ИВМС

Л. З. РУМШИСКИЙ

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ИЗБРАННЫЕ ГЛАВЫ
ВЫСШЕЙ МАТЕМАТИКИ
ДЛЯ ИНЖЕНЕРОВ
И СТУДЕНТОВ ВТУЗОВ

Л. З. РУМШИСКИЙ

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

ИЗДАНИЕ ПЯТОЕ, ПЕРЕРАБОТАННОЕ

Допущено Министерством высшего и среднего специального образования СССР в качестве учебного пособия для студентов высших учебных заведений

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
Москва 1976

517. 8
Р 86
УДК 519.2

Элементы теории вероятностей, Л. З. Румшиский, изд. 5-е, перераб. Главная редакция физико-математической литературы изд-ва «Наука», 1976.

В книге излагаются в доступной форме понятия вероятности случайного события, распределения вероятностей случайных величин различных типов, даются их статистические толкования. Подробно рассмотрены отдельные законы распределений, важные для приложений, приведены примеры таких приложений. Много внимания уделено числовым характеристикам распределения, а также вопросам оценки этих характеристик. Все главы снабжены упражнениями для самостоятельного решения (с ответами и указаниями в конце книги).

Книга предназначается для студентов массовых инженерно-технических, технологических и инженерно-экономических специальностей в качестве учебного пособия по разделу «Основы теории вероятностей» курса высшей математики.

Помимо студентов, книга может быть полезна инженерам и экономистам, в особенности тем, кто интересуется вероятностными методами в связи с обработкой результатов эксперимента.

Р $\frac{20203-044}{053(02)-76}$ 20-76

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Еще сравнительно недавно применение математики в большинстве естественных и технических наук ограничивалось классическими разделами математики, такими, как дифференциальное исчисление, интегральное исчисление, дифференциальные уравнения. Вероятностные идеи и методы находили применение только в новых областях физики и в некоторых технических науках, например в радиотехнике (обнаружение сигнала на фоне случайных помех). Последние десятилетия существенно изменили положение дела. Вероятностные и статистические методы широко проникли в технические, технологические, экономические науки, главным образом, в связи с развитием массовых процессов в производстве и экономике, а также в связи с развитием экспериментальной техники и необходимостью более тонкого анализа результатов эксперимента.

Это не могло не сказаться и на преподавании математики в вузах. Если 20—30 лет тому назад вероятностные и статистические методы излагались студентам очень немногих специальностей, то в настоящее время во всех инженерно-технических, технологических, экономических и многих других вузах читается специальный раздел теории вероятностей и математической статистики объемом 50—60 часов, а в некоторых вузах и более. И если не всем инженерам в своей практической деятельности придется применять эти методы, то уж во всяком случае им необходимо знакомство с основными идеями и понятиями, по крайней мере, чтобы понимать соответствующие выводы.

Цель настоящей книги — в доступной форме познакомить широкий круг студентов и инженеров с основными понятиями теории вероятностей, без которых немыслимо математическое образование современного инженера. Изложение ориентировано на студентов массовых вузов, прослушавших общий втузовский курс высшей математики. Как отбор материала, так и форма изложения определяются прикладной направленностью курса: подробно рассмотрены те распределения вероятностей, которые чаще встречаются в приложениях (главным образом, в математической обработке результатов эксперимента), даны статистические толкования основных понятий, много внимания уделено числовым характеристикам распределения, в первую очередь, математическому ожиданию и дисперсии, в тексте разобраны типичные примеры наиболее простых приложений. Специально методы математической статистики в книге не рассматриваются, отдельные задачи статистического характера приведены лишь для иллюстрации приложения вероятностных идей.

Для удобства ссылок параграфы нумеруются по главам, а формулы нумеруются в каждом параграфе отдельно. При ссылке на формулу в том же параграфе дается только номер формулы, при ссылке на формулу из другого параграфа указывается и номер параграфа, например, (1.5-4) означает ссылку на формулу 4 из параграфа 5 главы 1.

В конце каждой главы приведено небольшое число упражнений различной трудности для самостоятельного решения; более трудные упражнения отмечены звездочкой; в конце книги даны ответы к упражнениям и указания к более трудным из них.

Автор пользуется возможностью выразить искреннюю благодарность профессору Е. С. Вентцель за ряд весьма ценных замечаний.

ГЛАВА 1

СЛУЧАЙНЫЕ СОБЫТИЯ И ВЕРОЯТНОСТИ

§ 1.1. Введение

Для изучения физических явлений производят наблюдения или опыты. Их результаты обычно регистрируют в виде значений некоторых наблюдаемых величин. При повторении опытов мы обнаруживаем разброс их результатов. Например, повторяя измерения одной и той же величины одним и тем же прибором при сохранении определенных условий (температура, влажность и т. п.), мы получаем результаты, которые хоть немного, но все же отличаются друг от друга. Даже многократные измерения не дают возможности точно предсказать результат следующего измерения. В этом смысле говорят, что результат измерения есть величина случайная. Еще более наглядным примером случайной величины может служить номер выигрышного билета в лотерее. Можно привести много других примеров случайных величин. Все же и в мире случайностей обнаруживаются определенные закономерности. Математический аппарат для изучения таких закономерностей дает теория вероятностей. В настоящей книге мы познакомимся с теми разделами теории вероятностей, которые относятся к случайным величинам и их статистическим применениям.

Для простоты мы будем пользоваться одним термином *испытание* для таких понятий, как опыт (эксперимент), наблюдение, измерение и т. п. Мы будем считать, что испытание можно повторять неограниченное число раз. С испытанием мы будем связывать одну или несколько случайных величин или же будем просто выделять отдельные возможные исходы испытания в качестве случайных событий. В следующем параграфе мы введем меру возможности появления того или иного случайного события — его вероятность. Здесь же, пользуясь пока лишь интуитивными представлениями, приведем несколько иллюстративных примеров.

Пример 1. Испытание состоит в бросании игральной кости (куба, на гранях которого отмечены числа от 1 до 6). Случайная величина X — число очков, которое выпадает на верхней грани игральной кости. Возможными значениями этой величины служат числа 1, 2, 3, 4, 5 и 6. Если игральная кость сделана из однородного материала и обладает симметрией («правильная игральная кость»), то интуиция подсказывает нам равную возможность выпадения любого числа очков от 1 до 6. Эта равная возможность проявляется на практике в том, что при большом количестве бросаний игральной кости каждое число очков от 1 до 6 выпадает приблизительно одинаково часто.

В качестве случайных событий можно взять как элементарные исходы испытания: $X = m$ (на верхней грани выпадает m очков») ($m = 1, 2, 3, 4, 5, 6$), так и более сложные, например $X = 2k$ («на верхней грани выпадает четное число очков»), или $X > 4$ («на верхней грани выпадает число очков, большее четырех»).

Пример 2. Испытание состоит в наблюдении за правильностью соединений абонентов на телефонной станции за определенный промежуток времени. В качестве случайной величины рассматривается число неправильных соединений. Возможными значениями этой величины будут целые числа от 0 до наибольшего допустимого числа соединений (за рассматриваемый промежуток времени, например за 1 минуту). В этом примере явно не все значения случайной величины будут равновозможны, но для выяснения большей или меньшей их возможности мало одной интуиции.

Обозначим рассматриваемую случайную величину буквой Y. В качестве случайного события можно взять $Y \leqslant 3$: «За рассматриваемый промежуток времени произойдет не более трех неправильных соединений».

Пример 3. Испытание состоит в измерении некоторой величины (например, во взвешивании тела на аналитических весах). Интересующие нас случайные величины — результат измерения X и ошибка измерения $Z = X - a$, где a — истинное значение измеряемой величины. В отличие от предыдущих примеров здесь возможные значения случайной величины сплошь заполняют некоторый интервал (например, вес тела не может быть отрицательным и заведомо ограничен сверху; поэтому возможный результат взвешивания лежит в определенных границах; разумеется, эти границы не всегда можно указать заранее). Сравнивать бóльшую или меньшую возможность тех или иных результатов здесь еще труднее, чем в предыдущем примере 2, но это бывает нужно для обработки результатов измерений; в дальнейшем мы отметим те закономерности, которые позволяют выполнять такую обработку.

Укажем и здесь несколько примеров случайных событий, которые нам встретятся в дальнейшем: $Z > 0$ («ошибка измерения положительна»), $|Z| \leqslant \varepsilon$ («абсолютная величина ошибки измерения не превосходит ε»), $Z \in (a, b)$ («ошибка измерения попадает в интервал (a, b)») *).

Пример 4. Испытание — стрельба по плоской мишени. В качестве случайной величины можно рассматривать радиус-вектор точки попадания. Здесь мы встречаемся с величинами, которые уже не являются скалярными, как в предыдущем примере. Если ввести декартовы координаты в плоскости мишени, то можно рассматривать систему случайных величин X и Y, где X — абсцисса точки попадания, а Y — ее ордината. Для круглой мишени может оказаться более удобным ввести полярные координаты R и Φ; тогда, например, такое случайное событие, как попадание «в яблочко», можно записать просто в виде $R < r_0$, где r_0 — радиус центральной части мишени («яблочка»).

§ 1.2. Случайные события, их относительная частота и вероятность

Основной числовой характеристикой случайного события является его вероятность, как мера его объективной возможности. О наличии определенной меры

*) $\in$ — знак принадлежности, $Z \in (a, b)$ означает, что значение величины Z принадлежит интервалу (a, b).

объективной возможности случайного события можно судить следующим образом. Будем много раз повторять испытание и каждый раз отмечать, произошло ли интересующее нас случайное событие A или нет. Обозначим через n общее число произведенных испытаний, через n_A — число появлений события A за n испытаний, и назовем отношение $w_A = n_A/n$ — *относительной частотой* *) *случайного события* A.

Оказывается, что в различных сериях испытаний относительная частота случайного события A принимает близкие значения: если при большом числе n испытаний относительная частота w_A была равна, например, 0,2, то и в любой другой серии из достаточного большого числа n' испытаний относительная частота $w'_A = \frac{n'_A}{n'}$ будет близка к числу 0,2. Например, при многократном бросании правильной игральной кости (см. пример 1, стр. 8) относительная частота выпадения каждого числа очков от 1 до 6 всегда близка к числу 1/6. Отмеченное свойство относительных частот называют статистической устойчивостью. Таким образом, относительные частоты случайного события A в различных сериях испытаний группируются около определенного числа; это число называется *вероятностью случайного события* A и обозначается символом $\mathsf{P}(A)$.

Подчеркнем, что статистическая устойчивость относительной частоты события A и, следовательно, наличие определенной его вероятности $\mathsf{P}(A)$ является основным, определяющим свойством случайного события; события, не обладающие этим свойством, в теории вероятностей не рассматриваются.

Различные способы вычисления вероятностей случайных событий будут подробно рассмотрены в дальнейшем. Но в простейших случаях мы можем приписать случайным событиям определенные вероятности, исходя непосредственно из указанных выше соображений. Например, при бросании правильной игральной кости выпадение каждого числа очков от 1 до 6 имеет одну и ту же вероятность, равную 1/6. Это можно коротко

*) Вместо термина «относительная частота» употребляют также более короткий термин «частота».

записать так:

$$\mathrm{P}(X=1)=\mathrm{P}(X=2)=\ldots=\mathrm{P}(X=6)=\frac{1}{6},$$

где X — число очков, выпадающее на верхней грани игральной кости.

Отметим, что вероятность случайного события, как и его относительная частота, является безразмерной величиной — отвлеченным числом. Отметим еще, что при каждом испытании (т. е. воспроизведении определенного комплекса условий, существенных для возможности появления некоторого события A) вероятность случайного события A следует считать неизменной. В то же время относительная частота случайного события A в разных сериях по n повторений испытания может принимать различные значения, так что сама эта относительная частота выступает в качестве случайной величины (подробнее об этом см. § 2.3). В настоящей главе под относительной частотой события A мы понимаем какое-либо ее значение n_A/n в определенной серии из n повторений испытания.

Как уже было подчеркнуто выше, в теории вероятностей мы рассматриваем только такие случайные события, которые обладают определенными вероятностями. Нас будут интересовать количественные соотношения между вероятностями различных случайных событий, в первую очередь, соотношения, связывающие вероятности более сложных событий с вероятностями более простых событий, например элементарных исходов испытания. Такие соотношения являются абстрактным выражением соответствующих свойств относительных частот. А именно, если в системе случайных событий их частоты связаны определенными соотношениями (тождествами, неравенствами), то такими же соотношениями должны быть связаны и вероятности этих событий.

Сформулировав основные соотношения подробного рода в виде определенных правил (см. §§ 1.3—1.5) и рассчитав по этим правилам вероятность какого-либо сложного события через вероятности более простых, мы вправе ожидать, что его относительная частота будет близка к рассчитанной вероятности, коль скоро это имеет место для исходных событий. Практическое значение этого замечания можно иллюстрировать, например, следующей важной задачей: зная вероятности элементарных исходов испытания, выделить такие, связанные с этим испытанием, случайные события, которые имеют очень малую вероятность и которые, следовательно, очень редко происходят (так что возможностью их появления можно пренебречь; подробнее об этом см. § 1.8).

§ 1.3. Основные свойства вероятностей, правило сложения вероятностей

Мы будем предполагать, что рассматриваемая система случайных событий представляет собой множество всех возможных исходов испытания и что в этой системе имеют смысл все вводимые далее операции над случайными событиями (например, переход к противоположному событию). Мы будем предполагать также, что все рассматриваемые события обладают определенными вероятностями (как указано в § 1.2).

Нормировка вероятности. Число n_A появлений любого случайного события A не может быть отрицательным ($n_A \geqslant 0$). С другой стороны, если испытание повторяется n раз, то событие A не может произойти более чем n раз ($n_A \leqslant n$). Поэтому относительная частота случайного события всегда удовлетворяет неравенству

$$0 \leqslant \frac{n_A}{n} \leqslant 1.$$

Отсюда естественно вытекает следующее свойство вероятностей.

Вероятность случайного события заключена между 0 и 1:

$$0 \leqslant \mathsf{P}(A) \leqslant 1 \tag{1}$$

Достоверные и невозможные события. В рассматриваемую систему случайных событий целесообразно включать достоверные события, которые непременно происходят при каждом испытании, и невозможные события, которые не могут произойти ни при каком испытании. Например, если X — число очков, которое выпадает на верхней грани игральной кости, то событие $X \leqslant 6$ будет достоверным, а событие $X > 6$ — невозможным.

Для достоверного события U всегда $n_U = n$, для невозможного события V всегда $n_V = 0$. Поэтому естественно принять следующие соглашения.

Вероятность достоверного события равна единице, вероятность невозможного события равна нулю:

$$\mathsf{P}(U) = 1, \quad \mathsf{P}(V) = 0. \tag{2}$$

Сумма событий. Пусть событие C состоит в осуществлении хотя бы одного из двух событий A или B (безразлично, какого именно, или обоих, если это возможно). Такое событие C назовем *суммой событий A и B* и будем обозначать так:

$$C = A + B.$$

Применяются также обозначения $C = A$ или B, $C = A \cup B$.

Например, если событие A — выпадение 1 на верхней грани игральной кости, а событие B — выпадение 2, то событие $A + B$ — выпадение не более двух очков: это можно коротко записать так:

$$(X = 1) + (X = 2) = (X \leqslant 2),$$

где X — число очков, выпадающее на верхней грани игральной кости.

Подчеркнем, что понятие суммы событий вводится только для исходов одного и того же испытания. Лишь при этом условии имеет смысл говорить о вероятности суммы событий.

Назовем далее случайные события A и B несовместимыми, если они не могут произойти вместе при одном и том же испытании.

Например, в приведенных выше обозначениях несовместимы события

$$X < 2{,}8 \quad \text{и} \quad X > 2{,}2.$$

Правило сложения вероятностей. *Вероятность суммы двух несовместимых случайных событий равна сумме вероятностей этих событий.*

Короче: для несовместимых случайных событий A и B имеет место соотношение

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B). \tag{3}$$

Правило сложения вероятностей называют также *свойством аддитивности вероятностей*. Для обоснования этого правила достаточно проверить, что относительные частоты всегда удовлетворяют соотношению, аналогичному (3).

Пусть производится некоторое испытание, и пусть A и B — два несовместимых исхода этого испытания. Для наглядности будем говорить, что испытание заключается в вынимании одного шара из урны, событие A — в появлении красного шара, событие B — в появлении синего шара; несовместимость этих событий означает здесь отсутствие в урне красно-синих шаров. Далее, будем считать, что в урне нет других цветных шаров, кроме красных и синих; тогда событие $C = A + B$ состоит в появлении цветного шара (безразлично, красного или синего). Вынув шар из урны, мы отметим его цвет и бросим шар обратно в урну, чтобы повторить испытание в тех же условиях. При n-кратном повторении испытания количество вынутых цветных шаров равно сумме количеств вынутых красных и синих шаров:

$$n_C = n_A + n_B.$$

Поэтому для относительных частот всегда выполняется соотношение

$$\frac{n_C}{n} = \frac{n_A}{n} + \frac{n_B}{n},$$

аналогичное соотношению (3).

Подчеркнем, что правило сложения вероятностей в форме (3) справедливо только для несовместимых событий. Если, например, два стрелка одновременно стреляют по одной цели, то событие C — хотя бы одно попадание в цель — равно сумме двух событий:

A — попадание первого стрелка в цель,

B — попадание второго стрелка в цель.

Но здесь соотношение (3) не имеет места, так как события A и B — совместимы: вполне возможно, что оба стрелка одновременно поразят цель; пусть $\mathsf{P}(A) = 0{,}8$ и $\mathsf{P}(B) = 0{,}9$, тогда $\mathsf{P}(A) + \mathsf{P}(B) = 1{,}7$, что не может вообще служить вероятностью какого-либо события (решение задачи о вероятности поражения цели см. на стр. 34).

Расширение правила сложения. Выше мы ввели правило сложения вероятностей только для двух случайных событий. Это правило по индукции легко распространяется на сумму любого конечного числа случайных событий: *если случайные события* $A_1, A_2, \ldots, A_k$ *попарно несовместимы, то вероятность их суммы, т. е. вероятность появления хотя бы одного из этих событий, равна сумме их вероятностей*:

$$\mathsf{P}(A_1 + A_2 + \ldots + A_k) = \mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_k). \quad (4)$$

Вероятности в полной группе событий. Мы будем говорить, что *случайные события образуют полную группу, если они попарно несовместимы и если при каждом повторении испытания должно произойти хотя бы одно из них.*

Примеры. Пусть X — число очков, выпадающее на верхней грани игральной кости. События

$$X = 1,\ X = 2,\ X = 3,\ X = 4,\ X = 5,\ X = 6 \quad (5)$$

образуют полную группу. Точно так же полную группу образуют и события

$$X \text{ четно}, \quad X \text{ нечетно}$$

или события

$$X \leqslant 2, \quad X \geqslant 3.$$

Из этих примеров видно, что из системы событий, связанных с данным испытанием, можно различным образом конструировать полные группы событий.

Теорема. *Если случайные события* $A_1, A_2, \ldots, A_n$ *образуют полную группу, то сумма их вероятностей равна единице.*

Доказательство. Так как события $A_1, A_2, \ldots, A_n$ образуют полную группу, то событие $A_1 + A_2 + \ldots + A_n$ является достоверным, поэтому $\mathsf{P}(A_1 + A_2 + \ldots + A_n) = 1$; в то же время попарная несовместимость рассматриваемых событий позволяет здесь применить формулу (4), что дает требуемое соотношение

$$\mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n) = 1. \qquad (6)$$

Полученное соотношение часто применяется для контроля расчета вероятностей, с чем мы неоднократно встретимся в дальнейшем.

Противоположные события. Два случайных события называются противоположными (или взаимно противоположными), если появление одного из них равносильно непоявлению другого. Если одно из этих событий обозначено через A, то другое (противоположное) обозначают через $\bar{A}$ (читают «не A»). Вероятности двух противоположных событий в сумме дают единицу:

$$\mathsf{P}(A) + \mathsf{P}(\bar{A}) = 1. \qquad (7)$$

Это непосредственно следует из формулы (6), так как противоположные события, очевидно, образуют полную группу. Полная группа из двух событий представляет особый интерес потому, что по вероятности одного из противоположных событий непосредственно вычисляется вероятность другого: если вероятность события A равна $\mathsf{P}(A) = p$, то вероятность q противоположного события $\bar{A}$ равна

$$\mathsf{P}(\bar{A}) = q = 1 - p.$$

В частности, если оба противоположных события имеют одинаковую вероятность, то эта вероятность равна 1/2. По этой причине считают, например, что при бросании правильной монеты вероятность выпадения герба равна 1/2. Разумеется, это утверждение опирается на невысказанное предположение о том, что вероятности выпадения и невыпадения герба при бросании монеты одинаковы. Проверить это предположение можно было бы статистическим путем, повторяя много раз бросание монеты. Заметим, что столь же правдоподобным может показаться и предположение о равной вероятности рождения мальчика и девочки. Но обширная статистика не подтверждает это предположение: в некоторых странах проводились соответствующие наблюдения и относительные частоты рождения мальчиков оказались устойчивыми, но были больше половины и колебались около чисел 0,51 или 0,52.

Введенные выше понятия суммы случайных событий, несовместимости событий и противоположных событий, а также и соотношения между вероятностями можно наглядно иллюстрировать следующим образом (рис. 1).

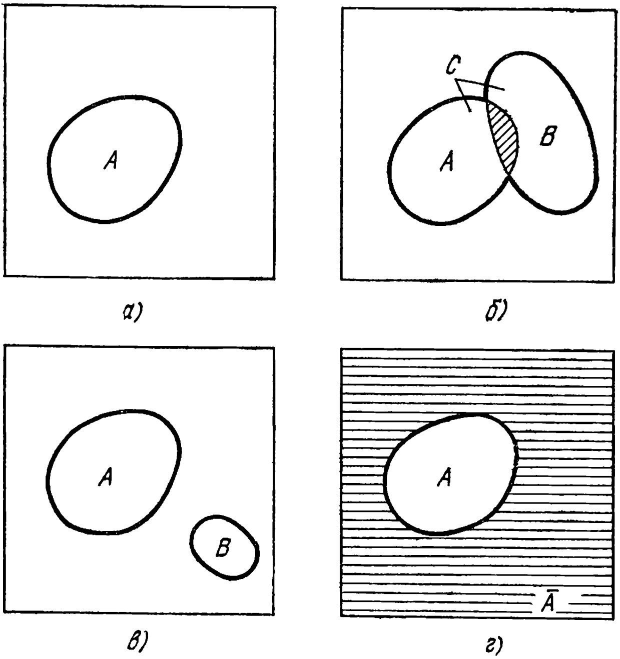


Рис. 1.

Будем толковать испытание как бросание точки в квадрат, а случайное событие A — как попадание точки в некоторую область (A) этого квадрата (рис. 1, *а*). Тогда рис. 1, *б* иллюстрирует понятие суммы $C = A + B$ (попадание в область, являющуюся объединением областей (A) и (B); ее граница проведена жирной линией), рис. 1, *в* — понятие несовместимости событий A и B (области (A) и (B) не имеют общих точек, поэтому никакая точка не может попасть одновременно в обе эти области), рис. 1, *г* — понятие противоположных событий (область $(\bar{A})$ — заштрихована).

При повторении испытания количество точек, попадающих в область $(A + B)$ на рис. 1, *в*, равно сумме количеств точек, попадающих в области (A) и (B), что иллюстрирует правило сложения вероятностей. В более общем случае (рис. 1, *б*) количество точек, попадающих в область $(A + B)$, может оказаться меньшим суммы количеств точек, попадающих в области (A) и (B) (за счет возможности попадания в их общую часть, подробнее об этом см. § 1.5); отсюда легко заключить, что в общем случае имеет место неравенство

$$\mathsf{P}(A + B) \leqslant \mathsf{P}(A) + \mathsf{P}(B). \tag{8}$$

Приведенная здесь иллюстрация становится особенно простой, если считать, что площадь всего квадрата равна 1, а вероятность попадания в некоторую область равна площади этой области. Для такой модели соотношения (3), (7) и (8) становятся очевидными.

§ 1.4. Вычисление вероятностей в классической модели (схема урн)

Классической моделью будем называть такую систему случайных событий, в которой каждое событие можно представить как сумму нескольких событий, называемых *элементарными событиями* (или *случаями*), причем

> элементарные события образуют полную группу и все элементарные события *равновозможны*.

Равная возможность элементарных событий на практике либо устанавливается по соображениям симметрии, когда нет оснований считать какой-либо случай более вероятным, чем любой другой, либо обеспечивается специальной организацией испытаний. Например, при бросании правильной игральной кости события (1.3-5) считаются равновозможными по соображениям симметрии; а при организации лотерей, случайных выборок и т. п. принимаются специальные меры, чтобы каждый объект или номер имел одинаковую с остальными вероятность быть выбранным.

Из равной возможности элементарных событий следует, что их число конечно (если бы множество элементарных событий в классической модели было бесконечным, то и сумма их вероятностей была бы бесконечной, что противоречит нормировке вероятностей).

Классическую модель называют также схемой урн, так как ее можно представить следующим образом. В урне находится N одинаковых на ощупь шаров, занумерованных числами от 1 до N. Испытание заключается в вынимании наугад одного шара. Случаем считается появление шара с заданным номером k $(k = 1, 2, \ldots, N)$. Если ввести случайную величину X — номер вынутого шара, то каждый случай можно записать в виде $X = k$. Равная вероятность этих случаев практически обеспечивается тем, что шары до вынимания тщательно перемешивают. Любое сложное событие A в этой схеме заключается в том, что номер вынутого шара принадлежит некоторому заданному множеству номеров. Если это множество содержит M номеров из набора $1, 2, \ldots, N$, то говорят, что «событию A благоприятствуют M случаев».

Т е о р е м а. *В классической модели вероятность любого события A равна отношению числа M случаев, благоприятствующих этому событию, к общему числу N всех случаев:*

$$\boxed{\mathsf{P}(A) = \frac{M}{N}.} \tag{1}$$

Д о к а з а т е л ь с т в о. По условию вероятность любого случая $X = k$ одна и та же, обозначим ее p:

$$\mathsf{P}(X = k) = p \qquad (k = 1, 2, \ldots, N).$$

Так как эти случаи образуют полную группу, то по формуле (1.3-6) имеем

$$N \cdot p = 1,$$

откуда $p = 1/N$. Далее, пусть для определенности, событию A благоприятствуют первые M случаев $(M \leqslant N)$.

Тогда по правилу сложения вероятностей получим

$$\mathsf{P}(A) = \mathsf{P}(X = 1) + \mathsf{P}(X = 2) + \ldots + \mathsf{P}(X = M) = \\ = M \cdot p = \frac{M}{N}.$$

Так как эта вероятность не зависит от того, какие именно случаи благоприятствуют событию A, то полученная формула сохраняется при любом выборе M случаев, благоприятствующих событию A, что и завершает доказательство.

Пример 1. Вычислить вероятность выпадения четного числа очков при бросании правильной игральной кости.

Решение. Здесь естественно строится классическая модель: случаями являются события (1.3-5). Число всех этих случаев равно $N = 6$. Число благоприятствующих случаев ($X = 2$, $X = 4$, $X = 6$) равно $M = 3$. Поэтому искомая вероятность равна $3/6 = 1/2$.

Пример 2. При бросании двух монет герб может выпасть 2 раза, 1 раз или 0 раз (не выпасть ни разу); вычислить вероятности этих трех случайных событий в предположении, что выпадение или невыпадение герба на каждой монете равновозможны.

Решение. Случайные события, перечисленные в условии задачи, очевидно, образуют полную группу. Но они не являются равновозможными и поэтому не образуют систему случаев. Для построения классической модели выделим элементарные исходы испытания, комбинируя каждый случай выпадения или невыпадения герба на первой монете со случаями выпадения или невыпадения герба на второй монете:

$$\text{ГГ},\quad \text{Г}\overline{\text{Г}},\quad \overline{\text{Г}}\text{Г},\quad \overline{\text{Г}}\overline{\text{Г}}.$$

По этим исходам испытания можно составить таблицу (см. стр. 21), в которой явно указано число выпавших гербов на двух монетах.

Перечисленные в этой таблице четыре исхода испытания естественно считать равновозможными по соображениям симметрии. И эти исходы снова составляют полную группу. Обозначая через X число выпадений

	Число выпадений герба			
Первая монета Вторая монета	1 1	1 0	0 1	0 0
Всего на двух монетах	2	1	1	0

герба на двух монетах и применяя формулу (1), получим

$$\mathsf{P}(X=2)=\frac{1}{4}, \quad \mathsf{P}(X=1)=\frac{2}{4}=\frac{1}{2}, \quad \mathsf{P}(X=0)=\frac{1}{4}.$$

Пример 3. Из полной колоды карт (52 карты) вынимаются наугад три карты (без возврата). Вычислить вероятность того, что среди вынутых карт будет точно один туз. Слово «наугад» надо понимать здесь так, что взятие любой тройки карт равновозможно.

Решение. Будем считать случаем взятие любой тройки карт. По условию эти случаи равновозможны. Для подсчета числа всех случаев и числа благоприятствующих случаев у нас есть два способа.

Первый способ. Будем различать тройки карт по порядку взятия карт, т. е. считать, например, тройку «туз, король, дама», отличной от тройки «король, туз, дама». Тогда число всех таких троек карт равно $N = A_{52}^3 = 52\cdot51\cdot50$. Из всех таких троек карт интересующие нас тройки содержат один туз и две другие карты; туз можно выбрать одним из четырех способов, а остальные две карты — одним из $48\cdot47$ способов. Если еще учесть, что тузом может быть первая, вторая или третья из выбранных карт, то число благоприятствующих случаев окажется равным $M = 3\cdot4\cdot48\cdot47$. Поэтому искомая вероятность равна

$$\mathsf{P}\text{ (точно один туз)} = \frac{3\cdot4\cdot48\cdot47}{52\cdot51\cdot50} = \frac{1128}{5525} = 0{,}204.$$

Второй способ. Будем считать различными тройки, которые различаются хотя бы одной картой, вне зависимости от порядка взятия карт. Тогда общее число всех возможных случаев будет равно числу сочетаний $N = C_{52}^3 = \frac{52 \cdot 51 \cdot 50}{1 \cdot 2 \cdot 3}$. Благоприятствуют нашему событию те и только те случаи, когда один из четырех тузов вынимается вместе с двумя из остальных 48 карт (снова независимо от порядка). Таких случаев будет

$$M = C_4^1 \cdot C_{48}^2 = 4 \cdot \frac{48 \cdot 47}{1 \cdot 2}.$$

Искомая вероятность оказывается равной

$$\mathsf{P}\ (\text{точно один туз}) = \frac{C_4^1 \cdot C_{48}^2}{C_{52}^3} = \frac{4 \cdot 48 \cdot 47 \cdot 3}{52 \cdot 51 \cdot 50} = 0{,}204,$$

что, разумеется, совпадает с ранее полученным числом.

Пример 4. В условиях предыдущего примера вычислить вероятность того, что среди вынутых карт будет хотя бы один туз.

Решение. Событие «хотя бы один туз» означает, что число вынутых тузов будет не менее 1. Эту задачу можно решить по правилу сложения: обозначив через X число вынутых тузов, имеем

$$\mathsf{P}\ (X \geqslant 1) = \mathsf{P}\ (X = 1) + \mathsf{P}\ (X = 2) + \mathsf{P}\ (X = 3).$$

Проще решить эту задачу, используя понятие противоположного события и формулу (1.3-7). Так как случайная величина X может принимать только одно из значений 0, 1, 2, 3, то события $(X \geqslant 1)$ и $(X = 0)$ будут противоположными и поэтому

$$\mathsf{P}\ (X \geqslant 1) = 1 - \mathsf{P}\ (X = 0).$$

Но вероятность того, что число вынутых тузов будет равно 0, т. е. того, что среди вынутых карт не будет ни одного туза, очевидно, равна

$$\mathsf{P}(X = 0) = \frac{C_{48}^3}{C_{52}^3} = \frac{48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot 50} = \frac{4324}{5525} = 0{,}783.$$

Следовательно, искомая вероятность равна

$$\mathsf{P}(X \geqslant 1) = 1 - \frac{4324}{5525} = \frac{1201}{5525} = 0{,}217.$$

Примечание В примерах 3 и 4 мы встретились с некоторыми *формулами комбинаторики*, которые часто оказываются полезными при вычислении вероятностей. Приведем для справки некоторые общие формулы этого типа.

Размещениями из n элементов по m называются такие соединения, которые различаются самими элементами или их порядком (например, размещения из трех элементов a, b, c по 2: ab, ac, bc, ba, ca, cb). Число всех размещений из n элементов по m равно

$$A_n^m = n(n-1)\dots(n-m+1) = \frac{n!}{(n-m)!}.$$

Перестановками из n элементов называются их соединения, различающиеся только порядком входящих в них элементов. Число всех перестановок из n различных элементов равно

$$A_n^n = n! = 1\cdot 2\cdot 3 \dots n.$$

Если среди n элементов a, b, c, ... имеются одинаковые (a повторяется α раз, b — β раз, c — γ раз, и т. д.), то число перестановок равно $\dfrac{n!}{\alpha!\,\beta!\,\gamma!\dots}$.

Сочетаниями из n элементов по m называются такие их соединения, которые различаются только своими элементами (например, сочетания из трех элементов a, b, c по 2: ab, bc, ca). Число сочетаний из n элементов по m равно

$$C_n^m = \frac{n!}{m!\,(n-m)!} = \frac{n(n-1)\dots(n-m+1)}{1\cdot 2\dots m} = C_n^{n-m}.$$

Последние числа равны коэффициентам разложения бинома $(1+x)^n$ по степеням x и называются *биномиальными коэффициентами*; другое их обозначение $C_n^m = \binom{n}{m}$. Для удобства полагают еще $C_n^0 = C_n^n = 1$. Для биномиальных коэффициентов имеются таблицы во многих математических справочниках (см., например, Б. И. Сегал и К. А. Семендяев «Пятизначные математические таблицы»).

§ 1.5. Правило умножения вероятностей и условные вероятности

Для дальнейшего изложения нам понадобится еще понятие совмещения случайных событий. Под совмещением случайных событий A и B понимают случайное событие, заключающееся в том, что в результате испытания произойдет и событие A и событие B. Это совмещение обозначим через $A\cdot B$ (другое обозначение (A и B); другое название — произведение случайных событий *).

Пусть, например, при игре в лото событие A состоит в вынимании бочонка с четным номером, а событие B — в вынимании бочонка с номером, кратным трем; тогда событие $A \cdot B$ состоит в вынимании бочонка с номером, кратным шести.

Нас интересует зависимость между вероятностями событий A и B и вероятностью их совмещения $A \cdot B$. Как и прежде, мы обратимся сначала к относительным частотам и попробуем установить зависимость между частотами рассматриваемых событий.

Для удобства описания будем представлять себе испытание как вынимание наугад шара из урны. Пусть нас интересуют два признака: цвет шара и наличие на нем рисунка (звездочки). Событие A состоит в том, что вынутый шар белый, событие B — в том, что он помечен звездочкой; тогда событие $A \cdot B$ состоит в том, что вынутый шар — белый со звездочкой. Будем повторять испытания, возвращая каждый раз вынутый шар в урну, чтобы не изменить вероятности рассматриваемых событий. После n-кратного повторения испытания мы получим какие-то относительные частоты n_A/n, n_B/n и n_{AB}/n событий A, B и $A \cdot B$ соответственно. Очевидна справедливость следующего тождества:

$$\frac{n_{A \cdot B}}{n} = \frac{n_A}{n} \cdot \frac{n_{A \cdot B}}{n_A} \tag{1}$$

Отношение n_{AB}/n_A есть доля шаров со звездочкой среди вынутых белых шаров. Это отношение можно рассматривать тоже как относительную частоту, а именно, относительную частоту появления шаров со звездочкой при дополнительном условии, что регистрируются лишь испытания, дающие белые шары. По-

*) Легко показать, что события $A \cdot B$ и $\bar{A} + \bar{B}$ противоположны (совмещение событий A и B не произойдет тогда и только тогда, когда не произойдет хотя бы одно из событий A или B; см. также пример на стр. 34). Поэтому для системы из двух событий, а значит, и для системы из любого конечного их числа введение понятия совмещения событий не требует новых предположений сверх тех, которые были высказаны в § 1.3.

лученное соотношение (1) позволяет сформулировать соответствующее правило умножения для вероятностей. Относительные частоты $n_{A\cdot B}/n$ и n_A/n заменим на вероятности $\mathsf{P}(A\cdot B)$ и $\mathsf{P}(A)$, а отношение $n_{A\cdot B}/n_A$ — на величину, которая называется у с л о в н о й в е р о я т н о с т ь ю с о б ы т и я B п р и у с л о в и и о с у щ е с т в л е н и я с о б ы т и я A и обозначается через $\mathsf{P}(B\,|\,A)$. Мы получим

$$\boxed{\mathsf{P}(A\cdot B)=\mathsf{P}(A)\cdot\mathsf{P}(B\,|\,A).} \qquad (2)$$

Это — п р а в и л о у м н о ж е н и я в е р о я т н о с т е й. Оно читается так: *вероятность совмещения двух случайных событий равна произведению вероятности одного из них на условную вероятность другого.*

Правило умножения вероятностей (2) можно легко распространить и на большее число случайных событий. Например, для трех событий A, B, C имеем

$$\mathsf{P}(A\cdot B\cdot C)=\mathsf{P}(A\cdot B)\,\mathsf{P}(C\,|\,A\cdot B)=$$
$$=\mathsf{P}(A)\;\mathsf{P}(B\,|\,A)\;\mathsf{P}(C\,|\,A\cdot B). \qquad (3)$$

П р и м е р. Из полной колоды карт (52 карты) вынимаются наугад три карты (без возврата). Вычислить вероятность того, что среди них не будет ни одного туза.

Р е ш е н и е. Обозначим через A_i событие, состоящее в том, что i-я вынутая карта не туз ($i=1,2,3$). Тогда интересующее нас событие можно представить как совмещение трех событий A_1, A_2, A_3. По правилу умножения имеем

$$\mathsf{P}(A_1\cdot A_2\cdot A_3)=\mathsf{P}(A_1)\,\mathsf{P}(A_2\,|\,A_1)\mathsf{P}(A_3\,|\,A_1\cdot A_2).$$

Так как всех карт 52, а тузов из них 4, то при вынимании первой карты мы можем воспользоваться схемой урн, считая $N=52$, $M=52-4=48$. Поэтому вероятность того, что первая вынутая карта не будет тузом, равна $\mathsf{P}(A_1)=\frac{48}{52}$.

При условии, что первая карта не туз, мы снова можем воспользоваться схемой урн, считая, что всех карт осталось 51, из них не тузов 51—4 =47; поэтому условная вероятность того, что вторая карта не будет тузом при условии, что первая карта не туз, равна

$$\mathsf{P}(A_2 \mid A_1) = \frac{47}{51}.$$

Аналогично,

$$\mathsf{P}(A_3 \mid A_1 \cdot A_2) = \frac{46}{50}.$$

Таким образом, искомая вероятность равна

$$\mathsf{P}(A_1 \cdot A_2 \cdot A_3) = \frac{48}{52} \cdot \frac{47}{51} \cdot \frac{46}{50} = \frac{4324}{5525} = 0{,}783.$$

Мы получаем здесь тот же результат, что и на стр. 22, но значительно более простым путем.

В рассмотренном примере вычисление условных вероятностей не составило труда, что позволило нам просто подсчитать вероятность совмещения случайных событий. Такая ситуация характерна для многих практических задач. В теоретических построениях иногда удобнее считать заданными как раз вероятности совмещения случайных событий. В этих ситуациях правило умножения может служить для определения условных вероятностей. Так, из формулы (2) при условии $\mathsf{P}(A) \neq 0$ мы получим

$$\mathsf{P}(B \mid A) = \frac{\mathsf{P}(A \cdot B)}{\mathsf{P}(A)}. \tag{4}$$

Аналогично определяется условная вероятность события A при условии осуществления события B:

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cdot B)}{\mathsf{P}(B)}, \quad \text{если} \quad \mathsf{P}(B) \neq 0.$$

В заключение этого параграфа приведем формулу, связывающую вероятность совмещения и вероятность суммы двух событий A и B, которые уже могут быть и совместимыми:

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cdot B). \tag{5}$$

Не приводя формального доказательства этого общего соотношения, мы ограничимся лишь его геометрической иллюстрацией на рис. 1, *б* (стр. 17). Напомним, что мы толковали там испытание как бросание точки на квадрат, а случайное событие A как попадание точки в некоторую область (A). Тогда совмещение случайных событий $A \cdot B$ можно толковать как попадание точки в общую часть областей (A) и (B) (на рис. 1, *б* эта общая часть заштрихована). Сумму событий $C = A + B$ при этом следует толковать как попадание в объединение областей (A) и (B) (на рис. 1, *б* эта область ограничена жирной линией).

При повторении испытания количество точек, попадающих в область (C), равно сумме количеств точек, попадающих в области (A) и (B), за вычетом количества точек, попадающих в их общую часть (так как эти последние точки при подсчете надо учитывать только один раз). Поэтому для относительных частот рассматриваемых событий мы всегда получаем соотношение того же типа, что и (5).

§ 1.6. Формула полной вероятности и формулы Байеса

Рассмотренные выше правила сложения и умножения вероятностей представляют собой основные правила, служащие для расчета вероятностей случайных событий через вероятности более простых событий. Во многих задачах оказывается полезным одно следствие из этих правил, которое получило название формулы полной вероятности. Пусть случайное событие A может произойти только с одним из событий $H_1, H_2, \ldots, H_n$, причем события $H_1, H_2, \ldots, H_n$ попарно несовместимы. Тогда вероятность события A может быть подсчитана по формуле полной вероятности

$$\mathsf{P}(A) = \mathsf{P}(H_1)\mathsf{P}(A \mid H_1) + \mathsf{P}(H_2)\mathsf{P}(A \mid H_2) + \ldots + \mathsf{P}(H_n)\mathsf{P}(A \mid H_n). \tag{1}$$

Для вывода этой формулы заметим прежде всего, что события $A \cdot H_1, A \cdot H_2, \ldots, A \cdot H_n$ тоже попарно несовместимы и что случайное событие A представляет собой их сумму. Поэтому к событию $A = A \cdot H_1 + A \cdot H_2 + \ldots + A \cdot H_n$ можно применить правило сложения вероятностей, что дает

$$\mathsf{P}(A) = \mathsf{P}(A \cdot H_1) + \mathsf{P}(A \cdot H_2) + \ldots + \mathsf{P}(A \cdot H_n).$$

Остается к каждому из слагаемых в последней формуле применить правило умножения вероятностей в виде

$$\mathsf{P}(A \cdot H_k) = \mathsf{P}(H_k)\,\mathsf{P}(A \mid H_k) \qquad (k = 1, 2, \ldots, n),$$

что приводит к формуле (1).

События $H_1, H_2, \ldots, H_n$, входящие в эту формулу, обычно называют *гипотезами*. Смысл этого названия мы поясним на примере.

Пример. Урны с шарами находятся в четырех комнатах, куда ведут ходы лабиринта, изображенного на рис. 2. Вошедший в лабиринт человек выбирает наугад один из возможных путей (как при входе, так и на развилке), доходит до комнаты и вынимает наугад один шар из урны.

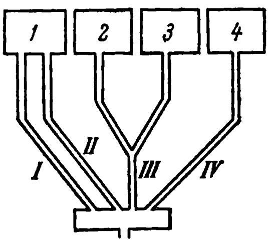


Рис. 2.

Вычислить вероятность того, что вынутый шар будет белым, если в первых трех урнах доли белых шаров составляют соответственно p_1, p_2, p_3, а в четвертой урне белых шаров нет.

Решение. *Первый метод.* В качестве гипотез введем выбор хода I, II, III или IV. Эти гипотезы равновероятны по условию, поэтому

$$\mathsf{P}(H_{\mathrm{I}}) = \mathsf{P}(H_{\mathrm{II}}) = \mathsf{P}(H_{\mathrm{III}}) = \mathsf{P}(H_{\mathrm{IV}}) = \frac{1}{4}.$$

Через A обозначим интересующее нас событие — вынимание белого шара. Тогда

$$\mathsf{P}(A) = \frac{1}{4}\mathsf{P}(A \mid H_{\mathrm{I}}) + \frac{1}{4}\mathsf{P}(A \mid H_{\mathrm{II}}) + \frac{1}{4}\mathsf{P}(A \mid H_{\mathrm{III}}) + \frac{1}{4}\mathsf{P}(A \mid H_{\mathrm{IV}}),$$

где $\mathsf{P}(A \mid H_{\mathrm{I}}) = \mathsf{P}(A \mid H_{\mathrm{II}}) = p_1$, $\mathsf{P}(A \mid H_{\mathrm{IV}}) = 0$. Для подсчета $\mathsf{P}(A \mid H_{\mathrm{III}})$ надо ввести еще две гипотезы о выборе пути после развилки и снова применить формулу полной вероятности, что дает

$$\mathsf{P}(A \mid H_{\mathrm{III}}) = \frac{1}{2} p_2 + \frac{1}{2} p_3.$$

Итак,

$$\mathsf{P}(A) = \frac{1}{4} p_1 + \frac{1}{4} p_1 + \frac{1}{4}\left(\frac{1}{2} p_2 + \frac{1}{2} p_3\right) = \\ = \frac{1}{2} p_1 + \frac{1}{8} p_2 + \frac{1}{8} p_3.$$

Второй метод. В качестве гипотез введем попадание в одну из комнат 1, 2, 3 или 4. Тогда по условию

$$\mathsf{P}(A \mid H_1) = p_1, \quad \mathsf{P}(A \mid H_2) = p_2, \quad \mathsf{P}(A \mid H_3) = p_3, \\ \mathsf{P}(A \mid H_4) = 0.$$

Остается найти вероятности введенных гипотез H_1, H_2, H_3, H_4. Непосредственный подсчет по правилам сложения и умножения дает

$$\mathsf{P}(H_1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \quad \mathsf{P}(H_2) = \mathsf{P}(H_3) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}, \\ \mathsf{P}(H_4) = \frac{1}{4}.$$

Итак,

$$\mathsf{P}(A) = \frac{1}{2} p_1 + \frac{1}{8} p_2 + \frac{1}{8} p_3.$$

Заметим, что ни вероятность гипотезы H_{IV}, ни вероятность гипотезы H_4 в ответ не входят, так как при этих гипотезах событие A не может произойти. Поэтому эти гипотезы можно было бы и не учитывать в формуле полной вероятности. Нам эти гипотезы понадобились только для подсчета вероятностей остальных гипотез по классической формуле (для построения полной группы случаев). Вообще же в формуле полной вероятности достаточно учитывать лишь те гипотезы, при которых событие A может осуществиться.

Выделим важный частный случай формулы полной вероятности. Противоположные гипотезы H и $\bar{H}$ всегда образуют полную группу. Поэтому всегда имеет место формула

$$\boxed{\mathsf{P}(A) = \mathsf{P}(H)\,\mathsf{P}(A \mid H) + \mathsf{P}(\bar{H})\,\mathsf{P}(A \mid \bar{H}).} \qquad (2)$$

Проиллюстрируем и эту формулу примером.

Пример. В урне было N шаров, из них M белых. Один шар укатился, цвет его неизвестен. Из оставшихся шаров наугад вынимается один шар. Вычислить вероятность того, что он будет белым.

Решение. Введем гипотезу H: утерянный шар — белый. Так как здесь применима схема урн, то вероятности гипотез равны соответственно

$$\mathsf{P}(H)=\frac{M}{N},\quad \mathsf{P}(\bar{H})=\frac{N-M}{N}.$$

Для условных вероятностей снова применима схема урн (с уменьшенными на единицу количеством всех шаров и количествами белых или небелых шаров в урне). Поэтому

$$\mathsf{P}(A\mid H)=\frac{M-1}{N-1},\quad \mathsf{P}(A\mid\bar{H})=\frac{M}{N-1}.$$

По формуле (2) находим искомую вероятность

$$\mathsf{P}(A)=\frac{M}{N}\cdot\frac{M-1}{N-1}+\frac{N-M}{N}\cdot\frac{M}{N-1}=\frac{M}{N},$$

которая оказывается такой же, как и вероятность вынуть белый шар из исходной урны (если бы шар не укатывался). Этот результат можно толковать еще и следующим образом. Из урны наугад один за другим вынимают шары; тогда, если неизвестен цвет шара, вынутого первым, то вероятность того, что второй шар будет белым, равна вероятности вынуть первым белый шар. Очевидно, что аналогичное утверждение будет верным и для последующих шаров.

Одно из интересных применений формулы полной вероятности связано с формулами Байеса. Если в выражении для условной вероятности

$$\mathsf{P}(H_1\mid A)=\frac{\mathsf{P}(A\cdot H_1)}{\mathsf{P}(A)}=\frac{\mathsf{P}(H_1)\,\mathsf{P}(A\mid H_1)}{\mathsf{P}(A)}$$

заменить вероятность $\mathsf{P}(A)$ по формуле полной вероятности (1), то мы получим формулу Байеса

$$\mathsf{P}(H_1\mid A)=\frac{\mathsf{P}(H_1)\,\mathsf{P}(A\mid H_1)}{\mathsf{P}(H_1)\,\mathsf{P}(A\mid H_1)+\ldots+\mathsf{P}(H_n)\,\mathsf{P}(A\mid H_n)}. \qquad (3)$$

Эта формула применяется для вычисления условной вероятности

$\mathsf{P}(H_1 \mid A)$ гипотезы H_1 после испытания, при котором произошло событие A. Аналогичные формулы имеют место для условных вероятностей $\mathsf{P}(H_k \mid A)$ всех гипотез H_k. Другими словами, формулы Байеса позволяют переоценить вероятности гипотез, принятые до испытания (*априорные*), по результатам уже произведенного испытания.

Пример. Пусть в условиях примера на стр. 28 известно, что вошедший человек вынул белый шар. Вычислить условные вероятности гипотез H_{I}, H_{II}, H_{III}, H_{IV} при этом дополнительном условии (т. е. вероятности того, что он пошел по одному из ходов I, II, III и IV, если он пришел с белым шаром).

Решение. По формулам Байеса находим

$$\mathsf{P}(H_{\mathrm{I}} \mid A) = \mathsf{P}(H_{\mathrm{II}} \mid A) = \frac{\frac{1}{4} p_1}{\frac{1}{2} p_1 + \frac{1}{8} p_2 + \frac{1}{8} p_3} = \frac{2p_1}{4p_1 + p_2 + p_3},$$

$$\mathsf{P}(H_{\mathrm{III}} \mid A) = \frac{\frac{1}{4}\left(\frac{1}{2} p_2 + \frac{1}{2} p_3\right)}{\frac{1}{2} p_1 + \frac{1}{8} p_2 + \frac{1}{8} p_3} = \frac{p_2 + p_3}{4p_1 + p_2 + p_3},$$

$$\mathsf{P}(H_{\mathrm{IV}} \mid A) = 0.$$

Таким образом, если до испытания вероятности всех гипотез H_{I}, H_{II}, H_{III}, H_{IV} были одинаковыми (ход выбирался наугад), то после испытания условные вероятности этих гипотез существенно зависят от распределения белых шаров по урнам. Если, например, $p_1 = p_2 = 0{,}1$, а $p_3 = 0{,}5$ то

$$\mathsf{P}(H_{\mathrm{I}} \mid A) = \mathsf{P}(H_{\mathrm{II}} \mid A) = 0{,}2, \text{ а } \mathsf{P}(H_{\mathrm{III}} \mid A) = 0{,}6;$$

другими словами, если при указанном распределении шаров в урнах человек пришел с белым шаром, то вероятность того, что он шел по ходу III, оказывается втрое большей, чем вероятность того, что он шел по ходу I или ходу II.

§ 1.7. Независимость случайных событий

Для одного важного класса задач правило умножения вероятностей позволяет выражать вероятность совмещения случайных событий непосредственно через вероятности самих этих событий. Этот класс задач связан с понятием независимости случайных событий.

Рассмотрим сначала простой пример для схемы урн. Пусть имеются две урны, в которых общее число шаров равно соответственно N_1 и N_2, а число белых шаров — M_1 и M_2. Испытание заключается в том, что из каждой урны наугад вынимается по одному шару. В качестве случайных событий A и B рассмотрим появление белого шара соответственно из первой или второй урны. Эти случайные события независимы в том смысле, что цвет шара, вынутого из одной урны, не может повлиять на цвет шара, вынутого из другой урны. Посмотрим, как связаны здесь вероятности случайных событий A и B с вероятностью их совмещения $A \cdot B$. Прежде всего по основной формуле (1.4-1) для схемы урн имеем

$$\mathsf{P}(A) = \frac{M_1}{N_1}, \quad \mathsf{P}(B) = \frac{M_2}{N_2}.$$

Для подсчета вероятности совмещения событий A и B, т. е. вероятности того, что оба вынутых шара окажутся белыми, нам надо построить новую схему урн, правильно перечислив все возможные исходы испытания. Каждый из N_1 результатов вынимания шара из первой урны может комбинироваться с каждым из N_2 результатов вынимания шара из второй урны, причем все такие комбинации будут равновозможными по соображениям симметрии. Поэтому все такие комбинации образуют систему случаев и число всех случаев равно $N = N_1 \cdot N_2$. По тем же соображениям число благоприятных случаев (вынимание двух белых шаров) будет равно $M = M_1 \cdot M_2$. Поэтому снова применима формула (1.4-1) и искомая вероятность совмещения $A \cdot B$ равна

$$\mathsf{P}(A \cdot B) = \frac{M_1 M_2}{N_1 N_2} = \frac{M_1}{N_1} \cdot \frac{M_2}{N_2}.$$

При этом оказывается, что

$$\mathsf{P}(A \cdot B) = \mathsf{P}(A)\,\mathsf{P}(B), \tag{1}$$

т. е. что вероятность совмещения событий A и B равна произведению их вероятностей. Полученное соотношение оказывается справедливым каждый раз, когда рассматриваемые случайные события никак не могут повлиять друг на друга по самому смыслу задачи. Это оправдывает следующее определение независимости двух случайных событий.

Определение 1. *Два случайных события A и B называются независимыми* (или *статистически независимыми*), *если вероятность их совмещения равна произведению их вероятностей*, т. е. если для них справедлива формула (1).

Другой подход к понятию независимости двух случайных событий связан с понятием условной вероятности. Сравнивая формулу (1) с формулой (1.5-2) или (1.5-4), мы замечаем, что формула (1) равносильна следующим формулам:

$$\left.\begin{aligned} \mathsf{P}(B\,|\,A) &= \frac{\mathsf{P}(A\cdot B)}{\mathsf{P}(A)} = \mathsf{P}(B), \\ \mathsf{P}(A\,|\,B) &= \frac{\mathsf{P}(A\cdot B)}{\mathsf{P}(B)} = \mathsf{P}(A). \end{aligned}\right\} \qquad (2)$$

Это значит, что *случайные события A и B независимы тогда и только тогда, когда условные вероятности этих событий совпадают с их безусловными вероятностями.*

Следующий пример иллюстрирует различие между зависимыми и независимыми случайными событиями. Пусть шары в урне распределены по двум признакам: по цвету (белые и цветные) и по рисунку (со звездочкой и без звездочки), причем среди цветных шаров встречаются как шары со звездочкой, так и без нее (на рис. 3 цветные шары отмечены штриховкой). Испытание состоит в вынимании из урны одного шара наугад, событие A — в появлении белого шара, событие B — в появлении шара со звездочкой. Пусть на 20 шаров приходится точно 10 белых шаров и 8 шаров со звездочкой, так что

$$\mathsf{P}(A) = \frac{10}{20} = 0{,}5, \qquad \mathsf{P}(B) = \frac{8}{20} = 0{,}4.$$

На рис. 3 схематически представлены два возможных распределения шаров по двум признакам:

а) пропорциональное распределение, при котором доля шаров со звездочкой среди белых шаров совпадает с долей шаров со звездочкой среди всех шаров в урне;

б) некоторое непропорциональное распределение.

В первом случае

$$P(B|A) = \frac{4}{10} = 0{,}4 = P(B),$$

так что дополнительная информация о цвете шара не меняет вероятности появления шара со звездочкой.

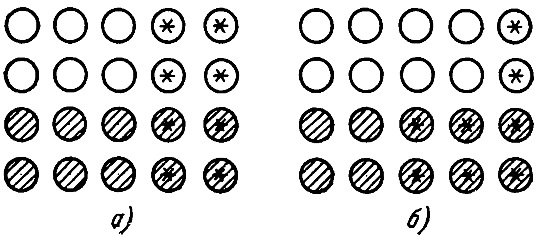


Рис. 3.

Здесь события A и B независимы. Во втором случае

$$P(B|A) = \frac{2}{10} = 0{,}2 \neq P(B),$$

так что дополнительная информация о цвете шара меняет вероятность появления шара со звездочкой. Здесь события A и B зависимы.

Полезно отметить, что независимость случайных событий A и B влечет за собой также и независимость противоположных событий $\bar{A}$ и $\bar{B}$. Это соображение в ряде задач позволяет существенно упростить расчеты вероятностей.

Пример. Два стрелка независимо друг от друга стреляют по одной и той же цели. Вероятность попадания для первого стрелка равна 0,8, для второго 0,9. Найти вероятность поражения цели, т. е. вероятность того, что хотя бы один из стрелков попадет в цель.

Решение. Обозначим случайные события: A — попадание первого стрелка в цель, B — попадание второго стрелка в цель. Нам надо найти вероятность события $A+B$, являющегося суммой совместимых событий A и B. Перейдем к противоположному событию. Таким событием является совмещение $\bar{A}\cdot\bar{B}$; действительно, $A+B$ есть событие «хотя бы один из стрелков попадет в цель», $\bar{A}\cdot\bar{B}$ есть событие «ни один из стрелков в цель не попадет». Так как в условии дана независимость событий A и B, а значит и независимость событий $\bar{A}$ и $\bar{B}$, то по правилу умножения имеем

$$\mathsf{P}(\bar{A}\cdot\bar{B}) = \mathsf{P}(\bar{A})\,\mathsf{P}(\bar{B}).$$

Остается воспользоваться формулой (1.3-7) для расчета вероятностей противоположных событий. Получим

$$\mathsf{P}(A) = 0{,}8, \quad \mathsf{P}(\bar{A}) = 1-0{,}8 = 0{,}2,$$
$$\mathsf{P}(B) = 0{,}9, \quad \mathsf{P}(\bar{B}) = 1-0{,}9 = 0{,}1.$$

Отсюда

$$\mathsf{P}(\bar{A}\cdot\bar{B}) = 0{,}2\cdot 0{,}1 = 0{,}02,$$

и следовательно, снова по формуле (1.3-7),

$$\mathsf{P}(A+B) = 1 - \mathsf{P}(\bar{A}\cdot\bar{B}) = 1-0{,}02 = 0{,}98.$$

Заметим, что тот же результат можно получить и непосредственно по формуле (1.5-5), переписав ее для независимых событий A и B в виде

$$\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A)\,\mathsf{P}(B);$$

в рассматриваемом примере это дает

$$\mathsf{P}(A+B) = 0{,}8 + 0{,}9 - 0{,}8\cdot 0{,}9 = 0{,}98.$$

На следующей странице мы рассмотрим пример, в котором переход к противоположному событию дает более существенную экономию в расчетах. Вообще говоря, переходить к противоположному событию выгоднее в тех задачах, где противоположное событие распадается на значительно меньшее число простых слагаемых, чем прямое.

Распространим теперь определение независимости на большее число случайных событий.

Определение 2. *События* $A_1, A_2, \ldots, A_n$ $(n > 2)$ *называются независимыми в совокупности, если вероятность совмещения любых двух, трех, ..., n из них равна произведению соответствующих вероятностей.*

Например, для трех событий A, B, C независимость означает выполнение четырех соотношений:

$$\mathsf{P}(A\cdot B) = \mathsf{P}(A)\,\mathsf{P}(B), \quad \mathsf{P}(A\cdot C) = \mathsf{P}(A)\,\mathsf{P}(C), \tag{3}$$
$$\mathsf{P}(B\cdot C) = \mathsf{P}(B)\,\mathsf{P}(C),$$
$$\mathsf{P}(A\cdot B\cdot C) = \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C). \tag{4}$$

Для n событий последнее соотношение имеет вид

$$\mathsf{P}(A_1 \cdot A_2 \ldots A_n) = \mathsf{P}(A_1)\,\mathsf{P}(A_2) \ldots \mathsf{P}(A_n). \qquad (5)$$

В частном случае, когда все события имеют одинаковую вероятность p, формула (5) дает

$$\mathsf{P}(A_1 \cdot A_2 \ldots A_n) = p^n. \qquad (6)$$

Пример. В некоторой системе имеется важный узел, вероятность безотказной работы которого (за определенный промежуток времени) равна p. Для повышения надежности системы этот узел можно дублировать так, чтобы система работала, если работает хоть один из дублирующих узлов. Сколько раз надо дублировать узел, чтобы вероятность безотказной работы системы превысила заданный уровень $\mathcal{P}$ ($\mathcal{P} > p$)?

Решение. Перейдем к противоположному событию — нарушению работы системы. Это нарушение наступит тогда, когда выйдут из строя все дублирующие узлы. Будем считать выходы из строя любых дублирующих узлов независимыми случайными событиями; вероятности этих событий равны $1 - p$. Вероятность выхода из строя n дублирующих узлов равна $(1 - p)^n$ по формуле (6). Поэтому вероятность безотказной работы системы равна

$$1 - (1 - p)^n.$$

Нам надо найти такое n, чтобы выполнялось неравенство

$$1 - (1 - p)^n > \mathcal{P}.$$

Решаем это неравенство относительно n:

$$(1 - p)^n < 1 - \mathcal{P}, \quad n \lg (1 - p) < \lg (1 - \mathcal{P}),$$

и так как $\lg (1 - p) < 0$, то

$$n > \frac{\lg (1 - \mathcal{P})}{\lg (1 - p)};$$

остается выбрать наименьшее целое число, удовлетворяющее этому неравенству.

Например, если надежность узла (т. е. вероятность его безотказной работы) равна $p = 0{,}8$, а требуемая надежность системы есть $\mathcal{P} = 0{,}998$, то мы получаем

$$n > \frac{\lg(1-0{,}998)}{\lg(1-0{,}8)}, \quad \text{т. е.} \quad n > \frac{\lg 0{,}002}{\lg 0{,}2} = \frac{-2{,}699}{-0{,}699} = 3{,}86$$

и выбираем $n = 4$. Таким образом, в данных условиях достаточно иметь четыре дублирующих узла. Заметим, что при четырех дублирующих узлах надежность системы достигнет уровня

$$1-(1-p)^4 = 1-0{,}2^4 = 0{,}9984.$$

§ 1.8. О случайных событиях с вероятностями, близкими к 0 или 1

Вернемся снова к частотному толкованию вероятностей, обращая особое внимание на случайные события с вероятностями, близкими к 0 или 1.

Если событие имеет очень малую вероятность, то оно происходит очень редко. Поэтому в практических расчетах возможностью его появления обычно пренебрегают. Вопрос сводится к назначению границы пренебрежимо малых вероятностей, т. е. такого уровня α, что можно пренебречь возможностью появления событий, вероятность которых меньше чем α. *Назначение такой границы не является математической задачей*; это назначение происходит *вне теории вероятностей* и связано с существом решаемых на практике задач, с тем, насколько важен для нас успех предсказания. Так, в примере, рассмотренном в конце предыдущего параграфа, надежность системы была задана величиной $\mathcal{P} = 0{,}998$; это означает, что условия задачи позволяют пренебрегать возможностью появления событий, вероятность которых меньше чем $\alpha = 1 - \mathcal{P} = 0{,}002$.

После того как граница α пренебрежимо малых вероятностей уже установлена, задача сводится к рассмотрению тех событий, которые имеют вероятности, меньшие чем α, или большие чем $1 - \alpha$. В отношении таких событий обычно формулируют «*принцип практической уверенности*»: если вероятность события A меньше, чем α, то имеется практическая уверенность, что событие A не произойдет (при единичном испытании). При этом событие A называют *практически невозможным*, а противоположное событие $\bar{A}$ — *практически достоверным.*

Указанный принцип лежит в основе многих практических приложений теории вероятностей (например, к контролю качества продукции).

УПРАЖНЕНИЯ

1. Вычислить вероятность того, что при бросании двух правильных игральных костей сумма очков на верхних гранях будет больше десяти.

2. Вычислить вероятность того, что при произвольном разбиении колоды из 52 карт на две части в каждой из них окажется по 13 черных и 13 красных карт.

3. При бросании трех правильных игральных костей сумма X очков может дать 11 при шести комбинациях $(6+4+1)$, $(6+3+2)$, $(5+5+1)$, $(5+4+2)$, $(5+3+3)$, $(4+4+3)$; и может дать 12 тоже при шести комбинациях $(6+5+1)$, $(6+4+2)$, $(6+3+3)$, $(5+5+2)$, $(5+4+3)$, $(4+4+4)$. Можно ли отсюда сделать заключение о равенстве вероятностей случайных событий $X=11$ и $X=12$? Каковы вероятности этих событий?

4. На двух станках обрабатываются однотипные детали; вероятность брака для станка № 1 составляет 0,03, а для станка № 2—0,02. Обработанные детали складываются в одном месте, причем станок № 1 обрабатывает вдвое больше деталей, чем станок № 2. Вычислить вероятность того, что взятая наудачу деталь не будет бракованной.

5. Из урны, в которой лежат 4 шара с номерами 1, 2, 3 и 123, вынимается наудачу один шар. Событие A_k состоит в том, что на вынутом шаре окажется цифра k ($k=1, 2, 3$). Будут ли события A_1, A_2, A_3 независимы в совокупности?

6. Ставится тот же вопрос, что и в задаче 5, при условии, что в урне лежат 8 шаров с номерами 1, 2, 3, 12, 13, 20, 30 и 123.

ГЛАВА 2

ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 2.1. Дискретная случайная величина, закон распределения вероятностей

Как уже упоминалось в главе 1, случайную величину мы связываем с определенным испытанием. В наиболее простых задачах удается построить полную группу исходов испытания таким образом, что каждому из упомянутых исходов отвечает некоторое значение рассматриваемой величины. При этом можно считать, что различным исходам испытания соответствуют различные значения случайной величины (если некоторым исходам соответствует одно и то же значение, то эти исходы можно объединить, после чего получится снова полная группа). Таким образом, в качестве исходов испытания (случайных событий) можно рассматривать принятие случайной величиной каких-либо из ее возможных значений.

Определение. *Величина X называется дискретной случайной величиной, если множество ее возможных (различных) значений представляет собой конечную или бесконечную последовательность чисел x_1, x_2, . . ., x_i, . . . и если каждое событие $X = x_i$ является случайным событием, т. е. имеет определенную вероятность p_i (события $X = x_i$* мы будем здесь называть *элементарными событиями*).

Законом распределения (вероятностей) случайной величины X мы будем называть любое правило, позволяющее находить все вероятности $\mathsf{P}(X = x_i) = p_i$ $(i = 1, 2, . . .)$. Другими словами, закон распределения задает вероятность p_i как функцию, определенную на множестве элементарных случайных событий $X = x_i$.

Эту функцию иногда удается выразить формулой вида $p_i = f(x_i)$ (см., например, §§ 2.2—2.4).

Если случайная величина X может принимать лишь конечное число различных значений, то часто выписывают все эти значения в таблицу, указывая против каждого значения x_i соответствующую ему вероятность p_i. Такая таблица называется *таблицей распределения* (вероятностей) случайной величины X:

(Возможное значение) X	x_1	x_2	. . .	x_i	. . .	x_k
(Вероятность)	p_1	p_2	. . .	p_i	. . .	p_k

(1)

Например, число X выпадений герба при бросании двух монет есть дискретная случайная величина со следующей таблицей распределения вероятностей (см. стр. 21)

X	0	1	2
	1/4	1/2	1/4

Если случайная величина X может принимать лишь конечное число различных значений $x_1, x_2, \ldots, x_k$, то элементарные события

$$X = x_1, \quad X = x_2, \ldots, \quad X = x_k$$

образуют полную группу случайных событий, и поэтому сумма их вероятностей равна единице:

$$p_1 + p_2 + \ldots + p_k = 1. \tag{2}$$

Если множество (различных) возможных значений величины X бесконечно, то конечную сумму в формуле (2) следует заменить бесконечным рядом $\sum_{i=1}^{\infty} p_i$, сумма которого тоже должна быть равна 1.

Всякое случайное событие, связанное с дискретной случайной величиной X, можно толковать как попадание ее значений в некоторое множество чисел D; вероятность такого события равна сумме вероятностей тех значений x_i, которые принадлежат множеству D.

Пример 1. Число X очков, выпадающее на верхней грани правильной игральной кости, есть дискретная случайная величина с таблицей распределения вероятностей

X	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

(3)

Если игральная кость не является правильной, то число очков, выпадающее на ее верхней грани, будет также дискретной случайной величиной с тем же набором возможных значений, но с другим законом распределения вероятностей.

Пример 2. Пусть случайное событие A имеет вероятность $\mathsf{P}(A) = p$. Назовем **индикатором события** A случайную величину I_A, которая принимает только два значения 1 и 0, причем $I_A = 1$ означает «событие A произошло», $I_A = 0$ означает «событие A не произошло». Тогда индикатор есть дискретная случайная величина с таблицей распределения вероятностей

I_A	1	0
	p	q

, где $q = 1 - p$. (4)

Пример 3. Из урны, в которой лежат 2 белых и 8 черных шаров, последовательно вынимают шары до тех пор, пока не появится черный шар. Число Y вынутых при этом шаров есть дискретная случайная величина. Найти закон распределения ее вероятностей.

Решение. Возможными значениями величины Y являются, очевидно, 1, 2, 3. Для нахождения соответствующих вероятностей воспользуемся схемой урн и правилом умножения вероятностей. Событие $Y = 1$ означает, что уже первый вынутый шар будет черным; поэтому его вероятность равна

$$\mathsf{P}(Y = 1) = \frac{8}{10}.$$

Событие $Y = 2$ означает, что приходится вынуть два шара, значит, первый шар будет белым, а второй — черным; по правилу умножения вероятностей получим

$$\mathsf{P}(Y = 2) = \mathsf{P}(бч) = \frac{2}{10}\cdot\frac{8}{9}.$$

Здесь буквами «*б*» и «*ч*» указана последовательность цветов вынимаемых шаров. Наконец, событие $Y = 3$ означает, что первые два шара оказались белыми; при этом третий шар непременно будет черным и поэтому

$$\mathsf{P}(Y = 3) = \mathsf{P}(ббч) = \frac{2}{10}\cdot\frac{1}{9}.$$

Сумма всех найденных вероятностей равна единице:

$$\frac{8}{10} + \frac{2}{10}\cdot\frac{8}{9} + \frac{2}{10}\cdot\frac{1}{9} = 1,$$

что может служить для контроля произведенных вычислений.

§ 2.2. Примеры дискретных законов распределения

Некоторые простые законы распределения вероятностей часто встречаются при решении различных задач. Такие законы распределения получили специальные названия: геометрическое распределение, гипергеометрическое распределение, биномиальное распределение, распределение Пуассона и т. д.

Распределение выборочного значения признака

Пусть имеется совокупность элементов (предметов, явлений), различающихся величиной некоторого количественного признака; например, можно рассматривать совокупность студентов некоторого института, а в качестве признака взять размер обуви. Обозначим через

$x_1, x_2, \ldots, x_k$ — различные значения признака у элементов совокупности. Пусть значение x_1 встречается у M_1 элементов совокупности, значение x_2 — у M_2 элементов, ..., значение x_k — у M_k элементов; при этом сумма $M_1 + M_2 + \ldots + M_k = N$ дает общее число элементов совокупности.

Из рассматриваемой совокупности выбирается наудачу один элемент (как уже было сказано выше, слово «наудачу» означает, что каждому элементу совокупности обеспечивается одинаковая с остальными возможность быть выбранным). Нас интересует *величина признака у выбираемого элемента* или *выборочное значение признака*. Очевидно, это есть дискретная случайная величина X с возможными значениями $x_1, x_2, \ldots$ $\ldots, x_k$. Закон распределения ее вероятностей находится здесь просто с помощью схемы урн: число всех возможных случаев равно N, а случайному событию $X = x_i$ благоприятствует M_i случаев, следовательно,

$$\boxed{\mathsf{P}(X = x_i) = \frac{M_i}{N} \qquad (i = 1, 2, \ldots, k).} \tag{1}$$

Эта формула и дает искомое распределение.

Геометрическое распределение

Рассмотрим сначала хорошо известную игру с набрасыванием кольца на стержень. Нас будет интересовать число X бросаний кольца до первого попадания на стержень при условии, что вероятность попадания при каждом бросании не зависит от результатов предыдущих бросаний и сохраняет постоянное значение p $(0 < p < 1)$. Величина X будет дискретной случайной величиной, возможными значениями которой служат натуральные числа. Найдем закон распределения ее вероятностей. Событие $X = 1$ означает попадание с первого бросания, поэтому его вероятность равна p: $\mathsf{P}(X = 1) = p$. Событие $X = 2$ означает попадание со второго бросания и, значит, промах при первом бросании. Применяя правило умножения вероятностей, в силу условия независимости получаем

$$\mathsf{P}(X = 2) = q \cdot p, \qquad \text{где} \qquad q = 1 - p.$$

Событие $X = 3$ означает попадание с третьего бросания и, значит, промахи при первых двух бросаниях; по указанным выше соображениям получаем

$$\mathsf{P}(X = 3) = q \cdot q \cdot p = q^2 p.$$

Продолжая эти рассуждения, приходим к общей формуле

$$\boxed{\mathsf{P}(X = m) = q^{m-1} p \qquad (m = 1, 2, \ldots),} \tag{2}$$

которая и выражает геометрический закон распределения вероятностей. Это название связано с тем, что ряд вероятностей (2) представляет собой бесконечно убывающую геометрическую прогрессию со знаменателем q; сумма этого ряда равна единице:

$$p + qp + q^2 p + \ldots + q^{m-1} p + \ldots = p \cdot \frac{1}{1-q} = 1.$$

Заметим, что в этой задаче испытание, связанное со случайной величиной X, представляет собой всю последовательность бросаний. Каждое отдельное бросание можно рассматривать как частичное испытание с двумя возможными исходами — попадание и промах. Если обозначить через A_k попадание при k-м бросании, то указанную в задаче независимость следует понимать в том смысле, что события $A_1, A_2, \ldots, A_k$ при любом $k = 2, 3, \ldots$ независимы в совокупности. При этом событие $X = 1$ совпадает с событием A_1, а событие $X = m$ при $m \geqslant 2$ является произведением событий $\bar{A}_1 \cdot \bar{A}_2 \ldots \bar{A}_{m-1} \cdot A_m$. Постоянство вероятности попадания означает, что

$$\mathsf{P}(A_1) = \mathsf{P}(A_2) = \ldots = p.$$

При таком толковании сразу видно, что

$$\mathsf{P}(X = m) = \mathsf{P}(\bar{A}_1 \cdot \bar{A}_2 \ldots \bar{A}_{m-1} \cdot A_m) = q^{m-1} p.$$

Примечание. В рассматриваемой задаче существенно, что игра ведется без ограничения числа бросаний. Если число бросаний ограничено, то меняется и формулировка задачи и закон распределения. Пусть, например, правила игры позволяют бросать кольцо не более трех раз. Сохраняя те же условия и те же обозначения, что и выше, мы снова получим

$$\mathsf{P}(X = 1) = p, \quad \mathsf{P}(X = 2) = qp.$$

Но событие $X = 3$ означает, что использованы все три бросания, т. е. означает просто два промаха в первых двух бросаниях; поэтому вероятность этого события равна

$$\mathsf{P}(X = 3) = q^2.$$

Таким образом, в этой задаче случайная величина X имеет конечное число возможных значений с таблицей распределения вероятностей

X	1	2	3
	p	qp	q^2

Конечно, и здесь сумма вероятностей равна единице:

$$p + qp + q^2 = p + q(p + q) = 1.$$

§ 2.3. Биномиальное распределение

Как уже известно, практическое представление о вероятности случайного события дает его относительная частота (в достаточно большой последовательности повторений испытания). Но сама относительная частота случайного события может рассматриваться как случайная величина, которая в разных сериях испытаний может принимать различные значения. Естественно поэтому, что распределение вероятностей относительной частоты случайного события играет большую роль и в теории вероятностей и в ее приложениях. Так как относительная частота принимает дробные значения, то обычно рассматривают абсолютную частоту, т. е. число появлений случайного события, которое только постоянным множителем отличается от его относительной частоты.

Рассмотрим сложное испытание, заключающееся в n-кратном повторении испытания, которое мы будем называть простым. Пусть с простым испытанием связано случайное событие A, имеющее определенную вероятность $\mathsf{P}(A) = p$ $(0 < p < 1)$. Появление события A при i-м повторении простого испытания есть случайное событие A_i в сложном испытании. Вероятности всех случайных событий A_i, очевидно, равны p. Мы будем предполагать, что все события $A_1, A_2, \ldots, A_n$ независимы в совокупности. Обычно это трактуют как независимость самих испытаний, составляющих сложное испытание, и говорят, что рассматривается *последовательность независимых испытаний по схеме Бернулли* или *по схеме повторной выборки* (по схеме возвращенного шара). Последнее название вызвано следующим толкованием.

В урне лежат неразличимые на ощупь шары, часть из них белого цвета, причем доля белых шаров в урне равна p. Случайное событие A заключается в том, что вынутый наугад из урны шар будет белого цвета; вероятность этого случайного события равна p — доле белых шаров в урне. Вынув из урны наугад один шар, отмечают, белый он или нет, затем вынутый шар возвращают в урну и шары тщательно перемешивают. После этого снова вынимают наугад один шар и так повторяют этот процесс n раз.

Интересующая нас случайная величина X — число появлений события A при n-кратном повторении испытания. Возможными значениями величины X служат числа $0, 1, 2, \ldots, n$. Найдем закон распределения вероятностей случайной величины X.

Событие $X = n$ означает появление события A во всех испытаниях; по правилу умножения вероятностей (с учетом условия независимости) получаем

$$\mathsf{P}(X = n) = \mathsf{P}(A_1 \cdot A_2 \ldots A_n) = p^n.$$

Аналогично находим $\mathsf{P}(X = 0) = q^n$, где $q = 1 - p = \mathsf{P}(\bar{A})$.

Событие $X = m$ $(m = 1, 2, \ldots, n - 1)$ означает, что за n испытаний случайное событие A наступит точно m раз и, значит, $n - m$ раз наступит противоположное событие $\bar{A}$. Вероятность того, что событие A наступит в первых m испытаниях и не наступит в остальных, подсчитаем по правилу умножения вероятностей:

$$\mathsf{P}(A_1 \cdot A_2 \ldots A_m \cdot \bar{A}_{m+1} \ldots \bar{A}_n) = p^m q^{n-m}.$$

Эта вероятность не зависит от того, в каких именно испытаниях наступит событие A. Поэтому по правилу сложения вероятностей искомая вероятность $\mathsf{P}(X = m)$ равна вероятности $p^m q^{n-m}$, умноженной на число

C_n^m способов выбора m испытаний, в которых наступит событие A, из общего числа n испытаний:

$$\boxed{\mathsf{P}(X = m) = C_n^m p^m q^{n-m}.} \qquad (1)$$

Здесь C_n^m есть число сочетаний из n элементов по m (см. примечание на стр. 22). Таким образом, мы получаем следующую таблицу распределения вероятностей:

X	0	1	$\ldots$	m	$\ldots$	$n-1$	n
	q^n	npq^{n-1}	$\ldots$	$C_n^m p^m q^{n-m}$	$\ldots$	$np^{n-1}q$	p^n

(2)

Этот закон распределения называется *биномиальным законом распределения вероятностей*. Название связано с тем, что вероятности (1) совпадают с соответствующими членами разложения бинома $(q+p)^n$ по степеням p:

$$(q+p)^n = q^n + npq^{n-1} + \ldots + C_n^m p^m q^{n-m} + \ldots + p^n.$$

Отсюда, кстати, сразу видно, что сумма всех вероятностей в таблице (2) равна единице, так как $q + p = 1$.

Примечание. Иногда в рассмотренной задаче вводят следующие термины. Появление события A называют «успехом», а его непоявление «неудачей». В этой терминологии формула (1) дает вероятность точно m «успехов» при n-кратном повторении испытания.

Пример. Из большой партии изделий берут на пробу 10 штук. Известно, что доля нестандартных изделий во всей партии составляет 25%. Требуется найти вероятность того, что более пяти отобранных изделий окажутся нестандартными.

Решение. Отбор каждого изделия будем считать испытанием, а обнаружение нестандартности у отобранного изделия — событием A. Вероятность p события A равна доле нестандартных изделий во всей партии, т. е. $p = 0{,}25$.

Количество X нестандартных изделий среди отобранных будет случайной величиной с биномиальным распределением вероятностей, если только изделия для пробы отбираются по схеме случайной повторной выборки (т. е. если изделия после проверки возвращают в исходную партию). При этом вероятности подсчитываются по формуле (1) при $n = 10$, $p = 0{,}25$, $q = 0{,}75$.

Таблица распределения вероятностей случайной величины X принимает вид, представленный в таблице 1 (с точностью до 0,0001).

Таблица 1

Биномиальное распределение вероятностей при $n = 10$, $p = 0{,}25$

m	$\mathsf{P}(X = m)$	m	$\mathsf{P}(X = m)$	m	$\mathsf{P}(X = m)$
0	0,0563	4	0,1460	8	0,0004
1	0,1877	5	0,0584	9	0,0000
2	0,2816	6	0,0162	10	0,0000
3	0,2503	7	0,0031		

В соответствии с правилом сложения вероятностей складываем вероятности $\mathsf{P}(X = m)$ при $m = 6, 7, 8, 9, 10$ и находим искомую вероятность

$$\mathsf{P}(X > 5) = \sum_{m=6}^{10} \mathsf{P}(X = m) \approx 0{,}020.$$

Полученная вероятность довольно мала, так что если бы в отобранном десятке изделий оказалось шесть или более нестандартных, то мы имели бы основание усомниться в том, действительно ли доля нестандартных изделий во всей партии составляет только 25%.

Примечание 1. Выше мы уже отмечали, что расчеты вероятностей в рассмотренном примере точны только при условии отбора изделий по схеме случайной повторной выборки. На практике изделия для проверки отбирают по схеме *случайной бесповторной выборки*, т. е. изделия из партии отбирают каждый раз случайно (обеспечивая для каждого изделия одинаковую с остальными возможность быть выбранным), но отобранные изделия не возвращают в партию (такой отбор особенно важен в тех задачах, где проверка изделия связана с его разрушением). Схема случайной бесповторной выборки приводит уже не к биномиальному, а к так называемому *гипергеометрическому закону* *распределения.* Не останавливаясь на выводе соответствующих формул, заметим только, что если число изделий в партии велико, а число отобранных изделий составляет весьма малую долю всей партии, то расчет интересующих нас вероятностей по формулам биномиального распределения дает достаточно хорошие приближения. Для иллюстрации этого утверждения произведем расчет вероятностей при таких условиях: партия содержит 1000 изделий, из них 25% нестандартных, что составляет 250 изделий; из партии случайно, но без возврата, отбирают 10 изделий. Сохраняя те же обозначения, что и выше, получим

$$\mathsf{P}(X=6)=\frac{C_{250}^{6}C_{750}^{4}}{C_{1000}^{10}}=$$

$$=\frac{250\cdot 249\cdot 248\cdot 247\cdot 246\cdot 245\cdot 750\cdot 749\cdot 748\cdot 747\cdot 10!}{6!\,4!\,1000\cdot 999\cdot 998\ldots 991}=0{,}0158$$

и, аналогично, $\mathsf{P}(X=7)=0{,}0030,\ldots$

В итоге мы получим $\mathsf{P}(X>5)\approx 0{,}019$, что мало отличается от значения вероятности, вычисленного по биномиальному распределению в примере на стр. 48 (но заметим, что расчет по биномиальному распределению значительно проще).

Примечание 2. Для биномиального закона составлены таблицы распределения вероятностей при различных n, p и m. Таблица 1 дает представление о типичном поведении вероятностей в этих таблицах. Просматривая таблицу 1, мы замечаем, что вероятности $\mathsf{P}(X=m)$ сначала возрастают, а затем убывают, так что значению $m=2$ соответствует наибольшая вероятность. Выясним, в какой мере это свойство присуще биномиальному распределению. Возьмем отношение двух последовательных вероятностей $\mathsf{P}(X=m)$ и $\mathsf{P}(X=m+1)$:

$$\frac{\mathsf{P}(X=m+1)}{\mathsf{P}(X=m)}=\frac{n-m}{m+1}\cdot\frac{p}{q}\qquad (m=0,1,2,\ldots,n-1).$$

При возрастании m от 0 до $n-1$ это отношение убывает от $\frac{np}{q}$ до $\frac{p}{nq}$. Если $np>q$ и $nq>p$, то рассматриваемое отношение переходит от значений, бо́льших 1, к значениям, меньшим 1. А это как раз и означает, что вероятности $\mathsf{P}(X=m)$ сначала возрастают, а затем убывают. И лишь в крайних случаях, когда $np<q$ или $nq<p$, вероятности $\mathsf{P}(X=m)$ изменяются монотонно. Примеры всех трех возможных случаев приведены на рис. 4.

Во всех случаях наиболее вероятное значение $X=m_0$ находится из неравенств

$$\frac{n-(m_0-1)}{m_0}\,\frac{p}{q}\geqslant 1,\qquad \frac{n-m_0}{m_0+1}\,\frac{p}{q}\leqslant 1,$$

откуда следует, что

$$np+p-1\leqslant m_0\leqslant np+p.$$

Этому неравенству удовлетворяет только одно целое число m_0, если только $np+p$ не является целым числом; в последнем случае имеются два наиболее вероятных значения $X = np + p - 1$ и $X = np + p$. В частном случае, когда произведение np — целое число, оно и является наиболее вероятным значением величины X.

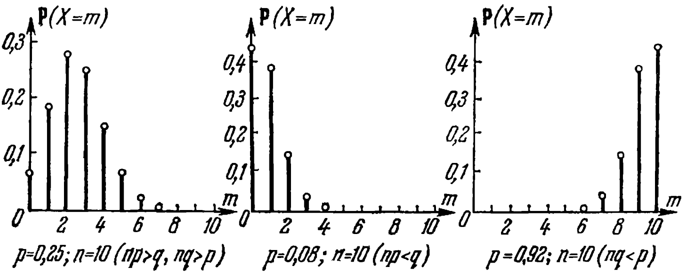


Рис. 4.

В заключение вернемся к поставленному в начале настоящего параграфа вопросу о законе распределения относительной частоты появления случайного события. Относительная частота только постоянным множителем отличается от числа X появлений случайного события, а именно, относительная частота есть X/n, где n — число испытаний. При этом очевидно, что события $X = m$ и $\frac{X}{n} = \frac{m}{n}$ равновероятны; поэтому таблица распределения вероятностей относительной частоты имеет вид

$\frac{X}{n}$	0	$\frac{1}{n}$	$\ldots$	$\frac{m}{n}$	$\ldots$	$\frac{n-1}{n}$	1
	q^n	npq^{n-1}	$\ldots$	$C_n^m p^m q^{n-m}$	$\ldots$	$np^{n-1}q$	p^n

(3)

§ 2.4. Распределение Пуассона

Распределение Пуассона часто встречается в задачах, связанных с простейшим потоком событий. Под *потоком событий* следует понимать последовательность событий, наступающих одно за другим в случайные моменты времени. Примерами могут служить: поток вызовов на телефонной станции, поток заявок в системе массового обслуживания, поток отказов при работе ЭВМ, последовательность распада частиц некоторого количества радия.

Простейший поток событий характеризуется следующими свойствами:

а) вероятность наступления того или иного числа событий за любой промежуток времени зависит только от длительности этого промежутка (а не от начала отсчета);

б) указанная вероятность не зависит от того, какое число событий наступило до начала рассматриваемого промежутка времени (отсутствие последействия);

в) за малый промежуток времени вероятность наступления одного события приближенно пропорциональна длительности такого промежутка, а вероятностью наступления двух или более событий можно пренебречь (это свойство мы уточним ниже, см. формулы (1) и (2)).

В качестве случайной величины X мы рассмотрим число событий простейшего потока, наступающих за фиксированный промежуток времени t. Значениями этой случайной величины могут быть любые целые числа $m = 0, 1, 2, 3, \ldots$ Соответствующие вероятности мы обозначим через

$$p_m(t) = \mathsf{P}(X = m);$$

$p_m(t)$ есть вероятность того, что за фиксированный промежуток времени t наступит ровно m событий простейшего потока.

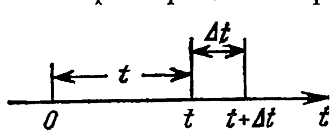


Рис. 5.

Для расчета вероятностей $p_m(t)$ мы применим метод вариации промежутка времени t, изменяя его на малую величину $\Delta t > 0$ и сравнивая вероятности $p_m(t)$ и $p_m(t + \Delta t)$ (рис. 5).

Прежде всего уточним свойство в) для вероятностей $p_m(\Delta t)$, относящихся к промежутку времени $\Delta t \to 0$:

$$p_1(\Delta t) = \lambda \Delta t + \theta_1, \qquad (1)$$

где

$$\lambda = \text{const} > 0, \qquad \lim_{\Delta t \to 0} \frac{\theta_1}{\Delta t} = 0;$$

$$\sum_{m=2}^{\infty} p_m(\Delta t) = \theta_2, \qquad \text{где } \lim_{\Delta t \to 0} \frac{\theta_2}{\Delta t} = 0. \tag{2}$$

Так как сумма всех вероятностей $p_m(\Delta t)$ для значений m от 0 до ∞ равна единице, то отсюда следует, что

$$p_0(\Delta t) = 1 - \lambda \Delta t + \theta_0, \qquad \text{где } \lim_{\Delta t \to 0} \frac{\theta_0}{\Delta t} = 0. \tag{3}$$

Применяя известное обозначение $o(\Delta t)$ для бесконечно малой величины более высокого порядка, чем Δt (при $\Delta t \to 0$), можно записать формулы (1) — (3) в виде

$$p_0(\Delta t) = 1 - \lambda \Delta t + o(\Delta t),$$
$$p_1(\Delta t) = \lambda \Delta t + o(\Delta t),$$
$$\sum_{m=2}^{\infty} p_m(\Delta t) = o(\Delta t).$$

Другими словами, с точностью до бесконечно малых более высокого порядка, чем Δt, можно приближенно считать, что $p_0(\Delta t)$ и $p_1(\Delta t)$ равны соответственно $1 - \lambda \Delta t$ и $\lambda \Delta t$, а вероятностями $p_m(\Delta t)$ при $m \geqslant 2$ можно пренебречь; это и есть уточнение свойства в), о котором было сказано выше.

Начнем с расчета вероятности $p_0(t)$, т. е. вероятности отсутствия событий за промежуток времени t. Так как отсутствие событий за промежуток времени $t + \Delta t$ означает отсутствие событий и за промежуток времени t и за промежуток времени Δt, то по правилу умножения вероятностей находим

$$p_0(t + \Delta t) = p_0(t) \cdot p_0(\Delta t)$$

(здесь мы использовали свойство отсутствия последействия). Отсюда с помощью формулы (3) получаем

$$p_0(t + \Delta t) = p_0(t)(1 - \lambda \Delta t + \theta_0)$$

и, значит,

$$\frac{p_0(t + \Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) + p_0(t) \frac{\theta_0}{\Delta t}.$$

В пределе при $\Delta t \to 0$ мы получаем для искомой вероятности $p_0(t)$ линейное дифференциальное уравнение первого порядка

$$p_0'(t) = -\lambda p_0(t). \tag{4}$$

Решая это уравнение при начальном условии $p_0(0) = 1$ (что вытекает из соотношения (3)), находим искомую вероятность отсутствия событий за промежуток времени t:

$$\boxed{p_0(t) = e^{-\lambda t}.} \tag{5}$$

Перейдем теперь к расчету вероятностей $p_m(t)$ при $m = 1, 2, 3, \ldots$ Этот расчет будет несколько сложнее, так как m событий за промежуток времени $t + \Delta t$ могут произойти при условиях, что за промежуток времени t произошло $m, m-1, \ldots, 1$ или 0 событий, а за промежуток времени Δt — соответственно $0, 1, \ldots, m-1$ или m событий. Обозначим через H_m, $H_{m-1}, \ldots, H_1, H_0$ гипотезы о том, что за промежуток времени t произошло соответственно $m, m-1, \ldots, 1, 0$ событий рассматриваемого потока, и воспользуемся формулой полной вероятности

$$\mathsf{P}(A) = \mathsf{P}(H_m)\mathsf{P}(A \mid H_m) + \mathsf{P}(H_{m-1})\mathsf{P}(A \mid H_{m-1}) + \\ + \ldots + \mathsf{P}(H_1)\mathsf{P}(A \mid H_1) + \mathsf{P}(H_0)\mathsf{P}(A \mid H_0).$$

Считая здесь событием A появление m событий рассматриваемого потока за промежуток времени $t + \Delta t$ и учитывая указанные выше условия простейшего потока, находим

$$\mathsf{P}(A) = p_m(t + \Delta t),$$

$$\mathsf{P}(H_k) = p_k(t); \quad \mathsf{P}(A \mid H_k) = p_{m-k}(\Delta t)$$

$$(k = m, m-1, \ldots, 1, 0)$$

(при условии, что за промежуток времени t произошло k событий, условная вероятность наступления m событий за промежуток времени $t + \Delta t$ равна вероятности наступления $m - k$ событий за оставшийся промежуток времени Δt). Поэтому

$$p_m(t + \Delta t) = \\ = p_m(t)\, p_0(\Delta t) + p_{m-1}(t)\, p_1(\Delta t) + \ldots + p_0(t) p_m(\Delta t);$$

отсюда с помощью формул (1) — (3) получаем

$$p_m(t+\Delta t) = p_m(t)(1-\lambda\Delta t) + p_{m-1}(t)\lambda\Delta t + \theta,$$

и, значит,

$$\frac{p_m(t+\Delta t) - p_m(t)}{\Delta t} = -\lambda p_m(t) + \lambda p_{m-1}(t) + \frac{\theta}{\Delta t},$$

где $\lim\limits_{\Delta t\to 0}\frac{\theta}{\Delta t} = 0$.

В пределе при $\Delta t \to 0$ мы получаем для искомой вероятности $p_m(t)$ линейное дифференциальное уравнение первого порядка

$$p'_m(t) + \lambda p_m(t) = \lambda p_{m-1}(t), \tag{6}$$

в правой части которого находится еще вероятность $p_{m-1}(t)$.

Другими словами, мы получаем *последовательность дифференциальных уравнений*

$$p'_m(t) + \lambda p_m(t) = \lambda p_{m-1}(t) \qquad (m = 1, 2, \ldots)$$

с одной известной функцией $p_0(t) = e^{-\lambda t}$. Начальные условия для этих уравнений находим из формул (1) и (2) при $\Delta t \to 0$:

$$p_m(0) = 0 \qquad (m = 1, 2, \ldots).$$

Для решения полученной системы удобно произвести замену

$$z_m(t) = p_m(t)\, e^{\lambda t}.$$

Тогда

$$z'_m(t) = [p'_m(t) + \lambda p_m(t)]\, e^{\lambda t}$$

и уравнения (6) сводятся к уравнениям

$$z'_m(t) = \lambda z_{m-1}(t),$$

которые решаются последовательным интегрированием. Учитывая, что при сделанной замене имеем

$$z_0(t) \equiv 1$$

и что начальные условия дают $z_m(0) = 0$ $(m = 1, 2, \ldots)$, последовательно находим

$$z_1(t) = \int_0^t \lambda z_0(t)\,dt = \int_0^t \lambda\,dt = \lambda t,$$

$$z_2(t) = \int_0^t \lambda z_1(t)\,dt = \int_0^t \lambda t \cdot \lambda\,dt = \frac{(\lambda t)^2}{2!}$$

и, вообще,

$$z_m(t) = \int_0^t \lambda z_{m-1}(t)\,dt = \int_0^t \frac{(\lambda t)^{m-1}}{(m-1)!}\lambda\,dt = \frac{(\lambda t)^m}{m!}.$$

Таким образом, мы получаем общую формулу для искомых вероятностей

$$\boxed{p_m(t) = \frac{(\lambda t)^m}{m!}e^{-\lambda t} \qquad (m = 0, 1, 2, \ldots).} \tag{7}$$

Эта формула и дает закон распределения случайной величины X — числа событий простейшего потока, наступающих за промежуток времени t. Статистический смысл параметра λ в формуле (7) можно выяснить из соотношения (1): если единицу времени разбить на N равных промежутков $\Delta t = 1/N$, то, заменив формулу (1) приближенным соотношением

$$p_1(\Delta t) \approx \frac{\lambda}{N},$$

мы можем толковать его правую часть как относительную частоту наступления некоторого события простейшего потока за промежуток времени наблюдения Δt; это означает, что λ есть среднее число наступлений таких событий за единицу времени. Другой подход к толкованию параметра λ мы рассмотрим далее в § 4.5.

Обычно произведение λt в формуле (7) обозначают одной буквой a и записывают закон распределения в виде

$$\boxed{\mathsf{P}(X = m) = \frac{a^m}{m!} e^{-a} \qquad (m = 0, 1, 2, \ldots).} \tag{8}$$

Этот закон распределения и называют *законом распределения Пуассона*. Сумма всех вероятностей (8) равна единице (это может служить для контроля произведенных расчетов):

$$\sum_{m=0}^{\infty} \mathsf{P}(X = m) = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!} = e^{-a} \cdot e^{a} = 1.$$

Применение распределения Пуассона в качестве приближения для биномиального распределения (при малых значениях p и больших значениях n).

Рассмотрим асимптотическое представление формулы биномиального распределения (2.3-1) при больших значениях n. Обозначим $np = a$, так что $p = a/n$, $q = 1 - p = 1 - a/n$, и разложим вероятность $\mathsf{P}(X = m) = C_n^m p^m q^{n-m}$ в степенной ряд по степеням $1/n$; ограничиваясь членами порядка $1/n$, получим

$$C_n^m p^m q^{n-m} =$$

$$= \frac{n(n-1)\ldots(n-m+1)}{m!} \left(\frac{a}{n}\right)^m \left(1 - \frac{a}{n}\right)^{n-m} =$$

$$= \frac{a^m}{m!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{m-1}{n}\right) e^{(n-m)\ln\left(1 - \frac{a}{n}\right)} =$$

$$= \frac{a^m}{m!} \left(1 - \frac{m(m-1)}{2n} + \ldots\right) e^{-a + \frac{1}{n}\left(ma - \frac{a^2}{2}\right) + \ldots} =$$

$$= \frac{a^m}{m!} \left(1 - \frac{m(m-1)}{2n} + \ldots\right) e^{-a} \left(1 + \frac{ma - \frac{a^2}{2}}{n} + \ldots\right) =$$

$$= \frac{a^m}{m!} e^{-a} \left[1 + \frac{1}{n} \left(ma - \frac{a^2}{2} - \frac{m(m-1)}{2}\right) + \ldots\right].$$

Первый член этого разложения дает как раз вероятность в распределении Пуассона, откуда и следует приближенная (асимптотическая) формула

$$C_n^m p^m q^{n-m} \approx \frac{a^m}{m!} e^{-a} \qquad (a = np);$$

второй член в квадратных скобках может служить для оценки относительной погрешности этой приближенной формулы. Точность получаемого приближения при небольших значениях m получается достаточно хорошей, если значение a невелико (обычно это приближение применяют при $a < 4$ и $p < 0,1$). Приведем пример расчета вероятностей в распределении Пуассона при $a = 1$ и для сравнения приведем значения вероятностей в биномиальном распределении при $p = 0,01$, $n = 100$ и $p = 0,02$, $n = 50$ (советуем обратить внимание на простоту расчета вероятностей в распределении Пуассона и на оценки относительных погрешностей, указанные в скобках в столбцах вероятностей биномиального распределения).

m	Распределение Пуассона $\mathsf{P}(X = m) = \frac{1}{m!}e^{-1}$	Биномиальное распределение	
		$p = 0,01$, $n = 100$	$p = 0,02$, $n = 50$
0	$\mathsf{P}(X = 0) = e^{-1} = 0,3679$	0,3660 (0,5%)	0,3642 (1%)
1	$\mathsf{P}(X = 1) = 1\ \mathsf{P}(X = 0) = 0,3679$	0,3697 (0,5%)	0,3716 (1%)
2	$\mathsf{P}(X = 2) = \frac{1}{2}\mathsf{P}(X = 1) = 0,1839$	0,1849 (0,5%)	0,1858 (1%)
3	$\mathsf{P}(X = 3) = \frac{1}{3}\mathsf{P}(X = 2) = 0,0613$	0,0610 (0,5%)	0,0606 (1%)
4	$\mathsf{P}(X = 4) = \frac{1}{4}\mathsf{P}(X = 3) = 0,0153$	0,0149 (2,5%)	0,0145 (5%)
5	$\mathsf{P}(X = 5) = \frac{1}{5}\mathsf{P}(X = 4) = 0,0031$	0,0029 (5,5%)	0,0028 (11%)
6	$\mathsf{P}(X = 6) = \frac{1}{6}\mathsf{P}(X = 5) = 0,0005$	0,0005 (9,5%)	0,0004 (19%)

Вероятности всех остальных значений m, не указанных в таблице, составляют в сумме менее 0,0001.

§ 2.5. Эмпирические распределения дискретных величин

В рассмотренных выше примерах нам каждый раз удавалось построить вероятностную модель и найти распределение вероятностей рассматриваемой случайной величины. Построение модели всегда связано с определенными допущениями и упрощениями. Например, распределение (2.1-3) для числа очков, выпадающих на верхней грани игральной кости, выведено в предположении, что кость правильная; при выводе распределения Пуассона мы приняли целый ряд допущений о «простейшем» характере потока случайных событий. В практических задачах кость может оказаться фальшивой, поток случайных событий — не простейшим. Поэтому на практике часто возникает вопрос о пригодности построенной модели для описания реальных явлений. Для случайного события статистическим аналогом вероятности служит относительная частота. Для дискретной случайной величины статистическим аналогом распределения вероятностей служит эмпирическое распределение относительных частот, которое можно получить, повторяя опыт или наблюдение n раз и регистрируя относительные частоты m_i/n появления различных значений x_i случайной величины X:

X	x_1	x_2	...	x_k
Относительная частота	$\frac{m_1}{n}$	$\frac{m_2}{n}$	...	$\frac{m_k}{n}$

Если построенная нами модель пригодна, т. е. если предполагаемое теоретическое распределение вероятностей согласуется с эмпирическим распределением относительных частот, то при достаточно большом числе n повторений испытания мы вправе ожидать, что относительные частоты m_i/n будут близки соответственно к вероятностям $p_i = \mathsf{P}(X = x_i)$ $(i = 1, 2, \ldots, k)$. (Разумеется, мы не ожидаем полного совпадения, но слишком большие расхождения заставляют подвергнуть сомнению пригодность исходной модели.)

Для иллюстрации рассмотрим два примера сравнения эмпирического распределения с распределением Пуассона. Классическим примером является рассмотренное В. Борткевичем распределение числа лиц, убитых ударом копыта в десяти армейских корпусах в Пруссии за 20 лет (1875—1894), на один корпус в год (см. табл. 2)

В правом столбце таблицы приведены вероятности, рассчитанные по распределению Пуассона (2.4-8), в котором за параметр a принято среднее число смертей за год на один корпус (см. §§ 2.4, 4.4 и 5.4):

$$a = \frac{122}{200} = 0{,}61 \qquad (e^{-0{,}61} = 0{,}5434).$$

Таблица 2

Число смертей за год	Частота таких случаев	Относительная частота	Вероятность
0	109	0,545	0,543
1	65	0,325	0,332
2	22	0,110	0,101
3	3	0,015	0,021
4 (и более)	1	0,005	0,003
Всего	200	1,000	1,000

В качестве второго примера рассмотрим распределение числа радиоактивных атомов, распадающихся в течение промежутка времени 1/8 минуты по данным $n = 2608$ наблюдений (см. табл. 3).

Таблица 3

Число распавшихся атомов	Частота	Относительная частота	Вероятность
0	57	0,022	0,021
1	203	0,078	0,081
2	383	0,147	0,156
3	525	0,201	0,201
4	532	0,204	0,195
5	408	0,156	0,151
6	273	0,105	0,097
7	139	0,053	0,054
8	45	0,017	0,026
9	27	0,010	0,011
10	10	0,004	0,004
11	4	0,002	0,002
12 (и более)	2	0,001	0,001
Всего	2608	1,000	1,000

В правом столбце приведены вероятности, рассчитанные по распределению Пуассона (2.4-8), в котором за параметр a принято среднее число атомов, распавшихся за выбранный промежуток времени в 1/8 минуты:

$$a = 3{,}87 \qquad (e^{-3{,}87} = 0{,}02086).$$

Здесь согласие между теоретическим и эмпирическим распределениями не такое хорошее, как в предыдущем примере, но все же вполне удовлетворительное.

Для более обоснованного суждения о согласованности предполагаемого теоретического распределения с эмпирическим существуют различные критерии согласия, излагаемые в курсах математической статистики.

§ 2.6. Независимость дискретных случайных величин

Во многих задачах приходится иметь дело с системой дискретных случайных величин. Для изучения системы случайных величин надо знать закон совместного распределения их вероятностей. Пусть, например, испытание состоит в бросании шарика на доску с круглыми отверстиями (лузами), причем шарик обязательно

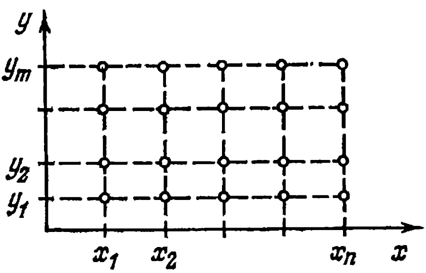


Рис. 6.

попадает в одну из луз. Пусть центры луз расположены в узлах декартовой сетки координат (рис. 6); обозначим абсциссы центров луз через $x_1, x_2, \ldots, x_n$, а ординаты — через $y_1, y_2, \ldots, y_m$. Тогда центр лузы, в которую попадает шарик, можно рассматривать как двумерную дискретную случайную величину (X, Y) с возможными значениями (x_i, y_j); такая случайная величина представляет собой систему двух случайных величин X и Y — ее декартовых координат. Задание закона совместного распределения величин X и Y означает задание вероятностей попадания шарика в каждую из луз с центрами $(x_i,\ y_j)$:

$$\mathsf{P}\,(X = x_i,\ Y = y_j) = p_{ij} \tag{1}$$
$$(i = 1,\ 2,\ \ldots,\ n;\ j = 1,\ 2,\ \ldots,\ m).$$

Эти вероятности могут быть любыми неотрицательными числами, сумма которых равна единице. Их можно записать в таблицу распределения с двумя входами:

X	Y				
	y_1	y_2	y_3	$\ldots$	y_m
x_1	p_{11}	p_{12}	p_{13}	$\ldots$	p_{1m}
x_2	p_{21}	p_{22}	p_{23}	$\ldots$	p_{2m}
x_3	p_{31}	p_{32}	p_{33}	$\ldots$	p_{3m}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_n	p_{n1}	p_{n2}	p_{n3}	$\ldots$	p_{nm}

(2)

Каждую вероятность p_{ij} можно рассматривать как вероятность совмещения случайных событий $X = x_i$ и $Y = y_j$. Независимость таких событий в соответствии с § 1.7 определяется соотношением

$$\mathsf{P}\,((X = x_i)\cdot(Y = y_j)) = \mathsf{P}\,(X = x_i)\,\mathsf{P}\,(Y = y_j).$$

Это оправдывает следующее

Определение. *Две дискретные случайные величины X и Y называются независимыми, если для всех их возможных значений x_i и y_j имеют место соотношения*

$$p_{ij} = \mathsf{P}\,(X = x_i)\,\mathsf{P}\,(Y = y_j). \tag{3}$$

Полезно обратить внимание на то, что для независимости дискретных случайных величин X и Y необходимо, чтобы вероятность совмещения *каждого* возможного значения x_i с *каждым* возможным значением y_j была отлична от нуля. Например, при расположении луз, указанном на рис. 7, абсцисса и ордината точки попадания не могут быть независимыми, так как

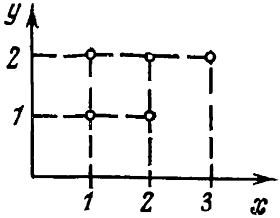


Рис. 7.

$$\mathsf{P}(X = 3,\ Y = 2) \neq 0,$$
$$\mathsf{P}(X = 3,\ Y = 1) = 0.$$

Данное выше определение естественно распространяется и на большее число дискретных случайных величин. Например, три дискретные случайные величины X, Y, Z с возможными значениями x_i, y_j, z_k и совместными вероятностями p_{ijk} будут независимы, если для всех их возможных значений имеют место соотношения

$$p_{ijk} = \mathsf{P}(X = x_i)\,\mathsf{P}(Y = y_j)\,\mathsf{P}(Z = z_k).$$

Понятия независимости случайных событий (§ 1.7) и независимости случайных величин согласованы между собой таким образом, что независимость случайных событий равносильна независимости их индикаторов (см. пример 2 на стр. 41). При большом числе событий условия независимости индикаторов могут оказаться более удобными для расчетов, чем условия независимости случайных событий. Мы проверим только, что из независимости двух случайных событий A_1 и A_2 вытекает независимость их индикаторов I_1 и I_2. Пусть вероятности событий A_1 и A_2 равны соответственно p_1 и p_2. Тогда индикаторы имеют следующие распределения вероятностей:

I_1	1	0
	p_1	q_1

, $\quad q_1 = 1 - p_1,$

I_2	1	0
	p_2	q_2

, $\quad q_2 = 1 - p_2.$

Пусть совместное распределение индикаторов задано таблицей

I_1	I_2 = 1	I_2 = 0
1	p_{11}	p_{10}
0	p_{01}	p_{00}

Из условия независимости случайных событий A_1 и A_2 вытекает, что $p_{11} = \mathsf{P}(I_1 = 1, I_2 = 1) = \mathsf{P}(A_1 \cdot A_2) = p_1 \cdot p_2$.

Далее, по правилу сложения вероятностей получаем

$$p_{11} + p_{10} = \mathsf{P}(I_1 = 1) = p_1,$$
$$p_{11} + p_{01} = \mathsf{P}(I_2 = 1) = p_2,$$
$$p_{01} + p_{00} = \mathsf{P}(I_1 = 0) = q_1,$$

откуда

$$p_{10} = p_1 - p_{11} = p_1(1 - p_2) = p_1 q_2,$$
$$p_{01} = p_2 - p_{11} = p_2(1 - p_1) = q_1 p_2,$$
$$p_{00} = q_1 - p_{01} = q_1(1 - p_2) = q_1 q_2,$$

что и завершает проверку. Аналогичную проверку для трех независимых событий предоставляем сделать читателю (упр. 6 в конце главы).

§ 2.7. Понятие функции дискретных случайных величин

Под функцией $f(X)$ случайной величины X понимают такую случайную величину Y, которая принимает значение $y = f(x)$ каждый раз, когда величина X принимает значение x. Аналогично вводится понятие функции от нескольких случайных величин. При этом, конечно, предполагается, что рассматриваемая функция определена для всех возможных значений аргументов.

Рассмотрим следующую задачу: найти распределение вероятностей функции $Y = f(X)$ от дискретной случайной величины X при условии, что для различных возможных значений x_i величины X значения функции $f(x_i)$ также различны. Возможными значениями случайной величины Y будут значения функции $f(x_i)$, причем величина Y примет значение $f(x_i)$ тогда и только тогда, когда величина X примет значение x_i; поэтому вероятности этих событий равны:

$$\mathsf{P}(Y = f(x_i)) = \mathsf{P}(X = x_i) = p_i \qquad (i = 1,\ 2, \ldots).$$

Следовательно, в рассматриваемой задаче таблицы распределения вероятностей величин X и $Y = f(X)$ отличаются только строкой возможных значений:

X	x_1	x_2	...
	p_1	p_2	...

Y	$f(x_1)$	$f(x_2)$	...
	p_1	p_2	...

(1)

Подобный случай встречался нам в § 2.3 при переходе от распределения (2.3-2) к распределению (2.3-3).

Рассмотрим несколько примеров.

Пример 1. Случайная величина X имеет таблицу распределения вероятностей

X	0	1	2
	1/2	1/3	1/6

Найти распределения вероятностей величин $X - 1$, X^2, $(X - 1)^2$.

Решение. Функции $X - 1$ и X^2 принимают в точках $X = 0, 1, 2$ различные значения, поэтому их таблицы распределения находятся по правилу (1):

$X - 1$	-1	0	1
	1/2	1/3	1/6

X^2	0	1	4
	1/2	1/3	1/6

Функция $(X - 1)^2$ принимает равные значения в точках $X = 0$ и $X = 2$, так что событие $(X - 1)^2 = 1$ есть сумма двух событий $X = 0$ и $X = 2$; по правилу сложения вероятностей получаем

$$\mathsf{P}((X - 1)^2 = 1) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}.$$

Таблица распределения вероятностей для $(X-1)^2$ будет иметь вид

$(X-1)^2$	0	1
	1/3	2/3

Пример 2. Найти распределение любой положительной степени индикатора случайного события A.

Решение. Индикатор I_A случайного события A принимает только два возможных значения 1 и 0:

I_A	1	0
	p	q

, где $p = \mathsf{P}(A)$, $q = \mathsf{P}(\bar{A}) = 1 - p$.

Поэтому любая его положительная степень имеет то же самое распределение вероятностей, так как $1^n = 1$, $0^n = 0$ $(n > 0)$.

Пример 3. Бросаются одновременно две правильные игральные кости. Найти распределение суммы очков на их верхних гранях.

Решение. Числа очков, выпадающие на верхних гранях первой и второй кости, представляют собой, очевидно, независимые случайные величины X_1 и X_2 с одинаковыми распределениями вероятностей

X_1	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

X_2	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

Поэтому закон их совместного распределения задается формулами

$$\mathsf{P}(X_1 = i, X_2 = j) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36} \quad (i, j = 1, 2, 3, 4, 5, 6).$$

Сумма $X_1 + X_2$ может принимать значения от 2 до 12, но некоторые значения она будет принимать при различных комбинациях значений X_1 и X_2, например, событие $X_1 + X_2 = 4$ может произойти

$$\begin{aligned} &\text{при } X_1 = 1, \quad X_2 = 3 \\ &\text{или при } X_1 = 2, \quad X_2 = 2, \\ &\text{или при } X_1 = 3, \quad X_2 = 1. \end{aligned}$$

По правилу сложения вероятностей надо складывать вероятности всех соответствующих комбинаций. Так как все комбинации имеют одну и ту же вероятность 1/36, надо умножить эту вероятность на число комбинаций. Получаем следующую таблицу распределения вероятностей суммы $X_1 + X_2$ (табл. 4).

Таблица 4

$X_1 + X_2$	2	3	4	5	6	7	8	9	10	11	12
	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Полезно обратить внимание на то, что хотя распределения величин X_1 и X_2 одинаковы, распределение их суммы $X_1 + X_2$ заметно отличается от распределения удвоенной величины $2X_1$.

Пример 4. Производятся два независимых испытания, в которых успех может появиться с вероятностями p_1 и p_2. Найти распределение вероятностей числа X успехов за два испытания.

Решение. Число X успехов за два испытания может принимать значения 0, 1 и 2. Для нахождения закона распределения вероятностей можно было бы воспользоваться правилами сложения и умножения вероятностей. Но удобнее представить случайную величину X в виде суммы двух более простых случайных величин — индикаторов успехов в первом и во втором испытаниях. Обозначим эти индикаторы через X_1 и X_2 (событие $X_1 = 1$ означает успех в первом испытании). Распределения их вероятностей даются таблицами

X_1	1	0
	p_1	q_1

X_2	1	0
	p_2	q_2

Независимость испытаний означает независимость этих индикаторов.

Интересующая нас случайная величина X равна сумме индикаторов X_1 и X_2:

$$X = X_1 + X_2,$$

так как эта сумма состоит из единиц и нулей, причем единиц в ней ровно столько, сколько раз наступит успех за два испытания.

Сказанное выше позволяет легко получить таблицу распределения величины X:

X	2	1	0
	p_1p_2	$p_1q_2 + p_2q_1$	q_1q_2

Приведенный здесь метод обобщается на любое число независимых испытаний. В частности, если вероятность успеха во всех испытаниях одна и та же и равна p, то при двух испытаниях получаем распределение величины $X = X_1 + X_2$ в виде

X	2	1	0
	p^2	$2pq$	q^2

При трех испытаниях получаем распределение величины $X = (X_1 + X_2) + X_3$ в виде

X	3	2	1	0
	p^3	$3p^2q$	$3pq^2$	q^3

и т. д. Методом математической индукции мы приходим снова к биномиальному распределению (см. § 2.3).

УПРАЖНЕНИЯ

1. Из урны, в которой лежат 2 белых и 8 черных шаров, вынимают 3 шара. Найти распределение вероятностей числа X вынутых белых шаров. Сравнить это гипергеометрическое распределение с соответствующим биномиальным распределением (когда шары после каждого вынимания возвращаются в урну).

2. Испытание заключается в бросании трех правильных игральных костей. Найти распределение вероятностей суммы очков, выпадающих на всех трех костях.

3. В урне лежат 20 шаров — 10 белых и 10 черных. Из урны вынимают 4 шара по схеме повторной выборки. Найти вероятность того, что не менее двух вынутых шаров будут белыми.

4. Тот же вопрос для вынимания шаров по схеме бесповторной выборки.

5. Что вероятнее выиграть у равносильного противника:

а) три партии из четырех или пять из восьми?

б) не менее трех партий из четырех или не менее пяти из восьми?

Под равносильным противником понимается такой, вероятность выиграть у которого равна 1/2.

6*. Проверить, что из независимости трех случайных событий вытекает независимость их индикаторов (ср. конец § 2.6).

7*. Независимые случайные величины X и Y имеют биномиальные законы распределения вероятностей с одним и тем же параметром p:

$$\mathsf{P}(X = m) = C_{n_1}^m p^m q^{n_1-m} \quad (m = 0, 1, 2, \ldots, n_1),$$
$$\mathsf{P}(Y = m) = C_{n_2}^m p^m q^{n_2-m} \quad (m = 0, 1, 2, \ldots, n_2),\ q = 1 - p.$$

Найти закон распределения их суммы $Z = X + Y$.

8*. Независимые случайные величины X и Y следуют законам распределения Пуассона с параметрами a и b. Доказать, что их сумма $Z = X + Y$ следует закону распределения Пуассона с параметром $c = a + b$.

ГЛАВА 3

НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 3.1. Непрерывные одномерные случайные величины. Плотность распределения вероятностей

В теории вероятностей и ее приложениях часто приходится иметь дело с такими случайными величинами, возможные значения которых сплошь заполняют некоторый интервал (α, β); таковы, например, ошибки измерения. Закон распределения вероятностей такой величины X должен позволять находить вероятность попадания ее значения в любой интервал (x_1, x_2), лежащий внутри (α, β); будем обозначать эту вероятность через $\mathsf{P}(x_1 < X < x_2)$.

Мы ограничимся только такими случайными величинами X, для которых вероятность попадания в интервал $(x, x + \Delta x)$ малой длины $\Delta x > 0$ можно приближенно считать пропорциональной длине этого интервала:

$$\mathsf{P}(x < X < x + \Delta x) \approx p(x)\,\Delta x \tag{1}$$

Приближенное равенство (1) надо понимать в том смысле, что при $\Delta x \to 0$ его левая часть отличается от правой на бесконечно малую величину более высокого порядка, чем Δx, т. е. $\mathsf{P}(x < X < x + \Delta x) = p(x)\,\Delta x + o(\Delta x)$. Точный смысл этого выражения дается предельным соотношением

$$\lim_{\Delta x \to 0} \frac{\mathsf{P}(x < X < x + \Delta x)}{\Delta x} = p(x). \tag{2}$$

Функция $p(x)$ в написанных соотношениях называется плотностью распределения вероятностей случайной величины X (или, короче, *плотностью распределения величины* X), а произведение $p(x)\,\Delta x = p(x)\,dx$ называется *элементом вероятности*

По элементу вероятности $p(x)\,dx$ с помощью интегрирования можно найти вероятность попадания значения X в любой интервал (x_1, x_2):

$$\boxed{\mathsf{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} p(x)\,dx.} \tag{3}$$

Определение. *Величина X называется непрерывной случайной величиной, если вероятность попадания ее значения в любой интервал (x_1, x_2) может быть представлена в виде интеграла* (3) *от некоторой функции $p(x)$ — плотности распределения вероятностей.* Подчеркнем, что плотность распределения вероятностей вполне определяет закон распределения непрерывной случайной величины X (или, как мы будем говорить, непрерывный закон распределения). При этом функция $p(x)$ должна быть *неотрицательной* (что связано с неотрицательностью вероятностей) и должна быть *нормирована условием*

$$\int_{-\infty}^{\infty} p(x)\,dx = 1, \tag{4}$$

отражающим достоверность события $(-\infty < X < \infty)$. Если все возможные значения случайной величины X сосредоточены в конечном интервале (α, β), то мы будем считать, что вне этого интервала плотность $p(x) \equiv 0$ и, значит, условие (4) сводится к условию

$$\int_{\alpha}^{\beta} p(x)\,dx = 1.$$

Что же касается основного свойства аддитивности вероятности, то для непрерывной случайной величины оно совпадает с известным свойством аддитивности интеграла: если точка x_3 лежит в интервале (x_1, x_2), то

$$\mathsf{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} p(x)\,dx = \int_{x_1}^{x_3} p(x)\,dx + \int_{x_3}^{x_2} p(x)\,dx =$$
$$= \mathsf{P}(x_1 < X < x_3) + \mathsf{P}(x_3 < X < x_2).$$

Следует подчеркнуть, что для непрерывной случайной величины имеет смысл рассматривать только такое событие, как попадание в интервал, а не попадание в отдельную точку. Так как вероятность попадания в малый интервал, по определению, пропорциональна длине интервала (с точностью до малых более высокого порядка), то вероятность попадания непрерывной случайной величины в любую заранее заданную точку равна нулю. В связи с этим, в частности, мы должны в случае непрерывного распределения считать равными следующие вероятности:

$$\mathsf{P}(x_1 < X < x_2) = \mathsf{P}(x_1 < X \leqslant x_2) = \mathsf{P}(x_1 \leqslant X \leqslant x_2).$$

Примечание. В схеме урн, где вероятности равны отношениям целых чисел M/N, равенство нулю вероятности случайного события равносильно его невозможности. Для непрерывной случайной величины попадание в точку неудобно считать невозможным событием, хотя его вероятность равна нулю. На практике это не приводит к недоразумениям, так как значение любой физической величины можно измерить лишь с некоторой ошибкой, что приводит к «размазыванию» этого значения в некотором интервале.

Кривая распределения вероятностей. Наглядное представление о непрерывном законе распределения вероятностей можно получить по графику плотности распределения $p(x)$, который называется кривой распределения вероятностей величины X (рис. 8). Этот график позволяет иллюстрировать вероятность $\mathsf{P}(x_1 < X < x_2)$, так как площадь заштрихованной на рис. 8 криволинейной трапеции выражается тем же интегралом (3), что и указанная вероятность.

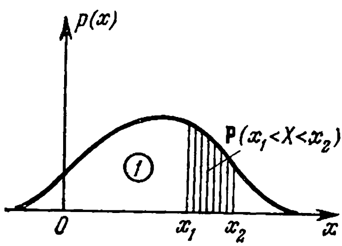


Рис. 8.

§ 3.2. Примеры непрерывных законов распределения

Равномерное распределение в интервале (α, β). Говорят, что случайная величина X распределена равномерно в конечном интервале (α, β), если все ее возможные значения сосредоточены на этом интервале и если плотность распределения ее вероятностей на этом интервале постоянна. Если эту постоянную обозначить буквой C, то плотность равномерного распределения задается формулой

$$\boxed{p(x)=\begin{cases} C & \text{при } \alpha<x<\beta, \\ 0 & \text{вне интервала } (\alpha,\ \beta).\end{cases}} \tag{1}$$

Для случайной величины X, равномерно распределенной в интервале (α, β), вероятность попадания в любой интервал (x_1, x_2), лежащий внутри интервала (α, β), пропорциональна длине этого интервала (см. рис. 9):

$$\mathsf{P}\,(x_1 < X < x_2) = C\,(x_2 - x_1). \tag{2}$$

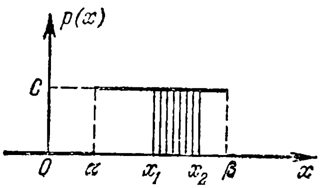


Рис. 9.

Параметр C определяется из условия нормировки:

$$\mathsf{P}\,(\alpha < X < \beta) = C\,(\beta - \alpha) = 1,$$

откуда

$$C = \frac{1}{\beta - \alpha}.$$

Подставляя полученное значение C в формулу (2), мы находим, что вероятность попадания значения величины X в интервал (x_1, x_2) равна отношению длины этого интервала к длине всего интервала (α, β):

$$\boxed{\mathsf{P}\,(x_1 < X < x_2) = \frac{x_2 - x_1}{\beta - \alpha}.} \tag{3}$$

Заметим, что обычно употребляемое выражение «выберем точку X наудачу в интервале (α, β)» означает, что рассматриваемая точка X представляет собой случайную величину с равномерным распределением вероятностей в интервале (α, β). Точно так же, если говорят, что выбирается наудачу направление на плоскости, то имеют в виду, что выбираемый угол есть случайная величина с равномерным распределением вероятностей в интервале $(0, 2\pi)$.

Показательное распределение. К показательному закону распределения вероятностей приводит задача о распределении промежутка времени T между двумя последовательными событиями в простейшем потоке, описанном в § 2.4. Случайная величина T может принимать только положительные значения $t > 0$. Обозначим плотность ее распределения через $p(t)$, тогда $p(t) = 0$ при $t \leqslant 0$. Пусть теперь $t > 0$. Подсчитаем вероятность того, что промежуток времени между двумя последовательными событиями меньше t; это означает, что за время t наступит хотя бы одно событие простейшего потока. Используя формулу (2.4-5), получим

$$\mathsf{P}(0 < T < t) = \mathsf{P}(X \geqslant 1) = 1 - \mathsf{P}(X = 0) = 1 - e^{-\lambda t} \qquad (4)$$

(здесь как и в § 2.4, через X обозначено число событий, наступающих за время t, а через λ — среднее число событий за единицу времени). Последнее соотношение можно представить в форме

$$\mathsf{P}(0 < T < t) = \int\limits_0^t \lambda e^{-\lambda t}\, dt,$$

откуда следует, что искомая плотность распределения вероятностей величины T равна

$$\boxed{p(t) = \lambda e^{-\lambda t} \quad \text{при} \quad t > 0.} \qquad (5)$$

Легко проверить, что условие нормировки здесь выполнено:

$$\int\limits_{-\infty}^{\infty} p(t)\, dt = \int\limits_0^{\infty} \lambda e^{-\lambda t}\, dt = 1.$$

Закон распределения вероятностей с плотностью (5) называется показательным; он встречается во многих задачах теории массового обслуживания. Кривая распределения вероятностей для этого закона представлена на рис. 10. Заметим, что хотя при показательном распределении возможны любые положительные значения величины T, но большие значения маловероятны из соотношения

$$P(T > t) = 1 - P(0 < T < t) = e^{-\lambda t}$$

видно, что вероятность $P(T > t)$ очень быстро убывает с ростом t; уже при $\lambda t = 7$ эта вероятность будет меньше 0,001.

Интересно отметить, что показательное распределение может служить непрерывным аналогом геометрического распределения (для дискретной величины, см § 2.2). В рассмотренной выше задаче о распределении промежутка времени между двумя последовательными событиями в простейшем потоке будем измерять время дискретно, а именно, регистрировать наступления события только в моменты времени, кратные Δt. Подсчитаем вероятность того, что, начиная отсчет времени от последнего происшедшего события, мы зарегистрируем следующее событие в момент $t_m = m\Delta t$; это есть вероятность того, что промежуток времени между событиями заключен между $t_{m-1} = (m - 1)\Delta t$ и t_m:

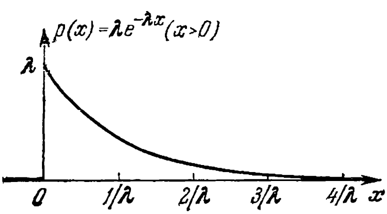


Рис. 10.

$$P(t_{m-1} < T < t_m) = \int_{t_{m-1}}^{t_m} \lambda e^{-\lambda t}\, dt = e^{-\lambda t_{m-1}} - e^{-\lambda t_m} =$$
$$= e^{-\lambda(m-1)\Delta t}(1 - e^{-\lambda \Delta t}) \qquad (m = 1, 2, \ldots).$$

Эту вероятность можно представить в виде

$$P(t_{m-1} < T < t_m) = q^{m-1}p,$$

где

$$p = 1 - e^{-\lambda\Delta t} = \int_0^{\Delta t} \lambda e^{-\lambda t}\, dt$$

есть вероятность регистрации события в момент времени Δt, т. е. вероятность того, что за промежуток времени Δt произойдет одно событие нашего потока (возможность наступления двух событий за промежуток времени

Δt мы здесь исключаем), а

$$q = e^{-\lambda \Delta t} = 1 - p.$$

Полученная формула совпадает с формулой для вероятностей в геометрическом распределении.

Обобщением показательного распределения является г а м м а - р а с п р е д е л е н и е. Плотность этого распределения определяется так:

$$\boxed{p(x) = Cx^{\alpha-1}e^{-\lambda x} \quad \text{при} \quad x > 0} \tag{6}$$

и $p(x)=0$ при $x \leqslant 0$. Параметры α и λ здесь могут быть любыми

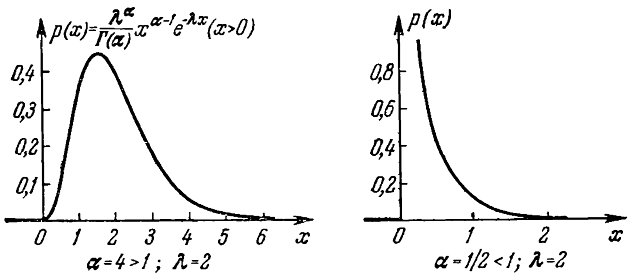


Рис. 11.

положительными числами Множитель C определяется из условия нормировки.

$$\mathrm{P}(0 < X < \infty) = \int_0^\infty Cx^{\alpha-1}e^{-\lambda x}\,dx = 1.$$

Сравнивая интеграл в последнем соотношении с гамма-функцией Эйлера

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1}e^{-z}\,dz,$$

замечаем, что условие нормировки можно записать в виде

$$C \cdot \frac{1}{\lambda^\alpha}\Gamma(\alpha) = 1,$$

откуда

$$C = \frac{\lambda^\alpha}{\Gamma(\alpha)}. \tag{7}$$

На рис. 11 показан вид кривых распределения вероятностей при значениях параметра $\alpha > 1$ и $\alpha < 1$ (при $\alpha = 1$ получаем показательное распределение, см. рис. 10). При значениях параметра $\alpha > 1$ кривая распределения имеет один максимум. Точку максимума найдем обычным способом, приравнивая нулю производную:

$$p'(x) = Ce^{-\lambda x}\left[(\alpha-1)x^{\alpha-2} - \lambda x^{\alpha-1}\right] = 0,$$

откуда

$$x = \frac{\alpha - 1}{\lambda} \qquad (\alpha > 1) \tag{8}$$

(на рис. 11 приняты значения параметров $\alpha = 4$, $\lambda = 2$, что дает для точки максимума значение $x = 3/2$). При значениях параметра $\alpha < 1$ плотность распределения убывает в интервале $(0, \infty)$. Три из задач, приводящих к гамма-распределению, даны в § 3.8, примеры 3, 4 и 5.

Примечание. Из свойств гамма-функции нам понадобятся лишь следующие легко проверяемые соотношения:

а) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$;

б) $\Gamma(1) = \int\limits_0^\infty e^{-z}\,dz = 1$;

в) для любого целого $n > 0$

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\ldots 2\cdot 1\Gamma(1) = n!;$$

г) $$\Gamma\left(\frac{1}{2}\right) = \int\limits_0^\infty e^{-z}\frac{dz}{\sqrt{z}} = 2\int\limits_0^\infty e^{-u^2}\,du = \sqrt{\pi};$$

д) для любого целого $n > 0$

$$\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\Gamma\left(n-\frac{1}{2}\right) = \frac{1\cdot 3\cdot 5\ldots(2n-1)}{2^n}\sqrt{\pi},$$

например, $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\sqrt{\pi}$.

В тех случаях, когда α отлично от целого числа или целого числа с половиной, значения $\Gamma(\alpha)$ можно найти в таблицах гамма-функции (которые приводятся во многих справочниках).

§ 3.3. Нормальный закон распределения

Среди всех непрерывных законов распределения вероятностей особую роль играет распределение вероятностей с плотностью

$$\boxed{\begin{array}{l}\varphi_{a;\,\sigma}(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp\left[-\dfrac{(x-a)^2}{2\sigma^2}\right],\\ \text{где } \exp(t) = e^t \quad (-\infty < x < \infty).\end{array}} \tag{1}$$

Такое распределение называется *нормальным распределением с параметрами* a, σ $(\sigma > 0)$. О случайной величине X с таким законом распределения вероятностей говорят, что она распределена нормально с параметрами a, σ. Множитель $\frac{1}{\sigma\sqrt{2\pi}}$ выбран так, чтобы удовлетворить условию нормировки:

$$\int_{-\infty}^{\infty} \varphi_{a;\,\sigma}(x)\,dx = 1.$$

Важность нормального закона распределения определяется тем, что к нему обычно приводят задачи, связанные с распределением сумм большого числа случайных величин. Откладывая рассмотрение таких задач до главы 6, мы отметим здесь только формулы для вычисления вероятностей в нормальном распределении и некоторые свойства соответствующей кривой распределения вероятностей (*кривой нормального распределения*).

Прежде всего выделим **стандартное нормальное распределение** с параметрами $a = 0$, $\sigma = 1$, т. е. с плотностью

$$\boxed{\varphi(x) = \varphi_{0;\,1}(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right).} \qquad (2)$$

Кривая стандартного нормального распределения вероятностей изображена на рис. 12. Она симметрична

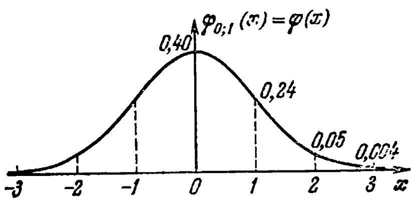


Рис. 12.

относительно оси ординат (в силу четности функции $\varphi(x)$), при $x = 0$ имеет единственный максимум, равный $\frac{1}{\sqrt{2\pi}} \approx 0{,}40$, и имеет две точки перегиба при $x = \pm 1$. При $x \to \pm\infty$ кривая распределения асимптотически приближается к оси абсцисс, причем приближается весьма быстро (например, уже $\varphi(3) = 0{,}0044$, $\varphi(4) = 0{,}00013$).

Интеграл от плотности $\varphi(x)$ не выражается через элементарные функции. Для расчета вероятностей случайных величин с нормальным распределением составлены таблицы специальной функции

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}}\, dx, \tag{3}$$

называемой *интегралом вероятностей*. В таблице 1 Приложения даются значения $\Phi(t)$ только для положительных значений t; значения $\Phi(t)$ для отрицательных значений t вычисляются с помощью свойства нечетности этой функции:

$$\Phi(-t) = \frac{2}{\sqrt{2\pi}} \int_0^{-t} \exp\left(-\frac{x^2}{2}\right) dx = -\Phi(t). \tag{4}$$

График функции $\Phi(t)$ приведен на рис. 13. При изменении t от 0 до $+\infty$ функция $\Phi(t)$ возрастает от значения $\Phi(0) = 0$ до значения $\Phi(+\infty) = 1$, причем возрастает очень быстро: уже $\Phi(3) = 0{,}9973$, $\Phi(4) = 0{,}999937$. С помощью интеграла вероятностей $\Phi(t)$ можно вычислить вероятность попадания случайной величины X со стандартным нормальным распределением в любой интервал (x_1, x_2):

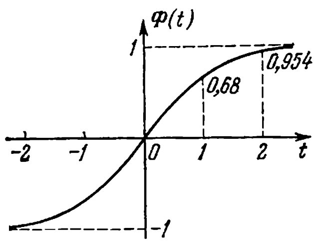


Рис. 13.

$$\mathsf{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} \varphi(x)\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{x_2} \exp\left(-\frac{x^2}{2}\right) dx - \frac{1}{\sqrt{2\pi}} \int_0^{x_1} \exp\left(-\frac{x^2}{2}\right) dx,$$

т. е.

$$\boxed{\mathsf{P}(x_1 < X < x_2) = \frac{1}{2}[\Phi(x_2) - \Phi(x_1)].} \tag{5}$$

В частности, для симметричного интервала $(-t, t)$

получаем

$$\mathsf{P}(-t < X < t) = \frac{1}{2}[\Phi(t) - \Phi(-t)] = \Phi(t). \quad (6)$$

Последнее соотношение позволяет толковать функцию $\Phi(t)$ при $t > 0$ как вероятность попадания в симметричный интервал $(-t, t)$ для случайной величины со стандартным нормальным законом распределения.

Перейдем теперь к общему нормальному распределению с параметрами a и σ. Влияние параметра σ на кривую нормального распределения видно из рис. 14.

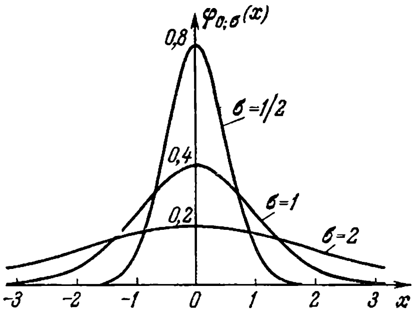


Рис. 14.

На этом рисунке приведены кривые нормального распределения при $a = 0$ и при значениях $\sigma = {}^1/_2$, 1 и 2. С уменьшением параметра σ кривая распределения сжимается вдоль оси абсцисс и вытягивается вдоль оси ординат, что приводит к увеличению вероятности попадания случайной величины X в любую фиксированную окрестность точки $x = 0$. Таким образом, *параметр σ характеризует рассеяние* случайной величины X; он называется **стандартным отклонением** или **стандартом**, а также **средним квадратическим отклонением** (смысл последнего названия мы поясним подробно далее в § 4.4). Параметр a определяет *сдвиг* вдоль оси абсцисс, как видно из формулы (1) и из рис. 15. При таком сдвиге кривая распределения вероятностей оказывается симметричной относительно прямой $x = a$, так что значение a является центром симметрии распределения.

Расчет вероятностей в общем нормальном распределении производится с помощью интеграла вероятностей $\Phi(t)$ следующим

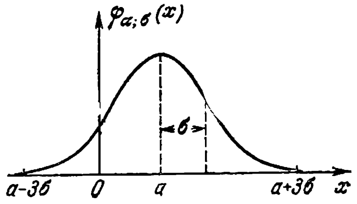


Рис. 15.

образом. Вероятность попадания случайной величины X в интервал (x_1, x_2) равна интегралу

$$\mathsf{P}(x_1 < X < x_2) = \int_{x_1}^{x_2} \varphi_{a;\,\sigma}(x)\,dx =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] dx;$$

с помощью замены $\frac{x-a}{\sigma} = u$, $\frac{dx}{\sigma} = du$ этот интеграл приводится к виду

$$\left.\begin{aligned} &\boxed{\begin{aligned} &\mathsf{P}(x_1 < X < x_2) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} \exp\left(-\frac{u^2}{2}\right) du = \frac{1}{2}\,[\Phi(t_2) - \Phi(t_1)], \end{aligned}} \\ &\text{где} \\ &t_1 = \frac{x_1 - a}{\sigma}, \quad t_2 = \frac{x_2 - a}{\sigma}. \end{aligned}\right\} \qquad (7)$$

В частности, вероятность отклонения величины X от центра a меньше, чем на $t\sigma$, равна

$$\mathsf{P}(|X - a| < t\sigma) = \mathsf{P}(a - t\sigma < X < a + t\sigma) = \Phi(t); \qquad (8)$$

это соотношение еще раз подчеркивает роль стандарта

σ как характеристики рассеяния случайной величины X. Например,

$$\mathsf{P}\,(|\,X - a\,| < 3\sigma) = \Phi\,(3) = 0{,}9973 > 1 - 0{,}003. \qquad (9)$$

В приложениях формулу (9) обычно трактуют следующим образом: для нормально распределенной случайной величины практически достоверно, что ее отклонения от центра окажутся меньше утроенного стандартного отклонения (так называемое *«правило трех сигм»*). Это означает, разумеется, что здесь пренебрегают возможностью появления событий, имеющих вероятности меньшие, чем 0,003.

§ 3.4. Функция распределения вероятностей

Законы распределения вероятностей дискретных и непрерывных величин мы задаем разными способами: для дискретных величин — вероятностями p_i отдельных возможных значений x_i, для непрерывных величин — плотностью распределения вероятностей $p\,(x)$ или элементом вероятности $p\,(x)\,dx$.

Можно ввести единый способ задания законов распределения вероятностей с помощью функции распределения вероятностей $F\,(x)$, которая определяется формулой

$$F\,(x) = \mathsf{P}\,(X < x), \qquad (1)$$

т. е. равна вероятности того, что случайная величина X примет значение, меньшее некоторого числа x.

Для непрерывной величины X в соответствии с основной формулой (3.1-3) имеем

$$F(x) = \int\limits_{-\infty}^{x} p(x)\,dx; \qquad (2)$$

отсюда следует, в частности, что производная от функции распределения вероятностей равна плотности распределения (в точках ее непрерывности)

$$F'\,(x) = p\,(x), \qquad (3)$$

а введенный на стр. 69 элемент вероятности есть дифференциал функции распределения

$$p\,(x)\,dx = dF\,(x).$$

Для дискретной величины X функция распределения вероятностей равна сумме вероятностей тех ее значений, которые меньше x:

$$F(x) = \sum_{x_k < x} p_k. \tag{4}$$

Например, для случайной величины с таблицей распределения

X	1	2	3
	0,3	0,5	0,2

функция распределения вероятностей имеет вид

$$F(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 1, \\ 0{,}3 & \text{при} \quad 1 < x \leqslant 2, \\ 0{,}8 & \text{при} \quad 2 < x \leqslant 3, \\ 1 & \text{при} \quad x > 3. \end{cases}$$

Для удобства записи таких функций вводят специальную единичную функцию

$$\eta(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 0, \\ 1 & \text{при} \quad x > 0. \end{cases}$$

Тогда функция распределения вероятностей для рассматриваемого примера запишется в виде

$$F(x) = 0{,}3\eta(x-1) + 0{,}5\eta(x-2) + 0{,}2\eta(x-3),$$

а в общем случае в виде

$$F(x) = \sum_i p_i \eta(x - x_i), \tag{5}$$

где сумма берется по всем возможным значениям x_i.

Из определения функции распределения и основных свойств вероятностей вытекает, что $F(x)$ есть неубывающая функция, изменяющаяся от значения $F(-\infty) = 0$ до значения $F(+\infty) = 1$.

Ее график называют *интегральной кривой распределения*. На рис. 16, 17, 18 и 19 приведены интегральные кривые распределения соответственно для равномерного распределения в интервале (α, β), для показательного распределения с параметром λ, для стандартного нормального распределения и для рассмотренного

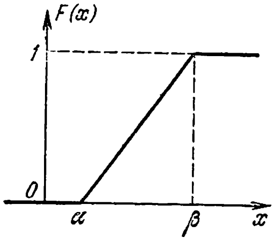


Рис. 16.

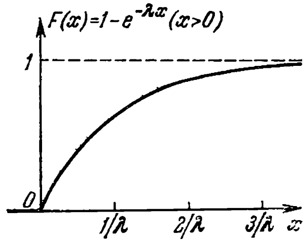


Рис. 17.

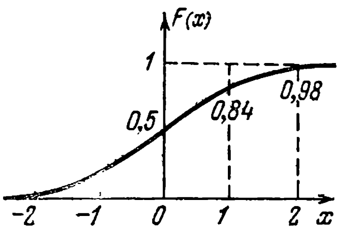


Рис. 18.

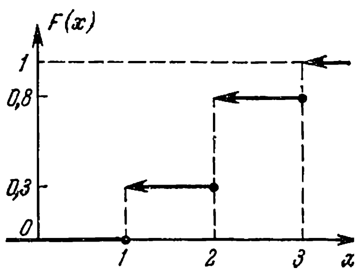


Рис. 19.

на стр. 82 примера дискретной случайной величины. Отметим, что для стандартного нормального закона функция распределения $F(x)$ связана с интегралом вероятностей $\Phi(x)$ простым соотношением

$$F(x) = \mathrm{P}(-\infty < X < x) = \frac{1}{2}[\Phi(x) - \Phi(-\infty)] = \frac{1}{2}[\Phi(x) + 1].$$

Во многих справочниках вместо таблицы интеграла вероятностей $\Phi(x)$ приводится таблица функции распределения $F(x)$ стандартного нормального закона.

Через функцию распределения вероятностей $F(x)$ можно выразить вероятность $\mathsf{P}(x_1 \leqslant X < x_2)$ следующим образом:

$$\mathsf{P}(x_1 \leqslant X < x_2) = \mathsf{P}(X < x_2) - \mathsf{P}(X < x_1) = F(x_2) - F(x_1).$$

§ 3.5. Эмпирическая функция распределения

При сопоставлении теории с опытом мы сравниваем вероятности с относительными частотами. Так как для непрерывной случайной величины X вероятности определяются для интервалов, а любые серии повторения испытания дают нам лишь наборы отдельных значений случайной величины $(x_1, x_2, \ldots, x_n)$, то для сопоставления предполагаемого теоретического распределения вероятностей с эмпирическим распределением относительных частот надо вычислять относительные частоты тоже для интервалов. Один из способов состоит в сравнении функции распределения $F(x)$, которая для любого x дает вероятность попадания в интервал $(-\infty, x)$, с так называемой эмпирической функцией распределения $F_n(x)$, равной относительной частоте тех эмпирических значений x_i, которые попадают в тот же интервал; эмпирическую функцию распределения можно записать в виде

$$F_n(x) = \frac{1}{n}[\eta(x - x_1) + \eta(x - x_2) + \ldots + \eta(x - x_n)]. \quad (1)$$

Действительно, из определения единичной функции $\eta(x)$ (§ 3.4) следует, что в квадратных скобках в формуле (1) будет столько единиц, сколько значений x_i окажутся меньшими, чем x; отношение этого числа к числу n повторений испытания дает относительную частоту события $X < x$. График эмпирической функции распределения (рис. 20, *а*) представляет собой ступенчатую линию со скачками величины $1/n$ в точках $x_1, x_2, \ldots, x_n$ (разумеется, если, скажем, два из значений x_i совпадают, то величина соответствующего скачка на самом деле будет уже $2/n$).

При больших значениях n мы вправе ожидать близости относительных частот к соответствующим вероятностям и, следовательно, незначительного расхождения между теоретической и эмпирической функциями распределения.

Другой способ сравнения теоретического и эмпирического распределений состоит в следующем. Весь диапазон (α, β) изменения случайной величины X разбивают на l интервалов точками $\gamma_1, \gamma_2, \ldots, \gamma_{l-1}$ $(\alpha = \gamma_0 < \gamma_1 < \gamma_2 < \ldots < \gamma_{l-1} < \gamma_l = \beta)$ и подсчитывают относительные частоты m_j/n значений x_i, которые попадают

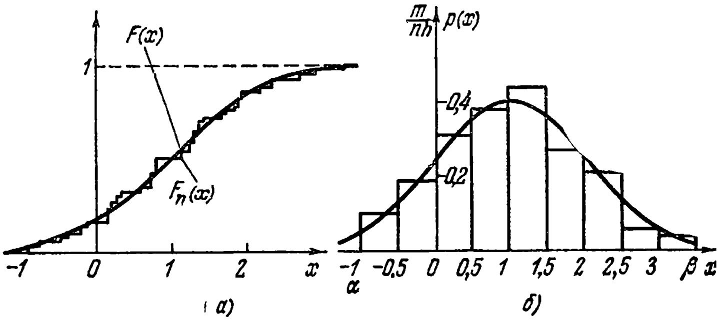


Рис. 20.

в каждый из интервалов $(\gamma_{j-1}, \gamma_j]$ $(j = 1, 2, \ldots, l)$. Эти относительные частоты сопоставляют с вероятностями p_j попадания в те же интервалы:

$$p_j = \mathrm{P}(\gamma_{j-1} < X \leqslant \gamma_j) = \int_{\gamma_{j-1}}^{\gamma_j} p(x)\,dx,$$

где $p(x)$ — плотность предполагаемого теоретического распределения.

Если предполагаемое теоретическое распределение хорошо согласуется с опытом (адекватно), то при достаточно большом числе n опытов относительные частоты m_j/n должны быть приближенно равны соответствующим вероятностям p_j $(j = 1, 2, \ldots, l)$.

Для наглядного графического сопоставления предполагаемого теоретического распределения с эмпирическим кривую распределения, т. е. график плотности $p(x)$, сравнивают с *гистограммой*, которая строится следующим образом. Каждую относительную частоту m_j/n делят на длину интервала $(\gamma_{j-1}, \gamma_j]$ и полученную *среднюю плотность относительной частоты попадания в интервал* относят к этому интервалу (на рис. 20, *б* длины всех интервалов выбраны одинаковыми, равными h, что позволяет на оси ординат откладывать значения m_j/nh). При таком построении относительные частоты попадания в интервалы будут равны площадям соответствующих столбиков гистограммы, подобно тому, как вероятности равны площадям соответствующих криволинейных трапеций Если предполагаемое теоретическое распределение хорошо согласуется с опытом, то при достаточно большом n и удачном выборе интервалов (γ_{j-1}, γ_j) гистограмма будет близка к кривой распределения. Иногда еще для наглядности сравнения строят *многоугольник распределения*, соединяя последовательно середины верхних оснований столбиков гистограммы.

Во всех случаях при обнаружении больших расхождений между предполагаемым теоретическим и эмпирическим распределениями возникает сомнение в правильности определения плотности распределения $p(x)$. Как уже упоминалось в § 2.5, для решения вопроса о допустимости тех или иных расхождений существуют критерии согласия, излагаемые в курсах математической статистики. Отметим только, что применение критериев согласия требует, как правило, большого числа опытных данных (порядка сотен и даже тысяч).

§ 3.6. Многомерные случайные величины

В § 1.1 мы уже упоминали о случайной точке попадания в мишень. Если координаты этой точки в плоскости мишени обозначить через X и Y, мы получим двумерную случайную величину $(X; Y)$ Иногда удобнее говорить о случайном радиусе-векторе или о системе двух случайных величин X и Y Закон распределения вероятностей непрерывной двумерной случайной величины $(X; Y)$ должен давать вероятность попадания ее значений $(x; y)$ в любую область на плоскости По аналогии с одномерными непрерывными величинами дадим следующее

Определение. *Величина* $(X;\ Y)$ *называется непрерывной двумерной случайной величиной, если вероятность попадания ее значения в любую область* D *на плоскости* $(x;\ y)$ *может быть представлена в виде двойного интеграла*

$$\boxed{\mathsf{P}((X;\ Y) \in D) = \iint\limits_{D} p(x,\ y)\,dx\,dy,} \qquad (1)$$

где $p(x,\ y)$ — некоторая неотрицательная функция, нормированная условием

$$\iint\limits_{-\infty}^{\infty} p(x,\ y)\,dx\,dy = 1. \qquad (2)$$

Здесь интеграл берется по всей плоскости; если же все возможные значения величины $(X;\ Y)$ сосредоточены лишь в области D_0, то вне этой области полагаем $p(x,\ y) = 0$.

Функция $p(x,\ y)$ называется *плотностью двумерного распределения вероятностей* или *плотностью совместного распределения случайных величин* X и Y. Элемент вероятности

$$p(x,\ y)\,dx\,dy$$

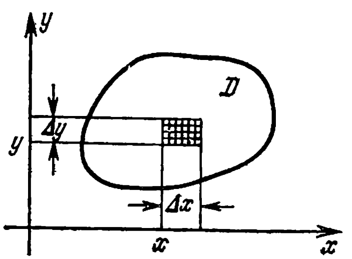


Рис. 21.

дает главную часть вероятности попадания точки $(X;\ Y)$ в прямоугольник, заштрихованный на рис. 21, т. е. главную часть вероятности совмещения случайных событий

$$x < X < x + \Delta x, \qquad y < Y < y + \Delta y$$
$$(dx = \Delta x;\ \ dy = \Delta y)$$

График функции $z = p(x,\ y)$ называется *поверхностью распределения вероятностей*.

Пример 1. Равномерное распределение в области D_0. Случайная величина $(X;\ Y)$ равномерно распределена в области D_0, если все ее возможные значения сосредоточены в области D_0 и если в этой области плотность распределения ее вероятностей постоянна:

$$p(x, y) = \begin{cases} C & \text{при} \quad (x, y) \in D_0, \\ 0 & \text{при} \quad (x, y) \notin D_0. \end{cases} \tag{3}$$

При этом вероятность попадания точки $(X;\ Y)$ в любую область D, лежащую внутри D_0, пропорциональна площади S_D этой области:

$$\mathsf{P}((X;\ Y) \in D) = C \cdot S_D.$$

Коэффициент C находится из условия нормировки

$$\mathsf{P}((X;\ Y) \in D_0) = C \cdot S_{D_0} = 1,$$

откуда

$$C = \frac{1}{S_{D_0}},$$

так что вероятность попадания в область D равна отношению площади области D к площади всей области D_0:

$$\mathsf{P}((X;\ Y) \in D) = \frac{S_D}{S_{D_0}}.$$

Когда говорят, что точку бросают «наудачу» в область D_0, то под этим всегда подразумевают равномерное распределение точки в области D_0.

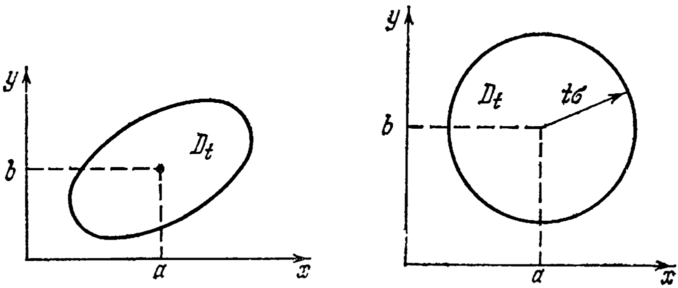
Пример 2. Нормальное распределение на плоскости. Двумерным нормальным распределением или нормальным распределением на плоскости называется распределение с плотностью

$$p(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp\left\{-\frac{1}{2}[A(x-a)^2 + \right.$$
$$\left. + 2B(x-a)(y-b) + C(y-b)^2]\right\}, \tag{4}$$

где A, B, C, a, b — некоторые постоянные, $A > 0$, $C > 0$, $AC - B^2 > 0$, а через $\exp(t) = e^t$ обозначена показательная функция (экспонента). Плотность (4) сохраняет постоянное значение на эллипсах

$$A(x-a)^2 + 2B(x-a)(y-b) + C(y-b)^2 = t^2 =$$
$$= \text{const}, \tag{5}$$

которые называются *эллипсами рассеяния* (рис. 22, *а*). Все эллипсы рассеяния подобны между собой и имеют центр в точке $(a;\ b)$, которая называется *центром рассеяния* или *центром распределения*.



а)

Рис. 22.

В частном случае, когда $B = 0$ и $A = C\ (= 1/\sigma^2)$, нормальное распределение на плоскости называется *круговым*; его плотность равна

$$p(x,\ y) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}[(x-a)^2 + (y-b)^2]\right\}. \qquad (6)$$

При этом эллипс рассеяния (5) обращается в окружность радиуса $t\sigma$ (рис. 22, *б*):

$$(x-a)^2 + (y-b)^2 = t^2\sigma^2.$$

Вероятность попадания точки $(X;\ Y)$ в круг D_t, ограниченный этой окружностью, легко вычислить с помощью перехода к полярным координатам $(r,\ \varphi)$ с полюсом в центре рассеяния $(a,\ b)$:

$$\mathsf{P}((X;\ Y) \in D_t) = \iint\limits_{D_t} p(x,\ y)\,dx\,dy =$$

$$= \frac{1}{2\pi\sigma^2}\int_0^{2\pi} d\varphi \int_0^{t\sigma} \exp\left(-\frac{r^2}{2\sigma^2}\right) r\,dr = 1 - e^{-\frac{t^2}{2}}. \qquad (7)$$

Отметим, что вероятность попадания точки в область D_t, ограниченную эллипсом рассеяния (5) при общем двумерном нормальном распределении, находится по той же формуле (7).

Распределения координат двумерной непрерывной случайной величины. Закон совместного распределения величин X и Y полностью определяет законы распределения каждой из величин X и Y. Пусть $p(x, y)$ — плотность совместного распределения вероятностей величин X и Y. Чтобы найти плотность $p_X(x)$ распределения величины X, рассмотрим вероятность попадания ее значения в любой интервал $(x_1 < X < x_2)$. Так как попадание абсциссы X в интервал (x_1, x_2) равносильно попаданию точки $(X; Y)$ в вертикальную полосу D, заштрихованную на рис. 23, то вероятности этих событий равны, следовательно,

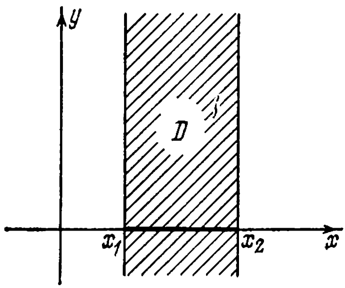


Рис. 23.

$$\mathsf{P}(x_1 < X < x_2) = \iint\limits_D p(x, y)\,dx\,dy = \int\limits_{x_1}^{x_2}\left[\int\limits_{-\infty}^{\infty} p(x, y)\,dy\right]dx.$$

Согласно определению плотности распределения вероятностей (см. § 3.1) отсюда следует, что искомая плотность равна

$$p_X(x) = \int\limits_{-\infty}^{\infty} p(x, y)\,dy. \tag{8}$$

Аналогично находится и плотность распределения вероятностей координаты Y:

$$p_Y(y) = \int\limits_{-\infty}^{\infty} p(x, y)\,dx. \tag{9}$$

Введенные выше понятия без труда обобщаются на k-мерные случайные величины при $k > 2$ или на системы более двух случайных величин. Так, для непрерывной трехмерной случайной величины закон распределения вероятностей задается плотностью

$p(x, y, z)$ и вероятность попадания точки $(X; Y; Z)$ в любую область D пространства выражается тройным интегралом:

$$\mathsf{P}((X; Y; Z) \in D) = \iiint\limits_D p(x, y, z)\,dx\,dy\,dz.$$

Распределение абсциссы X такой величины дается интегралом

$$p_X(x) = \iint\limits_{-\infty}^{\infty} p(x, y, z)\,dy\,dz.$$

Здесь уже можно говорить и о распределении подсистем случайных величин. Например, плотность совместного распределения величин X и Y выражается интегралом

$$p_{X,\,Y}(x, y) = \int\limits_{-\infty}^{\infty} p(x, y, z)\,dz.$$

Заметим еще, что система случайных величин может содержать как непрерывные, так и дискретные величины; такие ситуации в настоящем пособии не рассматриваются.

§ 3.7. Независимость непрерывных случайных величин

Обозначим плотности распределения величин X и Y через $p_X(x)$ и $p_Y(y)$, а плотность их совместного распределения через $p(x, y)$. Попадание точки $(X; Y)$ в прямоугольник $[x < X < x + \Delta x,\ y < Y < y + \Delta y)$, заштрихованный на рис. 21 (стр. 87), можно рассматривать как совмещение двух случайных событий: попадание абсциссы X в интервал $(x < X < x + \Delta x)$ и попадание ординаты Y в интервал $(y < Y < y + \Delta y)$. С точностью до бесконечно малых высших порядков вероятность попадания точки в указанный прямоугольник равна элементу вероятности $p(x, y)\,dx\,dy$ $(dx = \Delta x;\ dy = \Delta y)$, а вероятности попадания координат в указанные интервалы равны соответственно $p_X(x)\,dx$ и $p_Y(y)\,dy$. Поэтому для независимости величин X и Y естественно потребовать выполнения соотношения

$$p(x, y)\,dx\,dy = p_X(x)\,dx \cdot p_Y(y)\,dy,$$

т. е. соотношения

$$p(x, y) = p_X(x)\, p_Y(y) \tag{1}$$

для всех значений x и y. Распространяя эти соображения на любое число величин, дадим следующее

Определение. *Непрерывные случайные величины X, Y, Z, ... называются независимыми* (или *взаимно независимыми*), *если плотность их совместного распределения равна произведению плотностей этих величин*

$$p(x, y, z, \ldots) = p_X(x)\, p_Y(y)\, p_Z(z) \ldots \tag{2}$$

для всех значений x, y, z, ..

Заметим, что во многих задачах величины, не связанные между собой, считаются независимыми и в вероятностном смысле. В этом случае по плотностям распределения случайных величин можно просто найти плотность их совместного распределения, что позволяет существенно упростить расчеты.

Отметим также, что если плотность совместного распределения случайных величин X, Y, Z, ... допускает разделение переменных, т. е. может быть представлена в виде произведения

$$p(x, y, z, \ldots) = f_1(x) f_2(y) f_3(z) \ldots$$

каких-либо функций, каждая из которых зависит только от одного аргумента, то случайные величины X, Y, Z, ... независимы, причем плотности распределения их вероятностей лишь постоянными множителями отличаются от функций $f_1(x)$, $f_2(y)$, $f_3(z)$, .. соответственно. Доказательство этого утверждения не представляет труда (см. упр. 3 в конце главы).

Пример 1. Совместным распределением вероятностей величин X и Y служит равномерное распределение в прямоугольнике $[\alpha < X < \beta,\ \gamma < Y < \delta]$. Будут ли случайные величины X и Y независимы?

Решение. Плотность совместного распределения величин X и Y задается формулой:

$$p(x, y) = \begin{cases} \dfrac{1}{S} & \text{в прямоугольнике,} \\ 0 & \text{вне прямоугольника,} \end{cases}$$

где $S = (\beta - \alpha)(\delta - \gamma)$. Эту плотность можно представить в виде произведения плотностей двух равномерных распределений в интервалах ($\alpha < X < \beta$) и ($\gamma < Y < \delta$):

$$p_X(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{в интервале } (\alpha < x < \beta), \\ 0 & \text{вне этого интервала,} \end{cases}$$

$$p_Y(y) = \begin{cases} \dfrac{1}{\delta - \gamma} & \text{в интервале } (\gamma < y < \delta), \\ 0 & \text{вне этого интервала,} \end{cases}$$

а это и означает независимость величин X и Y.

Пример 2. Случайная величина (X; Y) имеет двумерное нормальное распределение вероятностей с плотностью (3.6-4). При каких значениях параметров A, B, C величины X и Y будут независимы?

Решение. Экспонента в формуле (3.6-4) допускает разделение переменных тогда и только тогда, когда $B = 0$. В этом случае плотность совместного распределения

$$p(x, y) = \frac{\sqrt{AC}}{2\pi} \exp\left\{-\frac{1}{2}\left[A(x-a)^2 + C(y-b)^2\right]\right\}$$

можно представить в виде произведения

$$p(x, y) = \frac{\sqrt{A}}{\sqrt{2\pi}} \exp\left[-\frac{A(x-a)^2}{2}\right] \cdot \frac{\sqrt{C}}{\sqrt{2\pi}} \exp\left[-\frac{C(y-b)^2}{2}\right],$$

откуда видно, что величины X и Y независимы и распределены нормально с параметрами соответственно $\left(a; \dfrac{1}{\sqrt{A}}\right)$ и $\left(b; \dfrac{1}{\sqrt{C}}\right)$. В частности, это утверждение справедливо и для кругового нормального распределения (3.6-6).

§ 3.8. Понятие функции непрерывных случайных величин

Понятие функции случайной величины было введено в § 2.7. Основной задачей здесь является нахождение закона распределения вероятностей функции по законам распределения вероятностей ее аргументов.

Общий метод решения такой задачи состоит в следующем. Случайное событие $u_1 < f(X, Y, Z, \ldots) < u_2$

можно представить как попадание случайной точки $(X;\ Y;\ Z;\ \dots)$ в множество D, определяемое указанным неравенством. В соответствии с определением непрерывной случайной величины вероятность такого события равна интегралу по множеству D от плотности совместного распределения вероятностей величин X, Y, Z, ... Для нахождения плотности $p_U(u)$ распределения функции $U = f(X,\ Y,\ Z,\ \dots)$ остается только преобразовать такой интеграл к виду $\int\limits_{u_1}^{u_2} p_U(u)\,du$. Ввиду сложности таких преобразований в общем случае мы ограничимся здесь монотонными функциями от одной случайной величины и приведем несколько примеров более сложных преобразований случайных величин.

Монотонная функция одной случайной величины. Пусть в интервале возможных значений случайной величины X функция $y = f(x)$ строго возрастает (кроме того, мы будем еще предполагать, что эта функция непрерывна вместе со своей первой производной и, следовательно, имеет дифференцируемую обратную функцию). При этом каждый внутренний интервал $(x_1,\ x_2)$ отображается взаимно однозначно на соответствующий интервал $(y_1,\ y_2)$ (рис. 24); поэтому вероятности попадания случайных величин X и Y в соответствующие интервалы равны между собой:

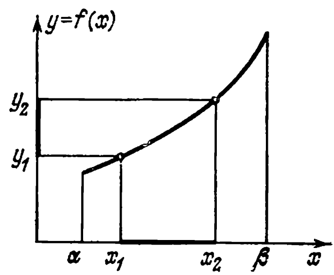


Рис. 24.

$$\mathsf{P}(y_1 < Y < y_2) = \mathsf{P}(x_1 < X < x_2).$$

Обозначим через $p_X(x)$ плотность распределения вероятностей величины X; тогда

$$\mathsf{P}(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} p_X(x)\,dx.$$

В последнем интеграле заменим x на функцию $g(y)$, обратную к функции $y = f(x)$; мы получим

$$\int_{x_1}^{x_2} p_X(x)\,dx = \int_{y_1}^{y_2} p_X[g(y)]\,g'(y)\,dy,$$

т. е.

$$\mathsf{P}(y_1 < Y < y_2) = \int_{y_1}^{y_2} p_X[g(y)]\,g'(y)\,dy.$$

Отсюда видно, что Y есть непрерывная случайная величина с плотностью распределения

$$\boxed{p_Y(y) = p_X[g(y)]\,g'(y) = p_X(x)\frac{dx}{dy}.} \qquad (1)$$

Последнее соотношение можно записать также в

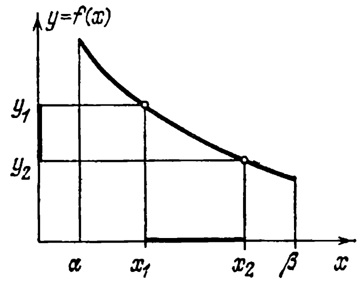


Рис. 25.

виде равенства элементов вероятностей

$$p_Y(y)dy = p_X(x)dx. \qquad (2)$$

Для строго убывающей функции $y = f(x)$ мы имеем $y_2 < y_1$ (рис. 25) и поэтому из соотношений

$$\mathsf{P}(y_2 < Y < y_1) = \mathsf{P}(x_1 < X < x_2) =$$

$$= \int_{x_1}^{x_2} p_X(x)\,dx = \int_{y_1}^{y_2} p_X[g(y)]\,g'(y)\,dy = -\int_{y_2}^{y_1} p_X[g(y)]\,g'(y)\,dy$$

получаем

$$p_Y(y) = -p_X[g(y)]g'(y). \qquad (3)$$

Формулы (1) и (3) можно объединить в одну:

$$p_Y(y) = p_X[g(y)]\,|g'(y)| = p_X(x)\left|\frac{dx}{dy}\right|. \qquad (4)$$

Формула (4) пригодна для любой монотонной функции.

Пример 1. Линейная функция. Случайная величина X имеет плотность распределения вероятностей $p_X(x)$. Найти плотность распределения вероятностей линейной функции $Y = aX + b$ $(a \neq 0)$.

Решение. При $a \neq 0$ функция $y = ax + b$ монотонна и обратной ей служит функция $x = \frac{y-b}{a}$; производная обратной функции равна $\frac{dx}{dy} = \frac{1}{a}$. Поэтому плотность распределения вероятностей линейной функции Y будет равна

$$p_Y(y) = p_X\left(\frac{y-b}{a}\right)\cdot\frac{1}{|a|}. \qquad (5)$$

Отметим интересные частные случаи.

а) Если случайная величина X имеет равномерное распределение в интервале (α, β), то случайная величина $Y = X - \frac{\alpha+\beta}{2}$ будет иметь равномерное распределение в симметричном относительно $y = 0$ интервале $(-\gamma, \gamma)$, где $\gamma = \frac{\beta-\alpha}{2}$, а случайная величина $Z = Y/\gamma$ будет иметь равномерное распределение в интервале $(-1, 1)$.

б) Если случайная величина X имеет показательное распределение (3.2-5) с параметром λ, то случайная величина $Y = aX$ при $a > 0$ будет иметь также показательное распределение с параметром λ/a.

в) *Если случайная величина X имеет нормальное распределение* (3.3-1) *с параметрами a и σ, то случайная величина $Y = \frac{X-a}{\sigma}$ имеет стандартное нормальное распределение* (3.3-2).

Проверку этих утверждений мы предоставляем читателю.

Пример 2. Случайная величина X имеет плотность распределения вероятностей

$$p_X(x) = \begin{cases} \frac{2}{a^2}x & \text{при } 0 < x < a, \\ 0 & \text{вне интервала } (0, a). \end{cases}$$

Найти плотность распределения вероятностей величины $Y = X^2$.

Решение. В интервале $(0, a)$ возможных значений величины X функция $y = x^2$ строго возрастает, имеет обратную функцию $x = \sqrt{y}$ с производной $\frac{1}{2\sqrt{y}}$ $(0 < y < a^2)$. Поэтому плотность распределения вероятностей величины Y равна

$$p_Y(y) = \begin{cases} \frac{2}{a^2}\sqrt{y}\,\frac{1}{2\sqrt{y}} = \frac{1}{a^2} & \text{при } 0 < y < a^2, \\ 0 & \text{вне интервала } (0, a^2), \end{cases}$$

т. е. величина Y распределена равномерно в интервале $(0, a^2)$ (см. рис. 26, где $a=2$).

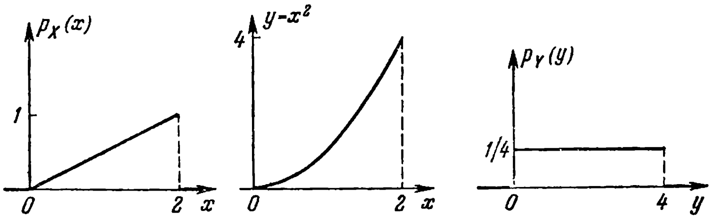


Рис. 26.

Рассмотрим теперь пример немонотонной функции.

Пример 3. Значения случайной величины X сосредоточены в симметричном интервале $(-\alpha, \alpha)$. Найти плотность распределения вероятностей величины $Y = X^2$, если плотность распределения вероятностей величины X в интервале $(-\alpha, \alpha)$ обозначена через $p_X(x)$.

Решение. Функция $y = x^2$ в интервале $(-\alpha, \alpha)$ не монотонна, поэтому непосредственно применять формулу (4) нельзя.

Как видно из рис. 27, событие $(y_1 < Y < y_2)$ при $0 < y_1 < y_2 < \alpha^2$ равносильно сумме двух несовместимых случайных событий $(x_1 < X < x_2)$ и $(-x_2 < X < -x_1)$, где $x_1 = \sqrt{y_1}$, $x_2 = \sqrt{y_2}$. Поэтому мы можем написать

$$\mathsf{P}(y_1 < Y < y_2) = \int_{x_1}^{x_2} p_X(x)\,dx + \int_{-x_2}^{-x_1} p_X(x)\,dx.$$

В первом из этих интегралов заменим $x = \sqrt{y}$, а во

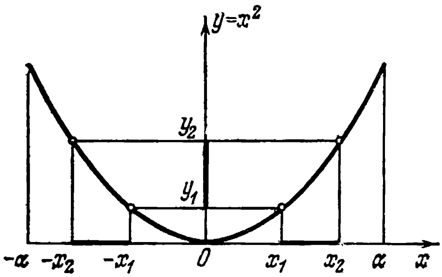


Рис. 27.

втором $x = -\sqrt{y}$; мы получим

$$\mathsf{P}(y_1 < Y < y_2) =$$
$$= \int_{y_1}^{y_2} p_X(\sqrt{y})\frac{dy}{2\sqrt{y}} + \int_{y_2}^{y_1} p_X(-\sqrt{y})\left(-\frac{dy}{2\sqrt{y}}\right) =$$
$$= \int_{y_1}^{y_2} [p_X(\sqrt{y}) + p_X(-\sqrt{y})]\frac{1}{2\sqrt{y}}\,dy.$$

Отсюда видно, что плотность распределения вероятностей величины Y равна

$$p_Y(y) = [p_X(\sqrt{y}) + p_X(-\sqrt{y})]\frac{1}{2\sqrt{y}} \quad \text{при } 0 < y < \alpha^2. \tag{6}$$

Вне интервала $(0, \alpha^2)$, очевидно, $p_Y(y) = 0$.

В частном случае, когда плотность $p_X(x)$ представляет собой четную функцию, формула (6) принимает более простой вид:

$$p_Y(y) = p_X(\sqrt{y})\frac{1}{\sqrt{y}} \quad \text{при } 0 < y < \alpha^2.$$

Так, если случайная величина X имеет стандартное нормальное распределение с плотностью

$$p_X(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \quad (-\infty < x < \infty),$$

то для величины $Y = X^2$ мы получим

$$p_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y}{2}}y^{-\frac{1}{2}} \quad \text{при } y > 0,$$

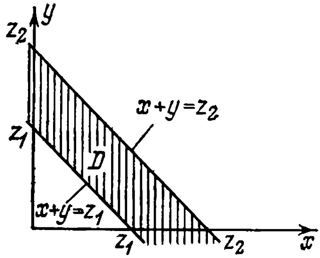


Рис. 28.

т. е. величина Y будет иметь гамма-распределение (3.2-6) с параметрами $\lambda = \frac{1}{2}$, $\alpha = \frac{1}{2}$.

Сумма двух случайных величин. Обозначим плотность совместного распределения вероятностей двух случайных величин X и Y через $p(x,y)$ и будем искать закон распределения вероятностей их суммы $Z = X + Y$. Случайное событие $z_1 < Z < z_2$, т. е. $z_1 < X + Y < z_2$, означает попадание случайной точки $(X;\ Y)$ в полосу D между прямыми $x + y = z_1$ и $x + y = z_2$ (рис. 28). Поэтому вероятность этого события равна

$$\mathsf{P}(z_1 < Z < z_2) = \iint_D p(x, y)\,dx\,dy = \int_{-\infty}^{\infty} dx \int_{z_1-x}^{z_2-x} p(x, y)\,dy.$$

Заменим во внутреннем интеграле $y = z - x$; мы получим

$$\mathsf{P}(z_1 < Z < z_2) = \int_{-\infty}^{\infty} dx \int_{z_1}^{z_2} p(x, z - x)\,dz.$$

Изменяя порядок интегрирования, запишем последнее соотношение в виде

$$\mathsf{P}(z_1 < Z < z_2) = \int_{z_1}^{z_2} \left[\int_{-\infty}^{\infty} p(x, z - x)\, dx \right] dz.$$

Отсюда видно, что Z есть непрерывная случайная величина с плотностью распределения вероятностей

$$p_Z(z) = \int_{-\infty}^{\infty} p(x, z - x)\, dx. \tag{7}$$

Особый интерес представляет с л о ж е н и е н е з а в и с и м ы х с л у ч а й н ы х в е л и ч и н X и Y. В этом случае $p(x, y) = p_X(x) p_Y(y)$ и формула (7) принимает следующий вид:

$$\boxed{p_Z(z) = \int_{-\infty}^{\infty} p_X(x)\, p_Y(z - x)\, dx.} \tag{8}$$

Закон распределения суммы независимых случайных величин называется *композицией их законов распределения*; интеграл в правой части формулы (8) называется *сверткой функций* $p_X(x)$ и $p_Y(y)$; таким образом, *композиция двух непрерывных законов распределения сводится к свертке их плотностей.*

Свертка обозначается символом $*$. Заметим, что операция образования свертки («свертывание») обладает легко проверяемыми свойствами:

переместительным $p_1 * p_2 = p_2 * p_1$,
сочетательным $(p_1 * p_2) * p_3 = p_1 * (p_2 * p_3)$.

В приложениях теории вероятностей большую роль играют такие законы распределения вероятностей, композиция которых сохраняет тип закона распределения. Этим свойством обладает, например, нормальный закон распределения, как это будет показано ниже (стр. 170).

П р и м е р 4. Независимые случайные величины X и Y имеют гамма-распределения (3.2-6) с одинаковыми значениями параметра λ. Найти плотность распределения их суммы $Z = X + Y$

Решение. Запишем плотности распределения вероятностей величин X и Y:

$$p_X(x) = C_1 e^{-\lambda x} x^{\alpha-1} \quad \text{при} \quad x > 0,$$
$$p_Y(y) = C_2 e^{-\lambda y} y^{\beta-1} \quad \text{при} \quad y > 0.$$

Плотность распределения суммы $Z = X + Y$ запишется в виде свертки

$$p_Z(z) = p_X(x) * p_Y(y) = \int_{-\infty}^{\infty} p_X(x)\, p_Y(z-x)\, dx.$$

Так как функция $p_X(x)$ обращается в нуль при $x < 0$, а функция $p_Y(y) = p_Y(z-x)$ — при $z - x < 0$, т. е. при $z < x$, то последний интеграл отличен от нуля только при $0 < x < z$ и при этом может быть записан так:

$$p_Z(z) = \int_0^z p_X(x)\, p_Y(z-x)\, dx =$$
$$= C_1 C_2 \int_0^z e^{-\lambda x} x^{\alpha-1} e^{-\lambda(z-x)} (z-x)^{\beta-1}\, dx.$$

Полученный интеграл преобразуем с помощью замены $x = zt$:

$$p_Z(z) = C_1 C_2 e^{-\lambda z} \int_0^1 t^{\alpha-1} z^{\alpha-1} (1-t)^{\beta-1} z^{\beta-1} z\, dt = C_3 e^{-\lambda z} z^{\alpha+\beta-1},$$

где

$$C_3 = C_1 C_2 \int_0^1 t^{\alpha-1} (1-t)^{\beta-1}\, dt.$$

Таким образом, случайная величина Z имеет тоже гамма-распределение с параметрами λ и $\alpha + \beta$, т. е. композиция двух гамма-распределений с одинаковыми значениями параметра λ дает снова гамма-распределение с тем же значением параметра λ. По индукции это свойство распространяется на любое число независимых слагаемых такого рода.

Заметим, что показательное распределение можно рассматривать как гамма-распределение с параметрами

λ и 1; поэтому композиция двух показательных распределений с одинаковыми параметрами λ приводит к гамма-распределению с параметрами λ и 2. Применяя метод индукции, приходим к выводу, что гамма-распределение с целым значением параметра $\alpha = k$ можно рассматривать как композицию k показательных распределений.

Пример 5. χ^2 — *распределение* (*распределение «хи-квадрат»*). Распределением χ^2 называется распределение суммы квадратов независимых случайных величин, следующих стандартному нормальному закону распределения. Найти плотность этого распределения.

Решение. Обозначим исходные случайные величины через X_1, X_2, ..., X_k, а сумму их квадратов через

$$U_k = X_1^2 + X_2^2 + \ldots + X_k^2.$$

Число слагаемых k называется *числом степеней свободы* величины U_k; сам закон распределения обозначается χ_k^2. Так как каждая величина X_i ($i = 1, 2, \ldots, k$) имеет стандартное нормальное распределение, то ее квадрат имеет гамма-распределение с параметрами $\lambda = 1/2$, $\alpha = 1/2$ (см. пример 3). Но при композиции k гамма-распределений с параметрами $\lambda = 1/2$, $\alpha = 1/2$ мы получим снова гамма-распределение с параметрами $\lambda = 1/2$, $\alpha = k \cdot 1/2$ (см. пример 4). Таким образом, χ^2 — распределение с числом степеней свободы k есть гамма-распределение с параметрами $\lambda = 1/2$ и $\alpha = k/2$; плотность этого распределения равна

$$p_{\chi_k^2}(u) = \begin{cases} C_k e^{-\frac{u}{2}} u^{\frac{k}{2}-1} & \text{при } u > 0, \\ 0 & \text{при } u < 0. \end{cases} \tag{9}$$

Здесь $C_k = \dfrac{\lambda^\alpha}{\Gamma(\alpha)} = \dfrac{1}{2^{k/2}\Gamma(k/2)}$; так при $k = 4$ имеем

$$C_4 = \frac{1}{2^2\Gamma(2)} = \frac{1}{4}, \quad p_{\chi_4^2}(u) = \frac{u}{4} e^{-\frac{u}{2}} \quad (u > 0) \text{ (рис. 29).}$$

Уже из самого определения χ^2-распределения непосредственно следует, что сумма независимых величин, имеющих χ^2-распределения, будет иметь снова χ^2-распределение, причем при сложении таких величин их числа степеней свободы складываются.

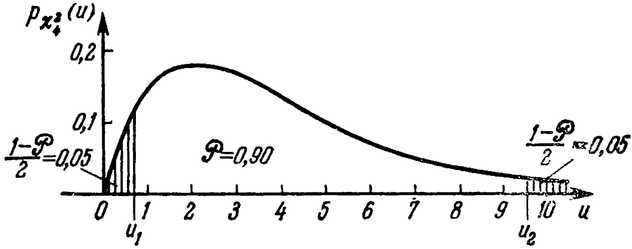


Рис. 29.

Пример 6. Плотность $p(x, y)$ совместного распределения величин X и Y зависит только от расстояния $r=\sqrt{x^2+y^2}$ точки (x, y) от начала координат: $p(x, y)=g(r)$. Найти закон распределения функции $R=\sqrt{X^2+Y^2}$, т. е. закон распределения расстояния точки $(X; Y)$ от начала координат.

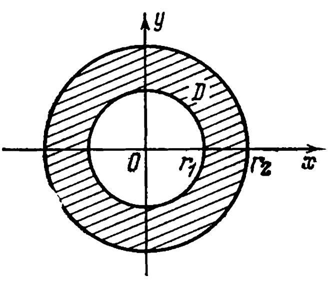


Рис. 30.

Решение. Случайное событие $r_1<R<r_2$ означает попадание точки $(X; Y)$ в кольцо D между окружностями радиусов r_1 и r_2 (рис. 30). Вероятность этого события равна

$$\mathsf{P}(r_1<R<r_2)=\iint_D p(x, y)\,dx\,dy.$$

Написанный интеграл целесообразно вычислять в полярных координатах r, φ, что дает

$$\mathsf{P}(r_1<R<r_2)=\int_0^{2\pi} d\varphi \int_{r_1}^{r_2} g(r)\,r\,dr = 2\pi \int_{r_1}^{r_2} g(r)\,r\,dr.$$

Отсюда видно, что R есть непрерывная случайная величина с плотностью распределения

$$p_R(r) = 2\pi g(r) r \quad \text{при} \quad r > 0. \tag{10}$$

Очевидно, что при $r < 0$ надо положить $p_R(r) = 0$.

В частности, если точка $(X;\ Y)$ распределена равномерно в круге K радиуса a, так что $p(x, y) = 1/\pi a^2$ в круге K, и $p(x, y) = 0$ вне круга K, то плотность распределения $p_R(r)$ равна

$$p_R(r) = \frac{2\pi r}{\pi a^2} = \frac{2r}{a^2} \text{ при } 0 < r < a \text{ и } p_R(r) = 0 \text{ при } r > a.$$

(Любопытно, что при этом величина R^2 распределена равномерно в $(0;\ a^2)$, см. пример 2, стр. 96.)

Отметим еще один частный случай; если точка $(X;\ Y)$ имеет круговое нормальное распределение (3.6-6) с центром в начале координат, так что

$$p(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

то плотность распределения расстояния R равна

$$p_R(r) = 2\pi r \frac{1}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad (r > 0).$$

Закон распределения с такой плотностью называется *законом Релея*. Он позволяет просто найти вероятность попадания в круг радиуса $t\sigma$:

$$\mathsf{P}(R < t\sigma) = \int_0^{t\sigma} \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr = 1 - e^{-\frac{t^2}{2}},$$

что совпадает с полученным выше результатом (3.6-7).

Пример 7. Пусть у случайного вектора $\{X;\ Y\}$ декартовы координаты X и Y независимы и следуют стандартному нормальному закону распределения вероятностей. Доказать, что при любом повороте системы координат новые координаты X' и Y' рассматриваемого случайного вектора будут снова независимы и будут распределены по стандартному нормальному закону.

Решение. В силу условия задачи плотность совместного распределения величин X и Y равна

$$p(x, y) = \varphi(x)\,\varphi(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \exp\left(-\frac{x^2+y^2}{2}\right).$$

Поворот системы координат на угол α (рис. 31) означает линейное преобразование координат по формулам

$$\begin{cases} X = X'\cos\alpha - Y'\sin\alpha, \\ Y = X'\sin\alpha + Y'\cos\alpha. \end{cases}$$

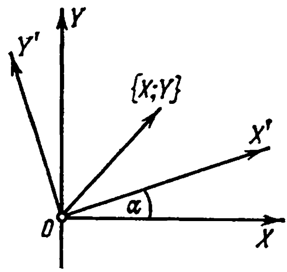


Рис. 31.

Нам нужны будут лишь два свойства этого преобразования: сохранение площадей (якобиан преобразования равен 1) и сохранение расстояний от начала координат $\sqrt{X^2+Y^2} = \sqrt{X'^2+Y'^2}$.

В силу первого свойства вероятность попадания точки $(X;\ Y)$ в любую область D на плоскости преобразуется при повороте так:

$$\mathsf{P}((X;\ Y) \in D) = \iint\limits_D p(x, y)\,dx\,dy = \iint\limits_D p(x, y)\,dx'\,dy';$$

откуда следует, что плотность $p_1(x', y')$ совместного распределения величин X' и Y' равна

$$p_1(x', y') = p(x, y),$$

т. е.

$$p_1(x', y') = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \exp\left(-\frac{x^2+y^2}{2}\right),$$

где

$$x = x'\cos\alpha - y'\sin\alpha, \quad y = x'\sin\alpha + y'\cos\alpha.$$

Но по второму свойству преобразования $x^2 + y^2 =$

$= x'^2 + y'^2$, и поэтому

$$p_1(x', y') = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \exp\left(-\frac{x'^2 + y'^2}{2}\right)$$

или

$$p_1(x', y') = \\ = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y'^2}{2}\right) = \varphi(x')\,\varphi(y').$$

Последнее соотношение и означает, что X' и Y' независимы и следуют стандартному нормальному закону распределения, что т. д.

Из хода доказательства ясно, что и для трехмерного случайного вектора при любом повороте системы координат сохраняется независимость и нормальность распределения его координат. Это свойство оказывается справедливым и для n-мерного случайного вектора $\{X_1, X_2, \ldots, X_n\}$, если под «поворотом» понимать ортогональное преобразование, т. е. линейное преобразование координат, сохраняющее все расстояния, а значит, и «объемы» и суммы квадратов координат. Это свойство имеет важные приложения в теории обработки результатов измерений.

УПРАЖНЕНИЯ

1. Точка случайно попадает на окружность радиуса a (распределение вероятностей предполагается равномерным по длине дуги). Найти распределение вероятностей проекции этой точки на диаметр.

2. Будут ли независимы случайные величины X и Y, если их совместное распределение является равномерным

а) в круге $x^2 + y^2 \leqslant 1$?

б) в треугольнике $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1 - x$?

в) в прямоугольнике $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 2$?

3. Плотность совместного распределения величин X и Y допускает разделение переменных, т. е. может быть представлена в виде $p(x, y) = f_1(x) f_2(y)$. Найти плотности распределения каждой из величин X и Y (ср. стр. 92).

4. Совместное распределение вероятностей величин X и Y является нормальным с плотностью

$$p(x, y) = \frac{1}{3\pi} \exp\left(-\frac{x^2 + 4y^2}{6}\right).$$

Найти вероятность попадания точки $(X;\ Y)$ в прямоугольник

$$[-1 < X < 1;\ -2 < Y < 2].$$

5. Совместное распределение вероятностей величин X и Y является нормальным с плотностью

$$p(x, y) = \frac{2}{\pi} \exp\left[-(x-3)^2 - 2(x-3)(y-4) - 5(x-4)^2\right].$$

Найти вероятность попадания точки $(X; Y)$ в полуплоскость $y > 5x - 11$.

6. Независимые случайные величины X и Y распределены по показательному закону с одним и тем же значением параметра λ. Найти законы распределения величин $U = 2X$ и $V = X + Y$.

7*. Найти композиции двух и трех случайных величин, распределенных равномерно в интервале $(-1, 1)$. Нарисовать соответствующие кривые распределения.

8. Найти композицию нормального распределения с параметрами 0 и σ и равномерного распределения в интервале $(-l, l)$.

9. Доказать, что если случайная величина X распределена равномерно в интервале $(0, 1)$, то абсолютная величина ее логарифма имеет показательное распределение вероятностей.

10*. Вывести закон распределения частного двух независимых положительных непрерывных величин X и Y с плотностями распределения $p_X(x)$ и $p_Y(y)$.

11. Найти распределение частного двух независимых случайных величин U и V, следующих законам распределения χ^2 с числами степеней свободы k и m.

12*. Кубики изготовляются с некоторой погрешностью. Найти распределение вероятностей их объемов, если все их линейные размеры независимы и имеют нормальное распределение вероятностей с одними и теми же параметрами a и σ.

13. *Логарифмически нормальным распределением* называется распределение величины Y, натуральный логарифм которой имеет нормальное распределение. Найти плотность логарифмически нормального распределения. Нарисовать кривую распределения.

14. *Задача о встрече.* Два лица условились встретиться в определенном месте в промежутке времени между 0 и 1 часами. Пришедший первым ждет другого 20 минут, после чего уходит. Найти вероятность встречи в предположении, что моменты прихода обоих лиц независимы и распределены равномерно в интервале $(0, 1)$.

ГЛАВА 4

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ РАСПРЕДЕЛЕНИЯ

§ 4.1. Математическое ожидание случайной величины и другие характеристики положения

В предыдущих главах мы рассмотрели ряд задач, в которых сравнительно просто удается найти законы распределения вероятностей изучаемых случайных величин.

Но отыскание закона распределения вероятностей часто связано с большими трудностями. Оказывается, что ряд практически важных задач можно решить с помощью немногих осредненных характеристик распределения.

Одной из самых важных характеристик распределения случайной величины является ее математическое ожидание или среднее значение.

Рассмотрим сначала дискретную случайную величину X, принимающую значения $x_1, x_2, \ldots, x_k$ с вероятностями $p_1, p_2, \ldots, p_k$ соответственно. Произведем n независимых испытаний для определения значений величины X, и пусть каждое значение x_i встретится при этом n_i раз ($i = 1, 2, \ldots, k$), так что $n_1 + n_2 + \ldots + n_k = n$. Как известно, среднее арифметическое из полученных значений определяется формулой

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n} \tag{1}$$

(каждое значение надо учитывать слагаемым столько раз, сколько раз оно встретилось!). Представим среднее арифметическое $\bar{x}$ в следующем виде:

$$\bar{x} = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \ldots + x_k \frac{n_k}{n} = \sum_{i=1}^{k} x_i \frac{n_i}{n}. \tag{2}$$

Так как при большом числе испытаний (n) относительные частоты n_i/n будут группироваться около соответствующих вероятностей

$$p_i = \mathsf{P}(X = x_i),$$

то значения среднего арифметического $\bar{x}$ будут группироваться около числа $\sum_{i=1}^{k} x_i p_i$. Это оправдывает следующее

Определение. *Математическим ожиданием $\mathsf{M}X$ дискретной случайной величины X называется сумма произведений всех ее возможных значений x_i на их вероятности p_i:*

$$\boxed{\mathsf{M}X = \sum_{i=1}^{k} x_i p_i.} \tag{3}$$

Если множество возможных значений дискретной случайной величины X бесконечно, то естественно заменить конечную сумму в формуле (3) суммой бесконечного ряда:

$$\boxed{\mathsf{M}X = \sum_{i=1}^{\infty} x_i p_i.} \tag{4}$$

При этом мы всегда будем предполагать, что этот бесконечный ряд абсолютно сходится (в противном случае говорят, что математическое ожидание случайной величины не существует).

Математическое ожидание $\mathsf{M}X$ называют также средним значением случайной величины X, подчеркивая тем самым статистический смысл этого понятия (статистическим аналогом математического ожидания служит среднее арифметическое из эмпирических значений случайной величины).

Заметим, однако, что в отличие от среднего арифметического значения (1), которое само является случайной величиной, так как зависит от испытаний, математическое ожидание является числом, которое связано только с законом распределения случайной величины.

Математическое ожидание MX называют еще и **центром распределения случайной величины** X. Это название вводят по аналогии с понятием центра тяжести для системы материальных точек, расположенных на одной прямой; если на оси x в k точках с координатами $x_1, x_2, \ldots, x_k$ сосредоточены массы $p_1, p_2, \ldots, p_k$, то координата x_c центра тяжести системы находится по формуле

$$x_c = \frac{\sum\limits_{i=1}^{k} x_i p_i}{\sum\limits_{i=1}^{k} p_i};$$

для распределения вероятностей $\sum\limits_{i=1}^{k} p_i = 1$ и написанное справа отношение сводится к сумме $\sum\limits_{i=1}^{k} x_i p_i$.

Обычно термин «центр распределения» (или даже «центр рассеяния») употребляют в тех случаях, когда хотят подчеркнуть, что математическое ожидание применяется в качестве характеристики положения случайной величины.

Отметим, что математическое ожидание дискретной случайной величины X может не совпадать ни с одним из ее возможных значений. Например, для распределения числа очков на верхней грани игральной кости, представленного таблицей (2.1-3), имеем

$$MX = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3{,}5.$$

Отметим также, что если величина X принимает только одно значение a с вероятностью 1, то ее математическое ожидание совпадает с этим значением; говорят, что математическое ожидание постоянной равно самой постоянной.

Математическое ожидание непрерывной случайной величины.

Распространим теперь понятие математического ожидания на непрерывные одномерные случайные величины X.

Определение. *Математическим ожиданием (средним значением, центром распределения) MX непрерывной случайной величины X называется интеграл от произведения ее значений x на плотность распределения вероятностей $p(x)$:*

$$\boxed{MX = \int_{-\infty}^{\infty} x p(x)\, dx.} \tag{5}$$

Несобственный интеграл (5) предполагается абсолютно сходящимся (в противном случае говорят, что математическое ожидание случайной величины не существует).

Полезно обратить внимание на то, что структура формулы (5) такова же, как и структура формулы (3): роль вероятностей p_i играют элементы вероятностей $p(x)\,dx$; поэтому практическим аналогом математического ожидания и здесь будет среднее арифметическое значение.

Если все возможные значения величины X сосредоточены на конечном интервале (α, β), т. е. если $p(x) = 0$ вне интервала (α, β), то

$$MX = \int_{\alpha}^{\beta} x p(x)\, dx;$$

в этом случае математическое ожидание случайной величины X лежит в интервале (α, β), так как

$$\alpha \int_{\alpha}^{\beta} p(x)\, dx < \int_{\alpha}^{\beta} x p(x)\, dx < \beta \int_{\alpha}^{\beta} p(x)\, dx,$$

т. е.

$$\alpha < MX < \beta.$$

Если математическое ожидание MX непрерывной случайной величины X существует и если ее плотность вероятностей $p(x)$ есть четная функция, то $MX = 0$ (так как в этом случае подынтегральная функция $xp(x)$ будет нечетной, а интервал интегрирования симметричным относительно начала $x = 0$). Вообще, если математическое ожидание MX случайной величины X

существует и если ее плотность $p(x)$ симметрична относительно некоторой точки $x = a$, то $\mathsf{M}X = a$. В частности, центр равномерного распределения в интервале (α, β) совпадает с серединой этого интервала ($\mathsf{M}X = (\alpha + \beta)/2$), а центр нормального распределения (3.3-1) совпадает с центром симметрии ($\mathsf{M}X = a$).

Подчеркнем, что *математическое ожидание случайной величины всегда имеет ту же размерность, что и значения случайной величины*, так как вероятности суть величины безразмерные.

Понятие о моде и медиане.

Кроме центра распределения иногда применяются и другие характеристики положения случайной величины. Мы укажем здесь две из таких характеристик — моду и медиану.

Для дискретной случайной величины модой называют такое ее значение, которое имеет наибольшую вероятность (наиболее вероятное значение случайной величины).

Например, как указывалось в § 2.3, случайная величина с биномиальным распределением вероятностей (2.3.2) имеет одно наиболее вероятное значение m_0, заключенное между числами $np + p - 1$ и $np + p$ (если только $np + p$ не является целым числом); это число m_0 и будет модой биномиального распределения.

Для непрерывной случайной величины модой называют точку максимума плотности распределения ее вероятностей. Распределения с одной модой называются одномодальными; они играют наиболее важную роль в приложениях. Иногда и здесь употребляют выражение «наиболее вероятное значение случайной величины»; это выражение надо понимать не в буквальном смысле (вероятность любого значения непрерывной случайной величины равна нулю!), а следующим образом: вероятность попадания значения случайной величины в достаточно малый интервал, содержащий моду, будет больше, чем вероятность попадания в любой другой интервал той же длины. Для гамма-распределения (3.2-6), рассмотренного в § 3.2, мы уже нашли моду при $\alpha > 1$; она оказалась равной $\frac{\alpha - 1}{\lambda}$ (см. стр. 75). Для нормального распределения (3.3-1) мода, очевидно, совпадает с центром распределения a.

Медианой непрерывной случайной величины X называется такое значение $x_{1/2}$, что вероятности $\mathsf{P}(X < x_{1/2})$ и $\mathsf{P}(X > x_{1/2})$ оказываются равными между собой:

$$\mathsf{P}(X < x_{1/2}) = \mathsf{P}(X > x_{1/2}) = \frac{1}{2}.$$

Иначе можно сказать, что в точке $x_{1/2}$ функция распределения $F(x)$ случайной величины X равна 1/2: $F(x_{1/2}) = 1/2$. Для симметричного распределения (например, равномерного, нормального) медиана, очевидно, совпадает с центром распределения. Найдем еще медиану показательного распределения (3.2-5), рассмотренного в § 3.2: из формулы

$$F(t) = \int_0^t \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda t} \qquad (t > 0)$$

получим $1 - e^{-\lambda t_{1/2}} = \frac{1}{2}$, откуда $t_{1/2} = \frac{\ln 2}{\lambda}$.

Для гамма-распределения (3.2-6) можно получить приближенную формулу $x_{1/2} \approx \frac{\alpha - \frac{1}{3}}{\lambda}$, точность которой возрастает с увеличением α. В частности, для χ^2-распределения с k степенями свободы $\left(\alpha = \frac{k}{2},\ \lambda = \frac{1}{2}\right)$ имеем $x_{1/2} \approx k - \frac{2}{3}$.

Для дискретных случайных величин понятие медианы не применяется, так как для них уравнение $F(x) = 1/2$ либо вовсе не имеет решений, либо решения этого уравнения заполняют целый интервал (см. рис. 19, стр. 83).

Математическое ожидание случайного вектора (центр многомерного распределения). Математическим ожиданием случайного вектора $\{X; Y\}$ называется вектор $\{\mathsf{M}X; \mathsf{M}Y\}$. Центром распределения (центром рассеяния) двумерной случайной величины $(X; Y)$ называется точка $(\mathsf{M}X; \mathsf{M}Y)$. Аналогично определяется математическое ожидание случайного вектора (центр распределения многомерной случайной величины) при размерности больше 2.

Отметим, что математическое ожидание случайного вектора $\{X; Y\}$ просто выражается через плотность совместного распределения вероятностей его координат. Обозначим через $p(x, y)$ плотность совместного распределения двух величин X и Y; учитывая формулу (3.6-8), получим

$$\mathsf{M}X = \int_{-\infty}^{\infty} x p_X(x)\,dx =$$

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

Аналогично получим

$$\mathsf{M}Y = \iint\limits_{-\infty}^{\infty} yp(x, y)\,dx\,dy.$$

Оба интеграла берутся по всей плоскости xOy.

Пример. Для равномерного распределения в плоской области D_0, где $p(x, y) = C = \text{const}$, получаем

$$a = \mathsf{M}X = \iint\limits_{D_0} xC\,dx\,dy = \frac{1}{S}\iint\limits_{D_0} x\,dx\,dy,$$

$$b = \mathsf{M}Y = \iint\limits_{D_0} yC\,dx\,dy = \frac{1}{S}\iint\limits_{D_0} y\,dx\,dy,$$

где S — площадь области D_0; эти формулы совпадают с формулами для координат центра тяжести однородной пластинки, занимающей область D_0.

Если плотность распределения $p(x, y)$ обладает симметрией относительно начала координат (т. е. не меняется при замене x на $-x$ и y на $-y$), то центр распределения системы X и Y совпадает с началом координат (0; 0). Например, для двумерного нормального распределения с плотностью

$$p(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp\left\{-\frac{1}{2}(Ax^2 + 2Bxy + Cy^2)\right\}$$

центр распределения есть точка (0; 0). Так как плотность распределения (3.6-4) отличается от только что написанной только сдвигом по осям координат, то центром общего двумерного нормального распределения (3.6-4) служит точка $(a; b)$.

§ 4.2. Математическое ожидание функции случайных величин

Функция $Y = f(X)$ сама является случайной величиной; поэтому ее математическое ожидание можно находить по общим формулам; например, для непрерывной величины Y

$$\mathsf{M}Y = \int\limits_{-\infty}^{\infty} yp_Y(y)\,dy.$$

Оказывается, что математическое ожидание функции $f(X)$ от случайной величины X можно выразить непосредственно через закон распределения самой величины X по формулам:

$$\boxed{\mathsf{M}f(X) = \int_{-\infty}^{\infty} f(x)\, p_X(x)\, dx} \tag{1}$$

для непрерывной величины X с плотностью распределения $p_X(x)$,

$$\boxed{\mathsf{M}f(X) = \sum_i f(x_i)\, p_i} \tag{2}$$

для дискретной величины X с возможными значениями x_i и соответствующими вероятностями p_i (при этом и интеграл (1) и ряд (2) предполагаются абсолютно сходящимися, см. стр. 109 и 111).

Доказательство формулы (2) непосредственно вытекает из определения понятий функции от случайной величины и математического ожидания. Если все значения $f(x_i)$ различны, то, как отмечалось в § 2.7, таблицы распределения вероятностей величин X и $Y = f(X)$ отличаются только строкой возможных значений:

X	x_1	x_2	. . .
	p_1	p_2	. . .

Y	$f(x_1)$	$f(x_2)$	. . .
	p_1	p_2	. . .

Поэтому

$$\mathsf{M}Y = \sum_i f(x_i)\, p_i,$$

что сразу приводит к формуле (2). Если же некоторые из значений $f(x_i)$ совпадают, то во второй из написанных выше таблиц надо объединить соответствующие столбцы, сложив вероятности p_i, но это сведется только к перестановке слагаемых в формуле (2), что снова подтверждает ее справедливость уже в общем случае.

Доказательство формулы (1) в общем случае несколько более громоздко. Мы проверим эту формулу только для простейшего случая строго возрастающей функции $f(x)$. В этом случае переход от формулы

$$\mathsf{M}Y = \int_{-\infty}^{\infty} y p_Y(y)\,dy$$

к формуле

$$\mathsf{M}Y = \int_{-\infty}^{\infty} f(x)\,p_X(x)\,dx$$

означает просто замену переменной $y = f(x)$ в первом интеграле, так как $p_Y(y)\,dy = p_X(x)\,dx$ в соответствии с ранее установленной формулой (3.8-2).

Аналогичное положение имеет место и для функции нескольких случайных величин. Например, математическое ожидание функции двух случайных величин с плотностью совместного распределения $p(x, y)$ можно выразить непосредственно через эту плотность по формуле

$$\boxed{\mathsf{M}f(X, Y) = \iint_{-\infty}^{\infty} f(x, y)\,p(x, y)\,dx\,dy} \qquad (3)$$

(интеграл берется по всей плоскости xOy; предполагается, что он сходится абсолютно). Мы не будем доказывать эту формулу в общем случае, а лишь проверим ее для функции $Z = X + Y$, рассмотренной в § 3.8. Пользуясь формулой (3.8-7), запишем

$$\mathsf{M}(X+Y) = \mathsf{M}Z = \int_{-\infty}^{\infty} z p_Z(z)\,dz = \int_{-\infty}^{\infty} z\,dz \int_{-\infty}^{\infty} p(x, z-x)\,dx.$$

Переменим порядок интегрирования

$$\mathsf{M}(X+Y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} zp(x, z-x)\,dz$$

и во внутреннем интеграле заменим z на $x + y$; мы получим

$$M(X+Y) = \\ = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} (x+y)\, p(x, y)\, dy = \iint_{-\infty}^{\infty} (x+y)\, p(x, y)\, dx\, dy$$

Приведем простой и интересный пример применения формул вида (3).

Пример. Случайная точка бросается наудачу в круг (или шар) радиуса a. Найти среднее расстояние точки от центра круга (или шара).

Решение. Бросание точки наудачу в круг означает равномерное распределение точки в круге, т. е. распределение с плотностью

$$p(x, y) = \begin{cases} C = \dfrac{1}{\pi a^2} & \text{внутри круга} \\ 0 & \text{вне круга.} \end{cases}$$

Расстояние точки $(X; Y)$ от центра круга (который мы примем за начало координат) равно $R = \sqrt{X^2 + Y^2}$. Задача состоит в нахождении математического ожидания этого расстояния. По формуле (3) имеем

$$MR = \iint_{-\infty}^{\infty} \sqrt{x^2 + y^2}\, p(x, y)\, dx\, dy.$$

Переходя к полярным координатам (r, φ), получим

$$MR = C \int_0^{2\pi} d\varphi \int_0^a r \cdot r\, dr = \frac{1}{\pi a^2} \cdot 2\pi \cdot \frac{a^3}{3} = \frac{2}{3} a.$$

Для шара получим аналогично: $R = \sqrt{X^2 + Y^2 + Z^2}$,

$$MR = \iiint_{-\infty}^{\infty} \sqrt{x^2 + y^2 + z^2}\, p(x, y, z)\, dx\, dy\, dz$$

что в сферических координатах (r, θ, φ) даст

$$MR = C_1 \int_0^{2\pi} d\varphi \int_0^{\pi} \sin\theta\, d\theta \int_0^{a} r \cdot r^2\, dr =$$

$$= \frac{1}{\frac{4}{3}\pi a^3} \cdot 2\pi \cdot 2 \cdot \frac{a^4}{4} = \frac{3}{4} a.$$

Полученные результаты можно обобщить на «шар» в n-мерном пространстве: интеграл, выражающий среднее значение расстояния точки от центра «шара», отличается от интеграла, выражающего «объем» этого «шара», только дополнительным множителем r при интегрировании по радиусу; поэтому мы получим

$$MR = \frac{\int_0^a r \cdot r^{n-1}\, dr}{\int_0^a r^{n-1}\, dr} = \frac{a^{n+1}/(n+1)}{a^n/n} = \frac{n}{n+1} a.$$

С увеличением размерности n пространства среднее расстояние точки от центра «шара» возрастает и стремится к радиусу «шара». Этот факт находит применение в задачах планирования экстремальных экспериментов.

§ 4.3. Свойства математического ожидания как операции осреднения

Для нахождения математического ожидания, так же как и для построения некоторых других числовых характеристик распределения, важны свойства самой операции отыскания математического ожидания, которую мы будем коротко называть операцией осреднения. Эти свойства одинаковы и для математических ожиданий дискретных случайных величин и для математических ожиданий непрерывных случайных величин и для средних арифметических значений (4.1-1) Мы будем доказывать формулируемые ниже свойства только для непрерывных случайных величин, предоставляя читателю самостоятельно провести доказательства для дискретных величин. При этом мы будем предполагать, что для рассматриваемых величин математические ожидания существуют, т. е. что соответствующие интегралы абсолютно сходятся.

Важнейшим свойством операции осреднения является линейность. *Математическое ожидание линейной комбинации случайных величин равно линейной комбинации их математических ожиданий:*

$$\mathsf{M}\,(C_1X_1 + C_2X_2 + \ldots + C_nX_n) = \\ = C_1\mathsf{M}X_1 + C_2\mathsf{M}X_2 + \ldots + C_n\mathsf{M}X_n, \quad (1)$$

где $C_1, C_2, \ldots, C_n$ — постоянные.

Это свойство вытекает непосредственно из следующих двух теорем.

1. Постоянный множитель C можно выносить за знак математического ожидания:

$$\mathsf{M}CX = C\mathsf{M}X. \quad (2)$$

2. Математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий:

$$\mathsf{M}\,(X + Y) = \mathsf{M}X + \mathsf{M}Y. \quad (3)$$

Доказательства.

1. Для доказательства формулы (2) воспользуемся формулой (4.2-1) и сразу получим

$$\mathsf{M}CX = \int_{-\infty}^{\infty} Cxp_X(x)\,dx = C\int_{-\infty}^{\infty} xp_X(x)\,dx = C\mathsf{M}X.$$

2. Для доказательства формулы (3) воспользуемся формулой (4.2-3)

$$\mathsf{M}(X + Y) = \iint_{-\infty}^{\infty} (x + y)\,p(x, y)\,dx\,dy = \\ = \int_{-\infty}^{\infty} x\,dx \int_{-\infty}^{\infty} p(x, y)\,dy + \int_{-\infty}^{\infty} y\,dy \int_{-\infty}^{\infty} p(x, y)\,dx.$$

Учитывая формулы (3.6-8) и (3.6-9), получим окончательно

$$\mathsf{M}\,(X + Y) = \int_{-\infty}^{\infty} xp_X(x)\,dx + \int_{-\infty}^{\infty} yp_Y(y)\,dy = \mathsf{M}X + \mathsf{M}Y.$$

В дальнейшем нам понадобится еще следующее свойство математического ожидания.

Теорема умножения математических ожиданий. *Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.* Например, для двух независимых случайных величин

$$\mathsf{M}(XY) = \mathsf{M}X \cdot \mathsf{M}Y. \tag{4}$$

Доказательство снова получается из формулы (4.2-3):

$$\mathsf{M}(XY) = \iint\limits_{-\infty}^{\infty} xyp(x, y)\,dx\,dy = \int\limits_{-\infty}^{\infty} x\,dx \int\limits_{-\infty}^{\infty} yp(x, y)\,dy. \tag{5}$$

Условие независимости случайных величин X и Y дает

$$p(x, y) = p_X(x)\,p_Y(y).$$

Подставив это выражение в написанный выше интеграл, получаем

$$\mathsf{M}(XY) = \int\limits_{-\infty}^{\infty} xp_X(x)\,dx \int\limits_{-\infty}^{\infty} yp_Y(y)\,dy = \mathsf{M}X \cdot \mathsf{M}Y.$$

§ 4.4. Дисперсия, среднее квадратическое отклонение и другие характеристики рассеяния

Наряду с характеристиками положения большую роль как в теории, так и в приложениях играют характеристики рассеяния. Ведь чем меньше рассеиваются значения случайной величины, тем точнее можно их предсказать! Мы уже обращали внимание на рассеяние случайной величины при рассмотрении нормального распределения в § 3.3; там мы выяснили, что как раз параметр σ нормального распределения (3.3-1) полностью характеризует его рассеяние. Теперь мы введем аналогичную характеристику для любой случайной величины.

Рассеяние случайной величины X связано с отклонением $X - a$ этой величины от ее центра распределения $a = \mathsf{M}X$. Непосредственное осреднение этого отклонения не может дать числовой характеристики рассеяния, так как

$$\mathsf{M}\,(X - a) = \mathsf{M}X - a = 0$$

(отклонения противоположных знаков в среднем взаимно погашаются).

Основной числовой характеристикой рассеяния случайной величины X служит *среднее квадратическое отклонение* $\sigma(X)$, определяемое по формуле

$$\boxed{\sigma(X) = \sqrt{\mathsf{M}\,(X - a)^2}.} \tag{1}$$

Расчеты удобнее производить с подкоренным выражением, которое получило специальное название *дисперсии случайной величины* X и обозначается $\mathsf{D}X$:

$$\boxed{\mathsf{D}X = \mathsf{M}\,(X - a)^2.} \tag{2}$$

Таким образом, дисперсия есть средний квадрат отклонения случайной величины от ее центра распределения. Иногда вводят специальное обозначение для отклонения $X - a$:

$$X - a = \mathring{X}; \tag{3}$$

величину $\mathring{X}$ называют *центрированной случайной величиной*, ее математическое ожидание равно нулю:

$$\mathsf{M}\mathring{X} = \mathsf{M}\,(X - a) = 0,$$

а дисперсия совпадает с дисперсией величины X:

$$\mathsf{D}\mathring{X} = \mathsf{M}\mathring{X}^2 = \mathsf{D}X.$$

Пользуясь формулами (4.2-1) и (4.2-2) для математического ожидания функции, запишем формулы для дисперсии непрерывной величины

$$\boxed{\mathsf{D}X = \int_{-\infty}^{\infty} (x - a)^2\, p(x)\, dx} \tag{4}$$

и для дисперсии дискретной величины

$$\boxed{\mathsf{D}X = \sum_i (x_i - a)^2\, p_i,} \tag{5}$$

где сумма берется по всем возможным значениям x_i величины X.

Как видно из написанных выше формул, дисперсия $\mathsf{D}X$ имеет квадратичную размерность; следовательно, среднее квадратическое отклонение $\sigma(X)$ имеет ту же размерность, что и сама случайная величина X.

Для вычисления дисперсии часто бывает удобна формула

$$\mathsf{D}X = \mathsf{M}(X-a)^2 = \mathsf{M}(X-C)^2 - (a-C)^2, \tag{6}$$

где C — любое число. Эта формула вытекает из свойства линейности математического ожидания:

$$\begin{aligned}\mathsf{M}(X-C)^2 &= \mathsf{M}[(X-a)+(a-C)]^2 = \\ &= \mathsf{M}(X-a)^2 + (a-C)^2 + 2(a-C)\,\mathsf{M}(X-a) = \\ &= \mathsf{M}(X-a)^2 + (a-C)^2,\end{aligned}$$

так как $\mathsf{M}(X-a) = 0$. В частном случае при $C = 0$ получаем

$$\mathsf{D}X = \mathsf{M}X^2 - a^2 = \mathsf{M}X^2 - (\mathsf{M}X)^2. \tag{7}$$

Из формулы (6) вытекает, что средний квадрат отклонения случайной величины X от ее центра распределения $a = \mathsf{M}X$ меньше, чем средний квадрат ее отклонения от любого другого числа C:

$$\mathsf{M}(X-a)^2 < \mathsf{M}(X-C)^2 \qquad (C \neq a).$$

Таким образом, *центр распределения минимизирует средний квадрат отклонения* $\mathsf{M}(X-C)^2$, *причем минимум этого среднего квадрата отклонения равен как раз дисперсии* $\mathsf{D}X$.

Отметим еще следующие правила вычисления дисперсий и средних квадратических отклонений.

1. При линейном преобразовании случайной величины X, т. е. для линейной функции

$$Y = kX + b,$$

дисперсия увеличивается в k^2 раз, а среднее квадратическое отклонение — в $|k|$ раз:

$$\mathsf{D}Y = \mathsf{D}(kX+b) = k^2\mathsf{D}X, \tag{8}$$

$$\sigma(Y) = \sigma(kX+b) = |k|\,\sigma(X). \tag{9}$$

Доказательство. В силу линейности математического ожидания имеем

$$\mathsf{M}Y = k\mathsf{M}X + b = ka + b.$$

Поэтому

$$\mathsf{D}Y = \mathsf{M}\,[Y - (ka + b)]^2 = \mathsf{M}\,(kX - ka)^2 = k^2\mathsf{M}\,(X - a)^2.$$

Доказанные формулы означают, что изменение начала отсчета значений случайной величины не влияет на ее дисперсию, а изменение масштаба приводит к соответствующему изменению среднего квадратического отклонения. В частности, если ввести так называемую *нормированную случайную величину*

$$X_{\text{норм}} = \frac{X - a}{\sigma(X)} = \frac{\overset{\circ}{X}}{\sigma(\overset{\circ}{X})}, \tag{10}$$

т. е. если вести отсчет значений случайной величины X от ее центра a, а за единицу масштаба принять $\sigma(X)$, то мы получим

$$\mathsf{M}X_{\text{норм}} = 0, \quad \mathsf{D}X_{\text{норм}} = \frac{\mathsf{D}X}{\sigma^2(X)} = 1. \tag{11}$$

2. Теорема сложения дисперсий: *если случайные величины X и Y независимы, то дисперсия их суммы равна сумме их дисперсий*:

$$\mathsf{D}\,(X + Y) = \mathsf{D}X + \mathsf{D}Y, \tag{12}$$

и следовательно,

$$\sigma(X + Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}. \tag{13}$$

Доказательство. Обозначим $\mathsf{M}X = a$, $\mathsf{M}Y = b$. В силу линейности математического ожидания имеем

$$\mathsf{M}\,(X + Y) = \mathsf{M}X + \mathsf{M}Y = a + b;$$

поэтому

$$\mathsf{D}\,(X + Y) = \mathsf{M}\,[(X + Y) - (a + b)]^2 = \mathsf{M}\,[(X - a) + (Y - b)]^2 = \mathsf{M}\,(X - a)^2 + \mathsf{M}\,(Y - b)^2 + 2\mathsf{M}\,(X - a)\,(Y - b),$$

или

$$\mathsf{D}\,(X+Y)=\mathsf{D}X+\mathsf{D}Y+2\mathsf{M}\,(X-a)(Y-b). \quad (14)$$

Далее остается только воспользоваться независимостью случайных величин X и Y и применить теорему умножения математических ожиданий:

$$\mathsf{M}\,(X-a)(Y-b)=\mathsf{M}\,(X-a)\cdot\mathsf{M}\,(Y-b)=0,$$

что и завершает доказательство.

Теорема сложения дисперсий без труда обобщается на любое число попарно независимых случайных величин, что во многих задачах позволяет существенно упростить расчеты (см. примеры в следующем параграфе).

Следствие. *Дисперсия линейной комбинации попарно независимых случайных величин* X_1, X_2, ..., X_n *может быть вычислена по формуле*

$$\mathsf{D}\,(C_1X_1+C_2X_2+\ldots+C_nX_n)=\\ =C_1^2\mathsf{D}X_1+C_2^2\mathsf{D}X_2+\ldots+C_n^2\mathsf{D}X_n. \quad (15)$$

Это непосредственно вытекает из формул (12) и (8).

В частности, если все величины $X_1, X_2, \ldots, X_n$ имеют одинаковую дисперсию

$$\mathsf{D}X_i=\sigma^2 \qquad (i=1, 2, \ldots, n),$$

то дисперсия их среднего арифметического равна

$$\mathsf{D}\,\frac{X_1+X_2+\ldots+X_n}{n}=\frac{\mathsf{D}X_1+\mathsf{D}X_2+\ldots+\mathsf{D}X_n}{n^2}=\frac{\sigma^2}{n} \quad (16)$$

и, следовательно, среднее квадратическое отклонение равно

$$\boxed{\sigma\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)=\frac{\sigma}{\sqrt{n}}.} \quad (17)$$

Последняя формула играет большую роль при обработке результатов измерений.

Некоторые другие характеристики рассеяния. Иногда вместо среднего квадратического отклонения применяется *среднее абсолютное отклонение от центра распределения*

$$\theta=\mathsf{M}\,|X-a|, \qquad \text{где } a=\mathsf{M}X$$

(или же среднее абсолютное отклонение от медианы). Хотя среднее абсолютное отклонение легче поддается статистическому толкованию, оно неудобно для теоретических расчетов и поэтому применяется редко.

Другой способ построения характеристик рассеяния связан с введением квантилей. *Квантилью* $x_{\mathscr{P}}$ непрерывной случайной величины X называется такое значение $x_{\mathscr{P}}$, для которого

$$F(x_{\mathscr{P}}) = \mathsf{P}(X < x_{\mathscr{P}}) = \mathscr{P}.$$

В частности, вводят *квартили*

$$x_{\frac{1}{4}},\ x_{\frac{2}{4}} = x_{\frac{1}{2}},\ x_{\frac{3}{4}};$$

прямые $x = x_{\frac{1}{4}}$, $x = x_{\frac{1}{2}}$, $x = x_{\frac{3}{4}}$ делят площадь под кривой распределения вероятностей на 4 равновеликие части. Квартиль $x_{\frac{1}{2}}$ есть медиана, которую мы уже встречали в § 4.1. Квартили $x_{\frac{1}{4}}$, $x_{\frac{3}{4}}$ применяются для характеристики рассеяния: их разность $x_{\frac{3}{4}} - x_{\frac{1}{4}}$ называется *полушириной распределения.*

Для дискретной случайной величины с конечными возможными значениями вместо этой характеристики применяют полуразность между наибольшим и наименьшим ее возможными значениями.

§ 4.5. Математические ожидания и дисперсии некоторых дискретных случайных величин

В настоящем параграфе мы найдем центры распределения и дисперсии случайных величин, рассмотренных в §§ 2.2—2.4, а также связанных с ними, и дадим им соответствующие статистические толкования.

Распределение выборочного значения признака и выборочное среднее значение. В § 2.2 мы рассмотрели значение признака у случайно выбранного элемента как случайную величину X, закон распределения которой дается формулой

$$\mathsf{P}(X = x_i) = \frac{M_i}{N} \qquad (i = 1, 2, \ldots, k). \tag{1}$$

Здесь $x_1, x_2, \ldots, x_k$ — различные значения признака у элементов той совокупности, из которой выбирается наудачу один элемент; такую исходную совокупность мы будем называть *генеральной совокупностью*. Через M_i обозначено число элементов генеральной совокупности, у которых значение признака равно x_i ($i = 1, 2, \ldots, k$), а через $N = M_1 + M_2 + \ldots + M_k$ — общее число элементов генеральной совокупности (*объем генеральной совокупности*). Найдем математическое ожидание величины X. По формуле (4.1-3) имеем

$$\mathsf{M}X = \sum_{i=1}^{k} x_i \frac{M_i}{N} = \frac{x_1 M_1 + x_2 M_2 + \ldots + x_k M_k}{N}. \qquad (2)$$

Это означает, что математическим ожиданием выборочного значения признака служит как раз среднее арифметическое значение признака в генеральной совокупности (*генеральное среднее значение*):

$$\mathsf{M}X = \bar{x}.$$

Точно так же дисперсия величины X

$$\mathsf{D}X = \sum_{i=1}^{k} (x_i - \bar{x})^2 \frac{M_i}{N} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 M_i}{N}$$

совпадает со средним квадратом отклонения значений признака от их среднего значения в генеральной совокупности.

Будем теперь повторять испытание по схеме повторной выборки: взяв наудачу один элемент из совокупности, мы отмечаем обнаруженное у него значение признака, возвращаем выбранный элемент в генеральную совокупность (чтобы не изменить состав генеральной совокупности) и затем выбираем наудачу следующий элемент. Этот процесс мы повторяем до получения n отметок; полученные отметки дают *случайную выборку объема n*. Среднее арифметическое значение признака в такой выборке называется *выборочным средним значением*. Найдем его математическое ожидание. Обозначим значение признака у первого выборочного элемента через X_1, у второго — через X_2, ..., у n-го —

через X_n. Выборочное среднее есть случайная величина

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}. \tag{3}$$

Так как элементы выбирались по схеме повторной выборки, то все случайные величины $X_1, X_2, \ldots, X_n$ имеют один и тот же закон распределения вероятностей (1), следовательно, одно и то же математическое ожидание (2):

$$\mathsf{M}X_j = \bar{x} \qquad (j = 1, 2, \ldots, n).$$

Математическое ожидание выборочного среднего (3) легко найдем с помощью свойства линейности математического ожидания

$$\mathsf{M}\overline{X} = \frac{\mathsf{M}X_1 + \mathsf{M}X_2 + \ldots + \mathsf{M}X_n}{n} = \bar{x}. \tag{4}$$

Таким образом, *математическим ожиданием выборочного среднего $\overline{X}$ служит генеральное среднее $\bar{x}$.*

Рассмотренная задача допускает следующее обобщение. Под *случайной выборкой объема n* понимают просто совокупность n независимых случайных величин $X_1, X_2, \ldots, X_n$ с одним и тем же законом распределения (не обязательно даже дискретным). Этот закон распределения называют *генеральным законом распределения* или *законом распределения признака в генеральной совокупности*, его математическое ожидание

$$\mathsf{M}X_1 = \mathsf{M}X_2 = \ldots = \mathsf{M}X_n = a$$

— *генеральным средним*, а его дисперсию

$$\mathsf{D}X_1 = \mathsf{D}X_2 = \ldots = \mathsf{D}X_n = \sigma^2$$

— *генеральной дисперсией*. При этом для выборочного среднего $\overline{X}$, определенного формулой (3), мы находим так же, как и выше,

$$\mathsf{M}\overline{X} = a,$$

т. е. математическое ожидание выборочного среднего совпадает с генеральным средним. Но, в соответствии с формулой (4.4-16) предыдущего параграфа, дисперсия выборочного среднего не равна генеральной дисперсии, а меньше ее в n раз:

$$\mathsf{D}\overline{X} = \frac{\sigma^2}{n}.$$

Геометрическое распределение. В § 2.2 геометрическое распределение (2.2-2) толковалось как распределение вероятностей числа X бросаний кольца до первого попадания в цель. Математическое ожидание случайной величины X можно при этом толковать как среднее число бросаний до первого попадания или как среднее число колец, израсходованных на одно попадание. Так как множество возможных значений величины X бесконечно (игра ведется без ограничения числа бросаний), то математическое ожидание величины X надо вычислять по формуле (4.1-4) как сумму бесконечного ряда

$$\mathsf{M}X = \sum_{m=1}^{\infty} m\mathsf{P}(X=m) =$$

$$= \sum_{m=1}^{\infty} mq^{m-1}p = p\sum_{m=1}^{\infty} mq^{m-1} \quad (p+q=1).$$

По известным формулам для степенных рядов находим

$$\sum_{m=0}^{\infty} q^m = 1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q},$$

$$\sum_{m=1}^{\infty} mq^{m-1} = 1 + 2q + 3q^2 + \ldots = \left(\frac{1}{1-q}\right) = \frac{1}{(1-q)^2}$$

и, следовательно,

$$\mathsf{M}X = p\frac{1}{(1-q)^2} = \frac{1}{p} \qquad (p>0). \tag{5}$$

Итак, среднее число бросаний до первого попадания есть число, обратное вероятности попадания при одном бросании. Например, если $p = 0{,}1$, то $\mathsf{M}X = 10$, что можно толковать так: если попадание в цель происходит в среднем 1 раз из 10 бросаний, то для попадания в цель требуется в среднем 10 бросаний (10 колец).

Пусть теперь при тех же условиях, что и выше, ведется игра до n-го попадания в цель. Найдем математическое ожидание числа Y произведенных бросаний (т. е. среднее число израсходованных колец на n попаданий). Обозначим через X_1 число бросаний до первого попадания, через X_2 — число бросаний от первого до второго попадания и т. д. Тогда мы сможем представить общее число Y бросаний до n-го попадания в виде суммы

$$Y = X_1 + X_2 + \ldots + X_n;$$

для вычисления его математического ожидания воспользуемся свойством линейности:

$$\mathsf{M}Y = \mathsf{M}X_1 + \mathsf{M}X_2 + \ldots + \mathsf{M}X_n.$$

Так как по условию задачи успешное попадание в цель не меняет вероятности попадания при следующих бросаниях, то для любой случайной величины X_j имеем одно и то же геометрическое распределение с одним и тем же математическим ожиданием

$$\mathsf{M}X_j = \frac{1}{p} \qquad (j = 1, 2, \ldots, n),$$

следовательно,

$$\mathsf{M}Y = n \cdot \frac{1}{p}.$$

Среднее число колец, израсходованных на n попаданий, в n раз превосходит среднее число колец, израсходованных на одно попадание.

Подсчитаем еще дисперсии величин X и Y, чтобы оценить характер рассеяния указанных средних. Для подсчета дисперсии величины X воспользуемся формулой (4.4-7) и известными формулами для сумм степенных рядов

$$\sum_{m=1}^{\infty} mq^{m-1} = \frac{1}{(1-q)^2}, \quad \sum_{m=2}^{\infty} m(m-1)\, q^{m-2} = \frac{2}{(1-q)^3}.$$

Мы получим

$$\mathsf{D}X = \mathsf{M}X^2 - (\mathsf{M}X)^2 = \sum_{m=1}^{\infty} m^2 q^{m-1} p - \left(\frac{1}{p}\right)^2 =$$

$$= p\left[\sum_{m=1}^{\infty} m(m-1)\, q^{m-1} + \sum_{m=1}^{\infty} m q^{m-1}\right] - \frac{1}{p^2} =$$

$$= p\left[q\frac{2}{(1-q)^3} + \frac{1}{(1-q)^2}\right] - \frac{1}{p^2} = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Для подсчета дисперсии величины Y можно воспользоваться теоремой сложения дисперсий, так как введенные выше величины X_1, X_2, ..., X_n, очевидно, независимы:

$$\mathsf{D}Y = \mathsf{D}X_1 + \mathsf{D}X_2 + \ldots + \mathsf{D}X_n = n \cdot \frac{1-p}{p^2}.$$

Полученные дисперсии тем меньше, чем ближе p к 1, т. е. чем точнее бросают кольцо. При увеличении n относительная ошибка в определении числа требуемых колец уменьшается. Например, если $p = 1/2$, $n = 100$, то

$$\mathsf{M}X = 2, \quad \mathsf{D}X = 2, \quad \sigma(X) = \sqrt{2} \approx 1{,}4,$$

$$\mathsf{M}Y = 200, \quad \mathsf{D}Y = 200, \quad \sigma(Y) = \sqrt{200} \approx 14.$$

Заметим, что примененный здесь метод подсчета математического ожидания и дисперсии величины Y может быть применен и в том случае, когда вероятность попадания в цель изменяется после каждого успешного попадания.

Биномиальное распределение (см. § 2.3). Непосредственное вычисление математического ожидания и тем более дисперсии биномиального распределения связано с громоздкими преобразованиями. Для упрощения расчетов представим рассматриваемую случайную величину X — число успехов при n-кратном повторении испытания — в виде суммы более простых случайных величин. В качестве таких величин мы возьмем индикаторы успехов, введенные на стр. 41. Индикатор X_k принимает значение 1 в случае успеха при k-м повторении испытания и значение 0 в противном случае. Поэтому

$$X = X_1 + X_2 + \ldots + X_n,$$

так как эта сумма состоит из единиц и нулей, причем число единиц в ней равно числу успехов при n-кратном повторении испытания (см. также пример 4 на стр. 66). Отсюда сразу следует, что

$$\mathsf{M}X = \mathsf{M}X_1 + \mathsf{M}X_2 + \ldots + \mathsf{M}X_n$$

(в силу свойства линейности математического ожидания). Так как биномиальное распределение связано с последовательностью независимых испытаний по схеме Бернулли, то индикаторы $X_1, X_2, \ldots, X_n$ — независимые случайные величины; поэтому можно применить и теорему сложения дисперсий, что дает

$$\mathsf{D}X = \mathsf{D}X_1 + \mathsf{D}X_2 + \ldots + \mathsf{D}X_n.$$

Остается найти математическое ожидание и дисперсию индикаторов. По условию задачи вероятность успеха при каждом повторении испытания одна и та же и равна p. Поэтому распределение вероятностей любого индикатора X_k дается таблицей

X_k	1	0
	p	q

, где $q = 1 - p$; $k = 1, 2, \ldots, n$.

Непосредственный подсчет математического ожидания и дисперсии индикатора X_k уже не составляет труда:

$$\mathsf{M}X_k = 1 \cdot p + 0 \cdot q = p;$$

$$\mathsf{D}X_k = \mathsf{M}(X_k - p)^2 = (1 - p)^2 \cdot p + (0 - p)^2 \cdot q = q^2 p + p^2 q = pq.$$

Следовательно, для биномиального распределения имеем

$$\boxed{\mathsf{M}X = np,} \qquad (6)$$

$$\boxed{\mathsf{D}X = npq,} \qquad (7)$$

и, значит,

$$\boxed{\sigma(X) = \sqrt{npq}.} \qquad (8)$$

Примечание. Метод разложения числа X успехов в сумму индикаторов X_k позволяет легко получить и более общий результат для задачи, где вероятность успеха от испытания к испытанию изменяется: если вероятность успеха в k-м испытании равна p_k, то

$$\mathsf{M}X_k = p_k,\quad \mathsf{D}X_k = p_k q_k,\quad \text{где}\quad q_k = 1 - p_k;$$

и

$$\mathsf{M}X = \sum_{k=1}^{n} p_k;\quad \mathsf{D}X = \sum_{k=1}^{n} p_k q_k.$$

Нахождение закона распределения вероятностей в такой задаче значительно сложнее (см. пример 4 на стр. 66, где $n = 2$).

Полученная формула (6) вполне согласуется с нашим интуитивным представлением о математическом ожидании числа успехов. Если, например, вероятность успеха равна $p = 0{,}2$ и испытание повторяется $n = 100$ раз, то мы ожидаем, что успех наступит примерно $np = 20$ раз.

Найдем еще математическое ожидание и дисперсию относительной частоты X/n:

$$\mathsf{M}\frac{X}{n} = \frac{1}{n}\mathsf{M}X = p, \tag{9}$$

$$\mathsf{D}\frac{X}{n} = \frac{1}{n^2}\mathsf{D}X = \frac{pq}{n},\quad \sigma\left(\frac{X}{n}\right) = \sqrt{\frac{pq}{n}}. \tag{10}$$

Статистическое толкование формулы (9) возвращает нас к исходной связи между понятиями относительной частоты и вероятности случайного события: *математическое ожидание относительной частоты случайного события есть вероятность этого события*. Формула (10) показывает, что рассеяние относительной частоты уменьшается с увеличением числа повторений испытания.

Распределение Пуассона (см. § 2.4). Число X событий, наступающих за время t в простейшем потоке, имеет закон распределения вероятностей вида

$$\mathsf{P}(X = m) = \frac{a^m}{m!}e^{-a}\qquad (m = 0, 1, 2, \ldots;\ a = \lambda t).$$

Его математическое ожидание можно подсчитать непосредственно по формуле (4.1-4)

$$\mathsf{M}X=\sum_{m=0}^{\infty} m\cdot\frac{a^m}{m!}e^{-a}=e^{-a}\cdot a\cdot\sum_{m=1}^{\infty}\frac{a^{m-1}}{(m-1)!}=e^{-a}\cdot a\cdot e^{a}=a.$$

Это позволяет выяснить статистический смысл параметра a в распределении Пуассона: параметр a есть среднее число событий, наступающих за время t в простейшем потоке. Отсюда уже можно дать статистическое толкование параметру λ в формуле (2.4-1): так как $\lambda = a/t$, то параметр λ есть среднее число событий, наступающих за единицу времени (см. также стр. 55).

Для подсчета дисперсии найдем сначала значения

$$\mathsf{M}X(X-1)=$$

$$=\sum_{m=0}^{\infty} m(m-1)\cdot\frac{a^m}{m!}e^{-a}=e^{-a}\cdot a^2\sum_{m=2}^{\infty}\frac{a^{m-2}}{(m-2)!}=a^2,$$

$$\mathsf{M}X^2=\mathsf{M}X(X-1)+\mathsf{M}X=a^2+a.$$

Далее по формуле (4.4-7) находим

$$\mathsf{D}X=\mathsf{M}X^2-(\mathsf{M}X)^2=(a^2+a)-a^2=a.$$

Следовательно, дисперсия распределения Пуассона совпадает с математическим ожиданием. Эти результаты хорошо согласуются с представлением распределения Пуассона в качестве предельного для биномиального распределения при $p\to 0$, $n\to\infty$, $np = a$ (см. стр. 56). Действительно, при этом

$$\mathsf{M}X=np=a,\quad \mathsf{D}X=npq=a(1-p)\to a.$$

§ 4.6. Математические ожидания и дисперсии некоторых непрерывных случайных величин

Мы начнем с нормального распределения, как наиболее важного из рассмотренных выше.

Стандартное нормальное распределение с плотностью

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

имеет центр распределения $\mathsf{M}X = 0$ в силу симметрии; его дисперсия равна

$$\mathsf{D}X = \mathsf{M}X^2 =$$
$$= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx.$$

Подстановка $x^2/2 = t$, $x\,dx = dt$ в последнем интеграле приводит его к гамма-функции (см. стр. 75):

$$\mathsf{D}X = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} dx =$$
$$= \frac{2}{\sqrt{\pi}} \int_0^{\infty} t^{\frac{1}{2}} e^{-t} dt = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 1.$$

Если случайная величина Y имеет общее нормальное распределение с плотностью

$$\varphi_{a;\sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-a)^2}{2\sigma^2}\right],$$

то ее можно рассматривать как линейную функцию от случайной величины X со стандартным нормальным распределением, а именно (см. стр. 96):

$$X = \frac{Y-a}{\sigma}, \quad Y = \sigma X + a;$$

отсюда сразу следует, что

$$\mathsf{M}Y = \sigma \mathsf{M}X + a = a, \tag{1}$$
$$\mathsf{D}Y = \sigma^2 \mathsf{D}X = \sigma^2. \tag{2}$$

Таким образом, в нормальном распределении оба параметра имеют простой статистический смысл: *параметр a есть центр распределения, параметр σ^2 есть дисперсия распределения*; следовательно, *стандартное отклонение совпадает здесь со средним квадратическим отклонением*. Напомним, что мы уже указывали на статистический смысл параметров нормального распределения еще до введения соответствующих числовых характеристик распределения в общем случае (см. § 3.3).

Равномерное распределение в интервале (α, β) (см. § 3.2). Как мы видели на стр. 112, центр равномерного распределения совпадает с серединой интервала (α, β):

$$\mathsf{M}X = \frac{\alpha+\beta}{2}. \tag{3}$$

Для подсчета дисперсии заметим, что центрированная величина $\mathring{X} = X - \frac{\alpha+\beta}{2}$ имеет равномерное распределение в симметричном интервале $(-\gamma, \gamma)$, где $\gamma = \frac{\beta-\alpha}{2}$. Поэтому

$$\mathsf{D}X = \mathsf{D}\mathring{X} = \mathsf{M}\mathring{X}^2 = \int_{-\gamma}^{\gamma} x^2 \cdot \frac{1}{2\gamma}\,dx = \frac{\gamma^2}{3} = \frac{(\beta-\alpha)^2}{12}. \tag{4}$$

Отсюда следует, что при равномерном распределении в интервале среднее квадратическое отклонение пропорционально длине интервала:

$$\sigma(X) = \frac{\beta-\alpha}{2\sqrt{3}}. \tag{5}$$

Показательное распределение (см. § 3.2). Для показательного распределения с плотностью (3.2-5) математическое ожидание вычисляется непосредственно по определению:

$$\mathsf{M}T = \int_{0}^{\infty} t \cdot \lambda e^{-\lambda t}\,dt = \frac{1}{\lambda}. \tag{6}$$

Если толковать T как промежуток времени между двумя последовательными событиями в простейшем потоке, то полученный результат соответствует установленному выше статистическому смыслу параметра λ как среднего числа событий, наступающих в единицу времени. Естественно, что средний промежуток времени между двумя последовательными событиями в простейшем потоке равен $1/\lambda$. Например, если за одну минуту в среднем наступает $\lambda = 5$ событий, то средний промежуток времени между двумя последовательными событиями равен $1/5$ минуты, что вполне согласуется с интуитивными представлениями. Дисперсия показательного распределения вычисляется по формуле (4.4-7)

$$\mathsf{D}T = \mathsf{M}T^2 - \left(\frac{1}{\lambda}\right)^2 = \int_0^\infty t^2 \cdot \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

так что среднее квадратическое отклонение здесь равно математическому ожиданию

$$\sigma(T) = \frac{1}{\lambda}. \tag{7}$$

Гамма-распределение (см. § 3.2). С помощью свойств гамма-функции (см. стр. 76) из формулы для плотности

$$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad (x > 0), \quad p(x) = 0 \quad (x \leqslant 0)$$

получаем

$$\mathsf{M}X = \int_0^\infty x p(x)\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda},$$

$$\mathsf{M}X^2 = \int_0^\infty x^2 \cdot p(x)\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-\lambda x}\, dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{(\alpha+1)\alpha}{\lambda^2} \tag{8}$$

и, следовательно,

$$\mathsf{D}X = \mathsf{M}X^2 - (\mathsf{M}X)^2 = \frac{(\alpha+1)\alpha}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}. \tag{9}$$

Из формул (8) и (9) следует, что параметры α и λ связаны с центром распределения и дисперсией соотношениями

$$\lambda = \frac{\mathsf{M}X}{\mathsf{D}X}, \quad \alpha = \frac{(\mathsf{M}X)^2}{\mathsf{D}X} \quad \text{или} \quad \sqrt{\alpha} = \frac{\mathsf{M}X}{\sigma(X)}. \tag{10}$$

Отношение $\frac{\sigma(X)}{\mathsf{M}X}$ среднего квадратического отклонения к математическому ожиданию случайной величины называют ее *коэффициентом вариации* и обозначают C_v; таким образом,

$$\sqrt{\alpha} = \frac{1}{C_v}.$$

В частности, для χ^2-распределения с k степенями свободы (т. е. при $\lambda = 1/2$, $\alpha = k/2$) формулы (8) и (9) дают $\mathsf{M}\chi^2 = k$,

$\mathsf{D}\chi^2 = 2k$; первая из этих формул может быть получена непосредственно из определения $\chi_k^2 = X_1^2 + X_2^2 + \ldots + X_k^2$, где X_1, X_2, ..., X_k — независимые случайные величины со стандартным нормальным распределением (см. стр. 102), так как

$$\mathsf{M}X_i^2 = \mathsf{D}X_i = 1 \quad (i = 1, 2, \ldots, k).$$

§ 4.7. Понятие о моментах распределения

Введенные выше две основные характеристики распределения — центр распределения $\mathsf{M}X = a$ и дисперсия $\mathsf{D}X = \mathsf{M}(X - a)^2$ — представляют собой частные случаи моментов распределения, введенных выдающимся русским математиком П. Л. Чебышевым для исследования законов распределения вероятностей.

Начальным моментом порядка k случайной величины X называется математическое ожидание k-й степени этой величины; начальные моменты обозначаются так:

$$\alpha_k = \mathsf{M}X^k. \tag{1}$$

Начальный момент первого порядка есть центр распределения случайной величины X:

$$\alpha_1 = \mathsf{M}X = a. \tag{2}$$

Центральным моментом порядка k случайной величины X называется начальный момент того же порядка центрированной случайной величины $\mathring{X} = X - a$; центральные моменты обозначаются так:

$$\mu_k = \mathsf{M}\mathring{X}^k = \mathsf{M}(X - a)^k. \tag{3}$$

Центральный момент первого порядка равен нулю:

$$\mu_1 = \mathsf{M}\mathring{X} = \mathsf{M}(X - a) = 0, \tag{4}$$

а центральный момент второго порядка есть дисперсия:

$$\mu_2 = \mathsf{M}(X - a)^2 = \mathsf{D}X. \tag{5}$$

Центральные моменты не зависят от начала отсчета значений случайной величины: при сдвиге $Y = X - C$ центр распределения тоже сдвигается ($\mathsf{M}Y = \mathsf{M}X - C$), и поэтому

$$\mathring{Y} = Y - \mathsf{M}Y = X - \mathsf{M}X = \mathring{X}.$$

Если распределение случайной величины X симметрично относительно ее математического ожидания $\mathsf{M}X = a$, т. е. если кривая распределения вероятностей симметрична относительно прямой $x = a$, то центральные моменты нечетных порядков равны нулю (разумеется, в тех случаях, когда соответствующие несобственные интегралы сходятся абсолютно). Действительно, в этом случае плотность распределения центрированной величины $\overset{\circ}{X} = X - a$ будет четной функцией $\psi(x)$, и поэтому

$$\mu_{2k+1} = \int_{-\infty}^{\infty} x^{2k+1} \psi(x)\,dx = 0 \qquad (k = 0, 1, 2, \ldots).$$

В частности, это справедливо для нормального и равномерного распределений.

Поэтому отличие центральных моментов нечетных порядков от нуля свидетельствует об асимметрии распределения. Обычно для характеристики асимметрии

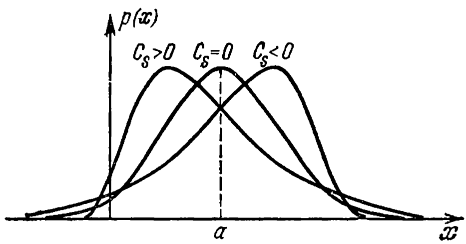


Рис. 32.

распределения используют центральный момент третьего порядка:

$$\mu_3 = \mathsf{M}\,(X - a)^3.$$

Чаще всего применяют безразмерный **коэффициент асимметрии**

$$C_s = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\mathsf{M}\,(X - a)^3}{\sigma^3(X)}. \tag{6}$$

Знак коэффициента асимметрии указывает на правостороннюю или левостороннюю асимметрию (рис. 32).

Иногда применяют еще центральный момент четвертого порядка $\mu_4 = \mathsf{M}(X - a)^4$ для характеристики сглаженности кривой распределения по отношению к кривой нормального распределения. Соответствующий безразмерный коэффициент называют **эксцессом** и определяют по формуле

$$E = \frac{\mu_4}{\sigma^4(X)} - 3.$$

(Для нормального распределения центральный момент четвертого порядка равен $\mu_4 = 3\sigma^4$, поэтому эксцесс равен нулю, см. ниже пример 1.)

Для расчета моментов часто оказываются полезными следующие соотношения между начальными и центральными моментами:

$$\left.\begin{aligned} \mu_2 &= \alpha_2 - \alpha_1^2, \\ \mu_3 &= \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3, \\ \mu_4 &= \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4, \end{aligned}\right\} \tag{7}$$

и т. д. Эти соотношения легко устанавливаются с помощью свойства линейности математического ожидания, например,

$$\mu_3 = \mathsf{M}(X - a)^3 = \mathsf{M}X^3 - 3a\mathsf{M}X^2 + 3a^2\mathsf{M}X - a^3 = \\ = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3.$$

Так же просто получаются и выражения начальных моментов через центральные, например,

$$\alpha_3 = \mathsf{M}X^3 = \mathsf{M}(X - a + a)^3 = \\ = \mathsf{M}(X - a)^3 + 3a\,\mathsf{M}(X - a)^2 + \\ + 3a^2\mathsf{M}(X - a) + a^3 = \mu_3 + 3a\mu_2 + a^3.$$

Пример 1. Найти центральные моменты нормального распределения.

Решение. Как мы уже знаем, в силу симметрии распределения все центральные моменты нечетных порядков равны нулю. Найдем центральные моменты порядка $2k$, где k — любое натуральное число. Пусть случайная величина X имеет стандартное нормальное распределение с параметрами $\mathsf{M}X = 0$ и $\mathsf{D}X = 1$. Тогда случайная величина $Y = \sigma X + a$, где $\sigma > 0$, имеет нормальное распределение с параметрами $\mathsf{M}Y = a$ и $\mathsf{D}Y = \sigma^2$ (см. стр. 134)

По определению центрального момента порядка $2k$ имеем

$$\mu_{2k} = \mathsf{M}(Y-a)^{2k} = \sigma^{2k}\mathsf{M}X^{2k} = \frac{2\sigma^{2k}}{\sqrt{2\pi}}\int_0^\infty x^{2k}\exp\left(-\frac{x^2}{2}\right)dx.$$

Подстановка $x^2/2 = t$, $x\,dx = dt$ приводит последний интеграл к гамма-функции, и мы получаем

$$\mu_{2k} = \sigma^{2k}\frac{\sqrt{2}}{\sqrt{\pi}}2^{k-\frac{1}{2}}\int_0^\infty t^{k-\frac{1}{2}}e^{-t}\,dt =$$

$$= \sigma^{2k}\frac{2^k}{\sqrt{\pi}}\Gamma\left(k+\frac{1}{2}\right) = \sigma^{2k}\cdot 1\cdot 3\cdot 5\cdots(2k-1). \qquad (8)$$

В частности,

$$\mu_2 = \sigma^2, \quad \mu_4 = 3\sigma^4, \quad \mu_6 = 15\sigma^6.$$

П р и м е р 2. Найти коэффициент асимметрии гамма-распределения (§ 3.2).

Р е ш е н и е. Найдем сначала начальный момент третьего порядка:

$$\alpha_3 = \mathsf{M}X^3 = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+2}e^{-\lambda x}dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\,\frac{\Gamma(\alpha+3)}{\lambda^{\alpha+3}} =$$

$$= \frac{(\alpha+2)(\alpha+1)\alpha}{\lambda^3}.$$

Теперь подсчитаем центральный момент по формуле (7)

$$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3 = \frac{(\alpha+2)(\alpha+1)\alpha}{\lambda^3} -$$

$$-3\frac{(\alpha+1)\alpha}{\lambda^2}\cdot\frac{\alpha}{\lambda} + 2\left(\frac{\alpha}{\lambda}\right)^3 = \frac{2\alpha}{\lambda^3}.$$

Наконец, учитывая формулу (4.6-9), получаем

$$C_s = \frac{\mu_3}{\sigma^3(X)} = \frac{2\alpha}{\alpha^{3/2}} = \frac{2}{\sqrt{\alpha}} > 0.$$

Заметим, что для гамма-распределения коэффициент асимметрии вдвое больше коэффициента вариации C_v, введенного на стр. 136.

В частности, при $\alpha = 1$ получаем коэффициент асимметрии для показательного распределения $C_s = 2$.

Отметим еще, что для χ^2-распределения с k степенями свободы коэффициент асимметрии $C_s = \frac{2}{\sqrt{k/2}} = \sqrt{\frac{2}{k}}$ стремится к нулю при $k \to \infty$.

Пример 3. Найти коэффициент асимметрии биномиального распределения.

Решение. В § 4.5 мы видели, что для подсчета моментов биномиального распределения удобно разложить рассматриваемую величину X — число успехов за n испытаний по схеме Бернулли в сумму индикаторов:

$$X = X_1 + X_2 + \ldots + X_n.$$

Центральный момент третьего порядка для индикатора X_i можно подсчитать непосредственно:

$$\mathsf{M}\,(X_i - p)^3 = (1 - p)^3 \cdot p + (0 - p)^3 \cdot q = \\ = q^3 p - p^3 q = pq\,(q^2 - p^2)$$

или, учитывая, что $p + q = 1$,

$$\mathsf{M}\,(X_i - p)^3 = pq\,(q - p).$$

Так как величины $X_1, X_2, \ldots, X_n$ независимы, то можно применить теорему сложения центральных моментов третьего порядка (см. упр. 11 в конце главы):

$$\mu_3 = \mathsf{M}\,(X - np)^3 = npq\,(q - p).$$

Таким образом, коэффициент асимметрии равен $C_s = \frac{npq\,(q - p)}{(npq)^{3/2}} = \frac{q - p}{\sqrt{npq}}$. В частном случае при $p = q = 1/2$ коэффициент асимметрии равен нулю. В общем случае при неограниченном увеличении числа n испытаний коэффициент асимметрии стремится к нулю.

Понятие о моментах распределения многомерных случайных величин

Мы здесь ограничимся определением моментов распределения только для двумерных случайных величин, распространение этих понятий на случайные величины размерности более двух (или системы более двух случайных величин) не представляет труда.

Начальные моменты порядка $k+l$ для случайной величины $(X;\ Y)$ (или системы случайных величин X и Y) определяются по формуле

$$\alpha_{k,l} = \mathsf{M}X^kY^l. \tag{9}$$

В частности, для непрерывной величины с плотностью совместного распределения $p(x, y)$:

$$\alpha_{k,l} = \iint\limits_{-\infty}^{\infty} x^k y^l p(x, y)\, dx\, dy.$$

Начальные моменты первого порядка

$$\alpha_{1,0} = \mathsf{M}X = a, \quad \alpha_{0,1} = \mathsf{M}Y = b$$

определяют точку (a, b) — центр распределения двумерной величины $(X;\ Y)$.

Центральные моменты порядка $k+l$ определяются по формуле

$$\mu_{k,l} = \mathsf{M}\,(X-a)^k\,(Y-b)^l. \tag{10}$$

Оба центральных момента первого порядка равны нулю:

$$\mu_{1,0} = \mathsf{M}\,(X-a) = 0, \quad \mu_{0,1} = \mathsf{M}\,(Y-b) = 0.$$

Из трех центральных моментов второго порядка два момента представляют собой дисперсии частных распределений величин X и Y:

$$\mu_{2,0} = \mathsf{M}\,(X-a)^2 = \mathsf{D}X,$$
$$\mu_{0,2} = \mathsf{M}\,(Y-b)^2 = \mathsf{D}Y$$

и лишь один момент $\mu_{1,1}$ связан с совместным распределением величин X и Y:

$$\mu_{1,1} = \mathsf{M}\,(X-a)\,(Y-b). \tag{11}$$

В частности, для непрерывной величины с плотностью совместного распределения $p(x, y)$:

$$\mu_{1,1} = \iint\limits_{-\infty}^{\infty} (x-a)(y-b)\,p(x, y)\,dx\,dy. \tag{12}$$

Момент $\mu_{1,1}$ получил название *смешанного момента второго порядка* или *корреляционного момента*.

Корреляционный момент называют также *ковариацией* величин X и Y и обозначают так: $\mu_{1,1} = \operatorname{cov}(X, Y)$.

Корреляционный момент можно выразить через начальные моменты так:

$$\mu_{1,1} = \mathsf{M}(XY) - a\mathsf{M}Y - b\mathsf{M}X + ab = \\ = \mathsf{M}(XY) - ab = \alpha_{1,1} - \alpha_{1,0}\cdot\alpha_{0,1}.$$

Значение корреляционного момента как характеристики зависимости между случайными величинами X и Y мы рассмотрим подробно в главе 7. Здесь же мы напомним только, что корреляционный момент входит в общую формулу сложения дисперсий (4.4-14):

$$\mathsf{D}(X+Y) = \mathsf{D}X + \mathsf{D}Y + 2\mathsf{M}(X-a)(Y-b) = \\ = \mu_{2,0} + \mu_{0,2} + 2\mu_{1,1}. \quad (13)$$

Случайные величины, для которых корреляционный момент равен нулю, называются некоррелированными. Как видно из написанной выше формулы (13), теорема сложения дисперсий распространяется и на некоррелированные величины. Заметим, что при доказательстве теоремы сложения дисперсий мы уже установили, что *из независимости случайных величин следует равенство нулю их корреляционного момента*, т. е. *их некоррелированность* (см. стр. 124). Обратное утверждение, вообще говоря, неверно: из некоррелированности случайных величин еще не следует их независимость. Например, если распределение величины X симметрично относительно точки $x = 0$, так что $\mathsf{M}X = 0$ и $\mathsf{M}X^3 = 0$, а величина $Y = X^2$, то корреляционный момент

$$\mu_{1,1} = \mathsf{M}(X-0)(Y-\mathsf{M}Y) = \\ = \mathsf{M}(X\cdot X^2) - \mathsf{M}X\cdot\mathsf{M}X^2 = 0$$

и случайные величины X и Y некоррелированы, хотя они связаны функциональной зависимостью.

Для характеристики зависимости между случайными величинами X и Y вводят еще безразмерный коэффициент — *коэффициент корреляции*

$$\rho = \rho(X, Y) = \frac{\mu_{1,1}}{\sigma_x\sigma_y} = \frac{\mathsf{M}(X-a)(Y-b)}{\sqrt{\mathsf{M}(X-a)^2}\sqrt{\mathsf{M}(Y-b)^2}} = \\ = \frac{\mathsf{M}(XY) - ab}{\sqrt{\mathsf{M}X^2 - a^2}\sqrt{\mathsf{M}Y^2 - b^2}}. \quad (14)$$

В главе 7 мы подробно рассмотрим свойства коэффициента корреляции. Моменты выше второго порядка для многомерных величин в настоящем пособии рассматриваться не будут.

УПРАЖНЕНИЯ

1. Посетитель тира покупает 30 пуль и решает стрелять по десяти фигурам, расходуя на каждую не более трех пуль. Предполагая, что вероятность попадания при каждом выстреле равна 0,5 (и не зависит от результатов предыдущих выстрелов), найти математическое ожидание числа израсходованных пуль, т. е. среднее число пуль, израсходованных на десять фигур.

2. В урне лежит $N = 100$ шаров, из них $M = 25$ белых. Из урны последовательно вынимаются 2 шара, X_1 — число белых шаров, появляющихся при первом вынимании, X_2 — при втором. Найти математические ожидания случайных величин X_1 и X_2, а также их произведения X_1X_2. Будут ли величины X_1 и X_2 независимы?

3. Найти центр распределения и среднее квадратическое отклонение для числа X очков, появляющихся на верхней грани правильной игральной кости при одном бросании.

4. То же для суммы чисел очков при двух и трех бросаниях.

5. В урне лежит 20 шаров, из них 4 белых. Из урны вынимаются наугад 5 шаров. Найти математическое ожидание числа X вынутых белых шаров.

6. Два лица условились встретиться в определенном месте в промежутке времени между 0 и 1 часами. Считая, что моменты их прихода независимы и распределены равномерно в интервале (0, 1), найти среднее время ожидания.

7*. В круге радиуса 1 выбирается наугад точка A и через нее в произвольном направлении проводится хорда BC. Найти среднюю длину хорды BC.

8. Найти центр и дисперсию распределения χ^2 (§ 3.8, пример 5, стр. 102).

9. Найти математическое ожидание функции $Y = e^X$, если случайная величина X имеет нормальное распределение с параметрами a и σ.

10. Найти коэффициенты асимметрии и эксцесса для распределения χ^2 (стр. 102).

11*. Вывести теорему сложения центральных моментов третьего порядка для независимых случайных величин (см. стр. 141).

12*. Говорят, что величина Y имеет логарифмически нормальное распределение вероятностей, если $Y = e^X$, где величина X имеет нормальное распределение вероятностей. Найти моду и медиану логарифмически нормального распределения, дать графическую иллюстрацию.

13*. Доказать, что если непрерывная случайная величина X принимает только неотрицательные значения и имеет математическое ожидание, то его можно представить в виде

$$\mathsf{M}X = \int_0^\infty dx \int_x^\infty p(x_1)\,dx_1 = \int_0^\infty [1 - F(x)]\,dx,$$

где $F(x)$ — функция распределения вероятностей, $p(x) = F'(x)$.

14*. Доказать, что если дискретная случайная величина X принимает только целые неотрицательные значения 0, 1, 2, ..., то ее математическое ожидание можно представить в виде

$$\mathsf{M}X = \sum_{i=1}^\infty \sum_{j=i}^\infty p_j, \qquad \text{где } p_j = \mathsf{P}(X = j).$$

Получить отсюда формулу (4.5-5).

ГЛАВА 5

ЗАКОН БОЛЬШИХ ЧИСЕЛ

§ 5.1. Неравенство Чебышева. Предел по вероятности

Случайный характер величины проявляется в том, что нельзя предвидеть, какое именно из своих значений она примет в итоге испытания. Но совокупное действие многих случайных причин может приводить к результату, почти не зависящему от случая. Так, при рассмотрении суммы большого числа случайных величин и их средних арифметических мы обнаруживаем, что частичное погашение отклонений при сложении вызывает уменьшение рассеяния средней арифметической и дает возможность предсказать ее поведение при неограниченном увеличении числа слагаемых. Закономерности такого рода и условия их возникновения составляют содержание ряда важных теорем, получивших общее название закона больших чисел. В исследовании этих вопросов значительную роль сыграли работы выдающегося русского математика академика П. Л. Чебышева (1821—1894) и его учеников.

В частности, для независимых случайных величин общий закон больших чисел дается теоремой Чебышева, доказанной им в 1867 г.

В настоящем параграфе мы рассмотрим одно неравенство, доказанное Чебышевым.

Неравенство Чебышева

Неравенство Чебышева оценивает вероятность того, что отклонение случайной величины X от ее математического ожидания $a = \mathsf{M}X$ превзойдет заданное положительное число ε:

$$\boxed{\mathsf{P}(|X-a| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}X}{\varepsilon^2}.} \qquad (1)$$

Оказывается, что эта вероятность тем меньше, чем меньше дисперсия $\mathsf{D}X$, что еще раз указывает на роль дисперсии в качестве характеристики рассеяния (разумеется, здесь предполагается, что и математическое ожидание и дисперсия случайной величины X конечны).

Мы приведем доказательство неравенства Чебышева для непрерывных случайных величин (для дискретных величин доказательство проводится аналогичным образом). Согласно определению закона распределения непрерывной случайной величины X, плотность вероятностей которой есть $p(x)$, интересующая нас вероятность равна

$$\mathsf{P}(|X-a|\geqslant\varepsilon)=\int\limits_{|x-a|\geqslant\varepsilon}p(x)\,dx,$$

где интеграл в правой части распространяется на интервалы $(-\infty, a-\varepsilon]$ и $[a+\varepsilon, \infty)$, в которых $|x-a|\geqslant\varepsilon$ (рис. 33). Так как в этих интервалах

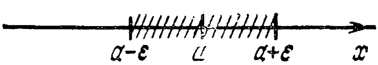


Рис. 33.

имеет место неравенство $1\leqslant\frac{(x-a)^2}{\varepsilon^2}$, а значит, и неравенство

$$p(x)\leqslant\frac{1}{\varepsilon^2}(x-a)^2p(x),$$

то

$$\int\limits_{|x-a|\geqslant\varepsilon}p(x)\,dx\leqslant\int\limits_{|x-a|\geqslant\varepsilon}\frac{1}{\varepsilon^2}(x-a)^2p(x)\,dx.$$

Теперь для завершения доказательства неравенства Чебышева (1) достаточно заметить, что в силу положительности подынтегральной функции имеет место неравенство

$$\int\limits_{|x-a|\geqslant\varepsilon}\frac{1}{\varepsilon^2}(x-a)^2p(x)\,dx\leqslant\frac{1}{\varepsilon^2}\int\limits_{-\infty}^{\infty}(x-a)^2p(x)\,dx=$$

$$=\frac{1}{\varepsilon^2}\mathsf{D}X.$$

Заменяя в неравенстве Чебышева ε на $t\sigma$, где $\sigma = \sqrt{\mathsf{D}X}$, мы можем записать его в виде

$$\mathsf{P}(|X - a| \geqslant t\sigma) \leqslant \frac{1}{t^2}. \qquad (2)$$

Заметим, что это неравенство, именно в силу своей универсальности, дает довольно грубую оценку вероятности случайного события $|X - a| \geqslant t\sigma$ и поэтому редко применяется для практических оценок. Например, при $t = 3$ неравенство (2) оценивает вероятность более чем трехсигмовых отклонений случайной величины от ее математического ожидания числом $1/9 = 0{,}111$, в то время как для нормально распределенной случайной величины эта вероятность равна (см. стр. 81)

$$1 - \Phi(3) < 0{,}003.$$

Предел по вероятности

Неравенство Чебышева оказывается особенно полезным при рассмотрении последовательности случайных величин, дисперсии которых стремятся к нулю. Пусть дана последовательность случайных величин

$$Z_1, Z_2, \ldots, Z_n, \ldots$$

с математическими ожиданиями $\mathsf{M}Z_n$ и дисперсиями $\mathsf{D}Z_n$, причем

$$\lim_{n\to\infty} \mathsf{D}Z_n = 0.$$

Применяя к величине Z_n неравенство Чебышева (1), получим

$$\mathsf{P}(|Z_n - \mathsf{M}Z_n| \geqslant \varepsilon) \leqslant \frac{\mathsf{D}Z_n}{\varepsilon^2}. \qquad (3)$$

Правая часть этого неравенства стремится к 0 при $n \to \infty$, а левая часть неотрицательна (как вероятность!). Поэтому из неравенства (3) следует, что вероятность $\mathsf{P}(|Z_n - \mathsf{M}Z_n| \geqslant \varepsilon)$ стремится к 0 при $n \to \infty$:

$$\lim_{n\to\infty} \mathsf{P}(|Z_n - \mathsf{M}Z_n| \geqslant \varepsilon) = 0, \qquad (4)$$

каково бы ни было заданное положительное число ε. Так как события $|Z_n - \mathsf{M}Z_n| \geqslant \varepsilon$ и $|Z_n - \mathsf{M}Z_n| < \varepsilon$ противоположны, то полученное соотношение (4) равносильно следующему:

$$\lim_{n\to\infty} \mathsf{P}(|Z_n - \mathsf{M}Z_n| < \varepsilon) = 1. \tag{5}$$

Эта ситуация описывается понятием сходимости по вероятности.

Определение. *Если для любого положительного числа ε вероятность неравенства*

$$|Z_n - a| < \varepsilon \tag{6}$$

стремится к 1 при $n \to \infty$, то говорят, что последовательность случайных величин Z_n сходится по вероятности к числу a (или имеет число a своим *пределом по вероятности*). Это записывают так:

$$\boxed{Z_n \xrightarrow[n\to\infty]{\text{вер}} a.} \tag{7}$$

Введенное понятие позволяет коротко сформулировать полученный выше результат (5) в виде следующей леммы.

Если для последовательности случайных величин Z_n дисперсии стремятся к 0 при $n \to \infty$, то последовательность центрированных величин $\overset{\circ}{Z}_n = Z_n - \mathsf{M}Z_n$ сходится по вероятности к нулю:

$$\overset{\circ}{Z}_n = Z_n - \mathsf{M}Z_n \xrightarrow[n\to\infty]{\text{вер}} 0.$$

§ 5.2. Теорема Я. Бернулли и устойчивость относительных частот

Рассмотрим последовательность испытаний по схеме Бернулли, т. е. последовательность независимых испытаний, в каждом из которых вероятность интересующего нас случайного события A («успеха») одна и та же и равна p. Обозначим относительную частоту успеха в серии из n испытаний через W_n. Для этой дискретной случайной величины в § 2.3 уже был найден закон распределения вероятностей (биномиальный!), а в § 4.5 были вычислены математическое ожидание и дисперсия. Напомним, что математическое ожидание относительной частоты успеха оказалось равным его вероятности p, а ее дисперсия оказалась равной $\frac{pq}{n}$ $(q = 1 - p)$:

$$\mathsf{M}W_n = p, \quad \mathsf{D}W_n = \frac{pq}{n}. \tag{1}$$

Пусть теперь число испытаний неограниченно увеличивается, $n \to \infty$. Тогда дисперсия относительной частоты, очевидно, стремится к 0. По лемме предыдущего параграфа отсюда непосредственно следует, что

$$\boxed{W_n \xrightarrow[n\to\infty]{\text{вер}} p.} \tag{2}$$

Этот результат составляет содержание замечательной теоремы Я. Бернулли, опубликованной в 1713 г. и явившейся первой теоремой из области закона больших чисел.

Т е о р е м а Я. Б е р н у л л и. *В последовательности испытаний по схеме Бернулли относительная частота случайного события сходится по вероятности к вероятности этого события в единичном испытании.*

Для статистического толкования теоремы Я. Бернулли используем принцип практической уверенности, сформулированный в § 1.8. Назначим границу α пренебрежимо малых вероятностей и в соответствии с определением понятия сходимости по вероятности выберем такое большое n, чтобы вероятность события $|W_n - p| < \varepsilon$ отличалась от 1 меньше, чем на α. Тогда можно быть практически уверенным в том, что указанное неравенство будет выполняться, т. е. что с точностью до заданного ε будет иметь место приближенное равенство

$$W_n \approx p.$$

Если фактически произвести серию из достаточно большого числа n испытаний по схеме Бернулли, то можно ожидать, что полученное в этой серии эмпирическое значение относительной частоты $W_n = n_A/n$ будет приближенно равно p с точностью до заданного малого числа $\varepsilon > 0$. Это обстоятельство и приводит к статистической устойчивости относительных частот в различных сериях испытаний.

Полученный результат снова напоминает нам о том, что выводы теории вероятностей можно применять только в таких ситуациях, когда относительные частоты всех рассматриваемых событий обладают статистической устойчивостью. Другими словами, не следует думать, что теорема Я. Бернулли автоматически обеспечивает устойчивость относительных частот в конкретных задачах практики,— эта устойчивость подлежит эмпирической проверке.

Заметим еще, что приближенное равенство $n_A/n \approx \approx \mathsf{P}(A)$ используется для приближенного вычисления вероятности случайного события A по опытным данным. При фиксированном значении числа n испытаний, из которых получена относительная частота n_A/n, указанное приближенное равенство нуждается в оценке точности. О том, как это можно сделать, мы расскажем в следующей главе.

§ 5.3. Теорема Чебышева

Рассмотрим последовательность попарно независимых случайных величин

$$X_1,\ X_2,\ \ldots,\ X_n,\ \ldots \tag{1}$$

Пусть все эти случайные величины имеют математические ожидания $\mathsf{M}X_n$ и дисперсии $\mathsf{D}X_n$ ($n = = 1, 2, 3, \ldots$), и пусть $\overline{X}_n$ — среднее арифметическое из первых n величин $X_1, X_2, \ldots, X_n$:

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{2}$$

Математическим ожиданием величины $\overline{X}_n$ служит среднее арифметическое из математических ожиданий величин $X_1, X_2, \ldots, X_n$:

$$\mathsf{M}\overline{X}_n = \frac{\mathsf{M}X_1 + \mathsf{M}X_2 + \cdots + \mathsf{M}X_n}{n}.$$

Дисперсия величины $\overline{X}_n$, в силу попарной независимости величин $X_1, X_2, \ldots, X_n$, равна

$$\mathsf{D}\overline{X}_n = \frac{\mathsf{D}X_1 + \mathsf{D}X_2 + \cdots + \mathsf{D}X_n}{n^2}.$$

Если все дисперсии $\mathsf{D}X_n$ ограничены одним и тем же числом

$$\mathsf{D}X_n \leqslant H \qquad (n = 1, 2, \ldots), \tag{3}$$

то

$$\mathsf{D}\overline{X}_n \leqslant \frac{H}{n}, \tag{4}$$

откуда видно, что дисперсия $\mathsf{D}\overline{X}_n$ стремится к 0 при $n \to \infty$ (рассеяние средних арифметических как бы затухает с увеличением числа n). По лемме § 5.1 отсюда непосредственно следует, что

$$\boxed{\overline{X}_n - \mathsf{M}\overline{X}_n \xrightarrow[n\to\infty]{\text{вер}} 0.} \tag{5}$$

Этот результат составляет содержание теоремы Чебышева.

Теорема Чебышева. *Если для последовательности попарно независимых случайных величин $\{X_n\}$ все дисперсии равномерно ограничены, то последовательность центрированных средних арифметических $\{\overline{X}_n - \mathsf{M}\overline{X}_n\}$ сходится по вероятности к нулю.*

Остановимся на одном частном случае теоремы Чебышева и на описании возможности ее усиления.

Прежде всего заметим, что если все случайные величины X_n имеют одно и то же математическое ожидание a:

$$\mathsf{M}X_n = a \qquad (n = 1, 2, \ldots), \tag{6}$$

то математическое ожидание среднего арифметического $\overline{X}_n$ также совпадает с a:

$$\mathsf{M}\overline{X}_n = \frac{\mathsf{M}X_1 + \mathsf{M}X_2 + \cdots + \mathsf{M}X_n}{n} = a;$$

поэтому теорема Чебышева принимает более простой вид

$$\overline{X}_n - a \xrightarrow[n\to\infty]{\text{вер}} 0,$$

что можно записать также в виде

$$\boxed{\overline{X}_n \xrightarrow[n\to\infty]{\text{вер}} a.} \tag{7}$$

Таким образом, в рассматриваемом случае *последовательность средних арифметических сходится по вероятности к общему центру распределения всех случайных величин* X_n (конечно, здесь по-прежнему предполагается попарная независимость этих величин и равномерная ограниченность их дисперсий).

Теорема Чебышева в виде (7) справедлива, в частности, для попарно независимых величин с одинаковым законом распределения. Если при этом предположить не только попарную независимость случайных величин $X_1, X_2, \ldots, X_n, \ldots$, но и их взаимную независимость, то теорема Чебышева может быть усилена в различных отношениях. Так, можно доказать, что *теорема Чебышева в виде* (7) *для взаимно независимых и одинаково распределенных случайных величин остается справедливой при одном только предположении о конечности математического ожидания* $\mathsf{M}X_n = a$, без всяких ограничений на дисперсию (которая может и не существовать),— эта теорема доказана советским математиком А. Я. Хинчиным.

В заключение настоящего параграфа отметим, что рассмотренная в § 5.2 теорема Я. Бернулли тоже является частным случаем теоремы Чебышева. Действительно, если в последовательности испытаний по схеме Бернулли с каждым n-м испытанием связать индикатор X_n успеха в этом испытании (как в § 4.5, стр. 130), то все индикаторы будут независимы и будут иметь одно и то же распределение вероятностей

X_n	1	0
	p	q

, $q = 1 - p$,

с центром $\mathsf{M}X_n = p$ и дисперсией $\mathsf{D}X_n = pq$ ($n = 1, 2, \ldots$). Относительная частота успеха за n испытаний будет как раз средним арифметическим из первых n индикаторов:

$$W_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Поэтому к величине W_n применима теорема Чебышева в виде (7), что приводит к теореме Бернулли:

$$W_n \xrightarrow[n\to\infty]{\text{вер}} p.$$

Приведенное здесь доказательство теоремы Я. Бернулли с помощью теоремы Чебышева позволяет глубже понять механизм сходимости (по вероятности) относительной частоты успеха к его вероятности.

§ 5.4. Устойчивость выборочных средних и метод моментов

Рассмотрим статистические толкования закона больших чисел для последовательности независимых одинаково распределенных случайных величин. Пусть исследуется одна случайная величина X. Для ее исследования производятся независимые испытания, каждое из которых дает одно значение этой величины. Будем предполагать, что с каждым испытанием связана своя случайная величина X_j, имеющая такое же распределение вероятностей, как и величина X; при этом независимость испытаний означает взаимную независимость случайных величин X_j $(j = 1, 2, \ldots)$, так что величины $X_1, X_2, \ldots, X_n$ образуют случайную выборку объема n из генеральной совокупности, законом распределения которой служит закон распределения величины X (см. стр. 127).

Каждый результат испытания может дать любое из возможных значений случайной величины X, и мы не можем предсказать, какое именно; но в отношении выборочного среднего $\overline{X}_n$ мы с большой уверенностью можем предсказать, что его значение будет близко к математическому ожиданию $\mathsf{M}X$. Заметим, что значение выборочного среднего $\overline{X}_n$ есть среднее арифметическое из результатов испытаний; если результат j-го испытания обозначить через x_j, то результаты n испытаний дают среднее арифметическое

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n},$$

и из теоремы Чебышева следует, что

$$\bar{x} \approx \mathsf{M}X, \tag{1}$$

т. е. что *приближенным значением математического ожидания случайной величины является среднее арифметическое ее эмпирических значений.*

В практических приложениях теории вероятностей последнее утверждение описывают еще следующим образом. Рассматривая случайные величины $X_1, X_2, \ldots, X_n$ как случайную выборку из генеральной совокупности, истолковывают их общее математическое ожидание $\mathsf{M}X$ как генеральное среднее (см. стр. 127); вытекающую из закона больших чисел практическую уверенность в близости $\bar{X}_n$ к $\mathsf{M}X$ при достаточно большом n формулируют как практическую уверенность в том, что *при достаточно большом объеме выборки выборочное среднее будет сколь угодно мало отличаться от генерального среднего.* Отсюда следует, что в двух случайных выборках достаточно больших объемов из одной и той же генеральной совокупности выборочные средние должны быть приближенно равны между собой (*статистическая устойчивость выборочных средних*).

Метод моментов. Пусть нас интересует некоторый параметр θ распределения вероятностей случайной величины X, и пусть для получения значений величины X произведено n независимых испытаний, так что их результаты $X_1, X_2, \ldots, X_n$ образуют случайную выборку. *Оценкой параметра* θ называется функция $f(X_1, X_2, \ldots, X_n)$ от случайной выборки, применяемая для приближенного представления параметра θ. Функция $f(X_1, X_2, \ldots, X_n)$ сама является случайной величиной. Если при неограниченном увеличении числа испытаний (при $n \to \infty$) последовательность функций $f(X_1, X_2, \ldots, X_n)$ сходится по вероятности к тому параметру θ, для которого построена оценка, то говорят, что функция $f(X_1, X_2, \ldots, X_n)$ есть *состоятельная оценка параметра* θ. Так, из закона больших чисел следует, что относительная частота случайного события есть состоятельная оценка его вероятности, что выборочное среднее арифметическое

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

есть состоятельная оценка математического ожидания $\mathsf{M}X$ (разумеется, если это математическое ожидание существует). Ценность состоятельной оценки для практики состоит в том, что, увеличивая число опытов, мы получаем возможность более точно оценить интересующий нас параметр.

Так как моменты распределения определяются через математические ожидания, то закон больших чисел позволяет дать состоятельные оценки начальных моментов

$$\alpha_k = \mathsf{M}X^k \tag{2}$$

в виде *выборочных начальных моментов* того же порядка

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{3}$$

(разумеется, если математическое ожидание $\mathsf{M}X^k$ существует). Действительно, из того, что случайные величины $X_1, X_2, \ldots, X_n$ взаимно независимы и одинаково распределены, следует, что величины $X_1^k, X_2^k, \ldots$ $\ldots, X_n^k$ тоже взаимно независимы и одинаково распределены; поэтому по закону больших чисел последовательность $\{\overline{X^k}\}$ будет сходиться по вероятности к начальному моменту α_k:

$$\overline{X^k} \xrightarrow[n\to\infty]{\text{вер}} \alpha_k. \tag{4}$$

Аналогично, состоятельными оценками центральных моментов

$$\mu_k = \mathsf{M}\,(X-a)^k,\ a = \mathsf{M}X,$$

служат *выборочные центральные моменты*

$$\overline{(X-a)^k} = \frac{1}{n}\sum_{i=1}^{n}(X_i-a)^k. \tag{5}$$

В частности, состоятельной оценкой дисперсии

$$\mathsf{D}X = \mathsf{M}\,(X-a)^2$$

служит выборочный центральный момент второго порядка

$$\overline{(X-a)^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2. \tag{6}$$

Полученная оценка дисперсии применима только в тех случаях, когда точно известно математическое ожидание $\mathsf{M}X = a$. В большинстве практических задач значение a заранее неизвестно и оценивается тоже по результатам опытов. Если в формуле (6) заменить математическое ожидание a его оценкой $\overline{X}$, то получающаяся при этом оценка

$$\widetilde{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{7}$$

тоже будет состоятельной оценкой дисперсии $\mathsf{D}X$, как мы покажем ниже. Но в одном отношении она будет существенно отличаться от приведенных выше оценок. Для всех предыдущих оценок математическое ожидание соответствующей функции $f(X_1, X_2, \ldots, X_n)$ от возможных результатов испытаний при любом фиксированном значении n совпадает с оцениваемым параметром:

$$\mathsf{M}\overline{X} = a, \quad \mathsf{M}\overline{X^k} = \alpha_k, \quad \mathsf{M}\,\overline{(X-a)^k} = \mu_k.$$

На языке математической статистики это формулируют так: выборочные моменты

$$\overline{X},\ \overline{X^k},\ \overline{(X-a)^k}$$

являются *несмещенными оценками* соответствующих теоретических моментов

$$a = \alpha_1,\ \alpha_k,\ \mu_k.$$

Оценка же (7) этим свойством не обладает. Естественно попытаться найти и состоятельную и несмещенную оценку дисперсии. С этой целью установим одно важное тождество, связывающее оценки (6) и (7).

Преобразуем сумму квадратов

$$\sum_{i=1}^{n} (X_i - a)^2 = \sum_{i=1}^{n} [(X_i - \overline{X}) + (\overline{X} - a)]^2 =$$
$$= \sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - a)^2 + 2(\overline{X} - a) \sum_{i=1}^{n} (X_i - \overline{X}).$$

Так как сумма отклонений от среднего равна нулю:

$$\sum_{i=1}^{n} (X_i - \overline{X}) = \sum_{i=1}^{n} X_i - n\overline{X} = 0,$$

то мы получаем

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - a)^2 - n(\overline{X} - a)^2 \tag{8}$$

или

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2 - (\overline{X} - a)^2. \tag{9}$$

Отсюда следует, во-первых, состоятельность оценки $\widetilde{S}^2$:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \xrightarrow[n\to\infty]{\text{вер}} \mathsf{D}X, \tag{10}$$

так как по доказанному выше

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2 \xrightarrow[n\to\infty]{\text{вер}} \mathsf{D}X \quad \text{и} \quad (\overline{X} - a)^2 \xrightarrow[n\to\infty]{\text{вер}} 0.$$

Во-вторых, формула (9) позволяет просто подсчитать математическое ожидание оценки $\widetilde{S}^2$ и убедиться в том, что она смещенная:

$$\mathsf{M} \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n} \mathsf{M} \sum_{i=1}^{n} (X_i - a)^2 - \mathsf{M}(\overline{X} - a)^2 =$$
$$= \frac{1}{n} \cdot n\mathsf{D}X - \mathsf{D}\overline{X} = \mathsf{D}X - \frac{1}{n}\mathsf{D}X = \frac{n-1}{n}\mathsf{D}X.$$

Для получения несмещенной оценки достаточно умножить оценку $\widetilde{S}^2$ на $n/(n-1)$. Так мы получаем несмещенную оценку дисперсии $\mathsf{D}X$:

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,} \tag{11}$$

которая снова является состоятельной оценкой дисперсии, так как множитель $n/(n-1)$ стремится к 1 при $n \to \infty$. Оценка (11) называется **выборочной** или **эмпирической дисперсией**; ее значение, вычисленное по фактическим результатам опытов:

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,}$$

обычно применяется в качестве приближенного значения дисперсии $\mathsf{D}X = \sigma^2(X)$.

Отметим, что для практических расчетов последнюю формулу обычно преобразуют к виду

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2\right];$$

при этом, разумеется, значения x_i можно отсчитывать от любого выбранного начала отсчета — это не влияет на значение s^2.

Так обстоит дело с оценками моментов распределения. В тех случаях, когда в качестве параметров распределения выступают его моменты, указанный выше метод позволяет непосредственно получать оценки этих параметров. В общем случае **оценки параметров по методу моментов** получают следующим образом. Пусть закон распределения величины X содержит подлежащие оценке параметры $\theta_1, \theta_2, \ldots$, например, пусть плотность распределения величины X есть $p(x; \theta_1, \theta_2)$. Тогда моменты распределения величины X, такие как

$$\alpha_1 = \mathsf{M}X = \int_{-\infty}^{\infty} xp(x;\ \theta_1,\ \theta_2)\,dx,$$

$$\mu_2 = \mathsf{D}X = \int_{-\infty}^{\infty} (x-\alpha_1)^2\,p(x;\ \theta_1,\ \theta_2)\,dx,$$

представляют собой некоторые функции от параметров θ_1, θ_2.

Рассматривая написанные выше соотношения как уравнения относительно θ_1 и θ_2, разрешаем их относительно θ_1 и θ_2, а затем в полученных выражениях заменяем теоретические моменты α_1, μ_2 соответствующими выборочными моментами — выборочным средним $\overline{X}$ и выборочной дисперсией S^2. Это и дает оценки θ_1 и θ_2 по методу моментов.

Пример 1. Распределение Пуассона (2.4-8) имеет один параметр a. Как было показано в § 4.4, этот параметр является математическим ожиданием случайной величины, следующей распределению Пуассона. Поэтому оценкой параметра a распределения Пуассона по методу моментов служит выборочное среднее арифметическое $\overline{X}$.

В примере, рассмотренном в § 2.5, где распределение числа смертей в год от удара копытом представлено таблицей на стр. 59, среднее арифметическое значение равно

$$\bar{x} = \frac{0\cdot 109 + 1\cdot 65 + 2\cdot 22 + 3\cdot 3 + 4\cdot 1}{200} = \frac{122}{200} = 0{,}61.$$

Это среднее арифметическое значение и принято за оценку параметра a в указанном примере и по этой оценке произведен расчет вероятностей распределения Пуассона в таблице на стр. 59.

Пример 2. Для случайной величины X, следующей общему нормальному распределению (3.3-1), в § 4.5 было показано, что $\mathsf{M}X = a$, $\mathsf{D}X = \sigma^2$. Поэтому, если в качестве параметров нормального закона рассматривать a и σ^2, то оценками этих параметров по методу моментов будут $\overline{X}$ и $\hat{S}^2$. Обычно предпочитают несмещенную оценку и вместо $\tilde{S}^2$ применяют выборочную дисперсию S^2.

Если в качестве параметра рассматривается не σ^2, а стандарт σ, то оценка его по методу моментов получается из соотношения $\sigma = \sqrt{\mathsf{D}X}$, что приводит к так называемому *эмпирическому стандарту* $S = \sqrt{S^2}$. Заметим, что хотя эта оценка и остается состоятельной, но она уже будет смещенной: можно показать, что

$$\mathsf{M}S = \sigma \cdot T_k, \text{ где } T_k = \sqrt{\frac{2}{k}}\,\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad (k = n-1)$$

(см. упр. 5 к гл. 6). Поэтому в качестве несмещенной оценки параметра σ можно взять величину S/T_k.

Пример 3. Показательное распределение (3.2-5) имеет один параметр λ. Как было показано в § 4.5, для случайной величины T, следующей показательному закону, $\mathsf{M}T = 1/\lambda$. Решая последнее уравнение относительно λ, находим $\lambda = 1/\mathsf{M}T$. Отсюда получаем оценку параметра λ по методу моментов в виде

$$\tilde{\lambda} = \frac{1}{\bar{T}} = \frac{n}{T_1 + T_2 + \cdots + T_n}. \tag{12}$$

Например, если в простейшем потоке случайных событий промежутки времени между последовательными случайными событиями составили 6, 8, 8, 7, 9, 7, 7, 6, 8, 8 минут, то, применяя показательное распределение времени T, мы можем в качестве оценки параметра λ взять

$$\tilde{\lambda} = \frac{10}{6\cdot 2 + 7\cdot 3 + 8\cdot 4 + 9\cdot 1} = \frac{10}{74} = 0{,}135.$$

Можно показать, что оценка $\bar{\lambda}$ смещенная:

$$\mathsf{M}\tilde{\lambda} = \frac{n}{n-1}\lambda.$$

Пример 4. Гамма-распределение (§ 3.2) имеет два параметра λ и α, которые связаны с моментами распределения формулами (4.6-10). Поэтому оценки этих параметров по методу моментов имеют вид

$$\tilde{\lambda} = \frac{\bar{X}}{S^2}, \qquad \tilde{\alpha} = \frac{(\bar{X})^2}{S^2}, \tag{13}$$

где $\bar{X}$ — среднее арифметическое результатов испытаний, S^2 — выборочная дисперсия. Эти оценки состоятельные, но смещенные.

Понятие о методе наибольшего (максимального) правдоподобия. На практике, где число испытаний обычно невелико, желательно пользоваться несмещенными оценками, имеющими возможно меньшее рассеяние. Оценки параметров по методу моментов этими свойствами обладают далеко не всегда. Поэтому последние годы широко применяют метод наибольшего правдоподобия, дающий в некотором смысле наилучшие оценки параметров.

Пусть непрерывная случайная величина X следует закону распределения с плотностью $p(x; \theta)$, где параметр θ подлежит оценке по результатам эксперимента. Как и выше, будем считать, что результаты эксперимента $X_1, X_2, \ldots, X_n$ образуют случайную выборку, т. е. что они независимы и следуют тому же закону распределения, что и X. Тогда плотность их совместного распределения есть

$$p(x_1, x_2, \ldots, x_n) = p(x_1; \theta)p(x_2; \theta) \ldots p(x_n; \theta). \qquad (14)$$

Оценкой параметра θ *по методу наибольшего правдоподобия* (*оценкой м.н.п.*) называется такое значение $\hat{\theta}$, которое обращает в максимум выражение (14) при $x_1 = X_1, x_2 = X_2, \ldots, x_n = X_n$. Обычно вместо плотности (14) рассматривают ее логарифм:

$$l = l(X_1, X_2, \ldots, X_n) = \ln p = \sum_{i=1}^{n} \ln p(X_i; \theta); \qquad (15)$$

приравнивая нулю производную от величины l по θ и решая полученное уравнение (*уравнение правдоподобия*), находят оценку $\hat{\theta}$.

Аналогично поступают при оценке нескольких параметров $\theta_1, \theta_2, \ldots$, заменяя в предыдущих формулах параметр θ вектором $\theta = \{\theta_1, \theta_2, \ldots\}$ и решая *систему уравнений правдоподобия*

$$\frac{\partial l}{\partial \theta_1} = 0; \qquad \frac{\partial l}{\partial \theta_2} = 0; \ldots$$

Для получения оценок м.н.п. параметров дискретного распределения в предыдущие формулы вместо плотности надо подставить саму вероятность.

Примеры. 1) Для показательного распределения с плотностью (3.2-5) имеем

$$l = \sum_{i=1}^{n} \ln(\lambda e^{-\lambda T_i}) = n \ln \lambda - \lambda \sum_{i=1}^{n} T_i.$$

Уравнение правдоподобия $\dfrac{dl}{d\lambda} = \dfrac{n}{\lambda} - \sum\limits_{i=1}^{n} T_i = 0$ дает

ту же оценку $\hat{\lambda} = n \Big/ \sum\limits_{i=1}^{n} T_i$, что и метод моментов.

2) Для нормального распределения (3.3-1) имеем

$$l=\sum_{i=1}^{n}\ln\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(X_i-a)^2}{2\sigma^2}\right)\right]=$$

$$=-n\ln\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-a)^2+\text{const}.$$

Система уравнений правдоподобия

$$\begin{cases}\dfrac{\partial l}{\partial a}=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(X_i-a)=0,\\[2ex] \dfrac{\partial l}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(X_i-a)^2=0\end{cases}$$

приводит к тем же оценкам $\hat{a}=\overline{X}$, $\hat{\sigma}^2=\widetilde{S}^2$, что и метод моментов.

3) Для гамма-распределения (3.2-6) имеем

$$l=n\alpha\ln\lambda-n\ln\Gamma(\alpha)+(\alpha-1)\sum_{i=1}^{n}\ln X_i-\lambda\sum_{i=1}^{n}X_i.$$

Система уравнений правдоподобия

$$\begin{cases}\dfrac{\partial l}{\partial\lambda}=\dfrac{n\alpha}{\lambda}-\sum\limits_{i=1}^{n}X_i=0,\\[2ex] \dfrac{\partial l}{\partial\alpha}=n\ln\lambda-n\,[\ln\Gamma(\alpha)]'+\sum\limits_{i=1}^{n}\ln X_i=0\end{cases}$$

приводит уже к другим оценкам, чем метод моментов, но эту систему можно решить только приближенно.

4) Для распределения Пуассона (2.4-8) вероятности запишем в виде $p(X)=\dfrac{a^X}{X!}e^{-a}$ $(X=0,1,2,\ldots)$. Тогда

$$l=\sum_{i=1}^{n}\ln p(X_i)=\sum_{i=1}^{n}X_i\ln a-na+\text{const}$$

и уравнение правдоподобия $\dfrac{dl}{da}=\dfrac{1}{a}\sum\limits_{i=1}^{n}X_i-n=0$ приводит к той же оценке $\hat{a}=\overline{X}$, что и метод моментов.

УПРАЖНЕНИЯ

1*. Распределением Коши называется непрерывное распределение с плотностью

$$p(x)=\frac{1}{\pi}\frac{1}{1+x^2} \qquad (-\infty<x<\infty).$$

Найти распределение среднего арифметического $\bar{X}$ из взаимно независимых случайных величин $X_1, X_2, \ldots, X_n$, подчиняющихся закону Коши, и доказать, что последовательность таких средних арифметических не сходится по вероятности ни к какому числу, т. е. что в этом случае закон больших чисел не применим.

2. Измерения 100 обработанных деталей дали следующие отклонения от номинального размера:

−2	2	1	2	−1	−2	3	1	−1	0
0	−1	3	1	2	−3	1	0	1	1
0	1	−1	1	0	2	2	1	0	−1
1	1	4	−1	1	1	−1	0	2	−2
2	0	−2	0	0	−1	1	4	−2	1
−3	0	0	1	4	0	−2	2	1	2
−1	1	0	−1	0	3	1	−2	3	−1
1	2	2	0	−2	1	0	−1	0	3
3	−2	−1	−2	1	0	0	−3	1	0
2	1	0	3	−1	2	1	0	−1	0

Найти оценки параметров соответствующего нормального распределения по методу моментов.

3. Найти оценки по методу моментов параметров гамма-распределения по следующим данным:

Значения x	0	1	2	3	4	5	6	
Частота m	1	33	41	18	5	1	1	100

4. Найти оценку м.н.п. параметра p биномиального распределения.

5. Найти оценку м.н.п. параметра p геометрического распределения.

6*. Найти оценку м.н.п. параметра θ равномерного распределения в интервале $(0, \theta)$.

ГЛАВА 6

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ И ОЦЕНКИ СРЕДНИХ

§ 6.1. Асимптотически нормальные распределения Понятие о центральной предельной теореме

В центральной предельной теореме рассматривается вопрос о законе распределения суммы взаимно независимых случайных величин при неограниченном увеличении числа слагаемых. Оказывается, что при некоторых весьма общих условиях закон распределения суммы приближается к нормальному закону; установление таких условий и составляет содержание центральной теоремы. Для некоторых частных условий центральная предельная теорема была доказана еще в XVIII веке Муавром и Лапласом. В общей форме задача была поставлена в исследованиях П. Л. Чебышева, но полученные им условия были довольно ограничительными. При весьма общих условиях центральная предельная теорема была доказана в 1900 г. учеником Чебышева академиком А. М. Ляпуновым.

Для удобства формулировки центральной предельной теоремы введем понятие об асимптотически нормальных распределениях. Пусть дана бесконечная последовательность случайных величин

$$Z_1,\ Z_2,\ \ldots,\ Z_n,\ \ldots$$

Говорят, что случайные величины Z_n имеют асимптотически нормальные распределения с параметрами γ_n и δ_n, если закон распределения вероятностей случайной величины $\frac{Z_n - \gamma_n}{\delta_n}$ при $n \to \infty$ стремится к стандартному нормальному закону распределения; это означает, что для любых чисел t_1 и $t_2 > t_1$ имеет место предельное соотношение

$$\lim_{n\to\infty} \mathsf{P}\left(t_1 < \frac{Z_n - \gamma_n}{\delta_n} < t_2\right) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \qquad (1)$$

в частности,

$$\lim_{n\to\infty} \mathsf{P}\left(\left|\frac{Z_n - \gamma_n}{\delta_n}\right| < t\right) = \Phi(t), \qquad (2)$$

где $\Phi(t)$ — интеграл вероятностей (3.3-3).

Если, сверх того, параметры γ_n и δ_n представляют собой математическое ожидание и среднее квадратическое отклонение случайной величины Z_n, т. е. если отношение $\frac{Z_n - \gamma_n}{\delta_n}$ есть нормированная случайная величина (с нулевым математическим ожиданием и единичной дисперсией), то говорят, что величина Z_n имеет *асимптотически нормальное распределение с центром* γ_n *и средним квадратическим отклонением* δ_n.

Как указывалось выше, в центральной предельной теореме рассматриваются суммы

$$Z_n = X_1 + X_2 + \ldots + X_n \qquad (n = 1, 2, \ldots) \qquad (3)$$

взаимно независимых случайных величин X_1, X_2, ..., X_n, ... и устанавливаются условия, при которых сумма Z_n имеет асимптотически нормальное распределение. Всюду в дальнейшем предполагается, что для каждой величины X_n существует математическое ожидание $\mathsf{M}X_n = a_n$ и дисперсия $\mathsf{D}X_n = \sigma_n^2$ $(n = 1, 2, \ldots)$, так что сумма Z_n имеет математическое ожидание $\mathsf{M}Z_n = a_1 + a_2 + \ldots + a_n$ и дисперсию $\mathsf{D}Z_n = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 = D_n$.

Формулировка центральной предельной теоремы в форме Ляпунова. *Если взаимно независимые случайные величины* X_1, X_2, ..., X_n, ... *имеют конечные абсолютные центральные моменты третьего порядка* $\mathsf{M}\,|X_i - a_i|^3$ $(i = 1, 2, \ldots)$ *и если эти моменты удовлетворяют* *условию*

$$\frac{\sum_{i=1}^{n} \mathsf{M}\,|X_i - a_i|^3}{D_n^{3/2}} \to 0 \quad \text{при} \quad n \to \infty, \tag{4}$$

то сумма Z_n *имеет асимптотически нормальное распределение с центром*

$$\gamma_n = \mathsf{M}Z_n$$

и средним квадратическим отклонением

$$\delta_n = \sqrt{D_n}.$$

Частные случаи. Если все случайные величины $X_1, X_2, \ldots, X_n, \ldots$ имеют одинаковые моменты первых трех порядков, т. е. если

$$\mathsf{M}X_n = a, \quad \mathsf{D}X_n = \sigma^2, \quad \mathsf{M}\,|X_n - a|^3 = \lambda$$
$$(n = 1, 2, \ldots),$$

то условие (4) выполняется автоматически, так как

$$D_n = n\sigma^2$$

и

$$\frac{\sum_{i=1}^{n} \mathsf{M}\,|X_i - a|^3}{D_n^{3/2}} = \frac{n\lambda}{n^{3/2}\sigma^3} = \frac{1}{\sqrt{n}}\,\frac{\lambda}{\sigma^3} \to 0 \quad \text{при} \quad n \to \infty.$$

В этом случае центральная предельная теорема утверждает, что сумма (3) имеет асимптотически нормальное распределение с центром na и средним квадратическим отклонением $\sigma\sqrt{n}$. При этом среднее арифметическое

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{Z_n}{n}$$

будет иметь асимптотически нормальное распределение с центром a и средним квадратическим отклонением $\sigma/\sqrt{n}$.

Заметим еще, что в случае одинаково распределенных слагаемых X_n для справедливости центральной предельной теоремы достаточно предположить конечность дисперсий $\mathsf{D}X_n = \sigma^2$, а конечность момента третьего порядка $\mathsf{M}\,|\,X_n - a\,|^{3}$ не требуется.

Мы не будем здесь доказывать центральную предельную теорему в форме Ляпунова в общем виде. Для случая одинаково распределенных слагаемых доказательство приведено в следующем параграфе. Общее доказательство может быть проведено тем же методом, но с более сложными оценками. В настоящее время известны необходимые и достаточные условия, при которых сумма взаимно независимых случайных величин имеет асимптотически нормальное распределение. Смысл этих условий грубо можно описать так: вклад каждой случайной величины в рассеяние суммы должен быть мал.

Центральная предельная теорема позволила объяснить пригодность модели нормального закона распределения во многих явлениях, где рассеяние изучаемых величин вызывается очень большим количеством случайных причин, влияние каждой из которых в отдельности ничтожно мало. В частности, эта теорема объясняет успешное применение нормального закона в теории ошибок наблюдений, при описании демографических и биологических явлений и т. п. С другой стороны, установление точных условий центральной предельной теоремы позволяет строго ограничить область применимости нормального закона распределения и тем самым избежать ошибочного применения его в тех задачах, где для этого нет оснований.

§ 6.2. Характеристические функции и доказательство центральной предельной теоремы для одинаково распределенных случайных величин

Для доказательства центральной предельной теоремы и для решения ряда других задач теории вероятностей весьма удобным оказался метод характеристических функций, разработанный А. М. Ляпуновым.

Характеристической функцией $f_X(u)$ случайной величины X называется математическое ожидание величины e^{iuX}:

$$\boxed{f_X(u) = \mathsf{M}e^{iuX},} \tag{1}$$

где u — действительный параметр, i — мнимая единица.

Для непрерывной случайной величины характеристическая функция совпадает с преобразованием Фурье от плотности распределения вероятностей $p(x)$:

$$f_X(u) = \mathsf{M}e^{iuX} = \int_{-\infty}^{\infty} e^{iux} p(x)\,dx. \tag{2}$$

Отсюда сразу следует оценка характеристической функции

$$|f_X(u)| \leqslant \int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{3}$$

В настоящем параграфе нам понадобятся следующие свойства характеристических функций.

1. *Характеристическая функция однозначно определяет распределение вероятностей случайной величины.* В частности, если случайная величина X имеет всюду непрерывную плотность распределения $p(x)$, то эта плотность может быть выражена через характеристическую функцию $f_X(u)$ с помощью обратного преобразования Фурье:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} f_X(u)\,du. \tag{4}$$

Это — известный факт из теории преобразований Фурье.

2. *Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых.*

Это свойство непосредственно следует из теоремы умножения математических ожиданий: независимость случайных величин X и Y влечет за собой независимость величин e^{iuX} и e^{iuY} и поэтому

$$\mathsf{M}e^{iu(X+Y)} = \mathsf{M}(e^{iuX} \cdot e^{iuY}) = \mathsf{M}e^{iuX} \cdot \mathsf{M}e^{iuY},$$

т. е.

$$f_{X+Y}(u) = f_X(u) \cdot f_Y(u). \tag{5}$$

Таким образом, при изучении сумм независимых случайных величин проще оперировать с характеристическими функциями (которые при этом перемножаются), чем с плотностями (которые свертываются). Конечно, переход к характеристическим функциям и обратно требует умения оперировать с преобразованиями Фурье, но эта часть работы существенно упрощается тем, что существуют весьма подробные таблицы преобразований Фурье.

3. При умножении случайной величины X на число C характеристическая функция преобразуется следующим образом:

$$f_{CX}(u) = \mathsf{M}e^{iu(CX)} = \mathsf{M}e^{i(Cu)X} = f_X(Cu). \tag{6}$$

4. Если характеристическая функция $f_X(u)$ *непрерывной* случайной величины X является пределом последовательности характеристических функций $f_{X_n}(u)$ каких угодно случайных величин X_n ($n = 1, 2, \ldots$), то функция распределения $F(x) = \mathsf{P}(X < x)$ является пределом последовательности функций распределений $F_n(x) = \mathsf{P}(X_n < x)$.

Это свойство является частным случаем известной теоремы о предельном переходе в преобразованиях Фурье; оно позволяет при изучении предельных распределений пользоваться методом характеристических функций, осуществляя предельный переход сначала в последовательности характеристических функций.

Пример. Для случайной величины X_0 со стандартным нормальным распределением вероятностей характеристической функцией является

$$f_0(u) = \int_{-\infty}^{\infty} e^{iux} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \exp\left(-\frac{u^2}{2}\right). \tag{7}$$

Всякая центрированная случайная величина X с нормальным распределением вероятностей отличается от величины X_0 только множителем σ:

$$X = X_0\sigma, \qquad \text{где} \quad \sigma^2 = \mathsf{D}X.$$

Поэтому ее характеристическая функция равна

$$f_X(u) = f_0(\sigma u) = \exp\left(-\frac{\sigma^2 u^2}{2}\right).$$

Если X_1 и X_2 — две независимые случайные величины, распределенные по нормальным законам с нулевыми математическими ожиданиями и дисперсиями, равными σ_1^2 и σ_2^2, то их сумма будет иметь характеристическую функцию

$$f_{X_1+X_2}(u) = \exp\left(-\frac{\sigma_1^2 u^2}{2}\right) \exp\left(-\frac{\sigma_2^2 u^2}{2}\right) = \exp\left[-\frac{(\sigma_1^2 + \sigma_2^2)u^2}{2}\right].$$

Отсюда следует, что эта сумма будет распределена тоже нормально с нулевым математическим ожиданием и дисперсией $\sigma_1^2 + \sigma_2^2$.

Таким образом, композиция нормальных законов распределения приводит снова к нормальному закону (см. стр. 100); доказательство этого свойства с помощью свертки требует значительно более сложных выкладок.

Из доказанного свойства следует, что любая линейная комбинация независимых нормально распределенных величин будет тоже иметь нормальное распределение вероятностей.

Центральная предельная теорема для одинаково распределенных случайных величин.

Напомним (см. стр. 167), что для последовательности независимых одинаково распределенных случайных величин X_1, X_2, ..., X_n, ... центральная предельная теорема формулируется следующим образом.

Сумма $X_1 + X_2 + \ldots + X_n$ имеет асимптотически нормальное распределение с центром na и средним квадратическим отклонением $\sigma\sqrt{n}$, где

$$a = \mathsf{M}X_i, \qquad \sigma^2 = \mathsf{D}X_i \qquad (i = 1, 2, \ldots). \tag{8}$$

Это значит, что распределение вероятностей нормированной суммы

$$Z_n = \frac{(X_1 + X_2 + \ldots + X_n) - na}{\sigma\sqrt{n}} \tag{9}$$

имеет своим пределом стандартное нормальное распределение вероятностей:

$$\lim_{n\to\infty} \mathsf{P}(|Z_n| < t) = \Phi(t),$$

где $\Phi(t)$ — интеграл вероятностей.

Доказательство. Мы приведем доказательство этой теоремы только для непрерывных случайных величин с помощью метода характеристических функций. Прежде всего заметим, что рассматриваемую теорему достаточно доказать только для нормированных случайных величин

$$Y_k = \frac{X_k - a}{\sigma} \qquad (k = 1, 2, \ldots),$$

так как очевидно, что

$$Z_n = \frac{Y_1 + Y_2 + \ldots + Y_n}{\sqrt{n}}. \tag{10}$$

Пусть $p(y)$ — плотность распределения вероятностей случайной величины Y_k, а $f_Y(u)$ — ее характеристическая функция

$$f_Y(u) = \int_{-\infty}^{\infty} e^{iuy}\, p(y)\, dy$$

(мы здесь пользуемся тем, что по условию плотность распределения одна и та же для всех величин X_k, а значит, и для всех величин Y_k).

Для нормированной суммы

$$Z_n = \frac{1}{\sqrt{n}} Y_1 + \frac{1}{\sqrt{n}} Y_2 + \ldots + \frac{1}{\sqrt{n}} Y_n$$

характеристическая функция находится по свойствам 2 и 3:

$$f_{Z_n}(u) = \underbrace{f_Y\left(\frac{1}{\sqrt{n}} u\right) \ldots f_Y\left(\frac{1}{\sqrt{n}} u\right)}_{n \text{ раз}} = \left[f_Y\left(\frac{1}{\sqrt{n}} u\right)\right]^n,$$

т. е.

$$f_{Z_n}(u) = \left[\int_{-\infty}^{\infty} e^{i\frac{u}{\sqrt{n}} y} p(y)\, dy\right]^n.$$

Разлагая подынтегральную функцию по степеням $1/\sqrt{n}$ (по формуле Тейлора) и ограничиваясь членами порядка $1/n$, получаем

$$f_{Z_n}(u) = \left[\int_{-\infty}^{\infty} \left(1 + i\frac{u}{\sqrt{n}} y - \frac{u^2}{2!\,n} y^2 + \ldots\right) p(y)\, dy\right]^n =$$

$$= \left[\int_{-\infty}^{\infty} p(y)\, dy + \frac{iu}{\sqrt{n}} \int_{-\infty}^{\infty} y p(y)\, dy - \frac{u^2}{2n} \int_{-\infty}^{\infty} y^2 p(y)\, dy + \ldots\right]^n.$$

Здесь

$$\int_{-\infty}^{\infty} p(y)\, dy = 1,$$

$$\int_{-\infty}^{\infty} y p(y)\, dy = \mathsf{M} Y_k = 0,$$

$$\int_{-\infty}^{\infty} y^2 p(y)\, dy = \mathsf{M} Y_k^2 = \mathsf{D} Y_k = 1,$$

а остаточный член в квадратных скобках представляет собой бесконечно малую величину более высокого порядка, чем $1/n$ (доказательство справедливости последнего утверждения в случае несобственных интегралов требует еще специальных оценок, которые мы опускаем).

Таким образом,

$$f_{Z_n}(u) = \left[1 - \frac{u^2}{2n} + \ldots\right]^n$$

и по известным правилам математического анализа последовательность характеристических функций $f_{Z_n}(u)$ имеет предел

$$\lim_{n\to\infty} f_{Z_n}(u) = e^{-\frac{u^2}{2}}.$$

Полученный предел, как было показано выше, как раз является характеристической функцией стандартного нормального распределения вероятностей.

В силу свойств 1 и 4 характеристических функций это означает, что распределение вероятностей величины Z_n стремится при $n \to \infty$ к стандартному нормальному распределению, что и требовалось доказать.

§ 6.3. Применения центральной предельной теоремы. Распределение случайных ошибок измерений. Теорема Муавра — Лапласа

Распределение случайных ошибок измерений. Результаты измерений, как правило, дают не точное, а лишь приближенные значения измеряемой величины. И чем большее число знаков нам удается снять с показаний приборов, тем нагляднее выступают различия между результатами измерений. По этим различным результатам нам надо судить об истинном значении a измеряемой величины, надо оценить это истинное значение. Многие методы обработки результатов измерений основаны на допущении, что результат измерения есть случайная величина. В настоящем параграфе нас будет интересовать закон распределения этой случайной величины в случае прямых измерений (когда приближенные значения измеряемой величины считываются непосредственно со шкалы прибора, как при взвешивании тела на аналитических весах). При этом, как обычно принято, мы будем записывать закон распределения не для самих результатов измерений, а для ошибок измерений. Под ошибкой измерения понимают разность $x - a$ между результатом измерения x и истинным значением a измеряемой величины. Мы не будем здесь говорить о грубых ошибках или промахах, когда нарушаются условия измерения или неправильно считываются показания прибора — устранение таких ошибок есть одна из задач организации измерений, а не теории вероятностей. Мы условимся далее считать, что интересующие нас результаты измерений свободны от так называемых систематических ошибок, вызванных постоянным или изменяющимся по известному закону смещением результатов измерений (для устранения систематических ошибок достаточно ввести соответствующие поправки во все результаты измерений). Математическим условием отсутствия систематических ошибок является равенство $MX = a$, где X — возможный результат измерения. Но при любом данном уровне техники измерений всегда остаются неизбежные, неустранимые ошибки — их принято называть случайными ошибками измерений. Эти ошибки вызываются многочисленными, трудно выявляемыми причинами, каждая из которых приводит лишь к незначительному разбросу результатов измерений (например, при взвешивании тела на аналитических весах к таким причинам относятся неконтролируемые колебания температуры и влажности воздуха, попадание соринок и т. п.). Каждая из указанных причин порождает свою, так называемую элементарную ошибку измерения; реально наблюдаемая ошибка измерения является суммой элементарных ошибок, и можно ожидать, что к ней применима центральная предельная теорема.

Для весьма многих видов измерений реально наблюдаемая ошибка измерения действительно более или менее точно следует нормальному закону распределения вероятностей.

Поэтому при математической обработке результатов прямых измерений считают, что *случайная ошибка измерения следует нормальному закону распределения вероятностей с параметрами* 0 *и* σ. Параметр σ этого распределения называют здесь *средней квадратической ошибкой измерения* или *стандартной ошибкой измерения* (иногда просто *стандартом*); стандартная ошибка характеризует точность измерений.

Результат измерения, как сумма истинного значения a и случайной ошибки, тоже следует при этом нормальному закону распределения вероятностей с тем же значением параметра σ, но уже с математическим ожиданием a.

Примечание. Принятие нормального закона распределения в качестве математической модели случайных ошибок измерения может вызывать различные сомнения. Так, случайная величина с нормальным законом принимает любые действительные значения, в то время как случайная ошибка измерения заведомо ограничена (например, вес тела не может быть отрицательным). Однако это не мешает хорошему согласию теории с практикой, если учесть, что вероятность выброса значений нормально распределенной величины за определенные границы очень мала. Можно было бы заметить также, что результаты измерений отсчитываются в единицах, связанных с ценой деления шкалы измерительного прибора, и потому выступают как дискретные величины; но если цена деления достаточно мала, то и это обстоятельство не мешает согласию с нормальным законом распределения.

Распределение ошибки округления суммы. При расчетах на электронных вычислительных машинах, когда производится огромное количество вычислений, а числа в машине неизбежно округляются (из-за ограниченного числа разрядов в машине), большое значение имеет вопрос правильной оценки накопленных ошибок при вычислениях. Классическая теория приближенных вычислений дает очень завышенную оценку накопленных ошибок, что часто ставит под сомнение полученные результаты. Например, при сложении n приближенных чисел, округленных до одного и того же разряда, классическая теория дает для суммы оценку ошибки в виде $n\varepsilon$, где ε — оценка ошибки каждого слагаемого; при этом не учитывается, что при сложении чисел ошибки разных знаков частично компенсируют друг друга.

Вероятностный подход к решению задачи заключается в том, что ошибки округления слагаемых считаются независимыми случайными величинами с одинаковыми распределениями, сосредоточенными в интервале $(-\varepsilon, \varepsilon)$. При сложении чисел накопленная ошибка — ошибка округления суммы — будет суммой указанных случайных величин. Если число слагаемых велико, то к этой сумме можно применить центральную предельную теорему для одинаково распределенных случайных величин, т. е. можно считать, что ошибка суммы более или менее точно следует нормальному закону распределения вероятностей. Найдем параметры этого нормального закона при условии, что ошибка X_k округления каждого слагаемого имеет равномерное распределение в интервале $(-\varepsilon, \varepsilon)$ (модель равномерного распределения ошибок округления принимается во многих практических задачах). При указанном условии математическое ожидание ошибки X_k равно нулю, а среднее квадратическое отклонение пропорционально длине интервала (см. стр. 135);

$$\sigma(X_k) = \frac{\varepsilon}{\sqrt{3}} \qquad (k = 1, 2, \ldots, n). \tag{1}$$

Для ошибки округления суммы (накопленной ошибки)

$$X = X_1 + X_2 + \ldots + X_n$$

математическое ожидание тоже будет равно нулю, а среднее квадратическое отклонение будет равно

$$\sigma(X) = \sqrt{n}\,\sigma(X_k) = \varepsilon\sqrt{\frac{n}{3}}. \tag{2}$$

Таким образом, величина X будет распределена приближенно нормально с параметрами 0 и $\varepsilon\sqrt{\frac{n}{3}}$. Это позволяет, в частности, дать обоснованные оценки величины накопленной ошибки при указанных условиях. Если воспользоваться «правилом трех сигм», т. е. пренебречь возможностью выхода величины X за трехсигмовые пределы, то можно считать, что накопленные ошибки не выйдут за границы интервала $(-\varepsilon\sqrt{3n}, \varepsilon\sqrt{3n})$; полученное правило известно как *правило Чеботарева.* Таким образом, хотя накопленная ошибка X распределена в интервале $(-n\varepsilon, n\varepsilon)$, почти все ее значения сосредоточены в значительно меньшем интервале, длина которого растет пропорционально $\sqrt{n}$ (а не пропорционально n). Например при $n = 3000$ классическая оценка накопленной ошибки дает 3000 ε, что означает потерю четырех знаков, а вероятностная оценка накопленной ошибки дает $\sqrt{9000}\varepsilon < 100\varepsilon$, что означает потерю только двух знаков. Подобные выводы обычно хорошо согласуются с практикой приближенных вычислений.

Предельная теорема Муавра—Лапласа. Весьма частным случаем центральной предельной теоремы является теорема Муавра — Лапласа о предельном распределении относительной частоты случайного события при неограниченном увеличении числа испытаний по схеме Бернулли.

Как мы уже знаем из § 2.3, число X успехов за n испытаний по схеме Бернулли есть дискретная случайная величина с биномиальным распределением вероятностей. В то же время эту величину можно представить в виде суммы

$$X = X_1 + X_2 + \ldots + X_n,$$

где X_k — индикатор успеха в k-м испытании ($k = 1, 2, \ldots, n$) (см. стр. 130). Все индикаторы независимы и имеют одно и то же распределение вероятностей с математическим ожиданием

$$\mathsf{M}X_k = p$$

и дисперсией $\mathsf{D}X_k = pq$, следовательно, с $\sigma(X_k) = \sqrt{pq}$, где p — вероятность успеха в каждом испытании, $q = 1 - p$.

По центральной предельной теореме случайная величина X — число успехов за n испытаний по схеме Бернулли — имеет асимптотически нормальное распределение с центром np и средним квадратическим отклонением $\sqrt{npq}$; относительная частота $W = X/n$ имеет тоже асимптотически нормальное распределение с центром $np/n = p$ и средним квадратическим отклонением $\frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$. Высказанное утверждение составляет содержание теоремы Муавра—Лапласа, доказанной Муавром для частного случая $p = q = 1/2$ (в 1730 г.) и Лапласом для общего случая любого p $(0 < p < 1)$ (в 1783 г.).

Эту теорему иногда формулируют так: *для биномиального распределения вероятностей существует предельное* (при $n \to \infty$) *распределение и это предельное распределение является нормальным.*

Значение теоремы Муавра—Лапласа определяется тем, что при большом числе испытаний (n) расчет вероятностей по точной биномиальной формуле (2.3-1) становится весьма затруднительным, особенно если речь идет не о вероятности отдельного равенства $X = m$, а о более важной для приложений вероятности неравенства $m_1 < X < m_2$, в частности, неравенства $\left|\frac{X}{n} - p\right| < \varepsilon$, дающего оценку отклонений относительной частоты от вероятности. По теореме Муавра—Лапласа при достаточно большом числе испытаний (и значениях p, не слишком близких к 0 или 1) мы можем вычислять указанные вероятности по приближенной формуле

$$\mathsf{P}(m_1 < X < m_2) \approx \frac{1}{2}[\Phi(t_2) - \Phi(t_1)], \qquad (3)$$

где $t_1 = \frac{m_1 - np}{\sqrt{npq}}$, $t_2 = \frac{m_2 - np}{\sqrt{npq}}$. В частности,

$$\mathsf{P}\left(\left|\frac{X - np}{\sqrt{npq}}\right| < t\right) \approx \Phi(t). \tag{4}$$

Замена дискретного биномиального распределения непрерывным нормальным распределением (и значит, замена сумм интегралами) позволяет значительно упростить расчеты.

Пример. Найти вероятность того, что при $n = 10\,000$ испытаниях по схеме Бернулли отклонение относительной частоты события от его вероятности не превзойдет $\varepsilon = 0{,}01$, если известно, что вероятность события равна $p = 0{,}2$.

Решение. Нам нужно найти вероятность неравенства

$$\left|\frac{X}{n} - p\right| \leqslant \varepsilon. \tag{5}$$

Это неравенство равносильно неравенству

$$np - n\varepsilon \leqslant X \leqslant np + n\varepsilon.$$

В нашей задаче $n = 10\,000$, $np = 2000$, $n\varepsilon = 100$ и точное значение искомой вероятности равно сумме

$$\sum_{m=1900}^{2100} \frac{10\,000!}{m!\,(10\,000 - m)!} (0{,}2)^m (0{,}8)^{10\,000 - m},$$

содержащей 201 слагаемое, каждое из которых весьма трудно вычислить. Объем работы здесь слишком велик.

Вычислим искомую вероятность по приближенной формуле (4). Для этого преобразуем неравенство (5) следующим образом:

$$\left|\frac{X - np}{\sqrt{npq}}\right| \leqslant \varepsilon \frac{n}{\sqrt{npq}} = t.$$

При числовых данных задачи имеем

$$t = \frac{\varepsilon \sqrt{n}}{\sqrt{pq}} = \frac{0{,}01 \cdot 100}{\sqrt{0{,}2 \cdot 0{,}8}} = 2{,}5,$$

и следовательно, искомая вероятность приближенно равна

$$\mathsf{P}\left(\left|\frac{X}{n}-0{,}2\right|\leqslant 0{,}01\right)\approx\Phi(2{,}5)=0{,}988.$$

Полученный результат оказывается верным до третьего знака, что объясняется довольно большим значением $n = 10\,000$.

§ 6.4. Построение доверительного интервала для математического ожидания нормального распределения с известной дисперсией

Рассмотрим ситуацию, которая возникает при обработке результатов измерений. Пусть, например, нас интересует длина какого-нибудь стержня, и мы пытаемся оценить ее истинное значение a с помощью n замеров. Если предположить, как это было сделано в предыдущем параграфе, что ошибки измерений независимы и распределены по нормальному закону с нулевым математическим ожиданием и некоторой дисперсией σ^2, то результаты измерений $X_1, X_2, \ldots, X_n$ суть независимые случайные величины, распределенные нормально с параметрами $\mathsf{M}X_i = a$, $\mathsf{D}X_i = \sigma^2$ $(i = 1, 2, \ldots, n)$. Будем считать, что величина σ, характеризующая точность измерений, нам известна. Тогда закон распределения случайных величин X_1, $X_2, \ldots, X_n$ известен с точностью до математического ожидания. В § 5.4 было показано, что параметр a можно оценить по результатам измерений, и была предложена для него несмещенная и состоятельная оценка $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$, для которой $\mathsf{M}\overline{X} = a$ и $\mathsf{P}(|\overline{X} - a| > \varepsilon) \to 0$ при $n \to \infty$ и любом фиксированном $\varepsilon > 0$.

Если, например, в результате 10 измерений длины стержня были получены следующие числа:

$$23,\ 24,\ 23,\ 25,\ 25,\ 26,\ 26,\ 25,\ 24,\ 25,$$

то оценка $\overline{X}$ приняла значение

$$X = (23\cdot 2 + 24\cdot 2 + 25\cdot 4 + 26\cdot 2) = 24{,}6.$$

Про оценку $\overline{X}$ говорят, что она точечная, ибо дает одно численное значение (одну точку на прямой). Все оценки, рассмотренные в § 5.4, тоже являются точечными, в результате эксперимента они дают числа.

В этом параграфе мы рассмотрим другой вид оценивания неизвестного параметра распределения. Мы поставим задачу указать некоторый интервал со случайными границами, который накрывает истинное значение a с заданной вероятностью $\mathcal{P} = 1 - \alpha$, где положительное число α выбирают достаточно малым, исходя из соображений, высказанных в § 1.8. Такой интервал называется доверительным, вероятность $\mathcal{P}$ — *доверительной вероятностью* или *надежностью*, α — *уровнем значимости*.

Итак, у нас есть независимые случайные величины $X_1, X_2, \ldots, X_n$, распределенные по нормальному закону с параметрами a и σ, причем параметр a нам неизвестен, и надо построить доверительный интервал, который накрывал бы истинное значение a с вероятностью $\mathcal{P}$. Учитывая симметрию нормального распределения относительно центра a, будем строить симметричный доверительный интервал, т. е. будем искать неравенство вида

$$|\overline{X} - a| < \varepsilon, \tag{1}$$

которое выполнялось бы с вероятностью $\mathcal{P}$. Так как при указанных выше условиях среднее арифметическое $\overline{X}$ имеет нормальное распределение с параметрами $\mathsf{M}\overline{X} = a$, $\sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$, то по формуле (3.3-8) имеем

$$\mathsf{P}\left(\left|\frac{\overline{X} - a}{\sigma/\sqrt{n}}\right| < t\right) = \mathsf{P}\left(|\overline{X} - a| < t\frac{\sigma}{\sqrt{n}}\right) = \Phi(t), \tag{2}$$

где $\Phi(t)$ — интеграл вероятностей (3.3-3). По заданной вероятности $\mathcal{P}$ мы найдем такое число $t = t(\mathcal{P})$, что

$$\Phi(t) = \mathcal{P}. \tag{3}$$

Тогда неравенство

$$\boxed{|\overline{X} - a| < t(\mathcal{P})\frac{\sigma}{\sqrt{n}},} \tag{4}$$

того же типа, что и (1), будет выполняться с вероятностью $\mathcal{P}$. Разрешив неравенство (4) относительно a, получим

$$\boxed{\bar{X} - t(\mathcal{P})\frac{\sigma}{\sqrt{n}} < a < \bar{X} + t(\mathcal{P})\frac{\sigma}{\sqrt{n}}\,\cdot} \tag{5}$$

Неравенство (5) дает решение поставленной задачи: интервал $\left(\bar{X} - t(\mathcal{P})\frac{\sigma}{\sqrt{n}},\ \bar{X} + t(\mathcal{P})\frac{\sigma}{\sqrt{n}}\right)$ накрывает истинное значение a с вероятностью $\mathcal{P}$, т. е. этот интервал есть доверительный интервал для математического ожидания нормального распределения с известной дисперсией σ^2 (известным стандартным отклонением σ).

В качестве примера построим доверительный интервал с вероятностью $\mathcal{P} = 0{,}95$ для параметра a по десяти измерениям длины стержня, приведенным в начале параграфа, при условии, что $\sigma = 1$. По таблице 2 Приложения для $\mathcal{P} = \Phi(t) = 0{,}95$ находим $t = 1{,}960$; следовательно, доверительный интервал для параметра a есть

$$\left(2{,}46 - \frac{1{,}960}{\sqrt{10}};\quad 24{,}6 + \frac{1{,}960}{\sqrt{10}}\right)$$

или (23,8; 25,4). Полученный результат нужно трактовать так: с вероятностью $\mathcal{P} = 0{,}95$ истинное значение a накрывается интервалом (23,8; 25,4).

В этом параграфе мы рассмотрели задачу построения доверительного интервала для математического ожидания нормального распределения при известном стандартном отклонении σ. Случай построения доверительного интервала для параметра a, когда стандартное отклонение неизвестно, существенно отличается от только что рассмотренного, и мы рассмотрим его в следующем параграфе.

§ 6.5. Доверительные оценки параметров нормального распределения

Нормальное распределение с плотностью

$$\varphi_{a;\,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] \tag{1}$$

вполне определяется двумя параметрами a и σ, которые представляют собой центр распределения $a = \mathsf{M}X$ и среднее квадратическое отклонение $\sigma = \sqrt{\mathsf{D}X}$. Несмещенными и состоятельными оценками a и σ^2 служат соответственно выборочное среднее и выборочная дисперсия, т. е. случайные величины

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2}$$

и

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2, \tag{3}$$

построенные по независимым случайным величинам $X_1, X_2, \ldots, X_n$, плотность вероятностей которых есть $\varphi_{a;\sigma}(x)$. Можно доказать, что случайные величины (2) и (3) независимы, и найти распределения их вероятностей. Величина $\overline{X}$, как мы уже знаем, имеет нормальное распределение с параметрами a и $\sigma/\sqrt{n}$. Величина

$$\frac{S^2(n-1)}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2}$$

может быть представлена в виде суммы квадратов $n-1$ независимых случайных величин со стандартным нормальным распределением, т. е. в виде

$$\frac{S^2(n-1)}{\sigma^2} = Y_1^2 + Y_2^2 + \ldots + Y_{n-1}^2,$$

где $Y_1, Y_2, \ldots, Y_{n-1}$ независимы и распределены нормально с параметрами 0 и 1. Это очевидно при $n = 2$:

$$\sum_{i=1}^{2} \frac{(X_i - \overline{X})^2}{\sigma^2} = \frac{\left(X_1 - \frac{X_1 + X_2}{2}\right)^2}{\sigma^2} + \frac{\left(X_2 - \frac{X_1 + X_2}{2}\right)^2}{\sigma^2} = \left(\frac{X_1 - X_2}{\sigma\sqrt{2}}\right)^2 = Y_1^2,$$

так как разность $X_1 - X_2$ имеет нормальное распределение с параметрами $a - a = 0$ и $\sqrt{\sigma^2 + \sigma^2} = \sigma\sqrt{2}$; доказательство в общем случае мы опускаем. Таким образом, величина $\frac{S^2(n-1)}{\sigma^2}$ имеет χ^2-распределение с $k = n - 1$ степенями свободы (стр. 102).

Из последнего утверждения можно сразу же получить *доверительную оценку неизвестного параметра* σ следующим образом. Зададим надежность $\mathscr{P}$ (или уровень значимости $\alpha = 1 - \mathscr{P}$) и по таблицам χ^2-распределения с числом степеней свободы $k = n - 1$ найдем такие числа u_1 и u_2, что

$$\int_0^{u_1} p_{\chi_k^2}(u)\,du = \int_{u_2}^{\infty} p_{\chi_k^2}(u)\,du = \frac{\alpha}{2}$$

(на рис. 29, стр. 103, указаны соответствующие квантили при $\mathscr{P} = 0{,}90$). Тогда вероятность неравенства

$$u_1 < \frac{S^2(n-1)}{\sigma^2} < u_2 \tag{4}$$

будет равна

$$\int_{u_1}^{u_2} p_{\chi_k^2}(u)\,du = 1 - \alpha = \mathscr{P}.$$

Разрешив неравенство (4) относительно σ, преобразуем его к виду

$$\boxed{S \cdot z_{\text{н}} < \sigma < S \cdot z_{\text{в}},} \tag{5}$$

где $z_{\text{в}} = \sqrt{\frac{n-1}{u_1}}$, $z_{\text{н}} = \sqrt{\frac{n-1}{u_2}}$ — верхняя и нижняя границы отношения σ/S. Это и дает *доверительный интервал для параметра* σ *с надежностью* $\mathscr{P}$. В таблице 4 Приложения приведены коэффициенты $z_{\text{н}}$ и $z_{\text{в}}$ при надежностях $\mathscr{P} = 0{,}95$ и $0{,}99$ и при различных значениях чисел степеней свободы k.

Доверительную оценку параметра a записывают обычно в форме

$$|\overline{X} - a| < t\frac{S}{\sqrt{n}}, \tag{6}$$

которая отличается от формулы (6.4-4) заменой σ на S и, естественно, изменением коэффициента t. Для расчета вероятности неравенства (6) надо найти распределение вероятностей отношения

$$T = \frac{|\bar{X} - a|}{S/\sqrt{n}}. \tag{7}$$

Преобразуем это отношение. Как указано выше на стр. 182, величина S^2 может быть представлена в виде $\frac{\sigma^2}{n-1}(Y_1^2 + Y_2^2 + \ldots + Y_{n-1}^2)$, где все величины $Y_1, Y_2, \ldots, Y_{n-1}$ независимы и имеют стандартное нормальное распределение, а величина $\frac{\bar{X} - a}{\sigma/\sqrt{n}} = Y_0$ независима от них и тоже имеет стандартное нормальное распределение. Поэтому отношение (7) можно представить в виде

$$T = \frac{Y_0}{\sqrt{Y_1^2 + Y_2^2 + \ldots + Y_{n-1}^2}} \cdot \sqrt{n-1},$$

откуда видно, что распределение отношения (7) не зависит ни от a, ни от σ. Плотность этого распределения — *распределения Стьюдента* — может быть подсчитана по правилам, указанным в § 3.8. Приведем без вывода эту плотность:

$$p_k(t) = C_k\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \tag{8}$$

где $k = n - 1$ — число степеней свободы, C_k — постоянная, определяемая из условий нормировки.

Интересно заметить, что при $k \to \infty$ плотность распределения Стьюдента стремится к плотности стандартного нормального распределения, так как

$$\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}} \to e^{-\frac{t^2}{2}} \quad \text{при} \quad k \to \infty.$$

Даже при сравнительно малых значениях k кривая распределения Стьюдента на вид мало отличается от кривой нормального распределения, как видно из рис. 34, на котором дан график плотности распределения Стьюдента при $k = 5$. Но «хвосты» этих распределений отличаются значительно, и это сказывается на соответствующих таблицах.

Зная плотность распределения Стьюдента, можно подсчитать вероятность

$$\mathsf{P}\left(\left|\frac{\bar{X}-a}{S/\sqrt{n}}\right|<t\right)=\mathsf{P}(|T|<t)=2\int_0^t p_k(t)\,dt=\mathscr{P}$$

и составить таблицу значений $t = t(\mathscr{P};\ k)$ в зависимости от заданной доверительной вероятности $\mathscr{P}$ и от чис-

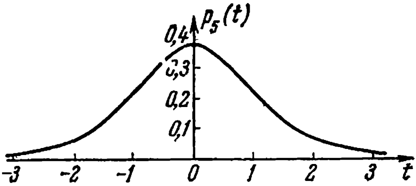


Рис. 34.

ла степеней свободы k. Такая таблица приведена в Приложении (табл. 3).

Итак, *доверительный интервал для неизвестного параметра* a *при неизвестном* σ *с надежностью* $\mathscr{P}$ есть

$$\boxed{\bar{X}-t(\mathscr{P};n-1)\frac{S}{\sqrt{n}}<a<\bar{X}+t(\mathscr{P};n-1)\frac{S}{\sqrt{n}}\,.} \tag{9}$$

Отличие оценки (9) от оценки (6.4-4) особенно заметно при малом числе n. Например, при $n = 5$ и $\mathscr{P} = 0{,}99$ таблица 3 дает $t(0{,}99;\ 4) = 4{,}604$, в то время как в случае известного значения σ мы нашли бы по таблице 2: $t(0{,}99) = 2{,}576$.

Приведенные выше доверительные оценки параметров a и σ нормального распределения применяются при обработке результатов измерений следующим образом. Если можно считать, что результаты измерений распределены нормально с математическим ожиданием, равным истинному значению a измеряемой величины, то оценка этого истинного значения дается формулой

(9), в которой через $X_1, X_2, \ldots, X_n$ обозначены результаты n независимых измерений. При этих же условиях формула (5) дает оценку точности измерений — оценку средней квадратической ошибки σ.

Пример. При определении величины заряда электрона $e_0 \cdot 10^{-10}$ (в единицах CGSE) Милликен получил $n=58$ результатов измерений X_i величины e_0, приведенных в следующей таблице:

4,781	4,764	4,777	4,809	4,761	4,769
4,795	4,776	4,765	4,790	4,792	4,806
4,769	4,771	4,785	4,779	4,758	4,779
4,792	4,789	4,805	4,788	4,764	4,785
4,779	4,772	4,768	4,772	4,810	4,790
4,775	4,789	4,801	4,791	4,799	4,777
4,772	4,764	4,785	4,788	4,799	4,749
4,791	4,774	4,783	4,783	4,797	4,781
4,782	4,778	4,808	4,740	4,790	
4,767	4,791	4,771	4,775	4,747	

Оценить истинное значение величины e_0 и стандартную ошибку измерения σ с надежностью $\mathcal{P} = 0{,}99$, предполагая, что измерения независимы и произведены без систематической ошибки, а случайные ошибки распределены по нормальному закону с параметрами 0 и σ.

Решение. По данным таблицы подсчитаем значения выборочного среднего

$$\overline{X} = \frac{1}{58}\sum_{i=1}^{58} X_i = 4{,}7808$$

и выборочной дисперсии

$$S^2 = \frac{1}{57}\sum_{i=1}^{58} (X_i - \overline{X})^2 = 234 \cdot 10^{-6}.$$

В качестве точечных оценок величин e_0 и σ примем соответственно $\overline{X} = 4{,}7808$ и $S = \sqrt{S^2} = 0{,}0153$.

Найдем теперь доверительные интервалы для величин e_0 и σ с надежностью $\mathcal{P} = 0{,}99$. Так как точное значение σ неизвестно, то доверительные границы для e_0

подсчитываем с помощью распределения Стьюдента. По таблице 3 Приложения для $\mathscr{P} = 0{,}99$ и числа степеней свободы $k = 58 - 1 = 57$ с помощью интерполяции находим

$$t = t\,(0{,}99;\ 57) = 2{,}665,$$

что позволяет найти соответствующую границу отклонения e_0 от $\overline{X}$:

$$|e_0 - \overline{X}| < 2{,}665 \cdot \frac{0{,}0153}{\sqrt{58}} = 0{,}0054.$$

Итак, доверительный интервал для e_0 с надежностью 0,99 есть $(4{,}7808 - 0{,}0054;\ 4{,}7808 + 0{,}0054)$ или $(4{,}7754;\ 4{,}7862)$. Округляя границы интервала, получаем $4{,}775 < e_0 < 4{,}786$.

Доверительный интервал для σ найдем по формуле (5). По таблице 4 Приложения при $\mathscr{P} = 0{,}99$ и числе степеней свободы $k = 57$ с помощью интерполяции находим нижнюю и верхнюю границы

$$z_{\text{н}} = 0{,}804, \quad z_{\text{в}} = 1{,}31,$$

что позволяет получить доверительный интервал для σ в виде $(0{,}804 \cdot 0{,}0153;\ 1{,}31 \cdot 0{,}0153)$ или $(0{,}0123;\ 0{,}0200)$. Округляя границы интервала, получаем окончательно $0{,}012 < \sigma < 0{,}020$ с надежностью 0,99.

О расчете выборочных средних и дисперсий. При наличии ЭВМ расчет выборочных средних и дисперсий не составляет труда. Но при небольшом числе результатов эксперимента часто приходится пользоваться ручным счетом. Чтобы облегчить расчет и уменьшить возможность ошибки в вычислениях, рекомендуется следующая схема расчета и контроля вычислений. Прежде всего, с помощью выбора подходящего начала отсчета c и масштаба h, т. е. с помощью линейного преобразования $(X - c)/h = U$ упрощают исходные данные для расчета. При этом формулы для расчета принимают вид

$$\overline{X} = c + h\overline{U}, \text{ где } \overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i,$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{h^2}{n-1}\sum_{i=1}^{n}(U_i - \overline{U})^2;$$

последнюю формулу удобнее записать в виде

$$S^2 = \frac{h^2}{n-1}\left(\sum_{i=1}^{n} U_i^2 - n(\overline{U})^2\right).$$

Далее можно упростить расчет, сгруппировав данные по интервалам и отнеся число данных, попавших в интервал, к середине интервала. При числе интервалов порядка 16 ÷ 20 ошибкой округления в рассматриваемой задаче можно пренебречь.

Проведем расчет для примера, рассмотренного выше. Все результаты измерений заключены в интервале [4,740; 4,810]. Выберем длину частичных интервалов $h = 0{,}004$, что даст 18 интервалов, скажем, от 4,7395 до 4,8115 (мы сдвигаем интервалы на 0,0005, чтобы результаты измерений не попали на концы интервалов; расчеты при этом нисколько не усложняются, так как они проводятся лишь с целыми значениями u_i.). Разнеся результаты измерений по интервалам, получаем следующую *расчетную таблицу*, где через y_i обозначены середины интервалов

Расчетная таблица

y_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$
4,7415	1	—9	—9	81
4,7455	1	—8	—8	64
4,7495	1	—7	—7	49
4,7535	0	—6	—	—
4,7575	1	—5	—5	25
4,7615	1	—4	—4	16
4,7655	5	—3	—15	45
4,7695	5	—2	—10	20
4,7735	6	—1	—6	6
4,7775	7	0	—	—
4,7815	5	1	5	5
4,7855	3	2	6	12
4,7895	10	3	30	90
4,7935	3	4	12	48
4,7975	3	5	15	75
4,8015	1	6	6	36
4,8055	2	7	14	98
4,8095	3	8	24	192
—	58	—	48	862

а через m_i — число результатов, попавших в интервал $\left(y_i - \frac{h}{2},\ y_i + \frac{h}{2}\right)$; $u_i = \frac{y_i - 4{,}7775}{0{,}004}$. В последней строке таблицы подсчитаны суммы, нужные для расчета:

$$\sum m_i = n = 58, \quad \sum u_i = 48, \quad \sum u_i^2 = 862.$$

По этим суммам подсчитаем

$$\bar{X} = 4{,}7775 + \frac{48}{58} \cdot 0{,}004 = 4{,}7808,$$

$$S^2 = 0{,}004^2 \cdot \frac{862 - 58\left(\frac{48}{58}\right)^2}{57} = 0{,}004^2 \cdot 14{,}4,$$

$$S = 0{,}004 \cdot 3{,}80 = 0{,}0152,$$

что мало отличается от значений $\bar{X}$ и S, приведенных выше.

Для контроля вычислений весь расчет повторяют с другим началом расчета, например, взяв за c не 4,7775, а 4,7815.

Замечание о доверительной оценке средней квадратической ошибки измерения.

Для получения приемлемой доверительной оценки параметра σ надо произвести очень большое число измерений. Как видно из таблицы 4 Приложения, даже при $n = 100$ относительная ошибка в определении параметра σ достигает 16% при $\mathcal{P} = 0{,}95$ и 22% при $\mathcal{P} = 0{,}99$. Увеличение числа измерений требует затраты времени и средств и не всегда целесообразно. В то же время часто встречается ситуация, когда одним и тем же измерительным прибором производят много серий измерений различных величин. Оказывается возможным по мере накопления информации о произведенных сериях измерений уточнять доверительные оценки точности прибора (если, разумеется, точность прибора за этот период времени не меняется).

Пусть одним и тем же прибором произведено m серий независимых измерений разных величин; обозначим $n_1, n_2, \ldots, n_m$ количества измерений в этих сериях, через $S_1^2, S_2^2, \ldots, S_m^2$ — соответствующие эмпирические дисперсии. Так как при композиции χ^2-распределений мы получаем снова χ^2-распределение с числом степеней свободы, равном сумме чисел степеней свободы слагаемых (см. стр. 102), то величина

$$\frac{S_1^2(n_1 - 1)}{\sigma^2} + \frac{S_2^2(n_2 - 1)}{\sigma^2} + \ldots + \frac{S_m^2(n_m - 1)}{\sigma^2}$$

будет иметь χ^2-распределение с числом степеней свободы

$$k = (n_1 - 1) + (n_2 - 1) + \ldots + (n_m - 1) = n - m,$$

где $n = n_1 + n_2 + \ldots + n_m$ — общее число измерений во всех сериях.

Отсюда следует, что *средняя эмпирическая дисперсия*

$$S^2 = \frac{S_1^2(n_1-1)+S_2^2(n_2-1)+\ldots+S_m^2(n_m-1)}{n-m} \tag{10}$$

может служить несмещенной оценкой дисперсии σ^2 и может быть использована для построения доверительной оценки параметра σ с большим числом степеней свободы. Такая доверительная оценка имеет вид

$$S\cdot z_{\text{н}}(\mathcal{P};\ n-m)<\sigma<S\cdot z_{\text{в}}(\mathcal{P};\ n-m). \tag{11}$$

В частном случае, когда в каждой серии число измерений одно и то же ($n_1 = n_2 = \ldots = n_m = l,\ n = ml$), получаем для S более простое выражение:

$$S=\sqrt{\frac{S_1^2+S_2^2+\ldots+S_m^2}{m}} \qquad (\text{при } k=m(l-1)).$$

УПРАЖНЕНИЯ

1. Найти характеристическую функцию индикатора случайного события и характеристическую функцию биномиального распределения.

2. Найти характеристическую функцию равномерного распределения в интервале $(-\alpha, \alpha)$ и характеристическую функцию суммы n независимых величин с равномерным распределением в интервале $\left(-\frac{\alpha}{n}, \frac{\alpha}{n}\right)$.

3. Доказать, что χ^2-распределение (стр. 102) является асимптотически нормальным распределением при числе степеней свободы $k \to \infty$; найти параметры этого распределения.

4. Было произведено 12 000 бросаний монеты, при этом герб выпадал 6019 раз. Насколько хорошо согласуется это с предположением о том, что вероятность выпадения герба равна 1/2?

5*. Для случайной выборки из нормального распределения величина S^2 является несмещенной оценкой параметра σ^2, а величина S — смещенной оценкой параметра σ. Зная, что величина $\frac{S^2k}{\sigma^2}$ ($k = n-1$) имеет χ^2-распределение с k степенями свободы, доказать, что

$$MS=\sigma\sqrt{\frac{2}{k}}\,\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)},$$

где Γ — гамма-функция.

6. Оценить с надежностью $\mathcal{P} = 0{,}99$ точность измерений по следующим результатам независимых измерений одним и тем же

прибором различных величин (предполагается, что ошибки измерений распределены нормально с параметрами $(0, \sigma)$).

1-я серия	2-я серия	3-я серия	4-я серия	5-я серия	6-я серия	7-я серия	8-я серия	9-я серия	10-я серия
4,16	6,04	5,71	4,94	4,50	3,84	5,56	5,33	7,89	4,23
4,19	6,05	5,71	4,92	4,48	3,84	5,54	5,32	7,88	4,21
4,15	6,06	5,70	4,93	4,51	3,83	5,56	5,31	7,89	4,22
4,17	6,03	5,71	4,95	4,49	3,84	5,56	5,34	7,87	4,20
4,18	6,05	5,69	4,94	4,52	3,82	5,55	5,32	7,90	4,22
4,17	6,05	5,71	4,93	4,50	3,84	5,57	5,33	7,89	4,21
4,17	6,05	5,71	4,94	4,50	3,85	5,56	5,32	7,91	4,20
4,16	6,06	5,69	4,94	4,50	3,84	5,55	5,35	7,88	4,21
4,17	6,05	5,71	4,93	4,49	3,84	5,56	5,33	7,89	4,22
4,17	6,04	5,73	4,95	4,50	3,86	5,56	5,34	7,90	4,21
4,17	6,05	5,71	4,96	4,49	3,83	5,57	5,33	7,89	4,21

ГЛАВА 7

УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ И РЕГРЕССИИ

§ 7.1. Условные распределения вероятностей

Выше мы рассматривали, главным образом, независимые случайные величины. Но часто встречаются и такие ситуации, когда случайные величины не являются независимыми и не связаны функционально. Примерами могут служить зависимость между осадками и урожаем или зависимость между толщиной снегового покрова зимой и объемом стока последующего половодья. При этом на практике чаще всего ограничиваются изучением изменения средних характеристик одной величины при изменении другой (скажем, большей толщине снегового покрова в среднем соответствует больший объем стока половодья).

Вероятностный подход к решению подобных задач исходит из предположения, что система рассматриваемых величин обладает определенным совместным распределением вероятностей. Для определенности рассмотрим сначала систему двух случайных величин X и Y. Фиксируя значение одной из них, мы определяем закон условного распределения другой, подобно тому, как определяли условные вероятности случайных событий.

Так, если совместное распределение вероятностей дискретно и вероятность попадания случайной величины $(X; Y)$ в точку (x_i, y_j) обозначить через $p_{ij} = \mathsf{P}(X = x_i, Y = y_j)$ $(i = 1, \ldots, n; j = 1, \ldots, m)$, то каждую вероятность p_{ij} можно представить по правилу умножения вероятностей в виде

$$p_{ij} = \mathsf{P}(X = x_i)\mathsf{P}(Y = y_j \mid X = x_i),$$

где

$$\mathsf{P}(X = x_i) = p_{i1} + p_{i2} + \ldots + p_{im}.$$

Отсюда находим условные вероятности

$$\mathsf{P}(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i1} + p_{i2} + \ldots + p_{im}} \quad (j = 1, \ldots, m), \tag{1}$$

которые и определяют *условное распределение вероятностей величины Y при условии $X = x_i$ ($i = 1, 2, \ldots, n$).*

Если совместное распределение вероятностей непрерывно с плотностью $p(x, y)$, то *плотность условного распределения вероятностей величины Y при условии $X = x$* определяется аналогичным образом по формуле

$$p_{Y|x}(y) = \frac{p(x, y)}{p_X(x)} \qquad (p_X(x) \neq 0), \tag{2}$$

где, в соответствии с формулой (3.6-8),

$$p_X(x) = \int_{-\infty}^{\infty} p(x, y)\,dy. \tag{3}$$

Аналогично вводится *плотность условного распределения величины X при условии $Y = y$*:

$$p_{X|y}(x) = \frac{p(x, y)}{p_Y(y)} = \frac{p(x, y)}{\int_{-\infty}^{\infty} p(x, y)\,dx} \qquad (p_Y(y) \neq 0).$$

Отсюда видно, что аналогом правила умножения для непрерывных величин служит формула

$$p(x, y) = p_X(x)\, p_{Y|x}(y) = p_Y(y)\, p_{X|y}(x). \tag{4}$$

Для систем более чем двух случайных величин можно ввести не только одномерные, но и многомерные условные распределения вероятностей. Например, для системы трех величин X, Y, Z с плотностью совместного распределения $p(x, y, z)$ можно ввести плотность условного распределения вероятностей величины X при условии $Y = y$, $Z = z$:

$$p_{X|y,z}(x) = \frac{p(x, y, z)}{p_{Y,Z}(y, z)} = \frac{p(x, y, z)}{\int_{-\infty}^{\infty} p(x, y, z)\,dx}$$

и плотность условного распределения вероятностей системы величин X, Y при условии $Z = z$:

$$p_{X,Y|z}(x, y) = \frac{p(x, y, z)}{p_Z(z)} = \frac{p(x, y, z)}{\int\limits_{-\infty}^{\infty}\int p(x, y, z)\,dx\,dy}.$$

§ 7.2. Условные математические ожидания. Регрессии, их основные свойства

Из числовых характеристик условного распределения вероятностей величины Y при условии $X = x$ нас будет интересовать прежде всего *центр этого условного распределения* или *условное математическое ожидание величины Y при значении $X = x$*. Ограничиваясь, для определенности, непрерывными распределениями, представим это условное математическое ожидание формулой

$$\boxed{\mathsf{M}_x Y = \int\limits_{-\infty}^{\infty} y p_{Y|x}(y)\,dy,} \tag{1}$$

где $p_{Y|x}(y)$ — плотность условного распределения вероятностей величины Y при условии $X = x$.

Условное математическое ожидание $\mathsf{M}_x Y$ есть функция от x, которая называется *регрессией величины Y на величину X* (или *функцией регрессии*); обозначим ее через

$$\boxed{f(x) = \mathsf{M}_x Y.} \tag{2}$$

Уравнение $y = f(x)$ называется *уравнением регрессии Y на X*, а график функции регрессии — *линией регрессии*. Линия регрессии Y на X показывает, как *в среднем* изменяется величина Y при изменении величины X.

Совершенно аналогично определяется регрессия величины X на величину Y:

$$g(y) = \mathsf{M}_y X = \int\limits_{-\infty}^{\infty} x p_{X|y}(x)\,dx.$$

Заметим, что функции регрессии $f(x)$ и $g(y)$, вообще говоря, не являются взаимно обратными (пример см. далее в § 7.3).

Основное свойство регрессии величины Y на величину X

Если $f(x)$ есть функция регрессии величины Y на величину X, то математическое ожидание квадрата отклонения величины Y от функции $f(X)$ меньше, чем от любой другой функции $h(X)$:

$$\boxed{\mathsf{M}\,[Y-f(X)]^2 \leqslant \mathsf{M}\,[Y-h(X)]^2.} \tag{3}$$

Доказательство основного свойства регрессии мы проведем в два этапа. Сначала докажем, что для любой функции $u(X)$

$$\mathsf{M}\,[u(X)\,Y] = \mathsf{M}\,[u(X)\,f(X)]. \tag{4}$$

Действительно, по формуле (4.2-3) для математического ожидания функции от случайных величин имеем

$$\mathsf{M}\,[u(X)\,Y] = \iint\limits_{-\infty}^{\infty} u(x)\,y p(x, y)\,dx\,dy.$$

Заменяя здесь по формуле (7.1-4)

$$p(x, y) = p_X(x)\,p_{Y|x}(y),$$

получаем

$$\mathsf{M}\,[u(X)\,Y] = \int\limits_{-\infty}^{\infty} u(x)\,p_X(x)\,dx \int\limits_{-\infty}^{\infty} y p_{Y|x}(y)\,dy,$$

что в соответствии с формулами (1) и (2) дает

$$\mathsf{M}\,[u(X)\,Y] = \int\limits_{-\infty}^{\infty} u(x)\,f(x)\,p_X(x)\,dx.$$

Последняя формула совпадает с формулой (4). Отметим, в частности, что

$$\mathsf{M}Y = \mathsf{M}f(X), \tag{5}$$

т. е. *математическое ожидание величины Y совпадает с математическим ожиданием функции $f(X)$.*

Теперь преобразуем математическое ожидание квадрата отклонения величины Y от функции $h(X)$

следующим образом:

$$\mathsf{M}\,[Y-h\,(X)]^2=\mathsf{M}\,[Y-f\,(X)+f\,(X)-h\,(X)]^2=$$
$$=\mathsf{M}\,[Y-f\,(X)]^2+\mathsf{M}\,[f\,(X)-h\,(X)]^2+$$
$$+2\mathsf{M}\,[Y-f\,(X)]\,[f\,(X)-h\,(X)].$$

Учитывая формулу (4) при $u\,(X)=f\,(X)-h\,(X)$, получим

$$\mathsf{M}\,[Y-f\,(X)]\,[f\,(X)-h\,(X)]=$$
$$=\mathsf{M}Yu\,(X)-\mathsf{M}f\,(X)\,u\,(X)=0,$$

и следовательно,

$$\mathsf{M}\,[Y-h\,(X)]^2=\mathsf{M}\,[Y-f\,(X)]^2+\mathsf{M}\,[f(X)-h(X)]^2. \tag{6}$$

Полученное соотношение и доказывает основное свойство регрессии Y на X. Аналогично формулируется и основное свойство регрессии X на Y:

$$\mathsf{M}\,[X-g\,(Y)]^2\leqslant\mathsf{M}\,[X-h_1\,(Y)]^2.$$

Отметим частный случай формулы (6). Полагая в этой формуле

$$h\,(X)=\mathsf{M}Y=\mathsf{M}f\,(X)=b,$$

получим

$$\mathsf{M}\,(Y-b)^2=\mathsf{M}\,[Y-f\,(X)]^2+\mathsf{M}\,[f\,(X)-b]^2,$$

т. е.

$$\mathsf{D}Y=\mathsf{D}f\,(X)+\mathsf{M}\,[Y-f\,(X)]^2. \tag{7}$$

Эта формула дает разложение общей дисперсии величины Y на сумму двух дисперсий: дисперсии функции $f\,(X)$ и среднего квадрата отклонения величины Y от этой функции. Из формулы (7) следует, что

$$\mathsf{D}Y\geqslant\mathsf{D}f\,(X),$$

т. е. дисперсия функции $f\,(X)$ меньше дисперсии величины Y (в то время как центры их распределений совпадают).

Отметим еще один частный случай формулы (4). Полагая в этой формуле $h\,(X)=X$, получим

$$\mathsf{M}XY=\mathsf{M}Xf(X). \tag{8}$$

Отсюда следует, что если функция регрессии $f(x)$ постоянна, т. е. если условное математическое ожидание $\mathsf{M}_x Y$ не зависит от значения x и, значит, совпадает с математическим ожиданием $\mathsf{M}Y = b$, то $f(x) = b$ и

$$\mathsf{M}XY = \mathsf{M}Xb = b\mathsf{M}X = \mathsf{M}X \cdot \mathsf{M}Y.$$

§ 7.3. Линейная корреляция

Наиболее простым случаем будет тот, когда *обе функции регрессии* $f(x) = \mathsf{M}_x Y$ *и* $g(y) = \mathsf{M}_y X$ линейны, так что обе линии регрессии будут прямыми линиями; они называются *прямыми регрессии.* В этом случае будем говорить о линейной корреляции между величинами X и Y.

Выведем уравнения прямых регрессии. Обозначим центры распределений величин X и Y соответственно через $a = \mathsf{M}X$ и $b = \mathsf{M}Y$, дисперсии — через $\sigma_x^2 = \mathsf{D}X$ и $\sigma_y^2 = \mathsf{D}Y$, корреляционный момент (см. стр. 142) — через $\mu_{1,1} = \mathsf{M}(X-a)(Y-b)$ и будем искать параметры A и B линейной регрессии Y на X в виде

$$f(x) = A(x-a) + B.$$

Подставляя функцию $f(X) = A(X-a) + B$ в формулы (7.2-5) и (7.2-8) предыдущего параграфа и учитывая, что $\mathsf{M}(X-a) = 0$, найдем последовательно

$$b = \mathsf{M}Y = \mathsf{M}f(X) = B,$$

$$\mu_{1,1} = \mathsf{M}(X-a)(Y-b) = \mathsf{M}(X-a)[f(X)-b] = \\ = A\mathsf{M}(X-a)^2 = A\sigma_x^2,$$

откуда

$$A = \frac{\mu_{1,1}}{\sigma_x^2}.$$

Таким образом, в случае линейной корреляции функция регрессии Y на X имеет вид

$$f(x) = \frac{\mu_{1,1}}{\sigma_x^2}(x-a) + b. \tag{1}$$

Аналогично, функция регрессии X на Y имеет вид

$$g(y)=\frac{\mu_{1,1}}{\sigma_y^2}(y-b)+a. \tag{2}$$

Коэффициенты регрессии $\frac{\mu_{1,1}}{\sigma_x^2}$ и $\frac{\mu_{1,1}}{\sigma_y^2}$ имеют тот же знак, что и корреляционный момент $\mu_{1,1}$. Поэтому, если величина Y в среднем возрастает при увеличении величины X (т. е. если $\mu_{1,1}>0$), то и величина X в среднем возрастает при увеличении величины Y.

Уравнения прямых регрессии можно записать в более симметричном виде, если воспользоваться введенным в § 4.7 (стр. 143) безразмерным коэффициентом корреляции

$$\rho=\rho(X,Y)=\frac{\mu_{1,1}}{\sigma_x\sigma_y}=\frac{\mathsf{M}(X-a)(Y-b)}{\sqrt{\mathsf{M}(X-a)^2}\sqrt{\mathsf{M}(Y-b)^2}}. \tag{3}$$

При этом уравнения прямых регрессии принимают вид

$$\boxed{y-b=\rho\frac{\sigma_y}{\sigma_x}(x-a)} \quad (\textit{прямая регрессии } Y \textit{ на } X), \tag{4}$$

$$\boxed{x-a=\rho\frac{\sigma_x}{\sigma_y}(y-b)} \quad (\textit{прямая регрессии } X \textit{ на } Y) \tag{5}$$

или

$$\frac{y-b}{\sigma_y}=\rho\frac{x-a}{\sigma_x}, \quad \frac{x-a}{\sigma_x}=\rho\frac{y-b}{\sigma_y}.$$

Из уравнений прямых регрессии видно, что обе эти прямые проходят через точку (a, b) — центр совместного распределения величин X и Y. Угловые коэффициенты прямых регрессии равны соответственно (обозначения см. на рис. 35):

$$\operatorname{tg}\alpha=\rho\frac{\sigma_y}{\sigma_x}, \quad \operatorname{tg}\beta=\frac{1}{\rho}\frac{\sigma_y}{\sigma_x}.$$

В следующем параграфе мы покажем, что всегда $|\rho|\leqslant 1$. Если $0<|\rho|<1$, то $|\rho|<\frac{1}{|\rho|}$, и поэтому прямая регрессии Y на X имеет меньший наклон к оси

Ox, чем прямая регрессии X на Y. Чем ближе $|\rho|$ к 1, тем меньше угол между этими прямыми; при $|\rho| = 1$ прямые регрессии сливаются.

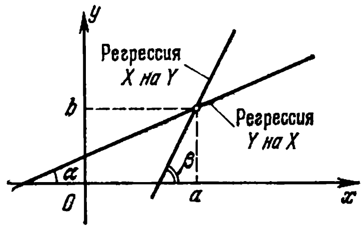


Рис. 35.

При $\rho = 0$ прямые регрессии имеют уравнения $y = b$ и $x = a$, так что обе они параллельны соответствующим осям координат; в этом случае, как было указано в § 4.7 величины X и Y называются *некоррелированными*, для них

$$\mathsf{M}_x Y = b = \mathsf{M}Y, \quad \mathsf{M}_y X = a = \mathsf{M}X,$$

т. е. условные математические ожидания совпадают с безусловными.

Нормальная корреляция

Говорят, что между величинами X и Y имеет место нормальная корреляция, если совместным распределением вероятностей величин X и Y служит нормальное распределение на плоскости (см. § 3.6, пример 2). Покажем, что *нормальная корреляция всегда линейна.*

Напомним, что плотность нормального распределения на плоскости задается формулой

$$p(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp\left\{-\frac{1}{2}\left[A(x-a)^2 + 2B(x-a)(y-b) + C(y-b)^2\right]\right\}, \quad (6)$$

где $A > 0$, $C > 0$, $AC - B^2 > 0$.

Найдем плотности распределения вероятностей величин X и Y и плотности их условных распределений. По формуле (3.6-8) имеем

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

Выделим в показателе экспоненты (6) полный квадрат и запишем плотность $p(x, y)$ в виде

$$p(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp\left\{-\frac{C}{2}\left[(y-b) + \frac{B}{C}(x-a)\right]^2\right\} \times \exp\left\{-\frac{1}{2}\left(A - \frac{B^2}{C}\right)(x-a)^2\right\}.$$

Пользуясь известной формулой

$$\int_{-\infty}^{\infty} e^{-\frac{C}{2}(y-m)^2} dy = \int_{-\infty}^{\infty} e^{-\frac{C}{2}y^2} dy = \frac{\sqrt{2\pi}}{\sqrt{C}} \qquad (m = \text{const}),$$

получим

$$p_X(x) = \frac{\sqrt{AC - B^2}}{2\pi} \cdot \frac{\sqrt{2\pi}}{\sqrt{C}} \exp\left\{-\frac{1}{2}\left(A - \frac{B^2}{C}\right)(x-a)^2\right\} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{AC - B^2}{C}} \exp\left\{-\frac{1}{2}\frac{AC - B^2}{C}(x-a)^2\right\}.$$

Это означает, что величина X имеет нормальное распределение вероятностей с параметрами

$$MX = a, \quad \sigma(X) = \sigma_x = \sqrt{\frac{C}{AC - B^2}}. \tag{7}$$

Точно так же и величина Y имеет нормальное распределение вероятностей с параметрами

$$MY = b, \quad \sigma(Y) = \sigma_y = \sqrt{\frac{A}{AC - B^2}}. \tag{8}$$

Плотность условного распределения величины Y при $X = x$ найдем по формуле (7.1-2)

$$p_{Y|x}(y) = \frac{p(x, y)}{p_X(x)} = \frac{\sqrt{C}}{\sqrt{2\pi}} \exp\left\{-\frac{C}{2}\left[y - b + \frac{B}{C}(x-a)\right]^2\right\}. \tag{9}$$

Аналогично, плотность условного распределения величины X при $Y = y$ равна

$$p_{X|y}(x) = \frac{p(x, y)}{p_Y(y)} = \frac{\sqrt{A}}{\sqrt{2\pi}} \exp\left\{-\frac{A}{2}\left[x - a + \frac{B}{A}(y - b)\right]^2\right\}. \tag{9a}$$

Таким образом, условные распределения вероятностей также являются нормальными.

Нас интересуют функции регрессии, т. е. центры условных распределений. Так как для нормального распределения вероятностей центр непосредственно входит параметром в плотность распределения, то из формул (9), (9а) получаем

$$\left.\begin{aligned} \mathsf{M}_x Y = f(x) = b - \frac{B}{C}(x - a), \\ \mathsf{M}_y X = g(y) = a - \frac{B}{A}(y - b). \end{aligned}\right\} \tag{10}$$

Таким образом, мы доказали, что нормальная корреляция линейна: обе функции регрессии (10) линейны. Из полученных формул (10) видно, что коэффициенты регрессии здесь равны соответственно

$$\rho\frac{\sigma_y}{\sigma_x} = -\frac{B}{C}, \quad \rho\frac{\sigma_x}{\sigma_y} = -\frac{B}{A}.$$

Так как из формул (7), (6) следует, что

$$\frac{\sigma_y}{\sigma_x} = \sqrt{\frac{A}{C}},$$

то коэффициент корреляции оказывается равным

$$\rho = -\frac{B}{\sqrt{AC}}; \tag{11}$$

при этом

$$1 - \rho^2 = \frac{AC - B^2}{AC} > 0.$$

Формулы (7), (8), (11) позволяют выразить параметры A, B, C нормального распределения на плоскости через σ_x, σ_y, ρ, именно

$$A = \frac{1}{\sigma_x^2(1-\rho^2)},$$

$$C = \frac{1}{\sigma_y^2(1-\rho^2)},$$

$$B = -\rho\sqrt{AC} = -\frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}.$$

Это позволяет записать плотность нормального распределения на плоскости в следующей канонической форме:

$$\left.\begin{aligned} &p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right],\\ &\text{где}\quad u = \frac{x-a}{\sigma_x},\ v = \frac{y-b}{\sigma_y}. \end{aligned}\right\} \quad (12)$$

Отметим, что *при нормальной корреляции из некоррелированности случайных величин X и Y вытекает их независимость*, т. е. здесь некоррелированность совпадает с независимостью. Действительно, для нормального распределения коэффициент корреляции $\rho = \rho(X, Y)$ равен нулю тогда и только тогда, когда $B = 0$, что равносильно независимости случайных величин X и Y (см § 3.7, пример 2).

§ 7.4. Коэффициент корреляции, его основные свойства

Коэффициент корреляции $\rho = \rho(X, Y)$ представляет собой безразмерную характеристику, как видно из формулы (7.3-3). Его можно представить также в виде математического ожидания произведения нормированных отклонений величин X и Y от их центров распределения:

$$\rho = \mathrm{M}\left(\frac{X-a}{\sigma_x}\cdot\frac{Y-b}{\sigma_y}\right) = \mathrm{M}\left[\frac{X-\mathrm{M}X}{\sigma(X)}\cdot\frac{Y-\mathrm{M}Y}{\sigma(Y)}\right]. \quad (1)$$

Отсюда непосредственно следует, что коэффициент корреляции между X и Y не изменяется ни при изменении начала отсчета, ни при изменении масштаба измерения случайных величин X и Y. Действительно, изменение масштаба и начала отсчета величины X означает линейное преобразование вида

$$X = x_0 + hX', \quad \text{где} \quad h > 0;$$

при таком линейном преобразовании мы имеем

$$\mathsf{M}X = x_0 + h\mathsf{M}X', \ \sigma(X) = h\sigma(X'),$$

и следовательно, нормированное отклонение величины X не изменяется:

$$\frac{X - \mathsf{M}X}{\sigma(X)} = \frac{(x_0 + hX') - (x_0 + h\mathsf{M}X')}{h\sigma(X')} = \frac{X' - \mathsf{M}X'}{\sigma(X')}. \tag{2}$$

Иногда оказывается удобным представить коэффициент корреляции ρ не через центральные моменты, как в формуле (1), а через начальные моменты (см. конец § 4.7)

$$\rho = \frac{\mathsf{M}(XY) - ab}{\sqrt{\mathsf{M}X^2 - a^2}\sqrt{\mathsf{M}Y^2 - b^2}} = \\ = \frac{\mathsf{M}(XY) - \mathsf{M}X \cdot \mathsf{M}Y}{\sqrt{\mathsf{M}(X^2) - (\mathsf{M}X)^2}\sqrt{\mathsf{M}(Y^2) - (\mathsf{M}Y)^2}}. \tag{3}$$

Как мы знаем из § 4.7, случайные величины, для которых коэффициент корреляции равен нулю (а значит, и корреляционный момент равен нулю), называются некоррелированными. Так как из независимости случайных величин следует их некоррелированность, то отличие коэффициента корреляции от нуля свидетельствует о наличии зависимости между величинами. Как же именно коэффициент корреляции характеризует зависимость между случайными величинами? Оказывается, что *коэффициент корреляции является мерой линейной зависимости между величинами* (говорят также, мерой прямолинейности): он показывает, насколько хорошо в среднем может быть представлена каждая из величин в виде линейной функции от другой.

Поясним это подробнее. Представим величину Y в виде суммы некоторой линейной функции от X и остатка Z:

$$Y = [A(X - a) + B] + Z. \tag{4}$$

Остаток $Z = Y - [A(X - a) + B]$ можно рассматривать как ошибку приближения величины Y линейной функцией. Докажем, что всегда существует *линейная функция наилучшего среднеквадратического приближения*, для которой $\mathsf{M}(Z^2)$ принимает наименьшее значение, и что для такой линейной функции отношение остаточной дисперсии $\sigma^2(Z)$ к дисперсии $\sigma^2(Y)$ зависит только от коэффициента корреляции $\rho(X, Y)$.

Для доказательства преобразуем математическое ожидание квадрата остатка $\mathsf{M}Z^2$ следующим образом:

$$\begin{aligned}\mathsf{M}Z^2 &= \mathsf{M}[(Y-b)-A(X-a)+(b-B)]^2 = \\ &= \mathsf{M}(Y-b)^2 + A^2\mathsf{M}(X-a)^2 + (b-B)^2 - \\ &\qquad\qquad - 2A\mathsf{M}(Y-b)(X-a),\end{aligned}$$

так как

$$\mathsf{M}(Y-b)=0,\ \mathsf{M}(X-a)=0.$$

Учитывая, что $\mathsf{M}(X-a)(Y-b) = \mu_{1,1} = \rho\sigma_x\sigma_y$, запишем полученное выражение в виде

$$\begin{aligned}\mathsf{M}Z^2 &= \sigma_y^2 + A^2\sigma_x^2 + (b-B)^2 - 2A\rho\sigma_x\sigma_y = \\ &= \sigma_y^2(1-\rho^2) + (A\sigma_x - \rho\sigma_y)^2 + (b-B)^2. \qquad (5)\end{aligned}$$

Отсюда видно, что математическое ожидание квадрата остатка достигает наименьшего значения при

$$A\sigma_x - \rho\sigma_y = 0 \quad \text{и} \quad b - B = 0,$$

т. е.

$$A = \rho\frac{\sigma_y}{\sigma_x} \quad \text{и} \quad B = b.$$

Таким образом, линейная функция наилучшего среднеквадратического приближения есть

$$\rho\frac{\sigma_y}{\sigma_x}(X-a)+b. \qquad (6)$$

Для этой функции математическое ожидание, очевидно, равно b; поэтому математическое ожидание остатка равно нулю:

$$\mathsf{M}Z = \mathsf{M}\left[Y - b - \rho\frac{\sigma_y}{\sigma_x}(X-a)\right] = 0;$$

из формулы (5) видно, что при $A = \rho \frac{\sigma_y}{\sigma_x}$ и $B = b$ дисперсия остатка (*остаточная дисперсия*) равна

$$\mathsf{D}Z = \mathsf{M}Z^2 = \sigma_y^2 (1 - \rho^2).$$

Отношение остаточной дисперсии $\mathsf{D}Z = \sigma^2(Z)$ к дисперсии $\sigma^2(Y)$ величины Y равно

$$\frac{\sigma^2(Z)}{\sigma^2(Y)} = 1 - \rho^2. \tag{7}$$

Отсюда следует, в частности, что

$$\left.\begin{aligned} 1 - \rho^2 \geqslant 0,& \\ -1 \leqslant \rho \leqslant 1 \text{ или } |\rho| \leqslant 1,& \end{aligned}\right\} \tag{8}$$

т. е. *коэффициент корреляции между случайными величинами по абсолютной величине не превосходит* 1.

Из формулы (7) видно, что $1 - \rho^2$ характеризует относительную величину остаточной дисперсии, т. е. дисперсии остатка от приближения величины Y линейной функцией. Чем ближе $|\rho|$ к 1, тем меньше относительная величина остаточной дисперсии, тем теснее линейная зависимость между X и Y, тем лучше можно приблизить Y линейной функцией от X.

Рассмотрим еще крайние случаи $\rho = \pm 1$. В этих случаях $1 - \rho^2 = 0$ и, следовательно, $\sigma^2(Z) = 0$. Но дисперсия $\sigma^2(Z)$ характеризует рассеяние случайной величины Z относительно центра ее распределения $\mathsf{M}Z = 0$. Поэтому равенство нулю остаточной дисперсии $\sigma^2(Z)$ означает, что (с вероятностью 1) остаток $Z = 0$ и, следовательно, Y есть линейная функция от X. Таким образом, *коэффициент корреляции принимает крайние значения* ± 1 *тогда и только тогда, когда между величинами* X *и* Y *имеется линейная функциональная зависимость.*

Мы подробно рассмотрели приближение величины Y линейной функцией величины X. Аналогичные результаты получаются и для приближения величины X линейной функцией величины Y. Линейная функция наилучшего среднеквадратического приближения к величине X есть

$$\rho \frac{\sigma_x}{\sigma_y}(Y - b) + a.$$

Остаток этого приближения

$$Z_1 = X - \left[\rho \frac{\sigma_x}{\sigma_y}(Y - b) + a\right]$$

имеет математическое ожидание $\mathsf{M}Z_1 = 0$ и остаточную дисперсию

$$\mathsf{D}Z_1 = \mathsf{M}Z_1^2 = \sigma_x^2\,(1 - \rho^2),$$

так что и здесь

$$\frac{\sigma^2(Z_1)}{\sigma^2(X)} = 1 - \rho^2.$$

Отметим, что параметры линейных функций наилучшего среднеквадратического приближения совпадают с параметрами прямых регрессии в случае линейной корреляции (ср. § 7.3). Это вполне соответствует основному свойству регрессии: в случае линейной корреляции прямая регрессии, скажем, Y на X, должна давать наилучшее среднеквадратическое приближение к величине Y среди всех других функций от X, в том числе и линейных.

§ 7.5. Оценки коэффициента корреляции и прямых регрессии по результатам эксперимента

Для оценки коэффициента корреляции между двумя случайными величинами X и Y и параметров прямых регрессии производят ряд независимых испытаний (экспериментов, наблюдений), исходом каждого из которых является пара величин (X_i, Y_i) $(i = 1, 2, \ldots, n)$; двумерные величины $(X_1;\ Y_1)$, $(X_2;\ Y_2)$, $\ldots$, $(X_n;\ Y_n)$ независимы и следуют тому же закону распределения, что и двумерная величина $(X;\ Y)$.

Как мы уже знаем из § 5.4, несмещенными и состоятельными оценками центров распределений a, b и дисперсий σ_x^2, σ_y^2 служат соответствующие эмпирические (выборочные) средние и дисперсии:

$$\left.\begin{aligned} &\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \\ &S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \\ &S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2. \end{aligned}\right\} \tag{1}$$

В качестве несмещенной и состоятельной оценки корреляционного момента $\mu_{1,1}$ может служить *эмпирический корреляционный момент* $m_{1,1}$:

$$m_{1,1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}). \tag{2}$$

Проверим несмещенность этой оценки.

Прежде всего, по условию имеем

$$\begin{aligned} &\mathsf{M}\,(X_i - a)\,(Y_i - b) = \mu_{1,1}, \\ &\mathsf{M}\,(X_i - a)\,(Y_j - b) = 0 \quad (i \neq j). \end{aligned} \tag{3}$$

Преобразуем теперь сумму $\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$ следующим образом:

$$\sum_{i=1}^{n}(X_i - \overline{X})\,(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - a)(Y_i - b) - n(\overline{X} - a)(\overline{Y} - b)$$

это вытекает из тождеств:

$$\begin{aligned} (X_i - \overline{X})\,(Y_i - \overline{Y}) = {} & (X_i - a)\,(Y_i - b) - (\overline{X} - a)\,(\overline{Y} - b) - \\ & - (X_i - \overline{X})\,(\overline{Y} - b) - (Y_i - \overline{Y})\,(\overline{X} - a), \end{aligned}$$

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0, \quad \sum_{i=1}^{n}(Y_i - \overline{Y}) = 0).$$

Вычислим математическое ожидание

$$\mathsf{M}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})=$$
$$=\sum_{i=1}^{n}\mathsf{M}(X_i-a)(Y_i-b)-n\mathsf{M}(\overline{X}-a)(\overline{Y}-b)=$$
$$=n\mu_{1,1}-n\mathsf{M}(\overline{X}-a)(\overline{Y}-b).$$

Так как $\overline{X}-a=\dfrac{(X_1-a)+\ldots+(X_n-a)}{n}$, то в силу свойства линейности математического ожидания и формул (3) имеем

$$\mathsf{M}(\overline{X}-a)(\overline{Y}-b)=$$
$$=\frac{1}{n^2}\mathsf{M}\left[(X_1-a)+\ldots+(X_n-a)\right]\left[(Y_1-b)+\ldots\right.$$
$$\left.\ldots+(Y_n-b)\right]=\frac{1}{n^2}\left[\mathsf{M}(X_1-a)(Y_1-b)+\ldots\right.$$
$$\left.\ldots+\mathsf{M}(X_n-a)(Y_n-b)\right]=\frac{\mu_{1,1}}{n}.$$

Таким образом,

$$\mathsf{M}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})=n\mu_{1,1}-n\frac{\mu_{1,1}}{n}=(n-1)\mu_{1,1}$$

и

$$\mathsf{M}m_{1,1}=\mu_{1,1},$$

что и означает несмещенность оценки (2).

С помощью полученных оценок S_x, S_y, $m_{1,1}$ параметров σ_x, σ_y, $\mu_{1,1}$ рассчитывают *эмпирический* (или *выборочный*) *коэффициент корреляции*

$$r=\frac{m_{1,1}}{S_xS_y}=\frac{\sum\limits_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\sum\limits_{i=1}^{n}(Y_i-\overline{Y})^2}},\qquad(4)$$

который и применяют в качестве точечной оценки коэффициента корреляции ρ. Можно показать, что эмпирический коэффициент корреляции r дает смещенную оценку теоретического коэффициента корреляции ρ; смещение оценки убывает с увеличением числа экспериментов n и стремится к нулю при $n \to \infty$.

Заменяя теперь в уравнениях прямых регрессии все теоретические моменты на значения соответствующих эмпирических моментов, мы получаем *уравнения эмпирических* (или *выборочных*) *прямых регрессии*:

$$y - \overline{Y} = r\frac{S_y}{S_x}(x - \overline{X}) \quad \text{(\textit{эмпирическая прямая регрессии } } Y \text{ \textit{на} } X\text{)}, \tag{5}$$

$$x - \overline{X} = r\frac{S_x}{S_y}(y - \overline{Y}) \quad \text{(\textit{эмпирическая прямая регрессии } } X \text{ \textit{на} } Y\text{)}. \tag{6}$$

Оказывается, что эмпирические прямые регрессии, определенные уравнениями (5) и (6), являются прямыми наилучшего среднеквадратического приближения к эмпирическим точкам $(X_i;\ Y_i)$, если понимать это утверждение в следующем смысле. Сумма квадратов отклонений эмпирических значений Y_i от эмпирической прямой регрессии (5) меньше, чем сумма квадратов отклонений их от любой другой прямой:

$$\sum_{i=1}^{n}\left[Y_i - \overline{Y} - r\frac{S_y}{S_x}(X_i - \overline{X})\right]^2 \leqslant \sum_{i=1}^{n}[Y_i - (AX_i + B)]^2.$$

Аналогично, сумма квадратов отклонений эмпирических значений X_i от эмпирической прямой регрессии (6) меньше, чем сумма квадратов отклонений их от любой другой прямой:

$$\sum_{i=1}^{n}\left[X_i - \overline{X} - r\frac{S_x}{S_y}(Y_i - \overline{Y})\right]^2 \leqslant \sum_{i=1}^{n}[X_i - (AY_i + B)]^2.$$

На рис. 36 показано, какие именно отклонения имеются в виду в обоих случаях.

Высказанные утверждения относительно эмпирических прямых регрессии можно доказать тем же путем, каким выводились формулы для линейной функции наилучшего среднеквадратического приближения в § 7.4 (с заменой операции M на операцию Σ). При этом легко получить вывод о том, что и эмпирический коэффициент корреляции r по абсолютной величине не превосходит 1, причем он достигает крайних значений ± 1 только в том случае, когда все экспериментальные точки лежат на одной прямой.

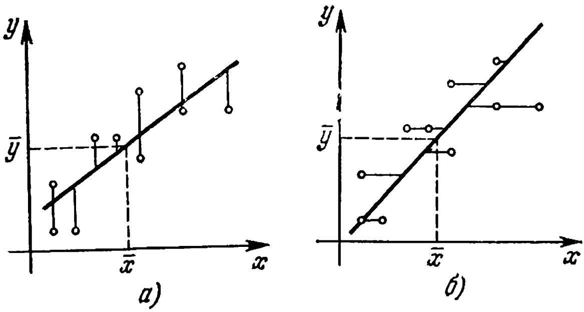


Рис. 36

Вычисления параметров эмпирических прямых регрессии часто можно упростить с помощью выбора подходящих начал отсчета и масштабов измерений значений X_i, Y_i, т. е. с помощью линейного преобразования вида

$$X = x_0 + h_1 u, \quad Y = y_0 + h_2 v,$$

или

$$u = \frac{X - x_0}{h_1}, \qquad v = \frac{Y - y_0}{h_2} \qquad (h_1 > 0, \ h_2 > 0).$$

При этом

$$X_i - \overline{X} = h_1 (u_i - \bar{u}), \quad Y_i - \overline{Y} = h_2 (v_i - \bar{v}),$$

$$S_x = h_1 \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} u_i^2 - n(\bar{u})^2\right]},$$

$$S_y = h_2 \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} v_i^2 - n(\bar{v})^2\right]},$$

и наконец,

$$r = \frac{h_1 h_2 \sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v})}{(n-1) S_x S_y} =$$

$$= \frac{\sum_{i=1}^{n} (u_i v_i) - n\bar{u}\bar{v}}{\sqrt{\sum_{i=1}^{n} u_i^2 - n(\bar{u})^2} \sqrt{\sum_{i=1}^{n} v_i^2 - n(\bar{v})^2}},$$

где

$$\bar{u} = \frac{1}{n} \sum_{i=1}^{n} u_i, \quad \bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i.$$

П р и м е р. В следующей таблице 1 приведены результаты $n=26$ экспериментов; числа m показывают, сколько раз встретилась пара соответствующих значений X и Y.

Таблица 1

X	Y	m	X	Y	m
23,0	0,48	2	26,0	0,51	1
24,0	0,50	4	26,0	0,53	2
24,5	0,49	3	26,5	0,50	1
24,5	0,50	2	26,5	0,52	1
25,0	0,51	1	27,0	0,54	2
25,5	0,52	1	27,0	0,52	1
26,0	0,49	2	28,0	0,53	3

Рассчитать эмпирические прямые регрессии Y на X и X на Y, изобразить их графически вместе с эмпирическими точками.

Р е ш е н и е. Выберем начала отсчета $x_0 = 26{,}0$, $y_0 = 0{,}50$ и масштабные коэффициенты $h_1 = 0{,}5$, $h_2 = 0{,}01$, т. е. произведем линейное преобразование рассматриваемых величин X и Y по формулам

$$u = \frac{X - 26{,}0}{0{,}5}, \quad v = \frac{Y - 0{,}50}{0{,}01}.$$

Составим расчетную таблицу 2 для вычисления нужных нам сумм

$$\Sigma u, \quad \Sigma v, \ \Sigma u^2, \ \Sigma v^2, \ \Sigma uv,$$

причем заметим, что каждое слагаемое следует учитывать столько раз, сколько оно встретилось в эксперименте, т. е. следует учитывать числа m в качестве множителей. В приведенной ниже расчетной таблице жирным шрифтом выделены выбранные начала отсчета, которым соответствуют нулевые значения u или v.

Таблица 2

X	Y	m	u	um	u^2m	v	vm	v^2m	uvm
23,0	0,48	2	−6	−12	72	−2	−4	8	24
24,0	0,50	4	−4	−16	64	0	0	0	0
24,5	0,49	3	−3	−9	27	−1	−3	3	9
24,5	0,50	2	−3	−6	18	0	0	0	0
25,0	0,51	1	−2	−2	4	1	1	1	−2
25,5	0,52	1	−1	−1	1	2	2	4	−2
26,0	0,49	2	0	0	0	−1	−2	2	0
26,0	0,51	1	0	0	0	1	1	1	0
26,0	0,53	2	0	0	0	3	6	18	0
26,5	0,50	1	1	1	1	0	0	0	0
26,5	0,52	1	1	1	1	2	2	4	2
27,0	0,54	2	2	4	8	4	8	32	16
27,0	0,52	1	2	2	4	2	2	4	4
28,0	0,53	3	4	12	48	3	9	27	36
Суммы		26	—	−26	248	—	22	104	87

В последней строке расчетной таблицы подсчитаны нужные суммы (уже с учетом множителей m):

$$\Sigma u = -26, \quad \Sigma u^2 = 248;$$
$$\Sigma v = 22, \quad \Sigma v^2 = 104, \quad \Sigma uv = 87$$

С помощью этих сумм вычисляем средние

$$\bar{u} = \frac{-26}{26} = -1, \quad \bar{X} = 26{,}0 + 0{,}5\,(-1) = 25{,}5,$$

$$\bar{v} = \frac{22}{26} = 0{,}846, \quad \bar{Y} = 0{,}50 + 0{,}01 \cdot 0{,}846 = 0{,}508,$$

эмпирические стандарты

$$S_x = 0{,}5\sqrt{\frac{248 - 26}{25}} = 1{,}49,$$

$$S_y = 0{,}01\sqrt{\frac{104 - 26 \cdot 0{,}846^2}{25}} = 0{,}0185,$$

и эмпирический коэффициент корреляции

$$r = \frac{87 - 26\,(-1) \cdot 0{,}846}{\sqrt{248 - 26}\,\sqrt{104 - 18{,}62}} = 0{,}793.$$

Эмпирические прямые регрессии имеют уравнения

$$y - 0{,}508 = 0{,}793\,\frac{0{,}0185}{1{,}49}\,(x - 25{,}5),$$

$$x - 25{,}5 = 0{,}793\,\frac{1{,}49}{0{,}0185}\,(y - 0{,}508),$$

или, после необходимого округления,

$$y - 0{,}51 = 0{,}0098\,(x - 25{,}5),$$
$$x - 25{,}5 = 64\,(y - 0{,}51).$$

Эти эмпирические прямые регрессии приведены на рис. 37 вместе с эмпирическими точками.

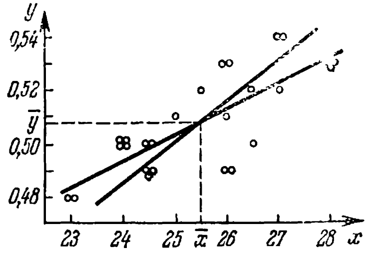


Рис. 37.

Контроль вычислений. Для контроля вычислений можно рекомендовать повторение расчета с другими началами отсчета. В приведенной ниже контрольной таблице 3 за начала отсчета приняты соответственно $x_0 = 25{,}5$, $y_0 = 0{,}51$.

Таблица 3

X	Y	m	u	um	u^2m	v	vm	v^2m	uvm
23,0	0,48	2	−5	−10	50	−3	−6	18	30
24,0	0,50	4	−3	−12	36	−1	−4	4	12
24,5	0,49	3	−2	−6	12	−2	−6	12	12
24,5	0,50	2	−2	−4	8	−1	−2	2	4
25,0	0,51	1	−1	−1	1	0	0	0	0
25,5	0,52	1	0	0	0	1	1	1	1
26,0	0,49	2	1	2	2	−2	−4	8	−4
26,0	0,51	1	1	1	1	0	0	0	0
26,0	0,53	2	1	2	2	2	4	8	4
26,5	0,50	1	2	2	4	−1	−1	1	−2
26,5	0,52	1	2	2	4	1	1	1	2
27,0	0,54	2	3	6	18	3	6	18	18
27,0	0,52	1	3	3	9	1	1	1	3
28,0	0,53	3	5	15	75	2	6	12	30
Суммы		26	—	0	222	—	−4	86	109

Суммы в контрольной таблице

$$\Sigma u = 0, \ \Sigma u^2 = 222,$$
$$\Sigma v = -4, \ \Sigma v^2 = 86, \ \Sigma uv = 109$$

позволяют проконтролировать средние:

$$\bar{u} = 0, \ \bar{X} = 25{,}5,$$
$$\bar{v} = \frac{-4}{26} = -0{,}154, \quad \bar{Y} = 0{,}51 - 0{,}01 \cdot 0{,}154 = 0{,}508,$$

эмпирические стандарты:

$$S_x = 0{,}5 \sqrt{\frac{222}{25}} = 1{,}49,$$
$$S_y = 0{,}01 \sqrt{\frac{86 - 26 \cdot 0{,}154^2}{25}} = 0{,}0185$$

и эмпирический коэффициент корреляции:

$$r = \frac{109}{\sqrt{222}\sqrt{86-0,62}} = 0,793.$$

Совпадение полученных результатов убеждает нас в отсутствии ошибок при вычислениях.

УПРАЖНЕНИЯ

1. В урне лежит $N = 100$ шаров, из них $M = 25$ белых. Из урны последовательно вынимают два шара. Пусть X_i — число белых шаров, появившихся при вынимании i-го шара ($i = 1, 2$). Найти коэффициент корреляции между величинами X_1 и X_2.

2. Точка $(X; Y)$ равномерно распределена в треугольнике

$$0 < X < 5,\ 0 < Y < 2 - \frac{2}{5}X.$$

Найти регрессию Y на X и регрессию X на Y.

3. Произвести расчет линейной корреляции по результатам эксперимента, представленным в следующей таблице (в столбце «m» указано, сколько раз встретилась соответствующая точка $(X; Y)$):

X	Y	m	X	Y	m
2,3	7,1	5	2,9	7,7	5
2,3	7,3	4	3,2	7,5	4
2,6	7,3	12	3,2	7,7	7
2,6	7,5	8	3,5	7,7	2
2,6	7,7	1	3,5	7,9	1
2,9	7,5	5	3,8	7,9	1

ОТВЕТЫ И УКАЗАНИЯ К УПРАЖНЕНИЯМ

К главе 1

1. $\frac{3}{36} = \frac{1}{12}$.

2. $\frac{C_{26}^{13} C_{26}^{13}}{C_{52}^{26}} = \frac{8 \cdot 17 \cdot 19^2 \cdot 23^2 \cdot 25^2}{27 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47} = 0{,}218.$

3. Вероятности событий $X = 11$ и $X = 12$ не равны:

$$\mathsf{P}(X = 11) = \frac{27}{216}, \quad \mathsf{P}(X = 12) = \frac{25}{216}.$$

Указание. Перечисленные в упражнении комбинации не равновероятны: вероятности каждой из комбинаций

$(6 + 4 + 1)$, $(6 + 3 + 2)$, $(5 + 4 + 2)$,
$(6 + 5 + 1)$, $(6 + 4 + 2)$, $(5 + 4 + 3)$

равны 6/216, вероятности каждой из комбинаций

$(5 + 5 + 1)$, $(5 + 3 + 3)$, $(4 + 4 + 3)$, $(6 + 3 + 3)$, $(5 + 5 + 2)$

равны 3/216, наконец, вероятность комбинации $(4 + 4 + 4)$ равна лишь 1/216.

4. $$p = \frac{2}{3}(1 - 0{,}03) + \frac{1}{3}(1 - 0{,}02) = 0{,}973.$$

5. Нет, так как

$$\mathsf{P}(A_1) = \frac{2}{4}, \ \mathsf{P}(A_2) = \frac{2}{4}, \ \mathsf{P}(A_3) = \frac{2}{4}, \ \mathsf{P}(A_1 \cdot A_2 \cdot A_3) = \frac{1}{4},$$

и поэтому $\mathsf{P}(A_1 \cdot A_2 \cdot A_3) \neq \mathsf{P}(A_1)\,\mathsf{P}(A_2)\,\mathsf{P}(A_3)$.

6. Нет, так как

$$\mathsf{P}(A_1) = \frac{4}{8} = \frac{1}{2}, \ \mathsf{P}(A_2) = \frac{1}{2}, \ \mathsf{P}(A_3) = \frac{1}{2}, \ \mathsf{P}(A_2 \cdot A_3) = \frac{1}{8},$$

и поэтому $\mathsf{P}(A_2 \cdot A_3) \neq \mathsf{P}(A_2)\,\mathsf{P}(A_3)$, хотя здесь

$$\mathsf{P}(A_1 \cdot A_2 \cdot A_3) = \frac{1}{8} = \mathsf{P}(A_1)\,\mathsf{P}(A_2)\,\mathsf{P}(A_3).$$

К главе 2

1. Распределение вероятностей числа X вынутых белых шаров дается таблицей

X	0	1	2
	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$

При вынимании шаров по схеме возвращенного шара число Y вынутых белых шаров может уже быть равно 0, 1, 2 или 3. Биномиальное распределение здесь $\left(\text{при } p = \frac{2}{10} = 0{,}2\right)$ дается таблицей

Y	0	1	2	3
	0,512	0,384	0,096	0,008

2. Распределение вероятностей суммы S очков, выпадающих на трех костях, дается таблицей

S	3	4	5	6	7	8	9	10	11	12	13	...	18
	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	...	$\frac{1}{216}$

3. Вероятность вынуть белый шар $p = \frac{10}{20} = \frac{1}{2}$. Число испытаний $n = 4$. Если X — число вынутых белых шаров, то

$$\mathsf{P}(X \geqslant 2) = \mathsf{P}(X = 2) + \mathsf{P}(X = 3) + \mathsf{P}(X = 4) =$$

$$= C_4^2 \left(\frac{1}{2}\right)^4 + C_4^3 \left(\frac{1}{2}\right)^4 + C_4^4 \left(\frac{1}{2}\right)^4 = \frac{11}{16}$$

или, иначе,

$$\mathsf{P}(X \geqslant 2) = 1 - \mathsf{P}(X = 0) - \mathsf{P}(X = 1) = \\ = 1 - \frac{1}{16} - \frac{4}{16} = \frac{11}{16} = 0{,}687.$$

4. При обозначениях предыдущей задачи

$$\mathsf{P}(X = 0) = \frac{10}{20} \cdot \frac{9}{19} \cdot \frac{8}{18} \cdot \frac{7}{17},$$

$$\mathsf{P}(X = 1) = 4 \cdot \frac{10}{20} \cdot \frac{10}{19} \cdot \frac{9}{18} \cdot \frac{8}{17},$$

и поэтому

$$\mathsf{P}(X \geqslant 2) = 1 - \mathsf{P}(X = 0) - \mathsf{P}(X = 1) = \frac{229}{323} = 0{,}709.$$

5. Пусть X — число выигранных партий. При $n = 4$

$$\mathsf{P}(X = 3) = C_4^3 \left(\frac{1}{2}\right)^4 = \frac{1}{4},$$

$$\mathsf{P}(X \geqslant 3) = \mathsf{P}(X = 3) + \mathsf{P}(X = 4) = \frac{5}{16}.$$

При $n = 8$

$$\mathsf{P}(X = 5) = C_8^5 \left(\frac{1}{2}\right)^8 = \frac{7}{32} < \frac{1}{4},$$

$$\mathsf{P}(X \geqslant 5) = \mathsf{P}(X = 5) + \mathsf{P}(X = 6) + \mathsf{P}(X = 7) + \mathsf{P}(X = 8) = \\ = \frac{93}{256} > \frac{5}{16}.$$

Таким образом, более вероятно выиграть:

а) три партии из четырех, чем пять из восьми;

б) не менее пяти партий из восьми, чем не менее трех из четырех.

6*. *Указание.* Независимость индикаторов I_1, I_2, I_3 означает выполнение восьми соотношений:

$$\begin{aligned}
&\mathsf{P}(I_1 = 1,\ I_2 = 1,\ I_3 = 1) = p_1p_2p_3,\\
&\mathsf{P}(I_1 = 0,\ I_2 = 1,\ I_3 = 1) = q_1p_2p_3,\ q_1 = 1 - p_1,\\
&\mathsf{P}(I_1 = 1,\ I_2 = 0,\ I_3 = 1) = p_1q_2p_3,\ q_2 = 1 - p_2,\\
&\mathsf{P}(I_1 = 0,\ I_2 = 0,\ I_3 = 1) = q_1q_2p_3,\\
&\mathsf{P}(I_1 = 1,\ I_2 = 1,\ I_3 = 0) = p_1p_2q_3,\ q_3 = 1 - p_3,\\
&\mathsf{P}(I_1 = 0,\ I_2 = 1,\ I_3 = 0) = q_1p_2q_3,\\
&\mathsf{P}(I_1 = 1,\ I_2 = 0,\ I_3 = 0) = p_1q_2q_3,\\
&\mathsf{P}(I_1 = 0,\ I_2 = 0,\ I_3 = 0) = q_1q_2q_3,
\end{aligned}$$

где $p_1 = \mathsf{P}(I_1 = 1)$, $p_2 = \mathsf{P}(I_2 = 1)$, $p_3 = \mathsf{P}(I_3 = 1)$.

Требуется доказать эквивалентность этих восьми соотношений следующим четырем:

$$\mathsf{P}(I_1 = 1,\ I_2 = 1) = p_1p_2,$$
$$\mathsf{P}(I_1 = 1,\ I_3 = 1) = p_1p_3,$$
$$\mathsf{P}(I_2 = 1,\ I_3 = 1) = p_2p_3,$$
$$\mathsf{P}(I_1 = 1,\ I_2 = 1,\ I_3 = 1) = p_1p_2p_3.$$

Переход от первых восьми соотношений к последним четырем опирается на правило сложения вероятностей, например:

$$\mathsf{P}(I_1 = 1, I_2 = 1, I_3 = 1) + \mathsf{P}(I_1 = 0, I_2 = 1, I_3 = 1) = \\ = \mathsf{P}(I_2 = 1,\ I_3 = 1),\ \text{т. е.}\ p_1p_2p_3 + q_1p_2p_3 = p_2p_3.$$

При обратном переходе эти преобразования выполняются в обратном порядке.

7*. Композиция двух биномиальных законов с одинаковыми значениями параметра p дает снова биномиальный закон:

$$\mathsf{P}(X + Y = m) = C_{n_1+n_2}^{m} p^m q^{n_1+n_2-m},\quad q = 1 - p.$$

Указание. Воспользоваться представлением величин X и Y в виде сумм индикаторов случайных событий (см. стр. 66).

8*. *Указание.* Воспользоваться формулой полной вероятности, которая здесь, в силу независимости величин X и Y, совпадает с формулой свертки для дискретных величин

$$\mathsf{P}(X + Y = m) = \sum_{k=0}^{m} \mathsf{P}(X = k)\,\mathsf{P}(Y = m - k).$$

К главе 3

1. Плотность распределения проекции равна

$$p(x) = \begin{cases} \dfrac{1}{\pi\sqrt{a^2 - x^2}} & \text{при}\ |x| < a, \\ 0 & \text{при}\ |x| > a. \end{cases}$$

2. а) Величины X и Y зависимы, так как плотности их распределений равны соответственно

$$p_X(x) = \begin{cases} 2C\sqrt{1 - x^2} & \text{при}\ |x| < 1, \\ 0 & \text{при}\ |x| > 1, \end{cases}$$

$$p_Y(y) = \begin{cases} 2C\sqrt{1 - y^2} & \text{при}\ |y| < 1, \\ 0 & \text{при}\ |y| > 1, \end{cases}$$

а плотность совместного распределения равна

$$p(x, y) = C = \frac{1}{\pi}\ \text{в круге}\ x^2 + y^2 < 1.$$

б) Величины X и Y зависимы, так как плотности их распределений равны соответственно

$$p_X(x) = \begin{cases} C(1-x) & \text{при } 0<x<1, \\ 0 & \text{вне интервала } (0, 1), \end{cases}$$

$$p_Y(y) = \begin{cases} C(1-y) & \text{при } 0<y<1, \\ 0 & \text{вне интервала } (0, 1), \end{cases}$$

а плотность совместного распределения равна

$p(x, y) = C = 2$ в треугольнике $0 < x < 1, 0 < y < 1 - x$.

в) Величины X и Y независимы, так как плотности их распределений равны соответственно

$$p_X(x) = \begin{cases} 2C & \text{при } 0<x<1, \\ 0 & \text{вне интервала } (0, 1), \end{cases}$$

$$p_Y(y) = \begin{cases} C & \text{при } 0<y<2, \\ 0 & \text{вне интервала } (0, 2), \end{cases}$$

а плотность совместного распределения равна

$$p(x, y) = \begin{cases} C = \frac{1}{2} & \text{в прямоугольнике } 0<x<1, 0<y<2, \\ 0 & \text{вне прямоугольника,} \end{cases}$$

так что $p(x, y) = p_X(x)\, p_Y(y)$.

3. $p_X(x) = k_1 f_1(x)$, $p_Y(y) = k_2 f_2(y)$, где

$$k_1 = \int_{-\infty}^{\infty} f_2(y)\,dy, \quad k_2 = \int_{-\infty}^{\infty} f_1(x)\,dx, \quad k_1 k_2 = 1.$$

4. По правилу умножения вероятностей

$$\mathsf{P}(-1<X<1, -2<Y<2) = \mathsf{P}(|X|<1)\,\mathsf{P}(|Y|<2) =$$
$$= \Phi\left(\frac{1}{\sqrt{3}}\right)\Phi\left(\frac{4}{\sqrt{3}}\right) = 0{,}4361\cdot 0{,}9791 = 0{,}427.$$

Указание. Величины X и Y независимы и распределены нормально с центром 0 и стандартами $\sigma_x = \sqrt{3}$, $\sigma_y = \sqrt{3}/2$.

5. 0,5. *Указание.* Прямая $y = 5x - 11$ проходит через точку (3; 4) — центр распределения, поэтому распределение симметрично относительно этой прямой.

6. Плотности распределения величин U и V равны

$$p_U(u) = \begin{cases} \frac{\lambda}{2} e^{-\frac{\lambda}{2}u} & \text{при } u>0, \\ 0 & \text{при } u<0 \end{cases}$$

(показательное распределение с параметром $\lambda/2$),

$$p_V(v) = \begin{cases} \lambda^2 e^{-\lambda v} v & \text{при} \quad v > 0, \\ 0 & \text{при} \quad v < 0 \end{cases}$$

(гамма-распределение с параметрами λ и $\alpha = 2$).

7*. Плотность распределения суммы двух величин равна

$$p_2(x) = \begin{cases} \frac{1}{4}(2 - |x|) & \text{при} \quad |x| \leqslant 2, \\ 0 & \text{при} \quad |x| > 2. \end{cases}$$

Плотность распределения суммы трех величин равна

$$p_3(x) = \begin{cases} \frac{1}{8}(3 - x^2) & \text{при} \quad |x| \leqslant 1, \\ \frac{1}{16}(3 - |x|)^2 & \text{при} \quad 1 < |x| \leqslant 3, \\ 0 & \text{при} \quad |x| > 3. \end{cases}$$

Кривые распределения приведены на рис. 38.

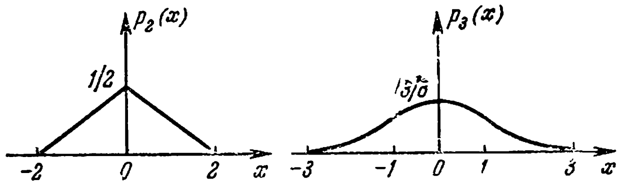


Рис. 38.

8. Плотность распределения суммы равна

$$p_{X+Y}(z) = \int_{-l}^{l} \frac{1}{2l} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z-x)^2}{2\sigma^2}\right] dx =$$

$$= \frac{1}{4l}\left[\Phi\left(\frac{z+l}{\sigma}\right) - \Phi\left(\frac{z-l}{\sigma}\right)\right].$$

9. Если $Y = |\ln X| = -\ln X$ $(0 < X < 1)$, то

$$p_Y(y) = p_X(e^{-y})\, e^{-y} = \begin{cases} e^{-y} & \text{при} \quad y > 0, \\ 0 & \text{при} \quad y < 0. \end{cases}$$

Если же $Z = |\log_a X| = \frac{1}{|\ln a|} Y$, то

$$p_Z(z) = \begin{cases} \lambda e^{-\lambda z} & \text{при} \quad z > 0, \\ 0 & \text{при} \quad z < 0, \end{cases}$$

где $\lambda = |\ln a|$.

10*. Плотность распределения частного $U = Y/X$ равна

$$p_U(u) = \int_0^\infty p_X(x)\, p_Y(ux)\, x\, dx \quad \text{при} \quad u > 0$$

и равна 0 при $u < 0$. *Указание.* Вероятность неравенства

$$u_1 < \frac{Y}{X} < u_2 \qquad (u_1 > 0,\ u_2 > 0)$$

записать в виде интеграла

$$\mathsf{P}\left(u_1 < \frac{Y}{X} < u_2\right) = \mathsf{P}(u_1 X < Y < u_2 X) =$$

$$= \int_0^\infty dx \int_{u_1 x}^{u_2 x} p(x, y)\, dy = \int_0^\infty p_X(x)\, dx \int_{u_1}^{u_2} p_Y(ux)\, x\, du$$

$$(y = xu,\ dy = x\, du)$$

и переменить порядок интегрирования.

11. Плотность распределения частного $Z = U/V$ равна

$$p_Z(z) = \frac{\Gamma\left(\frac{m+k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{k}{2}\right)} z^{\frac{k}{2}-1} (1+z)^{-\frac{m+k}{2}} \qquad (z > 0),$$

$$p_Z(z) = 0 \qquad (z < 0).$$

12*. Плотность распределения объема $V = X^3$ равна

$$p(v) = \frac{1}{3\sqrt[3]{v^2}\,\sigma\sqrt{2\pi}} \exp\left[-\frac{(\sqrt[3]{v} - a)^2}{2\sigma^2}\right].$$

13. $$p_Y(y) = \varphi_{a;\,\sigma}(\ln y)\frac{d\ln y}{dy} =$$

$$= \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln y - a)^2}{2\sigma^2}\right] \qquad (y > 0),$$

$$p_Y(y) = 0 \quad (y < 0).$$

Кривая распределения дана на рис. 39.

14. 5/9. *Указание.* Если X и Y — моменты прихода соответственно первого и второго лица, то искомая вероятность равна

$$\mathsf{P}\left(|X - Y| < \frac{1}{3}\right)$$

при условии, что точка $(X;\ Y)$ распределена равномерно в квадрате $0 < X < 1$, $0 < Y < 1$. Эта вероятность численно равна

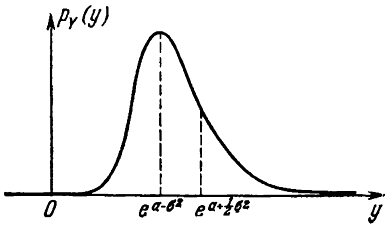


Рис. 39.

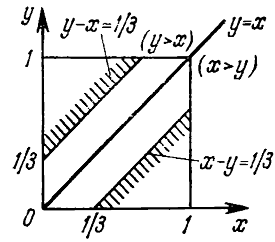


Рис. 40.

площади фигуры, выделенной на рис. 40 между прямыми $y - x = 1/3$ и $x - y = 1/3$.

К главе 4

1. 17,5. *Указание.* Если X — число пуль, израсходованных при стрельбе по одной фигуре, то $MX = 1 \cdot 0{,}5 + 2 \cdot 0{,}25 + 3 \cdot 0{,}25 = 1{,}75$.

2. $MX_1 = \frac{25}{100}$, $MX_2 = \frac{25}{100}$, $M(X_1 \cdot X_2) = \frac{25}{100} \cdot \frac{24}{99}$. Величины X_1 и X_2 зависимы, так как $M(X_1 \cdot X_2) \neq MX_1 \cdot MX_2$.

3. $MX = 3{,}5$, $\sigma(X) = \sqrt{\frac{35}{12}} = 1{,}71$.

4. $M(X_1 + X_2) = 3{,}5 \cdot 2 = 7$, $M(X_1 + X_2 + X_3) = 3{,}5 \cdot 3 = 10{,}5$,

$$\sigma(X_1 + X_2) = \sqrt{2 \cdot \frac{35}{12}} = 2{,}42, \quad \sigma(X_1 + X_2 + X_3) = \sqrt{3 \cdot \frac{35}{12}} = 2{,}96.$$

5. $MX = 1$. *Указание.* $X = \sum_{i=1}^{5} X_i$, где X_i — число вынутых белых шаров при i-м вынимании; $MX_i = \frac{4}{20} = \frac{1}{5}$.

6. 20 минут. *Указание.* Время ожидания $T = |X - Y|$, где X и Y — моменты прихода. Так как величины X и Y независимы и распределены равномерно в интервале (0, 1), то

$$MT = \int_0^1 \int_0^1 |x - y| \cdot 1 \cdot dx\,dy = \frac{1}{3} \text{ (часа)}.$$

7*. 16/3π. *Указание.* Длина L хорды BC равна (см. рис. 41)

$$L = 2\sqrt{1 - H^2} = 2\sqrt{1 - R^2 \sin^2 \Phi},$$

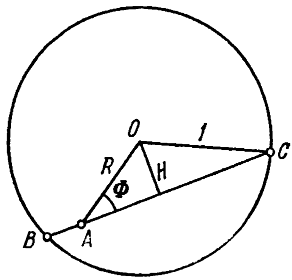


Рис. 41.

где R и Φ независимы и их плотности равны соответственно

$$p_R(r) = \begin{cases} 2\pi r \cdot \dfrac{1}{\pi} = 2r & \text{при } 0 < r < 1, \\ 0 & \text{вне } (0, 1), \end{cases}$$

$$p_\Phi(\varphi) = \begin{cases} \dfrac{2}{\pi} & \text{при } 0 < \varphi < \dfrac{\pi}{2}, \\ 0 & \text{вне } \left(0, \dfrac{\pi}{2}\right), \end{cases}$$

поэтому

$$\mathsf{M}L = \int_0^{\frac{\pi}{2}} \frac{2}{\pi} \, d\varphi \int_0^1 2\sqrt{1 - r^2 \sin^2 \varphi}\, 2r\, dr = \frac{16}{3\pi}.$$

8. $\mathsf{M}\chi_k^2 = \mathsf{M}(X_1^2 + X_2^2 + \ldots + X_k^2) = k,$

$\mathsf{D}\chi_k^2 = \mathsf{D}(X_1^2 + X_2^2 + \ldots + X_k^2) = 2k.$

Указание. Величины $X_1, X_2, \ldots, X_k$ независимы и имеют стандартное нормальное распределение; поэтому

$\mathsf{M}X_i^2 = \mathsf{D}X_i = 1,$

$\mathsf{D}X_i^2 = \mathsf{M}(X_i^2 - 1)^2 = \mathsf{M}X_i^4 - 2\mathsf{M}X_i^2 + 1 = 3 - 2 + 1 = 2$

$(i = 1, 2, \ldots, k).$

9. $$\mathsf{M}e^X = \int_{-\infty}^{\infty} e^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$

$$= e^a \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2} + \sigma t} dt = e^{a + \frac{1}{2}\sigma^2} \quad \left(t = \frac{x-a}{\sigma}\right).$$

10. $C_s = \dfrac{2\sqrt{2}}{\sqrt{k}}$, $E = \dfrac{12}{k}$.

11*. Если величины X и Y независимы, $\mathsf{M}X = a$, $\mathsf{M}Y = b$, то

$$\mathsf{M}\,[(X+Y)-(a+b)]^3 = \mathsf{M}(X-a)^3 + 3\,\mathsf{M}\,(X-a)^2\mathsf{M}\,(Y-b) + 3\,\mathsf{M}\,(X-a)\,\mathsf{M}\,(Y-b)^2 + \mathsf{M}\,(Y-b)^3 = \mathsf{M}\,(X-a)^3 + \mathsf{M}\,(Y-b)^3,$$

так как $\mathsf{M}\,(X-a) = 0$, $\mathsf{M}\,(Y-b) = 0$.

12*. Мода $e^{a-\sigma^2}$ (см. рис. 39, стр. 223); медиана e^a.
Указание.

$$\int\limits_0^{e^a} p_Y(y)\,dy = \int\limits_{-\infty}^{a} \varphi_{a;\,\sigma}(x)\,dx = \frac{1}{2} \qquad (y = e^x).$$

13*. *Указание.* В интеграле $\mathsf{M}X = \int\limits_0^\infty x\,p(x)\,dx = \int\limits_0^\infty p(x)\,dx \int\limits_0^x dy$ изменить порядок интегрирования.

К главе 5

1*. При любом n среднее арифметическое $\overline{X}$ имеет распределение Коши с плотностью $p(x) = \dfrac{1}{\pi}\dfrac{1}{1+x^2}$. Поэтому для любого числа C вероятность

$$\mathsf{P}\,(|\overline{X} - C| < \varepsilon) = \int\limits_{C-\varepsilon}^{C+\varepsilon} p(x)\,dx = \frac{1}{\pi}\,[\operatorname{arctg}(C+\varepsilon) - \operatorname{arctg}(C-\varepsilon)]$$

есть число, меньшее единицы (и не стремящееся к 1 при $n \to \infty$).
Указание. Для суммы двух величин $X_1 + X_2$ плотность найти по правилу композиции (свертки):

$$p_{X_1+X_2}(x) = \frac{1}{\pi^2}\int\limits_{-\infty}^{\infty} \frac{1}{1+t^2}\,\frac{1}{1+(x-t)^2}\,dt = \frac{1}{\pi}\,\frac{2}{x^2+4}.$$

Для суммы $X_1 + X_2 + \ldots + X_n$ по индукции получить плотность

$$p_{X_1+X_2+\ldots+X_n}(x) = \frac{1}{\pi}\,\frac{n}{x^2+n^2}$$

и воспользоваться соотношением $p_{\overline{X}}(x) = p_{X_1+\ldots+X_n}(nx)\cdot n$.

2. $\tilde{a} = \overline{X} = 0{,}40$, $\tilde{\sigma} = S = 1{,}6$.

3. $\tilde{\alpha} = 4$, $\tilde{\lambda} = 2$. *Указание.* Найти оценки математического ожидания $\overline{X} = 2{,}0$ и дисперсии $S^2 = 1{,}0$ и положить

$$\frac{\tilde{\alpha}}{\tilde{\lambda}} = \overline{X} = 2{,}0, \quad \frac{\tilde{\alpha}}{\tilde{\lambda}^2} = S^2 = 1{,}0.$$

4. $\hat{p} = \dfrac{X}{n}$, где X — число успехов при n-кратном повторении испытания.

5. $\hat{p} = \dfrac{1}{X}$, где X — число бросаний до первого попадания.

6*. Оценка м. н. п. $\hat{\theta} = \max\limits_{i=1,2,\ldots,n} X_i$; заметим, что оценка по методу моментов дает другое значение: $\tilde{\theta} = 2\overline{X} = \dfrac{2}{n}\sum\limits_{i=1}^{n} X_i$.

Указание. $p(X_1, X_2, \ldots, X_n) = \dfrac{1}{\theta^n}(\max X_i \leqslant \theta)$; $\mathsf{M}X = \dfrac{\theta}{2}$.

К главе 6

1. Для индикатора случайного события A, имеющего вероятность $\mathsf{P}(A) = p$, характеристическая функция

$$f(u) = pe^{iu} + q, \ q = 1 - p.$$

Для биномиального распределения характеристическая функция

$$f(u) = (pe^{iu} + q)^n.$$

Указание. Случайную величину с биномиальным распределением представить как сумму индикаторов (см. стр. 130).

2. Для равномерного распределения в интервале $(-\alpha, \alpha)$ характеристическая функция

$$f(u) = \int\limits_{-\alpha}^{\alpha} e^{iux}\frac{1}{2\alpha}\,dx = \frac{\sin\alpha u}{\alpha u}.$$

Для суммы n независимых величин, равномерно распределенных в $\left(-\dfrac{\alpha}{\sqrt{n}}, \dfrac{\alpha}{\sqrt{n}}\right)$, характеристическая функция

$$f(u) = \left(\frac{\sin\dfrac{\alpha u}{\sqrt{n}}}{\alpha u/\sqrt{n}}\right)^n.$$

Интересно проверить, что эта характеристическая функция стремится к характеристической функции нормального распределения при $n \to \infty$:

$$\left(\frac{\sin \frac{\alpha u}{\sqrt{n}}}{\alpha u/\sqrt{n}}\right)^n = \left(1 - \frac{\alpha^2 u^2}{3!n} + \ldots\right)^n \to \exp\left(-\frac{\alpha^2 u^2}{6}\right)$$

3. $\mathsf{M}\chi_k^2 = k$, $\mathsf{D}\chi_k^2 = 2k$. *Указание.* $\chi_k^2 = X_1^2 + X_2^2 + \ldots + X_k^2$, где $X_1^2, X_2^2, \ldots, X_k^2$ независимы и одинаково распределены с параметрами $\mathsf{M}X_i^2 = 1$, $\mathsf{D}X_i^2 = 2$ $(i = 1, 2, \ldots, k)$, см. ответ к задаче 8 главы 4.

4. При гипотезе $p = 1/2$ вероятность наблюденного или большего отклонения относительной частоты X/n от $1/2$ равна

$$\mathsf{P}\left(\left|\frac{X}{n} - \frac{1}{2}\right| \geqslant \frac{19}{12\,000}\right) \approx 1 - \Phi\left(\frac{\frac{19}{12\,000}\sqrt{12\,000}}{\sqrt{0,5\cdot 0,5}}\right) = 0,729.$$

Эта вероятность не мала, так что произведенный опыт не дает оснований сомневаться в справедливости гипотезы $p = 1/2$ (т. е. в симметрии монеты).

5*. *Указание.* Представить $S = \frac{\sigma}{\sqrt{k}}\sqrt{\chi_k^2}$ и воспользоваться плотностью χ^2-распределения:

$$\mathsf{M}S = \frac{\sigma}{\sqrt{k}} \int_0^\infty \sqrt{u}\, \frac{1}{2^{k/2}\Gamma(k/2)}\, e^{-\frac{u}{2}} u^{\frac{k}{2}-1}\, du.$$

6. Эмпирические дисперсии для каждой из десяти серий даны в таблице

i	1	2	3	4	5	6	7	8	9	10
$S_i^2\cdot 10^4$	1,09	0,76	1,22	1,29	1,16	1,09	0,76	1,29	1,20	0,88

Средняя эмпирическая дисперсия равна

$$S^2 = \frac{S_1^2 + S_2^2 + \ldots + S_{10}^2}{10} = 1,074\cdot 10^{-4},$$

откуда $S = 0,01036$. Число степеней свободы для нее равно $k = 110 - 10 = 100$. При надежности $\mathscr{P} = 0,99$ по таблице 4 Приложения находим нижнюю и верхнюю границы для доверительной оценки:

$$z_{\text{н}} = 0,845, \quad z_{\text{в}} = 1,22.$$

Доверительный интервал для средней квадратической ошибки измерений:

$$0{,}845\cdot 0{,}01036 < \sigma < 1{,}22\cdot 0{,}01036$$

или $0{,}0088 < \sigma < 0{,}0126$.

К главе 7

1.

$$\rho(X_1, X_2) = \frac{MX_1X_2 - MX_1MX_2}{\sigma(X_1)\sigma(X_2)} = \frac{\frac{25}{100}\cdot\frac{24}{99} - \left(\frac{25}{100}\right)^2}{\frac{25}{100}\cdot\frac{75}{100}} = -\frac{1}{99}.$$

2. Уравнение регрессии Y на X:

$$y = 1 - \frac{1}{5}x \quad \text{при} \quad 0 < x < 5;$$

уравнение регрессии X на Y:

$$x = \frac{5}{2} - \frac{5}{4}y \quad \text{при} \quad 0 < y < 2.$$

3. $\overline{X} = 2{,}80$, $\overline{Y} = 7{,}47$, $S_x = 0{,}37$, $S_y = 0{,}2$, $r = 0{,}83$. Эмпирические прямые регрессии:

$$y - 7{,}47 = 0{,}47\ (x - 2{,}80),$$
$$x - 2{,}80 = 1{,}5\ (y - 7{,}47).$$

ПРИЛОЖЕНИЯ

Таблица 1

Интеграл вероятностей $\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx.$

t	$\Phi(t)$		t	$\Phi(t)$		t	$\Phi(t)$	
0,00	0,0000		1,00	0,6827		2,00	0,9545	
		399			236			51
0,05	0,0399		1,05	0,7063		2,05	0,9596	
		398			224			47
0,10	0,0797		1,10	0,7287		2,10	0,9643	
		395			212			41
0,15	0,1192		1,15	0,7499		2,15	0,9684	
		393			200			38
0,20	0,1585		1,20	0,7699		2,20	0,9722	
		389			188			34
0,25	0,1974		1,25	0,7887		2,25	0,9756	
		384			177			30
0,30	0,2358		1,30	0,8064		2,30	0,9786	
		379			166			26
0,35	0,2737		1,35	0,8230		2,35	0,9812	
		371			155			24
0,40	0,3108		1,40	0,8385		2,40	0,9836	
		365			144			21
0,45	0,3473		1,45	0,8529		2,45	0,9857	
		356			135			19
0,50	0,3829		1,50	0,8664		2,50	0,9876	
		348			125			16
0,55	0,4177		1,55	0,8789		2,55	0,9892	
		338			115			15
0,60	0,4515		1,60	0,8904		2,60	0,9907	
		328			107			13
0,65	0,4843		1,65	0,9011		2,65	0,9920	
		318			98			11
0,70	0,5161		1,70	0,9109		2,70	0,9931	
		306			90			9
0,75	0,5467		1,75	0,9199		2,75	0,9940	
		296			82			9
0,80	0,5763		1,80	0,9281		2,80	0,9949	
		284			76			7
0,85	0,6047		1,85	0,9357		2,85	0,9956	
		272			69			7
0,90	0,6319		1,90	0,9426		2,90	0,9963	
		260			62			5
0,95	0,6579		1,95	0,9488		2,95	0,9968	
		248			57			

Примечание к таблице 1. Между строками таблицы мелким шрифтом напечатаны табличные разности функции. Это сделано для удобства интерполяции; например, чтобы вычислить значение $\Phi(t)$ при $t = 1{,}717$, находим меньшее табличное значение $t_{\text{табл}} = 1{,}70$, затем табличную разность функции

$$90 \cdot 10^{-4} = 0{,}0090$$

и вычисляем

$$\Phi(1{,}717) = \Phi(1{,}70) + \frac{0{,}017}{0{,}05} \cdot 0{,}0090 = 0{,}9109 + 0{,}0031 = 0{,}9140.$$

Ошибка при линейной интерполяции не превышает 0,0002 в интервале $0{,}4 < t < 1{,}8$ и не превышает 0,0001 в остальной части таблицы 1.

При значениях $t \geqslant 3$ значения интеграла вероятностей $\Phi(t)$ приведены ниже.

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
3,0	0,9973	3,4	0,99933	3,8	0,99985
3,1	0,9981	3,5	0,99953	3,9	0,99990
3,2	0,9986	3,6	0,99968	4,0	0,99993
3,3	0,9990	3,7	0,99978		

Здесь ошибка линейной интерполяции не превышает 0,0001 при $t < 3{,}5$ и не превышает 0,00001 при $t > 3{,}5$.

Интеграл вероятностей $\Phi(t)$ применяется при расчете вероятностей в нормальном распределении: если распределение вероятностей случайной величины X нормально с параметрами a и σ (§ 3.3), то

$$\mathsf{P}(|X - a| < t\sigma) = \Phi(t) \quad (t > 0),$$

$$\mathsf{P}(x_1 < X < x_2) = \frac{1}{2}[\Phi(t_2) - \Phi(t_1)],$$

где

$$t_1 = \frac{x_1 - a}{\sigma}, \quad t_2 = \frac{x_2 - a}{\sigma}, \quad \Phi(-t) = -\Phi(t).$$

Ошибка при расчете вероятностей $\mathsf{P}(x_1 < X < x_2)$ с помощью таблицы 1 не превосходит 0,0004.

Таблица 2

Решения уравнения $\Phi(t) = \mathcal{P}$

Значения $t = t(\mathcal{P})$ удовлетворяют условию $\Phi(t) = \mathcal{P} = 1 - \alpha$, где $\Phi(t)$ — интеграл вероятностей (табл. 1).

$\mathcal{P}$	$\alpha = 1 - \mathcal{P}$	$t = t(\mathcal{P})$	$\mathcal{P}$	$\alpha = 1 - \mathcal{P}$	$t = t(\mathcal{P})$
0,95	0,05	1,960	0,996	0,004	2,878
0,96	0,04	2,054	0,997	0,003	2,968
0,97	0,03	2,170	0,998	0,002	3,090
0,98	0,02	2,326	0,999	0,001	3,291
0,99	0,01	2,576	0,9995	0,0005	3,481
0,991	0,009	2,612	0,9998	0,0002	3,719
0,992	0,008	2,652	0,9999	0,0001	3,891
0,993	0,007	2,697	0,99999	10^{-5}	4,417
0,994	0,006	2,748	0,999999	10^{-6}	4,892
0,995	0,005	2,807			

Таблица 2 применяется при определении доверительного интервала для *неизвестного* параметра a в нормальном распределении с *известным* стандартным отклонением σ (§ 6.4): если случайные величины $X_1, X_2, \ldots, X_n$ независимы и распределены нормально с параметрами a и σ, то с заданной вероятностью $\mathcal{P}$ имеет место неравенство

$$\bar{X} - t\frac{\sigma}{\sqrt{n}} < a < \bar{X} + t\frac{\sigma}{\sqrt{n}},$$

где $t = t(\mathcal{P})$, $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$, т. е. интервал

$$\left(\bar{X} - t\frac{\sigma}{\sqrt{n}},\ \bar{X} + t\frac{\sigma}{\sqrt{n}}\right)$$

накрывает неизвестное значение a.

Величина $t\dfrac{\sigma}{\sqrt{n}}$ оценивает абсолютную ошибку приближенного равенства $a \approx \bar{X}$ с заданной вероятностью $\mathcal{P}$.

При обработке результатов измерений под a понимают истинное значение измеряемой величины, под $X_1, X_2, \ldots, X_n$ — результаты независимых измерений величины a, произведенных без систематических ошибок и с известной средней квадратической ошибкой σ. При неизвестном значении σ надо пользоваться таблицей 3.

Таблица 3

Значения $t(\mathscr{P};\ k)$ для t-распределения Стьюдента

Значения $t = t(\mathscr{P};\ k)$ удовлетворяют условию

$$2\int_0^t p_k(t)\,dt = \mathscr{P},$$

где $p_k(t)$ — плотность распределения Стьюдента (6.5 8) с числом степеней свободы, равным k.

k	$\mathscr{P}$		k	$\mathscr{P}$	
	0,95	0,99		0,95	0,99
2	4,303	9,925	18	2,101	2,878
3	3,182	5,841	20	2,086	2,845
4	2,776	4,604	25	2,060	2,787
5	2,571	4,032	30	2,042	2,750
6	2,447	3,707	35	2,030	2,724
7	2,365	3,499	40	2,021	2,704
8	2,306	3,355	45	2,014	2,689
9	2,262	3,250	50	2,009	2,678
10	2,228	3,169	60	2,000	2,660
11	2,201	3,106	70	1,994	2,648
12	2,179	3,055	80	1,990	2,639
13	2,160	3,012	90	1,987	2,632
14	2,145	2,977	100	1,984	2,626
15	2,131	2,947	∞	1,960	2,576
16	2,120	2,921			

Интерполяция таблицы 3 допустима только по аргументу k. Ошибка линейной интерполяции для значений k от 17 до 59 не превосходит 0,002 при $\mathscr{P} = 0{,}95$ и 0,004 при $\mathscr{P} = 0{,}99$; ошибка линейной интерполяции для значений k от 61 до 99 не превосходит 0,001.

Для значений $k > 100$ значения $t(\mathscr{P},\ k)$ с точностью до 0,001 можно вычислять по формулам

$$t(0{,}95;\ k) = 1{,}960 + 2{,}4/k;$$

$$t(0{,}99;\ k) = 2{,}576 + 5{,}0/k.$$

Таблица 3 применяется при определении доверительного интервала для *неизвестного* параметра a в нормальном распределении с *неизвестным* стандартным отклонением σ (§ 6.5): если случайные величины $X_1, X_2, \ldots, X_n$ независимы и распределены нормально с параметрами a и σ, то с заданной вероятностью $\mathscr{P}$ имеет место неравенство

$$\bar{X} - t\frac{S}{\sqrt{n}} < a < \bar{X} + t\frac{S}{\sqrt{n}},$$

где $t = t(\mathscr{P}; k)$, $k = n - 1$,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Таблица 4

Коэффициенты $z_н = z_н(\mathscr{P}; k)$ и $z_в = z_в(\mathscr{P}; k)$ для доверительных границ среднего квадратического отклонения σ:

$$S \cdot z_н < \sigma < S \cdot z_в.$$

k	$\mathscr{P} = 0{,}95$		$\mathscr{P} = 0{,}99$		k	$\mathscr{P} = 0{,}95$		$\mathscr{P} = 0{,}99$	
	$z_н$	$z_в$	$z_н$	$z_в$		$z_н$	$z_в$	$z_н$	$z_в$
2	0,521	6,28	0,434	14,12	17	0,750	1,50	0,690	1,73
3	0,566	3,73	0,483	6,47	18	0,756	1,48	0,696	1,70
4	0,599	2,87	0,519	4,40	19	0,760	1,46	0,702	1,67
5	0,624	2,45	0,546	3,48	20	0,765	1,44	0,707	1,64
6	0,644	2,20	0,569	2,98	25	0,784	1,38	0,730	1,54
7	0,661	2,04	0,588	2,66	30	0,799	1,34	0,748	1,48
8	0,675	1,92	0,604	2,44	35	0,811	1,30	0,762	1,43
9	0,688	1,83	0,618	2,28	40	0,821	1,28	0,774	1,39
10	0,699	1,75	0,630	2,15	45	0,829	1,26	0,784	1,36
11	0,708	1,70	0,641	2,06	50	0,837	1,24	0,793	1,34
12	0,717	1,65	0,651	1,98	60	0,849	1,22	0,808	1,30
13	0,725	1,61	0,660	1,91	70	0,858	1,20	0,820	1,27
14	0,732	1,58	0,669	1,85	80	0,866	1,18	0,829	1,25
15	0,739	1,55	0,676	1,81	90	0,873	1,17	0,838	1,23
16	0,745	1,52	0,683	1,76	100	0,879	1,16	0,845	1,22

Интерполяция таблицы 4 допустима только по аргументу k. Ошибка линейной интерполяции не превосходит 0,2% для $z_н$ и 1% для $z_в$.

При больших значениях числа степеней свободы k коэффициенты $z_н$ и $z_в$ можно вычислять по приближенным формулам

$$z_н = \left(1 + \frac{t(\mathscr{P})\sqrt{2}}{\sqrt{k}}\right)^{-\frac{1}{2}}, \quad z_в = \left(1 - \frac{t(\mathscr{P})\sqrt{2}}{\sqrt{k}}\right)^{-\frac{1}{2}},$$

где $t(0{,}95) = 1{,}96$, $t(0{,}99) = 2{,}58$.

Погрешности этих приближенных формул убывают с увеличением k; при $k \approx 100$ они составляют 1% для $\mathscr{P} = 0{,}95$ и 2% для $\mathscr{P} = 0{,}99$.

Пример. При $\mathscr{P} = 0{,}99$ и $k = 400$ получаем

$$z_н = \left(1 + \frac{2{,}58\sqrt{2}}{20}\right)^{-\frac{1}{2}} = 0{,}920, \quad z_в = \left(1 - \frac{2{,}58\sqrt{2}}{20}\right)^{-\frac{1}{2}} = 1{,}11$$

с погрешностью менее 1%.

Таблица 4 применяется при определении доверительного интервала для *неизвестного параметра* σ в нормальном распределении (§ 6.5): если случайные величины $X_1, X_2, \ldots, X_n$ независимы и распределены нормально с параметрами a и σ, то с заданной вероятностью $\mathscr{P}$ имеет место неравенство

$$S \cdot z_н < \sigma < S \cdot z_в,$$

где

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

ЛИТЕРАТУРА

Для более полного ознакомления с основами теории вероятностей и некоторыми ее приложениями можно рекомендовать книгу

Е. С. Вентцель, Теория вероятностей, «Наука», 1971.

Читатели с повышенной математической подготовкой могут более серьезно изучить теорию по книге

Б. В. Гнеденко, Курс теории вероятностей, «Наука», 1970.

Различные приложения теории вероятностей к обработке результатов эксперимента и к более общим задачам математической статистики можно найти в книгах

Ю. В. Линник, Метод наименьших квадратов и основы теории обработки наблюдений, Физматгиз, 1962.

Н. В. Смирнов и *И. В. Дунин-Барковский*, Краткий курс математической статистики, Физматгиз, 1959.

А. Хальд, Математическая статистика с техническими приложениями, ИЛ, 1956.

Приобрести навыки в решении задач можно с помощью задачника

Е. С. Вентцель и *Л. А. Овчаров*, Теория вероятностей, «Наука», 1973.

Для справок о всех вероятностных распределениях, изложенных в данном пособии, о приложениях теории вероятностей к статистической обработке наблюдений, а также об основных понятиях математической статистики можно рекомендовать книгу

Л. Н. Большев и *Н. В. Смирнов*, Таблицы математической статистики, М., Изд-во ВЦ АН СССР, 1968.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

$\bar{A}$ — событие, противоположное событию A 16, 17
$A + B$, A или B, $A \bigcup B$ — сумма событий A и B 13, 17
$A \cdot B$, A и B — совмещение событий A и B 23, 24
$\operatorname{cov}(X, Y) = \mu_{1,1}$ — ковариация величин X и Y 143
C_s — коэффициент асимметрии 138
$\mathsf{D}X = \sigma^2(X)$ — дисперсия величины X 121
$\exp(t) = e^t$ — экспонента 76, 88
$f_X(u) = \mathsf{M}e^{iuX}$ — характеристическая функция 169
$F(x)$ — функция распределения величины X 81, 82
$\mathsf{M}X$ — математическое ожидание величины X 109, 111, 113
M_xY — регрессия величины Y на величину X 194
$\mathsf{P}(A)$ — вероятность события A 10
$\mathsf{P}(B \mid A)$ — условная вероятность события B при условии осуществления события A 25, 26
$p(x) = p_X(x)$ — плотность распределения вероятностей величины X 69, 94
$p(x, y)$ — плотность совместного распределения вероятностей величин X и Y 87
$p_{Y|x}(y)$ — плотность условного распределения 193
$\mathcal{P}$ — доверительная вероятность, надежность оценки 180
r — выборочный коэффициент корреляции 208
S — эмпирический (выборочный) стандарт 161
$\bar{x}$ — среднее арифметическое значение 108
$x_{1/2}$ — медиана 112, 125
$x_{\mathcal{P}}$ — квантили 125
$\bar{X}$ — среднее арифметическое случайных величин, выборочное среднее 151, 155
$\mathring{X}$ — центрированная случайная величина 121
$\alpha_k = \mathsf{M}X^k$ — начальный момент порядка k 137, 139
$\mu_k = \mathsf{M}\mathring{X}^k$ — центральный момент порядка k 137, 139
σ, $\sigma(X)$ — стандарт, среднее квадратическое отклонение 79, 121
$\varphi_{\alpha,\sigma}(x)$ — плотность нормального распределения 76
$\Phi(t)$ — интеграл вероятностей 78, 229
χ^2_k — «хи-квадрат»-распределение с k степенями свободы 102
$Z_n \xrightarrow[n\to\infty]{\text{вер}} a$ — предел по вероятности 149

Лев Зимонович Румшиский

ЭЛЕМЕНТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

(Серия: «Избранные главы высшей математики для инженеров и студентов втузов»)

М., 1976 г., 240 стр. с илл.

Редакторы *М. С. Никулин* и *М. М. Горячая*
Техн. редактор *С. Я. Шкляр*
Корректор *М. Л. Медведская*

Сдано в набор 8/I-1976 г.
Подписано к печати 5/III 1976 г.
Бумага 84×108¹/₃₂. Физ. печ. л. 7,5.
Усл. печ. л. 12,6 Уч.-изд. л. 11,72.
Тираж 100 000 экз.
Цена книги 41 коп. Заказ № 38

Издательство «Наука»
Главная редакция
физико-математической литературы
117071, Москва, В-71, Ленинский проспект, 15

2-я типография издательства «Наука»,
Москва, Шубинский пер., 10

Цена 41 коп.