) = \frac{1}{\sigma}\varphi\Big(\frac{x-a}{\sigma}\Big).$$

Плотность нормального распределения с параметрами (a, σ^2) симметрична относительно математического ожидания a и имеет единственный максимум при $x = a$. Плотность $\varphi(x)$ стандартного нормального распределения — четная функция, а функция распределения $\Phi(x)$ удовлетворяет равенству $\Phi(x) + \Phi(-x) = 1$. Для нормального распределения с параметрами a и σ^2 вероятность неравенства $|x - a| > t\sigma$ ($t > 0$) равна $1 - \Phi(t) + \Phi(-t)$ и убывает с ростом t (здесь $\sigma = +\sqrt{\sigma^2}$). Поскольку $\mathrm{P}\{|X - a| > 3\sigma\} < 0{,}03$, на практике обычно пренебрегают возможными отклонениями от среднего значения, превышающими 3σ («правило трех сигм»).

Теорема, доказанная в предыдущей части, указывает одну из возможностей появления нормального распределения как точного распределения некоторых случайных величин. Эта теорема относится к ряду теорем, называемых теоремами характеризации нормального распределения. Еще один пример такого рода теорем связан с именем Максвелла: если компоненты трехмерного случайного вектора $X = (X_1, X_2, X_3)$ в любой декартовой системе координат независимы и имеют нулевые математические ожидания, то слу-

чайные величины X_1, X_2, X_3 одинаково нормально распределены (см. [11, т. 2, гл. 3, § 4]).

Нормальное распределение играет важнейшую роль в теории вероятностей как приближенное распределение для сумм большого числа независимых случайных величин (см. гл. 3, § 3).

Моменты непрерывно распределенных случайных величин

Пусть R^n — пространство элементарных событий, $\omega = (x_1, \ldots, x_n)$, $p(\omega) = p(x_1, \ldots, x_n)$ — плотность распределения, $X(\omega)$ — случайная величина, определенная на R^n. Тогда по определению

$$\mathrm{E}X(\omega) = \int\limits_{\mathrm{R}^n} X(\omega)p(\omega)\,d\omega = \int\limits_{-\infty}^{\infty}\ldots\int\limits_{-\infty}^{\infty} X(x_1, \ldots, x_n)p(x_1, \ldots, x_n)\,dx_1\ldots dx_n,$$

если интеграл сходится абсолютно, есть математическое ожидание случайной величины $X(\omega)$. Если $p_X(x)$, $x \in \mathrm{R}$, — плотность распределения случайной величины $X(\omega)$, то можно получить выражения для математического ожидания случайной величины и любой функции от случайной величины через распределение вероятностей $X(\omega)$:

$$\mathrm{E}X = \int\limits_{-\infty}^{\infty} xp_X(x)\,dx, \quad \mathrm{E}f(X) = \int\limits_{-\infty}^{\infty} f(x)p_X(x)\,dx,$$

если интервалы сходятся абсолютно.

Аналогично определяются моменты высших порядков:

$$\mathrm{E}X^m = \int\limits_{-\infty}^{+\infty} x^m p_X(x)\,dx, \quad m \geqslant 1,\ m \text{ целое}.$$

Пример 5. Легко вычислить математическое ожидание $\mathrm{E}X$ и дисперсию $\mathrm{D}X$ нормально распределенной случайной величины X, с плотностью $p(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/2\sigma^2}$:

$$\mathrm{E}X = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} xe^{-(x-a)^2/2\sigma^2}\,dx = a,$$

$$\mathrm{D}X = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/2\sigma^2}\,dx = \sigma^2.$$

Тем самым выявлен смысл параметров нормального распределения: a — среднее значение, или математическое ожидание, σ^2 — дисперсия, σ — стандартное отклонение.

Глава 3

Суммирование независимых случайных величин. Предельные теоремы теории вероятностей

Мы приведем примеры отыскания точных распределений сумм независимых случайных величин и убедимся, что это не всегда просто. Кроме того, мы введем некоторые характеристики распределений случайных величин, которые заключают в себе достаточно полную информацию о распределении и при этом подчиняются простым закономерностям, например, при сложении независимых случайных величин перемножаются.

Напомним, что если независимые случайные величины X_1 и X_2, имеют плотности распределения $p_{X_1}(x_1)$ и $p_{X_2}(x_2)$, то плотность их совместного распределения равна $p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2)$. В свое время была получена формула для плотности суммы $X_1 + X_2$ случайных величин X_1 и X_2:

$$q(y) = \int_{-\infty}^{\infty} p(x, y-x)\,dx.$$

Поэтому если величины независимы, то плотность суммы является *сверткой* плотностей $p_{X_1}(x_1)$ и $p_{X_2}(x_2)$:

$$q(y) = \int_{-\infty}^{\infty} p_{X_1}(x)p_{X_2}(y-x)\,dx.$$

Пример 1. Пусть X_1 и X_2 — независимые случайные величины, распределенные нормально с параметрами a_1, σ_1^2 и a_2, σ_2^2 соответственно. Тогда сумма $X_1 + X_2$ имеет нормальное распределение с параметрами $a_1 + a_2$, $\sigma_1^2 + \sigma_2^2$ (см., например, [7, § 24]).

Пример 2. Пусть X_1 и X_2 — независимые одинаково нормально распределенные величины с параметрами (0, 1). Тогда случайная величина $Z_1 = X_1^2 + X_2^2$ (Z_1 — квадрат длины вектора, имеющего независимые нормально распределенные компоненты) имеет плотность

$$p_{Z_1}(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$$

Получена так называемая *показательная* плотность с параметром $\frac{1}{2}$. Соответствующая функция распределения равна

$$\mathrm{P}\{Z_1 < r^2\} = 1 - e^{-r^2/2}.$$

Пример 3. В предположениях примера 2 рассмотрим случайную величину $Z_2 = \frac{X_1^2}{a_1^2} + \frac{X_2^2}{a_2^2}$. Плотность $p_{Z_2}(x)$ можно вычислить по формуле свертки:

$$p_{Z_2}(x) = \frac{1}{2} a_1 a_2 e^{-x(a_1^2+a_2^2)/4} I_0\Big(\frac{x}{4}(a_1^2 - a_2^2)\Big),$$

где $I_0(z) = \sum\limits_{r=0}^{\infty} \frac{z^{ir}}{(r!)^2}$ — бесселева функция чисто мнимого аргумента и нулевого индекса. Тогда вычисление вероятности $\mathrm{P}\{Z_2 < r^2\} = \int\limits_0^{r^2} p_{Z_2}(x)\,dx$ может быть осуществлено сведением к некоторой табулированной функции.

Последний пример показывает, что с помощью формулы свертки можно определить явный вид распределения суммы независимых случайных величин, но эти формулы могут оказаться бесполезными, так как интеграл, входящий в формулу свертки, часто не выражается через элементарные функции, что затрудняет использование этой формулы. Следует также отметить, что в случае, когда складывается большее число независимых величин, многократное использование формулы свертки приводит к более сложным выражениям.

§1. Производящие и характеристические функции

Обратимся теперь к новым характеристикам распределений, таким как производящие и характеристические функции.

Определение 1. *Производящей функцией* целочисленной неотрицательной случайной величины X, принимающей значения $0, 1, 2, 3, \ldots$ с вероятностями $p_0, p_1, p_2, \ldots$ (соответственно $p_m = \mathrm{P}\{X = m\}$), называется функция

$$\varphi(z) = \sum_{k=0}^{\infty} p_k z^k = \mathrm{E} z^X$$

действительного аргумента z, при этом входящий в определение производящей функции степенной ряд абсолютно сходится при $|z| \leqslant 1$ в силу того, что сумма коэффициентов ряда равна 1.

Отметим важнейшие свойства производящих функций.

1. Производящая функция суммы независимых целочисленных случайных величин равна произведению производящих функций слагаемых.

Пусть X_1 и X_2 — независимые случайные величины с производящими функциями $\varphi_1(z)$ и $\varphi_2(z)$. Тогда производящая функция суммы X_1+X_2 равна

$$\varphi(z)=\mathrm{E}z^{X_1+X_2}=\mathrm{E}z^{X_1}\cdot\mathrm{E}z^{X_2}=\varphi_1(z)\varphi_2(z).$$

2. Целочисленная случайная величина полностью определяется производящей функцией, так как коэффициенты степенного ряда полностью определяются суммой ряда.

Таким образом, по производящей функции можно находить распределение суммы случайных величин.

Пример 1. Пусть случайные величины $X_1, X_2, \ldots, X_n$ взаимно независимы и распределены по закону Пуассона с параметрами $\lambda_1, \ldots, \lambda_n$. Производящая функция распределения Пуассона с параметром λ равна $e^{\lambda(z-1)}$. Производящая функция суммы $X_1+\ldots+X_n$ равна

$$\varphi(z)=\prod_{k=1}^{n}\varphi_k(z)=\prod_{k=1}^{n}e^{\lambda_k(z-1)}=\exp\left\{\sum_{k=1}^{n}\lambda_k(z-1)\right\}$$

и представляет собой не что иное, как производящую функцию распределения Пуассона с параметром $\sum\limits_{k=1}^{n}\lambda_k$.

Пример 2. Пусть $X_1, \ldots, X_n$ — взаимно независимые случайные величины с одним и тем же дискретным равномерным распределением

$$\mathrm{P}\{X_n=k\}=\frac{1}{6},\quad k=1,2,3,4,5,6$$

(это соответствует, например, n независимым бросаниям правильной игральной кости, причем X_n — число очков при n-м бросании). Найдем распределение суммы $S_n=X_1+\ldots+X_n$ (ясно, что $n\leqslant S_n\leqslant 6n$). Сделать это непосредственными вычислениями нелегко даже при малых значениях n, но мы воспользуемся производящими функциями.

Очевидно, что производящая функция каждого слагаемого равна

$$\mathrm{E}z^{X_n}=\frac{1}{6}z+\frac{1}{6}z^2+\ldots+\frac{1}{6}z^6=\frac{1}{6}\frac{z(1-z^6)}{1-z}.$$

Тогда

$$\mathrm{E}z^{S_n} = \mathrm{E}z^{X_1}\cdot\ldots\cdot\mathrm{E}z^{X_n} = (\mathrm{E}z^{X_k})^n = \frac{1}{6^n}\frac{z^n(1-z^6)^n}{(1-z)^n} =$$

$$= \frac{1}{6^n}z^n\left[\sum_{k=0}^{n}(-1)^k C_n^k z^{6k}\right]\left[\sum_{j=0}^{\infty} C_{n+j-1}^{n-1} z^j\right].$$

Теперь перемножим в правой части последнего равенства ряды и найдем коэффициент при z^m, который равен вероятности $\mathrm{P}\{S_n = m\}$:

$$\mathrm{P}\{S_n = m\} = \frac{1}{6^n}\sum_{k=0}^{k'} C_n^k (-1)^k C_{m-6k-1}^{n-1},$$

где k' — наибольшее целое число, меньшее $\frac{m-n}{6}$, $m \geqslant n$. Рассмотрим численный пример. Пусть $n=5$ и $m=15$. Тогда

$$\frac{m-n}{6} = \frac{10}{6}, \quad k' = 1, \quad k = 0 \text{ или } 1,$$

$$\mathrm{P}\{S_n = 15\} = \frac{1}{6^5}[C_5^0(-1)^0 C_{15-0-1}^4 - C_5^1 C_{15-6-1}^4] = \frac{651}{7776}.$$

Очевидно, однако, что при больших значениях n и m выведенные формулы для $\mathrm{P}\{s_n = m\}$ мало пригодны для численных расчетов.

Метод производящих функций при всех своих достоинствах имеет имеет ограниченное поле приложений, так как предназначен лишь для целочисленных неотрицательных случайных величин.

Можно, конечно, определить $a(z) = \sum_{-\infty}^{\infty} a_k z^k$, но гарантировать сходимость ряда можно лишь для комплекснозначных z на окружности радиуса $|z| = 1$, т. е. для значений $z = e^{it}$. Это приводит к так называемым характеристическим функциям, которые служат обобщением производящих функций.

Определение 2. *Характеристической функцией* случайной величины X называется комплекснозначная функция действительной переменной[1] t:

$$f(t; X) = \mathrm{E}e^{itX}, \quad -\infty < t < +\infty.$$

Очевидно, что характеристическая функция определена для любой случайной величины.

Остановимся на свойствах характеристических функций.

1. $f(0; X) = 1$.
2. $|f(t; X)| \leqslant 1$.

[1] Математическое ожидание комплекснозначной случайной величины $X(\omega) = X_1(\omega) + iX_2(\omega)$ по определению равно $\mathrm{E}X(\omega) = \mathrm{E}X_1(\omega) + i\mathrm{E}X_2(\omega)$.

3. Функция $f(t; X)$ непрерывна при $-\infty < t < +\infty$.

4. Если $f(t; X)$ — характеристическая функция величины X и $Y = aX + b$, то $f(t; Y) = e^{itb} f(at; X)$.

5. Если $X_1, \ldots, X_n$ — взаимно независимые случайные величины с характеристическими функциями $f(t; X_1), \ldots, f(t; X_n)$, то характеристическая функция суммы $X = \sum_{k=1}^{n} X_k$ равна произведению характеристических функций слагаемых: $f(t; X) = \prod_{k=1}^{n} f(t; X_k)$.

6. По характеристической функции однозначно определяется распределение вероятностей случайной величины. Общие утверждения подобного рода (так называемые *теоремы единственности*) доказываются сложно (см., например, [7], [4]). Рассмотрим пример, показывающий, как выполняется это общее утверждение.

Пусть X — целочисленная случайная величина и $f(t; X)$ — ее характеристическая функция. Тогда по определению

$$f(t; X) = \sum_{m=-\infty}^{+\infty} e^{itm} \mathrm{P}\{X = m\}.$$

Умножая обе части равенства на $\frac{e^{-itk}}{2\pi}$ и интегрируя в промежутке $(-\pi, \pi)$ по t, получим

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} f(t; X)\, dt = \sum_{m=-\infty}^{+\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-it(m-k)} \mathrm{P}\{X = m\}\, dt = \mathrm{P}\{X = k\},$$

так как

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itl}\, dt = \begin{cases} 1, & l = 0, \\ 0, & l \neq 0. \end{cases}$$

Представление типа $\mathrm{P}\{X = k\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} f(t; X)\, dt$ называется *формулой обращения*. По формуле обращения распределение полностью определяется своей характеристической функцией.

Если X имеет непрерывное распределение с плотностью $p_X(x)$, то для абсолютно интегрируемой характеристической функции $f(t; X)$ выполнено равенство

$$p_X(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx} f(t; X)\, dt.$$

7. Если $\mathrm{E}|X|^r < \infty$, то r-я производная характеристической функции существует, непрерывна и равна $f^{(r)}(0; X) = i^r \mathrm{E}X^r$, где $\mathrm{E}X^r$ — момент r-го порядка случайной величины X. В частности, если $\mathrm{E}X^2 < \infty$, то $f'(0; X) = i\mathrm{E}X$ и $f''(0; X) = -\mathrm{E}X^2$. При этом имеет место разложение функции $f(t) = f(t; X)$ в ряд Тейлора в окрестности точки 0:

$$f(\tau) = f(0) + \tau f'(0) + \frac{\tau^2}{2} f''(0) + o(\tau^2) = 1 + i\tau\mathrm{E}X - \frac{\tau^2}{2}\mathrm{E}X^2 + o(\tau^2).$$

Свойства 1—5 легко проверяются. В дальнейшем на эти свойства мы будем ссылаться, не доказывая их. Укажем характеристические функции некоторых важных распределений, которые понадобятся нам в дальнейшем.

1. Биномиальное распределение:

$$f(t; X) = \sum_{m=0}^{n} e^{itm} C_n^m p^m (1-p)^{n-m} = (1 - p + pe^{it})^n.$$

2. Распределение Пуассона:

$$f(t; X) = \sum_{m=0}^{\infty} e^{itm} e^{-\lambda} \frac{\lambda^m}{m!} = e^{\lambda(e^{it}-1)}.$$

3. Стандартное нормальное распределение:

$$f(t; X) = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \cos(tx) e^{-x^2/2}\, dx = e^{-t^2/2}$$

(пользуемся тем, что $e^{itx} = \cos tx + i \sin tx$ и что интеграл от функции $\sin(tx)e^{-x^2/2}$ равен нулю; затем пользуемся табличным интегралом

$$\int_0^{\infty} e^{-a^2x^2} \cos bx\, dx = \frac{\sqrt{\pi}}{2a} e^{-b^2/4a^2}, \quad a > 0,$$

при $a = \dfrac{1}{\sqrt{2}}$, $b = t$).

4. Распределение Коши:

$$f(t; X) = \frac{1}{\pi} \int_{-\infty}^{+\infty} e^{itx} \frac{1}{1+x^2}\, dx = e^{-|t|}$$

(пользуемся табличными интегралами

$$\int_0^{\infty} \frac{\cos ax}{1+x^2}\, dx = \frac{\pi}{2} e^a, \ a < 0, \quad \int_0^{\infty} \frac{\cos ax}{1+x^2}\, dx = \frac{\pi}{2} e^{-a}, \ a > 0).$$

5. Если $X_1, \ldots, X_n$ — взаимно независимые одинаково распределенные случайные величины с характеристической функцией $f(t; X) = f(t)$, то распределение нормированной суммы

$$S_n^* = \frac{X_1 + \ldots + X_n}{\sqrt{n}}$$

вычисляется следующим образом:

$$f(t; S_n^*) = f\Big(\frac{t}{\sqrt{n}}; X_1 + \ldots + X_n\Big) = \Big[f\Big(\frac{t}{\sqrt{n}}; X_1\Big)\Big]^n = f^n\Big(\frac{t}{\sqrt{n}}\Big).$$

§ 2. Предельная теорема Пуассона

Рассмотрим схему испытаний Бернулли, т. е. взаимно независимые случайные величины $X_1, \ldots, X_n$ с одинаковым распределением $\mathrm{P}\{X_k = 1\} = p$, $\mathrm{P}\{X_k = 0\} = 1 - p$, $0 < p < 1$. Пусть $S_n = X_1 + \ldots + X_n$, тогда, как известно, случайная величина S_n имеет биномиальное распределение, т. е.

$$\mathrm{P}\{S_n = m\} = C_n^m p^m (1-p)^{n-m}, \quad m = 0, \ldots, n.$$

Оказывается, если n велико, а p — мало и $\lambda = np$ (например, $p = 0{,}01$, $n = 300$, $\lambda = 1$), то

$$\mathrm{P}\{S_n = m\} \sim \frac{\lambda^m}{m!} e^{-\lambda}.$$

Этот факт составляет содержание предельной теоремы Пуассона (см. [7, гл. 2, § 15]; [11, т. 1, гл. 6, § 5]). Мы докажем не предельную теорему, а неравенство, преимущество которого станет позже очевидным, так как будет получена не только равномерная сходимость $\mathrm{P}\{S_n = m\}$ к $e^{-\lambda}\frac{\lambda^m}{m!}$, но и оценка ошибки при замене $\mathrm{P}\{S_n = m\}$ на $e^{-\lambda}\frac{\lambda^m}{m!}$ для конкретных значений n и p.

Напомним, что случайной величиной, распределенной по закону Пуассона с параметром $\lambda > 0$, называется такая целочисленная неотрицательная случайная величина X, что $\mathrm{P}\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, \ldots$ Математическое ожидание и дисперсия величины X равны λ, а характеристическая функция равна $f(t; X) = e^{-\lambda(e^{it}-1)}$. Подробнее о распределении Пуассона см. в книге [11, т. 1, гл. 6, § 5, 6, 7]. Мы еще вернемся к распределению Пуассона позже в связи с марковскими процессами.

Рассмотрим так называемую схему Пуассона — независимые случайные величины $X_1, \ldots, X_n$ с распределениями

$$\mathrm{P}\{X_k = 1\} = p_k, \quad \mathrm{P}\{X_k = 0\} = 1 - p_k, \quad 0 < p_k < 1, \quad k = 1, \ldots, n$$

(это соответствует последовательности испытаний с двумя исходами, вероятности которых могут меняться от испытания к испытанию). Очевидно, что при $p_1 = \ldots = p_n$ мы приходим к схеме испытаний Бернулли.

Теорема. *Пусть случайные величины $X_1, \ldots, X_n$ образуют схему Пуассона с параметрами $p_1, \ldots, p_n$, и пусть $\mu_n = X_1 + \ldots + X_n$. Пусть $\lambda = p_1 + \ldots + p_n$ и $\delta = p_1^2 + \ldots + p_n^2$. Тогда при $p_j \leqslant \frac{1}{2}$, $j = 1, \ldots, n$, имеет место неравенство*

$$\left| \mathrm{P}\{\mu_n = m\} - \frac{\lambda^m}{m!} e^{-\lambda} \right| \leqslant 2\delta.$$

Доказательство[1]**.** Было показано, что для целочисленной случайной величины X с характеристической функцией $f(t; X)$ имеет место формула обращения

$$\mathrm{P}\{X = m\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itm} f(t; X)\, dt.$$

Пользуясь тем, что характеристическая функция величины μ_n равна

$$f(t; \mu_n) = \prod_{j=1}^{n} (1 - p_j + p_j e^{it}),$$

с помощью формулы обращения получим

$$\mathrm{P}\{\mu_n = m\} - \frac{\lambda^m}{m!} e^{-\lambda} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itm} \left[\prod_{j=1}^{n} (1 - p_j + p_j e^{it}) - e^{\lambda(e^{it} - 1)} \right] dt.$$

Оценим модуль подынтегрального выражения. Пользуясь тем, что $\lambda = p_1 + \ldots + p_n$, перепишем разность в квадратных скобках следующим образом:

$$\prod_{j=1}^{n} (1 - p_j + p_j e^{it}) - e^{\lambda(e^{it} - 1)} = \prod_{j=1}^{n} (1 + p_j (e^{it} - 1)) - \prod_{j=1}^{n} e^{p_j (e^{it} - 1)}.$$

Дальнейшие оценки основаны на двух простых фактах из анализа.

Лемма 1. *Если комплексные числа a_j и b_j, $j = 1, \ldots, n$, таковы, что $|a_j| \leqslant 1$, $|b_j| \leqslant 1$, $j = 1, \ldots, n$, то*

$$\left| \prod_{j=1}^{n} a_j - \prod_{j=1}^{n} b_j \right| \leqslant \sum_{j=1}^{n} |a_j - b_j|.$$

[1] Этот вариант доказательства впервые появился в курсе лекций на факультете ВМК МГУ в 1972 году.

Лемма 2. *При* $|z| \leqslant 1$ *справедливо неравенство*

$$|e^z - 1 - z| \leqslant |z|^2.$$

Воспользуемся сначала леммой 1 при $a_j = 1 + p_j(e^{it} - 1)$, $b_j = e^{p_j(e^{it}-1)}$. Легко показать, что в этом случае $|a_j| \leqslant 1$ и $|b_j| \leqslant 1$. Тогда

$$\left|\prod_{j=1}^{n}(1 + p_j(e^{it} - 1)) - \prod_{j=1}^{n} e^{p_j(e^{it}-1)}\right| \leqslant \sum_{j=1}^{n}\left|e^{p_j(e^{it}-1)} - 1 - p_j(e^{it} - 1)\right|.$$

Ввиду того что $p_j \leqslant \frac{1}{2}$, последнее выражение по лемме 2 не превосходит $\sum\limits_{j=1}^{n} p_j^2 |e^{it} - 1|^2 = 2\delta(1 - \cos t)$. Окончательно после интегрирования (учитывая, что $|e^{-itm}| = 1$) имеем

$$\left|\mathrm{P}\{\mu_n = m\} - \frac{\lambda^m}{m!}e^{-\lambda}\right| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} 2\delta(1 - \cos t)\,dt = 2\delta,$$

что и требовалось доказать. □

Замечание. Если δ мало, то можно применять это приближение на практике. В частности, если $p_1 = \ldots = p_n = \frac{\lambda}{n}$, то $\delta = \frac{\lambda^2}{n}$ и очевидно, что δ мало при большом значении n.

§ 3. Центральная предельная теорема для независимых случайных величин

Этим названием объединяется целый ряд теорем, утверждающих, что при некоторых общих условиях распределение сумм случайных величин сходится к нормальному распределению. Мы докажем сравнительно простой вариант центральной предельной теоремы (ЦПТ) в следующей формулировке.

Теорема. *Пусть* $X_1, \ldots, X_n$ *— независимые ограниченные* ($|X_k| \leqslant L$) *одинаково распределенные случайные величины с математическим ожиданием* $a = \mathrm{E}X_k$ *и дисперсией* $\sigma^2 = \mathrm{D}X_k$. *Пусть* $S_n^* = \dfrac{S_n - \mathrm{E}S_n}{\sqrt{\mathrm{D}S_n}}$, *где* $S_n = X_1 + \ldots + X_n$. *Тогда, каковы бы ни были числа* α *и* β $(-\infty < \alpha < \beta < +\infty)$ *при* $n \to \infty$,

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \to \Phi(\beta) - \Phi(\alpha),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy$$

— функция стандартного нормального распределения.

Доказательство[1]. Предположим временно, что $a=0$ и $\sigma^2=1$. В этом случае $S_n^*=\frac{S_n}{\sqrt{n}}$.

1. Докажем сначала, что характеристические функции S_n^* сходятся к характеристической функции стандартного нормального распределения. Обозначим через $f(t)=f(t;X_k)$ характеристическую функцию случайной величины X_k. Указывалось, что в окрестности нуля

$$f(\tau)=1+i\tau\mathrm{E}X_k+\frac{\tau^2}{2}(-\mathrm{E}X_k^2)+o(\tau^2).$$

В нашем случае

$$f(\tau)=1-\frac{\tau^2}{2}+o(\tau^2).$$

Возьмем фиксированное значение t и положим $\tau=\frac{t}{\sqrt{n}}$ в разложении характеристической функции:

$$f\Big(\frac{t}{\sqrt{n}}\Big)=1-\frac{t^2}{2n}+o\Big(\frac{1}{n}\Big).$$

Тогда

$$f^n\Big(\frac{t}{\sqrt{n}}\Big)=\Big[1-\frac{t^2}{2n}+o\Big(\frac{1}{n}\Big)\Big]^n.$$

Очевидно, что правая часть в последнем соотношении при $n\to\infty$ стремится к $e^{-t^2/2}$. Покажем, что эта сходимость равномерна на каждом конечном интервале $[-T,T]$, $T>0$. Обозначим $g(t)=e^{-t^2/2n}$, тогда $e^{-t^2/2}=g^n(t)$. Для оценки $\Big|f^n\Big(\frac{t}{\sqrt{n}}\Big)-g^n(t)\Big|$ воспользуемся тем, что $|a^n-b^n|\leqslant n|a-b|$ при $|a|\leqslant 1$ и $|b|\leqslant 1$. Полагая $a=f\Big(\frac{t}{\sqrt{n}}\Big)$ и $b=g(t)$ (очевидно, что $\Big|f\Big(\frac{t}{\sqrt{n}}\Big)\Big|\leqslant 1$ и $|g(t)|\leqslant 1$, в силу того что это характеристические функции), получим

$$\Big|f^n\Big(\frac{t}{\sqrt{n}}\Big)-g^n(t)\Big|\leqslant n\Big|f\Big(\frac{t}{\sqrt{n}}\Big)-g(t)\Big|. \tag{1}$$

Далее, имеем,

$$f\Big(\frac{t}{\sqrt{n}}\Big)-g(t)=\Big[f\Big(\frac{t}{\sqrt{n}}\Big)-\Big(1-\frac{t^2}{2n}\Big)\Big]-\Big[e^{-t^2/2n}-\Big(1-\frac{t^2}{2n}\Big)\Big].$$

При оценке модуля второй разности воспользуемся тем, что при $z>0$ выполнено неравенство

$$|e^{-z}-(1-z)|\leqslant\frac{z^2}{2}$$

[1] Этот вариант доказательства ЦПТ был впервые приведен в курсе лекций Ю. В. Прохорова на механико-математическом факультете в 1966 г.

и, следовательно,

$$\left|e^{-t^2/2n}-\left(1-\frac{t^2}{2n}\right)\right|\leqslant\frac{t^4}{8n^2}. \tag{2}$$

Оценим модуль разности $f\left(\frac{t}{\sqrt{n}}\right)-\left(1-\frac{t^2}{2n}\right)$. Обозначая

$$\psi_1(t)=\left|\mathrm{E}\cos\frac{t}{\sqrt{n}}X_1-\left(1-\frac{t^2}{2n}\right)\right| \quad \text{и} \quad \psi_2(t)=\left|\mathrm{E}\sin\frac{t}{\sqrt{n}}X_1\right|$$

и пользуясь представлением характеристической функции

$$f(t)=\mathrm{E}\cos tX_1+i\mathrm{E}\sin tX_1,$$

получаем

$$\left|f\left(\frac{t}{\sqrt{n}}\right)-\left(1-\frac{t^2}{2n}\right)\right|=\sqrt{\psi_1^2(t)+\psi_2^2(t)}.$$

По известным неравенствам

$$\left|\cos u-\left(1-\frac{u^2}{2}\right)\right|\leqslant\frac{u^4}{24} \quad \text{и} \quad |\sin u-u|\leqslant\frac{u^3}{6},$$

оценим $\psi_1(t)$ и $\psi_2(t)$ в предположении, что $\mathrm{E}X_1=0$ и $\mathrm{D}X_1=1$:

$$\psi_1(t)=\left|\mathrm{E}\left(\cos\frac{t}{\sqrt{n}}X_1-\left(1-\frac{t^2X_1^2}{2n}\right)\right)\right|\leqslant\frac{t^4}{24n^2}\mathrm{E}X_1^4,$$

$$\psi_2(t)=\left|\mathrm{E}\left(\sin\frac{t}{\sqrt{n}}X_1-\frac{t}{\sqrt{n}}X_1\right)\right|\leqslant\frac{|t|^3}{6n^{3/2}}\mathrm{E}|X_1|^3.$$

Используя условие $|X_k|\leqslant L$, $k=1,\ldots,n$, получаем

$$\psi_1(t)\leqslant\frac{t^4}{24n^2}L^2 \quad \text{и} \quad \psi_1(t)\leqslant\frac{|t|^3}{6n^{3/2}}L.$$

Окончательно имеем[1]

$$\left|f\left(\frac{t}{\sqrt{n}}\right)-\left(1-\frac{t^2}{2n}\right)\right|\leqslant\sqrt{\left(\frac{t^4L^2}{24n^2}\right)^2+\left(\frac{|t|^3L}{6n^{3/2}}\right)^2}=\frac{|t|^3L}{36n^{3/2}}\sqrt{\frac{t^2L^2}{16n}+1}. \tag{3}$$

Наконец, объединим оценки (2) и (3):

$$\left|f\left(\frac{t}{\sqrt{n}}\right)-g(t)\right|\leqslant\frac{|t|^3L}{36n^{3/2}}\sqrt{\frac{t^2L^2}{16n}+1}+\frac{t^4}{8n^2},$$

[1] Если вместо условия $|X_k|\leqslant L$ предположить существование третьего абсолютного момента, можно прийти к оценке

$$\left|f\left(\frac{t}{\sqrt{n}}\right)-\left(1-\frac{t^2}{2n}\right)\right|\leqslant\frac{|t|^3}{6n^{3/2}}\mathrm{E}|X_k|^3.$$

а полученную оценку с неравенством (1):

$$\left|f^n\Big(\frac{t}{\sqrt{n}}\Big)-e^{-t^2/2}\right| \leqslant \frac{|t|^3L}{36\sqrt{n}}\sqrt{\frac{t^2L^2}{16n}+1}+\frac{t^4}{8n}. \tag{4}$$

При любом фиксированном значении $T>0$ обозначим

$$\delta(T)=\max_{|t|\leqslant T}\left|f^n\Big(\frac{t}{\sqrt{n}}\Big)-e^{-t^2/2}\right|.$$

Из неравенства (4) следует, что

$$\delta_n(T)\to 0 \quad \text{при } n\to\infty, \tag{5}$$

а это означает, что сходимость равномерна по t при $|t|\leqslant T$.

2. Пусть S_n^* — последовательность случайных величин и Z — стандартная нормальная случайная величина, причем характеристические функции $f(t;S_n^*)$ сходятся равномерно к характеристической функции $f(t;Z)$. Нужно доказать, что при любых α и β $(\alpha<\beta)$ при $n\to\infty$ выполнено соотношение

$$\mathrm{P}\{\alpha\leqslant S_n^*\leqslant\beta\}\to\mathrm{P}\{\alpha\leqslant Z\leqslant\beta\}.$$

Рассмотрим отрезок $\Delta=[\alpha,\beta]$ и функцию

$$H_\Delta(x)=\begin{cases}1, & x\in\Delta,\\ 0, & x\notin\Delta.\end{cases}$$

Тогда $H_\Delta(S_n^*)$ — случайная величина, принимающая лишь два значения:

$$H_\Delta(S_n^*)=\begin{cases}0, & S_n^*\notin\Delta,\\ 1, & S_n^*\in\Delta.\end{cases}$$

Очевидно, что $\mathrm{E}H_\Delta(S_n^*)=\mathrm{P}\{\alpha\leqslant S_n^*\leqslant\beta\}$. Подобно этому $\mathrm{E}H_\Delta(Z)=$ $=\mathrm{P}\{\alpha\leqslant Z\leqslant\beta\}$. Таким образом, нужно доказать, что

$$\mathrm{E}H_\Delta(S_n^*)\to\mathrm{E}H_\Delta(Z) \quad \text{при } n\to\infty. \tag{6}$$

3. Покажем сначала, что для достаточно гладких функций $H(x)$ имеет место сходимость

$$\mathrm{E}H(S_n^*)\to\mathrm{E}H(Z) \tag{7}$$

при условии, что характеристические функции $f(t;S_n^*)$ сходятся к характеристической функции $f(t;Z)$ и эта сходимость равномерна на каждом конечном интервале $|t|\leqslant T$. Воспользуемся следующим фактом из анализа Фурье: если функция $H(x)$ дважды непрерывно дифференцируема и обращается в нуль вне некоторого конечного интервала, то ее можно представить интегралом Фурье

$$H(x)=\int_{-\infty}^{+\infty}e^{itx}h(t)\,dt, \tag{8}$$

где $h(t)$ — абсолютно интегрируемая функция. Пусть $H(x)$ — функция, допускающая представление (8), тогда легко показать, что

$$\mathrm{E}H(S_n^*) = \int\limits_{-\infty}^{+\infty} f(t; S_n^*)h(t)\,dt.$$

Аналогично

$$\mathrm{E}H(Z) = \int\limits_{-\infty}^{+\infty} f(t; Z)h(t)\,dt,$$

где $f(t; Z) = e^{-t^2/2}$ — характеристическая функция величины Z. Таким образом, мы должны доказать, что при $n \to \infty$ для любой абсолютно интегрируемой функции $h(t)$ выполнено соотношение

$$\int\limits_{-\infty}^{+\infty} f(t; S_n^*)h(t)\,dt \to \int\limits_{-\infty}^{+\infty} f(t; Z)h(t)\,dt, \tag{9}$$

если характеристическая функция величины S_n^* равномерно сходится к характеристической функции величины Z. Для любого $T > 0$ в силу доказанного при $n \to \infty$ имеем

$$\delta_n(T) = \max_{|t| \leqslant T} |f(t; S_n^*) - f(t; Z)| \to 0.$$

Получаем оценку

$$\left| \int\limits_{-\infty}^{+\infty} [f(t; S_n^*) - f(t; Z)]h(t)\,dt \right| \leqslant$$

$$\leqslant \int\limits_{-\infty}^{+\infty} |f(t; S_n^*) - f(t; Z)| \cdot |h(t)|\,dt = J_1(T) + J_2(T),$$

где

$$J_1(T) = \int\limits_{-T}^{T} |f(t; S_n^*) - f(t; Z)| \cdot |h(t)|\,dt,$$

$$J_2(T) = \int\limits_{-\infty}^{-T} |f(t; S_n^*) - f(t; Z)| \cdot |h(t)|\,dt + \int\limits_{T}^{\infty} |f(t; S_n^*) - f(t; Z)| \cdot |h(t)|\,dt.$$

При оценке интеграла $J_1(T)$ заменим модуль разности характеристических функций на $\delta_n(T)$:

$$J_1(T) \leqslant \delta_n(T) \int\limits_{-\infty}^{+\infty} |h(t)|\,dt.$$

Интеграл $J_2(T)$ оценим следующим образом:

$$J_2(T) \leqslant 2\int\limits_{-\infty}^{-T} |h(t)|\,dt + 2\int\limits_{T}^{\infty} |h(t)|\,dt.$$

Пусть $\varepsilon > 0$ произвольно. Выберем $T_\varepsilon > 0$ так, чтобы выполнялось неравенство

$$2\int\limits_{-\infty}^{-T_\varepsilon} |h(t)|\,dt + 2\int\limits_{T_\varepsilon}^{\infty} |h(t)|\,dt \leqslant \frac{\varepsilon}{2}$$

(это возможно в силу абсолютной интегрируемости функции $h(t)$). Затем найдем такое $n_0(\varepsilon)$, что при $n > n_0(\varepsilon)$

$$\delta_n(T_\varepsilon)\int\limits_{-\infty}^{+\infty} |h(t)|\,dt \leqslant \frac{\varepsilon}{2}.$$

Окончательно имеем

$$\left|\int\limits_{-\infty}^{+\infty} (f(t;S_n^*) - f(t;Z))h(t)\,dt\right| \leqslant \varepsilon.$$

Итак, показано, что из сходимости характеристических функций S_n^* к характеристической функции Z следует, что

$$\mathrm{E}H(S_n^*) \to \mathrm{E}H(Z) \quad \text{при } n \to \infty$$

для всех функций $H(x)$, удовлетворяющих указанным выше условиям.

4. Докажем теперь, что при любых α, β $(\alpha < \beta)$ выполнено соотношение

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \to \mathrm{P}\{\alpha \leqslant Z \leqslant \beta\}.$$

Зададим $\varepsilon > 0$. Введем наряду с Δ отрезки

$$\Delta_1 = \left[\alpha - \frac{\varepsilon}{2}, \beta + \frac{\varepsilon}{2}\right], \quad \Delta_2 = \left[\alpha + \frac{\varepsilon}{2}, \beta - \frac{\varepsilon}{2}\right]$$

и рассмотрим наряду с $H_\Delta(x)$ такие функции $H_\varepsilon^+(x)$ и $H_\varepsilon^-(x)$, удовлетворяющие условиям гладкости, что

$$H_\varepsilon^+(x) = 1 \ \text{ при } x \in \Delta_1, \quad H_\varepsilon^+(x) = 0 \ \text{ при } x \leqslant a - \frac{\varepsilon}{2} \ \text{ и } \ x \geqslant b + \frac{\varepsilon}{2},$$

$$H_\varepsilon^-(x) = 1 \ \text{ при } x \in \Delta_2, \quad H_\varepsilon^-(x) = 0 \ \text{ при } x \leqslant a \ \text{ и } \ x \geqslant b$$

(см. рис. 1). Очевидно, что

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} = \mathrm{E}H_\Delta(S_n^*) \leqslant \mathrm{E}H_\varepsilon^+(S_n^*).$$

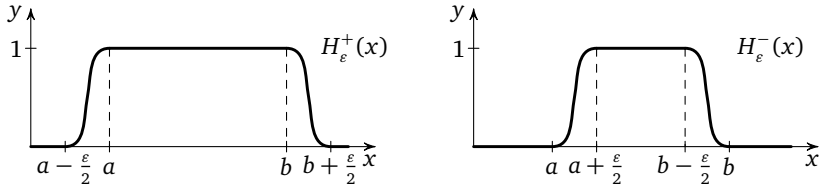


Рис. 1. $H_\varepsilon^+(x)$ и $H_\varepsilon^-(x)$: $H_\varepsilon^-(x) \leqslant H_\Delta(x) \leqslant H_\varepsilon^+(x)$

Оценим сверху $\mathrm{E}H_\varepsilon^+(S_n^*)$. В силу предположения, что для $H_\varepsilon^+(S_n^*)$ справедливо соотношение (7), заключаем, что для произвольно взятого $\varepsilon > 0$ существует такой номер $n_1 = n_1(\varepsilon)$, что при $n \geqslant n_1$ выполнено неравенство

$$\mathrm{E}H_\varepsilon^+(S_n^*) \leqslant \mathrm{E}H_\varepsilon^+(Z) + \frac{\varepsilon}{2}.$$

Отсюда получаем

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \leqslant \mathrm{E}H_\varepsilon^+(Z) + \frac{\varepsilon}{2} \leqslant$$

$$\leqslant \int\limits_{\alpha-\varepsilon/2}^{\beta+\varepsilon/2} H_\varepsilon^+(z)\varphi(z)\,dz + \frac{\varepsilon}{2} = \mathrm{P}\Big\{\alpha - \frac{\varepsilon}{2} \leqslant Z \leqslant \beta + \frac{\varepsilon}{2}\Big\} + \frac{\varepsilon}{2}.$$

В силу того что $\varphi(z) < \frac{1}{2}$, имеем

$$\mathrm{P}\Big\{\alpha - \frac{\varepsilon}{2} \leqslant Z \leqslant \beta + \frac{\varepsilon}{2}\Big\} =$$

$$= \int\limits_{\alpha-\varepsilon/2}^{\alpha} \varphi(z)\,dz + \int\limits_{\alpha}^{\beta} \varphi(z)\,dz + \int\limits_{\beta}^{\beta+\varepsilon/2} \varphi(z)\,dz \leqslant \int\limits_{\alpha}^{\beta} \varphi(z)\,dz + \frac{\varepsilon}{2}.$$

Таким образом, для любого $\varepsilon > 0$ существует такой номер n_1, что при $n \geqslant n_1$ имеет место неравенство

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \leqslant \mathrm{P}\{\alpha \leqslant Z \leqslant \beta\} + \varepsilon. \tag{10}$$

Аналогично, используя функцию $H_\varepsilon^-(S_n^*)$, докажем, что

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \geqslant \mathrm{P}\{\alpha \leqslant Z \leqslant \beta\} - \varepsilon \tag{11}$$

при $n \geqslant n_2(\varepsilon)$. Из неравенств (10) и (11) мы получаем следующий результат: при любых конечных $\alpha < \beta$ и $n \to \infty$

$$\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} \to \Phi(\beta) - \Phi(\alpha).$$

В начале доказательства мы ограничились случайными величинами с нулевым математическим ожиданием и единичной дисперсией. Если $\mathrm{E}X_k = a$ и $\mathrm{D}X_k = \sigma^2$, то нормированную сумму S_n^* можно представить в виде, для которого проводилось доказательство:

$$S_n^* = \frac{X_1 + \ldots + X_n - na}{\sigma\sqrt{n}} = \frac{\frac{X_1 - a}{\sigma} + \ldots + \frac{X_n - a}{\sigma}}{\sqrt{n}} = \frac{X_1' + \ldots + X_n'}{\sqrt{n}},$$

где $X_k' = \dfrac{X_k - a}{\sigma}$, причем $\mathrm{E}X_k' = 0$ и $\mathrm{D}X_k' = 1$. Таким образом, теорема доказана для любых независимых одинаково распределенных ограниченных случайных величин. □

§ 4. Замечания к центральной предельной теореме

1. Доказательство ЦПТ проводилось для конечных α и β. На самом деле предельное соотношение остается верным, если или $\alpha = -\infty$, или $\beta = +\infty$. Например, при $\beta = +\infty$ имеем

$$\mathrm{P}\{\alpha \leqslant S_n^*\} \to 1 - \Phi(\alpha).$$

Непосредственно из данного нами доказательства ЦПТ этот факт не следует, необходимо дополнительное рассуждение.

2. Производя при доказательстве ЦПТ более строгие оценки, можно получить представление о точности действия приближенной формулы. Так, например, можно получить следующее неравенство: при любых α и β если $|X_k| \leqslant L$, то

$$|\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} - (\Phi(\beta) - \Phi(\alpha))| \leqslant C\frac{L}{\sigma\sqrt{n}},$$

где[1] C — абсолютная постоянная, $0 < C < 1$. Таким образом, погрешность имеет порядок $\dfrac{1}{\sqrt{n}}$. Для практического использования это довольно грубая оценка, например, для того чтобы ошибка была меньше 0,01, нужно взять n порядка 10 000.

Следует отметить, что польза основного соотношения ЦПТ определяется не величиной абсолютной погрешности

$$|\mathrm{P}\{\alpha \leqslant S_n^* \leqslant \beta\} - (\Phi(\beta) - \Phi(\alpha))|,$$

а относительной ошибкой, т. е. отношением этой величины к $\Phi(\beta) - \Phi(\alpha)$, которое должно быть мало.

[1] Это неравенство есть частный случай неравенства Берри—Эссеена для отклонения распределения нормированных сумм от стандартного нормального распределения.

3. Если некоторые предположения ЦПТ нарушаются, то сам факт сходимости распределений сумм случайных величин к нормальному распределению может быть неверен. Пусть, например, случайные величины $X_1, \ldots, X_n$ одинаково распределены по стандартному закону Коши с плотностью $p(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$. Характеристическая функция распределения Коши равна $f(t)=e^{-|t|}$. Сумма $S_n=X_1+\ldots+X_n$ имеет характеристическую функцию $e^{-n|t|}$, а так как по характеристической функции распределение определяется однозначно, сумма S_n будет также распределена по закону Коши, но с другим параметром. Никакого сближения между характеристическими функциями распределения Коши и нормального распределения, а стало быть, и между самими распределениями не наблюдается (ср. графики характеристической функции Коши и нормального распределения на рис. 2, 3). Такой вывод объясняется нарушением условий ЦПТ, а именно, случайные величины с распределением Коши имеют не только бесконечную дисперсию, но и бесконечное математическое ожидание (напомним, что существование математического ожидания в этом случае равносильно существованию и конечности интеграла

$$\int\limits_{-\infty}^{+\infty} |x|p(x)\,dx = \frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\frac{|x|}{1+x^2}\,dx,$$

однако интеграл расходится). Отметим кстати, что по той же причине для последовательности случайных величин с плотностью Коши не выполняется закон больших чисел: арифметическое среднее $\frac{X_1+\ldots+X_n}{n}$ будет иметь характеристическую функцию, равную $e^{-|t|}$, т. е. такую же, какой обладает отдельное слагаемое, и, следовательно, арифметическое среднее $\frac{S_n}{n}$ случайных величин $X_1, \ldots, X_n$ будет иметь то же самое распределение Коши, что и каждая случайная величина X_k.

4. Рассмотрим в ЦПТ случайные величины с одним и тем же распределением

$$\mathrm{P}\{X_k=1\}=p, \quad \mathrm{P}\{X_k=0\}=1-p, \quad 0<p<1.$$

Тогда, очевидно, $a=\mathrm{E}X_k=p$, $\sigma^2=\mathrm{D}X_k=p(1-p)$ и при $n\to\infty$ выполнено соотношение

$$\mathrm{P}\Big\{\alpha\leqslant\frac{S_n-np}{\sqrt{np(1-p)}}\leqslant\beta\Big\}\to\Phi(\beta)-\Phi(\alpha).$$

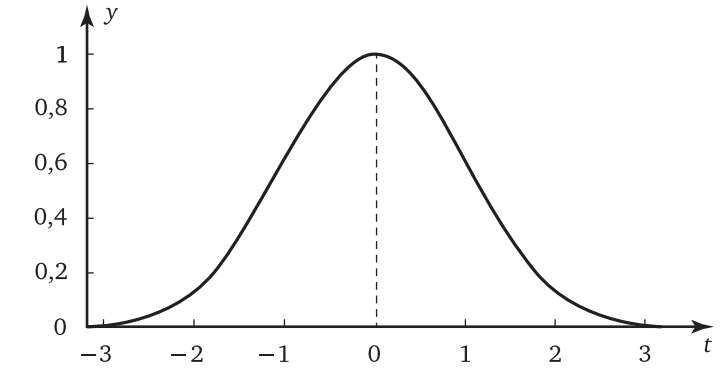


Рис. 2. Характеристическая функция $f(t)=\exp\left(-\frac{t^2}{2}\right)$, $-\infty<t<\infty$, нормального распределения с плотностью распределения вероятностей $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $-\infty<x<\infty$

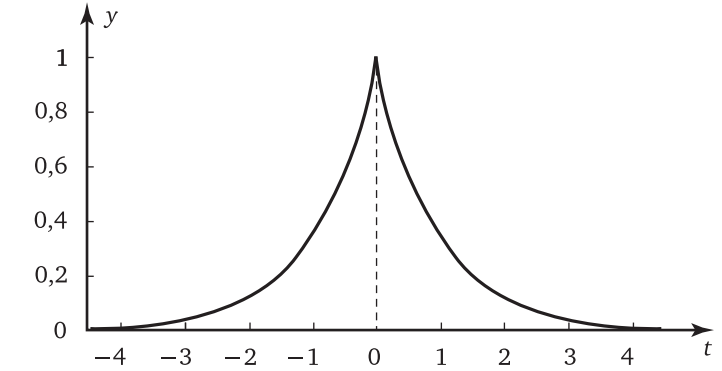


Рис. 3. Характеристическая функция $f(t)=e^{-|t|}$, $-\infty<t<\infty$, для плотности распределения Коши $p(x)=\frac{1}{\pi}\cdot\frac{1}{x^2+1}$, $-\infty<x<\infty$

В подобной частной форме теорема называется теоремой Лапласа и является исторически первым вариантом ЦПТ. Смысл теоремы Лапласа в том, что нормированное число «удач» в схеме Бернулли имеет асимптотически нормальное распределение с параметрами $(0, 1)$. Для этого случая получена более точная приближенная формула (см., например, [11, т. 1, гл. 7, § 3])

$$\mathrm{P}\left\{np+\alpha\sqrt{np(1-p)}+\frac{1}{2}\leqslant X_1+\ldots+X_n\leqslant np+\beta\sqrt{np(1-p)}-\frac{1}{2}\right\}\sim$$
$$\sim\Phi(\beta)-\Phi(\alpha),\quad(1)$$

где введена поправка на дискретность распределения, которая при больших n не существенна, а при малых n улучшает результат.

Пусть, например, в схеме Бернулли $n=100$, $p=\frac{1}{2}$, что соответствует 100 бросаниям симметричной монеты. Пусть $S_{100}=X_1+\ldots+X_{100}$ — число появлений «орла» при 100 бросаниях. Найдем с помощью теоремы Лапласа вероятность того, что $35\leqslant S_{100}\leqslant 65$. Очевидно, что $np=50$, $\sqrt{np(1-p)}=5$ и $\alpha=-3$, $\beta=3$. Тогда

$$\Phi(3)-\Phi(-3)=0{,}99730,$$

так как по таблицам нормального распределения $\Phi(3)=0{,}99865$, а $\Phi(-3)=1-\Phi(3)$. Можно увеличить точность, пользуясь формулой (1): так как в этом случае $\alpha=-3{,}1$, $\beta=3{,}1$, а $\Phi(3{,}1)=0{,}99903$, получаем, что

$$\Phi(3{,}1)-\Phi(-3{,}1)=0{,}99806.$$

О точности этих приближений можно судить, сравнивая их с вероятностью, вычисленной по таблицам биномиального распределения:

$$\mathrm{P}\{35\leqslant S_{100}\leqslant 65\}=0{,}99822.$$

§ 5. Применение центральной предельной теоремы к статистическим задачам[1]

1. Проверка *статистической гипотезы* о том, что вероятность «удачи» в схеме Бернулли равна заданному числу.

Пусть при 10 000 бросаний монеты «орел» появился 5400 раз. Согласуются ли опытные данные с гипотезой о том, что монета симметрична, т. е. $p=\frac{1}{2}$. Пусть S_n — случайное число «орлов». Будем считать, что данная гипотеза верна, и вычислим вероятность события $\{S_n\geqslant 5400\}$, в которое включается наблюдаемое нами событие $\{S_n=5400\}$: при $n=10\,000$ имеем

$$\mathrm{P}\{S_n\geqslant 5400\}=\mathrm{P}\{S_n^*\geqslant 8\}\sim 1-\Phi(8),$$

так как по теореме Лапласа величина $S_n^*=\dfrac{S_n-\mathrm{E}S_n}{\sqrt{\mathrm{D}S_n}}$ имеет асимптотически нормальное распределение. Вероятность, стоящая в правой части, во всяком случае меньше $16\cdot 10^{-16}$. Таким образом, мы наблюдаем весьма редкое событие и должны, скорее всего, отклонить гипотезу о симметричности монеты.

[1] Статистические задачи, или задачи математической статистики, и соответствующие понятия и термины составляют содержание гл. 5.

2. Построение доверительного интервала для неизвестной вероятности «удачи» в схеме Бернулли.

Пусть в схеме испытаний Бернулли с вероятностью «удачи» p, $0<p<1$, n — число испытаний, S_n — случайное число «удач», $p^*=\frac{S_n}{n}$ — частота «удачи». Пользуясь теоремой Лапласа, можно оценить вероятность отклонения частоты «удачи» от вероятности «удачи» p на заданную величину ε:

$$\mathrm{P}\{|p^*-p|<\varepsilon\}=\mathrm{P}\left\{\left|\frac{S_n}{n}-p\right|<\varepsilon\right\}\sim 2\Phi\Big(\varepsilon\sqrt{\frac{n}{p(1-p)}}\Big)-1.$$

Обозначим $\lambda=\varepsilon\sqrt{\frac{n}{p(1-p)}}$. По заданному α, $0<\alpha<1$, можно найти из таблиц нормального распределения такое значение λ, что $\Phi(\lambda)=1-\frac{\alpha}{2}$; корень этого уравнения обозначим $\lambda_{1-\alpha/2}$ (этот корень называют *квантилью* уровня $1-\frac{\alpha}{2}$). Таким образом, можно утверждать, что с вероятностью, близкой к $1-\alpha$, происходит событие

$$\Big\{|p^*-p|<\lambda_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\Big\}, \tag{1}$$

причем это верно при любом значении p, $0<p<1$. Обычно на практике p неизвестно, но известно, что $\{S_n=m\}$, где m — полученное число «удач», и, таким образом, частота p^* принимает значение $\frac{m}{n}$. Необходимо указать границы для неизвестной вероятности p, основываясь на значении частоты. Очевидно, что $0<p<1$. Указание других границ связано с риском совершить ошибку. Вероятность этой ошибки α, назначаемая заранее, называется *доверительным уровнем*, а вероятность $1-\alpha$ называется *доверительной вероятностью*. Так как p^* известно, неравенство (1), выполняющееся с вероятностью, близкой к $1-\alpha$, дает возможность найти пределы, в которых лежит неизвестное значение p. Эти границы находятся как корни $\widehat{p}_1$ и $\widehat{p}_2$ квадратного уравнения

$$(p^*-p)^2=\lambda^2_{1-\alpha/2}\frac{p(1-p)}{n}$$

и равны

$$\widehat{p}_{1,2}=\frac{n}{n+\lambda^2}\Big(p^*+\frac{\lambda}{2n}\mp\sqrt{\frac{p^*(1-p^*)}{n}+\frac{\lambda^2}{4n^2}}\Big).$$

Полученный интервал $(\widehat{p}_1,\widehat{p}_2)$ со случайными границами называется *доверительным интервалом* для неизвестного значения p, соответ-

ствующим доверительному уровню α. Следовательно, если утверждается, что интервал $(\widehat{p}_1, \widehat{p}_2)$ накрывает значение p, то вероятность ошибочного заключения равна α. С помощью доверительного интервала можно проверить гипотезу о том, что вероятность «удачи» в схеме Бернулли равна заданному числу p_0. Если $(\widehat{p}_1, \widehat{p}_2)$ — доверительный интервал, соответствующий доверительному уровню α, то проверка гипотезы $p = p_0$ проводится следующим образом: если оказывается, что $p_0 < \widehat{p}_1$ или $p_0 > \widehat{p}_2$, то гипотеза отвергается, если же $\widehat{p}_1 \leqslant p_0 \leqslant \widehat{p}_1$, то считаем, что гипотеза $p = p_0$ не противоречит полученным наблюдениям. При этом доля случаев, в которых мы будем отвергать на самом деле справедливую гипотезу, приближенно равна α. Рассмотрим данные пункта 1 на с. 75 и проверим с помощью доверительного интервала гипотезу $p = \frac{1}{2}$. Доверительному уровню $\alpha = 0{,}05$ соответствует доверительный интервал с границами $\widehat{p}_1 = 0{,}535$ и $\widehat{p}_2 = 0{,}545$. Так как $p = 0{,}5 \notin (\widehat{p}_1, \widehat{p}_2)$, мы отвергаем гипотезу $p = \frac{1}{2}$; при этом, поступая так, мы будем ошибаться в среднем в 5 случаях из 100.

Следующая табличка показывает, как при увеличении числа испытаний n сужается доверительный интервал (доверительного уровня 0,05 при значении частоты $p^* = 0{,}6$).

n	m	$\widehat{p}_1$	$\widehat{p}_2$
10	6	0,262	0,878
20	12	0,301	0,809
30	18	0,406	0,773
40	24	0,433	0,751
100	60	0,497	0,697

3. Различение двух гипотез о вероятности «удачи» в схеме Бернулли.

В простейшей формулировке эта задача о проверке двух гипотез ставилась в связи с оценкой доли белых шаров в урне (см. с. 12). Предположим, что в схеме Бернулли значение вероятности «удачи» p неизвестно, однако известно, что либо $p = p_1$ (гипотеза H_1), либо $p = p_2$ (гипотеза H_2), где p_1 и p_2 — заданные числа, причем $0 < p_1 < p_2 < 1$. Мы должны по результатам наблюдений, т. е. по числу «удач» m в n независимых испытаниях, сделать заключение о том, какая из двух гипотез верна. Безошибочный ответ на этот вопрос невозможен, поэтому выберем следующий способ действия:

фиксируем некоторое значение m_0, $0 < m_0 < n$ и если $m \leqslant m_0$, то принимаем H_1, а если $m > m_0$, то принимаем H_2.

При этом, очевидно, могут возникнуть ошибки. Ошибка 1-го рода: гипотеза H_1 отвергается, тогда как она верна. Вероятность ошибки 1-го рода равна вероятности получения $m > m_0$ «удач» в n испытаниях при условии, что $p = p_1$, т. е.

$$\alpha' = \sum_{m=m_0+1}^{n} C_n^m p_1^m (1-p_1)^{n-m}.$$

Ошибка 2-го рода: гипотеза H_1 принимается, тогда как верна не она, а гипотеза H_2. Вероятность ошибки 2-го рода равна вероятности получения $m \leqslant m_0$ «удач» в n испытаниях при условии, что $p = p_2$, т. е.

$$\beta' = \sum_{m=0}^{m_0} C_n^m p_2^m (1-p_2)^{n-m}.$$

Допустим, что наперед заданы малые числа α и β (например, $\alpha \leqslant 0{,}05$; $\beta \leqslant 0{,}05$). Каково должно быть наименьшее n, для которого существует такое число m_0, что можно различить две гипотезы о вероятности «удачи» p с вероятностями ошибок 1-го и 2-го рода, не превосходящими соответственно заданных чисел α и β? Итак, мы будет разыскивать n, начиная с которого

$$\alpha' \leqslant \alpha, \quad \beta' \leqslant \beta.$$

Укажем приближенное решение этой задачи с помощью теоремы Лапласа. Точное решение задачи см. в гл. 5, с. 130.

Зададим сначала квантиль x_γ как корень уравнения $\Phi(x_\gamma) = \gamma$, где $\Phi(x)$ — функция стандартного нормального распределения и $0 < \gamma < 1$. Так как $\Phi(x) + \Phi(-x) = 1$ и $\Phi(x_{1-\gamma}) = 1 - \gamma$, очевидно, имеем $-x_\gamma = x_{1-\gamma}$. По теореме Лапласа при $p = p_1$ заменим вероятность

$$\mathrm{P}\{S_n > m_0\} = \mathrm{P}\Big\{ \frac{S_n - np_1}{\sqrt{np_1(1-p_1)}} > \frac{m_0 - np_1}{\sqrt{np_1(1-p_1)}} \Big\}$$

величиной $1 - \Phi\Big(\dfrac{m_0 - np_1}{\sqrt{np_1(1-p_1)}} \Big)$. Определим m_0 из уравнения

$$\Phi\Big(\frac{m_0 - np_1}{\sqrt{np_1(1-p_1)}} \Big) = 1 - \alpha.$$

Следовательно,

$$x_{1-\alpha} = \frac{m_0 - np_1}{\sqrt{np_1(1-p_1)}}.$$

Таким образом,

$$m_0 = np_1 + x_{1-\alpha}\sqrt{np_1(1-p_1)}. \tag{2}$$

Аналогично при $p = p_2$ заменим вероятность

$$\mathrm{P}\{S_n \leqslant m_0\} = \mathrm{P}\Big\{\frac{S_n - np_2}{\sqrt{np_2(1-p_2)}} \leqslant \frac{m_0 - np_2}{\sqrt{np_2(1-p_2)}}\Big\}$$

величиной $\Phi\Big(\frac{m_0 - np_2}{\sqrt{np_2(1-p_2)}}\Big)$. Далее положим

$$\Phi\Big(\frac{m_0 - np_2}{\sqrt{np_2(1-p_2)}}\Big) = \beta,$$

тогда

$$x_\beta = -x_{1-\beta} = \frac{m_0 - np_2}{\sqrt{np_2(1-p_2)}}$$

и

$$m_0 = np_2 - x_{1-\beta}\sqrt{np_2(1-p_2)}. \tag{3}$$

Объединяя равенства (2) и (3), получаем

$$n(p_2 - p_1) = x_{1-\alpha}\sqrt{np_1(1-p_1)} + x_{1-\beta}\sqrt{np_2(1-p_2)}.$$

Окончательно значение n, при котором можно различить гипотетические значения p_1 и p_2 с заданными вероятностями ошибок 1-го и 2-го рода, определяется приближенным равенством

$$n = \frac{\big(x_{1-\alpha}\sqrt{np_1(1-p_1)} + x_{1-\beta}\sqrt{np_2(1-p_2)}\big)^2}{(p_2 - p_1)^2},$$

где α, β, p_1, p_2 заданы. Это значение числа n приближенное, так как получено с помощью теоремы Лапласа.

Пример. Приведем в следующей табличке результаты решения задачи различения гипотез о вероятности «удачи» в испытаниях Бернулли при $\alpha = \beta = 0{,}05$.

p_1	p_2	n	$m_0(n)$
0,1	0,5	13	3
0,3	0,5	67	26
0,1	0,2	135	19
0,05	0,1	248	21

Глава 4

Последовательности зависимых случайных величин. Цепи Маркова. Простейшие марковские процессы в задачах массового обслуживания

Последовательности зависимых случайных величин

Зависимость между двумя случайными величинами X и Y в полной мере характеризуется условными распределениями одной из этих величин при фиксированных значениях другой. Например, если в дискретном случае $\mathrm{P}\{X = x_i\}$, $i = 1, \ldots, s$, и $\mathrm{P}\{Y = y_j\}$, $j = 1, \ldots, r$, — распределения величин X и Y, а $\mathrm{P}\{X = x_i, Y = y_j\}$, $i = 1, \ldots, s$, $j = 1, \ldots, r$, — совместное распределение X и Y, то условное распределение Y при фиксированном значении x величины X определяется набором вероятностей

$$\mathrm{P}\{Y = y_i \mid X = x\} = \frac{\mathrm{P}\{X = x, Y = y_j\}}{\mathrm{P}\{X = x\}}$$

для всех y_j. Закономерности в изменении условного закона распределения Y при изменении значений величины X указывают на вид зависимости Y от X. В непрерывном случае условное распределение величины Y характеризуется условной плотностью, определяемой через плотность совместного распределения X и Y и плотность распределения величины X.

При переходе к нескольким случайным величинам $X_1, X_2, \ldots, X_n$ зависимость характеризуется условными распределениями вида $\mathrm{P}\{X_n \in A_n \mid X_1 \in A_1, X_2 \in A_2, \ldots, X_{n-1} \in A_{n-1}\}$, где $A_1, A_2, \ldots, A_{n-1}, A_n$ — фиксированные множества значений случайных величин $X_1, X_2, \ldots, X_{n-1}, X_n$. Ясно, что формы зависимости между величинами в последовательности $X_1, X_2, \ldots, X_n$ многообразны. В настоящей главе мы будем изучать самые простые формы зависимости, стараясь обобщить важные результаты, полученные для независимых случайных величин.

§1. Вариант закона больших чисел для зависимых случайных величин. Теорема Бернштейна

В гл. 2 был доказан закон больших чисел для последовательности взаимно независимых случайных величин. Остается ли спра-

ведливым закон больших чисел для последовательности зависимых случайных величин? Как уже отмечалось, для доказательства теоремы Чебышёва достаточно было предположить некоррелированность случайных величин. Можно ли в том же направлении ослабить достаточные условия для справедливости закона больших чисел? Для ответа на этот вопрос вернемся к неравенству Чебышёва, на которое опирается закон больших чисел в форме Чебышёва. Пусть случайные величины $X_1, \ldots, X_n$ имеют конечные математические ожидания $\mathrm{E}X_k$ и дисперсии $\mathrm{D}X_k$, и пусть

$$S_n = X_1 + \ldots + X_n.$$

Тогда для любого $\varepsilon > 0$ выполнено неравенство

$$\mathrm{P}\left\{\left|\frac{S_n}{n} - \mathrm{E}\frac{S_n}{n}\right| > \varepsilon\right\} < \frac{\mathrm{D}S_n}{n^2\varepsilon^2}.$$

Ясно, что если $\dfrac{\mathrm{D}S_n}{n^2} \to 0$ при $n \to \infty$, то ЗБЧ имеет место. В общем случае $\mathrm{D}S_n$ имеет при больших n порядок n^2 и $\dfrac{\mathrm{D}S_n}{n^2}$ к 0 не стремится. В следующей теореме будет дано обобщение теоремы Чебышёва на один класс зависимых случайных величин.

Теорема (Бернштейн (см. [1, с. 180—181])). *Пусть случайные величины $X_1, \ldots, X_n$ имеют математические ожидания $\mathrm{E}X_i$, дисперсии $\mathrm{D}X_i = \sigma_i^2$ и коэффициенты корреляции $R_{ij} = R(X_i, X_j)$. Если дисперсии всех рассматриваемых величин равномерно ограничены, т. е. $\sigma_i^2 \leqslant C$, $C > 0$, и коэффициенты корреляции R_{ij} между величинами X_i и X_j равномерно стремятся к нулю при $|i - j| \to \infty$, то к последовательности $X_1, \ldots, X_n$ применим ЗБЧ.*

Доказательство. Достаточно показать, что

$$\frac{\mathrm{D}(X_1 + \ldots + X_n)}{n^2} \to 0 \quad \text{при } n \to \infty.$$

По условию, каково бы ни было число $\alpha > 0$, найдется сколь угодно большое число h, не зависящее от n и такое, что при $|i - k| \geqslant h$ будет справедливо неравенство $|R_{ik}| < \alpha$. По ранее доказанному (см. теорему 5 на с. 38) дисперсия суммы произвольных случайных величин равна

$$\mathrm{D}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \sigma_k^2 + 2\sum_{i<j} R_{ij}\sigma_i\sigma_j.$$

Представим для удобства $\mathrm{D}(X_1 + \ldots + X_n)$ в виде суммы

$$\sum\nolimits_1 + \sum\nolimits_2 + \ldots + \sum\nolimits_n,$$

где

$$\sum_1 = \sigma_1^2 + R_{12}\sigma_1\sigma_2 + R_{13}\sigma_1\sigma_3 + \ldots + R_{1n}\sigma_1\sigma_n,$$
$$\sum_2 = R_{21}\sigma_2\sigma_1 + \sigma_2^2 + R_{23}\sigma_2\sigma_3 + \ldots + R_{2n}\sigma_2\sigma_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\sum_n = R_{n1}\sigma_n\sigma_1 + R_{n2}\sigma_n\sigma_2 + \ldots + \sigma_n^2.$$

Слагаемые любой суммы $\sum_k$ можно разбить на две группы: одна объединяет слагаемые с R_{ik}, для которых $|i-k| \geqslant h$, другая — с R_{ik}, для которых $|i-k| < h$. Слагаемых с $|i-k| < h$ всего $2h-1$, а число слагаемых с $|i-k| \geqslant h$ равно $n-2h+1$. Так как $\sigma_i^2 \leqslant C$ и $|R_{ik}| \leqslant 1$, для любых i и k имеем

$$|R_{ik}\sigma_i\sigma_k| \leqslant C,$$

а для i и k, удовлетворяющих условию $|i-k| \geqslant h$, имеем

$$|R_{ik}\sigma_i\sigma_k| < C\alpha.$$

Тогда при любом k выполнено неравенство

$$|\sum_k| < (n-2h+1)C\alpha + (2h-1)C$$

или во всяком случае

$$|\sum_k| < C(n\alpha + 2h).$$

Поэтому

$$\frac{\mathrm{D}(X_1 + \ldots + X_n)}{n^2} < 2C\frac{h}{n} + C\alpha.$$

Возьмем теперь произвольно малое число $\delta > 0$ и выберем α из условия $C\alpha < \frac{\delta}{2}$. После этого при определившемся значении $h = h(\alpha)$ возьмем число n столь большим, чтобы

$$2C\frac{h}{n} < \frac{\delta}{2}.$$

Тогда

$$\frac{\mathrm{D}(X_1 + \ldots + X_n)}{n^2} < \delta,$$

что и требовалось доказать. □

Замечание. Таким образом, если зависимость между X_i и X_k равномерно ослабевает с ростом $|i-k|$, то к величинам $X_1, \ldots, X_n$ применим ЗБЧ. При доказательстве теоремы существенно, чтобы положительные коэффициенты корреляции R_{ik} стремились к 0 при $|i-k| \to \infty$, так как слагаемые $R_{ik}\sigma_i\sigma_k$ с отрицательными R_{ik} можно отбросить.

§ 2. Цепи Маркова

Этот параграф посвящен важнейшему классу зависимых случайных величин — величин, связанных в цепь Маркова (по имени академика А. А. Маркова, который в 1907 г. впервые рассмотрел подобную зависимость). Понятие цепи Маркова, или марковской последовательности случайных величин, является естественным обобщением независимости.

Определение 1. Рассмотрим квадратную матрицу

$$\Pi = \begin{Vmatrix} p_{11} & \dots & p_{1s} \\ \dots & \dots & \dots \\ p_{s1} & \dots & p_{ss} \end{Vmatrix}$$

порядка $s \in \mathbb{N}$, элементы которой p_{ij} удовлетворяют следующим требованиям:

1) $p_{ij} \geqslant 0$, $i, j = 1, \dots, s$, 2) $\sum_{j=1}^{s} p_{ij} = 1$ при любом $i = 1, \dots, s$.

Матрицы со свойствами 1—2 называются *стохастическими*.

Определение 2. Пусть каждая из случайных величин $X_0, X_1, \dots, X_n$ принимает значения из конечного множества $\mathscr{X} = (x_1, x_2, \dots, x_s)$. Говорят, что эти величины связаны в *простую однородную цепь Маркова* с конечным числом значений, если существуют такие неотрицательные числа p_i, $i = 1, \dots, s$, $\sum_{i=1}^{s} p_i = 1$, и такая стохастическая матрица $\Pi = \|p_{ij}\|$ порядка s, что при любом наборе $x_{i_0}, \dots, x_{i_n} \in \mathscr{X}$ выполняется равенство

$$\mathrm{P}\{X_0 = x_{i_0}, X_1 = x_{i_1}, \dots, X_n = x_{i_n}\} = p_{i_0} p_{i_0, i_1} \dots p_{i_{n-1}, i_n}.$$

Очевидно, что эта формула задает совместное распределение случайных величин $X_0, \dots, X_n$ при $n \in \mathbb{N}$.

Пример 1. Как показал Марков, последовательность гласных и согласных звуков в русской речи представляет собой в первом приближении цепь Маркова. На материале 20 000 знаков «Евгения Онегина» Марковым были подсчитаны следующие характеристики:

- частота гласной = 0,4319,
- частота гласной после согласной = 0,663,
- частота гласной после гласной = 0,128.

Схема Бернулли независимых испытаний с постоянной вероятностью «удачи», очевидно, неприменима, поскольку эти вероятности существенно отличаются друг от друга.

Пример 2. Рассмотрим испытания Бернулли. Пусть событие A_k — «удача» в k-м испытании и $p = \mathrm{P}\{A_k\}$. Если ввести обозначение 1 для «удачи» и 0 для «неудачи», то результат n испытаний запишется последовательностью единиц и нулей длины n. Рассмотрим все возможные исходы двух последовательных испытаний:

$$11,\ 10,\ 01,\ 00.$$

Очевидно, что переход от одного события такого вида к другому зависит от того, каким было первое событие. Так, например, вероятность получить 00 после 11 равна нулю, а после 10 равна $1-p$. Все вероятности переходов равны p, $1-p$, 0, и поэтому соответствующая новой последовательности стохастическая матрица имеет вид

$$\begin{Vmatrix} p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \\ p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \end{Vmatrix}.$$

Следует отметить, что пары символов, которые не имеют общих элементов (иначе говоря, не имеют «зацеплений»), независимы. □

Перед тем как перейти к анализу цепей и выяснению смысла параметров, задающих цепь, приведем несколько вспомогательных формул.

Рассмотрим вероятностное пространство $(\Omega, \mathscr{A}, \mathrm{P})$.

I. *Формула полной вероятности для условных вероятностей.* Если события $A, A_1, \ldots, A_n \in \mathscr{A}$ таковы, что $A_i \cap A_j = \varnothing$ для любых $i \neq j$, $\bigcup_{k=1}^{n} A_k = \Omega$ и, кроме того, $\mathrm{P}\{A\} > 0$, то для любого $B \in \mathscr{A}$ выполнено равенство

$$\mathrm{P}\{B \mid A\} = \sum_{k=1}^{n} \mathrm{P}\{B \mid A_k \cap A\}\mathrm{P}\{A_k \mid A\}.$$

II. *Формула умножения вероятностей.* Для любого набора событий $A_1, \ldots, A_n \in \mathscr{A}$ выполнено равенство

$$\mathrm{P}\{A_1 \cap \ldots \cap A_n\} = \mathrm{P}\{A_1\}\mathrm{P}\{A_2 \mid A_1\}\mathrm{P}\{A_3 \mid A_1 \cap A_2\}\mathrm{P}\{A_n \mid A_1 \cap \ldots \cap A_{n-1}\}.$$

Эти формулы легко доказать самостоятельно.

Выясним вероятностный смысл параметров, задающих простую однородную цепь Маркова. Для случайных величин $X_0, \ldots, X_n$, связанных в цепь, вычислим вероятность $\mathrm{P}\{X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}\}$ по

совместному распределению $X_0, \ldots, X_n$. Имеем

$$\begin{aligned}\mathrm{P}\{X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}\} = \\ = \sum_{x_{i_n}} \mathrm{P}\{X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}, X_n = x_{i_n}\} = \\ = p_{i_0} p_{i_0,i_1} \cdots p_{i_{n-2},i_{n-1}} \sum_{i_n} p_{i_{n-1},i_n} = p_{i_0} p_{i_0,i_1} \cdots p_{i_{n-2},i_{n-1}},\end{aligned}$$

так как в силу свойства 2 элементов стохастической матрицы

$$\sum_{i_n} p_{i_{n-1},i_n} = 1.$$

Если n раз осуществить подобное суммирование, то в результате получим распределение вероятностей случайной величины X_0:

$$\mathrm{P}\{X_0 = x_{i_0}\} = p_{i_0}, \quad x_{i_0} \in (x_1, \ldots, x_s).$$

Определение 3. Распределение X_0, т. е. набор $(p_1, \ldots, p_s)$, где $\sum_{k=1}^{s} p_k = 1$ и $p_k = \mathrm{P}\{X_0 = x_k\}$, называется *начальным распределением вероятностей* цепи Маркова.

Далее, пользуясь определением цепи, находим для любого n условную вероятность:

$$\mathrm{P}\{X_n = x_{i_n} \mid X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}\} = \frac{p_{i_0} p_{i_0,i_1} \cdots p_{i_{n-2},i_{n-1}} p_{i_{n-1},i_n}}{p_0 p_{i_0,i_1} \cdots p_{i_{n-2},i_{n-1}}} = p_{i_{n-1},i_n}.$$

Таким образом, выписанная условная вероятность в случае, когда она определена, на самом деле зависит от значения, которое приняла случайная величина X_{n-1}, т. е. от $x_{i_{n-1}}$, и поэтому

$$p_{i_{n-1},i_n} = \mathrm{P}\{X_n = x_{i_n} \mid X_{n-1} = x_{i_{n-1}}\}.$$

Определение 4. Случайные величины $X_0, \ldots, X_n$ образуют *простую однородную цепь Маркова*, если

$$\begin{aligned}\mathrm{P}\{X_n = x_{i_n} \mid X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}\} = \\ = \mathrm{P}\{X_n = x_{i_n} \mid X_{n-1} = x_{i_{n-1}}\} = p_{i_{n-1},i_n},\end{aligned}$$

где p_{i_{n-1},i_n} — элементы стохастической матрицы П.

Определение корректно в случае, когда все условные вероятности, входящие в эти равенства, не обращаются в ноль.

Этими соотношениями разъясняется смысл элементов стохастической матрицы:

$$p_{ij} = \mathrm{P}\{X_n = x_j \mid X_{n-1} = x_i\}, \quad x_i, x_j \in \mathscr{X}.$$

Итак, двумя равносильными способами определена простая однородная цепь Маркова с конечным числом значений. Цепь Маркова называется *простой* в том смысле, что для предсказания значения случайной величины X_n достаточно знать значение предыдущей случайной величины X_{n-1}. *Однородность* цепи Маркова понимается в том смысле, что вероятность p_{ij} зависит лишь от значений x_i и x_j, но не зависит от n. Иными словами, для любых значений n и m и любых значений $x_i, x_j \in \mathscr{X}$ выполнено равенство

$$\mathrm{P}\{X_n = x_j \mid X_{n-1} = x_i\} = \mathrm{P}\{X_m = x_j \mid X_{m-1} = x_i\}.$$

Определение 5. Условные вероятности $p_{ij} = \mathrm{P}\{X_n = x_j \mid X_{n-1} = x_i\}$, образующие матрицу П, называются *вероятностями перехода за один шаг* от значения x_i к значению x_j, а сама матрица П называется *матрицей переходных вероятностей* для данной простой однородной цепи Маркова.

Замечание. Если $X_n = x_i$, то принято говорить, что в момент n некоторая система находится в состоянии x_i. Тогда вероятность $p_{ij} = \mathrm{P}\{X_n = x_j \mid X_{n-1} = x_i\}$ интерпретируется как вероятность того, что система в момент n находится в состоянии с номером j, при условии, что в предшествующий момент времени $n-1$ она находилась в состоянии с номером i.

Основные соотношения для простой однородной цепи Маркова

Пусть для цепи Маркова $X_0, X_1, \ldots, X_n$ со множеством значений $\mathscr{X} = (x_1, \ldots, x_s)$ заданы начальные вероятности $p_i(0) = \mathrm{P}\{X_0 = x_i\}$, $i = 1, \ldots, s$, и матрица переходных вероятностей $\Pi = \|p_{ij}\|$, $i, j = 1, \ldots, s$, где

$$p_{ij} = \mathrm{P}\{X_n = x_j \mid X_{n-1} = x_i\}, \quad i, j = 1, \ldots, s.$$

Введем теперь вероятности

$$p_j(n) = \mathrm{P}\{X_n = x_j\}, \quad j = 1, \ldots, s,$$

которые будем интерпретировать как вероятности того, что в момент n система находится в состоянии x_j. В совокупности эти вероятности $p_1(n), \ldots, p_s(n)$ образуют распределение вероятностей случайной величины X_n. Введем также условные вероятности $p_{ij}(m-k)$ перехода за $m-k$ $(m > k)$ шагов из состояния x_i в состояние x_j, которые в совокупности составляют матрицу $\Pi_{k,m}$ переходов за $m-k$ шагов:

$$\Pi_{k,m} = \|p_{ij}(m-k)\|, \quad p_{ij}(m-k) = \mathrm{P}\{X_m = x_j \mid X_k = x_i\}.$$

Легко проверить, что матрица $\Pi_{k,m}$ при любых k и m стохастическая: при любом фиксированном i

$$\sum_{j=1}^{s} p_{ij}(m-k) = \sum_{j=1}^{s} \mathrm{P}\{X_m = x_j \mid X_k = x_i\} = 1.$$

Покажем, что простая однородная цепь Маркова полностью определяется начальным распределением $p_1(0), \ldots, p_s(0)$ и матрицей Π переходных вероятностей за один шаг, что означает, что по этим параметрам можно вычислить вероятности любых событий, связанных с цепью, и, стало быть, вероятности $p_1(n), \ldots, p_s(n)$ и элементы матрицы $\Pi_{k,m}$ для любых n, k и m $(k < m)$.

Рассмотрим сначала две полезные леммы.

Лемма 1. *При любом* k, $1 \leqslant k \leqslant n-1$,

$$\mathrm{P}\{X_n = x_{i_n} \mid X_0 = x_{i_0}, \ldots, X_k = x_{i_k}\} = \mathrm{P}\{X_n = x_{i_n} \mid X_k = x_{i_k}\}.$$

Это свойство основано на определении и легко из него выводится. Очевидно, что при $k = n-1$ это соотношение совпадает с одним из определений цепи (см. определение 4).

Утверждение леммы означает, что вероятность перейти из состояния x_{i_k} в момент k в состояние x_{i_n} в момент n не зависит от того, что было до момента времени k. Более того, если $X_0, \ldots, X_n$ — простая однородная цепь Маркова, то при любом k, $1 \leqslant k \leqslant n-1$, имеем

$$\mathrm{P}\{X_{k-1} = x_{i_{k-1}}, X_{k+1} = x_{i_{k+1}} \mid X_k = x_{i_k}\} = \\ = \mathrm{P}\{X_{k-1} = x_{i_{k-1}} \mid X_k = x_{i_k}\} \cdot \mathrm{P}\{X_{k+1} = x_{i_{k+1}} \mid X_k = x_{i_k}\}.$$

Определение 6. Случайные величины образуют *простую однородную цепь Маркова*, если при любом значении k, $1 \leqslant k \leqslant n-1$, для всех значений $x_{i_0}, x_{i_1}, \ldots, x_{i_n} \in \mathscr{X}$, при которых все условные вероятности существуют, выполняются все равенства

$$\mathrm{P}\{X_0 = x_{i_0}, \ldots, X_{k-1} = x_{i_{k-1}}, X_{k+1} = x_{i_{k+1}}, \ldots, X_n = x_{i_n} \mid X_k = x_{i_k}\} = \\ = \mathrm{P}\{X_0 = x_{i_0}, \ldots, X_{k-1} = x_{i_{k-1}} \mid X_k = x_{i_k}\} \times \\ \times \mathrm{P}\{X_{k+1} = x_{i_{k+1}}, \ldots, X_n = x_{i_n} \mid X_k = x_{i_k}\}.$$

Иными словами, это свойство, называемое *марковским свойством* случайных последовательностей, означает, что «будущее» системы при фиксированном «настоящем» не зависит от «прошлого».

Найдем соотношение между матрицами Π и $\Pi_{k,m}$. Для этого рассмотрим три момента времени $k < l < m$.

Лемма 2. *При любых $k<l<m$ выполнено равенство*

$$\Pi_{k,m}=\Pi_{k,l}\cdot\Pi_{l,m}.$$

Доказательство. Рассмотрим все возможные переходы из состояния x_{i_k} в состояние x_{i_m} с учетом всех возможных состояний x_{i_l} в некоторый промежуточный момент времени l. По формуле полной вероятности для условных вероятностей и по лемме 1 имеем

$$\begin{aligned}\mathrm{P}\{X_m=x_{i_m}\mid X_k=x_{i_k}\}&=\\&=\sum_{i_l}\mathrm{P}\{X_m=x_{i_m}\mid X_k=x_{i_k},X_l=x_{i_l}\}\mathrm{P}\{X_l=x_{i_l}\mid X_k=x_{i_k}\}.\end{aligned}$$

Таким образом,

$$\mathrm{P}\{X_m=x_{i_m}\mid X_k=x_{i_k}\}=\sum_{i_l}\mathrm{P}\{X_m=x_{i_m}\mid X_l=x_{i_l}\}\mathrm{P}\{X_l=x_{i_l}\mid X_k=x_{i_k}\},$$

или

$$p_{i_k,i_m}(m-k)=\sum_{i_l}p_{i_k,i_m}(m-l)p_{i_k,i_l}(l-k).$$

В последнем соотношении слева стоит элемент матрицы $\Pi_{k,m}$, а справа — элемент матрицы, являющейся произведением матриц $\Pi_{l,m}$ и $\Pi_{k,l}$, следовательно, последнее соотношение представляет собой покомпонентную запись матричного равенства

$$\Pi_{k,m}=\Pi_{k,l}\cdot\Pi_{l,m}.$$

Как следствие, получаем

$$\Pi_{k,m}=\Pi^{m-k}.$$

В частности, при $n\geqslant 1$ имеем

$$\Pi_{0,n}=\Pi^n,$$

т. е. матрица перехода за n шагов является n-й степенью матрицы Π. □

Теорема (основное соотношение для простой однородной цепи Маркова с конечным числом значений). *Пусть $p_k(n)=\mathrm{P}\{X_n=x_k\}$ и $p(n)=(p_1(n),\ldots,p_s(n))$ — распределение случайной величины X_n в цепи Маркова $X_0,X_1,\ldots,X_n$, а $\Pi=\|p_{ij},\ i,j=1,\ldots,s\|$ — матрица переходных вероятностей цепи. Тогда*

$$\begin{aligned}p(m)&=p(m-1)\Pi,&&m\geqslant 1,\\p(m)&=p(k)\Pi^{m-k},&&0\leqslant k\leqslant m-1,\\p(n)&=p(0)\Pi^n,&&n\geqslant 1.\end{aligned}$$

Доказательство. По формуле полной вероятности имеем

$$\mathrm{P}\{X_m = x_j\} = \sum_{i=1}^{s} \mathrm{P}\{X_m = x_j \mid X_k = x_i\}\mathrm{P}\{X_k = x_i\}, \quad j = 1, \ldots, s,$$

$$\mathrm{P}\{X_n = x_j\} = \sum_{i=1}^{s} \mathrm{P}\{X_n = x_j \mid X_0 = x_i\}\mathrm{P}\{X_0 = x_i\}, \quad j = 1, \ldots, s,$$

и, переходя к принятым обозначениям, получаем

$$p_j(m) = \sum_{i=1}^{s} p_i(k)p_{ij}(m-k), \quad j = 1, \ldots, s,$$

$$p_j(n) = \sum_{i=1}^{s} p_i(0)p_{ij}(n), \qquad j = 1, \ldots, s,$$

где $p_{ij}(m-k)$ и $p_{ij}(n)$ — элементы матриц Π^{m-k} и Π^n соответственно. Используя лемму 2, получаем соответственно эти равенства в матричной форме:

$$p(m) = p(k)\Pi_{k,m} = p(k)\Pi^{m-k},$$
$$p(n) = p(0)\Pi_n = p(0)\Pi^n.$$

Из первой формулы при $k = m-1$ получаем, что $p(m) = p(m-1)\Pi$. Доказательство завершено. □

Рассмотрим отдельно случай, когда распределения $p(n)$ не зависят от n, а именно, $p(n) = q$ для любого значения n, где $q = (q_1, \ldots, q_s)$, $\sum_{k=1}^{s} q_k = 1$, $q_k = \mathrm{P}\{X_n = x_k\}$, $k = 1, \ldots, s$, т. е. $q = p(0) = p(1) = \ldots = p(n)$. Очевидно, что распределение q удовлетворяет системе уравнений $q = q\Pi$.

Определение 7. Распределение $q = (q_1, \ldots, q_s)$, удовлетворяющее матричному равенству $q = q\Pi$, или, что эквивалентно, системе линейных уравнений

$$q_j = \sum_{k=1}^{s} q_k p_{kj}, \quad j = 1, \ldots, s,$$

называется *стационарным* или *инвариантным* распределением цепи Маркова с данной матрицей переходных вероятностей Π.

Название определяется основным свойством этого распределения: если $p_j(0) = q_j$, $j = 1, \ldots, s$, то $p_j(1) = \sum_{i=1}^{s} p_j(0)p_{ij} = q_j$, $p_j(2) = \sum_{i=1}^{s} p_j(1)p_{ij} = q_j$ и вообще $p_j(n) = q_j$ для всех значений n, т. е. $q = q\Pi^n$.

Таким образом, если в качестве начального распределения вероятностей взять стационарное распределение, то все величины X_n в цепи Маркова будут одинаково распределены, оставаясь при этом зависимыми. В этом случае последовательность $X_1, \ldots, X_n$ называется *стационарной* цепью Маркова, и это есть простейший пример стационарной последовательности случайных величин (свойства стационарности в классических вероятностных моделях рассматриваются в гл. 2 на с. 22).

§ 3. Эргодические теоремы для цепей Маркова

В теории цепей Маркова наиболее важным является вопрос о предельном поведении вероятностей $p_j(n)$ и $p_{ij}(n)$ при $n \to \infty$ в зависимости от начального распределения.

Следующая теорема дает достаточные условия сходимости $p_j(n)$ при $n \to \infty$ к некоторому пределу. Теоремы такого типа называются *эргодическими*.

Теорема 1. *Пусть все элементы p_{ij} матрицы* Π *удовлетворяют условию*

$$\min_{1 \leqslant i,j \leqslant s} p_{ij} = \lambda > 0.$$

Тогда

1) *существуют такие неотрицательные числа $q_1, \ldots, q_s$, $\sum_{j=1}^{s} q_j = 1$, что $p_j(n) \to q_j$ при $n \to \infty$ независимо от начального распределения,*

2) *набор чисел $q_1, \ldots, q_s$ является единственным решением системы уравнений $q_j = \sum_{i=1}^{s} q_i p_{ij}$, $j = 1, \ldots, s$,*

3) $|p_j(n) - q_j| \leqslant (1 - s\lambda)^n$.

Заметим, что набор чисел q_j, $j = 1, \ldots, s$, представляет собой распределение вероятностей, определенное как решение уравнений $q_j = \sum_{i=1}^{s} q_i p_{ij}$, $j = 1, \ldots, s$. Это распределение было названо стационарным распределением вероятностей цепи Маркова с матрицей Π.

Отметим, что стационарное распределение может быть не единственным. Однако важно то, что единственное стационарное решение в условиях теоремы 1 является предельным в смысле сходимости $p_j(n) \to q_j$.

Доказательство теоремы 1. Докажем сначала, что элементы $p_{ij}(n)$ матрицы Π^n сходятся к некоторому пределу q_j при $n \to \infty$

независимо от значения i. При любом n определим

$$m_j(n) = \min_i p_{ij}(n), \quad M_j(n) = \max_i p_{ij}(n), \quad (m_j = m_j(1),\ M_j = M_j(1)).$$

Последовательность $m_j(n)$ является неубывающей, а последовательность $M_j(n)$ — невозрастающей в силу неравенств

$$m_j(n+1) = \min_i \sum_k p_{ik} p_{kj}(n) \geqslant \min_i \sum_k p_{ik} m_j(n) = m_j(n),$$
$$M_j(n+1) = \max_i \sum_k p_{ik} p_{kj}(n) \leqslant \max_i \sum_k p_{ik} M_j(n) = M_j(n).$$

Кроме того, для любых n_1 и n_2

$$m_j(n_1) \leqslant M_j(n_2).$$

Следовательно, обе последовательности сходятся. Для того чтобы показать, что последовательности сходятся к общему пределу, оценим разность $M_j(n) - m_j(n)$. Для любой фиксированной пары индексов r и t рассмотрим суммы

$$\sum_k{}' (p_{rk} - p_{tk}), \quad \sum_k{}'' (p_{rk} - p_{tk}),$$

где знаками $\sum_k{}'$ и $\sum_k{}''$ обозначено суммирование по множествам значений $\{k\colon p_{rk} \geqslant p_{tk}\}$ и $\{k\colon p_{rk} < p_{tk}\}$ соответственно. Легко показать, что

$$\sum_k{}' (p_{rk} - p_{tk}) + \sum_k{}'' (p_{rk} - p_{tk}) = 1$$

и

$$\sum_k{}' (p_{rk} - p_{tk}) \leqslant 1 - s\lambda.$$

Используя эти соотношения, получаем

$$M_j - m_j = \max_{r,t} (p_{rj} - p_{tj}) \leqslant 1 - s\lambda,$$

и при любом $n > 1$

$$M_j(n+1) - m_j(n) = \max_{r,t} \sum_k (p_{rk} - p_{tk}) p_{kj}(n) \leqslant$$
$$\leqslant \max_{r,t} \left[\sum_k{}' (p_{rk} - p_{tk}) M_j(n) + \sum_k{}'' (p_{rk} - p_{tk}) m_j(n) \right] =$$
$$= \max_{r,t} \sum_k{}' (p_{rk} - p_{tk}) (M_j(n) - m_j(n)) \leqslant (1 - s\lambda)(M_j(n) - m_j(n)).$$

Окончательно получаем

$$M_j(n) - m_j(n) \leqslant (1-s\lambda)^n.$$

Таким образом, последовательности $p_{ij}(n)$, $m_j(n)$, $M_j(n)$ имеют общий предел, который мы обозначим q_j, и при этом

$$|p_{ij}(n) - q_j| \leqslant M_j(n) - m_j(n) \leqslant (1-s\lambda)^n.$$

Легко доказать теперь утверждения 1 и 3 теоремы. Так как $p_j(n) = \sum_{i=1}^{s} p_i(0)p_{ij}(n)$, $p_{ij}(n) \to q_j$ и $\sum_{i=1}^{s} p_i(0) = 1$, получаем, что $p_j(n) \to q_i$ и при этом

$$|p_j(n) - q_j| \leqslant \sum_i p_i(0)|p_{ij}(n) - q_j| \leqslant (1-s\lambda)^n.$$

Покажем, что $q_1, \ldots, q_s$ — стационарное распределение вероятностей. Так как $q_j = \lim\limits_{n\to\infty} p_{ij}(n)$ и $\sum\limits_{j=1}^{s} p_{ij}(n) = 1$, имеем $\sum\limits_{j=1}^{s} q_j = 1$. В силу неравенств

$$\lambda = \min_{i,j} p_{ij} = \min_j(\min_i p_{ij}) = \min_j m_j \leqslant m_j \leqslant m_j(n) \leqslant p_{ij}(n)$$

очевидно, что $q_j \geqslant \lambda > 0$, $j = 1, \ldots, s$. Переходя к пределу при $n \to \infty$ в обоих частях равенства

$$p_j(n+1) = \sum_{i=1}^{s} p_i(n)p_{ij},$$

получим $q_j = \sum\limits_{i=1}^{s} q_i p_{ij}$.

Очевидно также, что распределение q_j, $j = 1, \ldots, s$, является единственным стационарным распределением. Пусть q'_j, $j = 1, \ldots, s$, — еще одно решение системы $q'_j = \sum\limits_{i=1}^{s} q'_i p_{ij}$, $j = 1, \ldots, s$. Возьмем в качестве начальных распределений оба стационарных распределения q_j и q'_j, $j = 1, \ldots, s$, т. е. $p_j(n) = \sum\limits_i q_i p_{ij}$ и $p'_j(n) = \sum\limits_i q'_i p_{ij}$. Тогда для любых n и j имеем $p_j(n) = q$, $p'_j(n) = q'$ и $p_j(n) - p'_j(n) = q_j - q'_j$. Предел левой части последнего равенства при $n \to \infty$ равен нулю, как следует из доказательства утверждения 1, поэтому $q_j = q'_j$, $j = 1, \ldots, s$. Таким образом, утверждение 2 теоремы также доказано. □

В теореме 1 условие $\min\limits_{1\leqslant i,j\leqslant s} p_{ij} = \lambda > 0$ означает, что все элементы матрицы Π положительны, т. е. все переходы из состояния в состояние осуществимы. Эргодическая теорема остается верной, если

условие теоремы 1 заменить следующим: существует такое значение $n_0 \geqslant 1$, что $\min\limits_{1\leqslant i,j\leqslant s} p_{ij}(n_0) = \lambda > 0$, или, что то же самое, все элементы матрицы Π^{n_0} положительны, т. е. все состояния достижимы из любого другого состояния за n_0 переходов.

Как показывает следующая теорема, достаточное условие сходимости $p_j(n)$ к стационарному распределению может быть еще ослаблено.

Теорема 2. *Пусть при некотором целом $n_0 \geqslant 1$ матрица Π^{n_0} такова, что существует подмножество $J = \{j_1, \ldots, j_{s_1}\}$ множества $\{1, \ldots, s\}$ значений индекса j $(1 \leqslant s_1 \leqslant s)$, для которого*

$$\min_{1\leqslant i\leqslant s,\ j\in J} p_{ij}(n_0) = \lambda > 0.$$

Тогда будут выполнены утверждения 1 *и* 2 *теоремы* 1, *а утверждение* 3 *заменится на*

3′) $|p_j(n) - q_j| \leqslant (1 - s_1\lambda)^{[n/n_0]}$,

где $[a]$ *обозначает целую часть* a.

Доказательство теоремы 2 ничем существенным не отличается от доказательства теоремы 1 (см. [8]).

Итак, смысл эргодической теоремы в том, что вероятность $\mathrm{P}\{X_n = x_j\}$ стремится к стационарной вероятности q_j при $n \to \infty$ и любом фиксированном значении $j = 1, \ldots, s$ независимо от начального распределения цепи Маркова.

Замечание (к теоремам 1 и 2)**.** Очевидно, что

$$\Pi^n \to Q = \begin{Vmatrix} q_1, & \ldots, & q_s \\ q_1, & \ldots, & q_s \\ q_1, & \ldots, & q_s \end{Vmatrix} \quad \text{при } n \to \infty.$$

Замечание. Рассмотрим цепь $X_0, X_1, \ldots, X_m, \ldots, X_{m+n}, \ldots, X_{m+n+r}$ и предположим, что все элементы матрицы переходных вероятностей положительны. Пусть событие A определяется по первым $m+1$ величинам $X_0, \ldots, X_m$, а событие B определяется по величинам $X_{m+n}, \ldots, X_{m+n+r}$, где $m > 0$, $n > 0$, $r > 0$. Если бы величины были независимы, то было бы справедливо равенство $\mathrm{P}\{A \cap B\} = \mathrm{P}\{A\} \cdot \mathrm{P}\{B\}$. В данном же случае

$$|\mathrm{P}\{A \cap B\} - \mathrm{P}\{A\}\mathrm{P}\{B\}| \leqslant C\delta^n \mathrm{P}\{A\},$$

где $\delta < 1$, а $C > 0$ — постоянная. Таким образом, зависимость между событиями A и B ослабевает с ростом n. Справедливость этого факта основана на эргодической теореме.

§ 4. Закон больших чисел для цепей Маркова

Рассмотрим последовательность случайных величин $X_0, \ldots, X_n, \ldots$, связанных в цепь Маркова.

Рассмотрим сначала случай стационарной цепи Маркова: начальное распределение цепи совпадает со стационарным, и потому все случайные величины одинаково распределены и их распределение является стационарным:

$$\mathrm{P}\{X_i = x_k\} = q_k, \quad k = 1, \ldots, s; \quad \sum_{i=1}^{s} q_k = 1.$$

Тогда математическое ожидание величины X_i равно $\mathrm{E}X_i = \sum_{i=1}^{s} x_k q_k$, дисперсия величины X_i равна $\mathrm{D}X_i = \sum_{k<j} q_k q_j (x_k - x_j)^2$, а ковариация между X_i и X_{i+n} равна

$$\operatorname{cov}(X_i, X_{i+n}) = \sum_{k,j} x_k x_j q_k (p_{kj}(n) - q_j).$$

Легко видеть, что дисперсии $\mathrm{D}X_i$ равномерно ограничены, а в выражении для ковариации каждое слагаемое суммы содержит разность, которая в соответствии с теоремой 1 стремится к нулю при $n \to \infty$. Таким образом, выполняются условия теоремы Бернштейна: $\mathrm{D}X_i \leqslant C$ и $\operatorname{cov}(X_i, X_{i+n}) \to 0$ при $n \to \infty$, и поэтому выполняется ЗБЧ: для любого $\varepsilon > 0$ при $n \to \infty$ имеем

$$\mathrm{P}\left\{\left|\frac{X_0 + \ldots + X_n}{n+1} - a\right| > \varepsilon\right\} \to 0.$$

Это означает, что среднее арифметическое $\dfrac{X_0 + \ldots + X_n}{n+1}$ случайных величин $X_0, \ldots, X_n$ сходится по вероятности к постоянной $a = \sum_{k=1}^{s} x_k q_k$:

$$\frac{X_0 + \ldots + X_n}{n+1} \xrightarrow[n\to\infty]{\mathrm{P}} a.$$

Покажем, что в общем случае условия эргодической теоремы достаточны для того, чтобы к цепи Маркова был применим ЗБЧ.

Теорема. *Пусть $X_0, \ldots, X_n$ — простая однородная цепь Маркова с конечным числом значений $(x_1, \ldots, x_s)$. Пусть матрица Π переходных вероятностей состоит из положительных элементов, и пусть $f(x)$ — такая функция на множестве $\mathscr{X} = (x_1, \ldots, x_s)$, что $|f(x)| \leqslant L$, $L > 0$. Тогда к последовательности $f(X_0), \ldots, f(X_n), \ldots$ применим ЗБЧ.*

Доказательство. Для того чтобы использовать неравенство Чебышёва, оценим

$$B_n = \mathrm{D}[f(X_0) + \ldots + f(X_n)].$$

Рассмотрим ковариацию C_m для произвольных X_k и X_{k+m}:

$$\begin{aligned} C_m &= \mathrm{E}f(X_k)f(X_{k+m}) - \mathrm{E}f(X_k)\mathrm{E}f(X_{k+m}) = \\ &= \sum_{i,j} f(x_i)f(x_j)\mathrm{P}\{X_k = x_i, X_{k+m} = x_j\} - \\ &\quad - \sum_{i,j} f(x_i)f(x_j)\mathrm{P}\{X_k = x_i\}\mathrm{P}\{X_{k+m} = x_j\} = \\ &= \sum_{i,j} f(x_i)f(x_j)[\mathrm{P}\{X_{k+m} = x_j \mid X_k = x_i\} - \mathrm{P}\{X_{k+m} = x_j\} \cdot \mathrm{P}\{X_k = x_i\}]. \end{aligned}$$

В силу эргодической теоремы 1 (по предположению $p_{ij} \geqslant \lambda > 0$) очевидно, что все разности в последнем выражении стремятся к нулю с ростом m. Используя оценку из теоремы 1, находим

$$\sum_j |\mathrm{P}\{X_{k+m} = x_j \mid X_k = x_i\} - \mathrm{P}\{X_{k+m} = x_j\}| \leqslant 2s(1-s\lambda)^m,$$

так как и $p_{ij}(m)$, и $p_{j(k+m)}$ при $m \to \infty$ стремятся к одному и тому же пределу q_j — стационарной вероятности события $\{X_n = x_j\}$. Так как функция $f(x)$ задана на конечном множестве $(x_1, \ldots, x_s)$ и

$$|f(x_j)| \leqslant L,$$

а также

$$\left|\sum_i \sum_j a_i b_j\right| \leqslant \sum_i |a_i| \sum_j |b_j|,$$

получаем, что

$$|C_m| \leqslant 2sL(1-s\lambda)^m \sum_j |f(x_j)|\mathrm{P}\{X_k = x_j\} \leqslant 2sL^2(1-s\lambda)^m.$$

Рассмотрим случайные величины

$$Y_k = f(X_k) - \mathrm{E}f(X_k).$$

По доказанному

$$|C_m| = |\mathrm{E}Y_k Y_{k+m}| \leqslant 2sL^2(1-s\lambda)^m.$$

Очевидно, что (поскольку $\mathrm{E}Y_k = 0$)

$$B_n = \mathrm{E}(Y_0 + \ldots + Y_n)^2 = \sum_{k=0}^{n} \mathrm{E}Y_k^2 + 2\sum_{k=0}^{n-1} \mathrm{E}Y_k Y_{k+1} + \ldots + 2\mathrm{E}Y_0 Y_k.$$

Используя далее оценки $|\mathrm{E}Y_kY_{k+m}|$ для $m=1, 2, \ldots, k$, получаем

$$B_n \leqslant 4(n+1)L^2[1+(1-s\lambda)+(1-s\lambda)^2+\ldots] = \frac{4(n+1)L^2}{\lambda}.$$

По неравенству Чебышёва при любом $\varepsilon > 0$

$$\mathrm{P}\left\{\left|\frac{f(X_0)+\ldots+f(X_n)}{n+1} - \frac{\mathrm{E}f(X_0)+\ldots+\mathrm{E}f(X_n)}{n+1}\right| > \varepsilon\right\} \leqslant \frac{2L^2}{\lambda}\cdot\frac{1}{(n+1)\varepsilon^2}.$$

Но правая часть неравенства при $n\to\infty$ стремится к нулю при любом фиксированном $\varepsilon > 0$ и постоянных λ и L. □

Следствие. *При $n\to\infty$ и любом $\varepsilon > 0$ выполнено соотношение*

$$\mathrm{P}\left\{\left|\frac{f(X_0)+\ldots+f(X_n)}{n+1} - \sum_j f(x_j)q_j\right| > \varepsilon\right\} \to 0.$$

Доказательство. В силу эргодической теоремы имеем

$$\mathrm{E}f(X_k) = \sum_j f(x_j)\mathrm{P}\{X_k = x_j\} \to \sum_j f(x_j)q_j \quad \text{при } k\to\infty.$$

Используем тот факт (лемма Тёплица), что если $\alpha_l > 0$ и $\alpha_l \to \alpha$ при $l\to\infty$, то

$$\frac{\alpha_1+\ldots+\alpha_n}{n} \to \alpha \quad \text{при } n\to\infty.$$

Таким образом, при $n\to\infty$

$$\frac{\mathrm{E}f(X_0)+\ldots+\mathrm{E}f(X_n)}{n+1} \to \sum_j f(x_j)q_j. \qquad \square$$

Замечание. Пусть $A = \{x_{i_1}, \ldots, x_{i_r}\}$ — какое-либо подмножество множества состояний $(x_1, \ldots, x_s)$, и пусть

$$I_s(x) = \begin{cases} 1, & x\in A,\\ 0, & x\notin A.\end{cases}$$

Тогда сумма $T_A(n+1) = I_A(X_0)+\ldots+I_A(X_n)$ случайных величин $I_A(X_0), \ldots, I_A(X_n)$ равна числу X_k, $0\leqslant k\leqslant n$, со значениями x_j из множества A. Иначе говоря, $T_A(n+1)$ можно интерпретировать как «время пребывания в множестве состояний A». Обозначим через Q_A стационарную вероятность множества состояний A:

$$Q_A = \sum_{x_j\in A} q_j.$$

Из доказанного следствия при $f(x) = I_A(x)$ получаем

$$\mathrm{P}\left\{\left|\frac{T_A(n+1)}{n+1} - Q_A\right| > \varepsilon\right\} \to 0$$

при $\varepsilon > 0$ и $n \to \infty$. Эта формула разъясняет вероятностный смысл стационарного распределения: относительное время (частота) пребывания в множестве состояний A стремится по вероятности к стационарной вероятности множества A.

Замечание (о центральной предельной теореме для цепей Маркова)**.** Рассмотрим случайные величины $X_0, X_1, \ldots, X_n$, связанные в цепь Маркова. При этом пусть X_k при любом k принимает только два значения 1 и 0, и пусть

$$\Pi = \left\| \begin{matrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{matrix} \right\|, \quad 0 < p_{00} < 1, \quad 0 < p_{10} < 1.$$

Тогда в условиях эргодической теоремы для любых α и β $(\alpha < \beta)$ при $n \to \infty$ имеем

$$\mathrm{P}\left\{ \alpha < \frac{X_0 + \ldots + X_n - (n+1)q_1}{\sqrt{nq_1(1-q_1)\frac{1+\delta}{1-\delta}}} < \beta \right\} \to \Phi(\beta) - \Phi(\alpha),$$

где $\delta = p_{00} - p_{10}$ — коэффициент корреляции между X_{n+1} и X_n, а q_1 — стационарная вероятность того, что $\{X_n = 1\}$. Это непосредственное обобщение теоремы Лапласа. Если случайные величины $X_0, \ldots, X_n$ независимы, то $\delta = 0$ и $q_1 = \mathrm{P}\{X_n = 1\}$, $1 - q_1 = \mathrm{P}\{X_n = 0\}$ (см. [7]).

§ 5. Понятие о марковском случайном процессе. Пуассоновский процесс

Определение 1. *Случайным* (стохастическим, вероятностным) процессом на вероятностном пространстве $(\Omega, \mathscr{A}, \mathrm{P})$ называется семейство случайных величин

$$X_t = X(t, \omega), \quad \omega \in \Omega,$$

зависящих от действительного параметра $t \in T$.

Параметр t обычно интерпретируется как время. Если фиксировать значение ω, то получаем функцию $X(t)$, называемую *выборочной функцией* или *траекторией* процесса. Если фиксировать t, то получим обычную случайную величину $X(\omega)$, значения которой описывают состояние некоторой системы в момент времени t (фазовое состояние).

Определение 2. Процессы, для которых T есть часть последовательности целых чисел $\ldots, -1, 0, 1, \ldots$ или вся эта последовательность, называются *процессами с дискретным временем* или *случайными последовательностями*. Процессы, для которых t меняется непрерывно, называются *процессами с непрерывным временем*.

Пример. Последовательность независимых случайных величин и последовательность величин, связанных в цепь Маркова, являются случайным процессом, для которого $T=\{0, 1, 2, \ldots\}$.

Определение 3. Случайный процесс называется *марковским*, если для любого фиксированного момента времени t_1 случайная величина $X(t)$, $t > t_1$, каково бы ни было фиксированное значение $x_1 = X(t_1)$, не зависит от $X(t_2)$, $t_2 < t_1$.

Определение 4. Марковский процесс $X(t)$ является *однородным* (по времени), если условные вероятности $p_{x,y}(t_1, t)$ перехода из состояния x в момент t_1 в состояние y в момент t зависят лишь от разности $t - t_1 > 0$ и не зависят от t_1 и t:

$$p_{x,y}(t_1, t) = p_{x,y}(t - t_1).$$

Таким образом, рассмотренная нами выше цепь Маркова — это однородный марковский процесс с дискретным временем и конечным числом возможных значений (состояний). Рассмотрим теперь пример однородного марковского процесса с непрерывным временем и счетным множеством возможных состояний, который естественным образом возникает во многих задачах, например в теории массового обслуживания.

Предположим, что начиная с момента времени $t=0$ наступают некоторые события. Будем говорить, что система находится в (фазовом) состоянии $E_k(t)$, $k=0, 1, \ldots$, в момент $t>0$, если за время $(0, t)$ наступило k событий. Обозначим через $p_{k,j}(\Delta t)$ условные вероятности перехода из состояния $E_k(t)$ в состояние $E_j(t+\Delta t)$ за время Δt и сделаем следующие предположения о поведении системы за время Δt $(\lambda > 0)$:

$$\begin{cases} p_{k,k}(\Delta t) = 1 - \lambda \Delta t + o(\Delta t), & k \geqslant 0, \\ p_{k,k+1}(\Delta t) = \lambda \Delta t + o(\Delta t), & \\ p_{k,k+l}(\Delta t) = o(\Delta t) & \forall\, l > 1, \\ p_{k,j}(\Delta t) = 0, & j < k. \end{cases}$$

Пусть $p_k(t) = \mathrm{P}\{E_k(t)\}$, причем $p_0(0) = 1$, $p_k(0) = 0$, $k > 0$ (начальные условия). По формуле полной вероятности получаем следующие разностные уравнения для вероятностей $p_k(t)$:

$$\begin{cases} p_0(t+\Delta t) = p_0(t) \cdot p_{0,0}(\Delta t) = p_0(t)(1 - \lambda \Delta t + o(\Delta t)), \\ p_k(t+\Delta t) = p_{k-1}(t) p_{k-1,k}(\Delta t) + p_k(t) p_{k,k}(\Delta t) = \\ \qquad = p_{k-1}(t)(\lambda \Delta t + o(\Delta t)) + p_k(t)(1 - \lambda \Delta t + o(\Delta t)), \quad k \geqslant 1. \end{cases}$$

Далее, имеем

$$\begin{cases} \dfrac{p_0(t+\Delta t)-p_0(t)}{\Delta t} = -\lambda p_0(t) + \dfrac{o(\Delta t)}{\Delta t}, \\ \dfrac{p_k(t+\Delta t)-p_k(t)}{\Delta t} = -\lambda p_k(t) + \lambda p_{k-1}(t) + \dfrac{o(\Delta t)}{\Delta t}, \quad k \geqslant 1. \end{cases}$$

Перейдем к пределу при $\Delta t \to 0$[1]. Получим систему дифференциальных уравнений при $t \in [0, +\infty)$:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t), \\ \dfrac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t), \quad k \geqslant 1. \end{cases}$$

Решим эту систему с помощью производящих функций $\varphi(z,t)$, определенных по крайней мере для $|z|<1$:

$$\varphi(z,t) = \sum_{k=0}^{\infty} p_k(t) z^k, \quad p_0(0) = 1, \quad p_k(0) = 0.$$

Вычислим

$$\begin{aligned} \frac{d\varphi(z,t)}{dt} &= \sum_{k=0}^{\infty} \frac{dp_k(t)}{dt} z^k = \frac{dp_0(t)}{dt} + \sum_{k=1}^{\infty} \frac{dp_k(t)}{dt} z^k = \\ &= \sum_{k=0}^{\infty} (-\lambda p_0(t)) z^k + z \sum_{k=0}^{\infty} \lambda p_k(t) z^k = \\ &= -\lambda \varphi(z,t) + \lambda z \varphi(z,t) = -\lambda(1-z)\varphi(z,t). \end{aligned}$$

Отсюда, интегрируя, получаем

$$\varphi(z,t) = C(z) e^{-\lambda(1-z)t}.$$

Находим $C(z)$ из начальных условий:

$$1 - p_0(0) = \varphi(z,0) = C(z).$$

Таким образом,

$$\varphi(z,t) = e^{-\lambda t} e^{\lambda z t} = \sum_{k=0}^{\infty} p_k(t) z^k.$$

Так как

$$e^{\lambda z t} = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} z^k,$$

получаем

$$p_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

[1] Здесь, вообще говоря, правая производная, но она совпадает с двусторонней.

что при фиксированном $t \in [0, +\infty)$ представляет собой распределение Пуассона с параметром λt. Если $X(t)$ — число случайных событий, наступивших за время Δt, то соответствующий случайный процесс $\{X(t), t \geqslant 0\}$, для которого

$$\mathrm{P}\{X(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

называется *пуассоновским процессом с параметром* $\lambda > 0$. Также говорят о пуассоновском «потоке» случайных событий (имеются в виду события E_k, $k = 0, 1, 2, \ldots$).

Замечание о показательном распределении в связи с пуассоновским процессом.

Определение 5. Случайная величина X имеет *показательное распределение* с параметром α, если оно задано плотностью

$$p(x) = \begin{cases} \alpha e^{-\alpha x}, & x > 0,\ \alpha > 0, \\ 0, & x \leqslant 0, \end{cases}$$

или функцией распределения

$$\mathscr{F}(x) = \mathrm{P}\{X < x\} = \begin{cases} 1 - e^{-\alpha x}, & x > 0, \\ 0, & x \leqslant 0. \end{cases}$$

Математическое ожидание и дисперсия величины X равны

$$\mathrm{E}X = \frac{1}{\alpha}, \quad \mathrm{D}X = \frac{1}{\alpha^2}.$$

Остановимся на важнейшем характеристическом свойстве показательного распределения, которое называется свойством *отсутствия последействия*. Вычислим при фиксированном t условную вероятность

$$\mathrm{P}\{t \leqslant X < t + \Delta t \mid X \geqslant t\} = \frac{\mathrm{P}\{t \leqslant X < t + \Delta t\}}{\mathrm{P}\{X \geqslant t\}} = \frac{\mathscr{F}(t + \Delta t) - \mathscr{F}(t)}{1 - \mathscr{F}(t)} =$$
$$= \frac{e^{-\alpha t} - e^{-\alpha(t+\Delta t)}}{e^{-\alpha t}} = 1 - e^{-\alpha \Delta t} = \lambda \Delta t + o(\Delta t).$$

Очевидно, что эта вероятность не зависит от t. Можно доказать, что показательное распределение — это единственное непрерывное распределение, обладающее свойством отсутствия последействия, или, более общим образом, марковским свойством (см., например, [11, т. 1, гл. 17, § 6; т. 2, гл. 1]). Указанным характеристическим свойством в значительной мере объясняется роль, которую показательное распределение играет в задачах теории массового обслуживания.

§ 6. Пример системы обслуживания с ожиданием

Предположим, что на некоторую систему обслуживания с одной линией обслуживания (или одним обслуживающим прибором) поступает поток вызовов или требований и при этом:

1) вызовы начинают поступать в начальный момент времени $t=0$;

2) если система свободна, то поступающий вызов обслуживается;

3) если система занята обслуживанием какого-либо вызова, то другие вызовы ожидают своей очереди;

4) поток вызовов соответствует пуассоновскому процессу с параметром $\lambda>0$;

5) случайные времена обслуживания отдельных вызовов независимы и имеют одинаковое показательное распределение с параметром $\mu>0$ (в этом случае условная вероятность того, что обслуживание продлится еще время Δt, если оно не окончилось до момента времени t, не зависит от t).

Будем изучать распределение числа вызовов, поступивших в систему к моменту времени t (распределение длины очереди). Для того чтобы воспользоваться марковским свойством процесса, выберем следующие фазовые состояния системы: будем считать, что система находится в состоянии $E_k(t)$, $k=0,1,2,\ldots$, в момент времени t, если к этому моменту времени в систему обслуживания поступило k вызовов. Пользуясь тем, что поток вызовов пуассоновский, а время обслуживания показательное, определим условные вероятности перехода из состояния в состояние за время Δt следующим образом: если в момент времени t система находится в состоянии E_0, то

$$\begin{cases} \mathrm{P}\{E_1(t+\Delta t)\mid E_0(t)\}=\lambda\Delta t+o(\Delta t),\\ \mathrm{P}\{E_0(t+\Delta t)\mid E_0(t)\}=1-\lambda\Delta t+o(\Delta t),\\ \mathrm{P}\{E_k(t+\Delta t)\mid E_0(t)\}=o(\Delta t),\quad k>1; \end{cases}$$

если в момент t система в состоянии $E_k(t)$, $k>0$, то

$$\begin{cases} \mathrm{P}\{E_{k+1}(t+\Delta t)\mid E_k(t)\}=\lambda\Delta t+o(\Delta t),\\ \mathrm{P}\{E_{k-1}(t+\Delta t)\mid E_k(t)\}=\mu\Delta t+o(\Delta t),\\ \mathrm{P}\{E_{k+l}(t+\Delta t)\mid E_k(t)\}=o(\Delta t),\quad l>1,\\ \mathrm{P}\{E_k(t+\Delta t)\mid E_k(t)\}=1-\lambda\Delta t-\mu\Delta t+o(\Delta t). \end{cases}$$

Таким образом, изменения в системе описываются марковским процессом со счетным множеством состояний $E_0, E_1, \ldots$ Пусть $p_k(t)$ —

вероятность того, что система находится в состоянии $E_k(t)$ в момент времени t. Составим по формуле полной вероятности следующие разностные соотношения:

$$\begin{cases} p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\mu\Delta t + o(\Delta t), \\ p_k(t+\Delta t) = p_{k-1}(t)\lambda\Delta t + p_k(t)(1-\lambda\Delta t - \mu\Delta t) + \\ \qquad\qquad + p_{k+1}(t)\mu\Delta t + o(\Delta t), \quad k \geqslant 1. \end{cases}$$

Имеем

$$\begin{cases} \dfrac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t) + \dfrac{o(\Delta t)}{\Delta t}, \\ \dfrac{p_k(t+\Delta t) - p_k(t)}{\Delta t} = \lambda p_{k-1}(t) - (\lambda+\mu) p_k(t) + \mu p_{k+1}(t) + \dfrac{o(\Delta t)}{\Delta t}. \end{cases}$$

При $\Delta t \to 0$ переходим к системе дифференциальных уравнений для $p_k(t)$:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t), \\ \dfrac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda+\mu) p_k(t) + \mu p_{k+1}(t), \quad k \geqslant 1. \end{cases}$$

Эту систему можно решить с помощью производящих функций, но на этом пути возникают осложнения, и сам результат имеет слишком громоздкий для исследования вид.

Нас интересует поведение системы обслуживания, работающей в стационарном режиме, и мы значительно упростим исследование, если предположим, что существует стационарное (не зависящее от t) решение p_k системы уравнений, удовлетворяющее условию $p_k(t) \to p_k$ при $t \to \infty$, и найдем его непосредственным вычислением, не затрагивая проблемы его существования и сходимости к нему.

Пусть $p_0 = \text{const}$. Из того, что $\dfrac{dp_0}{dt} = 0$, следует, что $\lambda p_0 = \mu p_1$. Отсюда получаем, что $p_1 = \dfrac{\lambda}{\mu} p_0$. Далее, для любого $k > 1$ имеем $p_k = \left(\dfrac{\lambda}{\mu}\right)^k p_0$. Для определения p_0 воспользуемся условием

$$\sum_{k=0}^{\infty} p_k = 1.$$

Если $\dfrac{\lambda}{\mu} < 1$, то $p_0 = 1 - \dfrac{\lambda}{\mu}$, и, таким образом,

$$p_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k, \quad k = 0, 1, \ldots,$$

что представляет собой геометрическое распределение с параметром $\frac{\lambda}{\mu}$. Если же $\frac{\lambda}{\mu} \geqslant 1$, то в этом случае или все p_k обращаются в нуль, или ряд $\sum\limits_{k=0}^{\infty} p_k$ расходится, и, следовательно, стационарного распределения не существует.

Замечание 1. Еще раз отметим, что наш вывод не является строго доказанным, так как существование стационарного решения здесь не обосновано.

Замечание 2. Ясно, что средний промежуток между соседними вызовами равен $\frac{1}{\lambda}$, а среднее время обслуживания одного вызова равно $\frac{1}{\mu}$. Условие, при котором существует стационарное распределение длины очереди, соответствует тому, что среднее время обслуживания меньше среднего промежутка времени между ближайшими вызовами: $\frac{1}{\lambda} > \frac{1}{\mu}$. Средняя длина очереди равна

$$\sum_{k=1}^{\infty} (k-1) p_k = \frac{\left(\frac{\lambda}{\mu}\right)^2}{1 - \frac{\lambda}{\mu}}.$$

Очевидно, если λ сближается с μ, то средняя длина очереди неограниченно возрастает.

Замечание 3. Если число линий обслуживания в системе больше 1, то для стационарного распределения длины очереди имеют место аналогичные формулы (см. [11, т. 1, гл. 17, § 7]).

§ 7. Пример системы обслуживания с отказами

Рассмотрим систему с одной линией обслуживания при следующих допущениях:

1) поток вызовов пуассоновский с параметром $\lambda > 0$;

2) времена обслуживания отдельных вызовов независимы и имеют одинаковое показательное распределение с параметром $\mu > 0$;

3) вызовы, приходящие в тот момент, когда линия свободна, начинают обслуживаться, а когда линия занята обслуживанием какого-либо вызова, получают отказ и теряются.

Нас будет интересовать вероятность того, что линия занята, или вероятность отказа. Для того чтобы процесс был марковским, введем два фазовых состояния системы: $E_0(t)$ — линия свободна в момент t, $E_1(t)$ — линия занята в момент t. Соответствующие вероятности обозначим через $p_0(t)$ и $p_1(t)$. Тогда в силу допущений 1—3

можем найти условные вероятности перехода из состояния в состояние за время Δt:

$$\begin{cases} \mathrm{P}\{E_1(t+\Delta t) \mid E_0(t)\} = \lambda\Delta t + o(\Delta t), \\ \mathrm{P}\{E_0(t+\Delta t) \mid E_0(t)\} = 1 - \lambda\Delta t + o(\Delta t), \\ \mathrm{P}\{E_0(t+\Delta t) \mid E_1(t)\} = \mu\Delta t + o(\Delta t), \\ \mathrm{P}\{E_1(t+\Delta t) \mid E_1(t)\} = 1 - \mu\Delta t + o(\Delta t). \end{cases}$$

По формуле полной вероятности получаем

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\mu\Delta t + o(\Delta t)$$

и

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t) + \frac{o(\Delta t)}{\Delta t}.$$

При $\Delta t \to 0$ переходим к дифференциальному уравнению

$$\frac{dp_0}{dt} = -\lambda p_0(t) + \mu p_1(t) = -(\lambda+\mu)p_0(t) + \mu,$$

используя тот факт, что $p_1(t) = 1 - p_0(t)$. Полученное неоднородное уравнение имеет решение, составленное из частного решения $\dfrac{\mu}{\lambda+\mu}$ и общего решения соответствующего однородного уравнения

$$Ce^{-(\lambda+\mu)t}, \quad C > 0.$$

Таким образом, общее решение дается формулами

$$p_0(t) = \frac{\mu}{\lambda+\mu} + Ce^{-(\lambda+\mu)t} \quad \text{и} \quad p_1(t) = \frac{\lambda}{\lambda+\mu} - Ce^{-(\lambda+\mu)t}.$$

Постоянная C определяется начальными условиями. Если допустить, что в начальный момент времени линия свободна, т. е. $p_0(0) = 1$, то $C = \dfrac{\lambda}{\lambda+\mu}$. Однако независимо от значения C при $t \to \infty$ имеем

$$p_0(t) \to \frac{\mu}{\lambda+\mu}, \quad p_1(t) \to \frac{\lambda}{\lambda+\mu}.$$

Следовательно, стационарные вероятности того, что линия свободна или занята, равны соответственно

$$p_0 = \frac{1}{1+\dfrac{\lambda}{\mu}} \quad \text{и} \quad p_1 = \frac{\dfrac{\lambda}{\mu}}{1+\dfrac{\lambda}{\mu}},$$

где отношение $\dfrac{\lambda}{\mu}$ интерпретируется как среднее число вызовов, поступивших в систему за время обслуживания одного вызова.

Предположим теперь, что в системе обслуживания s линий, и при этом условимся, что если хотя бы одна линия свободна, то вызов начинает обслуживаться, если же все линии заняты, то вызов получает отказ. Все остальные предположения остаются теми же, что и для одной линии. Вводим фазовые состояния системы: $E_k(t)$ — в момент времени t занято ровно k линий, $0 \leqslant k \leqslant s$, в частности, $E_s(t)$ — система занята и $E_0(t)$ — система свободна. Задаваясь условными вероятностями переходов за время Δt:

$$\begin{cases} \mathrm{P}\{E_0(t+\Delta t) \mid E_0(t)\} = 1-\lambda\Delta t+o(\Delta t), \\ \mathrm{P}\{E_{k+1}(t+\Delta t) \mid E_k(t)\} = \lambda\Delta t+o(\Delta t), \quad 0 \leqslant k < s, \\ \mathrm{P}\{E_{k-1}(t+\Delta t) \mid E_k(t)\} = k\mu\Delta t+o(\Delta t), \quad 0 < k \leqslant s, \\ \mathrm{P}\{E_k(t+\Delta t) \mid E_k(t)\} = 1-\lambda\Delta t-k\mu\Delta t+o(\Delta t), \\ \mathrm{P}\{E_s(t+\Delta t) \mid E_s(t)\} = 1-s\mu\Delta t+o(\Delta t), \end{cases}$$

получаем марковский процесс, описывающий все изменения в системе. Далее стандартным путем приходим к системе дифференциальных уравнений для определения вероятностей $p_k(t)$:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t)+\mu p_1(t), \\ \dfrac{dp_k(t)}{dt} = -(\lambda+\mu)p_k(t)p_{k-1}(t)+\mu(k+1)p_{k+1}(t), \quad 0 < k < s, \\ \dfrac{dp_s(t)}{dt} = -\mu s p_s(t)+\lambda p_{s-1}(t). \end{cases}$$

Снова будем искать решение этой системы, не зависящее от t, полагая производные равными нулю. На этом пути получаем формулы для стационарных вероятностей:

$$p_k = \frac{\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k}{\sum\limits_{k=0}^{s}\left(\frac{\lambda}{\mu}\right)^k\frac{1}{k!}}, \quad k = 0, \ldots, s.$$

Эти формулы называются *формулами Эрланга*.

Нас интересует стационарная вероятность p_s занятости системы. Выясним ее вероятностный смысл.

Рассмотрим систему обслуживания с отказами в промежутке времени $(0, T)$. Пусть случайная величина $T_{\text{зан}}$ — это суммарное время, в течение которого система занята. Тогда имеет место утверждение, обобщающее следствие из закона больших чисел для цепей

Маркова (см. § 4 на с. 96):

$$\mathrm{P}\left\{\left|\frac{T_{\text{зан}}}{T} - p_s\right| > \varepsilon\right\} \to 0, \quad T \to \infty.$$

Итак, стационарная вероятность p_s указывает относительное время занятости системы при большом времени наблюдения.

Замечание (о формулах Эрланга)**.** Формулы Эрланга остаются справедливыми и при более общих предположениях.

1. Поток вызовов пуассоновский с параметром λ, а времена обслуживания независимы и имеют одно и то же непрерывное распределение вероятностей с плотностью $p(x)$, причем

$$\mu = \frac{1}{\int\limits_0^\infty xp(x)\,dx}.$$

2. Промежутки между последовательными вызовами — независимые и одинаково распределенные случайные величины с плотностью $p(x) > 0$, причем

$$\lambda = \frac{1}{\int\limits_0^\infty xp(x)\,dx},$$

а времена обслуживания независимы и имеют одинаковое показательное распределение с параметром μ.

Подсчет доли потерянных вызовов

Рассмотрим более подробно случай, когда промежутки между вызовами — независимые случайные величины $X_1, X_2, \ldots$ с плотностью распределения $p(x)$ и математическим ожиданием

$$\frac{1}{\lambda} = \int\limits_0^\infty xp(x)\,dx,$$

а времена обслуживания $Y_1, Y_2, \ldots$ независимы и имеют показательное распределение с параметром μ. Обозначим $S_0 = 0$, $S_n = X_1 + \ldots + X_n$. Пусть ν — такой номер, что $S_\nu < Y_1$, $S_{\nu+1} \geqslant Y_1$, т. е. ν — номер последнего вызова, который пришел в течение времени обслуживания первого вызова Y_1. Найдем математическое ожидание случайной величины ν:

$$\mathrm{E}\nu = \sum_{N=0}^{\infty} N\mathrm{P}\{\nu = N\} = \sum_{N=1}^{\infty} \mathrm{P}\{\nu \geqslant N\} = \sum_{N=1}^{\infty} \mathrm{P}\{S_N < Y_1\},$$

так как события $\{\nu \geqslant N\}$ и $\{S_N < Y_1\}$ равносильны. Далее будем использовать тот факт, что Y_1 имеет показательное распределение, а слу-

чайная величина S_N как сумма независимых непрерывно распределенных величин сама непрерывно распределена. Обозначим плотность распределения величины S_N через $p_{S_N}(x)$. Найдем вероятность $\mathrm{P}\{S_N < Y_1\}$ с помощью формулы свертки — аналога формулы свертки для плотностей (см. с. 53 и, например, [11, т. 2, гл. 5, § 4]): в силу независимости S_N и Y_1 имеем

$$\mathrm{P}\{S_N - Y_1 < z\} = \int_0^\infty \mathscr{F}_{S_N}(z+y)\, d\mathscr{F}_{Y_1}(y),$$

где $\mathscr{F}_{Y_1}(y) = \mathrm{P}\{Y_1 < y\}$ и $\mathscr{F}_{S_N}(y) = \mathrm{P}\{S_N < y\}$ — функции распределения случайных величин Y_1 и S_N соответственно. Интегрируя по частям и полагая $z = 0$, получим

$$\mathrm{P}\{S_n - Y_1 < 0\} = \int_0^\infty (1 - \mathscr{F}_{Y_1}(y))\, d\mathscr{F}_{S_N}(y),$$

а так как

$$\mathscr{F}_{Y_1}(y) = 1 - e^{-\mu y}, \quad \mathscr{F}_{S_N}(y) = \int_0^y p_{S_N}(x)\, dx,$$

имеем

$$\mathrm{P}\{S_n < Y_1\} = \int_0^\infty e^{-\mu y} p_{S_N}(y)\, dy.$$

Последний интеграл представляет собой математическое ожидание случайной величины $e^{-\mu S_N}$: $\mathrm{P}\{S_n < Y_1\} = \mathrm{E}e^{-\mu S_N}$. Далее, в силу независимости $X_1, X_2, \ldots$ окончательно получаем

$$\mathrm{E}\nu = \sum_{N=1}^\infty (\mathrm{E}e^{-\mu X_1})^N = \frac{\mathrm{E}e^{-\mu X_1}}{1 - \mathrm{E}e^{-\mu X_1}}.$$

Очевидно, что $\mathrm{E}\nu$ — это среднее число вызовов, потерянных за время обслуживания одного вызова. Так как среднее число вызовов равно $1 + \mathrm{E}\nu$, отношение $\dfrac{\mathrm{E}\nu}{1+\mathrm{E}\nu}$ равно $\mathrm{E}e^{-\mu X_1}$.

Заметим, что величина $\mathrm{E}e^{-\mu X_1}$ выражает долю потерянных вызовов за любое время работы системы обслуживания. Рассмотрим большое число периодов занятости n, и пусть $\nu_1, \ldots, \nu_n$ — числа вызовов, потерянных за 1-й, ..., n-й периоды занятости соответственно. Тогда доля вызовов, потерянных за все время работы, равна

$$\frac{\nu_1 + \ldots + \nu_n}{(1+\nu_1) + \ldots + (1+\nu_n)} = \frac{\dfrac{\nu_1 + \ldots + \nu_n}{n}}{\dfrac{(1+\nu_1) + \ldots + (1+\nu_n)}{n}}.$$

При $n \to \infty$ числитель последней дроби сходится по вероятности к $\mathrm{E}\nu_1$, а знаменатель сходится по вероятности к $1+\mathrm{E}\nu_1$. Поэтому[1]

$$\frac{\nu_1+\ldots+\nu_n}{(1+\nu_1)+\ldots+(1+\nu_n)} \xrightarrow{\mathrm{P}} \frac{\mathrm{E}\nu}{1+\mathrm{E}\nu} = \mathrm{E}e^{-\mu X_1} \quad \text{при } n \to \infty.$$

Рассмотрим два примера.

Пример 1. Если поток вызовов пуассоновский с параметром λ, то случайная величина X_1 имеет показательное распределение с параметром $\frac{1}{\lambda} = \mathrm{E}X_1$ и поэтому

$$\mathrm{E}e^{-\mu X_1} = \int_0^\infty e^{-\mu x}\lambda e^{-\lambda x}\,dx = \frac{\lambda}{\lambda+\mu} = \frac{\frac{\lambda}{\mu}}{1+\frac{\lambda}{\mu}}.$$

Таким образом, в этом случае доля потерянных вызовов определяется стационарной вероятностью занятости системы, т. е. относительным временем занятости.

Пример 2. Если поток вызовов *регулярный*, т. е. вызовы поступают через равные промежутки времени $\frac{1}{\lambda}$, то $\mathrm{P}\left\{X_1 = \frac{1}{\lambda}\right\} = 1$. Так как $\mathrm{E}X_1 = \frac{1}{\lambda}$, доля потерянных вызовов равна

$$\mathrm{E}e^{-\mu X_1} = e^{-\mu/\lambda}.$$

В силу неравенства $e^z \geqslant 1+z$ при $z \geqslant 0$ имеем

$$e^{\mu/\lambda} \geqslant 1+\frac{\mu}{\lambda},$$

откуда следует, что

$$e^{-\mu/\lambda} \leqslant \frac{1}{1+\frac{\mu}{\lambda}} = \frac{\frac{\lambda}{\mu}}{1+\frac{\lambda}{\mu}}.$$

Таким образом, регулярный поток вызовов дает меньшую долю потерянных вызовов, чем пуассоновский поток. Можно показать, что регулярный поток вообще является наилучшим в рассматриваемом смысле:

$$\inf \mathrm{E}e^{-\mu X_1} = e^{-\mu/\lambda},$$

где инфинум берется по всем непрерывным распределениям с математическим ожиданием $\frac{1}{\lambda}$.

[1] По очевидному свойству: если $U_n \xrightarrow{\mathrm{P}} U$ и $V_n \xrightarrow{\mathrm{P}} V$, $V \neq 0$, при $n \to \infty$, то $\frac{U_n}{V_n} \xrightarrow{\mathrm{P}} \frac{U}{V}$.

Глава 5

Элементы математической статистики

§1. Постановка задач в математической статистике

Математическая статистика — раздел математики, посвященный математическим методам обработки и интерпретации статистических данных. Статистические методы исследования основаны на теории вероятностей, поскольку исследование генеральной совокупности совершается по схеме случайного выбора, обладающего свойствами репрезентативности и беспристрастности. Наиболее отчетливо связь между теорией вероятностей и математической статистикой проявляется при статистическом исследовании явлений, которые подчинены вероятностным закономерностям. Мы уже знаем, что вероятностные закономерности получают статистическое выражение в силу закона больших чисел. В конце гл. 1 приводился пример соотношения между задачами теории вероятностей и математической статистики как прямой и обратной задач. Остановимся на двух важнейших разделах математической статистики: теории статистических оценок и теории проверки статистических гипотез, не выходя, как правило, за пределы тех задач, которые уже появлялись в части курса, посвященной собственно теории вероятностей.

Рассмотрим в качестве пространства элементарных событий Ω евклидово пространство R^n, $\omega = (x_1, \ldots, x_n)$, или его подмножество, и пусть $\mathscr{A} = \{A, A \subset \Omega\}$ — класс событий, т. е. подмножеств пространства Ω, для которых определены вероятности $\mathrm{P}\{A\}$. Распределение вероятностей $\mathrm{P}\{A\}$ задается в дискретном случае как

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

где $p(\omega) \geqslant 0$, $\omega \in \Omega$, $\sum\limits_{\omega} p(\omega) = 1$, и в непрерывном случае — как

$$\mathrm{P}\{A\} = \int\limits_{A} p(\omega)\, d\omega,$$

где $p(\omega)$ — плотность распределения, $p(\omega) \geqslant 0$ и $\int\limits_{\Omega} p(\omega)\, d\omega = 1$.

Вероятностное пространство $(\Omega, \mathscr{A}, \mathrm{P})$ служит вероятностной моделью эксперимента со случайными исходами и позволяет вычислять вероятности всех событий, связанных с данным экспериментом.

Статистической моделью называют пространство $(\Omega, \mathscr{A}, \mathscr{P})$, где $\mathscr{P}$ — целое семейство распределений, принадлежащих к одному типу, или, иначе, связанных каким-либо одним свойством или закономерностью. В статистической литературе пространство Ω чаще всего называют *выборочным пространством*. Каждый элемент выборочного пространства представляет собой конечную совокупность результатов наблюдений или измерений, или, более общим образом, эмпирических или статистических данных. Обычно эти данные в статистической модели называют *выборочными* значениями, а их совокупность называют *выборкой*. Количество этих данных называют *размером* или *объемом* выборки.

Наиболее часто в математической статистике используется следующая *параметрическая* модель: имеется выборочное пространство Ω, а семейство распределений вероятностей $\mathrm{P}_\theta\{A\}$ зависит от параметра $\theta \subset \mathrm{R}^m$. Параметр θ принадлежит некоторому фиксированному параметрическому множеству $\Theta \subset \mathrm{R}^m$. Например, для семейства нормальных распределений $\theta = (a, \sigma^2)$, где a — математическое ожидание, а σ^2 — дисперсия, $\Theta = \{-\infty < a < \infty,\ \sigma^2 > 0\}$. Вид распределения $\mathrm{P}_\theta\{A\}$ предполагается известным, но заранее неизвестно, какое значение имеет параметр θ, и это нужно выяснить по результатам наблюдений, вернее, нужно решить, какое значение приписать параметру θ.

Каждому возможному результату наблюдения (элементу выборочного пространства) ω ставится в соответствие некоторое решение $d(\omega)$ о значении параметра θ (*статистическое решение*), и вводится функция $L[d(\omega), \theta]$, называемая *функцией потерь*, которой измеряется качество решения (здесь θ — истинное значение параметра). Предполагается, что функция потерь — неотрицательная и выпуклая функция. Ввиду того что статистическое решение $d(\omega)$ связано с результатом наблюдения, функция потерь является случайной величиной, и имеет смысл говорить о функции риска, или о среднем риске $R(d; \theta)$:

$$R(d, \theta) = \mathrm{E}_\theta L[d(\omega), \theta]$$

(знак θ у математического ожидания указывает на то, что оно вычисляется по распределению с истинным значением параметра θ).

Так как риск ошибки, с которым связано принятие решения о значении θ, должен быть невелик, естественно поставить цель — выбрать из допустимого множества статистических решений то решение, которое дает минимальный средний риск. Так как функция риска зависит от неизвестного значения параметра, следует искать такое решение, которое обладает наименьшим средним риском для всех значений θ из Θ. Нужно иметь в виду, что наилучшее в этом смысле статистическое решение не всегда существует.

Дальнейшие статистические процедуры различаются по типу статистического решения в зависимости от конкретной цели: выбор единственного значения θ (точечная статистическая оценка, см. § 1, 2), интервал значений θ (интервальная статистическая оценка, см. § 5), выбор θ из двух непересекающихся множеств значений (проверка статистической гипотезы, см. § 6). В иной, непараметрической постановке рассматриваются задачи оценки и проверки гипотез о типе распределения (см. [9, гл. 25, 30]).

§ 2. Статистическая оценка параметра биномиального распределения

Рассмотрим в качестве иллюстрации конкретную задачу оценки параметра биномиального распределения. Пусть в схеме испытаний Бернулли — независимых испытаний с двумя исходами — вероятность «удачи» p, $0 < p < 1$, неизвестна, но известно значение случайной величины S_n — числа «удач» в n испытаниях. Как лучше оценить p по значению S_n? Пусть точками выборочного пространства $\Omega = \{\omega\}$ являются возможные значения $0, 1, 2, \ldots, n$ числа «удач». Роль параметра θ играет p, а параметрического множества Θ — интервал $(0, 1)$. На Ω задано семейство биномиальных распределений

$$\mathrm{P}_p(m) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \ldots, n.$$

Оценивание параметра p равносильно тому, что каждому результату наблюдения ω сопоставляется некоторое значение функции $d(\omega)$, которую и будем называть оценкой параметра p. Таким образом, статистическая оценка параметра p — это числовая функция, определенная на множестве $(0, 1, 2, \ldots, n)$. Например, в качестве оценки можно взять частоту «удачи»

$$d^*(\omega) = \frac{S_n(\omega)}{n}.$$

Покажем, что в пределах некоторых условий частота $d^*(\omega)$ — наилучшая из допустимых оценок параметра p. Рассмотрим так называемую квадратичную функцию потерь

$$L(d(\omega), p) = (d(\omega) - p)^2.$$

Средний риск в этом случае равен математическому ожиданию квадрата отклонения оценки от истинного значения параметра:

$$R(d, p) = \mathrm{E}_p[d(\omega) - p]^2.$$

Например, для $d^*(\omega)$ имеем

$$R(d^*, p) = \mathrm{E}_p\Big[\frac{S_n}{n} - p\Big]^2 = \frac{p(1-p)}{n}.$$

Докажем, что при некотором дополнительном предположении о множестве допустимых оценок значение среднего риска $R(d^*p)$ минимально.

Определение. Оценка $d(\omega)$ параметра θ называется *несмещенной*, если ее математическое ожидание равно истинному значению параметра:

$$\mathrm{E}_\theta d(\omega) = \theta, \quad \text{или} \quad \mathrm{E}_p(d(\omega) - \theta) = 0.$$

Очевидно, что $\mathrm{E}_p d^*(\omega) = p$. Разность $d(\omega) - \theta$ называют *смещением* оценки $d(\omega)$.

Теорема. *Среди всех несмещенных оценок параметра p частота $d^*(\omega)$ является наилучшей оценкой в смысле минимальности среднего квадратичного риска.*

Доказательство. Рассмотрим произвольную несмещенную оценку $d(\omega)$ со значением $d(m)$, $m = 0, 1, \ldots, k$. При любом p из $(0, 1)$ справедливы следующие два соотношения:

$$\sum_{m=0}^{n} \mathrm{P}_p(m) \equiv 1, \qquad \sum_{m=0}^{n} d(m)\mathrm{P}_p(m) \equiv p$$

(второе тождество — это условие несмещенности). Продифференцируем оба тождества по параметру p и используем тот факт, что

$$\frac{\partial \mathrm{P}_p(m)}{\partial p} = \frac{\partial \ln \mathrm{P}_p(m)}{\partial p} \cdot \mathrm{P}_p(m).$$

Получим

$$\sum_{m=0}^{n} \frac{\partial \ln \mathrm{P}_p(m)}{\partial p} \cdot \mathrm{P}_p(m) \equiv 0, \qquad \sum_{m=0}^{n} d(m) \frac{\partial \ln \mathrm{P}_p(m)}{\partial p} \cdot \mathrm{P}_p(m) \equiv 1.$$

Умножим первое из полученных тождеств на p и вычтем из второго:

$$\sum_{m=0}^{n}[d(m)-p]\frac{\partial \ln \mathrm{P}_p(m)}{\partial p}\cdot \mathrm{P}_p(m)\equiv 1.$$

Воспользуемся далее неравенством Коши—Буняковского

$$\left(\sum_{m=0}^{n} a_m b_m\right)^2 \leqslant \left(\sum_{m=0}^{n} a_m^2\right)\left(\sum_{m=0}^{n} b_m^2\right)$$

с $a_m=\dfrac{\partial \ln \mathrm{P}_p(m)}{\partial p}\sqrt{\mathrm{P}_p(m)}$ и $b_m=[d(m)-p]\sqrt{\mathrm{P}_p(m)}$. Тогда получим

$$1\equiv\left(\sum_m [d(m)-p]\frac{\partial \ln \mathrm{P}_p(m)}{\partial p}\mathrm{P}_p(m)\right)^2\leqslant$$

$$\leqslant\left\{\sum_m [d(m)-p]^2\mathrm{P}_p(m)\right\}\left\{\sum_m\Big(\frac{\partial \ln \mathrm{P}_p(m)}{\partial p}\Big)^2\mathrm{P}_p(m)\right\}.$$

Очевидно, что

$$\sum_m [d(m)-p]^2\mathrm{P}_p(m)=\mathrm{D}_p d(\omega)$$

в силу несмещенности $d(\omega)$, а

$$\sum_m\Big(\frac{\partial \ln \mathrm{P}_p(m)}{\partial p}\Big)^2\mathrm{P}_p(m)=\mathrm{E}_p\Big(\frac{\partial \ln \mathrm{P}_p(\omega)}{\partial p}\Big)^2.$$

Итак, для любой несмещенной оценки $d(\omega)$ ее дисперсия удовлетворяет неравенству

$$\mathrm{D}_p d(\omega)\geqslant\frac{1}{\mathrm{E}_p\Big(\frac{\partial \ln \mathrm{P}_p(\omega)}{\partial p}\Big)^2}.$$

Математическое ожидание в знаменателе легко вычисляется:

$$\mathrm{E}_p\Big(\frac{\partial \ln \mathrm{P}_p(\omega)}{\partial p}\Big)^2=\frac{n}{p(1-p)}>0.$$

Таким образом, для любой несмещенной оценки $d(\omega)$ справедливо неравенство

$$\mathrm{D}_p d(\omega)\geqslant\frac{p(1-p)}{n}=\mathrm{D}_p\Big(\frac{S_n}{n}\Big),$$

правая часть которого совпадает с дисперсией частоты $d^*(\omega)$. Важно то, что ни для какой другой оценки, кроме частоты, равенство в полученном неравенстве не достигается. Это следует из того, что неравенство Коши—Буняковского обращается в равенство тогда и только

тогда, когда $b_m = \lambda a_m$ при всех m. Доказанное неравенство является частным случаем неравенства Рао—Крамера, или неравенства информации (см. теорему в § 3). □

Замечание 1. Хотя наилучшей оценкой параметра p, как доказано, служит частота $d^*(\omega)$, однако при дополнительной априорной информации о параметре могут существовать оценки, лучшие в смысле минимального среднего квадратичного риска. Пусть, например, $n = 2$ и $\Omega = (0, 1, 2)$. Оценим параметр p в предположении, что $\frac{1}{3} \leqslant p \leqslant \frac{2}{3}$. Частота $d^*(\omega)$ принимает значения 0, $\frac{1}{2}$, 1, при этом $\mathrm{E}_p[d^*(\omega) - p]^2 = \frac{p(1-p)}{2}$. Рассмотрим другую оценку $d(\omega)$ со значением $\frac{1}{2}$ во всех точках 0, 1, 2. Очевидно, что

$$\mathrm{E}_p d(\omega) = \frac{1}{2} \quad \text{и} \quad \mathrm{E}_p[d(\omega) - p]^2 = \left(p - \frac{1}{2}\right)^2.$$

Легко видеть (см. рис. 4), что $\mathrm{E}_p[d^*(\omega) - p]^2 > \mathrm{E}_p[d(\omega) - p]^2$ при $\frac{1}{3} \leqslant$ $\leqslant p \leqslant \frac{2}{3}$. Следовательно, в этом случае качество тривиальной оценки $d(\omega)$ лучше (отметим, что $d(\omega)$ имеет систематическое смещение).

Замечание 2. Оценка $d^*(\omega)$ определена при каждом конечном n. Отметим, что $d^*(\omega)$ при $n \to \infty$ сходится по вероятности к оцениваемому параметру p. Подобное свойство статистических оценок называется *состоятельностью* (см. § 3). Этот факт следует из закона больших чисел (см. теорему Бернулли в § 6 гл. 2).

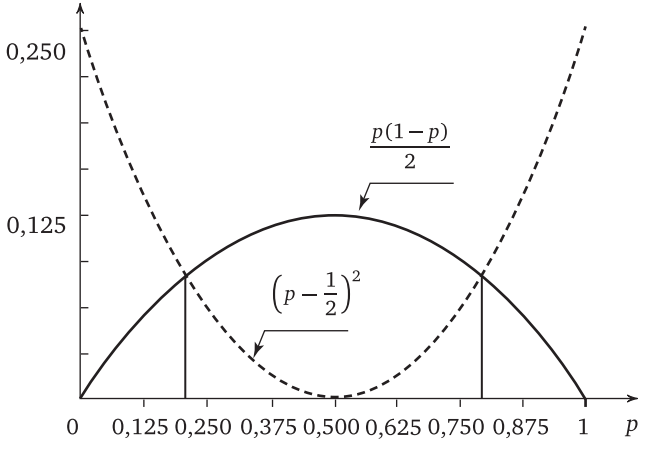


Рис. 4. Сравнение оценок параметра биномиального распределения по значению среднего квадратичного риска $\mathrm{E}_p[d^* - p]^2 = \frac{p(1-p)}{n}$ и $\mathrm{E}_p[d - p]^2 = \left(p - \frac{1}{2}\right)^2$

§ 3. Общие проблемы теории статистических оценок

Рассмотрим евклидово пространство R^n в качестве выборочного пространства. Удобно представлять результаты наблюдения в терминах случайных величин, а именно в виде случайного вектора $X = (X_1, \ldots, X_n)$, где независимые одинаково распределенные случайные величины $X_1, \ldots, X_n$ являются координатными величинами: $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$, $x = (x_1, \ldots, x_n) \in \mathrm{R}^n$. (Вектор X называется также случайной выборкой объема n.)

Предположим, что каждая из величин $X_1, \ldots, X_n$ распределена с плотностью $p(x_k; \theta)$, зависящей от параметра $\theta \in \Theta$. Плотность совместного распределения величин $X_1, \ldots, X_n$ в силу независимости равна $p(x; \theta) = \prod_{k=1}^{n} p(x_k; \theta)$. Будем рассматривать несмещенные оценки $\widehat{\theta}(X_1, \ldots, X_n)$ параметра θ по выборке $(X_1, \ldots, X_n)$, т. е. такие оценки, что

$$\mathrm{E}_\theta \widehat{\theta}(X_1, \ldots, X_n) = \theta,$$

и будем измерять качество оценок средним квадратичным риском $\mathrm{E}_\theta(\widehat{\theta} - \theta)^2$. Очевидно, что для несмещенных оценок $\mathrm{E}_\theta(\widehat{\theta} - \theta)^2 = \mathrm{D}_\theta \widehat{\theta}$. Естественно поэтому считать наилучшими оценки, имеющие наименьшее рассеяние около истинного значения параметра.

Теорема (Рао—Крамер). *Для любой несмещенной оценки*

$$\widehat{\theta}(X_1, \ldots, X_n)$$

параметра θ справедливо неравенство

$$\mathrm{D}_\theta \widehat{\theta} \geqslant \frac{1}{I_\theta}, \tag{1}$$

где

$$I_\theta = \int \Big(\frac{\partial \ln p(x; \theta)}{\partial \theta} \Big)^2 p(x; \theta)\, dx,$$

если выполнены следующие условия регулярности:

1) *плотность $p(x; \theta)$, $x \in \mathrm{R}^n$, дифференцируема по $\theta \in \Theta \subset \mathbb{R}^n$, и множество тех x, для которых $p(x; \theta) \neq 0$, не зависит от θ*;

2) *тождества*

$$\int p(x; \theta)\, dx \equiv 1 \tag{2}$$

и

$$\int \widehat{\theta}(x) p(x; \theta)\, dx \equiv \theta \tag{3}$$

можно дифференцировать под знакам интеграла по θ;

3) $I_\theta > 0$.

Исходными соотношениями для доказательства неравенства (1) служат тождества (2) и (3) по параметру θ, а само доказательство теоремы принципиально повторяет доказательство теоремы в § 2 и основано на интегральном варианте неравенства Коши—Буняковского. Неравенство (1) называют *неравенством Рао—Крамера* или *неравенством информации*, а I_θ называют *информацией Фишера*.

Замечание 1. Легко показать, что

$$I_\theta = n\int\limits_{\mathrm{R}^1}\Big(\frac{\partial \ln p(x_1;\theta)}{\partial\theta}\Big)^2 p(x_1;\theta)\,dx_1.$$

В самом деле, если $Y_k = \dfrac{\partial \ln p(x_k;\theta)}{\partial\theta}$, то

$$I_\theta = \mathrm{E}(Y_1+\ldots+Y_n)^2.$$

Так как

$$\mathrm{E}_\theta Y_k = \int\limits_{\mathrm{R}^1}\frac{\partial \ln p(x_k;\theta)}{\partial\theta}p(x_k;\theta)\,dx_k = \frac{\partial}{\partial\theta}\int\limits_{\mathrm{R}^1}p(x_k;\theta)\,dx_k = 0,$$

а $Y_1, \ldots, Y_n$ независимы, получаем, что

$$\mathrm{E}_\theta(Y_1+\ldots+Y_n)^2 = n\mathrm{E}_\theta Y_1^2,$$

где

$$\mathrm{E}_\theta Y_1^2 = \int\limits_{\mathrm{R}^1}\Big(\frac{\partial \ln p(x_1;\theta)}{\partial\theta}\Big)^2 p(x_1;\theta)\,dx_1 = \frac{1}{n}I_\theta.$$

Замечание 2. Равенство в формуле (1) при всех $\theta\in\Theta$ достигается в том и только том случае, когда

$$\frac{\partial \ln p(x;\theta)}{\partial\theta} = k(\theta)[\widehat{\theta}(x)-\theta], \tag{4}$$

где $k(\theta)$ не зависит от x.

Определение 1. Несмещенные оценки $\widehat{\theta}(X_1,\ldots,X_n)$, для которых неравенство Рао—Крамера обращается в равенство при всех значениях параметра θ, называются *эффективными* или, точнее, эффективными оценками по Рао—Крамеру, поскольку это неравенство выполняется при указанных выше условиях[1] 1—3.

[1] Следует учесть, что в разных источниках такое важное понятие, как эффективность статистической оценки, определяется неодинаково. А. А. Боровков [4] предложил рассматривать эффективные оценки в каждом выделенном классе допустимых оценок, например в классе всех оценок с фиксированным смещением. В контексте настоящего параграфа эффективность определяется условиями регулярности по Рао—Крамеру.

Следует, однако, проводить различие между эффективными оценками в указанном смысле, которые могут и не существовать, и несмещенными оценками с наименьшей дисперсией. Рассмотрим несколько примеров.

Пример 1. Пусть независимые наблюдения $X_1, \ldots, X_n$ имеют нормальное распределение с параметрами $\mathrm{E}X_k = a$ и $\mathrm{D}X_k = \sigma^2$. Предположим, что дисперсия σ^2 известна, а математическое ожидание a подлежит оценке. Сравним следующие оценки параметра a:

а) выборочное среднее

$$\widehat{a}_1(X_1, \ldots, X_n) = \frac{X_1 + \ldots + X_n}{n};$$

б) медиану выборки (при нечетном $n = 2m + 1$)

$$\widehat{a}_2(X_1, \ldots, X_n) = X_{(k+1)},$$

где $X_{(k+1)}$ определяется тем, что в расположении величин $X_1, \ldots, X_n$ в порядке их возрастания $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ слева и справа от $X_{(k+1)}$ лежит одинаковое число значений;

в) $\widehat{a}_3(X_1, \ldots, X_n) = \dfrac{X_{(1)} + X_{(2)}}{2}$, где $X_{(1)} = \min\limits_{1 \leqslant j \leqslant n} X_j$, $X_{(n)} = \max\limits_{1 \leqslant j \leqslant n} X_j$.

Все указанные оценки являются несмещенными:

$$\mathrm{E}_a \widehat{a}_1 = \mathrm{E}_a \widehat{a}_2 = \mathrm{E}_a \widehat{a}_3 = a.$$

Однако

$$\mathrm{D}_a \widehat{a}_1 = \frac{\sigma^2}{n},$$

$$\mathrm{D}_a \widehat{a}_2 = \frac{\sigma^2 \pi}{2n} + o\Big(\frac{1}{n}\Big),$$

$$\mathrm{D}_a \widehat{a}_3 = \frac{\sigma^2}{4 \ln n}\Big(\frac{\pi^2}{6} - 1\Big) + O\Big(\frac{1}{\ln^2 n}\Big)$$

(см. [9, п. 28.5 и 28.6]). Очевидно, что при больших n дисперсия оценки $\widehat{a}_2$ имеет тот же порядок, что и $\widehat{a}_1$ (но при этом $\mathrm{D}_a \widehat{a}_1 < \mathrm{D}_a \widehat{a}_2$), а оценка $\widehat{a}_3$ значительно хуже в этом отношении, чем $\widehat{a}_1$ и $\widehat{a}_2$. Легко показать, что для любой несмещенной оценки $\widehat{a}$ параметра a в силу неравенства Рао—Крамера

$$\mathrm{D}_a \widehat{a} \geqslant \frac{\sigma^2}{n}.$$

Вычислим значение нижней границы $\mathrm{D}_a \widehat{a}_1$. Воспользовавшись тем, что плотность нормального распределения с параметрами a и σ^2 равна

$$p(x; a; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{-\frac{(x-a)^2}{2\sigma^2}\Big\},$$

найдем

$$\frac{\partial \ln p(x;a;\sigma^2)}{\partial a} = \frac{x-a}{\sigma^2}.$$

Тогда (см. замечание 1 на с. 116)

$$I_a = n\int\limits_{\mathrm{R}^1} \frac{(x-a)^2}{\sigma^4} p(x;a;\sigma^2)\,dx = \frac{n}{\sigma^4}\int\limits_{\mathrm{R}^1} (x-a)^2 p(x;a;\sigma^2)\,dx = \\ = \frac{n}{\sigma^4}\sigma^2 = \frac{n}{\sigma^2},$$

а значит, $\frac{1}{I_a} = \frac{\sigma^2}{n}$. Поэтому $\mathrm{D}_a\widehat{a} \geqslant \mathrm{D}_a\widehat{a}_1$, т. е. эффективной оценкой параметра a является выборочное среднее $\widehat{a}_1$. Отметим, что хотя оценка $\widehat{a}_2$ несколько хуже, но вычислять ее на практике проще.

Пример 2. Предположим, что результаты наблюдений $X_1, \ldots, X_n$ распределены равномерно на отрезке $\left[a-\frac{1}{2}, a+\frac{1}{2}\right]$, где значение параметра a неизвестно. Очевидно, что $\mathrm{E}X_k = a$. В качестве оценок параметра a снова рассмотрим $\widehat{a}_1, \widehat{a}_2, \widehat{a}_3$, определенные в примере 1. Эти оценки являются несмещенными, но в данном случае

$$\mathrm{D}_a\widehat{a}_1 = \frac{1}{12n},$$
$$\mathrm{D}_a\widehat{a}_2 = \frac{1}{4(n+2)},$$
$$\mathrm{D}_a\widehat{a}_3 = \frac{1}{2(n+1)(n+2)}$$

(см. [9, п. 28.5 и 28.6]). Дисперсии оценок $\widehat{a}_1$ и $\widehat{a}_2$ имеют один порядок $\frac{1}{n}$ при больших n, а дисперсия $\widehat{a}_3$ — порядок $\frac{1}{n^2}$, т. е. оценка $\widehat{a}_3$ значительно точнее, чем $\widehat{a}_1$ и $\widehat{a}_2$. Следовательно, в этом случае выборочное среднее наблюдений не является наилучшей оценкой — существует несмещенная оценка $\widehat{a}_3$, дисперсия которой при всех n меньше. Такой результат непосредственно связан с тем, что для равномерного распределения, как легко проверить, не выполнены условия регулярности в теореме Рао—Крамера.

Пример 3. Пусть результаты наблюдения $X_1, \ldots, X_n$ распределены с показательной плотностью

$$p_\theta(x) = \begin{cases} \exp\{-(x-\theta)\}, & x \geqslant \theta, \\ 0, & x < \theta, \end{cases}$$

зависящей от параметра θ. Рекомендуется в качестве упражнения показать, что этот случай также не является регулярным в смысле условий 1—3 теоремы Рао—Крамера.

Пример 4. Рассмотрим случайные величины $X_1, \ldots, X_n$ в условиях примера 1. Но на этот раз предположим, что значение параметра σ^2 неизвестно. Оценим параметр σ^2 сначала в предположении, что значение a известно. В качестве оценки σ^2 рассмотрим величину

$$s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2.$$

Несложно показать, что $\mathrm{E}s_0^2 = \sigma^2$ и $\mathrm{D}s_0^2 = \frac{2\sigma^4}{n}$. Теорема Рао—Крамера дает в этом случае нижнюю границу для дисперсии всех несмещенных оценок параметра σ^2, равную $\frac{2\sigma^4}{n}$, т. е. s_0^2 есть эффективная оценка параметра σ^2 по Рао—Крамеру.

Если же a неизвестно, то, заменяя a несмещенной эффективной оценкой $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$, получим для оценки σ^2 величину

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}),$$

называемую выборочной дисперсией, для которой $\mathrm{E}s^2 = \frac{n-1}{n}\sigma^2$, $\mathrm{D}s^2 = \frac{2(n-1)}{n^2}\sigma^4$. Оценка s^2 является смещенной, но, устраняя смещение, перейдем к величине

$$s_1^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

для которой $\mathrm{E}s_1^2 = \sigma^2$, $\mathrm{D}s_1^2 = \frac{2\sigma^4}{n-1}$. Из очевидного неравенства

$$\frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} > \frac{2(n-1)\sigma^4}{n^2}$$

следует, что несмещенная оценка s_1^2 не является эффективной, а оценка s^2 хотя и смещена, но ее дисперсия меньше, чем нижняя граница в неравенстве Рао—Крамера. □

Итак, если $\widehat{\theta}(X)$ — несмещенная оценка параметра θ и выполнены условия регулярности 1—3, то качество оценки определяется близостью ее дисперсии к границе $\frac{1}{I_\theta}$.

Введенные понятия несмещенности и эффективности характеризуют качество оценки при любом конечном объеме выборки n. При отсутствии этих качеств существенными становятся асимпто-

тические свойства, такие, например, как асимптотические несмещенность и эффективность (см. об этом подробнее в [9]). Самым важным требованием, предъявляемым к оценкам, является их состоятельность.

Определение 2. Оценка $\widehat{\theta}_n = \widehat{\theta}(X_1, \ldots, X_n)$ параметра θ по n наблюдениям $X_1, \ldots, X_n$ называется *состоятельной*, если $\widehat{\theta}_n$ сходится по вероятности к θ при $n \to \infty$.

Любая несмещенная оценка параметра θ, дисперсия которой стремится к нулю при $n \to \infty$, является состоятельной (в силу неравенства Чебышёва). Все предложенные в этом параграфе оценки состоятельны.

Определение 3. Статистическая оценка $\widehat{\theta}_n$ параметра θ называется *асимптотически несмещенной*, если $M\widehat{\theta}_n \to \theta$ при $n \to \infty$ и любом $\theta \in \Theta$.

Например, выборочная дисперсия $s_n^2 = \frac{1}{n}\sum\limits_{k=1}^{n}(X_k - \bar{X}_n)^2$, являющаяся оценкой дисперсии σ^2 исходной модели, имеет при любом значении n смещение, но является асимптотически несмещенной оценкой, так как $Ms_n^2 \to \sigma^2$ при $n \to \infty$.

Асимптотическая эффективность обычно определяется в рамках условий, при которых выполняется неравенство Рао—Крамера. Для этого вводится коэффициент эффективности оценки $\widehat{\theta}_n$:

$$e_n(\widehat{\theta}_n) = \frac{\frac{1}{I(\theta)}}{\mathrm{D}\widehat{\theta}_n},$$

где $\mathrm{D}\widehat{\theta}_n$ — дисперсия оценки, а $I(\theta)$ — информация Фишера в выборке. Для эффективной оценки в этом случае $e_n(\widehat{\theta}_n) = 1$.

Определение 4. Оценка $\widehat{\theta}_n$ параметра θ называется *асимптотически эффективной*, если $e_n(\widehat{\theta}_n) \to 1$ при $n \to \infty$.

Например, несмещенная оценка дисперсии $s^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k - \bar{x}_n)^2$ не является эффективной оценкой, но обладает свойством асимптотической эффективности.

Во многих случаях, когда трудно обнаружить указанные выше свойства, нас может вполне удовлетворить свойство асимптотической нормальности.

Определение 5. Оценка $\widehat{\theta}_n$ параметра θ называется *асимптотически нормальной*, если распределение величины $\dfrac{\widehat{\theta}_n - \theta}{\sqrt{\mathrm{D}\widehat{\theta}_n}}$ стремит-

ся к стандартному нормальному распределению. В таком случае оценка $\widehat{\theta}_n$ является асимптотически несмещенной.

Во многих практических задачах определение 4 можно уточнить. Допустим, что существует возрастающая последовательность $c_n(\theta)$, сходящаяся при $n \to \infty$ и любом значении θ к нулю. Тогда оценку параметра θ называют асимптотически нормальной, если распределение $\dfrac{\widehat{\theta}_n - \theta}{c_n}$ стремится к стандартному нормальному распределению. Обычно $c_n = \dfrac{\sigma(\theta)}{\sqrt{n}}$, где $\sigma(\theta)$ — постоянная, не зависящая от выборочных значений, но, возможно, зависящая от параметра θ. В этом случае оценка $\widehat{\theta}_n$ является асимптотически нормальной оценкой параметра с асимптотическим средним θ и асимптотической дисперсией $\dfrac{\sigma^2(\theta)}{n}$.

§ 4. Методы получения точечных статистических оценок

Классическими методами нахождения оценок являются *метод моментов* и *метод максимального правдоподобия*. Метод моментов, наиболее простой в практическом отношении, заключается в следующем. По результатам наблюдений $X_1, \ldots, X_n$ определяются выборочные моменты $\dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^r$ или $\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^r$ порядка $r > 0$. Эти моменты рассматриваются как оценки соответствующих моментов распределения, зависящих от параметров. Предположим, что параметр θ s-мерный, т. е. $\theta = (\theta_1, \ldots, \theta_s) \in \mathrm{R}^s$. Тогда s выборочных моментов следует приравнять соответствующим моментам распределения. Полученные уравнения решаются относительно неизвестных параметров, и решения принимаются за оценки параметров по методу моментов.

Например, для оценки параметра p, $0 < p < 1$, биномиального распределения методом моментов составим уравнение $\bar{X}_n = p$, где $\bar{X}_n = \dfrac{X_1 + \ldots + X_n}{n}$ — выборочное среднее, или, иначе, выборочный момент первого порядка, а $p = \mathrm{E}X_1$ — математическое ожидание, или момент первого порядка теоретического распределения. В этом случае $\bar{X}_n = \widehat{p}_n$ есть оценка параметра p, и, как мы знаем, эта состоятельная оценка является несмещенной и эффективной.

Для нормального распределения метод моментов приводит к оценкам $\bar{X}$ и s^2 для a и σ^2, в других случаях найденные оценки не обладают наилучшими качествами. Можно лишь утверждать, что стати-

стические оценки по методу моментов при некоторых условиях регулярности обладают свойством состоятельности, хотя некоторые из них могут быть и эффективными оценками. Известно следующее утверждение.

Пусть параметр $\theta = (\theta_1, \ldots, \theta_s)$ является векторной величиной: $\theta \in \Theta \subset \mathrm{R}^s$. Рассмотрим $\alpha_k(\theta_1, \ldots, \theta_s)$ — момент порядка $k > 0$ распределения случайных величин $X_1, \ldots, X_n$ и $\widehat{\alpha}_{k,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ — выборочный момент того же порядка k, $1 \leqslant k \leqslant s$. В соответствии с методом моментов полагается составить s соотношений, приравнивая теоретические и выборочные моменты одного порядка:

$$\alpha_k(\theta_1, \ldots, \theta_s) = \widehat{\alpha}_{k,n}, \quad 1 \leqslant k \leqslant s,$$

и решить полученную систему уравнений относительно $\theta_1, \ldots, \theta_s$. Если существует единственное решение этой системы

$$\widehat{\theta}_k = f_k(\widehat{\alpha}_{1,n}, \ldots, \widehat{\alpha}_{s,n}), \quad 1 \leqslant k \leqslant s,$$

где $f_k(t_1, \ldots, t_s)$ — непрерывные функции всех своих аргументов, то оценки $\widehat{\theta}_1, \ldots, \widehat{\theta}_k$ являются состоятельными оценками соответствующих значений параметров, т. е. $\widehat{\theta}_k \xrightarrow{\mathrm{P}} \theta_k$ при $n \to \infty$ и любом значении k, $1 \leqslant k \leqslant s$. Для нормального распределения с параметром $\theta = (a, \sigma^2)$ полученные методом моментов оценка среднего значения $\widehat{a} = \bar{X}_n$ и оценка дисперсии $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_k - \bar{X}_n)^2$ являются состоятельными оценками.

Перейдем теперь к методу максимального правдоподобия, который в широком круге практических задач приводит к наилучшим оценкам. Рассмотрим задачу оценки параметра θ по результатам наблюдений $X_1, \ldots, X_n$. Введем функцию

$$L(\theta; X_1, \ldots, X_n) = p_\theta(X_1, \ldots, X_n),$$

которая при $X_1 = x_1, \ldots, X_n = x_n$ и фиксированном значении θ совпадает со значением плотности совместного распределения $p_\theta(x_1, \ldots, x_n)$ в точке $(x_1, \ldots, x_n)$ случайных величин $X_1, \ldots, X_n$ в непрерывном случае и со значением вероятности события $\{X_1 = x_1, \ldots, X_n = x_n\}$ в дискретном случае.

Определение 1. Случайная величина $L(\theta; X_1, \ldots, X_n)$, принимающая значения $L(\theta; x_1, \ldots, x_n)$ в точках $(x_1, \ldots, x_n)$, называется *функцией правдоподобия* вектора наблюдений $X_1, \ldots, X_n$. При конкретных значениях $x_1, \ldots, x_n$ величина $L(\theta) = L(\theta; x_1, \ldots, x_n)$ есть функция только параметра θ.

Определение 2. *Оценкой максимального правдоподобия* параметра θ называется такое значение $\theta \in \Theta$, при котором функция правдоподобия, рассматриваемая как функция параметра θ, достигает наибольшего возможного значения.

Таким образом, определение 2 соответствует принципу максимального правдоподобия: в качестве оценки $\widehat{\theta}(X_1, \ldots, X_n)$ выбирается такое значение параметра, при котором данные результаты наблюдения «наиболее вероятны». Если существует такое значение $\widehat{\theta}$, что

$$L(\widehat{\theta}) = \max_{\theta} L(\theta),$$

то $\widehat{\theta}$ — оценка максимального правдоподобия. Предположим, что функция $L(\theta)$ дифференцируема и точка ее максимума лежит внутри допустимого интервала значений θ, тогда в точке максимума

$$\frac{\partial L(\theta)}{\partial \theta} = 0$$

и $\theta = \widehat{\theta}$ есть решение этого уравнения. Во многих случаях функция $L(\theta)$ такова, что $\widehat{\theta}$ является единственным корнем указанного уравнения, называемого *уравнением правдоподобия*.

Используя логарифмическую производную функции $L(\theta)$, это уравнение удобно записать в виде

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0 \tag{1}$$

(мы не принимаем во внимание решения уравнения правдоподобия, которые не зависят от выборки, т. е. решения вида $\widehat{\theta} = \text{const}$). Для изучения свойств оценок максимального правдоподобия воспользуемся неравенством Рао—Крамера. Отмечалось, что это неравенство обращается в равенство тогда и только тогда, когда

$$\frac{\partial \ln L(\theta)}{\partial \theta} = k(\theta)(\widehat{\theta} - \theta), \tag{2}$$

где $k(\theta)$ зависит от θ, но не зависит от $x = (x_1, \ldots, x_n)$, а $\widehat{\theta}$ — оценка параметра θ. Если для оценки $\widehat{\theta}$ выполняется равенство (2), то $\widehat{\theta}$ — эффективная оценка параметра θ. Если существует эффективная оценка параметра θ, то в условиях регулярности 1—3 Рао—Крамера она может быть найдена методом максимального правдоподобия. В самом деле, оценка $\widehat{\theta}$ максимального правдоподобия является корнем уравнения (1), а из уравнения (2) следует, что в этом случае $\widehat{\theta} = \widehat{\theta}$. Очевидно, что $k(\theta) \neq 0$, так как иначе нарушается условие регулярности 3 (см. теорему в § 3).

Пример 1. Пусть случайные величины $X_1, \ldots, X_n$ представляют схему Бернулли, так что $\mathrm{P}\{X_k = 1\} = p$, $\mathrm{P}\{X_k = 0\} = 1 - p$. Оценим неизвестный параметр p. Так как $\mathrm{P}\{X_k = x_k\} = p^{x_k}(1-p)^{1-x_k}$, функция правдоподобия имеет вид

$$L(p) = p^{\sum\limits_{k=1}^{n} x_k}(1-p)^{n-\sum\limits_{k=1}^{n} x_k}.$$

Решаем уравнение правдоподобия

$$\frac{\partial \ln L(p)}{\partial p} = \frac{\sum\limits_{k=1}^{n} x_k}{p} - \frac{n - \sum\limits_{k=1}^{n} x_k}{1-p} = 0$$

и получаем оценку для p. Это выборочное среднее $\bar{X}$, т. е. $\widehat{p} = \dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ иными словами, частота появления «1» (ср. с § 2).

Пример 2. Если случайные величины $X_1, X_2, \ldots, X_n$ нормально распределены с параметрами a и σ^2, то логарифм функции правдоподобия равен

$$\ln L(a, \sigma) = -\frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(X_i - a)^2.$$

Решая уравнение правдоподобия, получим следующие оценки:

$$\widehat{a} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{для } a;$$

$$\widehat{\sigma}^2 = s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 \quad \text{для } \sigma^2,$$

если a известно. Если неизвестны и a, и σ^2, то оценкой σ^2 максимального правдоподобия является $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$. Соотнося полученные результаты с результатами примеров 1 и 3 из § 3, можно сделать вывод о том, что оценка $\widehat{a} = \bar{X}_n$, полученная методом максимального правдоподобия, является эффективной оценкой параметра a, а оценка максимального правдоподобия s^2 имеет смещение, но, поскольку при $n \to \infty$ смещение оценки стремится к нулю, являет собой пример асимптотически несмещенной оценки.

Метод максимального правдоподобия не всегда приводит к наилучшим оценкам при фиксированных n, однако при $n \to \infty$ оценки максимального правдоподобия обладают важнейшими свойствами: при некоторых условиях эти оценки состоятельны, асимптотически нормальны и асимптотически эффективны (см. [9, п. 33.3]).

Наиболее общим методом отыскания наилучших оценок, имеющим важное теоретическое и прикладное значение, является метод, использующий понятие *достаточной статистики*. См., например, [6, гл. 8.].

§ 5. Интервальные статистические оценки

Точечные статистические оценки, теории которых посвящены предыдущие параграфы, обладают одним существенным недостатком: даже для наилучших оценок невозможно в каждом конкретном случае оценить ошибку, возникающую при замене истинного значения параметра его оценкой. Выход из этого положения в определенной мере можно найти с помощью интервальных статистических оценок. Здесь мы рассмотрим лишь интервальные оценки, называемые доверительными интервалами, теория которых была разработана Нейманом и Пирсоном.

Определение. Пусть $X_1, \ldots, X_n$ — результаты наблюдений с распределением, зависящим от параметра θ. Если для произвольного заданного числа α, $0 < \alpha < 1$, две функции от результатов наблюдений $\widehat{\theta}_1(X_1, \ldots, X_n)$ и $\widehat{\theta}_2(X_1, \ldots, X_n)$ таковы, что

$$\mathrm{P}\{\widehat{\theta}_1(X_1, \ldots, X_n) \leqslant \theta \leqslant \theta_2(X_1, \ldots, X_n)\} \geqslant 1 - \alpha,$$

то интервал $[\widehat{\theta}_1, \widehat{\theta}_2]$ называется *доверительным интервалом* для параметра θ, соответствующим *доверительному уровню* α или *доверительной вероятности* $1 - \alpha$ (α также называют *вероятностью ошибки*).

В § 5 гл. 3 с помощью центральной предельной теоремы был построен приближенный доверительный интервал для неизвестного параметра биномиального распределения. Для подобных доверительных интервалов мы не можем указать границы $\widehat{\theta}_1$ и $\widehat{\theta}_2$, для которых вероятность события $\widehat{\theta}_1 \leqslant \theta \leqslant \widehat{\theta}_2$ в точности равна $1 - \alpha$. Построим теперь точные доверительные интервалы для параметров нормального распределения a и σ^2. Рассмотрим несколько случаев.

1. Пусть a неизвестно, а σ^2 известно. В этом случае, как было показано, несмещенной эффективной оценкой параметра a служит выборочное среднее

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n}.$$

Случайная величина $n\bar{X}$, как сумма независимых случайных величин с одинаковым нормальным распределением, имеет нормальное рас-

пределение с параметрами na и $n\sigma^2$, случайная величина $\bar{X}$ имеет нормальное распределение с параметрами a и $\frac{\sigma^2}{n}$, а случайная величина $\frac{\bar{X}-a}{\frac{\sigma}{\sqrt{n}}}$ — стандартное нормальное распределение с параметрами 0 и 1. Следовательно, при любом $\varepsilon>0$ и любых a и σ^2 имеем

$$\mathrm{P}_{a,\sigma^2}\left\{\frac{|\bar{X}-a|}{\frac{\sigma}{\sqrt{n}}}\leqslant\varepsilon\right\}=\frac{1}{\sqrt{2\pi}}\int\limits_{-\varepsilon}^{\varepsilon}e^{-y^2/2}\,dy=2\Phi(\varepsilon)-1,$$

а значит,

$$\mathrm{P}_{a,\sigma^2}\left\{\bar{X}-\frac{\sigma}{\sqrt{n}}\varepsilon\leqslant a\leqslant\bar{X}+\frac{\sigma}{\sqrt{n}}\varepsilon\right\}=2\Phi(\varepsilon)-1.$$

Как следует из последнего соотношения, вероятность того, что интервал $\left[\bar{X}-\frac{\sigma}{\sqrt{n}}\varepsilon,\bar{X}+\frac{\sigma}{\sqrt{n}}\varepsilon\right]$ накрывает неизвестное значение параметра a, не зависит от a. Пусть α, $0<\alpha<1$, произвольно. Найдем значение $\varepsilon_{1-\frac{\alpha}{2}}$ как корень уравнения (квантиль уровня $1-\frac{\alpha}{2}$)

$$\Phi(\varepsilon_{1-\frac{\alpha}{2}})=1-\frac{\alpha}{2},$$

пользуясь таблицами стандартного нормального распределения. Тогда при любом a получим

$$\mathrm{P}_{a,\sigma^2}\left\{\bar{X}-\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}}\leqslant a\leqslant\bar{X}+\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}}\right\}=1-\alpha.$$

Интервал $\left[\bar{X}-\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}},\bar{X}+\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}}\right]$ называется доверительным интервалом для параметра a доверительного уровня α. Назначая заранее доверительный уровень α, мы тем самым примиряемся с тем, что в среднем в доле случаев α мы можем ошибиться, принимая решение о том, что

$$\bar{X}-\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}}\leqslant a\leqslant\bar{X}+\frac{\sigma}{\sqrt{n}}\varepsilon_{1-\frac{\alpha}{2}}.$$

Замечание. Перед тем как продолжить построение доверительных интервалов, укажем два важных типа распределения вероятностей, определяемых по нормальному распределению. Пусть случайные величины $X_1,\ldots,X_n$ независимы и имеют стандартное нормальное распределение с параметрами 0 и 1. Тогда сумма $\chi_n^2=X_1^2+\ldots+X_n^2$ имеет распределение с плотностью

$$p_{\chi_n^2}(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}x^{n/2-1}e^{-x/2}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

где $\Gamma(m) = \int_0^\infty y^{m-1}e^{-y}\,dy$ — гамма-функция Эйлера. Распределение с такой плотностью называется *распределением хи-квадрат* (*распределением* χ^2) *с n степенями свободы*. Пусть, далее, случайная величина X распределена нормально с параметрами 0 и 1, а величина Y подчиняется распределению хи-квадрат с n степенями свободы. Если величины X и Y независимы, то случайная величина $T = \frac{X}{\sqrt{Y}}$ имеет распределение с плотностью

$$S_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{n}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty < x < \infty,$$

— так называемое *распределение Стьюдента с n степенями свободы*. Ввиду их особой важности в математической статистике распределения хи-квадрат и Стьюдента табулированы (см. [3]).

2. Построим теперь доверительный интервал для a при неизвестном значении σ^2. В качестве оценки σ^2 возьмем

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

и рассмотрим статистику

$$T_{n-1} = \sqrt{n-1}\,\frac{\bar{X} - a}{s_1}.$$

Легко показать, что распределение T_{n-1} не зависит от a и σ^2:

$$T_{n-1} = \sqrt{n-1}\,\frac{\frac{\bar{X}-a}{\sigma}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{(X_i-a)}{\sigma} - \frac{(\bar{X}-a)}{\sigma}\right]^2}} = \sqrt{n-1}\,\frac{\bar{X}'}{s_1'},$$

где $\bar{X}'$ и $(s_1')^2$ — это те же $\bar{X}$ и s_1^2, но для случайных величин $X_1', \ldots, X_n'$, нормально распределенных с параметрами 0 и 1. Труднее показать, что величина T_{n-1} имеет распределение Стьюдента с $n-1$ степенями свободы. Это заключение основано на известной *лемме Фишера*, в которой доказано, что для независимых случайных величин $X_1, \ldots, X_n$, имеющих нормальное распределение с параметрами a и σ^2, случайные величины

$$\bar{X}_n = \frac{X_1 + \ldots + X_n}{n} \quad \text{и} \quad s^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X}_n)^2$$

независимы (см. доказательство в книгах [9] и [12]). Из леммы Фишера следует, что статистика $(n-1)\frac{s_1^2}{\sigma^2}$ имеет распределение хи-квадрат с $n-1$ степенями свободы, и потому статистика $T_{n-1} = \sqrt{n-1}\frac{\bar{X}_n - a}{s_1}$ имеет распределение Стьюдента с $n-1$ степенями свободы (здесь $s_1 = +\sqrt{s_1^2}$). С помощью таблицы распределения Стьюдента для любого α, $0 < \alpha < 1$, можно найти такое значение $t_{1-\alpha}$, что

$$\mathrm{P}\left\{|T_{n-1}| \leqslant t_{1-\alpha}\right\} = \int\limits_{-t_\alpha}^{t_\alpha} S_{n-1}(t)\,dt = 1-\alpha.$$

Отсюда получаем доверительный интервал для a доверительного уровня α:

$$\mathrm{P}\left\{\bar{X} - t_{1-\alpha}\frac{s}{\sqrt{n-1}} \leqslant a \leqslant \bar{X} + t_{1-\alpha}\frac{s}{\sqrt{n-1}}\right\} = 1-\alpha.$$

3. Построим доверительный интервал для параметра σ^2 нормального распределения при неизвестном значении a. В качестве оценки дисперсии снова возьмем $s_1^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$. Рассмотрим случайную величину $(n-1)\frac{s_1^2}{\sigma^2}$, распределение которой не зависит от a и σ^2. Действительно,

$$(n-1)\frac{s_1^2}{\sigma^2} = (n-1)\left[\frac{1}{n-1}\sum_{k=1}^{n}\frac{(X_k - \bar{X})^2}{\sigma^2}\right] = (n-1)\left[\frac{1}{n-1}\sum_{k=1}^{n}(X_k' - \bar{X}')^2\right],$$

где $X_1', \ldots, X_n'$ — нормально распределенные случайные величины с параметрами 0 и 1, а $\bar{X}'$ — их арифметическое среднее. Так как случайная величина $(n-1)\frac{s_1^2}{\sigma^2}$ подчиняется распределению хи-квадрат с $n-1$ степенями свободы, для заданного α, $0 < \alpha < 1$, можно по таблице распределения хи-квадрат определить такие $g'_{\frac{\alpha}{2}}$ и $g''_{1-\frac{\alpha}{2}}$ $(g'_{\frac{\alpha}{2}} < g''_{1-\frac{\alpha}{2}})$, что

$$\int\limits_0^{g'_{\frac{\alpha}{2}}} p_{\chi^2_{n-1}}(x)\,dx = \frac{\alpha}{2}, \qquad \int\limits_{g''_{1-\frac{\alpha}{2}}}^{\infty} p_{\chi^2_{n-1}}(x)\,dx = 1-\frac{\alpha}{2}.$$

Поэтому, каковы бы ни были a и σ^2, при доверительном уровне α имеем

$$\mathrm{P}_{a,\sigma^2}\left\{g'_{\frac{\alpha}{2}} \leqslant (n-1)\frac{s_1}{\sigma^2} \leqslant g''_{1-\frac{\alpha}{2}}\right\} = 1-\alpha,$$

и, следовательно,

$$\mathrm{P}_{a,\sigma^2}\Big\{\frac{(n-1)s_1^2}{g''_{1-\frac{\alpha}{2}}} \leqslant \sigma^2 \leqslant \frac{(n-1)s_1^2}{g'_{\frac{\alpha}{2}}}\Big\} = 1-\alpha.$$

Процедура построения доверительных интервалов во всех предложенных случаях сравнительно проста. Это объясняется тем, что распределения выборочных характеристик вычисляются по нормальному распределению. Для других законов распределения задача нахождения точных распределений функций от результатов наблюдений оказывается трудноразрешимой. В этих случаях часто строят приближенные доверительные интервалы, используя предельные теоремы для соответствующих распределений, в частности центральную предельную теорему (см. [3]).

§ 6. Проверка статистических гипотез

В гл. 1 была сформулирована задача различения двух гипотез о вероятности «удачи» в схеме Бернулли с заданными ошибками 1-го и 2-го рода, которая затем была решена приближенно с помощью теоремы Лапласа (см. § 5 гл. 3).

Вернемся к задаче различения двух гипотез о параметре p в схеме Бернулли. Если имеются две гипотезы о значении параметра p, а именно $\mathrm{H}_1\colon p = p_1$ и $\mathrm{H}_2\colon p = p_2$ $(p_1 < p_2)$, и необходимо по результатам наблюдений выбрать одну из них, то возникает вопрос, каким должно быть число наблюдений, достаточное для того, чтобы различить эти гипотезы с вероятностями ошибок 1-го и 2-го рода, не превосходящими малых чисел α и β. В § 5 гл. 3 было дано приближенное решение этой задачи. Здесь предлагается точное решение поставленной проблемы.

Итак, случайная величина S_n — число «удач» в n испытаниях Бернулли — подчинена биномиальному распределению с параметром p, т. е.

$$\mathrm{P}\{S_n = m\} = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \ldots, n.$$

Пусть

$$P_1(m) = C_n^m p_1^m (1-p_1)^{n-m} \quad \text{и} \quad P_2(m) = C_n^m p_2^m (1-p_2)^{n-m}$$

— значения биномиальных вероятностей, которые соответствуют двум гипотезам о значении параметра p. Выбор между гипотезами H_1 и H_2 произведем на основании следующего правила.

Пусть для некоторого значения n существует такое число m^*, $0 < m^* < n$, что при $m \leqslant m^*$ мы полагаем, что $p = p_1$, т. е. верна гипотеза H_1, а при $m > m^*$ считаем, что $p = p_2$, т. е. верна гипотеза H_2. Очевидно, что при любом выборе значения m^*, которое естественно назвать *критическим*, могут быть допущены ошибки. Зададим произвольно малые числа α, β, $0 < \alpha < \frac{1}{2}$ и $0 < \beta < \frac{1}{2}$, и потребуем выполнения неравенств

$$\mathrm{P}_{\mathrm{H}_1}\{S_n > m^*\} \leqslant \alpha \quad \text{и} \quad \mathrm{P}_{\mathrm{H}_2}\{S_n \leqslant m^*\} \leqslant \beta.$$

Левые части этих неравенств представляют собой вероятности возможных ошибок: $\mathrm{P}_{\mathrm{H}_1}\{S_n > m^*\}$ — это вероятность принять гипотезу H_2 в случае, когда верна гипотеза H_1 (ошибка 1-го рода), $\mathrm{P}_{\mathrm{H}_2}\{S_n \leqslant m^*\}$ — это вероятность принять гипотезу H_1, если верна гипотеза H_2 (ошибка 2-го рода).

Мы ставим цель — получить неравенство с нижней оценкой числа n наблюдений, позволяющего решить поставленную задачу. Для этого воспользуемся неравенством Иенсена для выпуклых функций, а именно для функции $\ln x$, $x > 0$:

$$\sum_m (\ln x_m)\frac{\lambda_m}{\Lambda} \leqslant \ln\left(\sum_m x_m \cdot \frac{\lambda_m}{\Lambda}\right) \tag{1}$$

при $x_m > 0$, $\lambda_m > 0$, $\Lambda = \sum\limits_m \lambda_m$ и $m = 0, \ldots, n$. Введем необходимые обозначения:

$$\alpha' = \mathrm{P}_{\mathrm{H}_1}\{S_n > m^*\}, \qquad 1 - \alpha' = \mathrm{P}_{\mathrm{H}_1}\{S_n \leqslant m^*\},$$

$$\beta' = \mathrm{P}_{\mathrm{H}_2}\{S_n \leqslant m^*\}, \qquad 1 - \beta' = \mathrm{P}_{\mathrm{H}_2}\{S_n > m^*\},$$

$$S_1 = \sum_{m > m^*} \ln(x_m)\mathrm{P}_1(m), \qquad S_2 = \sum_{m \leqslant m^*} (\ln x_m)\mathrm{P}_1(m),$$

$$\sigma_1 = \sum_{m > m^*} (\ln x_m)\mathrm{P}_2(m), \qquad \sigma_2 = \sum_{m \leqslant m^*} (\ln x_m)\mathrm{P}_2(m),$$

$$S = S_1 + S_2 = \sum_m \ln(x_m)\mathrm{P}_1(m), \quad \sigma = \sigma_1 + \sigma_2 = \sum_m \ln(x_m)\mathrm{P}_2(m).$$

Используем функцию Кульбака $\varphi(a; b) = b \ln \frac{b}{1-a} + (1-b) \ln \frac{1-b}{a}$, которая в области определения неотрицательна и убывает по каждой переменной a и b при $0 < a < \frac{1}{2}$, $0 < b < \frac{1}{2}$. Несложно показать, что

$$S = \sum_m \left(\ln \frac{\mathrm{P}_2(m)}{\mathrm{P}_1(m)}\right)\mathrm{P}_1(m) = -n\varphi(1 - p_2, p_1)$$

и

$$\sigma = \sum_m \Big(\ln \frac{\mathrm{P}_2(m)}{\mathrm{P}_1(m)}\Big)\mathrm{P}_2(m) = n\varphi(1-p_1, p_2).$$

Подставим в неравенство Иенсена (1) значения $x_m = \dfrac{\mathrm{P}_2(m)}{\mathrm{P}_1(m)}$, $\lambda = \mathrm{P}_1(m)$, $\Lambda = \alpha'$ и получим неравенства

$$S_1 \leqslant \alpha' \ln \frac{1-\beta'}{\alpha'} \quad \text{и} \quad S_2 \leqslant (1-\alpha') \ln \frac{\beta'}{1-\alpha'}.$$

Поскольку $S = S_1 + S_2$ и $S = -n\varphi(1-p_2, p_1)$, то

$$-n\varphi(1-p_2, p_1) = S \leqslant \alpha' \ln \frac{1-\beta'}{\alpha'} + (1-\alpha') \ln \frac{\beta'}{1-\alpha'},$$

а это означает, что $-n\varphi(1-p_2, p_1) \leqslant -n\varphi(\beta', \alpha')$, откуда следует, что

$$n \geqslant \frac{\varphi(\beta', \alpha')}{\varphi(1-p_2, p_1)} \geqslant \frac{\varphi(\beta, \alpha)}{\varphi(1-p_2, p_1)}. \tag{2}$$

Последнее неравенство есть следствие монотонности функции $\varphi(a, b)$. Аналогичным образом выводим неравенство

$$n \geqslant \frac{\varphi(\alpha, \beta)}{\varphi(1-p_1, p_2)}. \tag{3}$$

В итоге получаем неравенство для n:

$$n \geqslant \max(\gamma_1, \gamma_2), \tag{4}$$

где $\gamma_1 = \dfrac{\varphi(\beta, \alpha)}{\varphi(1-p_2, p_1)}$, $\gamma_2 = \dfrac{\varphi(\alpha, \beta)}{\varphi(1-p_1, p_2)}$. Поскольку α, β, p_1, p_2 заданы, это неравенство дает нам представление о числе наблюдений, достаточном для того, чтобы решить задачу различения двух гипотез H_1 и H_2 с вероятностями ошибок 1-го и 2-го рода, не превосходящими заданных значений. Некоторые практические результаты наших расчетов можно увидеть в таблице (наименьшее целое значение n обозначено через n_0).

H_1	H_2	$\alpha = 0{,}05$	$\beta = 0{,}25$
p_1	p_2		
0,4	0,5	$n_0 = 150$	$m^* = 70$
0,2	0,4	$n_0 = 36$	$m^* = 11$
0,01	0,02	$n_0 = 1000$	$m^* = 15$

Можно добавить к предыдущему графический метод проверки двух гипотез $\mathrm{H}_1\colon p = p_1$ и $\mathrm{H}_2\colon p = p_2$ в схеме Бернулли (этот метод

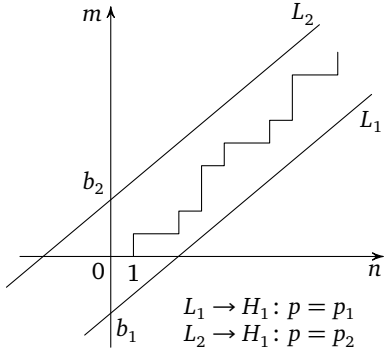


Рис. 5. Графическая иллюстрация к проверке гипотез $H_1 : p = p_1$ и $H_1 : p = p_2$ по методу Вальда

предложен А. Вальдом, см. [5]). На рис. 5 число испытаний n откладывается по горизонтальной оси. Если испытание заканчивается «неудачей» («0»), то точка на графике сдвигается вправо на единицу по горизонтали, если же испытание заканчивается «удачей» («1»), то точка на графике сдвигается вверх на единицу. Полученная ломаная соответствует накопленному числу «удач» за n испытаний. Проведем две параллельные прямые L_1 и L_2 с угловым коэффициентом

$$k = \frac{\ln \dfrac{1-p_1}{1-p_2}}{\ln \dfrac{p_2(1-p_1)}{(1-p_2)p_1}}.$$

Точки пересечения этих прямых с вертикальной осью соответственно равны

$$b_1 = \frac{\ln \dfrac{\beta}{1-\alpha}}{\ln \dfrac{p_2(1-p_1)}{(1-p_2)p_1}} \quad \text{и} \quad b_2 = \frac{\ln \dfrac{1-\beta}{\alpha}}{\ln \dfrac{p_2(1-p_1)}{(1-p_2)p_1}},$$

где α и β — вероятности ошибок 1-го и 2-го рода соответственно. Эти прямые наносятся на график перед началом эксперимента, и число n заранее неизвестно. На каждом шаге при проведении испытаний мы или принимаем гипотезу H_1, или принимаем гипотезу H_2, или продолжаем испытания. Таким образом, процедура проверки гипотез продолжается до тех пор, пока ломаная не выходит за пределы полосы между прямыми L_1 и L_2, но, как только точка лома-

ной попадает на линию L_1 или ниже ее, принимаем гипотезу H_1, а если точка оказывается на прямой L_2 или выше ее, то принимаем гипотезу H_2. Для числа наблюдений сохраняется та же нижняя граница (2)—(4), но при этом число наблюдений значительно сокращается по сравнению с фиксированным заранее и этот результат практически неулучшаем.

Рассмотрим более общую постановку задачи о различении гипотез. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с распределением P_θ, зависящим от числового параметра $\theta \in \Theta \subseteq R^1$. Вся совокупность наблюдений $X = (X_1, \ldots, X_n)$ есть точка выборочного пространства Ω. Будем проверять гипотезу H_0 о том, что $\theta \in \Theta_0$, где Θ_0 — заданное подмножество параметрического множества Θ. Выделим в Ω до эксперимента некоторое подмножество S, которое будем называть *критическим множеством* в следующем смысле; если $(X_1, \ldots, X_n) \in S$, то гипотеза H_0 отклоняется; если же $(X_1, \ldots, X_n) \in \overline{S}$, то H_0 принимается. При этом, как мы уже знаем, возможны ошибки. Ошибка 1-го рода состоит в том, что отклоняется на самом деле верная гипотеза H_0, ошибка 2-го рода — в том, что принимается неверная гипотеза H_0. Очевидно, что для фиксированного множества S вероятности этих ошибок равны соответственно $P_\theta\{X \in S\}$, $\theta \in \Theta_0$, и $P_\theta\{X \in \overline{S}\}$, $\theta \notin \Theta_0$. Поставленная задача решается следующим образом: фиксируется малое число α, $0 < \alpha < 1$, и из условия $P_\theta\{X \in S\} \leqslant \alpha$ определяется критическое множество S; если $X \in S$, то гипотеза H_0 отклоняется, при этом вероятность ошибки не превосходит α. Легко понять, что для фиксированного α множество S можно выбрать не единственным способом.

Пример. Пусть величины $X_1, \ldots, X_n$ нормально распределены с параметрами a и σ^2, при этом пусть a неизвестно, а $\sigma^2 = 1$. Предположим, что $-\infty < a < \infty$, и проверим гипотезу H_0: $a = 0$. Как мы знаем, случайная величина $\sqrt{n}(\bar{X} - a)$, где $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$, распределена нормально с параметрами 0 и 1 при любом значении a из параметрического множества $(-\infty, +\infty)$. В том случае, если гипотеза H_0 верна, величина $\sqrt{n}\bar{X}$ распределена нормально с параметрами 0 и 1. Поэтому для любого заданного числа α, $0 < \alpha < 1$, можно по таблице нормального распределения найти такое значение u_α, что

$$P_a\{|\sqrt{n}\bar{X}| > u_\alpha\} = \frac{2}{\sqrt{2\pi}} \int_{u_\alpha}^{\infty} e^{-z^2/2}\, dz = \alpha.$$

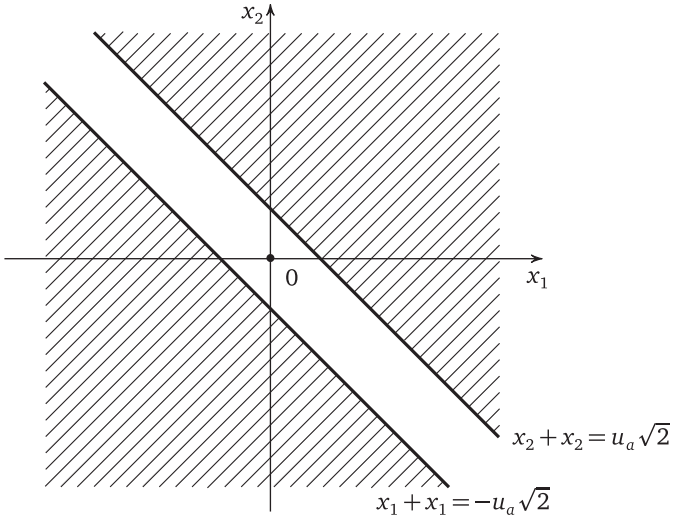


Рис. 6. Критическое множество (заштрихованная область) для проверки гипотезы H_0: $a = 0$ при $n = 2$

В соответствии с договоренностью устанавливаем следующее правило принятия решения: если $|\sqrt{n}\bar{X}| \leqslant u_\alpha$, то H_0 принимается, а если $|\sqrt{n}\bar{X}| > u_\alpha$, то H_0 отклоняется. При этом вероятность отклонить гипотезу H_0 в случае, если она на самом деле справедлива, будет в точности равна α. Очевидно, критическое множество S совпадает здесь с множеством всех точек $(x_1, \ldots, x_n)$ пространства R^n, для которых $|x_1 + \ldots + x_n| > u_\alpha\sqrt{n}$ (например, при $n = 2$ это внешность указанной на рис. 6 полосы, образованной прямыми $x_1 + x_2 = -u_\alpha\sqrt{2}$ и $x_1 + x_2 = u_\alpha\sqrt{2}$).

Если параметрическое множество имеет вид $[0, +\infty)$, то критическое множество при заданном α для проверки гипотезы H_0: $a = 0$ находится из соотношения

$$\mathrm{P}_a\{\sqrt{n}\bar{X} > u'_\alpha\} = \frac{1}{\sqrt{2\pi}} \int\limits_{u'_\alpha}^{\infty} e^{-z^2/2}\, dz = \alpha.$$

Отметим, что в рассуждениях, предшествующих примеру, и в самом примере мы оценивали лишь возможность ошибки при отклонении справедливой гипотезы H_0 и игнорировали тот случай, когда гипотеза H_0, которую мы считали согласующейся с результатами

эксперимента, оказывалась неверна. Критерии, на основании которых решается лишь вопрос о согласии между гипотезой H_0 и экспериментальными данными, называются *критериями согласия*. Для того чтобы оценить возможность ошибки 2-го рода, необходимо фиксировать не только основную гипотезу H_0, но и множество допустимых альтернативных гипотез (в рассмотренном примере такими альтернативными гипотезами могут быть $H_1: a \neq 0$ или $H_1: a > 0$). Предположим, что проверяется гипотеза $H_0: \theta \in \Theta_0$ при альтернативной гипотезе $H_1: \theta \in \Theta_1$, $\Theta_1 = \Theta \backslash \Theta_0$. При выборе критерия для проверки гипотезы H_0 сначала задается число ε, $0 < \varepsilon < 1$, а затем рассматриваются все критические множества S, $S \subset \Omega$, для которых

$$\sup_{\theta \in \Theta} \mathrm{P}_\theta\{X \in S\} = \varepsilon.$$

Число ε при этом называется *уровнем значимости* критерия с критическим множеством S. Поскольку каждый критерий определяется критическим множеством, можно отождествлять критерий с критическим множеством S и говорить о S-критерии. Если S фиксировано, то вероятность $\mathrm{P}_\theta\{X \in S\} = \varphi(\theta)$ является функцией лишь параметра θ. После того как ε задано, можно выбирать различные критерии, т. е. различные критические множества, или, что то же самое, различные функции $\varphi(\theta)$. Естественно выбрать тот критерий, который с большой вероятностью отклоняет проверяемую гипотезу H_0, если она неверна. Если, например, две функции $\varphi_1(\theta)$ и $\varphi_2(\theta)$, соответствующие разным критериям, таковы, что $\varphi_1(\theta) > \varphi_2(\theta)$ для всех $\theta \in \Theta_1$, то критерий с функцией $\varphi_1(\theta)$ считается лучшим по сравнению с критерием, определяемым функцией $\varphi_2(\theta)$.

Функция $\varphi(\theta)$ называется *функцией мощности* критерия, а значения $\varphi(\theta)$ при $\theta \in \Theta_1$ характеризуют *мощность* критерия как вероятность отклонения неверной гипотезы H_0 в пользу верной гипотезы H_1, т. е. вероятность принятия верной альтернативной гипотезы H_1. Принцип сравнения критериев: критерии одного уровня значимости сравниваются по мощности. Заметим, что в этом смысле идеален был бы тот критерий, который отклонял бы гипотезу H_0 с вероятностью 1 при $\theta \in \Theta_1$, т. е. в том случае, когда справедлива гипотеза H_1, однако, вообще говоря, такие критерии не существуют. Таким образом, из всех критериев, для которых вероятность ошибки 1-го рода не превосходит ε:

$$\alpha = \mathrm{P}_\theta\{X \in S\} \leqslant \varepsilon, \quad \theta \in \Theta_0,$$

мы должны выбрать тот критерий, для которого функция мощности максимальна для всех $\theta \in \Theta_1$, или, что то же самое, минимальна вероятность ошибки 2-го рода

$$\beta = \mathrm{P}_\theta\{X \in \bar{S}\}, \quad \theta \in \Theta_0.$$

Критерий проверки гипотезы H_0, имеющий наибольшую мощность в пользу гипотезы H_1 среди всех критериев с одинаковым уровнем значимости, называется *наиболее мощным критерием*.

§ 7. Фундаментальная лемма Неймана—Пирсона

Для того чтобы выяснить возможность построения наиболее мощных критериев, проведем разграничение статистических гипотез на типы. Статистическая гипотеза называется *простой*, если она определяет единственное значение параметра и, следовательно, единственный закон распределения. Все другие статистические гипотезы являются *сложными* и могут быть представлены как множество простых гипотез.

Рассмотрим случай, когда параметрическое множество состоит из двух точек θ_0 и θ_1 ($\theta_0 < \theta_1$) и необходимо проверить две простые гипотезы:

$$H_0\colon \theta = \theta_0; \quad H_1\colon \theta = \theta_1.$$

Как следует из приведенной ниже леммы Неймана—Пирсона, в этом случае существует наиболее мощный критерий.

Фундаментальная лемма (Нейман—Пирсон)**.** *Пусть вектор наблюдений $X = (X_1, \ldots, X_n)$ распределен с плотностью вероятности $p_\theta(x)$. Обозначим $p_0(x) = p_\theta(x)$ при $\theta = \theta_0$ и $p_1(x) = p_\theta(x)$ при $\theta = \theta_1$. Тогда при проверке гипотезы $H_0\colon \theta = \theta_0$ против альтернативной гипотезы $H_1\colon \theta = \theta_1$ среди всех критериев уровня значимости ε наиболее мощным является критерий с критическим множеством*

$$S^* = \left\{x\colon \frac{p_1(x)}{p_0(x)} > C\right\}, \tag{1}$$

где $C > 0$ определено условием[1]

$$\mathrm{P}_\theta\{X \in S^* \mid H_0\} = \varepsilon. \tag{2}$$

Отношение в правой части равенства (1), определяющее критическое множество, называется *отношением правдоподобий*, так как

[1] Обозначение $\mathrm{P}_\theta\{X \in S^* \mid H_0\}$ следует понимать как вероятность $\mathrm{P}_\theta\{X \in S^*\}$ события $\{X \in S^*\}$, вычисленную при условии, что $\theta = \theta_0$.

в числителе и знаменателе стоят функции правдоподобия выборки $(X_1, \ldots, X_n)$, соответствующие гипотезам H_1 и H_0. Критерий с критической областью S^* называется *критерием отношения правдоподобий*.

Мы докажем лемму Неймана—Пирсона в упрощенном варианте.

Теорема. *Пусть* $S^*(C) = \left\{x\colon \dfrac{p_1(x)}{p_0(x)} \geqslant C\right\}$ *при любом* $C > 0$ *для всех* x, *для которых* $p_0(x) > 0$. *Пусть* $0 < \alpha < 1$ *и существует такое значение* C_α, *что*[1]

$$\alpha = \mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_0\}. \tag{3}$$

Тогда критерий с критическим множеством $S^*(C_\alpha)$ *является наиболее мощным при проверке гипотезы* H_0 *против* H_1 *среди всех критериев с ошибкой* 1*-го рода* α.

Доказательство. Нужно доказать, что для любого множества S такого, что $p_0(x) > 0$, $x \in S$, и

$$\mathrm{P}_\theta\{X \in S \mid H_0\} \leqslant \alpha, \tag{4}$$

будет выполнено неравенство

$$\mathrm{P}_\theta\{X \in S \mid H_1\} \leqslant \mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_1\}.$$

По определению $S^*(C_\alpha)$ — подмножество выборочного пространства Ω, содержащее все точки x, для которых $p_0(x) > 0$ и $\dfrac{p_1(x)}{p_0(x)} > C_\alpha$, и удовлетворяющее условию (3). Пусть множество S удовлетворяет условию (4). Рассмотрим разность

$$R = \mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_1\} - \mathrm{P}_\theta\{X \in S \mid H_1\}.$$

Пусть $\Delta = S^*(C_\alpha) \cap S$. Тогда

$$\mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_1\} = \int\limits_{S^*(C_\alpha)} p_1(x)\,dx = \int\limits_{S^*(C_\alpha)\setminus\Delta} p_1(x)\,dx + \int\limits_{\Delta} p_1(x)\,dx$$

и

$$\mathrm{P}_\theta\{X \in S \mid H_1\} = \int\limits_{S} p_1(x)\,dx = \int\limits_{S\setminus\Delta} p_1(x)\,dx + \int\limits_{\Delta} p_1(x)\,dx.$$

Очевидно, что

$$R = \int\limits_{S^*(C_\alpha)\setminus\Delta} p_1(x)\,dx - \int\limits_{S\setminus\Delta} p_1(x)\,dx.$$

[1] Здесь уровень значимости совпадает с вероятностью ошибки 1-го рода.

Так как для всех $x \in S^*(C_\alpha)\backslash\Delta$ справедливо неравенство $\dfrac{p_1(x)}{p_0(x)} > C_\alpha$, получаем, что

$$R \geqslant C_\alpha \int\limits_{S^*(C_\alpha)\backslash\Delta} p_0(x)\,dx - \int\limits_{S\backslash\Delta} p_1(x)\,dx. \tag{5}$$

Если x' — это те точки x, которые принадлежат $S\backslash\Delta$, то $x' \notin S^*(C_\alpha)$. Тогда в точках x' имеем

$$\frac{p_1(x')}{p_0(x')} \leqslant C_\alpha.$$

Из неравенства (5) при этом получаем

$$R \geqslant C_\alpha \int\limits_{S^*(C_\alpha)\backslash\Delta} p_0(x)\,dx - C_\alpha \int\limits_{S\backslash\Delta} p_0(x)\,dx.$$

Далее прибавим и вычтем в правой части $C_\alpha \int\limits_\Delta p_0(x)\,dx$. Имеем

$$R \geqslant C_\alpha\left[\int\limits_{S^*(C_\alpha)\backslash\Delta} p_0(x)\,dx + \int\limits_\Delta p_0(x)\,dx\right] -$$

$$- C_\alpha\left[\int\limits_{S\backslash\Delta} p_0(x)\,dx + \int\limits_\Delta p_0(x)\,dx\right] = C_\alpha\left[\int\limits_{S^*(C_\alpha)} p_0(x)\,dx - \int\limits_S p_0(x)\,dx\right] =$$

$$= C_\alpha[\mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_0\} - \mathrm{P}_\theta\{X \in S \mid H_0\}],$$

но с учетом предположений (3) и (4)

$$\mathrm{P}_\theta\{X \in S^*(C_\alpha) \mid H_0\} - \mathrm{P}_\theta\{X \in S \mid H_0\} \geqslant 0,$$

откуда в силу того, что $C_\alpha > 0$, получаем окончательный результат, который доказывает теорему. □

Замечание. В доказанной теореме постулировалось существование для α такой постоянной C_α, что

$$\alpha = \mathrm{P}\Big\{X\colon \frac{p_1(x)}{p_0(x)} > C_\alpha\Big\},$$

так как не для всякого α постоянная C с таким свойством существует. В отличие от этой теоремы, в фундаментальной лемме Неймана—Пирсона утверждается существование наиболее мощного критерия, для которого вероятность ошибки 1-го рода в точности равна наперед заданному α. Достигается это путем перехода к так называемым *рандомизированным* критериям, которые обычно задаются

при помощи *критической* функции $\psi(x)$, $0 \leqslant \psi(x) \leqslant 1$, так, что если $\psi(x) = 0$ или $\psi(x) = 1$, то гипотеза H_0 соответственно принимается или отклоняется; если же $0 < \psi(x) < 1$, то гипотеза H_0 принимается с вероятностью $1 - \psi(x)$ и отклоняется с вероятностью $\psi(x)$. В классе рандомизированных критериев наиболее мощный критерий всегда существует. При выполнении некоторых ограничений на функцию $f(C) = \mathrm{P}\left\{\frac{p_1(x)}{p_0(x)} > C\right\}$ (например, непрерывности, см. пример ниже) наиболее мощный критерий будет нерандомизированным для любых значений α. См. об этом подробнее в книге [10].

Пример. Пусть $X_1, \ldots, X_n$ — независимые нормально распределенные величины с параметрами a и σ^2. Положим для определенности $\sigma^2 = 1$ и проверим простую гипотезу $H_0\colon a = 0$ при простой альтернативной гипотезе $H_1\colon a = a_1 > 0$. По лемме Неймана—Пирсона наиболее мощным будет критерий, критическое множество которого определяется неравенством

$$\exp\Big(a_1 n\bar{X} - \frac{na_1^2}{2}\Big) > C,$$

где $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$. Очевидно, что гипотеза H_0 отклоняется, когда

$$\bar{X} > \frac{a_1}{2} + \frac{\ln C}{na_1}.$$

Постоянную C определим из условия, что вероятность ошибки 1-го рода равна α, $0 < \alpha < 1$. При гипотезе H_0 случайная величина $\sqrt{n}\bar{X}$ имеет нормальное распределение со средним 0 и дисперсией 1. Поэтому

$$\mathrm{P}_a\Big\{\sqrt{n}\bar{X} > \frac{a_1\sqrt{n}}{2} + \frac{\ln C}{a_1\sqrt{n}}\Big\} = 1 - \Phi\Big(\frac{a_1\sqrt{n}}{2} + \frac{\ln C}{a_1\sqrt{n}}\Big),$$

где $\Phi(x)$ — функция стандартного нормального распределения. Обозначим через $u_{1-\alpha}$ корень уравнения $1 - \Phi(u) = \alpha$. Тогда C определяется из равенства $u_{1-\alpha} = \frac{a_1\sqrt{n}}{2} + \frac{\ln C}{a_1\sqrt{n}}$. Вероятность ошибки 2-го рода при этом равна

$$\beta = \mathrm{P}_a\Big\{\bar{X} \leqslant \frac{1}{2}a_1 + \frac{\ln C}{a_1 n}\Big\}$$

при условии, что справедлива гипотеза H_1. В этом случае величина $\sqrt{n}(\bar{X} - a_1)$ имеет нормальное распределение с параметрами 0 и 1 и

$$\mathrm{P}_a\Big\{\bar{X} \leqslant \frac{1}{2}a_1 + \frac{\ln C}{a_1 n}\Big) = \Phi\Big(\frac{\ln C}{a_1\sqrt{n}} - \frac{a_1\sqrt{n}}{2}\Big\}.$$

Снова обозначая через u_β корень уравнения $\Phi(u)=\beta$, получаем $u_\beta=\dfrac{\ln C}{a_1\sqrt{n}}-\dfrac{a_1\sqrt{n}}{2}$ и $u_\beta=u_{1-\alpha}-a_1\sqrt{n}$, откуда находим β. В случае, если n неизвестно, а значение β задано, имеем $\ln C=\dfrac{u_\alpha^2-u_\beta^2}{2}$ ($u_{1-\alpha}=-u_\alpha$) и можно найти приближенное значение n:

$$n=\frac{(u_\alpha+u_\beta)^2}{a_1^2}.$$

Замечание. Если альтернативная гипотеза H_1 сложная, т. е. представима как множество простых гипотез, то естественным образом вводится понятие равномерно наиболее мощного критерия, который является наиболее мощным для всех простых гипотез, составляющих H_1. Такой критерий, однако, не всегда существует. Например, в задаче проверки гипотезы о равенстве нулю среднего значения нормально распределенных случайных величин $H_0\colon a=0$ против сложной альтернативы $H_1\colon a>0$ равномерно наиболее мощным является критерий с критическим множеством, которое задается неравенством $\sqrt{n}\overline{X}>u_{1-\varepsilon}$, где $u_{1-\varepsilon}$ — корень уравнения $1-\Phi(u)=\varepsilon$; если же эта гипотеза проверяется при альтернативе $H_1\colon a\neq 0$, то равномерно наиболее мощного критерия не существует. По поводу проверки простых и сложных гипотез при сложных альтернативах, в частности по поводу условий существования равномерно наиболее мощных критериев, см. книгу [10].

Литература

1. *Бернштейн С. Н.* Курс теории вероятностей. М.: Гостехиздат, 1948.
2. *Бернштейн С. Н.* Сочинения. Т. 3. М.: АН СССР, 1960.
3. *Большев Л. Н.*, *Смирнов Н. В.* Таблицы математической статистики. М.: Наука, 1965.
4. *Боровков А. А.* Теория вероятностей. М.: Наука, 1976.
5. *Вальд А.* Последовательный анализ. М.: Гостехиздат, 1980.
6. *Ван дер Варден Б. Л.* Математическая статистика. М.: ИЛ, 1960.
7. *Гнеденко Б. В.* Курс теории вероятностей. М.: Наука, 1969.
8. *Дуб Дж. Л.* Вероятностные процессы. М.: ИЛ, 1965.
9. *Крамер Г.* Математические методы статистики. М.: Мир, 1975.
10. *Леман Э.* Проверка статистических гипотез. М.: Наука, 1964.
11. *Феллер В.* Теория вероятностей и ее приложения: В 2 т. М.: Мир, 1967.
12. *Севастьянов Б. А.* Теория вероятностей и математическая статистика. М.; Ижевск: ИКИ, 2004.