И. И. Баврин

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

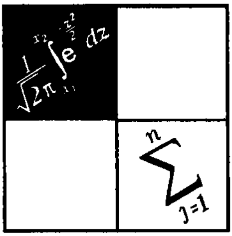
и математическая статистика

$$\frac{1}{2\pi}\int_{x_1}^{x_2} e^{-\frac{z^2}{2}}\,dz$$

$$\frac{p(H_i)\,P_{H_i}(A)}{\sum_{j=1}^{n} p(H_j)\,P_{H_j}(A)}$$

$$p(H_i)=\frac{p(H_i)\,P_{H_i}(A)}{p(A)}$$

$$\int_{x_1}^{x_2}$$



И.И. Баврин

ТЕОРИЯ

ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Допущено Министерством образования
Российской Федерации в качестве учебника
для студентов высших педагогических учебных заведений,
обучающихся по направлению «Естественнонаучное образование»
и специальностям «Математика»,
«Физика», «Химия», «Биология» и «География»

Москва
«Высшая школа»
2005

УДК 519.2
ББК 22.17
Б13

Рецензенты

кафедра высшей математики Московского государственного открытого педагогического университета им М А Шолохова (зав кафедрой — д-р пед наук, академик Академии информатизации образования *А И Нижников*), д-р физ -мат наук, проф *В И Гаврилов* (Московский государственный университет им М В Ломоносова)

Баврин, И. И.
Б13 Теория вероятностей и математическая статистика: Учебник/И. И. Баврин.— М.: Высш. шк., 2005.— 160 с.: ил.

ISBN 5-06-005322-9

Изложены основы теории вероятностей и математической статистики в приложении к физике, химии, биологии, географии, экологии, приведены упражнения для самостоятельной работы Все основные понятия и положения иллюстрируются разобранными примерами и задачами

Для студентов естественнонаучных специальностей педагогических вузов Может быть использован студентами других вузов

УДК 519 2
ББК 22 17

ISBN 5-06-005322-9

ПРЕДИСЛОВИЕ

Часто приходится изучать явления, для которых практически трудно или принципиально невозможно отыскать все причины, порождающие их, и тем более количественно их выразить. Такие явления невозможно описать функционально.

Например, при бросании монеты нельзя предсказать, какой стороной она упадет: для этого необходимо было бы учесть слишком много различных факторов: работу мышц руки, участвующей в бросании, малейшие отклонения в распределении массы монеты, движение воздуха и т. д. Результат бросания монеты случаен. Но, оказывается, при достаточно большом числе бросаний монеты существует определенная закономерность (герб и цифра выпадут приблизительно поровну).

Закономерности, которым подчиняются случайные события, изучаются в разделах математики, которые называются теорией вероятностей и математической статистикой.

Методы теории вероятностей и математической статистики широко применяются в естествознании, технике, экономике и других областях.

Настоящее издание ставит своей целью изложение основ теории вероятностей и математической статистики и их приложений. Оно написано в соответствии с действующими программами и с учетом профессиональной направленности будущих учителей специальностей «Математика», «Физика», «Биология», «Химия» и «Информатика». В нем особое внимание уделено понятиям и методам, имеющим прикладное значение. Это отражено в большом числе разобранных примеров и задач из естественнонаучных дисциплин.

Книга состоит из четырех глав и приложения. В ней рассмотрены случайные события (глава I), случайные величины, включая двумерные величины (главы II и III), элементы математической статистики (глава IV). В приложении приведены соответствующие таблицы, необходимые для выполнения упражнений, в том числе для удобства читателя, и таблицы производных и интегралов.

К каждой главе составлены упражнения для первоначальной самостоятельной работы студентов, а в конце книги приведены дополнительные упражнения, позволяющие закрепить полученные знания. В упражнениях, там где в этом есть необходимость, сразу после текста даны ответы (в квадратных скобках).

Автор

ГЛАВА I

СЛУЧАЙНЫЕ СОБЫТИЯ И ВЕРОЯТНОСТИ

§ 1.1. Случайные события. Классическое определение вероятности

1. Понятие о случайном событии. Опыт, эксперимент, наблюдение явления называют *испытанием*. Испытаниями, например, являются: бросание монеты, выстрел из винтовки, бросание игральной кости (кубика с нанесенным на каждую грань числом очков — от одного до шести).

Результат, исход испытания называется *событием*. Событиями являются: выпадение герба или цифры, попадание в цель или промах, появление того или иного числа очков на брошенной игральной кости.

Для обозначения событий используются большие буквы латинского алфавита: A, B, C и т. д.

Определение 1. Два события называются *совместимыми*, если появление одного из них не исключает появление другого в одном и том же испытании.

Пример 1. Испытание: однократное бросание игральной кости. Событие A — появление четырех очков, событие B — появление четного числа очков. События A и B совместимые.

Определение 2. Два события называются *несовместимыми*, если появление одного из них исключает появление другого в одном и том же испытании.

Пример 2. Испытание: однократное бросание монеты. Событие A — выпадение герба, событие B — выпадение цифры. Эти события несовместимы, так как появление одного из них исключает появление другого.

Несовместимость более чем двух событий в данном испытании означает их попарную несовместимость.

Пример 3. Испытание: однократное бросание игральной кости. Пусть события A_1, A_2, A_3, A_4, A_5, A_6 — соответственно выпадение одного очка, двух, трех и т. д. Эти события являются несовместимыми.

Определение 3. Два события A и B называются *противоположными*, если в данном испытании они несовместимы и одно из них обязательно происходит.

Событие, противоположное событию A, обозначают через $\bar{A}$.

П р и м е р 4. Испытание: однократное бросание монеты. Событие A — выпадение герба, событие B — выпадение цифры. Эти события противоположны, так как исходами бросания могут быть лишь они и появление одного из них исключает появление другого, т. е. $A=\overline{B}$ или $\overline{A}=B$.

О п р е д е л е н и е 4. Событие называется *достоверным*, если в данном испытании оно является единственно возможным его исходом, и невозможным, если в данном испытании оно заведомо не может произойти.

П р и м е р 5. Испытание: извлечение шара из урны, в которой все шары белые. Событие A — вынут белый шар — достоверное событие; событие B — вынут черный шар — невозможное событие.

Заметим, что достоверное и невозможное события в данном испытании являются противоположными.

О п р е д е л е н и е 5. Событие A называется *случайным*, если оно объективно может наступить или не наступить в данном испытании.

П р и м е р 6. Событие A_6 — выпадение шести очков при бросании игральной кости — случайное. Оно может наступить, но может и не наступить в данном испытании.

П р и м е р 7. Событие A_{98} — прорастание девяноста восьми зерен пшеницы из ста — случайное. Это событие может наступить, но, может быть, прорастет зерен больше или меньше.

2. Алгебра событий.

О п р е д е л е н и е 1. *Суммой событий* A и B называется событие $C=A+B$, состоящее в наступлении по крайней мере одного из событий A или B.

П р и м е р 1. Испытание: стрельба двух стрелков (каждый делает по одному выстрелу). Событие A — попадание в мишень первым стрелком, событие B — попадание в мишень вторым стрелком. Суммой событий A и B будет событие $C=A+B$, состоящее в попадании в мишень по крайней мере одним стрелком.

Аналогично суммой конечного числа событий A_1, A_2, ..., A_k называется событие $A=A_1+A_2+...+A_k$, состоящее в наступлении хотя бы одного из событий A_i ($i=1$, ..., k).

Из определения 1 непосредственно следует, что $A+B=B+A$. Справедливо также и сочетательное свойство. Однако $A+A=A$ (а не $2A$, как в алгебре).

О п р е д е л е н и е 2. *Произведением событий* A и B называется событие $C=AB$, состоящее в том, что в результате испытания произошли и событие A, и событие B.

Аналогично произведением конечного числа событий A_1, A_2, ..., A_k называется событие $A=A_1 A_2 ... A_k$, состоящее в том, что в результате испытания произошли все указанные события.

В условиях предыдущего примера произведением событий A и B будет событие $C = AB$, состоящее в попадании в мишень двумя стрелками.

Из определения 2 непосредственно следует, что $AB = BA$.

Справедливы также сочетательный и дистрибутивный законы. Однако $AA = A$ (а не A^2).

3. Классическое определение вероятности. Всякое испытание влечет за собой некоторую совокупность исходов — результатов испытания, т. е. событий. Во многих случаях возможно перечислить все события, которые могут быть исходами данного испытания.

Определение 1. Говорят, что совокупность событий образует *полную группу событий* для данного испытания, если его результатом обязательно становится хотя бы одно из них.

Примеры полных групп событий — выпадение герба и выпадение цифры при одном бросании монеты; попадание в цель и промах при одном выстреле; выпадение одного, двух, трех, четырех, пяти и шести очков при одном бросании игральной кости.

Рассмотрим полную группу попарно несовместимых событий U_1, U_2, ..., U_n, связанную с некоторым испытанием. Предположим, что в этом испытании осуществление каждого из событий U_i ($i = 1, 2, ..., n$) равновозможно, т. е. условия испытания не создают преимуществ в появлении какого-либо события перед другими возможными.

Определение 2. События U_1, U_2, ..., U_n, образующие полную группу попарно несовместимых и равновозможных событий, называют *элементарными событиями.*

Пример 1. Вернемся к опыту с подбрасыванием игральной кости. Пусть U_i — событие, состоящее в том, что кость выпала гранью с цифрой i. Как уже отмечалось (п. 1, 3), события U_1, U_2, ..., U_6 образуют полную группу попарно несовместимых событий. Так как кость предполагается однородной и симметричной, то события U_1, U_2, ..., U_6 являются и равновозможными, т. е. элементарными.

Определение 3. Событие A называется *благоприятствующим* событию B, если наступление события A влечет за собой наступление события B.

Пример 2. Пусть при бросании игральной кости события U_2, U_4 и U_6 — появление соответственно двух, четырех и шести очков, а A — событие, состоящее в появлении четного числа очков; события U_2, U_4 и U_6 благоприятствуют событию A.

Определение 4 (классическое определение вероятности). *Вероятностью* $P(A)$ события A называется отношение m/n числа элементарных событий, благоприятствующих событию A, к числу всех элементарных событий, т. е.

$$P(A) = m/n.$$

П р и м е р 3. Вычислим вероятность выпадения герба при одном бросании монеты. Очевидно, событие A — выпадение герба — и событие B — выпадение цифры — образуют полную группу несовместимых и равновозможных событий для данного испытания. Значит, здесь $n = 2$. Событию A благоприятствует лишь одно событие — само A, т. е. здесь $m = 1$. Поэтому $P(A) = \frac{1}{2}$.

П р и м е р 4. Очевидно, что при одном бросании игральной кости (вероятность выпадения какой-либо цифры от 1 до 6 будет равна

$$P(U_i) = \frac{1}{6},\ i = 1,\ 2,\ \ldots,\ 6.$$

П р и м е р 5. Найдем вероятность того, что при однократном бросании игральной кости выпадет число очков, делящееся на 2 (событие A).

Число элементарных событий здесь 6. Число благоприятствующих элементарных событии 3 (выпадение 2, 4 и 6). Поэтому

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

П р и м е р 6. При составлении команды космического корабля возникает вопрос о психологической совместимости отдельных членов экипажа. Допустим, что надо составить команду из трех человек: командира, инженера и врача. На место командира есть три кандидата: a_1, a_2, a_3; на место инженера — четыре кандидата: b_1, b_2, b_3, b_4; на место врача — два кандидата: c_1, c_2. Проведенная проверка показала психологическую несовместимость командира a_2 с инженерами b_3, b_4 и с врачом c_2, а также инженера b_2 с врачом c_2. Будем для простоты считать, что без учета фактора несовместимости все варианты составления команды равновозможны. Какова в этом случае вероятность того, что будет составлен экипаж, все члены которого психологически совместимы друг с другом?

Представим все варианты команды, при которых члены экипажа совместимы друг с другом в виде «дерева» (рис. 1). Число ветвей этого дерева, т. е. исходов, благоприятствующих событию A, равно 16, а общее число возможных комбинаций по правилу произведения равно $4 \cdot 3 \cdot 2 = 24$. Искомая вероятность $P(A) = \frac{16}{24} = \frac{2}{3}$.

З а д а ч а (Вероятности рождения мальчиков и девочек). Будем предполагать, что случаи рождения мальчика и девочки — равновозможные события.

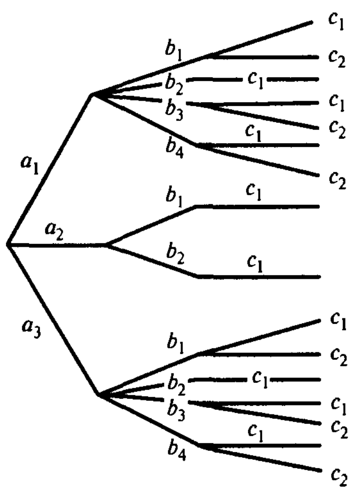


Рис. 1

Пусть в семье двое детей. Какова вероятность, что оба ребенка — мальчики? Если известно, что один мальчик, какова вероятность, что оба ребенка — мальчики?

На первый вопрос ответить нетрудно. Имеется четыре равновозможных исхода: *ММ*, *МД*, *ДМ*, *ДД* (*М* — мальчик, *Д* — девочка). Исходы *МД* и *ДМ* различны, так как в первом из них сначала родился мальчик, а потом девочка, во втором — наоборот. Из этих четырех исходов только один *ММ* благоприятствует нашему событию. Отсюда следует, что $P(MM)=\frac{1}{4}$.

Если дополнительно известно, что один ребенок — мальчик, то событие *ДД* исключается. Из трех равновозможных событий *ММ*, *МД*, *ДМ* по-прежнему только одно *ММ* благоприятствует желаемому исходу. Поэтому $P(MM)=\frac{1}{3}$.

Если известно, что старший ребенок — мальчик, то исключаются исходы *ДМ* и *ДД*. В этом случае $P(MM)=\frac{1}{2}$.

Из приведенного классического определения вероятности вытекают следующие ее свойства:

1. *Вероятность достоверного события равна единице.*

Действительно, достоверному событию должны благоприятствовать все n элементарных событий, т. е. $m=n$ и, следовательно,

$$P(A)=m/n=n/n=1.$$

2. *Вероятность невозможного события равна нулю.*

В самом деле, невозможному событию не может благоприятствовать ни одно из элементарных событий, т. е. $m=0$, откуда

$$P(A)=m/n=0/n=0.$$

3. *Вероятность случайного события есть положительное число, заключенное между нулем и единицей.*

Действительно, случайному событию благоприятствует лишь часть из общего числа элементарных событий. Поэтому в этом случае $0<m<n$ и, значит, $0<m/n<1$. Следовательно, $0<P(A)<1$.

Таким образом, вероятность любого события удовлетворяет двойному неравенству

$$0\leqslant P(A)\leqslant 1.$$

Замечание. Из определения вероятности следует, что элементарные события являются *равновероятными*, т. е. обладают одной и той же вероятностью.

4. Применение элементов комбинаторики к нахождению вероятностей.

Комбинаторика — раздел математики, занимающийся вопросами о том, сколько комбинаций определенного типа можно получить из данных предметов (элементов).

Как при решении задач с использованием классического определения вероятности, так и в дальнейшем могут нам понадобиться некоторые определения и формулы комбинаторики. Приведем наиболее употребительные из них.

Определение 1. *Размещениями* из n различных элементов по m элементов ($m \leqslant n$) называются комбинации, составленные из данных n элементов по m элементов, которые отличаются либо самими элементами, либо порядком элементов.

Например, из трех элементов a, b, c можно составить по два элемента следующие размещения:

$$ab,\ ac,\ bc,\ ba,\ ca,\ cb.$$

Число A_n^m размещений из n элементов $a_1, a_2, \ldots, a_n$ по m равно

$$A_n^m = n(n-1)(n-2)\ldots(n-m+1). \tag{1.1}$$

Пусть $a_{\alpha_1}a_{\alpha_2}\ldots a_{\alpha_m}$ ($1 \leqslant \alpha_k \leqslant n$; $k = 1, \ldots, m$) — всевозможные размещения, содержащие m элементов. Будем эти размещения строить последовательно. Сначала определим a_{α_1} — первый элемент размещения. Очевидно, из данной совокупности n элементов его можно выбрать n различными способами. После выбора первого элемента a_{α_1} для второго элемента a_{α_2} остается $n-1$ способов выбора и т. д. Так как каждый такой выбор дает новое размещение, то все эти выборы можно свободно комбинировать между собой. Поэтому общее число размещений равно указанному произведению (1.1).

Пример 1. Сколько можно составить сигналов из 6 флажков различного цвета, взятых по 2?

Искомое число сигналов $A_6^2 = 6 \cdot 5 = 30$.

Определение 2. *Перестановками* из n различных элементов называются размещения из этих n элементов по n.

Перестановки можно рассматривать как частный случай размещений при $m = n$, поэтому общее число перестановок из n элементов равно

$$P_n = n(n-1)(n-2)\ldots 3 \cdot 2 \cdot 1 = n! \tag{1.2}$$

Пример 2. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, если каждая цифра входит в изображение числа только один раз?

Искомое количество трехзначных чисел $P_3 = 3! = 1 \cdot 2 \cdot 3 = 6$.

Определение 3. *Сочетаниями* из n различных элементов по m элементов называются комбинации, составленные из данных n элементов по m элементов, которые отличаются хотя бы одним элементом.

Отметим разницу между сочетаниями и размещениями: в первых не учитывается порядок элементов.

Обозначим через C_n^m число сочетаний из n элементов по m.

Рассмотрим все допустимые сочетания элементов $a_{\alpha_1}a_{\alpha_2}\ldots a_{\alpha_m}$. Делая в каждом из них $m!$ возможных перестановок их элементов, очевидно, получим общее число рамещений из n элементов по m. Таким образом, $C_n^m \cdot m! = A_n^m$; отсюда

$$C_n^m = \frac{A_n^m}{m!} = \frac{n(n-1)\dots(n-m+1)}{m!}. \tag{1.3}$$

Формулу (3) можно представить также в виде

$$C_n^m = \frac{n!}{m!(n-m)!}.$$

C_n^m обладает очевидной особенностью

$$C_n^m = C_n^{n-m},$$

которая также верна и при $m=0$, если принять $C_n^0 = 1$.

Этой особенностью удобно пользоваться, когда $m > \frac{n}{2}$.

Числа C_n^m являются коэффициентами в формуле *бинома Ньютона*

$$(p+q)^n = p^n + C_n^1 p^{n-1}q + C_n^2 p^{n-2}q^2 + \dots + q^n$$

и поэтому часто называются *биномиальными коэффициентами.*

Пример 3. Сколькими способами можно выбрать две детали из ящика, содержащего 10 деталей?

Искомое число способов

$$C_{10}^2 = \frac{10!}{2!8!} = \frac{9 \cdot 10}{2} = 45.$$

При решении задач комбинаторики можно использовать следующие правила:

Правило суммы. *Если некоторый элемент A может быть выбран из совокупности элементов m способами, а другой элемент B — n способами, то выбрать либо A, либо B можно $m+n$ способами.*

Правило произведения. *Если элемент A можно выбрать из совокупности элементов m способами и после каждого такого выбора элемент B можно выбрать n способами, то пара элементов (A, B) в указанном порядке может быть выбрана mn способами.*

Эти правила справедливы и для любого конечного числа элементов.

Приведем, наконец, примеры применения формул комбинаторики к нахождению вероятностей событий.

Пример 4. Набирая номер телефона, абонент забыл две последние цифры и, помня лишь, что эти цифры различны, набрал их наудачу. Какова вероятность того, что номер набран правильно?

Две последние цифры можно набрать A_{10}^2 способами, а благоприятствовать событию M (цифры набраны правильно) будет только один способ. Поэтому

$$P(M) = \frac{1}{A_{10}^2} = \frac{1}{10 \; 9} = \frac{1}{90} \simeq 0{,}011.$$

Пример 5. Партия из 10 деталей содержит одну нестандартную. Какова вероятность, что при случайной выборке 5 деталей из этой партии все они будут стандартными (событие A)?

Здесь число всех случайных выборок $n = C_{10}^5$, а число выборок, благоприятствующих событию A, есть $m = C_9^5$. Таким образом, искомая вероятность

$$P(A) = \frac{C_9^5}{C_{10}^5} = \frac{9!}{5!4!} \cdot \frac{5!5!}{10!} = \frac{5}{10} = \frac{1}{2} = 0{,}5$$

Пример 6 (Задача о Генуэзской лотерее*). Разыгрывается 90 номеров, из которых выигрывают пять. По условию можно ставить ту или иную сумму на любой из 90 номеров или на любую совокупность двух, трех, четырех или пяти номеров. Если участник лотереи ставил на один номер, то он получал при выиграше в 15 раз больше ставки; если на два номера (амбо), то в 270 раз больше; если на три номера (терн), то в 5500 раз больше, если на четыре номера (катерн) — в 75 000 раз больше; если на пять номеров (квин) — в 1 000 000 раз больше, чем ставка. Какова вероятность выигрыша в каждом из указанных пяти случаев?

В первом случае вероятность выигрыша оказывалась

$$p_1 = \frac{C_5^1}{C_{90}^1} = \frac{5}{90} = \frac{1}{18} \simeq 0{,}055,$$

во втором, третьем, четвертом и пятом случаях вероятности выигрыша были соответственно равны:

$$p_2 = \frac{C_5^2}{C_{90}^2} = \frac{5 \cdot 4}{90 \cdot 89} = \frac{2}{801} \simeq 0{,}0025,$$

$$p_3 = \frac{C_5^3}{C_{90}^3} = \frac{5 \cdot 4 \cdot 3}{90 \cdot 89 \cdot 88} = \frac{1}{11\,748} \simeq 8{,}5 \cdot 10^{-5},$$

$$p_4 = \frac{C_5^4}{C_{90}^4} = \frac{5 \cdot 4 \cdot 3 \cdot 2}{90 \cdot 89 \cdot 88 \cdot 87} = \frac{1}{511\,038} \simeq 1{,}9 \cdot 10^{-6},$$

$$p_5 = \frac{C_5^5}{C_{90}^5} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86} = \frac{1}{43\,949\,268} \simeq 2{,}3 \cdot 10^{-8}.$$

§ 1.2. Геометрическая вероятность. Статистическое и аксиоматическое определения вероятности

1. Геометрическая вероятность. Классическое определение вероятности предполагает, что число всех элементарных событий конечно. Но на практике часто встречаются опыты, для которых множество таких событий бесконечно. Например, пусть на отрезке [0; 1] числовой прямой ставят наудачу точку. Что подсказывает нам интуиция о вероятностях событий «точка попала на правую половину

---

* В XVII веке в Генуе пользовалась большим успехом знаменитая лотерея В XVIII столетии она разыгрывалась также во Франции, Германии и других странах

отрезка» и «точка попала на левую половину отрезка»? Поскольку точка ставится наудачу, то естественно считать эти события равновероятными — вероятность каждого 0,5 (поскольку это противоположные события). Ну, а если мы разделим отрезок на 10 равных отрезков и рассмотрим события «точка попала на левый отрезок», «точка попала на второй слева отрезок», ..., «точка попала на правый отрезок»? Это опять равновероятные события. А вероятность каждого из них оказывается равной 0,1, поскольку это совокупность всех элементарных событий нашего опыта. Поставим теперь вопрос: «Какова вероятность попадания точки на отрезок [0,3; 0,7]?» Поскольку этому событию благоприятствуют четыре из указанных выше элементарных события, то искомая вероятность равна 0,4, т. е. длине отмеченного отрезка. В общем случае смысл выражения «точка поставлена наудачу на отрезок длины 1» состоит в том, что вероятность попадания точки на часть этого отрезка длины l равна этому числу l (если вместо отрезка [0; 1] взять отрезок $[0; s]$, $s > l$, то искомая вероятность будет равна l/s).

Аналогично уясняется смысл выражения «точка поставлена наудачу в квадрат со стороной 1 (или в прямоугольник площадью 1)»,— это значит, что вероятность попадания точки на любую часть этого квадрата (или прямоугольника) равна площади этой части.

В более сложных случаях (на плоскости) может оказаться, что при геометрической интерпретации получится такая картина: имеется фигура площадью s, и на нее наудачу ставится точка. Тогда вероятность попадания точки на часть этой фигуры, имеющую площадь q, оказывается равной q/s.

Аналогично в трехмерном случае (в пространстве) здесь берется отношение соответствующих объемов. Такое определение вероятности получило название *геометрического*.

Пример. В окружность вписан квадрат. В круг наудучу ставят точку. Какова вероятность того, что эта точка попадет в квадрат?

Отношение площадей квадрата и круга дает искомую вероятность:

$$p = \frac{a^2}{\pi R^2} = \frac{2R^2}{\pi R^2} = \frac{2}{\pi}.$$

2. Относительная частота. Статистическое определение вероятности. Классическое определение вероятности оказывается непригодным для изучения произвольных случайных событий. Так, оно неприемлемо, если результаты испытания не равновозможны. Например, при бросании неправильной игральной кости выпадение ее различных граней не равновозможно.

В таких случаях используется так называемое статистическое определение вероятности.

Пусть произведено n испытаний, при этом некоторое событие A наступило m раз.

Определение 1. Число m называется *абсолютной частотой* (или просто *частотой*) события A, а отношение

$$P^*(A) = \frac{m}{n}$$

называется *относительной частотой* события A.

Пример 1. При транспортировке из 10 000 арбузов испортилось 26. Здесь $m = 26$ — абсолютная частота испорченных арбузов, а

$$P^*(A) = \frac{26}{10\,000} = 0{,}0026$$

относительная.

Результаты многочисленных опытов и наблюдений, многие из которых описаны, например, в работах [1—4], помогают заключить: при проведении серий из n испытаний, когда число n сравнительно мало, относительная частота $P^*(A)$ принимает значения, которые могут довольно сильно отличаться друг от друга. Но с увеличением n — числа испытаний в сериях — относительная частота

$$P^*(A) = m/n$$

приближается к некоторому числу $P(A)$, стабилизируясь возле него и принимая все более устойчивые значения.

Пример 2. Было проведено 10 серий бросаний монеты, по 1000 бросаний в каждой. Относительные частоты выпадения герба оказались равными 0,501; 0,485; 0,509; 0,536; 0,485; 0,488; 0,500; 0,497; 0,494; 0,484 (см. [4]). Эти частоты группируются около числа 0,5.

Определение 2 (статистическое определение вероятности). Вероятностью события A в данном испытании называется число $P(A)$, около которого группируются значения относительной частоты при больших n.

В условиях только что приведенного примера указанная вероятность равна 0,5.

Пример 3. По данным шведской статистики, относительные частоты рождения девочек по месяцам одного года характеризуются следующими числами (расположены в порядке следования месяцев, начиная с января): 0,486; 0,489; 0,490; 0,471; 0,478; 0,482; 0,462; 0,484; 0,485; 0,491; 0,482; 0,473 (см. [2]). Эти частоты группируются около числа 0,482.

Таким образом, относительная частота события приближенно совпадает с его вероятностью, если число испытаний достаточно велико. Имеется огромный опытный материал по проверке последнего утверждения. Приведем еще один такой пример с бросанием монеты (см. [2]).

Экспериментатор	Число бросаний	Число выпадений герба	Относительная частота
Бюффон	4 040	2 048	0,5080
К. Пирсон	12 000	6 019	0,5016
К. Пирсон	24 000	12 012	0,5005

Здесь относительные частоты незначительно отличаются от числа 0,5, причем тем меньше, чем больше число испытаний. При 4040 испытаниях отклонение равно 0,008, а при 24 000 — 0,0005.

Таким образом, относительная частота события приближенно совпадает с его вероятностью в статистическом смысле, если число испытаний достаточно велико.

С этой точки зрения величина $m = np$ представляет собой среднее значение числа появления события A при n испытаниях.

При широких предположениях доказывается, что вероятности события в классическом и статистическом смысле совпадают между собой.

3. Аксиоматическое определение вероятности. В современных математических курсах вероятность определяется аксиоматически. При аксиоматическом построении теории вероятностей исходят из свойств вероятности событий, к которым применимо классическое или статистическое определение. Отдельные свойства вероятности известны из предыдущего изложения. Поэтому естественно принять следующие аксиомы.

Аксиома 1. *Каждому событию A поставлено в соответствие неотрицательное число $P(A)$, называемое его вероятностью.*

Аксиома 2. *Вероятность достоверного события равна единице.*

Аксиома 3. *Вероятность суммы попарно несовместимых событий равна сумме вероятностей этих событий.*

Последняя аксиома называется *аксиомой сложения* вероятностей.

Исходя из этих аксиом, свойства вероятностей и зависимости между ними выводят в качестве теорем.

Большая заслуга в аксиоматическом построении теории вероятностей принадлежит советскому математику А. Н. Колмогорову (1903—1987), работы которого положили начало созданию современной теории вероятностей как строгой математической науки [5].

§ 1.3. Свойства вероятности

1. Теорема сложения вероятностей несовместимых событий.

Теорема. *Вероятность суммы двух несовместимых событий A и B равна сумме вероятностей этих событий:*

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

Доказательство. Используем классическое определение вероятности. Предположим, что в данном испытании число всех элементарных событий равно n, событию A благоприятствуют k элементарных событий, событию B — l элементарных событий. Так как A и B — несовместимые события, то ни одно из элементарных событий U_1, U_2, ..., U_n не может одновременно благоприятствовать и событию A, и событию B. Следовательно, событию $A+B$ будет благоприятствовать $k+l$ элементарных событий. По определению вероятности

$$P(A)=k/n,\ P(B)=l/n,\ P(A+B)=(k+l)/n,$$

откуда и следует утверждение теоремы.

Совершенно так же теорема формулируется и доказывается для любого конечного числа попарно несовместимых событий.

Следствие 1. *Если события A_1, A_2, ..., A_n образуют полную группу попарно несовместимых событий, то сумма их вероятностей равна единице:*

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1.$$

Доказательство. Так как события A_1, A_2, ..., A_n образуют полную группу, то появление хотя бы одного из них — достоверное событие, и, значит,

$$P(A_1+A_2+\ldots+A_n)=1.$$

А так как эти события и несовместимые, то

$$P(A_1+A_2+\ldots+A_n)=P(A_1)+P(A_2)+\ldots+P(A_n),$$

что и приводит к искомому равенству.

Следствие 2. *Сумма вероятностей противоположных событий A и $\overline{A}$ равна единице:*

$$P(A)+P(\overline{A})=1. \tag{1.5}$$

Это следствие — частный случай следствия 1.

Пример. В урне 10 шаров: 3 красных, 5 синих и 2 белых. Какова вероятность вынуть цветной шар, если вынимается один шар?

Вероятность вынуть красный шар $P(A)=3/10$, синий $P(B)=5/10$. Так как события A и B несовместимы, то по доказанной выше теореме

$$P(A+B)=P(A)+P(B)=0{,}3+0{,}5=0{,}8.$$

2. Теорема умножения вероятностей.

Определение 1. Два события A и B называют *независимыми*, если вероятность появления каждого из них не зависит от того, появилось другое событие или нет*. В противном случае события A и B называют *зависимыми*.

Пример 1. Пусть в урне находятся 2 белых и 2 черных шара. Пусть событие A — вынут белый шар. Очевидно, $P(A)=\frac{1}{2}$. После

* Несколько событий A_1, , A_k называют *независимыми в совокупности* (или просто *независимыми*), если вероятность появления любого из них не зависит от того, произошли какие-либо другие рассматриваемые события или нет.

первого испытания вынутый шар кладется обратно в урну, шары перемешиваются и снова вынимается шар. Событие B — во втором испытании вынут белый шар — также имеет вероятность $P(B)=\frac{1}{2}$, т. е. события A и B — независимые.

Предположим теперь, что вынутый шар в первом испытании не кладется обратно в урну. Тогда, если произошло событие A, т. е. в первом испытании вынут белый шар, то вероятность события B уменьшается и оказывается равно одной трети, если в первом испытании был вынут черный шар, то вероятность события B увеличивается и становится равно двум третям.

Итак, вероятность события B существенно зависит от того, произошло или не произошло событие A, в таких случаях события A и B — зависимые.

Определение 2. Пусть A и B — зависимые события. *Условной вероятностью* $P_A(B)$ события B называют вероятность события B, найденную в предположении, что событие A уже наступило.

Так, в только что рассмотренном примере $P_A(B)=\frac{1}{3}$.

Условие независимости события B от события A можно записать в виде

$$P_A(B)=P(B),$$

а условие зависимости — в виде

$$P_A(B)\neq P(B).$$

Теорема 1. *Вероятность произведения двух зависимых событий A и B равна произведению вероятности одного из них на условную вероятность другого, найденную в предположении, что первое событие уже наступило:*

$$P(AB)=P(A)P_A(B). \tag{1.6}$$

Доказательство. Пусть из всего числа n элементарных событий k благоприятствуют событию A и пусть из этих k событий l благоприятствуют событию B, а, значит, и событию AB. Тогда

$$P(AB)=l/n=k/n\cdot l/k=P(A)P_A(B),$$

что и доказывает искомое равенство (1.6).

Замечание. Применив формулу (1.6) к событию BA, получим

$$P(BA)=P(B)P_B(A). \tag{1.6$'$}$$

Так как $AB=BA$ (§ 1.1, п. 2), то

$$P_B(A)=\frac{P(AB)}{P(B)}, \tag{1.7}$$

а сравнивая (1.6) и (1.6′), получаем равенство

$$P(A)P_A(B) = P(B)P_B(A). \tag{1.8}$$

П р и м е р 2. Пусть в урне находятся 2 белых и 2 черных шара. Пусть событие A — вынут белый шар. Событие B — во втором испытании вынут белый шар. Рассмотрим тот случай, когда вынутый шар в первом испытании не кладется обратно в урну. Поставим следующий вопрос: какова вероятность вынуть первый и второй раз белые шары? По формуле (1.6) имеем:

$$P(AB) = 1/2 \cdot 1/3 = 1/6.$$

П р и м е р 3. В терапевтическом отделении больницы 70% пациентов — женщины, а 21% — курящие мужчины. Наугад выбирают пациента. Он оказывается мужчиной. Какова вероятность того, что он курит?

Пусть M означает, что пациент — мужчина, а K — что пациент курит. Тогда в силу условия задачи $P(M) = 0{,}3$, а $P(MK) = 0{,}21$. Поэтому с учетом формулы (1.6) искомая условная вероятность

$$P_M(K) = \frac{P(MK)}{P(M)} = \frac{0{,}21}{0{,}3} = 0{,}7.$$

П р и м е р 4. В группе туристов 20% детей, причем 12% девочки. Наугад выбирают ребенка. Какова вероятность того, что это девочка? Какова вероятность того, что это мальчик?

Пусть A означает, что турист — ребенок, $Ж$ — турист женского пола, M — мужского. Тогда по условию

$$P(A) = 0{,}2, \; P(ЖA) = 0{,}12, \; P(MA) = 0{,}08.$$

Следовательно,

$$P_A(Ж) = \frac{P(ЖA)}{P(A)} = \frac{0{,}12}{0{,}2} = 0{,}6,$$

$$P_A(M) = \frac{P(MA)}{P(A)} = \frac{0{,}08}{0{,}2} = 0{,}4.$$

З а д а ч а (курение и случай заболевания легких). В группе обследуемых 1000 человек. Из них 600 курящих и 400 некурящих. Среди курящих 240 человек имеют те или иные заболевания легких. Среди некурящих легочных больных 120 человек. Являются ли курение и заболевание легких независимыми событиями?

Р е ш е н и е. Пусть событие A — обследуемый курит, событие B — обследуемый страдает заболеванием легких.

Тогда, согласно условию задачи,

$$P(B) = \frac{240 + 120}{1000} = 0{,}36; \; P_A(B) = \frac{240}{600} = \frac{2}{5} = 0{,}4.$$

Так как $0{,}36 \neq 0{,}4$, события A и B зависимы.

П р и м е р 5. Предположим, что вероятности встретить реку, загрязняемую постоянным фактором A — $P(A)$, временным фактором B — $P(B)$ и обоими факторами — $P(AB)$, равны соответственно 0,4; 0,1 и 0,05.

Найдем:

1) вероятность того, что река, загрязняемая временным фактором, будет к тому же загрязнена и постоянным фактором, т. е. $P_B(A)$;

2) вероятность того, что река, загрязняемая постоянным фактором, будет еще загрязнена и временным фактором, т. е. $P_A(B)$.

Имеем, согласно (1.7):

$$P_B(A) = \frac{P(AB)}{P(B)},$$

откуда

$$P_B(A) = \frac{0{,}05}{0{,}1} = 0{,}5.$$

Аналогично, используя формулу (1.6), находим

$$P_A(B) = \frac{0{,}05}{0{,}4} = 0{,}125.$$

Т е о р е м а 2. ***Вероятность произведения двух независимых событий A и B равна произведению вероятностей этих событий***

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

Действительно, если A и B — независимые события, то $P_A(B) = P(B)$ и формула (1.6) превращается в формулу (1.9).

В случае независимых событий в совокупности эта теорема распространяется на любое конечное число их, т. е. имеет место равенство

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n). \tag{1.10}$$

З а м е ч а н и е 1. Если события $A_1, A_2, \dots, A_n$ независимы в совокупности, то и противоположные им события $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ также независимы в совокупности.

П р и м е р 6. Найдем вероятность одновременного поражения цели двумя орудиями, если вероятность поражения цели первым орудием (событие A) равна 0,8, а вторым (событие B) — 0,7.

События A и B независимы, поэтому искомая вероятность

$$P(AB) = 0{,}7 \cdot 0{,}8 = 0{,}56.$$

П р и м е р 7. Вероятность выживания одной клетки в течение 20 минут $P = 0{,}7$. В пробирке с благоприятными для существования этих клеток условиями находятся только что разделившиеся две клетки. Какова вероятность того, что через 20 минут они будут жизнеспособны?

Пусть событие A — первая клетка жизнеспособна через 20 мин, событие B — вторая клетка жизнеспособна через 20 мин. Будем считать, что между клетками нет внутривидовой конкуренции, т. е. события A и B независимы. Событие, что обе клетки жизнеспособны, есть событие AB.

$$P(AB) = 0{,}7 \quad 0{,}7 = 0{,}49.$$

Пример 8. Пусть у нас перемешаны записи нейронной активности 10 клеток из одной области мозга (у 5 клеток зарегистрирована активность, характерная для клеток «внимания», у 5 — другой вид активности) и 20 из другой области (у 15 — активность типа клеток «внимания», у 5 — другого вида). Выясним, зависимы ли события A — «выбранная наугад запись сделана в первой области» и B — на «выбранной наугад записи зарегистрирована активность, характерная для клеток «внимания»». Имеем

$$P(A) = 10/30 = 1/3; \ P(B) = 20/30 = 2/3;$$

$$P(AB) = 5/30 = 1/6; \ P(AB) \neq P(A)P(B).$$

Следовательно, события A и B зависимы.

Теорема 3. *Если события A_1, A_2, ..., A_n независимы в совокупности, то вероятность наступления хотя бы одного из этих событий (т. е. вероятность суммы) вычисляется по формуле*

$$P(A_1 + A_2 + \ldots + A_n) = 1 - P(\overline{A}_1)P(\overline{A}_2) \ldots P(\overline{A}_n). \tag{1.11}$$

Доказательство. Событие $\overline{A}_1 \cdot \overline{A}_2 \ldots \overline{A}_n$ состоит в том, что не произошло ни одно из событий A_i ($i = 1, 2, \ldots, n$). Оно противоположно событию, состоящему в том, что произошло хотя бы одно из событий A_i, т. е. сумме событий $A_1 + A_2 + \ldots + A_n$. Поэтому, согласно формуле (1.5),

$$P(A_1 + A_2 + \ldots + A_n) + P(\overline{A}_1 \cdot \overline{A}_2 \ldots \overline{A}_n) = 1,$$

откуда

$$P(A_1 + A_2 + \ldots + A_n) = 1 - P(\overline{A}_1 \cdot \overline{A}_2 \ldots \overline{A}_n).$$

Но с учетом замечания 1 (п. 2) и формулы (1.10)

$$P(\overline{A}_1 \cdot \overline{A}_2 \ldots \overline{A}_n) = P(\overline{A}_1)P(\overline{A}_2) \ldots P(\overline{A}_n),$$

что и приводит к искомому равенству (1.11).

3. Теорема сложения вероятностей совместимых событий.

Теорема. *Вероятность суммы двух совместимых событий A и B равна сумме вероятностей этих событий минус вероятность их произведения:*

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

Доказательство. Пусть из всего числа n элементарных событий k благоприятствуют событию A, l — событию B и m —

одновременно событиям A и B. Отсюда событию $A+B$ благоприятствуют $k+l-m$ элементарных событий. Тогда

$$P(A+B)=\frac{k+l-m}{n}=\frac{k}{n}+\frac{l}{n}-\frac{m}{n}=P(A)+P(B)-P(AB).$$

Замечание 1. При использовании формулы (1.12) следует иметь в виду, что события A и B могут быть как независимыми, так и зависимыми.

Для независимых событий

$$P(A+B)=P(A)+P(B)-P(A)P(B); \tag{1.13}$$

для зависимых событий

$$P(A+B)=P(A)+P(B)-P(A)P_A(B).$$

Замечание 2. Если события A и B несовместимы, то их произведение AB есть невозможное событие и, следовательно, $P(AB)=0$, т. е. формула (1.4) является частным случаем формулы (1.12).

Пример. Вероятности попадания в цель при стрельбе первого и второго орудий соответственно равны: $P(A)=0{,}7$ и $P(B)=0{,}8$. Найдем вероятность попадания при одном залпе (из обоих орудий) хотя бы одним из орудий.

Очевидно, события A и B совместимы и независимы. Поэтому

$$P(A+B)=P(A)+P(B)-P(AB)=0{,}7+0{,}8-0{,}7\ \ 0{,}8=1{,}5-0{,}56=0{,}94.$$

4. Формула полной вероятности.

Теорема. *Вероятность события A, которое может наступить лишь при условии появления одного из n попарно несовместимых событий B_1, B_2, ..., B_n, образующих полную группу, равна сумме произведений вероятностей каждого из этих событий на соответствующую условную вероятность события A:*

$$P(A)=P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A)+\ldots+P(B_n)P_{B_n}(A) \tag{1.14}$$

(формула полной вероятности).

События B_1, B_2, ..., B_n будем называть *гипотезами.*

Доказательство. Событие A может наступить лишь при условии наступления одного из событий B_1, B_2, ..., B_n, т. е. $A=B_1A+$ $+B_2A+\ldots+B_nA$, причем ввиду несовместимости событий B_1, B_2, ..., B_n события B_1A, B_2A, ..., B_nA также несовместимы. Поэтому на основании теорем сложения и умножения вероятностей имеем

$$P(A)=P(B_1A)+P(B_2A)+\ldots+P(B_nA)=$$
$$=P(B_1)P_{B_1}(A)+P(B_2)P_{B_2}(A)+\ldots+P(B_n)P_{B_n}(A).$$

Пример 1. Имеются три одинаковых по виду ящика. В первом находятся две белые мыши и одна серая, во втором — три белые и одна серая, в третьем — две белые и две серые мыши. Какова вероятность того, что из наугад выбранного ящика будет извлечена белая мышь?

Обозначим B_1 — выбор первого ящика, B_2 — выбор второго ящика, B_3 — выбор третьего ящика, A — извлечение белой мыши.

Так как все ящики одинаковы, то $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$. Если выбран первый ящик, то $P_{B_1}(A) = \frac{2}{3}$. Аналогично $P_{B_2}(A) = \frac{3}{4}$, $P_{B_3}(A) = \frac{1}{2}$. Наконец, по формуле (1.14) получаем

$$P(A) = \frac{1}{3}\;\frac{2}{3} + \frac{1}{3}\cdot\frac{3}{4} + \frac{1}{3}\;\frac{1}{2} = \frac{23}{36}.$$

Пример 2. В санатории 30% пациентов — мужчины (M) и 70% — женщины ($Ж$). Болезни сердца среди мужчин встречаются в два раза чаще, чем среди женщин. Какова вероятность того, что наугал выбранный пациент сердечник?

Обозначив C — начилие заболевания сердца, запишем:

$$P(M) = 0{,}3,\;\; P(Ж) = 0{,}7,\;\; P_M(C) = \frac{2}{3},\;\; P_Ж(C) = \frac{1}{3}.$$

Подставляя эти числа в формулу полной вероятности **(1.14)**, **по**лучим

$$P(C) = 0{,}3\cdot\frac{2}{3} + 0{,}7\cdot\frac{1}{3} \approx 0{,}2 + 0{,}23 = 0{,}43.$$

Задача (смог над городом). На город примерно 100 дней в году дует ветер с севера и 200 дней в году — с запада. Промышленные предприятия, расположенные на севере, производят выброс вредных веществ каждый третий день, а расположенные на западе — в последний день каждой недели. Как часто город подвергается воздействию вредных выбросов? Иными словами, какова вероятность того, что в наугад выбранный день город будет накрыт промышленным смогом?

Обозначив C — ветер с севера, $З$ — ветер с запада и B — воздействие вредных выбросов на город, можем записать:

$$P(C) = \frac{100}{365} = \frac{20}{73} \simeq 0{,}27;\;\; P(З) = \frac{200}{365} = \frac{40}{73} \simeq 0{,}55;$$

$$P_C(B) = \frac{1}{3} \simeq 0{,}33;\;\; P_З(B) = \frac{1}{7} \simeq 0{,}14.$$

Отсюда по формуле полной вероятности

$$P(B) = P(C)P_C(B) + P(З)P_З(B) = \frac{20}{73}\cdot\frac{1}{3} + \frac{40}{73}\cdot\frac{1}{7} \approx 0{,}09 + 0{,}08 = 0{,}17.$$

Таким образом, около двух месяцев в году город накрыт смогом.

5. Формулы Байеса. Пусть в условиях рассуждения, относящегося к формуле полной вероятности, осуществлено одно испытание, в результате которого произошло событие A. Спрашивается, как изменились (в связи с тем, что событие A уже произошло) вероятности гипотез, т. е. величины $P(B_k)$, $k = 1, 2, \ldots, n$?

Найдем условную вероятность $P_A(B_k)$. По формуле (1.8) (см. п. 2) имеем

$$P(AB_k) = P(A)P_A(B_k) = P(B_k)P_{B_k}(A).$$

Отсюда

$$P_A(B_k) = \frac{P(B_k)\, P_{B_k}(A)}{P(A)}.$$

Наконец, используя формулу полной вероятности, находим:

$$P_A(B_k) = \frac{P(B_k)\, P_{B_k}(A)}{\sum\limits_{j=1}^{n} P(B_j)\, P_{B_j}(A)}, \quad k = 1,\ 2,\ \ldots,\ n. \qquad (1.15)$$

Выражения (1.15) называют *формулами Байеса**.

Пример. Партия деталей изготовлена тремя рабочими, причем первый рабочий изготовил 25% всех деталей, второй — 35%, третий — 40%. В продукции первого рабочего брак составляет 5%, в продукции второго — 4% и в продукции третьего — 2%. Случайно выбранная для контроля деталь оказалась бракованной. Какова вероятность того, что она изготовлена вторым рабочим?

Введем обозначения для событий: A — выбранная для контроля деталь оказалась бракованной; B_1, B_2, B_3 — эта деталь изготовлена соответственно первым, вторым и третьим рабочим. Имеем:

$$P(B_1) = 0{,}25;\ P(B_2) = 0{,}35;\ P(B_3) = 0{,}40;$$

$$P_{B_1}(A) = 0{,}05;\ P_{B_2}(A) = 0{,}04;\ P_{B_3}(A) = 0{,}02.$$

По формуле Байеса находим

$$P_A(B_2) = \frac{0{,}35\ \ 0{,}04}{0{,}25\ \ 0{,}05 + 0{,}35\ \ 0{,}04 + 0{,}40\ \ 0{,}02} = \frac{28}{69} \simeq 0{,}4.$$

Как здесь, так и в ряде других примеров для облегчения вычислений можно использовать калькулятор.

§ 1.4. Случайные события в физике, химии, биологии

1. Цепь приборов. Рассмотрим участок электрической цепи, содержащий два последовательно соединенных прибора: A и B (рис. 2, *а*).

Предположим, что приборы работают независимо один от другого, и каждый из них может либо пропустить ток (прибор исправен), либо не пропустить (прибор неисправен). Обозначим $P(A)$ и $P(B)$ вероятности исправности приборов A и B соответственно. Для

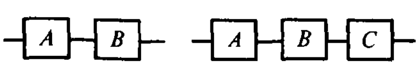


Рис. 2

* Томас Байес (1702—1761) — английский математик.

того чтобы по участку цепи прошел ток, нужно, чтобы и прибор A, и прибор B были исправны, т. е. нужно совмещение исправности приборов. Так как приборы работают независимо, то по формуле умножения вероятностей вероятность прохождения тока выразится произведением

$$P = P(A) \cdot P(B). \tag{1.16}$$

Совершенно аналогично для трех последовательно соединенных и независимо работающих приборов A, B, C (рис. 2, *б*) вероятность прохождения тока по участку цепи выразится произведением

$$P = P(A) \cdot P(B) \ P(C),$$

а для n приборов A_1, A_2, ..., A_n — произведением

$$P = P(A_1) \cdot P(A_2) \ldots \ P(A_n).$$

В частности, если приборы однотипны, точнее говоря, если вероятности их исправности равны $P(A_1) = P(A_2) = \ldots = P(A_n) = p$, то вероятность прохождения тока $P = p^n$.

Можно поставить в некотором смысле обратную задачу. Предположим, что вероятность исправности первого прибора $P(A)$ известна. После испытаний установили вероятность прохождения тока по всему участку P. Тогда из формулы (1.16) можно найти вероятность исправности второго прибора $P(B)$. Например, если $P(A) = 0{,}9$; $P = 0{,}72$, то в силу (1.16) $P(B) = P/P(A) = 0{,}72/0{,}9 = 0{,}8$.

2. Цепь реакций. Цепной называют химическую реакцию, которая представляет собой цепочку одинаковых звеньев. Звеном может быть одна, две, реже — несколько стадий. Например, звено

$$\dot{\mathrm{R}} + \mathrm{O_2} \to \mathrm{R\dot{O}_2}, \ \mathrm{R\dot{O}_2} + \mathrm{RH} \to \mathrm{ROOH} + \dot{\mathrm{R}},$$

начавшись с появления свободного радикала углеводорода $\dot{\mathrm{R}}$, во второй стадии снова выделяет этот радикал и тем самым создает возможность повторения такого же звена.

На некотором этапе цепная реакция может оборваться. Причиной обрыва может служить захват свободного радикала стенкой сосуда, действие ингибитора и т. п. Таким образом, на каждом этапе существует некоторая вероятность p продолжения цепи и вероятность $q = 1 - p$ обрыва цепи.

Какова вероятность, что цепная реакция содержит n звеньев? Для осуществления такой реакции нужно, чтобы n раз произошло продолжение реакции и после этого произошел обрыв. Так как процессы продолжения и обрыва независимы, то по формуле умножения вероятностей для $P(n)$ — вероятности появления цепи длины n, т. е. содержащей n звеньев,— можем написать

$$P(n) = \underbrace{p \cdot p \ldots p}_{n \text{ раз}} \cdot q = p^n q = p^n (1 - p).$$

3. Молекула полимера. Процесс полимеризации состоит в том, что к звену-мономеру присоединяется такой же мономер, к этому звену — еще один такой же мономер и т. д. Присоединение происходит с некоторой вероятностью p и, следовательно, не происходит с вероятностью $q = 1 - p$. Так как каждое следующее присоединение происходит независимо от предыдущих, то вероятность образования молекулы, содержащей n мономеров, как и в предыдущем примере, вычисляется по формуле

$$P(n) = \underbrace{p\ p \ldots p}_{n\ \text{раз}} \cdot q = p^n q = p^n(1 - p).$$

4. Параллельное соединение.

Цепь последовательно соединенных приборов — один из крайних, наиболее простых типов соединений. Другим простейшим типом является параллельное соединение.

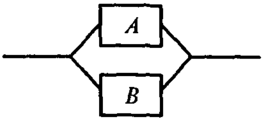


Рис. 3

Рассмотрим участок цепи, содержащий два прибора A и B, соединенных параллельно (рис. 3). Предположим, что приборы работают независимо и $P(A)$ — вероятность прохождения сигнала по прибору A, а $P(B)$ — по прибору B. Например, сигнал проходит по прибору, если прибор исправен, и не проходит — в противном случае. Очевидно, сигнал пройдет, если будет исправен хотя бы один прибор. Таким образом, вероятность прохождения сигнала по участку цепи — это вероятность $P(A + B)$, где сумма $A + B$ означает исправную работу хотя бы одного из приборов. Так как приборы работают независимо, то эту вероятность можно вычислить по формуле (1.13)

$$P(A + B) = P(A) + P(B) - P(A)P(B). \tag{1.17}$$

Например, если $P(A) = 0{,}8$, $P(B) = 0{,}9$, то

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

Можно поставить и обратную задачу. Предположим, что один из приборов — эталонный и вероятность его безотказной работы (т. е. вероятность прохождения по нему сигнала) известна. После испытаний установили вероятность прохождения сигнала по всему участку. Тогда из формулы (1.17) можно найти вероятность безотказной работы второго прибора. Например, если $P(A) = 0{,}8$, $P(A + B) = 0{,}95$, то подставив это в (1.17), будем иметь

$$0{,}95 = 0{,}8 + P(B) - 0{,}8 \cdot P(B).$$

Отсюда

$$P(B) = \frac{0{,}15}{0{,}2} = 0{,}75.$$

Если участок цепи состоит из n независимо работающих приборов, соединенных параллельно (рис. 4), и A_i означает, что сигнал прошел по i-му прибору, т. е. что i-й прибор исправен, то вероятность прохождения сигнала по участку — это вероятность исправной работы хотя бы одного прибора, т. е. вероятность суммы $A_1 + A_2 + \ldots + A_n$. Следовательно (см. п. 2, формула (1.8)),

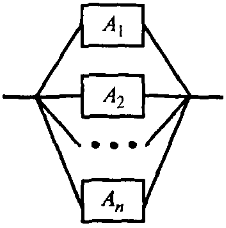


Рис. 4

$$P(A_1 + A_2 + \ldots + A_n) = 1 - P(\overline{A}_1)\ P(\overline{A}_2) \ldots P(\overline{A}_n).$$

В частности, если все вероятности равны $P(A_1) = P(A_2) = \ldots = P(A_n) = p$, то $P(\overline{A}_i) = 1 - P(A_i) = 1 - p = q$ и

$$P(A_1 + A_2 + \ldots + A_n) = 1 - q^n.$$

Видно, что даже при малой вероятности p, т. е. при q, близкой к единице, выбирая достаточно большое n, можно сделать вероятность $P(A_1 + \ldots + A_n)$ достаточно близкой к единице.

Параллельно соединенными могут работать не только приборы. Вместо разобранной схемы можно рассмотреть систему химических реакций, участки нервной или кровеносной систем и т. п.

5. Последовательные и параллельные соединения.

В предыдущих пунктах мы рассмотрели порознь последовательные и параллельные соединения приборов и установили, как вычисляется вероятность прохождения сигнала по участку схемы в том и другом случае. На практике приходится иметь дело с различными сочетаниями соединений обоих типов. Рассмотрим два характерных примера.

Предположим, что сигнал проходит по участку схемы, состоящему из двух параллельных блоков A и B, первый из которых состоит из одного прибора A, а второй содержит два последовательно соединенных прибора B_1 и B_2 (рис. 5, *а*). Пусть возможность отказа одного из приборов не зависит от работы остальных. Сигнал проходит, если хотя бы один из блоков исправен, а каждый из блоков выходит из строя, если хотя бы один из его приборов отказал.

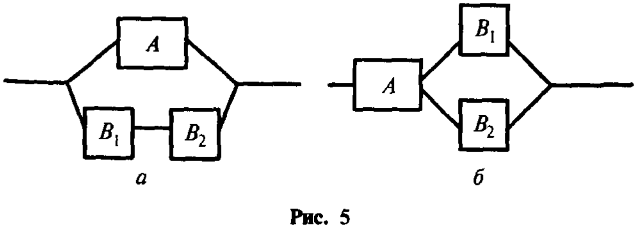


Рис. 5

Обозначим $P(A)$, $P(B_1)$ и $P(B_2)$ — вероятности безотказной ра-боты соответствующих приборов, $P(B)$ — вероятность исправности блока B (вероятность исправности блока A, очевидно, равна $P(A)$), $P(A+B)$ — вероятность прохождения сигнала по цепи Тогда, используя формулы сложения и умножения вероятностей, можно записать

$$P(A+B)=P(A)+P(B)-P(A)\;\;P(B)=$$
$$=P(A)+P(B_1)\;\;P(B_2)-P(A)\;\;P(B_1)\;\;P(B_2) \qquad (1\ 18)$$

На практике чаще задается не вероятность безотказной работы, а вероятность отказа, т е $P(\overline{A})$, $P(\overline{B}_1)$, $P(\overline{B}_2)$ и т п Так как отказ и безотказная работа — взаимно противоположные события, то $P(A)=1-P(\overline{A})$, $P(B_1)=1-P(\overline{B}_1)$, $P(B_2)=1-P(\overline{B}_2)$ и мы снова можем применить формулу (1 18)

Например, если $P(\overline{A})=0{,}1$, $P(\overline{B}_1)=0{,}07$, $P(\overline{B}_2)=0{,}08$, то $P(A)=0{,}9$, $P(B_1)=0{,}93$, $P(B_2)=0{,}92$ Поэтому $P(A+B)=0{,}9+0{,}93\;\;0{,}92-0{,}9\;\;0{,}93\;\;0{,}92=0{,}986$

Теперь предположим, что участок схемы состоит из двух последовательно соединенных блоков A и B, один из которых состоит из одного прибора A, а другой содержит два параллельно соединенных прибора B_1 и B_2 (рис 5, *б*) Пусть по-прежнему приборы работают независимо Блок B выходит из строя, если отказали оба его прибора Сигнал проходит, если оба блока A и B исправны Обозначив $P(AB)$ — вероятность прохождения сигнала по цепи и сохранив остальные обозначения для вероятностей, можем написать $P(AB)=P(A)\;\;P(B)=P(A)[P(B_1)+P(B_2)-P(B_1)\;\;P(B_2)]$ В частности, для данных предыдущего примера

$$P(AB)=0{,}9(0{,}93+0{,}92-0{,}93\;\;0{,}92)=0{,}895$$

Аналогично рассматриваются случаи, когда блок A состоит из двух или большего числа приборов, соединенных параллельно или последовательно, когда блоков более, чем два и т п

Разумеется, и в этом случае вместо приборов могут быть рассмотрены выходы химических реакций, жизнеспособность популяций и т п

6. Законы Менделя. Известно, что в простейших случаях передача некоторого признака по наследству зависит от определенного гена В половых клетках гены, отвечающие за некоторый признак, находятся парами Например, в клетках гороха имеется пара генов, отвечающих за окраску цветков потомства — красную и белую Эти гены могут находиться в двух состояниях — доминантном (обозначается буквой *A*) и рецессивном (обозначается буквой *a*) Поэтому пары генов могут быть такими

$$AA,\ Aa \text{ или } aA,\ aa$$

выписанные возможности определяют генотипы данной особи: первый — доминантный, второй — смешанный, третий — рецессивный. Оказывается, что наследование признака зависит от генотипа особи. Например, для гороха красная окраска цветков — доминантный признак, а белая — рецессивный.

Экспериментально установлен *I закон Менделя*: особи доминантного и смешанного генотипов в фенотипе* обладают доминантным признаком, и только особи рецессивного генотипа в фенотипе обладают рецессивным признаком.

Согласно этому закону, для гороха особи доминантного и смешанного генотипов имеют красную окраску цветков и только особи с рецессивным генотипом имеют белые цветки.

Пусть имеется популяция чистых линий с генотипами AA и aa — поколение F_0 (родительские формы).

После скрещивания особей с генотипом AA с особями с генотипом aa поколения F_0 образуется поколение гибридов с генотипом Aa. Это поколение в генетике принято обозначать F_1. В поколении F_1 других генотипов, кроме генотипа Aa, нет.

При случайном скрещивании особей поколения F_1 образуется поколение F_2, в котором одинаково часто встречаются 4 генотипа: AA, Aa, aA, aa.

Экспериментально получен *II закон Менделя*: в поколении F_2 происходит расщепление фенотипов в отношении $3:1$ (3 части составляют особи с доминантным признаком в фенотипе, 1 часть приходится на особи с рецессивным признаком в фенотипе).

Из этого закона следует, что для поколения F_2 вероятность того, что в фенотипе особи проявляется доминантный признак, равна 3/4, а вероятность того, что в фенотипе особи проявится рецессивный признак, равна 1/4.

7. Закон Харди.** Пусть в популяции встречаются три генотипа: AA, Aa, aa. Доля особей генотипа AA равна u, доля особей генотипа Aa равна $2v$ и доля особей генотипа aa равна w. Коротко будем говорить о структуре популяции и записывать ее так:

$$\begin{matrix} AA & Aa & aa \\ u & 2v & w. \end{matrix} \tag{1.19}$$

Под этим мы понимаем следующее: если популяция содержит N особей, то особей генотипа AA в ней uN, особей смешанного генотипа Aa в ней $2vN$ и особей рацессивного генотипа aa в ней wN. При этом, так как

$$uN + 2vN + wN = N,$$

* Фенотип — внешнее проявление признака.

** Об этом законе и других приложениях теории вероятностей в биологии подробнее см., например, в [4].

то

$$u+2v+w=1. \tag{1.20}$$

Подсчитаем число генов A в популяции. Все особи доминантного генотипа имеют $2uN$ генов A (у каждой особи два гена A, и всех особей uN), особи смешанного генотипа имеют $2vN$ генов A (у каждой особи один ген A, и всех особей $2vN$), у особей рецессивного генотипа генов A нет. Следовательно, в популяции число доминантных генов A равно:

$$2uN+2vN=2N(u+v),$$

или, короче, $2Np$, где

$$p=u+v.$$

Число p имеет простой вероятностный смысл — это есть $P(A)$, т. е. вероятность того, что выбранный наудачу ген доминантен. Действительно, доминантных генов $2Np$, и всех генов $2N$ (у каждой особи популяции два гена). Следовательно,

$$P(A)=\frac{2Np}{2N}=p. \tag{1.21}$$

Аналогично подсчитывается, что число всех рецессивных генов a в популяции равно:

$$2Nq,$$

где

$$q=w+v.$$

При этом число q имеет аналогичный вероятностный смысл:

$$P(a)=\frac{2Nq}{2N}=q.$$

Из вероятностного смысла чисел p и q, следует, что

$$p+q=1.$$

(В этом можно убедиться и подстановкой значений p и q.) Заметим, что числа u, $2v$ и w тоже имеют простой вероятностный смысл (подсчет аналогичен проведенному выше подсчету для доминантных генов):

$$P(AA)=\frac{uN}{N}=u,$$

$$P(Aa)=\frac{2vN}{N}=2v,$$

$$P(aa)=\frac{wN}{N}=w.$$

($P(AA)$ — вероятность того, что выбранная наудачу особь имеет генотип AA, аналогично $P(Aa)$ и $P(aa)$.)

Теперь определим, какова будет структура потомства. Пусть потомство имеет структуру:

$$\begin{array}{ccc} AA & Aa & aa \\ u_1 & 2v_1 & w_1 \end{array}$$

(это понимается так же, как и при задании структуры популяции (1.19)). Подсчитаем u_1, $2v_1$ и w_1. Числа u_1, $2v_1$ и w_1 есть вероятности того, что взятый наудачу потомок имеет соответственно генотип AA, Aa и aa (см. соответствующие формулы). Так как скрещивания происходят независимым образом, то вероятность u_1 может рассматриваться как вероятность следующего события· выбрали наудачу и независимым образом из всего запаса два гена A. Так как выбрать каждый ген A можно с вероятностью p (формула (1.21), то в силу теоремы умножения вероятностей независимых событий (§ 1.3, п. 2) интересующая нас вероятность равна p^2, т е.

$$u_1 = p^2.$$

Аналогично для w_1 получаем

$$w_1 = q^2.$$

Вероятность генотипа Aa в популяции потомков складывается из двух возможностей — либо ген A получен от отца, а ген a от матери, либо ген A получен от матери, а ген a от отца — соответствующие вероятности есть pq и qp. Следовательно, вероятность генотипа Aa в популяции потомков равна $2pq$, т. е. $2v_1 = 2pq$. Отсюда

$$v_1 = pq.$$

Следовательно, структура потомства имеет следующий вид:

$$\begin{array}{ccc} AA & Aa & aa \\ p^2 & 2pq & q^2. \end{array}$$

Самое замечательное состоит в том, что если для потомства взять $u_1 + v_1$ и $w_1 + v_1$, как это делалось для родителей, то получим те же самые числа p и q. Действительно, согласно полученным формулам, имеем:

$$u_1 + v_1 = p^2 + pq = p(p+q) = p,$$

$$w_1 + v_1 = q^2 + pq = q(q+p) = q.$$

Так как структура потомства вычислена только с использованием этих сумм, то потомки популяции также будут иметь ту же структуру. При этом говорят, что рассчитанная структура стационарна, т. е. от поколения к поколению не меняется.

Этот замечательный факт, что со второго поколения устанавливается стационарная структура популяции, является непосредственным обобщением второго закона Менделя и называется *законом Харди*.

На практике возможно отклонение, однако для больших популяций закон Харди остается в силе.

Для гороха вероятность получения белой особи равна q^2 (рецессивный признак), вероятность получения красной особи равна $1 - q^2$ (как для противоположного события) и отношение числа красных и белых особей равно $(1 - q^2) : q^2$.

Для описанного в пункте 1 случая $q = \frac{1}{2}$, и мы опять получаем $3 : 1$ (см. II закон Менделя).

Упражнения

1. В ящике имеется 100 яиц, из них 5 некачественных. Наудачу вынимают одно яйцо. Найдите вероятность того, что вынутое яйцо некачественное. [0,05]

2. Брошена игральная кость. Найдите вероятность того, что выпадет четное число очков. [0,5]

3. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100. Найдите вероятность того, что номер первого наудачу извлеченного жетона не содержит цифры 5. [0,81]

4. Набирая номер телефона, абонент забыл последнюю цифру и набрал ее наудачу. Какова вероятность того, что номер набран правильно? [0,1]

5. В партии из 100 деталей отдел технического контроля обнаружил 5 нестандартных деталей. Чему равна относительная частота появления нестандартных деталей? [0,05]

6. В окружность вписан правильный треугольник. В круге наугад ставят точку. Какова вероятность того, что она попадет в треугольник? $\left[3\sqrt{3}/4\pi\right]$

7. При стрельбе по мишени вероятность сделать отличный выстрел равна 0,3, а вероятность выстрела на оценку «хорошо» равна 0,4. Какова вероятность получить за сделанный выстрел оценку не ниже «хорошо»? [0,7]

8. Вероятность того, что человек умрет на 71-м году жизни, равна 0,04 Какова вероятность того, что человек все-таки не умрет на 71-м году? [0,96]

9. Бросается один раз игральная кость. Определите вероятность выпадения 3 или 5 очков. [1/3]

10. В урне 30 шаров: 15 белых, 10 красных и 5 синих. Какова вероятность вынуть цветной шар, если вынимается один шар? [0,5]

11. В денежно-вещевой лотерее на серию в 1000 билетов приходится 120 денежных и 80 вещевых выигрышей. Какова вероятность какого-либо выигрыша на один лотерейный билет? [0,2]

12. В урне 3 белых и 3 черных шара. Из урны дважды вынимают по одному шару, не возвращая их обратно. Найдите вероятность появления белого шара при втором испытании, если при первом испытании был извлечен черный шар. [0,6]

13. В колоде 36 карт. Наудачу вынимаются из колоды 2 карты. Определите вероятность того, что вторым вынут туз, если первым тоже вынут туз. [3/35]

14. В урне 2 белых и 3 черных шара. Из урны вынимают подряд два шара. Найдите вероятность того, что оба шара белые. [0,1]

15. Какова вероятность того, что из колоды в 36 карт будут вынуты подряд два туза? [1/105]

16. Два стрелка стреляют по цели. Вероятность поражения цели первым стрелком при одном выстреле равна 0,8, вторым стрелком — 0,7. Найдите вероятность поражения цели двумя пулями в одном залпе. [0,56]

17. Найдите вероятность одновременного появления герба при одном бросании двух монет. [0,25]

18. Имеется два ящика, содержащих по 10 деталей. В первом ящике 8, во втором 7 стандартных деталей. Из каждого ящика наудачу вынимают по одной детали. Найдите вероятность того, что обе вынутые детали окажутся стандартными. [0,56]

19. В семье двое детей. Принимая события, состоящие в рождении мальчика и девочки равновероятными, найдите вероятность того, что в семье: а) все девочки; б) дети одного пола. [а) 0,25; б) 0,5]

20. Пусть всхожесть семян оценивается вероятностью 0,7. Какова вероятность того, что из двух посеянных семян взойдет какое-либо одно? [0,91]

21. Из колоды в 36 карт наудачу вынимается одна. Какова вероятность того, что будет вынута пика или туз? [1/3]

22. Брошена игральная кость. Найдите вероятность того, что выпадет четное или кратное трем число очков. [2/3]

23. Имеется два набора деталей. Вероятность того, что деталь первого набора стандартна, равна 0,8, а второго — 0,9. Найдите вероятность того, что взятая наудачу деталь (из наудачу взятого набора) — стандартная. [0,85]

24. В первой коробке содержится 20 транзисторов, из них 18 стандартных; во второй коробке — 10 транзисторов, из них 9 стандартных. Из второй коробки наудачу взят транзистор и переложен в первую. Найдите вероятность того, что транзистор, наудачу извлеченный из первой коробки, будет стандартным. [0,9]

25. Студент M может заболеть гриппом (событие A) только в результате либо переохлаждения (событие B), либо контакта с другим больным (событие C). Требуется найти $P(A)$, если $P(B)=0,5$, $P(C)=0,5$, $P_B(A)=0,3$, $P_C(A)=0,1$ при условии несовместимости B и C. [$P(A)=0,2$]

26. В коробке находятся 6 новых и 2 израсходованные батарейки для карманного фонарика. Какова вероятность того, что две вынутые из коробки наудачу батарейки окажутся новыми? [15/28]

27. На трех карточках написаны буквы У, К, Ж. После тщательного перемешивания берут по одной карточке и кладут последовательно рядом. Какова вероятность того, что получится слово «ЖУК»? [1/6]

28. Слово «керамзит» составлено из букв разрезной азбуки. Затем карточки с буквами перемешивают, и из них извлекают по очереди четыре карточки. Какова вероятность, что эти четыре карточки в порядке выхода составят слово «река»? [1/1680]

29. Слово «Константинополь» составлено из букв А, И, К, Л, Н, Н, Н, О, О, О, П, С, Т, Т, Ь. Какова вероятность случайного составления этого слова из перечисленных букв?

$$\left[\frac{1}{15}\cdot\frac{3}{14}\cdot\frac{3}{13}\cdot\frac{1}{12}\cdot\frac{2}{11}\cdot\frac{1}{10}\cdot\frac{2}{9}\cdot\frac{1}{8}\cdot\frac{1}{7}\cdot\frac{1}{6}\cdot\frac{2}{5}\cdot\frac{1}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot 1=\frac{2^3\cdot 3^2}{15!}\right]$$

30. Имеется 12 различных точек. Одна из них обозначена буквой C. Из множества отрезков, концами которых являются любые две из данных 12 точек, наугад выбирается один. Какова вероятность того, что точка C не является концом выбранного отрезка?

$$\left[\frac{5}{6}\right]$$

31. Незнайка сдавал устный экзамен по математике. В каждом билете — по два вопроса. Всего — 25 экзаменационных билетов. Из них 5 билетов Незнайка вообще не учил. В каждом из оставшихся 20 билетов он хотя бы один вопрос выучил, причем в 18-ти билетах Нейзнайка выучил первый вопрос и в 15-ти билетах — второй вопрос. К каждому билету дается задача, а задач Незнайка решать не умеет. Поэтому он может получить удовлетворительную оценку, если вытащит такой билет, оба вопроса которого он знает. Какова вероятность того, что Незнайка сдаст экзамен, если он первым тянет билет?

$$\left[\frac{13}{25}\right]$$

32. Какова вероятность при шести бросаниях игральной кости получить все шесть граней в таком порядке: при первом бросании одно очко, при втором — два и т. д.?

$$\left[\frac{1}{6^6}\right]$$

ГЛАВА II

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 2.1. Дискретные случайные величины

1. Понятие «случайные величины».

Определение 1. *Случайной величиной* называется переменная величина, которая в зависимости от исхода испытания случайно принимает одно значение из множества возможных значений.

Примеры. 1). Число очков, выпавших при однократном бросании игральной кости, есть случайная величина, она может принять одно из значений: 1, 2, 3, 4, 5, 6;

2) прирост массы домашнего животного за месяц есть случайная величина, которая может иметь значение из некоторого числового промежутка;

3) число родившихся мальчиков среди пяти новорожденных есть случайная величина, которая может принять значения 0, 1, 2, 3, 4, 5;

4) расстояние, которое пролетит снаряд при выстреле из орудия, есть случайная величина, возможные значения которой принадлежат некоторому промежутку.

Случайные величины обычно обозначают прописными буквами X, Y, Z, а их возможные значения — соответствующими строчными буквами x, y, z. Например, если случайная величина X имеет три возможных значения, то они будут обозначены так:

$$x_1,\ x_2,\ x_3.$$

Определение 2. Случайная величина, принимающая различные значения, которые можно записать в виде конечной или бесконечной последовательности, называется *дискретной* случайной величиной.

Рассмотрим дискретные случайные величины, множество допустимых значений которых конечно.

Случайные величины из примеров 1) и 3) дискретные.

Определение 3. Случайная величина, которая может принимать все значения из некоторого числового промежутка, называется *непрерывной* случайной величиной.

Случайные величины из примеров 2) и 4) являются непрерывными.

Определение 4. Под *суммой (произведением) случайных величин* X и Y понимают случайную величину $Z=X+Y$ $(Z=XY)$, возможные значения которой состоят из сумм (произведений) каждого возможного значения величины X и каждого возможного значения величины Y.

2. Законы распределения дискретных случайных величин. Рассмотрим дискретную случайную величину X с конечным множеством возможных значений. Величина X считается заданной, если перечислены все ее возможные значения, а также вероятности, с которыми величина X может принимать эти значения. Указанный перечень всех ее возможных значений и их вероятностей называется *законом распределения дискретной случайной величины.* Закон распределения дискретной случайной величины может быть задан с помощью таблицы:

X	x_1	x_2	x_3	...	x_{n-1}	x_n
p	p_1	p_2	p_3	...	p_{n-1}	p_n

В верхней строке выписываются все возможные значения $x_1, x_2, ..., x_n$ величины X, в нижней строке выписываются вероятности $p_1, p_2, ..., p_n$ значений $x_1, x_2, ..., x_n$. Читается таблица следующим образом: случайная величина X может принимать значения x_i с вероятностями p_i $(i=1, 2, ..., n)$.

Так как события $X=x_i$ $(i=1, 2, ..., n)$ образуют полную группу несовместимых событий, то

$$p_1+p_2+...+p_n=1.$$

Пример. В денежной лотерее раньше разыгрывались: 1 выигрыш в 1000 р., 10 выигрышей по 100 р. и 100 выигрышей по 1 р. при общем числе билетов 10 000. Найдем закон распределения случайного выигрыша X для владельца одного лотерейного билета.

Здесь возможные значения для X есть: $x_1=0$, $x_2=1$, $x_3=100$, $x_4=1000$. Вероятности их будут: $p_2=0{,}01$, $p_3=0{,}001$, $p_4=0{,}0001$, $p_1=1-0{,}01-0{,}001-0{,}0001=0{,}9889$. Следовательно, закон распределения выигрыша X может быть задан таблицей:

X	0	1	100	1000
p	0,9889	0,01	0,001	0,0001

В заключение отметим так называемую «механическую» интерпретацию представленной таблицы. Представим себе, что некоторая масса, равная единице, распределена по оси абсцисс так, что в n отдельных точках $x_1, x_2, ..., x_n$ сосредоточены соответственно массы $p_1, p_2, ..., p_n$. Тогда эта таблица описывает систему материальных точек, размещенных на оси абсцисс.

§ 2.2. Математическое ожидание дискретной случайной величины

1. Понятие математического ожидания. Закон распределения полностью задает дискретную случайную величину. Однако часто встречаются случаи, когда закон распределения случайной величины неизвестен. В таких случаях случайную величину изучают по ее числовым характеристикам. Одной из таких характеристик является математическое ожидание.

Пусть некоторая дискретная случайная величина X с конечным числом своих значений задана законом распределения:

X	x_1	x_2	.	x_n
p	p_1	p_2	.	p_n

Определение. *Математическим ожиданием* $M(X)$ дискретной случайной величины X называется сумма произведений всех возможных значений величины X на соответствующие вероятности:

$$M(X)=x_1p_1+x_2p_2+\ldots+x_np_n. \qquad (2.1)$$

Пример. Найдем математическое ожидание выигрыша X в примере из § 2.1 (п. 2).

Используя полученную там таблицу, имеем

$$M(X)=0\cdot 0{,}9889+1\cdot 0{,}01+100\cdot 0{,}001+1000\cdot 0{,}0001=0{,}21 \text{ (руб.)}.$$

Очевидно, $M(X)=21$ коп. есть справедливая стоимость одного лотерейного билета.

Теорема. *Математическое ожидание дискретной случайной величины X приближенно равно среднему арифметическому всех ее значений (при достаточно большом числе испытаний).*

Доказательство. Предположим, что произведено n испытаний, в которых дискретная случайная величина X приняла значения $x_1, \ldots, x_k$ соответственно $m_1, \ldots, m_k$ раз, так что $m_1+\ldots+m_k=n$. Тогда среднее арифметическое всех значений, принятых величиной X, выразится равенством

$$x_{\text{ср}}=\frac{x_1m_1+x_2m_2+\ldots+x_km_k}{n},$$

или

$$x_{\text{ср}}=x_1\frac{m_1}{n}+x_2\frac{m_2}{n}+\ldots x_k\frac{m_k}{n}.$$

Так как коэффициент m_i/n является относительной частотой события «величина X приняла значение x_i» ($i=1, 2, \ldots, k$), то

$$x_{\text{ср}}=x_1p_1^*+x_2p_2^*+\ldots+x_kp_k^*.$$

Из статистического определения вероятности следует, что при достаточно большом числе испытаний $p_i^* \approx p_i$ $(i = 1, 2, ..., k)$. Поэтому

$$x_{\text{ср}} \approx x_1 p_1 + x_2 p_2 + ... + x_k p_k,$$

или

$$x_{\text{ср}} \approx M(X).$$

Таким образом, математическое ожидание случайной величины можно приближенно считать ее *средним значением*, что и делают на практике.

Обратимся теперь к механической интерпретации математического ожидания дискретной случайной величины X. Пусть на оси абсцисс расположены точки с абсциссами $x_1, x_2, ..., x_n$, в которых сосредоточены соответственно массы $p_1, p_2, ..., p_n$, причем $p_1 + p_2 + ... + p_n = 1$. Тогда математическое ожидание $M(X)$, определяемое формулой (2.1), есть не что иное, как абсцисса центра масс данной системы материальных точек.

2. Свойства математического ожидания дискретной случайной величины.

1. *Математическое ожидание* постоянной величины C равно этой величине.*

Постоянную C можно рассматривать как дискретную случайную величину, принимающую лишь одно значение C с вероятностью $p = 1$. Поэтому $M(C) = C \cdot 1 = C$.

2. *Постоянный множитель можно выносить за знак математического ожидания,* т. е. $M(CX) = CM(X)$.

Используя соотношение (2.1), имеем:

$$M(CX) = Cx_1 p_1 + Cx_2 p_2 + ... + Cx_n p_n =$$
$$= C(x_1 p_1 + x_2 p_2 + ... + x_n p_n) = CM(X).$$

3. *Математическое ожидание суммы двух случайных величин X и Y равно сумме их математических ожиданий:*

$$M(X + Y) = M(X) + M(Y).$$

Доказательство. Пусть X и Y имеют законы распределения:

$$\left.\begin{array}{ccccccc} X & x_1 & x_2 & \quad & Y & y_1 & y_2 \\ p & p_1 & p_2 & \quad & p & q_1 & q_2 \end{array}\right\}. \qquad (2.2)$$

Для упрощения доказательства мы ограничиваемся лишь двумя возможными значениями каждой из случайных величин (в общем случае доказательство аналогично).

* В дальнейшем часто ради краткости вместо слов «математическое ожидание» будем писать МО

Составим все возможные значения величины $X+Y$, для чего к каждому возможному значению X прибавим каждое возможное значение Y; получим x_1+y_1, x_1+y_2, x_2+y_1 и x_2+y_2 (их вероятности обозначим соответственно через p_{11}, p_{12}, p_{21} и p_{22}).

Докажем, что $p_{11}+p_{12}=p_1$. Событие, состоящее в том, что X примет значение x_1 (его вероятность равна p_1), влечет за собой событие, состоящее в том, что $X+Y$ примет значение x_1+y_1 или x_1+y_2 (вероятность этого события по теореме сложения вероятностей несовместимых событий (см. § 1.3, п. 1) равна $p_{11}+p_{12}$). Поэтому $p_{11}+p_{12}=p_1$. Аналогично доказываются равенства $p_{11}+p_{21}=q_1$; $p_{21}+p_{22}=p_2$; $p_{12}+p_{22}=q_2$.

Наконец, согласно формуле (2.1), имеем

$$M(X+Y)=(x_1+y_1)p_{11}+(x_1+y_2)p_{12}+(x_2+y_1)p_{21}+(x_2+y_2)p_{22}=$$

$$=x_1(p_{11}+p_{12})+x_2(p_{21}+p_{22})+y_1(p_{11}+p_{21})+y_2(p_{12}+p_{22})=$$

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

Определение. Случайные величины X и Y называются *независимыми*, если закон распределения каждой из них не зависит от того, какое возможное значение приняла другая величина.

Примером двух независимых случайных величин могут служить суммы выигрышей по каждому из двух лотерейных билетов по двум различным денежно-вещевым лотереям. Здесь ставший известным размер выигрыша по билету одной лотереи не влияет на ожидаемый размер выигрыша и соответствующую ему вероятность по билету другой лотереи.

Несколько случайных величин называются независимыми, если закон распределения любой из них не зависит от того, какие возможные значения приняли остальные случайные величины.

4. *Математическое ожидание произведения двух независимых случайных величин X и Y равно произведению их математических ожиданий:*

$$M(XY)=M(X)M(Y).$$

Пусть независимые случайные величины X и Y заданы законами распределения (2.2). Для упрощения выкладок мы ограничиваемся лишь двумя возможными значениями каждой из случайных величин (в общем случае доказательство аналогично).

Составим все возможные значения величины XY: x_1y_1, x_1y_2, x_2y_1, x_2y_2 (их вероятности обозначим соответственно через p_{11}, p_{12}, p_{21}, p_{22}).

По теореме умножения вероятностей независимых событий (см. § 1.3, п. 2) вероятность того, что XY примет значение x_1y_1, равна произведению вероятностей таких событий: X принимает значение x_1, а Y — значение y_1, т. е. $p_{11}=p_1q_1$. Аналогично $p_{12}=p_1q_2$, $p_{21}=p_2q_1$, $p_{22}=p_2q_2$.

Согласно формуле (2.1), получим:

$$M(XY) = x_1 y_1 p_1 q_1 + x_1 y_2 p_1 q_2 + x_2 y_1 p_2 q_1 + x_2 y_2 p_2 q_2 =$$
$$= x_1 p_1 (y_1 q_1 + y_2 q_2) + x_2 p_2 (y_1 q_1 + y_2 q_2) = (x_1 p_1 + x_2 p_2)(y_1 q_1 + y_2 q_2) =$$
$$= M(X) M(Y).$$

Следствием свойств 2 и 3 является свойство 5.

5. *Математическое ожидание разности двух случайных величин X и Y равно разности их математических ожиданий:*

$$M(X - Y) = M(X) - M(Y).$$

Примечание 1. Свойства 3 и 4 имеют место для любого конечного числа случайных величин.

Примечание 2. Если множество возможных значений дискретной случайной величины X бесконечно, то математическое ожидание $M(X)$ определяется суммой числового ряда

$$M(X) = \sum_{k=1}^{\infty} x_k p_k$$

при условии, что этот ряд абсолютно сходится. Перечисленные свойства математического ожидания остаются в силе (см. [2]) и для таких случайных величин.

Пример 1. Найдем математическое ожидание случайной величины $Z = X + 2Y$, если известны математические ожидания случайных величин X и Y: $M(X) = 5$, $M(Y) = 3$.

Используя свойства 3 и 2 математического ожидания, получим.

$$M(Z) = M(X + 2Y) = M(X) + M(2Y) =$$
$$= M(X) + 2M(Y) = 5 + 2 \cdot 3 = 11.$$

Пример 2. Найдем математическое ожидание случайной величины $Z = 2X - Y$, если заданы математические ожидания случайных величин X и Y: $M(X) = 4$; $M(Y) = 6$.

Используя свойства 5 и 2 математического ожидания, получаем

$$M(Z) = M(2X - Y) = M(2X) - M(Y) =$$
$$= 2M(X) - M(Y) = 2 \cdot 4 - 6 = 2.$$

Пример 3. Пусть независимые случайные величины заданы законами распределения:

X	1	2
p	0,2	0,8

Y	0,5	1
p	0,3	0,7

Требуется найти математическое ожидание случайной величины XY.

Сначала найдем математические ожидания каждой из данных величин:

$$M(X) = 1 \cdot 0{,}2 + 2 \cdot 0{,}8 = 1{,}8;$$
$$M(Y) = 0{,}5 \cdot 0{,}3 + 1 \cdot 0{,}7 = 0{,}15 + 0{,}7 = 0{,}85.$$

Случайные величины X и Y независимы, поэтому искомое математическое ожидание

$$M(XY) = M(X)M(Y) = 1{,}8 \cdot 0{,}85 = 1{,}53.$$

§ 2.3. Дисперсия дискретной случайной величины

1. Понятие дисперсии. Математическое ожидание не дает полной характеристики закона распределения случайной величины. Покажем это на примере. Пусть заданы две дискретные случайные величины X и Y своими законами распределения:

X	−2	0	2
p	0,4	0,2	0,4

Y	−100	0	100
p	0,3	0,4	0,3

$$M(X) = -2 \cdot 0{,}4 + 0 \cdot 0{,}2 + 2 \cdot 0{,}4 = 0{,}0;$$
$$M(Y) = -100 \cdot 0{,}3 + 0 \cdot 0{,}4 + 100 \cdot 0{,}3 = 0{,}0.$$

Несмотря на то что МО величин X и Y одинаковы, возможные значения величин X и Y «разбросаны» или «рассеяны» около своих МО (средних значений) по-разному: возможные значения величины X расположены гораздо ближе к своему МО, чем значения величины Y.

Вот еще один пример. При одинаковой средней величине годовых осадков одна местность может быть засушливой и неблагоприятной для сельскохозяйственных работ (нет дождей весной и летом), а другая — благоприятной для ведения сельского хозяйства.

Настоятельным является необходимость введения новой числовой характеристики случайной величины, по которой можно было бы судить о «рассеянии» возможных значений этой случайно величины.

Пусть задана дискретная случайная величина X:

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Определение 1. *Отклонением* случайной величины X от ее математического ожидания $M(X)$ (или просто отклонением случайной величины X) называется случайная величина $X - M(X)$.

Видно, что для того, чтобы отклонение случайной величины X приняло значение $x_1 - M(X)$, достаточно, чтобы случайная величина X приняла значение x_1. Вероятность же этого события равна p_1; следовательно, и вероятность того, что отклонение случайной величины X примет значение $x_1 - M(X)$, также равна p_1. Аналогично обстоит дело и для остальных возможных значений отклонения случайной величины X. Используя это, запишем закон распределения отклонения случайной величины X:

$X - M(X)$	$x_1 - M(X)$	$x_2 - M(X)$	...	$x_n - M(X)$
p	p_1	p_2	...	p_n

Вычислим теперь МО отклонения $X - M(X)$. Пользуясь свойствами 5 и 1 (§ 2.1, п. 2), получим

$$M[X - M(X)] = M(X) - M(X) = 0.$$

Следовательно, справедлива следующая теорема.

Теорема. *Математическое ожидание отклонения $X - M(X)$ равно нулю:*

$$M[X - M(X)] = 0.$$

Из теоремы видно, что с помощью отклонения $X - M(X)$ не удается определить среднее отклонение возможных значений величины X от ее математического ожидания, т. е. оценить степень рассеяния величины X. Это объясняется взаимным погашением положительных и отрицательных возможных значений отклонения. Однако можно освободиться от этого недостатка, если рассматривать квадрат отклонения случайной величины X.

Запишем закон распределения случайной величины $[X - M(X)]^2$ (рассуждения те же, что и в случае случайной величины $X - M(X)$):

$[X - M(X)]^2$	$[x_1 - M(X)]^2$	$[x_2 - M(X)]^2$	.	$[x_n - M(X)]^2$
p	p_1	p_2	...	p_n

Определение 2. *Дисперсией* $D(X)$ дискретной случайной величины X называется математическое отклонение квадрата отклонения случайной величины X от ее математического отклонения (среднего значения):

$$D(X) = M[(X - M(X))^2].$$

Из закона распределения величины $[X - M(X)]^2$ следует, что

$$D(X) = [x_1 - M(X)]^2 p_1 + [x_2 - M(X)]^2 p_2 + ... + [x_n - M(X)]^2 p_n.$$

Пример. Пусть случайная величина X задана своим законом распределения:

X	1	3	6	7	9
p	0,2	0,2	0,3	0,1	0,2

Найдем $D(X)$.

Имеем:

$$M(X) = 1 \cdot 0{,}2 + 3 \cdot 0{,}2 + 6 \cdot 0{,}3 + 7 \cdot 0{,}1 + 9 \cdot 0{,}2 = 5{,}1;$$

$$[x_1 - M(X)]^2 = [1 - 5{,}1]^2 = 16{,}81;$$

$$[x_2 - M(X)]^2 = [3 - 5{,}1]^2 = 4{,}41;$$

$$[x_3 - M(X)]^2 = [6 - 5{,}1]^2 = 0{,}81;$$

$$[x_4 - M(X)]^2 = [7 - 5{,}1]^2 = 3{,}61;$$

$$[x_5 - M(X)]^2 = [9 - 5{,}1]^2 = 15{,}21.$$

Таким образом, закон распределения случайной величины $[X - M(X)]^2$ выразится таблицей:

$[X - M(X)]^2$	16,81	4,41	0,81	3,61	15,21
p	0,2	0,2	0,3	0,1	0,2

Отсюда

$$D(X) = 16{,}81 \cdot 0{,}2 + 4{,}41 \cdot 0{,}2 + 0{,}81 \cdot 0{,}3 + 3{,}61 \cdot 0{,}1 + 15{,}21 \cdot 0{,}2 = 7{,}89.$$

2. Свойства дисперсии дискретной случайной величины.

1. *Дисперсия дискретной случайной величины X равна разности между математическим ожиданием квадрата величины X и квадратом ее математического ожидания:*

$$D(X) = M(X^2) - M^2(X).$$

Действительно, используя свойства математического ожидания, имеем:

$$D(X) = M[(X - M(X))^2] = M[X^2 - 2XM(X) + M^2(X)] =$$

$$= M(X^2) - 2M(X) \cdot M(X) + M^2(X) =$$

$$= M(X^2) - 2M^2(X) + M^2(X) = M(X^2) - M^2(X).$$

С помощью этого свойства и свойств математического ожидания устанавливаются и другие свойства.

2. *Дисперсия постоянной величины C равна нулю.*

Действительно,

$$D(C) = M(C^2) - M^2(C) = C^2 - C^2 = 0.$$

3. *Постоянный множитель можно выносить за знак дисперсии, возводя его в квадрат:*

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

В самом деле,

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

4. *Дисперсия суммы двух независимых случайных величин X и Y равна сумме дисперсий этих величин:*

$$D(X+Y) = D(X) + D(Y).$$

Действительно,

$$D(X+Y) = M[(X+Y)^2] - M^2(X+Y) =$$
$$= M(X^2 + 2XY + Y^2) - [M(X) + M(Y)]^2 = M(X^2) + 2M(X)M(Y) +$$
$$+ M(Y^2) - M^2(X) - 2M(X)M(Y) - M^2(Y) =$$
$$= [M(X^2) - M^2(X)] + [M(Y^2) - M^2(Y)] = D(X) + D(Y).$$

Используя метод математической индукции, это свойство можно распространить и на случай любого конечного числа слагаемых.

Следствием свойств 3 и 4 является следующее свойство.

5. *Дисперсия разности двух независимых случайных величин X и Y равна сумме их дисперсий:*

$$D(X-Y) = D(X) + D(Y).$$

Пример 1. Используя свойство 1 дисперсии, найдем дисперсию случайной величины X, имеющей следующий закон распределения:

X	1	2	3	4	5
p	0,1	0,2	0,3	0,3	0,1

Находим математические ожидания случайной величины X и ее квадрата:

$$M(X) = 1 \cdot 0,1 + 2 \cdot 0,2 + 3 \cdot 0,3 + 4 \cdot 0,3 + 5 \cdot 0,1 =$$
$$= 0,1 + 0,4 + 0,9 + 1,2 + 0,5 = 3,1;$$
$$M(X^2) = 1^2 \cdot 0,1 + 2^2 \cdot 0,2 + 3^2 \cdot 0,3 + 4^2 \cdot 0,3 + 5^2 \cdot 0,1 =$$
$$= 0,1 + 0,8 + 2,7 + 4,8 + 2,5 = 10,9.$$

Отсюда в силу свойства 1 дисперсии

$$D(X) = 10,9 - (3,1)^2 = 10,9 - 9,61 = 1,29.$$

П р и м е р 2. Дисперсия случайной величины X равна 3. Найдем дисперсию следующих величин: а) $-3X$; б) $4X+3$.

Согласно свойствам 2, 3 и 4 дисперсии, имеем:

$$\text{а)}\ D(-3X) = 9D(X) = 9 \cdot 3 = 27;$$

$$\text{б)}\ D(4X+3) = D(4X) + D(3) = 16D(X) + 0 = 16 \cdot 3 = 48.$$

П р и м е ч а н и е. Если множество возможных значений дискретной случайной величины X бесконечно, то ее дисперсия определяется суммой сходящегося числового ряда

$$D(X) = \sum_{k-1}^{\infty} [x_k - M(X)]^2 p_k.$$

3. Среднее квадратическое отклонение.

О п р е д е л е н и е. *Средним квадратическим отклонением* $\sigma(X)$ случайной величины X называется корень квадратный из ее дисперсии:

$$\sigma(X) = \sqrt{D(X)}.$$

Необходимость введения среднего квадратического отклонения объясняется тем, что дисперсия измеряется в квадратных единицах относительно размерности самой случайной величины. Например, если возможные значения некоторой случайной величины измеряются в метрах, то ее дисперсия измеряется в квадратных метрах. В тех случаях, когда нужно иметь числовую характеристику рассеяния возможных значений в той же размерности, что и сама случайная величина, и используется среднее квадратическое отклонение.

П р и м е р. Случайная величина X — число очков, выпавших при однократном бросании игральной кости. Определим $\sigma(X)$. Имеем:

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3\ \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5;$$

$$D(X) = (1-3{,}5)^2 \cdot \frac{1}{6} + (2-3{,}5)^2 \cdot \frac{1}{6} + (3-3{,}5)^2 \cdot \frac{1}{6} +$$

$$+ (4-3{,}5)^2 \cdot \frac{1}{6} + (5-3{,}5)^2\ \frac{1}{6} + (6-3{,}5)^2 \cdot \frac{1}{6} = \frac{35}{12} \simeq 2{,}92;$$

$$\sigma(X) = \sqrt{\frac{35}{12}} \approx 1{,}71.$$

4. Понятие о моментах распределения.

О п р е д е л е н и е 1. *Начальным моментом порядка* k случайной величины X называется математическое ожидание случайной величины X^k, где k — натуральное число:

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

Следовательно, если X имеет распределение

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

то

$$\nu_k = x_1^k p_1 + x_2^k p_2 + \ldots + x_n^k p_n.$$

Математическое ожидание и дисперсию случайной величины X можно выразить через начальные моменты первого и второго порядков:

$$\left.\begin{aligned} M(X) &= \nu_1; \\ D(X) &= M(X^2) - M^2(X) = \nu_2 - \nu_1^2. \end{aligned}\right\} \tag{2.3}$$

Определение 2. *Центральным моментом порядка k* случайной величины X называется математическое ожидание величины $[X - M(X)]^k$:

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

Из определения 2, согласно установленной выше теореме (п. 1) и определения дисперсии, следует, что

$$\left.\begin{aligned} \mu_1 &= M[X - M(X)] = 0; \\ \mu_2 &= M[(X - M(X))^2] = D(X). \end{aligned}\right\} \tag{2.4}$$

Сравнивая соотношения (2.3) и (2.4), получим

$$\mu_2 = \nu_2 - \nu_1^2. \tag{2.5}$$

Пример 1. Пусть дискретная случайная величина X задана законом распределения:

X	1	3
p	0,4	0,6

Требуется найти начальные моменты первого и второго порядков. Найдем начальный момент первого порядка:

$$\nu_1 = M(X) = 1 \cdot 0{,}4 + 3 \cdot 0{,}6 = 2{,}2.$$

Запишем закон распределения величины X^2:

X^2	1	9
p	0,4	0,6

Найдем начальный момент второго порядка:

$$\nu_2 = M(X^2) = 1 \cdot 0{,}4 + 9 \cdot 0{,}6 = 5{,}8.$$

Пример 2. Пусть дискретная случайная величина X задана законом распределения, приведенным в предыдущем примере. Найдем центральный момент второго порядка.

Как установлено в предыдущем примере, $\nu_1=2{,}2$ и $\nu_2=5{,}8$. Поэтому, согласно формуле (2.5),

$$\mu_2 = 5{,}8 - 2{,}2^2 = 5{,}8 - 4{,}84 = 0{,}96.$$

§ 2.4. Основные законы распределения дискретных случайных величин

1. Биномиальное распределение. Пусть осуществляется n испытаний, причем вероятность появления события A в каждом испытании равна p и не зависит от исхода других испытаний (независимые испытания). Такая последовательность испытаний называется *схемой Бернулли**. Так как вероятность наступления события A в одном испытании равна p, то вероятность его ненаступления равна $q=1-p$.

Найдем вероятность того, что при n испытаниях событие A наступит m раз ($m \leqslant n$).

Пусть событие A наступило в первых m испытаниях m раз и не наступило во всех последующих испытаниях. Это сложное событие можно записать в виде произведения:

$$\underbrace{AA\ldots A}_{m \text{ раз}}\ \underbrace{\bar{A}\bar{A}\ldots\bar{A}}_{n-m \text{ раз}}.$$

Общее число сложных событий, в которых событие A наступает m раз, равно числу сочетаний из n элементов по m элементов. При этом вероятность каждого сложного события оказывается равной $p^m q^{n-m}$. Так как указанные сложные события являются несовместимыми, то вероятность их суммы равна сумме их вероятностей. Итак, если $P_n(m)$ есть вероятность появления события A m раз в n испытаниях, то

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

или

$$P_n(m) = \frac{n!}{m!(n-m)!} p^m q^{n-m}. \tag{2.6}$$

Формула (2.6) называется *формулой Бернулли.*

Пример 1. Пусть всхожесть семян данного растения составляет 90%. Найдем вероятность того, что из четырех посеянных семян взойдут: а) три; б) не менее трех.

а) В данном случае $n=4$, $m=3$, $p=0{,}9$, $q=1-p=0{,}1$.

* Якоб Бернулли (1654—1705) — швейцарский математик.

Применяя формулу Бернулли, получим

$$P_4(3) = \frac{4!}{3!1!}(0{,}9)^3 \cdot 0{,}1 = 0{,}2916.$$

б) Искомое событие A состоит в том, что из четырех семян взойдут или три, или четыре. По теореме сложения вероятностей $P(A) = P_4(3) + P_4(4)$. Но $P_4(4) = (0{,}9)^4 = 0{,}6561$. Поэтому $P(A) = 0{,}2916 + 0{,}6561 = 0{,}9477$.

Снова рассмотрим n независимых испытаний, в каждом из которых наступает событие A с вероятностью p. Обозначим через X случайную величину, равную числу появлений события A в n испытаниях.

Понятно, что событие A может вообще не наступить, наступить 1 раз, 2 раза и т. д. и, наконец, наступить n раз. Следовательно, возможными значениями величины X будут числа $0, 1, 2, \ldots, n-1, n$.

По формуле Бернулли можно найти вероятности этих значений:

$$P_n(0) = C_n^0 q^n = q^n;$$

$$P_n(1) = C_n^1 q^{n-1} p;$$

$$\ldots$$

$$P_n(n) = p^n.$$

Запишем полученные данные в виде таблицы распределения:

X	0	1	.	m	. .	n
P	q^n	$C_n^1 pq^{n-1}$	.	$C_n^m p^m q^{n-m}$	.	p^n

Построенный закон распределения дискретной случайной величины X называется *законом биномиального распределения.*

Найдем $M(X)$ для биноминального распределения. Очевидно, что X_i — число появлений события A в каждом испытании — представляет собой случайную величину со следующим распределением:

X_i	0	1
p_i	q	p

Поэтому $M(X_i) = 0 \cdot q + 1 \cdot p = p$. Но так как $X = X_1 + \ldots + X_n$, то $M(X) = np$.

Найдем далее $D(X)$ и $\sigma(X)$. Так как величина X_i^2 имеет распределение

X_i^2	0^2	1^2
p_i	q	p

то $M(X_i^2) = 0^2 q + 1^2 p = p$. Поэтому

$$D(X_i) = M(X_i^2) - M^2(X_i) = p - p^2 = p(1-p) = pq.$$

Наконец, в силу независимости величин $X_1, X_2, \ldots, X_n$

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

Отсюда для биномиального распределения

$$\sigma(X) = \sqrt{npq}.$$

Пример 2. Монета брошена 2 раза. Напишем в виде таблицы закон распределения случайной величины X — числа выпадений герба.

Вероятность появления герба в каждом бросании монеты $p = 1/2$. Следовательно, вероятность непоявления герба $q = 1 - 1/2 = 1/2$. При двух бросаниях монеты герб может либо совсем не появиться, либо появиться 1 раз, либо появиться 2 раза. Таким образом, возможные значения X таковы: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$. Найдем вероятность этих возможных значений по формуле Бернулли:

$$P_2(0) = \frac{2!}{2!}\left(\frac{1}{2}\right)^2 = \frac{1}{4};$$

$$P_2(1) = \frac{2!}{1!1!} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2};$$

$$P_2(2) = \frac{2!}{2!}\left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Тогда искомый закон распределения будет иметь вид:

X	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Пример 3. (Первая игра де Мерэ*.) Игральная кость бросается четыре раза. Рыцарь бился об заклад, что при этом хотя бы один раз выпадет шесть очков. Какова вероятность выигрыша для рыцаря?

Решение. Вероятность выпадения шестерки при одном бросании игральной кости равна $\frac{1}{6}$ (см. § 1.1, п. 3, пример 4). Значит, вероятность ее невыпадения при одном бросании равна $1 - \frac{1}{6} = \frac{5}{6}$ (противоположное событие). Тогда вероятность невыпадения шестерки при четырех бросаниях, согласно формуле Бернулли,

$$P_4(4) = \frac{4!}{4!(4-4)!}\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)^{4-4} = \left(\frac{5}{6}\right)^4 = \frac{625}{1296} \simeq 0{,}48.$$

Значит, вероятность выигрыша для рыцаря есть

$$1 - \frac{625}{1296} = \frac{671}{1296} \simeq 0{,}518.$$

---

* Страстный игрок в кости рыцарь де Мерэ хотел разбогатеть при помощи этой игры, и для этого он придумывал различные усложненные правила игры.

Это значит, что чем больше рыцарь играл, тем больше он выигрывал. Оказываясь постоянно в проигрыше, противники рыцаря перестали играть с ним по этим правилам.

Пример 4. (Вторая игра де Мерэ.) Две игральные кости бросаются 24 раза. Рыцарь бился об заклад, что при этом хотя бы один раз выпадут две шестерки. Какова вероятность проигрыша для рыцаря?

Решение. При одном одновременном бросании двух игральных костей вероятность выпадения двух шестерок равна $\frac{1}{36}$, а вероятность того, что не выпадут две шестерки, равна $\frac{35}{36}$. Тогда вероятность того, что при 24-х одновременных бросаниях двух игральных костей ни разу не выпадут две шестерки, согласно формуле Бернулли,

$$P_{24}(24) = \frac{24!}{24!(24-24)}\left(\frac{35}{36}\right)^{24}\left(\frac{1}{36}\right)^{24-24} = \left(\frac{35}{36}\right)^{24} \approx 0{,}509,$$

т. е. вероятность проигрыша для рыцаря была больше $\frac{1}{2}$. Это значит, что чем больше рыцарь будет играть, тем больше он будет проигрывать. Когда так и случилось, рыцарь разорился и обратился к Паскалю за разъяснениями. Паскаль успешно раскрыл математические тайны правил двух игр рыцаря де Мерэ.

Пример 5. Применяемый метод лечения приводит к выздоровлению в 80% случаев. Какова вероятность того, что из 5 больных выздоровят 4?

В данном случае $p=0{,}8$, $q=1-p=0{,}2$, $n=5$, $m=4$. Поэтому по формуле Бернулли

$$P_5(4) = \frac{5!}{4!(5-4)!}(0{,}8)^4(0{,}2)^{5-4} = \frac{5\cdot 8^4\cdot 2}{10^5} = \frac{8^4}{10^4} = 0{,}4096 \approx 0{,}41.$$

Пример 6. В условии предыдущего примера найдем вероятность того, что из 5 больных выздоровят не менее 4.

Искомая вероятность есть сумма вероятностей $P_5(4)+P_5(5)$. Имеем:

$$P_5(4)+P_5(5) = 0{,}4096 + (0{,}8)^5 = 0{,}4096 + 0{,}32768 = 0{,}73728 \approx 0{,}74.$$

Задача об экстрасенсе. Обычный человек примерно в половине случаев правильно угадывает, в какой руке спрятан мелкий предмет.

Предположим, что верный ответ получен в трех случаях из четырех. Случайно ли это? Или при таком результате можно говорить о необычных способностях угадывающего?

Если принять вероятность угадывания в норме $p=\frac{1}{2}$, то по формуле Бернулли

$$P_4(3) = C_4^3p^3q,$$

где $q=1-p$,

или

$$P_4(3) = 4 \cdot \left(\frac{1}{2}\right)^3 \cdot \frac{1}{2} = \frac{1}{4}.$$

Как видим, каждый четвертый нормальный человек правильно угадывает в трех случаях из четырех.

Допустим, что верный ответ получен в девяти случаях из десяти. Какова вероятность такого угадывания у нормального человека?

По формуле Бернулли

$$P_{10}(9) = C_{10}^{9}\left(\frac{1}{2}\right)^9 \cdot \frac{1}{2} = C_{10}^{1} \cdot \frac{1}{2^{10}} = 10 \cdot \frac{1}{2^{10}} \approx 0{,}01.$$

Таким образом, нормальный человек лишь в одном случае из 100 может случайно продемонстрировать такой результат. И если подобное угадывание происходит чаще, то можно, по-видимому, говорить, что угадыватель — экстрасенс (или мистификатор).

2. Распределение Пуассона*. Пусть проводится серия n независимых испытаний ($n = 1, 2, 3, \ldots$), причем вероятность появления данного события A в этой серии $P(A) = p_n > 0$ зависит от ее номера n и стремится к нулю при $n \to \infty$ (последовательность «редких событий»). Предположим, что для каждой серии среднее значение числа появлений события A постоянно, т. е.

$$np_n = \mu = \text{const}.$$

Отсюда

$$p_n = \mu/n.$$

На основании формулы Бернулли (2.6) для вероятности появления события A в n-й серии ровно m раз имеет место формула

$$P_n(m) = C_n^m p_n^m (1 - p_n)^{n-m} = C_n^m \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{n-m}.$$

Пусть m фиксировано. Тогда

$$\lim_{n\to\infty} C_n^m \left(\frac{\mu}{n}\right)^m = \lim_{n\to\infty} \frac{n(n-1)(n-2)\ldots[n-(m-1)]}{m!\,n^m}\mu^m =$$

$$= \frac{\mu^m}{m!}\lim_{n\to\infty}\left[1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{m-1}{n}\right)\right] = \frac{\mu^m}{m!};$$

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n-m} = \lim_{n\to\infty}\left\{\left[\left(1-\frac{\mu}{n}\right)^{-\frac{n}{\mu}}\right]^{-\mu}\left(1-\frac{\mu}{n}\right)^{-m}\right\} = e^{-\mu}\cdot 1 = e^{-\mu}$$

(здесь использован второй замечательный предел $\lim\limits_{x\to\infty}(1+1/x)^x = e$).

* Симеон Пуассон (1781—1840) — французский физик и математик.

Поэтому

$$\lim_{n \to \infty} P_n(m) = \frac{\mu^m}{m!} e^{-\mu}.$$

Если n велико, то в силу определения предела вероятность $P_n(m)$ сколь угодно мало отличается от $\frac{1}{m!}\mu^m e^{-\mu}$. Отсюда при больших n для искомой вероятности $P_n(m)$ имеем приближенную *формулу Пуассона* (для простоты знак приближенного равенства опущен).

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

где $\mu = np_n$.

Пример. Завод отправил на базу 500 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,002. Найдем вероятность того, что на базу придут три негодных изделия.

По условию $n = 500$, $p_n = 0{,}002$, $m = 3$. Поэтому $\mu = 500 \cdot 0{,}002 = 1$ и искомая вероятность

$$P_{500}(3) = \frac{1}{3!} e^{-1} \approx 0{,}06.$$

Определение. Говорят, что случайная величина X распределена по *закону Пуассона*, если эта величина задана таблицей

X	0	1	2	3	
P	$e^{-\mu}$	$\mu e^{-\mu}$	$\frac{\mu^2}{2!} e^{-\mu}$	$\frac{\mu^3}{3!} e^{-\mu}$	

где μ — фиксированное положительное число (разным значениям μ отвечают разные распределения Пуассона).

Полезно проверить, что для приведенной таблицы сумма всех вероятностей равна единице. Действительно, с учетом известного разложения для e^x имеем

$$e^{-\mu} + \mu e^{-\mu} + \frac{\mu^2}{2!} e^{-\mu} + \ldots = e^{-\mu}\left(1 + \mu + \frac{\mu^2}{2!} + \ldots\right) = e^{-\mu} e^{\mu} = 1.$$

Распределение Пуассона заслуживает особого внимания, так как из всех дискретных распределений оно наиболее часто встречается в приложениях.

Найдем математическое ожидание дискретной величины X, распределенной по закону Пуассона. Согласно определению математического ожидания (§ 2.2, п. 2, примечание 2), имеем

$$M(X) = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} e^{\mu} = \mu.$$

Таким образом, параметр μ в распределении Пуассона есть не что иное, как математическое ожидание величины X.

Найдем далее $D(X)$. Сначала найдем начальный момент второго порядка (§ 2.3, п. 4):

$$\nu_2 = M(X^2).$$

Запишем закон распределения величины X^2:

X^2	0	1	4	9	.
P	$e^{-\mu}$	$\mu e^{-\mu}$	$\frac{\mu^2}{2!}e^{-\mu}$	$\frac{\mu^3}{3!}e^{-\mu}$	...

Отсюда

$$\nu_2 = M(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} = \mu e^{-\mu} \sum_{k=1}^{\infty} k \frac{\mu^{k-1}}{(k-1)!} =$$

$$= \mu e^{-\mu} \left[\sum_{k=1}^{\infty} k \frac{\mu^{k-1}}{(k-1)!} - \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} \right] =$$

$$= \mu e^{-\mu} \left[\sum_{k=2}^{\infty} (k-1) \frac{\mu^{k-1}}{(k-1)!} + e^{\mu} \right] = \mu e^{-\mu} \left[\mu \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} + e^{\mu} \right] =$$

$$= \mu^2 (e^{-\mu} e^{\mu}) + \mu (e^{-\mu} e^{\mu}) = \mu^2 + \mu.$$

Затем по известной формуле (§ 2.3, п. 4) вычислим дисперсию

$$D(X) = \nu_2 - \nu_1^2 = \mu^2 + \mu - \mu^2 = \mu.$$

Задача. (Редкие болезни.) Многие болезни достаточно редки или становятся таковыми после принятия профилактических и лечебных мер. Однако даже при самых благоприятных условиях в больших популяциях все же встречается некоторое число больных редкими болезнями. Например, при введении вакцины против полиомиелита иммунитет создается в 99,99% случаев. Какова вероятность того, что из 10 000 вакцинированных детей заболеет один?

Вероятность заболеть $p = 1 - 0{,}9999 = 0{,}0001$, число испытаний $n = 10\,000$. Поэтому $\mu = 0{,}0001 \cdot 10\,000 = 1$, и по формуле Пуассона имеем

$$P_{10\,000}(1) = \frac{1}{1!} e^{-1} = 0{,}368.$$

Аналогично, вероятность, что заболеют 2 ребенка

$$P_{10\,000}(2) = \frac{1^2}{2!} e^{-1} = 0{,}184,$$

а вероятности заболевания 3 и 4 детей соответственно равны

$$P_{10\,000}(3) = \frac{e^{-1}}{3!} \approx 0{,}061;$$

$$P_{10\,000}(4) = \frac{e^{-1}}{4!} = 0{,}015.$$

§ 2.5. Непрерывные случайные величины

1. Интегральная функция распределения. Для непрерывной случайной величины в отличие от дискретной нельзя построить таблицу распределения. Поэтому непрерывные случайные величины изучаются другим способом, который мы и будем рассматривать.

Пусть X — непрерывная случайная величина с возможными значениями из некоторого интервала $(a;\ b)$ и x — действительное число. Под выражением $X<x$ понимается событие «случайная величина X приняла значение, меньшее x». Вероятность этого события $P(X<x)$ есть некоторая функция переменной x:

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

Определение. *Интегральной функцией распределения* (или кратко *функцией распределения*) непрерывной случайной величины X называется функция $F(x)$, равная вероятности того, что X приняла значение, меньшее x:

$$F(x) = P(X<x). \tag{2.7}$$

$F(x)$ — это геометрический смысл этого равенства: вероятность того, что случайная величина X примет значение, которое изображается на числовой оси точкой, лежащей левее точки x.

Отметим, что функция распределения совершенно так же определяется для дискретных случайных величин.

Укажем свойства, которыми обладает функция $F(x)$.

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

Это свойство следует из того, что $F(x)$ есть вероятность.

2. $F(x)$ — *неубывающая функция*, т. е. если $x_1<x_2$, то $F(x_1) \leqslant F(x_2)$.

Доказательство. Предположим, что $x_1<x_2$. Событие «X примет значение, меньшее x_2» можно представить в виде суммы двух несовместимых событий: «X примет значение, меньшее x_1» и «X примет значение, удовлетворяющее неравенствам $x_1 \leqslant X<x_2$». Обозначим вероятности последних двух событий соответственно через $P(X<x_1)$ и $P(x_1 \leqslant X<x_2)$. По теореме о вероятности суммы двух несовместимых событий имеем

$$P(X<x_2) = P(X<x_1) + P(x_1 \leqslant X<x_2),$$

откуда с учетом равенства (2.7)

$$P(x_1 \leqslant X<x_2) = F(x_2) - F(x_1). \tag{2.8}$$

Так как вероятность любого события есть число неотрицательное, то $P(x_1 \leqslant X<x_2) \geqslant 0$ и, значит,

$$F(x_2) \geqslant F(x_1).$$

Формула (2.8) утверждает свойство 3.

3. *Вероятность попадания случайной величины X в полуинтервал $[a;\ b)$ равна разности между значениями функции распределения в правом и левом концах интервала $(a;\ b)$:*

$$P(a \leqslant X < b) = F(b) - F(a). \tag{2.9}$$

В частности, в случае полуинтервала $[x;\ x+\Delta x)$

$$P(x \leqslant X < x+\Delta x) = F(x+\Delta x) - F(x). \tag{2.9'}$$

Пример. Пусть случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0 \text{ при } x \leqslant -1; \\ \frac{x}{4} + \frac{1}{4} \text{ при } -1 < x \leqslant 3; \\ 1 \text{ при } x > 3. \end{cases}$$

Найдем вероятность того, что в результате испытания X **примет** значение, принадлежащее полуинтервалу $[0;\ 2)$.

Так как на полуинтервале $[0;\ 2)$ $F(x) = \frac{x}{4} + \frac{1}{4}$, то

$$P(0 \leqslant X < 2) = F(2) - F(0) = \frac{1}{2} + \frac{1}{4} - \frac{1}{4} = \frac{1}{2}.$$

В дальнейшем случайную величину будем называть *непрерывной*, если ее функция распределения непрерывна с непрерывной или кусочно-непрерывной производной.

4. *Вероятность того, что непрерывная случайная величина X примет какое-либо заранее заданное значение, равна нулю:*

$$P(X = x_1) = 0. \tag{2.10}$$

Доказательство. Положив в (2.9′) $x_2 = x_1 + \Delta x$, будем иметь

$$P(x_1 \leqslant X < x_1 + \Delta x) = F(x_1 + \Delta x) - F(x_1). \tag{2.11}$$

Так как $F(x)$ — непрерывная функция, то, перейдя в (2.11) к пределу при $\Delta x \to 0$, получим искомое равенство (2.10).

Из свойства 4 следует свойство 5.

5. *Вероятности попадания непрерывной случайной величины в интервал, сегмент и полуинтервал с одними и теми же концами одинаковы:*

$$P(a < X < b) = P(a \leqslant X \leqslant b) = P(a \leqslant X < b) = P(a < X \leqslant b). \tag{2.12}$$

6. *Если возможные значения случайной величины X принадлежат интервалу $(a;\ b)$, то:* 1) $F(x) = 0$ *при* $x \leqslant a$; 2) $F(x) = 1$ *при* $x \geqslant b$.

Доказательство. 1) Пусть $x_1 \leqslant a$. Тогда событие $X < x_1$ невозможно, и, следовательно, вероятность его равна нулю. 2) Пусть $x_2 \geqslant b$. Тогда событие $X < x_2$ достоверно, и, следовательно, вероятность его равна 1.

С л е д с т в и е. *Если возможные значения непрерывной случайной величины расположены на всей числовой оси, то справедливы следующие предельные соотношения:*

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0; \quad F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

2. Дифференциальная функция распределения. *Дифференциальной функцией распределения* непрерывной случайной величины X (или ее *плотностью вероятности*, или ее *плотностью распределения*) называется функция $f(x)$, равная производной интегральной функции

$$f(x) = F'(x).$$

Так как $F(x)$ — неубывающая функция, то $f(x) \geqslant 0$.

Из равенства (2.9′) с учетом неравенства $F(x+\Delta x) - F(x) \approx F'(x)\Delta x$, справедливого для малых $|\Delta x|$, и свойства 5 (п. 1) имеем

$$P(x < X < x + \Delta x) \approx F'(x)\Delta x$$

или

$$P(x < X < x + \Delta x) \approx f(x)\Delta x$$

(для малых Δx), т. е. вероятность попадания случайной величины X в интервал $(x;\ x + \Delta x)$ при малых Δx приближенно равна произведению ее плотности вероятности в точке x на длину этого итервала.

Имеет место и следующая теорема.

Т е о р е м а. *Вероятность попадания непрерывной случайной величины X в интервал $(a;\ b)$ равна определенному интегралу от ее плотности вероятности, взятому в пределах от a до b:*

$$P(a < X < b) = \int_a^b f(x)\,dx. \qquad (2.13)$$

Д о к а з а т е л ь с т в о. Так как $F(x)$ является первообразной для $f(x)$, то на основании формулы Ньютона — Лейбница имеем

$$\int_a^b f(x)\,dx = F(b) - F(a). \qquad (2.14)$$

Теперь с учетом соотношений (2.9), (2.12), (2.14) получим искомое равенство.

Из (2.13) следует, что геометрически вероятность $P(a < X < b)$ представляет собой площадь криволинейной трапеции, ограниченной графиком плотности вероятности $y = f(x)$ и отрезками прямых $y = 0$, $x = a$ и $x = b$.

С л е д с т в и е. В частности, *если $f(x)$ — четная функция и концы интервала симметричны относительно начала координат, то*

$$P(-a < X < a) = P(|X| < a) = 2\int_0^a f(x)\,dx. \qquad (2.15)$$

Действительно,

$$\int_{-a}^{a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_{0}^{a} f(x)\,dx = -\int_{a}^{0} f(-x)\,dx + \int_{0}^{a} f(x)\,dx =$$

$$= \int_{0}^{a} f(x)\,dx + \int_{0}^{a} f(x)\,dx = 2\int_{0}^{a} f(x)\,dx.$$

Пример 1. Пусть задана плотность вероятности случайной величины X

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 2x & \text{при } 0 \leqslant x \leqslant 1; \\ 0 & \text{при } x > 1. \end{cases}$$

Найдем вероятность того, что в результате испытания X примет значение, принадлежащее интервалу (0,5; 1).

Согласно формуле (2.13), искомая вероятность

$$P(0{,}5 < X < 1) = 2\int_{0,5}^{1} x\,dx = x^2 \Big|_{0,5}^{1} = 0{,}75.$$

Заменяя в формуле (2.14) a на $-\infty$ и b на x, получим

$$F(x) - F(-\infty) = \int_{-\infty}^{x} f(x)\,dx,$$

откуда в силу приведенного выше следствия (п. 1)

$$F(x) = \int_{-\infty}^{x} f(x)\,dx. \tag{2.16}$$

Выражение (2.16) позволяет найти интегральную функцию распределения $F(x)$ по ее плотности вероятности.

Заметим, что из формулы (2.16) и отмеченного следствия вытекает, что

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{2.17}$$

Пример 2. Пусть плотность вероятности случайной величины X задана так:

$$f(x) = \frac{A}{1 + x^2}\,(-\infty < x < +\infty).$$

Требуется найти коэффициент A, функцию распределения $F(x)$ и вероятность попадания случайной величины X в интервал (0; 1).

Коэффициент A найдем, воспользовавшись соотношением (2.17). Так как

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{+\infty} \frac{A\,dx}{1+x^2} = \int_{-\infty}^{0} \frac{A\,dx}{1+x^2} + \int_{0}^{+\infty} \frac{A\,dx}{1+x^2} =$$

$$= A \operatorname{arctg} x \,\Big|_{-\infty}^{0} + A \operatorname{arctg} x \,\Big|_{0}^{+\infty} = A[\operatorname{arctg}(+\infty) - \operatorname{arctg}(-\infty)] = A\pi,$$

то $A\pi = 1$, откуда $A = 1/\pi$.

Применяя формулу (2.16), получим функцию распределения $F(x)$:

$$F(x) = \int_{-\infty}^{x} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi} \operatorname{arctg} x \,\Big|_{-\infty}^{x} =$$

$$= \frac{1}{\pi}[\operatorname{arctg} x - \operatorname{arctg}(-\infty)] = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x.$$

Наконец, формулы (2.9) и (2.12) с учетом найденного значения функции $F(x)$ дают

$$P(0 < X < 1) = F(1) - F(0) = 0{,}25.$$

§ 2.6. Математическое ожидание и дисперсия непрерывной случайной величины

Пусть непрерывная случайная величина X задана плотностью вероятности $f(x)$. Допустим, что все возможные значения X принадлежат отрезку $[a;\ b]$. Точками $x_0 = a < x_1 < x_2 \ldots < x_{n-1} < x_n = b$ разобьем его на n частичных отрезков, длины которых обозначим через Δx_1, Δx_2, ..., Δx_n. Наибольшую из этих длин обозначим через λ.

Предполагая определить математическое ожидание непрерывной случайной величины по аналогии с дискретной, составим сумму

$$\sum_{k=1}^{n} x_k f(x_k) \Delta x_k$$

[напомним, что произведение $f(x_k)\Delta x_k$ при малых Δx_k приближенно равно вероятности попадания случайной величины X в интервал $(x_k;\ x_k + \Delta x_k)$ см. § 2.5, п. 2]. Перейдя в этой сумме к пределу при $\lambda \to 0$, получим определенный интеграл $\int_a^b x f(x)\,dx$, который и называют *математическим ожиданием непрерывной случайной величины* X, все возможные значения которой принадлежат отрезку $[a;\ b]$:

$$M(X) = \int_a^b x f(x)\,dx.$$

Если возможные значения непрерывной случайной величины X принадлежат всей числовой оси, то математическое ожидание определяется интегралом

$$M(X)=\int_{-\infty}^{+\infty} x f(x)\,dx. \tag{2.18}$$

При этом предполагается, что несобственный интеграл сходится абсолютно, т. е. интеграл $\int_{-\infty}^{+\infty}|x|f(x)\,dx$ существует.

По аналогии с дисперсией дискретной случайной величины определяется и дисперсия непрерывной случайной величины.

Дисперсией непрерывной случайной величины X называют математическое ожидание квадрата ее отклонения.

Если все возможные значения X принадлежат отрезку $[a;\ b]$, то

$$D(X)=\int_a^b [x-M(X)]^2 f(x)\,dx,$$

если возможные значения X принадлежат всей числовой оси, то

$$D(X)=\int_{-\infty}^{+\infty}[x-M(X)]^2 f(x)\,dx \tag{2.19}$$

при условии, что последний несобственный интеграл сходится.

Заметим, что свойства математического ожидания и дисперсии дискретных случайных величин сохраняются и для непрерывных случайных величин.

Наконец, для непрерывной случайной величины X *среднее квадратическое отклонение* $\sigma(X)$ определяется, как и для дискретной величины, формулой $\sigma(X)=\sqrt{D(X)}$.

Пример. Пусть случайная величина X задана плотностью вероятности

$$f(x)=\begin{cases}0 & \text{при } x<0;\\ \frac{x}{2} & \text{при } 0\leqslant x\leqslant 2;\\ 0 & \text{при } x>2.\end{cases}$$

Определим математическое ожидание, дисперсию и среднее квадратическое отклонение величины X.

Согласно формулам (2.18) и (2.19), имеем:

$$M(X)=\int_{-\infty}^{+\infty} x f(x)\,dx=\frac{1}{2}\int_0^2 x^2dx=\frac{x^3}{6}\Big|_0^2=4/3\simeq 1{,}33.$$

$$D(X)=\int_{-\infty}^{+\infty}\left(x-\frac{4}{3}\right)^2 f(x)\,dx=\int_0^2\left(x-\frac{4}{3}\right)^2\frac{1}{2}x\,dx=$$

$$=\frac{1}{2}\int_0^2\left(x^2-\frac{8}{3}x+\frac{16}{9}\right)x\,dx=\frac{1}{2}\int_0^2\left(x^3-\frac{8}{3}x^2+\frac{16}{9}x\right)dx=$$

$$=\frac{1}{2}\left(\frac{x^4}{4}-\frac{8}{9}x^3+\frac{8}{9}x^2\right)\Bigg|_0^2=\frac{1}{2}\left(4-\frac{64}{9}+\frac{32}{9}\right)=\frac{2}{9}\simeq 0{,}22$$

и наконец,

$$\sigma(X)=\sqrt{D(X)}=\frac{\sqrt{2}}{3}\approx 0{,}47.$$

§ 2.7. Основные законы распределения непрерывных случайных величин

1. Равномерное распределение. Распределение вероятностей непрерывной случайной величины X, принимающей все свои значения из отрезка $[a;\ b]$, называется *равномерным*, если ее плотность вероятности на этом отрезке постоянна, а вне его равна нулю, т. е.

$$f(x)=\begin{cases}0 & \text{при } x<a;\\ c & \text{при } a\leqslant x\leqslant b;\\ 0 & \text{при } x>b.\end{cases}$$

Отсюда

$$\int_{-\infty}^{+\infty}f(x)\,dx=\int_a^b c\,dx=c(b-a). \tag{2.20}$$

Но, как известно (см. § 2.5, п. 2),

$$\int_{-\infty}^{+\infty}f(x)\,dx=1. \tag{2.21}$$

Из сравнения равенств (2.20) и (2.21) получаем $c=\dfrac{1}{b-a}$.

Итак, плотность вероятности непрерывной случайной величины X, распределенной равномерно на отрезке $[a;\ b]$, имеет вид

$$f(x)=\begin{cases}0 & \text{при } x<a;\\ \dfrac{1}{b-a} & \text{при } a\leqslant x\leqslant b;\\ 0 & \text{при } x>b.\end{cases}$$

Пример. На отрезке $[a;\ b]$ наугад указывают точку. Какова вероятность того, что эта точка окажется в левой половине отрезка?

Обозначим через X случайную величину, равную координате выбранной точки. X распределена равномерно (в этом и состоит точный смысл слов: «наугад указывают точку»), а так как середина отрезка $[a;\ b]$ имеет координату $\frac{a+b}{2}$, то искомая вероятность равна (см. § 2.5, п. 2):

$$P\left(a < X < \frac{a+b}{2}\right) = \int_a^{\frac{a+b}{2}} f(x)\,dx = \int_a^{\frac{a+b}{2}} \frac{1}{b-a}\,dx =$$

$$= \frac{1}{b-a}\left(\frac{a+b}{2} - a\right) = \frac{1}{2}.$$

Впрочем, этот результат был ясен с самого начала (см. § 1.2, п. 1).

2. Нормальный закон распределения. Закон распределения вероятностей непрерывной случайной величины X называется *нормальным законом*, или *законом Гаусса**, если ее плотность вероятности есть

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (2.22)$$

где σ и a — постоянные, причем $\sigma > 0$.

Убедимся, что функция (2.22) удовлетворяет условию (2.17). Действительно, перейдя в интеграле

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \qquad (2.23)$$

к новой переменной

$$t = \frac{x-a}{\sigma\sqrt{2}}, \qquad (2.24)$$

получим интеграл

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^{+\infty} e^{-t^2} dt.$$

Но

$$\int_0^{+\infty} e^{-t^2} dt = \frac{\sqrt{\pi}}{2} \text{ (см. приложение 1).}$$

Следовательно,

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = 1. \qquad (2.25)$$

Значит, интеграл (2.23) тоже равен единице.

* Карл Гаусс (1777—1855) — немецкий математик.

Покажем, что $M(X) = a$, $\sigma = \sqrt{D(X)}$, или $\sigma^2 = D(X)$.

Согласно формуле (2.18), получаем

$$M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Введя новую переменную t по формуле (2.24), с учетом равенства (2.25) получим

$$M(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{-\infty} \left(a + t\sigma\sqrt{2}\right) e^{-t^2} \sigma\sqrt{2} \cdot dt =$$

$$= \frac{a}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt + \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2} dt = a - \frac{\sigma}{\sqrt{2\pi}} e^{-t^2} \Big|_{-\infty}^{+\infty} = a.$$

Далее, в соответствии с формулой (2.19)

$$D(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Воспользовавшись подстановкой (2.24), получим:

$$D(X) = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt.$$

Применяя здесь метод интегрирования по частям ($t = u$, $te^{-t^2}dt = dv$), получим с учетом (2.25)

$$D(X) = -\frac{\sigma^2}{\sqrt{\pi}} \left(te^{-t^2}\right) \Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = 0 + \sigma^2 = \sigma^2.$$

График функции $y = e^{-x^2}$ (кривая *Гаусса*) имеет вид (рис. 6). С учетом графика этой функции график функции (2.22) будет иметь вид (рис. 7). Причем его максимальная ордината равна $1/\left(\sigma\sqrt{2\pi}\right)$. Значит, эта ордината убывает с возрастанием значения σ (кривая «растягивается» к оси Ox — рис. 8) и возрастает с убыванием зна-

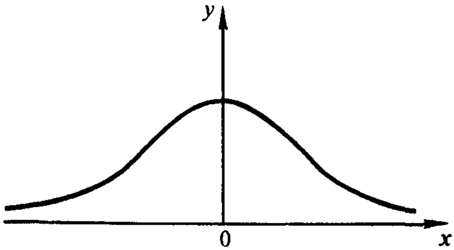


Рис. 6

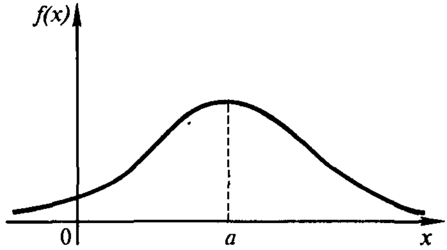


Рис. 7

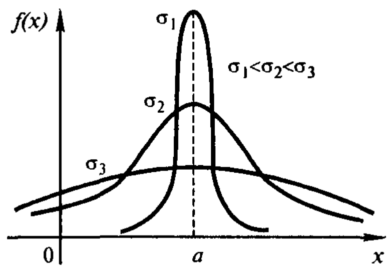


Рис. 8

чения σ (кривая «сжимается» в положительном направлении оси Oy). Изменение значений параметра a (при неизменном значении σ) не влияет на форму кривой.

Нормальное распределение с параметрами $a=0$ и $\sigma=1$ называется *нормированным*. Плотность вероятности в случае такого распределения оказывается равной

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Вероятность попадания случайной величины, распределенной по нормальному закону, в заданный интервал.

Пусть случайная величина X распределена по нормальному закону. Тогда вероятность того, что X примет значение, принадлежащее интервалу (α, β), согласно теореме из п. 2 § 2.5

$$P(\alpha<X<\beta)=\frac{1}{\sigma\sqrt{2\pi}}\int_{\alpha}^{\beta}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

Проведя в этом интеграле замену переменной $t=\frac{x-a}{\sigma}$, получим

$$P(\alpha<X<\beta)=\frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}}e^{-\frac{t^2}{2}}dt.$$

Учитывая, что функция $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ является первообразной для $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, и используя формулу Ньютона — Лейбница, будем иметь

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \qquad (2.26)$$

Пример 1. Пусть случайная величина X распределена по нормальному закону с параметрами $a = 30$ и $\sigma = 10$. Найдем вероятность того, что X примет значение, принадлежащее инвервалу (10; 50).

Пользуясь формулой (2.26), получим

$$P(10 < X < 50) = \Phi\left(\frac{50-30}{10}\right) - \Phi\left(\frac{10-30}{10}\right) = \Phi(2) - \Phi(-2) = 2\Phi(2).$$

По таблице приложения 3 находим $\Phi(2) = 0,4772$. Отсюда искома вероятность

$$P(10 < X < 50) = 2 \quad 0,4772 = 0,9544.$$

Вычисление вероятности заданного отклонения.

Часто требуется определить вероятность того, что отклонени нормально распределенной случайной величины X от ее математи ческого ожидания по абсолютной величине меньше заданного по ложительного числа δ, т. е. нужно найти $P(|X - a| < \delta)$.

Используя формулу (2.26) и учитывая, что функция $\Phi(x)$ нечет ная, имеем

$$P(|X - a| < \delta) = P(a - \delta < X < a + \delta) =$$

$$= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right) = 2\Phi\left(\frac{\delta}{\sigma}\right),$$

т. е.

$$P(|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right). \qquad (2.27)$$

Пример 2. Пусть случайная величина X распределена по нормальному закону с параметрами $a = 20$ и $\sigma = 10$. Найдем $P(|X - 20| < 3)$.

Используя выражение (2.27), имеем

$$P(|X - 20| < 3) = 2\Phi\left(\frac{3}{10}\right).$$

По таблице приложения 3 находим $\Phi(0,3) = 0,1179$. Поэтому

$$P(|X - 20| < 3) = 0,2358.$$

П р а в и л о   т р е х   с и г м.

Полагая в выражении (2.27) $\delta = 3\sigma$, получим

$$P(|X-a|<3\sigma)=2\Phi(3).$$

Но $\Phi(3)=0{,}49865$ (см. таблицу приложения 3) и, значит,

$$P(|X-a|<3\sigma)=0{,}9973. \qquad (2.28)$$

Формула (2.28) означает, что событие, состоящее в осуществлении неравенства $|X-a|<3\sigma$, имеет вероятность, близкую к единице, т. е. является почти достоверным. Эта формула выражает так называемое правило трех сигм: *если случайная величина распределена по нормальному закону, то модуль ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения.*

В заключение заметим, что нормальное распределение вероятностей имеет в теории вероятностей большое значение. Нормальному закону подчиняется вероятность при стрельбе по цели, его используют в теории погрешностей физических измерений и т. п.

§ 2.8. Закон больших чисел

1. Неравенство Чебышева.

Л е м м а. *Пусть X — случайная величина, принимающая только неотрицательные значения, тогда*

$$P(X\geqslant 1)\leqslant M(X). \qquad (2.29)$$

Д о к а з а т е л ь с т в о. Для простоты докажем это утверждение для дискретной величины X, принимающей случайные значения x_1, x_2, ..., x_n, при условии $x_i\geqslant 0$. По теореме сложения вероятностей для несовместимых событий (§ 1.3, п. 1) имеем

$$P(X\geqslant 1)=\sum_{x_i\geqslant 1}P(X=x_i),$$

где суммирование распространено на все значения x_i, большие или равные единице. Но для $x_i\geqslant 1$ очевидно,

$$P(X=x_i)\leqslant x_i P(X=x_i).$$

Поэтому

$$P(X\geqslant 1)=\sum_{x_i\geqslant 1}P(X=x_i)\leqslant\sum_{x_i\geqslant 1}x_iP(X=x_i). \qquad (2.30)$$

Добавим к правой части неравенства (2.30) сумму $\sum\limits_{x_i<1}x_iP(X=x_i)$, где $x_i<1$. Эта сумма неотрицательна, так как $x_i\geqslant 0$ по условию, а вероятность $P(X=x_i)\geqslant 0$. Поэтому

$$\sum_{x_i\geqslant 1}x_iP(X=x_i)\leqslant\sum_{x_i\geqslant 1}x_iP(X=x_i)+\sum_{x_i<1}x_iP(X=x_i)=\sum_{i=1}^{n}x_iP(X=x_i). \qquad (2.31)$$

Последняя сумма распространена на все значения x_i, принимаемые случайной величиной X. Следовательно (см. § 2.2, п. 1),

$$\sum_{i=1}^{n} x_i P(X = x_i) = M(X).$$

Отсюда, сопоставляя соотношения (2.30) и (2.31), получаем искомое неравенство (2.29).

Теорема. *Для любой случайной величины X при каждом положительном числе ε имеет место неравенство*

$$P(|X - M(X)| \geqslant \varepsilon) \leqslant \frac{D(X)}{\varepsilon^2}. \tag{2.32}$$

Неравенство (2.32) называется *неравенством Чебышева**.

Доказательство. Так как событие $|X - M(X)| \geqslant \varepsilon$ равносильно событию

$$\frac{(X - M(X))^2}{\varepsilon^2} \geqslant 1,$$

то

$$P\left[|X - M(X)| \geqslant \varepsilon\right] = P\left[\frac{(X - M(X))^2}{\varepsilon^2} \geqslant 1\right].$$

Случайная величина $[X - M(X)]^2/\varepsilon^2$ неотрицательна, и, значит, согласно лемме, свойству 2 математического ожидания (§ 2.2, п. 2) и определению дисперсии (§ 2.3, п. 1)

$$P\left[\frac{(X - M(X))^2}{\varepsilon^2} \geqslant 1\right] \leqslant M\left[\frac{(X - M(X))^2}{\varepsilon^2}\right] =$$

$$= \frac{1}{\varepsilon^2} M[(X - M(X))^2] = \frac{D(X)}{\varepsilon^2}.$$

Поэтому

$$P\left[|X - M(X)| \geqslant \varepsilon\right] \leqslant \frac{D(X)}{\varepsilon^2}.$$

Пример. Пусть случайная величина X имеет $D(X) = 0{,}001$. Какова вероятность того, что она отличается от $M(X)$ более чем на 0,1?

По неравенству Чебышева

$$P(|X - M(X)| \geqslant 0{,}1) \leqslant \frac{D(X)}{0{,}1^2} = \frac{0{,}001}{0{,}01} = 0{,}1.$$

Примечание. Отметим другую форму неравенства Чебышева. Так как событие, выражаемое неравенством $|X - M(X)| < \varepsilon$ противоположно событию, выражаемому неравенством $|X - M(X)| \geqslant \varepsilon$, то (§ 1.3, п. 1, следствие 2)

$$P(|X - M(X)| < \varepsilon) + P(|X - M(X)| \geqslant \varepsilon) = 1.$$

* П. Л. Чебышев (1821—1894) — выдающийся русский математик.

Отсюда с учетом неравенства (2.32) получаем такую форму неравенства Чебышева:

$$P(|X - M(X)| < \varepsilon) \geqslant 1 - \frac{D(X)}{\varepsilon^2}. \qquad (2.33)$$

2. Закон больших чисел Чебышева. Докажем закон больших чисел в широкой и удобной для практики форме, полученной П. Л. Чебышевым.

Теорема (теорема Чебышева; закон больших чисел). *Если дисперсии независимых случайных величин X_1, X_2, ..., X_n ограничены одной и той же постоянной c, $D(X_i) \leqslant c$, $i = 1, 2, ..., n$, то каково бы ни было $\varepsilon > 0$, вероятность выполнения неравенства $|\overline{X} - M(\overline{X})| < \varepsilon$,* где $\overline{X} = X_{\text{ср}} = (X_1 + X_2 + ... + X_n)/n$, *будет сколь угодно близка к единице, если число случайных величин n достаточно велико, т. е.*

$$\lim_{n\to\infty} P(|\overline{X} - M(\overline{X})| < \varepsilon) = 1. \qquad (2.34)$$

Доказательство. Применяя неравенство Чебышева (2.33) к величине X, имеем

$$P(|\overline{X} - M(\overline{X})| < \varepsilon) \geqslant 1 - \frac{D(\overline{X})}{\varepsilon^2}. \qquad (2.35)$$

Пользуясь свойствами дисперсии (§ 2.3, п. 2) и условием теоремы, получим

$$D(\overline{X}) = \frac{1}{n^2}[D(X_1) + ... + D(X_n)] \leqslant \frac{nc}{n^2} = \frac{c}{n}.$$

Отсюда с учетом неравенства (2.35) и того, что вероятность любого события не превосходит единицы (§ 1, п. 3), получим

$$1 \geqslant P(|\overline{X} - M(\overline{X})| < \varepsilon) \geqslant 1 - \frac{c}{\varepsilon^2 n}. \qquad (2.36)$$

Наконец, переходя в неравенстве (2.36) к пределу при $n \to \infty$, приходим к искомому соотношению (2.34).

Частный случай теоремы Чебышева. Если все X_k имеют одинаковое математическое ожидание $M(X_1) = ... = M(X_n) = a$ и $D(X_k) < c$, $k = 1, ..., n$, то

$$\lim_{n\to\infty} P(|\overline{X} - a| < \varepsilon) = 1. \qquad (2.37)$$

Действительно, в условиях рассматриваемого частного случая равенство (2.34) имеет вид (2.37).

Сущность теоремы Чебышева состоит в следующем. Несмотря на то, что каждая из независимых случайных величин X_k может принять значение, далекое от математического ожидания $M(X_k)$, среднее арифметическое $\overline{X}$ достаточно большого числа случайных величин с большой вероятностью весьма близко к среднему арифметическому их математических ожиданий.

Теорема Чебышева имеет громаднейшее практическое значение. Пусть, например, измеряется некоторая физическая величина. Обычно принимают в качестве искомого значения измеряемой величины среднее арифметическое результатов нескольких измерений. Можно ли считать такой подход верным? Теорема Чебышева (ее частный случай) отвечает на этот вопрос положительно.

На теореме Чебышева основан широко применяемый в статистике выборочный метод, согласно которому по сравнительно небольшой случайной выборке выносят суждение, касающееся всей совокупности исследуемых объектов.

Из теоремы Чебышева (частный случай) следует теорема Бернулли, являющаяся простейшей формой закона больших чисел.

Теорема Бернулли. *Пусть m — число наступлений события A в n независимых испытаниях и p есть вероятность наступления события A в каждом из испытаний. Тогда, каково бы ни было положительное число ε,*

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1. \tag{2.38}$$

Доказательство. Обозначим через X_k случайную величину, равную числу наступлений события A в k-м испытании, где $k=1$, 2, ..., n. Тогда имеем (§ 2.4, п. 1)

$$m = X_1 + X_2 + \ldots + X_n;$$

$$M(X_k) = p;\ D(X_k) = pq \leqslant \frac{1}{4},$$

и все условия частного случая теоремы Чебышева выполнены. Равенство (2.37) превращается в равенство (2.38).

Практический смысл теоремы Бернулли следующий: при постоянстве вероятности случайного события A во всех испытаниях при неограниченном возрастании числа испытаний можно с вероятностью, сколь угодно близкой к единице (т. е. как угодно близко к достоверности), утверждать, что наблюдаемая относительная частота случайного события будет как угодно мало отклоняться от его вероятности.

§ 2.9. Предельные теоремы теории вероятностей

1. Центральная предельная теорема. Как уже отмечалось, нормально распределенные случайные величины имеют широкое распространение на практике. Объяснение этому дает центральная предельная теорема, один из вариантов формулировки которой принадлежит русскому математику А. М. Ляпунову (1857—1918). Суть центральной предельной теоремы состоит в следующем: *если случайная величина X представляет собой сумму очень большого числа независимых случайных величин, влияние каждой из которых на всю* *сумму ничтожно мало, то X имеет распределение, близкое к нормальному.*

Приведем без доказательства (доказательство см. в работе [3]) центральную предельную теорему для случая одинаково распределенных случайных величин.

Теорема. *Если X_1, X_2, ..., X_n — независимые случайные величины, имеющие одно и то же распределение с математическим ожиданием a и дисперсией σ^2, то при неограниченном увеличении n закон распределения суммы $X = X_1 + X_2 + ... + X_n$ неограниченно приближается к нормальному.*

2. Локальная и интегральная предельные теоремы Лапласа.

Если число испытаний n велико, то вычисления по формуле Бернулли становятся затруднительными. Лаплас* получил важную приближенную формулу для расчета вероятности $P_n(m)$ появления события A точно m раз, если n — достаточно большое число. Им же получена приближенная формула и для суммы вида $\sum_{m=k}^{l} P_n(m)$.

Локальная предельная теорема Лапласа. *Пусть $p = P(A)$ — вероятность события A, причем $0 < p < 1$. Тогда вероятность того, что в условиях схемы Бернулли событие A при n испытаниях появится точно m раз, выражается приближенной формулой*

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\,\varphi\left(\frac{m - np}{\sqrt{npq}}\right), \qquad (2.39)$$

где $q = 1 - p$;

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Для функции $\varphi(x)$ составлена таблица (см. приложение 2) ее значений для положительных значений x [функция $\varphi(x)$ четная].

Выражение (2.39) называют *формулой Лапласа.*

Пример 1. Вероятность поражения цели стрелком при одиночном выстреле $p = 0{,}2$. Какова вероятность того, что при 100 выстрелах цель будет поражена ровно 20 раз?

Здесь $p = 0{,}2$, $q = 0{,}8$, $n = 100$ и $m = 20$. Отсюда $\sqrt{npq} = \sqrt{100 \cdot 0{,}2 \cdot 0{,}8} = 4$, и, следовательно,

$$\frac{m - np}{\sqrt{npq}} = \frac{20 - 100\ \ 0{,}2}{4} = 0.$$

Учитывая, что $\varphi(0) = 1/\sqrt{2\pi} \approx 0{,}40$, из формулы (2.39) получаем

$$P_{100}(20) \approx 0{,}40\frac{1}{4} = 0{,}10.$$

* Пьер Лаплас (1749—1827) — французский математик, механик и астроном.

Перейдем к интегральной теореме Лапласа. Поставим следующий вопрос: какова вероятность того, что в условиях схемы Бернулли событие A, имеющее вероятность $P(A)=p(0<p<1)$ при n испытаниях (как и прежде число испытаний велико), появится не менее k раз и не более l раз? Эту искомую вероятность обозначим $P_n(k, l)$.

На основании теоремы сложения вероятностей для несовместимых событий (§ 1.3, п. 1) получим

$$P_n(k, l) = \sum_{m=k}^{l} P_n(m). \tag{2.40}$$

Получим приближенную формулу Лапласа для подсчета суммы (2.40) при больших m и n. Используя локальную теорему Лапласа, приближенно будем иметь

$$P_n(k, l) \approx \sum_{m=k}^{l} \frac{1}{\sqrt{npq}} \varphi(x_m),$$

где

$$x_m = \frac{m-np}{\sqrt{npq}}, \quad m=k, \ k+1, \ \ldots, \ l; \tag{2.41}$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Далее, в силу равенства (2.41) имеем

$$\Delta x_m = x_{m+1} - x_m = \frac{(m+1)-np}{\sqrt{npq}} - \frac{m-np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}} \tag{2.42}$$

и потому

$$P_n(k, l) \approx \sum_{m=k}^{l} \varphi(x_m)\, \Delta x_m.$$

Здесь сумма справа является интегральной суммой для функции $\varphi(x)$ на отрезке $x_k \leqslant x \leqslant x_l$ причем, как следует из равенства (2.42), при $n \to \infty$ $\Delta x_m \to 0$. Следовательно, при $n \to \infty$ предел указанной интегральной суммы есть определенный интеграл

$$\int_{x_k}^{x_l} \varphi(x)\, dx.$$

Поэтому

$$P_n(k, l) \approx \int_{x_k}^{x_l} \varphi(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{x_k}^{x_l} e^{-\frac{x^2}{2}}\, dx, \tag{2.43}$$

где

$$x_k = \frac{k-np}{\sqrt{npq}}; \ x_l = \frac{l-np}{\sqrt{npq}}. \tag{2.44}$$

Выражение (2.43) при условии (2.44) и составляет содержание *интегральной предельной теоремы Лапласа.* Нами уже была введена функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, \qquad (2.45)$$

называемая *функцией Лапласа*, или *интегралом вероятности.* Очевидно, $\Phi(x)$ есть первообразная для функции $\varphi(x)$. Поэтому на основании формулы Ньютона — Лейбница из формулы (2.43) получим

$$P_n(k, l) \approx \Phi(x_l) - \Phi(x_k) \qquad (2.46)$$

(интегральная формула Лапласа).

Как известно, интеграл $\left(1/\sqrt{2\pi}\right)\int e^{-\frac{x^2}{2}} dx$ не берется в элементарных функциях. Поэтому для функции (2.45) составлена таблица (см. приложение 3) ее значений для положительных значений x, так как $\Phi(0) = 0$ и функция $\Phi(x)$ нечетная:

$$\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_0^{-x} e^{-\frac{t^2}{2}} dt = -\frac{1}{\sqrt{2\pi}} \int_0^{x} e^{-\frac{z^2}{2}} dz = -\Phi(x)$$

$$(t = -z,\ dt = -dz).$$

Пример 2. Вероятность того, что изделие не прошло проверку ОТК, $p = 0{,}2$. Найдем вероятность того, что среди 400 случайно отобранных изделий окажутся непроверенными от 70 до 100.

Здесь $n = 400$, $k = 70$, $l = 100$, $p = 0{,}2$, $q = 0{,}8$. Поэтому в силу равенств (2.44) $x_k = -1{,}25$, $x_l = 2{,}5$ и, согласно формуле (2.46),

$$P_{400}(70, 100) \approx \Phi(2{,}5) - \Phi(-1{,}25) = 0{,}4938 + 0{,}3944 = 0{,}8882.$$

Замечание. Отметим, что локальную и интегральную предельные теоремы Лапласа иногда еще называют локальной и интегральной предельными теоремами Муавра* — Лапласа.

3. Распределение случайных ошибок измерения. Пусть проводится измерение некоторой величины. Разность $x - a$ между результатом измерения x и истинным значением a измеряемой величины называется *ошибкой измерения.* Вследствие воздействия на измерение большого количества факторов, которые невозможно учесть (случайные изменения температуры, колебание прибора, ошибки, возникающие при округлении и т. п.), ошибку измерения можно считать суммой большого числа независимых случайных величин, которая по центральной предельной теореме должна быть распределена нормально. Если при этом нет систематически действующих факторов (например, неисправности приборов, завышающих при

* Авраам Муавр (1667—1754) — английский математик.

каждом измерении показания), приводящих к систематическим ошибкам, то математическое ожидание случайных ошибок равно нулю.

Итак, принимается положение: при отсутствии систематически действующих факторов ошибка измерения есть случайная величина (обозначим ее через T), распределенная нормально, причем ее математическое ожидание равно нулю, т. е. плотность вероятности величины T равна

$$\psi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}},$$

где σ — среднеквадратическое отклонение величины T, характеризующее разброс результатов измерения вокруг измеряемой величины.

Результат измерения также есть случайная величина (обозначим ее через X), связанная с T зависимостью $X = a + T$. Отсюда: $M(X) = a$, $\sigma(X) = \sigma(T) = \sigma$ и X имеет нормальный закон распределения.

Заметим, что случайная ошибка измерения, как и результаты измерения, всегда выражается в некоторых целых единицах, связанных с шагом шкалы измерительного прибора; в теории удобнее считать случайную ошибку непрерывной случайной величиной, что упрощает расчеты.

При измерении возможны две ситуации:

а) известно σ (это характеристика прибора и комплекса условий, при которых проводятся измерения), требуется по результатам измерений оценить a;

б) σ не известно, требуется по результатам измерений оценить a и σ.

Рассмотрению этих ситуаций при проведении физических измерений будет посвящен § 4.3.

Упражнения

1. Пусть случайная величина X — число очков, выпавших при подбрасывании игральной кости. Найдите закон распределения случайной величины X.

X	1	2	3	4	5	6
p	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

2. В денежной лотерее выпущено 100 билетов. Разыгрываются 1 выигрыш в 500 р. и 10 выигрышей по 10 р. Найдите закон распределения случайного выигрыша X для владельца одного лотерейного билета.

X	0	10	500
p	0,89	0,1	0,01

3. Закон распределения случайной величины X задан таблицей:

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

Найдите математическое ожидание случайной величины X. [2,2]

4. Найдите математическое ожидание выигрыша X в упражнении 2. [6 р.]

5. Найдите математическое ожидание случайной величины X, зная закон ее распределения:

X	2	3	5
p	0,3	0,1	0,6

[3,9]

6. Проводятся 2 выстрела с вероятностями попадания в цель, равными $p_1 = 0,4$; $p_2 = 0,3$. Найдите математическое ожидание общего числа попаданий. [0,7]

7. Найдите математическое ожидание суммы числа очков, которые могут выпасть при одном бросании двух игральных костей. [7]

8. Найдите математическое ожидание произведения числа очков, которые могут выпасть при одном бросании двух игральных костей. [12,25]

9. Независимые случайные величины X и Y заданы следующими законами распределения:

X	2	4	5
p	0,1	0,3	0,6

и

Y	7	9
p	0,8	0,2

Найдите математическое ожидание случайной величины XY. [32,56]

10. Найдите дисперсию случайной величины X, которая задана следующим законом распределения:

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

[2,01]

11. Известны дисперсии двух независимых случайных величин X, Y: $D(X) = 4$, $D(Y) = 3$. Найдите дисперсию суммы этих величин. [7]

12. Дисперсия случайной величины X равна 5. Найдите дисперсию следующих величин: а) $X - 1$; б) $-2X$; в) $3X + 6$.

[а) 5; б) 20; в) 45]

13—15. Найдите математическое ожидание и дисперсию случайных величин, заданных различными законами распределения:

13.

X	–2	–1	0	1	2
p	0,1	0,2	0,3	0,3	0,1

[$M(X) = 0,1$ и $D(X) = 1,29$]

14.

X	1	3	4	6	7
p	0,1	0,1	0,3	0,4	0,1

[$M(X) = 4,7$ и $D(X) = 3,01$]

15.

X	5	7	10	15
p	0,2	0,5	0,2	0,1

[$M(X) = 8$ и $D(X) = 8$]

16. К случайной величине прибавили постоянную a. Как при этом изменяется ее: а) математическое ожидание; б) дисперсия?

[а) прибавится a; б) не изменится]

17. Случайную величину умножили на a. Как при этом изменятся: а) математическое ожидание; б) дисперсия?

[а) умножится на a; б) умножится на a^2]

18. Случайная величина X принимает только 2 значения: 1 и –1, каждое с вероятностью 0,5. Найдите дисперсию $D(X)$ и среднее квадратическое отклонение $\sigma(X)$. [$D(X) = 1$; $\sigma(x) = 1$]

19. Дисперсия случайной величины $D(X) = 6,25$. Найдите среднее квадратическое отклонение $\sigma(X)$. [2,5]

20. Пусть закон распределения случайной величины X задан таблицей:

X	4	10	20
p	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Определите математическое ожидание $M(X)$, дисперсию $D(X)$ и среднее квадратическое отклонение $\sigma(X)$.

[$M(X) = 11$; $D(X) = 33$; $\sigma(X) \approx 5,75$]

21. Дискретная случайная величина X задана законом распределения:

X	3	5
p	0,2	0,8

Найдите начальные моменты первого и второго порядков.

[$\nu_1 = 4,6$; $\nu_2 = 21,8$]

22. Дискретная случайная величина X задана законом распределения, приведенным в предыдущем примере. Найдите центральный момент второго порядка. $[\mu_2 = 0{,}64]$

23. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 \text{ при } x \leqslant -1; \\ \frac{x}{3} + \frac{1}{3} \text{ при } -1 < x \leqslant 2; \\ 1 \text{ при } x > 2. \end{cases}$$

Найдите вероятность того, что в результате испытания случайная величина X примет значение, заключенное в интервале (0; 1).

$\left[\frac{1}{3}\right]$

24. Случайная величина X задана функцией распределеиия

$$F(x) = \begin{cases} 0 \text{ при } x \leqslant 2; \\ \frac{x}{2} - 1 \text{ при } 2 < x \leqslant 4; \\ 1 \text{ при } x > 4. \end{cases}$$

Найдите вероятность того, что в результате испытания X примет значение, заключенное в интервале (2; 3). [0,5]

25. Случайная величина X задана плотностью вероятности

$$f(x) = \begin{cases} 0 \text{ при } x < 0; \\ \frac{3}{32}(4x - x^2) \text{ при } 0 \leqslant x \leqslant 4; \\ 0 \text{ при } x > 4. \end{cases}$$

Найдите вероятность попадания случайной величины X на отрезок [–2; 3].

$\left[\frac{27}{32}\right]$

26. Плотность вероятности случайной величины X задана выражением

$$f(x) = \frac{1}{\pi(1 + x^2)} (-\infty < x < +\infty).$$

Найдите вероятность того, что величина X попадает на интервал (–1; 1).

[0,5]

27. Случайная величина задана плотностью вероятности

$$f(x) = \begin{cases} 0 \text{ при } x < -\frac{\pi}{2}; \\ a \cos x \text{ при } -\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2}; \\ 0 \text{ при } x > \frac{\pi}{2}. \end{cases}$$

Найдите коэффициент a. $[a = 0{,}5]$

28. Дана дифференциальная функция распределения непрерывной случайной величины X

$$f(x)=\begin{cases}0 \text{ при } x\leqslant 0;\\ \cos x \text{ при } 0<x\leqslant\frac{\pi}{2};\\ 0 \text{ при } x>\frac{\pi}{2}.\end{cases}$$

Найдите интегральную функцию распределения $F(x)$.

$$\left[F(x)=\begin{cases}0 \text{ прн } x\leqslant 0;\\ \sin x \text{ при } 0<x\leqslant\frac{\pi}{2};\\ 1 \text{ при } x>\frac{\pi}{2}.\end{cases}\right]$$

29. Дана дифференциальиая функция распределения непрерывной случайной величины X

$$f(x)=\begin{cases}0 \text{ при } x\leqslant 0;\\ \sin x \text{ при } 0<x\leqslant\frac{\pi}{2};\\ 0 \text{ при } x>\frac{\pi}{2}.\end{cases}$$

Найдите интегральную функцию распределения $F(x)$.

$$\left[F(x)=\begin{cases}0 \text{ при } x\leqslant 0;\\ 1-\cos x \text{ при } 0<x\leqslant\frac{\pi}{2};\\ 1 \text{ прн } x>\frac{\pi}{2}.\end{cases}\right]$$

30. Функция

$$f(x)=\frac{2A}{e^x+e^{-x}}\ (-\infty<x<+\infty)$$

является плотностью вероятности случайной величины X. Найдите коэффициент A и функцию распределения $F(x)$.

$$\left[A=\frac{1}{\pi};\ F(x)=\frac{2}{\pi}\operatorname{arctg} e^x\right]$$

31. Найдите математическое ожидание случайной величины X, заданной плотностью вероятности

$$f(x)=\begin{cases}0 \text{ при } x\leqslant 0;\\ \frac{1}{4} \text{ при } 0<x\leqslant 4;\\ 0 \text{ при } x>4.\end{cases}$$

$[M(X)=2]$

32. Случайная величина X задана плотностью вероятности

$$f(x) = \begin{cases} 0 & \text{при } x \leqslant 0; \\ 1 & \text{при } 0 < x \leqslant 1; \\ 0 & \text{при } x > 1. \end{cases}$$

Найдите математическое ожидание и дисперсию случайной величины X.

$\left[M(X) = \frac{1}{2};\ D(X) = \frac{1}{12}\right]$

33. В хлопке 75% длинных волокон. Какова вероятность того, что среди взятых наудачу трех волокон окажутся 2 длинных волокна?

$\left[\frac{27}{64}\right]$

34. При некоторых условиях стрельбы вероятность попадания в цель равна $^1/_3$. Проводится 6 выстрелов. Какова вероятность ровно двух попаданий?

$\left[\frac{80}{243}\right]$

35. Игральная кость бросается 5 раз. Найдите вероятность того, что 2 раза появится число очков, кратное трем.

$\left[\frac{80}{243}\right]$

36. Монета подбрасывается 5 раз. Какова вероятность того, что герб появится не менее двух раз?

$\left[\frac{13}{16}\right]$

37. Пусть всхожесть семян данного растения составляет 80%. Найдите вероятность того, что из трех посеянных семян взойдут: а) два; б) не менее двух. [а) 0,384; б) 0,896]

38. По мишени проводятся 3 выстрела, причем вероятность попадания при каждом выстреле равна 0,8. Рассматривается случайная величина X — число попаданий в мишень. Найдите закон ее распределения.

X	0	1	2	3
p	0,008	0,096	0,384	0,512

39. Принимая вероятности рождения мальчика и девочки одинаковыми, найдите вероятность того, что среди четырех новорожденных 2 мальчика. [0,375]

40. Вероятность попадания в цель при стрельбе из орудия $p = 0{,}6$. Найдите математическое ожидание общего числа попаданий, если проводят 10 выстрелов. [6 попаданий]

41. Найдите математическое ожидание числа лотерейных билетов, на которые выпадут выигрыши, если приобретено 20 билетов, причем вероятность выигрыша по одному билету равна 0,3. [6 билетов]

42. Найдите дисперсию случайной величины X — числа появлений события A в 100 независимых испытаниях, в каждом из которых вероятность наступления события A равна 0,7. [21]

43. Найдите: а) математическое ожидание и б) дисперсию числа бракованных изделий в партии из 5000 изделий, если каждое изделие может оказаться бракованным с вероятностью 0,02.

[а) 100 изделий; б) 98]

44. Проводится 10 независимых испытаний, в каждом из которых вероятность появления события A равна 0,6. Найдите дисперсию случайной величины X — числа появлений события A в этих испытаниях. [2,4]

45. Найдите дисперсию случайной величины X — числа появлений события A в двух независимых испытаниях, если $M(X)=0{,}8$. [0,48]

46. Рост взрослой женщины является случайной величиной, распределенной по нормальному закону с параметрами: $a=164$ см, $\sigma=5{,}5$ см. Найдите плотность вероятности этой величины.

$$\left[f(x)=\frac{1}{5{,}5\sqrt{2\pi}}e^{-\frac{(x-164)^2}{60{,}5}}\right]$$

47. Случайная величина X распределена по нормальному закону. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 0 и 2. Найдите вероятность того, что X примет значение, принадлежащее интервалу (–2; 3). [0,77453]

48. Случайная величина X распределена по нормальному закону. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 6 и 2. Найдите вероятность того, что X примет значение, принадлежащее интервалу (4; 8). [0,6826]

49. Пусть масса пойманной рыбы подчиняется нормальному закону с параметрами: $a=375$ г; $\sigma=25$ г. Найдите вероятность того, что масса пойманной рыбы будет от 300 до 425 г. [0,9759]

50. Диаметр детали, изготовляемой в цеху, является случайной величиной, распределенной по нормальному закону. Дисперсия ее равна 0,0001, а математическое ожидание — 2,5 мм. Найдите границы, в которых с вероятностью 0,9973 заключен диаметр наудачу взятой детали.

[2,47; 2,53]

51. Случайная величина X распределена по нормальному закону. Среднее квадратическое отклонение этой величины равно 0,4. Найдите вероятность того, что отклонение случайной величины X от ее математического ожидания по абсолютной величине будет меньше 0,3. [0,5468]

52. Случайная величина X распределена по нормальному закону. Среднее квадратическое отклонение этой величины равно 2. Найдите вероятность того, что отклонение случайной величины X от ее математического ожидания по абсолютной величине будет меньше 0,1. [0,03988]

53. Случайная величина X подчиняется нормальному закону распределения с математическим ожиданием 30 и дисперсией 100. Найдите вероятность того, что значение случайной величины заключено в интервале (10; 50).

[0,954]

54. Найдите дисперсию случайной величины X, заданной таблицей распределения:

X	2	3	5
p	0,1	0,6	0,3

[1,05]

55. При выработке некоторой массовой продукции вероятность появления одного нестандартного изделия составляет 0,01. Какова вероятность того, что в партии из 100 изделий этой продукции 2 изделия будут нестандартными? [0,184]

56. На завод прибыла партия деталей в количестве 1000 шт. Вероятность того, что одна деталь окажется бракованной, равна 0,001. Какова вероятность того, что среди прибывших деталей будет 5 бракованных? [0,003]

57. Игральную кость бросают 80 раз. Определите вероятность того, что цифра 3 появится 20 раз. [0,0162]

58. При установившемся технологическом режиме завод выпускает в среднем 70% продукции первого сорта. Определите вероятность того, что из 1000 изделий число первосортных заключено между 652 и 760. [0,99945]

59. Вероятность наступления случайного события при отдельном испытании равна p. Определите вероятность того, что в n испытаниях событие наступит подряд k раз. [$(n-k+1)\,p^k(1-p)^{n-k}$]

60. Подсчитайте при одновременном бросании n игральных костей количество исходов, в которых определенная грань встречается k раз. [$5^{n-k}C_n^k$]

ГЛАВА III

ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 3.1. Понятие о двумерной случайной величине

В различных практических приложениях встречаются случайные величины, возможные значения которых определяются не одним числом, а несколькими. Так, при вытачивании на станке цилиндрической детали ее размеры (диаметр основания и высота) являются случайными величинами. Таким образом, здесь мы имеем дело с совокупностью (системой) двух случайных величин, называемой *двумерной случайной величиной* и обозначаемой через (X, Y). Каждая из величин X и Y называется *составляющей (компонентой)* такой системы.

Различают дискретные (составляющие этих величин дискретные величины) и непрерывные (составляющие этих величин непрерывные величины) двумерные случайные величины.

Аналогично n-мерную случайную величину можно рассматривать как систему n случайных величин. Например, трехмерная величина (X, Y, Z) определяет систему трех случайных величин X, Y, Z.

На практике чаще приходится встречаться с двумерными случайными величинами. Поэтому ограничимся их рассмотрением, хотя все положения, касающиеся двумерных случайных величин, могут быть распространены и на n-мерные случайные величины. Геометрически двумерная случайная величина (X, Y) интерпретируется как случайная точка на плоскости.

Аналогично одномерной дискретной случайной величине, *законом распределения дискретной двумерной случайной величины* (X, Y) называют перечень возможных значений этой величины, т. е. пар чисел (x_i, y_j), и их вероятностей $P(X=x_i,\ Y=y_j)=p(x_i,\ y_j)=p_{ij}$ $(i=1,\ 2,\ \ldots,\ n;\ j=1,\ 2,\ \ldots,\ m)$. Обычно закон распределения этой величины задают в виде таблицы. Ее общий вид

$X \backslash Y$	y_1	y_2	.	y_m
x_1	p_{11}	p_{12}	…	p_{1m}
x_2	p_{21}	p_{22}	…	p_{2m}
…	…	…	…	…
x_n	p_{n1}	p_{n2}	…	p_{nm}

Здесь, например, p_{12} есть вероятность того, что двумерная величина (X, Y) примет значение (x_1, y_2).

Так как события $(X=x_i, Y=y_j)$ $(i=1, 2, ..., n; j=1, 2, ..., m)$ образуют полную группу несовместимых событий, то сумма всех вероятностей, помещенных в таблице, равна единице (§ 1.3, п. 1, следствие 1), т. е.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1.$$

Зная закон распределения двумерной дискретной случайной величины, можно найти законы распределения каждой из составляющих. Действительно, например, события $(X=x_1, Y=y_1)$, $(X=x_1, Y=y_2)$, ..., $(X=x_1, Y=y_m)$ несовместны, поэтому вероятность $P(X=x_1)=p(x_1)$ того, что X примет значение x_1, по теореме сложения вероятностей (см. § 1.3, п. 1) такова:

$$p(x_1) = p_{11} + p_{12} + \ldots + p_{1m}.$$

Таким образом, вероятность того, что X примет значение x_1, равна сумме вероятностей «строки x_1». Аналогично в случае других возможных значений величины X. Сложив же вероятности «столбца y_j» $(j=1, 2, ..., m)$, получим вероятность $P(Y=y_j)=p(y_j)$.

Пример. Найдем законы распределения составляющих двумерной случайной величины, заданной законом распределения:

X \ Y	y_1	y_2
x_1	0,10	0,06
x_2	0,30	0,18
x_3	0,20	0,16

Решение. Сложив вероятности по строкам, получим вероятности возможных значений X:

$$P(X=x_1)=0,16; \quad P(X=x_2)=0,48; \quad P(X=x_3)=0,36.$$

Отсюда закон распределения X:

$$\begin{array}{cccc} X & x_1 & x_2 & x_3 \\ p & 0,16 & 0,48 & 0,36\,. \end{array}$$

Сложив вероятности по столбцам, получим

$$P(Y=y_1)=0,60; \quad P(Y=y_2)=0,40,$$

и, значит, закон распределения составляющей Y:

$$\begin{array}{ccc} Y & y_1 & y_2 \\ p & 0,60 & 0,40\,. \end{array}$$

§ 3.2. Функция распределения двумерной случайной величины

1. Определение функции распределения двумерной случайной величины и ее свойства.

Определение. *Функцией распределения* двумерной случайной величины (X, Y) (безразлично, дискретной или непрерывной) называют функцию $F(x, y)$, определяющую для каждой пары действительных чисел x, y вероятность того, что X примет значение, меньшее x, и при этом Y примет значение, меньшее y:

$$F(x, y) = P(X < x,\ Y < y).$$

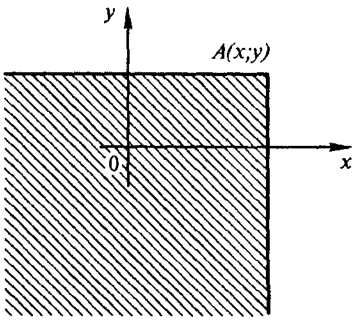


Рис. 9

Геометрически (рис. 9) это равенство представляет собой вероятность того, что случайная точка (X, Y) попадет в бесконечный квадрант с вершиной (x, y), расположенный левее и ниже этой вершины.

Пример. Найдите вероятность того, что в результате испытания составляющая X двумерной случайной величины (X, Y) примет значение $X < 4$ и при этом составляющая Y примет значение $Y < 5$, если известна функция распределения этой величины

$$F(x, y) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{x}{4} + \frac{1}{3}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{y}{5} + \frac{1}{3}\right).$$

Решение. Имеем

$$P(X < 4, Y < 5) = F(4, 5) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{4}{4} + \frac{1}{3}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{5}{5} + \frac{1}{3}\right) =$$

$$= \left(\frac{1}{\pi}\cdot\frac{\pi}{4} + \frac{1}{3}\right)\left(\frac{1}{\pi}\cdot\frac{\pi}{4} + \frac{1}{3}\right) = \frac{7}{12}\cdot\frac{7}{12} = \frac{49}{144}.$$

Укажем свойства, которыми обладает функция $F(x, y)$.

1. $0 \leqslant F(x, y) \leqslant 1$.

Это свойство следует из того, что $F(x, y)$ вероятность.

2. $F(x, y)$ — *неубывающая функция по каждому аргументу, т. е.*

$$F(x_1, y) \leqslant F(x_2, y),\ \text{если}\ x_1 < x_2;$$

$$F(x, y_1) \leqslant F(x, y_2),\ \text{если}\ y_1 < y_2.$$

Свойство становится наглядным, если воспользоваться геометрической интерпретацией функции распределения.

3. Имеют место предельные соотношения:

$$F(-\infty, y) = \lim_{x\to-\infty} F(x, y) = 0,$$

$$F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0;$$

$$F(-\infty, -\infty) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0;$$

$$F(+\infty, +\infty) = \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1.$$

Эти соотношения также следуют из геометрической интерпретации функции распределения.

4. При $y \to +\infty$ $(x \to +\infty)$ функция распределения системы стремится к функции распределения составляющей $X(Y)$:

$$\lim_{y \to +\infty} F(x, y) = F(x, +\infty) = F_1(x);$$

$$\lim_{x \to +\infty} F(x, y) = F(+\infty, y) = F_2(y).$$

Действительно, при $y \to +\infty$ $(x \to +\infty)$ бесконечный квадрант с вершиной в точке $A(x, y)$ превращается в полуплоскость, вероятность попадания в которую есть функция распределения составляющей $X(Y)$.

2. Вероятности попадания случайной точки в полуполосу и прямоугольник.

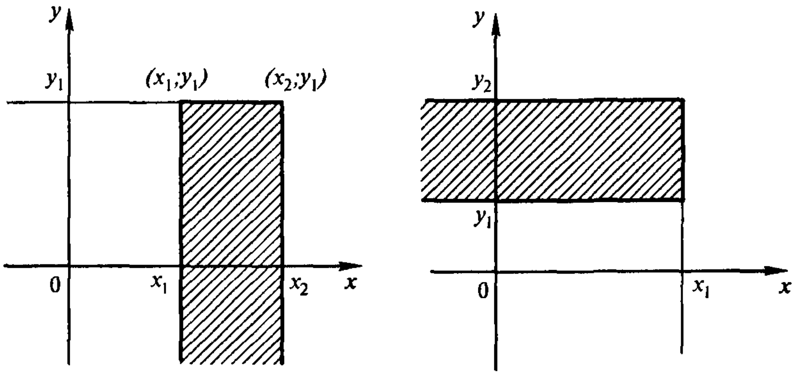
1. Найдем вероятность того, что в результате испытания случайная точка попадет в полуполосу $x_1 < X < x_2$ и $Y < y_1$ (рис. 10, *а*) или в полуполосу $X < x_1$ и $y_1 < Y < y_2$ (рис. 10, *б*).

Вычитая из вероятности попадания случайной точки в квадрант с вершиной $(x_2; y_1)$ вероятность попадания этой точки в квадрант с вершиной $(x_1; y_1)$ (рис. 10, *а*) получим

$$P(x_1 \leqslant X < x_2,\ Y < y_1) = F(x_2, y_1) - F(x_1, y_1).$$

Аналогично

$$P(X < x_1,\ y_1 \leqslant Y < y_2) = F(x_1, y_2) - F(x_1, y_1).$$



а *б*

Рис. 10

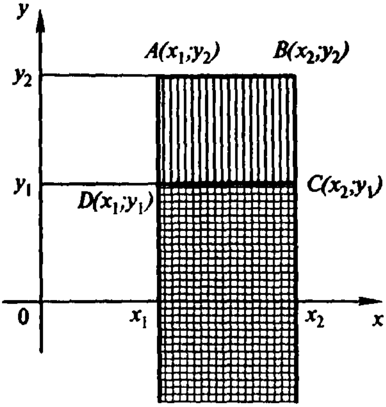


Рис. 11

Итак, *вероятность попадания случайной точки в полуполосу равна приращению функции распределения по одному из аргументов.*

2. Найдем вероятность попадания случайной точки $(X;\ Y)$ в прямоугольник $ABCD$ (рис. 11).

Для этого из вероятности попадания случайной точки в полуполосу AB с вертикальной штриховкой вычтем вероятность попадания этой точки в полуполосу DC с горизонтальной штриховкой. Получим

$$P(x_1 \leqslant X < x_2,\ y_1 \leqslant Y < y_2) =$$
$$= (F(x_2,\ y_2) - F(x_1,\ y_2)) -$$
$$- (F(x_2,\ y_1) - F(x_1,\ y_1)). \qquad (3.1)$$

Пример. Найдем вероятность попадания случайной точки $(X,\ Y)$ в прямоугольник, ограниченный прямыми $x = \frac{\pi}{6}$, $x = \frac{\pi}{2}$, $y = \frac{\pi}{6}$, $y = \frac{\pi}{2}$, если известна функция распределения

$$F(x,\ y) = \sin x \sin y \ \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right).$$

Решение. В данном примере в выражении (3.1) $x_1 = \frac{\pi}{6}$, $x_2 = \frac{\pi}{2}$, $y_1 = \frac{\pi}{6}$, $y_2 = \frac{\pi}{2}$ и, значит,

$$P\left(\frac{\pi}{6} < X < \frac{\pi}{2},\ \frac{\pi}{6} < Y < \frac{\pi}{2}\right) =$$
$$= \left(\sin\frac{\pi}{2}\cdot\sin\frac{\pi}{2} - \sin\frac{\pi}{6}\cdot\sin\frac{\pi}{2}\right) - \left(\sin\frac{\pi}{2}\cdot\sin\frac{\pi}{6} - \sin\frac{\pi}{6}\cdot\sin\frac{\pi}{6}\right) =$$
$$= \left(1\cdot 1 - \frac{1}{2}\cdot 1\right) - \left(1\cdot\frac{1}{2} - \frac{1}{2}\cdot\frac{1}{2}\right) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

§ 3.3. Плотность вероятности двумерной случайной величины

1. Двумерная плотность вероятности и ее свойства.

Определение. *Плотностью вероятности (плотностью распределения)* непрерывной двумерной случайной величины $(X,\ Y)$ называется вторая смешанная частная производная ее функции распределения, т. е.

$$f(x,\ y) = F''_{xy}(x,\ y). \qquad (3.2)$$

Геометрически плотность вероятности двумерной случайной величины $(X,\ Y)$ представляет собой поверхность в пространстве $Oxyz$, которую называют поверхностью распределения.

Плотность вероятности $f(x, y)$ обладает свойствами, аналогичными свойствам плотности вероятности одномерной случайной величины.

1. *Плотность вероятности двумерной случайной величины неотрицательна:*

$$f(x, y) \geqslant 0.$$

Действительно, разделив обе части равенства (3.1) на площадь прямоугольника $ABCD$ $\Delta x \Delta y$ ($\Delta x = x_2 - x_1$, $\Delta y = y_2 - y_1$) и дважды воспользовавшись формулой Лагранжа, получим

$$\frac{P(x_1 \leqslant X < x_2,\ y_1 \leqslant Y < y_2)}{\Delta x \Delta y} = F''_{xy}(x_1 + \theta\Delta x,\ y_1 + \theta_1 \Delta y).$$

Перейдя здесь к пределу при $\Delta x \to 0$ и $\Delta y \to 0$, находим

$$F''_{xy}(x, y) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x_1 \leqslant X < x_2,\ y_1 \leqslant Y < y_2)}{\Delta x \Delta y},$$

откуда и следует свойство 1.

2. Обозначим событие, состоящее в попадании случайной точки (X, Y) в область D, так: $(X, Y) \subset D$. Тогда *вероятность попадания непрерывной двумерной случайной величины* (X, Y) *в область* D (аналогично одномерному случаю) *равна*

$$P((X, Y) \subset D) = \iint\limits_D f(x, y)\, dx dy. \qquad (3.3)$$

Пример 1. Найдем плотность вероятности $f(x, y)$ случайной величины (X, Y) по известной функции распределения

$$F(x, y) = \left(\frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2}\right)\left(\frac{1}{\pi} \operatorname{arctg} y + \frac{1}{2}\right).$$

Решение. Имеем

$$F'_x(x, y) = \frac{1}{\pi(1 + x^2)}\left(\frac{1}{\pi} \operatorname{arctg} y + \frac{1}{2}\right),$$

$$F''_{xy}(x, y) = \frac{1}{\pi^2(1 + x^2)(1 + y^2)}.$$

Отсюда, согласно формуле (3.2),

$$f(x, y) = \frac{1}{\pi^2(1 + x^2)(1 + y^2)}.$$

3. $$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x, y)\, dx dy = 1. \qquad (3.4)$$

Это свойство следует из того, что интеграл слева в последнем равенстве есть вероятность попадания случайной точки (X, Y) во всю плоскость xOy, т. е. вероятность достоверного события.

П р и м е р 2. Двумерная плотность вероятности двух случайных величин X, Y

$$f(x, y) = \frac{C}{(9 + x^2)(16 + y^2)}.$$

Найдем величину C.

Р е ш е н и е. Согласно формуле (3.3), имеем

$$C\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{dxdy}{(9 + x^2)(16 + y^2)} = 1.$$

Но

$$C\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{dxdy}{(9 + x^2)(16 + y^2)} = \int_{-\infty}^{+\infty}\frac{dx}{9 + x^2}\int_{-\infty}^{+\infty}\frac{dy}{16 + y^2} =$$

$$= \frac{1}{12}\left(\operatorname{arctg}\frac{x}{3}\right)\Bigg|_{-\infty}^{+\infty}\left(\operatorname{arctg}\frac{y}{4}\right)\Bigg|_{-\infty}^{+\infty} = \frac{1}{3}\cdot\frac{\pi}{2}\cdot\frac{\pi}{2} = \frac{\pi^2}{12}$$

и, значит,

$$C\cdot\frac{\pi^2}{12} = 1,$$

откуда

$$C = \frac{12}{\pi^2}.$$

П р и м е р 3. В круге $x^2 + y^2 \leqslant 4$ плотность вероятности двумерной случайной величины $f(x, y) = \frac{3}{8\pi}\left(2 - \sqrt{x^2 + y^2}\right)$, а вне его $f(x, y) = 0$.

Найдем вероятность попадания случайной точки (X, Y) в круг $D\{x^2 + y^2 < 1\}$.

Р е ш е н и е. Согласно формуле (3.3), имеем

$$P((X, Y) \subset D) = \frac{3}{8\pi}\iint_D\left(2 - \sqrt{x^2 + y^2}\right)dxdy.$$

Перейдя здесь к полярным координатам, найдем

$$P((X, Y) \subset D) = \frac{3}{8\pi}\int_0^{2\pi}d\varphi\int_0^1(2 - \rho)\rho d\rho =$$

$$= \frac{3}{8\pi}\cdot 2\pi\left(\rho^2 - \frac{\rho^3}{3}\right)\Bigg|_0^1 = \frac{3}{4}\cdot\frac{2}{3} = \frac{1}{2}.$$

2. Отыскание функции распределения двумерной случайной величины по известной двумерной плотности вероятности.

Из формулы (3.2) имеем

$$\int_{-\infty}^{y}f(x, y)dy = F_x'(x, y) - F_x'(x, -\infty),$$

откуда

$$\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x, y)dxdy = F(x, y) - F(-\infty, y) - F(x, -\infty) + F(-\infty, -\infty).$$

Но (см. § 3.2, п. 1)

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

Следовательно,

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x, y)dxdy. \tag{3.5}$$

Пример. Пусть задана двумерная плотность вероятности случайной величины (X, Y)

$$f(x, y) = \frac{1}{(4 + x^2)(25 + y^2)}.$$

Найдем функцию распределения.

Решение. Согласно формуле (3.5),

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} \frac{dxdy}{(4 + x^2)(25 + y^2)} =$$

$$= \int\limits_{-\infty}^{x} \frac{dx}{4 + x^2} \int\limits_{-\infty}^{y} \frac{dy}{25 + y^2} = \frac{1}{2}\left(\operatorname{arctg}\frac{x}{2}\right)\Bigg|_{-\infty}^{x} \frac{1}{5}\left(\operatorname{arctg}\frac{y}{5}\right)\Bigg|_{-\infty}^{y} =$$

$$= \frac{1}{10}\left(\operatorname{arctg}\frac{x}{2} + \frac{\pi}{2}\right)\left(\operatorname{arctg}\frac{y}{5} + \frac{\pi}{2}\right).$$

§ 3.4. Нахождение плотностей вероятности составляющих двумериой случайной величины

Пусть известна двумерная плотность вероятности $f(x, y)$ случайной величины (X, Y). Тогда функция распределения $F(x, y)$ определяется формулой

$$F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x, y)dxdy,$$

откуда

$$F(x, +\infty) = \int\limits_{-\infty}^{x} dx \int\limits_{-\infty}^{+\infty} f(x, y)\, dy. \tag{3.6}$$

С другой стороны (см. § 3.2, п. 1)

$$F(x, +\infty) = F_1(x), \tag{3.7}$$

где $F_1(x)$ — функция распределения составляющей X. Из равенств (3.6) и (3.7) находим

$$F_1(x) = \int\limits_{-\infty}^{x} dx \int\limits_{-\infty}^{+\infty} f(x, y)\, dy.$$

Отсюда

$$\frac{dF_1(x)}{dx} = \int_{-\infty}^{+\infty} f(x, y)\,dy$$

или

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y)\,dy, \tag{3.8}$$

где $f_1(x)$ — плотность вероятности составляющей X.

Аналогично получим формулу для плотности вероятности составляющей Y:

$$f_2(y) = \int_{-\infty}^{+\infty} f(x, y)\,dx. \tag{3.8'}$$

Пример. Двумерная случайная величина (X, Y) задана плотностью распределения

$$f(x, y) = \frac{3\sqrt{3}}{\pi} e^{-4x^2-6xy-9y^2}.$$

Найдем плотности распределения составляющих X и Y.

Решение. Найдем плотность распределения составляющей X по формуле (3.8)

$$f_1(x) = \frac{3\sqrt{3}}{\pi}\int_{-\infty}^{+\infty} e^{-4x^2-6xy-9y^2}dy = \frac{3\sqrt{3}}{\pi}e^{-3x^2}\int_{-\infty}^{+\infty} e^{-(x+3y)^2}dy =$$

$$= \frac{2\sqrt{3}}{\pi}e^{-3x^2}\int_{0}^{+\infty} e^{-(x+3y)^2}d(3y) = \sqrt{\frac{3}{\pi}}e^{-3x^2},$$

так как интеграл

$$\int_{0}^{+\infty} e^{-t^2}dt = \frac{\sqrt{\pi}}{2}$$

(см. приложение 1).

Аналогично используя формулу (3.8′), получим

$$f_2(y) = \frac{3}{2}\sqrt{\frac{3}{\pi}}e^{-\frac{27}{4}y^2}.$$

§ 3.5. Условные законы распределения составляющих двумерных дискретных и непрерывных случайных величии

1. Условные законы распределения составляющих двумерных дискретных случайных величин.

Рассмотрим дискретную двумерную случайную величину (X, Y). Пусть возможные значения составляющих $x_1, x_2, \ldots, x_n$; $y_1, y_2, \ldots, y_m$.

Допустим, что в результате испытания величина Y приняла значение $Y = y_1$; при этом X примет одно из своих возможных значений: x_1 или x_2, ..., или x_n. Обозначим условную вероятность того, что X примет, например, значение x_1 при условии, что $Y=y_1$ через $p(x_1|y_1)$. В общем случае условные вероятности составляющей будем обозначать так:

$$p(x_i|y_j) \quad (i=1,\ 2,\ ...,\ n;\quad j=1,\ 2,\ ...,\ m).$$

Определение. *Условным распределением составляющей* X при $Y=y_1$ называют совокупность условных вероятностей

$$p(x_1|y_1),\ p(x_2|y_2),\ ...,\ p(x_n|y_1)^*$$

вычисленных в предположении, что событие $Y=y_1$ уже наступило.

Так же определяются и условные распределения X при $Y=y_2$, $Y=y_3$, ..., $Y=y_m$.

Аналогично определяются условные распределения составляющей Y.

Зная закон распределения двумерной дискретной случайной величины, можно, воспользовавшись формулой

$$P_B(A)=\frac{P(AB)}{P(B)} \quad (\text{см. § 1.3, п. 2, формула (1.7)}),$$

получить условные законы распределения составляющих. Так, условный закон распределения X в предположении, что событие $Y=y_1$ уже произошло, может быть найден по формуле

$$p(x_i|y_1)=\frac{p(x_i,\ y_1)}{p(y_1)},\quad i=1,\ 2,\ ...,\ n. \tag{3.9}$$

Аналогично находят условные законы распределения составляющей Y. Например, условный закон распределения Y в предположении, что событие $X=x_2$ уже произошло, есть

$$p(y_j|x_2)=\frac{p(x_2,\ y_j)}{p(x_2)},\quad j=1,\ 2,\ ...,\ m.$$

Замечание. Сумма вероятностей условного распределения авна единице. Действительно, например,

$$\sum_{j=1}^{m}p(y_j|x_1)=\sum_{j=1}^{m}\frac{p(x_1,\ y_j)}{p(x_1)}=\frac{1}{p(x_1)}\sum_{j=1}^{m}p(x_1|y_j)=\frac{p(x_1)}{p(x_i)}=1.$$

Это свойство используют для контроля вычислений.

Пример. Дискретная двумерная случайная величина задана аблицей

X \ Y	y_1	y_2
x_1	0,10	0,06
x_2	0,30	0,18
x_3	0,20	0,16

* Кратко обозначают $P_{y_1}(X)$.

Найдем условный закон распределения составляющей X при условии, что составляющая Y приняла значение y_1.

Решение. Искомый закон определяется совокупностью условных вероятностей:

$$p(x_1|y_1),\ p(x_2|y_1),\ p(x_3|y_1).$$

Воспользовавшись формулой (3.9) и приняв во внимание данные указанной таблицы ($p(x_1,\ y_1)=0{,}10$, $p(x_2,\ y_1)=0{,}30$, $p(x_3,\ y_1)=0{,}20$) и что $p(y_1)=0{,}60$ (§ 3.1, пример), имеем:

$$p(x_1|y_1)=\frac{p(x_1,y_1)}{p(y_1)}=\frac{0{,}10}{0{,}60}=\frac{1}{6},$$

$$p(x_2|y_1)=\frac{p(x_2,y_1)}{p(y_1)}=\frac{0{,}30}{0{,}60}=\frac{1}{2},$$

$$p(x_3|y_1)=\frac{p(x_3,y_1)}{p(y_1)}=\frac{0{,}20}{0{,}60}=\frac{1}{3}.$$

2. Условные законы распределения составляющих двумерных непрерывных случайных величин.

Пусть $(X,\ Y)$ — непрерывная двумерная случайная величина.

Определение. *Условной плотностью* $\varphi(x\,|\,y)$ *распределения составляющей* X при данном значении $Y=y$ называют отношение двумерной плотности вероятности $f(x,\ y)$ к плотности вероятности $f_2(y)$ составляющей Y:

$$\varphi(x|y)=\frac{f(x,y)}{f_2(y)}. \tag{3.10}$$

Отличие условной плотности $\varphi(x|y)$ от плотности $f_1(x)$ составляющей X состоит в том, что функция $\varphi(x|y)$ дает распределение X при условии, что составляющая Y приняла значение $Y=y$; функция же $f_1(x)$ дает распределение X независимо от того, какие из возможных значений приняла составляющая Y.

Аналогично определяется условная плотность составляющей Y при данном значении $X=x$:

$$\psi(y|x)=\frac{f(x,y)}{f_1(x)}. \tag{3.11}$$

Формулы (3.10) и (3.11) с учетом формул (3.8′) и (3.8) могут быть переписаны и в следующем виде:

$$\varphi(x|y)=\frac{f(x,y)}{\int\limits_{-\infty}^{+\infty} f(x,y)\,dx},$$

$$\psi(y|x)=\frac{f(x,y)}{\int\limits_{-\infty}^{+\infty} f(x,y)\,dy}.$$

Заметим, что, как и всякая плотность, условные плотности обладают свойствами:

$$\varphi(x|y) \geqslant 0, \quad \int_{-\infty}^{+\infty} \varphi(x|y)\,dx = 1,$$

$$\psi(y|x) \geqslant 0, \quad \int_{-\infty}^{+\infty} \psi(y|x)\,dy = 1.$$

П р и м е р. Пусть двумерная случайная величина (X, Y) задана плотностью вероятности

$$f(x, y) = \frac{3\sqrt{3}}{\pi} e^{-4x^2 - 6xy - 9y^2}.$$

Требуется найти условные плотности вероятности составляющих X и Y.

Р е ш е н и е. Ранее (см. § 3.4, пример) были найдены плотности вероятности составляющих X и Y

$$f_1(x) = \sqrt{\frac{3}{\pi}} e^{-3x^2}, \; f_2(y) = \frac{3}{2}\sqrt{\frac{3}{\pi}} e^{-\frac{27}{4}y^2}.$$

Поэтому, согласно формулам (3.10) и (3.11), найдем:

$$\varphi(x|y) = \frac{2}{\sqrt{\pi}} e^{-\left(2x + \frac{3}{2}y\right)^2}$$

$$\psi(y|x) = \frac{3}{\sqrt{\pi}} e^{-(x+3y)^2}.$$

§ 3.6. Независимость случайных величин

Т е о р е м а. *Для того чтобы случайные величины X и Y были независимыми, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функции ее составляющих:*

$$F(x, y) = F_1(x)F_2(y).$$

Д о к а з а т е л ь с т в о.

Н е о б х о д и м о с т ь. Пусть X и Y независимы. Тогда события $X<x$ и $Y<y$ независимы, следовательно, вероятность совмещения этих событий равна произведению их вероятностей.

$$P(X < x, \; Y < y) = P(X < x)\,P(Y < y),$$

или

$$F(x, y) = F_1(x)F_2(y).$$

Д о с т а т о ч н о с т ь. Пусть $F(x, y) = F_1(x)F_2(y)$. Отсюда следует, что вероятность совмещения событий $X<x$ и $Y<y$ равна произведению вероятностей этих событий. Следовательно, величины X и Y независимы.

Следствие. *Для того чтобы непрерывные случайные величины X и Y были независимыми, необходимо и достаточно, чтобы плотность вероятности системы (X, Y) была равна произведению плотностей вероятности составляющих X и Y:*

$$f(x, y) = f_1(x) f_2(y).$$

Пример. Двумерная непрерывная случайная величина (X, Y) задана плотностью вероятностей

$$f(x, y) = \begin{cases} \frac{1}{4} \sin x \sin y \text{ в квадрате } S\{0 \leqslant x \leqslant \pi,\ 0 \leqslant y \leqslant \pi\}, \\ 0 \text{ вне квадрата } S. \end{cases}$$

Докажите, что составляющие X и Y независимы.

Решение. Согласно формуле (3.8)

$$f_1(x) = \frac{1}{4} \int_0^\pi \sin x \sin y \, dy = -\frac{1}{4} \cos y \Big|_0^\pi \cdot \sin x = \frac{1}{2} \sin x.$$

Аналогично согласно формуле (3.8′)

$$f_2(y) = \frac{1}{2} \sin y$$

и, значит,

$$f(x, y) = f_1(x) f_2(y),$$

т. е. случайные величины X и Y независимы.

§ 3.7. Элементы теории корреляции

1. Корреляционная зависимость. Часто приходится иметь дело с более сложной зависимостью, чем функциональная. Такова, например, связь между осадками и урожаем или связь между толщиной снегового покрова зимой и объемом стока последующего половодья. Здесь каждому значению одной величины соответствует множество возможных значений другой величины. Подобного рода зависимости относятся к корреляционным зависимостям.

Определение 1. Две случайные величины X и Y находятся в *корреляционной* зависимости, если каждому значению любой из этих величин соответствует определенное распределение вероятностей другой величины.

Определение 2. *Условным математическим ожиданием* дискретной случайной величины X при $Y = y$ (y — определенное возможное значение Y) называется сумма произведений возможных значений величины X на их условные вероятности:

$$M_y(X) = \sum_{i=1}^{n} x_i p(x_i | y),$$

где $p(x_i|y)$ — условная вероятность равенства $X = x_i$ при **условии, что** $Y = y$.

Для непрерывных величин

$$M_y(X) = \int_{-\infty}^{+\infty} x\varphi(x|y)\,dx,$$

где $\varphi(x|y)$ — плотность вероятности случайной непрерывной величины X при условии $Y=y$.

Условное математическое ожидание $M_y(X)$ есть функция от y: $M_y(X)=f(y)$, которую называют *функцией регрессии* величины X на величину Y.

Аналогично определяются условное математическое ожидание случайной величины Y и функция регрессии Y на X:

$$M_x(Y)=g(x).$$

Уравнение $x=f(y)$ $(y=g(x))$ называется *уравнением регрессии* X на Y (Y на X), а линия на плоскости, соответствующая этому уравнению, называется *линией регрессии.*

Линия регрессии Y на X (X на Y) показывает, как в среднем зависит Y от X (X от Y).

Пример 1. Пусть X и Y независимы, $M(X)=a$, $M(Y)=b$. Тогда $g(x)=M_x(Y)=M(Y)=b$; $f(y)=M_y(X)=M(X)=a$. Линии регрессии изображены на рис. 12.

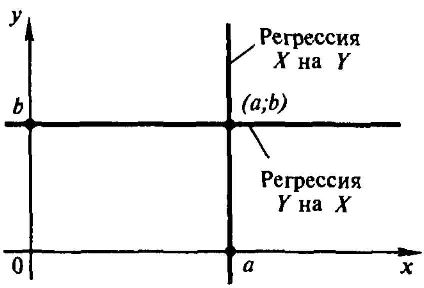


Рис. 12

Пример 2. X и Y связаны линейной зависимостью: $Y=AX+B$, $A\neq 0$. Тогда функция регрессии Y на X будет иметь вид

$$g(x)=M_x(Y)=M(Ax+B)=Ax+B.$$

Так как $X=\frac{1}{A}(Y-B)$, то функция регрессии X на Y имеет вид

$$f(y)=M_y(X)=M\left[\frac{1}{A}(y-B)\right]=\frac{1}{A}(y-B).$$

Значит, линия регрессии X на Y: $x=(y-B)/A$, т. е. $y=Ax+B$. Таким образом, в случае линейной зависимости X и Y линии регрессии X на Y и Y на X совпадают, и эта линия прямая.

2. Корреляционный момент и коэффициент корреляции.

Для характеристики корреляционной зависимости между вели чинами используются корреляционный момент и коэффициен корреляции.

Определение 2. *Корреляционным моментом* μ_{xy} случайных ве личин X и Y называют математическое ожидание произведени: отклонений этих величин

$$\mu_{xy} = M[(X - M(X))(Y - M(Y))].$$

Для вычисления корреляционного момента дискретных величи используется выражение

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))(y_j - M(Y))\,p(x_i,\ y_j), \qquad (3.12)$$

а для непрерывных — выражение

$$\mu_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - M(X))(y - M(Y))\,f(x, y)\,dxdy. \qquad (3.13)$$

Замечание. Корреляционный момент μ_{xy} может быть переписан в виде

$$\mu_{xy} = M(XY) - M(X)M(Y). \qquad (3.14)$$

Действительно, используя свойства математического ожидания (см. §§ 2.2; 2.6), имеем

$$M[(X - M(X))(Y - M(Y))] =$$
$$= M(XY - YM(X) - XM(Y) + M(X)M(Y)) =$$
$$= M(XY) - M(Y)M(X) - M(X)M(Y) + M(X)M(Y) =$$
$$= M(XY) - M(X)M(Y).$$

Теорема. *Корреляционный момент двух независимых случайны величин X и Y равен нулю.*

Доказательство. Согласно замечанию

$$\mu_{xy} = M(XY) - M(X)M(Y),$$

а так как X и Y независимые случайные величины, то (см. §§ 2.2. 2.6)

$$M(XY) = M(X)M(Y)$$

и, значит, $\mu_{xy} = 0$.

Из определения корреляционного момента следует, что он имее размерность, равную произведению размерностей величин X и Y т. е. его величина зависит от единиц измерения случайных величин Поэтому для одних и тех же двух величин величина корреляцион ного момента может иметь различные значения в зависимости о того, в каких единицах были измерены величины. Для устранения этого недостатка условились за меру связи (зависимости) двух случайных величин X и Y принять безразмерную величину

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}, \tag{3.15}$$

где $\sigma_x = \sigma(X)$, $\sigma_y = \sigma(Y)$, называемую *коэффициентом корреляции.*

Пример 1. Пусть двумерная дискретная случайная величина (X, Y) задана законом распределения:

$X \backslash Y$	1	2	3
1	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{36}$
2	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{18}$
3	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$

Найдем корреляционный момент и коэффициент корреляции случайных величин X и Y.

Решение. Сложив вероятности по строкам, получим вероятности возможных значений X:

$$P(X = x_1) = \frac{1}{6},\ P(X = x_2) = \frac{1}{3},\ P(X = x_3) = \frac{1}{2}.$$

Отсюда закон распределения X:

X	1	2	3
p	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

и, значит, $M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{2} = \frac{7}{3} = 2\frac{1}{3}$.

Сложив же вероятности по столбцам, найдем вероятности возможных значений Y:

$$P(Y = y_1) = \frac{1}{3},\ P(Y = y_2) = \frac{1}{2},\ P(Y = y_3) = \frac{1}{6}.$$

Отсюда закон распределения Y:

Y	1	2	3
p	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

и, значит, $M(Y) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{6} = \frac{11}{6} = 1\frac{5}{6}$.

Наконец, применяя формулу (3.12), получим:

$$\mu_{xy}=\left(-\frac{4}{3}\right)\left(-\frac{5}{6}\right)\frac{1}{18}+\left(-\frac{4}{3}\right)\frac{1}{6}\cdot\frac{1}{12}+\left(-\frac{4}{3}\right)\frac{7}{6}\cdot\frac{1}{36}+\left(-\frac{1}{3}\right)\left(-\frac{5}{6}\right)\frac{1}{9}+\left(-\frac{1}{3}\right)\frac{1}{6}\cdot\frac{1}{6}+$$

$$+\left(-\frac{1}{3}\right)\frac{7}{6}\cdot\frac{1}{18}+\frac{2}{3}\left(-\frac{5}{6}\right)\frac{1}{6}+\frac{2}{3}\cdot\frac{1}{6}\cdot\frac{1}{4}+\frac{2}{3}\cdot\frac{7}{6}\cdot\frac{1}{12}=-\frac{4}{3}\left(-\frac{5}{108}+\frac{1}{72}+\frac{7}{216}\right)-$$

$$-\frac{1}{3}\left(-\frac{5}{54}+\frac{1}{36}+\frac{7}{108}\right)+\frac{2}{3}\left(-\frac{5}{36}+\frac{1}{24}+\frac{7}{72}\right)=-\frac{4}{3}\cdot 0-\frac{1}{3}\cdot 0+\frac{2}{3}\cdot 0=0.$$

Следовательно, и коэффициент корреляции $r_{xy}=0$.

Пример 2. Пусть система случайных величин (X, Y) имеет закон распределения с плотностью

$$f(x, y)=\begin{cases}\frac{1}{2}\sin(x+y) & \text{в квадрате } S\left\{0\leqslant x\leqslant\frac{\pi}{2},\ 0\leqslant y\leqslant\frac{\pi}{2}\right\},\\ 0 & \text{вне } S.\end{cases}$$

Найдем корреляционный момент и коэффициент корреляции случайных величин X и Y.

Решение. Плотности вероятности составляющих X и Y найдем по формулам (3.8) и (3.8′).

$$f_1(x)=\frac{1}{2}\int_0^{\frac{\pi}{2}}\sin(x+y)\,dy=-\frac{1}{2}\cos(x+y)\Big|_0^{\frac{\pi}{2}}=\frac{1}{2}(\sin x+\cos x),$$

$$f_2(y)=\frac{1}{2}\int_0^{\frac{\pi}{2}}\sin(x+y)\,dx=\frac{1}{2}(\sin y+\cos y)$$

и, значит,

$$M(X)=\frac{1}{2}\int_0^{\frac{\pi}{2}}x(\sin x+\cos x)\,dx,$$

$$M(Y)=\frac{1}{2}\int_0^{\frac{\pi}{2}}y(\sin y+\cos y)\,dy.$$

Далее имеем

$$\frac{1}{2}\int_0^{\frac{\pi}{2}}\underbrace{x}_{u}\underbrace{(\sin x+\cos y)\,dx}_{dv}=\frac{1}{2}[x(\sin x-\cos x)]_0^{\frac{\pi}{2}}-\frac{1}{2}\int_0^{\frac{\pi}{2}}(\sin x-\cos x)\,dx=$$

$$=\frac{\pi}{4}+\frac{1}{2}(\sin x+\cos x)\Big|_0^{\frac{\pi}{2}}=\frac{\pi}{4}.$$

Следовательно, $M(X)=M(Y)=\frac{\pi}{4}$.

Найдем теперь μ_{xy}. Согласно формуле (3.13), имеем

$$\mu_{xy} = \frac{1}{2}\int_0^{\frac{\pi}{2}}\left(x-\frac{\pi}{4}\right)dx\int_0^{\frac{\pi}{2}}\underbrace{\left(y-\frac{\pi}{4}\right)}_{u}\underbrace{\sin(x+y)\,dy}_{dv}.$$

Но

$$\int_0^{\frac{\pi}{2}}\underbrace{\left(y-\frac{\pi}{4}\right)}_{u}\underbrace{\sin(x+y)\,dy}_{dv} = -\left[\left(y-\frac{\pi}{4}\right)\cos(x+y)\right]\Big|_0^{\frac{\pi}{2}} + \int_0^{\frac{\pi}{2}}\cos(x+y)\,dy =$$

$$= -\frac{\pi}{4}\cos\left(x+\frac{\pi}{2}\right) - \frac{\pi}{4}\cos x + \sin(x+y)\Big|_0^{\frac{\pi}{2}} =$$

$$= \frac{\pi}{4}\sin x - \frac{\pi}{4}\cos x + \sin\left(x+\frac{\pi}{2}\right) - \sin x = \left(1-\frac{\pi}{4}\right)(\cos x - \sin x).$$

Отсюда

$$\mu_{xy} = \frac{1}{2}\left(1-\frac{\pi}{4}\right)\int_0^{\frac{\pi}{2}}\underbrace{\left(x-\frac{\pi}{4}\right)}_{u_1}\underbrace{(\cos x - \sin x)\,dx}_{dv_1}.$$

Затем, интегрируя по частям, найдем

$$\mu_{xy} = \frac{8\pi - \pi^2 - 16}{16}.$$

Для определения коэффициента корреляции r_{xy}, предварительно найдем σ_x^2 и σ_y^2. Имеем

$$\sigma_x^2 = M(X^2) - M^2(X),$$

$$\sigma_y^2 = M(Y^2) - M^2(Y),$$

$$M(X^2) = \frac{1}{2}\int_0^{\frac{\pi}{2}} x^2(\sin x + \cos x)\,dx,$$

$$M(Y^2) = \frac{1}{2}\int_0^{\frac{\pi}{2}} y^2(\sin y + \cos y)\,dy,$$

т. е. $M(X^2) = M(Y^2)$.

Но

$$\frac{1}{2}\int_0^{\frac{\pi}{2}} x^2(\sin x + \cos x)\,dx = \frac{\pi^2}{8} + \frac{\pi}{2} - 2$$

(два раза применяли операцию интегрирования по частям).

Следовательно,

$$\sigma_x^2 = \sigma_y^2 = \sigma_x\sigma_y = \frac{\pi^2}{8} + \frac{\pi}{2} - 2 - \frac{\pi^2}{16} = \frac{\pi^2 + 8\pi - 32}{16}.$$

Таким образом, коэффициент корреляции

$$r_{xy} = \frac{8\pi - \pi^2 - 16}{\pi^2 + 8\pi - 32}.$$

Теорема. *Абсолютная величина корреляционного момента двух случайных величин не превосходит произведения их средних квадратических отклонений:*

$$|\mu_{xy}| \leqslant \sigma_x\sigma_y \quad [\sigma_x = \sigma(X),\ \sigma_y = \sigma(Y)].$$

Доказательство. Введя в рассмотрение случайную величину $Z_1 = \sigma_y X - \sigma_x Y$, где $\sigma_x = \sigma(X)$, $\sigma_y = \sigma(Y)$, найдем ее дисперсию. Имеем

$$D(Z) = M[Z_1 - M(Z_1)]^2 = M[\sigma_y(X - M(X)) - \sigma_x(Y - M(Y))]^2 =$$
$$= M[\sigma_y^2(X^2 - 2XM(X) + M^2(X)) - \sigma_x(Y^2 - 2YM(Y) + M^2(Y)) -$$
$$- 2\sigma_x\sigma_y(X - M(X))(Y - M(Y))] = \sigma_y^2(M(X^2) - M^2(X)) +$$
$$+ \sigma_x^2(M(Y^2) - M^2(Y)) - 2\sigma_x\sigma_y M[(X - M(X))(Y - M(Y))] =$$
$$= \sigma_y^2\sigma_x^2 + \sigma_x^2\sigma_y^2 - 2\sigma_x\sigma_y\mu_{xy} = 2\sigma_x^2\sigma_y^2 - 2\sigma_x\sigma_y\mu_{xy} \geqslant 0$$

(любая дисперсия неотрицательна).

Отсюда

$$\mu_{xy} \leqslant \sigma_x\sigma_y.$$

Введя случайную величину $Z_2 = \sigma_y X + \sigma_x Y$, аналогично найдем

$$\mu_{xy} \geqslant -\sigma_x\sigma_y.$$

В результате имеем

$$-\sigma_x\sigma_y \leqslant \mu_{xy} \leqslant \sigma_x\sigma_y$$

или

$$|\mu_{xy}| \leqslant \sigma_x\sigma_y. \tag{3.16}$$

Определение 2. Случайные величины X и Y называются некоррелированными, если $r_{xy} = 0$, и коррелированными, если $r_{xy} \neq 0$.

Пример 1. Независимые случайные величины X и Y являются некоррелированными, так как в силу соотношения (3.12) $r_{xy} = 0$.

Пример 2. Пусть случайные величины X и Y связаны линейной зависимостью $Y = AX + B$, $A \neq 0$. Найдем коэффициент корреляции. Имеем:

$$\mu_{xy} = M(XY) - M(X)M(Y) = M(AX^2 + BX) - M(X)M(AX + B) =$$
$$= AM(X^2) + BM(X) - AM^2(X) - BM(X) =$$
$$= A(M(X^2) - M^2(X)) = A\sigma_x^2,$$
$$\sigma_y^2 = D(Y) = D(AX + B) = D(AX) = A^2D(X) = A^2\sigma_x^2,$$

откуда

$$\sigma_y = |A|\sigma_x.$$

Поэтому

$$|r_{xy}| = \frac{|A|\sigma_x^2}{|A|\sigma_x^2} = 1.$$

Таким образом, коэффициент корреляции случайных величин, связанных линейной зависимостью, равен ± 1 (точнее, $r_{xy} = 1$, если $A > 0$ и $r_{xy} = -1$, если $A < 0$).

Отметим некоторые свойства коэффициента корреляции.

Из примера 1 следует:

1) *Если X и Y — независимые случайные величины, то коэффициент корреляции равен нулю.*

Заметим, что обратное утверждение, вообще говоря, неверно. (Доказательство см. в работе [2].)

2) *Абсолютная величина коэффициента корреляции не превосходит единицы:*

$$|r_{xy}| \leqslant 1.$$

Действительно, разделив обе части неравенства (3.16) на произведение $\sigma_x\sigma_y$, приходим к искомому неравенству.

3) Как видно из формулы (3.15) с учетом формулы (3.14), коэффициент корреляции характеризует относительную величину отклонения математического ожидания произведения $M(XY)$ от произведения математических ожиданий $M(X)$ $M(Y)$ величин X и Y. Так как это отклонение имеет место только для зависимых величин, то можно сказать, что *коэффициент корреляции характеризует тесноту зависимости между X и Y.*

3. Линейная корреляция. Этот вид корреляционной зависимости встречается довольно часто.

Определение. Корреляционная зависимость между случайными величинами X и Y называется *линейной корреляцией*, если обе функции регрессии $f(y)$ и $g(x)$ являются линейными. В этом случае обе линии регрессии являются прямыми, их называют *прямыми регрессии.*

Выведем уравнения прямой регрессии Y на X, т. е. найдем коэффициент линейной функции $g(x) = Ax + B$.

Обозначим $M(X) = a$, $M(Y) = b$, $M[(X-a)^2] = \sigma_x^2$, $M[(Y-b)^2] = \sigma_y^2$. С использованием свойств МО (§§ 2.2; 2.6) находим:

$$M(Y) = M[g(X)] = M(AX + B) = AM(X) + B,$$

т. е. $b = Aa + B$, откуда $B = b - Aa$.

Далее, с помощью тех же свойств математического ожидания имеем

$$M(XY) = M[Xg(X)] = M(AX^2 + BX) =$$
$$= AM(X^2) + BM(X) = AM(X^2) + (b - Aa)a,$$

откуда

$$A = \frac{\mu_{xy}}{M(X^2) - a^2}$$

или, согласно свойству 1 дисперсии (§§ 2.3; 2.6),

$$A = \frac{\mu_{xy}}{\sigma_x^2}.$$

Полученный коэффициент называется *коэффициентом регрессии Y на X* и обозначается через $\rho(Y/X)$:

$$\rho(Y/X) = \frac{\mu_{xy}}{\sigma_x^2}. \qquad (3.17)$$

Таким образом, уравнение прямой регрессии Y на X имеет вид

$$y = \rho(Y/X)(x - a) + b. \qquad (3.18)$$

Аналогично можно получить уравнение прямой регрессии X на Y

$$x = \rho(X/Y)(y - b) + a, \qquad (3.19)$$

где

$$\rho(X/Y) = \frac{\mu_{xy}}{\sigma_y^2} \qquad (3.20)$$

есть коэффициент регрессии X на Y.

Уравнения прямых регрессии можно записать в более симметричном виде, если воспользоваться коэффициентом корреляции. С учетом этого коэффициента имеем:

$$\rho(Y/X) = r_{xy}\frac{\sigma_y}{\sigma_x}; \; \rho(X/Y) = r_{xy}\frac{\sigma_x}{\sigma_y} \qquad (3.21)$$

и поэтому уравнения прямых регрессии принимают вид:

$$y - b = r_{xy}\frac{\sigma_y}{\sigma_x}(x - a);$$

$$x - a = r_{xy}\frac{\sigma_x}{\sigma_y}(y - b).$$

Из уравнений прямых регрессии видно, что обе эти прямые проходят через точку (a; b); угловые коэффициенты прямых регрессии равны соответственно (рис. 13):

$$\operatorname{tg}\alpha = r_{xy}\frac{\sigma_y}{\sigma_x}; \; \operatorname{tg}\beta = \frac{1}{r_{xy}}\frac{\sigma_y}{\sigma_x}.$$

Так как $|r_{xy}| \leqslant 1$, то $|\operatorname{tg}\alpha| \leqslant |\operatorname{tg}\beta|$. Это означает, что прямая регрессии Y на X имеет меньший наклон к оси абсцисс, чем прямая регрессии X на Y. Чем ближе $|r_{xy}|$ к единице, тем меньше угол между прямыми регрессии. Эти прямые сливаются тогда и только тогда, когда $|r_{xy}| = 1$.

При $r_{xy} = 0$ прямые регрессии описываются уравнениями $y = b$; $x = a$.

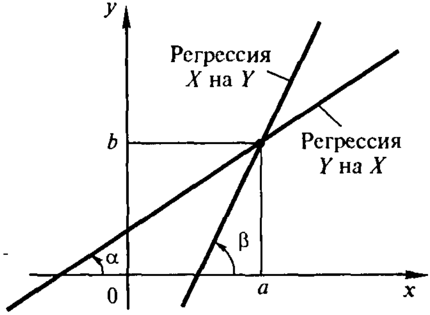


Рис. 13

В этом случае $M_x(Y) = b = M(Y)$; $M_y(X) = a = M(X)$.

Из формул (3.21) видно, что коэффициенты регрессии имеют тот же знак, что и коэффициент корреляции r_{xy}, и связаны соотношением

$$\rho(Y/X)\rho(X/Y) = r_{xy}^2.$$

4. Нормальное распределение двумерной случайной величины.

Определение. Распределение двумерной случайной величины (X, Y) называется *нормальным*, если ее плотность вероятности определяется выражением:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r_{xy}^2}}\, e^{-\frac{1}{2(1-r_{xy}^2)}\left[\frac{(x-a_1)^2}{\sigma_x^2}+\frac{(y-a_2)^2}{\sigma_y^2}-2r_{xy}\frac{x-a_1}{\sigma_x}\,\frac{y-a_2}{\sigma_y}\right]}.$$

Нормальное распределение зависит от пяти параметров a_1, a_2, σ_x, σ_y и r_{xy}. Можно доказать, что a_1 и a_2 — математические ожидания случайных величин X и Y, σ_x и σ_y — их средние квадратические отклонения и r_{xy} — коэффициент корреляции этих величин.

Покажем, что если составляющие двумерной нормально распределенной случайной величины некоррелированы, то они и независимы. Действительно, если X и Y некоррелированы, то $r_{xy} = 0$ и, следовательно,

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\, e^{-\frac{(x-a_1)^2}{2\sigma_x^2}-\frac{(y-a_2)^2}{2\sigma_y^2}} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{(x-a_1)^2}{2\sigma_x^2}}\, \frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{(y-a_2)^2}{2\sigma_y^2}} = f_1(x)f_2(y),$$

отсюда и следует независимость составляющих X и Y (см. § 3.6, следствие).

Справедливо и обратное утверждение.

Таким образом, понятия «некоррелированные величины» и «независимые величины» для случая нормального распределения равносильны.

Замечание. Опираясь на выражения (3.8) и (3.8′), можно доказать, что если двумерная случайная величина распределена нормально с параметрами a_1, a_2, σ_x, σ_y и r_{xy}, то ее составляющие также распределены нормально с параметрами, соответственно равными a_1, σ_x и a_2, σ_y.

Упражнения

1. Найдите законы распределения составляющих дискретной двумерной случайной величины, заданной законом распределения в виде таблицы

X \ Y	y_1	y_2
x_1	0,12	0,10
x_2	0,18	0,11
x_3	0,10	0,39

X	x_1	x_2	x_3	Y	y_1	y_2
p	0,22	0,29	0,49	p	0,40	0,60

2. Найдите вероятность того, что составляющая X двумерной случайной величины примет значение $X < \frac{1}{3}$ и при этом составляющая Y примет значение $Y < \frac{1}{2}$, если функция распределения величины (X, Y)
$F(x,y) = \left(\frac{1}{\pi}\operatorname{arctg} 3x + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg} 2y + \frac{1}{2}\right)$. $\left[\frac{9}{16}\right]$

3. Найдите вероятность попадания случайно поставленной точки (X, Y) в прямоугольник, ограниченный прямыми $x = 0$, $x = \frac{\pi}{2}$, $y = 0$, $y = \frac{\pi}{6}$, если функция распределения двумерной случайной величины $F(x, y) = \sin x \sin y$ $\left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right)$. $\left[\frac{1}{2}\right]$

4. Найдите плотность вероятности $f(x, y)$ двумерной случайной величины по известной функции распределения $F(x, y) = (1 - e^{-x})(1 - e^{-y})$ $(x \geqslant 0,\ y \geqslant 0)$. $[f(x, y) = e^{-x-y}]$

5. Плотность вероятности двумерной случайной величины определяется выражением

$$f(x, y) = \begin{cases} a(x^2 + y^2), & \text{если } x^2 + y^2 \leqslant 1; \\ 0, & \text{если } x^2 + y^2 > 1. \end{cases}$$

Найдите коэффициент a. $\left[\frac{2}{\pi}\right]$

6. Плотность вероятности двумерной случайной величины определяется выражением

$$f(x, y) = \frac{C}{\pi^2(16 + x^2)(25 + y^2)}.$$

Определите величину C и найдите функцию распределения $F(x, y)$.

$$\left[C = 20;\ F(x, y) = \left(\frac{1}{\pi}\operatorname{arctg}\frac{x}{4} + \frac{1}{2}\right)\left(\frac{1}{\pi}\operatorname{arctg}\frac{y}{5} + \frac{1}{2}\right)\right]$$

7. Плотность вероятности двумерной случайной величины определяется выражением

$$f(x, y) = \frac{1}{\pi}e^{-\frac{x^2 + 2xy + 5y^2}{2}}.$$

Найдите плотности распределения составляющих.

$$\left[\begin{array}{l} f_1(x) = \sqrt{\dfrac{2}{5\pi}}\,e^{-0{,}4x^2}; \\ f_2(x) = \sqrt{\dfrac{2}{\pi}}\,e^{-2y^2}. \end{array}\right]$$

8. **Дискретная** двумерная случайная величина задана таблицей

X \ Y	y_1	y_2
x_1	0,15	0,05
x_2	0,30	0,12
x_3	0,35	0,03

Найдите: а) условный закон распределения составляющей X при условии, что составляющая Y приняла значение y_1; б) условный закон распределения составляющей Y при условии, что составляющая X приняла значение x_2.

$$\left[\begin{array}{lccc} \text{а) } X & x_1 & x_2 & x_3 \\ \quad P_{y_1}(X) & \frac{3}{16} & \frac{3}{8} & \frac{7}{16} \end{array} \quad \begin{array}{lcc} \text{б) } Y & y_1 & y_2 \\ \quad P_{x_2}(Y) & \frac{5}{7} & \frac{2}{7} \end{array}\right]$$

9. В условиях, изложенных в упражнении 7, найдите условные законы распределения вероятностей составляющих.

$$\left[\begin{array}{l} \varphi_y(x\,|\,y) = \dfrac{1}{\sqrt{2\pi}}\,e^{-0{,}5(x+y)^2}, \\ \psi_x(y\,|\,x) = \sqrt{\dfrac{5}{2\pi}}\,e^{-0{,}1(x+5y)^2}. \end{array}\right]$$

10. Плотность распределения непрерывной двумерной случайной величины $(X,\ Y)$ задана выражением

$$f(x, y) = \begin{cases} 2\cos x\cos y \ \text{ в квадрате } S\left\{0 < x < \frac{\pi}{4},\ 0 < y < \frac{\pi}{4}\right\}, \\ 0 \ \text{ вне квадрата } S. \end{cases}$$

Докажите, что составляющие X и Y независимы.

11. Дана таблица, определяющая закон распределения двумерной дискретной случайной величины

$X \backslash Y$	20	40	60
10	$\frac{3}{20}$	$\frac{1}{20}$	0
20	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{10}$
30	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{4}$

Найдите коэффициент корреляции r_{xy}. [$r_{xy} = 0{,}56$]

12. Задана плотность вероятности непрерывной двумерной случайной величины

$$f(x, y) = \begin{cases} \frac{1}{4}\sin x \sin y & \text{в квадрате } S\left\{0 \leqslant x \leqslant \pi,\ 0 \leqslant y \leqslant \pi\right\}, \\ 0 & \text{вне квадрата } S \end{cases}$$

Найдите корреляционный момент μ_{xy} и коэффициент корреляции r_{xy}.

[$\mu_{xy} = 0,\ r_{xy} = 0$]

ГЛАВА IV

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

§ 4.1. Генеральная совокупность и выборка

Приступим к изучению элементов математической статистики, в которой разрабатываются научно обоснованные методы сбора статистических данных и их обработки.

1. Генеральная совокупность и выборка. Пусть требуется изучить множество однородных объектов (это множество называют *статистической совокупностью*) относительно некоторого качественного или количественного признака, характеризующего эти объекты. Например, если имеется партия деталей, то качественным признаком может служить соответствие детали стандартам, а количественным — контролируемый размер детали.

Лучше всего осуществить сплошное обследование, т. е. изучить каждый объект. Однако в большинстве случаев по разным причинам это сделать невозможно. Препятствовать сплошному обследованию может большое число объектов, их недоступность и т. п. Если, например, нужно знать среднюю глубину воронки при взрыве снаряда из опытной партии, то, проводя сплошное обследование, мы должны будем уничтожить всю партию.

Если сплошное обследование невозможно, то из всей совокупности выбирают для изучения часть объектов.

Статистическая совокупность, из которой отбирают часть объектов, называется *генеральной совокупностью.* Множество объектов, случайно отобранных из генеральной совокупности, называется *выборкой.*

Число объектов генеральной совокупности и выборки называется соответственно *объемом генеральной совокупности и объемом выборки.*

Пример. Плоды одного дерева (200 шт.) обследуют на наличие специфического для данного сорта вкуса. Для этого отбирают 10 шт. Здесь 200 — объем генеральной совокупности, а 10 — объем выборки.

Если выборку отбирают по одному объекту, который обследуют и снова возвращают в генеральную совокупность, то выборка называется *повторной*. Если объекты выборки уже не возвращаются в генеральную совокупность, то выборка называется *бесповторной*. На практике чаще используется бесповторная выборка. Если объем выборки составляет небольшую долю объема генеральной совокупности, то разница между повторной и бесповторной выборками незначительна

Свойства объектов выборки должны правильно отражать свойства объектов генеральной совокупности, или, как говорят, выборка должна быть *репрезентативной* (представительной). Считается, что выборка репрезентативна, если все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку, т. е. выбор осуществляется случайно. Например, для того чтобы оценить будущий урожай, можно сделать выборку из генеральной совокупности еще не созревших плодов и исследовать их характеристики (массу, качество и пр.). Если вся выборка будет взята с одного дерева, то она не будет репрезентативной. Репрезентативная выборка должна состоять из случайно выбранных плодов со случайно выбранных деревьев.

2. Статистическое распределение выборки. Полигон. Гистограмма. Пусть из генеральной совокупности извлечена выборка, причем x_1 наблюдалось n_1 раз, x_2 — n_2 раз, ..., x_k — n_k раз и $n_1+n_2+\ldots+n_k=n$ — объем выборки. Наблюдаемые значения x_1, x_2, ..., x_k называются *вариантами*, а последовательность вариант, записанная в возрастающем порядке,— *вариационным рядом*. Числа наблюдений n_1, n_2, ..., n_k называют частотами, а их отношения к объему выборки $\frac{n_1}{n}=p_1^*$, $\frac{n_2}{n}=p_2^*$, ..., $\frac{n_k}{n}=p_k^*$ — *относительными частотами*. Отметим, что сумма относительных частот равна единице:

$$p_1^*+p_2^*+\ldots+p_k^*=\frac{n_1}{n}+\frac{n_2}{n}+\ldots+\frac{n_k}{n}=\frac{n_1+n_2+\ \ +n_k}{n}=\frac{n}{n}=1.$$

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот. Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот *(непрерывное распределение)*. В качестве частоты, соответствующей интервалу, принимают сумму частот вариант, попавших в этот интервал.

Заметим, что в теории вероятностей под распределением понимают соответствие между возможными значениями случайной величины и их вероятностями, а в математической статистике — соответствие между наблюдаемыми вариантами и их частотами или относительными частотами.

Пример 1. Перейдем от частот к относительным частотам в следующем распределении выборки объема $n=20$:

Варианта x_i	2	6	12
Частота n_i	3	10	7

Найдем относительные частоты:

$$p_1^* = \frac{3}{20} = 0{,}15; \; p_2^* = \frac{10}{20} = 0{,}50; \; p_3^* = \frac{7}{20} = 0{,}35.$$

Поэтому получаем следующее распределение:

Варианта x_i	2	6	12
Относительная частота p_i^*	0,15	0,50	0,35

Для графического изображения статистического распределения используются *полигоны* и *гистограммы*.

Для построения полигона в декартовых координатах на оси Ox откладывают значения вариант x_i, на оси Oy — значения частот n_i (относительных частот p_i^*).

Пример 2. Рис. 14 представляет собой полигон следующего распределения:

Варианта x_i	1	2	3	5
Относительная частота p_i^*	0,4	0,2	0,3	0,1

Полигоном обычно пользуются в случае небольшого количества вариант. В случае большого количества вариант и в случае непрерывного распределения признака чаще строят гистограммы. Для этого интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько частичных интервалов шириной h и находят для каждого частичного интервала n_i — сумму частот вариант, попавших в i-й интервал. Затем на этих интервалах как на основаниях строят прямоугольники с высотами $\frac{n_i}{h}$ (или $\frac{n_i}{nh}$, где n — объем выборки). Площадь i-го частичного прямоугольника равна $\frac{hn_i}{h} = n_i$

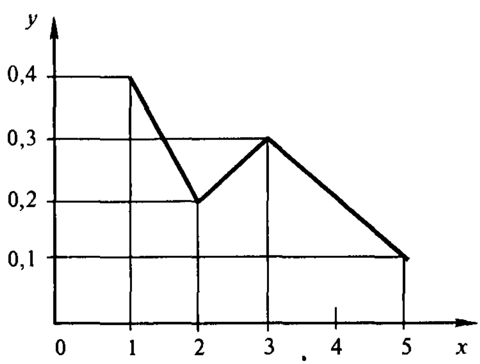


Рис. 14

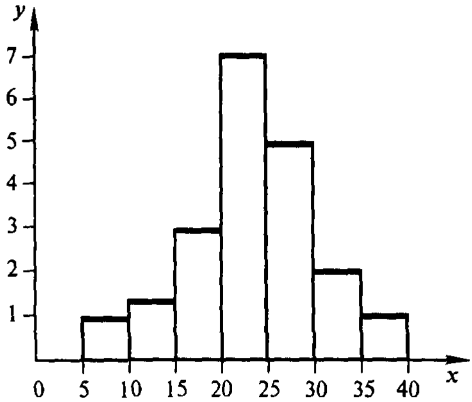


Рис. 15

$\left(\text{или } \frac{hn_i}{nh} = \frac{n_i}{n} = p_i^*\right)$. Следовательно, площадь гистограммы равна сумме всех частот (или относительных частот), т. е. объему выборки (или единице).

Пример 3. Рис. 15 показывает гистограмму непрерывного распределения объема $n = 100$, заданного следующей таблицей:

Частичный интервал, h	Сумма частот вариант частичного интервала, n_i	$\frac{n_i}{h}$
5—10	4	0,8
10—15	6	1,2
15—20	16	3,2
20—25	36	7,2
25—30	24	4,8
30—35	10	2,0
35—40	4	0,8

§ 4.2. Оценки параметров генеральной совокупности по ее выборке

1. Выборка как набор случайных величин. Пусть имеется некоторая генеральная совокупность, каждый объект которой наделен количественным признаком X. При случайном извлечении объекта из генеральной совокупности становится известным значение x признака X этого объекта. Таким образом, мы можем рассматривать извлечение объекта из генеральной совокупности как испытание, X — как случайную величину, а x — как одно из возможных значений X.

Допустим, что из теоретических соображений удалось установить, к какому типу распределений относится признак X. Естественно, возникает задача оценки (приближенного определения) параметров, которыми описывается это распределение. Например, если известно, что изучаемый признак распределен в генеральной совокупности нормально, то необходимо оценить, т. е. приближенно найти математическое ожидание и среднее квадратическое отклонение, так как эти два параметра полностью определяют нормальное распределение.

Обычно в распоряжении исследователя имеются лишь данные выборки генеральной совокупности, например значения количественного признака $x_1, x_2, \ldots, x_n$, полученные в результате n наблюдений (здесь и далее наблюдения предполагаются независимыми). Через эти данные и выражают оцениваемый параметр.

Опытные значения признака X можно рассматривать и как значения разных случайных величин $X_1, X_2, \ldots, X_n$ с тем же распределением, что и X, и, следовательно, с теми же числовыми характеристиками, которые имеет X. Значит, $M(X_i) = M(X)$ и $D(X_i) = D(X)$. Величины $X_1, X_2, \ldots, X_n$ можно считать независимыми в силу независимости наблюдений. Значения $x_1, x_2, \ldots, x_n$ в этом случае называются *реализациями* случайных величин $X_1, X_2, \ldots, X_n$. Отсюда и из предыдущего следует, что найти оценку неизвестного параметра — это значит найти функцию от наблюдаемых случайных величин $X_1, X_2, \ldots, X_n$, которая и дает приближенное значение оцениваемого параметра.

2. Генеральная и выборочная средние. Методы их расчета. Пусть изучается дискретная генеральная совокупность объема N относительно количественного признака X.

Определение 1. *Генеральной средней* $\bar{x}_{\text{г}}$ (или a) называется среднее арифметическое значений признака генеральной совокупности.

Если все значения $x_1, x_2, \ldots, x_N$ признака генеральной совокупности объема N различны, то

$$x_{\text{г}} = \frac{1}{N}(x_1 + x_2 + \ldots + x_N).$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $N_1, N_2, \ldots, N_k$, причем $N_1 + N_2 + \ldots + N_k = N$, то

$$\bar{x}_{\text{г}} = \frac{1}{N}(x_1 N_1 + x_2 N_2 + \ldots + x_k N_k),$$

или

$$\bar{x}_{\text{г}} = \frac{1}{N}\sum_{i=1}^{k} x_i N_i. \qquad (4.1)$$

Как уже отмечалось (п. 1), извлечение объекта из генеральной совокупности есть наблюдение случайной величины X.

Пусть все значения $x_1, x_2, ..., x_N$ различны. Так как каждый объект может быть извлечен с одной и той же вероятностью $1/N$, то

$$M(X) = x_1 \frac{1}{N} + x_2 \frac{1}{N} + \ldots + x_N \frac{1}{N} = \bar{x}_\text{г},$$

т. е.

$$M(X) = \bar{x}_\text{г}. \tag{4.2}$$

Такой же итог следует, если значения $x_1, x_2, ..., x_k$ имеют соответственно частоты $N_1, N_2, ..., N_k$.

В случае непрерывного распределения признака X по определению полагают $\bar{x}_\text{г} = M(X)$.

Пусть для изучения генеральной совокупности относительно количественного признака X произведена выборка объема n.

Определение 2. Выборочной средней $\bar{x}_\text{в}$ называется среднее арифметическое значений признака выборочной совокупности.

Если все значения $x_1, x_2, ..., x_n$ признака выборки объема n различны, то

$$\bar{x}_\text{в} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n). \tag{4.3}$$

Если же значения признака $x_1, x_2, ..., x_k$ имеют соответственно частоты $n_1, n_2, ..., n_k$, причем $n_1 + n_2 + ... + n_k = n$, то

$$\bar{x}_\text{в} = \frac{1}{n}(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k),$$

или

$$\bar{x}_\text{в} = \frac{1}{n}\sum_{i=1}^{k} x_1 n_i. \tag{4.4}$$

Пример 1. *Выборочным путем* были получены следующие данные о массе 20 морских свинок при рождении (в г): 30, 30, 25, 32, 30, 25, 33, 32, 29, 28, 27, 36, 31, 34, 30, 23, 28, 31, 36, 30. Найдем выборочную среднюю $\bar{x}_\text{в}$.

Согласно формуле (4.4), имеем:

$$\bar{x}_\text{в} = \frac{30\ \ 5 + 25\ \ 2 + 32\ \ 2 + 33 + 29 + 28\ \ 2 + 27 + 36\ \ 2 + 31\ \ 2 + 34 + 23}{20} = 30.$$

Итак, $\bar{x}_\text{в} = 30$.

Далее, не уменьшая общности рассуждений, будем считать значения $x_1, x_2, ..., x_n$ признака различными.

Разумеется, выборочная средняя для различных выборок того же объема n из той же генеральной совокупности будет получаться, вообще говоря, различной. И это не удивительно — ведь извлечение i-го по счету объекта есть наблюдение случайной величины X_1, а их среднее арифметическое

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

есть тоже случайная величина.

Таким образом, всевозможные получающиеся выборочные средние есть возможные значения случайной величины $\overline{X}$, которая называется выборочной средней случайной величиной.

Найдем $M(\overline{X})$, пользуясь тем, что $M(X_i) = M(X)$ (см. п. 1).

С учетом свойств математического ожидания (см. гл. II) получаем:

$$M(\overline{X}) = M\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right] = \frac{1}{n}[M(X_1) + M(X_2) + \ldots + M(X_n)] =$$

$$= \frac{1}{n}[M(X) + M(X) + \ldots + M(X)] = \frac{1}{n}na = a.$$

Итак, $M(\overline{X})$ (математическое ожидание выборочной средней) совпадает с a (генеральной средней).

Теперь найдем $D(\overline{X})$. Так как $D(X_i) = D(X)$ (п. 1) и $X_1, X_2, \ldots, X_n$ независимы, то, согласно свойствам дисперсии (см. гл. II), получаем

$$D(\overline{X}) = D\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right] = \frac{1}{n^2}[D(X_1) + D(X_2) + \ldots + D(X_n)] =$$

$$= \frac{1}{n^2}[D(X) + D(X) + \ldots + D(X)] = \frac{1}{n^2}nD(X) = \frac{D(X)}{n},$$

т. е.

$$D(\overline{X}) = \frac{D(X)}{n}. \qquad (4.5)$$

Наконец, отметим, что если варианты x_i — большие числа, то для облегчения вычисления выборочной средней применяют следующий прием. Пусть C — константа.

Так как

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}(x_i - C) + nC,$$

то формулу (4.3) можно преобразовать к виду

$$\bar{x}_в = C + \frac{1}{n}\sum_{i=1}^{n}(x_i - C). \qquad (4.6)$$

За константу C (так называемый ложный нуль) берут некоторое среднее значение между наименьшим и наибольшим значениями x_i ($i = 1, 2, \ldots, n$).

Пример 2. Имеется выборка:

$$x_1 = 71{,}88;\; x_2 = 71{,}93;\; x_3 = 72{,}05;\; x_4 = 72{,}07;$$

$$x_5 = 71{,}90;\; x_6 = 72{,}02;\; x_7 = 71{,}93;\; x_8 = 71{,}77;$$

$$x_9 = 72{,}11;\; x_{10} = 71{,}96.$$

Требуется найти $\bar{x}_в$.

Возьмем $C = 72{,}00$ и вычислим разности $\alpha_i = x_i - C$:

$$\alpha_1 = -0{,}12;\; \alpha_2 = -0{,}07;\; \alpha_3 = 0{,}05;\; \alpha_4 = 0{,}07;$$

$$\alpha_5 = -0{,}10;\; \alpha_6 = 0{,}02;\; \alpha_7 = -0{,}07;\; \alpha_8 = -0{,}23;$$

$$\alpha_9 = 0{,}11;\; \alpha_{10} = -0{,}04.$$

Их сумма: $\alpha_1 + \alpha_2 + \ldots + \alpha_{10} = -0{,}38$; их среднее арифметическое $\frac{1}{10}(\alpha_1 + \alpha_2 + \ldots + \alpha_{10}) = -0{,}038 \approx -0{,}04$. Выборочная средняя

$$\bar{x}_в = 72{,}00 - 0{,}04 = 71{,}96.$$

3. Генеральная и выборочная дисперсии. Для того чтобы охарактеризовать рассеяние значений количественного признака X генеральной совокупности вокруг своего среднего значения, вводят следующую характеристику — генеральную дисперсию.

Определение 1. *Генеральной дисперсией* $D_г$ называется среднее арифметическое квадратов отклонений значений признака X генеральной совокупности от генеральной средней $\bar{x}_г$.

Если все значения $x_1, x_2, \ldots, x_N$ признака генеральной совокупности объема N различны, то

$$D_г = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_г)^2.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $N_1, N_2, \ldots, N_k$, причем $N_1 + N_2 + \ldots + N_k = N$, то

$$D_г = \frac{1}{N}\sum_{i=1}^{k}(x_i - \bar{x}_г)^2 N_i. \qquad (4.7)$$

Пример 1. Генеральная совокупность задана таблицей распределения:

x_i	2	4	5	6
N_i	8	9	10	3

Найдем генеральную дисперсию.

Согласно формулам (4.1) и (4.7), имеем:

$$\bar{x}_г = \frac{2\cdot 8 + 4\cdot 9 + 5\cdot 10 + 6\cdot 3}{8 + 9 + 10 + 3} = \frac{120}{30} = 4;$$

$$D_г = \frac{(2-4)^2\cdot 8 + (4-4)^2\cdot 9 + (5-4)^2\cdot 10 + (6-4)^2\cdot 3}{30} = \frac{54}{30} = 1{,}8.$$

Генеральным средним квадратическим отклонением (стандартом) называется $\sigma_г = \sqrt{D_г}$.

Пусть все значения $x_1, x_2, \ldots, x_N$ различны.

Найдем дисперсию признака X, рассматриваемого как случайная величина:

$$D(X) = M[(X - M(X))^2].$$

Так как $M(X) = \bar{x}_г$ и $P\{X = x_i\} = \frac{1}{N}$ (см. п. 2), то

$$D(X) = (x_1 - \bar{x}_г)^2 \frac{1}{N} + (x_2 - \bar{x}_г)^2 \frac{1}{N} + \ldots + (x_N - \bar{x}_г)^2 \frac{1}{N} = D_г,$$

т. е.

$$D(\bar{X}) = D_г.$$

Таким образом, дисперсия $D(X)$ равна $D_г$.

Такой же итог можно получить, если значения $x_1, x_2, \ldots, x_k$ имеют соотвественно частоты $N_1, N_2, \ldots, N_k$.

В случае непрерывного распределения признака X по определению полагают

$$D_г = D(X). \tag{4.8}$$

С учетом формулы (4.8) формула (4.5) (п. 2) перепишется в виде

$$D(\bar{X}) = \frac{D_г}{n},$$

откуда $\sqrt{D(\bar{X})} = \sqrt{D_г}/\sqrt{n}$ или $\sigma(\bar{X}) = \sigma_г/\sqrt{n}$. Величина $\sigma(\bar{X})$ называется *средней квадратической ошибкой.*

Для того чтобы охарактеризовать рассеяние наблюдаемых значений количественного признака выборки вокруг своего среднего значения $\bar{x}_в$, вводят выборочную дисперсию.

Определение 2. *Выборочной дисперсией* $D_в$ называется среднее арифметическое квадратов отклонений наблюдаемых значений признака X от выборочной средней $\bar{x}_в$.

Если все значения $x_1, x_2, \ldots, x_n$ признака выборки объема n различны, то

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2. \tag{4.9}$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k$, причем $n_1 + n_2 + \ldots + n_k = n$, то

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2 n_i. \tag{4.10}$$

Пример 2. Пусть выборочная совокупность задана таблицей распределения:

x_i	1	2	3	4
n_i	20	15	10	5

Найдем выборочную дисперсию. Согласно формулам (4.4) и (4.10), имеем:

$$\bar{x}_в = \frac{1 \cdot 20 + 2 \cdot 15 + 3 \cdot 10 + 4 \cdot 5}{20 + 15 + 10 + 5} = \frac{100}{50} = 2;$$

$$D_в = \frac{(1-2)^2 \cdot 20 + (2-2)^2 \cdot 15 + (3-2)^2 \cdot 10 + (4-2)^2 \cdot 5}{50} = \frac{50}{50} = 1.$$

Выборочным средним квадратическим отклонением (стандартом) называется квадратный корень из выборочной дисперсии:

$$\sigma_в = \sqrt{D_в}.$$

В условиях примера 2 получаем, что $\sigma_в = \sqrt{D_в} = \sqrt{1} = 1$.

Далее, не уменьшая общности рассуждений, будем считать значения $x_1, x_2, \ldots, x_n$ признака различными.

Выборочную дисперсию, рассматриваемую нами как случайная величина, будем обозначать $\widetilde{S}^2$:

$$\widetilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Т е о р е м а. Математическое ожидание выборочной дисперсии равно $((n-1)/n)D_г$, т. е.

$$M(\widetilde{S}^2) = \frac{n-1}{n} D_г.$$

Д о к а з а т е л ь с т в о. С учетом свойств математического ожидания (см. гл. II) получаем

$$M(\widetilde{S}^2) = M\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{1}{n}\sum_{i=1}^{n} M[(X_i - \bar{X})^2].$$

Вычислим одно слагаемое $M[(X_i - \bar{X})^2]$. Имеем

$$M[(X_i - \bar{X})^2] = M(X_i^2 - 2X_i\bar{X} + \bar{X}^2) =$$

$$= M(X_i^2) - 2M(X_i\bar{X}) + M(\bar{X}^2).$$

Вычислим по отдельности эти математические ожидания.

Согласно свойству 1 дисперсии (см. гл. II) и формулам (4.2), (4.8) имеем

$$M(X_i^2) = M(X^2) = D(X) + M^2(X) = D_г + a^2.$$

Далее, с учетом свойства 4 математического ожидания (см. гл. II)

$$M(X_i\bar{X}) = M\left[X_i \frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right] =$$

$$= \frac{1}{n}[M(X_iX_1) + M(X_iX_2) + \ldots + M(X_iX_n)],$$

но слагаемое этой суммы, у которого второй индекс равен i, т. е. $M(X_iX_i)$, равно $M(X_i^2)=D_г+a^2$. У всех остальных слагаемых $M(X_iX_j)$ индексы разные. Поэтому в силу независимости X_i и X_j (см. гл. II)

$$M(X_iX_j)=M(X_i)M(X_j)=M(X)M(X)=M^2(X)=a^2.$$

Так как имеется $n-1$ таких слагаемых, то

$$M(X_i\overline{X})=\frac{1}{n}[D_г+a^2+(n-1)a^2]=a^2+\frac{D_г}{n}.$$

В силу свойства 1 дисперсии (см. гл. II) получаем

$$M(\overline{X}^2)=D(\overline{X})+M^2(\overline{X}).$$

Нами уже найден (см. пп. 2 и 3):

$$M(\overline{X})=M(X)=a;\ \ D(\overline{X})=\frac{1}{n}D_г.$$

Поэтому

$$M(\overline{X}^2)=\frac{D_г}{n}+a^2.$$

Таким образом,

$$M[(X_i-\overline{X})^2]=D_г+a^2-2\left(a^2+\frac{D_г}{n}\right)+\frac{D_г}{n}+a^2=\frac{n-1}{n}D_г$$

и не зависит от индекса суммирования i. Поэтому

$$M(\widetilde{S}^2)=\frac{1}{n}n\frac{n-1}{n}D_г=\frac{n-1}{n}D_г.$$

Что и требовалось доказать.

В заключение этого пункта отметим, что если варианты x_i — большие числа, то для облегчения вычисления выборочной дисперсии $D_в$ формулу (4.9) преобразуют к следующему виду:

$$D_в=\frac{1}{n}\sum_{i=1}^{n}(x_i-C)^2-(\overline{x}_в-C)^2, \tag{4.11}$$

где C — ложный нуль.

Действительно, с учетом формулы (4.3) имеем

$$\sum_{i=1}^{n}(x_i-C)^2-n(\overline{x}_в-C)^2=\sum_{i=1}^{n}x_i^2-2C\sum_{i=1}^{n}x_i+nC^2-n\overline{x}_в^2+2nC\overline{x}_в-nC^2=$$
$$=\sum_{i=1}^{n}x_i^2-2nC\overline{x}_в-n\overline{x}_в^2+2nC\overline{x}_в=\sum_{i=1}^{n}x_i^2-n\overline{x}_в^2=\sum_{i=1}^{n}x_i^2-2n\overline{x}_в^2+n\overline{x}_в^2=$$
$$=\sum_{i=1}^{n}x_i^2-2x_в\sum_{i=1}^{n}x_i+n\overline{x}_в^2=\sum_{i=1}^{n}(x_i^2-2x_i\overline{x}_в+\overline{x}_в^2)=\sum_{i=1}^{n}(x_i-\overline{x}_в)^2,$$

откуда

$$D_в=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x}_в)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-C)^2-(\overline{x}_в-C)^2.$$

Пример 3. Для выборки, указанной в примере 2 из п. 2, найдем $D_в$ (ложный нуль остается прежним $C = 72{,}00$)

$$\sum_{i=1}^{10}(x_i - C)^2 = \sum_{i=1}^{10}\alpha_i^2 = 0{,}0144 + 0{,}0049 + 0{,}0025 + 0{,}0049 +$$

$$+ 0{,}0100 + 0{,}0004 + 0{,}0049 + 0{,}0529 + 0{,}0121 + 0{,}0016 = 0{,}1086;$$

$$(\bar{x}_в - C)^2 = (-0{,}038)^2 \approx 0{,}0014.$$

Наконец, согласно формуле (4.11)

$$D_в \approx \frac{1}{10}\cdot 0{,}1086 - 0{,}0014 = 0{,}0094.$$

4. Оценки параметров распределения. Одной из задач статистики является оценка параметров распределения случайной величины X по данным выборки. При этом в теоретических рассуждениях считают, что генеральная совокупность бесконечна. Это делается для того, чтобы можно было переходить к пределу при $n \to \infty$, где n — объем выборки. Для оценки параметров распределения X из данных выборки составляют выражения, которые должны служить оценками неизвестных параметров. Например, $\bar{X}$ (см. п. 2) является оценкой генеральной средней, а $\tilde{S}^2$ (см. п. 3) — оценкой генеральной дисперсии $D_г$. Обозначим через Θ оцениваемый параметр, через $\tilde{\Theta}_n$ — оценку этого параметра [$\tilde{\Theta}_n$ является выражением, составленным из $X_1, X_2, \ldots, X_n$ (см. п. 1)]. Для того чтобы оценка $\tilde{\Theta}_n$ давала хорошее приближение, она должна удовлетворять определенным требованиям. Укажем эти требования.

Несмещенной называют оценку $\tilde{\Theta}_n$, математическое ожидание которой равно оцениваемому параметру Θ, т. е. $M(\tilde{\Theta}_n) = \Theta$, в противном случае оценка называется *смещенной.*

Пример 1. Оценка $\bar{X}$ является несмещенной оценкой генеральной средней a, так как $M(\bar{X}) = a$ (см. п. 2).

Пример 2. Оценка $\tilde{S}^2$ является смещенной оценкой генеральной дисперсии $D_г$, так как, согласно установленной выше теореме (см. п. 3),

$$M(\tilde{S}^2) = \frac{n-1}{n}D_г \neq D_г.$$

Пример 3. Наряду с выборочной дисперсией $\tilde{S}^2$ рассматривают еще так называемую исправленную дисперсию $S^2 = \frac{n}{n-1}\tilde{S}^2$, которая является также оценкой генеральной дисперсии. Для S^2 с учетом установленной выше теоремы (см. п. 3) имеем

$$M(S^2) = M\left(\frac{n}{n-1}\tilde{S}^2\right) = \frac{n}{n-1}M(\tilde{S}^2) = \frac{n}{n-1}\frac{n-1}{n}D_г = D_г.$$

Таким образом, оценка S^2 в отличие от оценки $\widetilde{S}^2$ является несмещенной оценкой генеральной дисперсии. Явное выражение для S^2 имеет вид

$$S^2 = \frac{n}{n-1}\widetilde{S}^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

т. е.

$$S^2 = \frac{n}{n-1}\sum_{i-1}^{n}(X_i - \overline{X})^2. \qquad (4.12)$$

Естественно в качестве приближенного неизвестного параметра брать несмещенные оценки для того, чтобы не делать систематической ошибки в сторону завышения или занижения.

Состоятельной называют такую оценку $\widetilde{\Theta}_n$ параметра Θ, что для любого наперед заданного числа $\varepsilon > 0$ вероятность $P\{|\widetilde{\Theta}_n - \Theta| < \varepsilon\}$ при $n \to \infty$ стремится к единице*. Это значит, что при достаточно больших n можно с вероятностью, близкой к единице, т. е. почти наверное, утверждать, что оценка Θ_n отличается от оцениваемого параметра Θ меньше, чем на ε.

Очевидно, такому требованию должна удовлетворять всякая оценка, пригодная для практического использования.

Заметим, что несмещенная оценка Θ_n будет состоятельной, если при $n \to \infty$ ее дисперсия стремится к нулю: $D(\widetilde{\Theta}_n) \to 0$. Это следует из неравенства Чебышева ((2.33) см. § 2.8, п. 1).

Пример 4. Как было установлено (см. п. 3), $D(\overline{X}) = \frac{D_{\text{г}}}{n}$. Отсюда следует, что несмещенная оценка $\overline{X}$ является и состоятельной, так как

$$\lim_{n\to\infty} D(\overline{X}) = \lim_{n\to\infty}\frac{D_{\text{г}}}{n} = D_{\text{г}}\lim_{n\to\infty}\frac{1}{n} = 0.$$

Можно показать, что несмещенная оценка S^2 является также состоятельной. Поэтому в качестве оценки генеральной дисперсии принимают исправленную дисперсию. Заметим, что оценки S^2 и $\widetilde{S}^2$ отличаются множителем $\frac{n}{n-1}$, который стремится к 1 при $n \to \infty$. На практике $\widetilde{S}^2$ и S^2 не различают при $n > 30$.

Для оценки генерального среднего квадратического отклонения используют исправленное среднее квадратическое отклонение, которое равно квадратному корню из исправленной дисперсии:

$$S = \sqrt{\frac{n}{n-1}\sum_{i-1}^{n}(X_i - \overline{X})^2}. \qquad (4.13)$$

Левые части формул (4.12), (4.13), в которых случайные величины $X_1, X_2, \ldots, X_n$ заменены их реализациями $x_1, x_2, \ldots, x_n$ и $\overline{X}$ —

* В таком случае говорят, что $\widetilde{\Theta}_n$ сходится к Θ по вероятности.

выборочной средней $\bar{x}_в$, будем обозначать соответственно через s^2 и s.

Отметим, что если варианты x_i — большие числа, то для облегчения вычисления s^2 формулу для s^2 аналогично формуле (4.9) преобразуют к виду

$$s^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - C)^2 - (\bar{x}_в - C)^2\right], \qquad (4.14)$$

где C — ложный нуль.

Оценки, обладающие свойствами несмещенности и состоятельности, при ограниченном числе опытов могут отличаться дисперсиями.

Ясно, что чем меньше дисперсия оценки, тем меньше вероятность грубой ошибки при определении приближенного значения параметра. Поэтому необходимо, чтобы дисперсия оценки была минимальной. Оценка, обладающая таким свойством, называется *эффективной.*

Из отмеченных требований, предъявляемых к оценке, наиболее важными являются требования несмещенности и состоятельности.

Пример 5. С плодового дерева случайным образом отобрано 10 плодов. Их массы x_1, x_2, ..., x_{10} (в граммах) записаны в первой колонке приведенной ниже таблицы. Обработаем статистические данные выборки. Для вычисления $\bar{x}_в$ и s по формулам (4.6) и (4.14) введем ложный нуль $C = 250$ и все необходимые при этом вычисления сведем в указанную таблицу:

i	x_i	$x_i - C$	$(x_i - C)^2$
1	225	–25	625
2	274	24	576
3	305	55	3025
4	253	3	9
5	220	–30	900
6	245	–5	25
7	211	–39	1521
8	234	–16	256
9	230	–20	400
10	231	–19	261
Сумма		–72	7598

Следовательно,

$$\bar{x}_в = 250 + \frac{1}{10}\sum_{i=1}^{10}(x_i - 250) = 250 + \frac{1}{10}(-72) = 250 - 7{,}2 \approx 243 \text{ (г)}.$$

$$s=\sqrt{\frac{10}{9}\left[\frac{1}{10}\sum_{i=1}^{10}(x_i-250)^2-(250-7{,}2-250)^2\right]}=$$

$$=\sqrt{\frac{10}{9}[759{,}8-(-7{,}2)^2]}\approx 28 \text{ (г)}.$$

Отсюда $\frac{s}{\sqrt{10}}\approx 9$ (г).

Итак, оценка генеральной средней массы плода равна 243 г со средней квадратической ошибкой 9 г.

Оценка генерального среднего квадратического отклонения массы плода равна 28 г.

П р и м е р 6. Через каждый час измерялось напряжение в электросети. Результаты измерений (в вольтах) представлены в следующей таблице:

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	222	219	224	220	218	217	221	220	215	218	223	225
i	13	14	15	16	17	18	19	20	21	22	23	24
x_i	220	226	221	216	211	219	220	221	222	218	221	219

Найти оценки для математического ожидания и дисперсии результатов измерений. Оценки для математического ожидания и дисперсии найдем по формулам (6) и (14), положив $C=220$. Все необходимые вычисления приведены в нижеследующей таблице:

i	x_i-C	$(x_i-C)^2$	i	x_i-C	$(x_i-C)^2$	i	x_i-C	$(x_i-C)^2$
1	2	4	9	–5	25	17	1	1
2	–1	1	10	–2	4	18	–1	1
3	4	16	11	3	9	19	0	0
4	0	0	12	5	25	20	1	1
5	–2	4	13	0	0	21	2	4
6	–3	9	14	6	36	22	–2	4
7	1	1	15	1	1	23	1	1
8	0	0	16	–4	16	24	–1	1
Сумма	1	35		4	116		1	13

Следовательно,

$$\bar{x}_в=220+\frac{1}{24}\sum_{i=1}^{24}(x_i-220)=220+\frac{6}{24}=220{,}25 \text{ (В)}.$$

$$s^2=\frac{24}{23}\left[-\frac{1}{24}-\sum_{i=1}^{24}(x_i-220)^2-(220{,}25-220)^2\right]\approx 7{,}06 \text{ (В}^2\text{)}.$$

§ 4.3. Доверительные интервалы для параметров нормального распределеиия

1. Надежность. Доверительные интервалы. Пусть Θ — оцениваемый параметр, $\widetilde{\Theta}_n$ — его оценка, составленная из $X_1, X_2, \ldots, X_n$.

Если известно, что оценка $\widetilde{\Theta}_n$ является несмещенной и состоятельной, то по данным выборки вычисляют значение $\widetilde{\Theta}_n$ и считают его приближением истинного значения Θ. При этом среднее квадратическое отклонение (если его вообще вычисляют) оценивает порядок ошибки. Такие оценки называются *точечными*. Например, в предыдущем параграфе речь шла о точечных оценках генеральной средней и генеральной дисперсии. В общем случае, когда о распределении признака X ничего неизвестно, это уже немало.

Если же о распределении имеется какая-либо информация, то можно сделать больше.

Здесь речь будет идти об оценке параметров a и σ случайной величины, имеющей нормальное распределение. Это очень важный случай. Например (см. § 2.7), результат измерения имеет нормальное распределение. В этом случае становится возможным применять так называемое *интервальное оценивание*, к изложению которого мы и переходим.

Пусть $\delta > 0$ — некоторое число. Если выполняется неравенство $|\Theta - \widetilde{\Theta}_n| < \delta$, т. е. $-\delta < \Theta - \widetilde{\Theta}_n < \delta$, что можно записать в виде $\widetilde{\Theta}_n - \delta < \Theta < \widetilde{\Theta}_n + \delta$, то говорят, что интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$ покрывает параметр Θ. Однако невозможно указать оценку $\widetilde{\Theta}_n$ такую, чтобы событие $\{|\Theta - \widetilde{\Theta}_n| < \delta\}$ было достоверным, поэтому мы будем говорить о вероятности этого события. Число δ называется точностью оценки $\widetilde{\Theta}_n$.

Определение. *Надежностью* (доверительной вероятностью) оценки $\widetilde{\Theta}_n$ параметра Θ для заданного $\delta > 0$ называется вероятность γ того, что интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$ покроет параметр Θ, т. е.

$$\gamma = P\{\widetilde{\Theta}_n - \delta < \Theta < \widetilde{\Theta}_n + \delta\} = P\{|\widetilde{\Theta}_n - \Theta| < \delta\}.$$

Заметим, что после того, как по данным выборки вычислена оценка $\widetilde{\Theta}_n$, событие $\{|\widetilde{\Theta}_n - \Theta| < \delta\}$ становится или достоверным, или невозможным, так как интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$ или покрывает Θ, или нет. Но дело в том, что параметр Θ нам неизвестен. Поэтому мы называем надежностью γ уже вычисленной оценки $\widetilde{\Theta}_n$ вероятность того, что интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$, найденный для произвольной выборки, покроет Θ. Если мы сделаем много выборок объема n и для каждой из них построим интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$, то доля тех выборок, чьи интервалы покроют Θ, равна γ.

Иными словами, γ есть мера нашего доверия вычисленной оценке $\widetilde{\Theta}_n$.

Ясно, что, чем меньше число δ, тем меньше надежность γ.

Определение. *Доверительным интервалом* называется найденный по данным выборки интервал $(\widetilde{\Theta}_n - \delta,\ \widetilde{\Theta}_n + \delta)$, который покрывает параметр Θ с заданной надежностью γ.

Надежность γ обычно принимают равной 0,95 или 0,99, или 0,999.

Конечно, нельзя категорически утверждать, что найденный доверительный интервал покрывает параметр Θ. Но в этом можно быть уверенным на 95% при $\gamma = 0{,}95$, на 99% при $\gamma = 0{,}99$ и т. д. Это значит, что если сделать много выборок, то для 95% из них (если, например, $\gamma = 0{,}95$) вычисленные доверительные интервалы действительно покроют Θ.

2. Доверительный интервал для математического ожидания при известном σ. В некоторых случаях среднее квадратическое отклонение σ ошибки измерения (а вместе с нею и самого измерения) бывает известно. Например, если измерения осуществляются одним и тем же прибором при одних и тех же условиях.

Итак, пусть случайная величина X распределена нормально с параметрами a и σ, причем σ известно. Построим доверительный интервал, покрывающий неизвестный параметр a с заданной надежностью γ. Данные выборки есть реализации случайных величин $X_1, X_2, \ldots, X_n$, имеющих нормальное распределение с параметрами a и σ (§ 4.2, п. 1). Оказывается, что и выборочная средняя случайная величина $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ тоже имеет нормальное распределение (это мы примем без доказательства). При этом (см. § 4.2, пп. 2, 3)

$$M(\overline{X}) = a;\ \ \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Потребуем, чтобы выполнялось соотношение $P(|\overline{X} - a| < \delta) = \gamma$, где γ — заданная надежность. Пользуясь формулой (2.27) (§ 2.7, п. 2), получим

$$P(|\overline{X} - a| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right),$$

или

$$P(|\overline{X} - a| < \delta) = 2\Phi(t),$$

где

$$t = \frac{\delta\sqrt{n}}{\sigma}. \tag{4.15}$$

Найдя из равенства (4.15) $\delta = \frac{t\sigma}{\sqrt{n}}$, можем написать

$$P\left(|\overline{X} - a| < \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t).$$

Так как P задана и равна γ, то окончательно имеем (для получения рабочей формулы выборочную среднюю заменяем на $\bar{x}_в$):

$$P\left(\bar{x}_в - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_в + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma.$$

Смысл полученного соотношения таков: *с надежностью γ можно утверждать, что доверительный интервал* $\left(\bar{x}_в - \frac{t\sigma}{\sqrt{n}};\ \bar{x}_в + \frac{t\sigma}{\sqrt{n}}\right)$ *покрывает неизвестный параметр a; точность оценки* $\delta = \frac{t\sigma}{\sqrt{n}}$. Здесь число t определяется из равенства $\Phi(t) = \frac{\gamma}{2}$ [оно следует из $2\Phi(t) = \gamma$] по таблице приложения 3.

Как уже упоминалось, надежность γ обычно принимают равной или 0,95 или 0,99, или 0,999.

Пример. Признак X распределен в генеральной совокупности нормально с известным $\sigma = 0{,}40$. Найдем по данным выборки доверительный интервал для a с надежностью $\gamma = 0{,}99$, если $n = 20$, $\bar{x}_в = 6{,}34$.

Для $\Phi(t) = \frac{\gamma}{2} = \frac{0{,}99}{2} = 0{,}495$ находим по таблице приложения 3 $t = 2{,}58$. Следовательно, $\delta = 2{,}58 \cdot \frac{0{,}40}{\sqrt{20}} \approx 0{,}23$. Границы доверительного интервала $6{,}34 - 0{,}23 = 6{,}11$ и $6{,}34 + 0{,}23 = 6{,}57$. Итак, доверительный интервал $(6{,}11;\ 6{,}57)$ покрывает a с надежностью 0,99.

3. Доверительный интервал для математического ожидания при неизвестном σ. Пусть случайная величина X имеет нормальное распределение с неизвестными нам параметрами a и σ. Оказывается, что случайная величина (ее возможные значения будем обозначать через t)

$$T = \frac{\bar{X} - a}{S/\sqrt{n}},$$

где n — объем выборки; $\bar{X}$ — выборочная средняя; S — исправленное среднее квадратическое отклонение, имеет распределение, не зависящее от a и σ. Оно называется распределением Стьюдента*.

Плотность вероятности распределения Стьюдента дается формулой

$$S(t, n) = B_n\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}},$$

где коэффициент B_n зависит от объема выборки.

---

* Стьюдент — псевдоним английского статистика И. О. Госсета.

Потребуем, чтобы выполнялось соотношение

$$P(|T| < t_\gamma) = \gamma,$$

где γ — заданная надежность.

Так как $S(t, n)$ — четная функция от t, то, пользуясь формулой (2.15) (см. § 2.5), получим

$$P\left(\frac{|\overline{X} - a|\sqrt{n}}{S} < t_\gamma\right) = 2\int_0^{t_\gamma} S(t, n)\,dt = \gamma.$$

Отсюда

$$P\left(\overline{X} - \frac{t_\gamma S}{\sqrt{n}} < a < \overline{X} + \frac{t_\gamma S}{\sqrt{n}}\right) = \gamma.$$

Следовательно, приходим к утверждению: *с надежностью γ можно утверждать, что доверительный интервал* $\left(\bar{x}_{\text{в}} - \frac{t_\gamma s}{\sqrt{n}};\ \bar{x}_{\text{в}} + \frac{t_\gamma s}{\sqrt{n}}\right)$ *покрывает неизвестный параметр a, точность оценки* $\delta = \frac{t_\gamma s}{\sqrt{n}}$. Здесь случайные величины $\overline{X}$ и S заменены неслучайными величинами $\bar{x}_{\text{в}}$ и s, найденными по выборке.

В приложении 4 приведена таблица значений $t_\gamma = t(\gamma, n)$ для различных значений n и обычно задаваемых значений надежности.

Заметим, что при $n \geqslant 30$ распределение Стьюдента практически не отличается от нормированного нормального распределения (см. § 2.7, п. 2). Это связано с тем, что $\lim\limits_{n \to \infty}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} = e^{-\frac{t^2}{2}}$.

Пример. Признак X распределен в генеральной совокупности нормально. Найдем доверительный интервал для $\bar{x}_{\text{г}}$ с надежностью $\gamma = 0{,}99$, если $n = 20$; $\bar{x}_{\text{в}} = 6{,}34$; $s = 0{,}40$. Для надежности $\gamma = 0{,}99$ и $n = 20$ находим по таблице приложения 4 $t_\gamma = 2{,}861$. Следовательно, $\delta = 2{,}861 \cdot \frac{0{,}40}{\sqrt{20}} \approx 0{,}26$. Концы доверительного интервала $6{,}34 - 0{,}26 = 6{,}08$ и $6{,}34 + 0{,}26 = 6{,}60$. Итак, доверительный интервал $(6{,}08;\ 6{,}60)$ покрывает $\bar{x}_{\text{г}}$ с надежностью $0{,}99$.

4. Доверительный интервал для среднего квадратического отклонения. Для нахождения доверительного интервала для среднего квадратического отклонения σ будем использовать следующее предложение, устанавливаемое аналогично двум предыдущим (пп. 2 и 3).

С надежностью γ можно утверждать, что *доверительный интервал $(s - sq;\ s + sq)$ покрывает неизвестный параметр σ; точность оценки $\delta = sq$.*

В приложении 5 приведена таблица значений $q = q(\gamma, n)$ для различных значений n и обычно задаваемых значений надежности γ.

Пример 1. Признак X распределен в генеральной совокупности нормально. Найдем доверительный интервал для $\sigma_г$ с надежностью $\gamma=0{,}95$, если $n=20$, $s=0{,}40$.

Для надежности $\gamma=0{,}95$ и $n=20$ находим в таблице приложения 5 $q=0{,}37$. Далее, $sq=0{,}40 \quad 0{,}37\approx 0{,}15$. Границы доверительного интервала $0{,}40-0{,}15=0{,}25$ и $0{,}40+0{,}15=0{,}55$. Итак, доверительный интервал $(0{,}25;\ 0{,}55)$ покрывает $\sigma_г$ с надежностью 0,95.

Пример 2. На ферме испытывалось влияние витаминов на прибавку в массе телят. С этой целью было осмотрено 20 телят одного возраста. Средняя масса их оказалась равной 340 кг, а «исправленное» среднее квадратическое отклонение — 20 кг.

Определим: 1) доверительный интервал для математического ожидания a с надежностью 0,95; 2) доверительный интервал для среднего квадратического отклонения с той же надежностью.

При решении задачи будем исходить из предположения, что данные пробы взяты из нормальной генеральной совокупности.

Решение. 1) Согласно условиям задачи, $x_в=340$; $s=20$; $\gamma=0{,}95$; $n=20$.

Пользуясь распределением Стьюдента, для надежности $\gamma=0{,}95$ и $n=20$ находим в таблице приложения 4 $t_\gamma=2{,}093$. Следовательно, $\delta=2{,}093\ \frac{20}{\sqrt{20}}=9{,}4$. Границы доверительного интервала $340-9{,}4=$ $=330{,}6$ и $340+9{,}4=349{,}4$. Итак, доверительный интервал $(330{,}6;\ 349{,}4)$ покрывает a с надежностью 0,95.

Можно считать, что в данном случае истинная масса измерена достаточно точно (отклонение порядка $\frac{9{,}4}{340}\approx 0{,}03$).

2) Для надежности $\gamma=0{,}95$ и $n=20$ находим в таблице приложения 5 $q=0{,}37$. Далее, $sq=20\cdot 0{,}37\approx 7{,}4$. Границы доверительного интервала $20-7{,}4=12{,}6$ и $20+7{,}4=27{,}4$. Таким образом, $12{,}6<\sigma<27{,}4$, откуда можно заключить, что σ определено неудовлетворительно (отклонение порядка $\frac{sq}{s}=q\approx 0{,}4$ — почти половина!). Чтобы сузить доверительный интервал при той же надежности, необходимо увеличить число проб n.

Примечание. Выше предполагалось, что $q<1$. Если $q>1$, то, учитывая, что $\sigma>0$, получаем $0<\sigma<s+sq$. Значения q и в этом случае определяются по таблице приложения 5.

Пример 3. Признак X генеральной совокупности распределен нормально. По выборке объема $n=10$ найдено «исправленное» среднее квадратическое отклонение $s=0{,}16$. Найдем доверительный интервал для $\sigma_г$ с надежностью 0,999.

Для надежности $\gamma=0{,}999$ и $n=10$ по таблице приложения 5 находим $q=1{,}80$.

Следовательно, искомый доверительный интервал таков·

$$0<\sigma<0{,}16+0{,}16\ \ 1{,}80,$$

или

$$0<\sigma<0{,}448.$$

5. Оценка истинного значения измеряемой величины. Пусть проводится n независимых равноточных измерений* некоторой физической величины, истинное значение a которой неизвестно. Будем рассматривать результаты отдельных измерений как случайные величины X_1, X_2, ..., X_n. Эти величины независимы (измерения независимы), имеют одно и то же математическое ожидание a (истинное значение измеряемой величины), одинаковые дисперсии σ^2 (измерения равноточны) и распределены нормально (такое допущение подтверждается опытом). Таким образом, все предположения, которые были сделаны при выводе доверительных интервалов в пп. 2 и 3 настоящего параграфа, выполняются, следовательно, мы вправе использовать полученные в них предложения. Так как обычно σ неизвестно, следует пользоваться предложением, найденным в п. 3 данного параграфа.

Пример. По данным девяти независимых равноточных измерений физической величины найдены среднее арифметическое результатов отдельных измерений $\bar{x}_в=42{,}319$ и «исправленное» среднее квадратическое отклонение $s=5{,}0$. Требуется оценить истинное значение a измеряемой величины с надежностью $\gamma=0{,}99$.

Истинное значение измеряемой величины равно ее математическому ожиданию. Поэтому задача сводится к оценке математического ожидания (при неизвестном σ) при помощи доверительного интервала

$$\bar{x}_в-\frac{t_\gamma s}{\sqrt{n}}<a<\bar{x}_в+\frac{t_\gamma s}{\sqrt{n}},$$

покрывающего a с заданной надежностью $\gamma=0{,}99$.

Пользуясь таблицей приложения 4 по $\gamma=0{,}99$ и $n=9$, находим $t_\gamma=3{,}36$.

Найдем точность оценки:

$$\delta=\frac{t_\gamma s}{\sqrt{n}}=3{,}36\ \ \frac{5}{\sqrt{9}}=3{,}36\ \ \frac{5}{3}=5{,}60.$$

Границы доверительного интервала

$$42{,}319-5{,}60=36{,}719$$

и

$$42{,}319+5{,}60=47{,}919.$$

Итак, с надежностью $\gamma=0{,}99$ истинное значение измеренной величины a заключено в доверительном интервале $36{,}719<a<47{,}919$.

---

* То есть измерений, проводимых в одинаковых условиях Эти условия считают выполненными, если измерения проводятся одним прибором

6. Оценка точности измерений. В теории ошибок принято точность измерений (точность прибора) характеризовать с помощью среднего квадратического отклонения σ случайных ошибок измерений. Для оценки σ используют «исправленное» среднее квадратическое отклонение s. Поскольку обычно результаты измерений независимы, имеют одно и то же математическое ожидание (истинное значение измеряемой величины) и одинаковую дисперсию (в случае равноточных измерений), то утверждение, приведенное в п. 4, применимо для оценки точности измерений.

Пример. По 16 независимым равноточным измерениям найдено «исправленное» среднее квадратическое отклонение $s=0{,}4$. Найдем точность измерений с надежностью $\gamma=0{,}99$.

Как отмечено выше, точность измерений характеризуется средним квадратическим отклонением σ случайных ошибок измерений. Поэтому задача сводится к отысканию доверительного интервала $(s-sq;\ s+sq)$, покрывающего σ с заданной надежностью $\gamma=0{,}99$ (см. п. 4). По таблице приложения 5 по $\gamma=0{,}99$ и $n=16$ найдем $q=0{,}70$. Следовательно, искомый доверительный интервал таков:

$$0{,}4(1-0{,}70)<\sigma<0{,}4(1+0{,}70),$$

или

$$0{,}12<\sigma<0{,}68.$$

§ 4.4. Проверка статистических гипотез

Пусть по выборке объема n получено эмпирическое распределение с равноотстоящими вариантами:

Варианты	x_1	x_2	...	x_m
Эмпирические (наблюдаемые) частоты	n_1	n_2	...	n_m

По данным наблюдения выдвигают гипотезу о законе распределения генеральной совокупности, например, предполагают, что генеральная совокупность распределена равномерно или нормально. Такие гипотезы называются *статистическими*. Затем для тех же объектов, которые попали в выборку, вычисляют частоты, уже исходя из теоретической гипотезы. В результате получаются частоты (их называют *выравнивающими частотами*), которые, вообще говоря, отличаются от наблюдавшихся. Как определить, правильно или нет выдвинута гипотеза, т. е. случайны ли расхождения наблюдавшихся и выравнивающих частот или эти расхождения являются следствием неправильности гипотезы? Для решения этого вопроса применяют критерии согласия эмпирических наблюдений к выдвинутой гипотезе. Имеется несколько критериев согласия: χ^2 («хи-квадрат»)

Пирсона, критерий Колмогорова, критерий Смирнова и др. Мы познакомимся с критерием согласия χ^2 («хи-квадрат») Пирсона.

Предположим, что на основе приведенного выше распределения выдвинута гипотеза H: генеральная совокупность имеет нормальное распределение. Для вычисления выравнивающих частот поступают следующим образом:

1) находят значения $\bar{x}_в$, $\sigma_в = \sqrt{D_в}$;

2) выравнивающие частоты n'_i ищут по формуле

$$n'_i = \frac{nh}{\sigma_в}\varphi(u_i),$$

где n — сумма наблюдавшихся частот; h — разность между двумя соседними вариантами; $u_i = \frac{x_i - \bar{x}_в}{\sigma_в}$ и $\varphi(u) = \frac{1}{\sqrt{2\pi}}e^{\frac{-u^2}{2}}$.

В результате получают множество выравнивающих частот:

$$n'_1,\ n'_2,\ \ldots,n'_m.$$

Обозначим через χ^2 сумму квадратов разностей между эмпирическими и выравнивающими частотами, деленных на соответствующие выравнивающие частоты:

$$\chi^2 = \sum_{i=1}^{m}\frac{(n_i - n'_i)^2}{n'_i}\ * \qquad (4.16)$$

(это обозначение и для распределения χ^2).

Для данной выборки по формуле (4.16) находим значение случайной величины χ^2. Обозначим его через χ^2_0. Затем определяется число $k = m - 3$, называемое числом степеней свободы, где m — число различных вариант выборки.

Теперь проверка гипотезы H проводится так. Задаются достаточно малой вероятностью p, называемой *уровнем значимости* (обычно в качестве p берут либо 0,05, либо 0,01, либо 0,001). Считается, что событие с такой вероятностью является практически невозможным. По таблице значений χ^2 (приложение 6, здесь речь идет о так называемых критических точках распределения χ^2) по заданному уровню значимости p и числу степеней свободы k находят значение $\chi^2(p;\ k)$. Если окажется, что $\chi^2_0 > \chi^2\ (p;\ k)$, то гипотеза H отвергается на уровне значимости p, так как произошло событие, которое не должно было произойти при верной гипотезе H; если же $\chi^2_0 < \chi^2$ $(p;\ k)$, то H принимается на уровне значимости p.

Пример. При уровне значимости 0,05 проверим гипотезу о нормальном распределении генеральной совокупности, если известны:

эмпирические частоты... 6 13 38 74 106 85 30 14

теоретические частоты... 3 14 42 82 99 76 37 13

---

* Из этой формулы видно, что чем меньше различие между эмпирическими и выравнивающими частотами, тем меньше будет χ^2.

Вычислим χ_0^2, для чего составим расчетную таблицу:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	6	3	3	9	3
2	13	14	–1	1	0,07
3	38	42	–4	16	0,38
4	74	82	–8	64	0,78
5	106	99	7	49	0,49
6	85	76	9	81	1,07
7	30	37	–7	49	1,32
8	14	13	1	1	0,08
					$\chi_0^2 = 7{,}19$

Найдем число степеней свободы, учитывая, что число различных вариант $m = 8$. Имеем: $k = 8 - 3 = 5$. По уровню значимости $p = 0{,}05$ и числу степеней свободы $k = 5$ по таблице значений χ^2 (приложение 6) находим: $\chi^2(0{,}05;\ 5) = 11{,}1$. Так как $\chi_0^2 < \chi^2\ (0{,}05;\ 5)$, нет оснований отвергнуть гипотезу H.

§ 4.5. Расчет прямых регрессии

Пусть проведено n опытов, в результате которых получены следующие значения величин (X, Y): (x_i, y_i), $i = 1, 2, \ldots, n$. За приближенные значения $M(X)$, $M(Y)$, $D(X)$ и $D(Y)$ принимают их выборочные значения:

$$\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i;\quad \bar{y}_в = \frac{1}{n}\sum_{i=1}^{n} y_i;\quad s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2;$$

$$s_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_в)^2.$$

Оценкой для μ служит величина

$$\mu_в = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)(y_i - \bar{y}_в).$$

Заменяя в соотношениях (3.15), (3.17), (3.20) величины μ_{xy}, σ_x, σ_y их выборочными значениями $\mu_в$, s_1, s_2, получим приближенные значения коэффициента корреляции и коэффициентов регрессии:

$$r_в \simeq \frac{\mu_в}{s_1 s_2};\ \rho(Y/X) = \frac{\mu_в}{s_1^2};\ \rho(Y/X) \approx \frac{\mu_в}{s_2^2}.$$

Подставляя в уравнения (3.18) и (3.19) вместо a, b, $\rho(Y/X)$ и $\rho(X/Y)$ их приближенные значения, получим выборочные уравнения прямых регрессий:

$$y - \bar{y}_{\text{в}} = \frac{\mu_{\text{в}}}{s_1^2}(x - \bar{x}_{\text{в}});$$

$$x - \bar{x}_{\text{в}} = \frac{\mu_{\text{в}}}{s_2^2}(y - \bar{y}_{\text{в}}).$$

Пример. Найдем выборочное уравнение прямой регрессии Y на X по данным $n = 10$ наблюдений. Результаты наблюдений и результаты вычислений собраны в таблице ($C = 70$ и $C' = 9{,}0$ — ложные нули).

x_i	y_i	$x_i - C$	$x_i - C'$	$x_i - \bar{x}_{\text{в}}$	$(x_i - \bar{x}_{\text{в}})^2$	$y_i - \bar{y}_{\text{в}}$	$(x_i - \bar{x}_{\text{в}}) \times$ $\times (y_i - \bar{y}_{\text{в}})$
71	8,6	1	–0,4	–4,5	20,25	–0,48	2,16
72	8,9	2	–0,1	–3,5	12,25	–0,18	0,63
73	8,9	3	–0,1	–2,5	6,25	–0,18	0,45
74	9,0	4	0,0	–1,5	2,25	0,08	–0,12
75	9,1	5	0,1	–0,5	0,25	0,02	–0,01
76	9,2	6	0,2	0,5	0,25	0,12	0,06
77	9,2	7	0,2	1,5	2,25	0,12	0,18
78	9,2	8	0,2	2,5	6,25	0,12	0,30
79	9,3	9	0,3	3,5	12,25	0,22	0,77
80	9,4	10	0,4	4,5	20,25	0,32	1,44
Сумма		55	0,8		82,5		5,86
$\bar{x}_{\text{в}} = 75{,}5$	$\bar{y}_{\text{в}} = 9{,}08$				$s_1^2 = 9{,}17$		$\mu_{\text{в}} = 0{,}65$

Вычисляем:

$$\rho(Y/X) = \frac{\mu_{\text{в}}}{s_1^2} = \frac{0{,}65}{9{,}17} \approx 0{,}071.$$

Уравнение искомой прямой имеет вид

$$y - 9{,}08 = 0{,}071(x - 75{,}5),$$

или

$$y = 0{,}071x + 3{,}72.$$

Упражнения

1. Постройте полигон следующего распределения:

x_i	1	3	5	7	9
n_i	10	15	30	33	12

2. Постройте полигон распределения по размеру проданной мужской обуви по данным следующей таблицы:

Размер обуви (варианта)	36	37	38	39	40	41	42	43
Число проданных пар (частота)	1	1	5	8	17	21	18	8

3. Постройте гистограмму следующего распределения:

Частичный интервал длиной h	Сумма частот вариант частичного интервала n_i
2—5	9
5—8	10
8—11	25
11—14	6

4. Постройте гистограмму распределения коров по проценту жирности молока по данным следующей таблицы:

Жирность молока, %	Число коров
3,45—3,55	1
3,55—3,65	1
3,65—3,75	3
3,75—3,85	4
3,85—3,95	7
3,95—4,05	5
4,05—4,15	2
4,15—4,25	1
4,25—4,35	1

5. Генеральная совокупность задана таблицей распределения:

x_i	1000	1200	1400
N_i	1000	6000	3000

Найдите генеральную среднюю $\bar{x}_г$ и генеральную дисперсию $D_г$.

[$\bar{x}_г = 1240$; $D_г = 14\,400$]

6. Найдите выборочную среднюю по данным следующей таблицы:

Длина хоботка у шести пчел (в мм)	
6,54	6,69
6,71	6,70
6,70	6,62

[6,66 мм]

7. Выборочная совокупность задана таблицей распределения:

x_i	4	7	10	15
n_i	10	15	20	5

Найдите выборочные среднюю $\bar{x}_в$ и дисперсию $D_в$.

[$\bar{x}_в = 8{,}4$; $D_в = 9{,}84$]

8. Выборочная совокупность задана таблицей распределения:

x_i	1	2	5	8	9
n_i	3	4	6	4	3

Найдите выборочную и исправленную дисперсии.

[$D_в = 8{,}4$; $s^2 = 8{,}84$]

9. Выборочным путем были получены следующие данные об урожайности ржи в районе:

Урожайность, ц/га	Число гектаров
18	10
20	20
21	20

Определить выборочную среднюю $\bar{x}_в$ и исправленное среднее квадратическое отклонение s. [$\bar{x}_в = 20$ ц; $s \approx 1{,}1$ ц]

10. По выборке объема $n = 51$ найдена выборочная дисперсия $D_в = 5$. Найдите исправленную дисперсию. [$s^2 = 5{,}1$]

11. В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8; 9; 11; 12. Найдите: а) выборочную среднюю результатов измерений; б) выборочную и исправленную дисперсии ошибок прибора.

[а) $\bar{x}_в = 10$; б) $D_в = 2{,}5$; $s^2 = {}^{10}/_3$]

12. Найдите выборочную среднюю по данному распределению выборки объема $n = 100$:

x_i	340	360	375	380
n_i	20	50	18	12

[361,1]

13. Найдите выборочную дисперсию по данному распределению выборки объема $n = 10$:

x_i	0,1	0,4	0,6
n_i	3	2	5

[0,0469]

14—16. Заданы среднее квадратическое отклонение, выборочная средняя и объем выборки нормально распределенного признака генеральной совокупности. Найдите границы доверительных интервалов для оценки генеральной средней $\bar{x}_г$ с заданной надежностью.

14. $\sigma=3$; $\bar{x}_в=4,1$; $n=36$; $\gamma=0,95$. [$3,12<x_г<5,08$]

15. $\sigma=2$; $\bar{x}_в=5,4$; $n=10$; $\gamma=0,95$. [$4,16<\bar{x}_г<6,64$]

16. $\sigma=3$; $\bar{x}_в=20,12$; $n=25$; $\gamma=0,99$. [$18,57<\bar{x}_г<21,67$]

17—19. Заданы «исправленное» среднее квадратическое отклонение, выборочная средняя и объем выборки нормально распределенного признака генеральной совокупности. Найдите, пользуясь распределением Стьюдента, доверительные интервалы для оценки генеральной средней $\bar{x}_г$ с заданной надежностью.

17. $s=0,8$; $\bar{x}_в=20,2$; $n=16$; $\gamma=0,95$. [$19,774<\bar{x}_г<20,626$]

18. $s=1,5$; $\bar{x}_в=16,8$; $n=12$; $\gamma=0,95$. [$15,85<\bar{x}_г<17,75$]

19. $s=2,4$; $\bar{x}_в=14,2$; $n=9$; $\gamma=0,99$. [$11,512<\bar{x}_г<16,888$]

20. Признак X генеральной совокупности распределен нормально. По выборке объема $n=25$ найдено «исправленное» среднее квадратическое отклонение $s=0,8$. Найдите доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95.

[$0,544<\sigma<1,056$]

21. По данным девяти независимых равноточных измерений физической величины найдены среднее арифметическое результатов отдельных измерений $\bar{x}_в=42,319$ и «исправленное» среднее квадратическое отклонение $s=5$. Требуется оценить истинное значение a измеряемой величины с надежностью 0,95. [$38,469<a<46,169$]

22. По 15 равноточным измерениям найдено «исправленное» среднее квадратическое отклонение $s=0,12$. Найдите точность измерений σ с надежностью 0,99. [$0,03<\sigma<0,21$]

23. По данным 16 независимых равноточных измерений физической величины найдены среднее арифметическое результатов отдельных измерений $\bar{x}_в=23,161$ и «исправленное» среднее квадратическое отклонение $s=0,4$. Требуется оценить истинное значение a измеряемой величины и точность измерений σ с надежностью 0,95.

$$\left[\begin{array}{c}22,948<a<23,374;\\ 0,224<\sigma<0,576\end{array}\right]$$

24—26. При уровне значимости 0,05 проверьте гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

24.

Эмпирические частоты	6	12	16	40	13	8	5
Теоретические частоты	4	11	15	43	15	6	6

[$\chi^2_0=2,5$; $\chi^2(0,05;\ 4)=9,5$. Нет оснований отвергнуть гипотезу]

25.

Эмпирические частоты	5	6	14	32	43	39	30	20	6	5
Теоретические частоты	4	7	12	29	48	35	34	18	7	6

[$\chi_0^2 \approx 3$; $\chi^2(0{,}05;\ 7) = 14{,}07$. Нет оснований отвергнуть гипотезу]

26.

Эмпирические частоты	5	13	12	44	8	12	6
Теоретические частоты	2	20	12	35	15	10	6

[$\chi_0^2 \approx 13$; $\chi^2(0{,}05;\ 4) = 9{,}5$. Гипотеза отвергается]

27. Найдите выборочное уравнение прямой регрессии Y на X по данным $n = 5$ наблюдений:

x_i	1,00	1,50	3,00	4,50	5,00
y_i	1,25	1,40	1,50	1,75	2,25

[$y \approx 0{,}202x + 1{,}024$]

28. Вычислите выборочный коэффициент корреляции по данным следующей таблицы:

x_i	92	91	90	86	85	85	85	83	80	78	80	83
y_i	84	85	84	81	76	77	75	79	78	78	76	75

[$r_в \approx 0{,}792$]

ЗАКЛЮЧЕНИЕ

Теория вероятностей возникла в середине XVII столетия в связи с подсчетом различных вероятностей, связанных с азартными играми в карты и кости. Первую такую задачу пытался решить еще в 1494 г. итальянский математик Лука Пачоли (1450—1520), но по-настоящему первые решения теоретико-вероятностных задач принадлежат французским математикам Блезу Паскалю (1623—1662), Пьеру Ферма (1601—1665) и голландскому математику Христиану Гюйгенсу (1629—1695). Именно к этому времени относится возникновение классического определения вероятности.

Основы теории игр изложены Гюйгенсом в сочинении «О расчетах в азартных играх», вышедшем в 1657 г. Остановимся, в частности, на задачах о справедливом разделении ставки, решения которых можно найти в переписке ученых Паскаля и Ферма, в которой, как обычно считают, и зародилась теория вероятностей.

1. Как разделить ставку при игре до трех выигрышных партий, если один игрок выиграл две партии, а другой — одну, и каждым вложено в игру по 32 пистоля?

Свое решение задачи Паскаль наиболее полно изложил в письме к Ферма от 29 июля 1654 г.: «Предположим, что один выиграл две партии, а другой — одну. Они играют еще одну партию, если ее выигрывает первый, то он получает всю сумму в 64 пистоля...; если же эту партию выигрывает второй, то каждый игрок будет иметь 2 выигранные партии, и, следовательно, если они намерены произвести раздел, каждый должен получить обратно свой вклад в 32 пистоля... Если же игроки не намерены рисковать... и хотят произвести раздел, то первый должен сказать: «Я имею 32 пистоля верных, ибо в случае проигрыша я их также получил бы, но остальные 32 пистоля могут быть получены либо мной, либо вами. Случайности равны. Разделим же эти 32 пистоля пополам, и дайте мне, кроме того, бесспорную сумму в 32 пистоля». Как видно из рассуждений Паскаля, первый игрок должен получить 48 пистолей, а второй — 16.

2. Как разделить ставку при игре до трех выигрышных партий, если один игрок выиграл две партии, а другой — ни одной, и каждым вложено в игру по 32 пистоля?

Ответы, предложенные Паскалем, таковы: первый игрок должен получить 56 пистолей, а второй — 8 пистолей. Рассуждения при решении подобны тем, которые были проведены при решении предыдущей задачи: если бы первый игрок выиграл еще одну партию, то ему причиталось бы 64 пистоля, если бы проиграл — 48 пистолей, остаток 16 делится поровну.

3. Как разделить ставку при игре до трех выигрышных партий, если первый игрок выиграл одну партию, а второй — ни одной, и каждым вложено в игру по 32 пистоля?

Пусть игроки сыграют еще одну партию. Если ее выигрывает первый, то он будет иметь, как и в предыдущем случае, 56 пистолей. Если он ее проигрывает, то у обоих окажется по одной выигранной партии, и первому следует получить 32 пистоля. Первый игрок может сказать: «Если вы не хотите играть эту партию, дайте мне мой бесспорный выигрыш в 32 пистоля, а остаток от 56 разделим поровну, т. е. возьмем каждый по 12, что с 32 составит 44». Значит, первый игрок должен получить 44 пистоля, а второй — 20.

В книге Якоба Бернулли (1713 г.) «Искусство предположений» были установлены все основные свойства вероятностей, рассмотрена схема независимых испытаний и выведена соответствующая формула. Кроме того, здесь доказана теорема о связи между вероятностью и частотой наступления события, которую сейчас называют теоремой Бернулли, или законом больших чисел в форме Бернулли. Это была первая из теорем этого типа, играющих сейчас большую роль в теории вероятностей.

Следующий период истории теории вероятностей — XVIII в. и начало XIX в.— связан главным образом с именами французских математиков А. Муавра, П. Лапласа, С. Пуассона и А. Лежандра и немецкого математика К. Гаусса. В это время в теории вероятностей, кроме понятия случайного события, рассматривается и понятие случайной величины. Теория вероятностей начала применяться уже в ряде научных областей — теории ошибок измерений, теории стрельбы и т. п.

В третьем периоде развития теории вероятностей, который относится ко второй половине XIX столетия, важнейшую роль играли работы русских ученых П. Л. Чебышева, А. М. Ляпунова и А. А. Маркова. В это время был доказан целый ряд предельных теорем и различных форм закона больших чисел. Марков рассмотрел одно из первых обобщений схемы Бернулли на случай зависимых испытаний, получивших название *цепей Маркова.*

Современный период истории теории вероятностей характеризуется возникновением и развитием многих новых областей и направлений, например теории информации и теории игр. В этот же период получает развитие и математическая статистика (возникновение ее относится к XVII в.). Круг применения теории вероятностей в различных областях науки и техники расширился настолько (теоретическая физика, радиоэлектроника, теория автоматического регулирования, экономика, биология, медицина и т. д.), что сейчас ее по праву можно считать одной из наиболее прикладных частей математики.

Большое значение для современной теории вероятностей имеют работы представителей советской школы, в частности А. Н. Колмогорова (1903—1987), С. Н. Бернштейна (1880—1968), А. Я. Хинчина (1894—1959), Ю. В. Линника (1915—1972) и Б. В. Гнеденко.

Дополнительные упражнения

К главе I

1. Найдите вероятность выпадения цифры при одном бросании монеты. [0,5]

2. Брошена игральная кость. Найдите вероятность того, что выпадет нечетное число очков. [0,5]

3. В шар вписан куб. Точка ставится наугад в шар. Какова вероятность того, что она попадет в куб?

$$\left[\frac{2\sqrt{3}}{3\pi}\right]$$

4. В книге 500 страниц. Чему равна вероятность того, что наугад открытая страница имеет порядковый номер, кратный 7? [0,142]

5. В урне 3 синих, 8 красных и 9 белых шаров. Какова вероятность появления: а) синего; б) красного; в) белого шара при одном вынимании шара из урны? [а) 0,15; б) 0,4; в) 0,45]

6. Студент из 30 экзаменационных билетов усвоил 24. Какова вероятность его успешного ответа на экзамене на билет при одном кратном извлечении билета? [0,8]

7. В сосуд вместимостью 10 л попала одна болезнетворная бактерия. Какова вероятность зачерпнуть ее при наборе из этого сосуда стакана воды (200 см^3)? [0,02]

8. Отдел технического контроля обнаружил 3 бракованные книги в партии из случайно отобранных 100 книг. Найдите относительную частоту появления бракованных книг. [0,03]

9. По цели произведено 20 выстрелов, причем зарегистрировано 18 попаданий. Найдите относительную частоту попаданий в цель. [0,9]

10. Среди 1000 новорожденных оказалось 517 мальчиков. Найдите относительную частоту рождения мальчиков. [0,517]

11. В День физкультурника Сизов пошел на стадион. Можно было купить билет на футбол с вероятностью 0,3 или купить билет на волейбол с вероятностью 0,2. Какова вероятность того, что Сизов попал на соревнования? [0,5]

12. Стрелок стреляет по мишени, разделенной на 3 области. Вероятность попасть в первую область равна 0,45; во вторую — 0,35. Найдите вероятность того, что стрелок при одном выстреле попадет либо в первую, либо во вторую область. [0,8]

13. Какова вероятность выпадения 4 или 6 очков при однократном бросании игральной кости?

$$\left[\frac{1}{3}\right]$$

14. В урне 2 синих, 6 красных и 12 белых шаров. Какова вероятность вынуть цветной шар, если вынимается один шар? [0,4]

15. Вероятность того, что стрелок при одном выстреле выбьет 10 очков, равна 0,1, вероятность выбить 9 очков равна 0,3; вероятность выбить 8 или меньше очков равна 0,6. Найдите вероятность того, что при одном выстреле стрелок выбьет не менее 9 очков. [0,4]

16. В урне 4 белых и 3 синих шара. Каждое испытание состоит в том, что наудачу извлекают один шар, не возвращая его обратно. Найдите вероятность появления синего шара при втором испытании, если при первом испытании был извлечен белый шар. [0,5]

17. В читальном зале имеется 6 учебников по теории вероятностей, из которых 3 в переплете. Библиотекарь наудачу взял 2 учебника. Найдите вероятность того, что оба учебника окажутся в переплете. [0,2]

18. В цехе работают 7 мужчин и 3 женщины. По табельным номерам наудачу отобрали двух человек. Найдите вероятность того, что все отобранные лица окажутся мужчинами.

$$\left[\frac{7}{15}\right]$$

19. Студент знает 20 из 25 вопросов программы. Найдите вероятность того, что студент знает предложенные ему экзаменатором два вопроса.

$$\left[\frac{19}{30}\right]$$

20. В мешочке содержится 10 одинаковых кубиков с номерами от 1 до 10. Наудачу извлекают по одному 2 кубика. Найдите веростность того, что последовательно появятся кубики с номерами 1, 2, если кубики извлекаются без возвращения.

$$\left[\frac{1}{90}\right]$$

21. Для некоторой местности среднее число пасмурных дней в июле равно шести. Найдите вероятность того, что первого и второго июля будет ясная погода.

$$\left[\frac{20}{31}\right]$$

22. Среди 50 электрических лампочек три нестандартные. Найдите вероятность того, что две взятые подряд лампочки окажутся нестандартными.

$$\left[\frac{3}{1225}\right]$$

23. Работница обслуживает два станка, работающие независимо друг от друга. Вероятность того, что в течение часа станок не потребует внимания работницы, для первого станка равна 0,9, для второго — 0,8. Найдите вероятность того, что в течение часа ни один из двух станков не потребует внимания работницы. [0,72]

24. Брошены две игральные кости. Какова вероятность того, что на обеих костях выпадут шестерки?

$\left[\frac{1}{36}\right]$

25. Экзаменационный билет содержит три вопроса. Вероятности того, что студент ответит на первый и второй вопросы билета, равны по 0,9, на третий — 0,8. Найдите вероятность того, что студент сдаст экзамен, если для этого необходимо ответить на все вопросы. [0,648]

26. Пусть вероятность того, что покупателю необходима обувь 41-го размера, равна 0,2. Найдите вероятность того, что пять первых покупателей потребуют обувь 41-го размера. [0,00032]

27. В семье трое детей. Принимая события, состоящие в рождении мальчика и девочки, равновероятными, найдите вероятность того, что в семье: а) все мальчики; б) дети одного пола. [а) 0,125; б) 0,25]

28. В мешке смешаны нити, среди которых 30% белых, а остальные красные. Определите вероятность того, что вынутые наудачу две нити будут: а) белые; б) красные; в) одного цвета. [а) 0,09; б) 0,49; в) 0,58]

29. Пусть вероятность того, что человек умрет на 71-м году жизни, равна 0,04. Какова вероятность того, что из трех человек семидесяти лет через год все будут живы? [0,88]

30. Вероятность установления в данной местности устойчивого снежного покрова с октября равна 0,1. Определите вероятность того, что в ближайшие три года в этой местности устойчивый снежный покров с октября не установится ни разу. [0,729]

31. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом, втором, третьем справочнике, соответственно равны 0,6; 0,7; 0,8. Найдите вероятность того, что формула содержится во всех трех справочниках. [0,336]

32. Брошены две игральные кости. Какова вероятность того, что хотя бы на одном из кубиков выпадет 6 очков?

$\left[\frac{11}{36}\right]$

33. Производятся два выстрела по одной и той же мишени. Вероятность попадания при первом выстреле равна 0,6, для второго — 0,8. Найдите вероятность того, что в мишени будет хотя бы одна пробоина.

[0,92]

34. В специализированную больницу поступают в среднем 50% больных с заболеванием K, 30% — с заболеванием L, 20% — с заболеванием M. Вероятность полного излечения болезни K равна 0,7; для болезней L и M эти вероятности соответственно равны 0,8 и 0,9. Найдите вероятность того, что больной, поступивший в больницу, будет выписан здоровым.

[0,77]

35. Для посева заготовлена смесь семян пшеницы 4 сортов. Зерен первого сорта 96%, второго — 1%, третьего — 2% и четвертого сорта — 1%. Вероятности того, что из зерна каждого сорта вырастает колос, содержащий не менее 50 зерен, соответственно равны: 0,50; 0,15; 0,20; 0,05. Какова вероятность того, что колос, выросший из произвольно взятого из заготовленной смеси зерна, будет содержать не менее 50 зерен?

[0,4995]

36. В группе спортсменов 20 лыжников, 6 велосипедистов и 4 бегуна. Вероятность выполнить квалификационную норму такова: для лыжника — 0,9, для велосипедиста — 0,8 и для бегуна — 0,75. Найдите вероятность того, что спортсмен, выбранный наудачу, выполнит норму. [0,86]

37. Партия электрических лампочек на 20% изготовлена первым заводом, на 30% — вторым и на 50% — третьим заводом. Для первого завода вероятность выпуска бракованной лампочки равна 0,01, для второго — 0,005 и для третьего — 0,006. Какова вероятность того, что наудачу взятая из партии лампочка окажется бракованной? [0,0065]

38. Из 50 деталей 18 изготовлены в первом цехе, 20 — во втором, а остальные — в третьем. Первый и третий цехи изготавливают продукцию отличного качества с вероятностью 0,9, второй — с вероятностью 0,6. Какова вероятность того, что взятая наудачу деталь будет отличного качества?

[0,78]

39. На четырех карточках написаны буквы А, Е, П, Р. Карточки перемешиваются и раскладываются в ряд. Какова вероятность того, что получится слово РЕПА?

$\left[\frac{1}{24}\right]$

40. В урне находятся 15 шаров, из них 9 красных и 6 синих. Найдите вероятность того, что вынутые наугад два шара оба окажутся красными.

$\left[\frac{12}{35}\right]$

41. Брошены две игральные кости. Найдите вероятность того, что сумма выпавших очков равна 4.

$\left[\frac{1}{12}\right]$

42. В ящике имеется n шаров: m белых и $n-m$ черных. Какова вероятность того, что среди r наудачу вынутых шаров окажется k белых?

$\left[\frac{C_m^k C_{n-m}^{r-k}}{C_n^r}\right]$

К главе II

1. В лотерее на 100 билетов разыгрываются две вещи, стоимости которых 2100 и 600 р. Составьте закон распределения суммы выигрыша для лица, имеющего один билет.

Сумма выигрыша	0	600	2100
Вероятность	0,98	0,01	0,01

2. Случайная величина X задана законом распределения:

X	0	2	4	5
p	0,1	0,6	?	0,1

Какова вероятность того, что она примет значение 4? [0,2]

3. Найдите математическое ожидание случайной величины X, заданной законом распределения:

а)

X	–4	6	10
p	0,2	0,3	0,5

б)

X	0,21	0,54	0,61
p	0,1	0,5	0,4

в)

X	1	3	6	8
p	0,2	0,1	0,4	0,3

[а) 6; б) 0,535; в) 5,3]

4. Найдите математическое ожидание случайной величины Z, если известны математические ожидания X и Y:

а) $Z = X + Y$, $M(X) = 5$, $M(Y) = 3$;

б) $Z = 3X + 4Y$, $M(X) = 2$, $M(Y) = 6$.

[а) 8; б) 30]

5. Производятся 3 выстрела с вероятностями попадания в цель, равными $p_1 = 0{,}4$; $p_2 = 0{,}3$ и $p_3 = 0{,}6$. Найдите математическое ожидание общего числа попаданий. [1,3]

6. Независимые случайные величины X и Y заданы следующими законами распределения:

X	1	3	6	8
p	0,2	0,1	0,4	0,3

Y	2	4	5	6
p	0,3	0,1	0,2	0,4

Найдите математическое ожидание случайной величины XY. [23,32]

7. Дискретная случайная величина X принимает три возможных значения: $x_1 = 4$ с вероятностью $p_1 = 0{,}5$; $x_2 = 6$ с вероятностью $p_2 = 0{,}3$ и x_3 с вероятностью p_3. Найдите x_3 и p_3, зная, что $M(X) = 8$.

[$x_3 = 21$, $p_3 = 0{,}2$]

8. Дан перечень возможных значений дискретной случайной величины X: $x_1 = -1$; $x_2 = 0$, $x_3 = 1$, а также известны математические ожидания этой величины и ее квадрата: $M(X) = 0{,}1$, $M(X^2) = 0{,}9$. Найдите вероятности p_1, p_2, p_3, соответствующие возможным значениям x_1, x_2, x_3.

[$p_1 = 0{,}4$; $p_2 = 0{,}1$; $p_3 = 0{,}5$]

9. Найдите дисперсию случайной величины X, заданной законом распределения:

а)

X	-1	0	1
p	0,2	0,3	0,5

б)

X	0	1	2	3
p	0,1	0,2	0,3	0,4

[а) $D(X) = 0{,}61$; б) $D(X) = 1$]

10. Определите математическое ожидание и дисперсию случайной величины X, если закон распределения имеет вид:

X	0	1	3	4
p	0,2	0,1	0,3	0,4

[$M(X) = 2{,}6$; $D(X) = 2{,}44$]

11. Случайная величина X распределена по закону:

X	2	4	6	8	10
p	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$

Найдите $M(X)$, $D(X)$ и $\sigma(X)$. [$M(X) = 6$; $D(X) = 9$; $\sigma(X) = 3$]

12. Найдите дисперсии и среднее квадратическое отклонение дискретной случайной величины X, заданной законом распределения:

X	-5	2	3	4
p	0,4	0,3	0,1	0,2

[$D(X) = 15{,}21$; $\sigma(X) = 3{,}9$]

13. Известны дисперсии двух независимых случайных величин X и Y: $D(X) = 5$, $D(Y) = 6$. Найдите дисперсию суммы этих величин. [11]

14. Найдите дисперсии следующих величин: а) $2X$; б) $-3X$, если $D(X) = 4$.

[а) 16; б) 36]

15. Дискретная случайная величина X задана законом распределения:

X	2	3	5
p	0,1	0,4	0,5

Найдите начальные моменты первого и второго порядков.

[$\nu_1 = 3{,}9$; $\nu_2 = 16{,}5$]

16. Дискретная случайная величина X задана законом распределения:

X	1	2	4
p	0,1	0,3	0,6

Найдите центральный момент второго порядка. $[\mu_2 = 1{,}29]$

17. Случайная величина X задана интегральной функцией

$$F(x) = \begin{cases} 0 \text{ при } x \leqslant -1; \\ \frac{3}{4}x + \frac{3}{4} \text{ при } -1 < x \leqslant \frac{1}{3}; \\ 1 \text{ при } x > \frac{1}{3}. \end{cases}$$

Найдите вероятность того, что в результате испытания величина X примет значение, заключенное в интервале $(0;\ ^1/_3)$. [0,25]

18. Случайная величина X на всей оси Ox задана интегральной функцией $F(x) = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x$. Найдите вероятность того, что в результате испытания величина X примет значение, заключенное в интервале (0; 1). [0,25]

19. Случайная величина X задана интегральной функцией

$$F(x) = \begin{cases} 0 \text{ при } x \leqslant -2; \\ \frac{1}{2} + \frac{1}{\pi}\arcsin\frac{x}{2} \text{ при } -2 < x \leqslant 2; \\ 1 \text{ при } x > 2. \end{cases}$$

Найдите вероятность того, что в результате испытания величина X примет значение, заключенное в интервале (–1; 1).

$$\left[\frac{1}{3}\right]$$

20. Функция

$$f(x) = \frac{4a}{e^x + e^{-x}} \quad (-\infty < x < +\infty)$$

является плотностью вероятности случайной величины X. Найдите коэффициент a и функцию распределения $F(x)$.

$$\left[a = \frac{1}{2\pi};\ F(x) = \frac{2}{\pi}\operatorname{arctg} e^x\right]$$

21. Случайная величина X задана по всей оси Ox плотностью вероятности

$$f(x) = \frac{2a}{1 + x^2}.$$

Найдите постоянный параметр a. $\left[a = \frac{1}{2\pi}\right]$

22. Случайная величина X задана плотностью вероятности

$$f(x)=\begin{cases}0 \text{ при } x<-\frac{\pi}{2};\\ \frac{a}{2}\cos x \text{ при } -\frac{\pi}{2}\leqslant x\leqslant\frac{\pi}{2};\\ 0 \text{ при } x>\frac{\pi}{2}.\end{cases}$$

Найдите коэффициент a. [$a=1$]

23. Случайная величина X задана плотностью вероятности $f(x)=\frac{2}{3}\sin 3x$ в интервале $\left(0;\frac{\pi}{3}\right)$, вне этого интервала $f(x)=0$. Найдите вероятность того, что X примет значение, принадлежащее интервалу $\left(\frac{\pi}{6};\frac{\pi}{4}\right)$.

$\left[\frac{\sqrt{2}}{9}\right]$

24. Случайная величина X задана плотностью вероятности

$$f(x)=\begin{cases}0 \quad \text{при } x\leqslant 0;\\ 2x \text{ при } 0<x\leqslant 1;\\ 0 \quad \text{при } x>1.\end{cases}$$

Найдите математическое ожидание случайной величины X. $\left[\frac{2}{3}\right]$

25. Случайная величина X задана плотностью вероятности

$$f(x)=\begin{cases}0 \text{ при } x\leqslant -2;\\ \frac{1}{4} \text{ при } -2<x\leqslant 2;\\ 0 \text{ при } x>2.\end{cases}$$

Найдите математическое ожидание и дисперсию случайной величины X.

[$M(X)=0$; $D(X)={}^4/_3$]

26. Два равносильных шахматиста играют в шахматы. Что вероятнее выиграть: одну партию из двух или две партии из четырех (ничьи во внимание не принимаются)?

[Вероятнее выиграть одну партию из двух]

27. Пусть всхожесть семян данного растения составляет 70%. Определите вероятность того, что из трех посеянных семян взойдут: а) два; б) не менее двух. [а) 0,441; б) 0,784]

28. В семье пятеро детей. Найдите вероятность того, что среди этих детей два мальчика. Вероятность рождения мальчика принять равной 0,51.

[0,31]

29. Монету бросали четыре раза. Чему равна при этом вероятность выпадения герба два раза? [0,375]

30. Монета подбрасывается три раза. Какова вероятность того, что герб появится не менее двух раз? [0,5]

31. Монета подбрасывается три раза. Рассматривается случайная величина X — число появлений герба. Найдите закон распределения случайной величины X.

X	0	1	2	3
p	0,125	0,375	0,375	0,125

32. Найдите математическое ожидание числа бракованных изделий в партии из 10 000 изделий, если каждое изделие может оказаться бракованным с вероятностью 0,005. [50 изделий]

33. Из всей выпускаемой фабрикой продукции 98% составляют изделия со Знаком качества. Найдите: а) математическое ожидание и б) дисперсию числа изделий со Знаком качества в партии из 5000 изделий.
[а) 4900; б) 98]

34. Подлежат исследованию 1200 проб руды. Пусть вероятность промышленного содержания металла в каждой пробе равна 0,09. Найдите: а) математическое ожидание и б) дисперсию числа проб с промышленным содержанием металла. [а) 108; б) 98,28]

35. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X соответственно равны 3 и 2. Найдите плотность вероятности случайной величины X $f(x)$.

$$\left[f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{8}}\right]$$

36. Напишите дифференциальную функцию нормально распределенной случайной величины X, зная, что $M(X) = 3$, $D(X) = 16$.

$$\left[f(x) = \frac{1}{4\sqrt{2\pi}} e^{\frac{-(x-3)^2}{32}}\right]$$

37. Нормально распределенная случайная величина X задана дифференциальной функцией $f(x) = \dfrac{1}{5\sqrt{2\pi}} e^{-\frac{(x-1)^2}{50}}$.

Найдите математическое ожидание и дисперсию X.
$[M(X) = 1,\ D(X) = 25]$

38. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X соответственно равны 10 и 2. Найдите вероятность того, что в результате испытания X примет значение, заключенное в интервале (12; 14). [0,1359]

39. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X соответственно равны 20 и 5. Найдите вероятность того, что в результате испытания X примет значение в интервале (15; 25). [0,6826]

40. Детали, выпускаемые цехом, по размеру диаметра распределяются по нормальному закону с параметрами: математическое ожидание равно 5 см, а среднее квадратическое отклонение равно 0,9 см. Найдите вероятность того, что отклонение диаметра наудачу взятой детали от математического ожидания по абсолютной величине будет меньше 2 см.

[0,9736]

41. Проводится взвешивание некоторого вещества без систематических ошибок. Случайные ошибки взвешивания подчинены нормальному закону со средним квадратическим отклонением $\sigma = 20$ г. Найдите вероятность того, что взвешивание будет проведено с ошибкой, которая по абсолютной величине меньше 10 г. [0,383]

42. АТС получает в среднем за час 300 вызовов. Какова вероятность того, что за данную минуту она получит точно 2 вызова? [0,09]

43. Среди 1000 человек приблизительно 8 левшей. Какова вероятность того, что среди сотни наугад выбранных человек не окажется ни одного *левши*? [0,4493]

44. Монету подбрасывают 100 раз. Какова вероятность того, что при этом герб выпадет ровно 50 раз? [0,08]

45. Какова вероятность того, что при 200-кратном бросании монеты число случаев выпадения герба S_{200} удовлетворяет неравенству $95 \leqslant S_{200} \leqslant 105$?

[0,5224]

46. Вероятность получения по лотерее проигрышного билета равна 0,1. Какова вероятность того, что среди 500 наугад купленных билетов не менее 48 и не более 55 безвыигрышных? [0,3913]

47. Какова вероятность того, что при подбрасывании монеты три раза герб выпадет по крайней мере один раз?

$\left[\frac{7}{8}\right]$

К главе III

1. Найдите законы распределения составляющих дискретной двумерной случайной величины, заданной законом распределения

X \ Y	$y_1 = 0{,}4$	$y_2 = 0{,}8$
$x_1 = 2$	0,15	0,05
$x_2 = 5$	0,30	0,12
$x_3 = 8$	0,35	0,03

X	2	5	8	Y	0,4	0,8
p	0,20	0,42	0,38	p	0,80	0,20

2. Найдите вероятность того, что составляющая X двумерной случайной величины (X, Y) примет значение $X < 2$ и при этом составляющая Y примет значение $Y < 4$, если известна функция распределения величины (X, Y).

$$F(x, y) = (1 - e^{-x})(1 - e^{-y}) \quad (x \geqslant 0,\ y \geqslant 0).$$

[0,849]

3. Найдите вероятность попадания случайной точки $(X,\ Y)$ в прямоугольник с вершинами $A\left(\frac{\pi}{6};\ \frac{\pi}{4}\right)$, $B\left(\frac{\pi}{6};\ \frac{\pi}{3}\right)$, $C\left(\frac{\pi}{2};\ \frac{\pi}{3}\right)$, $D\left(\frac{\pi}{2};\ \frac{\pi}{4}\right)$, если известна функция распределения $F(x,\ y) = \sin x \sin y \left(0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right)$.

[0,08]

4. Найдите плотность вероятности $f(x,\ y)$ двумерной случайной величины $(X,\ Y)$ по известной функции распределения

$$F(x,\ y) = (1 - e^{-2x})(1 - e^{-3y})\ \ (x \geqslant 0,\ y \geqslant 0).$$

$[f(x,\ y) = 6e^{-(2x+3y)}]$

5. Плотность вероятности двумерной случайной величины

$$f(x,\ y) = \frac{C}{(9 + x^2)(16 + y^2)}.$$

Найдите величину C. $\left[C = \frac{12}{\pi^2}\right]$

6. Задана двумерная плотность вероятности случайной величины $(X,\ Y)$

$$f(x,\ y) = \begin{cases} \frac{1}{2}\sin(x + y) \text{ в квадрате } S\left\{0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right\}, \\ 0 \text{ вне квадрата } S. \end{cases}$$

Найдите функцию распределения величины $(X,\ Y)$.

$\left[F(x,\ y) = \frac{1}{2}(\sin x + \sin y - \sin(x + y))\right]$

7. Задана плотность вероятности двумерной случайной величины

$$f(x,\ y) = \frac{\sqrt{3}}{\pi}e^{-x^2 - 2xy - 4y^2}.$$

Найдите плотности распределения составляющих.

$$\left[\begin{aligned} f_1(x) &= \frac{1}{2}\sqrt{\frac{3}{\pi}}\, e^{-0{,}75x^2}, \\ f_2(y) &= \sqrt{\frac{3}{\pi}}\, e^{-3y^2} \end{aligned}\right]$$

8. Дискретная двумерная случайная величина задана таблицей

$X \backslash Y$	y_1	y_2
x_1	0,15	0,30
x_2	0,06	0,10
x_3	0,25	0,03
x_4	0,04	0,07

Найдите условный закон распределения составляющей X при условии, что составляющая Y приняла значение y_1.

X	x_1	x_2	x_3	x_4
$P_{y_1}(X)$	0,3	0,12	0,5	0,08

9. В упражнении 7 найдите условные законы распределения вероятностей составляющих

$$\left[\begin{aligned}\varphi_y(x) &= \frac{1}{\sqrt{\pi}}e^{-(x+y)^2},\\ \psi_x(y) &= \frac{2}{\sqrt{\pi}}e^{-0,25(x+4y)^2}\end{aligned}\right]$$

10. Задана плотность распределения непрерывной двумерной случайной величины

$$f(x, y) = \begin{cases}\cos x \cos y \text{ в квадрате } S\left\{0 < x < \frac{\pi}{2},\ 0 < y < \frac{\pi}{2}\right\},\\ 0 \text{ вне квадрата } S.\end{cases}$$

Докажите, что составляющие X и Y независимы.

11. Дана таблица, определяющая закон распределения двумерной случайной величины

$X \backslash Y$	-1	1
-1	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$

Найдите коэффициент корреляции. $[r_{xy} = 0]$

12. Задана плотность распределения непрерывной двумерной случайной величины

$$f(x, y) = \begin{cases}\cos x \cos y \text{ в квадрате } S\left\{0 \leqslant x \leqslant \frac{\pi}{2},\ 0 \leqslant y \leqslant \frac{\pi}{2}\right\},\\ 0 \text{ вне квадрата } S.\end{cases}$$

Найдите корреляционный момент и коэффициент корреляции.
$[\mu_{xy} = 0,\ r_{xy} = 0]$

К главе IV

1. Перейдите от частот к относительным частотам в следующем распределении выборки объема $n = 10$:

Варианта x_i	2	5	7
Частота n_i	1	3	6

[

Варианта x_i	2	5	7
Относительная частота p_i^*	0,1	0,3	0,6

]

2. Перейдите от частот к относительным частотам в следующем распределении выборки объема $n=20$:

Варианта x_i	4	7	8	12
Частота n_i	5	2	3	10

[
Варианта x_i	4	7	8	12
Относительная частота p_i^*	0,25	0,10	0,15	0,50
]

3. Постройте полигон по данному распределению:

а)

x_i	1	4	5	7
n_i	20	10	14	6

б)

x_i	2	3	5	6
n_i	10	15	5	20

в)

x_i	2	4	5	7	10
p_i^*	0,15	0,2	0,1	0,1	0,45

г)

x_i	1	4	5	8	9
p_i^*	0,15	0,25	0,3	0,2	0,1

4. Постройте гистограмму по данному распределению выборки объема $n=100$:

Частичный интервал	Сумма частот вариант частичного интервала n_i
1—5	10
5—9	20
9—13	50
13—17	12
17—21	8

5. Постройте гистограмму (с переходом к относительным частотам) по данному распределению выборки:

Частичный интервал	Сумма частот вариант частичного интервала n_i
0—2	20
2—4	30
4—6	50

6. Генеральная совокупность задана таблицей распределения:

x_i	2560	2600	2620	2650	2700
N_i	2	3	10	4	1

Найдите генеральную среднюю $\bar{x}_г$. [$\bar{x}_г = 2621$]

7. Генеральная совокупность задана таблицей распределения:

x_i	186	192	194
N_i	2	5	3

Найдите генеральную среднюю $\bar{x}_г$ и генеральную дисперсию $D_г$.
[$\bar{x}_г = 191{,}4$; $D_г = 8{,}04$]

8. Из генеральной совокупности извлечена выборка объема $n = 60$:

Варианта x_i	1	3	6	26
Частота n_i	8	40	10	2

Найдите выборочную среднюю. [$\bar{x}_в = 4$]

9. Выборочная совокупность задана таблицей распределения:

x_i	2	5	7	10
n_i	16	12	8	14

Найдите выборочную среднюю. [$\bar{x}_в = 5{,}76$]

10. Найдите выборочную среднюю по данному распределению выборки объема $n = 10$:

x_i	1250	1270	1280
n_i	2	5	3

[$\bar{x}_в = 1269$]

11. Найдите выборочную среднюю по следующим данным: а) длина крыла у 6 пчел (в мм): 9,68; 9,81; 9,77; 9,60; 9,61; 9,55; б) длина листьев садовой земляники (в см): 5,2; 5,6; 7,1; 6,6; 8,6; 8,2; 7,7; 7,8.
[а) 9,67 мм; б) 7,1 см]

12. Найдите выборочную дисперсию по данному распределению выборки объема $n = 10$:

x_i	0,01	0,04	0,08
n_i	5	3	2

[$D_в \approx 0{,}0007$]

13. По выборке объема $n=41$ найдена выборочная дисперсия $D_в=3$. Найдите исправленную дисперсию. $[s^2 \approx 3{,}075]$

14. Найдите исправленную выборочную дисперсию по данному распределению выборки объема $n=10$:

x_i	102	104	108
n_i	2	3	5

$[s^2 = 6{,}93]$

15. Найдите исправленную выборочную дисперсию по данному распределению выборки объема $n=20$:

x_i	0,1	0,5	0,07	0,9
n_i	6	12	1	1

$[s^2 = 0{,}0525]$

16. В итоге пяти измерений длины стержня одной линейной (без систематических ошибок) получены следующие результаты (в мм): 92, 94, 103, 105, 106. Найдите выборочную среднюю длины стержня, а также выборочную и исправленную дисперсии ошибок измерения линейкой.
$[\bar{x}_в = 100;\ D_в = 34;\ s^2 = 42{,}5]$

17. Получены результаты измерения роста случайно отобранных 100 студентов и студенток:

Рост, см	154—158	158—162	162—166	166—170	170—174	174—178	178—182
Число студентов	10	14	26	28	12	8	2

Найдите выборочную среднюю и выборочную дисперсию роста обследованных студентов.

Указание. Найдите середины интервалов и примите их в качестве вариант.

$[\bar{x}_в = 166;\ D_в = 33{,}44]$

18. Найдите доверительный интервал для оценки с надежностью 0,95 неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если даны генеральное среднее квадратическое отклонение $\sigma = 5$, выборочная средняя $\bar{x}_в = 14$ и объем выборки $n=25$. $[12{,}04 < a < 15{,}96]$

19. Найдите доверительный интервал для оценки с надежностью 0,99 неизвестного математического ожидания a нормально распределенного признака X генеральной совокупности, если даны генеральное среднее квадратическое отклонение σ, выборочная средняя $\bar{x}_в$ и объем выборки n:

а) $\sigma = 4$, $\bar{x}_в = 10{,}2$, $n = 16$;

а) $\sigma = 5$, $\bar{x}_в = 16{,}8$, $n = 25$.

[а) $7{,}63 < a < 12{,}77$; б) $14{,}23 < a < 19{,}37$]

20. Из генеральной совокупности извлечена выборка объема $n = 10$:

Варианта x_i	–2	1	2	3	4	5
Частота n_i	2	1	2	2	2	1

Пользуясь распределением Стьюдента, оцените с надежностью 0,95 математическое ожидание a нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала. [$0{,}3 < a < 3{,}7$]

21. Из генеральной совокупности извлечена выборка объема $n = 12$:

x_i	–0,5	–0,4	–0,2	0	0,2	0,6	0,8	1	1,2	1,5
n_i	1	2	1	1	1	1	1	1	2	1

Пользуясь распределением Стьюдента, оцените с надежностью 0,95 математическое ожидание a нормально распределенного признака генеральной совокупности при помощи доверительного интервала.

[$-0{,}04 < a < 0{,}88$]

22. По данным 16 независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\bar{x}_в = 42{,}8$ и исправленное среднее квадратическое отклонение $s = 8$. Оцените истинное значение a измерений величины с надежностью $\gamma = 0{,}999$.

[$34{,}66 < a < 50{,}94$]

23. По данным выборки объема $n = 16$ из генеральной совокупности найдено исправленное среднее квадратическое отклонение $s = 1$ нормально распределенного количественного признака. Найдите доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,95. [$0{,}56 < a < 1{,}44$]

24. По данным выборки объема n из генеральной совокупности нормально распределенного количественного признака найдено исправленное среднее квадратическое отклонение s. Найдите доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,999, если:

а) $n = 10$, $s = 5{,}1$; б) $n = 50$, $s = 14$.

[а) $0 < \sigma < 14{,}28$; б) $7{,}98 < \sigma < 20{,}02$]

25. Проведено 12 измерений одним прибором (без систематических ошибок) некоторой физической величины, причем исправленное среднее квадратическое отклонение s случайных ошибок измерений оказалось равным 0,6. Найдите точность прибора с надежностью 0,99.

[$0{,}06 < \sigma < 1{,}14$]

26. Проведено 10 измерений одним прибором (без систематических ошибок) некоторой физической величины, причем исправленное среднее квадратическое отклонение случайных ошибок измерений оказалось равным 0,8. Найдите точность прибора с надежностью 0,95.

[$0{,}28 < \sigma < 1{,}32$]

27. При уровне значимости 0,01 проверьте гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

Эмпирические частоты	8	16	40	72	36	18	10
Теоретические частоты	6	18	36	76	39	18	7

[$\chi_0^2 = 3{,}068$; $\chi^2(0{,}01;\ 4) = 13{,}3$. Нет оснований отвергнуть гипотезу]

28. В следующих упражнениях при уровне значимости 0,05 проверьте гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты:

а)

Эмпирические частоты	5	10	40	8	7
Теоретические частоты	6	14	18	7	5

[$\chi_0^2 = 2{,}47$; $\chi^2(0{,}05;\ 2) = 6{,}0$. Нет оснований отвергнуть гипотезу]

б)

Эмпирические частоты	6	8	13	15	20	16	10	7	5
Теоретические частоты	5	9	14	16	18	16	9	6	7

[$\chi_0^2 = 1{,}52$; $\chi^2(0{,}05;\ 6) = 12{,}6$. Нет оснований отвергнуть гипотезу]

в)

Эмпирические частоты	14	18	32	70	20	36	10
Теоретические частоты	10	24	34	80	18	22	12

[$\chi_0^2 = 13{,}93$; $\chi^2(0{,}05;\ 4) = 9{,}5$. Гипотеза отвергается]

г)

Эмпирические частоты	5	7	15	14	21	16	9	7	6
Теоретические частоты	6	6	14	15	22	15	8	8	6

[$\chi_0^2 = 0{,}83$; $\chi^2(0{,}05;\ 6) = 12{,}6$. Нет оснований отвергнуть гипотезу]

29. Найдите выборочное уравнение прямой регрессии Y на X по данным следующей таблицы:

x_i	23,0	24,0	24,5	24,5	25,0	25,5	26,0	26,0	26,5	26,5	27,0	27,0	28,0
y_i	0,48	0,50	0,49	0,50	0,51	0,52	0,51	0,53	0,50	0,52	0,54	0,52	0,53

[$y = 0{,}0098x + 0{,}2581$]

30. По данным таблицы, приведенной в предыдущем упражнении, найдите выборочное уравнение прямой регрессии X на Y.

[$x = 64y - 7{,}012$]

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1

ТАБЛИЦЫ ПРОИЗВОДНЫХ И ИНТЕГРАЛОВ

Таблица производных

1. $(u^{\alpha})' = \alpha u^{\alpha - 1} u'$.
2. $(a^{u})' = a^{u} u' \ln a$.
3. $(e^{u})' = e^{u} u'$.
4. $(\log_a u)' = \dfrac{u'}{u \ln a}$.
5. $(\ln u)' = \dfrac{u'}{u}$.
6. $(\sin u)' = u' \cos u$.
7. $(\cos u)' = -u' \sin u$.
8. $(\operatorname{tg} u)' = \dfrac{u'}{\cos^2 u}$.
9. $(\operatorname{ctg} u)' = -\dfrac{u'}{\sin^2 u}$.
10. $(\arcsin u)' = \dfrac{u'}{\sqrt{1 - u^2}}$.
11. $(\arccos u)' = \dfrac{-u'}{\sqrt{1 - u^2}}$.
12. $(\operatorname{arctg} u)' = \dfrac{u'}{1 + u^2}$.
13. $(\operatorname{arcctg} u)' = \dfrac{-u'}{1 + u^2}$.

Здесь $u = u(x)$ и $v = v(x)$ — дифференцируемые функции.

Таблица интегралов

1. $\int x^{\mu} dx = \dfrac{x^{\mu + 1}}{\mu + 1} + C \quad (\mu \neq -1)$.
2. $\int \dfrac{dx}{x} = \ln |x| + C$.
3. $\int a^{x} dx = \dfrac{a^{x}}{\ln a} + C$.

4. $\int e^{kx}dx = \frac{e^{kx}}{k} + C.$

5. $\int \sin axdx = -\frac{1}{a}\cos ax + C.$

6. $\int \cos axdx = \frac{1}{a}\sin ax + C$

7. $\int \frac{dx}{\cos^2 ax} = \frac{1}{a}\operatorname{tg} ax + C.$

8. $\int \frac{dx}{\sin^2 ax} = -\frac{1}{a}\operatorname{ctg} ax + C.$

9. $\int \frac{xdx}{x^2 + a} = \frac{1}{2}\ln|x^2 + a| + C.$

10. $\int \frac{dx}{x^2 + a^2} = \frac{1}{a}\operatorname{arctg}\frac{x}{a} + C.$

11. $\int \frac{dx}{\sqrt{a^2 - x^2}} = \arcsin\frac{x}{a} + C.$

12. $\int \frac{dx}{\sqrt{x^2 + a}} = \ln|x + \sqrt{x^2 + a}|+C.$

13. $\int \frac{dx}{x^2 - a^2} = \frac{1}{2a}\ln\left|\frac{x - a}{x + a}\right| + C.$

14. $\int\limits_0^{+\infty} e^{-x^2}dx = \frac{\sqrt{\pi}}{2}$ (интеграл Эйлера — Пуассона).

ПРИЛОЖЕНИЕ 2

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315

x	0	1	2	3	4	5	6	7	8	9
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

ПРИЛОЖЕНИЕ 3

Таблица значений функции $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,0000	0,09	0,0359	0,18	0,0714	0,27	0,1064
0,01	0,0040	0,10	0,0398	0,19	0,0753	0,28	0,1103
0,02	0,0080	0,11	0,0438	0,20	0,0793	0,29	0,1141
0,03	0,0120	0,12	0,0478	0,21	0,0832	0,30	0,1179
0,04	0,0160	0,13	0,0517	0,22	0,0871	0,31	0,1217
0,05	0,0199	0,14	0,0557	0,23	0,0910	0,32	0,1255
0,06	0,0239	0,15	0,0596	0,24	0,0948	0,33	0,1293
0,07	0,0279	0,16	0,0636	0,25	0,0987	0,34	0,1331
0,08	0,0319	0,17	0,0675	0,26	0,1026	0,35	0,1368

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,36	0,1406	0,79	0,2852	1,22	0,3888	1,65	0,4505
0,37	0,1443	0,80	0,2881	1,23	0,3907	1,66	0,4515
0,38	0,1480	0,81	0,2910	1,24	0,3925	1,67	0,4525
0,39	0,1517	0,82	0,2939	1,25	0,3944	1,68	0,4535
0,40	0,1554	0,83	0,2967	1,26	0,3962	1,69	0,4545
0,41	0,1591	0,84	0,2995	1,27	0,3980	1,70	0,4554
0,42	0,1628	0,85	0,3023	1,28	0,3997	1,71	0,4564
0,43	0,1664	0,86	0,3051	1,29	0,4015	1,72	0,4573
0,44	0,1700	0,87	0,3078	1,30	0,4032	1,73	0,4582
0,45	0,1736	0,88	0,3106	1,31	0,4049	1,74	0,4591
0,46	0,1772	0,89	0,3133	1,32	0,4066	1,75	0,4599
0,47	0,1808	0,90	0,3159	1,33	0,4082	1,76	0,4608
0,48	0,1844	0,91	0,3186	1,34	0,4099	1,77	0,4616
0,49	0,1879	0,92	0,3212	1,35	0,4115	1,78	0,4625
0,50	0,1915	0,93	0,3238	1,36	0,4131	1,79	0,4633
0,51	0,1950	0,94	0,3264	1,37	0,4147	1,80	0,4641
0,52	0,1985	0,95	0,3289	1,38	0,4162	1,81	0,4649
0,53	0,2019	0,96	0,3315	1,39	0,4177	1,82	0,4656
0,54	0,2054	0,97	0,3340	1,40	0,4192	1,83	0,4664
0,55	0,2088	0,98	0,3365	1,41	0,4207	1,84	0,4671
0,56	0,2123	0,99	0,3389	1,42	0,4222	1,85	0,4678
0,57	0,2157	1,00	0,3413	1,43	0,4236	1,86	0,4686
0,58	0,2190	1,01	0,3438	1,44	0,4251	1,87	0,4693
0,59	0,2224	0,02	0,3461	1,45	0,4265	1,88	0,4699
0,60	0,2257	1,03	0,3485	1,46	0,4279	1,89	0,4706
0,61	0,2291	1,04	0,3508	1,47	0,4292	1,90	0,4713
0,62	0,2324	1,05	0,3531	1,48	0,4306	1,91	0,4719
0,63	0,2357	1,06	0,3554	1,49	0,4319	1,92	0,4726
0,64	0,2389	1,07	0,3577	1,50	0,4332	1,93	0,4732
0,65	0,2422	1,08	0,3599	1,51	1,4345	1,94	0,4738
0,66	0,2454	1,09	0,3621	1,52	0,4357	1,95	0,4744
0,67	0,2486	1,10	0,3643	1,53	0,4370	1,96	0,4750
0,68	0,2517	1,11	0,3665	1,54	0,4382	1,97	0,4756
0,69	0,2549	1,12	03686	1,55	0,4394	1,98	0,4761
0,70	0,2580	1,13	0,3708	1,56	0,4406	1,99	0,4767
0,71	0,2611	1,14	0,3729	1,57	0,4418	2,00	0,4772
0,72	0,2642	1,15	0,3749	1,58	0,4429	2,02	0,4783
0,73	0,2673	1,16	0,3770	1,59	0,4441	2,04	0,4793
0,74	0,2703	1,17	0,3790	1,60	0,4452	2,06	0,4803
0,75	0,2734	1,18	0,3810	1,61	0,4463	2,08	0,4812
0,76	0,2764	1,19	0,3830	1,62	0,4474	2,10	0,4821
0,77	0,2794	1,20	0,3949	1,63	0,4484	2,12	0,4830
0,78	0,2823	1,21	0,3869	1,64	0,4495	2,14	0,4838

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,16	0,4846	2,42	0,4922	2,68	0,4963	2,94	0,4984
2,18	0,4854	2,44	0,4927	2,70	0,4965	2,96	0,4985
2,20	0,4861	2,46	0,4931	2,72	0,4967	2,98	0,4986
2,22	0,4868	2,48	0,4934	2,74	0,4969	3,00	0,49865
2,24	0,4875	2,50	0,4938	2,76	0,4971	3,20	0,49931
2,26	0,4881	2,52	0,4941	2,78	0,4973	3,40	0,49966
2,28	0,4887	2,54	0,4945	2,80	0,4974	3,60	0,499841
2,30	0,4893	2,56	0,4948	2,82	0,4976	3,80	0,499928
2,32	0,4898	2,58	0,4951	2,84	0,4977	4,00	0,499968
2,34	0,4904	2,60	0,4953	2,86	0,4979	4,50	0,499997
2,36	0,4909	2,62	0,4956	2,88	0,4980	5,00	0,500000
2,38	0,4913	2,64	0,4959	2,90	0,4981		
2,40	0,4918	2,66	0,4961	2,92	0,4982		

ПРИЛОЖЕНИЕ 4

Таблица значений $t_\gamma = t(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,861	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,991	2,640	3,418
15	2,15	2,98	1,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

ПРИЛОЖЕНИЕ 5

Таблица значений $q = q(\gamma, n)$

n \ γ	0,95	0,99	0,999	n \ γ	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,221
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

ПРИЛОЖЕНИЕ 6

Таблица значений χ^2 в зависимости от p и k

k \ p	0,05	0,01	0,001
1	3,84	6,64	10,83
2	5,99	9,21	13,82
3	7,82	11,34	16,27
4	9,49	13,28	18,46
5	11,07	15,09	20,5
6	12,59	16,81	22,5
7	14,07	18,48	24,3
8	15,51	20,1	26,1
9	16,92	21,7	27,9
10	18,31	23,2	29,6
11	19,68	24,7	31,3
12	21,0	26,2	32,9
13	22,4	27,7	34,6
14	23,7	29,1	36,1
15	25,0	30,6	37,7
16	26,3	32,0	39,3
17	27,6	33,4	40,8

k \ p	0,05	0,01	0,001
18	28,9	34,8	42,3
19	30,1	36,2	43,8
20	31,4	37,6	45,3
21	32,7	38,9	46,8
22	33,9	40,3	48,3
23	35,2	41,6	49,7
24	36,4	43,0	51,2
25	37,7	44,3	52,6
26	38,9	45,6	54,1
27	40,1	47,0	55,5
28	41,3	48,3	56,9
29	42,6	49,6	58,3
30	43,8	50,9	59,7

ПРИЛОЖЕНИЕ 7

Латинский алфавит

Буквы	Название	Буквы	Название
Aa	а	*Nn*	эн
Bb	бэ	*Oo*	о
Cc	цэ	*Pp*	пэ
Dd	дэ	*Qq*	ку
Ee	э	*Rr*	эр
Ff	эф	*Ss*	эс
Gg	гэ (жэ)*	*Tt*	тэ
Hh	ха (аш)	*Uu*	у
Ii	и	*Vv*	вэ
Jj	йот (жи)	*Ww*	дубль-вэ
Kk	ка	*Xx*	икс
Ll	эль	*Yy*	игрек
Mm	эм	*Zz*	зэт

* В скобках даны французские названия этих букв.

Греческий алфавит

Буквы	Название	Буквы	Название
Αα	альфа	Νν	ню
Ββ	бета	Ξξ	кси
Γγ	гамма	Οο	омикрон
Δδ	дельта	Ππ	пи
Εε	эпсилон	Ρρ	ро
Ζζ	дзета	Σσ	сигма
Ηη	эта	Ττ	тау
Θθ	тета	Υυ	ипсилон
Ιι	йота	Φφ	фи
Κκ	каппа	Χχ	хи
Λλ	ламбда	Ψψ	пси
Μμ	мю	Ωω	омега

ЛИТЕРАТУРА

1. Гнеденко Б. В., Хинчин А. Я. Элементарное введение в теорию вероятностей.— М.: Наука, 1970.
2. Гнеденко Б. В. Курс теории вероятностей.— М.: Наука, 1988.
3. Крамер Г. Математические методы статистики.— М.: Мир, 1975.
4. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1.— М.: Мир, 1984.
5. Колмогоров А. Н. Осиовные понятия теории вероятностей.— М.: Наука, 1974.

ОГЛАВЛЕНИЕ

Учебное издание

Баврин Иван Иванович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Редактор *А. Г. Гаврилов*
Художник *Ю. С. Сметанина*
Художественный редактор *А. Ю. Войткевич*
Технический редактор *Н. И. Тростянская*
Компьютерный набор и верстка *Л. В. Коростылева*
Корректоры *В. А. Жилкина, О. Н. Шебашова*

Лицензия ИД № 06236 от 09.11.01.

Изд. № РЕНТ-308. Подп. в печать 26.04.05
Формат $60 \times 88\,^1/_{16}$. Бум. газетная. Гарнитура «Ньютон». Печать офсетная.
Объем 9,8 усл. печ. л., 10,3 усл. кр.-отт.
Тираж 3000 экз. Зак. 4857.

ФГУП «Издательство «Высшая школа», 127994, Москва, ГСП-4,
Неглинная ул., 29/14.

Тел.: (095) 200-04-56
http://www.v-shkola.ru. E-mail: info@v-shkola.ru

Отдел реализации: (095) 200-07-69, 200-31-47, факс: (095) 200-34-86.
E-mail: sales@v-shkola.ru

Отпечатано на ФГУП ордена «Знак Почета»
Смоленская областная типография им. В. И. Смирнова.
214000, г. Смоленск, пр-т им. Ю. Гагарина, 2.