

НАЦИОНАЛЬНОЕ ЭКОНОМИЧЕСКОЕ ОБРАЗОВАНИЕ

Л. Н. Фадеева, А. В. Лебедев

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Под редакцией доктора экономических наук,
профессора Л. Н. Фадеевой

3-е издание, переработанное и дополненное

Допущено УМО по классическому университетскому образованию
в качестве **учебного пособия** для студентов высших учебных заведений,
обучающихся по направлению 080100 «Экономика»

МОСКВА

ЧИТАЙ!

Рид Групп
2011

УДК 519.2(075.8)
ББК 22.17я73
Ф 15

- Фадеева Л. Н.**
Ф 15 Теория вероятностей и математическая статистика: учеб. пособие / Л. Н. Фадеева, А. В. Лебедев; под ред. д-ра экон. наук, проф. Л. Н. Фадеевой. — М. : Рид Групп, 2011. — 496 с. — (Национальное экономическое образование).

ISBN 978-5-4252-0390-8

Книга представляет собой учебно-методический комплекс, объединяющий теоретический материал, задачи и краткое руководство к разработке методов принятия решения в условиях неопределенности, рекомендации и выводы на основе анализа статистических данных, научно обоснованного прогнозирования случайных явлений и их взаимосвязи, построения математических моделей реальных экономических ситуаций. Книга входит в состав учебного комплекса «Математика для экономистов», специально созданного для экономических вузов страны экономическим факультетом МГУ имени М. В. Ломоносова. Цель данного издания — в удобной для восприятия форме дать студенту-экономисту весь объем математических знаний в части теории вероятностей и математической статистики.

Для студентов и преподавателей экономических специальностей вузов, слушателей послевузовского образования.

УДК 519.2(075.8)
ББК 22.17я73

ISBN 978-5-4252-0390-8

© Л. Н. Фадеева, А. В. Лебедев, 2011
© ООО «Рид Групп», 2011

Оглавление

Предисловие	7

Часть I ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ГЛАВА 1. Элементы комбинаторного анализа	11
1.1. Основные понятия и теоремы комбинаторики.	11
1.2. Разбиение множества на группы.	18
<i>Задачи для самостоятельного решения</i>	<i>19</i>
ГЛАВА 2. Классическая вероятностная модель. Геометрическая вероятность	23
2.1. Частотная интерпретация вероятности. Свойство устойчивости частот	23
2.2. Пространство элементарных исходов. Событие и его вероятность.	25
2.3. Статистики Бозе—Эйнштейна, Ферми—Дирака, Максвелла—Больцмана	32
2.4. Геометрическая вероятность.	33
<i>Задачи для самостоятельного решения</i>	<i>37</i>
ГЛАВА 3. Основные формулы теории вероятностей	41
3.1. Операции над событиями	41
3.2. Теоремы сложения вероятностей	43
3.3. Условная вероятность и теорема умножения.	46
3.4. Независимость событий	48
3.5. Формула полной вероятности	50
3.6. Формула Байеса.	53
3.7. Аксиоматическое построение теории вероятностей.	54
<i>Задачи для самостоятельного решения</i>	<i>58</i>
ГЛАВА 4. Повторные независимые испытания.	65
4.1. Испытания Бернулли	65
4.2. Наивероятнейшее число успехов.	67
4.3. Предельные теоремы и приближенные формулы	70
4.5. Полиномиальные испытания	74
<i>Задачи для самостоятельного решения</i>	<i>76</i>
ГЛАВА 5. Дискретные случайные величины	82
5.1. Случайная величина и ее закон распределения	82
5.2. Функция распределения случайной величины	84
5.3. Случайный вектор в дискретном вероятностном пространстве	86
5.4. Совместная функция распределения случайного вектора.	90

5.5. Числовые характеристики дискретных случайных величин . . .	93
5.6. Основные дискретные распределения и их характеристики . .	96
5.7. Ковариация. Коэффициент корреляции	97
5.8. Условные распределения и математические ожидания (дискретный случай)	102
<i>Задачи для самостоятельного решения</i>	105
ГЛАВА 6. Непрерывные случайные величины	115
6.1. Плотность и функция распределения непрерывной случайной величины	115
6.2. Числовые характеристики непрерывной случайной величины . .	118
6.3. Производящая функция моментов	121
6.4. Примеры непрерывных случайных величин	122
<i>Задачи для самостоятельного решения</i>	135
ГЛАВА 7. Функции от случайных величин. Непрерывный случайный вектор	140
7.1. Функции от случайных величин	140
7.2. Совместный закон распределения непрерывных случайных величин	142
7.3. Плотность суммы двух непрерывных случайных величин . .	147
7.4. Условные распределения и условные математические ожидания непрерывной случайной величины	149
<i>Задачи для самостоятельного решения</i>	152
ГЛАВА 8. Закон больших чисел. Центральная предельная теорема . . .	159
8.1. Неравенство Чебышева	159
8.2. Закон больших чисел	161
8.3. Центральная предельная теорема (ЦПТ)	164
<i>Задачи для самостоятельного решения</i>	167
ГЛАВА 9. Цепи Маркова	172
9.1. Основные определения	172
9.2. Цепи Маркова с конечным числом состояний и дискретным временем	173
9.3. Цепи Маркова с непрерывным временем. Системы массового обслуживания	177
<i>Задачи для самостоятельного решения</i>	181

Часть II МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ГЛАВА 10. Основные понятия и задачи математической статистики . .	187
10.1 Генеральная и выборочная совокупность	187
10.2. Графическое представление статистических рядов	191
10.3. Эмпирические законы распределения	195
<i>Задачи для самостоятельного решения</i>	200

ГЛАВА 11. Точечные оценки параметров законов распределения	205
11.1. Выборочные характеристики и точечные оценки	205
11.2. Статистическая устойчивость основных выборочных характеристик	207
11.3. Асимптотически нормальный характер основных выборочных характеристик	213
11.4. Эффективность оценок. Неравенство Рао—Фреше—Крамера	214
11.5. Оценка математического ожидания по неравноотчным наблюдениям	220
<i>Задачи для самостоятельного решения</i>	<i>222</i>
ГЛАВА 12. Функции и распределения в математической статистике	230
12.1. Бета- и гамма-функции	230
12.2. Квантили, процентные и критические точки	234
12.3. Распределения хи-квадрат (закон Пирсона).	237
12.4. Распределения Стьюдента	240
12.5. Распределение Фишера.	242
12.6. Гамма-распределение.	244
12.7. Бета-распределение	250
12.8. Приложения распределений в математической статистике. Теорема Фишера.	252
<i>Задачи для самостоятельного решения</i>	<i>257</i>
ГЛАВА 13. Методы построения оценок	260
13.1. Метод моментов	260
13.2. Метод максимального правдоподобия	265
13.3. Метод наименьших квадратов. Линейная регрессия.	273
<i>Задачи для самостоятельного решения</i>	<i>280</i>
Глава 14. Доверительные интервалы	292
14.1. Основные понятия	292
14.2. Точные доверительные интервалы	293
14.3. Асимптотические доверительные интервалы	298
14.4. Интервальная оценка коэффициента корреляции.	302
<i>Задачи для самостоятельного решения</i>	<i>304</i>
Глава 15. Проверка статистических гипотез	312
15.1. Основные определения.	312
15.2. Критерий отношения правдоподобия	315
15.3. Проверка гипотез для одной выборки	319
15.4. Проверка гипотез для двух выборок. Зависимые выборки: парные наблюдения	327
15.5. Проверка гипотез для двух выборок. Независимые выборки	328
15.6. Проверка гипотез о равенстве дисперсий для нескольких выборок. Критерий Бартлетта и Кокрена.	336
<i>Задачи для самостоятельного решения</i>	<i>338</i>

Глава 16. Критерии согласия	352
16.6. Критерии согласия Пирсона и Фишера (хи-квадрат)	352
16.2. Критерий согласия Колмогорова	362
<i>Задачи для самостоятельного решения</i>	<i>367</i>
Глава 17. Элементы анализа временных рядов.	372
17.1. Основные понятия в анализе временных рядов.	372
17.2. Простые методы анализа и прогнозирование временных рядов	374
17.3. Стационарность. Автокорреляция. Периодограмма	378
17.4. Модели авторегрессии и скользящего среднего	381
<i>Задачи для самостоятельного решения</i>	<i>384</i>
Глава 18. Элементы линейного регрессионного и корреляционного анализа	388
18.1. Функциональная, статистическая и корреляционная зависимости	388
18.2. Регрессионные модели как инструмент анализа и прогнозирования экономических моделей	391
18.3. Выборочные коэффициенты корреляции	392
<i>Задачи для самостоятельного решения</i>	<i>396</i>
Глава 19. Элементы дисперсионного анализа	399
19.1. Основные понятия и определения	399
<i>Задачи для самостоятельного решения</i>	<i>402</i>

ПРИЛОЖЕНИЯ

<i>Приложение 1. Примеры контрольных и экзаменационных работ</i>	<i>404</i>
<i>Приложение 2. Таблицы</i>	<i>447</i>
Ответы и указания	459
Литература	494

Предисловие

Одна из важнейших сфер приложения теории вероятностей и математической статистики — экономика. В настоящее время для прогнозирования экономических явлений и принятия правильных управленческих решений необходимо хорошо разбираться в вопросах эконометрического моделирования, регрессионного анализа, трендовых и сглаживающих моделей и др.

Курс теории вероятностей и математической статистики входит в цикл фундаментальных дисциплин, изучение которых является обязательным для студентов экономических вузов и факультетов.

Основное достоинство данного издания, подготовленного на основе цикла лекций, читаемых авторами на экономическом факультете Московского государственного университета имени М.В. Ломоносова, состоит в том, что оно соответствует образовательным стандартам третьего поколения и представляет собой учебно-методический комплекс, включающий теоретический материал и большое количество практических заданий, упражнений и задач.

Книга состоит из двух частей: «Теория вероятностей» и «Математическая статистика», каждая из которых разбита на главы, состоящие из отдельных параграфов.

Для лучшего восприятия и усвоения достаточно сложного теоретического материала все определения, утверждения и до-

казательства сопровождаются многочисленными практически-ми примерами. Для закрепления полученных знаний к каждой главе разработаны задачи, большинство из которых построены на конкретном экономическом материале. Упражнения, предлагаемые для самостоятельного решения, дают возможность студенту овладеть основными методами обработки и анализа статистических данных, получить навыки построения экономических моделей вероятностно-статистическими методами, научиться пользоваться статистическими пакетами прикладных программ для проведения необходимых вычислений, уметь интерпретировать результаты анализа и строить на их основе научные прогнозы. Издание содержит варианты контрольных и экзаменационных работ, а также теоретические тесты для самоконтроля полученных знаний.

Часть I

**ТЕОРИЯ
ВЕРОЯТНОСТЕЙ**

Глава 1

ЭЛЕМЕНТЫ КОМБИНАТОРНОГО АНАЛИЗА

Одной из основных задач комбинаторного анализа (комбинаторики) является подсчет числа элементов конечных множеств, заданных каким-либо описательным условием. Для этого разработаны различные формулы и правила.

1.1. Основные понятия и теоремы комбинаторики

Пусть имеется k групп $A_1, A_2, \dots, A_k$, причем i -я группа содержит n_i элементов. Тогда справедливы следующие правила.

Правило умножения (основная формула комбинаторики). Общее число N способов, которыми можно выбрать по одному элементу из каждой группы и расставить их в определенном порядке (т.е. получить упорядоченную совокупность $(a_1, a_2, \dots, a_k)$, где $a_i \in A_i$), равно

$$N = n_1 n_2 \dots n_k.$$

Это правило распространяется и на ситуации, когда новые группы образуются в процессе выбора элементов, если численности этих групп не зависят от того, какие именно элементы были выбраны.

Правило сложения. Если любые две группы A_i и A_j не имеют общих элементов, то выбор одного элемента или из A_1 , или из A_2 , ..., или из A_k можно осуществить $N = n_1 + n_2 + \dots + n_k$ способами.

Задача 1. В группе 30 студентов. Необходимо выбрать старосту, заместителя старосты и профорга. Сколько существует способов это сделать?

Решение. Старостой может быть выбран любой из 30 студентов, заместителем — любой из оставшихся 29, а профоргом — любой из оставшихся 28 студентов, т.е. $n_1 = 30$, $n_2 = 29$, $n_3 = 28$. По правилу умножения общее число N способов выбора старосты, его заместителя и профорга равно $N = n_1 \cdot n_2 \cdot n_3 = 30 \cdot 29 \cdot 28 = 24\,360$.

Задача 2. Два почтальона должны разнести 10 писем по 10 адресам. Сколькими способами они могут распределить работу?

Решение. Первое письмо имеет $n_1 = 2$ альтернативы: либо его относит к адресату первый почтальон, либо второй. Для второго письма также есть $n_2 = 2$ альтернативы и т.д., т.е. $n_1 = n_2 = \dots = n_{10} = 2$. Следовательно, в силу правила умножения общее число способов распределений писем между двумя почтальонами

$$N = n_1 n_2 \dots n_{10} = \underbrace{2 \cdot 2 \cdot \dots \cdot 2}_{10} = 2^{10} = 1024.$$

Задача 3. В ящике 100 деталей, из них 30 деталей 1-го сорта, 50 — 2-го, остальные — 3-го. Сколько существует способов извлечения из ящика одной детали 1-го или 2-го сорта?

Решение. Деталь 1-го сорта может быть извлечена $n_1 = 30$ способами, 2-го сорта — $n_2 = 50$ способами. По правилу суммы существует $N = n_1 + n_2 = 30 + 50 = 80$ способов извлечения одной детали 1-го или 2-го сорта.

Рассмотрим следующие типовые ситуации, которые являются частными случаями общей схемы. Пусть имеется некоторая конечная совокупность элементов $\{a_1, a_2, \dots, a_n\}$, называемая *генеральной совокупностью*, и n — объем этой совокупности. Пусть эксперимент состоит в том, что из генеральной совокупности последовательно выбирают k элементов и располагают их в порядке выбора. Возможны следующие ситуации.

Размещение без повторений, когда отобранный элемент перед отбором следующего не возвращается в генеральную совокупность. Такой выбор называется *последовательным выбором без возвращения*, а его результат — *размещением без повторений из n элементов по k* .

Размещения — это упорядоченные совокупности элементов, отличающиеся друг от друга либо составом, либо порядком элементов.

► **Теорема 1.** Число различных способов, которыми можно произвести последовательный выбор без возвращения k элементов из генеральной совокупности объема n , равно

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

Доказательство. Очевидно, что первый элемент можно выбрать $n_1 = n$ способами, и поскольку отобранный элемент не возвращается в генеральную совокупность, то следующий элемент выбирается из совокупности, объем которой на один элемент меньше, т.е. $n_2 = n - 1$, и т.д., так что $n_k = n - (k - 1)$. Тогда по правилу умножения общее число способов равно $N = n(n - 1) \dots (n - (k - 1)) = n!/(n - k)!$

Пример 1. Все размещения без повторений двух элементов из множества $\{a, b, c\}$: ab, ba, ac, ca, bc, cb .

В частном случае, когда выбирают все элементы генеральной совокупности, т.е. когда $k = n$, размещения называются *перестановками*.

Перестановки — это упорядоченные совокупности, отличающиеся друг от друга только порядком элементов. Число всех перестановок множества из n элементов равно $P_n = n!$

Пример 2. Все перестановки множества $\{a, b, c\}$: $abc, bac, cba, acb, cab, bca$.

Размещения с повторениями, когда отобранный элемент перед отбором следующего возвращается в генеральную совокупность. Такой выбор называется *последовательным выбором с*

возвращением, а его результат — размещением с повторениями из n элементов по k .

► **Теорема 2.** Общее число различных способов, которыми можно произвести выбор с возвращением k элементов из генеральной совокупности объема n , равно $\tilde{A}_n^k = n^k$.

Доказательство. Так как каждый раз отобранный элемент перед отбором следующего возвращается в генеральную совокупность, то выбор на каждом шаге производится из совокупности объема n , и можно считать, что выбор производится из k групп и все группы состоят из одинакового числа элементов $n_1 = n_2 = \dots = n_k = n$. Тогда, в силу основной формулы комбинаторики, число таких способов выбора равно $N = n^k$.

Пример 3. Все размещения с повторениями двух элементов из множества $\{a, b, c\}$: $aa, ab, ac, ba, bb, bc, ca, cb, cc$.

Задача 4. Расписание одного дня состоит из 5 различных уроков. Определите число вариантов расписания при выборе из 11 дисциплин.

Решение. Каждый вариант расписания представляет набор 5 дисциплин из 11, отличающихся от других вариантов как составом, так и порядком следования, поэтому

$$N = A_{11}^5 = \frac{11!}{(11-5)!} = \frac{11!}{6!} = 7 \cdot 8 \cdot 9 \cdot 10 \cdot 11 = 55\,440.$$

Задача 5. Порядок выступления 7 участников конкурса определяется жребием. Сколько различных вариантов жеребьевки при этом возможно?

Решение. Каждый вариант жеребьевки отличается только порядком участников конкурса, т.е. является перестановкой из 7 элементов. Их число равно $P_7 = 7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$.

Задача 6. В конкурсе по 5 номинациям участвуют 10 кинофильмов. Сколько существует вариантов распределения призов, если по каждой номинации установлены различные премии?

Решение. Каждый из вариантов распределения призов представляет собой комбинацию 5 фильмов из 10, отличающуюся от других комбинаций как составом, так и их порядком. Так как каждый фильм может получить призы как по одной, так и по нескольким номинациям, то одни и те же фильмы могут повторяться. Поэтому число таких комбинаций равно числу размещений с повторениями из 10 элементов по 5: $N = \overline{A}_{10}^5 = 10^5 = 100\,000$.

Сочетания без повторений. Пусть из генеральной совокупности берется сразу несколько элементов (либо элементы берут последовательно, но порядок их появления не учитывается). В результате такого *одновременного неупорядоченного выбора* k элементов из генеральной совокупности объема n получаются комбинации, которые называются *сочетаниями без повторений из n элементов по k* .

Сочетаниями C_n^k *из n элементов по k* называются неупорядоченные совокупности, отличающиеся друг от друга хотя бы одним элементом.

► **Теорема 3.** Число сочетаний из n элементов по k равно:

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

Доказательство. Среди A_n^k размещений без повторений имеется по $k!$ наборов каждого состава, представляющих собой всевозможные перестановки из k элементов этого состава.

Поэтому $C_n^k = A_n^k / k!$

Пример 4. Все сочетания без повторений двух элементов из множества $\{a, b, c\}$:

$$\{a, b\}, \{a, c\}, \{b, c\}.$$

Отметим некоторые полезные свойства числа сочетаний без повторений:

$$\begin{aligned}
C_n^0 &= C_n^n = 1; \\
C_n^k &= C_n^{n-k}; \\
C_n^k + C_n^{k+1} &= C_{n+1}^{k+1}; \\
C_n^0 + C_n^1 + C_n^2 + \dots + C_n^n &= 2^n; \\
A_n^k &= P_k C_n^k.
\end{aligned}$$

Числа C_n^k называют также *биномиальными коэффициентами*, поскольку они участвуют в разложении бинома Ньютона:

$$(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}.$$

Сочетания с повторениями. Пусть из генеральной совокупности объема n выбирается k элементов, один за другим, причем каждый отобранный элемент перед отбором следующего возвращается в генеральную совокупность. При этом ведется запись, какие элементы появились и сколько раз, однако порядок их появления не учитывается. Получившиеся совокупности называются *сочетаниями с повторениями из n элементов по k* .

► **Теорема 4.** Число сочетаний с повторениями из n элементов по k равно:

$$\bar{C}_n^k = C_{n+k-1}^k.$$

Пример 5. Все сочетания с повторениями двух элементов из множества $\{a, b, c\}$:

$$\{a, a\}, \{a, b\}, \{a, c\}, \{b, b\}, \{b, c\}, \{c, c\}.$$

Задача 7. В шахматном турнире участвуют 16 человек. Сколько партий должно быть сыграно в турнире, если между любыми участниками должна быть сыграна одна партия?

Решение. Каждая партия играется двумя участниками из 16 и отличается от других только составом пар участников, т.е. представляет собой сочетания без повторений из 16 элементов

по 2. Их число равно: $C_{16}^2 = \frac{16!}{14!2!} = \frac{15 \cdot 16}{1 \cdot 2} = 120$.

Задача 8. В условиях задачи 1.6 определите, сколько существует вариантов распределения призов, если по каждой номинации установлены одинаковые призы.

Решение. Если по каждой номинации установлены одинаковые призы, то порядок фильмов в комбинации 5 призов значения не имеет, и число вариантов представляет собой число сочетаний с повторениями из 10 элементов по 5, определяемое по формуле $\bar{C}_{10}^5 = C_{10+5-1}^5 = C_{14}^5 = \frac{10 \cdot 11 \cdot 12 \cdot 13 \cdot 14}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 2002$.

Задача 9. Садовник должен в течение 3 дней посадить 6 деревьев. Сколькими способами он может распределить по дням работу, если будет сажать не менее одного дерева в день?

Решение. Предположим, что садовник сажает деревья в ряд и может принимать различные решения относительно того, после какого по счету дерева остановиться в первый день и после какого — во второй. Таким образом, можно представить себе, что деревья разделены двумя перегородками, каждая из которых может стоять на одном из 5 мест (между деревьями). Перегородки должны стоять там по одной, поскольку иначе в какой-то день не будет посажено ни одного дерева. Таким образом, надо выбрать 2 элемента из 5 (без повторений). Следовательно, число способов $C_5^2 = 10$.

Задача 10. Сколько существует четырехзначных чисел (возможно, начинающихся с нуля), сумма цифр которых равна 5?

Решение. Представим число 5 в виде суммы последовательных единиц, разделенных на группы перегородками (каждая группа в сумме образует очередную цифру числа). Понятно, что таких перегородок понадобится 3. Мест для перегородок имеется 6 (до всех единиц, между ними и после). Каждое место может занимать одна или несколько перегородок (в последнем случае между ними нет единиц, и соответствующая сумма равна нулю). Рассмотрим эти места в качестве элементов множества. Таким образом, надо выбрать 3 элемента из 6 (с повторениями). Следовательно, искомое количество чисел

$$\bar{C}_6^3 = C_8^3 = \frac{8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3} = 56.$$

1.2. Разбиение множества на группы

Пусть множество из n различных элементов разбивается на k групп так, что в первую группу попадают n_1 элементов, во вторую — n_2 элементов, в k -ю группу — n_k элементов, причем $n_1 + n_2 + \dots + n_k = n$. Такую ситуацию называют *разбиением множества на группы*.

► **Теорема 5.** Число разбиений на k групп, когда в первую группу попадают n_1 элементов, во вторую — n_2 элементов, в k -ю группу — n_k элементов равно:

$$N_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Доказательство. В первую группу могут попасть любые n_1 элементов из имеющихся n элементов первоначально, что можно сделать $C_n^{n_1}$ способами. Вторую группу надо заполнить n_2 элементами из оставшихся $n - n_1$ элементов. Это можно сделать $C_{n-n_1}^{n_2}$ различными способами. Продолжая эту процедуру и используя основную формулу комбинаторики, получаем, что число способов разбиения n элементов по k групп так, что в первую группу попадают n_1 элементов, во вторую — n_2 элементов, в k -ю группу — n_k элементов, равно:

$$\begin{aligned} C_n^{n_1} C_{n-n_1}^{n_2} \dots C_{n-n_1-\dots-n_{k-1}}^{n_k} &= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \dots \times \\ &\times \frac{(n-n_1-\dots-n_{k-1})!}{n_k!0!} = \frac{n!}{n_1! n_2! \dots n_k!}, \end{aligned}$$

т.е. если множество из n элементов разбивается на m групп так, что в первую группу попадает n_1 элементов, во вторую — n_2 элементов, в m -ю группу — n_m элементов, то число таких разбиений равно:

$$N_n(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!}.$$

Пример 6. Перечислим разбиения множества из 4 элементов a, b, c, d на 2 группы по 2 элемента (6 штук): $\{\{a, b\}, \{c, d\}\}$, $\{\{a, c\}, \{b, d\}\}$, $\{\{a, d\}, \{b, c\}\}$, $\{\{c, d\}, \{a, b\}\}$, $\{\{b, d\}, \{a, c\}\}$, $\{\{b, c\}, \{a, d\}\}$.

Задача 11. Сколькими способами можно разбить группу из 25 студентов на три подгруппы A, B и C по 6, 9 и 10 человек соответственно?

Решение. Здесь $n = 25$, $k = 3$, $n_1 = 6$, $n_2 = 9$, $n_3 = 10$. Согласно формуле число таких разбиений $N_{25}(6, 9, 10) = \frac{25!}{6!9!10!}$.

Задача 12. Сколько существует семизначных чисел, состоящих из цифр 4, 5 и 6, в которых цифра 4 повторяется 3 раза, а цифры 5 и 6 — по 2 раза?

Решение. Каждое семизначное число отличается от другого порядком следования цифр, при этом фактически все семь мест в этом числе делятся на три группы: на одни места ставится цифра 4, на другие места — цифра 5, а на третьи места — цифра 6. Таким образом множество состоит из 7 элементов ($n = 7$), причем $n_1 = 3$, $n_2 = 2$, $n_3 = 2$, и, следовательно, количество таких чисел равно:

$$N_7(3; 2; 2) = \frac{7!}{3!2!2!} = 210.$$

Заметим, что порядок элементов при разбиении на группы не важен, а вот порядок групп (какая из них считается первой, какая — второй и т.д.) существен.

Задачи для самостоятельного решения

1. Монету подбросили 3 раза. Сколько различных результатов бросаний можно ожидать?
2. Доступ к файлу открывается, только если введен правильный пароль — определенный трехзначный номер из нечетных цифр. Каково максимальное число возможных попыток угадать пароль?
3. Группу из 10 человек требуется разбить на две непустые подгруппы *A* и *B*. Сколькими способами это можно сделать?
4. Два наборщика должны набрать 16 текстов. Сколькими способами они могут распределить эту работу между собой?
5. Поезд метро делает 16 остановок, на которых выходят пассажиры. Сколькими способами могут распределиться между этими остановками 100 пассажиров, вошедших в поезд на конечной остановке?
6. Шесть студентов-переводников нужно распределить по трем группам. Сколькими способами это можно сделать?
7. Лифт останавливается на 7 этажах. Сколькими способами могут выйти на этих этажах 6 пассажиров, находящихся в кабине лифта?

8. Сколько существует пятизначных чисел, в которых есть цифры 1 и 2 (считаем, что число может начинаться с нуля)?
9. Байт — это машинное слово, состоящее из восьми бит. Каждый бит равен либо 0, либо 1. Сколько символов можно закодировать с помощью байта?
10. Автомобильный номер состоит из трех букв и трех цифр. Сколько различных номеров можно составить, используя 30 букв и 10 цифр?
11. Сколько четырехзначных чисел, составленных из неповторяющихся нечетных цифр, содержит цифру 3?
12. Из цифр от 1 до 9 составляются всевозможные пятизначные числа, не содержащие одинаковых цифр. Определить количество чисел, в которых есть цифры 2, 4 и 5 одновременно.
13. Есть 3 билета в различные театры. Сколькими способами они могут быть распределены среди 25 студентов группы, если каждый студент может получить только один билет?
14. Сколькими способами можно расположить на полке 10 томов энциклопедии?
15. Сколькими способами можно расположить на полке 10 томов энциклопедии так, чтобы девятый и десятый тома не стояли рядом?
16. Шесть групп занимаются в шести расположенных подряд аудиториях. Сколько существует вариантов расписания, при которых группы 1 и 2 находились бы в соседних аудиториях?
17. Сколькими способами можно расставить группу из 10 человек в очередь так, чтобы между двумя студентами *A* и *B* было два человека?
18. Из бригады в 14 врачей ежедневно в течение 7 дней назначают двух дежурных врачей. Определить количество различных расписаний дежурства, если каждый врач дежурит ровно один раз в неделю.
19. Слово «ЭКОНОМИКА» разрезали на буквы, перемешали и выбирают их по одной, без возвращения. Сколько различных последовательностей букв может получиться?
20. Слово «МЕНЕДЖМЕНТ» разрезали на буквы, перемешали и выбирают их по одной, без возвращения. Сколько различных последовательностей букв может получиться?
21. В ящике 5 красных и 4 зеленых яблока. Сколькими способами можно выбрать 3 яблока из ящика?

22. Из ящика, в котором лежат 10 красных и 5 зеленых яблок, выбирают одно красное и два зеленых яблока. Сколькими способами это можно сделать?
23. Десять человек при встрече обмениваются рукопожатиями (каждый с каждым). Сколько всего рукопожатий будет сделано?
24. В группе из 25 студентов нужно выбрать старосту и 3 членов студенческого комитета. Сколькими способами это можно сделать?
25. Акционерное собрание компании выбирает из 50 человек президента компании, председателя совета директоров и 10 членов совета директоров. Сколькими способами это можно сделать?
26. Из фирмы, в которой работают 10 человек, 5 сотрудников должны уехать в командировку. Сколько может быть составов этой группы, если директор фирмы, его заместитель и главный бухгалтер одновременно уезжать не должны?
27. В телевизионной студии работают 3 режиссера, 4 звукорежиссера, 5 операторов, 7 корреспондентов и 2 музыкальных редактора. Сколькими способами можно составить съемочную группу, состоящую из одного режиссера, двух операторов, одного звукорежиссера и двух корреспондентов?
28. Студенческая группа из 27 человек пишет контрольную работу из трех вариантов (по 9 человек в каждом). Сколькими способами можно выбрать 5 человек из группы так, чтобы среди них оказались писавшие все три варианта?
29. На группу из 25 человек выделены 3 пригласительных билета на вечер. Сколькими способами они могут быть распределены (не более одного билета в руки)?
30. Имеются 7 билетов: 3 — в один театр и 4 — в другой. Сколькими способами они могут быть распределены между студентами группы из 25 человек?
31. Садовник должен в течение трех дней посадить 10 одинаковых деревьев. Сколькими способами он может распределить по дням работу, если будет сажать не менее одного дерева в день?
32. Инвестор намеревается купить 12 акций 4 компаний (не менее чем по одной акции каждой). Сколькими способами он может распределить свой капитал?
33. Инвестор намеревается купить 11 акций 5 компаний (причем от покупки акций одной или более из этих компаний он может отказаться). Сколькими способами он может распределить свой капитал?
34. Сколько существует трехзначных чисел (возможно, начинающихся с нуля), сумма цифр которых равна 6?

35. Сколько существует четырехзначных чисел (возможно, начинающихся с нуля), сумма цифр которых равна 9?
36. Группу из 10 человек требуется разбить на две подгруппы так, чтобы в первой группе было 6 человек, а во второй — 4 человека. Сколькими способами это можно сделать?
37. Группу из 16 человек требуется разбить на 3 подгруппы, в первой из которых должно быть 5 человек, во второй — 7 человек, в третьей — 4 человека. Сколькими способами это можно сделать?
38. Студенческую группу из 24 человек (12 девушек и 12 юношей) разбивают на две равные подгруппы так, чтобы в каждой подгруппе юношей и девушек было поровну. Сколькими способами это можно сделать?
39. Семь яблок и три апельсина надо положить в два пакета так, чтобы в каждом пакете был хотя бы один апельсин и чтобы количество фруктов в них было одинаковым. Сколькими способами это можно сделать? Пакеты считаем различными.
40. Лифт, в котором находятся 9 пассажиров, может останавливаться на 10 этажах. На одном этаже выходят 2 человека, на другом — 3 и еще на одном — 4. Сколькими способами пассажиры могут выйти из лифта?
41. Восемь авторов должны написать книгу из 16 глав. Сколькими способами можно распределить материал между авторами, если первые два автора напишут по три главы, следующие четыре — по две и последние два — по одной главе книги?
42. Сколькими способами можно расположить на шахматной доске две ладьи так, чтобы одна не могла взять другую? (Одна ладья может взять другую, если она находится с ней на одной горизонтали или на одной вертикали шахматной доски).
43. Сколько существует двузначных чисел, кратных либо 2, либо 5, либо тому и другому числу одновременно?
44. Десяти ученикам выданы два варианта контрольной работы. Сколькими способами можно посадить учеников в два ряда, чтобы у сидящих рядом не было одинаковых вариантов, а у сидящих друг за другом был один и тот же вариант?
45. Три машины № 1, 2, 3 должны доставить товар в шесть магазинов. Сколькими способами можно использовать машины, если грузоподъемность каждой из них позволяет взять товар сразу для всех магазинов и если две машины в один и тот же магазин не направляются? Сколько вариантов маршрута возможно, если решено использовать только машину № 1?

Глава 2

КЛАССИЧЕСКАЯ ВЕРОЯТНОСТНАЯ МОДЕЛЬ. ГЕОМЕТРИЧЕСКАЯ ВЕРОЯТНОСТЬ

2.1. Частотная интерпретация вероятности. Свойство устойчивости частот

Теория вероятностей — это наука о закономерностях случайных событий. Под случайным событием в теории вероятностей понимается всякое явление, которое может произойти или не произойти при осуществлении определенного комплекса условий. Каждое такое осуществление называется *испытанием*, *опытом* или *экспериментом*.

События можно подразделить на достоверные, невозможные и случайные.

Достоверным называется событие, которое обязательно произойдет при испытании.

Невозможным называется событие, которое заведомо не произойдет при испытании.

Случайным называется событие, которое в результате эксперимента может либо произойти, либо не произойти (в зависимости от случайных обстоятельств).

Такое определение событий можно назвать **эмпирическим**. Более строгие, математические (*теоретико-множественные*) определения будут даны позже.

Предметом теории вероятностей являются закономерности массовых случайных событий, где под массовостью мы понимаем многократную повторяемость.

Рассмотрим несколько событий:

1. A — появление герба при бросании монеты;
2. B — появление трех гербов при трехкратном бросании монеты;
3. C — попадание в цель при выстреле;
4. D — выигрыш по билету денежно-вещевой лотереи.

Очевидно, что каждое из этих событий обладает какой-то степенью возможности. Для того чтобы количественно сравнивать между собой события по степени их возможности, нужно с каждым событием связать определенное число.

Вероятность события есть численная мера степени объективной возможности этого события. В качестве единицы измерения вероятности принята вероятность достоверного события. Вероятность невозможного события равна нулю. Вероятность любого случайного события обозначается $P(A)$ и изменяется в диапазоне от нуля до единицы: $0 \leq P(A) \leq 1$.

Пусть проведена серия из n испытаний (n называют длиной серии), в каждом из которых может произойти или не произойти событие A . Подсчитаем, сколько раз в этой серии эксперимент заканчивался наступлением события A , и обозначим это число через $n(A)$. Поделив его на общее число n всех повторений эксперимента, получим величину $P_n(A) = \frac{n(A)}{n}$, которая называется

относительной частотой события A . При небольшом числе экспериментов относительная частота события носит случайный характер и может заметно меняться от одной группы опытов к другой. При увеличении числа экспериментов случайные обстоятельства, свойственные каждому отдельному эксперименту, в массе взаимно погашаются, и частота $P_n(A)$ проявляет тенденцию стабилизироваться, приближаясь к некоторой средней величине.

Этот эмпирический факт называется **свойством статистической устойчивости частот**: *по мере неограниченного увеличения числа однородных и независимых испытаний относительная частота события A стремится к некоторой постоянной величине*.

Число, к которому приближается относительная частота события при неограниченном увеличении числа экспериментов, можно принять за вероятность события A , т.е. **частотная интерпретация вероятности** состоит в том, что относительную частоту

ту события принимают за приближенное значение вероятности этого же события.

Математически данное свойство записывается в виде предела:

$$\lim_{n \rightarrow \infty} P_n(A) = P(A).$$

Однако приближение частоты события к его вероятности не является обычной сходимостью к пределу. Для описания сходимости частоты события к его вероятности вводится понятие сходимости по вероятности: последовательность X_n сходится по вероятности к величине a , если при любом сколь угодно малом $\varepsilon > 0$ вероятность неравенства $|X_n - a| < \varepsilon$ с увеличением n неограниченно приближается к единице, т.е. $\lim_{n \rightarrow \infty} P(|X_n - a| < \varepsilon) = 1$ и обозначается $X_n \xrightarrow{P} a$. Тогда справедлива

► **Теорема Бернулли.** *При неограниченном увеличении числа однородных и независимых испытаний частота события сходится по вероятности к вероятности появления события в отдельном испытании.*

Частота события A отличается от вероятности этого события тем, что вероятность величина детерминированная, а частота — величина случайная и до опыта неизвестная.

В качестве примера укажем на опыт Бюффона, в котором симметричная монета подбрасывалась 4040 раз, а герб выпадал 2048 раз. Частота появления герба в данной серии наблюдений равна $2048/4040 = 0,507$, что близко к интуитивно ожидаемому значению вероятности 0,5.

К сожалению, частотная интерпретация вероятности несовершенна как с логической, так и с практической точки зрения. Далеко не всегда возможно или желательно провести большое количество экспериментов, а кроме того, существует необходимость в предсказании вероятностей событий, которые еще не происходили. Поэтому далее мы рассмотрим другие определения.

2.2. Пространство элементарных исходов.

Событие и его вероятность

Для того чтобы формально описать некоторый эксперимент¹, нужно указать все возможные варианты результатов, ко-

¹ Под экспериментом имеем в виду не обязательно научный эксперимент, а любое действие или наблюдение либо их последовательность.

торыми этот эксперимент может закончиться. Предполагается, что эксперимент может закончиться одним и только одним исходом.

Множество Ω всех возможных исходов эксперимента называют **пространством элементарных исходов**, а каждый элемент пространства называется **элементарным событием**.

Если все возможные исходы можно перечислить, то пространство элементарных исходов называют **дискретным** (конечным или счетным):

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}, \text{ или } \Omega = \{\omega_1, \omega_2, \dots\}.$$

Пример 1. При бросании симметричной монеты возможны два исхода: выпадение решки или герба. Они удовлетворяют всем трем указанным выше условиям, поэтому в этом случае пространство элементарных исходов представляется в виде $\Omega = \{P, \Gamma\}$, где буквами P и Γ обозначены решка и герб соответственно.

Пример 2. При одновременном бросании двух монет исходы представляют собой упорядоченные пары, состоящих из символов P и Γ . Первый элемент этой пары — результат, выпавший на первой монете, второй элемент — результат на второй монете. Очевидно, что таких пар — четыре:

$$\Omega = \{PP, P\Gamma, \Gamma P, \Gamma\Gamma\}.$$

Пример 3. В случае бросания игральной кости может выпасть любое из чисел 1, 2, 3, 4, 5, 6. Поэтому пространство элементарных исходов $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Пример 4. При одновременном бросании двух игральных костей элементарные исходы представляют собой пары (x, y) , где x — число очков, выпавшее на первой кости, а y — число очков на второй кости. Всего таких пар — 36:

$$\Omega = \{(x, y): x = 1, \dots, 6, y = 1, \dots, 6\}.$$

Событием в случае дискретного пространства элементарных исходов называется любое его подмножество $A \subseteq \Omega$. Говорят, что «событие A произошло», если эксперимент закончился одним из элементарных исходов $\omega \in A$.

Простым событием называется множество, состоящее из одного элементарного исхода. *Сложным событием* называется множество, состоящее из двух или более элементарных исходов.

Введем определение вероятности. Говорят, что в дискретном пространстве элементарных событий $\Omega = \{\omega_1, \omega_2, \dots\}$ задана вероятность, если каждому элементарному событию ω_i поставлено в соответствие неотрицательное число $p_i \geq 0$, называемое его *вероятностью*, такое, что сумма (конечная или бесконечная) вероятностей всех элементарных исходов равна единице:

$$\sum_{i=1}^n p_i = 1 \quad (\text{или} \quad \sum_{i=1}^{\infty} p_i = 1).$$

Вероятностью события A называется сумма вероятностей всех элементарных исходов, входящих в A , т.е. $P(A) = \sum_{\omega_i \in A} P(\omega_i)$. Из этого определения вероятности события следует, что всегда выполняется неравенство $0 \leq P(A) \leq 1$.

Таким образом, **вероятность события** — это числовая функция, определенная на подмножествах пространства элементарных исходов.

Из определения вероятности события следует, что:

1. $P(\Omega) = 1$, где Ω — пространство элементарных исходов;
2. $P(\emptyset) = 0$, если $\emptyset$ — пустое множество.

Простейшим пространством элементарных исходов является так называемая «классическая модель», в которой пространство конечно и все исходы эксперимента:

- 1) равновозможны;
- 2) взаимно несовместны (никакие исходы не могут произойти одновременно);
- 3) все исходы образуют полную группу событий, т.е. никакие другие исходы, кроме перечисленных, не могут произойти.

Такое пространство называют **симметричным**.

Если $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ — симметричное пространство, то вероятности элементарных исходов равны между собой:

$P(\omega_i) = p_i = p$ для любого $i = 1, 2, \dots, n$ и

$$\sum_{i=1}^n P(\omega_i) = \sum_{i=1}^n p_i = np = 1,$$

так что вероятность элементарного исхода равна $p_i = p = \frac{1}{n}$ для любого $i = 1, 2, \dots, n$. Отсюда по определению вероятность события $A = \{\omega_1, \omega_2, \dots, \omega_m\}$ равна $P(A) = \sum_{\omega_i \in A} P(\omega_i) = \frac{1}{n} \cdot m = \frac{|A|}{|\Omega|}$, где

$n = |\Omega|$ — число элементов во множестве Ω , которое обычно называют *общим числом исходов*, а $m = |A|$ — число элементов во множестве A , называемое *числом исходов, благоприятствующих событию A* .

Итак, в случае симметричного пространства вероятности события A определяется как отношение числа случаев, благоприятствующих событию A , к общему числу случаев:

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

Это классическое определение вероятности события.

Событие $\bar{A}$, состоящее из всех элементарных исходов, не входящих в A , называется *противоположным событием* к событию A . Оно происходит тогда и только тогда, когда событие A не произошло. Очевидно, что

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

Это равенство используется для вычисления вероятности события A в случае, когда вероятность противоположного события известна или легко может быть найдена. Тогда $P(A) = 1 - P(\bar{A})$.

Таким образом, для вычисления вероятности в каждой задаче важно определить, в чем состоит эксперимент, правильно построить соответствующее пространство элементарных исходов Ω и выделить в нем требуемое событие A . Затем, используя методы комбинаторики, подсчитать число элементов в Ω и A .

Задача 1. В ящике 5 апельсинов и 4 яблока. Наудачу выбираются 3 фрукта. Какова вероятность, что все три фрукта — апельсины?

Решение. Элементарными исходами здесь являются выборки, включающие 3 фрукта. Поскольку порядок здесь безразличен, будем считать выборки неупорядоченными (и разумеется, неповторными). Общее число элементарных исходов $n = |\Omega|$ равно числу способов выбрать 3 элемента из 9, т.е. числу сочетаний C_9^3 . Число благоприятствующих исходов $m = |A|$ будет равно числу способов выбора трех апельсинов из имеющихся 5, т.е. числу сочетаний трех элементов из 5, или C_5^3 . Тогда искомая вероятность

$$P(A) = \frac{C_5^3}{C_9^3} = \frac{\frac{5!}{2!3!}}{\frac{9!}{6!3!}} = 0,12.$$

Задача 2. Преподаватель предлагает каждому из трех студентов задумать любое число от 1 до 10. Считая, что выбор каждым из студентов любого числа из заданных равновозможен, найти вероятность того, что у какой-то пары из них задуманные числа совпадут.

Решение. Вначале подсчитаем общее количество исходов. Первый из студентов выбирает одно из 10 чисел и имеет $n_1 = 10$ возможностей, второй тоже имеет $n_2 = 10$ возможностей, наконец, третий также имеет $n_3 = 10$ возможностей. В силу основной формулы комбинаторики общее число способов будет равно: $n = n_1 \cdot n_2 \cdot n_3 = 10^3 = 1000$, т.е. все пространство содержит 1000 элементарных исходов.

Подсчет количества благоприятствующих исходов более сложен. Заметим, что совпадение задуманных чисел может произойти у любой пары студентов (или даже одновременно у всех троих). Чтобы не разбирать отдельно все эти случаи, удобно перейти к противоположному событию, т.е. подсчитать количество тех случаев, когда все три студента задумывают разные числа. Первый из них по-прежнему имеет $m_1 = 10$ способов выбора числа. Второй студент имеет теперь лишь $m_2 = 9$ возможностей, поскольку ему приходится заботиться о том, чтобы его число не совпало с задуманным числом первого студента $m_2 \neq m_1$. Третий студент еще более ограничен в выборе — у него всего $m_3 = 8$ возможностей. Из 10 возможных для m_3 исключаются два числа: $m_3 \neq m_1$, $m_3 \neq m_2$. Поэтому общее количество комбинаций задуманных чисел, в которых нет совпадений, равно в силу той же основной формулы $m = 10 \cdot 9 \cdot 8 = 720$. Остальные 280 случаев характеризуются наличием хотя бы одного совпадения. Следовательно, искомая вероятность совпадения равна $P = 280/1000 = 0,28$.

Задача 3. Найти вероятность того, что в 8-значном числе ровно 4 цифры совпадают, а остальные различны.

Решение. Событие $A = \{\text{восьмизначное число содержит 4 одинаковые цифры}\}$. Из условия задачи следует, что в числе пять различных цифр, одна из них повторяется. Число способов ее выбора равно числу способов выбора одной цифры из 10 цифр. Эта цифра занимает любые 4 места в числе, что возможно сделать C_8^4 способами, так как порядок здесь не важен. Оставшиеся 4 места занимают различные цифры из не-

использованных девяти, и так как число зависит от порядка расположения цифр, то число способов выбора четырех цифр равно числу размещений A_9^4 . Тогда число благоприятствующих исходов $|A| = 10C_8^4 A_9^4$. Всего же способов составления 8-значных чисел $|\Omega| = 10^8$. Искомая вероятность

$$P = \frac{|A|}{|\Omega|} = \frac{10C_8^4 A_9^4}{10^8} = \frac{8!}{4!4!} \cdot \frac{9!}{5!} \cdot \frac{1}{10^7} = 0,021168.$$

Задача 4. Шесть клиентов случайным образом обращаются в 5 фирм. Найти вероятность того, что хотя бы в одну фирму никто не обратится.

Решение. Рассмотрим противоположное событие $\bar{A}$, состоящее в том, что в каждую из 5 фирм обратился клиент, тогда в какую-то из них обратились два человека, а в остальные 4 фирмы — по одному клиенту. Таких возможностей $|\bar{A}| = 5 \cdot N_6(2, 1, 1, 1, 1) = \frac{5 \cdot 6!}{1!1!1!1!2!}$. Всего же способов распределить 6 клиентов по 5 фирмам $|\Omega| = 5^6$. Отсюда $P(\bar{A}) = \frac{5 \cdot 6!}{1!1!1!1!2!} \cdot \frac{1}{5^6} = 0,1152$, следовательно, $P(A) = 1 - P(\bar{A}) = 0,8848$.

Задача 5. Среди 25 экзаменационных билетов имеется 5 «счастливых» и 20 «несчастливых». Студенты подходят за билетами один за другим по очереди. У кого больше вероятность вытащить «счастливый» билет: у того, кто подошел первым, или у того, кто подошел вторым?

Решение. Пусть «счастливые» билеты имеют номера 1, 2, 3, 4, 5. Обозначим через i_1 номер билета, взятого первым студентом, через i_2 — номер билета, взятого вторым студентом, тогда элементарным исходом будет пара (i_1, i_2) , а пространство элементарных исходов

$$\Omega = \{(i_1, i_2) : i_1 = 1, \dots, 25, i_2 = 1, \dots, 25, i_1 \neq i_2\},$$

где все элементарные исходы равновероятны. Событие $A = \{\text{первый студент взял «счастливый» билет}\}$ имеет вид

$$A = \{(i_1, i_2) : i_1 = 1, \dots, 5, i_2 = 1, \dots, 25, i_1 \neq i_2\},$$

а событие $B = \{\text{второй студент взял «счастливый» билет}\}$ имеет вид

$$B = \{(i, i_2) : i_1 = 1, \dots, 25, i_2 = 1, \dots, 5, i_1 \neq i_2\}.$$

Каждое из событий A и B содержит $|A| = |B| = C_{25}^1 C_{24}^1 = 120$ элементов, а все пространство Ω имеет $|\Omega| = C_{25}^1 C_{24}^1 = 600$ элементов. Следовательно, $P(A) = P(B) = 1/5$. Вероятность не зависит от того, кто подошел первым, кто вторым и т.п.

Задача 2.6. Пусть в урне имеется N шаров, из них M белых и $N - M$ черных. Из урны извлекается n шаров. Найти вероятность того, что среди них окажется ровно m белых шаров.

Решение. Так как порядок элементов здесь несуществен, число всех возможных наборов объема n из N элементов равно числу сочетаний C_N^n . Число испытаний, которые благоприятствуют событию A — « m белых шаров, $n - m$ черных», равно $C_M^m C_{N-M}^{n-m}$,

и, следовательно, искомая вероятность равна $P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$.

Описанная ситуация представляет собой пример «урновой модели». Говорят также, что случайное число белых шаров в выборке имеет **гипергеометрическое распределение**.

В общем случае предположим, что имеется $N = n_1 + n_2 + \dots + n_k$ различных частиц, причем n_1 частиц первого типа, n_2 — второго типа, ..., n_k — k -го типа. Случайным образом из этих N частиц выбирается m частиц. Найдем вероятность события A , состоящего в том, что среди выбранных окажется ровно $m_1 < n_1$ частиц первого типа, $m_2 < n_2$ — второго типа, ..., $m_k < n_k$ — k -го типа, так что $m = m_1 + m_2 + \dots + m_k$.

Поскольку порядок выбора несуществен, то при определении общего числа исходов и числа благоприятных исходов необходимо пользоваться числом сочетаний. Общее число элементарных исходов равно C_N^m . Далее, m_1 частиц первого типа можно выбрать $C_{n_1}^{m_1}$ способами, m_2 частиц второго типа — $C_{n_2}^{m_2}$ способами, ..., m_k частиц k -го типа — $C_{n_k}^{m_k}$ способами. При этом любой выбор частиц определенного типа комбинируют с любыми выборами частиц остальных типов и, следовательно, число благоприятствующих событию A исходов равно $C_{n_1}^{m_1} C_{n_2}^{m_2} \dots C_{n_k}^{m_k}$. Поэтому вероятность равна:

В статистике Ферми—Дирака n неразличимых частиц распределяются по m ячейкам ($n \leq m$), однако в каждой ячейке не может находиться более одной частицы. Число различных элементарных исходов совпадает с числом способов, которыми можно выбрать n занятых ячеек из общего числа ячеек m , и так как порядок выбора несуществен, то число способов равно C_m^n . Найдем вероятность того, что заняты k фиксированных ячеек. Пусть событие A — заняты фиксированные k ячеек ($k \leq n$). Тогда оставшиеся $m - k$ ячеек должны быть заполнены $n - k$ частицами, а это можно сделать C_{m-k}^{n-k} способами. Поэтому искомая вероятность равна:

$$P(A) = \frac{C_{m-k}^{n-k}}{C_m^n}.$$

Предполагая, что n различных частиц распределяются по m ячейкам без ограничений на число попавших в каждую ячейку частиц, получаем статистику Максвелла—Больцмана. Поскольку каждая из n частиц может попасть в любую из m ячеек, то общее число элементарных исходов равно m^n . Событие A заключается в том, что в первую ячейку попало n_1 частиц, во вторую — n_2 , ..., в m -ю — n_m частиц ($n_1 + n_2 + \dots + n_m = n$). Число благоприятных для события A исходов равно числу разбиений множества из n элементов на m групп так, что в первую группу попадает n_1 элементов, во вторую — n_2 элементов, в m -ю группу — n_m элементов, т.е. числу

$$N_n(n_1, n_2, \dots, n_m) = \frac{n!}{n_1! n_2! \dots n_m!}.$$

Таким образом, искомая вероятность

$$P(A) = \frac{n!}{n_1! n_2! \dots n_m!} \frac{1}{m^n}.$$

Статистика Максвелла—Больцмана представляет собой частный случай так называемой полиномиальной схемы.

2.4. Геометрическая вероятность

Рассмотрим n -мерное вещественное пространство R^n . Пусть в какую-то ограниченную область $\Omega \subset R^n$ наудачу бросили точку. Слово «наудачу» означает, что в таком эксперименте все

точки области Ω «равновозможны». В этом случае вероятность попадания этой точки в какую-то подобласть $A \subseteq \Omega$ определяется формулой

$$P(A) = \frac{V(A)}{V(\Omega)},$$

где $V(A)$ и $V(\Omega)$ — n -мерные объемы областей A и Ω соответственно. Здесь *элементарными исходами* называются точки множества Ω (которое играет роль пространства элементарных исходов), а *благоприятствующими исходами* — точки множества A .

Задача 7. Точку наудачу бросили на отрезок $[0; 2]$. Какова вероятность попадания этой точки на отрезок $[0,5; 1,4]$?

Решение. Здесь пространство элементарных исходов — весь отрезок $\Omega = [0; 2]$, а множество благоприятствующих исходов $A = [0,5; 1,4]$, при этом длины этих отрезков равны $l(\Omega) = 2$ и $l(A) = 0,9$. Поэтому вероятность попадания брошенной точки в указанный отрезок $P(A) = \frac{l(A)}{l(\Omega)} = \frac{0,9}{2} = 0,45$.

Задача 8. На отрезок $[0; 2]$ бросили наудачу и поочередно две точки. Какова вероятность, что первая точка лежит правее второй точки?

Решение. Обозначим получившиеся координаты точек через x и y . Элементарным исходом в таком бросании двух точек будет пара (x, y) , а пространством элементарных исходов — квадрат $\Omega = \{(x, y) : x, y \in [0; 2]\}$. Событие $A = \{\text{первая точка лежит правее второй точки}\}$ равносильно условию $x > y$, следовательно, $A = \{(x, y) : x, y \in [0; 2], x > y\}$, т.е. представляет собой треугольник (рис. 2.2). Площади квадрата и треугольника равны соответственно $S(\Omega) = 4$ и $S(A) = 2$, а потому вероятность $P(A) = \frac{S(A)}{S(\Omega)} = \frac{2}{4} = 0,5$.

Задача 9. Стержень (рис. 2.3) разламывается на две части в случайной точке, равномерно распределенной по длине стержня. Найти вероятность того, что меньший обломок имеет длину, не превосходящую одной трети длины стержня.

Решение. Обозначим длину стержня L , а расстояние точки разлома от одного (например, левого) конца стержня — x .

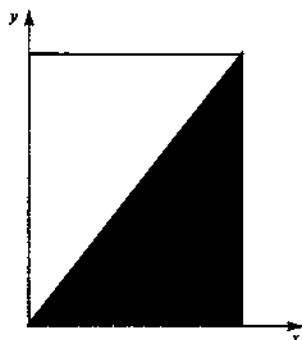


Рис. 2.2

Тогда описанное событие произойдет при условии, если либо $x \leq \frac{L}{3}$, либо $x \geq \frac{2L}{3}$. Искомая вероятность равна отношению

$$P(A) = \frac{\frac{L}{3} + \frac{L}{3}}{L} = \frac{2}{3}.$$

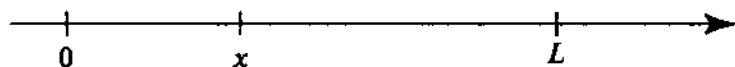


Рис. 2.3

Задача 10 (задача о встрече). Два лица A и B условились встретиться в определенном месте между 12 и 13 часами. Пришедший первым ждет другого в течение 20 минут, после чего уходит. Чему равна вероятность встречи лиц A и B , если приход каждого из них может произойти наудачу в течение указанного часа и моменты прихода независимы?

Решение. Обозначим моменты прихода лица A через x и лица B через y . Для того чтобы встреча произошла, необходимо и достаточно, чтобы $|x - y| \leq 20$. Изобразим x и y как координаты на плоскости, в качестве единицы масштаба выберем минуту. Всевозможные исходы представляются точками квадрата со стороной 60, а благоприятствующие встрече располагаются в заштрихованной области. Искомая вероятность равна отношению площади заштрихованной фигуры (рис. 2.4) к площади всего квадрата: $p = (60^2 - 40^2)/60^2 = 5/9$.

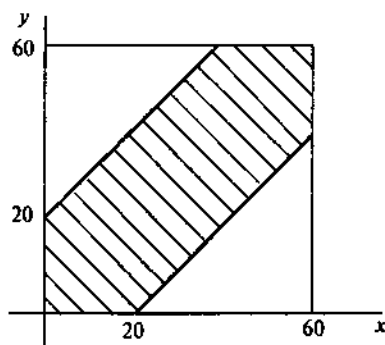


Рис. 2.4

Задача 11 (задача Бюффона). Плоскость разграфлена параллельными прямыми, отстоящими одна от другой на расстоянии $2a$. На плоскость наудачу бросается игла длиной $2l$ ($l < a$). Найти вероятность того, что игла пересечет какую-нибудь прямую.

Решение. Если игла бросается с достаточной высоты и ее начальное положение случайно, то под словом «наудачу» подразумевается, во-первых, что центр иглы наудачу попадет на отрезок длиной $2a$, во-вторых, что угол φ между прямой и иглой равномерно распределен на отрезке $[0; \pi]$, и в-третьих, что на величину угла не влияет расстояние от центра до прямой. Поэтому изобразим результат бросания точкой с координатами (φ, x) , лежащей внутри прямоугольника со сторонами a и π , где x — расстояние от центра иглы до ближайшей прямой. Из рис. 2.5,а видно, что пересечение иглы с прямой происходит тогда и только тогда, когда $x < l \sin \varphi$. Искомая вероятность равна отношению площади заштрихованной области A к площади прямоугольника на рис. 2.5,б:

$$P(A) = \frac{1}{\pi a} \int_0^{\pi} l \sin \varphi d\varphi = \frac{2l}{\pi a}.$$

Отметим, что полученную формулу можно применить для приближенного вычисления числа π . Действительно, получаем: $\pi = 2l/(aP(A))$. Проводя многократные эксперименты (бросания иглы), можно приблизить вероятность $P(A)$ относительной частотой $P_n(A)$ и соответственно найти приближенное значение $\pi_n = 2l/(aP_n(A))$. Подобное вычисление детерминированных ве-

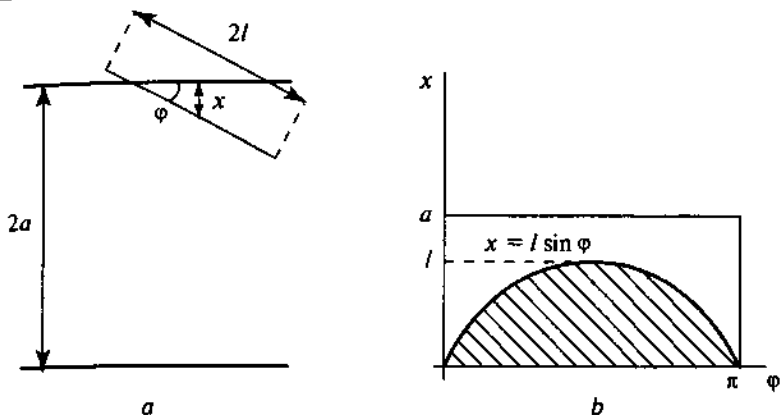


Рис. 2.5

личин с помощью последовательности испытаний со случайными исходами называется **методом Монте-Карло**. Разумеется, в современных исследованиях для этого используется компьютер.

Задачи для самостоятельного решения

1. Бросают две игральные кости. Чему равна вероятность того, что сумма очков, выпавших на обеих костях, не превзойдет 5?
2. Какова вероятность того, что в 4 бросания кости хотя бы один раз выпадет «единица»?
3. Четыре человека вошли в лифт на первом этаже шестизэтажного дома. Найти вероятности следующих событий: а) все пассажиры выйдут на шестом этаже; б) все пассажиры выйдут на одном и том же этаже; в) все пассажиры выйдут на разных этажах.
4. Семь человек вошли в лифт на первом этаже восьмиэтажного дома. Какова вероятность, что хотя бы на одном этаже вышли по крайней мере два человека?
5. Лифт начинает движение с 7 пассажирами и останавливается на 10 этажах. Найти вероятность того, что три пассажира вышли на одном этаже, два — на другом этаже и еще два — на еще одном этаже.
6. Пять яблок раскладываются в четыре ящика. Какова вероятность, что в двух ящиках будет по два яблока, в одном — одно яблоко и один ящик будет пустой?
7. Шесть шаров случайным образом раскладываются по 3 ящикам. Найти вероятность того, что в первом ящике лежит 4 шара.

8. Пять клиентов случайным образом обращаются в 5 фирм. Найти вероятность того, что ровно в одну фирму никто не обратится.
9. Шесть клиентов обращаются в 3 фирмы равновероятно. Найти вероятность того, что ровно в одну фирму обратятся 2 клиента.
10. Шесть клиентов обращаются в 4 фирмы равновероятно. Найти вероятность того, что в какие-то две фирмы обратятся по два клиента и в какие-то две — по одному.
11. N клиентов обращается в M фирм равновероятно. Найти вероятность того, что хотя бы в одну фирму никто не обратится, если: а) $N = 5$, $M = 4$; б) $N = 6$, $M = 4$; в) $N = 7$, $M = 5$.
12. N клиентов обращаются в M фирм равновероятно. Найти вероятность того, что во все фирмы обратятся разное число клиентов (включая, возможно, ноль), если а) $N = 6$, $M = 3$; б) $N = 7$, $M = 3$; в) $N = 7$, $M = 4$.
13. В каждой упаковке товара фирмы «Икс» имеется одна из букв «И», «К», «С» (равновероятно). Какова вероятность собрать все буквы, купив 5 упаковок товара?
14. В каждой из упаковок товара равновероятно встречаются буквы «П», «Р», «И», «З». Найти вероятность того, что можно будет собрать слово «ПРИЗ», купив а) 6 упаковок товара; б) 7 упаковок товара.
15. В каждой упаковке имеется одна из 5 различных наклеек (равновероятно). Какова вероятность собрать их все, купив 7 упаковок товара?
16. Найти вероятность того, что в шестизначном номере три цифры совпадают, а остальные различны (считаем, что номера могут начинаться с нуля).
17. Найти вероятность того, что в восьмизначном числе ровно три цифры совпадают, а остальные различные (считаем, что числа могут начинаться с нуля).
18. Найти вероятность того, что в пятизначном числе имеются 2 четные цифры и 3 нечетные, при условии, что все они различны (считаем, что число может начинаться с нуля).
19. Семь человек становятся случайным образом в очередь один за другим. Какова вероятность того, что два определенных человека, A и B , встанут рядом?
20. В очередь в булочную случайным образом встали восемь женщин и двое мужчин. Какова вероятность того, что между мужчинами будут стоять две женщины?
21. В очередь в кассу стоят 9 человек (трое мужчин, четыре женщины и двое детей). Какова вероятность, что между некоторыми двумя мужчинами будут стоять двое детей и одна женщина?

22. В очереди стоят 4 женщины и 3 мужчины. Какова вероятность, что все мужчины стоят рядом?
23. В урне 5 белых и 4 черных шара. Из урны наугад вынимают два шара. Какова вероятность того, что это будут: а) два белых шара; б) два черных шара; в) один черный и один белый.
24. В партии из 8 изделий имеется 3 изделия высшего качества. Найти вероятность того, что среди отобранных (без возвращения) 4 изделий — ровно одно изделие высшего качества.
25. В лотерее из 50 билетов 5 выигрышных. Какова вероятность того, что среди 5 купленных выбранных билетов ровно 2 будут выигрышными?
26. В лотерее из N билетов имеется M выигрышных. Найти вероятность того, что среди K купленных билетов окажется хотя бы один выигрышный, если: а) $N = 50$, $M = 4$, $K = 10$; б) $N = 60$, $M = 5$, $K = 10$.
27. Из N проданных за неделю телевизоров M имеют скрытые дефекты. Найти вероятность того, что среди случайно выбранных K телевизоров (из числа проданных за неделю) окажется ровно L без скрытых дефектов, если: а) $N = 10$, $M = 4$, $K = 5$, $L = 2$; б) $N = 16$, $M = 4$, $K = 6$, $L = 4$.
28. Из N проданных за неделю автомобилей M имеют скрытые дефекты. Найти вероятность того, что среди случайно выбранных K автомобилей (из числа проданных за неделю) окажется не более L со скрытыми дефектами, если: а) $N = 12$, $M = 4$, $K = 7$, $L = 1$; б) $N = 15$, $M = 3$, $K = 5$, $L = 1$; в) $N = 11$, $M = 5$, $K = 7$, $L = 2$; г) $N = 13$, $M = 4$, $K = 7$, $L = 2$.
29. Группа из N студентов пишет контрольную работу из M вариантов (одинаковое число студентов в каждом варианте). Найти вероятность того, что среди случайно выбранных K студентов есть писавшие каждый вариант, если а) $N = 18$, $M = 3$, $K = 5$; б) $N = 20$, $M = 4$, $K = 5$; в) $N = 24$, $M = 4$, $K = 5$; г) $N = 32$, $M = 4$, $K = 6$.
30. В трех студенческих группах 72 человека (по 24 человека в группе: 12 юношей и 12 девушек). Наудачу выбраны 5 человек. Какова вероятность, что среди них будут девушки из всех трех групп?
31. На группу из 10 человек предоставлено для производственной практики 6 мест в лаборатории № 1 и 4 места — в лаборатории № 2. Какова вероятность того, что при случайном распределении мест двое неразлучных друзей из этой группы попадут на практику в одну лабораторию?
32. Работа каждого из четырех студентов заочного отделения может проверяться одним из четырех преподавателей. Какова вероятность, что все четыре работы проверены разными преподавателями?

33. В пакете с леденцами лежит 4 красных, 6 желтых и 10 зеленых конфет. Найти вероятность вынуть наудачу подряд 3 конфеты одного цвета.
34. В урне K белых шаров, L красных и M черных. Найти вероятность достать (без возвращения) три шара, не все цвета которых одинаковы, если а) $K = 9$, $L = 7$, $M = 4$; б) $K = 10$, $L = 8$, $M = 6$.
35. В ящике находятся 5 белых, 3 красных и 2 черных шара. Наудачу выбирают 6 шаров. Найти вероятность того, что выборка будет содержать 3 белых, 2 красных и 1 черный шар, если: а) выборка производится без возвращения (все 6 шаров отбираются сразу); б) выборка производится с возвращением (фиксируется цвет выбранного шара, после чего он возвращается в ящик).
36. В точке C , любое положение которой на телефонной линии AB длиной 10 км равновозможно, произошел разрыв. Определить вероятность того, что точка C удалена от точки A , где находится ремонтная станция, на расстояние, не меньшее 1 км.
37. Какова вероятность, что дуэль состоится, если каждый из дуэлянтов приходит на место дуэли в случайный момент времени между 5 и 6 часами и ждет противника в течение 5 минут?
38. Две подружки договорились встретиться в условленном месте в промежутке от 17 до 19 часов. Пришедшая первой ждет другую не более 15 минут. Какова вероятность, что подружки не встретятся?
39. На отрезок $[2, 5]$ наудачу бросаются две точки. Какова вероятность, что расстояние между ними меньше 2?
40. На отрезок $[-1, 2]$ наудачу брошены две точки. Какова вероятность, что расстояние между ними больше 1?
41. Два теплохода должны подойти к одному и тому же причалу. Время прихода обоих теплоходов независимо и равновозможно в течение данных суток. Найти вероятность того, что ни одному из теплоходов не придется ожидать освобождения причала, если время стоянки первого теплохода 1 час, а второго — 2 часа.
42. Студент может добраться до факультета либо на автобусе, интервал движения которого составляет 7 минут, либо на троллейбусе, интервал движения которого составляет 10 минут. Найти вероятность того, что студенту, пришедшему на остановку в случайный момент времени, придется ждать не более трех минут.
43. На шахматную доску случайным образом поставлены две ладьи. Какова вероятность того, что они не будут бить друг друга?
44. Наудачу взяты два положительных числа X и Y , каждое из которых не превышает единицы. Найти вероятность того, что сумма $X + Y$ не превышает 1, а произведение XY не меньше 0,09.
45. Найти вероятность того, что из трех наудачу взятых отрезков длиной не более L можно построить треугольник.

Глава 3

ОСНОВНЫЕ ФОРМУЛЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

3.1. Операции над событиями

Одной из основных задач теории вероятностей является вычисление вероятностей сложных событий, когда известны вероятности каких-то других событий. Это возможно, если эти новые события можно выразить через исходные с помощью различных операций.

Суммой (или объединением) двух событий A и B называется событие $A \cup B$ ($A + B$), заключающееся в том, что произойдет хотя бы одно из событий A или B (либо событие A , либо событие B , либо A и B одновременно).

Произведением (или пересечением) двух событий A и B называется событие $A \cap B$ (или AB), состоящее в одновременном появлении и события A , и события B .

Отрицанием (или противоположным событием) для события A называется событие $\bar{A}$, которое происходит тогда и только тогда, когда не происходит событие A .

Симметрической разностью событий A и B называется событие $C = A \nabla B$, в которое входят те элементарные события, которые входят или в A , или в B , но не входят в их пересечение: $A \nabla B = (A \setminus B) \cup (B \setminus A)$.

Поскольку все события рассматриваются как подмножества пространства элементарных исходов Ω , то и операции над ними — это соответствующие операции над множествами (объединение, пересечение, дополнение). Все пространство Ω соответствует *достоверному событию* (поскольку эксперимент всегда

заканчивается каким-то элементарным исходом), а пустое множество $\emptyset$ — невозможному событию (поскольку в нем нет ни одного возможного исхода).

Справедливы следующие соотношения:

$$1. A \cup \Omega = \Omega, \quad A \cup \emptyset = A, \quad A \cup A = A.$$

$$2. A \cap \Omega = A, \quad A \cap \emptyset = \emptyset, \quad A \cap A = A.$$

$$3. \overline{\overline{A}} = A, \quad \overline{\overline{\Omega}} = \emptyset, \quad \overline{\overline{\emptyset}} = \Omega.$$

4. $\overline{A \cap B} = \overline{A} \cup \overline{B}$, $\overline{A \cup B} = \overline{A} \cap \overline{B}$ (принцип двойственности, или формулы де Моргана).

5. $A \cup B = B \cup A$, $A \cap B = B \cap A$ (коммутативность операций объединения и пересечения).

6. $A \cup (B \cap C) = (A \cup B) \cap C$, $A \cap (B \cup C) = (A \cap B) \cup C$ (ассоциативность операций объединения и пересечения).

7. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (дистрибутивность операции объединения относительно пересечения).

8. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (дистрибутивность операции пересечения относительно объединения).

Для наглядности соотношений между событиями используют графическую модель, называемую **диаграммой Вьенна** (рис. 3.1).

Пример 1. Бросают две игральные кости. Пусть A — событие, состоящее в том, что сумма очков нечетная, B — событие, заключающееся в том, что хотя бы на одной из костей выпала двойка. Опишем события $A \cup B$ и $A \cap B$.

Пространство элементарных исходов может быть представлено в виде: $\Omega = \{(1, 1), (1, 2), \dots, (2, 1), \dots, (6, 6)\}$; $|\Omega| = 36$.

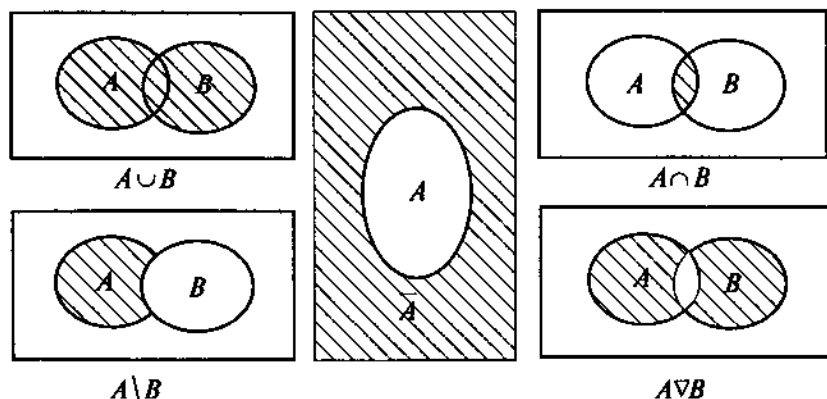


Рис. 3.1

Согласно условию задачи события A и B состоят из следующих элементарных исходов:

$$A = \{(1, 2), (1, 4), (1, 6), (2, 1), (2, 3), \dots, (6, 1), (6, 3), (6, 5)\};$$

$$B = \{(1, 2), (2, 1), (2, 3), (3, 2), (2, 4), (4, 2), (5, 2), (2, 5), (2, 6), (6, 2), (2, 2)\}.$$

Объединение $A \cup B$ представляет собой событие, состоящее в наступлении хотя бы одного из событий A и B , т.е. событие $A \cup B$ означает, что либо сумма выпавших очков нечетна, либо на одной из костей выпала двойка:

$$A \cup B = \{(1, 2), (1, 4), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \dots, (6, 1), (6, 3), (6, 5)\}.$$

Пересечение $A \cap B$ представляет собой событие, состоящее в одновременном наступлении событий A и B , т.е. в том, что на одной из костей выпала двойка, а сумма очков — нечетное число:

$$A \cap B = \{(1, 2), (2, 1), (2, 3), (3, 2), (2, 5), (5, 2)\}.$$

События A и B называются **несовместными (непересекающимися)**, если они не могут произойти одновременно: $A \cap B = \emptyset$.

События $A_1, A_2, \dots, A_n$ образуют **полную группу событий**, если они несовместны и в сумме образуют все пространство Ω , т.е.

$$\sum_{i=1}^n A_i = \Omega, \quad A_i \cap A_j = \emptyset, \quad i \neq j.$$

Это означает, что в результате эксперимента обязательно произойдет одно из данных событий, и только одно.

Пример 2. События A и $\bar{A}$ несовместны и образуют полную группу.

3.2. Теоремы сложения вероятностей

Пусть заданы вероятности некоторых событий и требуется найти вероятности их объединения.

► **Теорема 1 (теорема сложения вероятностей несовместных событий).** Вероятность объединения двух несовместных событий равна сумме их вероятностей, т.е. если $A \cap B = \emptyset$, то

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

Доказательство проведем для случая конечного числа исходов. Пусть пространство элементарных событий $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ содержит n элементарных исходов, из них n_1 благоприятствуют событию A и n_2 благоприятствуют событию B и нет исходов, благоприятствующих одновременно A и B , так как события несовместны, тогда $P(A) = \sum_{i: \omega_i \in A} P(\omega_i)$ и $P(B) = \sum_{k: \omega_k \in B} P(\omega_k)$. Отсюда следует, что событию $A \cup B$ благоприятствуют $n_1 + n_2$ исходов и вероятность этого события вычисляется по формуле

$$P(A \cup B) = \sum_{\omega_j \in A \cup B} P(\omega_j) = \sum_{\omega_i \in A} P(\omega_i) + \sum_{\omega_k \in B} P(\omega_k) = P(A) + P(B).$$

Доказательство без труда переносится на случай счетного пространства $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$, когда вместо конечных сумм рассматриваются суммы со счетным числом слагаемых — сходящиеся ряды.

Следствия

1. Методом математической индукции эту теорему можно распространить на любое конечное число слагаемых, т.е. если все события A_i несовместны, то

$$P\left(\bigcup_{i=1}^n A_i\right) = P(A_1) + P(A_2) + \dots + P(A_n).$$

2. Если события $A_1, A_2, \dots, A_n$ образуют полную группу событий, то сумма их вероятностей равна единице. В частности, поскольку противоположные события несовместны и в сумме образуют Ω , отсюда следует формула $P(A) = 1 - P(\bar{A})$.
3. Если $A \subset B$, то $P(A) \leq P(B)$.

► **Теорема 2 (теорема сложения вероятностей произвольных событий).** Для любых событий A и B верно равенство

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Доказательство. Для любых событий A и B событие $A \cup B$ наступит тогда, когда наступит одно из несовместных событий $A \cap B$, $A \cap \bar{B}$ или $\bar{A} \cap B$. По теореме сложения для несовместных событий:

$$P(A \cup B) = P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B).$$

Событие A наступит, если наступит хотя бы одно из двух несовместных событий: $A \cap B$ или $A \cap \bar{B}$. Тогда вероятность события A по теореме сложения для несовместных событий равна $P(A) = P(A \cap B) + P(A \cap \bar{B})$. Аналогично событие B наступит, если наступит хотя бы одно из несовместных событий $A \cap B$ или $\bar{A} \cap B$, и вероятность события B равна $P(B) = P(A \cap B) + P(\bar{A} \cap B)$. Отсюда получаем:

$$\begin{aligned} P(A \cup B) &= P(A) + P(\bar{A} \cap B) + P(A \cap B) - P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Следствия

1. Вероятность пересечения двух событий A и B вычисляется по формуле $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.
2. Вероятность суммы любого числа событий вычисляется по формуле включения-исключения:

$$P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i A_j) + \dots + (-1)^{n-1} P(A_1 \dots A_n).$$

3. Из теорем 1 и 2 для любых событий A и B следует, что $P(A \cup B) \leq P(A) + P(B)$.

Задача 1. В ящике 10 красных и 5 синих пуговиц. Вынимаются наудачу две пуговицы. Какова вероятность, что пуговицы будут одноцветными?

Решение. Событие $A = \{\text{вынуты пуговицы одного цвета}\}$ можно представить в виде суммы $A = A_1 + A_2$, где события A_1 и A_2 означают выбор пуговиц красного и синего цвета соответственно.

но. Вероятность вытащить две красные пуговицы $P(A_1) = \frac{C_{10}^2}{C_{15}^2}$,

а вероятность вытащить две синие пуговицы $P(A_2) = \frac{C_5^2}{C_{15}^2}$. Так

как события A_1 и A_2 не могут произойти одновременно, то в силу теоремы сложения

$$P(A) = \frac{C_{10}^2 + C_5^2}{C_{15}^2} = \frac{\frac{10!}{2!8!} + \frac{5!}{2!3!}}{\frac{15!}{2!13!}} = 0,524.$$

Задача 2. Среди сотрудников фирмы 28% знают английский язык, 30% — немецкий, 42% — французский; английский и немецкий — 8%, английский и французский — 10%, немецкий и французский — 5%, все три языка — 3%. Найти вероятность того, что случайно выбранный сотрудник фирмы: а) знает английский или немецкий; б) знает английский, немецкий или французский; в) не знает ни один из перечисленных языков.

Решение. Обозначим через A , B и C события, заключающиеся в том, что случайно выбранный сотрудник фирмы владеет английским, немецким или французским соответственно. Очевидно, доли сотрудников фирмы, владеющих теми или иными языками, определяют вероятности этих событий. Получаем:

- а) $P(A \cup B) = P(A) + P(B) - P(AB) = 0,28 + 0,3 - 0,08 = 0,5$;
 б) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - (P(AB) + P(AC) + P(BC)) + P(ABC) = 0,28 + 0,3 + 0,42 - (0,08 + 0,1 + 0,05) + 0,03 = 0,8$;
 в) $1 - P(A \cup B \cup C) = 0,2$.

3.3. Условная вероятность и теорема умножения

Говорить о вероятности $P(A)$ как о мере возможности появления случайного события A имеет смысл только при осуществлении определенного комплекса условий эксперимента, в рамках которого событие может произойти. При изменении условий эксперимента, вообще говоря, изменится и вероятность события A . Поэтому помимо обычной (безусловной) вероятности события A рассматривают так называемую *условную вероятность* события A , вычисляемую при условии, что произошло некоторое событие B .

Условной вероятностью события A при условии, что произошло событие B ($P(B) > 0$), называется число $P(A|B)$, которое вычисляется по формуле

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Аналогично определяется условная вероятность события B :

$$P(B|A) = \frac{P(AB)}{P(A)}, \quad P(A) > 0.$$

Из определения условной вероятности вытекает следующая теорема.

► **Теорема 3 (теорема умножения).** Вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое имело место:

$$P(AB) = P(B)P(A|B), \text{ или } P(AB) = P(A)P(B|A).$$

Несмотря на тривиальность доказательства этой теоремы, она имеет огромное практическое значение, так как используется для построения сложных вероятностных моделей.

Следствие. Вероятность пересечения нескольких событий равна произведению вероятностей этих событий, причем вероятность каждого последующего вычисляется при условии, что имели место все предыдущие, т.е. верно равенство

$$P\left[\bigcap_{i=1}^n A_i\right] = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_1A_2\dots A_{n-1}).$$

Для условной вероятности выполняются следующие свойства:

- 1) $0 \leq P(A|B) \leq 1$;
- 2) $P(\bar{A}|B) = 1 - P(A|B)$;
- 3) $P(A|A) = 1$;
- 4) $P((A_1 \cup A_2)|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$;
- 5) если $A \cap B = \emptyset$, то $P(A|B) = 0$;
- 6) если $A \subset B$, то $P(B|A) = 1$.

В случае, когда имеем дело с пространством равновероятных исходов, условную вероятность можно найти по формуле

$$P(A|B) = \frac{|AB|}{|B|}.$$

Задача 3. В семье двое детей. Какова вероятность того, что старший ребенок — мальчик, если известно, что в семье есть дети обоего пола?

Решение. Пусть $A = \{\text{старший ребенок — мальчик}\}$, $B = \{\text{в семье есть дети обоего пола}\}$. Будем считать, что рождение мальчика и рождение девочки — равновероятные события. Если рождение мальчика обозначить буквой M , а рождение девоч-

ки — D , то пространство всех элементарных исходов состоит из четырех пар: $\Omega = \{MM, MD, DM, DD\}$. В этом пространстве лишь два исхода (MD и DM) отвечают событию B . Событие AB означает, что в семье есть дети обоего пола. Старший ребенок — мальчик, следовательно, второй (младший) ребенок — девочка. Этому событию AB отвечает один исход — MD . Таким образом, $|AB| = 1$, $|B| = 2$ и

$$P(A|B) = \frac{|AB|}{|B|} = \frac{1}{2} = 0,5.$$

Задача 4. Мастер, имея 10 деталей, из которых 3 — нестандартные, берет и проверяет детали одну за другой, пока ему не попадется стандартная. Какова вероятность, что он проверит ровно две детали?

Решение. Событие $A = \{\text{мастер проверил ровно две детали}\}$ означает, что при такой проверке первая деталь оказалась нестандартной, а вторая — стандартной. Значит, $A = A_1A_2$, где $A_1 = \{\text{первая деталь оказалась нестандартной}\}$ и $A_2 = \{\text{вторая деталь — стандартная}\}$. Очевидно, что вероятность события A_1 равна $P(A_1) = 3/10$, кроме того, $P(A_2|A_1) = 7/9$, так как перед взятием второй детали у мастера осталось 9 деталей, из которых только 2 нестандартные и 7 стандартных. По теореме умножения

$$P(A) = P(A_1A_2) = P(A_1)P(A_2|A_1) = \frac{3}{10} \cdot \frac{7}{9} = 7/30.$$

3.4. Независимость событий

Событие A не зависит от события B , если появление B не меняет значения вероятности события A , т.е. условная вероятность равна безусловной: $P(A|B) = P(A)$. Тогда из теоремы умножения $P(A|B) = P(A)P(B|A) = P(B)P(A|B)$ в предположении, что $P(A) > 0$, получим, что условная вероятность события B равна его безусловной вероятности $P(B|A) = P(B)$. Следовательно, события A и B независимы, если появление одного из них не меняет вероятности появления другого.

Из определения независимости двух событий следует:

► **Теорема 4.** Два события A и B независимы, если справедливо равенство

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

Это равенство используется как критерий при практической проверке независимости двух событий.

Понятие независимости обобщается на любое конечное число событий.

События $A_1, A_2, \dots, A_n$ называются **независимыми в совокупности**, если для любого их набора $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ имеет место равенство

$$P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

для любых k от 1 до n и любых несовпадающих номеров $i_1, i_2, \dots, i_k$.

События $A_1, \dots, A_n$ называются **попарно независимыми**, если для любых $i \neq j$, $i, j \in \{1, \dots, n\}$ события A_i и A_j независимы.

Из данных определений следует, что из независимости в совокупности следует попарная независимость, но из попарной независимости не следует независимости в совокупности.

Пример Бернштейна. На плоскость бросается правильный тетраэдр, три грани которого покрашены в цвета: красный, синий и зеленый, а на четвертую грань нанесены все три цвета. Событие A — при бросании выпала красная грань, событие B — синяя грань, событие C — зеленая грань. Вероятности этих событий равны между собой:

$$P(A) = P(B) = P(C) = 2/4 = 1/2.$$

Найдем вероятности их попарных произведений:

$$P(AB) = 1/4 = 1/2 \cdot 1/2 = P(A)P(B);$$

$$P(AC) = P(A)P(C) = 1/4 = 1/2 \cdot 1/2;$$

$$P(BC) = 1/4 = 1/2 \cdot 1/2 = P(B)P(C).$$

Отсюда следует, что они попарно независимы. Однако вероятность появления всех трех цветов равна:

$$P(ABC) = 1/4 \neq P(A)P(B)P(C) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8,$$

т.е. вероятность произведения всех трех событий не равна произведению вероятностей этих событий и, следовательно, они зависимы в совокупности.

Задача 5. В одном ящике 3 белых и 5 черных шаров, в другом ящике — 6 белых и 4 черных шара. Найти вероятность того, что хотя бы из одного ящика будет вынут один белый шар, если из каждого ящика вынута по одному шару.

Решение. Событие $A = \{\text{хотя бы из одного ящика вынут белый шар}\}$ можно представить в виде суммы $A = A_1 + A_2$, где события A_1 и A_2 означают появление белого шара из первого и второго ящика соответственно. Вероятность вытащить белый шар из первого ящика равна $P(A_1) = 3/8$, а вероятность вытащить белый шар из второго ящика $P(A_2) = 6/10$. Кроме того, в силу независимости A_1 и A_2 имеем:

$$P(A_1 A_2) = P(A_1)P(A_2) = \frac{3}{8} \cdot \frac{6}{10} = 9/40.$$

По теореме сложения получаем:

$$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2) = 3/8 + 6/10 - 9/40 = 3/4.$$

3.5. Формула полной вероятности

Пусть событие A может быть реализовано только при условии появления одного из событий H_i , $i = 1, \dots, n$. Предположим, что события H_i образуют полную группу, и вероятности их до опыта известны. Такие события H_i называются *гипотезами*.

► **Теорема 5 (формула полной вероятности).** Вероятность события A вычисляется по формуле

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

Доказательство. Поскольку гипотезы H_i , $i = 1, 2, \dots, k$, несовместны, то несовместны и их пересечения с событием A , т.е. несовместны комбинации $A \cap H_i$ и $A \cap H_j$ при $i \neq j$. Обозначим $H = \bigcup_{i=1}^n H_i$. Так как $H_1, \dots, H_k$ образуют полную группу событий, то $H = \Omega$ и $A \cap H = A$.

$$P(A) = P\left(A \cap \left(\bigcup_{i=1}^n H_i\right)\right) = P\left(\bigcup_{i=1}^n (A \cap H_i)\right) = \sum_{i=1}^n P(A \cap H_i).$$

Применяя теорему умножения вероятностей для каждого слагаемого, получаем окончательный результат:

$$P(A) = \sum_{i=1}^n P(A \cap H_i) = \sum_{i=1}^n P(H_i)P(A|H_i).$$

Задача 6. Три экзаменатора принимают экзамен по некоторому предмету у группы в 30 человек, причем первый опрашивает 6 студентов, второй — 3 студентов, а третий — 21 студента (выбор студентов производится случайным образом из списка). Отношение трех экзаменаторов к слабо подготовившимся различное: шансы таких студентов сдать экзамен у первого преподавателя равны 40%, у второго — только 10%, у третьего — 70%. Найти вероятность того, что слабо подготовившийся студент сдаст экзамен.

Решение. Обозначим через H_1, H_2, H_3 гипотезы, состоящие в том, что слабо подготовившийся студент отвечал первому, второму и третьему экзаменатору соответственно. По условию задачи

$$P(H_1) = 6/30 = 0,2; \quad P(H_2) = 3/30 = 0,1; \quad P(H_3) = 21/30 = 0,7.$$

Пусть событие $A = \{\text{слабо подготовившийся студент сдал экзамен}\}$. Тогда в силу условия задачи

$$P(A|H_1) = 0,4, \quad P(A|H_2) = 0,1, \quad P(A|H_3) = 0,7.$$

По формуле полной вероятности получаем

$$P(A) = 0,4 \cdot 0,2 + 0,1 \cdot 0,1 + 0,7 \cdot 0,7 = 0,58.$$

Для решения задач такого типа удобно использовать так называемое «дерево» вероятностей. Из формулы полной вероятности следует, что для вычисления вероятности события A необходимо осуществить перебор всех путей, ведущих к результирующему событию A ; вычислить и расставить на соответствующих путях вероятности $P(H_i)$ того, что движение будет происходить по данному пути, и условные вероятности $P(A|H_i)$ того, что на данном пути будет достигнуто конечное событие A . Затем вероятности, стоящие на одном пути, перемножаются, а результаты, полученные для различных путей, складываются.

Каждое из условий может, в свою очередь, делиться на несколько дополнительных условий или гипотез, т.е. на каждом

этапе оно допускает неограниченное число ветвлений схемы. Поэтому при решении задач удобнее пользоваться не самой формулой полной вероятности, а графической схемой полной вероятности, которую называют «деревом» вероятностей (см. рис. 3.2 для задачи 6).

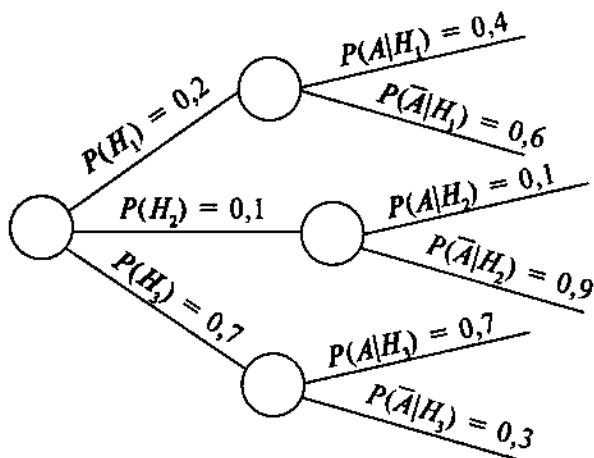


Рис. 3.2

Задача 7. Фирма имеет три источника поставки комплектующих — фирмы A , B , C . На долю фирмы A приходится 50% общего объема поставок, B — 30% и C — 20%. Из практики известно, что среди поставляемых фирмой A деталей 10% бракованных, фирмой B — 5% и фирмой C — 6%. Какова вероятность, что взятая наугад деталь окажется годной?

Решение. Пусть событие G — появление годной детали. Вероятности гипотез о том, что деталь поставлена фирмами A , B , C , равны соответственно $P(A) = 0,5$; $P(B) = 0,3$; $P(C) = 0,2$ (рис. 3.3). Условные вероятности появления при этом годной детали равны $P(G|A) = 0,9$, $P(G|B) = 0,95$, $P(G|C) = 0,94$ (как вероятности противоположных событий к появлению бракованной). По формуле полной вероятности, используя «дерево» вероятностей, получаем:

$$P(G) = 0,5 \cdot 0,9 + 0,3 \cdot 0,95 + 0,2 \cdot 0,94 = 0,923.$$

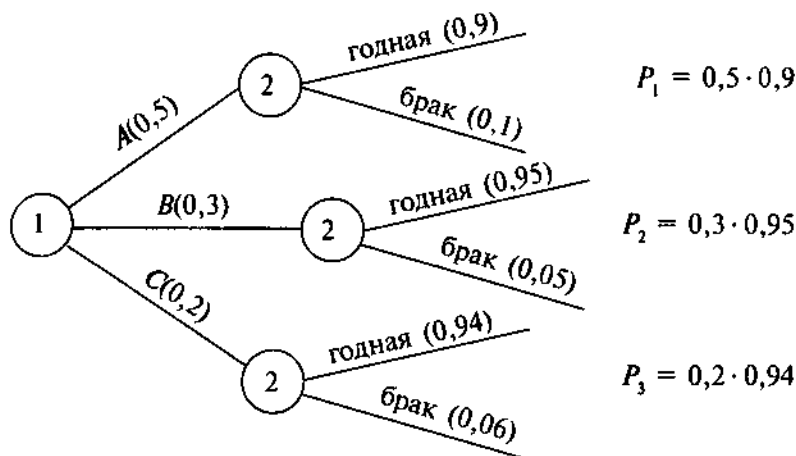


Рис. 3.3

3.6. Формула Байеса

Предположим другую ситуацию: пусть известно, что событие A произошло. Требуется найти вероятность того, что событие A произошло именно путем H_k . Эти условные вероятности вычисляются с помощью следующей теоремы.

► **Теорема 6 (формула Байеса):**

$$P(H_k | A) = \frac{P(H_k) \cdot P(A | H_k)}{\sum_{i=1}^n P(H_i) \cdot P(A | H_i)}.$$

Отметим, что в знаменателе этой формулы записана вероятность $P(A)$, вычисленная по формуле полной вероятности.

Доказательство. По определению условной вероятности

$$P(H_k | A) = \frac{P(A \cap H_k)}{P(A)} = \frac{P(H_k)P(A | H_k)}{P(A)},$$

где вероятность $P(A \cap H_k)$ найдена с помощью теоремы умножения.

Задача 8 (см. задачу 6). Пусть известно, что студент не сдал экзамен, т.е. получил оценку «неудовлетворительно». Кому из трех преподавателей вероятнее всего он отвечал?

Решение. Вероятность получить «неуд» равна $P(\bar{A}) = 1 - P(A) = 1 - 0,58 = 0,42$. Требуется вычислить условные вероятности $P(H_i | \bar{A})$, $i = 1, 2, 3$. По формуле Байеса получаем:

$$P(H_1 | \bar{A}) = \frac{P(H_1) \cdot P(\bar{A} | H_1)}{P(\bar{A})} = \frac{0,2 \cdot 0,6}{0,42} = 0,285,$$

аналогично

$$P(H_2 | \bar{A}) = \frac{0,1 \cdot 0,9}{0,42} = 0,214, \quad P(H_3 | \bar{A}) = \frac{0,7 \cdot 0,3}{0,42} = 0,5.$$

Отсюда следует, что, вероятнее всего, слабо подготовившийся студент сдавал экзамен третьему экзаменатору.

3.7. Аксиоматическое построение теории вероятностей

Ограниченность «классического» определения вероятностей заложена в предположении равной возможности исходов. Многие реальные случайные эксперименты не укладываются также в рамки дискретной модели с конечным или счетным пространством Ω . Не всегда может помочь и геометрическая интерпретация. Кроме того, возникают различные парадоксы. Например, можно построить так называемые *неизмеримые* множества, при попытке определить вероятность которых мы приходим к противоречию.

Пример Вители. Пусть точку наудачу бросают на окружность. Положение точки на окружности определяется углом (от 0 до 2π). Выберем иррациональное число $\alpha > 0$ и поставим в соответствие каждой точке x класс точек A_x , получаемых из нее поворотами на угол $2\pi n\alpha$, $n = 0, 1, 2, \dots$. Поскольку α иррационально, у нас никогда не получится целое число оборотов, а значит, все точки класса различны.

Для различных точек x классы A_x могут либо совпадать (когда одна точка переходит в другую поворотом вида $2\pi n\alpha$), либо не пересекаться нигде. Возьмем все непересекающиеся классы, выберем из каждого по одной точке и объединим эти точки в одно множество B_0 . Обозначим через B_n множество, получающе-

еся из B_0 поворотом на угол $2\pi n$. Тогда множества B_n , $n = 0, 1, 2, \dots$, не пересекаются, а их объединение дает всю окружность. Следовательно, они образуют полную группу событий. Понятно, что вероятности B_n равны между собой. Если предположить, что они равны нулю, то в сумме получается нуль, хотя должна быть единица. Если предположить, что они больше нуля, в сумме получается бесконечность, что также неверно.

Поэтому в случае бесконечного пространства Ω построение современной теории вероятностей базируется на подходе, предложенном великим русским математиком А.Н. Колмогоровым. Его основная идея заключается в том, что не все подмножества пространства Ω рассматриваются как события. Предполагается, что события — это некоторые подмножества из пространства элементарных событий, совокупность которых замкнута относительно операций конечного или счетного числа объединений и пересечений.

Пусть Ω — произвольное пространство элементарных исходов, а $\mathfrak{Z}$ — некоторый класс подмножеств множества Ω .

Алгеброй событий $\mathfrak{Z}$ называется любая непустая система подмножеств пространства Ω , удовлетворяющая следующим аксиомам:

1) если подмножество A принадлежит $\mathfrak{Z}$ (является событием), то его дополнение $\bar{A}$ также принадлежит $\mathfrak{Z}$ (также является событием);

2) если подмножества A и B принадлежат $\mathfrak{Z}$ (являются событиями), то и их объединение $A \cup B$ принадлежит $\mathfrak{Z}$ (также является событием).

Поскольку любую из операций над подмножествами можно получить, используя формулы де Моргана, с помощью только двух операций дополнения и объединения

$$A \cap B = \overline{\bar{A} \cup \bar{B}}, \quad A \setminus B = A \cap \bar{B} = \overline{\bar{A} \cup B},$$

пересечение и разность двух событий также будут событиями: $A \cap B \in \mathfrak{Z}$, $A \setminus B \in \mathfrak{Z}$ при любых $A \in \mathfrak{Z}$, $B \in \mathfrak{Z}$. Отсюда следует, что $\Omega \in \mathfrak{Z}$ и $\emptyset = \Omega \setminus \Omega \in \mathfrak{Z}$ тоже события.

Алгебра событий $\mathfrak{Z}$ называется **σ -алгеброй**, если объединение счетного числа элементов из $\mathfrak{Z}$ также является элементом $\mathfrak{Z}$, т.е.

из того, что $A_n \in \mathfrak{Z}$, $n = 1, 2, \dots$, следует $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{Z}$.

Таким образом, σ -алгебру событий $\mathfrak{Z}$ можно определить как систему подмножеств пространства элементарных исходов

Ω , замкнутую относительно счетного числа теоретико-множественных операций. Тривиальная σ -алгебра событий состоит из полного и пустого множеств: $\mathfrak{Z} = \{\emptyset, \Omega\}$. Любая σ -алгебра событий является одновременно и алгеброй событий. Обратное, вообще говоря, неверно, т.е. существуют алгебры событий, не являющиеся σ -алгебрами.

Теперь согласно аксиоматике Колмогорова можно ввести общее понятие вероятности события.

Вероятностью события, или **вероятностной мерой**, называется числовая функция, заданная на σ -алгебре событий $\mathfrak{Z}$, которая каждому событию $A \in \mathfrak{Z}$ ставит в соответствие число $P(A)$ так, что выполняются следующие четыре аксиомы:

1. **Аксиома неотрицательности:** $P(A) \geq 0$ для всех $A \in \mathfrak{Z}$.

2. **Аксиома нормированности:** $P(\Omega) = 1$.

3. **Аксиома конечной аддитивности:** $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$ для любых $A_i \cap A_j = \emptyset$, $i \neq j$, и $A_i \in \mathfrak{Z}$ для любого $i = 1, 2, \dots, n$.

4. **Аксиома счетной аддитивности:** $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, если события A_i в последовательности $A_1, A_2, \dots, A_n, \dots$ несовместны, т.е. для любых $A_i \cap A_j = \emptyset$, $i \neq j$, $A_i \in \mathfrak{Z}$ для любого $i = 1, 2, \dots$

Очевидно, что вероятность, определенная в дискретном вероятностном пространстве условием $P(A) = \sum_{\omega \in A} P(\omega)$, является счетно-аддитивной.

Заметим также, что введенные аксиомы в случае дискретного пространства превращаются в доказуемые утверждения.

Конечно-аддитивная вероятность $P(A_n)$, заданная на σ -алгебре множеств $\mathfrak{Z}$, называется *непрерывной*, если для любой убывающей последовательности множеств $A_n \subseteq \mathfrak{Z}$, $n = 1, 2, \dots$, такой, что $A_{n+1} \subseteq A_n$, имеющих пустое пересечение $\bigcap_{n=1}^{\infty} A_n = \emptyset$, имеет место равенство $\lim_{n \rightarrow \infty} P(A_n) = 0$.

Аксиома непрерывности

Если последовательность событий $A_1, A_2, \dots, A_n, \dots$ такова, что каждое последующее вложено в предыдущее, а пересечение всех событий A_n пусто, то $P(A_n) \rightarrow 0$ при $n \rightarrow \infty$.

► **Теорема 7.** *Аксиома счетной аддитивности равносильна аксиоме непрерывности.*

Доказательство. Докажем сначала, что из непрерывности следует счетная аддитивность. Рассмотрим последовательность несовместных событий B_n , и пусть $B = \bigcup_{i=1}^{\infty} B_i$. Введем события вида $A_n = \bigcup_{i=n+1}^{\infty} B_i$, тогда они удовлетворяют условиям аксиомы непрерывности и $P(A_n) \rightarrow 0$ при $n \rightarrow \infty$. По теореме сложения получаем

$$P(B) = P\left(\bigcup_{i=1}^n B_i\right) + P\left(\bigcup_{i=n+1}^{\infty} B_i\right) = \sum_{i=1}^n P(B_i) + P(A_n),$$

откуда

$$\sum_{i=1}^n P(B_i) = P(B) - P(A_n) \rightarrow P(B) = P\left(\bigcup_{i=1}^{\infty} B_i\right), \quad n \rightarrow \infty,$$

что и означает счетную аддитивность.

Докажем теперь, что из счетной аддитивности следует непрерывность. Пусть последовательность A_n удовлетворяет условиям аксиомы непрерывности. Введем события $C_n = A_n \setminus A_{n+1}$. Тогда эти события несовместны, причем допустимы представления:

$$A_1 \setminus A_n = \bigcup_{i=1}^{n-1} C_i, \quad A_1 = \bigcup_{i=1}^{\infty} C_i.$$

По теореме сложения получаем

$$P(A_1) = P(A_1 \setminus A_n) + P(A_n) = P\left(\bigcup_{i=1}^{n-1} C_i\right) + P(A_n) = \sum_{i=1}^{n-1} P(C_i) + P(A_n),$$

откуда

$$P(A_n) = P(A_1) - \sum_{i=1}^{n-1} P(C_i) = P\left(\bigcup_{i=1}^{\infty} C_i\right) - \sum_{i=1}^{n-1} P(C_i) \rightarrow 0, \quad n \rightarrow \infty,$$

так что аксиома непрерывности выполняется.

Тройка $(\Omega, \mathfrak{Z}, P)$, где Ω — пространство элементарных событий, $\mathfrak{Z}$ — σ -алгебра подмножеств Ω , называемых событиями, P — вероятностная мера, определенная на событиях, называется *вероятностным пространством*.

Далее будем всюду неявно предполагать, что любые рассматриваемые множества относятся к некоторой σ -алгебре $\mathfrak{Z}$, а вероятность удовлетворяет всем необходимым аксиомам.

Задачи для самостоятельного решения

1. Рабочий обслуживает три независимо работающих станка. Событие $A_i = \{i\text{-й станок в течение часа потребует наладки}\}$, $P(A_i) = 0,2$; $i = 1, 2, 3$. Выразить события: а) ровно два станка потребуют наладки; б) не более двух потребуют наладки; в) хотя бы один потребует наладки. Найти вероятность события в).
2. Стрелок делает три выстрела, при этом он поражает цель с вероятностью 0,6 при одном выстреле. Событие $A_i = \{i\text{-я пуля попала в цель}\}$, $i = 1, 2, 3$. Выразить события: а) было хотя бы одно попадание; б) ровно одно попадание; в) не более двух попаданий. Найти вероятность события в).
3. В коробке четыре детали. Мастер извлекает детали до тех пор, пока не вытащит годную (или пока они не кончатся). Событие $A_i = \{i\text{-я извлеченная деталь является годной}\}$, $P(A_i) = 0,9$, $i = 1, 2, 3, 4$. Выразить события, состоящие в том, что мастер сделал: а) ровно одно извлечение; б) ровно два извлечения; в) не менее двух извлечений. Найти вероятность б).
4. Сколько раз надо бросить игральную кость, чтобы не менее чем на 95% быть уверенным в том, что хотя бы при одном бросании появится «шестерка»?
5. Бросают три кубика. Какова вероятность того, что хотя бы на одном из них выпадет «шестерка», если известно, что на всех кубиках выпали разные грани?
6. В пакете с леденцами лежат 4 красных, 5 желтых и 6 зеленых конфет. Найти вероятность вынуть наудачу подряд 3 конфеты одного цвета.
7. В партии из 20 изделий 4 бракованных. Найти вероятность того, что в выборке из 5 изделий не более одного бракованного.
8. Известно, что в пятизначном числе все цифры разные. Какова вероятность при этом условии, что среди них ровно одна цифра четная (считаем, что номер может начинаться с нуля)?
9. Известно, что в пятизначном числе все цифры разные. Найти вероятность того, что среди них есть цифры 1 и 2 (считаем, что число может начинаться с нуля).
10. На обувной фабрике в отдельных цехах производят подметки, каблуки и верхи ботинок. При изготовлении каждого ботинка соединяют подметку, каблук и верх, выбирая их случайным образом. Дефекты имеют 1% каблуков, 2% подметок и 3% верхов. С какой вероятностью случайно выбранная пара обуви будет иметь дефекты?
11. Среди клиентов туристической фирмы 32% ездили в Турцию, 18% — в Египет, 10% — в Турцию и Египет. Найти вероят-

ность того, что случайно выбранный клиент ездил в Турцию или Египет.

12. Среди клиентов туристической фирмы 30% ездили в Турцию, 20% — в Египет, 10% — в Грецию; в Турцию и Египет — 12%, в Египет и Грецию — 5%, в Турцию и Грецию — 6%, во все три страны — 4%. Найти вероятность того, что случайно выбранный клиент: а) ездил в Турцию или Египет, б) ездил в Египет или Грецию, в) ездил в Турцию, Египет или Грецию, г) не ездил ни в одну из перечисленных стран.
13. В продукции птицефабрики 70% яиц стандартных, 20% большего объема и 10% двухжелтковых. С какой вероятностью среди 5 случайно выбранных яиц найдутся хотя бы одно большего объема и хотя бы одно двухжелтковое (вместе)?
14. Среди покупателей магазина 60% женщин, 30% мужчин, 10% детей. Найти вероятность того, что среди 4 случайно выбранных покупателей найдется хотя бы один мужчина и хотя бы один ребенок.
15. Студент в состоянии решить 20 задач из 25 в первом туре экзамена и 15 из 20 во втором. Найти вероятность сдачи им экзамена, если в каждом туре дается 3 задачи и достаточно решить хотя бы 2 из них.
16. Студент в состоянии решить 25 задач из 30 в первом туре экзамена и 18 из 24 во втором. Найти вероятность сдачи им экзамена, если в каждом туре дается четыре задачи и достаточно решить три из них.
17. В лифт девятиэтажного дома на первом этаже входят 6 человек. Для каждого человека равновероятен выход на любом из остальных 8 этажей. Известно, что все вышли на разных этажах. При этом условии найти вероятность, что на первых трех этажах (из восьми) вышли два человека.
18. В лифт на цокольном этаже входят 5 человек. Считая для каждого человека равновероятным выход на любом из 9 этажей, найти вероятность того, что двое из них выйдут на одном этаже, а остальные на разных.
19. Три пассажира садятся в поезд, случайно выбирая любой из 6 вагонов. Какова вероятность, что хотя бы один из них сядет в первый вагон, если известно, что все они сели в разные вагоны?
20. Семь пассажиров случайным образом выбирают один из 9 вагонов поезда. Известно, что они сели в разные вагоны. При этом условии найти вероятность того, что в первых трех вагонах поезда будут ехать два человека.

21. Пять человек случайным образом (независимо друг от друга) выбирают любой из 7 вагонов поезда. Известно, что ровно 2 вагона остались пустыми. При этом условии найти вероятность того, что первый и второй вагоны заняты.
22. Двое шахматистов равной силы играют 4 партии (без ничьих). Найти вероятность, что в результате победил первый, если известно, что в процессе игры каждый выиграл хотя бы один раз.
23. В ящике 12 красных, 8 зеленых и 10 синих шаров. Наудачу вынимаются два шара. Какова вероятность, что вынутые шары разного цвета, если известно, что не вынут синий шар?
24. Шесть шаров случайным образом раскладывают в три ящика. Найти вероятность, что во всех ящиках разное число шаров при условии, что все ящики не пустые.
25. Пять шаров распределены по трем ящикам. Известно, что нет пустых ящиков. При этом условии найти вероятность, что в первом ящике лежит один шар.
26. В урне 5 белых и 10 черных шаров. Извлечены 6 шаров (с возвращением). Известно, что среди них есть белые шары. При этом условии найти вероятность того, что среди них будут также не менее двух черных шаров.
27. В магазине было проведено исследование продаж некоторого товара. Выяснилось, что этот товар покупают 25% женщин, 10% мужчин и 20% детей. Среди покупателей магазина 60% женщин, 30% мужчин и 10% детей. Найти вероятность того, что случайный покупатель приобретет этот товар.
28. В центральную бухгалтерию корпорации поступили пачки накладных для проверки и обработки. 90% пачек были признаны удовлетворительными: они содержали 1% неправильно оформленных накладных. Остальные 10% накладных были признаны неудовлетворительными, так как они содержали 5% неправильно оформленных накладных. Какова вероятность того, что взятая наугад накладная окажется неправильно оформленной?
29. Фирма занимается строительством домов по одному из двух типовых проектов. При строительстве по первому проекту нарушение технологий происходит с вероятностью 0,3, а по второму — 0,2. При этом дома первого и второго типа составляют соответственно 40% и 60% общего объема строительства. Какова вероятность того, что случайно выбранный дом построен с нарушением технологии?
30. Фирма имеет трех поставщиков, каждый из которых надежен с вероятностью 0,8. В случае отказа одного поставщика фирма разоряется с вероятностью r_1 , двух — r_2 , трех — r_3 . Найти ве-

роятность разорения фирмы, если: а) $r_1 = 0,2$, $r_2 = 0,6$, $r_3 = 1$; б) $r_1 = 0,25$, $r_2 = 0,75$, $r_3 = 0,95$.

31. Фирма участвует в четырех проектах, каждый из которых может закончиться неудачей с вероятностью 0,1. В случае неудачи одного проекта вероятность разорения фирмы равна 20%, двух — 50%, трех — 70%, четырех — 90%. Найти вероятность разорения фирмы.
32. Два аудитора проверяют N фирм (равное число каждый), у M из которых имеются нарушения. Вероятность обнаружения нарушений первым аудитором равна p_1 , вторым — p_2 . Найти вероятность того, что все фирмы-нарушители будут выявлены, если: а) $N = 6$, $M = 2$, $p_1 = 0,7$, $p_2 = 0,8$; б) $N = 6$, $M = 3$, $p_1 = 0,7$, $p_2 = 0,9$; в) $N = 8$, $M = 3$, $p_1 = 0,6$, $p_2 = 0,8$; г) $N = 10$, $M = 2$, $p_1 = 0,8$, $p_2 = 0,9$.
33. В прибор входит комплект из двух независимых деталей, вероятность для которых выйти из строя в течение года соответственно равна 0,1 и 0,2. Если детали исправны, то прибор работает в течение года с вероятностью 0,99. Если выходит из строя только первая деталь, то прибор работает с вероятностью 0,7, а если только вторая, то с вероятностью 0,8. Если выходят из строя обе детали, прибор будет работать с вероятностью 0,1. Какова вероятность, что прибор будет работать в течение года?
34. Электроэнергия поступает в город через три электролинии, каждая из которых может быть отключена с вероятностью 0,1. Если отключена одна электролиния, город испытывает недостаток электроэнергии с вероятностью 0,2. Если отключены две электролинии, недостаток электроэнергии ощущается с вероятностью 0,5. Если же отключены все три электролинии, то недостаток электроэнергии наступает с вероятностью 1. В случае, когда работают все электролинии, недостатка энергии нет. Какова вероятность того, что город испытывает недостаток электроэнергии?
35. Из 10 лотерейных билетов 3 выигрышных. При подготовке вечера 2 билета потеряли, и было решено добавить 1 выигрышный. Какой стала вероятность вытянуть выигрышный билет?
36. В урне 8 красных шаров и 10 зеленых. Извлекли три шара (без возвращения), затем положили их обратно и добавили в урну два шара того цвета, который оказался в меньшинстве среди извлеченных. С какой вероятностью два шара, извлеченных после этого, будут зелеными?
37. В урне 7 красных шаров и 9 синих. Извлекли три шара (без возвращения), затем положили их обратно и добавили в урну два шара того цвета, который оказался в большинстве среди

извлеченных. С какой вероятностью два шара, вынутых после этого, окажутся синими?

38. В одной коробке 4 красных шара и 6 синих, а во второй 8 красных и 2 синих. Из первой во вторую переложили два шара, а затем извлекли из второй два шара без возвращения. Найти вероятность, что последние оказались одного цвета.
39. В первой урне лежат 1 белый и 3 черных шара, а во второй урне — 2 белых и 1 черный шар. Из первой урны во вторую перекладывается, не глядя, один шар, а затем один шар перекладывается из второй урны в первую. После этого из первой урны вынули один шар. Найти вероятность, что он белый.
40. В первой урне лежат 4 белых и 6 черных шаров, во второй — 2 белых и 8 черных. Из первой урны во вторую перекладывается, не глядя, один шар, а затем один шар перекладывается из второй урны в первую. После этого из первой урны вынули один шар. Найти вероятность, что он черный.
41. Есть две упаковки орешков, в каждой из которых 5 орехов с белой глазурью и 4 с черной. Из первой упаковки достали два орешка, после чего ее смешали со второй упаковкой. Какой стала вероятность достать орех с белой глазурью?
42. Фирма нарушает закон с вероятностью p . Аудитор обнаруживает нарушения с вероятностью r (если они есть). Проведенная им проверка не выявила нарушений. Найти вероятность, что на самом деле они есть, при: а) $p = 0,25$, $r = 0,75$; б) $p = 0,1$, $r = 0,8$; в) $p = 0,2$, $r = 0,9$.
43. Известно, что проверяемая фирма может уходить от налогов с вероятностью 40% и выбрать для этого одну из трех схем уклонения (равновероятно). Найти вероятность, что фирма уходит от налогов по третьей схеме, если при проверке по первым двум схемам нарушений не обнаружено.
44. Изделие имеет скрытые дефекты с вероятностью d . В течение года выходит из строя доля p изделий со скрытыми дефектами и r изделий без дефектов. Найти вероятность, что изделие имело скрытые дефекты, если оно вышло из строя в течение года, при: а) $d = 0,2$, $p = 0,75$, $r = 0,15$; б) $d = 0,1$, $p = 0,6$, $r = 0,15$.
45. В цехе работают N мастеров и M учеников (каждый работник изготавливает одинаковое количество изделий). Вероятность брака мастера — p_1 , вероятность брака ученика — p_2 . Изделие оказалось бракованным. Найти вероятность, что его изготовил ученик, если: а) $N = 7$, $M = 3$, $p_1 = 0,01$, $p_2 = 0,05$; б) $N = 10$, $M = 4$, $p_1 = 0,01$, $p_2 = 0,04$.

46. Производственный брак составляет 4%. Каждое изделие равновероятным образом поступает к одному из двух контролеров, первый из которых обнаруживает брак с вероятностью 0,92, второй — 0,98. Какова вероятность, что признанное годным изделие является бракованным?
47. Имеются три партии по 20 деталей в каждой. Число стандартных деталей в первой, второй и третьей партиях соответственно равно 20, 15 и 10. Из наудачу выбранной партии извлечена деталь, оказавшаяся стандартной. Деталь вернули в партию и вторично из той же партии наугад извлекли деталь, которая оказалась стандартной. Найти вероятность того, что детали были извлечены из третьей партии.
48. На заводе установлена аварийная сигнализация, которая при наличии аварии срабатывает с вероятностью p . Однако с вероятностью r сигнал также может возникнуть, когда аварии нет. Вероятность аварии равна s . Найти вероятность того, что случилась авария, если сигнализация сработала, при: а) $p = 0,99$, $r = 0,001$, $s = 0,005$; б) $p = 0,99$, $r = 0,002$, $s = 0,007$.
49. К системному администратору обращаются пользователи. Среди них доля начинающих — p , опытных — q . Вероятность того, что начинающий пользователь обратится за помощью — r , что обратится опытный — s . Найти вероятность того, что очередной пользователь, обратившийся за помощью, окажется начинающим, если: а) $p = 0,6$, $q = 0,4$, $r = 0,8$, $s = 0,1$; б) $p = 0,6$, $q = 0,4$, $r = 0,85$, $s = 0,15$; в) $p = 0,7$, $q = 0,3$, $r = 0,8$, $s = 0,1$.
50. Мимо бензоколонки проезжают легковые и грузовые машины. Среди них доля грузовых машин — p . Вероятность того, что проезжающая машина подъедет заправиться, для грузовой машины равна 0,1; для легковой — 0,2. Найти вероятность того, что очередная машина, подъехавшая на заправку, окажется грузовой, если: а) $p = 0,75$; б) $p = 0,7$.
51. Фирма A занимает 20% рынка электронной техники, фирма B — 50%, фирма B — 30%. Доля мобильных телефонов в поставках фирмы A составляет 20%, в поставках фирмы B — 40%, в поставках фирмы B — 70%. Покупатель приобрел мобильный телефон. Какова вероятность того, что этот телефон произведен фирмой A ?
52. В город поступают товары трех фирм в соотношении 2:3:5. В поставках первой фирмы 60% товара высшего качества, второй — 40%, а третьей — 20%. Куплен товар высшего качества. Найти вероятность, что он изготовлен второй фирмой.
53. В город поступают товары трех фирм в количественном соотношении 3:4:6. В поставках первой фирмы 30% товара высшего

- качества, второй — 20%, а третьей — 25%. Куплен товар высшего качества. Найти вероятность, что он изготовлен второй или третьей фирмой.
54. В компании 70% менеджеров работают в центральном офисе, 30% — в региональных. Вероятность того, что менеджеру центрального офиса потребуется консультация специалиста, равна 0,3, менеджеру регионального офиса — 0,5. Одному из менеджеров потребовалась консультация. Какова вероятность того, что он работает в центральном офисе?
55. Стрелок A поражает мишень с вероятностью 0,6, стрелок B — с вероятностью 0,5 и стрелок B — с вероятностью 0,4. Стрелки дали залп по мишени, и две пули попали в цель. Что вероятнее: попал стрелок B в мишень или нет?
56. Из урны, где было 4 белых и 6 черных шаров, вынут один шар неизвестного цвета. После этого из урны извлечены (без возвращения) два шара, оказавшиеся белыми. При этом условии найти вероятность, что сначала вынут был черный шар.
57. Из урны, где было 11 красных, 5 белых и 10 черных шаров, вынут один шар. После этого из урны извлечены (без возвращения) 2 шара, оказавшиеся красным и белым. При этом условии найти вероятность, что сначала был вынут черный шар.
58. Из урны, где было 10 красных, 6 желтых и 4 синих шара, вынут один шар. После этого из урны извлечены (без возвращения) 2 шара, оказавшиеся красным и желтым. При этом условии найти вероятность, что сначала был вынут синий шар.
59. Из урны, где было 6 белых и 4 черных шара, переложено вынутый наудачу шар в урну, содержащую 7 белых и 2 черных шара. После этого из второй урны был вынут шар, оказавшийся белым. Найти вероятность, что сначала был переложен белый шар.
60. Из урны, где было 5 белых и 3 черных шара, переложено вынутый наудачу шар в урну, содержащую 4 белых и 2 черных шара. После этого из второй урны был вынут шар, оказавшийся белым. Найти вероятность, что сначала был переложен черный шар.

Глава 4

ПОВТОРНЫЕ НЕЗАВИСИМЫЕ ИСПЫТАНИЯ. ТЕОРЕМА БЕРНУЛЛИ

4.1. Испытания Бернулли

Теория вероятностей имеет дело с такими экспериментами, которые можно повторять (по крайней мере теоретически) неограниченное число раз. Пусть некоторый эксперимент повторяется n раз, причем результаты каждого повторения не зависят от исходов предыдущих повторений. Такие серии повторений называют **независимыми испытаниями**. Частным случаем таких испытаний являются независимые испытания Бернулли, которые характеризуются двумя условиями:

1) результатом каждого испытания является один из двух возможных исходов, называемых соответственно «успехом» или «неудачей»;

2) вероятность «успеха» в каждом последующем испытании не зависит от результатов предыдущих испытаний и остается постоянной.

► **Теорема Бернулли.** Если производится серия из n независимых испытаний Бернулли, в каждом из которых «успех» появляется с вероятностью p , то вероятность того, что «успех» в n испытаниях появится ровно m раз, выражается формулой

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

где $q = 1 - p$ — вероятность «неудачи».

Эта формула называется *формулой Бернулли*.

Схему испытаний Бернулли называют также *биномиальной схемой*, а соответствующие вероятности — *биномиальными*, что связано с использованием биномиальных коэффициентов C_n^m .

Доказательство. Каждое испытание Бернулли описывается пространством элементарных исходов $\Omega = \{Y, H\}$, состоящим из двух элементов: Y (успех) и H (неудача), а также их вероятностями $P(Y) = p$, $P(H) = q$, $p + q = 1$. Успехом в испытании будем считать появление события A .

Составной эксперимент (серия из n испытаний) задается пространством Ω_n , каждый элемент которого представляет собой упорядоченный n -мерный набор конкретных результатов этих испытаний. Обозначим через B_m событие, состоящее в том, что в n опытах событие A появилось ровно m раз. Разложим событие B_m в сумму произведений событий, состоящих в появлении и непоявлении события A в отдельных опытах, при этом обозначим через A_i появление события A в i -м опыте и $\bar{A}_i$ — непоявление A в i -м опыте. Тогда каждый вариант события B_m состоит из m появлений события A и $n - m$ непоявлений события A , т.е.

$$B_m = A_1 A_2 \dots \bar{A}_{n-1} \bar{A}_n + A_1 \bar{A}_2 \dots \bar{A}_{n-1} A_n + \dots + \bar{A}_1 \bar{A}_2 \dots A_{n-1} A_n.$$

Число всех комбинаций такого рода равно числу способов, какими можно из n элементов одновременно выбрать m элементов, соответствующих m появлениям события A , т.е. числу сочетаний C_n^m . Вероятность каждой такой комбинации (каждого слагаемого) по теореме умножения независимых событий равна $p^m q^{n-m}$, а так как составляющие событие B_m являются несовместными событиями, то согласно теореме сложения несовместных событий $P(B_m) = P_n(m) = C_n^m p^m q^{n-m}$.

Задача 1. Игральная кость брошена 6 раз. Найти вероятность того, что ровно 3 раза выпадет «шестерка».

Решение. Шестикратное бросание кости можно рассматривать как последовательность независимых испытаний с вероятностью успеха («шестерки»), равной $1/6$, и вероятностью неудачи — $5/6$. Искомую вероятность вычисляем по формуле

$$P_6(3) = C_6^3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^3 = 0,053.$$

Задача 2. Монета бросается 6 раз. Найти вероятность того, что герб выпадет не более чем 2 раза.

Решение. Искомая вероятность равна сумме вероятностей трех событий, состоящих в том, что герб не выпадет ни разу, либо один раз, либо два раза:

$$\begin{aligned} P(A) &= P_6(0) + P_6(1) + P_6(2) = \\ &= C_6^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 + C_6^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 + C_6^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = 0,344. \end{aligned}$$

Задача 3. Аудитор обнаруживает финансовые нарушения у проверяемой фирмы с вероятностью 0,9. Найти вероятность того, что среди 4 фирм-нарушителей будет выявлено больше половины.

Решение. Событие состоит в том, что из 4 фирм-нарушителей будет выявлено три или четыре, т.е.

$$P(A) = P_4(3) + P_4(4) = C_4^3 0,9^3 \cdot 0,1 + C_4^4 0,9^4 = 0,9^3(0,4 + 0,9) = 0,9477.$$

4.2. Наивероятнейшее число успехов

Число m , при котором биномиальные вероятности $P_n(m)$ достигают своего максимального значения (при фиксированном числе испытаний n), называют обычно *наиболее вероятным* (наивероятнейшим) *числом успехов*. Справедливо следующее утверждение.

► **Теорема 2.** Наивероятнейшее число успехов m^* в серии из n независимых испытаний Бернулли (с вероятностью успеха p в одном испытании) определяется соотношением $np - q \leq m^* \leq np + p$, причем:

- 1) если число $np - q$ — дробное, то существует одно наивероятнейшее число m^* ;
- 2) если число $np - q$ — целое, то существует два наивероятнейших числа: $m^* = np - q$, $m^* = np + p$;
- 3) если np — целое число, то наивероятнейшее число $m^* = np$.

Доказательство. Рассмотрим отношение двух соседних вероятностей. Если отношение $\frac{P_n(m+1)}{P_n(m)} = \frac{(n-m)p}{(m+1)q}$ больше

единицы, то последующая вероятность $P_n(m+1)$ превышает предыдущую $P_n(m)$. Если же $P_n(m+1) < P_n(m)$, то отно-

шение $\frac{P_n(m+1)}{P_n(m)} = \frac{(n-m)p}{(m+1)q}$ меньше единицы. Для нахождения

наивероятнейшего числа m^* надо уловить тот момент, когда отношение, большее единицы, станет меньше единицы, т.е. найти такое m^* , для которого одновременно выполняются не-

равенства $\frac{P_n(m+1)}{P_n(m)} \leq 1$, $\frac{P_n(m)}{P_n(m-1)} \geq 1$. Тогда из неравенства

$\frac{P_n(m^*+1)}{P_n(m^*)} = \frac{(n-m^*)p}{(m^*+1)q} \leq 1$ получаем $m^* \geq np - q$, а из неравенства

$\frac{P_n(m^*)}{P_n(m^*-1)} = \frac{(n-m^*+1)p}{m^*q} \geq 1$ получаем $m^* \leq np + p$. Таким об-

разом, получаем, что m^* лежит в интервале единичной длины $np - q \leq m^* \leq np + p$, причем, обозначив через $m_0 = [np - q]$ целую часть числа $np - q$, получим:

1) если число $np - q$ — дробное, то имеется единственное целое число $m^* = (m_0 + 1)$, принадлежащее промежутку $[np - q; np + p]$, для которого вероятность $P_n(m)$ достигает своего максимального значения: $P_n(m^*) = \max P_n(m)$, $m = 0, 1, \dots, n$;

2) если число $np - q$ — целое, то имеются две точки максимума $m^* = np - q$ и $m^* + 1 = np + p$:

$$\max P_n(m) = P_n(m^*) = P_n(m^* + 1), m = 0, 1, \dots, n.$$

Последнее равенство следует из непосредственной проверки того, что отношение $\frac{P_n(m^*+1)}{P_n(m^*)} = \frac{(n-m^*)p}{(m^*+1)q}$ равно единице, если заменить m^* на $np - q$, а q на $1 - p$;

3) если np целое, то наивероятнейшее число $m^* = np$. Действительно, если np — целое, то в промежутке $np - q \leq m \leq np + p$, длиной единица ($p + q = 1$), содержится единственное целое число — np .

Задача 4. Монета подбрасывается 3 раза. Найти наиболее вероятное число успехов (выпадений герба).

Решение. Возможными значениями для числа успехов в трех рассматриваемых испытаниях являются $m = 0, 1, 2$ или 3. Пусть A_m — событие, состоящее в том, что при трех подбрасываниях монеты герб появляется m раз. По формуле Бернулли легко найти вероятности событий A_m .

m	0	1	2	3
$P_s(m)$	1/8	3/8	3/8	1/8

Из таблицы видно, что наиболее вероятными значениями являются числа 1 и 2 (их вероятности равны 3/8). Этот же результат можно получить и из теоремы 2. Действительно, $n = 3$, $p = 1/2$, $q = 1/2$. Тогда

$$3 \cdot \frac{1}{2} - \frac{1}{2} \leq m^* \leq 3 \cdot \frac{1}{2} + \frac{1}{2}, \text{ т.е. } 1 \leq m^* \leq 2.$$

Задача 5. Вероятность получения удачного результата при производстве сложного химического опыта равна 3/4. Найти наимвероятнейшее число удачных опытов, если общее их число равно 10.

Решение. В этом примере $n = 10$, $p = 3/4 = 0,75$, $q = 1/4 = 0,25$. Тогда неравенство для наиболее вероятного числа успехов имеет вид: $10 \cdot 0,75 - 0,25 \leq m^* \leq 10 \cdot 0,75 + 0,75$ или $7,25 \leq m^* \leq 8,25$. Существует только одно целое решение этого неравенства, а именно $m^* = 8$.

Задача 6. В результате каждого визита страхового агента договор заключается с вероятностью 0,1. Найти наимвероятнейшее число заключенных договоров после 25 визитов.

Решение. Имеем $n = 25$, $p = 0,1$, $q = 0,9$. Неравенство для наиболее вероятного числа успехов принимает вид: $25 \cdot 0,1 - 0,9 \leq m^* \leq 25 \cdot 0,1 + 0,9$ или $1,6 \leq m^* \leq 2,6$. У этого неравенства только одно целое решение: $m^* = 2$.

4.3. Предельные теоремы и приближенные формулы

При больших значениях n непосредственное нахождение вероятностей $P_n(m)$ по формуле Бернулли сопряжено с трудностями вычислительного характера, поэтому в таких случаях используют различные варианты приближенных формул, основанных на предельных теоремах Пуассона и Муавра—Лапласа.

Приближенная формула Пуассона используется в том случае, когда число испытаний n велико, а вероятность успеха в отдельном испытании мала ($p < 0,1$) и при этом np невелико ($np < 10$).

► **Теорема Пуассона.** Пусть $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \lambda = \text{const}$. Тогда

$$P_n(m) \rightarrow \frac{\lambda^m}{m!} e^{-\lambda}.$$

Доказательство. По формуле Бернулли, после умножения числителя и знаменателя на n^m и некоторых преобразований, получаем:

$$\begin{aligned} P_n(m) &= C_n^m p^m q^{n-m} = \frac{n(n-1)\dots(n-m+1)p^m}{m!} (1-p)^{n-m} = \\ &= \frac{n(n-1)\dots(n-m+1)n^m}{m!n^m(1-p)^m} p^m \left(1 - \frac{np}{n}\right)^n = \\ &= \frac{1 \cdot \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{m-1}{n}\right)}{m!} (np)^m \left(1 - \frac{np}{n}\right)^n \frac{1}{(1-p)^m} \rightarrow \frac{\lambda^m}{m!} e^{-\lambda}, \quad n \rightarrow \infty, \quad p \rightarrow 0. \end{aligned}$$

Предельные вероятности $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$ называются *пуассоновскими*. Поскольку при больших n верно $np \approx \lambda$, то можно считать, что $\lambda = np$.

Формула выражает закон распределения Пуассона вероятностей *массовых* (n велико) и *редких* (p мало) явлений. Отсюда название закона Пуассона — **закон редких явлений**. Закон Пуассона широко применяют в теории информации, в теории массового обслуживания при изучении потока событий.

В силу определенной «симметричности» понятий «успех» и «неудача» приближенная формула Пуассона может использоваться в схеме независимых испытаний Бернулли при больших n также и в случае, когда p близко к единице, т.е. при $q < 0,1$ и $nq < 10$:

$$P_n(n-m) = C_n^{n-m} p^{n-m} q^m = C_n^m p^{n-m} q^m \approx \frac{\lambda^m}{m!} e^{-\lambda}, \quad \lambda = nq.$$

Задача 7. Известно, что процент брака для некоторой детали равен 0,5%. Контролер проверяет 1000 деталей. Какова вероятность обнаружить ровно три бракованные детали? Какова вероятность обнаружить не меньше трех бракованных деталей?

Решение. Имеем 1000 испытаний Бернулли с вероятностью «успеха» $p = 0,005$. Применяя пуассоновское приближение с $\lambda = np = 5$, получаем:

$$1) P_{1000}(3) \approx \frac{5^3}{3!} e^{-5};$$

2) $P_{1000}(m \geq 3) = 1 - P_{1000}(m < 3) = 1 - [P_{1000}(0) + P_{1000}(1) + P_{1000}(2)] \approx$
 $\approx 1 - \sum_{m=0}^2 \frac{5^m}{m!} e^{-5}$, и по таблице значений функции Пуассона находим $P_{1000}(3) \approx 0,14$; $P_{1000}(m \geq 3) \approx 0,875$.

Приближенные формулы Муавра–Лапласа. Предположим, что в схеме независимых испытаний Бернулли число испытаний n велико, а вероятности «успеха» и «неудачи» не малы (например, $0,1 < p < 0,9$) или np и nq не малы ($np > 10$, $nq > 10$),

► **Локальная теорема Муавра–Лапласа.** Пусть $p = \text{const}$, $n \rightarrow \infty$, тогда

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}, \quad x = \frac{m - np}{\sqrt{npq}},$$

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

Здесь приближенное равенство понимается в смысле асимптотической эквивалентности при $n \rightarrow \infty$ для ограниченных x .

Функция $\varphi(x)$ — четная, и для положительных значений x составлена таблица ее значений (приложение 2).

Задача 8. Вероятность покупки при посещении клиентом магазина составляет $p = 0,75$. Найти вероятность того, что при 100 посещениях клиент совершит покупку ровно 80 раз.

Решение. В данном случае $n = 100$, $m = 80$, $p = 0,75$, $q = 0,25$.

Найдем $x = \frac{80 - 100 \cdot 0,75}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = 1,16$, и по табл. 1 приложения 2 определяем $\varphi(x) = 0,2036$, тогда искомая вероятность $P_{100}(80) = \frac{0,2036}{\sqrt{100 \cdot 0,75 \cdot 0,25}} = 0,047$.

► **Интегральная теорема Муавра–Лапласа.** Пусть $p = \text{const}$, $n \rightarrow \infty$, тогда

$$P\left(\frac{m - np}{\sqrt{npq}} < x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz,$$

где m — число успехов для любого числа x .

Отсюда следует, что для вычисления вероятности $P_n(m_1, m_2) = P(m_1 \leq m \leq m_2)$ события, состоящего в том, что число успехов m в n испытаниях Бернулли окажется заключенным в пределах от m_1 до m_2 , можно использовать приближенную формулу

$$P_n(m_1, m_2) = \Phi_0(x_2) - \Phi_0(x_1),$$

где $x_1 = \frac{m_1 - np}{\sqrt{npq}}$, $x_2 = \frac{m_2 - np}{\sqrt{npq}}$, а $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$ — функция Лапласа.

Функция $\Phi_0(x)$ равна 0 при $x = 0$: $\Phi_0(0) = 0$, и нечетная: $\Phi_0(-x) = -\Phi_0(x)$ для всех x . Для функции $\Phi_0(x)$ составлены специальные таблицы при некоторых положительных значениях аргумента (табл. 2 приложения 2). При $x > 5$ можно считать, что $\Phi_0(x) = 0,5$.

Задача 9. Страховая компания заключила 40 000 договоров. Вероятность страхового случая по каждому из них в течение года составляет 2%. Найти вероятность того, что таких случаев будет не более 870.

Решение. По условию задачи $n = 40\,000$, $p = 0,02$, $np = 800$, $\sqrt{npq} = 28$. Для вычисления вероятности $P(m \leq 870)$ воспользуемся интегральной теоремой Муавра–Лапласа:

$$P(0 < m \leq 870) = \Phi_0(x_2) - \Phi_0(x_1), \text{ где } x_1 = \frac{0-800}{28} = -28,57 \text{ и } x_2 = \frac{870-800}{28} = 2,5.$$

Находим по таблице значений функции Лапласа:

$$P(0 < m \leq 870) = \Phi_0(x_2) - \Phi_0(x_1) = \Phi_0(2,5) - \Phi_0(-28,57) = 0,4938 + 0,5 = 0,9938.$$

Следствие (интегральной теоремы Муавра–Лапласа).

Вероятность того, что относительная частота появления успеха в n независимых испытаниях Бернулли (т.е. число m/n) отклонится от вероятности успеха не более чем на $\varepsilon > 0$, может быть найдена по формуле

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi_0\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

Доказательство получаем из следующей цепочки очевидных равенств:

$$\begin{aligned} P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) &= P\left(p - \varepsilon \leq \frac{m}{n} \leq p + \varepsilon\right) = P(np - n\varepsilon \leq m \leq np + n\varepsilon) \approx \\ &\approx \Phi_0\left(\frac{np + n\varepsilon - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{np - n\varepsilon - np}{\sqrt{npq}}\right) = \Phi_0\left(\varepsilon \sqrt{\frac{n}{pq}}\right) - \Phi_0\left(-\varepsilon \sqrt{\frac{n}{pq}}\right) = 2\Phi_0\left(\varepsilon \sqrt{\frac{n}{pq}}\right). \end{aligned}$$

Задача 10. Каждый из 900 посетителей оптового рынка случайным образом обращается в один из 10 ларьков. В каких границах с вероятностью 0,95 лежит число клиентов отдельно взятого ларька?

Решение. Из условия задачи следует $n = 900$, $p = 0,1$ и заданная вероятность $P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 2\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 0,95$. Отсюда следует, что $\Phi_0\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) = 0,475$. По таблице функции Лапласа находим, что аргумент функции равен $\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} = 1,96$, и получаем значение, $\varepsilon = 0,0196$. Тогда из условия $\left|\frac{m}{900} - \frac{1}{10}\right| \leq 0,0196$ найдем проме-

жуток для m : $|m-90| \leq 17,64$ или $72,34 \leq m \leq 107,64$, и так как m — число целое, то $73 \leq m \leq 107$.

Приближенную формулу можно использовать и в следующей «урновой» схеме: из генеральной совокупности объема N , содержащей M белых и $N-M$ черных шаров, осуществляется последовательный выбор n шаров без возвращения. Вероятность того, что в полученной выборке окажется ровно m белых шаров, вычисляется по формуле

$$P_{M,N}(m, n) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Если объем генеральной совокупности и число белых шаров достаточно велики ($N \rightarrow \infty$, $M \rightarrow \infty$, $M/N \rightarrow p = \text{const}$), то «урновую» схему можно приближенно заменить схемой Бернулли:

$$P_{M,N}(m, n) \approx P_n(m), \quad \text{где } P_n(m) = C_n^m p^m q^{n-m}.$$

Далее, при необходимости, можно использовать формулы Пуассона и Муавра—Лапласа.

4.4. Полиномиальные испытания

От схемы независимых последовательных испытаний с двумя исходами (схема Бернулли или биномиальная схема) можно перейти к полиномиальной схеме, т.е. к схеме последовательных независимых испытаний, в каждом из которых возможны k исходов, $k > 2$, с вероятностями $p_1, p_2, \dots, p_k$, $0 < p_i < 1$, $\sum_{i=1}^k p_i = 1$.

В этом случае пространство элементарных событий содержит k^n таких событий, а вероятность того, что из n испытаний m_1 закончатся первым исходом, m_2 — вторым исходом, ..., m_k — k -м исходом, равна:

$$P_n(m_1, \dots, m_k) = \frac{n!}{m_1! m_2! \dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}.$$

Эта формула описывает полиномиальный закон распределения.

Полиномиальную схему можно трактовать как обобщение *статистики Максвелла—Больцмана* на случай, когда вероятности попадания каждой частицы в различные ячейки различны.

Задача 11. Шесть рукописей раскладываются случайным образом в пять папок. Какова вероятность, что ни одна папка не останется пустой?

Решение. На раскладку 6 рукописей в папки можно смотреть как на серию шести полиномиальных испытаний с 5 исходами (попадание в i -ю папку — это i -й исход). Вероятности исходов (папок) совпадают и равны $p_1 = p_2 = \dots = p_k = 1/5$. Событие $A = \{\text{ни одна папка не останется пустой}\}$ означает, что в одну папку попадут две рукописи, а в остальные папки — по одной рукописи. Следовательно, вероятность того, что в первую папку попадут две рукописи, а в остальные папки — по одной рукописи, равна:

$$P_6(2, 1, 1, 1, 1) = \frac{6!}{2! 1! 1! 1! 1!} \left(\frac{1}{5}\right)^2 \left(\frac{1}{5}\right)^1 \left(\frac{1}{5}\right)^1 \left(\frac{1}{5}\right)^1 \left(\frac{1}{5}\right)^1 = \frac{6!}{2!} \frac{1}{5^5},$$

а вероятность искомого события A (для которого неважно, в какую из 5 папок попадают две рукописи) равна:

$$P(A) = 5P_6(2, 1, 1, 1, 1) = 5 \cdot \frac{6!}{2!} \cdot \frac{1}{5^5} = \frac{6!}{2!} \cdot \frac{1}{5^5} = 0,1152.$$

Задача 12. Курс акции за день может подняться на 1 пункт с вероятностью 50%, опуститься на 1 пункт с вероятностью 30% и остаться неизменным с вероятностью 20%. Найти вероятность того, что за 5 дней торгов курс поднимется на 2 пункта.

Решение. Возможны только следующие два варианта развития событий:

- 1) курс растет 2 дня, ни разу не падает, не меняется 3 дня;
- 2) курс растет 3 дня, падает 1 день, не меняется 1 день.

Таким образом,

$$P(A) = P_5(2, 0, 3) + P_5(3, 1, 1) = \frac{5!}{2! 0! 3!} 0,5^2 \cdot 0,3^0 \cdot 0,2^3 + \\ + \frac{5!}{3! 1! 1!} 0,5^3 \cdot 0,3^1 \cdot 0,2^1 = 0,02 + 0,15 = 0,17.$$

Задачи для самостоятельного решения

1. Ежедневно новая сделка совершается с вероятностью 0,2 (но не более одной в день). Какова вероятность того, что за 5 дней будет совершено 3 сделки?
2. В результате каждого визита страхового агента договор заключается с вероятностью 0,25. Какова вероятность, что из 10 визитов страхового агента 5 закончатся заключением договора?
3. Вероятность поражения мишени стрелком равна 0,9. Найти вероятность, что он поразит мишень ровно два раза, сделав 5 выстрелов.
4. Для вычислительной лаборатории приобретено 9 компьютеров, причем вероятность брака для одного компьютера равна 0,1. Какова вероятность, что придется заменить более двух компьютеров?
5. Зачетная работа по предмету состоит из 6 задач, при этом зачет считается сданным, если студент решил хотя бы 3 задачи. Студент Иванов может решить каждую задачу с вероятностью 0,6 независимо от других. Какова вероятность, что он сдаст зачет?
6. В коробке 3 детали, вероятность брака для каждой детали равна 0,1. Какова вероятность, что среди 10 коробок будет не менее восьми не содержащих бракованных деталей?
7. Два равносильных противника играют в шахматы. Что вероятнее: а) выиграть одну партию из двух или две партии из четырех; б) выиграть не менее двух партий из четырех или не менее трех партий из пяти? Ничьи во внимание не принимаются.
8. Мастер и ученик играют шахматный матч. Мастер выигрывает матч, если он выиграл все партии в матче, ученик выигрывает матч, если он выиграл хотя бы одну партию в матче. Из скольких партий должен состоять матч, чтобы шансы на победу у мастера и ученика были равны, если вероятность победы мастера в одной партии равна 0,9, а ученика — 0,1?
9. Фонд инвестирует средства в три компании. Вероятность получения прибыли от каждой из компаний равна 0,7. Фонд зарабатывает деньги, если хотя бы две компании дают прибыль. Известно, что фонд оказался в выигрыше. Какова вероятность того, что при этом условии вложения во все три компании оказались прибыльными?

10. Курс акции за день растет или падает на 1 пункт равновероятно. Найти вероятность, что за 10 дней торгов курс: а) упадет на 4 пункта; б) поднимется на 2 пункта; в) поднимется на 6 пунктов.
11. В коробке 4 детали. Вероятность того, что деталь стандартна, равна 0,9. Сколько надо взять коробок, чтобы с вероятностью не менее 0,99 среди них нашлась хотя бы одна коробка, не содержащая брака?
12. Вероятность хотя бы одного попадания при двух выстрелах равна 0,96. Найти вероятность трех попаданий при четырех выстрелах.
13. Вероятность того, что в партии из 8 изделий имеется хотя бы одно бракованное изделие, составляет 57%. Найти вероятность того, что там не более одного бракованного изделия.
14. В люстре 3 лампы. Вероятность того, что они перегорят в течение года, составляет 0,8%. Найти вероятность, что в течение года перегорит ровно одна лампа.
15. На ежегодную вечеринку приглашены 12 человек, причем каждый из них может прийти с вероятностью 0,7 независимо от других. Найти наиболее вероятное число гостей и его вероятность.
16. Система состоит из 6 независимо работающих элементов. Вероятность отказа каждого элемента равна 0,3. Найти: а) наименее вероятное число отказавших элементов; б) вероятность наименее вероятного числа отказавших элементов системы; в) вероятность отказа системы, если для этого достаточно, чтобы отказали хотя бы пять элементов.
17. Игральная кость бросается 16 раз. Найти наименее вероятное число бросаний, в которых выпало число очков, кратное трем, и вычислить его вероятность.
18. Известно, что в среднем доля p деталей имеет дефекты. Сколько деталей надо проверить, чтобы с вероятностью r среди них оказалась хотя бы одна с дефектами? Найти наименее вероятное число деталей с дефектами в этом случае. Решить задачу, если: а) $p = 0,01$, $r = 0,95$; б) $p = 0,02$, $r = 0,97$; в) $p = 0,02$, $r = 0,99$.
19. При визите страхового агента вероятность заключения договора равна p . Найти наименее вероятное число заключенных договоров после N визитов и вероятность того, что их будет заключено не больше найденного числа. Решить задачу, если: а) $p = 0,2$, $N = 10$; б) $p = 0,15$, $N = 12$.

20. Страховой агент при каждом визите заключает договор с вероятностью 30%. При каком числе визитов наивероятнейшее число договоров будет равно 10?
21. Сколько надо сделать выстрелов с вероятностью попадания в цель 0,7, чтобы наивероятнейшее число попаданий в цель было равно 15?
22. Сколько раз надо бросить игральную кость, чтобы наивероятнейшее число появлений четного числа очков было равно 6?
23. Сколько надо сыграть партий в шахматы с вероятностью победы в одной партии, равной $1/3$, чтобы наивероятнейшее число побед было равно 5?
24. Найти вероятность того, что в серии из N бросаний монеты числа «орлов» и «решек» совпадают, если: а) $N = 100$; б) $N = 400$.
25. Посетитель магазина совершает покупку с вероятностью 0,75. Какова вероятность, что из 300 посетителей сделают покупки ровно: а) 230; б) 240.
26. Каждый из 100 компьютеров в интернет-кафе занят клиентами в среднем в течение 80% рабочего времени. Какова вероятность, что в некоторый момент клиентами будет занято: а) от 70 до 90 компьютеров; б) не менее 80 компьютеров?
27. На научную конференцию приглашены 100 человек, причем каждый из них прибывает с вероятностью 0,7. В гостинице для гостей заказано 65 мест. Какова вероятность, что все приехавшие будут поселены в гостинице?
28. В партии из N арбузов каждый арбуз оказывается неспелым с вероятностью p . Найти наиболее вероятное число спелых арбузов и вероятность того, что их реальное число будет находиться в пределах от K до L , если: а) $N = 768$, $p = 0,25$, $K = 564$, $L = 600$; б) $N = 800$, $p = 0,2$, $K = 600$, $L = 650$; в) $N = 400$, $p = 0,2$, $K = 300$, $L = 360$; г) $N = 625$, $p = 0,2$, $K = 490$, $L = 520$.
29. Страховая фирма заключила 10 000 договоров. Вероятность страхового случая по каждому в течение года составляет 2%. Найти вероятность, что таких случаев будет: а) не более 230; б) не более 250.
30. Страховая фирма заключила N договоров. Вероятность страхового случая по каждому договору в течение года составляет 2%. Найти наивероятнейшее число страховых случаев и вероятность того, что реальное число случаев отклонится от него не более чем на K , если: а) $N = 10\,000$, $K = 30$; б) $N = 8100$, $K = 25$; в) $N = 1000$, $K = 10$.

31. Консалтинговая фирма обслуживает N компаний. Каждая компания обращается в эту фирму в среднем 4 раза в год. Какова вероятность, что в течение месяца фирме придется обслужить не менее M компаний? Решить задачу, если: а) $N = 120$, $M = 50$; б) $N = 240$, $M = 100$.
32. Каждый из 1000 адресатов, получивших рекламное объявление, обращается в фирму с вероятностью 5%. Найти вероятность, что число обратившихся составит не менее 40 человек.
33. Каждый из 500 адресатов, получивших рекламное объявление, откликается на него с вероятностью 5%. Найти наименьшее число откликов и вероятность, что число откликов будет отличаться от него не более чем на 10.
34. Посетитель магазина совершает покупку с вероятностью 0,8. Какова вероятность, что из 400 посетителей сделают покупки от 300 до 350 человек?
35. Посетитель магазина совершает покупку с вероятностью 0,4. Какова вероятность, что из 600 посетителей сделают покупки от 200 до 250 человек?
36. Известно, что вероятность рождения мальчика равна 0,515, а девочки — 0,485. Какова вероятность того, что среди 1000 новорожденных детей мальчиков будет больше, чем девочек?
37. Производится 500 подбрасываний симметричной монеты. В каких пределах будет находиться отклонение частоты выпадения герба от 0,5 с вероятностью 0,99?
38. При изготовлении деталей наблюдается в среднем 10% брака. Исследуется партия в 1500 деталей. Найти границы, в которых число бракованных деталей лежит с вероятностью 0,9.
39. Каждый из 400 клиентов фирмы случайным образом обращается к одному из 5 менеджеров. В каких границах с вероятностью 0,95 лежит число клиентов отдельно взятого менеджера?
40. Каждый из 900 посетителей оптового рынка случайным образом обращается в один из 10 ларьков. В каких границах с вероятностью 0,95 лежит число клиентов отдельно взятого ларька?
41. Страховая фирма заключила 10000 договоров. Вероятность страхового случая по каждому в течение года составляет 1%. В каких границах с вероятностью 0,95 лежит количество страховых случаев?
42. Стиральным порошком «Снежок» пользуется 25% населения. Сколько людей надо опросить, чтобы определить эту долю с точностью 0,05 с вероятностью 0,95?

43. Зубной пастой «Чистозуб» пользуется 10% населения. Сколько человек надо опросить, чтобы определить эту долю с точностью 0,03 с вероятностью 0,99?
44. На выборах кандидата в мэры поддерживает 40% населения. При изучении общественного мнения было опрошено 1000 человек. С какой вероятностью можно утверждать, что доля избирателей из этой выборки, поддерживающих кандидата, отличается от истинной доли не более, чем на 0,05?
45. Доля населения региона, занятого в промышленности, равна 0,4. В каких границах с вероятностью 0,95 лежит число занятых в промышленности среди 10 000 случайно отобранных людей?
46. Известно, что вероятность «зависания» компьютера равна 0,6%. Какова вероятность, что из 200 компьютеров «зависнут»: а) ровно 6 компьютеров; б) не более 5 компьютеров?
47. В среднем 1% документов содержит ошибки. Какова вероятность, что из 100 документов будет не более двух с ошибками?
48. При наборе текста наборщик делает ошибку в слове с вероятностью p . Какова вероятность, что в тексте из N слов будет не более M ошибок? Решить задачу, если: а) $p = 0,001$, $N = 5000$, $M = 5$; б) $p = 0,002$, $N = 2000$, $M = 4$.
49. При наборе текста наборщик делает ошибку в слове с вероятностью p . Найти наивероятнейшее число ошибок в тексте из N слов и вероятность того, что оно не будет превышено. Решить задачу, если: а) $p = 0,002$, $N = 1500$; б) $p = 0,003$, $N = 1300$.
50. Сборник задач содержит 400 задач с ответами. В каждом ответе может быть ошибка с вероятностью 0,01. Какова вероятность, что ровно 99% всех ответов даны без ошибок?
51. Вероятность правильно прочитать закодированное слово равна 99,75%. С какой вероятностью в тексте длиной 2000 слов будет сделано более двух ошибок?
52. Известно, что вероятность выпуска дефектной детали равна 0,02. Детали укладываются в коробки по 100 штук. Чему равна вероятность того, что: а) в коробке нет дефектных деталей; б) число дефектных деталей в коробке не более двух?
53. Производители калькуляторов знают из опыта, что в среднем 1% проданных калькуляторов имеет дефекты. Аудиторская фирма купила 500 калькуляторов. Какова вероятность, что придется заменить ровно 4 калькулятора?

54. Изделие является годным с вероятностью p . Какова вероятность, что в партии из N деталей будет не более M бракованных? Решить задачу, если: а) $p = 0,95$, $N = 100$, $M = 2$; б) $p = 0,96$, $N = 100$, $M = 3$; в) $p = 0,95$, $N = 120$, $M = 4$.
55. Вероятность, что в партии из 100 изделий имеется брак, составляет 63,2%. Найти вероятность, что там не более трех бракованных изделий.
56. Вероятность того, что пассажир опоздает к поезду, равна 0,004. Найти наиболее вероятное число опоздавших, если всего билет купили 750 человек, и вероятность того, что их будет именно столько.
57. Два шахматиста, A и B , встречались за доской 50 раз, причем 15 раз выиграл A , 10 раз выиграл B и 25 партий закончились вничью. Найти вероятность, что в матче из 10 партий между этими шахматистами 3 партии выиграет A , 2 партии выиграет B , а 5 партий закончатся вничью. Использовать результаты предыдущих матчей для оценки вероятностей выигрыша каждого.
58. В магазине имеются в продаже: один костюм первого роста, два костюма второго роста, три костюма третьего роста. Костюм первого роста спрашивается с вероятностью 0,2, костюм второго роста — с вероятностью 0,3, костюм третьего роста — с вероятностью 0,5. В магазин обратилось три покупателя. Найти вероятность, что хотя бы один из них ушел без покупки.
59. Курс акции за день торгов может подняться или опуститься на 1 пункт либо остаться неизменным (все три случая равновероятны). Найти вероятность, что за 5 дней торгов курс поднимется на 2 пункта.
60. Курс акции за день торгов может подняться на 1 пункт с вероятностью 50%, опуститься на 1 пункт с вероятностью 30% и остаться неизменным с вероятностью 20%. Найти вероятность, что за 5 дней торгов курс: а) поднимется на 3 пункта; б) упадет на 2 пункта.

Глава 5

ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

5.1. Случайная величина и ее закон распределения

Случайной величиной ξ называется любая действительная функция¹ $\xi = \xi(\omega)$, $\omega \in \Omega$, определенная на пространстве элементарных событий Ω . Если множество значений функции конечно или счетно, то такую случайную величину называют дискретной.

В результате опыта случайная величина может принять то или иное значение, причем заранее неизвестно, какое именно. Примерами случайных величин являются: колебания курсов валют или цен товаров, прибыль или убытки фирмы, время выполнения работы, ожидания транспорта и т.д.

Пример 1. При двукратном подбрасывании монеты возможны следующие исходы: $\omega_1 = \text{РР}$, $\omega_2 = \text{РГ}$, $\omega_3 = \text{ГР}$, $\omega_4 = \text{ГГ}$, т.е. пространство элементарных событий имеет вид $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, причем каждый элементарный исход имеет вероятность $1/4$. Пусть $\xi(\omega)$ — число выпадений герба при двукратном бросании монеты, тогда $\xi(\omega_1) = 0$, $\xi(\omega_2) = 1$, $\xi(\omega_3) = 1$, $\xi(\omega_4) = 2$. Зная вероятности для элементарных исходов, можно вычислить вероятности для соответствующих значений случайной величины ξ :

$$P(\xi = 0) = P(\omega_1) = 1/4;$$

$$P(\xi = 1) = P(\omega_2, \omega_3) = 1/4 + 1/4 = 1/2;$$

$$P(\xi = 2) = P(\omega_4) = 1/4.$$

¹ В аксиоматике Колмогорова требуется, чтобы функция была измеримой, т.е. все события $\{\xi(\omega) < t\}$ принадлежали σ -алгебре τ вероятностного пространства (Ω, τ, P) (см. § 3.7). Далее будем считать это условие выполненным по умолчанию.

Полученные вероятности можно свести в следующую таблицу (в первой строке перечислены значения случайной величины, а во второй — их вероятности):

ξ	0	1	2
P	1/4	1/2	1/4

Такая таблица уже не содержит информацию о том, на каком вероятностном пространстве определена случайная величина, в ней приведены лишь значения случайной величины (в первой строке) и их вероятности (во второй строке).

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Простейшей формой закона распределения дискретной случайной величины является ряд распределения.

Рядом распределения дискретной случайной величины ξ называется следующая таблица, в которой перечислены все возможные значения $x_1, x_2, \dots, x_n$ этой случайной величины и соответствующие им вероятности $p_i = P(\xi = x_i)$:

ξ	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

При этом сумма вероятностей равна единице: $\sum_{i=1}^n p_i = 1$.

Если множество значений случайной величины счетно, то эта таблица является бесконечной справа, а суммой вероятностей

является ряд $\sum_{i=1}^{\infty} p_i = 1$.

Задача 1. В связке из трех ключей только один ключ подходит к двери. Ключи перебирают до тех пор, пока не отыщется подходящий. Построить закон распределения для случайной величины ξ — числа опробованных ключей.

Решение. Число опробованных ключей может равняться 1, 2 или 3. Если испытали только один ключ, это означает, что этот первый ключ сразу подошел к двери, а вероятность такого события равна $1/3$. Итак, $P(\xi = 1) = 1/3$. Далее, если опробованных ключей было 2, т.е. $\xi = 2$, это значит, что первый ключ не подошел, а второй — подошел. Вероятность этого события

равна $2/3 \cdot 1/2 = 1/3$, т.е. $P(\xi = 2) = 1/3$. Аналогично вычисляется вероятность $P(\xi = 3) = 1/3$. В результате получается следующий ряд распределения случайной величины:

ξ	1	2	3
P	1/3	1/3	1/3

5.2. Функция распределения случайной величины

Функцией распределения случайной величины ξ называется функция $F_{\xi}(x)$, определенная для любого действительного x и выражающая вероятность того, что случайная величина ξ примет значение, меньшее x :

$$F_{\xi}(x) = P(\xi < x).$$

Функция распределения обладает следующими свойствами:

1. Для любого $x \in R$ справедливо неравенство $0 \leq F_{\xi}(x) \leq 1$.
2. Функция распределения является неубывающей функцией, т.е. если $x_2 \leq x_1$, то $F_{\xi}(x_1) \leq F_{\xi}(x_2)$.
3. Вероятность того, что случайная величина примет значение из полуинтервала $[x_1, x_2)$, равна разности значений функции распределения на концах, т.е.

$$P(x_1 \leq \xi < x_2) = F_{\xi}(x_2) - F_{\xi}(x_1).$$

4. Справедливо равенство: $P(\xi \geq x) = 1 - F_{\xi}(x)$.
5. Справедливы следующие предельные соотношения:

$$\lim_{x \rightarrow -\infty} F_{\xi}(x) = 0, \quad \lim_{x \rightarrow \infty} F_{\xi}(x) = 1.$$

6. Функция распределения непрерывна слева, т.е.

$$\lim_{x \rightarrow a-0} F_{\xi}(x) = F_{\xi}(a).$$

Доказательство.

- 1) Очевидно, поскольку это вероятность события.
- 2) Представим событие, состоящее в том, что случайная величина примет значение, меньшее x_2 , в виде суммы несовместных событий

$$\{\omega: \xi(\omega) < x_2\} = \{\omega: \xi(\omega) < x_1\} \cup \{\omega: x_1 \leq \xi(\omega) < x_2\}.$$

По теореме сложения для несовместных событий:

$$F_{\xi}(x_2) = P(\xi(\omega) < x_2) = P(\xi(\omega) < x_1) + P(x_1 \leq \xi(\omega) < x_2) = F_{\xi}(x_1) + P(x_1 \leq \xi(\omega) < x_2),$$

и так как вероятность $P(x_1 \leq \xi(\omega) \leq x_2) \geq 0$, то получаем неравенство $F_\xi(x_2) \geq F_\xi(x_1)$.

3) Свойство вытекает из равенства

$$F_\xi(x_2) = F_\xi(x_1) + P(x_1 \leq \xi(\omega) < x_2).$$

4) Событие $\{\omega: \xi(\omega) \geq x\}$ является противоположным событию $\{\omega: \xi(\omega) < x\}$ и, следовательно,

$$P(\xi \geq x) = 1 - P(\xi < x) = 1 - F_\xi(x).$$

5) Рассмотрим убывающую последовательность $a_n \rightarrow -\infty$ и множества $A_n = \{\omega: \xi(\omega) < a_n\}$. Они удовлетворяют условиям аксиомы непрерывности (см. § 3.7), так что $F_\xi(a_n) = P(A_n) \rightarrow 0$, $n \rightarrow \infty$. Для возрастающей последовательности $a_n \rightarrow +\infty$ и множеств $A_n = \{\omega: \xi(\omega) \geq a_n\}$ получаем $F_\xi(a_n) = 1 - P(A_n) \rightarrow 1$, $n \rightarrow \infty$.

6) Рассмотрим возрастающую последовательность $a_n < a$, $a_n \rightarrow a$ и множества $A_n = \{\omega: a_n \leq \xi(\omega) < a\}$. Они удовлетворяют условиям аксиомы непрерывности, и поскольку $F_\xi(a) = F_\xi(a_n) + P(A_n)$, то $F_\xi(a_n) = F_\xi(a) - P(A_n) \rightarrow F(a)$, $n \rightarrow \infty$.

Таким образом, каждая функция распределения является неубывающей, непрерывной слева и удовлетворяющей условиям $F(-\infty) = 0$ и $F(+\infty) = 1$ функцией. Верно и обратное: каждая функция, удовлетворяющая перечисленным условиям, может рассматриваться как функция распределения некоторой случайной величины.

Функция распределения является универсальным законом распределения случайной величины. Все ее свойства остаются верными и тогда, когда пространство элементарных событий не является дискретным.

Функция распределения любой дискретной величины разрывна, возрастает скачками при тех значениях x , которые являются возможными значениями ξ . Величина скачка функции $F(x)$ в точке x , равна p_i .

Задача 2. Построить функцию распределения $F_\xi(x)$ для случайной величины ξ из задачи 1.

Решение. Случайная величина ξ имеет три значения 1, 2, 3, которые делят всю числовую ось на четыре промежутка: $(-\infty, 1]$, $(1, 2]$, $(2, 3]$, $(3, +\infty)$. Если $x \leq 1$, то неравенство $\xi < x$ невозможно (левее x нет значений случайной величины ξ) и значит, для такого x функция $F_\xi(x) = 0$.

Если $1 < x \leq 2$, то неравенство $\xi < x$ возможно только если $\xi = 1$, а вероятность такого события равна $1/3$, поэтому для таких x функция распределения $F_{\xi}(x) = 1/3$.

Если $2 < x \leq 3$, неравенство $\xi < x$ означает, что или $\xi = 1$, или $\xi = 2$, поэтому в этом случае вероятность $P(\xi < x) = P(\xi = 1) + P(\xi = 2) = 2/3$, т.е. $F_{\xi}(x) = 2/3$.

И наконец, в случае $x > 3$ неравенство $\xi < x$ выполняется для всех значений случайной величины ξ , поэтому $P(\xi < x) = P(\xi = 1) + P(\xi = 2) + P(\xi = 3) = 1$, т.е. $F_{\xi}(x) = 1$.

Итак, мы получили следующую функцию:

$$F_{\xi}(x) = \begin{cases} 0, & x \leq 1; \\ 1/3, & 1 < x \leq 2; \\ 2/3, & 2 < x \leq 3; \\ 1, & x > 3. \end{cases}$$

5.3. Дискретный случайный вектор

Пусть на дискретном вероятностном пространстве Ω задано несколько случайных величин $\xi_1(\omega)$, $\xi_2(\omega)$, ..., $\xi_n(\omega)$, где $\omega \in \Omega$. Такой упорядоченный набор называется *многомерным случайным вектором*, или *n-мерной случайной величиной*, и обозначается $\xi(\omega) = (\xi_1, \xi_2, \dots, \xi_n)$.

Рассмотрим случай $n = 2$, т.е. когда на пространстве элементарных исходов заданы две дискретные случайные величины ξ и η , принимающие значения x_i ($i = 1, 2, \dots$) и y_j ($j = 1, 2, \dots$) соответственно. Упорядоченная пара (ξ, η) называется *двумерным случайным вектором*, или *двумерной случайной величиной*. Сами величины ξ и η называются в этом случае *составляющими* (или *компонентами*) случайного вектора.

Геометрически совокупность двух случайных величин можно рассматривать как случайную точку с координатами (ξ, η) на плоскости XOY или как случайный вектор, направленный из начала координат в точку (ξ, η) , составляющие которого случайные величины ξ и η . Совокупность трех случайных величин изображается случайной точкой или случайным вектором в трехмерном пространстве, совокупность n случайных величин — случайной точкой или случайным вектором в пространстве n измерений.

Любое соотношение между возможными значениями случайного вектора и их вероятностями называется *совместным законом распределения*.

Совместный закон распределения вероятностей дискретных величин ξ и η задается набором вероятностей p_{ij} одновременного осуществления событий $\{\xi = x_i\}$ и $\{\eta = y_j\}$, т.е. $p_{ij} = P\{\xi = x_i, \eta = y_j\}$, и представляется в виде таблицы:

ξ	η			
	y_1	y_2	...	y_m
x_1	p_{11}	p_{12}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2m}
...	...	...	...	...
x_n	p_{n1}	p_{n2}	...	p_{nm}

Первый столбец таблицы содержит все возможные значения составляющей ξ , а первая строка — все возможные значения составляющей η . В соответствующей клетке таблицы указана вероятность того, что двумерная случайная величина приняла значение (x_i, y_j) . Так как события $\{\omega: \xi(\omega) = x_i, \eta(\omega) = y_j\}$ образуют полную группу событий, то сумма вероятностей, помещенных во всех клетках таблицы, равна единице: $\sum_i \sum_j P(\xi = x_i, \eta = y_j) = \sum_i \sum_j p_{ij} = 1$.

Такая таблица называется *рядом распределения вектора* (ξ, η) .

Вероятность события типа $\{(\xi, \eta) \in B\}$ — «случайная точка (ξ, η) попадает в область B » — вычисляется по формуле $P((\xi, \eta) \in B) = \sum_{(x_i, y_j) \in B} P(\xi = x_i, \eta = y_j)$, где суммирование происходит по

всем возможным парам (x_i, y_j) значений случайных величин ξ и η , для которых случайная точка (x_i, y_j) входит в область B .

Частным законом распределения случайной величины ξ называется набор вероятностей событий $\{\xi = x_i\}$. Если задан совместный закон распределения, то частный закон распределения для ξ можно получить с помощью формулы

$$p_{\xi}(x_i) = P\{\xi = x_i\} = \sum_j P\{\xi = x_i, \eta = y_j\} = \sum_j p_{ij}.$$

Действительно, событие $\{\omega: \xi = x_i\}$ может появиться с одним из событий $\{\xi = x_i, \eta = y_1\}, \{\xi = x_i, \eta = y_2\}, \dots, \{\xi = x_i, \eta = y_m\}$, которые несовместны и их объединение равно событию $\{\omega: \xi = x_i\}$, т.е. $\{\xi = x_i\} = \bigcup_{j=1}^m \{\xi = x_i, \eta = y_j\}$. Отсюда в силу теоремы сложения несовместных событий следует, что $P(\xi = x_i) = \sum_{j=1}^m p_{ij}$.

Аналогично частным законом распределения η называется набор вероятностей событий $\{\eta = y_j\}$, которые также можно вычислить с помощью формулы

$$p_{\eta}(y_j) = P(\eta = y_j) = \sum_i P(\xi = x_i, \eta = y_j) = \sum_i p_{ij}.$$

Таким образом, распределение каждой случайной величины восстанавливается с помощью совместного закона распределения вероятностей.

Случайные величины ξ и η называются *независимыми*, если для любых множеств A и B выполняется условие: $P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B)$, т.е. независимы события $\{\xi \in A\}$ и $\{\eta \in B\}$. Справедлива следующая теорема.

► **Теорема.** Дискретные случайные величины ξ и η независимы тогда и только тогда, когда события $\{\xi = x_i\}$ и $\{\eta = y_j\}$ независимы для всех значений x_i и y_j .

Доказательство. Пусть $P(\xi = x_i, \eta = y_j) = P(\xi = x_i)P(\eta = y_j)$, тогда для всех значений x_i и y_j получаем

$$\begin{aligned} P(\xi \in A, \eta \in B) &= \sum_{x_i \in A, y_j \in B} P(\xi = x_i, \eta = y_j) = \sum_{x_i \in A, y_j \in B} P(\xi = x_i)P(\eta = y_j) = \\ &= \sum_{x_i \in A} P(\xi = x_i) \sum_{y_j \in B} P(\eta = y_j) = P(\xi \in A)P(\eta \in B). \end{aligned}$$

Задача 3. Совместный закон распределения случайных величин ξ и η задан с помощью таблицы:

ξ	η	
	1	2
-1	1/16	3/16
0	1/16	3/16
1	1/8	3/8

Вычислить частные законы распределения составляющих величин ξ и η . Определить, зависимы ли они. Вычислить вероятность $P(\xi + \eta \geq 2)$.

Решение. Частное распределение для ξ получается суммированием вероятностей в строках:

$$P(\xi = -1) = P(\xi = -1, \eta = 1) + P(\xi = -1, \eta = 2) = 1/16 + 3/16 = 1/4;$$

$$P(\xi = 0) = P(\xi = 0, \eta = 1) + P(\xi = 0, \eta = 2) = 1/16 + 3/16 = 1/4;$$

$$P(\xi = 1) = P(\xi = 1, \eta = 1) + P(\xi = 1, \eta = 2) = 1/8 + 3/8 = 1/2.$$

Аналогично получается частное распределение для η :

$$P(\eta = 1) = 1/16 + 1/16 + 1/8 = 1/4;$$

$$P(\eta = 2) = 3/16 + 3/16 + 3/8 = 3/4.$$

Полученные вероятности можно записать в ту же таблицу напротив соответствующих значений случайных величин.

ξ	η		
	1	2	P_{ξ}
-1	1/16	3/16	1/4
0	1/16	3/16	1/4
1	1/8	3/8	1/2
P_{η}	1/4	3/4	1

Теперь ответим на вопрос о независимости случайных величин ξ и η . С этой целью для каждой клетки совместного распределения вычислим произведение $P(\xi = x_i)P(\eta = y_j)$ (т.е. сумм по соответствующей строке и столбцу) и сравним его со значением вероятности $P(\xi = x_i, \eta = y_j)$ в этой клетке. Например, в клетке для значений $\xi = -1$ и $\eta = 1$ стоит вероятность 1/16, а произведение соответствующих частных вероятностей $1/4 \cdot 1/4 = 1/16$, т.е. совпадает с совместной вероятностью. Это условие так же проверяется в оставшихся пяти клетках, и оно оказывается верным везде. Следовательно, случайные величины ξ и η независимы.

Заметим, что если бы наше условие нарушалось хотя бы в одной клетке, то величины следовало бы признать зависимыми.

Для вычисления вероятности $P(\xi + \eta \geq 2)$ отметим клетки, для которых выполнено условие $\xi + \eta \geq 2$. Таких клеток всего три, и соответствующие вероятности в этих клетках равны $1/8$, $3/16$, $3/8$. Их сумма $11/16$, это и есть искомая вероятность. Вычисление этой вероятности можно записать так:

$$\begin{aligned} P(\xi + \eta \geq 2) &= P(\xi = 1, \eta = 1) + P(\xi = 0, \eta = 2) + P(\xi = 1, \eta = 2) = \\ &= 1/8 + 3/16 + 3/8 = 11/16. \end{aligned}$$

5.4. Совместная функция распределения случайного вектора

Рассмотрим вероятностное пространство Ω , на котором определен вектор $\xi = (\xi, \eta)$. Обозначим через $\{\xi < x, \eta < y\}$ множество тех элементарных событий ω , для которых одновременно выполняются эти неравенства.

Совместной функцией распределения вероятностей случайных величин ξ и η или случайного вектора (ξ, η) называется функция двух аргументов $F(x, y)$, равная вероятности $P(\xi < x, \eta < y) = F(x, y)$.

Это общее определение верно как в дискретном, так и в непрерывном случаях.

Частными функциями распределения называют функции распределения составляющих ξ и η : $F_\xi(x)$ и $F_\eta(y)$.

Многомерные функции распределения обладают аналогичными свойствами, что и одномерные:

1. $0 \leq F(x, y) \leq 1$;
2. $F(x, y)$ есть неубывающая функция по каждому из аргументов;
3. $F(x, y)$ непрерывна слева по каждому из аргументов;
4. $F(x, y)$ удовлетворяет соотношениям: $F(+\infty, +\infty) = 1$ и $F(x, -\infty) = F(-\infty, y) = 0$.

Геометрически функция распределения $F(x, y) = P(\xi < x, \eta < y)$ определяет вероятность попадания случайной точки в бесконечный угол с вершиной в точке (x, y) .

Геометрическая интерпретация функции распределения $F(x, y)$ позволяет дать простое пояснение предельным свойствам функции распределения: если $x \rightarrow -\infty$ (или $y \rightarrow -\infty$), то правая граница (или верхняя граница) бесконечного квадранта неограниченно смещается влево (или вниз), и вероятность по-

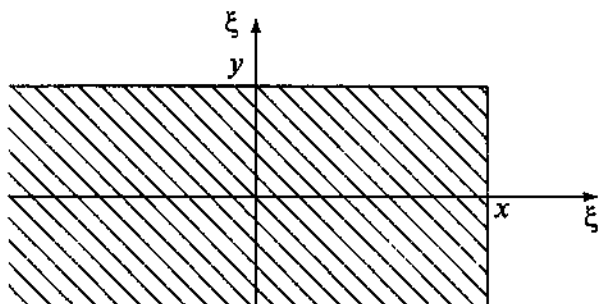


Рис. 5.1

падения случайной точки в квадрант стремится к нулю. При $x \rightarrow +\infty$ и $y \rightarrow +\infty$ бесконечный квадрант превращается во всю плоскость XOY и попадание случайной точки в эту плоскость является достоверным событием.

Если один из аргументов равен $+\infty$, то функция распределения вектора превращается в функцию распределения случайной величины, соответствующей другому аргументу: $F(x, +\infty) = F_\xi(x)$, $F(+\infty, y) = F_\eta(y)$. Действительно, поскольку событие $\{\omega: \eta(\omega) < +\infty\}$ достоверно, $F(x, +\infty)$ определяет вероятность события $\{\omega: \xi(\omega) < x\}$, т.е. представляет собой функцию распределения составляющей ξ . Аналогично $F(+\infty, y) = F_\eta(y)$. Геометрически функция распределения составляющих есть вероятность попадания случайной точки в полуплоскость, ограниченную прямой $X = x$, для $F(x, +\infty) = F_\xi(x)$, или прямой $Y = y$ для $F(+\infty, y) = F_\eta(y)$ (рис. 5.2).

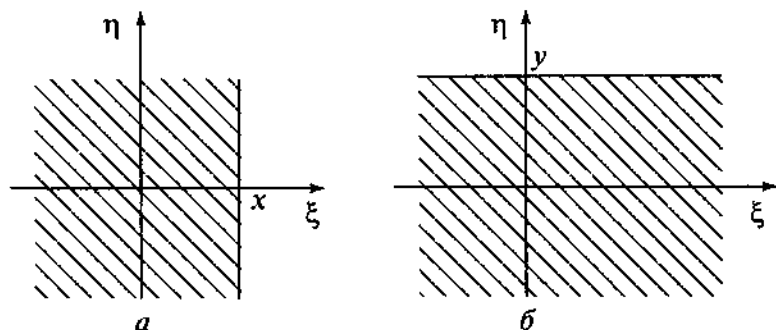


Рис. 5.2

Используя геометрическую интерпретацию функции распределения $F(x, y)$, можно вычислить вероятность попадания случайного вектора $\xi(\omega) = (\xi(\omega), \eta(\omega))$ в прямоугольник $\Pi = \{(x, y): x \in [a_1, b_1], y \in [a_2, b_2]\}$. Действительно (рис. 5.3), $P[(\xi, \eta) \in \Pi] = P(\xi < b_1, \eta < b_2) - P(\xi < b_1, \eta < a_2) + P(\xi < a_1, \eta < a_2) - P(\xi < a_1, \eta < b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$.

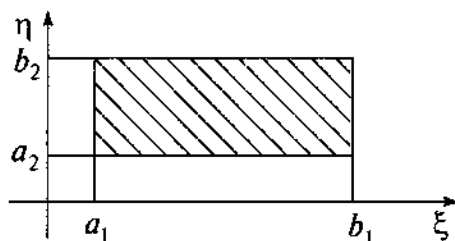


Рис. 5.3

Если вектор (ξ, η) непрерывный, то эта формула верна для любого интервала, полуинтервала или отрезка, так как вероятности точек и прямых в этом случае нулевые.

В двумерном случае приведенных четырех свойств недостаточно, чтобы функция $F(x, y)$ была функцией распределения. Необходимо, чтобы при любых a_1 и b_1 следующее выражение было неотрицательно:

$$\begin{aligned} P(a_1 \leq \xi < b_1, a_2 \leq \eta < b_2) &= P[(\xi, \eta) \in \Pi] = \\ &= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0. \end{aligned}$$

Из определения независимости случайных величин ξ и η следует, что их совместная функция распределения равна произведению функций распределения составляющих: $F(x, y) = F_\xi(x)F_\eta(y)$.

Справедливо и обратное: если $F(x, y) = F_\xi(x)F_\eta(y)$ верно для любых x и y , то случайные величины ξ и η независимы.

Все свойства, доказанные для функции распределения двумерной случайной величины, остаются справедливыми и для функции распределения $F(x_1, x_2, \dots, x_n)$ в случае n аргументов.

5.5. Числовые характеристики дискретных случайных величин

Пусть ξ — дискретная случайная величина со значениями $x_1, x_2, \dots, x_n$ и их вероятностями $p_i = P(\xi = x_i)$, $i = 1, 2, \dots, n$.

Математическим ожиданием (или *средним значением*) дискретной случайной величины ξ называется число

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

Если множество значений случайной величины ξ счетно, то математическое ожидание определяется как бесконечный ряд

$$M\xi = \sum_{i=1}^{\infty} x_i p_i$$

в случае, когда он абсолютно сходится. Если ξ — дискретная величина и $\varphi(x)$ — некоторая функция, то математическое ожидание величины $\eta = \varphi(\xi)$ можно вычислить по формуле

$$M\varphi(\xi) = \sum_i \varphi(x_i) p_i$$

в случае конечной суммы либо при условии, что ряд абсолютно сходится.

Если заданы совместное распределение вероятностей случайных величин ξ и η и функция $\varphi(x, y)$ двух аргументов, то

$$M\varphi(\xi, \eta) = \sum_{i,j} \varphi(x_i, y_j) p_{ij}.$$

Математическое ожидание обладает следующими свойствами:

- 1) $MC = C$ (C — константа);
- 2) $M(C\xi) = CM\xi$ для любой константы C ;
- 3) $M(\xi + \eta) = M\xi + M\eta$;
- 4) $M(\xi\eta) = (M\xi)(M\eta)$, если ξ и η независимы.

Доказательство.

1) константу C можно рассматривать как случайную величину, принимающую единственное значение C с вероятностью единица, так что $MC = C \times 1 = C$;

2) случайная величина $C\xi$ принимает значения Cx_i вместо x_i с вероятностями p_i , таким образом, сумма увеличивается в C раз;

3) используя ранее введенные обозначения, получаем

$$\begin{aligned}
 M(\xi + \eta) &= \sum_{i,j} (x_i + y_j) p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \\
 &= \sum_i x_i p_i(x_i) + \sum_j y_j p_j(y_j) = M\xi + M\eta
 \end{aligned}$$

4) аналогичным образом получаем

$$M(\xi\eta) = \sum_{i,j} x_i y_j p_{ij} = \sum_{i,j} x_i y_j p_i(x_i) p_j(y_j) = \sum_i x_i p_i(x_i) \cdot \sum_j y_j p_j(y_j) = M\xi \cdot M\eta.$$

Методом математической индукции можно доказать, что математическое ожидание суммы n случайных величин $\xi_1, \xi_2, \dots, \xi_n$ равно сумме их математических ожиданий:

$$M(\xi_1 + \xi_2 + \dots + \xi_n) = M\xi_1 + M\xi_2 + \dots + M\xi_n.$$

Дисперсией случайной величины ξ называется число $D\xi = M(\xi - M\xi)^2$. Величина $\sigma = \sqrt{D\xi}$ называется *средним квадратическим отклонением*.

Из определения дисперсии вытекают формулы для вычисления дисперсии дискретной случайной величины:

$$D\xi = \sum_{i=1}^n (x_i - M\xi)^2 p_i,$$

если случайная величина принимает конечное число значений $x_1, x_2, \dots, x_n$ и

$$D\xi = \sum_{i=1}^{\infty} (x_i - M\xi)^2 p_i$$

при условии абсолютной сходимости ряда.

Для дисперсии справедливы следующие свойства:

1. $D\xi = M\xi^2 - (M\xi)^2$.
2. $DC = 0$ (дисперсия постоянной равна нулю).
3. $D(C\xi) = C^2 D\xi$.
4. $D(\xi + C) = D\xi$.
5. Если случайные величины ξ и η независимы, то $D(\xi + \eta) = D\xi + D\eta$.

Доказательство.

1. Из свойств математического ожидания получаем

$$D\xi = M(\xi - M\xi)^2 = M(\xi^2 - 2\xi M\xi + (M\xi)^2) = M\xi^2 - 2(M\xi)^2 + (M\xi)^2 = M\xi^2 - (M\xi)^2$$

(при условии существования конечных $M\xi^2$ и $M\xi$).

$$2. DC = (C - MC)^2 = (C - C)^2 = 0.$$

$$3. D(C\xi) = M(C\xi - M(C\xi))^2 = M(C\xi - CM\xi)^2 = C^2 M(\xi - M\xi)^2 = C^2 D\xi.$$

$$4. D(\xi + C) = M((\xi + C) - M(\xi + C))^2 = M((\xi + C) - (M\xi + C))^2 = \\ = M(\xi - M\xi)^2 = D\xi.$$

$$5. D(\xi + \eta) = M((\xi + \eta) - M(\xi + \eta))^2 = M((\xi - M\xi) + (\eta - M\eta))^2 = \\ = D\xi + 2M((\xi - M\xi)(\eta - M\eta)) + D\eta,$$

$$\text{где } M((\xi - M\xi)(\eta - M\eta)) = M(\xi\eta - \xi M\eta - \eta M\xi + M\xi M\eta) = M(\xi\eta) - \\ - 2M\xi M\eta + M\xi M\eta = 0.$$

Методом математической индукции можно доказать, что если $\xi_1, \xi_2, \dots, \xi_n$ попарно независимы, то дисперсия их суммы равна сумме дисперсий:

$$D(\xi_1 + \xi_2 + \dots + \xi_n) = D\xi_1 + D\xi_2 + \dots + D\xi_n.$$

Дисперсия является мерой разброса значений случайной величины вокруг ее математического ожидания.

Модой случайной величины называют ее наиболее вероятное значение. В частности, наивероятнейшее значение числа успехов в схеме Бернулли — это *мода биномиального распределения*.

Медианой случайной величины ξ называют число X_{med} , такое что $P(\xi < X_{\text{med}}) = P(\xi > X_{\text{med}}) = 1/2$. Для дискретной случайной величины ξ это число может не совпадать ни с одним из значений ξ . Поэтому медиану дискретной случайной величины определяют как любое число X_{med} , лежащее между двумя соседними возможными значениями x_i и x_{i+1} , такими, что $F(x_i) < 1/2$; $F(x_{i+1}) \geq 1/2$.

Задача 4. Пусть случайная величина ξ имеет следующий закон распределения:

ξ	-1	0	2
P	1/4	1/4	1/2

Вычислить математическое ожидание $M\xi$, дисперсию $D\xi$ и среднее квадратическое отклонение σ .

Решение. По определению математическое ожидание ξ равно:

$$M\xi = \sum_{i=1}^3 x_i p_i = -1 \cdot 1/4 + 0 \cdot 1/4 + 2 \cdot 1/4 = 1/4.$$

Далее:

$$M\xi^2 = \sum_{i=1}^3 x_i^2 p_i = (-1)^2 \cdot 1/4 + 0^2 \cdot 1/4 + 2^2 \cdot 1/4 = 5/4,$$

а потому

$$D\xi = M\xi^2 - (M\xi)^2 = 5/4 - 1/16 = 19/16.$$

Среднее квадратическое отклонение $\sigma = \sqrt{D\xi} = \sqrt{19}/4$.

Задача 5. Для пары случайных величин из задачи 3 вычислить $M(\xi\eta)$.

Решение. Воспользуемся формулой $M(\xi\eta) = \sum_{i,j} x_i y_j p_{ij}$, а имен-

но: в каждой клетке таблицы выполняем умножение соответствующих значений x_i и y_j , результат умножаем на вероятность p_{ij} , и все это суммируем по всем клеткам таблицы. В итоге получаем:

$$M(\xi\eta) = -1 \cdot 1 \cdot 1/16 + (-1) \cdot 2 \cdot 3/16 + 0 \cdot 1 \cdot 1/16 + 0 \cdot 2 \cdot 3/16 + \\ + 1 \cdot 1 \cdot 1/8 + 1 \cdot 2 \cdot 3/8 = -1/16 - 3/8 + 1/8 + 3/4 = 7/16.$$

5.6. Основные дискретные распределения и их характеристики

Распределение Бернулли (или *биномиальное распределение*), определяется как закон распределения случайной величины равной числу успехов в n испытаниях Бернулли. Эта случайная величина ξ может принять любое из значений $0, 1, 2, \dots, n$, и их вероятности определяются формулой Бернулли: если p — вероятность успеха, q — вероятность неудачи, то

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

Для распределения Бернулли $M\xi = np$, $D\xi = npq$.

Распределение Пуассона. Случайная величина, распределенная по закону Пуассона, может принять любое из значений $0, 1, 2, \dots$ (счетное множество значений), а их вероятности задаются формулой

$$P(\xi = m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, 2, \dots, \lambda > 0.$$

Для распределения Пуассона $M\xi = \lambda$, $D\xi = \lambda$.

Геометрическое распределение имеет случайная величина ξ , равная числу испытаний Бернулли до первого «успеха» (включительно) с вероятностью «успеха» в одном испытании, равной p . Такая случайная величина ξ принимает значения 1, 2, 3, ..., а их вероятности задаются формулой

$$P(\xi = m) = pq^{m-1}, \quad m = 1, 2, \dots, \quad 0 < p < 1, \quad q = 1 - p.$$

Для геометрического распределения $M\xi = 1/p$, $D\xi = q/p^2$.

Гипергеометрическое распределение возникает, например, в задаче о выборке деталей. Пусть имеется N деталей, из которых M — стандартные. Делается выборка из n деталей. Случайная величина ξ определяется как число стандартных деталей в такой выборке. Оно может равняться любому числу от 0 до n , но, не больше, чем M . С другой стороны, число нестандартных деталей в выборке не больше, чем $N - M$, поэтому число стандартных не меньше $n - (N - M)$. Вероятности этих значений определяются гипергеометрической формулой

$$P(\xi = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad m = \max(0, n - N + M), \dots, \min(n, M).$$

Для гипергеометрического распределения $M\xi = np$, $D\xi = (N - n)npq/(N - 1)$, где $p = M/N$, $q = 1 - p$.

5.7. Ковариация. Коэффициент корреляции

Ковариацией случайных величин ξ и η называется число

$$\text{cov}(\xi, \eta) = M[(\xi - M\xi)(\eta - M\eta)]$$

(в предположении существования всех математических ожиданий).

Из определения ковариации вытекают следующие ее свойства:

1. Ковариацию можно вычислить по формуле

$$\text{cov}(\xi, \eta) = M(\xi\eta) - (M\xi)(M\eta).$$

2. Если ξ и η — независимые случайные величины, то $\text{cov}(\xi, \eta) = 0$.

$$3. \text{cov}(\xi, \eta) = \text{cov}(\eta, \xi).$$

$$4. \text{cov}(C\xi, \eta) = C \text{cov}(\xi, \eta); \text{cov}(\xi, C\eta) = C \text{cov}(\xi, \eta).$$

$$5. \operatorname{cov}(\xi_1 + \xi_2, \eta) = \operatorname{cov}(\xi_1, \eta) + \operatorname{cov}(\xi_2, \eta); \operatorname{cov}(\xi, \eta_1 + \eta_2) = \operatorname{cov}(\xi, \eta_1) + \operatorname{cov}(\xi, \eta_2).$$

$$6. \operatorname{cov}(\xi, \xi) = D\xi.$$

7. Если случайные величины ξ и η имеют конечные дисперсии $D\xi$ и $D\eta$, то дисперсия суммы этих случайных величин существует и равна

$$D(\xi + \eta) = D\xi + D\eta + 2\operatorname{cov}(\xi, \eta).$$

$$8. -\sqrt{D\xi D\eta} \leq \operatorname{cov}(\xi, \eta) \leq \sqrt{D\xi D\eta}.$$

9. Равенство $\operatorname{cov}(\xi_1, \xi_2) = \pm\sqrt{D\xi_1 D\xi_2}$ достигается тогда и только тогда, когда случайные величины ξ_1 и ξ_2 линейно зависимы.

Доказательство.

$$1. \operatorname{cov}(\xi, \eta) = M[(\xi - M\xi)(\eta - M\eta)] = M[(\xi\eta - \xi M\eta - \eta M\xi + M\xi M\eta)] = M(\xi\eta) - 2M\xi M\eta + M\xi M\eta = M(\xi\eta) - (M\xi)(M\eta).$$

$$2. \text{ В случае независимых случайных величин } M(\xi\eta) - (M\xi)(M\eta) = 0.$$

3. От перемены мест сомножителей произведение не меняется.

$$4. \operatorname{cov}(C\xi, \eta) = M[(C\xi - M(C\xi))(\eta - M\eta)] = M[(C\xi - CM\xi) \times (\eta - M\eta)] = CM[(\xi - M\xi)(\eta - M\eta)] = C\operatorname{cov}(\xi, \eta). \text{ Аналогично делается для } \eta.$$

5. Воспользуемся свойством математического ожидания:

$$\begin{aligned} \operatorname{cov}(\xi_1 + \xi_2, \eta) &= M(((\xi_1 + \xi_2) - M(\xi_1 + \xi_2))(\eta - M\eta)) = \\ &= M((\xi_1 - M\xi_1) + (\xi_2 - M\xi_2))(\eta - M\eta) = \\ &= M((\xi_1 - M\xi_1)(\eta - M\eta)) + M((\xi_2 - M\xi_2)(\eta - M\eta)) = \\ &= \operatorname{cov}(\xi_1, \eta) + \operatorname{cov}(\xi_2, \eta). \end{aligned}$$

Аналогично для η_1 и η_2 .

6. Получаем по определению дисперсии.

$$\begin{aligned} 7. D(\xi + \eta) &= M(\xi + \eta)^2 - (M(\xi + \eta))^2 = M(\xi^2 + 2\xi\eta + \eta^2) - \\ &- (M\xi + M\eta)^2 = M\xi^2 + 2M\xi\eta + M\eta^2 - (M\xi)^2 - 2M\xi M\eta - \\ &- (M\eta)^2 = D\xi + D\eta + 2\operatorname{cov}(\xi, \eta). \end{aligned}$$

8. Найдем дисперсию случайной величины $\eta_x = x\xi_1 - \xi_2$, где x — произвольное число. По свойствам дисперсии получим

$$D\eta_x = x^2 D\xi - 2x \operatorname{cov}(\xi, \eta) + D\eta.$$

Как функция от x дисперсия $D\eta_x$ представляет собой квадратный трехчлен. Но дисперсия любой случайной величины не может быть меньше нуля, так что $D\eta_x \geq 0$ для любого x . Поскольку $D\xi_1 \geq 0$, то дискриминант уравнения $D\eta_x = 0$ должен быть неположителен, т.е.

$$[\text{cov}(\xi, \eta)]^2 - D\xi D\eta \leq 0 \text{ или } |\text{cov}(\xi, \eta)| \leq \sqrt{D\xi D\eta}.$$

9. Предположим, что дискриминант равен нулю, тогда уравнение

$$x^2 D\xi - 2x \text{cov}(\xi, \eta) + D\eta = 0$$

имеет решение x_0 и $D\eta_{x_0} = 0$. Это означает, что $\eta_{x_0} \equiv c = \text{const}$, т.е. величина постоянная, и случайные величины ξ и η связаны линейной функциональной зависимостью $\eta = x_0 \xi - c$, причем если коэффициент пропорциональности x_0 положителен, то $\text{cov}(\xi, \eta) = \sqrt{D\xi D\eta}$, а если x_0 отрицателен, то $\text{cov}(\xi, \eta) = -\sqrt{D\xi D\eta}$.

Свойствами 1–5 удобно пользоваться при вычислении ковариации от сложных выражений. Например:

$$\begin{aligned} \text{cov}(2\xi + \eta, 3\xi - 4\eta) &= 2\text{cov}(\xi, 3\xi - 4\eta) + \text{cov}(\eta, 3\xi - 4\eta) = \\ &= 6\text{cov}(\xi, \xi) - 8\text{cov}(\xi, \eta) + 3\text{cov}(\eta, \xi) - 4\text{cov}(\eta, \eta) = \\ &= 6D\xi - 5\text{cov}(\xi, \eta) - 4D\eta. \end{aligned}$$

Итак, ковариацию можно считать мерой зависимости случайных величин, так как для независимых случайных величин ковариация равна нулю. Существенным недостатком ковариации является то, что ее размерность совпадает с произведением размерностей случайных величин. Естественно, желательно иметь безразмерную характеристику зависимости. Таковой является коэффициент корреляции.

Коэффициентом корреляции случайных величин ξ и η (с положительными дисперсиями) называется число

$$\rho = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi} \sqrt{D\eta}}.$$

Коэффициент корреляции является одной из важных мер зависимости случайных величин. Как следует из свойств ковариации, он принимает значения от -1 до $+1$, отражая как силу зависимости (по абсолютной величине), так и характер (положительная или отрицательная).

Чем ближе $|\rho|$ к единице, тем с большим основанием можно считать, что ξ и η находятся в линейной зависимости, т.е. коэффициент корреляции характеризует не всякую зависимость, а только так называемую *линейную вероятностную зависимость*, которая заключается в том, что при возрастании одной случайной величины другая имеет тенденцию изменяться по линейному закону. Можно сказать, что коэффициент корреляции ρ отражает степень линейной зависимости случайных величин. С возрастанием ξ случайная величина η имеет тенденцию к увеличению при $\rho > 0$ и к уменьшению при $\rho < 0$. Поэтому при $\rho > 0$ говорят о *положительной корреляционной зависимости* ξ и η , при $\rho < 0$ — об *отрицательной*.

Для независимых случайных величин коэффициент корреляции равен нулю. Если $\rho = 0$, то случайные величины называются *некоррелированными*. Из независимости случайных величин следует их некоррелированность, но наоборот — не всегда.

Для линейно зависимых величин, т.е. в случае $\eta = a\xi + b$, где a и b — константы, коэффициент корреляции равен $+1$ при $a > 0$ и -1 при $a < 0$.

Задача 6. Для пары случайных величин из задачи 3 вычислить ковариацию $\text{cov}(\xi, \eta)$.

Решение. В предыдущей задаче уже было вычислено математическое ожидание $M\xi\eta = 19/16$. Используя полученные в решении задачи 3 частные законы распределения, получаем

$$M\xi = -1 \cdot 1/4 + 0 \cdot 1/4 + 1 \cdot 1/2 = 1/4; \quad M\eta = 1 \cdot 1/4 + 2 \cdot 3/4 = 7/4$$

и

$$\text{cov}(\xi, \eta) = M(\xi\eta) - M\xi \cdot M\eta = 19/16 - 1/4 \cdot 7/4 = 0,$$

чего и следовало ожидать вследствие независимости случайных величин.

Задача 7. Случайный вектор (ξ, η) принимает значения $(0, 0)$, $(1, 0)$, $(-1, 0)$, $(0, 1)$ и $(0, -1)$ равновероятно (рис. 5.4). Вычислить ковариацию случайных величин ξ и η . Показать, что они зависимы.

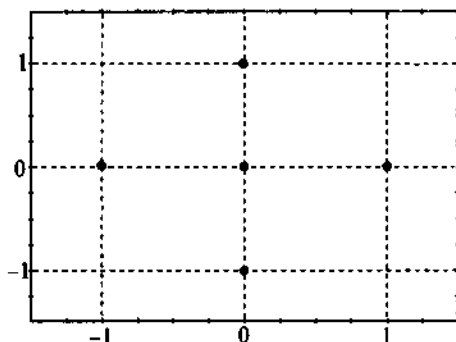


Рис. 5.4

Решение. Поскольку $P(\xi = 0) = 3/5$, $P(\xi = 1) = 1/5$, $P(\xi = -1) = 1/5$; $P(\eta = 0) = 3/5$, $P(\eta = 1) = 1/5$, $P(\eta = -1) = 1/5$, то $M\xi = 3/5 \cdot 0 + 1/5 \cdot 1 + 1/5 \cdot (-1) = 0$ и $M\eta = 0$;

$$M(\xi\eta) = 0 \cdot 0 \cdot 1/5 + 1 \cdot 0 \cdot 1/5 - 1 \cdot 0 \cdot 1/5 + 0 \cdot 1 \cdot 1/5 - 0 \cdot 1 \cdot 1/5 = 0.$$

Получаем $\text{cov}(\xi, \eta) = M(\xi\eta) - M\xi M\eta = 0$, и случайные величины некоррелированы. Однако они зависимы. Пусть $\xi = 1$, тогда условная вероятность события $\{\eta = 0\}$ равна $P(\eta = 0 | \xi = 1) = 1$ и не равна безусловной $P(\eta = 0) = 3/5$, или вероятность совместного появления $\{\xi = 0, \eta = 0\}$ не равна произведению вероятностей: $P(\xi = 0, \eta = 0) = 1/5 \neq P(\xi = 0)P(\eta = 0) = 9/25$. Следовательно, ξ и η зависимы.

Задача 8. Случайные приращения цен акций двух компаний за день ξ и η имеют совместное распределение, заданное таблицей:

ξ	η	
	-1	+1
-1	0,3	0,2
+1	0,1	0,4

Найти коэффициент корреляции.

Решение. Прежде всего вычислим $M\xi\eta = 0,3 - 0,2 - 0,1 + 0,4 = 0,4$. Далее находим частные законы распределения ξ и η :

ξ	η		
	-1	+1	p_i
-1	0,3	0,2	0,5
+1	0,1	0,4	0,5
p_j	0,4	0,6	1,0

Определяем $M\xi = 0,5 - 0,5 = 0$; $M\eta = 0,6 - 0,4 = 0,2$; $D\xi = 1$
 $D\eta = 1 - 0,2^2 = 0,96$; $\text{cov}(\xi, \eta) = 0,4$. Получаем

$$\rho = \frac{0,4}{\sqrt{1} \sqrt{0,96}} \approx 0,408.$$

Задача 9. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 1$ и $D\eta = 2$, а коэффициент их корреляции $\rho = 0,7$. Найти дисперсию приращения цены портфеля из 5 акций первой компании и 3 акций второй компании.

Решение. Используя свойства дисперсии, ковариации и определение коэффициента корреляции, получаем

$$D(5\xi + 3\eta) = 5^2 D\xi + 3^2 D\eta + 2 \cdot 5 \cdot 3 \rho \sqrt{D\xi} \sqrt{D\eta} = \\ = 25 \cdot 1 + 9 \cdot 2 + 30 \cdot 0,7 \cdot 1 \cdot \sqrt{2} \approx 72,7.$$

Замечание. Дисперсия приращений цены портфеля акций часто используется на практике как мера риска вложений: чем больше дисперсия, тем больше риск. Поэтому оценка данной величины имеет важное значение для инвесторов.

5.8. Условные распределения и математические ожидания (дискретный случай)

Пусть на одном и том же пространстве элементарных исходов Ω заданы две случайные величины ξ и η .

Условным законом распределения случайной величины ξ при условии $\eta = y$ называется любое соотношение, ставящее в соответствие значениям случайной величины ξ условные вероятности их принятия при условии $\eta = y$.

Рассмотрим здесь случай дискретных случайных величин ξ и η , принимающих значения x_i ($i = 1, 2, \dots$) и y_j ($j = 1, 2, \dots$)

соответственно. Тогда условное распределение ξ при условии $\eta = y_j$ ставит в соответствие значениям x_i вероятности

$$P_{\xi|\eta}(x_i | y_j) = P(\xi = x_i | \eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)}.$$

При этом предполагается, что $P(\eta = y_j) > 0$.

Если случайные величины независимы, то их условные распределения совпадают с исходными, и значение одной величины не влияет на распределение другой.

Условной функцией распределения случайной величины ξ при условии $\eta = y$ называется функция, ставящая в соответствие любому числу x условную вероятность события $\{\xi < x\}$ при условии $\eta = y$. В дискретном случае получаем:

$$F_{\xi|\eta}(x | y_j) = P(\xi < x | \eta = y_j) = \frac{P(\xi < x, \eta = y_j)}{P(\eta = y_j)}.$$

Условным математическим ожиданием случайной величины ξ при условии $\eta = y$ называется математическое ожидание условного распределения ξ при условии $\eta = y$. В дискретном случае получаем:

$$M(\xi | \eta = y_j) = \sum_i x_i P_{\xi|\eta}(x_i | y_j).$$

Аналогичным образом можно определить условную дисперсию.

Функцией регрессии случайной величины ξ по η называется функция, ставящая в соответствие числу y условное математическое ожидание ξ при условии $\eta = y$:

$$\varphi_{\xi|\eta}(y) = M(\xi | \eta = y).$$

Функция регрессии определена только на области возможных значений η .

На практике обычно невозможно точно предсказать значение одной случайной величины на основе знания другой, однако можно сделать это «в среднем». Функция регрессии как раз и характеризует среднее значение одной (неизвестной) величины при известном значении другой. В случае если разброс возможных значений не очень велик, это может принести большую пользу.

Условным математическим ожиданием ξ по η называется случайная величина, равная $\varphi_{\xi|\eta}(\eta)$, которая обозначается $M(\xi|\eta)$.

Условное математическое ожидание обладает следующими свойствами:

$$1) M(c|\eta) \equiv c;$$

$$2) M(ax + b|\eta) = aM(x|\eta) + b;$$

$$3) M(\xi_1 + \xi_2|\eta) = M(\xi_1|\eta) + M(\xi_2|\eta);$$

$$4) M\xi = M[M(\xi|\eta)];$$

5) $M[f(\xi)h(\eta)|\eta] = h(\eta)M[f(\xi)|\eta]$, где $f(\xi)$ и $h(\eta)$ — произвольные функции от случайных величин ξ и η ;

$$6) M(\xi|\eta) \equiv M(\xi), \text{ если } \xi \text{ и } \eta \text{ независимы.}$$

Задача 10. Распределение двумерной случайной величины задано таблицей:

η	ξ			
	1	3	4	8
3	0,15	0,06	0,25	0,04
6	0,30	0,10	0,03	0,07

Найти условное распределение и условное математическое ожидание η при $\xi = 1$.

Решение. Условное математическое ожидание равно

$$M(\eta|\xi = x_1) = y_1 P(y_1|x_1) + y_2 P(y_2|x_1).$$

Из условия задачи найдем распределение составляющих η и ξ (последний столбец и последняя строка таблицы):

η	ξ				P_η
	1	3	4	8	
3	0,15	0,06	0,25	0,04	0,50
6	0,30	0,10	0,03	0,07	0,50
P_ξ	0,45	0,16	0,28	0,11	1,00

Так как $P_\xi(x_1) = P(x_1, y_1) + P(x_1, y_2) = 0,15 + 0,30 = 0,45$, то условные вероятности находятся по формулам

$$P(y_1|x_1) = \frac{P(x_1, y_1)}{P_\xi(x_1)} = \frac{0,15}{0,45} = \frac{1}{3}, \quad P(y_2|x_1) = \frac{P(x_1, y_2)}{P_\xi(x_1)} = \frac{0,30}{0,45} = \frac{2}{3},$$

а искомое условное математическое ожидание равно

$$M(\eta|\xi = 1) = 3 \cdot \frac{1}{3} + 6 \cdot \frac{2}{3} = 5.$$

Задачи для самостоятельного решения**Теоретические задачи**

1. Докажите, что для случайной величины ξ , распределенной по закону Пуассона с параметром λ , математическое ожидание $M\xi = \lambda$, а дисперсия $D\xi = \lambda$.
2. Докажите, что для случайной величины ξ , распределенной по закону Бернулли с параметрами n и p , математическое ожидание $M\xi = np$, а дисперсия $D\xi = np(1-p)$.
3. Докажите, что для случайной величины ξ , распределенной по геометрическому закону с параметром p , математическое ожидание $M\xi = \frac{1}{p}$, а дисперсия $D\xi = \frac{1-p}{p^2}$.

Вычислительные задачи

4. Монету подбросили 3 раза. Найти распределение вероятностей числа появлений герба.
5. Вероятность, что лотерейный билет окажется выигрышным, равна 0,1. Покупатель купил 5 билетов. Найти распределение числа выигрышей у владельца этих 5 билетов.
6. Три стрелка с вероятностями попадания в цель при отдельном выстреле 0,7, 0,8 и 0,9 соответственно делают по одному выстрелу. Найти распределение вероятностей общего числа попаданий.
7. Два стрелка поражают мишень с вероятностями 0,8 и 0,9 соответственно (при одном выстреле). Найти распределение общего числа попаданий в мишень, если первый стрелок выстрелил один раз, а второй — два раза.
8. Два станка выпускают детали с вероятностями брака 0,01 и 0,05 соответственно. В выборке одна деталь выпущена первым станком и две — вторым. Найти распределение числа бракованных деталей в выборке.
9. Прибор комплектуется из двух деталей, вероятность брака для первой детали — 0,1, а для второй — 0,05. Выбрано 4 прибора. Прибор считается бракованным, если в нем есть хотя бы одна бракованная деталь. Найти распределение числа бракованных приборов среди выбранных.
10. Стрелок поражает мишень с вероятностью 0,7 при одном выстреле. Стрелок стреляет до первого попадания, но делает не более 3 выстрелов. Найти распределение числа выстрелов.

11. Монета подбрасывается до тех пор, пока герб не выпадет 2 раза, но при этом делается не более 4 бросаний. Найти распределение числа подбрасываний.
12. С конвейера поступило N деталей. Вероятность брака для каждой детали равна p . Детали проверяют одну за другой, пока не наберут M годных или пока они не кончатся. Найти распределение числа проверенных деталей, если: а) $N = 4$, $p = 0,1$, $M = 2$; б) $N = 5$, $p = 0,1$, $M = 3$; в) $N = 6$, $p = 0,05$, $M = 3$; г) $N = 6$, $p = 0,2$, $M = 2$; д) $N = 6$, $p = 0,1$, $M = 2$.
13. Среди N деталей — M нужного размера. Детали извлекают поочередно, пока не найдут K деталей нужного размера или пока не будет сделано L проб. Найти распределение числа извлеченных деталей, если: а) $N = 10$, $M = 3$, $K = 2$, $L = 4$; б) $N = 12$, $M = 6$, $K = 2$, $L = 5$; в) $N = 13$, $M = 6$, $K = 3$, $L = 5$; г) $N = 12$, $M = 7$, $K = 2$, $L = 5$; д) $N = 8$, $M = 3$, $K = 2$, $L = 5$; е) $N = 10$, $M = 4$, $K = 2$, $L = 5$; ж) $N = 10$, $M = 5$, $K = 2$, $L = 4$.
14. Среди 5 ключей два подходят к двери. Ключи пробуют один за другим, пока не откроют дверь. Найти распределение числа опробованных ключей.
15. Среди 6 ключей три подходят к двери. Ключи подбираются до тех пор, пока дверь не будет открыта. Найти распределение числа опробованных ключей.
16. Курс акции в течение дня торгов может подняться или опуститься на один пункт, либо остаться неизменным (все три варианта равновероятны). Найти распределение изменения курса акции: а) за 2 дня; б) за 3 дня.
17. Курс акции в течение дня торгов может подняться на один пункт с вероятностью p , опуститься на один пункт с вероятностью r либо остаться неизменным с вероятностью s . Найти распределение изменения цены акции за 2 дня. Вычислить математическое ожидание. Решить задачу, если: а) $p = 0,6$, $r = 0,3$, $s = 0,1$; б) $p = 0,5$, $r = 0,4$, $s = 0,1$; в) $p = 0,7$, $r = 0,2$, $s = 0,1$.
18. В телеигре игроку задают вопросы. Если игрок правильно отвечает на вопрос, ему задают следующий; если неправильно, то игрок выбывает из игры. Всего задается не более трех вопросов. Вероятность ответить на первый вопрос равна 0,9; на второй — 0,3; на третий — 0,1. Найти распределение числа правильных ответов и математическое ожидание выигрыша, если за один правильный ответ платят 100 руб., за два — 400 руб. и за три — 1000 руб.
19. На викторине задаются 3 вопроса. Вероятность правильно ответить на первый — p , на второй — r , на третий — s . После

- неправильного ответа игрок выбывает из игры. Найти распределение числа правильных ответов. Вычислить математическое ожидание. Решить задачу, если: а) $p = 0,7$, $r = 0,6$, $s = 0,3$; б) $p = 0,8$, $r = 0,4$, $s = 0,1$.
20. На викторине задаются 3 вопроса. Вероятность правильно ответить на первый — $0,7$, на второй — $0,5$, на третий — $0,2$. После неправильного ответа игрок выбывает из игры. Найти распределение числа заданных вопросов. Вычислить математическое ожидание.
21. На викторине задаются 4 вопроса. Вероятность правильно ответить на первый — p , на второй — r , на третий — s , на четвертый — w . После неправильного ответа игрок выбывает из игры. Найти распределение числа правильных ответов. Вычислить математическое ожидание. Решить задачу, если: а) $p = 0,9$, $r = 0,8$, $s = 0,4$, $w = 0,2$; б) $p = 0,9$, $r = 0,8$, $s = 0,3$, $w = 0,2$; в) $p = 0,9$, $r = 0,7$, $s = 0,4$, $w = 0,2$.
22. В офисе проводится собеседование с N кандидатами на некоторую должность (по очереди). Если подходящий человек найден (принято решение о приеме его на работу), то с оставшимися кандидатами собеседование не проводится. Вероятность того, что кандидат подходит, равна p . Найти распределение числа кандидатов, с которыми беседовали, и его математическое ожидание, если: а) $N = 4$, $p = 0,2$; б) $N = 5$, $p = 0,3$; в) $N = 6$, $p = 0,2$.
23. В игровом автомате три окошка, в которых случайным образом появляются цифры от 0 до 9, равновероятно и независимо друг от друга. Если две цифры совпали, игрок получает 10 руб., если все три — 100 руб. Чтобы начать игру, он платит 5 руб. Найти распределение выигрыша игрока и его математическое ожидание.
24. В игровом автомате три окошка, в которых случайным образом появляются 7 цветов, равновероятно и независимо друг от друга. Если два цвета совпали, игрок получает 20 руб., если три — 100 руб. Чтобы начать игру, он платит 10 руб. Найти распределение выигрыша игрока. Вычислить математическое ожидание.
25. В экзаменационном билете три задачи. Вероятность правильного решения студентом первой задачи равна p , второй — r и третьей — s . Найти распределение числа правильно решенных задач и найти его математическое ожидание, если: а) $p = 0,8$, $r = 0,7$, $s = 0,3$; б) $p = 0,9$, $r = 0,5$, $s = 0,3$; в) $p = 0,8$, $r = 0,6$, $s = 0,3$.
26. Фирма участвует в трех независимых проектах. Вероятность успешного завершения первого равна p , второго — r , третьего — s . Найти распределение числа успешных проектов. Вычислить

математическое ожидание. Решить задачу, если: а) $p = 0,9$, $r = 0,7$, $s = 0,6$; б) $p = 0,8$, $r = 0,7$, $s = 0,5$.

27. Фирма участвует в четырех независимых проектах. Вероятность успешного завершения первого и второго равна 0,8, а третьего и четвертого — 0,7. Найти распределение числа успешных проектов. Вычислить математическое ожидание.
28. В папке находятся 4 документа: два заявления и две анкеты. В заявлении могут быть ошибки с вероятностью 0,1. В анкете могут быть ошибки с вероятностью 0,2. Найти распределение числа документов с ошибками в папке и его математическое ожидание.
29. Стрелок стреляет по движущейся мишени до первого попадания в нее, причем успевает сделать не более 4 выстрелов. Найти математическое ожидание и дисперсию числа сделанных выстрелов, если вероятность попадания при каждом выстреле равна 0,6.
30. Найти математическое ожидание и дисперсию суммы выпавших очков при бросании четырех игральных костей.
31. На станцию обслуживания поступают заявки в соответствии с распределением Пуассона с параметром $\lambda = 2$ (в единицу времени). Мощность станции позволяет обслуживать не более 2 заявок в единицу времени. Найти вероятность того, что в течение данной единицы времени: а) станция не справится с потоком заказов и образуется очередь; б) станция обслуживания будет простаивать или работать не на полную мощность; в) на станции обслуживания не образуется очередь. Решить задачу без учета заявок, возможно, оставшихся с прошлой единицы времени.
32. В процессе производства изделие высшего качества удается получить только с вероятностью 0,2. С конвейера берутся наугад детали до тех пор, пока не будет взято изделие высшего качества. Найти математическое ожидание числа проверенных изделий.
33. Сдача экзамена по математике производится до получения положительного результата. Шансы сдать экзамен остаются неизменными и составляют 25%. Найти математическое ожидание числа попыток сдачи экзамена.
34. ОТК должен проверить 100 комплектов, состоящих из 4 детали каждый. Найти математическое ожидание числа комплектов, состоящих только из стандартных деталей, если каждая деталь стандартна с вероятностью 0,8.

35. Игральная кость подбрасывается до а) второго; б) третьего появления грани с номером «три». Найти среднее число подбрасываний.
36. В каждой упаковке товара имеется одна из N различных наклеек (равновероятно). Вычислить математическое ожидание и дисперсию числа упаковок, сколько понадобится купить, чтобы собрать все наклейки. Решить задачу, если: а) $N = 3$; б) $N = 4$; в) $N = 5$.
37. Закон распределения случайной величины x имеет вид:

ξ	0	1	2	3
P	1/8	3/8	3/8	1/8

Найти функцию распределения случайной величины ξ , вычислить ее математическое ожидание, дисперсию и среднее квадратическое отклонение. Вычислить вероятность $P\{-1 < \xi < 3/2\}$.

38. Закон распределения случайной величины ξ имеет вид:

ξ	-1	2	3	5
P	1/4	1/2	1/8	1/8

Найти функцию распределения случайной величины ξ , вычислить ее математическое ожидание, дисперсию и среднее квадратическое отклонение. Вычислить вероятность $P\{5/2 < \xi < 5\}$.

39. Закон распределения случайной величины ξ имеет вид:

ξ	0	1	3	4
P	1/9	2/9	1/6	1/2

Найти функцию распределения случайной величины ξ , вычислить ее математическое ожидание, дисперсию и среднее квадратическое отклонение. Вычислить вероятность $P\{1/2 < \xi < 7/2\}$.

40. Закон распределения случайной величины ξ имеет вид:

ξ	-2	-1	1	3
P	1/7	3/7	2/7	1/7

Найти функцию распределения случайной величины ξ , вычислить ее математическое ожидание, дисперсию и среднее квадратическое отклонение. Вычислить вероятность $P\{-3/2 < \xi < 2\}$.

41. Совместное распределение ξ и η задано следующей таблицей:

ξ	η	
	0	1
0	$\frac{1}{4}$	$\frac{1}{2} - C$
1	$\frac{1}{4}$	C

Найти C , при котором случайные величины независимы.

42. Совместное распределение ξ и η задано следующей таблицей:

ξ	η	
	-1	1
0	$\frac{1}{3} - C$	$\frac{1}{6}$
2	$\frac{1}{6} + C$	$\frac{1}{3}$

Найти C , при котором случайные величины независимы.

43. Совместное распределение ξ и η задано следующей таблицей:

ξ	η	
	-1	1
-1	$\frac{1}{3}$	C
1	$\frac{1}{2} - C$	$\frac{1}{6}$

Найти C , при котором случайные величины независимы.

44. Совместное распределение ξ и η задано следующей таблицей:

ξ	η	
	0	1
1	$\frac{1}{3} + C$	$\frac{1}{4}$
2	$\frac{1}{6}$	$\frac{1}{4} - C$

Найти C , при котором случайные величины независимы.

45. Совместный закон распределения пары (ξ, η) задан таблицей.

$\xi \backslash \eta$	-1	0	1
0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$
1	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{5}$

Найти распределение вероятностей случайной величины $\xi - \eta$ и вычислить $\text{cov}(\xi + \eta, \xi - \eta)$. Исследовать вопрос о зависимости случайных величин ξ и η .

46. Совместный закон распределения пары (ξ, η) задан таблицей.

$\xi \backslash \eta$	-1	0	1
-1	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{6}$
1	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{6}$

Найти закон распределения вероятностей случайной величины $\xi\eta$ и вычислить $\text{cov}(2\xi - 3\eta, \xi + 2\eta)$. Исследовать вопрос о зависимости случайных величин ξ и η .

47. Совместный закон распределения пары (ξ, η) задан таблицей.

$\xi \backslash \eta$	0	1	2
-1	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{3}$
1	$\frac{1}{9}$	$\frac{1}{18}$	$\frac{2}{9}$

Найти закон распределения вероятностей случайной величины $\xi + \eta$ и вычислить $\text{cov}(2\eta + \xi, \eta + \xi)$. Исследовать вопрос о зависимости случайных величин ξ и η .

48. Совместный закон распределения пары (ξ, η) задан таблицей.

$\xi \backslash \eta$	-1	0	2
0	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{3}{8}$

Найти закон распределения вероятностей случайной величины $\xi + \eta$ и вычислить $\text{cov}(\xi - \eta, 2\xi + \eta)$. Исследовать вопрос о зависимости случайных величин ξ и η .

49. Совместный закон распределения пары (ξ, η) задан таблицей

ξ	η		
	-1	1	2
-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{7}{18}$

Найти закон распределения вероятностей случайной величины $\xi\eta$ и вычислить $\text{cov}(2\xi + \eta, 3\xi - \eta)$. Исследовать вопрос о зависимости случайных величин ξ и η .

50. Совместное распределение случайных величин ξ и η задано таблицей

ξ	η		
	-2	0	3
0	0,1	0,2	0,2
1	0,1	0,1	0,3

Зависимы ли эти случайные величины? Найти распределение их суммы. Вычислить $\text{cov}(\eta, \eta + 4\xi)$.

51. Совместный закон распределения пары ξ, η задан таблицей

ξ	η		
	-2	-1	0
-1	0,15	0,1	0,05
2	0,4	0,2	0,1

Зависимы ли ξ и η ? Найти распределение их разности. Вычислить $\text{cov}(\xi + 5\eta, 2\xi - 7\eta)$.

52. Совместный закон распределения пары ξ, η задан таблицей

ξ	η		
	-1	0	2
0	$1/7$	$2/7$	$1/7$
1	$1/7$	$1/7$	$1/7$

Зависимы ли ξ и η ? Найти распределение их разности. Вычислить $\text{cov}(\xi - 3\eta, \eta + 6\xi)$.

53. Совместный закон распределения пары
- ξ, η
- задан таблицей

ξ	η		
	-1	0	2
-2	2/15	1/15	2/15
1	4/15	1/15	2/5

Зависимы ли ξ и η ? Найти распределение их разности. Вычислить $\text{cov}(4\xi - \eta, 2\eta + \xi)$.

54. Совместный закон распределения пары
- ξ, η
- задан таблицей

ξ	η		
	1	2	4
0	1/6	1/6	1/12
1	1/4	1/12	1/4

Зависимы ли ξ и η ? Найти распределение их разности. Вычислить $\text{cov}(2\xi - \eta, \eta - 3\xi)$.

55. Случайные приращения цен акций двух компаний за день
- ξ
- и
- η
- имеют совместное распределение, заданное таблицей

$\xi \backslash \eta$	-1	+1
-1	0,2	0,1
+1	0,2	0,5

Найти коэффициент корреляции.

56. Случайные приращения цен акций двух компаний за день
- ξ
- и
- η
- имеют совместное распределение, заданное таблицей.

$\xi \backslash \eta$	-1	+1
-1	0,4	0,1
+1	0,1	0,4

Найти коэффициент корреляции.

57. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 1,21$ и $D\eta = 2,56$, а коэффициент их корреляции $\rho = 0,5$. Найти дисперсию приращения цены портфеля из: а) 4 акций первой компании и 6 акций второй компании; б) 7 акций первой компании и 3 акций второй компании; в) 9 акций первой компании и 1 акции второй компании.
58. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 2$ и $D\eta = 3$, причем они некоррелированы.

- Инвестор намеревается приобрести 10 акций. Сколько акций каждой компании он должен купить, чтобы минимизировать риск вложений, т.е. дисперсию приращения цены портфеля?
59. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 1$ и $D\eta = 3$, причем они некоррелированы. Инвестор намеревается приобрести 12 акций. Сколько акций каждой компании он должен купить, чтобы минимизировать риск вложений?
60. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 4$ и $D\eta = 9$, а их коэффициент корреляции 0,5. В каких долях следует разделить капитал при вложении в акции, чтобы минимизировать риск?
61. Случайные приращения цен акций двух компаний за день имеют дисперсии $D\xi = 1$ и $D\eta = 4$, а их коэффициент корреляции 0,7. В каких долях следует разделить капитал при вложении в акции, чтобы минимизировать риск?
62. По таблице совместного распределения из задачи 45: а) найти условное распределение ξ при условии $\eta = 0$; б) найти условное распределение η при условии $\xi = 0$; в) функцию регрессии ξ по η ; г) функцию регрессии η по ξ .
63. По таблице совместного распределения из задачи 46: а) найти условное распределение ξ при условии $\eta = -1$; б) найти условное распределение η при условии $\xi = 1$; в) функцию регрессии ξ по η ; г) функцию регрессии η по ξ .
64. По таблице совместного распределения из задачи 47: а) найти условное распределение ξ при условии $\eta = 2$; б) найти условное распределение η при условии $\xi = -1$; в) функцию регрессии ξ по η ; г) функцию регрессии η по ξ .
65. Две независимые случайные величины ξ_1 и ξ_2 имеют распределения Пуассона с параметрами λ_1 и λ_2 соответственно. Найти условное распределение ξ_1 при условии, что $\eta = \xi_1 + \xi_2 = n$, $n > 0$, и функцию регрессии ξ_1 по η .

ГЛАВА 6

НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

6.1. Плотность и функция распределения непрерывной случайной величины

Случайная величина ξ называется *непрерывной*, если ее функция распределения $F_{\xi}(x) = P(\xi < x)$ непрерывна. Множество значений непрерывной случайной величины несчетно и обычно представляет собой некоторый промежуток (конечный или бесконечный). Функция распределения непрерывной случайной величины обладает теми же основными свойствами, что и в дискретном случае (§ 5.2).

Случайная величина ξ называется *абсолютно непрерывной*, если существует неотрицательная функция $p_{\xi}(x)$, такая, что при любых x функцию распределения $F_{\xi}(x)$ можно представить в виде интеграла

$$F_{\xi}(x) = P(\xi < x) = \int_{-\infty}^x p_{\xi}(t) dt.$$

Функция $p_{\xi}(x)$ называется *плотностью распределения*.

Имеют место следующие свойства:

1. $p_{\xi}(x) \geq 0$.
2. В точках непрерывности плотность распределения равна производной функции распределения: $p_{\xi}(x) = F'_{\xi}(x)$.

3. Интеграл по всей числовой прямой от плотности распределения равен единице: $\int_{-\infty}^{\infty} p_{\xi}(x) dx = 1$.

4. Плотность распределения определяет закон распределения случайной величины, так как определяет вероятность попадания случайной величины на любой полуинтервал $[a, b)$:

$$P\{\xi \in [a, b)\} = P\{a \leq \xi < b\} = F_{\xi}(b) - F_{\xi}(a) = \int_a^b p_{\xi}(t) dt.$$

5. Вероятность того, что непрерывная случайная величина примет конкретное значение a , равна нулю: $P(\xi = a) = 0$. Поэтому справедливы следующие равенства:

$$P\{a \leq \xi < b\} = P\{a < \xi \leq b\} = P\{a \leq \xi \leq b\} = P\{a < \xi < b\} = F_{\xi}(b) - F_{\xi}(a).$$

Доказательство.

1. Следует из определения.
2. Следует из взаимосвязи между производной и интегралом.

$$3. \int_{-\infty}^{\infty} p_{\xi}(x) dx = \lim_{x \rightarrow +\infty} \int_{-\infty}^x p_{\xi}(x) dx = \lim_{x \rightarrow +\infty} F_{\xi}(x) = 1.$$

4. Используем формулу Ньютона—Лейбница.

5. Представим событие $A = \{\omega: \xi(\omega) = a\}$ в виде произведения $\bigcap_{n=1}^{\infty} \{a \leq \xi < a + 1/n\}$ и воспользуемся аксиомой непрерывности (§ 3.7), тогда

$$P(A) = P(\xi = a) = \lim_{n \rightarrow \infty} P(a \leq \xi < a + 1/n) = \lim_{n \rightarrow \infty} \int_a^{a+\frac{1}{n}} p_{\xi}(x) dx = 0.$$

Заметим, что плотность может быть разрывна в некоторых точках, где ее можно определить произвольным образом: это никак не повлияет на функцию распределения и другие числовые характеристики. Поэтому можно считать, что множество значений случайной величины совпадает с промежутком, на котором плотность распределения отлична от нуля.

График плотности распределения называется *кривой распределения*. Площадь, ограниченная кривой распределения и осью абсцисс, как следует из свойства 3, равна единице. Тогда значение функции распределения $F_{\xi}(x)$ в точке x_0 геометрически есть площадь, ограниченная кривой распределения и осью абсцисс и лежащая левее точки x_0 (рис. 6.1).

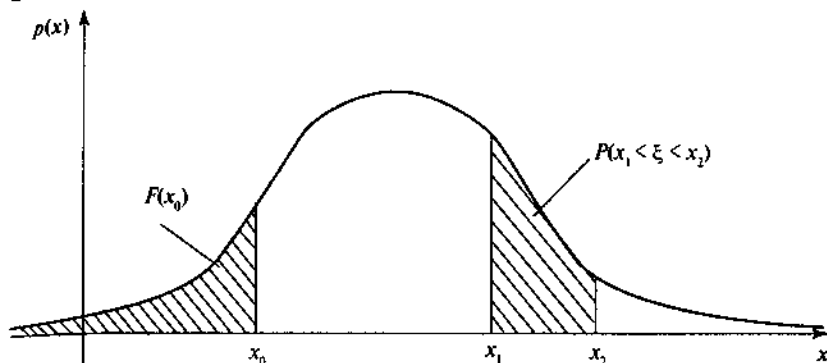


Рис. 6.1

Задача 1. Плотность распределения непрерывной случайной величины имеет вид:

$$p_{\xi}(x) = \begin{cases} 0, & x \notin [0, 2]; \\ Cx^2, & x \in [0, 2]. \end{cases}$$

Определить константу C , построить функцию распределения $F_{\xi}(x)$ и вычислить вероятность $P(-1 \leq \xi \leq 1)$.

Решение. Константа C находится из условия $\int_{-\infty}^{\infty} p_{\xi}(x) dx = 1$. В результате имеем

$$1 = \int_{-\infty}^{\infty} p_{\xi}(x) dx = \int_0^2 Cx^2 dx = C \frac{x^3}{3} \Big|_0^2 = \frac{8C}{3}, \quad \text{откуда } C = 3/8.$$

Чтобы построить функцию распределения $F_{\xi}(x)$, отметим, что отрезок $[0, 2]$ делит область значений аргумента x (числовую ось) на три части: $(-\infty, 0)$, $[0, 2]$, $(2, \infty)$. Рассмотрим каждый из этих промежутков. В первом случае (когда $x < 0$) вероятность события $\{\xi < x\}$ вычисляется так:

$$F_{\xi}(x) = P(\xi < x) = \int_{-\infty}^x p_{\xi}(t) dt = \int_{-\infty}^x 0 dt = 0,$$

так как плотность ξ на полуоси $(-\infty, 0)$ равна нулю. Во втором случае

$$F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(t) dt = \int_{-\infty}^0 p_{\xi}(t) dt + \int_0^x p_{\xi}(t) dt = 0 + \frac{3}{8} \int_0^x t^2 dt = \frac{x^3}{8}.$$

Наконец, в последнем случае, когда $x > 2$,

$$F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(t) dt = \int_{-\infty}^0 p_{\xi}(t) dt + \int_0^2 p_{\xi}(t) dt + \int_2^x p_{\xi}(t) dt = 0 + \frac{3}{8} \int_0^2 t^2 dt = 0 + 1 + 0 = 1,$$

так как плотность $f_{\xi}(x)$ обращается в нуль на полуоси $(2, \infty)$.
Итак, получена функция распределения

$$F_{\xi}(x) = \begin{cases} 0, & x < 0; \\ \frac{x^3}{8}, & 0 \leq x \leq 2; \\ 1, & x > 2. \end{cases}$$

Следовательно, $P\{-1 \leq \xi \leq 1\} = F(1) - F(-1) = 1/8 - 0 = 1/8$.

6.2. Числовые характеристики непрерывной случайной величины

Математическое ожидание для непрерывно распределенных случайных величин определяется по формуле

$$M\xi = \int_{-\infty}^{\infty} x p_{\xi}(x) dx,$$

при этом интеграл, стоящий справа, должен абсолютно сходиться.

Если $\varphi(x)$ — некоторая функция и ξ имеет плотность $p(x)$, то математическое ожидание случайной величины $\varphi(\xi)$ вычисляется по формуле

$$M\varphi(\xi) = \int_{-\infty}^{\infty} \varphi(x) p(x) dx,$$

если интеграл абсолютно сходится.

Дисперсия непрерывной случайной величины ξ вычисляется по формуле

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2 p(x) dx,$$

или, как и в дискретном случае, по формуле

$$D\xi = M\xi^2 - (M\xi)^2, \text{ где } M\xi^2 = \int_{-\infty}^{\infty} x^2 p(x) dx.$$

Все свойства математического ожидания и дисперсии, ковариации и коэффициента корреляции, приведенные в гл. 5 для дискретных случайных величин, справедливы и для непрерывных случайных величин.

Задача 2. Для случайной величины ξ из задачи 1 вычислить математическое ожидание и дисперсию.

Решение.

$$M\xi = \int_{-\infty}^{\infty} x p_{\xi}(x) dx = \int_{-\infty}^0 x \cdot 0 dx + \frac{3}{8} \int_0^2 x \cdot x^2 dx + \int_2^{\infty} x \cdot 0 dx = \frac{3}{8} \cdot \frac{x^4}{4} \Big|_0^2 = \frac{3}{2}.$$

Далее

$$M\xi^2 = \int_{-\infty}^{\infty} x^2 p_{\xi}(x) dx = \frac{3}{8} \int_0^2 x^2 \cdot x^2 dx = \frac{3}{8} \cdot \frac{x^5}{5} \Big|_0^2 = \frac{12}{5}, \text{ и значит,}$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{12}{5} - \left(\frac{3}{2}\right)^2 = 0,15.$$

В качестве других характеристик, описывающих свойства распределения случайной величины, используются начальные и центральные моменты.

Начальным моментом k -го порядка называется математическое ожидание k -й степени случайной величины: $\alpha_k = M\xi^k$.

Очевидно, что $M\xi = \alpha_1$.

Центральным моментом k -го порядка μ_k называют математическое ожидание k -й степени отклонения случайной величины от ее математического ожидания: $\mu_k = M(\xi - M\xi)^k$.

Из определения следует, что $\mu_2 = M(\xi - M\xi)^2 = D\xi$.

Справедливы формулы:

$$\mu_0 = 1;$$

$$\mu_1 = 0;$$

$$\mu_2 = \alpha_2 - \alpha_1^2;$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3;$$

$$\mu_4 = \alpha_4 + 6\alpha_1^2\alpha_2 - 4\alpha_1\alpha_3 - 3\alpha_1^4, \dots$$

Если распределение симметрично относительно математического ожидания, то все центральные моменты нечетного порядка равны нулю.

Величина $\beta_\xi = \mu_3/\sigma^3$ называется *коэффициентом асимметрии* случайной величины ξ . Он характеризует «скошенность» распределения по отношению к математическому ожиданию. Для симметричных распределений $\mu_3 = 0$, поэтому коэффициент асимметрии равен нулю. Для распределений скошенных влево $\beta < 0$, для скошенных вправо $\beta > 0$.

Величина $\gamma_\xi = \mu_4/\sigma^4 - 3$ называется *коэффициентом эксцесса* (или просто эксцесс) случайной величины ξ . Он характеризует «сглаженность», или «крутость» распределения по отношению к нормальному, так как для нормального закона распределения $\mu_4/\sigma^4 = 3$ и, следовательно, $\gamma_\xi = 0$. Для более островершинных распределений $\gamma_\xi > 0$, для менее островершинных $\gamma_\xi < 0$ (рис. 6.2).

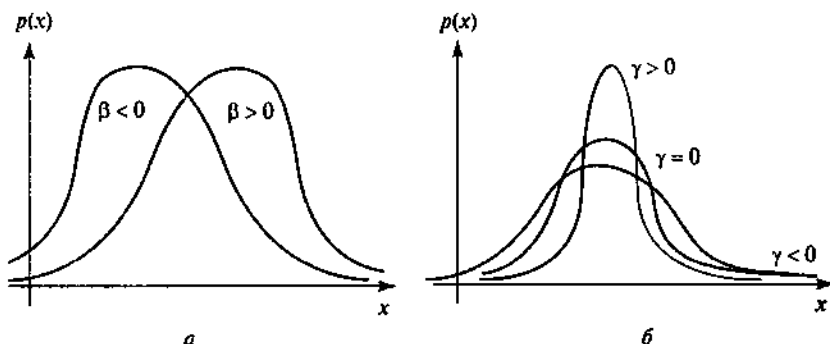


Рис. 6.2

Кроме начальных и центральных моментов на практике применяются так называемые *абсолютные моменты*, определяемые формулами

$$\bar{\alpha}'_k = M|\xi|^k, \quad \bar{\mu}'_k = M|\xi - M\xi|^k.$$

Очевидно, что абсолютные моменты четного порядка совпадают с обычными моментами. Чаще других применяется 1-й абсолютный момент $\mu'_1 = M|\xi - M\xi|$, называемый *средним абсолютным отклонением*.

Для непрерывной величины *модой* $X_{\text{мод}}$ называют точку локального максимума функции плотности распределения вероятностей. Если имеется один максимум, то распределение

называется *унимодальным*, более одного максимума — *мульти-модальным* (в частности, распределение, имеющее две моды, называется *бимодальным*). Распределение, имеющее минимум плотности, называется *антимодальным*.

Медианой непрерывной случайной величины ξ называют такое ее значение X_{med} , что $P(\xi < X_{\text{med}}) = P(\xi > X_{\text{med}}) = 1/2$. Геометрически медиана — это абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам (рис. 6.3).

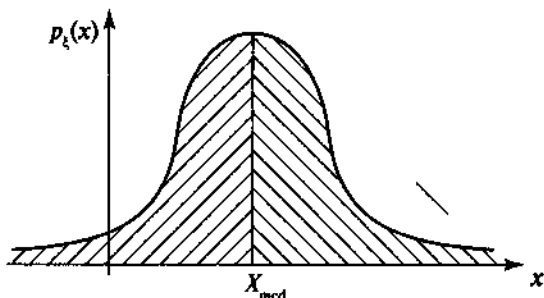


Рис. 6.3

Математическое ожидание, мода, медиана, начальные и центральные моменты являются основными числовыми характеристиками случайной величины. На практике числовыми характеристиками часто пользуются для приближенной замены одного закона распределения другим, но так, чтобы сохранились неизменными несколько важнейших моментов.

6.3. Производящая функция моментов

Производящей функцией моментов случайной величины ξ называется функция параметра t , равная математическому ожиданию функции $e^{t\xi}$: $m_\xi(t) = Me^{t\xi}$.

Наиболее важным ее свойством является то, что производящая функция $m_\xi(t)$ содержит в себе сведения обо всех начальных моментах («производит» моменты). Действительно, $m_\xi'(0) = Me^{t\xi}|_{t=0} = M\xi = \alpha_1$. В общем случае для любого k получаем $m_\xi^{(k)}(0) = Me^{t\xi}|_{t=0} = \alpha_k$, т.е. k -я производная от производящей функции при $t = 0$ равна начальному моменту k -го порядка, а любой центральный момент можно выразить через начальные моменты.

Свойства производящей функции следующие:

$$1. m_{\xi+a}(t) = Me^{(at+a)} = m_{\xi}(ct)e^a.$$

2. Производящая функция суммы независимых случайных величин равна произведению производящих функций этих величин.

Действительно, пусть $\xi = \sum_{i=1}^n \xi_i$, тогда

$$m_{\xi}(t) = M \exp \left(\sum_{i=1}^n t\xi_i \right) = M \prod_{i=1}^n \exp (t\xi_i) = \prod_{i=1}^n M \exp (t\xi_i) = \prod_{i=1}^n m_{\xi_i}(t).$$

По производящей функции можно определить функцию распределения, содержащую все сведения о случайной величине. В этом смысле производящая функция и функция распределения являются эквивалентными обобщающими характеристиками случайной величины. К сожалению, производящая функция моментов может быть определена не при всех значениях параметра (или вообще определена только в нуле).

6.4. Примеры непрерывных случайных величин

Равномерное распределение. Непрерывная случайная величина ξ имеет равномерное распределение на отрезке $[a, b]$, если плотность распределения $p_{\xi}(x)$ сохраняет постоянное значение на этом промежутке:

$$p_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Графики функции и плотности равномерного распределения показаны на рис. 6.4.

Функция распределения $F_{\xi}(x)$ равномерно распределенной случайной величины равна:

$$F_{\xi}(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b; \end{cases}$$

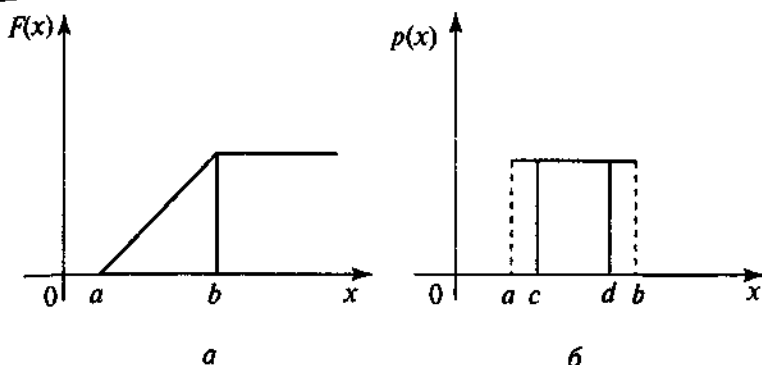


Рис. 6.4. Функция и плотность равномерного распределения

Математическое ожидание и дисперсия соответственно равны: $M\xi = \frac{a+b}{2}$; $D\xi = \frac{(b-a)^2}{12}$.

В силу симметричности равномерного распределения относительно математического ожидания асимметрия равна нулю, $X_{\text{med}} = M\xi$, а моды равномерное распределение не имеет.

Для определения эксцесса вычислим четвертый центральный момент: $\mu_4 = \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2}\right)^4 dx = \frac{(b-a)^4}{80}$. Отсюда эксцесс равен $\gamma_\xi = \frac{\mu_4}{\sigma^4} - 3 = -1,2$.

Вероятность попадания случайной величины ξ на отрезок $[c, d] \subset [a, b]$ равна $P(c \leq \xi \leq d) = F(d) - F(c) = \frac{c-d}{b-a}$, т.е. зависит только от длины отрезка и не зависит от того, где этот отрезок расположен. Таким образом, равномерное распределение реализует принцип геометрической вероятности при бросании точки на отрезок $[a, b]$.

Показательное (экспоненциальное) распределение. Непрерывная случайная величина ξ , принимающая неотрицательные значения, имеет показательное распределение с параметром $\lambda > 0$, если плотность распределения вероятностей случайной величины равна

$$p_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Функция показательного распределения имеет вид

$$F_{\xi}(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Графики плотности и функции стандартного показательного распределения представлены на рис. 6.5 (при $\lambda = 1$), общий вид — на рис. 6.6.

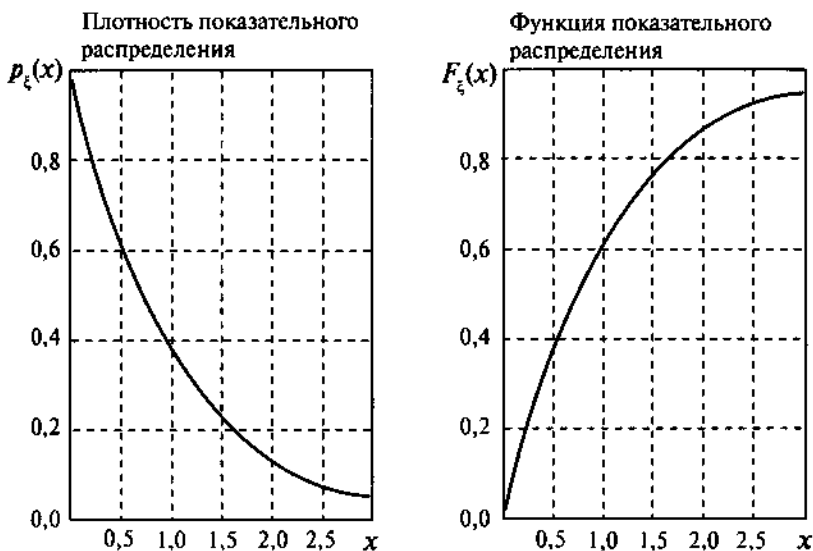


Рис. 6.5

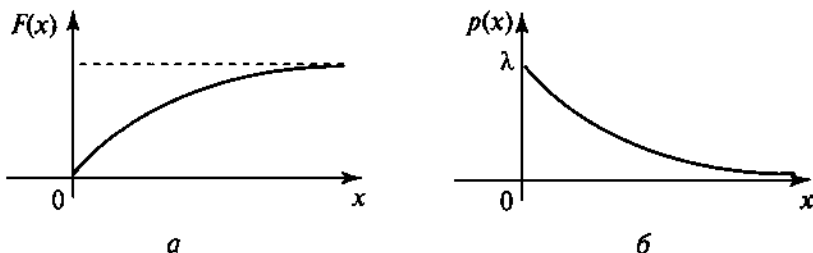


Рис. 6.6

Математическое ожидание и дисперсия равны:

$$M\xi = \frac{1}{\lambda}, \quad D\xi = \frac{1}{\lambda^2}.$$

Действительно, имеем

$$\begin{aligned} M\xi &= \int_0^{\infty} \lambda e^{-\lambda t} dt = |\lambda t = z, dz = \lambda dt| = \frac{1}{\lambda} \int_0^{\infty} z e^{-z} dz = \\ &= \frac{1}{\lambda} (-ze^{-z}|_0^{\infty} + \int_0^{\infty} e^{-z} dz) = \frac{1}{\lambda}; \end{aligned}$$

$$\begin{aligned} D\xi &= M\xi^2 - (M\xi)^2 = \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \int_0^{\infty} z^2 e^{-z} dz - \frac{1}{\lambda^2} = \\ &= 2 \frac{1}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

Можно показать также, что $\mu_3 = \frac{6}{\lambda^3}$, асимметрия $\beta_\xi = 2$, эксцесс $\gamma_\xi = 6$, $X_{\text{mod}} = 0$, $X_{\text{med}} = \frac{1}{\lambda} \ln 2$. Производящая функция моментов показательного распределения описывается формулой $m_\xi(t) = \frac{\lambda}{\lambda - t}$ при $t < \lambda$ (в противном случае производящей функции не существует).

Нормальное распределение (распределение Гаусса). Непрерывная случайная величина называется *распределенной по нормальному закону с параметрами a и σ^2* , если ее плотность распределения

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

Множество случайных величин, распределенных по нормальному закону с параметрами a и σ^2 , обозначается через $N(a, \sigma^2)$. В частном случае, когда $a = 0$ и $\sigma^2 = 1$, нормальное распределение называется *стандартным*, и класс таких случайных величин обозначается $N(0, 1)$.

Функция распределения нормально распределенной случайной величины

$$F_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Графики плотности и функции стандартного нормального распределения представлены на рис. 6.7, в общем случае — на рис. 6.8.

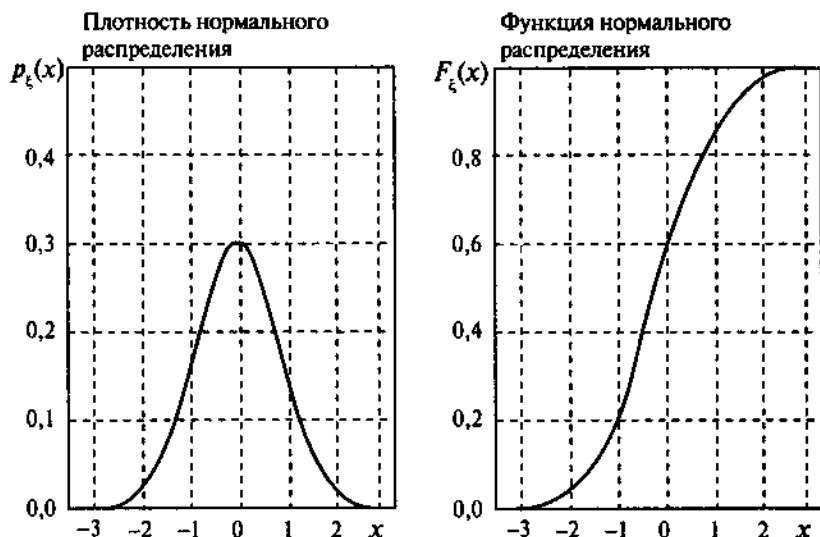


Рис. 6.7

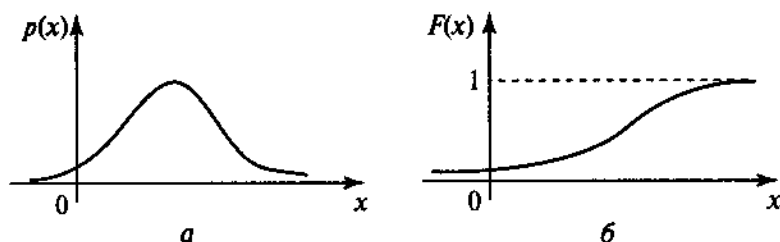


Рис. 6.8

Параметры нормального распределения суть математическое ожидание $M\xi = a$ и дисперсия $D\xi = \sigma^2$.

Плотность нормального распределения симметрична относительно $x = a$, поэтому $X_{\text{med}} = X_{\text{mod}} = M\xi = a$. Асимметрия и эксцесс нормального распределения равны нулю.

Плотность стандартного распределения равна

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

а функция распределения

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

Такой интеграл не вычисляем аналитически, но функция $\Phi(x)$ связана с функцией Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

соотношением

$$\Phi(x) = \frac{1}{2} + \Phi_0(x).$$

В случае же произвольных значений параметров a и σ^2 функция распределения $F_\xi(x)$ случайной величины $\xi \in N(a, \sigma^2)$ связана с функцией Лапласа с помощью соотношения

$$F_\xi(x) = \frac{1}{2} + \Phi_0\left(\frac{x-a}{\sigma}\right).$$

Поэтому вероятность попадания нормально распределенной случайной величины $\xi \in N(a, \sigma^2)$ на полуинтервал $[c_1, c_2)$ можно вычислять по формуле

$$P\{c_1 \leq \xi < c_2\} = \Phi_0\left(\frac{c_2-a}{\sigma}\right) - \Phi_0\left(\frac{c_1-a}{\sigma}\right).$$

Напомним, что значения функции Лапласа затабулированы в табл. 2 приложения 2.

Задача 3. Пусть задана случайная величина $\xi \in N(1, 4)$. Вычислить вероятность $P\{0 < \xi < 3\}$.

Решение. Здесь $a = 1$ и $\sigma = 2$. Согласно указанной выше формуле

$$\begin{aligned} P(0 < \xi < 3) &= \Phi_0\left(\frac{3-1}{2}\right) - \Phi_0\left(\frac{0-1}{2}\right) = \Phi_0(1) - \Phi_0(-0,5) = \\ &= \Phi_0(1) + \Phi_0(0,5) = 0,3413 + 0,1915 = 0,5328. \end{aligned}$$

Важными свойствами нормального распределения являются следующие:

1. Если $\eta = A\xi + B$, где $\xi \in N(a, \sigma^2)$, то $\eta \in N(Aa + B, A^2\sigma^2)$.

В частности, случайная величина $\xi \in N(a, \sigma^2)$ может быть представлена в виде $\xi = a + \sigma\xi_0$, где $\xi_0 \in N(0, 1)$.

2. Сумма двух независимых нормально распределенных случайных величин имеет нормальный закон распределения. При этом их математические ожидания и дисперсии суммируются.

Производящая функция моментов нормального распределения имеет вид $m_\xi(t) = e^{at + \frac{\sigma^2 t^2}{2}}$ (существует при любом значении параметра t).

Для центральных моментов любого порядка нормально распределенной случайной величины можно вывести рекуррентное соотношение

$$\mu_k = (k-1)\sigma^2\mu_{k-2},$$

позволяющее выражать моменты высших порядков через моменты низших порядков. Поскольку $\mu_1 = 0$, все нечетные моменты нормального распределения равны нулю. Для четных моментов получаем следующие выражения:

$$\mu_0 = 1, \mu_2 = \sigma^2, \mu_4 = 3\sigma^4, \dots, \mu_{2k} = (2k-1)!!\sigma^{2k}.$$

Логарифмически нормальное (логнормальное) распределение случайная величина ξ имеет в том случае, когда ее логарифм имеет нормальное распределение. Соответственно ξ может быть представлена как показательная функция от нормально распределенной случайной величины. Для описания логнормального распределения используется различная параметризация и соответственно по-разному выражаются математическое ожидание и дисперсия.

Если $\xi = ae^\eta$, $\eta \in N(0, \sigma^2)$, как это предполагается в [1], то $M\xi = ae^{\sigma^2/2}$ и $D\xi = a^2e^{\sigma^2}(e^{\sigma^2} - 1)$.

Если $\xi = e^\eta$, $\eta \in N(0, \sigma^2)$, что будем обозначать через $\xi \in \exp\{N(a, \sigma^2)\}$, то $M\xi = Ae^{\sigma^2/2}$ и $D\xi = a^2e^{\sigma^2}(e^{\sigma^2} - 1)$, где $A = e^a$.

Другие числовые характеристики:

мода — $X_{\text{mod}} = Ae^{-\sigma^2}$;

медиана — $X_{\text{med}} = A$;

асимметрия — $\beta_\xi = (e^{\sigma^2} - 1)^{-1/2}(e^{\sigma^2} + 2)$;

эксцесс — $\gamma_\xi = (e^{\sigma^2} - 1)(e^{3\sigma^2} + 3e^{2\sigma^2} + 6e^{\sigma^2} + 6)$.

Из этих формул видно, что асимметрия и эксцесс логарифмически нормального распределения всегда положительны

и тем ближе к нулю, чем ближе к нулю σ . Мода и медиана стремятся к слиянию по мере стремления к нулю величины σ .

Графики плотности и функции распределения при некоторых значениях параметров представлены на рис. 6.9 (при $A = 1$, $\sigma = 1$) и 6.10.

Значения логарифмически нормальной случайной величины образуются как «случайные искажения» некоторого «истинного значения» A , которое является не средним значением, а медианой. Значения логарифмически нормальной случайной величины оказываются характерными для многих конкретных физических и социально-экономических ситуаций (размеры и вес

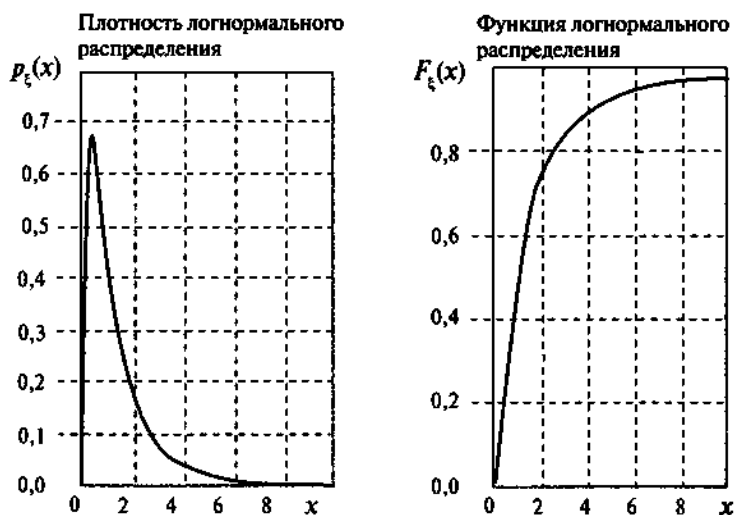


Рис. 6.9

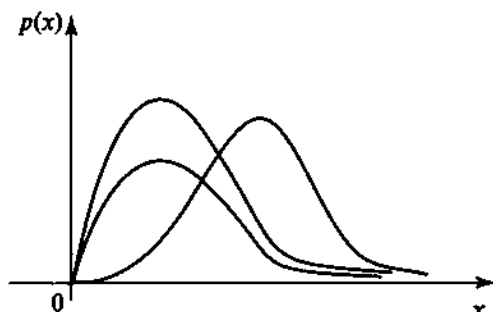


Рис. 6.10

частиц, образующихся при дроблении; заработная плата работника; доход семьи; размеры космических образований; долговечность изделия, работающего в режиме износа и старения, и др.).

Распределение Лапласа задается плотностью

$$p_{\xi}(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \quad -\infty < x < \infty$$

(двусторонняя показательная плотность).

Функция плотности распределения симметрична относительно нуля, т.е. четна, и поэтому математическое ожидание $M\xi = 0$. Дисперсия в два раза больше дисперсии случайной величины, распределенной по показательному закону: $D\xi = \frac{2}{\lambda^2}$. Действительно

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\lambda}{2} \int_{-\infty}^{\infty} x^2 e^{-\lambda|x|} dx = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Симметричная унимодальная функция плотности этого закона с острым максимумом в точке ноль иногда используется для описания распределений остаточных случайных компонент (ошибок) в моделях регрессионного типа.

В силу симметрии имеем $X_{\text{med}} = X_{\text{mod}} = 0$, $\beta_{\xi} = 0$. Эксцесс равен $\gamma_{\xi} = 3$.

Графики плотности и функции стандартного распределения Лапласа представлены на рис. 6.11 (при $\lambda = 1$), общий вид плотности — на рис. 6.12.

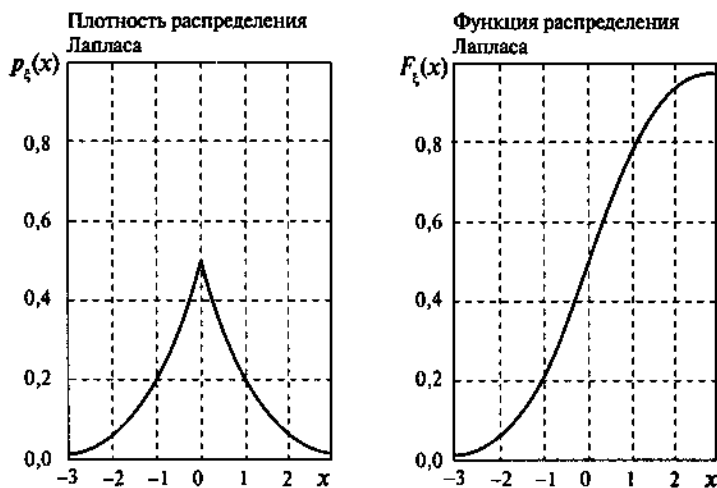


Рис. 6.11

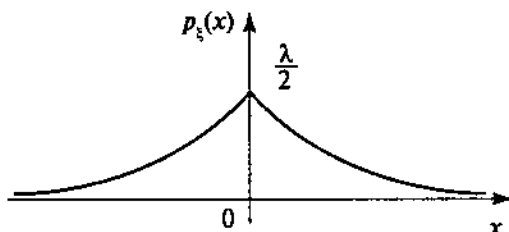


Рис. 6.12

Распределение Вейбулла может быть задано плотностью

$$p_{\xi}(x) = \begin{cases} \lambda \alpha x^{\alpha-1} e^{-\lambda x^{\alpha}}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad \alpha > 0.$$

Функция распределения имеет вид

$$F_{\xi}(x) = \begin{cases} 1 - e^{-\lambda x^{\alpha}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Графики плотности и функции распределения Вейбулла при некоторых значениях параметров представлены на рис. 6.13 ($\alpha = 2$, $\lambda = 1$) и 6.14.

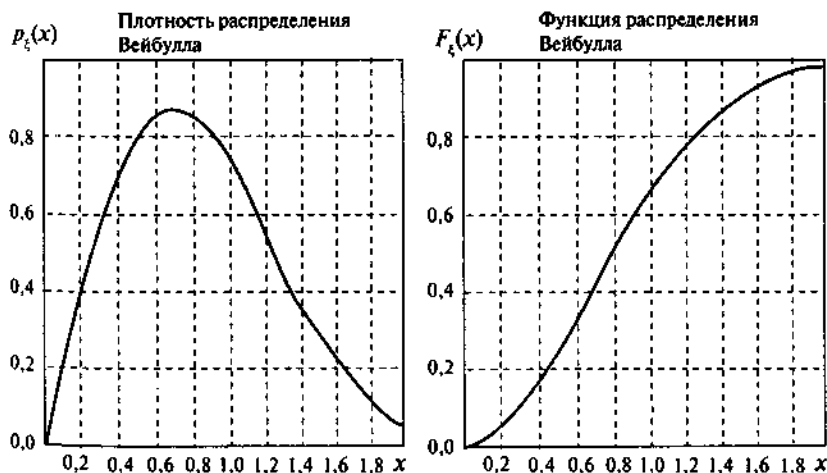


Рис. 6.13

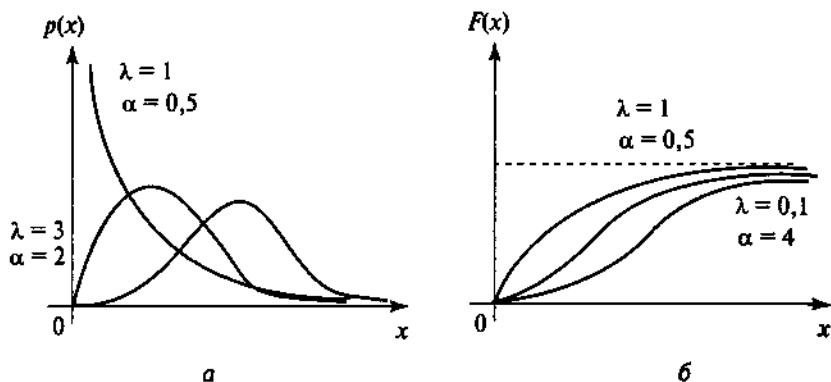


Рис. 6.14

Распределению Вейбулла подчиняется время безотказной работы многих технических устройств. В задачах этого профиля важной характеристикой является *интенсивность отказа (коэффициент смертности)* $\lambda(t)$ исследуемых элементов возраста t , определяемый соотношением $\lambda(t) = \frac{p_t(t)}{1 - F_t(t)}$. Для распределения

Вейбулла этот показатель принимает достаточно простой вид степенной функции: $\lambda(t) = \lambda \alpha t^{\alpha-1}$.

Если $\alpha = 1$, то распределение Вейбулла превращается в показательное распределение, а если $\alpha = 2$ — в так называемое **распределение Рэлея**.

Математическое ожидание распределения Вейбулла $M\xi = \lambda^{-\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)$ и дисперсия $D\xi = \lambda^{-\frac{2}{\alpha}} \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right]$, где $\Gamma(a)$ — гамма-функция Эйлера.

$$\text{Мода имеет вид } X_{\text{mod}} = \begin{cases} 0, & \alpha \leq 1; \\ \lambda^{-\frac{1}{\alpha}} \left(1 - \frac{1}{\alpha}\right)^{\frac{1}{\alpha}}, & \alpha \geq 1; \end{cases}$$

$$\text{медиана } X_{\text{med}} = \left(\frac{\ln 2}{\lambda} \right)^{1/\alpha}.$$

В различных задачах прикладной статистики часто встречаются так называемые «усеченные» распределения. Например, налоговые органы интересуются распределением доходов тех лиц, годовой доход которых превосходит некоторый порог c_0 ,

установленный законами о налогообложении. Эти распределения приближаются распределением Парето.

Распределение Парето задается функциями распределения и плотности (рис. 6.15, $c_0 = 1$, $\alpha = 2$)

$$F_{\xi}(x) = 1 - \left(\frac{c_0}{x}\right)^{\alpha}; \quad p_{\xi}(x) = \frac{\alpha}{c_0} \left(\frac{c_0}{x}\right)^{\alpha+1},$$

где $\alpha > 0$, а $x > c_0$. Функция плотности имеет вид монотонно убывающей кривой, выходящей из точки $(c_0, \alpha/c_0)$.

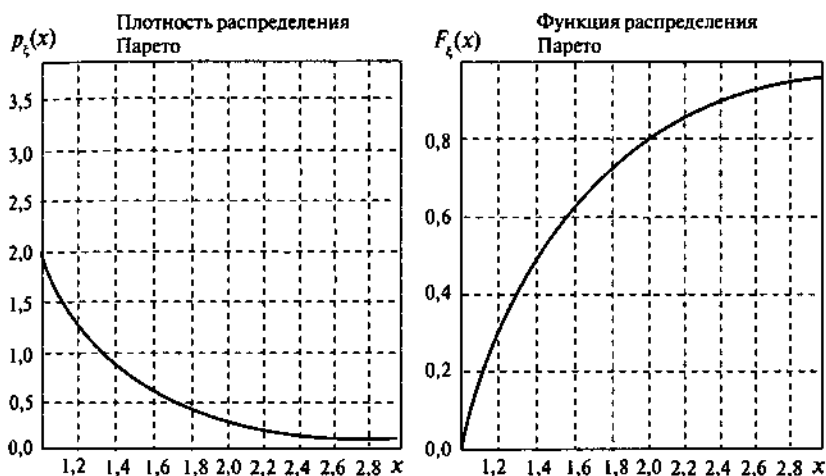


Рис. 6.15

Основные числовые характеристики этого распределения существуют не всегда, а лишь при соблюдении определенных требований к значению параметра α :

математическое ожидание $M\xi = \frac{\alpha c_0}{\alpha - 1}$ существует при $\alpha > 1$;

дисперсия $D\xi = \frac{\alpha c_0^2}{(\alpha - 1)^2 (\alpha - 2)}$ существует при $\alpha > 2$;

мода $X_{\text{mod}} = c_0$ и медиана $X_{\text{med}} = c_0 2^{\frac{1}{\alpha}}$ существуют всегда;

момент k -го порядка $M\xi^k = \frac{\alpha c_0^k}{\alpha - k}$ существует при $\alpha > k$.

Распределение Коши задается функцией плотности распределения

$$p_{\xi}(x) = \frac{c}{\pi(c^2 + (x - \alpha)^2)},$$

где $c > 0$ — параметр масштаба и α — параметр сдвига, определяющий значение моды и медианы. Функция распределения задается формулой

$$F_{\xi}(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x - \alpha}{c}.$$

Стандартное распределение Коши соответствует случаю $\alpha = 0$, $c = 1$.

Для параметров $\alpha = 0$, $c = 1$ графики приведены на рис. 6.16.

Отметим два важных свойства (самовоспроизводимости) распределения Коши.

1. Если случайная величина ξ имеет распределение Коши с параметрами c и α , то любая линейная функция $b_0 + b_1\xi$ имеет распределение того же типа с параметрами $c' = |b_1|c$ и $\alpha' = b_1\alpha + b_0$.

2. Если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и имеют одинаковое распределение Коши, то среднее арифметическое $\bar{\xi} = (\xi_1 + \xi_2 + \dots + \xi_n)/n$ имеет то же распределение Коши, что и ξ_1 .



Рис. 6.16

Задачи для самостоятельного решения

Теоретические задачи

1. Случайная величина ξ имеет нормальное распределение с параметрами a и σ^2 . Показать, что величина $\frac{\xi - a}{\sigma}$ нормально распределена с параметрами 0 и 1.
2. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и имеют одинаковую функцию плотности распределения $p(x) = \begin{cases} ae^{-ax}, & x \geq 0, \\ 0, & x < 0. \end{cases}$

Найти функцию и плотность распределения величин:

- а) $\eta_1 = \min\{\xi_1, \xi_2, \dots, \xi_n\}$; б) $\eta_2 = \max\{\xi_1, \xi_2, \dots, \xi_n\}$.
3. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и равномерно распределены на отрезке $[a, b]$. Найти функции распределения и плотности величин $\eta_1 = \min\{\xi_1, \xi_2, \dots, \xi_n\}$ и $\eta_2 = \max\{\xi_1, \xi_2, \dots, \xi_n\}$. Доказать, что $M(\eta_1 + \eta_2) = a + b$.
 4. Случайная величина распределена по закону Коши $p(x) = \frac{a}{1+x^2}$. Найти: а) коэффициент a ; б) функцию распределения; в) вероятность попадания на интервал $(-1, 1)$. Показать, что математическое ожидание ξ не существует.
 5. Случайная величина подчинена закону Лапласа с параметром λ ($\lambda > 0$): $p(x) = ae^{-\lambda|x|}$. Найти коэффициент a , математическое ожидание $M\xi$ и дисперсию $D\xi$, вероятности событий $\{|\xi| < \sqrt{D\xi}\}$ и $\{|\xi| < 3\sqrt{D\xi}\}$.
 6. Найти $P(|\xi - M\xi| < 3\sqrt{D\xi})$, если ξ имеет: а) нормальное распределение с параметрами a и σ^2 ; б) показательное распределение с параметром λ ; в) равномерное распределение на отрезке $[-1; 1]$.
 7. Случайная величина ξ подчинена закону Симпсона на отрезке $[-a, a]$, т.е. график ее плотности распределения имеет вид, показанный на рис. 6.17. Написать формулу для плотности распределения, найти $M\xi$ и $D\xi$.

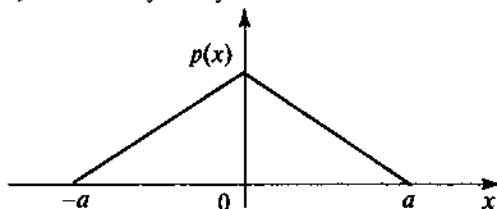


Рис. 6.17

8. Случайная точка A имеет в круге радиуса R равномерное распределение. Найти математическое ожидание и дисперсию расстояния ρ точки до центра круга. Показать, что величина ρ^2 равномерно распределена на отрезке $[0, R^2]$.
9. Найти производящую функцию моментов для: а) равномерного распределения на отрезке $[a, b]$; б) распределения Лапласа с параметром λ .
10. Доказать, что для распределений: а) логнормального; б) Вейбулла с $\alpha < 1$; в) Парето — производящая функция моментов определена только при $t \leq 0$; г) для распределения Коши — только при $t = 0$.

Вычислительные задачи

11. Плотность распределения случайной величины ξ равна $p_\xi(x)$. Вычислить константу C , функцию распределения $F(x)$, $M\xi$ и вероятность $P(A)$, если:

$$\text{а) } p_\xi(x) = \begin{cases} C(x+1), & x \in [-1, 2] \\ 0, & x \notin [-1, 2] \end{cases}; \quad A = \{\xi^2 < 1\};$$

$$\text{б) } p_\xi(x) = \begin{cases} C(5-x), & x \in [-2, 1] \\ 0, & x \notin [-2, 1] \end{cases}; \quad A = \{0 < \xi < 3\};$$

$$\text{в) } p_\xi(x) = \begin{cases} C(x-2), & x \in [3, 5] \\ 0, & x \notin [3, 5] \end{cases}; \quad A = \{4 < \xi < 6\};$$

$$\text{г) } p_\xi(x) = \begin{cases} C(x+4), & x \in [-1, 3] \\ 0, & x \notin [-1, 3] \end{cases}; \quad A = \{2 < \xi < 5\}.$$

$$\text{д) } p_\xi(x) = \begin{cases} C(3-x), & x \in [-5, 2] \\ 0, & x \notin [-5, 2] \end{cases}; \quad A = \{0 < \xi < 5\}.$$

12. Плотность распределения случайной величины ξ равна $p_\xi(x)$. Вычислить константу C , функцию распределения $F(x)$, $M\xi$ и вероятность $P(A)$, если:

$$\text{а) } p_\xi(x) = \begin{cases} C(x^2-1), & x \in [1, 3] \\ 0, & x \notin [1, 3] \end{cases}; \quad A = \{2 < \xi < 5\};$$

$$\text{б) } p_\xi(x) = \begin{cases} C(x^2+3), & x \in [-2, 1] \\ 0, & x \notin [-2, 1] \end{cases}; \quad A = \{-1 < \xi < 2\};$$

$$\text{в) } p_{\xi}(x) = \begin{cases} C(9-x^2), & x \in [-3, 2]; \\ 0, & x \notin [-3, 2]; \end{cases} \quad A = \{0 < \xi < 3\};$$

$$\text{г) } p_{\xi}(x) = \begin{cases} C(x^2+x), & x \in [1, 4]; \\ 0, & x \notin [1, 4]; \end{cases} \quad A = \{3 < \xi < 6\};$$

$$\text{д) } p_{\xi}(x) = \begin{cases} C(x^2-2x), & x \in [2, 5]; \\ 0, & x \notin [2, 5]; \end{cases} \quad A = \{1 < \xi < 4\};$$

$$\text{е) } p_{\xi}(x) = \begin{cases} C(x^3+1), & x \in [2, 6]; \\ 0, & x \notin [2, 6]; \end{cases} \quad A = \{5 < \xi < 7\};$$

$$\text{ж) } p_{\xi}(x) = \begin{cases} C(x^3+2), & x \in [1, 3]; \\ 0, & x \notin [1, 3]; \end{cases} \quad A = \{0 < \xi < 2\}.$$

13. Плотность распределения случайной величины ξ равна $p_{\xi}(x)$.

Вычислить константу C , функцию распределения $F(x)$, $M\xi$ и вероятность $P(A)$, если:

$$\text{а) } p_{\xi}(x) = \begin{cases} C/(x-4), & x \in [5, 8]; \\ 0, & x \notin [5, 8]; \end{cases} \quad A = \{3 < \xi < 6\};$$

$$\text{б) } p_{\xi}(x) = \begin{cases} C/(x+2), & x \in [-1, 3]; \\ 0, & x \notin [-1, 3]; \end{cases} \quad A = \{0 < \xi < 5\};$$

$$\text{в) } p_{\xi}(x) = \begin{cases} C/(8-x), & x \in [-2, 3]; \\ 0, & x \notin [-2, 3]; \end{cases} \quad A = \{2 < \xi < 5\};$$

$$\text{г) } p_{\xi}(x) = \begin{cases} C/(x+6), & x \in [-1, 4]; \\ 0, & x \notin [-1, 4]; \end{cases} \quad A = \{0 < \xi < 5\}.$$

14. Плотность распределения случайной величины ξ имеет вид:

$$p_{\xi}(x) = \begin{cases} C(x+1)^{-3/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Вычислить константу C , функцию распределения $F(x)$, $M\xi$, $D\xi$ и вероятность $P\{|\xi - 1/3| < 1\}$.

15. Плотность распределения случайной величины ξ имеет вид:

$$p_{\xi}(x) = \begin{cases} C(1-x^2), & |x| \leq 1, \\ 0, & |x| > 1. \end{cases}$$

Вычислить константу C , функцию распределения $F(x)$, $M\xi$, $D\xi$ и вероятность $P\{|\xi - 1/2| < 1/4\}$.

16. Плотность распределения случайной величины ξ имеет вид:

$$p_{\xi}(x) = \begin{cases} Ce^x, & x \leq 0, \\ 0, & x > 0. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, $M\xi$, $D\xi$ и вероятность $P\{-2 < \xi < 1\}$.

17. Случайная величина ξ имеет функцию распределения

$$F(x) = \begin{cases} \frac{1}{2}e^x, & x \leq 0, \\ 1 - \frac{1}{2}e^{-x}, & x > 0. \end{cases}$$

Найти плотность случайной величины, $M\xi$, $D\xi$ и вероятность $P\{-1 < \xi < 3\}$.

18. Проверить, что функция $F(x) = \begin{cases} 0, & x \leq 0, \\ 2x - x^2, & 0 < x < 1, \\ 1, & x \geq 1 \end{cases}$ может быть

функцией распределения случайной величины. Найти $M\xi$ и $D\xi$.

19. Стержень длины 24 см ломают на две части; будем считать, что точка излома распределена равномерно по всей длине стержня. Чему равна средняя длина большей части стержня?
20. Отрезок длины 12 см случайным образом разрезается на две части. Точка разреза равномерно распределена по всей длине отрезка. Чему равна средняя длина малой части отрезка?
21. Пассажир приходит в случайный момент на остановку, где может сесть на автобус или троллейбус (смотря что придет раньше). Автобус ходит с интервалами в 15 минут, троллейбус — 10 минут (независимо друг от друга). Найти функцию распределения времени ожидания и его среднее значение.
22. Пассажир приходит в случайный момент на остановку, где может сесть на автобус, троллейбус или трамвай (смотря что придет раньше). Автобус ходит с интервалами 20 минут, троллейбус — 15 минут, трамвай — 10 минут. Найти функцию распределения времени ожидания и его среднее значение.
23. У жителей поселка Соткино доходы имеют распределение Парето, минимальный доход составляет 100 тыс. руб., а средний — 200 тыс. руб. Определить долю жителей с доходами: а) свыше 300 тыс. руб.; б) от 200 до 500 тыс. руб.

24. Известно, что срок службы изделия распределен показательно со средним 1 год. Найти вероятность того, что изделие прослужит не менее: а) 6 месяцев; б) 2 лет.
25. Относительное изменение курса акции за день (в процентах) имеет распределение Лапласа с $D\xi = 8$. Найти вероятность того, что курс акции: а) упадет более чем на 5%; б) поднимется более чем на 3%.
26. Два сотрудника разделили работу поровну. Время выполнения доли работы каждого равномерно распределено от 1 до 2 часов. Найти среднее время выполнения работы.
27. Работа состоит из двух частей, которые выполняются независимо и параллельно. Время выполнения первой части равномерно распределено от 1 до 2 часов, второй — от 1 до 3 часов. Найти среднее время выполнения работы.
28. Работа состоит из двух частей, которые выполняются независимо и параллельно. Время выполнения первой части показательно распределено со средним 1 час, второй — 2 часа. Найти среднее время выполнения работы.
29. Прибор состоит из двух блоков, которые могут выйти из строя независимо друг от друга. Прибор выходит из строя, если выходит из строя хотя бы один блок. Срок службы первого блока показательно распределен со средним 2 года, второго — 4 года. Найти средний срок службы прибора.
30. Прибор состоит из двух блоков, которые могут выйти из строя независимо друг от друга. Прибор выходит из строя, если выходит из строя хотя бы один блок. Срок службы первого блока имеет распределение Рэлея со средним 3 года, второго — 4 года. Найти средний срок службы прибора.

ГЛАВА 7

ФУНКЦИИ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН. НЕПРЕРЫВНЫЙ СЛУЧАЙНЫЙ ВЕКТОР

7.1. Функции от случайных величин

Пусть задана функция плотности $p_{\xi}(x)$ случайной величины ξ и монотонная дифференцируемая функция $y = \psi(x)$. Тогда плотность распределения случайной величины $\eta = \psi(\xi)$ равна

$$p_{\eta}(y) = p_{\xi}(\psi^{-1}(y)) \left| \frac{d\psi^{-1}(y)}{dy} \right|.$$

Здесь $\psi^{-1}(y)$ — функция, обратная к функции $y = \psi(x)$.

Доказательство. Пусть $\psi(x)$ — возрастающая функция, тогда обратная ей функция также возрастающая, и функцию распределения случайной величины $\eta = \psi(\xi)$ можно представить в виде $F_{\eta}(x) = F_{\xi}(\psi^{-1}(x))$, так как

$$F_{\eta}(x) = P(\psi(\xi) < x) = P(\xi < \psi^{-1}(x)) = F_{\xi}(\psi^{-1}(x)).$$

Дифференцируя, получаем

$$p_{\eta}(x) = F'_{\eta}(x) = \frac{dF_{\xi}(\psi^{-1}(x))}{d\psi^{-1}(x)} \cdot \frac{d\psi^{-1}(x)}{dx} = p_{\xi}(\psi^{-1}(x)) \frac{d\psi^{-1}(x)}{dx},$$
$$\frac{d\psi^{-1}(x)}{dx} > 0.$$

Пусть теперь $\psi(x)$ — убывающая функция, тогда обратная ей также убывающая, и

$$F_{\eta}(x) = P(\psi(\xi) < x) = P(\xi > \psi^{-1}(x)) = 1 - F_{\xi}(\psi^{-1}(x)).$$

Отсюда получаем

$$\begin{aligned} p_{\eta}(x) &= F'_{\eta}(x) = (1 - F_{\xi}(\psi^{-1}(x)))' = -\frac{dF_{\xi}(\psi^{-1}(x))}{d\psi^{-1}(x)} \cdot \frac{d\psi^{-1}(x)}{dx} = \\ &= -p_{\xi}(\psi^{-1}(x)) \frac{d\psi^{-1}(x)}{dx}, \end{aligned}$$

причем здесь $\frac{d\psi^{-1}(x)}{dx} < 0$.

Объединяя формулы, полученные в случаях возрастающей и убывающей функции ψ , получаем исходное утверждение.

Если плотность $p_{\xi}(x)$ отлична от нуля на некотором промежутке (конечном или бесконечном), то границы соответствующего промежутка для $p_{\eta}(y)$ находятся подстановкой исходных границ в функцию ψ . Заметим, что порядок следования границ может меняться, когда ψ — убывающая функция.

Задача 1. Случайная величина ξ равномерно распределена на отрезке $[0, 2]$. Найти плотность случайной величины $\eta = -\sqrt{\xi + 1}$.

Решение. Из условия задачи следует, что

$$p_{\xi}(x) = \begin{cases} 0, & x \notin [0, 2], \\ \frac{1}{2}, & x \in [0, 2]. \end{cases}$$

Далее, функция $y = -\sqrt{x+1}$ является монотонной и дифференцируемой функцией на отрезке $[0, 2]$ и имеет обратную функцию $x = \psi^{-1}(y) = y^2 - 1$, производная которой равна $\frac{d\psi^{-1}(y)}{dy} = 2y$. Кроме того, $\psi(0) = -1$, $\psi(2) = -\sqrt{3}$. Следовательно,

$$\begin{aligned} p_{\eta}(y) &= p_{\xi}(\psi^{-1}(y)) \left| \frac{d\psi^{-1}(y)}{dy} \right| = p_{\xi}(\psi^{-1}(y)) \cdot 2|y| = \\ &= 2|y| \cdot \begin{cases} 0, & y \notin [-\sqrt{3}, -1], \\ \frac{1}{2}, & y \in [-\sqrt{3}, -1]. \end{cases} \end{aligned}$$

Значит, $p_{\eta}(y) = \begin{cases} 0, & y \notin [-\sqrt{3}, -1], \\ -y, & y \in [-\sqrt{3}, -1]. \end{cases}$

7.2. Совместный закон распределения непрерывных случайных величин

Пусть на вероятностном пространстве Ω заданы две непрерывные случайные величины ξ и η . Тогда упорядоченная пара (ξ, η) определяет «случайную» точку на плоскости и называется *двумерным случайным вектором*, или *двумерной случайной величиной*, так же как в дискретном случае.

Совместной функцией распределения непрерывных случайных величин ξ и η называется функция $F(x, y) = \{ \xi < x, \eta < y \}$, определяющая вероятность попадания случайного вектора (ξ, η) в бесконечный угол на плоскости с вершиной в точке (x, y) , лежащий ниже и левее этой точки. Это общее определение верно как в дискретном, так и в непрерывном случаях, и свойства совместной функции распределения для дискретных величин остаются справедливыми и для непрерывных случайных величин. Напомним, если случайные величины ξ и η независимы, то совместная функция распределения равна произведению функций распределения составляющих: $F(x, y) = F_{\xi}(x)F_{\eta}(y)$, и наоборот: если выполнено равенство $F(x, y) = F_{\xi}(x)F_{\eta}(y)$, то случайные величины ξ и η независимы.

Случайный вектор называется *непрерывным*, если его совместная функция распределения непрерывна. Случайный вектор называется *абсолютно непрерывным*, если существует такая неотрицательная функция $p_{\xi, \eta}(x, y)$, называемая *совместной плотностью распределения* случайных величин ξ и η (или вектора), что имеет место равенство

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p_{\xi, \eta}(u, v) du dv.$$

Смысл определения совместной плотности распределения заключается в следующем. Вероятность того, что «случайная точка» (ξ, η) попадет в область $B \subseteq R^2$ на плоскости, вычисляется как объем трехмерной фигуры — «криволинейного» цилиндра, ограниченного поверхностью $z = p_{\xi, \eta}(x, y)$ и плоскостью $z = 0$, основанием которого является множество B . Аналитически этот факт записывается с помощью двойного интеграла:

$$P\{(\xi, \eta) \in B\} = \iint_B p_{\xi, \eta}(x, y) dx dy.$$

Совместная плотность распределения обладает также следующими свойствами:

$$1) p(x, y) \geq 0;$$

$$2) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = 1;$$

$$3) p(x, y) = F''_{xy}(x, y).$$

Простейшим примером совместного распределения двух случайных величин является двумерное *равномерное распределение на множестве A* . Пусть задано ограниченное множество A с площадью $S(A)$. Тогда указанное распределение определяется как распределение пары (ξ, η) , задаваемое с помощью следующей совместной плотности:

$$p_{\xi, \eta}(x, y) = \begin{cases} 0, & (x, y) \notin A; \\ \frac{1}{S(A)}, & (x, y) \in A. \end{cases}$$

Задача 2. Пусть двумерный случайный вектор (ξ, η) равномерно распределен внутри треугольника $\Delta = \{(x, y): x > 0, y > 0, x + y < 2\}$. Вычислить вероятность неравенства $\xi > \eta$.

Решение. Площадь указанного треугольника Δ равна $S(\Delta) = 2$ (рис. 7.1). В силу определения двумерного равномерного распределения совместная плотность случайных величин ξ, η равна:

$$p_{\xi, \eta}(x, y) = \begin{cases} 0, & (x, y) \notin \Delta; \\ \frac{1}{2}, & (x, y) \in \Delta. \end{cases}$$

Событие $\{\xi > \eta\}$ соответствует множеству $B = \{(x, y): x > y\}$ на плоскости, т.е. полуплоскости. Тогда вероятность

$$P(B) = P\{(\xi, \eta) \in B\} = \iint_B p_{\xi, \eta}(x, y) dx dy.$$

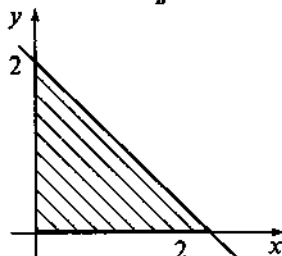


Рис. 7.1

На полуплоскости B совместная плотность $p_{\xi,\eta}(x, y)$ равна нулю вне множества Δ и $1/2$ — внутри множества Δ . Таким образом, полуплоскость B разбивается на два множества: $B_1 = B \cap \Delta$ и $B_2 = B \cap \bar{\Delta}$. Следовательно, двойной интеграл по множеству B представляется в виде суммы интегралов по множествам B_1 и B_2 , причем второй интеграл равен нулю, так как там совместная плотность равна нулю. Поэтому

$$\begin{aligned} P\{(\xi, \eta) \in B\} &= \iint_B p_{\xi,\eta}(x, y) dx dy = \iint_{B_1} \frac{1}{2} dx dy + \iint_{B_2} 0 dx dy = \\ &= \frac{1}{2} S(B_1) = \frac{1}{2} \cdot 1 = \frac{1}{2}. \end{aligned}$$

Если задана совместная плотность распределения $p_{\xi,\eta}(x, y)$ пары (ξ, η) , то плотности $p_\xi(x)$ и $p_\eta(y)$ составляющих ξ и η называются *частными плотностями* и вычисляются по формулам:

$$\begin{aligned} p_\xi(x) &= \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dy, \\ p_\eta(y) &= \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dx. \end{aligned}$$

Для непрерывно распределенных случайных величин с плотностями $p_\xi(x)$, $p_\eta(y)$ независимость означает, что при любых x и y выполнено равенство

$$p_{\xi,\eta}(x, y) = p_\xi(x) p_\eta(y).$$

Задача 3. В условиях предыдущей задачи определить, независимы ли составляющие случайного вектора ξ и η .

Решение. Вычислим частные плотности $p_\xi(x)$ и $p_\eta(y)$. Имеем:

$$p_\xi(x) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dy = \begin{cases} 0, & x \notin (0, 2), \\ \int_0^{2-x} \frac{1}{2} dy, & x \in (0, 2) \end{cases} = \begin{cases} 0, & x \notin (0, 2); \\ \frac{2-x}{2}, & x \in (0, 2). \end{cases}$$

Аналогично

$$p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dx = \begin{cases} 0, & y \notin (0, 2); \\ \frac{2-y}{2}, & y \in (0, 2). \end{cases}$$

Очевидно, что в нашем случае $p_{\xi,\eta}(x,y) \neq p_{\xi}(x)p_{\eta}(y)$, и потому случайные величины ξ и η зависимы.

Числовые характеристики для случайного вектора (ξ, η) можно вычислять с помощью следующей общей формулы. Пусть $p_{\xi,\eta}(x, y)$ — совместная плотность величин ξ и η , а $\psi(x, y)$ — функция двух аргументов, тогда

$$M\psi(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \psi(x, y) p_{\xi,\eta}(x, y) dx dy.$$

В частности,

$$M\xi = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x p_{\xi,\eta}(x, y) dx dy;$$

$$M\eta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y p_{\xi,\eta}(x, y) dx dy;$$

$$M\xi\eta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy p_{\xi,\eta}(x, y) dx dy.$$

Задача 4. В условиях предыдущей задачи вычислить $M\xi\eta$.

Решение. Согласно указанной выше формуле имеем

$$M\xi\eta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy p_{\xi,\eta}(x, y) dx dy = \iint_{\Delta} xy \cdot \frac{1}{2} dx dy.$$

Представив треугольник Δ в виде

$$\Delta = \{(x, y): 0 < x < 2; 0 < y < 2 - x\},$$

двойной интеграл можно вычислить как повторный:

$$\begin{aligned} M\xi\eta &= \iint_{\Delta} xy \cdot \frac{1}{2} dx dy = \frac{1}{2} \int_0^2 x dx \int_0^{2-x} y dy = \frac{1}{2} \int_0^2 x dx \left(\frac{y^2}{2} \Big|_0^{2-x} \right) = \\ &= \frac{1}{4} \int_0^2 x(2-x)^2 dx = \frac{1}{3}. \end{aligned}$$

Важную роль в приложениях играет **двумерное нормальное распределение**. Пусть η_1 и η_2 — независимые случайные величины, имеющие стандартное нормальное распределение.

Тогда двумерная случайная величина $\eta = (\eta_1, \eta_2)$ имеет *стандартное двумерное нормальное распределение*. Его совместная плотность равна:

$$p_{\eta_1, \eta_2}(x_1, x_2) = \frac{1}{2\pi} \exp \left(-\frac{x_1^2 + x_2^2}{2} \right).$$

По определению, случайная величина $\xi = (\xi_1, \xi_2)$ имеет *двумерное нормальное распределение*, если ее можно представить в виде

$$\xi = a + B\eta,$$

где $a = (a_1, a_2)$ — вектор математических ожиданий,
 B — некоторая матрица.

В случае, когда коэффициент корреляции $\rho = \rho(\xi_1, \xi_2)$ отличен по модулю от единицы, распределение имеет *двумерную нормальную плотность*:

$$p_{\xi_1, \xi_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x_1 - a_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} \right) \right\}.$$

При этом составляющие имеют нормальные распределения:

$$\xi_1 \in N(a_1, \sigma_1^2), \quad \xi_2 \in N(a_2, \sigma_2^2).$$

Двумерное нормальное распределение обладает рядом полезных свойств, например:

- 1) если $\rho = 0$, то ξ_1 и ξ_2 независимы;
- 2) любая линейная комбинация $c_1\xi_1 + c_2\xi_2$ имеет нормальное распределение;
- 3) условные распределения каждой составляющей по другой также нормальны;
- 4) функции регрессии составляющих линейны.

Поэтому на практике, если исследуется пара случайных величин, каждая из которых имеет нормальное распределение, то можно предположить, что их совместное распределение является двумерным нормальным распределением.

7.3. Плотность суммы двух непрерывных случайных величин

Пусть ξ и η — независимые случайные величины с плотностями $p_\xi(x)$ и $p_\eta(y)$. Плотность случайной величины $\xi + \eta$ вычисляется по формуле свертки

$$p_{\xi+\eta}(x) = \int_{-\infty}^{\infty} p_\xi(x-y)p_\eta(y)dy.$$

Доказательство. Используем представление совместной функции распределения через двойной интеграл:

$$F_{\xi+\eta}(z) = P\{\xi + \eta < z\} = \iint_{D_z} p_{\xi,\eta}(x,y)dx dy,$$

где множество $D_z = \{(x, y): x + y < z\}$ (рис. 7.2).

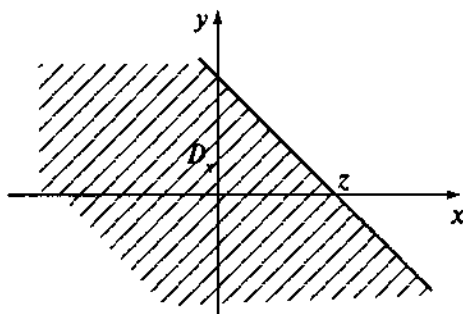


Рис. 7.2

Переходя от двойного интеграла к повторному, получаем

$$F_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} p_{\xi,\eta}(x,y)dx \right) dy.$$

Произведем замену во внутреннем интеграле, положив $x = t - y$, и поменяем порядок интегрирования:

$$F_{\xi+\eta}(z) = \int_{-\infty}^z \left(\int_{-\infty}^{\infty} p(t-y,y)dy \right) dt$$

Дифференцируя под знаком интеграла, получим

$$p_{\xi+\eta}(z) = F'_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} p_{\xi,\eta}(z-y,y)dy.$$

Поскольку величины ξ и η независимы, $p_{\xi, \eta}(x, y) = p_{\xi}(x)p_{\eta}(y)$.

Подставляя последнее равенство в полученное выражение для плотности суммы, приходим к формуле свертки (с точностью до обозначения аргумента):

$$p_{\xi+\eta}(z) = \int_{-\infty}^{\infty} p_{\xi}(z-y)p_{\eta}(y)dy.$$

Разумеется, в силу симметрии справедлива также формула

$$p_{\xi+\eta}(y) = \int_{-\infty}^{\infty} p_{\xi}(x)p_{\eta}(y-x)dx.$$

Задача 5. Пусть ξ и η — независимые случайные величины, распределенные по показательному закону с параметром $\lambda = 2$. Вычислить плотность суммы $\xi + \eta$.

Решение. Поскольку ξ и η распределены по показательному закону с параметром $\lambda = 2$, их плотности равны:

$$p_{\xi}(x) = p_{\eta}(x) = \begin{cases} 2e^{-2x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Следовательно,

$$p_{\xi}(x-y) = \begin{cases} 2e^{-2(x-y)}, & x \geq y; \\ 0, & x < y. \end{cases}$$

Поэтому

$$p_{\xi+\eta}(x) = \int_{-\infty}^{\infty} p_{\xi}(x-y)p_{\eta}(y)dy = \int_0^{\infty} p_{\xi}(x-y) \cdot 2e^{-2y}dy.$$

Если $x < 0$, то в этой формуле аргумент функции $p_{\xi}(x-y)$ отрицателен, и поэтому $p_{\xi}(x-y) = 0$. Следовательно, $p_{\xi+\eta}(x) = 0$. Если же $x \geq 0$, то имеем

$$\begin{aligned} p_{\xi+\eta}(x) &= \int_0^{\infty} p_{\xi}(x-y) \cdot 2e^{-2y}dy = \int_0^x 2e^{-2(x-y)} \cdot 2e^{-2y}dy = \\ &= 4 \int_0^x e^{-2(x-y)} \cdot e^{-2y}dy = 4e^{-2x} \int_0^x dy = 4xe^{-2x}. \end{aligned}$$

Таким образом, получен ответ:

$$p_{\xi+\eta}(x) = \begin{cases} 0, & x < 0; \\ 4xe^{-2x}, & x \geq 0. \end{cases}$$

7.4. Условные распределения и условные математические ожидания непрерывной случайной величины

Пусть на одном и том же пространстве элементарных исходов Ω заданы две абсолютно непрерывные случайные величины ξ и η .

Условным законом распределения случайной величины ξ при условии $\eta = y$, так же как и для дискретных случайных величин, называется любое соотношение, ставящее в соответствие значениям случайной величины ξ условные вероятности их принятия при условии $\eta = y$.

В общем случае условную функцию распределения случайной величины ξ при $\eta = y$ также естественно было бы определить формулой

$$F_{\xi|\eta}(x|y) = \frac{P(\xi < x, \eta = y)}{P(\eta = y)}.$$

Однако это невозможно, поскольку для непрерывной случайной величины $P(\eta = y) = 0$. Поэтому вместо события $\{\eta = y\}$ рассматривают событие $\{y \leq \eta < y + \Delta y\}$ и переходят к пределу по $\Delta y \rightarrow 0$. Таким образом, получаем формулу

$$F_{\xi|\eta}(x|y) = \lim_{\Delta y \rightarrow 0} \frac{P(\xi < x, y \leq \eta < y + \Delta y)}{P(y \leq \eta < y + \Delta y)} = \frac{F'_y(x, y)}{p_\eta(y)}.$$

Наиболее важен для приложений случай, когда вектор (ξ, η) представляет собой двумерную непрерывную случайную величину с совместной плотностью распределения $p_{\xi, \eta}(x, y)$. Тогда

$$F_{\xi|\eta}(x|y) = \int_{-\infty}^x \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)} du.$$

Тогда условная функция распределения имеет производную по x , т.е. существует условная плотность распределения ξ при условии $\eta = y$, равная

$$p_{\xi|\eta}(x|y) = \frac{\partial F_{\xi|\eta}(x|y)}{\partial x} = \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)}.$$

Условное математическое ожидание непрерывной случайной величины ξ при известном η вычисляется по формуле

$$M(\xi|\eta = y) = \int_{-\infty}^{+\infty} x p_{\xi|\eta}(x|y) dx = \int_{-\infty}^{+\infty} x \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)} dx$$

и называется *функцией регрессии* ξ по η :

$$\varphi_{\xi|\eta}(y) = M(\xi|\eta = y),$$

Функция регрессии $\varphi_{\xi|\eta}(y)$ определена в области возможных значений η , т.е. при $p_{\eta}(y) > 0$.

Задача 6. Двумерный случайный вектор (ξ, η) равномерно распределен внутри треугольника $\Delta = \{(x, y): x > 0, y > 0, x + y < 2\}$. Найти условное распределение ξ при $\eta = y$ и функцию регрессии $\varphi_{\xi|\eta}(y)$.

Решение. Как уже было показано (см. задачи 2 и 3),

$$p_{\xi,\eta}(x, y) = \begin{cases} 0, & (x, y) \notin \Delta, \\ \frac{1}{2}, & (x, y) \in \Delta \end{cases} \quad \text{и} \quad p_{\eta}(x) = \begin{cases} 0, & y \notin (0, 2); \\ \frac{2-y}{2}, & y \in (0, 2). \end{cases}$$

Поделив первую плотность на вторую, получаем условную плотность:

$$p_{\xi|\eta}(x|y) = \begin{cases} 0, & x \notin (0, 2-y); \\ \frac{1}{2-y}, & x \in (0, 2-y). \end{cases}$$

Таким образом, речь идет о равномерном распределении на промежутке $(0, 2-y)$. Функцию регрессии вычисляем как математическое ожидание равномерного распределения. Получаем $\varphi_{\xi|\eta}(y) = (2-y)/2$, $0 < y < 2$.

Задача 7. Точку бросают случайным образом в круг радиуса R с центром в начале координат. Найти условную плотность распределения случайной величины ξ — абсциссы точки падения — при условии, что ордината η приняла значение y .

Решение. Естественно, поскольку точка не может попасть за пределы круга, $p_{\xi,\eta}(x, y) = 0$ при $x^2 + y^2 > R^2$. Для каждой области внутри круга вероятность попадания пропорциональна площади этой области. Поэтому $p_{\xi,\eta}(x, y) = A$ при $x^2 + y^2 \leq R^2$, т.е. плотность внутри круга постоянна. Определим константу A :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y) dx dy = \iint_{x^2+y^2 \leq R^2} A dx dy = \pi A R^2 = 1.$$

Отсюда $A = 1/(\pi R^2)$ и плотность совместного распределения

$$p_{\xi, \eta}(x, y) = \begin{cases} 0, & x^2 + y^2 > R^2, \\ \frac{1}{\pi R^2}, & x^2 + y^2 \leq R^2 \end{cases} = \begin{cases} 0, & |x| > \sqrt{R^2 - y^2}, \\ \frac{1}{\pi R^2}, & |x| \leq \sqrt{R^2 - y^2}; \end{cases}$$

$$p_{\eta}(y) = \begin{cases} 0, & |y| > R, \\ \frac{2\sqrt{R^2 - y^2}}{\pi R^2}, & |y| \leq R, \end{cases}$$

и при $|y| \leq R$ получаем условную плотность

$$p_{\xi|\eta}(x|y) = \begin{cases} 0, & |x| > \sqrt{R^2 - y^2}, \\ \frac{1}{2\sqrt{R^2 - y^2}}, & |x| \leq \sqrt{R^2 - y^2}. \end{cases}$$

Таким образом, случайная величина ξ при условии $\eta = y$ равномерно распределена на отрезке $[-\sqrt{R^2 - y^2}; +\sqrt{R^2 - y^2}]$. Ее условное математическое ожидание тождественно равно нулю. Интересно отметить, что условная плотность распределения случайной величины ξ при условии $\eta = y$ равномерна, в то время как безусловная плотность ξ таковой не является. И в этом примере случайные величины ξ и η зависимы между собой.

В заключение отметим, что в таком важном для приложений случае, когда случайная пара $\xi = (\xi_1, \xi_2)$ имеет двумерное нормальное распределение

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-a_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right)\right\},$$

функции регрессии оказываются линейными и имеют вид:

$$\varphi_{\xi_1|\xi_2}(x_1) = a_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - a_1); \quad \varphi_{\xi_2|\xi_1}(x_2) = a_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - a_2),$$

что также можно вывести из представленных выше формул.

Задачи для самостоятельного решения

Теоретические задачи

1. Найти плотности распределения: а) суммы; б) разности; в) произведения; г) частного двух независимых случайных величин, имеющих равномерное распределение на $[0; a]$.
2. Случайные величины ξ_1 и ξ_2 независимы и имеют нормальные распределения с параметрами a_1, σ_1^2 и a_2, σ_2^2 соответственно. Доказать, что $\xi_1 + \xi_2$ имеет нормальное распределение, и найти его параметры.
3. Показать, что если ξ имеет непрерывную функцию распределения $F(x) = P(\xi < x)$, то случайная величина $\eta = F(\xi)$ имеет равномерное распределение на отрезке $[0, 1]$.

Вычислительные задачи

4. Плотность распределения ξ равна:
$$p(x) = \begin{cases} Cx^{-\frac{3}{2}}, & x \geq 1; \\ 0, & x < 1. \end{cases}$$

Найти постоянную C , плотность распределения $\eta = 1/\xi$ и вероятность $P(0,25 < \eta < 0,64)$.

5. Случайная величина ξ имеет функцию распределения

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$$

Найти функцию плотности случайной величины $\eta = \frac{1}{1-\xi}$.

6. Случайная величина ξ равномерно распределена на отрезке I . Найти плотность распределения случайной величины η , если:
 - а) $I = [1, 3], \eta = \xi^2 + 1$;
 - б) $I = [0, 3], \eta = 10 - \xi^2$;
 - в) $I = [1, 3], \eta = (\xi + 2)^2$;
 - г) $I = [0, 2], \eta = (3 - \xi)^3$;
 - д) $I = [-1, 2], \eta = \sqrt{\xi + 3}$;
 - е) $I = [1, 2], \eta = \sqrt{\xi - 1}$;
 - ж) $I = [-1, 2], \eta = 1/(\xi + 1)$;
 - з) $I = [1, 2], \eta = 6/(\xi + 1)$;
 - и) $I = [-1, 1], \eta = 1/(\xi + 1)^2$;
 - к) $I = [-1, 1], \eta = -\ln(\xi + 2)$.

7. Случайная величина ξ имеет показательное распределение с параметром $\lambda=2$. Найти плотность распределения случайной величины η , если:

а) $\eta = \xi^2 - 1$; б) $\eta = \xi^3 + 1$; в) $\eta = 1 + 3/\xi$; г) $\eta = e^{\xi} - 1$; д) $\eta = \ln(\xi + 1)$.

8. Случайная величина ξ имеет показательное распределение с параметром λ . Найти функции плотности распределения случайных величин:

а) $\eta_1 = \lambda \xi$; б) $\eta_2 = \xi^2$; в) $\eta_3 = \sqrt{\xi}$; г) $\eta_4 = 1 - e^{-\lambda \xi}$.

9. Случайная величина ξ распределена по нормальному закону с параметрами a и σ^2 . Найти плотность распределения случайной величины η , если:

а) $a = 2, \sigma = 2, \eta = (\xi - 2)^3$;

б) $a = -1, \sigma = 3, \eta = \xi^3 + 2$;

в) $a = 2, \sigma = 1, \eta = (\xi - 1)^3$;

г) $a = 1, \sigma = 2, \eta = e^{\xi}$;

д) $a = 1, \sigma = 3, \eta = \sqrt[3]{\xi}$.

10. Случайная величина ξ имеет стандартное распределение Коши. Найти плотность распределения случайной величины η , если:

а) $\eta = \xi^3$; б) $\eta = \operatorname{arctg} \xi$; в) $\eta = e^{\xi}$.

11. Плотность распределения случайной величины ξ равна $p_{\xi}(x)$. Найти плотность распределения η , если:

а) $p_{\xi}(x) = \begin{cases} \frac{x+4}{20}, & x \in [-1, 3] \\ 0, & x \notin [-1, 3] \end{cases}; \quad \eta = (\xi + 1)^3$;

б) $p_{\xi}(x) = \begin{cases} \frac{x-3}{4}, & x \in [4, 6] \\ 0, & x \notin [4, 6] \end{cases}; \quad \eta = \xi^2 - 6$;

в) $p_{\xi}(x) = \begin{cases} \frac{2(3-x)}{63}, & x \in [-5, 2] \\ 0, & x \notin [-5, 2] \end{cases}; \quad \eta = 1/(\xi + 5)^2$;

г) $p_{\xi}(x) = \begin{cases} \frac{2(5-x)}{55}, & x \in [2, 5] \\ 0, & x \notin [2, 5] \end{cases}; \quad \eta = \ln(\xi + 4)$;

д) $p_{\xi}(x) = \begin{cases} \frac{8-x}{28}, & x \in [-1, 3] \\ 0, & x \notin [-1, 3] \end{cases}; \quad \eta = e^{\xi} - 1$.

12. Плотность распределения случайной величины ξ равна $p_{\xi}(x)$. Найти плотность распределения η , если:

$$\text{а) } p_{\xi}(x) = \begin{cases} \frac{3(x^2-1)}{20}, & x \in [1, 3]; \\ 0, & x \notin [1, 3] \end{cases}; \quad \eta = \sqrt{\xi+1};$$

$$\text{б) } p_{\xi}(x) = \begin{cases} \frac{x^2+1}{24}, & x \in [1, 4]; \\ 0, & x \notin [1, 4] \end{cases}; \quad \eta = \sqrt{\xi-1};$$

$$\text{в) } p_{\xi}(x) = \begin{cases} \frac{3(9-x^2)}{100}, & x \in [-3, 2]; \\ 0, & x \notin [-3, 2] \end{cases}; \quad \eta = 1/(\xi+3);$$

$$\text{г) } p_{\xi}(x) = \begin{cases} \frac{2(x^2+x)}{57}, & x \in [1, 4]; \\ 0, & x \notin [1, 4] \end{cases}; \quad \eta = \ln(2\xi+1);$$

$$\text{д) } p_{\xi}(x) = \begin{cases} \frac{x^2-2x}{18}, & x \in [2, 5]; \\ 0, & x \notin [2, 5] \end{cases}; \quad \eta = \ln(\xi-1);$$

$$\text{е) } p_{\xi}(x) = \begin{cases} \frac{x^3+1}{324}, & x \in [2, 6]; \\ 0, & x \notin [2, 6] \end{cases}; \quad \eta = 2/(1-\xi);$$

$$\text{ж) } p_{\xi}(x) = \begin{cases} \frac{1}{(\ln 2)(x+5)}, & x \in [-1, 3]; \\ 0, & x \notin [-1, 3] \end{cases}; \quad \eta = e^{3-\xi};$$

$$\text{з) } p_{\xi}(x) = \begin{cases} \frac{1}{(\ln 2)(8-x)}, & x \in [-2, 3]; \\ 0, & x \notin [-2, 3] \end{cases}; \quad \eta = e^{\xi+2}.$$

13. Продавец прохладительных напитков установил, что его доход (за день) в летний период зависит от среднесуточной температуры воздуха экспоненциальным образом и при повышении температуры от 20°C до 30°C растет от 1000 до 2000 денежных единиц. В предположении, что температура равномерно распределена от 20°C до 30°C , найти: а) вероятность того, что доход составит более 1500 денежных единиц; б) средний доход; в) среднее квадратическое отклонение дохода.

14. Предприниматель взял кредит в 100 денежных единиц под сложные проценты (0,1% в день) на время реализации некоторого проекта. Определить, какую сумму в среднем ему придется выплатить, если время реализации проекта: а) равномерно распределено от 100 до 200 дней; б) показательно распределено со средним 150 дней.
15. Прибыль X производителя зависит от рыночной цены p на выпускаемый товар по закону: $X = 90p^2$. В предположении, что возможная цена равномерно распределена от 10 до 12 (денежных единиц), найти: а) вероятность того, что прибыль составит более 10 тыс. денежных единиц; б) среднюю прибыль; в) среднее квадратическое отклонение прибыли.
16. Прибыль X производителя зависит от рыночной цены p на расходуемое сырье по закону: $X = 10^6/p^3$. В предположении, что возможная цена равномерно распределена от 4 до 8 (денежных единиц), найти: а) вероятность того, что прибыль составит более 5 тыс. денежных единиц; б) среднюю прибыль.
17. Найти плотность распределения суммы двух независимых величин ξ и η , равномерно распределенных на отрезках I и J соответственно, если:
- а) $I = [1, 3]$, $J = [0, 1]$;
 - б) $I = [0, 2]$, $J = [3, 4]$;
 - в) $I = [0, 4]$, $J = [1, 2]$;
 - г) $I = [1, 3]$, $J = [2, 4]$;
 - д) $I = [0, 1]$, $J = [0, 7]$.
18. Найти плотность распределения суммы независимых случайных величин ξ и η , где ξ имеет равномерное распределение на отрезке I , а η имеет показательное распределение с параметром λ , если:
- а) $I = [0, 1]$, $\lambda = 2$;
 - б) $I = [0, 3]$, $\lambda = 2$;
 - в) $I = [0, 2]$, $\lambda = 5$;
 - г) $I = [-1, 1]$, $\lambda = 3$;
 - д) $I = [1, 2]$, $\lambda = 2$.
19. Найти плотность распределения суммы независимых показательно распределенных случайных величин ξ и η с параметрами λ_ξ и λ_η соответственно, если:
- а) $\lambda_\xi = 2$, $\lambda_\eta = 3$;
 - б) $\lambda_\xi = 2$, $\lambda_\eta = 5$;
 - в) $\lambda_\xi = 4$, $\lambda_\eta = 6$;
 - г) $\lambda_\xi = 1$, $\lambda_\eta = 5$;
 - д) $\lambda_\xi = 4$, $\lambda_\eta = 5$.

20. Работа состоит из двух частей, которые выполняются независимо и последовательно. Время выполнения первой части равномерно распределено от 1 до 2 часов, второй — от 1 до 3 часов. Найти: а) среднее время выполнения работы; б) вероятность того, что работа будет выполнена за 4 часа.
21. Работа состоит из двух частей, которые выполняются независимо и последовательно. Время выполнения первой части равномерно распределено от 1 до 3 часов, второй — от 2 до 4 часов. Найти: а) среднее время выполнения работы; б) вероятность того, что работа будет выполнена за 5 часов.
22. Работа состоит из двух частей, которые выполняются независимо и последовательно. Время выполнения первой части показательным распределено со средним 1 час, второй — со средним 2 часа. Найти вероятность того, что работа будет выполнена за 4 часа.
23. Работа состоит из двух частей, которые выполняются независимо и последовательно. Время выполнения первой части показательным распределено со средним 2 часа, второй — со средним 3 часа. Найти вероятность того, что работа будет выполнена за 6 часов.
24. Двумерная случайная величина (ξ, η) равномерно распределена на множестве A . Проверить величины ξ и η на зависимость, найти их плотности и $F(C)$, если:
- а) $A = \{(x - 1)^2 + y^2 \leq 9, y \geq 0\}$, $C = (1, 3)$;
 - б) $A = \{x^2 + (y - 3)^2 \leq 64\}$, $C = (0, 3)$;
 - в) $A = \{(x - 2)^2 + y^2 \leq 25, y \geq 0\}$, $C = (2, 5)$;
 - г) $A = \{x^2 + (y - 5)^2 \leq 49\}$, $C = (0, 5)$;
 - д) $A = \{x^2 + (y - 2)^2 \leq 9, x \geq 0\}$, $C = (3, 2)$.
25. Двумерная случайная величина (ξ, η) равномерно распределена на множестве A . Проверить величины ξ и η на зависимость, найти их плотности и $F(C)$, если:
- а) $A = \{|x| + |y - 3| \leq 10\}$, $C = (5, 8)$;
 - б) $A = \{|x| + |y - 4| \leq 6\}$, $C = (3, 7)$;
 - в) $A = \{|x - 1| + |y| \leq 4\}$, $C = (0, -1)$;
 - г) $A = \{|x| + 2|y| \leq 6\}$, $C = (-1, -1)$;
 - д) $A = \{|x| \leq 3; 0 \leq y \leq 3 - x\}$, $C = (1, 1)$;
 - е) $A = \{0 \leq x \leq y \leq 2\}$, $C = (1, 3/2)$;
 - ж) $A = \{-1 \leq y \leq 7 - 2|x - 1|\}$, $C = (2, 5)$;
 - з) $A = \{|x| + 2|y - 3| \leq 6\}$, $C = (4, 4)$.

26. Двумерная случайная величина (ξ, η) равномерно распределена на четырехугольнике, ограниченном прямыми $y = 0$, $y = 2$, $y = 2x + 2$, $y = 14 - 2x$. Проверить величины ξ и η на зависимость, найти их плотности и $F(1, 1)$.
27. Двумерная случайная величина (ξ, η) равномерно распределена на множестве A , ограниченном параболой и осью абсцисс: $A = \{0 < y < 4 - (x - 2)^2\}$. Проверить величины ξ и η на зависимость, найти их плотности и $F(2, 4)$.
28. Двумерная случайная величина (ξ, η) равномерно распределена в квадрате $K = \{(x, y): |x| + |y| \leq 2\}$. Найти вероятность

$$P\left(\frac{1}{2} \leq \xi \leq 1, \frac{1}{2} \leq \eta \leq 1\right).$$

29. Пара случайных величин ξ и η равномерно распределена внутри треугольника $K = \{(x, y): x + y \leq 1, x \geq 0, y \geq 0\}$. Найти вероятность $P(\xi \leq 3/4)$.
30. Пара случайных величин ξ и η равномерно распределена внутри трапеции с вершинами в точках $(-6, 0)$, $(-3, 4)$, $(3, 4)$, $(6, 0)$. Найти совместную плотность распределения для этой пары случайных величин и плотности составляющих. Зависимы ли ξ и η ?
31. Случайная пара (ξ, η) равномерно распределена внутри $K = \{(x, y): x \geq 0; x \leq y \leq 1\}$. Найти плотности ξ и η , исследовать вопрос об их зависимости. Найти $M(\xi \eta)$.
32. Совместная плотность двух случайных величин ξ и η равна:

$$f(x, y) = \begin{cases} 4e^{2y}, & y \leq 0, x \leq 0, y \leq x; \\ 0, & \text{иначе.} \end{cases}$$

Найти частные плотности ξ , η . Исследовать вопрос о зависимости ξ и η .

33. Дана совместная плотность случайных величин ξ и η :

$$p(x, y) = \begin{cases} A \sin(x + y), & 0 \leq x \leq \frac{\pi}{2}; 0 \leq y \leq \frac{\pi}{2}; \\ 0, & \text{иначе.} \end{cases}$$

Найти: а) коэффициент A ; б) совместную функцию распределения; в) частные плотности и функции распределения; г) условную плотность распределения ξ при условии $\eta = y$.

34. Функция совместного распределения случайного вектора (ξ, η) имеет вид

$$F(x, y) = xy(1 + \alpha(1 - x)(1 - y)), \quad x, y \in [0, 1].$$

- Найти: а) совместную и частные плотности распределения; б) при каких α данная формула действительно задает функцию распределения; в) коэффициент корреляции; г) функцию регрессии ξ по η .
35. Случайная пара (ξ, η) равномерно распределена в единичном квадрате $K = \{(x, y): 0 \leq x \leq 1; 0 \leq y \leq 1\}$. Пусть ρ — расстояние от точки (ξ, η) до нуля. Найти функцию условного распределения ρ при условии $\xi = x, 0 \leq x \leq 1$.
36. Пара случайных величин ξ и η равномерно распределена в треугольнике с вершинами в точках $(0, 0), (1, 1), (2, -1)$. Найти: а) функцию регрессии ξ по η ; б) функцию регрессии η по ξ .
37. Две независимые случайные величины ξ_1 и ξ_2 имеют показательное распределение с параметром λ . Найти условное распределение ξ_1 при условии, что $\eta = \xi_1 + \xi_2 = t, t > 0$.
38. Случайные приращения цен акций А и Б (в процентах) имеют дисперсии 1 и 4 соответственно и нулевые средние, а их коэффициент корреляции 0,7. Как в среднем должен измениться курс акции: а) Б, если курс акции А поднимется на 2%; б) А, если курс акции Б опустится на 4%. Решить задачу, используя приближение двумерным нормальным распределением.
39. Случайный выпуск продукта в циклическом технологическом процессе на химическом производстве имеет среднее 50 и среднее квадратическое отклонение 12 (весовых единиц), причем выпуск в каждом цикле зависит от предыдущего с коэффициентом корреляции $-0,4$. Спрогнозировать выпуск в следующем цикле, если в текущем цикле выпуск составил: а) 45; б) 65 весовых единиц. Решить задачу, используя приближение двумерным нормальным распределением.
40. Два экономических показателя, X и Y , имеют средние 10 и 20, дисперсии 25 и 64 соответственно, а их коэффициент корреляции 0,5. Найти: а) среднее Y при условии $X = 15$; б) среднее X при условии $Y = 28$. Решить задачу, используя приближение двумерным нормальным распределением.

Глава 8

ЗАКОН БОЛЬШИХ ЧИСЕЛ. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

8.1. Неравенство Чебышева

► **Теорема 1 (неравенство Маркова).** Пусть для случайной величины ξ существует $M|\xi|$, тогда для любого $\varepsilon > 0$ верно

$$P(|\xi| \geq \varepsilon) \leq \frac{M|\xi|}{\varepsilon}.$$

Доказательство. Проведем доказательство отдельно в дискретном и в абсолютно непрерывном случае.

Пусть ξ — дискретная случайная величина и $P(\xi = x_i) = p_i$, $i = 1, 2, \dots, n$. Тогда вероятность $P(|\xi| \geq \varepsilon)$ равна сумме вероятностей p_i для которых x_i находятся вне промежутка $(-\varepsilon, \varepsilon)$.

Очевидно, для таких x_i имеет место неравенство $\frac{|x_i|}{\varepsilon} \geq 1$. Учитывая это неравенство, получаем

$$\begin{aligned} P(|\xi| \geq \varepsilon) &= \\ &= \sum_{|x_i| \geq \varepsilon} p_i \leq \sum_{|x_i| \geq \varepsilon} \frac{|x_i|}{\varepsilon} p_i \leq \sum_{|x_i| \geq \varepsilon} \frac{|x_i|}{\varepsilon} p_i + \sum_{|x_i| < \varepsilon} \frac{|x_i|}{\varepsilon} p_i = \frac{1}{\varepsilon} \sum_{i=1}^n |x_i| p_i = \frac{1}{\varepsilon} M|\xi|. \end{aligned}$$

Пусть ξ — абсолютно непрерывная случайная величина с плотностью распределения $p_\xi(x)$. Тогда вероятность $P(|\xi| \geq \varepsilon)$

равна сумме интегралов от плотности распределения по промежуткам $(-\infty, -\varepsilon)$ и $(\varepsilon, +\infty)$. На этих промежутках $\frac{|x|}{\varepsilon} \geq 1$. Воспользовавшись формулой для математического ожидания получаем неравенство

$$\begin{aligned} P(|\xi| \geq \varepsilon) &= \int_{-\infty}^{-\varepsilon} p_{\xi}(x) dx + \int_{\varepsilon}^{\infty} p_{\xi}(x) dx \leq \frac{1}{\varepsilon} \int_{-\infty}^{-\varepsilon} |x| p_{\xi}(x) dx + \\ &+ \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} |x| p_{\xi}(x) dx \leq \frac{1}{\varepsilon} \int_{-\infty}^{-\varepsilon} |x| p_{\xi}(x) dx + \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} |x| p_{\xi}(x) dx + \\ &+ \frac{1}{\varepsilon} \int_{\varepsilon}^{\infty} |x| p_{\xi}(x) dx = \frac{1}{\varepsilon} \int_{-\infty}^{\infty} |x| p_{\xi}(x) dx = \frac{1}{\varepsilon} M|\xi|. \end{aligned}$$

В общем случае подобное доказательство проводится с использованием интеграла Лебега, который не входит в программу настоящего курса.

Следствие. Пусть для случайной величины ξ существует $M\xi^2$, тогда для любого $\varepsilon > 0$ верно

$$P(|\xi| \geq \varepsilon) \leq \frac{M\xi^2}{\varepsilon^2}.$$

Доказательство. Заметим, что событие $\{\omega: |\xi| \geq \varepsilon\}$ равносильно событию $\{\omega: \xi^2 \geq \varepsilon^2\}$, и применим неравенство Маркова к случайной величине ξ^2 .

► **Теорема 2 (неравенство Чебышева).** Пусть для случайной величины ξ существует математическое ожидание $M\xi$ и дисперсия $D\xi$. Тогда для любого $\varepsilon > 0$ верно

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}.$$

Доказательство. Применяем следствие неравенства Маркова к случайной величине $|\xi - M\xi|$, замечая, что $M(|\xi - M\xi|)^2 = D\xi$.

Неравенство Чебышева часто используют в виде

$$P(|\xi - M\xi| < \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2}.$$

► **Теорема 3.** Пусть для случайных величин $\xi_1, \xi_2, \dots, \xi_n$ существуют математические ожидания $M\xi_i$ и дисперсии $D\xi_i$, тогда для любого $\varepsilon > 0$ верно

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i - \frac{1}{n}\sum_{i=1}^n M\xi_i\right| < \varepsilon\right) \geq 1 - \frac{1}{n^2\varepsilon^2} \sum_{i=1}^n D\xi_i.$$

Доказательство. Применяем неравенство Чебышева к случайной величине $\zeta = \frac{1}{n}\sum_{i=1}^n \xi_i$, которая имеет $M\zeta = \frac{1}{n}\sum_{i=1}^n M\xi_i$ и $D\zeta = \frac{1}{n^2}\sum_{i=1}^n D\xi_i$ (согласно свойствам математического ожидания и дисперсии).

Задача 1. В 400 испытаниях Бернулли вероятность успеха в каждом испытании равна 0,8. С помощью неравенства Чебышева оценить вероятность того, что разница между числом успехов в этих испытаниях и средним числом успехов будет меньше 20.

Решение. Число успехов в этих испытаниях распределено по закону Бернулли, поэтому среднее число успехов равно $M\xi = np = 400 \cdot 0,8 = 320$, а дисперсия $D\xi = npq = 400 \cdot 0,8 \times 0,2 = 64$. Тогда в силу неравенства Чебышева имеем

$$P(|\xi - 320| < 20) \geq 1 - \frac{D\xi}{20^2} = 1 - \frac{64}{400} = 0,84.$$

Вычислим эту же вероятность с помощью приближенной (интегральной) формулы Муавра—Лапласа (см. гл. 4):

$$\begin{aligned} P(|\xi - 320| < 20) &= P(|\xi - np| < \varepsilon) = P\left(-\frac{\varepsilon}{\sqrt{npq}} < \frac{\xi - np}{\sqrt{npq}} < \frac{\varepsilon}{\sqrt{npq}}\right) = \\ &= 2\Phi_0\left(\frac{\varepsilon}{\sqrt{npq}}\right) = 2\Phi_0\left(\frac{20}{\sqrt{64}}\right) = 2\Phi_0(2,5) = 2 \cdot 0,4938 = 0,9876. \end{aligned}$$

Последнее вычисление показывает, что неравенство Чебышева дает довольно грубые оценки вероятностей.

8.2. Закон больших чисел

► **Теорема 4 (закон больших чисел).** Пусть задана бесконечная последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, для которых существуют мате-

математическое ожидание $M\xi_i = a$ и дисперсия $D\xi_i = \sigma^2$. Тогда для любого $\varepsilon > 0$ верно

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right) = 0.$$

Доказательство. Применяем неравенство Чебышева к случайной величине $\zeta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$, которая имеет $M\zeta_n = \frac{1}{n} \sum_{i=1}^n M\xi_i = a$ и $D\zeta_n = \frac{1}{n^2} \sum_{i=1}^n D\xi_i = \frac{\sigma^2}{n}$. Получаем

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right) = P(|\zeta_n - a| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Суть закона больших чисел состоит в том, что при возрастании числа слагаемых (одинаково распределенных случайных величин) среднее арифметическое этих слагаемых мало отличается от математического ожидания a . Иначе говоря, любое положительное отклонение среднего арифметического случайных величин от числа a становится при достаточно большом числе слагаемых маловероятным.

Пример 1. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность случайных величин, каждая из которых равна числу успехов в одном испытании Бернулли (т.е. 1 — в случае успеха и 0 — в случае неудачи). Закон распределения каждой такой случайной величины имеет вид:

ξ_i	0	1
P	q	p

Тогда математическое ожидание $M\xi_i = p$ и дисперсия $D\xi_i = pq$. Среднее арифметическое $\bar{x} = \frac{1}{n} \sum_{i=1}^n \xi_i$ равно частоте успехов в n испытаниях, и закон больших чисел утверждает, что эта частота успехов стремится к вероятности успеха p , когда число слагаемых (т.е. число испытаний) неограниченно возрастает. Закон больших чисел может выполняться и в случае попарному распределенных случайных величин.

Теорема 5. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых случайных величин, для которых существуют математические ожидания $M\xi_i$ и дисперсии $D\xi_i$, удовлетворяющие условию $\frac{1}{n^2} \sum_{i=1}^n D\xi_i \rightarrow 0, n \rightarrow \infty$. Тогда для любого $\varepsilon > 0$ справедливо равенство

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| < \varepsilon \right) = 1.$$

Доказательство. В силу теоремы 3 имеем

$$1 - \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n D\xi_i \leq P \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| < \varepsilon \right) \leq 1.$$

Переходя к пределу в неравенстве при $n \rightarrow \infty$, получаем утверждение теоремы.

Отметим, что для независимых одинаково распределенных случайных величин верно также следующее утверждение, более сильное, чем закон больших чисел. При этом можно даже отказаться от условия существования дисперсии.

► **Теорема 6 (усиленный закон больших чисел Колмогорова).** Пусть задана бесконечная последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, для которых существует математическое ожидание $M\xi_i = a$, тогда

$$P \left(\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right) = a \right) = 1.$$

В общем случае сходимость случайной последовательности X_n к пределу X вида

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1$$

называется *сходимостью с вероятностью единица* (или *почти на-верное*) и обозначается так:

$$X_n \xrightarrow{n.s.} X.$$

Из сходимости с вероятностью единица следует сходимость по вероятности, обратное верно не всегда.

8.3. Центральная предельная теорема (ЦПТ)

► **Теорема 7 (центральная предельная теорема).** Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — независимые одинаково распределенные случайные величины с $M\xi_i = a$ и $D\xi_i = \sigma^2$, $i = 1, 2, \dots, n, \dots$ Тогда для любого числа x верно

$$\lim_{n \rightarrow \infty} P\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n - na}{\sigma\sqrt{n}} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Смысл центральной предельной теоремы заключается в том, что сумма $S_n = \sum_{i=1}^n \xi_i$ случайных величин при надлежащем линейном преобразовании с увеличением числа слагаемых ($n \rightarrow \infty$) ведет себя почти как случайная величина со стандартным нормальным распределением $N(0, 1)$. Вероятность попадания в любой интервал имеет следующий предел:

$$\lim_{n \rightarrow \infty} P\left(c_1 < \frac{S_n - na}{\sigma\sqrt{n}} < c_2\right) = \frac{1}{\sqrt{2\pi}} \int_{c_1}^{c_2} e^{-\frac{y^2}{2}} dy = \Phi(c_2) - \Phi(c_1) = \Phi_0(c_2) - \Phi_0(c_1),$$

откуда следует приближенная формула

$$P(A < S_n < B) \approx \Phi_0\left(\frac{B - na}{\sigma\sqrt{n}}\right) - \Phi_0\left(\frac{A - na}{\sigma\sqrt{n}}\right),$$

которой имеет смысл пользоваться, если A и B не очень далеки от na .

Центральная предельная теорема может оставаться справедливой и в случае, когда случайные слагаемые не являются одинаково распределенными.

► **Теорема 8 (Ляпунова).** Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, имеющие конечный третий абсолютный центральный момент, и пусть $a_k = M\xi_k$, $\sigma_k^2 = D\xi_k$, $c_k^3 = M|\xi_k - a_k|^3$,

$A_n = \sum_{k=1}^n a_k$, $B_n^2 = \sum_{k=1}^n \sigma_k^2$, $C_n^3 = \sum_{k=1}^n c_k^3$. Если при $n \rightarrow \infty$ отношение

$\frac{C_n}{B_n} \rightarrow 0$, то для любого числа x верно

$$\lim_{n \rightarrow \infty} P\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n - A_n}{B_n} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

В общем случае сходимость случайной последовательности X_n к пределу (случайной величине) X вида

$$\lim_{n \rightarrow \infty} P(X_n < x) = P(X < x)$$

в точках непрерывности функции распределения X называется *сходимостью по распределению* и обозначается так:

$$X_n \xrightarrow{d} X.$$

Из сходимости по вероятности следует сходимость по распределению, обратное верно не всегда (но верно в случае сходимости к константе).

Пример 2. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность случайных величин, удовлетворяющая условиям предыдущего примера 1. В этом случае сумма $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ есть число успехов m в n испытаниях Бернулли. Из ЦПТ следует, что

$$\lim_{n \rightarrow \infty} P\left(c_1 < \frac{m - np}{\sqrt{npq}} < c_2\right) = \frac{1}{\sqrt{2\pi}} \int_{c_1}^{c_2} e^{-\frac{x^2}{2}} dx = \Phi_0(c_2) - \Phi_0(c_1),$$

где $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{y^2}{2}} dy$ — функция Лапласа.

Тогда вероятность того, что число успехов заключено между m_1 и m_2 , равна

$$\begin{aligned} P(m_1 \leq m \leq m_2) &= P\left(\frac{m_1 - np}{\sqrt{npq}} \leq \frac{m - np}{\sqrt{npq}} \leq \frac{m_2 - np}{\sqrt{npq}}\right) \approx \\ &\approx \Phi_0\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{m_1 - np}{\sqrt{npq}}\right). \end{aligned}$$

Этот результат называется *интегральной теоремой Муавра—Лапласа* и уже встречался нам раньше (см. § 4.3).

Задача 2. В продукции цеха детали отличного качества составляют 50%. Детали укладываются в коробки по 200 шт. в каждой. Какова вероятность того, что число деталей отличного качества в коробке отличается от 100 не более чем на 5?

Решение. Пусть ξ_i — случайное число деталей отличного качества в i -й коробке, тогда при $n = 200$, $p = q = \frac{1}{2}$ получим:

$$P(95 \leq m \leq 105) = P\left(-\frac{5}{\sqrt{50}} \leq \frac{m-np}{\sqrt{npq}} \leq \frac{5}{\sqrt{50}}\right) \approx \Phi_0(0,71) - \Phi_0(-0,71) \approx 0,52.$$

Задача 3. Используя условия задачи 2, указать, в каких границах с вероятностью 0,997 находится число деталей отличного качества в коробке.

Решение. По табл. 2 приложения 2 при условии

$P\left(\left|\frac{m-np}{\sqrt{npq}}\right| \leq u\right) \approx 0,997$ находим $u = 3$, и следовательно, S_n лежит в пределах $np \pm 3\sqrt{npq}$, т.е. число деталей отличного качества в коробке с вероятностью 0,997 находится в пределах 100 ± 21 .

Задача 4. Используя условия задачи 2, определить, сколько деталей надо положить в коробку, чтобы с вероятностью, не меньшей 0,99, можно было утверждать, что число деталей отличного качества в коробке не менее 100.

Решение. Обозначим $u = \frac{100-np}{\sqrt{npq}}$. Используя нормальное приближение, получаем

$$P(m \geq 100) = P\left(\frac{m-np}{\sqrt{npq}} \geq u\right) \approx 1 - \Phi(u) = \frac{1}{2} - \Phi_0(u) \geq 0,99.$$

Отсюда $\Phi_0(u) \leq -0,49$, а из табл. 2 приложения 2 и свойств функции Лапласа получаем неравенство $u \leq -2,32$. Обозначив $x = \sqrt{n} > 0$, с учетом $p = q = \frac{1}{2}$, приходим к квадратному неравенству $x^2 - 2,3x - 200 \geq 0$, решая которое, получаем $n \geq 236$.

Можно предложить и другой метод. Пусть ξ_i — число деталей, которые пришлось перебрать, чтобы найти i -ю деталь отличного качества (включая ее саму). Случайные величины имеют геометрическое распределение с параметром $p = 1/2$. Можем вычислить $M\xi = 1/p = 2$, $D\xi = (1-p)/p^2 = 2$. Используя ЦПТ, получаем неравенство

$$P(S_{100} \leq n) = \Phi\left(\frac{n-100 \cdot 2}{\sqrt{2 \cdot 100}}\right) = \frac{1}{2} + \Phi_0\left(\frac{n-200}{14,14}\right) \geq 0,99,$$

откуда следует $n \geq 200 + 14,14 \cdot 2,32 = 232,8$, или, округляя, $n \geq 234$.

Результаты получаются близкие, но первый метод более точен и потому предпочтительнее. Вторым методом лучше пользоваться, если нужно определить границы, в которых лежит неизвестное число деталей.

Задача 5. Доходы (в месяц) жителей города имеют математическое ожидание 10 тыс. руб. и среднее квадратическое отклонение 2 тыс. руб. Найти вероятность того, что средний доход 100 случайно выбранных жителей составит от 9,5 до 10,5 тыс. руб.

Решение. Переформулируем условие задачи для суммарного дохода: он должен составлять от 950 до 1050 тыс. руб. Используя ЦПТ, получаем:

$$P(950 < S_{100} < 1050) = \Phi_0\left(\frac{1050 - 100 \cdot 10}{2\sqrt{100}}\right) - \Phi_0\left(\frac{950 - 100 \cdot 10}{2\sqrt{100}}\right) = \\ = 2\Phi_0(2,5) = 0,9876.$$

Задача 6. Срок службы электрической лампы имеет показательное распределение с математическим ожиданием 1000 ч. Найти вероятность того, что средний срок службы для 100 ламп составит не менее 900 ч.

Решение. Примем для простоты 1000 ч за единицу времени. Вспомним числовые характеристики показательного распределения: $M\xi = \frac{1}{\lambda}$, $D\xi = \frac{1}{\lambda^2}$. Отсюда следует, что среднее квадратическое отклонение совпадает с математическим ожиданием (и оба они здесь равны единице). Переформулируя условие задачи для суммарного срока службы и используя ЦПТ, получаем:

$$P(S_{100} \geq 90) = 1 - \Phi\left(\frac{90 - 100}{\sqrt{100}}\right) = \frac{1}{2} + \Phi_0(1) = 0,8413.$$

Задачи для самостоятельного решения

Теоретические задачи

1. Проверить, выполняется ли закон больших чисел для последовательности независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, имеющих следующие законы распределения:

а)

ξ_n	$-\sqrt{n}$	0	$\sqrt{n}$
P	$1/n$	$1 - \frac{2}{n}$	$1/n$

б)

ξ_n	-2^n	0	2^n
P	$2^{-(2n+1)}$	$1 - 2^{-2n}$	$2^{-(2n+1)}$

в)

ξ_n	$-\sqrt{n}$	0	$\sqrt{n}$
P	$1/\sqrt{n}$	$1 - 2/\sqrt{n}$	$1/\sqrt{n}$

г)

ξ_n	$-n^{2/3}$	0	$n^{2/3}$
P	$1/n$	$1 - \frac{2}{n}$	$1/n$

2. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — независимые, одинаково распределенные случайные величины с функцией распределения $F(x)$. Пусть заданы случайные величины $\eta_i = \begin{cases} 1, & \xi_i < x, \\ 0, & \xi_i \geq x. \end{cases}$

Выполняется ли для последовательности $\eta_1, \eta_2, \dots, \eta_i, \dots$ закон больших чисел?

3. Доказать закон больших чисел в «обобщенной форме»: пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых случайных величин, у которых существуют математические ожидания $M\xi_i$ и дисперсии $D\xi_i$, причем все дисперсии ограничены сверху одной константой $C > 0$. Тогда для любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n M\xi_i \right| \geq \varepsilon \right) = 0.$$

4. Пусть случайная величина ξ имеет нормальное распределение с параметрами $M\xi = a$, $D\xi = \sigma^2$. Найти вероятности $P(|\xi - a| \geq \sigma)$ и $P(|\xi - a| \geq 3\sigma)$, пользуясь таблицами функции Лапласа. Затем оцените те же вероятности с помощью неравенства Чебышева.
5. Пусть случайная величина ξ имеет распределение Лапласа, т.е. ее плотность равна $P(x) = \frac{\lambda}{2} e^{-\lambda|x|}$, $\lambda > 0$. Найти вероятности $P(|\xi| < \sigma)$ и $P(|\xi| < 3\sigma)$, где σ — среднее квадратическое отклонение, и сравнить их с оценками, получаемыми с помощью неравенства Чебышева.

Вычислительные задачи

6. Средний размер вклада в отделении банка равен 6000 руб. Оценить вероятность, что случайно взятый вклад не превысит 10 000 руб.
7. Среднее количество вызовов, поступающих на АТС завода в течение часа, равно 300. Оценить вероятность того, что в течение следующего часа число вызовов на коммутатор: а) превысит 400; б) будет не более 600.
8. Отделение банка обслуживает в среднем 100 клиентов в день. Оценить вероятность того, что сегодня в отделении банка будет обслужено: а) не более 200 клиентов; б) более 150 клиентов.
9. Среднее абсолютное приращение цены акции компании в течение биржевых торгов составляет 0,3%. Оценить вероятность того, что на ближайших торгах курс изменится более чем на 3%.
10. Среднее квадратическое отклонение приращения цены акции компании в течение биржевых торгов составляет 2%, а среднее приращение равно нулю. Оценить вероятность того, что на ближайших торгах курс изменится более чем на 5%.
11. В каждой упаковке товара имеется одна из N различных наклеек (равновероятно). С помощью неравенства Чебышева оценить вероятность того, что удастся собрать все наклейки, купив не более M упаковок товара. Решить задачу, если: а) $N = 4$, $M = 20$; б) $N = 5$, $M = 30$.
12. По статистическим данным, в среднем 87% новорожденных доживают до 50 лет. С помощью неравенства Чебышева оценить вероятность того, что из 1000 новорожденных доля доживших до 50 лет будет отличаться от вероятности этого события не более чем на 0,04 (по абсолютной величине).
13. В среднем 10% работоспособного населения некоторого региона — безработные. Оценить с помощью неравенства Чебышева вероятность того, что уровень безработицы среди случайно выбранных 10 000 работоспособных жителей составит от 9 до 11% (включительно).
14. Опыт страховой компании показывает, что страховой случай приходится примерно на каждый пятый договор. Оценить с помощью неравенства Чебышева количество договоров, при котором с вероятностью 0,9 можно утверждать, что доля страховых случаев отклонится от 0,2 не более чем на 0,01 (по абсолютной величине). Уточнить ответ с помощью теоремы Муавра—Лапласа.

15. В продукции цеха детали отличного качества составляют 80%. В каких границах будет находиться с вероятностью 0,99 число деталей отличного качества среди 10 000 деталей? Сделать оценку с помощью неравенства Чебышева и с помощью теоремы Муавра—Лапласа.
16. Ежедневно новая сделка заключается с вероятностью 0,2 (но не более одной в день). За сколько дней с вероятностью 0,9 можно ожидать заключения не менее 50 сделок?
17. Аппаратура состоит из 100 одинаково надежных и независимо работающих элементов, каждый из которых может отказать в течение суток с вероятностью 0,01. На обнаружение отказавшего элемента и его замену требуется 20 минут, в течение которых аппаратура простаивает. Найти: а) вероятность того, что время простоя составит не более 40 минут в сутки; б) среднее время простоя аппаратуры в сутки.
18. На заводе 1000 станков, каждый из которых в среднем в течение 24 дней в месяц потребляет электроэнергию независимо от других станков с интенсивностью 10 единиц в день. Какое количество электроэнергии необходимо заводу ежедневно, чтобы недостаток электроэнергии наблюдался в среднем не чаще двух раз за 100 дней?
19. Предприятие выпускает 30% изделий стоимостью 100 руб., 30% изделий — стоимостью 200 руб. и 40% изделий — стоимостью 300 руб. Какова вероятность получить за 1000 случайно отобранных изделий не менее 215 тыс. руб.?
20. Известно, что треть всех деталей, сходящих с конвейера, подвергается выборочному контролю на основании некоторого случайного признака. Пусть через контроль прошло 100 деталей. В каких границах с вероятностью 0,99 лежит общее число деталей, сошедших с конвейера?
21. Кадровое агентство трудоустроило 500 специалистов. В каких границах с вероятностью 95% находилось общее число соискателей на работу, если каждый соискатель устраивается на работу с вероятностью 0,2?
22. Изготовление детали занимает случайное время, равномерно распределенное от 10 до 15 минут. Найти вероятность того, что на изготовление 100 деталей понадобится не менее 20,5 часа.
23. Изготовление детали занимает случайное время, равномерно распределенное от 4 до 8 минут. Изготовление скольких деталей за 25 рабочих часов можно гарантировать с вероятностью 95%?

24. В бункер помещается не более 150 деталей. Ежеминутно в бункер поступает случайное число деталей, имеющее распределение Пуассона с параметром $\lambda = 2$; количества деталей, поступающих в непересекающиеся промежутки времени, независимы. Каждый час все находящиеся в бункере детали перегружают на тележку и увозят. В начальный момент времени бункер свободен. Найти вероятность того, что за 20 часов непрерывной работы не произойдет ни одного переполнения бункера.
25. Число посетителей магазина (в день) имеет распределение Пуассона с математическим ожиданием 289. Найти вероятность того, что за 100 рабочих дней суммарное число посетителей составит от 28 550 до 29 250 человек.
26. По статистике страховой компании, число страховых случаев в день имеет распределение Пуассона, а за год в среднем происходит 400 страховых случаев. Найти вероятность того, что за год их произойдет не более 450.
27. Доходы жителей города имеют математическое ожидание 15 тыс. руб. и среднее квадратическое отклонение — 4 тыс. руб. (в месяц). Найти вероятность того, что средний доход 100 случайно выбранных жителей составит от 14 до 16 тыс. руб.
28. Срок службы изделия имеет показательное распределение с математическим ожиданием 2500 часов. Найти вероятность того, что средний срок службы для 100 изделий составит не менее 2300 часов.
29. Вес яблока имеет математическое ожидание 200 г и среднее квадратическое отклонение 50 г. Найти вероятность того, что в 100 кг окажется не менее 490 яблок.
30. Вес арбуза имеет математическое ожидание 10 кг и среднее квадратическое отклонение 2 кг. Найти вероятность того, что в 1 тонне окажется не более 105 арбузов.

Глава 10

ОСНОВНЫЕ ПОНЯТИЯ И ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

10.1. Генеральная и выборочная совокупность

Генеральной совокупностью называется множество объектов произвольной природы, обладающих признаками, доступными для наблюдения и количественного измерения.

Например, в случае социально-экономических исследований это может быть население какого-то города, региона или страны, а измеряемыми признаками — доходы, расходы или объем сбережений отдельно взятого человека. Если какой-то признак имеет качественный характер (например, пол, национальность, социальное положение, род деятельности и т.п.), но принадлежит к конечному множеству вариантов, он может быть также закодирован числом (как это часто делают в анкетах).

Объекты, входящие в генеральную совокупность, называются ее **элементами**, а их общее число — ее **объемом**.

Пусть число элементов генеральной совокупности равно N . Примем каждый из них за элементарный исход ω некоторого вероятностного пространства Ω и припишем всем исходам одинаковую вероятность $1/N$. Тогда соответствие между объектами и значениями какого-либо их признака задает случайную величину $\xi = \xi(\omega)$ как функцию на вероятностном простран-

б) время дискретно.

Далее будем изучать цепи Маркова с конечным или счетным числом состояний, как с дискретным, так и с непрерывным временем.

9.2. Цепи Маркова с конечным числом состояний и дискретным временем

В этом параграфе рассмотрим цепи Маркова:

а) с пространством состояний S из конечного числа элементов m ;

б) с дискретным временем n , принимающем значения $0, 1, 2$ и т.д.

Обычно полагают, что элементы S занумерованы числами $1, 2, \dots, m$.

Для таких цепей марковское свойство может быть записано в следующей форме:

$$\begin{aligned} P(\xi(n+1)=i_{n+1} | \xi(n)=i_n, \xi(n-1)=i_{n-1}, \dots, \xi(0)=i_0) = \\ = P(\xi(n+1)=i_{n+1} | \xi(n)=i_n). \end{aligned}$$

Однородными называются цепи Маркова, для которых условные вероятности $P(\xi(n+1)=j | \xi(n)=i)$ не зависят от n . Таким образом, однородная цепь ведет себя с любого момента так же, как с начала ($n=0$). Далее будем рассматривать только однородные цепи Маркова.

Вероятности $p_{ij} = P(\xi(n+1)=j | \xi(n)=i)$ называются *переходными вероятностями*, а составленная из них матрица P — *матрицей переходных вероятностей*:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{pmatrix}.$$

Поскольку в каждой строке матрицы записаны вероятности всех возможных переходов из выбранного состояния (в том числе и вероятности того, что система останется в нем), эти переходы образуют полную группу событий. Поэтому для каждой строки имеет место равенство $\sum_{j=1}^m p_{ij} = 1$ для любого $i = 1, 2, \dots, m$.

Распределение $\xi(n)$ задается вектором $p(n) = (p_1(n), \dots, p_m(n))$ где $p_i(n) = P(\xi(n) = i)$.

По формуле полной вероятности получаем

$$p(n+1) = p(n)P.$$

Начальным распределением цепи Маркова ξ называется распределение $\xi(0)$, заданное соответствующим вектором $p(0)$. Зная начальное распределение, можно найти $p(n)$ при любом $n > 0$, последовательно вычисляя $p(1) = p(0)P$, $p(2) = p(1)P$ и т.д. В общем случае получаем формулу

$$p(n) = p(0)P^n.$$

Отсюда следует, что вероятности $P_{ij}(n)$ перехода из состояния i в состояние j за время n можно найти как элементы n -й степени матрицы P , т.е. $(P_{ij}(n)) = P^n$.

Говорят, что из состояния i можно перейти в состояние j , если существует n , такое, что $P_{ij}(n) > 0$. Если из состояния i можно перейти в состояние j , такое, что обратно вернуться нельзя, то состояние i называют *несущественным*; в противном случае — *существенным*. Если из состояния i можно перейти в j , а из j — в i , то такие состояния называют *сообщающимися*. Все существенные состояния цепи разбиваются на *классы сообщающихся состояний* (внутри каждого класса все состояния сообщаются между собой, но не сообщаются с состояниями других классов).

Класс называется *периодическим* с периодом d , если для любого состояния i из этого класса возможные времена возвращения в него (т.е. такие n , что $P_{ii}(n) > 0$) кратны некоторому числу $d \geq 2$. В противном случае класс называется *апериодическим*.

Эргодической называется цепь Маркова, для которой существует *предельное распределение* $\pi = (\pi_1, \dots, \pi_m)$:

$$\pi_j = \lim_{n \rightarrow \infty} P_{ij}(n), \quad \sum_{j=1}^m \pi_j = 1.$$

Если имеется лишь один класс сообщающихся состояний и он апериодический, то цепь Маркова является эргодической.

Предельное распределение имеет важное практическое значение, однако найти его из приведенной формулы затруднительно, поэтому используется иной подход.

Стационарным распределением цепи Маркова называется такое распределение, которое, будучи задано в качестве начального, в дальнейшем останется неизменным. В таком случае говорят, что система находится в *стационарном режиме*.

Для эргодических цепей Маркова стационарное распределение существует и единственно, а самое главное — совпадает с предельным. Таким образом, если система изначально не находится в стационарном режиме, она со временем (при $n \rightarrow \infty$) выходит на него.

Стационарное распределение $\pi = (\pi_1, \dots, \pi_m)$ легко найти из системы уравнений:

$$\pi = \pi P, \quad \sum_{j=1}^m \pi_j = 1.$$

Следует отметить, что это система из $m + 1$ уравнений с m неизвестными, так что одно из уравнений (любое из первых m) можно исключить.

Пример 1. Простейшая форма контроля качества продукции с переменным планом заключается в следующем. Устанавливается m различных объемов выборок (планов) в порядке убывания: $n_1 > n_2 > \dots > n_m$, $m \geq 2$. Если в выборке из партии изделий обнаружено хотя бы одно дефектное, партия бракуется. Схема контроля производства строится по следующему алгоритму. Первая партия изделий проверяется выборкой максимального объема n_1 . Если она оказалась принятой, делается вывод о нормальном ходе производства, и для следующей партии переходят на выборку меньшего объема n_2 ; в противном случае объем остается максимальным. При контроле партии выборками промежуточного объема n_k , $1 < k < m$, в случае приемки партии переходят на выборку меньшего объема n_{k+1} ; в случае отбраковки — на выборку большего объема n_{k-1} . При контроле партии выборкой минимального объема n_m в случае приемки этот объем сохраняется, в случае отбраковки переходят на выборку объема n_{m-1} .

В рассматриваемой модели номер используемого плана (объема выборки) образует цепь Маркова, так как при известном текущем значении его следующие значения не зависят от предыдущих. Пространство состояний в данном случае $S = \{1, 2, \dots, m\}$. Обозначим через r_i вероятность отклонить партию при объеме выборки n_i , $1 \leq i \leq m$. Тогда переходные вероятности принимают вид:

$$p_{ij} = \begin{cases} r_i, & j = i - 1, \quad i > 1; \\ 1 - r_i, & j = i + 1, \quad i < m; \\ r_i, & j = i = 1; \\ 1 - r_m, & j = i = m. \end{cases}$$

Заметим, что если вероятность дефекта изделия ε мала, то $r_i \approx \varepsilon n_i$. Поэтому для простоты далее будем полагать $r_i = \varepsilon n_i$.

Задача 1. На производстве используется система приемочного контроля с переменным планом, где $n_1 = 20$, $n_2 = 10$, $\varepsilon = 0,01$. Построить матрицу переходных вероятностей и найти стационарное распределение для номера плана.

Решение. Здесь $S = \{1, 2\}$. Вычисляем $r_1 = 0,2$; $r_2 = 0,1$. Матрица переходных вероятностей имеет вид

$$P = \begin{pmatrix} 0,2 & 0,8 \\ 0,1 & 0,9 \end{pmatrix}.$$

Решаем систему уравнений для вектора $\pi = (\pi_1, \pi_2)$:

$$\pi = \pi P, \quad \pi_1 + \pi_2 = 1,$$

или, более подробно,

$$\begin{cases} \pi_1 = 0,2\pi_1 + 0,1\pi_2; \\ \pi_2 = 0,8\pi_1 + 0,9\pi_2; \\ \pi_1 + \pi_2 = 1. \end{cases}$$

Решая первое (или второе) уравнение совместно с третьим, получаем $\pi = (1/9, 8/9)$.

Задача 2. В городе Ромашкино каждый год 1% жителей переселяются в пригород, а 4% жителей пригорода — в город. Найти стационарное распределение в предположении, что общая численность населения остается постоянной.

Решение. Будем считать, что живущий в городе находится в состоянии 1, а живущий в пригороде — в состоянии 2. Матрица переходных вероятностей имеет вид

$$P = \begin{pmatrix} 0,99 & 0,01 \\ 0,04 & 0,96 \end{pmatrix}.$$

Решая систему уравнений

$$\begin{cases} \pi_1 = 0,99\pi_1 + 0,04\pi_2; \\ \pi_2 = 0,01\pi_1 + 0,96\pi_2; \\ \pi_1 + \pi_2 = 1, \end{cases}$$

получаем $\pi = (0,8; 0,2)$. Таким образом, в стационарном режиме 80% живут в городе, 20% — в пригороде.

Задача 3. Магазин электротоваров торгует холодильниками. Случайный спрос на холодильники за неделю имеет распределение, заданное таблицей:

Спрос	0	1	2
P	0,2	0,5	0,3

Если холодильники заканчиваются, делается заказ на 2 штуки, который прибывает на следующей неделе. Построить матрицу переходных вероятностей. Найти стационарное распределение числа холодильников в магазине.

Решение. В данном случае будем нумеровать состояния от нуля (по числу холодильников), так что $S = \{0, 1, 2\}$. Матрица переходных вероятностей имеет вид

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0,8 & 0,2 & 0 \\ 0,3 & 0,5 & 0,2 \end{pmatrix}.$$

Предполагаем, что неудовлетворенный спрос пропадает и не переносится на следующую неделю (т.е. если он составлял 2 штуки, а в магазине был только один холодильник, то система переходит в состояние 0, без каких-либо последствий).

Решая систему уравнений

$$\begin{cases} \pi_0 = 0,8\pi_1 + 0,3\pi_2; \\ \pi_1 = 0,2\pi_1 + 0,5\pi_2; \\ \pi_2 = \pi_0 + 0,2\pi_2; \\ \pi_0 + \pi_1 + \pi_2 = 1, \end{cases}$$

получаем $\pi \approx (0,330; 0,258; 0,412)$. Следует отметить, что используемая в данном случае политика пополнения запасов не выглядит эффективной. Действительно, 33% времени товар отсутствует, и магазин упускает возможную прибыль от продаж.

9.3. Цепи Маркова с непрерывным временем. Системы массового обслуживания

В этом параграфе рассмотрим цепи Маркова:

а) с пространством состояний S из конечного или счетного числа элементов;

б) с непрерывным временем $t \geq 0$.

Обычно полагают, что элементы S занумерованы числами 1, 2, ...

Как и ранее, будем рассматривать только однородные цепи Маркова, свойства которых не зависят от времени.

Интенсивностью перехода из состояния i в состояние j ($i \neq j$) называется число λ_{ij} , такое, что вероятность перехода из состояния i в состояние j за промежуток времени Δt равна $\lambda_{ij}\Delta t + o(\Delta t)$ при $\Delta t \rightarrow 0$. Предполагается, что вероятность более чем одного перехода за это время составляет $o(\Delta t)$ при $\Delta t \rightarrow 0$.

Определим также формально величины

$$\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij},$$

имеющие смысл полных интенсивностей *выхода* из соответствующих состояний.

Тогда для вероятностей $p_i(t) = P(\xi(t) = i)$ имеет место система дифференциальных уравнений, называемых **уравнениями Колмогорова**:

$$\frac{dp_i(t)}{dt} = \sum_{k=1}^m \lambda_{ki} p_k(t).$$

Эти уравнения могут быть решены исходя из начального распределения $p(0)$.

Определим матрицу $\Lambda = (\lambda_{ij})$, называемую *матрицей интенсивностей переходов*, тогда уравнения переписутся в виде

$$p'(t) = p(t)\Lambda.$$

Как и ранее, нас будут интересовать эргодические цепи Маркова и предельные (стационарные) распределения. Поскольку в стационарном режиме распределение не меняется (и производные равны нулю), оно может быть найдено из системы уравнений

$$\pi\Lambda = 0, \quad \sum_{j=1}^m \pi_j = 1.$$

Если в цепи Маркова с непрерывным временем имеется только один класс сообщающихся состояний, а стационарное распределение существует и единственно, то такая цепь оказывается эргодической, и ее предельное распределение совпадает со стационарным.

Типичной областью применения цепей Маркова с непрерывным временем является *теория массового обслуживания*, занимающаяся изучением *систем массового обслуживания*, т.е. таких систем, в которых, с одной стороны, возникают массовые тре-

бования на выполнение каких-либо услуг, а с другой — происходит удовлетворение этих требований (по мере возможности).

Система массового обслуживания включает в себя источник требований (внешний или внутренний) и обслуживающие приборы (каналы обслуживания).

Эту терминологию не следует понимать буквально. Так, в качестве требований могут выступать люди, предметы, документы, файлы и пакеты данных и др.; в качестве обслуживающих приборов — люди, станки, организации, компьютеры и др.

Простейшим потоком событий называется такая последовательность событий (в непрерывном времени), когда интенсивность наступления очередного события постоянна. Иначе говоря, существует такое λ , что вероятность наступления события за промежуток времени Δt равна $\lambda \Delta t + o(\Delta t)$ при $\Delta t \rightarrow 0$. Предполагается, что вероятность наступления более чем одного события за это время составляет $o(\Delta t)$ при $\Delta t \rightarrow 0$. Отсюда получается, что время до наступления события распределено показательным с тем же параметром λ . Верно и обратное.

Таким образом, если требования поступают в систему через показательным распределенные промежутки времени и времена обслуживания требований также показательно распределены, причем параметры остаются постоянными во времени, то данную систему можно описать цепью Маркова¹. Далее будем полагать все случайные времена показательно распределенными по умолчанию.

Пример 2. Машина требует наладки в среднем один раз в θ единиц времени. Наладка занимает в среднем τ единиц времени. Требуется найти стационарное распределение вероятностей заставить машину в рабочем (1) и нерабочем (0) состоянии.

Перейдем к интенсивностям: $\lambda = 1/\theta$, $\mu = 1/\tau$. Переход из состояния 0 в состояние 1 (наладка) имеет интенсивность μ , а переход из состояния 1 в состояние 0 (поломка) — интенсивность λ . Матрица интенсивностей переходов имеет вид

$$A = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}.$$

Решая систему уравнений $\pi A = 0$, $\pi_0 + \pi_1 = 1$, получаем

¹ На практике эти условия выполняются далеко не всегда. Однако получаемые при указанных предположениях результаты можно использовать в качестве предварительных грубых оценок показателей работы системы.

$$\pi_0 = \frac{\lambda}{\lambda + \mu} = \frac{\tau}{\tau + \theta}, \quad \pi_1 = \frac{\mu}{\lambda + \mu} = \frac{\theta}{\tau + \theta}.$$

В общем случае, если выходят из строя или обслуживаются параллельно несколько объектов, соответствующие интенсивности суммируются.

Пример 3. В систему из n обслуживающих приборов поступает поток требований с интенсивностью λ . Среднее время обслуживания равно τ единиц времени. Если при поступлении требования все приборы заняты, оно становится в очередь.

Состояния будем нумеровать по числу требований, находящихся в системе (как в очереди, так и на обслуживании). Пространство состояний $S = \{0, 1, 2, \dots\}$ в данном случае бесконечно и счетно.

Введем интенсивность обслуживания $\mu = 1/\tau$ и величину $\rho = \lambda/\mu = \lambda\tau$, называемую *загрузкой* системы. Она описывает, сколько в среднем новых требований поступит за время обслуживания одного. Если $\rho \geq n$, то очередь требований растет до бесконечности, если же $\rho < n$, то система выходит на стационарный режим.

Интенсивности переходов имеют вид

$$\lambda_j = \begin{cases} \lambda, & j = i + 1; \\ \min(i, n)\mu, & j = i - 1, \quad i > 0. \end{cases}$$

Для данной системы рассчитаны, в частности, следующие показатели:

1) вероятность того, что все приборы свободны:

$$\pi_0 = \left[\sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!(1-\rho/n)} \right]^{-1};$$

2) вероятность того, что в системе находится ровно k требований:

$$\pi_k = \frac{\rho^k}{k!} \pi_0 \quad \text{при } 1 \leq k \leq n; \quad \pi_k = \frac{\rho^k}{n! n^{k-n}} \pi_0 \quad \text{при } k > n;$$

3) вероятность того, что все приборы заняты:

$$\pi_{\geq n} = \frac{\rho^n}{n!(1-\rho/n)} \pi_0;$$

4) средняя длина очереди:

$$L = \frac{\rho \pi_{\geq n}}{n - \rho} = \frac{\rho^{n+1}}{n! n (1 - \rho/n)^2} \pi_0.$$

Пример 4. Бригада из n рабочих обслуживает m станков ($n < m$). Каждый станок требует наладки в среднем один раз в θ единиц времени. Наладка занимает в среднем τ единиц времени. Если все рабочие заняты, наладка станка откладывается (ставится в очередь).

Состояния будем нумеровать по числу требований (т.е. требующих наладки станков). Пространство состояний $S = \{0, 1, 2, \dots, m\}$ в данном случае конечно.

Как и ранее, положим $\lambda = 1/\theta$, $\mu = 1/\tau$, $\rho = \lambda/\mu = \tau/\theta$.

Интенсивности переходов имеют вид:

$$\lambda_{ij} = \begin{cases} (m-i)\lambda, & j=i+1, \quad i < m; \\ \min(i, n)\mu, & j=i-1, \quad i > 0. \end{cases}$$

Для данной системы рассчитаны, в частности, следующие показатели:

1) вероятность того, что все станки исправны:

$$\pi_0 = \left[\sum_{k=0}^n \frac{m!}{k!(m-k)!} \rho^k + \sum_{k=n+1}^m \frac{m!}{n^{k-n} n!(m-k)!} \rho^k \right]^{-1};$$

2) вероятность того, что в системе находится ровно k требований:

$$\pi_k = \frac{m!}{k!(m-k)!} \rho^k \pi_0 \quad \text{при } 1 \leq k \leq n;$$

$$\pi_k = \frac{m!}{n^{k-n} n!(m-k)!} \rho^k \pi_0 \quad \text{при } k > n;$$

3) средняя длина очереди:

$$L = \sum_{k=n+1}^m (k-n) \pi_k = \sum_{k=n+1}^m \frac{(k-n)m!}{n^{k-n} n!(m-k)!} \rho^k \pi_0.$$

Задачи для самостоятельного решения

1. На производстве используется система приемочного контроля с переменным планом, где $n_1 = 20$, $n_2 = 10$, $n_3 = 5$. Найти стационарное распределение, если: а) $\varepsilon = 0,01$; б) $\varepsilon = 0,02$.
2. Трудоспособное население (постоянной численности) делится на работающих и безработных. Вероятность потерять работу в течение месяца составляет 2%, а вероятность ее найти — 16%. Найти стационарный уровень безработицы. Во сколько раз сократится доля безработных, если благодаря государственной

программе занятости вероятность найти работу в течение месяца увеличится вдвое?

3. Рабочие города Молотково трудятся на трех заводах. Вероятности ухода с каждого завода в течение месяца составляют 1, 6 и 9% соответственно. В случае ухода с одного завода рабочий переходит на один из двух других равновероятно. Найти стационарное распределение. Общая численность рабочих предполагается постоянной.
4. Решить уже рассмотренную задачу о холодильниках в предположении, что вводится дополнительное правило: если в магазине остается один холодильник, делается заказ на еще один.
5. Мебельный магазин торгует шкафами. Случайный спрос на шкафы за неделю имеет распределение, заданное таблицей.

Спрос	0	1	2
P	0,3	0,6	0,1

Если шкафы заканчиваются, делается заказ на 2 штуки, если остается один — на 1 штуку. Заказы прибывают на следующей неделе. Найти стационарное распределение числа шкафов в магазине.

6. В городе Традицию каждый взрослый мужчина имеет одну из трех профессий А, Б и В. Сыновья сохраняют профессии отцов с вероятностями $3/5$, $2/3$ и $1/4$ соответственно или выбирают любую из двух других равновероятно. Найти: а) распределение по профессиям в следующем поколении, если в нынешнем профессию А имело 20%, Б — 30%, В — 50%; б) стационарное распределение по профессиям. Предполагается, что у каждого отца ровно один сын.
7. Торговец из города Олино ездит продавать товар в города Алино и Валино. Выехав из Олино, он направляется в Алино с вероятностью 40%, в Валино — с вероятностью 60%. Продав товар, он возвращается в Олино. Найти стационарное распределение вероятностей застать предпринимателя в каждом из городов.
8. Решить предыдущую задачу в предположении, что, посетив Алино или Валино, торговец возвращается в Олино с вероятностью 80% или едет в другой из двух городов (продавать оставшийся товар) с вероятностью 20%.
9. Фирма оценивает недельный объем сбыта как удовлетворительный (1), хороший (2) или отличный (3). Матрица переходных вероятностей имеет вид:

$$а) \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,3 & 0,6 & 0,1 \\ 0,1 & 0,7 & 0,2 \end{pmatrix}; \quad б) \begin{pmatrix} 0,4 & 0,5 & 0,1 \\ 0,1 & 0,7 & 0,2 \\ 0,6 & 0,2 & 0,2 \end{pmatrix}.$$

Найти стационарное распределение объема сбыта фирмы.

10. Продуктивность фермерского хозяйства в каждом году оценивается как хорошая (1), удовлетворительная (2) или плохая (3). Матрица переходных вероятностей имеет вид:

$$а) \begin{pmatrix} 0,3 & 0,6 & 0,1 \\ 0,1 & 0,6 & 0,3 \\ 0,1 & 0,4 & 0,5 \end{pmatrix}; \quad б) \begin{pmatrix} 0,3 & 0,5 & 0,2 \\ 0,2 & 0,6 & 0,2 \\ 0,1 & 0,5 & 0,4 \end{pmatrix}.$$

Найти стационарное распределение продуктивности хозяйства.

11. В процессе ежемесячного погашения кредита клиент банка может оказаться в одном из следующих состояний:
- 1) ежемесячный платеж погашен в срок в полном объеме;
 - 2) ежемесячный платеж погашен в полном объеме, но с временной задержкой;
 - 3) ежемесячный платеж погашен в срок в неполном объеме, с переходом остатка долга на следующий месяц;
 - 4) ежемесячный платеж погашен с временной задержкой и в неполном объеме, с переходом остатка долга на следующий месяц.
- Матрица переходных вероятностей имеет вид:

$$а) \begin{pmatrix} 0,3 & 0,4 & 0,1 & 0,2 \\ 0 & 0,2 & 0,5 & 0,3 \\ 0 & 0 & 0,4 & 0,6 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad б) \begin{pmatrix} 0,5 & 0,3 & 0,2 & 0 \\ 0 & 0,4 & 0,4 & 0,2 \\ 0 & 0 & 0,3 & 0,7 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Вычислить вероятности состояний через три месяца, если в начале клиент находился в состоянии а).

12. Машина требует наладки в среднем один раз в 2 часа. Наладка занимает в среднем 15 минут. Найти стационарное распределение вероятностей застать машину в рабочем (1) и нерабочем (0) состоянии.
13. Имеется 2 станка, каждый из которых требует наладки в среднем один раз в 2 часа. Наладка занимает в среднем 10 минут. Найти стационарное распределение числа работающих станков. Предполагается, что станки требуют наладки и обслуживаются независимо один от другого.
14. Машина состоит из трех узлов, каждый из которых отказывает с интенсивностью $\lambda = 2$. Интенсивность восстановления отказавшего узла равна $\mu = 3$. Найти стационарное распределение числа работающих узлов.
15. Машина состоит из трех узлов. Среднее время безотказной работы каждого узла составляет 20 часов, а среднее время ремонта узла — 5 часов. Найти среднюю производительность

машины, если при трех работающих узлах она равна 100%, при двух — 50%, а при одном или менее машина вообще не работает.

16. В ремонтной мастерской трудятся 5 мастеров. В течение дня на ремонт поступает в среднем 10 изделий, а каждый мастер успевает отремонтировать в среднем 2,5 изделия. Предполагая, что мастерская работает в стационарном режиме, найти: а) вероятность того, что все мастера свободны; б) вероятность того, что все мастера заняты; в) среднюю длину очереди; г) среднее число мастеров, свободных от работы.
17. В ремонтную мастерскую, где работают 3 мастера, поступает в среднем 4 заказа в час. Среднее время выполнения заказа составляет полчаса. Определить среднее число заказов, ожидающих начала выполнения.
18. Рабочий обслуживает группу из 3 автоматов. Каждый автомат требует обслуживания в среднем раз в полчаса. Обслуживание занимает в среднем 12 минут. Найти: а) стационарное распределение числа неисправных автоматов; б) среднее число автоматов, ожидающих обслуживания; в) среднее число простаивающих автоматов.
19. Двое рабочих обслуживают машину, состоящую из 4 блоков. Каждый блок требует обслуживания в среднем один раз в 2 часа. Обслуживание занимает в среднем 15 минут. Найти: а) стационарное распределение числа работающих блоков; б) среднюю длину очереди; в) среднее число занятых рабочих.
20. В кассу обращаются клиенты, в среднем по одному за 10 минут. Сколько в среднем времени должно занимать обслуживание одного клиента, чтобы средняя длина очереди в стационарном режиме не превышала: а) 2 человека; б) 5 человек?
21. В конторе три сотрудника принимают посетителей в порядке общей очереди. В среднем обращается по 5 человек в час, обслуживание каждого занимает в среднем 18 минут. Найти: а) вероятность того, что все сотрудники свободны; б) вероятность того, что все сотрудники заняты; в) среднюю длину очереди.
21. На железнодорожную сортировочную станцию поступает в среднем 2 состава в час. Обслуживание (расформирование) состава занимает в среднем 20 минут. В парке прибытия станции есть два пути, на которых производится обслуживание; если оба пути заняты, составы ожидают на внешних путях. Найти: а) вероятность того, что все пути свободны; б) среднюю длину очереди.

Часть II

**МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА**

Глава 9

ЦЕПИ МАРКОВА

9.1. Основные определения

Случайным процессом называется функция двух переменных $\xi(t, \omega)$, где $t \in T$ — время, $\omega \in \Omega$ — элементарный исход. Время может быть непрерывным или дискретным (целочисленным).

Случайный процесс ставит в соответствие каждому моменту времени $t \in T$ случайную величину $\xi(t) = \xi(t, \omega)$ как функцию от $\omega \in \Omega$. Поэтому процесс обычно для краткости обозначают $\xi(t)$.

Пусть заданы моменты времени $0 \leq t_1 < t_2 < \dots < t_n < t_{n+1} < \dots < t_{n+m}$, $n, m > 0$, и A — любое событие (утверждение), относящееся к случайным величинам $\xi(t_1), \dots, \xi(t_{n-1})$, B — относящееся к $\xi(t_{n+1}), \dots, \xi(t_{n+m})$, C — относящееся к $\xi(t_n)$. Говорят, что процесс обладает *марковским свойством*, если для него всегда выполняется равенство

$$P(B|AC) = P(B|C),$$

или (в эквивалентной форме)

$$P(AB|C) = P(A|C)P(B|C).$$

Марковское свойство означает, что будущее поведение процесса не зависит от его прошлого при условии, что известно настоящее (т.е. текущее значение процесса).

Марковским процессом называется случайный процесс, обладающий марковским свойством.

Пространством состояний S случайного процесса называется множество всех возможных значений функции $\xi(t, \omega)$.

Цепью Маркова называется марковский процесс, для которого выполнено хотя бы одно из следующих двух условий:

- а) пространство состояний конечно или счетно;

стве (согласно аксиоматике А.Н. Колмогорова). Числовые характеристики введенной формально случайной величины отражают важные свойства совокупности исследуемых объектов. В частности, ее математическое ожидание $M\xi$ равно среднему значению признака, а функция распределения $F_\xi(x) = P(\xi < x)$ показывает долю объектов, для которых значение признака меньше x .

Например, в социально-экономических исследованиях нас могут интересовать средний доход на душу населения, доля людей с доходами меньше прожиточного минимума и т.п.

Распределение ξ часто называют **распределением генеральной совокупности**.

Будем последовательно извлекать из генеральной совокупности ее элементы, выбирая их случайным образом (наудачу), измерять и записывать значения некоторого признака для них: $x_1, x_2, \dots, x_n$. Эти значения называются **наблюдениями** (признака), их набор — **выборкой**, а число сделанных наблюдений — **объемом выборки** n . Понятно, что наблюдения представляют собой случайные величины, в силу случайности нашего выбора объекта. Они заданы уже на другом вероятностном пространстве — на множестве всех вариантов выбора n элементов из генеральной совокупности. Полученные данные называют **наблюдениями случайной величины** ξ , а также говорят, что случайная величина ξ «принимает значения» $x_1, x_2, \dots, x_n$.

Основная задача математической статистики — сделать научно обоснованные выводы о распределении одной или более неизвестных случайных величин или их взаимосвязи между собой.

Выборочным методом называется метод решения этой задачи посредством анализа выборки, полученной в результате многократных наблюдений.

Для того чтобы характеристики случайной величины, полученные выборочным методом, были объективны, необходимо, чтобы выборка была **репрезентативной**, т.е. достаточно хорошо представляла исследуемую величину. В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществить случайно, т.е. все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку. Для этого существуют различные виды отбора выборки.

1. *Простым* случайным отбором называется отбор, при котором объекты извлекаются по одному из всей генеральной совокупности (случайным образом).

2. *Механическим* называется такой, при котором генеральную совокупность делят на столько частей, сколько объектов должно войти в выборку, и из каждой группы случайным образом отбирают один объект.

3. *Серийным* называется отбор, при котором объекты из генеральной совокупности отбираются «сериями», которые подвергаются сплошному обследованию.

4. *Стратифицированный (расслоенный)* отбор заключается в том, что исходная генеральная совокупность объема N подразделяется на подсовокупности (страты) $N_1, N_2, \dots, N_k$, так что $N_1 + N_2 + \dots + N_k = N$. Когда страты определены, из каждой из них извлекается простая случайная выборка объема $n_1, n_2, \dots, n_k$. Частным случаем стратифицированного отбора является *типический отбор*, при котором объекты отбирают не из всей генеральной совокупности, а из каждой типической ее части.

Комбинированный отбор сочетает в себе сразу несколько видов отбора, образующих различные фазы выборочного обследования. Существуют и другие методы организации выборки.

Выборка называется **повторной**, если отобранный объект перед выбором следующего возвращается в генеральную совокупность. **Выборка** называется **бесповторной**, если отобранный объект в генеральную совокупность не возвращается. Для конечной генеральной совокупности случайный отбор без возвращения на каждом шаге приводит к зависимости отдельных наблюдений, случайный равновозможный выбор с возвращением — к независимости наблюдений.

На практике обычно имеют дело с бесповторными выборками. Тем не менее, когда объем генеральной совокупности N во много раз больше, чем объем выборки n (например, в сотни или тысячи раз), зависимостью наблюдений можно пренебречь.

Таким образом, в математической статистике наблюдения $x_1, x_2, \dots, x_n$ моделируются независимыми случайными величинами с одинаковым распределением — тем же, что и у исходной случайной величины ξ .

Иногда рассматриваются по-разному распределенные наблюдения (например, в случае неравноточных измерений), а иногда — и зависимые наблюдения, но все эти случаи оговариваются особо.

Бывают ситуации, когда затруднительно или невозможно описать объекты, чьи признаки мы наблюдаем. Речь идет об измерении какой-либо величины, которая меняет свое значение случайным образом от одного наблюдения к другому. Это могут быть, например, колебания в ценах акций, курсах валют, случайные ошибки измерения и т.п. В качестве «объекта» выступает некоторое стечение обстоятельств, которые в принципе могли сложиться различным образом. В этих случаях тем не менее все равно применяют изложенную выше статистическую модель, хотя «выбор» здесь осуществляется без личного участия исследователя, какими-то внешними силами.

Как уже было отмечено, случайная величина ξ имеет определенную функцию распределения $F_{\xi}(x)$ и другие числовые характеристики, которые будем называть **теоретическими**, в отличие от **выборочных**, которые определяются по наблюдениям.

Ряд наблюдений, упорядоченных по возрастанию, называют **вариационным рядом**. Его члены обычно обозначают $x_{(1)}, x_{(2)}, \dots, x_{(n)}$. Наименьшее и наибольшее значения (минимум и максимум) обозначают $x_{\min}$ и $x_{\max}$ и называют их **крайними членами вариационного ряда**. Различные значения признака, появившиеся в процессе наблюдения, называют **вариантами**. Варианты, как и наблюдения, обозначаются через $x_1, \dots, x_n$.

Когда мы наблюдаем дискретную случайную величину, она может принимать одни и те же значения по много раз. Поэтому для экономии места и времени каждое значение записывают только один раз, с указанием, сколько всего раз оно появилось. Число n_i , показывающее, сколько раз появилось значение x_i в n наблюдениях, называют **частотой** данного значения, а отношение $w_i = n_i/n$ — **относительной частотой**. Число k различных значений в n наблюдениях всегда конечно, и $k \leq n$. Очевидно, имеют место равенства $\sum_{i=1}^k n_i = n$ и $\sum_{i=1}^k w_i = 1$. Результаты можно записать в таблицу:

x_1	x_2	...	x_k
w_1	w_2	...	w_k

В случае непрерывной случайной величины на практике часто применяют **группировку**. Это означает, что весь интервал наблюдаемых значений разбивают на k частичных интервалов $[c_0, c_1), [c_1, c_2), \dots, [c_{k-1}, c_k]$ равной длины h и затем подсчитыва-

ются числа попаданий наблюдений в эти интервалы, которые принимают за частоты n_i (для некоторой новой, уже дискретной случайной величины). В качестве новых значений вариант x_i обычно берут середины интервалов (либо в таблице указывают сами интервалы). Группировка может применяться и в случае дискретных случайных величин, если шаг, с которым меняются их значения, кажется нам слишком мелким.

Согласно формуле Стерджеса рекомендуемое число интервалов разбиения $k \approx 1 + \log_2 n$, а длины частичных интервалов $h = (x_{\max} - x_{\min})/k$. Предполагается, что весь интервал имеет вид $[x_{\min}, x_{\max}]$.

Понятно, что группировка связана с потерей части полезной информации, заключенной в выборке. Однако она имеет и свои преимущества. Оценим величину экономии, например, для $n = 10^6$ наблюдений. Рекомендуемое число интервалов $k = 21$, и от нас требуется сохранить и обработать лишь $2k = 42$ числа вместо миллиона!

10.2. Графическое представление статистических рядов

Набор вариант x_i (или частичных интервалов) и их относительных частот w_i называют **статистическим рядом**. Графически статистические ряды могут быть представлены в виде полигона, гистограммы или графика накопленных частот.

Полигоном частот называют ломаную линию, отрезки которой соединяют точки (x_1, n_1) , (x_2, n_2) , ..., (x_k, n_k) . Полигоном относительных частот называют ломаную, отрезки которой соединяют точки (x_1, w_1) , (x_2, w_2) , ..., (x_k, w_k) . Полигоны обычно служат для изображения выборки в случае дискретных случайных величин (рис. 10.1).

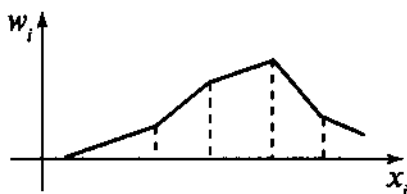


Рис. 10.1

Задача 1. Построить полигон частот по данному распределению выборки:

x_i	2	3	5	6
n_i	10	15	5	20



Рис. 10.2

Решение. Отложим на оси абсцисс (рис. 10.2) варианты x_p , а на оси ординат — соответствующие им частоты n_p , затем соединим последовательно точки (x_p, n_p) .

Гистограммой относительных частот¹ (или просто гистограммой) называется ступенчатая фигура, состоящая из прямоугольников, основанием которых служат частичные интервалы длиной h , а высоты равны w_i/h . Гистограмма обычно служит для изображения выборки в случае непрерывных случайных величин. Площадь гистограммы равна единице (рис. 10.3).

Поэтому гистограмму можно рассматривать как график эмпирической (выборочной) плотности распределения $p_n(x)$. Если у теоретического распределения F существует конечная плотность, то эмпирическая плотность является некоторым приближением для теоретической. В этом и состоит практическая польза гистограммы.

При построении гистограмм в реальных исследованиях следует понимать, что формула Стерджеса (как и любая другая) для числа интервалов разбиения k дает лишь рекомендацию, а не строгое правило. Проблема выбора этого числа заключается

¹ На практике гистограммами также называют ступенчатые фигуры с высотами w_i (без деления на h), как на рис. 10.3.

в следующем. При слишком малых k гистограмма получается слишком грубой, «смазанной», плохо отражающей свойства распределения. При слишком больших k гистограмма становится «колючей» и в конце концов распадается на отдельные «иглы» (узкие столбцы) попеременно с пустыми интервалами. Оптимальное значение в общем случае неизвестно — оно зависит как от типа распределения, так и от конкретной выборки. Что касается концов интервалов и значений вариантов, то для человеческого восприятия удобнее, чтобы они выражались более или менее «круглыми» числами.

Поскольку гистограммы теперь обычно строят не вручную, а на компьютере, исследователь легко может варьировать параметры гистограммы (нижнюю и верхнюю границы интервала, число частичных) и в конечном счете выбрать тот вариант, при котором, по его мнению, график выглядит лучше всего.

Графиком накопленных частот называется фигура, строящаяся аналогично гистограмме с той разницей, что для расчета высот прямоугольников берутся не простые, а **накопленные относительные частоты**, т.е. величины $w_i^C = \sum_{j=1}^i w_j$. Эти величины не убывают, и таким образом, график накопленных частот имеет вид ступенчатой «лестницы» (от 0 до 1).

График накопленных частот и эмпирическая функция распределения на практике используются для приближения теоретической функции распределения (рис. 10.4).

Задача 2. Анализируется выборка из 100 малых предприятий региона. Цель обследования — измерение коэффициента соотношения заемных и собственных средств (x_i) на каждом i -м предприятии. Результаты представлены в табл. 10.1.

Таблица 10.1

**Коэффициенты соотношений заемных и собственных
средств предприятий**

5,56	5,45	5,48	5,45	5,39	5,37	5,46	5,59	5,61	5,31
5,46	5,61	5,11	5,41	5,31	5,57	5,33	5,11	5,54	5,43
5,34	5,53	5,46	5,41	5,48	5,39	5,11	5,42	5,48	5,49
5,36	5,40	5,45	5,49	5,68	5,51	5,50	5,68	5,21	5,38

Окончание табл. 10.1

5,58	5,47	5,46	5,19	5,60	5,63	5,48	5,27	5,22	5,37
5,33	5,49	5,50	5,54	5,40	5,58	5,42	5,29	5,05	5,79
5,79	5,65	5,70	5,71	5,85	5,44	5,47	5,48	5,47	5,55
5,67	5,71	5,73	5,05	5,35	5,72	5,49	5,61	5,57	5,69
5,54	5,39	5,32	5,21	5,73	5,59	5,38	5,25	5,26	5,81
5,27	5,64	5,20	5,23	5,33	5,37	5,24	5,55	5,60	5,51

Требуется построить гистограмму и график накопленных частот.

Решение. Построим группированный ряд наблюдений.

1. Определим в выборке $x_{\min} = 5,05$ и $x_{\max} = 5,85$.

2. Разобьем весь диапазон $[x_{\min}, x_{\max}]$ на k равных интервалов: $k \approx 1 + \log_2 100 = 7,62$; $k \approx 8$, откуда длина интервала

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{5,85 - 5,05}{8} = 0,1.$$

Таблица 10.2

Сгруппированный ряд наблюдений

Номер интервала	Интервал	Середина интервала x_i	w_i	w_i^c	$f_n(x)$
1	5,05—5,15	5,1	0,05	0,05	0,5
2	5,15—5,25	5,2	0,08	0,13	0,8
3	5,25—5,35	5,3	0,12	0,25	1,2
4	5,35—5,45	5,4	0,20	0,45	2,0
5	5,45—5,55	5,5	0,26	0,71	2,6
6	5,55—5,65	5,6	0,15	0,86	1,5
7	5,65—5,75	5,7	0,10	0,96	1,0
8	5,75—5,85	5,8	0,04	1,00	0,4

На рис. 10.3 и 10.4, построенных по данным табл. 10.1 с помощью статистического пакета STATISTICA, представлены гистограмма и график накопленных частот. Кривые соответствуют плотности и функции нормального распределения, «подобранного» к данным.


Рис. 10.4

10.3. Эмпирические законы распределения

Эмпирической функцией распределения (или функцией распределения выборки) называют функцию $F_n(x)$, определяющую для каждого числа x относительную частоту события $\xi < x$, т.е. $F_n(x) = n_x/n$, где n_x — число наблюдений, меньших x , n — объем выборки.

Иначе говоря, эмпирической функцией распределения называют функцию, определяющую для каждого числа x долю из n наблюдений, меньших x .

Из определения следует, что значения эмпирической функции распределения при каждом x являются случайными величинами. В отличие от эмпирической функции распределения, функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*.

Эмпирическая функция распределения обладает всеми свойствами обычной функции распределения, а также некоторыми специфическими:

1. $0 \leq F_n(x) \leq 1$.
2. $F_n(x)$ — неубывающая функция.
3. $F_n(x)$ непрерывна слева.
4. $F_n(x) = 0$ при $x \leq x_{\min}$ и $F_n(x) = 1$ при $x > x_{\max}$.
5. $P(\limsup_{n \rightarrow \infty} |F_n(x) - F(x)| = 0) = 1$ (теорема Гливленко—Кантелли).

Доказательство теоремы Гливленко—Кантелли является довольно сложным, поэтому докажем следующий ее упрощенный вариант.

► **Теорема 1.** При любом $\varepsilon > 0$ верно $\lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| < \varepsilon) = 1$ для любого x .

Доказательство. По определению $F_n(x) = \frac{n_x}{n}$, где n_x — число наблюдений, меньших x . Рассмотрим наблюдения как n независимых испытаний Бернулли, в каждом из которых возможны два исхода: $\{x_k < x\}$ или $\{x_k \geq x\}$. Вероятности этих событий равны $p = P(\xi < x) = F(x)$ и $q = P(\xi \geq x) = 1 - F(x)$ соответственно. Событие $\{x_k < x\}$ можем называть успехом, тогда n_x — число успехов в n независимых испытаниях Бернулли. Следовательно, математическое ожидание $Mn_x = np$ и дисперсия $Dn_x = npq$. Отсюда

$$MF_n(x) = \frac{1}{n} Mn_x = \frac{nF(x)}{n} = F(x);$$

$$DF_n(x) = \frac{1}{n^2} Dn_x = \frac{nF(x)(1-F(x))}{n^2} = \frac{F(x)(1-F(x))}{n}.$$

В силу неравенства Чебышева для любого $\varepsilon > 0$ верна оценка

$$P(|F_n(x) - F(x)| \geq \varepsilon) \leq \frac{DF_n(x)}{\varepsilon^2},$$

поэтому

$$P(|F_n(x) - F(x)| \geq \varepsilon) \leq \frac{F(x)(1-F(x))}{n\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty,$$

откуда следует утверждение теоремы.

Смысл теоремы Гливенко—Кантелли заключается в том, что при увеличении объема выборки n у эмпирической функции распределения исчезают свойства случайности, и она приближается к теоретической функции распределения. Эмпирическая функция распределения служит оценкой функции распределения генеральной совокупности.

График эмпирической функции распределения есть неубывающая ступенчатая кривая со скачками, равными $1/n$ в точках вариационного ряда (если значения не совпадают). Если n_i точек вариационного ряда совпадают и равны x_i , то скачок в точке x_i равен n_i/n .

Задача 3. Задана таблица наблюдений случайной величины ξ :

x_i	2	3	5
n_i	75	20	5
w_i	0,75	0,2	0,05

Эмпирическая функция распределения имеет вид:

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq 2; \\ 0,75, & \text{если } 2 < x \leq 3; \\ 0,95, & \text{если } 3 < x \leq 5; \\ 1, & \text{если } x > 5. \end{cases}$$

Ее график представлен на рис. 10.5.

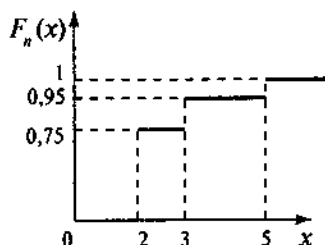


Рис. 10.5

Для непрерывной случайной величины лучшую оценку плотности распределения $p(x)$ дает полигон относительных частот. Таким образом, закон распределения выборки является некоторым приближением закона распределения генеральной совокупности.

Задача 4. Пусть $x_1, x_2, \dots, x_n$ — выборка независимых наблюдений из непрерывной генеральной совокупности с функцией распределения $F(x)$ и плотностью распределения $p(x)$. Найти функции распределения и плотности распределения крайних членов вариационного ряда: $x_{\min}$ и $x_{\max}$.

Решение. Из определения функции распределения следует, что

$$F_{x_{\max}}(y) = P(x_{\max} < y) = P(x_1 < y, x_2 < y, \dots, x_n < y) = P^n(x_i < y) = F_{\xi}^n(y).$$

$$\text{Тогда } p_{x_{\max}}(y) = [F_{x_{\max}}(y)]' = [F_{\xi}^n(y)]' = nF_{\xi}^{n-1}(y) \cdot F_{\xi}'(y) = nF_{\xi}^{n-1}(y) \cdot p_{\xi}(y).$$

Аналогично

$$F_{x_{\min}}(y) = P(x_{\min} < y) = 1 - P(x_{\min} \geq y) = 1 - P(x_1 \geq y, x_2 \geq y, \dots, x_n \geq y) = 1 - [1 - F_{\xi}(y)]^n.$$

Отсюда получаем функцию плотности

$$p_{x_{\min}}(y) = -n[1 - F_{\xi}(y)]^{n-1} \cdot (-F_{\xi}'(y)) = n[1 - F_{\xi}(y)]^{n-1} \cdot p_{\xi}(y).$$

Задача 5. Построить гистограмму частот по данному распределению выборки:

Номер интервала i	Частичный интервал	Число наблюдений, попавших в интервал, n_i
1	2—7	5
2	7—12	10
3	12—17	25
4	17—22	6
5	22—27	4

Решение. Найдём сначала плотности частот, т.е. величины n_i/h . Для данного примера $h = 5$.

Таким образом, получаем:

Номер интервала i	Плотность частоты n_i/h
1	1
2	2
3	5
4	1,2
5	0,8

Отложим на оси абсцисс (рис. 10.6) интервалы длиной $h = 5$ каждый, а затем проведем над ними отрезки, параллельные оси x , на расстояниях от нее, равных соответствующим значениям плотности частоты (ось ординат).

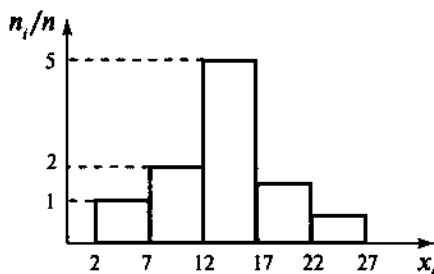


Рис. 10.6

Задача 6. Построить гистограмму относительных частот по заданному распределению выборки:

Номер интервала i	Частичный интервал	Число наблюдений, попавших в интервал, n_i
1	10–15	2
2	15–20	4
3	20–25	8
4	25–30	4
5	30–35	2

Решение. Найдем относительные частоты и плотности относительных частот:

Частота w_i	Плотность относительных частот w_i/h
$w_1 = n_1/n = 2/20 = 0,1$	$w_1/h = 0,1/5 = 0,02$
$w_2 = n_2/n = 4/20 = 0,2$	$w_2/h = 0,2/5 = 0,04$
$w_3 = n_3/n = 8/20 = 0,4$	$w_3/h = 0,4/5 = 0,08$
$w_4 = n_4/n = 4/20 = 0,2$	$w_4/h = 0,2/5 = 0,04$
$w_5 = n_5/n = 2/20 = 0,1$	$w_5/h = 0,1/5 = 0,02$

Построим на оси абсцисс (рис. 10.6) частичные интервалы $h = 5$, затем проведем параллельно им отрезки, отстоящие от оси x на соответствующие значения плотности относительной частоты.

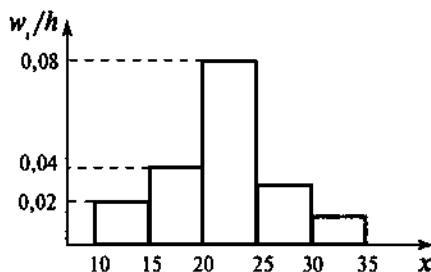


Рис. 10.7

Задачи для самостоятельного решения

- Сколько частичных интервалов можно порекомендовать в случае объема выборки:
а) $n = 200$; б) $n = 500$?

- Построить полигон частот по следующему распределению выборки:

x_i	15	20	25	30	35
n_i	10	15	30	20	25

- В таблице представлены данные по числу сделок на фондовой бирже за квартал для 400 инвесторов.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

Построить полигон относительных частот.

4. Построить гистограмму частот по следующему распределению выборки.

Номер интервала i	Частичный интервал	Число наблюдений, появившихся в интервал, n_i
1	3—5	4
2	5—7	6
3	7—9	20
4	9—11	40
5	11—13	20
6	13—15	4
7	15—17	6

5. Построить гистограмму относительных частот по следующему распределению выборки:

Номер интервала i	Частичный интервал	Число наблюдений, появившихся в интервал, n_i
1	2—5	6
2	5—8	10
3	8—11	4
4	11—14	5

6. Построить полигон относительных частот по следующему распределению выборки:

x_i	2	4	5	7	10
w_i	0,15	0,2	0,1	0,1	0,45

7. Для изучения распределения заработной платы работников определенной отрасли обследовано 100 человек. Результаты представлены в следующей таблице.

Заработная плата в долларах США	Число человек	Заработная плата в долларах США	Число человек
190—192	1	200—202	19
192—194	5	202—204	11
194—196	9	204—206	4
196—198	22	206—208	1
198—200	28	208—210	0

Построить гистограмму и график накопленных частот.

8. В ОТК были измерены диаметры 300 валков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала (мм) даны в таблице:

Границы отклонений	Середина интервала	Число валков	Границы отклонений	Середина интервала	Число валков
-30...-25	-27,5	3	0-5	2,5	55
-25...-20	-22,5	8	5-10	7,5	30
-20...-15	-17,5	15	10-15	12,5	25
-15...-10	-12,5	35	15-20	17,5	14
-10...-5	-7,5	40	20-25	22,5	8
-5...0	-2,5	60	25-30	27,5	7

Построить гистограмму и график накопленных частот.

9. В таблице представлены данные о месячном доходе жителя региона (руб.) по выборке из 1000 жителей:

x_i	Менее 500	500-1000	1000-1500	1500-2000	2000-2500	Свыше 2500
n_i	58	96	239	328	147	132

Построить гистограмму и график накопленных частот.

Указание. В качестве верхней границы последнего интервала использовать 3000.

10. В таблице представлены данные об удое 100 коров на молочной фирме за лактационный период (ц):

x_i	4-6	6-8	8-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26
n_i	1	3	6	11	15	20	14	12	10	6	2

Построить гистограмму и график накопленных частот.

11. Проведено исследование посещаемости популярного интернет-сайта. Много часов подряд регистрируется число посетителей, посетивших сайт в течение данного часа. Результаты исследования представлены в таблице:

Число посетителей	Время, ч	Число посетителей	Время, ч
0	57	7	139
1	203	8	45
2	383	9	27
3	525	10	10
4	532	11	4
5	408	12	1
6	273	14	1

Построить полигон относительных частот.

12. Проведено исследование посещаемости популярного интернет-сайта. Много часов подряд регистрируется число посетителей, посетивших сайт в течение данного часа. Результаты исследования представлены в таблице:

Число посетителей	Время, ч	Число посетителей	Время, ч
0	12	7	103
1	108	8	24
2	316	9	13
3	551	10	2
4	632	11	0
5	492	12	0
6	273	14	0

Построить полигон относительных частот.

13. Построить гистограмму и график накопленных частот по данному распределению выборки.

Номер интервала i	Частичный интервал	Число наблюдений, попавших в интервал, n_i
1	2—7	5
2	7—12	10
3	12—17	25
4	17—22	7
5	22—27	5

14. Результаты исследования прочности 200 образцов на сжатие представлены в виде интервального статистического ряда в таблице:

Номер интервала i	Частичный интервал	Число наблюдений, попавших в интервал, n_i
1	190—200	10
2	200—210	26
3	210—220	58
4	220—230	64
5	230—240	30
6	240—250	14

Построить гистограмму и график накопленных частот.

15. Построить гистограмму, полигон относительных частот и график накопленных частот по заданному распределению выборки:

Номер интервала i	Частичный интервал	Число наблюдений, попавших в интервал, n_i
1	10—15	2
2	15—20	4
3	20—25	8
4	25—30	5
5	30—35	3

16. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение 5 поездок фиксировал свое время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 (мин.). Найти эмпирическую функцию распределения и построить ее график.
17. При измерении веса 10 шоколадных батончиков (г) получены значения: 49,1; 50,0; 49,7; 50,5; 48,1; 50,3; 49,7; 51,6; 48,1; 50,1. Найти эмпирическую функцию распределения и построить ее график.

Глава 11

ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

11.1. Выборочные характеристики и точечные оценки

Выборочными характеристиками называются функции от наблюдений, приближенно оценивающие соответствующие *числовые характеристики* случайной величины. В случае равноточных измерений в качестве оценок математического ожидания, дисперсии, начальных и центральных моментов используются следующие выборочные характеристики:

1) выборочное среднее — $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$;

2) выборочные дисперсии — $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$;

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2;$$

3) выборочные начальные моменты k -го порядка —

$$\hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^n x_i^k;$$

4) выборочные центральные моменты k -го порядка —

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k.$$

В качестве других используемых на практике выборочных характеристик можно назвать **выборочную моду** X_{mod} , равную значению варианты с наибольшей частотой, и **выборочную медиану** X_{med} , равную значению, стоящему в середине вариационного ряда (либо полусумме двух значений, с номерами k и $k + 1$, при четном числе наблюдений $n = 2k$).

Иногда рассматриваются такие числовые характеристики распределения, как коэффициент вариации $V = \sigma/M\xi$, асимметрия $\beta = \mu_3/\sigma^3$ и эксцесс $\gamma = \mu_4/\sigma^4 - 3$. Им также соответствуют выборочные характеристики: выборочный коэффициент вариации $\hat{V} = s/\bar{x}$, выборочная асимметрия $\hat{\beta} = \hat{\mu}_3/s^3$, выборочный эксцесс $\hat{\gamma} = \hat{\mu}_4/s^4 - 3$.

Все эти характеристики не совпадают с соответствующими характеристиками генеральной совокупности, поскольку являются случайными величинами. Распределение указанных случайных величин однозначно определяется распределением генеральной совокупности.

Заметим также, что вычисление выборочных характеристик для какого-либо набора полученных в исследовании данных может быть полезно даже без предположения, что наблюдения представляют собой независимые и одинаково распределенные случайные величины.

Точечными оценками параметров распределения называются функции от наблюдений, предназначенные для приближенного оценивания этих параметров. Если распределение параметризуется какими-то числовыми характеристиками (например, нормальное распределение однозначно задается своими математическим ожиданием и дисперсией), то соответствующие выборочные характеристики являются их точечными оценками.

Чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям. Эти требования заключаются в том, что оценка должна быть состоятельной, несмещенной и желательна эффективной.

Оценка $\hat{\theta}_n$ называется **состоятельной**, если при неограниченном увеличении выборки она сходится по вероятности к оцениваемому параметру: $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$ для любого $\varepsilon > 0$.

Свойство состоятельности — асимптотическое, оно может проявляться и при столь больших объемах выборки, которых на практике не встречается.

Оценка $\hat{\theta}_n$ называется **несмещенной** (оценкой без систематической ошибки), если ее математическое ожидание при любом n равно оцениваемому параметру: $M\hat{\theta}_n = \theta$. Несмещенность оценки характеризует ее «доасимптотические» свойства, т.е. является показателем ее «хороших» свойств при любом конечном объеме выборки.

Оценка называется **эффективной** (в некотором классе оценок), если она имеет минимальную дисперсию в этом классе.

Задача 1. Найти выборочное среднее по выборке объема $n = 20$.

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

Решение. Для упрощения расчетов перейдем к условным вариантам $u_i = x_i - 2620$.

u_i	-60	-20	0	30	80
n_i	2	3	10	4	1

Тогда $\bar{u} = [2(-60) + 3(-20) + 10 \cdot 0 + 1 \cdot 80] / 20 = 1$ и $\bar{x} = 2620 + \bar{u} = 2621$.

Замечание. В качестве числа, которое вычитается при переходе к условным вариантам (**условный нуль**), обычно выбирается варианта, стоящая в середине ряда, либо та, для которой частота максимальна (выборочная мода). В данном примере они совпадают.

11.2. Статистическая устойчивость основных выборочных характеристик

На практике важно знать, насколько выборочные характеристики отличаются от истинных значений характеристик генеральной совокупности.

Пусть ξ имеет характеристики $M\xi = a$, $D\xi = \sigma^2$, $M\xi^k = \alpha_k$, $M(\xi - M\xi)^k = \mu_k$, $F(x) = P(\xi < x)$. Соответствующими выборочными характеристиками будут $\bar{x}$, $\hat{\sigma}^2$, s^2 , $\hat{\alpha}_k$, $\hat{\mu}_k$, $F_n(x)$.

Докажем, что $\bar{x}$ является несмещенной и состоятельной оценкой математического ожидания. При этом воспользуемся известным неравенством Чебышева (см. гл. 8)

$$P(|\eta - M\eta| \geq \varepsilon) \leq \frac{D\eta}{\varepsilon^2},$$

справедливым для любой случайной величины η (имеющей математическое ожидание и дисперсию) и любого фиксированного $\varepsilon > 0$.

Вычислим математическое ожидание и дисперсию выборочного среднего:

$$M\bar{x} = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n Mx_i = \frac{na}{n} = a,$$

так как $Mx_i = M\xi$, т.е. выборочное среднее является несмещенной оценкой математического ожидания. Далее,

$$D\bar{x} = D\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n Dx_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \rightarrow 0 \text{ при } n \rightarrow \infty,$$

так как $Dx_i = D\xi$.

Отсюда в силу неравенства Чебышева для любого фиксированного $\varepsilon > 0$

$$P(|\bar{x} - a| \geq \varepsilon) \leq \frac{D\bar{x}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Итак, среднее арифметическое выборки $\bar{x}$ сходится по вероятности к математическому ожиданию случайной величины, т.е. является состоятельной оценкой математического ожидания. Аналогично доказывается состоятельность оценок других начальных моментов $\hat{\alpha}_k$.

Вычислим математическое ожидание выборочных начальных моментов: $M\hat{\alpha}_k = M\left[\frac{1}{n} \sum_{i=1}^n x_i^k\right] = \frac{1}{n} \sum_{i=1}^n Mx_i^k = \frac{n\alpha_k}{n} = \alpha_k$.

Следовательно, выборочные начальные моменты являются несмещенными оценками теоретических начальных моментов. Очевидно, что для любого k справедливо равенство

$$D(\xi^k) = M(\xi^{2k}) - (M\xi^k)^2 = \alpha_{2k} - \alpha_k^2.$$

Учитывая, что x_i независимые случайные величины, получим соотношение

$$D\hat{\alpha}_k = D\left(\frac{1}{n} \sum_{i=1}^n x_i^k\right) = \frac{1}{n^2} \sum_{i=1}^n D x_i^k = \frac{\alpha_{2k} - \alpha_k^2}{n} \rightarrow 0 \text{ при } n \rightarrow \infty,$$

что является достаточным условием для сходимости по вероятности. Действительно, в силу неравенства Чебышева для любого фиксированного $\varepsilon > 0$

$$P(|\hat{\alpha}_k - \alpha_k| \geq \varepsilon) \leq \frac{D\hat{\alpha}_k}{\varepsilon^2} = \frac{\alpha_{2k} - \alpha_k^2}{n\varepsilon^2} \rightarrow 0 \text{ при } n \rightarrow \infty,$$

т.е. выборочный начальный момент $\hat{\alpha}_k$ k -го порядка является состоятельной оценкой начального момента α_k генеральной совокупности.

Покажем, что выборочная дисперсия $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ является смещенной оценкой дисперсии $D\xi = \sigma^2$ генеральной совокупности.

Поскольку выборочная дисперсия, как и теоретическая, не изменяется от прибавления к значениям случайной величины фиксированного числа, то для любого числа b выборочную дисперсию можно записать в виде

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - (\bar{x} - b)^2.$$

Действительно, доказательство утверждения вытекает из следующей цепочки равенств:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n [(x_i - b) - (\bar{x} - b)]^2 = \\ &= \frac{1}{n} \sum_{i=1}^n [(x_i - b)^2 - 2(x_i - b)(\bar{x} - b) + (\bar{x} - b)^2] = \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - \frac{2(\bar{x} - b)}{n} \sum_{i=1}^n (x_i - b) + \frac{1}{n} \sum_{i=1}^n (\bar{x} - b)^2 = \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - \frac{\bar{x} - b}{n} [2 \sum_{i=1}^n x_i - 2nb - n\bar{x} + nb] = \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - \frac{\bar{x} - b}{n} [2 \sum_{i=1}^n x_i - nb - \sum_{i=1}^n x_i] = \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - (\bar{x} - b) \left[\frac{1}{n} \sum_{i=1}^n x_i - b \right] = \frac{1}{n} \sum_{i=1}^n (x_i - b)^2 - (\bar{x} - b)^2. \end{aligned}$$

Следствие.

Если $M\xi = a$ есть среднее генеральной совокупности, то

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2 - (\bar{x} - a)^2.$$

Вычислим математическое ожидание выборочной дисперсии:

$$\begin{aligned} M\hat{\sigma}^2 &= M\left[\frac{1}{n} \sum_{i=1}^n (x_i - a)^2\right] - M(\bar{x} - a)^2 = \frac{1}{n} \sum_{i=1}^n M(x_i - a)^2 - D\bar{x} = \\ &= \frac{n\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Отсюда следует, что $\hat{\sigma}^2$ является смещенной оценкой дисперсии.

Несмещенной оценкой дисперсии будет $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, так как

$$Ms^2 = M\left(\frac{n}{n-1} \hat{\sigma}^2\right) = \frac{n}{n-1} M\hat{\sigma}^2 = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2.$$

Поэтому оценку $\hat{\sigma}^2$ (при получении которой сумма делится на n) называют **смещенной**, или **неисправленной выборочной дисперсией**, а оценку s^2 (при получении которой сумма делится на $n-1$) называют **несмещенной**, или **исправленной выборочной дисперсией**. **Выборочным средним квадратическим отклонением (исправленным или неисправленным)** называют положительный корень из соответствующей выборочной дисперсии. Эта оценка теоретического среднего квадратического отклонения $\sigma = \sqrt{D\xi}$, к сожалению, в обоих случаях является смещенной. Далее по умолчанию мы будем иметь в виду исправленные оценки.

Обычный закон больших чисел неприменим к центрированным величинам, так как после центрирования они становятся зависимыми. Однако с помощью теоремы Слуцкого можно доказать, например, что выборочная дисперсия сходится по вероятности к теоретической дисперсии, если последняя существует.

► **Теорема 1 (теорема Слуцкого).** Пусть функция $f(x, y)$ непрерывна в точке (a, b) и случайные последовательности X_n и Y_n сходятся по вероятности соответственно к числам a и b . Тогда $f(X_n, Y_n)$ сходится по вероятности к $f(a, b)$.

Доказательство. Если $f(x, y)$ непрерывна в точке (a, b) , то для любого $\varepsilon > 0$ существует $\delta > 0$, такое, что при $|x - a| < \delta$ и $|y - b| < \delta$ выполняется неравенство $|f(x, y) - f(a, b)| < \varepsilon$. Тогда, если $|f(x, y) - f(a, b)| \geq \varepsilon$, то справедливо хотя бы одно из неравенств: $|x - a| \geq \delta$ или $|y - b| \geq \delta$. Рассмотрим событие $\{\omega: |f(X_n, Y_n) - f(a, b)| \geq \varepsilon\} \subseteq \{\omega: |X_n - a| \geq \delta\} \cup \{\omega: |Y_n - b| \geq \delta\}$.

В силу теоремы сложения получаем

$$P(|f(X_n, Y_n) - f(a, b)| \geq \varepsilon) \leq P(\{|X_n - a| \geq \delta\} \cup \{|Y_n - b| \geq \delta\}) = \\ = P(|X_n - a| \geq \delta) + P(|Y_n - b| \geq \delta) - P(|X_n - a| \geq \delta, |Y_n - b| \geq \delta) \leq \\ \leq P(|X_n - a| \geq \delta) + P(|Y_n - b| \geq \delta),$$

причем $P(|X_n - a| \geq \delta) \rightarrow 0$ и $P(|Y_n - b| \geq \delta) \rightarrow 0$ при $n \rightarrow \infty$. Следовательно, вся сумма стремится к нулю, т.е.

$$\lim_{n \rightarrow \infty} P(|f(X_n, Y_n) - f(a, b)| \geq \varepsilon) = 0 \text{ или } f(X_n, Y_n) \xrightarrow{p} f(a, b).$$

Для доказательства сходимости по вероятности выборочной дисперсии к теоретической достаточно рассмотреть функцию $f(x, y) = x - y^2$, положив

$$X_n = \frac{1}{n} \sum_{i=1}^n x_i^2 \rightarrow \alpha_2 = a^2 + \sigma^2, \quad Y_n = \bar{x} \rightarrow a \quad (\text{при } n \rightarrow \infty),$$

и применить теорему Слуцкого, поскольку выборочная дисперсия представима в виде

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Конечно, сходимость выборочной дисперсии можно доказать и проще. Однако подобным образом доказывается сходимость по вероятности всех выборочных центральных моментов $\hat{\mu}_k \xrightarrow{p} \mu_k$, если существует $\mu_k = M(\xi - a)^k$. Действительно, из сходимости по вероятности случайных величин $\xi_i(n)$ к некоторым постоянным числам a_i следует, в силу теоремы Слуцкого, сходимость по вероятности любой непрерывной функции $\varphi(\xi_1(n), \xi_2(n), \dots, \xi_k(n))$ к ее значению в точке $(a_1, a_2, \dots, a_k)$, т.е. если $\xi_i(n) \xrightarrow{p} a_i$, то $\varphi(\xi_1(n), \xi_2(n), \dots, \xi_k(n)) \xrightarrow{p} \varphi(a_1, a_2, \dots, a_k)$ при $n \rightarrow \infty$.

Поскольку все выборочные центральные моменты (а также некоторые их модификации — асимметрия, эксцесс и т.п.) являются непрерывными функциями от наблюдений, они схо-

дятся по вероятности к соответствующим теоретическим значениям.

Справедлива следующая теорема.

► **Теорема 2.** Если $\hat{\theta}_n$ — несмещенная оценка параметра θ и ее дисперсия стремится к нулю, $D(\hat{\theta}_n) \rightarrow 0$ при $n \rightarrow \infty$, то оценка состоятельная.

Доказательство. Пусть $\hat{\theta}_n$ — несмещенная оценка параметра θ , т.е. $M\hat{\theta}_n = \theta$. Тогда для любого $\varepsilon > 0$ из неравенства Чебышева получаем $P(|\hat{\theta}_n - \theta| \geq \varepsilon) \leq \frac{D\hat{\theta}_n}{\varepsilon^2}$.

По условию $D\hat{\theta}_n \rightarrow 0$ при $n \rightarrow \infty$ и, следовательно, при каждом фиксированном $\varepsilon > 0$ получаем $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0$. Теорема доказана.

Задача 2. Найти неисправленную выборочную дисперсию по выборке объема $n = 50$.

x_i	18,4	18,9	19,3	19,6
n_i	5	10	20	15

Решение. Перейдем к условным вариантам $u_i = 10(x_i - 19,3)$.

Тогда $\hat{\sigma}^2(u) = 100\hat{\sigma}^2(u)$ и

u_i	-9	-4	0	3
n_i	5	10	20	15

Найдем выборочную дисперсию для новой варианты u_i :

$$\begin{aligned}\hat{\sigma}^2(u) &= \frac{1}{n} \sum_{i=1}^4 n_i u_i^2 - \left(\frac{\sum_{i=1}^4 n_i u_i}{n} \right)^2 = \\ &= \frac{1}{50} (5(-9)^2 + 10(-4)^2 + 20 \cdot 0^2 + 15 \cdot 3^2) - \\ &- \left[\frac{1}{50} (-45 - 40 + 0 + 45) / 50 \right]^2 = 13,36.\end{aligned}$$

Тогда, переходя к первоначальной вариане x_i :

$$\hat{\sigma}^2(x) = \hat{\sigma}^2(u)/100 = 13,36/100 = 0,1336.$$

Задача 3. По выборке объема $n = 50$ найдена смещенная оценка $\hat{\sigma}^2 = 9,8$ теоретической дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

Решение. Несмещенная оценка дисперсии связана со смещенной следующей формулой:

$$s^2 = \hat{\sigma}^2 \cdot \frac{n}{n-1} = 9,8 \cdot 50/49 = 10.$$

11.3. Асимптотически нормальный характер основных выборочных характеристик

Свойства выборки объема n зависят от распределения генеральной совокупности, но по мере увеличения n ($n \rightarrow \infty$) выборочные характеристики начинают вести себя одинаковым образом независимо от специфики генеральной совокупности. Поэтому характер поведения выборочных характеристик следует рассматривать в двух вариантах: при фиксированном n (ограниченном объеме выборки) и при $n \rightarrow \infty$ (асимптотические свойства выборки). При фиксированном n свойства выборок будут различны для разных типов генеральной совокупности (нормальной, экспоненциальной, равномерной, пуассоновской и т.д.). В условиях асимптотики ($n \rightarrow \infty$) общий характер поведения числовых характеристик практически не зависит от типа анализируемой генеральной совокупности. Введем следующее определение.

Последовательность случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ называется **асимптотически нормальной**, если существуют числовые последовательности $A_1, A_2, \dots$ и $B_1, B_2, \dots$ ($B_i > 0$ для всех i), такие, что

$$P\left(\frac{\xi_i - A_i}{\sqrt{B_i}} < x\right) \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad \text{при } n \rightarrow \infty.$$

Здесь $\Phi(x)$ — функция стандартного нормального распределения. Числа A_i и B_i называются *параметрами асимптотически*

нормально распределенной случайной величины ξ_i . Условие, что последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ асимптотически нормальна, записывается в виде $\xi_i \xrightarrow{\epsilon} N(A, B)$.

Используя введенный термин, центральную предельную теорему можно сформулировать следующим образом: пусть $\eta_1, \eta_2, \dots, \eta_n, \dots$ — независимые одинаково распределенные случайные величины с конечными моментами первого и второго порядков: $M\eta_i = a$; $D\eta_i = \sigma^2$. Тогда если $S_n = \eta_1 + \eta_2 + \dots + \eta_n$, то $S_n \xrightarrow{\epsilon} N(na, n\sigma^2)$, $n \rightarrow \infty$.

► **Теорема 3.** Если распределение генеральной совокупности имеет конечные математическое ожидание a и дисперсию σ^2 , то при $n \rightarrow \infty$ основные выборочные характеристики являются асимптотически нормальными:

$$1) \bar{x} \xrightarrow{\epsilon} N\left(a, \frac{\sigma^2}{n}\right);$$

$$2) \hat{\sigma}^2, s^2 \xrightarrow{\epsilon} N\left(\sigma^2, \frac{\mu_4 - \sigma^4}{n}\right);$$

$$3) F_n(x) \xrightarrow{\epsilon} N\left(F(x), \frac{F(x)(1-F(x))}{n}\right).$$

Другими словами, при больших объемах выборки все основные выборочные характеристики можно считать практически нормально распределенными. Доказательство следует из центральной предельной теоремы.

11.4. Эффективность оценок.

Неравенство Рао—Фреше—Крамера

Для одних и тех же параметров распределения существует бесконечно много различных несмещенных и состоятельных оценок. Поэтому важной задачей является сравнение их между собой и поиск наилучшей среди них. Естественным критерием такого поиска является дисперсия, как мера разброса значений случайных величин вокруг среднего значения. Наилучшей оценкой является оценка с минимальной дисперсией. Однако

для смещенной оценки $\hat{\theta}_n$ дисперсия служит мерой близости не к оцениваемому параметру θ , а к математическому ожиданию этой же оценки $M\hat{\theta}_n$. Поэтому естественно искать наилучшие оценки с наименьшей дисперсией только среди несмещенных оценок.

Получая ту или иную оценку, мы должны иметь возможность определить, обладает ли она минимальной дисперсией из всех возможных, или нет. С этой целью вводится понятие эффективности оценки и используется неравенство Рао—Фреше—Крамера.

Информацией Фишера о неизвестном параметре θ , содержащейся в одном из независимых наблюдений случайной величины ξ , называется величина

$$I(\theta) = M \left(\frac{\partial \ln p(\xi, \theta)}{\partial \theta} \right)^2,$$

где в качестве $p(x, \theta)$ берется либо плотность в точке x (для непрерывных случайных величин), либо вероятность принять значение x (для дискретных случайных величин). Говорят, что величина $I(\theta)$ определяет количество информации Фишера.

► **Теорема 4 (Рао—Фреше—Крамера).**

Пусть плотность $p(x, \theta)$ удовлетворяет следующим условиям регулярности:

1) область $G_n = \{x: p(x, \theta) > 0\}$ всех возможных значений случайной величины, для которых $p(x, \theta) \neq 0$, не зависит от параметра θ ;

2) в тождествах $\int_{R^n} \hat{\theta}_n(X) p(X, \theta) dX = M\hat{\theta}_n$ и $\int_{R^n} p(x, \theta) dx \equiv 1$ допустимо дифференцирование по параметру θ под знаком интеграла;

3) информация Фишера $I(\theta)$ конечна и положительна.

Тогда для произвольной несмещенной оценки $\hat{\theta}_n$, построенной по выборке объема n , выполняется неравенство (Рао—Фреше—Крамера):

$$D\hat{\theta}_n \geq \frac{1}{nI(\theta)}.$$

Доказательство. Пусть $\hat{\theta}_n$ — несмещенная оценка параметра θ , т.е.

$$M\hat{\theta}_n = R^n \int_{R^n} \hat{\theta}_n(X) p(X, \theta) dX = \theta, \quad (11.1)$$

где $X = (x_1, x_2, \dots, x_n)$ и $p(X, \theta)$ — плотность распределения, так что

$$\int_{R^n} p(X, \theta) dX = 1. \quad (11.2)$$

Дифференцируя по θ равенства (11.1) и (11.2), получим

$$\int_{R^n} \hat{\theta}_n(X) \frac{\partial p(X, \theta)}{\partial \theta} dX = 1 \text{ и } \int_{R^n} \frac{\partial p(X, \theta)}{\partial \theta} dX = 0.$$

Умножим второе равенство на θ и вычтем его из первого:

$$1 = \int_{R^n} (\hat{\theta}_n(X) - \theta) \frac{\partial p(X, \theta)}{\partial \theta} dX. \quad (11.3)$$

По условию на множестве $G_n = G^n$ плотность $p(X, \theta) > 0$, поэтому можно записать, что

$$\frac{\partial p(X, \theta)}{\partial \theta} = \frac{\partial \ln p(X, \theta)}{\partial \theta} p(X, \theta).$$

Подставим полученное выражение в равенство (11.3) и, используя неравенство Коши—Буняковского, находим

$$\begin{aligned} 1 &= \left[\int_{R^n} (\hat{\theta}_n(X) - \theta) \frac{\partial \ln p(X, \theta)}{\partial \theta} p(X, \theta) dX \right]^2 \leq \\ &\leq \left[\int_{R^n} (\hat{\theta}_n(X) - \theta)^2 p(X, \theta) dX \right] \cdot \left[\int_{R^n} \left[\frac{\partial \ln p(X, \theta)}{\partial \theta} \right]^2 p(X, \theta) dX \right], \end{aligned}$$

или

$$1 \leq D\hat{\theta}_n \cdot M \left(\frac{\partial \ln p(X, \theta)}{\partial \theta} \right)^2.$$

Учитывая независимость и одинаковый закон распределения наблюдений $x_1, x_2, \dots, x_n$, можно записать, что

$$M \left(\frac{\partial \ln p(X, \theta)}{\partial \theta} \right)^2 = n M \left(\frac{\partial \ln p(x_i, \theta)}{\partial \theta} \right)^2 = n I(\theta).$$

Подставляя это выражение в последнее неравенство, окончательно получаем утверждение теоремы: $D\hat{\theta}_n \geq \frac{1}{nI(\theta)}$.

Теорема верна и в дискретном случае, если в условии 2 заменить интегралы на суммы (по всем возможным значениям случайной величины).

Заметим, что информацию Фишера можно также представить в виде

$$I(\theta) = D\left(\frac{\partial \ln p(\xi, \theta)}{\partial \theta}\right) \text{ и } I(\theta) = -M\left(\frac{\partial^2 \ln p(\xi, \theta)}{\partial \theta^2}\right).$$

Обозначим правую часть неравенства Рао—Фреше—Крамера через $\Delta_n = \frac{1}{nI(\theta)}$. Эта величина является нижней гранью всех возможных дисперсий оценок.

Эффективностью несмещенной оценки $\hat{\theta}_n$ называется отношение минимально возможного значения дисперсии оценки в классе всех несмещенных оценок параметра θ к дисперсии рассматриваемой оценки:

$$e(\hat{\theta}_n) = \frac{\Delta_n}{D\hat{\theta}_n} = \frac{1}{nI(\theta)D\hat{\theta}_n}.$$

Из определения следует, что эффективность любой несмещенной оценки удовлетворяет неравенству $0 \leq e(\hat{\theta}_n) \leq 1$, и чем ближе она к единице, тем лучше оценка.

Несмещенная оценка $\hat{\theta}_n$ называется **эффективной**, если $e(\hat{\theta}_n) = 1$.

Асимптотической эффективностью оценки называется предел

$$e_0(\hat{\theta}) = \lim_{n \rightarrow \infty} e(\hat{\theta}_n),$$

если он существует.

Оценку $\hat{\theta}_n$ называют **асимптотически эффективной**, если $e_0(\hat{\theta}) = 1$.

Кроме того, для асимптотически нормальных оценок понятие асимптотической эффективности иногда трактуется более широко, а именно: для асимптотически нормальной оценки $\hat{\theta}_n \xrightarrow{\epsilon} N(\theta, \sigma^2/n)$ при $n \rightarrow \infty$, полагают

$$e_0(\hat{\theta}) = \frac{1}{I(\theta)\sigma^2}.$$

Задача 4. Доказать, что выборочное среднее является эффективной оценкой математического ожидания нормального распределения, когда дисперсия известна.

Решение. Выпишем функцию плотности для нормального распределения:

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

Прологарифмировав ее, получим

$$\ln p(x, a) = \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x-a)^2}{2\sigma^2},$$

при этом производная будет равна

$$\frac{\partial \ln p(x, a)}{\partial a} = \frac{2(x-a)}{2\sigma^2} = \frac{x-a}{\sigma^2}, \quad \left(\frac{\partial \ln p(x, a)}{\partial a} \right)^2 = \frac{(x-a)^2}{\sigma^4}.$$

Отсюда найдем информацию Фишера:

$$I(a) = M \left(\frac{\partial \ln p(\xi, a)}{\partial a} \right)^2 = \frac{1}{\sigma^4} M(x-a)^2 = \frac{1}{\sigma^2}.$$

Получаем значение $\Delta_n = \sigma^2/n$. С другой стороны, $D\bar{x} = \sigma^2/n$, так что $D\bar{x} = \Delta_n$. Таким образом, оценка является эффективной.

Из доказанного следует, что чем больше дисперсия нормальной случайной величины, тем меньше информации о значении среднего этой величины заключено в одном наблюдении.

Задача 5. Доказать, что относительная частота успеха в качестве оценки неизвестной вероятности θ в схеме Бернулли является эффективной оценкой.

Решение. Оценкой неизвестной вероятности является относительная частота успеха $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i$, где x_i — успех (1) или неудача (0) в i -м испытании. Покажем, что оценка несмещенная:

$$M\hat{\theta} = \frac{1}{n} M(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n Mx_i = M\xi = \theta.$$

Дисперсия имеет вид:

$$D\hat{\theta} = D \left(\frac{1}{n} \sum_{i=1}^n x_i \right) = \frac{1}{n^2} \sum_{i=1}^n Dx_i = \frac{D\xi}{n} = \frac{\theta(1-\theta)}{n}.$$

Найдем информацию Фишера, причем в данном случае наблюдаемая величина принимает всего два значения: 0 и 1 с вероятностями $P(0; \theta) = 1 - \theta$ и $P(1; \theta) = \theta$.

$$I(\theta) = \left[\frac{\partial \ln P(0; \theta)}{\partial \theta} \right]^2 P(0; \theta) + \left[\frac{\partial \ln P(1; \theta)}{\partial \theta} \right]^2 P(1; \theta) = \\ = \left(-\frac{1}{1-\theta} \right)^2 (1-\theta) + \left(\frac{1}{\theta} \right)^2 \theta = \frac{1}{\theta(1-\theta)}.$$

Таким образом, $e(\hat{\theta}_n) = \frac{1}{nI(\theta)D(\hat{\theta}_n)} = 1$.

Задача 6. Пусть выборка $x_1, x_2, \dots, x_n$ произведена из генеральной совокупности с равномерным распределением на интервале $(0; \theta)$. Проверить на эффективность оценку $\hat{\theta} = \frac{n+1}{n} x_{\max}$ для неизвестного параметра θ .

Решение. Функция распределения $F_{\max}(x)$ максимума $x_{\max}$ задается формулой

$$F_{\max}(x) = P(x_{\max} < x) = P(x_1 < x, \dots, x_n < x) = P(x_1 < x) \dots P(x_n < x) = \left(\frac{x}{\theta} \right)^n$$

на отрезке $0 \leq x \leq \theta$. Отсюда получаем

$$M\hat{\theta} = \frac{n+1}{n} \int_0^\theta nx \frac{x^{n-1}}{\theta^n} dx = \theta.$$

Значит, оценка $\hat{\theta}$ несмещенная. Найдём дисперсию этой оценки:

$$M\hat{\theta}^2 = \left(\frac{n+1}{n} \right)^2 \int_0^\theta nx^2 \frac{x^{n-1}}{\theta^n} dx = \frac{(n+1)^2}{n(n+2)} \theta^2;$$

$$D\hat{\theta} = M\hat{\theta}^2 - (M\hat{\theta})^2 = \frac{\theta^2}{n(n+2)}.$$

Видно, что дисперсия оценки $\hat{\theta}_n$ при $n \rightarrow \infty$ убывает как $\frac{1}{n^2}$.

Такая оценка оказалась лучше эффективной, поскольку дисперсия эффективной оценки имеет порядок убывания только $\frac{1}{n}$. Разгадка парадокса в том, что для данного семейства распределений не выполнены условия теоремы Рао—Фреше—Крамера, а именно: область значений случайной величины зависит от параметра θ .

Подобные оценки называют **сверхэффективными**.

11.5. Оценка математического ожидания по неравноточным наблюдениям

Ранее предполагалось, что все наблюдения равноточны, т.е. имеют одинаковую дисперсию (и среднее квадратическое отклонение). Однако на практике встречаются и ситуации *неравноточных* наблюдений. А именно пусть выполнено

$Mx_1 = Mx_2 = \dots = Mx_n = a$; $Dx_1 = \sigma_1^2$, $Dx_2 = \sigma_2^2$, ..., $Dx_n = \sigma_n^2$, и надо найти наилучшую (в смысле несмещенности и эффективности) оценку для a .

Классом линейных несмещенных оценок параметра θ называется класс оценок вида $\hat{\theta}_n = c_1x_1 + c_2x_2 + \dots + c_nx_n$, для которых $M\hat{\theta}_n = \theta$.

Числовые коэффициенты $c_1, c_2, \dots, c_n$ называются **весами** наблюдений. Из требования несмещенности оценки следует, что должно выполняться равенство $c_1 + c_2 + \dots + c_n = 1$. Чтобы получить эффективную оценку (в классе линейных несмещенных оценок), надо минимизировать дисперсию $D\hat{a}_n = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \dots + c_n^2\sigma_n^2$.

Действительно, из требования несмещенности оценки

$$M\hat{a}_n = M \sum_{i=1}^n c_i x_i = \sum_{i=1}^n c_i (Mx_i) = a \sum_{i=1}^n c_i = a$$

получаем, что оценка будет несмещенной, если $\sum_{i=1}^n c_i = 1$.

По определению, оценка будет эффективной, если она имеет минимальную дисперсию. Коэффициенты c_i надо определить так, чтобы дисперсия была минимальна, но при условии, что $\sum_{i=1}^n c_i = 1$. Вычислим дисперсию $D\hat{a}_n$:

$$D\hat{a}_n = D \sum_{i=1}^n c_i x_i = \sum_{i=1}^n c_i^2 Dx_i = \sum_{i=1}^n c_i^2 \sigma_i^2,$$

поскольку случайные величины x_i независимы и дисперсия суммы равна сумме дисперсий.

Задача свелась к нахождению таких коэффициентов c_i , при которых функция $f(c) = \sum_{i=1}^n c_i^2 \sigma_i^2$ имеет минимум при условии $g(c) = \sum_{i=1}^n c_i - 1 = 0$. Это задача на условный экстремум. Функция

Лагранжа имеет вид $L(c) = f(c) - \lambda g(c)$. Для определения коэффициентов приравняем к нулю производные:

$$\begin{cases} \frac{\partial L}{\partial c_i} = 2c_i\sigma_i^2 - \lambda = 0, \\ \frac{\partial L}{\partial \lambda} = -g(c) = -\left(\sum_{i=1}^n c_i - 1\right) = 0; \end{cases} \Rightarrow \begin{cases} c_i = \frac{\lambda}{2\sigma_i^2}, \\ \sum_{i=1}^n c_i = 1; \end{cases} \Rightarrow \begin{cases} c_i = \frac{\lambda}{2\sigma_i^2}, \\ \frac{\lambda}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} = 1. \end{cases} \Rightarrow$$

$$\begin{cases} c_i = \frac{\lambda}{2\sigma_i^2}, \\ \lambda = 2\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^{-1}. \end{cases}$$

Из этих условий получаем значения для коэффициентов $c_i = \sigma_i^{-2} / \sum_{j=1}^n \sigma_j^{-2}$. Таким образом, веса наблюдений должны быть обратно пропорциональны их дисперсиям (менее точные наблюдения, имеющие большую дисперсию, входят с меньшим весом, более точные — с большим). В результате эффективной оценкой математического ожидания оказывается средневзвешенное значение, имеющее вид

$$\hat{a}_n = \frac{\sum_{i=1}^n \frac{1}{\sigma_i^2} x_i}{\sum_{i=1}^n \frac{1}{\sigma_i^2}},$$

с дисперсией, равной $D\hat{a}_n = 1 / \sum_{j=1}^n \sigma_j^{-2}$.

Если дисперсии всех наблюдений равны (т.е. наблюдения *равноточные*), то $c_i = 1/n$ для любого $i = 1, 2, \dots, n$ и $\hat{a}_n = \frac{1}{n} \sum_{i=1}^n x_i$, т.е. среднее арифметическое выборки является эффективной оценкой математического ожидания в классе линейных несмещенных оценок (при любом распределении, имеющем конечное математическое ожидание и дисперсию).

Задачи для самостоятельного решения

Теоретические задачи

1. Доказать состоятельность выборочного коэффициента вариации в случае $M\xi > 0$, выборочных коэффициентов асимметрии и эксцесса.
2. Пусть $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ — случайные выборки объема n и m из нормально распределенной генеральной совокупности $N(a, \sigma^2)$ с исправленными выборочными дисперсиями s_x^2, s_y^2 . Доказать, что

$$s_{x,y}^2 = \frac{1}{n+m-2} [(n-1)s_x^2 + (m-1)s_y^2]$$

является несмещенной оценкой параметра σ^2 .

3. Вывести формулы, связывающие третий и четвертый центральные моменты с начальными моментами. Построить соответствующие формулы для выборочных моментов и доказать их состоятельность.
4. Доказать, что если $M\hat{\theta}_n \rightarrow \theta$ и $D\hat{\theta}_n \rightarrow 0$, то $\hat{\theta}_n$ — состоятельная оценка параметра θ .
5. Доказать, что для показательного распределенной с функцией плотности

$$p(x, \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x \geq 0, \\ 0, & x < 0 \end{cases}$$

выборочное среднее является эффективной оценкой параметра θ .

6. Доказать, что выборочное среднее является эффективной оценкой параметра λ в распределении Пуассона: $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.
7. Доказать, что для распределения Бернулли $P(\xi = m) = C_N^m p^m q^{N-m}$ оценка $\hat{p} = \bar{x} / N$ является эффективной оценкой вероятности p .
8. Доказать, что для геометрического распределения $P(\xi = m) = p(1-p)^{m-1}$ с параметром $p = 1/\theta, \theta > 1$, выборочное среднее является эффективной оценкой параметра θ .
9. Доказать, что эмпирическая функция распределения $F_n(x)$ является эффективной оценкой теоретической функции распределения $F(x)$ при каждом x .

10. Пусть $F(x) = e^{x-\theta}$, $x < \theta$. Построить несмещенную оценку $\hat{\theta}$ на основе $x_{\max}$ и доказать ее сверхэффективность.
11. Пусть $F(x) = (x/\theta)^\beta$, $0 \leq x \leq \theta$, $\beta > 0$. При известном значении β построить несмещенную оценку $\hat{\theta}$ на основе $x_{\max}$ и доказать ее сверхэффективность.
12. В случае распределения Парето

$$F(x) = \begin{cases} 0, & x < \theta; \\ 1 - (x/\theta)^{-2}, & x \geq \theta \end{cases}$$

оценивается параметр θ . Проверить эффективность оценки $\hat{\theta} = \frac{\bar{x}}{2}$.

Вычислительные задачи

13. Найти исправленную выборочную дисперсию по выборке объема $n = 50$

x_i	0,1	0,5	0,6	0,8
n_i	5	15	20	10

14. В итоге четырех измерений некоторой величины одним прибором (без систематических ошибок) получены следующие результаты: 8, 9, 11, 12. Найти: а) выборочное среднее результатов измерений; б) смещенную и исправленную выборочные дисперсии ошибок прибора.
15. С помощью измерительного прибора, не имеющего систематической ошибки, было сделано 8 независимых измерений некоторой величины.

Номер измерения	1	2	3	4	5	6	7	8
x_i	2504	2486	2525	2495	2515	2528	2492	2494

Найти несмещенные оценки математического ожидания и дисперсии.

16. При измерении веса 20 шоколадных батончиков (с номинальным весом 50 г) получены следующие значения (г): 49,1; 50,0; 49,7; 50,5; 48,1; 50,3; 49,7; 51,6; 49,8; 50,1; 49,7; 48,8; 51,4; 49,1; 49,6; 50,9; 48,5; 52,0; 50,7; 50,6. Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную медиану, крайние члены вариационного ряда.
17. В ОТК были измерены диаметры 300 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала (нм) даны в таблице:

Границы отклонений	Середина интервала	Число валиков	Границы отклонений	Середина интервала	Число валиков
-30...-25	-27,5	3	0...5	2,5	55
-25...-20	-22,5	8	5...10	7,5	30
-20...-15	-17,5	15	10...15	12,5	25
-15...-10	-12,5	35	15...20	17,5	14
-10...-5	-7,5	40	20...25	22,5	8
-5...0	-2,5	60	25...30	27,5	7

Найти выборочное среднее, выборочные дисперсии, выборочную моду.

18. В таблице представлены результаты наблюдений случайной величины ξ . Найти выборочное среднее $\bar{x}$, исправленную выборочную дисперсию s^2 , выборочное среднее квадратическое отклонение s .

а) ξ — число сделок на фондовой бирже за квартал; $n = 400$ (инвесторов).

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

б) ξ — месячный доход жителя региона (руб.); $n = 1000$ (жителей).

x_i	Менее 500	500—1000	1000—1500	1500—2000	2000—2500	Свыше 2500
n_i	58	96	239	328	147	132

Указание. В качестве верхней границы последнего интервала использовать 3000.

в) ξ — удой коров на молочной ферме за лактационный период (ц); $n = 100$ (коров).

x_i	4—6	6—8	8—10	10—12	12—14	14—16	16—18	18—20	20—22	22—24	24—26
n_i	1	3	6	11	15	20	14	12	10	6	2

19. В таблице приведено распределение 50 рабочих по производительности труда ξ (единиц за смену), разделенных на две группы: 30 и 20 человек.

Прошедшие техническое обучение (группа I)						Не прошедшие техническое обучение (группа II)				
x_i	85	90	95	100	105	85	90	95	100	105
n_i	2	5	11	8	4	2	6	8	3	1

Вычислить общие и групповые выборочные средние и дисперсии.

20. Расстояние до цели определено двумя способами. Точность первого способа характеризуется средним квадратическим отклонением $\sigma_1 = 30$ м, результат измерения $x_1 = 1480$ м; точность второго — средним квадратическим отклонением $\sigma_2 = 40$ м, результат измерения $x_2 = 1560$ м. Определить приближенное значение расстояния до цели и оценить его точность.
21. Диаметр втулки определен четырьмя способами, точность которых $\sigma_1^2 = 1,6$ мм², $\sigma_2^2 = 2$ мм², $\sigma_3^2 = 2,5$ мм², $\sigma_4^2 = 3$ мм². Результаты измерений: $x_1 = 19$ мм, $x_2 = 18$ мм, $x_3 = 20$ мм, $x_4 = 21$ мм. Определить приближенное значение диаметра втулки и оценить его точность.
22. При измерении диаметра детали одним прибором установлен средний диаметр $\bar{x}_1 = 10$ мк, $n_1 = 8$. При измерении другим равноточным прибором — $\bar{x}_2 = 12$ мк, $n_2 = 16$. Определить наиболее точную оценку диаметра по измерениям двух приборов.
23. Для изучения распределения заработной платы работников некоторой отрасли обследовано 100 человек. Результаты представлены в следующей таблице.

Заработная плата, долл.	Число человек	Заработная плата, долл.	Число человек
190—192	1	200—202	19
192—194	5	202—204	11
194—196	9	204—206	4
196—198	22	206—208	1
198—200	28	208—210	0

Найти несмещенные оценки математического ожидания и дисперсии зарплаты.

24. Производство зерна в России в 1996—2002 гг. представлено таблицей:

Год	1996	1997	1998	1999	2000	2001	2002
Производство, млн т	69,3	88,6	47,9	54,7	65,5	85,2	86,6

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную медиану, крайние члены вариационного ряда.

25. Производство пшеницы в России в 1995–2001 гг. представлено таблицей:

Год	1995	1996	1997	1998	1999	2000	2001
Производство, млн т	30,1	34,9	44,3	27,0	31,0	34,5	47,0

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную медиану, крайние члены вариационного ряда.

26. Урожайность зерновых культур в России в 1992–2001 гг. представлена таблицей:

Год	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Урожайность, ц/га	18,0	17,1	15,3	13,1	14,9	17,8	12,9	14,4	15,6	19,4

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную медиану, крайние члены вариационного ряда.

27. Численность населения городов России с числом жителей более 1 млн чел. на 2002 г. представлена таблицей:

Город	Население, тыс. чел.
Волгоград	1013
Екатеринбург	1293
Казань	1105
Москва	10 358
Нижний Новгород	1311
Новосибирск	1426
Омск	1134
Пермь	1000
Ростов-на-Дону	1070
Самара	1158
Санкт-Петербург	4669
Уфа	1042
Челябинск	1078

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную медиану, крайние члены вариационного ряда.

28. Решить предыдущую задачу, исключив из выборки Москву и Санкт-Петербург, как города, имеющие федеральный статус.
29. В таблице приведены сгруппированные данные о коэффициентах соотношения заемных и собственных средств на 100 малых предприятиях региона.

Номер интервала	Интервал	Середина интервала x_i	n_i
1	5,05—5,15	5,1	5
2	5,15—5,25	5,2	8
3	5,25—5,35	5,3	12
4	5,35—5,45	5,4	20
5	5,45—5,55	5,5	26
6	5,55—5,65	5,6	15
7	5,65—5,75	5,7	10
8	5,75—5,85	5,8	4

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения.

30. Проведено исследование посещаемости популярного интернет-сайта. Много часов подряд регистрируется число посетивших сайт в течение данного часа. Результаты исследования представлены в таблице:

Число посетителей	Часы	Число посетителей	Часы
0	57	7	139
1	203	8	45
2	383	9	27
3	525	10	10
4	532	11	4
5	408	12	1
6	273	14	1

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную моду.

31. Проведено исследование посещаемости популярного интернет-сайта. Много часов подряд регистрируется число посетивших сайт в течение данного часа. Результаты исследования представлены в таблице.

Число посетителей	Часы	Число посетителей	Часы
0	12	7	103
1	108	8	24
2	316	9	13
3	551	10	2
4	632	11	0
5	492	12	0
6	273	14	0

Найти выборочное среднее, выборочные дисперсии, средние квадратические отклонения, выборочную моду.

32. В таблице приведены результаты измерения роста (см) случайно отобранных 100 студентов.

Рост	154—158	158—162	162—166	166—170	170—174	174—178	178—182
Число студентов	10	14	26	28	12	8	2

Найти выборочное среднее, выборочные дисперсии, выборочную моду.

33. Трое исследователей провели выборочное обследование доходов населения. Первый обследовал 10 семей и определил средний годовой доход 2400 у.е., второй — 25 семей и 2350 у.е., третий — 15 семей и 2450 у.е. Построить наиболее точную оценку среднего годового дохода.
34. Трое исследователей провели независимые выборочные обследования доходов населения. Первый определил средний годовой доход 3200 у.е. со средним квадратическим отклонением $\sigma_1 = 100$ у.е., второй — 2900 у.е. с $\sigma_2 = 50$ у.е., третий — 3000 у.е. с $\sigma_3 = 40$ у.е. Построить наиболее точную оценку среднего годового дохода и найти ее среднее квадратическое отклонение.

35. Трое исследователей провели независимые выборочные обследования доходов населения. Первый определил средний годовой доход 3200 у.е. со средним квадратическим отклонением $\sigma_1 = 200$ у.е., второй — 2900 у.е. с $\sigma_2 = 100$ у.е., третий — 3070 у.е. с $\sigma_3 = 50$ у.е. Построить наиболее точную оценку среднего годового дохода и найти ее среднее квадратическое отклонение.

Глава 12

ФУНКЦИИ И РАСПРЕДЕЛЕНИЯ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

12.1. Бета- и гамма-функции

Бета-функцией, или интегралом Эйлера первого рода, называется интеграл вида $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$, где параметры $a > 0$, $b > 0$.

Введенный интеграл сходится для любых положительных значений параметров.

Свойства бета-функции следующие:

1. $B(a, b) = B(b, a)$.

2. $B(a, b) = \frac{b-1}{a+b-1} B(a, b-1)$, $B(1, 1) = 1$.

Отсюда, в частности, следует, что $B(m, n) = \frac{(n-1)!(m-1)!}{(n+m-1)!}$.

3. $B(a, b) = \int_0^\infty \frac{y^{a-1}}{(1+y)^{a+b}} dy$.

В частности, $B(a, 1-a) = \int_0^\infty \frac{y^{a-1}}{1+y} dy = \frac{\pi}{\sin a\pi}$.

Гамма-функцией, или интегралом Эйлера второго рода, называется функция вида $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, где интеграл сходится для любого значения параметра $\alpha > 0$.

Свойства гамма-функции:

1. $\Gamma(a+1) = a\Gamma(a)$, откуда следует

$$\Gamma(a+n) = (a+n-1)(a+n-2)\dots(a+1)\Gamma(a).$$

2. $\Gamma(1) = 1$.

3. $\Gamma(n+1) = n!$, если $n \in \mathbb{N}$.

$$4. \Gamma(1/2) = \sqrt{\pi}.$$

5. $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ — формула связи гамма- и бета-функции.

$$6. \Gamma(a)\Gamma(1-a) = \frac{\pi}{\sin a\pi} \quad (\text{формула дополнения}).$$

$$7. \Gamma(x+1) \sim \left(\frac{x}{e}\right)^x \sqrt{2\pi x}, \quad x \rightarrow \infty \quad (\text{формула Стирлинга}).$$

Докажем свойства 1–4.

1) Вычислим интеграл $\Gamma(\alpha+1) = \int_0^{\infty} x^{\alpha} e^{-x} dx$, интегрируя по

$$\begin{aligned} \text{частям. Получим } \Gamma(\alpha+1) &= \int_0^{\infty} x^{\alpha} e^{-x} dx = -xe^{-x} \Big|_0^{\infty} + \\ &+ \alpha \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha). \end{aligned}$$

2) Очевидно,

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1.$$

3) Применим принцип математической индукции. Утверждение верно для $n=0$. Пусть оно выполнено для $n=k-1$: $\Gamma(k) = (k-1)!$. Тогда по свойству 1: $\Gamma(k+1) = k\Gamma(k) = k(k-1)! = k!$

4) Вычислим интеграл методом замены переменной ($y = x^{1/2}$):

$$\Gamma(1/2) = \int_0^{\infty} x^{-1/2} e^{-x} dx = 2 \int_0^{\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi} \quad (\text{интеграл}$$

Эйлера—Пуассона).

Поскольку гамма-функция не выражается через элементарные функции, ее значения табулированы. Обычно в таблицах представлены значения $\Gamma(x)$ для $1 \leq x \leq 2$, и этого достаточно для вычисления функции при любых $x > 0$ (с помощью свойства 1).

Гамма-функция стремится к бесконечности при $x \rightarrow 0$ и при $x \rightarrow +\infty$.

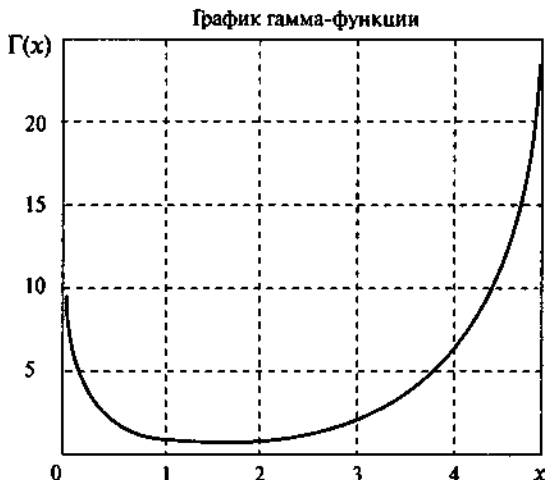


Рис. 12.1

На рис. 12.1 представлен график гамма-функции на отрезке $[0,01; 5]$.

Примеры

1. $\Gamma(3/2) = \Gamma(1 + 1/2) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$;
2. $\Gamma(4,7) = 3,7 \cdot 2,7 \cdot 1,7 \cdot \Gamma(1,7) = 3,7 \cdot 2,7 \cdot 1,7 \cdot 0,9086 = 15,43$;
3. $\Gamma(0,7) = \Gamma(1,7)/0,7 = 1,298$;
4. Пользуясь свойствами гамма-функции, можем вычислить:

$$\begin{aligned}
 B\left(\frac{7}{2}, \frac{3}{2}\right) &= \int_0^1 x^{\frac{5}{2}}(1-x)^{\frac{1}{2}} dx = \frac{\Gamma\left(\frac{7}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{7}{2} + \frac{3}{2}\right)} = \frac{\frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{\Gamma(5)} = \\
 &= \frac{5 \cdot 3 \cdot 1}{4! \cdot 2^4} \pi = \frac{5\pi}{2^7}.
 \end{aligned}$$

Задача 1. Доказать формулу, связывающую бета- и гамма-функции:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Решение. По определению бета-функция равна интегралу

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx. \text{ Сделаем замену переменных, положив}$$

$$y = \frac{x}{1-x}, \text{ тогда } x = \frac{y}{1+y} \text{ и}$$

$$B(a, b) = \int_0^{+\infty} \frac{y^{a-1}}{(1+y)^{a-1}} \left(1 - \frac{y}{1+y}\right)^{b-1} \cdot \frac{1}{(1+y)^2} dy = \int_0^{+\infty} \frac{y^{a-1}}{(1+y)^{a+b}} dy.$$

Для доказательства равенства преобразуем гамма-функцию, сделав в интегральном выражении для $\Gamma(\alpha)$ замену $x = ty$, тогда

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \int_0^{+\infty} t^{\alpha-1} y^{\alpha-1} e^{-ty} t dy.$$

Отсюда следует равенство $\frac{\Gamma(\alpha)}{t^\alpha} = \int_0^{+\infty} y^{\alpha-1} e^{-y} dy$. Заменим в полученном равенстве t на $t+1$ и положим $\alpha = a+b$. Получим

$$\frac{\Gamma(a+b)}{(1+t)^{a+b}} = \int_0^{+\infty} y^{(a+b)-1} e^{-(1+t)y} dy.$$

Умножив это равенство на t^{a-1} и проинтегрировав по t от 0 до $+\infty$, получим следующее равенство:

$$\Gamma(a+b) \cdot \int_0^{+\infty} \frac{t^{a-1}}{(1+t)^{a+b}} dt = \int_0^{+\infty} t^{a-1} dt \int_0^{+\infty} y^{(a+b)-1} e^{-y(1+t)} dy;$$

$$\begin{aligned} \Gamma(a+b) \cdot B(a, b) &= \int_0^{+\infty} y^{(a+b)-1} e^{-y} dy \int_0^{+\infty} t^{a-1} e^{-ty} dt = \int_0^{+\infty} y^{(a+b)-1} e^{-y} dy \cdot \frac{\Gamma(a)}{y^a} = \\ &= \Gamma(a) \int_0^{+\infty} y^{b-1} e^{-y} dy = \Gamma(a) \Gamma(b). \end{aligned}$$

Задача 2. Вычислить интегралы:

$$1. \int_0^{+\infty} \frac{e^{-x}}{\sqrt{x}} dx = \int_0^{+\infty} x^{\frac{1}{2}-1} e^{-x} dx = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

$$2. \int_0^{+\infty} x^4 e^{-x} dx = \int_0^{+\infty} x^{5-1} e^{-x} dx = \Gamma(5) = 4! = 24.$$

$$3. \int_0^1 x^3 (1-x)^2 dx = \int_0^1 x^{4-1} (1-x)^{3-1} dx = B(4, 3) = \frac{\Gamma(4)\Gamma(3)}{\Gamma(7)} = \frac{3! \cdot 2!}{6!} = \frac{1}{60}.$$

$$4. \int_0^1 x^{\frac{5}{2}} (1-x)^{\frac{1}{2}} dx = \frac{\Gamma\left(\frac{7}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma(5)} = \frac{\frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{4!} = \frac{5\pi}{64}.$$

$$\begin{aligned} 5. \int_0^2 x^{\frac{5}{2}} (2-x)^3 dx &= |x=2y, dx=2dy| = \int_0^1 (2y)^{\frac{5}{2}} (2-2y)^3 2dy = \\ &= 2^{\frac{5}{2}+4} \int_0^1 y^{\frac{5}{2}} (1-y)^3 dy = 2^{\frac{13}{2}} B\left(\frac{7}{2}, 4\right) = 2^{\frac{13}{2}} \frac{\Gamma\left(\frac{13}{2}\right)\Gamma(4)}{\Gamma\left(\frac{13}{2}+4\right)} = \\ &= \frac{2^{\frac{13}{2}} \cdot 3! \cdot \frac{11}{2} \cdot \frac{9}{2} \cdot \frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{\frac{19}{2} \cdot \frac{17}{2} \cdot \frac{15}{2} \cdot \frac{13}{2} \cdot \frac{11}{2} \cdot \frac{9}{2} \cdot \frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right)} = \frac{2^{22}}{19 \cdot 17 \cdot 5}. \end{aligned}$$

Задача 3. Найти $B\left(\frac{7}{2}, \frac{3}{2}\right)$.

$$\begin{aligned} \text{Решение. } B\left(\frac{7}{2}, \frac{3}{2}\right) &= \int_0^1 x^{\frac{5}{2}} (1-x)^{\frac{1}{2}} dx = \frac{\Gamma\left(\frac{7}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{7}{2}+\frac{3}{2}\right)} = \\ &= \frac{\frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{\Gamma(5)} = \frac{5 \cdot 3 \cdot 1}{4! \cdot 2^4} \pi = \frac{5\pi}{2^7}. \end{aligned}$$

12.2. Квантили, процентные и критические точки

Квантилью уровня p , или p -квантилью, непрерывной случайной величины ξ с функцией распределения $F(x)$ называется такое возможное значение x_p этой случайной величины, для которого вероятность события $\xi < x_p$ равна заданной величине p : $P(\xi < x_p) = p$, $0 < p < 1$, т.е. из определения следует, что x_p есть решение, по предположению единственное, уравнения $F(x_p) = p$, $0 < p < 1$.

Геометрически x_p есть такое значение случайной величины ξ , при котором площадь криволинейной трапеции, ограниченная графиком плотности распределения и осью абсцисс, и лежащая левее x_p , равна p (рис. 12.2).

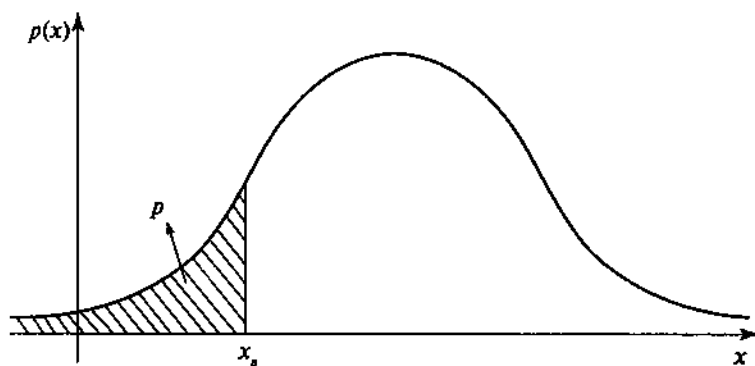


Рис. 12.2

Процентной точкой уровня q , или $q\%$ -ной точкой (при $0 \leq q \leq 100$), для непрерывной случайной величины ξ с функцией распределения $F(x)$ называется такое значение v_q случайной величины, что вероятность события $\xi \geq v_q$ равна $q/100$, т.е. $1 - F(v_q) = P(\xi \geq v_q) = q/100$.

Геометрически $q\%$ -ная точка — это значение случайной величины, при котором площадь криволинейной трапеции, ограниченной графиком плотности распределения, осью абсцисс и лежащая правее v_q , равна $q/100$.

На рис. 12.3 показана квантиль уровня 0,8 (20%-ная точка) для стандартного нормального распределения (на графиках плотности и функции распределения).

К понятию процентной точки близко понятие критической точки, широко используемое в задачах проверки гипотез. Критические точки для заданного распределения определяют границы, за пределы которых случайная величина выходит достаточно редко. Например, если нас интересуют большие положительные значения случайной величины ξ , то критическая точка $t_{кр}$ может быть определена из условия $P(\xi > t_{кр}) = \alpha$, где α мало. Если же нас интересуют значения, большие по абсолютной величине (положительные или отрицательные), то можно определить критическую точку из условия $P(|\xi| > t_{кр}) = \alpha$.



Рис. 12.3

Нижней критической границей $\underline{u}_\alpha$ и верхней критической границей $\bar{u}_\alpha$, соответствующими заданному уровню значимости α , называются значения случайной величины, для которых выполнены условия:

$$P(\xi < \underline{u}_\alpha) = F(\bar{u}_\alpha) = \alpha/2;$$

$$P(\underline{u}_\alpha \leq \xi < \bar{u}_\alpha) = F(\bar{u}_\alpha) - F(\underline{u}_\alpha) = 1 - \alpha;$$

$$P(\xi \geq \bar{u}_\alpha) = 1 - F(\bar{u}_\alpha) = \alpha/2.$$

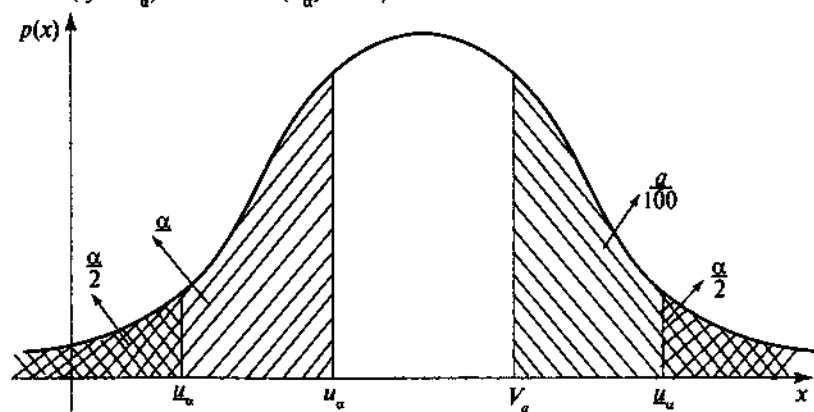


Рис. 12.4

Между критическими границами и квантилями для симметричного распределения существуют следующие соотношения $\underline{u}_\alpha = x_{\frac{\alpha}{2}}$; $\bar{u}_\alpha = x_{1-\frac{\alpha}{2}}$ (рис. 12.4).

Конкретные значения критических точек для различных распределений и уровней значимости можно найти в таблицах. Эмпирическими аналогами теоретических квантилей будут члены вариационного ряда.

12.3. Распределение хи-квадрат (закон Пирсона)

Распределением хи-квадрат с n степенями свободы (обозначается χ_n^2) называется распределение суммы квадратов n независимых случайных величин со стандартным нормальным распределением, т.е. $\chi_n^2 = \sum_{i=1}^n \xi_i^2$, если $\xi_i \in N(0, 1)$.

Такое же распределение будет иметь величина

$$\chi_n^2 = \sum_{i=1}^n \frac{(\xi_i - a)^2}{\sigma^2}, \text{ если } \xi_i \in N(a, \sigma^2).$$

Плотность распределения случайной величины χ_n^2 с n степенями свободы имеет вид:

$$p_{\chi^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Функция распределения:

$$F(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Основные числовые характеристики распределения χ_n^2 следующие:

математическое ожидание — $M\chi_n^2 = n$;

дисперсия — $D\chi_n^2 = 2n$;

$$\text{асимметрия} - \beta_{\chi^2} = \sqrt{\frac{8}{n}};$$

$$\text{эксцесс} - \gamma_{\chi^2} = 12/n.$$

На рис. 12.5 представлен график плотности распределения хи-квадрат с 5 степенями свободы.



Рис. 12.5

Плотность χ_n^2 распределения зависит от одного параметра n — числа степеней свободы. При $n \leq 2$ функция плотности убывает, а при $n > 2$ имеет единственный максимум в точке $X_{\text{mod}} = n - 2$. С ростом числа степеней свободы n распределение χ_n^2 приближается к нормальному закону распределения со средним n и дисперсией $2n$ (в смысле асимптотической нормальности). Общий вид графика представлен на рис. 12.6.

Заштрихованная область соответствует вероятности α и определяет *квантиль уровня α* распределения хи-квадрат.

Задача 4. Доказать асимптотическую эффективность несмещенной оценки генеральной дисперсии в случае нормального распределения.

Решение. Как известно (см. § 8 гл. 12), несмещенная оценка генеральной дисперсии и ее дисперсия выглядят следующим

$$\text{образом: } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad Ds^2 = \frac{2\sigma^4}{n-1}.$$

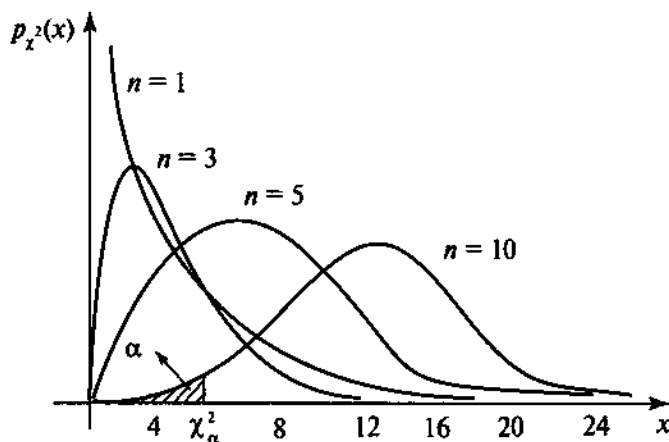


Рис. 12.6

Для того чтобы доказать эффективность оценки, необходимо найти информацию Фишера. Выпишем функцию плотности нормального распределения:

$$p_{\xi}(x, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \text{ и прологарифмируем ее:}$$

$$\ln p_{\xi}(x, \sigma^2) = \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x-a)^2}{2\sigma^2} = -\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{(x-a)^2}{2\sigma^2}.$$

Продифференцируем по параметру $\theta = \sigma^2$:

$$\frac{d \ln p_{\xi}(x, \theta)}{d\theta} = -\frac{1}{2\theta} + \frac{(x-a)^2}{2\theta^2}; \quad \left(\frac{d \ln p_{\xi}(x, \theta)}{d\theta} \right)^2 = \frac{1}{4\theta^4} ((x-a)^2 - \theta)^2.$$

Тогда информация Фишера равна:

$$\begin{aligned} I(\theta) &= \frac{1}{4\theta^4} M((\xi-a)^2 - \theta)^2 = \frac{1}{4\theta^4} D(\xi-a)^2 = \frac{1}{4\theta^4} \theta^2 D\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2 = \\ &= \frac{1}{4\theta^2} D\chi_1^2 = \frac{1}{2\theta^2} = \frac{1}{2\sigma^4}, \end{aligned}$$

так как случайная величина $\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2 = \chi_1^2$ имеет распределение хи-квадрат с одной степенью свободы, и тогда $D\chi_1^2 = 2$. Отсюда

получим $\Delta_n = \frac{2\sigma^4}{n}$. С другой стороны, $Ds^2 = \frac{2\sigma^4}{n-1}$, и в результате

$$\text{имеем } \frac{\Delta_n}{Ds^2} = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} \rightarrow 1 \text{ при } n \rightarrow \infty.$$

Таким образом, несмещенная оценка генеральной дисперсии является асимптотически эффективной.

12.4. Распределение Стьюдента

Пусть $n + 1$ случайных величин $\xi_0, \xi_1, \dots, \xi_n$ независимы и имеют стандартное нормальное распределение: $\xi_i \in N(0, 1)$. Пусть

$$\eta = \sqrt{\frac{1}{n} \sum_{i=1}^n \xi_i^2}. \text{ Случайная величина } t_n = \frac{\xi_0}{\eta} \text{ называется безразмер-}$$

ной дробью Стьюдента, а ее распределение — **распределением Стьюдента с n степенями свободы**.

Плотность распределения t_n имеет вид:

$$p_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Плотность распределения Стьюдента зависит от одного параметра — числа степеней свободы. С ростом числа степеней свободы n распределение Стьюдента сходится к стандартному нормальному распределению.

Основные числовые характеристики:

мода и медиана равны математическому ожиданию и равны нулю, т.е. распределение унимодально и симметрично относительно точки $x = 0$; математическое ожидание существует при $n > 1$ и тоже равно нулю.

дисперсия — $Dt_n = n/(n-2)$ и существует только при $n > 2$;

асимметрия — $\beta = 0$;

эксцесс — $\gamma_n = 6/(n-4)$.

На рис. 12.7 представлен график плотности распределения Стьюдента с тремя степенями свободы.

Общий вид кривой распределения представлен на рис. 12.8.



Рис. 12.7

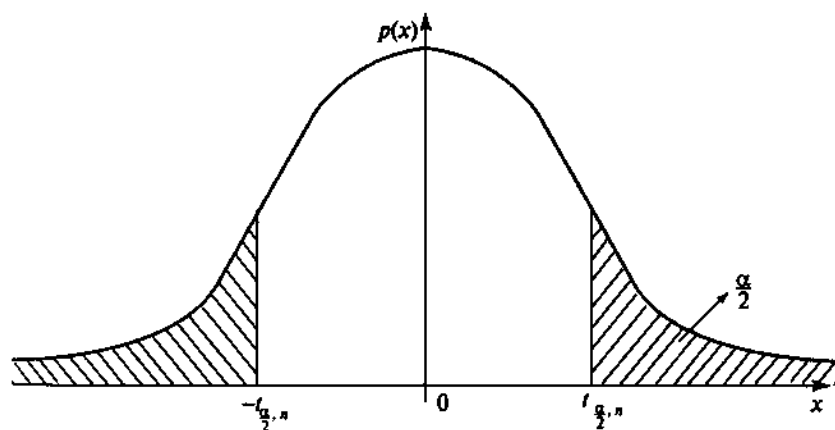


Рис. 12.8

Заштрихованным областям (слева и справа) соответствуют вероятности по $\alpha/2$, в сумме дающие вероятность α и определяющие *критические точки* распределения Стьюдента для двусторонней области.

12.5. Распределение Фишера

Если χ_n^2 и χ_m^2 — независимые случайные величины, распределенные по закону χ^2 с числами степеней свободы n и m соответственно, то случайная величина

$$F(n, m) = \frac{\chi_n^2/n}{\chi_m^2/m}$$

имеет распределение, которое называют **распределением Фишера—Снедекора** (F -распределением) с числами степеней свободы n и m , или **распределением дисперсионного отношения**.

Функция распределения и функция плотности имеют вид:

$$F_{n,m}(x) = \frac{\Gamma\left(\frac{n+m}{2}\right) n^{\frac{n}{2}} m^{\frac{m}{2}}}{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{m}{2}\right)} \int_0^x \frac{u^{\frac{n}{2}-1}}{(nu+m)^{\frac{n+m}{2}}} du;$$

$$p_{n,m}(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{m}{2}\right)} \cdot \frac{n^{\frac{n}{2}} m^{\frac{m}{2}} x^{\frac{n}{2}-1}}{(nx+m)^{\frac{n+m}{2}}}, \quad x > 0.$$

Распределение Фишера определяется двумя параметрами n и m , называемыми числами степеней свободы.

Основные числовые характеристики:

математическое ожидание $M_F = \frac{m}{m-2}$ существует только при $m > 2$;

дисперсия $D_F = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ существует только при $m > 4$;

мода $X_{\text{mod}} = \frac{m(n-2)}{n(m+2)}$ при $n > 1$;

асимметрия $\beta_F = \frac{(2n+m-2)\sqrt{8(m-4)}}{(m-6)\sqrt{n(n+m-2)}}$ при $m > 6$;

эксцесс $\gamma_F = \frac{3(m-6)(2+\frac{1}{2}\beta_F^2)}{m-8} - 3$ при $m > 8$.

Из этих формул следует, что F -распределение всегда имеет модальное значение, меньшее единицы, и среднее значение, большее единицы, а также положительную асимметрию.

На рис. 12.9 представлен график плотности распределения Фишера с числами степеней свободы $n = 10$ и $m = 15$.

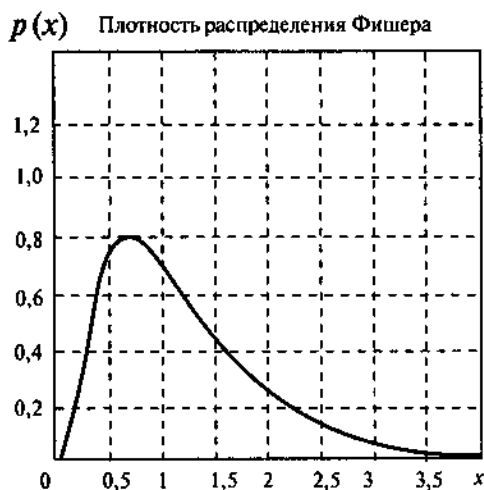


Рис. 12.9

Общий вид графика плотности распределения Фишера при различных значениях параметров представлен на рис. 12.10.

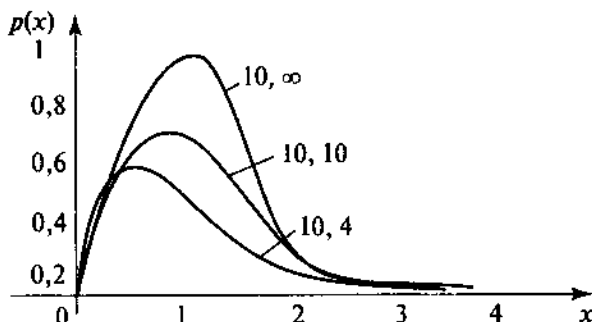


Рис. 12.10

12.6. Гамма-распределение

Двухпараметрический закон гамма-распределения случайной величины $\xi \in \gamma(\alpha, \lambda)$ описывается функцией плотности:

$$p_{\gamma}(x) = \begin{cases} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0 \end{cases}$$

где $\Gamma(\alpha)$ — гамма-функция Эйлера, $\alpha > 0$ — параметр «формы», $\lambda > 0$ — параметр «масштаба». Если $\lambda = 1$, случайная величина $\gamma(\alpha, 1)$ зависит от одного параметра α и подчинена **однопараметрическому закону гамма-распределения**.

Основные свойства гамма-распределения следующие:

1. Если случайная величина имеет гамма-распределение $\xi \in \gamma(\alpha, \lambda)$, то $c\xi \in \gamma(\alpha, \lambda/c)$, в частности $\lambda\xi \in \gamma(\alpha, 1)$.

Доказательство. Действительно, пусть $\eta = c\xi$. Найдем функцию распределения $F_{\eta}(y)$. По определению имеем

$$\begin{aligned} F_{\eta}(y) &= P(\eta < y) = P(c\xi < y) = P\left(\xi < \frac{y}{c}\right) = F_{\xi}\left(\frac{y}{c}\right) = \int_0^{\frac{y}{c}} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \\ &= \left| cx = u, cdx = du \right| = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{\frac{y}{c}} \left(\frac{u}{c}\right)^{\alpha-1} e^{-\frac{\lambda u}{c}} \frac{1}{c} du = \frac{\left(\frac{\lambda}{c}\right)^{\alpha}}{\Gamma(\alpha)} \int_0^y u^{\alpha-1} e^{-\left(\frac{\lambda}{c}\right)u} du. \end{aligned}$$

Следовательно, функция плотности случайной величины $\eta = c\xi$ имеет вид

$$p_{\eta}(y) = \begin{cases} \frac{\left(\frac{\lambda}{c}\right)^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\lambda}{c}y}, & y \geq 0; \\ 0, & y < 0. \end{cases}$$

Отсюда следует, что если $c = \lambda$, то $\lambda\xi \in \gamma(\alpha, 1)$.

2. Сумма любого числа независимых гамма-распределенных случайных величин с одинаковым параметром масштаба λ и параметрами формы $\alpha_1, \alpha_2, \dots, \alpha_n$ также подчиняется гамма-распределению с параметрами $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_n$ и λ .

Доказательство. Рассмотрим случай, когда ξ и η являются однопараметрическими гамма-распределенными случайными величинами с параметрами α и β соответственно ($\alpha > 0, \beta > 0$),

т.е. $\lambda = 1$. Покажем, что если ξ и η независимые случайные величины, то случайная величина $\tau = \xi + \eta$ подчиняется гамма-распределению с параметром $\alpha + \beta$.

Пусть $\xi \in \gamma(\alpha, 1)$ и $\eta \in \gamma(\beta, 1)$. Функции плотностей распределения случайных величин ξ и η соответственно равны:

$$p_{\xi}(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \quad p_{\eta}(x) = \frac{1}{\Gamma(\beta)} x^{\beta-1} e^{-x}, \quad x > 0.$$

Поскольку ξ и η независимы, совместная плотность распределения равна произведению функций плотностей сомножителей: $p(x, y) = p_{\xi}(x)p_{\eta}(y)$, и функция плотности суммы вычисляется с помощью свертки

$$p_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx = \int_{-\infty}^{+\infty} p_{\xi}(x) p_{\eta}(z-x) dx.$$

Подставляя в последнюю формулу выражения для плотностей составляющих, получим

$$p_{\xi+\eta}(z) = \int_{-\infty}^{+\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} I_{[0, \infty)}(x) \frac{1}{\Gamma(\beta)} (z-x)^{\beta-1} e^{-(z-x)} I_{[0, \infty)}(z-x) dx,$$

где $I_{[0, \infty)}$ — индикаторная функция множества $D \subseteq \mathbb{R}^1$, определяемая соотношением

$$I_D(x) = \begin{cases} 1, & x \in D; \\ 0, & x \notin D. \end{cases}$$

Отсюда следует, что

$$I_{[0, \infty)}(z-x) = \begin{cases} 1, & 0 \leq z-x < +\infty, \\ 0, & z-x < 0 \end{cases} = \begin{cases} 1, & x \leq z, \\ 0, & x > z. \end{cases}$$

Таким образом, получаем

$$\begin{aligned} p_{\xi+\eta}(z) &= \int_0^{+\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \frac{1}{\Gamma(\beta)} (z-x)^{\beta-1} e^{-z+x} I_{[0, z]}(x) dx = \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} e^{-z} dx = \left| \frac{x}{z} = u, dx = z du \right| = \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 (zu)^{\alpha-1} (z-zu)^{\beta-1} e^{-z} z du = \frac{z^{\alpha+\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du = \\ &= \frac{z^{\alpha+\beta-1} e^{-z}}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \end{aligned}$$

Поэтому функция плотности распределения суммы $\xi + \eta$, равная

$$p_{\xi+\eta}(z) = \frac{1}{\Gamma(\alpha+\beta)} z^{(\alpha+\beta)-1} e^{-z},$$

соответствует гамма-распределению с параметром $\alpha + \beta$, т.е. $\xi + \eta \in \gamma(\alpha + \beta, 1)$.

Из первого свойства следует, что $\xi_1 = \frac{1}{\lambda} \xi \in \gamma(\alpha, \lambda)$ и $\eta_1 = \frac{1}{\lambda} \eta \in \gamma(\beta, \lambda)$. Тогда в силу доказанного получаем, что случайная величина $\xi_1 + \eta_1 = \frac{1}{\lambda} (\xi + \eta) \in \gamma(\alpha + \beta, \lambda)$.

По индукции теперь можно доказать, что сумма любого числа независимых гамма-распределенных случайных величин с одинаковым параметром масштаба λ и параметрами формы $\alpha_1, \alpha_2, \dots, \alpha_n$ также подчиняется гамма-распределению с параметрами $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_n$ и λ .

3. Распределение χ_n^2 является частным случаем гамма-распределения с параметрами $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$, т.е. $\chi_n^2 = \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Доказательство. Достаточно подставить параметры $\alpha = \frac{n}{2}$, $\lambda = \frac{1}{2}$ в формулу плотности гамма-распределения и сравнить с формулой плотности распределения хи-квадрат.

Основные числовые характеристики случайной величины $\gamma(\alpha, \lambda)$ следующие:

1) математическое ожидание — $M_\gamma = \frac{\alpha}{\lambda}$;

2) мода — $X_{\text{mod}} = \frac{\alpha-1}{\lambda}$ при $\alpha \geq 1$;

3) дисперсия — $D_\gamma = \frac{\alpha}{\lambda^2}$;

4) асимметрия — $\beta_\gamma = \frac{2}{\sqrt{\alpha}}$;

5) эксцесс — $\gamma_\gamma = \frac{6}{\alpha}$.

Гамма-распределение иногда используют при моделировании реальных ситуаций. Например, с его помощью описывается распределение доходов или сбережений населения в некоторых определенных случаях.

Задача 5. Пусть случайная величина ξ имеет плотность распределения вида

$$p(x, \alpha) = \begin{cases} Cx^{\alpha-1}e^{-x}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Найти константу C , математическое ожидание $M\xi$ и дисперсию $D\xi$.

Решение. По свойству плотности распределения имеем

$$\int_{-\infty}^{+\infty} p(x) dx = C \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = C\Gamma(\alpha) = 1, \quad \alpha > 0.$$

Отсюда получаем, что $C = \frac{1}{\Gamma(\alpha)}$.

Вычислим числовые характеристики распределения:

$$M\xi = \int_{-\infty}^{+\infty} xp(x) dx = \int_0^{+\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha} e^{-x} dx = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^{(\alpha+1)-1} e^{-x} dx = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha.$$

Для нахождения дисперсии вычислим второй начальный момент:

$$\begin{aligned} M\xi^2 &= \int_{-\infty}^{+\infty} x^2 p(x) dx = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^{(\alpha+2)-1} e^{-x} dx = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \\ &= \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)} = \alpha(\alpha+1). \end{aligned}$$

Тогда дисперсия равна:

$$D\xi = M\xi^2 - (M\xi)^2 = \alpha(\alpha+1) - \alpha^2 = \alpha.$$

Общий вид графика плотности гамма-распределения при различных значениях параметров представлен на рис. 12.11.

На рис. 12.12 представлен график плотности гамма-распределения при $\alpha = 3$ и $\lambda = 1$.

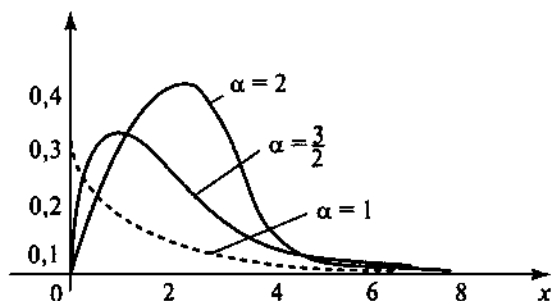


Рис. 12.11



Рис. 12.12

Задача 6. Пусть случайная величина η подчиняется стандартному нормальному закону распределения: $\eta \in N(0, 1)$. Найти функцию плотности случайной величины $\xi = \frac{\eta^2}{2}$. К какому параметрическому семейству распределений относится ξ ?

Решение. Поскольку случайная величина $\eta \in N(0, 1)$, ее плотностью распределения будет функция $p_\eta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Найдем функцию распределения случайной величины ξ .

$$F_\xi(x) = P(\xi < x) = P\left(\frac{\eta^2}{2} < x\right) = P(\eta^2 < 2x) =$$

$$\begin{aligned}
 &= \begin{cases} P(|\eta| < \sqrt{2x}), & x \geq 0, \\ 0, & x < 0 \end{cases} = P(-\sqrt{2x} < \eta < \sqrt{2x}) = \int_{-\sqrt{2x}}^{\sqrt{2x}} p_{\eta}(t) dt = \\
 &= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2x}} \exp\left(-\frac{t^2}{2}\right) dt = \left| \frac{t^2}{2} = u; \quad t^2 = 2u; \quad 2t dt = 2du; \quad dt = \frac{du}{t} = \frac{du}{\sqrt{2u}} \right| = \\
 &= \frac{2}{\sqrt{2\pi}} \int_0^x e^{-u} \frac{du}{\sqrt{2u}} = \frac{1}{\sqrt{\pi}} \int_0^x u^{-\frac{1}{2}} e^{-u} du.
 \end{aligned}$$

Из полученного равенства и функции плотности распределения получаем плотность распределения случайной величины ξ в виде

$$p_{\xi}(x) = \frac{1}{\sqrt{\pi}} x^{\frac{1}{2}} e^{-x},$$

а эта функция является плотностью однопараметрического гамма-распределения с параметром, равным $\alpha = 1/2$, т.е. $\gamma(1/2, 1) = \gamma(1/2)$.

Задача 7. Пусть распределение случайной величины ξ задано плотностью

$$p(x, \alpha, \lambda) = \begin{cases} Cx^{\alpha-1}e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Найти константу C , математическое ожидание и дисперсию ξ .

Решение. Поскольку $\int_{-\infty}^{+\infty} p(x) dx = 1$, имеем $\int_0^{+\infty} Cx^{\alpha-1}e^{-\lambda x} dx = 1$.

Сделаем замену, положив $\lambda x = u$, $\lambda dx = du$, получим

$$\frac{C}{\lambda} \int_0^{+\infty} \left(\frac{u}{\lambda}\right)^{\alpha-1} e^{-u} du = \frac{C}{\lambda^{\alpha}} \int_0^{+\infty} u^{\alpha-1} e^{-u} du = \frac{C}{\lambda^{\alpha}} \Gamma(\alpha) = 1.$$

Отсюда следует, что $C = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}$. Вычислим теперь математическое ожидание:

$$\begin{aligned}
 M\xi &= \int_{-\infty}^{+\infty} xp(x) dx = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha} e^{-\lambda x} dx = \left| \lambda x = u, \quad dx = \frac{du}{\lambda} \right| = \\
 &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \cdot \frac{1}{\lambda^{\alpha+1}} \int_0^{+\infty} u^{\alpha} e^{-u} du = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} = \frac{\alpha}{\lambda}.
 \end{aligned}$$

Для вычисления дисперсии аналогично находим второй начальный момент:

$$\begin{aligned} M\xi^2 &= \int_{-\infty}^{+\infty} x^2 p(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha+1} e^{-\lambda x} dx = \left| \lambda x = u, \quad dx = \frac{du}{\lambda} \right| = \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{\lambda^{\alpha+2}} \int_0^{+\infty} u^{\alpha+1} e^{-u} du = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2} = \frac{\alpha(\alpha+1)}{\lambda^2}. \end{aligned}$$

Тогда дисперсию находим по формуле

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.$$

12.7. Бета-распределение

Случайная величина $\xi \in \beta(a, b)$, подчиняющаяся закону бета-распределения (β -распределение) с параметрами $a > 0$ и $b > 0$, имеет плотность распределения

$$p_\xi(x) = \begin{cases} \frac{\Gamma(a+b)x^{a-1}(1-x)^{b-1}}{\Gamma(a)\Gamma(b)}, & 0 \leq x \leq 1, \\ 0, & x \notin [0, 1]. \end{cases}$$

Основные свойства бета-распределения следующие:

1. Если $\xi_1 \in \gamma(\alpha_1, \lambda)$ и $\xi_2 \in \gamma(\alpha_2, \lambda)$ — две независимые гамма-распределенные случайные величины, то отношение $\eta = \frac{\xi_1}{\xi_1 + \xi_2}$ имеет бета-распределение с параметрами α_1 и α_2 : $\eta \in \beta(\alpha_1, \alpha_2)$.
2. Случайная величина $\beta(1, 1)$ распределена равномерно на отрезке $[0, 1]$.
3. Функция распределения квадрата студентовской величины t_m^2 связана с функцией распределения случайной величины β соотношением

$$F_{t_m^2}(x) = F_{\beta\left(\frac{1}{2}, \frac{m}{2}\right)}\left(\frac{x}{m+x}\right).$$

4. Функция распределения случайной величины $F(n, m)$ связана с функцией распределения случайной величины β соотношением

$$F_{n,m}(x) = F_{\beta\left(\frac{n}{2}, \frac{m}{2}\right)}\left(\frac{nx}{m+nx}\right).$$

5. Между функцией распределения случайной величины β и биномиальным распределением существуют соотношения

$$F_{\beta(m, n-m+1)}(x) = \sum_{k=m}^n C_n^k x^k (1-x)^{n-k}.$$

6. Имеет место симметрия: $p_{\beta(a,b)}(x) = p_{\beta(b,a)}(1-x)$ и $F_{\beta(a,b)}(x) = 1 - F_{\beta(b,a)}(1-x)$.

Основные числовые характеристики случайной величины $\beta(a, b)$:

1) математическое ожидание (среднее) — $M\beta(a, b) = \frac{a}{a+b}$;

2) мода — $X_{\text{mod}} = \frac{a-1}{a+b-2}$ при $a > 1, b > 1$;

3) дисперсия — $D_{\beta(a,b)} = \frac{ab}{(a+b)^2(a+b+1)}$;

4) асимметрия — $\beta_p = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$;

5) эксцесс — $\gamma_p = \frac{3(a+b+1)[2(a+b)^2 + ab(a+b-6)]}{ab(a+b+2)(a+b+3)} - 3$.

Бета-распределение используется для описания некоторых реальных распределений, сосредоточенных на отрезке $[0,1]$, например, для описания распределения величин субъективных вероятностей, полученных в ходе экспертного опроса.

На рис. 12.13 представлен график плотности бета-распределения с $a = 3, b = 5$.



Рис. 12.13

12.8. Приложения распределений в математической статистике.

Теорема Фишера

При построении экономико-математической модели особенно важным является исследование свойств конечной выборки при фиксированном n , когда выборка $x_1, x_2, \dots, x_n$ сделана из нормальной генеральной совокупности $N(a, \sigma^2)$. Справедлива теорема Фишера, описывающая доасимптотические свойства оценок. Предварительно рассмотрим следующие леммы.

Лемма 1. Если $x_1, x_2, \dots, x_n$ — независимые одинаково нормально распределенные случайные величины: $x_i \in N(0, \sigma^2)$, и $y_1, y_2, \dots, y_n$ получены из $x_1, x_2, \dots, x_n$ ортогональным преобразованием $y_i = \sum_{k=1}^n C_{ik} x_k$, $i = 1, 2, \dots, n$, где $C = (c_{ij})$ — ортогональная матрица, то $y_1, y_2, \dots, y_n$ независимые, одинаково нормально распределенные случайные величины: $y_i \in N(0, \sigma^2)$.

Доказательство. Пусть $x_i \in N(0, \sigma^2)$ и $y_i = \sum_{k=1}^n c_{ki} x_k$, $i = 1, 2, \dots, n$, где матрица C ортогональна. Тогда существует обратная матрица C^{-1} тоже ортогональная и, следовательно $|\det C^{-1}| = 1$. В векторной форме это преобразование можно записать так: $Y = CX$ и обратное преобразование $X = C^{-1}Y = X(Y)$. Функция плотности случайного вектора Y будет равна $p_y(y_1, y_2, \dots, y_n) = p_x(x_1(Y), x_2(Y), \dots, x_n(Y)) |\det J|$, но так как преобразование $X \rightarrow Y$ линейное, легко видеть, что якобиан $J = C^{-1}$. Кроме того, квадратичная форма инвариантна относительно ортогонального преобразования:

$$\begin{aligned} y_1^2 + y_2^2 + \dots + y_n^2 &= Y^T Y = (CX)^T (CX) = X^T (C^T C) X = X^T X = \\ &= x_1^2 + x_2^2 + \dots + x_n^2. \end{aligned}$$

Поэтому

$$p_y(y_1, y_2, \dots, y_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\frac{-1}{2\sigma^2}(x_1^2(y) + x_2^2(y) + \dots + x_n^2(y))} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-y_i^2}{2\sigma^2}},$$

откуда следует независимость и нормальное распределение $y_i \in N(0, \sigma^2)$.

Лемма 2 (Фишера). Пусть задано некоторое число $p < n$ линейных функций $y_1, y_2, \dots, y_p$ от $x_1, x_2, \dots, x_n$: $y_i = c_{i1}x_1 + c_{i2}x_2 + \dots + c_{in}x_n$, $i = 1, 2, \dots, p$, где строки матрицы $C = (c_{ij})$ удовлетворяют условию ортогональности $\sum_{j=1}^n c_{ij}c_{kj} = \delta_{ik} = \begin{cases} 1, & i=k \\ 0, & i \neq k \end{cases}$. Если $x_1, x_2, \dots, x_n$ независимые одинаково нормально распределенные случайные величины $x_i \in N(0, \sigma^2)$, то случайная величина $\xi = \sum_{i=1}^n x_i^2 - \sum_{i=1}^p y_i^2$ распределена по закону $\sigma^2 \chi_{n-p}^2$ с $n - p$ степенями свободы, причем ξ и $y_1^2 + y_2^2 + \dots + y_p^2$ взаимно независимы.

Доказательство. Известно, что для прямоугольной матрицы $C_{(p \times n)}$ с ортогональными строками можно подобрать дополнительно $n - p$ строк и дополнить ее до квадратной ортогональной матрицы $C_{(n \times n)}$. Если к вектору X применить преобразование CX , то получим вектор Y , у которого первые p компонентов совпадают с заданными случайными величинами $y_1, y_2, \dots, y_p$, и так как квадратичная форма инвариантна относительно ортогонального преобразования, получим $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$. Отсюда следует, что $\xi = \sum_{i=1}^n x_i^2 - \sum_{i=1}^p y_i^2 = \sum_{i=p+1}^n y_i^2$ и ξ как функция от $y_{p+1}, \dots, y_n$ не зависит от $y_1, y_2, \dots, y_p$, т.е. не зависит от суммы $\sum_{i=1}^p y_i^2$.

В силу леммы 1 случайные величины $y_1, y_2, \dots, y_p$ имеют нормальное распределение $y_i \in N(0, \sigma^2)$, а сумма квадратов нормально распределенных случайных величин имеет χ^2 -распределение:

$\sum_{i=p+1}^n \left(\frac{y_i}{\sigma}\right)^2 \in \chi_{n-p}^2$, следовательно, $\sum_{i=p+1}^n y_i^2 \in \sigma^2 \chi_{n-p}^2$. Лемма доказана.

► **Теорема Фишера.** Пусть $\bar{x}$ и s^2 — соответственно выборочное среднее и выборочная дисперсия, построенные по выборке $x_1, x_2, \dots, x_n$ из нормальной генеральной совокупности $N(a, \sigma^2)$. Тогда при любом фиксированном объеме выборки n их совместный закон распределения описывается следующим образом:

1) $\bar{x}$ распределено по нормальному закону $\bar{x} \in N(a, \sigma^2/n)$;

2) статистика $\frac{(n-1)s^2}{\sigma^2}$ распределена по закону χ^2 с $n - 1$ степенями свободы;

3) случайные величины $\bar{x}$ и s^2 независимы.

Доказательство. Пусть $x_1, x_2, \dots, x_n$ — случайная выборка из нормальной генеральной совокупности $N(a, \sigma^2)$. Нормальность распределения выборочного среднего $\bar{x}$ вытекает из факта нормального распределения любой линейной комбинации независимых нормально распределенных случайных величин.

Введем случайную величину $z_i = x_i - a$. Тогда $Mz_i = 0$, $Dz_i = Dx_i = \sigma^2$. Отсюда следует, что $s_z^2 = s_x^2$, т.е. без ограничения общности можно полагать $a = 0$.

Итак, $z_i = x_i - a \in N(0, \sigma^2)$ и $s_x^2 = s_z^2 = \frac{1}{n-1} \left(\sum_{i=1}^n z_i^2 - n\bar{z}^2 \right)$. Так как $a = 0$, имеем

$$\bar{x}^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 = \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \right)^2$$

$$\text{Отсюда получим } n\bar{x}^2 = \left(\frac{z_1}{\sqrt{n}} + \frac{z_2}{\sqrt{n}} + \dots + \frac{z_n}{\sqrt{n}} \right)^2.$$

Обозначим $y_1 = \sqrt{n} \cdot \bar{z} = \frac{z_1}{\sqrt{n}} + \frac{z_2}{\sqrt{n}} + \dots + \frac{z_n}{\sqrt{n}}$, т.е. случайная величина y_1 представлена в виде линейной функции

$$y_1 = C_{11}z_1 + C_{12}z_2 + \dots + C_{1n}z_n,$$

где $C_{1i} = \frac{1}{\sqrt{n}}$, а $C_{11}^2 + C_{12}^2 + \dots + C_{1n}^2 = 1$, т.е. строка $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$ удовлетворяет условию ортогональности. Применяя лемму Фишера к $y_1 = \sqrt{n} \cdot \bar{z}$ при $p = 1$, получаем, что $y_1^2 = n\bar{z}^2$ и случайная величина $\xi = (n-1)s_z^2 = \sum_{i=1}^n z_i^2 - y_1^2$ в случае выборки из нормальной генеральной совокупности независимы, причем $y_1 = \sqrt{n} \cdot \bar{z} \in N(0, \sigma^2)$, и $(n-1)s^2 \in \sigma^2 \chi_{n-1}^2$.

Таким образом доказано, что $\bar{x} \in N(0, \frac{\sigma^2}{n})$, $\frac{(n-1)s^2}{\sigma^2} \in \chi_{n-1}^2$.

Следствие 1. Дисперсия исправленной выборочной дисперсии s^2 , построенной по случайной выборке объема n из нормальной генеральной совокупности $N(a, \sigma^2)$, определяются формулой

$$Ds^2 = \frac{2\sigma^4}{n-1}.$$

Поскольку $D\chi_{n-1}^2 = 2(n-1)$, то $D\left(\frac{(n-1)s^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4} Ds^2 = 2(n-1)$.

Поэтому $Ds^2 = \frac{2\sigma^4}{n-1}$.

Следствие 2. Если $\bar{x}$ и s^2 соответственно выборочное среднее и дисперсия, построенные по выборке из нормальной генеральной совокупности $N(a, \sigma^2)$, то статистика

$$T = \frac{\bar{x} - a}{s} \sqrt{n}$$

распределена по закону Стьюдента с $n - 1$ степенями свободы.

Для доказательства достаточно представить статистику T в виде

$$T = \frac{\frac{\bar{x} - a}{\sigma} \sqrt{n}}{\sqrt{\frac{ns^2}{(n-1)\sigma^2}}},$$

и, так как $\frac{\bar{x} - a}{\sigma} \sqrt{n} \in N(0, 1)$, $\frac{(n-1)s^2}{\sigma^2} \in \chi_{n-1}^2$, по определению, статистика T имеет распределение Стьюдента с $n - 1$ степенями свободы.

Следствие 3. Пусть случайная выборка $x_1, x_2, \dots, x_n$ произведена из нормальной генеральной совокупности $N(a_1, \sigma_1^2)$, а выборка $y_1, y_2, \dots, y_m$ — из генеральной совокупности $N(a_2, \sigma_2^2)$, и эти выборки независимы. Тогда

$$\bar{x} - \bar{y} \in N\left(a_1 - a_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

В частности, если обе выборки взяты из одной и той же генеральной совокупности ($a_1 = a_2, \sigma_1 = \sigma_2$), то

$$\bar{x} - \bar{y} \in N\left(0, \sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

Поскольку линейная комбинация независимых нормально распределенных случайных величин распределена нормально, то для доказательства достаточно вычислить математическое ожидание и дисперсию разности независимых случайных величин x и y .

Следствие 4. Пусть случайная выборка $x_1, x_2, \dots, x_n$ произведена из нормальной генеральной совокупности $N(a_1, \sigma_1^2)$, а выборка $y_1, y_2, \dots, y_m$ из генеральной совокупности $N(a_2, \sigma_2^2)$, и эти выборки независимы. Тогда при одинаковых (возможно, неизвестных) дисперсиях $\sigma_1^2 = \sigma_2^2 = \sigma^2$ случайная величина

$$T = \frac{(\bar{x} - \bar{y}) - (a_1 - a_2)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{\frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}}},$$

где s_1^2 и s_2^2 — исправленные оценки дисперсий, имеет распределение Стьюдента с $(n + m - 2)$ степенями свободы.

Из следствия 3 вытекает, что

$$\eta_1 = \frac{(\bar{x} - \bar{y}) - (a_1 - a_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \in N(0, 1).$$

По теореме Фишера $\frac{(n-1)s_1^2}{\sigma^2} \in \chi_{n-1}^2$, $\frac{(m-1)s_2^2}{\sigma^2} \in \chi_{m-1}^2$, причем s_1^2, s_2^2 независимы.

Следовательно,

$$\eta_2 = \frac{(n-1)s_1^2}{\sigma^2} + \frac{(m-1)s_2^2}{\sigma^2} \in \chi_{n+m-2}^2,$$

так как по свойству χ^2 -распределения сумма $\chi_{n-1}^2 + \chi_{m-1}^2$ распределена тоже по закону χ^2 с $n + m - 2$ степенями свободы. Тогда

случайная величина $T = \frac{\eta_1}{\sqrt{\frac{\eta_2}{n+m-2}}}$ по определению распределена по закону Стьюдента с $n + m - 2$ степенями свободы.

Следствие 5. Пусть случайная выборка $x_1, x_2, \dots, x_n$ произведена из нормальной генеральной совокупности $N(a_1, \sigma_1^2)$, а выборка $y_1, y_2, \dots, y_m$ — из генеральной совокупности $N(a_2, \sigma_2^2)$, и эти выборки независимы. Тогда при одинаковых (возможно, неизвестных) дисперсиях $\sigma_1^2 = \sigma_2^2 = \sigma^2$ случайная величина

$$F = \frac{s_1^2}{s_2^2}$$

распределена по закону Фишера с $n - 1$ и $m - 1$ степенями свободы.

Доказательство вытекает из того факта, что случайные величины

$$\frac{(n-1)s_1^2}{\sigma^2} \in \chi_{n-1}^2, \quad \frac{(m-1)s_2^2}{\sigma^2} \in \chi_{m-1}^2$$

распределены по закону χ^2 и независимы, так как независимы соответствующие выборки.

Задачи для самостоятельного решения

- Вычислить значения: а) $\Gamma(7/2)$; б) $\Gamma(9/2)$; в) $B(5/2, 7/2)$; г) $B(3/2, 9/2)$.
- Доказать свойства 1–3 бета-функции.
- Доказать, что сумма двух независимых χ^2 -распределенных случайных величин подчиняется распределению χ^2 , а именно $\chi_{n+m}^2 = \chi_n^2 + \chi_m^2$.
- Найти математическое ожидание и дисперсию распределения χ_n^2 .
- Доказать, что распределение χ_n^2 сходится к нормальному закону со средним n и дисперсией $2n$ при $n \rightarrow \infty$ (в смысле асимптотической нормальности).
- Пусть $\xi_0, \xi_1, \dots, \xi_n$ — независимые, нормально распределенные случайные величины, $\xi_i \in N(0, \sigma^2)$, $i = 0, 1, 2, \dots, n$ и $\eta = \sqrt{\frac{1}{n} \sum_{i=1}^n \xi_i^2}$.
Найти распределение случайной величины $t = \frac{\xi_0}{\eta}$.
- Найти дисперсию распределения Стьюдента при $n > 2$.
- Найти эксцесс распределения Стьюдента при $n > 4$.
- Пусть случайная величина ξ имеет распределение Стьюдента с n степенями свободы. Какое распределение имеет случайная величина $\eta = \xi^2$?
- Доказать, что распределение Стьюдента с ростом числа степеней свободы ($n \rightarrow \infty$) сходится к стандартному нормальному распределению.
- Найти моду гамма-распределения $\gamma(\alpha, \lambda)$.

12. Доказать, что если $\xi_1 \in \gamma(\alpha_1, \lambda)$ и $\xi_2 \in \gamma(\alpha_2, \lambda)$ — две независимые гамма-распределенные случайные величины, то отношение $\eta = \frac{\xi_1}{\xi_1 + \xi_2}$ имеет бета-распределение с параметрами α_1 и α_2 ; $\eta \in \beta(\alpha_1, \alpha_2)$.
13. Найти математическое ожидание и дисперсию бета-распределения $\beta(a, b)$.
14. Найти моду бета-распределения $\beta(a, b)$.
15. Пусть $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ — независимые, нормально распределенные случайные величины — $\xi_i \in N(0, \sigma^2)$; $\eta_j \in N(0, \sigma^2)$.

Найти закон распределения случайной величины $\gamma = \frac{\frac{1}{n} \sum_{i=1}^n \xi_i^2}{\frac{1}{m} \sum_{j=1}^m \eta_j^2}$.

16. Определить, к какому распределению сходится распределение Фишера $F(n, m)$ при $m \rightarrow \infty$.
17. Найти математическое ожидание, моду и дисперсию распределения Фишера.
18. Пусть даны две независимые случайные величины $\xi, \eta \in \gamma(n; \lambda)$.
Найти распределение случайной величины $\frac{\xi}{\eta}$.

19. Доказать, что в случае нормального распределения $N(a, \sigma^2)$ не-исправленная выборочная дисперсия имеет дисперсию

$$D\hat{\sigma}^2 = \frac{2\sigma^4(n-1)}{n^2}.$$

20. Пусть $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m$ — независимые случайные выборки объема n и m из нормально распределенной генеральной совокупности $N(a, \sigma^2)$ с исправленными выборочными дисперсиями s_x^2, s_y^2 . Доказать, что оценка

$$s_{x,y}^2 = \frac{1}{n+m-2} [(n-1)s_x^2 + (m-1)s_y^2]$$

является эффективной в классе всех линейных несмещенных оценок параметра σ^2 , построенных по наблюдениям s_x^2, s_y^2 .

21. Пусть имеется M независимых случайных выборок объемов $n_1, n_2, \dots, n_M$ из нормально распределенной генеральной сово-

купности $N(a, \sigma^2)$ с исправленными выборочными дисперсиями $s_1^2, s_2^2, \dots, s_M^2$. Построить эффективную оценку (в классе всех линейных несмещенных оценок) для дисперсии σ^2 .

22. Доказать, что смещенная выборочная дисперсия нормального распределения является асимптотически эффективной (согласно определению для асимптотически нормальных оценок).

23. Для показательного распределения с плотностью $p_\lambda(x, \lambda) = \lambda e^{-\lambda x}$, $x \geq 0$ доказать несмещенность и асимптотическую эффективность оценки $\hat{\lambda} = \frac{n-1}{nx}$ параметра λ .

24. Для оценивания параметра σ нормального распределения $N(0, \sigma^2)$ намереваются использовать оценку вида $\sigma^* = \frac{c}{n} \sum_{i=1}^n |x_i|$.

Найти константу c , при которой оценка σ^* является несмещенной. Вычислить ее эффективность.

25. Для оценивания параметра σ нормального распределения $N(0, \sigma^2)$ намереваются использовать оценку вида $\sigma = c_n s$, где s — исправленное выборочное среднее квадратическое отклонение. Найти последовательность чисел c_n , при которых оценка σ является несмещенной. Проверить ее на асимптотическую эффективность.

Глава 13

МЕТОДЫ ПОСТРОЕНИЯ ОЦЕНОК

13.1. Метод моментов

Метод моментов, предложенный английским статистиком Карлом Пирсоном в 1894 г., заключается в приравнении определенного числа выборочных моментов к соответствующим теоретическим, которые являются функциями неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$. Рассматривая количество моментов, равное числу k неизвестных параметров, подлежащих определению, и решая полученные уравнения относительно этих параметров, получаем искомые оценки. Иначе говоря, оценки параметров $\theta_1, \theta_2, \dots, \theta_k$ являются решениями систем уравнений $\alpha_i(\theta_1, \theta_2, \dots, \theta_k) = \hat{\alpha}_i$ или $\mu_i(\theta_1, \theta_2, \dots, \theta_k) = \hat{\mu}_i$ для некоторых $i = i_1, i_2, \dots, i_k$.

Метод моментов содержит неопределенность, поскольку можно получить уравнения для неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$, используя как начальные, так и центральные моменты, а также некоторые их модификации типа асимметрии или эксцесса.

Пример. Функция

$$p(x) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & x > 0; \\ 0, & x \leq 0 \end{cases}$$

задает плотность распределения Рэлея (см. § 6.4). Требуется оценить параметр θ по выборке $x_1, x_2, \dots, x_n$.

Найдем оценку параметра θ , приравнявая начальные выборочные и теоретические моменты. Первый начальный момент

имеет вид: $a_1 = \sqrt{\pi\theta/2}$. Приравнявая, получаем первую оценку параметра: $\sqrt{\pi\theta_n/2} = \bar{x}$, откуда $\hat{\theta}_n = \frac{2\bar{x}^2}{\pi}$.

Приравнявая вторые начальные моменты, можем получить другую оценку: $\hat{\theta}_n = \frac{1}{2n} \sum_{i=1}^n x_i^2$, а из уравнения, которое получится при использовании второго центрального момента (дисперсии), — третью оценку: $\hat{\theta}_n = \frac{2s^2}{4 - \pi}$.

Часто полагают, что для нахождения оценки одного параметра следует брать первый момент, для двух — первые два момента и т.п. По возможности действительно имеет смысл поступать так, поскольку это проще всего. Однако такой подход годится не всегда. Он не проходит, например, если некоторые моменты равны нулю или не зависят от нужных параметров.

В общем случае система уравнений для моментов может не иметь решения в элементарных функциях (и тогда можно искать решение приближенными методами) или вообще оказаться неразрешимой (несовместной).

Оценки, полученные методом моментов, часто оказываются смещенными. К достоинствам метода моментов следует отнести его простую вычислительную реализацию, а также то, что оценки являются функциями от выборочных моментов.

В силу теоремы Слуцкого любая непрерывная функция от выборочных моментов функции сходится по вероятности к постоянной, получаемой подстановкой в эту функцию теоретических моментов, если они существуют и если получаемая таким образом постоянная конечна. Для определенности рассмотрим функцию $H(\alpha_1, \alpha_2)$ от двух моментов (начальных или центральных), хотя ее можно обобщить на любое конечное число аргументов, в том числе и на случай, когда H зависит только от одного аргумента.

► **Теорема 1. (Крамера).** Пусть в некоторой окрестности точки (α_1, α_2) функция $H(\alpha_1, \alpha_2)$ непрерывна и имеет непрерывные частные производные первого и второго порядка:

$$H_1 = \frac{\partial H(\alpha_1, \alpha_2)}{\partial \alpha_1}, \quad H_2 = \frac{\partial H(\alpha_1, \alpha_2)}{\partial \alpha_2} \quad \text{и} \quad H_0 = H(\alpha_1, \alpha_2);$$

Тогда для любой выборки, по которой найдены оценки $\hat{\alpha}_1, \hat{\alpha}_2$, случайная величина $H(\hat{\alpha}_1, \hat{\alpha}_2)$ асимптотически нормальна при $n \rightarrow \infty$ со следующими параметрами:

$$H(\hat{\alpha}_1, \hat{\alpha}_2) \xrightarrow{d} N(H_0, D\hat{\alpha}_1 H_1^2 + 2H_1 H_2 \text{cov}(\hat{\alpha}_1, \hat{\alpha}_2) + D\hat{\alpha}_2 H_2^2).$$

Иногда оценки, получаемые с помощью метода моментов, принимаются в качестве первого приближения, по которому можно построить другими методами оценки более высокого качества.

Оценки метода моментов используются также, когда существует необходимость оценить не сами параметры распределения (которые часто представляют собой некие абстракции), а определенные практически значимые показатели, зависящие от этих параметров функционально: $G = g(\theta_1, \theta_2, \dots, \theta_k)$. Самый простой (хотя и не самый точный) способ такого оценивания — подставить полученные оценки в соответствующую функцию: $\hat{G} = g(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$.

Если распределение определяется одним параметром, то для построения оценки один теоретический момент приравнивают к одному эмпирическому моменту того же порядка (обычно первого).

Задача 1. Случайная величина ξ (число появлений события A в t независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром p . Ниже приведено эмпирическое распределение числа появлений события в 10 опытах по 5 испытаний в каждом (в первой строке указано число x_i появлений события A в одном опыте; во второй строке указана частота n_i — количество опытов, в которых наблюдалось столько появлений события A).

x_i	0	1	2	3	4
n_i	5	2	1	1	1

Найти методом моментов точечную оценку параметра p биномиального распределения. Оценить вероятность $p_0 = P(\xi = 0)$.

Решение. Математическое ожидание биномиального распределения известно: $M\xi = tp$. Приравняв математическое ожидание к выборочному среднему, получим уравнение: $t\hat{p} = \bar{x}$, откуда $\hat{p} = \bar{x}/t$. Для рассматриваемого примера имеем:

$$\bar{x} = (0 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 1) / 10 = 1,1;$$

$$\hat{p} = 1,1/5 = 0,22; \hat{p}_0 = (1 - 0,22)^5 \approx 0,29.$$

Если распределение определяется двумя параметрами, то для построения их оценок два теоретических момента приравняют двум соответствующим эмпирическим моментам тех же порядков (обычно первым двум).

Задача 2. Случайная величина X (отклонение контролируемого размера изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами a и σ . Ниже приведена таблица наблюдаемых отклонений от номинала, подвергнутых группировке, для $n = 200$ изделий. В первой строке указаны середины интервалов отклонений x_i (мм); во второй строке приведена частота n_i — число наблюдений, попадающих в данный интервал.

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,2	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки неизвестных параметров a и σ нормального распределения. Оценить долю изделий с отклонением менее 1,5 мм в генеральной совокупности, используя нормальное приближение.

Решение. Для нахождения двух неизвестных параметров необходимо два уравнения. Первое получаем, приравняв начальный теоретический момент первого порядка к начальному эмпирическому моменту первого порядка, а второе — приравняв центральный теоретический момент второго порядка к центральному эмпирическому моменту второго порядка:

$$a' = \bar{x}, (\sigma')^2 = s^2.$$

Найдем величины $\bar{x}$ и $\hat{\sigma}^2$ по данным выборки:

$$\begin{aligned} \bar{x} &= \frac{1}{200} \sum_{i=1}^{11} n_i x_i = (0,3 \cdot 6 + 0,5 \cdot 9 + 0,7 \cdot 26 + 0,9 \cdot 25 + 1,1 \cdot 30 + 1,3 \cdot 26 + \\ &+ 1,5 \cdot 21 + 1,7 \cdot 24 + 1,9 \cdot 20 + 2,2 \cdot 8 + 2,3 \cdot 5) / 200 = 1,266. \end{aligned}$$

Для нахождения выборочной дисперсии перейдем к условным вариантам $u_i = 10x_i$:

$$\hat{\sigma}^2(u) = \frac{1}{200} \sum_{i=1}^{11} u_i^2 n_i - \left(\frac{1}{200} \sum_{i=1}^{11} u_i n_i \right)^2 = (6 \cdot 9 + 9 \cdot 25 + 26 \cdot 49 + 25 \cdot 81 + 30 \cdot 121 + 26 \cdot 169 + 21 \cdot 225 + 24 \cdot 289 + 20 \cdot 361 + 8 \cdot 484 + 5 \cdot 529) / 200 - 1,266^2 \cdot 100 = 24,72;$$

$\hat{\sigma}^2(x) = \hat{\sigma}^2(u)/100 = 0,2472$; $s^2 = \frac{200}{199} \hat{\sigma}^2 = 0,2484$ (при больших n исправленная и неисправленная выборочные дисперсии мало различаются); $s = 0,498$.

Таким образом, получаем: $a^* \approx 1,27$ (мм); $\sigma^* \approx 0,5$ (мм).

Оценим долю изделий с отклонением менее 1,5 мм как вероятность для нормальной ξ :

$$\hat{p} = \hat{P}(\xi < 1,5) = \Phi\left(\frac{1,5 - 1,27}{0,5}\right) \approx 0,68.$$

Замечание. Эту долю можно оценить также непосредственно по таблице: поскольку значение 1,5 мм делит соответствующий интервал пополам, получается, что 122 из 200 изделий имеет отклонение меньше заданного, что дает близкую оценку 0,61.

Задача 3. Предполагается, что выполнение некоторой работы занимает случайное время с распределением Симпсона на отрезке $[a, b]$. Хронометраж 20 испытаний дал среднее время работы 30 мин. и исправленную выборочную дисперсию 24 мин.². Определить параметры a и b методом моментов. Оценить, за какое время работа будет выполняться с вероятностью 98%.

Решение. Для распределения Симпсона (плотность которого имеет вид равнобедренного треугольника с основанием на заданном отрезке) имеем

$$M\xi = \frac{a+b}{2}; \quad D\xi = \frac{(b-a)^2}{24}.$$

Параметры распределения можно выразить через математическое ожидание и дисперсию:

$$a = M\xi - \sqrt{6D\xi}; \quad b = M\xi + \sqrt{6D\xi}.$$

Подставляя вместо теоретических моментов выборочные, получаем оценки

$$\hat{a} = \bar{x} - \sqrt{6s^2}; \quad \hat{b} = \bar{x} + \sqrt{6s^2},$$

откуда $\hat{a} = 18$ (мин.), $\hat{b} = 42$ (мин.).

Функция распределения Симпсона имеет вид:

$$F(x) = \begin{cases} 0, & x < a; \\ 2(x-a)^2/(b-a)^2, & a \leq x \leq (a+b)/2; \\ 1-2(b-x)^2/(b-a)^2, & (a+b)/2 \leq x \leq b; \\ 1, & x > b. \end{cases}$$

Решая уравнение $F(x) = 0,98$, находим искомое время $T = b - 0,1(b-a)$.

Подставляя полученные оценки в формулу вместо теоретических параметров, получаем $T = 42 - 0,1(42 - 18) = 39,6$ (мин.).

13.2. Метод максимального правдоподобия

Одним из основных методов получения оценок параметров генеральной совокупности по данным выборки является метод максимального правдоподобия, предложенный Р. Фишером. Основу метода составляет функция правдоподобия, выражающая плотность вероятности (либо вероятность) совместного появления результатов выборки $x_1, x_2, \dots, x_n$.

Результаты выборки рассматривают как одну из возможных реализаций n -мерной случайной величины $(\xi_1, \xi_2, \dots, \xi_n)$, компоненты которой независимы и имеют одну и ту же функцию распределения.

Совместное распределение этих величин задается в виде произведения частных распределений (поскольку предполагается, что наблюдения независимы) и, следовательно, функция правдоподобия имеет вид

$$L(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n P(x_i, \theta)$$

в случае дискретного распределения, заданного вероятностями $P(x, \theta)$, и

$$L(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n p(x_i, \theta)$$

в случае непрерывного распределения с плотностью $p(x, \theta)$.

Из определения функции правдоподобия следует, что чем вероятнее (правдоподобнее) набор значений $(x_1, x_2, \dots, x_n)$ случайной величины ξ при фиксированном θ , тем больше значение функции правдоподобия. Поэтому в качестве оценки неизвестного параметра θ принимается такое значение $\hat{\theta}_n$, которое максимизирует функцию правдоподобия.

Если параметр $\theta \in \Omega$, где Ω — замкнутая область допустимых значений параметра, то получаем задачу математического программирования:

найти такое $\hat{\theta}_n$, чтобы $L(x_1, \dots, x_n, \hat{\theta}_n) = \max_{\theta \in \Omega} L(x_1, \dots, x_n, \theta)$.

Поиск оценки упрощается, если максимизировать не саму функцию правдоподобия, а ее логарифм, потому что максимумы $L(X, \theta)$ и $\ln L(X, \theta)$ достигаются при одном и том же значении параметра θ . Функцию $l_n(x_1, \dots, x_n, \theta) = \ln L(X, \theta)$ называют **логарифмической функцией правдоподобия**.

Если максимум функции $l_n(x_1, \dots, x_n, \theta)$ достигается внутри допустимой области Ω , то в точке максимума $\hat{\theta}_n$ выполняются необходимые условия экстремума:

$$\frac{\partial l_n(x_1, \dots, x_n, \theta)}{\partial \theta_i} = 0 \text{ для любого } i = 1, 2, \dots, k.$$

Полученная система уравнений называется **уравнениями правдоподобия**. Решения этой системы могут соответствовать максимуму (минимуму) функции $L(x_1, \dots, x_n, \theta)$, а также являться точками перегиба. Необходимо проверить, что полученное решение есть точка максимума функции правдоподобия.

Отсюда алгоритм для отыскания оценки параметра θ : решают уравнение или систему уравнений правдоподобия, получаемых приравниванием производной (частных производных) по параметру (параметрам) θ к нулю: $\frac{\partial \ln L}{\partial \theta} = 0$. Затем отбирают то решение, которое соответствует именно максимуму функции $\ln L$, т.е. вторая производная в данной точке должна быть отрицательной: $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$. Иногда функция правдоподобия имеет несколько максимумов и приходится искать наибольший из них.

Бывают также случаи, когда этот алгоритм не действует, поскольку функция правдоподобия достигает максимума не во внутренней точке, а на границе некоторой области, либо ког-

да она просто недифференцируема в точке максимума. Такие случаи называют **нерегулярными**.

Достаточные условия регулярности (в одномерном случае) следующие.

1. В интервале возможных значений параметра θ плотность $p(x, \theta)$ трижды дифференцируема по θ , причем $\left| \frac{\partial^3 p}{\partial \theta^3} \right| \leq H(x)$, где $MH(\xi) \leq C$, C не зависит от θ .

2. Тождество $\int_{-\infty}^{\infty} p(x, \theta) dx \equiv 1$ можно дважды дифференцировать под знаком интеграла по параметру θ .

3. Математическое ожидание $I(\theta) = M \left(\frac{\partial \ln P(\xi, \theta)}{\partial \theta} \right)^2$ конечно и положительно.

► **Теорема 2.** Если выполнены условия регулярности 1–3, то

- 1) решение $\hat{\theta}_n$ уравнения правдоподобия существует;
- 2) $\hat{\theta}_n$ — состоятельная оценка параметра θ ;
- 3) распределение оценки $\hat{\theta}_n$ асимптотически нормально с параметрами θ и $\frac{1}{nI(\theta)}$;
- 4) оценка максимального правдоподобия асимптотически эффективна.

Если для оценок максимального правдоподобия выполнены условия теоремы Рао—Фреше—Крамера (см. § 11.4), то справедлива следующая

► **Теорема 3.** Если эффективная (по Рао—Фреше—Крамеру) оценка существует, то она является оценкой максимального правдоподобия.

Однако это не означает, что любая оценка максимального правдоподобия эффективна. Но если оценка оказывается неэффективной, это значит, что эффективных оценок вообще нет, хотя при этом могут существовать оценки с дисперсией, сколь угодно близкой к Δ_n .

Следует также отметить, что метод максимального правдоподобия иногда дает те же оценки, что и метод моментов, а иногда — другие. Бывает, что ни один из этих методов не дает хороших оценок, и приходится использовать другие методы.

А. Дискретные распределения

Задача 4. Найти методом максимального правдоподобия оценку вероятности «успеха» θ в схеме испытаний Бернулли.

Решение. Рассмотрим случайную величину

$$\xi = \begin{cases} 1, & \text{в случае «успеха»;} \\ 0, & \text{в случае «неудачи»,} \end{cases}$$

тогда функция вероятностей случайной величины ξ запишется в виде

$$P(x, \theta) = \begin{cases} \theta, & x=1; \\ 1-\theta, & x=0. \end{cases}$$

Логарифмическая функция правдоподобия для одного испытания будет иметь вид

$$l_1(\theta) = \ln P(x, \theta) = \begin{cases} \ln \theta, & x=1; \\ \ln(1-\theta), & x=0. \end{cases}$$

Для n испытаний и m «успехов» в n испытаниях получаем

$$l_n(\theta) = m \ln \theta + (n - m) \ln(1 - \theta).$$

Отсюда $l'_n(\theta) = \frac{m}{\theta} - \frac{n-m}{1-\theta} = 0$ и $\theta_n = \frac{m}{n}$. Проверим знак

второй производной: $l''_n(\theta) = -\frac{m}{\theta^2} - \frac{n-m}{(1-\theta)^2} < 0$.

Таким образом, относительная частота появления события является оценкой вероятности «успеха» в одном испытании Бернулли, найденной методом максимального правдоподобия. Поскольку $M\hat{\theta}_n = \theta$, то оценка $\hat{\theta}_n$ является несмещенной оценкой вероятности.

Задача 5. Случайная величина ξ (число поврежденных стеклянных изделий в одном контейнере) распределена по закону Пуассона с неизвестным параметром λ . В таблице приведено эмпирическое распределение числа поврежденных изделий в 500 контейнерах (в первой строке указано число x_i поврежденных изделий в одном контейнере, во второй строке указана частота n_i — число контейнеров, содержащих x_i поврежденных изделий).

x_i	0	1	2	3	4	5	6	7
n_i	199	169	87	31	9	3	1	1

Найти методом максимального правдоподобия точечную оценку неизвестного параметра λ распределения Пуассона.

Решение. Выпишем функцию правдоподобия

$$L = P(x_1; \lambda) P(x_2; \lambda) \dots P(x_n; \lambda) = \frac{1}{x_1! x_2! \dots x_n!} \lambda^{x_1 + x_2 + \dots + x_n} e^{-n\lambda}.$$

Найдем точку максимума логарифмической функции правдоподобия, для чего приравняем к нулю ее первую производную по λ :

$$\ln L = \left(\sum_{i=1}^n x_i \right) \ln \lambda - n\lambda - \ln(x_1! x_2! \dots x_n!), \quad \frac{\partial \ln L}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0.$$

Получаем $\lambda^* = \bar{x}$. Убедимся, что полученное значение λ является точкой максимума. Для этого найдем вторую производную и проверим ее знак в точке λ^* : $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i$. Если в последнее уравнение подставить $\lambda^* = \bar{x}$, то вторая производная

будет отрицательна, значит $\bar{x}$ является точкой максимума.

Найдем значение λ^* для рассматриваемого примера:

$$\lambda^* = \frac{1}{n} \sum_{i=1}^8 n_i x_i = (169 + 2 \cdot 87 + 3 \cdot 31 + 4 \cdot 9 + 5 \cdot 3 + 6 + 7) / 500 = 1.$$

Б. Непрерывные распределения

Задача 6. Найти методом максимального правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечные оценки параметров a и σ нормального распределения, плотность которого

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

Решение. Выпишем функцию правдоподобия:

$$L(x_1, x_2, \dots, x_n, a, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2\right).$$

Логарифмическая функция правдоподобия имеет вид

$$\ln L = -n \ln \sigma - n \ln \sqrt{2\pi} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2.$$

Найдем точку максимума, решив систему из двух уравнений, получающихся путем приравнивания первых двух частных производных по неизвестным параметрам к нулю:

$$\begin{cases} \frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a) = 0; \\ \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - a)^2 = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^n (x_i - a) = 0; \\ \frac{n}{\sigma} = \frac{\sum_{i=1}^n (x_i - a)^2}{\sigma^3} \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \hat{a} = \frac{1}{n} \sum_{i=1}^n x_i; \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \end{cases}$$

Проверим, является ли точка $(\hat{a}, \hat{\sigma}^2)$ точкой максимума функции правдоподобия:

$$\begin{cases} \frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{\sigma^2}; \\ \frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (x_i - a)^2; \\ \frac{\partial^2 \ln L}{\partial a \partial \sigma} = -\frac{2}{\sigma^3} \sum_{i=1}^n (x_i - a). \end{cases}$$

Как известно из математического анализа, чтобы функция $f(x_1, x_2, \dots, x_n)$ достигала максимума в некоторой точке, достаточно, чтобы матрица второго дифференциала функции d^2f в этой точке была отрицательно определена. По критерию Сильвестра для этого необходимо и достаточно, чтобы ее главные миноры чередовались по знаку, а именно $\Delta_1 < 0$, $\Delta_2 > 0$. Рассмотрим матрицу производных

$$A = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{2}{\sigma^3} \sum_{i=1}^n (x_i - a) \\ -\frac{2}{\sigma^3} \sum_{i=1}^n (x_i - a) & \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (x_i - a)^2 \end{pmatrix}.$$

Отсюда

$$\Delta_1 = -\frac{n}{\sigma^2} < 0;$$

$$\Delta_2 = -\frac{n^2}{\sigma^4} + \frac{3n}{\sigma^6} \sum_{i=1}^n (x_i - a)^2 - \frac{4}{\sigma^6} \left(\sum_{i=1}^n (x_i - a) \right)^2 = -\frac{n^2}{\sigma^4} + \frac{3n^2}{\sigma^4} - 0 = \frac{2n^2}{\sigma^4} > 0.$$

Следовательно, точка $(\hat{a}, \hat{\sigma}^2)$ — действительно точка максимума, и полученные оценки являются оценками максимального правдоподобия.

Задача 7. Найти оценку максимального правдоподобия для параметра сдвига θ распределения Коши, заданного плотностью $p(x) = \frac{1}{\pi(1+(x-\theta)^2)}$, по выборке из двух наблюдений, если:

а) $x_1 = -1, x_2 = 1$; б) $x_1 = -2, x_2 = 2$.

Решение. Функция правдоподобия для двух наблюдений имеет вид

$$L(\theta) = \frac{1}{\pi(1+(x_1-\theta)^2)} \cdot \frac{1}{\pi(1+(x_2-\theta)^2)}.$$

Введем функцию $R(\theta) = \frac{1}{\pi^2 L(\theta)}$. Тогда задача максимизации функции правдоподобия эквивалентна задаче минимизации $R(\theta)$:

а) если $x_1 = -1, x_2 = 1$, то $R(\theta) = (1+(1+\theta)^2)(1+(\theta-1)^2) = \theta^4 + 4$. Функция R достигает минимума в точке $\theta = 0$, так что это и есть оценка максимального правдоподобия;

б) если $x_1 = -2, x_2 = 2$, то $R(\theta) = (1+(2+\theta)^2)(1+(\theta-2)^2) = \theta^4 - 6\theta^2 + 25$, и производная имеет три нуля: в точках $\theta = 0$ и $\theta = \pm\sqrt{3}$. При этом точка $\theta = 0$ оказывается точкой максимума функции R . Точкам $\theta = \pm\sqrt{3}$ соответствует минимум функции R , причем в обеих этих точках величина R одинакова. Таким образом, оба значения $\pm\sqrt{3}$ являются в данном случае оценками максимального правдоподобия.

Замечание. Ни метод моментов, ни метод максимального правдоподобия не могут дать хороших оценок для параметра сдвига распределения Коши. Тем не менее существует простая оценка для него — выборочная медиана: $\hat{\theta}_n = x_{\text{мед}}$, поскольку $R(\theta) = 1/2$.

В. Нерегулярные случаи

Задача 8. Найти методом максимального правдоподобия оценки параметров a и b равномерного закона распределения:

$$p(x; a, b) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Решение. Для равномерного закона не выполняется одно из условий регулярности, а именно: плотность разрывна по параметрам. В подобных ситуациях оценку следует искать другим способом.

Выпишем функцию правдоподобия для равномерного распределения:

$$L(x_1, x_2, \dots, x_n) = \frac{1}{(b-a)^n}, \text{ если все } x_i \in [a, b];$$

$$L(x_1, x_2, \dots, x_n) = 0 \text{ в остальных случаях.}$$

Условие, что все наблюдения принадлежат отрезку $[a, b]$, можно выразить через неравенства для крайних членов вариационного ряда: $a \leq x_{\min}$, $b \geq x_{\max}$. При фиксированном значении a функция правдоподобия убывает по b при $b \geq x_{\max}$ и, следовательно, принимает максимальное значение при $b = x_{\max}$. При фиксированном значении b функция правдоподобия возрастает по a при $a \leq x_{\min}$ и, следовательно, принимает максимальное значение при $a = x_{\min}$.

Таким образом, оценками максимального правдоподобия будут крайние члены вариационного ряда $\hat{a} = x_{\min}$ и $\hat{b} = x_{\max}$.

Задача 9. Построить оценку методом максимального правдоподобия параметра сдвига θ для распределения Лапласа с плотностью $p(x) = \frac{1}{2} \exp(-|x - \theta|)$, $-\infty < x < \infty$.

Решение. Функция правдоподобия имеет вид

$$L(\theta) = \frac{1}{2^n} \exp\left(-\sum_{i=1}^n |x_i - \theta|\right).$$

Логарифмируя, получаем

$$\ln L(\theta) = -n \ln 2 - \sum_{i=1}^n |x_i - \theta|.$$

Заметим, что эта функция недифференцируема во всех точках $x_1, x_2, \dots, x_n$, а в остальных точках производная имеет вид

$$\frac{d \ln L}{d \theta} = \sum_{i=1}^n \begin{cases} +1, & \theta < x_i; \\ -1, & \theta > x_i. \end{cases}$$

Отсюда следует, что функция правдоподобия возрастает, если слева от значения θ находится меньше членов вариационного ряда, чем справа, и убывает в противном случае. Следовательно, максимума она достигает посередине вариационного ряда. Если $n = 2k + 1$, то это происходит в точке $x_{(k)}$. Если $n = 2k$, то функция постоянна на интервале $(x_{(k)}, x_{(k+1)})$, где принимает наибольшее значение, и в качестве оценки можно взять середину этого интервала. Таким образом, оценкой максимального правдоподобия оказывается выборочная медиана $\hat{\theta}_n = x_{\text{мед}}$.

13.3. Метод наименьших квадратов.

Линейная регрессия

Метод наименьших квадратов заключается в том, что оценка определяется из условия минимизации суммы квадратов отклонений выборочных данных от определяемой оценки. Метод наименьших квадратов получил самое широкое распространение в практике статистических исследований, так как, во-первых, не требует знания закона распределения выборочных данных, во-вторых, достаточно хорошо разработан в отношении вычислительной реализации.

Первоначально метод был разработан для обработки данных с нормальными ошибками. В простейшем случае речь идет о данных вида $x_i = \theta + \varepsilon_i$, где $\varepsilon_i \in N(0, \sigma^2)$. Оценка θ строится методом максимального правдоподобия. Функция правдоподобия для нормального распределения равна

$$L(\theta) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2} = C e^{-hT},$$

где $C = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n$, $h = \frac{1}{2\sigma^2} > 0$, $T = \sum_{i=1}^n (x_i - \theta)^2$.

Следовательно, функция правдоподобия имеет максимум тогда, когда $T = \sum_{i=1}^n (x_i - \theta)^2$ достигает минимума. Если на пара-

метр θ не накладываются ограничения, то оценка, полученная методом наименьших квадратов, совпадает с оценкой максимального правдоподобия для нормального распределения $\hat{\theta} = \bar{x}$.

В более общем случае пусть Y — некоторый экономический показатель, объективный закон которого описывается функциональной зависимостью $Y = \varphi(X, \theta)$, где $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ — параметр; X — многомерная неслучайная переменная. Пусть в результате i -го наблюдения мы получили значение y_i функции $\varphi(x_i, \theta)$ со случайной ошибкой ε_i , т.е. $y_i = \varphi(x_i, \theta) + \varepsilon_i$. Требуется по наблюдениям $(x_1, y_1), \dots, (x_n, y_n)$ оценить значения параметров $\theta_1, \theta_2, \dots, \theta_k$.

Если в результате наблюдений получено n пар значений (x_i, y_i) , где x_i — значение аргумента, а y_i — значение функции, то параметры $(\theta_1, \theta_2, \dots, \theta_k)$ аппроксимирующей функции выбираются так, чтобы обратилась в минимум сумма

$$S = \sum_{i=1}^n [y_i - \varphi(x_i, \theta)]^2.$$

Обоснование метода следующее. Пусть измерения независимы и распределение ошибок нормальное: $\varepsilon_i \in N(0, \sigma^2)$. Величина y_i как сумма постоянной величины $\varphi(x_i, \theta)$ и случайной величины ε_i является случайной величиной, и ее распределение также нормальное: $y_i \in N(\varphi(x_i, \theta), \sigma^2)$. Плотность распределения y_i будет иметь вид

$$p(y_i, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \varphi(x_i, \theta))^2}{2\sigma^2}}.$$

Функция правдоподобия для наблюдаемых значений $y_1, y_2, \dots, y_n$ будет равна

$$L(Y, \theta) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - \varphi(x_i, \theta)]^2}.$$

Отсюда функция правдоподобия $L(\theta)$ при изменении θ имеет максимум тогда и только тогда, когда статистика

$$T_n(\theta) = \sum_{i=1}^n [y_i - \varphi(x_i, \theta)]^2$$

достигает минимума.

Задача свелась к задаче математического программирования: найти такое значение $\hat{\theta}_n$, которое минимизировало бы квадратичную форму

$$T_n(\hat{\theta}_n) = \min_{\theta_1, \dots, \theta_k} \sum_{i=1}^n [y_i - \varphi(x_i, \theta)]^2,$$

где $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ — случайная выборка. Так как $\hat{\theta}_n = (\theta_1, \theta_2, \dots, \theta_k)$ является точкой минимума статистики $T_n(\theta)$, то, приравнявая к нулю ее частные производные

$$\frac{\partial T_n(\hat{\theta}_n)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k,$$

получаем систему нормальных уравнений, решения которой и являются оценками $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$ неизвестных параметров, найденными методом наименьших квадратов.

Система нормальных уравнений всегда имеет решение, так как положительный квадратичный многочлен всегда достигает минимума. Однако решение не обязательно является единственным. Может случиться, что нормальные уравнения однозначно разрешимы лишь для некоторых определенных линейных комбинаций параметров $\theta_1, \theta_2, \dots, \theta_k$, а относительно самих параметров однозначного решения нет. Такие линейные комбинации индийский статистик Рао назвал *допускающими оценку*.

На практике метод применяется теперь гораздо шире, в частности, если об ошибках ε_i известно лишь, что $M\varepsilon_i = 0$ и $D\varepsilon_i < \infty$. Бывает, что метод применяется даже в случаях, когда об ошибках нельзя сказать, что они являются случайными величинами, независимыми и одинаково распределенными.

Одной из основных задач математической статистики является исследование зависимости между двумя или несколькими переменными. Строгая функциональная зависимость реализуется редко, так как одна или обе величины подвержены еще и воздействию случайных факторов. *Статистической* называется зависимость, при которой изменение одной из величин влечет за собой изменение распределения другой.

Поскольку обычно невозможно точно предсказать неизвестное значение одной величины при известном значении другой, это желательно сделать хотя бы «в среднем». Поэтому естественно использовать *функции регрессии* (см. § 7.4). Напримср, в случае двумерного нормального распределения с плотностью

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-a_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right)\right\}$$

функции регрессии оказываются линейными и имеют вид:

$$\varphi_{\xi_1|\xi_2}(x_1) = a_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - a_1), \quad \varphi_{\xi_2|\xi_1}(x_2) = a_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - a_2).$$

Однако на практике модель линейной зависимости между величинами используется гораздо шире, без предположения нормальности совместного распределения.

Линейной регрессией называется сведение наблюдаемой на опыте зависимости некоторой переменной (**зависимой** или **объясняемой**) от одной или более других переменных (**независимых** или **объясняющих**) к линейной (в предположении, что строгая линейная зависимость между ними нарушается случайными ошибками). Для проведения линейной регрессии часто используется метод наименьших квадратов.

В простейшем случае речь идет о двух переменных. Пусть x — независимая переменная, y — зависимая, и между ними существует следующая связь:

$$y_i = a + bx_i + \varepsilon_i,$$

где a и b — числовые коэффициенты, ε_i — случайные ошибки, $M\varepsilon_i = 0$ и $D\varepsilon_i < \infty$. Задача состоит в том, чтобы по имеющимся наблюдениям $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ построить оценки для a и b .

Согласно методу наименьших квадратов необходимо решить следующую математическую задачу:

$$T = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min.$$

Решим задачу, вычислив частные производные суммы квадратов по каждому из коэффициентов и приравняв эти производные к нулю. Получим систему **нормальных уравнений**:

$$\begin{cases} \frac{\partial T}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0; \\ \frac{\partial T}{\partial b} = -2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0. \end{cases}$$

Решая данную систему относительно параметров a и b , получим оценки:

$$\hat{b} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad \hat{a} = \bar{y} - \hat{b}\bar{x}.$$

Уравнение вида $\hat{y} = \hat{a} + \hat{b}x$ называется **уравнением линейной регрессии**, а получаемые из него значения $\hat{y}_i = \hat{a} + \hat{b}x_i$ — **предсказанными значениями**, в отличие от **наблюдаемых значений** y_i . Заметим, что уравнение линейной регрессии часто бывает удобно записать в виде $\hat{y} = \bar{y} + \hat{b}(x - \bar{x})$. Соответствующая прямая всегда проходит через точку выборочных средних $(\bar{x}, \bar{y})$. Числитель и знаменатель в формуле оценки параметра b можно вычислять по следующим эквивалентным формулам:

$$\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}; \quad \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2.$$

Здесь также можно перейти к условным вариантам $u_i = x_i - c$, $v_i = y_i - d$, и оценка b от этого не изменится.

В некоторых случаях, например когда величина y (по смыслу) представляет собой некую долю от x , рассматривают более простой тип линейной зависимости:

$$y_i = bx_i + \varepsilon_i$$

т.е. полагают $a = 0$. В этом случае оценка b методом наименьших квадратов имеет вид

$$\hat{b} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}.$$

Важным и практически значимым результатом линейной регрессии является то, что она позволяет «предсказывать» значения зависимой переменной даже для таких значений независимых, которые не наблюдались реально. Таким образом, например, можно строить прогнозы на будущее.

Задача 10. Затраты x на развитие производства и y — величина годовой прибыли фирмы в течение 5 лет представлены в условных единицах следующей таблицей:

x	6	3	7	5	10
y	33	27	32	28	42

На величину прибыли влияют случайные факторы. Предполагается, что имеет место линейная зависимость $y_i = a + bx_i + \varepsilon_i$ между затратами x и прибылью y . Среднее значение ε_i равно нулю и дисперсия конечна. Каждый год случайное влияние не коррелировано с предыдущими годами. Оценить параметры a

и b . Оценить годовую прибыль в том случае, если на развитие производства будет затрачено 12 у.е.

Решение. Перейдем к условным вариантам $u_i = x_i - 6$, $v_i = y_i - 33$.

u	0	-3	1	-1	4
v	0	-6	-1	-5	9

Получаем $\bar{u} = (0 - 3 + 1 - 1 + 4)/5 = 0,2$; $\bar{v} = (0 - 6 - 1 - 5 + 9)/5 = -0,6$.

Отсюда $\bar{x} = 6,2$; $\bar{y} = 32,4$.

Далее вычисляем следующие суммы:

$$\sum_{i=1}^n u_i v_i - n\bar{u}\bar{v} = ((-3) \cdot (-6) + 1 \cdot (-1) + (-1) \cdot (-5) + 4 \cdot 9) - 5 \cdot 0,2 \cdot (-0,6) = 18 - 1 + 5 + 36 + 0,6 = 58,6.$$

$$\sum_{i=1}^n u_i^2 - n\bar{u}^2 = (0 + (-3)^2 + 1^2 + (-1)^2 + 4^2) - 5 \cdot 0,2^2 = 9 + 1 + 1 + 16 - 0,2 = 26,8.$$

Получаем $\hat{b} = 58,6/26,8 \approx 2,187$; $\hat{a} = 32,4 - 2,187 \cdot 6,2 \approx 18,843$.

Имеем $\hat{y}(12) \approx 45$ (у.е.).

На рис. 13.1 представлены данные задачи и прямая линейной регрессии.

Замечание. На самом деле по столь небольшому числу точек нельзя делать серьезные выводы на будущее.



Рис. 13.1

Задача 11. В таблице представлены данные о годовых доходах и расходах на личное потребление (долл. США) для 10 семей:

Годовой доход	Расходы на личное потребление
2508	2406
2572	2464
2408	2336
2522	2281
2700	2641
2531	2385
2390	2297
2595	2416
2524	2460
2685	2448

Провести линейную регрессию расходов по доходам в виде $y = bx$. Оценить параметр b . Оценить величину расходов для семьи с годовым доходом 2500 долл. США.

Решение. Суммируем доходы x_i , расходы y_i и делим одно на другое. Получаем значение параметра $\hat{b} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = 24\,134 / 25\,435 \approx 0,949$.

Для семьи с доходом 2500 долл. США получаем оценку расходов в 2372 долл. США.

На рис. 13.2 представлены данные задачи и прямая линейной регрессии.

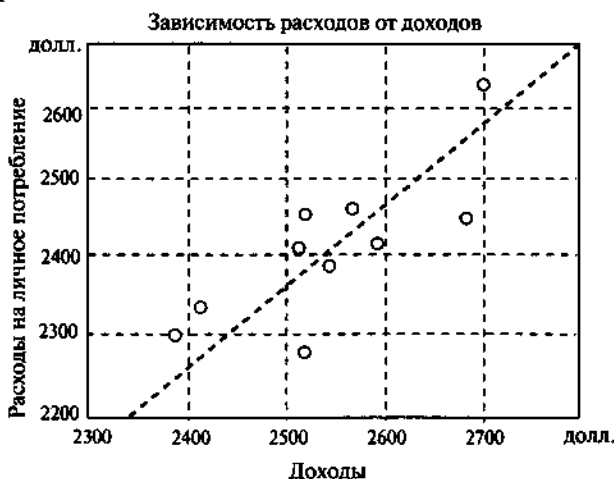


Рис. 13.2

Задачи для самостоятельного решения

Теоретические задачи

1. Найти методом моментов оценку параметра λ распределения Пуассона.
2. Найти методом моментов оценку параметра p (вероятности «успеха») для геометрического распределения.
3. Найти методом моментов оценку параметра θ для геометрического распределения с вероятностью «успеха» $p = 1/(1 + \theta)$, $\theta \geq 0$. Доказать ее несмещенность.
4. В случае сдвинутого показательного распределения $p(x, \theta, \mu) = \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}$, $x > \mu$ ($\theta, \mu > 0$) с помощью метода моментов найти оценки θ и μ параметров θ и μ соответственно.
5. Для гамма-распределения $\gamma(\alpha, \lambda)$ методом моментов найти оценку параметра α при известном λ . Доказать ее несмещенность и найти дисперсию этой оценки.
6. Для гамма-распределения $\gamma(\alpha, \lambda)$ методом моментов найти оценки неизвестных параметров α и λ .
7. Найти оценку для числа степеней свободы r распределения χ_r^2 методом моментов. Доказать несмещенность и найти дисперсию этой оценки.
8. Найти оценку для числа степеней свободы r распределения Стьюдента методом моментов. При каких r это возможно?
9. Пусть случайная величина ξ равномерно распределена на отрезке $[0, \theta]$. Найти методом моментов оценку для параметра θ .
10. Пусть случайная величина ξ равномерно распределена на отрезке $[a, b]$. Найти методом моментов оценки для a и b .
11. Пусть случайная величина ξ равномерно распределена на $[c - \sqrt{d}; c + \sqrt{d}]$. Найти методом моментов оценки для c и d . Доказать их несмещенность.
12. В случае равномерного распределения на отрезке $[\theta_1, \theta_1 + \theta_2]$ с помощью метода моментов найти оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ для параметров θ_1 и θ_2 .
13. Случайная величина ξ равна сумме двух независимых случайных величин, имеющих показательное распределение с параметрами λ_1 и λ_2 соответственно. По наблюдениям случайной

величины ξ оценить параметры λ_1 и λ_2 методом моментов, в предположении, что $\lambda_1 < \lambda_2$.

14. По наблюдениям случайной величины с распределением Парето

$$F(x) = \begin{cases} 0, & x < 1; \\ 1 - x^{-\alpha}, & x \geq 1 \end{cases} \quad \text{оценить параметр } \alpha \text{ методом моментов.}$$

При каких значениях параметра α это возможно?

15. По наблюдениям случайной величины с распределением $F(x) = x^\beta$, $0 \leq x \leq 1$, оценить параметр β методом моментов.
16. По наблюдениям случайной величины с распределением $F(x) = (x/\theta)^\beta$, $0 \leq x \leq \theta$, оценить методом моментов а) параметр β при известном θ ; б) параметр θ при известном β .
17. По наблюдениям логнормальной случайной величины $\xi \in \exp\{N(a, \sigma^2)\}$ найти оценки параметров a и σ^2 методом моментов.
18. Найти оценку методом моментов для параметра λ распределения Лапласа, заданного функцией плотности $p(x) = \frac{\lambda}{2} e^{-\lambda|x|}$.
19. Найти по выборке $x_1, x_2, \dots, x_n$ методом максимального правдоподобия точечную оценку параметра p геометрического распределения: $P(\xi = k) = p(1 - p)^{k-1}$, $k \geq 1$.
20. Найти методом максимального правдоподобия оценку параметра θ для геометрического распределения с вероятностью «успеха» $p = 1/(1 + \theta)$, $\theta \geq 0$. Доказать ее несмещенность.
21. По выборке $x_1, x_2, \dots, x_n$ в случае биномиального распределения $P(\xi = m) = C_N^m p^m q^{N-m}$ при известном N методом максимального правдоподобия найти оценку параметра p . Совпадает ли эта оценка с оценкой, полученной с помощью метода моментов? Исследовать оценку на несмещенность и состоятельность.
22. По выборке $x_1, x_2, \dots, x_n$ в случае равномерного распределения на отрезке $[0, \theta]$ методом максимального правдоподобия найти оценку параметра θ . Вычислить математическое ожидание полученной оценки и построить несмещенную оценку на ее основе.
23. Случайная величина равномерно распределена на $[0, 2\theta]$, $\theta > 0$. Найти оценку параметра θ методом максимального правдоподобия.
24. Случайная величина равномерно распределена на $[-\theta, 2\theta]$, $\theta > 0$. Найти оценку параметра θ методом максимального правдоподобия.

25. Случайная величина равномерно распределена на $[1 - \theta, 1 + \theta]$, $0 < \theta < 1$. Найти оценку параметра θ методом максимального правдоподобия.
26. Найти методом максимального правдоподобия оценку параметра a для распределения, заданного следующей функцией плотности:

$$p(x) = \begin{cases} 1, & a \leq x \leq a+1; \\ 0, & x \notin [a, a+1]. \end{cases}$$

Построить несмещенную оценку на основе оценки максимального правдоподобия.

27. Найти оценку методом максимального правдоподобия для параметра λ распределения Лапласа, заданного функцией плотности $P(x) = \frac{\lambda}{2} e^{-\lambda|x|}$.

28. Оценить с помощью метода максимального правдоподобия параметр сдвига θ в сдвинутом показательном распределении, заданном плотностью $p(x) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta; \\ 0, & x < \theta. \end{cases}$

29. Пусть случайная величина ξ имеет функцию распределения

$$F(x) = \begin{cases} e^{x-\theta}, & x < \theta, \\ 0, & x \geq \theta. \end{cases}$$

Методом максимального правдоподобия оценить параметр θ .

30. Распределение Кэптейна имеет плотность вида

$$p(x) = \frac{g'(x)}{\sigma\sqrt{2\pi}} e^{-\frac{(g(x)-a)^2}{2\sigma^2}},$$

где $g(x)$ — дифференцируемая функция. Найти методом максимального правдоподобия точечную оценку параметра a , если параметр σ известен.

31. Для распределения Кэптейна найти методом максимального правдоподобия точечную оценку параметра σ , если параметр a известен.
32. Для гамма-распределения $\gamma(\alpha, \lambda)$ методом максимального правдоподобия найти оценку параметра λ при известном α .

33. Для показательного распределения с плотностью $p(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$, $x \geq 0$, $\mu > 0$, найти оценку параметра μ методом максимального правдоподобия.

34. Функция распределения случайной величины ξ имеет вид $F(x) = e^{-\frac{\theta}{x^2}}$, $x > 0$, $\theta > 0$. Найти оценку параметра θ методом максимального правдоподобия.

35. Для распределения Вейбулла вида $F(x) = 1 - e^{-\lambda x^2}$, $x \geq 0$, $\lambda > 0$, найти оценку параметра λ методом максимального правдоподобия.

36. Для сдвинутого показательного распределения с плотностью $p(x) = \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}$, $x \geq \mu$, $\theta > 0$, методом максимального правдоподобия найти оценки параметров θ и μ .

37. Для распределения Парето вида

$$F(x) = \begin{cases} 0, & x < 1; \\ 1 - x^{-\alpha}, & x \geq 1, \end{cases}$$

оценить параметр α методом максимального правдоподобия.

38. По наблюдениям случайной величины с распределением $F(x) = x^\beta$, $0 \leq x \leq 1$, оценить параметр β методом максимального правдоподобия.

39. По наблюдениям случайной величины с распределением $F(x) = (x/\theta)^\beta$, $0 \leq x \leq \theta$, оценить методом максимального правдоподобия: а) параметр β при известном θ ; б) параметр θ при известном β .

40. По наблюдениям логнормальной случайной величины $\xi \in \exp\{N(a, \sigma^2)\}$ найти оценки параметров a и σ^2 методом максимального правдоподобия.

41. По наблюдениям случайной величины, равномерно распределенной на отрезке $[a, b]$, найти оценки параметров a и b методом максимального правдоподобия. Найти их математическое ожидание и построить несмещенные оценки.

42. Доказать эффективность оценки, полученной методом максимального правдоподобия, для параметра θ гамма-распределения (с фиксированным параметром α), параметризованного в виде

$$p(x, \alpha, \theta) = \frac{x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta}, \quad x > 0.$$

43. Доказать, что оценка методом моментов для параметра p биномиального распределения $P(\xi = m) = C_N^m p^m q^{N-m}$ является эффективной (при известном N).

44. Доказать, что не существует эффективных оценок для параметра λ показательного распределения $p_i(x, \lambda) = \lambda e^{-\lambda x}$, $x > 0$.
45. Доказать, что не существует эффективных оценок для параметра λ гамма-распределения (с фиксированным параметром α), параметризованного в виде
- $$p(x, \alpha, \theta) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, \quad x > 0.$$
46. Доказать, что не существует эффективных оценок для параметра σ^2 нормального распределения $N(a, \sigma^2)$.

Вычислительные задачи

47. Случайная величина ξ (число семян сорняков в пробе зерна) распределена по закону Пуассона: $P(l) = \frac{\lambda^l}{l!} e^{-\lambda}$. Ниже приведено распределение семян сорняков в $n = 1000$ пробах зерна (в первой строке указано количество x_i сорняков в одной пробе; во второй строке указана частота n_i — число проб, содержащих x_i семян сорняков).

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

Найти методом моментов точечную оценку параметра λ . Оценить вероятность того, что в пробе зерна не будет сорняков.

48. Случайная величина ξ (срок службы изделия) имеет показательное распределение $p(x) = \lambda e^{-\lambda x}$ ($x \geq 0$). В таблице приведены сгруппированные данные по срокам службы (ч) для $n = 200$ изделий.

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Найти методом моментов точечную оценку неизвестного параметра λ показательного распределения. Оценить время, которое изделие прослужит с вероятностью 90%.

49. В таблице приведены сгруппированные данные о коэффициентах соотношения заемных и собственных средств на 100 малых предприятиях региона.

Номер интервала	Интервал	Середина интервала x_i	n_i
1	5,05–5,15	5,1	5
2	5,15–5,25	5,2	8
3	5,25–5,35	5,3	12
4	5,35–5,45	5,4	20
5	5,45–5,55	5,5	26
6	5,55–5,65	5,6	15
7	5,65–5,75	5,7	10
8	5,75–5,85	5,8	4

Оценить долю малых предприятий с коэффициентом не более 5,5 на основе оценок методом моментов (используя нормальное приближение) и непосредственно по таблице.

50. В ОТК были измерены диаметры 300 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала (нм) даны в таблице.

Границы отклонений	Середина интервала	Число валиков	Границы отклонений	Середина интервала	Число валиков
–30...–25	–27,5	3	0...5	2,5	55
–25...–20	–22,5	8	5...10	7,5	30
–20...–15	–17,5	15	10...15	12,5	25
–15...–10	–12,5	35	15...20	17,5	14
–10...–5	–7,5	40	20...25	22,5	8
–5...0	–2,5	60	25...30	27,5	7

Оценить долю изделий, для которых отклонение не превосходит 15 нм по абсолютной величине, с применением метода моментов (используя нормальное приближение) и непосредственно по таблице.

51. В таблице представлены данные по числу сделок на фондовой бирже за квартал для 400 инвесторов:

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

В предположении, что случайное число сделок описывается распределением Пуассона, оценить параметр λ методом моментов. Определить вероятность того, что число сделок за квартал будет не менее двух, применяя метод моментов, и непосредственно по таблице.

52. Для изучения распределения заработной платы работников определенной отрасли обследовано 100 человек. Результаты представлены в таблице.

Заработная плата, долл. США	Число человек	Заработная плата, долл. США	Число человек
190–192	1	200–202	19
192–194	5	202–204	11
194–196	9	204–206	4
196–198	22	206–208	1
198–200	28	208–210	0

Определить долю работников с заработной платой менее 200 долл. США на основе оценок методом моментов (используя нормальное приближение) и непосредственно по таблице.

53. При измерении веса 20 шоколадных батончиков (с номинальным весом 50 г) получены следующие значения (г): 49,1; 50,0; 49,7; 50,5; 48,1; 50,3; 49,7; 51,6; 49,8; 50,1; 49,7; 48,8; 51,4; 49,1; 49,6; 50,9; 48,5; 52,0; 50,7; 50,6.

Оценить долю батончиков с весом менее 49 г на основе оценок методом моментов (используя нормальное приближение) и непосредственно по их доле в выборке.

54. Уровень воды в реке во время паводка (по сравнению с нормой) описывается гамма-распределением с параметрами α и λ . Наблюдения за паводками показали выборочное среднее 165 см и выборочное среднее квадратическое отклонение 83 см. Найти методом моментов оценки α и λ . Вычислить модальное значение уровня воды.
55. Время выполнения работы описывается гамма-распределением с параметрами α и λ . Хронометраж показал выборочное среднее N мин и выборочное среднее квадратическое отклонение M мин. Найти методом моментов оценки α и λ . Вычислить модальное значение времени выполнения работы. Решить задачу при: а) $N = 60$, $M = 30$; б) $N = 63$, $M = 21$.
56. В поселке Червонцево все жители имеют доход не менее 100 тыс. руб. в месяц. Выборочное обследование доходов 10 человек дало средний доход 200 тыс. руб. В предположении, что случайная величина дохода имеет распределение Парето вида

$$F(x) = \begin{cases} 0, & x < x_0; \\ 1 - (x/x_0)^{-\alpha}, & x \geq x_0, \end{cases}$$

где $x_0 = 100$ тыс. руб., оценить параметр α и средний доход жителей с доходами свыше 500 тыс. руб. с использованием метода моментов.

57. В поселке Полтинниково все жители имеют доход не менее 50 тыс. руб. в месяц. Выборочное исследование дало средний доход 250 тыс. руб. В предположении, что случайная величина дохода имеет распределение Парето с $x_0 = 50$ (тыс. руб.), оценить параметр α методом моментов и долю жителей с доходами свыше 800 тыс. руб.
58. Известно, что некоторая работа занимает время, состоящее из обязательного периода μ и случайной задержки, распределенной показательно со средним значением θ . Хронометраж рабочего времени в 10 испытаниях показал среднее время 37 мин. при исправленной выборочной дисперсии 49 мин². Оценить параметры μ и θ методом моментов. Оценить срок, за который работа будет выполнена с вероятностью 99%.
59. Известно, что некоторая работа занимает время, состоящее из постоянного периода и случайной задержки, распределенной показательно. Хронометраж рабочего времени показал, что работа занимает в среднем N мин. со средним квадратическим отклонением M мин. С помощью метода моментов оценить вероятность того, что работа будет закончена за час, а также время, за которое работа будет выполнена с вероятностью 95%. Решить задачу при: а) $N = 45$, $M = 15$; б) $N = 40$, $M = 8$.
60. Прибор состоит из двух блоков: основного и резервного. Если основной блок выходит из строя, включается резервный. Времена службы блоков показательно распределены со средними значениями θ_1 и θ_2 . Выборочные испытания для 10 приборов показали средний срок службы 35 часов и среднее квадратическое отклонение 25 часов. Оценить средние времена службы основного и резервного блоков методом моментов в предположении, что $\theta_1 > \theta_2$.
61. В группе людей, имеющих доходы с логнормальным распределением $\exp\{N(a, \sigma^2)\}$, проведено выборочное обследование. По выборке из 10 человек получен средний доход 5000 руб. при среднем квадратическом отклонении 300 руб. Найти оценки параметров a и σ^2 методом моментов. Оценить долю людей с доходами от 4500 до 5500 руб.
62. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение 5 поездок фиксировал свое время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 (мин.). Известно, что

автобус ходит с интервалами по 9 мин. Оценить θ методом моментов.

63. Ежедневный спрос на некоторый товар имеет распределение Симпсона на отрезке $[a, b]$. За 25 рабочих дней спрос составлял в среднем 100 кг с исправленной выборочной дисперсией 54 кг². Определить параметры a и b методом моментов. Оценить, сколько нужно товара, чтобы удовлетворить ежедневный спрос с вероятностью 92%.
64. Ежедневный спрос на некоторый товар имеет равномерное распределение на отрезке $[a, b]$. За период наблюдения спрос составлял в среднем N кг с исправленной выборочной дисперсией M кг². С помощью метода моментов оценить, сколько нужно товара, чтобы удовлетворить ежедневный спрос с вероятностью 90%. Решить задачу, если: а) $N = 100$, $M = 192$; б) $N = 125$, $M = 432$.
65. Известно, что доля возвратов по кредитам в банке имеет распределение $F(x) = x^{\beta}$, $0 \leq x \leq 1$. Наблюдения доли показали выборочное среднее w . Методом моментов оценить параметр β и вероятность того, что доля опустится ниже r . Решить задачу при: а) $w = 0,9$, $r = 0,75$; б) $w = 0,8$, $r = 0,6$.
66. Срок службы некоторого изделия имеет распределение Вейбулла с параметром $\alpha = 2$. Наблюдения показали, что в среднем он составляет N часов. Найти вероятность того, что изделие прослужит не менее M часов. Решить задачу, если: а) $N = 1000$, $M = 1500$; б) $N = 1500$, $M = 2000$.
67. Случайные относительные изменения цены акции за день описываются распределением Лапласа. Наблюдения показали, что выборочное среднее квадратическое отклонение составляет 1%. Оценить границы, в которых относительное изменение цены акции за день находится с вероятностью 90%.
68. Случайные относительные изменения цены акции за день описываются распределением Лапласа. Наблюдения показали, что выборочное среднее абсолютное значение составляет 2%. Оценить границы, в которых относительное изменение цены акции за день находится с вероятностью 95%.
69. В июне ежедневный спрос на мороженое в киоске составляет в среднем 700 порций с выборочным средним квадратическим отклонением 50 порций. Оценить количество порций, удовлетворяющих потребность на один день с вероятностью 95% (используя нормальное приближение).

70. Рукопись проверяют два редактора, независимо друг от друга. Один нашел 70 ошибок, другой — 50, причем 25 найденных ошибок были одни и те же (т.е. они обнаружены обоими редакторами). Оценить число ошибок, которые они еще не нашли.
71. В поселке Полтинниково все жители имеют доход не менее 50 тыс. руб. в месяц. Выборочное обследование доходов 10 человек дало следующие результаты: 54; 60; 59; 79; 71; 92; 53; 54; 78; 56 (тыс. руб.). В предположении, что случайная величина дохода имеет распределение Парето с $x_0 = 50$ (тыс. руб), оценить параметр a и средний доход жителей методом максимального правдоподобия. Оценить долю жителей с доходами свыше 100 тыс. руб.
72. Известно, что некоторая работа занимает время, состоящее из обязательного периода μ и случайной задержки, распределенной показательной со средним значением θ . Хронометраж рабочего времени в 10 случаях дал следующие результаты: 32; 30; 37; 35; 42; 39; 34; 32; 31; 35 (мин.). Оценить параметры μ и θ методом максимального правдоподобия. Оценить срок, за который работа будет выполнена с вероятностью 99%.
73. Известно, что некоторая работа занимает время, состоящее из постоянного периода и случайной задержки, распределенной показательной. Хронометраж рабочего времени показал, что работа занимает как минимум N мин., а в среднем — M мин. С помощью метода максимального правдоподобия оценить вероятность того, что работа будет закончена за час, а также время, за которое работа будет выполнена с вероятностью 95%. Решить задачу, если: а) $N = 32$, $M = 39$; б) $N = 36$, $M = 48$.
74. В группе людей, имеющих доходы с логнормальным распределением $\exp\{N(a, \sigma^2)\}$, проведено выборочное обследование. По выборке из 10 человек получены следующие результаты: 4722; 2907; 4974; 2763; 3659; 5493; 3161; 3175; 4521; 3698 (руб.). Найти оценки параметров a и σ^2 методом максимального правдоподобия. Оценить долю людей с доходами от 4000 до 5000 руб. на основе оценок максимального правдоподобия.
75. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение 5 поездок фиксировал свое время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 (мин). Известно, что автобус ходит с интервалами по θ мин. Оценить параметр θ методом максимального правдоподобия. Вычислить несмещенную оценку.

76. Ежедневный спрос на некоторый товар равномерно распределен на отрезке $[a, b]$. За 6 рабочих дней спрос составлял: 104; 80; 96; 120; 113; 82 (кг). Оценить a и b , используя несмещенные оценки на основе оценок максимального правдоподобия. Оценить, сколько товара нужно для удовлетворения ежедневного спроса с вероятностью 90%.
77. В таблице представлены данные о производстве электроэнергии в России за 1998–2003 гг.

Год	1998	1999	2000	2001	2002	2003
Производство, млрд кВт·ч	827	846	878	891	891	915

Провести линейную регрессию производства по годам и сделать прогноз на 2004 г.

78. В таблице представлены данные о производстве всех зерновых культур и производстве пшеницы в России за 1998–2002 гг.

Год	1998	1999	2000	2001	2002
Производство зерновых, млн т	47,9	54,7	65,5	85,2	86,6
Производство пшеницы, млн т	27,0	31,0	34,5	47,0	57,7

Провести линейную регрессию и сделать прогнозы на 2003 г. для: а) производства зерновых по годам; б) производства пшеницы по годам.

79. В условиях предыдущей задачи провести линейную регрессию производства пшеницы по производству всех зерновых культур в виде: а) $y = a + bx$; б) $y = bx$ ($a = 0$).

В обоих случаях оценить производство пшеницы, если производство всех зерновых составит 90 млн т.

80. В таблице представлены данные о годовых доходах и расходах на личное потребление (долл.), для 10 семей.

Годовой доход, долл. США	Расходы на личное потребление, долл. США
2435	2311
2354	2278
2404	2240
2381	2181

Окончание табл.

Годовой доход, долл. США	Расходы на личное потребление, долл. США
2581	2408
2529	2379
2562	2378
2624	2554
2407	2232
2448	2356

Провести линейную регрессию расходов по доходам в виде $y = bx$. Оценить параметр b . Определить величину расходов для семьи с годовым доходом 2600 долл.

Глава 14

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

14.1. Основные понятия

Точечная оценка неизвестного параметра, найденная по выборке объема n , не указывает, какую ошибку допускают, принимая вместо точного значения параметра θ его приближенное значение. Поэтому вводят интервальную оценку, которая определяется двумя числами — концами интервала, внутри которого с определенной вероятностью находится неизвестное значение параметра θ , причем границы интервала не должны зависеть от искомого параметра.

Доверительным интервалом, или интервальной оценкой, называется интервал $(\hat{\theta}_1, \hat{\theta}_2)$, который покрывает неизвестный параметр θ с заданной доверительной вероятностью $0 < \gamma < 1$ (ее называют также надежностью доверительного интервала). Часто доверительный интервал может быть представлен в виде $(\hat{\theta} - \delta, \hat{\theta} + \delta)$, тогда величина δ (половина длины интервала) называется точностью оценки (точностью доверительного интервала). При заданном значении γ точность δ зависит от объема выборки n . Понятно, что чем меньше длина доверительного интервала, тем точнее оценка.

Однако сама интервальная оценка конструируется вокруг точечной оценки, вид которой определяется законом распределения случайной величины, который зависит от неизвестного параметра, т.е. границы доверительного интервала зависят от значения неизвестного параметра θ , следовательно, пользовать-

ся такими границами нельзя. Существуют два подхода к преодолению этой трудности.

1. Классический метод состоит в искусственном подборе статистик $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \dots, x_n)$, $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \dots, x_n)$, не зависящих от неизвестного параметра и таких, что $\hat{\theta}_1 < \hat{\theta}_2$ при любых $x_1, x_2, \dots, x_n$, θ , и $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = \gamma$.

2. Другой метод основан на асимптотических свойствах оценок (как правило, на асимптотической нормальности). Поэтому такой метод дает приближенные результаты и пригоден только при достаточно больших объемах выборок.

В обоих случаях бывает, что границы доверительных интервалов, построенных формально из каких-то теоретических соображений, выходят за рамки возможного (например, становятся отрицательными для положительных по смыслу величин). В таких случаях их «округляют» до разумных пределов.

14.2. Точные доверительные интервалы

Точные доверительные интервалы строятся, как правило, в предположении нормальности данных. Следует понимать, что реальные данные, на основании которых мы строим эти интервалы, могут совсем не выглядеть нормальными (например, это целые положительные числа, в то время как нормальное распределение непрерывно и рассредоточено по всей действительной прямой). Тем не менее широкое практическое применение описываемых методов дает неплохие результаты (это объясняется, в частности, асимптотической нормальностью оценок), и такое несоответствие не должно нас смущать в дальнейшем.

Предположим, что наблюдается случайная величина $\xi \in N(a, \sigma^2)$. Для параметров строятся следующие точные доверительные интервалы.

1. Для неизвестного среднего a при известной дисперсии σ^2 :

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}} u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

2. Для неизвестного среднего a при неизвестной дисперсии σ^2 :

$$\bar{x} - \frac{s}{\sqrt{n}} t_\gamma < a < \bar{x} + \frac{s}{\sqrt{n}} t_\gamma,$$

где t_γ — критическая точка распределения Стьюдента (для двусторонней области) с $n - 1$ степенями свободы и уровнем значимости $\alpha = 1 - \gamma$.

3. Для неизвестной дисперсии σ^2 :

$$\frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}, n-1}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2},$$

где $\chi_{\frac{\alpha}{2}, n-1}^2$ — критические точки χ^2 -распределения с $n - 1$ степенями свободы и соответствующими уровнями значимости, $\alpha = 1 - \gamma$.

Доказательство. 1. По теореме Фишера (см. 12.8) имеем $\bar{x} \in N(a, \sigma^2/n)$. Случайная величина $\eta = \frac{\bar{x} - a}{\sigma}$ тогда имеет стандартное нормальное распределение $N(0, 1)$ и

$$P(|\eta| < u) = P(-u < \eta < u) = \Phi(u) - \Phi(-u) = \Phi_0(u) - \Phi_0(-u) = 2\Phi_0(u).$$

Выбирая u_γ из соотношения $\Phi_0(u_\gamma) = \gamma/2$, получаем, что $P(|\eta| < u_\gamma) = \gamma$. Таким образом, с вероятностью γ верно $\left| \frac{\bar{x} - a}{\sigma} \right| < u_\gamma$, откуда следует

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_\gamma < a < \bar{x} + \frac{\sigma}{\sqrt{n}} u_\gamma.$$

2. По следствию 2 из теоремы Фишера (см. 12.8) статистика $T = \frac{\bar{x} - a}{s/\sqrt{n}}$ имеет распределение Стьюдента с $n - 1$ степенями свободы. Заметим, что критическая точка распределения Стьюдента (для двусторонней области) определяется таким образом, чтобы выполнялось $P(|T| > t_{\alpha/2}) = \alpha$. Выбирая $\alpha = 1 - \gamma$, получаем $P(|T| < t_{\alpha/2}) = \gamma$. Таким образом, с вероятностью γ верно $\left| \frac{\bar{x} - a}{s/\sqrt{n}} \right| < t_\gamma$, откуда следует

$$\bar{x} - \frac{s}{\sqrt{n}} t_\gamma < a < \bar{x} + \frac{s}{\sqrt{n}} t_\gamma.$$

3. По теореме Фишера имеем $\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$. Заметим, что критические точки распределения $\chi^2_{p,n-1}$ определяются из условия $P(\chi^2 > \chi^2_{p,n-1}) = p$. Выбирая $\alpha = 1 - \gamma$, получаем

$$P(\chi^2_{1-\frac{\alpha}{2}, n-1} < \chi^2 < \chi^2_{\frac{\alpha}{2}, n-1}) = P(\chi^2 > \chi^2_{1-\frac{\alpha}{2}, n-1}) - P(\chi^2 > \chi^2_{\frac{\alpha}{2}, n-1}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha = \gamma.$$

Таким образом, с вероятностью γ верно $\chi^2_{1-\frac{\alpha}{2}, n-1} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}, n-1}$, откуда следует

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}.$$

Можно также по выборке $x_1, x_2, \dots, x_n$ построить доверительный интервал для следующего, $(n+1)$ -го наблюдения (т.е. определить границы, в которых оно будет лежать с заданной вероятностью γ). А именно имеем

$$\bar{x} - st_r \sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_r \sqrt{1 + \frac{1}{n}}.$$

Понятно, что это может быть полезно в качестве прогноза на будущее.

Доказательство. Заметим, что случайные величины x_{n+1} и $\bar{x}$ независимы, причем $x_{n+1} \in N(a, \sigma^2)$ и $\bar{x} \in N(a, \sigma^2/n)$, следовательно, $x_{n+1} - \bar{x} \in N\left(0, \left(1 + \frac{1}{n}\right)\sigma^2\right)$. Тогда величина $\frac{x_{n+1} - \bar{x}}{\sigma \sqrt{1 + \frac{1}{n}}}$ имеет стандартное нормальное распределение $N(0, 1)$.

Введем статистику

$$T^* = \frac{x_{n+1} - \bar{x}}{\sigma \sqrt{1 + \frac{1}{n}}} \bigg/ \sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}}.$$

Поскольку $\frac{(n-1)s^2}{\sigma^2} \in \chi^2_{n-1}$, статистика T имеет распределение Стьюдента с $n - 1$ степенями свободы. Рассуждая как

ранее в доказательстве 2, получаем, что с вероятностью γ верно $\left| \frac{x_{n+1} - \bar{x}}{s \sqrt{1 + \frac{1}{n}}} \right| < t_\gamma$, откуда следует $\bar{x} - st_\gamma \sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_\gamma \sqrt{1 + \frac{1}{n}}$.

Задача 1. Найти минимальный объем выборки, при котором с надежностью 0,925 точность оценки математического ожидания нормально распределенной случайной величины (по выборочному среднему $\bar{x}$) равна $\delta = 0,2$, если известно среднее квадратическое отклонение $\sigma = 1,5$.

Решение. Формула, определяющая точность оценки математического ожидания по выборочному среднему, выглядит следующим образом: $\delta = \frac{u_\gamma \sigma}{\sqrt{n}}$. Отсюда следует $n = \frac{u_\gamma^2 \sigma^2}{\delta^2}$ (при этом обычно n округляют в большую сторону для надежности). По таблице функции Лапласа находим u_γ для данного примера, учитывая, что функция принимает значение $\Phi_0(u_\gamma) = 0,925/2 = 0,4625$. Таким образом, $u_\gamma = 1,78$. Подставив данные задачи, получим искомый объем выборки: $n = \frac{1,78^2 \cdot 1,5^2}{0,2^2} \approx 178,22$. Принимаем округленно $n = 179$.

Задача 2. Из генеральной совокупности извлечена выборка объема $n = 12$:

Варианта x_i	-0,5	-0,4	-0,2	0	0,2	0,6	0,8	1	1,2	1,5
Частота n_i	1	2	1	1	1	1	1	1	2	1

Оценить с надежностью 0,95 математическое ожидание a нормально распределенной случайной величины с помощью доверительного интервала.

Решение. Найдем выборочное среднее $\bar{x}$ и исправленное выборочное среднее квадратическое отклонение s . Пусть $u_\gamma = 10x_\gamma$, тогда

$$\bar{u} = \frac{1}{n} \sum_{i=1}^{10} n_i u_i = 4,2; \quad \bar{x} = \frac{\bar{u}}{10} = 0,42;$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^{10} n_i x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{10} n_i x_i \right)^2 \right] \approx 0,52.$$

Находим для уровня значимости $\alpha = 0,05$ и числа степеней свободы $n - 1 = 11$ по таблице распределения Стьюдента кри-

тическую точку $t_\gamma = 2,23$ и определяем границы доверительного интервала:

$$\bar{x} - \frac{t_\gamma s}{\sqrt{n}} = 0,42 - \frac{2,23 \cdot 0,72}{\sqrt{12}} = -0,04; \quad \bar{x} + \frac{t_\gamma s}{\sqrt{n}} = 0,42 + \frac{2,23 \cdot 0,72}{\sqrt{12}} = 0,88.$$

Таким образом, искомый доверительный интервал:
 $-0,04 < a < 0,88$.

Задача 3. Для отрасли, включающей 1200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что исправленное среднее квадратическое отклонение для числа работающих на фирме составляет $s = 25$ (человек). Построить 90%-ный доверительный интервал для среднего квадратического отклонения числа работающих на фирме по всей отрасли.

Решение. Доверительный интервал для параметра σ имеет вид

$$s \sqrt{\frac{(n-1)}{\chi_{\frac{\alpha}{2}, n-1}^2}} < \sigma < s \sqrt{\frac{(n-1)}{\chi_{1-\frac{\alpha}{2}, n-1}^2}},$$

где $\chi_{\frac{\alpha}{2}, n-1}^2$ и $\chi_{1-\frac{\alpha}{2}, n-1}^2$ находят по таблице критических точек распределения хи-квадрат. По таблице определяем для данного примера $\chi_{0,05;18}^2 = 28,9$; $\chi_{0,95;18}^2 = 9,39$. Подставив в формулу необходимые величины, получаем искомый доверительный интервал: $25\sqrt{18/28,9} < \sigma < 25\sqrt{18/9,39}$, откуда $19,74 < \sigma < 34,61$ (человек).

Задача 4. За последние 5 лет годовой рост цены актива А составлял в среднем 20% со средним квадратическим отклонением (исправленным) 5%. Построить доверительный интервал с вероятностью 95% для цены актива в конце года, если в начале года она равна 100 денежных единиц.

Решение. Рассмотрим величины относительного прироста цены актива за год. Будем пользоваться нормальным приближением¹. Применяем формулу

$$\bar{x} - st_\gamma \sqrt{1 + \frac{1}{n}} < x_{n+1} < \bar{x} + st_\gamma \sqrt{1 + \frac{1}{n}},$$

¹ Такое приближение и соответствующие оценки являются довольно грубыми. На практике распределение относительного прироста цены обычно далеко от нормального и, к сожалению, не описывается ни одной из классических формул.

где t_γ находим из таблицы критических точек распределения Стьюдента (для двусторонней области): $t_\gamma = t_{\text{кр}}(0,05; 4) = 2,78$.

Получаем $0,2 - 0,05 \cdot 2,78 \sqrt{1,2} < x_0 < 0,2 + 0,05 \cdot 2,78 \sqrt{1,2}$, откуда $0,05 < x_0 < 0,35$. Таким образом, цена актива в следующем году составит от 105 до 135 денежных единиц.

Помимо случаев построения доверительных интервалов для параметров одной выборки иногда рассматривают и случай двух выборок. Например, когда имеются две выборки $x_1, x_2, \dots, x_n$

и $y_1, y_2, \dots, y_m$ из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$ соответственно, и надо построить доверительный интервал для разности средних.

1. При известных дисперсиях

$$\bar{x} - \bar{y} - \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_\gamma < a_1 - a_2 < \bar{x} - \bar{y} + \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

Доказательство основано на следствии 3 теоремы Фишера (см. § 12.8).

2. При неизвестных, но равных дисперсиях

$$\begin{aligned} \bar{x} - \bar{y} - \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} \sqrt{\frac{n+m}{nm}} t_\gamma < a_1 - a_2 < \bar{x} - \bar{y} + \\ + \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} \sqrt{\frac{n+m}{nm}} t_\gamma, \end{aligned}$$

где t_γ — критическая точка распределения Стьюдента (для двусторонней области) с $n + m - 2$ степенями свободы и уровнем значимости $\alpha = 1 - \gamma$.

Доказательство основано на следствии 4 теоремы Фишера.

14.3. Асимптотические доверительные интервалы

Асимптотическим доверительным интервалом при оценивании параметра θ называется такой интервал $(\hat{\theta}_1, \hat{\theta}_2)$, что $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \rightarrow \gamma$ (при $n \rightarrow \infty$).

Предположим, мы решили учитывать тот факт, что наши наблюдения имеют распределение, отличное от нормального. Пусть это распределение зависит только от одного параметра θ ,

для которого надо построить доверительный интервал, а также известно, что оценка $\hat{\theta}_n$ асимптотически нормальна и верно $\hat{\theta} \xrightarrow{\epsilon} N(\theta, \sigma^2(\theta)/n)$ при $n \rightarrow \infty$. Приведем два метода построения асимптотических доверительных интервалов.

1. Подстановка оценки параметра в формулу для дисперсии. Получаем:

$$\hat{\theta} - \frac{\sigma(\hat{\theta})}{\sqrt{n}} u_\gamma < \theta < \hat{\theta} + \frac{\sigma(\hat{\theta})}{\sqrt{n}} u_\gamma,$$

где u_γ определяется из соотношения $\Phi_0(u_\gamma) = \gamma/2$.

2. Использование функционального преобразования. Определим функцию

$$g(u) = \int \frac{du}{\sigma(u)}, \text{ тогда верно } g(\hat{\theta}) \xrightarrow{\epsilon} N(g(\theta), 1/n) \text{ при } n \rightarrow \infty.$$

Следовательно, можно использовать асимптотическое неравенство

$$g(\hat{\theta}) - \frac{u_\gamma}{\sqrt{n}} < g(\theta) < g(\hat{\theta}) + \frac{u_\gamma}{\sqrt{n}}.$$

Решая его относительно θ , получаем асимптотический доверительный интервал.

Например, доверительный интервал для вероятности «успеха» p в n испытаниях Бернулли обычно строится первым методом. Поскольку $M\xi = p$ и $D\xi = p(1-p)$, то из ЦПТ получаем $w \xrightarrow{\epsilon} N(p, p(1-p)/n)$, и доверительный интервал имеет вид

$$w - \sqrt{\frac{w(1-w)}{n}} u_\gamma < p < w + \sqrt{\frac{w(1-w)}{n}} u_\gamma,$$

где w — относительная частота события.

Асимптотические доверительные интервалы рекомендуется применять, когда объем выборки достаточно велик (порядка сотен и более).

Доказательство (обоснование метода функционального преобразования). Если оценка $\hat{\theta}_n$ асимптотически нормальна и верно $\hat{\theta} \xrightarrow{\epsilon} N(\theta, \sigma^2(\theta)/n)$, то для любой функции g , гладкой в окрестности θ , верно $g(\hat{\theta}) \xrightarrow{\epsilon} N(g(\theta), (g'(\theta))^2 \sigma^2(\theta)/n)$.

Чтобы дисперсия не зависела от θ , необходимо функцию g подобрать так, чтобы выражение $\sigma(\theta)g'(\theta)$ было постоянным, например, положив $\sigma(\theta)g'(\theta) = 1$.

Задача свелась к выбору такой функции g , чтобы она являлась решением дифференциального уравнения $g'(\theta) = 1/\sigma(\theta)$.

Решение этого уравнения имеет вид $g(\theta) = \int \frac{d\theta}{\sigma(\theta)}$, при этом произвольная постоянная в неопределенном интеграле выбирается из соображений простоты окончательного выражения и обычно полагается равной нулю. Тогда получается $g(\hat{\theta}) \xrightarrow{E} N(g(\theta), 1/n)$ при $n \rightarrow \infty$.

Задача свелась к построению доверительного интервала для математического ожидания нормально распределенной случайной величины при известной дисперсии с заданной надежностью γ . Согласно ранее доказанному такой интервал имеет вид

$$g(\hat{\theta}_n) - \frac{1}{\sqrt{n}} u_\gamma < g(\theta) < g(\hat{\theta}_n) + \frac{1}{\sqrt{n}} u_\gamma.$$

Задача 5. Произведено 300 испытаний, в каждом из которых неизвестная вероятность p появления события A постоянна. Событие A появилось в 250 испытаниях. Найти доверительный интервал, покрывающий неизвестную вероятность p с надежностью 0,95.

Решение. Число испытаний $n = 300$ достаточно велико, поэтому можем воспользоваться следующими формулами для границ доверительного интервала:

$$p_1 = w - \sqrt{\frac{w(1-w)}{n}} u_\gamma; \quad p_2 = w + \sqrt{\frac{w(1-w)}{n}} u_\gamma.$$

Значение u_γ находим из соотношения $\Phi_0(u_\gamma) = \gamma/2 = 0,475$ по таблице функции Лапласа, в данном случае $u_\gamma = 1,96$. Относительная частота события A составляет $w = 250/300 \approx 0,83$. Подставим это значение w в формулы для p_1 и p_2 :

$$p_1 = 0,83 - 1,96 \sqrt{\frac{0,83 \cdot 0,17}{300}} \approx 0,79; \quad p_2 = 0,83 + 1,96 \sqrt{\frac{0,83 \cdot 0,17}{300}} \approx 0,87.$$

Итак, получаем искомый доверительный интервал:
 $0,79 < p < 0,87$.

Задача 6. Построить доверительный интервал для вероятности «успеха» в испытаниях Бернулли методом функционального преобразования.

Решение. Необходимо построить функцию $g(u) = \int \frac{du}{\sigma(u)}$, где $\sigma(p) = \sqrt{p(1-p)}$.

Для рассматриваемого случая имеем

$$\begin{aligned} g(p) &= \int \frac{dp}{\sqrt{p(1-p)}} = \left| p = t^2, dp = 2tdt \right| = 2 \int \frac{tdt}{\sqrt{t^2(1-t^2)}} = 2 \int \frac{dt}{\sqrt{1-t^2}} = \\ &= 2 \arcsin t = 2 \arcsin \sqrt{p}. \end{aligned}$$

С учетом вида функции $g(p)$ асимптотическое неравенство примет вид

$$2 \arcsin \sqrt{w} - \frac{u_\gamma}{\sqrt{n}} < 2 \arcsin \sqrt{p} < 2 \arcsin \sqrt{w} + \frac{u_\gamma}{\sqrt{n}}.$$

Для функции $y = g(x) = 2 \arcsin \sqrt{x}$ при $0 \leq x \leq 1$ обратной функцией будет $x = g^{-1}(y) = \sin^2(y/2)$, где $0 \leq y \leq \pi$. Поэтому если выполняются неравенства

$$2 \arcsin \sqrt{w} - \frac{u_\gamma}{\sqrt{n}} \geq 0, \quad 2 \arcsin \sqrt{w} + \frac{u_\gamma}{\sqrt{n}} \leq \pi,$$

то, применив обратное преобразование g^{-1} , получим доверительный интервал

$$\sin^2\left(\arcsin \sqrt{w} - \frac{u_\gamma}{2\sqrt{n}}\right) < p < \sin^2\left(\arcsin \sqrt{w} + \frac{u_\gamma}{2\sqrt{n}}\right).$$

Замечание. Очевидно, первым способом получается гораздо более простой и удобный в применении доверительный интервал, хотя и менее точный.

Задача 7. Построить асимптотический доверительный интервал для параметра λ показательного закона распределения (двумя способами).

Решение. В данном случае удобно перейти к новому параметру $\theta = 1/\lambda$. Эффективной оценкой для него является выборочное среднее $\hat{\theta} = \bar{x}$. Имеем $M\hat{\theta} = \theta$, $D\hat{\theta} = \theta^2/n$, откуда $\sigma(\theta) = \theta$.

Первым способом, подставляя оценку в формулу для дисперсии, получаем

$$\bar{x} - \frac{\bar{x}}{\sqrt{n}} u_\gamma < \theta < \bar{x} + \frac{\bar{x}}{\sqrt{n}} u_\gamma.$$

Возвращаясь к параметру $\lambda = 1/\theta$, приходим к неравенству

$$\frac{1}{\bar{x}} \left(1 + \frac{u_\gamma}{\sqrt{n}} \right)^{-1} < \lambda < \frac{1}{\bar{x}} \left(1 - \frac{u_\gamma}{\sqrt{n}} \right)^{-1}.$$

Вторым способом, определив функцию $g(u) = \int \frac{du}{u} = \ln u$, получаем асимптотическое неравенство

$$\ln \bar{x} - \frac{u_\gamma}{\sqrt{n}} < \ln \theta < \ln \bar{x} + \frac{u_\gamma}{\sqrt{n}},$$

откуда следует

$$\bar{x} \exp \left(-\frac{u_\gamma}{\sqrt{n}} \right) < \theta < \bar{x} \exp \left(\frac{u_\gamma}{\sqrt{n}} \right)$$

и

$$\frac{1}{\bar{x}} \exp \left(-\frac{u_\gamma}{\sqrt{n}} \right) < \lambda < \frac{1}{\bar{x}} \exp \left(\frac{u_\gamma}{\sqrt{n}} \right).$$

Заметим, что по ходу решения задачи получены доверительные интервалы для математического ожидания θ показательного распределения. Этот результат имеет и самостоятельную ценность.

В дальнейших задачах асимптотические интервалы по умолчанию строятся методом функционального преобразования (кроме задач на вероятности «успехов»).

14.4. Интервальная оценка коэффициента корреляции

Пусть $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ — независимые наблюдения над двумерной нормальной случайной величиной. Построим асимптотический доверительный интервал для коэффициента корреляции ρ , соответствующий надежности γ .

Точечной оценкой методом моментов для коэффициента корреляции является выборочный коэффициент корреляции

$$\hat{\rho}_n = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right]}}.$$

Известно, что оценка асимптотически нормальна со следующими параметрами:

$$\hat{\rho}_n \xrightarrow{\epsilon} N\left(\rho, \frac{(1-\rho^2)^2}{n}\right), \quad n \rightarrow \infty.$$

Отсюда следует, что $\sigma(\rho) = 1 - \rho^2$ и

$$g(\rho) = \int \frac{d\rho}{1-\rho^2} = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} = \operatorname{arth} \rho.$$

Получено так называемое *z-преобразование Фишера* для коэффициента корреляции. Это преобразование хорошо исследовано, и для него известны следующие соотношения:

$$M(\operatorname{arth} \hat{\rho}_n) \approx \operatorname{arth} \rho + \frac{\rho}{2(n-1)}, \quad D(\operatorname{arth} \hat{\rho}_n) = \frac{1}{n-3}.$$

Построим доверительный интервал для $\operatorname{arth} \rho$. Заменяя n на $n-3$ и пренебрегая в математическом ожидании при достаточно большом n величиной $\frac{\rho}{2(n-1)}$, получаем

$$P(\operatorname{arth} \hat{\rho}_n - \frac{1}{\sqrt{n-3}} u_\gamma < \operatorname{arth} \rho < \operatorname{arth} \hat{\rho}_n + \frac{1}{\sqrt{n-3}} u_\gamma) = \gamma.$$

Функция $\operatorname{th} x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (гиперболический тангенс), обратная функции $\operatorname{arth} x$, однозначна и монотонно возрастает, поэтому после преобразования доверительный интервал имеет вид

$$\operatorname{th}(\operatorname{arth} \hat{\rho}_n - \frac{1}{\sqrt{n-3}} u_\gamma) < \rho < \operatorname{th}(\operatorname{arth} \hat{\rho}_n + \frac{1}{\sqrt{n-3}} u_\gamma),$$

где u_γ определяется соотношением $\Phi_0(u_\gamma) = \gamma/2$.

Найденный приближенный доверительный интервал настолько мало отличается от истинного, что может применяться уже для выборок объема $n \geq 10$.

Задачи для самостоятельного решения¹

Теоретические задачи

1. Построить асимптотический доверительный интервал для параметра λ распределения Пуассона: $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.
2. Построить асимптотический доверительный интервал для параметра θ случайной величины ξ , равномерно распределенной на отрезке $[0, \theta]$. Точечную оценку параметра θ найти методом моментов.
3. Построить асимптотический доверительный интервал для неизвестного параметра λ при известном α , если случайная величина ξ имеет гамма-распределение: $\xi \in \gamma(\alpha, \lambda)$. *Указание:* перейти к новому параметру $\theta = \alpha/\lambda$.
4. Построить асимптотический доверительный интервал для неизвестного параметра α при известном λ , если случайная величина ξ имеет гамма-распределение: $\xi \in \gamma(\alpha, \lambda)$.
5. Построить асимптотический доверительный интервал для числа степеней свободы распределения хи-квадрат.
6. Построить асимптотический доверительный интервал для p с известным числом испытаний m : $P_m(k) = C_m^k p^k (1-p)^{m-k}$ (двумя способами).
7. Построить асимптотический доверительный интервал для параметра θ (математического ожидания) геометрического распределения с $p = 1/\theta$ (методом подстановки).
8. Построить доверительный интервал для параметра α распределения Парето вида

$$F(x) = \begin{cases} 0, & x < 1; \\ 1 - x^{-\alpha}, & x \geq 1, \end{cases}$$
 при $\alpha > 2$ (методом подстановки).
Указание: перейти к новому параметру $\theta = \alpha/(\alpha - 1)$.
9. Построить асимптотический доверительный интервал для параметра θ нормального распределения $N(\theta, \theta^4)$, $\theta > 0$.
10. Построить асимптотический доверительный интервал для параметра θ нормального распределения $N(\theta, \theta^6)$, $\theta > 0$.

¹ В задачах выборочная дисперсия и выборочное среднее квадратическое отклонение предполагаются исправленными по умолчанию, если не сказано обратное.

Вычислительные задачи

11. Импортёр упаковывает чай в пакеты. Известно, что наполняющая машина работает со средним квадратическим отклонением, равным 1 г. Выборка 50 пакетов показала средний вес 125,18 г. Найти доверительный интервал для среднего веса пакета с вероятностью 95%. Найти объем выборки n , при котором точность интервала составила бы 0,2 г. Генеральная совокупность распределена нормально.
12. Автомат упаковывает товар в пакеты со средним квадратическим отклонением веса, равным A г. Выборка 100 пакетов показала средний вес B г. Найти доверительный интервал для среднего веса пакета с вероятностью 95%. Решить задачу, если: а) $A = 1,5$, $B = 100,2$; б) $A = 1$, $B = 100,5$; в) $A = 0,8$, $B = 100,1$.
13. Автомат упаковывает товар в пакеты со средним квадратическим отклонением веса, равным A г. Выборка N пакетов показала средний вес B г. Найти доверительный интервал для среднего веса пакета с вероятностью 90%. Решить задачу, если: а) $A = 1$, $N = 50$, $B = 100,2$; б) $A = 1,2$, $N = 40$, $B = 100,4$; в) $A = 0,8$, $N = 100$, $B = 99,95$.
14. Станок-автомат штампует валики. По выборке объема $n = 100$ вычислено выборочное среднее диаметров изготовленных валиков. Найти с надежностью 0,95 точность δ , с которой выборочное среднее оценивает математическое ожидание диаметров изготавливаемых валиков, зная, что их среднее квадратическое отклонение $\sigma = 2$ мм. Предполагается, что диаметры валиков распределены нормально.
15. Фирма коммунального хозяйства желает на основе выборки оценить среднюю квартплату за квартиры определенного типа с надежностью 99% и точностью до 10 денежных единиц (д.е.). Предполагая, что квартплата имеет нормальное распределение со средним квадратическим отклонением 35 д.е., найдите необходимый объем выборки.
16. Для исследования средней заработной платы работающих жителей города сделана выборка из N человек, при этом оказалось, что средняя заработная плата в выборке равна A у.е., а выборочная дисперсия B . С надежностью 95% найти доверительный интервал для средней заработной платы жителей города. Определить, какой объем выборки нужен, чтобы достигнуть точности в 1 у.е. Решить задачу, если: а) $N = 121$, $A = 300$, $B = 324$; б) $N = 256$, $A = 250$, $B = 289$; в) $N = 400$, $A = 280$, $B = 625$.

17. Для отрасли, включающей 1200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что в фирме в среднем работают 77,5 человека при выборочном среднем квадратическом отклонении 25 человек. Пользуясь 95%-ным доверительным интервалом, оцените среднее число работающих в фирме по всей отрасли и общее число работающих в отрасли. Предполагается, что случайное число работников фирмы описывается нормальным распределением.
18. Производитель автомобильных шин заинтересован в получении оценки средней износоустойчивости шин особой модели. Он произвел случайную выборку объемом 10 шин и подверг их специальному испытанию. Средняя износоустойчивость по данным выборки оказалась равна 22 500 миль с неисправленным средним квадратическим отклонением 3000 миль. Построить доверительный интервал с вероятностью 99% для средней износоустойчивости всего выпуска шин этого типа. Генеральная совокупность распределена нормально.
19. По данным испытаний на износ 25 автомобильных шин выборочная средняя износоустойчивость составила N тысяч километров с выборочным средним квадратическим отклонением M тысячи километров. Построить доверительный интервал для средней износоустойчивости с надежностью 99%. Решить задачу, если: а) $N = 27$, $M = 4$; б) $N = 32$, $M = 5$.
20. В деревне Простоквашино проведено выборочное исследование доходов. По выборке из N человек получен выборочный средний доход A рублей с выборочным средним квадратическим отклонением B руб. Построить доверительный интервал для среднего дохода с надежностью 95%. Решить задачу, если: а) $N = 25$, $A = 4750$, $B = 150$; б) $N = 16$, $A = 4250$, $B = 125$; в) $N = 25$, $A = 4760$, $B = 230$.
21. Бухгалтер компании решил предпринять выборочную проверку и выбрал 18 из 1200 компонент, продававшихся в прошлом месяце. Стоимости отобранных компонент: 82, 30, 98, 116, 80, 150, 200, 88, 70, 90, 160, 100, 86, 76, 90, 140, 76, 68 (д.е.). Найти оценку средней стоимости всех компонент и построить для нее доверительный интервал с надежностью 0,95. Оценить, какой объем выборки необходим для достижения точности $\delta = 3$ (д.е.).
22. Среднее время сборки изделия было 90 мин. Инженер изобрел новый метод сборки изделия. При испытаниях времена сборки 10 изделий новым способом составили: 79, 74, 112, 95, 83, 96,

- 77, 84, 70, 90 (мин.). Построить доверительный интервал для нового среднего времени сборки с надежностью 95%.
23. При измерении веса 20 шоколадных батончиков получены выборочное среднее 50,01 г и выборочное среднее квадратическое отклонение 1,02 г. Найти доверительный интервал для среднего веса с надежностью 95%.
24. Произведено 12 измерений одним прибором (без систематической ошибки) некоторой физической величины, причем выборочное среднее квадратическое отклонение случайных ошибок измерения оказалось равным 0,6. Найти точность (среднее квадратическое отклонение) прибора с надежностью 0,95. Предполагается, что результаты измерений распределены нормально.
25. Выборочное среднее квадратическое отклонение ежесуточного дохода случайно выбранных 10 магазинов некоторой фирмы оказалось равно 100 д.е. Построить доверительный интервал для среднего квадратического отклонения с надежностью 90%. Предполагается, что доход — нормально распределенная величина.
26. Выборочное среднее квадратическое отклонение ежесуточного дохода 30 киосков некоторой фирмы оказалось равно N д.е. Построить доверительный интервал для среднего квадратического отклонения с надежностью 90%. Решить задачу, если: а) $N = 80$; б) $N = 75$.
27. Проведены испытания новой машины, упаковывающей товар в пакеты. По выборке из 30 пакетов получено выборочное среднее квадратическое отклонение веса A г. Построить доверительный интервал с надежностью $B\%$ для дисперсии веса пакета при упаковке машиной. Решить задачу, если: а) $A = 2$, $B = 90$; б) $A = 1,5$, $B = 90$; в) $A = 0,7$, $B = 95$; г) $A = 0,4$, $B = 95$.
28. Проведены испытания новой машины, упаковывающей товар в пакеты. По выборке из N пакетов получено выборочное среднее квадратическое отклонение веса A г. Построить доверительный интервал с надежностью $B\%$ для среднего квадратического отклонения веса пакета при упаковке машиной. Решить задачу, если: а) $N = 16$, $A = 3$, $B = 90$; б) $N = 20$, $A = 4$, $B = 90$; в) $N = 25$, $A = 1,5$, $B = 95$; г) $N = 20$, $A = 0,9$, $B = 90$; д) $N = 25$, $A = 0,5$, $B = 95$.
29. По выборке из 25 упаковок товара средний вес составил 101 г с выборочным средним квадратическим отклонением 3 г.

- Построить доверительные интервалы для среднего и дисперсии с вероятностью 90%.
30. За последние 9 лет годовой рост цены актива *A* составлял в среднем 22% с выборочным средним квадратическим отклонением 6%. Построить доверительный интервал с вероятностью 90% для средней цены актива в конце года, если в начале года она равна 200 денежным единицам.
 31. За последние 7 лет годовой рост цены актива *B* составлял в среднем 15% с выборочным средним квадратическим отклонением 3%. Построить доверительный интервал с вероятностью 95% для средней цены актива в конце года, если в начале года она равна 120 денежным единицам.
 32. Производство зерна в России в 1996–2002 гг. составляло в среднем 71,11 млн т в год с выборочным средним квадратическим отклонением 16,26 млн т. Построить доверительные интервалы для среднего объема производства и для следующего наблюдения с надежностью 95%, используя нормальное приближение.
 33. Производство пшеницы в России в 1995–2001 гг. составляло в среднем 35,54 млн т с выборочным средним квадратическим отклонением 7,45 млн т. Построить доверительные интервалы для среднего объема производства и для следующего наблюдения с надежностью 95%, используя нормальное приближение.
 34. Урожайность зерновых культур в России в 1992–2001 гг. составляла в среднем 15,85 ц/га с выборочным средним квадратическим отклонением 2,17 ц/га. Построить доверительные интервалы для средней урожайности и для следующего наблюдения с надежностью 95%, используя нормальное приближение.
 35. По результатам наблюдений за 100 рабочих дней, ежедневные доходы магазина *A* составляли в среднем 150 д.е. с выборочным средним квадратическим отклонением 26 д.е., а магазина *B* — в среднем 120 д.е. с выборочным средним квадратическим отклонением 32 д.е. Приняв оценки дисперсий за их точные значения, построить доверительный интервал для разности средних ежесуточных доходов магазинов *A* и *B* с надежностью 95%.
 36. По результатам наблюдений за 21 рабочий день, ежедневные доходы магазина *A* составляли в среднем 140 д.е. с выборочным средним квадратическим отклонением 22 д.е., а магазина *B* — в среднем 100 д.е. с выборочным средним квадратическим отклонением 24 д.е. В предположении, что неизвестные дис-

персии равны, построить доверительный интервал для разности средних ежесуточных доходов магазинов A и B с надежностью 90%.

37. Изготовлен экспериментальный игровой автомат, который должен обеспечить появление выигрыша в одном случае из 100 бросаний монеты в автомат. Для проверки пригодности автомата произведено 400 испытаний, причем выигрыш появился 5 раз. Построить доверительный интервал, покрывающий неизвестную вероятность появления выигрыша с надежностью $\gamma = 0,99$.
38. При испытаниях 1000 элементов зарегистрировано 100 отказов. Найти доверительный интервал, покрывающий неизвестную вероятность p отказа элемента с надежностью 0,95.
39. В случайной выборке из 300 аспирантов, специализирующихся по управлению предприятиями, составленной по нескольким основным университетам, 180 человек оказались детьми бизнесменов. Построить доверительные интервалы с надежностью 90% для доли аспирантов, являющихся детьми бизнесменов, и числа таких аспирантов среди 2000 аспирантов.
40. Среди 100 электрических ламп в течение 1000 часов вышло из строя 36 штук. Построить доверительный интервал с надежностью 95% для вероятности того, что лампа прослужит не менее 1000 часов.
41. Среди N изделий в течение года вышли из строя M . Построить доверительный интервал с надежностью $K\%$ для вероятности того, что изделие прослужит не менее года. Решить задачу, если: а) $N = 100$, $M = 25$, $K = 95$; б) $N = 200$, $M = 25$, $K = 90$; в) $N = 100$, $M = 20$, $K = 95$; г) $N = 250$, $M = 45$, $K = 90$; д) $N = 400$, $M = 50$, $K = 95$.
42. Фирма разослала 1000 новых рекламных каталогов и получила N заказов. Построить доверительный интервал для эффективности рекламы (вероятности отклика) с надежностью $M\%$. Решить задачу, если: а) $N = 120$, $M = 95$; б) $N = 140$, $M = 90$; в) $N = 180$, $M = 95$; г) $N = 130$, $M = 90$.
43. По выборке из 400 зрителей рейтинг телепрограммы оценивается в 9%. Построить доверительный интервал для истинного рейтинга с надежностью 95%.
44. Проведена случайная выборка заемных счетов в банке. Из 1000 отобранных счетов 60 оказались с задолженностью по возврату ссуды сроком до трех месяцев. Найти доверительный интервал

- с вероятностью 90% для доли счетов в банке, которые имеют задолженности до трех месяцев.
45. В ходе аудиторской проверки была проведена случайная выборка записей по счетам. В выборке из 500 записей 10 содержали некоторые ошибки в самой записи или процедуре. Найти доверительный интервал для доли ошибок во всей генеральной совокупности с надежностью 0,95. Оценить объем выборки, которую надо произвести аудитору, чтобы определить долю ошибок с точностью до 0,5% и надежностью 95%.
 46. Согласно социологическому опросу среди 100 человек, 20 из них пользуются стиральным порошком фирмы А. Сколько еще людей надо опросить, чтобы с вероятностью 99% получить результат (долю людей, пользующихся этим стиральным порошком) с точностью до 1%?
 47. По данным социологического опроса среди 100 человек 25 из них пользовались шампунем «Колобок». Оценить, сколько всего человек надо опросить, чтобы определить долю пользующихся данным шампунем с точностью до 2% и надежностью 95%.
 48. По данным социологического опроса среди 100 человек 16 из них пользовались моющим средством «Мойлодыр». Оценить, сколько всего человек надо опросить, чтобы определить долю пользующихся данным моющим средством с точностью до 1% с надежностью 90%.
 49. В выборке из 100 изделий средний срок службы изделия составил 1000 часов. Построить доверительный интервал для среднего срока службы изделия с вероятностью 90%. Время службы изделия распределено по показательному закону.
 50. В выборке из 400 изделий средний срок службы составил N часов. Известно, что срок службы изделия имеет показательное распределение. Построить доверительный интервал для среднего срока службы изделия с вероятностью 95%. Решить задачу, если: а) $N = 1500$; б) $N = 2000$; в) $N = 2500$.
 51. За 100 рабочих дней в магазин обращалось в среднем N покупателей в день. Известно, что число покупателей в день имеет распределение Пуассона. Построить доверительный интервал для среднего числа покупателей с вероятностью $M\%$. Решить задачу, если: а) $N = 256$, $M = 99$; б) $N = 289$, $M = 90$.
 52. Число семян сорняков в пробе зерна распределено по закону Пуассона. Выборка из 1000 проб зерна показала в среднем

0,9 семян в пробе. Построить доверительный интервал для среднего числа семян с надежностью 95%.

53. В поселке Соткино все жители имеют доход не менее 100 тыс. руб. в месяц. Выборочное обследование доходов 100 человек дало средний доход 150 тыс. руб. В предположении, что случайная величина дохода имеет распределение Парето вида

$$F(x) = \begin{cases} 0, & x < x_0; \\ 1 - (x/x_0)^{-\alpha}, & x \geq x_0, \end{cases}$$

где $x_0 = 100$ (тыс. руб.), построить доверительный интервал для параметра α с надежностью 90%.

54. Анализ выборки из 100 заказов в интернет-магазине показал, что в среднем делается по K покупок. Построить доверительный интервал для математического ожидания θ с надежностью 95% в предположении, что число покупок в заказе имеет геометрическое распределение. Решить задачу методом подстановки, если: а) $K = 1,8$; б) $K = 2,5$.
55. По выборке из N наблюдений двумерной нормальной случайной величины получен выборочный коэффициент корреляции A . Построить доверительный интервал для коэффициента корреляции с надежностью 90%. Решить задачу, если: а) $N = 50$, $A = 0,5$; б) $N = 100$, $A = 0,7$.

Глава 15

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

15.1. Основные определения

Статистической гипотезой называется любое предположение относительно генеральной совокупности. Гипотеза называется **параметрической**, если в ней содержится некоторое утверждение о параметрах распределения случайной величины (когда сам закон распределения считается известным) и **непараметрической** в иных случаях. В этой главе будем иметь дело с параметрическими гипотезами.

Метод использования выборки для проверки истинности статистической гипотезы называется статистическим доказательством истинности выдвинутой гипотезы.

Наряду с выдвинутой гипотезой рассматривают альтернативную гипотезу. Если выдвинутая гипотеза отвергается, то вместо нее принимается альтернативная. Поэтому гипотезы делятся на нулевые и альтернативные.

Нулевой (основной) гипотезой H_0 называется предположение, которого мы придерживаемся изначально. При проверке параметрических гипотез в качестве основной гипотезы обычно предполагается равенство одного или нескольких параметров заданным значениям (или между собой).

Альтернативной (конкурирующей) гипотезой H_1 называется гипотеза, которая противоречит основной гипотезе H_0 и которую мы принимаем, если отвергаем основную гипотезу. При этом предполагается, что другие возможности (кроме H_0 и H_1) исключены.

Для проверки нулевой гипотезы используют специально подобранную случайную величину — выборочную статистику, точное или приближенное распределение которой известно.

Случайная величина K , построенная по наблюдениям для проверки нулевой гипотезы, называется **статистикой критерия**.

Схема построения критерия такова: все выборочное пространство делится на две взаимодополняющие области — область S отклонения основной гипотезы H_0 и область $\bar{S}$ принятия этой гипотезы. Область S , при попадании в которую выборочной точки $x = (x_1, x_2, \dots, x_n)$ основная гипотеза отклоняется, называется **критической**.

При проверке гипотез могут быть ошибки двух типов.

Ошибка первого рода состоит в том, что основная гипотеза отвергается, хотя на самом деле она верна. Ее вероятность обозначают обычно α .

Ошибка второго рода состоит в том, что основная гипотеза принимается, хотя на самом деле она неверна. Ее вероятность обозначают обычно β .

Вероятность α совершить ошибку первого рода называют также **уровнем значимости**, или **размером критерия**. Вероятность $1 - \beta$ не совершить ошибку второго рода называют **мощностью критерия**.

Критерий называется **наиболее мощным**, если из всех возможных критериев с заданным уровнем значимости α он обладает наибольшей мощностью.

Пример. Пусть определена статистика критерия K и пусть функция плотности вероятностей выборочной статистики K при условии истинности нулевой гипотезы H_0 равна $p(K|H_0)$, медиана K равна K_0 . По заданному уровню значимости α определяют квантили $K_{\alpha/2}$ и $K_{1-\alpha/2}$ из условия

$$P(K \leq K_{\alpha/2}) = \int_{-\infty}^{K_{\alpha/2}} p(K|H_0) dK = \alpha/2;$$

$$P(K \geq K_{1-\alpha/2}) = \int_{K_{1-\alpha/2}}^{\infty} p(K|H_0) dK = \alpha/2,$$

где α полагают достаточно малым, чтобы попадание случайной величины K за пределы интервала $(K_{\alpha/2}; K_{1-\alpha/2})$ можно было считать маловероятным событием. Область $(K_{\alpha/2}; K_{1-\alpha/2})$ и является областью допустимых значений, т.е. областью принятия нулевой гипотезы. Промежутки $(-\infty; K_{\alpha/2})$ и $(K_{1-\alpha/2}; +\infty)$ образуют критическую область критерия, при попадании в которую наблюдаемого значения K нулевую гипотезу отвергают. Точки, отделяющие критические области от области принятия гипотезы, называются **критическими точками** (рис. 15.1).

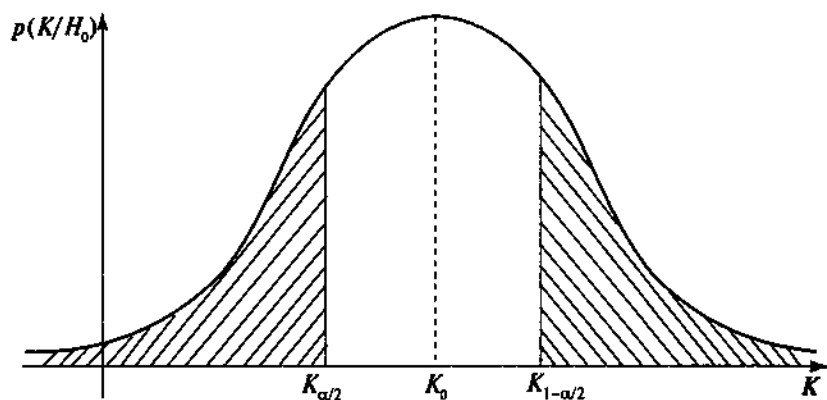


Рис. 15.1

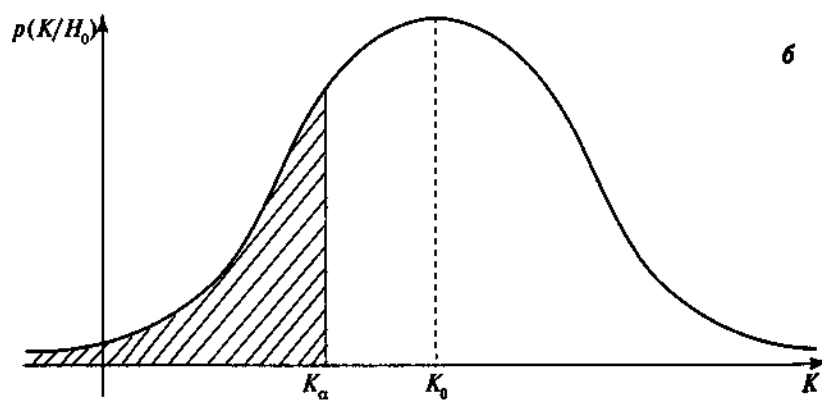
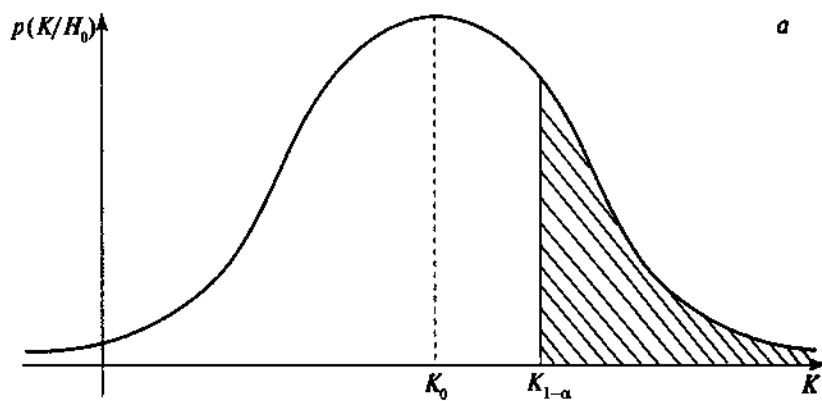


Рис. 15.2

Критическая область называется *двусторонней*, если она располагается слева и справа от медианы K_0 (см. рис. 15.1), *правосторонней*, если $P(K > K_{1-\alpha}) = \alpha$, и *левосторонней*, если $P(K < K_\alpha) = \alpha$ (рис. 15.2).

Итак, основной принцип проверки статистической гипотезы можно сформулировать так: если наблюдаемое значение статистики критерия принадлежит критической области, нулевую гипотезу отвергают; если наблюдаемое значение принадлежит области допустимых значений, нулевую гипотезу принимают.

Заметим также, что можно установить связь между задачами проверки гипотез и задачами построения доверительных интервалов (см. гл. 14). Например, пусть построен доверительный интервал $(\hat{\theta}_1, \hat{\theta}_2)$ для параметра θ с надежностью γ . Тем самым утверждается, что истинное значение параметра лежит в интервале с вероятностью γ , а вне этого интервала — с малой вероятностью $\alpha = 1 - \gamma$. Таким образом, если мы проверяем гипотезу $H_0: \theta = \theta_0$ против какой-либо альтернативной гипотезы, то в качестве критерия можно взять $S = \{x: \theta_0 \notin (\hat{\theta}_1, \hat{\theta}_2)\}$. Уровень значимости в этом случае будет равен α . Однако такой критерий не обязательно будет оптимальным.

15.2. Критерий отношения правдоподобия

Основным методом построения наиболее мощных статистических критериев (по теореме Неймана—Пирсона) является **метод отношения правдоподобия**, суть которого заключается в следующем: пусть ξ — непрерывная случайная величина, имеющая плотность распределения $p_0(x)$ при условии истинности нулевой гипотезы H_0 и $p_1(x)$ при выполнении гипотезы H_1 . Функции правдоподобия в точке x соответственно равны

$$L_0(x) = p_0(x_1)p_0(x_2)\dots p_0(x_n); \quad L_1(x) = p_1(x_1)p_1(x_2)\dots p_1(x_n).$$

О правдоподобию выборки в отношении гипотез H_0 и H_1 будем судить по отношению правдоподобия L_1/L_0 ($L_0 \neq 0$): чем правдоподобнее выборка в условиях истинности гипотезы H_0 , тем меньше L_1 по сравнению с L_0 , и тем меньше отношение L_1/L_0 .

► **Теорема 1 (Неймана—Пирсона).** Критическая область S наиболее мощного критерия имеет вид

$$S = \left\{ x: L_0(x) = 0 \cup \frac{L_1(x)}{L_0(x)} > C, \quad L_0(x) \neq 0 \right\},$$

где константа $C = C(\alpha)$ является решением уравнения¹.

$$P\left(\frac{L_1(x)}{L_0(x)} > C \mid H_0\right) = \alpha.$$

Подобный метод построения критической области, использующий отношение правдоподобия, дает нам **критерий отношения правдоподобия**. В дискретном случае построение проводится аналогично (только вместо плотностей берутся вероятности).

Статистика критерия здесь имеет вид $K = L_1/L_0$ (при $L_0 = 0$ полагаем $K = +\infty$), тогда критическая область $S = \{K > C\}$, а область допустимых значений $\bar{S} = \{K \leq C\}$.

Задача 1. Пусть случайная величина $\xi \in N(a, \sigma^2)$, причем значение параметра a неизвестно, а дисперсия σ^2 известна. Требуется на уровне значимости α проверить нулевую гипотезу $H_0: a = a_0$, если альтернативная гипотеза $H_1: a = a_1 > a_0$. Построить критерий отношения правдоподобия. Вычислить объем выборки n , необходимый для достижения ошибки второго рода, равной β , при уровне значимости α .

Решение. Если верна гипотеза H_0 , т.е. $\xi \in N(a_0, \sigma^2)$, то функция правдоподобия в точке $x = (x_1, x_2, \dots, x_n)$ равна

$$L_0(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i - a_0)^2}{2\sigma^2}\right).$$

Если же верна гипотеза H_1 , т.е. $\xi \in N(a_1, \sigma^2)$, то функция правдоподобия равна

$$L_1(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i - a_1)^2}{2\sigma^2}\right).$$

Отношение правдоподобия имеет вид

$$\frac{L_1}{L_0} = \exp\left(\frac{(a_1 - a_0)(2\bar{x} - a_1 - a_0)n}{2\sigma^2}\right).$$

Поскольку $a_1 > a_0$, то это отношение является монотонно возрастающей функцией от $\bar{x}$, и поскольку $L_0(x) \neq 0$, то неравен-

¹ Бывает, что данное уравнение не имеет решения, и в таких случаях используются так называемые *рандомизированные критерии*, которые мы здесь рассматривать не будем.

ство $L_1/L_0 > C$ равносильно неравенству $\bar{x} > \bar{C}$, где C и $\bar{C}$ — некоторые константы. Поэтому критическая область имеет вид $S = \{x: \bar{x} > \bar{C}\}$, где $P(\bar{x} > \bar{C} | H_0) = \alpha$.

При условии истинности нулевой гипотезы H_0 имеем $\bar{x} \in N(a_0, \frac{\sigma^2}{n})$, поэтому

$$\alpha = P(\bar{x} > \bar{C} | H_0) = 1 - \Phi\left(\frac{\bar{C} - a_0}{\sigma} \sqrt{n}\right) = \frac{1}{2} - \Phi_0\left(\frac{\bar{C} - a_0}{\sigma} \sqrt{n}\right).$$

$$\text{Отсюда } \Phi_0\left(\frac{\bar{C} - a_0}{\sigma} \sqrt{n}\right) = \frac{1}{2} - \alpha.$$

Обозначим через u_α решение уравнения $\Phi_0(u_\alpha) = 1/2 - \alpha$, тогда константа имеет вид $\bar{C} = a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$. Заметим, что величина u_α является квантилью уровня $1 - \alpha$ для стандартного нормального распределения и выступает здесь в качестве критической точки. Значение ее можно найти по таблице функции Лапласа.

Итак, наиболее мощным критерием проверки гипотез: $H_0: a = a_0$ при альтернативной $H_1: a = a_1 > a_0$ оказывается следующий:

если $\bar{x} < a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$, то H_0 принимается;

если $\bar{x} > a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}$, то H_0 отклоняется (и принимается H_1).

По определению ошибка второго рода равна

$$\beta = P(\bar{x} \leq \bar{C} | H_1) = \Phi\left(\frac{\bar{C} - a_1}{\sigma} \sqrt{n}\right). \text{ Отсюда } \Phi\left(\frac{a_1 - \bar{C}}{\sigma} \sqrt{n}\right) = 1 - \beta.$$

Получаем, что должно выполняться равенство

$$\bar{C} = a_1 - u_\beta \frac{\sigma}{\sqrt{n}} = a_0 + u_\alpha \frac{\sigma}{\sqrt{n}}.$$

Решая уравнение относительно n , получаем объем выборки, необходимый для получения ошибки второго рода β , при заданном уровне значимости α :

$$n = \frac{(u_\alpha + u_\beta)^2}{(a_1 - a_0)^2} \sigma^2.$$

Полученное значение обычно округляется до целого в большую сторону, для уменьшения вероятностей ошибок.

Мощность критерия в данном случае равняется

$$1 - \beta = \Phi\left(\frac{a_1 - \bar{C}}{\sigma} \sqrt{n}\right) = \Phi\left(\frac{a_1 - a_0}{\sigma} \sqrt{n} - u_\alpha\right).$$

Задача 2. Крупная торговая фирма желает открыть в новом районе города филиал. Известно, что фирма будет работать прибыльно, если средний месячный доход жителей района превышает 400 долл. Также известно, что среднее квадратическое отклонение дохода σ составляет 20 долл. Проводится выборочное обследование населения по величине доходов, чтобы принять решение об открытии филиала.

1) Определите правило принятия решения, с помощью которого, основываясь на выборке $n = 100$ (человек) и уровне значимости $\alpha = 0,05$, можно установить, что филиал будет работать прибыльно.

2) Рассчитайте вероятность того, что при применении правила принятия решения, полученного при ответе на вопрос п. 1, будет совершена ошибка второго рода, если в действительности средний доход достигает 406 долл.

3) Считая альтернативное значение генерального среднего месячного дохода равным 410 долл., рассчитайте объем выборки, при котором ошибка первого рода не превысит 2,5%, а ошибка второго рода не превысит 5%.

Решение. 1) Фирма не откроет филиал, если средний доход жителей не превысит 400 долл. Поэтому будем считать, что $H_0: a = a_0 = 400$, а $H_1: a = a_1 > 400$. Значение дисперсии σ^2 дохода известно. Находим u_α по таблице функции Лапласа исходя из равенства $\Phi_0(u_\alpha) = 1/2 - \alpha$. Поскольку $a_0 = 400$ и $u_\alpha = 1,65$, то H_1 принимают и, следовательно, филиал открывают, если средний месячный доход 100 жителей

$$\bar{x} > 400 + \frac{20}{\sqrt{100}} \cdot 1,65 = 403,3.$$

2) Вероятность ошибки второго рода β можно легко найти, зная мощность критерия, которая равна $1 - \beta$. Получаем

$$1 - \beta = \Phi\left(\frac{406 - 400}{20} \sqrt{100} - 1,65\right) = \Phi(1,35) = \frac{1}{2} + \Phi_0(1,35) \approx 0,91.$$

Отсюда вероятность ошибки второго рода $\beta \approx 0,09$.

3) Используем формулу для необходимого объема выборки. Для этого найдем квантили u_α и u_β с помощью таблицы функции Лапласа и получим

$$n \geq \frac{(1,96 + 1,65)^2}{100} 20^2 = 52,13, \text{ откуда } n = 53.$$

Замечание. Приведенная формула может давать слишком малые значения n , так как основывается на строго нормальном распределении. На практике n должно быть достаточно велико, чтобы пользоваться асимптотической нормальностью оценок.

15.3. Проверка гипотез для одной выборки

Рассмотрим простые методы проверки параметрических гипотез в случае нормального распределения (которые являются формально точными), а также гипотезы о вероятности «успеха» в испытаниях Бернулли (на основе асимптотической нормальности). Как и ранее, нас не будет смущать тот факт, что реальные данные, по которым проверяются гипотезы, могут совсем не выглядеть нормальными (например, это целые положительные числа, в то время как нормальное распределение непрерывно и рассредоточено по всей действительной прямой). Тем не менее широкое практическое применение описываемых методов дает неплохие результаты (это объясняется, в частности, асимптотической нормальностью оценок).

Следующие три типа гипотез проверяются для нормальных данных: $\xi \in N(a, \sigma^2)$.

1. Гипотезы о неизвестном среднем a при известной дисперсии σ^2 .

Основная гипотеза $H_0: a = a_0$, альтернативная гипотеза H_1 может быть трех видов: а) $a \neq a_0$; б) $a > a_0$; в) $a < a_0$. Во всех трех случаях для проверки используется статистика критерия

$$U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}.$$

В случае а) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = (1 - \alpha)/2$. Если $|U| < u_{кр}$, гипотеза H_0 принимается, если $|U| > u_{кр}$ — отвергается. Таким образом, в данном случае имеет место двусторонняя критическая область.

В случаях б) и в) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$.

В случае б) если $U < u_{кр}$, то гипотеза H_0 принимается, если $U > u_{кр}$ — отвергается.

В случае в) если $U > -u_{кр}$, то гипотеза H_0 принимается, если $U < -u_{кр}$ — отвергается.

Здесь имеют место односторонние критические области (*правосторонняя* и *левосторонняя* соответственно).

Замечание. Этим методом можно пользоваться и в случае неизвестной дисперсии при больших объемах выборки (порядка сотен), когда оценку дисперсии можно принять за ее точное значение.

2. Гипотезы о неизвестном среднем a при неизвестной дисперсии σ^2 .

Основная гипотеза $H_0: a = a_0$, альтернативная гипотеза H_1 может быть трех видов: а) $a \neq a_0$; б) $a > a_0$; в) $a < a_0$. Во всех трех случаях для проверки используется статистика критерия

$$T = \frac{\bar{x} - a_0}{s} \sqrt{n}.$$

Для проверки берутся критические точки $t_{кр}$ распределения Стьюдента с $n - 1$ степенями свободы и уровнем значимости α , причем в случае а) — для *двусторонней* критической области, в случаях б) и в) — для *односторонней* критической области.

В случае а) если $|T| < t_{кр}$, то гипотеза H_0 принимается, если $|T| > t_{кр}$ — отвергается.

В случае б) если $T < t_{кр}$, то гипотеза H_0 принимается, если $T > t_{кр}$ — отвергается.

В случае в) если $T > -t_{кр}$, то гипотеза H_0 принимается, если $T < -t_{кр}$ — отвергается.

3. Гипотезы о неизвестной дисперсии σ^2 .

Обычно предполагается, что хотя дисперсия неизвестна, но дана ее несмещенная оценка s^2 . Основная гипотеза $H_0: \sigma^2 = \sigma_0^2$, альтернативная гипотеза H_1 может быть трех видов: а) $\sigma^2 \neq \sigma_0^2$; б) $\sigma^2 > \sigma_0^2$; в) $\sigma^2 < \sigma_0^2$. Во всех трех случаях для проверки используется статистика критерия

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}.$$

Для проверки берутся критические точки распределения хи-квадрат с $n - 1$ степенями свободы (и различными уровнями значимости).

В случае а) если $\chi^2_{1-\alpha/2; n-1} < \chi^2 < \chi^2_{\alpha/2; n-1}$, то гипотеза H_0 принимается, иначе она отвергается. Заметим, что границы области *несимметричны* относительно оценки s^2 .

В случае б) если $\chi^2 < \chi^2_{\alpha; n-1}$, то гипотеза H_0 принимается, иначе отвергается.

В случае в) если $\chi^2 > \chi^2_{1-\alpha; n-1}$, то гипотеза H_0 принимается, иначе отвергается.

Следующая гипотеза проверяется приближенно, на основе асимптотической нормальности оценки.

4. Гипотеза о неизвестной вероятности «успеха» в испытаниях Бернулли.

Основная гипотеза $H_0: p = p_0$, альтернативная гипотеза H_1 может быть трех видов: а) $p \neq p_0$; б) $p > p_0$; в) $p < p_0$. Во всех трех случаях для проверки используется статистика критерия

$$U = \frac{w - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}, \text{ где } w \text{ — относительная частота «успехов» в}$$

n наблюдениях.

Далее критические точки и области для проверки выбираются так же, как и при проверке гипотезы о неизвестном среднем при известной дисперсии.

Замечание. Этим методом можно пользоваться только при больших объемах выборки (порядка нескольких десятков или сотен).

Доказательство. 1) Если нулевая гипотеза справедлива, статистика критерия $U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}$ имеет стандартное нормальное распределение $N(0, 1)$.

В случае справедливости гипотезы а) эта статистика может стремиться как к $+\infty$, так и к $-\infty$, в зависимости от того, $a > a_0$ или $a < a_0$. Поэтому область принятия гипотезы H_0 следует ограничить с двух сторон, справа и слева. Логично это сделать симметричным образом, в силу симметрии нормального распределения. Имеем $P(|U| < u) = 2\Phi_0(u)$, поэтому, выбирая $u_{\alpha/2}$ из условия $\Phi_0(u_{\alpha/2}) = (1 - \alpha)/2$, получаем $P(|U| < u_{\alpha/2}) = 1 - \alpha$ и $P(|U| > u_{\alpha/2}) = \alpha$, что обеспечивает уровень значимости α .

В случае справедливости гипотезы б) $a > a_0$ статистика U стремится к $+\infty$, так что область принятия гипотезы H_0 следу-

ет ограничить справа. Имеем $P(U < u) = \Phi(u) = 1/2 + \Phi_0(u)$, поэтому, выбирая $u_{кр}$ из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$, получаем $P(U < u_{кр}) = 1 - \alpha$ и $P(U > u_{кр}) = \alpha$, что обеспечивает уровень значимости α .

В случае справедливости гипотезы в) $a < a_0$ статистика U стремится к $-\infty$, так что область принятия гипотезы H_0 следует ограничить слева. Имеем $P(U > -u) = 1 - \Phi(-u) = 1/2 + \Phi_0(u)$, поэтому, выбирая $u_{кр}$ из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$, получаем $P(U > -u_{кр}) = 1 - \alpha$ и $P(U < -u_{кр}) = \alpha$, что обеспечивает уровень значимости α .

2) Если нулевая гипотеза справедлива, статистика критерия $T = \frac{\bar{x} - a_0}{s} \sqrt{n}$ имеет распределение Стьюдента с $n - 1$ степенями свободы. Критическая область выбирается в зависимости от альтернативной гипотезы из тех же соображений, что и в случае 1.

3) Если нулевая гипотеза справедлива, статистика критерия $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ имеет распределение хи-квадрат с $n - 1$ степенями свободы. Далее рассуждаем аналогично.

4) Если нулевая гипотеза справедлива, из центральной предельной теоремы (см. гл. 8) следует, что относительная частота успехов w асимптотически нормальна, с математическим ожиданием p_0 и дисперсией $p_0(1 - p_0)/n$, так что статистика

$U = \frac{w - p_0}{\sqrt{p_0(1 - p_0)}} \sqrt{n}$ сходится по распределению к стандартной

нормальной величине при $n \rightarrow \infty$. Значит, при достаточно больших n можно считать, что она имеет стандартное нормальное распределение. Далее рассуждаем, как в случае 1.

Описанные критерии проверки гипотез можно представить в виде табл. 15.1.

Задача 3. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 5$ извлечена выборка объема $n = 100$ и по ней найдено выборочное среднее 26,5. Требуется на уровне значимости 0,05 проверить гипотезу $H_0: a = a_0 = 25$ против альтернативной гипотезы $H_1: a \neq 25$. Изменится ли результат, если заменим альтернативную гипотезу на $H_1: a > 25$?

Решение. Найдем значение статистики критерия:

$$U = \frac{26,5 - 25}{5} \sqrt{100} = 3.$$

При проверке гипотезы $H_1: a \neq 25$ из соотношения $\Phi_0(u_{кр}) = 0,95/2 = 0,475$ находим $u_{кр} = 1,96$ и $|U| > u_{кр}$, так что основная гипотеза отвергается. При проверке гипотезы $H_1: a > 25$ из соотношения $\Phi_0(u_{кр}) = 0,45$ находим $u_{кр} = 1,65$ и $U > u_{кр}$, так что основная гипотеза отвергается. В обоих случаях результат одинаков.

Таблица 15.1

H_0	Предположения	Статистика критерия	H_1	Область принятия H_0
$a = a_0$	σ^2 известно	$U = \frac{\bar{x} - a_0}{\sigma} \sqrt{n}$	$a > a_0$ $a < a_0$ $a \neq a_0$	$U < u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $U > -u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $ U < u_{кр}, \Phi_0(u_{кр}) = (1 - \alpha)/2$
	σ^2 неизвестно	$T = \frac{\bar{x} - a_0}{s} \sqrt{n}$	$a > a_0$ $a < a_0$ $a \neq a_0$	$T < t_{кр}(\alpha, n - 1)$ для односторонней области $T > -t_{кр}(\alpha, n - 1)$ для односторонней области $ T < t_{кр}(\alpha, n - 1)$ для двусторонней области
$\sigma^2 = \sigma_0^2$	a неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi_{\alpha; n-1}^2$ $\chi^2 > \chi_{1-\alpha; n-1}^2$ $\chi_{1-\alpha/2; n-1}^2 < \chi^2 < \chi_{\alpha/2; n-1}^2$
$p = p_0$	n порядка нескольких десятков или сотен	$U = \frac{w - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n},$ где $w = m/n$	$p > p_0$ $p < p_0$ $p \neq p_0$	$U < u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $U > -u_{кр}, \Phi_0(u_{кр}) = 1/2 - \alpha$ $ U < u_{кр}, \Phi_0(u_{кр}) = (1 - \alpha)/2$

Задача 4. По выборке объема $n = 16$, извлеченной из нормальной генеральной совокупности, найдены выборочное среднее, равное 12,4, и исправленное среднее квадратическое отклонение, равное 1,2. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: a = 11,8$ при конкурирующей гипотезе $H_1: a \neq 11,8$.

Решение. Найдем наблюдаемое значение статистики критерия:

$$T = \frac{(\bar{x} - a_0)\sqrt{n}}{s} = \frac{(12,4 - 11,8) \cdot 4}{1,2} = 2.$$

Поскольку конкурирующая гипотеза имеет вид $a \neq a_0$, искомая критическая область двусторонняя. По таблице критических точек распределения Стьюдента найдем по уровню значимости $\alpha = 0,05$ и числу степеней свободы $n - 1 = 15$ критическую точку $t_{\text{кр}} = t_{\text{кр}}(0,05; 15) = 2,13$.

В силу того что $|T| < t_{\text{кр}}$, нет оснований отвергнуть нулевую гипотезу.

Задача 5. Точность работы станка-автомата проверяется по дисперсии размеров изделий, которая не должна превышать $\sigma_0^2 = 0,01$ (мм²). По выборке из 25 изделий получена исправленная выборочная дисперсия $s^2 = 0,02$ (мм²). На уровне значимости 0,05 проверить, обеспечивает ли станок необходимую точность.

Решение. Найдем значение статистики критерия:

$$\chi^2 = \frac{24 \cdot 0,02}{0,01} = 48.$$

Альтернативной гипотезой в данном случае является $H_1: \sigma^2 > \sigma_0^2$.

По таблице находим критическую точку распределения хи-квадрат: $\chi_{0,05;24}^2 = 36,4$.

Поскольку $48 > 36,4$, то основная гипотеза отвергается. Следовательно, станок не обеспечивает необходимой точности.

Задача 6. Из нормальной генеральной совокупности извлечена выборка объема $n = 31$. В следующей таблице представлены сгруппированные данные.

Вариант x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
Частота n_i	1	3	7	10	6	3	1

Требуется на уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma^2 = 0,18$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 0,18$.

Решение. Перейдем к условным вариантам $u_i = 10x_i - 110$.

u_i	-9	-7	-4	2	5	8	10
n_i	1	3	7	10	6	3	1

Получаем:

$$s_u^2 = \frac{1}{n-1} \left[\sum_{i=1}^7 n_i u_i^2 - \frac{1}{n} \left(\sum_{i=1}^7 n_i u_i \right)^2 \right] = \frac{1}{30} \left(822 - \frac{26^2}{31} \right) \approx 26,67;$$

$$s_x^2 = \frac{s_u^2}{100} \approx 0,27; \quad \chi^2 = \frac{(n-1)s_x^2}{\sigma_0^2} = \frac{30 \cdot 0,27}{0,18} = 45.$$

По таблице критических точек распределения хи-квадрат по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = n - 1 = 30$ находим $\chi_{кр}^2(0,05; 30) = 43,8$.

Поскольку $45 > 43,8$, основная гипотеза H_0 отвергается.

Задача 7. Партия изделий принимается, если доля брака составляет не более 2%. Среди случайно отобранных 500 изделий оказалось 13 бракованных. Следует ли принять партию (на уровне значимости 0,05)?

Решение. Относительная частота брака составляет $w = 13/500 = 0,026$.

Найдем значение статистики критерия

$$U = \frac{0,026 - 0,02}{\sqrt{0,02 \cdot 0,98}} \sqrt{500} \approx 0,96.$$

Альтернативной гипотезой в данном случае является $H_1: p > p_0$.

Из соотношения $\Phi_0(u_{кр}) = 0,45$ находим $u_{кр} = 1,65$ и получаем $U < u_{кр}$, так что основная гипотеза принимается. Таким образом, партию изделий можно принять.

Задача 8. Торговец утверждает, что он получает заказы в среднем, по крайней мере, от 30% предполагаемых клиентов. Можно ли при 5%-ном уровне значимости считать это утверждение неверным, если торговец получил заказы от 20 из 100 случайно отобранных потенциальных клиентов?

Решение. В данном случае нулевая гипотеза будет выглядеть как $H_0: p = p_0 = 0,3$, а конкурирующая — как $H_1: p < 0,3$.

Найдем значение статистики критерия, учитывая, что относительная частота для данной задачи равна $w = 20/100 = 0,2$:

$$U = \frac{(w - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} = \frac{(0,2 - 0,3) \cdot 10}{\sqrt{0,3 \cdot 0,7}} \approx -2,18.$$

Из соотношения $\Phi_0(u_{\text{кр}}) = 0,45$ находим $u_{\text{кр}} = 1,65$, и $U < -u_{\text{кр}}$, так что нулевая гипотеза отвергается и с утверждением торговца согласиться нельзя.

Задача 9. По выборке объема n , извлеченной из нормальной генеральной совокупности с известным средним квадратическим отклонением σ , найдено выборочное среднее. При уровне значимости α требуется найти функцию мощности критерия проверки нулевой гипотезы $H_0: a = a_0$ при конкурирующей гипотезе $H_1: a = a_1 \neq a_0$.

Решение. Поскольку конкурирующая гипотеза имеет вид $a \neq a_0$, критическая область двусторонняя. Она определяется неравенством: $|U| > u_{\text{кр}}$, где $u_{\text{кр}}$ находится из соотношения $\Phi_0(u_{\text{кр}}) = 1/2 - \alpha$.

Найдем мощность рассматриваемого критерия, что по определению есть вероятность попадания статистики критерия в критическую область при допущении, что справедлива конкурирующая гипотеза:

$$1 - \beta = P\left\{\left|\frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}\right| > u_{\text{кр}} \mid a = a_1\right\}.$$

Преобразуем выражение, стоящее под знаком модуля:

$$\frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(\bar{x} - a_1)\sqrt{n}}{\sigma} + \frac{(a_1 - a_0)\sqrt{n}}{\sigma} = b + \lambda,$$

где $b = \frac{(\bar{x} - a_1)\sqrt{n}}{\sigma}$, $\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}$.

Используя данные соотношения, найдем мощность критерия:

$$\begin{aligned} 1 - \beta &= P(|b + \lambda| > u_{\text{кр}}) = P(b + \lambda > u_{\text{кр}}) + P(b + \lambda < -u_{\text{кр}}) = \\ &= P(b > u_{\text{кр}} - \lambda) + P(b < -u_{\text{кр}} - \lambda) = [1 - P(b < u_{\text{кр}} - \lambda)] + \end{aligned}$$

$$+ P(b < -u_{\text{кр}} - \lambda) = [1/2 - \Phi_0(u_{\text{кр}} - \lambda)] + [1/2 + \Phi_0(-u_{\text{кр}} - \lambda)] = \\ = 1 - \Phi_0(u_{\text{кр}} - \lambda) - \Phi_0(u_{\text{кр}} + \lambda).$$

Поскольку каждому значению a_1 будет соответствовать свое значение мощности критерия (в силу того, что λ есть функция от a_1), то мощность критерия также является функцией от a_1 . Обозначив мощность критерия через π , получим

$$\pi(a_1) = 1 - [\Phi_0(u_{\text{кр}} - \lambda) + \Phi_0(u_{\text{кр}} + \lambda)].$$

15.4. Проверка гипотез для двух выборок.

Зависимые выборки: парные наблюдения

Под случаем «зависимых выборок» обычно имеют в виду ситуацию, когда речь идет об одном и том же наборе объектов до и после какого-либо воздействия на них. Предполагается, что воздействие может повлиять на признаки, сдвинув их средние значения в большую или меньшую сторону, и это необходимо проверить.

Вначале признаки объектов принимают значения x_i после воздействия — значения y_i . Такие наблюдения называются **парными**. Вычислим их разности: $d_i = y_i - x_i$. Тогда ставится следующая задача: по наблюдениям $d_1, d_2, \dots, d_n$ проверить гипотезу о равенстве нулю генерального среднего ($H_0: a_d = 0$) при неизвестной дисперсии σ_d^2 . Предполагается, что случайные изменения признаков распределены нормально. Тогда гипотезу можно проверить, как это описано в предыдущем параграфе.

Задача 10. Физическая подготовка 9 спортсменов была проверена при поступлении в спортивную школу, а затем после недели тренировок. Итоги проверки в баллах приведены в таблице (в первой строке указано число баллов, полученных каждым спортсменом при поступлении в школу; во второй строке — после обучения).

x_i	76	71	57	49	70	69	26	65	59
y_i	81	85	52	52	70	63	33	83	62

Требуется на уровне значимости 0,05 проверить, существенно ли улучшилась физическая подготовка спортсменов (используя нормальное приближение).

Решение. Вычислим разности $d_i = y_i - x_i$

d_i	5	14	-5	3	0	-6	7	18	3

Получаем

$$\sum_{i=1}^9 d_i = 5 + 14 - 5 + 3 - 6 + 7 + 18 + 3 = 39;$$

$$\sum_{i=1}^9 d_i^2 = 25 + 196 + 25 + 9 + 36 + 49 + 324 + 9 = 673.$$

Отсюда находим выборочное среднее и исправленное среднее квадратическое отклонение:

$$\bar{d} = \frac{1}{9} \sum_{i=1}^9 d_i = \frac{39}{9} \approx 4,33;$$

$$s_d = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n d_i^2 - \frac{1}{n} \left(\sum_{i=1}^n d_i \right)^2 \right]} = \sqrt{\frac{673 - 169}{8}} \approx 7,94.$$

Проверяем гипотезу $H_0: a_d = 0$ против $H_1: a_d > 0$.

Найдем значение статистики критерия:

$$T = \frac{\bar{d} \sqrt{n}}{s_d} = \frac{4,33 \cdot 3}{7,94} \approx 1,47.$$

По таблице критических точек распределения Стьюдента для односторонней области по уровню значимости 0,05 и числу степеней свободы $k = n - 1 = 8$ определяем $t_{кр}(0,05; 8) = 1,86$. Поскольку $T < t_{кр}$, нет оснований отвергнуть нулевую гипотезу.

Таким образом, нельзя утверждать, что подготовка спортсменов существенно улучшилась.

15.5. Проверка гипотез для двух выборок. Независимые выборки

Пусть имеются две независимые выборки: $x_1, x_2, \dots, x_n$ и $y_1, y_2, \dots, y_m$, имеющие нормальное распределение с параметрами (a_x, σ_x^2) и (a_y, σ_y^2) соответственно. Обычно ставится задача проверки их однородности, т.е. равенства обоих параметров, либо надо проверить равенство параметров по отдельности.

1. Гипотеза о равенстве дисперсий двух выборок.

Предположим, что известны исправленные выборочные дисперсии для обеих выборок — s_x^2 и s_y^2 . Проверяем гипотезу $H_0: \sigma_x^2 = \sigma_y^2$. Альтернативная гипотеза H_1 может быть трех видов: а) $\sigma_x^2 \neq \sigma_y^2$; б) $\sigma_x^2 > \sigma_y^2$; в) $\sigma_x^2 < \sigma_y^2$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно.

В случае а) делят *большую* выборочную дисперсию на *меньшую*:

$$F = \frac{s_{\max}^2}{s_{\min}^2}.$$

Обозначим через $n_{\min}$ объем выборки с меньшей выборочной дисперсией и через $n_{\max}$ — с большей. По таблице для распределения Фишера находим критическую точку с уровнем значимости $\alpha/2$ и числами степеней свободы $n_{\max} - 1$ и $n_{\min} - 1$. Если $F < F_{\text{кр}}$, то основная гипотеза принимается, иначе отвергается.

В случае б) делят первую выборочную дисперсию на вторую: $F = \frac{s_x^2}{s_y^2}$.

По таблице распределения Фишера находим критическую точку с уровнем значимости α и числами степеней свободы $n - 1$ и $m - 1$. Если $F < F_{\text{кр}}$, то основная гипотеза принимается, иначе отвергается.

2. Гипотеза о равенстве средних при известных дисперсиях.

Проверяем гипотезу $H_0: a_x = a_y$. Альтернативная гипотеза H_1 может быть трех видов: а) $a_x \neq a_y$; б) $a_x > a_y$; в) $a_x < a_y$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно. Во всех случаях вычисляют статистику критерия

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

В случае а) критическая точка $u_{\text{кр}}$ выбирается из условия $\Phi_0(u_{\text{кр}}) = (1-\alpha)/2$. Если $|U| < u_{\text{кр}}$, гипотеза H_0 принимается, если $|U| > u_{\text{кр}}$ — отвергается.

В случае б) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$. Если $U < u_{кр}$, то гипотеза H_0 принимается, если $U > u_{кр}$ — отвергается.

Замечание. Гипотеза о средних обычно проверяется таким образом и в случае неизвестных дисперсий для больших выборок (объемом порядка сотен), когда оценки дисперсий можно принять за их точные значения.

3. Гипотеза о равенстве средних при неизвестных равных дисперсиях.

Проверяем гипотезу $H_0: a_x = a_y$. Альтернативная гипотеза H_1 может быть трех видов: а) $a_x \neq a_y$; б) $a_x > a_y$; в) $a_x < a_y$, однако случай в) сводится к б) перестановкой x и y и не будет рассматриваться отдельно. Во всех случаях вычисляют статистику критерия:

$$T = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}},$$

$$\text{где } s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}.$$

Величина s^2 является объединенной оценкой дисперсии (общей для выборок).

Эту же формулу можно представить в виде

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Для проверки берут критические точки $t_{кр}$ распределения Стьюдента с $n + m - 2$ степенями свободы и уровнем значимости α , причем в случае а) — для двусторонней критической области, в случае б) — для односторонней критической области.

В случае а) если $|T| < t_{кр}$, то гипотеза H_0 принимается, если $|T| > t_{кр}$ — отвергается.

В случае б) если $T < t_{кр}$, то гипотеза H_0 принимается, если $T > t_{кр}$ — отвергается.

Замечание. Поскольку для проверки гипотезы требуется равенство дисперсий у двух выборок, сначала необходимо проверить гипотезу о равенстве дисперсий. В противном случае данный метод применять нельзя.

4. Гипотеза о равенстве вероятностей «успеха» в двух сериях испытаний Бернулли.

Гипотеза проверяется на основе асимптотической нормальности относительных частот, так что данный метод может применяться только при больших объемах выборок (порядка нескольких десятков или сотен). Пусть в одной серии из n_1 испытаний получили m_1 «успехов», в другой серии из n_2 испытаний получили m_2 «успехов». Проверяем гипотезу $H_0: p_1 = p_2$. Альтернативная гипотеза может быть трех видов: а) $p_1 \neq p_2$; б) $p_1 > p_2$; в) $p_1 < p_2$, однако случай в) сводится к б) перестановкой индексов и не будет рассматриваться отдельно.

Во всех случаях вычисляют статистику критерия

$$U = \frac{w_1 - w_2}{\sqrt{w(1-w) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}},$$

где $w = \frac{m_1 + m_2}{n_1 + n_2}$.

В случае а) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = (1 - \alpha)/2$. Если $|U| < u_{кр}$, гипотеза H_0 принимается, если $|U| > u_{кр}$ — отвергается.

В случае б) критическая точка $u_{кр}$ выбирается из условия $\Phi_0(u_{кр}) = 1/2 - \alpha$. Если $U < u_{кр}$, то гипотеза H_0 принимается, если $U > u_{кр}$ — отвергается.

Доказательство. 1) Если основная гипотеза справедлива, то по следствию 5 теоремы Фишера (см. 12.8) статистика $F = \frac{s_x^2}{s_y^2}$ имеет распределение Фишера с числами степеней свободы $n - 1$ и $m - 1$.

Если справедлива гипотеза б), то F стремится к $+\infty$, так что область принятия нулевой гипотезы следует ограничить справа. Критические точки распределения Фишера определяются из условия: $P(F > F_{кр}) = \alpha$. Выбирая соответственно критическую область, получаем уровень значимости α .

Если справедлива гипотеза а), то F может стремиться как к $+\infty$, так и к 0, так что область принятия нулевой гипотезы следует ограничить и справа, и слева. Но условие $\frac{s_{\max}^2}{s_{\min}^2} < C$ эк-

вивалентно $\frac{1}{C} < \frac{s_x^2}{s_y^2} < C$ при $C > 1$. С учетом того, что для распределения Фишера верно $F_\alpha(n, m) = \frac{1}{F_{1-\alpha}(m, n)}$, получаем, что при выборе критической точки $F_{кр}$, соответствующей уровню значимости $\alpha/2$, будет выполняться $P\left(\frac{s_{\max}^2}{s_{\min}^2} > F_{кр}\right) = \alpha$, так что достигается нужный нам уровень значимости α .

2) Если основная гипотеза справедлива, то по следствию 3 теоремы Фишера статистика $U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$ имеет стандартное

нормальное распределение. Далее рассуждаем так же, как при проверке гипотезы для одной выборки (см. § 15.3).

3) Если основная гипотеза справедлива, то по следствию 4 теоремы Фишера статистика $T = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}$

имеет распределение Стьюдента с $n + m - 2$ степенями свободы. Далее рассуждаем так же, как при проверке гипотезы для одной выборки.

4) Если основная гипотеза справедлива, то из центральной предельной теоремы следует, что относительные частоты w_1 и w_2 асимптотически нормальны с одинаковым средним $p = p_1 = p_2$ и дисперсиями $p(1-p)/n_1$ и $p(1-p)/n_2$. Кроме того, они независимы. Следовательно, их разность $w_1 - w_2$ асимптотически нормальна с нулевым средним и дисперсией $p(1-p)(1/n_1 + 1/n_2)$. Однако p неизвестно. Наилучшей оценкой (несмещенной, состоятельной, эффективной в классе линейных несмещенных оценок) для p является $w = \frac{m_1 + m_2}{n_1 + n_2}$. Заменяя p на w , получаем, что распределение статистики

$$U = \frac{w_1 - w_2}{\sqrt{w(1-w)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

сходится к стандартному нормальному распределению при $n \rightarrow \infty$. Далее рассуждаем так же, как при проверке гипотезы для одной выборки.

Описанные критерии проверки гипотез можно представить в виде табл. 15.2.

Таблица 15.2

H_0	Предположения	Статистика критерия	H_1	Область принятия H_0
$\sigma_x = \sigma_y$	σ_x^2 и σ_y^2 известны	$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$	$a_x > a_y$ $a_x < a_y$ $a_x \neq a_y$	$U < u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = 1/2 - \alpha$ $U > -u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = 1/2 - \alpha$ $ U < u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = (1 - \alpha)/2$
	σ_x^2 и σ_y^2 неизвестны, но равны	$T = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}, \text{ где } s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$	$a_x > a_y$ $a_x < a_y$ $a_x \neq a_y$	$T < t_{\alpha/2}(\alpha, n+m-2)$ для односторонней области $T > -t_{\alpha/2}(\alpha, n+m-2)$ для односторонней области $ T < t_{\alpha/2}(\alpha, n+m-2)$ для двусторонней области
$\sigma_x^2 \neq \sigma_y^2$	a_x и a_y неизвестны	$F = \frac{s_x^2}{s_y^2}, \text{ где } s_x^2 > s_y^2$	$\sigma_x^2 > \sigma_y^2$ $\sigma_x^2 \neq \sigma_y^2$	$F < F_{\alpha}(\alpha, n-1, m-1)$ $F < F_{\alpha/2}(\alpha/2, n-1, m-1)$
$p_1 = p_2$	n_1 и n_2 порядка нескольких десятков или сотен	$U = \frac{w_1 - w_2}{\sqrt{w(1-w)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ где } w = \frac{m_1 + m_2}{n_1 + n_2}$	$p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$	$U < u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = 1/2 - \alpha$ $U > -u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = 1/2 - \alpha$ $ U < u_{\alpha/2}, \Phi_0(u_{\alpha/2}) = (1 - \alpha)/2$

Задача 11. По выборке объема $n = 30$ найден средний вес изделий, равный 130 г, изготовленных на первом станке; по выборке объема $m = 40$ найден средней вес изделий, равный 125 г, изготовленных на втором станке. Генеральные дисперсии известны: $\sigma_x^2 = 60 \text{ г}^2$, $\sigma_y^2 = 80 \text{ г}^2$. Требуется на уровне значимости 0,05 проверить нулевую гипотезу $H_0: a_x = a_y$ при конкурирующей гипотезе $H_1: a_x \neq a_y$. Предполагается, что случайные величины распределены нормально и выборки независимы.

Решение. Найдем значение статистики критерия:

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} = \frac{130 - 125}{\sqrt{\frac{60}{30} + \frac{80}{40}}} = 2,5.$$

По таблице функции Лапласа найдем критическую точку из равенства $\Phi(u_{\text{кр}}) = (1-\alpha)/2 = 0,475$, получаем $u_{\text{кр}} = 1,96$. Так как $|U| > u_{\text{кр}}$, гипотеза H_0 отвергается. Таким образом, нельзя утверждать, что средние значения веса изделий двух станков совпадают.

Задача 12. По двум независимым выборкам, объемы которых $n = 9$ и $m = 16$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии $s_x^2 = 34,02$ и $s_y^2 = 12,15$. На уровне значимости 0,01 проверить нулевую гипотезу $H_0: \sigma_x^2 = \sigma_y^2$ против конкурирующей гипотезы $H_1: \sigma_x^2 > \sigma_y^2$.

Решение. Рассчитаем значение статистики критерия:

$$F = \frac{s_x^2}{s_y^2} = \frac{34,02}{12,15} = 2,8.$$

Числа степеней свободы $k_1 = n - 1 = 8$, $k_2 = m - 1 = 15$. По таблице критических точек распределения Фишера—Снедекора по заданному уровню значимости $\alpha = 0,01$ и числам степеней свободы находим $F_{\text{кр}}(0,01; 8; 15) = 4$. Поскольку $F < F_{\text{кр}}$, нулевая гипотеза принимается.

Задача 13. Реклама утверждает, что из двух типов пластиковых карт «Русский Экспресс» и «Супер-Понт» богатые люди предпочитают первый. С целью проверки этого утверждения были обследованы среднемесячные платежи $n = 16$ обладателей «Русского Экспресса» и $m = 11$ обладателей «Супер-Понта». Выяснилось, что платежи по картам «Русский Экспресс» составляют в среднем 563 долл. с исправленным средним квадратическим отклонением 178 долл., а по картам «Супер-Понт» — в среднем 485 долл. с исправленным средним квадратическим отклонением 196 долл.

Предварительный анализ законов распределения месячных расходов как среди обладателей «Русского Экспресса», так и

среди обладателей «Супер-Понта» показал, что они достаточно хорошо описываются нормальным приближением.

Проверить утверждение рекламы на уровне значимости 10%.

Решение. В данном случае речь идет о проверке гипотезы о средних при неизвестных дисперсиях (объемы выборок малы). Поэтому прежде всего необходимо проверить гипотезу о равенстве дисперсий, а лишь затем двигаться дальше. Имеем

$$F = \frac{s_{\max}^2}{s_{\min}^2} = \frac{196^2}{178^2} = \frac{38\,416}{31\,684} \approx 1,21.$$

По таблице критических точек распределения Фишера—Снедекора по уровню значимости $\alpha/2 = 0,05$ и числам степеней свободы $k_1 = n_{\max} - 1 = 10$ и $k_2 = n_{\min} - 1 = 15$ найдем критическую точку $F_{\text{кр}} = 2,55$. Так как $1,21 < 2,55$, принимаем гипотезу о равенстве дисперсий двух выборок.

Теперь можем воспользоваться критерием Стьюдента для проверки гипотезы о равенстве средних. Имеем

$$s^2 = \frac{10 \cdot 38\,416 + 15 \cdot 31\,684}{25} = 34\,376,8; \quad s \approx 185,4.$$

Вычисление статистики критерия дает:

$$T = \frac{563 - 485}{185,4 \sqrt{1/11 + 1/16}} \approx 1,074.$$

Из таблиц критических точек распределения Стьюдента (для односторонней области) по уровню значимости $\alpha = 0,1$ и числу степеней свободы 25 находим $t_{\text{кр}} = 1,32$.

Поскольку $T < t_{\text{кр}}$, принимается основная гипотеза (о равенстве средних). Таким образом, утверждение рекламы не подтверждается имеющимися данными.

Задача 14. В партии из 500 деталей, изготовленных первым станком-автоматом, оказалось 60 нестандартных; из 600 деталей второго станка — 42 нестандартных. На уровне значимости $\alpha = 0,01$ проверить нулевую гипотезу $H_0: p_1 = p_2$ о равенстве вероятностей изготовления нестандартной детали обоими станками против конкурирующей гипотезы $H_1: p_1 \neq p_2$.

Решение. Имеем $w_1 = 60/500 = 0,12$; $w_2 = 42/600 = 0,07$; $w = (60 + 42)/(500 + 600) \approx 0,09$.

Найдем значение статистики критерия

$$U = \frac{0,12 - 0,07}{\sqrt{0,09 \cdot 0,91 \left(\frac{1}{500} + \frac{1}{600} \right)}} \approx 2,85.$$

Найдем критическую точку из соотношения: $\Phi_0(u_{кр}) = 0,495$, откуда $u_{кр} = 2,57$. Поскольку $|U| > u_{кр}$, нулевая гипотеза отвергается. Значит, вероятности изготовления нестандартных деталей на двух станках различны.

15.6. Проверка гипотез о равенстве дисперсий для нескольких выборок. Критерии Бартлетта и Кокрена

Пусть генеральные совокупности $X_1, X_2, \dots, X_M$ распределены нормально. Из этих совокупностей извлечены независимые выборки объемов $n_1, n_2, \dots, n_M$. По выборкам найдены исправленные выборочные дисперсии $s_1^2, s_2^2, \dots, s_M^2$. Требуется на уровне значимости α проверить нулевую гипотезу о равенстве дисперсий для всех выборок, т.е. $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_M^2$ (против гипотезы, что какие-то из дисперсий не равны).

Опишем критерий Бартлетта, позволяющий проверить такую гипотезу.

Введем обозначения: пусть $k_i = n_i - 1$ — число степеней свободы s_i^2 ;

$$k = \sum_{i=1}^M k_i \text{ — сумма чисел степеней свободы;}$$

$$s^2 = \frac{1}{k} \sum_{i=1}^M k_i s_i^2 \text{ — среднее арифметическое исправленных дисперсий, взвешенное по степеням свободы;}$$

$$V = k \ln s^2 - \sum_{i=1}^M k_i \ln s_i^2; \quad C = 1 + \frac{1}{3(M-1)} \left(\sum_{i=1}^M \frac{1}{k_i} - \frac{1}{k} \right).$$

Статистикой критерия Бартлетта является величина $B = V/C$. При условии, что нулевая гипотеза верна, эта статистика распределена примерно как хи-квадрат с $M-1$ степенями свободы. Для применения критерия необходимо, чтобы все $n_i \geq 4$.

Если $B < \chi_{\alpha; M-1}^2$, то H_0 принимается, иначе отвергается.

Иногда нет необходимости вычислять величину C . А именно, если оказывается, что $V < \chi^2_{\alpha; M-1}$, то этого достаточно для выполнения $B < \chi^2_{\alpha; M-1}$, поскольку $C > 1$.

В случае, когда размеры выборок одинаковы, предпочтительнее использовать критерий Кокрена. Статистикой критерия Кокрена является отношение максимальной исправленной выборочной дисперсии к их сумме:

$$G = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_M^2}.$$

В предположении, что нулевая гипотеза верна, распределение этой статистики зависит только от числа степеней свободы $k = n - 1$ и числа выборок M .

По таблице критических точек распределения Кокрена находят критическую точку $G_{\text{кр}} = G_{\text{кр}}(\alpha; k; M)$. Если $G < G_{\text{кр}}$, то H_0 принимается, иначе отвергается.

Если гипотеза о равенстве дисперсий у всех выборок принимается, то в качестве оценки этой общей дисперсии можно использовать величину s^2 .

Задача 15. По четырем независимым выборкам, объемы которых $n_1 = 17$, $n_2 = 20$, $n_3 = 15$, $n_4 = 16$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии, соответственно равные 2,5; 3,6; 4,1; 5,8. Требуется: 1) на уровне значимости 0,05 проверить гипотезу о равенстве дисперсий; 2) оценить генеральную дисперсию.

Решение.

1) Вычислим вначале числитель V критерия Бартлетта:

$$V = k \ln s^2 - \sum_{i=1}^4 k_i \ln s_i^2, \quad \text{где } k = \sum_{i=1}^4 k_i = 16 + 19 + 14 + 15 = 64.$$

Имеем

$$s^2 = \frac{1}{k} \sum_{i=1}^4 k_i s_i^2 = \frac{1}{64} (16 \cdot 2,5 + 19 \cdot 3,6 + 14 \cdot 4,1 + 15 \cdot 5,8) = 3,95.$$

Следовательно, $V = 64 \cdot \ln 3,95 - (16 \cdot \ln 2,5 + 19 \cdot \ln 3,6 + 14 \cdot \ln 4,1 + 15 \cdot \ln 5,8) \approx 2,79$.

По таблице критических точек распределения хи-квадрат для уровня значимости 0,05 и числа степеней свободы $M - 1 = 3$ находим критическую точку $\chi^2_{\text{кр}} = 7,8$. Так как $V < \chi^2_{\text{кр}}$, то $B = V/C < \chi^2_{\text{кр}}$, и принимаем нулевую гипотезу о равенстве средних.

2) При условии равенства дисперсий в качестве оценки генеральной дисперсии принимаем среднее арифметическое исправленных дисперсий, взвешенное по числам степеней свободы: $s^2 = 3,95$.

Задача 16. По четырем независимым выборкам одинакового объема $n = 17$ найдены исправленные выборочные дисперсии: 0,21; 0,25; 0,34; 0,40. Требуется: 1) на уровне значимости 0,05 проверить гипотезу о равенстве дисперсий; 2) оценить генеральную дисперсию.

Решение.

1) Находим значение статистики критерия Кокрена:

$$G = \frac{0,4}{0,21 + 0,25 + 0,34 + 0,4} \approx 0,33.$$

По таблице критических точек распределения Кокрена по уровню значимости 0,05, числу степеней свободы 16 и числу выборок 4 находим $G_{кр} = G_{кр}(0,05; 16; 4) = 0,437$. Поскольку $G < G_{кр}$, нулевая гипотеза принимается.

2) При условии равенства дисперсий в качестве оценки генеральной дисперсии принимают среднее арифметическое исправленных дисперсий:

$$s^2 = \frac{0,21 + 0,25 + 0,34 + 0,4}{4} = 0,3.$$

Задачи для самостоятельного решения¹

Теоретические задачи

1. Пусть случайная величина $\xi \in N(a, \sigma^2)$, причем значение параметра a известно, а дисперсия σ^2 неизвестна. Требуется на уровне значимости α проверить нулевую гипотезу $H_0: \sigma = \sigma_0$, если альтернативная гипотеза $H_1: \sigma = \sigma_1 > \sigma_0$. Построить критерий отношения правдоподобия.
2. Пусть имеется последовательность независимых испытаний Бернулли с неизвестной вероятностью «успеха» p . Требуется на уровне значимости α проверить нулевую гипотезу $H_0: p = p_0$, если альтернативная гипотеза $H_1: p = p_1 > p_0$. Построить критерий отношения правдоподобия.

¹ В задачах выборочная дисперсия и выборочное среднее квадратическое отклонение предполагаются исправленными по умолчанию.

рий отношения правдоподобия, используя нормальное приближение для числа успехов. Вычислить объем выборки n , необходимый для достижения заданных ошибок первого и второго рода α и β .

3. Пусть случайная величина имеет распределение Пуассона с параметром λ . Требуется на уровне значимости α проверить нулевую гипотезу $H_0: \lambda = \lambda_0$, если альтернативная гипотеза $H_1: \lambda = \lambda_1 > \lambda_0$. Построить критерий отношения правдоподобия, используя нормальное приближение. Вычислить объем выборки n , необходимый для достижения заданных ошибок первого и второго рода α и β .
4. Пусть случайная величина имеет показательное распределение с параметром λ . Требуется на уровне значимости α проверить нулевую гипотезу $H_0: \lambda = \lambda_0$, если альтернативная гипотеза $H_1: \lambda = \lambda_1 > \lambda_0$. Построить критерий отношения правдоподобия.
5. Проверяется нулевая гипотеза о том, что случайная величина равномерно распределена на отрезке $[-1, 1]$ против гипотезы, что она имеет нормальное распределение $N(0, \sigma^2)$. Построить критерий отношения правдоподобия в общем виде и в частном случае при $n = 2$, $\alpha < \pi/4$.

Вычислительные задачи

6. По выборке объема $n = 9$, извлеченной из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 4$, на уровне значимости $\alpha = 0,05$ проверяется нулевая гипотеза $H_0: a = a_0 = 15$. Требуется: а) найти мощность критерия для проверки гипотезы $H_1: a = a_1 = 17$; б) найти объем выборки n_1 , при котором мощность критерия равна 0,8.
7. По выборке объема $n = 25$, извлеченной из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 10$, найдено выборочное среднее, равное 18. При уровне значимости 0,05 требуется: а) найти критическую область для проверки нулевой гипотезы $H_0: a = a_0 = 16$ при конкурирующей гипотезе $H_1: a = a_1 > 16$ и провести проверку; б) найти мощность критерия при $a_1 = 20$.
8. Крупная торговая фирма желает открыть в новом районе города филиал. Известно, что фирма будет работать прибыльно, если средний месячный доход жителей района превышает 300 долл. Также известно, что среднее квадратическое отклонение дохода σ составляет 50 долл. Проводится выборочное обследование

населения по величине доходов, чтобы принять решение об открытии филиала.

- а) Определите правило принятия решения, с помощью которого, основываясь на выборке $n = 100$ (человек) и уровне значимости $\alpha = 0,05$, можно установить, что филиал будет работать прибыльно.
 - б) Рассчитайте вероятность того, что при применении правила принятия решения, полученного при ответе на вопрос п. а), будет совершена ошибка второго рода, если в действительности средний доход достигает 310 долл.
 - в) Считая альтернативное значение генерального среднего месячного дохода равным 320 долл., рассчитайте объем выборки, при котором ошибка первого рода не превысит 2,5%, а ошибка второго рода не превысит 5%.
9. Путем выборочного опроса проверяется гипотеза о том, что стиральным порошком фирмы *A* пользуется 30% населения, против гипотезы, что им пользуется только 20%. Оценить объем выборки, необходимый для проверки гипотезы с ошибкой первого рода не более 5% и второго рода — не более 2,5%.
 10. Выборочные испытания 10 изделий показали, что их средний срок службы составляет 950 час. В предположении, что срок службы изделия распределен показательно, проверить гипотезу о том, что генеральное среднее составляет 1000 ч (против гипотезы, что оно меньше) на уровне значимости 5%.
 11. Рафинированный сахар упаковывается в пакеты с номинальным весом 1,0 кг со средним квадратическим отклонением, равным 0,01 кг. Случайная выборка $n = 16$ пакетов готовой продукции выявила средний вес 1,01 кг. При уровне значимости 5% проверить нулевую гипотезу о том, что средний вес пакета соответствует номиналу.
 12. Штамповочный пресс делает отверстия в металлических шайбах с нормативным размером 4,00 мм и средним квадратическим отклонением $\sigma = 0,20$ мм. Случайная выборка $n = 25$ шайб показала, что средний размер 3,88 мм. При уровне значимости 1% проверить нулевую гипотезу о том, что средняя величина отверстия соответствует нормативу.
 13. Автомат производит детали с известным средним квадратическим отклонением размера A мм. Выборка из 100 деталей показала выборочный средний размер B мм. Можно ли утверждать на уровне значимости 10%, что средний размер детали соответствует номиналу в C мм (против гипотезы, что не соответ-

- ствует)? Решить задачу, если: а) $A = 0,05$; $B = 7,02$; $C = 7$; б) $A = 0,08$; $B = 5,01$; $C = 5$; в) $A = 0,09$; $B = 5,02$; $C = 5$.
14. В прошлом году средняя заработная плата жителей города Булкино составляла N руб. В этом году выборка из 400 человек показала, что средняя заработная плата составляет A руб. при выборочном среднем квадратическом отклонении B руб. Можно ли на уровне значимости 5% утверждать, что заработная плата жителей города увеличилась? Решить задачу, если: а) $N = 7300$, $A = 7320$, $B = 150$; б) $N = 7770$, $A = 7779$, $B = 120$; в) $N = 7500$; $A = 7530$, $B = 120$; г) $N = 9500$, $A = 9514$, $B = 180$.
15. В прошлом году средняя заработная плата жителей города Сушкино составляла N руб. В этом году выборка из 250 человек показала, что средняя заработная плата составляет A руб. при выборочном среднем квадратическом отклонении B руб. Можно ли на уровне значимости 5% утверждать, что заработная плата жителей города увеличилась? Решить задачу, если: а) $N = 8500$, $A = 8520$, $B = 150$; б) $N = 7500$, $A = 7550$, $B = 500$; в) $N = 9500$, $A = 9515$, $B = 350$.
16. По выборке из 25 упаковок товара средний вес составил 101 г с выборочным средним квадратическим отклонением 3 г. На уровне значимости 5% проверить гипотезу о том, что средний вес упаковки товара составляет 100 г.
17. При измерении веса 20 шоколадных батончиков (с номинальным весом 50 г) получены следующие значения (г): 40,1; 50,0; 49,7; 50,5; 48,1; 50,3; 49,7; 51,6; 49,8; 50,1; 49,7; 48,8; 51,4; 49,1; 49,6; 50,9; 48,5; 52,0; 50,7; 50,6. Согласуются ли полученные данные с номиналом (на уровне значимости 10%)?
18. Средний доход фирмы в день составлял A единиц. После реорганизации выборочный средний доход в день за N рабочих дней составил B единиц с выборочным средним квадратическим отклонением C единиц. Можно ли утверждать (на уровне значимости 5%), что реорганизация привела к увеличению среднего дохода? Решить задачу, если: а) $A = 1020$, $N = 30$, $B = 1070$, $C = 90$; б) $A = 500$, $N = 25$, $B = 530$, $C = 30$; в) $A = 1500$, $N = 25$, $B = 1506$, $C = 20$; г) $A = 300$, $N = 16$, $B = 320$, $C = 15$; д) $A = 250$, $N = 16$, $B = 254$, $C = 12$.
19. Средний дневной объем продаж в магазине составлял 500 единиц. После реорганизации, за период в 25 рабочих дней, выборочный средний дневной объем продаж составил 520 единиц с выборочным средним квадратическим отклонением 40 еди-

- ниш. Можно ли утверждать (на уровне значимости 10%), что реорганизация привела к увеличению среднего объема продаж?
20. Среднее время сборки изделия было 90 мин. Инженер изобрел новый метод сборки этого изделия. В результате времена сборки 10 изделий новым способом составили: 79, 74, 112, 95, 83, 96, 77, 84, 70, 90 мин. Можно ли утверждать, что среднее время сборки сократилось (на уровне значимости 5%)?
21. Среднее время сборки изделия было N мин. При испытаниях нового метода сборки по выборке из M изделий получено выборочное среднее время сборки A мин. с выборочным средним квадратическим отклонением B мин. Можно ли утверждать на уровне значимости 5%, что среднее время сборки сократилось? Решить задачу, если: а) $N = 60$, $M = 25$, $A = 51$, $B = 6$; б) $N = 50$, $M = 25$, $A = 48$, $B = 7$; в) $N = 60$, $M = 16$, $A = 57$, $B = 9$; г) $N = 80$, $M = 16$, $A = 76$, $B = 5$.
22. Фирма утверждает, что средний срок службы ее изделия составляет N часов. По выборке из 25 изделий получен выборочный средний срок службы A часов с выборочным средним квадратическим отклонением B часов. Подтверждается ли заявление фирмы на уровне значимости 5% (против гипотезы, что средний срок службы на самом деле меньше)? Решить задачу, если: а) $N = 200$, $A = 192$, $B = 11$; б) $N = 300$, $A = 290$, $B = 15$; в) $N = 250$, $A = 235$, $B = 16$; г) $N = 350$, $A = 347$, $B = 13$.
23. В селе Сидорово проведено выборочное обследование доходов жителей. По выборке из 16 человек получено среднее 2620 руб. и среднее квадратическое отклонение 150 руб. Можно ли утверждать на уровне значимости 10%, что средний доход жителей составляет более 2500 руб.?
24. В селе Петрово проведено выборочное обследование доходов жителей. По выборке из 25 человек получено среднее 2380 руб. и среднее квадратическое отклонение 90 руб. Можно ли утверждать на уровне значимости 5%, что средний доход жителей составляет менее 2500 руб.?
25. В книжном магазине проведено исследование продаж детективов писателя Горшкова «Хромой против Косого». В течение 16 рабочих дней роман продавался ежедневно в среднем по 57 экземпляров с выборочным средним квадратическим отклонением 12 экземпляров. Можно ли утверждать на уровне значимости 10%, что этот роман расходуется лучше, чем предыдущий роман автора, «Хромой против Слепого», если тот расходился в среднем по 50 экземпляров в день?

26. В книжном магазине проведено исследование продаж фантастического романа писателя Бурьяненко «Танцы в пустоте». В течение 25 рабочих дней роман продавался ежедневно в среднем по 64 экземпляра с выборочным средним квадратическим отклонением 10 экземпляров. Можно ли утверждать на уровне значимости 5%, что этот роман расходуется хуже, чем предыдущий роман автора, «Звездная жуть», если тот расходился в среднем по 70 экземпляров в день?
27. В книжном магазине проведено исследование продаж нового романа писательницы Пунцовой. В течение 30 рабочих дней роман продавался ежедневно в среднем по 48 экземпляров с выборочным средним квадратическим отклонением 15 экземпляров. Можно ли утверждать на уровне значимости 5%, что этот роман расходуется лучше, чем предыдущий роман автора, если тот расходился в среднем по 40 экземпляров в день?
28. Из нормальной генеральной совокупности извлечена выборка объема $n = 17$ и по ней найдена выборочная дисперсия $\sigma^2 = 0,24$. Требуется на уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^2 = 0,18$, приняв в качестве конкурирующей гипотезы $H_1: \sigma^2 > 0,18$.
29. В результате длительного хронометража времени сборки узла различными сборщиками установлено, что дисперсия этого времени $\sigma_0^2 = 2$ мин². Результаты 20 наблюдений за работой новичка приведены в таблице (x_i — время сборки одного узла в минутах, середины интервалов; n_i — частота).

x_i	56	58	60	62	64
n_i	1	4	10	3	2

Можно ли на уровне значимости 0,05 считать, что дисперсия затрачиваемого им времени существенно не отличается от дисперсии времени остальных сборщиков?

30. Инвестор считает вложения в активы с дисперсией доходности более 0,04 слишком рискованными. За последние 10 лет выборочная дисперсия доходности актива A составила 0,06. Следует ли делать вложения в актив A , принимая решение на уровне значимости 5%?
31. Партия изделий принимается, если дисперсия размеров не превышает 0,2. Выборочная дисперсия для 30 изделий оказалась равна 0,3. Можно ли принять партию на уровне значимости 5%?

32. Старая наполняющая машина работала со средним квадратическим отклонением веса упаковки A (г). Были проведены испытания новой машины. По выборке из N упаковок товара получено выборочное среднее квадратическое отклонение B (г). Можно ли утверждать на уровне значимости 5%, что для новой машины среднее квадратическое отклонение меньше? Решить задачу, если: а) $A = 1,5$, $N = 30$, $B = 0,8$; б) $A = 0,7$, $N = 25$, $B = 0,6$; в) $A = 1,1$, $N = 25$, $B = 0,4$; г) $A = 1$, $N = 30$, $B = 0,8$.
33. По 100 независимым испытаниям найдена относительная частота $w = 0,15$. На уровне значимости 0,05 проверить нулевую гипотезу $H_0: p = 0,17$ при конкурирующей гипотезе $H_1: p \neq 0,17$.
34. Независимому статистику поручено проверить информацию маркетинговой службы некоторого туристического бюро о том, что 70% клиентов выбирают в качестве формы обслуживания полупансион. Статистик провел опрос 150 случайно выбранных туристов, из них полупансион предпочли 84 человека. К какому выводу пришел статистик при проверке гипотезы $H_0: p = 0,7$ при альтернативе $H_1: p \neq 0,7$ на уровне значимости критерия $\alpha = 0,05$?
35. Статистику необходимо проверить экспертную оценку о том, что 75% отечественных предприятий уклоняется (частично) от уплаты налогов. По результатам неофициального опроса руководителей предприятий 140 из 200 случайно отобранных директоров подтвердили, что используют различные схемы для ухода от уплаты налогов. Можно ли при уровне значимости 0,05 согласиться с приведенной экспертной оценкой?
36. Партия изделий принимается, если вероятность того, что изделие окажется бракованным, не превышает 0,03. Среди случайно отобранных 400 изделий оказалось 18 бракованных. Можно ли принять партию на уровне значимости 0,05?
37. Фирма разослала N новых рекламных каталогов и получила M заказов. Можно ли утверждать (на уровне значимости 5%), что эффективность рекламы повысилась, если ранее она составляла в среднем $A\%$? Решить задачу, если: а) $N = 1000$, $M = 120$, $A = 10$; б) $N = 500$, $M = 75$, $A = 13$; в) $N = 500$, $M = 70$, $A = 12$.
38. В таблице приведены данные по годовым поступлениям в региональные бюджеты доходов от налогообложения субъектов малого предпринимательства (юридических лиц) до и после введения в рассматриваемых регионах единого налога на вмененный доход (млн руб.):

Год «до»	32	25	42	35	50	40	38	19	15
Год «после»	35	35	25	19	40	40	75	45	18

Требуется установить на уровне значимости 0,05, существенно ли изменились объемы поступлений в региональные бюджеты в сторону увеличения (используя нормальное приближение).

39. В таблице представлены месячные объемы продаж в магазинах фирмы до и после проведения ее рекламной компании (у.е.):

Магазины	1	2	3	4	5	6	7
До рекламы	10	15	12	13	14	11	15
После рекламы	17	20	13	11	16	10	18

Можно ли утверждать на уровне значимости 0,10, что рекламная кампания привела к существенному увеличению объемов продаж?

40. В таблице представлены данные о производительности труда группы сотрудников фирмы до и после обучения на курсах повышения квалификации:

Сотрудник	1	2	3	4	5	6	7
До обучения	20	18	15	16	20	14	17
После обучения	26	27	20	23	19	22	28

Можно ли утверждать на уровне значимости 5%, что обучение привело к существенному увеличению производительности труда?

41. За последние N лет выборочная дисперсия доходности актива A составила U , актива B – V . Есть ли основания утверждать (на уровне значимости 5%), что вложения в B более рискованны, чем в A ? Решить задачу, если: а) $N = 5$, $U = 0,04$, $V = 0,05$; б) $N = 7$, $U = 0,05$, $V = 0,08$; в) $N = 10$, $U = 0,03$, $V = 0,12$; г) $N = 10$, $U = 0,08$, $V = 0,14$.
42. По двум независимым выборкам с объемами n_x и n_y , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии σ_x^2 и σ_y^2 . При уровне значимости α проверить нулевую гипотезу $\sigma_x^2 = \sigma_y^2$ против конкурирующей гипотезы $H_1: \sigma_x^2 > \sigma_y^2$, если:
- а) $n_x = 10$; $n_y = 16$; $\sigma_x^2 = 3,6$; $\sigma_y^2 = 2,4$; $\alpha = 0,05$;
- б) $n_x = 13$; $n_y = 18$; $\sigma_x^2 = 0,72$; $\sigma_y^2 = 0,20$; $\alpha = 0,01$.

43. Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объемы которых $n = 10$ и $m = 8$. В результате измерения контролируемого размера отобранных изделий получены следующие результаты.

x_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
y_i	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,38		

Можно ли считать, что станки обладают одинаковой точностью, если принять уровень значимости $\alpha = 0,1$ и в качестве конкурирующей гипотезы взять $H_1: \sigma_x^2 \neq \sigma_y^2$?

44. По выборке объема $n = 50$ найден средний размер диаметра валиков, изготовленных автоматом № 1, равный 20,1 мм; по выборке объема $m = 50$ найден средний размер диаметра валиков, изготовленных автоматом № 2, равный 19,8 мм. Генеральные дисперсии известны: $\sigma_x^2 = 1,750 \text{ мм}^2$, $\sigma_y^2 = 1,375 \text{ мм}^2$. Требуется на уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma_x^2 = \sigma_y^2$ при конкурирующей гипотезе $H_1: \sigma_x^2 \neq \sigma_y^2$. Предполагается, что случайные величины распределены нормально и выборки независимы.
45. По двум независимым выборкам с объемами $n = 10$ и $m = 10$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние, равные 14,3 и 12,2 соответственно. Генеральные дисперсии известны: $\sigma_x^2 = 22$, $\sigma_y^2 = 18$. На уровне значимости 0,05 проверить нулевую гипотезу $H_0: \sigma_x^2 = \sigma_y^2$ при конкурирующей гипотезе $H_1: \sigma_x^2 > \sigma_y^2$.
46. По двум независимым выборкам с объемами $n = 50$ и $m = 50$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние, равные 142 и 150 соответственно. Генеральные дисперсии известны: $\sigma_x^2 = 28,2$, $\sigma_y^2 = 22,8$. На уровне значимости 0,01 проверить нулевую гипотезу $H_0: \sigma_x^2 = \sigma_y^2$ при конкурирующей гипотезе $H_1: \sigma_x^2 < \sigma_y^2$.
47. Компания по производству сахарного песка имеет две производственные линии для наполнения пакетов сахарным песком по 1 кг. Среднее квадратическое отклонение веса пакетов, поставляемых с первой линии, $\sigma_1 = 0,02 \text{ кг}$, а со второй линии — $\sigma_2 = 0,04 \text{ кг}$. С первой линии была взята случайная выборка объема $n_1 = 10$ пакетов и найден средний вес $\bar{x}_1 = 1,018 \text{ кг}$. Подобная выборка $n_2 = 12$ пакетов была взята со второй линии и найден средний вес $\bar{x}_2 = 0,989 \text{ кг}$. Есть ли основание считать, что средний вес пакетов первой и второй линий различаются? Проверить гипотезу при уровне значимости 0,01.

48. В течение 100 рабочих дней магазин *A* посещало в среднем 198 покупателей в день, магазин *B* — 202. Есть ли основания утверждать на уровне значимости 5%, что магазин *B* более популярен, чем *A*, если числа покупателей в день имеют дисперсии, равные 256?
49. В течение *K* дней в фирму *A* обращалось в среднем *N* клиентов в день, в фирму *B* — *M*. Есть ли основания утверждать (на уровне значимости 5%), что фирма *B* более популярна, чем фирма *A*, если числа клиентов в день имеют известные дисперсии *U* и *V* соответственно? Решить задачу, если: а) $K = 64$, $N = 87$, $M = 93$, $U = 124$, $V = 132$; б) $K = 64$, $N = 90$, $M = 93$, $U = 125$, $V = 199$; в) $K = 25$, $N = 49$, $M = 58$, $U = 119$, $V = 137$; г) $K = 100$, $N = 109$, $M = 112$, $U = 150$, $V = 250$.
50. В городах Усатово и Полосатово проведены выборочные обследования доходов жителей. По выборкам из 100 человек получены следующие результаты: в Усатово — средний доход 4050 руб., среднее квадратическое отклонение 105 руб.; в Полосатово — средний доход 3970 руб., среднее квадратическое отклонение 85 руб. Можно ли утверждать на уровне значимости 5%, что в Усатово живут в среднем так же, как и в Полосатово?
51. В городах Сквородкино и Кастрюлино проведены выборочные обследования доходов жителей. По выборкам из 200 человек получены следующие результаты: в Сквородкино — средний доход 5230 руб., среднее квадратическое отклонение 250 руб.; в Кастрюлино — средний доход 5180 руб., среднее квадратическое отклонение 190 руб. Можно ли утверждать на уровне значимости 5%, что в Сквородкино живут в среднем богаче, чем в Кастрюлино?
52. В городах Бубликово и Баранкино проведены выборочные обследования доходов жителей. По выборкам из 100 человек получено, что в Бубликово средний доход *A* руб. с выборочным средним квадратическим отклонением *B* руб., в Баранкино средний доход *C* руб. с выборочным средним квадратическим отклонением *D* руб. Можно ли утверждать на уровне значимости 5%, что в Бубликово живут в среднем богаче, чем в Баранкино? Решить задачу, если: а) $A = 8030$, $B = 240$, $C = 7960$, $D = 180$; б) $A = 9010$, $B = 200$, $C = 8985$, $D = 150$; в) $A = 8120$, $B = 140$, $C = 7990$, $D = 180$.

53. В таблице представлены сгруппированные данные о расходе сырья на одно изделие для двух различных технологий изготовления.

Показатель	Старая технология			Новая технология			
Расход сырья	304	307	308	303	304	306	308
Число изделий	1	4	4	2	6	4	1

В предположении, что расход сырья как при старой, так и при новой технологии, имеет нормальное распределение, выяснить, влияет ли изменение технологии на средний расход сырья на одно изделие. Принять $\alpha = 0,1$.

54. По двум независимым выборкам с объемами $n = 10$ и $m = 8$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние, равные 142,3 и 145,3, и исправленные выборочные дисперсии, равные 2,7 и 3,2 соответственно. На уровне значимости 0,1 проверить нулевую гипотезу $H_0: \alpha_x = \alpha_y$ при конкурирующей гипотезе $H_1: \alpha_x \neq \alpha_y$.
55. Компания «Рекс», производящая батареек, утверждает, что в среднем период службы их батареек дольше, чем батареек их конкурента — компании «Полкан». Ассоциация потребителей взяла случайную выборку $n = 10$ батареек производства компании «Рекс», и, испытав их, определила средний срок службы $\bar{x} = 198$ час. с исправленным средним квадратическим отклонением $s_x = 8,7$ час. Такая же проверка продукции «Полкан» для выборки объема $m = 15$ дала величины $\bar{y} = 194$ и $s_y = 5,8$ час. Подтверждают ли эти результаты заявление компании «Рекс»? Уровень значимости — 0,1.
56. Проведено исследование батареек фирм «Винтик» и «Шпунтик». По выборке из N батареек фирмы «Винтик» получен средний срок службы A часов с выборочным средним квадратическим отклонением B часов, по выборке из M батареек фирмы «Шпунтик» — средний срок службы C часов с выборочным средним квадратическим отклонением D часов. Можно ли утверждать на уровне значимости 10%, что в среднем срок службы батареек этих двух фирм одинаков? Решить задачу, если: а) $N = 11$, $M = 16$, $A = 207$, $B = 9$, $C = 195$, $D = 8$; б) $N = 10$, $M = 17$, $A = 256$, $B = 13$, $C = 252$, $D = 11$; в) $N = 12$, $M = 15$, $A = 205$, $B = 8$, $C = 195$, $D = 7$.
57. В селах Сидорово и Петрово проведены выборочные обследования доходов жителей. По выборкам из 10 человек получены следующие результаты: в Сидорово — средний доход 2620 руб., среднее квадратическое отклонение 150 руб.; в Петрово —

средний доход 2380 руб., среднее квадратическое отклонение 90 руб. Можно ли утверждать на уровне значимости 10%, что в Петрово живут в среднем беднее, чем в Сидорово?

58. В книжном магазине проведено исследование продаж детективов писателя Горшкова «Хромой против Косого» и «Рябой против Глухого». В течение 10 рабочих дней первый роман продавался ежедневно в среднем по 57 экземпляров со средним квадратическим отклонением 12 экземпляров; второй роман — в среднем по 62 экземпляра со средним квадратическим отклонением 15 экземпляров. Можно ли утверждать на уровне значимости 10%, что второй роман расходуется лучше первого?
59. В книжном магазине проведено исследование продаж фантастических романов писателя Бурьяненко «Ночной досмотр» и «Дневной просмотр». В течение 12 рабочих дней первый роман продавался ежедневно в среднем по 70 экземпляров со средним квадратическим отклонением 14 экземпляров; второй роман — в среднем по 63 экземпляра со средним квадратическим отклонением 11 экземпляров. Можно ли утверждать на уровне значимости 10%, что второй роман расходуется хуже первого?
60. Урожайность зерновых культур в России в 1992–2001 гг. представлена таблицей.

Год	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Урожайность, ц/га	18,0	17,1	15,3	13,1	14,9	17,8	12,9	14,4	15,6	19,4

Можно ли утверждать, что урожайность в 1992–1996 гг. и 1997–2001 гг. была в среднем одинаковая (на уровне значимости 0,1)?

61. Производство пшеницы в России в 1995–2002 гг. представлено таблицей:

Год	1995	1996	1997	1998	1999	2000	2001	2002
Производство, млн т	30,1	34,9	44,3	27,0	31,0	34,5	47,0	57,7

Можно ли утверждать, что производство пшеницы в 1992–1998 гг. и 1999–2002 гг. было в среднем одинаковое (на уровне значимости 0,1)?

62. Для оценки качества изделий, изготовленных двумя заводами, взяты выборки по $n_1 = 200$ и $n_2 = 300$ изделий. В этих выборках оказалось соответственно $m_1 = 20$ и $m_2 = 15$ бракованных изделий. При уровне значимости 0,05 проверить нулевую гипотезу

$H_0: p_1 = p_2$ о равенстве вероятностей изготовления бракованного изделия обоими заводами при конкурирующей гипотезе $H_1: p_1 > p_2$.

63. Для оценки качества изделий, изготовленных двумя заводами, взяты выборки из 100 изделий первого и 200 изделий второго. В них оказалось по N и M бракованных изделия соответственно. Можно ли утверждать на уровне значимости 10%, что вероятность изготовления брака на обоих заводах одинакова? Решить задачу, если: а) $N = 18$, $M = 24$; б) $N = 15$, $M = 14$; в) $N = 10$, $M = 26$.
64. Фирма разослала 1000 рекламных каталогов и получила N заказов. Можно ли утверждать на уровне значимости 5%, что эффективность рекламы повысилась, если в прошлом году фирма получила M заказов (при том же объеме рассылки)? Решить задачу, если: а) $N = 160$, $M = 120$; б) $N = 120$, $M = 100$; в) $N = 150$, $M = 130$; г) $N = 140$, $M = 100$; д) $N = 180$, $M = 160$.
65. Аудиторы компании интересуются системой обработки счетов доходов. Они взяли случайную выборку из 100 законченных счетов, в которой 8 счетов оказались дефектными. Тогда аудиторы предложили некоторые модификации в процедуре и через определенное время провели случайную выборку из 150 завершенных счетов, где обнаружили 6 дефектных счетов. Имеется ли основание предполагать на уровне значимости 5%, что новые процедуры уменьшают ошибку?
66. Можно ли воспользоваться критерием Бартлетта для проверки гипотезы об однородности дисперсий по выборкам объема $n_1 = 15$, $n_2 = 25$, $n_3 = 10$, $n_4 = 3$? Дайте пояснения.
67. По трем независимым выборкам, объемы которых $n_1 = 9$, $n_2 = 13$, $n_3 = 15$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии, равные 3,2; 3,8; 6,3 соответственно. Требуется: а) на уровне значимости 0,05 проверить гипотезу о равенстве дисперсий; б) оценить генеральную дисперсию.
68. Четыре фасовочных автомата настроены на отвешивание одинакового веса. На каждом автомате отвесили 10 проб, затем эти же пробы взвесили на точных весах и нашли исправленные выборочные дисперсии отклонений от норматива: 0,01; 0,02; 0,04; 0,03. Можно ли утверждать на уровне значимости 5%, что автоматы обеспечивают одинаковую точность взвешивания?

69. Инвестор имеет данные по доходности актива A за 5 лет, актива B — за 7 лет, актива C — за 10 лет. Исправленные выборочные дисперсии доходностей равны 0,01; 0,02 и 0,03 соответственно. Можно ли утверждать на уровне значимости 5%, что риск вложений в эти активы одинаков?
70. Инвестор имеет данные за 10 лет по доходности активов A , B , C и D . Исправленные выборочные дисперсии доходностей равны 0,012; 0,021; 0,025; 0,032 соответственно. Можно ли утверждать на уровне значимости 5%, что риск вложений в эти активы одинаков?

Глава 16

КРИТЕРИИ СОГЛАСИЯ

Критерии проверки гипотезы о предполагаемом виде закона распределения случайной величины называют **критериями согласия**.

Следует понимать, что проверяют не тот факт, что случайная величина *действительно* имеет определенный закон распределения (например, нормальный), а проверяется лишь, достаточно ли хорошо наблюдаемые данные *согласуются* с некоторым законом распределения, чтобы можно было использовать этот закон для прогнозирования поведения рассматриваемой случайной величины.

Гипотезы могут быть как простыми, так и сложными. Гипотеза называется *простой*, если проверяется соответствие некоторому закону распределения с заданными параметрами. Гипотеза называется *сложной*, если проверяется соответствие некоторому закону распределения с произвольными параметрами. В этом случае параметры оценивают по выборке.

Наиболее часто используемые критерии согласия — это критерии Пирсона, Фишера и Колмогорова.

16.1. Критерии согласия Пирсона и Фишера (хи-квадрат)

Критерий согласия Пирсона (хи-квадрат) идеально подходит для проверки гипотез в полиномиальной схеме.

Пусть проводится n независимых испытаний, каждое из которых может иметь r различных исходов $A_1, A_2, \dots, A_r$. Требуется проверить гипотезу о том, что вероятности этих исходов равны $p_1, p_2, \dots, p_r$, если в последовательности испытаний они встретились $m_1, m_2, \dots, m_r$ раз.

► **Теорема 1 (Пирсона).** Если основная гипотеза верна, то распределение статистики хи-квадрат

$$\chi^2 = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i}$$

при $n \rightarrow \infty$ стремится к распределению хи-квадрат с $r - 1$ степенями свободы.

В противном случае эта статистика стремится к бесконечности.

Отсюда получаем критерий (применимый при больших n): если $\chi^2 < \chi_{\alpha, r-1}^2$, то основная гипотеза принимается, иначе отвергается.

Дадим краткое обоснование критерия Пирсона.

Будем считать наступление события A_i успехом, а его отсутствие — неудачей.

Тогда m_i — это число успехов в n испытаниях Бернулли, m_i/n — относительная частота события A_i , которая является асимптотически нормальной, несмещенной и состоятельной оценкой вероятности этого события. Поскольку $M\left(\frac{m_i}{n}\right) = p_i$, при больших n можно считать справедливым приближенное равенство $\frac{m_i}{n} \approx p_i$, т.е. вектор частот $\frac{m}{n} = \left(\frac{m_1}{n}, \frac{m_2}{n}, \dots, \frac{m_r}{n}\right)$ является оценкой гипотетических вероятностей $p = (p_1, p_2, \dots, p_r)$.

В качестве меры расхождения гипотетической и теоретической вероятностей рассматривается сумма квадратов отклонений

$$\sum_{i=1}^r \gamma_i \left(\frac{m_i}{n} - p_i \right)^2,$$

где γ_i — веса отклонений. Например, в методе наименьших квадратов принято $\gamma_i = 1$, но согласно теории ошибок Гаусса каждое слагаемое должно входить в сумму со своей точностью. Пирсон показал, что если положить $\gamma_i = n/p_i$, то полученная статистика критерия будет иметь известный закон распределения. Действительно,

$$\sum_{i=1}^r \gamma_i \left(\frac{m_i}{n} - p_i \right)^2 = \sum_{i=1}^r \frac{n}{p_i} \left(\frac{m_i - np_i}{n} \right)^2 = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i},$$

величина $y_i = \frac{m_i - np_i}{\sqrt{np_i}}$ асимптотически нормальна, т.е.

$y_i \xrightarrow{c} N(0, 1 - p)$. Если бы y_i имели распределение $N(0, 1)$ и были независимы, то случайная величина $\chi^2 = \sum_{i=1}^r y_i^2$ имела бы

предельное распределение χ_r^2 . Однако между случайными величинами y_i для различных i существует линейная зависимость. Действительно, поскольку

$$\sum_{i=1}^r \sqrt{p_i} y_i = \sum_{i=1}^r \frac{\sqrt{p_i} (m_i - np_i)}{\sqrt{np_i}} = \frac{1}{\sqrt{n}} \sum_{i=1}^r (m_i - np_i) = \frac{1}{\sqrt{n}} (n - n) = 0,$$

то $y_i = \frac{m_i - np_i}{\sqrt{np_i}}$ линейно зависимы, причем их связь описывается всего одним уравнением. С учетом этих фактов, при $n \rightarrow \infty$ случайная величина $\chi^2 = \sum_{i=1}^r y_i^2$ имеет распределение χ^2 с $r - 1$ степенями свободы.

Если вероятности $p_1, p_2, \dots, p_r$ зависят от неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$, то при вычислении статистики хи-квадрат используют оценки этих вероятностей, найденные по данным эксперимента $m_1, m_2, \dots, m_r$, как правило, методом максимального правдоподобия. Но в этом случае предельное распределение статистики хи-квадрат имеет уже $r - k - 1$ степеней свободы. Затем принимают решение: если $\chi^2 < \chi_{\alpha; r-k-1}^2$, то гипотеза принимается, иначе отвергается.

Критерий хи-квадрат для простой гипотезы (т.е. в случае известных параметров) называют также критерием хи-квадрат Пирсона, а критерий хи-квадрат для сложной гипотезы (с оцениванием параметров) — критерием хи-квадрат Фишера.

Критерий хи-квадрат можно применять и в более общей схеме, для проверки распределений случайных величин. В этом случае в качестве исходов $A_1, A_2, \dots, A_r$ берут попадания наблюдений в некоторые множества $\Delta_1, \Delta_2, \dots, \Delta_r$. Для дискретных величин это могут быть отдельные значения или их объединения. Для непрерывных величин используют обычную группировку, т.е. подсчитывают числа попаданий в некоторые интервалы.

Если распределение не ограничено слева или справа, то крайние интервалы продолжают до бесконечности. Если числа попаданий в какие-то интервалы слишком малы (например, мень-

ше 5), то такие интервалы объединяют с соседними интервалами. Всего желательно иметь не менее 50 наблюдений в выборке.

В результате есть множества $\Delta_1, \Delta_2, \dots, \Delta_r$, находят числа $m_1, m_2, \dots, m_r$ попаданий наблюдений в эти множества и теоретические вероятности $p_i = P(\xi \in \Delta_i)$, $i = 1, 2, \dots, r$, после чего применяют критерий хи-квадрат.

Рассмотрим более подробно следующие случаи.

I. Пусть $x_1, x_2, \dots, x_n$ — независимые наблюдения некоторой случайной величины ξ с неизвестной функцией распределения $F(x)$. Требуется по выборке $x_1, x_2, \dots, x_n$ проверить нулевую гипотезу о том, что генеральная совокупность имеет функцию распределения $F_0(x)$, если известны значения параметров закона распределения, т.е. имеет место простая гипотеза.

Для проверки этой гипотезы область наблюдаемых значений $x_1, x_2, \dots, x_n$ случайной величины ξ произвольным образом разбивают на r непересекающихся областей Δ_i , $i = 1, 2, \dots, r$. Обычно это последовательность непересекающихся интервалов и полуинтервалов вида $(-\infty, C_1); [C_1, C_2); \dots; [C_{r-1}, +\infty)$; $C_0 = -\infty$; $C_r = +\infty$.

Если справедлива основная гипотеза, т.е. случайные величины x_k имеют своей функцией распределения функцию $F_0(x)$, то можно найти теоретические вероятности попадания случайной величины в частичные интервалы из условия

$$p_i = P(C_{i-1} \leq \xi < C_i) = F_0(C_i) - F_0(C_{i-1}), \text{ где } p_i > 0, \sum_{i=1}^r p_i = 1.$$

Со случайными величинами $x_1, x_2, \dots, x_n$ естественно связана полиномиальная схема с n испытаниями, в которой результатом k -го испытания является попадание значения x_k в какой-либо интервал. Обозначим через $m_i = m_i(x_1, x_2, \dots, x_n)$ число значений среди $x_1, x_2, \dots, x_n$, попавших в промежуток $\Delta_i = [C_{i-1}, C_i)$.

По теореме Пирсона получаем, что если $x_1, x_2, \dots, x_n$ — выборка из генеральной совокупности с функцией распределения $F_0(x)$, то статистика $\chi^2 = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i}$ имеет при $n \rightarrow \infty$

(т.е. при достаточно больших n) распределение хи-квадрат с $r - 1$ степенями свободы, если основная гипотеза верна. В противном случае статистика стремится к бесконечности. Поэтому в качестве критической области выбирают область больших значений.

Поскольку односторонний критерий более «жестко» отвергает нулевую гипотезу, чем двусторонний, построим правостороннюю критическую область исходя из требования, что вероятность попадания критерия в эту область в предположении истинности нулевой гипотезы должна быть равна принятому уровню значимости α : $P(\chi^2 \geq \chi^2_{\alpha, r-1}) = \alpha$.

Алгоритм проверки гипотезы следующий.

1. Из генеральной совокупности производят выборку объема n ($n > 50$).

2. Составляют сгруппированный статистический ряд.

3. Весь диапазон наблюдаемых значений разбивают на r частичных интервалов (в каждом из которых должно быть минимум 5–8 наблюдений, иначе интервалы объединяются; хорошие результаты получают при $np_i \geq 10$).

4. На основании гипотетической функции распределения $F_0(x)$ вычисляют вероятности попадания случайной величины ξ в частичные интервалы:

$$p_i = P(C_{i-1} < \xi < C_i) = F_0(C_i) - F_0(C_{i-1}), \quad i = 1, 2, \dots, r.$$

5. Умножая полученные вероятности p_i на объем выборки, получаем теоретические частоты np_i , т.е. частоты, которые следует ожидать, если нулевая гипотеза справедлива.

6. Вычисляют статистику хи-квадрат: $\chi^2 = \sum_{i=1}^r \frac{(m_i - np_i)^2}{np_i}$.

7. По таблице критических точек распределения хи-квадрат по заданному уровню значимости α и числу степеней свободы $r - 1$ находим критические точки $\chi^2_{\alpha, r-1}$.

8. Сравнивая наблюдаемые значения статистики χ^2 с критическим значением $\chi^2_{\alpha, r-1}$, принимаем одно из двух решений:

а) если $\chi^2 > \chi^2_{\alpha, r-1}$, то нулевая гипотеза отвергается в пользу альтернативной, т.е. считается, что гипотетическая функция распределения не согласуется с опытными данными;

б) если $\chi^2 < \chi^2_{\alpha, r-1}$, то нет основания для отклонения нулевой гипотезы, т.е. гипотетическая функция $F_0(x)$ согласуется с опытными данными.

II. Если значения параметров гипотетической функции распределения $F_0(x)$ неизвестны, то имеем сложную гипотезу.

Основная гипотеза H_0 заключается в том, что функция распределения имеет вид $F_0(x) = F(x, \theta_1, \dots, \theta_k)$ при некоторых

неизвестных значениях параметров $\theta_1, \theta_2, \dots, \theta_k$. В этом случае вероятности $p_1, p_2, \dots, p_r$ также зависят от параметров.

Статистика хи-квадрат принимает вид:

$$\chi^2 = \sum_{i=1}^r \frac{(m_i - np_i(\theta_1, \theta_2, \dots, \theta_k))^2}{np_i(\theta_1, \theta_2, \dots, \theta_k)}.$$

При известных значениях параметров имел бы место первый случай. Но так как истинные значения $\theta_1, \theta_2, \dots, \theta_k$ неизвестны, то, подставляя их оценки, найденные методом максимального правдоподобия, получаем статистику χ^2 с меньшим числом степеней свободы, а именно $\nu = r - k - 1$, где r — число интервалов, на которые разбит весь диапазон наблюдаемых значений, k — число параметров гипотетической функции распределения.

Сравнивая наблюдаемое значение статистики χ^2 с критическим значением $\chi_{\alpha, r-k-1}^2$, по приведенной схеме делаем заключение об истинности нулевой гипотезы: гипотеза принимается, если $\chi^2 < \chi_{\alpha, r-k-1}^2$, и отвергается в противном случае.

Задача 1. В следующей таблице представлены данные о числе сделок, заключенных на фондовой бирже за квартал, для 517 инвесторов.

i	0	1	2	3	4	5	6	7
m_i	112	168	130	68	32	5	1	1

В первой строке приведено число сделок, во второй — число инвесторов, заключивших указанное число сделок за квартал.

Проверить, используя критерий Пирсона, что на уровне значимости $\alpha = 0,05$ число сделок, заключенных одним инвестором за квартал, распределено по закону Пуассона с параметром $\lambda = 1,5$.

Решение. Поскольку распределение Пуассона дискретно, в качестве различных исходов здесь можно принять сами значения случайной величины. Заметим, что два последних значения (6 и 7) встретились слишком много раз, поэтому их следует объединить с предыдущим (5). Кроме того, распределение Пуассона не ограничено справа, и следует учесть все значения, превышающие число 7 (которые не встретились ни разу).

Таким образом, в качестве множеств Δ_i выберем значения $\{0\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5, +\infty\}$. Здесь $r = 6$.

Найдем теоретические вероятности по формуле распределения Пуассона:

$$P(\xi = j) = \frac{\lambda^j}{j!} e^{-\lambda}, \quad j = 0, 1, 2, \dots$$

При $\lambda = 1,5$ получаем:

$$p_1 = P(\xi = 0) \approx 0,2231;$$

$$p_2 = P(\xi = 1) \approx 0,3347;$$

$$p_3 = P(\xi = 2) \approx 0,2510;$$

$$p_4 = P(\xi = 3) \approx 0,1255;$$

$$p_5 = P(\xi = 4) \approx 0,0471;$$

$$p_6 = P(\xi \geq 5) \approx 0,0186.$$

Умножим эту величину на число инвесторов $n = 517$ и составим таблицу (табл. 16.1).

Таблица 16.1

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2 / (np_i)$
0	112	115,34	-3,34	0,10
1	168	173,04	-5,04	0,15
2	130	129,77	0,23	0,00
3	68	64,88	3,12	0,15
4	32	24,35	7,65	2,40
≥ 5	7	9,62	-2,62	0,71

Суммируя значения в последнем столбце, получаем значение статистики хи-квадрат ($\chi^2 = 3,51$).

По таблице критических точек распределения хи-квадрат при уровне значимости $\alpha = 0,05$ и числе степеней свободы $r - 1 = 5$ находим критическую точку $\chi^2_{кр} = 11,1$. Поскольку $\chi^2 < \chi^2_{кр}$, можно считать, что число сделок, заключенных одним инвестором за квартал, распределено по закону Пуассона с параметром $\lambda = 1,5$.

Замечание. Если бы значение параметра $\lambda = 1,5$ было оценено по самой выборке, следовало бы задать число степеней свободы $r - 2 = 4$. Тогда имеем $\chi^2_{кр} = 9,5$, следовательно, гипотеза тоже принимается.

Задача 2. В табл. 16.2 приведены сгруппированные данные о коэффициентах соотношения заемных и собственных средств на 100 малых предприятиях региона.

Таблица 16.2

Номер интервала	Интервал	Середина интервала	Число наблюдений в интервале
1	5,05—5,15	5,1	5
2	5,15—5,25	5,2	8
3	5,25—5,35	5,3	12
4	5,35—5,45	5,4	20
5	5,45—5,55	5,5	26
6	5,55—5,65	5,6	15
7	5,65—5,75	5,7	10
8	5,75—5,85	5,8	4

На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что коэффициенты можно описать нормальным распределением.

Решение. В рассматриваемом случае параметры распределения явно не заданы, их следует оценить по сгруппированным данным. Находим выборочное среднее (5,46) и выборочную дисперсию (0,03).

Теоретические вероятности оцениваем по формуле:

$$P(c_{i-1} < \xi < c_i) = \Phi\left(\frac{c_i - \bar{x}}{s}\right) - \Phi\left(\frac{c_{i-1} - \bar{x}}{s}\right), \quad i = 1, 2, \dots, 8.$$

Следует продолжить крайние интервалы и положить $c_0 = -\infty$, $c_8 = +\infty$, поскольку нормальное распределение не ограничено с обеих сторон.

С учетом полученных значений строим таблицу (табл. 16.3).

Таблица 16.3

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2 / (np_i)$
$\leq 5,15$	5	3,67	1,32	0,47
5,15—5,25	8	7,59	0,41	0,02

Окончание табл. 16.3

Δ_i	m_i	np_i	$m_i - np_i$	$(m_i - np_i)^2 / (np_i)$
5,25—5,35	12	15,00	-3,00	0,60
5,35—5,45	20	21,43	-1,43	0,09
5,45—5,55	26	22,14	3,86	0,67
5,55—5,65	15	16,53	-1,53	0,14
5,65—5,75	10	8,93	1,07	0,13
$\geq 5,75$	4	4,70	-0,70	0,10

Суммируя значения в последнем столбце, получаем значение статистики хи-квадрат ($\chi^2 = 2,22$).

С помощью таблицы критических точек распределения хи-квадрат по уровню значимости $\alpha = 0,05$ и числу степеней свободы $r - 1 - k = 8 - 1 - 2 = 5$ находим критическую точку $\chi^2_{кр} = 11,1$. Поскольку $\chi^2 < \chi^2_{кр}$, можно считать, что коэффициенты хорошо описываются нормальным распределением.

Замечание. В принципе здесь можно было бы объединить крайние интервалы с соседними. Вычисления показывают, что и в этом случае гипотеза принимается.

III. Экономисты также используют критерий χ^2 в качестве критерия однородности.

Пусть имеется $k \geq 2$ независимых выборок, содержащих соответственно $n_1, n_2, \dots, n_k$ независимых наблюдений: $(x_1, x_2, \dots, x_{n_1})$; $(y_1, y_2, \dots, y_{n_2})$; $\dots$; $(z_1, z_2, \dots, z_{n_k})$.

Гипотеза об однородности предполагает, что генеральные совокупности, из которых извлечены выборки, одинаковы (или все выборки произведены из одной генеральной совокупности) и им соответствуют одинаковые функции распределения.

Наиболее часто в приложениях встречается случай, когда $k = 2$. Пусть имеется два ряда наблюдений некоторого признака, и каждый ряд разбит на r групп по значениям этого признака. Сгруппированный ряд имеет вид:

$$\begin{array}{ccccccc}
 n_1: & m_1 & m_2 & \dots & m_i & \dots & m_r \\
 n_2: & l_1 & l_2 & \dots & l_i & \dots & l_r
 \end{array}$$

Пусть m_i и l_i — количество выборочных значений в i -й группе для первого и второго ряда наблюдений. Тогда статистический критерий для проверки истинности нулевой гипотезы принимает вид

$$\chi^2 = n_1 n_2 \sum_{i=1}^r \frac{\left(\frac{m_i}{n_1} - \frac{l_i}{n_2} \right)^2}{\frac{m_i + l_i}{n_1 + n_2}},$$

и в случае истинности основной гипотезы при $n \rightarrow \infty$ он имеет предельное распределение χ^2_{r-1} с $r - 1$ степенями свободы. Критическими точками, соответствующими уровню значимости α , будут $\chi^2_{\alpha, r-1}$, и проверка гипотезы проводится по общей схеме:

если $\chi^2 < \chi^2_{\alpha, r-1}$, то нулевая гипотеза принимается, в противном случае отвергается.

Если положить $w_i = \frac{m_i}{m_i + l_i}$ и $w = \frac{n_1}{n_1 + n_2}$, то формулу можно представить в виде, часто более удобном для практических расчетов:

$$\chi^2 = \frac{(n_1 + n_2)^2}{n_1 n_2} \left(\sum_{i=1}^r \frac{m_i^2}{m_i + l_i} - \frac{n_1^2}{n_1 + n_2} \right) = \frac{1}{w(1-w)} \left(\sum_{i=1}^r m_i w_i - n_1 w \right).$$

Задача 3. Распределение доходов среди рабочих и служащих согласно шведской переписи 1930 г. приведено в таблице. С помощью критерия Пирсона проверьте гипотезу об одинаковости распределения доходов у двух возрастных групп для заводских мастеров и всех рабочих и служащих.

Доходы, 1000 крон	Все рабочие и служащие в промышленности			Заводские мастера		
	возрастные группы			возрастные группы		
	m_i	l_i	w_i	m_i	l_i	w_i
	40–50	50–60		40–50	50–60	
0–1	7831	7558	0,508869	71	54	0,568000
1–2	26 740	20 685	0,563837	430	324	0,570291
2–3	35 572	24 186	0,595267	1072	894	0,545269

Окончание табл.

3-4	20 009	12 280	0,629784	1609	1202	0,572394
4-6	11 527	6776	0,629787	1178	903	0,566074
> 6	6919	4222	0,621039	158	112	0,585185

Решение. Рассматривается гипотеза однородности: обе выборки (по возрастным группам) извлечены из одной генеральной совокупности. Статистика критерия при проверке основной гипотезы имеет вид

$$\chi^2 = n_1 n_2 \sum_{i=1}^6 \frac{\left(\frac{m_i}{n_1} - \frac{l_i}{n_2} \right)^2}{\frac{m_i}{n_1} + \frac{l_i}{n_2}}.$$

Для заводских мастеров получаем $\chi^2 = 4,27$. Отсюда $\chi^2 < \chi_{0,05;5}^2 = 11,1$ и можно считать, что две выборки извлечены из одной генеральной совокупности. Но если сравнить распределение доходов у возрастных групп всех промышленных рабочих и служащих, то получаем $\chi^2 = 840,62$, что указывает на очень высокую степень различия между распределениями. При этом видно, что числа w_i имеют тенденцию возрастать с ростом доходов.

16.2. Критерий согласия Колмогорова

Критерий согласия Колмогорова применяют для проверки гипотез о законах распределения только непрерывных случайных величин. Проверяем гипотезу $H_0: F(x) = F_0(x)$ против альтернативной $H_1: F(x) \neq F_0(x)$.

Критерий основан на том факте, что распределение супремума разности между теоретической и эмпирической функциями распределения $D_n = \sup |F_n(x) - F(x)|$ одинаково для любой $F(x)$. Величину D_n называют **статистикой Колмогорова**.

При малых n для статистики Колмогорова существуют таблицы критических точек $D_{кр}$. Если $D_n < D_{кр}$, то гипотеза H_0 принимается, иначе отвергается. При больших n используют **предельное распределение Колмогорова**. Имеет место следующая теорема.

► Теорема 2 (Колмогорова)

$$P(\sqrt{n}D_n < x) \rightarrow Q(x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2}, \quad n \rightarrow \infty.$$

Для распределения Колмогорова $Q(x)$, предельного для статистики $\lambda = \sqrt{n}D_n$, также существуют таблицы критических точек $\lambda_{кр}$. Практически их используют уже при $n > 20$. Если $\lambda < \lambda_{кр}$, то гипотеза H_0 принимается, иначе отвергается.

Покажем, что распределение статистики Колмогорова D_n не зависит от вида неизвестной функции распределения $F(x)$.

Рассмотрим преобразование $t' = F(t)$, $0 < t' < 1$. Можно предположить, что $F(t)$ — строго возрастающая функция. Тогда существует обратное преобразование $t = F^{-1}(t')$. Разность функций распределения равна

$$F(t) - F_n(t) = t' - F_n(F^{-1}(t')),$$

где $F_n(F^{-1}(t')) = \frac{\text{число } x_i < F^{-1}(t')}{n}$. Однако из неравенства $x_i < F^{-1}(t')$

следует, что $F(x_i) < t'$. Пусть $F(x_i) = \eta_i$ и обозначим через $G_n(t')$ выборочную функцию распределения для $\eta_1, \eta_2, \dots, \eta_n$.

Теоретическая функция распределения для η следующая:

$$G(t') = P(\eta_i < t') = P(F(x_i) < t') = P(x_i < F^{-1}(t')) = F(F^{-1}(t')) = t',$$

и в силу свойств вероятности можно записать

$$G(t') = \begin{cases} 0, & t' \leq 0; \\ t', & 0 < t' < 1; \\ 1, & t' \geq 1 \end{cases}$$

Следовательно, $G(t')$ является функцией распределения равномерной случайной величины, заданной на отрезке $[0, 1]$, т.е. $F(x_i)$ имеет равномерное распределение на отрезке $[0, 1]$.

Тогда статистика

$$D_n = \sup_{-\infty < t < \infty} |F(t) - F_n(t)| = \sup_{0 < t' < 1} |t' - F_n(F^{-1}(t'))| = \sup_{0 < t' < 1} |t' - G_n(t')|$$

имеет ту же величину, но уже для известного (равномерного) распределения. Таким образом доказано, что распределение статистики D_n не зависит от вида неизвестной функции распределения $F(x)$.

Функция распределения $F(x)$ является разрывной, причем она имеет разрывы только первого рода, являющиеся скачками,

и поэтому выборочную статистику D_n в общем случае определяют с помощью точной верхней границы (sup):

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|.$$

Введем также статистики $D_n^- = \sup_{-\infty < x < \infty} [F_0(x) - F_n(x)]$ и

$$D_n^+ = \sup_{-\infty < x < \infty} [F_n(x) - F_0(x)], \text{ тогда } D_n = \max[D_n^-, D_n^+].$$

Статистики D_n^- , D_n^+ называют статистиками Смирнова.

На практике статистику Колмогорова можно вычислить по эквивалентным формулам:

$$D_n = \max_{1 \leq i \leq n} \left\{ \left| F_0(x_{(i)}) - \frac{i-1}{2n} \right|, \left| \frac{i}{n} - F_0(x_{(i)}) \right| \right\} \text{ и } D_n = \max_{1 \leq i \leq n} \left| F_0(x_{(i)}) - \frac{2i-1}{2n} \right| + \frac{1}{2n},$$

где $x_{(i)}$ — члены вариационного ряда.

Заметим, что критерий Колмогорова, строго говоря, *нельзя* применять в случаях сгруппированных данных при неизвестных параметрах распределения. Тем не менее, его иногда применяют на практике и в подобных ситуациях. Однако при этом статистики критерия получаются *заниженными*, что увеличивает ошибку первого рода. В таких случаях предпочтительнее пользоваться критерием хи-квадрат Пирсона.

Проверка гипотезы однородности с помощью критерия Колмогорова—Смирнова состоит в следующем. Пусть $x_1, x_2, \dots, x_{n_1}$ и $y_1, y_2, \dots, y_{n_2}$ — выборки из двух генеральных совокупностей. Требуется проверить нулевую гипотезу H_0 о совпадении законов распределения генеральных совокупностей, из которых произведены выборки.

Определим эмпирические функции распределения $F_1^{(n_1)}$ и $F_2^{(n_2)}$. Для проверки гипотезы вводятся статистики:

$$D_{n_1, n_2}^- = \sup_{-\infty < x < \infty} (F_1^{(n_1)}(x) - F_2^{(n_2)}(x));$$

$$D_{n_1, n_2}^+ = \sup_{-\infty < x < \infty} (F_2^{(n_2)}(x) - F_1^{(n_1)}(x));$$

$$D_{n_1, n_2} = \max(D_{n_1, n_2}^+, D_{n_1, n_2}^-) = \sup_{-\infty < x < \infty} |F_1^{(n_1)}(x) - F_2^{(n_2)}(x)|.$$

Пусть $n_1, n_2 \rightarrow \infty$, $n_0 = \frac{n_1 n_2}{n_1 + n_2}$ и предельные функции $F_1(x)$ и $F_2(x)$ непрерывны. Тогда в условиях истинности нулевой гипотезы

тезы статистика $D_{n_1, n_2} \sqrt{n_0}$ имеет в пределе распределение Колмогорова. Критической областью является область больших значений, т.е. основная гипотеза отклоняется, если $D_{n_1, n_2} \sqrt{n_0} > \lambda_\alpha$, где λ_α — критическая точка распределения Колмогорова, соответствующая уровню значимости α .

Задача 4. Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение пяти поездок фиксировал время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 (мин.). Проверить гипотезу о том, что время ожидания равномерно распределено на отрезке $[0; 10]$ на уровне значимости $\alpha = 0,05$.

Решение. Составим вариационный ряд: 1,2; 3,7; 4,8; 5,1; 9,2.

С учетом того, что в данном случае $F_0(x) = x/10$, $0 \leq x \leq 10$, построим таблицу:

i	$x_{(i)}$	$F_0(x_{(i)})$	$(2i - 1)/(2n)$	$ F_0(x_{(i)}) - (2i - 1)/(2n) $
1	1,2	0,12	0,1	0,02
2	3,7	0,37	0,3	0,07
3	4,8	0,48	0,5	0,02
4	5,1	0,51	0,7	0,19
5	9,2	0,92	0,9	0,02

Таким образом, значение статистики Колмогорова составляет $D_5 = 0,19 + 0,1 = 0,29$.

По таблице критических точек при $\alpha = 0,05$ и $n = 5$ находим $D_{кр} = 0,56$. Поскольку $D < D_{кр}$, нулевая гипотеза (о равномерности распределения) принимается.

Замечание. На самом деле по таким небольшим выборкам, конечно же, нельзя делать далеко идущие выводы.

Задача 5. Выборка из 10 наблюдений приведена в таблице. Проверить с помощью критерия Колмогорова гипотезу о том, что эта выборка из генеральной совокупности, равномерно распределенной на отрезке $[0, 1]$. Уровень значимости $\alpha = 0,05$.

Решение. Пусть $F(x_{(i)}) = t_i$

i	$x_{(i)}$	$F_n(x) = \frac{i}{n}$	$F_0(x_i) = t_i$	$\frac{i}{n} - t_i$	$t_i - \frac{i-1}{n}$
1	0,01	0,1	0,01	0,09	0,01
2	0,09	0,2	0,09	0,11	-0,01
3	0,1	0,3	0,1	0,2	-0,1
4	0,25	0,4	0,25	0,15	-0,05
5	0,33	0,5	0,33	0,17	-0,07
6	0,35	0,6	0,35	0,25	-0,15
7	0,52	0,7	0,52	0,18	-0,08
8	0,73	0,8	0,73	0,07	0,03
9	0,76	0,9	0,76	0,14	-0,04
10	0,86	1	0,86	0,14	-0,04

Из таблицы получаем $D_{10} = 0,25$.

По заданному уровню значимости $\alpha = 0,05$ находим критическую точку $D_{кр} = 0,41$. Поскольку $D_{10} < D_{кр}$, то гипотеза принимается.

Задача 6. В таблице приведены условные данные о заработной плате $n_1 = 100$ и $n_2 = 100$ служащих двух отраслей народного хозяйства. Проверить с помощью критерия Колмогорова—Смирнова гипотезу о том, что распределение заработной платы служащих первой отрасли ($F_1(x)$) совпадает с распределением заработной платы служащих второй отрасли ($F_2(x)$). Уровень значимости $\alpha = 0,05$.

Решение.

Заработная плата	x_i	y_i	i	Заработная плата	$F_1(x_{(i)})$	$F_2(x_{(i)})$	$F_1(x_{(i)}) - F_2(x_{(i)})$
130—150	4	1	1	150	0,04	0,01	0,03
150—170	4	1	2	170	0,08	0,02	0,06
170—200	15	8	3	200	0,23	0,10	0,13
200—250	51	43	4	250	0,74	0,53	0,21
250—300	22	34	5	300	0,96	0,87	0,09
300—350	3	7	6	350	0,99	0,94	0,05
350—400	1	3	7	400	1,00	0,97	0,03
400—500	—	3	8	500	1,00	1,00	0,00

Используя значения эмпирических функций распределения в правых концах интервалов, получаем данные для расчета критических статистик. Получаем

$$\lambda = D_{100,100} \sqrt{50} = 0,21 \sqrt{50} \approx 1,4849,$$

так как

$$n_0 = \frac{n_1 n_2}{n_1 + n_2} = \frac{n}{2} = 50.$$

По таблице критических точек распределения Колмогорова находим $\lambda_{кр} = 1,36$. Поскольку $\lambda > \lambda_{кр}$, гипотеза однородности отвергается.

Задачи для самостоятельного решения

Теоретические задачи

1. В городе 17 036 семей имеют двоих детей. В 4529 семьях — два мальчика, в 4019 — две девочки, в 8488 семьях — мальчик и девочка. Можно ли на уровне значимости 0,05 считать, что количество мальчиков в семьях с двумя детьми имеет биномиальное распределение с вероятностью рождения мальчика 0,515?
2. В следующей таблице приведены данные о числе покупок, сделанных в магазине (в первой строке — число покупок, во второй — количество покупателей, сделавших такое число покупок).

x_i	1	2	3	4	5	6	7	8
n_i	298	152	77	35	20	11	4	3

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что число покупок для одного покупателя имеет геометрическое распределение с параметром $p = 1/2$.

3. В следующей таблице представлены данные о числе крупных сделок, заключенных на фондовой бирже за квартал, для 400 инвесторов.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что число крупных сделок, заключенных инвестором за квартал, имеет распределение Пуассона.

4. Проведено исследование посещаемости популярного интернет-сайта. Долгое время регистрируется число людей, посетивших сайт в течение данного часа. Результаты исследования представлены в таблице.

Число посетителей	Часы	Число посетителей	Часы
0	57	7	139
1	203	8	45
2	383	9	27
3	525	10	10
4	532	11	4
5	408	12	1
6	273	14	1

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что посещаемость сайта можно описать распределением Пуассона.

5. В следующей таблице представлены сгруппированные данные о заработной плате (в тыс. руб.) мужчин и женщин, работающих на крупной фирме.

Зарплата	Женщины	Мужчины
10–30	80	60
30–50	160	180
50–100	120	240
> 100	40	120

Проверить гипотезу однородности на уровне значимости $\alpha = 0,05$.

6. В следующей таблице представлены сгруппированные данные о продолжительности жизни мужчин и женщин в некотором городе.

Продолжительность жизни, лет	Женщины	Мужчины
< 40	39	197
40–50	56	111
50–60	103	136
60–70	168	161
70–80	258	185
> 80	376	210

Проверить гипотезу однородности на уровне значимости $\alpha = 0,05$.

7. В следующей таблице приведены сгруппированные данные о моментах прибытия машин к бензоколонке в течение 10 ч наблюдений (с 8 до 18 ч).

Время	Число машин
8—9	12
9—10	40
10—11	22
11—12	16
12—13	28
13—14	6
14—15	11
15—16	33
16—17	18
17—18	14

На уровне значимости $\alpha = 0,01$ проверить гипотезу, что моменты прибытия машин распределены равномерно.

8. В следующей таблице приведены сгруппированные данные по срокам службы (ч) для 1000 изделий:

x_i	0—10	10—20	20—30	30—40	40—50	50—60	60—70
n_i	365	245	150	100	70	45	25

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что срок службы изделия имеет показательное распределение.

9. В следующей таблице представлены данные о месячных доходах жителей региона (тыс. руб.) для 1000 жителей:

x_i	< 5	5—10	10—15	15—20	20—25	> 25
n_i	58	96	239	328	147	132

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что доходы жителей региона можно описать нормальным распределением.

7. В следующей таблице представлены данные об удоях коров на молочной ферме за лактационный период (ц) для 100 коров.

x_i	4—6	6—8	8—10	10—12	12—14	14—16	16—18	18—20	20—22	22—24	24—26
n_i	1	3	6	11	15	20	14	12	10	6	2

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что удон коров можно описать нормальным распределением.

11. В ОТК были измерены диаметры 300 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала (нм) даны в таблице.

Границы отклонений	Середина интервала	Число валиков	Границы отклонений	Середина интервала	Число валиков
-30...-25	-27,5	3	0...5	2,5	55
-25...-20	-22,5	8	5...10	7,5	30
-20...-15	-17,5	15	10...15	12,5	25
-15...-10	-12,5	35	15...20	17,5	14
-10...-5	-7,5	40	20...25	22,5	8
-5...0	-2,5	60	25...30	27,5	7

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что отклонения диаметров от номинала можно описать нормальным распределением.

12. За 6 рабочих дней спрос на некоторый товар составлял: 104; 80; 96; 120; 113; 82 кг. На уровне значимости $\alpha = 0,1$ проверить гипотезу о том, что спрос равномерно распределен на отрезке [75; 125].
13. В поселке Полтинниково все жители имеют доход не менее 50 тыс. руб. в месяц. Выборочное обследование доходов 10 человек дало следующие результаты: 54; 60; 59; 79; 71; 92; 53; 54; 78; 56 тыс. руб. На уровне значимости $\alpha = 0,05$ проверить гипотезу, что случайная величина дохода имеет распределение Парето вида

$$F(x) = \begin{cases} 0, & x < x_0; \\ 1 - (x/x_0)^{-\alpha}, & x \geq x_0, \end{cases}$$

где $x_0 = 50$ (тыс. руб.), $\alpha = 3$.

14. При измерении веса 20 шоколадных батончиков (с номинальным весом 50 г) получены следующие значения (г): 49,1; 50,0; 49,7; 50,5; 48,1; 50,3; 49,7; 51,6; 49,8; 50,1; 49,7; 48,8; 51,4; 49,1; 49,6; 50,9; 48,5; 52,0; 50,7; 50,6.

На уровне значимости $\alpha = 0,05$ проверить гипотезу, что вес батончиков имеет нормальное распределение со средним 50 г и дисперсией 1 г².

15. Урожайность зерновых культур в России в 1992–2001 гг. представлена таблицей.

Год	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Урожай- ность, ц/га	18,0	17,1	15,3	13,1	14,9	17,8	12,9	14,4	15,6	19,4

Проверить на уровне значимости $\alpha = 0,1$ гипотезу, что урожайность можно описать нормальным распределением с параметрами $\alpha = 16$, $\sigma = 2$ (ц/га).

Глава 17

ЭЛЕМЕНТЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

17.1. Основные понятия в анализе временных рядов

Временным рядом называется ряд наблюдений некоторой величины, сделанных в последовательные моменты времени (обычно через равные интервалы).

В экономических приложениях это могут быть, например, объемы производства или продаж каких-либо товаров, цены акций, курсы валют и т.п. Данные могут регистрироваться ежегодно, ежемесячно или ежедневно, а учитывая современные компьютерные технологии — даже ежечасно и ежеминутно.

Основными целями анализа временных рядов является выявление присущих им закономерностей и прогнозирование на будущее (с учетом этих закономерностей).

Существует великое множество методов анализа временных рядов, которым посвящена обширная литература. Эти методы могут применяться каждый в отдельности, в различных комбинациях и в разной последовательности. В зависимости от этого могут получаться различные результаты. Однако для «хороших» временных рядов они будут не очень сильно различаться между собой.

Предполагается, что временной ряд можно разделить на три составляющие:

- 1) **тренд**, т.е. некая общая закономерность роста или убывания;
- 2) **сезонная (или циклическая) компонента**;

3) случайная компонента (шум).

В качестве тренда обычно берут достаточно простую функцию. Наиболее часто используют линейный и экспоненциальный тренды. Первое означает, что величина в среднем растет или убывает в арифметической прогрессии (на столько-то в единицу времени), второе — что она меняется в геометрической прогрессии (во столько-то раз или на столько-то процентов в единицу времени). Иногда рассматривают и более сложные тренды, например полиномиальные и т.п.

Линейный тренд легко построить с помощью линейной регрессии, применяя метод наименьших квадратов к исходным данным. Экспоненциальный тренд можно построить, если перейти к логарифмам наблюдаемых величин, провести линейную регрессию для них, а затем взять экспоненту от полученной линейной функции.

Сезонной компоненты во временном ряду может не быть совсем, их может быть несколько или одна. Например, во временном ряду из ежемесячных данных скорее всего будет сезонная компонента с периодом 12 (годовой цикл). В ряду ежедневных наблюдений могут быть компоненты с периодом 7 (неделя), около 30 (месяц) и 365 (год) и т.д. В ежегодных наблюдениях могут присутствовать многолетние циклы (например, экономические циклы Кондратьева).

Сезонная компонента может быть аддитивной (когда ее амплитуда относительно «среднего» значения остается в среднем постоянной) и мультипликативной (когда колебания происходят пропорционально текущему «среднему», и их амплитуда может расти или убывать со временем согласно тренду). Для описания сезонной компоненты часто используют **сезонные индексы**, показывающие, на сколько (или во сколько раз) значение в данной точке цикла отличается от некоторого «среднего» значения.

Следует отметить, что существуют временные ряды с *ложными* трендами и *ложными* циклами, когда закономерности, выявленные на одном отрезке наблюдений, не подтверждаются на других. Однако здесь такие ряды рассматривать не будем.

Общая схема анализа временного ряда обычно такова, что выявленные в нем закономерности исключают путем некоторых преобразований и далее изучают получившиеся величины, называемые **остатками**, пытаясь выявить закономерности уже

в их поведении, и т.д. Так продолжают до тех пор, пока не останутся чисто случайные колебания (независимые случайные величины).

17.2. Простые методы анализа и прогнозирования временных рядов

Рассмотрим некоторые простые методы анализа и прогнозирования временных рядов, не требующие глубокой теории, на примере нескольких практических задач.

Задача 1. В таблице представлены данные об урожайности (ц/га) зерновых культур в СССР за 1951–1970 гг.

Год	Урожайность	Год	Урожайность
1951	7,4	1961	10,7
1952	8,6	1962	10,9
1953	7,8	1963	8,3
1954	7,7	1964	11,4
1955	8,4	1965	9,5
1956	9,9	1966	13,7
1957	8,4	1967	12,1
1958	11,1	1968	14,0
1959	10,4	1969	13,2
1960	10,9	1970	15,6

Построить линейный тренд.

Решение. Методом наименьших квадратов находим $\hat{r}(t) = 6,97 + 0,34(t - 1950)$, где t — год наблюдения. На рис. 17.1 изображен график урожайности и линейный тренд. Видно, что в поведении остатков (отклонений от тренда) не проявляется какой-либо закономерности, и их можно считать чисто случайными.



Рис. 17.1

Задача 2. В таблице представлены данные о производстве молока (тыс. т) в России с января 1992 г. по октябрь 1996 г. (рис. 17.2).

Месяц/год	1992	1993	1994	1995	1996
Январь	2015	1759	1510	1172	1038
Февраль	2123	1773	1484	1226	1104
Март	2624	2361	1988	1651	1439
Апрель	2891	2649	2211	1859	1521
Май	3335	3203	2559	2392	1827
Июнь	4071	3936	3209	2864	2446
Июль	4040	3861	3204	2714	2369
Август	3392	3321	2687	2420	2081
Сентябрь	2467	2438	2031	1925	1577
Октябрь	2092	1760	1506	1338	1081
Ноябрь	1494	1299	1050	984	
Декабрь	1562	1345	1054	1020	



Рис. 17.2

Построить линейный тренд, вычислить сезонные индексы по данным 1992–1995 гг. Сделать прогноз на 1996 г. и сравнить его с реальными данными.

Решение. Методом наименьших квадратов получаем $\hat{tr}(t) = 2899,9 - 26,64t$, где t — номер месяца (от 1 до 58). Производство молока имеет тенденцию к уменьшению, обусловленную сокращением поголовья молочного скота, и подвержено сильным сезонным колебаниям с максимумом летом и минимумом зимой. При этом величина сезонных колебаний пропорциональна среднему уровню производства.

Вычислим значения $x(t)/\hat{tr}(t)$, выразим их в процентах, а затем усредним для каждого месяца за 1992–1995 гг., и таким образом получим сезонные индексы.

Месяц/год	1992	1993	1994	1995	Сезонный индекс (среднее)
Январь	70,13	68,88	67,59	61,22	66,96
Февраль	74,58	70,16	67,23	64,95	69,23
Март	93,05	94,43	91,16	88,71	91,84

Окончание табл.

Месяц/год	1992	1993	1994	1995	Сезонный индекс (среднее)
Апрель	103,50	107,09	102,64	101,34	103,64
Май	120,54	130,89	120,29	132,32	126,01
Июнь	148,57	162,62	152,75	160,80	156,18
Июль	148,89	161,29	154,47	154,69	154,83
Август	126,25	140,30	131,23	140,06	134,46
Сентябрь	92,74	104,17	100,50	113,16	102,64
Октябрь	79,44	76,06	75,52	79,90	77,73
Ноябрь	57,31	56,79	53,37	59,71	56,80
Декабрь	60,54	59,49	54,30	62,91	59,31

Теперь, чтобы построить прогноз для какого-либо месяца последующих лет, надо умножить значение, получаемое из тренда, на соответствующий сезонный индекс. Для 1996 г. получаем таблицу (результаты округлены до целых).

Месяц	Тренд на 1996 г.	Прогноз на 1996 г.	Реальные данные
Январь	1595	1068	1038
Февраль	1568	1085	1104
Март	1541	1415	1439
Апрель	1515	1569	1521
Май	1488	1875	1827
Июнь	1461	2282	2446
Июль	1435	2221	2369
Август	1408	1893	2081
Сентябрь	1381	1418	1577
Октябрь	1355	1053	1081

Приведенные данные свидетельствуют о хорошем согласии прогноза с реальными данными в первые 5 месяцев — относительная погрешность не превышает 3%. В последующие 4 месяца прогноз ниже реальных данных на 6–10%. Относительная ошибка за октябрь не превышает 3%. Заметим, что наибольшие

различия прогноза с реальными данными наблюдаются в летние месяцы, сезонный индекс которых подвержен наибольшим колебаниям. Эти колебания из года в год имеют порядок 10–20%. С учетом этого можно признать в целом хорошее согласие прогноза с реальными данными.

Задача 3. Годовые прибыли фирмы (тыс. долл.) за 5 лет представлены в таблице:

Год	1	2	3	4	5
Прибыль	99	112	120	135	144

Построить экспоненциальный тренд и сделать прогноз на следующий год.

Решение. Построим временной ряд из логарифмов $y(t) = \ln x(t)$:

Год, t	1	2	3	4	5
$y(t)$	4,60	4,72	4,79	4,91	4,97

Проведя линейную регрессию, получаем линейный тренд: $\hat{tr}_y(t) = 4,52 + 0,09t$.

Возвращаясь к исходным величинам, получаем

$$\hat{tr}_x(t) = \exp(4,52 + 0,09t).$$

В качестве прогноза на следующий год имеем $\hat{tr}_x(6) \approx 157,6$ (тыс. долл.).

17.3. Стационарность. Автокорреляция. Периодограмма

Временной ряд можно рассматривать как последовательность наблюдений некоторого случайного процесса $X(t)$, $t > 0$, в целочисленные моменты времени.

Случайный процесс $X(t)$ называется **стационарным в узком смысле**, если для любого $s > 0$ и любых $0 < t_1 < t_2 < \dots < t_n$ распределение многомерного вектора $(X(t_1), X(t_2), \dots, X(t_n))$ совпадает с распределением вектора $(X(t_1 + s), X(t_2 + s), \dots, X(t_n + s))$.

Таким образом, стационарный в узком смысле случайный процесс ведет себя с любого момента так же (в статистиче-

ском смысле), как и с начала. Однако проверить это условие на практике бывает затруднительно.

Определим характеристики случайного процесса: функцию математического ожидания $m(t) = MX(t)$ и ковариационную функцию $B(t, s) = \text{cov}(X(t), X(s))$, $t, s > 0$.

Отсюда можем выразить дисперсию $DX(t) = B(t, t)$ и среднее квадратическое отклонение $\sigma(t) = \sqrt{DX(t)} = \sqrt{B(t, t)}$.

Корреляционной функцией называется

$$R(t, s) = B(t, s) / (\sigma(t) \cdot \sigma(s)).$$

Случайный процесс $X(t)$ называется **стационарным в широком смысле**, если для него $m(t) = \text{const}$ и $B(t_1 + s, t_2 + s) = B(t_1, t_2)$ для любых $t_{1,2}, s > 0$. Это условие уже проще проверить статистическими методами. Из него также следует, что $R(t_1 + s, t_2 + s) = R(t_1, t_2)$ для любых $t_{1,2}, s > 0$.

Автокорреляционной функцией стационарного случайного процесса называется функция $r(t) = R(s, s + t)$, где $s > 0$ — любое.

Пример 1. Пусть $X(t) = \xi \cos(\omega t + \phi)$, где $\omega > 0$ — число, $\xi \in N(0, 1)$, а ϕ равномерно распределено от 0 до 2π , и эти случайные величины независимы. Данный процесс стационарен как в широком, так и в узком смысле, причем $r(t) = \cos(\omega t)$.

Действительно, поскольку $M\xi = 0$ и $M\xi^2 = 1$, то $m(t) \equiv 0$ и

$$\begin{aligned} B(t_1, t_2) &= MX(t_1)X(t_2) = M\{\xi^2 \cos(\omega t_1 + \phi) \cos(\omega t_2 + \phi)\} = \\ &= \int_0^{2\pi} \cos(\omega t_1 + u) \cos(\omega t_2 + u) \frac{du}{2\pi} = \\ &= \frac{1}{2} \int_0^{2\pi} \{\cos(\omega(t_1 - t_2)) + \cos(\omega(t_1 + t_2) + 2u)\} \frac{du}{2\pi} = \\ &= \frac{1}{2} \cos(\omega(t_1 - t_2)) + \frac{1}{4\pi} \sin(\omega(t_1 + t_2) + 2u) \Big|_0^{2\pi} = \frac{1}{2} \cos(\omega(t_1 - t_2)), \end{aligned}$$

так что $B(t_1 + s, t_2 + s) = B(t_1, t_2)$ для любых $t_{1,2}, s > 0$. Кроме того, $\sigma(t) = \sqrt{B(t, t)} = 1/\sqrt{2}$ и $R(t_1, t_2) = 2B(t_1, t_2) = \cos(\omega(t_1 - t_2))$, откуда следует $r(t) = \cos(\omega t)$.

В случае временных рядов имеем дело с автокорреляционной функцией от целых чисел, т.е. $r(k)$. Ее можно оценить статистически следующим образом.

Выборочной автокорреляционной функцией называют

$$\hat{r}(k) = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Этой формулой имеет смысл пользоваться только при достаточно больших размерах n (не менее 50). Отметим, что точность приближения снижается с ростом k , поэтому на практике обычно ограничиваются изучением ограниченного числа первых значений $r(k)$ (как правило, не более $n/4$). Кроме того, полученная оценка оказывается смещенной (со смещением порядка $1/n$).

График выборочной автокорреляционной функции называют **коррелограммой**.

Построение коррелограммы для реальных временных рядов может давать различные результаты.

Если все значения выборочной автокорреляционной функции достаточно велики и испытывают слабую тенденцию к убыванию, это может свидетельствовать о наличии тренда (или общей нестационарности ряда).

Если после удаления тренда и анализа остатков получаем коррелограмму, похожую на график периодической функции, это может свидетельствовать о наличии одной сезонной (циклической) компоненты с соответствующим периодом.

Если после удаления тренда и сезонной компоненты коррелограмма имеет хаотический характер, причем значения выборочной автокорреляционной функции малы и попадают (все или почти все) в 95%-ный доверительный интервал $-1/n \pm 2/\sqrt{n}$, это свидетельствует о том, что остатки можно считать за «белый шум» (последовательность независимых одинаково распределенных случайных величин).

Некоторые другие случаи поведения коррелограмм будут рассмотрены в следующем параграфе.

Бывают ситуации, когда временные ряды содержат несколько циклических компонент с разными периодами, которые накладываются одна на другую и создают сложную картину. Для их анализа можно применить следующий метод, предложенный А. Шустером в 1898 г. Введем функции

$$A(\lambda) = \frac{2}{n} \sum_{k=1}^n x_k \cos \frac{2\pi k}{\lambda}, \quad B(\lambda) = \frac{2}{n} \sum_{k=1}^n x_k \sin \frac{2\pi k}{\lambda}, \quad S^2(\lambda) = A^2(\lambda) + B^2(\lambda).$$

График функции $S^2(\lambda)$ называют **периодограммой**.

Построенная функция должна принимать большие значения (иметь локальные максимумы — пики) для тех значений λ , которые являются периодами имеющихся у временного ряда циклических составляющих. На практике, в силу наличия случайности временного ряда, это не всегда так, и некоторые максимумы могут не иметь отношения к реальным периодам. Поэтому к анализу периодограммы следует подходить с осторожностью. В настоящее время разработаны также другие, более точные и более сложные методы.

17.4. Модели авторегрессии и скользящего среднего

До сих пор предполагалось, что случайная компонента временного ряда представляет собой «белый шум» (последовательность независимых одинаково распределенных случайных величин). Однако на практике это может оказаться не так, и величины могут быть зависимы. Эта зависимость может возникать и в результате каких-то преобразований ряда, что также необходимо учитывать. Для описания этих явлений используются, в частности, следующие модели.

Процессом скользящего среднего порядка q называется процесс вида

$$X(k) = \xi_k + \theta_1 \xi_{k-1} + \dots + \theta_q \xi_{k-q},$$

где θ_k — некоторые коэффициенты, а ξ_k образуют последовательность независимых одинаково распределенных случайных величин. Определенный таким образом процесс всегда является стационарным.

В коррелограмме процесса скользящего среднего достаточно велики только значения $\hat{r}(k)$ при k от 1 до q (те, для которых коэффициенты θ_k отличны от нуля), а остальные малы и хаотичны.

Процессом авторегрессии порядка q называется процесс вида

$$X(k) = \varphi_1 X(k-1) + \dots + \varphi_q X(k-q) + \xi_k,$$

где ϕ_k — некоторые коэффициенты, а ξ_k образуют последовательность независимых одинаково распределенных случайных величин. Такой процесс является стационарным только при дополнительном условии: все корни характеристического уравнения

$$\lambda^p = \phi_1 \lambda^{p-1} + \dots + \phi_p$$

должны быть по модулю меньше единицы.

Коррелограмма процесса авторегрессии имеет вид экспоненциального убывания или затухающих колебаний (при $p > 1$).

В простейшем случае, при $p = 1$, автокорреляционная функция представляет собой просто геометрическую прогрессию: $r(k) = \phi_1^k$.

Как правило, рассматриваются процессы скользящего среднего и авторегрессии небольших порядков (первого или второго).

В случае, когда применяется модель авторегрессии первого порядка, т.е.

$$X(k) = \phi_1 X(k-1) + \xi_k,$$

можно также провести линейную регрессию $X(k)$ по $X(k-1)$.

Пример 2. Временной ряд описывает значения выхода 70 последовательных партий продукта в химическом производственном процессе (данные расположены по столбцам).

47	71	51	50	48	38	68
64	35	57	71	55	59	38
23	57	50	56	45	55	50
71	40	60	74	57	41	60
38	58	45	50	50	53	39
64	44	57	58	62	49	59
55	80	50	45	44	34	40
41	55	45	54	64	35	57
59	37	25	36	43	54	54
48	74	59	54	52	45	23

Выборочное среднее значение выхода продукции составляет 51,13 единицы, а выборочное среднее квадратическое отклонение — 11,91 единицы. Тренд и циклические компоненты отсутствуют.

Видно, что в ряду имеется отчетливая тенденция к чередованию «вверх-вниз», однако невозможно предсказать точно выход следующей партии. Проведем линейную регрессию $X(k)$ по $X(k-1)$. Результат представлен на рисунке 17.3.

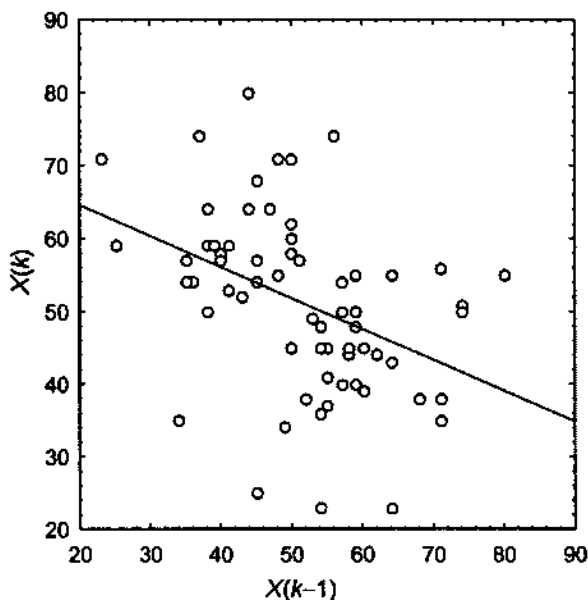


Рис. 17.3

В качестве оценки коэффициента ϕ_1 (как автокорреляции) получаем $\hat{\phi}_1 \approx -0,4$. Подобная отрицательная зависимость часто наблюдается в данных о выпуске продукции и является следствием эффекта «переноса». В данном примере высокий выход цикла производства приводил к получению дегтярных остатков, которые не полностью устранялись из емкости и отрицательно влияли на выход следующего цикла.

Задачи для самостоятельного решения

1. В таблице представлены данные по объемам продаж товаров на оптовом рынке Кировской области за 2002–2003 гг. (тыс. руб.) (рис. 17.4).

Месяц	Год	
	2002	2003
Январь	2174	3174
Февраль	2078	3570
Март	2608	3788
Апрель	2701	3694
Май	2834	3421
Июнь	2852	3522
Июль	2840	4056
Август	3033	4338
Сентябрь	2895	4312
Октябрь	3244	4693
Ноябрь	3133	4621
Декабрь	3126	4746

Построить экспоненциальный тренд. Насколько хорошо он приближает данные? Построить прогноз на 2004 г.



Рис. 17.4

2. В таблице представлены данные по продажам, мороженого «Пломбир» фирмой в Нижнем Новгороде за два года (руб./день) (рис. 17.5)

Месяц	Объем продаж	Месяц	Объем продаж
Июль	8174,40	Июль	8991,84
Август	5078,33	Август	5586,16
Сентябрь	4507,20	Сентябрь	4957,92
Октябрь	2257,19	Октябрь	2482,91
Ноябрь	3400,69	Ноябрь	3740,76
Декабрь	2968,71	Декабрь	3265,58
Январь	2147,14	Январь	2361,85
Февраль	1325,56	Февраль	1458,12
Март	2290,95	Март	2520,05
Апрель	2953,34	Апрель	3248,67
Май	4216,28	Май	4637,91
Июнь	8227,57	Июнь	9050,33

Построить линейный тренд и вычислить сезонные индексы.
Построить прогноз на следующий год.



Рис. 17.5

3. В таблице представлены данные по розничной торговле бензином в США в 1991–1995 гг. (млн долл.) (рис. 17.6).

Месяц	1991	1992	1993	1994	1995
Январь	11 297	10 508	10 839	10 966	11 981
Февраль	10 064	10 071	10 498	10 652	11 443
Март	10 883	10 725	11 476	11 800	12 790
Апрель	11 052	10 885	11 684	11 842	12 701
Май	11 960	11 836	12 346	12 491	13 937
Июнь	11 846	11 874	12 291	12 835	14 210
Июль	12 091	12 225	12 638	13 207	14 013
Август	12 406	12 218	12 418	13 710	14 186
Сентябрь	11 350	11 569	11 679	12 854	13 213
Октябрь	11 678	12 002	12 237	12 983	13 190
Ноябрь	11 360	11 418	11 806	12 647	12 650
Декабрь	11 308	11 619	11 785	12 880	12 931

Построить линейный тренд и вычислить сезонные индексы.



Рис. 17.6

4. В таблице представлены данные по месячным перевозкам пассажиров на международных авиалиниях в 1954–1958 гг. (тыс. чел.) (рис. 17.7).

Месяц	1954	1955	1956	1957	1958
Январь	204	242	284	315	340
Февраль	188	233	277	301	318
Март	235	267	317	356	362
Апрель	227	269	313	348	348
Май	234	270	318	355	363
Июнь	264	315	374	422	435
Июль	302	364	413	465	491
Август	293	347	405	467	505
Сентябрь	259	312	355	404	404
Октябрь	229	274	306	347	359
Ноябрь	203	237	271	305	310
Декабрь	229	278	306	336	337

Построить экспоненциальный тренд и вычислить сезонные индексы.



Рис. 17.7

5. В таблице представлены данные о производстве электроэнергии в России за 1998–2003 гг.

Год	1998	1999	2000	2001	2002	2003
Производство, млрд кВт·ч	827	846	878	891	891	915

Построить экспоненциальный тренд и сделать прогноз на 2004 г.

Глава 18

ЭЛЕМЕНТЫ ЛИНЕЙНОГО РЕГРЕССИОННОГО И КОРРЕЛЯЦИОННОГО АНАЛИЗА

18.1. Функциональная, статистическая и корреляционная зависимости

Для описания, анализа и прогнозирования явлений и процессов в экономике применяют математические модели в форме уравнений или функций. Модели экономического объекта, отражая основные его свойства и абстрагируясь от второстепенных, позволяют судить о его поведении в определенных условиях.

Одной из основных задач математической статистики является исследование зависимости между двумя или несколькими переменными. Две переменные X и Y могут быть независимыми или связанными функциональной либо статистической зависимостью. Строгая функциональная зависимость реализуется редко, так как хотя бы одна из переменных подвержена случайным факторам.

Статистической зависимостью называется такая зависимость, при которой изменение одной из величин влечет за собой изменение распределения другой.

Если при изменении одной из величин изменяется среднее значение другой, то в этом случае статистическая зависимость

является корреляционной. **Корреляционной зависимостью** между двумя переменными величинами называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

Основная цель изучения зависимостей между случайными величинами заключается в прогнозировании с данной вероятностью области значений одной случайной величины на основании наблюдаемых значений другой случайной величины.

На практике при исследовании зависимости между случайными величинами X и Y часто ограничиваются исследованием зависимости между X и условным математическим ожиданием Y .

Напомним, что зависимости такого рода называются **регрессионными** зависимостями. Функцию

$$\varphi(x) = \int_{-\infty}^{+\infty} yp(y|x)dy = M(Y|X=x)$$

называют функцией регрессии первого рода (или модельной функцией регрессии) Y по X , а ее график — линией регрессии Y на X (см. § 7.4).

Аналогично функцию

$$\psi(y) = \int_{-\infty}^{+\infty} xp(x|y)dx = M(X|Y=y)$$

называют функцией регрессии X по Y .

Например, уравнение регрессии Y по X двумерной случайной величины (X, Y) , распределенной по нормальному закону, имеет вид

$$\varphi(x) = \hat{a}_y + r \frac{\sigma_y}{\sigma_x} (x - \hat{a}_x),$$

т.е. функция является линейной и график ее имеет вид прямой линии, проходящей через центр распределения $C(a_x, a_y)$. Аналогично линией регрессии X по Y будет прямая

$$\psi(y) = \hat{a}_x + r \frac{\sigma_x}{\sigma_y} (y - \hat{a}_y).$$

Уравнение регрессии $\hat{y} = \varphi(x)$ делает возможными точечные предсказания значений условных математических ожиданий составляющей Y двумерной случайной величины (X, Y) по зна-

чению составляющей $X = x$. Однако для такого прогноза необходимо знать закон распределения двумерной случайной величины (X, Y) , который на практике, как правило, неизвестен.

Поэтому исследование зависимости случайной величины от ряда неслучайных и случайных величин приводит к моделям регрессии на базе выборочных данных. В качестве оценок условных математических ожиданий принимаются условные средние, которые находятся по данным выборки (x_i, y_i) , $i = 1, 2, \dots, n$, объема n .

Условным средним $\bar{y}_x$ называется среднее арифметическое наблюдавшихся значений y , соответствующих значению $X = x$.

Например, если при $x = 2$ величина y принимает значения $y_1 = 5, y_2 = 6, y_3 = 10, y_4 = 8$, тогда $\bar{y}_x = \frac{5+6+10+8}{4} = \frac{29}{4}$.

Условным средним $\bar{x}_y$ называется среднее арифметическое наблюдавшихся значений X , соответствующих $Y = y$.

Если результаты выборки представить в виде точек (x_i, y_i) , $i = 1, 2, \dots, n$, в декартовой системе координат, то получим точечную диаграмму, называемую корреляционным полем (рис. 18.1).

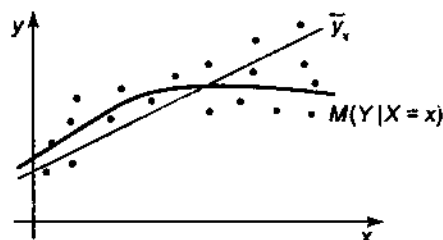


Рис. 18.1

Оценку функции регрессии называют эмпирической регрессией, или функцией регрессии второго рода. Поскольку функции регрессии обладают свойством минимальности, т.е. среднее квадратическое отклонение случайной величины Y от функции $\varphi(x)$ является наименьшим, то для нахождения эмпирических уравнений регрессии $\hat{y} = \varphi(x, a, \dots, c)$ применяется метод наименьших квадратов (см. § 13.3).

Регрессионный анализ — это анализ функций регрессии первого и второго рода. Перед проведением регрессионного анализа необходимо по статистическим данным выбрать об-

щий вид эмпирической функции регрессии. На практике часто считают, что функция регрессии — это линейная функция: $\hat{y} = bx + a$, где b — коэффициент регрессии y на x . Проверка этой гипотезы основывается на результатах оценки коэффициента корреляции.

Если установлено, что зависимость между некоторыми наблюдаемыми величинами существует, то на практике важно знать, какая она: сильная или слабая, положительная или отрицательная. Для выяснения этих обстоятельств используется корреляционный анализ.

Корреляционный анализ — это анализ оценок коэффициента корреляции. Он позволяет ответить на вопрос, существует ли линейная функциональная зависимость между случайными величинами X и Y , и позволяет измерить степень близости статистической зависимости к функциональной.

18.2. Регрессионные модели как инструмент анализа и прогнозирования экономических явлений

В случае применения регрессионных моделей результат действия экономической системы или объекта в виде одного или нескольких выходных показателей можно представить как функцию влияющих на него факторов. Некоторые из них оказывают существенное влияние на результат, другие — весьма незначительное. Как правило, существенных факторов мало, в то время как несущественных — достаточно много, поэтому пренебрегать полностью ими нельзя.

К числу основных факторов при изучении экономического объекта обычно относятся настоящий труд (или трудовые ресурсы) и прошлый труд (энергия, сырье, материалы, оборудование, здания, сооружения и т.д.). Вместе с тем труд прилагается при определенном состоянии внешней среды, т.е. при определенных экономических и природных условиях. Поэтому соответствующие факторы также должны найти отражение в модели.

В процессе производства и распределения продукции имеют место массовые многократно повторяющиеся события. Несмотря на развитие экономики и ее отдельных частей, на протяжении относительно небольших временных периодов и в пределах отдельных экономических подсистем имеет место стабильность в

условиях совершения массовых событий. Так, прежде всего в процессе прогнозирования подразумевается возможность многократного повторения производственной ситуации, при наличии других существенных и несущественных факторов, однако при относительно стабильном комплексе внешних условий.

Таким образом, приходим к следующей теоретико-вероятностной схеме. Результирующий показатель y является функцией существенных $(x_1, x_2, \dots, x_k)$ и несущественных факторов $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$: $y = F(x_1, \dots, x_k, \varepsilon_1, \dots, \varepsilon_n)$.

Несмотря на инертность поведения экономических систем, особенно больших, возможно их резкое изменение в связи с достижениями научно-технического прогресса. В новых условиях, вообще говоря, следует рассматривать другую функцию. Однако исходя из предположения эволюторности поведения экономической системы на относительно небольшом временном интервале и при сравнительно малых изменениях переменных в принципе с помощью полиномов можно с высокой степенью приближения описать любую эволюторную функцию. В случае же невозможности линеаризировать функцию (разлагая в ряд или преобразуя ее) при небольшом числе коэффициентов используют методы нелинейного оценивания.

Линейная регрессия двух переменных и ее различные экономические приложения более подробно разобраны в § 13.3.

18.3. Выборочные коэффициенты корреляции

Основными характеристиками, описывающими степень связи между составляющими X и Y двумерной случайной величины (X, Y) являются ковариация $\text{cov}(X, Y) = \mu_{xy} = M(X - a_x)(Y - a_y)$ и коэффициент корреляции $\rho_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$, который является мерой

линейной зависимости между X и Y .

Так как в случае статистических данных закон распределения двумерной случайной величины неизвестен, для оценки тесноты связи применяется эмпирическая (или выборочная) ковариация $\hat{\mu}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ и эмпирический (или выборочный) коэффициент корреляции.

Выборочным коэффициентом корреляции для выборки вида $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ называется выборочная характеристика

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Выборочный коэффициент корреляции может принимать значения от -1 до $+1$. Он сходится к теоретическому коэффициенту корреляции соответствующих случайных величин, если тот существует. По абсолютной величине и знаку коэффициента можно судить о степени зависимости (сильная или слабая) и характере (положительная или отрицательная).

Выборочный коэффициент корреляции обычно используется в предположении нормального закона распределения данных (нормальность данных). Как известно, в этом случае из равенства нулю теоретического коэффициента ρ следует независимость случайных величин (в более общем случае это неверно). В случае нормального распределения можно проверить гипотезу $\rho = 0$. Пусть

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Если гипотеза $\rho = 0$ верна, то T имеет распределение Стьюдента с $n - 2$ степенями свободы. При уровне значимости α выберем критическую точку $t_{\text{кр}} = t_{\text{кр}}(\alpha; n - 2)$ для двусторонней области. Если $|T| < t_{\text{кр}}$, гипотеза $\rho = 0$ принимается, иначе — отвергается.

В случае, когда нормальность данных нарушается, применение выборочного коэффициента корреляции может вести к ошибкам: либо «не заметим» зависимость между величинами, либо получим *ложную* корреляцию. Существуют коэффициенты и методы, свободные от предположения о нормальности.

Наблюдения всегда можно упорядочить по возрастанию какой-либо переменной (x или y). **Рангом наблюдения** называется его номер в таком ряду. Если какое-то значение переменной встречается несколько раз, ему приписывается *средний* ранг. Обозначим ранги наблюдений по возрастанию x и y через r_i и s_i соответственно. Пусть

$$S = \sum_{i=1}^n (r_i - s_i)^2.$$

Выборочным коэффициентом ранговой корреляции Спирмена называется величина

$$r_s = 1 - \frac{6S}{n^3 - n}.$$

Этот коэффициент также может принимать значения от -1 до $+1$. Аналогичным образом он отражает силу и характер зависимости между величинами. Для проверки гипотезы о независимости случайных величин существуют специальные таблицы критических точек. Однако при больших n можно проверять гипотезу так же, как для обычного выборочного коэффициента корреляции.

Заметим, что с помощью коэффициента Спирмена можно анализировать и ситуации, когда некоторый признак объекта («качество», «привлекательность» и т.п.) нельзя строго выразить численно, но можно упорядочить объекты по его возрастанию или убыванию, т.е. **поранжировать** их.

Задача 1. В табл.18.1 представлены средние цены на растительное масло и сахар-песок в 12 городах Центрального района России на июнь 1996 г. (руб.) (рис. 18.2).

Таблица 18.1

Город	Цена на масло	Цена на сахар
Брянск	7726	3410
Владимир	7880	3183
Иваново	6182	3209
Калуга	8237	3400
Кострома	8750	3600
Москва	11 024	4418
Орел	8456	3634
Рязань	9172	4033
Смоленск	8320	3909
Тверь	7083	3416
Тула	8259	3486
Ярославль	7991	3938

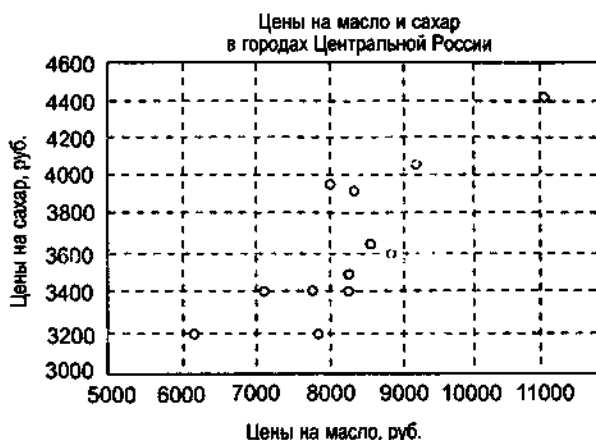


Рис. 18.2

Вычислить выборочный коэффициент корреляции между ценами на растительное масло и сахар. Проверить нулевую гипотезу на уровне значимости $\alpha = 0,1$.

Решение. Для проверки нулевой гипотезы вычисляем выборочный коэффициент корреляции, равный $r = 0,82$, и на-

блюдаемое значение статистики $T = \frac{0,82\sqrt{10}}{\sqrt{1-0,82^2}} \approx 4,53$. Из табли-

цы распределения Стьюдента по заданному уровню значимости $\alpha = 0,1$ и числу степеней свободы $n = 10$ находим критическую точку $t_{кр}(0,1; 10) = 1,81$. Поскольку $|T| > t_{кр}$, нулевая гипотеза отвергается. Связь между ценами на растительное масло и сахар оказывается довольно сильной и положительной.

Задача 2. Два эксперта проранжировали 11 фирм в порядке их привлекательности для инвестиций. Получены следующие последовательности рангов фирм:

r_i	1	2	3	4	5	6	7	8	9	10	11
s_i	1	2	3	5	4	9	8	11	6	7	10

Проверить с помощью коэффициента Спирмена, насколько согласуются мнения экспертов. Проверить нулевую гипотезу на уровне значимости $\alpha = 0,05$.

Решение. Вычисляем $S = 1^2 + 1^2 + 3^2 + 1^2 + 3^2 + 3^2 + 3^2 + 1^2 = 40$, отсюда $r_s = 1 - \frac{6 \cdot 40}{11^3 - 11} \approx 0,82$. Для проверки нулевой гипотезы

вычисляем $T = \frac{0,82\sqrt{9}}{\sqrt{1-0,82^2}} \approx 4,29$ и из таблицы распределе-

ния Стьюдента по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $n = 9$ находим критическую точку $t_{кр}(0,05; 9) = 2,26$. Поскольку $|T| > t_{кр}$ нулевая гипотеза отвергается. Мнения экспертов достаточно хорошо согласуются между собой.

Задачи для самостоятельного решения

1. В табл.18.2 представлены средние цены на растительное масло, сахар-песок, говядину и белый хлеб высшего сорта в 12 городах Центрального района России на июнь 1996 г. (руб.):

Таблица 18.2

Город	Цена на масло	Цена на сахар	Цена на говядину	Цена на хлеб
Брянск	7726	3410	12 500	4875
Владимир	7880	3183	13 857	7125
Иваново	6182	3209	14 150	4998
Калуга	8237	3400	12 697	5170
Кострома	8750	3600	13 000	5476
Москва	11 024	4418	14 120	6466
Орел	8456	3634	10 678	4200
Рязань	9172	4033	12 163	4720
Смоленск	8320	3909	12 833	4354
Тверь	7083	3416	14 400	5440
Тула	8259	3486	12 083	5140
Ярославль	7991	3938	14 397	5283

Вычислить выборочные коэффициенты корреляции между ценами:

- а) на масло и говядину;
- б) масло и хлеб;
- в) сахар и говядину;
- г) сахар и хлеб;
- д) говядину и хлеб.

В каждом случае проверить нулевую гипотезу на уровне значимости 0,05. Сделать вывод о зависимости цен.

2. Решить предыдущую задачу, вычислив выборочные коэффициенты ранговой корреляции Спирмена для цен на товары.
3. Два эксперта проранжировали 9 проектов реорганизации фирмы по их предполагаемой эффективности. Получены следующие последовательности рангов.

r_i	1	2	3	4	5	6	7	8	9
s_j	4	1	5	3	2	6	9	8	7

Проверить с помощью коэффициента Спирмена, насколько согласуются мнения экспертов. Проверить нулевую гипотезу на уровне значимости $\alpha = 0,1$.

4. В таблице представлены данные о производстве всех зерновых культур и производстве пшеницы в России за 1996–2002 гг.

Год	1996	1997	1998	1999	2000	2001	2002
Производство зерновых, млн т	69,3	88,6	47,9	54,7	65,5	85,2	86,6
Производство пшеницы, млн т	34,9	44,3	27,0	31,0	34,5	47,0	57,7

Вычислить выборочный коэффициент корреляции и выборочный коэффициент ранговой корреляции Спирмена между показателями.

5. В табл. 18.3 представлены данные о годовых доходах и расходах на личное потребление для 10 семей (долл.).

Таблица 18.3

Годовой доход	Расходы на личное потребление
2508	2406
2572	2464
2408	2336
2522	2281
2700	2641
2531	2385
2390	2297
2595	2416
2524	2460
2685	2448

Вычислить выборочный коэффициент корреляции и выборочный коэффициент ранговой корреляции Спирмена между показателями.

ГЛАВА 19

ЭЛЕМЕНТЫ ДИСПЕРСИОННОГО АНАЛИЗА

19.1. Основные понятия и определения

Предположим, что имеется k выборок с объемами $n_1, n_2, \dots, n_k$, $N = n_1 + n_2 + \dots + n_k$, и наблюдения можно представить в виде $x_{ij} = a_j + \varepsilon_{ij}$, где i — номер наблюдения в выборке; j — номер выборки; a_j — групповые математические ожидания; ε_{ij} — случайные ошибки с $M\varepsilon_{ij} = 0$, о которых предполагается, что они независимы и одинаково распределены.

Подобная ситуация возникает, когда существует некий фактор, принимающий k различных значений (называемых уровнями), и каждая группа объектов, чьи признаки мы измеряем, подвергается воздействию определенного уровня этого фактора. Методы математической статистики, изучающие воздействие одного фактора на объекты и их признаки, называют в совокупности **однофакторным анализом**.

Предположим, что ошибки нормально распределены: $\varepsilon_{ij} \in N(0, \sigma^2)$. Тогда можно изучать влияние фактора, вычисляя дисперсии некоторых величин. Совокупность этих методов называют **однофакторным дисперсионным анализом**.

Основной гипотезой, нуждающейся в проверке, является гипотеза о равенстве групповых средних — $H_0: a_1 = a_2 = \dots = a_k$. Иными словами, проверяют гипотезу о том, что фактор *вообще не влияет* на наблюдения. В случае нормальных ошибок

ее можно проверить, вычислив две разные оценки дисперсии σ^2 , а именно:

$$\sigma^{2*} = \frac{1}{N-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - x_{.j})^2; \quad \sigma^{2**} = \frac{1}{k-1} \sum_{j=1}^k n_j (x_{.j} - \bar{x})^2,$$

где через $x_{.j}$ обозначены групповые выборочные средние. Первую величину называют **остаточной дисперсией**, а вторую — **факторной**.

При нарушении гипотезы H_0 оценка σ^{2**} имеет тенденцию к возрастанию, тем большему, чем больше отклонение от H_0 . Проверить гипотезу можно с помощью уже знакомого нам дисперсионного отношения Фишера: $F = \sigma^{2**}/\sigma^{2*}$, которое должно иметь распределение Фишера с $k-1$ и $N-k$ степенями свободы. При заданном уровне значимости α можно найти критическую точку $F_{\text{кр}} = F_{\text{кр}}(\alpha; k-1; N-k)$. Если $F < F_{\text{кр}}$, то нулевая гипотеза принимается, иначе отвергается. Заметим, что если $F < 1$, то следует сразу принять гипотезу H_0 , поскольку $F_{\text{кр}}$ всегда больше единицы.

Если гипотеза о равенстве средних не подтверждается, имеет смысл оценивать величины a_j по отдельности. Получаем для них следующие доверительные интервалы (с надежностью γ):

$$x_{.j} - \frac{\sigma^*}{\sqrt{n_j}} t_\gamma < a_j < x_{.j} + \frac{\sigma^*}{\sqrt{n_j}} t_\gamma,$$

где t_γ — критическая точка распределения Стьюдента с $N-k$ степенями свободы и уровнем значимости $\alpha = 1 - \gamma$ (для двусторонней области): $t_\gamma = t_{\text{кр}}(1 - \gamma; N-k)$.

Существуют и более сложные методы однофакторного анализа, не требующие предположения о нормальности наблюдений. Однако здесь будем придерживаться классического дисперсионного анализа.

Задача 1. Банк имеет по четыре отделения в трех городах. Текущие объемы денежных вкладов (в условных единицах) представлены в следующей таблице:

Отделение	Город		
	1	2	3
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что в среднем дела идут одинаково хорошо во всех трех городах? Построить доверительные интервалы для средних по городам с надежностью 90%.

Решение. Вычисляем групповые средние $x_{\cdot 1} = 35$, $x_{\cdot 2} = 25$, $x_{\cdot 3} = 27$ и общее среднее $\bar{x} = 29$. Затем вычисляем оценки дисперсий: $\sigma^{2**} = 112$, $\sigma^{2*} \approx 22,67$. Получаем $F \approx 4,94$. По таблице находим критическую точку: $F_{\text{кр}}(0,05; 2; 9) = 4,26$. Имеем $F > F_{\text{кр}}$, следовательно, групповые средние различаются. Чтобы построить доверительные интервалы, находим $\sigma^* \approx 4,76$ и $t_{\gamma} = t_{\text{кр}}(0,1; 9) = 1,83$. Точность доверительных интервалов во всех группах составляет $\sigma^* t_{\gamma} / \sqrt{n_j} \approx 4,36$, отсюда $30,64 < a_1 < 39,36$; $20,64 < a_2 < 29,36$; $22,64 < a_3 < 31,36$. Таким образом, лучше всего дела идут в первом городе, с явным отрывом от двух других, где ситуация примерно одинакова (доверительные интервалы перекрываются).

Задача 2. В цеху работает четыре бригады по семь рабочих. Выработка каждого (в условных единицах) представлена в следующей таблице:

Рабочий	Бригада			
	1	2	3	4
1	51	52	56	54
2	59	58	56	58
3	53	66	58	62
4	59	69	58	64
5	63	70	70	66
6	69	72	74	67
7	72	74	78	69

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что все бригады работают в среднем одинаково?

Решение. Вычисляем групповые средние $x_{\cdot 1} = 60,9$; $x_{\cdot 2} = 65,9$; $x_{\cdot 3} = 64,3$; $x_{\cdot 4} = 62,9$ и общее среднее $\bar{x} = 63,5$. Затем вычисляем оценки дисперсий: $\sigma^{2**} = 31,67$; $\sigma^{2*} \approx 60,25$. Получаем $F < 1$, так что следует принять гипотезу о равенстве групповых средних. Все бригады работают в среднем одинаково.

Задачи для самостоятельного решения

1. Фирма имеет по пять филиалов в четырех регионах страны. Прибыли за отчетный период (в условных единицах) представлены в таблице:

Филиал	Регион			
	1	2	3	4
1	36	56	52	39
2	47	61	57	57
3	50	64	59	63
4	58	66	58	61
5	67	66	79	65

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что дела во всех регионах идут в среднем одинаково?

2. Фирма имеет 13 магазинов в трех округах Москвы: 4 — в Северном округе, 6 — в Южном, 3 — в Восточном. Объемы продаж за отчетный период (в условных единицах) по каждому магазину представлены в таблице:

Филиал	Регион		
	Северный округ	Южный округ	Восточный округ
1	37	60	69
2	47	86	100
3	40	67	98
4	60	92	
5		95	
6		98	

Можно ли утверждать на уровне значимости $\alpha = 0,01$, что продажи во всех округах идут в среднем одинаково? Построить доверительные интервалы для средних по регионам с надежностью 95%.

3. В целях изучения влияния материального вознаграждения на производительность труда проведено обследование четырех групп рабочих по 5 человек, упорядоченных по величине вознаграждения (от меньшей к большей). Выработка (в условных единицах) представлена в таблице:

Рабочий	Группа			
	1	2	3	4
1	8	12	12	24
2	10	17	15	16
3	16	14	16	22
4	13	9	16	18
5	12	16	19	20

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что вознаграждение влияет на производительность труда? Построить доверительные интервалы для средних по группам с надежностью 90%.

4. В продуктовом магазине продается четыре вида соков – яблочный, апельсиновый, персиковый и ананасовый (за одну цену). Данные о продажах за пять рабочих дней представлены в таблице (л):

День	Сок			
	яблочный	апельсиновый	персиковый	ананасовый
1	18	28	26	20
2	24	31	28	29
3	25	32	30	32
4	29	33	29	30
5	34	33	40	33

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что спрос на разные виды соков в среднем одинаковый?

5. Изучается популярность еженедельной передачи у трех возрастных групп населения. Рейтинги, %, за 4 недели представлены в таблице:

Неделя	Группа		
	молодежь	средний возраст	пожилые
1	27	24	22
2	23	20	21
3	29	26	36
4	29	30	37

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что передача пользуется одинаковой популярностью у всех трех групп населения?

ПРИЛОЖЕНИЯ

Приложение 1

1. Примеры вариантов контрольных и экзаменационных работ

Контрольная работа № 1 по теории вероятностей.

Случайные события

Вариант 1

1. Основные правила комбинаторики. Сочетания, размещения, перестановки. Разбиение множества на группы. Дискретное пространство элементарных исходов. Простые и сложные события. Классическое определение вероятности.

2. В урне 9 белых шаров, 7 красных и 4 черных. Найти вероятность достать (без возвращения) три шара, не все цвета которых одинаковы.

3. Два аудитора проверяют 8 фирм (по 4 фирмы каждый), у трех из которых имеются нарушения. Вероятность обнаружения нарушений первым аудитором равна 0,6, вторым — 0,9. Найти вероятность, что все фирмы-нарушители будут выявлены.

4. В цехе работают 7 мастеров и 3 ученика, производящие одинаковое число изделий. Мастер допускает брак в 1% случаев, а ученик — в 5% случаев. Изделие оказалось бракованным. Найти вероятность, что его сделал ученик.

5. Деталь является бракованной с вероятностью 2%. Сколько деталей надо взять, чтобы с вероятностью 99% среди них оказалась хотя бы одна бракованная? Найти наименьшее число бракованных деталей в этом случае.

Вариант 2

1. Сигма-алгебры и аксиомы вероятности. Доказать эквивалентность счетной аддитивности и непрерывности. Геометрическая вероятность.

2. В партии из 30 изделий 5 бракованных. Найти вероятность, что в случайной выборке из 10 изделий содержится не более одного бракованного.

3. В одной коробке 4 красных шара и 6 синих, а во второй — 8 красных и 2 синих. Из первой во вторую переложили два шара, а затем извлекли из второй два шара без возвращения. Найти вероятность, что последние оказались одного цвета.

4. На заводе установлена система аварийной сигнализации, которая при наличии аварии срабатывает с вероятностью 99%. Однако в 0,2% случаев, когда аварии нет, сигнал также может возникнуть. Найти вероятность, что случилась авария, если сигнализация сработала. Вероятность аварии 0,007.

5. Страховой агент при каждом визите заключает договор с вероятностью 30%. При каком числе визитов наивероятнейшее число договоров будет равно 10?

Вариант 3

1. Независимые испытания Бернулли. Формула Бернулли и неравенство для наивероятнейшего числа успехов (с доказательствами).

2. Шесть клиентов обращаются в 3 фирмы равновероятно. Найти вероятность, что ровно в одну фирму обратятся 2 клиента.

3. Из продукции птицефабрики 75% яиц стандартные, 20% большего объема и 5% двухжелтковых. С какой вероятностью среди 3 случайно выбранных яиц окажутся хотя бы одно большего объема и хотя бы одно двухжелтковое (вместе)?

4. К системному администратору обращаются за помощью пользователи. Среди них начинающих — 60%, опытных — 40%. Вероятность обращения начинающего пользователя — 85%, опытного — 15%. Найти вероятность, что очередной пользователь, обратившийся за помощью, окажется начинающим.

5. Стиральным порошком фирмы А пользуется 25% населения. Сколько людей надо опросить, чтобы получить эту долю с точностью 0,05 с вероятностью 0,95?

Вариант 4

1. Условная вероятность. Полная группа событий. Формула полной вероятности и формула Байеса (с доказательствами).

2. Студент в состоянии решить 20 задач из 25 в первом туре экзамена и 15 из 20 во втором. Найти вероятность сдачи им экзамена, если в каждом туре дается 3 задачи и достаточно решить хотя бы 2 из них.

3. Фирма участвует в четырех независимых проектах, вероятности успеха которых составляют 0,5; 0,6; 0,7 и 0,8. Найти вероятность того, что хотя бы два проекта увенчаются успехом.

4. В город поступают товары трех фирм в соотношении 2:3:5. В поставках первой фирмы 60% товара высшего качества, второй — 40%, а третьей — 20%. Куплен товар высшего качества. Найти вероятность, что он изготовлен второй фирмой.

5. Вероятность того, что пассажир опоздает к поезду, равна 0,005. Найти наиболее вероятное число опоздавших, если всего билет купило 600 человек, и вероятность того, что их будет именно столько.

Вариант 5

1. Операции над событиями. Формулы де Моргана. Независимость событий. Вероятность противоположного события. Теоремы сложения и умножения вероятностей (с доказательствами).

2. В каждой упаковке товара фирмы «Том» имеется одна из букв «Т», «О», «М» равновероятно. Какова вероятность собрать все буквы, купив 5 упаковок товара?

3. В урне 7 красных шаров и 9 синих. Извлекли три шара (без возвращения), затем положили их обратно и добавили в урну два шара того цвета, который оказался в большинстве среди извлеченных. С какой вероятностью два шара, вынутых после этого, окажутся синими?

4. Мимо бензоколонки проезжают легковые и грузовые машины. Среди них грузовых машин 75%. Вероятность того, что проезжающая машина подъедет заправиться, для грузовой машины равна 0,1, для легковой — 0,2. Найти вероятность того, что очередная машина, подъехавшая на заправку, окажется грузовой.

5. 400 клиентов фирмы обращаются к 5 менеджерам равновероятно и независимо друг от друга. В каких границах с вероятностью 0,95 лежит число клиентов отдельно взятого менеджера?

Вариант 6

1. Теорема Муавра—Лапласа и следствия. Теорема Пуассона (с доказательством).

2. Группа из 27 студентов пишет контрольную из 3 вариантов (по 9 человек в каждом). Найти вероятность, что среди случайно выбранных 5 студентов есть писавшие каждый вариант.

3. Фирма имеет трех поставщиков, каждый из которых надежен с вероятностью 0,8. В случае отказа одного поставщика вероятность разорения фирмы равна 0,25, двух — 0,75, трех — 0,95. Найти вероятность разорения фирмы.

4. Из урны, где было 10 красных, 6 белых и 4 синих шара, вынут один шар. После этого из урны извлечены (без возвращения) 2 шара, оказавшиеся красным и белым. При этом условии найти вероятность, что сначала был вынут синий шар.

5. Страховая фирма заключила 2500 договоров. Вероятность страхового случая по каждому в течение года составляет 2%. Найти вероятность того, что число таких случаев не превзойдет наивероятнейшее число более чем на 20.

Контрольная работа № 2 по теории вероятностей.

Случайные величины. Законы распределения

Вариант 1

1. Два станка выпускают некоторую деталь соответственно с вероятностями брака $p_1 = 0,01$ и $p_2 = 0,05$. В выборке одна деталь, выпущенная первым станком, и две со второго станка. Найти распределение вероятностей случайной величины ξ — числа бракованных деталей в выборке.

2. Совместный закон распределения пары (ξ, η) задан таблицей:

ξ	η		
	-1	0	1
-1	1/16	1/8	1/16
1	3/16	3/8	3/16

Найти распределения вероятностей $\xi + \eta$ и ковариацию $\text{cov}(\xi - 3\eta, \xi + \eta)$; проверить зависимость ξ и η .

3. Плотность распределения случайной величины ξ имеет вид $p_{\xi}(x) = \begin{cases} C(x+1), & x \in [-1, 2], \\ 0, & x \notin [-1, 2]. \end{cases}$

Найти C , функцию распределения $F(x)$, $P(\xi^2 < 1)$, $M\xi$.

4. Случайная величина равномерно распределена на отрезке $[1, 3]$. Найти плотность распределения вероятностей $\eta = \xi^2 + 1$.

5. Двумерная случайная величина (ξ, η) равномерно распределена на множестве $\{(x, y)/x^2 + y^2 \leq 1, y \geq 0\}$. Найти плотность распределения составляющих ξ и η и ответить на вопрос о зависимости ξ и η (ответ обосновать).

Вариант 2

1. Два стрелка поражают мишень соответственно с вероятностями 0,7 и 0,8 при одном выстреле. Найти распределение вероятностей числа попаданий, если первый стрелок выстрелил один раз, а второй — два раза.

2. Совместный закон распределения пары (ξ, η) задан таблицей:

ξ	η		
	-1	0	1
-1	1/8	1/4	1/8
1	3/8	1/4	1/8

Найти распределение $2\xi + \eta$ и ковариацию $\text{cov}(\xi - \eta, 2\eta)$; проверить зависимость ξ и η .

3. Плотность распределения случайной величины ξ имеет

$$\text{вид } p_{\xi}(x) = \begin{cases} C(x+1)^{-4}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Найти C , функцию $F(x)$, $P(|\xi - 1/3| \leq 1)$ и $M\xi^2$.

4. Случайные величины ξ и η независимы и имеют показательное распределение с параметром $\lambda = 2$. Найти $P(\xi + \eta \geq 1)$.

5. Двумерная случайная величина (ξ, η) равномерно распределена в квадрате с вершинами $(-2, 0)$, $(0, -2)$, $(2, 0)$, $(0, 2)$. Найти значение функции совместного распределения $F(x, y)$ в точке $(1, -1)$. Ответить на вопрос о зависимости ξ и η (ответ обосновать).

Вариант 3

1. Среди 5 ключей два подходят к двери. Ключи пробуют один за другим, пока не откроют дверь. Найти распределение вероятностей ξ — числа опробованных ключей.

2. Совместный закон распределения пары (ξ, η) задан таблицей:

ξ	η		
	-1	0	1
-1	1/10	1/5	1/5
1	1/5	1/10	1/5

Найти распределение $\xi - \eta$ и ковариацию $\text{cov}(2\xi + \eta; \xi - \eta)$; проверить зависимость ξ и η .

3. Плотность распределения случайной величины ξ имеет

$$\text{вид } p_{\xi}(x) = \begin{cases} C(1-x^2), & |x| \leq 1; \\ 0, & |x| > 1 \end{cases}$$

Найти C , функцию распределения $F(x)$, $P(|\xi - 1/2| \leq 1/4)$, $M\xi$.

4. Случайная величина ξ имеет равномерное распределение на отрезке $[-1; 1]$. Найти плотность случайной величины $\eta = \text{tg}(\pi\xi/2)$.

5. Двумерная случайная величина равномерно распределена в треугольнике $(0, 4)$, $(4, 0)$ и $(4, 4)$. Найти плотность распределения составляющих ξ и η . Ответить на вопрос о зависимости случайных величин ξ и η (ответ обосновать).

Вариант 4

1. С конвейера поступило 4 детали. Вероятность брака для каждой детали p . Детали проверяют одну за другой, пока не наберут 2 доброкачественные. Найти распределение вероятностей числа ξ проверенных деталей.

2. Совместный закон распределения пары (ξ, η) задан таблицей:

ξ	η		
	-1	0	1
-1	1/12	1/6	1/4
1	1/4	1/12	1/6

Найти закон распределения вероятностей $\xi\eta$, ковариацию $\text{cov}(\xi - \eta, \xi + \eta)$; проверить зависимость случайных величин ξ и η .

3. Плотность распределения случайной величины ξ имеет вид

$$p_{\xi}(x) = \begin{cases} C(x^3 + 2), & x \in [1, 3]; \\ 0, & x \notin [1, 3]. \end{cases}$$

Найти C , функцию распределения $F(x)$ и $M\xi$, $P(-1 < \xi < 2)$.

4. Случайная величина ξ имеет равномерное распределение на отрезке $[-1, 1]$. Найти плотность распределения $\eta = -\ln(2 + \xi)$.

5. Случайная пара (ξ, η) имеет равномерное распределение на множестве $\{x, y: |x| + |y| \leq 2\}$. Найти плотности распределения ξ и η . Являются ли они зависимыми? (ответ обосновать).

Вариант 5

1. Со склада получено 4 прибора. Каждый прибор комплектуется из двух деталей, вероятность брака для первой детали — 0,1, а для второй — 0,05. Найти распределение числа бракованных приборов.

2. Совместный закон распределения пары (ξ, η) задан таблицей:

ξ	η		
	-1	0	1
-1	1/16	1/16	1/8
1	3/16	3/16	3/8

Найти распределение $\xi + \eta$, ковариацию $\text{cov}(\xi + 2\eta, 2\xi + \eta)$; проверить зависимость ξ и η .

3. Случайная величина ξ имеет плотность распределения

$$p_{\xi}(x) = \begin{cases} C(9 - x^2), & x \in [-3, 2]; \\ 0, & x \notin [-3, 2]. \end{cases}$$

Найти C , функцию распределения $F(x)$ и $M\xi$, вероятность $P(-1 < \xi < 3)$.

4. Случайная величина ξ равномерно распределена на отрезке $[-1/2, 1/2]$. Найти плотность распределения случайной величины $\eta = \sin \frac{\pi \xi}{2}$.

5. Случайная пара (ξ, η) равномерно распределена на множестве $\{x, y: |x| + |y| \leq 1; y \geq 0\}$. Найти плотности распределения ξ и η . Ответить на вопрос о зависимости ξ и η (ответ обосновать).

**2. Примеры вариантов экзаменационных работ
по теории вероятностей
(специальность «Экономика»)**

Вариант 1

1. Пространство элементарных исходов. События (простые и сложные). Операции над событиями. Формулы де Моргана. Классическое определение вероятности.

2. Ковариация, коэффициент корреляции и их свойства.

3. Группа из 20 студентов пишет контрольную работу из 4 вариантов (по 5 человек в каждом). Найти вероятность того, что среди случайно выбранных 5 студентов окажутся писавшие каждый вариант.

4. Двумерная случайная величина (ξ, η) равномерно распределена на множестве $\{0 \leq y \leq 4 - x^2\}$. Проверить величины ξ и η на зависимость, найти их частные плотности.

5. Плотность случайной величины ξ равна:

$$p_{\xi}(x) = \begin{cases} C(5 - x^2), & x \in [0, 2]; \\ 0, & x \notin [0, 2]. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероятность $P(-2 < \xi < 1)$, математическое ожидание $M\xi$ и плотность $\eta = \xi^2 + 1$.

Вариант 2

1. Непрерывная случайная величина. Плотность непрерывной случайной величины и ее свойства. Совместная плотность распределения двумерной случайной величины. Условия независимости случайных величин.

2. Нормальный закон распределения, его функция распределения и плотность. Вычислить математическое ожидание и дисперсию нормально распределенной случайной величины.

3. Семь клиентов случайным образом обращаются в 5 фирм. Найти вероятность, что хотя бы в одну фирму никто не обратится.

4. Найти плотность распределения суммы независимых случайных величин ξ и η , если они равномерно распределены на отрезках $[0, 2]$ и $[1, 4]$ соответственно.

5. Плотность случайной величины ξ равна:

$$p_{\xi}(x) = \begin{cases} C(x+4), & x \in [-1, 3]; \\ 0, & x \notin [-1, 3]. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероятность $P(1 < \xi < 4)$, математическое ожидание $M\xi$ и плотность

$$\eta = \frac{1}{3 - \xi}.$$

Вариант 3

1. Независимость событий, попарная и в совокупности. Независимые испытания Бернулли. Теорема о распределении вероятностей числа успехов в n независимых испытаниях Бернулли. Формула Бернулли. Наивероятнейшее число успехов.

2. Дать определение показательного распределения. Вычислить его математическое ожидание и дисперсию.

3. Найти вероятность, что в 8-значном числе ровно 3 цифры совпадают, а остальные различны (число может начинаться с нуля).

4. Двумерная случайная величина равномерно распределена на множестве $\{0 \leq y \leq 4 - (x-1)^2\}$. Проверить величины ξ и η на независимость, найти их плотности.

5. Плотность случайной величины ξ равна:

$$p_{\xi}(x) = \begin{cases} C(x-1)^{-5}, & x > 2; \\ 0, & x \leq 2. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероятность $P(0 < \xi < 4)$, плотность $\eta = \frac{1}{\xi - 1}$, математическое ожидание $M\xi$.

Вариант 4

1. Полная группа событий. Формула полной вероятности, формула Байеса (с доказательством).

2. Дать определение распределения Бернулли. Вычислить его математическое ожидание и дисперсию.

3. Фирма нарушает закон с вероятностью 0,25. Аудитор обнаруживает нарушение с вероятностью 0,75. Проведенная им проверка не выявила нарушений. Найти вероятность, что на самом деле они есть.

4. Найти плотность распределения суммы независимых случайных величин ξ и η , если ξ равномерно распределена на отрезке $[0, 1]$, а η показательно распределена с параметром $\lambda = 2$.

5. Плотность случайной величины ξ равна:

$$p_{\xi}(x) = \begin{cases} C(x+2)^2, & x \in [0, 2]; \\ 0, & x \notin [0, 2]. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероят-

ность $P(-1 < \xi < 1)$, плотность $\eta = \frac{2}{\xi+3}$, математическое ожидание $M\xi$.

Вариант 5

1. Определение вероятности события. Вероятность противоположного события. Условная вероятность. Теоремы сложения и умножения. Теорема умножения для независимых событий.

2. Дать определение распределения Пуассона. Вычислить его математическое ожидание и дисперсию.

3. Курс акции за день растет или падает на 1 пункт равновероятно. Найти вероятность, что за 10 дней торгов курс акции упадет на 4 пункта.

4. Двумерная случайная величина (ξ, η) равномерно распределена на множестве $\{(x-2)^2 + y^2 \leq 9, y \geq 0\}$. Проверить величины ξ и η на зависимость, найти их частные плотности.

5. Плотность случайной величины ξ равна:

$$p_{\xi}(x) = \begin{cases} Cx^{-5}, & x \geq 1; \\ 0, & x < 1. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероятность $P(0 < \xi < 2)$, плотность $\eta = \sqrt{\xi-1}$, математическое ожидание $M\xi$.

Вариант 6

1. Неравенство Чебышева. Закон больших чисел в форме Чебышева. Теорема Бернулли как частный случай теоремы Чебышева. Формулировка центральной предельной теоремы.

2. Случайная величина и ее законы распределения. Функция распределения случайной величины и ее свойства.

3. Два аудитора проверяют 8 фирм (по 4 фирмы каждый), у двух из которых есть нарушения. Вероятность обнаружения нарушений первым аудитором равна 0,8, вторым – 0,9. Найти вероятность, что все фирмы-нарушители будут выявлены.

4. Найти плотность распределения суммы независимых случайных величин ξ и η , если они показательны распределены с параметрами $\lambda_\xi = 1$ и $\lambda_\eta = 2$ соответственно.

5. Плотность случайной величины ξ равна:

$$p_\xi(x) = \begin{cases} C(x+7), & x \in [-2, 1]; \\ 0, & x \notin [-2, 1]. \end{cases}$$

Найти константу C , функцию распределения $F(x)$, вероятность $P(\xi^2 < 1/4)$, математическое ожидание $M\xi$ и плотность $\eta = (\xi + 2)^2$.

3. Примеры вариантов контрольных по математической статистике

Контрольная работа № 3. Математическая статистика (специальность «Менеджмент»)

Вариант 1

1. Точечные оценки и их свойства. Эмпирическая функция распределения, выборочное среднее, выборочные дисперсии, их свойства. Теорема Фишера.

2. Найти оценки методом моментов для параметров a и b равномерного распределения на $[a, b]$.

3. Найти оценку максимального правдоподобия для параметра θ распределения Вейбулла $F(x) = 1 - e^{-\theta x^2}$, $x > 0$, $\theta > 0$.

4. По данным испытаний 25 автомобильных шин выборочная средняя износостойчивость составила 32 тыс. км с выборочным средним квадратическим отклонением 5 тыс. км. Построить доверительный интервал для средней износостойчивости с надежностью 99%.

5. Автомат производит детали с известным средним квадратическим отклонением размера 0,09 мм. Выборка из 100 деталей показала выборочный средний размер 5,02 мм. Можно ли утверждать на уровне значимости 10%, что средний размер детали соответствует номиналу в 5 мм (против гипотезы, что не соответствует)?

Вариант 2

1. Распределения хи-квадрат и Стьюдента, их определения, свойства, графики, математические ожидания и дисперсии.

2. Найти оценку методом моментов для параметра θ распределения $F(x) = (x/\theta)^2$, $0 \leq x \leq \theta$.

3. Найти оценку методом максимального правдоподобия для параметра p геометрического распределения.

4. Автомат упаковывает товар в пакеты с известным средним квадратическим отклонением веса 1,5 г. По выборке из 25 пакетов получен выборочный средний вес 1,01 кг. Построить доверительный интервал для среднего веса пакета с надежностью 95%.

5. Средний дневной доход фирмы составлял 250 единиц. После реорганизации за 16 рабочих дней выборочный средний доход составил 258 единиц с выборочным средним квадратическим отклонением 12 единиц. Можно ли утверждать на уровне значимости 5%, что реорганизация привела к увеличению среднего дохода фирмы?

Вариант 3

1. Метод моментов. Метод максимального правдоподобия.

2. Найти оценку методом моментов для параметра β распределения $F(x) = x^\beta$, $0 \leq x \leq 1$.

3. Найти оценку максимального правдоподобия для параметра λ распределения Пуассона.

4. Для исследования средней заработной платы работающих жителей города сделана выборка из 256 человек, при этом оказалось, что средняя заработная плата в выборке равна 289 у.е., а выборочная дисперсия 324. С надежностью 95% найти доверительный интервал для средней заработной платы жителей города.

5. Фирма утверждает, что средний срок службы ее изделия составляет 250 часов. По выборке из 25 изделий получен выборочный средний срок службы 243 часа с выборочным средним квадратическим отклонением 11 часов. Подтверждается ли заявление фирмы на уровне значимости 5% (против гипотезы, что средний срок службы на самом деле меньше)?

Вариант 4

1. Информация Фишера. Неравенство Рао—Фреше—Крамера. Эффективность оценки.

2. Найти оценку методом моментов для параметров θ и μ равномерного распределения на $[\theta, \theta + \mu]$.

3. Найти оценку методом максимального правдоподобия для параметра α распределения Парето: $F(x) = 1 - x^{-\alpha}$, $x \geq 1$, $\alpha > 0$.

4. В деревне Сидорово проведено выборочное исследование доходов. По выборке из 16 человек получен выборочный средний доход 5270 руб. с выборочным средним квадратическим отклонением 115 руб. Построить доверительный интервал для среднего дохода с надежностью 95%.

5. Среднее время сборки изделия было 50 мин. При испытаниях нового метода сборки по выборке из 25 изделий получено выборочное среднее время сборки 42 мин. с выборочным средним квадратическим отклонением 7 мин. Можно ли утверждать на уровне значимости 5%, что среднее время сборки сократилось?

Вариант 5

1. Понятие доверительного интервала. Точность и надежность. Доверительные интервалы для параметров нормального распределения.

2. Найти оценку методом моментов для параметра α распределения Парето: $F(x) = 1 - x^{-\alpha}$, $x \geq 1$, $\alpha > 1$.

3. Найти оценку методом максимального правдоподобия для параметра θ равномерного распределения на $[0, \theta]$.

4. Выборочное среднее квадратическое отклонение ежесуточного дохода 30 киосков некоторой фирмы оказалось равно 75 единиц. Построить доверительный интервал для среднего квадратического отклонения с надежностью 90%.

5. В книжном магазине проведено исследование продаж нового романа писателя Бородачева. В течение 25 рабочих дней роман продавался ежедневно в среднем по 36 экземпляров с выборочным средним квадратическим отклонением 9 экземпляров. Можно ли утверждать на уровне значимости 5%, что новый роман расходуется лучше, чем предыдущий, если тот расходился в среднем по 30 экземпляров в день?

Вариант 6

1. Проверка статистических гипотез. Основные понятия. Критерий отношения правдоподобия. Теорема Неймана—Пирсона.

2. Найти оценку методом моментов для параметра θ равномерного распределения на $[-\theta, \theta]$.

3. Найти оценку методом максимального правдоподобия для параметра β распределения $F(x) = x^\beta$, $0 \leq x \leq 1$.

4. Проведены испытания нового автомата по упаковке товара. При взвешивании 16 упаковок было получено выборочное среднее квадратическое отклонение 2,5 г. Построить доверительный интервал для среднего квадратического отклонения с надежностью 95%.

5. В прошлом году средняя заработная плата жителей города составляла 370 у.е. В этом году выборка из 400 человек показала, что средняя выборочная заработная плата равна 380 у.е. при выборочном среднем квадратическом отклонении 25 у.е. Можно ли на 5%-ном уровне значимости утверждать, что средняя заработная плата жителей города увеличилась?

Контрольная работа № 1 по математической статистике.

Точечные оценки (специальность «Экономика») -

Вариант 1

1. Бета- и гамма-функции: определения и свойства. Найти $\Gamma(11/2)$ и $\Gamma(9/2, 5/2)$.

2. Вывести формулу, связывающую третий центральный момент с начальными моментами. Построить соответствующую формулу для выборочного момента и доказать его состоятельность.

3. Доказать эффективность оценки методом максимального правдоподобия для математического ожидания в случае нормального распределения.

4. Найти оценку методом моментов для параметра β распределения $F(x) = (x/5)^\beta$, $0 \leq x \leq 5$.

5. Трое исследователей провели независимые выборочные обследования доходов населения. Первый определил средний годовой доход 3500 у.е. со средним квадратическим отклонением $\sigma_1 = 120$ у.е., второй — 3100 у.е. с $\sigma_2 = 60$ у.е., третий — 3200 у.е. с $\sigma_3 = 90$ у.е. Построить наиболее точную оценку среднего годового дохода и найти ее среднее квадратическое отклонение.

Вариант 2

1. Распределения хи-квадрат, Стьюдента и Фишера: определения, свойства, характеристики, графики. Определить, какое распределение имеет случайная величина ξ^2 , если ξ имеет распределение Стьюдента с n степенями свободы.

2. Доказать несмещенность и состоятельность исправленной выборочной дисперсии как оценки теоретической дисперсии.

3. Доказать эффективность оценки максимального правдоподобия параметра θ показательного распределения $p(x) = (1/\theta)e^{-x/\theta}$, $x \geq 0$.

4. Найти оценки методом моментов для параметров a и b равномерного распределения на отрезке $[a, b]$.

5. В поселке Семеркино все жители имеют доход не менее 70 тыс. руб в месяц. Выборочное исследование дало средний доход 210 тыс. руб. В предположении, что случайная величина дохода имеет распределение Парето $F(x) = 1 - (x/x_0)^{-\alpha}$, $x \geq x_0$, где $x_0 = 70$ (тыс. руб.), оценить параметр α методом моментов и долю жителей с доходами свыше 280 тыс. руб.

Вариант 3

1. Генеральная совокупность. Выборка и вариационный ряд. Эмпирическая функция распределения и ее свойства. Теорема Гливенко—Кантелли. Доказать несмещенность и состоятельность эмпирической функции распределения как оценки теоретической функции распределения (в любой точке).

2. Найти дисперсии исправленной и неисправленной выборочных дисперсий в случае нормального распределения.

3. Для распределения $F(x) = (x/\theta)^2$, $0 \leq x \leq \theta$, построить несмещенную оценку параметра θ на основе $x_{\max}$ и доказать ее сверхэффективность.

4. Найти оценку методом моментов для числа степеней свободы r распределения Стьюдента. Определить, при каких r это возможно.

5. Ежедневный спрос на некоторый товар имеет равномерное распределение на отрезке $[a, b]$. За период наблюдения спрос составлял в среднем 128 кг с исправленной выборочной дисперсией 363 кг². С помощью метода моментов оценить, сколько потребуется товара, чтобы удовлетворить ежедневный спрос с вероятностью 90%.

Вариант 4

1. Выборочные характеристики и их свойства. Точечные оценки параметров. Несмещенность, состоятельность, эффективность. Теорема Слуцкого (с доказательством). Асимптотическая нормальность оценок.

2. Найти функции распределения и плотности $x_{\min}$ и $x_{\max}$ для выборки $x_1, \dots, x_n$ из генеральной совокупности с функцией распределения $F(x)$ и плотностью $p(x)$.

3. Доказать эффективность оценки методом моментов для вероятности p распределения Бернулли с числом испытаний N .

4. Найти оценку максимального правдоподобия для параметра θ распределения Вейбулла $F(x) = 1 - e^{-\theta x^3}$, $x > 0$, $\theta > 0$.

5. Известно, что некоторая работа занимает время, состоящее из постоянного периода и случайной задержки, распределенной показательно. Хронометраж рабочего времени показал, что работа занимает в среднем 50 мин. со средним квадратическим отклонением 20 мин. С помощью метода моментов оценить время, за которое работа будет выполнена с вероятностью 95%.

Вариант 5

1. Бета- и гамма-распределения: определения, свойства, характеристики, графики. Найти математическое ожидание гамма-распределения.

2. Даны две выборки $x_1, \dots, x_n$ и $y_1, \dots, y_m$ из общего нормального распределения $N(a, \sigma^2)$. Построить оценку дисперсии, эффективную в классе линейных несмещенных оценок, по исправленным выборочным дисперсиям σ_x^2, σ_y^2 .

3. Доказать эффективность оценки методом максимального правдоподобия для параметра λ распределения Пуассона.

4. Найти оценку методом моментов для параметра α распределения Парето $F(x) = 1 - (x/2)^{-\alpha}$, $x \geq 2$.

5. Известно, что некоторая работа занимает время, состоящее из постоянного периода и случайной задержки, распределенной показательно. Хронометраж рабочего времени показал, что работа занимает как минимум 30 мин., а в среднем — 40 мин. С помощью метода максимального правдоподобия оценить вероятность того, что работа будет выполнена за час.

Вариант 6

1. Леммы и теорема Фишера. Следствия из теоремы (с доказательствами).

2. Построить оценку математического ожидания, эффективную в классе линейных несмещенных оценок, в случае неравнооточных наблюдений. Вычислить ее дисперсию.

3. Доказать асимптотическую эффективность исправленной выборочной дисперсии как оценки теоретической дисперсии в случае нормального распределения.

4. Найти оценку максимального правдоподобия для параметра θ равномерного распределения на отрезке $[\theta, 3\theta]$, $\theta > 0$.

5. Известно, что доля возвратов по кредитам банка имеет распределение $F(x) = x^\theta$, $0 \leq x \leq 1$. Наблюдения показали, что в среднем она составляет 80%. С помощью метода моментов оценить параметр θ и вероятность того, что доля возвратов опустится ниже 70%.

Контрольная работа № 2 по математической статистике.

Доверительные интервалы. Статистическая гипотеза
(специальность «Экономика»)

Вариант 1

1. Основные понятия в проверке гипотез: основная и альтернативная гипотезы, статистика критерия и критическая область, ошибки первого и второго рода, уровень значимости и мощность критерия.

2. Построить асимптотический доверительный интервал для параметра θ нормального распределения $N(\theta, \theta^6)$, $\theta > 0$.

3. Фирма разослала 1000 рекламных каталогов и получила 180 заказов. Построить доверительный интервал для эффективности рекламы (вероятности заказа по каталогу) с надежностью 95%.

4. В прошлом году средняя заработная плата жителей города Мушкино составляла 8500 руб. В этом году выборка из 300 человек показала, что средняя заработная плата составляет 8560 руб. при выборочном среднем квадратическом отклонении 120 руб. Можно ли на 5%-ном уровне значимости утверждать, что заработная плата жителей города увеличилась?

5. За последние 10 лет выборочная дисперсия годовой доходности актива А составила 0,08, актива Б — 0,14. Можно ли утверждать на уровне значимости 5%, что вложения в актив Б более рискованны, чем в А (т.е. дисперсия там больше)?

Вариант 2

1. Критерий отношения правдоподобия. Теорема Неймана—Пирсона. Применение к гипотезе о математическом ожидании для нормального распределения.

2. Построить асимптотический доверительный интервал для неизвестного параметра α при известном λ , если случайная величина имеет гамма-распределение.

3. По данным социологического опроса среди 100 человек, 18 из них пользовались моющим средством «Чистомой». Оценить, сколько всего человек надо опросить, чтобы с вероятностью 95% определить долю пользующихся моющим средством с точностью до 2%.

4. В книжном магазине проведено исследование продаж нового романа писателя Бородачева. В течение 25 рабочих дней роман продавался ежедневно в среднем по 55 экземпляров с выборочным средним квадратическим отклонением 12 экземпляров. Можно ли утверждать на уровне значимости 10%, что этот роман расходуется лучше, чем предыдущий роман автора, если тот продавался в среднем по 48 экземпляров в день?

5. В течение 100 рабочих дней в фирму А обращалось в среднем 89 клиентов в день, в фирму Б — 96. Есть ли основания утверждать на уровне значимости 5%, что фирма Б более популярна, чем фирма А, если числа клиентов в день для фирм А и Б имеют известные дисперсии 120 и 136 соответственно?

Вариант 3

1. Проверка гипотез о математическом ожидании и дисперсии для нормальной выборки.

2. Построить асимптотический доверительный интервал для неизвестного параметра λ при известном α , если случайная величина имеет гамма-распределение.

3. Проведены испытания новой машины, упаковывающей товар в пакеты. По выборке из 30 пакетов получено выборочное среднее квадратическое отклонение веса 0,6 г. Построить доверительный интервал с надежностью 95% для дисперсии веса пакета при упаковке машиной.

4. Фирма разослала 500 рекламных каталогов и получила 75 заказов. Можно ли утверждать на уровне значимости 5%, что эффективность рекламы возросла, если раньше она составляла 13%?

5. В городах Бусино и Гусино проведены выборочные обследования доходов жителей. По выборкам из 250 человек получено, что в Бусино средний доход 8020 руб. с выборочным средним квадратическим отклонением 195 руб., в Гусино средний доход 7960 руб. с выборочным средним квадратическим отклонением 140 руб. Можно ли утверждать на уровне значимости 5%, что в Бусино живут в среднем богаче, чем в Гусино?

Вариант 4

1. Понятие доверительного интервала. Точные доверительные интервалы для параметров нормального распределения.

2. Построить асимптотический доверительный интервал для параметра θ равномерного распределения на отрезке $[0, \theta]$ (использовать оценку методом моментов).

3. Импортёр упаковывает чай в пакеты. Известно, что наполняющая машина работает со средним квадратическим отклонением, равным 0,7 г. Выборка 16 пакетов показала средний вес 100,2 г. Найти доверительный интервал для среднего веса пакета с вероятностью 95%.

4. Средний дневной объем продаж фирмы составлял 1200 единиц. После реорганизации за 25 рабочих дней выборочный средний объем продаж составил 1270 единиц с выборочным средним квадратическим отклонением 30 единиц. Можно ли утверждать на уровне значимости 5%, что реорганизация привела к увеличению среднего объема продаж фирмы?

5. Для оценки качества изделий, изготовленных двумя заводами, взяты выборки из 100 изделий первого и 200 изделий второго. В них оказалось по 18 и 24 бракованных изделия соответственно. Можно ли утверждать на уровне значимости 10%, что вероятность изготовления брака на обоих заводах одинакова?

Вариант 5

1. Проверка гипотез о математических ожиданиях и дисперсиях для двух нормальных выборок.

2. Построить асимптотический доверительный интервал для параметра λ показательного распределения.

3. В деревне Валенково проведено выборочное исследование доходов. По выборке из 25 человек получен выборочный средний доход 4760 руб. с выборочным средним квадратическим отклонением 230 руб. Построить доверительный интервал для среднего дохода с надежностью 95%.

4. Среднее время сборки изделия было 60 мин. При испытаниях нового метода сборки по выборке из 16 изделий получено выборочное среднее время сборки 53 мин. с выборочным средним квадратическим отклонением 9 мин. Можно ли утверждать на уровне значимости 5%, что среднее время сборки сократилось?

5. Проведено исследование батареек фирм «Винтик» и «Шпунтик». По выборке из 11 батареек фирмы «Винтик» получен средний срок службы 207 ч с выборочным средним квадратическим отклонением 9 ч, по выборке из 16 батареек фирмы «Шпунтик» — средний срок службы 195 ч с выборочным средним квадратическим отклонением 8 ч. Можно ли утверждать на уровне значимости 10%, что в среднем срок службы батареек этих двух фирм одинаков?

Вариант 6

1. Понятие асимптотического доверительного интервала. Методы построения асимптотических доверительных интервалов.

2. Построить асимптотический доверительный интервал для параметра λ распределения Пуассона.

3. Среди 400 изделий в течение года вышли из строя 50. Построить доверительный интервал с уровнем надежности 95% для вероятности того, что изделие прослужит не менее года.

4. Старая наполняющая машина работала со средним квадратическим отклонением веса упаковки 1,1 г. Были проведены испытания новой машины. По выборке из 25 упаковок товара получено выборочное среднее квадратическое отклонение 0,4 г. Можно ли утверждать на уровне значимости 5%, что для новой машины среднее квадратическое отклонение меньше?

5. Фирма разослала 1000 рекламных каталогов и получила 180 заказов. Можно ли утверждать на уровне значимости 5%, что эффективность рекламы повысилась, если в прошлом году фирма получила 160 заказов (при том же объеме рассылки)?

4. Примеры вариантов экзаменационных работ по математической статистике (специальность «Экономика»)

Вариант 1

1. Бета-распределение: определение, свойства, характеристики, график. Вычислить математическое ожидание.

2. Критерий отношения правдоподобия. Теорема Неймана—Пирсона. Применение к гипотезе о среднем для нормального распределения.

3. Доказать эффективность оценки методом максимального правдоподобия для математического ожидания в случае нормального распределения.

4. В выборке из 400 изделий средний срок службы составил 1500 ч. Известно, что срок службы изделия имеет показательное распределение. Построить доверительный интервал для среднего срока службы с надежностью 95%.

5. Средний дневной объем продаж фирмы составлял 860 единиц. После реорганизации за 25 рабочих дней выборочный средний объем продаж составил 900 единиц с выборочным средним квадратическим отклонением 35 единиц. Можно ли утверждать на уровне значимости 5%, что реорганизация привела к увеличению среднего объема продаж?

Вариант 2

1. Гамма-распределение: определение, свойства, характеристики, график. Вычислить математическое ожидание и дисперсию.

2. Понятие доверительного интервала. Точные доверительные интервалы для параметров нормального распределения.

3. Доказать эффективность оценки максимального правдоподобия параметра θ показательного распределения $p(x) = (1/\theta)e^{-x/\theta}$, $x > 0$.

4. Проведены испытания новой машины, упаковывающей товар в пакеты. По выборке из 30 пакетов получено выборочное среднее квадратическое отклонение веса 0,7 г. Построить доверительный интервал с надежностью 95% для дисперсии веса пакета при упаковке машиной.

5. В прошлом году средняя заработная плата жителей города Булкино составляла 7300 руб. В этом году выборка из 400 человек показала, что средняя заработная плата составляет 7320 руб. при выборочном среднем квадратическом отклонении

150 руб. Можно ли на уровне значимости 5% утверждать, что заработная плата жителей города увеличилась?

Вариант 3

1. Распределение Стьюдента: определение, свойства, характеристики, график. Доказать сходимость при $n \rightarrow \infty$ к нормальному распределению $N(0, 1)$.

2. Понятие асимптотического доверительного интервала. Методы построения асимптотических доверительных интервалов. Примеры построения интервалов для вероятности «успеха» в испытаниях Бернулли (двумя методами).

3. Доказать несмещенность, состоятельность и эффективность эмпирической функции распределения как оценки теоретической функции распределения.

4. Импортёр упаковывает чай в пакеты. Известно, что наполняющая машина работает со средним квадратическим отклонением, равным 0,8 г. Выборка 100 пакетов показала средний вес 99,9 г. Найти доверительный интервал для среднего веса пакета с надежностью 90%.

5. Проведено исследование батареек фирм «Нептун» и «Плутон». По выборке из 10 батареек фирмы «Нептун» получен средний срок службы 207 ч с выборочным средним квадратическим отклонением 11 ч, по выборке из 16 батареек фирмы «Плутон» — средний срок службы 194 ч с выборочным средним квадратическим отклонением 8 ч. Можно ли утверждать на уровне значимости 10%, что в среднем срок службы батареек этих двух фирм одинаков?

Вариант 4

1. Распределение Фишера: определение, свойства, характеристики, график. Проверить, к какому распределению оно сходится при $m \rightarrow \infty$.

2. Основные понятия в проверке гипотез: основная и альтернативная гипотезы, статистика критерия и критическая область, ошибки первого и второго рода, уровень значимости и мощность критерия.

3. Доказать эффективность оценки максимального правдоподобия для вероятности «успеха» p распределения Бернулли с числом испытаний m .

4. За 100 рабочих дней в магазин в среднем обращалось 289 покупателей в день. Известно, что число покупателей в

день описывается распределением Пуассона. Построить доверительный интервал для среднего числа покупателей в день с надежностью 90%.

5. Старая наполняющая машина работала со средним квадратическим отклонением веса упаковки 1 г. Были проведены испытания новой машины. По выборке из 30 упаковок товара получено выборочное среднее квадратическое отклонение 0,7 г. Можно ли утверждать на уровне значимости 5%, что для новой машины среднее квадратическое отклонение меньше?

Вариант 5

1. Распределение хи-квадрат: определение, свойства, характеристики, график. Доказать, что $\chi_n^2 = \gamma(n/2, 1/2)$.

2. Метод наименьших квадратов. Система нормальных уравнений. Теорема Гаусса—Маркова.

3. Доказать эффективность оценки методом моментов для параметра λ распределения Пуассона.

4. В деревне Ложкино проведено выборочное исследование доходов. По выборке из 25 человек получен выборочный средний доход 4750 руб. с выборочным средним квадратическим отклонением 150 руб. Построить доверительный интервал для среднего дохода с надежностью 95%.

5. Фирма разослала 1000 рекламных каталогов и получила 140 заказов. Можно ли утверждать на уровне значимости 5%, что эффективность рекламы повысилась, если в прошлом году фирма получила 100 заказов (при том же объеме рассылки)?

Вариант 6

1. Бета- и гамма-функции: определения и свойства. Найти $\Gamma(11/2)$ и $\Gamma(9/2, 5/2)$.

2. Критерии согласия Пирсона (хи-квадрат) и Колмогорова.

3. Построить оценку математического ожидания, эффективную в классе линейных несмещенных оценок, в случае неравноточных наблюдений. Вычислить ее дисперсию.

4. Фирма разослала 1000 рекламных каталогов и получила 140 заказов. Построить доверительный интервал для эффективности рекламы (вероятности заказа по каталогу) с надежностью 90%.

5. В городах Глашино и Дашино проведены выборочные обследования доходов жителей. По выборкам из 100 человек получено, что в Глашино средний доход 8020 руб. с выборочным средним квадратическим отклонением 190 руб., в Дашино сред-

ний доход 7960 руб. с выборочным средним квадратическим отклонением 160 руб. Можно ли утверждать на уровне значимости 5%, что в Глашино живут в среднем богаче, чем в Дашино?

**5. Примеры вариантов экзаменационных работ
по теории вероятностей и математической статистике
(специальность «Менеджмент»)¹**

Вариант 1. Теоретическая часть.

1. Формула полной вероятности имеет вид:

а) $P(AB) = P(A|B)P(B)$;

б) $P(A) = \sum_{k=1}^n P(H_k | A)P(H_k)$;

в) $P(A) = \sum_{k=1}^n P(A|H_k)P(H_k)$;

г) $P(H_k | A) = P(A|H_k)P(H_k) / P(A)$.

2. Свойством испытаний Бернулли является следующее:

- а) все исходы испытаний равновероятны;
- б) испытания заканчиваются одним из двух исходов;
- в) вероятность успеха определяется результатом одного произвольного испытания;
- г) все приведенные выше ответы верны.

3. Размещением m элементов из n называется:

- а) упорядоченный набор n элементов из множества, содержащего m элементов;
- б) упорядоченный набор m элементов из множества, содержащего n элементов;
- в) неупорядоченный набор m элементов из множества, содержащего n элементов;
- г) операция перемены мест m элементов во множестве, содержащем n элементов.

4. Случайная величина называется дискретной, если она:

- а) зависит от случая;
- б) принимает конечное или счетное число значений;

¹ Каждый вариант состоит из теоретического теста и вычислительных задач.

- в) равна числу успехов в схеме Бернулли;
- г) задается своей функцией распределения.

5. Какое свойство НЕ является обязательным для функции распределения:

- а) $F(x)$ не убывает;
- б) $\lim_{x \rightarrow +\infty} F(x) = 1$;
- в) $P(a \leq \xi < b) = F(b) - F(a)$;
- г) $F(x)$ непрерывна?

6. Случайная величина распределена по показательному закону, если:

- а) ее закон распределения имеет вид $P(\xi = k) = \lambda^k e^{-\lambda} / k!$;
- б) ее плотность $p(x) = (1/\sqrt{2\pi})e^{-x^2/2}$;
- в) ее плотность $p(x) = \lambda e^{-\lambda x}, x \geq 0$;
- г) ее функция распределения $F(x) = e^{-\lambda x}$ при $x \geq 0$.

7. Дисперсию дискретной случайной величины можно найти по формуле:

- а) $D\xi = M(\xi - M\xi)$;
- б) $D = \text{cov}(\xi, M\xi)$;
- в) $D\xi = M(\xi - M\xi)^2$;
- г) ни один из ответов не верен.

8. Метод моментов для нахождения оценок заключается в:

- а) поиске распределения случайной величины;
- б) приравнивании выборочных моментов к теоретическим;
- в) максимизации функции моментов;
- г) вычислении s^2 .

9. Доверительный интервал для параметра — это интервал:

- а) в который параметр попадает с максимальной вероятностью;
- б) в котором параметр лежит с заданной вероятностью;
- в) в котором лежат все возможные значения параметра;
- г) в котором выборочное среднее лежит с заданной вероятностью.

10. Распределение Фишера имеет:

- а) один параметр – число степеней свободы;
- б) один параметр – уровень значимости;
- в) два параметра – два числа степеней свободы;
- г) два параметра – уровень значимости и число степеней свободы.

Задачи

1. В очередь в булочную случайным образом встали 5 женщин и 2 подростка. Какова вероятность того, что два подростка стоят в очереди рядом?

2. Фирма нарушает закон с вероятностью 0,25. Аудитор обнаруживает нарушение с вероятностью 0,85. Проведенная им проверка не выявила нарушений. Найти вероятность того, что на самом деле они есть.

3. Плотность распределения случайной величины ξ имеет вид

$$p(x) = \begin{cases} 0, & x \notin [1, 4]; \\ C(x-1), & x \in [1, 4]. \end{cases}$$

Найти C , $F(x)$, $M\xi$, $P(\xi > 3)$.

4. Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	-1	0	2
1	0,1	0,1	0,2
2	0,2	0,3	0,1

Зависимы ли эти величины? Вычислить $P(\xi + \eta < 2)$ и $\text{cov}(\xi, \eta)$.

5. По выборке из 25 упаковок товара средний вес составил 105 г при выборочном среднем квадратическом отклонении (исправленном) 5 г. На уровне значимости 10% проверьте гипотезу о том, что средний вес упаковок товара в генеральной совокупности равен 100 г.

Вариант 2. Теоретическая часть

1. Теорема умножения вероятностей имеет вид:

- а) $P(AB) = P(A|B)P(B)$;
- б) $P(AB) = P(A)P(B)$;
- в) $P(AB) = P(B|A)P(B)$;
- г) $P(A)P(B) = P(A|B)P(B|A)$.

2. События A и B независимы, если:

- а) $P(AB) = 0$;
- б) $P(A|B) = P(B)$;
- в) $P(AB) = P(A)P(B)$;
- г) A и B не пересекаются.

3. Каково классическое определение вероятности:

- а) $P(A) = N(A) / N$;
- б) $P(A)$ равна 1, если A произошло, и 0, если A не произошло;
- в) $P(\xi = m) = C_n^m p^m q^{n-m}$;
- г) $P(\xi = m) = pq^{m-1}$?

4. Случайная величина распределена по Пуассону, если:

- а) ее закон распределения имеет вид $P(\xi = k) = \lambda^k e^{-\lambda}$;
- б) ее плотность $p(x) = \lambda e^{-\lambda x}$, $x \geq 0$;
- в) ее закон распределения имеет вид $P(\xi = k) = p^k q^{n-k}$;
- г) ее закон распределения имеет вид $P(\xi = k) = \lambda^k e^{-\lambda} / k!$.

5. Приближенная формула Пуассона используется в случае, когда:

- а) n велико, $np < 10$;
- б) n велико, $np > 10$;
- в) n любое, $np > 10$;
- г) недостаточно данных для точного подсчета.

6. Математическое ожидание произвольной дискретной случайной величины определяется по формуле:

- а) $\int_{-\infty}^{+\infty} x p(x) dx$;
- б) $\sum_k p_k$;
- в) $\sum_k x_k p_k$;
- г) $\frac{1}{n}(x_1 + \dots + x_n)$.

7. Какое из перечисленных ниже свойств НЕ является свойством функции плотности $p(x)$:

а) $P(\xi < x) = \int_{-\infty}^x p(x) dx$;

б) $\lim_{x \rightarrow +\infty} p(x) = 1$;

в) $p(x) \geq 0$;

г) $\int_{-\infty}^{+\infty} p(x) dx = 1$?

8. Распределение Стьюдента имеет случайная величина:

а) $\xi_1^2 + \dots + \xi_n^2$, где $\xi_1, \dots, \xi_n \in N(0, 1)$ и независимы;

б) $\sqrt{(\xi_1^2 + \dots + \xi_n^2) / n}$ где $\xi_1, \dots, \xi_n \in N(0, 1)$ и независимы;

в) $\xi_0 / \sqrt{(\xi_0^2 + \dots + \xi_n^2) / n}$ где $\xi_0, \dots, \xi_n \in N(0, 1)$ и независимы;

г) $\xi_0 / \sqrt{(\xi_1^2 + \dots + \xi_n^2) / n}$ где $\xi_0, \dots, \xi_n \in N(0, 1)$ и независимы.

9. Ошибкой 1-го рода при проверке статистических гипотез называется ошибка, при которой:

а) отвергается неверная гипотеза H_0 ;

б) отвергается правильная гипотеза H_0 ;

в) отвергается правильная альтернативная гипотеза H_1 ;

г) вероятность отклонения H_0 становится меньше уровня значимости.

10. Несмещенной оценкой для неизвестной дисперсии является:

а) $\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2$;

б) $\frac{1}{n} (x_1 + \dots + x_n)$;

в) $\frac{1}{n-1} \sum_{k=1}^n (x_k^2 - \bar{x}^2)$;

г) $\frac{1}{n-1} \sum_{k=1}^n (\bar{x} - x_k)^2$.

Задачи

1. Пять рукописей случайным образом раскладывают в 4 папки. Какова вероятность того, что ровно одна папка останется пустой.

2. Два аудитора проверяют 4 фирмы (по 2 каждый), у двух из которых имеются нарушения. Вероятность обнаружения нарушений первым аудитором равна 0,7, вторым – 0,9. Найти вероятность, что все фирмы-нарушители будут выявлены.

3. Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	0	1	2
-1	1/9	2/9	1/18
1	1/9	1/3	1/16

Зависимы ли эти величины? Вычислить $P(\xi + \eta \geq 2)$ и $\text{cov}(\xi, \eta)$.

4. Плотность распределения случайной величины ξ имеет вид

$$p(x) = \begin{cases} 0, & x \notin [-1, 1]; \\ C(1+x), & x \in [-1, 1]. \end{cases}$$

Найти C , $F(x)$, $M\xi$, $P(\xi > 0)$.

5. Средний вес таблеток сильного действия должен быть 0,50 мг. Выборочная проверка 121 таблетки показала, что средний вес таблетки равен 0,53 мг. На уровне значимости 0,01 проверьте гипотезу о том, что выборочные данные согласуются с нормативом. Считать, что вес таблеток распределен по нормальному закону со средним квадратическим отклонением 0,11 мг.

Вариант 3. Теоретическая часть

1. Формула Байеса имеет вид:

а) $P(AB) = P(A|B)P(B)$;

б) $P(A) = \sum_{k=1}^n P(A|H_k)P(H_k)$;

в) $P(H_k|A) = P(A|H_k)P(H_k)/P(A)$;

г) $P(H_k|A) = P(A|H_k)P(H_k) / \sum_{j=1}^n P(H_j|A)P(H_j)$.

2. Сочетанием m элементов из n называется:

- а) упорядоченный набор n элементов из множества, содержащего m элементов;
- б) неупорядоченный набор n элементов из множества, содержащего m элементов;
- в) упорядоченный набор m элементов из множества, содержащего n элементов;
- г) неупорядоченный набор m элементов из множества, содержащего n элементов.

3. Вероятность того, что в n испытаниях Бернулли произойдет m успехов, равна:

- а) pq^{m-1} ;
- б) $p^n q^m$;
- в) $\frac{m!}{n!} p^m q^{n-m}$;
- г) $\frac{n!}{m!(n-m)!} p^m q^{n-m}$.

4. Случайная величина является непрерывной, если:

- а) она непрерывно изменяется;
- б) она принимает бесконечно много значений;
- в) ее плотность равна интегралу от вероятности;
- г) ее функция распределения равна интегралу от плотности.

5. Какое свойство НЕ всегда справедливо для дисперсии:

- а) $Dc = 0$;
- б) $D(\xi + \eta) = D\xi + D\eta$;
- в) $D\xi = M\xi^2 - (M\xi)^2$;
- г) $D\xi = \text{cov}(\xi, \xi)$?

6. Случайная величина распределена равномерно, если

- а) ее закон распределения имеет вид $P(\xi = k) = pq^{k-1}$;
- б) ее значения равномерно распределены по прямой;
- в) ее плотность $p(x) = 1/(b-a)$, $x \in [a, b]$;
- г) ее закон распределения имеет вид $P(\xi = x) = 1/(b-a)$, $x \in [a, b]$.

7. Какое из перечисленных ниже свойств справедливо для дискретной случайной величины ξ , принимающей значения $x_1 \leq \dots \leq x_n$:

а) $D\xi = \sum_{k=1}^n x_k^2 P(\xi = x_k)$;

б) $P(\xi = x_k) \leq P(\xi = x_{k+1})$;

в) $\sum_{k=1}^n P(\xi = x_k) = 1$;

г) все ответы верны?

8. Метод максимального правдоподобия заключается в:

а) приравнении функции правдоподобия к 0;

б) нахождении оценки, ближайшей к параметру;

в) максимизации функции правдоподобия;

г) поиске наивероятнейшего значения функции правдоподобия.

9. Распределение Фишера имеет случайная величина:

а) $\xi_1^2 + \dots + \xi_n^2$, где $\xi_1, \dots, \xi_n \in N(0, 1)$ и независимы;

б) χ_m^2 / χ_n^2 , где χ_m^2, χ_n^2 независимы и имеют χ^2 -распределение;

в) $\frac{\chi_m^2 / m}{\chi_n^2 / n}$, где χ_m^2, χ_n^2 независимы и имеют χ^2 -распределение;

г) $\frac{1}{\sqrt{n}} \chi_n^2$, где χ_n^2 имеет χ^2 -распределение.

10. Доверительный интервал для параметра a нормального распределения при известной дисперсии имеет вид:

а) $\bar{x} - \sigma u_\gamma / n < a < \bar{x} + \sigma u_\gamma / n$;

б) $\bar{x} - \sigma u_\gamma / \sqrt{n} < a < \bar{x} + \sigma u_\gamma / \sqrt{n}$;

в) $\bar{x} - st_\gamma / \sqrt{n} < a < \bar{x} + st_\gamma / \sqrt{n}$;

г) $(\bar{x} - st_\gamma) / \sqrt{n} < a < (\bar{x} + st_\gamma) / \sqrt{n}$.

Задачи

1. Шесть пассажиров садятся в поезд, случайно выбирая один из шести вагонов. Какова вероятность того, что ровно в один вагон никто не сядет?

2. Изделие имеет скрытые дефекты с вероятностью 0,15. В течение года выходит из строя 70% изделий со скрытыми дефектами и 10% изделий без дефектов. Найти вероятность того, что изделие имело скрытые дефекты, если оно вышло из строя в течение года.

3. Плотность распределения случайной величины ξ имеет вид $p(x) = \begin{cases} 0, & x \notin [-2, 1]; \\ C(1-x), & x \in [-2, 1]. \end{cases}$

Найти C , $F(x)$, $M\xi$, $P(0 < \xi < 3)$.

4. Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	-1	1	2
0	1/8	1/6	1/12
1	1/12	1/6	3/8

Зависимы ли они? Вычислить $P(\xi + \eta < 2)$ и $\text{cov}(\xi, \eta)$.

5. Взвесив 50 пакетов товара, нашли выборочное среднее 502 г и выборочное среднее квадратическое отклонение 10 г. На уровне значимости 5% проверьте гипотезу о том, что вес пакета равен 500 г.

Вариант 4. Теоретическая часть

1. Формула для числа сочетаний (без повторений) m элементов из n имеет вид:

а) $\frac{n!}{m!}$;

б) $\frac{n!}{m!(n-m)!}$;

в) $\frac{n!}{(n-m)!}$;

г) ни один из приведенных выше ответов не верен.

2. Теорема сложения вероятностей имеет вид:

а) $P(A+B) = P(A) + P(B)$;

б) $P(A+B) = P(A) + P(B) + P(AB)$;

в) $P(A+B) = P(A) + P(B) - P(AB)$;

г) $P(A+B) = P(A|B)P(B)$.

3. События A и B называются несовместными, если:

а) $P(AB) = P(A)P(B)$;

б) $P(AB) = 1$;

в) $A \cup B = \emptyset$;

г) $A \cap B = \emptyset$.

4. Наиболее вероятное число успехов m^* в схеме Бернулли определяется по формуле:

а) $np - p \leq m^* \leq np + p$;

б) $np - q \leq m^* \leq np + p$;

в) $np - q \leq m^* \leq np + q$;

г) $np - p \leq m^* \leq np + q$.

5. Математическое ожидание непрерывной случайной величины находится по формуле:

а) $\int_{-\infty}^{\infty} p(x) dx$;

б) $\int_{-\infty}^{\infty} x p(x) dx$;

в) $\sum_k p_k$;

г) $\sum_k x_k p_k$.

6. Какое из свойств НЕ является свойством ковариации:

а) $\text{cov}(c\xi, \eta) = \text{cov}(c\eta, \xi)$;

б) $\text{cov}(\xi + \eta, \xi) = \text{cov}(\xi, \eta) + \text{cov}(\xi, \xi)$;

в) $\text{cov}(\xi, \eta) = 0$ для независимых ξ, η ;

г) $\text{cov}(c\xi, \xi) = c^2 \text{cov}(\xi, \xi)$?

7. Случайная величина распределена по Бернулли, если:

а) ее закон распределения имеет вид $P(\xi = m) = C_n^m p^m q^{n-m}$;

б) ее закон распределения имеет вид $P(\xi = m) = p^m q^{n-m}$;

в) ее закон распределения имеет вид $P(\xi = m) = pq^{m-1}$;

г) ее плотность $p(m) = C_n^m p^m q^{n-m}$.

8. Оценка параметра называется несмещенной, если:
- ее отклонение от оцениваемого параметра равно 0;
 - ее математическое ожидание равно оцениваемому параметру;
 - она равна выборочному среднему;
 - ее дисперсия минимальна.
9. Ошибкой 2-го рода при проверке статистических гипотез называется ошибка, при которой:
- отвергается неверная гипотеза H_0 ;
 - отвергается правильная гипотеза H_0 ;
 - отвергается неверная альтернативная гипотеза H_1 ;
 - отвергается правильная альтернативная гипотеза H_1 .
10. Распределение χ^2 имеет случайная величина:
- равная квадрату нормальной случайной величины $N(a, \sigma^2)$;
 - равная квадрату суммы независимых нормальных случайных величин $N(0, 1)$;
 - равная сумме квадратов независимых нормальных случайных величин $N(0, 1)$;
 - ни один из приведенных ответов не верен.

Задачи

- Группа из 20 человек пишет контрольную работу из 4 вариантов (по пять человек в каждом). Найти вероятность того, что среди 5 случайно отобранных работ есть каждый вариант.
- Фирма участвует в трех проектах, каждый из которых может закончиться не удачей с вероятностью 0,2. В случае неудачи одного проекта вероятность разорения фирмы равна 20%, двух — 50%, трех — 90%. Определить вероятность разорения фирмы.
- Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	-1	0	2
1	0,2	0,3	0
2	0	0,1	C

Определить константу C. Исследовать вопрос о независимости ξ и η . Вычислить $\text{cov}(\xi, \eta)$.

4. Плотность распределения случайной величины ξ имеет вид $p(x) = \begin{cases} 0, & x \notin [0, 5]; \\ Cx, & x \in [0, 5]. \end{cases}$

Найти C , $F(x)$, $M\xi$, $P(2 < \xi < 3)$.

5. Средний доход фирмы в день составлял 1020 у.е. После реорганизации фирмы выборочный средний доход в день за 30 рабочих дней составил 1070 у.е. с выборочным стандартным отклонением 90 у.е. Можно ли утверждать (на уровне значимости 5%), что реорганизация фирмы привела к увеличению среднего дохода?

Вариант 5. Теоретическая часть

1. В определение классического вероятностного пространства входит условие:

- а) исходы равновероятны;
- б) исходы независимы;
- в) испытания заканчиваются одним из двух исходов;
- г) все приведенные выше ответы верны.

2. Условная вероятность $P(A|B)$ события A при условии B равна:

- а) $P(A)/P(B)$;
- б) $P(B)/P(A)$;
- в) $P(AB)/P(A)$;
- г) $P(AB)/P(B)$.

3. Локальная теорема Муавра—Лапласа используется в случае, когда:

- а) n велико, $np < 10$;
- б) n велико, $np > 10$;
- в) n любое, $np < 10$;
- г) не хватает данных для точного подсчета.

4. Случайная величина — это:

- а) набор вероятностей, зависящих от случая;
- б) числовая функция, заданная на вероятностном пространстве.
- в) случайный исход эксперимента.
- г) набор случайных событий, отвечающих заданному условию.

5. Какое свойство НЕ выполняется для независимых случайных величин ξ, η :

- а) $\text{cov}(\xi, \eta)$;
- б) $D(\xi + \eta) = D\xi + D\eta$;
- в) $M(\xi\eta) = M\xi M\eta$;
- г) $D(\xi\eta) = D\xi D\eta$?

6. Пусть случайная величина ξ распределена по показательному закону. Какое из следующих свойств верно:

- а) при $x \geq 0$ $P(\xi = x) = \lambda e^{-\lambda x}$;
- б) при $x \geq 0$ ее функция распределения $F(x) = e^{-\lambda x}$;
- в) при $x \geq 0$ ее функция распределения $F(x) = e^{\lambda x}$;
- г) для любого x $P(\xi = x) = 0$?

7. Дисперсия непрерывной случайной величины равна:

- а) $D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 dx$;
- б) $D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 p(x) dx$;
- в) $D\xi = \int_{-\infty}^{+\infty} (x - \bar{x})^2 p(x) dx$;
- г) $D\xi = \frac{1}{n} \int_{-\infty}^{+\infty} (x - \bar{x})^2 dx$.

8. Оценкой для неизвестного параметра называется:

- а) истинное значение параметра;
- б) угаданное значение параметра;
- в) функция от выборки;
- г) число, отличающееся от параметра не более чем на заданную величину.

9. Доверительный интервал для параметра a нормального распределения при неизвестной дисперсии имеет вид:

- а) $\bar{x} - \sigma u_{\gamma} / n < a < \bar{x} + \sigma u_{\gamma} / n$;
- б) $\bar{x} - \sigma u_{\gamma} / \sqrt{n} < a < \bar{x} + \sigma u_{\gamma} / \sqrt{n}$;
- в) $\bar{x} - st_{\gamma} / \sqrt{n} < a < \bar{x} + st_{\gamma} / \sqrt{n}$;
- г) $(\bar{x} - st_{\gamma}) / \sqrt{n} < a < (\bar{x} + st_{\gamma}) / \sqrt{n}$.

10. Критической областью при проверке статистических гипотез называется:

а) область значений наблюдаемой статистики, в которой верна основная гипотеза H_0 ;

б) область значений наблюдаемой статистики, в которой верна альтернативная гипотеза H_1 ;

в) область значений наблюдаемой статистики, в которой отклоняется основная гипотеза H_0 ;

г) область значений наблюдаемой статистики, в которой отклоняется альтернативная гипотеза H_1 .

Задачи

1. В лифт семиэтажного дома на первом этаже входят пять человек. Какова вероятность того, что на первых трех этажах (начиная со второго) вышли два человека при условии, что все вышли на разных этажах?

2. Поставки продуктов в город производятся из трех областей в количествах 20, 45, 35%. Доля продукта высшего качества в поставках составляет 45, 40, 25%. Покупатель купил продукт высшего качества. Какова вероятность того, что он произведен в первой области?

3. Плотность распределения случайной величины ξ имеет вид $p(x) = \begin{cases} 0, & x \notin [0, 4]; \\ Cx^2, & x \in [0, 4]. \end{cases}$

Найти C , $F(x)$, $M\xi$, $P(1 < \xi < 7)$.

4. Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	-1	0	2
1	1/6	1/6	1/3
2	1/12	1/6	1/12

Исследовать вопрос о независимости ξ и η . Вычислить $P(\xi + \eta < 2)$ и $\text{cov}(\xi, \eta)$.

5. Яблоки упаковываются в коробки весом по 10 кг. Случайная выборка 9 коробок с яблоками выявила, что средний вес коробки равен 9,5 кг с выборочной дисперсией 0,04 кг². На

уровне значимости 5% проверить гипотезу о том, что средний вес коробки с яблоками равен 10 кг.

Вариант 6. Теоретическая часть

1. Число размещений (без повторений) m элементов из n равно:

- а) $n!/m!$;
- б) m/n ;
- в) $n!/(n-m)!$;
- г) $\frac{n!}{m!(n-m)!}$.

2. Какое из свойств условной вероятности НЕ верно:

- а) $P(A|B)P(B) = P(B|A)P(A)$;
- б) $P(A|\bar{B}) = 1 - P(A|B)$;
- в) $P(A|B) = 0$, если $A \cap B = \emptyset$;
- г) $P(A|B) = 1$, если $A \subset B$?

3. Приближенная формула Пуассона имеет вид:

- а) $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$, где $\lambda = np$;
- б) $P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}$, где $\lambda = mp$;
- в) $P_n(m) \approx \frac{n^m}{m!}$;
- г) $P_n(m) \approx \frac{n!}{m!(n-m)!}$.

4. Пусть событие A может произойти совместно с некоторыми событиями (гипотезами) H_i . Какое условие нужно для применения формулы полной вероятности:

- а) события H_i образуют полную группу событий;
- б) события H_i независимы;
- в) события H_i содержат в себе один из двух исходов;
- г) все приведенные выше ответы верны?

5. Пусть ξ — непрерывная случайная величина с плотностью $p(x)$ и функцией распределения $F(x)$. Какое из следующих свойств верно:

- а) $P(a < \xi < b) = F(b) - F(a)$;
- б) $P(a < \xi < b) = p(b) - p(a)$;
- в) $P(a < \xi < b) = p(a) - p(b)$;
- г) $P(a < \xi < b) = F(a) - F(b)$?

6. Случайная величина распределена по геометрическому закону, если:

- а) ее плотность $p(x) = p(1-p)^{x-1}$ при $x \geq 0$;
- б) ее закон распределения имеет вид $P(\xi = m) = (1-p)^{m-1}p$;
- в) ее закон распределения имеет вид $P(\xi = m) = (1-p)^{m-1}$;
- г) ее плотность $p(x) = \lambda e^{-\lambda x}$ при $x \geq 0$.

7. Дисперсия дискретной случайной величины равна:

- а) $D\xi = \sum_k (x_k - M\xi)^2$;
- б) $D\xi = \sum_k (x_k - M\xi)^2 p_k$;
- в) $D\xi = \frac{1}{n} \sum_k (x_k - \bar{x})^2$;
- г) $D\xi = \frac{1}{n-1} \sum_k (x_k - \bar{x})^2$.

8. Оценка называется состоятельной, если:

- а) среди всех оценок она наиболее точно описывает параметр;
- б) ее математическое ожидание равно оцениваемому параметру;
- в) с ростом числа наблюдений она сходится по вероятности к параметру;
- г) с ростом числа наблюдений она сходится к вероятности успеха.

9. Надежностью доверительного интервала называется:

- а) вероятность того, что оцениваемый параметр попадет в интервал;
- б) вероятность того, что оцениваемый параметр не попадет в интервал;
- в) длина доверительного интервала;
- г) половина длины доверительного интервала.

10. Пусть основная гипотеза $H_0: a = a_0$ отклоняется. В этом случае:

- а) делается вывод, что истинное значение параметра a отличается от a_0 ;
- б) нужно дополнительно посчитать выборочное среднее $\bar{x}$;
- в) мы совершаем ошибку 1-го рода;
- г) мы совершаем ошибку 2-го рода.

Задачи

1. Семь клиентов случайным образом обращаются в 5 фирм. Найти вероятность того, что хотя бы в одну фирму никто не обратится.

2. В ящике 20 теннисных мячей, в том числе 12 новых и 8 иггранных Из ящика извлекаются наугад два мяча для игры и после игры возвращаются в ящик. После этого из ящика вынимают два мяча для следующей игры. Найти вероятность того, что эти оба эти мяча будут новыми (т.е. неигранными).

3. Совместное распределение двух дискретных случайных величин ξ и η задано таблицей:

ξ	η		
	1	2	3
1	C	0,2	0
2	0,2	0,4	0,2

Найти C. Вычислить $\text{cov}(\xi, \eta)$ и $P(\xi + \eta > 3)$. Исследовать вопрос о независимости ξ и η .

4. Плотность распределения случайной величины ξ имеет вид

$$p(x) = \begin{cases} 0, & x \notin [0, 1]; \\ Cx^3, & x \in [0, 1]. \end{cases}$$

Найти C, $F(x)$, $M\xi$, $P(1/2 < \xi < 3/2)$.

5. Средний дневной объем продаж в магазине составлял 500 единиц. После реорганизации выборочный средний дневной объем продаж за 25 рабочих дней составил 520 единиц с выборочным средним квадратическим отклонением 40 единиц. Можно ли утверждать (на уровне значимости 5%), что реорганизация привела к увеличению среднего объема продаж?

6. Пример задания по теории вероятностей для самостоятельной работы

1. На фирме работают 10 сотрудников, из которых знают английский — 7, немецкий — 6, французский — 5, английский и немецкий — 3, английский и французский — 3, немецкий и французский — 4, все три языка — 2. Найти вероятность того, что команда из трех случайно выбранных сотрудников сможет осуществить перевод со всех трех языков.

2. На фирме работают 3 менеджера по продажам. Каждый из 10 клиентов обращается к любому из менеджеров равновероятно и независимо от других. Найти вероятность того, что у какого-либо из менеджеров окажется больше клиентов, чем у двух других вместе взятых, при условии, что к каждому менеджеру обратился хотя бы один клиент.

3. В урне имеется 5 шаров: 2 черных, 1 белый, 1 красный и 1 синий. Два игрока поочередно вынимают шары из урны (начиная с первого игрока). Игрок, вынувший белый шар, выигрывает. Если вынут красный шар, игра оканчивается ничьей. Если вынут черный шар, игра продолжается. Если вынут синий шар, все шары возвращаются в урну и игра начинается с начала. Найти вероятности выигрышей каждого из игроков и ничьей.

4. В каждой двенадцатой упаковке товара фирмы «Рекс» вложена одна из букв «Р», «Е», «К», «С» равновероятно. Покупатель хочет собрать все буквы, чтобы выиграть приз. Найти математическое ожидание и дисперсию числа упаковок, которые ему придется купить.

5. Случайные доходности по акциям компаний «Астра» и «Василек» имеют дисперсии 4 и 9 соответственно, а их коэффициент корреляции равен $1/2$. В каких долях следует разделить капитал при вложении в акции, чтобы минимизировать риск (дисперсию доходности)?

6. Случайное время изготовления каждой детали равномерно распределено от 4 до 8 мин. и не зависит от других. Изготовление скольких деталей за 25 рабочих часов можно гарантировать с вероятностью 95?

7. Случайные величины A , B , C независимы и равномерно распределены на отрезке $[0, 1]$. Найти вероятность того, что уравнение $Ax^2 + 2Bx + C = 0$ имеет решения.

8. Случайная величина ξ имеет показательное распределение с математическим ожиданием 3. Найти вероятность $P(\sin \xi < 1/2)$.

9. Случайные величины α и β независимы и имеют стандартное нормальное распределение. Найти распределение $\xi = \alpha/\beta$.

10. Построить пример двух зависимых случайных величин ξ и η , каждая из которых имеет равномерное распределение на отрезке $[0, 1]$, а их коэффициент корреляции равен любому заданному $\rho \in [-1, 1]$.

7. Пример задания по математической статистике для самостоятельной работы

1. Вычислить дисперсию и моду распределения Стьюдента.
2. Вычислить математическое ожидание и дисперсию распределения Фишера.

3. Вычислить математическое ожидание, моду и дисперсию бета-распределения.

4. Показать, что выборочные средние квадратические отклонения, полученные по исправленной и неисправленной дисперсии для нормального распределения, являются смещенными оценками. Вычислить их математические ожидания.

5. Показать, что оценка максимального правдоподобия для параметра λ показательного распределения является смещенной. Построить на ее основе несмещенную оценку.

6. Для оценивания параметра σ нормального распределения $N(0, \sigma^2)$ используют оценку вида $\sigma^* = \left(\frac{c}{n}\right) \sum_{i=1}^n |x_i|$. Найти константу c , при которой оценка окажется несмещенной. Вычислить эффективность оценки.

7. Для равномерного распределения на отрезке $[0, 3\theta]$ найти оценку максимального правдоподобия, вычислить ее математическое ожидание. Построить на основе этой оценки несмещенную оценку и доказать ее сверхэффективность.

8. Случайная величина принимает значения $-1, 0$ и 1 с вероятностями $(1 - \theta)/3; 1/3; (1 + \theta)/3$ соответственно, $\theta \in [-1; 1]$. Построить для параметра θ оценки методом моментов и методом максимального правдоподобия, а также асимптотический доверительный интервал (на основе метода моментов).

9. Для распределения Коши с плотностью $p(x) = \frac{1}{\pi(1 + (x - \theta)^2)}$

строится оценка максимального правдоподобия параметра θ по двум наблюдениям. Определить, при каких x_1 и x_2 эта оценка существует и единственна.

10. Проверяется нулевая гипотеза о том, что случайная величина равномерно распределена на отрезке $[-1; 1]$, против гипотезы, что она имеет распределение Лапласа. Построить критерий отношения правдоподобия при $n = 2$ с уровнем значимости $\alpha < 1/2$.

Приложение 2. Таблицы

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

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491

Окончание табл. 1

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
x	Десятые доли x									
	0	2	4	6	8					
4	0,0001338	0000589	0000249	0000101	0000040					
5	0,0000015									

Таблица 2. Значения функции Лапласа

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$	x	$\Phi_0(x)$
0,00	0,0000	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,01	0,0040	0,33	0,1293	0,65	0,2422	0,97	0,3340
0,02	0,0080	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,03	0,0120	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,04	0,0160	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,05	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,06	0,0239	0,38	0,1480	0,70	0,2580	1,02	0,3461
0,07	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,08	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,09	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,10	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,11	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,12	0,0478	0,44	0,1700	0,76	0,2764	1,08	0,3599
0,13	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,14	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643
0,15	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,16	0,0636	0,48	0,1844	0,80	0,2881	1,12	0,3686
0,17	0,0675	0,49	0,1879	0,81	0,2910	1,13	0,3708
0,18	0,0714	0,50	0,1915	0,82	0,2939	1,14	0,3729
0,19	0,0753	0,51	0,1950	0,83	0,2967	1,15	0,3749
0,20	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,3770
0,21	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,3790
0,22	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,3810
0,23	0,0910	0,55	0,2088	0,87	0,3078	1,19	0,3830
0,24	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,25	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,26	0,1026	0,58	0,2190	0,90	0,3159	1,22	0,3883
0,27	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,28	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,29	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,30	0,1179	0,62	0,2324	0,94	0,3264		
0,31	0,1217	0,63	0,2357	0,95	0,3289		

Окончание табл. 2

x	$\Phi_p(x)$	x	$\Phi_p(x)$	x	$\Phi_p(x)$	x	$\Phi_p(x)$
1,26	0,3962	1,59	0,4441	1,92	0,4726	2,50	0,4938
1,27	0,3980	1,60	0,4452	1,93	0,4732	2,52	0,4941
1,28	0,3997	1,61	0,4463	1,94	0,4738	2,54	0,4945
1,29	0,4015	1,62	0,4474	1,95	0,4744	2,56	0,4948
1,30	0,4032	1,63	0,4484	1,96	0,4750	2,58	0,4951
1,31	0,4049	1,64	0,4495	1,97	0,4756	2,60	0,4953
1,32	0,4066	1,65	0,4505	1,98	0,4761	2,62	0,4956
1,33	0,4082	1,66	0,4515	1,99	0,4767	2,64	0,4959
1,34	0,4099	1,67	0,4525	2,00	0,4772	2,66	0,4961
1,35	0,4115	1,68	0,4535	2,02	0,4783	2,68	0,4963
1,36	0,4131	1,69	0,4545	2,04	0,4793	2,70	0,4965
1,37	0,4147	1,70	0,4554	2,06	0,4803	2,72	0,4967
1,38	0,4162	1,71	0,4564	2,08	0,4812	2,74	0,4969
1,39	0,4177	1,72	0,4573	2,10	0,4821	2,76	0,4971
1,40	0,4192	1,73	0,4582	2,12	0,4830	2,78	0,4973
1,41	0,4207	1,74	0,4591	2,14	0,4838	2,80	0,4974
1,42	0,4222	1,75	0,4599	2,16	0,4846	2,82	0,4976
1,43	0,4236	1,76	0,4608	2,18	0,4854	2,84	0,4977
1,44	0,4251	1,77	0,4616	2,20	0,4861	2,86	0,4979
1,45	0,4265	1,78	0,4625	2,22	0,4868	2,88	0,4980
1,46	0,4279	1,79	0,4633	2,24	0,4875	2,90	0,4981
1,47	0,4292	1,80	0,4641	2,26	0,4881	2,92	0,4982
1,48	0,4306	1,81	0,4649	2,28	0,4887	2,94	0,4984
1,49	0,4319	1,82	0,4656	2,30	0,4893	2,96	0,4985
1,50	0,4332	1,83	0,4664	2,32	0,4898	2,98	0,4986
1,51	0,4345	1,84	0,4671	2,34	0,4904	3,00	0,49865
1,52	0,4357	1,85	0,4678	2,36	0,4909	3,20	0,49931
1,53	0,4370	1,86	0,4686	2,38	0,4913	3,40	0,49966
1,54	0,4382	1,87	0,4693	2,40	0,4918	3,60	0,499841
1,55	0,4394	1,88	0,4699	2,42	0,4922	3,80	0,499928
1,56	0,4406	1,89	0,4706	2,44	0,4927	4,00	0,499968
1,57	0,4418	1,90	0,4713	2,46	0,4931	4,50	0,499997
1,58	0,4429	1,91	0,4719	2,48	0,4934	5,00	0,499997

Таблица 3. Значения функции Пуассона $P(m, \lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$

m/\lambda	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,90484	0,81873	0,74082	0,67032	0,60653	0,54881	0,49659	0,44033	0,40657	0,36788
1	0,09048	0,16375	0,22225	0,26813	0,30327	0,32929	0,34761	0,35946	0,36591	0,36788
2	0,00452	0,01637	0,03334	0,05363	0,07582	0,09879	0,12166	0,14379	0,16466	0,18394
3	0,00015	0,00109	0,00333	0,00715	0,01264	0,01976	0,02839	0,03834	0,04940	0,06131
4		0,00005	0,000250	0,00072	0,00158	0,00296	0,00497	0,00767	0,01111	0,01533
5			0,00002	0,00006	0,00016	0,00036	0,00070	0,00123	0,00200	0,00307
6					0,00001	0,00004	0,00008	0,00016	0,00030	0,00051
7							0,00001	0,00002	0,00004	0,00007
8										0,00001
m/\lambda	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0
0	0,22313	0,13534	0,08208	0,04979	0,03020	0,01832	0,01111	0,00674	0,00409	0,00248
1	0,33740	0,27067	0,20521	0,14936	0,10569	0,07326	0,04999	0,03369	0,02248	0,01487
2	0,25102	0,27067	0,25652	0,22404	0,18496	0,14653	0,11248	0,08422	0,06181	0,04462
3	0,12551	0,18045	0,21376	0,22404	0,21579	0,19537	0,16872	0,14037	0,11332	0,08924
4	0,04707	0,09022	0,13360	0,16803	0,18881	0,19537	0,18981	0,17547	0,15582	0,13385

Окончание табл. 3

[illegible]

Таблица 4. Критические точки распределения хи-квадрат

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	10,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Таблица 5. Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)						
	0,2	0,1	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,7	318,3	637,0
2	1,89	2,92	4,30	6,97	9,92	22,33	31,6
3	1,64	2,35	3,18	4,54	5,84	10,22	12,9
4	1,53	2,13	2,78	3,75	4,60	7,17	8,61
5	1,48	2,01	2,57	3,37	4,03	5,89	6,86
6	1,44	1,94	2,45	3,14	3,71	5,21	5,96
7	1,42	1,89	2,36	3,00	3,50	4,79	5,40
8	1,40	1,86	2,31	2,90	3,36	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	4,03	4,44
12	1,36	1,78	2,18	2,68	3,05	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,69	4,01
17	1,33	1,74	2,11	2,57	2,90	3,65	3,95
18	1,33	1,73	2,10	2,55	2,88	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,58	3,88
20	1,33	1,73	2,09	2,53	2,85	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,49	3,77
24	1,32	1,71	2,06	2,49	2,80	3,47	3,74
25	1,32	1,71	2,06	2,49	2,79	3,45	3,72
26	1,32	1,71	2,06	2,48	2,78	3,44	3,71
27	1,31	1,71	2,05	2,47	2,77	3,42	3,69
28	1,31	1,70	2,05	2,46	2,76	3,40	3,66
29	1,31	1,70	2,05	2,46	2,76	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,28	1,65	1,96	2,33	2,58	3,09	3,29
Число степеней свободы k	Уровень значимости α (односторонняя критическая область)						
	0,1	0,05	0,025	0,01	0,005	0,001	0,0005

Таблица 6. Критические точки распределения Фишера—Снедекора

 $(k_1$ — число степеней свободы в числителе, k_2 — число степеней свободы в знаменателе)Уровень значимости $\alpha = 0,01$

$k_2 \backslash k_1$	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
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

Уровень значимости $\alpha = 0,05$

$k_1 \backslash k_2$	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	18,51	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

Таблица 7. Критические точки для статистики Колмогорова D_n

Объем выборки n	Уровень значимости α			
	0,10	0,05	0,02	0,01
1	0,95	0,98	0,99	0,995
2	0,78	0,84	0,90	0,93
3	0,64	0,71	0,78	0,83
4	0,57	0,62	0,69	0,73
5	0,51	0,56	0,62	0,67
6	0,47	0,52	0,58	0,62
7	0,44	0,48	0,54	0,58
8	0,41	0,45	0,51	0,54
9	0,39	0,43	0,48	0,51
10	0,37	0,41	0,46	0,49
11	0,35	0,39	0,44	0,47
12	0,34	0,38	0,42	0,45
13	0,33	0,36	0,40	0,43
14	0,31	0,35	0,39	0,42
15	0,30	0,34	0,38	0,40
16	0,29	0,33	0,37	0,39
17	0,29	0,32	0,36	0,38
18	0,28	0,31	0,34	0,37
19	0,27	0,30	0,34	0,36
20	0,26	0,29	0,33	0,35

Критические точки распределения Колмогорова

$$Q(\lambda) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 \lambda^2}$$

α	0,10	0,05	0,02	0,01
$\lambda_{кр}$	1,23	1,36	1,52	1,63

Таблица 8. Критические точки распределения Кокрена
 (k — число степеней свободы, l — количество выборок)
 Уровень значимости $\alpha = 0,05$

l/k	1	2	3	4	5	6	7	8	9	10
2	0,999	0,975	0,939	0,906	0,877	0,853	0,833	0,816	0,801	0,788
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633	0,617	0,603
4	0,907	0,768	0,684	0,629	0,590	0,560	0,537	0,518	0,502	0,488
5	0,841	0,634	0,598	0,544	0,506	0,478	0,456	0,439	0,424	0,412
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	0,382	0,368	0,357
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	0,338	0,326	0,315
8	0,680	0,516	0,438	0,391	0,360	0,336	0,319	0,304	0,293	0,283
9	0,639	0,478	0,403	0,358	0,329	0,307	0,290	0,277	0,266	0,257
10	0,602	0,445	0,373	0,331	0,303	0,282	0,267	0,254	0,244	0,235

ОТВЕТЫ И УКАЗАНИЯ

Глава 1. Элементы комбинаторного анализа

1. 8. 2. 125. 3. 1022. 4. 2^{16} . 5. 16^{100} . 6. 3^6 . 7. 7^6 . 8. 14 670.
9. 256. 10. $30^3 \cdot 10^3$. 11. $A_5^4 - A_4^4 = 96$. 12. 1800. 13. $A_{25}^3 = 13\,800$.
14. $10!$. 15. $9! \cdot 8$. 16. $5 \cdot 2 \cdot 4! = 240$. 17. $14 \cdot 8!$. 18. $\frac{14!}{2^7}$. 19. $\frac{9!}{4}$.
20. $\frac{10!}{24}$. 21. 84. 22. 100. 23. 45. 24. 50 600. 25. $50 \cdot 49 \cdot C_{48}^{10}$.
26. $C_{10}^5 - C_7^2 = 231$. 27. 2520. 28. 210. 29. $C_{25}^3 = 2300$.
30. $C_{25}^3 C_{12}^4 = 16\,824\,500$. 31. $C_9^2 = 36$. 32. 165. 33. 1365. 34. 28. 35. 220.
36. 210. 37. $\frac{16!}{5!7!4!}$. 38. $(C_{12}^6)^2$. 39. 210. 40. $A_{10}^3 N_9(2,3,4) = \frac{10!}{4}$.
41. $\frac{16!}{288}$. 42. 3136. 43. 54. 44. $2(5!)^2$. 45. 1). 3^6 ; 2). 6!

Глава 2. Классическая вероятностная модель. Геометрическая вероятность

1. 5/18. 2. 0,52. 3. а) $1/5^4$; б) $1/5^3$; в) $24/5^3$. 4. 0,994. 5. 0,00756.
6. 0,35. 7. 0,082. 8. 0,384. 9. 50/81. 10. 0,264. 11. а) 0,765;
б) 0,619; в) 0,215. 12. а) $2/3$; б) 0,461; в) 0,154. 13. 50/81. 14.
а) 0,381; б) 0,513. 15. 0,215. 16. 0,1008. 17. 0,085. 18. 25/63.
19. $2/7$. 20. 0,155. 21. 0,0476. 22. $1/7$. 23. а) $5/18$; б) $3/18$;
в) $5/9$. 24. 0,43. 25. 0,067. 26. а) 0,603; б) 0,612. 27. а) 0,238;
б) 0,124. 28. а) 0,152; б) 0,758; в) 0,197; г) 0,657. 29. а) 0,725;
б) 0,322; в) 0,305; г) 0,459. 30. 0,17. 31. 0,47. 32. 0,094. 33. 0,126.
34. а) 0,892; б) 0,903. 35. а) $2/7$; б) 0,135. 36. 0,9. 37. $1 - (1/12)^2$.
38. 49/64. 39. 8/9. 40. 4/9. 41. 0,88. 42. 0,6. 43. 49/63. 44. 0,2.
Указание: запишите условие задачи в виде неравенства, изобразите графически события и вычислите площади с помощью интегралов ($\ln 9 \approx 2,2$). 45. 0,5. Введите пространственную систему координат. Возможные значения x, y, z — от 0 до L . События, благоприятствующие условию задачи, $x < y + z, y < x + z, z < x + y$.

Глава 3. Основные формулы теории вероятностей

1. а) $A_1 A_2 \bar{A}_3 + \bar{A}_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3$; б) $\bar{A}_1 \bar{A}_2 \bar{A}_3$; в) $\bar{A}_1 \bar{A}_2 \bar{A}_3$, $p = 0,488$.
 2. а) $\bar{A}_1 \bar{A}_2 \bar{A}_3$; б) $A_1 \bar{A}_2 \bar{A}_3 + \bar{A}_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3$; в) $\bar{A}_1 \bar{A}_2 \bar{A}_3$, $p = 0,784$.
 3. а) A_1 ; б) $\bar{A}_1 A_2$; в) $\bar{A}_1$; $p = 0,09$. 4. $n \geq 17$. 5. 0,5. 6. 0,075.
 7. 0,75. 8. 0,099. 9. 2/9. 10. 0,114. 11. 0,4. 12. а) 0,38; б) 0,25;
 в) 0,41; г) 0,59. 13. 0,2499. 14. 0,2334. 15. 0,781. 16. 0,66.
 17. 0,535. 18. 0,512. 19. 0,5. 20. 0,5. 21. 0,476. 22. 2/7. 23. 0,51.
 24. 0,67. 25. 7/15. 26. 0,98. 27. 0,2. 28. 0,014. 29. 0,24. 30. а) 0,1424;
 б) 0,1756. 31. 0,085. 32. а) 0,562; б) 0,5072; в) 0,34; г) 0,722.
 33. 0,9148. 34. 0,0631. 35. 17/45. 36. 0,282. 37. 0,3098. 38.
 0,58. 39. 0,328. 40. 34/55. 41. 5/9. 42. а) 0,077; б) 1/46; в) 1/41.
 43. 0,182. 44. а) 5/9; б) 0,308. 45. а) 15/22; б) 8/13. 46. 0,00207.
 47. 0,138. 48. а) 0,83; б) 0,777. 49. а) 12/13; б) 51/57; в) 56/59.
 50. а) 3/5; б) 7/13. 51. 4/45. 52. 6/17. 53. 23/32. 54. Стрелок В
 попал в мишень с вероятностью 10/19 и не попал с вероятностью
 9/19. Более вероятно, что он попал. 55. Стрелок В попал в мишень
 с вероятностью 10/19 и не попал с вероятностью 9/19. Более
 вероятно, что он попал. 56. 0,75. 57. 5/12. 58. 2/9. 59. 12/19.
 60. 12/37.

Глава 4. Повторные независимые испытания

1. 0,05. 2. 0,058. 3. 0,0081. 4. 0,05. 5. 0,82. 6. 0,463. 7. а) Ве-
 роятнее выиграть одну партию из двух; б) вероятнее выиграть две
 партии из четырех. 8. 7. 9. 7/16. 10. а) 15/128; б) 0,205; в) 0,044.
 11. $n \geq 5$. 12. 0,4096. 13. 0,813. 14. 0,384. 15. 9; 0,24. 16. а) 2;
 б) 0,324; в) 0,0109. 17. 5; 0,2078. 18. а) 299; 2 и 3; б) 174; 3
 в) 228; 4. 19. а) 2; 0,678 б) 1; 0,443. 20. $n = 33, 34, 35$. 21. 21.
 22. $n = 11, 12, 13$. 23. $n = 14, 15, 16, 17$. 24. а) 0,079; б) 0,04.
 25. а) 0,042; б) 0,007. 26. а) $2\Phi_0(2,5) = 0,9876$; б) 0,5.
 27. $0,5 - \Phi_0(1,09) = 0,1379$. 28. а) 576; $\Phi_0(2) + \Phi_0(1) = 0,8185$;
 б) 640; $\Phi_0(0,88) + \Phi_0(3,54) \approx 0,8104$; в) 320; $0,5 + \Phi_0(2,5) = 0,9938$;
 г) 500; $\Phi_0(2) + \Phi_0(1) = 0,8185$. 29. а) $\Phi_0(2,14) + 0,5 = 0,9838$;
 б) $\Phi_0(3,57) + 0,5 \approx 0,9998$. 30. а) 200; $2\Phi_0(2,14) = 0,9676$;
 б) 162; $2\Phi_0(1,98) = 0,9522$; в) 20; $2\Phi_0(2,26) = 0,9762$.
 31. а) $0,5 - \Phi_0(1,94) = 0,0262$; б) $0,5 - \Phi_0(2,74) = 0,0031$.
 32. $0,5 + \Phi_0(1,45) = 0,9265$. 33. 25; $2\Phi_0(2,05) \approx 0,96$. 34. $\Phi_0(3,75) +$

+ $\Phi_0(2,5) \approx 0,9937$. 35. $\Phi_0(0,83) + \Phi_0(3,33) \approx 0,7964$. 36. $0,5 + \Phi_0(0,89) = 0,8133$. 37. $\varepsilon = 0,057$. 38. От 131 до 169. 39. От 65 до 95. 40. От 73 до 107. 41. От 81 до 119. 42. 289. 43. 666. 44. 0,998. 45. От 3904 до 4096. 46. а) 0,00125; б) 0,998. 47. 0,92. 48. а) 0,616; б) 0,63. 49. а) 3; 0,647 б) 3; 0,452. 50. 0,195. 51. 0,875. 52. а) 0,135; б) 0,676. 53. 0,175. 54. а) 0,125; б) 0,433; в) 0,285. 55. 0,981. $P\{m \geq 1\} = 1 - P(0) = 1 - e^{-\lambda} = 0,632$. Отсюда $\lambda \approx 1$ и $P\{m \leq 3\} = P(0) + P(1) + P(2) + P(3) = 0,981$ (формула Пуассона). 56. 3; 0,224. 57. 0,08505. 58. 0,131. 59. 10/81. 60. а) 0,14375; б) 0,0612.

Глава 5. Дискретные случайные величины

Теоретические задачи

1. Воспользуйтесь разложением $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$. 2. Воспользуйтесь представлением $\xi = \xi_1 + \dots + \xi_n$, где $\xi_i = 0$ в случае неудачи в i -м испытании, и $\xi_i = 1$ в случае успеха. 3. Воспользуйтесь тем, что $np^{n-1} = (p^n)'$.

Вычислительные задачи

4.

ξ	0	1	2	3
P	1/8	3/8	3/8	1/8

5.

ξ	0	1	2	3	4	5
P	0,59049	0,32805	0,0729	0,0081	0,00045	0,00001

6.

ξ	0	1	2	3
P	0,006	0,092	0,398	0,504

7.

ξ	0	1	2	3
P	0,002	0,044	0,306	0,648

8.

ξ	0	1	2	3
P	0,893475	0,103075	0,003425	0,000025

9.

ξ	0	1	2	3	4
P	0,53439	0,36251	0,0922	0,0104	0,00044

10.

ξ	1	2	3
P	0,7	0,21	0,09

11.

ξ	2	3	4
P	1/4	1/4	1/2

12. а)

ξ	2	3	4
P	0,81	0,162	0,028

б)

ξ	3	4	5
P	0,729	0,2187	0,0523

в)

ξ	3	4	5	6
P	0,8574	0,1286	0,0129	0,0011

г)

ξ	2	3	4	5	6
P	0,64	0,256	0,0768	0,02048	0,00672

д)

ξ	2	3	4	5	6
P	0,81	0,162	0,0243	0,00324	0,00046

13. а)

ξ	2	3	4
P	1/15	7/60	49/60

б)

ξ	2	3	4	5
P	5/22	3/11	5/22	3/11

в)

ξ	3	4	5
P	10/143	21/143	112/143

г)

ξ	2	3	4	5
P	7/22	7/22	7/33	5/33

д)

ξ	2	3	4	5
P	3/28	5/28	3/14	1/2

е)

ξ	2	3	4	5
P	2/15	1/5	3/14	19/42

ж)

ξ	2	3	4
P	2/9	5/18	1/2

14.

ξ	1	2	3	4
P	0,4	0,3	0,2	0,1

15.

ξ	1	2	3	4
P	1/2	3/10	3/20	1/20

16. а)

ξ	-2	-1	0	1	2
P	1/9	2/9	1/3	2/9	1/9

б)

ξ	-3	-2	-1	0	1	2	3
P	1/27	1/9	2/9	7/27	2/9	1/9	1/27

17. а) $M\xi = 0,6$

ξ	-2	-1	0	1	2
P	0,09	0,06	0,37	0,12	0,36

б) $M\xi = 0,2$

ξ	-2	-1	0	1	2
P	0,16	0,08	0,41	0,1	0,25

в) $M\xi = 1$

ξ	-2	-1	0	1	2
P	0,04	0,04	0,29	0,14	0,49

18. $M\xi = 187,2$ (руб.)

ξ	0	1	2	3
P	0,1	0,63	0,243	0,027

19. а) $M\xi = 1,264$

ξ	0	1	2	3
P	0,3	0,28	0,294	0,126

б) $M\xi = 1,152$

ξ	0	1	2	3
P	0,2	0,48	0,288	0,032

20. $M\xi = 2,05$

ξ	1	2	3
P	0,3	0,35	0,35

21. а) $M\xi = 1,9656$

ξ	0	1	2	3	4
P	0,1	0,18	0,432	0,2304	0,0576

б) $M\xi = 1,8792$

ξ	0	1	2	3	4
P	0,1	0,18	0,504	0,1728	0,0432

в) $M\xi = 1,8324$

ξ	0	1	2	3	4
P	0,1	0,27	0,378	0,2016	0,0504

22. а) $M\xi = 2,952$

ξ	1	2	3	4
P	0,2	0,16	0,128	0,512

б) $M\xi = 2,7731$

ξ	1	2	3	4	5
P	0,3	0,21	0,147	0,1029	0,2401

в) $M\xi = 3,6893$

ξ	1	2	3	4	5	6
P	0,2	0,16	0,128	0,1024	0,08192	0,32768

23. $M\xi = -1,3$ (руб.)

ξ	-5	5	95
P	0,72	0,27	0,01

24. $M\xi = -30/49$ (руб.)

ξ	-10	10	90
P	30/49	18/49	1/49

25. а) $M\xi = 1,8$

ξ	0	1	2	3
P	0,042	0,284	0,506	0,168

б) $M\xi = 1,7$

ξ	0	1	2	3
P	0,035	0,365	0,465	0,135

в) $M\xi = 1,7$

ξ	0	1	2	3
P	0,056	0,332	0,468	0,144

26. а) $M\xi = 2,2$

ξ	0	1	2	3
P	0,012	0,154	0,456	0,378

б) $M\xi = 2$

ξ	0	1	2	3
P	0,03	0,22	0,47	0,28

27. $M\xi = 3$

ξ	0	1	2	3	4
P	0,0036	0,0456	0,2116	0,4256	0,3136

28. $M\xi = 0,6$

ξ	0	1	2	3	4
P	0,5184	0,3744	0,0964	0,0104	0,0004

29. $M\xi = 1,624$; $D\xi = 0,811$.

30. $M\xi = 14$; $D\xi = 35/3$.

31. а) $P\{\xi > 2\} = 1 - P\{\xi \leq 2\} = 1 - P(0) - P(1) - P(2) = 1 - 5e^{-2} = 0,323$;

б) $P\{\xi \leq 1\} = P(0) + P(1) = 3e^{-2} = 0,406$;

в) $P\{\xi \leq 2\} = 5e^{-2} = 0,677$.

32. 5 (геометрическое распределение).

33. 4 (геометрическое распределение).

34. 40,96. 35. а) 12; б) 18. 36. а) $M\xi = 11/2$, $D\xi = 27/4$;

б) $M\xi = 25/3$, $D\xi = 130/9$; в) $M\xi = 137/12$, $D\xi = 3625/144$.

37. $M\xi = 3/2$; $D\xi = 3/4$; $\sigma = \sqrt{3}/2$; $P\{-1 < \xi < 3/2\} = 1/2$;

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1/8, & 0 < x \leq 1, \\ 1/2, & 1 < x \leq 2, \\ 7/8, & 2 < x \leq 4, \\ 1, & x > 4. \end{cases}$$

38. $M\xi = 7/4$; $D\xi = 55/16$; $\sigma \approx 1,854$; $P\{5/2 < \xi < 5\} = 1/8$;

$$F(x) = \begin{cases} 0, & x \leq -1, \\ 1/4, & -1 < x \leq 2, \\ 3/4, & 2 < x \leq 3, \\ 7/8, & 3 < x \leq 5, \\ 1, & x > 5. \end{cases}$$

39. $M\xi = 49/18$; $D\xi \approx 2,312$; $\sigma \approx 1,52$; $P\{1/2 < \xi < 7/2\} = 7/18$;

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1/9, & 0 < x \leq 1, \\ 1/3, & 1 < x \leq 3, \\ 1/2, & 3 < x \leq 4, \\ 1, & x > 4. \end{cases}$$

40. $M\xi = 0$; $D\xi = 18/7$; $\sigma \approx 1,604$; $P\{-3/2 < \xi < 2\} = 5/7$;

$$F(x) = \begin{cases} 0, & x \leq -2, \\ 1/7, & -2 < x \leq -1, \\ 4/7, & -1 < x \leq 1, \\ 6/7, & 1 < x \leq 3, \\ 1, & x > 3. \end{cases}$$

41. $1/4$. 42. $1/6$. 43. $1/3$ и $1/6$. 44. $1/6$ и $-1/4$.

45. $\text{cov}(\xi + \eta, \xi - \eta) = -0,44$; ξ и η зависимы.

ξ	-1	0	1	2
P	1/5	2/5	1/5	1/5

46. $\text{cov}(2\xi - 3\eta, \xi + 2\eta) = -13/6$; ξ и η зависимы.

ξ	-1	0	1
P	5/12	1/3	1/4

47. $\text{cov}(2\eta + \xi, \eta + \xi) = 176/81$; ξ и η зависимы.

ξ	-1	0	1	2	3
P	1/9	1/6	4/9	1/18	2/9

48. $\text{cov}(\xi - \eta, 2\xi + \eta) = -47/36$; ξ и η зависимы.

ξ	-1	0	1	2	3
P	1/8	1/4	1/12	1/6	3/8

49. $\text{cov}(2\xi + \eta, 3\xi - \eta) = 406/81$; ξ и η зависимы.

x	-2	-1	1	2
P	1/6	2/9	2/9	7/18

50. $\text{cov}(\eta, \eta + 4\xi) = 4,69$

ξ	-2	-1	0	1	3	4
P	1/10	1/10	1/5	1/10	1/5	3/10

51. $\text{cov}(\xi + 5\eta, 2\xi - 7\eta) = -15,3$

ξ	-	0	1	2	3	4
P	0,05	0,1	0,15	0,1	0,2	0,4

52. $\text{cov}(\xi - 3\eta, \eta + 6\xi) = -143/49$

ξ	-2	-1	0	1	2
P	1/7	1/7	2/7	2/7	1/7

53. $\text{cov}(4\xi - \eta, 2\eta + \xi) = 1226/225$

ξ	-3	-2	0	1	3
P	2/15	1/5	2/5	1/15	1/5

54. $\text{cov}(2\xi - \eta, \eta - 3\xi) = -21/8$

ξ	-	-3	-2	-1	0
P	1/12	1/4	1/6	1/4	1/4

55. 0,356.

56. 0,6.

57. а) 153,76; б) 119,29; в) 116,41.

58. 6 акций первой компании и 4 акции второй компании.

59. 9 акций первой компании и 3 акции второй компании.

60. 6/7 – в акции первой компании, 1/7 – в акции второй компании.

61. Все в акции первой компании.

62. а)

x	0	1
$P_{\xi \eta}(x 0)$	2/3	1/3

б)

y	-1	0	1
$P_{\eta \xi}(y 0)$	0,2	0,4	0,4

в)

y	-1	0	1
$\phi_{\xi \eta}$	2/3	1/3	1/2

г)

x	0	1
$\phi_{\eta \xi}$	0,2	0

63. а)

x	-1	1
$P_{\xi \eta}(x -1)$	1/4	3/4

б)

y	-1	0	1
$P_{\eta \xi}(y 1)$	1/2	1/6	1/3

в)

y	-1	0	1
$\phi_{\xi \eta}$	1/2	-1/2	0

г)

x	-1	1
$\phi_{\eta \xi}$	1/6	-1/6

64. а)

x	-1	1
$P_{\xi \eta}(x 2)$	3/5	2/5

б)

y	0	1	2
$P_{\eta \xi}(y 1)$	2/11	3/11	6/11

в)

y	0	1	2
$\phi_{\xi \eta}$	0	-1/2	-1/5

г)

x	-1	1
$\phi_{\eta \xi}$	15/11	9/7

65. Биномиальное распределение с параметрами n и $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$,

функция регрессии $\phi_{\xi|\eta}(k) = \frac{k\lambda_1}{\lambda_1 + \lambda_2}$, $k = 0, 1, 2, \dots, n$.

Глава 6. Непрерывные случайные величины

Теоретические задачи

1. Воспользоваться формулой для плотности функции от случайной величины.

2. а) $P(\eta_1 < x) = 1 - P(\xi_1 \geq x, \xi_2 \geq x, \dots, \xi_n \geq x) = 1 - (P(\xi \geq x))^n = 1 - e^{-ax}$;

б) $P(\eta_2 < x) = P(\xi_1 < x, \xi_2 < x, \dots, \xi_n < x) = (P(\xi < x))^n = (1 - e^{-ax})^n$.

3. Найти функции распределения для η_1 и η_2 методом, использованным в задаче 2, и найти их математические ожидания.

4. а) $a = \frac{1}{\pi}$; б) $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctg x$; в) $P\{-1 < \xi < 1\} = 1/2$.

$$5. a = \frac{\lambda}{2}; M\xi = 0, D\xi = \frac{2}{\lambda^2}; P\{|\xi| < \sqrt{D\xi}\} = 1 - e^{-\sqrt{2}} \approx 0,75;$$

$$P\{|\xi| < 3\sqrt{D\xi}\} = 1 - e^{-3\sqrt{2}} \approx 0,98.$$

$$6. a) 0,9974; б) 0,9817; в) 1.$$

$$7. p(x) = \frac{1}{a} \left(1 - \frac{|x|}{a} \right), |x| < a; M\xi = 0; D\xi = \frac{a^2}{6}$$

8. Воспользоваться методом геометрической вероятности:

$$P(\rho < x) = P(\xi^2 + \eta^2 < x^2) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}; \quad x \in [0, R]; \quad M\rho = \frac{2}{3}R, D\rho = \frac{R^2}{18}.$$

$$9. a) m_{\xi}(t) = \begin{cases} \frac{e^{bt} - e^{at}}{t(b-a)}, & t \neq 0; \\ 1, & t = 0; \end{cases} \quad б) m_{\xi}(t) = \frac{\lambda^2}{\lambda^2 - t^2}, |t| < \lambda.$$

10. Указание. Изучить соответствующие интегралы на сходимость.

Вычислительные задачи

$$11. a) C = 2/9; F(x) = \begin{cases} 0, & x < -1; \\ \frac{(x+1)^2}{9}, & -1 \leq x \leq 2; \\ 1, & x > 2; \end{cases} M\xi = 1; P\{\xi^2 < 1\} = 4/9;$$

$$б) C = 2/33; F(x) = \begin{cases} 0, & x < -2; \\ \frac{-x^2 + 10x + 24}{33}, & -2 \leq x \leq 1; \\ 1, & x > 1; \end{cases}$$

$$M\xi = -7/11; P\{0 < \xi < 2\} = 3/11;$$

$$в) C = 1/4; F(x) = \begin{cases} 0, & x < 3; \\ \frac{x^2 - 4x + 3}{8}, & 3 \leq x \leq 5; \\ 1, & x > 5; \end{cases}$$

$$M\xi = 25/6; P\{4 < \xi < 6\} = 5/8;$$

$$\text{г) } C=1/20; F(x)=\begin{cases} 0, & x < -1; \\ \frac{x^2+8x+7}{40}, & -1 \leq x \leq 3; \\ 1, & x > 3; \end{cases}$$

$$M\xi=19/15; P\{2 < \xi < 5\}=13/40;$$

$$\text{д) } C=2/63; F(x)=\begin{cases} 0, & x < -5; \\ \frac{-x^2+6x+55}{63}, & -5 \leq x \leq 2; \\ 1, & x > 2; \end{cases}$$

$$M\xi=-65/27; P\{0 < \xi < 5\}=8/63.$$

$$12. \text{ а) } C=3/20; F(x)=\begin{cases} 0, & x < 1; \\ \frac{x^3-3x+2}{20}, & 1 \leq x \leq 3; \\ 1, & x > 3; \end{cases}$$

$$M\xi=12/5; P\{2 < \xi < 5\}=4/5;$$

$$\text{б) } C=1/12; F(x)=\begin{cases} 0, & x < -2; \\ \frac{x^3+9x+26}{36}, & -2 \leq x \leq 1; \\ 1, & x > 1; \end{cases}$$

$$M\xi=-11/16; P\{-1 < \xi < 2\}=5/9;$$

$$\text{в) } C=0,03; F(x)=\begin{cases} 0, & x < -3; \\ \frac{-x^3+27x+54}{100}, & -3 \leq x \leq 2; \\ 1, & x > 2; \end{cases}$$

$$M\xi=-3/16; P\{0 < \xi < 3\}=0,64;$$

$$\text{г) } C=2/57; F(x)=\begin{cases} 0, & x < 1; \\ \frac{2x^3+3x^2-5}{171}, & 1 \leq x \leq 4; \\ 1, & x > 4; \end{cases}$$

$$M\xi=113/38; P\{3 < \xi < 6\}=5/9;$$

$$д) C=1/18; F(x)=\begin{cases} 0, & x < 2; \\ \frac{x^3-3x^2+4}{54}, & 2 \leq x \leq 5; \\ 1, & x > 5; \end{cases}$$

$$M\xi = 33/8; P\{1 < \xi < 4\} = 10/27;$$

$$е) C=1/324; F(x)=\begin{cases} 0, & x < 2; \\ \frac{x^4+4x-24}{1296}, & 2 \leq x \leq 6; \\ 1, & x > 6; \end{cases}$$

$$M\xi = 652/135; P\{5 < \xi < 7\} = 25/48;$$

$$ж) C=1/24; F(x)=\begin{cases} 0, & x < 1; \\ \frac{x^4+8x-9}{96}, & 1 \leq x \leq 3; \\ 1, & x > 3; \end{cases}$$

$$M\xi = 47/20; P\{0 < \xi < 2\} = 23/96.$$

$$13. а) C=1/\ln 4; F(x)=\begin{cases} 0, & x < 5; \\ \frac{\ln(x-4)}{\ln 4}, & 5 \leq x \leq 8; \\ 1, & x > 8; \end{cases}$$

$$M\xi = 4 + \frac{3}{\ln 4}; P\{3 < \xi < 6\} = 1/2;$$

$$б) C=1/\ln 5; F(x)=\begin{cases} 0, & x < -1; \\ \frac{\ln(x+2)}{\ln 5}, & -1 \leq x \leq 3; \\ 1, & x > 3; \end{cases}$$

$$M\xi = \frac{4}{\ln 5} - 2; P\{0 < \xi < 2\} = 1 - \frac{\ln 2}{\ln 5};$$

$$в) C=1/\ln 2; F(x)=\begin{cases} 0, & x < -2; \\ \frac{\ln(10/(8-x))}{\ln 2}, & -2 \leq x \leq 3; \\ 1, & x > 3; \end{cases}$$

$$M\xi = 8 - \frac{5}{\ln 4}; P\{2 < \xi < 5\} = \frac{\ln(6/5)}{\ln 2};$$

$$r) C = 1/\ln 2; F(x) = \begin{cases} 0, & x < -1; \\ \frac{\ln((x+6)/5)}{\ln 2}, & -1 \leq x \leq 4; \\ 1, & x > 4; \end{cases}$$

$$M\xi = \frac{5}{\ln 2} - 6; P\{0 < \xi < 5\} = \frac{\ln(10/6)}{\ln 5}.$$

14. $C = 1/2; F(x) = \begin{cases} 0, & x < 0, \\ 1 - (x+1)^{-1/2}, & x \geq 0; \end{cases}$ математическое ожидание и дисперсия не существуют; $P\{|\xi - 1/3| < 1\} = 1 - (7/3)^{-1/2}$.

$$15. C = 3/4; F(x) = \begin{cases} 0, & x < -1; \\ \frac{3x - x^3 + 2}{4}, & -1 \leq x \leq 1; \\ 1, & x > 1; \end{cases}$$

$$M\xi = 0; D\xi = 1/5; P\{|\xi - 1/2| < 1/4\} = 35/128.$$

$$16. C = 1; F(x) = \begin{cases} e^x, & x < 0; \\ 1, & x \geq 0; \end{cases} M\xi = -1, D\xi = 1; P\{-2 < \xi < 1\} = 1 - e^{-2} = 0,86.$$

$$17. p(x) = \frac{1}{2}e^{-|x|}; M\xi = 0, D\xi = 2; P\{-1 < \xi < 3\} \approx 0,79.$$

$$18. p(x) = \begin{cases} 2 - 2x, & x \in [0, 1]; \\ 0, & x \notin [0, 1]; \end{cases} M\xi = 1/3, D\xi = 1/18.$$

19. Указание: доказать, что случайная величина $\eta = \max(\xi, 24 - \xi)$ распределена по равномерному закону на $[12; 24]$. Ответ: 18 см.

20. Указание: доказать, что случайная величина $\eta = \min(\xi, 12 - \xi)$ распределена по равномерному закону на $[0; 6]$. Ответ: 3 см.

$$21. F(x) = \begin{cases} 0, & x < 0; \\ \frac{x}{6} - \frac{x^2}{150}, & 0 \leq x \leq 10; \\ 1, & x > 10. \end{cases}$$

Среднее время ожидания — 3,89 минуты.

$$22. F(x) = \begin{cases} 0, & x < 0; \\ \frac{13}{60}x - \frac{3}{200}x^2 + \frac{1}{3000}x^3, & 0 \leq x \leq 10; \\ 1, & x > 10. \end{cases}$$

Среднее время ожидания — 3 мин 20 с.

23. а) 1/9; б) 0,21.

24. а) 0,607; б) 0,135.

25. а) 0,041; б) 0,112.

26. 1 ч 40 мин.

27. 2 ч 5 мин.

28. 2 ч 20 мин.

29. 1 год 4 мес.

30. 2,4 года.

Глава 7. Функции от случайных величин.

Непрерывный случайный вектор

Теоретические задачи

$$1. \text{ а) } p(x) = \begin{cases} 0, & x < 0, \\ x/a^2, & 0 \leq x \leq a, \\ (2a-x)/a^2, & a \leq x \leq 2a, \\ 0 & x > 2a; \end{cases} \quad \text{б) } p(x) = \begin{cases} 0, & x < -a, \\ (a+x)/a^2, & -a \leq x \leq 0, \\ (a-x)/a^2, & 0 \leq x \leq a, \\ 0, & x > a; \end{cases}$$

$$\text{в) } p(x) = \frac{1}{a^2} \ln \frac{a^2}{x}, 0 < x < a; \quad \text{г) } p(x) = \begin{cases} 0, & x < 0, \\ 1/2, & 0 < x < 1, \\ 1/(2x^2), & x \geq 1. \end{cases}$$

2. Достаточно доказать это свойство для «центрированных» случайных величин, т.е. для тех, у которых $M\xi_i = 0$. Воспользоваться

формулой свертки, учитывая соотношение $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$.

3. Рассмотреть функцию распределения для η .

Вычислительные задачи

$$4. C = \frac{1}{2}, \quad p(x) = \frac{1}{2\sqrt{x}} \text{ при } x \in (0, 1]; \quad P(0,25 < \eta < 0,64) = 0,3.$$

$$5. p(x) = \frac{2}{x^2} \left(1 - \frac{1}{x}\right) \text{ при } x \geq 1.$$

$$6. \text{ а) } p(x) = \frac{1}{4\sqrt{x-1}} \text{ при } x \in [2, 10];$$

$$\text{б) } p(x) = \frac{1}{6\sqrt{10-x}} \text{ при } x \in [2, 10];$$

$$\text{в) } p(x) = \frac{1}{4\sqrt{x}} \text{ при } x \in [9, 25];$$

$$\text{г) } p(x) = \frac{1}{6}x^{-2/3} \text{ при } x \in [1, 27];$$

$$\text{д) } p(x) = \frac{2}{3}x \text{ при } x \in [\sqrt{2}, \sqrt{5}];$$

$$\text{е) } p(x) = 2x \text{ при } x \in [0, 1];$$

$$\text{ж) } p(x) = \frac{1}{3x^2} \text{ при } x \geq 1/3;$$

$$\text{з) } p(x) = \frac{6}{x^2} \text{ при } x \in [2, 3];$$

$$\text{и) } p(x) = \frac{1}{4x^{3/2}} \text{ при } x \geq 1/4;$$

$$\text{к) } p(x) = \frac{1}{2}e^{-x} \text{ при } x \in [-\ln 3, 0].$$

$$7. \text{ а) } p_{\eta}(x) = \frac{e^{-2\sqrt{x+1}}}{\sqrt{x+1}}, x > -1;$$

$$\text{б) } p_{\eta}(x) = \frac{2}{3}(x-1)^{-2/3}e^{-2(x-1)^{1/3}}, x > 1;$$

$$\text{в) } p_{\eta}(x) = \frac{6}{(x-1)^2}e^{-6/(x-1)}, x > 1;$$

$$\text{г) } p(x) = \frac{2}{(x+1)^3}, x > 0;$$

$$\text{д) } p_{\eta}(x) = 2e^{-2(e^x-1)+x}, x \geq 0.$$

$$8. \text{ а) } p(x) = e^{-x} \text{ при } x \geq 0;$$

$$\text{б) } p(x) = \frac{\lambda}{2\sqrt{x}}e^{-\lambda\sqrt{x}} \text{ при } x > 0;$$

$$\text{в) } p(x) = 2\lambda xe^{-\lambda x^2} \text{ при } x \geq 0;$$

$$\text{г) равномерное распределение на отрезке } [0, 1].$$

$$9. \text{ а) } p(x) = \frac{1}{6\sqrt{2\pi}x^{2/3}} e^{-\frac{1}{8}x^{2/3}} \text{ при } x \neq 0;$$

$$\text{б) } p_{\eta}(x) = \frac{1}{9\sqrt{2\pi}}(x-1)^{-2/3} e^{-\frac{1}{18}(\sqrt[3]{x-1}+1)^2}, x \neq 2;$$

$$\text{в) } p_{\eta}(x) = \frac{1}{3\sqrt{2\pi}} x^{-2/3} e^{-\frac{1}{2}(\sqrt[3]{x}-1)^2}, x \neq 0;$$

$$\text{г) } p(x) = \frac{1}{2\sqrt{2\pi}x} e^{-\frac{1}{8}(\ln x)^2} \text{ при } x > 0;$$

$$\text{д) } p_{\eta}(x) = \frac{x^2}{\sqrt{2\pi}} e^{-\frac{1}{18}(x^3-1)^2}.$$

$$10. \text{ а) } p_{\eta}(x) = \frac{1}{3\pi(1+x^{2/3})x^{2/3}}, x \neq 0;$$

$$\text{б) } p_{\eta}(x) = 1/\pi, x \in (-\pi/2, \pi/2);$$

$$\text{в) } p_{\eta}(x) = \frac{1}{\pi(1+(\ln x)^2)x}, x > 0.$$

$$11. \text{ а) } p(x) = \frac{x^{-1/3} + 3x^{-2/3}}{60} \text{ при } x \in [0, 64];$$

$$\text{б) } p(x) = \frac{\sqrt{x+6}-3}{8\sqrt{x+6}} \text{ при } x \in [10, 30];$$

$$\text{в) } p(x) = \frac{1}{63x^{3/2}} \left(8 - \frac{1}{\sqrt{x}} \right) \text{ при } x \geq 1/49;$$

$$\text{г) } p(x) = \frac{2}{55}(9-e^x)e^x \text{ при } x \in [\ln 6, \ln 9];$$

$$\text{д) } p(x) = \frac{8-\ln(x+1)}{28(x+1)} \text{ при } x \in [e^{-1}-1, e^3-1].$$

$$12. \text{ а) } p(x) = \frac{3}{10}(x^5-2x^3) \text{ при } x \in [\sqrt{2}, 2];$$

$$\text{б) } p(x) = \frac{1}{12}x(x^4+2x+2) \text{ при } x \in [0, \sqrt{3}];$$

$$\text{в) } p(x) = \frac{3}{100x^2} \left(\frac{6}{x} - \frac{1}{x^2} \right) \text{ при } x \geq 1/5;$$

$$\text{г) } p(x) = \frac{1}{228} e^x (e^{2x} - 1) \text{ при } x \in [\ln 3, \ln 9];$$

$$\text{д) } p(x) = \frac{1}{18} (e^{2x} - 1) e^x \text{ при } x \in [0, \ln 4];$$

$$\text{е) } p(x) = \frac{1}{162x^2} \left(2 - \frac{6}{x} + \frac{12}{x^2} - \frac{8}{x^3} \right) \text{ при } x \in [-2, -2/5];$$

$$\text{ж) } p(x) = \frac{1}{(\ln 2)(8 - \ln x)x} \text{ при } x \in [1, e^4];$$

$$\text{з) } p(x) = \frac{1}{(\ln 2)(10 - \ln x)x} \text{ при } x \in [1, e^5].$$

13. а) 0,415; б) 1142,7; в) 287,5.

14. а) 116,2; б) 117,6.

15. а) 0,73; б) 10 920; в) 1143,5.

16. а) 0,46; б) 5859,375.

$$17. \text{ а) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 1 \\ \frac{x-1}{2}, & 1 < x < 2 \\ 1/2, & 2 < x < 3 \\ \frac{4-x}{2}, & 3 < x < 4 \\ 0, & x > 4; \end{cases} \quad \text{б) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 3 \\ \frac{x-3}{2}, & 3 < x < 4 \\ 1/2, & 4 < x < 5 \\ \frac{6-x}{2}, & 5 < x < 6 \\ 0, & x > 6; \end{cases}$$

$$\text{в) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 1 \\ \frac{x-1}{4}, & 1 < x < 2 \\ 1/4, & 2 < x < 5 \\ \frac{6-x}{4}, & 5 < x < 6 \\ 0, & x > 6; \end{cases} \quad \text{г) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 3 \\ \frac{x-3}{4}, & 3 < x < 5 \\ \frac{7-x}{4}, & 5 < x < 7 \\ 0, & x > 7; \end{cases}$$

$$\text{д) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{7}, & 0 < x < 1 \\ 1/7, & 1 < x < 7 \\ \frac{8-x}{7}, & 7 < x < 8 \\ 0, & x > 8. \end{cases}$$

$$18. \text{ а) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-2x}, & 0 < x < 1 \\ (e^2 - 1)e^{-2x}, & x > 1; \end{cases}$$

$$\text{б) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 0 \\ (1 - e^{-2x})/3, & 0 < x < 3 \\ (e^6 - 1)e^{-2x}/3, & x > 3; \end{cases}$$

$$\text{в) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 0 \\ (1 - e^{-5x})/2, & 0 < x < 2 \\ (e^{10} - 1)e^{-5x}/2, & x > 2; \end{cases}$$

$$\text{г) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < -1 \\ (1 - e^{-3(x+1)})/2, & -1 < x < 1 \\ (e^3 - e^{-3})e^{-3x}/2, & x > 1; \end{cases}$$

$$\text{д) } p_{\xi+\eta}(x) = \begin{cases} 0, & x < 1 \\ 1 - e^{2-2x}, & 1 < x < 2 \\ (e^4 - e^2)e^{-2x}, & x > 2. \end{cases}$$

$$19. \text{ а) } p(x) = 6(e^{-2x} - e^{-3x}) \text{ при } x \geq 0;$$

$$\text{б) } p(x) = \frac{10}{3}(e^{-2x} - e^{-5x}) \text{ при } x \geq 0;$$

$$\text{в) } p(x) = 12(e^{-4x} - e^{-6x}) \text{ при } x \geq 0; \text{ г) } p(x) = \frac{5}{4}(e^{-x} - e^{-5x}) \text{ при } x \geq 0;$$

$$\text{д) } p(x) = 20(e^{-4x} - e^{-5x}) \text{ при } x \geq 0.$$

$$20. \text{ а) } 3,5 \text{ ч; б) } 3/4. \quad 21. \text{ а) } 5 \text{ ч; б) } 1/2. \quad 22. 0,748. \quad 23. 0,694.$$

24. Во всех случаях случайные величины ξ и η зависимы.

$$\text{а) } p_{\xi}(x) = \frac{2}{9\pi} \sqrt{9 - (x-1)^2}, x \in [-2, 4];$$

$$p_{\eta}(y) = \frac{1}{9\pi} \sqrt{9 - y^2}, y \in [-3, 3]; F(1,3) = 1/2;$$

$$\text{б) } p_{\xi}(x) = \frac{1}{32\pi} \sqrt{64 - x^2}, x \in [-8, 8];$$

$$p_{\eta}(y) = \frac{1}{32\pi} \sqrt{64 - (y-3)^2}, y \in [-5, 11]; F(0,3) = 1/4;$$

$$в) p_{\xi}(x) = \frac{2}{25\pi} \sqrt{25 - (x-2)^2}, x \in [-3, 7];$$

$$p_{\eta}(y) = \frac{4}{25\pi} \sqrt{25 - y^2}, y \in [0, 5]; F(2, 5) = 1/2;$$

$$г) p_{\xi}(x) = \frac{2}{49\pi} \sqrt{49 - x^2}, x \in [-5, 5];$$

$$p_{\eta}(y) = \frac{2}{49\pi} \sqrt{49 - (y-5)^2}, y \in [-2, 12]; F(0, 5) = 1/4;$$

$$д) p_{\xi}(x) = \frac{4}{9\pi} \sqrt{9 - x^2}, x \in [0, 3];$$

$$p_{\eta}(y) = \frac{2}{9\pi} \sqrt{9 - (y-2)^2}, y \in [-1, 5]; F(1, 3) = 1/2.$$

25. Во всех случаях случайные величины ξ и η зависимы.

$$а) p_{\xi}(x) = \frac{10 - |x|}{100}, x \in [-10, 10];$$

$$p_{\eta}(y) = \frac{10 - |y-3|}{100}, y \in [-7, 13]; F(5, 8) = 3/4;$$

$$б) p_{\xi}(x) = \frac{6 - |x|}{36}, x \in [-6, 6];$$

$$p_{\eta}(y) = \frac{6 - |y-4|}{36}, y \in [-2, 10]; F(3, 7) = 3/4;$$

$$в) p_{\xi}(x) = \frac{4 - |x-1|}{16}, x \in [-3, 5];$$

$$p_{\eta}(y) = \frac{4 - |y|}{16}, y \in [-4, 4]; F(0, -1) = 1/16;$$

$$г) p_{\xi}(x) = \frac{6 - |x|}{36}, x \in [-6, 6];$$

$$p_{\eta}(y) = \frac{3 - |y|}{9}, y \in [-3, 3]; F(-1, -1) = 1/16;$$

$$д) p_{\xi}(x) = \frac{3 - |x|}{9}, x \in [-3, 3]; p_{\eta}(y) = \frac{6 - 2y}{9}, y \in [0, 3]; F(1, 1) = 7/18;$$

$$е) p_{\xi}(x) = \frac{2-x}{2}, x \in [0, 2]; p_{\eta}(y) = \frac{y}{2}, y \in [0, 2]; F(1, 3/2) = 1/2;$$

$$\text{ж) } p_{\xi}(x) = \frac{4 - |x - 1|}{16}, x \in [-3, 5]; p_{\eta}(y) = \frac{7 - y}{32}, y \in [-1, 7]; F(2, 5) = 21/32;$$

$$\text{з) } p_{\xi}(x) = \frac{6 - |x|}{36}, x \in [-6, 6]; p_{\eta}(y) = \frac{3 - |y - 3|}{9}, y \in [0, 6]; F(4, 4) = 13/18.$$

$$26. p_{\xi}(x) = \begin{cases} 0, & x < -1 \\ \frac{x+1}{7}, & -1 < x < 0 \\ 1/7, & 0 < x < 6 \\ \frac{7-x}{7}, & 6 < x < 7 \\ 0, & x > 7; \end{cases} \quad p_{\eta}(y) = \frac{8-y}{14}, y \in [0, 2]; F(1, 1) = 1/8.$$

Случайные величины ξ и η зависимы.

$$27. p_{\eta}(x) = \frac{3}{32}(4 - (x-2)^2), x \in [0, 4];$$

$$p_{\eta}(x) = \frac{3}{16}\sqrt{4-x}, y \in [0, 4]; F(2, 4) = 1/2.$$

Случайные величины ξ и η зависимы.

$$28. 1/32.$$

$$29. 15/16.$$

$$30. p_{\xi, \eta}(x, y) = 1/36 \text{ внутри трапеции, } p_{\xi}(x) = \begin{cases} 0, & x \notin [-6, 6] \\ \frac{x+6}{27}, & -6 < x < -3 \\ 1/9, & -3 < x < 3 \\ \frac{6-x}{27}, & 3 < x < 6; \end{cases}$$

$$p_{\eta}(y) = \begin{cases} 0, & y \notin [0, 4]; \\ \frac{8-y}{24}, & y \in [0, 4]. \end{cases}$$

Случайные величины ξ и η зависимы.

$$31. p_{\xi}(x) = 2(1-x), x \in [0, 1], p_{\eta}(y) = 2y, y \in [0, 1]; \text{случайные величины } \xi \text{ и } \eta \text{ зависимы; } M(\xi\eta) = 1/4.$$

$$32. p_{\xi}(x) = 2e^{2x}, x < 0, p_{\eta}(y) = -4ye^{2y}, y < 0, \text{случайные величины } \xi \text{ и } \eta \text{ зависимы.}$$

$$33. \text{а) } A = 1/2; \text{б) } F(x, y) = \frac{\sin x + \sin y - \sin(x+y)}{2};$$

в) $F_{\xi}(x) = \frac{1 - \cos x + \sin x}{2}$; $p_{\xi}(x) = \frac{\sin x + \cos x}{2}$; $x \in [0, \pi/2]$; случайная величина η распределена так же; г) $p_{\xi\eta}(x|y) = \frac{\sin(x+y)}{\sin y + \cos y}$, $x \in [0, \pi/2]$, $y \in [0, \pi/2]$.

34. а) $p(x, y) = 1 + \alpha(1 - 2x)(1 - 2y)$, $x, y \in [0, 1]$; $p_{\xi}(x) = 1$, $x \in [0, 1]$; $p_{\eta}(y) = 1$, $y \in [0, 1]$;

б) $\alpha \in [-1, 1]$; в) $\alpha/3$; г) $(2\alpha y - \alpha + 3)/6$.

35. $F_{p_{\xi}}(x|y) = \sqrt{x^2 - y^2}$, $x \in [y, \sqrt{y^2 + 1}]$.

36. а) $\phi_{\xi\eta}(y) = \begin{cases} \frac{3-7y}{5}, & -1 \leq y \leq 0; \\ \frac{2y+3}{5}, & 0 \leq y \leq 1; \end{cases}$ б) $\phi_{\eta\xi}(x) = \begin{cases} \frac{x}{4}, & 0 \leq x \leq 1; \\ \frac{6-5x}{4}, & 1 \leq x \leq 2. \end{cases}$

37. Равномерное распределение на отрезке $[0, 1]$.

38. а) поднимется на 2,8%; б) опустится на 1,4%.

39. а) 52; б) 44.

40. а) 24; б) 12,5.

Глава 8. Закон больших чисел. Центральная предельная теорема

Теоретические задачи

1. Везде — да, по теореме 5.

2. Выполняется.

3. Применить неравенство Чебышева к случайной величине

$$\xi = \frac{1}{n} \sum_{i=1}^n \xi_i.$$

4. Неравенство Чебышева дает оценки: $P\{|\xi - a| \geq \sigma\} \leq 1$, $P\{|\xi - a| \geq 3\sigma\} \leq 1/9 = 0,11$. Непосредственное вычисление этих вероятностей дает $P\{|\xi - a| \geq \sigma\} = 0,3174$, $P\{|\xi - a| \geq 3\sigma\} = 0,0027$.

$$5. M\xi = 0, \sigma = \frac{\sqrt{2}}{\lambda}; P\{|\xi| < \sigma\} = 1 - e^{-\sqrt{2}} = 0,75;$$

$P\{|\xi| < 3\sigma\} = 1 - e^{-3\sqrt{2}} = 0,98$. Неравенство Чебышева приводит к оценкам: $P\{|\xi| < \sigma\} \geq 0$; $P\{|\xi| < 3\sigma\} \geq 0,89$.

Вычислительные задачи

6. $\geq 0,4$. 7. а) $\leq 0,75$; б) $\geq 0,5$. 8. а) $\geq 0,5$; б) $\leq 2/3$. 9. $\leq 0,1$.
 10. $\leq 0,16$. 11. а) $\geq 0,89$; б) $\geq 0,93$. 12. $\geq 0,93$. 13. $\geq 0,91$. 14. Неравенство Чебышева дает оценку $n \geq 16\,000$, а следствие из интегральной формулы Муавра—Лапласа $n \geq 4356$. 15. $P(|m - np| \leq x) = 0,99$, отсюда $x \approx 2,57\sqrt{npq} \approx 103$. Величина m находится в пределах 8000 ± 103 по ЦПТ, и 8000 ± 400 по неравенству Чебышева. 16. $n \geq 294$.
 17. а) 0,92; б) 20 мин. 18. Не менее 8310 единиц. 19. 0,0287.
 20. От 237 до 363. *Указание:* применить ЦПТ, учитывая, что на каждую подвергнутую контролю деталь приходится геометрически распределенное ($p = 1/3$) число деталей, сошедших с конвейера.
 21. От 2304 до 2696. 22. 0,9177. 23. 245. 24. 0,94. 25. 0,9606.
 26. 0,9938. 27. 0,9876. 28. 0,7881. 29. 0,9649. 30. 0,9927.

Глава 9. Цепи Маркова

1. а) (0,0065; 0,0523; 0,9412); б) (1/28, 3/28, 6/7).
 2. Доля безработных — 1/9. Сократится в $17/9 \approx 1,89$ раза.
 3. (18/23, 3/23, 2/23).
 4. (0,079; 0,658; 0,263).
 5. (1/31, 20/31, 10/31).
 6. а) (143/400, 171/400, 86/400); б) (15/41, 18/41, 8/41).
 7. (0,5; 0,2; 0,3).
 8. (4/9, 13/54, 17/54).
 9. а) (25/54, 23/54, 1/9); б) (4/15, 42/75, 13/75).
 10. а) (1/8, 17/32, 11/32); б) (7/36, 5/9, 1/4).
 11. а) (0,027; 0,076; 0,217; 0,68); б) (0,125; 0,183; 0,242; 0,45).
 12. (1/9, 8/9).
 13. (1/169, 24/169, 144/169).
 14. (8/125, 36/125, 54/125, 27/125).
 15. 70,4%.
 16. а) 0,013; б) 0,554; в) 2,216; г) 1,001.
 17. 8/9.
 18. а) (0,282; 0,339; 0,271; 0,108); б) 0,49; в) 1,2.
 19. а) (0,00046; 0,00729; 0,05837; 0,31129; 0,62259); б) 0,00821; в) 0,44353.
 20. а) не более 7,32 мин.; б) не более 8,54 мин.
 21. а) 0,211; б) 0,237; в) 0,237.
 22. а) 1/2; б) 1/12.

Глава 10. Основные понятия и задачи математической статистики

1. а) $k = 9$; б) $k = 10$.

$$16. F(x) = \begin{cases} 0, & x \leq 1,2; \\ 0,2, & 1,2 < x \leq 3,7; \\ 0,4, & 3,7 < x \leq 4,8; \\ 0,6, & 4,8 < x \leq 5,1; \\ 0,8, & 5,1 < x \leq 9,2; \\ 1, & x > 9,2. \end{cases} \quad 17. F(x) = \begin{cases} 0, & x \leq 48,1; \\ 0,2, & 48,1 < x \leq 49,1; \\ 0,3, & 49,1 < x \leq 49,7; \\ 0,5, & 49,7 < x \leq 50,0; \\ 0,6, & 50,0 < x \leq 50,1; \\ 0,7, & 50,1 < x \leq 50,3; \\ 0,8, & 50,3 < x \leq 50,5; \\ 0,9, & 50,5 < x \leq 51,6; \\ 1, & x > 51,6. \end{cases}$$

Глава 11. Точечные оценки параметров законов распределения

Теоретические задачи

$$3. M(\xi - M\xi)^3 = M\xi^3 - 3M\xi^2 M\xi + 2(M\xi)^3; \quad \hat{\mu}_3 = \hat{\alpha}_3 - 3\hat{\alpha}_2 \bar{x} + 2\bar{x}^3;$$

$$M(\xi - M\xi)^4 = M\xi^4 - 4M\xi^3 M\xi + 6M\xi^2 (M\xi)^2 - 3(M\xi)^4;$$

$$\hat{\mu}_4 = \hat{\alpha}_4 - 4\hat{\alpha}_3 \bar{x} + 6\hat{\alpha}_2 \bar{x}^2 - 3\bar{x}^4.$$

$$10. \hat{\theta} = x_{\max} + \frac{1}{n}.$$

$$11. \hat{\theta} = \left(1 + \frac{1}{\beta n}\right) x_{\max}.$$

12. Оценка неэффективна, ее дисперсия бесконечна.

Вычислительные задачи

$$13. s^2 = 0,035.$$

$$14. а) \bar{x} = 10; б) \hat{\sigma}^2 = 2,5; s^2 = 3,33.$$

$$15. \bar{x} = 2504,88; s^2 = 254,41.$$

$$16. \bar{x} = 50,01; s^2 = 1,05; s = 1,02; x_{\text{med}} = 49,9; x_{\text{min}} = 48,1; x_{\text{max}} = 52.$$

$$17. \bar{x} = -0,4; s^2 = 128,85; X_{\text{mod}} = -2,5.$$

$$18. а) \bar{x} = 1,535; s^2 = 3,39, s = 1,84;$$

$$б) \bar{x} = 1653; s^2 = 446037, s = 667,86;$$

$$в) \bar{x} = 15,6; s^2 = 19,9; s = 4,38.$$

$$19. \text{В группе I: } \bar{x} = 96,17; s^2 = 30,49;$$

в группе II: $\bar{x} = 88,75$; $s^2 = 25,99$;

общие: $\bar{x} = 93,2$; $s^2 = 41,59$.

20. $\hat{a} = 1508,8$; $\sigma = 24$ (м).

21. $\hat{a} = 19,31$; $\sigma = 0,73$ (мм).

22. $\hat{a} = 11,33$ (мк).

23. $\bar{x} = 198,96$; $s^2 = 9,45$.

24. $\bar{x} = 71,11$; $s^2 = 264,52$; $s = 16,26$; $x_{\text{med}} = 69,3$; $x_{\text{min}} = 47,9$; $x_{\text{max}} = 88,6$.

25. $\bar{x} = 35,54$; $s^2 = 55,45$; $s = 7,45$; $x_{\text{med}} = 34,5$; $x_{\text{min}} = 27$; $x_{\text{max}} = 47$.

26. $\bar{x} = 15,85$; $s^2 = 4,69$; $s = 2,17$; $x_{\text{med}} = 15,45$; $x_{\text{min}} = 12,9$; $x_{\text{max}} = 19,4$.

27. $\bar{x} = 2127,46$; $s^2 = 7\,078\,286$; $s = 2660,5$; $x_{\text{med}} = 1134$; $x_{\text{min}} = 1000$; $x_{\text{max}} = 10\,358$.

28. $\bar{x} = 1148$; $s^2 = 18937,16$; $s = 137,61$; $x_{\text{med}} = 1105$; $x_{\text{min}} = 1000$; $x_{\text{max}} = 1426$.

29. $\bar{x} = 5,46$; $s^2 = 0,03$; $s = 0,17$.

30. $\bar{x} = 3,87$; $s^2 = 3,69$; $s = 1,92$; $x_{\text{mod}} = 4$.

31. $\bar{x} = 3,99$; $s^2 = 2,48$; $s = 1,57$; $x_{\text{mod}} = 4$.

32. $\bar{x} = 166$; $s^2 = 33,44$; $x_{\text{mod}} = 168$.

33. $\hat{a} = 2390$ (y.e.).

34. $\hat{a} = 2982,22$ (y.e.); $\sigma = 29,81$ (y.e.).

35. $\hat{a} = 3060,95$ (y.e.); $\sigma = 43,64$ (y.e.).

Глава 12. Функции и распределения в математической статистике

1. а) $\frac{15}{8}\sqrt{\pi}$; б) $\frac{105}{16}\sqrt{\pi}$; в) $\frac{3\pi}{256}$; г) $\frac{7\pi}{256}$.

6. Распределение Стьюдента с n степенями свободы.

9. Распределение Фишера—Снедекора $F(1, n)$.

15. Распределение Фишера—Снедекора $F(m, n)$.

16. Сходится к гамма-распределению $\gamma\left(\frac{n}{2}, \frac{n}{2}\right)$.

18. Распределение Фишера—Снедекора $F(2n, 2n)$.

21. $s^2 = \frac{1}{N-M} \sum_{i=1}^M (n_i - 1)s_i^2$, где $N = \sum_{i=1}^M n_i$.

24. $c = \sqrt{\pi/2}$; $e(\sigma^*) = \frac{1}{\pi-2} \approx 0,88$. Оценка неэффективна.

25. $c_n = \sqrt{\frac{n-1}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$. Оценка неэффективна, но асимптотически эффективна.

Глава 13. Методы построения оценок

Теоретические задачи

1. $\hat{\lambda} = \bar{x}$. 2. $\hat{p} = 1/\bar{x}$. 3. $\hat{\theta} = \bar{x} - 1$. 4. $\hat{\theta} = s, \hat{\mu} = \bar{x} - s$.
5. $\hat{\alpha} = \lambda \bar{x}, D\hat{\alpha} = \alpha/n$. 6. $\hat{\alpha} = \bar{x}^2/s^2, \hat{\lambda} = \bar{x}/s^2$. 7. $\hat{r} = \bar{x}, D\hat{r} = 2r/n$.
8. $\hat{r} = \frac{2s^2}{s^2 - 1}, r > 2$. 9. $\hat{\theta} = 2\bar{x}$. 10. $\hat{a} = \bar{x} - s\sqrt{3}, \hat{b} = \bar{x} + s\sqrt{3}$.
11. $\hat{c} = \bar{x}, \hat{d} = 3s^2$. 12. $\hat{\theta}_1 = \bar{x} - s\sqrt{3}, \hat{\theta}_2 = 2\sqrt{3}s$. 13. $\hat{\lambda}_{1,2} = \frac{2}{\bar{x} \pm \sqrt{2s^2 - \bar{x}^2}}$.
14. $\hat{\alpha} = \frac{\bar{x}}{\bar{x} - 1}, \alpha > 1$. 15. $\hat{\beta} = \frac{\bar{x}}{1 - \bar{x}}$. 16. а) $\hat{\beta} = \frac{\bar{x}}{\theta - \bar{x}}$; б) $\hat{\theta} = \frac{\beta + 1}{\beta} \bar{x}$.
17. $\hat{a} = 2\ln \bar{x} - \frac{1}{2} \ln \hat{\alpha}_2, \hat{\sigma}^2 = \ln \hat{\alpha}_2 - 2\ln \bar{x}$. 18. $\hat{\lambda} = \sqrt{2}/s$. 19. $\hat{p} = 1/\bar{x}$.
20. $\hat{\theta} = \bar{x} - 1$. 21. $\hat{p} = \bar{x}/N$. Оценка несмещенная, состоятельная.
22. $\hat{\theta} = x_{\max}$. Несмещенная оценка: $\tilde{\theta} = \frac{n+1}{n} x_{\max}$. 23. $\hat{\theta} = x_{\max}/2$.
24. $\hat{\theta} = \max(-x_{\min}, x_{\max}/2)$. 25. $\hat{\theta} = \max(1 - x_{\min}, x_{\max} - 1)$. 26. Оценкой максимального правдоподобия является любое число на интервале $(x_{\max} - 1, x_{\min})$. Несмещенная оценка — середина интервала: $\hat{a} = (x_{\max} + x_{\min} - 1)/2$.
27. $\hat{\lambda} = \left[\frac{1}{n} \sum_{i=1}^n |x_i| \right]^{-1}$. 28. $\hat{\theta} = x_{\min}$. 29. $\hat{\theta} = x_{\max}$.
30. $\hat{a} = \frac{1}{n} \sum_{i=1}^n g(x_i)$. 31. $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (g(x_i) - a)^2}$. 32. $\hat{\lambda} = \alpha/\bar{x}$. 33. $\hat{\mu} = \bar{x}$.
34. $\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i^2} \right)^{-1}$. 35. $\hat{\lambda} = \left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right]^{-1}$. 36. $\hat{\mu} = x_{\min}, \hat{\theta} = \bar{x} - x_{\min}$.

$$37. \hat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^n \ln x_i \right)^{-1}. 38. \hat{\beta} = - \left(\frac{1}{n} \sum_{i=1}^n \ln x_i \right)^{-1}. 39. а) \hat{\beta} = \left[\ln \theta - \frac{1}{n} \sum_{i=1}^n \ln x_i \right]^{-1};$$

$$б) \hat{\theta} = x_{\max}. 40. \hat{a} = \frac{1}{n} \sum_{i=1}^n \ln x_i; \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \hat{a})^2. 41. \hat{a} = x_{\min}, \hat{b} = x_{\max}.$$

$$\text{Несмещенные оценки: } \tilde{a} = x_{\min} - \frac{x_{\max} - x_{\min}}{n-1}; \tilde{b} = x_{\max} + \frac{x_{\max} - x_{\min}}{n-1}.$$

Вычислительные задачи

47. $\hat{\lambda} = 0,9$; $\hat{p}_0 = 0,407$, по таблице — 0,405. 48. $\hat{\lambda} = 0,2$; $\hat{T} = 11,51$ (часа). 49. На основе оценок методом моментов — 59%, по таблице — 58%. 50. На основе оценок методом моментов — 81%, по таблице — 82%. 51. $\hat{\lambda} = 1,535$. На основе оценки методом моментов — 46%, по таблице — 39%. 52. На основе оценок методом моментов — 63%, по таблице — 65%. 53. На основе оценок методом моментов — 16%, по доле в выборке — 20%. 54. $\hat{\alpha} = 3,95$, $\hat{\lambda} = 0,024$, $X_{\text{mod}} = 123$ (см). 55. $\hat{\alpha} = 4$, $\hat{\lambda} = 1/15$, $X_{\text{mod}} = 45$ (мин); б) $\hat{\alpha} = 9$, $\hat{\lambda} = 1/7$, $X_{\text{mod}} = 56$ (мин). 56. $\hat{\alpha} = 2$; доля жителей 4%. 57. $\hat{\alpha} = 5/4$; доля жителей 1/32. 58. $\hat{\mu} = 30$; $\hat{\theta} = 7$; $\hat{T} = 62,24$ (мин). 59. а) 0,86; 75 мин; б) 0,97; 56 мин. 60. $\hat{\theta}_1 = 20$; $\hat{\theta}_2 = 15$ (ч). 61. $\hat{a} = 8,156$; $\hat{\sigma}^2 = 0,003$; доля людей 92%. 62. $\hat{\theta} = 9,6$ (мин). 63. $\hat{a} = 82$; $\hat{b} = 118$ (кг); товара нужно 110,8 (кг). 64. а) 119,2 (кг); б) 153,8 (кг). 65. а) $\hat{\beta} = 9$, $\hat{p} = 0,075$; б) $\hat{\beta} = 4$, $\hat{p} = 0,1296$. 66. а) 0,17; б) 0,25. 67. $\pm 1,63\%$. 68. $\pm 6\%$. 69. 783 порции. 70. 45 ошибок. 71. $\hat{a} = 3,94$. Средний доход — 67 тыс. руб.; доля жителей 6,5%. 72. $\hat{\mu} = 30$; $\hat{\theta} = 4,7$; $\hat{T} = 51,64$ (мин). 73. а) 0,98; 53 мин; б) 0,86; 72 мин. 74. $\hat{a} = 8,244$; $\hat{\sigma}^2 = 0,059$; доля людей 29%. 75. $\hat{\theta} = 9,2$; $\hat{\theta} = 11,04$ (мин). 76. $\tilde{a} = 72$; $\tilde{b} = 128$ (кг); товара нужно 122,4 кг. 77. $\hat{y} = 866,27 + 16,8(x - 2000)$; $\hat{y}(2004) = 933,47$ (млрд кВт·ч). 78. а) $\hat{y} = 67,98 + 10,79(x - 2000)$; $\hat{y}(2003) = 100,35$ (млн т); б) $\hat{y} = 39,44 + 7,74(x - 2000)$; $\hat{y}(2003) = 62,66$ (млн т). 79. а) $\hat{y} = -7,26 + 0,69x$; $\hat{y}(90) = 54,84$ (млн т); б) $\hat{y} = 0,58x$; $\hat{y}(90) = 52,22$ (млн т). 80. $\hat{y} = 0,943x$; $\hat{y}(2600) \approx 2452$ (долл.).

Глава 14. Доверительные интервалы

Теоретические задачи

$$1. \left(\sqrt{\bar{x}} - \frac{u_\gamma}{2\sqrt{n}} \right)^2 < \lambda < \left(\sqrt{\bar{x}} + \frac{u_\gamma}{2\sqrt{n}} \right)^2, \text{ где } \Phi_0(u_\gamma) = \frac{\gamma}{2}.$$

$$2. 2\bar{x} \exp\left(-\frac{u_\gamma}{\sqrt{3n}}\right) < \theta < 2\bar{x} \exp\left(\frac{u_\gamma}{\sqrt{3n}}\right).$$

$$3. \frac{\alpha}{\bar{x}} \exp\left(-\frac{u_\gamma}{\sqrt{\alpha n}}\right) < \lambda < \frac{\alpha}{\bar{x}} \exp\left(\frac{u_\gamma}{\sqrt{\alpha n}}\right).$$

$$4. \left(\sqrt{\lambda \bar{x}} - \frac{u_\gamma}{2\sqrt{n}} \right)^2 < \alpha < \left(\sqrt{\lambda \bar{x}} + \frac{u_\gamma}{2\sqrt{n}} \right)^2.$$

$$5. \left(\sqrt{\bar{x}} - \frac{u_\gamma}{\sqrt{2n}} \right)^2 < r < \left(\sqrt{\bar{x}} + \frac{u_\gamma}{\sqrt{2n}} \right)^2.$$

$$6. v - u_\gamma \sqrt{\frac{v(1-v)}{mn}} < p < v + u_\gamma \sqrt{\frac{v(1-v)}{mn}};$$

$$\sin^2\left(\arcsin \sqrt{v} - \frac{u_\gamma}{2\sqrt{mn}}\right) < p < \sin^2\left(\arcsin \sqrt{v} + \frac{u_\gamma}{2\sqrt{mn}}\right), \text{ где } v = \frac{\bar{x}}{m}.$$

$$7. \bar{x} - u_\gamma \sqrt{\frac{\bar{x}(\bar{x}-1)}{n}} < \theta < \bar{x} + u_\gamma \sqrt{\frac{\bar{x}(\bar{x}-1)}{n}}.$$

$$8. 1 + \frac{1}{\bar{x}-1} \left(1 + u_\gamma \sqrt{\frac{\bar{x}}{n(2-\bar{x})}} \right)^{-1} < \alpha < 1 + \frac{1}{\bar{x}-1} \left(1 - u_\gamma \sqrt{\frac{\bar{x}}{n(2-\bar{x})}} \right)^{-1}.$$

$$9. \left(\frac{1}{\bar{x}} + \frac{u_\gamma}{\sqrt{n}} \right)^{-1} < \theta < \left(\frac{1}{\bar{x}} - \frac{u_\gamma}{\sqrt{n}} \right)^{-1}.$$

$$10. \left(\frac{1}{\bar{x}^2} + \frac{2u_\gamma}{\sqrt{n}} \right)^{-1/2} < \theta < \left(\frac{1}{\bar{x}^2} - \frac{2u_\gamma}{\sqrt{n}} \right)^{-1/2}.$$

Вычислительные задачи

$$11. 124,90 < a < 125,46 \text{ (r); } n = 96.$$

$$12. \text{ а) } 99,9 < a < 100,5; \text{ б) } 100,3 < a < 100,7; \text{ в) } 99,94 < a < 100,26 \text{ (r).}$$

$$13. \text{ а) } 99,97 < a < 100,43; \text{ б) } 100,09 < a < 100,71; \text{ в) } 99,97 < a < 100,03 \text{ (r).}$$

14. $\delta = 0,39$ (мм).
15. $n = 82$.
16. а) $269,79 < a < 303,21$ (у.е.); $n = 1245$;
б) $247,92 < a < 252,08$ (у.е.); $n = 1111$;
в) $277,55 < a < 282,45$ (у.е.); $n = 2401$.
17. $65,46 < a < 89,54$ (человек на фирме); $78\,552 < N < 107\,448$ (человек в отрасли).
18. $19\,250 < a < 25\,750$ (миль).
19. а) $24\,760 < a < 29\,240$; б) $29\,200 < a < 34\,800$ (км).
20. а) $4688,2 < a < 4811,8$; б) $4183,44 < a < 4316,56$;
в) $4665,24 < a < 4854,76$ (руб.)
21. $80,11 < a < 119,89$ (д.е.); $n = 792$.
22. $77,04 < a < 94,96$ (мин).
23. $49,53 < a < 50,49$ (г).
24. $0,43 < \sigma < 1,02$.
25. $72,98 < \sigma < 164,4$ (д.е.).
26. а) $66,01 < \sigma < 102,4$; б) $61,88 < \sigma < 96$ (д.е.).
27. а) $2,72 < \sigma^2 < 6,55$; б) $1,53 < \sigma^2 < 3,69$; в) $0,31 < \sigma^2 < 0,89$;
г) $0,10 < \sigma^2 < 0,29$ (г2).
28. а) $2,32 < \sigma < 4,31$; б) $3,18 < \sigma < 5,49$; в) $1,17 < \sigma < 2,09$;
г) $0,72 < \sigma < 1,23$; д) $0,39 < \sigma < 0,70$ (г).
29. $99,97 < a < 102,03$ (г); $5,93 < \sigma^2 < 15,65$ (г2).
30. От 220,5 до 276,5 д.е.
31. От 128,6 до 147,4 д.е.
32. $56,05 < a < 86,17$; $28,52 < x < 113,7$ (млн т).
33. $28,64 < a < 42,44$; $16,03 < x < 55,05$ (млн т).
34. $14,3 < a < 17,4$; $10,71 < a < 20,99$ (ц/га).
35. $21,92 < a_1 - a_2 < 38,08$ (д.е.).
36. $28,06 < a_1 - a_2 < 51,94$ (д.е.).
37. $0 < p < 0,03$.
38. $0,08 < p < 0,12$.
39. $0,55 < p < 0,65$; $1107 < N < 1293$.
40. $0,55 < p < 0,73$.
41. а) $0,67 < q < 0,83$; б) $0,84 < q < 0,91$; в) $0,72 < q < 0,88$;
г) $0,78 < q < 0,86$; д) $0,84 < q < 0,91$.

42. а) $0,10 < p < 0,14$; б) $0,12 < p < 0,16$; в) $0,16 < p < 0,20$;
 г) $0,11 < p < 0,15$.
43. От 6,2 до 11,8%.
44. $0,05 < p < 0,07$.
45. $0,01 < p < 0,03$; $n = 3012$ (записей).
46. $n = 10\ 551$.
47. $n = 1801$.
48. $n = 3659$.
49. $847,89 < \theta < 1179,39$ (ч).
50. а) $1360,0 < \theta < 1654,5$; б) $1813,3 < \theta < 2205,93$;
 в) $2262,6 < \theta < 2757,4$ (ч).
51. а) $251,9 < \lambda < 260,1$; б) $286,2 < \lambda < 291,8$ (покупателей).
52. $0,84 < \lambda < 0,96$ (сорняков).
53. $2,56 < \alpha < 3,80$. Этот пример показывает, что выборочное среднее плохо характеризует распределение Парето, и лучше использовать другие оценки.
54. а) $1,56 < \theta < 2,04$; б) $2,12 < \theta < 2,88$ (покупок)
55. а) $0,30 < p < 0,66$; б) $0,60 < p < 0,78$.

Глава 15. Проверка статистических гипотез

Теоретические задачи

- $S = \left\{ x: \frac{1}{\sigma_0^2} \sum_{i=1}^n (x_i - a)^2 > \chi_{\alpha, n}^2 \right\}$.
 - $S = \left\{ x: \bar{x} > p_0 + u_\alpha \sqrt{p_0(1-p_0)/n} \right\}$; $\Phi_0(u_\alpha) = 1/2 - \alpha$; $\Phi_0(u_\beta) = 1/2 - \beta$;

$$n = \frac{(u_\alpha \sqrt{p_0(1-p_0)} + u_\beta \sqrt{p_1(1-p_1)})^2}{(p_0 - p_1)^2}$$
.
 - $S = \left\{ x: \bar{x} > \lambda_0 + u_\alpha \sqrt{\lambda_0/n} \right\}$; $n = \frac{(u_\alpha \sqrt{\lambda_0} + u_\beta \sqrt{\lambda_1})^2}{(\lambda_0 - \lambda_1)^2}$.
 - $S = \left\{ x: \lambda_0 \sum_{i=1}^n x_i < \chi_{1-\alpha, 2n}^2 \right\}$.
 - В общем виде: $S = \left\{ x: |x|_{\max} > 1 \right\} \cup \left\{ x: \sum_{i=1}^n x_i^2 < C \right\}$.
- В частном случае: $S = \left\{ |x_1| > 1 \right\} \cup \left\{ |x_2| > 1 \right\} \cup \left\{ x_1^2 + x_2^2 < \frac{4}{\pi} \alpha \right\}$.

Вычислительные задачи

6. а) Мощность критерия 0,44; б) $n = 25$. 7. а) $\bar{x} > 19,3$. Нулевая гипотеза принимается; б) мощность критерия 0,6368. 8. а) $\bar{x} > 308,25$; б) $\beta = 0,3632$; в) $n = 82$. 9. $n = 237$. 10. Нулевая гипотеза отвергается. 11. Нулевая гипотеза отвергается. 12. Нулевая гипотеза отвергается. 13. а) нет; б) да; в) нет. 14. а) да; б) нет; в) да; г) нет. 15. а) да; б) нет; в) нет. 16. Нулевая гипотеза принимается. 17. Да. 18. а) да; б) нет; в) да; г) нет. 19. Да. 20. Нет. 21. а) да; б) нет; в) нет; г) да. 22. а) нет; б) нет; в) нет; г) да. 23. Да. 24. Да. 25. Да. 26. Да. 27. Да. 28. Нулевая гипотеза принимается. 29. Нет. 30. Да. 31. Нет. 32. а) да; б) нет; в) да; г) нет. 33. Нулевая гипотеза принимается. 34. Нулевая гипотеза отвергается. 35. Да. 36. Нет. 37. а) да; б) нет; в) нет. 38. Нет. 39. Да. 40. Да. 41. а) нет; б) нет; в) да; г) нет. 42. а) нулевая гипотеза принимается; б) нулевая гипотеза отвергается. 43. Да. 44. Нулевая гипотеза принимается. 45. Нулевая гипотеза принимается. 46. Нулевая гипотеза отвергается. 47. Нет. 48. Да. 49. а) да; б) нет; в) да; г) нет. 50. Нет. 51. Да. 52. а) да; б) нет; в) да. 53. Да. 54. Нулевая гипотеза отвергается. 55. Да. 56. а) нет; б) да; в) нет. 57. Да. 58. Нет. 59. Да. 60. Да. 61. Да. 62. Нулевая гипотеза отвергается. 63. а) да; б) нет; в) да. 64. а) да; б) нет; в) нет; г) да; д) нет. 65. Нет. 66. Нет, т.к. $n_4 < 4$. 67. а) Нулевая гипотеза принимается; б) 4,69. 68. Нет. 69. Да. 70. Да.

Глава 16. Критерии согласия

1. Да. 2. Да. 3. Нет. 4. Да. 5. Нет. 6. Нет. 7. Нет. 8. Нет. 9. Да. 10. Да. 11. Да. 12. Да. 13. Да. 14. Да. 15. Да.

Глава 17. Элементы анализа временных рядов

1. а) $\hat{tr}_x(t) = 2098 + 103,67t$, где t — номер месяца (от начала наблюдений); б) $\hat{tr}_x(t) = \exp(7,72 + 0,03t)$. Прогнозы на 2004 г. представлены в таблице:

Месяц	Прогноз (линейный)	Прогноз (экспонента)
Январь	4690	4770
Февраль	4793	4915
Март	4897	5064

Окончание табл.

Месяц	Прогноз (линейный)	Прогноз (экспонента)
Апрель	5001	5219
Май	5104	5378
Июнь	5208	5541
Июль	5312	5710
Август	5415	5884
Сентябрь	5519	6063
Октябрь	5622	6248
Ноябрь	5726	6438
Декабрь	5830	6634

2. $\hat{tr}_x(t) = \exp(8,2024 + 0,007t)$, где t — номер месяца (от начала наблюдений).

Месяц	Сезонные индексы, %	Прогноз
Июль	224	9725
Август	138	6042
Сентябрь	122	5362
Октябрь	60	2685
Ноябрь	90	4046
Декабрь	78	3532
Январь	56	2554
Февраль	34	1576
Март	59	2725
Апрель	76	3513
Май	108	5016
Июнь	208	9788

Здесь и далее сезонные индексы, как и прогнозируемые значения, округлены до целых.

3. $\hat{tr}_x(t) = 10\,836 + 39,285t$, где t — номер месяца (от начала наблюдений).

Месяц	Сезонные индексы, %
Январь	94
Февраль	89
Март	97
Апрель	97
Май	105
Июнь	105
Июль	106
Август	107
Сентябрь	100
Октябрь	102
Ноябрь	98
Декабрь	99

4. $\hat{tr}_x(t) = \exp(5,44 + 0,01t)$, где t — номер месяца (от начала наблюдений).

Месяц	Сезонные индексы, %
Январь	92
Февраль	87
Март	101
Апрель	98
Май	99
Июнь	114
Июль	128
Август	125
Сентябрь	107
Октябрь	92
Ноябрь	80
Декабрь	89

5. а) $\hat{tr}_x(t) = 815,87 + 16,8t$, где t — номер года (начиная с 1998).
Прогноз: 933,5;
б) $\hat{tr}_x(t) = \exp(6,7 + 0,02t)$. Прогноз: 934,5 (млрд кВт·ч).

Глава 18. Элементы линейного регрессионного и корреляционного анализа

1. а) $r = -0,18$; зависимости нет; б) $r = 0,24$; зависимость есть; в) $r = 0,01$; зависимости нет; г) $r = -0,05$; зависимости нет; д) $r = 0,61$; зависимость есть. 2. а) $r_s = -0,43$; б) $r_s = 0,07$; в) $r_s = -0,1$; г) $r_s = -0,18$; д) $r_s = 0,62$. 3. $r_s = 0,73$. Мнения экспертов достаточно хорошо согласованы. 4. $r = 0,90$; $r_s = 0,89$. 5. $r = 0,78$; $r_s = 0,76$.

Глава 19. Элементы дисперсионного анализа

1. Да. 2. Нет. $29,61 < a_1 < 62,39$; $69,62 < a_2 < 96,38$; $69,05 < a_3 < 108,95$. В Северном округе дела идут хуже, чем в Южном и Восточном (где примерно одинаково). 3. Да. 4. Да. 5. Да.

ЛИТЕРАТУРА

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. — М.: ЮНИТИ, 2001.
3. Бочаров П.П., Печинкин А.В. Теория вероятностей. Математическая статистика. — М.: Гардарика, 1998.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука, 1983.
5. Боровиков В. STATISTICA. Искусство анализа данных на компьютере. — СПб.: Питер, 2003.
6. Ван дер Варден Б. Математическая статистика. — М.: Высшая школа, 1975.
7. Венецкий Г.И., Кильдышев А.Н. Математическая статистика. — 3-е изд. — М.: Статистика, 1986.
8. Венцель Е.С. Теория вероятностей и математическая статистика. — М.: Мир, 1972.
9. Венцель Е.С. Исследование операций: задачи, принципы, методология. — М.: Наука, 1980.
10. Виленкин Н.Я., Виленкин А.Н., Виленкин П.А. Комбинаторика. — М.: МЦНМО, 2006.
11. Гнеденко Б.В. Курс теории вероятностей. — М.: Наука, 1960.
12. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1998.
13. Грачева М.В., Фадеева Л.Н., Черемных Ю.Н. Количественные методы в экономических исследованиях. — М.: ЮНИТИ, 2004.
14. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. — М.: Наука, 1989.
15. Дугерти К. Введение в эконометрику. — М.: ИНФРА-М, 2001.
16. Кибзун А.И., Горяинова Е.Р., Наумов А.В., Сиротин А.Н. Теория вероятностей и математическая статистика. — М.: Физматлит, 2002.
17. Колемаев В.А., Староверов О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1991.
18. Крамер Г. Математические методы статистики. — М.: Мир, 1975.
19. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
20. Кузнецов Ю.А. Оптимальное управление экономическими системами. — Изд-во Нижегородского госуниверситета, 2008.

21. Лукаш Е.Н., Медведев П.А., Пахомов В.Ф. Высшая математика. — М.: МГУ, 1987.
22. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. — М.: Дело, 1997.
23. Севастьянов В.А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.
24. Соколов Г.А., Чистякова Н.А. Теория вероятностей. Управляемые цепи Маркова в экономике. — М.: Физматлит, 2005.
25. Сулицкий В.Н. Методы статистического анализа в управлении. — М.: Дело, 2002.
26. Тутубалин В.Н. Теория вероятностей. — М.: Академия, 2008.
27. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. под ред. В.Э. Фигурнова. — М.: ИНФРА-М, 2003.
28. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Яценко И.В. Теория вероятностей и статистика. — М.: МЦНМО, 2004.
29. Тюрин Ю.Н., Макаров А.А., Симонова Г.И. Теория вероятностей. — М.: МЦНМО, 2009.
30. Фадеева Л.Н. Теория вероятностей и математическая статистика. — М.: Эксмо, 2006.
31. Фадеева Л.Н., Жуков Ю.В., Лебедев А.В. Математика для экономистов: Теория вероятностей и математическая статистика. Задачи и упражнения. — М.: Эксмо, 2006.
32. Фадеева Л.Н., Лебедев А.В. Теория вероятностей и математическая статистика. — М.: Эксмо, 2010.
33. Феллер В. Введение в теорию вероятностей и ее приложения. — М.: Мир, 1984.
34. Чистяков В.П. Курс теории вероятностей. — М.: Наука, 1982.
35. Ширяев А.Н. Вероятность. — М.: МЦНМО, 2007.
36. Ширяев А.Н. Задачи по теории вероятностей. — М.: МЦНМО, 2006.
37. Ширяев А.Н. Основы стохастической финансовой математики. — М.: ФАЗИС, 1998.
38. Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах. — М.: ЮНИТИ, 1999.
39. Эддаус М., Стэнсфилд Р. Методы принятия решения. — М.: ЮНИТИ, 1997.
40. Экономико-математические методы и прикладные модели / под ред. В.В. Федосеева. — М.: ЮНИТИ, 1999.

Учебное издание
НАЦИОНАЛЬНОЕ ЭКОНОМИЧЕСКОЕ ОБРАЗОВАНИЕ

**Фадеева Людмила Николаевна
Лебедев Алексей Викторович**

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебное пособие

Главный редактор *И. Федосова*
Ответственный редактор *В. Щербакова*
Художественный редактор *М. Левыкин*
Технический редактор *Н. Белова*
Компьютерная верстка *С. Терентьева*

ООО «Рид Групп»
119021, Москва, ул. Россолимо, д. 17, стр. 1, тел. 788-0075(76)

Свои пожелания и предложения по качеству и содержанию книг
Вы можете сообщить по эл. адресу editorial@readgroup.ru

Издание осуществлено при техническом содействии
ООО «Издательство АСТ»

Подписано в печать 30.06.2011. Формат 60х90/ 16.
Усл. печ. л. 31,0. Печать офсетная. Бумага писчая.
Тираж 2 000 экз. Заказ № 1494.

Отпечатано с готовых диапозитивов в типографии ООО «Полиграфиздат»
144003, г. Электросталь, Московская область, ул. Тевосяна, д. 25

