А.М. Карлов

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА ДЛЯ ЭКОНОМИСТОВ

Рекомендовано УМО по образованию в области финансов,
учета и мировой экономики
в качестве **учебного пособия** для студентов,
обучающихся по специальностям
«Финансы и кредит»,
«Бухгалтерский учет, анализ и аудит»

КноРус
МОСКВА
2011

УДК 591.2(075.8)
ББК 22.17я73
К23

Рецензенты:
Е.Н. Кикоть, проф. кафедры высшей математики Балтийской государственной академии рыбопромыслового флота, д-р пед. наук, проф.;
А.В. Сербулов, декан учетно-аналитического факультета Балтийского института экономики и финансов, д-р экон. наук, проф.;
В.А. Бывшев, проф. кафедры математического моделирования экономических процессов Финансовой академии при Правительстве РФ, д-р техн. наук

Карлов А.М.
К23 Теория вероятностей и математическая статистика для экономистов : учебное пособие / А.М. Карлов. — М. : КНОРУС, 2011. — 264 с.

ISBN 978-5-406-00267-4

В доступной форме приведено описание основных разделов теории вероятностей и математической статистики, предусмотренных учебной программой дисциплины в соответствии с государственным образовательным стандартом по специальностям 080105.65 «Финансы и кредит» и 080109.65 «Бухгалтерский учет, анализ и аудит». Изложение теории сопровождается большим количеством графического иллюстративного материала и решением примеров экономической направленности.

Предназначено для студентов вузов по экономическим специальностям всех форм обучения.

УДК 591.2(075.8)
ББК 22.17я73

Карлов Анатолий Михайлович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА ДЛЯ ЭКОНОМИСТОВ

Санитарно-эпидемиологическое заключение
№ 77.99.60.953.Д.006828.04.10 от 28.04.2010 г.

Изд. № 1835. Подписано в печать 12.07.2010. Формат 60×90/16.
Гарнитура «PetersburgC». Печать офсетная.
Усл. печ. л. 16,5. Уч.-изд. л. 12,0. Тираж 2000 экз. Заказ №

ООО «Издательство КноРус».
129110, Москва, ул. Большая Переяславская, 46, стр. 7
Тел.: (495) 680-7254, 680-0671, 680-1278.
E-mail: office@knorus.ru http://www.knorus.ru

Отпечатано в ОАО «Московская типография № 2».
129085, Москва, пр. Мира, 105.

ISBN 978-5-406-00267-4

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Теория вероятностей как раздел математической науки изучает общие закономерности в случайных явлениях.

Случайным явлением называют такое явление, которое при неоднократном воспроизведении одного и того же опыта каждый раз протекает несколько по-иному. Совершенно очевидно, что в природе, обществе, экономике нет ни одного явления, в котором в той или иной мере не присутствовали бы элементы случайности. Протекание (развитие) любого явления зависит от влияния множества факторов. Для того чтобы определить (предсказать) исход случайного явления, нужно оценить в виде математических функциональных зависимостей (моделей) влияние множества факторов на исход события. Однако во многих практических задачах интересующий нас исход опыта зависит от столь большого числа факторов, что практически невозможно все их зарегистрировать и учесть. Предположим, что из всего множества факторов Φ нам удалось выделить наиболее существенные Φ_1 факторы и получить математическую модель, позволяющую рассчитывать исход опыта. В этом состоит суть детерминистского подхода к изучаемому явлению. Но при этом остается множество $\Phi_2 = \Phi/\Phi_1$ неучтенных второстепенных факторов, которые имеют некоторое влияние на развитие явления. Так как мы не определили влияние множества факторов Φ_2 на развитие явления, это приводит к тому, что рассчитанный по полученной нами математической модели исход опыта будет не совсем точным. Таким образом, наличие множества неучтенных факторов Φ_2 вносит элемент случайности в развитие любого явления и приводит к невозможности однозначно предсказать его исход.

Первоначально теория вероятностей формировалась в основном на материале азартных игр, которые создавались именно так, чтобы исход опыта в них был независим от поддающихся наблюдению условий опыта, т.е. был чисто случайным. Схемы азартных игр представляют собой простые модели случайных явлений, дающие возможность наблюдать и изучать вероятностные законы этих явлений. Возможность в азартных играх неограниченно повторять один и тот же опыт позволяет экспериментально проверять справедливость этих законов в условиях действительности. В дальнейшем теория вероятностей нашла применение и в других областях человеческой деятельности.

Методы теории вероятностей позволяют количественно оценить влияние случайных неучтенных факторов на те или иные явления.

При вероятностном методе используются статистические (вероятностные) модели. Эти модели позволяют описывать процессы, явления и события, которые невозможно достоверно предсказать. Такие процессы, явления и события называются случайными. Несмотря на то что в каждом конкретном испытании нельзя предсказать заранее возможное значение или исход случайного процесса или события, при многократных наблюдениях (испытаниях) случайных процессов и явлений в одних и тех же условиях можно выявить определенные вероятностные закономерности. Знание вероятностных закономерностей случайных процессов и событий позволяет принимать решения, которые в конечном итоге приводят к благоприятному результату.

Переход экономики России от жестко регламентированных (при СССР) к рыночным отношениям привел к необходимости перехода от детерминированных моделей в экономике к вероятностным.

При старых хозяйственных отношениях производитель получал государственные плановые показатели и соответствующие средства. Его деятельность практически не зависела от того, потребляются ли произведенные им товары (услуги). При этом цена товаров (услуг) определялась государством и абсолютно не зависела от потребителя. Более того, изменение стоимости сырья, необходимого для производства товара (услуги), как правило, компенсировалось государством. Что касается потребителя, то он был лишен права выбора, поскольку цены на товары (услуги) практически были одинаковы независимо от места их приобретения. Такой жесткий детерминизм служил основой для государственного планирования производства товаров (услуг).

Картина коренным образом изменилась с переходом к рыночным отношениям. Для производителя жизненно важным стал вопрос продажи предлагаемых им товаров (услуг), а стало быть, вопрос цены на эти товары (услуги). При этом цена стала являться результатом компромисса между производителем и потребителем, т.е. стала зависеть от покупательной способности потребителя. Последняя в свою очередь зависит от множества разнообразных причин (зарплаты, инфляции, региона, наличия на рынке конкурирующих предложений и т.п.). Что касается себестоимости произведенных товаров (услуг), то она также напрямую стала зависеть от многочисленного числа факторов, включающих уровень инфляции, размер транспортных расходов, уровень использования новых технологий, регион производства, цены на необходимые для производства товары и услуги и т.д. Следствием является то, что себестоимость с течением времени не остается постоянной, а подвергается непрерывным, в том числе и случайным, изменениям. Это свидетельствует о том, что как цена, по которой может быть продан товар (услуга), так и стоимость средств, затраченных на его производство, с позиции производителя являются принципиально случайными величинами.

Как отмечают многие ученые и специалисты, одним из наиболее концептуально важных моментов в экономической политике на рынке является умение четко представлять стратегические цели и перспективы компаний, фирм, акционерных обществ, смешанных предприятий.

Для формулировки таких стратегий требуется достаточно ясно и четко представлять стратегические и тактические потенциальные возможности производителя, а также уметь оценивать возможность их реализации на рынке. Последнее требует очень четкого понимания процессов, происходящих на этом рынке, где помимо конкурирующих производителей, оказывающих аналогичные услуги, задействованы тысячи потребителей таких услуг, покупательная способность которых на предлагаемые услуги подвержена непрерывным и очень сильным флюктуациям.

Рыночные отношения породили принципиально иной класс задач, стоящих перед компаниями, решение которых может быть сформулировано исключительно на языке вероятностных оценок тех или иных экономических показателей. В качестве примера можно привести отдельные задачи, которые в условиях дорыночных отношений даже не могли быть поставлены. Формулировка их может выглядеть следующим образом:

1) определение средней прибыли, которая может быть получена компанией на рынке, при заданной цене на производимые ею товары (услуги);

2) определение среднеквадратичного отклонения (или его квадрата-дисперсии) истинного значения прибыли от ее среднего значения при заданной цене на производимые компанией товары (услуги);

3) определение наиболее вероятного значения прибыли при заданной цене на производимые компанией товары (услуги);

4) определение цены на производимые компанией товары (услуги), при которой прибыль превысит заданное значение с заданной вероятностью;

5) определение вероятности того, что при заданной цене на производимые компанией товары (услуги) ее прибыль превысит заданное значение.

Приведенный перечень задач, конечно, не исчерпывает весь их спектр. Главное здесь состоит в том, что эти задачи, во-первых, жизненны, а во-вторых, относятся к одному и тому же классу и требуют для своего решения одинакового методологического подхода.

Решение сформулированных задач возможно только на основе вероятностных методов анализа экономических показателей, которые в условиях рыночной экономики являются случайными.

ГЛАВА 1

ПОНЯТИЕ СЛУЧАЙНЫХ СОБЫТИЙ, ВЕРОЯТНОСТИ СОБЫТИЙ

1.1. Случайные события

Чтобы придать теории вероятностей точный определенный смысл, необходимо прежде всего условиться об основных используемых понятиях. Во введении мы уже дали определение такого понятия, как случайное явление, т.е. такого явления, которое при неоднократном повторении одного и того же опыта протекает каждый раз несколько по-иному. В силу различного протекания случайного явления при повторении опыта исход этих опытов каждый раз будет различным. Исходы случайного испытания (опыта) называют ***элементарными событиями***.

Все возможные исходы опыта называют пространством элементарных событий Ω. Во многих случаях пространство элементарных событий Ω может быть определено из анализа протекания самого случайного явления. Так, например, при игре в карты колодой в 36 карт при случайном выборе из колоды одной карты число возможных исходов опыта, а значит и пространство элементарных событий Ω, равно тридцати шести.

Если в результате каждого испытания наступает один и тот же исход, то испытание называют детерминированным неслучайным, а данный исход испытания — достоверным событием. Предположим, что при игре в игральные кости нас интересует такое событие, когда при бросании игральной кости число выпавших очков будет меньше или равно шести. Очевидно, что это событие будет выпадать при каждом бросании игральной кости и является достоверным событием.

Событие, которое не может произойти при данных условиях проведения опыта, называется ***невозможным событием***. Например, при игре в карты колодой в 36 карт событие, состоящее в том, что случайно вытянутая из колоды карта окажется джокером, является невозможным событием, так как джокера в колоде из 36 карт нет.

Случайным событием называют событие, которое в результате опыта может произойти, а может и не произойти. Например, событие состоящее в том, что наудачу вытянутая из колоды карта окажется пиковым тузом, является случайным событием.

В зависимости от количества возможных исходов опыта составляющих пространство элементарных событий Ω различают ***конечные, счетные*** или ***бесконечные, несчетные*** пространства элементарных событий. Например, при стрельбе по плоской мишени нас интересует событие, состоящее в попадании в точку с заданными координатами. Так как на плоскости бесконечное число точек с отличающимися друг от друга координатами, то число возможных исходов данного опыта, образующих пространство Ω, будет бесконечным.

При подбрасывании одной игральной кости пространство элементарных событий состоит из шести возможных исходов (выпавших очков)

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Будем обозначать случайные события большими буквами латинского алфавита A, B, C, D ... и рассмотрим возможные отношения между случайными событиями на примере бросания одной игральной кости. Обозначим каждый из шести элементарных исходов случайными событиями A_1, A_2, A_3, A_4, A_5, A_6. Предположим, что нас интересует случайное событие B, состоящее в том, что при бросании игральной кости выпадет четное число очков. Значит, событие B образует подмножество $B = \{A_2; A_4; A_6\}$.

Из сравнения множества Ω и подмножества B видно, что множество Ω включает в себя подмножество B. Это обозначают как отношение включения $B \subset \Omega$.

Если имеются два подмножества случайных событий B и C, между которыми выполняется отношение включения $B \subset C$, это значит, что событие C происходит всякий раз, когда происходит событие B. Схематично это отношение показано на рис. 1.1, *а*. Частным случаем отношения включения является отношение эквивалентности $B = C$. Это отношение имеет место, когда одновременно выполняются $B \subset C$ и $C \subset B$.

Следующим отношением является объединение подмножеств, или сумма событий $B+C$, состоящее в том, что в результате испытания произойдет хотя бы одно из событий B или C. Для рассматриваемого примера событие B является суммой трех элементарных событий:

$$B = A_2+A_4+A_6.$$

Если мы имеем два случайных подмножества $B = \{A_2;\ A_4;\ A_6\}$ и $C = \{A_4; A_5; A_6\}$ (число выпавших очков не менее четырех), то суммой этих двух событий будет случайное событие D, состоящее из следующего подмножества элементарных событий:

$$B + C = C + B = D = \{A_2; A_4; A_5; A_6\}.$$

Схематично объединение подмножеств показано на рис. 1.1, *б* заштрихованной областью.

Разностью событий B–C называют событие D, включающее множество элементарных событий, принадлежащих B, но не принадлежащих C. Событие D означает, что событие B произошло, а событие C не произошло (рис. 1.1, *в*). Для рассматриваемого примера получим, что событие D включает только один элементарный исход A_2:

$$D = B - C = \{A_2; A_4; A_6\} - \{A_4; A_5; A_6\} = A_2.$$

Из определения разности подмножеств вытекает определение противоположного события. Событие $\overline{B}$ является противоположным событию B, если выполняется соотношение $\overline{B} = \Omega - B$. Событие $\overline{B}$ состоит в том, что в результате испытания событие B не произойдет. Для рассматриваемого примера, если $B = \{A_2; A_4; A_6\}$ — выпадение четного числа очков, то событие $\overline{B} = \{A_1; A_3; A_5\}$ — выпадение нечетного числа очков (рис. 1.1, *г*).

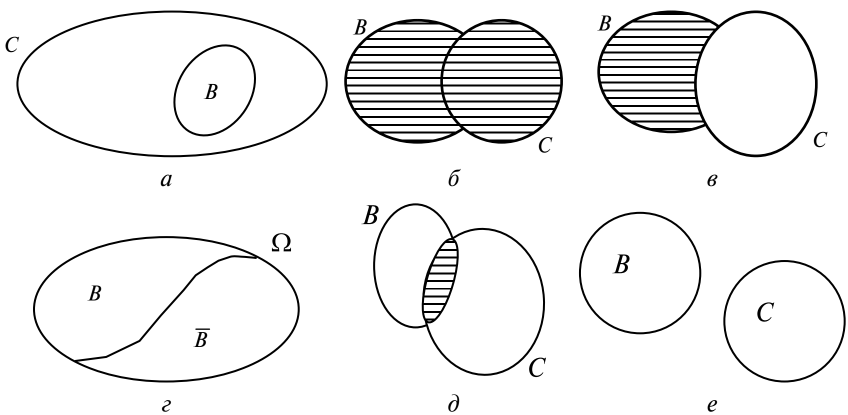


Рис. 1.1. Возможные отношения между случайными событиями: *а* — $B \subset C$; *б* — $B + C$; *в* — $B - C$; *г* — $\Omega - B$; *д* — $B \cdot C$; *е* — $B \cdot C = \varnothing$

Произведением случайных событий B и C (пересечение подмножеств) называют событие D, которое состоит в одновременном (совместном) наступлении событий B и C. Схематически событие $D = B \cdot C$

изображено заштрихованной областью на рис. 1.1, *д*. Для рассматриваемого примера с бросанием игральной кости получим:

$$D = B \cdot C = \{A_2; A_4; A_6\} \cdot \{A_4; A_5; A_6\} = \{A_4; A_6\}.$$

Если произведение двух случайных событий B и C образует пустое подмножество (рис. 1.1, *е*), то события B и C называются ***несовместными*** случайными событиями:

$$B \cdot C = \varnothing,$$

где символ $\varnothing$ обозначает пустое подмножество.

Из рисунка 1.1, *г* следует, что произведение события B на противоположное $\bar{B}$ также образует пустое подмножество: $B \cdot \bar{B} = \varnothing$.

1.2. Вероятность случайного события

Вероятность случайного события является количественной мерой степени случайности того или иного события. Наиболее продуктивными определениями вероятности оказались определения, основанные на понятии «равновозможности» всех исходов опыта (классическое определение) и на понятии частоты появления события в большом количестве испытаний (статистическое определение).

Два или несколько событий называются ***равновозможными***, если одинаков комплекс условий для их появления и нет никаких объективных причин считать, что в данных условиях какое-либо из этих событий имеет больше шансов появиться, чем другое.

Например, если игральная кость изготовлена из однородного материала, имеет форму правильного шестигранника и при подбрасывании падает на гладкую твердую горизонтальную поверхность, то можно с уверенностью утверждать, что все исходы данного опыта A_1, A_2, A_3, A_4, A_5, A_6 будут равновозможными.

1.2.1. Классическое определение вероятности

Пусть пространство элементарных событий состоит из конечного числа равновозможных попарно несовместимых исходов A_1, A_2, ..., A_i, ..., A_n. Нас интересует событие B, которое наступает при появлении определенного сочетания элементарных событий A_i. Предположим, что число сочетаний A_i, удовлетворяющих событию B, равно m (благоприятствующие исходы), а общее число исходов равно n. Вероятность наступления события B, в соответствии с классическим определением вероятности, можно определить по формуле

$$P(B)=\sum_{A_i\subset B}P(A_i)=\frac{m}{n}. \qquad (1.1)$$

Таким образом, ***вероятностью случайного события B*** называется отношение числа m исходов, благоприятствующих событию B, к общему числу n всех равновозможных исходов испытания, образующих полную группу несовместных событий.

Пример 1.1. Бросают две игральные кости. Найти вероятность того, что сумма выпавших очков на двух костях будет равна 5.

Решение. Обозначим интересующее нас событие через B. Определим общее число возможных исходов опыта. Так как при бросании одной игральной кости возможны только 6 исходов опыта, то при бросании двух игральных костей число возможных исходов опыта равно $n = 6 \cdot 6 = 36$. Число исходов опыта, благоприятствующих событию B, будет равно $m = 4$.

В соответствии с формулой (1.1) вероятность события B будет равна:

$$B=\{(1+4);\ (2+3);\ (3+2);\ (4+1)\}.$$

$$P(B)=\frac{m}{n}=\frac{4}{36}=\frac{1}{9}.$$

Из формулы (1.1) и рассмотренного примера можно отметить предельные значения вероятности событий.

Предположим, нас интересует событие B_1, состоящее в том, что при бросании двух игральных костей сумма выпавших очков будет равна 1. Число исходов опыта, благоприятствующих событию B_1, будет равно $m = 0$, так как на двух костях сумма очков всегда будет больше или равна 2. Тогда получим:

$$P(B_1)=\frac{m}{n}=\frac{0}{36}=0.$$

Предположим, нас интересует событие B_2, состоящее в том, что при бросании двух костей сумма выпавших очков будет меньше или равна 12. Число исходов опыта, благоприятствующих событию B_2, будет равно $m = 36$, так при любом из $n = 36$ возможных исходов сумма очков на двух игральных костях будет меньше или равна 12.

В этом случае получим:

$$P(B_2)=\frac{m}{n}=\frac{36}{36}=1.$$

При любом проведении опыта число m исходов, благоприятствующих появлению некоторого случайного события B, может принимать значения

$$0 \le m \le n,$$

поэтому вероятность наступления события может принимать значения в интервале от нуля до единицы

$$0 \le P(B) \le 1.$$

1.2.2. Статистическое определение вероятности

Данное определение вероятности основывается на результатах измерения частоты появления случайного события при многократном повторении опыта.

Вероятностью случайного события A называется некоторое положительное число $P(A)$, около которого группируется относительная частота этого события при возрастании числа испытаний.

Рассмотрим данное определение на примере бросания игральной кости. Будем интересоваться событием A_2, состоящим в том, что на игральной кости выпадет два очка.

Обозначим через n число повторений опыта, а через m — частоту появления события A_2 при n-кратном повторении опыта. В таблице 1.1 в первой и второй строке приведены возможные значения n и m соответственно.

Таблица 1.1

Результаты опыта

n	10	20	30	40	50	60	70	80	90	100
m	2	3	5	7	8	10	12	13	15	17
m/n	0,2	0,15	0,16(6)	0,175	0,16	0,16(6)	0,171	0,1625	0,16(6)	0,17
$m/n - P(A_2)$	0,03	–0,017	0	0,008	–0,007	0	0,004	–0,004	0	0,003

В третьей строке приведены значения относительной частоты m/n появления события A_2, из которых следует, что при увеличении числа опытов n относительная частота m/n все плотнее группируется относительно величины $P(A_2) = 1/6$ равной вероятности появления события A_2.

Таким образом, упрощенно можно записать:

$$P(A) = \lim_{n \to \infty} \frac{m}{n}. \tag{1.2}$$

Эта модель обычно используется для приближенного вычисления вероятности по результатам повторных испытаний.

Следует однако отметить, что характер приближения относительной частоты m/n к вероятности $P(A)$ при увеличении числа испытаний n несколько отличается от «стремления к пределу» в том понимании, о котором идет речь в математическом анализе. Дело в том, что при достаточно большом n ($n \to \infty$) возможны случаи, когда относительная частота будет иметь значительные отклонения от вероятности $P(A)$, но вероятность этих значительных отклонений будет очень маленькой. Поэтому стремление относительной частоты к $P(A)$ следует рассматривать как событие не абсолютно достоверное, а лишь практически достоверное.

1.2.3. Геометрическое определение вероятности

До сих пор мы рассматривали ситуации, когда пространство элементарных событий Ω является конечным (счетным) множеством числа возможных исходов опыта. На практике часто встречаются ситуации, когда пространство возможных событий имеет бесконечное число элементарных исходов. Так, например, время прихода транспорта с грузом, время поступления денег на счет фирмы в банке — все эти величины носят так называемый непрерывный характер (в отличие от ранее рассматриваемых дискретных), и их возможные значения целиком заполняют некоторый интервал. В этом случае простейшим примером определения вероятности при бесконечном числе равновозможных исходов является геометрический метод определения вероятности.

Пусть пространство возможных исходов опыта Ω определяется множеством точек конечной меры (длины, площади, объема). Предположим, что событие A наступает тогда, когда точка оказывается внутри некоторой области A (рис. 1.2). Если вероятность попадания точки в область A пропорциональна мере этой области, то вероятность события A определяется формулой

$$P(A) = \frac{\operatorname{mes} A}{\operatorname{mes} \Omega}, \tag{1.3}$$

где $\operatorname{mes} A$ — мера (длина, площадь, объем) области A; $\operatorname{mes} \Omega$ — мера пространства возможных исходов Ω.

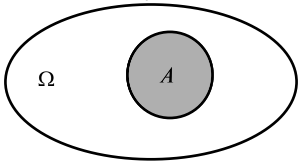


Рис. 1.2. Геометрическое определение вероятности

Пример 1.2. Два теплохода А и В могут подойти к причалу в любое время суток для разгрузки, продолжительность которой равна 6 часам. У причала может разгружаться только один теплоход. Найти вероятность того, что ни одному из них не придется ждать освобождения причала, если время подхода для каждого теплохода равновероятно в течение суток.

Решение. Обозначим через x и y время прихода первого и второго теплоходов соответственно. Так как время прихода случайно в течение суток, то $0 \le x \le 24$ и $0 \le y \le 24$. На плоскости $x0y$ произвольная точка $(x; y)$ своими координатами указывает время прибытия каждого теплохода, поэтому пространство элементарных событий в данном случае — это множество точек квадрата со стороной 24 (рис. 1.3):

$$\Omega = \{(x;\ y) / 0 \le x \le 24;\ 0 \le y \le 24\}.$$

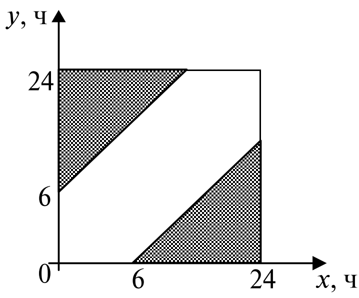


Рис. 1.3. Пространство электронных событий

Если $|x - y| > 6$, то теплоходам не придется ждать освобождения причала. Этому событию соответствуют на рис. 1.3 точки заштрихованной области. Отсюда искомая вероятность равна отношению площади заштрихованной части квадрата к площади всего квадрата:

$$P(|x - y| > 6) = 18^2 / 24^2 = (18/24)^2 = 0{,}5625.$$

1.3. Основные теоремы теории вероятностей

1.3.1. Теорема умножения вероятностей

Прежде чем сформулировать теорему умножения вероятностей, введем очень важное понятие, связанное с зависимостью случайных событий.

Два события называются ***независимыми***, если в ходе эксперимента появление или непоявление одного из них не влияет на появление или непоявление другого события. В противном случае данные события называются ***зависимыми***.

Рассмотрим зависимые и независимые случайные события на простом примере.

Пример 1.3. В черном ящике находятся 16 черных и 24 белых шара. Из ящика последовательно извлекаются два шара. Найти вероятность того, что второй извлеченный шар окажется белым, если: первый извлеченный шар был тоже белым; первый извлеченный шар был черным.

Задачу решим для двух случаев.

После извлечения первого шара фиксируется его цвет, шар возвращается в черный ящик, и шары перемешиваются.

Первый шар после извлечения в урну не возвращается.

Обозначим через A_1 и A_2 события, заключающиеся в том, что соответственно первый или второй шар оказались белыми.

Решение 1. Так как для любого шара извлечения из ящика являются равновозможными событиями, воспользуемся классическим определением вероятностей (1.1). Всего в ящике 40 шаров, значит, число возможных исходов опыта равно $n = 40$. В ящике 24 белых шара, значит число возможных исходов, благоприятствующих событию A_1, равно $m_б = 24$. Вероятность события A_1 будет равна

$$P(A_1) = \frac{m_б}{n} = \frac{24}{40} = 0{,}6.$$

Вероятность извлечения первым черного шара (событие $\bar{A}_1$) будет равна

$$P(\bar{A}_1) = \frac{m_ч}{n} = \frac{16}{40} = 0{,}4.$$

Так как после извлечения первый шар возвращается в ящик, то условия проведения опыта при извлечении второго шара не изменились. В ящике осталось 16 черных и 24 белых шара. Из этого следует, что вероятность извлечения второго белого шара будет равна

$$P(A_2) = \frac{m_б}{n} = \frac{24}{40} = 0{,}6,$$

т.е. вероятность события A_2 (второй шар белый) не зависит от того, какой шар был извлечен первым — белый A_1 или черный $\bar{A}_1$.

Таким образом, при данном опыте события A_1 и A_2 являются независимыми.

*Решение 2. Р*ешим задачу для случая, когда первый извлеченный шар в ящик не возвращается.

Определим вероятность события A_2 при условии, что первый извлеченный шар был белым (событие A_1). Эту условную вероятность обозначают $P(A_2/A_1)$. Если первый шар был белым, то перед вытаскиванием второго шара в ящике было всего 39 шаров (n = 39), из них 23 шара белых ($m_б$ = 23). Тогда условная вероятность события A_2 при условии A_1 будет равна

$$P(A_2 / A_1) = \frac{m_б}{n} = \frac{23}{39} \approx 0{,}5897.$$

Если первый извлеченный шар был черным (событие $\bar{A}_1$), то перед вытаскиванием второго шара в ящике было всего 39 шаров (n = 39), из них 24 шара белых ($m_б$ = 24). Тогда условная вероятность вытащить второй белый шар A_2 при условии, что первым был вынут черный шар $\bar{A}_1$, равна

$$P(A_2 / \bar{A}_1) = \frac{m_б}{n} = \frac{24}{39} \approx 0{,}6154.$$

Сравнивая значения $P(A_2/A_1)$ и $P(A_2 / \bar{A}_1)$, мы видим, что при данном проведении опыта вероятность события A_2 зависит от того, произошло событие A_1 или не произошло $\bar{A}_1$. В данном случае события A_2 и A_1 являются зависимыми.

Теорема умножения вероятностей. Вероятность произведения двух событий $A \cdot B$ равна произведению вероятности одного их них (A или B) на условную вероятность другого события при условии, что первое событие уже произошло:

$$P(A \cdot B) = P(A) \cdot P(B / A) = P(B) \cdot P(A / B). \tag{1.4}$$

Доказательство. Воспользуемся классическим определением вероятности. Пусть число возможных исходов некоторого опыта равно n

(рис. 1.4). Допустим, что событию A благоприятствуют m исходов, событию B благоприятствуют k исходов, а произведению событий $A \cdot B$ благоприятствуют r исходов.

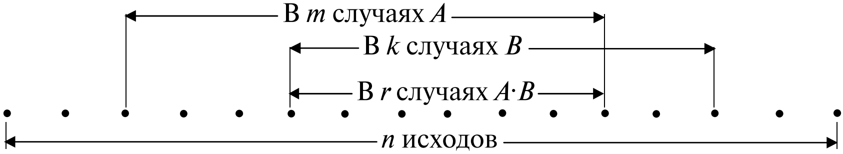


Рис. 1.4. К определению вероятности произведения событий

Для вероятности произведения событий $A \cdot B$ общее число исходов равно n, а число благоприятствующих исходов равно r, значит

$$P(A \cdot B) = \frac{r}{n}. \tag{1.5}$$

Умножим и разделим дробь (1.5) на m:

$$P(A \cdot B) = \frac{r \cdot m}{n \cdot m} = \frac{m}{n} \cdot \frac{r}{m}.$$

Отношение m/n равно $P(A)$ безусловной вероятности события A. Если событие A уже произошло, то для события B общее число возможных исходов будет равно m, из них только r случаев благоприятствуют событию B. Значит, отношение r/m равно условной вероятности события B при условии, что A уже произошло $P(B/A)$. Таким образом

$$P(A \cdot B) = P(A) \cdot P(B/A). \tag{1.6}$$

Умножим и разделим дробь (1.5) на k:

$$P(A \cdot B) = \frac{r \cdot k}{n \cdot k} = \frac{k}{n} \cdot \frac{r}{k}.$$

Отношение k/n равно $P(B)$ безусловной вероятности события B. Если событие B уже произошло, то для события A общее число возможных исходов равно k, из них только r случаев благоприятствуют событию A. Таким образом, условная вероятность $P(A/B)$ равна отношению r/k. Значит

$$P(A \cdot B) = P(B) \cdot P(A/B). \tag{1.7}$$

Из формул (1.6) и (1.7) слудует, что если известна вероятность произведения двух событий $A \cdot B$ и безусловные вероятности этих событий, то условные вероятности могут быть определены по формулам:

$$P(A/B) = \frac{P(A \cdot B)}{P(B)}; \qquad P(B/A) = \frac{P(A \cdot B)}{P(A)}. \tag{1.8}$$

Если события A и B независимы, то условные вероятности этих событий равны безусловным (см. пример 1.3):

$$P(A/B) = P(A);\ P(B/A) = P(B).$$

Тогда теорема умножения вероятностей двух независимых событий запишется в виде:

$$P(A \cdot B) = P(A) \cdot P(B). \tag{1.9}$$

Для трех случайных событий A, B, C формула (1.6) может быть записана в виде:

$$P(A \cdot B \cdot C) = P(A) \cdot P(B/A) \cdot P(C/A, B), \tag{1.10}$$

где $P(C/A, B)$ — условная вероятность события C при условии, что события A и B произошли.

Если события A, B, C независимы, то формула произведения вероятности по аналогии с (1.9) будет иметь вид:

$$P(A \cdot B \cdot C) = P(A) \cdot P(B) \cdot P(C). \tag{1.11}$$

На практике наиболее часто встречаются ситуации, когда случайные события являются независимыми.

Пример 1.4. Для предоставления кредитов трем клиентам банка выделена сумма S. Сумма кредита, запрашиваемая первым, вторым и третьим клиентом, соответственно равна $0{,}5S$, $0{,}4S$ и $0{,}3S$. Вероятность обращения за кредитом для первого, второго и третьего клиента соответственно равна 0,5, 0,6, 0,7. Определить вероятность того, что выделенная сумма S окажется недостаточной для предоставления кредита обратившимся в банк клиентам.

Решение. Обозначим через A_1, A_2, A_3 случайные события, состоящие в том, что за кредитом в банк обратился соответственно первый, второй и третий клиент. В соответствии с условием задачи вероятности этих событий равны: $P(A_1) = 0{,}5$, $P(A_2) = 0{,}6$, $P(A_3) = 0{,}7$.

Исходя из сумм кредита, запрашиваемых клиентами, выделенная под кредит сумма S окажется недостаточной, если в банк за кредитом обратятся одновременно три клиента:

$$0{,}5S + 0{,}4S + 0{,}3S > S.$$

Обозначим через B событие, состоящее в том, что в банк за кредитом обратились одновременно три клиента. Событие B является произведением событий A_1, A_2 и A_3:

$$B = A_1 \cdot A_2 \cdot A_3.$$

Так как события A_1, A_2, и A_3 являются попарно независимыми (клиенты действуют независимо друг от друга), то вероятность события B определяется по формуле произведения вероятностей (1.11):

$$P(B) = P(A_1) \cdot P(A_2) \cdot P(A_3) = 0{,}5 \cdot 0{,}6 \cdot 0{,}7 = 0{,}2.$$

Если события A и B несовместны (см. рис. 1.1, *е*), то вероятность произведения этих событий равна нулю. В формуле (1.4) если событие A произошло, то $P(B/A) = 0$, или если событие B произошло, то $P(A/B) = 0$.

1.3.2. Теорема сложения вероятностей

Теорема. Вероятность суммы двух событий $(A + B)$ равна сумме вероятностей этих событий минус вероятность произведения этих событий:

$$P(A+B) = P(A) + P(B) - P(A{\cdot}B). \tag{1.12}$$

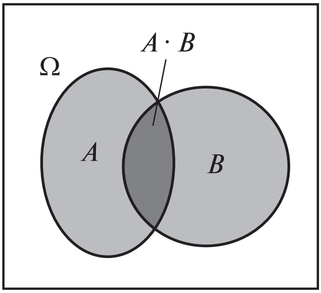


Рис. 1.5. К определению вероятности суммы событий

Доказательство. Для доказательства воспользуемся геометрическим определением вероятностей.

Предположим, что множество всех возможных исходов опыта определяется площадью квадрата S_Ω. Событию A удовлетворяет область, определяющаяся площадью овала S_A. Событию B удовлетворяет область, определяющаяся площадью овала S_B. Событию $(A+B)$ удовлетворяет область, определяющаяся площадью фигуры S_{A+B}, полученной объединением овалов A и B (обведена на рис. 1.5 жирной линией).

В соответствии с формулой (1.3) для вероятности суммы событий $A + B$ запишем:

$$P(A+B) = \frac{S_{A+B}}{\Omega}. \tag{1.13}$$

Площадь S_{A+B} можно записать в виде:

$$S_{A+B} = S_A + S_B - S_{A\cdot B},$$

где $S_{A\cdot B}$ — площадь фигуры, образующейся наложением двух овалов.

После подстановки данной формулы в (1.13) получим

$$P(A+B) = \frac{S_A}{S_\Omega} + \frac{S_B}{S_\Omega} - \frac{S_{A\cdot B}}{S_\Omega}.$$

В соответствии с формулой (1.3) первое слагаемое S_A/S_Ω есть вероятность $P(A)$, второе слагаемое S_B/S_Ω есть вероятность $P(B)$, а третье слагаемое (со знаком минус) $S_{A\cdot B}/S_\Omega$ есть вероятность произведения событий A и B — $P(A\cdot B)$. Таким образом, для суммы вероятностей двух событий мы приходим к формуле (1.12).

Для несовместных случайных событий A и B (см. рис. 1.1, *е*) произведение событий $A\cdot B = \varnothing$ образует пустое подмножество, и вероятность произведения этих событий равна

$$P(A\cdot B) = 0.$$

Поэтому для несовместных случайных событий A и B вероятность суммы этих событий равна сумме их вероятностей:

$$P(A+B) = P(A) + P(B). \tag{1.14}$$

Для случая n совместных случайных событий $A_1, A_2, \ldots, A_i, \ldots, A_n$ формула (1.12) преобразуется к виду

$$\begin{gathered} P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i=j=1;i\neq j}^{n} P(A_i)P(A_j) + \\ + \sum_{i,j,k=1,i\neq j\neq k}^{n} P(A_i)P(A_j)P(A_k) - \ldots + (-1)^{n-1} P(A_1 \cdot A_2 \ldots A_n). \end{gathered} \tag{1.15}$$

Если события A_i являются несовместными событиями, формула (1.15) значительно упрощается:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)\,. \tag{1.16}$$

Теорема сложения вероятностей для несовместных случайных событий имеет одно важное следствие.

A_1	A_2	A_3	A_4	A_5
A_6	A_7	...	...	...
...	A_i	A_{i+1}	...	...
...	...	...	A_{n-1}	A_n

Рис. 1.6. Полная группа событий A_i, $i = 1 - n$

Если несовместные случайные события A_i, $i = 1 \div n$ образуют полную группу событий (перекрывают всю область возможных исходов опыта Ω (рис. 1.6), то сумма вероятностей этих событий равна единице:

$$\sum_{i=1}^{n} P(A_i) = 1. \tag{1.17}$$

Доказательство. Для доказательства воспользуемся геометрическим определением вероятности (1.3):

$$\sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{S_{A_i}}{S_\Omega} = \frac{S_\Omega}{S_\Omega} = 1.$$

Условие (1.17) в теории вероятности часто называют условием нормировки.

Пример 1.5. Для задачи, приведенной в примере 1.4, определить вероятность того, что клиенты, обратившиеся в банк, получат запрашиваемый кредит.

Решение. Обозначим следующие случайные события:

B_0 — ни один клиент не обратится в банк за кредитом;

B_1 — один клиент обратится в банк за кредитом;

B_2 — два клиента обратятся в банк за кредитом;

B_3 — три клиента обратятся в банк за кредитом.

События B_0, B_1, B_2, B_3 можно выразить через события A_1, A_2, A_3 и противоположные события $\bar{A}_1, \bar{A}_2, \bar{A}_3$:

$$B_0 = \bar{A}_1 \cdot \bar{A}_2 \cdot \bar{A}_3\,;$$

$$B_1 = A_1\bar{A}_2\bar{A}_3 + \bar{A}_1 A_2 \bar{A}_3 + \bar{A}_1\bar{A}_2 A_3\,;$$

$$B_2 = A_1 A_2 \overline{A}_3 + A_1 \overline{A}_2 A_3 + \overline{A}_1 A_2 A_3;$$

$$B_3 = A_1 A_2 A_3.$$

Для вероятностей противоположных событий $\overline{A}_i$ получим $P(\overline{A}_i) = 1 - P(A_i)$:

$$P(\overline{A}_1) = 1 - 0{,}5 = 0{,}5;$$

$$P(\overline{A}_2) = 1 - 0{,}6 = 0{,}4;$$

$$P(\overline{A}_3) = 1 - 0{,}7 = 0{,}3.$$

Используя формулу умножения вероятностей (1.11) и формулу сложения вероятностей (1.16), для вероятностей событий B_i получим:

$$P(B_0) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3) = 0{,}5 \cdot 0{,}4 \cdot 0{,}3 = 0{,}06;$$

$$P(B_1) = P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3) =$$
$$= 0{,}5 \cdot 0{,}4 \cdot 0{,}3 + 0{,}5 \cdot 0{,}6 \cdot 0{,}3 + 0{,}5 \cdot 0{,}4 \cdot 0{,}7 = 0{,}29;$$

$$P(B_2) = P(A_1)P(A_2)P(\overline{A}_3) + P(A_1)P(\overline{A}_2)P(A_3) + P(\overline{A}_1)P(A_2)P(A_3) =$$
$$= 0{,}5 \cdot 0{,}6 \cdot 0{,}3 + 0{,}5 \cdot 0{,}4 \cdot 0{,}7 + 0{,}5 \cdot 0{,}6 \cdot 0{,}7 = 0{,}44;$$

$P(B_3) = 0{,}21$ (определено в примере 1.4).

События B_i составляют полную группу несовместных событий, и для них выполняется условие (1.17).

В условии задачи требуется определить вероятность события C, состоящего в том, что клиенты, обратившиеся в банк, получат запрашиваемый кредит. Значит, событие C наступит тогда, когда наступит событие B_1 или событие B_2, т.е. $C = B_1 + B_2$.

Таким образом, по формуле сложения вероятностей получим:

$$P(C) = P(B_1) + P(B_2) = 0{,}29 + 0{,}44 = 0{,}73.$$

1.3.3. Формула полной вероятности

Предположим, что некоторое событие A может произойти лишь с одним из n несовместных событий $H_1, H_2, ..., H_i, ..., H_n$, составляющих полную группу событий. Их обозначение буквой H обусловлено названием этих событий — гипотезы.

Допустим также, что известны вероятности наступления этих гипотез $P(H_1), P(H_2), ..., P(H_n)$. Кроме того, известно, что событие A будет зависимым от гипотез $H_1, H_2, ..., H_i, ..., H_n$, условные вероятности появления события A $P(A/H_1)$, $P(A/H_2)$, ..., $P(A/H_i)$, ..., $P(A/H_n)$ также известны.

Теорема. Если событие A в некотором испытании может произойти лишь с одной из гипотез H_i, $i = 1 \div n$, то безусловная вероятность наступления события A в этом испытании определяется по формуле.

$$P(A) = \sum_{i=1}^{n} P(H_i) P(A / H_i). \qquad (1.18)$$

Доказательство. Так как событие A происходит совместно с одной из гипотез H_i, $i = 1 \div n$, образующих полную группу событий, то для события A можно записать:

$$A = A \cdot H_1 + A \cdot H_2 + ... + A \cdot H_i + ... + A \cdot H_n.$$

Используя формулу умножения вероятностей для зависимых событий, для вероятностей произведения событий $A \cdot H_i$ получим:

$$P(A \cdot H_i) = P(H_i)\ P(A/H_i).$$

Так как события $AH_1, AH_2, ..., AH_i, ..., AH_n$ являются несовместными, то по формуле сложения вероятностей (1.16) для вероятности $P(A)$ получим формулу (1.18):

Пример 1.6. Для подготовки к установленному сроку бухгалтерской отчетности предприятия к выполнению этой работы могут привлечены один, два или три работника. Вероятности этих событий соответственно равны 0,5, 0,3 и 0,2. Вероятность подготовки бухгалтерской отчетности в установленный срок при привлечении одного работника равна 0,3, при привлечении двух работников — 0,6, при привлечении трех работников — 0,95. Определить вероятность подготовки бухгалтерской отчетности в установленный срок.

Решение. Обозначим через A событие, состоящее в подготовке бухгалтерской отчетности в установленный срок. Буквами H_1, H_2, H_3 обозначим события, состоящие в привлечении к работе соответственно одного, двух или трех работников, и назовем эти события гипотезами. По условию задачи вероятности гипотез равны:

$$P(H_1) = 0,5;\ P(H_2) = 0,3;\ P(H_3) = 0,2.$$

Видно, что гипотезы образуют полную группу событий:

$$\sum_{i=1}^{3} P(H_i) = 0,5 + 0,3 + 0,2 = 1.$$

Событие A зависит от гипотез H_i, и по условию задачи условные вероятности наступления события A при соответствующих гипотезах равны:

$$P(A/H_1) = 0{,}3;\ P(A/H_2) = 0{,}6;\ P(A/H_3) = 0{,}95.$$

В соответствии с формулой полной вероятности (1.18) для безусловной вероятности наступления события A получим:

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i) = 0{,}5\cdot 0{,}3 + 0{,}3\cdot 0{,}6 + 0{,}2\cdot 0{,}95 = 0{,}52\ .$$

1.3.4. Теорема (формула) Байеса

Условия применимости формулы Байеса формулируются в определенной мере аналогично условиям применимости формулы полной вероятности.

Пусть проводится испытание, в результате которого событие A может произойти лишь с одной из несовместных гипотез (событий) H_i, $i = 1 \div n$. Гипотезы образуют полную группу событий, и вероятности их наступления $P(H_i)$ известны. Известны также условные вероятности наступления события A при условии наступления соответствующих гипотез $P(A/H_i)$.

Далее формулировка задачи отличается от формулировки условий применения формулы полной вероятности.

Предположим, что в результате испытания событие A уже произошло, но с какой из гипотез — неизвестно. Необходимо определить, каковы будут вероятности гипотез H_i при условии, что событие A уже произошло. То есть нам необходимо определить условные вероятности $P(H_i/A)$.

Теорема. Пусть случайное событие A может произойти лишь с одним из несовместных событий (гипотез) H_i, $i = 1 \div n$, составляющих полную группу событий, и в результате испытания событие A уже произошло. Тогда вероятность того, что событие A произошло совместно с гипотезой H_i, определяется соотношением

$$P(H_i / A) = \frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}\ . \qquad (1.19)$$

Доказательство. События A и H_i являются зависимыми. По формуле умножения вероятностей (1.4) для зависимых событий можно записать вероятность произведения событий A и H_i:

$$P(A \cdot H_i) = P(A)\ P(H_i/A) = P(H_i)\ P(A/H_i).$$

Используя вторую часть равенства для условной вероятности $P(H_i/A)$, получим

$$P(H_i / A) = \frac{P(H_i) P(A / H_i)}{P(A)}. \quad (1.20)$$

В формуле неизвестной является только вероятность $P(A)$, которую можно определить по формуле полной вероятности (1.18). После подстановки формулы (1.18) в (1.20) мы получим формулу Байеса (1.19).

Формула Байеса имеет очень важное значение для принятия оптимальных управленческих решений.

Пример 1.7. Результаты финансово-экономического анализа деятельности обанкротившихся банков в период кризиса финансовой системы показали следующее. Причиной банкротства банков являлось принятие рискованных управленческих решений H_1, H_2, H_3. Вероятности принятия этих решений по анализируемым банкам соответственно равны $P(H_1) = 0,3$, $P(H_2) = 0,5$, $P(H_3) = 0,2$. Вероятности банкротства банков (событие A) при принятии первого H_1, второго H_2 и третьего H_3 управленческих решений соответственно равны:

$$P(A/H_1) = 0,7;\ P(A/H_2) = 0,8;\ P(A/H_3) = 0,9.$$

Определить, какое из принимаемых управленческих решений явилось наиболее вероятной причиной банкротства банков.

Решение. Так как банкротство банков (событие A) уже произошло, по условию задачи понятно, что для ее решения необходимо использовать формулу Байеса (1.19). Для ответа на сформулированный в задаче вопрос определим условные вероятности $P(H_i/A)$:

$$P(H_1/A) = \frac{P(H_1) P(A / H_1)}{\sum_{i=1}^{3} P(H_i) P(A / H_i)} = \frac{0,7 \cdot 0,3}{0,7 \cdot 0,3 + 0,5 \cdot 0,8 + 0,2 \cdot 0,9} = 0,2658;$$

$$P(H_2/A) = \frac{P(H_2) P(A / H_2)}{\sum_{i=1}^{3} P(H_i) P(A / H_i)} = \frac{0,5 \cdot 0,8}{0,7 \cdot 0,3 + 0,5 \cdot 0,8 + 0,2 \cdot 0,9} = 0,5063;$$

$$P(H_3/A) = \frac{P(H_3) P(A / H_3)}{\sum_{i=1}^{3} P(H_i) P(A / H_i)} = \frac{0,2 \cdot 0,9}{0,7 \cdot 0,3 + 0,5 \cdot 0,8 + 0,2 \cdot 0,9} = 0,2279.$$

Из сравнения полученных значений условных вероятностей делаем вывод, что наиболее вероятной причиной банкротства банков явилось принятие ими второго управленческого решения.

Если судить по условным вероятностям $P(A/H_i)$, то наиболее рискованным является принятие третьего управленческого решения. Но этот вывод делается без учета вероятностей принятия решений $P(H_i)$.

Наиболее вероятной причиной банкротства банка, несмотря на то что $P(A/H_2) < P(A/H_3)$, стала гипотеза H_2 в связи с тем, что вероятность принятия этого решения H_3 в банках была наибольшей.

1.3.5. Теорема о повторных независимых испытаниях (формула Бернулли)

Схема повторных испытаний является базовой моделью для решения многих практических задач, связанных с контролем качества продукции, оценкой надежности, с организацией работы систем массового обслуживания и т.п.

В общем виде схему повторных испытаний можно сформулировать следующим образом.

Производится n однотипных независимых испытаний, в каждом из которых может появиться интересующее нас событие A с вероятностью $P(A) = p$. Требуется определить вероятность того, что в n испытаниях событие A появится ровно m раз.

Для независимых испытаний вероятность появления события A в каждом опыте одинакова и в каждом опыте возможны два исхода:

A — появление события A с вероятностью p;

$\overline{A}$ — непоявление события A с вероятностью $q = 1 - p$.

Теорема. Пусть вероятность появления события A в каждом опыте постоянна и равна p. Тогда вероятность того, что в n независимых испытаниях событие A появится ровно m раз, определяется формулой

$$P_n(m) = C_n^m\, p^m q^{n-m}, \tag{1.21}$$

где $C_n^m = \dfrac{n!}{m!(n-m)!}$ — число сочетаний из n по m; $n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot (n-1) \cdot n$.

Доказательство. Обозначим через B_i возможные события, когда в n испытаниях событие A появится m раз:

$$B_1 = A_1 \cdot A_2 \cdot A_3 \cdot \ldots \cdot A_m \cdot \overline{A}_{m+1} \cdot \overline{A}_{m+2} \cdot \ldots \cdot \overline{A}_n. \tag{1.22}$$

Событие B_1 состоит в том, что при n-кратном проведении опыта в первых m испытаниях событие A появилось, а в последующих испытаниях событие A не появилось.

Возможны и другие события B_i, в которых событие A появилось m раз и не появилось $(n - m)$ раз, но номера испытаний, в которых появилось событие A, для различных B_i будут различными. Все возможные события B_i можно записать, если в произведении (1.22) осуществлять перестановку номеров испытаний, в которых событие A произошло.

Общее число таких перестановок, т.е. число событий B_i, будет равно числу сочетаний из n по m C_n^m. Так как каждое событие B_i состоит в том, что в n испытаниях событие A появилось m раз, а в $(n - m)$ испытаниях появилось противоположное событие $\bar{A}$, то вероятность всех этих событий B_i определяется по формуле умножения вероятностей и будет равна

$$P(B_i) = \prod_{j \subset m} P(A_j) \cdot \prod_{i \subset n-m} P(\bar{A}_i) = p^m q^{n-m} .$$

Событие C, состоящее в том, что в результате n испытаний интересующее нас событие A появится m раз, является суммой несовместных событий B_i:

$$C = \sum_{i=1}^{C\frac{m}{n}} B_i .$$

В соответствии с формулой сложения вероятностей для несовместных событий получим

$$P(C) = P_n(m) = \sum_{i=1}^{C\frac{m}{n}} p^m q^{n-m} = C_n^m p^m q^{n-m} .$$

Пример 1.8. Вероятность выпуска бракованного изделия на станке равна 0,2. Определить вероятность того, что из десяти выпущенных на данном станке деталей m деталей не будут бракованными (задачу решить для $0 \le m \le 10$).

Решение. Исходя из условия задачи, нас интересует событие A — выпуск небракованных деталей и нужно определить вероятность того, что это событие появится m раз в пяти испытаниях. В условии задачи дана вероятность появления события $\bar{A}$ — выпуск бракованной детали. Таким образом, в соответствии с условием задачи имеем $n = 10$, $q = 0{,}2$, $p = 0{,}8$. Определяем вероятность того, что все детали в партии бракованные $m = 0$:

$$P_{10}(0) = C_{10}^0\, p^0 q^{10} = \frac{10!}{0!10!} 0,8^0 \cdot 0,2^{10} \approx 10^{-7}.$$

Определяем вероятность того, что в партии только одна деталь небракованная:

$$P_{10}(1) = C_{10}^1\, p^1 q^9 = \frac{10!}{1!9!} 0,8 \cdot 0,2^9 = 2 \cdot 10^{-5}.$$

Вероятность того, что в партии две детали небракованные, получим

$$P_{10}(2) = C_{10}^2\, p^2 q^8 = \frac{10!}{2!8!} 0,8^2 \cdot 0,2^8 = 7,2 \cdot 10^{-5}.$$

Аналогичным образом рассчитаем вероятности $P_{10}(m)$ при $3 \le m \le 10$ и результаты вычислений сведем в табл. 1.2.

Таблица 1.2

Результаты вычислений

m	0	1	2	3	4	5	6	7	8	9	10
$P_{10}(m)$	10^{-7}	$2\times$ $\times 10^{-5}$	$7,2\times$ $\times 10^{-5}$	$7,8\times$ $\times 10^{-4}$	$5,5\times$ $\times 10^{-3}$	0,0264	0,0881	0,2013	0,302	0,2684	0,1074

Из приведенных расчетов плнятно, что при увеличении m от нуля до 8 вероятности $P_{10}(m)$ возрастают, а затем убывают.

Наибольшее значение вероятности получается при $m = 8$. Такое число появления события A в n испытаниях, которому соответствует наибольшая вероятность, называется модой или наивероятнейшим числом и обозначается m_0.

Для определения наивероятнейшего числа m_0 не обязательно вычислять все значения вероятностей $P_n(m)$, а достаточно знать число испытаний n и вероятность p появления события A в одном испытании. Значение m_0 может быть найдено из решения системы неравенств:

$$\begin{cases} P_n(m_0) \ge P_n(m_0 - 1); \\ P_n(m_0) \ge P_n(m_0 + 1). \end{cases}$$

Подставляя значения $P_n(m_0)$ и $P_n(m_0 - 1)$ в первое неравенство, получим:

$$\frac{n!}{m_0!(n - m_0)!} p^{m_0} q^{n - m_0} \ge \frac{n!}{(m_0 - 1)!(n - m_0 + 1)!} p^{m_0 - 1} q^{n - m_0 + 1}.$$

После сокращений данное неравенство преобразуется к виду:

$$\frac{p}{m_0} \geq \frac{q}{n-m_0+1},$$

или

$$np - m_0p + p \geq m_0q.$$

Решая данное неравенство относительно m_0 с учетом условия $p + q = 1$, получим $m_0 \leq np + p$.

Подставляя значение $P_n(m_0)$ и $P_n(m_0 + 1)$ во второе неравенство и проведя аналогичные преобразования, мы получим второе неравенство $m_0 \geq np - q$. Объединяя два полученных неравенства, получим, что наивероятнейшее число испытаний, в которых появляется событие A, лежит в интервале

$$np - q \leq m_0 \leq np + p. \qquad (1.23)$$

Для примера 1.8 по формуле (1.23) для наивероятнейшего числа небракованных изделий получим:

$$10 \cdot 0{,}8 - 0{,}2 \leq m_0 \leq 10 \cdot 0{,}8 + 0{,}8;$$
$$7{,}8 \leq m_0 \leq 8{,}8,$$

т.е. мода, или наивероятнейшее число небракованных изделий, будет равно 8, что совпадает с результатами табл. 1.2.

1.3.6. Теорема (формула) Пуассона

Формула Бернулли (1.21) является единственной точной формулой в схеме повторных независимых испытаний. Однако использовать ее можно в весьма ограниченных условиях, когда число испытаний n незначительно ($n \leq 15$), а вероятность p существенно отличается от 0 или 1.

На практике же мы часто встречаемся с ситуациями, когда число испытаний n исчисляется сотнями и тысячами, а вероятность p очень мала (или весьма близка к 1). Это так называемые «редкие события», особенно часто встречающиеся в теории массового обслуживания.

Пример 1.9. При транспортировке изделие может быть повреждено с вероятностью 0,002. Найти вероятность того, что в партии из 2000 изделий в пути окажутся поврежденными три изделия.

Решение. Имеем $n = 2000$, $p = 0{,}002$, $q = 1 - 0{,}002 = 0{,}998$, $m = 3$.

Пользуясь формулой Бернулли, следовало бы записать:

$$P_{2000}(3) = C_{2000}^{3} \cdot 0{,}002^3 \cdot 0{,}998^{1997}.$$

Совершенно ясно, что эта формула здесь непригодна, так как вычислять число сочетаний C_{2000}^{3} и последующие степени — занятие весьма трудоемкое. Здесь на помощь приходит приближенная формула Пуассона, в основе которой лежит следующая теорема.

Теорема. Если вероятность p наступления события A в каждом испытании стремится к нулю ($p \to 0$) и при увеличении числа испытаний произведение np стремится к постоянному числу $\lambda = \lim\limits_{n\to\infty} np$, то вероятность $P_n(m)$ того, что в n независимых испытаниях событие A появится m раз, удовлетворяет соотношению:

$$P_n(m) - \frac{\lambda^m e^{-\lambda}}{m!} \to 0 \text{ при } n \to \infty.$$

Данную теорему приводим без доказательства.

Таким образом, для приближенного вычисления вероятностей $P_n(m)$ при достаточно больших m и малых значениях p можно воспользоваться формулой

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda} , \tag{1.24}$$

где $\lambda = n \cdot p$ — параметр распределения Пуассона.

При этом возможно воспользоваться и специальной таблицей, в которой для каждой пары λ и m определена вероятность $P_n(m)$ (приложение 4).

В задаче из примера 1.9 $\lambda = 4$, $m = 3$. Из приложения 4 находим, что $P_{2000}(3) = 0{,}195367$. Точное значение этой вероятности равно 0,195464. Погрешность приближения незначительна — всего 0,05%, что объясняется действительно большим числом испытаний и очень малой вероятностью p. Если эти условия не соблюдены, то погрешность формулы Пуассона становится ощутимой. Так, при $n = 50$ и $p = 0{,}08$ ошибка приближения равна 2,02%, а при $n = 20$ и $p = 0{,}2$ ошибка уже равна 5,12% (табл. 1.3).

Таблица 1.3

Значения вероятностей

n	p	λ	m	Приближенное значение $P_n(m)$	Точное значение $P_n(m)$	Относительная погрешность, %
2 000	0,002	4	3	0,195367	0,195464	0,05
200	0,02	4	3	0,195367	0,196347	0,50
100	0,04	4	3	0,195367	0,197333	1,01
50	0,08	4	3	0,195367	0,199321	2,02
20	0,20	4	3	0,195367	0,205364	5,12

1.4. Примеры решения задач

Задача 1.1. Пять клиентов банка готовы вложить деньги по срочному депозитному договору на срок 1, 2 или 3 года с равными вероятностями. Определить вероятность того, что все клиенты банка вложат деньги по срочному депозитному договору: 1) на 2 года; 2) на один и тот же срок.

Решение. 1. Обозначим событие A, состоящее в том, что все пять клиентов банка заключат депозитный договор на 2 года.

Каждый клиент банка имеет три варианта заключить депозитный договор соответственно на 1, 2 или 3 года. Общее число возможных вариантов заключения депозитных договоров для пяти клиентов будет равно $N = 3^5$. Число вариантов, благоприятствующих событию A, равно $m_A = 1$. Таким образом, вероятность события A будет равна

$$P(A) = \frac{m_A}{N} = \frac{1}{3^5} = \frac{1}{243} \approx 0{,}0041.$$

2. Обозначим событие B, состоящее в том, что все клиенты банка заключают депозитный договор на один и тот же срок. Общее число возможных вариантов заключения депозитных договоров для пяти клиентов остается тем же — $N = 3^5$. Число вариантов заключения договоров, благоприятствующих событию B, будет равно $m_B = 3$ (все клиенты заключили договоры на 1 год, или на 2 года, или на 3 года). Поэтому вероятность события B будет равна

$$P(B) = \frac{m_B}{N} = \frac{3}{3^5} = \frac{1}{81} \approx 0{,}0123.$$

Задача 1.2. Кредитный портфель банка включает 100 кредитных договоров, из которых 6 с высоким кредитным риском. Кредитный портфель произвольно разделен на две равные части, одна из которых передана филиалу банка. Определить вероятность того, что договоры с высоким кредитным риском распределяются поровну в обе части кредитного портфеля.

Решение. Общее число возможных способов, какими 50 договоров можно выбрать из 100 договоров кредитного портфеля, будет равно $N = C_{100}^{50}$, где $C_n^k = \dfrac{n!}{k!(n-k)!}$.

Обозначим событие A, состоящее в том, что в каждой части кредитного портфеля по 3 кредитных договора с высоким кредитным риском.

Событию A будут благоприятствовать случаи, когда в одной части кредитного портфеля будут 47 безрисковых договоров из 94 безрисковых договоров во всем кредитном портфеле и 3 договора из 6 с высоким кредитным риском. Таким образом, число благоприятных случаев определяется произведением:

$$m_A = C_{94}^{47} \cdot C_6^3.$$

Для вероятности события A получим

$$P(A) = \frac{m_A}{N} = \frac{C_{94}^{47}}{C_{100}^{50}} = \frac{94!}{47!47!} \frac{6!}{3!3!} \frac{50!50!}{100!} =$$

$$= \frac{48 \cdot 49 \cdot 50 \cdot 48 \cdot 49 \cdot 50 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 95 \cdot 96 \cdot 97 \cdot 98 \cdot 99 \cdot 100} \cong 0{,}322.$$

Задача 1.3. Пакеты акций на рынке ценных бумаг могут дать их владельцу доход с вероятностью 0,5 (для каждого пакета акций). Сколько пакетов акций нужно приобрести, чтобы с вероятностью не менее 0,96875 можно ожидать доход хотя бы по одному пакету акций?

Решение. Обозначим событие A_i, состоящее в том, что приобретенный на рынке i-й пакет акций принесет доход. В соответствии с условием вероятность этого события равна $P(A_i) = 0{,}5$. Противоположное событие $\overline{A}_i$ состоит в том, что приобретенный i-й пакет акций не принесет доход, и вероятность этого события равна $P(\overline{A}_i) = 1 - P(A_i) = 0{,}5$. Обозначим событие $\overline{A}_N$, состоящее в том, что ни один пакет акций из приобретенных N пакетов не принесет дохода. Так как события $\overline{A}_i$ являются независимыми, то вероятность события $\overline{A}_N$ по формуле умножения вероятностей будет равна

$$P(\overline{A}_N) = \left[P(\overline{A}_i) \right]^N. \tag{1.25}$$

Вероятность $P(\overline{A}_N)$ может быть определена из условия задачи:

$$P(\overline{A}_N) = 1 - P(A_N),$$

где $P(A_N) = 0{,}96875$ — вероятность получения дохода хотя бы от одного пакета акций из N приобретенных пакетов.

С учетом данного равенства формулу (1.25) запишем в виде:

$$1 - P(A_N) = \left[1 - P(A_i) \right]^N. \tag{1.26}$$

Необходимое количество пакетов акций N, которое необходимо приобрести, можно найти, если прологарифмировать обе части уравнения (1.26):

$$\log[1-P(A_N)] = N\log[1-P(A_i)];$$

$$N \geq \frac{\lg[1-P(A_N)]}{\lg[1-P(A_i)]} = \frac{\lg 0{,}03125}{\lg 0{,}5} = \frac{\lg(0{,}5)^5}{\lg 0{,}5} = 5.$$

Для того чтобы получить доход хотя бы по одной акции с вероятностью не менее 0,96875, нужно приобрести не менее 5 пакетов акций.

Задача 1.4. Торговая организация закупает данный вид товара в трех фирмах в соотношении 2:5:3. Среди товара, поставляемого первой фирмой, стандартный товар составляет 75%, второй фирмой — 90%, третьей фирмой — 80%. Определить вероятность приобретения покупателем в данной торговой организации нестандартного товара.

Решение. Определим вероятности событий H_1, H_2 и H_3 приобретение товара соответственно у первой, второй и третьей фирм. Примем общий объем закупаемого товара равным 2 + 5 + 3 = 10. Тогда для искомых вероятностей получим:

$$P(H_1) = \frac{2}{10} = 0{,}2; \quad P(H_2) = \frac{5}{10} = 0{,}5; \quad P(H_3) = \frac{3}{10} = 0{,}3.$$

Обозначим событие A, состоящее в том, что товар стандартный.

В соответствии с условием товар, приобретенный торговой организацией у первой фирмы, будет стандартным с вероятностью $P(A/H_1) = 0{,}75$. Вероятность $P(A/H_1)$ является условной вероятностью стандартного товара при условии, что он поставлен первой фирмой.

Аналогичные условные вероятности для товара, поставленного второй и третьей фирмами, будут равны $P(A/H_2) = 0{,}9$; $P(A/H_3) = 0{,}8$.

Условные вероятности поставки нестандартного товара (событие $\overline{A}$) первой, второй и третьей фирмой будут равны:

$$P(\overline{A})/H_1) = 1 - P(A/H_1) = 1 - 0{,}75 = 0{,}25;$$

$$P(\overline{A})/H_2) = 1 - P(A/H_2) = 1 - 0{,}9 = 0{,}1;$$

$$P(\overline{A})/H_3) = 1 - P(A/H_3) = 1 - 0{,}8 = 0{,}2.$$

Вероятность приобретения покупателем в торговой организации нестандартного товара $P(\overline{A})$ определяется формулой полной вероятности:

$$P(\overline{A}) = \sum_{i=1}^{3} P(H_i) \cdot P(\overline{A}/H_i),$$

$$P(\overline{A}) = P(H_1) \cdot P(\overline{A} / H_1) + P(H_2) \cdot P(\overline{A} / H_2) + P(H_3) \cdot P(\overline{A} / H_3) =$$
$$= 0,2 \cdot 0,2 + 0,5 \cdot 0,1 + 0,3 \cdot 0,2 = 0,16.$$

Задача 1.5. При оценке рисков страховая компания распределяет застрахованных на три класса: первый класс — малый риск; второй класс — средний риск; третий класс — большой риск. Среди клиентов страховой компании 30% застрахованных относятся к первому классу риска, 25% — ко второму классу риска и 45% — к третьему классу риска. Вероятность выплаты страхового вознаграждения для клиентов первого класса риска равна 0,03, для второго класса риска — 0,08 и для третьего — 0,04. Определить вероятность того, что: 1) застрахованный клиент получит страховое вознаграждение; 2) получивший страховое вознаграждение клиент относится ко второму классу риска.

Решение. 1. Обозначим событие A, состоящее в том, что клиент страховой компании получит страховое вознаграждение, а события H_1, H_2 и H_3, состоящие в том, что клиенты страховой компании относятся к первому, второму и третьему классу риска соответственно. В соответствии с условием задачи вероятности событий (гипотез) H_1, H_2 и H_3 будут равны:

$$P(H_1) = 0,3;\ P(H_2) = 0,25;\ P(H_3) = 0,45.$$

Условные вероятности $P(A/H_1)$, $P(A/H_2)$ и $P(A/H_3)$ выплаты страхового вознаграждения клиентам, относящимся к первому, второму и третьему классу риска, по условию задачи равны:

$$P(A/H_1) = 0,03;\ P(A/H_2) = 0,08;\ P(A/H_3) = 0,04.$$

В задаче «1» требуется определить безусловную вероятность выплаты страхового вознаграждения $P(A)$.

Эта вероятность определяется по формуле полной вероятности:

$$P(A) = \sum_{i=1}^{3} P(H_i) \cdot P(A / H_i) = P(H_1) \cdot P(A / H_1) + P(H_2) \cdot P(A / H_2) +$$
$$+ P(H_3) \cdot P(A / H_3) = 0,3 \cdot 0,03 + 0,25 \cdot 0,08 + 0,45 \cdot 0,04 = 0,047.$$

2. В данном вопросе задачи накладывается условие, что страховое вознаграждение уже выплачено, т.е. событие A уже произошло. Нужно определить условную вероятность того, что страховое вознаграждение выплачено клиенту, относящемуся ко второму классу риска $P(H_2/A)$. Эта условная вероятность определяется формулой Байеса:

$$P(H_2 / A) = \frac{P(H_2)P(A / H_2)}{P(A)} = \frac{0,25 \cdot 0,08}{0,047} \approx 0,425.$$

Задача 1.6. Из опыта работы страховой компании известно, что страховая сумма выплачивается по 20% договоров.

Найти вероятность того, что из десяти договоров страховая сумма будет выплачена: 1) по трем договорам; 2) не более чем по двум договорам.

Решение. 1. Обозначим событие A, состоящее в том, что страховая сумма выплачивается страховой компанией. Вероятность этого события по условию задачи равна $P(A) = p = 0,2$. Вероятность противоположного события $\overline{A}$ невыплаты страховой суммы равна $P(\overline{A}) = q = 1 - p = 0,8$. В первом вопросе задачи требуется определить вероятность того, что событие A наступит по $m = 3$ договорам из $n = 10$ договоров. Данная вероятность $P_n(m)$ определяется по формуле Бернулли:

$$P_{10}(3) = C_{10}^3 p^3 q^{10-3} = \frac{10!}{3!7!} \cdot 0,2^3 \cdot 0,8^7 = 120 \cdot 0,2^3 \cdot 0,8^7 \approx 0,201.$$

2. Обозначим событие B, состоящее в том, что из $n = 10$ договоров событие A наступит не более чем по двум договорам $m \le 2$. Событию B удовлетворяют следующие значения: $m = 0$; $m = 1$; $m = 2$, и вероятность события B по теореме сложения вероятностей и формуле Бернулли будет равна

$$P(B) = \sum_{m=0}^{2} P_{10}(m_i) = \sum_{m=0}^{2} C_{10}^m p^m q^{10-m} = C_{10}^0 p^0 q^{10} + C_{10}^1 p^1 q^9 + C_{10}^2 p^2 q^8 =$$

$$= q^8 (C_{10}^0 p^2 + C_{10}^1 pq + C_{10}^2 p^2) = 0,8^8 (1 \cdot 0,8^2 + 10 \cdot 0,2 \cdot 0,8 + 45 \cdot 0,2^2) \approx 0,678.$$

Задача 1.7. У страховой компании имеются $n = 100\,000$ клиентов. Из опыта работы страховой компании известно, что вероятность наступления несчастного случая с выплатой страховой суммы равна $p = 4 \cdot 10^{-5}$. Определить вероятность того, что страховой случай с выплатой страховой суммы произойдет у $m = 6$ клиентов.

Решение. Так как число клиентов страховой компании велико ($n = 10^5$), а вероятность наступления страхового случая с выплатой страховой суммы мала ($p = 4 \cdot 10^{-5}$), то вероятность наступления страхового случая с выплатой страховой суммы у m клиентов можно определить по формуле Пуассона (1.24):

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda},$$

где $\lambda = n \cdot p = 10^5 \cdot 4 \cdot 10^{-5} = 4$ — параметр распределения Пуассона.

Данную вероятность можно найти по таблице значений функции Пуассона, приведенной в приложении 4. По таблице при $\lambda = 4$ и $m = 6$ находим:

$$P(m = 6) \approx 0{,}1042.$$

Вопросы для самоконтроля

1. Какие события называются достоверными, случайными, невозможными?
2. Какие события называют элементарными? Что такое пространство элементарных событий?
3. Дайте определение суммы, разности, произведения случайных событий. Что такое совместные и несовместные события?
4. Дайте классическое определение вероятности.
5. Дайте статистическое определение вероятности. Что называется относительной частотой случайного события?
6. Дайте определение вероятности случайного события для модели с несчетным числом возможных исходов.
7. Дайте определение зависимых и независимых случайных событий. Что такое условная вероятность? Приведите примеры.
8. Сформулируйте и докажите теорему умножения вероятностей для зависимых и независимых событий.
9. Сформулируйте и докажите теорему сложения вероятностей для совместных и несовместных событий.
10. Сформулируйте и докажите теорему (формулу) полной вероятности. Приведите примеры.
11. Сформулируйте и докажите теорему Байеса. Приведите примеры.
12. Сформулируйте и докажите теорему о повторении опытов (формулу) Бернулли.
13. Дайте определение и выведите формулу для нахождения наивероятнейшего числа наступления случайного события при повторении опытов.
14. Сформулируйте теорему Пуассона и условия ее применимости. Поясните правила вычисления вероятностей по формуле Пуассона.

ГЛАВА 2

ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Основные понятия о дискретных случайных величинах

Наряду со случайными событиями одним из основных объектов изучения в теории вероятностей являются случайные величины. Любая количественная характеристика той или иной сферы человеческой деятельности является случайной величиной. Примерами случайных величин в области экономики и финансов является разброс цен на товары и услуги, величина банковской кредитной ставки, курсы валют на валютной бирже, налоговые и неналоговые поступления в бюджет и т.п.

В теории вероятностей под ***случайной величиной*** понимают такую переменную величину, которая в результате опыта может принимать то или иное заранее неизвестное значение.

Понятия случайной величины и случайного события тесно связаны между собой. Так, например, при бросании игральной кости возможные исходы опыта можно рассматривать как случайные события, а можно — как случайные величины. Число очков, выпавших на игральной кости, можно рассматривать как случайную величину, которая в результате опыта может принять одно из шести значений — 1, 2, 3, 4, 5, 6, но заранее не известно какое.

По аналогии с пространством элементарных событий случайные величины можно характеризовать пространством (областью или множеством) возможных значений случайной величины Ω.

Случайные величины будем обозначать прописными латинскими буквами X, Y, Z, а значения случайных величин, которые они принимают в результате опыта, соответственно строчными буквами x, y, z.

По возможным значениям случайные величины подразделяют на два основных типа — дискретные и непрерывные случайные величины. ***Дискретной случайной величиной*** называют случайную величину, множество возможных значений которой либо конечное, либо бесконечное, но счетное, т.е. все возможные значения дискретной случайной величины могут быть пронумерованы натуральными числами. Например, случайная величина X, которая может принимать только целочисленные значения, является дискретной случайной величиной.

Множество возможных значений случайной величины Ω может быть как конечным, так и бесконечным множеством. Если множество значений дискретной случайной величины является конечным, то мы можем заранее перечислить все ее возможные значения. Например, если множество значений дискретной случайной величины X, принимающей целочисленные значения, является конечным

$$\Omega_X = \{5 \le x \le 10\},$$

то можно заранее записать все ее возможные значения:

$$x_1 = 5;\ x_2 = 6;\ x_3 = 7;\ x_4 = 8;\ x_5 = 9;\ x_6 = 10.$$

Если множество значений случайной величины является бесконечным

$$\Omega_X = \{-\infty < x < \infty\},$$

то перечислять все ее возможные значения бессмысленно, их число бесконечно. Но мы можем охарактеризовать возможные значения дискретной случайной величины общим признаком, например, случайная величина X может принимать только четные значения.

Задание области возможных значений случайной величины является необходимым, но не дает ее вероятностного описания. Для вероятностного описания случайных величин должен быть задан закон распределения вероятностей.

Закон распределения вероятностей устанавливает взаимосвязь между значениями случайных величин x и их вероятностями $P(x)$.

2.2. Закон распределения вероятностей дискретных случайных величин

Перечень всех возможных значений дискретной случайной величины и соответствующих этим значениям вероятностей появления этих значений случайной величины называют ***законом распределения вероятностей дискретной случайной величины***.

Данный закон распределения вероятностей называют рядом распределения дискретных случайных величин.

Предположим, что дискретная случайная величина X может принимать конечное множество всех ее возможных значений $\Omega = \{x_1, x_2, ..., x_i, ..., x_n\}$, причем известны вероятности того, что случайная величина X примет то или иное значение $P(X = x_i) = p_i;\ i = 1 \div n$. В этом случае соответствие между значениями случайной величины x_i и вероятностями их появления p_i (закон распределения) может быть задано в табличной форме (табл. 2.1).

Таблица 2.1

Пример задания закона распределения дискретной случайной величины в табличной форме

x_i	1	2	3	4	5	6	7
p_i	0,053	0,083	0,113	0,143	0,173	0,203	0,232

Значения x_i, указанные в табл. 2.1, включают всю область возможных значений дискретной случайной величины. В каждом исходе опыта дискретная случайная величина может принять только одно какое-то значение. Поэтому на основании следствия теоремы сложения вероятностей для несовместных событий должно выполняться условие нормировки:

$$\sum_{i=1}^{n} p_i = 1 .$$

Помимо табличной формы, закон распределения вероятностей может быть задан в виде графика (рис. 2.1).

При графическом задании закона распределения вероятностей по оси абсцисс откладывают возможные значения дискретной случайной величины, а по оси ординат — значения вероятностей их появления. На рисунке 2.1 приведен пример графического задания закона распределения случайной величины, соответствующего табл. 2.1. Если ординаты каждой точки соединяются между собой линией (на рис. 2.1 пунктирная линия), то полученный график называют многоугольником распределения. При этом следует помнить, что ординаты p_i соединяют лишь для наглядности, так как случайная величина X дискретна и не может принимать значения внутри интервалов $x_i \div x_{i+1}$. Вероятности значений x внутри этих интервалов $x_i < x < x_{i+1}$ равны нулю.

Кроме того, закон распределения случайных величин может быть задан в аналитическом виде (в виде формулы).

Например, для дискретной случайной величины X, принимающей целочисленные значения в интервале $0 \le x \le n$, возможный закон распределения вероятностей может описываться формулой Бернулли (1.21):

$$P(x) = C_n^x p^x (1-p)^{n-x}. \qquad (2.1)$$

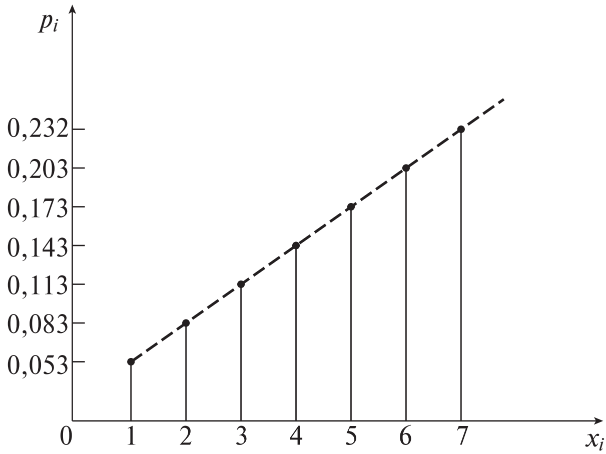


Рис. 2.1. Пример графического задания закона распределения вероятностей

В данной формуле переменная случайная величина x может принимать значения 0, 1, 2, ..., n. Значения параметров n и p, называемых параметрами распределения, должны быть заданы. При заданных значениях параметров n и p формула (2.1) позволяет для каждого значения x_i рассчитать вероятности появления этих значений p_i.

Если множество возможных значений случайной величины X является несчетным (бесконечное множество), то ее закон распределения вероятностей задается, как правило, в аналитическом виде. Табличное и графическое задание законов распределения для таких случайных величин используется лишь для наглядности.

Например, для дискретной случайной величины X, принимающей целочисленные значения в интервале $x \geq 0$, возможный закон распределения вероятностей может описываться формулой Пуассона (1.23):

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}. \qquad (2.2)$$

Данный закон распределения полностью определяется одним параметром распределения λ, который должен быть задан.

При заданном законе распределения вероятностей могут быть определены различные вероятностные характеристики случайной величины.

Пример 2.1. Для дискретной случайной величины, закон распределения вероятностей которой задан табл. 2.1, определить вероятность того, что случайная величина примет значения, принадлежащие интервалу $2 \le x \le 5$.

Решение. Будем рассматривать как случайное событие A_i то, что случайная величина примет значение x_i. Тогда событие B, состоящее в том, что случайная величина примет значение, принадлежащее интервалу $2 \le x \le 5$, будет являться суммой событий $B = A_2 + A_3 + A_4 + A_5$.

В каждом опыте случайная величина может принять только одно значение, значит события A_i являются несовместными. Тогда вероятность события B можно найти по формуле сложения вероятностей для несовместных событий:

$$P(B) = P(2 \le x \le 5) = \sum_{i=2}^{5} P(A_i) = \sum_{i=2}^{5} p_i. \qquad (2.3)$$

Подставляя в формулу (2.3) значения вероятностей p_i из табл. 2.1, получим:

$$P(2 \le x \le 5) = \sum_{i=2}^{5} p_i = 0{,}083 + 0{,}113 + 0{,}143 + 0{,}173 = 0{,}512.$$

2.3. Интегральная функция распределения дискретных случайных величин

Ряд распределения является очень удобной формой вероятностного описания дискретных случайных величин. Однако во многих случаях более удобно задавать закон распределения вероятностей не рядом распределения, а так называемой интегральной функцией распределения случайных величин.

Интегральной функцией распределения (или просто функцией распределения) случайной величины X называют вероятность того, что в результате испытания X примет значение меньше некоторого значения x.

Обозначается функция распределения как $F(x)$ и в соответствии с определением

$$F(x) = P(X < x). \qquad (2.4)$$

Графически определение функции распределения и формулу (2.4) можно проиллюстрировать рис. 2.2. Если X — случайная величина, которая может принимать некоторые значения, отмеченные на числовой оси, то функция распределения есть вероятность того, что значение случайной величины, полученное в результате опыта, окажется левее текущего значения переменной x.

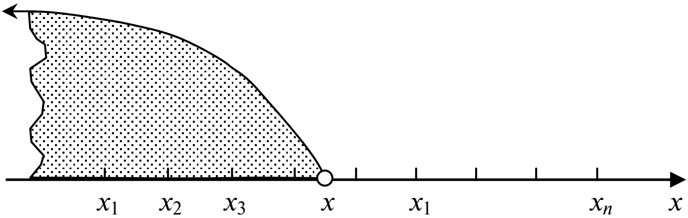


Рис. 2.2. К определению функции распределения

Для дискретной случайной величины X возможные ее значения образуют конечное множество $\{x_1, x_2, ..., x_i, ..., x_n\}$, и функцию распределения можно определить в виде суммы:

$$F(x) = \sum_{x_i < x} P(x_i)\,, \tag{2.5}$$

где неравенство $x_i < x$ означает, что суммирование осуществляется по всем значениям x_i, которые меньше некоторого текущего значения переменной x.

Функция распределения, как и ряд распределения, может быть задана в табличном, графическом или аналитическом виде.

В соответствии с определением функции распределения (2.4) и формулой (2.3) для дискретной случайной величины X можно записать:

$$F(x) = P(X < x) = \sum_{i=1}^{x} p_i\ . \tag{2.6}$$

Подставляя в формулу (2.6) значения вероятностей p_i из табл. 2.1, получим значения для функции распределения.

Таблица 2.2

Функция распределения дискретной случайной величины

x	1	2	3	4	5	6	7
$F(x)$	0,053	0,136	0,249	0,392	0,565	0,768	1

График функции распределения дискретной случайной величины (табл. 2.2) приведен на рис. 2.3.

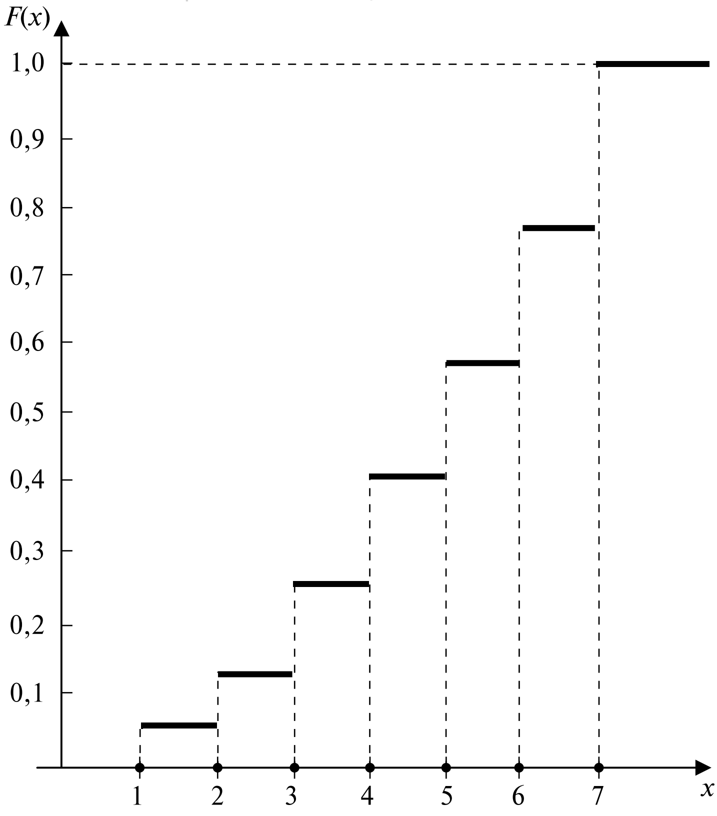


Рис. 2.3. График функции распределения дискретной случайной величины

При текущих значениях $x < 1$ функция распределения равна нулю, так как $P(X<1) = 0$. При текущих значениях переменной $1 \leq x < 2$ функция распределения равна $F(x) = P(X < 2) = p_1$. При $2 \leq x < 3$ $F(x) = P(X < 3) = p_1 + p_2$ и т.д.

В конечном итоге при текущих значениях $x \geq 7$ получим, что функция распределения будет равна:

$$F(x) = P(X \geq 7) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 = 1.$$

По функции распределения, как и по ряду распределения в примере 2.1, можно определить вероятность попадания случайной величины в заданный интервал. Предположим, нам нужно определить вероятность попадания случайной величины в интервал $x_1 \leq x < x_2$ (рис. 2.4).

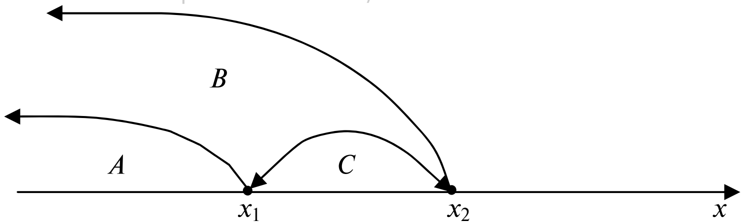


Рис. 2.4. К определению вероятности попадания случайной величины в заданный интервал

Обозначим событием A ситуацию, когда случайная величина X принимает значение меньше, чем x_1. Тогда в соответствии с определением функции распределения (2.4) вероятность события A равна значению функции распределения в точке x_1:

$$P(A) = F(x_1).$$

Событием B обозначим ситуацию, когда $X < x_2$. Для вероятности события B аналогично можно записать $P(B) = F(x_2)$. Из рисунка 2.4 следует, что событие C, состоящее в попадании значений случайной величины в интервал $[x_1; x_2)$, является разностью событий:

$$C = B - A \text{ или } C + A = B.$$

Тогда на основании теоремы сложения вероятностей для несовместных событий можно записать:

$$P(C) = P(x_1 \le x < x_2) = P(B) - P(A)$$

или

$$P(x_1 \le x < x_2) = F(x_2) - F(x_1). \tag{2.7}$$

Рассмотрим основные свойства функции распределения.

Свойство 1. Функция распределения $F(x)$ принимает значения в интервале от нуля до единицы:

$$0 \le F(x) \le 1.$$

Это свойство очевидно, так как $F(x)$ — это вероятность случайного события $X < x$, а вероятность может принимать значения в интервале от нуля до единицы.

Свойство 2. $\lim\limits_{x\to-\infty} F(x) = 0,\ \lim\limits_{x\to\infty} F(x) = 1$.

Первый предел равен нулю, так как событие $X < -\infty$ невозможно. Второй предел равен единице, так как событие $X < \infty$ является достоверным. Если множество возможных значений случайной величины X

ограничено снизу значением $x_{\min}$, а сверху $x_{\max}$, то это свойство можно записать в виде:

$$\lim_{x \to x_{\min}} F(x) = 0, \quad \lim_{x \to x_{\max}} F(x) = 1.$$

Свойство 3. Функция распределения является функцией неубывающей, т.е. для любых значений x_1 и x_2, если $x_1 < x_2$, выполняется неравенство $F(x_1) \le F(x_2)$.

Это свойство легко доказывается на основании формулы (2.7). Предположим обратное,

$$F(x_1) > F(x_2) \text{ при } x_1 < x_2. \tag{2.8}$$

Тогда для вероятности попадания случайной величины в интервал от x_1 до x_2 получим:

$$P(x_1 \le x < x_2) = F(x_2) - F(x_1) < 0.$$

Но вероятность отрицательной быть не может. Значит, наше предположение (2.8) неверно, а верно свойство 3.

Пример 2.2. Решить пример 2.1, если случайная величина X задана функцией распределения (табл. 2.2).

Решение. Для решения воспользуемся формулой (2.7). В примере 2.1 требуется определить вероятность того, что случайная величина X примет значение, принадлежащее интервалу $2 \le x \le 5$. В данном неравенстве нижняя граница, равная 2, входит в заданный интервал, а в формуле (2.7) нижняя граница x_1 не входит в заданный интервал, поэтому за x_1 мы должны взять ближайшее снизу к 2 возможное значение случайной величины X, т.е. $x_1 = 1$. Отсюда получим:

$$P(2 \le x \le 5) = P(1 < x \le 5) = F(5) - F(1) = 0{,}565 - 0{,}053 = 0{,}512.$$

Мы получили то же значение, что и в примере 2.1.

2.4. Числовые характеристики дискретных случайных величин

При решении большинства практических задач требуется знание не столько законов распределения случайных величин, сколько их числовых характеристик, дающих иногда более наглядное представление о свойствах и особенностях случайных величин. Числовые характеристики называют также моментами распределения случайных величин, так как они зависят от закона распределения вероятностей.

Наибольшее употребление в теории вероятностей имеют такие числовые характеристики, как математическое ожидание и дисперсия случайных величин.

2.4.1. Математическое ожидание дискретных случайных величин

Математическим ожиданием дискретной случайной величины X называется сумма произведений всех ее возможных значений x_i $(i = 1 \div n)$ на соответствующие им вероятности p_i. Обозначают математическое ожидание $M(X)$, m_x или $<X>$:

$$m_x = M(X) = <X> = \sum_{i=1}^{n} x_i p_i \,. \tag{2.9}$$

Математическое ожидание — это среднее значение случайной величины, вокруг которого группируются ее возможные значения, поэтому операцию нахождения математического ожидания называют операцией усреднения. Буква M или угловые скобки < > в формуле (2.9) говорят о том, что осуществляется операция усреднения той случайной величины, которая указана в скобках. Из формулы (2.9) видно, что операция усреднения (нахождения среднего значения) состоит в умножении каждого из возможных значений случайной величины x_i на их вероятности p_i и суммировании всех полученных произведений.

Пример 2.3. Закон распределения случайной величины X задан рядом распределения

x_i	0	1	2	3	4	5
p_i	0,3	0,2	0,15	0,2	0,05	0,1

Определить математическое ожидание случайной величины X.

Решение. Используя формулу (2.9), получим:

$$m_x = \sum_{i=1}^{5} x_i p_i = 0 \cdot 0{,}3 + 1 \cdot 0{,}2 + 2 \cdot 0{,}15 + 3 \cdot 0{,}2 + 4 \cdot 0{,}05 + 5 \cdot 0{,}1 = 1{,}8 \,.$$

Отметим основные свойства математического ожидания.

Свойство 1. Математическое ожидание постоянной величины C есть сама постоянная $M(C) = C$.

Действительно, если случайная величина X принимает одно-единственное значение $X = C$, то вероятность появления этого значения равна единице $p = 1$. Отсюда $M(C) = C \cdot 1 = C$.

Свойство 2. Постоянный множитель можно вынести за знак математического ожидания $M(CX) = CM(X)$.

Доказательство. Предположим, что дискретная случайная величина X задана рядом распределения, т.е. известны значения x_i и p_i. Найдем математическое ожидание случайной величины $Y = CX$:

$$M(Y) = M(CX).$$

Возможные значения Y равны $y_i = Cx_i$ и будут иметь те же значения вероятностей p_i, что и x_i, тогда

$$M(Y) = M(CX) = \sum_i y_i p_i = \sum_i Cx_i p_i = C\sum_i x_i p_i = CM(X).$$

Свойство 3. Математическое ожидание алгебраической суммы конечного числа случайных величин равно алгебраической сумме математических ожиданий случайных величин.

Докажем это свойство для двух дискретных случайных величин.

Предположим, что случайные величины X и Y заданы рядом распределения:

X	x_1	x_2
P_x	p_1	p_2

Y	y_1	y_2
P_y	q_1	q_2

Составим закон распределения случайной величины $Z = X \pm Y$. Возможные значения случайной величины Z будут определяться следующими элементами множества:

$$z_1 = x_1 \pm y_1;\ z_2 = x_1 \pm y_2;\ z_3 = x_2 \pm y_1;\ z_4 = x_2 \pm y_2.$$

Для того чтобы случайная величина Z приняла значение z_k, $k = 1 \div 4$, необходимо, чтобы одновременно произошли два события: $X = x_i$ и $Y = y_j$. Значит, вероятность того, что Z примет значение z_k, будет определяться по формуле умножения вероятностей:

$$P(Z = z_k) = p_i q_j.$$

Таким образом, ряд распределения случайной величины Z будет выглядеть так:

z_k	$x_1 \pm y_1$	$x_1 \pm y_2$	$x_2 \pm y_1$	$x_2 \pm y_2$
$P(z_k)$	p_1q_1	p_1q_2	p_2q_1	p_2q_2

Тогда для $M(Z)$ получим:

$$M(Z) = (x_1 \pm y_1)\, p_1q_1 + (x_1 \pm y_2)\, p_1q_2 + (x_2 \pm y_1)\, p_2q_1 + (x_2 \pm y_2)\, p_2q_2 =$$
$$= x_1 p_1q_1 \pm y_1 p_1q_1 + x_1 p_1q_2 \pm y_2 p_1q_2 + x_2 p_2q_1 \pm y_1 p_2q_1 + x_2 p_2q_2 \pm y_2 p_2q_2.$$

Сгруппируем слагаемые при одинаковых значениях x и y:

$$M(z) = x_1p_1(q_1 + q_2) + x_2p_2(q_1 + q_2) \pm y_1q_1(p_1 + p_2) \pm y_2q_2(p_1 + p_2).$$

По условию нормировки $p_1 + p_2 = 1$ и $q_1 + q_2 = 1$. Отсюда получим:

$$M(z) = x_1p_1 + x_2p_2 \pm y_1q_1 \pm y_2q_2 = M(X) \pm M(Y).$$

Свойство 4. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий.

Докажем это свойство для произведения двух независимых случайных величин.

Предположим, случайные величины X и Y заданы рядом распределения (см. таблицы из доказательства свойства 3).

Составим закон распределения случайной величины $Z = XY$. Возможные значения Z будут определяться элементами множества:

$$z_1 = x_1y_1;\ z_2 = x_1y_2;\ z_3 = x_2y_1;\ z_4 = x_2y_2.$$

Вероятности значений z_k, $k = 1 \div 4$ также будут определяться произведением вероятностей p_iq_j

z_k	$x_1\,y_1$	$x_1\,y_2$	$x_2\,y_1$	$x_2\,y_2$
$P(z_k)$	p_1q_1	p_1q_2	p_2q_1	p_2q_2

$$M(Z) = x_1y_1p_1q_1 + x_1y_2p_1q_2 + x_2y_1p_2q_1 + x_2y_2p_2q_2 =$$
$$= x_1p_1(y_1q_1 + y_2q_2) + x_2p_2(y_1q_1 + y_2q_2) = (x_1p_1 + x_2p_2)(y_1q_1 + y_2q_2) =$$
$$= M(X) \cdot M(Y).$$

2.4.2. Дисперсия дискретных случайных величин

Для полного представления о характере случайной величины знания одного лишь математического ожидания недостаточно. Может оказаться так, что две и более случайные величины имеют одно и то же среднее значение, но их рассеяние вокруг центра распределения будет различным. Поэтому целесообразно ввести еще одну — качественную — характеристику, дающую представление о степени разбросанности значений случайной величины вокруг ее среднего значения.

Пусть в результате испытания случайная величина X приняла значение $X = x$. Разность $x - m_x$ называется ***отклонением*** значения случайной величины от ее математического ожидания. Новая случайная величина $X - m_x$ называется ***центрированной*** и обозначается символом $\overset{\circ}{X}$. Ее математическое ожидание равно нулю:

$$M(\overset{\circ}{X}) = M(X - m_x) = M(X) - M(m_x) = 0.$$

Отсюда следует, что среднее значение отклонения $X - m_x$ не может быть использовано для оценки степени рассеяния случайной величины. Можно взять в качестве меры рассеяния среднее значение абсолютной величины отклонения $|X - m_x|$, эта характеристика, называемая ***абсолютным отклонением,*** используется на практике очень редко. Во-первых, ее применение в расчетах сопряжено с некоторыми трудностями, а во-вторых, как выяснилось, она плохо реагирует на большие отклонения, учет которых имеет в ряде случаев решающее значение.

Наиболее оптимальным в этом смысле оказалось среднее значение квадрата отклонения $X - m_x$; эта характеристика получила название ***дисперсии*** и обозначение $D(X)$, Dx, $< [x - m_x]^2 >$ или $M[(x - m^x)^2]$.

Дисперсией случайной величины X называется математическое ожидание квадрата отклонения этой величины от ее центра распределения:

$$D(X) = M[(X - m_x)^2] \text{ или } D(X) = M(\overset{\circ}{X})^2.$$

Если X — дискретная случайная величина, то из определения математического ожидания следует:

$$D(X) = \sum_{i=1}^{n} (x_i - m_x)^2 p_i. \tag{2.10}$$

На практике для вычисления дисперсии случайных величин удобнее пользоваться другой формулой

$$D(X) = \sum_{i=1}^{n} x_i^2 p_i - m_x^2, \tag{2.11}$$

которая легко получается из формулы (2.10) при возведении разности в квадрат и почленном суммировании:

$$D(X) = \sum_{i=1}^{n} (x_i^2 - 2x_i m_x + m_x^2) p_i = \sum_{i=1}^{n} x_i^2 p_i - 2m_x \sum_{i=1}^{n} x_i p_i + m_x^2 \sum_{i=1}^{n} p_i.$$

Так как $\sum_{i=1}^{n} x_i p_i = m_x$ и $\sum_{i=1}^{n} p_i = 1$, то после подстановки этих значений для дисперсии $D(X)$ получим формулу (2.11).

Дисперсия как мера рассеяния случайной величины все же имеет один недостаток: размерность ее квадратичная, а значит, со средним значением случайной величины дисперсия несоизмерима.

Поэтому вводится еще одна числовая характеристика — среднее квадратичное отклонение.

Средним квадратичным отклонением σ_x случайной величины X называется корень квадратный из ее дисперсии:

$$\sigma_x = \sqrt{D(X)}. \tag{2.12}$$

Так как $D(X) = \sigma_x^2$, то дисперсию часто обозначают символом σ_x^2.

Рассмотрим основные свойства дисперсии.

Свойство 1. Дисперсия постоянной величины равна нулю.

Действительно, так как $M(C) = C$, то $D(C) = M(C - C)^2 = M(0) = 0$.

Свойство 2. Постоянный множитель C можно вынести за знак дисперсии, возведя его в квадрат:

$$D(CX) = C^2 D(X).$$

Доказательство. На основании свойства 2 математического ожидания (см. подраздел 2.4.1) $M(CX) = C\,M(X) = Cm_x$. Значит,

$$D(CX) = M(CX - Cm_x)^2 = M[C^2(X - m_x)^2] = C^2 M(X - m_x)^2 = C^2 D(X).$$

Свойство 3. Дисперсия суммы (разности) независимых случайных величин равна сумме дисперсий слагаемых:

$$D(X_1 \pm X_2 \pm \ldots \pm) = D(X_1) + D(X_2) + \ldots + D(X_n).$$

Докажем свойство для двух независимых случайных величин X и Y:

$$D(Z) = D(X \pm Y) = D(X) + D(Y).$$

Обозначим $Z = X \pm Y$. По определению дисперсии

$$\begin{aligned} D(Z) &= D(X \pm Y) = M(Z - m_z)^2 = M(X - m_x \pm Y \mp m_y)^2 = \\ &= M(x - m_x)^2 \pm 2M[(X - m_x)(Y - m_y)] + M(Y - m_y)^2. \end{aligned} \tag{2.13}$$

Первое и третье слагаемые в выражении (2.13) равны соответственно

$$M(X - m_x)^2 = D_x;$$

$$M(Y - m_y)^2 = D_y.$$

Второе слагаемое в выражении (2.13) равно нулю, так как по условию X и Y являются независимыми случайными величинами, значит независимы и их центрированные величины. Отсюда на основании свойства 4 математического ожидания можно записать:

$$M[(X - m_x)(Y - m_y)] = M(X - m_x)M(Y - m_x) = 0.$$

Подставляя данное равенство в формулу (2.13), получим

$$D(z) = D(X \pm Y) = D(x) + D(y)$$

Пример 2.4. Для дискретной случайной величины X, закон распределения которой задан в примере 2.3, найти дисперсию D_x и среднеквадратичное отклонение.

Решение. Для решения задачи воспользуемся формулой (2.11). Математическое ожидание случайной величины X найдено в примере 2.3: $m_x = 1{,}8$. Найдем математическое ожидание квадрата случайной величины:

$$M(x^2) = \sum_{i=1}^{5} x_i^2 p_i = 0 \cdot 0{,}3 + 1 \cdot 0{,}2 + 4 \cdot 0{,}15 + 9 \cdot 0{,}2 + 16 \cdot 0{,}05 + 25 \cdot 0{,}1 = 5{,}9\ .$$

Отсюда для дисперсии D_x по формуле (2.11) получим

$$D_x = M(x^2) - m_x^2 = 5{,}9 - 3{,}24 = 2{,}66\ .$$

Среднеквадратичное отклонение случайной величины X будет равно

$$\sigma_x = \sqrt{D_x} \cong 1{,}63.$$

2.4.3. Моменты распределения дискретных случайных величин

Математическое ожидание $M(X)$ и дисперсия $D(X)$ являются частными случаями более общего понятия, называемого моментом распределения случайной величины.

В теории вероятностей моменты служат для описания основных свойств распределения случайных величин.

Моментом распределения k-го порядка случайной величины X относительно некоторой постоянной C называется математическое ожидание k-й степени отклонения X от постоянной C:

$$M(X-C)^k.$$

Если постоянная $C = 0$, то момент называется ***начальным*** и обозначается:

$$\nu_k = M(X^k).$$

Начальный момент k-го порядка абсолютной величины X называется ***абсолютным*** моментом:

$$M\left(|X|^k\right).$$

Если в качестве C берется математическое ожидание m_x случайной величины X, т.е. центр ее распределения, то момент называется ***центральным***:

$$\mu_k = M(X-m_x)^k.$$

При $k = 0$ начальный и центральный моменты равны 1:

$$\nu_0 = M(X^0) = M(1) = 1;\ \mu_0 = M(X-m_x)^0 = M(1) = 1.$$

При $k = 1$ имеем: $\nu_1 = M(X^1) = m_x$, т.е. начальный момент первого порядка есть не что иное, как математическое ожидание случайной величины; $\mu_1 = M(X-m_x) = 0$ — центральный момент первого порядка, равен нулю.

При $k = 2$: $\nu_2 = M(X^2)$ — математическое ожидание квадрата случайной величины X — известное выражение из формулы (2.11) для вычисления дисперсии; $\mu_2 = M(X-m_x)^2 = D(X)$ — центральный момент второго порядка есть дисперсия случайной величины X.

Иногда для более полной характеристики распределений используются еще центральные моменты третьего и четвертого порядков:

$$\mu_3 = M(X-m_x)^3;\ \mu_4 = M(X-m_x)^4.$$

Отношение центрального момента третьего порядка μ_3 к кубу среднеквадратичного отклонения называют ***коэффициентом асимметрии*** закона распределения вероятностей случайной величины:

$$a = \frac{\mu_3}{\sigma_x^3}.$$

Коэффициент асимметрии является мерой симметричности ряда распределения относительно математического ожидания. На рисунке 2.5 приведены графики многоугольников распределения дискретных случайных величин.

Для случайной величины X, закон распределения вероятностей которой приведен на рис. 2.5, *а*, математическое ожидание будет равно $m_x = 4$. Из графика закона распределения (см. рис. 2.5, *а*) следует, что он симметричен относительно математического ожидания. Если для данного закона распределения вероятности посчитать μ_3 и коэффициент асимметрии, то увидим, что $\mu_3 = 0$, $a = 0$. То есть при $\mu_3 = 0$ и $a = \frac{\mu_3}{\sigma_x^3} = 0$ график закона распределения вероятностей является симметричным относительно математического ожидания m_x. Для случайных величин, закон распределения вероятностей которых приведен на рис. 2.5, *б* и 2.5, *в*, для математических ожиданий соответственно получим:

$$m_б = \sum_{i=1}^{7} x_i p_{iб} = 4{,}7;\ m_в = \sum_{i=1}^{7} x_i p_{iв} = 3{,}3.$$

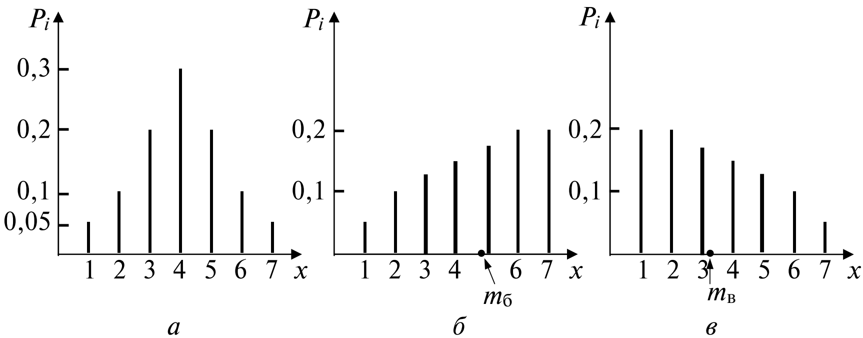


Рис. 2.5. Законы распределения дискретных случайных величин с различными коэффициентами асимметрии:
а — математическое ожидание $m_x = 4$; *б* — математическое ожидание $m_x = 4{,}7$; *в* — математическое ожидание $m_x = 3{,}3$

Из приведенных графиков видно, что закон распределения вероятностей для этих случайных величин является несимметричным относительно математического ожидания.

Для закона распределения вероятностей, приведенного на рис. 2.5, *в*, центральный момент третьего порядка μ_3 и коэффициент асимметрии будут иметь отрицательные значения.

Для закона распределения вероятностей, приведенного на рис. 2.5, *б*, центральный момент третьего порядка и коэффициент асимметрии будут иметь положительные значения.

2.5. Основные законы распределения дискретных случайных величин

В своей практической деятельности человек имеет дело с самыми разнообразными случайными величинами. Это разнообразие определяется различными условиями выполнения того или иного вида практической работы, различными природными или экономическими факторами, оказывающими влияние на развитие той или иной ситуации. Однако несмотря на свое разнообразие, случайные величины, формирующиеся под влиянием большого числа факторов, могут быть описаны тем или иным хорошо изученным законом распределения вероятностей. Рассмотрим наиболее часто встречающиеся в теории вероятностей законы распределения дискретных случайных величин.

2.5.1. Равномерный закон распределения

Равномерный закон распределения дискретных случайных величин используется для вероятностного описания дискретных случайных величин, множество возможных значений которых ограничено минимальным $x_{\min}$ и максимальным значением $x_{\max}$. В пределах этого интервала случайная величина принимает n возможных значений, а вероятности появления каждого из возможных значений случайной величины одинаковы. В таблице 2.3 приведен пример ряда распределения случайной величины, имеющей равномерный закон распределения вероятностей.

Таблица 2.3

Пример ряда распределения случайной величины

X	$x_1 = x_{\min}$	x_2	x_3	...	x_i	...	$x_n = x_{\max}$
$P(x_i)$	$1/n$	$1/n$	$1/n$		$1/n$		$1/n$

Так как для каждого x_i вероятности одинаковы $P(x_i) = p$, то из условия нормировки получим:

$$\sum_{i=1}^{n} p(x_i) = \sum_{i=1}^{n} p_i = np_i = 1 \text{ или } p_i = 1/n.$$

Функция распределения равномерно распределенной дискретной случайной величины может быть записана в виде:

$$F(x)=\begin{cases}0 \text{ при } x<x_{\min}=x_1\\ \dfrac{m}{n} \text{ при } x_1\le x<x_n=x_{\max}\\ 1 \text{ при } x\ge x_n=x_{\max},\end{cases},$$

где m определяет число значений дискретной случайной величины, попадающих в интервал $x_{\min}\div x$.

На рисунке 2.6 приведены графическое представление ряда распределения и функции распределения равномерно распределенной дискретной случайной величины при $n = 5$ и $x_{\min} = 2$, $x_{\max} = 6$.

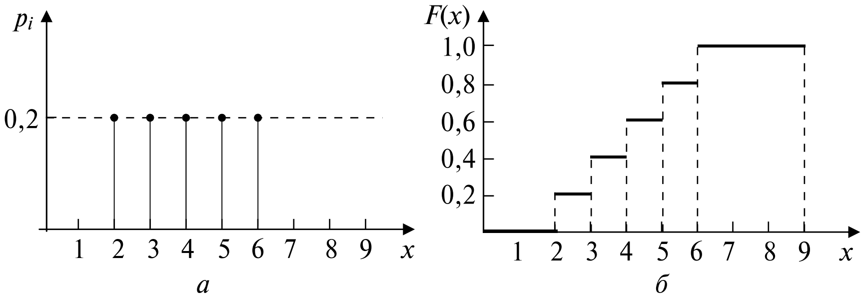


Рис. 2.6. Ряд распределения (*а*) и функция распределения (*б*) равномерно распределенной дискретной случайной величины

Если дискретная случайная величина X имеет равномерный закон распределения и принимает n возможных значений в интервале от $x_{\min}$ до $x_{\max}$ с шагом

$$h=\frac{x_{\max}-x_{\min}}{n-1},$$

то математическое ожидание и дисперсия данной случайной величины будут определяться формулами:

$$m_x=\frac{x_{\max}+x_{\min}}{2};\ D_x=\frac{(x_{\max}-x_{\min}+h)^2-h^2}{12}.$$

2.5.2. Биномиальный закон распределения (распределение Бернулли)

Условия применимости биномиального закона распределения аналогичны условиям применения теоремы о повторении опытов (см. подраздел 1.3.5). Проводится n независимых повторных испытаний, в каждом из которых событие A может появиться с вероятностью p. Число появлений события A при n испытаниях является дискретной случайной величиной, принимающей целочисленные значения в интервале $0 \le x \le n$. Вероятность того или иного значения дискретной случайной величины X в этом случае вычисляется по формуле Бернулли:

$$P(x = m) = P_n(m) = C_n^m p^m q^{n-m}\ ,\ m = 0,\ 1,\ 2,\ ...,\ n.$$

Значения вероятностей $P(x)$ и функции распределения $F(x)$ приведены в табл. 2.4, а их графики — на рис. 2.7 для биномиального закона распределения при $p = 0{,}4$ и $n = 6$.

Таблица 2.4

Значения параметров вероятности

x	0	1	2	3	4	5	6
$P(x)$	0,047	0,187	0,311	0,276	0,138	0,037	0,004
$F(x)$	0,047	0,234	0,545	0,821	0,959	0,996	1,0

Найдем математическое ожидание и дисперсию случайной величины, имеющей биномиальный закон распределения вероятностей. Для этого введем вспомогательную случайную величину Y_i, которая равна числу появления события A в одном i-м опыте ($i = 1 \div n$). Случайная величина Y_i может принимать значения 1 и 0 с вероятностями p и $q = 1 - p$ соответственно. Тогда для математического ожидания $M(Y_i)$ получим

$$M(Y_i) = 0 \cdot q + 1 \cdot p = p,$$

для дисперсии $D(Y_i)$

$$D(Y_i) = M(Y_i^2) - [M(Y_i)]^2 = p - p^2 = p(1 - p) = pq.$$

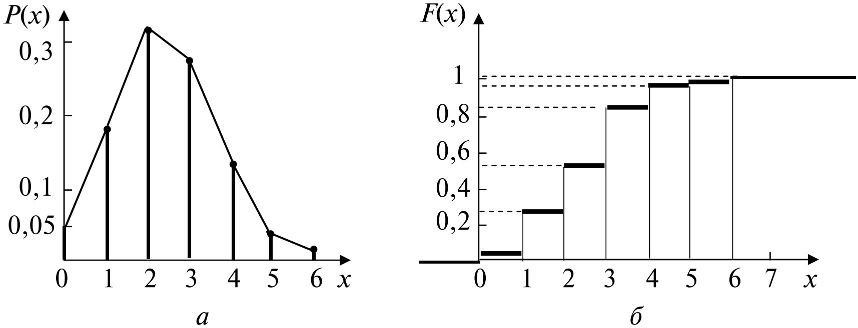


Рис. 2.7. Многоугольник (*а*) и функция (*б*) распределения при биномиальном законе

Случайная величина X является суммой случайных величин Y_i, $i = (1 \div n)$ при n-кратном повторении опыта $X = \sum_{i=1}^{n} Y_i$. Отсюда на основании свойства 3 (см. подраздел 2.4.1) математического ожидания $M(x)$ получим:

$$M(X) = M(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} M(Y_i) = \sum_{i=1}^{n} p = np.$$

Случайные величины Y_i независимы, поэтому на основании свойства 3 (см. подраздел 2.4.2) дисперсии $D(X)$ получим:

$$D(X) = D(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} D(Y_i) = \sum_{i=1}^{n} pq = npq.$$

Можно показать, что коэффициент асимметрии биноминального закона распределения равен

$$a = \frac{\mu_3}{\sigma^3} = \frac{1-2p}{\sqrt{npq}}.$$

Из приведенного соотношения следует, что при $p = 0{,}5$, $a = 0$, значит многоугольник распределения будет симметричным относительно точки $m_x = np = \frac{n}{2}$. При $p > 0{,}5$ коэффициент асимметрии меньше нуля, математическое ожидание m_x сдвигается в сторону больших значений X и многоугольник распределения становится несимметричным.

2.5.3. Геометрическое распределение

Данное распределение получается, когда случайная величина X является числом испытаний до первого появления интересующего нас события A. Если вероятность появления события A в каждом из опытов постоянна и равна p, то вероятность первого появления случайного события при n-кратном повторении опытов будет равна:

$$P(X = n) = q^{n-1} p; \quad n = 1, 2, 3, 4, \ldots; \quad q = 1 - p.$$

В данном случае случайная величина X может принимать целочисленные значения в неограниченном сверху интервале $0 < x < \infty$.

Математическое ожидание случайной величины X определяется формулой

$$m_x = \frac{1}{1-q}.$$

Значения вероятностей геометрического распределения при $p = 0{,}3$ и функции распределения $F(x)$ приведены в табл. 2.5, а соответствующие им графики — на рис. 2.8.

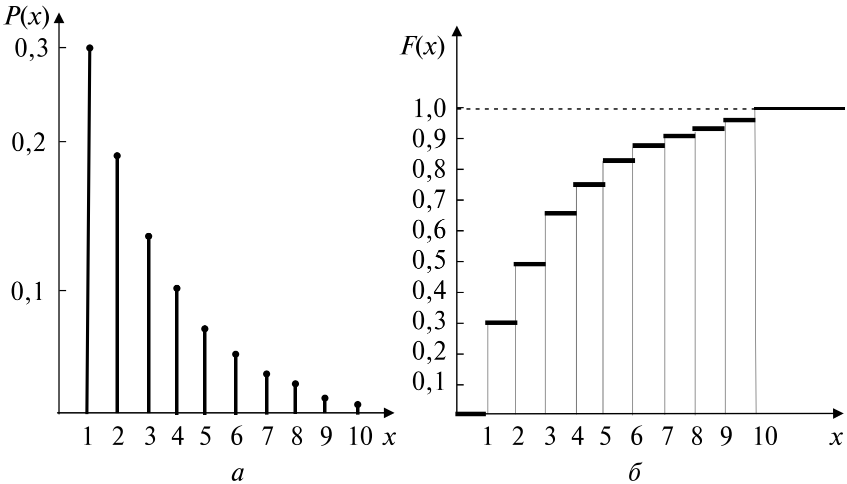


Рис. 2.8. Многоугольник и функция распределения при геометрическом законе распределения вероятностей ($p = 0{,}3$)

Таблица 2.5

Значения вероятностей геометрического распределения и функции распределения

x	1	2	3	4	5	6	7	8	9	10
$P(x)$	0,3	0,21	0,147	0,103	0,072	0,05	0,035	0,025	0,017	0,012
$F(x)$	0,3	0,51	0,657	0,76	0,832	0,882	0,917	0,942	0,959	1,0

2.5.4. Распределение Пуассона

Распределение Пуассона является предельным случаем формулы Бернулли (биномиального распределения при бесконечном повторении опытов). Если в каждом из опытов вероятность появления интересующего нас события A очень мала и равна p и при бесконечном повторении опытов $n \to \infty$ произведение $n \cdot p = \lambda$, то вероятность того, что событие A появится X раз, определяется распределением Пуассона:

$$P(x) = \frac{\lambda^x}{x!}e^{-\lambda},$$

где $x = 0, 1, 2, 3, ...$, т.е. случайная величина X, может принимать целочисленные значения в неограниченном сверху интервале $0 \le x < \infty$.

Из приведенной формулы следует, что распределение Пуассона полностью определяется одним параметром λ, значение которого должно быть задано.

Значения вероятностей $P(x)$ распределения Пуассона при $\lambda = 2$ и функции распределения $F(x)$ приведены в табл. 2.6, а соответствующие им графики — на рис. 2.9.

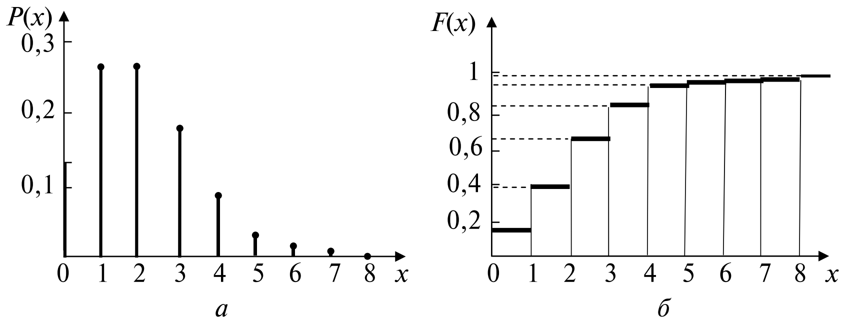


Рис. 2.9. Многоугольник (*а*) и функция (*б*) распределения для распределения Пуассона при $\lambda = 2$

Таблица 2.6

Значения вероятностей распределения Пуассона и функции распределения

x	0	1	2	3	4	5	6	7	8
$P(x)$	0,135	0,271	0,271	0,180	0,090	0,036	0,012	0,003	0,001
$F(x)$	0,135	0,406	0,677	0,857	0,947	0,989	0,995	0,999	1

Определим математическое ожидание и дисперсию случайной величины X, имеющей распределение Пуассона.

По формуле (2.9) при $n \to \infty$ получим:

$$M(X) = \sum_{i=0}^{\infty} x_i p(x_i) = \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!}.$$

Полученный ряд является разложением показательной функции

$$\sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} = e^{\lambda}.$$

С учетом этого равенства окончательно для математического ожидания получим

$$m_x = M(X) = \lambda .$$

Дисперсию D_x будем определять по формуле (2.11): $D_x = M(X^2) - m_x^2$. Для начального момента второго порядка $M(X^2)$ можно записать:

$$M(X^2) = \sum_{i=0}^{\infty} x_i^2 p(x_i) = \sum_{m=0}^{\infty} m^2 \frac{\lambda^m}{m!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{m\lambda^{m-1}}{(m-1)!} .$$

Полученный ряд является разложением произведения e^{λ} на $(1+\lambda)$:

$$\sum_{m=1}^{\infty} \frac{m\lambda^{m-1}}{(m-1)!} = (1+\lambda)e^{\lambda} .$$

С учетом приведенной формулы для начального момента второго порядка получим

$$M(X^2) = \lambda(1+\lambda) = \lambda + \lambda^2.$$

С учетом формулы (2.11) и полученного значения для математического ожидания $m_x = \lambda$ окончательно получим, что дисперсия случайной величины X будет равна

$$D_x = \lambda + \lambda^2 - \lambda^2 = \lambda .$$

Из полученных результатов следует, что если случайная величина X имеет распределение Пуассона, то ее математическое ожидание m_x и дисперсия D_x имеют одно и то же значение, равное параметру распределения Пуассона λ:

$$m_x = D_x = \lambda.$$

2.6. Закон распределения системы дискретных случайных величин

Процессы, происходящие в окружающей нас действительности, представляют собой результат взаимодействия большого количества случайных величин. При изучении сложных процессов приходится рассматривать две, три и более случайные величины, выступающие в системе как единое целое, с учетом их взаимосвязи.

Управление финансовой деятельностью предприятия связано с исследованием некоторого комплекса показателей, например таких, как объем выпускаемой продукции, ее себестоимость, цена и спрос на данную продукцию на рынке. Каждый из названных показателей, например себестоимость продукции, зависит от множества факторов и для каждого предприятия является случайной величиной. Кроме того, можно видеть, что рассматриваемые показатели являются зависимыми друг от друга случайными величинами.

Рассмотрим способы задания законов распределения системы случайных величин на примере двух дискретных случайных величин X и Y, область возможных значений которых задана соответственно дискретными множествами $x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$ и $y_1, y_2, y_3, \ldots, y_j, \ldots$ $\ldots, y_m$. Для рассматриваемой системы двух случайных величин возможные исходы характеризуются совместным появлением двух значений случайных величин $x_i y_j$, где i принимает значения от 1 до n, а j от 1 до m. Общее число возможных исходов будет равно $n \cdot m$. Для вероятностного описания системы двух случайных величин должны быть заданы все совместные вероятности, характеризующие вероятность того, что случайные величины X и Y примут значения, соответственно равные x_i и y_j:

$$P(X = x_i;\ Y = y_j) = p_{ij}. \tag{2.14}$$

Закон распределения системы случайных величин, устанавливающий соответствие между возможными их значениями x_i и y_j и вероятностью появления этих значений, может быть задан в табличном виде (табл. 2.7).

Таблица 2.7

Закон распределения системы случайных величин

$Y \backslash X$	x_1	x_2	…	x_i	…	x_n
y_1	p_{11}	p_{12}	…	p_{1i}	…	p_{1n}
y_2	p_{21}	p_{22}	…	p_{2i}	…	p_{2n}
…			…		…	
y_j	p_{j1}	p_{j2}	…	p_{ji}	…	p_{jn}
…	…	…	…	…	…	…
y_m	p_{m1}	p_{m2}	…	p_{mi}	…	p_{mn}

В соответствии с теорией сложения вероятностей закон распределения вероятностей двух дискретных величин должен удовлетворять следующим свойствам.

При суммировании совместных вероятностей (2.14) по всем возможным значениям x_i мы получим вероятность того, что случайная величина Y примет значение y_j:

$$P(Y = y_j) = \sum_{i=1}^{n} p_{ji}\,. \tag{2.15}$$

В соответствии с выражением (2.15) для получения вероятности появления значения y_2 случайной величины Y нужно просуммировать вероятности p_{2i}, входящие во вторую строку табл. 2.7.

Аналогично при суммировании совместных вероятностей (2.14) по всем возможным значениям y_j мы получим вероятность того, что случайная величина X примет значение x_i:

$$P(X = x_i) = \sum_{j=1}^{m} p_{ji}\,. \tag{2.16}$$

В соответствии с выражением (2.16) вероятность $P(X = x_1)$ можно получить суммированием всех вероятностей p_{ji}, входящих в первый столбец табл. 2.7.

При суммировании совместных вероятностей (2.14) по всем возможным значениям x_i и y_j в соответствии с условием нормировки мы получим единицу:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = \sum_{j=1}^{m} p(y_j) = \sum_{i=1}^{n} P(x_i) = 1.$$

В соответствии с теоремой умножения вероятностей по заданному закону распределения системы случайных величин X и Y (табл. 2.7) можно найти условные вероятности:

$$p(x_i / y_j) = \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)} = \frac{p_{ji}}{p(y_j)}; \tag{2.17}$$

$$p(y_j / x_i) = \frac{P(X = x_i; Y = y_j)}{P(X = x_i)} = \frac{p_{ji}}{p(x_i)}. \tag{2.18}$$

Условная вероятность $p(x_i/y_j)$ характеризует вероятность того, что случайная величина X примет значение x_i при условии, что случайная величина Y уже равна (приняла значение) y_j.

Сравнивая значения $p(x_i)$ и $p(y_j)$, вычисленные по формулам (2.15) и (2.16), с соответствующими значениями условных вероятностей $p(x_i/y_j)$ и $p(y_j/x_i)$, можно судить о том, являются ли случайные величины X и Y зависимыми между собой.

При $p(x_i) = p(x_i/y_j)$ или $p(y_j) = p(y_j/x_i)$ случайные величины X и Y независимы. Если указанные равенства не выполняются, то случайные величины будут зависимы друг от друга.

Условные вероятности (2.17) и (2.18) должны удовлетворять условию нормировки:

$$\sum_{i=1}^{n} p(x_i / y_j) = 1; \ \sum_{j=1}^{m} p(y_j / x_i) = 1.$$

Это свойство легко доказывается с учетом формул (2.15) — (2.18). Например,

$$\sum_{i=1}^{n} p(x_i / y_j) = \sum_{i=1}^{n} \frac{p_{ji}}{p(y_j)} = \frac{p(y_j)}{p(y_j)} = 1.$$

Знание условных вероятностей (2.17) и (2.18) дает возможность вычислить условное математическое ожидание и условную дисперсию:

$$\begin{aligned} m_x(y) &= m_{x/y_j} = \sum_{i=1}^{n} x_i p(x_i / y_j) = \sum_{i=1}^{n} \frac{x_i p_{ji}}{p(y_j)}; \\ m_y(x) &= m_{y/x_i} = \sum_{j=1}^{m} y_j p(y_j / x_i) = \sum_{j=1}^{m} \frac{y_j p_{ji}}{p(x_i)}; \end{aligned} \qquad (2.19)$$

$$\begin{aligned} D_x(y) &= D_{x/y_j} = \sum_{i=1}^{n} (x_i - m_{x/y_i})^2 p(x_i / y_j); \\ D_y(x) &= D_{y/x_i} = \sum_{j=1}^{m} (y_j - m_{y/x_i})^2 p(y_j / x_i). \end{aligned} \qquad (2.20)$$

Равенство $m_x(y) = m_{x/yi}$ говорит о том, что условное математическое ожидание $m_{x/yi}$ зависит от значений случайной величины Y. То есть условное математическое ожидание $m_x(y) = m_{x/yi}$ является функцией аргумента y. Аналогично условное математическое ожидание $m_y(x) = m_{y/xi}$ есть функция аргумента x, т.е. зависит от значений случайной величины X.

Уравнение $m_{x/yi} = m_x(y)$ называют ***уравнением регрессии*** случайной величины Y на случайную величину X. Аналогично уравнение $m_{y/xi} = m_y(x)$ называют уравнением регрессии случайной величины X на случайную величину Y.

Пример графического изображения линий регрессии, построенных по данным уравнениям, показан на рис. 2.10. Взаимное расположение линий регрессии характеризует степень зависимости одной случайной величины от другой. Если случайные величины X и Y независимы друг от друга, то выполняются равенства:

$$p(x_i/y_j) = p(x_i); \quad p(y_j/x_i) = p(y_j),$$

на основе которых видно, что условные математические ожидания равны безусловным:

$$m_{x/yi} = m_x; \; m_{y/xi} = m_y.$$

Для независимых случайных величин X и Y линии регрессии будут параллельны осям x и y. Помимо условных вероятностей и условных математических ожиданий, для количественной характеристики степени взаимосвязи между случайными величинами используют смешанный центральный момент второго порядка, который называют корреляционным моментом и определяют по формуле

$$k_{xy} = M[(X - m_x)(Y - m_y)] = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - m_x)(y_i - m_y)\,p_{ij}. \tag{2.21}$$

Если перемножить два выражения в скобках и провести усреднение полученных слагаемых, получим более удобную формулу для определения корреляционного момента:

$$k_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ji} - m_x \sum_{i=1}^{n}\sum_{j=1}^{m} y_j p_{ji} - m_y \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p_{ji} + m_x m_y.$$

Суммы во втором и третьем слагаемых данной формулы соответственно равны m_y и m_x, поэтому для корреляционного момента получим:

$$k_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ji} - m_x m_y. \tag{2.22}$$

Используя формулу произведения вероятностей для зависимых случайных величин

$$p_{ji} = p(x_i)\,p(y_j / x_i) = p(y_j)\cdot p(x_i / y_j)$$

в формуле (2.22), можно избавиться от двойной суммы:

$$k_{xy} = \sum_{i=1}^{n} x_i\, p(x_i) \sum_{j=1}^{m} y_j\, p(y_j / x_i) - m_x m_y.$$

Вторая сумма в первом слагаемом равна условному математическому ожиданию $m_{y/xi}$. С учетом этого для корреляционного момента получим более простую формулу:

$$k_{xy} = \sum_{i=1}^{n} x_i m_{y/x_i}\, p(x_i) - m_x m_y. \tag{2.23}$$

Аналогично можно показать, что

$$k_{xy} = \sum_{j=1}^{m} y_j m_{x/y_j}\, p(y_j) - m_x m_y. \tag{2.23 а}$$

Полученные формулы позволяют упростить вычисление корреляционного момента при известных значениях условных математических ожиданий.

Рассмотрим некоторые частные случаи.

1. Если случайные величины X и Y независимы друг от друга, то в соответствии с теоремой умножения вероятностей совместная вероятность p_{ji} будет равна произведению вероятностей $p(x_i)$ и $p(y_j)$. С учетом этого для корреляционного момента получим:

$$k_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p(x_i)\, p(y_j) - m_x m_y = \sum_{i=1}^{n} x_i p(x_i) \sum_{j=1}^{m} y_j p(y_j) - m_x m_y =$$
$$= m_x m_y - m_x m_y = 0.$$

Таким образом, для независимых случайных величин корреляционный момент равен нулю.

2. Если случайная величина Y имеет прямую зависимость от X, т.е. $Y = X$, то в этом случае $m_y = m_x$. Производя в формуле (2.21) соответствующие подстановки для корреляционного момента, получим:

$$\begin{aligned} k_{y=x} &= M[(X - m_x)(X - m_x)] = M[(X - m_x)^2] = \\ &= \sum_{i=1}^{n} (x_i - m_x)^2 P(x_i) = D_x. \end{aligned} \tag{2.24}$$

Таким образом, при ($Y = X$) корреляционный момент линейно зависимых случайных величин X и Y имеет максимальное значение, равное дисперсии.

3. Когда случайные величины X и Y связаны обратной линейной зависимостью, например $X = -Y$, то в отличие от предыдущего случая увеличение значений случайной X приводит к уменьшению случайной величины Y. Для математических ожиданий можно записать: $m_x = -m_y$. Производя в формуле (2.21) соответствующие подстановки для корреляционного момента, получим:

$$k_{x=-y} = M[(-Y+m_y)(Y-m_y)] = -M[(Y-m_y)^2] = \\ = -\sum_{j=1}^{m} y_j P(y_j) = -D_y. \qquad (2.25)$$

Таким образом, при обратной линейной зависимости случайных величин ($X = -Y$) корреляционный момент отрицательный и имеет минимальное значение, по абсолютной величине равное дисперсии.

Из рассмотренных частных случаев наглядно видно, что корреляционный момент дает количественную оценку степени линейной взаимосвязи между значениями случайных величин. Из формул (2.22) и (2.23) следует также, что корреляционный момент зависит от дисперсий (среднеквадратичных отклонений) случайных величин. Чем больше разброс возможных значений случайных величин X и Y от математических ожиданий, тем больше по абсолютной величине максимальное и минимальное значение корреляционного момента.

Для того чтобы исключить зависимость корреляционного момента от среднеквадратичных отклонений случайных величин и оставить только количественную зависимость, характеризующую их степень линейной взаимосвязи, используют понятие коэффициента корреляции ρ_{xy}, который определяется как нормированное к среднеквадратичным отклонениям σ_x и σ_y значение корреляционного момента:

$$\rho_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y} = \frac{1}{\sigma_x \sigma_y}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ji} - m_x m_y\right]. \qquad (2.26)$$

Из формул (2.24) — (2.25) следует, что коэффициент корреляции может принимать значения в интервале $-1 \le \rho \le 1$.

Если случайные величины X и Y независимы, то $\rho_{xy} = 0$.

Можно показать, что при $Y = aX + b$ (прямая линейная зависимость) $\rho_{xy} = 1$, а при $Y = -aX + b$ (обратная линейная зависимость) $\rho_{xy} = -1$.

Пример 2.5. Закон распределения системы двух дискретных случайных величин X и Y приведен в табл. 2.8. Необходимо:

1. Определить закон распределения случайных величин X и Y.
2. Выяснить по условным вероятностям, являются ли случайные величины зависимыми или независимыми.
3. Определить условные математические ожидания и построить линии регрессии одной случайной величины на другую.
4. Вычислить коэффициент корреляции случайных величин X и Y.

Таблица 2.8

Закон распределения дыух случайных величин

i	1	2	3	4	5
y_j \ x_i	2	2,5	3	3,5	4
$y_1 = 0$	0,023	0,015	0,01	0,007	0,005
$y_2 = 1$	0,05	0,032	0,020	0,010	0,008
$y_3 = 2$	0,07	0,048	0,030	0,020	0,012
$y_4 = 3$	0,083	0,066	0,045	0,03	0,016
$y_5 = 4$	0,10	0,1	0,1	0,06	0,04

Решение. 1. Для определения законов распределения случайной величины X воспользуемся формулой (2.16). Для вероятности $P(X = 2)$ получим:

$$P(X=2)=P(x_1)=\sum_{j=1}^{5}p_{j1}=0,023+0,05+0,07+0,083+0,1=0,326.$$

Производя аналогичные вычисления для $P(x_2)$, $P(x_3)$, $P(x_4)$ и $P(x_5)$, полученные значения занесем в табл. 2.9, которая определяет закон распределения случайной величины X.

Таблица 2.9

Закон распределения случайной величины

i	1	2	3	4	5
x_i	2	2,5	3	3,5	4
$P(x_i)$	0,326	0,261	0,205	0,127	0,081

При правильном вычислении $P(x_i)$ они должны удовлетворять условию нормировки $\sum_{i=1}^{n}P(x_i)=1$. Вычислим значения математического ожидания и дисперсии случайной величины X соответственно по формулам (2.9) и (2.11):

$$m_x=\sum_{i=1}^{5}x_iP(x_i)=2\cdot 0,326+2,5\cdot 0,261+3\cdot 0,205+3,5\cdot 0,127+4\cdot 0,081=2,688;$$

$$D_x=\sum_{i=1}^{5}x_i^2P(x_i)-m_x^2=4\cdot 0,326+6,25\cdot 0,261+9\cdot 0,205+12,25\cdot 0,127+$$

$$+16\cdot 0,081-(2,688)^2\approx 0,407;$$

$$\sigma_x=\sqrt{D_x}\approx 0,638.$$

Вероятности $P(y_j)$, определяющие закон распределения случайной величины Y, вычисляются по формуле (2.15). Для вероятности $P(y_1)$ получим:

$$P(y_1) = P(Y = 0) = \sum_{i=1}^{5} p_{ij} = 0,023 + 0,015 + 0,01 + 0,007 + 0,005 = 0,06 .$$

Аналогично вычисляются вероятности всех возможных значений случайной величины Y. Результаты вычислений приведены в табл. 2.10.

Таблица 2.10

Значение случайной величины Y

j	1	2	3	4	5
y_j	0	1	2	3	4
$P(y_j)$	0,06	0,12	0,18	0,24	0,4

Из таблицы следует, что для полученных значений $P(y_j)$ условие нормировки выполняется.

Вычислим математическое ожидание, дисперсию и среднеквадратичное отклонение случайной величины Y:

$$m_y = \sum_{j=1}^{5} y_j P(y_j) = 0 \cdot 0,06 + 1 \cdot 0,12 + 2 \cdot 0,18 + 3 \cdot 0,24 + 4 \cdot 0,4 = 2,8;$$

$$D_y = \sum_{j=1}^{5} y_j^2 P(y_j) - m_y^2 = 0 \cdot 0,06 + 1 \cdot 0,12 + 4 \cdot 0,18 +$$

$$+ 9 \cdot 0,24 + 16 \cdot 0,4 - (2,8)^2 \approx 1,56;$$

$$\sigma_y = \sqrt{D_y} \approx 1,249.$$

2. Вычислим условные вероятности $P(x_i/y_j)$. Для этого воспользуемся формулой (2.17) и табл. 2.8 и 2.10:

$$P(x_1 / y_1) = \frac{p_{11}}{P(y_1)} = \frac{0,023}{0,06} \approx 0,383;$$

$$P(x_1 / y_2) = \frac{p_{21}}{P(y_2)} = \frac{0,05}{0,12} \approx 0,417;$$

$$P(x_1 / y_3) = \frac{p_{31}}{P(y_3)} = \frac{0,07}{0,18} \approx 0,389;$$

$$P(x_1 / y_4) = \frac{p_{41}}{P(y_4)} = \frac{0,083}{0,24} \approx 0,346;$$

$$P(x_1 / y_5) = \frac{p_{51}}{P(y_5)} = \frac{0,1}{0,4} \approx 0,25.$$

Аналогично вычисляются и другие условные вероятности. Результаты вычисления $P(x_i/y_j)$ по формуле (2.17) и $P(y_j/x_i)$ по формуле (2.18) представлены в табл. 2.11 и 2.12.

Таблица 2.11

Значения условных вероятностей $P(x_i/y_j)$

y_j \ x_i	$x_1 = 2$	$x_2 = 2,5$	$x_3 = 3$	$x_4 = 3,5$	$x_5 = 4$
$y_1 = 0$	0,383	0,25	0,167	0,117	0,083
$y_2 = 1$	0,417	0,266	0,167	0,083	0,067
$y_3 = 2$	0,389	0,267	0,167	0,111	0,066
$y_4 = 3$	0,346	0,275	0,187	0,125	0,067
$y_5 = 4$	0,25	0,25	0,25	0,15	0,1

Таблица 2.12

Значения условных вероятностей $P(y_j/x_i)$

x_i \ y_j	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$y_4 = 3$	$y_5 = 4$
$x_1 = 2$	0,0706	0,1534	0,2147	0,2546	0,3067
$x_2 = 2,5$	0,0575	0,1226	0,1839	0,2529	0,3831
$x_3 = 3$	0,0488	0,0976	0,1463	0,2195	0,4878
$x_4 = 3,5$	0,0551	0,0787	0,1575	0,2362	0,4725
$x_5 = 4$	0,0617	0,0988	0,1482	0,1975	0,4938

Правильность вычисления $P(x_i/y_j)$ и $P(y_j/x_i)$ можно проверить по условию нормировки, в соответствии с которым сумма условных вероятностей по каждой из строк табл. 2.11 и 2.12 должна быть равна единице.

Из полученных значений условных вероятностей слеудет, что $P(x_i/y_j)$ имеет различные значения для различных значений y_i и не выполняется условие

$$P(x_i/y_j) = P(x_i);\ P(y_j/x_i) = P(y_j).$$

Из этого следует, что случайные величины X и Y являются зависимыми.

3. Для определения условных математических ожиданий воспользуемся формулами (2.19) и результатами расчетов условных вероятностей, представленных в табл. 2.11 и 2.12:

$$m_{x/y_1} = \sum_{i=1}^{5} x_i P(x_i / y_1) = 2 \cdot 0{,}383 + 2{,}5 \cdot 0{,}25 + 3 \cdot 0{,}167 + 3{,}5 \cdot 0{,}117 +$$
$$+4 \cdot 0{,}083 \approx 2{,}633;$$

$$m_{x/y2} = \sum_{i=1}^{5} x_i P(x_i / y_2) = 2 \cdot 0{,}417 + 2{,}5 \cdot 0{,}266 + 3 \cdot 0{,}167 + 3{,}5 \cdot 0{,}083 +$$
$$+4 \cdot 0{,}067 \approx 2{,}561.$$

Используя таблицу 2.11, аналогично вычисляются остальные значения условного математического ожидания $m_x(y)$. Результаты расчетов приведены в табл. 2.13.

Таблица 2.13

Математическое ожидание $m_x(y)$

j	1	2	3	4	5
y_j	0	1	2	3	4
$m_x(y)$	2,6335	2,5585	2,599	2,646	2,8

Расчеты условного математического ожидания $m_y(x)$ производятся аналогично с использованием табл. 2.12. Полученные результаты приведены в табл. 2.14.

Таблица 2.14

Математическое ожидание $m_x(x)$

i	1	2	3	4	5
x_j	2	2,5	3	3,5	4
$m_y(x)$	2,5734	2,7815	2,9999	2,9923	2,9629

По результатам расчетов, представленных в табл. 2.13 и 2.14, на рис. 2.10 приведено графическое изображение линий регрессии случайных величин X и Y. По взаимному расположению линий регрессии можно предположить, что между случайными величинами X и Y наблюдается очень слабая взаимная связь:

$$k_{xy} = \sum_{j=1}^{m} y_j m_{x/y_j} P(y_j) - m_x \cdot m_y = 1 \cdot 2{,}5585 \cdot 0{,}12 + 2 \cdot 2{,}599 \cdot 0{,}18 +$$
$$+3 \cdot 2{,}646 \cdot 0{,}24 + 4 \cdot 2{,}8 \cdot 0{,}4 - 2{,}688 \cdot 2{,}8 \approx 0{,}101.$$

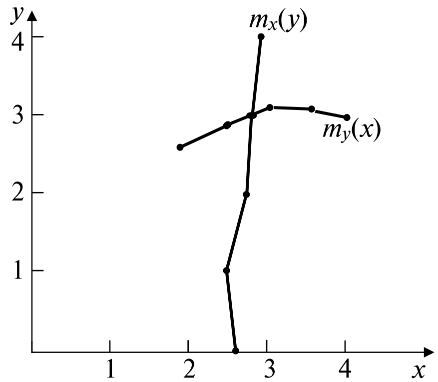


Рис. 2.10. Графическое изображение линий регрессии случайных величин

4. Для количественной оценки степени линейной взаимосвязи между случайными величинами X и Y вычислим коэффициент корреляции ρ_{xy}. Для этого сначала вычислим по формуле (2.23) значение корреляционного момента. При вычислениях k_{xy} воспользуемся результатами вычислений, приведенных в табл. 2.10 и 2.13:

С учетом вычисленных значений среднеквадратичных отклонений для коэффициента корреляции получим:

$$\rho_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y} = \frac{0{,}101}{0{,}638 \cdot 1{,}249} \approx 0{,}127 .$$

Так как ρ_{xy} имеет значение, близкое к нулю, можно сделать вывод о том, что случайные величины X, Y слабо зависимы друг от друга.

2.7. Примеры решения задач

Задача 2.1. Инвестиционная компания приобрела пакеты акций трех предприятий. Вероятности получения дохода по пакетам акций первого, второго и третьего предприятия соответственно равны 0,6, 0,65 и 0,7. Найти закон распределения дохода инвестиционной компании, если ожидаемый доход по каждому пакету приобретенных акций равен 1 тыс. дол. Построить функцию распределения данной случайной величины, определить математическое ожидание, дисперсию и среднеквадратическое отклонение.

Решение. Обозначим события A_1, A_2 и A_3, состоящие в том, что получен доход по акциям соответственно первого, второго и третьего предприятия. Вероятности наступления этих событий по условию задачи равны:

$$P(A_1) = 0{,}6;\ P(A_2) = 0{,}65;\ P(A_3) = 0{,}7.$$

Вероятности противоположных событий $\overline{A}_1$; $\overline{A}_2$ и $\overline{A}_3$ (неполучения дохода) будут равны:

$$P(\overline{A}_1) = 1 - P(A_1) = 1 - 0{,}6 = 0{,}4;$$

$$P(\overline{A}_2) = 1 - P(A_2) = 1 - 0{,}65 = 0{,}35;$$

$$P(\overline{A}_3) = 1 - P(A_3) = 1 - 0{,}7 = 0{,}3.$$

Искомая случайная величина x может принимать значения 0, 1, 2 и 3 тыс. дол.

При $x = 0$ инвестиционная компания не получит доход ни по одному из трех пакетов акций. При $x = 1$ инвестиционная компания получит доход по одному пакету акций из трех приобретенных. При $x = 2$ инвестиционная компания получит доход по двум из трех пакетов акций. При $x = 3$ инвестиционная компания получит доход по всем трем приобретенным пакетам акций. Вероятности всех возможных значений случайной величины x определяются исходными вероятностями $P(A_i)$ и $P(\overline{A}_i)$ при $i = 1, 2, 3$:

$$P(x=0) = P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot P(\overline{A}_3) = 0{,}4\cdot 0{,}35\cdot 0{,}3 = 0{,}042;$$

$$P(x=1) = P(A_1)\cdot P(\overline{A}_2)\cdot P(\overline{A}_3) + P(\overline{A}_1)\cdot P(A_2)\cdot P(\overline{A}_3) + P(\overline{A}_1)\times$$

$$\times P(\overline{A}_2)\cdot P(A_3) = 0{,}6\cdot 0{,}35\cdot 0{,}3 + 0{,}4\cdot 0{,}65\cdot 0{,}3 + 0{,}4\cdot 0{,}35\cdot 0{,}7 = 0{,}239;$$

$$P(x=2) = P(A_1)\cdot P(A_2)\cdot P(\overline{A}_3) + P(A_1)\cdot P(\overline{A}_2)\cdot P(A_3) +$$

$$+P(\overline{A}_1)\cdot P(A_2)\cdot P(A_3) = 0{,}6\cdot 0{,}65\cdot 0{,}3 + 0{,}6\cdot 0{,}35\cdot 0{,}7 + 0{,}4\cdot 0{,}65\cdot 0{,}7 = 0{,}446;$$

$$P(x=3) = P(A_1)\cdot P(A_2)\cdot P(A_3) = 0{,}6\cdot 0{,}65\cdot 0{,}7 = 0{,}273.$$

Закон распределения дохода инвестиционной компании может быть представлен рядом распределения, приведенным в табл. 2.15.

Таблица 2.15

Закон распределения дохода иньестиционной компании

x_i, тыс. дол.	0	1	2	3
$P(x_i)$	0,042	0,239	0,446	0,273

Правильность полученного закона распределения подтверждается выполнением условия нормировки вероятностей $\sum_{i=0}^{3} P(x_i) = 1$.

В соответствии с определением функции распределения $F(x)$ получим:

$$
\begin{array}{lll}
\text{при} & x<0 & F(x)=0; \\
 & 0\le x<1 & F(x)=0,042; \\
 & 1\le x<2 & F(x)=0,281; \\
 & 2\le x<3 & F(x)=0,727; \\
 & x\ge 1 & F(x)=1.
\end{array}
$$

График функции распределения приведен на рис. 2.11.

Математическое ожидание дохода инвестиционной компании находим по формуле (2.9):

$$m_x=\sum_{i=0}^{3}x_iP(x_i)=0\cdot 0,042+1\cdot 0,239+2\cdot 0,446+3\cdot 0,273=1,95 \text{ тыс. дол.}$$

Дисперсию дохода инвестиционной компании определим по формуле (2.11):

$$D(x)=\sum_{i=0}^{3}x_i^2P(x_i)-m_x^2=0\cdot 0,042+1\cdot 0,239+4\cdot 0,446+$$
$$+9\cdot 0,273-(1,95)^2=0,6775\,[\text{тыс. дол.}]^2.$$

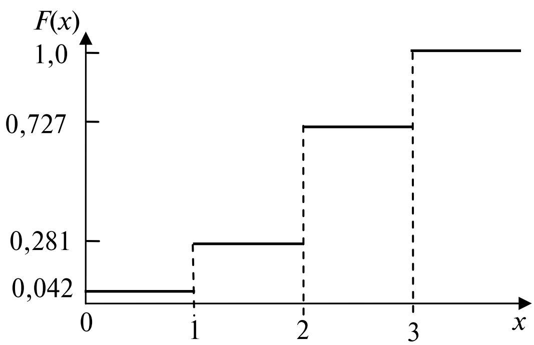


Рис. 2.11. График функции распределения

Среднее квадратичное отклонение дохода инвестиционной компании от математического ожидания будет равно:

$$\sigma_x=\sqrt{D(x)}=\sqrt{0,6775}\approx 0,823 \text{ тыс. дол.}$$

Задача 2.2. Составить закон распределения прибыли П = В – З, если выручка В и затраты З независимы и заданы распределениями:

i	1	2	3	i	1	2	3
$В_i$	4	5	6	$З_i$	2	3	4
P_i	0,25	0,4	0,35	P_i	0,3	0,5	0,2

Определить математическое ожидание прибыли и среднее квадратичное отклонение прибыли.

Решение. При заданных значениях выручки и затрат прибыль может принимать следующие значения:

$$П_0 = В_1 - З_3 = 4 - 4 = 0;$$
$$П_1 = В_1 - З_2 = 4 - 3 = 1 \text{ или } П_1 = В_2 - З_3 = 5 - 4 = 1;$$
$$П_2 = В_1 - З_1 = 4 - 2 = 2 \text{ или } П_2 = В_3 - З_3 = 6 - 4 = 2;$$
$$П_3 = В_2 - З_1 = 5 - 2 = 3 \text{ или } П_3 = В_3 - З_2 = 6 - 3 = 3;$$
$$П_4 = В_3 - З_1 = 4.$$

Так как по условию задачи выручка и затраты независимы, то вероятности указанных значений прибыли будут равны:

$$P(П_0) = P(В_1) \cdot P(З_3) = 0,25 \cdot 0,2 = 0,05;$$
$$P(П_1) = P(В_1) \cdot P(З_2) + P(В_2) \cdot P(З_3) = 0,25 \cdot 0,5 + 0,4 \cdot 0,2 = 0,205;$$
$$P(П_2) = P(В_1) \cdot P(З_1) + P(В_2) \cdot P(З_2) + P(В_3) \cdot P(З_3) =$$
$$= 0,25 \cdot 0,3 + 0,4 \cdot 0,5 + 0,35 \cdot 0,2 = 0,345;$$
$$P(П_3) = P(В_2) \cdot P(З_1) + P(В_3) \cdot P(З_2) = 0,4 \cdot 0,3 + 0,35 \cdot 0,5 = 0,295;$$
$$P(П_4) = P(В_3) \cdot P(З_1) = 0,35 \cdot 0,3 = 0,105.$$

Закон распределения прибыли в виде ряда распределения приведен в табл. 2.16.

Таблица 2.16

Закон распределения прибыли

$П_i$	0	1	2	3	4
P_i	0,05	0,205	0,345	0,295	0,105

Для определения математического ожидания прибыли можно воспользоваться формулой (2.9) или свойством 3 математического ожидания.

Так как П = В – З, то математическое ожидание прибыли можно найти как разность математического ожидания выручки и математического ожидания затрат:

$$m_{\text{П}} = m_{\text{В}} - m_{\text{З}};$$

$$m_{\text{В}} = \sum_{i=1}^{3} \text{В}_i \, P_i = 4 \cdot 0,25 + 5 \cdot 0,4 + 6 \cdot 0,35 = 5,1;$$

$$m_{\text{З}} = \sum_{i=1}^{3} \text{З}_i \, P_i = 2 \cdot 0,3 + 3 \cdot 0,5 + 4 \cdot 0,2 = 2,9;$$

$$m_{\text{П}} = 5,1 - 2,9 = 2,2.$$

По формуле (2.9) и табл. 2.16 для математического ожидания прибыли получим:

$$m_{\text{П}} = \sum_{i=0}^{4} \text{П}_i P_i = 0 \cdot 0,05 + 1 \cdot 0,205 + 2 \cdot 0,345 +$$

$$+3 \cdot 0,295 + 4 \cdot 0,105 = 2,2.$$

Как видно, оба варианта расчета дают одинаковые результаты.

Для определения среднеквадратичного отклонения прибыли определим дисперсию прибыли. Дисперсию также можно определить двумя способами: по формуле (2.11) и табл. 2.16 или по свойству 3 дисперсии, в соответствии с которым

$$D(\text{П}) = D(\text{В}) + D(\text{З}).$$

Определим дисперсию выручки и дисперсию затрат:

$$D(\text{В}) = \sum_{\text{i}=1}^{3} \text{В}_i^2 \; P_i - (m_{\text{В}})^2 = 16 \cdot 0,25 + 25 \cdot 0,4 + 36 \cdot 0,35 - (5,1)^2 = 0,59;$$

$$D(\text{З}) = \sum_{\text{i}=1}^{3} \text{З}_i^2 \; P_i - (m_{\text{З}})^2 = 4 \cdot 0,3 + 9 \cdot 0,5 + 16 \cdot 0,2 - (2,9)^2 = 0,49.$$

Дисперсия и среднеквадратичное отклонение прибыли будут равны:

$$D(\text{П}) = D(\text{В}) + D(\text{З}) = 0,59 + 0,49 = 1,08;$$

$$\sigma_{\text{П}} = \sqrt{D(\text{П})} = \sqrt{1,08} \approx 1,039.$$

Определим дисперсию и среднеквадратичное отклонение прибыли по формуле (2.11) и табл. 2.16:

$$D(\Pi) = \sum_{i=0}^{4} \Pi_i^2 \, P_i - (m_\Pi)^2 = 0 \cdot 0{,}05 + 1 \cdot 0{,}205 + 4 \cdot 0{,}345 +$$

$$+ \, 9 \cdot 0{,}295 + 16 \cdot 0{,}105 - (2{,}2)^2 = 1{,}08;$$

$$\sigma_\Pi = \sqrt{D(\Pi)} = \sqrt{1{,}08} \approx 1{,}039.$$

Из полученных значений следует, что оба способа расчета дают одинаковые результаты.

Задача 2.3. Фирма, занимающаяся установкой оконных стеклопакетов, раскладывает рекламные листки по почтовым ящикам. По данным отдела маркетинга фирмы, число поступивших заказов по данной рекламной кампании подчиняется закону распределения Пуассона с параметром распределения $\lambda = 9$. Определить математическое ожидание числа заказов, полученных фирмой, и вероятность того, что число заказов будет не менее десяти.

Решение. В данной задаче случайной величиной m является число заказов, полученных фирмой по данной рекламной кампании, которая распределена по закону Пуассона:

$$P(m) = \frac{\lambda^m}{m!} e^{-\lambda}; \; \lambda = 9.$$

Математическое ожидание случайной величины, распределенной по закону распределения Пуассона, равно

$$M(m) = \lambda = 9.$$

Вероятность того, что число заказов, полученных фирмой, будет не менее десяти, определится по формуле

$$P(m \geq 10) = 1 - \sum_{m=0}^{9} P(m) = 1 - \sum_{m=0}^{9} \frac{\lambda^m}{m!} e^{-\lambda}.$$

Для вычисления суммы, входящей в данную формулу, можно воспользоваться таблицей значений функции Пуассона в приложении 4 при $\lambda = 9$ и $m \leq 9$.

$$P(m \geq 10) = 1 - 0{,}000123 - 0{,}001111 - 0{,}004998 - 0{,}014994 - 0{,}033737 -$$
$$- \, 0{,}060727 - 0{,}09109 - 0{,}11716 - 0{,}131756 \approx 0{,}4125.$$

Задача 2.4. Экзаменатор задает студенту вопросы, пока тот отвечает правильно. Как только студент ответит неправильно, экзаменатор прекращает задавать вопросы. Вероятность правильного ответа студента на вопрос равна 0,8. Определить математическое ожидание числа заданных студенту вопросов и вероятность того, что число заданных студенту вопросов будет не больше пяти.

Решение. В данной задаче случайная величина X принимает целочисленные значения $X \geq 1$ и является числом испытаний до первого появления интересующего нас события A. Таким образом, событием A по условию задачи является неправильный ответ студента на заданный вопрос. Случайная величина X подчиняется геометрическому распределению:

$$P(X = n) = q^{n-1} \cdot p;\ n = 1, 2, 3, 4, \ldots,$$

где p — вероятность появления события A в каждом из опытов; $q = 1 - p$ — вероятность противоположного события $\overline{A}$.

В условии задачи задана вероятность правильного ответа на вопрос, т.е. вероятность противоположного события $\overline{A}$:

$$q = 1 - p = 0{,}8 \Rightarrow p = 1 - q = 0{,}2.$$

Тогда вероятность числа заданных студенту вопросов определится по формуле

$$P(X = n) = 0{,}8^{n-1} \cdot 0{,}2;\ n = 1, 2, 3, 4, \ldots .$$

Математическое ожидание числа заданных вопросов при геометрическом распределении определяется формулой

$$m_x = \frac{1}{1-q} = \frac{1}{1-0{,}8} = 5.$$

Вероятность того, что число заданных студенту вопросов будет не больше пяти, определится суммой

$$P(X \leq 5) = \sum_{n=1}^{5} q^{n-1} \cdot p =$$
$$= 0{,}2(1 + 0{,}8 + 0{,}64 + 0{,}512 + 0{,}4096) = 0{,}67232.$$

Задача 2.5. Двумерный закон распределения дискретных случайных величин В — выручки от реализации и З — затрат на производство продукции — приведен в табл. 2.17. Найти вероятность того, что прибыль от реализации продукции будет больше или равна 3.

$$P(\text{П} = \text{В} - \text{З} \geq 3) = ?$$

Таблица 2.17

Распределение величины выручки и затрат на производство

$В_i$ \ $З_j$	$З_1 = 2$	$З_2 = 3$	$З_3 = 4$
$В_1 = 4$	0,05	0,1	0,05
$В_2 = 5$	0,05	0,2	0,05
$В_3 = 6$	0,2	0,2	0,1

Решение. Вычислим одномерные распределения вероятностей выручки $P(В_i)$ и затрат $P(З_j)$ по формулам (2.15) и (2.16):

$$P(В_1) = \sum_{j=1}^{3} P(В_1;\, З_j) = 0{,}05 + 0{,}1 + 0{,}05 = 0{,}2;$$

$$P(В_2) = \sum_{j=1}^{3} P(В_2;\, З_j) = 0{,}05 + 0{,}2 + 0{,}05 = 0{,}3;$$

$$P(В_3) = \sum_{j=1}^{3} P(В_3;\, З_j) = 0{,}2 + 0{,}2 + 0{,}1 = 0{,}5.$$

Результаты вычислений занесены в таблицу:

i	1	2	3
$В_i$	4	5	6
$P(В_i)$	0,2	0,3	0,5

Для распределения вероятностей затрат получим:

$$P(З_1) = \sum_{i=1}^{3} P(В_i;\, З_1) = 0{,}05 + 0{,}05 + 0{,}2 = 0{,}3;$$

$$P(З_2) = \sum_{i=1}^{3} P(В_i;\, З_2) = 0{,}1 + 0{,}2 + 0{,}2 = 0{,}5;$$

$$P(З_3) = \sum_{i=1}^{3} P(В_i;\, З_3) = 0{,}05 + 0{,}05 + 0{,}1 = 0{,}2.$$

Результаты вычислений занесены в таблицу:

j	1	2	3
$З_j$	2	3	4
$P(З_j)$	0,3	0,5	0,2

Вычислим условные вероятности $P(B_i/З_j)$ по формуле (2.17):

$$P(\text{В}_1/\text{З}_1)=\frac{P(\text{В}_1;\text{З}_1)}{P(\text{З}_1)}=\frac{0{,}05}{0{,}3}=\frac{1}{6};$$

$$P(\text{В}_2/\text{З}_1)=\frac{P(\text{В}_2;\text{З}_1)}{P(\text{З}_1)}=\frac{0{,}05}{0{,}3}=\frac{1}{6};$$

$$P(\text{В}_3/\text{З}_1)=\frac{P(\text{В}_3;\text{З}_1)}{P(\text{З}_1)}=\frac{0{,}2}{0{,}3}=\frac{2}{3}.$$

Аналогично рассчитываются условные вероятности $P(\text{В}_i/\text{З}_2)$, $i = 1, 2, 3$ и $P(\text{В}_i/\text{З}_3)$, $i = 1, 2, 3$.

Результаты расчетов условных вероятностей $P(\text{В}_i/\text{З}_j)$ сведены в табл. 2.19.

Таблица 2.19

Результаты расчетов

В_i \ З_j	$\text{З}_1 = 2$	$\text{З}_2 = 3$	$\text{З}_3 = 4$
$\text{В}_1 = 4$	1/6	0,2	0,25
$\text{В}_2 = 5$	1/6	0,4	0,25
$\text{В}_3 = 6$	2/3	0,4	0,5
Σ	1	1	1

Из приведенных расчетов видно, что выручка от реализации В_i и затраты на производство З_j продукции являются зависимыми дискретными случайными величинами, так как $P(\text{В}_i/\text{З}_j) \neq P(\text{В}_i)$.

По данным таблицы 2.19 вычислим условные вероятности того, что прибыль будет больше или равна 3, при различных значениях затрат:

$$P(\Pi\geq 3/\text{З}_1)=P(\text{В}_2/\text{З}_1)+P(\text{В}_3/\text{З}_1)=1/6+\frac{2}{3}=\frac{5}{6};$$

$$P(\Pi\geq 3/\text{З}_2)=P(\text{В}_3/\text{З}_2)=0{,}4;$$

$$P(\Pi\geq 3/\text{З}_3)=0.$$

Безусловная вероятность того, что прибыль будет больше или равна 3, определяется по формуле

$$P(\Pi\geq 3)=\sum_{j=1}^{3}P(\text{З}_j)P(\Pi\geq 3/\text{З}_j)=\frac{5}{6}\cdot 0{,}3+0{,}4\cdot 0{,}5+0\cdot 0{,}2=0{,}45.$$

Вопросы для самоконтроля

1. Дайте определение дискретной случайной величины. Поясните область возможных значений. Сформулируйте понятия конечное и бесконечное множество возможных значений дискретной случайной величины.
2. Докажите закон распределения вероятностей дискретных случайных величин. Дайте понятие ряду распределения, многоугольнику распределения, аналитическому заданию закона распределения вероятностей. Приведите условие нормировки.
3. Каким образом определяют интегральную функцию распределения дискретной случайной величины? Раскройте смысл табличного и графического задания функции распределения дискретной случайной величины.
4. Каковы свойства интегральной функции распределения? Дайте определение вероятности попадания дискретной случайной величины в заданный интервал.
5. Дайте определение математическому ожиданию дискретной случайной величины. Объясните сущность понятия математического ожидания.
6. Сформулируйте свойства математического ожидания случайных величин.
7. Докажите, что математическое ожидание разности двух независимых случайных величин равно разности их математических ожиданий.
8. Сформулируйте понятия «дисперсия» и «среднеквадратичное отклонение дискретных случайных величин». Объясните сущности этих понятий.
9. Какие свойства дисперсии случайных величин вы знаете?
10. Докажите, что дисперсия разности двух независимых случайных величин равна сумме дисперсий этих случайных величин.
11. Дайте определение моментам распределения случайных величин и коэффициенту асимметрии закона распределения.
12. Сформулируйте равномерный закон распределения дискретных случайных величин. Дайте определение дефинициям: ряд распределения, функция распределения, математическое ожидание и дисперсия.
13. Сформулируйте биноминальный закон распределения. Дайте понятие аналитическому заданию, графику многоугольника

распределения и функции распределения. Приведите формулы для математического ожидания, дисперсии, коэффициента асимметрии.

14. Геометрическое распределение. Аналитическое задание. Графики многоугольника распределения и функции распределения.
15. Поясните распределение Пуассона. Аналитическое задание. Графики многоугольника и функции распределения. Математическое ожидание и дисперсия.
16. Сформулируйте закон распределения системы двух случайных величин. Поясните порядок вычисления одномерных законов распределения каждой из случайных величин. Сформулируйте условие нормировки двумерного распределения.
17. Каким образом определить условные вероятности по заданному закону распределения двух случайных величин? Сформулируйте условие независимости двух случайных величин. Сформулируйте условие нормировки для условных вероятностей.
18. Дайте определение условному математическому ожиданию и условной дисперсии случайной величины. Поясните особенности уравнения регрессии и линии регрессии одной случайной величины на другую. Поясните условие независимости двух случайных величин.
19. В чем заключается корреляционный момент случайных величин? Каким образом вычисляется корреляционный момент по двумерному закону распределения и по условным математическим ожиданиям?
20. Поясните определение корреляционного момента для трех частных случаев. Что характеризует корреляционный момент? Какова зависимость корреляционного момента от дисперсий случайных величин?
21. Что такое коэффициент корреляции случайных величин? Какой смысл введения нормировки корреляционного момента? Какова область возможных значений коэффициента корреляции, что он характеризует?

ГЛАВА 3

НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.1. Основные понятия о непрерывных случайных величинах

Напомним еще раз общее определение случайных величин. ***Случайной величиной*** называют такую переменную величину, которая в результате опыта может принимать то или иное заранее не известное значение.

В отличие от дискретных случайных величин, для которых можно заранее указать все возможные значения, для непрерывных случайных величин этого сделать невозможно, даже если область возможных значений непрерывной случайной величины ограничена некоторым интервалом $x_{\min} \le X \le x_{\max}$. Число возможных значений непрерывной случайной величины будет бесконечно большим. Каждое возможное значение непрерывной случайной величины можно охарактеризовать положением точки, находящейся на отрезке длиной $x_{\max} - x_{\min}$. Так как точка на числовой оси имеет длину, равную нулю, то на отрезке $x_{\max} - x_{\min}$ располагается бесконечно большое число точек, определяющих возможные значения непрерывной случайной величины. Область возможных значений непрерывной случайной величины может быть ограниченной или неограниченной при $x_{\min} \to -\infty$ и (или) $x_{\max} \to \infty$, но число возможных ее значений будет бесконечно большим.

Примерами непрерывных случайных величин могут служить случайный момент времени изменения курса валют на валютной бирже, случайное время нахождения в плавании транспортного судна, осуществляющего транспортировку интересующего нас груза, объем топлива, затраченного на эту транспортировку, и его стоимость, себестоимость единицы выпускаемой продукции и т.п. Любые физические и экономические величины имеют свои единицы измерения, и при их оценке, как правило, осуществляется округление получаемых значений до принятых единиц измерения. Это создает иллюзию дискретности случайных величин, хотя по своей природе они являются непрерывными случайными величинами.

То, что непрерывные случайные величины имеют бесконечное число возможных значений, накладывает особенности на задание их законов распределения вероятностей. Для дискретных случайных величин закон распределения вероятностей определяет вероятности появления каждого возможного значения случайной величины $P(x_i)$. В соответствии с условием нормировки сумма этих вероятностей равна единице:

$$\sum_{i=1}^{n} P(x_i) = 1,$$

где n — число возможных значений случайной величины.

Условие нормировки должно выполняться и для непрерывных случайных величин, для которых $n \to \infty$. Для того чтобы сумма бесконечного числа слагаемых при $n \to \infty$ была равна единице, сами эти слагаемые $P(x_i)$ должны иметь бесконечно малые значения, практически равные нулю. Таким образом, задание закона распределения непрерывной случайной величины в виде вероятностей появления каждого возможного ее значения не имеет смысла, так как вероятность того, что непрерывная случайная величина примет некоторое значение x_i, будет равна нулю.

Вследствие данного факта для непрерывных случайных величин закон распределения задается в виде распределения плотности вероятности непрерывных случайных величин.

3.2. Плотность вероятности и функция распределения непрерывных случайных величин

Рассмотрим некоторую непрерывную случайную величину, область возможных значений которой ограничена значениями $x_{\min} \le X < x_{\max}$ (рис. 3.1).

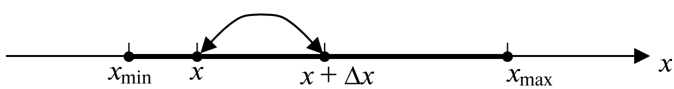


Рис. 3.1. Область возможных значений

В соответствии с условием нормировки вероятность того, что случайная величина X попадет в интервал от $x_{\min}$ до $x_{\max}$, равна единице:

$$P(x_{\min} \le X < x_{\max}) = 1.$$

Рассмотрим вероятность попадания случайной величины X на интервал от x до $x + \Delta x$. Число всех возможных значений непрерывной случайной величины на интервале $x_{\min} \div x_{\max}$ будет бесконечно большим $n \to \infty$. Число m значений непрерывной случайной величины, принадлежащих интервалу Δx, также будет бесконечно большим $m \to \infty$. Так как $\Delta x < x_{\max} - x_{\min}$, то, сравнивая эти две бесконечно большие величины, можно записать $m < n$. По аналогии с дискретной случайной величиной для вероятности попадания непрерывной случайной величины X на интервал Δx запишем:

$$P(x < X \le x + \Delta x) = \sum_{i=1}^{m\to\infty} P(x_i) < 1.$$

Под знаком суммы стоят бесконечно малые величины $P(x_i) \to 0$, число слагаемых в сумме равно бесконечности, поэтому результат суммирования дает конечное число, равное искомой вероятности $P(x \le \le X \le x + \Delta x)$. Причем эта вероятность будет меньше единицы, так как $\Delta x < x_{\max} - x_{\min}$ и $m < n$. Значение вероятности $P(x \le X < x + \Delta x)$ в общем случае может зависеть как от интервала Δx, так и от координат точки x, определяющей положение интервала Δx на отрезке $x_{\min} \div x_{\max}$.

Плотность распределения вероятностей $W(x)$ непрерывной случайной величины по области возможных значений можно характеризовать пределом отношения

$$\lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x} = W(x); \quad x_{\min} \le x \le x_{\max}. \tag{3.1}$$

При $\Delta x \to 0$ вероятность $P(x < X \le x + \Delta x)$ также будет стремиться к нулю, но предел неопределенности вида $0/0$ имеет конечное значение $W(x)$, зависящее от значения x и называемое ***плотностью вероятности непрерывной случайной величины***.

Для непрерывной случайной величины X закон распределения вероятностей считается заданным, если известна зависимость плотности вероятности от текущих значений x.

Интегральная функция распределения $F(x)$ непрерывных случайных величин определяется так же, как и для дискретных случайных величин. Значение функции распределения $F(x)$ в точке x равно вероятности того, что случайная величина X примет значение меньше, чем значение x:

$$F(x) = P(X < x).$$

В соответствии со свойством функции распределения (2.7) для вероятности попадания случайной величины X на интервал Δx можно записать:

$$P(x \le X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Подставляя данное равенство в формулу (3.1), получим:

$$\lim_{\Delta x \to \infty} \frac{F(x+\Delta x) - F(x)}{\Delta x} = W(x).$$

В левой части данного равенства записан предел приращения функции $F(x)$ к приращению аргумента Δx при $\Delta x \to 0$. По определению этот предел равен производной функции распределения $dF(x)/dx$. Значит, плотность вероятности непрерывной случайной величины равна производной от интегральной функции распределения:

$$W(x) = \frac{dF(x)}{dx}. \tag{3.2}$$

В силу данного определения (3.2) плотность вероятности иногда называют дифференциальным законом распределения непрерывной случайной величины. Формула (3.2) позволяет получить аналитическое выражение для плотности вероятности при известной функции распределения.

В соответствии с определением неопределенного интеграла можно получить обратную зависимость, позволяющую определить функцию распределения при заданной плотности вероятности:

$$F(x) + C = \int W(x)dx .$$

Постоянная интегрирования C должна определяться из условий соблюдения свойств функции распределения (см. подраздел 2.3). С учетом области возможных значений непрерывной случайной величины и свойств функции распределения от постоянной интегрирования C можно избавиться, если в данной формуле от неопределенного интеграла перейти к интегралу с переменным верхним пределом интегрирования:

$$F(x) = \int_{x_{\min}}^{x} W(x)dx. \tag{3.3}$$

Из формулы (3.3) следует, что функция распределения определяется как интеграл от плотности вероятности непрерывной случайной величины, поэтому функцию распределения называют интегральной функцией распределения.

Поставим задачу определения вероятности попадания непрерывной случайной величины X в интервал $a \le X < b$. В соответствии со свойством (2.7) функции распределения можно записать:

$$P(a \le x < b) = F(b) - F(a).$$

С другой стороны, основываясь на определении определенного интеграла (формула Ньютона — Лейбница), можно записать:

$$P(a \le X < b) = F(b) - F(a) = \int_a^b W(x)\,dx. \tag{3.4}$$

Из формулы (3.4) следует, что при заданной плотности вероятности $W(x)$ вероятность попадания случайной величины в интервал $a \le X < b$ определяется определенным интегралом от плотности вероятности в пределах интегрирования от «a» до «b».

Значит, определенный интеграл с переменным верхним пределом интегрирования в формуле (3.3) будет равен вероятности того, что случайная величина X будет меньше некоторого текущего значения x ($P(x_{\min} \le X < x)$), что совпадает с определением интегральной функции распределения.

Как и функция распределения $F(x)$, плотность вероятности $W(x)$ непрерывной случайной величины должна удовлетворять определенным условиям, которые можно сформулировать следующим образом.

1. Плотность вероятности может принимать только положительные значения:

$$W(x) \ge 0. \tag{3.5}$$

Для доказательства данного свойства сделаем обратное предположение. Предположим, что на некотором интервале $c < X < d$ плотность вероятности принимает отрицательные значения. Тогда определенный интеграл от плотности вероятности в этих пределах в соответствии с геометрическим смыслом определенного интеграла будет отрицательным числом:

$$\int_c^d W(x)\,dx < 0.$$

В соответствии с формулой (3.4) данный определенный интеграл равен вероятности попадания случайной величины X в интервал $c \div d$.

Но вероятность по определению является сугубо положительной величиной. Значит, сделанное нами предположение неверно и справедливо условие (3.5).

2. Второе условие можно записать следующим образом:

$$W(x)=0 \text{ при } x<x_{\min}; \\ W(x)=0 \text{ при } x>x_{\max}, \tag{3.6}$$

т.е. при значениях текущей переменной x, находящихся вне области возможных значений непрерывной случайной величины, плотность вероятности равна нулю.

Для доказательства этого свойства определим вероятность попадания случайной величины X в интервал $c \div d$, который лежит выше максимально возможного значения $x_{\max} < c < d$. Так как интервал $c \div d$ находится вне области возможных значений, то событие, состоящее в том, что $c < X \le d$, является невозможным событием, и вероятность этого события равна нулю

$$P(c<x\le d)=\int_c^d W(x)dx=0, \text{ при } x_{\max}<c<d.$$

Для того чтобы определенный интеграл был равен нулю и выполнялось условие (3.5), плотность вероятности должна быть равна нулю. Для непрерывных случайных величин с неограниченной областью возможных значений $x_{\min}\to-\infty$; $x_{\max}\to\infty$. Данное условие записывается в следующем виде:

$$\lim_{x\to-\infty} W(x)=0; \ \lim_{x\to\infty} W(x)=0. \tag{3.6а}$$

3. Третье условие называют условием нормировки плотности вероятности. Записывается оно в следующем виде:

$$\int_{-\infty}^{\infty} W(x)dx=1. \tag{3.7}$$

В геометрическом смысле это условие означает, что площадь под кривой плотности вероятности всегда одинакова и равна единице.

В вероятностном смысле определенный интеграл (3.7) характеризует вероятность того, что случайная величина примет значение, принадлежащее области возможных значений. Данное событие является достоверным, а вероятность достоверного события равна единице.

Пример 3.1. Плотность вероятности непрерывной случайной величины X, область возможных значений которой ограничена $0 \le x \le \pi$, задана аналитическим выражением $W(x) = C \sin x$.

Необходимо определить значение постоянного множителя C, аналитическое выражение для функции распределения $F(x)$, построить графики плотности вероятности и функции распределения, определить вероятность попадания случайной величины в интервал $\pi / 6 \div \pi / 2$.

Решение. В соответствии с условием 2 можно записать $W(x) = 0$ при $x < 0$ и $W(x) = 0$ при $x > \pi$. Тогда условие нормировки для случайной величины с ограниченной областью возможных значений можно записать в виде:

$$\int_0^{\pi} W(x)\,dx = 1.$$

Для рассматриваемого примера это условие запишется:

$$\int_0^{\pi} C \sin x\,dx = 1.$$

В результате интегрирования получим:

$$-C\cos x\Big|_0^{\pi} = -c[-1-1] = 2C = 1.$$

Таким образом, $C = \dfrac{1}{2}$, и для плотности вероятности можно записать:

$$W(x) = \begin{cases} 0 & \text{при } x < 0; \\ \dfrac{1}{2}\sin x & \text{при } 0 \le x \le \pi; \\ 0 & \text{при } x > \pi. \end{cases}$$

Функцию распределения определяем по формуле (3.3). При $x < 0$ $W(x) = 0$ и $F(x) = 0$ при $x < 0$. При $0 \le x \le \pi$ получим:

$$F(x) = \int_0^{x} \frac{1}{2}\sin x\,dx = -\frac{1}{2}\cos x\Big|_0^{x} = \frac{1}{2}(1 - \cos x).$$

При $x > \pi$ можно записать:

$$F(x) = \int_0^{\pi} w(x)\,dx + \int_{\pi}^{x} w(x)\,dx.$$

Первый интеграл равен единице из условия нормировки, второй интеграл равен нулю в соответствии с условием 2 для плотности вероятности. Таким образом, $F(x) = 1$ при $x > \pi$.

Окончательно для функции распределения можно записать:

$$F(x) = \begin{cases} 0 & \text{при} \quad x < 0; \\ \frac{1}{2}(1 - \cos x) & \text{при} \quad 0 \le x < \pi; \\ 1 & \text{при} \quad x \ge \pi. \end{cases}$$

Графики плотности вероятности и функции распределения для рассматриваемого примера приведены на рис. 3.2.

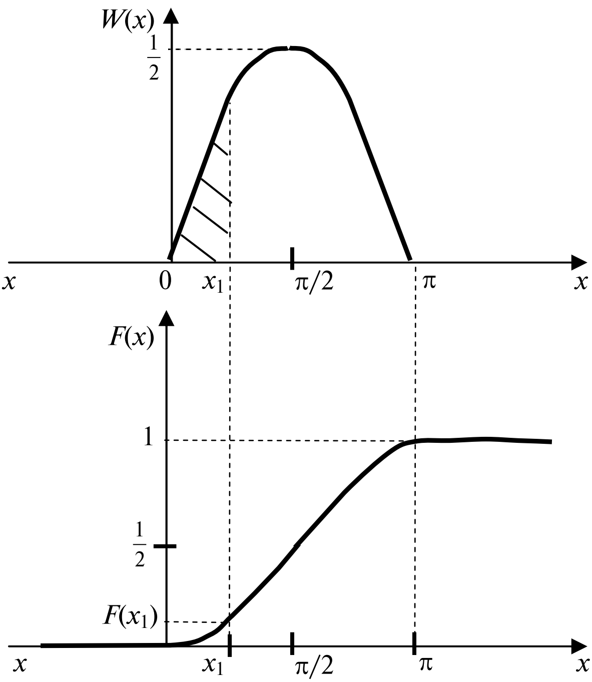


Рис. 3.2. Графики плотности вероятности и функции распределения непрерывной случайной величины

В соответствии с определением функции распределения значение ее $F(x_1)$ в произвольной точке x_1 равно площади под кривой плотности вероятности, находящейся левее данной точки x_1.

Вероятность попадания случайной величины в интервал $\frac{\pi}{6} \le x < \frac{\pi}{2}$ определим по плотности вероятности и по функции распределения, (формула (3.4)):

$$P\left(\frac{\pi}{6} \le X < \frac{\pi}{2}\right) = \int_{\pi/6}^{\pi/2} \frac{1}{2} \sin x \, dx = -\frac{1}{2} \cos x \Big|_{\pi/6}^{\pi/2} = -\frac{1}{2}\left[0 - \frac{\sqrt{3}}{2}\right] = \frac{\sqrt{3}}{4}.$$

$$P\left(\frac{\pi}{6} \le x < \frac{\pi}{2}\right) = F\left(\frac{\pi}{2}\right) - F\left(\frac{\pi}{6}\right) = \frac{1}{2}\left[1 - \cos\frac{\pi}{2}\right] - \frac{1}{2}\left[1 - \cos\frac{\pi}{6}\right] =$$

$$= \frac{1}{2} - \frac{1}{2} + \frac{\sqrt{3}}{4} = \frac{\sqrt{3}}{4}.$$

3.3. Числовые характеристики непрерывных случайных величин

Как и дискретные случайные величины, непрерывные случайные величины можно характеризовать их числовыми характеристиками, в общем называемыми моментами распределения непрерывных случайных величин. Из наиболее употребляемых числовых характеристик следует выделить начальный момент первого порядка, или математическое ожидание непрерывной случайной величины. Математическое ожидание непрерывной случайной величины есть среднее значение, вокруг которого группируются возможные значения непрерывной случайной величины. Для непрерывных случайных величин операция нахождения среднего значения (математического ожидания), или операция усреднения, записывается следующим образом:

$$m_x = M(X) = \langle X \rangle = \int_{-\infty}^{\infty} x W(x) dx. \tag{3.8}$$

Операция усреднения, определяемая формулой (3.8), полностью аналогична операции усреднения дискретных случайных величин (2.9). Покажем эту аналогию на примере определения математического ожидания непрерывной случайной величины, область возможных значений которой ограничена интервалом $x_0 \le X < x_n$, а плотность вероятности $W(x)$ известна (приведена на рис. 3.3).

С учетом условия (3.6) формулу (3.8) для данной случайной величины можно записать в виде:

$$m_x = \int_{x_0}^{x_n} x W(x) dx. \tag{3.9}$$

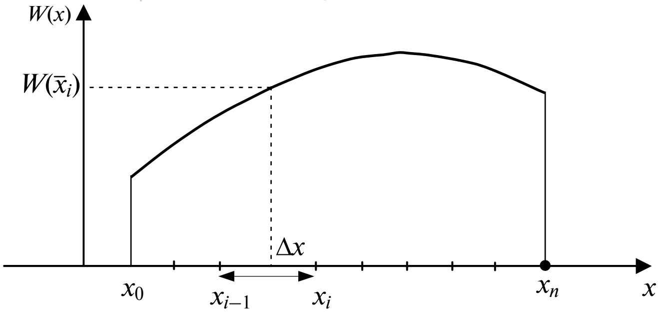


Рис. 3.3. Плотность вероятности случайной величины

Разобъем область возможных значений непрерывной случайной величины на n интервалов Δx:

$$\Delta x = \frac{x_0 - x_n}{n}; \quad \Delta x = x_i - x_{i-1}$$

и представим формулу (3.9) в виде суммы интегралов:

$$m_x = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} xW(x)dx. \tag{3.10}$$

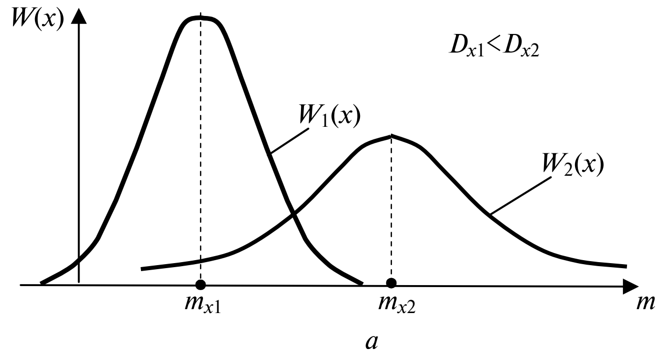
При большом числе n можно приближенно считать, что в пределах каждого интервала плотность вероятности остается постоянной и равной $W(\bar{x}_i)$, где $\bar{x}_i = \frac{1}{2}(x_i + x_{i-1})$ — середина i-го интервала. С учетом данного приближения после интегрирования в формуле (3.10) получим:

$$m_x \approx \sum_{i=1}^{n} \bar{x}_i W(\bar{x}_i)\Delta x = \sum_{i=1}^{n} \bar{x}_i P(\bar{x}_i).$$

Полученный результат аналогичен формуле (2.9). Таким образом, для непрерывных случайных величин операция усреднения означает умножение усредняемой величины на плотность вероятности $W(x)$ и интегрирование произведения по области всех возможных значений случайной величины. На рисунке 3.4 приведены графики плотностей вероятности $W_1(x)$ и $W_2(x)$ двух случайных величин, имеющих разные математические ожидания $m_{x1} < m_{x2}$.

Дисперсия определяется как среднее значение квадрата отклонения случайной величины от математического ожидания:

$$D_x = M[(x - m_x)^2] = \left\langle (x - m_x)^2 \right\rangle .$$



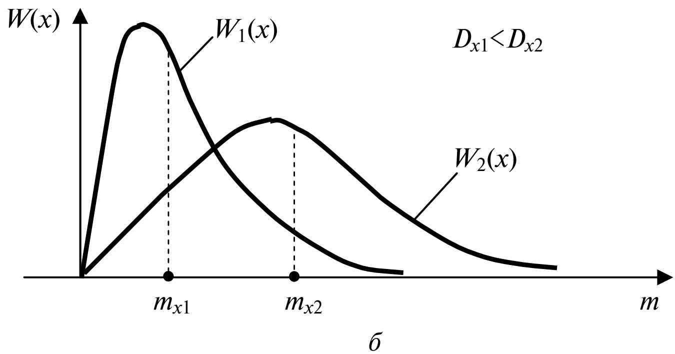


Рис. 3.4. Плотности вероятности случайных величин, имеющих разные значения математических ожиданий и дисперсий

В результате усреднения (умножения $(x - mx)^2$ на плотность вероятности и интегрирования) для дисперсии непрерывной случайной величины получим формулу

$$D_x = \int_{-\infty}^{\infty} (x - m_x)^2 W(x) dx. \tag{3.11}$$

Если в формуле (3.11) возвести выражение в скобках в квадрат и выполнить интегрирование трех слагаемых, получим более простую формулу для вычисления дисперсии непрерывной случайной величины:

$$D_x = \int_{-\infty}^{\infty} x^2 W(x) dx - m_x^2. \tag{3.12}$$

Интеграл в формуле (3.12) является начальным моментом второго порядка распределения непрерывной случайной величины:

$$\nu_2 = \int_{-\infty}^{\infty} x^2 W(x)\,dx;$$
$$\nu_2 = D_x + m_x^2. \tag{3.13}$$

На рисунке 3.4 приведены графики плотности вероятностей $W_1(x)$ и $W_2(x)$ двух случайных величин, имеющих разные дисперсии $D_{x1} < D_{x2}$.

Математическое ожидание имеет те же единицы измерения, что и случайная величина. Из формулы (3.12) видно, что единицы измерения дисперсии равны квадрату единиц измерения случайной величины. Поэтому для сравнения среднего значения отклонения случайной величины от математического ожидания со значением математического ожидания используют среднеквадратичное отклонение σ_x:

$$\sigma_x = \sqrt{D_x}\ .$$

Из приведенных на рис. 3.4 графиков плотностей вероятностей видно, что для второй случайной величины (график $W_2(x)$) область наиболее вероятных значений X_2 (область x, где $W_2(x) > 0$) шире, чем для первой случайной величины. Значит, для второй случайной величины возможны бóльшие отклонения ее значений от математического ожидания. Это говорит о том, что дисперсия и среднеквадратичное отклонение для второй случайной величины будут больше, чем для первой $D_{x_1} < D_{x_2}$; $\sigma_{x_1} < \sigma_{x_2}$.

Для всех графиков плотностей вероятностей, приведенных на рис. 3.4, должно выполняться условие нормировки (3.7), т.е. площадь под кривыми $W_1(x)$ и $W_2(x)$ одинакова и равна единице.

В заключение отметим, что математическое ожидание и дисперсия непрерывных случайных величин обладают теми же свойствами, что математическое ожидание и дисперсия дискретных случайных величин (см. подразделы 2.4.1 и 2.4.2).

3.4. Законы распределения непрерывных случайных величин

В данном подразделе мы рассмотрим наиболее распространенные законы распределения непрерывных случайных величин. Знание законов распределения непрерывных случайных величин позволит использовать их в качестве математических моделей для аналитического задания законов распределения различных экономических показателей.

Особое место среди всех законов распределения занимает нормальный закон распределения, играющий исключительно важную роль в теории вероятностей. Особое положение нормального закона распределения объясняется тем, что, во-первых, он является предельным законом, к которому асимптотически приближаются другие законы распределения, во-вторых, его основу составляет так называемая центральная предельная теорема теории вероятностей. Эта теорема утверждает, что плотность вероятности суммы большого числа случайных величин, вклад каждой из которых в сумму очень мал, независимо от законов распределения каждого из слагаемых приближается к нормальному закону распределения. В экономике значения тех или иных показателей формируются как результат взаимодействия большого числа независимых случайных факторов, каждый из которых не оказывает превалирующего влияния на интересующий нас показатель, т.е. описанная ситуация соответствует условию применимости центральной предельной теоремы теории вероятностей.

3.4.1. Нормальный закон распределения

Нормальным называется распределение непрерывной случайной величины X, плотность вероятности которой определяется формулой

$$W(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left[-\frac{(x - m_x)^2}{2\sigma_x^2} \right], \tag{3.14}$$

где $-\infty < x < \infty$; $-\infty < m_x < \infty$; $\sigma_x > 0$.

Из формулы (3.14) следует, что для построения графика плотности вероятности должны быть численно заданы два параметра: m_x и σ_x. Эти параметры имеют уже известный нам вероятностный смысл: m_x — это математическое ожидание случайной величины, а σ_x — ее среднеквадратичное отклонение.

В соответствии с формулами (3.8) и (3.14) для математического ожидания получим:

$$M(X) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left[\frac{(x - m_x)^2}{2\sigma_x^2} \right] dx.$$

Переходя к новой переменной интегрирования $y = \frac{x - m_x}{\sigma_x}$, данный интеграл преобразуется к виду:

$$M(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma_x y+m_x)e^{-\frac{y^2}{2}}dy=\frac{\sigma_x}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ye^{-\frac{y^2}{2}}dy+\frac{m_x}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy.$$

В данном выражении первый интеграл равен нулю как интеграл от нечетной функции, вычисляемый в симметричных пределах интегрирования. Интеграл во втором слагаемом является интегралом Пуассона и равен $\sqrt{2\pi}$. Таким образом, получим:

$$M(X)=\frac{m_x}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=m_x .$$

Дисперсию нормально распределенной случайной величины (3.14) определим по формуле (3.11):

$$D_x=\frac{1}{\sigma_x\sqrt{2\pi}}\int(x-m_x)^2\exp\left[-\frac{(x-m_x)^2}{2\sigma_x^2}\right]dx.$$

Сделаем ту же самую замену переменной $y=\dfrac{x-m_x}{\sigma_x}$ и получим

$$D_x=\frac{\sigma_x^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}y^2e^{-\frac{y^2}{2}}dy.$$

Далее воспользуемся методом интегрирования по частям, для этого введем обозначения:

$$U=y,\ d\nu=ye^{-\frac{y^2}{2}}dy,\ \text{отсюда}\ \nu=\int ye^{-\frac{y^2}{2}}dy=-e^{-\frac{y^2}{2}},\ dU=dy.$$

С учетом введенных обозначений получим

$$D_x=\frac{\sigma_x^2}{\sqrt{2\pi}}\left[-\lim_{c\to\infty}ye^{-\frac{y^2}{2}}\Bigg|_{-c}^{c}+\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\right].$$

Первое слагаемое в квадратных скобках равно нулю, а второе является интегралом Пуассона, равным $\sqrt{2\pi}$. С учетом этого получим

$$D_x=\frac{\sigma_x^2}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=\sigma_x^2.$$

Для нормального закона распределения удается получить в общем виде все центральные моменты произвольного порядка. Все центральные моменты нечетного порядка будут равны нулю, а для центральных моментов четного порядка справедлива формула

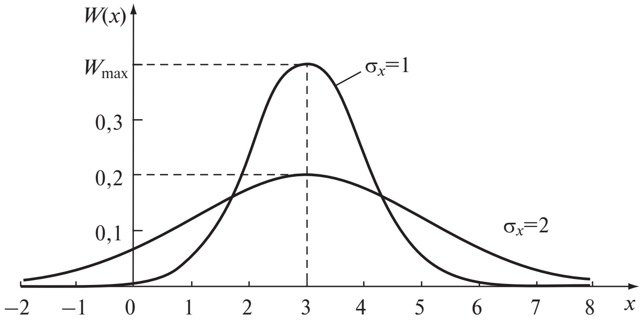
$$\mu_{2k} = \frac{(2k)!}{k!}\left(\frac{\sigma^2}{2}\right)^k.$$

При $k = 1$ получаем центральный момент второго порядка, равный дисперсии

$$\mu_2 = D_x = \sigma_x^2.$$

График плотности вероятности нормально распределенной случайной величины при $m_x = 3$, $\sigma_x = 1$ и $\sigma_x = 2$ приведен на рис. 3.5. Из формулы (3.14) и приведенных графиков видно, что плотность вероятности является симметричной относительно точки $x = m_x$. В данной точке плотность вероятности имеет максимальное значение, равное

$$W(m_x) = \frac{1}{\sigma_x\sqrt{2\pi}}.$$



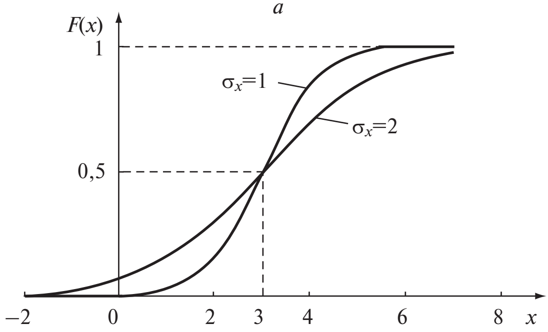


Рис. 3.5. Плотность вероятности и функция распределения при нормальном законе распределения: *а* — $\sigma_x = 1$; *б* — $\sigma_x = 2$; $W_{\max} = \dfrac{1}{\sqrt{2\pi}\,\sigma_x}$

На приведенных графиках показано, что уменьшение дисперсии случайной величины $D_x = \sigma_x^2$ приводит к увеличению максимального значения плотности вероятности и более резкому уменьшению значений $W(x)$ при отклонении значений x от m_x. Изменение математического ожидания m_x не влияет на форму кривой $W(x)$, а лишь смещает ее вправо при увеличении m_x и влево при уменьшении m_x. Площадь под кривой плотности вероятности не зависит от значений параметров m_x и σ_x и всегда равна единице (условие нормировки).

Для вычисления вероятности попадания нормально распределенной случайной величины в заданный интервал проще использовать функцию распределения $F(x)$. В соответствии с формулами (3.3) и (3.4) для функции распределения можно записать:

$$F(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(x-m_x)^2}{2\sigma_x^2}\right] dx. \qquad (3.15)$$

Интеграл (3.15) через элементарные функции не выражается, а результаты численного интегрирования зависят от значений параметров m_x и σ_x. Для исключения этой зависимости сделаем в выражении (3.15) замену переменной интегрирования $z = \dfrac{x-m_x}{\sigma_x}$:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m_x}{\sigma_x}} \exp\left(-\frac{z^2}{2}\right) dx = \Phi\left(\frac{x-m_x}{\sigma_x}\right). \qquad (3.16)$$

Интеграл в формуле (3.16) зависит только от верхнего предела интегрирования $z = \dfrac{x-m_x}{\sigma_x}$ и называется интегралом вероятности $\Phi(z)$:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$$

Интеграл вероятности является специальной табулированной функцией, значения которой приведены в таблице приложения 1. Графики интегральной функции распределения $F(x)$ приведены на рис. 3.5, *б*.

Для вычисления вероятности попадания случайной величины в интервал $x_1 \le X < x_2$ необходимо вычислить значения переменных z_1 и z_2, соответствующие границам заданного интервала:

$$z_1 = \frac{x_1 - m_x}{\sigma_x};\; z_2 = \frac{x_2 - m_x}{\sigma_x}.$$

Далее, по таблице интеграла вероятности (см. приложение 1) находим значение интегральной функции распределения на границах интервала:

$$F(x_1) = \Phi(z_1);\ F(x_2) = \Phi(z_2).$$

Если при вычислении значений z_1 и z_2 одно или оба значения оказались отрицательными, то следует учитывать следующее свойство интеграла вероятности:

$$\Phi(-z) = 1 - \Phi(z). \qquad (3.17)$$

По найденным значениям функции распределения легко вычисляется вероятность попадания нормальной случайной величины в заданный интервал:

$$P(x_1 \le X < x_2) = F(x_2) - F(x_1) = \Phi(z_2) - \Phi(z_1). \qquad (3.18)$$

Определим вероятность того, что отклонение нормальной случайной величины от математического ожидания не превысит заданной величины ε:

$$P(|X - m_x| < \varepsilon) \text{ или } P(m_x - \varepsilon \le X < m_x + \varepsilon).$$

Для этого определим значения переменной z на границах интервала:

$$z_1 = \frac{m_x - \varepsilon - m_x}{\sigma_x} = -\frac{\varepsilon}{\sigma_x};\ z_2 = \frac{m_x + \varepsilon - m_x}{\sigma_x} = \frac{\varepsilon}{\sigma_x}.$$

Отношение $\frac{\varepsilon}{\sigma_x}$ называют ***кратностью отклонения***. Находим значения функции распределения на границах интервала:

$$F(m_x - \varepsilon) = \Phi\left(-\frac{\varepsilon}{\sigma_x}\right) = 1 - \Phi\left(\frac{\varepsilon}{\sigma_x}\right);$$

$$F(m_x + \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma_x}\right).$$

Для искомой вероятности получим:

$$P(|X - m_x| < \varepsilon) = P(m_x - \varepsilon < X < m_x + \varepsilon) =$$

$$= F(m_x + \varepsilon) - F(m_x - \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 + \Phi\left(\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1. \qquad (3.19)$$

Пример 3.2. Найти вероятность того, что отклонение случайной величины от математического ожидания не превысит кратность отклонения $\dfrac{\varepsilon}{\sigma_x} = 1$, 2 и 3.

Решение. В соответствии с формулой (3.19) находим табличные значения $\Phi\left(\dfrac{\varepsilon}{\sigma_x}\right)$:

$$\Phi\left(\frac{\varepsilon}{\sigma_x}\right) = \Phi(1) = 0{,}8413;$$

$$\Phi(2) = 0{,}97725;$$
$$\Phi(3) = 0{,}99865.$$

Для искомых вероятностей получим:

$$P(|X - m_x| < \sigma_x) = 2\Phi(1) - 1 = 2 \cdot 0{,}8413 - 1 = 0{,}6826\ ;$$

$$P(|X - m_x| < 2\sigma_x) = 2\Phi(2) - 1 = 2 \cdot 0{,}97725 - 1 = 0{,}9545$$

$$P(|X - m_x| < 3\sigma_x) = 2\Phi(3) - 1 = 2 \cdot 0{,}99865 - 1 = 0{,}9973.$$

По вычисленным вероятностям видно, что все значения нормальной случайной величины группируются около математического ожидания, практически не удаляясь от него на величину больше чем $\pm 2\sigma_x$.

Пример 3.3. Непрерывная случайная величина распределена по нормальному закону с математическим ожиданием $m_x = 8{,}5$ и среднеквадратическим отклонением $\sigma_x = 1{,}6$. Найти вероятность того, что случайная величина примет значения, не принадлежащие интервалу $\Delta x = (7{,}3 \div 10{,}9)$, $P(X \not\subset \Delta x)$.

Решение. Вычисляем значение переменной z на границах интервала Δx:

$$z_1 = \frac{x_1 - m_x}{\sigma_x} = \frac{7{,}3 - 8{,}5}{1{,}6} = -0{,}75;$$

$$z_2 = \frac{x_2 - m_x}{\sigma_x} = \frac{10{,}9 - 8{,}5}{1{,}6} = 1{,}5.$$

По таблице значений интеграла вероятности находим:

$$\Phi(z_2) = \Phi(1{,}5) = 0{,}9332;$$
$$\Phi(z_1) = \Phi(-0{,}75) = 1 - \Phi(0{,}75) = 1 - 0{,}7734 = 0{,}2266.$$

Определяем вероятность попадания случайной величины X в интервал Δx:

$$P(X \subset \Delta x) = \Phi(z_2) - \Phi(z_1) = 0{,}9332 - 0{,}2266 = 0{,}7066.$$

Вероятность противоположного события $P(X \not\subset \Delta x)$, являющаяся искомой вероятностью, определяется по формуле

$$P(X \not\subset \Delta x) = 1 - P(X \subset \Delta x) = 1 - 0{,}7066 = 0{,}2934 .$$

3.4.2. Распределение Эрланга

Распределение Эрланга является двухпараметрическим законом распределения, используемым для вероятностного задания положительных непрерывных случайных величин. Плотность вероятности случайной величины, имеющей распределение Эрланга, определяется формулой

$$W(x) = \frac{\lambda^{k+1}}{k!} x^k e^{-\lambda x}, \; x \ge 0, \lambda > 0. \qquad (3.20)$$

Как следует из формулы (3.20), плотность вероятности зависит от значения двух параметров k и λ. Параметр k называют порядком распределения Эрланга, и он может иметь целочисленные значения $k = 0, 1, 2, \ldots$ На рисунке 3.6, *а* приведены графики плотности вероятности распределения Эрланга нулевого $k = 0$, первого $k = 1$ и второго $k = 2$ порядка при $\lambda = 2$.

Математическое ожидание и дисперсия случайной величины, имеющей распределение Эрланга, определяется формулами:

$$m_x = \frac{k+1}{\lambda}; \; D_x = \frac{k+1}{\lambda^2} . \qquad (3.21)$$

Получим функцию распределения $F(x)$ для случайной величины, имеющей распределение Эрланга первого порядка $k = 1$. Подставляя формулу (3.20) в (3.3), для $F(x)$ получим:

$$F(x) = \int_0^x \lambda^2 x e^{-\lambda x} dx.$$

Используем метод интегрирования по частям. Для этого введем обозначения:

$$u = x, \; dv = \lambda e^{-\lambda x} dx,$$

отсюда

$$v = \int \lambda e^{-\lambda x} dx = -e^{-\lambda x}.$$

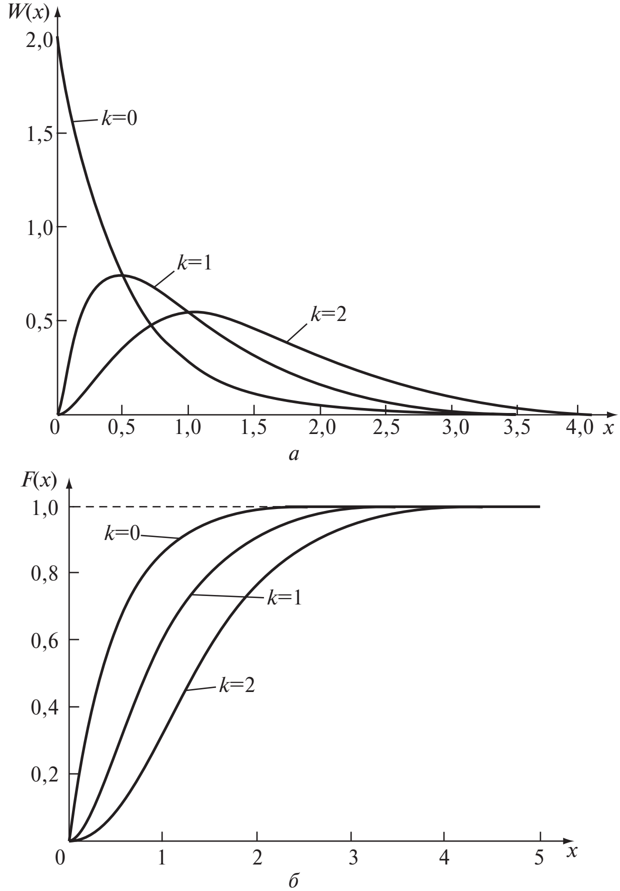


Рис. 3.6. Распределение Эрланга

В соответствии с принятыми обозначениями получим:

$$F(x) = \lambda\left[-xe^{-\lambda x}\Big|_0^x + \int_0^x e^{-\lambda x}\,dx\right] = \lambda\left[-xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x} + \frac{1}{\lambda}\right] \Rightarrow$$

$$\Rightarrow F(x) = 1 - (1 + \lambda x)e^{-\lambda x}. \qquad (3.22)$$

Формула (3.22) позволяет легко определить вероятность попадания в заданный интервал непрерывной случайной величины, имеющей распределение Эрланга первого порядка. Графики функции распределения случайной величины, имеющей распределение Эрланга, приведены на рис. 3.6, *б*. При более высоком порядке распределения Эрланга $k \geq 2$ формула для функции распределения получается более сложной и здесь не приводится.

Экспоненциальное (показательное) распределение является частным случаем распределения Эрланга при $k = 0$. Плотность вероятности случайной величины, имеющей экспоненциальное распределение, определяется формулой

$$W(x) = \lambda e^{-\lambda x}, \ x \geq 0, \ \lambda > 0. \tag{3.23}$$

Экспоненциальное распределение является однопараметрическим законом распределения, т.е. определяется только одним параметром λ.

Математическое ожидание и дисперсия экспоненциально распределенной случайной величины определяется формулами:

$$m_x = \frac{1}{\lambda}; \quad D_x = \frac{1}{\lambda^2}. \tag{3.24}$$

Характерным признаком экспоненциального распределения является то, что при любых значениях параметра распределения λ математическое ожидание равно среднеквадратичному отклонению случайной величины:

$$m_x = \sigma_x = \frac{1}{\lambda}.$$

Интегральная функция распределения при экспоненциальном распределении случайной величины определяется формулой

$$F(x) = 1 - e^{-\lambda x}; \quad x \geq 0. \tag{3.25}$$

На рисунке 3.7 приведены графики плотности вероятности и функции распределения случайной величины, имеющей экспоненциальный закон распределения при значениях параметра $\lambda = 0{,}5,\ 1$ и 2.

Еще одним частным случаем распределения Эрланга является так называемое ***хи-квадрат*** (χ^2-распределение). Это распределение нашло широкое применение в математической статистике. Случайная величина X имеет χ^2-распределение, если ее плотность вероятности определяется формулой

$$W(x) = \frac{1}{\sqrt{2^n}\,\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ x \geq 0, \ n = 1, 2, 3, \ldots, \tag{3.26}$$

где $\Gamma\left(\frac{n}{2}\right)$ — специальная функция, называемая гамма-функцией.

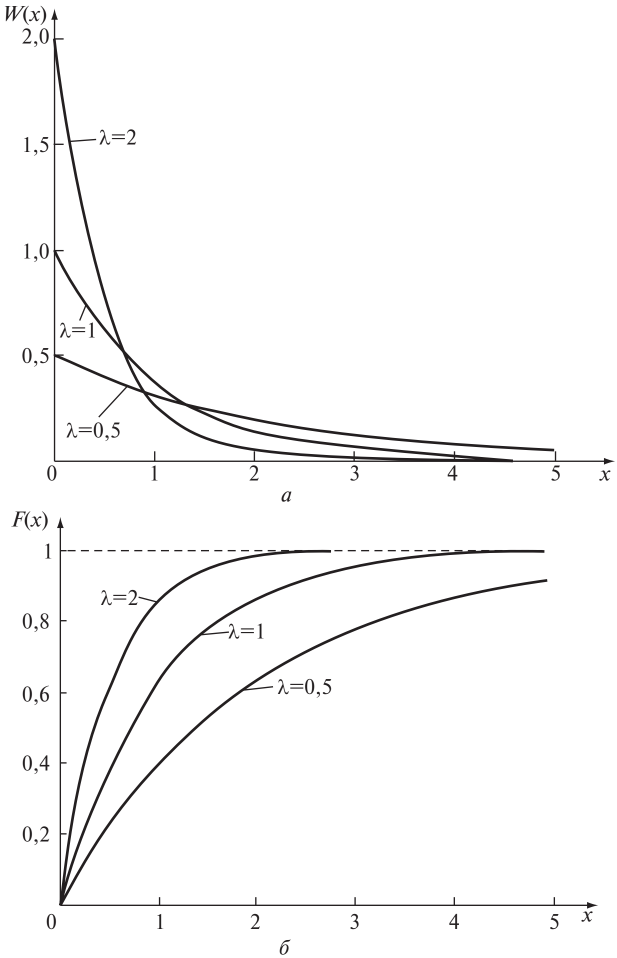


Рис. 3.7. Плотность вероятности и функция распределения экспоненциально распределенной случайной величины

При четных значениях $n \geq 2$ гамма-функция будет равна:

$$\Gamma\left(\frac{n}{2}\right)=\left(\frac{n}{2}-1\right)!=1\cdot 2\cdot 3\cdot 4\ldots\left(\frac{n}{2}-1\right).$$

При нечетных значениях n гамма-функция определяется формулой

$$\Gamma\left(\frac{n}{2}\right)=\frac{\sqrt{\pi}}{2^n}(2n-1)!!,$$

где $(2n-1)!!=1\cdot3\cdot5\cdot7\cdot\ldots\cdot(2n-1)$ — произведение нечетных чисел от 1 до $(2n-1)$.

Хи-квадрат распределение получается, если в законе распределения Эрланга принять $\lambda={}^1/_2$ и $k=\frac{n}{2}-1.$ Оно относится к классу однопараметрических законов распределения, так как плотность вероятности полностью определяется одним параметром n, который называют числом степеней свободы.

Математическое ожидание и дисперсия случайной величины, имеющей χ^2-распределение, определяются простыми формулами, которые приведем без вычислений:

$$m_x=n;\ D_x=\sigma_x^2=2n. \tag{3.27}$$

Из формулы (3.26) следует, что при $n=2$ χ^2-распределение совпадает с экспоненциальным распределением при $\lambda={}^1/_2$. На рисунке 3.8 приведены графики плотности вероятности и функции распределения случайной величины, имеющей χ^2 распределение при $n=4, 6, 10$. Интегральная функция распределения при χ^2-распределении определяется через специальные неполные гамма-функции $\Gamma\left(\frac{n}{2};\frac{x}{2}\right)$:

$$F(x)=\begin{cases}0 & \text{при} \quad x<0;\\ \Gamma\left(\frac{n}{2};\frac{x}{2}\right)/\Gamma\left(\frac{n}{2}\right) & \text{при} \quad x\ge 0.\end{cases} \tag{3.28}$$

Специальные функции $\Gamma\left(\frac{n}{2};\frac{x}{2}\right)$ и $\Gamma\left(\frac{n}{2}\right)$ являются табулированными функциями, что позволяет определять вероятность попадания случайной величины, имеющей χ^2-распределение, в заданный интервал.

3.4.3. Распределение Вейбула

В некоторых случаях распределение Эрланга не может быть применено для вероятностного описания случайных величин из-за слишком медленного убывания плотности вероятности при увеличении значений x. В этих случаях более адекватной моделью закона распределения случайной величины может служить распределение Вейбула, для которого плотность вероятности определяется формулой

$$W(x)=\begin{cases}\dfrac{\alpha}{\sigma}\left(\dfrac{x}{\sigma}\right)^{\alpha-1}\exp\left(-\dfrac{x^{\alpha}}{\sigma^{\alpha}}\right) \text{ при } x\geq 0\ . \\ 0 \quad \text{при } x<0. \end{cases} \tag{3.29}$$

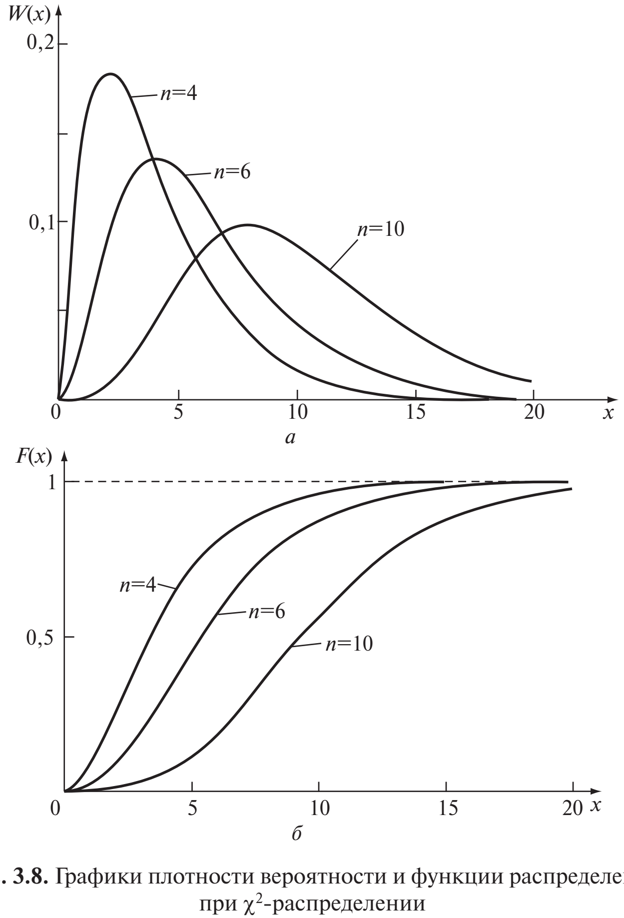


Рис. 3.8. Графики плотности вероятности и функции распределения при χ^2-распределении

Интегральная функция распределения Вейбула достаточно просто определяется по плотности вероятности:

$$F(x)=\int_0^x \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1}\exp\left(-\frac{x^{\alpha}}{\sigma^{\alpha}}\right)dx.$$

Если в данном интеграле сделать замену переменной интегрирования $y = \frac{x^\alpha}{\sigma^\alpha}$, то получим:

$$F(x) = \int_0^{\frac{x^\alpha}{\sigma^\alpha}} \exp(-y)\,dy = -e^{-y}\Big|_0^{\frac{x^\alpha}{\sigma^\alpha}};$$
$$F(x) = 1 - \exp\left(-\frac{x^\alpha}{\sigma^\alpha}\right). \qquad (3.30)$$

Из формул (3.29) и (3.30) следует, что распределение Вейбула является двухпараметрическим законом распределения, так как плотность вероятности и функция распределения полностью определяются двумя параметрами α и σ. При введенном обозначении параметр σ не следует путать со среднеквадратичным отклонением случайной величины σ_x. В зависимости от значений параметров α и σ формула (3.29) позволяет получить широкое разнообразние графиков плотности вероятности случайных величин. Графики плотности вероятности и функции распределения случайной величины, имеющей закон распределения Вейбула при $\sigma = 1$ и $\alpha = 1, 2, 4$, приведены на рис. 3.9.

Из формулы (3.29) следует, что при $\alpha = 1$ распределение Вейбула совпадает с экспоненциальным распределением при $\lambda = \frac{1}{\sigma}$. При $\alpha = 2$ получим другой частный случай распределения Вейбула, называемый ***распределением Рэлея.*** В своей стандартной записи плотность вероятности распределения Рэлея определяется формулой

$$W(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right). \qquad (3.31)$$

Интегральная функция распределения Рэлея определяется формулой:

$$F(x) = 1 - \exp\left(\frac{-x^2}{2\sigma^2}\right). \qquad (3.32)$$

Графики плотности вероятности и функции распределения случайной величины, имеющей закон распределения Рэлея при $\sigma = 1;\ \sqrt{2};\ 2$, приведены на рис. 3.10. Вычисления математического ожидания, начального момента второго порядка и дисперсии случайной величины, распределенной по закону Рэлея, дают следующие результаты:

$$m_x = \sqrt{\frac{\pi}{2}}\sigma;$$
$$\nu_2 = 2\sigma^2; \tag{3.33}$$

$$D_x = \nu_2 - m_x^2 = \left(2 - \frac{\pi}{2}\right)\sigma^2;$$
$$\sigma_x = \sqrt{2 - \frac{\pi}{2}}\,\sigma. \tag{3.34}$$

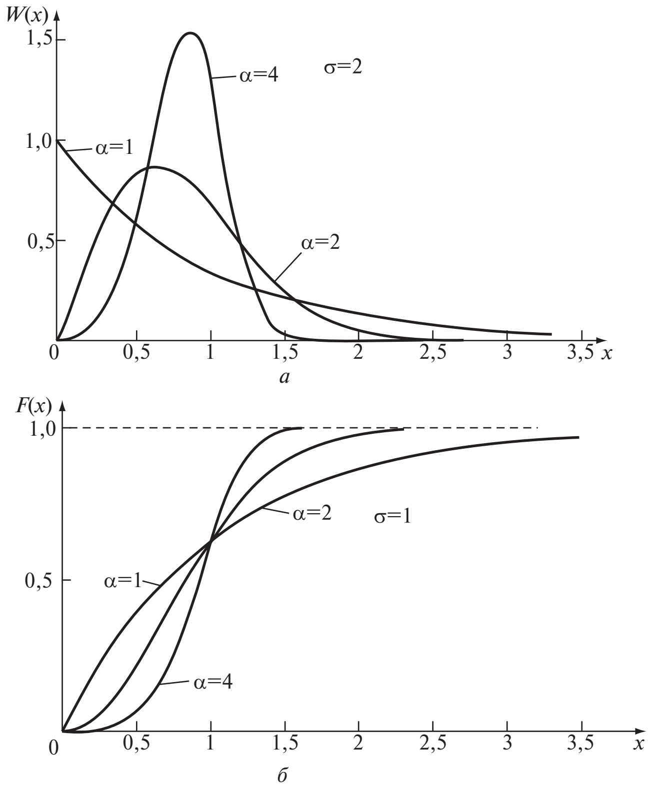


Рис. 3.9. Плотность вероятности и функция распределения Вейбула

Графики, приведенные на рис. 3.10 показывают, что с увеличением значения параметра распределения σ увеличивается математическое ожидание m_x и среднеквадратичное отклонение σ_x случайной величины.

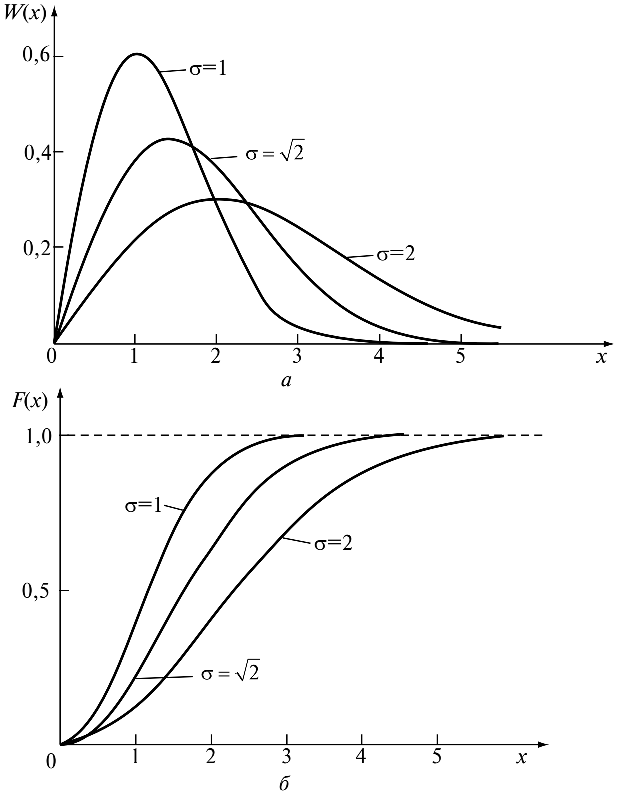


Рис. 3.10. Плотность вероятности и функция распределения Рэлея

3.4.4. Равномерное распределение

Равномерное распределение используется для вероятностного описания непрерывных случайных величин, область возможных значений которых ограничена некоторым интервалом $[a; b]$ и вероятность попадания случайной величины X в элементарный интервал Δx внутри области возможных значений распределена равномерно. То есть при равномерном распределении плотность вероятности случайной величины X внутри интервала $[a; b]$ имеет постоянное значение:

$$W(x) = \begin{cases} c & \text{при } a \le x < b \\ 0 & \text{при } x < a;\ x \ge b. \end{cases}$$

Значение плотности вероятности $W(x) = c$ внутри интервала $[a; b]$ можно найти из условия нормировки:

$$\int_{-\infty}^{\infty} W(x)\,dx = \int_{a}^{b} c\,dx = 1.$$

В результате интегрирования получим:

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

Таким образом, плотность вероятности при равномерном распределении случайной величины в интервале $[a; b]$ имеет вид:

$$W(x) = \begin{cases} \dfrac{1}{b-a} & \text{при} \quad a \le x < b; \\ 0 & \text{при} \quad x < a;\, x \ge b. \end{cases} \tag{3.35}$$

Для интегральной функции распределения в соответствии с формулой (3.3) получим:

$$F(x) = \int_{a}^{x} \frac{1}{b-a}\,dx = \frac{x-a}{b-x} \text{ при } a \le x < b.$$

Окончательно с учетом свойств интегральной функции распределения для $F(x)$ получим формулу

$$F(x) = \begin{cases} 0 & \text{при} \quad x < a; \\ \dfrac{x-a}{b-a} & \text{при} \quad a \le x < b; \\ 1 & \text{при} \quad x \ge b. \end{cases} \tag{3.36}$$

Из формул (3.35) и (3.36) следует, что равномерное распределение является двухпараметрическим законом распределения, так как $W(x)$ и $F(x)$ полностью определяются двумя параметрами a и b, ограничивающими нижнюю и верхнюю границу области возможных значений случайной величины.

Определим математическое ожидание и дисперсию случайной величины, имеющей равномерное распределение:

$$m_x = \int_{-\infty}^{\infty} xW(x)\,dx = \int_{a}^{b} \frac{x}{b-a}\,dx = \frac{b^2 - a^2}{2(b-a)} \Rightarrow m_x = \frac{b+a}{2}. \tag{3.37}$$

Из формулы (3.37) следует, что при равномерном распределении математическое ожидание случайной величины равно середине интервала, определяющего область ее возможных значений.

Найдем дисперсию по формуле

$$D_x = \int_{-\infty}^{\infty} x^2 W(x)dx - m_x^2 = \int_a^b \frac{x^2}{b-a}dx - m_x^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(b+a)^2}{4}.$$

После разложения разности кубов на сомножители $b^3 - a^3 = (b - a)(b^2 + ab + a^2)$ и вычитания дробей для дисперсии получим

$$D(x) = \frac{(b-a)^2}{12}. \qquad (3.38)$$

Среднеквадратичное отклонение равномерно распределенной случайной величины будет равно

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}. \qquad (3.39)$$

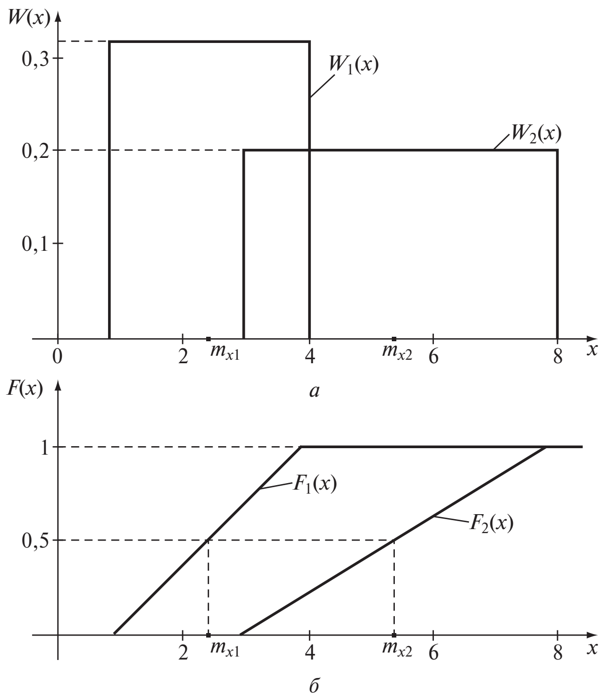


Рис. 3.11. Плотность вероятности и функция распределения равномерно распределенной случайной величины

На рисунке 3.11 приведены графики зависимости плотностей вероятности и функций распределения двух равномерно распределенных случайных величин, имеющих разные математические ожидания $m_{x1} < m_{x2}$ и дисперсии $D_{x1} < D_{x2}$. Из графиков видно, что область возможных значений первой случайной величины ограничена интервалом [1; 4], а второй случайной величины — интервалом [3; 8]. Соответственно внутри этих интервалов плотности вероятности первой и второй случайной величины будут равны: $W_1(x) = {}^1/_3$, $W_2(x) = {}^1/_5$. Математические ожидания и дисперсии этих случайных величин будут равны: $m_{x1} = 2{,}5$, $m_{x2} = 5{,}5$, $D_{x1} = {}^3/_4$, $D_{x2} = {}^{25}/_{12}$.

Пример 3.4. Непрерывная случайная величина имеет равномерный закон распределения. Известны значения математического ожидания m_x и дисперсии D_x данной случайной величины. Вычислить значения границ интервала $[a; b]$, определяющего область возможных значений данной случайной величины.

Решение. В соответствии с формулами (3.37) и (3.39) для границ интервала $[a; b]$ можно записать:

$$\begin{cases} a + b = 2m_x; \\ b - a = 2\sqrt{3}\sigma_x. \end{cases} \tag{3.40}$$

В данных уравнениях неизвестными являются границы интервала a и b, значения которых могут быть найдены из решения приведенной системы двух уравнений (3.40). При суммировании и вычитании друг из друга уравнений (3.40) мы получим решения:

$$\begin{aligned} b &= m_x + \sqrt{3}\sigma_x; \\ a &= m_x - \sqrt{3}\sigma_x. \end{aligned} \tag{3.41}$$

Подставляя в эти формулы значения математического ожидания m_x и значения среднеквадратичного отклонения $\sigma_x = \sqrt{D_x}$, можно вычислить значения границ области возможных значений непрерывной случайной величины при равномерном законе ее распределения.

3.5. Закон распределения системы двух непрерывных случайных величин

Рассмотрим способы задания законов распределения системы двух непрерывных случайных величин X и Y. Обозначим возможные значения этих случайных величин через x и y. Их можно характеризовать некоторой точкой M на плоскости (рис. 3.12) с координатами $M\ (x; y)$. Тогда вся область возможных значений Ω может быть отображена некоторой фигурой, границы которой ограничивают возможные значения системы случайных величин X и Y. Если область возможных значений случайных величин X, Y не ограничена, то Ω представляет собой всю координатную плоскость $-\infty < x < \infty,\ \ -\infty < y < \infty$.

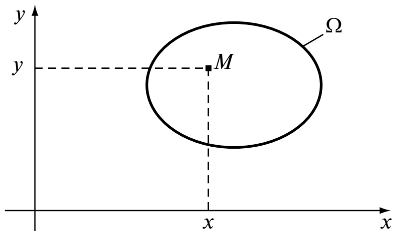


Рис. 3.12. Область возможных значений системы двух случайных величин

Одной из форм задания закона распределения системы двух случайных величин является задание двумерной функции распределения $F(x; y)$. Двумерная функция распределения является функцией двух независимых переменных x и y, и ее значение равно совместной вероятности выполнения двух условий: $X < x$ и $Y < y$:

$$F(x, y) = P(X < x,\ Y < y). \tag{3.42}$$

Для системы случайных величин с неограниченной областью возможных значений функция распределения $F(x; y)$ характеризует вероятность попадания точки в заштрихованную на рис. 3.13 область.

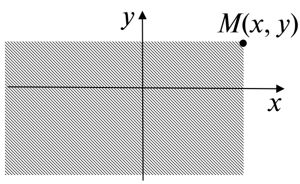


Рис. 3.13. К определению двумерной функции распределения

Рассмотрим основные свойства двухмерной функции распределения непрерывных случайных величин.

Свойство 1.

$$0 \le F(x, y) \le 1.$$

Это свойство очевидно из определения (3.42) функции распределения; так, вероятность может иметь значения $0 \le P \le 1$.

Свойство 2.

$$\lim_{\substack{x\to\infty\\ y\to\infty}} F(x;\, y) = F(\infty;\, \infty) = 1. \tag{3.43}$$

В соответствии с определением (3.42) функция распределения

$$F(\infty;\, \infty) = P(X < \infty;\, Y < \infty).$$

Событие, заключающееся в том, что в результате опыта случайные величины X и Y примут значения меньше, чем бесконечность, является достоверным событием, а вероятность достоверного события равна единице.

Условие (3.43) называют условием нормировки двумерной функции распределения.

Свойство 3.

$$\begin{aligned} &\lim_{y\to\infty} F(x;\, y) = F(x;\, \infty) = P(X < x;\, Y < \infty) = F(x);\\ &\lim_{x\to\infty} F(x;\, y) = F(\infty;\, y) = P(X < \infty;\, Y < y) = F(y). \end{aligned} \tag{3.44}$$

Формулы (3.44) позволяют легко получить одномерные функции распределения случайных величин X и Y путем предельного перехода от двумерной функции распределения $F(x, y)$.

Свойство 4.

$$\lim_{x\to-\infty} F(x;\, y) = \lim_{y\to-\infty} F(x;\, y) = \lim_{\substack{x\to-\infty\\ y\to-\infty}} F(x;\, y) = 0.$$

Данное свойство является очевидным, так как $P(X < -\infty) = 0$ $P(Y < -\infty) = 0$.

Свойство 5. Функция распределения $F(x;\, y)$ является неубывающей функцией по каждому аргументу x и y.

Для доказательства этого свойства рассмотрим на плоскости две точки — $A(x_1;\, y)$ и $B(x_2;\, y)$, где $x_1 < x_2$. Вычислим вероятность попадания случайной точки $(X;\, Y)$ в полосу на координатной плоскости $MABN$ (рис. 3.14).

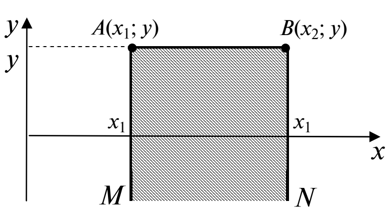


Рис. 3.14. Полоса координатной плоскости

$$P(x_1 \le X < x_2;\ Y < y) = P(X < x_2;\ Y < y) - P(X < x_1;\ Y < y) = \\ = F(x_2; y) - F(x_1; y). \qquad (3.45)$$

Так как вероятность любого события неотрицательна:

$$P(x_1 \le X < x_2;\ Y < y) \ge 0,$$

из этого следует, что

$$F(x_2; y) \ge F(x_1; y).$$

Аналогично доказывается, что при $y_1 < y_2$ и любом x выполняется условие $F(x; y_2) \ge F(x; y_1)$.

Свойство 6. Функция распределения $F(x; y)$ системы непрерывных случайных величин X и Y непрерывна в каждой точке области их возможных значений.

Данное свойство вытекает из того, что для непрерывных случайных величин вероятность событий $X = x$ и $Y = y$ равна нулю.

Если известно аналитическое выражение для двумерной функции распределения $F(x; y)$, то закон распределения системы случайных величин считается заданным. Закон распределения системы случайных величин X и Y может быть задан также их плотностью вероятности $W(x; y)$.

Плотностью вероятности $W(x; y)$ системы случайных величин $(X; Y)$ называется предел отношения вероятности попадания случайной точки $(X; Y)$ в некоторую область ΔS к площади этой области ΔS при условии, что эта область произвольно стягивается в точку.

Зададим область ΔS в виде прямоугольника со сторонами Δx и Δy (рис. 3.15). Тогда в соответствии с определением для двухмерной плотности вероятности $W(x; y)$ системы случайных величин X и Y можно записать:

$$W(x; y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x < X < x + \Delta x;\ y < Y < y + \Delta y)}{\Delta x \Delta y}. \qquad (3.46)$$

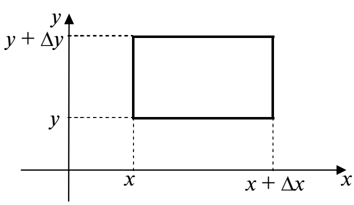


Рис. 3.15. Область ΔS

Определим вероятность попадания случайных величин X и Y в заданную прямоугольную область. В соответствии с формулой (3.45) получим:

$$P(x \le X < x+\Delta x;\ y \le Y < y+\Delta y) =$$
$$=\left[F(x+\Delta x; y+\Delta y)-F(x; y+\Delta y)\right]-\left[F(x+\Delta x; y)-F(x; y)\right]. \quad (3.47)$$

После подстановки формулы (3.47) в (3.46) получим:

$$W(x;y)=$$
$$=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\left\{\frac{\left[F(x+\Delta x; y+\Delta y)-F(x; y+\Delta y)\right]}{\Delta x\,\Delta y}-\frac{\left[F(x+\Delta x; y)-F(x; y)\right]}{\Delta x\,\Delta y}\right\} \quad (3.48)$$

По определению предел в правой части равенства (3.48) есть производная от двумерной функции распределения $F(x; y)$, взятая по каждой из переменных, т.е.

$$W(x;y)=\frac{\partial^2 F(x;y)}{\partial x\,\partial y}. \quad (3.49)$$

Таким образом, чтобы найти двумерную плотность вероятности, нужно дважды продифференцировать двумерную функцию распределения сначала по одной, а затем по другой переменной. Графически функция распределения $F(x; y)$ и плотность вероятности $W(x; y)$ представляют собой некоторые поверхности в трехмерном пространстве (рис. 3.16), уравнения которых определяются аналитическими выражениями для $F(x; y)$ и $W(x; y)$.

В соответствии с формулой (3.46) при достаточно малых значениях Δx и Δy вероятность попадания случайных величин X и Y в прямоугольную область, изображенную на рис. 3.15, приближенно можно определить произведением:

$$P(x < X < x+\Delta x;\ y < Y < y+\Delta y) \approx W(x;\ y)\Delta x\Delta y.$$

Из рисунка 3.16, *а* видно, что данная вероятность равна объему параллелепипеда с основанием площадью $\Delta S=\Delta x\Delta y$ и высотой $W(x; y)$. Таким образом, по аналогии с формулой (3.4) вероятность попадания случайных величин в прямоугольную область $a < X < b$ и $c < Y < d$ можно определить путем вычисления двойного интеграла:

$$P(a < X < b; c < Y < d)=\int_a^b\int_c^d W(x;\ y)\,dx\,dy. \quad (3.50)$$

Из приведенных рассуждений и определения плотности вероятности следует, что $W(x; y)$ должна удовлетворять следующим свойствам.

Свойство 1. Двумерная плотность вероятности случайных величин X; Y есть функция неотрицательная $W(x; y) \ge 0$.

Свойство 2. Двумерная функция распределения системы случайных величин (X; Y) связана с двумерной плотностью вероятности соотношением

$$F(x;y)=\int_{-\infty}^{x}\int_{-\infty}^{y}W(x;y)\,dx\,dy. \tag{3.51}$$

Свойство 3. Данное свойство называют условием нормировки двумерной плотности вероятности:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W(x;y)\,dx\,dy=1. \tag{3.52}$$

Свойство 4. Для системы случайных величин (X; Y) с неограниченной областью возможных значений это свойство записывается следующим образом:

$$\lim_{x\to-\infty}W(x;\,y)=\lim_{y\to-\infty}W(x;\,y)=\lim_{\substack{x\to-\infty\\ y\to-\infty}}W(x;\,y)=0;$$

$$\lim_{x\to\infty}W(x;\,y)=\lim_{y\to\infty}W(x;\,y)=\lim_{\substack{x\to\infty\\ y\to\infty}}W(x;\,y)=0.$$

Для системы случайных величин (X; Y), возможные значения которых ограничены некоторой областью Ω, данное свойство записывается в виде:

$$W(x;\,y)=0,\ \text{при}\ x\not\subset\Omega\ \text{и (или)}\ y\not\subset\Omega,$$

т.е. при значениях x и (или) y, не принадлежащих области значений случайных величин (X; Y), плотность вероятности равна нулю.

Пример 3.5. В качестве примера рассмотрим систему случайных величин (X; Y), область возможных значений которых ограничена прямоугольником $a < X < b$; $c < Y < d$ (см. рис. 3.16), а вероятность попадания случайных величин (X; Y) в элементарную площадку $\Delta S=\Delta x\Delta y$ распределена по области возможных значений равномерно.

Построим графики распределения двумерной плотности вероятности и функции распределения этих случайных величин. Равномерное распределение вероятностей предполагает, что в пределах области возможных значений Ω двумерная плотность вероятности равна некоторой постоянной величине C:

$$W(x;\,y)=C\ \ \text{при}\ x\subset\Omega\ \text{и}\ y\subset\Omega.$$

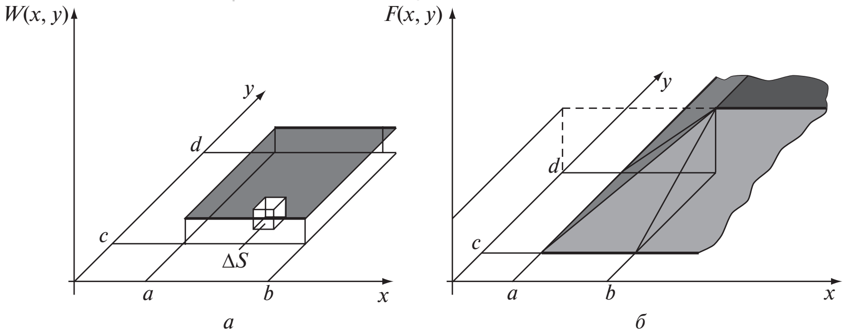


Рис. 3.16. Двумерные плотность вероятности и функция распределения системы равномерно распределенных случайных величин

В соответствии со свойством 4 за областью Ω $W(x; y) = 0$, и условие нормировки (3.52) для рассматриваемой системы случайных величин можно записать в виде:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x)dx = \int_{a}^{b}\int_{c}^{d} c\,dx\,dy = c(b-a)(d-c) = 1 \;\Rightarrow\; c = \frac{1}{(b-a)(d-c)}.$$

Отсюда для двумерной плотности вероятности получим:

$$W(x;y) = \begin{cases} \dfrac{1}{(b-a)(d-c)} & \text{при} \quad a \le x < b \text{ и } c \le y < d; \\ 0 & \text{при} \quad x < a,\ x \ge b \text{ и(или) } y < c,\ y \ge d. \end{cases} \qquad (3.53)$$

Плотность вероятности $W(x;\ y)$ для рассматриваемой системы случайных величин $(X;\ Y)$ показана заштрихованной плоскостью на рис. 3.16, *а*.

Аналитическое выражение для двумерной функции распределения $F(x;\ y)$ получим подстановкой формулы (3.53) в (3.51):

$$F(x;\ y) = \int_{a}^{x}\int_{c}^{y} \frac{1}{(b-a)(d-c)} dx\,dy.$$

В результате поочередного интегрирования по x и по y и подстановки пределов интегрирования получим:

$$F(x;\ y) = \frac{1}{(b-a)(d-c)} x \cdot y \Big|_{a}^{x} \Big|_{c}^{y} \quad \Rightarrow$$

$$\Rightarrow \quad F(x;\,y)=\begin{cases}0 & \text{при} \quad x<a \text{ и(или) } y<c;\\ \dfrac{(x-a)(y-c)}{(b-a)(d-c)} & \text{при} \quad a\le x<b \text{ и } c\le y<d;\\ 1 & \text{при} \quad x\ge b \text{ и } y\ge d.\end{cases} \qquad (3.54)$$

Двумерная функция распределения $F(x;\,y)$ рассматриваемой системы случайных величин $(X;\,Y)$ показана на рис. 3.16, *б* заштрихованной плоскостью. Из формулы (3.54) и графика функции распределения (см. рис. 3.16, *б*) следует, что для полученного аналитического выражения $F(x;\,y)$ выполняются все свойства двумерной функции распределения.

3.6. Плотности вероятности составляющих системы непрерывных случайных величин

Знание закона распределения системы непрерывных случайных величин $(X;\,Y)$, заданного двумерной функцией распределения или двумерной плотностью вероятности, позволяет определить одномерные плотности вероятности случайных величин X и Y. Предположим, что нам известны аналитические выражения для $W(x;\,y)$ и $F(x;\,y)$. На основании свойства 3, формулы (3.44) по двумерной функции распределения $F(x;\,y)$ мы легко можем определить одномерные функции распределения случайных величин X и Y:

$$F(x)=F(x;\,y=\infty);$$

$$F(y)=F(x=\infty;\,y)\,.$$

С учетом формулы (3.51) эти равенства можно переписать в виде:

$$F(x)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{\infty}W(x;\,y)\,dx\,dy;$$

$$F(y)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{y}W(x;\,y)\,dx\,dy.$$

Одномерные плотности вероятности $W(x)$ и $W(y)$ можно найти дифференцированием функций распределения $F(x)$ и $F(y)$ по соответствующим переменным (3.2). Используя правила дифференцирования интегралов с переменным верхним пределом интегрирования, для одномерных плотностей вероятности $W(x)$ и $W(y)$ получим:

$$W(x) = \frac{dF(x)}{dx} = \int_{-\infty}^{\infty} W(x;\, y)\,dy;$$
$$W(y) = \frac{dF(y)}{dy} = \int_{-\infty}^{\infty} W(x;\, y)\,dx. \tag{3.55}$$

Таким образом, для нахождения одномерных плотностей $W(x)$ и $W(y)$ нужно проинтегрировать в бесконечных пределах двумерную плотность вероятности по лишней переменной.

В некоторых случаях после вычисления $W(x)$ и $W(y)$ можно увидеть, что двумерная плотность вероятности равна произведению одномерных:

$$W(x;\, y) = W(x)W(y). \tag{3.56}$$

При соблюдении равенства (3.56) непрерывные случайные величины X и Y являются независимыми. Для зависимых непрерывных случайных величин равенство (3.56) не выполняется. По аналогии с теоремой умножения вероятностей для зависимых случайных событий для зависимых непрерывных случайных величин можно записать:

$$W(x;\, y) = W(x)W(y/x) = W(y)W(x/y), \tag{3.57}$$

где $W(x/y)$ — условная плотность вероятности случайной величины X при условии, что случайная величина Y принимает фиксированное значение $Y = y$; $W(y/x)$ — условная плотность вероятности случайной величины Y при условии, что случайная величина X принимает фиксированное значение $X = x$.

Из формулы (3.57) следует, что условные плотности вероятности можно определить как результат деления двумерной плотности вероятности на одномерную:

$$W\left(x/y\right) = \frac{W(x;\, y)}{W(y)};$$
$$W\left(x/y\right) = \frac{W(x;\, y)}{W(x)}. \tag{3.58}$$

Для зависимых случайных величин условные плотности вероятности не равны одномерным плотностям вероятности:

$$W(x/y) \neq W(x);\; W(y/x) \neq W(y). \tag{3.59}$$

Таким образом, условие (3.59) является условием зависимости случайных величин X и Y. Если $W(x/y) = W(x)$ и $W(y/x) = W(y)$, то равенство (3.57) преобразуется в равенство (3.56), являющееся условием независимости случайных величин.

Условные плотности вероятности должны удовлетворять всем свойствам одномерных плотностей вероятности (см. подраздел 3.2). Покажем это на примере условия нормировки, которое для условной плотности вероятности $W(y/x)$ можно записать в виде:

$$\int_{-\infty}^{\infty} W(y/x)dy = 1. \qquad (3.60)$$

Для доказательства подставим в подынтегральное выражение (3.60) правую часть соответствующего равенства (3.58):

$$\int_{-\infty}^{\infty} \frac{W(x;y)}{W(x)}dy = \frac{1}{W(x)}\int_{-\infty}^{\infty} W(x;y)dy = \frac{W(x)}{W(x)} = 1.$$

Плотность вероятности $W(x)$ мы выносим за знак интеграла, так как она не зависит от переменной интегрирования y. Оставшийся интеграл в соответствии с формулой (3.55) будет равен $W(x)$.

Определим условные плотности вероятности для случайных величин, рассматриваемых в примере 3.5.

Определим вначале одномерные плотности вероятности. На основании формул (3.53) и (3.55) получим:

$$W(x) = \int_{c}^{d} \frac{1}{(b-a)(d-c)}dy = \frac{(d-c)}{(b-a)(d-c)} = \frac{1}{(b-a)}; \quad a \le x \le b;$$

$$W(y) = \int_{a}^{b} \frac{1}{(b-a)(d-c)}dx = \frac{(b-a)}{(b-a)(d-c)} = \frac{1}{(d-c)}; \quad c \le y < d.$$

Из полученных формул для одномерных плотностей вероятности и формулы (3.53) следует, что выполняется условие (3.56):

$$W(x;y) = W(x)\cdot W(y) = \frac{1}{(b-a)(d-c)} \quad \text{при } c \le y < d \text{ и } a \le x < b.$$

Из этого можно сделать вывод, что случайные величины X и Y, имеющие равномерное распределение совместной плотности вероятности $W(x;\ y)$ в прямоугольной области возможных значений $c \le y < d$ и $a \le x < b$, являются независимыми случайными величинами.

Рассмотрим другой случай.

Пример 3.6. Рассмотрим систему случайных величин (X; Y), равномерно распределенную внутри области возможных значений Ω, представляющую собой круг радиуса r с центром в начале координат (рис. 3.17).

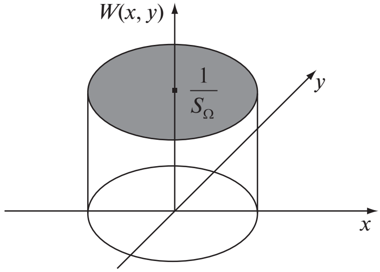


Рис. 3.17. Совместная плотность вероятности системы случайных величин (X; Y)

Так как в пределах области возможных значений Ω совместная (двумерная) плотность вероятности имеет постоянное значение $W(x; y) = c$, то условие нормировки совместной плотности вероятности этих случайных величин можно записать в виде:

$$\iint_{\Omega} W(x;\, y)\,dx\,dy = cS_\Omega = 1;\;\; W(x;\, y) = c = \frac{1}{S_\Omega} = \frac{1}{\pi r^2}\,,$$

где S_Ω — площадь круга, определяющего область возможных значений случайных величин X и Y.

Граница области возможных значений Ω описывается уравнением окружности $x^2 + y^2 = r^2$. Исходя из этого, для плотности вероятности $W(x; y)$ можно записать

$$W(x;\, y) = \begin{cases} \dfrac{1}{\pi r^2} & \text{при} \quad x^2 + y^2 \le r^2; \\ 0 & \text{при} \quad x^2 + y^2 > r^2. \end{cases}$$

Определим одномерные плотности вероятности случайных величин X и Y, для этого проинтегрируем $W(x; y)$ по лишним переменным. Из уравнения границы области возможных значений Ω при интегрировании по x пределы интегрирования будут равны:

$$-\sqrt{r^2 - y^2} \le x < \sqrt{r^2 - y^2}\,,$$

а при интегрировании по y:

$$-\sqrt{r^2 - x^2} \le y < \sqrt{r^2 - x^2}\,.$$

В соответствии с формулой (3.55) и указанными пределами интегрирования для одномерной безусловной плотности вероятности $W(x)$ получим:

$$W(x)=\int\limits_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{1}{\pi r^2}dy=\frac{2\sqrt{r^2-x^2}}{\pi r^2} \quad \text{при } -r\le x<r; \tag{3.61}$$

$$W(x)=0 \quad \text{при } -r>x\ge r.$$

Аналогично получается одномерная безусловная плотность вероятности $W(y)$:

$$W(y)=\begin{cases}\dfrac{2\sqrt{r^2-y^2}}{\pi r^2} & \text{при} \quad -r\le y<r;\\ 0 & \text{при} \quad -r>y\ge r.\end{cases} \tag{3.61 a}$$

На рисунке 3.18 приведен график плотности вероятности $W(y)$ при $r = 1$. Из формул, полученных для $W(x)$ и $W(y)$, следует, что график плотности вероятности $W(x)$ будет иметь такой же вид, как и график плотности вероятности $W(y)$.

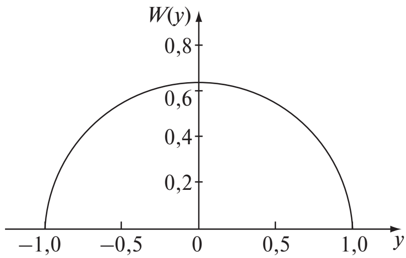


Рис. 3.18. Одномерная плотность вероятности $W(y)$

В примере 3.5 система случайных величин $(X;\ Y)$ имела равномерное распределение совместной плотности вероятности $W(x;\ y)$. Одномерные плотности вероятности $W(x)$ и $W(y)$ также имели равномерное распределение. В данном примере совместная плотность вероятности $W(x;\ y)$ также имеет равномерное распределение, но одномерные вероятности $W(x)$ и $W(y)$ распределены в пределах области возможных значений x и y неравномерно.

Определим условные плотности вероятности $W(x/y)$ и $W(y/x)$ по формулам (3.58):

$$W(x/y)=\begin{cases}\dfrac{1}{\pi r^2}:\dfrac{2\sqrt{r^2-y^2}}{\pi r^2}=\dfrac{1}{2\sqrt{r^2-y^2}} & \text{при} \quad x^2+y^2\le r^2;\\ 0 & \text{при} \quad x^2+y^2>r^2.\end{cases}$$

Для $W(y/x)$ получим аналогичную формулу:

$$W(y/x) = \begin{cases} \dfrac{1}{2\sqrt{r^2 - x^2}} & \text{при} \quad x^2 + y^2 \le r^2; \\ 0 & \text{при} \quad x^2 + y^2 > r^2. \end{cases} \tag{3.62}$$

Графики условной плотности вероятности $W(y/x)$ при $r = 1$ и фиксированных значениях случайной величины X, равных $x_1 = 0{,}4$ и $x_2 = 0{,}8$, приведены на рис. 3.19.

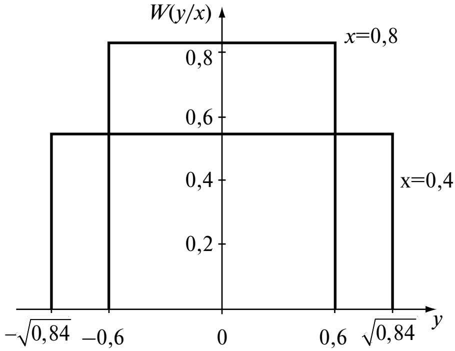


Рис. 3.19. Условная плотность вероятности $W(y/x)$

Из формулы (3.62) следует, что условная плотность вероятности $W(y/x)$ не зависит от y и, значит, имеет равномерный закон распределения (x имеет некоторое фиксированное значение). При $r = 1$ и $x = 0{,}4$ область возможных значений y будет определяться интервалом:

$$-\sqrt{r^2 - x^2} \le y < \sqrt{r^2 - x^2} \Rightarrow -\sqrt{0{,}84} \le y < \sqrt{0{,}84}.$$

При $x = 0{,}8$ область возможных значений y определяется интервалом $-0{,}6 \le y \le 0{,}6$. Соответственно для условных плотностей $W(y/x)$ внутри указанных интервалов получим:

$$W(y/x = 0{,}4) = \frac{1}{2\sqrt{0{,}84}};$$

$$W(y/x = 0{,}8) = \frac{1}{2\sqrt{0{,}36}} = \frac{1}{1{,}2}.$$

Аналогичным образом можно построить графики условной плотности вероятности $W(x/y)$, которая также имеет равномерный закон распределения.

Из сравнения формул (3.61) и (3.62), а также графиков плотностей вероятностей, приведенных на рис. 3.18 и 3.19, видно, что $W(y/x) \neq \neq W(y)$. Отсюда следует, что в рассматриваемом примере 3.6 непрерывные случайные величины X и Y являются зависимыми.

3.7. Моменты распределения системы непрерывных случайных величин

Моментом μ_{kl} порядка $(k + l)$ системы случайных величин $(X; Y)$ относительно постоянных c_1 и c_2 называется математическое ожидание произведения k-й степени отклонения $(X - c_1)$ на l-ю степень отклонения $(Y - c_2)$:

$$\mu_{kl} = M[(X - c_1)^k (Y - c_2)^l] .$$

Если $c_1 = 0$ и $c_2 = 0$, то моменты называются начальными и обозначаются ν_{kl}. В данной формуле под знаком усреднения M находятся две случайные величины X и Y, поэтому для нахождения математического ожидания (усреднения) произведения $X^k Y^l$ это произведение нужно умножить на совместную плотность вероятности этих случайных величин $W(x; y)$ и полученное произведение проинтегрировать по переменным x и y:

$$\nu_{kl} = M\left[X^k Y^l\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^l W(x; y)\,dx\,dy. \qquad (3.63)$$

Нетрудно видеть, что

$$\nu_{10} = M\left[X^1 Y^0\right] = M\left[X\right] = m_x;\ k = 1;\ l = 0;$$

$$\nu_{01} = M\left[X^0 Y^1\right] = M\left[Y\right] = m_y;\ k = 0;\ l = 1,$$

т.е. начальные моменты первого порядка ν_{10} и ν_{01} равны соответственно математическим ожиданиям случайных величин X и Y. Это также следует из формулы (3.63):

$$\nu_{10} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^1 y^0 W(x; y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x W(x, y)\,dx\,dy = \int_{-\infty}^{\infty} x W(x)\,dx = m_x;$$

$$\nu_{01} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y W(x; y)\,dx\,dy = \int_{-\infty}^{\infty} y W(y)\,dy = m_y. \qquad (3.64)$$

При $c_1 = m_x$ и $c_2 = m_y$ моменты называются центральными и обозначаются буквой μ_{kl}:

$$\mu_{kl} = M\left[(X-m_x)^k (Y-m_y)^l\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_x)^k (y-m_y)^l W(x;\, y)\,dx\,dy.$$

Из данной формулы следует, что центральные моменты второго порядка μ_{20} и μ_{02} будут равны дисперсиям случайных величин X и Y:

$$\begin{aligned}\mu_{20} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_x)^2 W(x;\, y)\,dx\,dy = \int_{-\infty}^{\infty} (x-m_x)^2 W(x)\,dx = D_x;\ l=0;\\ \mu_{02} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-m_y)^2 W(x;\, y)\,dx\,dy = \int_{-\infty}^{\infty} (y-m_y)^2 W(y)\,dy = D_y;\ k=0.\end{aligned} \tag{3.65}$$

Наряду с безусловными моментами распределения (3.64) и (3.65) для системы непрерывных случайных величин используются также условные моменты распределения. Чаще всего применяется условное математическое ожидание случайных величин:

$$m_{x/y} = M[x/y],\ m_{y/x} = M[y/x].$$

При нахождении условных математических ожиданий усреднение случайной величины осуществляется по условной плотности вероятности:

$$\begin{aligned}m_{x/y} &= \int_{-\infty}^{\infty} xW(x\,/\,y)\,dx = m_x(y);\\ m_{y/x} &= \int_{-\infty}^{\infty} yW(y\,/\,x)\,dy = m_y(x).\end{aligned} \tag{3.66}$$

При определении условных математических ожиданий в результате вычисления определенного интеграла исчезает зависимость от переменной интегрирования и остается зависимость от условной переменной. То есть численные значения $m_{x/y}$ будут зависеть от значений условной переменной y. Аналогичное утверждение справедливо для $m_{y/x}$. Так как $m_{x/y}$ есть некоторое среднее значение случайной величины X, $m_{y/x}$ — некоторое среднее значение случайной величины Y, то зависимость условных математических ожиданий от условной переменной можно записать в виде уравнений:

$$x = m_x(y);\quad y = m_y(x).$$

Данные уравнения называют уравнениями регрессии, а графики зависимости, построенные по этим уравнениям, называются линиями регрессии.

Если случайные величины X и Y независимы, то $W(x/y) = W(x)$ и $W(y/x) = W(y)$. В этом случае, как видно из выражений (3.66) и (3.64), условные математические ожидания будут равны безусловным. То есть для независимых случайных величин

$$m_{x/y} = m_x;\ m_{y/x} = m_y. \tag{3.67}$$

Но обратное утверждение несправедливо. Если выполняются условия (3.67), мы не можем на основании этого утверждать, что случайные величины X и Y независимы. Необходимым и достаточным условием независимости случайных величин является условие (3.56).

Это наглядно видно на примере 3.6. Рассматриваемые в этом примере случайные величины X и Y являются зависимыми. Вместе с тем из графиков, приведенных на рис. 3.18 и 3.19 для безусловной и условной плотностей вероятности, следует, что

$$m_y = m_{y/x} = 0.$$

Как и для дискретных случайных величин (подраздел 2.6, формула (2.21)), особое место среди числовых характеристик занимает смешанный центральный момент второго порядка μ_{11}, который называют корреляционным моментом и обозначают k_{xy}. Для непрерывных случайных величин корреляционный момент определяется по формуле

$$\begin{gathered}\mu_{11} = k_{xy} = M\left[(X - m_x)(Y - m_y)\right] \Rightarrow \\ k_{xy} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x - m_x)(y - m_y)W(x;\, y)\,dx\,dy.\end{gathered} \tag{3.68}$$

Корреляционный момент характеризует степень линейной взаимосвязи случайных величин X и Y. Предположим, что эти случайные величины связаны между собой детерминированной линейной зависимостью

$$Y = aX + b, \tag{3.69}$$

где a и b — постоянные величины.

Если случайная величина X имеет математическое ожидание m_x и дисперсию D_x, то в соответствии со свойствами математического ожидания и дисперсии можно записать:

$$m_y = M\left(aX + b\right) = am_x + b;$$
$$D_y = M\left[(aX + b - am_x - b)^2\right] = a^2 D_x; \qquad (3.70)$$
$$\sigma_y = a\sigma_x.$$

Предположим, что случайная величина X в результате опыта приняла значения x_1, x_2, x_3, ..., x_i, ..., x_n, отмеченные точками на оси x (рис. 3.20).

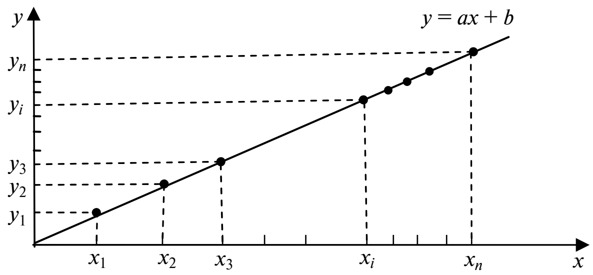


Рис. 3.20. *Р*асположение значений случайных величин X и Y при их линейной зависимости

Тогда случайная величина Y примет значения y_1, y_2, y_3, ..., y_i, ..., y_n, отмеченные точками на оси y. Система случайных величин X и Y характеризуется точками на плоскости с координатами x_i и y_i, которые в соответствии с зависимостью (3.68) будут располагаться на прямой $y = ax + b$.

Определим корреляционный момент случайных величин X, Y. Подставляя значения Y (3.69) и m_y (3.70) в формулу (3.68), получим:

$$k_{xy} = \int_{-\infty}^{\infty} (x - m_x)(ax - am_x)W(x)dx =$$
$$= a\int_{-\infty}^{\infty} (x - m_x)^2 W(x)dx = aD_x = \sigma_x \sigma_y. \qquad (3.71)$$

Аналогично можно показать, что при обратной линейной взаимосвязи случайных величин X и Y, когда

$$Y = -aX + b,$$

корреляционный момент будет равен

$$k_{xy} = -\sigma_x \sigma_y.$$

Если случайные величины X и Y будут независимыми, то $W(x; y) = W(x)W(y)$ и формула (3.68) преобразуется к виду:

$$k_{xy} = \int_{-\infty}^{\infty} (x - m_x) W(x) dx \int_{-\infty}^{\infty} (y - m_y) W(y) dy.$$

В данном выражении первый и второй интегралы являются центральными моментами первого порядка случайных величин X и Y. Эти моменты равны нулю, поэтому корреляционный момент для независимых величин будет равен нулю.

Для того чтобы исключить зависимость корреляционного момента от среднеквадратических отклонений случайных величин X и Y, аналогично дискретным случайным величинам (2.26) вводят безразмерный показатель ρ_{xy} — коэффициент линейной корреляции непрерывных случайных величин:

$$\rho_{xy} = \frac{k_{xy}}{\sigma_x \sigma_y}. \tag{3.71 а}$$

Определим значение корреляционного момента для системы случайных величин, рассматриваемых в примере 3.6. Из формул (3.61) и графика на рис. 3.18 следует, что распределение случайных величин X и Y симметрично относительно начала координат, поэтому математические ожидания $m_x = m_y = 0$. Подставляя в выражение (3.68) двумерную плотность вероятности случайных величин X, Y, найдем:

$$W(x; y) = \frac{1}{\pi r^2}; \; m_x = m_y = 0,$$

для корреляционного момента получим

$$k_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyW(x; y) dx\, dy = \int_{-r}^{r} \frac{x}{\pi r^2} dx \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} y\, dy.$$

Для всех x и y, удовлетворяющих условию $x_2 + y_2 = r_2$, осуществляется интегрирование по x и y нечетных функций в симметричных пределах. Поэтому в результате интегрирования получим, что корреляционный момент рассматриваемых зависимых случайных величин равен нулю. Отсюда можно сделать вывод, что равенство нулю корреляционного момента не является достаточным условием независимости случайных величин. В этом случае говорят, что случайные величины X и Y зависимы, но некоррелированы ($k_{xy} = 0$).

3.8. Двумерная плотность вероятности системы нормально распределенных случайных величин

Система случайных величин (X; Y) имеет нормальный закон распределения, если двумерная плотность вероятности этих случайных величин определяется формулой

$$W(x;y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\cdot\exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\times\right.$$
$$\left.\times\left[\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2\rho_{xy}(x-m_x)(y-m_y)}{\sigma_x\sigma_y}+\frac{(y-m_y)^2}{\sigma_y^2}\right]\right\}, \qquad (3.72)$$
$$-\infty<x<\infty;\quad -\infty<y<\infty;$$

где $m_x, m_y, \sigma_x, \sigma_y$ — математические ожидания и среднеквадратичные отклонения случайных величин X и Y; ρ_{xy} — коэффициент линейной корреляции случайных величин X и Y.

Вычислим одномерные безусловные и условные плотности вероятности случайных X и Y. Для безусловной плотности вероятности $W(y)$ можно записать:

$$W(y)=\frac{\exp\left[-\frac{(y-m_y)^2}{2\sigma_y^2(1-\rho_{xy}^2)}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\times\right.$$
$$\left.\times\left[\frac{(x-m_x)^2}{\sigma_x^2}-\frac{2\rho_{xy}(x-m_x)(y-m_y)}{\sigma_x\sigma_y}\right]\right\}dx. \qquad (3.73)$$

В подынтегральной функции показатель степени экспоненты дополним до полного квадрата. Для этого в квадратных скобках показателя экспоненты добавим слагаемые

$$\frac{\rho_{xy}^2(y-m_x)^2}{\sigma_y^2}+\left(-\frac{\rho_{xy}^2(y-m_y)^2}{\sigma_y^2}\right).$$

В результате выражение (3.73) можно преобразовать к виду:

$$W(y) =$$

$$= \frac{\exp\left[-\frac{(y-m_y)^2 - \rho_{xy}^2 (y-m_y)^2}{2\sigma_y^2 (1-\rho_{xy}^2)}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \cdot \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{2}\left[\frac{(x-mx)}{\sigma_x\sqrt{1-\rho_{xy}^2}} - \frac{\rho_{xy}(y-m_y)}{\sigma_y\sqrt{1-\rho_{xy}^2}}\right]^2\right\} dx.$$

После замены переменной получим:

$$\frac{x-m_x}{\sigma_x\sqrt{1-\rho_{xy}}} - \frac{\rho_{xy}(y-m_y)}{\sigma_y\sqrt{1-\rho_{xy}}} = t,$$

интеграл в полученном выражении преобразуется в интеграл Пуассона:

$$W(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{(y-m_y)^2}{2\sigma_y^2}\right] \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt .$$

В данной формуле интеграл равен $\sqrt{2\pi}$, поэтому для безусловной одномерной плотности вероятности $W(y)$ окончательно получим:

$$W(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{(y-m_y)^2}{2\sigma_y^2}\right]. \tag{3.74}$$

Данная формула полностью совпадает с формулой (3.14) для нормального закона распределения непрерывной случайной величины.

Выполнив аналогичные преобразования при интегрировании (3.72) по y, получим одномерную плотность вероятности случайной величины X.

$$W(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-m_x)^2}{2\sigma_x^2}\right]. \tag{3.75}$$

Поделив двумерную плотность вероятности $W(x; y)$ формула (3.72) на одномерные плотности вероятности (3.74) и (3.75), получим условные плотности вероятности случайных величин X и Y.

$$W(x / y) =$$

$$= \frac{W(x; y)}{w(y)} = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho_{xy}^2}} \cdot \exp\left\{-\frac{\left[x-m_x-\rho_{xy}\frac{\sigma_x}{\sigma_y}(y-m_y)\right]^2}{2\sigma_x^2(1-\rho_{xy}^2)}\right\}; \tag{3.76}$$

$$W(y/x) =$$

$$\frac{W(x;y)}{w(x)} = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho_{xy}^2}} \cdot \exp\left\{-\frac{\left[y - m_y - \rho_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x)\right]^2}{2\sigma_y^2(1-\rho_{xy}^2)}\right\}. \quad (3.77)$$

Сравнивая формулы (3.76) и (3.77) с нормальным законом распределения (3.74) и (3.75), видим, что условные плотности вероятности $W(x/y)$ и $W(y/x)$ также имеют нормальный закон распределения с условными математическими ожиданиями

$$m_{x/y} = m_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(y - m_y);$$
$$m_{y/x} = m_y + \rho_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x) \quad (3.78)$$

и условными среднеквадратичными отклонениями

$$\sigma_{x/y} = \sigma_x\sqrt{1-\rho_{xy}^2};$$
$$\sigma_{y/x} = \sigma_y\sqrt{1-\rho_{xy}^2}.$$

Формулы (3.78) позволяют записать уравнение регрессии одной случайной величины на другую:

$$x = m_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(y - m_y);$$
$$y = m_y + \rho_{xy}\frac{\sigma_y}{\sigma_x}(x - m_x). \quad (3.79)$$

Из данных уравнений регрессии и условных плотностей вероятности (3.76) и (3.77) можно сделать следующий вывод. При значении случайной величины $x(y)$, вычисленной по первому (второму) уравнению регрессии, условная плотность вероятности $W(x/y)$ ($W(y/x)$ имеет максимальное значение.

По формулам (3.79) при $m_x = 2$; $m_y = 1$; $\sigma_x = 1$; $\sigma_y = 2$ произведен расчет и построены линии регрессии для коэффициентов корреляции $\rho_{xy} = 0{,}7$ и $\rho_{xy} = -0{,}7$. Графики линий регрессии приведены на рис. 3.21.

На приведенных графиках показано, что линии регрессии пересекаются в точке с координатами $x = m_x$; $y = m_y$. При положительном коэф-

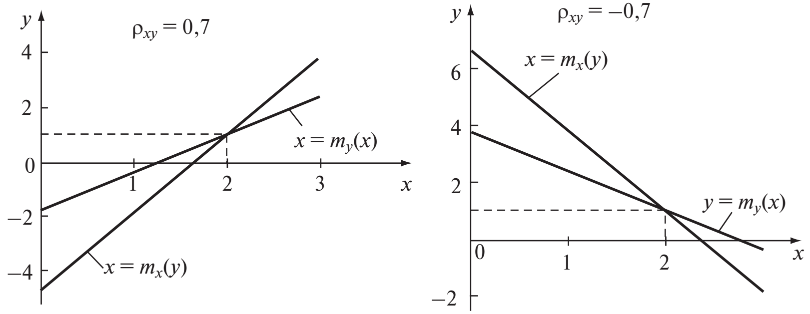


Рис. 3.21. Линии регрессии нормально распределенных случайных величин: *а* — $\rho_{xy} = 0,7$; *б* — $\rho_{xy} = -0,7$

фициенте корреляции линии регрессии имеют положительный наклон к оси *0x*. При отрицательном коэффициенте корреляции тангенс угла наклона линий регрессии к оси *0x* также отрицательный. Из формул (3.79) следует, что для некоррелированных случайных величин *X* и *Y*, когда $\rho_{xy} = 0$, уравнения регрессии имеют вид:

$$x = m_x; \quad y = m_y,$$

т.е. при $\rho_{xy} = 0$ линии регрессии параллельны осям *0x*, *0y* и перпендикулярны между собой.

Из формулы (3.72) и формул (3.74) и (3.75) следует, что при $\rho_{xy} = 0$ двумерная плотность вероятности $W(x; y)$ равна произведению одномерных:

$$W(x; y) = W(x)W(y).$$

Из этого следует, что для нормально распределенных случайных величин равенство нулю коэффициента корреляции свидетельствует также о независимости этих случайных величин. Для случайных величин, закон распределения которых отличается от нормального, это утверждение несправедливо.

На рисунке 3.22 приведены графики зависимости двумерной плотности вероятности, рассчитанные по формуле (3.72) при различных значениях m_x, σ_x, m_y, σ_y и значениях коэффициента корреляции ρ_{xy}.

Двумерная плотность вероятности $W(x; y)$ представляет собой криволинейную поверхность в трехмерном пространстве. Форма этой поверхности зависит от значений математических ожиданий (m_x; m_y); среднеквадратичных отклонений ($\sigma_x; \sigma_y$) и коэффициента корреляции случайных величин.

Если поверхность двумерной плотности вероятности $W(x; y)$ пересечь плоскостью $F(x; y) = C$, параллельной плоскости $x0y$, то линия пересечения будет представлять собой эллипс, который называют эллипсом равной плотности вероятности или эллипсом рассеяния случайных величин X и Y. Для получения уравнения эллипса рассеяния приравняем правую часть формулы (3.72) постоянной величине C и получим:

$$2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xx}^2}C =$$
$$= \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho_{xy}(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]\right\}.$$

Если прологарифмировать левую и правую части полученного выражения, то для эллипса рассеяния получим уравнение

$$\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho_{xy}(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} = 2(1-\rho_{xy}^2)C_1, \quad (3.80)$$

где $C_1 = -\ln\left|2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}C\right|$ — значение постоянной C, выбирается таким, чтобы произведение, стоящее под логарифмом, было меньше единицы.

В этом случае в правой части уравнения (3.80) получится положительное число.

Из уравнения (3.80) следует, что центр эллипса рассеяния расположен в точке с координатами $x = m_x$, $y = m_y$ (рис. 3.23). При $\rho_{xy} = 0$, т.е. если случайные величины X и Y являются некоррелированными и независимыми, эллипс рассеяния определяется уравнением

$$\frac{(x-m_x)^2}{\sigma_x^2} + \frac{(y-m_y)^2}{\sigma_y^2} = -2\ln(2\pi\sigma_x\sigma_y C).$$

При $\rho_{xy} = 0$ оси эллипса рассеяния параллельны координатным осям $0x$ и $0y$, а размеры полуосей зависят от среднеквадратичных отклонений σ_x и σ_y.

При $\rho_{xy} = \pm 1$ уравнение (3.80) преобразуется к виду:

$$\frac{(x-m_x)}{\sigma_x} \mp \frac{(y-m_y)}{\sigma_y} = 0,$$

или

$$y = \pm\frac{\sigma_y}{\sigma_x}(x-m_x) + m_y.$$

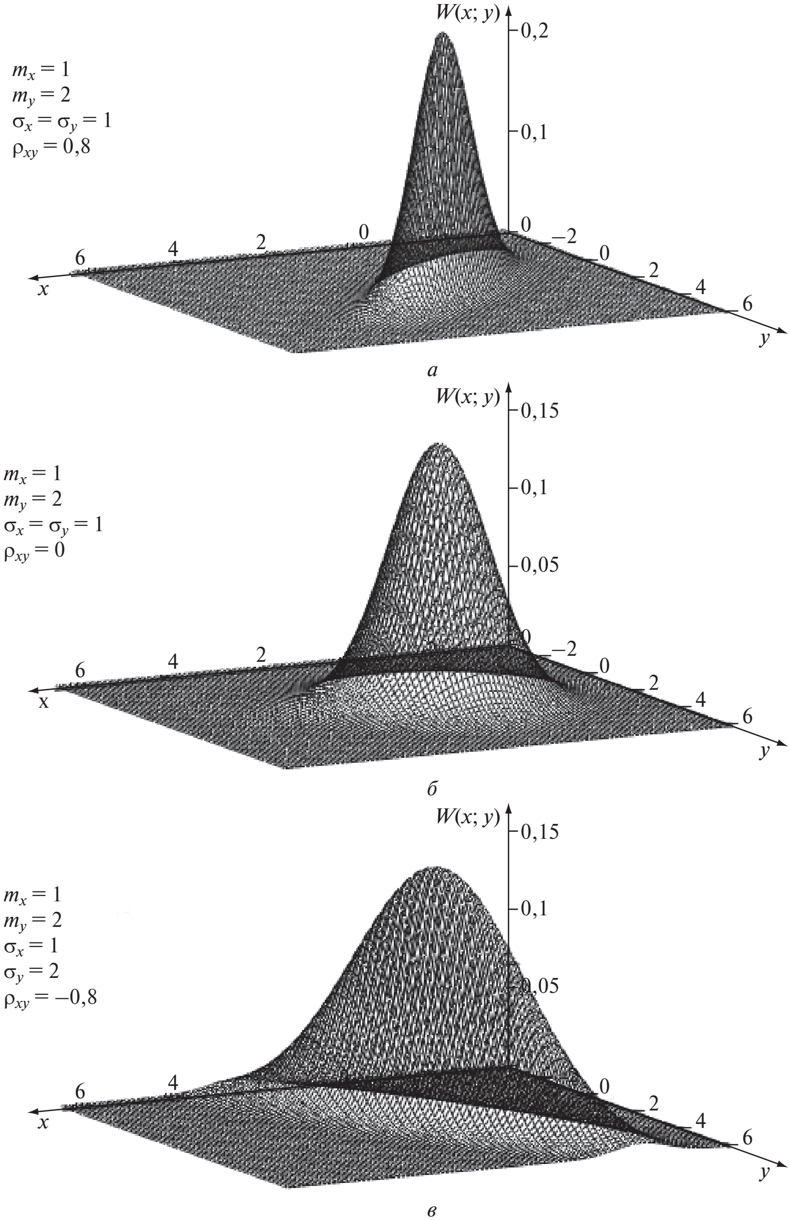


Рис. 3.22. Графики двухмерной плотности вероятности нормально распределенных случайных величин

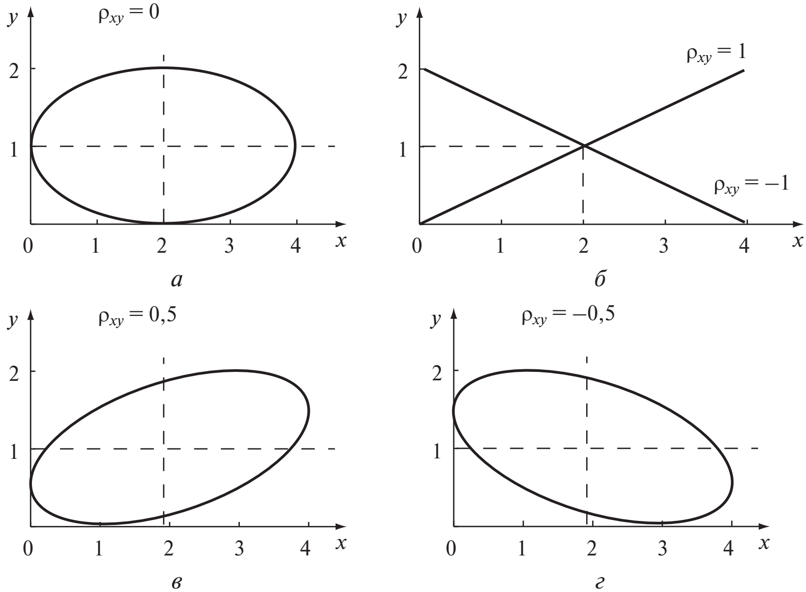


Рис. 3.23. Эллипсы рассеяния нормально распределенных случайных величин при различных значениях коэффициента корреляции:
а — $\rho_{xy} = 0$; *б* — $\rho_{xy} = 1$; $\rho_{xy} = -1$; *в* — $\rho_{xy} = 0{,}5$; *г* — $\rho_{xy} = -0{,}5$

То есть при $\rho_{xy} = \pm 1$ эллипс рассеяния преобразуется в прямую линию, проходящую через точку с координатами $x = m_x$, $y = m_y$, с тангенсом угла наклона к оси $0x$, равным $\pm\sigma_y / \sigma_x$.

При $|\rho_{xy}| < 1$ ориентация эллипса относительно координатных осей зависит от знака и значения коэффициента корреляции.

На рисунке 3.23 приведены графики эллипсов рассеяния, рассчитанные по формуле (3.80) для $m_x = 2$; $\sigma_x = 2$; $m_y = 1$; $\sigma_y = 1$ и $\rho_{xy} = 1$; 0,5; 0; –0,5.

3.9. Примеры решения задач

Задача 3.1. При проведении маркетинговых исследований установлено, что цена на некоторый вид товара имеет равномерное распределение в интервале от 1200 до 1800 руб. Определить функцию распределения и вероятность того, что цена товара будет меньше 1400 руб. Определить математическое ожидание, дисперсию и среднее квадратичное отклонение цены товара.

Решение. Плотность вероятности случайной величины x, имеющей равномерное распределение, определяется формулой (3.35)

$$w(x)=\begin{cases}\dfrac{1}{b-a} & \text{при} \quad a\le x<b;\\ 0 & \text{при} \quad x<a,\ x\ge b.\end{cases}$$

По условию задачи минимальная цена товара равна $a = 1200$ руб., а максимальная $b = 1800$ руб., и плотность вероятности цены товара x можно записать в виде:

$$w(x)=\begin{cases}\dfrac{1}{600} & \text{при} \quad 1200\le x<1800;\\ 0 & \text{при} \quad x<1200,\ x\ge 1800.\end{cases}$$

Функция распределения равномерно распределенной случайной величины определяется формулой (3.36) и для решаемой задачи может быть записана в виде:

$$F(X)=\begin{cases}0 & \text{при} \quad x<1200;\\ \dfrac{x-1200}{600} & \text{при} \quad 1200\le x<1800;\\ 1 & \text{при} \quad x\ge 1800.\end{cases}$$

Вероятность попадания случайной величины в некоторый интервал $c \le x < d$ определяется формулой (3.4). По условию решаемой задачи с $\to -\infty$; $d = 1400$ руб.

$$P(x<d)=\int_a^d w(x)dx=F(d).$$

В интеграле нижний предел взят равным $a = 1200$, так как при $x < a = 1200$ и $W(x) = 0$:

$$P(x<1400)=\int_{1200}^{1400}\frac{1}{600}dx=\frac{1400-1200}{600}=\frac{1}{3}.$$

Вычисляя вероятность по функции распределения, получим тот же результат.

Математическое ожидание равномерно распределенной случайной величины определяется формулой (3.37). Для математического ожидания цены товара получили:

$$m(x)=\frac{a+b}{2}=\frac{1200+1800}{2}=1500 \text{ руб.}$$

Дисперсию цены товара найдем по формуле (3.38):

$$D(X)=\frac{(b-a)^2}{12}=\frac{(1800-1200)^2}{12}=\frac{36\cdot 10^4}{12}=3\cdot 10^4 \text{ (руб.)}^2.$$

Среднее квадратичное отклонение цены товара от математического ожидания будет равно:

$$\sigma_x=\sqrt{D(X)}\approx 173{,}2 \text{ руб.}$$

Задача 3.2. Текущая цена акции *X* изменяется случайным образом и может быть описана нормальным законом распределения с математическим ожиданием $m_x = 100$ руб. и средним квадратичным отклонением $\sigma_x = 1{,}2$ руб.

Найти вероятность того, что текущая цена акции будет находиться в пределах $102 \le x < 105$. С помощью правила трех сигм найти границы, в которых будет находиться текущая цена акции и вероятность нахождения цены в этих границах.

Решение. Плотность вероятности нормально распределенной случайной величины определяется формулой (3.14). Для решаемой задачи плотность вероятности запишется в виде:

$$w(x)=\frac{1}{1{,}2\sqrt{2\pi}}\exp\left[-\frac{(x-100)^2}{2{,}88}\right].$$

Вероятность нахождения текущей цены акции в интервале от $x_1 = 102$ до $x_2 = 105$ руб. можно определить по формуле (3.18):

$$P(102\le X<105)=\Phi(z_2)-\Phi(z_1),$$

где $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{t^2}{2}}dt$ — табулированный интеграл вероятности (см. приложение 1):

$$z_1=\frac{102-m_x}{\sigma_x}=\frac{102-100}{1{,}2}\approx 1{,}67;$$

$$z_2=\frac{105-m_x}{\sigma_x}=\frac{105-100}{1{,}2}\approx 4{,}17.$$

По таблице значений интеграла вероятности (см. приложение 1) находим:

$$\Phi(z_2) = 0{,}99998477;\quad \Phi(z_1) = 0{,}95254.$$

Для искомой вероятности получим:

$$P(102 \le X < 105) = 0{,}99998477 - 0{,}95254 \approx 0{,}047.$$

Границы интервала, в которых будет находиться текущая цена акций, по правилу трех сигм записываются в виде:

$$m_x - 3\sigma_x \le X < m_x + 3\sigma_x.$$

Для решаемой задачи данный интервал будет равен:

$$96{,}4 \le X < 103{,}6.$$

Для вероятности нахождения текущей цены акций в данном интервале получим:

$$P(m_x - 3\sigma_x \le X < m_x + 3\sigma_x) =$$
$$= \Phi\left(\frac{m_x + 3\sigma_x - m_x}{\sigma_x}\right) - \Phi\left(\frac{m_x - 3\sigma_x - m_x}{\sigma_x}\right) = \Phi(3) - \Phi(-3).$$

По формуле (3.17) $\Phi(-3) = 1 - \Phi(3)$, отсюда

$$P(m_x - 3\sigma_x \le X < m_x + 3\sigma_x) = 2\Phi(3) - 1 = 2 \cdot 0{,}99865 - 1 = 0{,}9973.$$

Таким образом, вероятность нахождения нармально распределенной случайной величины в интервале от $(m_x - 3\sigma_x)$ до $(m_x + 3\sigma_x)$ равна 0,9973.

Задача 3.3. При подаче заявки на получение кредита в банке время t принятия решения по рассматриваемой заявке является случайной величиной и может быть смоделировано экспоненциальным (показательным) законом распределения с математическим ожиданием $m_t = 32$ ч. Определить время T, в течение которого ($t < T$) решение по рассматриваемой заявке будет принято с вероятностью $P(t < T) = 0{,}9$.

Решение. Плотность вероятности и функция распределения экспоненциально распределенной случайной величины определяются формулами (3.23) и (3.25):

$$W(t) = \lambda e^{-\lambda t};\ F(t) = 1 - e^{-\lambda t};\ t \ge 0,$$

где λ — параметр экспоненциального распределения.

Значение параметра λ может быть найдено из формулы (3.24) по известному значению математического ожидания:

$$m_t = \frac{1}{\lambda};\ \lambda = \frac{1}{m_t} = \frac{1}{32}.$$

Вероятность $P(t < T)$ может быть найдена по функции распределения:

$$P(t < T) = F(T) = 1 - e^{-\lambda T};$$
$$0{,}9 = 1 - e^{-\lambda T};$$
$$e^{-\lambda T} = 1 - 0{,}9.$$

Прологарифмируем обе части уравнения и получим:

$$\lambda T = \ln(1 - 0{,}9);$$

$$T = -\frac{1}{\lambda}\ln 0{,}1 = -32\ln 0{,}1 \approx 74 \text{ ч.}$$

Задача 3.4. Сумму денежных средств на расчетных счетах юридических лиц — клиентов банка — можно рассматривать как случайную величину x, распределенную по закону Рэлея. Математическое ожидание этой случайной величины равно $m_x = 100$ млн руб. Какую сумму остатков на расчетных счетах юридических лиц банк может использовать в качестве ресурса $x_р$ для проведения активных операций, чтобы риск неудовлетворения заявок юридических лиц на текущие платежи не превышал 0,25. Сумму заявок юридических лиц на текущие платежи можно считать постоянной и равной $x_т = 20$ млн руб.

Решение. Плотность вероятности и функция распределения случайной величины, имеющей закон распределения Рэлея, определяются формулами (3.31) и (3.32):

$$W(x) = \frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right); \; F(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

где σ — параметр распределения Рэлея, который связан с математическим ожиданием m_x формулой (3.33):

$$m_x = \sqrt{\frac{\pi}{2}}\sigma, \text{ или } \sigma = \sqrt{\frac{\pi}{2}}m_x.$$

Риск неудовлетворения заявок юридических лиц на текущие платежи $x_т$ с учетом того, что часть средств в сумме $X_р$ используется банком, определяется вероятностью соблюдения неравненства

$$P(x - X_р < X_т).$$

В данном неравенстве $(x - X_р)$ — остатки на расчетных счетах юридических лиц с учетом того, что сумма $X_р$ изъята банком и вложена в активные операции. Риск неплатежеспособности банка наступает, если эти остатки окажутся меньше, чем требования юридических лиц по текущим платежам. Эту вероятность также можно записать в виде:

$$P(x < X_т + X_р) = F(X_т + X_р) \le 0,25,$$

где в правой части равенства функция распределения Рэлея при $x = X_т + X_р$

$$1 - \exp\left[-\frac{(X_т + X_р)^2}{2\sigma^2}\right] \le 0,25;$$

$$\exp\left[-\frac{(X_т + X_р)^2}{2\sigma^2}\right] \ge 0,75.$$

После логарифмирования левой и правой частей неравенства получим:

$$-\frac{(X_т + X_р)^2}{2\sigma^2} \ge \ln 0,75;$$

$$X_т + X_р \le \sqrt{2\sigma^2 \ln 0,75}.$$

Подставляя значение $\sigma = \sqrt{\frac{2}{\pi}} m_x$ для суммы, которую банк может использовать как ресурс для активных операций, получим:

$$X_р \le \sqrt{\frac{4}{\pi} m_x^2 \ln 0,75} - X_т.$$

После подстановки численных значений, указанных в условии задачи, $m_x = 10^8$ руб., $X_т = 2 \cdot 10^7$ руб., получим:

$$X_р \le \sqrt{\frac{4}{\pi} 10^{16} \ln 0,75} - 2 \cdot 10^7 \approx 0,605 \cdot 10^8 - 2 \cdot 10^7 = 40,5 \text{ млн руб.}$$

Для того чтобы вероятность риска неплатежеспособности банка по требованиям текущих платежей юридических лиц не превышала 0,25, банк может использовать в качестве ресурса для своих активных операций не более 40,5 млн руб. из суммы средств на расчетных счетах юридических лиц — клиентов банка.

Задача 3.5. Двумерный закон распределения выручки X от реализации и затрат Y на производство продукции определяется плотностью вероятности:

$$W(x; y) =$$

$$= \frac{1}{2\pi\sqrt{0,162}} \exp\left\{-\frac{1}{0,72}\left[\frac{(x-5)^2}{0,9} - \frac{1,2(x-5)(y-3)}{\sqrt{0,45}} + \frac{(y-3)^2}{0,5}\right]\right\}. \qquad (3.81)$$

Определить математическое ожидание и среднее квадратическое отклонение выручки X и затрат Y на производство продукции, коэффициент корреляции ρ_{xy} данных случайных величин. Записать уравнение регрессии затрат на выручку от реализации продукции.

Решение. По виду приведенной двумерной плотности вероятности можно сделать вывод, что случайные величины X и Y имеют нормальный закон распределения, двумерная плотность вероятности которого определяется формулой (3.72). Из сравнения формул (3.72) и (3.81) можно выписать равенства, из которых очень просто определяются математические ожидания, дисперсии и коэффициент корреляции случайных величин X и Y:

$$\frac{(x-m_x)^2}{\sigma_x^2}=\frac{(x-5)^2}{0,9};$$

$$\frac{(y-m_y)^2}{\sigma_y^2}=\frac{(y-3)^2}{0,5};$$

$$\frac{1}{2(1-\rho_{xy}^2)}=\frac{1}{0,72}.$$

Из данных равенств следует, что математическое ожидание выручки от реализации и затрат на производство продукции соответственно равно $m_x = 5$; $m_y = 3$.

Среднее квадратичное отклонение выручки σ_x и затрат σ_y на производство продукции соответственно равно $\sigma_x=\sqrt{0,9}$; $\sigma_y=\sqrt{0,5}$.

Для коэффициента корреляции случайных величин X и Y получим:

$$2(1-\rho_{xy}^2)=0,72 \Rightarrow 1-\rho_{xy}^2=0,36 \Rightarrow \rho_{xy}=\sqrt{1-0,36}=0,8.$$

Уравнение регрессии затрат на производство Y на выручку от реализации продукции X определяется формулой (3.79), которая для рассматриваемого примера может быть записана в виде:

$$x=5+0,8\sqrt{\frac{0,9}{0,5}}(y-3);\quad x\approx 1,073y+1,78.$$

График уравнения регрессии затрат на производство Y на выручку от реализации продукции X приведен на рис. 3.24.

Из графика и полученного уравнения следует, что для увеличения выручки от реализации продукции необходимо увеличивать затраты на ее производство. При увеличении затрат на одну денежную единицу $\Delta y = 1$ можно с наибольшей вероятностью ожидать, что выручка от реализации продукции возрастет на $\Delta x = 1,073$ денежных единиц.

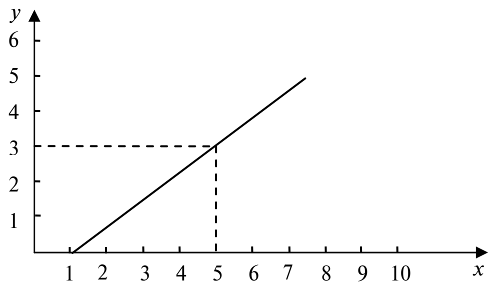


Рис. 3.24. График уравнения регрессии

Вопросы для самоконтроля

1. Приведите понятие и примеры непрерывных случайных величин. Почему для непрерывной случайной величины вероятность $P(X = x_i)$ равна нулю?
2. Как определяется плотность вероятности непрерывной случайной величины, ее связь с функцией распределения?
3. Назовите свойства плотности вероятности непрерывной случайной величины.
4. Что такое интегральная функция распределения непрерывной случайной величины и какова ее связь с плотностью вероятности?
5. Дайте определение математическому ожиданию непрерывной случайной величины.
6. Что такое дисперсия и начальный момент второго порядка непрерывной случайной величины?
7. Что такое нормальный закон распределения непрерывных случайных величин? Дайте характеристику понятиям «плотность вероятности», «математическое ожидание» и «дисперсия».
8. Как определяется функция распределения непрерывной случайной величины, имеющей нормальный закон распределения?
9. Что такое распределение Эрланга? Плотность вероятности, функция распределения, математическое ожидание и дисперсия для распределения Эрланга.
10. Что такое экспоненциальное распределение? Плотность вероятности, функция распределения, математическое ожидание и дисперсия для экспоненциального распределения.
11. Что такое хи-квадрат распределение? Плотность вероятности, функция распределения, математическое ожидание и дисперсия для хи-квадрат распределения.

12. Что такое распределение Вейбула? Плотность вероятности, функция распределения для распределения Вейбула.
13. Что такое распределение Рэлея? Плотность вероятности, функция распределения, математическое ожидание и дисперсия распределения Рэлея.
14. Что такое равномерное распределение? Плотность вероятности, функция распределения, математическое ожидание и дисперсия для равномерного распределения.
15. Сформулируйте закон распределения системы двух непрерывных случайных величин. Что такое двумерная функция распределения? Каковы ее свойства?
16. Дайте понятие двумерной плотности вероятности непрерывных случайных величин. Ее связь с функцией распределения. Каковы свойства двумерной плотности вероятности?
17. Поясните алгоритм определения вероятности попадания системы случайных величин X и Y в заданную область по плотности вероятности и функции распределения.
18. Что такое условные плотности вероятности случайных величин? Раскройте условие зависимости и независимости двух случайных величин.
19. Дайте определение моментам распределения системы двух непрерывных случайных величин.
20. Что такое условные математические ожидания случайных величин? Какие уравнения и линии регрессии одной случайной величины на другую вы знаете?
21. Дайте понятие корреляционному моменту и коэффициенту линейной корреляции двух случайных величин.
22. Что такое двумерная плотность вероятности системы нормально распределенных случайных величин?
23. Дайте комментарий по условным плотностям вероятностей нормально распределенных случайных величин вы знаете?
24. Приведите уравнение и линии регрессии нормально распределенных случайных величин. Какова зависимость положения линий регрессии от коэффициента корреляции и дисперсий случайных величин?
25. Что такое эллипс рассеяния непрерывных случайных величин? Каково положение эллипса рассеяния на плоскости $x0y$ в зависимости от математических ожиданий, дисперсий и коэффициента корреляции случайных величин?

ГЛАВА 4

ФУНКЦИОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

В реальной жизни экономисты при исследовании экономической эффективности производства товаров или услуг часто используют формальные подходы, основанные на количественной взаимосвязи между различными элементами выручки. Приведем некоторые из используемых соотношений:

$$С = КЗ_{пер} + З_{пост.} \tag{4а}$$

Данное соотношение устанавливает взаимосвязь между себестоимостью продукции С, количеством выпускаемой продукции К, переменными затратами на единицу продукции $З_{пер}$ и постоянными затратами на организацию выпуска данного вида продукции $З_{пост}$.

Оптовая цена выпускаемой продукции $Ц_{опт}$ может быть определена по формуле

$$Ц_{опт} = С\,(1+N_р), \tag{4б}$$

где $N_р$ — норма рентабельности.

В условиях рыночной экономики величины, входящие в приведенные соотношения, являются случайными. Если нам известны законы распределения некоторых случайных величин, например К, $З_{пер}$, $З_{пост}$, то, используя методы функционального преобразования непрерывных случайных величин, мы можем определить законы распределения других случайных величин, например С и $Ц_{опт}$, и далее решать задачи, связанные с обеспечением требуемой вероятности безубыточного выпуска продукции.

4.1. Плотность вероятности функции одного случайного аргумента

Предположим, что нам известен закон распределения непрерывной случайной величины X, заданный, например, плотностью вероятности

$W(x)$. Но нас интересует закон распределения случайной величины Y, которая связана со случайной величиной X известной функциональной зависимостью

$$Y = f(X). \tag{4.1}$$

Предположим, что функция (4.1) монотонна, непрерывна и дифференцируема на всем интервале возможных значений случайной величины X. На рисунке 4.1 в качестве примера приведен график функции $y = f(x)$, удовлетворяющий вышеперечисленным условиям.

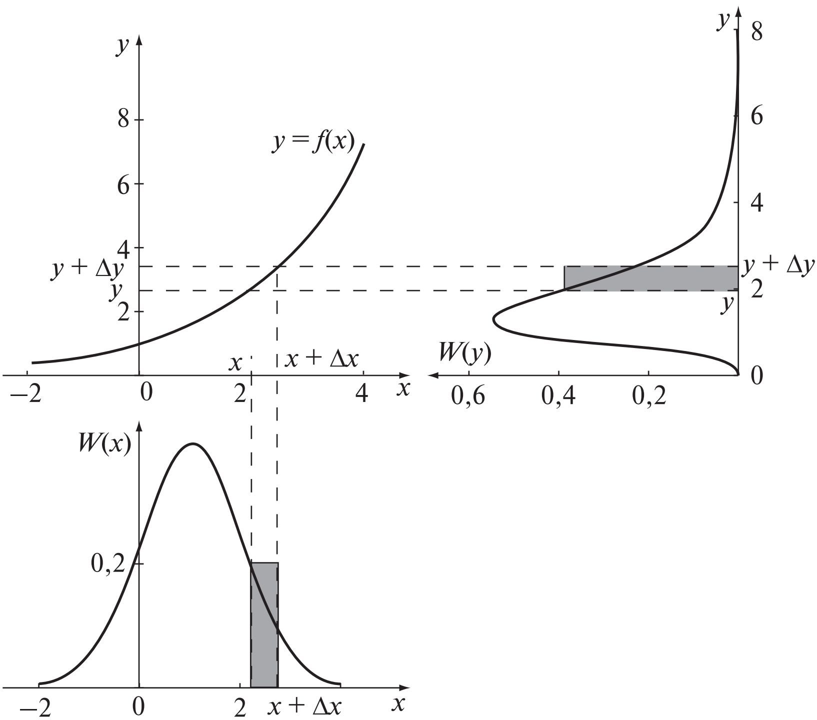


Рис. 4.1. Функциональное преобразование плотности вероятности случайных величин

Для определения плотности вероятности $W(y)$ вычислим вероятность попадания случайной величины X в произвольно выбранный интервал Δx. При достаточно малом интервале Δx для искомой вероятности можно записать:

$$P(x \le X < x + \Delta x) = \int_{x}^{x+\Delta x} W(x)dx \approx W(x)\Delta x. \tag{4.2}$$

При значениях X, принадлежащих интервалу Δx ($X \subset \Delta x$), значения случайной величины Y будут принадлежать интервалу Δy $(Y \subset \Delta y)$, т.е. данные события будут иметь одинаковые вероятности:

$$P(x \le X < x + \Delta x) = P(y \le Y < y + \Delta y). \tag{4.3}$$

Если предположить, что нам известна плотность вероятности $W(y)$, то по аналогии с формулой (4.2) можно записать:

$$P(y \le Y < y + \Delta y) = \int_{y}^{y+\Delta y} W(y)dy \approx W(y)\Delta y. \tag{4.4}$$

На основании приближенных равенств (4.2) и (4.4) равенство (4.3) может быть записано в виде:

$$W(y)\Delta y = W(x)\Delta x,$$

или

$$W(y) = W(x)\frac{\Delta x}{\Delta y}. \tag{4.5}$$

Для получения окончательной формулы для $W(y)$ нужно в правой части второго равенства (4.5) от переменной x перейти к переменной y и устремить приращения Δx и Δy к нулю. Это можно сделать, если из решения уравнения (4.1) мы найдем обратную функцию

$$X = g(Y). \tag{4.6}$$

С учетом найденной обратной функции (4.6) для отношения приращений $\Delta x / \Delta y$ при $\Delta y \to 0$ получим:

$$\lim_{\Delta y \to 0} \frac{\Delta x}{\Delta y} = \lim_{\Delta y \to 0} \frac{g(y + \Delta y) - g(y)}{\Delta y} = \frac{d\,g(y)}{dy}. \tag{4.7}$$

Подставляя выражения (4.6) и (4.7) во второе равенство (4.5), для плотности вероятности $W(y)$ получим:

$$W(y) = W_x\left[g(y)\right] \cdot \left|\frac{d\,g(y)}{dy}\right|, \tag{4.8}$$

где $W_x\left[g(y)\right]$ — аналитическое выражение для плотности вероятности $W(x)$, в которое вместо аргумента x подставлена обратная функция $x = g(y)$.

Производная от обратной функции в уравнении (4.8) берется по абсолютной величине, так как она может иметь отрицательное значение, а плотность вероятности $W(y)$ в соответствии с ее свойствами является функцией неотрицательной $W(y) \geq 0$.

Область возможных значений случайной величины Y находится по формуле (4.1) с учетом области возможных значений случайной величины X.

Пример 4.1. Случайная величина X имеет нормальный закон распределения с математическим ожиданием m_x и дисперсией σ_x^2. Определить плотность вероятности случайной величины Y, если известно, что

$$Y = e^{kX},\ k > 0. \tag{4.9}$$

Решение. Находим обратную функцию $x = g(y)$:

$$x = \frac{1}{k}\ln y. \tag{4.10}$$

Вычисляем производную от обратной функции:

$$\frac{dg(y)}{dy} = \frac{dx}{dy} = \frac{1}{ky}. \tag{4.11}$$

Найдем область возможных значений случайной величины Y. В соответствии с формулой (4.9) получим при $x \to -\infty,\ y \to e^{-k\infty} \to 0,$ при $x \to \infty,\ y \to e^{k\infty} \to \infty.$

Таким образом, случайная величина Y может принимать значения

$$0 < y < \infty.$$

Для нахождения $W(y)$ в аналитическое выражение для $W(x)$ (формула (3.14)) подставляем обратную функцию (4.10):

$$W_x\left[g(y)\right] = \frac{1}{\sqrt{2\pi}\,\sigma_x}\exp\left[-\frac{\left(\frac{1}{k}\ln y - m_x\right)^2}{2\sigma_x^2}\right]. \tag{4.12}$$

Подставляя формулы (4.11) и (4.12) в формулу (4.8), для плотности вероятности $W(y)$ получим:

$$W(y) = \frac{1}{\sqrt{2\pi}\,\sigma_x ky}\exp\left[-\frac{\left(\frac{1}{k}\ln y - m_x\right)^2}{2\sigma_x^2}\right].$$

Графическая иллюстрация решения данной задачи представлена на рис. 4.1. Графики плотностей вероятности $W(x)$ и $W(y)$ построены при $m_x = 1,\ \sigma_x = 1,\ k = 0{,}5$.

Из приведенного примера следует, что закон распределения случайной величины в результате ее функционального преобразования изменяется. Исключением из данного правила является только линейное функциональное преобразование между случайными величинами. При линейном функциональном преобразовании соотношение (4.1) может быть записано в виде:

$$y = ax + b.$$

Для обратной функции и ее производной получим:

$$g(y) \Rightarrow x = \frac{1}{a}(y-b);\ \frac{d\,g(y)}{dy} = \frac{1}{a}.$$

Поэтому на основании формулы (4.8) плотность вероятности $W(y)$ будет равна:

$$W(y) = \left|\frac{1}{a}\right| W_x\left[\frac{1}{a}(y-b)\right]. \qquad (4.13)$$

Из формулы (4.13) следует, что закон распределения случайной величины Y остался таким же, как и для случайной величины X. Произошел только сдвиг плотности вероятности вправо на величину b и изменился масштаб по осям на величину $1/a$, т.е. изменились только параметры закона распределения.

Пример 4.2. Себестоимость выпускаемой продукции С имеет нормальный закон распределения с математическим ожиданием m_c и среднеквадратичным отклонением σ_c. Определить плотность вероятности, математическое ожидание и среднеквадратичное отклонение оптовой цены выпускаемой продукции при норме рентабельности $N_р = 0{,}3$.

Решение. Плотность вероятности себестоимости выпускаемой продукции в соответствии с условием определяется формулой

$$W(c) = \frac{1}{\sqrt{2\pi}\,\sigma_c}\exp\left[-\frac{(c-m_c)^2}{2\sigma_c^2}\right],$$

где m_c и σ_c — математическое ожидание и среднеквадратичное отклонение себестоимости выпускаемой продукции.

Оптовая цена выпускаемой продукции Ц связана с себестоимостью С и нормой рентабельности $N_р$ функциональным соотношением (4б). Эта функциональная зависимость является линейной, поэтому плотность вероятности оптовой цены определяется формулой (4.13), в которой $a = (1 + N_р)$, $b = 0$. Найдем

$$W(Ц) = \frac{1}{\sqrt{2\pi}\,\sigma_с(1+N_р)}\exp\left[-\frac{\left(\frac{Ц}{1+N_р}-m_с\right)^2}{2\sigma_с^2}\right] =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_с(1+N_р)}\exp\left\{-\frac{\left[Ц-(1+N_р)m_с\right]^2}{2\sigma_с^2(1+N_р)^2}\right\}.$$

Из полученного выражения видно, что оптовая цена выпускаемой продукции имеет также нормальный закон распределения. Математическое ожидание (среднее значение) и среднеквадратичное отклонение оптовой цены будут равны:

$$m_ц = (1+ N_р)m_с = 1{,}3m_с;$$
$$\sigma_ц = (1+N_р)\sigma_с = 1{,}3\sigma_с.$$

4.2. Функциональное преобразование системы непрерывных случайных величин

Предположим, что система случайных величин (X; Y) в результате функционального преобразования преобразуется в систему случайных величин (V; Z):

$$\begin{cases} V = f_1(X;Y); \\ Z = f_2(X;Y). \end{cases} \tag{4.14}$$

Будем считать, что функции f_1 и f_2, входящие в систему уравнений (4.14), известны и удовлетворяют условиям монотонности, непрерывности и дифференцируемости по обеим переменным x и y. Также будем считать известным аналитическое выражение для совместной плотности вероятности $W(x; y)$.

По аналогии с формулами (4.2) и (4.4) запишем соотношения для определения вероятности попадания случайных величин X и Y в интервалы Δx, Δy и вероятности попадания случайных величин V и Z в соответствующие интервалы Δv, Δz:

$$P(x \le X < x+\Delta x;\ y \le Y < y+\Delta y) = \int\limits_{x}^{x+\Delta x}\int\limits_{y}^{y+\Delta y} W(x;\ y)\,dxdy \approx W(x;\ y)\Delta x \Delta y;$$

$$P(v \le V < v+\Delta v;\ z \le Z < z+\Delta z) = \int\limits_{v}^{v+\Delta v}\int\limits_{z}^{z+\Delta z} W(v;\ z)\,dv\,dz \approx W(v;\ z)\Delta v \Delta z.$$

В этом случае интервалы Δv и Δz определяются разностью:

$$\Delta v = f_1(x+\Delta x;\ y+\Delta y) - f_1(x;\ y);$$
$$\Delta z = f_2(x+\Delta x;\ y+\Delta y) - f_2(x;\ y).$$

Если случайные величины X и Y принимают значения, принадлежащие интервалам $X \subset \Delta x$ и $Y \subset \Delta y$, то случайные величины V и Z будут иметь значения принадлежащие интервалам $V \subset \Delta v$, $Z \subset \Delta z$, а вероятности этих событий будут одинаковы:

$$W(v;\ z)\,\Delta v\ \ \Delta z\ = W(x;\ y)\ \ \Delta x\ \ \Delta y\ ;$$

$$W(v;\ z) \approx W(x;\ y)\frac{\Delta x \Delta y}{\Delta v \Delta z}. \qquad (4.15)$$

Для того чтобы равенство (4.15) было не приближенным, а точным, необходимо потребовать, чтобы интервалы Δv и Δz, Δx и Δy были бесконечно малыми:

$$\lim_{\substack{\Delta v \to 0\\ \Delta z \to 0}} \frac{\Delta x \Delta y}{\Delta v \Delta z} = \frac{\partial g_1(v;\ z)\partial g_2(v;\ z)}{\partial v \partial z}, \qquad (4.16)$$

где $x = g_1(v;\ z)$ и $y = g_2(v;\ z)$ — функции, обратные функциям (4.14), которые определяются из решения системы уравнений (4.14) относительно x и y.

Формула (4.16) представляет собой так называемый якобиан преобразования переменных $\Delta(v;\ z)$, который вычисляется как определитель матрицы частных производных

$$\Delta(v;\ z) = \frac{\partial g_1(v;\ z)\partial g_2(v;\ z)}{\partial v \partial z} = \begin{vmatrix} \dfrac{\partial g_1(v;\ z)}{\partial v} & \dfrac{\partial g_1(v;\ z)}{\partial z} \\ \dfrac{\partial g_2(v;\ z)}{\partial v} & \dfrac{\partial g_2(v;\ z)}{\partial z} \end{vmatrix}. \qquad (4.17)$$

После подстановки в аналитическое выражение для $W(x;\ y)$ вместо переменных x и y обратных функций $x = g_1(v;\ z)$ и $y = g_2(v;\ z)$ с учетом якобиана преобразования переменных (4.16) и (4.17) формула (4.15) может быть записана в виде:

$$W(v,z) = W_{xy}\left[g_1(v;\,z);g_2(v;\,z)\right]\left|\Delta(v;\,z)\right|. \tag{4.18}$$

Полученная формула (4.18) позволяет решать частную задачу, которая часто встречается на практике и заключается в определении закона распределения случайной величины z, являющейся функцией двух случайных величин X и Y:

$$Z = f(X;\ Y).$$

Чтобы для решения данной задачи воспользоваться формулой (4.18), введем вспомогательную случайную величину $V = X$. Тогда систему функциональных преобразований случайных величин можно записать в виде:

$$\begin{cases} V = X; \\ Z = f(X; Y). \end{cases}$$

Из решения данной системы уравнений запишем обратные функции:

$$\begin{cases} X = V; \\ Y = g(V; Z). \end{cases} \tag{4.19}$$

При условии дифференцируемости функций $y = g(v;\, z)$ на основании формулы (4.17) запишем якобиан преобразования переменных:

$$\Delta(v;\, z) = \begin{vmatrix} 1 & 0 \\ \dfrac{\partial g(v;\, z)}{\partial v} & \dfrac{\partial g(v;\, z)}{\partial z} \end{vmatrix} = \frac{\partial g(v;\, z)}{\partial z}. \tag{4.20}$$

Для плотности вероятности случайных величин v и z с учетом обратных функций (4.19) и якобиана преобразования переменных (4.20) формула (4.18) преобразуется к виду:

$$W(v;\, z) = W\left[v;\, g(v;\, z)\right]\left|\frac{\partial g(v;\, z)}{\partial z}\right|. \tag{4.21}$$

В постановке задачи нам необходимо определить плотность вероятности случайной величины Z. Случайная величина V и соответственно переменная v введена как вспомогательная и в полученном решении (4.21) является лишней. Искомую одномерную плотность вероятности $W(z)$ можно определить, интегрируя правую часть равенства (4.21) по лишней переменной:

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left[v;\, g(v;\, z)\right]\left|\frac{\partial g(v;\, z)}{\partial z}\right| dv. \tag{4.22}$$

В качестве примера рассмотрим применение формулы (4.22) на простейших функциональных преобразованиях случайных величин.

1. Плотность вероятности суммы двух непрерывных случайных величин $Z = X + Y$. Введем вспомогательную случайную величину $v = X$ и запишем систему функциональных преобразований:

$$\begin{cases} v = X; \\ Z = X + Y. \end{cases}$$

Обратные функциональные преобразования определяются формулами:

$$\begin{cases} X = v; \\ Y = Z - v. \end{cases}$$

Якобиан преобразования переменных будет равен

$$\Delta(v;\, z) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1.$$

Отсюда для плотности вероятности $W(z)$ получим:

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left[v;\,(z-v)\right]dv. \qquad (4.23)$$

Если бы в качестве вспомогательной переменной ввели $v = Y$, то для плотности $W(z)$ получили бы формулу

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left[(z-v);\, v\right]dv,$$

которая полностью равносильна формуле (4.23).

2. Точно так же можно показать, что плотность вероятности разности двух случайных величин $Z = X - Y$ будет равна

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left[v;\,(v-z)\right]dv. \qquad (4.24)$$

3. Для определения плотности вероятности произведения двух случайных величин $Z = XY$ введем вспомогательную переменную $v = X$. Системы прямых и обратных функциональных преобразований запишутся в виде:

$$\begin{cases} v = x; \\ z = x \cdot y; \end{cases} \Leftrightarrow \begin{cases} x = v; \\ y = z / v. \end{cases}$$

Якобиан преобразования переменных будет равен

$$\Delta(v;\, z) = \begin{vmatrix} 1 & 0 \\ -\dfrac{z}{v^2} & \dfrac{1}{v} \end{vmatrix} = \frac{1}{v}.$$

Плотность вероятности произведения двух случайных величин будет определяться формулой

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left(v;\, z/v\right)\left|\frac{1}{v}\right| dv\,. \qquad (4.25)$$

4. Для частного двух случайных величин $Z = X/Y$ при вспомогательной переменной $v = X$ системы прямых и обратных функциональных преобразований запишутся в виде:

$$\begin{cases} v = x; \\ z = x\,/\,y; \end{cases} \Leftrightarrow \begin{cases} x = v; \\ y = v/z. \end{cases}$$

Якобиан преобразования переменных будет равен

$$\Delta(v;\, z) = \begin{vmatrix} 1 & 0 \\ \dfrac{1}{z} & -\dfrac{v}{z^2} \end{vmatrix} = -\frac{v}{z^2}.$$

Тогда для плотности вероятности частного двух случайных величин получим формулу

$$W(z) = \int_{-\infty}^{\infty} W_{xy}\left(v;\, \frac{v}{z}\right)\left|\frac{v}{z^2}\right| dv. \qquad (4.26)$$

Если в качестве вспомогательной переменной в рассматриваемых примерах принимать $v = y$, то мы получим формулы, которые после интегрирования по лишней переменной v дадут тот же результат, что и формулы (4.23) — (4.26).

Пример 4.3. Рассмотрим систему независимых случайных величин X и Y, каждая из которых имеет нормальный закон распределения с нулевыми математическими ожиданиями $m_x = m_y = 0$ и равными дисперсиями $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Требуется найти закон распределения случайной величины $Z = \sqrt{X^2 + Y^2}$, характеризующей величину случайного отклонения z от начала координат (рис. 4.2).

Решение. В соответствии с условием и формулой (3.14) плотности вероятности случайных величин X, Y будут равны:

$$W(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}};$$

$$W(y)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}}.$$

Так как случайные величины X и Y независимы, совместная плотность вероятности $W(x;\ y)$ будет равна произведению одномерных плотностей вероятности:

$$W(x;\ y)=\frac{1}{\sqrt{2\pi}\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \tag{4.27}$$

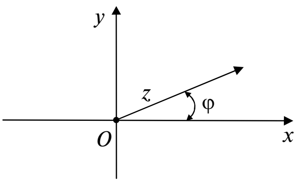


Рис. 4.2. Преобразование системы случайных величин X и Y в полярную систему координат

Введем дополнительную переменную φ (рис. 4.2), которая представляет собой угол между вектором z и осью Ox. Тогда систему обратных функциональных преобразований можно записать в виде:

$$\begin{cases} x = z\cos\varphi; \\ y = z\sin\varphi. \end{cases}$$

Эти уравнения характеризуют преобразование случайных величин X и Y, представленных в декартовой системе координат, в случайные величины z и φ, представленные в полярной системе координат. Определим область возможных значений случайных величин z и φ. Из рисунка 4.2 видно, что при изменении случайных величин X и Y в пределах $-\infty < x < \infty; -\infty < y < \infty$ область возможных значений случайных величин z и φ будет определяться неравенством:

$$z \ge 0;$$
$$-\pi \le \varphi < \pi .$$

Определим якобиан преобразования переменных

$$\Delta(z;\ \varphi)=\begin{vmatrix}\cos\varphi & -z\sin\varphi \\ \sin\varphi & z\cos\varphi\end{vmatrix}=z\cos^2\varphi+z\sin^2\varphi=z .$$

Для двумерной плотности вероятности случайных величин z и φ получим:

$$W(z,\varphi)=\frac{z}{2\pi\sigma^2}\exp\left[-\frac{z^2\cos^2\varphi+z^2\sin^2\varphi}{2\sigma^2}\right]=\frac{z}{2\pi\sigma^2}e^{-\frac{z^2}{2\sigma^2}}.$$

Интегрируя полученное выражение по лишней переменной φ в пределах области ее возможных значений, получим искомую плотность вероятности $W(z)$:

$$W(z)=\int_{-\pi}^{\pi}\frac{z}{2\pi\sigma^2}e^{-\frac{z}{2\sigma^2}}d\varphi=\frac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}.$$

Из полученного выражения видно, что случайное отклонение независимых нормально распределенных случайных величин от начала координат $z=\sqrt{x^2+y^2}$ при $m_x = m_y = 0$ и $\sigma_x^2=\sigma_y^2=\sigma^2$ имеет распределение Рэлея (3.31), а случайная фаза φ имеет равномерное распределение в интервале

$$\pi \le \varphi < \pi .$$

4.3. Математическое ожидание и дисперсия функции случайных величин

Начнем с решения задачи определения математического ожидания случайной величины $z = f(x; y)$, являющейся функцией случайных величин X и Y:

$$m_z = M(Z) = \langle z \rangle. \tag{4.28}$$

При этом будем считать, что совместная плотность вероятности случайных величин $W(x; y)$ нам известна.

Для нахождения m_z нужно усреднить случайную величину Z. В соответствии с ранее приведенным определением для усреднения непрерывной случайной величины z ее нужно умножить на плотность вероятности $W(z)$ и полученное произведение проинтегрировать по области возможных значений z. Для осуществления этой операции нужно предварительно определить плотность вероятности $W(z)$. Возможно определение математического ожидания m_z без вычисления $W(z)$. Если случайная величина Z зависит от одной из случайных величин, например X:

$$z = f(x).$$

Подставляя это функциональное преобразование в формулу (4.28), получим

$$m_z = M[f(x)] = \langle f(x) \rangle. \qquad (4.29)$$

В выражении (4.29) под знаком усреднения (M[·], < >) стоит функция $f(x)$, в которой случайной является только величина X. В соответствии с определением операции усреднения нужно эту функцию $f(x)$ умножить на плотность вероятности $W(x)$ и полученное произведение проинтегрировать по области возможных значений X:

$$m_z = \int_{-\infty}^{\infty} f(x) W(x)\,dx. \qquad (4.30)$$

Одномерная плотность вероятности $W(x)$ легко находится интегрированием двумерной плотности вероятности $W(x; y)$ по лишней переменной (3.55).

Если X является дискретной случайной величиной, то по аналогии с формулой (2.9) для m_z можно записать:

$$m_z = \sum_{i=1}^{n} f(x_i) p_i. \qquad (4.31)$$

Если случайная величина z зависит от двух случайных величин

$$z = f(x; y),$$

то, подставляя это функциональное преобразование в формулу (4.28), получим

$$m_z = M[f(x; y)] = \langle f(x; y) \rangle. \qquad (4.32)$$

В формуле (4.32) под знаком усреднения стоит функция $f(x; y)$, которая зависит от двух случайных величин X и Y, значит, усреднять эту функцию нужно по обеим случайным величинам. Для этого функция $f(x; y)$ умножается на двумерную плотность вероятности $W(x; y)$ и полученное произведение интегрируется по обеим переменным, получим

$$m_z = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x; y) W(x; y)\,dx\,dy. \qquad (4.33)$$

Если X и Y являются дискретными случайными величинами, то операция усреднения по x_i и y_j записывается в виде:

$$m_z = \sum_{i=1}^{n} \sum_{j=1}^{m} f(x_i\, y_j)\, p_{ij}, \tag{4.34}$$

где p_{ij} — совместная вероятность появления значений x_i и y_j дискретных случайных величин X и Y.

Приведем два свойства математического ожидания случайных величин, не рассматриваемые в подразделе 2.4.1.

Свойство 1. Математическое ожидание линейного преобразования случайных величин

$$z = \sum_{i=1}^{n} a_i X_i + b$$

равно тому же линейному преобразованию математических ожиданий.

В данном случае обозначение X_i означает, что мы имеем n случайных величин $X_1, X_2, ..., X_n$.

Доказательство. На основании формулы (4.33) запишем:

$$m_z = \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n\ раз} \left[\sum_{i=1}^{n} a_i x_i + b\right] W_n(x_1; x_2; ...; x_n)\, dx_1\, dx_2 ... dx_n\ ,$$

где $W_n(x_1; x_2 ... x_n)$ — n-мерная плотность вероятности случайных величин X_i; $i = 1 \div n$.

Так как интеграл от суммы равен сумме интегралов, данное выражение (4.35) можно записать в виде:

$$m_z = b \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n\ раз} W_n(x_1; x_2; ...; x_n)\, dx_1\, dx_2 ... dx_n +$$

$$+ a_1 \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n\ раз} x_1 W_n(x_1; x_2; ...; x_n)\, dx_1\, dx_2 ... dx_n +$$

$$+ a_2 \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n\ раз} x_2 W_n(x_1; x_2; ...; x_n)\, dx_1\, dx_2 ... dx_n +$$

$$+ a_n \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n\ раз} x_n W_n(x_1; x_2; ...; x_n)\, dx_1\, dx_2 ... dx_n.$$

Интеграл в первом слагаемом является условием нормировки n-мерной плотности вероятности и равен единице. Во всех остальных слагаемых интегралы равны математическим ожиданиям m_{x1}, m_{x2}, ..., m_{xn}. На основании этого можно записать:

$$m_z = \sum_{i=1}^{n} a_i m_{xi} + b.$$

Свойство 2. Математическое ожидание произведения зависимых случайных величин X и Y равно сумме произведений их математических ожиданий и корреляционного момента:

$$M(XY) = m_x m_y + k_{xy}. \qquad (4.35)$$

Доказательство. По определению корреляционного момента (3.67)

$$k_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_x)(y - m_y) W(x, y)\, dx\, dy.$$

Перемножим выражения в скобках в подынтегральном выражении и выполним почленное интегрирование слагаемых:

$$k_{xy} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyW(x;\, y)\, dx\, dy + m_y \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xW(x;\, y)\, dx\, dy -$$

$$-m_x \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yW(x;\, y)\, dx\, dy = m_x m_y \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(x;\, y)\, dx\, dy.$$

Первое слагаемое в данном выражении равно M (XY). Интегралы во втором и третьем слагаемом соответственно равны m_x и m_y. Интеграл в четвертом слагаемом является условием нормировки двумерной плотности вероятности и равен единице. Таким образом, для корреляционного момента получим

$$k_{xy} = M(XY) - m_x m_y,$$

или

$$M(XY) = m_x m_y + k_{xy}. \qquad (4.35)$$

Из формулы (4.35) следует, что если случайные величины X и Y являются независимыми, то их корреляционный момент равен нулю и математическое ожидание произведения этих случайных величин равно произведению их математических ожиданий.

Аналогичные соотношения можно получить для дисперсии функции случайных величин. По определению, дисперсия есть среднее значение квадрата отклонения случайной величины от математического ожидания (формулы (3.11), (3.12):

$$D_z = M[(z - m_z)^2] = M(z^2) - m_z^2. \tag{4.36}$$

Если Z является функцией одной случайной величины $z = f(x)$, то для дисперсии D_z получим:

$$D_z = M\left[f^2(x)\right] - m_z^2 = \int_{-\infty}^{\infty} \left[f(x)\right]^2 W(x)\,dx - m_z^2. \tag{4.37}$$

Математическое ожидание m_z определяется формулой (4.30).

Если z является функцией двух случайных величин $z = f(x; y)$, то дисперсия D_z будет определяться по формуле

$$D_z = M\left[f^2(x, y)\right] - m_z^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^2(x, y)\, W(x; y)\,dx\,dy - m_z^2. \tag{4.38}$$

В этом случае математическое ожидание m_z определяется по формуле (4.33).

Для дискретных случайных величин формулы (4.37) и (4.38) преобразуются к виду:

$$\begin{aligned} D_z &= \sum_{i=1}^{n} \left[f(x_i)\right]^2 p_i - m_z^2 \quad при\ z_i = f(x_i); \\ D_z &= \sum_{i=1}^{n}\sum_{j=1}^{m} \left[f(x_i; y_j)\right]^2 p_{ij} - m_z^2 \quad при\ z_i = f(x_i y_i). \end{aligned} \tag{4.39}$$

В дополнение к свойствам дисперсии, приведенным в подразделе 2.4.2, приведем еще три.

Свойство 1. Дисперсия суммы (разности) двух случайных величин равна сумме дисперсий этих величин плюс (минус) удвоенное значение их корреляционного момента:

$$D(X \pm Y) = D_x + D_y \pm 2k_{xy}. \tag{4.40}$$

Доказательство. Если случайная величина z равна сумме случайных величин X и Y, то в соответствии с формулой (4.38) можно записать

$$D_z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x \pm y)^2 W(x, y)\,dx\,dy - m_z^2.$$

После возведения $(x \pm y)$ в квадрат и почленного интегрирования слагаемых получим

$$D_z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 W(x;\, y)\,dx\,dy \pm 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyW(x;\, y)\,dx\,dy + \\ + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 W(x;\, y)\,dx\,dy - m_z^2. \qquad (4.41)$$

Из формулы (3.12) следует, что первое и третье слагаемые в данном выражении будут соответственно равны $D_x + m_x^2$ и $D_y + m_y^2$. Второе слагаемое в соответствии с формулой (4.35) будет равно $k_{xy} + m_x m_y$. После подстановки этих значений в формулу (4.41) получим

$$D_z = D_x + D_y \pm 2k_{xy} + m_x^2 \pm 2m_x \cdot m_y + m_y^2 - m_z^2. \qquad (4.42)$$

Математическое ожидание суммы (разности) двух случайных величин будет равно

$$m_z = m_x \pm m_y.$$

После возведения m_z в квадрат и подстановки в формулу (4.42) мы получим формулу (4.40). Из формулы (4.40) следует, что дисперсия суммы некоррелированных случайных величин равна сумме их дисперсий:

$$D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i). \qquad (4.43)$$

На основании свойства (4.40) можно сформулировать и доказать одно важное положение. Для любых случайных величин X и Y их корреляционный момент по абсолютной величине не больше, чем произведение среднеквадратичных отклонений этих случайных величин:

$$\left|k_{xy}\right| \le \sigma_x \sigma_y\,. \qquad (4.43а)$$

Для доказательства вычислим дисперсию линейного преобразования случайных величин X и Y:

$$Z = aX + bY.$$

На основании формулы (4.40) и свойств дисперсии можно записать:

$$D_z = a^2 D_x + b^2 D_y + 2abk_{xy}.$$

Будем считать, что $a = \sigma_y$, $b = \pm\sigma_x$. В этом случае для D_z получим:

$$D_z = \sigma_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2 \pm 2\sigma_y\sigma_x k_{xy} = 2\sigma_y\sigma_x(\sigma_y\sigma_x \pm k_{xy}).$$

Так как $D_z \geq 0$, $\sigma_x \geq 0$ и $\sigma_y \geq 0$, то должно выполняться неравенство

$$\sigma_y \sigma_x \pm k_{xy} \geq 0,$$

откуда $-\sigma_x \sigma_y \leq k_{xy} \leq \sigma_x \sigma_y$, или $\left| k_{xy} \right| \leq \sigma_x \sigma_y$. Из этого также следует, что $\left| \rho_{xy} \right| \leq 1$.

Свойство 2. Дисперсия произведения двух независимых случайных величин определяется сотношением

$$D(XY) = D_x D_y + m_x^2 D_y + m_y^2 D_x. \tag{4.44}$$

Доказательство. Рассмотрим случайную величину $Z = XY$. В силу независимости случайных величин X и Y будут независимы и их квадраты. Определим вначале начальный момент второго порядка случайной величины Z:

$$M(Z^2) = M[X^2 Y^2] = M(X^2)M(Y^2). \tag{4.45}$$

Для начальных моментов второго порядка справедливы соотношения (3.13):

$$M(Z^2) = D_z + m_z^2;$$

$$M(X^2) = D_x + m_x^2$$

$$M(Y^2) = D_y + m_y^2.$$

Подставляя данные соотношения в выражение (4.45), получим:

$$D_z + m_z^2 = D_x D_y + m_x^2 D_y + m_y^2 D_x + m_x^2 m_y^2.$$

Так как для независимых случайных величин $m_z = m_x m_y$, можно видеть, что последнее полученное соотношение совпадает с формулой (4.44).

Свойство 3. Дисперсия суммы линейных преобразований некоррелированных случайных величин X_i, $i = 1 \div n$:

$$Z = \sum_{i=1}^{n} a_i X_i + b$$

определяется формулой

$$Dz = D(\sum_{i=1}^{n} a_i X_i + b) = \sum_{i=1}^{n} a_i^2 D_{x_i}. \tag{4.46}$$

Данное свойство доказывается аналогично свойству 1 для математического ожидания.

Пример 4.4. Производится n независимых измерений некоторой физической величины. Результат каждого измерения есть случайная величина X_i ($i = 1 \div n$). Дисперсии всех результатов измерения одинаковы и равны σ^2. Найти дисперсию среднего арифметического $X_{\text{ср}}$ результатов n измерений:

$$X_{\text{ср}} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Так как случайные величины X_i, $i = 1 \div n$ независимы, то по формуле (4.46) получим

$$D(X_{\text{ср}}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D_{x_i} = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}. \qquad (4.47)$$

Из формулы (4.47) следует вывод: для того чтобы повысить точность измерения некоторой величины, нужно провести n независимых измерений этой величины и за результат измерения принять значение $X_{\text{ср}}$. В этом случае ошибка измерения будет равна $\pm\frac{\sigma}{\sqrt{n}}$.

Пример 4.5. В результате финансово-экономической деятельности отраслевых предприятий по выпуску некоторого вида продукции получены данные о средних значениях и величинах среднеквадратичных отклонений по следующим показателям:

1) k — количество выпускаемой продукции данного вида $m_k = 2000$, $\sigma_k = 1000$;

2) $з_1$ – переменные затраты на единицу выпускаемой продукции $m_{з_1} = 1120$ руб., $\sigma_{з_1} = 330$ руб.;

3) $з_2$ — постоянные затраты по выпуску данного вида продукции $m_{з_1} = 3$ млн руб.; $\sigma_{з_2} = 270$ тыс. руб.

Считая случайные величины k, $з_1$ и $з_2$ независимыми, определить среднее значение и среднеквадратичное отклонение себестоимости данного вида продукции по предприятиям отрасли.

Решение. Для решения задачи воспользуемся формулой (4а), устанавливающей функциональную взаимосвязь между случайными величинами, приведенными в условии задачи, и себестоимостью продукции.

В соответствии с формулами (4а) и (4.28) для математического ожидания себестоимости продукции можно записать:

$$m_с = M(с) = M(kз_1+з_2) = M(kз_1) + M(з_2).$$

Так как случайные величины k и $з_1$ независимы, окончательно получим:

$$m_с = m_k\, m_{з_1} + m_{з_2} = 20\,000 \cdot 1120 + 3000 \cdot 10^3 = 25{,}4 \text{ млн руб.}$$

Для определения среднеквадратичного отклонения себестоимости продукции определим начальный момент второго порядка этой случайной величины:

$$M(с^2) = M[(k\,з_1 + з_2)^2] = M\left(k^2\, з_1^2\right) + 2M\left(k\, з_1\, з_2\right) + M\left(з_2^2\right). \quad (4.47а)$$

Так как случайные величины k, $з_1$ и $з_2$ независимы, то $M(k^2\, з_1^2) = M(k^2) \cdot M(з_1^2)$; $M(k\, з_1\, з_2) = M(k)M(з_1)M(з_2) = m_k m_{з_1} m_{з_2}$.

С учетом этих соотношений, а также равенств

$$M(с^2) = D_с + m_с^2;$$

$$M(k^2) = D_k + m_k^2;$$

$$M(з_1^2) = D_{з_1} + m_{з_1};$$

$$M(з_2^2) = D_{з_2} + m_{з_2}$$

формулу (4.47а) можно записать в виде

$$D_с + m_с^2 = D_k D_{з_1} + m_k^2 D_{з_1} + m_{з_1}^2 D_k +$$
$$+ D_{з_2} + m_k^2 m_{з_1}^2 + 2m_k m_{з_1} m_{з_2} + m_{з_2}^2.$$

С учетом полученного в данном примере соотношения для $m_с$ окончательно для дисперсии $D_с$ получим:

$$D_с = D_k D_{з_1} + m_k^2 D_{з_1} + m_{з_1}^2 D_k + D_{з_2} = (1000)^2 \cdot (330)^2 + (20\,000)^2 \cdot (330)^2 +$$
$$+ (1120)^2 \cdot (1000)^2 + (270\,000)^2 = 4499{,}62 \cdot 10^{10}\,(\text{руб.})^2.$$

Среднеквадратичное отклонение себестоимости продукции будет равно

$$\sigma_с = \sqrt{D_с} \approx \sqrt{4499{,}62 \cdot 10^{10}} = 67 \cdot 10^5 \text{ руб.} = 6{,}7 \text{ млн руб.}$$

Если считать закон распределения себестоимости выпуска продукции $W(c)$ нормальным, то в соответствии с примером 3.2 можно сказать, что с вероятностью 0,68 себестоимость выпускаемой продукции на предприятиях отрасли будет находиться в пределах

$$m_c - \sigma_c \le c \le m_c + \sigma_c,$$

или

$$18{,}7 \text{ млн руб.} \le c \le 32{,}1 \text{ млн руб.}$$

4.4. Корреляционный момент случайных величин после их функционального преобразования

Рассмотрим две случайные величины Z_1 и Z_2, являющиеся функциями случайных величин X и Y:

$$\begin{aligned} Z_1 &= f_1(X;Y); \\ Z_2 &= f_2(X;Y). \end{aligned} \tag{4.48}$$

В соответствии с формулой (4.35) для корреляционного момента $k_{z_1 z_2}$ можно записать:

$$k_{z_1 z_2} = M(Z_1 Z_2) - m_{z_1} m_{z_2}. \tag{4.49}$$

С учетом функциональных преобразований (4.48) для $M(Z_1Z_2)$ получим:

$$M(Z_1 Z_2) = M\left[f_1(X;Y) f_2(X;Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x;y) f_2(x;y) W(x;y)\,dxdy - m_{z_1} m_{z_2}.$$

Значения математических ожиданий m_{z_1} и m_{z_2} определяются формулой (4.33).

Дальнейшее рассмотрение проведем для двух частных случаев.

1. Будем считать, что функциональные преобразования (4.48) имеют вид:

$$\begin{gathered} Z_1 = a_1 X + b_1; \\ Z_2 = a_2 Y + b_2, \end{gathered}$$

т.е. Z_1 и Z_2 являются линейными преобразованиями случайных величин X и Y.

Тогда для $M(Z_1Z_2)$ получим:

$$\begin{gathered} M(Z_1 Z_2) = M[(a_1 X + b_1)(a_2 Y + b_2)] = \\ = a_1 a_2 M(XY) + b_2 a_1 M(X) + b_1 a_2 M(Y) + b_1 b_2. \end{gathered}$$

В данном выражении $M(XY) = k_{xy} + m_x m_y$; $M(Y) = m_y$; $M(X) = m_x$, поэтому

$$M(Z_1Z_2) = a_1a_2k_{xy} + a_1a_2m_xm_y + b_1a_2m_y + a_1b_2m_x + b_1b_2. \qquad (4.50)$$

Математические ожидания m_{z_1} и m_{z_2} будут равны:

$$m_{z_1} = a_1m_x + b_1;$$

$$m_{z_2} = a_2my + b_2.$$

Перемножая m_{z_1} и m_{z_2} и подставляя результат перемножения и формулу (4.50) в уравнение (4.49), для корреляционного момента $k_{z_1,\, z_2}$ получим:

$$k_{z_1,\, z_2} = a_1a_2k_{xy}.$$

2. Будем считать, что функциональные преобразования (4.48) имеют вид:

$$Z_1 = a_1X + b_1Y;$$
$$Z_2 = a_2X + b_2Y,$$

т.е. Z_1 и Z_2 являются взвешенными суммами случайных величин X и Y.

После подстановки этих равенств для $M(Z_1Z_2)$ получим:

$$M(Z_1Z_2) = M[(a_1X + b_1Y)(a_2X + b_2Y)] =$$
$$= a_1a_2M(X^2) + b_1a_2M(XY) + a_1b_2M(XY) + b_1b_2M(Y^2).$$

В данном выражении $M(XY) = k_{xy} + m_xm_y$, $M(X^2) = Dx + m_x^2$;

$$M(Y^2) = D_y + m_y^2,$$

поэтому

$$M\left(Z_1Z_2\right) = a_1a_2D_x + b_1b_2D_y + k_{xy}\left(a_1b_2 + a_2b_1\right) + a_1a_2m_x^2 + \\ + m_xm_y\left(a_1b_2 + a_2b_1\right) + b_1b_2m_y^2. \qquad (4.51)$$

Математические ожидания m_{z_1} и m_{z_2} будут равны $m_{z_1} = a_1m_x + b_1m_y$ и $m_{z_2} = a_2m_x + b_2m_y$, и для произведения $m_{z_1}\, m_{z_2}$ получим:

$$m_{z_1}\, m_{z_2} = a_1a_2\, m_x^2 + m_xm_y\,(a_1b_2 + a_2b_1) + b_1b_2\, m_y^2\,. \qquad (4.52)$$

После подстановки равенств (4.51) и (4.52) в формулу (4.49) для корреляционного момента $k_{z_1 z_2}$ получим:

$$k_{z_1 z_2} = k_{xy}(a_1b_2 + a_2b_1) + a_1a_2D_x + b_1b_2D_y.$$

Эта формула является более общим случаем формулы (4.40).

4.5. Примеры решения задач

Задача 4.1. Выручка фирмы X в тыс. дол. США имеет нормальный закон распределения, заданный плотностью вероятности:

$$W(x)=\frac{1}{\sqrt{2\pi}\cdot 1{,}2}e^{-\frac{(x-15)^2}{2{,}88}}. \qquad (4.53)$$

Найти закон распределения выручки фирмы в тысячах рублей $Z = XY$ в пересчете по курсу доллара Y, если курс доллара есть постоянная величина $Y = 30$ руб./дол. и выручка фирмы в долларах X не зависит от курса доллара. Найти математическое ожидание и среднее квадратичное отклонение выручки фирмы в тыс. руб.

Решение. Сравнивая формулу (4.53) с плотностью вероятности нормально распределенной случайной величины (3.14), определим математическое ожидание, дисперсию и среднее квадратичное отклонение выручки фирмы в тыс. дол.:

$$m_x = 15 \text{ тыс. дол.; } \sigma_x = 1{,}2 \text{ тыс. дол.;}$$

$$D_x = \sigma_x^2 = 1{,}44 \text{ (тыс. дол.)}^2.$$

По свойству 1 математического ожидания (см. подраздел 4.3)) математическое ожидание выручки фирмы в тыс. рублей будет равно

$$M_z = Ym_x = 30 \cdot 15 = 450 \text{ тыс. руб.}$$

По свойству 3 дисперсии (см. подраздел 4.3 формула (4.46)) дисперсия выручки фирмы Z будет равна

$$D_z = Y^2 D_x = 30^2 \cdot 1{,}44 = 1296 \text{ (тыс. руб.)}^2.$$

Среднее квадратичное отклонение выручки фирмы в тысячах рублей будет равно

$$\sigma_z = \sqrt{D_z} = \sqrt{Y^2\sigma_x^2} = Y\sigma_x = 30 \cdot 1{,}2 = 36 \text{ тыс. руб.}$$

Плотность вероятности выручки фирмы в тысячах рублей может быть определена по формуле (4.13), где $a = Y$; $b = 0$:

$$W(z) = \frac{1}{Y}W_x\left[\frac{z}{Y}\right] = \frac{1}{\sqrt{2\pi}\cdot 1{,}2\cdot 30}\exp\left[-\frac{\left(\frac{z}{30}-15\right)^2}{2{,}88}\right] =$$

$$= \frac{1}{\sqrt{2\pi}\cdot 36}\exp\left[-\frac{(z-450)^2}{2\cdot 1296}\right].$$

Данную плотность вероятности можно было записать исходя из того, что при линейном преобразовании $z = YX$ нормальный закон распределения случайной величины z сохраняется, а значения его параметров m_z, D_z и σ_z были определены выше.

Задача 4.2. Себестоимость выпускаемой продукции С определяется суммой постоянных X и переменных Y затрат на производство продукции $C = X + Y$. Определить математическое ожидание, дисперсию, среднее квадратичное отклонение и плотность вероятности оптовой цены $Z = (1 + N_p)C = (1 + N_p)(X + Y)$ выпускаемой продукции при норме рентабельности $N_p = 0{,}2$, если плотность вероятности постоянных и переменных затрат определяется формулами:

$$W(x) = \frac{1}{\sqrt{2\pi}\cdot 0{,}2}e^{-\frac{(x-2)}{0{,}08}};\;\; W(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-5)}{2}}.$$

Решение. Из приведенных выражений для плотностей вероятностей постоянных X и переменных Y затрат следует, что они имеют нормальный закон распределения с математическими ожиданиями $m_x = 2$ и $m_y = 5$ (денежных единиц), средними квадратичными отклонениями $\sigma_x = 0{,}2$ и $\sigma_y = 1$ (денежных единиц) и дисперсиями, равными $D_x = 0{,}04$ и $D_y = 1$.

Оптовая цена выпускаемой продукции определяется линейным преобразованием постоянных и переменных затрат $Z = (1 + N_p)(X + Y)$, которые имеют нормальный закон распределения. Из этого делаем вывод о том, что оптовая цена Z выпускаемой продукции будет иметь нормальный закон распределения:

$$W(z) = \frac{1}{\sqrt{2\pi}\sigma_z}\exp\left[-\frac{(z-m_z)^2}{2\sigma_z^2}\right].$$

Математическое ожидание оптовой цены m_z находим по свойству 1 (подраздел 4.3) математического ожидания:

$$m_z = (1 + N_p)(m_x + m_y) = 1{,}2(2 + 5) = 8{,}4.$$

Дисперсию оптовой цены определяем по формуле (4.46):

$$D_z = (1 + N_p)^2(D_x + D_y) = 1{,}2^2(0{,}2 + 1) = 1{,}728.$$

Среднее квадратичное отклонение оптовой цены будет равно

$$\sigma_z = \sqrt{D_z} = \sqrt{1{,}728} \approx 1{,}314.$$

С учетом вычисленных значений m_z и D_z для плотности вероятности оптовой цены выпускаемой продукции получим

$$W(z) = \frac{1}{\sqrt{2\pi \cdot 1{,}728}} \exp\left[-\frac{(2-8{,}4)^2}{2 \cdot 1{,}728}\right].$$

Задача 4.3. Для исходных данных, приведенных в задаче 4.1, найти математическое ожидание и среднее квадратичное отклонение выручки фирмы в тыс. руб. при условии, что курс доллара Y является случайной величиной, плотность вероятности которой определяется формулой

$$W(y) = \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{(y-30)^2}{8}}.$$

Решение. Из приведенной формулы для плотности вероятности курса доллара видно, что он имеет нормальный закон распределения с математическим ожиданием m_y = 30 руб./дол., средним квадратичным отклонением σ_y = 2 руб./дол. Так как выручка фирмы по условию задачи определяется в тысячах долларов и тысячах рублей, приведем числовые характеристики распределения курса доллара к этим же денежным единицам: m_y = 30 тыс. руб./тыс. дол.; σ_y = 2 тыс. руб./тыс. дол.; D_y = 4 (тыс. руб./тыс. дол.)^2. По условию задачи, если выручка фирмы в долларах и курс доллара независимые величины, то математическое ожидание и дисперсию выручки фирмы в тысячах рублей можно определить по формуле (4.35) при $k_{xy} = 0$ и по свойству 2 (подраздел 4.3) дисперсии произведения двух независимых величин (формула (4.44)).

Для математического ожидания выручки фирмы в тыс. руб. получим:

$$m_z = m_y m_x = 30 \cdot 15 = 450 \text{ тыс. руб.}$$

Для дисперсии выручки по формуле (4.44) получим:

$$D_z = D_x D_y + m_x^2 D_y + m_y^2 D_x =$$
$$= 1{,}44 \cdot 4 + (15)^2 \cdot 4 + (30)^2 \cdot 1{,}44 = 2201{,}76 \text{ (тыс. руб.)}^2.$$

Среднее квадратичное отклонение выручки фирмы в тысячах рублей будет равно:

$$\sigma_z = \sqrt{D_z} = \sqrt{2201{,}76} \approx 46{,}923 \text{ тыс. руб.}$$

Из сравнения полученных значений m_z, D_z и σ_z со значениями, полученными в задаче 4.1, можно сделать вывод, что случайный характер курса доллара при $m_y = 30$ не влияет на значение математического ожидания m_z, но приводит к существенному увеличению дисперсии D_z и среднего квадратичного отклонения σ_z.

Задача 4.4. Сумму денежных средств на расчетных счетах клиентов банка — юридических лиц — можно рассматривать как случайную величину X, распределенную по нормальному закону распределения с математическим ожиданием $m_x = 200$ млн руб. и средним квадратичным отклонением $\sigma_x = 20$ млн руб. Заявки юридических лиц на текущие платежи можно также считать случайной величиной Y, имеющей нормальный закон распределения с математическим ожиданием $m_y = 20$ млн руб. и средним квадратичным отклонением $\sigma_y = 4$ млн руб.

Определить плотность вероятности случайной величины $Z = X - Y$ — остатков средств на расчетных счетах после удовлетворения заявок юридических лиц на текущие платежи.

Определить, какую сумму $Z_{\text{р}}$ из остатков средств на расчетных счетах юридических лиц банк может использовать в качестве дополнительного ресурса средств для проведения активных операций, чтобы риск неудовлетворения заявок юридических лиц на текущие платежи не превышал 0,15.

Решение. Так как остатки на расчетных счетах Z определяются как разность нормально распределенных случайных величин X и Y, то случайная величина Z также будет иметь нормальный закон распределения:

с математическим ожиданием

$$m_z = m_x - m_y = 200 - 20 = 180 \text{ млн руб.},$$

дисперсией

$$D_z = D_x + D_y = \sigma_x^2 + \sigma_y^2 = 400 + 16 = 416 \text{ (млн руб.)}^2$$

и средним квадратичным отклонением

$$\sigma_z = \sqrt{D_z} = \sqrt{416} \approx 20{,}4 \text{ млн руб.}$$

С учетом найденных значений числовых характеристик распределения случайной величины z плотность вероятности остатков средств на счетах юридических лиц будет определяться выражением

$$W(z) = \frac{1}{\sqrt{2\pi \cdot 416}} \exp\left[-\frac{(z-180)^2}{2 \cdot 416}\right].$$

Риск неудовлетворения заявок на текущие платежи возникает при выполнении неравенства $Z - Zp < 0$.

Вероятность выполнения этого неравенства равна вероятности риска неудовлетворения заявок на текущие платежи:

$$P(z - z_{\text{р}} < 0) = P(z < z_{\text{р}}) = F(z_{\text{р}}) < 0{,}15,$$

где $F(z_{\text{р}})$ — функция распределения нормально распределенной случайной величины.

В соответствии с формулой (3.16) это неравенство можно записать в виде:

$$F(z_{\text{р}}) = \Phi\left(\frac{z_{\text{р}} - m_z}{\sigma_z}\right) = \Phi(u) < 0{,}15,$$

где $u = \dfrac{z_{\text{р}} - m_z}{\sigma_z}$; $\Phi(u) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt$ — интеграл вероятности.

В таблице интеграла вероятности, приведенной в приложении 1, таких значений нет. Из этого следует, что $z_{\text{р}} < m_z$ и переменная $u < 0$, т.е. отрицательная. С учетом свойства интеграла вероятности

$$\Phi(-u) = \Phi\left(\frac{m_z - z_{\text{р}}}{\sigma_z}\right) = 1 - \Phi(u),$$

последнее неравенство можно записать в виде:

$$\Phi\left(\frac{m_z - z_{\text{р}}}{\sigma_z}\right) > 0{,}85.$$

Из таблицы интеграла вероятности находим, что данное неравенство выполняется при:

$$\frac{m_z - z_{\text{р}}}{\sigma_z} > 1{,}04;$$

$$z_{\text{р}} < m_z - 1{,}04 \cdot \sigma_z;$$

$$z_{\text{р}} < 180 - 1{,}04 \cdot 20{,}4;$$

$$z_{\text{р}} < 158{,}78 \text{ млн руб.}$$

Таким образом, если банк из остатков средств на расчетных счетах юридических лиц в качестве дополнительного ресурса средств для проведения активных операций будет использовать не более 158,7 млн руб., то вероятность риска неудовлетворения заявок по текущим платежам не превысит 0,15.

Задача 4.5. По условию задачи 3.5 определить плотность вероятности прибыли предприятия $Z = X - Y$. Определить вероятность того, что предприятие в результате работы получит убытки ($z < 0$), что прибыль предприятия будет больше 2,5 денежных единиц ($z > 2{,}5$).

Решение. Прибыль предприятия z определяется разностью нормально распределенных случайных величин выручки от реализации X и затрат Y на производство продукции, поэтому случайная величина Z также будет иметь нормальный закон распределения. Для определения плотности вероятности прибыли определим ее математическое ожидание прибыли с учетом решения задачи 3.5, получим

$$m_z = m_x - m_y = 5 - 3 = 2 \text{ ден. ед.}$$

Для дисперсии прибыли с учетом зависимости случайных величин и результатов решения задачи 3.5 по формуле (4.40) получим

$$D_z = D_x + D_y - 2k_{xy} = D_x + D_y - 2\sqrt{D_x D_y}\rho_{xy} =$$
$$= 0{,}9 + 0{,}5 - 2\sqrt{0{,}9 \cdot 0{,}5} \cdot 0{,}8 \approx 0{,}8633.$$

Среднее квадратичное отклонение прибыли будет равно

$$\sigma_z = \sqrt{D_z} = \sqrt{0{,}8633} \approx 0{,}929.$$

С учетом вычисленных числовых значений распределения прибыли для плотности вероятности получим

$$W(z) = \frac{1}{\sqrt{2\pi \cdot 0{,}8633}} \exp\left[-\frac{(z-2)^2}{2 \cdot 0{,}8633}\right].$$

Вероятность того, что предприятие в результате работы получит убытки $P(z < 0)$, определяется функцией распределения при $z = 0$:

$$P(z \le 0) = F(z = 0) = \Phi\left(\frac{-2}{0{,}929}\right) = 1 - \Phi(2{,}15) = 1 - 0{,}98422 = 0{,}01578.$$

Вероятность того, что прибыль предприятия составит больше 2,5 денежных единиц ($z > 2{,}5$), будет равна

$$P(z > 2{,}5) = \int_{2{,}5}^{\infty} W(x)dx = 1 - \int_{-\infty}^{2{,}5} W(x)dx = 1 - \Phi\left(\frac{2{,}5 - m_z}{\sigma_z}\right) = 1 - \Phi\left(\frac{2{,}5 - 2}{0{,}929}\right) =$$
$$= 1 - \Phi(0{,}54) = 1 - 0{,}7054 = 0{,}2946.$$

Вопросы для самоконтроля

1. Каким образом определить плотность вероятности случайной величины после ее функционального преобразования?
2. Каким образом определить плотность вероятности случайной величины после линейного функционального преобразования?
3. Как найти плотность вероятности системы двух случайных величин после их функционального преобразования?
4. Как найти плотность вероятности случайной величины, являющейся функцией двух случайных величин?
5. Как определить плотность вероятности суммы, разности двух случайных величин?
6. Каким образом вычислить плотность вероятности произведения, частного двух случайных величин?
7. Как вычислить математическое ожидание функции дискретных и непрерывных случайных величин?
8. Как найти математическое ожидание линейного преобразования случайных величин?
9. Как высчитывается математическое ожидание произведения зависимых случайных величин?
10. Как рассчитать дисперсию функции дискретных и непрерывных случайных величин?
11. Чему равна дисперсия суммы, разности двух случайных величин?
12. Докажите, что корреляционный момент двух случайных величин по абсолютной величине не больше, чем произведение их среднеквадратичных отклонений?
13. Чему равна дисперсия суммы линейных преобразований некоррелированных случайных величин?
14. Как найти корреляционный момент случайных величин после их функционального преобразования?
15. Чему равен корреляционный момент случайных величин после их линейных преобразований?
16. Как находится коэффициент корреляции случайных величин, являющихся взвешенными суммами двух случайных величин?

ГЛАВА 5

ОСНОВЫ ТЕОРИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

5.1. Основные понятия и классификация случайных процессов

В предыдущих главах предметом изучения являлось вероятностное описание случайных событий и случайных величин.

Случайная величина характерна тем, что в результате испытания всегда принимает только одно, заранее не известное, но единственное значение. Это имеет место, когда изучаемое нами явление предстает в «застывшем виде», в фиксированных постоянных условиях эксперимента. В реальной жизни условия протекания случайного явления не остаются постоянными, а изменяются во времени, поэтому случайные величины как количественное проявление тех или иных характеристик случайного явления также изменяются во времени. Случайный процесс характеризует изменение случайной величины во времени. При этом для случайного процесса не представляется возможным описать изменение случайной величины во времени какими-либо детерминированными функциями.

Основным отличием случайных процессов является то, что его функциональная зависимость от времени имеет не детерминированный, а случайный характер.

Случайным процессом называют процесс изменения случайной величины во времени, когда значения процесса, взятые в фиксированные моменты времени, являются случайными величинами.

Реализацией случайного процесса $X(t)$ называется некоторая случайная функция $X(t)$, полученная в результате испытания в заданных условиях. Серия таких испытаний одинаковой длительности дает совокупность, или пучок, реализаций (рис. 5.1), которые дают представление о случайном процессе $X(t)$.

Примером реализации случайного процесса $X(t)$ может служить изменение во времени прироста численности населения по одному из регионов Российской Федерации.

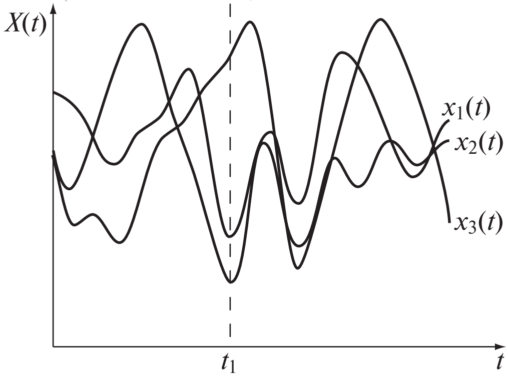


Рис. 5.1. Пучок реализаций случайного процесса $X(t)$

Если рассматривать изменение во времени прироста численности населения по нескольким регионам, мы получим пучок реализаций процесса $X(t)$. Если пучок реализаций пересечь прямой $t = t_1$, то получим совокупность значений случайного процесса $X(t)$: $x_1(t_1)$, $x_2(t_1)$, $x_3(t_1)$, $x_4(t_1)$, ..., $x_n(t_1)$, которые являются значениями случайной величины $X(t_1)$. То есть в каждый момент времени t_i значение случайного процесса $X(t_i)$ является случайной величиной.

Случайные процессы классифицируют по двум признакам: по характеру изменения процесса во времени и по области возможных значений $X(t)$.

Если аргумент t (время) процесса $X(t)$ является непрерывной переменной, то такой процесс называют ***случайным процессом.***

Если же время t дискретно, то процесс $X(t)$ называют ***случайной последовательностью.*** В реальной жизни мы имеем дело со случайными последовательностями, когда нас интересуют значения случайного процесса в некоторые характерные фиксированные моменты времени. Например, данные бухгалтерского баланса (или другой финансовой отчетности) на конец каждого месяца или квартала представляют собой реализации случайных последовательностей.

По области возможных значений процесс делят точно так же, как случайные величины, на дискретные и непрерывные.

В соответствии с названными двумя признаками случайные процессы можно разделить на четыре основных класса (рис. 5.2):

1. ***Непрерывный случайный процесс*** — это процесс, непрерывный по времени и непрерывный по области возможных значений (рис. 5.2, *а*).

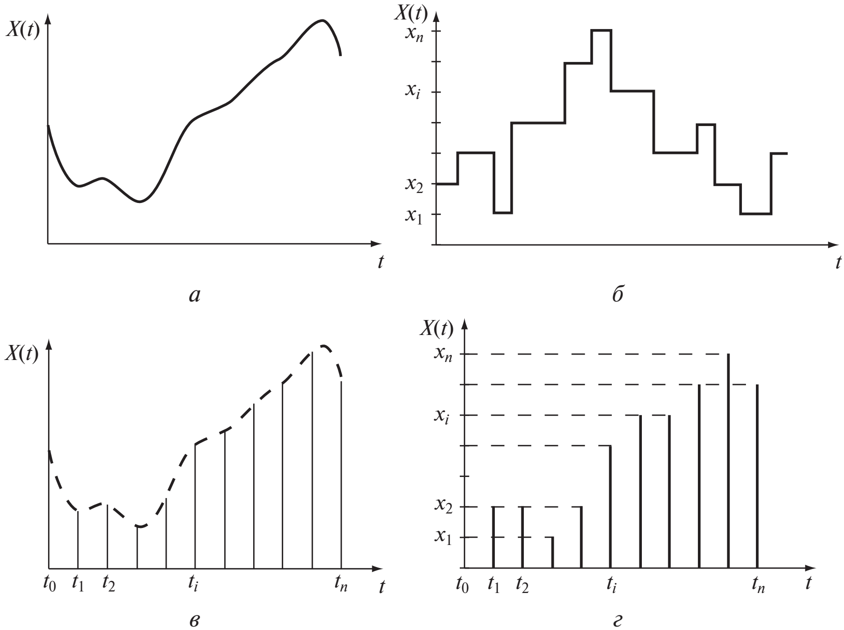


Рис. 5.2. Реализации случайных процессов различных классов

2. ***Дискретный случайный процесс*** — это процесс, непрерывный по времени и дискретный по области возможных значений (рис. 5.2, *б*).

3. ***Непрерывная случайная последовательность*** — это процесс, дискретный по времени и непрерывный по области возможных значений (рис. 5.2, *в*).

4. ***Дискретная случайная последовательность*** — это процесс, дискретный по времени и дискретный по области возможных значений (рис. 5.2, *г*).

В данной главе далее мы будем рассматривать вопросы вероятностного описания непрерывных случайных процессов.

5.2. Вероятностное описание непрерывных случайных процессов

Предположим, что мы имеем пучок реализаций случайного процесса $X(t)$ (см. рис. 5.1) $x_1(t)$, $x_2(t)$, ..., $x_i(t)$, ..., $x_n(t)$ и их число n достаточно велико. Значения случайного процесса в произвольный момент времени t_1 будут значениями случайной величины $X(t_1)$. Как будет показано в главе 6, по этим значениям $x_1(t_1)$, $x_2(t_1)$, ..., $x_i(t_1)$, ..., $x_n(t_1)$ можно найти плотность вероятности случайной величины $X(t_1)$. Обозначим ее $W(x_1/t_1)$. Символ x_1 в аргументе плотности вероятности говорит о том, что это случайное значение процесса $X(t)$ в момент времени t_1.

Найденная плотность вероятности $W(x_1/t_1)$ дает вероятностное описание процесса $X(t)$ только в момент времени $t = t_1$ и не дает вероятностных характеристик процесса в другие моменты времени.

Для вероятностного описания случайного процесса $X(t)$ на некотором интервале наблюдения $0 < t \le T$ выделим на этом интервале n отсчетов времени t_1, t_2, ..., t_i, ..., t_n, отстоящих друг от друга на величину $\tau = t_{j+1} - t_j$ (рис. 5.3).

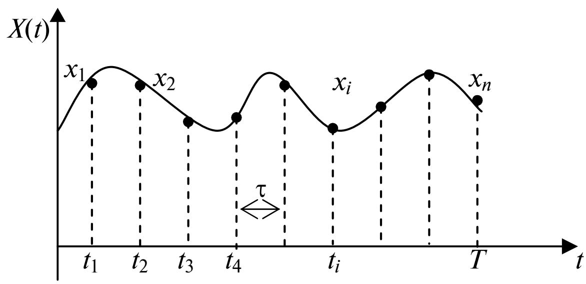


Рис. 5.3. Временные отсчеты случайного процесса $X(t)$

Значения случайного процесса в эти фиксированные моменты времени будут случайными величинами $x_1 = X(t_1)$, $x_2 = X(t_2)$, $x_i = X(t_i)$, $x_n = X(t_n)$. Вероятностное описание системы из n случайных величин дается n-мерной плотностью вероятности

$$W_n(x_1, x_2, ..., x_i, ..., x_{n/}\, t_1; \tau). \tag{5.1}$$

Данная плотность вероятности характеризует случайный процесс в n моментов времени.

Чем больше число отсчетов времени n и соответственно мерность совместной плотности вероятности (формула (5.1)), тем более исчерпывающее вероятностное описание случайного процесса $X(t)$ мы будем иметь на интервале наблюдения $0 < t \le T$. Во-первых, получить такую n-мерную плотность вероятности практически невозможно. Во-вторых, даже если мы получим аналитическое выражение для формулы (5.1), использовать его для решения практических задач также нереально. Поэтому для описания случайных процессов необходимо использовать более простые аналитические модели плотности вероятности. Наиболее простой аналитической моделью плотности вероятности случайного процесса, которая содержит информацию о характере изменения случайного процесса во времени, является двумерная плотность вероятности

$$W(x_1, x_{2/}t_{1;}t_2) = W(x_1, x_{2/}t_1;\tau), \tag{5.2}$$

где $\tau = t_2 - t_1$.

Плотность вероятности (5.2) дает вероятностное описание процесса в два момента времени t_1 и $t_2 = t_1 + \tau$:

$$x_1 = X(t_1);\ x_2 = X(t_2).$$

Плотность вероятности (5.2) является функцией аргументов x_1 и x_2, характеризующих возможные значения случайного процесса, а также функцией аргументов t_1 и t_2 или t_1 и τ. При изменении значений τ можно видеть, как изменяется плотность вероятности (5.2) при удалении или приближении друг к другу по времени отсчетов случайного процесса $X(t_1)$ и $X(t_2)$. При изменении значений аргумента t_1 можно видеть, как изменяется плотность вероятности (5.2) при перемещении отсчетов по времени внутри интервала наблюдения $0 < t \leq T$.

Если случайный процесс $X(t)$ является нормальным случайным процессом, то двумерная плотность вероятности (5.2) определяется аналитическим выражением

$$W(x_1; x_2 / t_1; \tau) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2(t_1;\tau)}} \times$$
$$\times \exp\left\{-\frac{1}{2[1-\rho^2(t_1;\tau)]}\left[\frac{(x_1-m_1)^2}{\sigma_1^2} - \right.\right.$$
$$\left.\left. - \frac{2\rho(t_1;\tau)(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right]\right\}, \tag{5.3}$$

где $m_1 = m_x(t_1)$; $m_2 = m_x(t_2)$ — математические ожидания случайного процесса $X(t)$ в моменты времени t_1 и t_2; $\sigma_1 = \sigma_x(t_1); \sigma_2 = \sigma_x(t_2)$ — среднеквадратичные отклонения случайного процесса в моменты времени t_1 и t_2; $\rho(t_1;\tau) = \rho(t_1;t_2)$ — коэффициент корреляции между значениями случайного процесса $X(t_1)$ и $X(t_2)$, которые являются случайными величинами.

Знание двумерной плотности вероятности (см. подраздел 3.6) позволяет определить плотности вероятности составляющих системы случайных величин $x_1 = X(t_1)$ и $x_2 = X(t_2)$, являющихся временными отсчетами случайного процесса.

Одномерные плотности вероятности значений случайного процесса в моменты времени t_1 и t_2:

$$W(x_1/t_1)=\int_{-\infty}^{\infty}W(x_1;x_2/t_1;t_2)dx_2;$$
$$W(x_2/t_2)=\int_{-\infty}^{\infty}W(x_1;x_2/t_1;t_2)dx_1. \quad (5.4)$$

По аналогии с формулами (3.74) и (3.75) можно показать, что для нормального случайного процесса, двумерная плотность вероятности которого определяется формулой (5.3), одномерные плотности вероятности (5.4) будут равны:

$$W(x_1/t_1)=\frac{1}{\sqrt{2\pi}\sigma_x(t_1)}\exp\left\{-\frac{[x_1-m_x(t_1)]^2}{2\sigma_x^2(t_1)}\right\};$$
$$W(x_2/t_2)=\frac{1}{\sqrt{2\pi}\sigma_x(t_2)}\exp\left\{-\frac{[x_2-m_x(t_2)]^2}{2\sigma_x^2(t_2)}\right\}. \quad (5.5)$$

Для вероятностного описания случайных процессов, как и для случайных величин, используются моменты распределения, из которых наиболее часто используются математическое ожидание, дисперсия и корреляционный момент.

5.3. Моменты распределения случайных процессов

Рассмотрим случайный процесс $X(t)$, пучок возможных реализаций которого приведен на рис. 5.4. Для произвольно взятого момента времени t_i значения реализаций случайного процесса в данный момент времени являются значениями $x_1(t_i)$, $x_2(t_i)$, ..., $x_n(t_i)$ случайной величины $X(t_i)$. Эта случайная величина $X(t_i)$ описывается одномерной плотностью вероятности $W(x_i/t_i)$ и имеет математическое ожидание и дисперсию, определяющиеся формулами:

$$m_x(t_i)=\int_{-\infty}^{\infty}xW(x/t_i)dx; \quad (5.6)$$

$$D_x(t_i)=\int_{-\infty}^{\infty}x^2W(x/t_i)dx-m_x^2(t_i). \quad (5.7)$$

В другой момент времени t_j случайная величина $X(t_j)$ может иметь другое, отличное от $m_x(t_i)$, значение математического ожидания $m_x(t_j)$ и другое значение дисперсии $D_x(t_j)$, которые определяются формулами (5.6) и (5.7) по плотности вероятности $W(x/t_j)$.

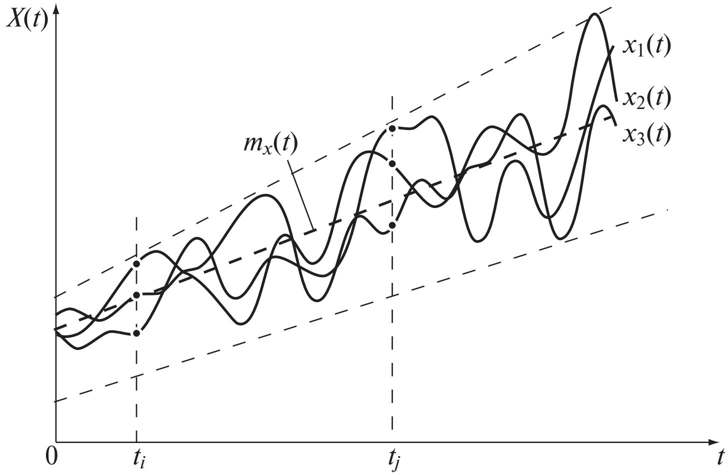


Рис. 5.4. Пучок реализаций случайного процесса $X(t)$

Таким образом, математическое ожидание случайного процесса $X(t)$ есть некоторая функция $m_x(t)$, которая при каждом значении аргумента t равна математическому ожиданию случайной величины, получающейся в сечении случайного процесса в данный момент времени t. По своему смыслу это некоторая средняя функция, вокруг которой с теми или иными случайными отклонениями группируются возможные реализации $x_i(t)$ $(i = 1 \div n)$ случайного процесса $X(t)$.

Аналогично дисперсия $D_x(t)$ или среднеквадратичное отклонение $\sigma_x(t) = \sqrt{D_x(t)}$ — это некоторая функция, характеризующая, как изменяется среднее значение квадрата отклонения случайной величины $X(t)$ от математического ожидания $m_x(t)$ при изменении аргумента t.

Математическое ожидание и дисперсия случайного процесса обладают теми же свойствами, присущими данным числовым характеристикам случайных величин.

Из реализаций $x_i(t)$ случайного процесса $X(t)$, приведенных на рис. 5.4, следует, что математическое ожидание $m_x(t)$ и дисперсия $D_x(t) = M[(X(t) - m_x(t))^2]$ данного случайного процесса увеличиваются с увеличением времени t.

Эти числовые характеристики дают некоторое представление о характере изменения случайного процесса во времени, но возможны ситуации, когда при одинаковых зависимостях $m(t)$ и $D(t)$ случайные процессы во времени ведут себя по-разному. На рисунке 5.5 приведен пучок возможных реализаций случайного процесса $Y(t)$, для которого

$$m_y(t) \approx m_x(t);\ D_y(t) \approx D_x(t),$$

но характер изменения во времени процесса $Y(t)$ отличается от процесса $X(t)$ (см. рис. 5.4).

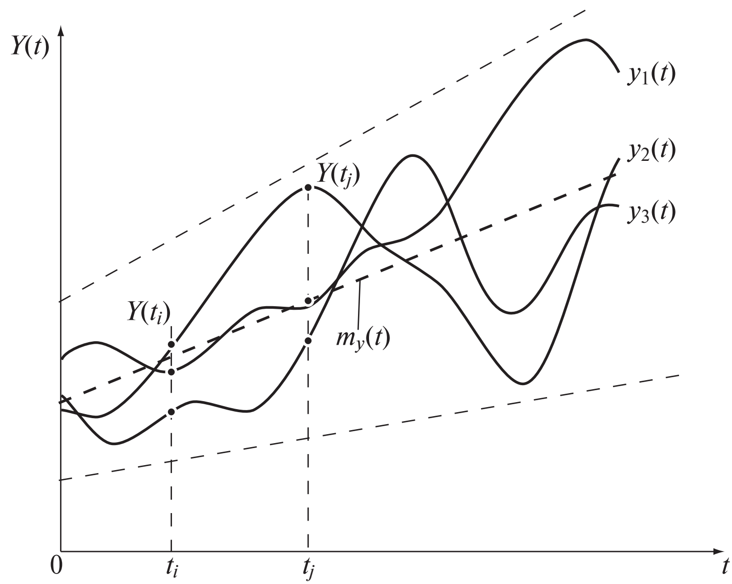


Рис. 5.5. Пучок реализаций случайного процесса $Y(t)$

Из приведенных на рис. 5.5 реализаций случайного процесса $Y(t)$ видно, что если нам известно значение $Y(t_i)$ случайного процесса в момент времени t_i, то через некоторый промежуток времени в точке t_j возможно предсказать значение случайного процесса $Y(t_j)$.

Для процесса $X(t)$ (см. рис. 5.4) такое предсказание сделать весьма проблематично. Для процесса $Y(t)$ мы можем предсказать значение $Y(t_j)$ по $Y(t_i)$, потому что эти случайные величины достаточно сильно связаны между собой, имеют сильную корреляционную связь. Для случайных значений $X(t_i)$ $X(t_j)$ процесса $X(t)$ корреляционная связь между этими случайными значениями практически равна нулю, и случайные значения $X(t_i)$, $X(t_j)$ практически независимы, поэтому предсказать значение $X(t_j)$ по известному значению $X(t_i)$ невозможно.

Точно так же, как для случайных величин — формула (3.68), корреляционную связь между двумя значениями случайного процесса $X(t)$ и $X(t+\tau)$, отстоящими друг от друга на интервал времени τ, можно определить по формуле

$$k_x(t;\tau)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}[x-m_x(t)][x_\tau-m(t+\tau)]W(x;x_\tau/t;\tau)\,dx\,dx_\tau, \quad (5.8)$$

где $x=X(t)$; $x_\tau=X(t+\tau)$.

Корреляционный момент $k_x(t;\tau)$ определяемый для случайного процесса по формуле (5.8), называют корреляционной функцией.

Корреляционной функцией случайного процесса $X(t)$ называется неслучайная функция двух аргументов $k_x(t_1;t_2)$, которая при фиксированных значениях моментов времени t_1 и $t_2=t_1+\tau$ равна корреляционному моменту случайных величин $X(t_1)$ и $X(t_2)$.

$$\begin{aligned}k_x(t_1;t_2)&=M\left\{\left[X(t_1)-m_x(t_1)\right]\left[X(t_2)-m_x(t_2)\right]\right\}=\\&=M\left[X(t_1)X(t_2)\right]-m_x(t_1)m_x(t_2).\end{aligned} \quad (5.9)$$

Укажем свойства корреляционной функции случайных процессов.

Свойство 1. $k_x(t_1;t_2)=k_x(t_2;t_1)$ — это так называемое свойство симметрии функции корреляции. Оно вытекает из определения функции корреляции (5.9). Это свойство может быть также записано в виде:

$$k_x(t;t+\tau)=k_x(t;t-\tau). \quad (5.10)$$

Свойство 2. Корреляционная функция случайного процесса в совпадающие моменты времени $t=t_1=t_2$ при $\tau=0$ равна дисперсии случайного процесса:

$$\lim_{\tau\to 0}k(t;t+\tau)=D_x(t). \quad (5.11)$$

Доказательство. При $t_1=t_2=t$ получим $X(t_1)=X(t_2)=X(t)$, $m_x(t_1)=$ $=m_x(t_2)=m_x(t)$. Подставляя эти значения в выражение (5.9), получим

$$\lim_{\tau\to 0}k_x(t;t+\tau)=M\left\{\left[X(t)-m_x(t)\right]^2\right\}=D_x(t).$$

Свойство 3. При τ, не равном нулю, функция корреляции случайного процесса по абсолютной величине не больше, чем произведение среднеквадратичных отклонений:

$$\left|k_x(t;t+\tau)\right|\le\sigma_x(t)\sigma_x(t+\tau). \quad (5.12)$$

Данное свойство доказывается аналогично доказательству формулы (4.43а) для случайных величин.

Свойство 4. Если $Y(t) = X(t) + \varphi(t)$, где $\varphi(t)$ — неслучайная (детерминированная) функция, то

$$k_y(t_1; t_2) = k_x(t_1; t_2). \tag{5.13}$$

Доказательство. В соответствии со свойствами математического ожидания

$$m_y(t) = m_x(t) + \varphi(t),$$

отсюда следует равенство

$$Y(t) - m_y(t) = X(t) - m_x(t),$$

а значит, и равенство (5.13).

Свойство 5. Если $Y(t) = \varphi(t)X(t)$, где $\varphi(t)$ — детерминированная, неслучайная функция, то

$$k_y(t_1; t_2) = \varphi(t_1)\,\varphi(t_2)k_x(t_1; t_2). \tag{5.14}$$

Доказательство. Из свойств математического ожидания следует

$$m_y(t) = \varphi(t)m_x(t).$$

Подставляя соотношения для $Y(t)$ и $m_y(t)$ в формулу (5.9), получим

$$k_y(t_1; t_2) = M\{[\varphi(t_1)X(t_1) - \varphi(t_1)m_x(t_1)]\cdot[\varphi(t_2)X(t_2) - \varphi(t_2)m_x(t_2)]\} =$$
$$= \varphi(t_1)\varphi(t_2)M\{[X(t_1) - m_x(t_1)][X(t_2) - m_x(t_2)]\} = \varphi(t_1)\varphi(t_2)k_x(t_1; t_2).$$

Свойство 6. При $\tau \to \infty$ функция корреляции случайного процесса стремится к нулю:

$$\lim_{\tau\to\infty} k_x(t; t+\tau) = 0. \tag{5.15}$$

Доказательство. При $\tau \to \infty$ мы рассматриваем два значения случайного процесса, удаленные друг от друга на бесконечно большой интервал времени τ. За это время случайный процесс $X(t)$ может многократно изменяться в сторону уменьшения и увеличения относительно своего значения $X(t_1)$ в момент времени t_1 (см. рис. 5.4). На основании этого можно считать значения случайного процесса $X(t_1)$ и $X(t_1+\tau)$ при $\tau \to \infty$ независимыми друг от друга случайными величинами. Для независимых случайных величин двумерная плотность вероятности равна произведению одномерных:

$$W(x_1; x_2/t_1; t_2) = W(x_1/t_1)\,W(x_2/t_2),$$
$$\text{при } \tau = t_2 - t_1 \to \infty.$$

Подставляя данное равенство в формулу (5.8), получим

$$k_x(t_1 t_2) =$$
$$= \int_{-\infty}^{\infty} [x_1 - m_x(t_1)] W(x_1 / t_1) dx_1 \cdot \int_{-\infty}^{\infty} [x_2 - m_x(t_2)] W(x_2 / t_2) dx_2.$$

В данном выражении каждый из интегралов представляет собой математическое ожидание центрированной случайной величины, которое равно нулю. Из этого следует, что при $\tau \to \infty$ выполняется условие (5.15).

В соответствии со свойствами функции корреляции можно привести возможные графики изменения функции корреляции случайных процессов (рис. 5.6).

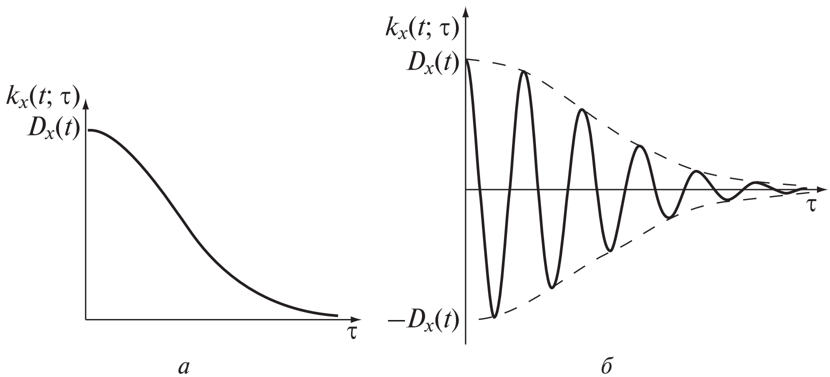


Рис. 5.6. Графики изменения функции корреляции случайных процессов

Из свойств 2 и 3 следует, что максимальное значение функции корреляции случайного процесса равно его дисперсии $D_x(t)$. Для исключения зависимости от дисперсии случайного процесса на практике часто используют нормированную функцию корреляции

$$\rho_x(t_1; t_2) = \frac{k_x(t_1; t_2)}{\sigma_x(t_1)\sigma_x(t_2)}. \qquad (5.16)$$

Нормированная функция корреляции случайного процесса $\rho_x(t_1; t_2)$ по смыслу аналогична коэффициенту корреляции случайных величин. Она характеризует степень линейной взаимосвязи между двумя значениями случайного процесса, отстоящими друг от друга на интервал времени $\tau = t_2 - t_1$.

5.4. Стационарные случайные процессы

Стационарными называют случайные процессы, для которых многомерная плотность вероятности зависит только от взаимного расположения моментов времени t_1, t_2, ..., t_n, но не от самих значений этих величин, определяющихся точкой начала отсчетов t_1. То есть для стационарных случайных процессов n-мерная плотность вероятности должна удовлетворять условию

$$W_n(x_1; x_2; ...; x_n/t_1; t_2; ...; t_n) = \\ = W_n(x_1; x_2; ... x_n/t_1 - t_0; t_2 - t_0; ...; t_n - t_0). \quad (5.17)$$

Равенство (5.17) нужно понимать следующим образом. Если мы определим n-мерную плотность вероятности $Wn(x_1; x_2; ...; x_n/t_1; t_2; ...; t_n)$ на интервале наблюдения $0 < t < T$ (см. рис. 5.3), то при сдвиге этого интервала влево или вправо по оси времени на произвольную величину t_0 n-мерная плотность вероятности не должна меняться.

Условие (5.17) является очень жестким условием, поэтому данное определение является определением стационарности случайных процессов в узком смысле.

В подразделе 5.2 было показано, что на практике для вероятностного описания случайных процессов используют двумерные плотности вероятности $W(x_1; x_2/t_1; t_2)$. При $t_0 = t_1$ в соответствии с формулой (5.17) для двумерной плотности вероятности получим

$$W(x_1; x_2/t_1; t_2) = W(x_1; x_2/0; t_2 - t_1) = W(x_1; x_2/\tau). \quad (5.18)$$

Из формулы (5.18) следует, что для стационарных случайных процессов двумерная плотность вероятности не зависит от начального момента времени t_1, а зависит только от интервала $\tau = t_2 - t_1$.

Для одномерной плотности вероятности $W(x/t_1)$ при $t_0 = t_1$ в соответствии с формулой (5.17) получим

$$W(x/t_1) = W(x/t_1 - t_1) = W(x). \quad (5.19)$$

То есть для стационарного случайного процесса одномерная плотность вероятности имеет одно и то же аналитическое выражение $W(x)$ для любого момента времени t (не зависит от времени t).

В соответствии с формулой (5.19) для математического ожидания и дисперсии стационарного случайного процесса получим:

$$m_x(t)=\int\limits_{-\infty}^{\infty} xW(x)dx = m_x = \text{const};$$
$$D_x(t)=\int\limits_{-\infty}^{\infty} (x-m_x)^2 W(x)dx = D_x = \text{const}. \qquad (5.20)$$

Из формулы (5.18) для функции корреляции стационарного случайного процесса получим

$$k_x(t;\, t+\tau)=\int\limits_{-\infty}^{\infty} (x-m_x)(x_\tau - m_x)W(x;\, x_\tau / \tau)\, dxdx_\tau \Rightarrow$$
$$k_x(t;\, t+\tau)=k_x(\tau). \qquad (5.21)$$

Условия (5.20) и (5.21) являются менее жесткими по сравнению с условием (5.17). Этим условиям может удовлетворять более широкий перечень случайных процессов, поэтому стационарными в широком смысле называют случайные процессы, математическое ожидание и дисперсия которых являются постоянными величинами (5.20) (не зависят от времени), а корреляционная функция зависит только от временного сдвига τ и не зависит от выбора начального момента времени t (5.21).

Для стационарных случайных процессов свойства 1 и 3 функции корреляции, рассмотренные в подразделе 5.3, формулируются в несколько ином виде.

Свойство 1. Корреляционная функция случайного процесса является четной функцией:

$$k_x(-\tau)=k_x(\tau)\,. \qquad (5.22)$$

Свойство 2. Для любого $\tau \ge 0$ абсолютная величина корреляционной функции не превышает дисперсии случайного процесса:

$$|k_x(\tau)| \le D_x. \qquad (5.23)$$

Нормированная корреляционная функция (5.16) для стационарных случайных процессов с учетом формулы (5.23) будет определяться формулой

$$\rho_x(\tau)=\frac{k_x(\tau)}{D_x};$$
$$-1 \le \rho_x(\tau) \le 1. \qquad (5.24)$$

Как уже отмечалось в подразделе 5.3, нормированная корреляционная функция не зависит от дисперсии случайного процесса и характеризует только степень линейной взаимосвязи между двумя значениями случайного процесса, отстоящими друг от друга на интервал времени τ.

Для случайных процессов, функция корреляции которых имеет вид, показанный на рис. 5.6, *а*, часто используют еще одну числовую характеристику, называемую временем корреляции τ_k. Время корреляции случайного процесса определяется по формуле

$$\tau_k = \int_0^{\infty} \rho(\tau)\,d\tau = \frac{1}{D_x}\int_0^{\infty} k_x(\tau)\,d\tau. \qquad (5.25)$$

Формулу (5.25) можно переписать в виде:

$$D_x \tau_k = \int_0^{\infty} k_x(\tau)\,d\tau.$$

Определенный интеграл в правой части данного равенства равен площади под кривой функции корреляции. Произведение $D_x\tau_k$ в левой части равенства равно площади прямоугольника (пунктирная линия на рис. 5.7). Таким образом, в соответствии с формулой (5.25) время корреляции случайного процесса τ_k равно основанию прямоугольника высотой D_x, площадь которого равна площади под кривой функции корреляции.

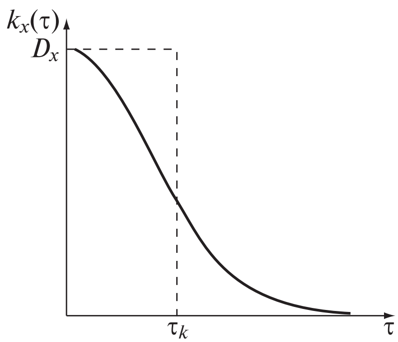


Рис. 5.7. Определение времени корреляции случайного процесса

По своему смыслу время корреляции τ_k есть некоторое граничное значение. При $\tau < \tau_k$ нормированная функция корреляции имеет значения ближе к единице, и два значения случайного процесса x и x_τ, отстоящие друг от друга на интервал τ, имеют сильную взаимосвязь, т.е. сильно коррелированны. При $\tau > \tau_k$ нормированная функция корреляции имеет значения, близкие к нулю, поэтому значения случайного процесса x и x_τ будут слабокоррелированны.

5.5. Эргодическое свойство стационарных случайных процессов

Значительная часть стационарных случайных процессов обладает одним замечательным свойством, существенно облегчающим на практике определение основных числовых характеристик случайных процессов. Это свойство заключается в том, что для эргодических случайных процессов их числовые характеристики могут определяться по одной достаточно длинной во времени реализации.

Математическое ожидание, дисперсия и функция корреляции стационарных случайных процессов определяются формулами (5.20) и (5.21) путем усреднения по плотности вероятности.

В подразделе 5.2 отмечалось, что плотность вероятности случайных процессов может быть определена по ансамблю (пучку) реализаций случайного процесса, поэтому иногда говорят, что в формулах (5.20) и (5.21) усреднение осуществляется по ансамблю реализаций $x_i(t)$ ($i = 1 \div n$).

При наличии одной достаточно длинной реализации случайного процесса математическое ожидание, дисперсия и функция корреляции при усреднении по времени могут быть вычислены по формулам:

$$\begin{gathered} m_x = \bar{X} = \frac{1}{T}\int_0^T x(t)\,dt; \\ D_x = \overline{(X - m_x)^2} = \frac{1}{T}\int_0^T \left[x(t) - m_x\right]^2 dt; \\ k_x(\tau) = \\ = \overline{\left[X(t) - m_x\right]\left[X(t+\tau) - m_x\right]} = \frac{1}{T}\int_0^T \left[x(t) - m_x\right]\left[x(t+\tau) - m_x\right] dt, \end{gathered} \tag{5.26}$$

где $x(t)$ — это любая из возможных реализаций случайного процесса $X(t)$.

В формулах (5.26) усреднение по времени обозначено чертой над усредняемой величиной $\overline{X(t)}$, $\overline{(X(t) - m_x)^2}$.

Эргодическими называют стационарные случайные процессы, для которых усреднение по ансамблю реализаций (5.20) и (5.21) и усреднение по времени (формулы (5.26)) при $T \to \infty$ дает один и тот же результат:

$$\begin{gathered}
\int_{-\infty}^{\infty} xW(x)\,dx = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)\,dt;\\
\int_{-\infty}^{\infty} (x-m_x)^2 W(x)\,dx = \lim_{T\to\infty}\frac{1}{T}\int_0^T \left[x(t)-m_x\right]^2 dt;\\
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_x)(x_\tau-m_x)W(x;x_\tau/\tau)\,dx\,dx_\tau =\\
=\lim_{T\to\infty}\frac{1}{T}\int_0^T \left[x(t)-m_x\right]\times\left[x(t+\tau)-m_x\right]dt.
\end{gathered} \tag{5.27}$$

Стационарность случайного процесса не является достаточным условием эргодичности случайных процессов.

В качестве примера рассмотрим случайный процесс $Z(t) = X(t) + Y$, где $X(t)$ — стационарный случайный процесс, обладающий свойством эргодичности, а Y — не меняющаяся во времени случайная величина.

На рисунке 5.8 приведены реализации случайного процесса $z_1(t)$, $z_2(t)$ и $z_3(t)$ при трех различных значениях случайной величины y_1; y_2; y_3:

$$z_1(t) = x(t) + y_1;\ z_2(t) = x(t) + y_2;\ z_3(t) = x(t) + y_3.$$

При усреднении по ансамблю реализаций для математического ожидания случайного процесса $Z(t)$ получим:

$$m_z = M[Z(t)] = M[X(t)] + M[Y] = m_x + m_y.$$

При усреднении по времени, как показывает рис. 5.8, результат усреднения будет зависеть от того, по какой из реализаций $Z(t)$ проводилось усреднение:

$$\overline{Z_1(t)} = \frac{1}{T}\int_0^T z_1(t)\,dt = \frac{1}{T}\int_0^T x(t)\,dt + \frac{1}{T}\int_0^T y_1\,dt = m_x + y_1;$$

$$\overline{Z_2(t)} = \frac{1}{T}\int_0^T z_2(t)\,dt = m_x + y_2.$$

Из сравнения полученных результатов усреднения по ансамблю реализаций и по времени мы видим, что процесс $Z(t)$ не является эргодическим.

Достаточным условием эргодичности стационарного случайного процесса относительно моментов первого и второго порядка является условие (5.15), т.е. функция корреляции стационарного случайного процесса должна стремиться к нулю при $\tau \to \infty$.

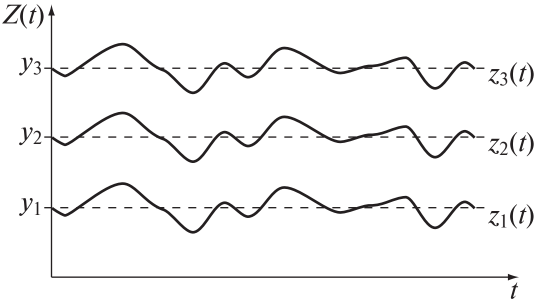


Рис. 5.8. Реализация случайного процесса

На практике условие $T \to \infty$, входящее в формулы (5.27), не является достаточно жестким. Усреднение по времени может быть произведено и при конечном значении времени усреднения T. Реализация случайного процесса считается достаточно длинной по времени, если она задана на интервале наблюдения (интервале усреднения) $0 \div T$ значительно большем, чем время корреляции случайного процесса $T >> \tau_k$.

На основе этого можно сделать вывод о том, что формулы (5.26) могут применяться для определения числовых характеристик нестационарных случайных процессов. Но это возможно только в том случае, если $m_x(t)$, $D_x(t)$ и $k_x(t; \tau)$ являются медленно меняющимися функциями аргумента t и в пределах интервала усреднения T их можно считать постоянными, не зависящими от времени t.

5.6. Взаимная функция корреляции случайных процессов

При решении задач экономического прогнозирования часто приходится рассматривать системы случайных процессов. Из множества показателей экономического развития регионов выделим для примера, два — валовый региональный продукт $X(t)$ и инвестиции в основной капитал $Y(t)$. Эти показатели определяются результатами каждодневной деятельности хозяйствующих субъектов, региональных и местных органов исполнительной власти и многими другими факторами, поэтому являются случайными процессами. По итогам каждого года в регионах производится оценка этих показателей, и эта оценка представляет собой отсчеты случайных процессов $X(t)$ и $Y(t)$ на конец каждого года.

Из определения сущности понятий валового регионального продукта и инвестиций в основной капитал можно увидеть, что эти показатели являются зависимыми друг от друга. Прогнозирование значений этих процессов (показателей) $X(t)$ и $Y(t)$ на будущие годы осуществляется на основе их реализаций $x(t)$ и $y(t)$ за предыдущие годы с учетом взаимосвязи этих случайных процессов.

В общем случае взаимную связь случайных процессов $X(t)$ и $Y(t)$ по аналогии с корреляционным моментом двух случайных величин можно определить функцией взаимной корреляции случайных процессов $X(t)$ и $Y(t)$:

$$k_{xy}(t_1;t_2)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left[x_1-m_x(t_1)\right]\left[y_2-m_y(t_2)\right]W(x_1;\,y_2\,/\,t_1;t_2)dx_1dy_2, \qquad (5.28)$$

где $x_1 = X(t_1)$ и $y_2 = Y(t_2)$ — случайные значения процессов $X(t)$ и $Y(t)$ соответственно в моменты времени t_1 и t_2; $m_x(t_1)$ и $m_y(t_2)$ — математические ожидания случайных процессов $X(t)$ и $Y(t)$ в моменты времени t_1 и t_2; $W(x_1;\, y_2/t_1;\, t_2)$ — совместная плотность вероятности случайных значений процессов $x(t_1)$ и $y(t_2)$.

Из формулы (5.28) следует, что при $t_1 = t_2 = t$ функция взаимной корреляции процессов будет равна

$$k_{xy}(t)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left[x-m_x(t)\right]\left[y-m_y(t)\right]W(x;\,y\,/\,t)\,dx\,dy. \qquad (5.29)$$

Случайные значения $x = X(t)$ и $y = Y(t)$ являются случайными величинами и $k_{xy}(t)$ является корреляционным моментом случайных процессов $X(t)$ и $Y(t)$, определяемым в совпадающие моменты времени $t_1 = t_2 = t$.

Функция взаимной корреляции $k_{xy}(t_1;\, t_2)$ является корреляционным моментом случайных отсчетов процессов $X(t)$ и $Y(t)$, сдвинутых относительно друг друга на интервал времени $\tau = t_2 - t_1$.

Для случайных величин X и Y было показано (формулы (3.71) и (4.43а), что корреляционный момент k_{xy} имеет максимальное значение при линейной взаимосвязи этих случайных величин. Предположим, что случайные процессы $X(t)$ и $Y(t)$ также связаны линейной функциональной зависимостью вида

$$Y(t) = aX(t - \tau) + b. \qquad (5.30)$$

Из соотношения (5.30) следует, что значение случайного процесса $Y(t)$ в момент времени t определяется линейной зависимостью от значения случайного процесса $X(t)$ в предшествующий момент времени $t - \tau$.

В соответствии со свойствами математического ожидания и дисперсии для числовых характеристик $Y(t)$ можно записать:

$$m_y(t) = am_x(t-\tau) + b; \\ D_y(t) = a^2 D_x(t-\tau). \tag{5.31}$$

Определим функцию взаимной корреляции процессов $X(t)$ и $Y(t)$ при их линейной функциональной взаимосвязи (5.30). В соответствии с формулами (5.28), (5.30) и (5.31) для функции взаимной корреляции получим

$$\begin{aligned} k_{xy}(t-\tau;\, t) &= M\{[X(t-\tau) - m_x(t-\tau)]\times \\ &\times[aX(t-\tau) + b - am_x(t-\tau) - b]\} = \\ &= M\{a[X(t-\tau) - m_x(t-\tau)]^2\} = aD_x(t-\tau) = \sigma_x(t-\tau)\sigma_y(t) \end{aligned} \tag{5.32}$$

Из формулы (5.32) следует, что при линейной взаимосвязи случайных процессов вида (5.30) функция их взаимной корреляции имеет максимальное значение, равное $\sigma_x(t-\tau)\sigma_y(t)$.

Для случайных процессов $X(t)$ и $Y(t)$, не имеющих между собой детерминированной функциональной взаимосвязи, функция взаимной корреляции стремится к нулю при $\tau \to \infty$:

$$\lim_{\tau\to\infty} k_{xy}(t;\, t+\tau) = 0\, . \tag{5.33}$$

На рисунке 5.9 приведен один из возможных видов графика взаимной функции корреляции двух случайных процессов $X(t)$ и $Y(t)$. По графику функции взаимной корреляции можно сделать следующие выводы.

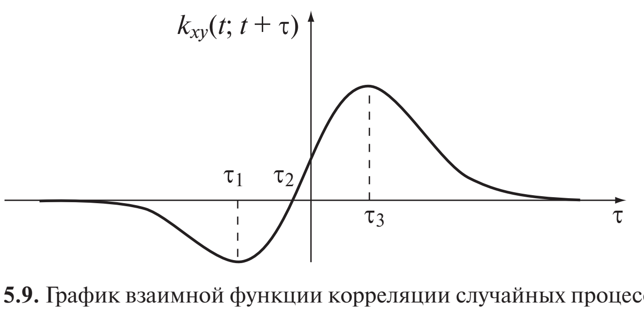


Рис. 5.9. График взаимной функции корреляции случайных процессов

При $\tau = \tau_2$ случайные процессы $X(t)$ и $Y(t)$ будут некоррелированы, так как $k_{xy}(t;\, t-\tau_2) = 0$. При $\tau = \tau_3$ функция взаимной корреляции принимает максимальное (положительное) значение. Значит, между случайными процессами $X(t)$ и $Y(t)$ при $\tau = \tau_3$ существует максимальная прямая зависимость, т.е. увеличение значений случайного процесса $X(t)$ приводит к увеличению значений процесса $Y(t)$.

При $\tau = \tau_1$ функция взаимной корреляции имеет минимальное отрицательное значение. Значит, между процессами $X(t)$ и $Y(t)$ при $\tau = \tau_1$ существует максимальная обратная зависимость, т.е. увеличение значений $X(t)$ приводит к уменьшению значений $Y(t)$.

5.7. Примеры решения задач

Задача 5.1. При организации выпуска нового вида продукции и увеличении объема выпуска данной продукции приток денежных средств $y(t)$ от реализации данной продукции моделируется нормальным случайным процессом с плотностью вероятности:

$$W(y;t) = \frac{1}{\sqrt{2\pi D_y(t)}} \exp\left\{-\frac{\left[y(t) - m_y(t)\right]^2}{2D_y(t)}\right\},$$

где $m_y(t) = m_y(1 - e^{-\lambda t})$ — математическое ожидание притока денежных средств; $D_y(t) = D_y(1 - e^{-\lambda t})$ — дисперсия притока денежных средств.

Отток денежных средств $x(t)$, определяющийся издержками на производство данного вида продукции, также моделируется нормальным случайным процессом с плотностью вероятности:

$$W(x;t) = \frac{1}{\sqrt{2\pi D_x(t)}} \exp\left\{-\frac{\left[x(t) - m_x(t)\right]^2}{2D_x(t)}\right\},$$

где $m_x(t) = m_x(1 - e^{-\lambda t})$ — математическое ожидание оттока денежных средств; $D_x(t) = D_x(1 - e^{-\lambda t})$ — дисперсия оттока денежных средств.

Построить график зависимости от времени $t > 0$ вероятности того, что чистый денежный поток $z(t) = y(t) - x(t)$ будет иметь положительные значения $P[z(t) \geq 0]$.

При расчетах принять, что $m(x) = 0{,}1$, $m(y) = 1{,}3m_x$, $D_y = 1{,}69D_x$, $D_x = 0{,}01$, а также то, что приток денежных средств $y(t)$ и отток денежных средств $x(t)$ являются независимыми случайными процессами.

Решение. Из приведенных выражений для плотностей вероятностей процессов $y(t)$ и $x(t)$ следует, что это нестационарные случайные процессы с изменяющимися во времени математическими ожиданиями и дисперсиями:

$$\frac{m_y(t)}{m_y} = \frac{m_x(t)}{m_x} = \frac{D_y(t)}{D_y} = \frac{D_x(t)}{D_x} = 1 - e^{-\lambda t},$$

где параметр λ характеризует время $t_{\text{ст}}$, в течение которого процессы x(t) и $x(t)$ приходят в стационарное состояние $t_{\text{ср}} \approx \frac{1}{\lambda}$.

График зависимости нормированных значений математических ожиданий и дисперсий этих процессов от времени приведен на рис. 5.10.

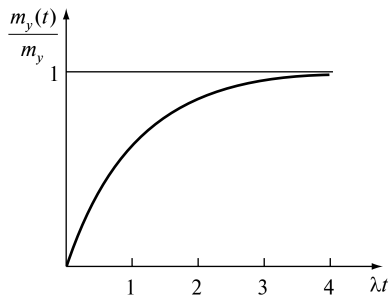


Рис. 5.10. График зависимости нормированных значений математических ожиданий и дисперсий этих процессов от времени

Так как чистый денежный поток $z(t)$ является разностью потоков $y(t)$ и $x(t)$, имеющих нормальную плотность вероятности, то поток $z(t)$ будет иметь плотность вероятности вида:

$$W(z, t) = \frac{1}{\sqrt{2\pi D_z(t)}} \exp\left\{-\frac{\left[z(t) - m_z(t)\right]^2}{2D_z(t)}\right\},$$

где $m_z(t)$ — математическое ожидание; $D_z(t)$ — дисперсия чистого денежного потока определяются формулами:

$$m_z(t) = m_y(t) - m_x(t) = m_z(1 - e^{-\lambda t});$$
$$m_z = m_y - m_x = 1{,}3m_x - m_x = 0{,}3m_x;$$
$$D_z(t) = D_y(t) + D_x(t) = D_z(1 - e^{-\lambda t});$$
$$D_z = D_y + D_x = 1{,}69D_x + D_x = 2{,}69D_x.$$

Среднее квадратичное отклонение чистого денежного потока от математического ожидания $\sigma_z(t) = \sqrt{D_z(t)}$ будет равно:

$$\sigma_z(t) \approx 1{,}64\,\sigma_x (1 - e^{-\lambda t})^{1/2},$$

$$\sigma_x = \sqrt{D_x} = 0{,}1.$$

Вероятность того, что чистый денежный поток $z(t)$ имеет положительные значения, будет равна:

$$P(z(t) \geq 0) = \int_0^{\infty} w(z;t)dz = 1 - \int_{-\infty}^{0} w(z;t)dz = 1 - \Phi\left[-\frac{m_z(t)}{\sigma_z(t)}\right] =$$

$$= \Phi\left[\frac{m_z(t)}{\sigma_z(t)}\right] = \Phi\left[\frac{0,3}{0,164}(1-e^{-\lambda t})^{1/2}\right].$$

Результаты расчетов $P[z(t) \geq 0]$ приведены в табл. 5.1.

Таблица 5.1

Результаты расчетов

λt	0,125	0,25	0,5	1,0	1,5	2,0	2,5	$\lambda t \to \infty$
$P(z(t) \geq 0)$	0,7356	0,8051	0,8729	0,9265	0,9463	0,9554	0,9599	0,9664

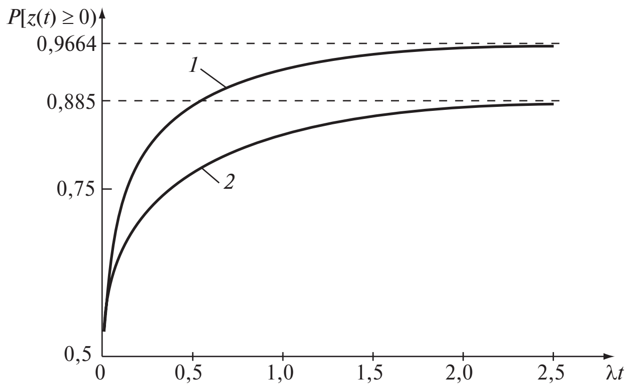


Рис. 5.11. Кривые вероятностей

Из приведенных расчетов следует, что максимальное значение вероятности $P[z(t) \geq 0]$ при $t \to \infty$ определяется отношением $\frac{m_y - m_x}{\sqrt{D_y + D_x}}$:

$$P[z(t) \geq 0]_{t \to \infty} = \Phi\left[\frac{m_y - m_x}{\sqrt{D_y + D_x}}\right].$$

Кривая *1* на рис. 5.11 получена для рассматриваемого в данной задаче случая, когда $\frac{m_y - m_x}{\sqrt{D_y + D_x}} \approx \frac{0,3}{0,164} \approx 1,83$. Кривая *2* рассчитана для случая, когда $\frac{m_y - m_x}{\sqrt{D_y + D_x}} = 1,2$.

В данной задаче рассматривалась ситуация, когда потоки $y(t)$ и $x(t)$ являются независимыми случайными процессами. В следующей задаче рассмотрим ситуацию, когда приток $y(t)$ и отток $x(t)$ денежных средств являются зависимыми случайными процессами.

Задача 5.2. При организации выпуска нового вида продукции поток $x(t)$, направляемый на финансирование затрат на производство данного вида продукции, при $t \geq 0$ моделируется нормальным стационарным случайным процессом с математическим ожиданием $m_x(t) = m_x = 10$ ден. ед., средним квадратичным отклонением $\sigma_x(t) = \sigma_x = 2$ ден. ед. и функцией корреляции $k_x(\tau) = \sigma^2{}_x e^{-\alpha|\tau|}$.

Приток денежных средств, получаемых от реализации данного вида продукции, определяется формулой:

$$\begin{gathered} y(t) = ax(t-T) \text{ при } t \geq T, \\ y(t) = 0 \text{ при } t > T, \end{gathered} \tag{5.34}$$

где T — время определяется длительностью производственно-финансового цикла; $a = (1 + N_{\text{р}})$ — коэффициент определяется нормой рентабельности $N_{\text{р}} = 0{,}3$.

Записать выражения для плотностей вероятностей оттока денежных средств $x(t)$, притока денежных средств $y(t)$ и чистого денежного потока $z(t)$:

$$\begin{gathered} z(t) = y(t) - x(t) = ax(t-T) - x(t) \text{ при } t \geq T; \\ z(t) = -x(t) \text{ при } t < T. \end{gathered} \tag{5.35}$$

Определить вероятность того, что чистый денежный поток $z(t)$ будет иметь положительные значения $P(z \geq 0)$ при $t \geq T$.

Решение. Из формулы (5.35) следует, что чистый денежный поток $z(t)$ определяется значениями случайного процесса $x(t)$ в два момента времени t и $(t - T)$. Поэтому для вероятностного описания процесса $x(t)$ запишем совместную плотность вероятности двух значений этого процесса в моменты времени t и $(t + \tau)$.

Так как $x(t)$ по условию задачи является нормальным стационарным процессом, то плотность вероятности значений случайного процесса в два момента времени $x(t) = x$ и $x(t + \tau) = x_\tau$ по аналогии с формулами (3.72) и (5.3) можно записать в виде:

$$w(x; x_\tau) = \frac{1}{2\pi\sigma_x^2\sqrt{1-\rho_x^2(\tau)}} \times$$
$$\times \exp\left\{-\frac{(x-m_x)^2 - 2\rho_x(\tau)(x-m_x)(x_\tau - m_x) + (x_\tau - m_x)^2}{2\sigma_x^2[1-\rho_x^2(\tau)]}\right\},$$

где $m_x = 10$ ден. ед.; $\sigma_x = 2$ ден. ед.; $\rho_x(\tau) = e^{-\alpha|\tau|}$ — коэффициент корреляции, характеризующий взаимосвязь двух значений процесса $x(t)$ и два момента времени, отстоящих друг от друга на интервал времени τ.

Время корреляции τ_k, в течение которого значения случайного процесса $x(t)$ и $x(t + \tau_k)$ можно считать зависимыми, равно (5.25) $\tau_k = \int\limits_0^\infty e^{-\alpha\tau} d\tau = \dfrac{1}{\alpha}$.

Пример возможной реализации $x(t)$ приведен на рис. 5.12, *а*, примеры реализации процессов $y(t)$ и $z(t)$ приведены на рис. 5.12, *б* и 5.12, *в*.

Так как $x(t)$ является нормальным стационарным процессом, то, как следует из формул (5.34) и (5.35), приток денежных средств $y(t)$ и чистый денежный поток $z(t)$ на интервале $t \geq T$ также будут нормальными стационарными случайными процессами. Для записи их плотностей вероятности определим математическое ожидание и дисперсию процессов $y(t)$ и $z(t)$ на интервале времени $t \geq T$.

На основании свойства 2 (см. подраздел 2.4.1) для математического ожидания m_y получим

$$m_y = M[y(t)] = aM[x(t-T)] = am_x. \tag{5.36}$$

На основании свойства 2 (подраздел 2.4.2) для дисперсии D_y получим

$$D_y = D[y(t)] = a^2 D[x(t-T)] = a^2\sigma_x^2. \tag{5.37}$$

На основании свойств 2, 3 (см. подраздел 2.4.1) и формулы (5.31) для математического ожидания m_z процесса $z(t)$ получим

$$\begin{gathered} m_z = M[z(t)] = aM[x(t-T)] - M[x(t)] = \\ = am_x - m_x = m_x(a-1) = N_\text{р} m_x. \end{gathered} \tag{5.38}$$

На основании формул (4.37) и (4.40) для дисперсии D_z процесса $z(t)$ получим:

$$D_z = M[z^2(t)] - m_z^2 = M\{[ax(t-T) - x(t)]^2\} - m_z^2 = a^2 M[x^2(t-T)] -$$
$$-2aM[x(t)x(t-T)] + M[x^2(t)] - am_x^2 + 2am^2 - m_x^2 = a^2 D_x + D_x - 2ak_x(T),$$

где $k_x(T) = \sigma_x^2 e^{-\alpha|T|}$ — функция корреляции процесса $x(t)$ при $\tau = T$.

Окончательно для дисперсии процесса $z(t)$ можно записать:

$$D_z = \sigma_x^2\left(1 + a^2 - 2ae^{-\alpha|T|}\right). \tag{5.39}$$

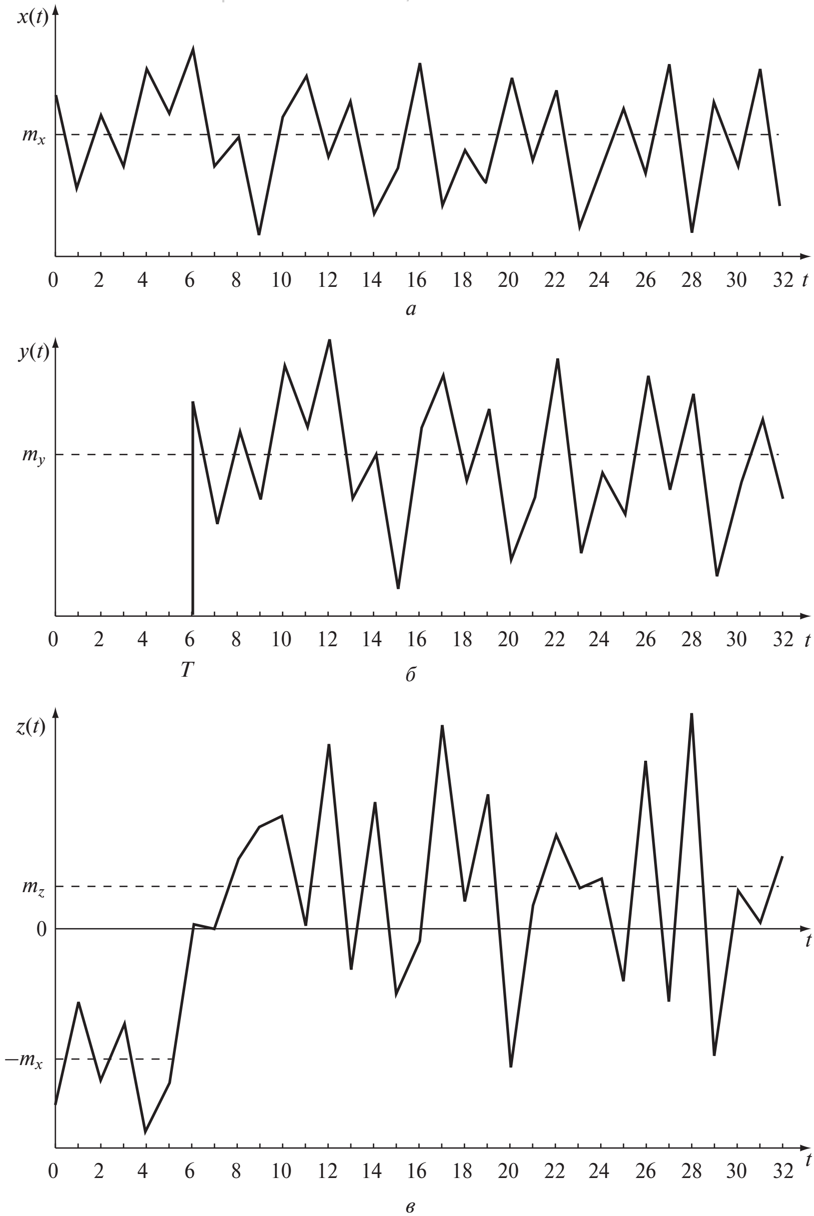


Рис. 5.12. Возможные реализации:
а — $x(t)$; *б* — $y(t)$; *в* — $z(t)$

С учетом полученных значений математических ожиданий и дисперсий плотности вероятности процессов $y(t)$ и $z(t)$ на интервале времени $t \geq T$ будут определяться формулами:

$$w(y) = \frac{1}{\sqrt{2\pi} a \sigma_x} \exp\left[-\frac{(y - am_x)^2}{2a^2 \sigma_x^2} \right];$$

$$w(z) = \frac{1}{\sqrt{2\pi D_z}} \exp\left[-\frac{(z - N_p m_x)^2}{2D_z} \right],$$

где D_z — определяется формулой (5.39).

Вероятность того, что чистый денежный поток $z(t)$ на интервале времени $t \geq T$ будет иметь положительные значения $P(z \geq 0)$, будет равна:

$$P(z \geq 0) = \int_0^\infty w(z)dz = 1 - \int_\infty^0 w(z)dz = 1 - F_z(0),$$

где $F_z(0)$ — функция распределения процесса $z(t)$ при значении аргумента $z = 0$.

Для нормального закона распределения (формула (3.16)) функция распределения равна табулированному интегралу вероятности:

$$F(z) = \Phi\left(\frac{z - m_z}{\sigma_z} \right).$$

Отсюда для вероятности $P(z \geq 0)$ получим:

$$P(z \geq 0) = \Phi\left(\frac{m_z}{\sigma_z} \right).$$

После подстановки значений m_z и σ_z из формул (5.38) и (5.39) для вероятности того, что чистый денежный поток будет иметь положительные значения, окончательно получим:

$$P(z \geq 0) = \Phi\left[\frac{N_\text{p} m_z}{\sigma_z \sqrt{1 + a^2 - 2ae^{-\alpha|T|}}} \right].$$

Помимо численных значений, заданных в условии задачи N_p, m_x, σ_x, искомая вероятность зависит от соотношения длительности производственно-финансового цикла и времени корреляции процесса $x(t)$. Данное соотношение определяется как $\dfrac{T}{\tau_k} = \alpha T$.

Значения вероятностей $P(z \geq 0)$ при различных значениях соотношения $\dfrac{T}{\tau_k} = \alpha T$ приведены в табл. 5.2, а на рис. 5.13 приведен график зависимости $P(z \geq 0)$ от αT.

Таблица 5.2

Значения вероятностей $P(z \geq 0)$

$\dfrac{T}{\tau_k} = \alpha T$	0	$\dfrac{1}{8}$	$\dfrac{1}{4}$	$\dfrac{1}{2}$	1	2	3	∞
$P(z \geq 0)$	≈ 1	0,9934	0,9664	0,9222	0,8729	0,8365	0,8264	0,8198

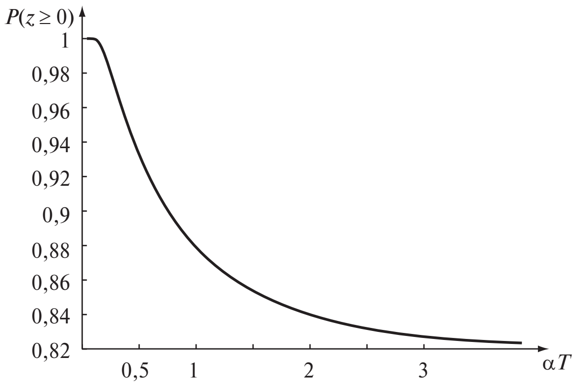


Рис. 5.13. График зависимости $P(z \geq 0)$ от αT

Из результатов расчета следует, что при $T > \tau_k$ или $\alpha T > 1$ (см. рис. 5.12), когда отток денежных средств $x(t)$ несколько раз изменяется относительно математического ожидания m_x за длительность производственно-финансового цикла T, вероятность того, что чистый денежный поток будет отрицательным, $P(z < 0) = 1 - P(z \geq 0)$, имеет существенное значение $P(z < 0) > 0{,}13$. Для уменьшения этой вероятности $P(z < 0)$ необходимо, чтобы время корреляции процесса $x(t)$ было больше длительности производственно-финансового цикла $\tau_k \geq T$, т.е. необходимо, чтобы случайные изменения оттока денежных средств происходили как можно медленнее во времени.

Вопросы для самоконтроля

1. Дайте основные понятия о случайных процессах. Что такое реализация случайного процесса, пучок (ансамбль) реализаций?

2. Что такое вероятностное описание непрерывных случайных процессов? В чем смысл многомерной и двумерной плотности вероятности случайного процесса?
3. Дайте понятие двумерной и одномерной плотности вероятности случайного процесса, имеющего нормальный закон распределения.
4. Что такое математическое ожидание и дисперсия случайного процесса?
5. Что характеризует функция корреляции случайных процессов?
6. Каковы свойства функции корреляции случайных процессов?
7. Зачем нормированная функция корреляции вводится и что характеризует?
8. Дайте определение времени корреляции случайного процесса по функции корреляции и нормированной функции корреляции. Что оно характеризует?
9. Каким образом провести определение стационарных случайных процессов в узком и широком смысле?
10. Чем характеризуется эргодическое свойство стационарных случайных процессов? Что такое усреднение случайных процессов по ансамблю реализаций и по времени?
11. Взаимная корреляционная функция случайных процессов. Как она определяется, что она характеризует?

ГЛАВА 6

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

6.1. Основные задачи математической статистики

В предыдущих главах рассматривались законы распределения случайных величин, случайных процессов и их систем. При этом не затрагивался вопрос, как эти законы получены. Как и само понятие вероятности, так и законы распределения вероятностей определяются на основе исследований развития случайных явлений и количественных значений случайных величин, получаемых в результате опыта. Разработка методов этих исследований — методов регистрации, систематизации, описания и анализа статистических данных, получаемых в результате наблюдения массовых случайных явлений, — составляет предмет математической статистики.

Математическая статистика оперирует статистическими данными, которые получены в процессе наблюдения в виде количественных значений некоторой случайной величины. Множество всех возможных значений случайной величины (конечное или бесконечное) в математической статистике называется ***генеральной совокупностью***. Естественно, что в результате опыта мы не можем получить все возможные значения случайной величины. Значения случайной величины, полученные в результате опыта, являются некоторой случайной, ограниченной по числу значений, выборкой из генеральной совокупности и называются ***выборочной совокупностью***. Количество элементов (значений случайной величины), вошедших в выборочную совокупность, называется ***объемом выборки*** и обозначается буквой n. Объем генеральной совокупности обозначается буквой N и может быть как конечным, так и бесконечным. В математической статистике чрезвычайно важным является вопрос о том, какой объем выборки n будет достаточным (репрезентативным) для получения достоверных представлений о распределении случайной величины. При выборке малого объема обработка статистических данных может дать искаженное представление о вероятностных характеристиках случайной величины. При увеличении объема выборки $n \to N$ точность статистической оценки вероятностных характеристик случайной величины будет повышаться, но при этом увеличиваются также затраты на проведение эксперимента и обработку его результатов.

Таким образом, объем выборки снизу ограничен требованиями к ее репрезентативности, а сверху — экономической целесообразностью. Вопрос о достаточности объема выборки решается в каждом случае отдельно.

Рассмотрим основные задачи математической статистики, часто решаемые на практике.

1. Построение закона распределения исследуемой случайной величины по полученным экспериментальным данным, т.е. по выборочной совокупности. Эта задача заключается в том, чтобы в результате статистической обработки выборочной совокупности значений случайной величины построить графики ее распределения. При обработке выборочных значений может ставиться задача построения экспериментальной (статистической) функции распределения $F^*(x) = P^*(X < x)$ или экспериментальной (статистической) плотности вероятности $W^*(x)$ случайной величины. Используемая в обозначениях $F^*(x)$ и $W^*(x)$ звездочка (*) свидетельствует о том, что данная функция распределения и плотность вероятности получены на основе статистической обработки экспериментальных данных. На практике исследователю приходится часто довольствоваться ограниченным объемом выборки, поэтому результат наблюдений и их обработки всегда содержит больший или меньший элемент случайности. Возникает вопрос: какие элементы полученных экспериментальных законов распределения являются постоянными, устойчивыми закономерностями, а какие элементы обусловлены случайностью и ограниченностью выборки?

Естественно, что элементы полученных экспериментальных законов распределения, обусловленные случайными отклонениями из-за ограниченности объема выборки, необходимо исключить. Это можно сделать путем сглаживания полученных экспериментальных графиков $F^*(x)$ и $W^*(x)$ в виде некоторой кривой, описываемой известным аналитическим выражением. Таких сглаживающих кривых, описываемых известными аналитическими выражениями, может быть подобрано несколько, и в связи с этим возникает следующая задача математической статистики — статистическая проверка гипотез.

2. Статистическая проверка гипотез. Семейство сглаживающих экспериментальные законы распределения $F^*(x)$ и $W^*(x)$ кривых, описываемых известными аналитическими зависимостями, называют ***гипотезами о теоретическом законе распределения случайной величины***. Решение задачи статистической проверки гипотез предполагает получение ответа на вопрос, какая из гипотез (сглаживающих кривых) наилучшим образом согласуется с экспериментальными данными. Если мы получили ответ на этот вопрос, то данная гипотеза принимается в качестве теоретического закона распределения $F(x)$ или $W(x)$ исследуемой случайной величины.

В математической статистике разработан ряд специальных приемов статистической проверки различных гипотез. Эти приемы предполагают решение задачи определения параметров теоретических законов распределения, выбранных в качестве гипотез, и определение критериев, в соответствии с которыми из всех гипотез выбирается одна, наилучшая.

3. Задача нахождения неизвестных параметров теоретического распределения. Каждый из выбранных в качестве гипотезы теоретических законов распределения определяется значениями соответствующих параметров.

Для нормального закона распределения такими параметрами являются математическое ожидание и дисперсия случайной величины. Для экспоненциального распределения это параметр λ. В подразделе 3.4 показано, что моменты распределения различных порядков (например, математическое ожидание и дисперсия) определяются через параметры данного закона распределения. Вполне естественным является следующий алгоритм решения поставленной задачи. Вначале по выборочным значениям производят оценку моментов распределения полученных экспериментальных законов $\nu_n^*, \mu_n^*,$ затем полученные оценки ν_n^*, μ_n^* приравнивают соответствующим моментам теоретического распределения и из полученных уравнений определяют параметры теоретических законов распределения. Например, выбранный в качестве гипотезы закон распределения Эрланга (3.19) полностью определяется двумя параметрами k и λ. Для оценки двух неизвестных параметров k и λ достаточно определить по выборочным данным два момента распределения, например начальный момент первого порядка, равный математическому ожиданию $\nu_1^* = m_x^*$, и центральный момент второго порядка, равный дисперсии $\mu_2^* = D_x^*$. Для выбранного теоретического закона данные моменты определяются формулами (3.20), которые позволяют составить следующую систему уравнений:

$$\begin{cases} m_x^* = \dfrac{k+1}{\lambda}; \\ D_x^* = \dfrac{k+1}{\lambda^2}. \end{cases}$$

Из решения данной системы двух уравнений легко находятся два неизвестных параметра k и λ.

Но в задаче нахождения неизвестных параметров распределения остался открытым вопрос, с помощью каких статистических методов можно получить наилучшую оценку моментов ν_n^* и μ_n^* по имеющимся выборочным данным. Это как раз и является одной из основных задач математической статистики. Один из возможных методов статистической оценки моментов распределения по выборочным данным будет рассмотрен в подразделе 6.3.

4. Задача оценки корреляционных связей системы случайных величин. Для системы зависимых друг от друга случайных величин важной является задача определения корреляционной связи между этими случайными величинами. В подразделе 3.7 было показано, что основными характеристиками, определяющими зависимость случайных величин друг от друга, являются их корреляционный момент k_{xy} (или коэффициент корреляции ρ_{xy}) и уравнения регрессии одной случайной величины на другую.

Разработка методов оценки названных статистических характеристик системы случайных величин по выборочным данным составляет суть данной задачи математической статистики.

6.2. Построение экспериментальных законов распределения случайных величин

Предположим, что в результате экспериментальных наблюдений получена выборочная совокупность значений некоторой случайной величины X с объемом выборки, равным n. Первоначально выборочные данные представляются в табл. 6.1 в порядке их поступления в зависимости от порядкового номера испытания $i = 1 \div n$.

Данная выборка является неупорядоченной, так как с увеличением порядкового номера испытания значения x_1, x_2, ..., x_i, ..., x_n могут случайным образом увеличиваться или уменьшаться. Для удобства статистической обработки по полученной выборке (см. табл. 6.1) строят

Таблица 6.1

Первоначальные выборочные данные

Номер испытания	1	2	3	…	i	…	n
Выборочное значение случайной величины	x_1	x_2	x_3		x_i		x_n

упорядоченную выборку. Для ее построения выборочные значения случайной величины заносят в табл. 6.2 в порядке их возрастания и производят их перенумерацию.

Таблица 6.2

Упорядоченные выборочные данные

Номер испытания	1	2	…	j	…	n
Выборочное значение случайной величины	x_1	x_2		x_j		x_n

В таблице 6.2 первое значение x_1 равно минимальному выборочному значению из табл. 6.1 $x_1 = x_{\min}$, а последнее значение x_n равно максимальному выборочному значению из табл. 6.1 $x_n = x_{\max}$. Промежуточные значения записываются в порядке возрастания:

$$x_2 \geq x_1; \quad x_3 \geq x_2, \ \ldots, \quad x_j \geq x_{j-1.}$$

Если в таблице 6.1 встречаются повторяющиеся выборочные значения, например $x_k = x_m = x_l$, то эти значения записываются в табл. 6.2 последовательно друг за другом в произвольном порядке.

Если объем выборки невелик и среди значений x_j в табл. 6.2 нет одинаковых значений случайной величины, то можно ограничиться данной таблицей и перейти к построению экспериментальной функции распределения $F^*(x)$.

В соответствии со статистическим определением вероятности (1.2) как относительной частоты наступления некоторого события для экспериментальной функции распределения $F^*(x)$ можно записать

$$F^*(x_j) = P^*(X < x_j) = \frac{m_j}{n}, \tag{6.1}$$

где m_j — число значений случайной величины X в табл. 6.2, которые меньше, чем x_j.

Изменяя значения x_j от $x_1 = x_{\min}$ до $x_n = x_{\max}$, получим график экспериментальной функции распределения $F^*(x)$.

Пример 6.1. По результатам наблюдения в течение одного рабочего дня за размерами вкладов населения на счета физических лиц получена упорядоченная выборка, приведенная в табл. 6.3. Требуется построить экспериментальную функцию распределения $F^*(x)$ исследуемой случайной величины.

Таблица 6.3

Результаты наблюдения

Номер испытания	1	2	3	4	5	6	7
Выборочное значение случайной величины	200	300	350	500	550	600	1 000

Решение. Объем упорядоченной выборки (см. табл. 6.3) равен $n = 7$. В соответствии с формулой (6.1) находим значения экспериментальной функции распределения

$$F^*(200) = 0;\ F^*(300) = \frac{1}{7};\ F^*(350) = \frac{2}{7};$$

$$F^*(500) = \frac{3}{7};\ F^*(550) = \frac{4}{7};\ F^*(600) = \frac{5}{7};$$

$$F^*(1000) = \frac{6}{7};\ F^*(1000 + \varepsilon) = \frac{7}{7} = 1.$$

В последнем значении $F^*(1000 + \varepsilon)$ величина ε является сколь угодно малой величиной. По полученным данным строим экспериментальную функцию распределения (рис. 6.1).

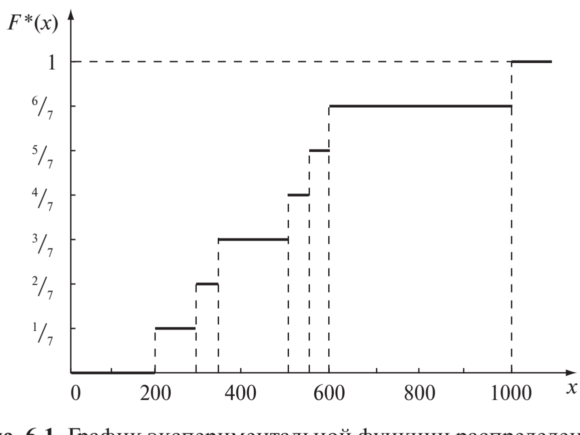


Рис. 6.1. График экспериментальной функции распределения

Рассмотренный в данном примере метод построения экспериментальной функции распределения используют для статистического исследования дискретных случайных величин или, как указывалось выше, в случае ограниченного объема выборки. Для выборки большого объема табл. 6.3 выглядела бы довольно длинной и неудобной для использования, тем более что в ней встречались бы повторяющиеся, равные друг другу выборочные значения случайной величины. В этом случае для дальнейшей статистической обработки экспериментальных данных по упорядоченной выборке (см. табл. 6.2) формируется сгруппированная выборка, в которой вся выборочная совокупность значений случайной величины группируется (разбивается) по k интервалам (группам). Для этого размах вариации $R = x_{\max} - x_{\min}$ делится на k интервалов и определяется шаг интервала h:

$$h = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k}. \tag{6.2}$$

Весь размах вариации разбивается на k интервалов точками (см. рис. 6.2) $x^{(0)}, x^{(1)}, x^{(2)}, \ldots, x^{(k)}$, которые определяют начало и конец каждого интервала. Так, для i-го интервала значение $x^{(i-1)}$ определяет его начало, а x^{i} — его окончание. В соответствии с формулой (6.2) для границ j-го интервала можно записать формулы:

$$j = 1;\ x^{(0)} = x_{\min};\ x^{(1)} = x_{\min} + h;$$
$$j = 2;\ x^{(1)} = x_{\min} + h;\ x^{(2)} = x_{\min} + 2h;$$
$$j = i;\ x^{(i-1)} = x_{\min} + (i-1)h;\ x^{(i)} = x_{\min} + ih;$$
$$j = k;\ x^{(k-1)} = x_{\min} + (k-1)h;\ x^{(k)} = x_{\min} + kh = x_{\max}.$$

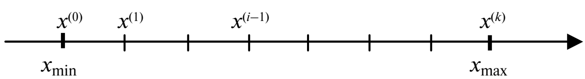


Рис. 6.2. Разбиение размаха вариации на интервалы

Далее подсчитывают, сколько выборочных значений случайной величины попадает в каждый j-й интервал. Это число m_j называют частотой попадания случайной величины в j-й интервал. Результаты расчетов заносят в табл. 6.4 (строки 1, 2, 3).

При правильных расчетах должно выполняться условие $\sum_{j=1}^{k} m_j = n$.

Число интервалов разбиения k ориентировочно определяется из неравенства

$$k \le 5 \lg n \tag{6.3}$$

Таблица 6.4

Результаты расчетов

Номер интервала	1	2	...	k
Границы интервала	$x^{(0)} \div x^{(1)}$	$x^{(1)} \div x^{(2)}$		$x^{(k-1)} \div x^{(k)}$
Частота m_j	m_1	m_2		m_k
Относительная частота $P_j^* = \dfrac{m_j}{n}$	P_1^*	P_2^*		P_k^*
Плотность вероятности $W^*(x_j)$	$W^*(x_1)$	$W^*(x_2)$		$W^*(xk)$

и обычно варьирует в пределах от 6 до 20. Следует иметь в виду, при что очень малом значении k результаты статистической обработки могут быть недостаточно точными, при больших значениях k увеличивается объем вычислительных процедур, и результаты обработки также могут быть недостаточно точными, если относительная частота m_j/n попадания случайной величины в какие-то j-е интервалы будет иметь очень маленькие значения.

По сгруппированной выборке, приведенной в строках 1, 2 и 3 табл. 6.4, можно построить экспериментальные законы распределения $F^*(x)$ и $W^*(x)$ исследуемой случайной величины.

Рассмотрим алгоритм построения экспериментальной плотности вероятности случайной величины $W^*(x)$, график которой называют гистограммой распределения.

По данным сгруппированной выборки вычисляются относительные частоты попадания выборочных значений в j-е интервалы, которые в соответствии с формулой (1.2) равны экспериментальным значениям P_j^* вероятности попадания случайной величины в j-й интервал:

$$P_j^* = m_j/n.$$

Результаты вычисления значений P_j^* $(j = 1 \div k)$ заносятся в табл. 6.4. При правильных расчетах должно выполняться условие $\sum_{j=1}^{k} P_j^* = 1.$ При построении гистограммы распределения $W^*(x)$ предполагается, что в пределах каждого j-го интервала плотность вероятности не меняется и равна некоторому постоянному числу

$$W^*(x) = W^*(x_j) = \text{const при } x^{(j-1)} < x < x^{(j)}. \tag{6.4}$$

С учетом предположения (6.4) и формулы (4.2) для вероятностей P_j^* можно записать:

$$P_j^* = W^*(x_j)h \text{ или } W^*(x_j) = \frac{P_j^*}{h}. \qquad (6.5)$$

Результаты расчетов $W^*(x_j)$ по формуле (6.5) заносятся в табл. 6.4, и по ним строится гистограмма распределения $W^*(x)$. Гистограмма распределения (рис. 6.3) представляет собой ступенчатую функцию, состоящую из прямоугольников, основанием которых служат интервалы $x^{(j-1)} \div x^{(j)}$, $j = 1 \div k$, а высоты равны вычисленным значениям плотности вероятности $W^*(x_j)$.

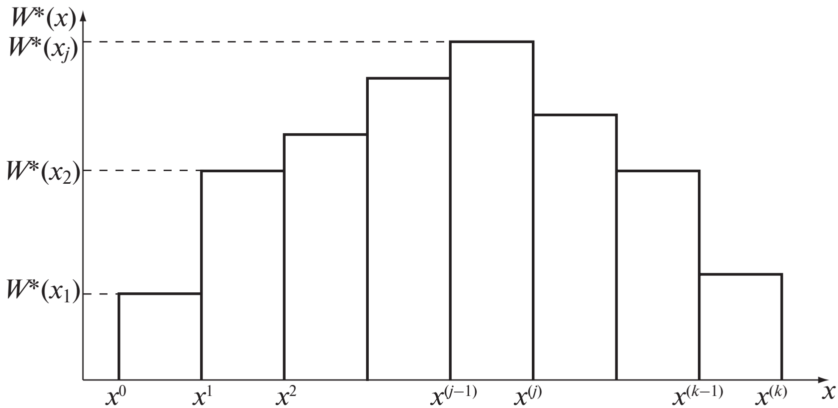


Рис. 6.3. Гистограмма распределения $W^*(x)$

Результаты расчетов, занесенные в табл. 6.4, позволяют построить также экспериментальную функцию распределения $F^*(x)$. Для этого рассчитаем значения функции распределения на границах интервала:

$$F^*(x_0) = P^*(X < x_0) = 0;$$
$$F^*(x_1) = P^*(X < x_1) = P_1^*;$$
$$F^*(x_2) = P^*(X < x_2) = P_1^* + P_2^*;$$
$$F^*(x_j) = P^*(X < x_j) = P_1^* + P_2^* + \ldots + P_j^*;$$
$$F^*(x_k) = P^*(X < x_k) = \sum_{j=1}^{k} P_j^* = 1.$$

Так как в пределах каждого j-го интервала $W^*(x_j)$ имеет постоянное значение, то функция распределения в пределах каждого интервала будет линейно возрастать от значения $F^*(x^{(j-1)})$ до значения $F^*(x^{(j)})$. График экспериментальной функции распределения, соответствующий гистограмме распределения (см. рис. 6.3), приведен на рис. 6.4.

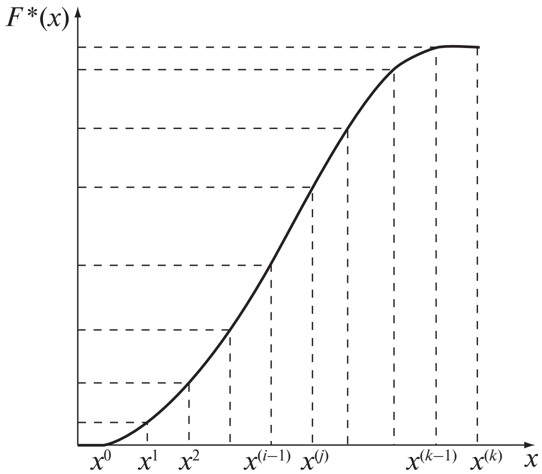


Рис. 6.4. Экспериментальная функция распределения

Гистограмма распределения $W^*(x)$ и экспериментальная функция распределения $F^*(x)$ дают наглядное представление о характере распределения случайной величины. Если гистограмму распределения сгладить плавной кривой, то по ее виду можно судить, какой из теоретических законов распределения может быть выбран в качестве гипотезы о законе распределения исследуемой случайной величины.

Пример 6.2. В таблице приложения 5 приведены выборочные значения валового регионального продукта на душу населения X.

Построить гистограмму распределения данной случайной величины.

Решение. По таблице (см. приложение 5) находим размах вариации:

$$R = x_{\max} - x_{\min} = 37{,}1 - 3{,}1 = 34.$$

Объем выборки равен $n = 78$. По формуле (6.3) находим ориентировочное число интервалов разбиения:

$$k \leq 5 \lg n = 5 \lg 78 = 5 \cdot 1{,}89 = 9{,}45.$$

Выбираем $k = 8$.

По формуле (6.2) определяем шаг интервала

$$h = \frac{R}{k} = \frac{34}{8} = 4{,}25\,.$$

Записываем границы интервалов в табл. 6.5.

Таблица 6.5

Границы интервалов

Номер интервала	1	2	3	4	5	6	7	8
Границы интервала	3,1÷ 7,35	7,35÷ 11,6	11,6÷ 15,85	15,85 ÷20,1	20,1÷ 24,35	24,35 ÷28,6	28,6÷ 32,85	32,85 ÷37,1
Частота m_j	12	28	21	6	7	1	2	1
Относительная частота $P_j^* = \frac{m_j}{n}$	0,154	0,359	0,269	0,077	0,09	0,013	0,025	0,013
Плотность вероятности $W^*(x_j)$	0,036	0,084	0,063	0,08	0,021	0,003	0,006	0,003

По таблице приложения 5 определяем частоту m_j попадания случайной величины в каждый интервал и заносим в табл. 6.5.

Правильность вычисления значений m_j проверяем по условию:

$$\sum_{j=1}^{k} m_j = n;$$

$$\sum_{j=1}^{k} m_j = 12 + 28 + 21 + 6 + 7 + 1 + 2 + 1 = 78.$$

Первые три строки табл. 6.5 представляют сгруппированную выборку. Определяем относительные частоты попадания случайной величины в j-е интервалы $j = (1 \div 8)$:

$$P_j^* = \frac{m_j}{n};$$

$$P_1^* = \frac{12}{78} \approx 0{,}154\,;\; P_2^* = \frac{28}{78} \approx 0{,}359\,;\; P_3^* = \frac{21}{78} \approx 0{,}269;$$

$$P_4^* = \frac{2}{78} \approx 0{,}077\,;\; P_5^* = \frac{7}{78} \approx 0{,}09\,;\; P_6^* = \frac{1}{78} \approx 0{,}013;$$

$$P_7^* = \frac{2}{78} \approx 0{,}025\,;\; P_8^* = \frac{1}{78} \approx 0{,}013.$$

Эти значения равны вероятностям попадания выборочных значений случайной величины в j-е интервалы. При правильном вычислении и округлении результатов должно выполняться условие

$$\sum_{j=1}^{k} P_j^* = 1;$$

$$\sum_{j=1}^{8} P_j^* = 0{,}154 + 0{,}359 + 0{,}269 + 0{,}077 + 0{,}09 + 0{,}013 + 0{,}025 + 0{,}013 = 1.$$

По формуле (6.5) определяем значения экспериментальной плотности вероятности:

$$W^*(xj) = P_j^* / h;$$

$$W^*(x_1) = \frac{0{,}154}{4{,}25} \approx 0{,}036; \quad W^*(x_2) = \frac{0{,}359}{4{,}25} \approx 0{,}084;$$

$$W^*(x_3) = \frac{0{,}269}{4{,}25} \approx 0{,}063; \quad W^*(x_4) = \frac{0{,}077}{4{,}25} \approx 0{,}018;$$

$$W^*(x_5) = \frac{0{,}09}{4{,}25} \approx 0{,}021; \quad W^*(x_6) = W^*(x_8) = \frac{0{,}013}{4{,}25} \approx 0{,}003;$$

$$W^*(x_7) = \frac{0{,}25}{4{,}25} \approx 0{,}006.$$

Результаты вычислений заносим в табл. 6.5 и строим гистограмму распределения (рис. 6.5).

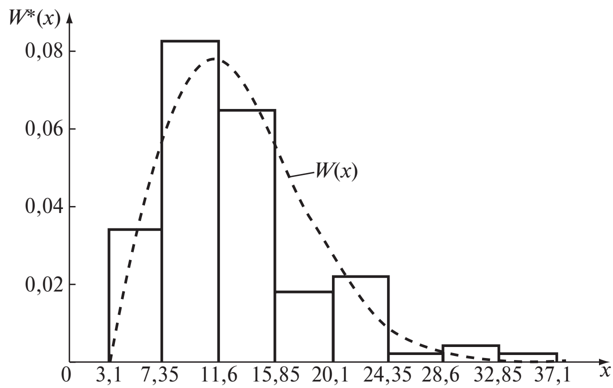


Рис. 6.5. Гистограмма распределения валового регионального продукта на душу населения в регионах Российской Федерации

Пунктирной линией на рис. 6.5 показан график зависимости плотности вероятности $W(x)$, сглаживающий полученную гистограмму распределения $W^*(x)$.

6.3. Точечные оценки параметров распределения

Как было показано в подразделе 6.1, параметры теоретического закона распределения могут быть определены по числовым характеристикам выборочной совокупности исследуемой случайной величины. Как правило, число параметров теоретического закона распределения не превышает двух, поэтому для оценки достаточно определить математическое ожидание и дисперсию выборочной совокупности.

Математическое ожидание выборочной совокупности называют выборочной средней и обозначают $\tilde{X}$. Выборочное среднее значение представляет собой статистический начальный момент первого порядка и определяется как среднее арифметическое всех выборочных значений случайной величины

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{6.6}$$

Статистический центральный момент второго порядка называется выборочной дисперсией S^2 и определяется как среднее арифметическое квадрата отклонения выборочных значений от средней $\tilde{X}$:

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{X})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i - (\tilde{X})^2. \tag{6.7}$$

Формулы (6.6) и (6.7) легкоприменимы при небольшом объеме выборки n. При увеличении n вычислительные процедуры для выборочного среднего $\tilde{X}$ и выборочной дисперсии S_x^2 усложняются.

Для уменьшения объема вычислений при больших n для определения выборочного среднего и выборочной дисперсии воспользуемся сгруппированной выборкой (табл. 6.4). Сгруппированную выборку можно представить как дискретный вариационный ряд, устанавливающий взаимосвязь между значением дискретной случайной величины и вероятностью ее появления. Примем за значения дискретной случайной величины середины j-х интервалов:

$$x_{j\,\text{ср}} = \frac{x^{(j-1)} + x^{(j)}}{2}. \tag{6.8}$$

Тогда таблицу 6.4 можно представить в виде вариационного ряда (табл. 6.6):

Таблица 6.6

Таблица вариационного ряда

Номер интервала j	1	2	...	i	...	k
Середина интервала	$x_{1\text{ср}}$	$x_{2\text{ср}}$	...	$x_{i\text{ср}}$	...	$x_{k\text{ср}}$
Относительная частота P_j^*	P_1^*	P_2^*	...	P_i^*	...	P_k^*

По вариационному ряду и формулам (2.9) и (2.11) можно значительно проще получить оценку выборочного среднего m_x^* и выборочной дисперсии D_x^*:

$$m_x^* = \sum_{j=1}^{k} x_{jcp} P_j^*; \quad D_x^* = \sum_{j=1}^{k} x_{jcp}^2 p_j^* - (m_x^*)^2. \tag{6.9}$$

Проверим, совпадают ли оценки m_x^* и D_x^*, вычисленные по формулам (6.9), с выборочным средним $\tilde{X}$ (формула (6.6)) и выборочной дисперсией (формула (6.7)).

Представим сумму в формуле (6.6) в виде двойной суммы

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{m_j} x_{ij}.$$

При таком представлении вначале суммируются все выборочные значения x_{ij} в пределах j-х интервалов, а затем полученные результаты суммируются по всем интервалам.

Далее сумму $\sum_{i=1}^{m_j} x_{ij}$ умножим и разделим на m_j:

$$\tilde{X} = \sum_{j=1}^{k} \frac{m_j}{n} \frac{1}{m_j} \sum_{i=1}^{m_j} x_{ij} = \sum_{j=1}^{k} p_j^* x_{j\,\text{ср}}, \tag{6.10}$$

где $x_{j\,\text{ср}} = \frac{1}{m_j}\sum_{i=1}^{m_j} x_{ij}$.

Так как при построении гистограммы в пределах каждого интервала плотность вероятности постоянна, то можно записать:

$$x_{j\,\text{ср}} = \frac{1}{m_j}\sum_{i=1}^{m_j} x_{ij} = \frac{x^{(j-1)} + x^{(j)}}{2}.$$

Из формулы (6.10) следует, что вычисления выборочного среднего по формуле (6.6) и по первой формуле (6.9) дают один и тот же результат

$$\tilde{X} = m_x^*.$$

Проведем аналогичные преобразования для выборочной дисперсии

$$S_x^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{m_j} x_{ij}^2 - (\tilde{X})^2 = \sum_{j=1}^{k}\frac{m_j}{n}\frac{1}{m_j}\sum_{i=1}^{m_j} x_{ij}^2 - (\tilde{X})^2. \quad (6.11)$$

В данном выражении внутренняя сумма равна начальному моменту второго порядка выборочных значений x_i, попадающих в j-й интервал

$$\frac{1}{m_j}\sum_{i=1}^{m_j} x_{ij}^2 = \nu_2^{(j)} = D_{xj} + (x_{jcp})^2\ . \quad (6.12)$$

При равномерном распределении x_i в пределах j-го интервала ($W^*(x_j) = \text{const}$) дисперсия D_{xj} будет равна

$$D_{xj} = \frac{\left(x^{(j)} - x^{(j-1)}\right)^2}{12} = \frac{h^2}{12}. \quad (6.13)$$

После подстановки формул (6.13) и (6.12) в (6.11) получим

$$S_x^2 = \frac{h^2}{12}\sum_{j=1}^{k} p_j^* + \sum_{j=1}^{k} p_j^* x_{j\,\text{ср}}^2 - (\tilde{X})^2.$$

В данной формуле сумма в первом слагаемом равна единице по условию нормировки. Второе и третье слагаемые в соответствии с формулой (6.9) равны D_x^*. Окончательно для выборочной дисперсии получим

$$S_x^2 = D_x^* + \frac{h^2}{12}\ . \quad (6.14)$$

Таким образом, формулы (6.9) и (6.14) позволяют значительно упростить вычисления выборочного среднего и выборочной дисперсии при большом объеме выборки.

Для нахождения параметров теоретического закона распределения составим систему уравнений

$$\begin{cases} m_x = \tilde{X}; \\ D_x = S_x^2. \end{cases}, \tag{6.15}$$

где m_x и D_x — математическое ожидание и дисперсия, определяемые по формулам (3.8) и (3.12) для выбранного теоретического закона распределения.

Система уравнений (6.15) означает, что за математическое ожидание случайной величины m_x мы принимаем его оценку по выборочному среднему $\tilde{X}$, а за дисперсию случайной величины принимаем ее оценку по выборочной дисперсии S_x^2.

Так как объем выборки n, как правило, значительно меньше объема генеральной совокупности N, то оценки $\tilde{X}$ и S_x^2, полученные по данной выборочной совокупности, будут случайными величинами. Если мы проведем второй опыт, то из-за случайного отбора выборочных значений из генеральной совокупности вторая выборочная совокупность будет отличаться от первой случайным образом. Значит и оценки $\tilde{X}$ и S_x^2 по формулам (6.6) и (6.7) будут случайными величинами при повторении опыта.

Для того чтобы точечная оценка $\tilde{X}$ была репрезентативной, т.е. достаточно точно представляла параметр m_x, она должна обладать тремя свойствами.

Свойство 1. Несмещенность оценки.

Точечная оценка называется ***несмещенной***, если ее математическое ожидание равно неизвестному оцениваемому параметру

$$M(\tilde{X}) = m_x. \tag{6.16}$$

В противном случае ***оценка называется смещенной***.

Свойство 2. Состоятельность оценки.

Оценка $\tilde{X}$ называется ***состоятельной,*** если она сходится по вероятности к оцениваемому параметру m_x при бесконечном повторении опытов $n \to \infty$

$$\lim_{n\to\infty} P(\left|\tilde{X} - m_x\right| < \varepsilon) = 1, \tag{6.17}$$

где ε — сколь угодно малое положительное число.

Свойство 3. Эффективность оценки.

Несмещенная и состоятельная оценка $\tilde{X}$ называется ***эффективной***, если ее дисперсия $D(\tilde{X})$ имеет минимальное значение

$$D(\tilde{X}) = D_{\min}. \qquad (6.18)$$

Аналогичным образом можно записать условия несмещенности, состоятельности и эффективности для оценки S_x^2.

Рассмотрим, удовлетворяют ли оценки $\tilde{X}$ и S_x^2 первому свойству, т.е. проверим, являются они смещенными или несмещенными. Будем считать, что генеральная совокупность случайной величины X имеет математическое ожидание m_x и дисперсию D_x.

Вычислим математическое ожидание оценки $\tilde{X}$:

$$M(\tilde{X}) = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(x_i).$$

Последнее равенство написано на основании свойств математического ожидания.

Так как x_i являются случайными значениями, выбранными из генеральной совокупности случайной величины X, то $M(x_i) = m_x$, и в результате получим

$$M(\tilde{X}) = \frac{1}{n}\sum_{i=1}^{n} m_x = \frac{1}{n} n m_x = m_x .$$

Из данных вычислений следует, что для оценки $\tilde{X}$ выполняется условие (6.16), т.е. выборочное среднее является несмещенной оценкой математического ожидания m_x.

Для оценки S_x^2 условие несмещенности можно записать в виде

$$M(S_x^2) = D_x. \qquad (6.19)$$

В соответствии с формулой (6.7) для математического ожидания оценки S_x^2 получим

$$M(S_x^2) = M\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n} M(x_i^2) - \frac{1}{n^2} M\left(\sum_{i=1}^{n} x_i\right)^2 . \qquad (6.20)$$

В первом слагаемом в полученном выражении (6.20) под знаком суммы стоит начальный момент второго порядка случайной величины X:

$$M(x_i^2) = D_x + m_x^2 . \qquad (6.21)$$

Второе слагаемое в выражении (6.20) при возведении суммы в квадрат можно записать в виде:

$$M\left(\sum_{i=1}^{n} x_i\right)^2 = M\left(\sum_{i=1}^{n} x_i^2\right) + M\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} x_i\, x_j\right) = \sum_{i=1}^{n} M(x_i^2) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} M(x_i x_j) =$$

$$= \sum_{i=1}^{n}(D_x + m_x^2) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}(k_{ij} + m_x^2), \qquad (6.22)$$

где k_{ij} — корреляционный момент выборочных значений $x_i x_j$ случайной величины X.

Так как выборка значений x_1, x_2, ..., x_i, ..., x_n из генеральной совокупности производится независимо друг от друга, то корреляционный момент k_{ij} будет равен нулю. С учетом этого после суммирования формула (6.22) преобразуется к виду

$$M\left(\sum_{i=1}^{n} x_i\right)^2 = n(D_x + m_x^2) + n(n-1)m_x^2. \qquad (6.23)$$

Подставляя формулы (6.21) и (6.23) в (6.20), получим

$$M(S_x^2) = D_x + m_x^2 - \frac{1}{n}(D_x + m_x^2) - \frac{n-1}{n}m_x^2 = \frac{n-1}{n}D_x +$$

$$+\frac{n-1}{n}m_x^2 - \frac{n-1}{n}m_x^2, \qquad (6.24)$$

или

$$M(S_x^2) = \frac{n-1}{n}D_x.$$

Из формулы (6.24) следует, что математическое ожидание оценки S_x^2 не совпадает с дисперсией случайной величины X, т.е. условие (6.19) не выполняется и выборочная дисперсия S_x^2 дает смещенную оценку D_x. Из формулы (6.24) следует, что несмещенной оценкой дисперсии D_x случайной величины X будет так называемая исправленная выборочная дисперсия $\widehat{S}_x^2$:

$$D_x = \frac{n}{n-1}S_x^2 = \widehat{S}_x^2. \qquad (6.25)$$

При больших значениях n отличие в значениях S_x^2 и $\widehat{S}_x^2$ невелико, при малом объеме выборки n оценка дисперсии D_x должна осуществляться по исправленной выборочной дисперсии.

С учетом полученных результатов система уравнений (6.15) для оценки параметров теоретического закона распределения должна быть записана в виде:

$$\begin{cases} m_x = \tilde{x}; \\ D_x = \widehat{S}_x^2 = \dfrac{n}{n-1} S_x^2. \end{cases} \tag{6.26}$$

Таким образом, алгоритм точечной оценки параметров теоретического закона распределения сводится к следующему.

1. В зависимости от объема выборки по формулам (6.6) и (6.7) или по формулам (6.9), (6.14) вычисляют значения выборочного среднего $\tilde{X}$ и выборочной дисперсии S_x^2.

2. С учетом множителя $\dfrac{n}{n-1}$ по формуле (6.25) определяют значение исправленной выборочной дисперсии $\widehat{S}_x^2$.

3. По формулам (3.8) и (3.12) для выбранного закона распределения $W(x)$ рассчитывают значения m_x и D_x как функции от параметров λ_1, λ_2 выбранного теоретического закона распределения $m_x(\lambda_1;\lambda_2)$, $D_x^2(\lambda_1;\lambda_2)$.

4. По вычисленным значениям $\tilde{X}$, $\widehat{S}_x^2$ и $m_x(\lambda_1;\lambda_2)$, $D_x^2(\lambda_1;\lambda_2)$ составляют систему уравнений (6.26), из решения которой находят параметры λ_1 и λ_2 выбранного теоретического закона распределения.

Пример 6.3. Для условий примера 6.2 вычислить выборочное среднее $\tilde{X}$ и исправленную выборочную дисперсию $\widehat{S}_x^2$.

Решение. Объем выборки в примере 6.2 достаточно велик: $n = 78$, поэтому для упрощения вычислений воспользуемся формулами (6.9).

Вычисляем середины интервалов и составляем вариационный ряд (табл. 6.7):

$$x_{1\text{ср}} = \frac{x^{(o)} + x^{(1)}}{2} = \frac{3{,}1 + 7{,}35}{2} = 5{,}225;$$

$$x_{2\text{ср}} = \frac{x^{(1)} + x^{(2)}}{2} = \frac{7{,}35 + 11{,}6}{2} = 9{,}475;$$

$$x_{3\text{ср}} = \frac{x^{(2)} + x^{(3)}}{2} = \frac{11{,}6 + 15{,}85}{2} = 13{,}725; \quad x_{4\text{ср}} = 17{,}975; \; x_{5\text{ср}} = 22{,}225;$$

$$x_{6\text{ср}} = 26{,}475; \; x_{7\text{ср}} = 30{,}725; \; x_{8\text{ср}} = 34{,}975.$$

Таблица 6.7

Вариационный ряд

Номер интервала j	1	2	3	4	5	6	7	8
Середина интервала $x_{j\text{ср}}$	5,225	9,475	13,725	17,975	22,225	26,475	30,725	34,975
Относительная частота P_j^*	0,154	0,359	0,269	0,077	0,090	0,013	0,025	0,013

Для выборочного среднего получим:

$$\tilde{X} = m_x^* = \sum_{j=1}^{8} x_{j\,\text{ср}} P_j^* = 5{,}225 \cdot 0{,}154 + 9{,}475 \cdot 0{,}359 +$$
$$+13{,}725 \cdot 0{,}269 + 17{,}975 \cdot 0{,}077 + 22{,}225 \cdot 0{,}09 + 26{,}475 \cdot 0{,}013 +$$
$$+30{,}725 \cdot 0{,}025 + 34{,}975 \cdot 0{,}013 = 12{,}818.$$

Рассчитываем оценку дисперсии D_x^* по выборочному ряду

$$D_x^* = \sum_{j=1}^{8} x_{j\,\text{ср}}^2 P_j^* - (m_x^*)^2 = (5{,}225)^2 \cdot 0{,}154 + (9{,}475)^2 \cdot 0{,}359 + (13{,}725)^2 \cdot 0{,}269 +$$
$$+(17{,}975)^2 \cdot 0{,}077 + (22{,}225)^2 \cdot 0{,}09 + (26{,}475)^2 \cdot 0{,}013 + (30{,}725)^2 \cdot 0{,}025 +$$
$$+(34{,}975)^2 \cdot 0{,}013 - (12{,}818)^2 = 40{,}755.$$

С учетом поправки (6.14) рассчитываем выборочную дисперсию S_x^2:

$$S_x^2 = D_x^* + \frac{h^2}{12} = 40{,}755 + \frac{(4{,}25)^2}{12} = 42{,}26.$$

С учетом поправочного множителя $\dfrac{n}{n-1}$ по формуле (6.25) рассчитываем исправленную выборочную дисперсию

$$S_x^2 = \frac{n}{n-1} S_x^2 = \frac{78}{78-1} \cdot 42{,}26 = 42{,}809.$$

6.4. Статистическая проверка гипотезы о законе распределения

Задачу о статистической проверке гипотез рассмотрим на примере проверки гипотезы о законе распределения случайной величины. В этом случае речь идет об идентификации закона распределения исследуемой случайной величины с одним из известных законов распределения.

Предположим, что в результате статистической обработки выборочной совокупности мы получили сгруппированную выборку (см. табл. 6.4) значения P_j^* и построили гистограмму распределения (см. рис. 6.3). По виду гистограммы мы должны выбрать теоретический закон распределения $W(x)$ и принять его в качестве гипотезы. Обозначим эту гипотезу H_0. Количественную оценку соответствия или несоответствия принятой гипотезы H_0 закону распределения случайной величины X будем осуществлять по величине χ^2 — хи-квадрат:

$$\chi^2 = n\sum_{j=1}^{k}\frac{(P_j - P_j^*)^2}{P_j^*}, \tag{6.27}$$

где P_j^* — вероятность попадания выборочных значений случайной величины X в j-й интервал, определяемая по сгруппированной выборке (табл. 6.4); k — количество интервалов разбиения; n — объем выборки; P_j — вероятность попадания случайной величины X в j-й интервал, определяемая по теоретическому закону распределения, выбранному в качестве гипотезы H_0.

$$P_j = \int_{x^{(j-1)}}^{x^{(j)}} W(x)dx = F(x^{(j)}) - F(x^{(j-1)}). \tag{6.28}$$

Из формулы (6.27) следует, что чем больше отличаются значения P_j^* и P_j, тем хуже выбранная гипотеза согласуется с результатами статистической обработки выборочной совокупности. Причем при разных знаках разности $P_j -$ P_j^* за счет ее возведения в квадрат, имеющиеся различия суммируются по всем интервалам с одинаковыми знаками.

Формула (6.27) может быть преобразована также к виду:

$$\chi^2 = \sum_{j=1}^{k}\frac{n^2(P_j - P_j^*)^2}{nP_j^*} = \sum_{j=1}^{k}\frac{(m_j - m_j^*)^2}{m_j^*}, \tag{6.29}$$

где $m_j^* = nP_j^*$ — частота попадания выборочных значений в j-й интервал, определяемая по сгруппированной выборке; $m_j = nP_j$ — частота попадания случайной величины в j-й интервал, определяемая по теоретическому закону распределения.

При многократном повторении опытов по исследованию случайной величины X мы получим несколько выборочных совокупностей, которые будут отличаться друг от друга случайным образом. При статистической обработке этой группы выборочных совокупностей найденные значения P_j^* и m_j^* будут так же случайно изменяться от выборки к выборке, т.е. по своей природе значения P_j^* и m_j^* являются случайными величинами.

Величина хи-квадрат, определяемая по формулам (6.27) или (6.29), также является случайной величиной, плотность вероятности которой имеет χ^2-распределение (3.26) (рис. 6.6).

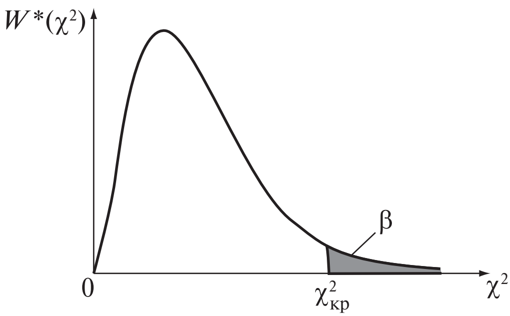


Рис. 6.6. Плотность вероятности χ^2-распределения

Плотность вероятности χ^2-распределения определяется одним параметром, называемым числом степеней свободы v:

$$v = k - s - 1,$$

где k — число интервалов сгруппированной выборки; s — число параметров теоретического закона распределения, принятого в качестве гипотезы H_0.

По результатам статистической обработки выборочной совокупности и выбранной гипотезе о теоретическом законе распределения по формулам (6.27) или (6.29) мы можем рассчитать численное значение величины хи-квадрат $\chi^2_{эксп}$. Чем больше вычисленное значение $\chi^2_{эксп}$, тем хуже выбранная гипотеза (теоретический закон распределения) согласуется с экспериментальными данными. Для ответа на вопрос, может ли гипотеза H_0 быть принята или должна быть отвергнута, вычисленное значение $\chi^2_{эксп}$ нужно сравнить с некоторой критической величиной $\chi^2_{кр}$.

Если $\chi^2_{эксп} \geq \chi^2_{кр}$, то разница между P_j^* и P_j слишком велика и выбранная гипотеза о теоретическом законе распределения должна быть отвергнута.

Если $\chi^2_{эксп} < \chi^2_{кр}$, то гипотеза о выбранном теоретическом законе распределения принимается, так как различия в P_j^* и P_j считаются несущественными.

Критическое значение $\chi^2_{\text{кр}}$ определяют по заданному значению вероятности:

$$\beta = P(\chi^2 \geq \chi^2_{\text{кр}}) = \int_{\chi^2_{\text{кр}}}^{\infty} W(\chi^2)\,d\chi^2 . \tag{6.30}$$

Вероятность β называют ***уровнем значимости***, графически она определяется заштрихованной площадью на рис. 6.6.

Вероятность $\alpha = 1 - \beta$, равную

$$\alpha = P(\chi^2 < \chi^2_{\text{кр}}) = \int_{0}^{\chi^2_{\text{кр}}} W(\chi^2)\,d\chi^2 ,$$

называют ***доверительной вероятностью.***

Результаты расчетов $\chi^2_{\text{кр}}$ по формуле (6.30) с учетом (3.26) для различных значений уровня значимости β и числа степеней свободы v приведены в таблице приложения 2.

Из рисунка 6.6 следует, что при увеличении значения уровня значимости β пороговое (критическое) значение $\chi^2_{\text{кр}}$ будет уменьшаться и требования по соответствию теоретического и экспериментального законов распределения ($\chi^2_{\text{эксп}} < \chi^2_{\text{кр}}$) становятся жестче.

Рассмотрим алгоритм принятия решения о законе распределения исследуемой случайной величины на конкретном примере.

Пример 6.4. Для случайной величины, рассматриваемой в примерах 6.2 и 6.3, выбрать гипотезу H_0 о теоретическом законе ее распределения $W(x)$ и проверить ее на соответствие экспериментальному закону распределения $W^*(x)$ при уровне значимости $\beta = 0,05$.

Решение. При решении примера 6.2 был построен экспериментальный закон распределения $W^*(x)$, представленный в виде гистограммы распределения на рис. 6.5. Сравнивая данную гистограмму с графиками плотности вероятности $W(x)$ теоретических законов распределения, можно видеть, что в качестве гипотезы H_0 может быть принята гипотеза о законе распределения Рэлея (3.31) для случайной величины $Z = X - x_{\min}$:

$$W(z) = \frac{z}{\sigma^2} \exp\left(\frac{z^2}{2\sigma^2}\right), \quad z \geq 0. \tag{6.31}$$

Для исследуемой случайной величины $X = Z + x_{\min} = Z + 3,1$ плотность вероятности будет определяться формулой

$$W(x) = \frac{x-3{,}1}{\sigma^2}\exp\left[-\frac{(x-3{,}1)^2}{2\sigma^2}\right], \quad x \geq 3{,}1. \tag{6.32}$$

Плотности вероятности (6.31) и (6.32) полностью определяются одним параметром σ, поэтому для его определения достаточно провести точечную оценку одного из выборочных моментов, например выборочного среднего.

В примере 6.3 было получено, что выборочное среднее для исследуемой случайной величины равно $\tilde{X} = 12{,}818$. Математическое ожидание m_z при распределении Рэлея определяется формулой (3.33):

$$m_z = \sqrt{\frac{2}{\pi}}\sigma .$$

Используя функциональную взаимосвязь между случайными величинами Z и X, можно записать:

$$m_x = M(z + x_{\min}) = m_z + x_{\min} = m_z + 3{,}1.$$

Приравнивая значение математического ожидания m_x и выборочного среднего $\tilde{X}$, определяем значение параметра σ выбранного теоретического закона распределения:

$$m_x = \tilde{X} = m_z + 3{,}1;$$

$$\tilde{X} = \sqrt{\frac{\pi}{2}}\sigma + 3{,}1;$$

$$\sigma = \sqrt{\frac{2}{\pi}}(12{,}818 - 3{,}1) \approx 7{,}756;$$

$$\sigma^2 \approx 60 .$$

С учетом вычисленного значения параметра σ закона распределения Рэлея для плотности вероятности случайной величины X можно записать:

$$W(x) = \frac{x-3{,}1}{60}\exp\left[-\frac{(x-3{,}1)^2}{120}\right], \quad x \geq 3{,}1. \tag{6.33}$$

Результаты расчетов плотности вероятности $W(x)$ по формуле (6.33) в виде графика приведены на рис. 6.5 пунктирной линией.

Проверим гипотезу (6.33) на соответствие экспериментальному закону распределения $W^*(x)$ по критерию согласия хи-квадрат.

Для этого вычислим вероятности P_j попадания случайной величины X в j-е интервалы по теоретическому закону распределения (6.33):

$$P_j = \int_{x^{(j-1)}}^{x^{(j)}} W(x)dx = \int_{z^{(j-1)}}^{z^{(j)}} W(z)dz = F(z^{(j)}) - F(z^{(j-1)}), \qquad (6.34)$$

где $z^{(j)} = x^{(j)} - 3{,}1$; $F(z)$ — функция распределения Рэлея, определяющаяся формулой (3.32):

$$P_1 = F(z^{(1)}) - F(z^{(0)}) = F(4{,}25) - F(0) = 1 - \exp\left[-\frac{(4{,}25)^2}{120}\right] \approx 0{,}141;$$

$$P_2 = F(8{,}5) - F(4{,}25) = 1 - \exp\left[-\frac{(8{,}5)^2}{120}\right] - 0{,}141 \approx 0{,}311;$$

$$P_3 = F(12{,}75) - F(8{,}5) = 1 - \exp\left[-\frac{(12{,}75)^2}{120}\right] - 0{,}452 \approx 0{,}29;$$

$$P_4 = F(17) - F(12{,}75) = 1 - \exp\left[-\frac{(17)^2}{120}\right] - 0{,}742 \approx 0{,}168;$$

$$P_5 = F(21{,}25) - F(17) = 1 - \exp\left[-\frac{(21{,}25)^2}{120}\right] - 0{,}91 \approx 0{,}066;$$

$$P_6 = F(25{,}5) - F(21{,}25) = 1 - \exp\left[-\frac{(25{,}5)^2}{120}\right] - 0{,}976 \approx 0{,}019;$$

$$P_7 = F(29{,}75) - F(25{,}5) = 1 - \exp\left[-\frac{(29{,}75)^2}{120}\right] - 0{,}995 \approx 0{,}0044;$$

$$P_8 = F(\infty) - F(29{,}75) \approx 1 - 0{,}9994 = 0{,}0006.$$

Результаты расчетов P_j должны удовлетворять условию нормировки

$$\sum_{j=1}^{8} P_j = 1\,.$$

Полученные результаты для P_j с учетом значений P_j^* (см. табл. 6.5) позволяют по формуле (6.27) вычислить значение $\chi^2_{\text{эксп}}$:

$$\chi^2_{\text{эксп}} = 78\left[\frac{(0{,}141 - 0{,}154)^2}{0{,}154} + \frac{(0{,}311 - 0{,}359)^2}{0{,}359} + \frac{(0{,}29 - 0{,}269)^2}{0{,}269} + \right.$$

$$+\frac{(0{,}168-0{,}077)^2}{0{,}077}+\frac{(0{,}066-0{,}09)^2}{0{,}09}+\frac{(0{,}019-0{,}013)^2}{0{,}013}+\frac{(0{,}0044-0{,}025)^2}{0{,}025}+$$

$$\left.+\frac{(0{,}0006-0{,}013)^2}{0{,}013}\right]\approx 12{,}06.$$

Определяем число степеней свободы χ^2-распределения

$$v=k-s-1=8-1-1=6.$$

По таблице приложения 2 для v = 6 и уровня значимости $\beta=0{,}05$ находим критическое значение $\chi^2_{кр}=12{,}59$.

Сравнивая значения $\chi^2_{эксп}$ и $\chi^2_{кр}$, видим, что выполняется неравенство

$$\chi^2_{эксп}<\chi^2_{кр}\;\Rightarrow\;12{,}06<12{,}59.$$

На основании этого можно сделать вывод о том, что теоретическое распределение (6.33) достаточно хорошо согласуется с экспериментальными данными и плотность вероятности (6.33) может быть принята как закон распределения исследуемой случайной величины.

6.5. Корреляционный анализ выборочной совокупности

Одной из наиболее часто встречающихся задач, связанных с исследованием системы случайных величин, является задача выявления зависимости между этими случайными величинами. Наиболее простая — детерминированная функциональная зависимость между случайными величинами — на практике встречается редко. Как правило, между случайными величинами имеет место более сложная зависимость, когда изменение одной случайной величины обусловлено не только изменением другой, но влиянием множества других факторов, зачастую не поддающихся учету. Такая зависимость, при которой изменение одной величины приводит к изменению характеристик распределения другой величины, называется статистической.

Таким образом, статистическая зависимость между случайными величинами X и Y может состоять из двух компонентов: функционального, обусловленного жесткой, детерминированной зависимостью между этими случайными величинами, и случайного, связанного с влиянием как раздельных для X и Y случайных факторов, так и общих.

Взаимосвязь между случайными величинами определяется их корреляционным моментом k_{xy} или коэффициентом корреляции ρ_{xy}, а также уравнениями регрессии одной случайной величины на другую.

Рассмотрим методы определения названных характеристик системы случайных величин X и Y по их выборочным совокупностям.

6.5.1. Оценка выборочного корреляционного момента системы двух случайных величин

Рассмотрим систему двух случайных величин X и Y, для которых получены выборочные совокупности объемом n, приведенные в табл. 6.8.

Таблица 6.8

Выборочные совокупности

Номер испытания j	1	2	...	j	...	n
X	x_1	x_2	...	x_i	...	x_n
Y	y_1	y_2	...	y_i	...	y_n

В таблице 6.8 выборочные значения случайных величин x_i, y_j связаны между собой принадлежностью к общему j-му признаку. В данном случае это номер испытания.

Если объем выборки невелик, то выборочные средние и исправленные выборочные дисперсии случайных величин X и Y можно сравнительно просто определить по формулам:

$$\tilde{X} = \frac{1}{n}\sum_{j=1}^{n} x_j; \quad \tilde{Y} = \frac{1}{n}\sum_{j=1}^{n} y_j; \tag{6.35}$$

$$\begin{aligned} \widehat{S}_x^2 &= \frac{1}{n-1}\sum_{j=1}^{n} x_j^2 - \frac{n}{n-1}(\tilde{X})^2; \\ \widehat{S}_y^2 &= \frac{1}{n-1}\sum_{j=1}^{n} y_j^2 - \frac{n}{n-1}(\tilde{Y})^2. \end{aligned} \tag{6.36}$$

В формулах (6.35), (6.36) осуществляется усреднение по объему выборки. Используя аналогичное усреднение по объему выборки для выборочного корреляционного момента случайных величин X и Y, можно записать:

$$K_{xy}=\frac{1}{n}\sum_{j=1}^{n}(x_j-\tilde{X})(y_j-\tilde{Y})=\frac{1}{n}\sum_{j=1}^{n}x_j y_j-\tilde{X}\tilde{Y}. \tag{6.37}$$

Выборочный коэффициент корреляции случайных величин будет определяться формулой

$$r_{xy}=\frac{K_{xy}}{\sqrt{\widehat{S}_x^2\,\widehat{S}_y^2}}. \tag{6.38}$$

Рассмотрим методику определения выборочных значений корреляционного момента и коэффициента корреляции на конкретном примере.

Пример 6.5. Из таблицы приложения 5 по распределению валового регионального продукта Y и количества предприятий и организаций X по регионам Российской Федерации произведена случайная выборка значений X и Y, приведенная в табл. 6.9.

Таблица 6.9

Случайная выборка значений *X* и *Y*

Номер региона j	1	2	3	4	5	6	7	8	9	10
Количество предприятий X, тыс.	4,3	7,3	10,2	15,7	19,9	22,4	37,4	40	47,8	68,7
ВРП на душу населения Y, тыс. руб.	7,3	5,7	6,9	23,3	19,8	13,5	20,2	17,8	14,0	15,8

Определить выборочный корреляционный момент и выборочный коэффициент корреляции случайных величин X и Y.

Решение. По формулам (6.35) определяем выборочные средние случайных величин:

$$\tilde{X}=\frac{1}{n}\sum_{j=1}^{n}x_j=\frac{1}{10}[4{,}3+7{,}3+10{,}2+15{,}7+19{,}9+$$
$$+22{,}4+37{,}4+40+47{,}8+68{,}7]=27{,}37;$$

$$\tilde{Y}=\frac{1}{n}\sum_{j=1}^{n}y_j=\frac{1}{10}[7{,}3+5{,}7+6{,}9+23{,}8+19{,}8+13{,}5+$$
$$+20{,}2+17{,}8+14{,}0+18{,}8]=14{,}43.$$

По формулам (6.36) определяем исправленные выборочные дисперсии:

$$\hat{S}_y^2 = \frac{1}{n-1}\sum_{j=1}^{n} x_j^2 - \frac{n}{n-1}(\tilde{X})^2 = \frac{1}{9}\big(4,3^2 + 7,3^2 + 10,2^2 + 15,7^2 + 19,9^2 +$$

$$+22,4^2 + 37,4^2 + 40^2 + 47,8^2 + 68,7^2\big) - \frac{10}{9}(27,37)^2 \approx 425,8;$$

$$\hat{S}_x^2 = \frac{1}{n-1}\sum_{j=1}^{n} y_j^2 - \frac{n}{n-1}(\tilde{Y})^2 = \frac{1}{9}\big(7,3^2 + 5,7^2 + 6,9^2 + 23,3^2 + 19,8^2 + 13,5^2 +$$

$$+20,2^2 + 17,8^2 + 14,0^2 + 15,8^2\big) - \frac{10}{9}(14,43)^2 \approx 37,65.$$

По формуле (6.37) определяем выборочный корреляционный момент

$$K_{xy} = \frac{1}{n}\sum_{j=1}^{n} x_j y_j - \tilde{X}\cdot\tilde{Y} = \frac{1}{10}\big(4,3\cdot 7,3 + 7,3\cdot 5,7 + 10,2\cdot 6,9 +$$
$$+15,7\cdot 23,3 + 19,9\cdot 19,8 + 22,4\cdot 13,5 + 37,4\cdot 20,2 + 40\cdot 17,8 +$$
$$+47,8\cdot 14 + 68,7\cdot 15,8\big) - 27,37\cdot 14,43 \approx 50,83.$$

По формуле (6.38) определяем выборочный коэффициент корреляции случайных величин

$$r_{xy} = \frac{K_{xy}}{\sqrt{S_x^2 S_y^2}} = \frac{50,83}{\sqrt{425,8\cdot 37,65}} \approx 0,4.$$

При большом объеме выборки n, когда вычисления по формулам (6.35) — (6.37) становятся слишком громоздкими, корреляционный анализ можно проводить по сгруппированным выборкам. Для этого в соответствии с методикой, изложенной в подразделе 6.2, определяется размах вариации случайных величин X и Y:

$$R_x = x_{\max} - x_{\min};$$
$$R_y = y_{\max} - y_{\min}.$$

По формуле (6.3) определяется ориентировочно число интервалов разбиения вариаций X и Y. Даже при одинаковых объемах выборки случайных величин X и Y фактическое число интервалов разбиения k_x и k_y, исходя из условий удобства и повышения точности вычислений, может быть $k_x \neq k_y$.

Далее по формуле (6.2) определяется шаг интервалов для случайных величин X, Y:

$$h_x = \frac{R_x}{k_x} = \frac{x_{\max} - x_{\min}}{k_x};$$

$$h_y = \frac{R_y}{k_y} = \frac{y_{\max} - y_{\min}}{k_y},$$

записываются границы каждого интервала по X и по Y:

$$x_{i\,\min} = x_{\min} + (i-1)h_x;$$
$$x_{i\,\max} = x_{\min} + ih_x;$$
$$y_{j\,\min} = y_{\min} + (j-1)h_y;$$
$$y_{ji\,\max} = y_{\min} + jh_y$$

и определяются середины каждого интервала:

$$x_{i\,\text{ср}} = \frac{x_{i\,\min} + x_{i\,\max}}{2};$$

$$y_{j\,\text{ср}} = \frac{y_{j\,\min} + y_{j\,\max}}{2}.$$

По вычисленным значениям границ интервалов случайных величин X, Y выборочная совокупность преобразуется в вариационный ряд системы случайных величин X, Y (табл. 6.10).

Таблица 6.10

Вариационный ряд систем случайных величин

	Номер интервала i	1	2	...	i	...	k_x
Номер интервала j	$x_{i\,\text{ср}}$ / $y_{j\,\text{ср}}$	$x_{1\text{ср}}$	$x_{2\text{ср}}$	...	$x_{i\text{ср}}$	...	$x_{k\text{ср}}$
1	$y_{1\text{ср}}$	m_{11}	m_{12}	...	m_{1i}	...	m_{1k}
2	$y_{2\text{ср}}$	m_{21}	m_{22}	...	m_{2i}	...	m_{2k}
...	...	...	...	...	...	...	...
j	$y_{j\text{ср}}$	m_{j1}	m_{j2}	...	m_{ji}	...	m_{jk}
...	...	...	...	...	...	...	...
k_y	$y_{k\text{ср}}$	m_{l1}	m_{l2}	...	m_{li}	...	m_{lk}

В поле таблицы 6.10 указываются частоты m_{ji}, которые характеризуют число выпадения случаев, когда одновременно выполняются два условия: $y_{j\,\min} < Y < y_{j\,\max}$ и $x_{i\,\min} < X < x_{i\,\max}$. При правильном преобразовании выборочной совокупности в вариационный ряд системы случайных величин X, Y должны выполняться следующие условия.

1. Сумма частот m_{ji} каждой строки равна частоте попадания случайной величины в j-й интервал m_j:

$$\sum_{i=1}^{k_x} m_{ji} = m_{yj}. \tag{6.39}$$

2. Сумма частот m_{ji} каждого столбца равна частоте попадания случайной величины X в i-й интервал m_{xi}:

$$\sum_{j=1}^{k_y} m_{ji} = m_{xi}. \tag{6.40}$$

3. Сумма частот m_{xi} или m_{yj} должна быть равна объему выборки n:

$$\sum_{j=1}^{k_y} m_{yj} = \sum_{i=1}^{k_x} m_{xi} = \sum_{j=1}^{k_y}\sum_{i=1}^{k_x} m_{ji} = n. \tag{6.41}$$

Вариационный ряд (табл. 6.10) системы случайных величин X, Y является аналогом совместного распределения системы дискретных случайных величин $x_{i\,\text{ср}}$; $y_{j\,\text{ср}}$, так как относительные частоты

$$P_{ji}^* = \frac{m_{ji}}{n} \tag{6.42}$$

характеризуют совместную вероятность выполнения двух условий: $y_{j\,\min} < Y < y_{j\,\max}$; $x_{i\,\min} < X < x_{i\,\max}$. Тогда с учетом формул (6.39), (6.40) и (6.42) формулы (6.9) и (6.14) для выборочных величин преобразуются к виду:

1) выборочные средние:

$$\begin{aligned} \tilde{X} &= \frac{1}{n}\sum_{i=1}^{k_x} x_{i\,\text{ср}} m_{xi} = \frac{1}{n}\sum_{j=1}^{k_y}\sum_{i=1}^{k_x} x_{i\,\text{ср}} m_{ji}; \\ \tilde{Y} &= \frac{1}{n}\sum_{j=1}^{k_y} y_{j\,\text{ср}} m_{yj} = \frac{1}{n}\sum_{j=1}^{k_y}\sum_{i=1}^{k_x} y_{j\,\text{ср}} m_{ji}; \end{aligned} \tag{6.43}$$

2) выборочные дисперсии:

$$\begin{aligned} S_x^2 &= \frac{1}{n}\sum_{i=1}^{k_x} x_{i\,\text{ср}}^2 m_{xi} - (\tilde{X})^2 + \frac{h_x^2}{12}; \\ S_y^2 &= \frac{1}{n}\sum_{i=1}^{k_y} y_{j\,\text{ср}}^2 m_{yj} - (\tilde{Y})^2 + \frac{h_y^2}{12}; \end{aligned} \tag{6.44}$$

3) для выборочного корреляционного момента по аналогии с формулой (2.22) можно записать:

$$K_{xy} = \frac{1}{n}\sum_{i=1}^{k_x}\sum_{j=1}^{k_y} x_i\, y_j\, m_{ji} - \tilde{X}\tilde{Y}; \qquad (6.45)$$

4) выборочный коэффициент корреляции r_{xy} определяется формулой

$$r_{xy} = \frac{K_{xy}}{\sqrt{S_x^2 S_y^2}}. \qquad (6.46)$$

Как всякая выборочная оценка, выборочный коэффициент корреляции, рассчитанный по формулам (6.43) — (6.46), является случайной величиной.

Значение выборочного коэффициента корреляции, рассчитанное по конкретной выборочной совокупности, может быть отлично от нуля, даже если случайные величины X и Y независимы. Значит, для проверки гипотезы об отсутствии (или наличии) корреляции между случайными величинами X и Y необходимо установить, значимо ли полученное отличие выборочного коэффициента корреляции r_{xy} от нуля. Для этого проверим гипотезу H_0 о равенстве нулю коэффициента корреляции ρ_{xy} генеральной совокупности случайных величин X и Y. Проверку гипотезы H_0 можно выполнить по случайной величине

$$T = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}},$$

которая имеет распределение Стьюдента с $v = n - 2$ степенями свободы (рис. 6.7).

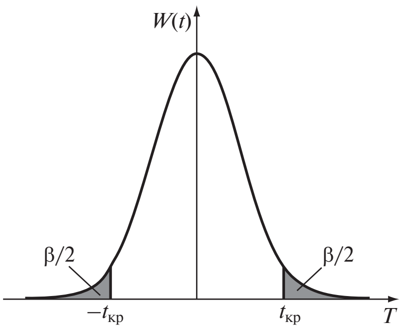


Рис. 6.7. Плотность вероятности распределения Стьюдента

Величину T называют T-статистикой Стьюдента. Если при заданном уровне значимости β

$$\beta = 1 - \int_{-t_{\text{кр}}}^{t_{\text{кр}}} W(T)\,dT,$$

вычисленное экспериментальное значение

$$t_{\text{эксп}} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} \qquad (6.47)$$

по абсолютной величине не превышает критического значения $t_{\text{кр}}(\beta;\,\nu)$

$$\left|t_{\text{эксп}}\right| < t_{\text{кр}}(\beta;\,\nu),$$

то принимается гипотеза H_0 о равенстве нулю коэффициента корреляции ρ_{xy} генеральной совокупности случайных величин X, Y.

Если $\left|t_{\text{эксп}}\right| > t_{\text{кр}}(\beta;\,\nu)$, то корреляционную связь между случайными величинами X, Y следует признать существенной. Критические значения $t_{\text{кр}}(\beta;\,\nu)$ приведены в приложении 3.

Пример 6.6. Результаты исследований по объемам выполняемых работ Y (млн руб.) и накладным расходам X (млн руб.) при выполнении этих работ сведены в вариационный ряд, приведенный в табл. 6.11.

Требуется:

1) вычислить выборочный коэффициент корреляции случайных величин X, Y;

2) проверить гипотезу о значимости полученного коэффициента корреляции r_{xy} при уровне значимости $\beta = 0{,}05$.

Таблица 6.11

Вариационный ряд

$y_{i\,\text{ср}}$ \ $x_{i\,\text{ср}}$	1,5	2,5	3,5	4,5	5,5	6,5	7,5	8,5	m_{yj}
15	4	5	0	0	0	0	0	0	9
25	1	3	1	0	0	0	0	0	5
35	2	3	6	5	3	1	0	0	20
45	0	5	6	19	8	7	2	1	51
55	0	1	2	7	16	9	4	2	41
65	0	0	1	5	6	4	2	2	20
75	0	0	0	0	0	0	1	3	4
m_{xi}	7	17	19	36	33	21	9	8	150

Решение. По формулам (6.39) и (6.40) вычисляем частоты попадания случайной величины Y в j-е интервалы m_{yj} и случайной величины X в i-е интервалы m_{xi}. *Р*езультаты вычисления занесены в табл. 6.11.

Из формулы (6.41) определяем объем выборки:

$$n = \sum_{j=1}^{7} m_{yj} = 9 + 5 + 20 + 51 + 41 + 20 + 4 = 150;$$

$$n = \sum_{i=1}^{8} m_{xi} = 7 + 17 + 19 + 36 + 33 + 21 + 9 + 8 = 150.$$

По формулам (6.43) определяем выборочные средние случайных величин X и Y:

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{8} x_{i\,\text{ср}} m_{xi} = \frac{1}{150}(1{,}5\cdot 7 + 2{,}5\cdot 1{,}7 + 3{,}5\cdot 19 + 4{,}5\cdot 36 + 5{,}5\cdot 33 + 6{,}5\cdot 21 +$$
$$+ 7{,}5\cdot 9 + 8{,}5\cdot 8) = 4{,}9;$$

$$\tilde{Y} = \frac{1}{n}\sum_{j=1}^{7} y_{j\,\text{ср}} m_{y\,j} = \frac{1}{150}\left[15\cdot 9 + 25\cdot 5 + 35\cdot 20 + 45\cdot 51 + 55\cdot 41 + 65\cdot 20 +\right.$$
$$\left. + 75\cdot 4\right] = 47{,}4.$$

Определяем шаг интервалов по случайным величинам X, Y:

$$h_x = x_i - x_{i-1} = 1;$$
$$h_y = y_i - y_{i-1} = 10.$$

По формулам (6.44) определяем выборочные дисперсии:

$$S_x^2 = \frac{1}{150}\left(1{,}5^2\cdot 7 + 2{,}5^2\cdot 17 + 3{,}5^2\cdot 19 + 4{,}5^2\cdot 36 + 5{,}5^2\cdot 33 + 6{,}5^2\cdot 21 +\right.$$
$$\left. + 7{,}5^2\cdot 9 + 8{,}5^2\cdot 8\right) - (4{,}9)^2 + \frac{1}{12} = 3{,}096;$$

$$S_y^2 = \frac{1}{150}\left(15^2\cdot 9 + 25^2\cdot 5 + 35^2\cdot 20 + 45^2\cdot 51 + 55^2\cdot 41 + 65^2\cdot 20 + 75^2\cdot 4\right) -$$
$$-(47{,}4)^2 + \frac{(10)^2}{12} = 187{,}906.$$

По формуле (6.45) определяем выборочный корреляционный момент

$$K_{xy} = \frac{1}{150}\sum_{i=1}^{8}\sum_{j=1}^{7} x_{i\,\text{ср}} y_{j\,\text{ср}} m_{ji} - 4{,}9\cdot 47{,}4 = \frac{1}{150}\ \cdot\ 37\,175 - 232{,}26 \approx 15{,}57.$$

Выборочный коэффициент корреляции в соответствии с формулой (6.46) будет равен

$$r_{xy} = \frac{K_{xy}}{\sqrt{S_x^2\, S_y^2}} = \frac{15{,}57}{\sqrt{3{,}096 \cdot 187{,}906}} \approx 0{,}646\,.$$

Проверим значимость выборочного коэффициента корреляции. Для этого по формуле (6.47) определяем экспериментальное значение t-статистики Стьюдента:

$$t_{\text{эксп}} = 0{,}646\sqrt{\frac{150-2}{1-(0{,}646)^2}} \approx 10{,}3.$$

По таблицам распределения Стьюдента (см. приложение 3) при уровне значимости $\beta = 1 - \alpha = 0{,}05$ и числе степеней свободы $v = n - 2 = 148$ находим критическое значение t-статистики:

$$1{,}96 < t_{\text{кр}}(\beta;\, v) \ < 1{,}98.$$

По найденным значениям $t_{\text{эксп}}$ и $t_{\text{кр}}(\beta;\, v)$ можно сделать вывод, что выполняется условие

$$t_{\text{эксп}} > t_{\text{кр}}(\beta;\, v)\,.$$

Из этого следует, что случайные величины X и Y имеют существенную корреляционную взаимосвязь. Коэффициент корреляции этих случайных величин может быть принят равным $r_{xy} = 0{,}646$.

6.5.2. Регрессионный анализ двух случайных величин

Задачами регрессионного анализа являются выбор аналитической зависимости (модели) между случайными величинами, определение оптимальных значений параметров выбранной модели и оценка точности полученной модели.

Предположим, что мы имеем выборочные совокупности двух случайных величин X, Y (например, табл. 6.9). Величину X будем считать независимой переменой, а Y — зависимой. Тогда пары значений x_i, y_i системы двух случайных величин можно отобразить точками на координатной плоскости (рис. 6.8).

Из рисунка 6.8 следует, что между случайными величинами X и Y имеется аналитическая зависимость, которая заключается в том, что при увеличении случайной величины X случайная величина Y имеет также тенденцию к увеличению.

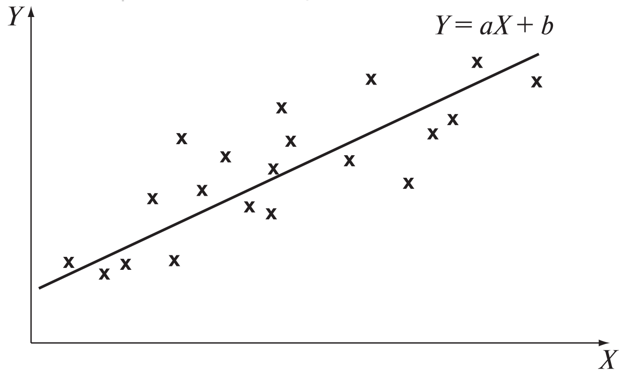


Рис. 6.8. Выборочные значения системы случайных величин X и Y

Будем рассматривать линейные модели аналитической зависимости между случайными величинами:

$$Y = aX + b. \tag{6.48}$$

В соответствии с выбранной моделью (6.48) зависимость между значениями системы двух случайных величин можно записать в виде:

$$y_i = ax_i + b + \varepsilon_i, \tag{6.49}$$

где ε_i характеризует случайный разброс выборочных значений y_i от уравнения прямой (6.48).

Так как мы ограничились линейной моделью аналитической зависимости между случайными величинами, то теперь в рамках данной модели мы должны наилучшим образом выбрать значения параметров a и b.

Оптимальные значения (оценку) параметров модели будем определять из условия минимизации среднего квадрата случайной величины ε_i:

$$\min \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - ax_i - b\right]^2 = F. \tag{6.50}$$

Запишем необходимые условия экстремума функционала F (6.50):

$$\frac{\partial F}{\partial b} = -\frac{2}{n}\sum_{i=1}^{n}\left[y_i - ax_i - b\right] = 0; \tag{6.51}$$

$$\frac{\partial F}{\partial a} = -\frac{2}{n}\sum_{i=1}^{n}x_i\left(y_i - ax_i - b\right) = 0. \tag{6.52}$$

Из уравнения (6.51) после раскрытия скобок получим

$$b = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{a}{n}\sum_{i=1}^{n} x_i,$$

или

$$b = \tilde{Y} - a\tilde{X}, \qquad (6.53)$$

где $\tilde{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$; $\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

После раскрытия скобок в уравнении (6.52) получим

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{a}{n}\sum_{i=1}^{n} x_i^2 - \frac{b}{n}\sum_{i=1}^{n} x_i = 0.$$

Подставляя в данное уравнение значение b из формулы (6.53), получим

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \tilde{X}\tilde{Y} - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\tilde{X}_i)^2\right) = 0. \qquad (6.54)$$

Первые два слагаемых в уравнении (6.54) равны выборочному корреляционному моменту K_{xy}. Слагаемые в скобках равны выборочной дисперсии S_x^2. С учетом этих подстановок решение уравнения (6.54) можно записать в виде:

$$\hat{a} = \frac{K_{xy}}{S_x^2} = \sqrt{\frac{S_y^2}{S_x^2}} r_{xy}. \qquad (6.55)$$

Подставляя формулу (6.55) в (6.53) для оценки параметра b модели (6.49) получим формулу

$$\hat{b} = \tilde{Y} - \sqrt{\frac{S_y^2}{S_x^2}} r_{xy} \tilde{X}. \qquad (6.56)$$

Описанная выше методика определения оптимальных значений параметров a и b линейной модели (6.49) называется ***методом наименьших квадратов***.

После вычисления оценок $\hat{a}$ и $\hat{b}$ по формулам (6.55) и (6.56) аналитическую зависимость между случайными величинами X и Y можно записать в виде

$$\hat{y}_i = \hat{a}x_i + \hat{b}, \qquad (6.57)$$

где $\hat{y}_i$ — оценка значений случайной величины Y по известным выборочным значениям случайной величины X.

В математической статистике доказана теорема Гаусса — Маркова, в соответствии с которой оценки $\hat{a}$ и $\hat{b}$, полученные по методу наименьших квадратов (МНК), имеют наименьшую дисперсию в классе всех линейных несмещенных осценок.

Для оценки точности регрессионной модели (6.57) вычислим математическое ожидание и дисперсию ошибок регрессии e_i:

$$e_i = y_i - \hat{y}_i = y_i - \hat{a}x_i - \hat{b}. \tag{6.58}$$

Усредняя уравнение (6.58) по объему выборки для выборочного среднего ошибок регрессии e_i, получим

$$\tilde{e}_i = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{1}{n}\sum_{i=1}^{n} y_i - a\frac{1}{n}\sum_{i=1}^{n} x_i - b = \tilde{Y} - \hat{a}\tilde{X} - \hat{b}.$$

Подставляя в полученное равенство формулы (6.55) и (6.56) для $\hat{a}$ и $\hat{b}$, получим

$$\tilde{e}_i = \tilde{Y} - \frac{K_{xy}}{S_x^2}\tilde{X} - \tilde{Y} + \frac{K_{xy}}{S_x^2}\tilde{X} = 0. \tag{6.59}$$

Таким образом, выборочное среднее ошибок регрессии равно нулю. После возведения правой части равенства (6.58) в квадрат и усреднения по объему выборки выборочная дисперсия ошибок регрессии S_e^2 запишется в виде:

$$S_e^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - 2\hat{a}\frac{1}{n}\sum_{i=1}^{n} x_i y_i + \hat{a}^2\frac{1}{n}\sum_{i=1}^{n} x_i^2 -$$

$$-2\hat{b}\frac{1}{n}\sum_{i=1}^{n} y_i + 2\hat{a}\hat{b}\frac{1}{n}\sum_{i=1}^{n} x_i + \hat{b}^2 = S_y^2 + \tilde{Y}^2 -$$

$$-2\hat{a}\left[K_{xy} + \tilde{X}\tilde{Y}\right] + \hat{a}^2\left[S_x^2 + \tilde{X}^2\right] - 2\hat{b}\tilde{Y} + 2\hat{a}\hat{b}\tilde{X} + \hat{b}^2 \tag{6.60}$$

При подстановке в формулу (6.60) значений оптимальных оценок параметров $\hat{a}$ и $\hat{b}$ из формул (6.55) и (6.56) после несложных преобразований для выборочной дисперсии ошибок регрессии получим

$$S_e^2 = S_y^2(1 - r_{xy}^2). \tag{6.61}$$

Из формулы (6.61) следует, что дисперсия ошибок регрессии S_e^2 определяется выборочной дисперсией случайной величины Y и коэффициентом корреляции случайных величин X и Y. При $r_{xy} = 1$ между случайными величинами X и Y существует детерминированная линейная зависимость и поэтому дисперсия ошибок регрессии будет равна нулю. При $r_{xy} = 0$ случайные величины X и Y некоррелированы и дисперсия ошибок регрессии будет максимальной и равной S_y^2, так как при $r_{xy} = 0$ аналитическая зависимость (6.57) преобразуется к виду $\hat{y}_i = \tilde{Y}$.

В математической статистике для оценки точности регрессионных моделей часто используют показатель, называемый коэффициентом детерминации R^2. Для определения коэффициента детерминации рассмотрим выборочную дисперсию случайной величины Y:

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{Y})^2. \qquad (6.62)$$

С учетом оценки, получаемой из регрессионной модели (6.57), правую часть равенства (6.62) можно записать в виде:

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \tilde{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}e_i^2 + \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{Y})^2 - 2\frac{1}{n}\sum_{i=1}^{n}e_i(\hat{y}_i - \tilde{Y}). \qquad (6.63)$$

Так как ошибки регрессии e_i и выборочные значения y_i являются некоррелированными случайными величинами, можно легко показать, что третье слагаемое в формуле (6.63) равно нулю. С учетом этого для выборочной дисперсии S_y^2 получим

$$S_y^2 = S_e^2 + \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{Y})^2. \qquad (6.64)$$

Из формулы (6.64) следует, что выборочная дисперсия случайной величины Y состоит из двух слагаемых. Первое слагаемое S_e^2 обусловлено случайными отклонениями выборочных значений y_i от уравнения регрессии (6.57). Второе слагаемое равно среднему квадрату отклонения детерминированной функции (6.57) от выборочного среднего $\tilde{Y}$ и является как бы объяснимой долей дисперсии S_y^2.

Коэффициентом детерминации R^2 называется отклонение объяснимой доли дисперсии к общей дисперсии случайной величины Y:

$$R^2 = \frac{1}{S_y^2 n}\sum_{i=1}^{n}(\hat{y}_i - \tilde{y})^2.$$

В соответствии с формулой (6.64) коэффициент детерминации может быть определен соотношением

$$R^2 = 1 - \frac{S_e^2}{S_y^2}. \tag{6.65}$$

Рассмотрим два случая.

1. Предположим, что в качестве регрессионной модели мы выбрали функцию $\hat{y}_i = \tilde{Y}$. В данной модели нет зависимости y_i от x_i, т.е. никакого предсказания значений y_i по значениям x_i эта модель не дает. В этом случае второе слагаемое в выражении (6.64) будет равно нулю и коэффициент детерминации также будет равен нулю.

2. Предположим, что в качестве регрессионной модели мы подобрали некоторую сложную нелинейную функцию $\hat{y}_i = f(x)$ так, что все выборочные значения x_i, y_i находятся на этой линии регрессии. В этом случае разность $y_i - \hat{y}_i = e_i$ будет равна нулю для всех i-х значений выборки, поэтому $S_e^2 = 0$ и коэффициент детерминации будет равен единице.

В исследовательской практике использование очень сложных регрессионных моделей нецелесообразно, поэтому значение коэффициента детерминации находится в пределах $0 < R^2 < 1$.

Из условия (6.50), составляющего суть метода наименьших квадратов, следует, что оптимальные значения параметров $\hat{a}; \hat{b}$ уравнения регрессии обеспечивают минимальное значение дисперсии ошибок регрессии ε_i. Таким образом, в рамках выбранной модели уравнения регрессии определение параметров модели по методу наименьших квадратов обеспечивает максимальное значение коэффициента детерминации.

Для линейной модели уравнения регрессии после подстановки формулы (6.61) в (6.65) получим

$$R^2 = 1 - \frac{S_y^2(1 - r_{xy}^2)}{S_y^2} = r_{xy}^2.$$

Иными словами, для линейной модели уравнения регрессии максимальное значение коэффициента детерминации равно квадрату выборочного коэффициента корреляции случайных величин X и Y.

6.6. Примеры решения задач

Задача 6.1. Результаты наблюдения за числом сделок x_i, совершаемых инвесторами ($n = 1000$) на фондовой бирже за месяц, приведены в табл. 6.12.

Таблица 6.12

Результаты наблюдения

Число сделок x_i	0	1	2	3	4	5	6	7
Число инвесторов n_i	131	275	265	184	93	34	13	5
P_i^*	0,131	0,275	0,265	0,184	0,093	0,034	0,013	0,005

Построить полигон относительных частот числа совершаемых сделок. По критерию согласия хи-квадрат при уровне значимости $\beta = 0{,}05$ проверить гипотезу о том, что случайная величина x имеет распределение Пуассона.

Решение. Из условия задачи следует, что случайная величина x является дискретной случайной величиной, принимающей целочисленные значения $x \geq 0$.

Вычислим относительные частоты значений случайной величины x_i, равные эмпирическим значениям вероятностей, по формуле (6.1):

$$P_i^* = P^*(x_i) = \frac{n_i}{n},$$

где $n = \sum_{i=0}^{7} n_i = 1000.$

Результаты вычисления сведем в табл. 6.13 (строка 2). На рисунке 6.9 приведен полигон относительных частот (эмпирических значений вероятностей) для значений случайной величины x_i. При правильных вычислениях сумма вероятностей P_i^* по строке 2 должна равняться единице.

Проверим гипотезу о том, что закон распределения исследуемой случайной величины x может быть описан законом Пуассона:

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda},$$

где $x = 0, 1, 2, 3, \ldots$; λ — параметр распределения равен математическому ожиданию $\lambda = m_x$.

Для определения значения параметра λ закона Пуассона вычислим выборочное среднее значение случайной величины x (6.9):

$$m_x^* = \sum_{i=0}^{7} x_i P_i^*.$$

Занесем произведения $x_i P_i^*$ в строку 3 таблицы 6.13, тогда сумма по строке 3 даст выборочное значение математического ожидания $m_x^* = 2{,}012$. Отсюда значение параметра λ теоретического закона распределения можно принять равным:

$$\lambda = m_x^* \approx 2.$$

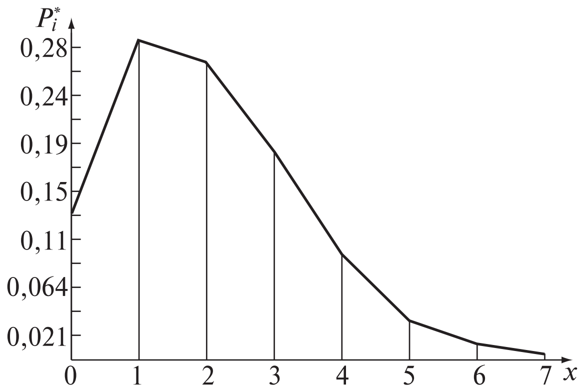


Рис. 6.9. Полигон относительных частот

Из таблицы приложения 4 при λ = 2 в табл. 6.13 записываем вероятности P_i, определяемые по закону Пуассона для значений случайной величины x_i.

Для проверки заданной в условии задачи гипотезы вычислим экспериментальное значение хи-квадрат по формуле (6.27):

$$\chi^2_{\text{эксп}} = n \sum_{i=0}^{7} \frac{\left(P_i - P_i^*\right)^2}{P_i^*}.$$

Значения числителя дроби под знаком суммы приведены в строке 5, а значения слагаемых — в строке 6 табл. 6.13, отсюда для экспериментального значения хи-квадрат получим:

$$\chi^2_{\text{эксп}} = 1000 \cdot 149{,}5 \cdot 10^{-5} = 1{,}495.$$

Определим число степеней свободы χ^2-распределения:

$$\nu = k - s - 1,$$

где $k = 8$ — число значений случайной величины; $s = 1$ — число параметров теоретического закона распределения.

Таблица 6.13

Результаты вычислений

x_i	0	1	2	3	4	5	6	7	Σ
P_i^*	0,131	0,275	0,265	0,184	0,093	0,034	0,013	0,005	1
$x_i P_i^*$	0	0,275	0,53	0,552	0,372	0,17	0,078	0,035	2,012
P_i	0,135	0,271	0,271	0,18	0,09	0,036	0,012	0,003	0,998
$(P_i - P_i^*)^2$	16×10^{-6}	16×10^{-6}	36×10^{-6}	16×10^{-6}	9×10^{-6}	$4{,}0 \times 10^{-6}$	1×10^{-6}	4×10^{-6}	—
$\frac{(P_i - P_i^*)^2}{P_i^*}$	$12{,}2 \times 10^{-5}$	$5{,}8 \times 10^{-5}$	$13{,}6 \times 10^{-5}$	$8{,}7 \times 10^{-5}$	$9{,}7 \times 10^{-5}$	$11{,}8 \times 10^{-5}$	$7{,}7 \times 10^{-5}$	80×10^{-5}	$149{,}5 \times 10^{-5}$

При заданном уровне значимости $\beta = 0{,}05$ и числе степеней свободы $\nu = 8 - 1 - 1 = 6$ находим критическое значение χ^2-распределения по таблице приложения 2:

$$\chi^2_{\text{кр}}(\nu = 6; \beta = 0{,}05) = 12{,}59.$$

Так как $\chi^2_{\text{эксп}} < \chi^2_{\text{кр}}$, то гипотеза о том, что случайная величина χ имеет закон распределения Пуассона, принимается.

Задача 6.2. Результаты анализа производительности труда Y $\left(\frac{\text{ед. прод.}}{\text{ч}}\right)$ и уровня механизации работ X (%) для $n = 15$ однотипных предприятий приведены в табл. 6.14.

Таблица 6.14

Производительность труда и уровень механизации работ

x_i	30	32	36	40	41	47	50	54	55	56	60	61	67	69	76
y_i	24	20	28	30	31	33	34	37	40	34	38	41	43	45	48

Выбрать аналитическую зависимость между величинами X и Y (уравнение регрессии), определить оптимальные значения параметров выбранной зависимости и оценить ее точность.

Решение. Из условия задачи ясно, что случайную величину X следует выбрать в качестве независимой переменной, а Y — зависимой переменной. На рисунке 6.10 графически изображено расположение точек $[x_i; y_i]$, соответствующее значениям этих случайных величин, приведенных в табл. 6.14. Из рисунка 6.10 видна тенденция увеличения производительности труда Y при увеличении уровня механизации работ. Эта зависимость между величинами X и Y может быть отображена линейной зависимостью (уравнением регрессии):

$$Y = aX + b.$$

В рамках линейного уравнения регрессии оптимальные (по критерию минимума среднего квадрата ошибки) значения параметров $\hat{a}$ и $\hat{b}$ определяются формулами (6.55) и (6.56). Вычислим выборочные значения дисперсии S_x^2 и корреляционного момента K_{xy} по формулам:

$$K_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \tilde{X}\tilde{Y};$$

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\tilde{X})^2.$$

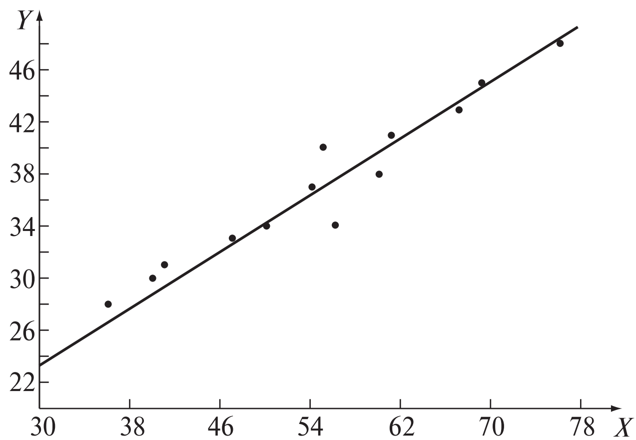


Рис. 6.10. Уровень регрессии

Вычислим выборочные средние значения случайных величин $\tilde{X}$ и $\tilde{Y}$:

$$\tilde{X} = \frac{1}{n}\sum_{i=1}^{n} x_i =$$

$$= \frac{30+32+36+40+41+47+50+54+55+56+60+61+67+69+76}{15} \approx 51{,}6;$$

$$\tilde{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i =$$

$$= \frac{24+20+28+30+31+33+34+37+40+34+38+41+43+45+48}{15} \approx 35{,}07;$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i y_i =$$

$$= \frac{720+640+1008+1200+1271+1551+1700+1998+2200+1904+2280}{15} +$$

$$+ \frac{2501+2881+3105+3648}{15} = \frac{28\,607}{15} \approx 1907{,}13.$$

Для выборочных значений S_x^2 и K_{xy} получим:

$$K_{xy} = 1907{,}13 - 51{,}6 \cdot 35{,}07 \approx 97{,}6;$$

$$S_x^2 =$$

$$= \frac{900+1024+1296+1600+1681+2209+2500+2916+3025+3136+3600}{15} +$$

$$+ \frac{3721+4489+4761+5776}{15} - (51{,}6)^2 = \frac{42\,634}{15} - 2662{,}56 \approx 179{,}7.$$

По формуле (6.55) находим значение углового коэффициента $\hat{a}$ уравнения регрессии

$$\hat{a} = \frac{K_{xy}}{S_x^2} = \frac{97{,}6}{179{,}7} \approx 0{,}543.$$

По формуле (6.56) находим значение параметра $\hat{b}$:

$$\hat{b} = \tilde{Y} - \hat{a}\tilde{X} = 35{,}07 - 0{,}543 \cdot 51{,}6 \approx 7{,}05.$$

Таким образом, аналитическая зависимость между величинами X и Y может быть описана уравнением регрессии

$$Y = 0{,}543X + 7{,}05.$$

График данной линейной зависимости приведен на рис. 6.10.

Точность полученного уравнения регрессии оценим по коэффициенту детерминации R^2. В подразделе 6.5.2 показано, что для линейной модели уравнения регрессии значение коэффициента детерминации равно квадрату выборочного коэффициента корреляции:

$$R^2 = r_{xy}^2 = \frac{K_{xy}^2}{S_x^2 \cdot S_y^2}.$$

Вычислим выборочное значение дисперсии случайной величины Y:

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - (\tilde{Y}) =$$

$$= \frac{576+400+784+900+961+1089+1156+1369+1600+1156}{15} +$$

$$+ \frac{1444+1681+1849+2025+2304}{15} - (35{,}07)^2 = \frac{19\,292}{15} - 1229{,}9 \approx 56{,}2.$$

Для коэффициента детерминации R^2 получим:

$$r_{xy}^2 = R^2 = \frac{(97{,}6)^2}{179{,}7 \cdot 56{,}2} \approx 0{,}943.$$

Для оценки точности полученной регрессионной модели вычислим дисперсию ошибок регрессии S_e^2 по формуле (6.61):

$$S_e^2 = S_y^2(1 - r_{xy}^2) = 56{,}2(1 - 0{,}943) \approx 3{,}2.$$

Так как значение коэффициента детерминации R^2 близко к единице, то полученное уравнение регрессии достаточно точно отражает зависимость случайных величин X и Y.

Вопросы для самоконтроля

1. Раскройте основные понятия и основные задачи математической статистики.
2. Что такое выборочная совокупность, упорядоченная выборка? Как осуществляется построение функции распределения при малом объеме выборки?
3. Как осуществляется построение сгруппированной выборки при выборочной совокупности большого объема?
4. Как осуществляется построение гистограммы распределения по сгруппированной выборке?
5. Как осуществляется построение функции распределения по сгруппированной выборке и гистограмме распределения?
6. Как определяются точечные оценки параметров распределения при малом объеме выборки?
7. Как определяются точечные оценки параметров распределения по сгруппированной выборке?
8. Каковы условия несмещенности, состоятельности и эффективности точечных оценок параметров распределения? Что такое исправленная выборочная дисперсия?

9. Каков алгоритм точечной оценки параметров теоретического закона распределения?
10. Что такое проверка гипотезы о теретическом законе распределения случайной величины по критерию согласия χ^2-распределения?
11. Как проводится оценка выборочного корреляционного момента двух случайных величин при малом объеме выборки?
12. Как строится вариационный ряд системы двух случайных величин при большом объеме выборки?
13. Как определяются оценка выборочного коэффициента корреляции по вариационному ряду системы двух случайных величин?
14. Как осуществляется проверка значимости коэффициента корреляции по критерию *t*-статистики Стьюдента?
15. Что такое регрессионный анализ двух случайных величин? Как проводится оценка параметров линейного уравнения регрессии по методу наименьших квадратов?
16. Как осуществляется определение выборочного среднего и выборочной дисперсии ошибок регрессии?
17. Как проводится сравнение моделей уравнения регрессии двух случайных величин по коэффициенту детерминации?

Список литературы

1. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика. Основы моделирования и первичная обработка данных. М. : Финансы и статистика, 1983.
2. *Бочаров П.П., Печенкин А.В.* Теория вероятностей и математическая статистика. М. : Гардарика, 1998.
3. *Вентцель Е.С.* Теория вероятностей : учеб. пособие. М. : Высш. шк., 1998.
4. *Вентцель Е.С., Овчаров Л.А.* Теория случайных процессов и ее инженерные приложения. М. : Наука, 1991.
5. *Гмурман В.Е.* Теория вероятностей и математическая статистика : учеб. пособие. 12-е изд., перераб. М. : Высшее образование, 2007.
6. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие. 11-е изд., перераб. М. : Высшее образование, 2006.
7. *Гнеденко Б.В.* Курс теории вероятностей. М. : Наука, 1971.
8. *Гурский Е.Н.* Теория вероятностей с элементами математической статистики. М. : Высш. шк., 1971.
9. *Иванова В.М., Калинина В.Н.* Математическая статистика. М. : Высшая школа, 1981.
10. *Карлов А.М.* Теория вероятностей и математическая статистика для экономистов : учеб. пособие. Калининград : БИЭФ, 2005.
11. *Ковалев В.В.* Сборник задач по финансовому анализу : учеб. пособие. М. : Финансы и статистика, 2003.
12. *Колемаев В.А., Староверов О.В., Турундаевский В.В.* Теория вероятностей и математическая статистика. М. : Высш. шк., 1991.
13. *Кремер Н.Ш.* Математическая статистика. М. : Энциклопедическое образование, 1992.
14. *Кремер Н.Ш.* Теория вероятностей и математическая статистика : учебник для вузов. М. : ЮНИТИ-ДАНА, 2003.
15. *Львовский Е.Н.* Статистические методы построения эмпирических формул. М. : Высш. шк., 1982.
16. *Митропольский А.К.* Техника статистических вычислений. М. : Физматгиз, 1971.
17. *Мхитарян В.С.* Теория вероятностей и прикладная статистика. М. : ЮНИТИ-ДАНА, 2001.
18. *Пустыльник Е.И.* Статистические модели анализа и обработки данных. М. : Наука, 1968.
19. *Репина О.В., Мелетиев А.Н.* Стохастические модели рынка транспортных услуг. М. : Экономический факультет МГУ : ТЕИС, 2003.
20. *Розанов Ю.А.* Лекции по теории вероятностей. М. : Наука, 1986.

21. *Свешников А.А.* Прикладные методы теории случайных функций. М. : Наука, 1968.
22. *Смирнов Н.В., Дунин-Барковский И.В.* Курс теории вероятностей и математической статистики для технических приложений : учеб. пособие. М. : Наука, 1969.
23. Теория вероятностей и математическая статистика в задачах : учеб. пособие / В.А. Ватутин, Г.И. Ивченко, Ю.И. Медведев [и др.]. 3-е изд., испр. М. : Дрофа, 2005.
24. *Тюрин Ю.Н., Макаров Л.А.* Статистический анализ данных на компьютерах / под ред. В.Э. Фигурнова. М. : ИНФРА-М, 1998.
25. *Хан Г., Шапиро С.* Статистические модели в инженерных задачах. М. : Мир, 1969.
26. *Чистяков В.П.* Курс теории вероятностей. 4-е изд. М. : Агар, 1996.
27. *Черняк В.З.* Сборник задач по экономике. М. : Экзамен, 2008.
28. *Юсупов Р.А.* Теория вероятностей и математическая статистика : учеб. пособие. Астрахань : Изд-во АГТУ, 2000.
29. *Яковлева А.В.* Теория вероятностей и математическая статистика : ответы на экзаменационные вопросы : учеб. пособие. М. : Экзамен, 2006.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица интеграла вероятности $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$ **для** $0,00 \le z \le 4,99$; $\Phi(-z)=1-\Phi(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0 7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,90147

Продолжение

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0 96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,920097*	0,920358	0,920613	0,920863	0,921106	0,92344	0,921576
2,4	0,921802	0,922024	0,922240	0,922451	0,922656	0,922857	0,923053	0,923244	0,923431	0,923613
2,5	0,923790	0,923963	0,924132	0,924297	0,924457	0,924614	0,924766	0,924915	0,925060	0,925201
2,6	0,925339	0,925473	0,925604	0,925731	0,925885	0,925975	0,926093	0,926207	0,926319	0,926427
2,7	0,926533	0,926636	0,926736	0,926833	0,926928	0,927020	0,927110	0,927197	0,927282	0,927365
2,8	0,927445	0,927523	0,927599	0,927673	0,927744	0,927814	0,927882	0,927948	0,928012	0,928074
2,9	0,928134	0,928193	0,928250	0,928305	0,928359	0,928411	0,928462	0,928511	0,928559	0,928605
3,0	0,928650	0,928694	0,928736	0, 928777	0,928817	0,928856	0,928893	0,928930	0,928965	0,928999
3,1	0,930324**	0,930646	0,930957	0,931260	0,931553	0,931836	0,932112	0,932378	0,932636	0,932886

Окончание

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,2	0,933129	0,933363	0,933590	0,933810	0,934024	0,934230	0,934429	0,934623	0,934810	0,934991
3,3	0,935166	0,935335	0,935499	0,935658	0,935811	0,935959	0,936103	0,936242	0,936376	0,936505
3,4	0,936631	0,936752	0,936869	0 936982	0,937091	0,937197	0,937299	0,937398	0,937493	0,937585
3,5	0,937674	0,937759	0,937842	0,937922	0,937999	0,938074	0,938146	0,938215	0,938282	0,938347
3,6	0,938409	0,938469	0,938527	0,938583	0,938637	0,938689	0,938739	0,938787	0,938834	0,938879
3,7	0,938922	0,938964	0,940039	0,940426	0,940799	0,941158	0,941504	0,941838	0,942159	0,942468
3,8	0,942765	0,943052	0,943327	0,943593	0,943848	0,944094	0,944331	0,944558	0,944777	0,944988
3,9	0,945190	0,945385	0,945573	0,945753	0,945926	0,946092	0,946253	0,946406	0,946554	0,946696
4,0	0,946833	0,946964	0,947090	0,947211	0,947327	0,947439	0,947546	0,947649	0,947748	0,947843
4,1	0,947934	0,948022	0,948106	0,948186	0,948263	0,948338	0,948409	0,948477	0,948542	0,948605
4,2	0,948665	0,948723	0,948778	0,948832	0,948882	0,948931	0,948978	0,950226	0,950655	0,951066
4,3	0,951460	0,951837	0,952199	0,952545	0,952876	0,953193	0,953497	0,953788	0,954066	0,954332
4,4	0,954587	0,954841	0,955065	0,955288	0,955502	0,955706	0,955902	0,956089	0,956268	0,956439
4,5	0,956602	0,956759	0,956908	0,957051	0,957187	0,957318	0,957442	0,957561	0,957675	0,957784
4,6	0,957888	0,957987	0,958081	0,958172	0,958258	0,958340	0,958419	0,958494	0,958566	0,958634
4,7	0,958699	0,958761	0,958821	0,958877	0,958931	0,958983	0,960320	0,960789	0,961235	0,961661
4,8	0,962067	0,962453	0,962822	0,963173	0,963508	0,963827	0,964131	0,964420	0,964696	0,964958
4,9	0,965208	0,965446	0,965673	0,965889	0,966094	0,966289	0,966475	0,966652	0,966821	0,966981

Пример: Ф (3,57) = 0,938215 = 0,9998215.

* 0,920097 = 0,990097;

** 0,930324 = 0,990324.

Приложение 2

Процентные точки распределения $\chi^2 : P\left[\chi^2 > \chi^2_{\nu;\beta}\right] = \beta$

	β							
ν	0,990	0,975	0,950	0,900	0,10	0,05	0,025	0,010
1	0,00016	0,00098	0,0039	0,0158	2,71	3,84	5,02	6,63
2	0,0201	0,0506	0,103	0,211	4,61	5,99	7,38	9,21
3	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,34
4	0,297	0,484	0,711	1,06	7,78	9,49	11,14	13,28
5	0,554	0,831	1,15	1,61	9,24	11,07	12,83	15,09
6	0,872	1,24	1,64	2,20	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21
11	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,73
12	4,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00
17	6,41	7,56	8,67	10,08	24,77	27,59	30,19	33,41
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,88	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89
40	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,69
60	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38
120	86,92	91,58	95,70	100,62	140,23	146,57	152,21	158,95

Приложение 3

Процентные точки *t*-распределения Стьюдента $P\left[|T| > t_{\nu;\,\beta}\right] = \beta$

	β				
ν	0,2	0,1	0,05	0,02	0,01
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1 337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,731	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2 831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,39	2,58

Приложение 4

Значение функции Пуассона $P(X=m)=\frac{1}{m!}\lambda^m e^{-\lambda}$

m	λ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812	0,496585	0,449329	0,406570
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287	0,347610	0,359463	0,365913
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786	0,121663	0,143785	0,164661
3	0,000151	0,001092	0,003334	0,007150	0,012636	0,019757	0,028388	0,028388	0,049398
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964	0,004968	0,007669	0,011115
5		0,000002	0,000015	0,000057	0,000158	0,000356	0,000696	0,001227	0,002001
6			0,000001	0,000004	0,000013	0,000036	0,000081	0,000164	0,000300
7					0,000001	0,000003	0,000008	0,000019	0,000039
8							0,000001	0,000002	0,000004

m	λ								
	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	0,367879	0,135335	0,049787	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,367879	0,270671	0,149361	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111
2	0,183940	0,270671	0,224042	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998
3	0,061313	0,180447	0,224042	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994
4	0,015328	0,090224	0,168031	0,195367	0,175467	0,133853	0,091226	0,057252	0,033737
5	0,003066	0,036089	0,100819	0,156293	0,175467	0,160623	0,127717	0,091604	0,060727
6	0,000511	0,012030	0,050409	0,104196	0,146223	0,160623	0,149003	0,122138	0,091090
7	0,000073	0,003437	0,021604	0,059540	0,104445	0,136777	0,149003	0,139587	0,117116
8	0,000009	0,000859	0,008102	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756
9	0,000001	0,000191	0,002701	0,013231	0,036266	0,068838	0,101405	0,124077	0,131756

Окончание

m	λ								
	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0
10	0,000038	0,000810	0,005292	0,018133	0,041303	0,070983	0,099262	0,118580	0,125110
11	0,000007	0,000221	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020	0,113740
12	0,000001	0,000055	0,000642	0,003434	0,011264	0,026350	0,048127	0,072765	0,094780
13		0,000013	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376	0,072908
14		0,000003	0,000056	0,000472	0,002228	0,007094	0,016924	0,032384	0,052077
15		0,000001	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431	0,034718
16			0,000004	0,000049	0,000334	0,001448	0,004513	0,010930	0,021699
17			0,000001	0,000014	0,000118	0,000596	0,002124	0,005786	0,012764
18				0,000004	0,000039	0,000232	0,000944	0,002893	0,007091
19				0,000001	0,000012	0,000085	0,000397	0,001370	0,003732
20					0,000004	0,000030	0,000159	0,000617	0,001866
21					0,000001	0,000010	0,000061	0,000264	0,000889
22						0,000003	0,000022	0,000108	0,000404
23						0,000001	0,000008	0,000042	0,000176
24							0,000003	0,000016	0,000073
25							0,000001	0,000006	0,000029
26								0,000002	0,000011
27								0,000001	0,000004
28									0,000001

Приложение 5

Экономические показатели развития регионов Российской Федерации (1997 г.)

№ п/п	Республика, область	ВРП*	Количество предприятий, тыс.	№ п/п	Республика, область	ВРП*	Количество предприятий, тыс.
1	Республика Ингушетия	3,1	6,5	15	Курганская область	8,2	18,0
2	Республика Дагестан	4,4	21,9	16	Алтайский край	8,2	38,3
3	Республика Северная Осетия — Алания	5,1	12,3	17	Пензенская область	8,3	16,7
4	Республика Калмыкия	5,6	9,9	18	Псковская область	8,4	11,8
5	Республика Адыгея	5,7	7,3	19	Брянская область	8,4	14,3
6	Республика Тыва	5,8	6,0	20	Чувашская Республика	8,5	13,9
7	Еврейская автоном. область	6,3	2,9	21	Калининградская область	9,0	23,3
8	Карачаево-Черкесская Республика	6,3	8,2	22	Тульская область	9,2	20,7
9	Кабардино-Балкарская Республика	6,9	10,2	23	Владимирская область	9,3	19,7
10	Ивановская область	7,1	15,2	24	Ставропольский край	9,6	51,1
11	Тамбовская область	7,3	14,0	25	Краснодарский край	9,7	92,6
12	Республика Алтай	7,3	4,3	26	Орловская область	9,8	11,2
13	Ростовская область	7,9	72,8	27	Республика Мордовия	9,8	11,2
14	Республика Мари-Эл	8,1	9,1	28	Тверская область	9,9	23,9

Окончание

29	Читинская область	9,9	13,5	54	Нижегородская область	14,3	42,1
30	Калужская область	10,0	16,5	55	Ярославская область	14,7	20,9
31	Воронежская область	10,3	28,9	56	Московская область	14,8	89,3
32	Смоленская область	10,4	16,0	57	Амурская область	15,2	13,3
33	Новгородская область	10,5	11,9	58	Вологодская область	15,5	16,9
34	Кировская область	10,7	21,7	59	Омская область	15,5	36,6
35	Астраханская область	10,9	13,7	60	Республика Башкортостан	15,7	47,0
36	Рязанская область	11,0	20,2	61	Свердловская область	15,8	68,7
37	Республика Бурятия	11,0	14,8	62	г. Санкт-Петербург	15,9	127,9
38	Костромская область	11,1	12,7	63	Кемеровская область	16,1	36,1
39	Ульяновская область	11,1	16,0	64	Пермская область	17,2	34,2
40	Курская область	11,5	16,3	65	Республика Татарстан	17,8	40,0
41	Ленинградская область	11,6	22,0	66	Мурманская область	18,6	14,4
42	Саратовская область	11,7	42,9	67	Томская область	19,8	19,9
43	Волгоградская область	12,0	45,8	68	Хабаровский край	20,2	23,0
44	Белгородская область	12,3	20,1	69	Иркутская область	20,2	37,4
45	Липецкая область	12,6	12,3	70	Камчатская область	20,4	8,9
46	Архангельская область	12,8	15,5	71	Красноярский край	21,2	41,6
47	Республика Карелия	12,9	12,2	72	Сахалинская область	21,3	12,9
48	Республика Удмуртия	13,5	22,4	73	Самарская область	21,9	50,6
49	Республика Хакасия	13,7	7,5	74	Республика Коми	23,3	15,7
50	Оренбургская область	13,7	29,9	75	Магаданская область	25,8	8,0
51	Приморский край	13,7	36,6	76	Чукотский автоном. округ	28,7	1,2
52	Челябинская область	14,0	47,8	77	Республика Саха-Якутия	29,7	17,5
53	Новосибирская область	14,2	54,5	78	г. Москва	37,1	480

* ВРП — валовой региональный продукт, тыс. руб. на душу населения.