УДК 519.2(075.8)
ББК 22.17я73
Д33

Рецензенты:
Й. Микешь, д-р физ.-мат. наук, проф. (Университет им. Ф. Палацкого, Чешская Республика),
И.А. Александрова, доц. Департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве РФ, канд. физ.-мат. наук, доц.

Авторы (Департамент анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве РФ):
И.Е. Денежкина, канд. техн. наук, доц.,
С.Е. Степанов, д-р физ.-мат. наук;
И.И. Цыганок, канд. физ.-мат. наук

Денежкина, Ирина Евгеньевна.
Д33 Теория вероятностей и математическая статистика в вопросах и задачах : учебное пособие / И.Е. Денежкина, С.Е. Степанов, И.И. Цыганок.

Содержит краткий теоретический материал по курсу «Теория вероятностей и математическая статистика», читаемый для бакалавров направлений «Экономика», «Менеджмент», «Управление персоналом», подробные решения типовых задач, большую подборку задач для самостоятельного решения по каждой теме курса, удобную при рейтинговом контроле, а также при дистанционном обучении.
Соответствует ФГОС ВО последнего поколения.

Для студентов бакалавриата высших учебных заведений, преподавателей теории вероятностей и математической статистики и всех интересующихся этим разделом курса математики.

Ключевые слова: случайные события и величины; основы математической статистики.

УДК 519.2(075.8)
ББК 22.17я73

Денежкина Ирина Евгеньевна, Степанов Сергей Евгеньевич
Цыганок Ирина Ивановна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В ВОПРОСАХ И ЗАДАЧАХ

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Быстро меняющийся рынок труда, связанный с развитием цифровой экономики, будет предполагать растущий спрос на специалистов в области анализа данных, математического моделирования, финансовых технологий. Такие специалисты должны иметь серьезную математическую подготовку. Конечно, изучение математики имеет мировоззренческое значение, но для будущих экономистов и менеджеров математика должна стать инструментом анализа, организации, управления и контроля экономических и социальных систем. Рыночные отношения, сложные экономические связи, научно-технический прогресс способствуют интенсивному развитию производительных сил и приводят к существенным изменениям методов труда. Принимаемые решения в таких условиях должны быть обоснованными, методы руководства — гибкими и направленными на повышение эффективности использования кадрового потенциала. Разработка стратегий, обеспечивающих повышение продуктивности ресурсов и способствующих росту производительности труда, также требует не только качественной подготовки специалистов, но и эффективного управления персоналом. Оценка такой эффективности предполагает изучение многих факторов, что невозможно без правильно организованного сбора и обработки информации. Теория вероятностей и математическая статистика знакомят студентов с существующими приемами и методами, позволяющими решать задачи, связанные с анализом данных в области экономики, финансовой деятельности, менеджмента, в частности, управления персоналом.

Выпускники-бакалавры по роду своей деятельности редко встречаются с ситуациями, требующими решения серьезных вероятностных или статистических задач. Но даже при их решении часто используются более простые задачи теории вероятностей и математической статистики. Целью этого пособия является формирование у студентов общих представлений об основных понятиях теории вероятностей и математической статистики, которые позволят в дальнейшем обоснованно выбирать методы для эффективного анализа данных. Необходимый теоретический материал, большое количество примеров решенных задач, а также список задач для самостоятельного решения после каждого параграфа будут способствовать достижению поставленной цели.


ВЫЧИСЛИТЕЛЬНЫЕ
И ЭКСПЕРИМЕНТАЛЬНЫЕ
МЕТОДЫ
НАУЧНОГО
ЭКСПЕРИМЕНТА
Н.Ю. АФАНАСЬЕВА
КНОРУС

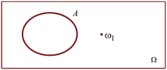


Рис. 1.1. Пространство событий

Если события не могут наступить одновременно в одном и том же испытании, то их называют ***несовместными*** (рис. 1.2).

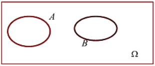


Рис. 1.2. Несовместные события

Если события могут наступить одновременно в одном и том же испытании, то их называют ***совместными*** (рис. 1.3).

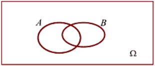


Рис. 1.3. Совместные события

Если в одном и том же испытании при наступлении события A событие B наступает обязательно, то говорят, что A ***благоприятствует*** B (рис. 1.4).

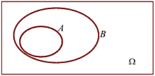


Рис. 1.4. Событие A благоприятствует событию B

Если событие A благоприятствует B и, одновременно, B благоприятствует A, то события A и B определяются как равные или ***тождественные***.

С событиями выполнимы операции сложения, вычитания и умножения. Результаты действий операций определяются как новые события. Их области наступления можно отмечать штриховкой на диаграмме. Для новых событий можно заполнять таблицу значимости, в которой 1 обозначает наступление события, а 0 отмечает, что событие не наступило.

Суммой $A + B$ двух событий называют новое событие, состоящее в том, что наступило ***хотя бы одно*** из событий A или B (рис. 1.5).

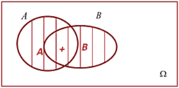


Рис. 1.5. Сумма событий A и B

Таблица значимости для суммы событий имеет вид:

A	B	$A+B$
1	1	1
1	0	1
0	1	1
0	0	0

Разностью $A - B$ двух событий называют новое событие, состоящее в том, что A наступило, а B — не наступило (рис. 1.6).

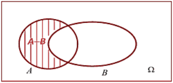


Рис. 1.6. Разность событий A и B

Таблица значимости для разности событий имеет вид:

A	B	$A - B$
1	1	0
1	0	1
0	1	0
0	0	0

Можно рассмотреть разность $\Omega - A$. Это будет новое событие, состоящее в том, что A не наступило. Такое событие называют ***противоположным событию*** A и обозначают $\overline{A}$ (рис. 1.7).

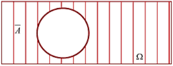


Рис. 1.7. Противоположные события

Таблица значимости для противоположного события имеет вид:

A	$\overline{A}$
1	0
0	1

Произведением AB двух событий называют новое событие, состоящее в том, что наступило ***каждое*** из событий A или B (рис. 1.8).

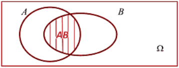


Рис. 1.8. Произведение событий A и B

Таблица значимости для произведения событий имеет вид:

A	B	AB
1	1	1
1	0	0
0	1	0
0	0	0

Отметим, что определение суммы и произведения событий обобщаются на любое количество компонент. Так, суммой событий $A_1, A_2, \ldots, A_n$ называют новое событие $A_1 + A_2 + \ldots + A_n$, состоящее в том, что наступило ***хотя бы одно*** из событий $A_1, A_2, \ldots, A_n$. Аналогично произведением событий $A_1, A_2, \ldots, A_n$ называют новое событие $A_1 \cdot A_2 \cdot \ldots \cdot A_n$, состоящее в том, что наступило ***каждое*** из событий $A_1, A_2, \ldots, A_n$.

Для рассмотренных операций выполнимы следующие равенства:

1. $A + B = B + A$.
2. $A + (B + C) = (A + B) + C$.
3. $A + \varnothing = A$.
4. $A + \Omega = \Omega$.
5. $AB = BA$.
6. $A \cdot (BC) = (AB) \cdot C$.
7. $A \cdot \varnothing = \varnothing$.
8. $A \cdot \Omega = A$.
9. $A + BC = (A + B) \cdot (A + C)$.
10. $A \cdot (B + C) = AB + AC$.
11. $A - BC = (A - B) + (A - C)$.
12. $A - (B + C) = (A - B) \cdot (A - C)$.
13. $A - A = \varnothing$.
14. $A + \overline{A} = \Omega$.
15. $A \cdot \overline{A} = \varnothing$.
16. $\overline{A + B} = \overline{A} \cdot \overline{B}$.
17. $\overline{A \cdot B} = \overline{A} + \overline{B}$.

Доказательство равенств можно проводить, используя как диаграммы, так и таблицы значимости.

Пример 1.1. Доказать равенство $A - (B + C) = (A - B) \cdot (A - C)$.

Решение.

Первый способ использует диаграммы.

Отметим штриховкой области наступления событий из левой и правой частей доказываемого равенства, при этом будем считать, что исходные события A, B и C совместны в совокупности (рис. 1.9):

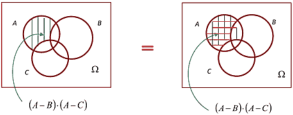


Рис. 1.9. Области наступления событий $A-(B+C)$ и $(A-B)\cdot(A-C)$

Отмеченные области одинаковы, что доказывает выполнимость заданного равенства.

Второй способ использует таблицы значимости.

Составим таблицы значимости для событий из левой и правой частей доказываемого равенства:

A	B	C	$B+C$	$A-(B+C)$	$A-B$	$A-C$	$(A-B)\cdot(A-C)$
1	1	1	1	0	0	0	0
1	0	1	1	0	1	0	0
0	1	1	1	0	0	0	0
0	0	1	1	0	0	0	0
1	1	0	1	0	0	1	0
1	0	0	0	1	1	1	1
0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0

Сравнивая построчно выделенные столбцы, приходим к выводу, что рассматриваемые события $A-(B+C)$ и $(A-B)\cdot(A-C)$ наступают и не наступают в одних и тех же случаях, следовательно, являются равными.

Пример 1.2. Проверить выполнимость равенства

$$AB+C=(B+C)-\overline{A}\cdot B\cdot\overline{C}$$

при условиях того, что: A благоприятно для B, A совместно с C, B совместно с C.

Решение.
Мы проведем доказательство с помощью таблиц значимости. Но перед заполнением таблицы представим, как дополнительные условия определяют положение областей наступления событий. Поэтому начнем с построения диаграммы (рис. 1.10).

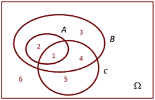


Рис. 1.10. Диаграмма к примеру 1.2

Таким образом, при составлении таблиц значимости мы должны учесть тот факт, что наступление события A обязательно определяет и наступление события B, т.е. ситуации, когда A наступает, а B не наступает, мы из таблицы должны исключить:

	A	B	C	AB	$AB+C$	$B+C$	$\overline{A}$	$\overline{C}$	$\overline{A}\cdot B\cdot\overline{C}$	$(B+C)-\overline{A}\cdot B\cdot\overline{C}$
1	1	1	1	1	1	1	0	0	0	1
2	1	1	0	1	1	1	0	1	0	1
3	0	1	0	0	0	1	1	1	1	0
4	0	1	1	0	1	1	1	0	0	1
5	0	0	1	0	1	1	1	0	0	1
6	0	0	0	0	0	0	1	1	0	0

В первом столбике указаны номера областей, отмеченных на диаграмме.

Сравнивая построчно выделенные столбцы, приходим к выводу, что рассматриваемые события $AB+C$ и $(B+C)-\overline{A}\cdot B\cdot\overline{C}$ наступают и не наступают в одних и тех же случаях, следовательно, являются равными.

Задачи для самостоятельного решения

В задачах **1.1–1.10** для любых событий A, B и C, совместных в совокупности, необходимо доказать равенство.

1.1. $A+BC=(A+B)\cdot(A+C)$.

1.2. $AB + C = (A + C)\cdot(B + C)$.

1.3. $A\cdot(B + C) = AB + AC$.

1.4. $(A + B)\cdot C = AC + BC$.

1.5. $\overline{A + B} = \overline{A}\cdot\overline{B}$.

1.6. $\overline{AB} = \overline{A} + \overline{B}$.

1.7. $A - BC = (A - B) + (A - C)$.

1.8. $AB - C = (A - C)\cdot(B - C)$.

1.9. $(A + B) - C = (A - C) + (B - C)$.

1.10. $\overline{A}\cdot\overline{BC} = \left(\overline{A + B}\right) + \left(\overline{A + C}\right)$.

В задачах **1.11–1.20** необходимо проверить выполнимость равенства при дополнительных условиях на события A, B и C:

1.11. $C - B = (C - A)\cdot\overline{B}$ при условии, что A благоприятствует B и B благоприятствует C.

1.12. $A + B = C\cdot(A + B)$ при условии, что A благоприятствует C, B благоприятствует C, а события A и B совместны.

1.13. $(A + C)\cdot B = B$ при условии, что B благоприятствует A, событие C совместно с A и несовместно с B.

1.14. $(B + C)\cdot A = A$ при условии, что A благоприятствует B, C благоприятствует B, а события A и C несовместны.

1.15. $A + B + C = B$ при условии, что A благоприятствует B, C благоприятствует B, и события A и C совместны.

1.16. $A + B + C = B + C$ при условии, что A благоприятствует B, а C не совместно с B.

1.17. $A\cdot(B + C) = B + C$ при условии, что B благоприятствует A, C благоприятствует A и события B и C несовместны.

1.18. $A + B = A + BC$ при условии, что B благоприятствует C, а событие A совместно с событием C, но несовместно с B.

1.19. $(C + B) - A = B - A$ при условии, что C благоприятствует B, а событие A совместно и с событием B, и с событием C.

1.20. $(A+B)\cdot(B+C)=B$ при условии, что A и C — несовместны, а событие B совместно и с событием A, и с событием C.

1.2. Классическое определение вероятности

Пусть некоторый опыт имеет n равновозможных исходов, k из которых благоприятствуют наступлению события A. Классической вероятностью события A называют отношение числа благоприятных для A исходов к общему числу исходов испытания:

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

Если событие достоверное, то его вероятность равна 1.

Для невозможного события вероятность равна 0.

Если же событие A случайное, то его вероятность подчиняется условию

$$0<P(A)<1.$$

При вычислениях значений n и k нам будут полезны некоторые понятия ***комбинаторики*** и вычислительные формулы, которые мы приведем ниже.

Комбинаторика изучает различные соединения элементов каких-либо множеств, а также отвечает на вопросы о том, сколько различных соединений, подчиненных определенным условиям, из этих элементов можно составить.

Если некоторое множество M состоит из n элементов (это записывают как $m(M)=n$), то один элемент из этого множества можно выбрать n способами. Существуют два основных комбинаторных правила.

Правило суммы. Если один элемент из множества A можно выбрать $m(A)$ способами, один элемент из множества B можно выбрать $m(B)$ способами, то один элемент из объединения $A\bigcup B$ множеств A и B можно выбрать $m(A)+m(B)$ способами при условии, что множества A и B не пересекаются: $A\bigcap B=\varnothing$.

Пример 1.3. Банк имеет 10 филиалов в европейской части России и 15 филиалов в азиатской части России. Правление банка принимает решение о проверке работы всех своих филиалов. Сколькими различными способами можно выбрать один филиал для проверки с участием председателя правления банка?

Решение.

Множество европейских филиалов обозначим A, тогда $m(A) = 10$, множество азиатских филиалов обозначим B, тогда $m(B) = 15$. Понятно, что условие $A \bigcap B = \varnothing$ выполнено.

Существует 25 различных вариантов выбора филиала банка для проверки с участием председателя правления банка.

Ответ: 25.

Правило суммы применяется в тех случаях, когда речь идет о выборе *одного элемента* из какой-либо совокупности, представленной объединением непересекающихся множеств, поэтому его можно обобщать на любое количество таких множеств.

В сложных задачах правило суммы применяется при необходимости подсчета количества независимых исходов, связанных с составлением каких-либо соединений элементов с заданными свойствами (если проведена детализация определенного свойства).

Правило произведения. Если элемент x из множества A можно выбрать $m(A)$ способами, элемент y из множества B можно выбрать $m(B)$ способами, то упорядоченную пару вида $(x; y)$ можно выбрать $m(A) \cdot m(B)$ способами.

Напомним, что набор элементов считается ***упорядоченным***, если элементы выбираются из различных множеств; если набор формируется из элементов одного и того же множества, то упорядоченность понимается в привычном смысле: важно, какой элемент какое место по порядку в наборе занимает. Во втором случае, оценивая возможности выбора какого-либо элемента в наборе, считают, что все предыдущие элементы уже выбраны, и оценку ведут на основе оставшихся вариантов.

Пример 1.4. Банк имеет 10 филиалов в европейской части России и 15 филиалов в азиатской части России. Правление банка принимает решение о проверке работы всех своих филиалов. Председатель правления банка хочет принять участие в проверке только двух филиалов банков — одного европейского и одного азиатского. Сколькими различными способами можно выбрать такую пару филиалов для проверки с участием председателя правления банка?

Решение.

Множество европейских филиалов обозначим A, тогда $m(A)=10$, множество азиатских филиалов обозначим B, тогда $m(B)=15$.

Европейский филиал x из множества A можно выбрать 10 способами, азиатский филиал y из множества B можно выбрать 15 способами, пару филиалов $(x; y)$ для проверки с участием председателя правления банка можно выбрать $m(A)\cdot m(B)=10\cdot 15=150$ способами.

Ответ: 150.

Пример 1.5. Банк имеет 10 филиалов в европейской части России и 15 филиалов в азиатской части России. Правление банка принимает решение о проверке работы всех своих филиалов. Председатель правления банка хочет принять участие в проверке только двух европейских филиалов банков в течение одной командировки. Сколькими различными способами может быть выбрана пара филиалов, если для председателя банка важно, какой из филиалов будет первым при посещении, а какой — вторым?

Решение.

Множество европейских филиалов обозначим A, тогда $m(A)=10$, Европейский филиал x для посещения первым из множества A можно выбрать 10 способами, филиал y для посещения вторым тоже выбираем из множества A, но уже при условии того, что филиал для x выбран, то есть выбор y осуществляется из оставшихся $10-1=9$ элементов. Пару филиалов $(x; y)$ для проверки с участием председателя правления банка рассматриваем как упорядоченную: председателю важно, какой филиал он посещает первым, а какой — вторым. Поэтому для ответа на вопрос задачи применяем правило произведения: $10\cdot 9=90$. Пара филиалов для проверки с участием председателя правления банка с учетом выдвинутых председателем условий может быть выбрана 90 способами.

Ответ: 90.

Правило произведения обобщается на выбор упорядоченных наборов любой длины.

Теперь определим основные понятия комбинаторики и приведём соответствующие примеры.

Размещением из n элементов по k называют любой упорядоченный набор длины k, составленный из элементов n-элементного множества. Если все элементы в наборе различны, то размещение определяется как ***размещение без повторений***, если же среди элементов набора могут быть одинаковые, то размещение определяется как ***размещение с по-***

вторениями. Число всех размещений вычисляется по правилу произведения.

Для размещений без повторений имеет место формула:

$$A_n^k = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1).$$

Пример 1.6. Из 10 членов правления необходимо выбрать председателя и его первого и второго заместителей. Сколькими различными способами это можно сделать?

Решение.

Председателя, его первого заместителя и второго заместителя можно рассматривать как упорядоченный набор длины 3, составленный из элементов 10-элементного множества, при этом элементы в наборе не могут быть одинаковыми (не может председатель быть одновременно и заместителем), поэтому для ответа на вопрос задачи необходимо найти число всех размещений без повторений из 10 элементов по 3:

$A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$.

Ответ: 720.

Для размещений с повторениями имеет место формула:

$$\overline{A_n^k} = n^k.$$

Пример 1.7. Десяти членам правления предложена анкета, в которой только два варианта ответа: «поддерживаю существующую стратегию развития предприятия» и «стратегия развития требует изменений». Сколько существует различных вариантов заполнения такой анкеты десятью членами правления?

Решение.

У каждого члена правления всего две возможности выбора ответа, при этом выбранные разными членами ответы могут быть одинаковыми (например, все поддерживают выбранную стратегию). Поэтому варианты заполнения анкеты десятью членами правления можно рассматривать как упорядоченный набор длины 10, составленный из элементов двухэлементного множества, при этом элементы, входящие в набор, могут быть одинаковыми. Значит, для решения задачи будем использовать формулу размещений с повторениями:

$\overline{A_2^{10}} = 2^{10} = 1024$.

Ответ: 1024.

Перестановкой из n элементов без повторений называют любое размещение из n элементов по n без повторений.

Число всех перестановок из n элементов без повторений находят по формуле:

$$P_n = n!$$

Выражение $n!$ читается как «n — факториал» и определяется как произведение натуральных чисел от 1 до n включительно:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n\,.$$

По определению полагают

$$1! = 1 \text{ и } 0! = 1\,.$$

Известны «оценочные» формулы для $n!$:

$$n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

при $n \to \infty$ (*формула Стирлинга*) и

$$e^{\frac{1}{12n+1}} < \frac{n!}{\sqrt{2pn} \cdot n^n \cdot e^{-n}} < e^{\frac{1}{12n}}\,.$$

Пример 1.8. В начале рабочего дня председатель правления изучает новости по пяти экономическим изданиям. Сколько существует различных вариантов для такого начала рабочего дня?

Решение.

В данном случае важно, в какой последовательности председатель правления будет выбирать экономические издания (например, если сначала он изучит материалы в газете «Коммерсантъ», а затем — в «РБК», то это один способ начать рабочий день, если же сначала он будет работать с «РБК», а затем с газетой «Коммерсантъ», то это уже другой способ для начала рабочего дня). Поэтому для решения задачи применяют формулу перестановок без повторений из пяти элементов:

$P_5 = 5! = 120\,.$

Ответ: 120.

Если в перестановке из n элементов есть одинаковые, то она определяется как ***перестановка с повторениями***.

Число всех перестановок с повторениями вычисляют по формуле:

$$\overline{P_n}\left(n_1, n_2, \ldots, n_k\right) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}\,,$$

где n определяет длину перестановки, n_1 определяет, сколько раз элемент a_1 встречается в перестановке, n_2 определяет, сколько раз элемент a_2 встречается в перестановке, ..., n_k определяет, сколько раз элемент a_k встречается в перестановке. Понятно, что $n_1 + n_2 + ... + n_k = n$. .

Пример 1.9. Сколько существует различных способов перестановки букв в слове «экономика»?

Решение.

Так как в слове «экономика» есть одинаковые буквы, для решения задачи применим формулу перестановок с повторениями, при этом n определяет длину перестановки и равно 9, $n_1 = 1$ (определяет, сколько раз элемент a_1 — буква «э» встречается в перестановке), $n_2 = 2$ (определяет, сколько раз элемент a_2 — буква «к» встречается в перестановке), $n_3 = 2$ (определяет, сколько раз элемент a_3 — буква «о» встречается в перестановке), $n_4 = 1$ (определяет, сколько раз элемент a_4 — буква «н» встречается в перестановке), $n_5 = 1$ (определяет, сколько раз элемент a_5 — буква «м» встречается в перестановке), $n_6 = 1$ (определяет, сколько раз элемент a_6 — буква «и» встречается в перестановке), $n_7 = 1$ (определяет, сколько раз элемент a_7 — буква «а» встречается в перестановке):

$$\overline{P_9}(1,2,2,1,1,1,1) = \frac{9!}{1!\cdot 2!\cdot 2!\cdot 1!\cdot 1!\cdot 1!\cdot 1!} = 90720.$$

Ответ: 90 720.

Если в составляемых наборах порядок следования элементов неважен, то мы имеем дело с сочетаниями.

Сочетанием из n элементов по k без повторений называют любое k-элементное подмножество n-элементного множества.

При формировании подмножества порядок выбора элементов неважен.

Число всех сочетаний из n элементов по k без повторений вычисляют по формуле:

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

Пример 1.10. Из десяти сотрудников отдела троих необходимо отправить в командировку. Сколькими различными способами это можно сделать?

Решение.

Так как при выборе командируемых формируются просто подмножества (а не упорядоченные наборы элементов, когда важно, кто первым будет выбран для поездки, кто — вторым, а кто — третьим), ответ на вопрос задачи находят, применяя формулу сочетаний без повторений:

$$C_{10}^{3} = \frac{10!}{3! \cdot (10-3)!} = \frac{10!}{3! \cdot 7!} = \frac{8 \cdot 9 \cdot 10}{6} = 120 .$$

Ответ: 120.

Из свойств сочетаний без повторений наиболее важными являются следующие:

1) $C_n^k = C_n^{n-k}$;

2) $C_n^k = C_{n-1}^k + C_{n-1}^{k-1}$;

3) $C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n = 2^n$;

4) $(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + C_n^2 a^{n-2} b^2 + \ldots + C_n^{n-1} a^1 b^{n-1} + C_n^n a^0 b^n$.

Если в выбираемых соединениях порядок следования элементов неважен, но могут встречаться одинаковые элементы, то в таком случае говорят о ***сочетаниях из n элементов по k с повторениями***.

Формула для вычисления числа всех сочетаний из n элементов по k с повторениями имеет вид:

$$\overline{C_n^k} = C_{n+k-1}^k .$$

Пример 1.11. На бирже торгуют акциями десяти различных предприятий. Сколькими способами можно осуществить покупку 3 акций?

Решение.

Так как при покупке акций их порядок неважен (неважно, какую акцию покупают первой, какую — второй, какую — третьей), речь идет о сочетаниях. Поскольку можно купить все три акции одного предприятия, следовательно, мы имеем дело с сочетаниями с повторениями. Поэтому при решении задачи используется формула сочетаний с повторениями:

$$\overline{C_{10}^3} = C_{10+3-1}^3 = C_{12}^3 = \frac{12!}{3! \cdot 9!} = \frac{10 \cdot 11 \cdot 12}{6} = 220 .$$

Ответ: 220.

При решении более сложных задач важно уметь выделить простые составляющие и использовать правила суммы и произведения.

Теперь приведем пример решения задачи на вычисление вероятности события на основе формулы классической вероятности.

Пример 1.12. В коробке находятся три белых и семь черных шаров. Наугад из коробки извлекают три шара. Какова вероятность того, что среди извлеченных шаров будет только один белый?

Решение.

Рассмотрим событие A: «Из трех извлеченных из коробки шаров — только один белый» и найдем его вероятность, используя формулу классической вероятности. Для этого необходимо знать, сколькими способами может закончиться опыт и в скольких случаях наступит событие A.

Всего в коробке находится десять шаров. Опыт состоит в том, что из этих десяти шаров извлекают три. Количество исходов опыта находим по формуле сочетаний:

$$n = C_{10}^{3} = \frac{10!}{3! \cdot 7!} = \frac{8 \cdot 9 \cdot 10}{6} = 120 \,.$$

Благоприятные исходы связаны с наступлением события A, значит, необходимо знать, сколькими способами формируется выборка из одного белого и двух черных шаров. Понятно, что белые шары извлекаются из белых, а черные — из черных, затем полученные значения перемножаются:

$$k = C_{3}^{1} \cdot C_{7}^{2} = \frac{3!}{1! \cdot 2!} \cdot \frac{7!}{2! \cdot 5!} = 3 \cdot 21 = 63.$$

Теперь можно вычислить вероятность события A:

$$P(A) = \frac{k}{n} = \frac{63}{120} = \frac{21}{40} = 0{,}525.$$

Ответ: 0,525.

Задачи для самостоятельного решения

В задачах **1.21–1.30** при решении следует использовать формулу классической вероятности.

1.21. Из 50 деталей пять имеют дефект. Наугад выбирают три детали. Какова вероятность того, что среди выбранных деталей только одна будет иметь дефект?

1.22. Из 20 предприятий региональный банк кредитовал восемь предприятий. Пять предприятий объявили о банкротстве. Какова вероятность того, что все они брали кредиты в региональном банке?

1.23. Из всех возможных двузначных чисел наугад выбирают три числа. Какова вероятность того, что среди выбранных будет только одно четное число?

1.24. Какова вероятность того, что при случайной перестановке букв слова ВЕКТОР первые две буквы будут гласными?

1.25. Какова вероятность того, что при случайной перестановке букв слова МАТЕМАТИКА все буквы А окажутся рядом?

1.26. Группу из 20 мужчин и 10 женщин произвольным образом делят на две подгруппы по 15 человек в каждой. Какова вероятность того, что все женщины окажутся в одной подгруппе?

1.27. Из коробки, содержащей пять белых и десять черных шаров, случайным образом извлекают три шара. Какова вероятность того, что среди извлеченных будет не более двух белых?

1.28. Из множества всех возможных четырехзначных чисел, составленных из цифр 1, 2, 3, 4, 5, 6, 7, 8, 9, наугад выбирают одно. Какова вероятность того, что в выбранном числе две одинаковые цифры не стоят радом?

1.29. Какова вероятность того, что из всех возможных трехзначных чисел случайным образом будет выбрано три числа, каждое из которых делится на 10?

1.30. Из цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 составлены четырехзначные числа, все цифры в записи которых различны. Из этих чисел случайным образом выбирают четыре числа. Какова вероятность того, что среди выбранных чисел будет не более двух четных чисел?

1.3. Геометрическое определение вероятности

Если число исходов опыта несчетно, то при вычислении вероятности события применяют формулу геометрической вероятности:

$$P(A)=\frac{\text{мера } g}{\text{мера } G},$$

где G обозначает область осуществления испытания, а g — область наступления события A.

Пример 1.13. Внутри квадрата со стороной 10 см находится круг радиуса 4 см (рис. 1.11). Случайным образом с квадрат вбрасывается точка. Какова вероятность того, что точка не попадет в круг?

Решение.

Так как множество точек квадрата несчетно, при решении задачи будет применяться формула геометрической вероятности. Для этого необходимо вычислить меры (площади) области осуществления испытания и области наступления события A: «Точка, вброшенная в квадрат, не попадает в круг».

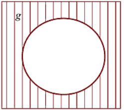


Рис. 1.11. Диаграмма к примеру 1.13

Мерой области осуществления испытания является площадь квадрата:

мера $G = 10^2 = 100$ см^2.

Область наступления события отмечена штриховкой, ее меру можно найти вычитанием из площади квадрата площадь круга:

мера $g = 100 - \pi \cdot 4^2 = 100 - 16\pi$ см^2.

Теперь может быть вычислена вероятность события A:

$$P(A) = \frac{\text{мера } g}{\text{мера } G} = \frac{100 - 16\pi}{100} = 1 - 0{,}16\pi \approx 1 - 0{,}503 = 0{,}497 .$$

Ответ: 0,497.

Задачи для самостоятельного решения

В задачах **1.31–1.40** при решении использовать формулу геометрической вероятности.

1.31. На отрезок длиной 10 см наугад вброшена точка. Какова вероятность того, что она окажется на далее 1 см от середины отрезка?

1.32. На отрезок длиной 10 см наугад вброшена точка. Какова вероятность того, что она окажется удаленной от концов отрезка больше, чем на 1 см?

1.33. Внутри квадрата со стороной 10 см расположен круг радиуса 2 см. Какова вероятность того, что точка, наугад вброшенная в квадрат, попадет в круг?

1.34. В круг радиуса 5 см вписан квадрат, в этот квадрат, в свою очередь, вписан круг. Какова вероятность того, что вброшенная наугад в большой круг точка попадет и в малый?

1.35. В квадрат со стороной 10 см вписан круг, а в круг вписан правильный треугольник. Случайным образом в квадрат вбрасывается точка. Какова вероятность того, что эта точка не попадет в треугольник?

1.36. Мишень образована пятью концентричными окружностями радиусов 2 см, 4 см, 6 см, 8 см и 10 см. Какова вероятность того, что при одном выстреле по мишени не попадут в самый маленький круг? Какова вероятность того, что местом попадания при одном выстреле будет внешнее кольцо (границами являются окружности радиусами 8 см и 10 см)?

1.37. На плоскости проведены параллельные прямые, расстояние между которыми равны 12 см. Какова вероятность того, что брошенный наугад на эту плоскость круг радиуса 5 см не будет пересечен ни одной из этих прямых.

1.38. На плоскости расположены параллельные прямые, расстояния между которыми чередуется и равно 1,5 см и 8 см. Какова вероятность того, что брошенный на эту плоскость круг радиуса 2,5 см не будет пересечен ни одной из этих прямых.

1.39. Два друга договорились о встрече в период от 10:00 до 10:15, пришедший на эту встречу первым ждет другого не более 5 минут, после чего уходит. Считая, что приход каждого из участников встречи возможен в любой момент указанного промежутка, найти вероятность того, что встреча состоится.

1.40. В квадратном уравнении $x^2 + px + q = 0$ коэффициенты p и q выбираются произвольным образом в интервале $(0;1)$. Какова вероятность того, что при выбранных коэффициентах данное уравнение будет иметь действительные корни?

1.4. Основные правила вычисления вероятностей

Для вычисления вероятностей суммы и произведения событий существуют следующие правила.

Правило 1. Если события *A* и *B* несовместны, то вероятность суммы событий *A* + *B* равна сумме вероятностей этих событий:

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

Это утверждение обобщается на любое количество попарно несовместных событий.

Правило 2.

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

если $A_i \cap A_j = \varnothing$ **при** $i \neq j$.

Пример 1.14. Из коробки, содержащей пять красных, три зеленых и два белых шара, извлекают один. Какова вероятность того, что извлеченный шар не будет белым?

Решение.

Событие из вопроса задачи «Извлеченный шар не будет белым» может быть заменено эквивалентным событием «Извлеченный шар — красный или зеленый», что представляет собой сумму попарно несовместных событий A: «Извлеченный шар — красный» и B: «Извлеченный шар зеленый», поэтому для ответа на вопрос задачи применим **Правило 1**: $P(A+B) = P(A) + P(B) = \frac{5}{10} + \frac{3}{10} = \frac{8}{10} = 0{,}8.$

Ответ: 0,8.

Поскольку события A и $\overline{A}$ попарно несовместны, а в сумме дают достоверное событие, то для них выполнимы следующие утверждения:

Правило 3. $P\left(A + \overline{A}\right) = 1$.

Правило 4. $P\left(A\right) + P\left(\overline{A}\right) = 1$.

Правило 5. $P\left(\overline{A}\right) = 1 - P\left(A\right)$.

В рассмотренном примере 1.14 можно ввести событие A: «Извлеченный шар белый», тогда событие из вопроса задачи «Извлеченный шар не будет белым» представляет собой противоположное событие $\overline{A}$, вероятность которого можно найти, используя **Правило 5**:

$$P\left(\overline{A}\right) = 1 - P\left(A\right) = 1 - \frac{2}{10} = \frac{8}{10} = 0{,}8.$$

Попарно несовместные события, сумма которых образует все пространство событий, называют ***полной группой испытания***.

Если события $H_1, H_2, \ldots, H_n$ образуют полную группу испытания, то для них всегда выполнены условия:

1) $H_i \cap H_j = \varnothing$ при $i \neq j$;

2) $H_1 + H_2 + \ldots + H_n = \Omega$;

3) $P\left(H_1\right) + P\left(H_2\right) + \ldots + P\left(H_n\right) = 1$.

Например, если опыт состоит в однократном подбрасывании кубика, то полная группа испытания может формироваться по-разному. Шесть событий, связанных с выпадением 1, 2, 3, 4, 5 или 6 очков образуют полную группу, как и всего для события «Выпало четное число очков» и «Выпало нечетное число очков». В первом случае события являются элементарными исходами опыта, во втором — составными событиями. Но и в первом, и во втором случае выполнены все условия, необходимые для формирования полной группы испытания: попарная несовместность событий (условие 1), в сумме дающих все пространство событий (условие 2), и сумма вероятностей которых равна 1 (условие 3).

Если вероятность события A вычисляется при условии наступления события B, то такую вероятность называют ***условной вероятностью*** и обозначают $P(A|B)$ или $P_B(A)$.

Правило 6. Вероятность произведения двух событий *A* и *B* вычисляют по формуле

$$P(AB)=P(A)\cdot P(B|A)\ =P(B)\cdot P(A|B).$$

Это правило обобщается на любое количество множителей.

Правило 7.

$$P(A_1\cdot A_2\cdot A_3\cdot\ldots\cdot A_{n-1}\cdot A_n)=$$
$$=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cdot A_2)\cdot\ldots\cdot P(A_n|A_1\cdot A_2\cdot\ldots\cdot A_{n-1}).$$

Пример 1.15. Из коробки, содержащей пять красных и три зеленых шара, последовательно без возвращения извлекают два шара. Какова вероятность того, что извлечены разноцветные шары?

Решение.

Можно рассмотреть событие из вопроса задачи A «Извлечены разноцветные шары» как результат действий с более простыми событиями, моделирующими опыт, связанный с формированием выборки без возвращения:

$$A=I_к\cdot II_з+I_з\cdot II_к\ ,$$

где, например, $I_к$ определяет событие «Первым извлекли красный шар». При решении задачи надо учитывать, что при извлечении второго шара общее количество шаров в коробке уменьшается на один (первый извлеченный шар в коробку не возвращается):

$$P(A)=P(I_к\cdot II_з+I_з\cdot II_к)=\frac{5}{8}\cdot\frac{3}{7}+\frac{3}{8}\cdot\frac{5}{7}=\frac{30}{56}=\frac{15}{28}=0{,}5357.$$

Ответ: 0,5357.

Задача примера 1.15 могла быть решена и с применением формулы классической вероятности, однако в приведенном примере показано, какое влияние на решение оказывает условие формирования выборки. Мы должны прийти к выводу, что выборка без возвращения формирует зависимые события и при решении задачи в таких условиях необходимо использовать условные вероятности.

Если условные и безусловные вероятности событий совпадают, то события определяются как ***независимые***:

$$P(A)=P(A|B) \text{ и } P(B)=P(B|A).$$

Можно сказать, что для независимых событий наступление одного события не изменяет вероятности наступления другого.

Правила 6 и **7** для ***независимых*** событий примут вид:

Правило 8. $P(AB)=P(A)\cdot P(B)$.

Правило 9. $P(A_1\cdot A_2\cdot\ldots\cdot A_n)=P(A_1)\cdot P(A_2)\cdot\ldots\cdot P(A_n)$.

В последнем случае события должны быть независимыми в совокупности: наступление какого-либо из них не изменяет вероятности наступления оставшихся событий, а также любых их произведений.

Пример 1.16. Предприятие оформило кредиты в трех разных банках, оценив вероятность возврата кредита первому банку в 0,8, второму — 0,9 и третьему — 0,7. Какова вероятность того, что предприятие вернет все взятые кредиты? Какова вероятность того, что предприятие вернет только один из взятых кредитов?

Решение.

В условии задачи нет указаний на то, что вероятность возврата кредита одному банку как-то связана с вероятностью возврата кредита другому банку, т.е. события

A_1: «Предприятие вернуло кредит первому банку»,

A_2: «Предприятие вернуло кредит второму банку»,

A_3: «Предприятие вернуло кредит третьему банку»

являются независимыми. На основе этих событий можно описывать события из вопроса задачи.

Событие *А* «Предприятие вернет все взятые кредиты» равно произведению событий A_1, A_2, A_3. Поэтому вероятность события *А* вычислим с применением **Правила 9**:

$$P(A)=P(A_1\cdot A_2\cdot A_3)=P(A_1)\cdot P(A_2)\cdot P(A_3)=0{,}8\cdot 0{,}9\cdot 0{,}7=0{,}504.$$

Событие B: «Предприятие вернет только один из взятых кредитов» может быть представлено как

$$B = A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3.$$

При вычислении вероятности события B применяем **Правила 2**, **5** и **9**:

$$P(B) = P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3\right) =$$

$$= P\left(A_1 \cdot \overline{A_2} \cdot \overline{A_3}\right) + P\left(\overline{A_1} \cdot A_2 \cdot \overline{A_3}\right) + P\left(\overline{A_1} \cdot \overline{A_2} \cdot A_3\right) =$$

$$= 0{,}8 \cdot (1-0{,}9) \cdot (1-0{,}7) + (1-0{,}8) \cdot 0{,}9 \cdot (1-0{,}7) + (1-0{,}8) \cdot (1-0{,}9) \cdot 0{,}7 =$$

$$= 0{,}8 \cdot 0{,}1 \cdot 0{,}3 + 0{,}2 \cdot 0{,}9 \cdot 0{,}3 + 0{,}2 \cdot 0{,}1 \cdot 0{,}7 = 0{,}092.$$

Ответ: 0,504; 0,092.

Если событие A может наступить только с одним из событий $H_1, H_2, ..., H_n$, образующих полную группу испытания, то вероятность наступления события A вычисляют по ***формуле полной вероятности***:

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) + ... + P(H_n) \cdot P(A|H_n).$$

События $H_1, H_2, ..., H_n$ принято называть априорными гипотезами.

Если событие A уже наступило, то становится понятным, вместе с каким из событий $H_1, H_2, ..., H_n$ это произошло. В этом случае соответствующая гипотеза, обеспечившая наступление события A, называется апостериорной и ее вероятность может быть переоценена по ***формуле Байеса***:

$$P(H_i|A) = \frac{P(H_i) \cdot P(A|H_i)}{P(A)}.$$

В знаменателе — полная вероятность события A.

Пример 1.17. Три эксперта составили списки предприятий с нарушениями в финансовой отчетности. В списке, составленном первым экспертом, четыре предприятия, вторым — пять предприятий и третьим — два предприятия. Для проведения повторной проверки наугад выбирается список и из него наугад выбирается предприятие. Какова вероятность того, что предприятие для повторной проверки было выбрано из списка, составленного третьим экспертом?

Решение.

Так как список выбирается наугад, необходимо выдвинуть гипотезы:

H_1: «Выбран список, составленный первым экспертом»,

H_2: «Выбран список, составленный вторым экспертом»,

H_3: «Выбран список, составленный третьим экспертом»,

которые равновероятны: $P(H_1)=P(H_2)=P(H_3)=\dfrac{1}{3}$, и после этого рассуждать о вероятности выбора одного предприятия из соответствующего списка. Тогда вероятность события ***A*** «Для повторной проверки будет выбрано одно предприятие» вычисляется по формуле полной вероятности:

$$P(A)=P(H_1)\cdot P(A|H_1)+P(H_2)\cdot P(A|H_2)+(H_3)\cdot P(A|H_3)=$$

$$=\frac{1}{3}\cdot\frac{1}{4}+\frac{1}{3}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{1}{2}=\frac{5\cdot 2+4\cdot 2+4\cdot 5}{3\cdot 4\cdot 5\cdot 2}=\frac{38}{120}=\frac{19}{60}=0{,}3167.$$

Для ответа на вопрос задачи необходимо переоценить вероятность третьей гипотезы. Для этого используем формулу Байеса:

$$P(H_3|A)=\frac{P(H_3)\cdot P(A|H_3)}{P(A)}=\frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{38}{120}}=\frac{120}{6\cdot 38}=\frac{120}{228}=\frac{10}{19}=0{,}5263.$$

Ответ: 0,5263.

Пример 1.18. В трех списках, представленных экспертами, содержится перечень проверенных предприятий. В списке из десяти предприятий, проверенных первым экспертом, два имеют нарушения в финансовой отчетности, во втором списке из восьми предприятий нарушения отчетности отмечены у трех предприятий, а из 12 предприятий, проверенных третьим экспертом, нарушения выявлены у пяти предприятий. Из каждого списка наугад выбирают одно предприятие и затем из них выбирают одно для повторной проверки. Какова вероятность того, что в результате будет выбрано предприятие, не имеющее нарушений в финансовой отчетности?

Решение.

Так как неизвестно, какая тройка предприятий по одному из каждого списка будет выбрана, необходимо выдвижение следующих гипотез:

H_1: «Все три выбранных из соответствующих списков предприятия имеют нарушения в финансовой отчетности».

H_2: «Все три выбранных из соответствующих списков предприятия не имеют нарушений в финансовой отчетности».

H_3: «Только одно из трех выбранных из соответствующих списков предприятий имеет нарушения в финансовой отчетности».

H_4: «Только одно из трех выбранных из соответствующих списков предприятий не имеет нарушений в финансовой отчетности».

Вероятности гипотез вычисляем с применением формулы классической вероятности, а также **Правил 2**, **5** и **9**:

$$P(H_1)=\frac{2}{10}\cdot\frac{3}{8}\cdot\frac{5}{12}=\frac{30}{960}=\frac{1}{32},\quad P(H_2)=\frac{8}{10}\cdot\frac{5}{8}\cdot\frac{7}{12}=\frac{280}{960}=\frac{7}{24},$$

$$P(H_3)=\frac{2}{10}\cdot\frac{5}{8}\cdot\frac{7}{12}+\frac{8}{10}\cdot\frac{3}{8}\cdot\frac{7}{12}+\frac{8}{10}\cdot\frac{5}{8}\cdot\frac{5}{12}=\frac{438}{960}=\frac{73}{160},$$

$$P(H_4)=\frac{8}{10}\cdot\frac{3}{8}\cdot\frac{5}{12}+\frac{2}{10}\cdot\frac{5}{8}\cdot\frac{5}{12}+\frac{2}{10}\cdot\frac{3}{8}\cdot\frac{7}{12}=\frac{212}{960}=\frac{53}{240}.$$

Вероятность события A: «Предприятие, выбранное одно из трех, не будет иметь нарушений в финансовой отчетности» вычисляем по формуле полной вероятности в присутствии четырех гипотез:

$$P(A)=P(H_1)\cdot P(A|H_1)+P(H_2)\cdot P(A|H_2)+(H_3)\cdot P(A|H_3)+P(H_4)\cdot P(A|H_4)=$$

$$=\frac{1}{32}\cdot\frac{0}{3}+\frac{7}{24}\cdot\frac{3}{3}+\frac{73}{160}\cdot\frac{2}{3}+\frac{53}{240}\cdot\frac{1}{3}=\frac{0+840+876+212}{2880}=$$

$$=\frac{1928}{2880}=\frac{241}{360}=0{,}6694.$$

Ответ: 0,6694.

Задачи для самостоятельного решения

В задачах **1.41–1.60** при решении использовать формулы для вычисления вероятностей суммы или произведения событий.

1.41. Вероятность успешного завершения первого проекта равна 0,8, второго — 0,9 и третьего — 0,95. Найти вероятность того, что из трех проектов успешно завершится только один.

1.42. Чтобы принять партию готовых изделий, используют выборочный контроль: берут пять изделий и если среди них окажется хотя бы одно бракованное, то бракуется и вся партия. Вероятность выпуска бракованного изделия равна 0,01. Найти вероятность того, что проверяемая партия изделий будет забракована.

1.43. Имеются три коробки, в каждой из которых находятся по три белых и два черных шара. Из каждой коробки наугад выбирают по одному шару. Какова вероятность того, что среди выбранных шаров будет хотя бы один белый шар?

1.44. Вероятность попадания по мишени при одном выстреле равна 0,7. Сколько раз надо выстрелить по мишени, чтобы вероятность хотя бы одного попадания была $\geq 0{,}9$?

1.45. Вероятность хотя бы одного попадания по мишени при 10 выстрелах равна 0,7. Какова вероятность попадания по мишени при одном выстреле, если она не меняется от выстрела к выстрелу?

1.46. В коробке находятся один белый и пять черных шаров. Два игрока по очереди извлекают из коробки по одному шару. Выигрывает тот, кто первым извлечет белый шар. Какова вероятность выигрыша того игрока, который начинает игру первым? При решении задачи рассмотреть случай, когда извлеченный игроком шар возвращается в коробку, и случай, когда извлеченный шар в коробку не возвращается.

1.47. Игра продолжается до тех пор, пока на заданный вопрос дают правильный ответ. Вероятность правильного ответа на один вопрос равна 0,1 и не меняется от вопроса к вопросу. Какова вероятность того, что в процессе игры будет задано не более десяти вопросов?

1.48. На шахматную доску наугад ставят две ладьи. Вычислить вероятность того, что ладьи побьют друг друга при условии, что их поставили на клетки разного цвета.

1.49. Из натуральных чисел от 1 до 100 случайным образом выбирают три числа. Какова вероятность того, что третье число попадет в интервал, образованный первым и вторым числом при условии, что первое число меньше второго?

1.50. Известно, что $P(A)=0,7$, $P(B)=0,6$, $P(AB)=0,5$. Найти $P(A+B)$ и $P(\overline{A}+\overline{B})$.

1.51. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0,4$, $P(B)=0,6$, $P(C)=0,5$. Найти $P(AB+\overline{C})$.

1.52. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0,8$, $P(B)=0,5$, $P(C)=0,4$. Найти $P((A+B)\cdot(A+C))$.

1.53. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0,3$, $P(B)=0,4$, $P(C)=0,5$. Найти $P((A+B+C)\cdot(\overline{A}+C))$.

1.54. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0,9$, $P(B)=0,5$, $P(C)=0,6$. Найти $P((AB+BC)\cdot(AC))$.

1.55. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0,7$, $P(B)=0,6$, $P(C)=0,5$. Найти $P((A\cdot\overline{B}+BC)\cdot(A\cdot\overline{C}))$.

1.56. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0{,}4$, $P(B)=0{,}6$, $P(C)=0{,}5$. Найти $P\left((AB+\overline{C})\middle|(AC)\right)$.

1.57. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0{,}8$, $P(B)=0{,}5$, $P(C)=0{,}4$. Найти $P\left((A+B)\middle|\ (A+C)\right)$.

1.58. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0{,}3$, $P(B)=0{,}4$, $P(C)=0{,}5$. Найти $P\left((A+B+C)\middle|\ (\overline{A}+C)\right)$.

1.59. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0{,}9$, $P(B)=0{,}5$, $P(C)=0{,}6$. Найти $P\left((AB+BC)\middle|\ (AC)\right)$.

1.60. Известно, что события *A*, *B* и *C* являются независимыми и $P(A)=0{,}7$, $P(B)=0{,}6$, $P(C)=0{,}5$. Найти $P\left((A\cdot\overline{B}+BC)\middle|\ (A\cdot\overline{C})\right)$.

В задачах **1.61–1.80** при решении использовать формулу полной вероятности или формулу Бейеса:

1.61. Имеются две коробки. В первой содержится два белых и пять черных шаров, во второй — три белых и один черный. Из первой коробки извлекают один шар и перекладывают во вторую, после чего из второй коробки извлекают один шар. Какова вероятность того, что извлеченный из второй коробки шар окажется черным?

1.62. Имеются три коробки. В первой находятся два белых и три черных шара, во второй — один белый и два черных, в третьей — одинбелый и три черных. Из первой коробки извлекли один шар и переложили во вторую, затем из второй коробки извлекли один шар и переложили в третью. После этого из третьей коробки извлекают один шар. Какова вероятность того, что этот извлеченный из третьей коробки шар окажется белым?

1.63. Имеются две коробки. В первой содержится пять белых и пять черных шаров, во второй — три белых и три черных. Из первой коробки извлекают два шара и перекладывают во вторую, после чего из второй коробки извлекают один шар. Какова вероятность того, что этот извлеченный из второй коробки шар окажется черным?

1.64. Имеются две коробки. В первой содержится 5 белых и 8 черных шаров, во второй — три белых и пять черных. Из первой коробки извлекают три шара и перекладывают во вторую, после чего из второй коробки извлекают один шар. Какова вероятность того, что этот извлеченный из второй коробки шар окажется белым?

1.65. Имеются три одинаковых коробки. В первой находятся два белых и три черных шара, во второй — один белый и два черных, в третьей — один белый и три черных. Наугад выбирают коробку и затем из выбранной коробки наугад извлекают один шар. Какова вероятность того, что этот шар окажется черным?

1.66. Имеются три одинаковых коробки. В первой находятся десять белых и два черных шара, во второй — один белый и 11 черных, в третьей — три белых и девять черных. Из каждой коробки наугад выбирают по одному шару и затем из выбранных трех шаров наугад выбирают один. Какова вероятность того, что этот выбранный шар окажется белым?

1.67. Имеются три одинаковых коробки. В первой находятся десяь белых и два черных шара, во второй — один белый и 11 черных, в третьей — три белых и девять черных. Из каждой коробки наугад выбирают по два шара и затем из выбранных шести шаров наугад выбирают два. Какова вероятность того, что эти два выбранных шара окажутся разноцветными?

1.68. В коробку, содержащую пять шаров, вбросили один белый шар, после чего наугад извлекли один шар. Вычислить вероятность того, что извлеченный шар окажется белым, если равновероятны все возможные предположения о первоначальном количестве белых шаров в коробке.

1.69. Каждая буква слова ВЕКТОР написана на отдельной карточке. Две карточки потеряны. Из оставшихся карточек наугад извлекается одна. Какова вероятность того, что на извлеченной карточке будет написана гласная буква?

1.70. Имеется три одинаковых коробки. В первой находятся десять белых шаров, во второй — десять черных, в третьей — пять белых и пять черных. Коробку выбирают, подбрасывая кубик один раз. Если выпадет 6 очков, то выбирают первую коробку, если выпадет одно очко, то выбирают вторую коробку, во всех других случаях должна быть выбрана третья коробка. Из выбранной коробки извлекают один шар. Какова вероятность того, что он окажется белым?

1.71. Детали могут быть обработаны тремя автоматами. Первый автомат обрабатывает 30% выпускаемой продукции, второй — 40%, остальную часть обрабатывает третий автомат. Вероятность выпуска бракованной детали для первого автомата равна 0,01, для второго — 0,03 и для третьего — 0,01. Выбранная для контроля деталь оказалась бракованной. Какова вероятность того, что ее обрабатывали на втором автомате?

1.72. За потребительским кредитом клиент может обратиться в первый банк с вероятностью 0,8, а во второй банк — с вероятностью 0,2. Вероятность отказа в получении кредита в первом банке составляет 0,4, во втором — 0,7. Обратившись в банк, клиент получил кредит. Какова вероятность того, что этот кредит был получен во втором банке?

1.73. Из 25 студентов десять выучили все 50 вопросов к экзамену, десять студентов выучили 30 вопросов, а пять студентов выучили только десять вопросов. Выбранный случайным образом студент на экзамене ответил на вопрос. Какова вероятность того, что этот студент выучил к экзамену только десять вопросов?

1.74. Из пяти монет одна является бракованной — на ней герб отпечатан с двух сторон. Наугад выбранную монету, не разглядывая, подбрасывают пять раз, и она пять раз падает гербом вверх. Какова вероятность того, что в руках у подбрасывающего была обычная, а не бракованная монета?

1.75. 95% выпускаемых изделий являются изделиями высшего сорта, остальные — первого сорта. Система контроля качества признает изделие высшего сорта изделием высшего сорта с вероятностью 0,99, а изделие первого сорта изделием высшего сорта с вероятностью 0,02. Изделие при проведении контроля было признано изделием высшего сорта. Какова вероятность того, что оно является изделием первого сорта?

1.76. На складе хранится продукция трех производителей в отношении 2:3:5. Процент брака в выпускаемой продукции у первого производителя составляет 5%, у второго — 2%, у третьего — 1%. Случайным образом выбранная со склада продукция оказалась без брака. Что более вероятно: эта продукция второго производителя или третьего?

1.77. Банк разделил своих клиентов в отношении 1:8. Вероятность нарушения сроков погашения кредита для клиентов первой группы равна 0,3, а для клиентов второй группы — 0,01. Случайным образом выбранный клиент нарушил срок погашения кредита. Какова вероятность того, что он из второй группы?

1.78. Вероятность попадания по мишени для каждого из пяти стрелков первой группы равна 0,6, а вероятность попадания по мишени для каждого из трех стрелков второй группы равна 0,9. Случайным образом выбранный стрелок попал по мишени. Какова вероятность того, что он из первой группы?

1.79. В группе пять отличников, 15 хорошо успевающих студентов и пять студентов, занимающихся слабо. На предстоящем экзамене отличники могут получать только оценку «отлично», хорошо успевающие студенты с равной вероятностью могут получать оценки «отлично» или «хорошо», а слабо успевающие студенты с равной вероятностью могут получать оценки «хорошо», «удовлетворительно» или «неудовлетворительно». Наугад выбранный студент получил на экзамене оценку «хорошо». Какова вероятность того, что он — слабо успевающий студент?

1.80. Каждое из трех любимых мест для рыбалки рыболов выбирает с равной вероятностью. На первом месте при однократном забрасывании удочки рыба клюет с вероятностью 0,6, на втором месте — с вероятностью 0,8, а на третьем — 0,5. Выйдя в очередной раз на рыбалку, рыболов трижды закидывал удочку, но рыба клюнула только один раз. Какова вероятность того, что он рыбачил на втором месте?

1.5. Повторные независимые испытания

Говорят, что серия из n испытаний проходит по ***схеме Бернулли***, если выполнены условия:

– эти испытания независимы;
– каждое из них имеет только два исхода (A наступило или A не наступило);
– вероятность наступления события A в каждом испытании одна и та же и равна $p = P(A)$.

Наступление события A принято называть «успехом». «Неуспех» связан с тем, что событие A не наступило, при этом $P(\overline{A}) = 1 - p = q$.

Если испытания проходят по схеме Бернулли, то вероятность k «успехов» при n испытаниях вычисляют по формуле Бернулли:

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k},$$

где $C_n^k = \dfrac{n!}{k! \cdot (n-k)!}$, $p = P(A)$, $q = 1 - p$.

Пример 1.19. Испытываются пять независимо работающих одинаковых прибора. Вероятность отказа в работе для каждого прибора равна 0,2. Найти вероятность того, что в результате проводящихся испытаний откажет только один прибор.

Решение.

Испытания проходят по схеме Бернулли: они независимы и вероятность отказа остается неизменной от опыта к опыту, поэтому ответ на вопрос задачи найдем по формуле Бернулли при $n = 5$, $k = 1$, $p = 0{,}2$ и $q = 1 - 0{,}2 = 0{,}8$:

$$P_5(1) = C_5^1 \cdot 0{,}2^1 \cdot 0{,}8^{5-1} = 5 \cdot 0{,}2 \cdot 0{,}4096 = 0{,}4096.$$

Ответ: 0,4096.

Понятно, что для различных значений k соответствующие вероятности $P_n(k)$, вообще говоря, различны. Однако наиболее вероятное количество «успехов» k_0 всегда подчиняется неравенству:

$$np + p - 1 \le k_0 \le np + p\ .$$

Пример 1.20. Вероятность наступления события A в каждом из девяти независимых испытаний равна 0,5. Найти наиболее вероятное число наступлений события A.

Решение.

Можно, используя формулу Бернулли, вычислить соответствующие вероятности $P_n(k)$ для $n = 9$ и значений k от 0 до 9, затем выбрать то k_0, которому отвечает наибольшая вероятность. Но удобнее сразу использовать оценочное неравенство:

$$9 \cdot 0{,}5 + 0{,}5 - 1 \le k_0 \le 9 \cdot 0{,}5 + 0{,}5\ ,\quad 4 \le k_0 \le 5\ .$$

Таким образом, наибольшая вероятность отвечает двум значениям $k_0 = 4$ и $k_0 = 5$, для всех других значений k соответствующая вероятность $P_9(k)$ будет меньше.

Ответ: $k_0 = 4$, $k_0 = 5$.

Можно заметить, что если верхняя граница $np + p$ в неравенстве получается дробным числом, то наиболее вероятное число «успехов» k_0 определится как единственное.

При больших значениях n вычисления по формуле Бернулли весьма затруднительны, поэтому для решения задач используют приближенные формулы. Различают локальную формулу Муавра — Лапласа, интегральную формулу Лапласа и приближенную формулу Пуассона.

Последнюю применяют в тех случаях, когда $n \to \infty$, $p \to 0$, т.е. при условии, когда «успех» представляет собой ***очень редкое событие***. В этом случае произведение np^2 должно быть значительно меньше 1. Однако при решении задач допустимо использовать оценку произведения npq, которое должно быть значительно меньше 10.

Локальная приближенная формула Муавра — Лапласа имеет вид:

$$P_n(k) \approx \frac{\phi(x)}{\sqrt{npq}},$$

где $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ — функция Гаусса, значения которой табулированы (Приложение 1) для различных значений аргумента

$$x = \frac{k - np}{\sqrt{npq}}.$$

Из полезных для решения задач отметим следующие свойства функции $\phi(x)$:

1) $\phi(-x) = \phi(x)$ — функция Гаусса является четной;

2) $\lim\limits_{x \to \pm\infty} \phi(x) = 0$.

Интегральная приближенная формула Лапласа имеет вид:

$$P_n(k_1 \le k \le k_2) \approx \Phi(x_2) - \Phi(x_1),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$ — функция Лапласа, значения которой табулированы (Приложение 2), а аргументы вычисляют по формулам:

$$x_2 = \frac{k_2 - np}{\sqrt{npq}}, \; x_1 = \frac{k_1 - np}{\sqrt{npq}}.$$

Из полезных для решения задач отметим следующие свойства функции $\Phi(x)$:

1) $\Phi(-x) = -\Phi(x)$ — функция Лапласа является нечетной;

2) $\lim\limits_{x \to +\infty} \Phi(x) = 0{,}5$, $\lim\limits_{x \to -\infty} \Phi(x) = -0{,}5$.

Приближенная формула Пуассона имеет вид:

$$P_n(k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda},$$

где $\lambda = np$ и определяет среднее число «успехов» в серии из n испытаний, проходящих по схеме Бернулли. Для вероятностей $P_n(k)$ при разных значениях λ и k также составлены таблицы (Приложение 3).

Пример 1.21. Известно, что среди выпускаемых изделий изделия первого сорта составляют 20%, остальные — высшего сорта. Какова вероятность того, что из 400 проверенных изделий 60 будут изделиями первого сорта?

Решение.

Вероятность события A «Проверяемое изделие окажется изделием первого сорта» ($P(A) = \frac{20}{100} = 0{,}2$) остается постоянной в серии из 400 независимых испытаний. Поэтому имеет место схема испытаний Бернулли. Однако n = 400 достаточно велико и использование формулы Бернулли для вычисления $P_{400}(60)$ затруднительно, поэтому необходимо применить одну из приближенных формул. Для выбора нужной формулы оценим произведение npq:

$$npq = 400 \cdot 0{,}2 \cdot (1 - 0{,}2) = 400 \cdot 0{,}2 \cdot 0{,}8 = 64 > 10.$$

Из полученной оценки понятно, что для решения задачи должна быть применена локальная приближенная формула Муавра — Лапласа:

$$P_n(k) \approx \frac{\phi(x)}{\sqrt{npq}},$$

при n = 400 и k = 60.

Вычислим аргумент для функции Гаусса:

$$x = \frac{k - np}{\sqrt{npq}} = \frac{60 - 400 \cdot 0{,}2}{\sqrt{400 \cdot 0{,}2 \cdot 0{,}8}} = \frac{60 - 80}{\sqrt{64}} = -\frac{20}{8} = -2{,}5.$$

Теперь найдем ответ на вопрос задачи:

$$P_{400}(60) \approx \frac{\phi(-2{,}5)}{\sqrt{64}} = \frac{\phi(2{,}5)}{8} = \frac{0{,}0175}{8} = 0{,}00219.$$

При вычислениях использовано свойство четности функции Гаусса и ее табличное значение для аргумента x = 2,5.

Ответ: $P_{400}(60) \approx 0{,}00219$.

Пример 1.22. Вероятность выхода из строя для приборов определенного типа составляет 0,001. Найти вероятность того, что из 2000 проверяемых приборов выйдут из строя только 2 прибора.

Решение.

Вероятность события A «Проверяемый прибор выйдет из строя» $(P(A)=0{,}001)$ остается постоянной в серии из 2000 независимых испытаний. Поэтому имеет место схема испытаний Бернулли. Но $n = 2000$ очень велико и формула Бернулли в данном случае не применяется, необходимо применить одну из приближенных формул. Для выбора нужной формулы оценим произведение npq:

$$npq = 2000\cdot 0{,}001\cdot(1-0{,}001) = 2000\cdot 0{,}001\cdot 0{,}999 = 1{,}998 << 10$$

(почнее: $np^2 = 2000\cdot 0{,}001^2 = 0{,}002 << 1$).

Из полученной оценки понятно, что для решения задачи должна быть применена приближенная формула Пуассона

$$P_n(k) \approx \frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

при $n = 2000$, $k = 2$ и $\lambda = np = 2000\cdot 0{,}001 = 2$:

$$P_{2000}(2) \approx \frac{2^2}{2!}\cdot e^{-2} = 0{,}2707.$$

Мы использовали таблицу значений для функции $P_n(k) \approx \dfrac{\lambda^k}{k!}\cdot e^{-\lambda}$ при $k = 2$ и $\lambda = np = 2$.

Ответ: $P_{2000}(2) \approx$ 0,2707.

Пример 1.23. Известно, что среди выпускаемых изделий изделия первого сорта составляют 20%, остальные — высшего сорта. Какова вероятность того, что из 400 проверенных изделий не менее 50, но и не более 90 окажутся изделиями первого сорта?

Решение.

Вероятность события $\boldsymbol{A}$: «Проверяемое изделие окажется изделием первого сорта» $\left(P(A)=\dfrac{20}{100}=0{,}2\right)$ остается постоянной в серии из 400 независимых испытаний. Поэтому имеет место схема испытаний Бернулли. Однако при $n = 400$ формула Бернулли не используется. В **Примере 1.19** мы уже оценили произведение $npq = 400\cdot 0{,}2\cdot(1-0{,}2) = 64 > 10$ и пришли к выводу, что для решения задачи должна быть применена приближенная формула Лапласа. В случае рассматриваемого примера — это интегральная формула Лапласа

$$P_n\left(k_1 \le k \le k_2\right) \approx \Phi\left(x_2\right) - \Phi\left(x_1\right),$$

так как число «успехов» должно быть заключено в интервале от 50 до 90:

$$P_{400}\left(50 \le k \le 90\right) \approx \Phi\left(x_2\right) - \Phi\left(x_1\right).$$

Вычислим соответствующие значения аргументов для функции Лапласа:

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{90 - 400 \cdot 0{,}2}{\sqrt{64}} = 1{,}25,$$

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{50 - 400 \cdot 0{,}2}{\sqrt{64}} = -3{,}75.$$

Теперь найдем ответ на вопрос задачи:

$$P_{400}\left(50 \le k \le 90\right) \approx \Phi\left(1{,}25\right) - \Phi\left(-3{,}75\right) =$$

$$= \Phi\left(1{,}25\right) + \Phi\left(3{,}75\right) = 0{,}39435 + 0{,}49991 = 0{,}89426.$$

При вычислениях было использовано свойство нечетности функции Лапласа и ее табличные значения для аргументов 1,25 и 3,75.

Ответ: $P_{400}\left(50 \le k \le 90\right) \approx 0{,}89426$.

Пример 1.24. Вероятность выхода из строя для приборов определенного типа составляет 0,001. Найти вероятность того, что из 2000 проверяемых приборов выйдут из строя более двух приборов.

Решение.

В **Примере 1.22** мы уже обосновали необходимость применения приближенной формулы Пуассона, опираясь на оценку произведения npq:

$$npq = 2000 \cdot 0{,}001 \cdot (1 - 0{,}001) = 2000 \cdot 0{,}001 \cdot 0{,}999 = 1{,}998 << 10.$$

(точнее: $np^2 = 2000 \cdot 0{,}001^2 = 0{,}002 << 1$).

Поэтому просто найдем ответ на вопрос задачи:

$$P_{2000}\left(2 < k \le 2000\right) \approx 1 - \left(P_{2000}\left(0\right) + P_{2000}\left(1\right) + P_{2000}\left(2\right)\right) =$$

= 1 — (0,1353 + 0,2707 + 0,2707) = 0,3233.

При этом мы использовали табличные значения для функции

$$P_n\left(k\right) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda} \text{ при } \lambda = np = 2000 \cdot 0{,}001 = 2 \text{ и } k = 0,\ k = 1,\ k = 2.$$

Ответ: $P_{2000}\left(2 < k \le 2000\right) \approx 0{,}3233$.

Задачи для самостоятельного решения

В задачах **1.81–1.100** при решении использовать формулу Бернулли или приближенные формулы.

1.81. Тест состоит из пяти вопросов, на каждый вопрос приведено четыре ответа, из которых только один является правильным. Студент выбирает ответы наугад. Какова вероятность того, что он правильно ответит только на три вопроса? Какова вероятность того, что студент правильно ответит на все вопросы?

1.82. В горном районе работают четыре сейсмологические станции, вероятность выхода из строя в течение года для каждой из них равна 0,2. Какова вероятность того, что в течение года из строя выйдет только одна станция? Какова вероятность того, что в течение года из строя выйдет хотя бы одна станция?

1.83. Какова вероятность того, что в многодетной семье из восьми детей будет семь мальчиков и одна девочка? Какова вероятность того, что в многодетной семье из восьми детей будет четыре мальчика и четыре девочки?

1.84. Что более вероятно: выиграть две партии из четырех или три партии из шести у равного по силе партнера?

1.85. Вероятность ошибочной транзакции для некоторого банка составляет 0,002. Найти наиболее вероятное число ошибочных транзакций при проведении 4999 операций.

1.86. По шоссе за час проезжает 120 автомашин. Вероятность того, что проезжающей машине понадобится заправка в бензоколонке, равна 0,4. Какова вероятность того, что в течение часа на бензоколонке заправятся 20 машин?

1.87. В некотором регионе вероятность того, что летний день будет солнечным, равна 0,8. Какова вероятность того, что из 92 летних дней 75 будут солнечными?

1.88. Известно, что вероятность отчисления студента, попавшего в списки неуспевающих в семестре, равна 0,2. Какова вероятность того, что из 52 студентов из списка неуспевающих будут отчислены 15?

1.89. Вероятность бракованного изделия в продукции определенного вида равна 0,05. Какова вероятность того, что при проверке 100 изделий этого вида 90 окажутся без брака?

1.90. Вероятность того, что выданный кредит определенного вида будет погашен в срок, банк считает равной 0,9. Какова вероятность того, что из 150 выданных кредитов такого вида в срок будут погашены 140?

1.91. Вероятность ошибочной транзакции для некоторого банка составляет 0,001. Какова вероятность того, что из 2000 транзакций ошибочными будут две?

1.92. У определенного вида морских моллюсков вероятность обнаружить жемчужину равна 0,004. Какова вероятность того, что из 1000 моллюсков жемчужину обнаружили только у одного?

1.93. Вероятность укуса человека клещом определенного вида при прогулке по лесу в некотором регионе равна 0,01. Известно, что 10% клещей этого вида являются носителями вируса энцефалита. Какова вероятность того, что из 1000 человек, гулявших по лесу в этом регионе, энцефалитом заболеют 5?

1.94. В течение года из-за возможных изменений условий хранения (влажность, температура) вероятность выхода из строя изделия равна 0,0003. Какова вероятность того, что в течение года хранения на этом складе 2000 изделий из строя выйдет только одно?

1.95. При работе высокоточного станка-автомата на каждую 1000 изделий выходит 1 бракованное. Какова вероятность того, что из 5000 изделий, изготовленных на этом автомате, бракованными окажутся 2 изделия?

1.96. По шоссе за час проезжает 120 автомашин. Вероятность того, что проезжающей машине понадобится заправка на бензоколонке, равна 0,4. Какова вероятность того, что в течение часа на бензоколонке заправятся не более 20 машин?

1.97. В некотором регионе вероятность того, что летний день будет солнечным, равна 0,8. Какова вероятность того, что из 92 летних дней не менее 75 будут солнечными?

1.98. Вероятность бракованного изделия в продукции определенного вида равна 0,05. Какова вероятность того, что при проверке 100 изделий этого вида окажутся без брака от 90 до 95 изделий?

1.99. Вероятность ошибочной транзакции для некоторого банка составляет 0,001. Какова вероятность того, что из 2000 транзакций ошибочными будут не более двух?

1.100. В течение года из-за возможных изменений условий хранения (влажность, температура) вероятность выхода из строя изделия равна 0,0003. Какова вероятность того, что в течение года хранения на этом складе 2000 изделий из строя выйдет более одного изделия?

ГЛАВА 2

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

2.1. Закон распределения дискретной случайной величины

Обобщением понятия события является понятие случайной величины. ***Случайной*** называют величину, принимающую числовое значение, но заранее не известно — какое. Прибыль фирмы в наугад выбранном году, курс доллара в наугад выбранный день или время бесперебойной работы какого-либо прибора — это лишь немногие примеры из огромного количества случайных величин. Каждая из них принимает определенное числовое значение, но узнать заранее это значение не представляется возможным. Для изучения случайной величины в первую очередь важно знать, какие значения она может принимать. В зависимости от возможных значений, принимаемых величиной, различают дискретные и непрерывные случайные величины.

Если множество возможных значений величины конечно или счетно, то величина определяется как дискретная. Если же множество возможных значений величины представляет собой часть или все множество действительных чисел, то такая величина определяется как непрерывная.

Мы рассмотрим сначала дискретные случайные величины (ДСВ).

Существуют стандарты обозначений случайных величин как X, Y, Z... Соответствующие возможные значения величины обозначают малыми буквами с индексами: $x_1, x_2, \ldots$ — значения для величины X, $y_1, y_2, \ldots$ — значения для величины Y, $z_1, z_2, \ldots$ — значения для величины Z.

Чтобы полностью охарактеризовать дискретную случайную величину, необходимо не только указать все ее возможные значения, но и определить вероятности, с которыми величина эти значения может принять. Такая информация для ДСВ записывается в виде таблицы:

Значение X	x_1	x_2	...	x_n
Вероятность p_i	p_1	p_2	...	p_n

Здесь

$$p_1 = P(X = x_1)\ ,\ p_2 = P(X = x_2)\ ,\ ...,\ p_n = P(X = x_n)\ .$$

Вероятность того, что величина принимает определенное числовое значение, является событием. Поэтому если в первой строке таблицы мы предлагаем все возможные значения величины, то формирующиеся события образуют полную группу, и поэтому сумма всех вероятностей во второй строке таблицы должна быть равна 1. Это условие является определяющим для формирования закона распределения дискретной случайной величины.

Законом распределения дискретной случайной величины называют соответствие между возможными значениями этой величины и вероятностями, отвечающими этим значениям при необходимом условии равенства единице суммы этих вероятностей:

Значение X	x_1	x_2	...	x_n
Вероятность p_i	p_1	p_2	...	p_n

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

Пример 2.1. В коробке находятся пять белых и два черных шара. Случайным образом из коробки извлекают два шара. Составить закон распределения для величины X — числа белых шаров в выборке.

Решение.

Предлагаемый состав шаров в коробке допускает, чтобы среди выбранных двух шаров оба были черными (0 белых), или оба — белыми, или же выборка состояла из 1 белого и 1 черного шаров. Поэтому в качестве возможных значений величины X мы предлагаем 0, 1 и 2.

Соответствующие этим значениям вероятности будем вычислять как вероятности событий, связанных с принятием величиной определенного числового значения:

$p_1 = P(X = 0) = \dfrac{C_5^0 \cdot C_2^2}{C_7^2} = \dfrac{1}{21}$ (вычисляли вероятность события «Из двух выбранных из коробки шаров — оба черные»).

$p_2 = P(X=1) = \frac{C_5^1 \cdot C_2^1}{C_7^2} = \frac{10}{21}$ (вычисляли вероятность события «Из двух выбранных из коробки шаров один белый, один черный»).

$p_3 = P(X=2) = \frac{C_5^2 \cdot C_2^0}{C_7^2} = \frac{10}{21}$ (вычисляли вероятность события «Из двух выбранных из коробки шаров оба белые»).

Условие $\sum_{i=1}^{n} p_i = \frac{1}{21} + \frac{10}{21} + \frac{10}{21} = 1$ выполнено.

Поэтому закон распределения имеет вид:

Значение X	0	1	2
Вероятность p_i	$\frac{1}{21}$	$\frac{10}{21}$	$\frac{10}{21}$

Эту таблицу и можно считать ответом на вопрос задачи.

Пример 2.2. Банк выдал три кредита предприятию, оценив вероятность возврата каждого кредита в 0,9. Составить закон распределения величины X — числа возвращенных банку кредитов.

Решение.

В качестве возможных значений величины X мы предлагаем 0, 1, 2 и 3.

Соответствующие этим значениям вероятности будем вычислять как вероятности событий, связанных с принятием величиной определенного числового значения:

$$p_1 = P(X=0) = P_3(0) = C_3^0 \cdot 0{,}9^0 \cdot 0{,}1^3 = 0{,}001 .$$

(Вычисляли вероятность события «Из трех выданных кредитов возвратятся 0» по формуле Бернулли).

$$p_2 = P(X=1) = P_3(1) = C_3^1 \cdot 0{,}9^1 \cdot 0{,}1^2 = 0{,}027$$

(вычисляли вероятность события «Из трех выданных кредитов возвратится 1» по формуле Бернулли).

$$p_3 = P(X=2) = P_3(2) = C_3^2 \cdot 0{,}9^2 \cdot 0{,}1^1 = 0{,}243$$

(вычисляли вероятность события «Из трех выданных кредитов возвратятся 2» по формуле Бернулли).

$$p_4 = P(X=3) = P_3(3) = C_3^3 \cdot 0{,}9^3 \cdot 0{,}1^0 = 0{,}729$$

(вычисляли вероятность события «Из трех выданных кредитов возвратятся все 3» по формуле Бернулли).

Условие $\sum_{i=1}^{n} p_i = 0{,}001 + 0{,}027 + 0{,}243 + 0{,}729 = 1$ выполнено.

Поэтому закон распределения имеет вид:

Значение X	0	1	2	3
Вероятность p_i	0,001	0,027	0,243	0,729

Эта таблица и определяет искомый закон распределения.

Задачи для самостоятельного решения

В задачах **2.1–2.10** необходимо пояснить, как определяется закон распределения искомой величины X.

2.1. В коробке находятся 5 белых и 3 черных шара. Из нее извлекают (без возвращения) 4 шара. Составить закон распределения случайной величины X — количества черных шаров в выборке.

2.2. Тест состоит из 5 вопросов, на каждый вопрос приведено 4 ответа, из которых только один является правильным. Студент выбирает ответы наугад. Составить закон распределения случайной величины X — количества правильных ответов на вопросы теста.

2.3. У самолета имеется два одинаковых двигателя, вероятность отказа в полете для каждого двигателя равна 0,001. Составить закон распределения случайной величины X — количества отказавших в полете двигателей самолета.

2.4. Испытываются три независимо работающих прибора. Вероятность отказа первого прибора равна 0,1, второго — 0,2 и третьего — 0,15. Составить закон распределения случайной величины X — числа отказавших в процессе испытания приборов.

2.5. Проверено 25 контрольных работ, за четыре из них была поставлена оценка «отлично». Из всех проверенных работ наугад выбирают три. Составить закон распределения случайной величины X — числа работ, оцененных на «отлично», среди выбранных.

2.6. Отрезок длины 10 см разделен на два — длины 3 см и 7 см. На этот отрезок случайным образом вбрасывают 4 точки. Составить закон распределения случайной величины X — числа точек, не попавших на отрезок длины 3 см.

2.7. Баскетболист выполняет 2 штрафных броска. Вероятность попадания в кольцо при одном броске для него равна 0,8. Составить закон распределения случайной величины X — числа промахов при выполнении штрафных бросков.

2.8. Баскетболист выполняет броски до первого попадания, но не более 5 бросков. Составить закон распределения случайной величины X — количества осуществленных бросков.

2.9. Участнику викторины вопросы задают до тех пор, пока он на них правильно отвечает. Если участник ответил неправильно или количество правильных ответов стало равным 5, вопросы задавать перестают. Составить закон распределения случайной величины X — числа заданных участнику вопросов, если вероятность правильного ответа на каждый заданный вопрос викторины для участника равна 0,5.

2.10. В коробке находятся 5 белых и 3 черных шара. Из нее извлекают по одному шару (без возвращения) до тех пор, пока не появится белый шар, после этого опыт прекращают. Составить закон распределения случайной величины X — числа извлеченных из коробки шаров до первого появления белого шара.

2.2. Арифметические операции с независимыми дискретными случайными величинами

Со случайными величинами можно выполнять целый ряд арифметических действий (операций), при этом должна формироваться новая случайная величина. Если речь идет о дискретных случайных величинах, то результат выполнения операции должен иметь свой закон распределения. При этом возможны как действия, при которых меняются только значения величины, а вероятности остаются неизменными, так и действия, в результате которых изменятся и значения, и вероятности. Рассмотрим подробнее возможности выполнения действий с дискретными случайными величинами.

Если X — случайная величина с известным законом распределения, то случайной является и величина $Y = g(X)$ при условии, что $g(x)$ — непрерывная числовая функция.

Под действием функции $g(x)$ изменяются только значения величины X. Поэтому если закон распределения дискретной случайной величины X имеет вид:

Значение X	x_1	x_2	...	x_n
Вероятность p_i	p_1	p_2	...	p_n

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

то закон распределения дискретной случайной величины $Y = g(X)$ определяется следующим образом:

Значение $Y = g(X)$	$g(x_1)$	$g(x_2)$	...	$g(x_n)$
Вероятность p_i	p_1	p_2	...	p_n

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

Пример 2.3. Известен закон распределения величины X:

Значение X	–2	–1	0	1	2
Вероятность p_i	0,1	0,1	0,2	0,3	0,3

Составить законы распределений для величин:

а) $Y = X + 5$;

б) $Y = 3X$;

в) $Y = X^2$.

Решение.

а) величина $Y = X + 5$ определяется на основе функции $g(x) = x + 5$, под действием которой изменятся значения величины X. Найдем все возможные значения для величины $Y = X + 5$:

$y_1 = x_1 + 5 = -2 + 5 = 3$;

$y_2 = x_2 + 5 = -1 + 5 = 4$;

$y_3 = x_3 + 5 = 0 + 5 = 5$;

$y_4 = x_4 + 5 = 1 + 5 = 6$;

$y_5 = x_5 + 5 = 2 + 5 = 7$.

Поскольку вероятности при этом не меняются, закон распределения для новой величины $Y = X + 5$ будет иметь вид:

Значение $Y = X + 5$	3	4	5	6	7
Вероятность p_i	0,1	0,1	0,2	0,3	0,3

б) величина $Y = 3X$ определяется на основе функции $g(x) = 3x$, под действием которой изменятся значения величины X. Найдем все возможные значения для величины $Y = 3X$:

$y_1 = 3x_1 = -6$; $y_2 = 3x_2 = -3$; $y_3 = 3x_3 = 0$;

$y_4 = 3x_4 = 3$; $y_5 = 3x_5 = 6$.

И соответствующий закон распределения примет вид:

Значение $Y = 3X$	–6	–3	0	3	6
Вероятность p_i	0,1	0,1	0,2	0,3	0,3

в) величина $Y = X^2$ определяется на основе функции $g(x) = x^2$, под действием которой изменятся значения величины X. Найдем все возможные значения для величины $Y = X^2$:

$y_1 = (x_1)^2 = (-2)^2 = 4$; $y_2 = (x_2)^2 = (-1)^2 = 1$; $y_3 = (x_3)^2 = 0^2 = 0$;

$y_4 = (x_4)^2 = 1^2 = 1$; $y_5 = (x_5)^2 = 2^2 = 4$.

Мы видим, что среди вычисленных значений есть одинаковые. В таблицу, определяющую закон распределения, заносят только различные значения.

Поэтому среди возможных значений $Y = X^2$ отметим только 0, 1 и 4.

Вычислим теперь соответствующие этим значениям вероятности:

$P(Y = 4) = P(X^2 = 4) = P(X = -2) + P(X = 2) = 0{,}1 + 0{,}3 = 0{,}4$;

$P(Y = 1) = P(X^2 = 1) = P(X = -1) + P(X = 1) = 0{,}1 + 0{,}3 = 0{,}4$;

$P(Y = 0) = P(X^2 = 0) = P(X = 0) = 0{,}2$.

И закон распределения для $Y = X^2$ в итоге принимает вид:

Значение $Y = X^2$	0	1	4
Вероятность p_i	0,2	0,4	0,4

Ответ: Законы распределения искомых величин:

Значение $Y = X + 5$	3	4	5	6	7
Вероятность p_i	0,1	0,1	0,2	0,3	0,3

Значение $Y = 3X$	–6	–3	0	3	6
Вероятность p_i	0,1	0,1	0,2	0,3	0,3

Значение $Y = X^2$	0	1	4
Вероятность p_i	0,2	0,4	0,4

Выполнение операций сложения, умножения, вычитания и деления с дискретными случайными величинами позволяет получить новые дискретные случайные величины, называемые суммой, произведением, разностью и частным исходных ДСВ. Для формирования законов распределения этих новых ДСВ необходимо определять и новые значения, и соответствующие этим новым значениям вероятности. Проще всего задачи решаются в случае, когда исходные ДСВ являются независимыми.

Дискретные случайные величины X и Y являются ***независимыми*** в том случае, когда принятие величиной X какого-либо значения никак не влияет ни на возможные значения величины Y, ни не отвечающие им вероятности.

Именно для независимых величин X и Y выполняется равенство:

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) \qquad (*)$$

для всех возможных значений i, j.

Суммой независимых величин X и Y называется величина $Z = X + Y$, значения которой определяются возможными значениями $x_i + y_j$ и вероятностями, определяемыми условием (*), для всех возможных значений i, j.

Разностью независимых величин X и Y называется величина $Z = X - Y$, значения которой определяются возможными значениями $x_i - y_j$ и вероятностями, определяемыми условием (*), для всех возможных значений i, j.

Произведением независимых величин X и Y называется новая величина $Z = X \cdot Y$, значения которой определяются всеми возможными значениями $x_i \cdot y_j$ и вероятностями, определяемыми условием (*), для всех возможных значений i, j.

Частное от деления независимых величин X и Y определяется аналогично, но при дополнительном условии, что среди возможных значений Y нет равных нулю: $y_j \neq 0$ для всех возможных значений j.

Пример 2.4. Даны законы распределений независимых величин X и Y:

Значение X (x_i)	–1	0	1	2
Вероятность p_i	0,2	0,2	0,3	0,3

Значение Y (y_j)	–1	0	1
Вероятность p_j	0,2	0,3	0,5

1) Составить законы распределения для их суммы, разности и произведения.

2) Какое из событий наиболее вероятно: $A:\{X+Y=2\}$, $B:\{X-Y=2\}$ или $C:\{X\cdot Y=2\}$?

Решение.

1) а) Составим закон распределения $Z = X + Y$.

Для этого найдем все возможные суммы $x_i + y_j$ и соответствующие им вероятности по правилу: $P(X+Y=x_i+y_j)=P(X=x_i)\cdot P(Y=y_j)$, следующему из условия независимости величин X и Y:

x_i	y_j	$x_i \quad y_j$	$P(X+Y=x_i+y_j)=P(X=x_i)\cdot P(Y=y_j)$
–1	–1	–2	0,2 · 0,2 = 0,04
–1	0	**–1**	0,2 · 0,3 = **0,06**
–1	1	0	0,2 · 0,5 = 0,1
0	–1	**–1**	0,2 · 0,2 = **0,04**
0	0	0	0,2 · 0,3 = 0,06
0	1	1	0,2 · 0,5 = 0,1
1	–1	0	0,3 · 0,2 = 0,06
1	0	1	0,3 · 0,3 = 0,09
1	1	2	0,3 · 0,5 = 0,15
2	–1	1	0,3 · 0,2 = 0,06
2	0	2	0,3 · 0,3 = 0,09
2	1	3	0,3 · 0,5 = 0,15

В выделенном столбце находятся суммы $x_i + y_j$ для всех возможных значений индексов i, j. Мы видим, что среди значений есть одинаковые. В таблицу для закона распределения $Z = X + Y$ заносим только различные значения: –2, –1, 0, 1, 2 и 3. При вычислении вероятностей для этих значений используем правило: для одинаковых значений суммы $x_i + y_j$ вероятности складываются. Например (см. выделенные цветом значения и вероятности):

$$P(X+Y=-1)=P(X=-1)\cdot P(Y=\mathbf{0})+P(X=\mathbf{0})\cdot P(Y=-1)=$$
$$=0{,}06+0{,}04=0{,}1.$$

В итоге закон распределения величины $Z = X + Y$ имеет вид:

Значение величины $Z = X + Y$	–2	–1	0	1	2	3
Вероятности	0,04	0,1	0,22	0,25	0,24	0,15

Условие на закон распределения выполнен: сумма всех вероятностей во второй строке равна 1.

б) Составим закон распределения $Z = X - Y$.

Для этого найдем все возможные разности $x_i - y_j$ и соответствующие им вероятности по правилу: $P(X - Y = x_i - y_j) = P(X = x_i) \times$ $\times P(Y = y_j)$, следующему из условия независимости величин X и Y:

x_i	y_j	$x_i - y_j$	$P(X - Y = x_i - y_j) = P(X = x_i) \cdot P(Y = y_j)$
–1	–1	0	0,2 · 0,2 = 0,04
–1	0	–1	0,2 · 0,3 = **0,06**
–1	1	–2	0,2 · 0,5 = 0,1
0	–1	1	0,2 · 0,2 = 0,04
0	0	0	0,2 · 0,3 = 0,06
0	1	–1	0,2 · 0,5 = **0,1**
1	–1	2	0,3 · 0,2 = 0,06
1	0	1	0,3 · 0,3 = 0,09
1	1	0	0,3 · 0,5 = 0,15
2	–1	3	0,3 · 0,2 = 0,06
2	0	2	0,3 · 0,3 = 0,09
2	1	1	0,3 · 0,5 = 0,15

В выделенном столбце теперь находятся разности $x_i \quad y_j$ для всех возможных значений индексов i, j. Мы видим, что среди значений есть одинаковые. В таблицу для закона распределения $Z = X - Y$, как и прежде, заносим только различные значения: –2, –1, 0, 1, 2 и 3. При вычислении вероятностей для этих значений используем правило: для одинаковых значений разности $x_i - y_j$ вероятности складываются. Например (см. выделенные цветом значения и вероятности):

$$P(X - Y = -1) = P(X = -1) \cdot P(Y = 0) + P(X = 0) \cdot P(Y = 1) =$$
$$= 0{,}06 + 0{,}1 = \mathbf{0{,}16}.$$

В итоге закон распределения величины $Z = X - Y$ имеет вид:

Значение $Z = X - Y$	–2	–1	0	1	2	3
Вероятность	0,1	0,16	0,25	0,28	0,15	0,06

Условие на закон распределения и в этом случае выполнено: сумма всех вероятностей во второй строке равна 1.

в) Составим закон распределения $Z = X \cdot Y$.

Для этого найдем все возможные произведения $x_i \cdot y_j$ и соответствующие им вероятности по правилу: $P(X \cdot Y = x_i \cdot y_j) = P(X = x_i) \cdot P(Y = y_j)$, следующему из условия независимости величин X и Y:

x_i	y_j	$x_i \cdot y_j$	$P(X \cdot Y = x_i \cdot y_j) = P(X = x_i) \cdot P(Y = y_j)$
–1	–1	1	0,2 · 0,2 = 0,04
–1	0	0	0,2 · 0,3 = 0,06
–1	1	–1	0,2 · 0,5 = 0,1
0	–1	0	0,2 · 0,2 = 0,04
0	0	0	0,2 · 0,3 = 0,06
0	1	0	0,2 · 0,5 = 0,1
1	–1	–1	0,3 · 0,2 = 0,06
1	0	0	0,3 · 0,3 = 0,09
1	1	1	0,3 · 0,5 = 0,15
2	–1	–2	0,3 · 0,2 = 0,06
2	0	0	0,3 · 0,3 = 0,09
2	1	2	0,3 · 0,5 = 0,15

В выделенном столбце теперь находятся произведения $x_i \cdot y_j$ для всех возможных значений индексов i, j. Мы видим, что и среди этих значений есть одинаковые. В таблицу для закона распределения $Z = X \cdot Y$, по-прежнему заносим только различные значения: –2, –1, 0, 1 и 2. При вычислении вероятностей для этих значений используем правило: для одинаковых значений произведений $x_i \cdot y_j$ вероятности складываются.

Например (см. выделенные цветом значения и вероятности):

$$P(X \cdot Y = -1) = P(X = -1) \cdot P(Y = 1) + P(X = 1) \cdot P(Y = -1) =$$
$$= 0{,}1 + 0{,}06 = 0{,}16.$$

В итоге закон распределения величины $Z = X \cdot Y$ имеет вид:

Значение $Z = X \cdot Y$	–2	–1	0	1	2
Вероятность	0,06	0,16	0,44	0,19	0,15

Условие на закон распределения и в этом случае выполнено: сумма всех вероятностей во второй строке равна 1.

2) Из законов распределения $X + Y$, $X — Y$ и $X \cdot Y$ определяем:

$P(X + Y = 2) = 0{,}24;$

$P(X - Y = 2) = 0{,}15;$

$P(X \cdot Y = 2) = 0{,}15.$

Поэтому наиболее вероятно событие $A: \{X + Y = 2\}$.

Ответ:

1) Искомые законы распределений:

Значение $Z = X + Y$	–2	–1	0	1	2	3
Вероятность	0,04	0,1	0,22	0,25	0,24	0,15

Значение $Z = X - Y$	–2	–1	0	1	2	3
Вероятность	0,1	0,16	0,25	0,28	0,15	0,06

Значение $Z = X \cdot Y$	–2	–1	0	1	2
Вероятность	0,06	0,16	0,44	0,19	0,15

2) Наиболее вероятно событие $A: \{X + Y = 2\}$.

Задачи для самостоятельного решения

В задачах **2.11–2.20** заданы законы распределений независимых величин X и Y. Необходимо составить закон распределения случайной величины Z и ответить на вопросы задачи.

2.11. Случайные величины X и Y независимы:

Значение X	–2	–1	0	1
Вероятность	0,1	0,3	0,1	0,5

Значение Y	0	1	2
Вероятность	0,5	0,3	0,2

Составить закон распределения случайной величины $Z = |X| + 2Y - 1$ и найти $P(Z > 0)$.

2.12. Случайные величины X и Y независимы:

Значение X	1	2	3
Вероятность	0,1	0,2	0,7

Значение Y	–1	0	1	2
Вероятность	0,2	0,2	0,2	0,4

Составить закон распределения случайной величины $Z = Y^2 - 3X - 2$ и найти вероятность $P(Z + 2 > 0)$.

2.13. Случайные величины X и Y независимы:

Значение X	-2	-1	0	1	2
Вероятность	0,1	0,3	0,1	0,3	0,2

Значение Y	0	1
Вероятность	0,8	0,2

Составить закон распределения случайной величины $Z = X^2 - 5Y^2$ и найти вероятность $P\left(X^2 < 5Y^2\right)$.

2.14. Случайные величины X и Y независимы:

Значение X	-1	0	1
Вероятность	0,2	0,3	0,5

Значение Y	-2	-1	1	2
Вероятность	0,1	0,4	0,4	0,1

Составить закон распределения случайной величины $Z = (X+1)(Y-2)$ и найти вероятность $P(Z \neq 0)$.

2.15. Случайные величины X и Y независимы:

Значение X	1	2	4
Вероятность	0,1	0,1	0,8

Значение Y	-2	0	2	4
Вероятность	0,1	0,2	0,3	0,4

Составить закон распределения случайной величины $Z = 4 - Y^2 X$ и найти вероятность $P(Z < 0)$.

2.16. Случайные величины X и Y независимы:

Значение X	-3	-1	1	3
Вероятность	0,2	0,3	0,1	0,4

Значение Y	-1	0	1
Вероятность	0,2	0,6	0,2

Составить закон распределения случайной величины $Z = |X| - 3|Y|$ и сравнить вероятности $P(Z < 0)$ и $P(|X| = 3|Y|)$.

2.17. Случайные величины X и Y независимы:

Значение X	-2	-1	0	1	2
Вероятность	0,1	0,1	0,1	0,2	0,5

Значение Y	0	1
Вероятность	0,6	0,4

Составить закон распределения случайной величины $Z = X^2 - Y^3 - 2$ и найти вероятность $P\left(Z = 2 \middle| Z > 0\right)$.

2.18. Случайные величины X и Y независимы:

Значение X	−4	−2	0
Вероятность	0,2	0,5	0,3

Значение Y	−3	0	3	9
Вероятность	0,4	0,3	0,2	0,1

Составить закон распределения случайной величины $Z = X^3 \cdot Y$ и сравнить вероятности $P\left(Z < 0\right)$ и $P\left(Z > 0\right)$.

2.19. Случайные величины X и Y независимы:

Значение X	−3	−1	0	1	3
Вероятность	0,1	0,2	0,4	0,2	0,1

Значение Y	−1	1
Вероятность	0,5	0,5

Составить закон распределения случайной величины $Z = 2X - \left(Y + 1\right)^2$ и найти вероятность $P\left(Z = 0 \middle| Z \le 0\right)$.

2.20. Случайные величины X и Y независимы:

Значение X	−1	0
Вероятность	0,1	0,9

Значение Y	−3	−2	0	1	2
Вероятность	0,2	0,2	0,2	0,2	0,2

Составить закон распределения случайной величины $Z = 3X^2 \cdot Y$ и сравнить вероятности $P\left(Z = 0 \middle| Z \ge 0\right)$ и $P\left(Z = 0 \middle| Z \le 0\right)$.

2.3. Числовые характеристики дискретных случайных величин

Для решения многих практических задач совсем не обязательно знать все возможные значения величины и соответствующие этим значениям вероятности, а вполне достаточно ориентироваться лишь на отдельные числовые показатели, выражающие наиболее важные свойства случайной величины.

Эти показатели принято называть числовыми характеристиками случайной величины. Мы перечислим основные характеристики дискретных случайных величин и сформулируем их основные свойства.

Математическим ожиданием дискретной случайной величины с известным законом распределения

Значение X	x_1	x_2	...	x_n
Вероятность p_i	p_1	p_2	...	p_n

при выполненном условии $\sum_{i=1}^{n} p_i = 1$ называют сумму произведений значений величины на соответствующую вероятность:

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_n \cdot p_n .$$

Пример 2.5. Найти математическое ожидание величины, закон распределения которой имеет вид:

Значение X	–1	0	1	2
Вероятность p_i	0,2	0,2	0,3	0,3

Решение.

$$E(X) = \sum_{i=1}^{n} x_i \cdot p_i = -1 \cdot 0{,}2 + 0 \cdot 0{,}2 + 1 \cdot 0{,}3 + 2 \cdot 0{,}3 = 0{,}7 .$$

Ответ: $E(X) = 0{,}7$.

Если множество возможных значений величины счетно, то закон распределения такой величины имеет вид:

Значение X	x_1	x_2	...	x_n	...
Вероятность p_i	p_1	p_2	...	p_n	...

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

Существование математического ожидания $E(X)$ в таком случае определяется абсолютной сходимостью числового ряда:

$$\sum_{i=1}^{\infty} x_i \cdot p_i = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_n \cdot p_n + \ldots$$

Поэтому величины со счетным множеством значений могут и не иметь математического ожидания.

Пример 2.6. Найти математическое ожидание величины, закон распределения которой имеет вид:

Значение величины X	3	3^2	3^3	…
Вероятность p_i	$\frac{1}{3}$	$\frac{1}{3^2}$	$\frac{1}{3^3}$	…

Решение.

Вычислим

$$\sum_{i=1}^{\infty} x_i \cdot p_i = 3 \cdot \frac{1}{3} + 3^2 \cdot \frac{1}{3^2} + 3^3 \cdot \frac{1}{3^3} + \ldots = 1 + 1 + 1 + \ldots = \sum_{n=1}^{\infty} 1 = \infty,$$

ряд расходится, следовательно, математическое ожидание не существует.

Ответ: $E(X)$ не существует.

Перечислим ***свойства математического ожидания***, которые применяются при решении задач.

1. Математическое ожидание постоянной величины равно самой этой постоянной:

$$E(C) = C, \; C \text{ — const}.$$

2. Постоянный множитель можно выносить за знак математического ожидания:

$$E(kX) = k \cdot E(X).$$

3. Математическое ожидание от суммы или разности величин равно сумме или разности их математических ожиданий:

$$E(X \pm Y) = E(X) \pm E(Y).$$

4. Математическое ожидание произведения **независимых** величин равно произведению математических ожиданий этих величин:

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

5. Если $\phi(x)$ — числовая функция, а X — дискретная случайная величина с известным законом распределения, то

$$E(\phi(X)) = \phi(x_1) \cdot p_1 + \phi(x_2) \cdot p_2 + \ldots$$

Пример 2.7. Даны законы распределений независимых величин X и Y:

Значение X (x_i)	–1	0	1	2
Вероятность p_i	0,2	0,2	0,3	0,3

Значение Y (y_j)	–1	0	1
Вероятность p_j	0,2	0,3	0,5

Найти $E\left(X\cdot(3X-5Y)\right)$.

Решение.

Раскроем скобки под знаком математического ожидания и, используя свойства, преобразуем:

$$E\left(X\cdot(3X-5Y)\right)=$$

$$=E\left(3X^2-5XY\right)=3E\left(X^2\right)-5E\left(XY\right)=3E\left(X^2\right)-5E\left(X\right)\cdot E\left(Y\right).$$

Вычислим:

$$E\left(X\right)=-1\cdot 0{,}2+0\cdot 0{,}2+1\cdot 0{,}3+2\cdot 0{,}3=0{,}7;$$

$$E\left(Y\right)=-1\cdot 0{,}2+0\cdot 0{,}3+1\cdot 0{,}5=0{,}3;$$

$$E\left(X^2\right)=(-1)^2\cdot 0{,}2+0^2\cdot 0{,}2+1^2\cdot 0{,}3+2^2\cdot 0{,}3=1{,}7.$$

Теперь подставим:

$$E\left(X\cdot(3X-5Y)\right)=3E\left(X^2\right)-5E\left(X\right)\cdot E\left(Y\right)=$$

$$=3\cdot 1{,}7-5\cdot 0{,}7\cdot 0{,}3=4{,}05.$$

Ответ: $E\left(X\cdot(3X-5Y)\right)=4{,}05$

Дисперсией дискретной случайной величины называют математическое ожидание квадрата отклонения величины от своего математического ожидания:

$$D\left(X\right)=E\left(\left(X-E\left(X\right)\right)^2\right).$$

При известном законе распределения случайной величины

Значение X	x_1	x_2	...	x_n
Вероятность p_i	p_1	p_2	...	p_n

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

получим формулу, по которой можно вычислять дисперсию:

$$D(X)=\sum_{i=1}^{n}\left(x_i-E(X)\right)^2\cdot p_i\ ,$$

где $E(X)=\sum_{i=1}^{n}x_i\cdot p_i$.

Если воспользоваться свойствами математического ожидания, то можно получить еще одну формулу для вычисления дисперсии:

$$D(X)=E\left(X^2\right)-\left[E(X)\right]^2,$$

которая при известном законе распределения принимает вид:

$$D(X)=\left(\sum_{i=1}^{n}(x_i)^2\cdot p_i\right)-\left(\sum_{i=1}^{n}x_i\cdot p_i\right)^2.$$

Кроме того, как следствие определения дисперсии и свойств математического ожидания имеет место равенство:

$$E\left(X^2\right)=D(X)+\left[E(X)\right]^2,$$

которое часто применяется при решении задач.

Пример 2.8. Найти дисперсию величины, закон распределения которой имеет вид:

Значение X	–1	0	1	2
Вероятность p_i	0,2	0,2	0,3	0,3

Решение.

Первый способ (по определению). Вычислим математическое ожидание:

$$E(X)=\sum_{i=1}^{n}x_i\cdot p_i=-1\cdot 0{,}2+0\cdot 0{,}2+1\cdot 0{,}3+2\cdot 0{,}3=0{,}7\ .$$

Составим закон распределения на квадрат отклонения величины от своего математического ожидания:

Значение $\left(X-E(X)\right)^2$	$(-1-0{,}7)^2$	$(0-0{,}7)^2$	$(1-0{,}7)^2$	$(2-0{,}7)^2$
Вероятность p_i	0,2	0,2	0,3	0,3

Теперь вычислим математическое ожидание этой величины:

$$E\left(\left(X-E(X)\right)^2\right)=$$
$$=(-1-0,7)^2\cdot 0,2+(0-0,7)^2\cdot 0,2+(1-0,7)^2\cdot 0,3+(2-0,7)^2\cdot 0,3=$$
$= 2,89{\cdot}0,2 + 0,49{\cdot}0,2 + 0,09{\cdot}0,3 + 1,69{\cdot}0,3 =$
$= 0,578 + 0,098 + 0,027 + 0,507 = 1,21.$

По определению $D(X)=E\left(\left(X-E(X)\right)^2\right)$, поэтому $D(X)=1{,}21$.

Второй способ.
Воспользуемся расчетной формулой

$$D(X)=\left(\sum_{i=1}^{n}(x_i)^2\cdot p_i\right)-\left(\sum_{i=1}^{n}x_i\cdot p_i\right)^2:$$

$$D(X)=\left(\sum_{i=1}^{n}(x_i)^2\cdot p_i\right)-\left(\sum_{i=1}^{n}x_i\cdot p_i\right)^2=$$

$$=\left((-1)^2\cdot 0,2+0^2\cdot 0,2+1^2\cdot 0,3+2^2\cdot 0,3\right)-\left(-1\cdot 0,2+0\cdot 0,2+1\cdot 0,3+2\cdot 0,3\right)^2=$$

$= 1,7–0,7^2 = 1,7–0,49 =$ **1,21.**

Видим, что значение дисперсии не зависит от способа вычисления, но второй способ значительно удобнее.

Ответ: $D(X)=1{,}21$.

Перечислим основные ***свойства дисперсии*** случайной величины:

1. Дисперсия постоянной величины равна нулю:

$$D(C)=0\ ,\ C \text{ — const}.$$

2. Постоянный множитель из-под знака дисперсии выносится в квадрате:

$$D(kX)=k^2\cdot D(X)\ .$$

3. Дисперсия суммы или разности **независимых** величин всегда равна сумме дисперсий этих величин:

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

Дисперсия величины имеет размерность, не совпадающую с размерностью самой величины. Для того чтобы иметь характеристику для рассеивания величины той же размерности, что и сама величина, определили ***среднее квадратическое отклонение*** (стандартное) как квадратный корень из дисперсии:

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

Понятно, что если для величины не существует математическое ожидание, то не существует и дисперсия, а, следовательно, и среднее квадратическое (стандартное) отклонение.

К основным числовым характеристикам дискретной случайной величины относят начальные и центральные моменты порядка k.

Начальным моментом порядка k называют математическое ожидание k-й степени величины:

$$\nu_k = E\left(X^k\right).$$

Например, начальным моментом первого порядка является математическое ожидание величины: $\nu_1 = E\left(X\right)$, а начальным моментом второго порядка — математическое ожидание квадрата величины: $\nu_2 = E\left(X^2\right)$.

Центральным моментом порядка k называют математическое ожидание k-й степени отклонения величины от своего математического ожидания:

$$\mu_k = E\left(\left(X - E\left(X\right)\right)^k\right).$$

Центральный момент первого порядка всегда равен нулю:

$$\mu_1 = E\left(X - E\left(X\right)\right) = 0.$$

Центральным моментом второго порядка является дисперсия величины:

$$\mu_2 = E\left(\left(X - E\left(X\right)\right)^2\right) = D\left(X\right).$$

Для вычисления центральных моментов более высоких порядков используют следующие тождества:

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^2; \quad \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$

Центральные моменты третьего и четвертого порядков необходимы для определения асимметрии и эксцесса распределения величины.

Асимметрия определяется отношением вида

$$AsX = \frac{\mu_3}{\sigma^3},$$

где $\sigma = \sigma\left(X\right)$.

Эксцесс определяется как

$$EsX = \frac{\mu_4}{\sigma^4} - 3,$$

где $\sigma = \sigma(X)$.

При использовании некоторых свойств математического ожидания и дисперсии существенным условием являлась независимость величин. Поэтому нам осталось определить числовые характеристики, указывающие на зависимость величин, перечислить их свойства, а также уточнить свойства математического ожидания и дисперсии для зависимых величин.

Ковариацией величин X и Y называют математическое ожидание произведения отклонений величин от своих математических ожиданий:

$$\text{Cov}(X,Y) = E\left(\left(X - E(X)\right)\cdot\left(Y - E(Y)\right)\right).$$

Отметим основные ***свойства ковариации***:

1. $\text{Cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$.

2. $\text{Cov}(X,X) = D(X)$.

3. Если X и Y независимы, то $\text{Cov}(X,Y) = 0$.

4. $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

5. $\text{Cov}(\alpha X,Y) = \text{Cov}(X,\alpha Y) = \alpha\text{Cov}(X,Y)$.

6. $\text{Cov}(X+Y,Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$.

7. $\text{Cov}(X,Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$.

Если $\text{Cov}(X,Y) = 0$, то случайные величины X и Y называются ***некоррелированными***. Таким образом, по свойству **3** из независимости X и Y следует их некоррелированность. **Обратное неверно.**

Ковариация имеет размерность произведения величин. Для характеристики зависимости удобнее использовать безразмерную числовую характеристику, называемую ***коэффициентом корреляции***:

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}.$$

Важными являются следующие ***свойства коэффициента корреляции***:

1) $\rho(X,Y) = \rho(Y,X)$;

2) $\left|\rho(X,Y)\right| \le 1$;

3) $\left(\left|\rho(X,Y)\right| \le 1\right) \Leftrightarrow \left(Y = aX + b\right)$, $\quad a,\ b$ — const.

Уточним теперь свойство 4 математического ожидания и свойство 3 дисперсии величин, которые имеют место только для независимых величин. В случае зависимости величин эти свойства принимают вид:

4*. Математическое ожидание произведения **зависимых** величин равно сумме произведения математических ожиданий этих величин и их ковариации:

$$E(X\cdot Y) = E(X)\cdot E(Y) + \mathrm{Cov}(X,Y).$$

3*. Дисперсия суммы **зависимых** величин вычисляется по формуле:

$$D(X+Y) = D(X) + D(Y) + 2\mathrm{Cov}(X,Y).$$

Дисперсия разности **зависимых** величин вычисляется по формуле:

$$D(X-Y) = D(X) + D(Y) - 2\mathrm{Cov}(X,Y).$$

Пример 2.9. Найти $D(2X-3Y+5)$, если известны дисперсии $D(X)=4$, $D(Y)=1$ и коэффициент корреляции $\rho(X,Y)=0{,}2$.

Решение.

Используя свойства дисперсии и ковариации, а также определение коэффициента корреляции, преобразуем:

$$D(2X-3Y+5)=$$

$$= D(2X-3Y) = D(2X) + D(3Y) - 2\mathrm{Cov}(2X,3Y) =$$

$$= 2^2 D(X) + 3^2 D(Y) - 2\cdot 2\cdot 3\cdot \mathrm{Cov}(X,Y) =$$

$$= 4D(X) + 9D(Y) - 12\cdot\rho(X,Y)\cdot\sigma(X)\cdot\sigma(Y) =$$

$$= 4D(X) + 9D(Y) - 12\cdot\rho(X,Y)\cdot\sqrt{D(X)}\cdot\sqrt{D(Y)}.$$

С учетом данных условия задачи получим:

$$D(2X-3Y+5) = 4D(X) + 9D(Y) - 12\cdot\rho(X,Y)\cdot\sqrt{D(X)}\cdot\sqrt{D(Y)} =$$

$$= 4\cdot 4 + 9\cdot 1 - 12\cdot 0{,}2\cdot\sqrt{4}\cdot\sqrt{1} = 16 + 9 - 4{,}8 = 20{,}2.$$

Ответ: $D(2X-3Y+5) = 20{,}2$.

Задачи для самостоятельного решения

В задачах **2.21–2.30** необходимо использовать определение или свойства математического ожидания случайной величины.

2.21. Случайная величина X имеет закон распределения:

Значение X	–3	–1	0	1	3
Вероятность	0,1	0,1	0,2	0,2	0,4

Найти математические ожидания $E(X)$ и $E(X^2)$.

2.22. Случайная величина X имеет закон распределения:

Значение X	–1	0	1	2	3	4	5	6
Вероятность	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1

Найти математические ожидания $E(|X|)$ и $E\left(|X|^2\right)$.

2.23. Случайная величина X имеет закон распределения:

Значение X	–2	–1	0	1	2
Вероятность	0,1	0,2	0,2	0,2	0,3

Найти математические ожидания $E\left(2^X\right)$ и $E\left(\left(2^X\right)^2\right)$.

2.24. Случайная величина X имеет закон распределения:

Значение X	–1	0	1	2	4	8	16	32
Вероятность	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1

Найти математическое ожидание $E(X)$ и вероятность $P(X > E(X))$.

2.25. Случайная величина X имеет закон распределения:

Значение X	–3	–2	–1	0	1	2	3	4
Вероятность	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3

Найти математическое ожидание $E(X)$ и вычислить вероятность $P(|X - E(X)| < 2)$.

2.26. Известны законы распределений независимых дискретных случайных величин X и Y:

Значение X	–1	0
Вероятность	0,1	0,9

Значение Y	0	1	2
Вероятность	0,2	0,3	0,5

Найти $E\left(X^2 + 2Y - XY + 5\right)$.

2.27. Известны законы распределений независимых дискретных случайных величин X и Y:

Значение X	−1	0	1
Вероятность	0,1	0,8	0,1

Значение Y	0	1	2
Вероятность	0,1	0,3	0,6

Найти $E\left(\left(X - Y\right)^2\right)$.

2.28. Случайные величины X и Y независимы и могут принимать только значения 1 и 2, причем известно, что $P\left(X = 1\right) = 0{,}2$ и $P\left(Y = 2\right) = 0{,}3$. Найти $E\left(3^{X-Y}\right)$ и $E\left(XY\right)$.

2.29. Величины $X_1, X_2, \ldots, X_{10}$ независимы и могут принимать только целые значения от -10 до 11 с одинаковой вероятностью. Найти $E\left(X_1 \cdot X_2 \cdot \ldots \cdot X_{10}\right)$.

2.30. Случайные величины X, Y и Z независимы и могут принимать только значения 1 и 2 с одинаковой вероятностью. Найти $E\left(X^2 + Y^2 + Z^2\right)$ и $E\left(\left(X + Y\right) \cdot \left(Y + Z\right)\right)$.

В задачах **2.31–2.40** необходимо использовать определение или свойства дисперсии случайной величины.

2.31. Случайная величина X имеет закон распределения:

Значение X	−2	−1	0	1	2
Вероятность	0,1	0,2	0,1	0,2	0,4

Найти дисперсию $D\left(X\right)$ и стандартное (среднеквадратическое) отклонение $\sigma\left(X\right)$.

2.32. Случайная величина X имеет закон распределения:

Значение X	−1	0	1	2	3	4	5	6
Вероятность	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1

Найти вероятность $P\left(\left|X - E\left(X\right)\right| < \sigma\left(X\right)\right)$.

2.33. Известны законы распределений независимых дискретных случайных величин X и Y:

Значение X	-1	0
Вероятность	0,3	0,7

Значение Y	-1	0	1
Вероятность	0,3	0,3	0,4

Найти дисперсию $D(2X-3Y+5)$.

2.34. Известны законы распределений независимых дискретных случайных величин X и Y:

Значение X	-1	0	1
Вероятность	0,2	0,6	0,2

Значение Y	0	2	4
Вероятность	0,1	0,4	0,5

Найти дисперсию $D(XY-2)$.

2.35. Известно, что случайные величины X и Y независимы и имеют равные математические ожидания $E(X)=E(Y)=2$ и дисперсии $D(X)=D(Y)=1$. Найти $E\left(X^2-Y^2\right)$.

2.36. Случайные величины X и Y независимы и могут принимать только значения 0 и 1, причем известно, что $P(X=0)=0{,}2$ и $P(Y=1)=0{,}2$. Найти $E\left((X-Y)^2\right)$.

2.37. Величины $X_1, X_2, \ldots, X_{10}$ независимы и имеют одинаковые математические ожидания $E(X_1)=E(X_2)=\ldots=E(X_{10})=1$ и дисперсии $D(X_1)=D(X_2)=\ldots=D(X_{10})=1$. Найти $E\left(X_1^2 \cdot X_2^2 \cdot \ldots \cdot X_{10}^2\right)$.

2.38. Величины $X_1, X_2, \ldots, X_{10}$ независимы и могут принимать только значения -1 и 1, причем для всех величин X_i $(i=1,2,\ldots,10)$ известно, что $P(X_i=-1)=0{,}9$. Найти $E\left((X_1+X_2+\ldots+X_{10})^2\right)$.

2.39. Случайные величины X_1, X_2, X_3 имеют одинаковые законы распределения:

Значение X_i $(i=1,2,3)$	1	2
Вероятность	0,5	0,5

Для величины $Z = X_1 + X_2 + X_3$ найти вероятность $P\left(\left|Z - E(Z)\right| < \sigma(Z)\right)$.

2.40. Для случайной величины X известно, что математическое ожидание $E(X) = 5$ и дисперсия $D(X) = 109$, кроме того, известно, что $E(|X|) = 8$. Найти $D(|X|)$.

В задачах **2.41–2.50** необходимо учитывать зависимость величин при использовании свойств математического ожидания и дисперсии.

2.41. Известно, что для случайных величин X и Y $\mathrm{Cov}(X,Y) = 2$, $E(X) = -1$, $E(Y) = 1$. Найти математическое ожидание произведения $E(XY)$.

2.42. Для случайных величин X и Y известно, что $\mathrm{Cov}(X,Y) = -1$, $D(X) = 2$, $D(Y) = 3$. Найти дисперсию $D(X+Y)$.

2.43. Для случайных величин X и Y известно, что $\mathrm{Cov}(X,Y) = 3$, $D(X) = 1$, $D(Y) = 2$. Найти дисперсию $D(2X - 3Y)$.

2.44. Для случайных величин X и Y известно, что $D(X) = 1$, $D(Y) = 2$ и $\rho(X,Y) = 0{,}2$. Найти дисперсию $D(X - Y)$.

2.45. Для случайных величин X и Y известно, что $D(X) = 2$, $D(Y) = 5$ и $\rho(X,Y) = -0{,}9$. Найти дисперсию $D(3X - 2Y)$.

2.46. Случайные величины X и Y заданы законами распределений:

Значение X	-1	1
Вероятность	0,1	0,9

Значение Y	-2	0	2
Вероятность	0,3	0,4	0,3

Найти ковариацию $\mathrm{Cov}(X,Y)$, если известно, что $\rho(X,Y) = 0{,}7$.

2.47. Случайные величины X и Y заданы законами распределений:

Значение X	-2	0	2
Вероятность	0,2	0,3	0,5

Значение Y	1	2	3
Вероятность	0,1	0,2	0,7

Найти $E\left(\left(X+Y\right)^2\right)$, если коэффициент корреляции $\rho\left(X,Y\right) = -0{,}3$.

2.48. Случайные величины X и Y могут принимать только значения 1 и 2, при этом $P\left(X=1\right)=0{,}4$, $P\left(Y=1\right)=0{,}1$. Найти $E\left(\left(5X-4Y\right)^2\right)$, если известно, что $\rho\left(X,Y\right) = -0{,}4$.

2.49. Для случайных величин X и Y известно, что $D\left(X\right)=1$, $D\left(Y\right)=2$ и $\rho\left(X,Y\right)=0{,}6$. Найти дисперсию суммы и разности величин V и W, если $V=2X-Y$, $W=X-2Y$.

2.50. Известны числовые характеристики случайных величин X и Y: $E\left(X\right)=-1$, $D\left(X\right)=1$, $E\left(Y\right)=-2$, $D\left(Y\right)=2$ и $\rho\left(X,Y\right)=0{,}5$. Найти математическое ожидание произведения величин V и W, если $V=3X-Y$, $W=X+3Y$.

2.4. Непрерывные случайные величины

Если множество значений случайной величины представляет собой часть или даже все множество действительных чисел, то случайная величина определяется как ***непрерывная***. Понятно, что закон распределения в виде таблицы для такой величины не составить, поэтому непрерывные случайные величины задаются с помощью функции, устанавливающей связь между возможными значениями величины и вероятностями, с которыми эти значения принимаются.

Функцией распределения случайной величины X называется функция, определяемая условием:

$$F\left(x\right)=P\left(X<x\right).$$

Областью определения функции распределения является все множество действительных чисел. Среди свойств этой функции отметим наиболее важные:

1. $F\left(x\right)$ – ограниченная функция: $0\le F\left(x\right)\le 1$.

2. $F\left(x\right)$ — неубывающая функция: $\left(x_1<x_2\right)\Rightarrow\left(F\left(x_1\right)\le F\left(x_2\right)\right)$.

3. Поведение на бесконечности определяется пределами: $\lim\limits_{n\to-\infty} F\left(x\right)=0$; $\lim\limits_{n\to+\infty} F\left(x\right)=1$.

4. $F(x)$ — непрерывна слева в любой точке.

Для непрерывной случайной величины линия графика — сплошная.

Функцию распределения можно задавать и для дискретных случайных величин с известным законом распределения:

Значение X	x_1	x_2	x_3	x_4	...	x_n
Вероятность p_i	p_1	p_2	p_3	p_4	...	p_n

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

В этом случае график функции распределения не будет сплошной линией, функция распределения дискретной случайной величины имеет разрывы в виде скачков в точках возможных значений величины (рис. 2.1).

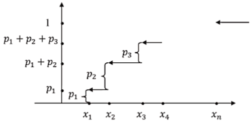


Рис. 2.1. График функции распределения дискретной случайной величины

Аналитически функция распределения дискретной случайной величины задается по принципу «накопленных вероятностей»:

$$F(x) = \begin{cases} 0, & x \le x_1 \\ p_1, & x_1 < x \le x_2 \\ p_1 + p_2, & x_2 < x \le x_3 \\ p_1 + p_2 + p_3, & x_3 < x \le x_4 \\ \dots\dots\dots\dots\dots\dots\dots\dots & \\ 1, & x > x_n \end{cases}$$

Конечно, при работе с дискретными случайными величинами удобнее использовать закон распределения в виде таблицы, но это не значит, что для таких величин функция распределения не определяется. Можно сказать, что общим способом задания случайной величины является указание вида ее функции распределения, и если эта функция имеет разрывы в виде скачков, то она определяет дискретную величину; если же функция распределения — непрерывная, то и определяет она непрерывную случайную величину.

Пример 2.10. Для заданной дискретной случайной величины с известным законом распределения:

Значение X	–1	0	1	2
Вероятность p_i	0,2	0,2	0,3	0,3

задать функцию распределения.

Решение.

Используем принцип «накопления вероятностей»:

$$F(x)=\begin{cases} 0, & x\le -1 \\ 0{,}2, & -1<x\le 0 \\ 0{,}2+0{,}2, & 0<x\le 1 \\ 0{,}2+0{,}2+0{,}3, & 1<x\le 2 \\ 1, & x>2 \end{cases}.$$

Ответ: $F(x)=\begin{cases} 0, & x\le -1 \\ 0{,}2, & -1<x\le 0 \\ 0{,}4, & 0<x\le 1 \\ 0{,}7, & 1<x\le 2 \\ 1, & x>2 \end{cases}$

Функция распределения позволяет ответить на вопросы, связанные принятием величиной определенного числового значения или попаданием значения величины в определенный интервал:

1. $P(X=x_1)=0$.
2. $P(x_1<X<x_2)=P(x_1\le X\le x_2)=$
   $=P(x_1\le X<x_2)=P(x_1<X\le x_2)=F(x_2)-F(x_1)$.
3. $P(X<x_1)=F(x_1)$.
4. $P(X>x_1)=1-P(X<x_1)=1-F(x_1)$.

Пример 2.11. Функция распределения величины X имеет вид:

$$F(x)=\begin{cases}0, & x\le 4\\ a(x-4), & 4<x\le 8\\ 1, & x>8\end{cases}$$, где a — некоторое число (параметр).

Найти значение параметра a и $P(5<X<9)$.

Решение.

Для вычисления параметра используем непрерывность функции распределения: $F(4)=0$ и $F(8)=1$.

В результате получаем:

$F(4)=0 \Rightarrow a(4-4)=0$ — тождество,

$F(8)=1 \Rightarrow a(8-4)=1 \Rightarrow 4a=1 \Rightarrow a=\dfrac{1}{4}$.

Для ответа на второй вопрос задачи воспользуемся условием **2**:

$P(x_1<X<x_2)=F(x_2)-F(x_1)$.

Тогда получим: $P(5<X<9)=F(9)-F(5)=$

$=1-\dfrac{1}{4}\cdot(5-4)=1-\dfrac{1}{4}=\dfrac{3}{4}=0{,}75$.

Ответ: $a=\dfrac{1}{4}$, $P(5<X<9)=0{,}75$.

Непрерывная случайная величина X называется ***абсолютно непрерывной***, если существует функция $f(x)$, удовлетворяющая условию:

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

Функция называется ***плотностью*** распределения X (или плотностью вероятности распределения X).

Вспомнив «геометрию» определенного интеграла, можно наглядно представить условие из определения (рис. 2.2):

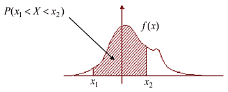


Рис. 2.2. Вероятность попадания значений величины в заданный интервал

Из свойств плотности вероятности непрерывной величины перечислим наиболее важные:

1. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ — свойство нормированности плотности.

2. $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ — функцию распределения $F(x)$ можно «восстанавливать», используя функцию плотности вероятности $f(x)$.

3. $\int\limits_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1)$.

Пример 2.12. Для непрерывной величины X функция плотности вероятности имеет вид:

$$f(x) = \begin{cases} 0, & x \le 3 \\ a(x-3), & 3 < x \le 5 \\ 0, & x > 5 \end{cases}, \text{ где } a \text{ — некоторое число (параметр)}$$

Найти значение параметра a и $P(4 < X < 8)$.

Решение.

Для вычисления параметра плотности вероятности используем свойство нормированности плотности: $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$.

В нашем случае имеем:

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{-\infty}^{3} f(x)dx + \int\limits_{3}^{5} f(x)dx + \int\limits_{5}^{+\infty} f(x)dx = \int\limits_{3}^{5} a(x-3)dx = 1.$$

Вычисляя последний интеграл, получим:

$$\int\limits_{3}^{5} a(x-3)dx = a \cdot \left.\frac{(x-3)^2}{2}\right|_3^5 = \frac{a}{2}(4-0) = 2a = 1 \;\Rightarrow\; a = \frac{1}{2} = 0{,}5.$$

Для ответа на второй вопрос задачи, используем определение плотности вероятности $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} f(x)dx$.

В нашем случае получим:

$P(4 < X < 8) =$

$$= \int\limits_{4}^{8} f(x)dx = \int\limits_{4}^{5} f(x)dx + \int\limits_{5}^{8} f(x)dx = \int\limits_{4}^{5} \frac{1}{2}(x-3)dx = \frac{1}{2} \cdot \left.\frac{(x-3)^2}{2}\right|_4^5 =$$

$$= \frac{1}{4} \cdot (4 - 1) = \frac{3}{4} = 0{,}75 .$$

Ответ: $a = 0{,}5$, $P(4 < X < 8) = 0{,}75$.

Числовые характеристики непрерывных случайных величин находят, применяя интегрирование:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx ,$$

$$D(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 \cdot f(x)dx ,$$

$$\nu_k = E(X^k) = \int_{-\infty}^{+\infty} x^k \cdot f(x)dx ,$$

$$\mu_k = E\Big((X - E(X))^k\Big) = \int_{-\infty}^{+\infty} (x - E(X))^k \cdot f(x)dx$$

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - E(X)) \cdot (y - E(Y))dxdy$$

Все рассмотренные свойства математического ожидания, дисперсии и ковариации сохраняются. Уточним только свойство для математического ожидания величины, функционально зависящей от X:

$$E(\phi(X)) = \int_{-\infty}^{+\infty} \phi(x) \cdot f(x)dx .$$

Понятно, что эти числовые характеристики для непрерывной случайной величины существуют только в случае сходимости несобственных интегралов, на основе которых они определяются.

Пример 2.13. Для непрерывной величины X функция плотности вероятности имеет вид:

$$f(x) = \begin{cases} 0, & x \le 3 \\ \frac{1}{2} \cdot (x - 3), & 3 < x \le 5 \\ 0, & x > 5 \end{cases}$$

Найти математическое ожидание и дисперсию величины X.

Решение.

Для вычисления математического ожидания применим формулу

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx\ .$$

В нашем случае имеем:

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx\ =$$

$$=\int_{-\infty}^{3} x\cdot f(x)dx+\int_{3}^{5} x\cdot f(x)dx+\int_{5}^{+\infty} x\cdot f(x)dx=\int_{3}^{5}\frac{1}{2}x(x-3)dx=$$

$$=\frac{1}{2}\int_{3}^{5}(x^2-3x)dx=\frac{1}{2}\cdot\left(\frac{x^3}{3}-\frac{3x^2}{2}\right)\Bigg|_3^5=\frac{1}{2}\cdot\left((\frac{125}{3}-\frac{75}{2})-(\frac{27}{3}-\frac{27}{2})\right)=\frac{13}{3}.$$

Для вычисления дисперсии применим формулу:

$$D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2\cdot f(x)dx\ .$$

В нашем случае получим:

$$D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2\cdot f(x)dx\ =$$

$$=\int_{-\infty}^{3}(x-\frac{13}{3})^2\cdot f(x)dx+\int_{3}^{5}(x-\frac{13}{3})^2\cdot f(x)dx+\int_{5}^{+\infty}(x-\frac{13}{3})^2\cdot f(x)dx=$$

$$=\int_{3}^{5}(x-\frac{13}{3})^2\cdot\frac{1}{2}(x-3)dx=\frac{1}{2}\int_{3}^{5}(x^2-\frac{26}{3}x+\frac{169}{9})\cdot(x-3)dx=$$

$$=\frac{1}{2}\int_{3}^{5}(x^3-\frac{26}{3}x^2+\frac{169}{9}x-3x^2+26x-\frac{169}{3})dx=$$

$$=\frac{1}{2}\left(\frac{x^4}{4}-\frac{35x^3}{9}+\frac{403x^2}{18}-\frac{169}{3}x\right)\Bigg|_3^5=\frac{1}{2}\left(\left(\frac{625}{4}-\frac{35\cdot125}{9}+\frac{403\cdot25}{18}-\frac{169\cdot5}{3}\right)-\right.$$

$$\left.-\left(\frac{81}{4}-\frac{35\cdot27}{9}+\frac{403\cdot9}{18}-\frac{169\cdot3}{3}\right)\right)=\frac{2}{9}.$$

Ответ: $E(X)=\dfrac{13}{3}$, $D(X)=\dfrac{2}{9}$.

Задачи для самостоятельного решения

В задачах **2.51–2.60** необходимо, используя данный закон распределения дискретной случайной величины *X*:

— задать функцию распределения $F(x)$ этой величины;
— построить ее график;
— найти значение $F(x_0)$ в указанной точке x_0.

2.51.

Значение X	-1	0	1	2	3	4
Вероятность	0,1	0,2	0,2	0,2	0,2	0,1

При $x_0 = 4{,}3$ вычислить значение функции $F(4{,}3)$.

2.52.

Значение X	-2	-1	0	1	2	3
Вероятность	0,2	0,2	0,1	0,2	0,1	0,2

При $x_0 = -0{,}3$ вычислить значение функции $F(-0{,}3)$.

2.53.

Значение X	-2	-1	$-0{,}5$	0	0,5	1
Вероятность	0,1	0,1	0,1	0,1	0,1	0,5

При $x_0 = -0{,}1$ вычислить значение функции $F(-0{,}1)$.

2.54.

Значение X	1	2	3	4	5	6
Вероятность	0,1	0,2	0,1	0,1	0,1	0,4

При $x_0 = 5{,}1$ вычислить значение функции $F(5{,}1)$.

2.55.

Значение X	-3	-2	-1	0	3	4
Вероятность	0,4	0,2	0,1	0,1	0,1	0,1

При $x_0 = -1{,}2$ вычислить значение функции $F(-1{,}2)$.

2.56.

Значение X	10	20	30	40	50	60
Вероятность	0,3	0,2	0,2	0,1	0,1	0,1

При $x_0 = 35$ вычислить значение функции $F(35)$.

2.57.

Значение X	-20	-15	-10	-5	0	5
Вероятность	0,1	0,1	0,1	0,1	0,1	0,5

При $x_0 = -2$ вычислить значение функции $F(-2)$.

2.58.

Значение X	$-2,1$	$-1,5$	$-0,3$	0	0,2	1,5
Вероятность	0,2	0,1	0,1	0,1	0,3	0,2

При $x_0 = -0,7$ вычислить значение функции $F(-0,7)$.

2.59.

Значение X	1,1	2,5	3,4	4	5	5,5
Вероятность	0,2	0,2	0,1	0,2	0,2	0,1

При $x_0 = 3,1$ вычислить значение функции $F(3,1)$.

2.60.

Значение X	-30	-20	-10	0	10	20
Вероятность	0,3	0,1	0,1	0,1	0,1	0,3

При $x_0 = -1,1$ вычислить значение функции $F(-1,1)$.

В задачах **2.61–2.70** необходимо:
— найти значение параметра (параметров) функции распределения $F(x)$ случайной величины X,
— построить график $F(x)$,
— найти функцию плотности вероятности $f(x)$,
— построить график $f(x)$,
— вычислить искомую вероятность.

2.61. $F(x) = \begin{cases} 0, & x \le 4 \\ a \cdot (x-4), & 4 < x \le 8 \\ 1, & x > 8 \end{cases}$, a — параметр.

Вычислить $P(5 < X < 9)$.

2.62. $F(x) = \begin{cases} 0, & x \le 1 \\ a \cdot (x-1)^2, & 1 < x \le 3 \\ 1, & x > 3 \end{cases}$, a — параметр.

Вычислить $P(X > 2)$.

2.63. $F(x) = \begin{cases} 0, & x \le 3 \\ a \cdot (x-3)^3, & 3 < x \le 4 \\ 1, & x > 4 \end{cases}$, a — параметр.

Вычислить $P(-2 < X < 2)$.

2.64. $F(x) = \begin{cases} 0, & x \le -1 \\ (x+a)^3, & -1 < x \le 0 \\ 1, & x > 0 \end{cases}$, a — параметр.

Вычислить $P\left(-\dfrac{1}{3} < X < 0\right)$.

2.65. $F(x) = \begin{cases} 0, & x \le 0 \\ a \cdot \arcsin\dfrac{x}{3}, & 0 < x \le 3 \\ 1, & x > 3 \end{cases}$, a — параметр.

Вычислить $P\left(0 < X < \dfrac{3}{2}\right)$.

2.66. $F(x) = a \cdot \mathrm{arctg}\, 2x + b$, $x \in (-\infty; +\infty)$, a, b — параметры.

Вычислить $P\left(-\dfrac{1}{2} < X < \dfrac{1}{2}\right)$.

2.67. $F(x) = \begin{cases} 0, & x \le 0 \\ ax^3 + bx, & 0 < x \le 2 \\ 1, & x > 2 \end{cases}$, a, b — параметры. Функция $F(x)$ имеет максимум при значении $x = 2$. Вычислить $P(1 < X < 3)$.

2.68. $F(x) = \begin{cases} 0, & x \le 1 \\ ax^2 + bx + c, & 1 < x \le 4 \\ 1, & x > 4 \end{cases}$, a, b и c — параметры. Функция $F(x)$ имеет максимум при значении $x = 4$. Вычислить $P(2 < X < 5)$.

2.69. $F(x) = \begin{cases} 0, & x \le 2 \\ a \cdot (x-2)^2, & 2 < x \le 6 \\ 1, & x > 6 \end{cases}$, a — параметр. Вычислить вероятность того, что при проведении 10 независимых испытаний случайная величина X три раза примет значение из интервала $(2; 3)$.

2.70. $F(x)=\begin{cases}0, & x\le 0\\ a\cdot\operatorname{arctg}\dfrac{x}{2}, & 0<x\le 2\\ 1, & x>2\end{cases}$, a — параметр. Вычислить вероятность того, что при проведении пяти независимых испытаний случайная величина X не более двух раз примет значение из интервала $(-1;1)$.

В задачах **2.71–2.80** необходимо:

— найти значение параметра (параметров) функции плотности вероятности $f(x)$ случайной величины X,

— построить график $f(x)$,

— найти функцию распределения $F(x)$,

— построить график $F(x)$,

— вычислить искомую вероятность.

2.71. $f(x)=\begin{cases}0, & x\le -3\\ a, & -3<x\le 5\\ 0, & x>5\end{cases}$, a — параметр.

Вычислить $P(-2<X<1)$.

2.72. $f(x)=\begin{cases}0, & x\le 3\\ a\cdot(x-3), & 3<x\le 5\\ 0, & x>5\end{cases}$, a — параметр.

Вычислить $P(2<X<6)$.

2.73. $f(x)=\begin{cases}0, & x\le -2\\ a\cdot x^2, & -2<x\le 2\\ 0, & x>2\end{cases}$, a — параметр.

Вычислить $P(-3<X<0)$.

2.74. $f(x)=\begin{cases}0, & x\le 0\\ a\cdot e^{-3x}, & x>0\end{cases}$, a — параметр.

Вычислить $P(3<X<5)$.

2.75. $f(x)=\begin{cases}0, & x\le 0\\ 5\cdot e^{-ax}, & x>0\end{cases}$, $a>0$ — параметр.

Вычислить $P(-1<X<3)$.

2.76. $f(x)=\dfrac{a}{1+x^2}$, $x\in(-\infty;+\infty)$, a — параметр.

Вычислить $P(-1<X<1)$.

2.77. $f(x)=\dfrac{a\cdot e^x}{1+e^{2x}}$, $x\in(-\infty;+\infty)$, a — параметр.

Вычислить $P(0<X<1)$.

2.78. $f(x)=\begin{cases}0, & x\le -3\\ ax^2+bx+c, & -3<x\le 3\\ 0, & x>3\end{cases}$, a,b и c — параметры. Известно, что на всей числовой прямой функция плотности вероятности $f(x)$ непрерывна. Вычислить $P(X<2)$.

2.79. $f(x)=\begin{cases}0, & x\le 0\\ a\cdot\sin bx+c, & 0<x\le \pi\\ 0, & x>\pi\end{cases}$, a,b и c — параметры. Известно, что на всей числовой прямой функция плотности вероятности $f(x)$ непрерывна, а значение параметра $b\in(0;2)$. Вычислить $P\left(0<X<\dfrac{2\pi}{3}\right)$.

2.80. $f(x)=\begin{cases}0, & |x|>a\\ \dfrac{1}{18}x^2, & |x|\le a\end{cases}$, $a>0$ — параметр. Вычислить $P(0<X<1)$.

В задачах **2.81–2.90** задана функция распределения $F(x)$ случайной величины X и плотность вероятности $f(y)$ случайной величины Y. Считая величины X и Y независимыми, вычислить математическое ожидание и дисперсию заданной величины Z.

2.81. $F(x)=\begin{cases}0, & x\le -2\\ \dfrac{1}{4}\cdot(x+2)^2, & -2<x\le 0\\ 1, & x>0\end{cases}$, $f(y)=\begin{cases}0, & y\le -2\\ \dfrac{1}{2}, & -2<y\le 0\\ 0, & y>0\end{cases}$, $Z=2X-Y$.

2.82. $F(x)=\begin{cases}0, & x\le 1\\ \frac{1}{49}\cdot(x-1)^2, & 1<x\le 8\\ 1, & x>8\end{cases}$, $f(y)=\begin{cases}0, & y\le 1\\ \frac{1}{7}, & 1<y\le 8\\ 0, & y>8\end{cases}$, $Z=X-3Y$.

2.83. $F(x)=\begin{cases}0, & x\le -5\\ \frac{1}{25}\cdot(x+5)^2, & -5<x\le 0\\ 1, & x>0\end{cases}$, $f(y)=\begin{cases}0, & y\le -5\\ \frac{1}{5}, & -5<y\le 0\\ 0, & y>0\end{cases}$,

$Z=3X+2Y$.

2.84. $F(x)=\begin{cases}0, & x\le 5\\ \frac{1}{16}\cdot(x-5)^2, & 5<x\le 9\\ 1, & x>9\end{cases}$, $f(y)=\begin{cases}0, & y\le 5\\ \frac{1}{4}, & 5<y\le 9\\ 0, & y>9\end{cases}$, $Z=X-Y^2$.

2.85. $F(x)=\begin{cases}0, & x\le 3\\ \frac{1}{9}\cdot(x-3)^2, & 3<x\le 6\\ 1, & x>6\end{cases}$, $f(y)=\begin{cases}0, & y\le 3\\ \frac{1}{3}, & 3<y\le 6\\ 0, & y>6\end{cases}$, $Z=X^2+2Y$.

2.86. $F(x)=\begin{cases}0, & x\le 1\\ \frac{1}{64}\cdot(x-1)^2, & 1<x\le 9\\ 1, & x>9\end{cases}$, $f(y)=\begin{cases}0, & y\le 1\\ \frac{1}{8}, & 1<y\le 9\\ 0, & y>9\end{cases}$, $Z=4X-e^Y$.

2.87. $F(x)=\begin{cases}0, & x\le 1\\ \frac{1}{6}\cdot(x^2-x), & 1<x\le 3\\ 1, & x>3\end{cases}$, $f(y)=\begin{cases}0, & y\le 1\\ \frac{1}{2}\cdot(y-1), & 1<y\le 3\\ 0, & y>3\end{cases}$,

$Z=4X^2+Y$.

2.88. $F(x)=\begin{cases}0, & x\le 1\\ \frac{1}{48}\cdot(x^3-x^2), & 1<x\le 4\\ 1, & x>4\end{cases}$, $f(y)=\begin{cases}0, & y\le 1\\ \frac{1}{6}\cdot\left(y-\frac{1}{2}\right), & 1<y\le 4\\ 0, & y>4\end{cases}$,

$Z=2X+Y^2$.

2.89. $F(x)=\begin{cases} 0, & \le 1 \\ \dfrac{}{25}\cdot\left(x-1\right) & , \quad 1<x\le 6 \\ 1, & 6 \end{cases}$, $f(y)=\begin{cases} 0, & y\le 1 \\ \dfrac{2}{25}\cdot\left(y-1\right), & 1<y\le 6 \\ 0, & y>6 \end{cases}$,

$Z=e^{X}-Y^{2}$.

2.90. $F(x)=\begin{cases} 0, & x\le -2 \\ \dfrac{1}{16}\cdot\left(x^{3}+8\right), & -2<x\le 2 \\ 1, & x>2 \end{cases}$, $f(y)=\begin{cases} 0, & y\le -2 \\ \dfrac{1}{8}\cdot\left(y+2\right), & -2<y\le 2 \\ 0, & y>2 \end{cases}$,

$Z=3X^{2}-Y^{2}$.

2.5. Специальные законы распределений случайных величин

Биномиальный закон распределения. Мы уже рассматривали ситуацию, когда осуществляется серия из n независимых испытаний с неизменной, равной $p=P\left(A\right)$, вероятностью наступления события A в каждом испытании, и говорили, что в таком случае испытания проходят по ***схеме Бернулли***. Вероятность наступления события A в k испытаниях из n осуществленных вычисляют по формуле Бернулли:

$$P_{n}(k)=C_{n}^{k}\cdot p^{k}\cdot q^{n-k},$$

где

$$C_{n}^{k}=\frac{n!}{k!\cdot(n-k)!},\quad p=P\left(A\right),\ q=1-p.$$

Понятно, что число наступлений события A в серии из n независимых испытаний будет определять дискретную случайную величину X. Закон распределения этой величины задается таблицей, в первой строке которой указываются все возможные значения величины, а во второй — соответствующие этим значениям вероятности. В качестве возможных значений надо предлагать все целые числа от 0 до n включительно (ведь событие A может не наступить ни разу, а может наступить в каждом из n испытаний). Соответствующие вероятности при этом надо вычислять по формуле Бернулли. Таким образом, закон распределения примет вид:

Значение X	0	1	2	3	...	n
Вероятность p_i	$P_n(0)$	$P_n(1)$	$P_n(2)$	$P_n(3)$	···	$P_n(n)$

Для того, чтобы таблица рассматривалась как закон распределения, необходимо выполнение условия $\sum_{i=1}^{n} p_i = 1$. Проверим его выполнимость:

$$\sum_{i=1}^{n} p_i = C_n^0 p^0 q^n + C_n^1 p^1 q^{n-1} + C_n^2 p^2 q^{n-2} + \ldots + C_n^{n-1} p^{n-1} q^1 + C_n^n p^n q^0 =$$

$$= \left(p+q\right)^n = 1^n = 1.$$

Для преобразований была использована формула, которую называют биномом Ньютона:

$$\left(a+b\right)^n = C_n^0 a^0 b^n + C_n^1 a^1 b^{n-1} + C_n^2 a^2 b^{n-2} + \ldots + C_n^{n-1} a^{n-1} b^1 + C_n^n a^n b^0,$$

поэтому распределение назвали ***биномиальным***.

Итак, величина X, определяющая число наступлений события в серии из n независимых испытаний, имеет биномиальный закон распределения.

Пример 2.14. Тест состоит из 5 вопросов, на каждый из них предлагают 3 варианта ответа, из которых только один — правильный. Составить закон распределения величины X — числа правильных ответов на вопросы теста? Тест заполняется наугад: ответы выбираются случайным образом.

Решение.

Для составления закона распределения необходимо указать все возможные значения величины и вычислить соответствующие этим значениям вероятности. Отметим, что условия опыта используют схему Бернулли: испытания независимы (в задаче не сказано, что вопросы теста связаны между собой), и в каждом из них с неизменной вероятностью $p = \frac{1}{3}$ наступает событие A: «Выбран правильный ответ», поэтому величина X — число правильных ответов на вопросы теста имеет биномиальное распределение с параметрами $n = 5$ и $p = \frac{1}{3}$:

Значение величины X	0	1	2	3	4	5
Вероятность p_i	$P_5(0)$	$P_5(1)$	$P_5(2)$	$P_5(3)$	$P_5(4)$	$P_5(5)$

Вычислим соответствующие вероятности по формуле Бернулли:

$$P_5(0)=C_5^0p^0q^5=\frac{5!}{0!\cdot(5-0)!}\cdot\left(\frac{1}{3}\right)^0\cdot\left(\frac{2}{3}\right)^5=1\cdot1\cdot\frac{32}{243}=\frac{32}{243};$$

$$P_5(1)=C_5^1p^1q^4=\frac{5!}{1!\cdot(5-1)!}\cdot\left(\frac{1}{3}\right)^1\cdot\left(\frac{2}{3}\right)^4=5\cdot\frac{1}{3}\cdot\frac{16}{81}=\frac{80}{243};$$

$$P_5(2)=C_5^2p^2q^3=\frac{5!}{2!\cdot(5-2)!}\cdot\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^3=10\cdot\frac{1}{9}\cdot\frac{8}{81}=\frac{80}{243};$$

$$P_5(3)=C_5^3p^3q^2=\frac{5!}{3!\cdot(5-3)!}\cdot\left(\frac{1}{3}\right)^3\cdot\left(\frac{2}{3}\right)^2=10\cdot\frac{1}{27}\cdot\frac{4}{9}=\frac{40}{243};$$

$$P_5(4)=C_5^4p^4q^1=\frac{5!}{4!\cdot(5-4)!}\cdot\left(\frac{1}{3}\right)^4\cdot\left(\frac{2}{3}\right)^1=5\cdot\frac{1}{81}\cdot\frac{2}{81}=\frac{10}{243};$$

$$P_5(5)=C_5^5p^5q^0=\frac{5!}{5!\cdot(5-5)!}\cdot\left(\frac{1}{3}\right)^5\cdot\left(\frac{2}{3}\right)^0=1\cdot\frac{1}{243}\cdot1=\frac{1}{243}.$$

Закон распределения величины X принимает вид:

Значение X	0	1	2	3	4	5
Вероятность p_i	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

Необходимое условие $\sum p$ выполнено.

Ответ:

Значение X	0	1	2	3	4	5
Вероятность p_i	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

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

Числовые характеристики величины X, имеющей биномиальный закон распределения, находят по формулам:

$$E(X)=np;\ D(X)=npq;\ \sigma(X)=\sqrt{npq}.$$

Мы видим, что если известны параметры опыта: n — число испытаний в серии и $p=P(A)$ — вероятность наступления события A при однократном осуществлении испытания (а, значит, и $q=P(\overline{A})=1-p$ —

вероятность ненаступления события A), то числовые характеристики легко находятся без составления самого закона распределения.

Пример 2.15. Банк выдал 700 кредитов определенного вида, оценив вероятность возврата каждого в 84%. Найти математическое ожидание и дисперсию числа возвращенных банку кредитов.

Решение.

Величина X — число возвращенных банку кредитов формируется в условиях серии испытаний, проходящих по схеме Бернулли, поэтому будет иметь биномиальное распределение с параметрами $n = 700$ и $p = 0{,}84$. Зная эти параметры, числовые характеристики легко найти:

$E(X) = np = 700 \cdot 0{,}84 = 588;$

$D(X) = npq = 700 \cdot 0{,}84 \cdot (1 - 0{,}84) = 94{,}08.$

Ответ: $E(X) = 588$; $D(X) = 94{,}08$.

Распределение Пуассона. Если испытания проходят по схеме Бернулли, но количество испытаний велико ($n \to \infty$), а вероятность наступления события A очень мала ($p \to 0$), то величина X — число наступлений события A в серии из n испытаний будет иметь распределение Пуассона:

Значение X	0	1	2	...	k	...
Вероятность p_i	$P_n(0)$	$P_n(1)$	$P_n(2)$	...	$P_n(k)$	...

Для величины, распределенной по закону Пуассона, вероятности, отвечающие определенным значениям, вычисляются по формуле Пуассона:

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

где $\lambda = np$.

Проверим выполнимость необходимого условия $\sum_{i=1}^{n} p_i = 1$ для закона распределения величины:

$$\sum_{i=1}^{n} p_i = \frac{\lambda^0}{0!} \cdot e^{-\lambda} + \frac{\lambda^1}{1!} \cdot e^{-\lambda} + \frac{\lambda^2}{2!} \cdot e^{-\lambda} + \ldots + \frac{\lambda^k}{k!} \cdot e^{-\lambda} + \ldots =$$

$$= e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^k}{k!} + \ldots\right) = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Для преобразований была использовано известное разложение функции $y = e^x$ в степенной ряд:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!} + \ldots,$$

которое при $x = \lambda$ принимает вид:

$$e^\lambda = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \ldots + \frac{\lambda^k}{k!} + \ldots$$

Поэтому условие $\sum_{i=1}^{n} p_i = 1$ для распределения Пуассона выполнено.

Пример 2.16. При печати книги в типографии вероятность появления бракованной страницы равна 0,001. Составить закон распределения для величины X — числа бракованных страниц в книге из 700 страниц и определить, что более вероятно: что книга выйдет без брака или что книга будет содержать не менее одной бракованной страницы.

Решение.

Величина X — число бракованных страниц в книге формируется в условиях испытания, проходящего по схеме Бернулли при $n = 700 \to \infty$ и $p = 0{,}001 \to 0$, поэтому будет иметь распределение Пуассона:

Значение X	0	1	2	…	k	…
Вероятность p_i	$P_{700}(0)$	$P_{700}(1)$	$P_{700}(2)$	…	$P_{700}(k)$	…

где соответствующие вероятности вычислим по формуле Пуассона:

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

при $n = 700$ $\lambda = np = 700 \cdot 0{,}001 = 0{,}7$;

$$P_{700}(0) = \frac{0{,}7^0}{0!} \cdot e^{-0{,}7} = 0{,}4966; \; P_{700}(1) = \frac{0{,}7^1}{1!} \cdot e^{-0{,}7} = 0{,}3476;$$

$$P_{700}(2) = \frac{0{,}7^2}{2!} \cdot e^{-0{,}7} = 0{,}1217; \ldots;$$

$$P_{700}(k) = \frac{0{,}7^k}{k!} \cdot e^{-0{,}7}, \ldots$$

Понятно, что все возможные значения числа бракованных страниц (от 0 до 700) в таблице просто не поместятся, тем не менее мы можем предложить искомый закон распределения:

Значение X	0	1	2	…
Вероятность p_i	0,4966	0,3476	0,1217	…

Используя ответим на второй вопрос задачи.

Условие, что книга выйдет без брака, фактически означает, что число бракованных страниц в книге равно нулю, а вероятность этого в законе указана: $P_{700}(0) = 0{,}4966$.

Условие, что книга будет содержать не менее одной бракованной страницы, предполагает, что количество таких страниц может быть равно 1, 2,…,700; вероятность этого можно вычислить как 1–0,4966 = = 0,5043.

Таким образом, более вероятно то, что книга будет содержать не менее одной бракованной страницы.

Ответ: более вероятно то, что книга будет содержать не менее одной бракованной страницы.

Числовые характеристики величины X, распределенной по закону Пуассона, находят по формулам:

$$E(X) = np = \lambda\ ;\ D(X) = np = \lambda\ ;\ \sigma(X) = \sqrt{np} = \sqrt{\lambda}\ .$$

Пример 2.17. Для величины X, распределенной по закону Пуассона, известно значение отношения вероятностей $\dfrac{P(X=8)}{P(X=7)} = 0{,}03$. Найти дисперсию величины X.

Решение:

Если величина распределена по закону Пуассона, то $D(X) = np = \lambda$.

Мы найдем неизвестное значение λ из условий:

$$P_n(8) = \frac{\lambda^8}{8!}\cdot e^{-\lambda}\ ;\ P_n(7) = \frac{\lambda^7}{7!}\cdot e^{-\lambda}$$

и заданного значения отношения вероятностей

$$\frac{P(X=8)}{P(X=7)} = 0{,}03\ .$$

Получим:

$$\frac{\dfrac{\lambda^8}{8!}e^{-\lambda}}{\dfrac{\lambda^7}{7!}e^{-\lambda}} = 0{,}03\ \Rightarrow\ \frac{\lambda}{8} = 0{,}03\ \Rightarrow\ \lambda = 0{,}24\ .$$

Следовательно,

$D(X) = \lambda = 0{,}24$.

Ответ: $D(X) = 0{,}24$.

Геометрическое и гипергеометрическое распределения. Схема испытаний Бернулли позволяет проявиться случайной величине X — числу наступлений события A, вероятность которого в рамках опыта остается постоянной в серии из n испытаний. В зависимости от условий опыта X может формироваться как биномиальное распределение или как распределение Пуассона, с ними мы только что познакомились. Оставаясь в рамках схемы Бернулли, можно наблюдать еще за одной интересной величиной X — числом испытаний, осуществленных до первого наступления события A, вероятность которого по-прежнему остается постоянной.

Такая величина имеет следующий закон распределения:

Значение X	1	2	3	...	k	...
Вероятность p_i	p	qp	$q^2 p$	...	$q^{k-1} p$	...

$P(X=1)=p$ — если величина X приняла значение, равное 1, то было осуществлено всего одно испытание, в результате которого событие A наступило с известной вероятностью $p=P(A)$ и опыт закончился.

$P(X=2)=qp$ — если величина X приняла значение, равное 2, то было осуществлено 2 испытания: в результате первого — A не наступило (с вероятностью $q=1-p$), в результате второго — A наступило с известной вероятностью $p=P(A)$ и опыт закончился.

$P(X=3)=q^2 p$ — если величина X приняла значение, равное 3, то было осуществлено 3 испытания: в результате первого — A не наступило (с вероятностью $q=1-p$), в результате второго — A не наступило (с вероятностью $q=1-p$), в результате третьего A наступило с известной вероятностью $p=P(A)$ и опыт закончился.

Понятно, что множество возможных значений величины X определяется множеством натуральных чисел, которое является бесконечным, но счетным. Поэтому X является дискретной случайной величиной.

Проверим выполнимость необходимого для закона распределения условия $\sum_{i=1}^{n} p_i = 1$:

$$\sum_{i=1}^{n} p_i = p + qp + q^2 p + \ldots + q^{k-1} p + \ldots = \frac{p}{1-q} = \frac{p}{p} = 1 .$$

Ряд $p+qp+q^2p+\ldots$ является геометрическим со знаменателем $q<1$.

Такой ряд является сходящимся, и его сумма определяется по формуле

$$S=\frac{a_1}{1-\frac{a_2}{a_1}},$$

которая в нашем случае принимает вид $S=\frac{p}{1-q}$, а поскольку $q=1-p$, то равенство единице становится очевидным.

Таким образом, условие $\sum_{i=1}^{n} p_i = 1$ для геометрического распределения выполнено. В процессе доказательства появляется геометрический ряд, что и определяет название соответствующего распределения.

Для геометрического распределения числовые характеристики находят по формулам:

$$E(X)=\frac{1}{p};\ D(X)=\frac{q}{p^2};\ \sigma(X)=\frac{\sqrt{q}}{p}.$$

Пример 2.18. Вероятность выигрыша в лотерею составляет 0,01. Петя решил покупать по одному билету в неделю этой лотереи до тех пор, пока ему не попадется выигрышный билет. Как долго Пете придется ждать выигрыша?

Решение.

Для ответа на вопрос задачи необходимо понять, что число купленных Петей билетов до первого выигрышного является случайной величиной, имеющей геометрическое распределение, и вычислить математическое ожидание этой величины:

$$E(X)=\frac{1}{p}=\frac{1}{0{,}01}=100.$$

Ответ: ожидание для Пети может продлиться 100 недель.

Пусть имеется коробка с n шарами, среди которых k белых, остальные — черные. Наугад из коробки извлекается 1 шар. Если извлеченный шар — белый, то опыт прекращается; если же извлеченный шар — черный, то он возвращается в коробку и опыт начинается снова. Величина X — количество извлечений шара из коробки до первого появления белого имеет геометрическое распределение, а ее закон имеет вид:

Значение X	1	2	3	...
Вероятность p_i	$\dfrac{k}{n}$	$\dfrac{n-k}{n}\cdot\dfrac{k}{n}$	$\left(\dfrac{n-k}{n}\right)^2\cdot\dfrac{k}{n}$	...

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

Здесь существенную роль играет условие возвращения извлеченного шара в коробку, это позволяет каждый раз начинать опыт при неизменных начальных условиях, что является важным требованием для испытаний, проходящих по схеме Бернулли.

Пусть теперь в коробке N шаров, из которых K — белых, остальные — черные. Наугад из коробки извлекают без возвращения n шаров. Составим закон распределения для величины X — числа белых шаров в выборке.

По условию естественны ограничения: $K \le N$, $n \le N$. Если обозначить k — число белых шаров в выборке, то естественно появляется еще одно ограничение: $k \le n$. Значения X определяются всеми возможными значениями k: 0, 1, 2, 3, 4,..., K. Для вычисления соответствующих вероятностей будем использовать классическую формулу и принцип выбора «белое — из белого, черное — из черного», тогда закон распределения примет вид:

Значение X	0	1	2	...	K
Вероятность	$\dfrac{C_K^0\cdot C_{N-K}^n}{C_N^n}$	$\dfrac{C_K^1\cdot C_{N-K}^{n-1}}{C_N^n}$	$\dfrac{C_K^2\cdot C_{N-K}^{n-2}}{C_N^n}$	...	$\dfrac{C_K^K\cdot C_{N-K}^{n-K}}{C_N^n}$

Условие $\sum\limits_{i=1}^{n} p_i = 1$ для такого распределения выполнено.

Полученное распределение называется ***гипергеометрическим***.

Числовые характеристики величины X, имеющей гипергеометрическое распределение, находят по формулам:

$$E(X) = n\cdot\frac{K}{N};\quad D(X) = n\cdot\frac{K}{N}\cdot\left(1-\frac{K}{N}\right)\cdot\left(1-\frac{n-1}{N-1}\right).$$

Равномерное распределение на отрезке. Перейдем к рассмотрению специальных видов распределений непрерывных случайных величин.

Напомним, что абсолютно непрерывная случайная величина X задается не только функцией распределения $F(x)=P(X<x)$, но и плотностью вероятности $f(x)=F'(x)$. Специальный вид распределения X предполагает специальный вид плотности вероятности.

Величина X ***равномерно распределена на*** некотором ***отрезке*** $[a;b]$, если ее плотность вероятности имеет вид:

$$f(x)=\begin{cases}0, x<a\\ \dfrac{1}{b-a}, a\le x\le b\\ 0, x>b\end{cases}.$$

График плотности вероятности изображен на рис. 2.3:

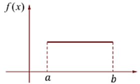


Рис. 2.3. Плотность вероятности равномерного распределения на $[a;b]$.

Функция распределения $F(x)=P(X<x)$ восстанавливается по функции плотности с использованием условия $F(x)=\int\limits_{-\infty}^{x} f(t)dt$ и принимает вид:

$$F(x)=\begin{cases}0, x<a\\ \dfrac{x-a}{b-a}, a\le x\le b\\ 1, x>b\end{cases}.$$

Можно построить и график функции распределения (рис. 2.4):

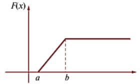


Рис. 2.4. Функция распределения равномерного распределения на $[a;b]$

Числовые характеристики для равномерного распределения на отрезке находят по соответствующим формулам для непрерывных случайных величин:

$$E(X)=\int_{-\infty}^{\infty}x\cdot f(x)dx=\int_{a}^{b}x\cdot\frac{1}{b-a}dx=$$
$$=\frac{1}{b-a}\cdot\frac{x^2}{2}\Big|_a^b=\frac{1}{b-a}\cdot\frac{b^2-a^2}{2}=\frac{a+b}{2};$$

$$D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2\cdot f(x)dx=\int_{-\infty}^{+\infty}x^2f(x)dx-\left(E(X)\right)^2=$$

$$=\int_a^b x^2\cdot\frac{1}{b-a}dx-\left(\frac{a+b}{2}\right)^2=\frac{1}{b-a}\cdot\frac{x^3}{3}\Big|_a^b-\frac{a^2+2ab+b^2}{4}=$$

$$=\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}-\frac{a^2+2ab+b^2}{4}=$$

$$=\frac{b^2+ab+a^2}{3}-\frac{a^2+2ab+b^2}{4}=\frac{4b^2+4ab+4a^2-3a^2-6ab-3b^2}{12}=$$

$$=\frac{b^2-2ab+a^2}{12}=\frac{(b-a)^2}{12};$$

$$\sigma(X)=\sqrt{\frac{(b-a)^2}{12}}=\frac{b-a}{2\sqrt{3}}.$$

Пример 2.19. Величины X и Y независимы и равномерно распределены на отрезках $[-2;4]$ и $[5;8]$ соответственно. Найти математическое ожидание и дисперсию разности этих величин.

Решение.

При вычислениях используем известные свойства математического ожидания и дисперсии, а также только что полученные числовые характеристики для равномерного распределения:

$$E(X-Y)=E(X)-E(Y)=\frac{-2+4}{2}-\frac{5+8}{2}=1-\frac{13}{2}=-\frac{11}{2}=-5{,}5\ ;$$

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

$$=\frac{(-2-4)^2}{12}+\frac{(5-8)^2}{12}=\frac{36+9}{12}=\frac{45}{12}=\frac{15}{4}=3{,}75$$

Ответ: $E(X)=-5{,}5$; $D(X)=3{,}75$.

Пример 2.20. Величина X равномерно распределена на отрезке $[-2;4]$. Найти $P(0<X<5)$.

Решение.

1 способ. Для ответа на вопрос задачи можно использовать функцию распределения величины X:

$$F(x)=\begin{cases}0, x<a\\ \dfrac{x-a}{b-a}, a\le x\le b\\ 1, x>b\end{cases},$$

которая в нашем случае примет вид: $F(x)=\begin{cases}0, x<-2\\ \dfrac{x+2}{6}, -2\le x\le 6\\ 1, x>6\end{cases}$,

и соответствующее свойство $P(x_1<X<x_2)=\ F(x_2)-F(x_1)$:

$$P(0<X<5)\ =F(5)-F(0)=\ 1-\frac{0+2}{6}=1-\frac{1}{3}=\frac{2}{3}\ .$$

2 способ. Для ответа на вопрос задачи можно использовать и функцию плотности вероятности величины X:

$$f(x)=\begin{cases}0, x<a\\ \dfrac{1}{b-a}, a\le x\le b\\ 0, x>b\end{cases},$$

которая в нашем случае примет вид: $f(x)=\begin{cases}0, x<-2\\ \dfrac{1}{6}, -2\le x\le 4\\ 0, x>4\end{cases}$, и соответствующее свойство $P(x_1 < X < x_2) = \int\limits_{x_1}^{x_2} f(x)dx$:

$$P(0 < X < 5) = \int\limits_0^5 f(x)dx = \int\limits_0^4 f(x)dx + \int\limits_4^5 f(x)dx = \int\limits_0^4 \frac{1}{6}dx = \frac{1}{6}x \Big|_0^4 = \frac{4}{6} - \frac{0}{6} = \frac{2}{3}.$$

Ответ: $P(0 < X < 5) = \dfrac{2}{3}$.

Показательное распределение. Величина X имеет показательное распределение с параметром $\lambda > 0$ в том случае, когда ее плотность вероятности имеет вид:

$$f(x) = \begin{cases} 0, & x \le 0 \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}.$$

Можно построить график плотности вероятности для такой величины (рис. 2.5):

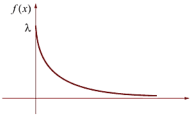


Рис. 2.5. Плотность вероятности показательного распределения

Функция распределения $F(x) = P(X < x)$ восстанавливается по функции плотности с использованием условия $F(x) = \int\limits_{-\infty}^{x} f(t)dt$:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_0^x \lambda e^{-\lambda t}dt = \lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda t}\Big|_0^x = -\left(e^{-\lambda x} - e^0\right) = 1 - e^{-\lambda x};$$

Таким образом, в итоге получаем:

$$F(x) = \begin{cases} 0, & x \le 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}.$$

График такой функции имеет вид (рис. 2.6):

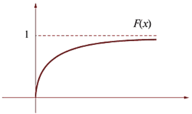


Рис. 2.6. Функция $F(x)$ показательного распределения

Найдем числовые характеристики для величины X, имеющей показательное распределение, используя соответствующие определения:

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\lim_{b\to+\infty}\int_0^b x\lambda e^{-\lambda x}dx=\left|\begin{array}{c} u=x \\ du=e^{-\lambda x} \\ du=dx \\ u=-\dfrac{1}{\lambda}e^{-\lambda x}\end{array}\right|=$$

(используем интегрирование по частям и определение несобственного интеграла)

$$=\lambda\lim_{b\to+\infty}\left(-\frac{1}{\lambda}e^{-\lambda x}\Bigg|_0^b+\frac{1}{\lambda}\int_0^b e^{-\lambda x}dx\right)=\lambda\lim_{b\to+\infty}\left(-\frac{1}{\lambda}be^{-\lambda b}-0-\frac{1}{\lambda^2}e^{-\lambda x}\Bigg|_0^b\right)=$$

$$=\lambda\lim_{b\to+\infty}\left(0-\frac{1}{\lambda^2}\left(\underbrace{e^{-\lambda b}}_{\to 1}-\underbrace{e^0}_{=1}\right)\right)=\lambda\left(-\frac{1}{\lambda^2}(-1)\right)=\frac{1}{\lambda};E(X)=\frac{1}{\lambda}.$$

В итоге имеем:

$$E(X)=\frac{1}{\lambda}.$$

Для нахождения дисперсии используем условия:

$$D(X)=\int_{-\infty}^{+\infty}(x-E(X))^2\cdot f(x)dx=\int_{-\infty}^{+\infty}x^2f(x)dx-\left(E(X)\right)^2,$$

поэтому сначала вычислим $\int_{-\infty}^{+\infty}x^2f(x)dx=E(X^2)$:

$$E\left(X^2\right)=\int\limits_0^{\infty}x^2\lambda e^{-\lambda x}dx=\lambda\lim_{b\to+\infty}\int\limits_0^{b}x^2e^{-\lambda x}dx=\left|\begin{array}{c}u=x^2\\ du=e^{-\lambda x}dx\\ du=2xdx\\ u=-\dfrac{1}{\lambda}e^{-\lambda x}\end{array}\right|=$$ (дважды используем интегрирование по частям и определение несобственного интеграла)

$$=\lambda\lim_{b\to+\infty}\left(-\frac{1}{\lambda}b^2\underbrace{e^{-\lambda b}}_{\to 0}-0+\frac{2}{\lambda}\cdot\frac{1}{\lambda^2}\right)=\lambda\cdot\frac{2}{\lambda^3}=\frac{2}{\lambda^2}.$$

Теперь подставим в формулу дисперсии и получим:

$$D(X)=\int\limits_{-\infty}^{+\infty}x^2f(x)dx-\left(E(X)\right)^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{1}{\lambda^2}.$$

Окончательно имеем:

$$D(X)=\frac{1}{\lambda^2}.$$

Для среднего квадратического (стандартного) отклонения получим:

$$\sigma(X)=\sqrt{D(X)}=\sqrt{\quad}=\frac{1}{\lambda}.$$

Пример 2.21. Величина X имеет показательное распределение и математическое ожидание, равное 2. Найти $P(-3<X<5)$.

Решение.

Если математическое ожидание для показательного распределения известно, то можно найти параметр λ, а затем и плотность вероятности, и функцию распределения случайной величины:

$$E(X)=\frac{1}{\lambda}=2\ \Rightarrow\ \lambda=\frac{1}{2};$$

$$f(x)=\begin{cases}0, & x\le 0\\ \lambda e^{-\lambda x}, & x>0\end{cases}\ \Rightarrow\ f(x)=\begin{cases}0, & x\le 0\\ \dfrac{1}{2}\cdot e^{-\frac{1}{2}x}, & x>0\end{cases};$$

$$F(x)=\begin{cases}0, & x\le 0\\ 1-e^{-\lambda x}, & x>0\end{cases}\ \Rightarrow\ F(x)=\begin{cases}0, & x\le 0\\ 1-e^{-\frac{1}{2}x}, & x>0\end{cases}.$$

Теперь можно найти ответ на вопрос задачи.

1-й способ. Для ответа на вопрос задачи можно использовать функцию распределения величины X и соответствующее свойство

$$P(x_1 < X < x_2) = F(x_2) - F(x_1).$$

Получаем:

$$P(-3 < X < 5) = F(5) - F(-3) = \left(1 - e^{-\frac{1}{2}5}\right) - 0 = 1 - e^{-\frac{5}{2}} =$$

$$= 1 - \frac{1}{e^2\sqrt{e}} = 0{,}9179.$$

2-й способ. Для ответа на вопрос задачи можно использовать функцию плотности вероятности величины X и соответствующее свойство

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx.$$

Получаем:

$$P(-3 < X < 5) = \int_{-3}^{5} f(x)dx = \int_{-3}^{0} f(x)dx + \int_{0}^{5} f(x)dx = \int_{0}^{5} \frac{1}{2} e^{-\frac{1}{2}x} dx =$$

$$= \frac{1}{2} \cdot (-2) \cdot e^{-\frac{1}{2}x} \Bigg|_{0}^{5} = -\left(e^{-\frac{5}{2}} - e^0\right) = 1 - e^{-\frac{5}{2}} = 0{,}9179.$$

Ответ: $P(-3 < X < 5) = 0{,}9179$.

Пример 2.22. Величины X и Y независимы и распределены по показательным законам. Известны $E(X) = 1$ и $E(Y) = 2$. Найти $E((X - Y)^2)$.

Решение.

Если математические ожидания для величин, имеющих показательные распределения, известны, то можно найти значение параметров λ, а также вычислить дисперсии этих величин:

$$E(X) = \frac{1}{\lambda} = 1 \;\Rightarrow\; \lambda = 1;\; D(X) = \frac{1}{\lambda^2} \;\Rightarrow\; D(X) = 1;$$

$$E(Y) = \frac{1}{\lambda} = 2 \;\Rightarrow\; \lambda = \frac{1}{2};\; D(Y) = \frac{1}{\lambda^2} \;\Rightarrow\; D(Y) = 4.$$

Теперь найдем ответ на вопрос задачи, используя свойства числовых характеристик:

$$E((X - Y)^2) = D(X - Y) + \left[E(X - Y)\right]^2 = D(X) + D(Y) + \left[E(X) - E(Y)\right]^2 =$$

$$= 1 + 4 + \left[1 - 2\right]^2 = 6.$$

Ответ: $E((X - Y)^2) = 6$.

Нормальное распределение. Величина X имеет нормальное распределение ($X \sim N(\mu, \sigma^2)$) с параметрами μ и σ, если ее плотность вероятности имеет вид:

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

Применив стандартную схему исследования, известную из курса математического анализа, можно определить экстремум, интервалы монотонности, точки перегиба и построить график плотности вероятности (рис. 2.7):

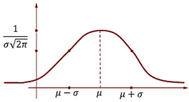


Рис. 2.7. Плотность вероятности нормального распределения $X \sim N(\mu, \sigma^2)$

Если $\mu = 0$, $\sigma = 1$, то распределение будет нормированным: $X \sim N(0, 1)$. При таких значениях параметров плотность вероятности становится известной нам функцией Гаусса:

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

а ее график — симметричным относительно оси Oy.

Функция распределения $F(x) = P(X < x)$ восстанавливается по функции плотности с использованием условия

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

После интегрирования на основе замены переменной получим:

$$F(x) = 0{,}5 + \Phi\left(\frac{x-\mu}{\sigma}\right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ — известная нам функция Лапласа, для значений которой составлены таблицы.

Зная вид функции распределения и ее свойства, можно получить важную для нормального распределения формулу:

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right).$$

Пример 2.23. Для величины $X \sim N(1, 4)$ найти $P(-1 < X < 3)$.

Решение.

Из определения нормально распределенной величины и данных условия задачи можно найти значение параметров распределения:

$X \sim N(\mu, \sigma^2) \Rightarrow X \sim N(1, 4) \Rightarrow$

$\mu = 1$ и $\sigma = \sqrt{\sigma^2} = \sqrt{4} = 2.$

Теперь по формуле

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)$$

находим ответ на вопрос задачи:

$$P(-1 < X < 3) = \Phi\left(\frac{3-1}{2}\right) - \Phi\left(\frac{-1-1}{2}\right) = \Phi(1) - \Phi(-1) =$$

$$= \Phi(1) + \Phi(1) = 2\Phi(1) = 2 \cdot 0{,}34134 = 0{,}68268.$$

При вычислениях использовано свойство нечетности функции Лапласа и таблица ее значений.

Ответ: $P(-1 < X < 3) = 0{,}68268$.

Параметры μ и σ нормального распределения величины $X \sim N(\mu, \sigma^2)$ имеют совершенно определенный вероятностный смысл, а именно:

$$\mu = E(X)\,; \ \sigma = \sqrt{D(X)}\,.$$

Поэтому если известна плотность вероятности нормально распределенной величины, то фактически сразу известны и ее математическое ожидание, и дисперсия, и среднее квадратическое (стандартное) отклонение.

Пример 2.24. Величины X и Y независимы и распределены по нормальным законам с известными плотностями вероятностей:

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-5)^2}{8}},$$

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

Найти $E\left(X^2 - Y^2\right)$.

Решение.

Найдем числовые характеристики величин X и Y, исходя из вероятностного смысла параметров:

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-5)^2}{8}} \Rightarrow E(X) = \mu = 5,\ D\left(X\right) = \sigma^2 = 4;$$

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-2)^2}{2}} \Rightarrow E(Y) = \mu = 2,\ D(Y) = \sigma^2 = 1.$$

Теперь будем искать ответ на вопрос задачи, используя свойства числовых характеристик:

$$E\left(X^2 - Y^2\right) = E(X^2) - E(Y^2) = \left(D(X) + \left[E(X)\right]^2\right) - \left(D(Y) + \left[E(Y)\right]^2\right) =$$

$$= \left(4 + 5^2\right) - \left(1 + 2^2\right) = 24.$$

Ответ: $E\left(X^2 - Y^2\right) = 24$.

Пример 2.25. Величина X имеет нормальное распределение с известной плотностью вероятности

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-5)^2}{8}}.$$

Найти $P\left(\left|X - E(X)\right| > \sigma(X)\right)$.

Решение.

Исходя из вероятностного смысла параметров, найдем числовые характеристики величины X:

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-5)^2}{8}} \Rightarrow E(X) = \mu = 5,\ \sigma = \sqrt{D(X)} = \sqrt{\sigma^2} = 2.$$

Теперь можно уточнить вопрос задачи:

$$P\left(\left|X - E(X)\right| > \sigma(X)\right) = P\left(\left|X - 5\right| > 2\right)$$

и найти на него ответ:

$$P\left(\left|X - 5\right| > 2\right) = 1 - P\left(\left|X - 5\right| < 2\right) =$$

$$= 1 - P\left(-2 < X - 5 < 2\right) = 1 - \left(P(3 < X < 7\right) =$$

$$= 1-\left(\Phi\left(\frac{7-5}{2}\right)-\Phi\left(\frac{3-5}{2}\right)\right) = 1-\left(\Phi(1)-\Phi(-1)\right) =$$

$$= 1-2\Phi(1) = 1-2\cdot 0{,}34134 = 1-0{,}68268 = 0{,}31732.$$

Ответ: $P\left(|X-5|>2\right) = 0{,}31732.$

Если X — нормально распределенная случайная величина, то и величина $Y = aX + b$, где $a,b - const$, тоже имеет нормальное распределение.

Линейная комбинация нормально распределенных величин сама является нормально распределенной величиной.

Логарифмически-нормальное распределение. Рассмотрим еще одно распределение, тесно связанное с нормальным — так называемое логарифмически-нормальное (или логнормальное) распределение. Величина X имеет ***логарифмически-нормальное распределение***, если ее логарифм $\ln X$ имеет нормальное распределение с параметрами μ и σ.

Плотность вероятности для такой величины имеет вид:

$$f(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, & x > 0 \end{cases}.$$

Функция распределения для логарифмически-нормальной случайной величины X стандартно определяется условием:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\ln x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt .$$

Числовые характеристики находят по формулам:

$$E(X) = e^{\mu+\frac{\sigma^2}{2}};\ D(X) = e^{2\mu+\sigma^2}\cdot\left(e^{\sigma^2}-1\right).$$

Произведение логарифмически-нормально распределенных случайных величин является логарифмически-нормально распределенной величиной. Это свойство является важнейшим для таких распределений.

Широкое применение логарифмически-нормальные случайные величины имеют в финансовой математике, а также в финансовой аналитике при изучении изменения цены актива в будущем. Для решения таких задач применение нормально распределенной величины нецелесообразно по причинам того, что относительно своей центральной оси оно может иметь как положительные, так и отрицательные значения, а цена актива не может быть отрицательной.

Кроме того, логарифмически-нормальное распределение асимметрично, имеет правостороннюю скошенность, что указывает на большую вероятность отклонения вверх. Это больше отвечает реалиям цены актива: инфляция оказывает давление на цены в сторону их повышения, да и временную сущность денег всегда надо иметь в виду — их вчерашняя стоимость меньше сегодняшней. Поэтому принято считать, что логарифмически-нормальное распределение лучше нормального дает первое приближение в случае активов, которыми торгуют на конкурентных рынках аукционного типа для длинных рассматриваемых периодов.

Пример 2.26. Пусть $S(n)$ обозначает цену акции к концу n-ой недели, $n \geq 1$. Предполагая, что отношения цен $\dfrac{S(n)}{S(n-1)}$, $n > 1$ являются независимыми случайными величинами, имеющими логарифмически-нормальные распределения с параметрами $\mu = 0{,}00331$ и $\sigma = 0{,}0513$, найти вероятность того, что цена акции в конце четвертой недели будет выше, чем в конце первой недели.

Решение.

Величины вида $\dfrac{S(n)}{S(n-1)}$ имеют логарифмически-нормальные распределения, это значит, что логарифмы этих величин $\ln\dfrac{S(n)}{S(n-1)}$ распределены нормально с заданными параметрами $\mu = 0{,}00331$ и $\sigma = 0{,}0513$. Кроме того, для разных значений n соответствующие величины будут независимы. Значит, для величины X можно записать:

$$X = \ln\left(\frac{S(4)}{S(3)} \cdot \frac{S(3)}{S(2)} \cdot \frac{S(2)}{S(1)}\right) = \ln\left(\frac{S(4)}{S(3)}\right) + \ln\left(\frac{S(3)}{S(2)}\right) + \ln\left(\frac{S(2)}{S(1)}\right).$$

Эта величина, как сумма независимых нормально распределенных величин, сама является нормально распределенной величиной. Параметры для нее найдем, используя свойства математического ожидания и дисперсии:

$$E(X) = \mu + \mu + \mu = 3 \cdot 0{,}00331 = 0{,}00993;$$

$$D(X) = \sigma^2 + \sigma^2 + \sigma^2 = 3 \cdot 0{,}0513^2 = 0{,}007895.$$

Теперь можем найти ответ на вопрос задачи:

$$P\left(\frac{S(4)}{S(1)}>1\right)=$$

$$=P\left(\left(\frac{S(4)}{S(3)}\cdot\frac{S(3)}{S(2)}\cdot\frac{S(2)}{S(1)}\right)>1\right)=P\left(\ln\left(\frac{S(4)}{S(3)}\right)+\ln\left(\frac{S(3)}{S(2)}\right)+\ln\left(\frac{S(2)}{S(1)}\right)>0\right)=$$

$$=P(X>0)=1-P(X<0)=1-\left(0,5+\Phi\left(\frac{0-0,00993}{\sqrt{0,007895}}\right)\right)=$$

$$=0,5-\Phi(-0,11)=0,5+\Phi(0,11)=0,5+0,0438=0,5438.$$

Ответ: $P\left(\frac{S(4)}{S(1)}>1\right)=0,5438.$

Задачи для самостоятельного решения

В задачах **2.91–2.140** при решении необходимо учитывать, что заданные случайные величины имеют распределения специального вида.

2.91. Случайная величина X распределена по биномиальному закону с параметрами $n=5$ и $p=0,2$. Найти вероятность $P(X<3)$.

2.92. Случайная величина X распределена по биномиальному закону с параметрами $n=10$ и $p=0,1$. Найти вероятность $P(|X-E(X)|<\sigma(X))$.

2.93. Случайная величина X распределена по биномиальному закону с параметрами $n=1000$ и $p=0,8$. Найти вероятность $P(X=500)$.

2.94. Случайная величина X распределена по биномиальному закону с параметрами $n=900$ и $p=0,7$. Найти вероятность $P(|X-E(X)|>3\sigma(X))$.

2.95. Случайные величины X и Y независимы и имеют биномиальные распределения с параметрами $n=20$ и $p=0,3$ для величины X и $n=30$ и $p=0,2$ для величины Y. Найти математическое ожидание и дисперсию величины $Z=2X-Y$.

2.96. Случайные величины X и Y имеют биномиальные распределения с параметрами $n=40$ и $p=0,2$ для величины X и $n=100$ и $p=0,1$ для величины Y. Найти математическое ожидание и дисперсию величины $Z = 10X - 2Y$, если известен коэффициент корреляции $\rho(X,Y) = -0{,}7$.

2.97. Случайная величина *X* распределена по закону Пуассона с параметром $\lambda = 1$. Найти вероятность $P\left(X < 2\right)$.

2.98. Случайная величина *X* распределена по закону Пуассона с параметром $\lambda = 2$. Найти вероятность $P\left(\left|X - E(X)\right| < \sigma(X)\right)$.

2.99. Величина *X* распределена по закону Пуассона. Известно, что $\dfrac{P(X=10)}{P(X=9)} = 0{,}1$. Найти дисперсию $D(X)$.

2.100. Величины X_1, X_2, X_3 независимы и имеют биномиальные распределения с одинаковыми параметрами $n = 10$ и $p = 0{,}1$. Величины Y_1, Y_2 распределены по закону Пуассона с одинаковым параметром $\lambda = 1$. Кроме того, известно, что $\rho(Y_1, Y_2) = -0{,}1$, но величины Y_1, Y_2 не зависят от величин X_1, X_2, X_3. Найти $E\left(X_1^2 + X_2^2 + X_3^2 + Y_1^2 + Y_2^2\right)$ и $E\left(\left(2\cdot(X_1 + X_2 + X_3) - 3(Y_1 + Y_2)\right)^2\right)$.

2.101. Случайная величина *X* распределена по геометрическому закону с параметром $p = 0{,}5$. Найти вероятность $P\left(X > 4\right)$.

2.102. Случайная величина *X* распределена по геометрическому закону с параметром $p = 0{,}3$. Найти вероятность $P\left(\left|X - E(X)\right| < \sigma(X)\right)$.

2.103. Величина *X* распределена по геометрическому закону. Известно, что $\dfrac{P(X=5)}{P(X=4)} = 0{,}1$. Найти дисперсию $D(X)$.

2.104. Случайные величины *X* и *Y* независимы и имеют геометрические распределения с параметрами $p = 0{,}5$ для величины *X* и $p = 0{,}4$ для величины *Y*. Найти математическое ожидание и дисперсию величины $Z = 2X - Y$.

2.105. Случайные величины *X* и *Y* имеют геометрические распределения с параметрами $p = 0{,}2$ для величины *X* и $p = 0{,}1$ для величины *Y*. Найти математическое ожидание и дисперсию величины $Z = X - 2Y$, если известен коэффициент корреляции $\rho(X,Y) = 0{,}8$.

2.106. Непрерывная случайная величина X равномерно распределена на отрезке $[-1;4]$. Найти вероятность $P(X<3)$.

2.107. Непрерывная случайная величина X равномерно распределена на отрезке $[-5;5]$. Найти вероятность $P(-1<X<7)$.

2.108. Непрерывные случайные величины X и Y независимы и равномерно распределены: величина X на отрезке $[-1;3]$, величина Y — на отрезке $[1;6]$. Найти математическое ожидание и дисперсию величины $Z=X-3Y$.

2.109. Непрерывные случайные величины X и Y равномерно распределены: величина X на отрезке $[-2;4]$, величина Y — на отрезке $[-1;8]$. Найти математическое ожидание и дисперсию величины $Z=2X+5Y$, если известен коэффициент корреляции $\rho(X,Y)=-0{,}1$.

2.110. Непрерывная случайная величина X равномерно распределена на отрезке $[0;1]$. Найти коэффициент корреляции для величин X и X^2.

2.111. Непрерывная случайная величина X равномерно распределена на отрезке $[-3;5]$. Найти математическое ожидание и дисперсию величины $\sqrt[3]{X^2}$.

2.112. Непрерывные случайные величины X и Y независимы и равномерно распределены: величина X на отрезке $[0;3]$, величина Y — на отрезке $[1;5]$. Найти дисперсию $D(XY)$ произведения этих величин.

2.113. Непрерывная случайная величина X равномерно распределена на отрезке $[-1;1]$. Найти вероятность того, что при проведении 10 независимых испытаний случайная величина X только три раза примет значение, меньшее своего математического ожидания.

2.114. Непрерывная случайная величина X равномерно распределена на отрезке $[1;3]$. Найти асимметрию $\mathrm{As}X$ величины X.

2.115. Непрерывная случайная величина X равномерно распределена на отрезке $[1;3]$. Найти эксцесс $\mathrm{Es}X$ величины X.

2.116. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 2$. Найти вероятность $P(X < 3)$.

2.117. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 5$. Найти вероятность $P(1 < X < 5)$.

2.118. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 1$. Найти вероятность $P(|X - E(X)| < \sigma(X))$.

2.119. Непрерывная случайная величина X распределена по показательному закону с математическим ожиданием $E(X) = \dfrac{2}{\ln 2}$. Найти вероятность $P(|X - E(X)| > 3\sigma(X))$.

2.120. Непрерывные случайные величины X и Y независимы и распределены по показательному закону с параметром $\lambda = 2$ для величины X и $\lambda = 1$ для величины Y. Найти ковариацию $\mathrm{Cov}(X + Y, X - Y)$ суммы и разности этих величин.

2.121. Непрерывные случайные величины X и Y распределены по показательному закону с параметром $\lambda = 1$ для величины X и $\lambda = 3$ для величины Y. Найти математическое ожидание и дисперсию величины $Z = X + 2Y$, если для величин X и Y известен коэффициент корреляции $\rho(X, Y) = -0{,}8$.

2.122. Величины $X_1, X_2, \ldots, X_{10}$ независимы и одинаково распределены по показательному закону с параметром $\lambda = 5$. Найти $E\left(X_1^2 + X_2^2 \ldots + X_{10}^2\right)$ и $E\left(\left(X_1 + X_2 + \ldots + X_{10}\right)^2\right)$.

2.123. Непрерывные случайные величины X и Y независимы. Величина X распределена по показательному закону с параметром $\lambda = 2$, величина Y равномерно распределена на отрезке $[-1; 4]$. Найти коэффициент корреляции $\rho(X \cdot Y, X + Y)$ произведения и суммы этих величин.

2.124. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 3$. Найти асимметрию $\mathrm{As}X$ величины X.

2.125. Непрерывная случайная величина X распределена по показательному закону с параметром $\lambda = 2$. Найти эксцесс $\mathrm{Es}X$ величины X.

2.126. Случайная величина X распределена по нормальному закону ($X \sim N(5,1^2)$) с параметрами $\mu = 5$ и $\sigma = 1$. Найти плотность вероятности величины X.

2.127. Случайная величина X распределена по нормальному закону ($X \sim N(2,3^2)$) с параметрами $\mu = 2$ и $\sigma = 3$. Найти функцию распределения величины X.

2.128. Случайная величина X распределена по нормальному закону: $X \sim N(1,4)$. Найти вероятность $P\left(X < 2\right)$.

2.129. Случайная величина X распределена по нормальному закону: $X \sim N(4,9)$. Найти вероятность $P\left(X > 4\right)$.

2.130. Случайная величина X распределена по нормальному закону: $X \sim N(1,4)$. Найти вероятность $P\left(\left|X - E(X)\right| < \sigma(X)\right)$.

2.131. Случайная величина X распределена по нормальному закону, известно, что $E(X) = 37{,}7$ и $P\left(X < 30\right) = 0{,}24196$. Найти дисперсию $D(X)$ величины X.

2.132. Случайная величина X распределена по нормальному закону, известно, что $D(X) = 81$ и $P\left(X < 37\right) = 0{,}97128$. Найти математическое ожидание $E(X)$ величины X.

2.133. Непрерывные случайные величины X и Y независимы и имеют нормальные распределения: $X \sim N(1,4)$, $Y \sim N(2,9)$. Найти вероятность $P\left(X + Y < 5\right)$.

2.134. Непрерывные случайные величины X и Y независимы и имеют нормальные распределения: $X \sim N(3,2)$, $Y \sim N(4,3)$. Найти вероятность $P\left(X > Y\right)$.

2.135. Непрерывные случайные величины X и Y независимы и имеют нормальные распределения: $X \sim N(1,9)$, $Y \sim N(3,4)$. Найти ковариацию $\mathrm{Cov}(X+Y, X-Y)$ суммы и разности этих величин.

2.136. Непрерывные случайные величины X и Y имеют нормальные распределения: $X \sim N(1,9)$, $Y \sim N(3,4)$. Найти математическое ожидание и дисперсию величины $Z = 2X - Y$, если для величин X и Y известен коэффициент корреляции $\rho(X,Y) = -0{,}8$.

2.137. Величины $X_1, X_2, \dots, X_8$ независимы и имеют нормальные распределения с математическими ожиданиями и дисперсиями, равными 1. Найти $E\left(X_1^2 + X_2^2 \dots + X_8^2\right)$ и $E\left(\left(X_1 + X_2 + \dots + X_8\right)^2\right)$.

2.138. Величины X_1, X_2, X_3, X_4 независимы и имеют нормальные распределения с математическими ожиданиями, равными 1, и дисперсиями, равными 2. Найти вероятность $P\left(5X_1 - X_2 - X_3 - X_4 < 2\right)$.

2.139. Непрерывная случайная величина X имеет нормальное распределение: $X \sim N(1, 4)$. Найти асимметрию $\mathrm{As}X$ величины X.

2.140. Непрерывная случайная величина X имеет нормальное распределение: $X \sim N(2, 1)$. Найти эксцесс $\mathrm{Es}X$ величины X.

2.6. Дискретные многомерные случайные величины

До сих пор мы изучали случайные величины, каждое возможное значение которых определялось одним числом. Такие величины принято называть одномерными. Однако часто приходится рассматривать испытания, результатом каждого возможного случая из которых является пара, тройка или более значений. В таких случаях определяется многомерная случайная величина, возможными значениями которой являются упорядоченные наборы чисел. Многомерные случайные величины еще принято называть векторами и обозначать $X = \left(X_1, X_2 \dots, X_n\right)$, где $X_1, X_2, \dots, X_n$ являются одномерными случайными величинами и определяются как компоненты многомерной случайной величины (вектора). Значения величины X определяются множеством наборов вида $\left(x_1, x_2, \dots, x_n\right)$, где x_1 – возможное значение величины X_1 , x_2 – возможное значение величины X_2 , ..., x_n – возможное значение величины X_n . Если все величины $X_1, X_2, \dots, X_n$ являются дискретными, то и многомерная случайная величина $X = \left(X_1, X_2 \dots, X_n\right)$ определяется как дискретная. Мы ограничимся рассмотрением двумерных дискретных случайных величин (векторов).

Упорядоченная пара случайных величин (X_1, X_2), определенных на одном и том же пространстве элементарных событий Ω, называется ***двумерной случайной величиной,*** или ***системой случайных величин,*** или ***двумерным случайным вектором***.

Пара чисел (x, y) называется возможным значением случайного вектора (X_1, X_2), если соответствует одному и тому же исходу в пространстве Ω.

Двумерную случайную величину (X_1, X_2) можно интерпретировать как случайную точку на плоскости или как случайный вектор, поэтому для нее чаще используют обозначение (X, Y) вместо (X_1, X_2). Мы тоже перейдем к такому обозначению.

Закон распределения двумерной дискретной случайной величины (X, Y) задается таблицей вида:

	$X = x_1$	$X = x_2$	...	$X = x_n$
$Y = y_1$	p_{11}	p_{21}		p_{n1}
$Y = y_2$	p_{12}	p_{22}		p_{n2}
...	...	...		...
$Y = y_k$	p_{1k}	p_{2k}		p_{nk}

Здесь $p_{ij} = P(X = x_i, Y = y_j)$ для всех возможных значений $i = 1, 2, \ldots, n$ и $j = 1, 2, \ldots, k$.

Необходимое условие для закона распределения выполнено:

$$\sum_{i=1}^{n} \sum_{j=1}^{k} p_{ij} = 1.$$

При известном законе распределения двумерной величины (X, Y) можно составить законы распределений для величин X и Y. Для этого необходимо указать возможные значения этих величин и вычислить соответствующие этим значениям вероятности.

Закон распределения величины X имеет вид:

Значение X	$X = x_1$	$X = x_2$	...	$X = x_n$
Вероятность p_i	p_1	p_2	...	p_n

Здесь

$$p_1 = P\left(X = x_1\right) = p_{11} + p_{12} + \ldots + p_{1n}\ ;\ p_2 = P\left(X = x_2\right) = p_{21} + p_{22} + \ldots + p_{2n}\ ;\ \ldots,$$

$$p_n = P\left(X = x_n\right) = p_{n1} + p_{n2} + \ldots + p_{nn}\ .$$

Необходимое условие для закона распределения $\sum_{i=1}^{n} p_i = 1$ выполнено.

Аналогично для величины Y будем иметь:

Значение Y	$Y = y_1$	$Y = y_2$	…	$Y = y_k$
Вероятности $p_j^{'}$	$p_1^{'}$	$p_2^{'}$		$p_k^{'}$

Здесь

$$p_1^{'} = P\left(Y = y_1\right) = p_{11} + p_{21} + \ldots + p_{n1}\ ;\ p_2^{'} = P\left(Y = y_2\right) = p_{12} + p_{22} + \ldots + p_{n2}\ ;\ \ldots,$$

$$p_k^{'} = P\left(Y = y_k\right) = p_{1k} + p_{2k} + \ldots + p_{nk}\ .$$

Необходимое условие для закона распределения $\sum_{j=1}^{k} p_k^{'} = 1$ и в этом случае выполнено.

Для величин X и Y можно находить числовые характеристики, используя их законы распределения.

Пример 2.27. Известен закон распределения двумерной величины (X,Y):

	$X = 10$	$X = 20$	$X = 30$	$X = 40$
$Y = 0$	0	0,2	0,1	0,1
$Y = 1$	0,3	0,2	0,1	0

Найти $E(X)$ и $D(Y)$.

Решение.

Найдем закон распределения величины X:

Значение X	10	20	30	40
Вероятность p_i	0,3	0,4	0,2	0,1

Здесь

$$p_1 = P\left(X = 10\right) = P(X = 10, Y = 0) + P(X = 10, Y = 1) = 0 + 0{,}3 = 0{,}3\ ;$$

$$p_2 = P\left(X = 20\right) = P(X = 20, Y = 0) + P(X = 20, Y = 1) = 0{,}2 + 0{,}2 = 0{,}4\ ;$$

$p_3 = P(X = 30) = P(X = 30, Y = 0) + P(X = 30, Y = 1) = 0{,}1 + 0{,}1 = 0{,}2$;

$p_4 = P(X = 40) = P(X = 40, Y = 0) + P(X = 40, Y = 1) = 0{,}1 + 0 = 0{,}1$.

Необходимое условие для закона распределения $\sum_{i=1}^{n} p_i = 1$ выполнено.

Зная закон распределения X , вычислим математическое ожидание $E(X)$:

$$E(X) = \sum_{i=1}^{4} x_i p_i = 10 \cdot 0{,}3 + 20 \cdot 0{,}4 + 30 \cdot 0{,}2 + 40 \cdot 0{,}1 = 21 .$$

Теперь найдем закон распределения величины Y :

Значение Y	0	1
Вероятности $p_j^{'}$	0,4	0,6

Здесь:

$$p_1^{'} = P(Y = 0) = P(X = 10, Y = 0) + P(X = 20, Y = 0) + \\ + P(X = 30, Y = 0) + P(X = 40, Y = 0) = \\ = 0 + 0{,}2 + 0{,}1 + 0{,}1 = 0{,}4$$

$$p_2^{'} = P(Y = 1) = P(X = 10, Y = 1) + P(X = 20, Y = 1) + \\ + P(X = 30, Y = 1) + P(X = 40, Y = 1) = \\ = 0{,}3 + 0{,}2 + 0{,}1 + 0 = 0{,}6$$

Необходимое условие $\sum_{j=1}^{k} p_k^{'} = 1$ и в этом случае выполнено.

Теперь можно вычислить дисперсию $D(Y)$:

$$D(Y) = 0^2 \cdot 0{,}4 + 1^2 \cdot 0{,}6 - (0 \cdot 0{,}4 + 1 \cdot 0{,}6)^2 = 0{,}6 - 0{,}36 = 0{,}24 .$$

Ответ: $E(X) = 21$; $D(Y) = 0{,}24$.

Если задан закон распределения двумерной величины (X,Y), то можно составить закон распределения для любой величины, полученной в результате арифметических действий с составляющими X и Y .

Пример 2.28. Известен закон распределения двумерной величины (X,Y) :

	$X = 0$	$X = 1$	$X = 3$	$X = 4$
$Y = -1$	0,1	0	0,1	0,1
$Y = 0$	0,1	0,1	0,1	0,1
$Y = 1$	0,1	0,1	0,1	0

Составить закон распределения величины $Z = X - Y$ и найти $E(Z)$.

Решение.

Найдем возможные значения разности $Z = X - Y$, это удобно выполнить, используя вспомогательную таблицу.

В нижней части каждой клетки сохраняем вероятности из заданного двумерного закона распределения, в верхней части каждой клетки считаем разность, соответствующую выбранным значениям X и Y (выделено цветом):

$Z = X - Y$	$X = 0$	$X = 1$	$X = 3$	$X = 4$
$Y = -1$	0–(–1)= **1**	1–(–1)= **2**	3–(–1)= **4**	4–(–1)= **5**
	0,1	0	0,1	0,1
$Y = 0$	0–0= **0**	1–0= **1**	3–0= **3**	4–0= **4**
	0,1	0,1	0,1	0,1
$Y = 1$	0–1= **–1**	1–1= **0**	3–1= **2**	4–1= **3**
	0,1	0,1	0,1	0

Закон распределения величины $Z = X - Y$ имеет вид:

Значения $Z = X - Y$	–1	0	1	2	3	4	5
Вероятности p_i	0,1	0,2	0,2	0,1	0,1	0,2	0,1

Здесь

$$p_1 = P(X - Y = -1) = P(X = 0, Y = 1) = 0,1;$$

$$p_2 = P(X - Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 1) = 0,1 + 0,1 = 0,2;$$

$$p_3 = P(X - Y = 1) = P(X = 0, Y = -1) + P(X = 1, Y = 0) = 0,1 + 0,1 = 0,2;$$

$$p_4 = P(X - Y = 2) = P(X = 1, Y = -1) + P(X = 3, Y = 1) = 0 + 0,1 = 0,1;$$

$$p_5 = P(X - Y = 3) = P(X = 3, Y = 0) + P(X = 4, Y = 1) = 0,1 + 0 = 0,1;$$

$$p_6 = P(X - Y = 4) = P(X = 3, Y = -1) + P(X = 4, Y = 0) = 0,1 + 0,1 = 0,2;$$

$$p_7 = P(X - Y = 5) = P(X = 4, Y = -1) = 0,1$$

Необходимое условие для закона распределения $\sum_{i=1}^{n} p_i = 1$ выполнено.

Теперь можем найти математическое ожидание $E(Z)$, используя определение и полученный закон распределения:

$E(Z) =$

$= \sum_{i=1}^{7} z_i p_i = -1 \cdot 0{,}1 + 0 \cdot 0{,}2 + 1 \cdot 0{,}2 + 2 \cdot 0{,}1 + 3 \cdot 0{,}1 + 4 \cdot 0{,}2 + 5 \cdot 0{,}1 = 1{,}9.$

Ответ: искомый закон распределения имеет вид:

Значение $Z = X - Y$	-1	0	1	2	3	4	5
Вероятности p_i	0,1	0,2	0,2	0,1	0,1	0,2	0,1

$E(Z) = 1{,}9.$

В пространстве Ω можно рассматривать разные события, связанные с тем, что компоненты X и Y величины (X,Y) принимают значения, удовлетворяющие заданным условиям, и анализировать их зависимость.

Напомним, что для независимых событий всегда выполнено условие:

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

Пример 2.29. Известен закон распределения двумерной величины (X,Y):

	$X = 0$	$X = 1$	$X = 3$	$X = 4$
$Y = -1$	0,1	0	0,1	0,1
$Y = 0$	0,1	0,1	0,1	0,1
$Y = 1$	0,1	0,1	0,1	0

Определить, будут ли зависимыми события A:"$Y < X$" и B:"$X > 0$".

Решение.

Отметим в заданной таблице «области наступления» событий A, B и их произведения $A \cdot B$ и вычислим соответствующие вероятности.

Событие A:"$Y < X$":

	$X = 0$	$X = 1$	$X = 3$	$X = 4$
$Y = -1$	0,1	0	0,1	0,1
$Y = 0$	0,1	0,1	0,1	0,1
$Y = 1$	0,1	0,1	0,1	0

$P(A) = P(Y < X) =$

$= P(Y=-1, X=0) + P(Y=-1, X=1) + P(Y=-1, X=3) + P(Y=-1, X=4) +$

$+ P(Y=0, X=1) + P(Y=0, X=3) + P(Y=0, X=4) +$

$+ P(Y=1, X=3) + P(Y=1, X=4) =$

$= 0{,}1 + 0 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0 =$ **0,7**;

Событие $B: "X > 0"$:

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

$P(B) = P(X > 0) =$

$= P(X=1, Y=-1) + P(X=3, Y=-1) + P(X=4, Y=-1) +$

$+ P(X=1, Y=0) + P(X=3, Y=0) + P(X=4, Y=0) +$

$+ P(X=1, Y=1) + P(X=3, Y=1) + P(X=4, Y=1) =$

$= 0 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0 =$ **0,7**.

Произведение событий $A: "Y < X"$ и $B: "X > 0"$:

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

$P(A \cdot B) = P(Y < X \text{ и } X > 0) =$

$= P(X=1, Y=-1) + P(X=3, Y=-1) + P(X=4, Y=-1) +$

$+ P(X=1, Y=0) + P(X=3, Y=0) + P(X=4, Y=0) +$

$+ P(X=3, Y=1) + P(X=4, Y=1) =$

$= 0 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0{,}1 + 0 =$ **0,6**.

Теперь проверим выполнимость равенства $P(A \cdot B) = P(A) \cdot P(B)$:

$0{,}6 \neq 0{,}7 \cdot 0{,}7$

Следовательно, условие независимости не выполнено, значит, события $A: "Y < X"$ и $B: "X > 0"$ являются зависимыми.

Ответ: события $A: "Y < X"$ и $B: "X > 0"$ являются зависимыми.

Если задан закон распределения двумерной величины (X, Y), то можно ответить и на вопрос о зависимости составляющих двумерную величину компонентах X и Y. Для этого считают ковариацию величин X и Y, и если ковариация не равна нулю, то величины X и Y являются зависимыми.

Если же ковариация величин X и Y равна нулю, то перед тем, как сделать вывод о независимости величин X и Y, **для каждой пары** значений $(X = x_i, Y = y_j)$ **проверяют выполнимость равенства**:

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

Пример 2.30. Известен закон распределения двумерной величины (X,Y):

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

Вычислить ковариацию $\mathrm{Cov}(X,Y)$.

Решение.

Для вычисления ковариации $\mathrm{Cov}(X,Y)$ используем свойство:

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

Чтобы найти математическое ожидание произведения, используем вспомогательную таблицу (см. ***Пример 2.28***):

$Z = X \cdot Y$	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	$0\cdot(-1)=$ **0**	$1\cdot(-1)=$ **–1**	$3\cdot(-1)=$ **–3**	$4\cdot(-1)=$ **–4**
	0,1	0	0,1	0,1
$Y=0$	$0\cdot 0=$ **0**	$1\cdot 0=$ **0**	$3\cdot 0=$ **0**	$4\cdot 0=$ **0**
	0,1	0,1	0,1	0,1
$Y=1$	$0\cdot 1=$ **0**	$1\cdot 1=$ **1**	$3\cdot 1=$ **3**	$4\cdot 1=$ **4**
	0,1	0,1	0,1	0

Для вычисления $E(X \cdot Y)$ просуммируем произведения полученных значений и соответствующих им вероятностей:

$E(X \cdot Y) = 0 \cdot 0{,}1 + (-1) \cdot 0 + (-3) \cdot 0{,}1 + (-4) \cdot 0{,}1 +$

$+\ 0 \cdot 0{,}1 + 0 \cdot 0{,}1 + 0 \cdot 0{,}1 + 0 \cdot 0{,}1 + \ + \ 0 \cdot 0{,}1 + 1 \cdot 0{,}1 + 3 \cdot 0{,}1 + 4 \cdot 0 = -0{,}3$.

Чтобы найти $E(X)$ и $E(Y)$, необходимо составить законы распределения для X и для Y, как мы это уже делали в **Примере 2.27:**

Значение X	$X=0$	$X=1$	$X=3$	$X=4$
Вероятность p_i	0,3	0,2	0,3	0,2

$E(X) = 0 \cdot 0{,}3 + 1 \cdot 0{,}2 + 3 \cdot 0{,}3 + 4 \cdot 0{,}2 = 1{,}9$.

Значение Y	$Y=-1$	$Y=0$	$Y=1$
Вероятность p_j	0,3	0,4	0,3

$E(Y) = -1 \cdot 0{,}3 + 0 \cdot 0{,}4 + 1 \cdot 0{,}3 = 0$.

Теперь ответим на вопрос задачи:

$\mathrm{Cov}(X,Y) = E(XY) - E(X) \cdot E(Y) = E(X \cdot Y) - E(X) \cdot E(Y) =$

$= -0{,}3 - 1{,}9 \cdot \quad = -0{,}3$.

Ответ: $\mathrm{Cov}(X,Y) = -0{,}3$.

В рассмотренном примере ковариация величин X и Y не равна нулю, значит, величины X и Y являются зависимыми.

Если известен закон распределения дискретной двумерной величины (X,Y), то, вычислив ковариацию величин X и Y, можно найти для них и коэффициент корреляции по известной формуле

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma(X) \cdot \sigma(Y)},$$

где

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

Задачи для самостоятельного решения

В задачах **2.141–2.150** известен закон распределения дискретной двумерной случайной величины (X,Y). Необходимо:

— составить закон распределения величины Z,

— вычислить $E(Z)$ и $D(Z)$,

— ответить на вопрос о вероятности.

2.141. Известен закон распределения двумерной величины (X,Y):

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

Составить закон распределения величины $Z=X+2Y$, найти ее математическое ожидание и дисперсию и вычислить $P(Z<3)$.

2.142. Известен закон распределения двумерной величины (X,Y):

	$X=4$	$X=5$	$X=6$	$X=7$
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0	0,1	0	0,1
$Y=2$	0,1	0,2	0,2	0

Составить закон распределения величины $Z=X-5Y$, найти ее математическое ожидание и дисперсию и вычислить $P(Z<0)$.

2.143. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,2	0,1	0,1
$Y=0$	0	0,1	0	0,2
$Y=2$	0,1	0	0,1	0

Составить закон распределения величины $Z=X\cdot Y+5$, найти ее математическое ожидание и дисперсию и вычислить $P(Z\geq 0)$.

2.144. Известен закон распределения двумерной величины (X,Y):

	$X=-3$	$X=-1$	$X=0$	$X=3$
$Y=-1$	0,1	0,2	0,1	0,2
$Y=1$	0	0,1	0	0,2
$Y=3$	0	0	0,1	0

Составить закон распределения величины $Z=X\cdot Y^2$, найти ее математическое ожидание и дисперсию и вычислить $P(Z\quad)$.

2.145. Известен закон распределения двумерной величины (X,Y):

	$X=1$	$X=2$	$X=3$	$X=4$
$Y=2$	0,1	0	0,2	0
$Y=3$	0,1	0,1	0	0,1
$Y=4$	0,1	0,1	0,2	0

Составить закон распределения величины $Z=\max(X,Y)$, найти ее математическое ожидание и дисперсию и вычислить $P(Z>3)$.

2.146. Известен закон распределения двумерной величины (X,Y):

	$X=5$	$X=6$	$X=7$	$X=8$
$Y=7$	0,1	0	0,1	0
$Y=8$	0	0,2	0	0,2
$Y=9$	0,1	0,1	0,2	0

Составить закон распределения величины $Z=\min(X,Y)$, найти ее математическое ожидание и дисперсию и вычислить $P(Z<7)$.

2.147. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,2	0	0,1	0,2
$Y=0$	0,1	0,1	0	0,1
$Y=1$	0,1	0	0,1	0

Составить закон распределения величины $Z=X^2-Y$, найти ее математическое ожидание и дисперсию и вычислить $P(X^2<Y)$.

2.148. Известен закон распределения двумерной величины (X,Y):

	$X=-4$	$X=-1$	$X=0$	$X=4$
$Y=-2$	0	0	0,1	0,2
$Y=0$	0,1	0,1	0.1	0,1
$Y=1$	0,1	0.1	0,1	0

Составить закон распределения величины $Z=\min(X+1,Y+2)$, найти ее математическое ожидание и дисперсию и вычислить $P(X<Y)$.

2.149. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=1$	0,2	0	0,1	0

Составить закон распределения величины $Z=\left|X+Y\right|$, найти ее математическое ожидание и дисперсию и вычислить $P\left(\left|Z-E(Z)\right|<\sigma(Z)\right)$.

2.150. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=2$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=2$	0,1	0	0,1	0,1

Составить закон распределения величины $Z=(X+1)\cdot(Y-2)$, найти ее математическое ожидание и дисперсию и вычислить $P\left(\left|Z-E(Z)\right|>\sigma(Z)\right)$.

В задачах **2.151–2.160** известен закон распределения дискретной двумерной случайной величины (X,Y). Необходимо вычислить математические ожидания $E(X), E(Y)$, дисперсии $D(X), D(Y)$, стандартные отклонения $\sigma(X), \sigma(Y)$, ковариацию $\mathrm{Cov}(X,Y)$ и коэффициент корреляции $\rho(X,Y)$.

2.151.

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

2.152.

	$X=4$	$X=5$	$X=6$	$X=7$
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0	0,1	0	0,1
$Y=2$	0,1	0,2	0,2	0

2.153.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,2	0,1	0,1
$Y=0$	0	0,1	0	0,2
$Y=2$	0,1	0	0,1	0

2.154.

	$X=-3$	$X=-2$	$X=0$	$X=3$
$Y=-1$	0,1	0,2	0,1	0,2
$Y=1$	0	0,1	0	0,2
$Y=3$	0	0	0,1	0

2.155.

	$X=1$	$X=2$	$X=3$	$X=4$
$Y=2$	0,1	0	0,3	0
$Y=3$	0,1	0,1	0	0,1
$Y=4$	0,1	0,1	0,1	0

2.156.

	$X=5$	$X=6$	$X=7$	$X=8$
$Y=7$	0,2	0	0,1	0
$Y=8$	0	0,2	0	0,1
$Y=9$	0,1	0,1	0,2	0

2.157.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,2	0,1	0
$Y=0$	0,1	0,1	0	0,2
$Y=1$	0,1	0	0,1	0

2.158.

	$X=-4$	$X=-1$	$X=0$	$X=4$
$Y=-2$	0	0	0,1	0,2
$Y=0$	0,1	0,1	0.1	0,1
$Y=1$	0,1	0.1	0,1	0

2.159.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=1$	0,2	0	0,1	0

2.160.

	$X=-3$	$X=-2$	$X=2$	$X=3$
$Y=-2$	0	0,3	0,1	0
$Y=0$	0	0,2	0	0,1
$Y=2$	0,2	0	0,1	0

При решении задач **2.161–2.170** для обоснования необходимо использовать определение независимых событий.

2.161. Известен закон распределения двумерной величины (X,Y):

	$X=0$	$X=1$	$X=3$	$X=3$
$Y=-1$	0,1	0	0,1	0,1
$Y=0$	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

Будут ли независимыми событие *A*: “$Y<X$” и событие *B*: “$X=Y$”?

2.162. Известен закон распределения двумерной величины (X,Y):

	$X=4$	$X=5$	$X=6$	$X=7$
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0	0,1	0	0,1
$Y=2$	0,1	0,2	0,2	0

Будут ли зависимыми событие *A*: “$Y=X-4$” и событие *B*: “*X* больше *Y* на 5”?

2.163. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,2	0,1	0,1
$Y=0$	0	0,1	0	0,2
$Y=2$	0,1	0	0,1	0

Будут ли зависимыми событие *A*: "$|X|=|Y|$" и событие *B*: "$X=Y$"?

2.164. Известен закон распределения двумерной величины (X,Y):

	$X=-3$	$X=-2$	$X=0$	$X=3$
$Y=-1$	0,1	0,2	0,1	0,2
$Y=1$	0	0,1	0	0,2
$Y=3$	0	0	0,1	0

Будут ли независимыми событие *A*: "$|X|=3|Y|$" и событие *B*: "$X>Y$"?

2.165. Известен закон распределения двумерной величины (X,Y):

	$X=1$	$X=2$	$X=3$	$X=4$
$Y=2$	0,1	0	0,2	0
$Y=3$	0,1	0,1	0	0,1
$Y=4$	0,1	0,1	0,2	0

Будут ли зависимыми событие *A*: "$XY<10$" и событие *B*: "$X>1$"?

2.166. Известен закон распределения двумерной величины (X,Y):

	$X=5$	$X=6$	$X=7$	$X=8$
$Y=7$	0,1	0	0,1	0
$Y=8$	0	0,2	0	0,2
$Y=9$	0,1	0,1	0,2	0

Будут ли независимыми события *A*: "$X+Y>15$" и *B*: "$Y>7$"?

2.167. Известен закон распределения двумерной величины (X, Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,2	0	0,1	0,2
$Y=0$	0,1	0,1	0	0,1
$Y=1$	0,1	0	0,1	0

Будут ли независимыми события A: “ $X \cdot Y > 0$ ” и B: “$Y \neq 0$”?

2.168. Известен закон распределения двумерной величины (X, Y):

	$X=-4$	$X=-1$	$X=0$	$X=4$
$Y=-2$	0	0	0,1	0,2
$Y=0$	0,1	0,1	0.1	0,1
$Y=1$	0,1	0.1	0,1	0

Будут ли зависимыми события A: “ $X \cdot Y \leq 0$ ” и B: “$Y \geq 0$ ”?

2.169. Известен закон распределения двумерной величины (X, Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=1$	0,2	0	0,1	0

Будут ли независимыми события A: “ $X \cdot Y \geq 0$ ” и B: “$X \geq 0$ ”?

2.170. Известен закон распределения двумерной величины (X, Y):

	$X=-3$	$X=-2$	$X=-1$	$X=0$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=1$	0,2	0	0,1	0

Будут ли зависимыми события A: “ $X+Y<0$ ” и B: “$Y - X > 0$ ”?

2.7. Непрерывные многомерные случайные величины

Как и прежде, при рассмотрении непрерывных многомерных величин мы ограничимся двумерным случаем. Если распределения величин X и Y непрерывны, то двумерная величина (X,Y) определяется как непрерывная и задается с помощью плотности вероятности — интегрируемой функции $f_{X,Y}(x,y)$, удовлетворяющей условию:

$$P((X,Y)\in G)=\iint_G f_{X,Y}(x,y)dxdy .$$

Здесь G — некоторая область плоскости xOy, вероятность попадания в которую точки (X,Y) и вычисляют с помощью двойного интеграла.

Функция распределения двумерной непрерывной величины $F_{X,Y}(x,y)$ определяется стандартно условием

$$F_{X\ Y}(x,y)=P(X<x,Y<y)$$

и через плотность вероятности выражается как:

$$F_{X,Y}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)dsdt$$

Для плотности вероятности $f_{X\ Y}(x\ y)$ двумерной непрерывной случайной величины (X,Y) имеют место следующие свойства:

1. $f_{X,Y}(x,y)\ \geq\ 0;$

2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy\ =1$ (свойство нормированности плотности вероятности);

3. $f_{X,Y}(x,y)\ =\ \dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x\partial y}$.

Для непрерывной двумерной случайной величины (X,Y) ее компоненты X и Y являются непрерывными величинами и плотности вероятностей для них находят по формулам:

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\ ;$$

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\ .$$

Пример 2.31. Плотность вероятности непрерывной двумерной случайной величины (X,Y) имеет вид:

$$f_{X,Y}(x,y)=\begin{cases} a(x+y), & 0\le x\le 5,\ 0\le y\le 5 \\ 0, & \text{в остальных точках} \end{cases}.$$

Найти значение параметра a и плотность вероятности величины X.

Решение.

Для того чтобы найти параметр плотности вероятности, необходимо использовать свойство нормированности плотности:

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1.$$

С учетом данных условия задачи имеем:

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \int\limits_0^5\int\limits_0^5 a(x+y)dxdy =$$

$$= a\int\limits_0^5 dx\left[\int\limits_0^5 (x+y)dy\right] = a\int\limits_0^5\left[xy+\frac{y^2}{2}\bigg|_0^5\right]dx =$$

$$= a\int\limits_0^5\left(5x+\frac{25}{2}\right)dx = a\left(5\cdot\frac{x^2}{2}+\frac{25}{2}x\bigg|_0^5\right) =$$

$$= a\left(\frac{125}{2}+\frac{125}{2}\right) = 125a = 1 \Rightarrow a=\frac{1}{125}.$$

Для того, чтобы найти плотность вероятности величины X, воспользуемся условием:

$$f_X(x)=\int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

С учетом данных условия задачи и найденного значения параметра $a=\dfrac{1}{125}$ получим:

$$f_X(x)=\int\limits_0^5\frac{1}{125}(x+y)dy = \frac{1}{125}\left(xy+\frac{y^2}{2}\bigg|_0^5\right) = \frac{1}{125}\left(5x+\frac{25}{2}\right) =$$

$$= \frac{x}{25}+\frac{1}{10} = 0{,}04x+0{,}1.$$

Ответ: $a=\dfrac{1}{125}$, $f_X(x)=\begin{cases} 0{,}04x+0{,}1; & 0\le x\le 5 \\ 0, & \text{в остальных точках} \end{cases}.$

Понятно, что, имея возможность определить вид плотности вероятности отдельно для величин X и Y, дальше можно вычислять их математические ожидания и дисперсии, а также решать задачи, связанные вероятностями попадания в определенные области значений этих величин.

Пример 2.32. Плотность вероятности непрерывной двумерной случайной величины (X,Y) имеет вид:

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{125}\cdot(x+y), & 0\le x\le 5,\ 0\le y\le 5\\ 0, & \text{в остальных точках}\end{cases}.$$

Найти математическое ожидание $E(X)$ и $P(X>1)$.

Решение.

Для непрерывной величины X математическое ожидание вычисляют по формуле:

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx .$$

Особенность нашего случая состоит в том, что X — один из компонентов случайной величины (X,Y), поэтому ее плотность вероятности должна быть найдена из условия:

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy .$$

Мы это уже сделали в ***примере 2.31***, поэтому воспользуемся полученным результатом

$$f_X(x)=\begin{cases}0{,}04x+0{,}1; & 0\le x\le 5\\ 0, & \text{в остальных точках}\end{cases}$$

и вычислим математическое ожидание величины X:

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx = \int_0^5 x\cdot\left(0{,}04x+0{,}1\right)dx = \int_0^5\left(0{,}04x^2+0{,}1x\right)dx =$$

$$= 0{,}04\cdot\frac{x^3}{3}+0{,}1\cdot\frac{x^2}{2}\Bigg|_0^5 = \frac{1}{3}+\frac{25}{20}=\frac{20+75}{60}=\frac{95}{60}=\frac{19}{12}=1{,}583 .$$

Ответ на второй вопрос задачи находим, используя определение плотности вероятности $P(x_1<X<x_2)=\int_{x_1}^{x_2} f(x)dx$, которое в нашем случае примет вид:

$$P(X>1) = P(1<X<+\infty)=\int_1^{+\infty} f(x)dx=\int_1^5\left(0{,}04x+0{,}1\right)dx =$$

$$= 0{,}04 \cdot \frac{x^2}{2} + 0{,}1x \Big|_1^5 = \left(0{,}02 \cdot 25 + 0{,}5\right) - \left(0{,}02 + 0{,}1\right) = 1 \ -0{,}12 \ = 0{,}88.$$

Ответ: $E(X) = 1{,}583$; $P(X > 1) = 0{,}88$.

Если двумерная непрерывная величина (X,Y) ***равномерно распределена*** в некоторой ***области*** G, то ее плотность вероятности имеет вид:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{площадь}\, G}, & (x,y) \in G \\ 0, \text{ в остальных точках} \end{cases}.$$

Для равномерно распределенной в области G двумерной величины (X,Y) всегда выполнено условие:

$$P\big((X,Y) \in D\big) = \frac{\text{площадь}\, D}{\text{площадь}\, G},$$

если D является некоторой частью области G.

Пример 2.33. Непрерывная двумерная случайная величина (X,Y) равномерно распределена в области $G: \{0 \le x \le 5; \quad 0 \le y \le 5\}$. Найти плотность вероятности этой величины и $P(X+1>Y)$.

Решение.

Область G является квадратом со стороной 5 (см). Площадь такого квадрата равна 25 (кв. см). Поэтому плотность вероятности сразу может быть записана как

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{25}, & (x,y) \in G \\ 0, \text{ в остальных точках} \end{cases}.$$

Для ответа на второй вопрос задачи отметим в координатной плоскости данные из условия (рис. 2.8).

Цветом выделена область D, отвечающая условию $X+1>Y$.

Так как величина (X,Y) равномерно распределена в области G, то ответ можно искать по формуле $P\big((X,Y) \in D\big) = \dfrac{\text{площадь}\, D}{\text{площадь}\, G}$:

$P(X+1>Y) =$

$$= P\big((X,Y) \in D\big) = \frac{\text{площадь}\, D}{\text{площадь}\, G} = \frac{25 - \frac{1}{2} \cdot 4 \cdot 4}{25} = \frac{17}{25} = 0{,}68.$$

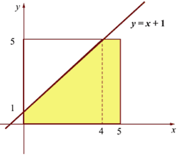


Рис. 2.8. К примеру 2.33

Площадь области D (в числителе дроби) мы нашли, вычитая из площади квадрата площадь прямоугольного треугольника, катеты которого равны 4 (см).

Ответ: $f_{X,Y}(x,y) = \begin{cases} \frac{1}{25}, & (x,y) \in G \\ 0, & \text{в остальных точках} \end{cases}$, $P(X+1>Y) = 0{,}68$.

Для непрерывной двумерной случайной величины (X,Y) составляющие величины X и Y могут быть независимыми. Независимость величин X и Y, образующих двумерную величину (X,Y), определяется выполнением равенств:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \text{ или } F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y).$$

В последнем условии функции распределения $F_X(x) = P(X < x)$ и $F_Y(y) = P(Y < y)$ находят по функции распределения $F_{X,Y}(x,y) = P(X < x, Y < y)$, используя пределы:

$$\lim_{y\to\infty} F_{X,Y}(x,y) = F_X(x) \text{ и } \lim_{x\to\infty} F_{X,Y}(x,y) = F_Y(y).$$

Пример 2.34. Известно, что величины X и Y независимы и равномерно распределены на отрезках $[3;5]$ и $[2;10]$ соответственно. Найти плотность вероятности и функцию распределения величины (X,Y).

Решение.

Используя условие равномерной распределенности величин X и Y, получим сразу для них плотности вероятностей и функции распределения:

$$f_X(x)=\begin{cases}\dfrac{1}{5-3}=\dfrac{1}{2}, & 3\le x\le 5\\ 0, \text{ в остальных точках}\end{cases};\quad F_X(x)=\begin{cases}0, & x<3\\ \dfrac{x-3}{5-3}=\dfrac{x-3}{2}, & 3\le x\le 5\\ 1, & x>5\end{cases};$$

$$f_Y(y)=\begin{cases}\dfrac{1}{10-2}=\dfrac{1}{8}, & 2\le y\le 10\\ 0, \text{ в остальных точках}\end{cases};\quad F_Y(y)=\begin{cases}0, & y<2\\ \dfrac{y-2}{10-2}=\dfrac{y-2}{8}, & 2\le y\le 10\\ 1, & y>10\end{cases}.$$

Из условия независимости находим:

$$f_{XY}(x,y)=f_X(x)\cdot f_Y(y)\ \Rightarrow$$

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{2}\cdot\dfrac{1}{8}=\dfrac{1}{16}, & 3\le x\le 5, 2\le y\le 10\\ 0, \text{ в остальных точках}\end{cases}.$$

$$F_{X,Y}(x,y)=F_X(x)\cdot F_Y(y)\ \Rightarrow$$

$$F_{X,Y}(x,y)=\begin{cases}0, & x<3, y<2\\ \dfrac{x-3}{2}\cdot\dfrac{y-2}{8}, & 3\le x\le 5, 2\le y\le 10\\ 1, & x>5, y>10\end{cases}.$$

Ответ:

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{16}, & 3\le x\le 5, 2\le y\le 10\\ 0, \text{ в остальных точках}\end{cases};$$

$$F_{X,Y}(x,y)=\begin{cases}0, & x<3, y<2\\ \dfrac{xy-2x-3y+6}{16}, & 3\le x\le 5, 2\le y\le 10\\ 1, & x>5, y>10\end{cases}.$$

В качестве характеристик двумерного распределения используют математические ожидания некоторых функций от рассматриваемых случайных величин.

Пусть задана плотность вероятности $f_{X,Y}(x,y)$ непрерывной двумерной случайной величины (X,Y) и некоторая функция двух аргументов $z = \phi(x, y)$, под действием которой формируется новая величина $Z = \phi(X, Y)$. Тогда для этой новой величины математическое ожидание находят по формуле:

$$E(Z) = E(\phi(X, Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi(x, y) \cdot f_{X,Y}(x, y) dxdy .$$

В частности, если

$$Z = \left(X - E(X)\right) \cdot (Y - E(Y)) ,$$

то

$$E(Z) = E\left((X - E(X) \cdot (Y - E(Y))\right) = \mathrm{Cov}(X, Y)$$

и тогда

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(x - E(X)\right) \cdot \left(y - E(Y)\right) \cdot f_{X,Y}(x, y) dxdy .$$

Пример 2.35. Плотность вероятности непрерывной двумерной случайной величины (X,Y) имеет вид:

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{125} \cdot (x + y), & 0 \le x \le 5,\ 0 \le y \le 5 \\ 0, \text{ в остальных точках} \end{cases} .$$

Найти ковариацию величин X и Y.

Решение.

Для того, чтобы использовать формулу для вычисления ковариации

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(x - E(X)\right) \cdot \left(y - E(Y)\right) \cdot f_{X,Y}(x, y) dxdy ,$$

необходимо найти математические ожидания $E(X)$ и $E(Y)$, а для этого, в свою очередь, нам понадобятся соответствующие плотности вероятностей $f_X(x)$ и $f_Y(y)$. Частично мы эту задачу решили (см. выше пример 2.31 и пример 2.32), поэтому будем использовать готовые результаты из этих примеров. Итак, имеем:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy = \begin{cases} 0{,}04x + 0{,}1; & 0 \le x \le 5 \\ 0, \text{ в остальных точках} \end{cases}$$

и

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_0^5 x \cdot \left(0{,}04x + 0{,}1\right) dx = \frac{19}{12} = 1{,}583 .$$

Аналогично для Y получим:

$$f_Y(y)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \;=\begin{cases}0{,}04y+0{,}1; & 0\le y\le 5\\ 0, & \text{в остальных точках}\end{cases}$$

и

$$E(Y)=\int_{-\infty}^{+\infty} y\cdot f_Y(y)dy \;=\int_0^5 y\cdot\left(0{,}04y+0{,}1\right)dy \;=\frac{19}{12}=1{,}583\,.$$

Теперь ответим на вопрос задачи:

$$\mathrm{Cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-E(X)\right)\cdot\left(y-E(Y)\right)\cdot f_{X,Y}(x,y)dxdy \;=$$

$$=\int_0^5\int_0^5\left(x-\frac{19}{12}\right)\cdot\left(y-\frac{19}{12}\right)\cdot\frac{1}{125}(x+y)dxdy \;=$$

$$=\int_0^5\int_0^5 xy\cdot\frac{1}{125}(x+y)dxdy-\frac{19}{12}\cdot\frac{19}{12}\;=\;\frac{1}{125}\int_0^5\int_0^5 xy\left(x+y\right)dxdy-\frac{361}{144}=$$

$$=\frac{1}{125}\int_0^5\int_0^5(x^2y+xy^2)dxdy-\frac{361}{144}=\frac{1}{125}\int_0^5 dx\left[\frac{x^2y^2}{2}+\frac{xy^3}{3}\Bigg|_0^5\right]-\frac{361}{144}\;=$$

$$=\frac{1}{125}\int_0^5\left(\frac{25x^2}{2}+\frac{125x}{3}\right)dx-\frac{361}{144}\;=\;\frac{1}{125}\left(\frac{25x^3}{6}+\frac{125x^2}{6}\Bigg|_0^5\right)-\frac{361}{144}\;=$$

$$=\frac{1}{125}\left(\frac{25\cdot125}{6}+\frac{125\cdot25}{6}\right)-\frac{361}{144}\;=\;\frac{25}{3}-\frac{361}{144}\;=$$

$$=\frac{25\cdot48-361}{144}=\frac{839}{144}=5{,}826\,.$$

Ответ: $\mathrm{Cov}(X,Y)=\;5{,}826.$

В процессе решения мы использовали равенство:

$$\mathrm{Cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-E(X)\right)\cdot\left(y-E(Y)\right)\cdot f_{X,Y}(x,y)dxdy \;=$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\cdot f_{X,Y}(x,y)dxdy-E(X)\cdot E(Y)\,,$$

полученное на основе преобразований под знаком двойного интеграла и определений математических ожиданий непрерывных случайных величин X и Y:

$$E\left(X\right)=\int_{-\infty}^{\infty} x\cdot f_{X,Y}\left(x,y\right)dxdy\,;\;\; E\left(Y\right)=\int_{-\infty}^{\infty} y\cdot f_{X,Y}\left(x,y\right)dxdy\,.$$

При известной ковариации случайных величин вычисляется и их коэффициент корреляции по известной формуле

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)},$$

где

$$\sigma(X)=\sqrt{D(X)},\ \sigma(Y)=\sqrt{D(Y)},$$

а дисперсии определяются как

$$D(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-E(X)\right)^2 f_{X,Y}(x,y)dxdy;$$

$$D(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(y-E(Y)\right)^2 f_{X,Y}(x,y)dxdy.$$

При решении задач из финансовых приложений часто используются ***ковариационная*** и к***орреляционная*** матрицы, которые в двумерном случае имеют вид:

$$C(X,Y)=\begin{pmatrix} D(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & D(Y)\end{pmatrix}$$

и

$$R(X,Y)=\begin{pmatrix} 1 & \rho(X,Y) \\ \rho(X,Y) & 1\end{pmatrix}.$$

Эти матрицы являются симметричными, а их определители — неотрицательны, более того, для корреляционной матрицы определитель всегда ≤ 1.

Задачи для самостоятельного решения

В задачах **271–280** задана плотность вероятности $f(x,y)$ двумерной непрерывной случайной величины (X,Y). Необходимо:

— найти параметр a плотности вероятности $f(x,y)$,

— плотности вероятностей $f_X(x)$ и $f_Y(y)$ величин X и Y,

— вычислить $E(X)$, $E(Y)$, $D(X)$, $D(Y)$,

— вычислить ковариацию $\mathrm{Cov}(X,Y)$ и коэффициент корреляции $\rho(X,Y)$,

— ответить на вопрос о вероятности.

2.171. $f_{X,Y}(x,y)=\begin{cases} ax+y, & 0\le x\le 2,\quad 0\le y\le 2 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P(X<1)$.

2.172. $f_{X,Y}(x,y)=\begin{cases} x+ay, & 0\le x\le 1,\quad 0\le y\le 1 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P\left(Y>\dfrac{1}{2}\right)$.

2.173. $f_{X,Y}(x,y)=\begin{cases} x+y, & 0\le x\le a,\quad 0\le y\le a \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P(X<Y)$.

2.174. $f_{X,Y}(x,y)=\begin{cases} a+x-y, & 0\le x\le 1,\quad 0\le y\le 1 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P(X>Y)$.

2.175. $f_{X,Y}(x,y)=\begin{cases} 1+ax-y, & 0\le x\le 2,\quad 0\le y\le 2 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P(X<1)$.

2.176. $f_{X,Y}(x,y)=\begin{cases} a(x-y)^2, & 0\le x\le 1,\quad 0\le y\le 1 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P\left(Y>\dfrac{1}{2}\right)$.

2.177. $f_{X,Y}(x,y)=\begin{cases} (ax-y)^2, & 0\le x\le 2,\quad 0\le y\le 2 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P(Y>1)$.

2.178. $f_{X,Y}(x,y)=\begin{cases} (x-ay)^2, & 0\le x\le 1,\quad 0\le y\le 1 \\ 0, & \text{в остальных точках}\end{cases}$, a — параметр.

Вычислить $P\left(X<\frac{1}{2}\right)$.

2.179. $f_{X,Y}(x,y)=\begin{cases}1+x-y, & 0\le x\le a, \quad 0\le y\le a\\ 0, \text{ в остальных точках}\end{cases}$, a — параметр.

Вычислить $P\left(Y>\frac{a}{3}\right)$.

2.180. $f_{X,Y}(x,y)=\begin{cases}(x-y)^2, & 0\le x\le a, \quad 0\le y\le a\\ 0, \text{ в остальных точках}\end{cases}$, a — параметр.

Вычислить $P\left(X<\frac{a}{4}\right)$.

В задачах **2.181–2.190**, используя условие равномерной распределенности в заданной области G двумерной непрерывной случайной величины (X,Y), найти плотность вероятности $f(x,y)$, вычислить $E(X)$, $E(Y)$, $D(X)$, $D(Y)$ и ответить на вопрос о вероятности.

2.181. Непрерывная двумерная случайная величина $(X\ Y)$ равномерно распределена в области G : $\{0\le x\le 10;\quad 0\le y\le 5\}$. Найти вероятность $P(X>Y)$.

2.182. Непрерывная двумерная случайная величина (X,Y) равномерно распределена в области G : $\{0\le x\le 10;\quad 0\le y\le 5\}$. Найти вероятность $P(X<Y)$.

2.183. Непрерывная двумерная случайная величина (X,Y) равномерно распределена в области G : $\{x\ge 0,\ y\ge 0,\ 2x+3y\le 12\}$. Найти вероятность $P(X<4)$.

2.184. Непрерывная двумерная случайная величина (X,Y) равномерно распределена в области G : $\{x\ge 0,\ y\ge 0,\ 3x+2y\le 6\}$. Найти вероятность $P(Y>1)$.

2.185. Непрерывная двумерная случайная величина (X,Y) равномерно распределена в области G : $\{0\le x\le 5;\quad 0\le y\le 5\}$.

Найти вероятность $P(X+Y>3)$.

2.186. Компоненты X и Y непрерывной двумерной случайной величины (X,Y) независимы и равномерно распределены: X — на отрезке $[2;4]$, Y- на отрезке $[3;5]$. Найти вероятность $P(X+Y<8)$.

2.187. Компоненты X и Y непрерывной двумерной случайной величины (X,Y) независимы и равномерно распределены: X — на отрезке $[0;5]$, Y — на отрезке $[1;6]$. Найти вероятность $P(X>Y)$.

2.188. Компоненты X и Y непрерывной двумерной случайной величины (X,Y) независимы и равномерно распределены: X — на отрезке $[0;4]$, Y — на отрезке $[0;4]$. Найти вероятность $P(2X<Y-1)$.

2.189. Компоненты X и Y непрерывной двумерной случайной величины (X,Y) независимы и равномерно распределены: X — на отрезке $[0;10]$, Y — на отрезке $[0;5]$. Найти вероятность $P(X<2Y)$.

2.190. Компоненты X и Y непрерывной двумерной случайной величины (X,Y) независимы и равномерно распределены: X — на отрезке $[0;6]$, Y — на отрезке $[0;6]$. Найти вероятность $P(2X<3Y)$.

2.8. Условные распределения

При изучении числовых характеристик двумерной случайной величины (X,Y) отдельное внимание должно быть уделено условным характеристикам, связанным с условными распределениями. Рассмотрим эти понятия подробнее.

Пусть задана двумерная случайная величина (X,Y). Мы можем изучать одну из ее компонент при условии, что другая компонента приняла значение из определенного числового интервала (или даже просто определенное числовое значение). В этом случае говорят об ***условном распределении***.

Рассмотрим сначала ситуацию, когда (X,Y) — дискретная случайная величина с известным законом распределения:

	$X=x_1$	$X=x_2$	…	$X=x_n$
$Y=y_1$	p_{11}	p_{21}		p_{n1}
$Y=y_2$	p_{12}	p_{22}		p_{n2}
…	…	…		…
$Y=y_k$	p_{1k}	p_{2k}		p_{nk}

Здесь $p_{ij}=P\left(X=x_i,Y=y_j\right)$ и $\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{k}p_{ij}=1$.

Условное распределение формируется под влиянием дополнительного условия для одного из компонентов и является случайной величиной со своим, ***условным***, законом распределения. В этом законе распределения должны быть указаны все возможные при дополнительном условии значения и соответствующие этим значениям вероятности, которые в таком случае вычисляются как условные:

$$P\left(X=x_i\middle|Y=y_j\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(Y=y_j\right)}=\frac{p_{ij}}{\sum\limits_{j}p_{ij}}$$

– если дополнительное условие связано с величиной Y, или

$$P\left(Y=y_j\middle|X=x_i\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(X=x_i\right)}=\frac{p_{ij}}{\sum\limits_{i}p_{ij}};$$

– если дополнительное условие связано с величиной X.

Пример 2.36. Дано распределение двумерной дискретной случайной величины (X,Y):

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	0	0,1	0,1	0,1
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0,1	0,1	0	0,2

Составить условные законы распределений для величин $Z=\left(Y\middle|\left|X\right|<2\right)$ и $T=\left(X\middle|Y\geq 0\right)$.

Решение.

Для того чтобы составить закон распределения $Z=\left(Y\middle|\left|X\right|<2\right)$, надо считать условие $\left|X\right|<2$ выполненным, при заданном законе распределения этому условию отвечают выделенные значения $X=-1$, $X=0$ и $X=1$:

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	**0**	**0,1**	**0,1**	0,1
$Y=0$	**0,1**	**0**	**0,1**	0,1
$Y=1$	**0,1**	**0,1**	**0**	0,2

Можем найти и $P(|X|<2)$, просуммировав все вероятности из выделенной части таблицы:

$P(|X|<2)$ = 0 + 0,1 +0,1 + 0,1 + 0 + 0,1 + 0,1 + 0,1 + 0 = 0,6.

Условие $|X|<2$ для заданной двумерной величины (X,Y) не ограничивает возможных значений для Y, поэтому среди возможных значений для $Z=(Y\,|\,|X|<2)$ мы должны указать все возможные значения Y:

Значение $Z=(Y\,|\,|X|<2)$	-1	0	1
Вероятность	$\dfrac{0+0{,}1+0{,}1}{0{,}6}$	$\dfrac{0{,}1+0+0{,}1}{0{,}6}$	$\dfrac{0{,}1+0{,}1+0}{0{,}6}$

Соответствующие определенным значениям вероятности мы вычисляем как условные:

$$P(Y=-1\,|\,|X|<2)=\frac{P(Y=-1 \text{ и } |X|<2)}{P(|X|<2)}=$$

$$=\frac{P(Y=-1,X=-1)+P(Y=-1,X=0)+P(Y=-1,X=1)}{P(|X|<2)}=$$

$$=\frac{0+0{,}1+0{,}1}{0{,}6}=\frac{0{,}2}{0{,}6}=\frac{1}{3};$$

$$P(Y=0\,|\,|X|<2)=\frac{P(Y=0 \text{ и } |X|<2)}{P(|X|<2)}=$$

$$=\frac{P(Y=0,X=-1)+P(Y=0,X=0)+P(Y=0,X=1)}{P(|X|<2)}=$$

$$=\frac{0{,}1+0+0{,}1}{0{,}6}=\frac{0{,}2}{0{,}6}=\frac{1}{3};$$

$$P\left(Y=1\,\middle|\,|X|<2\right)=\frac{P\left(Y=1 \text{ и } |X|<2\right)}{P\left(|X|<2\right)}=$$

$$=\frac{P(Y=1,X=-1)+P(Y=1,X=0)+P(Y=1,X=1)}{P(|X|<2)}=$$

$$=\frac{0{,}1+0{,}1+0}{0{,}6}=\frac{0{,}2}{0{,}6}=\frac{1}{3}.$$

Таким образом, условный закон распределения $Z=\left(Y\,\middle|\,|X|<2\right)$ имеет вид:

Значение $Z=\left(Y\,\middle|\,\lvert X\rvert<2\right)$	–1	0	1
Вероятность	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Необходимое для закона распределения условие $\sum_i p_i=1$ выполнено.

Аналогично поступим и в случае $T=\left(X\,\middle|\,Y\geq 0\right)$.

Будем считать условие $Y\geq 0$ выполненным:

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	0	0,1	0,1	0,1
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0,1	0,1	0	0,2

Тогда $P(Y\geq 0)=0{,}1+0+0{,}1+0{,}1+0{,}1+0{,}1+0+0{,}2=0{,}7$.

Заданное условие $Y\geq 0$ не ограничивает возможных значений X, но для нового закона соответствующие вероятности будем считать как условные:

Значение $T=\left(X\,\middle|\,Y\geq 0\right)$	–1	0	1	2
Вероятность	$\frac{0{,}1+0{,}1}{0{,}7}$	$\frac{0+0{,}1}{0{,}7}$	$\frac{0{,}1+0}{0{,}7}$	$\frac{0{,}1+0{,}2}{0{,}7}$

Здесь имеем:

$$P\left(X=-1\,\middle|\,Y\geq 0\right)=\frac{P\left(X=-1 \text{ и } Y\geq 0\right)}{P\left(Y\geq 0\right)}=$$

$$= \frac{P(X=-1, Y=0)+P(X=-1, Y=1)}{P(Y \geq 0)} = \frac{0{,}1+0{,}1}{0{,}7} = \frac{0{,}2}{0{,}7} = \frac{2}{7};$$

$$P\left(X=0 \middle| Y \geq 0\right) = \frac{P\left(X=0 \text{ и } Y \geq 0\right)}{P\left(Y \geq 0\right)} =$$

$$\frac{P(X=0, Y=0)+P(X=0, Y=1)}{P(Y \geq 0)} = \frac{0 \quad 0{,}1}{0{,}7} = \frac{0{,}1}{0{,}7} = \frac{1}{7};$$

$$P\left(X=1 \middle| Y \geq 0\right) = \frac{P\left(X=1 \text{ и } Y \geq 0\right)}{P\left(Y \geq 0\right)} =$$

$$= \frac{P(X=1, Y=0)+P(X=1, Y=1)}{P(Y \geq 0)} = \frac{0{,}1+0}{0{,}7} = \frac{0{,}1}{0{,}7} = \frac{1}{7};$$

$$P\left(X=2 \middle| Y \geq 0\right) = \frac{P\left(X=2 \text{ и } Y \geq 0\right)}{P\left(Y \geq 0\right)} =$$

$$= \frac{P(X=2, Y=0)+P(X=2, Y=1)}{P(Y \geq 0)} = \frac{0{,}1+0{,}2}{0{,}7} = \frac{0{,}3}{0{,}7} = \frac{3}{7}.$$

Таким образом, условный закон распределения $T=\left(X \middle| Y \geq 0\right)$ имеет вид:

Значение $T=\left(X \middle| Y \geq 0\right)$	-1	0	1	2
Вероятность	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$

и условие $\sum_i p_i = 1$ выполнено.

Ответ: искомые законы распределений имеют вид

Значение $Z=\left(Y \middle| \lvert X \rvert < 2\right)$	-1	0	1
Вероятность	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Значение $T=\left(X \middle| Y \geq 0\right)$	-1	0	1	2
Вероятность	–	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{3}{7}$

Особый интерес представляют условные распределения, в которых заданное условие связано с одним из значений величины.

Пример 2.37. Дано распределение двумерной дискретной случайной величины (X,Y):

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	0	0,1	0,1	0,1
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0,1	0,1	0	0,2

Составить условный закон распределения для величины $Z=(Y|X=-1)$.

Решение.

Поступая так, как показано в ***примере 2.36:***

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	0	0,1	0,1	0,1
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0,1	0,1	0	0,2

без труда получаем соответствующее распределение:

Значение $Z=(Y\|X=-1)$	-1	0	1
Вероятность	$\frac{0}{0,2}$	$\frac{0,1}{0,2}$	$\frac{0,1}{0,2}$

Значения, для которых возможная вероятность равна 0, из закона распределения исключаем, поэтому в итоге будем иметь:

Ответ:

Значения $Z=(Y\|X=-1)$	0	1
Вероятность	$\frac{1}{2}$	$\frac{1}{2}$

Новая величина имеет свои числовые характеристики — математическое ожидание и дисперсию, которые легко вычислить:

$$E(Y|X=-1)=0\cdot\frac{1}{2}+1\cdot\frac{1}{2}=\frac{1}{2}=0,5;$$

$$D(Y|X=-1)=0^2\cdot\frac{1}{2}+1^2\cdot\frac{1}{2}-\left(\frac{1}{2}\right)^2=\frac{1}{2}-\frac{1}{4}=\frac{1}{4}=0,25.$$

Понятно, что при изменении условия изменится и распределение, а значит — и числовые характеристики. Поэтому условные математические ожидания и условные дисперсии ведут себя как случайные величины $E(Y|X)$ и $D(Y|X)$ соответственно, у которых есть свои законы распределений.

Пример 2.38. Дано распределение двумерной дискретной случайной величины (X,Y):

	$X=-1$	$X=0$	$X=1$	$X=2$
$Y=-1$	0	0,1	0,1	0,1
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0,1	0,1	0	0,2

Составить законы распределений $E(Y|X)$ и $D(Y|X)$.

Решение.

Чтобы составить искомые законы распределений $E(Y|X)$ и $D(Y|X)$, необходимо перебрать все условия для возможных значений X, задавая условные распределения и вычисляя их математические ожидания и дисперсии. Полученные значения этих числовых характеристик и будут возможными значениями $E(Y|X)$ и $D(Y|X)$, а соответствующие вероятности будут определяться вероятностями условий.

Часть этой работы мы уже выполнили в ***примере 2.37***, поэтому воспользуемся результатом:

Значение $Z=(Y\|X=-1)$	0	1
Вероятность	$\frac{1}{2}$	$\frac{1}{2}$

$P(X=-1)=0{,}2$	$E(Y\|X=-1)=0{,}5$
	$D(Y\|X=-1)=0{,}25$

Поступая аналогично для условий $X=0$, $X=1$, $X=2$, получим:

Значение $Z=(Y\|X=0)$	-1	1
Вероятность	$\frac{1}{2}$	$\frac{1}{2}$

$P(X=0)=0{,}2$	$E(Y\|X=0)=0$
	$D(Y\|X=0)=1$

Значение $Z=(Y\|X=1)$	–1	**0**
Вероятность	$\frac{1}{2}$	$\frac{1}{2}$

$P(X=1)=0{,}2$	$E(Y\|X=1)=-0{,}5$
	$D(Y\|X=1)=0{,}25$

Значение $Z=(Y\|X=2)$	–1	0	1
Вероятность	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

$P(X=2)=0{,}4$	$E(Y\|X=2)=0{,}25$
	$D(Y\|X=2)=0{,}6875$

Теперь у нас есть все возможные значения для условных математических ожиданий и условных дисперсий, поэтому составим искомые законы распределений (для условного математического ожидания используемые условия выделены цветом):

Значение $E(Y\|X)$	–0,5	0	0,25	0,5
Вероятность	0,2	0,2	0,4	0,2

Закон распределения условной дисперсии $D(Y|X)$ имеет вид:

Значение $D(Y\|X)$	0,25	0,6875	1
Вероятность	0,4	0,2	0,4

В таблице учтено, что значение, равное 0,25, условная дисперсия принимает в двух случаях: при $X=1$ (с вероятностью $P(X=1)=0{,}2$) и при $X=-1$ (с вероятностью $P(X=-1)=0{,}2$). Соответствующие этим условиям вероятности необходимо сложить: $P(D(Y|X)=0{,}25)=0{,}2+0{,}2=0{,}4$.

Ответ:

Значение $E(Y\|X)$	–0,5	0	0,25	0,5
Вероятность	0,2	0,2	0,4	0,2

Значение $D(Y\|X)$	0,25	0,6875	1
Вероятность	0,4	0,2	0,4

Еще раз подчеркнем, что условное математическое ожидание $E(Y|X)$ и условная дисперсия $D(Y|X)$ являются случайными величинами, следовательно, для них можно находить числовые характеристики — математические ожидания и дисперсии.

Перечислим ***свойства условного математического ожидания.***

1. $E\left(c|X\right)=c, \qquad c-const$

2. $E\left(aY+b|X\right)=aE\left(Y|X\right)+b$, где a и b — постоянные.

3. $E\left(X+Y|Z\right)=E\left(X|Z\right)+E\left(Y|Z\right)$.

4. Для независимых величин X и Y выполняются равенства $E\left(Y|X\right)=E\left(Y\right)$ и $E\left(X|Y\right)=E\left(X\right)$.

Имеют место и такие равенства:

$$E(X)=E\Big(E(X|Y)\Big);\ E(Y)=E\Big(E(Y|X)\Big),$$

то есть математическое ожидание величины может быть найдено как математическое ожидание от условного математического ожидания.

Для решения задач будет полезно и такое равенство:

$$E(X\cdot Y)=E\Big(X\cdot E(Y|X)\Big)\ =\ E\Big(Y\cdot E(X|Y)\Big).$$

Имеют место и ***свойства условной дисперсии***:

1. $D\left(c|X\right)=0, \qquad c-const$

2. $D\left(aX+b|Y\right)=a\ D\left(X|Y\right)$, где a и b — постоянные.

3. Для независимых величин X и Y выполняются равенства

$$D\left(X|Y\right)=D\left(X\right) \text{ и } D\left(Y|X\right)=D\left(Y\right).$$

Кроме того, имеет место формула полной дисперсии:

$$D(X)=E\Big(D(X|Y)\Big)+D\Big(E(X|Y)\Big);\ D(Y)=E\Big(D(Y|X)\Big)+D\Big(E(Y|X)\Big).$$

Для непрерывной двумерной случайной величины $\left(X,Y\right)$ с известной плотностью вероятности $f(x,y)$ условные распределения определяются на основе условных плотностей вероятностей:

$$f_{X|Y}\left(x|y\right)=\frac{f\left(x,y\right)}{f_Y\left(y\right)},\ f_{Y|X}\left(y|x\right)=\frac{f\left(x,y\right)}{f_X\left(x\right)}.$$

Если случайные величины X и Y независимы, то их условные плотности равны безусловным:

$$f_{X|Y}\left(x|y\right)=f_X\left(x\right),\ f_{Y|X}\left(y|x\right)=f_Y\left(y\right).$$

Условное математическое ожидание для непрерывной величины Y при условии $X=x$ определяется соответствующим интегралом:

$$E(Y|X=x)=\int_{-\infty}^{+\infty} y\cdot f_{Y|X}(y|x)dy=\int_{-\infty}^{+\infty} y\cdot\frac{f_{X,Y}(x,y)}{f_X(x)}dy.$$

Из общей теории известно, что это условное математическое ожидание определяет ***регрессию*** Y по X.

Аналогично можно определять и регрессию X по Y.

Условная дисперсия для непрерывной величины Y при условии $X=x$ определяется соответствующим интегралом:

$$D(Y|X=x)=$$
$$=\int_{-\infty}^{+\infty}(y-E(Y|X=x))^2\cdot f_{Y|X}(y|x)dy=\int_{-\infty}^{+\infty}(y-E(Y|X=x))^2\cdot\frac{f_{X,Y}(x,y)}{f_X(x)}dy.$$

Важным примером непрерывной двумерной случайной величины $\left(X,Y\right)$ является двумерное нормальное распределение.

Для ***двумерного нормального распределения*** $\left(X,Y\right)$ плотность вероятности имеет вид:

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\cdot e^{-\frac{q(x,y)}{2}},$$

где

$$q(x,y)=\frac{1}{1-\rho^2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) —$$

квадратичная форма.

Параметры этого распределения определяют и его числовые характеристики:

$\mu_X=E(X)$ — математическое ожидание величины X,

$\mu_Y = E(Y)$ — математическое ожидание величины Y,

$\sigma_X = \sigma(X) = \sqrt{D(X)}$ — среднее квадратическое отклонение величины X,

$\sigma_Y = \sigma(Y) = \sqrt{D(Y)}$ — среднее квадратическое отклонение величины Y,

$\rho = \rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}$ — коэффициент корреляции величин X и Y.

Стандартно непрерывное двумерное распределение записывают как:

$$(X,Y) \sim N\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right).$$

Квадратичная форма $q(x,y)$ может быть представлена в матричном виде:

$$q(x,y) = \begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}^T \cdot C^{-1} \cdot \begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix},$$

где C^{-1} — матрица, обратная к ковариационной матрице:

$$C = C(X,Y) = \begin{pmatrix} D(X) & Cov(X,Y) \\ Cov(X,Y) & D(Y) \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & Cov(X,Y) \\ Cov(X,Y) & \sigma_Y^2 \end{pmatrix}.$$

Важно отметить, что компоненты X и Y двумерной нормальной величины (X,Y) имеют нормальные распределения:

$$X \sim N(\mu_X, \sigma_X^2),\ Y \sim N(\mu_Y, \sigma_Y^2).$$

Более того, условные распределения тоже подчиняются нормальному закону:

$$\left(X \middle| Y=y\right) \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y);\ \ \sigma_X^2\left(1-\rho^2\right)\right);$$

$$\left(Y \middle| X=x\right) \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(y-\mu_X);\ \ \sigma_Y^2\left(1-\rho^2\right)\right).$$

Отсюда заключаем, что

$$E\left(X \middle| Y=y\right) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y);\ D\left(X \middle| Y=y\right) = \sigma_X^2\left(1-\rho^2\right)$$

и

$$E\left(Y \middle| X=x\right) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\left(y-\mu_X\right);\ D\left(Y \middle| X=x\right) = \sigma_Y^2\left(1-\rho^2\right).$$

Доказательство этих красивых зависимостей выходит за рамки нашего курса.

2.9. Функции от случайных величин

Многие задачи теории вероятностей и математической статистики предполагают изучение суммы независимых величин. Остановимся подробнее на задачах нахождения и изучения закона распределения суммы большого числа независимых слагаемых.

Пусть X и Y — две дискретные независимые величины, формирующиеся в результате некоторого испытания, и пусть $Z = X + Y$.

Возможные значения для величины $Z = z$ определяются всеми возможными значениями $x + y$ значений величин $X = x$ и $Y = y$. Так как величины X и Y независимы, выполняется условие:

$$P(Z=z) = P(Z=x+y) = P(X=x, Y=y) = P(X=x)\cdot P(Y=y).$$

Принимая во внимание, что $y = z - x$, можно записать:

$$P(Z=z) = \sum_{\substack{x \\ x+y=z}} P(X=x)\cdot P(Y=z-x).$$

Суммирование ведется по всем значениям x при выполненном условии $y = z - x$.

При аналогичном условии относительно $P(X = z - y)$ можно записать:

$$P(Z=z) = \sum_{\substack{y \\ x+y=z}} P(Y=y)\cdot P(X=z-y).$$

Суммирование ведется по всем значениям y при выполненном условии $x = z - y$.

Полученные равенства определяют закон распределения $Z = X + Y$.

Мы уже рассматривали операции с независимыми дискретными величинами (см. пример 2.4), поэтому просто напомним, как найти закон распределения суммы независимых величин, предложив для этого вспомогательную таблицу несколько иного вида.

Пример 2.39. Даны законы распределений независимых величин X и Y:

Значение X (x_i)	1	2
Вероятность p_i	0,3	0,7

Значение Y (y_j)	–1	0
Вероятность p_j	0,2	0,8

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

Решение.

Составим вспомогательную таблицу, в которую включим данные распределения X и Y :

$Z = X + Y$		0,3	0,7
		$X = 1$	$X = 2$
0,2	$Y = -1$	$Z = 1 + (-1) = 0$	$Z = 2 + (-1) = 1$
		$0{,}3 \cdot 0{,}2 = 0{,}06$	$0{,}7 \cdot 0{,}2 = 0{,}14$
0,8	$Y = 0$	$Z = 1 + 0 = 1$	$Z = 2 + 0 = 2$
		$0{,}3 \cdot 0{,}8 = 0{,}24$	$0{,}7 \cdot 0{,}8 = 0{,}56$

Заполняем таблицу, вычисляя все возможные значения сумм значений X и Y , а соответствующие найденным значениям вероятности находим, используя условие независимости величин X и Y :

$$P(Z = z) = P(Z = x + y) = P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \ .$$

В итоге закон распределения величины $Z = X + Y$ примет вид:

Значение $Z = X + Y$	0	1	2
Вероятность	0,06	0,38	0,56

Поясним:

$$P(Z = X + Y = 0) = P(X = 1, Y = -1) = P(X = 1) \cdot P(Y = -1) = 0{,}3 \cdot 0{,}2 = 0{,}06;$$

$$P(Z = X + Y = 1) = P(X = 1, Y = 0 \text{ или } X = 2, Y = -1) =$$

$$= P(X = 1) \cdot P(Y = 0) + P(X = 2, Y = -1) =$$

$$= 0{,}3 \cdot 0{,}8 + 0{,}7 \cdot 0{,}2 = 0{,}38 ;$$

$$P(Z = X + Y = 2) = P(X = 2, Y = 0) = P(X = 2) \cdot P(Y = 0) = 0{,}7 \cdot 0{,}8 = 0{,}56.$$

Условие на закон распределения выполнен: сумма всех вероятностей во второй строке равна 1.

При составлении итоговой таблицы мы учли, что значение, равное 1, величина Z принимает два раза, поэтому применили условие $P(Z=z)=\sum\limits_{\substack{x \\ x+y=z}} P(X=x)\cdot P(Y=z-x)$, проще говоря, сложили вероятности, соответствующие одинаковым значениям суммы.

Ответ:

Значение $Z=X+Y$	0	1	2
Вероятность	0,06	0,38	0,56

Рассмотрим теперь случай непрерывных независимых величин X и Y. Если известна плотность вероятности $f_{X,Y}(x,y)$ двумерной величины (X,Y), то условие независимости принимает вид:

$$f_{X,Y}(x,y)=f_X(x)\cdot f_Y(y),$$

где

$$f_X(x)=\int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)dy;$$

$$f_Y(y)=\int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

Для величины $Z=X+Y$ плотность вероятности определяется условием

$$f_{X+Y}(z)=\int\limits_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx=\int\limits_{-\infty}^{\infty} f_{X,Y}(z-y,y)dy,$$

которое для независимых величин примет вид:

$$f_{X+Y}(z)=\int\limits_{-\infty}^{\infty} f_X(x)\cdot f_Y(z-x)dx=\int\limits_{-\infty}^{\infty} f_X(z-y)\cdot f_Y(y)dy.$$

Пример 2.40. Найти плотность вероятности суммы независимых величин X и Y, если известно, что $X\sim N(1;9)$, а Y равномерно распределена на отрезке $[-2;2]$.

Решение.

Для заданных величин найдем их плотности вероятностей:

$$f_X(x)=\frac{1}{\sqrt{2\pi}\cdot 3}e^{-\frac{(x-1)^2}{18}}.$$

Мы использовали стандартный вид плотности вероятности нормально распределенной величины и заданные по условию $X \sim N(1;9)$ параметры распределения: $E(X) = \mu = 1$ и $\sigma = \sigma(X) = \sqrt{D(X)} = \sqrt{9} = 3$.

Для равномерно распределенной на отрезке $[2;2]$ величины Y плотность вероятности имеет вид:

$$f_Y(y) = \begin{cases} 0, & y < -2 \\ \dfrac{1}{4}, & -2 \le y \le 2 \\ 0, & y > 2 \end{cases}.$$

Так как величины X и Y независимы, то для плотности вероятности их суммы $Z = X + Y$ применим условие:

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y) \cdot f_Y(y)dy$$

и учтем, что $f_Y(y) \neq 0$ только на отрезке $[-2;2]$, поэтому получим:

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y) \cdot f_Y(y)dy = \int_{-2}^{2} \frac{1}{4} \cdot \frac{1}{3\sqrt{2\pi}} e^{-\frac{(z-y-1)^2}{18}} dy =$$

$$= \frac{1}{4} \cdot \frac{1}{\sqrt{2\pi}} \int_{-2}^{2} \frac{1}{3} \cdot e^{-\frac{(z-y-1)^2}{18}} dy.$$

Последний интеграл вычисляют подстановкой $y = z - 1 + 3t$, $dy = 3dt$ при $z - const$:

$$\frac{1}{4} \cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{2+z-1}{3}}^{\frac{2-z+1}{3}} e^{-\frac{t^2}{2}} dt = \frac{1}{4}\left[\Phi\left(\frac{2-z+1}{3}\right) + \Phi\left(\frac{2+z-1}{3}\right)\right] =$$

$$= \frac{1}{4}\left[\Phi\left(\frac{3-z}{3}\right) + \Phi\left(\frac{1+z}{3}\right)\right].$$

В итоге имеем: $f_{X+Y}(z) = \dfrac{1}{4}\left[\Phi\left(\dfrac{3-z}{3}\right) + \Phi\left(\dfrac{1+z}{3}\right)\right]$.

Ответ: Плотность вероятности $Z = X + Y$ для заданных независимых величин X и Y имеет вид: $f_{X+Y}(z) = \dfrac{1}{4}\left[\Phi\left(\dfrac{3-z}{3}\right) + \Phi\left(\dfrac{1+z}{3}\right)\right]$.

Интересен пример с суммой независимых равномерно распределенных на отрезке $[0;1]$ величин X_i. Для наглядности мы рассмотрим только графики плотностей вероятностей таких сумм (рис. 2.9–2.11):

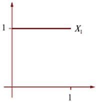


Рис. 2.9. График плотности вероятности величины X_1

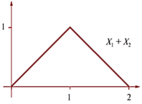


Рис. 2.10. График плотности вероятности суммы $X_1 + X_2$

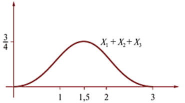


Рис. 2.11. График плотности вероятности суммы $X_1 + X_2 + X_3$

С увеличением количества слагаемых закон распределения их суммы неограниченно приближается к нормальному закону распределения. Эти закономерности имеют общий характер, устанавливаемый законом больших чисел и центральной предельной теоремой. Наш пример дает наглядное представление о том, как это происходит.

Задачи для самостоятельного решения

При решении задач **2.191–2.200** необходимо использовать понятия условного распределения и его числовых характеристик.

2.191. Известен закон распределения двумерной величины (X,Y):

	$X=0$	$X=1$	$X=3$	$X=4$
$Y=-1$	0,1	0	0,1	0,1
	0,1	0,1	0,1	0,1
$Y=1$	0,1	0,1	0,1	0

Составить условный закон распределения величины $Z=(Y|X>0)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E(Y|X)$ и $D(Y|X)$.

2.192. Известен закон распределения двумерной величины (X,Y):

	$X=4$	$X=5$	$X=6$	$X=7$
$Y=0$	0,1	0	0,1	0,1
$Y=1$	0	0,1	0	0,1
$Y=2$	0,1	0,2	0,2	0

Составить условный закон распределения величины $Z=(Y|X\le 6)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E(X|Y)$ и $D(X|Y)$.

2.193. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,2	0,1	0,1
$Y=0$	0	0,1	0	0,2
$Y=2$	0,1	0	0,1	0

Составить условный закон распределения величины $Z=(Y\|X|<2)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E(Y|X)$ и $D(Y|X)$.

2.194. Известен закон распределения двумерной величины (X,Y):

	$X=-3$	$X=-2$	$X=0$	$X=3$
$Y=-1$	0,1	0,2	0,1	0,2
$Y=1$	0	0,1	0	0,2
$Y=3$	0	0	0,1	0

Составить условный закон распределения величины $Z=(Y|X\le 0)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E(X|Y)$ и $D(X|Y)$.

2.195. Известен закон распределения двумерной величины (X,Y):

	$X=1$	$X=2$	$X=3$	$X=4$
$Y=2$	0,1	0	0,2	0
$Y=3$	0,1	0,1	0	0,1
$Y=4$	0,1	0,1	0,2	0

Составить условный закон распределения величины $Z=(Y|X\ne Y)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E(Y|X)$ и $D(Y|X)$.

2.196. Известен закон распределения двумерной величины (X,Y):

	$X=5$	$X=6$	$X=7$	$X=8$
$Y=7$	0,1	0	0,1	0
$Y=8$	0	0,2	0	0,2
$Y=9$	0,1	0,1	0,2	0

Составить условный закон распределения величины $Z=(Y|X=Y)$ и найти $E(Z)$ и $D(Z)$. Найти распределения $E(X|Y)$ и $D(X|Y)$.

2.197. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,2	0	0,1	0,2
$Y=0$	0,1	0,1	0	0,1
$Y=1$	0,1	0	0,1	0

Составить условный закон распределения величины $Z=\left(Y\middle|X\geq 0\right)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E\left(Y\middle|X\right)$ и $D\left(Y\middle|X\right)$.

2.198. Известен закон распределения двумерной величины (X,Y):

	$X=-4$	$X=-1$	$X=0$	$X=4$
$Y=-2$	0	0	0,1	0,2
$Y=0$	0,1	0,1	0.1	0,1
$Y=1$	0,1	0.1	0,1	0

Составить условный закон распределения величины $Z=\left(Y\middle||X|>|Y|\right)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E\left(X\middle|Y\right)$ и $D\left(X\middle|Y\right)$.

2.199. Известен закон распределения двумерной величины (X,Y):

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-1$	0,1	0,3	0,1	0
$Y=0$	0	0,1	0	0,1
$Y=1$	0,2	0	0,1	0

Составить условный закон распределения величины $Z=\left(Y\middle|X^2\leq Y^2\right)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E\left(Y\middle|X\right)$ и $D\left(Y\middle|X\right)$.

2.200. Известен закон распределения двумерной величины (X,Y):

	$X=-5$	$X=-4$	$X=-3$	$X=-2$
$Y=-1$	0,1	0,3	0,1	0
$Y=1$	0	0,1	0	0,1
$Y=4$	0,2	0	0,1	0

Составить условный закон распределения величины $Z=\left(Y\middle|X+Y<0\right)$ и найти $E(Z)$ и $D(Z)$. Составить законы распределений $E\left(X\middle|Y\right)$ и $D\left(X\middle|Y\right)$.

2.10. Закон больших чисел и предельные теоремы

Закон больших чисел понимается как совокупность утверждений, в каждом из которых при различных условиях рассматривается факт приближения средних характеристик к некоторым определенным постоянным при неограниченно большом числе испытаний. Эти утверждения позволяют утверждать, что средний результат действия большого числа испытаний практически перестает быть случайным и его можно предсказать с достаточной определенностью.

Если случайная величина X принимает только неотрицательные значения и для нее существует математическое ожидание $E(X)$, то для любого числа $\varepsilon > 0$ имеет место неравенство:

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Это неравенство называют ***неравенством Маркова*** и иногда записывают в другой форме:

$$P(X < \varepsilon) \geq 1 - \frac{E(X)}{\varepsilon}.$$

Неравенство Маркова позволяет получать первичные оценки вероятностей событий, связанных с тем, что значение случайной величины удовлетворяет известным ограничениям.

Пример 2.41. Отделение банка обслуживает в среднем 100 клиентов в день. Оценить вероятность того, что за день в этом отделении банка будет обслужено не менее 150 клиентов.

Решение.

Величина X — число клиентов, обслуживаемых банком за день, по условию задачи имеет математическое ожидание $E(X) = 100$. Для ответа на вопрос задачи применим неравенство Маркова

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

При $\varepsilon = 150$

$$P(X \geq 150) \leq \frac{100}{150} = \frac{2}{3}.$$

Ответ: вероятность того, что за день в отделении банка будет обслужено не менее 150 клиентов, будет не более $\frac{2}{3}$.

Прямым следствием неравенства Маркова является ***неравенство Чебышева***, которое выполняется для любой случайной величины X, имеющей математическое ожидание $E(X)$ и дисперсию $D(X)$:

$$P\left(\left|X - E(X)\right| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}.$$

Как и в неравенстве Маркова, ε является любым положительным числом.

Неравенство Чебышева может быть записано и в следующем виде:

$$P\left(\left|X - E(X)\right| < \varepsilon\right) \geq 1 - \frac{D(X)}{\varepsilon^2}.$$

Пример 2.42. Вероятность того, что средства, переданные клиентами на депозит банку, будут востребованы в течение полугода, равна 0,08 для каждого клиента. Оценить с помощью неравенства Чебышева вероятность того, что из 1000 клиентов от 70 до 90 в течение полугода затребуют возвращения вложенных средств.

Решение.

Величина X — число депозитов, отрытых банком, распределена по биномиальному закону, следовательно, можем найти ее числовые характеристики, используя известные формулы:

$$E(X) = np = 1000 \cdot 0{,}08 = 80;\ D(X) = npq = 1000 \cdot 0{,}08 \cdot 0{,}92 = 73{,}6.$$

Заданный для значения величины X интервал от 70 до 90 симметричен относительно математического ожидания, следовательно, выполнено условие:

$$P(70 < X < 90) = P\left(\left|X - 80\right| < 10\right),$$

поэтому значение ε определяется и равно 10. Теперь можно применить неравенство Чебышева и найти ответ на вопрос задачи:

$$P\left(\left|X - 80\right| < 10\right) \geq 1 - \frac{73{,}6}{10^2} = 0{,}264.$$

Ответ: вероятность того, из 1000 клиентов в течение полугода от 70 до 90 из них затребуют возвращения вложенных средств, составляет более 0,264.

Мы уже отметили, что закон больших чисел представляет собой совокупность утверждений. Для независимых случайных величин этот закон формулируется как ***теорема Чебышева.*** Если случайные величины $X_1, X_2, \ldots, X_n$ независимы, имеют одинаковые математические ожидания

$$E(X_1)=E(X_2)=\ldots=E(X_n)=a$$

и дисперсии, ограниченные одним и тем же числом C:

$$D(X_1)<C,\ D(X_2)<C,\ \ldots,\ D(X_n)<C\ ,$$

то для любого числа $\varepsilon>0$ будет выполнено условие:

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon\right)\geq 1-\frac{C}{n\varepsilon^2}\ ,$$

и если $n\to\infty$, то

$$\lim_{n\to\infty}P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon\right)=1\,.$$

Смысл теоремы Чебышева состоит в следующем: при большом числе случайных величин практически достоверно, что их средняя $\frac{X_1+X_2+\ldots+X_n}{n}$ — величина, по сути, **случайная**, бесконечно мало отличается от **неслучайной** постоянной величины a, то есть практически перестает быть случайной. Равный единице предел вероятности в общей теории означает сходимость по вероятности. В рассматриваемом случае можно утверждать, что среднее $\frac{X_1+X_2+\ldots+X_n}{n}$ для независимых величин $X_1,X_2,\ldots,X_n$ сходится по вероятности к значению a.

Пример 2.43. Дисперсия каждой из 1000 независимых величин $X_1,X_2,\ldots,X_{1000}$, имеющих одно и то же математическое ожидание, равна 4. Оценить вероятность того, что среднее арифметическое этих величин отклонится по модулю от среднего арифметического их математических ожиданий меньше, чем на 0,1.

Решение.

По условию $E(X_1)=E(X_2)=\ldots=E(X_{1000})=a-const$, и среднее арифметическое математических ожиданий этих величин тоже будет равно a:

$$\frac{E(X_1)+E(X_2)+\ldots+(X_n)}{n}=\frac{\underbrace{a+a+\ldots+a}_{n\ \text{раз}}}{n}=\frac{na}{n}=a.$$

Для ответа на вопрос задачи даже не обязательно знать значение a, достаточно воспользоваться теоремой Чебышева:

$$P\left(\left|\frac{X_1+X_2+\ldots+X_n}{n}-a\right|<\varepsilon\right)\geq 1-\frac{C}{n\varepsilon^2} \Rightarrow$$

$$\Rightarrow P\left(\left|\frac{X_1+X_2+\ldots+X_{1000}}{1000}-a\right|<0{,}1\right)\geq 1-\frac{4}{1000\cdot 0{,}1^2}=0{,}6 .$$

Ответ: вероятность того, что среднее этих величин отклонится по модулю от среднего арифметического их математических ожиданий меньше чем на 0,1, будет более 0,6.

Прямым следствием теоремы Чебышева является теорема Бернулли, которая дает теоретическое обоснование замене неизвестной вероятности события относительной частотой, полученной при n повторных независимых испытаниях, проводимых при одном и том же комплексе условий.

Теорема Бернулли. Если в каждом из n независимых испытаний вероятность $p=P(A)$ появления некоторого события A постоянна, то вероятность отклонения по модулю относительной частоты $\frac{k}{n}$ от вероятности $p=P(A)$ удовлетворяет условию для любого числа $\varepsilon>0$:

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2},$$

и если $n\to\infty$, то

$$\lim_{n\to\infty}P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)=1 .$$

Из условия теоремы Бернулли следует, что относительная частота наступления события $\frac{k}{n}$ сходится по вероятности к теоретической вероятности $p=P(A)$ этого события. Таким образом, обоснована замена неизвестной вероятности относительной частотой события.

Пример 2.44. Оценить вероятность того, что частота выпадения герба при 10000 подбрасываниях монеты отклонится по модулю от значения $\frac{1}{2}$ меньше чем на 0,01.

Решение.

Для ответа на вопрос задачи нам даже необязательно знать, сколько раз из 10000 подбрасываний выпал герб, чтобы точно определить относительную частоту выпадения герба, достаточно воспользоваться условием из теоремы Бернулли:

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2} \Rightarrow$$

$$\Rightarrow P\left(\left|\frac{k}{n}-\frac{1}{2}\right|<0{,}01\right)\geq 1-\frac{\frac{1}{2}\cdot\frac{1}{2}}{10000\cdot 0{,}01^2}=1-\frac{1}{4}=0{,}75\,.$$

Ответ: вероятность того, что частота выпадения герба при 10000 подбрасываниях монеты отклонится по модулю от $\frac{1}{2}$ меньше чем на 0,01, будет больше 0,75.

Центральная предельная теорема определяется группой теорем, устанавливающих условия возникновения нормального закона распределения. Важнейшее место в этой группе занимает теорема Ляпунова. Мы сформулируем ее в наиболее простой форме.

Теорема Ляпунова. Если случайные величины $X_1, X_2, ..., X_n$ независимы, имеют одинаковые математические ожидания и дисперсии:

$$E(X_1)=E(X_2)=...=E(X_n)=a\,;$$

$$D(X_1)=D(X_2)=...=D(X_n)=\sigma^2\,,$$

а также абсолютный центральный момент третьего порядка $\mu_3=E\left(\left|X-E(X)\right|^3\right)$, то при неограниченном увеличении n $(n\to\infty)$ закон распределения суммы этих величин

$$S_n=X_1+X_2+...+X_n$$

неограниченно приближается к нормальному закону.

Для решения задач удобно использовать «более аналитическую» формулировку этой теоремы, рассматривая переход от величины $S_n=X_1+X_2+...+X_n$ к центрированной и нормированной величине

$$S_n^{'}=\frac{S_n-E(S_n)}{\sigma(S_n)}\,.$$

После несложных преобразований с использованием свойств математического ожидания и дисперсии суммы независимых величин получим выражение для величины $S_n^{'}$:

$$S_n^{'} = \frac{S_n - na}{\sigma\sqrt{n}}.$$

Условие «неограниченно приближается к нормальному распределению» означает, что функция распределения величины $S_n^{'}$ в пределе даст функцию распределения нормально распределенной величины при $\mu = 0$ и $\sigma = 1$:

$$F(x) = P(X < x) = 0{,}5 + \Phi\left(\frac{x - 0}{1}\right) = 0{,}5 + \Phi(x).$$

В нашем случае имеем:

$$\lim_{n\to\infty} P(S_n^{'} < x) = 0{,}5 + \Phi(x)$$

Или подробнее:

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \ldots + X_n - na}{\sigma\sqrt{n}} < x\right) = 0{,}5 + \Phi(x).$$

Пример 2.45. Для независимых случайных величин X_1, X_2, ..., равномерно распределенных на отрезке [0; 15], найти

$$\lim_{n\to\infty} P\left(X_1 + X_2 + \ldots + X_n < \frac{15}{2}n + \sqrt{3n}\right).$$

Решение.

Рассмотрим величину

$S_n = X_1 + X_2 + \ldots + X_n$

и найдем ее числовые характеристики, учитывая, что все величины равномерно распределены на отрезке $[0; 15]$:

$$E(S_n) = E(X_1 + X_2 + \ldots + X_n) = \underbrace{\frac{0+15}{2} + \frac{0+15}{2} + \ldots + \frac{0+15}{2}}_{n} = \frac{15}{2}\cdot n;$$

$$D(S_n) = D(X_1 + X_2 + \ldots + X_n) = \underbrace{\frac{(15-0)^2}{12} + \frac{(15-0)^2}{12} + \ldots + \frac{(15-0)^2}{12}}_{n} =$$

$$= \frac{15^2}{12}\cdot n, \Rightarrow \sigma(S_n) = \sqrt{D(S_n)} = \sqrt{\frac{15^2}{12}\cdot n} = \frac{15\sqrt{n}}{2\sqrt{3}}.$$

Теперь определим величину $S_n^{'}$:

$$S_n^{'} = \frac{S_n - E(S_n)}{\sigma(S_n)} =$$

$$= \frac{X_1 + X_2 + ... + X_n - \dfrac{15}{2}n}{\dfrac{15\sqrt{n}}{2\sqrt{3}}}.$$

Для этой величины выполнена теорема Ляпунова:

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + ... + X_n - na}{\sigma\sqrt{n}} < x\right) = 0{,}5 + \Phi(x) \Rightarrow$$

$$\Rightarrow \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + ... + X_n - \dfrac{15}{2}n}{\dfrac{15\sqrt{n}}{2\sqrt{3}}} < x\right) = 0{,}5 + \Phi(x).$$

Преобразуем выражение под знаком предела из вопроса задачи таким образом, чтобы стало возможным применение теоремы Ляпунова:

$$\lim_{n\to\infty} P\left(X_1 + X_2 + ... + X_n < \frac{15}{2}n + \sqrt{3n}\right) =$$

$$= \lim_{n\to\infty} P\left(X_1 + X_2 + ... + X_n - \frac{15}{2}n < \sqrt{3n}\right) =$$

$$= \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + ... + X_n - \dfrac{15}{2}n}{\dfrac{15\sqrt{n}}{2\sqrt{3}}} < \frac{\sqrt{3n}}{\dfrac{15\sqrt{n}}{2\sqrt{3}}}\right) =$$

$$= \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + ... + X_n - \dfrac{15}{2}n}{\dfrac{15\sqrt{n}}{2\sqrt{3}}} < \frac{2}{5}\right) = 0{,}5 + \Phi(0{,}4) =$$

$= 0{,}5 + 0{,}15542 = 0{,}65542.$

Ответ. $\lim_{n\to\infty} P\left(X_1 + X_2 + ... + X_n < \frac{15}{2}n + \sqrt{3n}\right) = 0{,}65542.$

Задачи для самостоятельного решения

В задачах **2.201–2.210** при решении необходимо использовать закон больших чисел или центральную предельную теорему.

2.201. На студенческой научной конференции в среднем ежегодно выступают 100 человек. Оценить вероятность того, что в следующем году на студенческой конференции выступят более 120 человек.

2.202. В среднем 150 сотрудников предприятия ежегодно посещают курсы повышения квалификации. Оценить вероятность того, что в следующем году курсы повышения квалификации посетят не более 100 человек.

2.203. Средний возраст сотрудников предприятия составляет 35 лет, дисперсия равна 16. Оценить вероятность того, что возраст выбранного случайным образом сотрудника предприятия будет заключен в границах от 35 до 40 лет.

2.204. На предприятии брак составляет 5% от всего объема выпускаемой продукции. Оценить вероятность того, что при проверке партии из 2000 изделий выявится отклонения от установленного процента брака более чем на 2%.

2.205. Вероятность положительного исхода эксперимента составляет 0,8. Оценить вероятность того, что при проведении 1000 таких экспериментов отклонение доли положительных исходов отклонится от соответствующей вероятности по абсолютной величине не более чем на 0,05.

2.206. 2000 предприятий малого бизнеса некоторого города выбрали с целью определения среднего дохода. Известно, что дисперсия по каждому выбранному предприятию не превышает 5. Оценить вероятность того, что отклонение среднего дохода по выборке отличается от среднего дохода предприятий малого бизнеса по всему городу не более чем на 20 тыс.у.е.

2.207. Независимые случайные величины $X_1, X_2, \ldots$ одинаково распределены по геометрическому закону с математическим ожиданием, равным 3. Вычислить предел $\lim\limits_{n\to\infty} P\left(X_1 + X_2 + \ldots + X_n + \sqrt{6n} > 3n\right)$.

2.208. Независимые случайные величины $X_1, X_2, \ldots$ одинаково равномерно распределены на отрезке $[-3; 3]$. Вычислить предел $\lim\limits_{n\to\infty} P\left(\left|X_1 + X_2 + \ldots + X_n\right| > \sqrt{3n}\right)$.

2.209. Независимые случайные величины $X_1, X_2, \ldots$ одинаково распределены по закону Пуассона с математическим ожиданием, равным 3, независимые величины $Y_1, Y_2, \ldots$ тоже одинаково распределены по закону Пуассона, но с математическим ожиданием, равным 2. При этом величины $X_1, X_2, \ldots$ независимы от величин $Y_1, Y_2, \ldots$ Вычислить предел $\lim\limits_{n\to\infty} P\left(X_1 + X_2 + \ldots + X_n + n > Y_1 + Y_2 + \ldots + Y_n + \sqrt{5n}\right)$.

2.210. Независимые случайные величины $X_1, X_2, \ldots$ одинаково распределены по закону Пуассона с математическим ожиданием, равным 1, независимые величины $Y_1, Y_2, \ldots$ тоже одинаково распределены по закону Пуассона, но с математическим ожиданием, равным 2. При этом величины $X_1, X_2, \ldots$ независимы от величин $Y_1, Y_2, \ldots$ Вычислить предел $\lim\limits_{n\to\infty} P\left(X_1 + X_2 + \ldots + X_n + \sqrt{3n} < Y_1 + Y_2 + \ldots + Y_n + n\right)$.

ГЛАВА 3

ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

3.1. Эмпирические характеристики признака

Основной задачей математической статистики является разработка методов получения научно обоснованных выводов о массовых явлениях и процессах на основе данных, полученных в результате наблюдений или экспериментов. Эти выводы и заключения должны относиться не к отдельным испытаниям, а представлять собой утверждения об общих вероятностных характеристиках процесса или явления. Использование фактических данных для такого прогнозирования и является главной отличительной чертой статистического метода.

Для решения задач прогнозирования необходимо планирование статистических наблюдений и сбор статистических данных с дальнейшим анализом для принятия решений или прогнозирования случайных явлений.

Теоретической основой математической статистики является теория вероятностей. Чтобы качественно использовать теоретические положения теории вероятностей в математической статистике, вводят понятие ***признака*** — некоторой функции без фиксированной области определения. Например, в качестве признака X можно рассматривать размер заработной платы, значения которого могут определяться и на множестве работников бюджетной сферы, и на множестве менеджеров высшего звена компаний как в пределах региона, так и в масштабах всей страны.

Фиксированная область определения, связанная с признаком, называется ***статистической совокупностью***, а количество ее элементов — объемом. Если X — это признак, заданный на совокупности $\Omega = \{\overline{\omega}_1, \overline{\omega}_2, \ldots, \overline{\omega}_n\}$, то значения признака определятся как $x_1 = X(\overline{\omega}_1), x_2 = X(\overline{\omega}_2), \ldots, x_n = X(\overline{\omega}_n)$.

Для признаков определяется целый ряд эмпирических характеристик.

Средним значением признака (***эмпирическим средним***) X называют среднее арифметическое всех его значений в выбранной совокупности:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Дисперсией признака X в совокупности Ω (эмпирической дисперсией) называется среднее арифметическое квадратов отклонений значений признака от эмпирического среднего:

$$D(X) = \frac{\left(x_1 - \overline{x}\right)^2 + \left(x_2 - \overline{x}\right)^2 + \ldots + \left(x_n - \overline{x}\right)^2}{n},$$

а ***стандартным отклонением признака*** называют

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

Эмпирический начальный момент k-го порядка признака X определяется как

$$\nu_k(X) = \frac{x_1^k + x_2^k + \ldots + x_n^k}{n}.$$

Эмпирический центральный момент k-го порядка признака X определяется соответственно как

$$\mu_k(X) = \frac{\left(x_1 - \overline{x}\right)^k + \left(x_2 - \overline{x}\right)^k + \ldots + \left(x_n - \overline{x}\right)^k}{n}.$$

Можно определить и ***эмпирическую функцию распределения*** $F(x)$:

$$F(x) = \frac{\left\langle \textit{число элементов } \omega \in \Omega, \textit{ для которых } X(\omega) < x \right\rangle}{n}.$$

Признак ведет себя как случайная величина (мы даже сохранили обозначение). Если возможные значения этой величины считать равновероятными, то все эмпирические характеристики совпадут с числовыми характеристиками случайной величины, которые мы рассматривали в теории вероятностей.

Значения признака $x_1, x_2, \ldots, x_n$, необязательно различные, расположенные в неубывающем порядке (ранжированные), называют ***вариационным рядом***. Сами значения признака часто называют ***вариантами***.

Разность между наибольшей и наименьшей вариантой в вариационном ряду называют ***размахом признака,*** или ***размахом вариации***.

Порядковый номер середины (центра) вариационного ряда называется ***эмпирической медианой*** Me признака. В зависимости от количества вариант n (нечетного или четного) в вариационном ряду медиану вычисляют по-разному:

$$Me = \begin{cases} x_{k+1}, & n = 2k+1 \\ \dfrac{x_k + x_{k+1}}{2}, & n = 2k \end{cases}.$$

Например, для вариационного ряда $2; 2; 3; 4; 4; 5; 5$ количество вариант $n = 7$ — нечетное:

$$n = 2k+1 \;\Rightarrow\; 7 = 2\cdot 3+1 \;\Rightarrow\; k = 3,$$

и нас будет интересовать четвертое (по порядку) значение варианты в вариационном ряду, то есть $x_{k+1} = x_{3+1} = x_4 = 4$, поэтому

$$Me = x_{k+1} = 4.$$

Для вариационного ряда $2; 2; 4; 4; 5; 5$ количество вариант $n =$ $= 6$ — четное:

$$n = 2k \;\Rightarrow\; 6 = 2\cdot 3 \;\Rightarrow\; k = 3.$$

Медиану в таком случае вычисляем по формуле:

$$Me = \frac{x_k + x_{k+1}}{2} = \frac{x_3 + x_4}{2} = \frac{4+4}{2} = 4.$$

Если в совокупности Ω значение x_1 наблюдалось n_1 раз, значение $x_2 - n_2$ раза, ..., $x_k - n_k$ раз, то числа $n_1, n_2, ..., n_k$ называются частотами вариант $(n_1 + n_2 + ... + n_k = n)$. Таблица вида:

Значение признака (варианты)	x_1	x_2	...	x_k
Частота	n_1	n_2	...	n_k

определяет ***частотное распределение*** признака и называется статистическим распределением.

Относительной частотой варианты ω_i называют отношение $\dfrac{n_i}{n}$,

а таблица вида

Значение признака (варианты)	x_1	x_2	...	x_k
Относительная частота	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_k}{n}$

определяется как распределение относительных частот и называется ***эмпирическим распределением признака***.

Понятно, что для распределения относительных частот выполняется условие: $\sum_{i=1}^{k} \omega_i = \frac{n_1 + n_2 + \ldots + n_k}{n} = 1$.

Эмпирические характеристики при известном частотном (а, следовательно, и эмпирическом) распределении признака вычисляются по формулам:

$$\overline{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{n} = x_1\omega_1 + x_2\omega_2 + \ldots + x_k\omega_k ;$$

$$D(X) = \frac{\left(x_1 - \overline{x}\right)^2 \cdot n_1 + \left(x_2 - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(x_k - \overline{x}\right)^2 \cdot n_k}{n} =$$

$$= \left(x_1 - \overline{x}\right)^2 \cdot \omega_1 + \left(x_2 - \overline{x}\right)^2 \cdot \omega_2 + \ldots + \left(x_k - \overline{x}\right)^2 \cdot \omega_k .$$

Для эмпирической дисперсии верна формула

$$D(X) = \frac{x_1^2 \cdot n_1 + x_2^2 \cdot n_2 + \ldots + x_k^2 \cdot n_k}{n} - \left(\overline{x}\right)^2 .$$

Формулы для вычисления моментов принимают вид:

$$\nu_k = \frac{x_1^k \cdot n_1 + x_2^k \cdot n_2 + \ldots + x_k^k \cdot n_k}{n} ;$$

$$\mu_k = \frac{\left(x_1 - \overline{x}\right)^k \cdot n_1 + \left(x_2 - \overline{x}\right)^k \cdot n_2 + \ldots + \left(x_k - \overline{x}\right)^k \cdot n_k}{n} .$$

Отношение

$$\overline{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{n_1 + n_2 + \ldots + n_k}$$

часто называют ***взвешенным средним*** чисел $x_1, x_2, \ldots, x_k$, а значение n_i — ***весом*** варианты x_i.

Наиболее распространенным графическим представлением вариационного ряда является полигон частот — ломаная, соединяющая точки с координатами $(x_1;n_1),...,(x_k;n_k)$ (рис. 3.1).

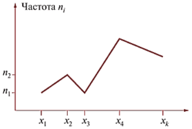


Рис. 3.1. Полигон частот вариационного ряда

Полигон частот дает наглядное представление о закономерности изменения значений наблюдаемого признака.

Модой Mo вариационного ряда называется значение варианты x_i, которому соответствует наибольшая частота n_i.

Моду вариационного ряда наряду с эмпирическим средним рассматривают как меру центральной тенденции вариационного ряда.

Для характеристики вариационного ряда привлекают асимметрию $\mathrm{As} = \frac{\mu_3}{\sigma^3}$ и эксцесс $\mathrm{Ex} = \frac{\mu_4}{\sigma^4} - 3$. Если асимметрия равна нулю, то говорят о симметричности вариационного ряда.

Пример 3.1. Изучается признак X — процент по ипотечному кредитованию в банках Москвы и Московской области. Получены следующие результаты:

11; 10,5; 11; 11; 13; 10,5; 11; 10,5; 12; 10,5.

Задать частотное распределение признака и эмпирическую функцию распределения, построить полигон частот и вычислить эмпирические характеристики признака.

Решение.

На основе полученных данных сформируем вариационный ряд, для этого наблюдаемые значения признака расположим в неубывающем порядке:

10,5; 10,5; 10,5; 10,5; 11; 11; 11; 11; 12; 13.

Теперь можно задать частотное распределение:

Значение признака (варианты)	10,5	11	12	13
Частота	4	4	1	1

$$n = 4 + 4 + 1 + 1 = 10$$

и построить полигон частот (рис. 3.2):

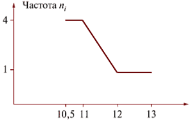


Рис. 3.2. Полигон частот к примеру 3.1

Перед тем, как задать эмпирическую функцию распределения, определим распределение относительных частот:

Значение признака (варианты)	10,5	11	12	13
Относительная частота $\frac{n_i}{n}$	0,4	0,4	0,1	0,1

$$\sum_i \frac{n_i}{n} = 0,4 + 0,4 + 0,1 + 0,1 = 1\ .$$

Эмпирическая функция распределения имеет вид:

$$F(x) = \begin{cases} 0, & x \le 10,5 \\ 0,4, & 10,5 < x \le 11 \\ 0,8, & 11 < x \le 12 \\ 0,9, & 12 < x \le 13 \\ 1, & x > 13 \end{cases}.$$

Теперь вычислим эмпирические характеристики признака:

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{n} = \frac{10,5 \cdot 4 + 11 \cdot 4 + 12 \cdot 1 + 13 \cdot 1}{10} = 11,1\ ;$$

$$D(X) = \frac{\left(x_1 - \bar{x}\right)^2 \cdot n_1 + \left(x_2 - \bar{x}\right)^2 \cdot n_2 + \ldots + \left(x_k - \bar{x}\right)^2 \cdot n_k}{n} =$$

$$= \frac{(10{,}5-11{,}1)^2\cdot 4+(11-11{,}1)^2\cdot 4+(12-11{,}1)^2\cdot 1+(13-11{,}1)^2\cdot 1}{10} =$$

$= 0{,}59;$

$\sigma(X)=\sqrt{D(X)}=\sqrt{0{,}59}=0{,}768$;

$Mo = 10{,}5$ и $Mo = 11$ — распределение бимодальное;

$Me=\dfrac{x_k+x_{k+1}}{2}=\dfrac{x_5+x_6}{2}=\dfrac{11+11}{2}=11$ (для четного значения $n=10$);

$$\text{As}=\frac{\mu_3}{\sigma^3}=\frac{(x_1-\bar{x})^3\cdot n_1+(x_2-\bar{x})^3\cdot n_2+\ldots+(x_k-\bar{x})^3\cdot n_k}{n\cdot\sigma^3}=$$

$$=\frac{(10{,}5-11{,}1)^3\cdot 4+(11-11{,}1)^3\cdot 4+(12-11{,}1)^3\cdot 1+(13-11{,}1)^3\cdot 1}{10\cdot 0{,}768^3}=$$

$= 1{,}485;$

$$\text{Ex}=\frac{\mu_4}{\sigma^4}-3=\frac{(x_1-\bar{x})^4\cdot n_1+(x_2-\bar{x})^4\cdot n_2+\ldots+(x_k-\bar{x})^4\cdot n_k}{n\cdot\sigma^4}-3=$$

$$=\frac{(10{,}5-11{,}1)^4\cdot 4+(11-11{,}1)^4\cdot 4+(12-11{,}1)^4\cdot 1+(13-11{,}1)^4\cdot 1}{10\cdot 0{,}768^4}-3=$$

$= 1{,}084.$

Мы ответили на все вопросы задачи.

Если количество значений вариант велико, то задается ***интервальный вариационный ряд*** распределения. Чтобы построить такой ряд, весь интервал варьирования значений признака разбивают на несколько интервалов и подсчитывают, какое количество значений признака попадает в каждый интервал.

Рекомендуемое количество интервалов вычисляют по формуле:

$$k = 1 + 3{,}322\lg n.$$

Длину интервала в этом случае находят по формуле:

$$h=\frac{x_{\max}-x_{\min}}{k}=\frac{x_{\max}-x_{\min}}{1+3{,}322\lg n}.$$

Если получаемое число — дробное, то в качестве длины интервала рассматривают ближайшее целое число или простую дробь (обыкновенную или десятичную).

Интервальное статистическое распределение имеет вид:

Интервал	$(a_1;a_2]$	$(a_2;a_3]$	...	$(a_k;a_{k+1}]$
Частота	n_1	n_2	...	n_k

Для вычисления эмпирических характеристик признака находят середину каждого интервала

Интервалы	$(a_1;a_2]$	$(a_2;a_3]$	...	$(a_k;a_{k+1}]$
Середина интервала x_i^*	$x_1^*=\dfrac{a_1+a_2}{2}$	$x_2^*=\dfrac{a_2+a_3}{2}$		$x_k^*=\dfrac{a_k+a_{k+1}}{2}$
Частота	n_1	n_2	...	n_k

и используют эти значения в качестве значений вариант с соответствующими частотами:

$$\overline{x}=\frac{x_1^*\cdot n_1+x_2^*\cdot n_2+\ldots+x_k^*\cdot n_k}{n};$$

$$D(X)=\frac{\left(x_1^*-\overline{x}\right)^2\cdot n_1+\left(x_2^*-\overline{x}\right)^2\cdot n_2+\ldots+\left(x_k^*-\overline{x}\right)^2\cdot n_k}{n};$$

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

Аналогично вычисляются эмпирические начальные и центральные моменты.

Для вычисления моды интервального вариационного ряда используют следующую формулу:

$$Mo=x_{Mo}+h\cdot\frac{n_{Mo}-n_{Mo-1}}{\left(n_{Mo}-n_{Mo-1}\right)+\left(n_{Mo}-n_{Mo+1}\right)},$$

где

x_{Mo} — нижняя граница модального интервала,

h — длина (шаг) интервала,

n_{Mo} — частота модального интервала,

n_{Mo-1} — частота интервала, предшествующего модальному,

n_{Mo+1} — частота интервала, следующего за модальным.

Модальным считается интервал с наибольшим количеством значений вариант.

Медиану интервального вариационного ряда вычисляют по формуле:

$$Me = x_{Me} + h \cdot \frac{\frac{n}{2} - S_{Me-1}}{n_{Me}},$$

где

x_{Me} — нижняя граница медианного интервала,

h — длина (шаг) интервала,

n_{Me} — частота медианного интервала,

n — общее количество значений вариант,

S_{Me-1} — сумма накопленных частот до медианного интервала.

Медианным считается интервал, которому принадлежит $\frac{n}{2}$-е (в случае четного значения n) или $\frac{n+1}{2}$-е (в случае нечетного значения n) значение варианты.

Пример 3.2. Изучается признак X — среднемесячная заработная плата работников в целом по экономике в марте 2014 года по 83 субъектам Российской Федерации. Получены следующие результаты:

Значение X (тыс. руб.)	18–25	25–32	32–39	39–46	46–53	53–60	60–67	67–74
Частота	44	22	5	2	2	4	2	2

Вычислить эмпирические характеристики признака.

Решение.

Найдем середины каждого интервала и дополним таблицу:

Значение X (тыс. руб.)	18–25	25–32	32–39	39–46	46–53	53–60	60–67	67–74
Середина x_i^*	21,5	28,5	35,5	42,5	49,5	56,5	63,5	70,5
Частота	44	22	5	2	2	4	2	2

$$n = 44 + 22 + 5 + 2 + 2 + 4 + 2 + 2 = 83.$$

Найдем эмпирические характеристики признака:

$$\overline{x} = \frac{x_1^* \cdot n_1 + x_2^* \cdot n_2 + \ldots + x_k^* \cdot n_k}{n} = \frac{21{,}5 \cdot 44 + \ldots + 70{,}5 \cdot 2}{83} = 29{,}256;$$

$$D(X) = \frac{\left(x_1^* - \overline{x}\right)^2 \cdot n_1 + \left(x_2^* - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(x_k^* - \overline{x}\right)^2 \cdot n_k}{n} =$$

$$= \frac{(21{,}5-29{,}256)^2 \cdot 44 + \ldots + (70{,}5-29{,}256)^2 \cdot 2}{83} = 153{,}508201;$$

$$\sigma(X) = \sqrt{D(X)} = 12{,}39.$$

Для вычисления моды определим модальный интервал. Это интервал с наибольшей частотой, то есть первый: от 18 до 25. Теперь воспользуемся формулой для вычисления моды:

$$Mo = x_{Mo} + h \cdot \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})} =$$

$$= 18 + 7 \cdot \frac{44-0}{(44-0)+(44-22)} = 22{,}67.$$

Для вычисления медианы найдем медианный интервал. Для этого определим, в каком интервале находится $\frac{83+1}{2} = 42$-е значение варианты (используем строку с частотами). Приходим к выводу, что это тоже первый интервал: от 18 до 25. Воспользуемся формулой для вычисления медианы:

$$Me = x_{Me} + h \cdot \frac{\frac{n}{2} - S_{Me-1}}{n_{Me}} = 18 + 7 \cdot \frac{\frac{83}{2} - 0}{44} = 24{,}602.$$

Последние характеристики позволяют сделать выводы о том, что в половине субъектов РФ среднемесячная заработная плата работников в целом по экономике в марте 2014 года была меньше 24,602 тысяч рублей, что ниже эмпирического среднего значения 29,256; модальное значение тоже получилось ниже эмпирического среднего значения. Используя середины интервалов в качестве значений вариант, вычисляем эмпирическую асимметрию

$$As = \frac{\mu_3}{\sigma^3} = \frac{(21{,}5-29{,}256)^3 \cdot 44 + \ldots + (70{,}5-29{,}256)^3 \cdot 2}{83 \cdot 12{,}39^3} = 0{,}953$$

и эмпирический эксцесс

$$Ex = \frac{\mu_4}{\sigma^4} - 3 = \frac{(21{,}5-29{,}256)^4 \cdot 44 + \ldots + (70{,}5-29{,}256)^4 \cdot 2}{83 \cdot 12{,}39^4} - 3 = 0{,}353.$$

Ответ: $\overline{x} = 29{,}256$; $D(X) = 153{,}508201$; $\sigma(X) = 12{,}39$;

$Mo = 22{,}67$; $Me = 24{,}602$; $As = 0{,}953$; $Ex = 0{,}353$.

В тех случаях, когда эмпирическая функция хорошо приближается функцией нормального распределения, для вычисления эмпирического стандартного отклонения применяют поправку Шепарда:

$$\sigma(X) \approx \sqrt{D(X) - \frac{h^2}{12}}.$$

Графическим изображением интервального статистического распределения является ***гистограмма***. Для ее построения по оси *Ox* откладывают интервалы шириной h, затем на каждом интервале строят прямоугольник высотой n_i — частотой, соответствующей выбранному интервалу.

Для интервального вариационного ряда из примера 3.2 гистограмма имеет вид (рис. 3.3):

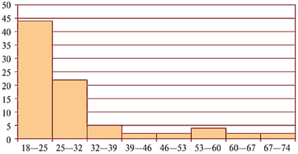


Рис. 3.3. Гистограмма к примеру 3.2

Рассмотрим признак *X*, наблюдаемый на совокупности Ω объема *n*, разбитой на *k* групп заданных объемов n_i ($i = 1, 2, \ldots, k$). Обозначим через x_{ij} значение признака на j-м элементе i-й группы. Эмпирическое среднее и эмпирическую дисперсию в таком случае можно вычислять, используя соответствующие характеристики каждой группы.

Например, если $\overline{x_i} = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$ — эмпирическое среднее в i-й группе, то общее среднее находят по формуле:

$$\overline{x} = \frac{\sum_{i=1}^{n} \overline{x_i} \cdot n_i}{n}.$$

Для общей дисперсии справедлива формула:

$$\sigma^2 = \overline{\sigma}^2 + \delta^2 ,$$

где $\sigma^2 = D(X)$ — эмпирическая дисперсия признака во всей совокупности Ω ;

$\overline{\sigma}^2 = \dfrac{\sum_{i=1}^{k} \sigma_i^2 \cdot n_i}{n}$ — средняя групповая дисперсия из дисперсий в группах;

$D_i(X) = \sigma_i^2 = \dfrac{\sum_{j=1}^{n_i} \left(x_{ij} - \overline{x_i}\right)^2}{n_i}$, $\delta^2 = \dfrac{\sum_{i=1}^{k} \left(\overline{x_i} - \overline{x}\right)^2 \cdot n_i}{n}$ — межгрупповая дисперсия (дисперсия средних).

Средняя групповая дисперсия характеризует изменчивость признака в каждой группе, поэтому иногда называется внутригрупповой дисперсией, а межгрупповая дисперсия характеризует разброс групповых средних от общего среднего.

Важной эмпирической характеристикой признака является ***коэффициент детерминации***:

$$\eta = \frac{\delta^2}{\sigma^2} ,$$

определяющий долю межгрупповой дисперсии в общей дисперсии.

Если на совокупности Ω объема *n* наблюдаются два признака *X* и *Y* и известно их совместное частотное распределение

	$X = x_1$	$X = x_2$	…	$X = x_k$
$Y = y_1$	n_{11}	n_{21}	…	n_{k1}
…	…	…	…	…
$Y = y_m$	n_{1m}	n_{2m}		n_{km}

(сумма всех частот n_{ij} равна объему совокупности *n*), то для них может быть вычислена ***эмпирическая ковариация*** по формуле:

$$\mathrm{Cov}(X,Y) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} \left(x_i - \overline{x}\right) \cdot \left(y_j - \overline{y}\right) \cdot n_{ij}}{n} .$$

Эмпирический коэффициент корреляции определяется формулой:

$$\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}.$$

Пример 3.3. На 16 фермерских хозяйствах изучали два признака: X — количество прямых договоров о поставках продукции и Y — количество кредитов, взятых в банках. Было получено совместное статистическое распределение признаков:

	$X=2$	$X=3$	$X=4$	$X=5$
$Y=3$	1	0	1	0
$Y=4$	2	4	4	2
$Y=5$	0	1	0	1

Найти эмпирический коэффициент корреляции $\rho(X,Y)$.

Решение.

На основе заданного совместного статистического распределения найдем статистические распределения признака X и признака Y:

Вариант x_i	2	3	4	5
Частота n_i	3	5	5	3

При составлении распределения частоты соответствующие определенному значению варианты суммируются (выделено в таблицах).

Аналогично для статистического распределения Y:

Вариант y_j	3	4	5
Частота n_j	2	12	2

Найдем эмпирические характеристики признаков X и Y:

$$\overline{x}=\frac{2\cdot3+3\cdot5+4\cdot5+5\cdot3}{16}=\frac{56}{16}=3{,}5;$$

$$D(X)=\frac{2^2\cdot3+3^2\cdot5+4^2\cdot5+5^2\cdot3}{16}-(3{,}5)^2=\frac{212}{16}-12{,}25=1;$$

$$\sigma(X)=\sqrt{D(X)}=1;$$

$$\overline{y}=\frac{3\cdot2+4\cdot12+5\cdot2}{16}=\frac{64}{16}=4;$$

$$D(Y)=\frac{3^2\cdot2+4^2\cdot12+5^2\cdot2}{16}-(4)^2=\frac{260}{16}-16=0{,}25;$$

$$\sigma(Y)=\sqrt{D(Y)}=0{,}5.$$

Из заданного совместного статистического распределения и с учетом средних значений признаков X и Y находим:

$$\mathrm{Cov}(X,Y)=\frac{\sum_{i=1}^{k}\sum_{j=1}^{m}\left(x_i-\overline{x}\right)\cdot\left(y_j-\overline{y}\right)\cdot n_{ij}}{n}=\frac{\sum_{i=1}^{k}\sum_{j=1}^{m}x_i y_j\cdot n_{ij}}{n}-\overline{x}\cdot\overline{y}=$$

$$=\frac{2\cdot3\cdot1+4\cdot3\cdot1+2\cdot4\cdot2+3\cdot4\cdot4+4\cdot4\cdot4+5\cdot4\cdot2+3\cdot5\cdot1+5\cdot5\cdot1}{16}-3{,}5\cdot4=$$

$$=14{,}125-14=0{,}125\,.$$

Теперь найдем ответ на вопрос задачи:

$$\rho\left(X,Y\right)=\frac{\mathrm{Cov}\left(X,Y\right)}{\sigma\left(X\right)\cdot\sigma\left(Y\right)}=\frac{0{,}125}{1\cdot0{,}5}=0{,}25\,.$$

Ответ: $\rho X,Y)=0{,}25$

3.2. Выборочный метод в статистике

Генеральной совокупностью называют все множество объектов или явлений, подлежащих изучению. Можно говорить о генеральной совокупности и как о совокупности всех возможных наблюдений, которые могли бы быть проведены при определенном комплексе условий. Прямое изучение генеральной совокупности очень трудоемко, да и не всегда возможно или оправдано. Генеральные совокупности изучаются на основе выборочных совокупностей или ***выборок***, которые должны достаточно хорошо отражать пропорции генеральной совокупности. В таком случае выборка называется ***репрезентативной***. Одна из задач математической статистики состоит в том, чтобы на основе анализа выборочной совокупности сделать обоснованные выводы относительно всей генеральной совокупности. Существуют два способа формирования выборки: ***повторный*** и ***бесповторный***. В первом случае выбираемый объект возвращается в генеральную совокупность (и может повторно попасть в выборку), во втором случае выбранный объект в генеральную совокупность уже не возвращается, поэтому бесповторная выборка состоит из различных объектов.

Генеральную совокупность принято обозначать Ω. Объем генеральной совокупности будем обозначать N. Выборку объема n из генеральной совокупности принято обозначать $\tilde{\Omega}$. Пусть X — признак, заданный в генеральной совокупности, тогда $x_{01},x_{02},\ldots,x_{0N}$ — все возможные значения этого признака в генеральной совокупности. Значения признака в выборочной совокупности ведут себя как случайные величины $X_1, X_2, ... X_n$ с множеством возможных значений $x_{01}, x_{02}, ..., x_{0N}$.

Характеристики признака в генеральной совокупности называют ***генеральными характеристиками***, они не являются случайными величинами и определяются как

$$\overline{x_0} = \frac{x_{01} + x_{02} + ... + x_{0N}}{N}$$ — генеральное среднее;

$$D(X) = \frac{\sum_{i=1}^{N}\left(x_{0i} - \overline{x_0}\right)^2}{N}$$ — генеральная дисперсия;

$$\sigma(X) = \sqrt{D(X)}$$ — генеральное стандартное отклонение.

Еще раз подчеркнем, что генеральные характеристики являются ***числами***, а не случайными величинами.

Эмпирические характеристики признака в выборочной совокупности называются ***выборочными характеристиками*** признака X. Понятно, что выборочные характеристики являются *случайными величинами*:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$ — выборочное среднее;

$$D_в(X) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}$$ — выборочная дисперсия;

$$\sigma_в(X) = \sqrt{D_в(X)}$$ — выборочное стандартное отклонение.

Закон больших чисел является теоретической основой для применения выборочного метода: при неограниченном увеличении объема выборки ее характеристики сколь угодно близко приближаются к характеристикам генеральной совокупности.

Существует доказанная связь между выборочными и генеральными характеристиками, при этом учитывается и способ формирования выборочной совокупности.

В случае ***повторной*** выборки имеют место равенства:

$$E\left(\overline{X}\right) = \overline{x_0}\ ,\ \ D\left(\overline{X}\right) = \frac{D(X)}{n}.$$

Здесь $\overline{X}$ — выборочное среднее, которое является случайной величиной, и поэтому можно найти ее математическое ожидание и дисперсию. Мы видим, что математическое ожидание выборочного среднего для повторной выборки равно генеральному среднему, а дисперсия выборочного среднего повторной выборки определяется генеральной дисперсией, деленной на объем выборки.

В случае ***бесповторной*** выборки имеют место равенства:

$$E\left(\overline{X}\right)=\overline{x_0}\ ,\ D\left(\overline{X}\right)=\frac{D(X)}{n}\cdot\frac{N-n}{N-1}\ .$$

И здесь $\overline{X}$ — выборочное среднее, которое является случайной величиной. Ее математическое ожидание и для бесповторной выборки равно генеральному среднему, а дисперсия выборочного среднего бесповторной выборки теперь определяется генеральной дисперсией, деленной на объем выборки и умноженной на отношение $\frac{N-n}{N-1}$, где N — объем генеральной совокупности.

Важно отметить, что при известных характеристиках генеральной совокупности с помощью указанных формул всегда можно оценить выборочное среднее, для этого необходимо знать только объем выборки и способ ее формирования.

Средней ошибкой выборки называют среднеквадратическую ошибку в равенстве $\overline{x_0}\approx\overline{X}$ и обозначают $\sigma\left(\overline{X}\right)$. Понятно, что $\sigma\left(\overline{X}\right)=\sqrt{D\left(\overline{X}\right)}$,

поэтому среднюю ошибку выборки в случае повторного отбора можно вычислить как

$$\sigma\left(\overline{X}\right)=\sqrt{D\left(\overline{X}\right)}=\sqrt{\frac{D(X)}{n}}=\frac{\sigma(X)}{\sqrt{n}},$$

а в случае бесповторного отбора средняя ошибка выборки определяется по формуле

$$\sigma\left(\overline{X}\right)=\sqrt{D\left(\overline{X}\right)}=\sqrt{\frac{D(X)}{n}\cdot\frac{N-n}{N-1}}=\sigma(X)\cdot\sqrt{\frac{N-n}{n\cdot(N-1)}}\ .$$

Напомним, что $\sigma(X)$ — генеральное стандартное отклонение.

При ***заданной*** средней ошибке $\sigma\left(\overline{X}\right)$ и известных генеральных характеристиках из полученных равенств определяется ***необходимый объем выборки***, при котором эта ошибка возможна.

В случае ***повторной*** выборки этот объем вычисляется как

$$n=\left(\frac{\sigma(X)}{\sigma(\overline{X})}\right)^2,$$

а в случае ***бесповторного*** отбора он определяется по формуле

$$n=\frac{D(X)\cdot N}{N\cdot\left(\sigma(\overline{X})\right)^2+D(X)}.$$

Напомним, что генеральная дисперсия $D(X)=\sigma^2(X)$.

Можно утверждать, что ***независимо от способа формирования выборки*** выполняется неравенство

$$\sigma(\overline{X})\le\frac{\sigma(X)}{\sqrt{n}},$$

где

$\overline{X}$ — выборочное среднее;

$\sigma(X)$ — генеральное стандартное отклонение;

n — объем выборочной совокупности.

Если в генеральной совокупности Ω объема N изучать два признака X и Y по выборке объема n, то, обозначив значения признаков в выборке как $X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n$, можно получить выражение для ковариации выборочных средних $\overline{X}$ и $\overline{Y}$:

$\mathrm{Cov}(\overline{X},\overline{Y})=\dfrac{\mathrm{Cov}(X,Y)}{n}$ — в случае повторной выборки;

$\mathrm{Cov}(\overline{X},\overline{Y})=\dfrac{\mathrm{Cov}(X,Y)}{n}\cdot\dfrac{N-n}{N-1}$ — в случае бесповторной выборки.

Пример 3.4. В генеральной совокупности из 100 торговых предприятий в течение месяца наблюдали за двумя признаками: X — количеством жалоб потребителей на качество продаваемой продукции и Y — количеством проверок торговой инспекции. Результаты оказались следующие:

	$X=5$	$X=10$	$X=15$
$Y=0$	10	20	40
$Y=1$	5	10	15

Из генеральной совокупности случайным образом выбирают 10 предприятий. Найти среднюю ошибку выборки для количества жалоб потребителей на качество продаваемой продукции, а также ковариацию выборочных средних $\overline{X}$ и $\overline{Y}$. Рассмотреть условие повторной и бесповторной выборки.

Решение.

Объем генеральной совокупности нам известен: $N = 100$.

Найдем генеральные характеристики:

$\overline{x_0} = \dfrac{x_{01} + x_{02} + \ldots + x_{0N}}{N} = \dfrac{5 \cdot 15 + 10 \cdot 30 + 15 \cdot 55}{100} = 12$ — генеральное среднее для X;

$D(X) = \dfrac{\sum\limits_{i=1}^{N}\left(x_{0i} - \overline{x_0}\right)^2}{N} = \dfrac{5^2 \cdot 15 + 10^2 \cdot 30 + 15^2 \cdot 55}{100} - 12^2 = 13{,}5$ — генеральная дисперсия для X;

$\sigma(X) = \sqrt{D(X)}\ \sqrt{13{,}5} = 3{,}674$ — генеральное стандартное отклонение для X;

$\overline{y_0} = \dfrac{y_{01} + y_{02} + \ldots + y_{0N}}{N} = \dfrac{0 \cdot 70 + 1 \cdot 30}{100} = 0{,}3$ — генеральное среднее для Y;

$D(Y) = \dfrac{\sum\limits_{i=1}^{N}\left(y_{0i} - \overline{y_0}\right)^2}{N} = \dfrac{0^2 \cdot 70 + 1^2 \cdot 30}{100} - 0{,}3^2 = 0{,}21$ — генеральная дисперсия для Y;

$\sigma(Y) = \sqrt{D(Y)} = \sqrt{0{,}21} = 0{,}458$ — генеральное стандартное отклонение для Y;

$$\mathrm{Cov}(X,Y) = \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m}\left(x_{0i} - \overline{x}\right)\cdot\left(y_{0j} - \overline{y}\right)\cdot n_{ij}}{N} = \frac{\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} x_{0i} y_{0j} \cdot n_{ij}}{N} - \overline{x_0}\cdot\overline{y_0} =$$

$$= \frac{0 + 0 + 0 + 5 \cdot 1 \cdot 5 + 10 \cdot 1 \cdot 10 + 15 \cdot 1 \cdot 15}{100} - 12 \cdot 0{,}3 = -0{,}1.$$

Теперь ответим на вопросы задачи.

Рассмотрим сначала условие повторной выборки.

Средняя ошибка для X определяется равенством

$$\sigma\left(\overline{X}\right) = \frac{\sigma(X)}{\sqrt{n}}$$

и в нашем случае равна

$$\sigma\left(\overline{X}\right) = \frac{3{,}674}{\sqrt{10}} = 1{,}162.$$

$$\operatorname{Cov}\left(\overline{X},\overline{Y}\right)=\frac{\operatorname{Cov}\left(X,Y\right)}{n}=\frac{-0{,}1}{10}=-0{,}01.$$

Для бесповторной выборки имеем:

$$\sigma\left(\overline{X}\right)=\sqrt{\frac{D\left(X\right)}{n}\cdot\frac{N-n}{N-1}}=\sigma\left(X\right)\cdot\sqrt{\frac{N-n}{n\cdot\left(N-1\right)}}=3{,}674\cdot\sqrt{\frac{100-10}{100-1}}=1{,}108,$$

$$\operatorname{Cov}\left(\overline{X},\overline{Y}\right)=\frac{\operatorname{Cov}\left(X,Y\right)}{n}\cdot\frac{N-n}{N-1}=\frac{-0{,}1}{10}\cdot\frac{100-10}{100-1}=-0{,}009.$$

Ответ: для повторной выборки средняя ошибка выборки для количества жалоб потребителей на качество продаваемой продукции равна 1,162 и ковариация средних $\operatorname{Cov}\left(\overline{X},\overline{Y}\right)=-0{,}01$;

для бесповторной выборки средняя ошибка выборки для количества жалоб потребителей на качество продаваемой продукции равна 1,108 и ковариация средних $\operatorname{Cov}\left(\overline{X},\overline{Y}\right)=-0{,}009$.

Пример 3.5. Каким должен быть минимальный объем выборки из генеральной совокупности в примере 3.1, чтобы средняя ошибка выборки для количества жалоб потребителей на качество продаваемой продукции была не больше 0,8? Рассмотреть условия повторной и бесповторной выборки.

Решение.

Нам известны генеральные характеристики и средняя ошибка выборки $\sigma\left(\overline{X}\right)=0{,}8$. Для ответа на вопрос задачи воспользуемся соответствующими формулами.

В случае повторной выборки

$$n=\left(\frac{\sigma\left(X\right)}{\sigma\left(\overline{X}\right)}\right)^2=\left(\frac{3{,}674}{0{,}8}\right)^2=21{,}091,$$

то есть начиная с $n = 22$, условие на среднюю ошибку выборки выполняется (объем выборки не может быть дробным числом).

В случае бесповторной выборки:

$$n=\frac{D\left(X\right)\cdot N}{N\cdot\left(\sigma\left(\overline{X}\right)\right)^2+D\left(X\right)}=\frac{13{,}5\cdot100}{100\cdot0{,}8^2+13{,}5}=17{,}42,$$

то есть уже начиная с $n = 18$ условие на среднюю ошибку выборки будет выполняться.

Ответ: для средней ошибки, не превосходящей 0,8, минимальный объем повторной выборки равен 22, минимальный объем бесповторной выборки равен 18.

Генеральной долей p значения x_i признака X называется отношение $\frac{N_i}{N}$; ***выборочной долей*** $\hat{p}$ значения x_i признака X называется отношение $\frac{n_i}{n}$, где, как и прежде, N – объем генеральной совокупности, а n – объем выборочной совокупности.

Так как выборочная доля ведет себя как случайная величина, можно вычислять для нее математическое ожидание и дисперсию. Имеют место следующие равенства:

$E\left(\hat{p}\right) = p$ и $D\left(\hat{p}\right) = \frac{p\cdot(1-p)}{n}$ — для повторной выборки;

$E\left(\hat{p}\right) = p$ и $D\left(\hat{p}\right) = \frac{p\cdot(1-p)}{n}\cdot\frac{N-n}{N-1}$ — для бесповторной выборки.

Средней ошибкой выборки называют среднеквадратическую ошибку в равенстве $\hat{p} \approx p$ и обозначают $\sigma\left(\hat{p}\right)$. Понятно, что $\sigma\left(\hat{p}\right) = \sqrt{D\left(\hat{p}\right)}$, поэтому в случае повторного отбора среднюю ошибку выборки вычисляют как:

$$\sigma\left(\hat{p}\right) = \sqrt{D\left(\hat{p}\right)} = \sqrt{\frac{pq}{n}},\ q = 1-p,$$

а в случае бесповторного отбора средняя ошибка выборки определяется по формуле:

$$\sigma\left(\hat{p}\right) = \sqrt{D\left(\hat{p}\right)} = \sqrt{\frac{pq}{n}\cdot\frac{N-n}{N-1}},\ q = 1-p.$$

При решении задач бывает полезно и приближенное равенство:

$$\sigma\left(\hat{p}\right) \approx \sqrt{\frac{pq}{n}\cdot\left(1-\frac{n}{N}\right)},\ q = 1-p.$$

Например, если объем генеральной совокупности равен $N = 1000$, а генеральная доля признака составляет $p = 0{,}25$, то при объеме выборки $n = 50$ средняя ошибка будет оцениваться как

$$\sigma\left(\hat{p}\right) \approx \sqrt{\frac{0{,}25\cdot 0{,}75}{50}\cdot\left(1-\frac{50}{1000}\right)} = 0{,}0597.$$

Задачи для самостоятельного решения

В задачах **3.1–3.10** даны результаты, полученные при изучении признака *X*. Необходимо задать частотное распределение признака и эмпирическую функцию распределения, построить полигон частот и вычислить эмпирические характеристики признака.

3.1. 4; 2; 3; 2; 2; 3; 4; 5; 4; 4; 5; 5; 4; 4; 5.

3.2. 0,2; 0,1; 0,1; 0,3; 0,2; 0,1; 0,2; 0,4; 0,5; 0,3; 0,4; 0,3; 0,5; 0,1; 0,1; 0,2.

3.3. 41; 42; 43; 42; 42; 43; 44; 45; 44; 44; 45; 45; 44; 44; 45.

3.4. 184; 182; 183; 182; 182; 183; 184; 185; 184; 184; 185; 185; 184; 194; 185.

3.5. 2; 1; 1; 3; 2; 1; 2; 4; 5; 3; 4; 3; 5; 1; 1; 2.

3.6. 202; 198; 197; 203; 202; 201; 202; 204; 198; 203; 204; 203; 197; 201; 201; 202.

3.7. 14; 12; 13; 12; 12; 13; 14; 15; 14; 14; 15; 15; 14; 14; 15.

3.8. 22; 18; 17; 23; 22; 21; 22; 24; 18; 23; 24; 23; 17; 21; 21; 22.

3.9. 401; 402; 403; 402; 402; 403; 404; 405; 404; 404; 405; 405; 404; 404; 405.

3.10. 0,3; 0,2; 0,2; 0,3; 0,5; 0,5; 0,5; 0,4; 0,5; 0,3; 0,4; 0,5; 0,5; 0,2; 0,2; 0,3.

В задачах **3.11–3.20** дано интервальное распределение признака *X*. Необходимо построить гистограмму и задать эмпирическую функцию распределения, а также вычислить эмпирические характеристики признака.

3.11.

Значение *X*	12–15	15–18	18–21	21–24	24–27	27–30
Частота	15	12	11	8	5	2

3.12.

Значение *X*	1–5	5–10	10–15	15–20	20–25	25–30
Частота	2	3	7	8	7	3

3.13.

Значение *X*	2–4	4–6	6–8	8–10	10–12	12–14
Частота	10	8	6	8	9	10

3.14.

Значение *X*	18–24	24–30	30–36	36–42	42–48
Частота	24	30	52	58	80

3.15.

Значение *X*	60–70	70–80	80–90	90–100	100–110	60–70
Частота	111	95	47	22	5	111

3.16.

Значение X	7–14	14–21	21–28	28–35	35–42	42–48
Частота	48	34	23	12	7	4

3.17.

Значение X	102–105	105–108	108–201	201–204	204–207	102–105
Частота	24	36	44	40	32	24

3.18.

Значение X	120–150	150–180	180–210	210–240	240–270
Частота	37	42	58	70	94

3.19.

Значение X	4–8	8–12	12–16	16–20	20–24	24–28	28–32
Частота	15	22	19	17	20	17	19

3.20.

Значение X	12–20	20–28	28–36	36–44	44–52	52–60
Частота	22	30	38	32	28	20

В задачах **3.21–3.30** дано совместное статистическое распределение двух признаков X и Y, изученных в генеральной совокупности.

Вычислить эмпирическую ковариацию $\mathrm{Cov}(X,Y)$ и эмпирический коэффициент корреляции $\rho(X,Y)$.

3.21.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-4$	25	11	0	12
$Y=-3$	14	20	0	13
$Y=-2$	0	10	15	0

3.22.

	$X=10$	$X=15$	$X=20$	$X=25$
$Y=5$	15	10	0	0
$Y=10$	20	10	0	20
$Y=15$	0	0	15	10

3.23.

	$X = 11$	$X = 12$	$X = 13$	$X = 14$
$Y = 1$	0	10	0	25
$Y = 2$	20	0	10	30
$Y = 3$	0	15	0	10

3.24.

	$X = 100$	$X = 110$	$X = 120$	$X = 130$
$Y = 10$	11	0	14	40
$Y = 20$	0	12	0	20
$Y = 30$	40	0	18	0

3.25.

	$X = -20$	$X = -5$	$X = 10$	$X = 40$
$Y = -10$	16	0	14	0
$Y = 10$	0	20	0	12
$Y = 20$	18	0	20	0

3.26.

	$X = -2$	$X = -1$	$X = 1$	$X = 4$
$Y = -100$	18	22	0	0
$Y = 150$	0	46	38	12
$Y = 300$	0	34	0	0

3.27.

	$X = 2$	$X = 4$	$X = 8$	$X = 32$
$Y = -1$	11	0	14	40
$Y = 0$	0	12	0	20
$Y = 10$	40	0	18	0

3.28.

	$X = 0$	$X = 5$	$X = 10$	$X = 15$
$Y = 0$	23	0	10	0
$Y = 5$	12	36	21	30
$Y = 10$	0	0	44	14

3.29.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 2$	11	15	0	0
$Y = 4$	12	36	24	20
$Y = 6$	2	0	20	0

3.30.

	$X = -3$	$X = -2$	$X = -1$	$X = 0$
$Y = -4$	15	36	0	0
$Y = -2$	44	0	0	28
$Y = 0$	0	0	22	15

В задачах **3.31–3.40** из заданного совместного статистического распределения двух признаков X и Y, изученных в генеральной совокупности объема N, извлекается выборка заданного объема n. Вычислить ковариацию выборочных средних $\overline{X}$ и $\overline{Y}$. Рассмотреть условия, когда выборка повторная и когда выборка бесповторная.

3.31.

	$X = -2$	$X = -1$	$X = 0$	$X = 1$
$Y = -4$	25	11	0	12
$Y = -3$	14	20	0	13
$Y = -2$	0	10	15	0

$N = 120$, $n = 20$.

3.32.

	$X = 10$	$X = 15$	$X = 20$	$X = 25$
$Y = 5$	15	10	0	0
$Y = 10$	20	10	0	20
$Y = 15$	0	0	15	10

$N = 100$, $n = 15$.

3.33.

	$X = 11$	$X = 12$	$X = 13$	$X = 14$
$Y = 1$	0	10	0	25
$Y = 2$	20	0	10	30
$Y = 3$	0	15	0	10

$N = 110$, $n = 20$.

3.34.

	$X = 100$	$X = 110$	$X = 120$	$X = 130$
$Y = 10$	11	0	14	40
$Y = 20$	0	12	0	20
$Y = 30$	40	0	18	0

$N = 155$, $n = 25$.

3.35.

	$X = -20$	$X = -5$	$X = 10$	$X = 40$
$Y = -10$	16	0	14	0
$Y = 10$	0	20	0	12
$Y = 20$	18	0	20	0

$N = 100$, $n = 10$.

3.36.

	$X = -2$	$X = -1$	$X = 1$	$X = 4$
$Y = -100$	18	22	0	0
$Y = 150$	0	46	38	12
$Y = 300$	0	34	0	0

$N = 170$, $n = 30$.

3.37.

	$X = 2$	$X = 4$	$X = 8$	$X = 32$
$Y = -1$	0	0	14	25
$Y = 0$	0	46	45	30
$Y = 10$	25	0	0	0

$N = 185$, $n = 15$.

338.

	$X = 0$	$X = 5$	$X = 10$	$X = 15$
$Y = 0$	23	0	10	0
$Y = 5$	12	36	21	30
$Y = 10$	0	0	44	14

$N = 190$, $n = 10$.

3.39.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 2$	11	15	0	0
$Y = 4$	12	36	24	20
$Y = 6$	2	0	20	0

$N = 140$, $n = 25$.

3.40.

	$X = -3$	$X = -2$	$X = -1$	$X = 0$
$Y = -4$	15	36	0	0
$Y = -2$	44	0	0	28
$Y = 0$	0	0	22	15

$N = 160$, $n = 20$.

В задачах **3.41–3.50** необходимо использовать определение средней ошибки выборки характеристики признака.

3.41. Значения признака в генеральной совокупности заданы распределением:

Значение X	10	11	12	13	14	15
Частота	15	12	8	8	5	2

Из этой генеральной совокупности извлекается повторная выборка объема 10. Найти среднюю ошибку выборки.

3.42. Значения признака в генеральной совокупности заданы распределением:

Значение X	10	11	12	13	14	15
Частота	15	12	8	8	5	2

Из этой генеральной совокупности извлекается бесповторная выборка объема 10. Найти среднюю ошибку выборки.

3.43. Значения признака в генеральной совокупности заданы распределением:

Значение X	1–5	5–10	10–15	15–20	20–25	25–30
Частота	2	3	7	8	7	3

Из этой генеральной совокупности извлекается повторная выборка объема 5. Найти среднюю ошибку выборки.

3.44. Значения признака в генеральной совокупности заданы распределением:

Значение X	1–5	5–10	10–15	15–20	20–25	25–30
Частота	2	3	7	8	7	3

Из этой генеральной совокупности извлекается бесповторная выборка объема 5. Найти среднюю ошибку выборки.

3.45. Значения признака в генеральной совокупности заданы распределением:

Значение X	10–14	14–18	18–22
Частота	5	9	11

Каким должен быть минимальный объем повторной выборки, чтобы ее средняя ошибка была не больше 4?

3.46. Значения признака в генеральной совокупности заданы распределением:

Значение X	10–14	14–18	18–22
Частота	5	9	11

Каким должен быть минимальный объем бесповторной выборки, чтобы ее средняя ошибка была не больше 4?

3.47. Из генеральной совокупности, имеющей генеральную дисперсию $D(X) = 45$, извлекли повторную выборку объема $n = 15$. Какая при этом получается средняя ошибка выборки $\sigma(\overline{X})$? Каким должен быть минимальный объем повторной выборки, чтобы ее средняя ошибка была не больше 1,5?

3.48. Средняя ошибка повторной выборки объема $n = 15$ из генеральной совокупности оказалась равной 4. Чему равна дисперсия этой генеральной совокупности? Каким должен быть объем повторной выборки из этой генеральной совокупности, чтобы ее ошибка не превосходила 3?

3.49. Из генеральной совокупности объема $N = 100$, имеющей генеральную дисперсию $D(X) = 15$, извлекли бесповторную выборку объема $n = 10$. Какая при этом получается средняя ошибка выборки $\sigma(\overline{X})$? Каким должен быть минимальный объем повторной выборки, чтобы ее средняя ошибка была не больше 1,5?

3.50. Средняя ошибка бесповторной выборки объема $n = 10$ из генеральной совокупности объема $N = 110$ оказалась равной 0,2.

Чему равна дисперсия этой генеральной совокупности? Каким должен быть объем бесповторной выборки из этой генеральной совокупности, чтобы ее ошибка не превосходила 0,3?

3.3. Статистическая оценка параметров распределения

Мы уже отмечали, что одной из задач математической статистики является разработка методов, используя которые, можно по выборочным данным охарактеризовать изучаемую генеральную совокупность. Если генеральное распределение признака зависит от некоторого параметра, то поставленную задачу можно решать, оценивая значение этого параметра. Пусть θ — параметр, от которого зависит генеральное распределение признака X. ***Статистической оценкой*** $\hat{\theta}$ ***параметра*** θ является функция, зависящая от результатов наблюдений за выборочными значениями признака: $\hat{\theta}=\hat{\theta}(X_1, X_2, \ldots, X_n)$ Если статистическая оценка определяется одним числом, то она называется ***точечной***, если же статистическая оценка определяется с помощью некоторого числового промежутка, то в таком случае говорят об ***интервальной*** оценке. Понятно, что статистическая оценка отличается от значения параметра θ генеральной совокупности — ведь ее значение определяется выборочными данными, меняющимися от выборки к выборке. Поэтому актуальным становится вопрос: насколько качественной является полученная статистическая оценка $\hat{\theta}$? Статистические оценки должны удовлетворять определенным требованиям, чтобы стать «хорошими» оценками параметров генеральной совокупности, а именно — быть состоятельными, эффективными и несмещенными.

Оценка $\hat{\theta}$ называется ***состоятельной***, если для любого сколь угодно малого числа $\varepsilon > 0$ выполняется равенство:

$$\lim_{n\to\infty}\left(P\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1$$

Это условие означает сходимость по вероятности полученной оценки к оцениваемому параметру.

Оценка $\hat{\theta}$ называется ***эффективной*** в классе оценок в том случае, когда в этом классе при фиксированном объеме выборки она имеет наименьшую дисперсию.

Оценка $\hat{\theta}$ называется ***несмещенной***, если ее математическое ожидание равно оцениваемому параметру:

$$E(\hat{\theta}) = \theta \cdot$$

Выполнимость этого условия гарантирует отсутствие систематических ошибок при проведении оценки параметра.

При решении задач математической статистики несостоятельные оценки никогда не применяются.

Для состоятельности оценки $\hat{\theta}$ достаточно выполнимости условий

$$\lim_{n\to\infty} D\left(\hat{\theta}\right) = 0\,; \quad E\left(\hat{\theta}\right) = \theta\,.$$

Для несмещенности оценки $\hat{\theta}$ достаточно, чтобы выполнялось только одно условие $\lim\limits_{n\to\infty} D\left(\hat{\theta}\right) = 0$, потому что второе — требование из определения.

Для любой выборки $X_1, X_2, \ldots, X_n$ из генеральной совокупности ***выборочное среднее*** $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ ***является несмещенной и состоятельной оценкой генеральной средней*** $\overline{x_0} = \dfrac{x_{01} + x_{02} + \ldots + x_{0N}}{N}$, то есть выполнены условия:

$$E\left(\overline{X}\right) = \overline{x_0}\,; \; \lim_{n\to\infty} D\left(\overline{X}\right) = 0\,.$$

Если генеральная совокупность распределена по нормальному закону, то выборочное среднее является и эффективной оценкой генеральной средней.

Относительная частота $\dfrac{k}{n}$ наступления события A в серии из n испытаний, проходящих по схеме Бернулли, является эффективной, состоятельной и несмещенной оценкой вероятности $p = P(A)$.

Исправленная выборочная дисперсия

$$S^2 = \frac{n}{n-1} D_в(X)$$

является состоятельной ***несмещенной оценкой генеральной дисперсии***.

При вычислении S^2 используют формулу:

$$S^2 = \frac{n}{n-1} D_в(X) = \frac{n}{n-1} \cdot \frac{\sum_{n=1}^{\infty}\left(X_i - \overline{X}\right)^2}{n} = \frac{\sum_{i=1}^{\infty}\left(X_i - \overline{X}\right)^2}{n-1}.$$

Если ***известно генеральное среднее*** $E(X) = \overline{x_0}$, то несмещенная состоятельная оценка вычисляется по формуле

$$S^2 == \frac{\sum_{i=1}^{\infty}\left(X_i - \overline{x_0}\right)^2}{n}.$$

Пример 3.6. Генеральная совокупность изучается по случайной выборке объема $n = 100$, результаты отражены в таблице:

Значение	$X = 5$	$X = 10$	$X = 15$
Частота	15	30	55

Найти несмещенные оценки генерального среднего и генеральной дисперсии.

Решение.

Несмещенной оценкой генерального среднего является выборочное среднее, а несмещенной оценкой генеральной дисперсии — величина , поэтому вычисляем:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{5 \cdot 15 + 10 \cdot 30 + 15 \cdot 55}{100} = 12,$$

$$S^2 = \frac{n}{n-1} D_в(X) = \frac{100}{99} \cdot \frac{(5-12)^2 \cdot 15 + (10-12)^2 \cdot 30 + (15-12)^2 \cdot 55}{100} =$$

$$= \frac{(5-12)^2 \cdot 15 + (10-12)^2 \cdot 30 + (15-12)^2 \cdot 55}{99} = \frac{1350}{99} = \frac{150}{11} = 13{,}636.$$

Ответ: несмещенная оценка генерального среднего равна 12, несмещенная оценка генеральной дисперсии равна 13,636.

Для нахождения точечных оценок параметров заданного распределения генеральной совокупности используют два основных метода — метод моментов и метод максимального правдоподобия.

Суть ***метода моментов*** для оценивания неизвестных параметров распределения генеральной совокупности состоит в приравнивании теоретических моментов и соответствующих выборочных моментов с последующим решением полученного уравнения или системы уравнений.

Если теоретическое распределение имеет один параметр, то для его оценки методом моментов необходимо решить уравнение:

$$E(X) = \overline{X},$$

приравняв начальные моменты первого порядка (для распределения генеральной совокупности это теоретическое математическое ожидание, а для выборочной совокупности это выборочное среднее).

Если теоретическое распределение имеет два параметра, то для их оценки методом моментов необходимо решить систему уравнений:

$$\begin{cases} E(X) = \overline{X} \\ D(X) = D_в(X) \end{cases},$$

приравняв не только начальные моменты первого порядка, но и центральные моменты второго порядка (для распределения генеральной совокупности это теоретическая дисперсия, а для выборочной совокупности это выборочная дисперсия).

Метод моментов наиболее прост при проведении оценки параметров и получаемые в результате оценки всегда являются состоятельными, этого бывает достаточно для решения многих задач математической статистики.

Пример 3.7. Величина X распределена по показательному закону с плотностью вероятности $f(x) = \begin{cases} 0, & x \le 0 \\ \lambda e^{-\lambda x}, & x > 0 \end{cases}$ и изучается с помощью выборки:

Значение	$X=0$	$X=1$	$X=2$	$X=3$
Частота	115	100	20	15

Найти методом моментов точечную оценку значения λ.

Решение.

Для показательного распределения λ является единственным параметром, поэтому при решении задачи будем использовать уравнение:

$E(X) = \overline{X}$.

Напомним, что математическое ожидание для величины, распределенной по показательному закону, вычисляется как $E(X) = \frac{1}{\lambda}$. Поэтому наше уравнение примет вид:

$\frac{1}{\lambda} = \overline{X}$.

Выборочное среднее будем вычислять, используя данные условия задачи, а в уравнении заменим λ на $\hat{\lambda}$ — оценку, получаемую по выборке:

$$\frac{1}{\hat{\lambda}}=\frac{0\cdot 115+1\cdot 100+2\cdot 20+3\cdot 15}{115+100+20+15} \Rightarrow \frac{1}{\hat{\lambda}}=\frac{185}{250} \Rightarrow \hat{\lambda}=\frac{250}{185}=1{,}351.$$

Ответ: $\hat{\lambda}=1{,}351$.

Пример 3.8. Величина X распределена по биномиальному закону с параметрами n и p и изучается с помощью выборки:

Значение	$X=0$	$X=1$	$X=3$	$X=4$	$X=5$	$X=7$
Частота	2	8	10	10	6	1

Найти методом моментов точечную оценку значений параметров.

Решение.

Биномиальное распределение зависит от двух параметров, поэтому для решения задачи будем использовать систему:

$$\begin{cases} E(X)=\overline{X} \\ D(X)=D_в(X) \end{cases}.$$

Мы знаем, как вычисляются математическое ожидание и дисперсия для величины, распределенной по биномиальному закону:

$$E(X)=np, \qquad D(X)=npq=np(1-p).$$

Поэтому система примет вид:

$$\begin{cases} np=\overline{X} \\ np(1-p)=D_в(X) \end{cases}.$$

По заданной выборке вычислим:

$$\overline{X}=\frac{0\cdot 2+1\cdot 8+3\cdot 10+4\cdot 10+5\cdot 6+7\cdot 1}{2+8+10+10+6+1}=\frac{115}{37};$$

$$D_в(X)=\frac{0^2\cdot 2+1^2\cdot 8+3^2\cdot 10+4^2\cdot 10+5^2\cdot 6+7^2\cdot 1}{37}-\left(\frac{115}{37}\right)^2=$$

$$=\frac{3684}{1369}$$

Подставим полученные значения в систему и переобозначим значения параметров:

$$\begin{cases} \hat{n}\hat{p}=\dfrac{115}{37} \\ \hat{n}\hat{p}(1-\hat{p})=\dfrac{3684}{1369} \end{cases} \Rightarrow \begin{cases} \hat{n}\hat{p}=\dfrac{115}{37} \\ \dfrac{115}{37}\cdot(1-\hat{p})=\dfrac{3684}{1369} \end{cases} \Rightarrow$$

$$\begin{cases} \hat{n}\hat{p} = \dfrac{115}{37} \\ \hat{p} = \dfrac{571}{4255} \end{cases} \Rightarrow \begin{cases} \hat{n} = \dfrac{489325}{21127} \\ \hat{p} = \dfrac{571}{4255} \end{cases}.$$

Параметр n биномиального распределения определяет количество независимых испытаний, проходящих по схеме Бернулли, поэтому не может быть дробным числом. Учтем это и ответим на вопрос задачи:

$$\begin{cases} \hat{n} = \dfrac{489325}{21127} = 23{,}161 \\ \hat{p} = \dfrac{571}{4255} = 0{,}134 \end{cases} \Rightarrow \hat{n} = 24;\ \hat{p} = 0{,}134.$$

Ответ: $\hat{n} = 24$; $\hat{p} = 0{,}134$.

Еще одним методом получения точечных оценок параметров является ***метод максимального правдоподобия***. В основе этого метода лежит функция правдоподобия. Если известна плотность вероятности $f(x,\theta)$ генеральной совокупности, но неизвестен параметр θ этого распределения, то при заданной выборке $X_1, X_2, \ldots, X_n$ функция правдоподобия определяется произведением вида:

$$L(x_1, x_2, \ldots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta).$$

Если закон распределения генеральной совокупности — дискретный, то функция правдоподобия определяется произведением вида:

$$L(x_1, x_2, \ldots, x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \ldots \cdot p(x_n, \theta),$$

где $p(x_i, \theta) = P(X = x_i, \theta)$.

Точечной оценкой $\hat{\theta}$ параметра θ называют значение θ, при котором функция правдоподобия $L(x_1, x_2, \ldots, x_n, \theta)$ достигает максимума. Доказано, что функции $L(x_1, x_2, \ldots, x_n, \theta)$ и $\ln L(x_1, x_2, \ldots, x_n, \theta)$ достигают максимума в одних и тех же точках, поэтому удобнее исследовать на максимум функцию $\ln L(x_1, x_2, \ldots, x_n, \theta)$.

Если распределение генеральной совокупности зависит от нескольких параметров, то функция правдоподобия исследуется на максимум как функция нескольких переменных.

Оценки параметров, полученные на основе метода максимального правдоподобия, являются состоятельными и эффективными, однако трудность решений уравнений правдоподобия и знание закона распределения генеральной совокупности являются основными недостатками этого метода.

Пример 3.9. Величина X распределена по показательному закону с плотностью вероятности $f(x)=\begin{cases}0, & x\le 0\\ \lambda e^{-\lambda x}, & x>0\end{cases}$ и изучается с помощью выборки:

Значение	$X=0$	$X=1$	$X=2$	$X=3$
Частота	115	100	20	15

Найти методом максимального правдоподобия точечную оценку значения параметра λ.

Решение.

Используя выборочные данные, составим функцию правдоподобия, при этом учитываем, что параметром распределения является λ $(\theta=\lambda)$:

$$L(x_1,x_2,...,x_n,\theta)=f(x_1,\theta)\cdot f(x_2,\theta)\cdot ...\cdot f(x_n,\theta)=$$

$$=\left(\lambda e^{-\lambda\cdot 0}\right)^{115}\cdot\left(\lambda e^{-\lambda\cdot 1}\right)^{100}\cdot\left(\lambda e^{-\lambda\cdot 2}\right)^{20}\cdot\left(\lambda e^{-\lambda\cdot 3}\right)^{15}=$$

$$=\lambda^{115+100+20+15}\cdot e^{-\lambda\cdot(0\cdot 115+1\cdot 100+2\cdot 20+3\cdot 15)}=$$

$$=\lambda^{250}\cdot e^{-185\lambda}.$$

Найдем $\ln L(x_1,x_2,...,x_n,\theta)$:

$$\ln L(x_1,x_2,...,x_n,\theta)=\ln\left(\lambda^{250}\cdot e^{-185\lambda}\right)=250\ln\lambda-185\lambda.$$

Исследуем $\ln L(x_1,x_2,...,x_n,\theta)$ на максимум, для этого вычислим производную первого порядка и приравняем ее нулю:

$$\left(\ln L(x_1,x_2,...,x_n,\theta)\right)'_\lambda=\left(250\ln\lambda-185\lambda\right)'_\lambda=\frac{250}{\lambda}-185;$$

$$\frac{250}{\lambda}-185=0\ \Rightarrow\ \lambda=\frac{250}{185}=1{,}351.$$

Используя известное из курса математического анализа достаточное условие экстремума, доказывается, что при $\lambda=1{,}351$ функция $\ln L(x_1,x_2,...,x_n,\theta)$ достигает максимума, а значит, при этом значении достигает максимума и функция правдоподобия $L(x_1,x_2,...,x_n,\theta)$. Полученное значение и является искомой точечной оценкой неизвестного параметра λ.

Ответ: $\hat{\lambda}=1{,}351$.

Точечные оценки параметров не позволяют говорить о том, какую ошибку мы совершаем, заменяя неизвестный параметр θ найденной оценкой $\hat{\theta}$. При малых объемах выборки такие ошибки могут быть весьма существенными. Интервальные оценки параметров позволяют оценить ошибку при такой замене.

Интервальной оценкой параметра θ называется числовой интервал $(\theta_1;\theta_2)$, накрывающий неизвестное значение параметра с заданной вероятностью γ. Границы интервала вычисляются по выборочным данным, поэтому ведут себя как случайные величины.

Интервал $(\theta_1;\theta_2)$ называется ***доверительным***, а соответствующая вероятность γ — ***доверительной вероятностью***. Доверительная вероятность зависит от решаемой задачи и выбирается заранее, чаще всего используется $\gamma = 0{,}95$ или $\gamma = 0{,}99$.

По условию $\gamma = P\left(\left|\theta - \hat{\theta}\right| < \delta\right)$, где δ — точность оценки. Это условие преобразуется к виду: $\gamma = P\left(\hat{\theta} - \delta < \theta < \theta + \delta\right)$ и при известном законе распределения генеральной совокупности позволяет формировать доверительные интервалы для оцениваемых параметров.

Напомним, что квантилем порядка (уровня) называется такое значение q_p абсолютно непрерывной случайной величины, что

$$\int_{-\infty}^{q_p} f(x)dx = p .$$

Геометрически это условие можно изобразить так (рис. 3.4):

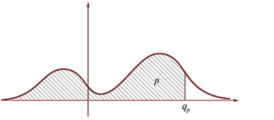


Рис. 3.4. Квантиль порядка

С понятием квантиля тесно связано понятие процентной точки. Например, 5%-ная точка является квантилем порядка 0,95.

Если q_p — квантиль распределения порядка p, то он также называется $100\cdot(1-p)$ -процентной точкой распределения. Если, например, $p=0{,}95$, то q_p — 5%-ная точка распределения.

Кроме нормального распределения генеральной совокупности с плотностью вероятности $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$ в математической статистике используют и другие законы распределения.

Рассмотрим распределение некоторых случайных величин, представляющих собой функции нормальных случайных величин, используемые в математической статистике.

Распределением χ^2 (***хи-квадрат***) с $k=n$ степенями свободы называется распределение суммы квадратов n независимых случайных величин, распределенных по стандартному нормальному закону

$$Z=Y_1^2+Y_2^2+\ldots+Y_n^2,$$

$$Y_i \sim N(0;1), \quad Z \sim \chi^2(n).$$

График плотности вероятности распределения имеет вид (рис. 3.5):

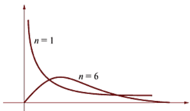


Рис. 3.5. Плотность вероятности распределения χ^2

Процентные точки распределения χ^2 табулированы.

Отметим, что $\chi^2_\alpha(k)\approx\dfrac{\left(Z_\alpha+\sqrt{2k-1}\right)^2}{2}$ при $k>30$.

Распределением Стьюдента, или t-распределением с k степенями свободы называется распределение случайной величины

$$Z = \frac{X}{\sqrt{\frac{Y}{k}}},$$

где X и Y независимы, $X \sim N(0;1)$, $Y \sim \chi^2(k)$. Распределение Стьюдента с k степенями свободы принято обозначать $Z \sim t(k)$.

График плотности вероятности распределения Стьюдента имеет вид (рис. 3.6):

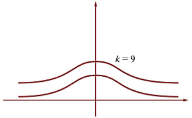


Рис. 3.6. Плотность вероятности распределения Стьюдента

Кривая распределения Стьюдента по сравнению с кривой нормального распределения является более пологой.

Числовые характеристики распределения Стьюдента определяются как:

$$E(t) = 0; \quad D(t) = \frac{k}{k-2} \quad \text{при } k > 2.$$

При $k \to \infty$ t — распределение приближается к стандартному нормальному. При $k > 30$ распределение Стьюдента и стандартное нормальное распределение уже практически неразличимы.

Процентные точки t-распределения табулированы.

Отметим, что $t_\alpha(k) \approx Z_\alpha$ при $k > 30$; $t_{1-\alpha}(k) = -t_\alpha(k)$.

Распределением Фишера или F — распределением с k_1, k_2 степенями свободы называется распределение случайной величины

$$Z = \frac{\frac{X}{k_1}}{\frac{Y}{k_2}},$$

где X и Y — независимые случайные величины, $X \sim \chi^2(k_1)$, $Y \sim$ $\sim \chi^2(k_2)$.

График плотности вероятности распределения Фишера имеет вид (рис. 3.7):

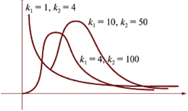


Рис. 3.7. Плотность вероятности распределения Фишера

С ростом числа степеней свободы распределение Фишера приближается к нормальному распределению.

Процентные точки распределения Фишера табулированы. Отметим, что

$$F_{1-\alpha}(k_1, k_2) = \frac{1}{F_\alpha(k_1, k_2)}.$$

Доверительный интервал для оценки неизвестного математического ожидания нормально распределенной генеральной совокупности ***при известной генеральной дисперсии*** $D(X) = \sigma^2$ имеет вид:

$$\overline{X} - z_\gamma \cdot \frac{\sigma}{\sqrt{n}} < E(X) < \overline{X} + z_\gamma \cdot \frac{\sigma}{\sqrt{n}},$$

где

$\overline{X}$ — выборочное среднее,

n — объем выборки,

σ — известное стандартное отклонение генеральной совокупности,

$\Phi(z_\gamma)=\dfrac{\gamma}{2}$, $\Phi(z)$ — функция Лапласа, значения которой табулированы.

Пример 3.10. Найти доверительный интервал для неизвестного математического ожидания нормально распределенной генеральной совокупности с доверительной вероятностью 0,95, если дисперсия генеральной совокупности равна 9, а выборочное среднее равно 1,1 при объеме выборки, равном 100.

Решение.

Так как генеральная совокупность распределена по нормальному закону и известна ее дисперсия $D(X)=\sigma^2=9$, то для формирования искомого доверительного интервала воспользуемся формулой:

$$\overline{X}-z_\gamma\cdot\frac{\sigma}{\sqrt{n}}<E(X)<\overline{X}+z_\gamma\cdot\frac{\sigma}{\sqrt{n}},$$

где $\overline{X}=1{,}1$, $n=100$, $\sigma=\sqrt{D(X)}=3$, $\Phi(z_\gamma)=\dfrac{\gamma}{2}=\dfrac{0{,}95}{2}=0{,}475 \Rightarrow$

$z_\gamma=1{,}96$, и искомый доверительный интервал определяется как:

$$1{,}1-1{,}96\cdot\frac{3}{\sqrt{100}}<E(X)<1{,}1+1{,}96\cdot\frac{3}{\sqrt{100}},$$

$$1{,}1-0{,}588<E(X)<1{,}1+0{,}588,$$

$$0{,}512<E(X)<1{,}688.$$

Ответ: $(0{,}512;1{,}688)$ — искомый доверительный интервал.

В рассмотренном примере доверительная вероятность $\gamma=0{,}95$ позволяет утверждать, что $0{,}95=P\left(\left|E(X)-\overline{X}\right|<0{,}588\right)$, где $0{,}588$ — точность оценки.

Пример 3.11. Каким должен быть минимальный объем выборки из нормально распределенной генеральной совокупности с дисперсией, равной 9, чтобы с доверительной вероятностью 0,95 можно было утверждать, что выборочное среднее, равное 1,1, отклонялось от генерального среднего не более чем на 0,1?

Решение.

Генеральная совокупность распределена по нормальному закону с известной дисперсией. Поэтому используем условия

$$0{,}95=P\left(\left|E(X)-\overline{X}\right|<\delta\right) \Rightarrow 0{,}95=P\left(\overline{X}-\delta<E(X)<\overline{X}+\delta\right),$$

$$\overline{X}-z_\gamma\cdot\frac{\sigma}{\sqrt{n}}<E(X)<\overline{X}+z_\gamma\cdot\frac{\sigma}{\sqrt{n}}.$$

Отсюда определим точность оценки (отклонение) δ и необходимый нам минимальный объем выборки:

$$\delta = z_\gamma \cdot \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n = \left(\frac{z_\gamma \cdot \sigma}{\delta}\right)^2.$$

Учитывая данные условия и ***примера 3.10***, получим:

$$n = \left(\frac{z_\gamma \cdot \sigma}{\delta}\right)^2 = \left(\frac{1{,}96 \cdot 3}{0{,}1}\right)^2 = 3457{,}44.$$

Так как объем выборки не может быть дробным числом, в ответ запишем $n = 3455$.

Ответ: минимальный объем выборки, соответствующий заданному отклонению 0,1 и доверительной вероятности 0,95, равен $n = 3455$.

Доверительный интервал для оценки неизвестного математического ожидания нормально распределенной генеральной совокупности ***при неизвестной генеральной дисперсии*** имеет вид:

$$\overline{X} - t_\gamma \cdot \frac{S}{\sqrt{n}} < E(X) < \overline{X} + t_\gamma \cdot \frac{S}{\sqrt{n}},$$

где

$\overline{X}$ — выборочное среднее,

n — объем выборки,

$S = \sqrt{S^2}$, $S^2 = \dfrac{n}{n-1} D_в(X)$ — несмещенная оценка дисперсии генеральной совокупности («исправленная» дисперсия),

t_γ — $\dfrac{\alpha}{2} \cdot 100\%$ -ная точка распределения Стьюдента с $(n-1)$ степенями свободы (рис. 3.8).

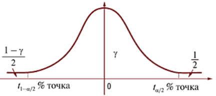


Рис. 3.8. Доверительный интервал для оценки неизвестного математического ожидания

При формировании доверительного интервала часто удобнее использовать табл. 3.1 значений t_γ, которые находят по заданным значениям n и γ.

Таблица 3.1

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	4,14	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				

Пример 3.12. Найти доверительный интервал для неизвестного математического ожидания нормально распределенной генеральной совокупности с доверительной вероятностью 0,99, если выборочное среднее и «исправленная» дисперсия равны 1,1 и 2,56 соответственно при объеме выборки 15.

Решение.

Так как генеральная совокупность распределена по нормальному закону, но генеральная дисперсия неизвестна, применим формулу:

$$\overline{X} - t_\gamma \cdot \frac{S}{\sqrt{n}} < E(X) < \overline{X} + t_\gamma \cdot \frac{S}{\sqrt{n}},$$

которая с учетом данных условия задачи и значения $t_\gamma = 2{,}98$, взятого из табл. 3.1 при $n = 15$ и $\gamma = 0{,}99$, примет вид:

$$1{,}1 - 2{,}98 \cdot \frac{1{,}6}{\sqrt{15}} < E(X) < 1{,}1 + 2{,}98 \cdot \frac{1{,}6}{\sqrt{15}}; \quad 1{,}1 - 1{,}231 < E(X) < 1{,}1 + 1{,}231;$$

$$-0{,}131 < E(X) < 2{,}331.$$

Ответ: искомый доверительный интервал $(-0{,}131; 2{,}331)$.

При формировании доверительного интервала для дисперсии генеральной совокупности при ***известном генеральном среднем*** $\overline{x_0} = E(X)$ используется несмещенная точечная оценка дисперсии $S^2 = \dfrac{\sum\limits_{i=1}^{\infty}\left(X_i - \overline{x_0}\right)^2}{n}$.

Случайная величина $\dfrac{n \cdot S^2}{\sigma^2}$ имеет распределение χ^2 с $k = n$ степенями свободы.

При заданной доверительной вероятности γ имеем

$$\chi^2_{1-\alpha/2}(n) < \frac{nS^2}{\sigma^2} < \chi^2_{\alpha/2}(n) \Rightarrow$$

$$\Rightarrow \quad \frac{nS^2}{\chi^2_{\alpha/2}(n)} < \sigma^2 < \frac{nS^2}{\chi^2_{1-\alpha/2}(n)},$$

искомый доверительный интервал для генеральной дисперсии $D(X) = \sigma^2$ (рис. 3.9).

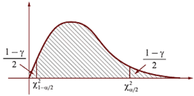


Рис. 3.9. Доверительный интервал для оценки неизвестной генеральной дисперсии

При формировании доверительного интервала для дисперсии генеральной совокупности при неизвестном генеральном среднем $\overline{x_0} = E(X)$ используется несмещенная точечная оценка дисперсии

$$S^2 = \frac{n}{n-1} D_в(X) = \frac{\sum\limits_{i=1}^{\infty}\left(X_i - \overline{X}\right)^2}{n-1}.$$

Случайная величина

$$\frac{(n-1)\cdot S^2}{\sigma^2}$$

имеет распределение χ^2 с $k=n-1$ степенями свободы. В этом случае искомый доверительный интервал имеет вид:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}.$$

При формировании доверительного интервала для среднего квадратического (стандартного) отклонения $\sigma=\sqrt{\sigma^2}=\sqrt{D(X)}$ генеральной совокупности часто используют более удобные формулы (без непосредственного привлечения распределения χ^2):

$$s\left(1-q\right)<\sigma<s\left(1+q\right),\ \text{если}\ q<1$$

или

$$0<\sigma<s\left(1+q\right),\ \text{если}\ q>1,$$

где значение q находят по заданным значениям n и γ из табл. 3.2:

Таблица 3.2

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	1,06	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	0,211
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

Например, если из нормально распределенной совокупности извлечена выборка объема $n = 25$ и «исправленная» дисперсия равна 4, то доверительный интервал для стандартного отклонения σ генеральной совокупности с доверительной вероятностью $\gamma = 0{,}95$ будет определяться как

$$\sqrt{4}\cdot(1-0{,}32)<\sigma<\sqrt{4}\cdot(1+0{,}32) \quad\Rightarrow\quad 1{,}36<\sigma<2{,}64\,.$$

Интервал $(1{,}36;\,2{,}64)$ накрывает стандартное отклонение генеральной совокупности с доверительной вероятностью 0,95.

Доверительный интервал для неизвестной вероятности p биномиального распределения формируется на основе относительной частоты $\omega = \dfrac{m}{n}$ и имеет вид:

$$\omega - z_\gamma\cdot\sqrt{\frac{\omega(1-\omega)}{n}}<p<\omega+z_\gamma\cdot\sqrt{\frac{\omega(1-\omega)}{n}}\,,$$

где

n — объем выборки,

$\Phi(z_\gamma)=\dfrac{\gamma}{2}$, $\Phi(z)$ — функция Лапласа, значения которой табулированы.

Например, доверительный интервал для неизвестной вероятности при относительной частоте $\omega = 0{,}07$ и объеме выборки $n = 100$ будет вычисляться как

$$0{,}07-1{,}96\cdot\sqrt{\frac{0{,}07\cdot(1-0{,}07)}{100}}<p<0{,}07+1{,}96\cdot\sqrt{\frac{0{,}07\cdot(1-0{,}07)}{100}}$$

с доверительной вероятностью 0,95 и будет равен $(0{,}02;\,0{,}12)$.

Аналогично можно построить доверительный интервал для неизвестной доли признака в генеральной совокупности W по известной выборочной доле признака $w = \dfrac{m}{n}$ в случае повторной выборки:

$$w-z_\gamma\cdot\sqrt{\frac{w(1-w)}{n}}<W<w+z_\gamma\cdot\sqrt{\frac{w(1-w)}{n}}\,,$$

где

n — объем выборки, $\Phi(z_\gamma)=\dfrac{\gamma}{2}$, $\Phi(z)$ — функция Лапласа.

Для бесповторной выборки доверительный интервал определяется условием

$$w - z_\gamma \cdot \sqrt{\frac{w(1-w)}{n} \cdot \frac{N-n}{N-1}} < W < w + z_\gamma \cdot \sqrt{\frac{w(1-w)}{n} \cdot \frac{N-n}{N-1}},$$

где N — объем генеральной совокупности.

При интервальной оценке неизвестной вероятности (или доли признака в генеральной совокупности в случае повторной выборки) точность оценки определяется значением

$$\delta = z_\gamma \cdot \sqrt{\frac{\omega(1-\omega)}{n}},$$

где $\omega = \frac{m}{n}$ и $\Phi(z_\gamma) = \frac{\gamma}{2}$.

Если использовать заданные доверительную вероятность γ и точность оценки δ, то можно определять минимально необходимый объем выборки по формуле: $n = \frac{z_\gamma^2}{4\delta^2}$. Например, чтобы относительная частота отклонялась от вероятности не более чем на 0,001 при доверительной вероятности $\gamma = 0{,}95$, минимально необходимый объем выборки должен составить $\frac{1{,}96^2}{4 \cdot 0{,}001^2} = 960400$, тогда как отклонение на 0,01 при той же доверительной вероятности предполагает объем, равный $\frac{1{,}96^2}{4 \cdot 0{,}01^2} = 9604$.

Задачи для самостоятельного решения

В задачах **3.51–3.60** генеральная совокупность изучается по заданной выборке. Используя выборочные данные, найти несмещенные оценки генерального среднего и генеральной дисперсии.

3.51.

Значение X	5	10	15	20	25	30
Частота	5	7	7	8	7	5

3.52.

Значение X	3	4	8	11	15	17
Частота	11	9	8	8	9	8

3.53.

Значение X	0	1	2	4	5	6
Частота	124	98	72	15	12	8

3.54.

Значение X	10	12	15	18	20	22
Частота	38	45	47	48	34	30

3.55.

Значение X	−11	−7	−4	−3	−2	−1
Частота	12	28	34	49	55	72

3.56.

Значение X	−15	−10	−8	−5	−2	0
Частота	12	35	44	27	16	8

3.57.

Значение X	11	14	15	18	22	25
Частота	42	37	41	39	38	40

3.58.

Значение X	1	2	3	4	5	6
Частота	15	18	17	17	20	16

3.59.

Значение X	0	1	2	3	4	5
Частота	155	120	45	32	18	5

3.60.

Значение X	12	22	26	28	31	35
Частота	43	45	39	32	40	42

В задачах **3.61–3.70** найти точечные оценки параметров распределения, используя метод моментов.

3.61. Случайная величина X распределена по биномиальному закону с параметрами n и p и изучается с помощью выборочной совокупности:

Значение X	0	1	3	4	5	6
Частота	11	18	22	23	16	8

Используя метод моментов и выборочные данные, найти точечные оценки параметров.

3.62. Случайная величина X распределена по биномиальному закону с параметрами n и p и изучается с помощью выборочной совокупности:

Значение X	10–15	15–20	20–25	25–30	30–35	35–40	40–45
Частота	4	7	24	33	38	40	36

Используя метод моментов и выборочные данные, найти точечные оценки параметров.

3.63. Случайная величина X распределена равномерно на отрезке $[a;b]$ и изучается с помощью выборочной совокупности:

Значение X	3	4	6	7	8	10	11
Частота	34	37	28	33	30	28	36

Используя метод моментов и выборочные данные, найти точечные оценки параметров a и b равномерного распределения.

3.64. Случайная величина X распределена равномерно на отрезке $[a;b]$ и изучается с помощью выборочной совокупности:

Значение X	3–8	8–13	13–18	18–23	23–28	28–33	33–38
Частота	4	7	24	33	38	40	36

Используя метод моментов и выборочные данные, найти точечные оценки параметров a и b равномерного распределения.

3.65. Случайная величина X распределена по закону Пуассона с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0	1	2	3	4	5
Частота	112	89	63	24	12	5

Используя метод моментов и выборочные данные, найти точечную оценку параметра λ.

3.66. Случайная величина X распределена по закону Пуассона с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0	1	2	3	4	5
Частота	15	12	8	5	3	1

Используя метод моментов и выборочные данные, найти точечную оценку параметра λ.

3.67. Случайная величина X распределена по геометрическому закону с параметром p и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6
Частота	12	11	9	8	5	2

Используя метод моментов и выборочные данные, найти точечную оценку параметра p.

3.68. Случайная величина X распределена по геометрическому закону с параметром p и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6	7
Частота	15	10	10	9	9	7	6

Используя метод моментов и выборочные данные, найти точечную оценку параметра p.

3.69. Случайная величина X распределена по показательному закону с параметром λ и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6
Частота	125	100	94	82	61	48

Используя метод моментов и выборочные данные, найти точечную оценку параметра λ.

3.70. Случайная величина X распределена по показательному закону с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0–2	2–4	4–6	6–8	8–10	10–12
Частота	140	112	95	79	65	32

Используя метод моментов и выборочные данные, найти точечную оценку параметра λ.

В задачах **3.71—3.80** найти точечные оценки параметров распределения, используя метод максимального правдоподобия.

3.71. Случайная величина X распределена по закону Пуассона с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0	1	2	3	4	5
Частота	112	89	63	24	12	5

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра λ.

3.72. Случайная величина X распределена по закону Пуассона с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0	1	2	3	4	5
Частота	15	12	8	5	3	1

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра λ.

3.73. Случайная величина X распределена по показательному закону с параметром λ и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6
Частота	125	100	94	82	61	48

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра λ.

3.74. Случайная величина X распределена по показательному закону с параметром λ и изучается с помощью выборочной совокупности:

Значение X	0–2	2–4	4–6	6–8	8–10	10–12
Частота	140	112	95	79	65	32

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра λ.

3.75. Случайная величина X распределена по геометрическому закону с параметром p и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6
Частота	12	11	9	8	5	2

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра p.

3.76. Случайная величина X распределена по геометрическому закону с параметром p и изучается с помощью выборочной совокупности:

Значение X	1	2	3	4	5	6	7
Частота	15	10	10	9	9	7	6

Используя метод максимального правдоподобия и выборочные данные, найти точечную оценку параметра p.

3.77. Случайная величина X распределена по биномиальному закону с параметрами n и p и изучается с помощью выборочной совокупности:

Значение X	0	1	3	4	5	6
Частота	11	18	22	23	16	8

Используя метод максимального правдоподобия и выборочные данные, найти точечные оценки параметров n и p.

3.78. Случайная величина X распределена по биномиальному закону с параметрами n и p и изучается с помощью выборочной совокупности:

Значение X	10–15	15–20	20–25	25–30	30–35	35–40	40–45
Частота	4	7	24	33	38	40	36

Используя метод максимального правдоподобия и выборочные данные, найти точечные оценки параметров n и p.

3.79. Случайная величина X распределена по нормальному закону с параметрами μ и σ и изучается с помощью выборочной совокупности:

Значение X	0	1	3	4	5	6
Частота	45	49	48	33	25	16

Используя метод максимального правдоподобия и выборочные данные, найти точечные оценки параметров μ и σ.

3.80. Случайная величина X распределена по нормальному закону с параметрами μ и σ и изучается с помощью выборочной совокупности:

Значение X	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Частота	14	17	24	30	28	19	15

Используя метод максимального правдоподобия и выборочные данные, найти точечные оценки параметров μ и σ.

В задачах **3.81—3.90** найти искомые доверительные интервалы для оценки указанных параметров распределения с заданной надежностью γ или, используя интервальную оценку, ответить на вопрос о минимальном объеме выборки.

3.81. Генеральная совокупность распределена по нормальному закону с дисперсией $D(X)=4$. Определить доверительный интервал для неизвестного математического ожидания с доверительной вероятностью $\gamma = 0{,}99$ по выборочному среднему $\overline{X} = 2{,}1$ при объеме выборки $n = 25$.

3.82. Каким должен быть минимальный объем выборки из нормально распределенной генеральной совокупности с дисперсией $D(X) = 4$, чтобы с доверительной вероятностью $\gamma = 0{,}95$ можно было утверждать, что выборочное среднее $\overline{X} = 2{,}1$ отклонится от генерального среднего не более чем на 0,2?

3.83. Средний стаж X работников на предприятиях некоторой отрасли подчиняется нормальному закону, дисперсия для которого известна и равна $D(X) = 25$. Случайная выборка объема $n = 80$ показала, что средний стаж работников, попавших в выборку, равен 16 годам. Указать доверительный интервал для среднего стажа работников в целом по отрасли с доверительной вероятностью $\gamma = 0{,}94$.

3.84. Каким должен быть минимальный объем выборки из распределения X — среднего стажа работников на предприятиях некоторой отрасли, подчиняющегося нормальному закону с известной дисперсией $D(X) = 25$, чтобы с доверительной вероятностью $\gamma = 0{,}95$ можно было утверждать, что выборочное среднее $\overline{X} = 16$ отклонится от генерального среднего не более чем на 1 год?

3.85. Для определения средней урожайности X зерновой культуры с площади 500 гектаров выбрали 30 участков, каждый площадью 0,5 гектара. Выборочная средняя урожайность составила 25 центнеров с гектара. Определить доверительный интервал при $\gamma = 0{,}9$ для средней урожайности по всей площади, если дисперсия средней урожайности по всей площади известна и равна $D(X) = 16$.

3.86. Каким должно быть минимальное количество выбираемых участков площадью 0,5 гектара каждый, чтобы с доверительной вероятностью $\gamma = 0{,}95$ можно было утверждать, что выборочная средняя урожайность 25 центнеров с гектара отклонится от средней урожайности по всей площади (500 гектаров) не более чем на 4 центнера с гектара? Дисперсия средней урожайности по всей площади равна $D(X) = 16$.

3.87. Определить доверительный интервал при $\gamma = 0{,}99$ для неизвестной вероятности возврата в срок кредита определенного вида, если из 1500 взятых кредитов в срок были возвращены 1470.

3.88. Какой минимальный объем выборки необходим, чтобы относительная частота возврата кредита отклонялась от вероятности возврата кредита не более чем на 0,008 при доверительной вероятности $\gamma = 0{,}95$?

3.89. Определить доверительный интервал при $\gamma = 0{,}92$ для неизвестной доли признака в генеральной совокупности, если доля признака в выборочной совокупности равна 0,7 при $n = 800$.

3.90. Какой минимальный объем выборки необходим, чтобы доля признака в генеральной совокупности отклонялась от выборочной доли признака не более чем на 0,05 при доверительной вероятности $\gamma = 0{,}95$?

3.4. Статистическая проверка гипотез

Мы уже не раз отмечали, что важнейшей задачей математической статистики является изучение генеральной совокупности по выборочным данным. Предположения о виде функции распределения генеральной совокупности или о параметре этой функции распределения принято называть ***статистическими гипотезами***.

Если статистическая гипотеза содержит предположение о виде функции распределения генеральной совокупности, то она называется ***непараметрической***, если же статистическая гипотеза содержит предположение о значении параметра, то она называется ***параметрической***.

Гипотеза, однозначно фиксирующая вид функции распределения или значение параметра, является ***простой***, в противном случае статистическая гипотеза определяется как ***сложная***.

В процессе решения задачи должны быть выдвинуты две гипотезы — основная H_0 и альтернативная H_1. Альтернативная гипотеза является конкурирующей с основной гипотезой таким образом, что если принимается основная гипотеза, то отвергается альтернативная, и, наоборот, альтернативная гипотеза отвергается в пользу основной.

Понятно, что должны существовать обоснованные правила, позволяющие статистически значимо подтвердить или опровергнуть основную гипотезу. Такие правила определяются как ***статистические***

критерии, они используют случайные величины, зависящие от значений выборки, называемые ***статистиками***. Значение такой величины, вычисленное для определенной выборки, называют ***наблюдаемым значением статистики***.

Множество всех возможных значений статистики можно разбить на две области — область принятия основной гипотезы (для этих значений H_0 выполняется) и ***критическую область*** (для этих значений H_0 не выполняется). Значения статистики, отделяющие область принятия H_0 от критической области, называют ***критическими точками***.

Так как в основе рассуждений лежит применение случайных величин, то на этапе вывода можно совершить ошибку. При статистической проверке гипотез возможны ошибки первого и второго рода.

Ошибка первого рода совершается в том случае, когда отвергается изначально верная основная гипотеза H_0, если же изначально неверная основная гипотеза принимается как верная, то говорят, что совершена ***ошибка второго рода***.

Вероятность допустить ошибку первого рода называют ***уровнем значимости критерия*** и обозначают α.

Вероятность допустить ошибку второго рода обозначают β. Число $1-\beta$ определяет ***мощность критерия***. Мощность критерия определяет вероятность того, что ошибка второго рода не будет совершена.

Желание минимизировать значения α и β при проверке статистических гипотез естественно, однако при фиксированном объеме выборки с уменьшением значения α значение β растет и наоборот, в этом случае обычно минимизируют уровень значимости α. Уровень доверия при фиксированном α определяется значением $\gamma = 1-\alpha$.

При одном и том же уровне значимости можно использовать разные критерии. Оптимальным является тот, которому соответствует меньшее значение β, то есть выбирается критерий наибольшей мощности.

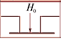
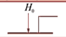
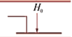
Задачи выбора критерия уже решены как самостоятельные задачи математической статистики, поэтому при проверке статистических гипотез мы будем использовать только те критерии, которым отвечает наибольшая мощность.

Выбранный критерий при заданном уровне значимости позволяет определять критические точки и сформировать критическую область, а, следовательно, сделать вывод о подтверждении или отклонении основной гипотезы.

При проверке статистических гипотез мы будем придерживаться определенного плана, который включает следующие этапы:

1) выдвижение основной гипотезы H_0 и альтернативной гипотезы H_1;
2) выбор уровня значимости α для проведения проверки;
3) выбор статистики (функции, значения которой зависят от выборки) и вычисление наблюдаемого значения статистики;
4) формирование критической области;
5) ответ на вопрос о принадлежности наблюдаемого значения статистики критической области;
6) вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной.

Формирование критической области напрямую зависит от того, как выдвигается альтернативная гипотеза. Для большинства выбираемых статистик критическая область может быть как двусторонней, так и односторонней.

Знак отношения, используемого для H_0	$=$	$\leq$	$\geq$
Знак отношения, используемого для H_1	$\neq$	$>$	$<$
Критическая область	двусторонняя	правосторонняя	левосторонняя
			

Правила, по которым вычисляются критические значения, зависят от выбранной статистики. Это могут быть *Z*-статистики, *T*-статистики, *F*-статистики и статистики χ^2.

Z-статистика имеет нормальный закон распределения, *T*-статистика имеет распределение Стьюдента, *F*-статистика имеет распределение Фишера, статистика χ^2 имеет распределение Пирсона (распределение χ^2).

Выбор статистики является важнейшей составляющей решения и диктуется поставленной задачей. Мы рассмотрим основные задачи, решаемые с помощью этих статистик.

В тех случаях, когда генеральная совокупность имеет нормальное распределение с известной дисперсией (в случае двух генеральных совокупностей это условие должно быть выполнено для каждой из них), при проверке статистических гипотез применяется Z-статистика.

Кроме того, Z-статистика применима и в тех случаях, когда изучают теоретическую вероятность какого-либо события. В таблице 3.3 перечислены основные задачи, при решении которых применяется Z-статистика.

Таблица 3.3

Задачи на проверку гипотез с использованием Z- статистики

Задача	Вид Z-статистики	
Сравнение выборочной средней с генеральной средней нормальной совокупности при известной дисперсии $D(X)=\sigma^2$	$Z=\dfrac{\overline{X}-\overline{x_0}}{\sqrt{\dfrac{D(X)}{n}}}$	$\overline{X}$ — выборочное среднее, $\overline{x_0}$ — генеральное среднее, задано по условию задачи, $D(X)$ — генеральная дисперсия, задана по условию задачи? n — объем выборки
Сравнение наблюдаемой относительной частоты с теоретической вероятностью события	$Z=\dfrac{\omega-p}{\sqrt{\dfrac{pq}{n}}}$	$\omega=\dfrac{m}{n}$ — наблюдаемая относительная частота, p — теоретическая вероятность, задана по условию, $q=1-p$, — объем выборки
Сравнение двух средних нормальных совокупностей при известных дисперсиях $D(X)=\sigma_X^2$ и $D(Y)=\sigma_Y^2$	$Z=\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{D(X)}{n}+\dfrac{D(Y)}{m}}}$	$\overline{X}$ — выборочное среднее из первой ГС, $\overline{Y}$ — выборочное среднее из второй ГС, $D(X)$ — дисперсия для первой ГС, задана по условию задачи

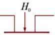
Окончание таблицы

Задача	Вид Z-статистики	
		n — объем выборки из первой ГС, $D(Y)$ — дисперсия для второй ГС, задана по условию задачи, m — объем выборки из второй ГС
Сравнение двух вероятностей биномиальных распределений генеральных совокупностей	$Z = \dfrac{\overline{\omega_1} - \overline{\omega_2}}{\sqrt{\omega(1-\omega)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $\omega = \dfrac{m_1 + m_2}{n_1 + n_2}$	$\overline{\omega_1} = \dfrac{m_1}{n_1}$ — наблюдаемая относительная частота для первой ГС, n_1 — объем выборки из первой ГС, $\overline{\omega_2}$ — наблюдаемая относительная частота для второй ГС, n_2 — объем выборки из второй ГС

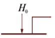
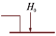
Если решение задачи предполагает использование Z-статистики, то для принятия или отклонения основной гипотезы H_0 необходимо формирование критической области, а для этого надо вычислить критическое значение, используя функцию Лапласа, выбранный уровень значимости α и определенные правила (табл. 3.4).

Таблица 3.4

Правила формирования критических областей при использовании Z-статистики

Двусторонняя критическая область		По таблице значений функции Лапласа из условия: $\Phi(z_{\text{крит}}) = \dfrac{1-\alpha}{2}$, $z_{\text{левое}} = -z_{\text{крит}}$, $z_{\text{правое}} = z_{\text{крит}}$

Окончание таблицы

Правосторонняя критическая область		По таблице значений функции Лапласа из условия: $\Phi(z_{\text{крит}}) = \dfrac{1-2\alpha}{2}$, $z_{\text{правое}} = z_{\text{крит}}$
Левосторонняя критическая область		По таблице значений функции Лапласа из условия: $\Phi(z_{\text{крит}}) = \dfrac{1-2\alpha}{2}$, $z_{\text{левое}} = -z_{\text{крит}}$

Пример 3.13. Для проверки эффективности новой технологии отобраны две группы рабочих: в первой группе численностью $n = 50$ человек, где применялась новая технология, выборочная средняя выработка составила $\overline{X} = 85$ (изделий), во второй группе численностью $m = 70$ человек, где технология осталась прежней, выборочная средняя составила $\overline{Y} = 78$ (изделий). Генеральные дисперсии выработки известны и равны $D(X) = 100$ и $D(Y) = 74$ соответственно. На уровне значимости $\alpha = 0{,}05$ выяснить, влияет ли новая технология на среднюю производительность, считая генеральные совокупности нормально распределенными.

Решение.

При решении задачи будем придерживаться нашего плана/

1. Выдвижение основной гипотезы H_0 и альтернативной гипотезы H_1.

Основная гипотеза должна состоять в том, что средние выработки по старой и новой технологиям равны, альтернативную гипотезу можно выбирать направленной, а именно — выработка при новой технологии больше:

$H_0:\ \overline{x_0} = \overline{y_0}$,

$H_1:\ \overline{x_0} > \overline{y_0}$.

2. Выбор уровня значимости α для проведения проверки.

Уровень значимости нам задан по условию задачи: $\alpha = 0{,}05$.

3. Выбор статистики и вычисление наблюдаемого значения статистики.

Мы находимся в рамках стандартной задачи сравнения двух средних нормальных совокупностей при известных дисперсиях $D(X) = 100$ и $D(Y) = 74$, поэтому выбираем статистику

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}},$$

для которой вычислим наблюдаемое значение, используя данные условия задачи:

$$Z_{\text{наблюдаемое}} = \frac{85 - 78}{\sqrt{\frac{100}{50} + \frac{74}{70}}} = 4{,}0035.$$

4. Формирование критической области.

Критическая область (рис. 3.10) будет формироваться как правосторонняя (определяется альтернативной гипотезой $H_1 : \overline{x_0} \quad \overline{y_0}$), а при вычислении критического значения используем соответствующее правило:

$$\Phi\left(z_{\text{критическое}}\right) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0{,}05}{2} = 0{,}45 \quad \Rightarrow \quad z_{\text{критическое}} = 1{,}64.$$

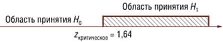


Рис. 3.10. Критическая область для примера 3.13

5. Ответ на вопрос о принадлежности наблюдаемого значения статистики критической области.

Наблюдаемое значение статистики $Z_{\text{наблюдаемое}} = 4{,}0035$ принадлежит критической области (4,0035>1,64).

6. Вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной.

Из принадлежности наблюдаемого значения критической области следует, что основная гипотеза отвергается в пользу альтернативной: новая технология позволяет увеличить среднюю выработку.

Ответ: при уровне значимости $\alpha = 0{,}05$ можно утверждать, что новая технология увеличивает среднюю выработку.

T-статистика применяется для решения похожих задач, но на выборках малых объемов и при неизвестных дисперсиях генеральных совокупностей.

T-статистики имеют распределение Стьюдента, для которого критические точки определяются по таблице с использованием уровня значимости α (для двусторонней или односторонней критической области) и числа степеней свободы k. Поэтому при перечислении задач, решаемых с привлечением T-статистики, обязательно указывают число степеней свободы k (табл. 3.5).

Таблица 3.5

Задачи на проверку гипотез с использованием *T*-статистики

Задача	Вид *T*-статистики	
Сравнение выборочной средней с генеральной средней нормальной совокупности при неизвестной дисперсии	$T=\dfrac{\overline{X}-\overline{x}_0}{\sqrt{\dfrac{S^2}{n-1}}}$, число степеней свободы $k=n-1$	$\overline{X}$ — выборочное среднее, $\overline{x}_0$ — генеральное среднее, задано по условию задачи, S^2 — «исправленная» дисперсия, n — объем выборки
Сравнение двух средних нормальных совокупностей при неизвестных, но равных дисперсиях $D(X)=D(Y)=\sigma^2$ генеральных совокупностей	$T=\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S^2}{n}+\dfrac{S^2}{m}}}$, $S^2=\dfrac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}$, число степеней свободы $k=n+m-2$	$\overline{X}$ — выборочное среднее из первой ГС, $\overline{Y}$ — выборочное среднее из второй ГС, S_X^2 — «исправленная» дисперсия для первой выборки, n — объем выборки из первой ГС, S_Y^2 — «исправленная» дисперсия для второй выборки, m — объем выборки из второй ГС
Сравнение двух средних нормальных совокупностей при неизвестных и неравных дисперсиях $D(X)\neq D(Y)$	$T=\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_X^2}{n}+\dfrac{S_Y^2}{m}}}$, число степеней свободы $k=\min(n-1;m-1)$	$\overline{X}$ — выборочное среднее из первой ГС, $\overline{Y}$ — выборочное среднее из второй ГС, S_X^2 — «исправленная» дисперсия для первой выборки

Окончание таблицы

Задача	Вид T-статистики	
генеральных совокупностей		n — объем выборки из первой ГС, S_Y^2 — «исправленная» дисперсия для второй выборки, m — объем выборки из второй ГС
Проверка значимости выборочного коэффициента корреляции	$T = \rho_в \cdot \dfrac{\sqrt{n-2}}{\sqrt{1-\rho_в^2}}$, число степеней свободы $k = n - 2$	$\rho_в$ — выборочный коэффициент корреляции, n — объем выборки

При использовании T-статистики критические значения определяют по табл. 3.6, используя выбранный уровень значимости α и определенные правила формирования критических областей (табл. 3.7).

Таблица 3.6

Критические точки распределения Стьюдента

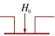
Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,1	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,3138	12,7062	31,8205	63,6567	318,3088	636,6192
2	2,9200	4,3027	6,9646	9,9248	22,3271	31,5991
3	2,3534	3,1824	4,5407	5,8409	10,2145	12,9240
4	2,1318	2,7764	3,7469	4,6041	7,1732	8,6103
5	2,0150	2,5706	3,3649	4,0321	5,8934	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,2076	5,9588
7	1,8946	2,3646	2,9980	3,4995	4,7853	5,4079
8	1,8595	2,3060	2,8965	3,3554	4,5008	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,1437	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370
12	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
	0,05	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)					

Окончание таблицы

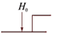
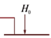
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (двусторонняя критическая область)					
14	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405
15	1,7531	2,1314	2,6025	2,9467	3,7328	4,0728
16	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
50	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602
70	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,1737	3,3905
	0,05	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)					

Таблица 3.7

Правила формирования критических областей при использовании *T*-статистики

Двусторонняя критическая область		По таблице критических точек распределения Стьюдента: $t_{\text{двустор. крит}}(\alpha, k)$, α — уровень значимости, k — число степеней свободы, $t_{\text{левое}} = -t_{\text{крит}}$,

Окончание таблицы

		$t_{\text{правое}} = t_{\text{крит}}$
Правосторонняя критическая область		По таблице критических точек распределения Стьюдента: $t_{\text{одностор. крит}}(\alpha, k)$, α — уровень значимости, k — число степеней свободы, $t_{\text{правое}} = t_{\text{крит}}$
Левосторонняя критическая область		По таблице критических точек распределения Стьюдента: $t_{\text{одностор. крит}}(\alpha, k)$, α — уровень значимости, k — число степеней свободы, $t_{\text{левое}} = -t_{\text{крит}}$

Пример 3.14. Менеджер предприятия, работающего в две смены, решил выяснить, существует ли различие в производительности труда рабочих дневной и вечерней смены. Случайно организованная выборка 10 рабочих дневной смены показала, что средний выпуск продукции составляет 74,3 ед./ч, а исправленная выборочная дисперсия оказалась равной S^2 = 16 ед.^{2}/ч. Выборка же 10 рабочих вечерней смены выявила, что средний выпуск продукции равнялся 69,7 ед./ч, а S^2 = 18 ед.^{2}/ч. При уровне значимости $\alpha = 0{,}02$ определить, существуют ли различия в производительности труда рабочих дневной и вечерней смены, если дисперсии производительности дневной и вечерней смены предположительно равны.

Решение.

При решении задачи будем придерживаться нашего плана.

1. Выдвижение основной гипотезы H_0 и альтернативной гипотезы H_1.

Основная гипотеза должна состоять в том, что различия в производительности труда рабочих дневной и вечерней смены не существуют, то есть производительности равны, альтернативную гипотезу можно выбирать ненаправленной, а именно — производительности не равны:

$H_0: \overline{x_0} = \overline{y_0}$,

$H_1: \overline{x_0} \neq \overline{y_0}$.

2. Выбор уровня значимости α для проведения проверки.

Уровень значимости нам задан по условию задачи: $\alpha = 0{,}02$.

3. Выбор статистики и вычисление наблюдаемого значения статистики.

Мы находимся в рамках стандартной задачи сравнения двух средних нормальных совокупностей при неизвестных, но равных (по условию) дисперсиях $D(X) = D(Y) = \sigma^2$, поэтому будем использовать статистику

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S^2}{n} + \frac{S^2}{m}}},$$

где $S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

и число степеней свободы $k = n+m-2$.

Вычислим наблюдаемое значение статистики:

$$S^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{9 \cdot 16 + 9 \cdot 18}{10+10-2} = 17,$$

$$T_{\text{наблюдаемое}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S^2}{n} + \frac{S^2}{m}}} = \frac{74{,}3 - 69{,}7}{\sqrt{\frac{17}{10} + \frac{17}{10}}} = 2{,}495.$$

Число степеней свободы $k = n+m-2 = 18$.

4. Формирование критической области.

Критическая область (рис. 3.11) будет формироваться как двусторонняя (определяется альтернативной гипотезой H_1: $\overline{x_0} \neq \overline{y_0}$), $a = 2{,}55$ находят по таблице, используя двустороннюю область при уровне значимости гипотезы $a = 0{,}02$ и числе степеней свободы $k = 18$:

Рис. 3.11. Критическая область для примера 3.14

5. Ответ на вопрос о принадлежности наблюдаемого значения статистики критической области.

Наблюдаемое значение статистики $T_{\text{наблюдаемое}} = 2{,}495$ не принадлежит критической области ($-2{,}55 < 2{,}495 < 2{,}55$).

6. Вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной. Из того что наблюдаемое значение статистики не принадлежит критической области, следует, что нет статистически значимых оснований отвергать основную гипотезу, и она принимается.

Ответ: при уровне значимости $\alpha = 0{,}02$ можно утверждать, что различия в производительности труда рабочих дневной и вечерней смены не существуют, полученные по выборкам различия носят случайный характер.

При изучении двух нормально распределенных генеральных совокупностей часто возникает необходимость сравнения их стандартных отклонений или дисперсий. Для решения такой задачи применяется F-статистика (табл. 3.8).

Таблица 3.8

Задача на проверку гипотезы с использованием *F*-статистики

Задача	Вид F-статистики	
Сравнение двух дисперсий нормально распределенных совокупностей	$F=\frac{S_1^2}{S_2^2}, S_1^2>S_2^2$, $k_1=n_1-1$ — число степеней свободы числителя, $k_2=n_2-1$ — число степеней свободы знаменателя	S_1^2 — «исправленная» дисперсия для выборки объема n_1, S_2^2 — «исправленная» дисперсия для выборки объема n_2

F-статистика имеет распределение Фишера со степенями свободы k_1, k_2, для которого составлены таблицы критических точек (табл. 3.9, 3.10):

Таблица 3.9

Критические точки распределения Фишера при уровне значимости α = 0,01 (k_1 — число степеней свободы большей дисперсии, k_2 — число степеней свободы меньшей дисперсии)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98.49	99.01	90.17	99.25	99.33	99.30	99.34	99.36	99.36	99.40	99.41	99.42
3	34.12	30.81	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.89
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11

Окончание таблицы

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
10	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71
11	9.86	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45

Таблица 3.10

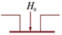
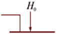
Критические точки распределения Фишера при уровне значимости α = 0,05 (k_1 — число степеней свободы большей дисперсии, k_2 — число степеней свободы меньшей дисперсии)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18.5	19.00	19.16	19.25	19:30	19.33	19.36	19.37	19.38	19.39	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38

Для формирования критической области при использовании F-статистики критические значения выбирают из таблицы, используя выбранный уровень значимости α и определенные правила (табл. 3.11).

Таблица 3.11

Правила формирования критических областей при использовании *F*-статистики

Двусторонняя критическая область		По таблице критических точек распределения Фишера: $F_{крит}(\frac{\alpha}{2}, k_1, k_2)$, α — уровень значимости, k_1, k_2 — число степеней свободы большей и меньшей дисперсии
Правосторонняя критическая область		По таблице критических точек распределения Фишера: $F_{крит}(\alpha, k_1, k_2)$, α — уровень значимости, k_1, k_2 — число степеней свободы большей и меньшей дисперсии
Левосторонняя критическая область		Не рассматривается

Пример 3.15. По выборкам объемов $n_X = 12$ и $n_Y = 11$ из двух нормально распределенных совокупностей получены «исправленные» дисперсии $S_X^2 = 6{,}3$ и $S_Y^2 = 8{,}5$. Можно ли на основе этих данных утверждать, что генеральные совокупности имеют равные дисперсии?

Решение.

При решении задачи будем придерживаться нашего плана.

1. Выдвижение основной гипотезы H_0 и альтернативной гипотезы H_1 .

Основная гипотеза должна состоять в том, дисперсии генеральных совокупностей равны, альтернативную гипотезу можно выбирать направленной:

H_0 : $D(X) = D(Y)$,

H_1 : $D(Y) > D(X)$.

2. Выбор уровня значимости α для проведения проверки.

Уровень значимости нам не задан по условию задачи, в таких случаях стандартно выбирают $\alpha = 0{,}05$.

3. Выбор статистики и вычисление наблюдаемого значения статистики.

Мы находимся в рамках стандартной задачи сравнения двух дисперсий нормальных совокупностей, поэтому будем использовать F-статистику

$$F = \frac{S_1^2}{S_2^2}, S_1^2 > S_2^2 .$$

По условию $S_Y^2 = 8{,}5 \quad > \quad S_X^2 = 6{,}3$, поэтому

$S_1^2 = 8{,}5$ и $k_1 = 11 - 1 = 10$,

$S_2^2 = 6{,}3$ и $k_2 = 12 - 1 = 11$.

Вычислим наблюдаемое значение статистики:

$$F_{\text{наблюдаемое}} = \frac{S_1^2}{S_2^2} = \frac{8{,}5}{6{,}3} = 1{,}349 .$$

4. Формирование критической области/

Критическая область (рис. 3.12) будет формироваться как правосторонняя (определяется альтернативной гипотезой $H_1: \; D(Y) > D(X)$), а при вычислении критического значения используем соответствующее правило и таблицу критических точек распределения Фишера:

$$F_{\text{крит}}(\alpha, k_1, k_2) = F_{\text{крит}}(0{,}05; 10; 11) = 2{,}86 .$$

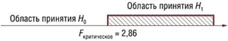


Рис. 3.12. Критическая область для дримера 3.15

5. Ответ на вопрос о принадлежности наблюдаемого значения статистики критической области/

Наблюдаемое значение статистики $F_{\textit{наблюдаемое}} = 1{,}349$ не принадлежит критической области (1,349 < 2,86).

6. Вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной:

Из того что наблюдаемое значение статистики не принадлежит критической области, следует, что основная гипотеза принимается.

Ответ: при уровне значимости $\alpha = 0{,}05$ можно утверждать, что дисперсии генеральных совокупностей равны.

Если имеется выборка из нормально распределенной генеральной совокупности и вычислена «исправленная» дисперсия, то, применяя статистику χ^2 , можно ответить на вопрос о соответствии выборочного значения дисперсии значению дисперсии генеральной совокупности. Статистика χ^2 применяется и в качестве критерия согласия — статистического критерия о предполагаемом виде закона распределения генеральной совокупности (табл. 3.12).

Напомним, что одной из задач математической статистики является установление теоретического закона распределения по эмпирическому (выборочному) распределению. Статистика χ^2 имеет распределение Пирсона (распределение χ^2), поэтому число степеней свободы указывается обязательно.

Таблица 3.12

Задачи на проверку гипотез с использованием статистики χ^2

Задача	Вид статистики χ^2	
Сравнение выборочной дисперсии с дисперсией генеральной совокупности $D(X)=\sigma^2$	$\chi \;= \dfrac{(n- \;)\cdot S}{\;}$ число степеней свободы $k=n-1$	S^2 — «исправленная» дисперсия выборки, n — объем выборки, $D(X)=\sigma^2$ — дисперсия генеральной совокупности (задана по условию)
Критерий согласия — проверка гипотезы о законе распределения генеральной совокупности	$\chi^2=\sum\limits_i \dfrac{\left(n_i-n_i^{'}\right)^2}{n_i^{'}}$ число степеней свободы $k=r-l-1$	n_i — эмпирические частоты, $n_i^{'}$ — теоретические частоты, r — количество групп (интервалов) наблюдений наполняемостью более 5 вариант, l — количество параметров теоретического распределения

Критические точки распределения χ^2 находят в табл. 3.13 по уровню значимости α и числу степеней свободы k.

При формировании критической области используют таблицу 3.13 и соответствующие правила.

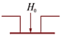
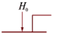
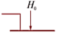
Таблица 3.13

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α								
	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99
1	6,634	5,023	3,841	2,705	0,454	0,015	0,003	0,0009	0,0002
2	9,210	7,377	5,991	4,605	1,386	0,210	0,102	0,050	0,020
3	11,34	9,348	7,814	6,251	2,365	0,584	0,351	0,215	0,114
4	13,27	11,14	9,487	7,779	3,356	1,063	0,710	0,484	0,297
5	15,08	12,83	11,07	9,236	4,351	1,610	1,145	0,831	0,554
6	16,81	14,44	12,59	10,64	5,348	2,204	1,635	1,237	0,872
7	18,47	16,01	14,06	12,01	6,345	2,833	2,167	1,689	1,239
8	20,09	17,53	15,50	13,36	7,344	3,489	2,732	2,179	1,646
9	21,66	19,02	16,91	14,68	8,342	4,168	3,325	2,700	2,087
10	23,20	20,48	18,30	15,98	9,341	4,865	3,940	3,246	2,558
11	24,72	21,92	19,67	17,27	10,34	5,577	4,574	3,815	3,053
12	26,21	23,33	21,02	18,54	11,34	6,303	5,226	4,403	3,570
13	27,68	24,73	22,36	19,81	12,33	7,041	5,891	5,008	4,106
14	29,14	26,11	23,68	21,06	13,33	7,789	6,570	5,628	4,660
15	30,57	27,48	24,99	22,30	14,33	8,546	7,260	6,262	5,229
16	31,99	28,84	26,29	23,54	15,33	9,312	7,961	6,907	5,812
17	33,40	30,19	27,58	24,76	16,33	10,08	8,671	7,564	6,407
18	34,80	31,52	28,86	25,98	17,33	10,86	9,390	8,230	7,014
19	36,19	32,85	30,14	27,20	18,33	11,65	10,11	8,906	7,632
20	37,56	34,16	31,41	28,41	19,33	12,44	10,85	9,590	8,260
21	38,93	35,47	32,67	29,61	20,33	13,23	11,59	10,28	8,897
22	40,28	36,78	33,92	30,81	21,33	14,04	12,33	10,98	9,542
23	41,63	38,07	35,17	32,00	22,33	14,84	13,09	11,68	10,19
24	42,97	39,36	36,41	33,19	23,33	15,65	13,84	12,40	10,85
25	44,31	40,64	37,65	34,38	24,33	16,47	14,61	13,11	11,52
26	45,64	41,92	38,88	35,56	25,33	17,29	15,37	13,84	12,19
27	46,96	43,19	40,11	36,74	26,33	18,11	16,15	14,57	12,87
28	48,27	44,46	41,33	37,91	27,33	18,93	16,92	15,30	13,56
29	49,58	45,72	42,55	39,08	28,33	19,76	17,70	16,04	14,25
30	50,89	46,97	43,77	40,25	29,33	20,59	18,49	16,79	14,95

Таблица 3.14

Правила формирования критических областей при использовании статистики χ^2

Двусторонняя критическая область		По таблице критических точек распределения χ^2: $\chi^2_{\text{левое крит}}(1-\frac{\alpha}{2},k)$, α — уровень значимости, k — число степеней свободы, $\chi^2_{\text{правое крит}}(\frac{\alpha}{2},k)$ — уровень значимости, k — число степеней свободы
Правосторонняя критическая область		По таблице критических точек распределения χ^2: $\chi^2_{\text{крит}}(\alpha,k)$, α — уровень значимости, k — число степеней свободы
Левосторонняя критическая область		По таблице критических точек распределения χ^2: $\chi^2_{\text{крит}}(1-\alpha,k)$, α — уровень значимости, k — число степеней свободы

Отметим, что при проверке гипотезы на соответствие закону распределения альтернативная гипотеза явно не выдвигается, а критическая область стандартно выбирается правосторонней.

Пример 3.16. Проверить, согласуется ли предположение о нормальном распределении генеральной совокупности с эмпирическим распределением выборки, полученной из этой совокупности:

x_i	190–200	200–210	210–220	220–230	230–240	240–250
n_i	10	26	56	64	30	14

Решение.

1. При использовании критерия согласия в основной гипотезе H_0 утверждается, что эмпирическое и теоретическое распределения согласованы, альтернативная гипотеза H_1 по умолчанию предполагает противоположное — распределения не согласованы.

2. Уровень значимости выбираем стандартно $\alpha = 0{,}05$.

3. Будем использовать статистику χ^2:

$$\chi^2 = \sum_i \frac{\left(n_i - n_i^{'}\right)^2}{n_i^{'}},$$

где n_i — эмпирические частоты,

$n_i^{'}$ — теоретические частоты.

Эмпирические частоты берем из данного эмпирического распределения, теоретические — вычисляем по формуле: $n_i^{'} = np_i$, где n — объем выборки, p_i — теоретическая вероятность попадания значения величины в i-й интервал. Так как мы проверяем гипотезу о соответствии нормальному распределению, то теоретическую вероятность p_i будем вычислять по формуле:

$$p_i = P(x_i < X < x_{i+1}) = \Phi\left(\frac{x_{i+1} - \overline{X}}{\sigma_в}\right) - \Phi\left(\frac{x_i - \overline{X}}{\sigma_в}\right).$$

Вычислим выборочное среднее $\overline{X}$ и выборочное стандартное отклонение $\sigma_в$:

$$\overline{X} = \frac{195\cdot 10 + 205\cdot 26 + 215\cdot 56 + 225\cdot 64 + 235\cdot 30 + 245\cdot 14}{10 + 26 + 56 + 64 + 30 + 14} = 221;$$

$$D_в(X) =$$

$$\frac{195^2\cdot 10 + 205^2\cdot 26 + 215^2\cdot 56 + 225^2\cdot 64 + 235^2\cdot 30 + 245^2\cdot 14}{200} - 221^2 =$$

$$= 152;$$

$$\sigma_в = \sqrt{D_в(X)} = \sqrt{152} = 12{,}33.$$

Теперь вычислим теоретические вероятности, учитывая, что нормальное распределение определено на всей числовой прямой $(-\infty; +\infty)$:

$$p_1 = P(-\infty < X < 200) = \Phi\left(\frac{200 - 221}{12{,}33}\right) - \Phi(-\infty) =$$
$$= \Phi(-1{,}7) - (-0{,}5) = -0{,}45543 + 0{,}5 = 0{,}04457,$$

$$p_2 = P(200 < X < 210) \quad = \Phi\left(\frac{210-221}{12{,}33}\right) - \Phi\left(\frac{200-221}{12{,}33}\right) =$$
$$= \Phi(-0{,}89) - \Phi(-1{,}7) = -0{,}31327 + 0{,}45543 = 0{,}14216,$$

$$p_3 = P(210 < X < 220) \quad = \Phi\left(\frac{220-221}{12{,}33}\right) - \Phi\left(\frac{210-221}{12{,}33}\right) =$$
$$= \Phi(-0{,}08) - \Phi(-0{,}89) = -0{,}03188 + 0{,}31327 = 0{,}28139,$$

$$p_4 = P(220 < X < 230) \quad = \Phi\left(\frac{230-221}{12{,}33}\right) - \Phi\left(\frac{220-221}{12{,}33}\right) =$$
$$= \Phi(0{,}73) - \Phi(-0{,}08) = 0{,}26730 + 0{,}03188 = 0{,}29919,$$

$$p_5 = P(230 < X < 240) \quad = \Phi\left(\frac{240-221}{12{,}33}\right) - \Phi\left(\frac{230-221}{12{,}33}\right) =$$
$$= \Phi(1{,}54) - \Phi(0{,}73) = 0{,}43822 - 0{,}26730 = 0{,}17092,$$

$$p_5 = P(240 < X < +\infty) \quad = \Phi(+\infty) - \Phi\left(\frac{240-221}{12{,}33}\right) =$$
$$= 0{,}5 - \Phi(1{,}54) = 0{,}5 - 0{,}26730 = 0{,}06178.$$

Условие $\sum_i p_i = 1$ выполняется.

Для того, чтобы вычислить наблюдаемое значение статистики, заполним вспомогательную таблицу 3.15:

Таблица 3.15

Вспомогательная таблица к примеру 3.16

Интервал значения x_i	190–200	200–210	210–220	220–230	230–240	240–250	$\sum$
n_i	10	26	56	64	30	14	200
p_i	0,04457	0,14216	0,28139	0,29918	0,17092	0,06178	1
$np_i = 200p_i$	8,914	28,432	56,278	59,836	34,184	12,356	200
$(n_i - np_i)^2$	$(10-8{,}914)^2 =$ =1,179	5,914	0,077	17,339	17,506	2,703	
$\frac{(n_i - np_i)^2}{np_i}$	$\frac{1{,}179}{8{,}914} =$ =0,132	0,208	0,001	0,290	0,512	0,219	–1,362

Наблюдаемое значение статистики выделено в таблице как сумма:

$$\chi^2_{\text{наблюдаемое}} = \sum_i \frac{\left(n_i - n_i'\right)^2}{n_i'} = \sum_i \frac{\left(n_i - np_i\right)^2}{np_i} = 1{,}362.$$

4. Формирование критической области.

Критическая область (рис. 3.13) в таких задачах всегда правосторонняя. При вычислении критического значения используем соответствующее правило (табл. 3.13) и таблицу 3.12 критических точек распределения χ^2:

$$\chi^2_{\text{крит}}(\alpha, k) = \chi^2_{\text{крит}}\left(0{,}05; 3\right) = 7{,}8\ .$$

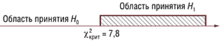


Рис. 3.13. Критическая область для примера 3.16

Число степеней свободы вычисляли по формуле $k = r - l - 1$ при $r = 6$ (количество интервалов в выборочной совокупности), $l = 2$ (количество параметров нормального распределения).

5. Ответ на вопрос о принадлежности наблюдаемого значения статистики критической области.

Наблюдаемое значение статистики $\chi^2_{\text{наблюдаемое}} =$ 1,362 не принадлежит критической области (1,362 < 7,8).

6. Вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной.

Из того что наблюдаемое значение статистики не принадлежит критической области, следует, что основная гипотеза принимается.

Ответ: при уровне значимости $\alpha = 0{,}05$ можно утверждать, что данное эмпирическое распределение согласуется с нормальным законом распределения.

В рассмотренном примере мы проверяли гипотезу о соответствии выборочных данных нормальному распределению, поэтому теоретические вероятности вычисляли с применением формул, имеющих место для нормального распределения. Если проверять гипотезы на соответствие другим видам распределений, то и соответствующие вычисления теоретических вероятностей необходимо выполнять, опираясь на те формулы, которые имеют место для проверяемого закона распределения. Кроме того, при вычислении критического значения важно помнить о количестве параметров проверяемого закона распределения — это позволит правильно определить число степеней свободы.

Задачи для самостоятельного решения

В задачах **3.91—3.100** осуществить проверку статистической гипотезы на заданном уровне значимости α.

3.91. Генеральная совокупность распределена по нормальному закону с известными числовыми характеристиками $E(X) = 12$ и $D(X) = 16$. Из генеральной совокупности извлекается выборка объема $n = 25$, для которой выборочное среднее $\overline{X} = 10{,}4$. На уровне значимости $\alpha = 0{,}01$ проверить гипотезу о соответствии выборочных данных данным генеральной совокупности.

3.92. Вероятность выпуска бракованного изделия на предприятии составляет 0,03. Из 1000 взятых на проверку изделий предприятия бракованными оказались 38. На уровне значимости $\alpha = 0{,}05$ проверить гипотезу на соответствие наблюдаемой относительной частоты появления бракованного изделия в выборке теоретической вероятности.

3.93. Из двух независимых генеральных совокупностей, распределенных по нормальному закону с известными дисперсиями $D(X) = 81$ и $D(Y) = 90$, извлечены выборки объемов $n = 20$ и $m = 25$ соответственно, для которых вычислены выборочные средние $\overline{X} = 15$ и $\overline{Y} = 14$. На уровне значимости $\alpha = 0{,}02$ проверить гипотезу о равенстве математических ожиданий генеральных совокупностей.

3.94. Из 2000 выбранных изделий, выпускаемых первым предприятием, бракованными оказались 60, а из 1500 выбранных изделий, выпускаемых вторым предприятием, бракованными оказались 48. Можно ли на уровне значимости $\alpha = 0{,}03$ утверждать, что вероятности выпуска бракованных изделий на первом и втором предприятиях совпадают?

3.95. Известно, что генеральная совокупность распределена по нормальному закону с математическим ожиданием $E(X) = 11$. Из нее извлекается выборка объема $n = 15$, для которой вычислены выборочное среднее $\overline{X} = 12$ и выборочная дисперсия $D_в(X) = 4{,}5$. На уровне значимости $\alpha = 0{,}01$ проверить гипотезу о соответствии выборочных данных данным генеральной совокупности.

3.96. Из двух независимых генеральных совокупностей, распределенных по нормальному закону с неизвестными, но равными дисперсиями $\left(D(X)=D(Y)\right)$, извлечены выборки объемов $n = 10$ и $m = 15$ соответственно, для которых вычислены выборочные средние $\overline{X} = 115$ и $\overline{Y} = 118$ и выборочные дисперсии $D_в(X) = 60$ и $D_в(Y) = 64$. На уровне значимости $\alpha = 0{,}01$ проверить гипотезу о равенстве математических ожиданий генеральных совокупностей.

3.97. Из двух независимых генеральных совокупностей, распределенных по нормальному закону, извлечены выборки объемов $n = 10$ и $m = 9$, для которых вычислены выборочные дисперсии $D_в(X) = 60$ и $D_в(Y) = 64$ соответственно. Проверить гипотезу о равенстве дисперсий генеральных совокупностей на уровне значимости $\alpha = 0{,}05$.

3.98. Из генеральной совокупности с известной дисперсией $D(X) = 100$ извлечена выборка объема $n = 35$ и вычислена выборочная дисперсия $D_в(X) = 134{,}2$. На уровне значимости $\alpha = 0{,}02$ проверить гипотезу о соответствии выборочных данных данным генеральной совокупности.

3.99. Проверить на уровне значимости $\alpha = 0{,}01$, согласуется ли предположение о показательном распределении генеральной совокупности с эмпирическим распределением выборки, полученной из этой совокупности:

Значение X	1	2	3	4	5	6
Частота	125	100	94	82	61	48

3.100. Проверить на уровне значимости $\alpha = 0{,}02$, согласуется ли предположение о нормальном распределении генеральной совокупности с эмпирическим распределением выборки, полученной из этой совокупности:

Значение X	25–35	35–45	45–55	55–65	65–75	75–85
Частота	38	64	94	88	72	29

3.5. Понятие о корреляционно-регрессионном анализе

При рассмотрении величин особое внимание уделяют изучению зависимости между ними. Если какая-либо величина определяется как *однозначная* функция каких-либо величин, то говорят о *функциональной* зависимости между этими величинами. Всем известны, например, зависимости между производительностью труда, объемом выпускаемой продукции и временем, в течение которого эта продукция была выпущена, или между скоростью движения, временем и расстоянием. Функциональная зависимость возникает всякий раз, когда существует функция $y = f(x_1, x_2, ..., x_n)$ и значение однозначно определяется для каждого возможного набора значений $(x_1, x_2, ..., x_n)$. Между случайными величинами тоже могут возникать функциональные зависимости, и мы о них уже говорили.

Однако между случайными величинами чаще возникает вероятностная, или *стохастическая* зависимость.

Стохастическая зависимость между двумя величинами Y и X обычно возникает в том случае, когда имеются как общие случайные факторы, влияющие на обе величины, так и другие факторы, неодинаковые для обеих случайных величин. В этом случае фиксированному значению величины X соответствует множество возможных значений величины Y, причем заранее не известно, какое именно из этих значений будет принято величиной.

Стохастические зависимости в рамках математико-статистического исследования изучаются с использованием законов распределений рассматриваемых величин, вероятностей, с которыми встречаются те или иные комбинации значений исследуемых величин и т.д. Можно говорить о том, что изучение стохастических зависимостей без разработанного в курсе теории вероятностей математического аппарата невозможно. Мы рассмотрим один из примеров стохастической зависимости — корреляционную зависимость между величинами. Теория корреляции позволяет решать задачи обоснованного прогноза: когда указаны пределы, в которых с заданной надежностью будут содержаться значения интересующей нас величины, если связанные с ней величины принимают определенные значения.

Если имеется стохастическая зависимость величины Y от величины X, то прежде всего обращают внимание на изменения условного математического ожидания $E(Y|X = x)$ при изменениях значения

$X = x$. Если при изменении x условные математические ожидания $E(Y|X = x)$ меняются, то говорят о наличии корреляционной зависимости Y от X. Функция, которая описывает эти изменения, называется функцией регрессии.

Пример 3.17. Известно распределение двумерной дискретной случайной величины (X,Y):

	$X=2$	$X=3$	$X=5$
$Y=0,3$	0,15	0,2	0,01
$Y=0,5$	0,05	0,45	0,14

Определить, существует ли между величинами Y и X стохастическая зависимость.

Решение.

О наличии стохастической зависимости будет говорить изменение значений условных математических ожиданий $E(Y|X = x)$ при изменении значений $X = x$.

Найдем условные законы распределений $(Y|X = x)$ при всех возможных значениях $X = x$ ($X = 2$, $X = 3$, $X = 5$) и вычислим соответствующие условные математические ожидания.

Значение $(Y\|X=2)$	0,3	0,5
Вероятность	0,75	0,25

При вычислении вероятностей использовали определение условной вероятности:

$$P(Y=0,3|X=2) = \frac{P(Y=0,3;X=2)}{P(X=2)} = \frac{0,15}{0,15+0,05} = \frac{0,15}{0,20} = \frac{3}{4} = 0,75,$$

$$P(Y=0,5|X=2) = \frac{P(Y=0,5;X=2)}{P(X=2)} = \frac{0,05}{0,15+0,05} = \frac{0,05}{0,20} = \frac{1}{4} = 0,25.$$

Найдем условное математическое ожидание

$E(Y|X = 2) = 0,3\cdot 0,75 + 0,5\cdot 0,25 = 0,35$.

Значение $(Y\|X=3)$	0,3	0,5
Вероятность	$\frac{4}{13}$	$\frac{9}{13}$

Вероятности вычисляем аналогично:

$$P\left(Y=0{,}3\middle|X=3\right)=\frac{P(Y=0{,}3;X=3)}{P(X\quad 3)}=\frac{0{,}2}{0{,}2+0{,}45}=\frac{0{,}2}{0{,}65}=\frac{4}{13},$$

$$P\left(Y=0{,}5\middle|X=3\right)=\frac{P(Y=0{,}5;X=3)}{P(X=3)}=\frac{0{,}45}{0{,}2+0{,}45}=\frac{0{,}45}{0{,}65}=\frac{9}{13}.$$

Найдем условное математическое ожидание:

$$E(Y|X=3)=0{,}3\cdot\frac{4}{13}+0{,}5\cdot\frac{9}{13}=\frac{57}{130}.$$

Значение $\left(Y\middle|X=5\right)$	0,3	0,5
Вероятность	$\frac{1}{15}$	$\frac{14}{15}$

Вероятности вычисляем аналогично:

$$P\left(Y=0{,}3\middle|X=5\right)=\frac{P(Y=0{,}3;X=5)}{P(X=5)}=\frac{0{,}01}{0{,}01+0{,}14}=\frac{0{,}01}{0{,}15}=\frac{1}{15},$$

$$P\left(Y=0{,}5\middle|X=5\right)=\frac{P(Y=0{,}5;X=5)}{P(X=5)}=\frac{0{,}14}{0{,}01+0{,}14}=\frac{0{,}14}{0{,}15}=\frac{14}{15}.$$

Найдем условное математическое ожидание:

$$E(Y|X=5)=0{,}3\cdot\frac{1}{15}+0{,}5\cdot\frac{14}{15}=\frac{73}{150}.$$

Таким образом, зависимость между величинами Y и X является стохастической, потому что величина Y может принимать значения 0,3 или 0,5 и нельзя сказать заранее — какое именно значение она примет, более того, вероятность $P(Y=0{,}3)$ изменяется при изменении значения $X=x$.

Мы видим, что с изменением значения X меняется и значение условного математического ожидания:

	$X=2$	$X=3$	$X=5$
$E(Y\|X=x)$	0,35	$\frac{57}{130}$	$\frac{73}{150}$

Это означает, что стохастическая зависимость между величинами Y и X является корреляционной, а составленная таблица задает функцию регрессии Y по X. Так как величины X и Y по условию — дискретные и принимают только определенные значения, то функция регрессии определяется только для допустимых значений X и Y.

Ответ: между величинами Y и X существует стохастическая зависимость, являющаяся корреляционной.

Информацию о связи величин X и Y дает и коэффициент корреляции

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

Если коэффициент корреляции равен 1 или −1, то между величинами существует линейная зависимость вида $Y=\alpha+\beta X$, где α,β — некоторые постоянные. Чем ближе значение $|\rho|$ к 1, тем сильнее корреляционная связь между величинами.

При неизвестном двумерном распределении (X,Y) прибегают к выборочным оценкам. Если доказано, что корреляционная связь между величинами средняя или сильная, то можно составить уравнение линейной регрессии Y по X, используя следующую формулу:

$$y=\overline{Y}+\rho_в(X,Y)\cdot\frac{\sigma_в(Y)}{\sigma_в(X)}\left(x-\overline{X}\right).$$

Коэффициентом регрессии $\beta_{Y|X}$ называют коэффициент при аргументе x:

$$\beta_{Y|X}=\rho_в(X,Y)\cdot\frac{\sigma_в(Y)}{\sigma_в(X)}.$$

Аналогично можно составлять уравнение линейной регрессии X по Y:

$$x=\overline{X}+\rho_в(X,Y)\cdot\frac{\sigma_в(X)}{\sigma_в(Y)}\left(y-\overline{Y}\right).$$

Коэффициент регрессии $\beta_{X|Y}$ в этом случае имеет вид:

$$\beta_{X|Y}=\rho_в(X,Y)\cdot\frac{\sigma_в(X)}{\sigma_в(Y)}.$$

Коэффициент регрессии фактически является угловым коэффициентом прямой и определяет тангенс угла наклона прямой регрессии к положительному направлению оси *Ox*.

Пример 3.18. Для изучения двумерной случайной величины (X,Y) из генеральной совокупности извлечена выборочная совокупность объема $n=10$ и получено эмпирическое распределение:

	$X=8$	$X=9$	$X=11$	$X=12$
$Y=5$	2	1	0	0
$Y=6$	2	1	0	0
$Y=7$	0	1	0	0
$Y=8$	0	0	0	1
$Y=10$	0	0	1	1

Проверить наличие корреляционной зависимости между величинами и в случае средней или сильной связи составить уравнение линейной регрессии Y по X.

Решение.

Для того, чтобы установить наличие корреляционной зависимости между величинами, вычислим выборочный коэффициент корреляции по формуле: $\rho_в(X,Y)=\dfrac{\operatorname{cov}_в(X,Y)}{\sigma_в(X)\cdot\sigma_в(Y)}$, где $\operatorname{cov}_в(X,Y)=\dfrac{\sum\limits_i\sum\limits_j x_i y_j n_{ij}}{n}-\overline{X}\cdot\overline{Y}$.

Найдем:

$$\overline{X}=\frac{8\cdot 4+9\cdot 3+11\cdot 1+12\cdot 2}{10}=9{,}4,$$

$$D_в(X)=\frac{8^2\cdot 4+9^2\cdot 3+11^2\cdot 1+12^2\cdot 2}{10}-9{,}4^2=2{,}44,$$

$$\sigma_в(X)=\sqrt{2{,}44}=1{,}562,$$

$$\overline{Y}=\frac{5\cdot 3+6\cdot 3+7\cdot 1+8\cdot 1+10\cdot 2}{10}=6{,}8,$$

$$D_в(Y)=\frac{5^2\cdot 3+6^2\cdot 3+7^2\cdot 1+8^2\cdot 1+10^2\cdot 2}{10}-6{,}8^2=6{,}36,$$

$$\sigma_в(Y)=\sqrt{6{,}36}=2{,}522,$$

$$\operatorname{cov}_в(X,Y)=\frac{\sum\limits_i\sum\limits_j x_i y_j n_{ij}}{n}-\overline{X}\cdot\overline{Y}=$$

$$=\frac{8\cdot 5\cdot 2+8\cdot 6\cdot 2+9\cdot 5\cdot 1+9\cdot 6\cdot 1+9\cdot 7\cdot 1+11\cdot 10\cdot 1+12\cdot 8\cdot 1+12\cdot 10\cdot 1}{10}-$$

$$-9{,}4\cdot 6{,}8=$$

$=2{,}48.$

$$\rho_в(X,Y)=\frac{\operatorname{cov}_в(X,Y)}{\sigma_в(X)\cdot\sigma_в(Y)}=\frac{2{,}48}{1{,}562\cdot 2{,}522}=0{,}63.$$

Полученный коэффициент корреляции говорит о средней связи между величинами, поэтому можно составить и уравнение линейной регрессии Y по X:

$$y = \overline{Y} + \rho_в(X,Y)\cdot\frac{\sigma_в(Y)}{\sigma_в(X)}\left(x-\overline{X}\right) \Rightarrow$$

$$y = 6{,}8 + 0{,}63\cdot\frac{2{,}522}{1{,}562}\left(x-9{,}4\right) \Rightarrow$$

$$y = 6{,}8 + 1{,}017\left(x-9{,}4\right) \Rightarrow$$

$y = 1{,}017x - 2{,}7598$ — искомое уравнение линейной регрессии.

Ответ: $y = 1{,}017x - 2{,}7598$.

Используя составленное уравнение, можно рассчитать по заданным значениям величины X соответствующие значения величины Y:

$$X = 8 \Rightarrow y = 1{,}017\cdot 8 - 2{,}7598 = 5{,}3762;$$

$$X = 9 \Rightarrow y = 1{,}017\cdot 9 - 2{,}7598 = 6{,}3932;$$

$$X = 11 \Rightarrow y = 1{,}017\cdot 11 - 2{,}7598 = 8{,}4272;$$

$$X = 12 \Rightarrow y = 1{,}017\cdot 12 - 2{,}7598 = 9{,}4442.$$

Мы видим, что вычисленные значения не совпадают с теми значениями Y, которые получены по выборке, однако коэффициенты в уравнение регрессии подбираются таким образом, чтобы все отклонения эмпирических значений от теоретических являлись минимальными (рис. 3.14).

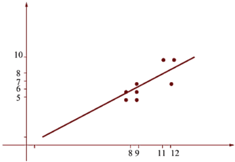


Рис. 3.14. Корреляционная и регрессионная зависимости из примера 3.18

Более того, прямая регрессии обязательно проходит через центр распределения — точку с координатами $\left(\overline{X},\overline{Y}\right)$, в нашем примере это точка $(9{,}4;\ 6{,}8)$.

Уравнение линейной регрессии составляется для прогнозирования зависимости величин Y и X в генеральной совокупности. Для обоснования такого прогноза необходимо удостовериться в значимости выборочного коэффициента корреляции. Такая задача решается с привлечением T-статистики

$$T = \rho_в \cdot \frac{\sqrt{n-2}}{\sqrt{1-\rho_в^2}},$$

где $\rho_в$ — выборочный коэффициент корреляции,

n — объем выборки и число степеней свободы $k = n-2$.

Если соответствующая проверка подтвердит значимость выборочного коэффициента корреляции, то его значение можно «перенести» в генеральную совокупность и формулировать выводы о зависимости величин Y и X в генеральной совокупности, предлагая для описания такой зависимости уравнение линейной регрессии.

Пример 3.19. Проверить значимость выборочного коэффициента корреляции, полученного при решении примера 3.18.

Решение.

Проверка значимости выборочного коэффициента корреляции является стандартной задачей на проверку статистической гипотезы, поэтому при решении будем придерживаться уже известного нам плана решения таких задач.

1. Выдвижение основной гипотезы H_0 и альтернативной гипотезы H_1.

Основная гипотеза H_0 — коэффициент корреляции в генеральной совокупности равен нулю, альтернативную гипотезу можно выбирать ненаправленной, а именно — коэффициент корреляции не равен нулю:

$H_0:\ \rho(X,Y)=0$,

$H_1:\ \rho(X,Y)\neq 0$.

2. Выбор уровня значимости α для проведения проверки:

Уровень значимости выбираем $\alpha = 0{,}01$.

3. Выбор статистики и вычисление наблюдаемого значения статистики.

Мы уже отметили, что для решения такого типа задачи применяется T-статистика с числом степеней свободы $k = n-2$:

$$T = \rho_в \cdot \frac{\sqrt{n-2}}{\sqrt{1-\rho_в^2}}.$$

Вычислим наблюдаемое значение статистики:

$$_{\text{наблюдаемое}} = \rho_в \cdot \frac{\sqrt{\ -2}}{\sqrt{1-\rho_в^2}} = 0{,}63 \cdot \frac{\sqrt{10-2}}{\sqrt{1-0{,}63^2}} = 2{,}29.$$

Число степеней свободы $k = n - 2 = 8$.

4. Формирование критической области.

Критическая область (рис. 3.15) будет формироваться как двусторонняя (определяется альтернативной гипотезой H_1: $\rho_в(X,Y) \neq \rho(X,Y)$,

$t_{крит} = 2{,}31$ находят по таблице, используя двустороннюю область при уровне значимости $a = 0{,}05$ и числе степеней свободы $k = 8$.

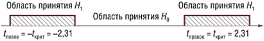


Рис. 3.15. Критическая область для примера 3.19

5. Ответ на вопрос о принадлежности наблюдаемого значения статистики критической области.

Наблюдаемое значение статистики $T_{наблюдаемое} = 2{,}29$ не принадлежит критической области ($-2{,}31 < 2{,}29 < 2{,}31$).

6. Вывод о принятии основной гипотезы или ее отклонении в пользу альтернативной.

Наблюдаемое значение статистики не принадлежит критической области, следовательно, основная гипотеза принимается.

Ответ: при уровне значимости $\alpha = 0{,}05$ можно утверждать, что выборочный коэффициент корреляции не является значимым: корреляционная зависимость, установленная по выборочным данным, носит случайный характер.

Полученный вывод позволяет нам теперь утверждать, что уравнение регрессии из примера 3.18 не «работает» для прогнозирования в генеральной совокупности на уровне $\alpha = 0{,}05$. Однако если уровень значимости определить $\alpha = 0{,}1$, то критическое значение будет равно 1,86 и наблюдаемое значение статистики $T_{наблюдаемое} = 2{,}29$ попадает в критическую область. Это уже позволяет сделать вывод о значимости выборочного коэффициента корреляции и существующей в генеральной совокупности зависимости, которую можно задать уравнением линейной регрессии.

Объяснить это можно тем, что значение выборочного коэффициента корреляции из примера 3.18 указывало на среднюю связь между величинами X и Y, наблюдаемые значения оказались «разбросанными» относительно прямой регрессии, в такой ситуации исследователь может настаивать на линейной зависимости в генеральной совокупности, но только с надежностью $\gamma = 1 - \alpha = 0{,}9$ и меньше.

Задачи для самостоятельного решения

В задачах **3.101—3.110,** изучив данное распределение двумерной величины (X,Y), определить наличие стохастической зависимости между величинами X и Y.

3.101.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-4$	0,2	0,1	0	0,1
$Y=-3$	0,1	0	0,1	0,1
$Y=-2$	0	0,1	0,1	0,1

3.102.

	$X=10$	$X=15$	$X=20$	$X=25$
$Y=10$	0	0	0,2	0,1
$Y=15$	0,2	0,3	0,1	0,1

3.103.

	$X=11$	$X=12$	$X=13$	$X=14$
$Y=1$	0	0,1	0	0,1
$Y=2$	0,3	0	0,2	0,1
$Y=3$	0	0,1	0	0,1

3.104.

	$X=100$	$X=110$	$X=120$	$X=130$
$Y=10$	0,2	0	0,2	0
$Y=20$	0	0,3	0	0,3

3.105.

	$X = -20$	$X = -5$	$X = 10$
$Y = -10$	0,1	0	0,4
$Y = 10$	0	0,3	0
$Y = 20$	0,1	0	0,1

3.106.

	$X = 2$	$X = 4$	$X = 8$	$X = 32$
$Y = -1$	0,09	0,17	0,08	0,03
$Y = 0$	0,01	0,02	0,07	0,1
$Y = 10$	0,25	0,15	0,02	0,01

3.107.

	$X = 0$	$X = 5$	$X = 10$	$X = 15$
$Y = 0$	0,19	0,1	0,07	0,02
$Y = 5$	0,01	0,09	0,07	0,02
$Y = 10$	0,26	0,13	0,02	0,02

3.108.

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 2$	0,1	0,1	0	0,1
$Y = 4$	0,1	0,1	0,1	0,1
$Y = 6$	0,1	0,1	0,1	0

3.109.

	$X = -3$	$X = -2$	$X = -1$	$X = 0$
$Y = -4$	0,05	0,1	0,02	0,03
$Y = -2$	0,2	0,05	0,03	0,01
$Y = 0$	0,25	0,15	0,05	0,06

3.110.

	$X = -2$	$X = -1$	$X = 1$	$X = 4$
$Y = -100$	0,05	0,05	0	0
$Y = 150$	0	0,1	0,25	0,35
$Y = 300$	0,1	0,1	0	0

В задачах **3.111–3.120** для изучения двумерной величины (X,Y) сформировали выборочную совокупность, распределение которой задано. Составить уравнение линейной регрессии Y по X и проверить значимость выборочного коэффициента корреляции.

3.111.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-4$	0	15	0	10
$Y=-3$	5	15	0	15
$Y=-2$	15	10	15	0

3.112.

	$X=10$	$X=15$	$X=20$	$X=25$
$Y=5$	20	10	0	0
$Y=10$	10	0	0	20
$Y=15$	0	0	30	10

3.113.

	$X=11$	$X=12$	$X=13$	$X=14$
$Y=1$	10	0	15	10
$Y=2$	0	9	5	5
$Y=3$	11	15	0	20

3.114.

	$X=100$	$X=110$	$X=120$	$X=130$
$Y=10$	19	0	12	15
$Y=20$	0	10	0	15
$Y=30$	11	0	18	0

3.115.

	$X=-20$	$X=-5$	$X=10$	$X=40$
$Y=-10$	4	2	11	15
$Y=10$	5	5	0	25

В задачах **3.111–3.120** для изучения двумерной величины (X,Y) сформировали выборочную совокупность, распределение которой задано. Составить уравнение линейной регрессии Y по X и проверить значимость выборочного коэффициента корреляции.

3.111.

	$X=-2$	$X=-1$	$X=0$	$X=1$
$Y=-4$	0	15	0	10
$Y=-3$	5	15	0	15
$Y=-2$	15	10	15	0

3.112.

	$X=10$	$X=15$	$X=20$	$X=25$
$Y=5$	20	10	0	0
$Y=10$	10	0	0	20
$Y=15$	0	0	30	10

3.113.

	$X=11$	$X=12$	$X=13$	$X=14$
$Y=1$	10	0	15	10
$Y=2$	0	9	5	5
$Y=3$	11	15	0	20

3.114.

	$X=100$	$X=110$	$X=120$	$X=130$
$Y=10$	19	0	12	15
$Y=20$	0	10	0	15
$Y=30$	11	0	18	0

3.115.

	$X=-20$	$X=-5$	$X=10$	$X=40$
$Y=-10$	4	2	11	15
$Y=10$	5	5	0	25

Список литературы

1. *Браилов А.В., Солодовников А.С.* Сборник задач по курсу «Математика в экономике». Часть 3. Теория вероятностей. М. : Финансы и статистика, 2010.
2. *Денежкина И.Е., Орлова М.Г., Швецов Ю.Н.* Основы математической статистики : учеб. пособие для самостоятельной работы бакалавров. М. : Финансовая академия при Правительстве РФ, 2010.
3. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. М. : ЮНИТИ, 2007.
4. *Солодовников А.С., Бабайцев В.А., Браилов А.В., Шандра И.Г.* Математика в экономике : учебник. В 3 ч. Ч. 3. Теория вероятностей и математическая статистика. М. : Финансы и статистика, 2008.

Приложение 1

Таблица значений функции Гаусса $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061

Окончание таблицы

	0	1	2	3	4	5	6	7	8	9
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Приложение 2

Таблица значений функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02892	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	035769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41308	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327

Окончание таблицы

	0	1	2	3	4	5	6	7	8	9
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49422	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49336	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995

Приложение 3

Таблица значений функции Пуассона $P_n(k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2223	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1216	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005

k \ λ	2.0	3.0	4.0	5.0	6,0	7,0	8,0	9,0	10,0
0	0,1353	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0001
1	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005
2	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023
3	0,1805	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076
4	0,0902	0,1681	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337	0,0189
5	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378
6	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631
7	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171	0,0901
8	0,0009	0,0081	0,0298	0,0653	0,1033	0,1304	0,1396	0,1318	0,1126
9	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318	0,1251
10	0,0000	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186	0,1251
11	0,0000	0,0002	0,0019	0,0082	0,0225	0,0452	0,0722	0,0970	0,1137
12	0,0000	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728	0,0948
13	0,0000	0,0000	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504	0,0729
14	0,0000	0,0000	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324	0,0521
15	0,0000	0,0000	0,0000	0,0002	0,0009	0,0033	0,0090	0,0194	0,0347
16	0,0000	0,0000	0,0000	0,0000	0,0003	0,0015	0,0045	0,0109	0,0217
17	0,0000	0,0000	0,0000	0,0000	0,0001	0,0006	0,0021	0,0058	0,0128
18	0,0000	0,0000	0,0000	0,0000	0,0000	0,0002	0,0009	0,0029	0,0071