ВЫСШЕЕ ОБРАЗОВАНИЕ – БАКАЛАВРИАТ
серия основана в 1996 г.

И.В. БЕЛЬКО
И.М. МОРОЗОВА
Е.А. КРИШТАПОВИЧ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Допущено
Министерством образования Республики Беларусь
в качестве учебного пособия для студентов
учреждений высшего образования
по экономическим специальностям

Минск
«Новое знание»

Москва
«ИНФРА-М»

2016

УДК 519.2(075.8)
ББК 22.17я73
Б44

ФЗ № 436-ФЗ	Издание не подлежит маркировке в соответствии с п. 1 ч. 4 ст. 11

Рецензенты:
кафедра общей математики и информатики механико-математического факультета Белорусского государственного университета (зав. кафедрой — доктор физико-математических наук, профессор *В.А. Еровенко*);
зав. кафедрой высшей математики Белорусского государственного экономического университета, доктор физико-математических наук, профессор *М.П. Дымков*

Белько, И.В.

Б44 Теория вероятностей, математическая статистика, математическое программирование : учеб. пособие / И.В. Белько, И.М. Морозова, Е.А. Криштапович. — Минск : Новое знание ; М. : ИНФРА-М, 2016. — 299 с. : ил. — (Высшее образование: Бакалавриат).

ISBN 978-985-475-759-9 (Новое знание).
ISBN 978-5-16-011748-5 (ИНФРА-М, print).
ISBN 978-5-16-104278-6 (ИНФРА-М, online).

Рассмотрены темы, изучаемые в третьей части курса «Высшая математика»: теория вероятностей, математическая статистика и математическое программирование. Каждая глава содержит теоретические основы с примерами, задачи с решениями и задачи для самостоятельной работы, снабженные ответами. Большинство примеров и задач имеют экономическую направленность. Пособие может быть использовано для самостоятельной подготовки по данным разделам курса высшей математики.

Для студентов и преподавателей экономических специальностей высших учебных заведений.

УДК 519.2(075.8)
ББК 22.17я73

ISBN 978-985-475-759-9 (Новое знание)
ISBN 978-5-16-011748-5 (ИНФРА-М, print)
ISBN 978-5-16-104278-6 (ИНФРА-М, online)

© Белько И.В., Морозова И.М., Криштапович Е.А., 2016
© ООО «Новое знание», 2016

Оглавление

Предисловие 8

ЧАСТЬ I. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Глава 1. События и вероятности 13
1.1. Элементы комбинаторики 13
1.2. События 16
1.3. Понятие вероятности 17
1.4. Действия над событиями 21
1.5. Теорема сложения вероятностей 24
1.6. Условные вероятности, теорема умножения вероятностей, независимость событий 27
1.7. Формула полной вероятности и формула Байеса 30
1.8. Аксиоматическое построение теории вероятностей 33
Задачи с решениями 36
Задачи для самостоятельного решения 43
Ответы 46

Глава 2. Схема испытаний Бернулли 47
2.1. Формула Бернулли 47
2.2. Формула Пуассона 50
2.3. Локальная формула Муавра — Лапласа 50
2.4. Интегральная формула Муавра — Лапласа 51
Задачи с решениями 52
Задачи для самостоятельного решения 55
Ответы 57

Глава 3. Случайные величины и законы их распределения 58
3.1. Общие сведения о случайных величинах 58
3.2. Дискретные случайные величины и операции над ними 60
3.2.1. Операции над случайными величинами 62
3.2.2. Математическое ожидание и дисперсия дискретной случайной величины 63
3.2.3. Функция распределения дискретной случайной величины 65
3.3. Непрерывные случайные величины 67
3.4. Математическое ожидание и дисперсия непрерывной случайной величины 69

Задачи с решениями 70
Задачи для самостоятельного решения 78
Ответы 81

Глава 4. Законы распределения случайных величин 83
4.1. Законы распределения дискретных случайных величин 83
4.1.1. Постоянная случайная величина (вырожденное распределение) 83
4.1.2. Альтернативная случайная величина (распределение Бернулли) 83
4.1.3. Биномиальное распределение 84
4.1.4. Геометрическое распределение 85
4.1.5. Гипергеометрическое распределение 85
4.1.6. Закон Пуассона 86
4.2. Законы распределения непрерывных случайных величин 88
4.2.1. Равномерное распределение $R[a, b]$ 88
4.2.2. Показательное (экспоненциальное) распределение 89
4.2.3. Нормальное распределение 91
4.2.4. Логнормальное распределение 94
Задачи с решениями 96
Задачи для самостоятельного решения 100
Ответы 103

Глава 5. Многомерные случайные величины 104
5.1. Дискретные двумерные случайные величины 105
5.2. Функция распределения двумерной случайной величины 107
5.3. Непрерывные двумерные случайные величины 108
5.4. Независимые случайные величины. Ковариация, коэффициент корреляции 109
5.5. Двумерное нормальное распределение 112
Задачи с решениями 113
Задачи для самостоятельного решения 117
Ответы 118

Глава 6. Закон больших чисел 120
6.1. Неравенства Чебышева и Маркова 120
6.2. Закон больших чисел 121
6.3. Теорема Бернулли 122
Задачи для самостоятельного решения 123
Ответы 124

ЧАСТЬ II. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Глава 7. Вариационные ряды и их характеристики 128
7.1. Генеральная и выборочная совокупности 128
7.2. Понятие вариационного ряда и его графические изображения 129
7.3. Числовые характеристики вариационных рядов 136
Задачи с решениями 138
Задачи для самостоятельного решения 141
Ответы 143

Глава 8. Оценка параметров генеральной совокупности 144
8.1. Точечные оценки параметров 144
8.2. Основные статистические распределения 150
8.3. Интервальные оценки параметров 152
8.4. Доверительные интервалы для генеральной средней и генеральной доли выборки 152
8.5. Объем выборки 155
8.6. Доверительный интервал для дисперсии 156
Задачи с решениями 157
Задачи для самостоятельного решения 161
Ответы 163

Глава 9. Проверка статистических гипотез 164
9.1. Основные понятия 164
9.2. Гипотезы о значениях числовых характеристик 168
9.2.1. Гипотеза о численной величине среднего значения ... 168
9.2.2. Гипотеза о числовом значении дисперсии 170
9.2.3. Гипотеза о числовом значении доли признака 171
9.3. Проверка гипотезы о равенстве средних значений 172
9.4. Критерии согласия 174
Задачи с решениями 176
Задачи для самостоятельного решения 180
Ответы 182

Глава 10. Однофакторный дисперсионный анализ 184
10.1. Условия применения 184
10.2. Разложение суммы квадратов отклонений 185
10.3. Проверка гипотезы о равенстве групповых средних 186
10.4. Коэффициент детерминации 187

10.5. Сводка формул ... 188
Задачи с решениями ... 189
Задачи для самостоятельного решения ... 191
Ответы ... 192

Глава 11. Корреляционный анализ ... 193
11.1. Представление данных в корреляционном анализе ... 194
11.2. Коэффициент корреляции ... 195
11.3. Корреляционное отношение ... 197
Задачи с решениями ... 198
Задачи для самостоятельного решения ... 200
Ответы ... 201

Глава 12. Регрессионный анализ ... 202
12.1. Корреляционное поле ... 203
12.2. Линейная регрессия ... 203
12.3. Статистический анализ уравнения регрессии ... 205
Задача с решением ... 206
Задачи для самостоятельного решения ... 208
Ответы ... 208

ЧАСТЬ III. МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Глава 13. Классические задачи линейного программирования ... 211
13.1. Задача о планировании производственной программы предприятия ... 211
13.2. Задача об оптимальной корзине продуктов (задача о диете) ... 214
13.3. Задача оптимального раскроя материалов ... 216
13.4. Транспортная задача ... 218
13.5. Задача о назначениях ... 219

Глава 14. Основные понятия задач линейного программирования и графический метод их решения ... 221
14.1. Постановка общей задачи линейного программирования ... 221
14.2. Формы записи задач линейного программирования ... 222
14.3. Базисные и свободные переменные. Опорный план ... 223
14.4. Графическое решение задачи линейного программирования, сводимой к двум переменным ... 224

Задачи с решениями ... 227
Задачи для самостоятельного решения ... 231
Ответы ... 233

Глава 15. Симплексный метод. Теория двойственности ... 234
15.1. Симплексный метод решения ЗЛП ... 234
15.2. Двойственность в линейном программировании ... 243
15.2.1. Понятия анализа устойчивости и чувствительности ... 243
15.2.2. Понятие двойственности ... 246
15.3. Основные теоремы теории двойственности и их экономический смысл ... 250
15.4. Информационные технологии в решении задач математического программирования ... 255
Задачи для самостоятельного решения ... 265
Симплексный метод ... 265
Двойственность в линейном программировании ... 268
Ответы ... 269

Глава 16. Транспортная задача ... 270
16.1. Постановка транспортной задачи ... 270
16.2. Построение начального опорного плана ... 273
16.3. Проверка опорного решения на оптимальность ... 278
16.4. Переход к следующему распределению поставок ... 280
Задача с решением ... 283
Задачи для самостоятельного решения ... 289
Ответы ... 292

Приложения ... 293

Литература ... 298

Предисловие

Теория вероятностей, математическая статистика и математическое программирование занимают важное место в математической подготовке экономистов. Теория вероятностей и математическая статистика служат основой для последующего изложения эконометрики, а третий раздел является введением в теорию экономико-математических методов и математических моделей.

Конечно, для глубокого и полного усвоения предмета при изложении со всей строгостью необходима серьезная математическая подготовка. Имеется достаточное количество пособий с таким уровнем строгости, хотя, согласно прежним стандартам, отсутствуют учебники, включающие совместно все три названных раздела курса.

Данное учебное пособие соответствует программе предмета, принятой в новом образовательном стандарте, и состоит из трех компонентов: теоретические основы с примерами, задачи с решениями и задачи для самостоятельной работы, снабженные ответами. Таким образом, пособие может одновременно служить и учебником, и сборником задач. Теоретическая часть пособия содержит основные понятия и свойства, приведенные с возможной полнотой и выделением идейных основ, насколько это позволяет принятое в пособии интуитивно-наглядное изложение. Для усвоения этого материала достаточно базовой математической подготовки.

Важным дополнением к теории, которое облегчает ее понимание, служат примеры, приведенные в пособии.

Особая роль отводится задачам с решениями. Для их усвоения достаточно сведений из теоретической части. Они позволяют глубже понять теоретический материал и получить навыки самостоятельного решения задач. Большинство приводимых задач и примеров имеет экономическую направленность. Набор задач для самостоятельной работы достаточен для практических занятий и домашних заданий. Отметим, что несоответствие в точности числовых данных при округлении вызвано тем, что в используемых таблицах данные приводятся с точностью до 0,0001. Такая высокая точность вряд ли уместна во многих вычислениях.

Содержание пособия традиционно и адаптировано к уровню подготовки студентов.

В I части после краткого знакомства с элементами комбинаторики на интуитивном уровне вводятся основные понятия —

событие и вероятность. Свойства операций над событиями и их связь с вероятностями интерпретируются на основе классической вероятности. Для полноты изложения и восприятия основ теории и свойств вероятностей приводится аксиоматика вероятностного пространства. Хотя, в принципе, для принятого уровня при отсутствии доказательств глубоких результатов саму аксиоматику можно опустить. При этом нужно все же учесть, что у читателей должно сформироваться представление о вероятностном пространстве. Основными примерами для этого служат классическая схема случаев и схема Бернулли (повторных независимых испытаний). Для вероятностей событий из схемы Бернулли даются формула Бернулли, а также приближенные формулы: Пуассона, локальная и интегральная формулы Муавра — Лапласа. Центральное место в I части пособия отводится случайным величинам. Рассматриваются два типа случайных величин: дискретные и непрерывные, приводятся способы их задания и их числовые характеристики: математическое ожидание и дисперсия. В качестве примеров рассматриваются следующие распределения случайных величин: постоянная случайная величина, альтернативная случайная величина, биномиальное, геометрическое и гипергеометрическое распределения, распределение Пуассона, равномерное, показательное и нормальное распределения. Рассмотрение многомерных случайных величин ограничено двумерным случаем. Вводятся ковариация и коэффициент корреляции пары случайных величин. Закон больших чисел приводится в упрощенных формулировках теорем Чебышева и Бернулли, пригодных для решения задач.

Во II части, посвященной математической статистике, вводятся основные понятия выборочного метода, повышенное внимание уделено заданию выборки в виде случайного вектора или его реализации. Строятся выборочные характеристики генеральных совокупностей, даются точечные и интервальные оценки их параметров. Описываются этапы проверки статистических гипотез, которые затем иллюстрируются гипотезами о средних значениях, дисперсии и доле признака. Приводится критерий Пирсона для проверки гипотезы о законе распределения случайной величины. Оценка зависимости между переменными (случайными и детерминированными) проводится на основе дисперсионного, корреляционного и регрессионного анализа.

III часть пособия является по существу введением в математическое программирование и содержит примеры задач линейного программирования и методы их решения. Графический метод решения задач линейного программирования с двумя переменными и сводящихся к ним задач служит интерпретацией идеи общего симплекс-метода. На основе симплекс-метода и соответствующих ему таблиц сначала проводится разбиение переменных на базисные и свободные, а затем находится оптимальное значение целевой функции. Анализ решения проводится на основе теории двойственности. Отдельно выделяется метод потенциалов для решения транспортной задачи. Приводятся другие типы задач математического программирования.

С учетом требований учебной программы в пособии даются решения задач линейного программирования с применением компьютерного пакета прикладных программ Excel.

Авторы выражают благодарность и признательность заведующему кафедрой высшей математики Белорусского государственного экономического университета Михаилу Пахомовичу Дымкову, коллективу кафедры высшей математики и информатики Белорусского государственного университета и ее заведующему Валерию Александровичу Еровенко. Их скрупулезное прочтение и замечания по содержанию рукописи значительно улучшили изложение материала пособия.

ЧАСТЬ I

Теория вероятностей

- События и вероятности
- Схема испытаний Бернулли
- Случайные величины и законы их распределения
- Законы распределения случайных величин
- Многомерные случайные величины
- Закон больших чисел

В этой части приведено краткое изложение основ теории вероятностей на интуитивно-наглядном уровне. Дана также аксиоматика вероятностей. Элементы комбинаторики служат базой для решения задач классического раздела теории. Для вводимых понятий и свойств приводятся примеры задач и задания для самостоятельной работы.

Особое внимание уделяется случайным величинам, которые составляют основу математической статистики.

Из всего их многообразия выделены основные величины, которые рассматриваются в классической теории вероятностей. Для многомерных случайных величин описывается только двумерный случай, в частности, двумерная нормальная величина. Достаточно полно введено понятие независимости случайных величин и их числовые характеристики.

1

СОБЫТИЯ И ВЕРОЯТНОСТИ

Для вычисления вероятностей в начале главы вводятся элементы комбинаторики. Связи между операциями над событиями и их вероятностями описываются теоремами сложения и умножения вероятностей. Для случая, когда пространство событий разбивается гипотезами, рассматриваются формула полной вероятности и формула Байеса.

1.1. Элементы комбинаторики

Комбинаторика — это раздел математики, в котором изучаются различные комбинации подмножеств в конечных множествах. Элементы комбинаторики будут использоваться при вычислениях классических вероятностей.

Упорядоченным множеством называется множество, в котором указан порядок следования его элементов. Так, например, множества $\{a, b, c\}$ и $\{a, c, b\}$ есть различные упорядоченные множества.

Сформулируем основные правила комбинаторики.

1. Правило суммы

Пусть из множества A элемент a_1 можно выбрать n_1 способами, элемент a_2 — другими n_2 способами и т.д., элемент a_k – n_k способами, отличными от предыдущих. Тогда выбор одного из элементов a_1, или a_2, ..., или a_k можно произвести $n_1 + n_2 + ... + n_k$ способами.

Пример 1.1. Пусть в корзине имеется 7 апельсинов, 5 бананов и 10 яблок. Тогда выбор одного из фруктов (или апельсина, или банана, или яблока) можно сделать 22 способами (22 = 7 + 5 + 10).

Характерным признаком использования правила суммы служит предлог «или».

2. Правило произведения

Пусть A — некоторое множество, из которого выбор элемента a_1 можно осуществить n_1 способами, после этого выбор элемента a_2

можно осуществить n_2 способами и т.д., наконец, после выбора элемента a_{k-1} элемент a_k можно выбрать n_k способами. Тогда одновременный выбор элементов $a_1, a_2, ..., a_k$ в указанном порядке можно осуществить $n_1 \cdot n_2 \cdot ... \cdot n_k$ способами.

Пример 1.2. Пусть в спортивном велосипеде имеются 3 ведущие звездочки и 4 ведомые. Сколько передач имеется в велосипеде?

Так как каждая передача определяется выбором одной ведущей и одной ведомой звездочки, то число всех передач совпадает с числом выборов одного элемента из 3 и другого элемента из 4. Поэтому число передач равно $3 \cdot 4 = 12$.

Характерным признаком использования правила произведения служит предлог «и».

Пусть некоторое множество A содержит n различных элементов. *Размещением из n элементов по k элементов* называется каждое его упорядоченное подмножество, состоящее из k элементов.

Через A_n^k обозначают число всех размещений из n элементов по k (читается: «A из n по k»).

Пример 1.3. Пусть в группе 27 студентов. Любые трое из них, расположенные в порядке возрастания номеров их зачетных книжек, образуют размещение из 27 по 3.

Пример 1.4. Перечислим все размещения из трех элементов $\{a, b, c\}$ по два: (a, b), (b, a), (a, c), (c, a), (b, c), (c, b). Их количество равно 6, т.е. $A_3^2 = 6$.

Для числа размещений справедлива формула

$$A_n^k = n(n-1)(n-2)...\big(n-(k-1)\big). \qquad (1.1)$$

Доказательство этой формулы легко получается по правилу произведения: первый элемент a_1 можно выбрать n способами, после этого для выбора элемента a_2 остается $(n-1)$ способ и т.д. Для последнего k-го элемента a_k остается $n-(k-1)$ способ.

Например, согласно формуле (1.1) число размещений из 27 по 3

$$A_{27}^3 = 27 \cdot 26 \cdot 25 = 17\,550.$$

Для сокращения записи произведение всех натуральных чисел от 1 до n обозначается $n!$ (n факториал), т.е. $n! = 1 \cdot 2 \cdot 3 \cdots n$. При этом условно считается, что $0! = 1$. Применяя это обозначение, после умножения и деления правой части равенства (1.1) на $(n-k)!$ получим

$$A_n^k = \frac{n!}{(n-k)!}. \tag{1.2}$$

Размещение из n элементов по n называется *перестановкой из n элементов*. Количество всех таких перестановок обозначается P_n. Из предыдущей формулы получим

$$P_n = A_n^n = n!. \tag{1.3}$$

Пример 1.5. Пусть в группе 27 студентов, расположение их в порядке возрастания номеров зачетных книжек является перестановкой из 27 элементов. Количество всех перестановок из 27 элементов равно $27! = 1 \cdot 2 \cdot 3 \cdots 27 = 10\,888\,869\,450\,418\,352\,160\,768\,000\,000$.

Пример 1.6. Перечислим все перестановки из трех элементов $\{a, b, c\}$: (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a). Их количество равно $P_3 = 3! = 1 \cdot 2 \cdot 3 = 6$.

Пусть задано множество, состоящее из n элементов. *Сочетанием из n элементов по k элементов* называется каждое его (неупорядоченное) подмножество, содержащее k элементов.

Пример 1.7. Любые трое студентов из группы в 27 человек образуют сочетание из 27 по 3.

Число всех сочетаний из n элементов по k элементов обозначается C_n^k («C из n по k»). Для числа сочетаний справедлива формула

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{(n-k)!\cdot k!} = \frac{n(n-1)(n-2)\cdots(n-k-1)}{k!}. \tag{1.4}$$

Для их доказательства заметим, что выбору сочетания из n по k (так как выбираются подмножества из k элементов) отвечает выбор $k!$ размещений из n по k (т.е. упорядоченных подмножеств из этих элементов). Поэтому справедливо соотношение

$$A_n^k = k!C_n^k.$$

Например, число сочетаний из 27 элементов по 3

$$C_{27}^{3} = \frac{27 \cdot 26 \cdot 25}{1 \cdot 2 \cdot 3} = 2925.$$

Таким будет количество троек элементов во множестве, состоящем из 27 элементов.

Число C_n^k обладает свойством

$$C_n^k = C_n^{n-k}, \tag{1.5}$$

справедливость которого следует из того, что после выбора подмножества из k элементов ($k < n$) в нем остается $n - k$ элементов и таких подмножеств одинаковое количество. Случаю $k = n$ соответствует выбор пустого подмножества:

$$C_n^0 = C_n^n = 1.$$

1.2. События

Построение основ теории вероятностей можно проводить формально-логическим аксиоматическим методом. Аксиоматику теории вероятностей, предложенную в начале XX века А.Н. Колмогоровым, мы приведем ниже. В своем изложении мы будем, как правило, оставаться на интуитивно-наглядном уровне, который позволит глубже освоить практические приложения теории. По этой причине основные определения носят скорее описательный характер, без формализованного уровня строгости.

Эксперимент, испытание, опыт, процесс — это возникновение или преднамеренное создание определенного комплекса условий, результатом которого является тот или иной исход.

Применение этого понятия в теории вероятностей требует наличия определенных закономерностей в появлении исхода испытаний.

Событием называется исход испытания. Это понятие является первичным в теории вероятностей. События обозначаются большими латинскими буквами A, B, C, ...

Среди событий выделяются достоверные и невозможные. *Достоверное* событие — это такое событие, которое всегда происходит

при выполнении данного комплекса условий. Оно обозначается Ω. *Невозможное* событие ∅ — это такое событие, которое не может произойти при выполнении комплекса условий.

Примеры событий: выпадение орла при бросании монеты, выигрыш по облигации, увеличение курса доллара в следующем месяце, появление заявки на телефонной станции и т.д.

Элементарными называются те из событий, которые нельзя разложить на составляющие их события. Элементарные события будем обозначать буквой ω. Любое событие A из пространства Ω можно составить из элементарных событий.

Пример 1.8. В опыте с бросанием игральной кости (кубика) элементарными событиями являются выпадения чисел 1, 2, 3, 4, 5, 6. В этом же опыте событиями являются выпадения четного или нечетного числа.

Пространство событий Ω, соответствующее определенному комплексу условий (эксперименту), удобно представлять в виде некоторого множества, элементы которого представляют элементарные события, а его подмножества — события (рис. 1.1).

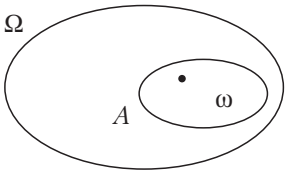


Рис. 1.1. События

1.3. Понятие вероятности

Наряду с понятием события вторым основным понятием теории вероятностей является понятие вероятности события. В этом параграфе мы приведем основные подходы к этому понятию.

Представление о случайности событий связано с невозможностью предсказать заранее исход того или иного эксперимента. Однако, например, при многократном бросании монеты выясняется, что примерно в половине случаев выпадает герб. При исследовании большого количества одинаковых испытаний обнаруживаются определенные закономерности, которые можно описать, используя понятие вероятности.

Под ***вероятностью события*** понимается некоторая числовая характеристика возможности наступления этого события. Существует несколько подходов к определению вероятности.

Статистическое определение вероятности

Пусть при проведении n испытаний некоторое событие A появилось m раз. Многочисленные эксперименты такого рода показывают, что при больших n отношение $\frac{m}{n}$, называемое ***частостью*** события A, остается примерно постоянным.

При статистическом определении ***вероятностью*** события A называется постоянная величина, вокруг которой колеблются значения частостей при неограниченном возрастании числа n.

Пример 1.9. Английский ученый Пирсон произвел 23 000 бросаний монеты. При этом герб появился 11 512 раз. Значит, частость появления герба равна

$$\frac{11\,512}{23\,000} = 0{,}5005.$$

Этот пример показывает, что за вероятность появления герба можно взять число 0,5.

Классическое определение вероятности

Классической схемой, или *схемой случаев*, называется испытание, при котором число элементарных исходов конечно и все они равновозможны.

Элементарное событие (исход) ω называется *благоприятствующим* событию A, если наступление ω влечет наступление A (т.е. ω входит в число элементов, составляющих A).

Например, при бросании игральной кости выпадение числа 2 является благоприятствующим выпадению четного числа очков.

Классической вероятностью события A, принадлежащего пространству событий Ω классической схемы, называется отношение числа m элементарных событий, благоприятствующих событию A, к числу n всех элементарных событий из этой схемы:

$$P(A) = \frac{m}{n}. \tag{1.6}$$

Из определения классической вероятности следует, что

$$P(\varnothing) = 0, \quad P(\Omega) = 1, \quad 0 \le P(A) \le 1. \tag{1.7}$$

Эти свойства вероятности входят в аксиоматику теории вероятностей.

Пусть пространство Ω состоит из n элементарных событий $\omega_1, \omega_2, ..., \omega_n$. Тогда для каждого ω_i благоприятным исходом будет только само ω_i и для него $m = 1$. Поэтому

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

Таким образом, если в классической схеме пространство Ω состоит из n элементарных равновозможных событий, то вероятность каждого из них равна $\frac{1}{n}$.

Пример 1.10. В опыте с бросанием игральной кости число всех исходов n равно 6 и все они равновозможны. Пусть событие A означает появление четного числа. Тогда для этого события благоприятными исходами будут появления чисел 2, 4, 6. Их количество m равно 3. Поэтому вероятность события A

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

Пример 1.11. В лифт девятиэтажного дома входят 4 человека. Какова вероятность того, что они выйдут на разных этажах?

Согласно классическому определению вероятности,

$$P = \frac{m}{n},$$

где число n — число всех возможных способов, m — число способов, благоприятствующих событию.

У первого пассажира лифта есть 8 возможностей выйти (на этажах 2–9), тогда у второго остается 7 возможностей, у третьего — 6 и у четвертого — 5. Значит, $m = A_8^4$. Так как каждый пассажир лифта может выйти на любом из этажей, то число всех возможностей по правилу произведения равно 8^4. Поэтому

$$P = \frac{A_8^4}{8^4} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{8^4} \approx 0{,}41.$$

Геометрическое определение вероятности

Для классической схемы пространство событий состоит из конечного числа равновозможных событий. Геометрическая вероятность является обобщением классической.

Пусть пространство событий Ω представляется в виде некоторого подмножества на евклидовой плоскости и пусть $S(A)$ означает

площадь подмножества $A \subset \Omega$ (мы оставляем в стороне вопросы, связанные с понятием и свойствами площади). ***Геометрической вероятностью*** события $A \subset \Omega$ называется отношение

$$P(A) = \frac{S(A)}{S(\Omega)}.$$

Аналогично определяется геометрическая вероятность для пространства событий на прямой или в евклидовом пространстве. В этом случае площадь заменяется длиной или объемом соответственно.

Пример 1.12. Два приятеля договорились о встрече между полуднем и часом дня. Каждый из них приходит на встречу случайным образом в указанном промежутке времени. Пришедший первым ждет не более 15 мин. С какой вероятностью они встретятся?

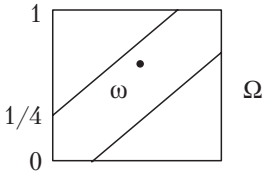


Рис. 1.2. Задача о встрече

Обозначим через x и y время прихода приятелей ($0 \le x \le 1$, $0 \le y \le 1$). Тогда пространство событий можно представить единичным квадратом, каждая точка которого (x, y) задает время прихода приятелей (рис. 1.2). Встреча приятелей состоится, если модуль разности времени их приходов не превышает 15 мин, или $1/4$ ч: $|x - y| \le 1/4$.

Указанное неравенство равносильно условию

$$-\frac{1}{4} \le x - y \le \frac{1}{4},$$

или системе

$$\begin{cases} y \ge x - \dfrac{1}{4}, \\ y \le x + \dfrac{1}{4}. \end{cases}$$

Имеем

$$S(\Omega) = 1, \quad S(A) = 1 - 2 \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} = \frac{7}{16}.$$

Ответ: $P(A) = \dfrac{7}{16}$.

1.4. Действия над событиями

Представление событий в виде подмножеств некоторого множества Ω позволяет перенести операции над подмножествами на операции над событиями.

Событие A влечет событие B, если из наступления события A следует наступление события B. Этот факт обозначается $A \subset B$. В терминах элементарных событий это означает, что каждое элементарное событие, входящее в A, входит также и в B.

Пример 1.13. Пусть производится выбор одного из чисел от 1 до 100. Пусть событие B означает выбор четного числа, событие A — выбор числа, кратного 10. Тогда $A \subset B$, так как каждое число, кратное 10, является четным.

Равенство событий $A = B$ означает, что $A \subset B$ и $B \subset A$, т.е. они состоят из одних и тех же элементарных событий.

Пример 1.14. Пусть производится выбор одного из чисел от 1 до 100. Если событие B — выбор числа, кратного 3, а событие A — это выбор числа, сумма цифр которого делится на 3, то $A = B$ по признаку делимости на 3.

Суммой событий A и B называется такое событие $C = A + B$, которое означает наступление или A, или B, т.е. хотя бы одного из них.

Пример 1.15. Пусть при выборе одного из чисел от 1 до 13 событие A означает выбор четного числа (т.е. событие A состоит из элементарных событий — выборов одного из чисел 2, 4, 6, 8, 10, 12), событие B — выбор числа, кратного 3 (т.е. событие B состоит из элементарных событий — выборов одного из чисел 3, 6, 9, 12). Тогда событие $A + B$ означает выбор числа, кратного 2 или 3, т.е. событие $A + B$ состоит из чисел 2, 3, 4, 6, 8, 9, 10, 12.

Произведением событий A и B называется событие, состоящее в совместном наступлении и A, и B, и обозначается $C = A \cdot B$.

Несовместными называются события A и B, произведение которых является невозможным событием, т.е. $A \cdot B = \varnothing$.

Дополнением к событию A называется событие $\overline{A}$, которое можно определить равенствами $A \cdot \overline{A} = \varnothing$ и $A + \overline{A} = \Omega$.

Разностью событий A и B называется событие, состоящее в наступлении A и ненаступлении B. Разность A и B обозначается $A - B$ и характеризуется условием $A - B = A \cdot \overline{B}$.

Пример 1.16. Пусть при выборе одного из чисел от 1 до 13 событие A означает выбор четного числа, событие B — выбор числа, кратного 3. Тогда событие $A \cdot B$ означает выбор числа, кратного и 2, и 3 одновременно, т.е. событие $A \cdot B$ состоит из элементарных событий — выборов одного из чисел 6, 12.

Пример 1.17. Пусть пространством событий является множество чисел $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Пусть $A = \{3, 4, 8, 9, 10\}$, $B = \{1, 2, 5, 6\}$. Тогда события A и B несовместны, т.е. $A \cdot B = \varnothing$ (невозможное событие). При этом $A + B = \{1, 2, 3, 4, 5, 6, 8, 9, 10\}$. Дополнением к событию А является событие $\overline{A} = \{1, 2, 5, 6, 7\}$. Разностью событий A и B будет событие $A - B = \{3, 4, 8, 9, 10\}$.

Операции над событиями можно представить как операции над множествами. При этом события представляются подмножествами некоторого множества Ω, обозначающего пространство всех элементарных событий. Сумме событий $A + B$ соответствует объединение $A \cup B$ этих подмножеств, а их произведению $A \cdot B$ — пересечение $A \cap B$. Достоверное событие представляется объемлющим множеством Ω, а невозможное событие — пустым подмножеством $\varnothing$ в нем. Несовместность событий A и B означает, что соответствующие подмножества A и B не пересекаются: $A \cap B = \varnothing$. Событие $\overline{A}$, противоположное событию А, является дополнением к событию A во множестве Ω. Эти операции в графическом виде иллюстрируются диаграммами Венна (рис. 1.3).

Приведем определения операций над множествами в формально-логической записи. Они понадобятся при формулировке аксиом

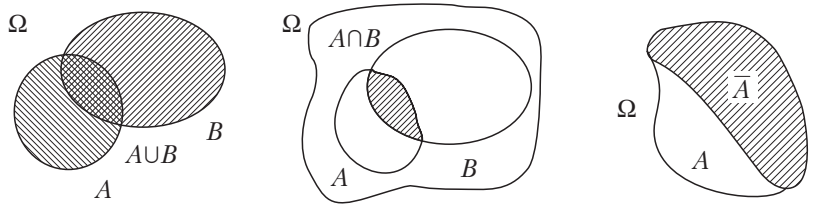


Рис. 1.3. Диаграммы Венна

теории вероятностей в § 1.8. Сами множества рассматриваются как подмножества в объемлющем пространстве Ω.

О б ъ е д и н е н и е множеств A и B, обозначаемое $A \cup B$, состоит из элементов, принадлежащих по крайней мере одному из них:

$$A \bigcup B = \{\omega \in \Omega \mid \omega \in A \vee \omega \in B\},$$

т.е.

$$\omega \in A \bigcup B \Leftrightarrow \omega \in A \vee \omega \in B,$$

где символ «$\vee$» означает «или».

Читается: «ω принадлежит объединению $A \cup B$ тогда и только тогда, когда ω принадлежит A или ω принадлежит B».

П е р е с е ч е н и е множеств A и B, обозначаемое $A \cap B$, состоит из элементов, принадлежащих и A, и B:

$$A \bigcap B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \in B\},$$

т.е.

$$\omega \in A \bigcap B \Leftrightarrow \omega \in A \wedge \omega \in B,$$

где символ «$\wedge$» означает «и».

Читается: «ω принадлежит пересечению $A \cap B$ тогда и только тогда, когда ω принадлежит A и ω принадлежит B».

Дополнение $\bar{A}$ к множеству A задается условием

$$\omega \in \bar{A} \Leftrightarrow \omega \notin A.$$

Знак «$\notin$» означает отрицание принадлежности.

Формулы де Моргана

$$\overline{A \cdot B} = \bar{A} + \bar{B}, \quad \overline{A + B} = \bar{A} \cdot \bar{B}. \tag{1.8}$$

Докажем эти формулы. Равенство двух множеств равносильно тому, что каждый элемент одного из них принадлежит и второму, и наоборот.

Имеем

$$\omega \in \overline{A \cdot B} \Leftrightarrow \omega \notin A \cdot B \Leftrightarrow \omega \notin A \vee \omega \notin B \Leftrightarrow$$
$$\Leftrightarrow \omega \in \bar{A} \vee \omega \in \bar{B} \Leftrightarrow \omega \in \bar{A} + \bar{B},$$

$$\omega \in \overline{A + B} \Leftrightarrow \omega \notin A + B \Leftrightarrow \omega \notin A \wedge \omega \notin B \Leftrightarrow$$
$$\Leftrightarrow \omega \in \bar{A} \wedge \omega \in \bar{B} \Leftrightarrow \omega \in \bar{A} \cdot \bar{B},$$

что и требовалось доказать.

Формулы де Моргана справедливы также для любого количества сомножителей или слагаемых.

1.5. Теорема сложения вероятностей

Свойства вероятности суммы событий, доказываемые здесь для классической схемы, в дальнейшем кладутся в основу аксиоматики вероятностей.

Пусть A и B несовместные события. Тогда вероятность их суммы равна сумме вероятностей:

$$P(A + B) = P(A) + P(B), \tag{1.9}$$

что называется *теоремой сложения вероятностей*.

Докажем это свойство для случая классической схемы.

Пусть n — число всех элементарных событий, m_A, m_B — число элементарных событий, благоприятствующих A и B соответственно. Так как A и B несовместны ($A \cap B = \varnothing$), то они не имеют общих благоприятствующих элементарных событий и $m_{A+B} = m_A + m_B$ — число всех благоприятствующих элементарных событий суммы $A + B$. Поэтому

$$P(A + B) = \frac{m_{A+B}}{n} = \frac{m_A + m_B}{n} = \frac{m_A}{n} + \frac{m_B}{n} = P(A) + P(B).$$

Формула (1.9) справедлива также и для любого конечного числа несовместных событий.

Следствием этой теоремы является связь между вероятностями противоположных событий:

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

Действительно,

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

так как A и $\bar{A}$ несовместны. Следовательно,

$$P(\bar{A}) = 1 - P(A). \tag{1.10}$$

Для произвольных событий верна более общая *теорема*:

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

Считая теорему (1.9) справедливой для любого пространства событий, докажем теорему (1.11). Для этого представим сумму $A + B$ в виде трех несовместных слагаемых: $A + B = A \cdot \bar{B} + \bar{A} \cdot B + A \cdot B$ (рис. 1.4).

По формуле (1.9) имеем

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

С другой стороны, A и B также представляются в виде суммы несовместных слагаемых:

$$A = A \cdot \bar{B} + A \cdot B,$$

$$B = B \cdot \bar{A} + B \cdot A = \bar{A} \cdot B + A \cdot B.$$

Поэтому

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

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

Отсюда следует, что

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

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

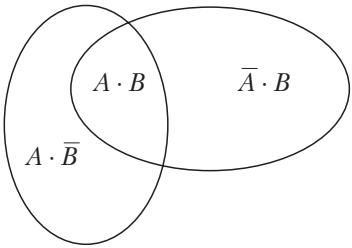


Рис. 1.4. Разложение суммы событий $A + B$

И теперь имеем

$$P(A + B) = P(A \cdot \bar{B}) + P(\bar{A} \cdot B) + P(A \cdot B) =$$

$$= P(A) - P(A \cdot B) + P(B) - P(A \cdot B) + P(A \cdot B) =$$

$$= P(A) + P(B) - P(A \cdot B),$$

что и требовалось доказать.

Эта теорема обобщается на случай трех и более слагаемых. Для любых трех событий справедлива формула

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

$$+ P(C) - P(A \cdot B) - P(A \cdot C) - P(B \cdot C) + P(A \cdot B \cdot C). \quad (1.12)$$

Пример 1.18. Пусть пространством событий является множество чисел $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, в котором все числа — элементарные события — равновозможны, вероятность каждого из них равна 1/10. Пусть $A = \{1, 3, 5, 9\}$, $B = \{2, 4, 10\}$, тогда A и B несовместны и $A + B = \{1, 2, 3, 4, 5, 9, 10\}$. При этом

$$P(A) = \frac{4}{10} = \frac{2}{5}, \quad P(B) = \frac{3}{10}, \quad P(A + B) = \frac{7}{10},$$

что соответствует формуле (1.9).

Пример 1.19. Пусть пространством событий является множество чисел $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, в котором все числа — элементарные события — равновозможны, вероятность каждого из них равна 1/10. Пусть $A = \{1, 3, 5, 9, 10\}$, $B = \{2, 5, 6, 8, 9\}$, тогда $A \cdot B = \{5, 9\}$, $A + B = \{1, 2, 3, 5, 6, 8, 9, 10\}$, $P(A) = \frac{5}{10}$, $P(B) = \frac{5}{10}$, $P(A \cdot B) = \frac{2}{10}$, $P(A + B) = \frac{8}{10}$, что соответствует формуле (1.11).

Пример 1.20. Пусть в группе из 27 туристов 17 человек владеют английским языком, 6 — французским, а 2 — обоими языками. Найти вероятность того, что случайно выбранный из группы турист владеет по крайней мере одним из этих языков.

Введем события: A — «выбранный турист владеет английским», B — «выбранный турист владеет французским». Тогда событие $A + B$ означает, что турист владеет хотя бы одним из этих языков. По условию задачи

$$P(A) = \frac{17}{27}, \quad P(B) = \frac{6}{27}, \quad P(A \cdot B) = \frac{2}{27},$$

по формуле (1.11)

$$P(A + B) = P(A) + P(B) - P(A \cdot B) = \frac{17}{27} + \frac{6}{27} - \frac{2}{27} = \frac{7}{9}.$$

1.6. Условные вероятности, теорема умножения вероятностей, независимость событий

Условная вероятность события связана с важным свойством событий, их независимостью.

На практике часто встречаются ситуации, когда наступление некоторого события значительно меняет возможности наступления других событий и их вероятности. Если произошло событие B, то новая вероятность события A называется *условной вероятностью* и обозначается $P(A|B)$. Читается «вероятность события A при условии B». При этом B оказывается достоверным событием и играет роль пространства элементарных событий Ω.

Условная вероятность $P(A|B)$ определяется формулой (при $P(B) > 0$)

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}. \tag{1.13}$$

В примере 1.19

$$P(A \cdot B) = \frac{2}{10}, \quad P(B) = \frac{5}{10},$$

поэтому условная вероятность

$$P(A|B) = \frac{\frac{2}{10}}{\frac{5}{10}} = \frac{2}{5}.$$

Эту же вероятность можно найти непосредственно по определению. В состав события B входит 5 чисел, только 2 из чисел, входящих в A, входят также и в B. По классической формуле (1.6)

$$P(A|B) = \frac{2}{5}.$$

Из (1.13) следуют симметричные формулы

$$P(A \cdot B) = P(B)P(A|B), \quad P(B) > 0, \tag{1.14}$$

$$P(A \cdot B) = P(A)P(B|A), \quad P(A) > 0, \tag{1.15}$$

которые носят название ***теорем умножения вероятностей***.

Независимыми называются такие события A и B, когда

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

Для независимых случайных событий из (1.14) следует равенство

$$P(A \cdot B) = P(A)P(B). \qquad (1.16)$$

Соотношение (1.16) также служит определением независимости событий, так что события называются *независимыми*, если вероятность их произведения равна произведению вероятностей.

Пример 1.21. Пусть события A и B несовместны и $P(A) \neq 0$, $P(B) \neq 0$. Тогда $P(A \cdot B) = P(\varnothing) = 0$, $P(A)P(B) \neq 0$. Поэтому $P(A \cdot B) \neq P(A)P(B)$. Это означает, что несовместные события A и B зависимы, если их вероятности отличны от нуля.

Пример 1.22. Пусть событие A влечет событие B, $A \subset B$ и $P(B) \neq 1$. Тогда $A \cdot B = A$, $P(A \cdot B) = P(A)$. Поэтому $P(A \cdot B) \neq P(A)P(B)$. Это означает, что если $A \subset B$, то A и B зависимы, если $P(A) \neq 0$, $P(B) \neq 1$.

Пример 1.23. Пусть пространством событий является множество чисел Ω из примера 1.18. Рассмотрим события $A = \{1, 3, 5, 7, 9\}$, $B = \{1, 2, 3, 5, 6, 8\}$. Тогда $A \cdot B = \{1, 3, 5\}$, $P(A \cdot B) = 3/10$, $P(A) = 5/10$, $P(B) = 6/10$. Значит, $P(A \cdot B) = P(A)P(B)$ и события A и B независимы.

Пример 1.24. Пусть событие A — извлечение наудачу из колоды карты пиковой масти, B — извлечение наудачу из колоды туза. Проверим независимость событий A и B. Имеем

$$P(A) = \frac{9}{36} = \frac{1}{4}, \quad P(B) = \frac{4}{36} = \frac{1}{9}, \quad P(A \cdot B) = \frac{1}{36}.$$

Поскольку $P(A \cdot B) = P(A)P(B)$, события A и B независимы.

Формулы (1.14) и (1.15) обобщаются на случай трех и более событий:

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

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_k) = P(A_1)\,P(A_2|A_1) \ldots P(A_k|A_1 \cdot A_2 \cdot \ldots \cdot A_{k-1}).$$

Несколько событий A_1, A_2, ..., A_k называются *независимыми в совокупности*, если они попарно независимы и независимы любое из них и произведение любых остальных событий.

Для независимых в совокупности событий A_1, A_2, ..., A_k имеют место равенства

$$P(A_i \cdot A_j) = P(A_i)P(A_j), \quad i \neq j;$$

$$P(A_{i_1} \cdot A_{i_2} \cdot \ldots \cdot A_{i_s}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_s}),$$

$$i_p \neq i_q; \quad i, j, i_1, \ldots, i_s = \overline{1, k}. \tag{1.17}$$

В отличие от случая двух событий, равенств (1.17) уже недостаточно для независимости в совокупности трех и более событий.

Следующий пример показывает, что из попарной независимости не следует независимость в совокупности и что свойство независимости в совокупности может оказаться труднопроверяемым. Однако в приложениях теории вероятностей используется в основном только равенство (1.17) для независимых в совокупности событий.

Пример 1.25. Имеется четыре флага: красный, зеленый, белый и трехцветный красно-зелено-белый. Наудачу выбирается флаг. Исследовать на независимость события, состоящие в выборе флага следующего цвета: К — красного, З — зеленого, Б — белого.

Для события К благоприятными исходами будут выбор красного и трехцветного флагов, поэтому

$$P(\text{К}) = \frac{2}{4} = \frac{1}{2}.$$

Аналогично имеем

$$P(\text{З}) = \frac{1}{2},$$

$$P(\text{Б}) = \frac{1}{2}.$$

Событие К · З · Б означает выбор только трехцветного флага (из четырех), поэтому

$$P(\text{К} \cdot \text{З} \cdot \text{Б}) = \frac{1}{4}.$$

Поскольку неравенство

$$P(\text{К} \cdot \text{З} \cdot \text{Б}) = \frac{1}{4} \neq \frac{1}{8} = P(\text{К})P(\text{З})P(\text{Б})$$

противоречит условию (1.17), события К · З · Б не являются независимыми в совокупности. В то же время

$$P(\text{К} \cdot \text{З}) = \frac{1}{4} = P(\text{К})P(\text{З}),$$

$$P(\text{К} \cdot \text{Б}) = \frac{1}{4} = P(\text{К})P(\text{Б}),$$

$$P(\text{З} \cdot \text{Б}) = \frac{1}{4} = P(\text{З})P(\text{Б}).$$

И события К, З, Б попарно независимы.

Приведем еще один пример попарно независимых событий, которые не являются независимыми в совокупности.

Пример 1.26 (пример Бернштейна). Пусть пространство событий состоит из четырех равновероятных событий $\{\omega_1, \omega_2, \omega_3, \omega_4\}$, $P(\omega_i) = 1/4$. Пусть $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$, $C = \{\omega_1, \omega_4\}$. Тогда

$$P(A) = P(B) = P(C) = \frac{1}{2},$$

$$P(A \cdot B) = \frac{1}{4} = P(A)P(B),$$

$$P(A \cdot C) = \frac{1}{4} = P(A)P(C),$$

$$P(B \cdot C) = \frac{1}{4} = P(B)P(C).$$

Следовательно, события A, B, C — попарно независимы. Однако

$$P(A \cdot B \cdot C) = P(\omega_1) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}.$$

1.7. Формула полной вероятности и формула Байеса

В случае, когда пространство событий представляется в виде суммы попарно несовместных событий (гипотез), вероятности которых заданы, то вероятность произвольного события A из этого

пространства выражается через вероятности гипотез и условные вероятности события A по формуле полной вероятности. Формула Байеса позволяет в этих условиях найти апостериорные вероятности гипотез после наступления события A.

События $H_1, H_2, ..., H_n$ образуют ***полную группу событий***, если они попарно несовместны, а их сумма является достоверным событием, т.е. $H_i \cdot H_j = \varnothing$ при $i \neq j$ и $H_1 + H_2 + ...+ H_n = \Omega$.

Такие события называются ***гипотезами***. Простейшим примером полной группы событий является произвольное событие A и его дополнение $\overline{A}$. По теореме сложения вероятностей для полной группы событий справедливо равенство

$$P(H_1) + P(H_2) + ... + P(H_n) = 1.$$

Пример 1.27. Пусть $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, как в примере 1.18. Рассмотрим события $H_1 = \{1, 3, 5, 7, 9\}$, $H_2 = \{2, 4, 6\}$, $H_3 = \{8, 10\}$. Они попарно несовместны, а их сумма является достоверным событием Ω. Значит, H_1, H_2, H_3 составляют полную группу событий. Для них $P(H_1) = 5/10$, $P(H_2) = 3/10$, $P(H_3) = 2/10$ и сумма их вероятностей равна 1.

Пример 1.28. В лесу растут деревья, среди которых 60 % берез, 10 % елей и 30 % сосен. Для замеров деревьев случайным образом выбирается одно из них. Введем события: H_1 — «выбранное дерево — береза», H_2 — «выбранное дерево — ель», H_3 — «выбранное дерево — сосна». Тогда события H_1, H_2, H_3 попарно несовместны, $P(H_1) = 0{,}6$, $P(H_2) = 0{,}1$, $P(H_3) = 0{,}3$. Значит, H_1, H_2, H_3 — полная группа событий.

Полную группу событий можно интерпретировать диаграммой (рис. 1.5). Полная группа событий H_i, $i = 1, 2, ..., n$ расслаивает пространство Ω на непересекающиеся множества.

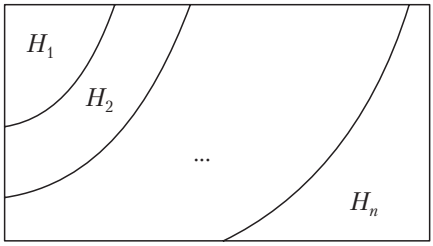


Рис 1.5. Полная группа событий

Пусть события H_1, H_2, ..., H_n образуют полную группу событий. Тогда для любого события A имеет место формула полной вероятности:

$$P(A) = \sum_{k=1}^{n} P(A|H_k)P(H_k). \qquad (1.18)$$

Пример 1.29. Сохраним условия примера 1.28. Пусть при замере диаметра деревьев он оказался больше 15 см для 30 % берез, 40 % елей и 70 % сосен. Введем событие A — «диаметр случайно выбранного дерева больше 15 см». Найдем вероятность этого события.

По условию примера условные вероятности события A $P(A|H_1) = 0{,}3$, $P(A|H_2) = 0{,}4$, $P(A|H_3) = 0{,}7$. Тогда по формуле полной вероятности (1.14) имеем:

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

$$= 0{,}6 \cdot 0{,}3 + 0{,}1 \cdot 0{,}4 + 0{,}3 \cdot 0{,}7 = 0{,}43.$$

В формулу полной вероятности входят вероятности $P(H_1)$, $P(H_2)$, ..., $P(H_n)$, которые называются априорными. Если событие A наступило, то эти вероятности изменяются. Это будут условные вероятности $P(H_1|A)$, $P(H_2|A)$, ..., $P(H_n|A)$. Они могут быть найдены по ***формуле Байеса***:

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

Пример 1.30. Пусть в условиях предыдущего примера диаметр случайно выбранного дерева оказался больше 15 см, т.е. событие A наступило. Найти вероятность того, что измеренное дерево — береза.

По формуле Байеса (1.19) имеем:

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A)} = \frac{0{,}3 \cdot 0{,}6}{0{,}43} \approx 0{,}42.$$

Доказательство равенств (1.18) и (1.19) легко получается из представления A в виде суммы несовместных событий

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

1.8. Аксиоматическое построение теории вероятностей

Дедуктивно-аксиоматический метод служит фундаментом для построения всех разделов математики и других естественных наук. Согласно этому методу научная теория строится по следующей схеме. Сначала вводятся основные (неопределяемые) объекты и основные отношения между ними. Содержание теории составляют выводы о свойствах объектов теории, полученные формально-логическим путем из основных отношений между ними. Особый вклад в аксиоматический метод внесла история построения геометрии, аксиоматика которой прошла многовековой путь от «Начал» Евклида до заключительных трудов Д. Гильберта. Аксиоматика теории вероятностей построена А.Н. Колмогоровым в начале XX в. Эта аксиоматика использует основы теории множеств и теории функций, основными ее объектами служат события и их вероятности, а основные отношения наследуют свойства классической и статистической вероятностей.

Приведем основные положения аксиоматики Колмогорова теории вероятностей.

Пространством элементарных событий называется некоторое множество Ω, элементы которого называются *элементарными событиями*. Далее предполагается, что задана некоторая (непустая) совокупность подмножеств множества Ω, называемых *событиями*. События обозначаются заглавными латинскими буквами A, B, C и т.д., а вся совокупность событий — буквой $\mathcal{F}$.

Совокупность событий $\mathcal{F}$ должна удовлетворять следующим *аксиомам событий*:

А1. Для любого конечного или счетного семейства событий их о б ъ е д и н е н и е также является событием.

А2. Для всякого события A его д о п о л н е н и е $\bar{A}$ до всего пространства событий также является событием.

В формальной записи аксиомы событий выглядят следующим образом:

А1. $\{A_i\}_{i\in\mathcal{J}}$ — конечное или счетное семейство, тогда $A_i \in \mathcal{F}\ \forall$ $\forall\, i \in \mathcal{J} \Rightarrow \bigcup\limits_{i\in\mathcal{J}} A_i \in \mathcal{F}$.

А2. $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$.

Если пространство событий Ω является *конечным множеством*, то семейство всех его подмножеств удовлетворяет аксиомам событий.

Для всякого события A объединение $A \cup \bar{A} = \Omega$, поэтому само Ω является событием. Событием также является пустое подмножество $\varnothing = \bar{\Omega}$. Из аксиом А1 и А2 также следует, что пересечение конечного или счетного семейства событий также является событием.

Совокупность $\mathcal{F}$ подмножеств данного множества Ω, удовлетворяющих аксиомам А1 и А2, называется σ-*алгеброй*.

Таким образом, σ-*алгебра* — это совокупность $\mathcal{F}$ подмножеств данного множества, в которой можно выполнять операции объединения и пересечения ее конечных и счетных семейств и переходить к дополнениям подмножеств.

Для полноты изложения приведем определения объединения и пересечения счетного семейства множеств.

Семейство множеств $\{A_i\}_{i\in\mathcal{J}}$ называется счетным (в частности, конечным), если его семейство индексов $\mathcal{J}$ является множеством натуральных чисел $\mathbb{N}$ (или его подмножеством). Счетное семейство подмножеств можно записывать в виде

$$\{A_i\}_{i\in\mathbb{N}} = \{A_1, A_2, ..., A_n, ...\}. \tag{1.20}$$

О б ъ е д и н е н и е м семейства (1.20) называется множество

$$\bigcup_{i\in\mathbb{N}} A_i = \left\{\omega \in \Omega \mid \exists i_0 \in \mathbb{N},\ \omega \in A_{i_0}\right\}$$

(знак «$\exists$» означает «существует»).

П е р е с е ч е н и е м семейства (1.20) называется множество

$$\bigcap_{i\in\mathbb{N}} A_i = \left\{\omega \in \Omega \mid \forall i \in \mathbb{N},\ \omega \in A_i\right\}$$

(знак «$\forall$» означает «для любого»).

Приведем теперь аксиомы, задающие вероятности.

Пусть $\mathcal{F}$ — σ-алгебра подмножеств (событий) пространства Ω. Тогда вероятности событий описываются следующими *аксиомами вероятностей*:

Р1. Каждому событию $A \in \mathcal{F}$ поставлено в соответствие неотрицательное число $P(A)$, называемое *вероятностью события A*.

Р2. $P(\Omega) = 1$ — *аксиома нормировки*.

Р3. Если $\{A_i\}_{i\in\mathbb{N}}$ — счетное семейство попарно несовместных событий, то $P\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \sum\limits_{i\in\mathbb{N}} P(A_i)$ — *аксиома счетной аддитивности*.

Для конечного пространства событий Ω аксиому счетной аддитивности можно заменить аксиомой сложения:

$$P(A_1 + A_2) = P(A_1) + P(A_2),$$

где $A_1 + A_2$ означает объединение $A_1 \cup A_2$.

Более общим образом в дальнейшем знак объединения $\cup$ заменяется знаком суммы (+), знак пересечения $\cup$ — знаком произведения ($\cdot$).

Вероятностным пространством называется тройка объектов $\{\Omega, \mathcal{F}, P\}$, состоящая из пространства событий Ω, σ-алгебры $\mathcal{F}$ его подмножеств и функции вероятности P, удовлетворяющих аксиомам А1, А2, Р1, Р2, Р3 событий и вероятностей.

Все теоремы теории вероятностей выводятся формально-логическим путем из этих аксиом.

Например, свойство $P(A) + P(\bar{A}) = 1$ легко следует из аксиом Р2 и Р3, поскольку $A + \bar{A} = \Omega$. При доказательстве правил сложения вероятностей (1.11) мы по существу пользовались только аксиомой сложения.

В качестве более содержательных теорем приведем без доказательства две теоремы о непрерывной зависимости вероятности $P(A)$ от события A, или о переходе к пределу под знаком вероятностей [11].

Теорема 1. Если $A_1, A_2, \ldots$ — последовательность таких событий, что каждое следующее из них является следствием предыдущего ($A_1 \subset A_2 \subset \ldots \subset A_n \subset A_{n+1} \subset \ldots$), то

$$\lim_{n\to\infty} P(A_n) = P(A_1 + A_2 + \ldots + A_n + \ldots).$$

Теорема 2. Если $A_1, A_2, \ldots$ — последовательность таких событий, что каждое предыдущее является следствием последующего ($A_1 \supset A_2 \supset \ldots \supset A_n \supset A_{n+1} \supset \ldots$), то

$$\lim_{n\to\infty} P(A_n) = P(A_1 \cdot A_2 \cdot \ldots \cdot A_n \cdot \ldots).$$

В заключение приведем общую модель вероятностного пространства, соответствующую классической схеме.

Классическим вероятностным пространством называется тройка $\{\Omega, \mathcal{F}, P\}$, где Ω — конечное множество, $\mathcal{F}$ — σ-алгебра всех его подмножеств (событий), а вероятностью события $A \in \mathcal{F}$ называется число $P(A) = |A|/|\Omega|$, где для произвольного конечного множества M число $|M|$ обозначает число элементов этого множества.

Задачи с решениями

Задача 1.1. Сколько прямых линий можно провести через 9 точек, из которых ровно 3 лежат на одной прямой?

Решение. Обозначим точки через $A_1, A_2, ..., A_9$ и предположим, что A_1, A_2, A_3 лежат на одной прямой. Через каждую пару точек из множества $\{A_1, A_2, ..., A_9\}$ проведем одну прямую. Число таких прямых равно $C_9^2 = 36$. Так как через точки A_1, A_2, A_3 проходит одна прямая (а не три), то искомое число прямых $n = 36 - 2 = 34$.

Задача 1.2. На железнодорожной станции имеется 10 путей. Сколькими способами можно расставить на них 3 состава?

Решение. Число способов поставить первый состав равно 10, второй — 9 (так как один путь уже занят), третий — 8. По правилу произведения искомое число равно числу размещений:

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

Задача 1.3. 9 человек размещаются в гостинице, состоящей из четырех-, трех- и двухместного номера. Сколько существует способов их размещения?

Решение. Число способов заполнить четырехместный номер равно C_9^4, после этого для заполнения трехместного номера остается C_5^3 способов и, наконец, для двухместного номера остается один способ. По правилу произведения число всех способов равно

$$C_9^4 \cdot C_5^3 \cdot C_2^2 = \frac{9 \cdot 8 \cdot 7 \cdot 6}{2 \cdot 3 \cdot 4} \cdot \frac{5 \cdot 4 \cdot 3}{2 \cdot 3} \cdot 1 = 1\,260.$$

Можно начинать заполнение с трехместного номера, можно с двухместного. Результат от этого не изменяется.

Задача 1.4. В лотерее «Спортлото» игрок должен зачеркнуть 6 из 49 возможных чисел от 1 до 49. Сколько существует всех возможных вариантов выбора для игрока?

Решение. Число вариантов равно числу выбора всех 6-элементных подмножеств из 49-элементного множества, т.е. числу сочетаний:

$$n = C_{49}^{6} = \frac{49!}{6! \cdot 43!} = 13\,983\,816.$$

Задача 1.5. Сколькими способами можно выбрать 3 цветка из вазы, в которой стоят 10 красных и 4 розовых гвоздики? А если выбрать красную гвоздику и 2 розовых?

Решение. Так как порядок выбора и цвет не учитываются, то выбор 3 цветков означает сочетание по 3 из 14. Всех возможностей будет

$$C_{14}^{3} = \frac{14 \cdot 13 \cdot 12}{1 \cdot 2 \cdot 3} = 364.$$

Выбрать красную гвоздику можно 10 способами, выбор же двух розовых гвоздик из четырех является сочетанием из 4 по 2. Их количество равно

$$C_{4}^{2} = \frac{4 \cdot 3}{1 \cdot 2} = 6.$$

По правилу произведения выбор одной красной и двух розовых гвоздик можно произвести $10 \cdot 6 = 60$ способами.

Задача 1.6. Сколько существует способов выпадения или менее 3, или более 10 очков при двух бросаниях игральной кости?

Решение. По правилу суммы число способов выпадения или менее 3, или более 10 очков равно сумме способов выпадения менее 3 и более 10 очков. Существует только один способ выпадения менее 3 очков (1 + 1). Выпадение более 10 очков возможно (по правилу суммы) 3 способами (или 5 + 6, или 6 + 5, или 6 + 6). Поэтому число искомых возможностей равно 1 + 3 = 4.

Задача 1.7. В партии из 100 изделий находится 5 бракованных. Для контроля было выбрано 5 изделий. Какова вероятность того, что среди них будет ровно одно бракованное?

Решение. Для того чтобы одно изделие было бракованным, его следует выбрать из 5 бракованных, число способов такого выбора равно C_5^1. Остальные 4 изделия выбираются из стандартных 95 изделий, это можно осуществить C_{95}^4 способами. По теореме умножения число способов, благоприятствующих событию, $m = C_5^1 \cdot C_{95}^4$. Число всех способов равно числу способов выбрать 5 элементов из 100, т.е. $n = C_{100}^5$. Поэтому искомая вероятность

$$p = \frac{C_{95}^4 \cdot C_5^1}{C_{100}^5} = \frac{95! \cdot 5 \cdot 5! \cdot 95!}{4! \cdot 91! \cdot 100!} = \frac{95 \cdot 94 \cdot 93 \cdot 92 \cdot 5 \cdot 5}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = 0{,}2114.$$

Задача 1.8. Из 1000 посаженных семян взошло 787. Какова статистическая вероятность всхожести семян?

Решение. Частость всхожести семян равна отношению

$$\frac{787}{1000} = 0{,}787.$$

Она же и принимается за статистическую вероятность всхожести семян.

Задача 1.9. Расстояние от пункта M до пункта N автобус проходит за 2 мин, а пешеход за 15 мин. Интервал движения автобусов 25 мин. Пешеход в случайный момент времени отправляется из M в N пешком. Найти вероятность того, что его в пути догонит автобус.

Решение. Поскольку интервал движения автобусов больше времени, в течение которого пешеход находится в пути ($25 > 15$), догнать пешехода может не более, чем один автобус. Обозначим через t_1 ($0 \le t_1 \le 15$) — время в пути пешехода после выхода из пункта M, t_2 ($0 \le t_2 \le 2$) — время в пути после отправления из пункта M того автобуса, который может догнать пешехода. Пространство всех событий $\{(t_1, t_2) \mid 0 \le t_1 \le 15,\ 0 \le t_2 \le 2\}$ изобразим точками прямоугольника со сторонами 15 и 2 (рис. 1.6).

Время, которое осталось пешеходу до прихода в пункт N, равно $15 - t_1$ мин, время, которое осталось автобусу до прибытия в пункт N, равно $2 - t_2$ мин. Автобус прибывает в N не позже пешехода, поэтому $2 - t_2 \le 15 - t_1$, или $t_2 \ge t_1 - 13$. Автобус отправляется из M

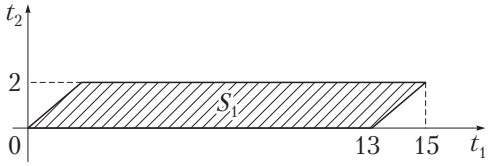


Рис. 1.6. Задача о движении

не раньше пешехода, поэтому $t_2 \le t_1$. Пространство благоприятных исходов изображается точками заштрихованного параллелограмма с площадью $S_1 = 13 \cdot 2 = 26$. Площадь всего прямоугольника равна $S = 30$. По формуле геометрической вероятности искомая вероятность

$$p = \frac{S_1}{S} = \frac{13}{15}.$$

Задача 1.10. 5 открыток случайным образом раскладываются по 3 конвертам. Какова вероятность того, что в каждом конверте окажется хотя бы одна открытка?

Решение. Обозначим через A_i, $i = 1, 2, 3$, событие, состоящее в том, что ни одна открытка не попала в i-й конверт, через A — интересующее нас событие. Тогда

$$A = \overline{A_1 + A_2 + A_3},$$

$$P(A) = P\left(\overline{A_1 + A_2 + A_3}\right) = 1 - P(A_1 + A_2 + A_3).$$

Для применения формулы суммы (1.20) вычислим нужные вероятности:

$$P(A_1) = P(A_2) = P(A_3) = \frac{2^5}{3^5} = \frac{32}{243} \approx 0{,}132,$$

так как число способов, по которым можно разложить 5 открыток по 3 конвертам, по правилу произведения равно 3^5, а число способов разложить открытки по двум конвертам равно 2^5.

$$P(A_1 \cdot A_2) = P(A_1 \cdot A_3) = P(A_2 \cdot A_3) = \frac{1}{243} \approx 0{,}004,$$

так как, например, событие $A_1 \cdot A_2$ означает попадание всех открыток в третий конверт, т.е. благоприятным будет только один исход.

$P(A_1 \cdot A_2 \cdot A_3) = 0$, так как событие $A_1 \cdot A_2 \cdot A_3$ является невозможным.

Теперь

$$P(\bar{A}) = 3 \cdot \frac{32}{243} - \frac{3}{243} = \frac{93}{243} = \frac{31}{81} \approx 0{,}383,$$

$$P(A) = 1 - P(\bar{A}) \approx 0{,}617.$$

Задача 1.11. В рекламной фирме 21 % работников получает высокую зарплату. Среди них отношение числа мужчин и женщин равно 14,6:6,4. Известно также, что в фирме работают 40 % женщин. Выяснить, существует ли в фирме дискриминация женщин в оплате труда.

Решение. Введем события: A — «случайно выбранный работник получает высокую зарплату», B — «случайно выбранный работник — женщина». Тогда по условию $P(A \cdot B) = 0{,}064$, $P(B) = 0{,}4$, $P(A) = 0{,}21$. Найдем вероятность $P(A|B)$. По определению условной вероятности

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} = 0{,}16.$$

Так как $P(A|B) < P(A)$, то труд женщин является менее оплачиваемым.

Задача 1.12. В коробке находится 4 белых, 3 синих и 2 черных шара. Наудачу последовательно вынимают 3 шара. Какова вероятность того, что 1-й шар будет белым, 2-й — синим, 3-й — черным?

Решение. Введем следующие события: A_1 — «первым вытащили белый шар», A_2 — «вторым — синий», A_3 — «третьим — черный». Тогда нужное нам событие A представится в виде $A = A_1 \cdot A_2 \cdot A_3$.

По правилу умножения вероятностей

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cdot A_2).$$

$$P(A_1) = \frac{4}{9};$$

$$P(A_2|A_1) = \frac{3}{8},$$

так как шаров осталось 8, а число благоприятных случаев для события A_2 равно 3;

$$P(A_3|A_1 \cdot A_2) = \frac{2}{7},$$

так как уже два шара (белый и синий) вытащены.

Следовательно,

$$P(A) = \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} = \frac{1}{21} \approx 0{,}05.$$

Задача 1.13. В сборочный цех завода поступает 40 % деталей из I цеха и 60 % — из II цеха. В I цехе производится 90 % стандартных деталей, а во II — 95 %. Найти вероятность того, что наудачу взятая сборщиком деталь окажется стандартной.

Решение. Взятие детали можно разбить на два этапа. Первый — это выбор цеха. Имеется две гипотезы: H_1 — деталь изготовлена I цехом, H_2 — II цехом. Второй этап — взятие детали. Событие A — «взятая наудачу деталь стандартна». Очевидно, события H_1 и H_2 образуют полную группу, $P(H_1) = 0{,}4$, $P(H_2) = 0{,}6$. Числа 0,90 и 0,95 являются условными вероятностями события A при условии гипотез H_1 и H_2 соответственно, т.е. $P(A|H_1) = 0{,}90$ и $P(A|H_2) = 0{,}95$. По формуле (1.20) находим

$$P(A) = \sum_{i=1}^{2} P(H_i)P(A|H_i) = 0{,}4 \cdot 0{,}90 + 0{,}6 \cdot 0{,}95 = 0{,}93.$$

Задача 1.14. В условиях задачи 1.13 найти вероятность того, что эта стандартная деталь изготовлена II цехом.

Решение. Определим вероятность гипотезы H_2 при условии, что событие A (взятая деталь стандартна) уже произошло, т.е. $P(H_2|A)$:

$$P(H_2|A) = \frac{P(H_2)P(A|H_2)}{P(A)} = \frac{0{,}6 \cdot 0{,}95}{0{,}93} = \frac{19}{31} \approx 0{,}613.$$

Задача 1.15. В первом ящике из 20 деталей 4 бракованных, во втором из 30 деталей 5 бракованных. Из первого ящика переложили во второй 2 детали. Найти вероятность того, что деталь, извлеченная после этого из второго ящика, бракованная.

Решение. Мы имеем 3 гипотезы: H_1 — обе переложенные детали бракованные, H_2 — одна из них бракованная, H_3 — обе переложенные детали качественные. Эти гипотезы составляют полную группу событий. Найдем вероятности этих гипотез:

$$P(H_1) = \frac{C_4^2}{C_{20}^2} = \frac{4 \cdot 3}{20 \cdot 19} = \frac{3}{95} \approx 0{,}032,$$

$$P(H_2) = \frac{C_4^1 \cdot C_{16}^1}{C_{20}^2} = \frac{4 \cdot 16}{10 \cdot 19} = \frac{32}{95} \approx 0{,}337,$$

$$P(H_3) = \frac{C_{16}^2}{C_{20}^2} = \frac{16 \cdot 15}{20 \cdot 19} = \frac{60}{95} \approx 0{,}632.$$

Обозначим через A искомое событие — «извлеченная после перекладывания деталь — бракованная». Найдем условные вероятности события A:

$$P(A|H_1) = \frac{7}{32} \approx 0{,}219,$$

$$P(A|H_2) = \frac{6}{32} \approx 0{,}188,$$

$$P(A|H_3) = \frac{5}{32} \approx 0{,}156.$$

По формуле полной вероятности (1.14)

$$P(A) = 0{,}219 \cdot 0{,}032 + 0{,}188 \cdot 0{,}337 + 0{,}632 \cdot 0{,}156 = 0{,}169.$$

Задача 1.16. Оператор радиолокационной станции фиксирует самолет противника с вероятностью 0,8 и принимает помеху за самолет с вероятностью 0,1. Практика показала, что в 15 % случаев на экран оператора попадает помеха. Оператор принял решение о наличии в воздушном пространстве самолета противника. Определить вероятность того, что сигнал получен действительно от самолета.

Решение. Введем события: A — «решение оператора о появлении самолета», H_1 — «появление самолета», H_2 — «появление помехи». Требуется найти вероятность $P(H_1|A)$.

События H_1 и H_2 образуют полную группу событий, так как $H_1 \cdot H_2 = \varnothing$ и $H_1 + H_2 = \Omega$. Тогда по формуле Байеса

$$P(H_1|A) = \frac{P(A|H_1)P(H_1)}{P(A)},$$

где $P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2)$.

По условию задачи

$$P(H_1) = 0{,}85, \quad P(H_2) = 0{,}15, \quad P(A|H_1) = 0{,}8, \quad P(A|H_2) = 0{,}1.$$

Тогда

$$P(A) = 0{,}8 \cdot 0{,}85 + 0{,}1 + 0{,}15 = 0{,}695.$$

Итак,

$$P(H_1|A) = \frac{0{,}8 \cdot 0{,}85}{0{,}695} = 0{,}9784.$$

Задачи для самостоятельного решения

1.1. Для разгрузки товаров директору маркета требуется выделить 5 из 20 имеющихся рабочих. Сколькими способами это можно сделать, осуществляя отбор в случайном порядке?

1.2. Отдел рекламы фирмы имеет средства на размещение рекламы только в 15 из 25 городских газет. Сколько существует способов для случайного отбора газет для помещения объявлений?

1.3. Номер машины состоит из 3 букв и 4 цифр. Сколько всего существует разных номеров, если алфавит содержит 32 буквы?

1.4. Сколько можно образовать различных инициалов, если каждый человек имеет одну фамилию, имя, отчество и для инициалов из алфавита используется 29 букв?

1.5. Для доступа в компьютерную сеть оператору необходимо набрать пароль из 4 цифр. Оператор забыл или не знает необходимого кода. Сколько всевозможных комбинаций он может составить для набора пароля: а) если цифры в коде не повторяются; б) если могут повторяться?

1.6. Сколькими способами можно выбрать старосту и профорга в группе из 25 студентов?

1.7. Имеется две урны. В первой — 10 красных и 6 черных шаров. Во второй — 4 красных и 6 черных шаров. Из каждой урны вынимается по шару. Найти вероятность того, что: а) оба шара будут красными; б) из первой урны будет вынут красный шар, а из второй — черный; в) хотя бы один из вынутых шаров черный.

1.8. Из 5 букв разрезной азбуки составлено слово «книга». Ребенок, не умеющий читать, рассыпал эти буквы, а затем собрал в произвольном порядке. Найти вероятность того, что у него снова получилось слово «книга».

1.9. В колоде 36 карт 4 мастей. После извлечения и возвращения одной карты колода перемешивается и снова извлекается одна карта. Определить вероятность того, что обе извлеченные карты одной масти.

1.10. 8 различных книг расставляются наугад на полке. Найти вероятность того, что две определенные книги окажутся поставленными рядом.

1.11. Среди изготовленных 15 деталей имеется 5 нестандартных. Определить вероятность того, что взятые наугад три детали окажутся стандартными.

1.12. В клетке содержится 18 кур. Из них 6 невакцинированы. Партию делят на две равные части. Какова вероятность того, что невакцинированные куры разделятся поровну?

1.13. Из 25 студентов группы 12 занимаются научной работой на кафедре бухгалтерского учета, 7 — экономического анализа, остальные — на кафедре статистики. Какова вероятность того, что два случайно отобранных студента занимаются научной работой на кафедре статистики?

1.14. Определить вероятность того, что участник лотереи «Спортлото — 5 из 36» угадает правильно все 5 номеров?

1.15. В группе 25 студентов, из них 10 юношей и 15 девушек. Какова вероятность того, что из вызванных наудачу трех студентов: а) все три девушки; б) первые две девушки, третий юноша; в) все трое юноши?

1.16. Вероятность попадания в цель при одном выстреле равна 0,7. Выстрелы производятся по одному до первого попадания.

Определить вероятность того, что придется производить четвертый выстрел.

1.17. Вероятность одного попадания в цель при одновременном залпе из двух орудий равна 0,44. Найти вероятность поражения цели при одном выстреле первым орудием, если для второго орудия эта вероятность равна 0,8.

1.18. Вероятность выигрыша по лотерейному билету равна 0,1. Приобретено три билета. Какова вероятность выиграть хотя бы по одному из них?

1.19. Брошены две игральные кости. Предполагается, что все комбинации выпавших очков равновероятны. Найти условную вероятность того, что выпали две пятерки, если известно, что сумма выпавших очков делится на 5.

1.20. Два игрока поочередно бросают две игральные кости. Выигрывает тот, у которого первым появится в сумме 12 очков. Найти вероятность выигрыша для каждого игрока.

1.21. Студент из 40 экзаменационных вопросов выучил только 30. Каким выгодней ему зайти на экзамен — первым или вторым?

1.22. Вероятность хотя бы одного попадания в мишень стрелком при трех выстрелах равна 0,784. Найти вероятность одного промаха при трех выстрелах.

1.23. В первой урне 10 деталей, из них 8 стандартных. Во второй — 6 деталей, из которых 5 стандартных. Из второй урны переложили в первую одну деталь. Какова вероятность того, что деталь, извлеченная после этого из второй урны, нестандартная?

1.24. В первой бригаде производится в три раза больше продукции, чем во второй. Вероятность того, что производимая продукция окажется стандартной для первой бригады, равна 0,7, для второй — 0,8. Определить: а) вероятность того, что взятая наугад единица продукции будет стандартной; б) вероятность, что она из второй бригады.

1.25. Расстояние от пункта M до пункта N автобус проходит за 3 мин, а пешеход за 15 мин. Интервал движения автобусов 30 мин. Пешеход в случайный момент времени отправляется из M в N пешком. Найти вероятность того, что его в пути догонит автобус.

Ответы

1.1. $n = 15\,504$.
1.2. $n = 3\,268\,760$.
1.3. $n = 327\,680\,000$.
1.4. $29^3 = 24\,389$.
1.5. а) $n = 5040$,
б) $n = 10\,000$.
1.6. $n = 25 \cdot 24 = 600$.
1.7. $p = 0{,}25$; $p = 0{,}375$; $p = 0{,}75$.
1.8. $p = \frac{1}{120}$.
1.9. $p = 0{,}25$.
1.10. $p = 0{,}25$.
1.11. $p = 0{,}26$.
1.12. $p = 0{,}38$.
1.13. $p = 0{,}05$.
1.14. $p = \frac{1}{376\,992}$.
1.15. а) $p = \frac{91}{460}$;
б) $p = \frac{7}{46}$; в) $p = \frac{6}{115}$.
1.16. $p = 0{,}027$.
1.17. $p = 0{,}6$.
1.18. $p = 0{,}271$.
1.19. $p = \frac{1}{7}$.
1.20. $p = \frac{36}{71}$; $p = \frac{35}{71}$.
1.21. Вероятность одинакова.
1.22. $p = 0{,}288$.
1.23. $p = \frac{1}{6}$.
1.24. а) $p = 0{,}725$;
б) $p = 0{,}276$.
1.25. $p = 0{,}8$.

2

СХЕМА ИСПЫТАНИЙ БЕРНУЛЛИ

Наряду с классическим вероятностным пространством важное место в теории вероятностей отводится специальной схеме испытаний Бернулли, справедливой для числа успехов в n независимых испытаниях. Однако ее применение для большого количества испытаний слишком громоздко. В этом случае используются приближенные формулы: формула Пуассона, а также локальная и интегральная формулы Муавра — Лапласа.

2.1. Формула Бернулли

Пусть проводится n независимых испытаний, в результате каждого из которых событие A может появиться с вероятностью p и не появиться с вероятностью $q = 1 - p$. Независимость испытаний означает независимость события A от порядка испытания. Появление события A называется успехом, а непоявление — неуспехом. Такая схема называется *последовательностью испытаний Бернулли*, или *схемой Бернулли*.

Вероятность события «ровно m успехов в n испытаниях», обозначаемая $P_n(m)$, вычисляется по ***формуле Бернулли***:

$$P_n(m) = C_n^m p^m q^{n-m}. \tag{2.1}$$

Идею доказательства этой формулы проиллюстрируем на примере.

Пример 2.1. Пусть трижды бросается игральная кость. С какой вероятностью ровно два раза выпадет 6 очков?

Имеем схему испытаний, в которой число испытаний n равно 3, а вероятность p события A (выпадение шестерки) в каждом испытании равна 1/6. Выпадению шестерки ровно 2 раза при 3 испытаниях соответствует событие

$$A \cdot A \cdot \bar{A} + A \cdot \bar{A} \cdot A + \bar{A} \cdot A \cdot A.$$

Слагаемые суммы являются несовместными событиями. Поэтому вероятность суммы равна сумме вероятностей слагаемых. Число слагаемых

равно числу способов разместить событие A 2 раза на 3 местах, т.е. равно числу $C_3^2 = 3$. При этом вероятность каждого слагаемого равна $p^2(1-p) = p^2q$. Поэтому

$$P_3(2) = 3 \cdot \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} = \frac{5}{72},$$

что соответствует формуле $C_3^2 p^2 q$.

Приведем обоснование формулы Бернулли (2.1).

В общем случае появление события A ровно m раз в n испытаниях задается суммой, в которой каждое слагаемое является произведением и содержит сомножителем событие A m раз, а событие $\overline{A}$ — $(n-m)$ раз. Слагаемые являются попарно несовместными событиями, а вероятность каждого из них в силу независимости сомножителей равна $p^m q^{n-m}$. Число слагаемых равно числу способов разместить A на m местах из всех n мест, т.е. C_n^m. Поэтому имеем $P_n(m) = C_n^m p^m q^{n-m}$.

Отметим, что

$$\sum_{m=0}^{n} P_n(m) = 1.$$

Пример 2.2. Вероятность выигрыша по лотерейному билету равна 0,05. Какова вероятность того, что среди купленных 10 билетов окажется два выигрышных?

Требуется найти вероятность двух успехов из $n = 10$ испытаний Бернулли с вероятностью успеха в одном испытании $p = 0{,}05$. По формуле Бернулли эта вероятность равна

$$C_{10}^2 \cdot 0{,}05^2 \cdot 0{,}95^8 = 0{,}0746.$$

Отметим, что в n испытаниях Бернулли всегда имеется одно или два числа успехов $\bar{m}$, вероятность которых больше вероятностей всех остальных вероятностей чисел успехов ($P(\bar{m}) > P_n(m)$). Эти числа удовлетворяют неравенствам

$$np - q \le \bar{m} \le np + p, \tag{2.2}$$

из которых их легко найти.

Поскольку разность между концами промежутка равна 1, на этом промежутке найдется или ровно одно целое число $\bar{m}$, или два.

Пример 2.3. Всхожесть семян данного растения имеет вероятность 0,8. Найти наиболее вероятное число проросших семян из 5 посеянных.

По формуле (2.2) наиболее вероятное число проросших семян удовлетворяет неравенствам

$$5 \cdot 0{,}8 - 0{,}2 \le \bar{m} \le 5 \cdot 0{,}8 + 0{,}8,$$

поэтому $\bar{m} = 4$.

Для сравнения вычислим по формуле Бернулли вероятности числа проросших семян:

$$P_5(0) = q^5 = 0{,}2^5 = 0{,}00032,$$

$$P_5(1) = C_5^1 pq^4 = 5 \cdot 0{,}8 \cdot 0{,}2^4 = 0{,}0064,$$

$$P_5(2) = C_5^2 p^2 q^3 = 6{,}4 \cdot 0{,}008 = 0{,}0512,$$

$$P_5(3) = C_5^3 p^3 q^2 = 10 \cdot 0{,}8^3 \cdot 0{,}2^2 = 0{,}2048,$$

$$P_5(4) = C_5^4 p^4 q = 5 \cdot 0{,}4096 \cdot 0{,}2 = 0{,}4096,$$

$$P_5(5) = C_5^5 p^5 q^0 = 0{,}3264.$$

Наиболее вероятное число успехов равно 4 и имеет вероятность 0,4096.

Пример 2.4. В каждом из двух карманов лежит по коробку спичек, в каждом из которых по 10 спичек. При каждом закуривании карман выбирается наудачу. При очередном закуривании коробок оказался пустым. Найти вероятность того, что в другом коробке 6 спичек.

Обозначим через A событие, состоящее в том, что при очередном закуривании мы выбрали первый карман. Тогда событие $\bar{A}$ означает, что выбран второй карман. Поскольку выбор кармана проводится наудачу, вероятность события A равна 0,5. Если во втором кармане осталось 6 спичек, то мы израсходовали 14 спичек, т.е. провели 14 независимых испытаний. Искомое событие означает появление A в 10 испытаниях из 14. По формуле Бернулли

$$P_{14}(10) = C_{14}^{10} \cdot 0{,}5^{10} \cdot 0{,}5^4 = 0{,}061.$$

2.2. Формула Пуассона

Для больших значений n формула Бернулли слишком громоздка. Например, для $n = 500$, $p = 0{,}02$, $m = 10$ имеем

$$P_{500}(10) = C_{500}^{10} \cdot 0{,}02^{10} \cdot 0{,}98^{490}.$$

Если в схеме испытаний Бернулли число испытаний n велико, вероятность p события A в одном испытании мала, а произведение $\lambda = np$ не превосходит 20, то вероятность $P_n(m)$ того, что событие A появится ровно в m испытаниях, находится по приближенной ***формуле Пуассона***:

$$P_n(m) \cong \frac{e^{-\lambda}\lambda^m}{m!} = p(m), \quad m = 0, ..., n. \tag{2.3}$$

Значения $P_n(m)$ имеются в таблицах распределения Пуассона.

Приближенное равенство (2.3) будет достаточно точным при условиях $npq < 10$, $p < 0{,}1$.

Пример 2.5. В ходе проверки аудитор случайным образом отбирает 10 счетов. Найти вероятность того, что он обнаружит один счет с ошибкой, если в среднем 3 % счетов содержат ошибки.

Используя формулу (2.3), в которой $n = 10$, $p = 0{,}03$, $\lambda = np = 0{,}3$, имеем

$$P_{10}(1) \approx \frac{e^{-\lambda}\lambda^1}{1!}.$$

По таблицам значений распределения Пуассона находим $P_{10}(1) \approx \approx 0{,}2223$.

2.3. Локальная формула Муавра — Лапласа

При больших n вероятность $P_n(m)$ появления события A ровно m раз в схеме из n испытаний Бернулли можно вычислять по приближенной ***формуле Муавра — Лапласа***:

$$P_n(m) \approx \frac{1}{\sqrt{npq}}\varphi(x), \tag{2.4}$$

где $x = \dfrac{m - np}{\sqrt{npq}}$, а $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ — малая функция Лапласа, или функция Гаусса, значения которой имеются в соответствующих таблицах. При этом следует отметить, что $\varphi(-x) = \varphi(x)$, $\varphi(x) \approx 0$ при $|x| > 4$.

Приближенное равенство (2.4) будет достаточно точным при условиях $npq \geq 10$, а вероятность p близка к 0,5.

Пример 2.6. Вероятность найти белый гриб среди прочих равна 0,25. Какова вероятность того, что среди 80 грибов белых будет 20?

По формуле (2.4) при $n = 80$, $m = 20$ имеем:

$$x = \frac{20 - 20}{\sqrt{80 \cdot 0{,}25 \cdot 0{,}75}} = 0, \quad \varphi(0) = 0{,}3989.$$

Тогда

$$P_n(m) \approx \frac{1}{\sqrt{80 \cdot 0{,}25 \cdot 0{,}75}} \cdot \varphi(0) = 0{,}103.$$

2.4. Интегральная формула Муавра — Лапласа

Вероятность того, что число m появления события A в схеме Бернулли находится в заданном промежутке $a \leq m \leq b$, при большом числе испытаний n приблизительно равна

$$P(a \leq m \leq b) \approx \frac{1}{2}\Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \frac{1}{2}\Phi\left(\frac{a - np}{\sqrt{npq}}\right), \qquad (2.5)$$

где

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}} dt \qquad (2.6)$$

— функция Лапласа, значения которой находятся по таблицам. При этом следует отметить, что $\Phi(-x) = -\Phi(x)$, $\Phi(x) \approx 1$ при $|x| > 5$.

Замечание. Наряду с формулой (2.6) в учебниках функция Лапласа часто определяется формулой, в которой в числителе вместо 2 стоит 1. Поэтому при пользовании такими таблицами значения функции $\Phi(x)$ нужно удвоить.

Отметим два важных следствия из формулы (2.5).

Для частости m/n появления события A в n испытаниях Бернулли справедлива приближенная формула

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right), \tag{2.7}$$

а для числа m появлений события A —

$$P\left(|m - np| \le \varepsilon\right) \approx \Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right). \tag{2.8}$$

Задачи с решениями

Задача 2.1. В партии из 768 арбузов каждый арбуз оказывается неспелым с вероятностью $q = 0{,}25$. Найти вероятность того, что количество спелых арбузов будет находиться в пределах от 564 до 600.

Решение. Будем считать, что имеется $n = 768$ испытаний Бернулли с вероятностью успеха (зрелый арбуз), равной 0,75. Если m — число успехов, то требуется найти вероятность $P(a \le m \le b)$, где $a = 564$, $b = 600$.

Поскольку

$$\frac{b - np}{\sqrt{npq}} = \frac{600 - 768 \cdot 0{,}75}{\sqrt{768 \cdot 0{,}75 \cdot 0{,}25}} = \frac{24}{11{,}97} = 2{,}0,$$

$$\frac{a - np}{\sqrt{npq}} = \frac{564 - 576}{11{,}97} = -1{,}03,$$

то по формуле (2.5) получим (используя таблицу значений функции Лапласа)

$$P(564 \le m \le 600) = \frac{1}{2}\Phi(2) - \frac{1}{2}\Phi(-1) =$$

$$= \frac{1}{2} \cdot 0{,}9545 + \frac{1}{2} \cdot 0{,}6827 = 0{,}8186.$$

Задача 2.2. Сколько нужно произвести бросаний монеты, чтобы с вероятностью 0,99 можно было утверждать, что частость выпадения герба отличается от 0,5 по модулю не более чем на $\varepsilon = 0{,}001$?

Решение. Используем формулу (2,7), в которой $p = 1/2$, $q = 1/2$, $\varepsilon = 0{,}001$. Из условия

$$P\left(\left|\frac{m}{n} - p\right| \le 0{,}001\right) = 0{,}99$$

получим

$$\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) = 0{,}99.$$

По значению функции Лапласа из таблиц находим соответствующий аргумент и получаем, что

$$\varepsilon\sqrt{\frac{n}{pq}} = 2{,}58,$$

тогда

$$n = \frac{2{,}58^2 \cdot 0{,}5 \cdot 0{,}5}{0{,}001^2} = 16\ 641.$$

Задача 2.3. Кандидата в высший орган власти поддерживает 80 % населения. В каких пределах с вероятностью 0,95 находится число проголосовавших «за» на выборах, если число избирателей равно 1 200 000?

Решение. Используем формулу (2.8), в которой $n = 1\ 200\ 000$, $p = 0{,}8$, $q = 0{,}2$:

$$P\left(|m - np| \le \varepsilon\right) = 0{,}95 = \Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

По значению функции Лапласа из таблиц находим, что

$$\frac{\varepsilon}{\sqrt{npq}} = 1{,}96.$$

Тогда

$$\varepsilon = 1{,}96\sqrt{npq} = 1{,}96 \cdot 490 = 960{,}4,$$

$$|m - np| = |m - 960\ 000| \le 960{,}4,$$

$$960\ 000 - 960{,}4 \le m \le 960\ 000 + 960{,}4,$$

$$959\ 039 \le m \le 960\ 961.$$

Задача 2.4. В страховой компании 10 000 клиентов. Страховой взнос каждого из них составляет 500 руб. При наступлении страхового случая, вероятность которого можно считать равной $p = 0{,}005$, страховая компания выплачивает клиенту 50 000 руб. На какую минимальную прибыль может рассчитывать страховая компания с надежностью 0,95?

Решение. Пусть m — число страховых случаев. Прибыль компании d равна разности между общим страховым взносом и суммарной выплатой, т.е.

$$d = 10\,000 \cdot 500 - m \cdot 50\,000.$$

Минимум прибыли соответствует максимальному числу m_0 страховых случаев — такому, что вероятность

$$P_{10\,000}(0 \le m \le m_0) = 0{,}95.$$

Эту вероятность можно найти по формуле Муавра — Лапласа (условие $npq = 10\,000 \cdot 0{,}005 \cdot 0{,}995 = 49{,}75 \ge 20$ выполнено). Тогда

$$P_{10\,000}(0 \le m \le m_0) = \frac{1}{2}\Phi(x_2) - \frac{1}{2}\Phi(x_1) = 0{,}95,$$

где $x_2 = \dfrac{m_0 - np}{\sqrt{npq}}$, $x_1 = \dfrac{0 - np}{\sqrt{npq}} = -7{,}09$.

Отсюда

$$m_0 = np + x_2\sqrt{npq} = 50 + x_2\sqrt{49{,}75}.$$

По таблицам

$$\Phi(x_1) = \Phi(-7{,}09) \approx -1.$$

Теперь имеем $\Phi(x_2) = 1{,}9 + \Phi(x_1) = 0{,}9$. По таблицам находим значение аргумента $x_2 = 1{,}645$. Следовательно,

$$m_0 = 50 + 1{,}645\sqrt{49{,}75} = 61{,}6,$$

а минимальная гарантированная прибыль

$$d = 5\,000\,000 - 61{,}6 \cdot 50\,000 = 1\,920\,000 \text{ руб.}$$

Задачи для самостоятельного решения

2.1. Вероятность банкротства одной из 6 фирм к концу года равна $p = 0{,}2$. Какова вероятность того, что к концу года обанкротится не более двух фирм?

2.2. Торговый агент в среднем контактирует с 8 потенциальными покупателями в день. Из опыта ему известно, что вероятность того, что потенциальный покупатель совершит покупку, равна 0,1. Чему равна вероятность того, что у агента будут хотя бы две продажи в течение дня?

2.3. В порту каждые сутки может появиться одно большегрузное судно с вероятностью 1/6. Вероятность появления более одного судна в течение суток пренебрежимо мала. Какова вероятность того, что за месяц (30 дней) порт посетит не более 4 судов?

2.4. Телефонная станция обслуживает 2000 абонентов. Вероятность позвонить любому абоненту в течение часа равна 0,003. Какова вероятность того, что в течение часа позвонят 5 абонентов?

2.5. Найти вероятность того, что при подбрасывании монеты 100 раз герб появится ровно 60 раз.

2.6. Для поступления в университет необходимо успешно сдать вступительные экзамены. В среднем их выдерживают 25 % абитуриентов. Предположим, что в приемную комиссию поступило 1800 заявлений. Чему равна вероятность того, что хотя бы 450 поступающих успешно сдадут экзамены (наберут проходной балл)?

2.7. Найти такое число m, чтобы с вероятностью 0,95 можно было бы утверждать, что среди 800 новорожденных будет более, чем m девочек. Считается, что вероятность рождения девочки равна 0,485.

2.8. Установлено, что виноградник поражен вредителями в среднем на 10 %. Определить вероятность того, что из 10 проверенных кустов винограда один будет поражен. Вычислить вероятности по формулам Бернулли, Лапласа, Пуассона. Сравнить результаты. Сделать выводы.

2.9. На факультете 900 студентов. Вероятность дня рождения каждого студента в данный день равна 1/365. Найти вероятность того, что найдутся ровно 3 студента с днями рождения: а) 8 марта, б) в любой день года.

2.10. В автопарке 70 машин. Вероятность поломки машины равна $p = 0{,}2$. Найти наиболее вероятное число исправных автомобилей и вероятность этого числа.

2.11. При проведении некоторого опыта вероятность появления ожидаемого результата равна 0,01. Сколько раз надо провести опыт, чтобы с вероятностью 0,5 можно было бы ожидать хотя бы одного появления этого результата?

2.12. Вероятность появления события в каждом из 400 независимых испытаний равна 0,8. Найти такое положительное число ε, чтобы с вероятностью 0,9876 абсолютная величина отклонения относительной частоты появления события от вероятности 0,8 не превысила ε.

2.13. На симпозиум приглашены 75 человек, причем каждый из них прибывает с вероятностью 0,8. В гостинице для гостей заказано 65 мест. Какова вероятность, что все приезжающие будут поселены в гостинице?

2.14. Производится 500 бросаний симметричной монеты. В каких пределах будет находиться отклонение частости выпадения герба от 0,5 с вероятностью 0,99?

2.15. На факультете 20 % студентов — выходцы из сельской местности. В каких пределах с вероятностью 0,9 находится число городских жителей среди студентов из группы в 28 человек?

2.16. До 72-летнего возраста в среднем доживает половина мужского населения города. Какова вероятность того, что из 5 товарищей двое доживут до этого возраста?

2.17. Сколько надо произвести бросаний симметричной монеты, чтобы с вероятностью 0,9 отклонение частости выпадения герба отличалось от 0,5 не более чем на 0,01?

2.18. При передаче текстовой информации слова кодируются в символы. Вероятность искажения каждого символа при передаче равна 0,009. При искажении двух и более символов слово не поддается дешифровке. Найти вероятность того, что слово, содержащее 10 символов, будет принято правильно.

Ответы

2.1. $P(m \le 2) = 0{,}9421$.
2.2. $p = 0{,}1868$.
2.3. $P(m \le 4) = 0{,}4405$.
2.4. Среднее число позвонивших в течение часа $n = 6$,

$$p_5 = \frac{6^5 \cdot e^{-6}}{5!} \approx 0{,}13.$$

2.5. $p = 0{,}0108$.
2.6. $P(m \ge 450) = 0{,}5$.
2.7. $m = 364$.
2.8. $p = 0{,}387$; $p = 0{,}420$; $p = 0{,}368$.
2.9. а) $p = 0{,}212$; б) $p = 0{,}212$.
2.10. $m = 56$; $p = 0{,}115$.
2.11. $m = 69$.
2.12. $\varepsilon = 0{,}05$.
2.13. $p = 0{,}962$.
2.14. $\left|\frac{m}{n} - p\right| \le 0{,}052$.
2.15. $18 \le m \le 26$.
2.16. $p = 0{,}3125$.
2.17. $n = 9604$.
2.18. $p = 0{,}9965$.

3

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ЗАКОНЫ ИХ РАСПРЕДЕЛЕНИЯ

Важнейшим понятием теории вероятности является понятие случайной величины. *Случайной величиной* называется числовая величина, значения которой зависят от исхода опыта или эксперимента. Таким образом, случайную величину можно понимать как числовую функцию, аргументами которой служат элементарные случайные события из заданного пространства событий. Хотя случайная величина определяется как функция, в теории вероятностей, в отличие от математического анализа, для случайных величин используются другие свойства функций. Особенно важными являются значения случайных величин и вероятности событий, для которых эти величины принимают данные значения.

3.1. Общие сведения о случайных величинах

Пусть Ω — пространство элементарных событий. Пусть на этом пространстве задана числовая функция X, сопоставляющая каждому элементарному событию $\omega \in \Omega$ число $X(\omega)$. Для функции X и всякого действительного числа x в пространстве Ω определяется подмножество $(X < x) = \{\omega \in \Omega \mid X(\omega) < x\}$, состоящее из всех элементарных событий ω, для которых значение функции $X(\omega)$ меньше числа x. Дадим более строгое определение случайной величины.

Случайной величиной называется такая числовая функция X на пространстве событий Ω, что для всякого числа x подмножество $(X < x)$ из Ω является событием, т.е. входит в число подмножеств, составляющих σ-алгебру пространства Ω.

Пример 3.1. При бросании игральной кости пространство элементарных событий Ω состоит из 6 элементов. Пусть X — функция на этом пространстве, значения которой равны числу выпавших очков. Эта функция

является случайной величиной, так как всякое подмножество из Ω является событием.

Если пространство элементарных событий Ω состоит из конечного числа элементов, то всякая функция X на этом пространстве является случайной величиной, поскольку в этом случае любое подмножество из Ω является событием.

Пример 3.2. Пусть бросаются две игральные кости, элементарными событиями при этом являются выпавшие пары количества очков. Пространство элементарных событий Ω можно отождествить с набором пар чисел $\{(i, j) \mid i, j = 1, 2, ..., 6\}$.

Пусть X – функция, которая каждой паре выпавших очков сопоставляет их сумму, тогда случайная величина X при выпадении пары (i, j) принимает значение, равное сумме $i + j$. Например, величина X принимает значение 6 при выпадении следующих 5 пар: (1, 5), (2, 4), (3, 3), (4, 2), (5, 1). Поскольку вероятность выпадения каждой из пар (i, j) равна $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$, величина X принимает значение 6 с вероятностью $\frac{5}{36}$. В следующей таблице приведены все возможные значения x_k, принимаемые величиной X (первая строка таблицы), и вероятности p_k этих значений (вторая строка таблицы):

x_k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Одним из способов задания случайной величины служит функция распределения. *Функцией распределения* случайной величины X называется функция $y = F(x)$, значения которой определяются формулой

$$F(x) = P(X < x), \tag{3.1}$$

где $P(X < x)$ — вероятность события $(X < x)$. Значение функции распределения в точке x равно вероятности того, что случайная величина примет значение меньшее, чем x.

Пример 3.3. Для случайной величины, равной количеству очков при бросании игральной кости, значение функции распределения в точке 7 равно 1, так как число выпавших очков всегда меньше 7, а событие $(X < 7)$ является достоверным.

Законом распределения случайной величины X называются ее значения вместе с соответствующими вероятностями (с функцией распределения).

При использовании случайных величин само пространство событий Ω уходит на задний план. Часто оно вообще не поддается описанию и зависит от многих факторов. Однако при применении случайной величины основную роль играют ее значения и вероятности этих значений. Одним из основных способов задания случайных величин является задание функции распределения $F(x)$. Различают два основных типа случайных величин: дискретные и непрерывные. Множество значений дискретной случайной величины конечно или счетно, а множество значений непрерывной случайной величины заполняет всю числовую прямую или промежуток на ней.

3.2. Дискретные случайные величины и операции над ними

Случайная величина X на пространстве Ω называется ***дискретной***, если множество ее значений является конечным или счетным.

Если само пространство Ω является конечным или счетным множеством, то любая функция на Ω будет дискретной случайной величиной.

Для заданной дискретной случайной величины X, принимающей значения $x_1, x_2, x_3, \ldots$, введем события

$$(X = x_i) = \{\omega \in \Omega \mid X(\omega) = x_i\}, \quad i = 1, 2, \ldots .$$

Распределением дискретной случайной величины X называется совокупность ее значений $x_1, x_2, \ldots, x_n, \ldots$ вместе с вероятностями событий $(X = x_1), (X = x_2), \ldots, (X = x_n), \ldots$. Распределение дискретной случайной величины задается в виде таблицы:

x_i	x_1	x_2	...	x_n	...
p_i	p_1	p_2	...	p_n	...

В этой таблице $p_i = P(X = x_i)$ — это вероятность события $(X = x_i)$, $i = 1, 2, \ldots$. Сумма всех этих вероятностей равна 1, $\sum p_i = 1$, поскольку в первой строке таблицы задаются все возможные значения X.

В примере 2.3 дан пример случайной величины X — числа растений, появившихся из 5 посеянных семян. Закон распределения этой величины X имеет вид:

x_k	0	1	2	3	4	5
p_k	0,01024	0,0768	0,2304	0,3456	0,2592	0,07776

Две дискретные случайные величины X и Y называются ***независимыми***, если события $(X = x_i)$ и $(Y = y_j)$, $i, j = 1, 2, \ldots$, независимы в совокупности. Это, в частности, означает (по теореме умножения вероятностей), что

$$P\left((X = x_i) \cdot (Y = y_j)\right) = P(X = x_i)P(Y = y_j), \quad i, j = 1, 2, \ldots . \quad (3.2)$$

Пусть даны две случайные величины X и Y, при этом закон распределения случайной величины X имеет вид

$$p_i = P(X = x_i), \quad (3.3)$$

а случайной величины Y —

$$q_j = P(Y = y_j). \quad (3.4)$$

Тогда можно ввести новые случайные величины. Величина $Z = kX$, где k — постоянная величина. Есть также случайная величина, которая принимает значения kx_i с теми же самыми вероятностями, что и случайная величина X принимает значения x_i, т.е.

$$P(Z = kx_i) = p_i. \quad (3.5)$$

Закон распределения случайной величины Z может быть записан в виде таблицы:

z_i	kx_1	kx_2	...	kx_n	...
p_i	p_1	p_2	...	p_n	...

3.2.1. Операции над случайными величинами

Сумма случайных величин X и Y — это новая случайная величина $Z = X + Y$, принимающая все значения вида $z_k = x_i + y_j$ с вероятностями

$$r_k = \sum_{i,j,(x_i + y_j = z_k)} p_{ij}, \tag{3.6}$$

где $p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big)$.

Если при этом случайные величины X и Y независимы, то

$$p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big) = p_i q_j. \tag{3.7}$$

Разностью случайных величин называется новая случайная величина $Z = X - Y$, для которой

$$r_k = P(Z = z_k) = \sum_{i,j,(x_i - y_j = z_k)} p_{ij}, \tag{3.8}$$

где $p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big)$.

Пример 3.4. Пусть случайные величины X и Y независимы и заданы законами распределения:

x_i	0	1	2	3
p_i	0,2	0,4	0,3	0,1

y_j	−1	0	1
q_j	0,4	0,2	0,4

Найти закон распределения случайной величины $Z = X - Y$.

Все возможные значения разности случайных величин $X - Y$ составляют множество чисел {−1, 0, 1, 2, 3, 4}. Определим их вероятности.

$$P(Z = -1) = P\big[(X = 0) \cdot (Y = 1)\big] = p_{13} = p_1 q_3 = 0{,}2 \cdot 0{,}4 = 0{,}08,$$

$$P(Z = 0) = P\big[(X = 0) \cdot (Y = 0)\big] + P\big[(X = 1) \cdot (Y = 1)\big] =$$

$$= p_{12} + p_{23} = 0{,}2 \cdot 0{,}2 + 0{,}4 \cdot 0{,}4 = 0{,}2,$$

$$P(Z = 1) = p_{11} + p_{22} + p_{33} = 0{,}2 \cdot 0{,}4 + 0{,}4 \cdot 0{,}2 + 0{,}3 \cdot 0{,}4 = 0{,}28,$$

$$P(Z = 2) = p_{21} + p_{32} + p_{43} = 0{,}4 \cdot 0{,}4 + 0{,}3 \cdot 0{,}2 + 0{,}1 \cdot 0{,}4 = 0{,}26,$$

$$P(Z = 3) = p_{31} + p_{42} = 0{,}3 \cdot 0{,}4 + 0{,}1 \cdot 0{,}2 = 0{,}14,$$

$$P(Z = 4) = p_{41} = 0{,}04.$$

Закон распределения случайной величины $Z = X - Y$ задается таблицей

z_k	–1	0	1	2	3	4
r_k	0,08	0,2	0,28	0,26	0,14	0,04

Произведением случайных величин называется новая случайная величина $Z = XY$, принимающая все значения вида $z_k = x_i y_j$ с вероятностями

$$r_k = \sum_{i,j,(x_i y_j = z_k)} p_{ij}, \tag{3.9}$$

где $p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big)$.

3.2.2. Математическое ожидание и дисперсия дискретной случайной величины

Часто множество значений случайной величины и их вероятностей труднообозримы в целом. Поэтому для описания поведения случайной величины используются ее обобщенные характерные признаки. Основными из них являются математическое ожидание и дисперсия.

Математическое ожидание MX дискретной случайной величины X определяется формулой

$$MX = \sum_k x_k p_k \tag{3.10}$$

и имеет смысл среднего значения случайной величины. Так, если число значений случайной величины конечно и равно n, а вероятности $p_k = \dfrac{1}{n}$, $k = 1, 2, ..., n$, то MX совпадает с обычным средним значением величин $x_1, x_2, ..., x_n$:

$$MX = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

Математическое ожидание обладает следующими свойствами:

1. $MC = C$, где $C = \text{const}$.
2. $M(CX) = C \cdot MX$, где $C = \text{const}$.
3. $M(X \pm Y) = MX \pm MY$, для любых X и Y.
4. $M(X \cdot Y) = MX \cdot MY$, если X и Y независимы.

Для оценки *степени рассеяния* значений случайной величины вокруг ее среднего значения вводится понятие дисперсии.

Дисперсией DX случайной величины X называется математическое ожидание квадрата разности $(X - MX)^2$ (квадрата отклонения величины от ее среднего значения):

$$DX = M(X - MX)^2 = \sum_k (x_k - MX)^2 p_k. \qquad (3.11)$$

Для вычисления дисперсии обычно используют более удобную формулу

$$DX = MX^2 - (MX)^2. \qquad (3.12)$$

Свойства дисперсии:

1. $DC = 0$, где $C = \text{const}$.
2. $D(CX) = C^2 DX$, где $C = \text{const}$.
3. $D(X \pm Y) = DX \pm DY$, если X и Y независимы.

Дисперсия произведения случайных величин будет вычисляться далее, после введения ковариации случайных величин. Там же свойства математического ожидания и дисперсии будут дополнены.

Средним квадратичным отклонением случайной величины X называется величина

$$\sigma X = \sqrt{DX}, \qquad (3.13)$$

которая также является мерой рассеяния случайной величины X.

Дисперсия DX имеет размерность квадрата размерности случайной величины X, величина же σX имеет ту же размерность, что и сама X. Иногда вместо σX пишут σ_X, а вместо $F(x)$ пишут $F_X(x)$.

Пример 3.5. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение случайной величины X — числа взошедших семян из примера 3.1.

По формуле (3.10) имеем

$$MX = 0 \cdot 0{,}01024 + 1 \cdot 0{,}0768 + 2 \cdot 0{,}2304 + 3 \cdot 0{,}3456 +$$
$$+ 4 \cdot 0{,}2592 + 5 \cdot 0{,}07776 = 3.$$

Для вычисления дисперсии сначала найдем

$$MX^2 = 0^2 \cdot 0{,}01024 + 1^2 \cdot 0{,}0768 +$$
$$+ 2^2 \cdot 0{,}2304 + 3^2 \cdot 0{,}3456 + 4^2 \cdot 0{,}2592 + 5^2 \cdot 0{,}07776 = 10{,}2.$$

По формуле (3.12) получим

$$DX = 10{,}2 - 9 = 1{,}2.$$

Тогда

$$\sigma X = \sqrt{1{,}2} \cong 1{,}095.$$

3.2.3. Функция распределения дискретной случайной величины

Для дискретной случайной величины X с законом распределения $P(X = x_k) = p_k$, $k = 1, 2, ...$, событие $(X < x)$, где x — некоторое число, состоит из всех событий $(X = x_k)$, для которых $x_k < x$. Поэтому согласно формуле (3.1) *функция распределения* случайной величины X имеет вид

$$F(x) = \sum_{k, x_k < x} p_k. \qquad (3.14)$$

Свойства функции распределения дискретной случайной величины:

1. Функция распределения монотонно не убывает.
2. Функция распределения непрерывна слева.
3. $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

Функция распределения дискретной случайной величины является кусочно-постоянной функцией. Скачки функции $F(x)$ находятся в точках разрыва $x = x_k$ и равны вероятностям $p_k = P(X = x_k)$, с которыми величина X принимает значения x_k. Поэтому по графику функции $F(x)$ легко восстановить таблицу распределения вероятностей случайной величины X.

Пример 3.6. Партия изделий содержит 10 % нестандартных. Пусть случайная величина X — число стандартных изделий в выборке из 5 изделий. Найти функцию распределения X и вероятность $P(X > 1)$. Построить график функции $F(x)$.

Случайная величина X может принимать значения $x_k = 0, 1, 2, 3, 4, 5$. Найдем их вероятности по формуле Бернулли $p_k = C_n^k p^k q^{n-k}$, где $n = 5$, $p = 0{,}9$, $q = 0{,}1$. Вычислив вероятности, получим: $p_0 = 0{,}00001$, $p_1 = 0{,}00045$, $p_2 = 0{,}0081$, $p_3 = 0{,}0729$, $p_4 = 0{,}32085$, $p_5 = 0{,}59049$.

По формуле (3.14) имеем

$$F(x) = P(X < x) = \sum_{k, x_k < x} p_k.$$

При $x \le 0$ очевидно, $F(x) = 0$,
при $0 < x \le 1$ $F(x) = p_0 = 0{,}00001$,
при $1 < x \le 2$ $F(x) = p_0 + p_1 = 0{,}00046$,
при $2 < x \le 3$ $F(x) = p_0 + p_1 + p_2 = 0{,}00856$,
при $3 < x \le 4$ $F(x) = p_0 + p_1 + p_2 + p_3 = 0{,}08146$,
при $4 < x \le 5$ $F(x) = p_0 + p_1 + p_2 + p_3 + p_4 = 0{,}40951$,
при $x > 5$ $F(x) = 1$.

График функции распределения $F(x)$ изображен на рис. 3.1.

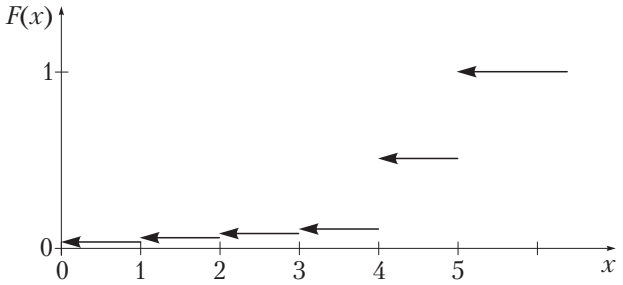


Рис. 3.1. График функции распределения дискретной случайной величины

Вероятность

$$P(X > 1) = 1 - P\big[(X = 0) + (X = 1)\big] = 1 - 0{,}00046 = 0{,}99954.$$

3.3. Непрерывные случайные величины

Непрерывной называется случайная величина X, для которой функция распределения непрерывна. Как уже отмечалось, множество значений такой величины заполняет целый промежуток.

Пример 3.7. Интервал движения метро равен 5 мин. Пусть X — время ожидания поезда пассажиром на станции, если появление пассажира равновозможно в любой момент. Величина X принимает значения от 0 до 5 мин.

Из определения функции распределения следует, что для любых чисел $a < b$ имеет место равенство

$$P(a \le X < b) = F(b) - F(a). \tag{3.15}$$

Из этой формулы следует, что вероятность каждого конкретного значения непрерывной случайной величины равна нулю:

$$P(X = x) = 0. \tag{3.16}$$

Действительно, для любого значения x случайной величины X представим $P(X = x)$ в виде

$$P(X = x) = \lim_{b \to x} P(x \le X < b) = \lim_{b \to x} \big[F(b) - F(x)\big] =$$

$$= \lim_{b \to x} F(b) - F(x) = F(x) - F(x) = 0$$

($\lim\limits_{b \to x} F(b) = F(x)$ в силу непрерывности функции F).

Распределением вероятностей непрерывной случайной величины называется совокупность вероятностей

$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

для любых действительных чисел $a < b$.

Распределение вероятностей непрерывной случайной величины однозначно определяется ее функцией распределения $F(x)$ в силу равенства (3.15).

Свойства 1, 2, 3 функции распределения дискретных случайных величин выполняются и для функции распределения непрерывных случайных величин. Ее графиком будет непрерывная линия вида кривой, изображенной на рис. 3.2.

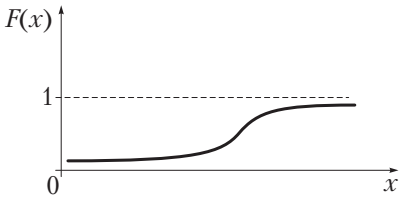


Рис. 3.2. График функции распределения непрерывной случайной величины

Теперь приведем другой способ задания непрерывных случайных величин.

Если существует такая неотрицательная функция $p(x) \geq 0$, что

$$F(x) = \int_{-\infty}^{x} p(t)dt, \tag{3.17}$$

то говорят, что случайная величина X имеет *плотность* $p(x)$.

Если случайная величина X имеет плотность, то ее функция распределения $F(x)$ является дифференцируемой, а сама величина X называется ***абсолютно непрерывной*** случайной величиной.

Плотность $p(x)$ однозначно определяет распределение случайной величины, поскольку

$$P(a \leq X \leq b) = \int_{a}^{b} p(x)dx. \tag{3.18}$$

Заметим, что

$$\int_{-\infty}^{\infty} p(x)dx = F(+\infty) = 1. \tag{3.19}$$

Из формулы (3.16) по свойству производной от интеграла с переменным верхним пределом следует, что

$$F'(x) = \frac{d}{dx}\int_{-\infty}^{x} p(t)dt = p(x). \tag{3.20}$$

Так что, если функция $F(x)$ дифференцируема, то соответствующая случайная величина имеет плотность, равную производной

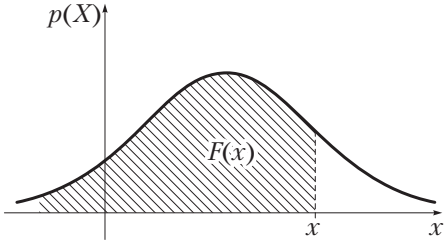


Рис 3.3. Связь функции распределения с функцией плотности

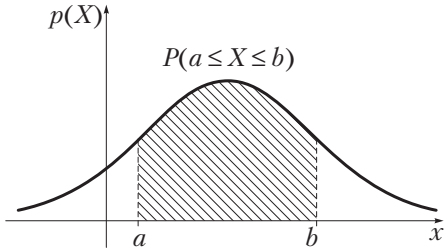


Рис. 3.4. Задание вероятностей функцией плотности

функции распределения. Свойства функции плотности и ее связь с функцией распределения отображены на рис. 3.3 и 3.4.

Для дискретной случайной величины используются два равносильных способа задания: с помощью функции распределения или таблицы распределения.

Для непрерывной случайной величины также используются два равносильных способа задания: с помощью функции распределения или функции плотности вероятностей.

3.4. Математическое ожидание и дисперсия непрерывной случайной величины

Математическое ожидание MX непрерывной случайной величины, имеющей плотность, определяется формулой

$$MX = \int_{-\infty}^{+\infty} xp(x)dx, \tag{3.21}$$

а дисперсия —

$$DX = M(X - MX)^2 = \int_{-\infty}^{+\infty} (x - MX)^2 p(x)dx. \qquad (3.22)$$

Для вычисления дисперсии также верна формула (3.12), откуда для непрерывной случайной величины получаем

$$DX = \int_{-\infty}^{+\infty} x^2 p(x)dx - (MX)^2. \qquad (3.23)$$

Все свойства математического ожидания и дисперсии для дискретных случайных величин переносятся и на случай непрерывных случайных величин.

Замечание. При более детальном описании распределений вероятностей случайных величин наряду с математическим ожиданием и дисперсией используются начальные и центральные моменты высших порядков:

$$\nu_k = M(X^k), \quad \mu_k = M(X - MX)^k, \quad k = 1, 2, \ldots .$$

Задачи с решениями

Задача 3.1. Пусть вероятность попадания при одном выстреле равна $p = 0{,}8$, X — случайная величина, значение которой — число выстрелов до первого попадания. Найти закон распределения этой случайной величины.

Решение. Нам надо определить значения случайной величины X и вероятности, с которыми она принимает эти значения. Очевидно, что величина X может принимать значения 1, 2, 3, ..., n, ..., т.е. эта величина является дискретной со счетным множеством значений. Обозначим через A событие, соответствующее попаданию при очередном выстреле. Тогда первое попадание при n выстрелах соответствует событию

$$\underbrace{\overline{A} \cdot \ldots \cdot \overline{A}}_{n-1} \cdot A.$$

Вероятность этого события в силу независимости сомножителей равна

$$\underbrace{0{,}2 \cdot \ldots \cdot 0{,}2}_{n-1} \cdot 0{,}8 = 0{,}2^{n-1} \cdot 0{,}8.$$

Таблица распределения вероятностей этой величины имеет вид:

x_i	1	2	3	...	n	...
p_i	0,8	0,16	0,032	...	$0{,}2^{n-1} \cdot 0{,}8$	...

Задача 3.2. Найти математическое ожидание и дисперсию случайной величины из задачи 3.1.

Решение. Случайная величина X принимает натуральные значения $k = 1, 2, 3, \ldots$. Событие $(X = k)$ означает, что в начале было $k - 1$ промахов, а при k-м выстреле — попадание. Считая результаты выстрелов независимыми событиями, получим закон распределения X:

$$p_k = P(X = k) = \underbrace{q \cdot q \cdot \ldots \cdot q}_{k-1} \cdot p = q^{k-1}p, \quad p + q = 1.$$

По определению математическое ожидание X

$$MX = \sum_{k=1}^{\infty} kp_k = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}.$$

Для его вычисления рассмотрим равенство

$$\sum_{k=1}^{\infty} q^k = \frac{q}{1-q}$$

(сумма геометрической прогрессии) и продифференцируем его по q:

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}.$$

Отсюда получим

$$MX = p\frac{1}{(1-q)^2} = \frac{1}{p},$$

т.е.

$$MX = \frac{1}{0,8} = 1,25.$$

Для вычисления дисперсии найдем вначале

$$MX^2 = \sum_{k=1}^{\infty} k^2 pq^{k-1}.$$

Умножив

$$\sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$$

на q, имеем

$$\sum_{k=1}^{\infty} kq^{k} = \frac{q}{(1-q)^2}.$$

Продифференцировав это равенство, получим

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{1+q}{(1-q)^3}.$$

Тогда

$$MX^2 = \frac{p(1+q)}{(1-q)^3}$$

и дисперсия

$$DX = MX^2 - (MX)^2 = \frac{2-p}{p^2} - \frac{1}{p^2},$$

т.е.

$$DX = \frac{q}{p^2}.$$

При $p = 0,8$ $DX = 0,3125$.

Задача 3.3. Пусть вероятность попадания при одном выстреле $p = 0,8$, X — случайная величина, значение которой — число выстрелов до первого попадания, когда имеется лишь 4 снаряда. Найти закон распределения этой случайной величины.

Решение. Указанная величина может принимать значения 1, 2, 3 при первом попадании с первого, второго или третьего раза соответственно. Ее значение равно 4, если первое попадание будет при 4-м выстреле или если закончатся снаряды, т.е. при первых четырех выстрелах был неуспех (табл. 3.1, рис. 3.5). Как и в задаче 3.1, $p_1 = 0{,}8$, $p_2 = 0{,}16$, $p_3 = 0{,}032$. Вероятность p_4 равна сумме вероятностей несовместных событий:

$$p_4 = 0{,}2^3 \cdot 0{,}8 + 0{,}2^4 = 0{,}0064 + 0{,}0016 = 0{,}008.$$

Таблица 3.1

x_i	1	2	3	4
p_i	0,8	0,16	0,032	0,008

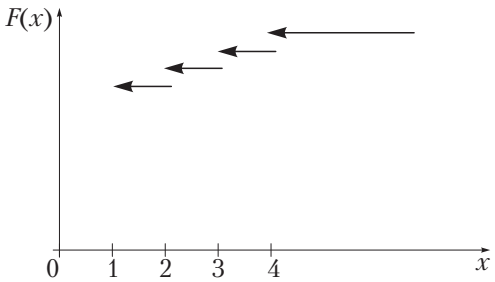


Рис. 3.5. График функции распределения

Задача 3.4. Найти математическое ожидание и дисперсию случайной величины из задачи 3.3.

Решение. По формуле 3.10 из табл. 3.1 получаем:

$$MX = 1 \cdot 0{,}8 + 2 \cdot 0{,}16 + 3 \cdot 0{,}032 + 4 \cdot 0{,}008 = 1{,}248.$$

Дисперсию DX найдем по формуле (3.12):

$$MX^2 = 1 \cdot 0{,}8 + 4 \cdot 0{,}16 + 9 \cdot 0{,}032 + 16 \cdot 0{,}008 = 1{,}856,$$

$$(MX)^2 = 1{,}248^2 = 1{,}5575,$$

$$DX = 1{,}856 - 1{,}5575 = 0{,}2985.$$

Задача 3.5. Пусть случайная величина X имеет функцию распределения $F(x) = c_1 \operatorname{arctg} x + c_2$. Определить константы c_1 и c_2, найти вероятность $p = P(X \geq 1)$.

Решение. Определим c_1 и c_2 из свойств функции распределения:

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

Получим

$$\lim_{x \to -\infty} (c_1 \operatorname{arctg} x + c_2) = -c_1 \frac{\pi}{2} + c_2 = 0,$$

$$\lim_{x \to +\infty} (c_1 \operatorname{arctg} x + c_2) = c_1 \frac{\pi}{2} + c_2 = 1.$$

Решая систему уравнений относительно c_1 и c_2, находим

$$c_1 = \frac{1}{\pi}, \quad c_2 = \frac{1}{2}.$$

Тогда

$$F(x) = \frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2}$$

и $p = P(X \geq 1) = 1 - P(X < 1) = 1 - F(1) = 1 - \left(\frac{1}{\pi} \operatorname{arctg} 1 + \frac{1}{2} \right) = \frac{1}{4}$.

Задача 3.6. Пусть случайная величина X имеет плотность $p(x)$. Найти плотность $g(x)$ случайной величины X^2.

Решение. Пусть $F(x)$ — функция распределения X, $G(x)$ — функция распределения X^2. Тогда $p(x) = F'(x)$, $g(x) = G'(x)$ и при $x > 0$ имеем:

$$G(x) = P(X^2 < x) = P\left(-\sqrt{x} < X < \sqrt{x}\right) = F\left(\sqrt{x}\right) - F\left(-\sqrt{x}\right),$$

$$g(x) = F'\left(\sqrt{x}\right)\frac{1}{2\sqrt{x}} + F'\left(-\sqrt{x}\right)\frac{1}{2\sqrt{x}} = \frac{1}{2\sqrt{x}}\left(p\left(\sqrt{x}\right) + p\left(-\sqrt{x}\right)\right).$$

Для всех $x \leq 0$ $G(x) = 0$ и $g(x) = 0$.

Задача 3.7. Пусть случайная величина X имеет плотность

$$p(x) = \frac{c}{1 + x^2}.$$

Определить константу c и найти вероятность $p = P(|X| < 1)$.

Решение. Определим константу c из условия для плотности

$$\int_{-\infty}^{\infty} p(x)dx = 1.$$

Получим

$$\int_{-\infty}^{\infty} \frac{cdx}{1+x^2} = c\,\mathrm{arctg}\,x\,|_{-\infty}^{\infty} = c\left(\frac{\pi}{2}+\frac{\pi}{2}\right) = c\pi.$$

Тогда

$$c = \frac{1}{\pi}, \quad p(x) = \frac{1}{\pi(1+x^2)}.$$

Так как для любых a и b

$$P(a < x < b) = \int_a^b p(x)dx,$$

то

$$p = P\left(|X| < 1\right) = P(-1 < X < 1) =$$

$$= \int_{-1}^{1} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\,\mathrm{arctg}\,x\,|_{-1}^{1} = \frac{1}{\pi}\left(\frac{\pi}{4}+\frac{\pi}{4}\right) = \frac{1}{2}.$$

Задача 3.8. Пусть случайная величина X имеет плотность

$$p(x) = \frac{6}{11}(x^2 + x + 1)$$

при $0 \le x \le 1$ и $p(x) = 0$ при $x < 0$ и $x > 1$. Найти математическое ожидание и дисперсию X.

Решение. Убедимся, что функция $p(x)$ действительно представляет плотность. Очевидно, что

$$p(x) \ge 0$$

и $$\int_{-\infty}^{\infty} p(x)dx = \frac{6}{11}\int_0^1 (x^2 + x + 1)dx = \frac{6}{11}\left(\frac{1}{3}+\frac{1}{2}+1\right) = 1,$$

так что $p(x)$ удовлетворяет необходимым и достаточным условиям плотности случайной величины. По определению математического ожидания

$$MX = \int_{-\infty}^{\infty} xp(x)dx = \frac{6}{11}\int_0^1 x(x^2 + x + 1)dx = \frac{6}{11}\int_0^1 (x^3 + x^2 + x)dx =$$

$$= \frac{6}{11}\left(\frac{1}{4} + \frac{1}{3} + \frac{1}{2}\right) = \frac{13}{22} = 0{,}5909.$$

Найдем

$$MX^2 = \int_{-\infty}^{+\infty} x^2 p(x)dx = \frac{6}{11}\int_0^1 x^2(x^2 + x + 1)dx =$$

$$= \frac{6}{11}\int_0^1 (x^4 + x^3 + x^2)dx = \frac{6}{11}\left(\frac{1}{5} + \frac{1}{4} + \frac{1}{3}\right) = \frac{47}{110}.$$

Тогда

$$DX = MX^2 - (MX)^2 = \frac{47}{110} - \left(\frac{13}{22}\right)^2 = 0{,}0781.$$

Задача 3.9. Пусть в результате некоторого испытания появляется значение случайной величины X, функция распределения которой имеет вид:

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{x^2}{2}, & 0 < x \le \sqrt{2}, \\ 1, & x > \sqrt{2}. \end{cases}$$

Найти вероятность того, что в результате 4 независимых испытаний случайная величина X трижды примет значение, принадлежащее интервалу $(0, 1)$.

Решение. Искомая вероятность является вероятностью появления события — попадания X в интервал $(0, 1)$, трижды в схеме Бернулли из 4 испытаний. Вероятность попадания X в интервал $(0, 1)$ равна

$$p = F_X(1) - F_X(0) = \frac{1}{2}.$$

Поэтому

$$P_4(3) = C_4^3 p^3 q = 4 \cdot \frac{1}{2^4} = \frac{1}{4}.$$

Задача 3.10. В лотерее имеется 1000 билетов, из них выигрышных: 10 по 500 руб., 50 по 50 руб., 100 по 10 руб., 150 по 1 руб. Найти математическое ожидание выигрыша на один билет.

Решение. Ряд распределения случайной величины X — суммы выигрыша на один билет таков $\left(\sum_{i=1}^{5} p_i = 1\right)$:

X	500	50	10	1	0
p	0,01	0,05	0,1	0,15	0,69

Тогда

$$MX = 500 \cdot 0{,}01 + 50 \cdot 0{,}05 + 10 \cdot 0{,}1 + 1 \cdot 0{,}15 + 0 \cdot 0{,}69 = 8{,}65 \text{ руб.}$$

Задача 3.11. Пусть X — сумма числа очков при двух бросаниях игральной кости. Найти ее математическое ожидание и дисперсию.

Решение. Представим величину X как сумму $X = X_1 + X_2$, где X_1 и X_2 — сумма очков при первом и втором бросании соответственно.

Величины X_1 и X_2 имеют одинаковые законы распределения:

x_k	1	2	3	4	5	6
p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Их математическое ожидание

$$MX_1 = MX_2 = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} =$$

$$= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2},$$

а дисперсия $DX_1 = M(X_1^2)$. Поскольку

$$M(X_1^2) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}, \quad (MX_1)^2 = \frac{441}{36},$$

имеем

$$DX_1 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Теперь

$$MX = MX_1 + MX_2 = 2 \cdot \frac{7}{2} = 7.$$

В силу независимости величин X_1 и X_2 по свойству дисперсии

$$DX = DX_1 + DX_2 = 2 \cdot \frac{35}{12} = \frac{35}{6}.$$

Задачи для самостоятельного решения

3.1. Вероятность рождения в семье мальчика равна 0,515. Составить закон распределения случайной величины X — числа мальчиков в семьях, имеющих 4 детей. Построить таблицу распределения вероятностей. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение.

3.2. Стрелок производит выстрелы по цели до первого попадания или пока не израсходует имеющиеся 5 патронов. Составить закон распределения случайной величины числа выстрелов, сделанных стрелком. Вероятность попадания при каждом выстреле равна 0,7.

3.3. Стрелок производит выстрелы по цели до первого попадания, когда число патронов неограничено. Составить закон распределения случайной величины X — числа выстрелов, сделанных стрелком. Вероятность попадания в цель при каждом выстреле составляет 0,7.

3.4. В группе, состоящей из $2n + 1$ студентов, n девушек. Составить закон распределения случайной величины X — числа девушек из случайно отобранных 3 студентов. Построить график функции распределения.

3.5. Игрок поочередно покупает билеты двух разных лотерей до первого выигрыша. Вероятность выигрыша по одному билету первой лотереи составляет 0,2, а второй — 0,3. Игрок вначале покупает билет первой лотереи. Составить закон распределения и найти математическое ожидание случайной величины X — числа куплен-

ных билетов, если он имеет возможность купить: а) только 5 билетов; б) неограниченное число билетов.

3.6. Вероятность попадания в цель первым стрелком равна 0,9, вторым — 0,8 и третьим — 0,7. Составить закон распределения случайной величины X — числа попаданий в цель, если каждый стрелок производит по одному выстрелу. Определить математическое ожидание случайной величины X.

3.7. Вероятность успешной сдачи экзамена первым студентом составляет 0,7, а вторым — 0,8. Составить закон распределения случайной величины X — числа студентов, успешно сдавших экзамен, если каждый из них может пересдать один раз экзамен в случае, если он его в первый раз не сдал. Найти математическое ожидание случайной величины X.

3.8. Дискретная случайная величина X задана законом распределения:

x_i	Значения вероятностей по вариантам											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0,1	0,2	0,05	0,15	0,1	0,2	0,25	0,1	0,4	0,05	0,15	0,12
3	0,2	0,25	0,15	0,2	0,3	0,4	0,3	0,15	0,3	0,1	0,25	0,23
5	0,4	0,3	0,2	0,25	0,3	0,3	0,2	0,25	0,2	0,15	0,25	0,46
7	0,2	0,15	0,4	0,25	0,2	0,05	0,15	0,35	0,08	0,25	0,2	0,14
9	0,1	0,1	0,2	0,15	0,1	0,05	0,1	0,15	0,02	0,45	0,15	0,05

По одному из 12 вариантов построить многоугольник распределения. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение случайной величины X.

3.9. Математическое ожидание случайной величины X равно 8. Найти математическое ожидание случайных величин: а) $X - 4$; б) $X + 6$; в) $3X - 4$; г) $4X + 3$.

3.10. Дисперсия случайной величины X равна 8. Найти дисперсию следующих величин: а) $X - 2$; б) $X + 6$; в) $3X - 2$; г) $2X + 7$.

3.11. Независимые случайные величины X и Y имеют следующие распределения:

x_i	2	4	6
p_i	0,3	0,5	0,2

y_i	3	4
p_i	0,4	0,6

Составить закон распределения случайных величин: а) $Z = X + Y$; б) $V = XY$. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение случайных величин Z и V.

3.12. Пусть X — случайная величина — сумма очков при двух бросаниях игральной кости. Найти ее математическое ожидание и дисперсию.

3.13. Вероятность выигрыша по облигации равна 0,05. Пусть случайная величина X — число выигрышных облигаций из 5. Составить закон распределения X, найти математическое ожидание и дисперсию.

3.14. В магазине имеется 10 телевизоров, из которых 4 дефектные. Пусть случайная величина X — число исправных телевизоров среди 3 выбранных. Найти закон распределения X, математическое ожидание и дисперсию.

3.15. Пусть $F(x)$ — функция распределения случайной величины X. Найти функцию распределения случайной величины $Y = aX + b$, $a > 0$.

3.16. В магазин поступила обувь с двух фабрик в соотношении 2:3. Куплено 4 пары обуви. Найти закон распределения числа купленных пар обуви, изготовленных первой фабрикой. Найти математическое ожидание, дисперсию и среднее квадратичное отклонение этой величины.

3.17. Для рекламы фирма вкладывает в каждую 10-ю единицу продукции приз в 1000 руб. Пусть случайная величина X — размер выигрыша при 5 сделанных покупках. Изобразить график функции распределения X, найти математическое ожидание.

3.18. В экзаменационном билете 3 задачи. Вероятность правильного решения студентом первой задачи равна 0,8, второй — 0,6 и третьей — 0,4. Найти математическое ожидание и дисперсию числа правильно решенных задач.

3.19. Плотность случайной величины X $p(x) = \frac{3}{4}(1 - x^2)$ при $x \in [-1, 1]$ и $p(x) = 0$ при $x \notin [-1, 1]$. Найти математическое ожидание и дисперсию.

3.20. Функция распределения случайной величины X:

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ x^3, & 0 < x \le 1, \\ 1, & x > 1. \end{cases}$$

Найти математическое ожидание и дисперсию.

3.21. Пусть X_1, X_2 — независимые, одинаково распределенные случайные величины и плотность X_i, $i = \overline{1,2}$, равна: $p(x) = 1$ при $x \in [0, 1]$, $p(x) = 0$ при $x \notin [0, 1]$. Пусть $X = \max\{X_1, X_2\}$. Найти плотность случайной величины X.

Указание. Функция распределения случайной величины X имеет вид

$$F_X(x) = P(X < x) = P\left(\max\{X_1, X_2\} < x\right) =$$

$$= P\left((X_1 < x) \cdot (X_2 < x)\right) = P(X_1 < x) \cdot P(X_2 < x) =$$

$$= \left\{x^2 \text{ при } x \in [0, 1], 0 \text{ при } x \notin [0, 1]\right\}.$$

Ответы

3.1.

x_i	0	1	2	3	4
p_i	0,055	0,235	0,374	0,265	0,071

$MX = 2{,}062$; $DX = 0{,}148$; $\sigma X = 0{,}385$.

3.2.

x_i	1	2	3	4	5
p_i	0,7	0,21	0,063	0,0189	0,0081

3.3.

x_i	1	2	3	4	5	6	7	...
p_i	0,7	0,21	0,063	0,0189	0,00567	0,0017	0,0005	...

3.4. $x_i = 0, 1, 2, 3$; $p_i = \dfrac{C_n^i C_{n+1}^{3-i}}{C_{2n+1}^3}$

3.5. $MX = 3{,}121$; $MX = 4{,}091$.
3.6. $MX = 2{,}4$.
3.7. $MX = 1{,}87$.
3.9. а) $MX = 4$; б) $MX = 14$; в) $MX = 20$; г) $MX = 35$.
3.10. а) $DX = 8$; б) $DX = 8$; в) $DX = 72$; г) $DX = 32$.
3.11. $MZ = 7{,}4$; $DZ = 2{,}2$; $\sigma Z = 1{,}48$; $MV = 13{,}68$; $DV = 29{,}3376$; $\sigma V = 5{,}4164$.
3.12. $MX = 7$; $DX = \dfrac{35}{6}$.
3.13. $MX = 0{,}25$; $DX = 0{,}2375$.
3.14. $MX = 1{,}8$; $DX = 0{,}56$.
3.15. $G(y) = F\left(\dfrac{y-b}{a}\right)$.

3.16. $MX = 1{,}6$; $DX = 0{,}96$; $\sigma X = 0{,}9799$.
3.17. $MX = 500$.
3.18. $MX = 1{,}8$; $DX = 0{,}64$.
3.19. $MX = 0$; $DX = 2$.
3.20. $MX = \frac{3}{4}$; $DX = \frac{3}{80}$.
3.21. $p(x) = \begin{cases} 0, & x \notin [0, 1] \\ 2x, & x \in [0, 1] \end{cases}$.

4

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

В этой главе рассмотрим основные законы распределения вероятностей случайных величин, наиболее часто возникающие на практике. Особое место среди них занимает нормальный закон распределения.

4.1. Законы распределения дискретных случайных величин

Мы знаем, что закон распределения дискретной случайной величины можно задавать либо таблицей распределения вероятностей, либо функцией распределения $F(x)$ и ее графиком.

4.1.1. Постоянная случайная величина (вырожденное распределение)

Постоянная случайная величина принимает только одно значение с вероятностью единица. Она задается таблицей

x	c
p	1

4.1.2. Альтернативная случайная величина (распределение Бернулли)

Альтернативная случайная величина принимает только два (любые) значения с вероятностями p и $(1 - p)$. Она задается таблицей

x_i	1	0
p_i	p	$1 - p$

Примером может служить величина, получаемая при однократном бросании монеты. Пусть выпадению решки соответствует число 1, а выпадению орла — число 0. Тогда таблица распределения имеет вид

x_i	1	0
p_i	1/2	1/2

4.1.3. Биномиальное распределение

Пусть имеется n испытаний Бернулли с вероятностью успеха в одном испытании, равной p, и вероятностью неуспеха q, $p + q = 1$. Дискретная случайная величина X, значения которой равны числу успехов, имеет распределение

$$p_k = P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, ..., n. \tag{4.1}$$

Это распределение называется *биномиальным* с параметрами p и n. Заметим, что сумма вероятностей по формуле Бернулли

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p + q)^n = 1.$$

Математическое ожидание и дисперсия биномиальной случайной величины:

$$MX = np, \quad DX = npq. \tag{4.2}$$

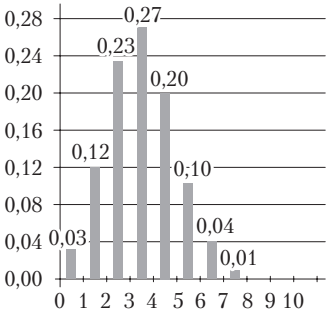


Рис. 4.1. Биномиальное распределение ($n = 10$, $p = 0{,}3$)

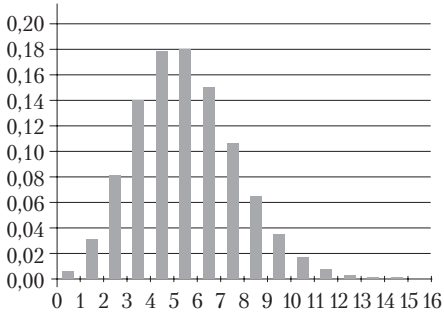


Рис. 4.2. Биномиальное распределение ($n = 100$, $p = 0{,}05$)

Графики биномиального распределения для $p = 0{,}3$, $n = 10$ и $p = 0{,}05$, $n = 100$ приведены на рис. 4.1 и 4.2.

4.1.4. Геометрическое распределение

Случайная величина X, имеющая *геометрическое распределение*, представляет собой число испытаний Бернулли до первого успеха.

Эта случайная величина принимает значения $k = 1, 2, 3, \ldots$ (счетное множество значений) с вероятностями

$$p_k = P(X = k) = pq^{k-1}, \quad 0 < p < 1, \quad q = 1 - p. \tag{4.3}$$

Определение является корректным, так как сумма вероятностей

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} pq^{k-1} = p\frac{1}{1-q} = 1.$$

Математическое ожидание и дисперсия X:

$$MX = 1/p, \quad DX = q/p^2. \tag{4.4}$$

Пример 4.1. В большой партии изделий вероятность брака равна p. Контроль качества проводится до первого появления бракованного изделия. В результате серии проверок обнаружилось, что бракованное изделие впервые появлялось в среднем при десятом испытании. Оценить вероятность p.

Пусть X — число испытаний до первого появления бракованного изделия. Эта случайная величина имеет геометрическое распределение. По условию ее среднее значение $MX = 10$. Так как $MX = 1/p$, то $p = 1/10$.

4.1.5. Гипергеометрическое распределение

Дискретная случайная величина X имеет *гипергеометрическое распределение*, если она принимает значения m с вероятностями

$$p_m = P(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \tag{4.5}$$

где $m = 0, 1, \ldots, k$, $k = \min\{n, M\}$, $M \le N$, $n \le N$.

Вероятность p_m является вероятностью появления m объектов, обладающих заданным свойством, в множестве n объектов, случайно извлеченных (без возврата) из совокупности N объектов, среди которых M объектов обладают заданным свойством.

Пример 4.2. Пусть в урне N шаров, среди которых M черных и $N - M$ белых. Наудачу выбираем (без возврата) n шаров.

Пусть X — случайная величина, значения которой равны числу m черных шаров среди n выбранных, $m = 0, 1, 2, ..., n$. Вероятности этих значений находятся по формуле (4.5).

Математическое ожидание и дисперсия случайной величины, имеющей гипергеометрическое распределение с параметрами n, M, N:

$$MX = n\frac{M}{N}, \quad DX = n\frac{M}{N-1}\left(1 - \frac{M}{N}\right)\left(1 - \frac{n}{N}\right). \tag{4.6}$$

Пример 4.3. Среди продукции цеха 10 электронных плат из партии 100 штук не удовлетворяют стандарту. При приемке продукции проверяют 10 плат. Какое среднее количество нестандартных плат обнаружат?

Количество нестандартных плат имеет гипергеометрическое распределение. Так как $n = 10$, $M = 10$, $N = 100$, то согласно (4.6)

$$MX = n\frac{M}{N} = 10 \cdot \frac{10}{100} = 1.$$

4.1.6. Закон Пуассона

Говорят, что случайная величина X *распределена по закону Пуассона* с параметром λ, если она принимает целые значения $k = 0, 1, 2, 3, ...$ с вероятностями

$$p_k = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \tag{4.7}$$

где $\lambda > 0$ — параметр распределения. При этом

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

Значения вероятностей p_k приводятся в таблицах распределения Пуассона. Распределение Пуассона иногда называют распре-

делением редких событий. Примерами величин, распределенных по закону Пуассона, могут служить: число несчастных случаев, число дефектов в производстве и т.п.

Математическое ожидание и дисперсия пуассоновской случайной величины равны параметру распределения:

$$MX = \lambda, \quad DX = \lambda. \tag{4.8}$$

Распределение Пуассона используется для приближенных вычислений в схеме Бернулли.

Если в испытаниях Бернулли n принимает большие значения, а вероятность успеха p относительно мала, так что число $\lambda = np$ и не мало, и не велико, то распределение Бернулли с достаточно высокой точностью можно заменить распределением Пуассона.

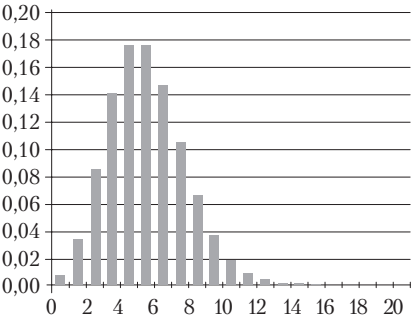


Рис. 4.3. Распределение Пуассона с параметром $\lambda = 5$

Сравните график распределения Пуассона с параметром $\lambda = 5$ (рис. 4.3) с графиком биномиального распределения при $p = 0{,}05$, $n = 100$ (рис. 4.2).

Пример 4.4. В ходе проверки аудитор случайным образом отбирает 10 счетов. Найти вероятность того, что он обнаружит один счет с ошибкой, если в среднем 3 % счетов содержат ошибки.

Используя формулу (2.3), в которой $n = 10$, $p = 0{,}03$, $\lambda = np = 0{,}3$, имеем

$$P_{10}(1) \approx \frac{e^{-\lambda}\lambda^1}{1!}.$$

По таблицам значений распределения Пуассона находим $P_{10}(1) \approx \approx 0{,}2223$.

Пример 4.5. При бросании игральной кости пространство элементарных событий Ω состоит из 6 элементов. Пусть X — функция на этом пространстве, значения которой равны числу выпавших очков. Эта функция является случайной величиной с таблицей распределения

x_i	1	2	3	4	5	6
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Такую случайную величину можно назвать дискретной равномерной с параметром $n = 6$, или классической случайной величиной.

4.2. Законы распределения непрерывных случайных величин

В отличие от дискретных случайных величин, задаваемых таблицей (или графиком), распределения непрерывных случайных величин задаются или функцией распределения $F(x)$, или функцией плотности вероятностей $p(x)$.

4.2.1. Равномерное распределение $R[a, b]$

Непрерывная случайная величина X распределена *равномерно* на отрезке $[a, b]$, если ее плотность вероятности $p(x)$ постоянна на этом отрезке и равна нулю вне его, т.е.

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & x < a,\ x > b. \end{cases} \tag{4.9}$$

Функция распределения случайной величины, распределенной по равномерному закону, имеет вид

$$F(x) = \begin{cases} 0, & x \le a, \\ \dfrac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases} \tag{4.10}$$

График $F(x)$ приведен на рис. 4.4.

Математическое ожидание и дисперсия равномерной случайной величины:

$$MX = \frac{a+b}{2}, \quad DX = \frac{(b-a)^2}{12}. \quad (4.11)$$

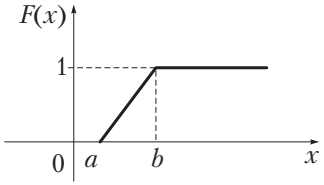


Рис. 4.4. График функции распределения случайной величины

Пример 4.6. Интервал движения автобуса равен 15 мин. Какова вероятность того, что пассажир на остановке будет ждать автобус не более 5 мин?

Пусть случайная величина X — это время ожидания автобуса. Она имеет равномерное распределение на отрезке [0, 15], если время прихода пассажира на остановку равновозможно в любой момент. Ее функция распределения имеет вид

$$F(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{x}{15}, & 0 < x \le 15, \\ 1, & x > 15. \end{cases}$$

Тогда по формуле (3.15)

$$P(0 \le X \le 5) = F(5) - F(0) = \frac{5}{15} - 0 = \frac{1}{3}.$$

4.2.2. Показательное (экспоненциальное) распределение

С законом Пуассона редких событий связана непрерывная случайная величина, распределение которой называется ***показательным*** (экспоненциальным). *Интенсивностью потока событий* называется число (параметр) λ — количество событий, происходящих в среднем за единицу времени. Если величина t задает время между последовательными наступлениями редких событий, имеющих интенсивность λ, то величина t имеет экспоненциальное распределение с параметром λ. Это распределение часто используется для описания интервалов времени между последовательными случайными редкими событиями.

Наоборот, если интервалы времени между последовательными редкими событиями независимы и имеют экспоненциальное распределение, то количество событий в единичном интервале времени имеет распределение Пуассона.

Непрерывная случайная величина X имеет *показательное распределение* с параметром λ, если ее плотность вероятностей имеет вид

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{4.12}$$

где $\lambda > 0$.

Функция распределения случайной величины, распределенной по показательному закону,

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{4.13}$$

Показательная случайная величина X с параметром λ на промежутке $[a, b]$, где $a \geq 0$, принимает значения с вероятностью

$$P(a \leq x \leq b) = e^{-a\lambda} - e^{-b\lambda}.$$

Графики функции распределения и плотности показательного распределения приведены на рис. 4.5.

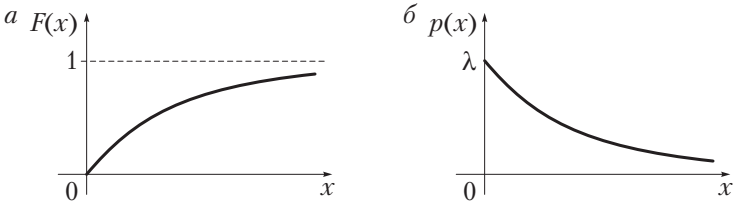


Рис. 4.5. Графики функции распределения (*а*) и плотности показательного распределения (*б*) случайной величины

Математическое ожидание и дисперсия показательно распределенной случайной величины X:

$$MX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}. \tag{4.14}$$

Пример 4.7. Установлено, что время T горения электрической лампочки является случайной величиной, распределенной по показательному закону. Считая, что среднее значение этой величины равно 6 месяцам, найти вероятность того, что лампочка будет гореть не менее года.

Так как $MT = \frac{1}{\lambda} = 6$, то $\lambda = \frac{1}{6}$, а функция распределения случайной величины T имеет вид

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{6}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Поэтому

$$P(T > 12) = P(12 < T < +\infty) = F(\infty) - F(12) =$$

$$= 1 - \left(1 - e^{-\frac{12}{6}}\right) = e^{-2} \approx 0{,}1353.$$

Биномиальное распределение, распределение Пуассона и показательное распределение тесно связаны между собой. Значения биномиальной случайной величины равны числу успехов в системе Бернулли с параметрами n и p. Среднее значение числа успехов $\lambda = np$. Распределение Пуассона является предельным случаем биномиального распределения, когда $n \to \infty$, $p \to 0$, так что их произведение $\lambda = np$ остается постоянным. Это распределение задает дискретную случайную величину с натуральными значениями, равными числу редких событий, появляющихся за единицу времени. Ее параметр λ, называемый интенсивностью потока событий, равен числу событий, появляющихся в среднем за единицу времени. Как отмечено выше, интервалы времени между последовательными появлениями событий в потоке Пуассона имеют показательное распределение.

Схожесть биномиального распределения и распределения Пуассона подтверждается рис. 4.1–4.3.

4.2.3. Нормальное распределение

Непрерывная случайная величина X имеет *нормальный закон распределения* (*закон Гаусса*) с параметрами a и σ, если ее плотность вероятности имеет вид

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (4.15)$$

где $\sigma > 0$.

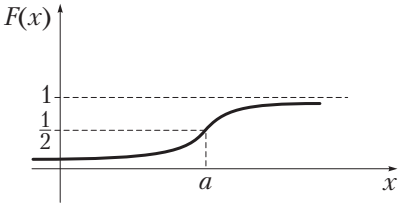


Рис. 4.6. График функции распределения нормальной случайной величины

Обозначим $N(a; \sigma)$ множество случайных величин, распределенных по нормальному закону с параметрами a и σ. Функция распределения нормальной случайной величины $X \in N(a; \sigma)$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt, \qquad (4.16)$$

а ее график представлен на рис. 4.6.

Математическое ожидание случайной величины, распределенной по нормальному закону, равно параметру a, а дисперсия равна σ^2, т.е.

$$MX = a, \quad DX = \sigma^2. \qquad (4.17)$$

На рис. 4.7 и 4.8 приведены графики плотностей вероятностей нормально распределенных случайных величин при различных значениях параметров a и σ. Параметр a характеризует положение, а σ — форму кривой нормального распределения.

При $a = 0$ и $\sigma = 1$ распределение называется *стандартным нормальным*. Множество случайных величин с таким распределением обозначим $N(0; 1)$.

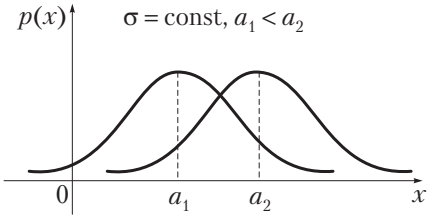


Рис. 4.7. Плотность нормального распределения при σ = const

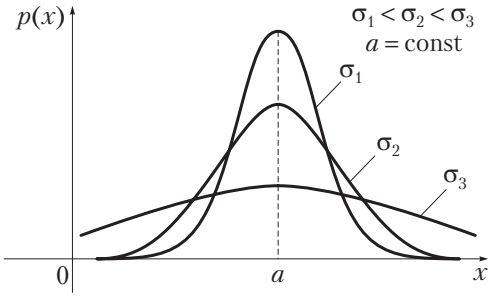


Рис. 4.8. Плотность нормального распределения при a = const

Для стандартного распределения плотность вероятностей

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

Введенная ранее [см. формулу (2.6)] функция Лапласа

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

задает вероятность попадания случайной нормальной стандартной величины X в интервал $(-x, x)$.

Функция $\Phi(x)$ связана с функцией $\varphi(x)$ соотношением

$$\Phi(x) = \int_{-x}^{x} \varphi(t)dt = 2\int_0^x \varphi(t)dt.$$

Геометрически $\Phi(x)$ представляет собой часть площади под кривой $\varphi(t)$ (рис. 4.9, заштрихованная область).

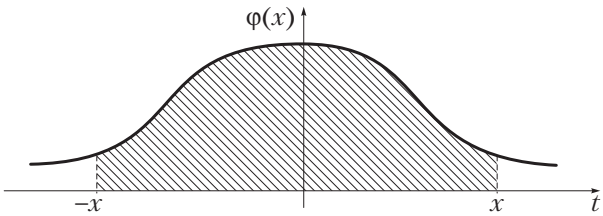


Рис. 4.9. Связь функции Лапласа с плотностью стандартного распределения

Пусть $X \in N(a; \sigma)$, тогда вероятность того, что X принимает значения из интервала (x_1, x_2), можно вычислить, используя функцию $\Phi(x)$:

$$P(x_1 < X < x_2) = \frac{1}{2}\Phi\left(\frac{x_2 - a}{\sigma}\right) - \frac{1}{2}\Phi\left(\frac{x_1 - a}{\sigma}\right). \qquad (4.18)$$

Замечание. Нормальное распределение возникает всегда при систематических отклонениях случайной величины от своего среднего, например при измерении физических величин, отклонениях заданных координат при стрельбе и т.д., т.е. в тех случаях, когда на величину влияет большое количество случайных факторов. При довольно общих предположениях нормальное распределение возникает при сложении множества случайных величин, чему дают пример предельные теоремы Муавра—Лапласа. А.М. Ляпуновым в 1901 г. была доказана общая теорема, сущность которой заключается в том, что при некоторых условиях нормировки распределение суммы n независимых величин с произвольными распределениями стремится к нормальному при $n \to \infty$.

Пример 4.8. Текущая цена ценной бумаги представляет собой нормально распределенную случайную величину X со средним значением 100 усл. ед. и дисперсией 9. Найти вероятность того, что цена актива будет находиться в пределах от 91 до 109 усл. ед.

Так как $a = 100$, $\sigma = \sqrt{9} = 3$, то

$$P(91 < X < 109) = \frac{1}{2}\Phi\left(\frac{109 - 100}{3}\right) - \frac{1}{2}\Phi\left(\frac{91 - 100}{3}\right) =$$

$$= \frac{1}{2}\Phi(3) - \frac{1}{2}\Phi(-3) = \Phi(3) = 0{,}9973.$$

Важными следствиями из формулы 4.18 являются следующие оценки отклонения значений нормальной случайной величины от ее среднего значения:

1) $P\left(|X - a| < \varepsilon\right) = \Phi\left(\frac{\varepsilon}{\sigma}\right)$;

2) правило трех сигм $P\left(|X - a| < 3\sigma\right) \approx 0{,}997$.

4.2.4. Логнормальное распределение

Распределение случайной величины X с положительными значениями называется *логнормальным*, если случайная величина $\ln X$ имеет нормальное распределение $N(a; \sigma)$. Логнормальное распре-

деление используется для распределения доходов, банковских вкладов, заработной платы, посевных площадей, долговечных изделий и др.

Найдем функцию распределения и функцию плотности вероятностей величины X. Для $x > 0$ неравенства $X < x$ и $\ln X < \ln x$ равносильны. Поэтому искомая функция распределения

$$F_X(x) = P(X < x) = P(\ln X < \ln x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt$$

в силу формулы (4.16) для нормального распределения. Функция плотности вероятностей равна производной от функции $F_X(x)$, которая имеет вид сложной функции. По правилу дифференцирования сложной функции имеем

$$\frac{d}{dx} f\left(\ln(x)\right) = \frac{df}{d(\ln x)} \cdot \frac{1}{x}.$$

Следовательно,

$$\frac{d}{dx} F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-a)^2}{2\sigma^2}} \cdot \frac{1}{x}.$$

Таким образом, функция плотности вероятностей логнормального распределения имеет вид

$$p_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-a)^2}{2\sigma^2}}.$$

Числовые характеристики логнормальной случайной величины X задаются следующими выражениями:

$$MX = e^{\left(a+\frac{\sigma^2}{2}\right)}, \quad DX = (e^{\sigma^2} - 1)e^{2a+\sigma^2}.$$

Задачи с решениями

Задача 4.1. 20 % изделий, выпускаемых предприятием, нуждаются в дополнительной регулировке. Наудачу отобрано 150 изделий. Найти среднее значение и дисперсию случайной величины X — числа изделий в выборке, нуждающихся в регулировке.

Решение. В данном случае мы имеем дело со схемой испытаний Бернулли, поэтому случайная величина X биномиально распределена. Используя формулы $MX = np$ и $DX = npq$, находим (при $n = 150$, $p = 0{,}2$, $q = 0{,}8$) $MX = 150 \cdot 0{,}2 = 30$, $DX = 150 \cdot 0{,}2 \cdot 0{,}8 = 24$.

Задача 4.2. Проверяется партия из 1000 изделий. Вероятность того, что изделие окажется бракованным, равна 0,002. Найти математическое ожидание и дисперсию числа бракованных изделий из этой партии. Найти вероятность того, что в партии есть хотя бы одно бракованное изделие.

Решение. Число опытов ($n = 1000$) достаточно велико, а вероятность ($p = 0{,}002$) «успеха» в каждом из них мала, поэтому можно считать, что случайная величина X — число бракованных изделий — распределена по закону Пуассона: ее возможные значения 0, 1, 2, ..., 1000, а соответствующие вероятности определяются по формуле

$$p_m = P\{X = m\} = \frac{\lambda^m e^{-\lambda}}{m!}.$$

По формуле $\lambda = np$ определяем параметр $\lambda = 2$ и математическое ожидание случайной величины X:

$$MX = \lambda = 1000 \cdot 0{,}002 = 2.$$

Дисперсия числа бракованных изделий

$$DX = npq = 1000 \cdot 0{,}002 \cdot 0{,}998 = 1{,}996,$$

т.е. $MX \approx DX$. Полагая, что $MX = DX = 2$, находим приближенно искомую вероятность события $A = \{$в партии содержится хотя бы одно бракованное изделие$\}$: $P(A) = 1 - P(\bar{A}) = 1 - P_{1000}(0)$. По таблице Пуассона $P_{1000}(0) \approx 0$, $P(A) \approx 1$.

Задача 4.3. Производится стрельба по цели до первого попадания. Вероятность попадания при каждом выстреле равна 0,2. Найти математическое ожидание и дисперсию случайной величины X — числа произведенных выстрелов, считая, что: а) стрелять можно неограниченное число раз; б) в наличии всего 5 патронов.

Решение. а) Случайная величина X распределена геометрически, ее ряд распределения имеет вид

x_i	1	2	3	...
p_i	p	qp	q^2p	...

Числовые характеристики этого распределения:

$$MX = \frac{1}{p}, \quad DX = \frac{q}{p^2}.$$

Следовательно, зная, что $p = 0{,}2$ и $q = 0{,}8$, имеем:

$$MX = \frac{1}{0{,}2} = 5, \quad DX = \frac{0{,}8}{0{,}04} = 20.$$

б) Ряд распределения случайной величины X имеет вид

x_i	1	2	3	4	5
p_i	p	qp	q^2p	q^3p	$q^4p + q$

Поэтому

$$MX = 1 \cdot p + 2 \cdot qp + 3 \cdot q^2 p + 4 \cdot q^3 p + 5 \cdot q^4.$$

При $p = 0{,}2$ и $q = 0{,}8$ имеем:

$$MX = 0{,}2 + 0{,}32 + 0{,}384 + 0{,}4096 + 2{,}048 = 3{,}3616,$$

$$DX = MX^2 - (MX)^2 =$$

$$= 1 \cdot p + 4 \cdot qp + 9 \cdot q^2 p + 16q^3 p + 25(q^4 p + q^5) - (MX)^2.$$

При $p = 0{,}2$ и $q = 0{,}8$ получим

$$DX = 0{,}2 + 0{,}64 + 1{,}152 + 1{,}6384 + 10{,}24 - (3{,}3616)^2 \approx$$

$$\approx 13{,}8704 - 11{,}3 = 2{,}57.$$

Задача 4.4. Случайная величина X распределена равномерно на отрезке $[a, b]$, т.е. $X \in R[a, b]$. Найти вероятность попадания случайной величины X на отрезок $[\alpha, \beta]$, целиком содержащийся внутри отрезка $[a, b]$.

Решение. Воспользуемся формулой

$$P\{\alpha \le X \le \beta\} = \int_{\alpha}^{\beta} p(x)dx,$$

где плотность вероятности случайной величины X имеет вид

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Следовательно,

$$P\{\alpha \le X \le \beta\} = P\{X \in [\alpha, \beta]\} = \int_{\alpha}^{\beta} \frac{1}{b-a} dx = \frac{1}{b-a} \cdot x\Big|_{\alpha}^{\beta} = \frac{\beta - \alpha}{b-a},$$

т.е. окончательно —

$$P\{X \in [\alpha, \beta]\} = \frac{\beta - \alpha}{b-a}.$$

Задача 4.5. Плотность вероятности непрерывной случайной величины X имеет вид

$$p(x) = \begin{cases} 0{,}25A, & x \in [0, 4], \\ 0, & x \notin [0, 4]. \end{cases}$$

Найти A, $F(x)$, MX, DX, σX, $P\{X \in [0; 1{,}1]\}$.

Решение. Коэффициент A найдем, используя следующее свойство плотности вероятности:

$$\int_{-\infty}^{\infty} p(x)dx = 1.$$

Имеем:

$$\int_{-\infty}^{\infty} p(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{4} 0{,}25Adx + \int_{4}^{\infty} 0dx = 1,$$

т.е. $0{,}25Ax\,|_0^4 = 1$.

Отсюда следует, что $A = 1$. Итак,

$$p(x) = \begin{cases} \frac{1}{4}, \ x \in [0,\ 4], \\ 0, \ x \notin [0,\ 4], \end{cases}$$

и, значит, случайная величина X равномерно распределена на отрезке $[0,\ 4]$. Функция распределения для случайной величины $X \in R[0,\ 4]$ имеет вид

$$F(x) = \begin{cases} 0, \ x \le 0, \\ \frac{x}{4}, \ 0 < x \le 4, \\ 4 < x. \end{cases}$$

Числовые характеристики этого распределения:

$$MX = \left[\frac{a+b}{2}\right] = \frac{0+4}{2} = 2,$$

$$DX = \left[\frac{(b-a)^2}{12}\right] = \frac{4^2}{12} = \frac{4}{3},$$

$$\sigma X = \left[\sqrt{DX}\right] = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{3}.$$

Вероятность попадания случайной величины X в промежуток $[0;\ 1{,}1]$ находим, используя формулу

$$P\{X \in [\alpha,\ \beta]\} = \frac{\beta - \alpha}{b - a}.$$

Отсюда

$$P\{X \in [0;\ 1{,}1]\} = \frac{1{,}1 - 0}{4 - 0} = \frac{1{,}1}{4} = 0{,}275.$$

Задача 4.6. Непрерывная случайная величина X имеет показательное распределение. Найти вероятность попадания случайной величины X в отрезок $[a,\ b]$, где $a \ge 0$, $b > a$.

Решение. Воспользовавшись формулой

$$P\{a \le X \le b\} = \int_a^b p(x)dx,$$

находим

$$P\{a \le X \le b\} = \int_a^b \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_a^b = -e^{-\lambda b} + e^{-\lambda a}.$$

Этот же результат можно получить, используя формулу

$$P\{a \le X \le b\} = F(b) - F(a).$$

Так как $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$, то

$$P\{a \le X \le b\} = (1 - e^{-\lambda b}) - (1 - e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b}.$$

Задача 4.7. Случайные ошибки измерения детали подчинены нормальному закону с параметром $\sigma = 20$ мм. Найти вероятность того, что измерение детали произведено с ошибкой, не превосходящей по модулю 25 мм.

Решение. Воспользуемся формулой

$$P\left\{|X - a| < \varepsilon\right\} = \Phi\left(\frac{\varepsilon}{\sigma}\right).$$

В нашем случае $\sigma = 20$, $\varepsilon = 25$, поэтому

$$P\left\{|X - a| < 25\right\} = \Phi\left(\frac{25}{20}\right) = \Phi(1{,}25) = 2 \cdot 0{,}3944 = 0{,}7888.$$

Задачи для самостоятельного решения

4.1. Найти среднее число лотерейных билетов, на которые выпадут выигрыши, если приобретено 20 билетов, а вероятность выигрыша одного билета равна 0,1. Найти дисперсию числа успехов в данном опыте.

4.2. Проводится 3 независимых испытания, в каждом из которых вероятность наступления некоторого события постоянна и равна p. Пусть X — число появлений события A в этом опыте. Найти DX, если известно, что $MX = 2{,}1$.

4.3. Вероятность поражения цели при одном выстреле равна 0,4. Сколько надо произвести выстрелов, чтобы можно было ожидать в среднем 80 попаданий в цель?

4.4. Дискретная случайная величина X распределена по закону Пуассона с параметром $\lambda = 0{,}324$. Найти математическое ожидание и среднее квадратичное отклонение этой случайной величины.

4.5. В магазин отправлено 1000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равна 0,002. Найти среднее число разбитых бутылок и вероятность того, что магазин получит более двух разбитых бутылок.

4.6. Игрок покупает лотерейные билеты до первого выигрыша. Вероятность выигрыша по одному билету равна 0,1. Найти MX, где случайная величина X — число купленных билетов, если игрок может купить: а) только 4 билета; б) неограниченное число билетов.

4.7. Некто ожидает телефонный звонок между 19.00 и 20.00. Время ожидания звонка есть непрерывная случайная величина X, имеющая равномерное распределение на отрезке [19, 20]. Найти вероятность того, что звонок поступит в промежутке от 19.22 до 19.46.

4.8. Средняя продолжительность телефонного разговора равна 3 мин. Найти вероятность того, что произвольный телефонный разговор будет продолжаться не более 9 мин, считая, что время разговора является случайной величиной X, распределенной по показательному закону.

4.9. Пусть $X = N(5; 0{,}5)$. Найти вероятность того, что при 3 независимых испытаниях случайной величины X хотя бы в одном из них X примет значение в интервале $(2, 4)$.

4.10. Известно, что $X \in N(50; \sigma)$, $P\{X \in (40, 60)\} = 0{,}7887$. Найти DX.

4.11. Рост взрослых мужчин является случайной величиной X, распределенной по нормальному закону: $X \in N(175; 10)$. Найти: плотность вероятности; функцию распределения этой случайной величины; вероятность того, что ни один из 3 наудачу выбранных мужчин не будет иметь рост менее 180 см.

4.12. Успеваемость студентов 1-го курса составляет 80 %. Найти математическое ожидание и дисперсию числа успевающих студентов среди 50 наудачу отобранных первокурсников.

4.13. Студент знает 30 из 40 вопросов. Экзаменатор задает вопросы студенту до тех пор, пока обнаружит незнание вопроса. Найти вероятность того, что число заданных вопросов больше двух.

4.14. Автобусы данного маршрута идут с интервалом 30 мин. Пассажир подходит к автобусной остановке в произвольный момент времени. Время ожидания автобуса есть непрерывная случайная величина X, имеющая равномерное распределение. Найти плотность вероятности; функцию распределения; математическое ожидание и дисперсию этой случайной величины; вероятность появления пассажира не ранее, чем через 17 мин после ухода предыдущего автобуса, но не позднее, чем за 1 мин до отхода следующего автобуса.

4.15. Найти математическое ожидание и дисперсию произведения двух независимых непрерывных случайных величин X и Y с равномерными законами распределениями: $X \in R[0, 1]$, $Y \in R[1, 3]$.

4.16. 90 % лампочек перегорает после 800 ч работы. Найти вероятность того, что лампочка перегорит в промежутке от 100 до 200 ч работы (использовать случайную величину T — время безотказной работы лампочки).

4.17. X — нормально распределенная случайная величина, причем $MX = 6{,}2$ и $\sigma X = 4{,}4$. Найти $P\{|X - MX| < 7{,}5\}$.

4.18. Ошибка измерения подчинена нормальному закону с параметрами $a = 50$ дм и $\sigma = 10$ дм. Найти вероятность того, что измеренное значение будет отклоняться от истинного не более, чем на 20 дм.

4.19. Срок безотказной работы телевизора представляет собой случайную величину $X \in N(12; 3)$. Найти вероятность того, что телевизор проработает: а) не менее 15 лет; б) от 6 до 9 лет; в) от 9 до 15 лет.

Ответы

4.1. $np = 2$; $DX = 1{,}8$.
4.2. $DX = 0{,}63$.
4.3. $n = 200$.
4.4. $MX = 0{,}324$; $\sigma X = 0{,}569$.
4.5. $n = 2$; $p \approx 0{,}323$.
4.6. а) $MX = 3{,}439$; б) $MX = 10$.
4.7. $p = 0{,}4$.
4.8. $p = 0{,}95$.
4.9. $p = 0{,}067$.
4.10. $DX = 64$.
4.11. $F(x) = 0{,}5 + 0{,}5\Phi\left(\dfrac{x-175}{10}\right)$;
$p(\mathrm{X} \geq 180) = 0{,}159$;
$p(x) = 0{,}1\varphi\left(\dfrac{x-175}{10}\right)$.
4.12. $MX = 40$; $DX = 8$.
4.13. $p = \dfrac{29}{52}$.
4.14. $f(x) = \begin{cases} \dfrac{1}{30}, & x \in [0,\ 30], \\ 0, & x \notin [0,\ 30]; \end{cases}$

$$F(x) = \begin{cases} 1 - 0, & x \leq 0, \\ \dfrac{x}{30}, & 0 < x \leq 30, \\ 1, & 30 < x; \end{cases}$$

$MX = 15$; $DX = 75$, $p = 0{,}4$.
4.15. $MXY = 1$, $DXY = \dfrac{4}{9}$.
4.16. $p = 0{,}1084$ ($\lambda = 0{,}0029$).
4.17. $p = 0{,}805$.
4.18. $p = 0{,}954$.
4.19. а) $p = 0{,}159$; б) $p = 0{,}136$;
в) $p = 0{,}683$.

5

МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Многомерной случайной величиной, или ***n-мерным случайным вектором***, называется упорядоченный набор $X = \{X_1, X_2, ..., X_n\}$ случайных величин, рассматриваемых как функции на одном и том же пространстве событий, вместе с совместным законом распределения вероятностей.

Многомерная случайная величина каждому элементарному событию ставит в соответствие n действительных чисел $(x_1, x_2, ..., x_n)$, которые приняли случайные величины $X_1, X_2, ..., X_n$ в результате испытания.

Случайные величины $X_1, X_2, ..., X_n$ называются *составляющими* случайного n-вектора X. Эти составляющие однозначно определяются вектором X. *Совместным законом распределения* составляющих $\{X_1, X_2, ..., X_n\}$ в случае дискретных случайных величин называются вероятности $P\big((X_1 = x_1) \cdot (X_2 = x_2) \cdot ... \cdot (X_n = x_n)\big)$ для всех возможных значений $(x_1, x_2, ..., x_n)$ этих величин. В случае непрерывных случайных величин $X = \{X_1, X_2, ..., X_n\}$ *совместным распределением* называется распределение вероятностей $P\big((X_1 < x_1) \cdot (X_2 < x_2) \cdot ... \cdot (X_n < x_n)\big)$ для всех возможных наборов чисел $(x_1, x_2, ..., x_n)$. Обычно при использовании случайных величин мы ограничиваемся значениями этих величин и соответствующими им вероятностями. Также и для многомерной случайной величины мы обычно рассматриваем ее значения и соответствующие совместные распределения вероятностей.

Отметим при этом, что совместное распределение однозначно определяет распределения вероятностей всех составляющих $\{X_1, X_2, ..., X_n\}$. Обратное утверждение будет справедливым только в случае независимых в совокупности случайных величин $\{X_1, X_2, ..., X_n\}$.

Ниже мы уточним это высказывание и приведем примеры совместных законов распределений.

Для простоты и наглядности изложения в дальнейшем будем рассматривать только двумерные случайные величины.

5.1. Дискретные двумерные случайные величины

Если случайные величины X и Y дискретны, то двумерная случайная величина $\{X, Y\}$ называется *дискретной*. Дискретная двумерная случайная величина может принимать конечное или счетное множество значений, при этом ее значениями являются пары действительных чисел (x, y).

Совместным законом распределения дискретной двумерной случайной величины называется множество ее значений и вероятностей этих значений. Эти вероятности обозначаются

$$p_{ij} = P\big((X = x_i) \cdot (Y = y_j)\big),$$

где $i = 1, 2, 3, \ldots, j = 1, 2, 3, \ldots$.

Закон распределения двумерной дискретной случайной величины с конечным числом значений можно представить в виде двумерной матрицы распределения, в каждой клетке которой содержатся вероятности p_{ij} произведения событий $(X = x_i) \cdot (Y = y_j)$ (табл. 5.1), которые и являются заданием совместного распределения вероятностей.

Пусть $A_{ij} = (X = x_i) \cdot (Y = y_j)$ — события, состоящие в том, что случайная величина X примет значение x_i, а случайная величина Y — значение y_j. Тогда очевидно, что события A_{ij} и A_{kl} несовместны, если хотя бы одно число пары (i, j) не совпадает с соответствующим числом пары (k, l). В совокупности эти события составляют все пространство Ω, т.е. они образуют полную группу событий. Поэтому

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1.$$

Таблица 5.1

x_i \ y_j	y_1	y_2	...	y_j	...	y_m	Σ
x_1	p_{11}	p_{12}	...	p_{1j}	...		$p_1 = \sum_{j=1}^{m} p_{1j}$
x_2	p_{21}	p_{22}	...	p_{2j}	...		$p_2 = \sum_{j=1}^{m} p_{2j}$

Окончание табл. 5.1

x_i \ y_j	y_1	y_2	...	y_j	...	y_m	Σ
...	...	...	...	...	...	...	...
	p_{i1}	p_{i2}	...	p_{ij}	...		$p_i = \sum_{j=1}^{m} p_{ij}$
...	...	...	...	...	...	...	...
x_n	p_{n1}	p_{n2}	...	p_{nj}	...		$p_n = \sum_{j=1}^{m} p_{nj}$
Σ	q_1	q_2	...	q_j	...	q_m	1

Таблица распределения двумерной случайной величины $\{X, Y\}$ (табл. 5.1) однозначно определяет распределения одномерных случайных величин X и Y, ее составляющих. Вероятности значений величин X и Y определяются формулами

$$p_i = P(X = x_i) = \sum_{j=1}^{m} p_{ij}, \tag{5.1}$$

$$q_j = P(Y = y_j) = \sum_{i=1}^{n} p_{ij}. \tag{5.2}$$

Таким образом, суммируя элементы матрицы распределения вероятностей по строкам, получим распределение случайной величины X, по столбцам — величины́ Y. Вероятности p_i и q_j записывают в последний столбец и последнюю строку таблицы распределения.

Если считать, что событие $(Y = y_j)$ произошло, то распределение случайной величины X называется ***условным распределением*** при условии $(Y = y_j)$. Из определения условной вероятности

$$p_j(x_i) = P((X = x_i) \mid (Y = y_j)) =$$

$$= \frac{P\left[(X = x_i) \cdot (Y = y_i)\right]}{P(Y = y_j)} = \frac{p_{ij}}{q_j}. \tag{5.3}$$

Аналогично задается условное распределение случайной величины X при условии $(X = x_i)$:

$$p_i(y_j) = \frac{p_{ij}}{p_i}. \tag{5.4}$$

Отметим, что совместное распределение вероятностей двумерной случайной величины $\{X, Y\}$ (см. табл. 5.1) однозначно определяет распределения ее составляющих. Однако задание распределений величин X и Y не определяет их совместное распределение. Для задания совместного распределения p_{ij}, как видно из формул (5.3), (5.4), нужно знать распределения составляющих и их условные распределения. В случае независимости составляющих совместные вероятности p_{ij} равны произведениям $p_{ij} = p_i \cdot q_j$.

5.2. Функция распределения двумерной случайной величины

По аналогии с одномерной случайной величиной функция распределения двумерной случайной величины задает вероятности попадания значений величины в квадрант на плоскости.

Пусть (x, y) — пара действительных чисел. Тогда множество

$$(X < x) \cdot (Y < y) = \left\{\omega \in \Omega \mid (X(\omega) < x) \cdot (Y(\omega) < y)\right\}$$

является событием, а его вероятность задает значение совместной функции распределения:

$$F(x, y) = P\left[(X < x) \cdot (Y < y)\right].$$

Функция распределения $F(x, y)$ является числовой функцией двух переменных.

Для дискретной двумерной случайной величины *функция распределения* определяется по формуле

$$F(x,y) = \sum_{i,\, x_i < x} \; \sum_{j,\, y_j < y} p_{ij}.$$

Свойства функции распределения:

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

2. Функция распределения есть неубывающая функция по каждому из своих аргументов, т.е.

$$x_1 < x_2 \Rightarrow F(x_1, y) \le F(x_2, y),$$

$$y_1 < y_2 \Rightarrow F(x, y_1) \le F(x, y_2).$$

3. $F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0$.

4. $F(x, +\infty) = F_1(x)$, $F(+\infty, y) = F_2(y)$, где $F_1(x)$ и $F_2(y)$ есть функции распределения случайных величин X и Y, т.е. $F_1(x) = P(X < x)$, $F_2(y) = P(Y < y)$.

5. $F(+\infty, +\infty) = 1$.

5.3. Непрерывные двумерные случайные величины

Непрерывной называется двумерная случайная величина $\{X, Y\}$, для которой функция распределения $F(x, y)$ непрерывна по обоим аргументам.

Если функцию распределения можно представить в виде двукратного интеграла от некоторой неотрицательной функции $p(x, y)$:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(t_1, t_2) dt_1 dt_2,$$

то говорят, что двумерная случайная величина $\{X, Y\}$ имеет *совместную плотность распределения* $p(x, y)$, а сама величина $\{X, Y\}$ называется *абсолютно непрерывной*. В этом случае функция $F(x, y)$ будет непрерывной. Такая функция распределения связана с функцией плотности формулой

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = p(x, y),$$

если $F(x, y)$ дважды дифференцируема.

Свойства совместной плотности:

1. $p(x, y) \ge 0$.

2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = 1$.

Если $S \subset R^2$ — область на плоскости, то вероятность попадания двумерной случайной величины в эту область выражается интегралом

$$P(\{X,\ Y\} \in S) = \iint\limits_S p(x,y)dxdy.$$

В частности, если S представляет собой прямоугольник с вершинами $A(a_1,\ b_1)$, $B(a_2,\ b_1)$, $C(a_2,\ b_2)$, $D(a_1,\ b_2)$, где $a_1 < a_2$, $b_1 < b_2$, то вероятность попадания случайной величины в S

$$P(\{X,\ Y\} \in S) = \int\limits_{a_1}^{a_2}\int\limits_{b_1}^{b_2} p(x,y)dxdy = \int\limits_{a_1}^{a_2} dx \int\limits_{b_1}^{b_2} p(x,y)dy.$$

Если $p(x,\ y)$ — совместная плотность двух случайных величин X и Y, то одномерные плотности $p(x)$ и $q(y)$ величин X и Y можно вычислить по формулам

$$p(x) = \int\limits_{-\infty}^{+\infty} p(x,y)dy, \quad q(y) = \int\limits_{-\infty}^{+\infty} p(x,y)dx. \tag{5.5}$$

5.4. Независимые случайные величины. Ковариация, коэффициент корреляции

Независимыми называются такие случайные величины X и Y, для которых совместная функция распределения $F(x,\ y)$ равна произведению функций распределений $F_1(x)$ и $F_2(y)$ случайных величин X и Y:

$$F(x,y) = F_1(x) \cdot F_2(y). \tag{5.6}$$

Если случайные величины X и Y независимы и имеют совместную плотность $p(x,\ y)$, то она равна произведению плотностей $p_1(x)$ и $p_2(y)$ величин X и Y:

$$p(x,y) = p_1(x) \cdot p_2(y). \tag{5.7}$$

Условие (5.7) также может служить определением независимости непрерывных случайных величин, имеющих плотности.

Одной из характеристик совместной вариации (зависимости) случайных величин X и Y служит их *ковариация*

$$\operatorname{cov}(X,\ Y) = M\left(\left(X - M(X)\right)\cdot\left(Y - M(Y)\right)\right) =$$

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

Если случайные величины независимы, то математическое ожидание их произведения равно произведению их математических ожиданий, поэтому их ковариация равна нулю. Чаще вместо ковариации используется коэффициент корреляции.

Коэффициентом корреляции двух случайных величин X и Y называется число $r(X,\ Y)$, определяемое формулой

$$r(X,Y) = \frac{M\left((X - MX)\cdot(Y - MY)\right)}{\sigma X\cdot\sigma Y} =$$

$$= \frac{M(X\cdot Y) - MX\cdot MY}{\sigma_X\cdot\sigma_Y}. \qquad (5.8)$$

Коэффициент корреляции $r(X,\ Y)$ обладает свойствами:

1. $-1 \le r(X,\ Y) \le 1$.
2. Если X и Y независимы, то $r(X,\ Y) = 0$.
3. Если $r(X,\ Y) \ne 0$, то X и Y являются зависимыми случайными величинами.
4. Если $Y = aX + b$, то $r(X,\ Y) = \operatorname{sgn} a$, где

$$\operatorname{sgn} a = \begin{cases} 1,\ a > 0, \\ 0,\ a = 0, \\ -1,\ a < 0. \end{cases}$$

Верно и обратное: если $|r(X,\ Y)| = 1$, то между величинами X и Y существует линейная зависимость.

Из свойств коэффициента корреляции следует, что $r(X,\ Y)$ является мерой линейной зависимости случайных величин X и Y.

Пример 5.1. Найти коэффициент корреляции между величинами X и Y, совместный закон распределения которых задан в табл. 5.2.

Таблица 5.2

x_i \ y_j	1	2	3	4	$\sum$
10	0,2	0,02	0,01	0	0,23
20	0,03	0,3	0,02	0	0,35
30	0,02	0,1	0,2	0,1	0,42
$\sum$	0,25	0,42	0,23	0,1	1

Последовательно находим

$$MX = \sum_i x_i p_i = 10 \cdot 0{,}23 + 20 \cdot 0{,}35 + 30 \cdot 0{,}42 = 21{,}9,$$

$$MY = \sum_j y_j q_j = 1 \cdot 0{,}25 + 2 \cdot 0{,}42 + 3 \cdot 0{,}23 + 4 \cdot 0{,}1 = 2{,}18,$$

$$MX^2 = \sum_i x_i^2 p_i = 100 \cdot 0{,}23 + 400 \cdot 0{,}35 + 900 \cdot 0{,}42 = 541,$$

$$M(X \cdot Y) = \sum_{i,j} x_i y_j p_{ij} = 10 \cdot 0{,}2 + 20 \cdot 0{,}02 + 30 \cdot 0{,}01 + 20 \cdot 0{,}03 +$$

$$+ 40 \cdot 0{,}3 + 60 \cdot 0{,}02 + 30 \cdot 0{,}02 + 60 \cdot 0{,}1 + 90 \cdot 0{,}2 + 120 \cdot 0{,}1 = 53{,}1.$$

Тогда

$$DX = MX^2 - (MX)^2 = 541 - 21{,}9^2 = 61{,}39,$$

$$\sigma_X = \sqrt{DX} = 7{,}8352, \quad \sigma_Y = 0{,}9207,$$

$$r(X,Y) = \frac{M(X \cdot Y) - MX \cdot MY}{\sigma_X \cdot \sigma_Y} = \frac{53{,}1 - 21{,}9 \cdot 2{,}18}{7{,}8352 \cdot 0{,}9207} = 0{,}7427.$$

Таким образом, между величинами X и Y тесная зависимость.

Если коэффициент корреляции двух случайных величин равен нулю, то величины X и Y называются *некоррелированными*.

Для нормального распределения понятия независимости и некоррелируемости совпадают, т.е. если две нормальные величины некоррелированы, то они независимы.

В общем случае это утверждение неверно: некоррелированные случайные величины могут быть зависимыми. Это подтверждается следующим примером.

Пример 5.2. Пусть совместное распределение величин X и Y задается табл. 5.3.

Таблица 5.3

X \ Y	0	1/6	1
–1	$\frac{1}{4}$	0	0
0	0	0	$\frac{1}{4}$
1	0	$\frac{1}{2}$	0

Тогда $MX = \frac{1}{3}$, $MY = \frac{1}{4}$,

$$M\big((X - MX)\cdot(Y - MY)\big) = \left(0 - \frac{1}{3}\right)\left(-1 - \frac{1}{4}\right)\cdot\frac{1}{4} +$$

$$+ \left(\frac{1}{6} - \frac{1}{3}\right)\left(1 - \frac{1}{4}\right)\cdot\frac{1}{2} + \left(1 - \frac{1}{3}\right)\left(0 - \frac{1}{4}\right)\cdot\frac{1}{4} = 0$$

$$\text{и} \quad r = \frac{M\big((X - MX)\cdot(Y - MY)\big)}{\sqrt{DX}\sqrt{DY}} = 0.$$

При этом

$$P\big((X > 0)\cdot(Y > 0)\big) = \frac{1}{2}, \quad P(X > 0) = \frac{3}{4}, \quad P(Y > 0) = \frac{1}{2}$$

$$\text{и} \quad P\big((X > 0)\cdot(Y > 0)\big) \neq P(X > 0)P(Y > 0).$$

Значит, величины X и Y зависимы, но некоррелированы.

5.5. Двумерное нормальное распределение

Двумерная случайная величина $\{X, Y\}$ называется *распределенной по двумерному нормальному закону*, если ее совместная плотность имеет вид

$$p(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a_x)^2}{\sigma_X^2} - 2\rho\frac{(x-a_x)(y-a_y)}{\sigma_X\sigma_Y} + \frac{(y-a_y)^2}{\sigma_Y^2}\right]\right\}.$$

Двумерный нормальный закон определяется параметрами a_x, a_y, σ_X, σ_Y, ρ, которые имеют следующий смысл:

$$a_x = MX, \quad a_y = MY, \quad \sigma_X^2 = DX, \quad \sigma_Y^2 = DY, \quad \rho = r(X, Y).$$

Одномерные плотности распределений составляющих X и Y находятся по формулам

$$p(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-a_x)^2}{2\sigma_X^2}}, \quad q(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{(y-a_y)^2}{2\sigma_Y^2}}.$$

Задачи с решениями

Задача 5.1. Пусть закон распределения двумерной случайной величины $\{X, Y\}$ задан табл. 5.2. Найти: а) законы распределения величин X и Y; б) условное распределение Y при условии $X = 30$.

Решение. а) Согласно формулам (5.1) и (5.2), суммируя вероятности в таблице по строкам и столбцам, получим распределения

x_i	10	20	30
p_i	0,23	0,35	0,42

y_i	1	2	3	4
q_i	0,25	0,42	0,23	0,1

б) Вычисляя вероятности $P_{(x=30)}(y_i)$ по формуле (5.3), получим условное распределение Y при условии $X = 30$:

y_i	1	2	3	4
$P_{(x=30)}(y_i)$	0,0476	0,2381	0,4762	0,2381

Задача 5.2. Пусть $F(x, y)$ — функция распределения двумерной случайной величины $\{X, Y\}$. Найти вероятность события $(\{X, Y\} \in D)$, где D — прямоугольник.

Решение. Введем события:

$$B_1 = \big((X < a_2) \cdot (Y < b_2)\big), \quad B_2 = \big((X < a_2) \cdot (Y < b_1)\big),$$

$$B_3 = \big((X < a_1) \cdot (Y < b_2)\big), \quad B_4 = \big((X < a_1) \cdot (Y < b_1)\big).$$

Тогда

$$P\big(\{X,Y\} \in D\big) = P(B_1) - P(B_2 + B_3) =$$

$$= P(B_1) - P(B_2) - P(B_3) + P(B_2 \cdot B_3) =$$

$$= P(B_1) - P(B_2) - P(B_3) + P(B_4) =$$

$$= F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1).$$

Задача 5.3. Пусть $p(x, y)$ — совместная плотность величин X и Y. Найти плотность $p(z)$ распределения величины $Z = X + Y$.

Решение. Найдем вначале функцию распределения $F(z)$ величины Z:

$$F(z) = P(Z < z) = P(X + Y < z).$$

Эта вероятность (см. § 5.3) равна

$$F(z) = \iint\limits_{x+y<z} p(x,y)dxdy = \int\limits_{-\infty}^{+\infty} dy \int\limits_{-\infty}^{z-y} p(x,y)dx.$$

Так как $p(z) = F'(z)$, то дифференцированием по z получим:

$$p(z) = \int\limits_{-\infty}^{+\infty} p(z - y, y)dy. \tag{5.9}$$

Формула (5.9) называется *формулой свертки*. Если X и Y независимы, то из (5.9) следует

$$p(z) = \int\limits_{-\infty}^{+\infty} p_1(z - y)p_2(y)dy, \tag{5.10}$$

где $p_1(x)$ и $p_2(y)$ — плотности случайных величин X и Y.

Задача 5.4. Пусть X_1 и X_2 независимы и распределены равномерно на интервалах $(0, a)$ и $(0, b)$ соответственно. Найти совместную плотность $p(x_1, x_2)$ величин X_1 и X_2.

Решение. Плотность величины X_1 равна

$$p_1(x_1) = \frac{1}{a} \quad \text{при} \quad 0 < x_1 < a,$$

плотность X_2 равна

$$p_2(x_2) = \frac{1}{b} \quad \text{при} \quad 0 < x_2 < b.$$

Пусть D — прямоугольник

$$D = \{(x_1, x_2) : 0 < x_1 < a,\ 0 < x_2 < b\}.$$

Тогда из (5.7) получим

$$p(x_1, x_2) = \frac{1}{ab} \quad \text{при} \quad (x_1, x_2) \in D$$

$$\text{и} \quad p(x_1, x_2) = 0 \quad \text{при} \quad (x_1, x_2) \notin D.$$

Задача 5.5. Найти коэффициент корреляции между величинами X и Y, совместный закон распределения которых задан табл. 5.2.

Решение. Последовательно находим

$$MX = \sum_i x_i p_i = 10 \cdot 0{,}23 + 20 \cdot 0{,}35 + 30 \cdot 0{,}42 = 21{,}9,$$

$$MY = \sum_j y_j q_j = 1 \cdot 0{,}25 + 2 \cdot 0{,}42 + 3 \cdot 0{,}23 + 4 \cdot 0{,}1 = 2{,}18,$$

$$MX^2 = \sum_i x_i^2 p_i = 100 \cdot 0{,}23 + 400 \cdot 0{,}35 + 900 \cdot 0{,}42 = 541,$$

$$M(X \cdot Y) = \sum_{i,j} x_i y_j p_{ij} = 10 \cdot 0{,}2 + 20 \cdot 0{,}02 + 30 \cdot 0{,}01 + 20 \cdot 0{,}03 +$$

$$+ 40 \cdot 0{,}03 + 60 \cdot 0{,}02 + 30 \cdot 0{,}02 + 60 \cdot 0{,}1 + 90 \cdot 0{,}2 + 120 \cdot 0{,}1 = 53{,}1.$$

Тогда

$$DX = MX^2 - (MX)^2 = 541 - 21{,}9^2 = 61{,}39,$$

$$\sigma_X = \sqrt{DX} = 7{,}8352, \quad \sigma_Y = 0{,}9207,$$

$$r(X,Y) = \frac{M(X \cdot Y) - MX \cdot MY}{\sigma_X \cdot \sigma_Y} = \frac{53{,}1 - 21{,}9 \cdot 2{,}18}{7{,}8352 \cdot 0{,}9207} = 0{,}7427.$$

Задача 5.6. Пусть X и Y — независимые нормальные стандартные величины. Показать, что

$$Z = X + Y \in N\left(0; \sqrt{2}\right).$$

Решение. По формуле свертки (5.9) плотность случайной величины Z

$$p(z) = \int_{-\infty}^{+\infty} \varphi(z-t)\varphi(t)dt,$$

где $\varphi(t)$ — плотность стандартного нормального распределения.

Вычислим интеграл

$$p(z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{(z-t)^2}{2}-\frac{t^2}{2}}dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{z^2-2tz+2t^2}{2}}dt.$$

Сделаем замену

$$\sqrt{2}\left(t-\frac{z}{2}\right) = u,$$

тогда

$$p(z) = \frac{1}{2\pi\sqrt{2}}\int_{-\infty}^{+\infty} e^{-\frac{u^2+\frac{z^2}{2}}{2}}du = \frac{e^{-\frac{z^2}{4}}}{2\pi\sqrt{2}}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}du = \frac{e^{-\frac{z^2}{4}}}{2\pi\sqrt{2}}\sqrt{2\pi} =$$

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

Очевидно, что $p(z)$ есть плотность нормального распределения с параметрами $a = 0$, $\sigma = \sqrt{2}$.

Задача 5.7. Двумерная случайная величина $\{X, Y\}$ распределена равномерно в круге $x^2 + y^2 \le 1$. Найти ее совместную плотность распределения и вероятность того, что расстояние от точки (x, y) до начала координат будет больше 0,6.

Решение. Равномерное распределение в области D на плоскости задается постоянной функцией плотности совместного распределения, равной $1/S_D$ (где S_D — площадь D) для точек из D, и нулю вне области D:

$$p(x,y) = \begin{cases} \dfrac{1}{S_D},\ (x,y) \in D, \\ 0,\ (x,y) \notin D. \end{cases}$$

В нашем случае область D является кругом с площадью, равной 1, поэтому $S_D = \pi$. Точки (x, y), для которых расстояние до начала координат больше 0,6 и $p(x, y) \neq 0$, образуют кольцо D_1 между окружностями радиусов 0,6 и 1. Площадь кольца D_1 равна $\pi - 0{,}36\pi = 0{,}64\pi$. Вероятность

$$p\big((x,y) \in D_1\big) = \iint\limits_{D_1} p(x,y)dxdy = \frac{1}{\pi} \cdot 0{,}64\pi = 0{,}64.$$

Задачи для самостоятельного решения

5.1. Найти совместную плотность распределения двумерной случайной величины $\{X, Y\}$, если ее функция распределения равна

$$F(x,y) = \sin x \cdot \sin y \left(0 \le x \le \frac{\pi}{2},\ 0 \le y \le \frac{\pi}{2}\right),$$

$F(x, y) = 0$ для остальных значений.

5.2. Независимые случайные величины X и Y распределены по показательным законам с параметрами $\lambda = 1/5$ и $\mu = 1/6$. Найти плотность распределения $p(z)$ их суммы $Z = X + Y$.

5.3. Двумерная дискретная случайная величина $\{X, Y\}$ задана законом распределения

x_i \ y_j	6	5	4	3
0,1	0,2	0,1	0,1	0
0,2	0,1	0,2	0,15	0
0,3	0	0	0,05	0,1

Найти условное распределение величины X при условии $Y = 4$. Являются ли величины X и Y независимыми?

5.4. Дискретная случайная величина X имеет закон распределения

x_k	0	1	2	3
p_k	0,1	0,3	0,4	0,2

Найти закон распределения случайной величины

$$Y = \sin\left(\frac{\pi}{2}X\right) + 1.$$

5.5. Случайная величина X может принимать ряд значений — 10, 20, 30 с вероятностью 1/3. Погрешности ее измерений Y не зависят от значений X и имеют закон распределения

y_k	0,2	0,3	0,4	0,5
q_k	0,2	0,3	0,2	0,3

Найти закон распределения относительной погрешности

$$Z = \frac{Y}{X}100\ \%.$$

5.6. Случайные величины X и Y независимы и распределены по показательным законам с параметрами $\lambda = 0{,}1$; $\lambda = 0{,}4$. Найти совместную плотность двумерной величины $\{X, Y\}$ и $M(X \cdot Y)$.

5.7. Случайная величина $\{X, Y\}$ распределена равномерно в квадрате $K = \{(x,y) \mid 0 \le x \le 1,\ 0 \le y \le 1\}$. Показать, что X и Y независимы.

5.8. Случайная величина называется распределенной по логарифмически нормальному закону, если $\ln X \in N(a; \sigma)$. Пусть $\ln X \in N(0; 1)$. Найти MX.

5.9. Совместная плотность величин X и Y равна

$$p(x,y) = \frac{3}{8}(x^2 + y^2), \quad |x| < 1, \quad |y| < 1.$$

Найти $r(x, y)$.

Ответы

5.1. $p(x,y) = \cos x \cdot \cos y \left(0 \le x \le \frac{\pi}{2},\ 0 \le y \le \frac{\pi}{2}\right)$.

5.2. $p(v) = e^{-\frac{z}{6}}\left(1 - e^{-\frac{z}{30}}\right)$.

5.3.

x_1	0,1	0,2	0,3
r_i	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Величины X и Y не являются независимыми.

5.4.

y_k	0	1	2
q_k	0,2	0,5	0,3

5.5.

$\frac{Y}{X}100\ \%$	$\frac{2}{3}$	1	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	2	$\frac{5}{2}$	3	4	5
p	$\frac{2}{30}$	$\frac{5}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{3}{30}$

5.6. $P_{\{X,Y\}}(x,y) = p_X(x) \cdot p_Y(y) = \begin{cases} 25e^{-0,1x-0,4y},\ x \geq 0,\ y \geq 0, \\ 0,\ \text{в остальных случаях;} \end{cases}$ $M(X \cdot Y) = 25.$

5.8. $MX = \sqrt{e}$.

5.9. $r(x, y) = 0$.

6

ЗАКОН БОЛЬШИХ ЧИСЕЛ

В большой серии независимых однотипных опытов исход каждого отдельного опыта является случайным, неопределенным, однако средний результат всей серии опытов обладает определенными закономерностями. Такого типа закономерности, свойственные массовым случайным событиям, и составляют предмет теории вероятностей. Например, статистическое определение вероятности события основано на использовании таких закономерностей. В теории вероятностей доказан ряд теорем об отклонениях средних характеристик большого числа случайных величин от некоторых определенных постоянных. Эти теоремы носят название закона больших чисел. В основе их доказательства лежит неравенство П.Л. Чебышева, доказанное им в 1845 г.

6.1. Неравенства Чебышева и Маркова

Если значения случайной величины X неотрицательны и существует математическое ожидание $MX = a$, то для любого $\varepsilon > 0$

$$P(X \le \varepsilon) \ge 1 - \frac{a}{\varepsilon}. \tag{6.1}$$

Неравенство (6.1) называется *неравенством Маркова*.

Если случайная величина имеет дисперсию DX, то справедливо *неравенство Чебышева*

$$P(|X - a| \le \varepsilon) \ge 1 - \frac{DX}{\varepsilon^2}. \tag{6.2}$$

Пример 6.1. Среднее число дождливых дней в году в данном районе равно 80. Оценить вероятность того, что в этом районе будет не более 100 дождливых дней в году.

Пусть X — число дождливых дней в году. По неравенству Маркова

$$P(X \le 100) \ge 1 - \frac{80}{100} \Leftrightarrow P(X \le 100) \ge 0{,}2.$$

6.2. Закон больших чисел

Рассмотрим закон больших чисел в упрощенной формулировке, пригодной для решения практических задач.

Если случайные величины X_i, $i = 1, 2, ..., n$, независимы и одинаково распределены со средними $MX_i = a$ и дисперсиями $DX_i = DX$, то справедлива *теорема Чебышева*:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \le \varepsilon\right) \ge 1 - \frac{DX}{n\varepsilon^2}. \tag{6.3}$$

Из неравенства (6.3) при $n \to \infty$ следует ***закон больших чисел***

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \le \varepsilon\right) = 1. \tag{6.4}$$

Смысл закона больших чисел заключается в том, что средние значения совокупности случайных величин стремятся по вероятности к их математическому ожиданию при $n \to \infty$. Отклонение средних значений от математического ожидания становится сколь угодно малым с вероятностью, близкой к единице, если n достаточно велико. Другими словами, вероятность любого отклонения средних значений от математического ожидания сколь угодно мала с ростом n.

Пример 6.2. Номинальное значение диаметра втулки равно 5 мм, а дисперсия, вследствие погрешностей изготовления, не превосходит 0,01. Оценить вероятность того, что размер втулки будет отличаться от номинала не более чем на 0,5 мм.

В силу неравенства Чебышева

$$P\left(|X - a| \le \varepsilon\right) \ge 1 - \frac{DX}{\varepsilon^2} = 1 - \frac{0{,}01}{0{,}5} = 0{,}98.$$

6.3. Теорема Бернулли

Частным случаем закона больших чисел является ***теорема Бернулли***.

Пусть имеется n испытаний Бернулли с вероятностью успеха p, $q = 1 - p$ и m — число успехов. Тогда для любого $\varepsilon > 0$

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \ge 1 - \frac{pq}{n\varepsilon^2}. \tag{6.5}$$

Отсюда следует ***закон больших чисел в форме Бернулли***:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 1. \tag{6.6}$$

Теорема Бернулли утверждает, что частость события A стремится к его вероятности, так что при больших n отклонение частости m/n от вероятности p становится сколь угодно малым. Эта теорема в некотором смысле обосновывает статистическое определение вероятности.

Пример 6.3. Сколько следует провести независимых испытаний, чтобы вероятность выполнения неравенства $\left|\frac{m}{n} - p\right| \le 0{,}05$ превысила 0,75, если вероятность появления данного события в отдельном испытании $p = 0{,}8$?

По теореме Бернулли получаем

$$P\left(\left|\frac{m}{n} - p\right| \le 0{,}08\right) > 1 - \frac{0{,}8 \cdot 0{,}2}{n \cdot 0{,}0064} > 0{,}75 \Rightarrow$$

$$\Rightarrow 0{,}25 > \frac{0{,}16}{0{,}0064 \cdot n} \Rightarrow n > 100.$$

Пример 6.4. Определить, сколько надо произвести замеров диаметров деревьев, чтобы средний диаметр деревьев отличался от истинного значения не более чем на 2 см с вероятностью, не меньшей 0,95. Известно, что на данном участке среднее квадратичное отклонение диаметров деревьев не превышает 10 см.

Из неравенства (6.3) получаем

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < 2\right) \ge 1 - \frac{10^2}{n \cdot 2^2} \ge 0{,}95 \Rightarrow 0{,}05 \ge \frac{100}{4n}, \quad n \ge 500.$$

Пример 6.5. Для определения урожайности поля площадью 200 га взяли выборку с каждого гектара. Известно, что по каждому гектару поля дисперсия не превышает 2. Оценить вероятность того, что отклонение средней выборочной урожайности от средней урожайности не превосходит 0,2 ц.

По теореме Чебышева искомая вероятность

$$P > 1 - \frac{D}{n\varepsilon^2} = 1 - \frac{2}{200 \cdot 0{,}04} = 0{,}75.$$

Задачи для самостоятельного решения

6.1. Средний вес клубня картофеля равен 120 г. Какова вероятность того, что наугад взятый клубень картофеля весит не более 300 г?

6.2. Случайная величина X имеет дисперсию $DX = 0{,}01$. Какова вероятность того, что случайная величина X отличается от $MX = a$ не более чем на 0,25?

6.3. Случайная величина X имеет закон распределения

x_k	2	3	4	4,5	5	6
p_k	0,1	0,4	0,3	0,05	0,05	0,1

Пользуясь неравенством Маркова, оценить вероятность того, что случайная величина X примет значение, не большее 4.

6.4. Всхожесть семян некоторого растения составляет 60 %. Найти вероятность того, что при посеве 10 000 семян отклонение доли взошедших семян от вероятности того, что взойдет каждое из них, не превосходит 0,01.

6.5. Расход технической воды на предприятии составляет 5000 л в день, а среднее квадратичное отклонение этой величины не превышает 1000 л. Оценить вероятность того, что расход воды в день не превышает 10 000 л.

6.6. Среднее изменение курса акций в течение месяца составляет 0,5 %. Оценить вероятность того, что в течение месяца курс акций изменится не более чем на 2 %.

6.7. Среднее число клиентов, обслуживаемых банком за день, равно 30. Оценить вероятность того, что в течение дня будет обслужено не более 60 человек.

6.8. Электронная система состоит из 45 элементов одинаковой надежности, равной 0,95, за время T. Найти вероятность того, что доля безотказно работавших элементов в течение времени T отличается от 0,95 не более чем на 0,1.

6.9. Известно, что $P(X < 1) = 0{,}95$. В каких пределах находится MX, если X — неотрицательная случайная величина?

6.10. Среднее количество звонков в офис фирмы в течение дня равно 30. Оценить вероятность того, что в течение дня будет более 90 звонков.

6.11. Опыт страховой компании показывает, что на каждый пятый договор приходится страховой случай. Сколько договоров нужно заключить, чтобы с вероятностью 0,95 можно было утверждать, что доля страховых случаев отклонится от 0,2 не более чем на 0,05?

6.12. Сколько надо произвести измерений данной величины, чтобы с вероятностью не менее 0,95 гарантировать отклонение средней арифметической этих измерений от истинного значения величины не более чем на 1, если среднее квадратичное отклонение (ошибка измерений) не превосходит 3?

6.13. Вероятность того, что акции, переданные на депозит, будут востребованы, равна 0,075. Оценить вероятность того, что из 1000 клиентов от 60 до 90 востребуют свои акции.

6.14. Известно, что $P\left(\left|X - MX\right| < 1\right) = 0{,}95$. В каких пределах находится среднее квадратичное отклонение σ_X?

Ответы

6.1. $p = 0{,}6$.
6.2. $p = 0{,}84$.
6.3. $p = 0{,}9188$.
6.4. $p = 0{,}76$.
6.5. $p \geq 0{,}96$.
6.6. $p \geq 0{,}75$.
6.7. $p \geq 0{,}5$.
6.8. $p \geq 0{,}8944$.
6.9. $MX \geq 0{,}05$.
6.10. $p \leq 2/3$.
6.11. $n \geq 1280$.
6.12. $n \geq 180$.
6.13. $p \geq 0{,}692$.
6.14. $\sigma_X \geq 0{,}0224$.

ЧАСТЬ II

Математическая статистика

- Вариационные ряды и их характеристики
- Оценка параметров генеральной совокупности
- Проверка статистических гипотез
- Однофакторный дисперсионный анализ
- Корреляционный анализ
- Регрессионный анализ

Предметом математической статистики является изучение случайных событий и случайных величин по результатам наблюдений. *Статистической совокупностью* называется совокупность предметов или явлений, объединенных каким-либо общим признаком, который сохраняет свои свойства при многократных наблюдениях. Результатом наблюдений над статистической совокупностью являются *статистические данные* — сведения о том, какие значения принял в итоге наблюдений интересующий нас признак (случайная величина X).

Обработка статистических данных методами математической статистики приводит к установлению определенных закономерностей, присущих массовым явлениям. При этом точность статистических выводов повышается с ростом числа наблюдений.

При изучении случайной величины мы будем основываться на двояком подходе к множеству ее значений. Опишем эти подходы.

Статистические данные, как правило, представляют собой ряд значений $\{x_1, x_2, ..., x_n\}$ некоторой случайной величины X, исследование которой начинается с обработки этого ряда значений. Затем строятся функции, характеризующие случайную величину. Эти функции называются статистиками.

Таким образом, *статистика* — это функция

$$T: (x_1, x_2, ..., x_n) \to T(x_1, x_2, ..., x_n),$$

которая набору значений (реализации) $\{x_1, x_2, ..., x_n\}$ случайной величины сопоставляет по некоторому правилу действительное число. Статистика является функцией от реализации случайной величины. Поскольку аргументы функции T являются значениями случайной величины, значения самой функции T также становятся реализацией случайной величины.

В теоретических исследованиях удобно рассматривать статистику T как функцию от случайных величин $X_1, X_2, ..., X_n$, каждая из которых имеет такое же распределение вероятностей, как и сама случайная величина X:

$$T = T(X_1, X_2, ..., X_n).$$

Таким образом, изучая случайную величину X, мы рассматриваем набор одинаково распределенных, независимых в совокупности случайных величин $\{X_1, X_2, ..., X_n\}$. В такой трактовке статистика становится случайной величиной и изучение ее распределения приводит к выводам о распределении самой случайной величины X.

Пусть X — случайная величина и имеется n ее наблюдений x_1, x_2, ..., x_n. Простейшей статистикой этой случайной величины является среднее значение

$$T(x_1, x_2, ..., x_n) = \frac{1}{n}(x_1 + x_2 + ... + x_n).$$

Вместе со случайной величиной X возникает набор случайных величин $\{X_1, X_2, ..., X_n\}$, распределенных так же, как и X, и новая случайная величина

$$T(X_1, X_2, ..., X_n) = \bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n).$$

7

ВАРИАЦИОННЫЕ РЯДЫ И ИХ ХАРАКТЕРИСТИКИ

Вариационный ряд является статистическим аналогом случайной величины и служит для ее изучения. Однако вариационные ряды имеют свои особенности, связанные с отсутствием полной информации. И здесь на первый план вместо точных значений выходят их оценки и методы получения.

7.1. Генеральная и выборочная совокупности

Генеральной совокупностью называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе.

В математической статистике генеральная совокупность часто понимается как совокупность всех мыслимых наблюдений, которые могли быть произведены при выполнении некоторых условий. Понятие генеральной совокупности аналогично понятию случайной величины (закону распределения вероятностей), так как обе они полностью определяются заданными условиями.

Генеральная совокупность может быть конечной или бесконечной.

Объемом генеральной совокупности называется число ее объектов (наблюдений).

Выборочной совокупностью называется часть объектов генеральной совокупности, используемая для исследования.

Пример 7.1. Число единиц товара N, произведенного фирмой в течение года, есть конечная генеральная совокупность. Для исследования качества продукции на практике рассматривается выборка, состоящая из n единиц товара. Признаком, или случайной величиной, может быть число единиц товара, удовлетворяющих сертификатным требованиям.

Сущность выборочного метода в математической статистике заключается в том, чтобы по определенной части генеральной совокупности (выборке) судить о ее свойствах в целом.

Выборочный метод является уместным особенно в случае бесконечной генеральной совокупности или когда исследование связано с уничтожением (гибелью) наблюдаемых объектов (например, исследование предельных режимов приборов, исследование действия вирусов на подопытных животных и т.д.). Для того чтобы по выборке можно было адекватно судить о случайной величине, она должна быть *представительной* (репрезентативной). Репрезентативность выборки обеспечивается объемом выборки и случайностью отбора ее элементов, при котором все элементы генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.

Имеются два основных способа образования выборки:

- *повторная выборка*, когда каждый элемент, случайно отобранный и исследованный, возвращается в общую совокупность и может быть отобран повторно;
- *бесповторная выборка*, когда отобранный элемент не возвращается в общую совокупность.

7.2. Понятие вариационного ряда и его графические изображения

Пусть некоторый признак генеральной совокупности описывается случайной величиной X. Рассмотрим выборку $\{x_1, x_2, ..., x_n\}$ объема n из генеральной совокупности. Элементы этой выборки представляют собой значения случайной величины X.

На первом этапе статистической обработки производят *ранжирование* выборки, т.е. упорядочивание чисел $x_1, x_2, ..., x_n$ по возрастанию.

Вариантами x_i называются различные элементы выборки.

Частотой варианты x_i называется число m_i, показывающее, сколько раз эта варианта встречается в выборке.

Частостью, *относительной частотой*, или *долей варианты* называется число

$$\omega_i = \frac{m_i}{n}. \tag{7.1}$$

Частоты и частости называются *весами*.

Пусть x некоторое число. Тогда количество вариант m_x, значения которых меньше x, называется *накопленной частотой*, т.е.

$$m_x = \sum_{x_i < x} m_i. \tag{7.2}$$

Отношение накопленной частоты к общему числу наблюдений n называется *накопленной частостью*:

$$\omega_x = \frac{m_x}{n} = \frac{1}{n} \sum_{x_i < x} m_i. \tag{7.3}$$

Вариационным рядом называется ряд вариант, расположенных в порядке возрастания их значений, с соответствующими им весами. Вариационные ряды бывают дискретными и интервальными.

Дискретным называется вариационный ряд, который представляет собой выборку значений дискретной случайной величины.

Общий вид дискретного вариационного ряда показан в табл. 7.1.

Таблица 7.1

Варианты	x_1	x_2	...	x_k
Частоты	m_1	m_2	...	m_k

Непрерывным (*интервальным*) называется ряд, который представляет выборку значений непрерывной случайной величины. Для построения интервального вариационного ряда разбивают множество значений вариант на полуинтервалы $[a_i, a_{i+1})$, т.е. производят их группировку. Рекомендуется количество интервалов k выбирать по формуле Стерджерса

$$k = 1 + 1{,}4 \ln n. \tag{7.4}$$

Длина интервала

$$\Delta = \frac{x_{\max} - x_{\min}}{k}. \tag{7.5}$$

Подсчитывая число значений, попавших в полуинтервал $[a_i, a_{i+1})$, получим значения частот m_i, $i = 1, 2, ..., k$. В итоге интервальный ряд можно представить таблицей (табл. 7.2):

Таблица 7.2

Варианты	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_k, a_{k+1})$
Частоты	m_1	m_2	...	m_k

Если варианта находится на границе интервала, то ее присоединяют к правому интервалу.

Для наглядности представления используют графические изображения вариационных рядов в виде полигона, гистограммы и кумулянты.

Полигон служит для изображения дискретного вариационного ряда и представляет собой ломаную, соединяющую точки плоскости с координатами (x_i, m_i), $i = 1, 2, ..., k$. Для интервального ряда также строится полигон, только его ломаная проходит через точки (c_i, m_i), где $c_i = (a_i + a_{i+1})/2$, $i = 1, 2, ..., k$.

Гистограмма служит только для представления интервальных вариационных рядов и имеет вид ступенчатой фигуры из прямоугольников с основаниями, равными длине интервалов Δ, и высотами, равными частотам m_i интервалов.

Кумулянта представляет собой ломаную, соединяющую точки с координатами (x_i, m_{x_i}) (где m_{x_i} — накопленные частоты) для дискретного ряда или точки с координатами (a_i, m_{a_i}) для интервального ряда.

Эмпирической функцией распределения $F_n(x)$ называется функция, значение которой в точке x равно накопленной частости (рис. 7.1):

$$F_n(x) = \omega_x = \frac{m_x}{n}. \qquad (7.6)$$

Для интервального ряда указываются не конкретные значения вариант, а только их частоты на интервалах. В этом случае эмпирическая функция распределения определена только на концах интервалов. Ее можно изобразить ломаной, проходящей через точки $(a_i, F_n(a_i))$, $i = 1, 2, ..., k$ (рис. 7.2).

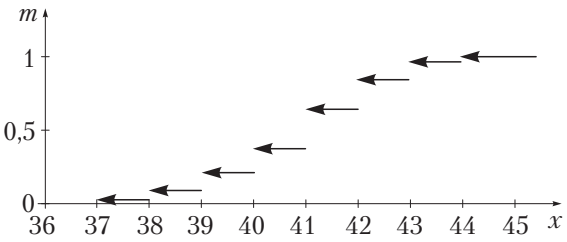


Рис. 7.1. Эмпирическая функция распределения дискретного ряда

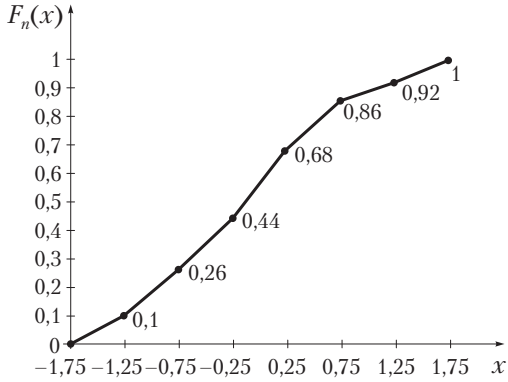


Рис. 7.2. Эмпирическая функция распределения интервального ряда

Эмпирической плотностью распределения непрерывного вариационного ряда называется функция

$$f_n(x) = \begin{cases} \dfrac{m_i}{n\Delta}, \text{ если } a_i \le x \le a_{i+1},\ i = 1, 2, ..., k, \\ 0, \text{ если } x < a_1 \text{ или } x > a_{k+1}. \end{cases} \qquad (7.7)$$

Функция $f_n(x)$ является аналогом плотности распределения случайной величины. Площадь области под графиком этой функции равна единице.

Пример 7.2. В магазине за день было продано 45 пар мужской обуви. Имеется выборка значений случайной величины X — размера обуви:

39, 41, 40, 42, 41, 40, 42, 44, 40, 43, 42, 41, 43, 39, 42,
41, 42, 39, 41, 37, 43, 41, 38, 43, 42, 41, 40, 41, 38, 44,
40, 39, 41, 40, 42, 40, 41, 42, 40, 43, 38, 39, 41, 41, 42.

Построить дискретный вариационный ряд, полигон, кумулянту и эмпирическую функцию распределения.

Для построения вариационного ряда различные значения признака располагаем в порядке их возрастания и под каждым из их значений записываем его частоту (табл. 7.3).

Таблица 7.3

x_i	37	38	39	40	41	42	43	44
m_i	1	3	5	8	12	9	5	2

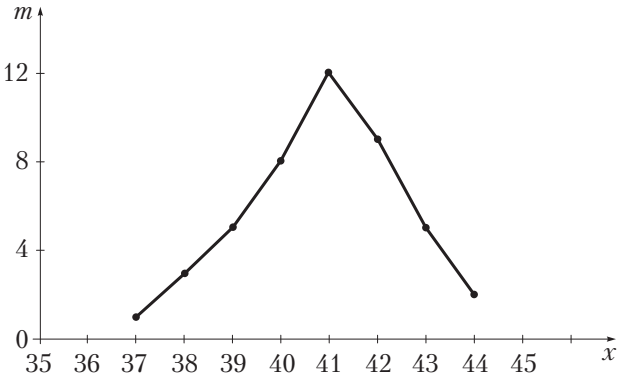


Рис. 7.3. Полигон

Полигон этого распределения изображен на рис. 7.3.

По данным табл. 7.3 находим накопленные частоты и частости (табл. 7.4).

Таблица 7.4

x_i	37	38	39	40	41	42	43	44	45
m_{x_i}	0	1	4	9	17	29	38	43	45
ω_{x_i}	0	0,022	0,089	0,2	0,378	0,644	0,844	0,978	1

На рис. 7.4 изображена кумулянта, а на рис. 7.5 — полигон и гистограмма интервального ряда.

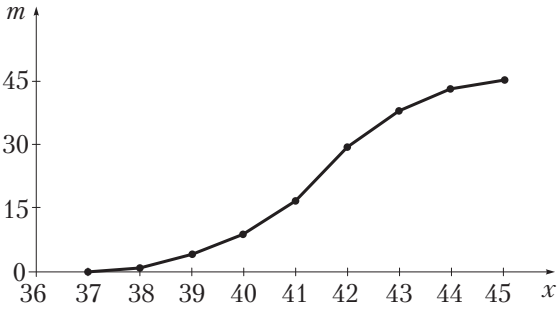


Рис. 7.4. Кумулянта

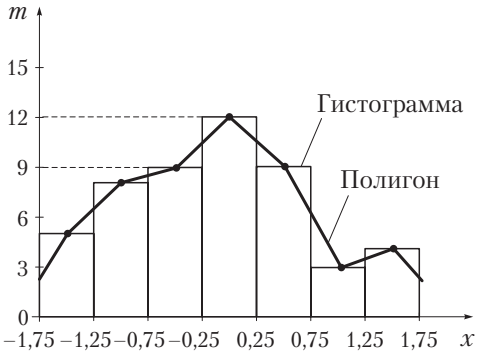


Рис. 7.5. Полигон и гистограмма интервального ряда

Пример 7.3. Результаты измерений отклонений от номинала диаметров 50 подшипников дали численные значения (мкм), приведенные в табл. 7.5.

Для данной выборки построить интервальный вариационный ряд, полигон, гистограмму, графики эмпирической функции распределения и эмпирической плотности распределения.

Таблица 7.5

–1,752	–0,291	–0,933	–0,450	0,512
–1,256	1,701	0,634	0,720	0,490
1,531	–0,433	1,409	1,730	–0,266
–0,058	0,248	–0,095	–1,488	–0,361
0,415	–1,382	0,129	–0,361	–0,087
–0,329	0,086	0,130	–0,244	–0,882
0,318	–1,087	0,899	1,028	–1,304
0,349	–0,293	–0,883	–0,056	0,757
–0,059	–0,539	–0,078	0,229	0,194
–1,084	0,318	0,367	–0,992	0,529

По данным табл. 7.5 определяем $x_{\min} = -1{,}752$; $x_{\max} = -1{,}73$. Разобьем множество значений выборки на интервалы. Число интервалов находим по формуле (7.4):

$$k = 1 + 1{,}4 \ln 50 \cong 6{,}447.$$

Выберем: $k = 7$, начало первого интервала $a_1 = -1{,}75$, конец последнего, седьмого, интервала $a_8 = 1{,}75$. При этом варианту $x_1 = -1{,}752$ отнесем в первый интервал. Длина каждого интервала будет равна

$$\Delta = \frac{a_8 - a_1}{k} = \frac{1{,}75 + 1{,}75}{7} = 0{,}5.$$

Подсчитывая число вариант, попадающих в каждый интервал, получим вариационный ряд (табл. 7.6).

Таблица 7.6

$[a_i, a_{i+1})$	[–1,75, –1,25)	[–1,25, –0,75)	[–0,75, –0,25)	[–0,25, 0,25)
m_i	5	8	9	12

$[a_i, a_{i+1})$	[0,25, 0,75)	[0,75, 1,25)	[1,25, 1,75)
m_i	9	3	4

Для построения эмпирической функции распределения вычислим накопленные частоты ω_{a_i} (табл. 7.7).

Таблица 7.7

a_i	–1,75	–1,25	–0,75	–0,25	0,25	0,75	1,25	1,75
ω_{a_i}	0	0,1	0,26	0,44	0,68	0,86	0,92	1

По формуле (7.7) вычислим значения

$$\rho_i = \frac{m_i}{n\Delta}$$

эмпирической плотности вероятности для каждого интервала (табл. 7.8).

Таблица 7.8

$[a_i, a_{i+1})$	[–1,75, –1,25)	[–1,25, –0,75)	[–0,75, –0,25)	[–0,25, 0,25)
ρ_i	5	8	9	12

$[a_i, a_{i+1})$	[0,25, 0,75)	[0,75, 1,25)	[1,25, 1,75)
ρ_i	9	3	4

7.3. Числовые характеристики вариационных рядов

Основной характеристикой вариационного ряда является его *средняя арифметическая*, называемая также *выборочным средним*.

Для дискретного вариационного ряда выборочная средняя

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{7.8}$$

для интервального ряда —

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \sum_{i=1}^{k} x_i \omega_i. \tag{7.9}$$

В последней формуле за x_i принимают середину i-го интервала.

Вариационным размахом называется число $R = x_{\max} - x_{\min}$.

Выборочной дисперсией называется средняя арифметическая квадратов отклонений вариант от их выборочной средней:

$$S^2 = \sum_{i=1}^{k}(x_i - \bar{x})^2 \frac{m_i}{n} = \sum_{i=1}^{k}(x_i - \bar{x})^2 \omega_i, \tag{7.10}$$

где x_i — это значение варианты для дискретного ряда или середина i-го интервала для интервального ряда.

Для практических вычислений S^2 более удобной является следующая формула:

$$S^2 = \overline{x^2} - (\bar{x})^2. \tag{7.11}$$

Свойства выборочной дисперсии:

1. Дисперсия постоянной равна нулю.

2. Если ко всем вариантам добавить постоянное число, то дисперсия не изменится.

3. Если все варианты умножить на одно и то же число k, то дисперсия умножится на k^2.

4. *Правило сложения дисперсий*. Пусть значения выборки разбиты на l групп. Обозначим через n_i, $i = 1, 2, ..., l$, количество различных вариант в i-й группе, через n_{ij} — частоту j-й варианты в этой группе. Тогда i-я группа ряда записывается в виде $\{x_{i1}, x_{i2},$

..., $\bar{x}_{in_i}\}$, $i = 1, 2, ..., l$, при этом значение x_{ij} повторяется n_{ij} раз. Обозначим через $\bar{x}_i$ групповые средние арифметические:

$$\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} \frac{n_{ij}}{n_i}.$$

Групповые дисперсии S_i^2

$$S_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \frac{n_{ij}}{n_i}. \tag{7.12}$$

Средняя арифметическая групповых дисперсий

$$\bar{S}_{(i)}^2 = \sum_{i=1}^{l} S_i^2 \cdot \frac{n_i}{n}. \tag{7.13}$$

Межгрупповая дисперсия

$$\delta^2 = \sum_{i=1}^{l} (\bar{x}_i - \bar{x})^2 \frac{n_i}{n}. \tag{7.14}$$

Правилом сложения дисперсий называется равенство

$$S^2 = \bar{S}_{(i)}^2 + \delta^2. \tag{7.15}$$

Равенство (7.15) служит основой для дисперсионного анализа.

Мерой вариации признака является *выборочное среднее квадратичное отклонение*, которое определяется как корень из дисперсии:

$$S = \sqrt{\sum_{i=1}^{k} (x_i - \bar{x})^2 \frac{m_i}{n}}. \tag{7.16}$$

В статистическом анализе рассматривается также *коэффициент вариации*, равный процентному отношению выборочного среднего квадратичного отклонения к выборочной средней:

$$\nu = \frac{S}{|\bar{x}|} \cdot 100, \quad \bar{x} \neq 0. \tag{7.17}$$

Задачи с решениями

Задача 7.1. В табл. 7.9 приведены данные об урожайности ржи на различных участках поля.

Таблица 7.9

Урожайность ржи, ц/га	9–12	12–15	15–18	18–21	21–24	24–27
Среднее урожайности, ц/га	10,5	13,5	16,5	19,5	22,5	25,5
Доля участка в общей площади, %	6	12	33	22	19	8

Найти выборочную дисперсию, коэффициент вариации и размах вариации признака X, обозначающего урожайность ржи.

Решение. Выборочное среднее вычислим по формуле (7.9):

$$\bar{x} = \frac{1}{100}(10{,}5 \cdot 6 + 13{,}5 \cdot 12 + 16{,}5 \cdot 33 + 19{,}5 \cdot 22 + 22{,}5 \cdot 19 + 25{,}5 \cdot 8) = 18{,}3.$$

По формуле (7.10) находим выборочную дисперсию:

$$S^2 = [(10{,}5 - 18{,}3)^2 \cdot 6 + (13{,}5 - 18{,}3)^2 \cdot 12 + (16{,}5 - 18{,}3)^2 \cdot 33 + (19{,}5 - 18{,}3)^2 \cdot 22 + (22{,}5 - 18{,}3)^2 \cdot 19 + (25{,}5 - 18{,}3)^2 \cdot 8]/100 = 15{,}3.$$

Тогда $S = 3{,}9$, коэффициент вариации по формуле (7.17) равен

$$\nu = \frac{3{,}9}{18{,}3} \cdot 100 = 21{,}3\ \%.$$

Размах вариации $R = 27 - 9 = 18$.

Задача 7.2. Выполнение норм выработки рабочих характеризуется данными, приведенными в табл. 7.10. Найти выборочную дисперсию и коэффициент вариации.

Таблица 7.10

Процент выполнения норм выработки, %	90–100	100–110	110–120	120–130	130–140
Число рабочих	10	160	100	60	20

Решение. Для упрощения расчетов введем новые варианты

$$U_i = \frac{x_i - C}{h},$$

где $C = 115 \approx \bar{x} = 112{,}71$, $h = 10$.

Тогда U_i принимают значения $U_1 = -2$, $U_2 = -1$, $U_3 = 0$, $U_4 = 1$, $U_5 = 2$,

$$\bar{U} = \frac{-2 \cdot 10 - 1 \cdot 160 + 1 \cdot 60 + 2 \cdot 20}{350} = -0{,}229,$$

$$\overline{U^2} = \frac{4 \cdot 10 + 160 + 60 + 4 \cdot 20}{350} = 0{,}971,$$

$$S_U^2 = \overline{U^2} - (\bar{U})^2 = 0{,}971 - 0{,}229^2 = 0{,}919.$$

$$\bar{x} = h\bar{U} + C = 10 \cdot (-0{,}229) + 115 = 112{,}71,$$

$$S^2 = h^2 \cdot S_U^2 = 100 \cdot (-0{,}229) = 91{,}9,$$

$$S = \sqrt{91{,}9} = 9{,}586.$$

Коэффицент вариации

$$\nu = \frac{S}{|\bar{x}|} \cdot 100 = \frac{9{,}586}{112{,}71} \cdot 100 = 8{,}5\ \%.$$

Задача 7.3. В табл. 7.11 приведено распределение $n = 50$ рабочих по производительности труда X (единиц за смену), разделенных на две группы. Найти общие групповые средние и проверить правило сложения дисперсий.

Таблица 7.11

Производительность труда	Прошедшие техническое обучение (группа 1)					Не прошедшие техническое обучение (группа 2)				
x_{ij}	34	85	96	102	103	63	69	83	89	106
m_{ij}	5	2	11	8	4	2	6	8	3	1

Решение. Для вычисления средних и дисперсий введем новые варианты

$$U_{ij} = \frac{x_{ij} - C}{h}, \quad i = 1, 2; \quad j = 1, 2, 3, 4, 5,$$

где $C = 80$, $h = 10$. Тогда распределение U задается табл. 7.12.

Таблица 7.12

U_{ij}	–4,6	0,5	1,5	2,2	2.3	–1,7	–1,1	0,3	0,9	2,6
m_{ij}	5	2	11	8	4	2	6	8	3	1

Определим групповые средние и дисперсии. Здесь $n = 50$, $n_1 = 30$, $n_2 = 20$, $l = 2$,

$$\overline{U_1} = \frac{-4,6 \cdot 5 + 0,5 \cdot 2 + 1,6 \cdot 11 + 2,2 \cdot 8 + 2,3 \cdot 4}{30} = 0,7476,$$

$$\overline{U_1^2} = \frac{-4,6^2 \cdot 5 + 0,5^2 \cdot 2 + 1,6^2 \cdot 11 + 2,2^2 \cdot 8 + 2,3^2 \cdot 4}{30} = 6,478,$$

$$S_{U_1}^2 = \overline{U_1^2} - \left(\overline{U_1}\right)^2 = 6,478 - 0,7476^2 = 5,92,$$

$$\overline{U_2} = \frac{-1,7 \cdot 2 - 1,1 \cdot 6 + 0,3 \cdot 8 + 0,9 \cdot 3 + 2,6}{20} = -0,115,$$

$$\overline{U_2^2} = \frac{-1,7^2 \cdot 2 - 1,1^2 \cdot 6 + 0,3^2 \cdot 8 + 0,9^2 \cdot 3 + 2,6^2}{20} = 1,1343,$$

$$\overline{x_1} = h \cdot \overline{U_1} + C = 10 \cdot 0,7467 + 80 = 87,467,$$

$$S_1^2 = h^2 \cdot S_{U_1}^2 = 100 \cdot 5,92 = 592,$$

$$\overline{x_2} = h \cdot \overline{U_2} + C = 10 \cdot (-0,115) + 80 = 78,85,$$

$$S_2^2 = h^2 \cdot S_{U_2}^2 = 100 \cdot 1,1343 = 113,432.$$

Находим общие среднюю и дисперсию:

$$\bar{U} = \frac{1}{50}(-4,6 \cdot 5 + 0,5 \cdot 2 + 1,6 \cdot 11 + 2,2 \cdot 8 + 2,3 \cdot 4 - 1,7 \cdot 2 - 1,1 \cdot 6 +$$

$$+ 0,3 \cdot 8 + 0,9 \cdot 3 + 2,6) = 0,402,$$

$$\overline{U^2} = \frac{1}{50}(-4,6^2 \cdot 5 + 0,5^2 \cdot 2 + 1,6^2 \cdot 11 + 2,2^2 \cdot 8 + 2,3^2 \cdot 4 -$$

$$- 1,7^2 \cdot 2 - 1,1^2 \cdot 6 + 0,3^2 \cdot 8 + 0,9^2 \cdot 3 + 2,6^2) = 4,3458,$$

$$S_U^2 = \overline{U^2} - (\bar{U})^2 = 4,3458 - 0,402^2 = 4,1842,$$

$$\bar{x} = h\bar{U} + C = 10 \cdot 0,402 + 80 = 84,02,$$

$$S^2 = h^2 \cdot S_U^2 = 100 \cdot 4{,}1842 = 418{,}42,$$

$$\overline{S_{(i)}^2} = 592{,}04 \cdot \frac{30}{50} = 113{,}432 \cdot \frac{20}{50} = 400{,}596,$$

$$\delta^2 = (87{,}467 - 84{,}02)^2 \cdot \frac{3}{5} + (78{,}85 - 84{,}02)^2 \cdot \frac{2}{5} = 17{,}8207.$$

Так как $\overline{S_{(i)}^2} + \delta^2 = 418{,}417$, то $S^2 = S_{(i)}^2 + \delta^2$, поэтому правило сложения дисперсий выполняется.

Задачи для самостоятельного решения

7.1. Ошибки l в третьем знаке при округлении чисел заданы в табл. 7.13.

Таблица 7.13

–1,000	–0,998	–0,993	–0,975	–0,911	–0,688	0,068	0,604	–0,986	0,645
0,746	0,677	–0,659	–0,048	–0,355	0,297	0,981	–0,786	0,452	–0,213
0,747	–0,162	–0,857	0,317	–0,382	0,853	0,555	–0,347	0,922	0,653
–0,383	–0,172	0,414	0,030	0,459	0,483	0,769	0,265	0,669	–0,365
–0,213	0,004	–0,059	–0,388	0,200	0,692	0,356	–0,091	0,245	0,295

Построить вариационный ряд и найти $\overline{l}, S_l^2$ для выборки случайной величины l.

7.2. В табл. 7.14 приведен ряд моментов t срока работы электрических лампочек (в годах). Построить вариационный ряд, найти $\overline{t}$ и S_t^2 выборки.

Таблица 7.14

0,001	0,001	0,003	0,012	0,046	0,169	0,763	1,620	0,007	1,728
2,067	1,824	0,187	0,646	0,389	1,046	4,672	0,113	1,295	0,500
1,754	0,543	0,074	1,075	0,370	2,612	1,504	0,396	3,245	1,751
0,369	0,534	1,227	0,724	1,307	1,353	2,157	1,000	1,800	0,382
0,500	0,697	0,636	0,365	0,916	1,871	1,134	0,606	0,975	1,043

7.3. Время $\overline{t}$ (в минутах) ожидания водителем зеленого света на перекрестке представлено случайной выборкой (табл. 7.15). Составить вариационный ряд, определить $\overline{t}$ и S_t^2.

Таблица 7.15

0,000	0,001	0,003	0,012	0,044	0,156	0,534	0,802	0,007	0,822
0,873	0,838	0,170	0,476	0,322	0,648	0,991	0,107	0,726	0,393
0,827	0,419	0,071	0,659	0,309	0,927	0,778	0,327	0,961	0,826
0,308	0,414	0,707	0,515	0,729	0,742	0,884	0,632	0,835	0,318
0,394	0,502	0,471	0,306	0,600	0,846	0,678	0,454	0,623	0,648

7.4. При обследовании диаметров карданных валов автомобиля, выпускаемых заводом, были зафиксированы отклонения от номинала Δd (мкм), которые приведены в табл. 7.16. Найти $\overline{\Delta d}$, $S^2_{\Delta d}$.

Таблица 7.16

–8,760	–1,455	–4,665	–2,250	2,560	–1,645	0,425	0,650	–1,220	–4,410
–6,280	8,550	3,170	0,360	2,450	1,590	–5,435	4,495	5,140	–6,520
7,655	–2,215	7,045	8,650	–1,330	1,745	–1,460	–4,415	–0,280	3,785
–4,790	1,240	–0,475	–7,440	–1,805	–0,295	–2,695	–0,390	1,145	0,970
2,075	–6,910	0,645	–11,805	–5,435	–5,420	1,590	1,835	–4,960	2,645

7.5. Интервал движения поездов метро составляет 2 мин. В табл. 7.17 приведены значения случайной величины X — времени ожидания пассажиром поезда. Составить интервальный вариационный ряд и найти среднее время ожидания.

Таблица 7.17

0,000	0,002	0,007	0,025	0,089	0,312	1,068	1,604	0,014	0,045
1,747	1,677	0,341	0,952	0,645	1,297	1,981	0,214	1,452	0,787
1,654	0,838	0,143	1,317	0,618	1,853	1,555	0,653	1,922	1,653
0,617	0,828	1,413	1,030	1,459	1,483	1,769	1,265	1,669	0,635
0,787	1,004	0,941	0,612	1,200	1,692	1,356	0,908	1,245	1,295

Ответы

7.1. $l = 0{,}052$; $S_l^2 = 0{,}34$.
7.2. $\overline{t} = 1{,}06$; $S_t^2 = 0{,}752$.
7.3. $\overline{t} = 0{,}5$; $S_t^2 = 0{,}0097$.
7.4. $\overline{\Delta d} = -0{,}83$; $S_{\Delta d}^2 = 18{,}6$.
7.5. $\overline{X} = 1{,}022$.

8

ОЦЕНКА ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

При задании случайной величины выборкой ее значений параметры случайной величины заменяются их точечными или интервальными оценками. Методом максимального правдоподобия находятся оценки среднего, дисперсии и доли. Качество оценок проверяется с использованием распределений Стьюдента, хи-квадрат и Фишера, построенных на основе нормального распределения.

8.1. Точечные оценки параметров

Пусть распределение случайной величины X (генеральной совокупности) задается вероятностями $P(x_i, \theta)$ (для дискретной случайной величины) или плотностью вероятности $P(x, \theta)$ (для непрерывной случайной величины), которые зависят от неизвестного параметра θ. Этим параметром может быть, например, параметр λ закона Пуассона или параметры a и σ нормального распределения. На практике о величине параметра σ можно судить по выборке $\{x_1, x_2, ..., x_n\}$ объема n из генеральной совокупности.

Оценкой $\overline{\theta}_n$ параметра θ называется функция от значений выборки $\overline{\theta}_n = \overline{\theta}_n(x_1, x_2, \ldots, x_n)$. Связь оценки $\overline{\theta}_n$ с θ уточняется после введения свойств оценки.

Заметим, что сам параметр θ является некоторым постоянным (неслучайным) числом, которое представляет истинное значение параметра генеральной совокупности. Статистику $\overline{\theta}_n$ можно рассматривать как функцию от случайных величин $X_1, X_2, ..., X_n$, таких, что x_i есть реализация случайной величины X_i:

$$\overline{\theta}_n = \overline{\theta}_n(X_1, X_2, ..., X_n). \qquad (8.1)$$

Таким образом, оценка $\overline{\theta}_n$ является случайной величиной и ее реализации заменяют неизвестные значения параметра θ. Ясно, что оценку $\overline{\theta}_n$ следует выбирать так, чтобы ее значения как можно точнее оценивали неизвестный параметр θ.

Качество оценки характеризуется тремя основными свойствами: *несмещенностью*, *состоятельностью* и *эффективностью*.

Оценка $\bar{\theta}_n$ называется *несмещенной*, если ее математичесое ожидание равно оцениваемому параметру

$$M\bar{\theta}_n = \theta. \tag{8.2}$$

Однако на практике иногда приходится использовать оценки, которые не являются несмещенными.

Например, выборочная дисперсия S^2, определяемая по стандартной формуле (7.10), не является несмещенной оценкой для дисперсии σ^2.

Если это требование не выполняется, то в среднем оценка $\bar{\theta}_n$ будет всегда давать значение θ с некоторым отклонением. Для несмещенных оценок отсутствует систематическая ошибка при оценке параметра θ. Статистика $\bar{\theta}_n$ зависит от объема выборки n, что отмечено в ее обозначении, и при удачном построении оценки естественно ожидать, что при больших n значение статистики приближается к истинному значению параметра θ. Это свойство оценки описывается с помощью понятия предела по вероятности.

Пределом по вероятности случайной величины $\bar{\theta}_n$ называется такое число θ, что для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\bar{\theta}_n - \theta\right| < \varepsilon\right\} = 1. \tag{8.3}$$

Для этого предела применяется запись

$$p \lim_{n\to\infty} \bar{\theta}_n = \theta.$$

Состоятельность оценки означает, что при больших n вероятность сколь угодно малого отклонения $\bar{\theta}_n$ от θ становится сколь угодно близкой к единице. Таким образом, практически достоверно, что состоятельная оценка $\bar{\theta}_n$ для больших n сколь угодно мало отклоняется от самого параметра θ.

Оценкой качества несмещенной оценки является ее дисперсия. Несмещенная оценка $\bar{\theta}_n$ называется *эффективной*, если ее дисперсия

$$D\bar{\theta}_n = M(\bar{\theta}_n - \theta)^2$$

является наименьшей среди дисперсий всех возможных оценок параметра θ, вычисленных по одному и тому же объему выборки n.

При этом оценка $\overline{\theta}_n$ предполагается выбранной из определенного класса функций (например, линейных).

Оценка $\overline{\theta}_n$ называется *точечной*, так как она оценивает одно численное значение параметра θ (точку). Таким образом, точечная оценка параметра — это случайная величина, построенная на основе выборок исходной случайной величины, значения которой служат для оценки величины параметра θ.

Одним из основных методов построения точечных оценок параметров является ***метод максимального правдоподобия***, применение которого будет описано в задаче 8.3. Суть метода заключается в следующем. На основе известной (или предполагаемой) функции плотности распределения вероятностей $p(x, \theta)$ генеральной совокупности X и выборки ее значений $(x_1, x_2, ..., x_n)$ строится функция правдоподобия

$$L(x_1, x_2, ..., x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot ... \cdot p(x_n, \theta).$$

В дискретном случае значения $p(x, \theta)$ задают вероятность принятия случайной величиной X возможного значения x. Значения функции правдоподобия при заданном θ характеризуют вероятность (правдоподобность) появления выборки $(x_1, x_2, ..., x_n)$. Поэтому для нахождения оценки параметра θ следует искать такие значения $\overline{\theta}_n$, которые дают максимум функции правдоподобия. Нахождение $\overline{\theta}_n$ упрощается, если искать максимум функции $\ln L$. *Уравнение правдоподобия* (или система уравнений, когда параметров несколько), которое является необходимым условием максимума, имеет вид

$$\frac{d \ln L}{d\theta} = \frac{1}{L} \frac{dL}{d\theta} = 0.$$

Таким образом, в качестве оценки $\overline{\theta}_n$ выбирается такое решение $\overline{\theta}_n = \overline{\theta}_n(x_1, ..., x_n)$ уравнения правдоподобия, которое задает максимум функции $\ln L$.

Рассмотрим повторную выборку $\{x_1, x_2, ..., x_n\}$ значений генеральной совокупности X. Эта выборка может рассматриваться как реализация системы случайных величин $\{X_1, X_2, ..., X_n\}$, которые будут предполагаться независимыми в совокупности и имеющими такое распределение вероятности, как и сама X. Пусть $MX = a$, $DX = \sigma^2$ — генеральные средняя и дисперсия случайной величи-

ны X. В качестве оценок для a и σ^2 рассмотрим среднюю арифметическую выборки

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8.4}$$

и выборочную дисперсию

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 m_i, \tag{8.5}$$

где k — число вариант выборки; m_i — часто́ты вариант.

Оценка $\bar{x}$ для математического ожидания $MX = a$ является несмещенной, состоятельной и эффективной. Ее дисперсия

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}. \tag{8.6}$$

Оценка S^2 для дисперсии $DX = \sigma^2$ является состоятельной, но смещенной, так как ее математическое ожидание не равно σ^2. Поэтому на практике часто пользуются несмещенной оценкой

$$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 m_i. \tag{8.7}$$

Исправленной выборочной дисперсией называется оценка $\overline{S^2}$ для дисперсии σ^2, определяемая формулой (8.7).

Для бесповторной выборки оценки (8.4), (8.7) также являются несмещенными и состоятельными, а дисперсия оценки $\bar{x}$

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right), \tag{8.8}$$

где N — объем генеральной совокупности.

При $N \to \infty$ бесповторная выборка неотличима от повторной и формула (8.8) переходит в формулу (8.6). В случае, когда дисперсия σ^2 неизвестна, она заменяется значением ее оценки $\overline{S^2}$.

Пусть генеральная совокупность содержит M элементов, обладающих некоторым признаком A.

Генеральной долей признака A называется величина

$$p = \frac{M}{N}. \tag{8.9}$$

Для доли p несмещенной и состоятельной оценкой будет *выборочная доля*

$$\omega = \frac{m}{n}, \tag{8.10}$$

где m — число элементов выборки, обладающих признаком A.

Дисперсия σ_ω^2 выборочной доли в случае бесповторной выборки равна

$$\sigma_\omega^2 = \frac{pq}{n}, \tag{8.11}$$

а в случае повторной выборки

$$\sigma_\omega^2 = \frac{pq}{n}\left(1 - \frac{n}{N}\right), \tag{8.12}$$

где $q = 1 - p$.

Если $n << N$ (n намного меньше N), то повторная выборка практически не отличается от бесповторной и формулы (8.11) и (8.12) дают приблизительно одинаковый результат. Если же $n = N$, то объем выборки равен объему генеральной совокупности и выборочная доля равна генеральной, тогда $\sigma_\omega^2 = 0$.

В случае, когда p неизвестно, его заменяют выборочным значением ω.

Для наглядности и удобства пользования сведем данные о точечных оценках параметров генеральной совокупности в таблицу (табл. 8.1).

Таблица 8.1

Параметр	Оценка	Дисперсия для повторной выборки	Дисперсия для бесповторной выборки
Средняя a	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{\sigma^2}{n}$	$\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)$
Дисперсия σ^2	$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 m_i$	—	—
Доля p	$\omega = \frac{m}{n}$	$\frac{\omega(1-\omega)}{n}$	$\frac{\omega(1-\omega)}{n}\left(1 - \frac{n}{N}\right)$

Пример 8.1. Из 1500 деталей отобрано 250, распределение которых по размеру задано в табл. 8.2. Найти оценки $\bar{x}$, $\overline{S^2}$ для среднего и дисперсии, а также дисперсию $\sigma^2_{\bar{x}}$ оценки $\bar{x}$ для повторного и бесповторного отбора.

Таблица 8.2

Размер детали	7,8–8,0	8,0–8,2	8,2–8,4	8,4–8,6	8,6–8,8	8,8–9,0
Количество деталей	5	20	80	95	40	10

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{250}(7{,}9 \cdot 5 + 8{,}1 \cdot 20 + 8{,}3 \cdot 80 +$$

$$+ 8{,}5 \cdot 95 + 8{,}7 \cdot 40 + 8{,}9 \cdot 10) = 8{,}44,$$

$$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 =$$

$$= \frac{1}{249}[(7{,}9 - 8{,}44)^2 \cdot 5 + (8{,}1 - 8{,}44)^2 \cdot 20 + (8{,}3 - 8{,}44)^2 \cdot 80 +$$

$$+ (8{,}5 - 8{,}44)^2 \cdot 9{,}5 + (8{,}7 - 8{,}44)^2 \cdot 40 + (8{,}9 - 8{,}44)^2 \cdot 10] = 0{,}042.$$

Для повторной выборки

$$\sigma^2_{\bar{x}} = \frac{\overline{S^2}}{n} = 0{,}00016,$$

для бесповторной выборки

$$\sigma^2_{\bar{x}} = \frac{\overline{S^2}}{n}\left(1 - \frac{n}{N}\right) = 0{,}00016 \cdot \left(1 - \frac{250}{1500}\right) = 0{,}00013.$$

Пример 8.2. Выборочно обследовали партию кирпича, поступившего на стройку. Из 100 проб в 12 случаях кирпич оказался бракованным. Найти оценку ω доли бракованного кирпича и σ^2_{ω}.

По формулам (8.10)–(8.12) для оценки доли и ее дисперсии имеем

$$m = 12, \quad n = 100, \quad \omega = \frac{m}{n} = 0{,}12,$$

$$\sigma^2_{\omega} = \frac{\omega(1-\omega)}{n} = \frac{0{,}12 \cdot 0{,}88}{100} = 0{,}001056, \quad \sigma_{\omega} = \sqrt{\sigma^2_{\omega}} = 0{,}031.$$

8.2. Основные статистические распределения

Для решения многих задач статистики используются специальные распределения случайных величин. Наиболее важным в статистических исследованиях является нормальное распределение. Рассмотрим несколько случайных величин, построенных на основе нормального распределения, которые наиболее часто встречаются в математической статистике.

1. ***Распределение χ_n^2 (хи-квадрат с n степенями свободы).***

Если $\xi_i \in N(0; 1)$, $i = 1, 2, ..., n$ — независимые стандартные нормальные случайные величины, то говорят, что случайная величина

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2 \tag{8.13}$$

имеет *распределение хи-квадрат* с n степенями свободы.

При этом $M\chi_n^2 = n$, а графики плотности распределения χ_n^2 при различных n изображены на рис. 8.1.

2. ***Распределение Стьюдента t_n.***

Пусть $\xi_0, \xi_1, ..., \xi_n$ — независимые стандартные нормальные случайные величины. Тогда случайная величина

$$t_n = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} \xi_i^2}} \tag{8.14}$$

имеет *распределение Стьюдента* с n степенями свободы. Для этой величины

$$Mt_n = 0, \quad Dt_n = \frac{n}{n-2}.$$

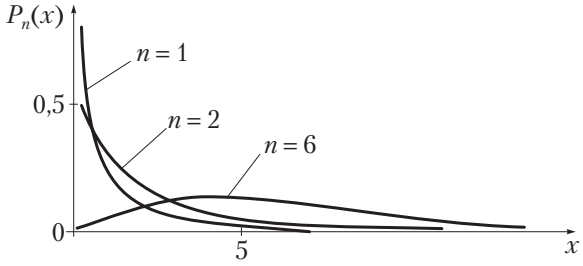


Рис. 8.1. Графики плотности распределения χ_n^2

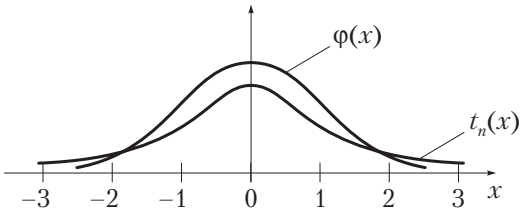


Рис. 8.2. График плотности распределения Стьюдента

График плотности $t_n(x)$ распределения Стьюдента изображен на рис. 8.2, где $\varphi(x)$ — плотность распределения случайной величины $\xi \in N(0; 1)$.

При $n \to \infty$ $t_n(x) \to \varphi(x)$, и уже при $n > 30$ распределение Стьюдента можно приближенно заменить на стандартное нормальное распределение.

3. ***Распределение Фишера.***

Пусть $\xi_i \in N(0; 1)$, $i = 1, 2, ..., m$, $\eta_j \in N(0; 1)$, $j = 1, 2, ..., n$ — независимые случайные величины. Тогда случайная величина

$$F_{m,n} = \frac{\frac{1}{m}\sum_{i=1}^{m}\xi_i^2}{\frac{1}{n}\sum_{j=1}^{n}\eta_j^2} \tag{8.15}$$

распределена по закону Фишера со степенями свободы m, n. Для больших m, n $F_{m,n}$ стремится к нормальному распределению (рис. 8.3). Кроме этого, $F_{1,n} = t_n^2$ для всех n.

Распределения случайных величин χ_n^2, t_n, $F_{m,n}$ имеются в статистических таблицах.

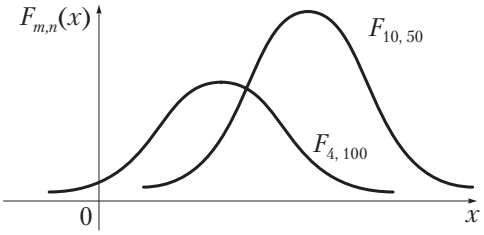


Рис. 8.3. График плотности распределения Фишера

8.3. Интервальные оценки параметров

Точечная оценка $\overline{\theta}_n$ параметра θ дает лишь некоторое приближенное значение θ. Чтобы получить представление о точности и надежности оценки, используют интервальную оценку параметра.

Интервальной оценкой параметра θ называется интервал (α, β), концы которого являются случайными величинами и который с заданной вероятностью γ накрывает неизвестное значение параметра θ (рис. 8.4).

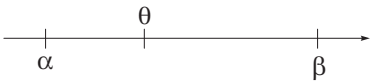


Рис 8.4. Доверительный интервал

Такой интервал (α, β) называется *доверительным интервалом*, вероятность γ — *доверительной вероятностью*, а $\alpha = 1 - \gamma$ — *уровнем надежности*. Обычно доверительный интервал выбирается симметричным относительно оценки $\overline{\theta}_n$, тогда он определяется формулой

$$P\left\{\left|\overline{\theta}_n - \theta\right| < \Delta\right\} = \gamma \qquad (8.16)$$

и имеет вид

$$(\overline{\theta}_n - \Delta,\ \overline{\theta}_n + \Delta),$$

т.е. неравенства $\overline{\theta}_n - \Delta < \theta < \overline{\theta}_n + \Delta$ выполняются с вероятностью γ.

Наибольшее отклонение Δ выборочного значения оценки $\overline{\theta}_n$ параметра от истинного значения параметра θ называется *предельной ошибкой выборки* при заданном объеме выборки и заданной доверительной вероятности.

8.4. Доверительные интервалы для генеральной средней и генеральной доли выборки

Пусть $\{x_1, x_2, ..., x_n\}$ — выборка из генеральной совокупности объема n, $\overline{x}$ — выборочное среднее, $\overline{S^2}$ — выборочная дисперсия, $\overline{S} = \sqrt{\overline{S^2}}$ — выборочное среднее квадратичное отклонение, $\omega = \frac{m}{n}$ — выборочная доля признака.

Доверительный интервал с доверительной вероятностью γ для генеральной средней a имеет вид

$$\bar{x} - \Delta < a < \bar{x} + \Delta, \tag{8.17}$$

где Δ — предельная ошибка выборки, зависящая от γ.

При $n > 30$ для повторной выборки

$$\Delta = t\frac{\bar{S}}{\sqrt{n}}, \tag{8.18}$$

для бесповторной выборки —

$$\Delta = t\frac{\bar{S}}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}. \tag{8.19}$$

В этих формулах t определяется из условия

$$\Phi(t) = \gamma, \tag{8.20}$$

где $\Phi(t)$ — функция Лапласа.

Если $n \le 30$, то доверительный интервал для среднего строится только для нормальной генеральной совокупности.

Для повторной выборки

$$\Delta = t_{n-1}\frac{\bar{S}}{\sqrt{n}}. \tag{8.21}$$

Здесь t_{n-1} определяется из условия

$$P\left\{|\xi| < t_{n-1}\right\} = \gamma, \quad \left(P\left\{|\xi| > t_{n-1}\right\} = \alpha = 1 - \gamma\right), \tag{8.22}$$

где случайная величина ξ имеет распределение Стьюдента с $n - 1$ степенями свободы, а t_{n-1} находится по таблицам распределения Стьюдента.

Для бесповторной выборки

$$\Delta = t_{n-1}\frac{\bar{S}}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}, \tag{8.23}$$

где t_{n-1} находится по таблицам распределения Стьюдента.

Доверительный интервал для генеральной доли p равен (ω − Δ, ω + Δ), т.е.

$$\omega - \Delta < p < \omega + \Delta, \tag{8.24}$$

где при $n > 30$ для повторной выборки

$$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}, \tag{8.25}$$

для бесповторной выборки

$$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}\sqrt{1-\frac{n}{N}}, \tag{8.26}$$

где t определяется равенством (8.20).

При $n \le 30$ рассматриваются только выборки из нормальной совокупности. Для повторной выборки Δ определяется по формуле (8.25), для бесповторной — по формуле (8.26).

Для удобства пользования приведенные выше формулы сведем в табл. 8.3.

Таблица 8.3

Параметр	Оценка	Предельная ошибка выборки			
		Повторная выборка		Бесповторная выборка	
		$n > 30$	$n \le 30$	$n > 30$	$n \le 30$
a	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\Delta = t\frac{\bar{S}}{\sqrt{n}}$	$\Delta = t_{n-1}\frac{\bar{S}}{\sqrt{n}}$	$\Delta = t\frac{\bar{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$	$\Delta = \frac{t_{n-1}\bar{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$
p	$\omega = \frac{m}{n}$	$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}$		$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}\sqrt{1-\frac{n}{N}}$	

Пример 8.3. Из партии в 5000 электрических лампочек было отобрано 300 по схеме бесповторной выборки. Средняя продолжительность горения лампочек в выборке оказалась равной 1450 ч, а дисперсия — 4000. Найти доверительный интервал для среднего срока a горения лампочки с вероятностью 0,9996.

Для $\gamma = 0{,}9996$ по таблицам находим $t = 3{,}57$. При $n > 30$ для бесповторной выборки

$$\Delta = \frac{3{,}57 \cdot \sqrt{40\,000}}{\sqrt{300}}\sqrt{1-\frac{300}{5000}} = 12{,}639.$$

Тогда искомый доверительный интервал имеет вид $1437 < a < 1463$.

Пример 8.4. В партии, содержащей 5000 изделий, проверено 400. Среди них оказалось 300 изделий высшего сорта. Найти с доверительной вероятностью 0,95 доверительный интервал для доли изделий высшего сорта в случаях повторной и бесповторной выборок.

Для $\gamma = 0{,}95$ определим t из условия $\Phi(t) = \gamma$. Получим $t = 1{,}96$. При $m = 300$, $n = 400$, $N = 5000$ найдем оценку для доли p,

$$\omega = \frac{m}{n} = 0{,}75.$$

Для случая повторной выборки предельная ошибка

$$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}} = 1{,}96\sqrt{\frac{0{,}75 \cdot 0{,}25}{400}} = 0{,}0424.$$

Тогда доверительный интервал

$$0{,}75 - 0{,}0424 < p < 0{,}75 + 0{,}0424 \Rightarrow 0{,}7076 < p < 0{,}7924.$$

Для бесповторной выборки

$$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}\sqrt{1-\frac{n}{N}} = 0{,}0424\sqrt{1-\frac{400}{5000}} = 0{,}0407$$

и доверительный интервал имеет вид $0{,}7093 < p < 0{,}7907$.

8.5. Объем выборки

При решении статистических задач часто требуется определить необходимый объем выборки для достижения заданной надежности доверительного интервала. Как следует из формул (8.18) или (8.23), объем выборки n определяется доверительной вероятностью γ и предельной ошибкой Δ. Объемы выборок для различных случаев приведены в табл. 8.4.

Таблица 8.4

Параметр	Объем выборки	
	Повторная выборка	Бесповторная выборка
a	$n = \dfrac{t^2\overline{S^2}}{\Delta^2}$	$n\dfrac{Nt\overline{S^2}}{N\Delta^2 + t^2\overline{S^2}}$
p	$n = \dfrac{\omega(1-\omega)t^2}{\Delta^2}$	$n = \dfrac{Nt^2\omega(1-\omega)}{N\Delta^2 + t^2\omega(1-\omega)}$

Пример 8.5. Найти объемы повторной и бесповторной выборок из 10 000 банок консервов для определения доли банок, не соответствующих стандарту. Предполагается, что предельная ошибка выборки не превосходит 0,05 с доверительной вероятностью 0,9995.

По условию задачи $N = 10\,000$, $\Delta = 0{,}05$, $\gamma = 0{,}9995$. Из условия $\Phi(t) = 0{,}0995$ находим $t = 3{,}5$.

Для повторной выборки получим

$$n = \frac{\omega(1-\omega)t^2}{\Delta^2} = \frac{3{,}5^2}{0{,}05^2}\omega(1-\omega) - 4900\omega(1-\omega).$$

Так как ω неизвестно, то выберем его таким, чтобы значение выражения $\omega(1-\omega)$ было максимальным, т.е. положим $\omega = 0{,}5$. Тогда завышенное значение n будет равно $n = 4900 \cdot 0{,}5 \cdot 0{,}5 = 1225$.

Для бесповторной выборки получим

$$n = \frac{Nt^2\omega(1-\omega)}{N\Delta^2 + t^2\omega(1-\omega)} = \frac{Nt^2}{\dfrac{N\Delta^2}{\omega(1-\omega)} + t^2}.$$

Очевидно, и в этом случае наибольшее значение выражения $\omega(1-\omega)$ соответствует максимальному n. Положим $\omega = 0{,}5$, тогда

$$n = \frac{10\,000 \cdot 3{,}5^2 \cdot 0{,}5 \cdot 0{,}5}{10\,000 \cdot 0{,}05^2 + 3{,}5^2 \cdot 0{,}5 \cdot 0{,}5} \cong 1091.$$

Замечание. Если $N \to \infty$ (бесконечный объем генеральной совокупности), то бесповторная выборка не отличается от повторной и формулы для бесповторной выборки в табл. 8.2, 8.3 переходят в формулы для повторной выборки.

8.6. Доверительный интервал для дисперсии

Если $\{x_1, x_2, ..., x_n\}$ — выборка из нормальной совокупности $N(a; \sigma)$, где a и σ неизвестны, то статистика

$$Z = \frac{n\overline{S^2}}{\sigma^2} \tag{8.27}$$

имеет распределение χ^2_{n-1}. Тогда доверительный интервал для σ^2 можно найти из соотношения

$$P\left\{\frac{n\overline{S^2}}{z_2} < \sigma^2 < \frac{n\overline{S^2}}{z_1}\right\} = \gamma,$$

где z_1 и z_2 определяются из условия $P\left\{z_1 < \chi^2_{n-1} < z_2\right\} = \gamma$ (рис. 8.5).

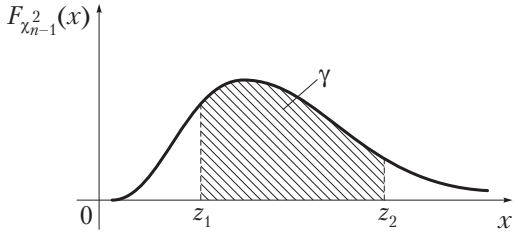


Рис. 8.5. Выбор параметров для доверительного интервала дисперсии

Обычно z_1 и z_2 выбираются таким образом, чтобы

$$P\left\{\chi^2_{n-1} < z_1\right\} = P\left\{\chi^2_{n-1} > z_2\right\} = \frac{1-\gamma}{2}.$$

Тогда

$$P\left\{\chi^2_{n-1} > z_1\right\} = \frac{1+\gamma}{2}, \quad P\left\{\chi^2_{n-1} > z_2\right\} = \frac{1-\gamma}{2} \qquad (8.28)$$

— условия для выбора величин z_1 и z_2, которые можно найти из таблиц распределения χ^2_{n-1}.

Задачи с решениями

Задача 8.1. В табл. 8.5 приведены результаты выборочного обследования заработной платы X 100 рабочих предприятия, на котором занято 1000 рабочих.

Таблица 8.5

Заработная плата, усл. ед.	100–120	120–140	140–160	160–180	180–200
Число рабочих	17	40	32	8	3

Найти выборочное среднее $\bar{x}$, исправленную выборочную дисперсию S_x^2 и дисперсию $\sigma^2_{\bar{x}}$ оценки $\bar{x}$.

Решение. $n = 100$, $N = 1000$.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{100}(110 \cdot 17 + 130 \cdot 40 + 150 \cdot 32 + 170 \cdot 8 + 190 \cdot 3) =$$

$$= \frac{1}{100}(1870 + 5200 + 4800 + 1360 + 570) = \frac{13\,800}{100} = 138.$$

$$\overline{S_x^2} = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{99}\Big[(110-138)^2 + (130-138)^2 +$$

$$+ (150-138)^2 + (170-138)^2 + (190-138)\Big]^2 =$$

$$= \frac{1}{99}(784 + 64 + 144 + 1024 + 2704) = \frac{4720}{99} = 47{,}6767, \quad \overline{S_x} = 6{,}9.$$

$$\sigma_{\bar{x}}^2 = \frac{\overline{S^2}}{n}\left(1 - \frac{n}{N}\right) = 0{,}4768 \cdot 0{,}9 = 0{,}429.$$

Задача 8.2. С целью определения средней суммы Q вкладов в банке, имеющем 2200 вкладчиков, проведено выборочное обследование (бесповторный отбор), результаты которого приведены в табл. 8.6.

Таблица 8.6

Сумма вклада, усл. ед.	10–30	30–50	50–70	70–90	90–110	110–130
Число вкладчиков	1	3	10	30	60	7

Найти с вероятностью 0,96 доверительные границы для Q.

Решение. $n = 111$, $N = 2200$, $\gamma = 0{,}96$, $n > 30$, $\Phi(t) = \gamma$; $t = 0{,}6629$.

$\bar{x}$ — точечная оценка для средней генеральной совокупности. Найдем выборочное среднее

$$\bar{x} = \frac{1}{111}(20 \cdot 1 + 40 \cdot 3 + 60 \cdot 10 + 80 \cdot 30 + 100 \cdot 60 + 120 \cdot 7) =$$

$$= \frac{9980}{111} = 89{,}91.$$

Таким образом, $\bar{Q} = 89{,}91$ — средний размер вклада.

$$\overline{S^2} = \frac{1}{110}\Big[(120-89{,}9)^2 + (40-89{,}9)^2 + (60-89{,}9)^2 +$$

$$+ (80-89{,}9)^2 + (100-89{,}9)^2 + (120-89{,}9)^2\Big] =$$

$$= \frac{9376}{110} = 85{,}236, \quad \bar{S} = 9{,}23.$$

$$\Delta = t\frac{\bar{S}}{\sqrt{n}}, \quad \Delta = 0{,}6629 \cdot \frac{9{,}23}{\sqrt{110}} = 0{,}588$$

для повторной выборки,

$$\Delta = 0{,}6629 \cdot \frac{9{,}23}{\sqrt{110}} \cdot \left(1 - \frac{110}{2200}\right) = 0{,}5586$$

для бесповторной выборки.

$$\bar{x} - \Delta < a < \bar{x} + \Delta,$$

$$89{,}9 - 0{,}5586 < Q < 89{,}9 + 0{,}5586,$$

$$89{,}34 < Q < 90{,}458.$$

Задача 8.3. Пусть $\{x_1, x_2, ..., x_n\}$ — выборка из нормального распределения $N(a; \sigma)$. Определить параметры a и σ методом максимального правдоподобия.

Решение. Составим функцию правдоподобия

$$L = \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right\}.$$

Тогда

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2.$$

Рассмотрим уравнения для определения максимума функции $\ln L$.

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0,$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - a)^2 = 0.$$

Из первого уравнения находим

$$\sum_{i=1}^{n} x_i - na = 0, \quad a = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

Из второго уравнения —

$$\sigma^2 n = \sum_{i=1}^{n}(x_1 - \bar{x})^2, \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2.$$

Задача 8.4. Для определения среднего процентного содержания белка в зернах пшеницы было отобрано 626 зерен, обследование которых показало, что выборочное среднее $\bar{x} = 16{,}8$, а выборочная дисперсия $\overline{S^2} = 4$. Чему равна с вероятностью 0,988 предельная ошибка выборки?

Решение. Считаем, что объем генеральной совокупности бесконечен, тогда выборку можно считать повторной. Так как $\gamma = 0{,}988$, то по таблицам функции $\Phi(x)$ определим t из условия $\Phi(t) = 0{,}988$, $t = 2{,}5$. Тогда предельная ошибка выборки

$$\Delta = t\frac{\bar{S}}{\sqrt{n}} = 2{,}5 \cdot \frac{2}{\sqrt{625}} = 0{,}2.$$

Задача 8.5. Выборочное среднее квадратичное отклонение десяти измерений некоторой величины равно 10 см. Найти с надежностью $\alpha = 0{,}6$ предельную ошибку выборки.

Решение. Здесь $n = 10 < 30$ и выборка повторная. По таблицам распределения Стьюдента находим для $\alpha = 0{,}6$ $t_9 = 0{,}543$. Тогда

$$\Delta = t_{n-1}\frac{\bar{S}}{\sqrt{n}} = 0{,}543\frac{10}{\sqrt{10}} = 1{,}72.$$

Задача 8.6. Среди стандартных изделий одной фабрики в среднем 15 % относится ко второму сорту. С какой вероятностью можно утверждать, что процент p изделий второго сорта среди 1000 стандартных изделий данной фабрики отличается от 15 % не более чем на 2 %?

Решение. Здесь $n = 1000$, $\omega = 0{,}15$, $\Delta = 0{,}02$. Вычислим

$$t = \Delta\sqrt{\frac{n}{\omega(1-\omega)}} = 0{,}02\sqrt{\frac{1000}{0{,}15 \cdot 0{,}85}} = 1{,}77.$$

Тогда

$$P\{|p - 0{,}15| \le 0{,}02\} = \Phi(1{,}77) = 0{,}9233.$$

Задача 8.7. Службой контроля проверен расход энергии в течение месяца в 10 квартирах 70-квартирного дома, в результате чего были получены значения (кВт · ч): 125, 78, 102, 140, 90, 45, 50, 125, 115, 112.

Определить с вероятностью 0,95 доверительный интервал для оценки среднего расхода электроэнергии в доме.

Решение. Считаем выборку бесповторной, тогда (при $n = 10 < 30$) предельная ошибка выборки равна

$$\Delta = t_{n-1}\frac{\overline{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}.$$

Здесь $n = 10$, $N = 70$. При $\gamma = 0{,}95$ получим $t = 2{,}26$. Находим

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 98{,}2 \quad \text{и} \quad \overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x)^2 = 1033{,}29;$$

$$\overline{S} = 32{,}145.$$

Тогда

$$\Delta = 2{,}26\frac{32{,}145}{\sqrt{10}}\sqrt{1-\frac{10}{70}} = 21{,}27.$$

Получим доверительный интервал $\bar{x} - \Delta < a < \bar{x} + \Delta$, следовательно, $76{,}93 < a < 119{,}47$.

Задачи для самостоятельного решения

8.1. В выборке объемом 500 единиц, произведенной для определения процента всхожести зерна, установлена частость доброкачественных зерен

$$\frac{m}{n} = 0{,}94.$$

Найти вероятность процента всхожести, если допустимая погрешность в его определении равна 2 %.

8.2. Из 400 коммерческих фирм налоговой инспекцией было проверено 100, данные о сокрытии налогов среди которых приведены в табл. 8.7.

Таблица 8.7

Сумма, тыс. руб.	40–42	42–44	44–46	46–48	48–50
Количество фирм	7	24	38	19	12

Найти вероятность того, что сокрытие налогов отличается от среднего значения не более чем на 0,4 тыс. руб.

8.3. По схеме бесповторной выборки произведено измерение выработки на земляных работах 145 рабочих из 500. В результате средняя выработка на одного рабочего оказалась равной $\bar{x} = 4{,}95$ м, а среднее квадратичное отклонение — $\bar{S} = 1{,}95$. Найти доверительные границы для генерального среднего a при доверительной вероятности $\gamma = 0{,}954$.

8.4. Определить численность выборки при обследовании остатков на расчетных счетах у клиентов банка, чтобы с вероятностью 0,683 предельная ошибка равнялась 5 усл. ед., если $\sigma = 120$ усл. ед.

8.5. При просмотре 1000 волокон из партии льна обнаружено 120 недозрелых. Сколько надо просмотреть волокон льна из этой партии, чтобы с вероятностью $\gamma = 0{,}997$ точность определения доли недозрелых волокон находилась в пределах 5 %?

8.6. Сколько лиц в возрасте от 19 до 24 лет надо опросить, чтобы установить средний процент студентов с точностью до 0,5 %?

8.7. Найти объем n_1 повторной и n_2 бесповторной выборок из 500 автомобилей для определения среднего пробега без ремонта, чтобы с вероятностью 0,99 предельная ошибка выборки не превосходила 25 тыс. км. Принять генеральное среднее квадратичное отклонение равным 150 тыс. км.

8.8. Для оценки уровня безработицы в городе были отобраны 100 человек рабочих специальностей. Из них 6 человек оказались безработными. Оценить с вероятностью 0,95 процент p безработных рабочих в городе.

8.9. При обследовании 100 малых предприятий оказалось, что 94 из них относится к негосударственной форме собственности. Определить с вероятностью 0,95 границы для доли p негосударственных малых предприятий в городе.

8.10. Для оценки стоимости древесины с целью последующей продажи была сделана выборка из 100 деревьев, подлежащих вырубке. В результате измерения диаметров d деревьев были получены среднее значение $\bar{d} = 50$ см и выборочная дисперсия $\sigma_{\bar{d}}^2 = 3{,}5$. Определить на уровне 0,95 доверительный интервал для диаметра d деревьев, подлежащих вырубке.

8.11. Из 25 студентов группы, проходивших флюорографию, у одного были обнаружены признаки легочных заболеваний. Найти с вероятностью 0,99 доверительный интервал для P — процента студентов, не страдающих легочными заболеваниями, если всего в университете учатся около 7000 студентов.

8.12. При выборке в 250 человек оказалось, что 105 из них поддерживают определенного кандидата в парламент. Сколько человек следует опросить, чтобы с вероятностью 0,99 можно было бы утверждать, что доля избирателей, поддерживающих данного кандидата, отличается от истинной доли не более чем на 0,05?

8.13. Из 10 орехов, взятых для проверки, 2 оказались пустыми. Найти 95%-ный доверительный интервал для доли пустых орехов в продаваемой партии.

Ответы

8.1. $p = 0{,}897$.
8.2. $p = 0{,}9997$.
8.3. $4{,}75 < a < 5{,}15$.
8.4. $n = 576$.
8.5. $n = 373$.
8.6. $n = 1089$.
8.7. $n_1 = 240$, $n_2 = 162$.
8.8. $1{,}3\ \% < p < 10{,}7\ \%$.
8.9. $0{,}8935 < p < 0{,}9865$.
8.10. $46{,}33 < d < 53{,}67$.
8.11. $92{,}7 < P < 99{,}3$.
8.12. $n = 32$.
8.13. $0 < p < 0{,}448$.

9

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Статистические методы позволяют находить лишь оценки (параметров, возможных зависимостей между признаками, видов распределения и др.). Поэтому возникает необходимость выдвинуть и проверить на заданном уровне доверия те или иные предположения о свойствах статистических совокупностей. Такая задача решается проверкой статистических гипотез.

9.1. Основные понятия

Статистической гипотезой называется любое предположение о свойствах распределения вероятностей, лежащего в основе наблюдаемых явлений.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов:

1. Гипотезы о виде закона распределения исследуемой случайной величины.

2. Гипотезы о числовых значениях параметров случайной величины.

3. Гипотезы об общем виде модели, описывающей статистическую зависимость между признаками.

4. Гипотезы о принадлежности некоторого признака к тому или иному классу.

Гипотезы обозначаются большими латинскими буквами $H_0, H_1, H_2, ..., H_k$. Как правило, через H_0 обозначается *основная* гипотеза, в справедливости которой нам важно убедиться (например, H_0 — благополучная посадка самолета, выздоровление больного, успех в коммерческой деятельности и т.д.).

Гипотезе H_0 противопоставляются гипотезы $H_1, H_2, ..., H_k$, которые называются *альтернативными*. В дальнейшем мы будем рассматривать только две гипотезы: H_0 — основную и H_1 — альтернативную (конкурирующую). Принятие гипотезы H_0 или ее аль-

тернативы основано на исследовании статистических данных, т.е. выборки $\{x_1, x_2, ..., x_n\}$ из некоторой генеральной совокупности.

Гипотеза H называется *простой*, если она полностью определяет теоретическое распределение случайной величины по имеющейся выборке ее значений. В противном случае гипотеза называется *сложной*. Обычно основная гипотеза H_0 является простой, в то время как ее альтернатива может быть простой или сложной.

Пусть имеется выборка значений случайной величины $\{x_1, x_2, ..., x_n\}$ из $N(a; \sigma)$ и σ известна. Тогда, если основная гипотеза H_0: $a = a_1$, альтернативная гипотеза H_1: $a = a_2$, где a_1 и a_2 — некоторые числа, то и H_0, и H_1 — простые гипотезы. Если H_0: $a = a_0$, H_1: $a \neq a_0$, то H_0 — простая гипотеза, H_1 — сложная.

Правило К, по которому принимается или отвергается гипотеза, называется критерием. Вероятность отвергнуть гипотезу H_0, если она верна, и принять гипотезу H_1 называется ***ошибкой 1-го рода***, или ***уровнем значимости***, и обозначается α:

$$P\{H_1 \mid H_0\} = \alpha. \tag{9.1}$$

Величина $1 - \alpha$, равная вероятности принять верную гипотезу H_0, называется ***уровнем доверия***, или ***доверительной вероятностью***:

$$P\{H_0 \mid H_0\} = 1 - \alpha = \gamma. \tag{9.2}$$

Вероятность принять основную гипотезу, если она неверна, называется *ошибкой 2-го рода* и обозначается β:

$$P\{H_0 \mid H_1\} = \beta. \tag{9.3}$$

Вероятность принять гипотезу H_1, если она верна, называется ***мощностью критерия***:

$$P\{H_1 \mid H_1\} = 1 - \beta. \tag{9.4}$$

Процедура построения критерия осуществляется следующим образом. По имеющимся данным выборки составляется статистика $\overline{\theta}_n$ так, чтобы в случае, если гипотеза H_0 верна, распределение случайной величины $\overline{\theta}_n = \overline{\theta}_n(X_1,...,X_n)$ было бы известным. Построение критерия, в зависимости от вида гипотезы H_0, заключается в выборе таких значений θ_1 и θ_2, что если $\theta_1 < \overline{\theta}_n < \theta_2$, то гипотеза H_0

принимается. При этом возможно, что $\theta_1 = -\infty$ или $\theta_2 = +\infty$. Значения θ_1 и θ_2 называются *критическими*, а область

$$T_д = \{\theta : \theta_1 < \theta < \theta_2\} \tag{9.5}$$

называется ***областью допустимых значений***. Так что, если $\overline{\theta}_n \in T_д$, то гипотеза H_0 принимается, если $\overline{\theta}_n \notin T_д$, то отвергается, т.е. принимается альтернатива H_1. При этом область

$$T_{кр} = \{\theta : \theta \le \theta_1, \theta \ge \theta_2\} \tag{9.6}$$

называется *критической областью*.

В зависимости от вида альтернативной гипотезы H_1 выбирают *правостороннюю*, *левостороннюю* или *двустороннюю область*. Для правосторонней критической области

$$P\{\overline{\theta}_n \ge \theta_2\} = \alpha,$$

для левосторонней

$$P\{\overline{\theta}_n \le \theta_1\} = \alpha,$$

для двусторонней

$$P\{\overline{\theta}_n \le \theta_1\} = P\{\overline{\theta}_n \ge \theta_2\} = \frac{\alpha}{2}.$$

Предположим, что если верна гипотеза H_0, то статистика $\overline{\theta}_n$ имеет распределение с плотностью $p = p_1(\theta)$, а если верна гипотеза H_1, то $p = p_2(\theta)$. Тогда описанные выше построения можно изобразить графически. На рис. 9.1 $\theta_1 = -\infty$ или $\theta_2 = \theta_{кр}$. Критической право-

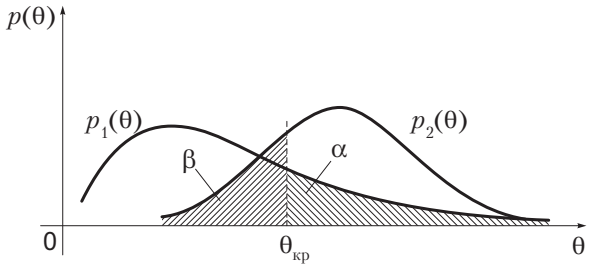


Рис. 9.1. Схема проверки гипотез

сторонней областью $T_{кр}$ является множество $\{\theta: \theta \geq \theta_{кр}\}$. Гипотеза H_0 принимается, если $\bar{\theta}_n \leq \theta_{кр}$. Площади заштрихованных областей представляют собой вероятности ошибок 1-го и 2-го рода α и β соответственно.

Схемы проверки гипотез удобно представить в виде таблицы (табл. 9.1).

Таблица 9.1

Гипотеза	Принята	Отклонена
Верна	Доверительная вероятность $P\{H_0 \mid H_0\} = 1 - \alpha = \gamma$	Ошибка 1-го рода (уровень значимости) $P\{H_1 \mid H_0\} = \alpha$
Неверна	Ошибка 2-го рода $P\{H_0 \mid H_1\} = \beta$	Мощность критерия $P\{H_1 \mid H_1\} = 1 - \beta$

Обычно при определении критического значения $\theta_{кр}$ задается доверительная вероятность $\gamma = 1 - \alpha$ или уровень значимости α (например, $\gamma = 0{,}95$, $\alpha = 0{,}05$) и $\theta_{кр}$ находится из условия

$$P\{\theta_1 < \theta < \theta_2\} = \gamma = 1 - \alpha. \qquad (9.7)$$

Очевидно, что выбор доверительной области не является однозначным, так как при заданном γ и известной плотности вероятностей статистики $\bar{\theta}_n$ существует много областей допустимых значений, удовлетворяющих условию (9.7).

Таким образом, для проверки статистической гипотезы H_0 сначала по выборочным данным строится статистика $\bar{\theta}_n$, имеющая известное распределение. Затем задается доверительная вероятность γ критерия К для принятия гипотезы H_0 и, наконец, делаются выводы о справедливости гипотезы H_0.

Способ проверки статистической гипотезы H_0 не дает точного вывода о ее верности или неверности, так как принятие гипотезы всегда происходит на некотором (субъективно) принятом уровне надежности и основывается на значениях конечной выборки. Принятие гипотезы означает, что на принятом уровне надежности гипотеза H_0 не противоречит имеющимся выборочным данным.

9.2. Гипотезы о значениях числовых характеристик

Гипотезы о равенстве среднего значения a и дисперсии σ^2 определенным числам a_0 и σ_0^2 являются простыми гипотезами, когда среднее значение и дисперсия полностью определяют теоретическое распределение (например, в случае нормального распределения). Такие гипотезы возникают, например, при проверке качества функционирования измерительных устройств. Если a_0 — номинальное значение измеряемого параметра и $a \neq a_0$, то это означает, что прибор дает систематическую ошибку. Точность прибора определяется значением σ_0, и если $\sigma > \sigma_0$, то это означает, что качество прибора не отвечает стандартным требованиям.

9.2.1. Гипотеза о численной величине среднего значения

Пусть случайная величина $X \in N(a; \sigma)$ и имеется выборка $\{x_1, x_2, ..., x_n\}$ ее значений. Рассмотрим гипотезу H_0: $a = a_0$, где a_0 — некоторое число, и ее альтернативу H_1: $a \neq a_0$. Проверим гипотезу H_0 на уровне доверия γ. Возможны два случая:

1. Дисперсия σ^2 известна.

Так как

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \in N\left(a, \frac{\sigma}{\sqrt{n}}\right),$$

то случайная величина (статистика)

$$t = \frac{(\bar{x} - a)\sqrt{n}}{\sigma} \tag{9.8}$$

будет иметь стандартное нормальное распределение, если H_0 верна (т.е. $a = a_0$). При заданном уровне доверия γ по таблицам функции $\Phi(t)$ находим такое $t_{кр}$, чтобы

$$P\{|\xi_0| < t_{кр}\} = \gamma, \tag{9.9}$$

где $\xi_0 \in N(0; 1)$.

Будем принимать гипотезу H_0, если $|t| < t_{кр}$. Тогда критическими значениями будут $t_1 = -t_{кр}$, $t_2 = t_{кр}$ (рис. 9.2).

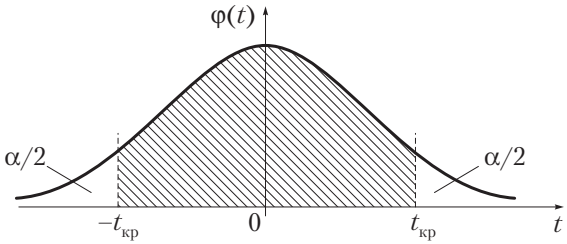


Рис. 9.2. Выбор критического значения

Критической областью будет двусторонняя область

$$T_{\text{кр}} = \left\{t: |t| \geq t_{\text{кр}}\right\},$$

а областью допустимых значений —

$$T_{\text{д}} = \left\{t: |t| < t_{\text{кр}}\right\}.$$

Пример 9.1. По результатам $n = 9$ замеров установлено, что выборочное среднее время (в секундах) изготовления детали $\bar{x} = 48$. Предполагая, что время изготовления — нормально распределенная случайная величина с дисперсией $\sigma^2 = 9$, рассмотреть на уровне 0,95 гипотезу H_0: $a = 49$ против конкурирующей гипотезы H_1: $a \neq 49$.

Здесь рассматривается простая гипотеза $a = 49$ против сложной $a \neq 49$. На уровне $\gamma = 0{,}95$ находим по таблицам нормального распределения $t_{\text{кр}} = 1{,}96$.

Гипотеза принимается при

$$|t| = \left|\frac{(\bar{x} - a)\sqrt{n}}{\sigma}\right| < t_{\text{кр}}.$$

Так как

$$t = \frac{(48 - 49)\sqrt{9}}{3} = -1,$$

то $|t| < t_{\text{кр}}$ и гипотеза принимается.

2. Дисперсия σ^2 неизвестна.

В этом случае в качестве статистики t берут величину

$$t = \frac{(\bar{x} - a)\sqrt{n}}{\bar{S}}, \tag{9.10}$$

где

$$\overline{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Статистика t имеет распределение Стьюдента с $n - 1$ степенями свободы. Для заданного уровня доверия γ по таблицам распределения Стьюдента определяется критическое значение из условия

$$P\left\{|t_{n-1}| < t_{\text{кр}}\right\} = \gamma. \tag{9.11}$$

Гипотеза H_0 принимается, если $|t| < t_{n-1,\text{кр}}$.

9.2.2. Гипотеза о числовом значении дисперсии

Пусть случайная величина $X \in N(a; \sigma)$ и имеется выборка $\{x_1, x_2, ..., x_n\}$ ее значений. Рассмотрим гипотезу H_0: $\sigma = \sigma_0$ против конкурирующей гипотезы H_1: $\sigma \neq \sigma_0$, при этом a может быть произвольным. В этом случае в качестве статистики выбираем величину

$$t = \frac{(n-1)\overline{S^2}}{\sigma_0^2}. \tag{9.12}$$

Известно, что статистика t имеет распределение χ^2_{n-1}. По таблицам распределения для заданного уровня доверия γ выбираем такие U и V, чтобы

$$P\left\{\chi^2_{n-1} \le U\right\} = P\left\{\chi^2_{n-1} \ge V\right\} = \frac{\gamma}{2}, \tag{9.13}$$

и гипотезу H_0 принимаем, если

$$U < t < V. \tag{9.14}$$

Пример 9.2. Точность работы станка-автомата проверяется по дисперсии σ^2 контролируемого размера изделий, которая не должна превышать 0,15. По данным из 25 отобранных изделий вычислена оценка дисперсии $\overline{S^2} = 0{,}25$. При уровне значимости $\alpha = 0{,}1$ выяснить, обеспечивает ли станок требуемую точность.

Предположим, что размер изделия — нормальная случайная величина. Рассмотрим гипотезу H_0: $\sigma^2 = 0{,}15$ и альтернативу H_1: $\sigma^2 > 0{,}15$. Найдем по таблицам χ^2 число V из условия $P\left\{\chi^2_{n-1} \geq V\right\} = 0{,}1$ при $n = 25$. Получим $V = 15{,}5$. Считаем $U = 0$, рассматривая при этом правостороннюю критическую область (рис. 9.3).

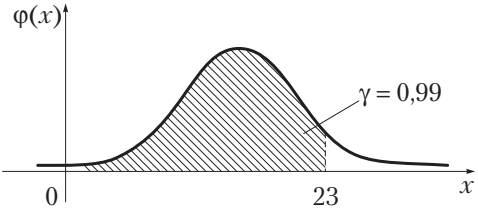


Рис. 9.3. Выбор критического значения для правосторонней области

Для проверки гипотезы вычислим по формуле (9.12) статистику

$$t = \frac{(n-1)\overline{S^2}}{\sigma_0^2} = \frac{24 \cdot 0{,}25}{0{,}15} = 40.$$

Так как $t = 40 > t_{кр} = V = 15{,}5$, то гипотеза о достижении станком требуемой точности отклоняется.

9.2.3. Гипотеза о числовом значении доли признака

Иногда возникает необходимость сравнения числового значения доли p признака A в некоторой генеральной совокупности X с заданным числом p_0. Само число p_0 определяется на основе дополнительной информации, полученной в предыдущих исследованиях. Для решения этой задачи составляется статистика

$$t = \frac{(\omega - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}, \qquad (9.15)$$

которая при больших n ($n > 30$) распределена по стандартному нормальному закону $N(0; 1)$. Гипотеза H_0: $p = p_0$ принимается на уровне доверия γ, если $|t| < t_{кр}$, где $\Phi(t_{кр}) = \gamma$.

Схемы проверки гипотез о числовых значениях характеристик сведем в табл. 9.2.

Таблица 9.2

Гипотеза H_0	Предположение	Статистика t	Критерий
$a = a_0$	σ^2 известна	$\dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$	$\lvert t\rvert < t_{кр}$
$a = a_0$	σ^2 неизвестна	$\dfrac{(\bar{x} - a_0)\sqrt{n}}{\overline{S}}$	$\lvert t\rvert < t_{n-1,кр}$
$\sigma = \sigma_0$	—	$\dfrac{(n-1)\overline{S^2}}{\sigma_0^2}$	$U < t < V$
$p = p_0$	$n > 30$	$\dfrac{(\omega - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}}$	$\lvert t\rvert < t_{кр}$

Пример 9.3. Партия изделий принимается в том случае, если вероятность того, что изделие окажется соответствующим стандарту, будет не менее 0,95. Среди случайно отобранных 100 изделий оказалось 98 соответствующих стандарту. Можно ли на уровне доверия $\gamma = 0{,}9$ принять партию?

В примере требуется принять гипотезу о том, что неизвестная генеральная доля не менее определенного числа. Пусть H_0: $P(p \geq p_0) = 0{,}95$, H_1: $P < 0{,}95$. Так как конкурирующая гипотеза — левосторонняя, то и критическая область левосторонняя. Вычислим статистику

$$t = \frac{(\omega - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(0{,}97 - 0{,}95)\sqrt{100}}{\sqrt{0{,}95 \cdot 0{,}05}} = 1{,}3768.$$

Из условия $P\{t < t_{кр}\} = 0{,}9$ получим $t_{кр} = 1{,}06$. Так как $t = 1{,}3768 > t_{кр}$, то на уровне доверия 0,9 гипотеза о принятии партии изделий принимается.

9.3. Проверка гипотезы о равенстве средних значений

На практике часто встречаются ситуации, когда среднее значение данных одного эксперимента отличается от среднего значения данных другого, хотя условия экспериментов являются схожими. Тогда возникает вопрос, можно ли считать это расхождение не-

значимым, т.е. случайным, или оно вызвано существенным различием двух генеральных совокупностей. Например, такие вопросы возникают при исследовании надежности технических систем, где результаты сравниваются с предыдущими измерениями; при контроле качества изделий, изготовленных на разных предприятиях; в финансах — при сравнении уровня доходности различных активов.

Пусть $X_1 \in N(a_1; \sigma_1)$, $X_2 \in N(a_2; \sigma_2)$ и дисперсии σ_1^2 и σ_2^2 известны. Имеются выборки $\{x_1, x_2, ..., x_{n_1}\}$ и $\{y_1, y_2, ..., y_{n_2}\}$ из генеральных совокупностей X_1 и X_2.

Гипотеза H_0 состоит в том, что средние распределений равны, т.е. H_0: $a_1 = a_2$, при этом H_1: $a_1 \neq a_2$.

Если выполняется гипотеза H_0, то статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \tag{9.16}$$

будет иметь стандартное нормальное распределение $N(0; 1)$. На заданном уровне доверия γ гипотеза проверяется по схеме п. 9.2.1, где t вычисляется по формуле (9.16).

Если же дисперсии σ_1^2 и σ_2^2 неизвестны, но равны, то доказано, что в случае справедливости гипотезы H_0 статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{n_1 S_1^2 + n_2 S_1^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \tag{9.17}$$

имеет распределение Стьюдента с $k = n_1 + n_2 - 2$ степенями свободы. Здесь S_1^2 и S_2^2 «неисправленные» выборочные дисперсии, т.е.

$$S_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2, \quad S_2^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}(y_j - \bar{y})^2.$$

Для заданного уровня доверия γ по таблицам распределения Стьюдента находим $t_{k,\text{кр}}$ из условия

$$P\left\{|t_k| < t_{k,\text{кр}}\right\} = \gamma \tag{9.18}$$

и гипотеза H_0 принимается, если полученное после вычислений по формуле (9.17) значение t_k удовлетворяет неравенству

$$|t_k| < t_{k,\text{кр}}.$$

Если $n > 30$, то распределения средних $\bar{x}$ и $\bar{y}$ можно считать приближенно нормальными, тогда схема проверки гипотез для средних, изложенная в п. 9.2.1, подходит для выборок из любых генеральных совокупностей.

Пример 9.4. Было произведено 12 измерений диаметра вала (мм). При этом оказалось, что среднее $\bar{x}_1 = 10{,}2$, а стандартное отклонение $S_1 = 0{,}05$. Затем вал поместили в условия с высокой температурой и произвели еще 8 измерений диаметра его оси. Среднее на этот раз оказалось равным 10,25, а стандартное отклонение — 0,06. Можно ли сделать вывод при 5%-ном уровне значимости, что диаметр вала существенно увеличивается при увеличении температуры?

По условию

$$\bar{x}_1 = 10{,}2, \quad \sigma_1 = 0{,}05, \quad n_1 = 12;$$

$$\bar{x}_2 = 10{,}25, \quad \sigma_2 = 0{,}06, \quad n_2 = 8.$$

Гипотеза H_0 состоит в том, что $a_1 = a_2$ против конкурирующей $a_1 \neq a_2$.

Будем считать, что результаты измерений диаметра вала являются нормально распределенными случайными величинами с известными дисперсиями σ_1^2 и σ_2^2. Для проверки гипотезы H_0 составим статистику

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \frac{10{,}2 - 10{,}25}{\sqrt{\dfrac{0{,}05^2}{12} + \dfrac{0{,}06^2}{8}}} = -1{,}9487.$$

Так как при уровне значимости $\alpha = 0{,}05$ критическое значение нормального распределения $t_{\text{кр}} = 1{,}96$ и $|t| < t_{\text{кр}}$, то гипотеза H_0 на уровне доверия 0,95 принимается и можно считать, что диаметр вала существенно не увеличивается в условиях повышенной температуры.

9.4. Критерии согласия

Одной из главных задач математической статистики является установление истинного закона распределения случайной величины на основании экспериментальных данных. Вид закона распре-

деления определяют из общих, технических, финансовых и других соображений, учитывая схожесть условий эксперимента с исследованными ранее, или из теоретических предпосылок.

На практике о виде закона распределения можно судить по графику выборочной плотности распределения вероятностей. Параметры закона распределения обычно неизвестны и их заменяют на выборочные значения. Однако как бы мы ни выбирали закон распределения и его параметры, полной уверенности в том, что в результате получится истинный закон распределения, к которому принадлежит имеющаяся выборка, не существует. Поэтому может ставиться вопрос лишь о том, что на определенном уровне доверия выбранный закон согласуется с данными выборки.

В соответствии с этим критерии, устанавливающие закон распределения, называются *критериями согласия*, к которым можно отнести критерии Колмогорова — Смирнова, Пирсона и др.

Рассмотрим *критерий Пирсона* на примерах и задачах.

Пусть $\{x_1, x_2, ..., x_n\}$ — выборка из некоторой генеральной совокупности X, $F(x)$ — предполагаемая функция теоретического распределения. На основании выборки построим интервальный ряд $\{\Delta_i, n_i\}$, $i = 1, 2, ..., m$, где n_i — число элементов выборки, попавших в интервал $\Delta_i = [a_i, a_{i+1})$. Для каждого интервала Δ_i вычислим теоретические вероятности p_i попадания в него случайной величины X:

$$p_i = P\{X \in \Delta_i\} = F(a_{i+1}) - F(a_i). \tag{9.19}$$

Числа n_i и np_i называются эмпирической и теоретической частотами соответственно. Доказано, что при $n \to \infty$ статистика

$$t = \sum_{i=1}^{n} \frac{(n_i - np_i)^2}{np_i} \tag{9.20}$$

имеет χ^2_k-распределение с $k = m - r - 1$ степенями свободы, где m — число интервалов вариационного ряда; r — число параметров теоретического распределения, вычисленных по экспериментальным данным.

Основная гипотеза H_0 состоит в том, что функцией распределения случайной величины X является выбранная теоретическая функция $F(x)$.

Для заданного уровня доверия γ по таблицам распределения χ^2_k находим критическое значение

$$\chi^2_{k,\text{кр}}: P\left\{\chi^2_k < \chi^2_{k,\text{кр}}\right\} = \gamma. \qquad (9.21)$$

Гипотеза H_0 о согласии экспериментальных данных с распределением $F(x)$ принимается, если

$$t < \chi^2_{k,\text{кр}}.$$

Отметим, что статистика t имеет распределение χ^2 при $n \to \infty$, поэтому критерий Пирсона следует применять только при больших n ($n > 30$).

Замечание. Наряду с приведенным критерием согласия Пирсона имеются и другие критерии, например критерий Колмогорова.

Задачи с решениями

Задача 9.1. Пользуясь критерием Пирсона, проверить гипотезу о нормальности распределения генеральной совокупности по выборке, заданной в табл. 9.3.

Таблица 9.3

–1,754	–0,290	–0,934	0,449	0,512	–0,329	0,085	0,130	–0,323	–0,882
–1,255	1,700	0,634	0,072	–0,490	0,318	–1,088	0,899	1,028	–1,302
1,531	–0,443	1,409	1,730	–0,267	0,349	–0,292	0,882	–0,056	0,757
–0,957	0,248	–0,095	–1,487	–0,361	–0,059	–0,540	–0,078	0,229	0,194
0,415	–1,382	0,129	–2,364	–1,088	–1,085	0,318	0,367	–0,992	0,529

Решение. Построим вариационный ряд распределения. Здесь $x_{\min} = -2{,}364$, $x_{\max} = 1{,}73$. Разобьем интервал $(-2, 2)$ на 5 интервалов одинаковой длины $\Delta = 0{,}8$. Значение $-2{,}364$ отнесем в крайний левый интервал. Подсчитывая число элементов выборки, попадающих в каждый интервал, получим вариационный ряд (табл. 9.4).

Таблица 9.4

x_i	[–2; 1,2)	[–1,2; 0,4)	[–0,4; 0,4)	[0,4; 1,2)	[1,2; 2,0)
n_i	6	11	21	7	5

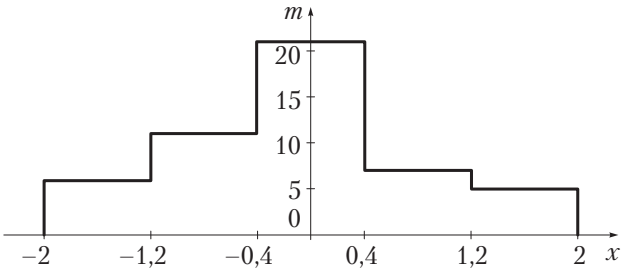


Рис. 9.4. Гистограмма распределения

Гистограмма распределения изображена на рис. 9.4.

По виду гистограммы можно сделать вывод о нормальности распределения случайной величины X. Вычислим выборочные среднюю и дисперсию.

$$\bar{x} = \frac{1}{50}(-1{,}6 \cdot 6 - 0{,}8 \cdot 11 + 0 \cdot 21 + 0{,}8 \cdot 7 + 1{,}6 \cdot 5) = -0{,}096,$$

$$\overline{x^2} = \frac{1}{50}(-1{,}6^2 \cdot 6 - 0{,}8^2 \cdot 11 + 0{,}8^2 \cdot 7 + 1{,}6^2 \cdot 5) = 0{,}7936,$$

$$S_x^2 = \overline{x^2} - (\bar{x})^2 = 0{,}7936 - 0{,}096^2 = 0{,}7844,$$

$$\overline{S_x^2} = \frac{n}{n-1} S_x^2 = \frac{50}{49} \cdot 0{,}7844 = 0{,}8004,$$

$$\bar{S}_x = \sqrt{\overline{S_x^2}} = 0{,}895.$$

Найдем теоретические частоты np_i попадания X в интервалы $[a_i, a_{i+1})$:

$$n = 50; \quad p_i = \frac{1}{2}\left[\Phi\left(\frac{a_{i+1} - \bar{x}}{\bar{S}_x}\right) - \Phi\left(\frac{a_i - \bar{x}}{\bar{S}_x}\right)\right].$$

Для удобства вычислений составим таблицу (табл. 9.5).

Таблица 9.5

i	a_i	$\frac{a_i - \bar{x}}{\bar{S}_x}$	$\Phi\left(\frac{a_i - \bar{x}}{\bar{S}_x}\right)$	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
1	–2,0	–2,127	–0,967	[–2, –1,2)	6	0,107	0,078
2	–1,2	–1,234	–0,779	[–1,2, –0,4)	11	0,256	0,253

Окончание табл. 9.5

i	a_i	$\frac{a_i - \bar{x}}{\bar{S}_x}$	$\Phi\left(\frac{a_i - \bar{x}}{\bar{S}_x}\right)$	Δ_i	n_i	p_i	$\frac{(n_i - np_i)^2}{np_i}$
3	–0,4	–0,340	–0,266	[–0,4, 0,4)	21	0,343	0,864
4	0,4	0,554	0,421	[0,4, 1,2)	7	0,216	1,337
5	1,2	1,448	0,852	[1,2, 2,0)	5	0,078	0,310
6	2	2,342	0,981	—	—	—	$\Sigma = 2{,}842$

По критерию Пирсона вычислим статистику

$$t = \sum_{i=1}^{5} \frac{(n_i - np_i)^2}{np_i} = 2{,}842.$$

Если гипотеза о согласии H_0 верна, то t имеет распределение χ^2 с $k = 5 - 2 - 1 = 2$ степенями свободы. Для $\alpha = 1 - \gamma = 1 - 0{,}9 = 0{,}1$ по таблицам распределения χ^2_2 находим $\chi^2_{2,\text{кр}} = 4{,}6$. Так как $t = 2{,}842 < t_{\text{кр}}$, то гипотеза H_0 принимается.

Задача 9.2. По утверждению руководства фирмы средний размер дебиторского счета равен 187,5 тыс. руб. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборки равна 175 тыс. руб при среднем квадратическом отклонении 35 тыс. руб. Может ли оказаться в действительности правильным объявленный размер дебиторского счета? Принять уровень значимости равным $\alpha = 0{,}05$.

Решение. Здесь $a_0 = 187{,}5$ тыс. руб., $\bar{x} = 175$ тыс. руб., $n = 10$, $\bar{S} = 35$. Поскольку $\alpha = 0{,}05$, то $\gamma = 0{,}95$. Так как дисперсия неизвестна, то для проверки гипотезы H_0: $\bar{x} = a$ воспользуемся распределением Стьюдента. Тогда

$$t = \frac{(\bar{x} - a_0)\sqrt{n}}{\bar{S}} = \frac{(175 - 187{,}5)\sqrt{10}}{35} = -1{,}129.$$

По таблицам распределения Стьюдента при $\gamma = 0{,}95$ находим $t_{\text{кр}} = 2{,}25$. Так как $|t| = 1{,}129 < t_{\text{кр}}$, то гипотеза H_0 о среднем размере дебиторского счета принимается на уровне доверия $\gamma = 0{,}95$.

Задача 9.3. Расходы сырья x_i и y_j на единицу продукции по старой и новой технологиям приведены в табл. 9.6.

Таблица 9.6

	По старой технологии				По новой технологии				
Расход сырья	x_i	304	307	308	y_j	303	304	306	308
Число изделий	n_i	1	4	4	n_j	2	6	4	1

Предполагается, что генеральные совокупности X и Y имеют нормальные распределения с одинаковыми дисперсиями и средними a_1, a_2. Требуется проверить гипотезу H_0: $a_1 = a_2$ против H_0: $a_1 \neq a_2$ на уровне значимости $\alpha = 0{,}1$.

Решение. Вычислим

$$\bar{x} = \frac{1}{n_x}\sum_{i=1}^{3} x_i n_i = 307{,}11,$$

$$\overline{S_x^2} = \frac{1}{n_x - 1}\sum_{i=1}^{3}(x_i - \bar{x})^2 = 1{,}1610,$$

$$\bar{y} = \frac{1}{n_y}\sum_{j=1}^{4} y_j n_j = 304{,}77,$$

$$\overline{S_y^2} = \frac{1}{n_y - 1}\sum_{j=1}^{4}(y_i - \bar{y})^2 = 1{,}2065.$$

Тогда статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{n_x \overline{S_x^2} + n_y \overline{S_y^2}}{n_x + n_y - 2}\left(\dfrac{1}{n_x} + \dfrac{1}{n_y}\right)}} =$$

$$= \frac{307{,}11 - 304{,}77}{\sqrt{\dfrac{9 \cdot 1{,}1610 + 13 \cdot 1{,}2065}{9 + 13 - 2}\left(\dfrac{1}{9} - \dfrac{1}{13}\right)}} = 4{,}722.$$

Если H_0 верна, то статистика t имеет распределение Стьюдента с $n_x + n_y = 20$ степенями свободы. При $\alpha = 0{,}1$ находим $t_{\text{кр}} = 1{,}725$. Так как $|t| > t_{\text{кр}}$, то гипотезу H_0 отвергаем, т.е. при переходе на вторую технологию происходит изменение среднего расхода сырья.

Задачи для самостоятельного решения

9.1. Для проверки эффективности нового лекарства были отобраны две случайные группы по 15 человек, страдающих гриппом. При применении старого лекарства средний срок выздоровления составлял 11 дней с выборочной дисперсией $S_1^2 = 3$, при применении нового — срок выздоровления составил 8 дней с $S_2^2 = 4$. Проверить на уровне 0,99 гипотезу о преимуществе нового лекарства.

9.2. В двух фирмах, выпускающих детское питание, производилась оценка качества продукции. В фирме A, где проверялось 30 единиц продукции, средняя сумма баллов оказалась равной 52. Во второй фирме B проверялось 36 единиц продукции и их средняя сумма баллов оказалась равной 47. Считая дисперсию балльной оценкой, равной 12, определить на уровне значимости $\alpha = 0{,}05$, какая фирма выпускает лучшую продукцию.

9.3. При $n = 4000$ бросаниях монеты Ж.Л. Бюффон получил $m = 2048$ выпадений герба. На каком уровне значимости можно принять гипотезу о том, что вероятность выпадения герба $P = 0{,}5$?

9.4. Точность работы двух станков оценивалась отклонениями от номинала производимой продукции. Из 10 единиц продукции первого станка отклонение от номинала (выборочная дисперсия) составило $S_1^2 = 3{,}5$, из 15 единиц продукции второго станка — $S_2^2 = 4{,}5$. Можно ли считать на уровне значимости $\alpha = 0{,}1$, что станки имеют одинаковую точность?

9.5. В прошлом году доля бракованных изделий, выпускаемых предприятием, равнялась 0,04. В этом году было проверено 300 изделий, из которых 19 оказались бракованными. Можно ли на уровне значимости 0,01 считать, что качество продукции осталось на прежнем уровне?

9.6. Учет времени сборки узла машины бригадой из 10 слесарей показал, что среднее время (в минутах) сборки узла равно $\bar{x} = 76$, а $S^2 = 15$. Предполагая распределение времени сборки нормальным, проверить на уровне значимости $\alpha = 0{,}1$ гипотезу о том, что 75 мин является нормативом (математическим ожиданием) трудоемкости.

9.7. Номинальная точность прибора равна $\sigma_0 = 2$ мкм. Из 10 замеров детали была получена выборочная дисперсия показаний

прибора, равная $\overline{S^2} = 0{,}9$. На уровне доверия $\gamma = 0{,}9$ проверить гипотезу H_0: $\sigma = \sigma_0$ при альтернативе H_1: $\sigma > \sigma_0$.

9.8. По техническим условиям средняя прочность троса составляет 2000 кг. В результате испытаний 20 кусков троса было установлено, что средняя прочность на разрыв равна 1955 кг при средней ошибке 25 кг. Удовлетворяет ли образец троса техническим условиям?

9.9. Для проверки новой технологии были выбраны две группы рабочих численностью $n_1 = 40$ и $n_2 = 50$ человек. В первой группе при применении старой технологии средняя выработка составила $\bar{x}_1 = 85$ (изделий), во второй, где применялась новая технология, $\bar{x}_2 = 95$. Дисперсии по группам $\sigma_1^2 = 100$, $\sigma_2^2 = 75$ были известны заранее. Выяснить на уровне значимости $\alpha = 0{,}05$ влияние новой технологии на производительность.

9.10. Средний годовой оборот 5 компаний в регионе *A* составил 4900 усл. ед., средний оборот 10 компаний в регионе *B* составил 5000 усл. ед. Выборочная дисперсия оборота компаний в регионе *A* оказалась равной 1000, а в регионе *B* — 4000. Считая дисперсии среднегодовых оборотов одинаковыми ($\sigma_1^2 = \sigma_2^2$), проверить на уровне значимости $\alpha = 0{,}05$ гипотезу о равенстве средних значений в регионах *A* и *B*.

9.11. Из 50 человек, покупающих в магазине кофе, 23 выбирают сорт «Арабика». Проверить на уровне значимости $\alpha = 0{,}05$ гипотезу о том, что половина покупателей выбирает данный сорт.

9.12. В результате проверки 10 продавцов одной из торговых точек города были обнаружены недовесы со средним значением $\bar{x}_1 = 150$ г и выборочной дисперсией $S_1^2 = 2500$. В другой точке недовесы характеризовались $\bar{x}_2 = 125$ г и $S_2^2 = 1600$ среди выборки из 15 продавцов. Выяснить на уровне доверия $\gamma = 0{,}95$, в какой точке предпочтительнее покупать продукцию.

9.13. При проверке размеров подшипников из двух партий по 10 штук в каждой, поставленных разными заводами, были обнаружены отклонения от номинала, характеризуемые выборочными дисперсиями $S_1^2 = 9$, $S_2^2 = 8{,}5$. Можно ли считать при уровне доверия $\alpha = 0{,}1$ одинаковой точность изготовления подшипников разными заводами?

9.14. При подбрасываниях игральной кости на верхней грани единица выпала 25 раз, двойка — 19 раз, тройка — 15 раз, четверка — 22 раза, пятерка — 15 раз, шестерка — 21 раз. На каком уровне значимости можно считать, что игральная кость имеет правильную форму?

9.15. Стрелок по летающим тарелкам попадал в цель в 80 % случаев. После тренировок в учебном центре он стал поражать мишени в 89 случаях из 100. Можно ли считать на уровне доверия 0,95, что его квалификация улучшилась?

9.16. Средний урожай пшеницы площадью 1 га составил $Q = 80$ ц/га. После рекультивации почвы был определен средний урожай с площади в 9 га, причем оказалось, что $\overline{Q} = 85$ ц/га, а $S^2_{\overline{Q}} = 4{,}5$. Можно ли считать на уровне доверия 0,05, что урожайность поля увеличилась?

9.17. В больнице скорой помощи фиксировалось количество X вызовов в час специализированных бригад. Наблюдения велись в течение 100 ч, их результаты приведены в табл. 9.6.

Таблица 9.6

Количество вызовов	0	1	2	3	4	5	6	7
Частота вызовов	6	27	26	20	10	5	5	1

Найти выборочный параметр $\bar{x}$ и по критерию Пирсона показать, что на уровне доверия $\alpha = 0{,}05$ случайная величина X распределена по закону Пуассона.

Ответы

9.1. $t_{кр} = 2{,}76$, $t = 4{,}24$.
9.2. $t_{кр} = 1{,}65$, $t = 1{,}652$. Лучшую продукцию выпускает фирма A.
9.3. $\alpha = 0{,}4$.
9.4. $U = 0{,}1$, $V = 7{,}1$, $t = 0{,}6$. Гипотеза об одинаковой точности станков принимается.
9.5. $t_{кр} = 2{,}58$, $t = 1{,}767$. Гипотеза принимается.
9.6. $t_{9,кр} = 1{,}83$, $t = 0{,}82$, $|t| < t_{9,кр}$. Гипотеза принимается.
9.7. $\chi^2_{9,\,кр} = 14{,}7$, $t = 2{,}025$, $|t| < \chi^2_{9,\,кр}$. Гипотеза принимается.
9.8. $t_{кр} = 1{,}96$, $t = 8{,}05$. Гипотеза отвергается, трос не удовлетворяет техническим условиям.

9.9. При $\alpha = 0{,}05$ $t_{кр} = 1{,}96$, $t = 5$. Гипотеза о равенстве средних отвергается, новая технология повышает производительность.
9.10. $t_{кр} = 1{,}8$, $t = -3{,}103$. Средний годовой оборот компании в регионе *B* больше, чем в регионе *A*.
9.11. $t_{кр} = 1{,}96$, $t = 0{,}566$. Гипотеза подтверждается.
9.12. $t_{кр} = 1{,}7$, $t = 1{,}326$. На данном уровне доверия недовесы в торговых точках одинаковы.
9.13. $U = 0{,}1$, $V = 7{,}1$, $t = 1{,}06$. Да.
9.14. $\alpha = 0{,}617$.
9.15. $t_{кр} = 1{,}96$, $t = 2{,}25$. Да.
9.16. $t_{кр} = 1{,}9$, $t = 7{,}071$. Гипотеза об увеличении урожайности принимается.
9.17. $\bar{x} = 2{,}41$, $t = 5{,}46$, $\chi^2_{кр} = 11$, $t < \chi^2_{кр}$. Гипотеза о пуассоновском распределении принимается.

10

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

В данной главе рассмотрен анализ зависимости между переменными на основе свойств дисперсии.

10.1. Условия применения

Дисперсионный анализ был предложен Р. Фишером для решения задач в области биологических исследований, в частности в сельскохозяйственной статистике. В настоящее время дисперсионный анализ определяется как статистический метод, предназначенный для оценки влияния различных факторов на результат эксперимента, в связи с чем область применения этого метода становится значительно шире. Результатом эксперимента является некоторая случайная величина X, называемая также *результативным признаком*. На значения случайной величины X влияет фактор A, состоящий из нескольких *уровней*, или *групп*, A_i, $i = 1, 2, ..., r$.

Пример 10.1. Директора фирмы интересует зависимость выполненных работ за смену от работающей на стройке бригады. Предположим, что на стройке работает r бригад. Объем выполненных работ является результативным признаком X. Работающую бригаду назовем фактором A, или группой фактора A, а через A_i обозначим i-й уровень, или группу фактора A (i-ю бригаду, $i = 1, 2, ..., r$).

В дисперсионном анализе наблюдаемые величины разбиваются на r групп, причем i-я группа содержит выборку из n_i величин $X_i \in N(a + m_i; \sigma_0)$, где σ_0 является постоянной, хотя и неизвестной величиной, не зависящей от i.

Обозначим через x_{ij} значение j-й величины в i-й группе. Модель *однофакторного дисперсионного анализа* можно записать в виде

$$x_{ij} = a + m_i + \varepsilon_{ij}, \qquad (10.1)$$

где a — генеральное среднее всех возможных результатов наблюдений, т.е. MX, m_i — эффект влияния на X, вызванный i-м уровнем фактора A, или, иначе, отклонение математического ожидания a_i результативного признака при i-м уровне фактора от общего математического ожидания a, т.е. $m_i = a_i - a$, ε_{ij} — случайный остаток, отражающий влияние на величину x_{ij} всех других неконтролируемых факторов.

Основными предпосылками дисперсионного анализа являются:

1) остатки ε_{ij} независимы в совокупности для любых i и j;

2) $\varepsilon_{ij} \in N(0; \sigma_0)$ и σ_0 не зависит от i.

Средние значения m_i в (10.1) могут меняться под влиянием некоторых факторов, например под влиянием различных способов обработки, различных видов животных или растений, неоднородности почвы и т.д. Целью эксперимента является исследование этой изменчивости средних значений (например, гипотезы H_0 об их равенстве).

10.2. Разложение суммы квадратов отклонений

Несмещенной оценкой для неизвестной дисперсии σ^2 является, как известно, сумма квадратов

$$Q = \frac{1}{n-1} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2, \tag{10.2}$$

где

$$n = \sum_{i=1}^{r} n_i \tag{10.3}$$

— количество всех наблюдений.

Основная идея дисперсионного анализа заключается в разбиении этой суммы квадратов отклонений на несколько компонент, каждая из которых соответствует предполагаемой причине изменения средних значений m_i.

Обозначим через

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \tag{10.4}$$

среднее арифметическое величин i-й группы, через

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} \tag{10.5}$$

среднее арифметическое всех величин. Тогда справедливо тождество

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2, \tag{10.6}$$

или

$$Q = Q_1 + Q_2. \tag{10.7}$$

Таким образом, полная сумма Q квадратов отклонений от общего среднего разбивается на две компоненты: Q_1 — сумма квадратов между группами, Q_2 — сумма квадратов внутри групп. Если поделить обе части равенства (10.6) на число наблюдений n, то получим известное правило сложения дисперсий.

10.3. Проверка гипотезы о равенстве групповых средних

Пусть H_0: $m_i = m$, $i = 1, 2, ..., r$. Заметим, что величина

$$\overline{S^2} = \frac{Q}{n-1},$$

являющаяся несмещенной оценкой для σ^2, всегда будет иметь распределение χ^2 с $n - 1$ степенями свободы и по ней можно построить доверительный интервал для σ^2 (см. главу 8).

Если гипотеза H_0 верна, то величины

$$\overline{S_\Phi^2} = \frac{1}{r-1}Q_1, \quad \overline{S_0^2} = \frac{1}{n-r}Q_2 \tag{10.8}$$

будут иметь распределение F с $r - 1$ и $n - r$ степенями свободы соответственно, при этом $\overline{S_\Phi^2}$ и $\overline{S_0^2}$ являются несмещенными оценками для межгрупповой дисперсии σ_0^2.

Отношение

$$t = \frac{\frac{1}{r-1}Q_1}{\frac{1}{n-1}Q_2} \tag{10.9}$$

называется *дисперсионным отношением*.

Если гипотеза H_0 верна, то статистика t имеет распределение Фишера с $r-1$ и $n-r$ степенями свободы. В этом случае эффекты влияния уровней фактора A будут нулевыми, т.е. $m_1 = m_2 = ... = m_r = 0$, а оценка параметра a равна общему среднему $\bar{x}$, вычисленному по формуле (10.5). Проверка гипотезы H_0 о равенстве групповых средних проводится по схеме, изложенной в главе 9. Если же гипотеза H_0 отвергается, то параметр a по-прежнему вычисляется по формуле (10.5), а оценка эффекта m_i влияния i-го уровня фактора

$$\bar{m}_i = \bar{x}_i - \bar{x}, \tag{10.10}$$

где $\bar{x}_i$ определяется по формуле (10.4), а $\bar{x}$ — по формуле (10.5).

10.4. Коэффициент детерминации

Предположим, что фактор A влияет на результативный признак X.

Для измерения степени этого влияния используют *выборочный коэффициент детерминации*

$$\bar{d} = \frac{Q_1}{Q}, \tag{10.11}$$

который показывает, какую долю выборочной дисперсии составляет дисперсия групповых средних, иначе говоря, какая доля общей дисперсии объясняется зависимостью результативного признака X от фактора A.

10.5. Сводка формул

Изложенные выше формулы для решения задач однофакторного анализа приведены в табл. 10.1.

Таблица 10.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Оценки дисперсии
Межгрупповая	$Q_1 = \sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2$	$r - 1$	$\overline{S_\Phi^2} = \frac{1}{r-1}$
Внутригрупповая	$Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$	$n - r$	$\overline{S_0^2} = \frac{1}{n-r}$
Общая	$Q = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2$	$n - 1$	$\overline{S^2} = \frac{1}{n-1}$

При вычислении сумм квадратов Q, Q_1, Q_2 часто удобно при $n_i = n_0$ использовать следующие формулы:

$$Q_1 = \frac{\sum_{i=1}^{r}\left(\sum_{j=1}^{n_0} x_{ij}\right)^2}{n_0} - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_0} x_{ij}\right)^2}{n_0 r}, \tag{10.12}$$

$$Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_0} x_{ij}^2 - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_0} x_{ij}\right)^2}{n_0}, \tag{10.13}$$

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{n_0} x_{ij}^2 - \frac{\left(\sum_{i=1}^{r}\sum_{j=1}^{n_0} x_{ij}\right)^2}{n_0 r}. \tag{10.14}$$

В заключение заметим, что мы рассматривали здесь только однофакторный дисперсионный анализ, хотя аналогичные методы имеют место и в случае, когда факторов A несколько (многофакторный дисперсионный анализ).

Задачи с решениями

Задача 10.1. В табл. 10.2 приведены данные по объемам работ, выполненных на стройке за смену для 4 бригад. Вычислить общую несмещенную дисперсию.

Таблица 10.2

Номер бригады	Объем выполненной работы				Групповое среднее	Выборочная дисперсия
1	140	144	142	145	142,75	3,69
2	150	149	152	152	150,25	1,19
3	148	149	146	147	147,50	1,25
4	150	155	154	152	152,75	3,69

Решение. Найдем несмещенные оценки дисперсий

$$\overline{S_1^2} = \frac{S_1^2 n_1}{n_1 - 1} = \frac{3{,}69 \cdot 4}{3} = 4{,}92, \quad \overline{S_2^2} = \frac{1{,}19 \cdot 4}{3} = 1{,}58,$$

$$\overline{S_3^2} = \frac{1{,}25 \cdot 4}{3} = 1{,}25, \quad \overline{S_4^2} = 4{,}92$$

и общую несмещенную дисперсию

$$\overline{S^2} = \frac{(n_1 - 1)S_1^2 + \ldots + (n_4 - 1)S_4^2}{(n_1 - 1) + \ldots + (n_4 - 1)} = \frac{3(4{,}92 + 1{,}58 + 1{,}67 + 4{,}92)}{3 \cdot 4} = 3{,}27.$$

Задача 10.2. Для данных задачи 10.1 проверить гипотезу H_0 о равенстве групповых средних: $m_1 = m_2 = m_3 = m_4$.

Решение. Для проверки гипотезы H_0 вычислим суммы Q_1 и Q_2. Общее выборочное среднее

$$\bar{x} = \frac{140 + 144 + 142 + \ldots + 154 + 152}{16} = 148{,}31.$$

Тогда

$$Q_1 = \sum_{i=1}^{4} 4(\bar{x}_i - \bar{x})^2 =$$

$$= 4\left[(142{,}75 - 148{,}31) + \ldots + (152{,}75 - 148{,}31)^2\right] = 220{,}19,$$

$$Q_2 = \sum_{i=1}^{4}\sum_{j=1}^{4}(x_{ij} - \bar{x}_i)^2 = (140 - 142{,}75)^2 + \ldots + (152 - 152{,}75)^2 = 39{,}27.$$

Вычислим статистику Фишера:

$$t = \frac{(n-r)Q_1}{(r-1)Q_2} = \frac{(16-4)\cdot 220{,}19}{(4-1)\cdot 39{,}27} = 22{,}43.$$

По таблицам Фишера для $\alpha = 0{,}05$ и степеней свободы $k_1 = 3$, $k_2 = 12$ найдем критическое значение $F_{кр} = 3{,}49$.

Так как $t > F_{кр}$, то гипотезу H_0 отклоняем, т.е. считаем, что объем ежедневной выработки зависит от работающей бригады. Оценим степень этой зависимости с помощью коэффициента детерминации. Для этого вычислим

$$Q = \sum_{i=1}^{4}\sum_{j=1}^{4}(x_{ij} - \bar{x}_i)^2 = (140 - 148{,}31)^2 + \ldots + (152 - 148{,}31)^2 = 259{,}46.$$

По формуле (10.11) получим

$$\bar{d} = \frac{Q_1}{Q} = \frac{220{,}19}{259{,}46} = 0{,}849.$$

Это означает, что 84,9 % общей вариации ежедневного объема выработки связано с работающей сменой.

Задача 10.3. Для данных задачи 10.2 найти оценки параметров модели (10.1) дисперсионного анализа.

Решение. Так как объем ежедневной выработки зависит от работающей смены, то оценка параметра a равна $\bar{x} = 148{,}31$.

Оценки параметров m_i

$$\bar{m}_i = \bar{x}_i - \bar{x},$$

т.е.

$$\bar{m}_1 = 142{,}75 - 148{,}31 = -5{,}56, \quad \bar{m}_2 = 150{,}15 - 148{,}31 = 1{,}94,$$

$$\bar{m}_3 = 147{,}5 - 148{,}31 = -0{,}81, \quad \bar{m}_4 = 152{,}75 - 148{,}31 = 4{,}44.$$

Оценка параметра σ_0^2

$$\overline{\sigma_0^2} = \frac{Q}{r-1} = \frac{220{,}19}{3} = 73{,}39.$$

Задачи для самостоятельного решения

10.1. В течение 6 лет использовались 4 различные технологии по выращиванию сельскохозяйственной культуры. Данные по эксперименту (ц/га) приведены в табл. 10.3.

Таблица 10.3

Технология (фактор A)	Урожайность						Число лет наблюдений
A_1	1,2	1,1	1,0	1,3	—	—	4
A_2	0,6	1,1	0,8	0,7	0,7	0,7	6
A_3	0,9	0,6	0,8	1,1	—	—	4
A_4	1,7	1,4	1,3	1,5	1,2	1,3	6

Показать возможность применения к экспериментальным данным дисперсионного анализа. Уровень значимости взять равным $\alpha = 0{,}05$.

10.2. По данным задачи (10.1) установить наличие влияния различных технологий на урожайность культуры при $\alpha = 0{,}05$. Найти оценки для m_1, m_2, m_3, m_4.

10.3. Имеются 4 партии сырья для промышленности. Из каждой партии отобрано по 5 образцов и проведены испытания на определение разрывной нагрузки. Результаты испытаний приведены в табл. 10.4.

Таблица 10.4

Номер партии	Разрывная нагрузка (кг/см)				
1	200	140	170	145	165
2	190	150	210	150	150
3	230	190	200	190	200
4	150	170	150	170	180

Выяснить на уровне значимости $\alpha = 0{,}05$, существенно ли влияние различных партий сырья на величину разрывной нагрузки.

10.4. На заводе установлены 4 линии по выпуску облицовочной плитки. С каждой линии случайным образом в течение смены отобрано по 5 плиток и сделаны замеры толщины (мм). Отклонения от номинального размера даны в табл. 10.5.

Таблица 10.5

Номер линии	Отклонение от номинала				
1	0,6	0,2	0,8	0,5	0,8
2	0,2	0,2	0,8	0,3	0,3
3	0,8	0,6	0,4	0,4	0,9
4	0,7	0,7	0,6	0,6	0,2

Требуется на уровне $\gamma = 0{,}95$ установить зависимость качества плиток от линии. Найти процентное выражение этой зависимости.

10.5. В цехе завода проводилось исследование влияния квалификации рабочих на их заработную плату (тыс. усл. ед.). Результаты опроса сведены в табл. 10.6.

Таблица 10.6

Квалификационный разряд	Заработная плата		
1	0,15	0,17	0,09
2	0,18	0,14	0,15
3	0,20	0,19	0,17
4	0,30	0,25	0,27
5	0,35	0,34	0,35

Установить на уровне значимости $\alpha = 0{,}05$ влияние квалификационного разряда рабочих на их заработную плату. Найти коэффициент детерминации и объяснить его.

Ответы

10.1. Статистика Фишера $t = 0{,}587$, гипотеза о равенстве групповых средних принимается.

10.2. Статистика Фишера $t = 1{,}582$, гипотеза отвергается. Влияние различных технологий на урожайность является незначимым. $\bar{m}_1 = 0{,}15$, $\bar{m}_2 = -0{,}23$, $\bar{m}_3 = -0{,}15$, $\bar{m}_4 = 0{,}40$.

10.3. Статистика Фишера $t = 3{,}65$, гипотеза H_0 отвергается. На уровне значимости $\alpha = 0{,}05$ влияние различных партий сырья на величину разрывной нагрузки существенно.

10.4. Статистика Фишера $t = 3{,}24$, гипотеза H_0 отвергается. На уровне 0,95 можно считать, что качество плиток зависит от линии. Эта зависимость выражается коэффициентом детерминации $\bar{d} = 55{,}1\ \%$.

10.5. Статистика Фишера $t = 38{,}75$, гипотеза H_0 отвергается. Квалификация рабочих очень значимо влияет на их заработную плату. Коэффициент детерминации $\bar{d} = 93{,}9\ \%$.

11

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Будем обозначать через X независимую переменную, а через Y зависимую от нее переменную. Зависимость Y от X называется *функциональной*, если каждому значению величины X соответствует единственное значение величины Y. Однако гораздо чаще в окружающем нас мире каждому значению переменной X соответствует не одно, а множество значений переменной Y, причем заранее сказать, какое именно значение примет величина Y, нельзя. Это объясняется тем, что на результирующую переменную действует не только контролируемый фактор X, но и множество других неконтролируемых случайных факторов.

Таким образом, каждому значению переменной X соответствует определенное (условное) распределение случайной переменной Y. Такая зависимость называется *стохастической*, или *вероятностной*.

Корреляционной зависимостью между двумя переменными, соответствующей стохастической зависимости между ними, называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой. Корреляционная зависимость может быть представлена в виде

$$\varphi(x) = M_x(Y), \tag{11.1}$$

или

$$\psi(y) = M_y(X). \tag{11.2}$$

Уравнения (11.1) и (11.2) называются уравнениями регрессии соответственно Y на X и X на Y, а их графики — *линиями регрессии*. Если

$$\varphi(x) = \text{const}, \quad \psi(y) = \text{const},$$

то говорят, что корреляционная связь отсутствует.

Переменная X в корреляционном анализе может быть как детерминированной, так и случайной величиной.

Основной задачей корреляционного анализа является выявление тесноты связи между переменными X и Y и ее количественная оценка.

Корреляционный анализ следует применять только в том случае, когда данные наблюдений или эксперимента можно считать случайными и выбранными из нормальной совокупности.

11.1. Представление данных в корреляционном анализе

Простейший случай задания экспериментальных данных — несгруппированные данные, т.е. набор пар чисел (x_i, y_i), $i = 1, 2, ..., n$, где $\{x_i, ..., x_n\}$ — выборка значений переменной X, $\{y_i, ..., y_n\}$ — выборка значений переменной Y. Однако очень часто экспериментальные данные задаются *корреляционной таблицей*. Вариант такого задания рассмотрим на следующем примере.

Пример 11.1. Для исследования зависимости годового объема производства Y от основных фондов X получены статистические данные по 20 предприятиям (табл. 11.1).

Таблица 11.1

y_j \ x_i	12,5	17,5	22,5	27,5	n_j
20,5	1	—	—	—	1
21,5	—	2	—	—	2
22,5	—	1	2	—	3
23,5	—	—	3	3	6
24,5	—	—	—	8	8
n_i	1	3	5	11	$n = 20$

В первой строке таблицы записаны значения переменной X, в первом столбце — интервалы изменения переменной Y, а во втором — середины интервалов из первого столбца. Для дальнейших расчетов необходимы только середины интервалов, поэтому сами интервалы могут отсутствовать. Центральную часть таблицы занимают частоты n_{ij} (число предприятий), соответствующие значениям переменных $X = x_i$ и $Y = y_j$. В последней строке записаны частоты появлений значений x_i

$$n_i = \sum_{j=1}^{m} n_{ij}, \quad i = 1, ..., l, \tag{11.3}$$

а в последнем столбце — частоты появлений значений y_j

$$n_j = \sum_{i=1}^{l} n_{ij}, \quad j = 1, ..., m, \tag{11.4}$$

здесь $l = 4$ — число средних значений величины X, $m = 5$ — число средних значений величины Y, $n = \sum n_i = \sum n_j = 20$ — число всех значений.

Таблица такого вида называется *корреляционной*.

11.2. Коэффициент корреляции

Основной оценкой для тесноты связи между переменными X и Y служит *выборочный коэффициент корреляции* r, который определяется формулой

$$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{S_x S_y}. \tag{11.5}$$

Свойства выборочного коэффициента корреляции аналогичны свойствам коэффициента корреляции между случайными величинами X и Y.

Свойства коэффициента корреляции:

1. Коэффициент корреляции принимает значения на отрезке $[-1, 1]$, т.е.

$$-1 \le r \le 1. \tag{11.6}$$

В зависимости от того, насколько $|r|$ приближается к 1, различают слабую, умеренную и сильную связь, т.е. чем ближе $|r|$ к 1, тем теснее связь.

2. Если переменные X и Y умножить на одно и то же число, то коэффициент корреляции не изменится.

3. Если $r = \pm 1$, то корреляционная связь между X и Y представляет собой линейную зависимость.

Запишем более подробные формулы для вычисления коэффициента корреляции

$$r = \frac{n\sum_{i=1}^{l}\sum_{j=1}^{m} x_i y_j n_{ij} - \left(\sum_{i=1}^{l} x_i n_i\right)\left(\sum_{j=1}^{m} y_j n_j\right)}{\sqrt{n\sum_{i=1}^{l} x_i^2 n_i - \left(\sum_{i=1}^{l} x_i n_i\right)^2} \cdot \sqrt{n\sum_{j=1}^{m} y_j^2 n_j - \left(\sum_{j=1}^{m} y_j n_j\right)^2}}, \tag{11.7}$$

$$r = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}(x_i - \bar{x})(y_j - \bar{y})n_{ij}}{nS_xS_y}. \qquad (11.8)$$

Если данные несгруппированные, то формулы (11.7), (11.8) упрощаются:

$$r = \frac{n\sum_{i=1}^{n}x_iy_i - \left(\sum_{i=1}^{n}x_i\right)\left(\sum_{j=1}^{n}y_j\right)}{\sqrt{n\sum_{i=1}^{n}x_x^2 - \left(\sum_{i=1}^{n}x_i\right)^2}\cdot\sqrt{n\sum_{j=1}^{n}y_j^2 - \left(\sum_{j=1}^{n}y_j\right)^2}}, \qquad (11.9)$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{nS_xS_y}. \qquad (11.10)$$

Так как r вычисляется по данным выборки, то, в отличие от генерального коэффициента корреляции ρ, r является величиной случайной. Если $r \neq 0$, то возникает вопрос, объясняется ли это действительно существующей линейной связью между X и Y или вызвано случайными факторами. Для выяснения значимости коэффициента корреляции проверяется гипотеза H_0 об отсутствии линейной корреляционной связи между переменными X и Y, т.е. гипотеза H_0: $\rho = 0$. При справедливости этой гипотезы статистика

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad (11.11)$$

имеет распределение Стьюдента с $k = n - 2$ степенями свободы. Поэтому гипотеза H_0 отвергается, если

$$|r| > t_{k,\text{кр}},$$

где $t_{k,\text{кр}}$ — табличное значение распределения Стьюдента, определенное на уровне значимости α при числе степеней свободы $k = n - 2$.

11.3. Корреляционное отношение

Введенный ранее коэффициент корреляции является показателем тесноты связи лишь в случае линейной зависимости. Для получения показателя тесноты связи при зависимости любого вида запишем правило сложения дисперсий в форме

$$S_y^2 = S_{\Phi y}^2 + S_0^2, \tag{11.12}$$

где

$$S_y^2 = \frac{1}{n}\sum_{j=1}^{m} y_j^2 n_j - \left(\frac{1}{n}\sum_{j=1}^{m} y_j n_j\right)^2 \tag{11.13}$$

— общая дисперсия переменной Y,

$$S_{\Phi y}^2 = \frac{\sum_{i=1}^{l} (\overline{y}_i - \overline{y})^2 n_i}{n} \tag{11.14}$$

— межгрупповая дисперсия,

$$S_0^2 = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m} (y_j - \overline{y}_i)^2 n_i}{n} \tag{11.15}$$

— остаточная дисперсия.

Здесь

$$\overline{y}_i = \frac{1}{n}\sum_{j=1}^{m} y_j n_{ij}$$

— групповые средние.

Остаточная дисперсия представляет собой ту часть дисперсии Y, которая обусловлена неучтенными случайными факторами. Межгрупповая дисперсия выражает ту часть дисперсии Y, которая вызвана изменчивостью X. Величина

$$\eta_{YX} = \frac{S_{\Phi y}}{S_y} \tag{11.16}$$

называется *эмпирическим корреляционным отношением*. Чем теснее связь, тем большее влияние на вариацию Y оказывает изменчивость X, тем больше η_{YX}. Из формулы (10.11) следует, что величина $\bar{d} = \eta^2_{YX}$ является эмпирическим коэффициентом детерминации и показывает, какая часть вариации Y обусловлена вариацией X. Чем ближе $\bar{d}$ к 1, тем теснее наблюдения примыкают к линии регрессии, тем лучше регрессия описывает зависимость переменных. Можно показать, что для линейной модели $\bar{d} = r^2$.

Задачи с решениями

Задача 11.1. Для данных табл. 11.1 найти выборочный коэффициент корреляции, проверить его значимость на уровне $\alpha = 0{,}05$.

Решение. Для вычислений составим табл. 11.2.

Находим суммы

$$\sum_{i=1}^{4} x_i n_i = 480, \quad \sum_{j=1}^{5} y_j n_j = 468,$$

$$\sum_{i=1}^{4} x_i^2 n_i = 11\,925, \quad \sum_{j=1}^{5} y_j^2 n_j = 10\,979$$

и заносим их в табл. 11.2. Вычислим

$$\sum_{i=1}^{4}\sum_{j=1}^{5} x_i y_j n_{ij} = 20{,}5 \cdot 12{,}5 + 21{,}5 \cdot 17{,}5 \cdot 2 + 22{,}5 \cdot 17{,}5 +$$

$$+ 22{,}5 \cdot 22{,}5 \cdot 2 + 23{,}5 \cdot 22{,}5 \cdot 3 + 23{,}5 \cdot 27{,}5 \cdot 3 + 24{,}5 \cdot 27{,}5 \cdot 8 = 11\,330.$$

Таблица 11.2

i, j	x_i	n_i	y_j	n_j	$x_i n_i$	$y_j n_j$	x_i^2	$x_i^2 n_i$	y_j^2	$y_j^2 n_j$
1	12,5	1	20,5	1	12,5	20,5	156,25	156,25	420,25	420,25
2	17,5	3	21,5	2	52,5	43,0	306,25	918,75	462,25	924,5
3	22,5	5	22,5	3	112,5	67,5	506,25	2 531,25	506,25	1 518,75
4	27,5	11	23,5	6	302,5	141,0	756,25	8 318,75	552,25	3 313,5
5	—	—	24,5	8	—	196,0	—	—	600,25	4802
Σ	—	—	—	—	480	468	—	11 925	—	10 979

Подставляя полученные значения сумм в (11.7), найдем выборочный коэффициент корреляции

$$r = \frac{20 \cdot 11\,330 - 480 \cdot 468}{\sqrt{20 \cdot 11\,925 - 480^2} \cdot \sqrt{20 \cdot 10\,979 - 468^2}} = \frac{1960}{90 \cdot 23{,}58} = 0{,}924.$$

Необходимо проверить значимость r на уровне $\alpha = 0{,}05$. Для этого вычислим статистику

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}924 \cdot \sqrt{18}}{\sqrt{1-0{,}924^2}} = 10{,}25.$$

По таблицам распределения Стьюдента при $k = n - 2 = 18$ находим критическое значение $t_{кр} = 2{,}1$. Так как $t > t_{кр}$, то считаем значение r значимым.

Задача 11.2. Для данных табл. 11.1 найти корреляционное отношение η_{YX}, сделать вывод.

Решение. Для вычисления эмпирического корреляционного отношения найдем групповые средние $\overline{y}_i$ и $\overline{y}$:

$$\overline{y}_1 = 20{,}5,$$

$$\overline{y}_2 = \frac{1}{3}(21{,}5 \cdot 2 + 22{,}5) = 21{,}83,$$

$$\overline{y}_3 = \frac{1}{5}(22{,}5 \cdot 2 + 23{,}5 \cdot 3) = 23{,}1,$$

$$\overline{y}_4 = \frac{1}{11}(23{,}5 \cdot 3 + 24{,}5 \cdot 8) = 24{,}23,$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \overline{y}_i n_i = \frac{1}{20}(20{,}5 + 21{,}83 \cdot 3 + 23{,}1 \cdot 5 + 24{,}23 \cdot 11) = 23{,}4.$$

Тогда

$$S^2_{\Phi y} = \frac{1}{n}\sum_{i=1}^{4} (\overline{y}_i - \overline{y})^2 n_i = \frac{1}{20}\big[(20{,}5 - 23{,}4)^2 + (21{,}83 - 23{,}4)^2 \cdot 3 +$$

$$+ (23{,}1 - 23{,}4)^2 \cdot 5 + (24{,}33 - 23{,}4)^2 \cdot 11\big] = 1{,}29,$$

$$S^2_y = \frac{1}{n}\sum_j y_j^2 n_j - \left(\frac{1}{n}\sum_j y_j n_j\right)^2 = \frac{1}{20} \cdot 10\,979 - \left(\frac{1}{20} \cdot 468\right)^2 = 1{,}39.$$

По формуле (11.16) получим

$$\eta_{YX} = \frac{S_{\Phi y}}{S_y} = \sqrt{\frac{1{,}29}{1{,}39}} = 0{,}963.$$

Значит, 96,3 % вариации Y объясняется вариацией X.

Задачи для самостоятельного решения

11.1. По имеющимся данным об уровне механизации труда X (%) и производительности труда Y (т/ч) для 14 предприятий (табл. 11.3) оценить тесноту связи между переменными с помощью коэффициента корреляции и проверить значимость коэффициента корреляции на уровне значимости $\alpha = 0{,}05$.

Таблица 11.3

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

11.2. В табл. 11.4 приводятся данные для розничного товарооборота Z (млрд руб.), средней численности населения X (млн чел.) и среднегодового дохода Y (млн руб.) для некоторого региона.

Таблица 11.4

Z	1,2	1,3	2,5	1,4	1,2	0,2	2,4	4,1	1,1
X	1,4	1,4	2,5	1,5	1,3	0,3	2,6	4,2	1,1
Y	1,3	1,3	1,4	1,8	1,5	1,6	1,8	1,9	1,6

Определить тесноту связи между Z и переменными X и Y.

11.3. Для исследования влияния объема капиталовложений X (млрд руб.) на получаемую годичную прибыль Y (млрд руб.) была собрана статистика по 20 крупным предприятиям (табл. 11.5).

Таблица 11.5

Y \ X	0–10	10–20	20–30	30–40	40–50	n_y
1,5–2,5	1	—	—	—	—	1
2,5–3,5	2	5	2	—	—	9
3,5–4,5	—	3	3	2	—	8
4,5–5,5	—	—	—	—	2	2
n_x	3	8	5	2	2	$n = 20$

Определить тесноту связи между переменными X и Y с помощью выборочного коэффициента корреляции. Оценить значимость коэффициента корреляции r и построить доверительный интервал для генерального значения коэффициента корреляции r на уровне значимости $\alpha = 0{,}05$.

11.4. Определить тесноту связи общего веса X (г) некоторого растения и веса Y (г) его семян на основе выборочных данных, представленных в табл. 11.6.

Таблица 11.6

X	40	50	60	70	80	90	100
Y	20	25	28	30	35	40	45

Проверить значимость выборочного коэффициента корреляции при $\alpha = 0{,}05$.

11.5. В корреляционной табл. 11.7 представлено распределение 50 гастрономических магазинов области по уровню издержек обращения X (%) и годовому доходу Y (млн руб.). Оценить тесноту связи с помощью коэффициента корреляции r и эмпирического корреляционного отношения η_{YX}.

Таблица 11.7

Y \ X	4–6	6–8	8–10	10–12	12–14
0,5–2,0	—	—	2	3	1
2,0–3,5	—	4	5	1	—
3,5–5,0	—	8	5	5	—
5,0–6,5	3	8	2	—	—
6,5–8,0	2	1	—	—	—

Ответы

11.1. $r = 0{,}969$, статистика $t = 2{,}18$, коэффициент корреляции значим.
11.2. $r = 0{,}552$, связь является «средней».
11.3. $r = 0{,}782$, статистика $t = 8{,}54$, коэффициент корреляции значим.
11.4. $r = 0{,}992$, статистика $t = 17{,}6$, коэффициент корреляции значим.
11.5. $r = -019{,}6$, $\eta_{YX} = 0{,}642$, оба коэффициента значимы.

12

РЕГРЕССИОННЫЙ АНАЛИЗ

При изучении стохастических зависимостей в задачах техники, экономики и т.д. одним из главных моментов является установление вида зависимости Y от $X_1, X_2, ..., X_n$, т.е. вида уравнения регрессии. Это связано, в первую очередь, с необходимостью прогнозирования исследуемых процессов.

Установление формы зависимости, оценка функции регрессии и ее параметров являются задачами ***регрессионного анализа***.

В регрессионном анализе изучаются модели вида

$$Y = \varphi(X) + \varepsilon, \tag{12.1}$$

где X — неслучайная независимая переменная, называемая *фактором*; Y — случайная зависимая переменная (результирующий признак); ε — случайная переменная, характеризующая отклонение от линии регрессии (остаточная переменная).

Основными предпосылками регрессионного анализа являются:

1. Для каждого из наблюдений с номером i, $i = 1, 2, ..., n$, $\varepsilon_i \in N(a; \sigma)$, где

$$\sigma^2 = DY_i = D\varepsilon_i = \text{const}. \tag{12.2}$$

2. Для разных наблюдений переменные ε_i и ε_j некоррелированны при $i \neq j$.

Уравнение регрессии (12.1) записывается в виде

$$y_x = \varphi(x, b_0, b_1, ..., b_p), \tag{12.3}$$

где x — значения величины X; $y_x = M_x Y$; $b_0, b_1, ..., b_p$ — параметры функции регрессии.

Таким образом, задача регрессионного анализа состоит в определении функции φ, ее параметров $b_0, b_1, ..., b_p$ и дальнейшего статистического исследования уравнения (12.3).

12.1. Корреляционное поле

Рассмотрим понятие корреляционного поля на примере данных табл. 11.1, где через x_i и y_j обозначены соответственно середины интервалов изменения переменных, а через n_i и n_j — их частоты.

Зависимость между X и Y изобразим точками координатной плоскости (x, y). На рис. 12.1 точки (x_i, y_j) откладываются соответственно значениям переменных X и Y.

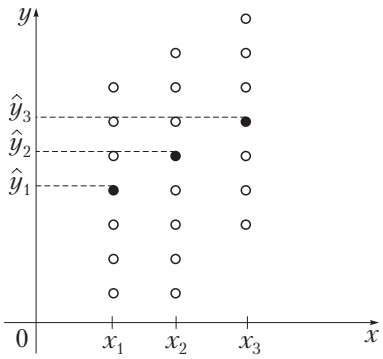


Рис. 12.1. Поле корреляции

Графическое изображение типа, представленного на рис. 12.1, называется *полем корреляции*.

Для каждого значения x_i, $i = 1, 2, ..., l$, вычислим групповое среднее $\overline{y}_i$. Соединив полученные точки ломаной линией, получим график *эмпирической линии* регрессии Y по X.

По виду эмпирической линии регрессии можно предположить, что связь между величинами Y и X является линейной.

12.2. Линейная регрессия

Если функция φ в уравнении (12.3) линейна по x, т.е.

$$y_x = a + bx, \tag{12.4}$$

то говорят, что имеет место *линейная регрессия* Y по X.

Для определения параметров a и b используем систему нормальных уравнений

$$\begin{cases} na + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i, \\ a\sum\limits_{i=1}^{n} - b\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases} \tag{12.5}$$

Это линейная система двух уравнений с двумя неизвестными a и b, из которой следует, что

$$a = \frac{\sum\limits_{i=1}^{n} y_i \sum\limits_{i=1}^{n} x_i^2 - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} x_i y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}, \tag{12.6}$$

$$b = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}. \tag{12.7}$$

Учитывая, что

$$n\bar{x} = \sum_{i=1}^{n} x_i, \quad n\bar{y} = \sum_{i=1}^{n} y_i, \quad n\overline{x^2} = \sum_{i=1}^{n} x_i^2, \quad n\bar{x} \cdot \bar{y} = \sum_{i=1}^{n} x_i y_i,$$

из равенств (12.6) и (12.7) получим

$$a = \frac{\bar{y} \cdot \overline{x^2} - \bar{x} \cdot \overline{xy}}{\bar{x}^2 - (\bar{x})^2}, \tag{12.8}$$

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2}. \tag{12.9}$$

Коэффициент b в уравнении регрессии (12.4) называется *коэффициентом регрессии* Y по X и обозначается b_{YX}.

Этот коэффициент находится по формуле

$$b_{YX} = r\frac{S_y}{S_x}. \tag{12.10}$$

Линейное уравнение регрессии можно записать в принятой в математической статистике форме:

$$y_x - \bar{y} = b_{YX}(x - \bar{x}) \tag{12.11}$$

или

$$y_x - \bar{y} = r\frac{S_y}{S_x}(x - \bar{x}). \tag{12.12}$$

Аналогично, уравнение регрессии X на Y имеет вид

$$x_y - \bar{x} = r\frac{S_x}{S_y}(y - \bar{y}). \tag{12.13}$$

12.3. Статистический анализ уравнения регрессии

Для того чтобы установить, соответствует ли выбранная регрессионная модель (12.1) экспериментальным данным, используют основное уравнение дисперсионного анализа:

$$Q = Q_\Phi + Q_0,$$

где

$$Q = \sum_{j=1}^{m}(y_j - \bar{y})^2 n_j \tag{12.14}$$

— общая сумма квадратов отклонений Y от средней,

$$Q_\Phi = \sum_{i=1}^{l}(y_{x_i} - \bar{y})^2 n_i \tag{12.15}$$

— сумма квадратов, обусловленная регрессией,

$$Q_0 = \sum_{i=1}^{l}\sum_{j=1}^{m}(y_j - y_{x_i})^2 n_{ij} \tag{12.16}$$

— остаточная сумма квадратов.

Для несгруппированной выборки формулы (12.14)–(12.16) несколько упрощаются:

$$Q = \sum_{i=1}^{n}(y_j - \bar{y})^2, \tag{12.17}$$

$$Q_{\Phi} = \sum_{i=1}^{n} (y_{x_i} - \bar{y})^2, \tag{12.18}$$

$$Q_0 = \sum_{i=1}^{n} (y_i - y_{x_i})^2. \tag{12.19}$$

Для заданного уровня α находим критическое значение $F_{\text{кр}}$ распределения Фишера при $k_1 = l - k$, $k_2 = n - l$ степенях свободы, где n — число наблюдений, l — число групп в корреляционной таблице или число оцениваемых параметров в несгруппированной выборке. Если статистика

$$t = \frac{Q_{\Phi}(n-l)}{Q_0(l-1)} > F_{\text{кр}}, \tag{12.20}$$

то уравнение регрессии считается значимым, т.е. соответствующим экспериментальным данным на уровне значимости α.

В случае линейной регрессии при $l = 2$ уравнение регрессии значимо на уровне α, если

$$t = \frac{Q_{\Phi}(n-2)}{Q_0} > F_{\text{кр}}. \tag{12.21}$$

Воздействие неучтенных случайных факторов в линейной модели определяется остаточной дисперсией σ_0^2. Оценкой этой дисперсии является выборочная остаточная дисперсия

$$S_0^2 = \frac{1}{n-l} Q_0. \tag{12.22}$$

Задача с решением

Для зависимости Y от X, заданной корреляционной табл. 11.1, найти оценки параметров a и b уравнения линейной регрессии $y_x = a + bx$, остаточную дисперсию.

Выяснить значимость уравнения регрессии при $\alpha = 0{,}05$.

Решение. Воспользуемся результатами задачи 11.1.

$$\bar{x} = \frac{1}{n}\sum_i x_i n_i = \frac{1}{20} \cdot 480 = 24, \quad \bar{y} = 23{,}4, \quad r = 0{,}934;$$

$$\overline{x^2} = \frac{1}{n}\sum_i x_i^2 n_i = \frac{1}{20}\cdot 11\,925 = 596{,}25, \quad S_x^2 = 20{,}25, \quad S_x = 4{,}5;$$

$$\overline{y^2} = \frac{1}{n}\sum_j y_j^2 n_j = \frac{1}{20}\cdot 10\,979 = 548{,}95, \quad S_y^2 = 1{,}39, \quad S_y = 1{,}179.$$

Согласно формуле (12.12) уравнение регрессии будет иметь вид

$$y_x - 23{,}4 = 0{,}934 \cdot \frac{1{,}179}{4{,}5}(x - 24),$$

отсюда $y_x = 17{,}524 + 0{,}2447x$, тогда $\alpha = 17{,}53$, $b = 0{,}245$.

Для выяснения значимости уравнения регрессии вычислим суммы Q_Φ и Q_0. Составим расчетную таблицу (табл. 12.1).

Таблица 12.1

x_i	$y_{x_i} = a + bx_i$	n_i	$(y_{x_i} - \bar{y})^2 n_i$	$\sum_{j=1}^{m}(y_j - y_{x_i})^2 n_{ij}$
12,5	20,59	1	7,90	0,0081
17,5	21,81	3	7,58	0,6683
22,5	23,03	5	0,68	1,2245
27,5	24,26	11	8,14	1,8128
Σ	—	20	24,30	3,7137

Из формул (12.15) и (12.16) по данным табл. 12.1 получим

$$Q_\Phi = 24{,}3, \quad Q_0 = 3{,}71.$$

При $k_1 = l - 1 = 4 - 1 = 3$, $k_2 = n - l = 20 - 4 = 16$ и $\alpha = 0{,}05$ по таблицам распределения Фишера находим критическое значение $F_{кр} = 3{,}24$. Вычислим статистику (12.20):

$$t = \frac{Q_\Phi(n-l)}{Q_0(l-1)} = \frac{24{,}3 \cdot 16}{3{,}71 \cdot 3} = 35.$$

Так как $t > F_{кр}$, то уравнение регрессии значимо. Остаточная дисперсия

$$S_0^2 = \frac{1}{n-l}Q_0 = \frac{3{,}71}{16} = 0{,}232.$$

Задачи для самостоятельного решения

Задача 12.1. В табл. 12.2 приведены данные о связи между ценой на нефть X (усл. ед.) и индексом нефтяных компаний Y (усл. ед.). Предполагая, что связь между величинами X и Y линейна, найти функцию регрессии, оценить ее значимость и построить доверительный интервал уровня $\alpha = 0{,}05$ для Y при $X = 12$.

Таблица 12.2

x	11,0	11,5	12,0	12,5	13,0	13,5
y	1,5	1,5	1,6	1,7	1,9	1,9

Задача 12.2. В табл. 12.3 приведены данные о X и Y. Предполагая, что связь между величинами X и Y линейна, найти функцию регрессии, оценить ее значимость и построить доверительный интервал уровня $\alpha = 0{,}05$ для Y при $X = 12$.

Таблица 12.3

x	80	100	120	140	160	180	200	220	240	260
y	55	65	79	80	102	110	120	135	137	150

Задача 12.3. По данным табл. 12.4:

Таблица 12.4

x	2	2,1	2,3	2,4	2,9	3,3	3,8	4,6
y	14,3	18,6	18,7	20,9	22,3	24,2	25,7	27

1. Построить уравнение регрессии, описывающее зависимость между среднесуточной производительностью труда Y (т) и стоимостью основных производственных фондов X (млн руб.).

2. Оценить степень тесноты связи между переменными с помощью коэффициента корреляции.

3. Найти доверительный интервал для уравнения регрессии в целом, а также для индивидуального прогнозного значения $y_{\text{пр}}$ для $x_{\text{пр}} = 3$.

Ответы

12.1. $\hat{y} = -0{,}677 + 0{,}189x$. Функция регрессии значима.
12.2. $\hat{y} = 18{,}32 + 0{,}59x$. Функция регрессии значима.
12.3. $\hat{y} = 9{,}101 + 4{,}226x$. Функция регрессии значима. $r = 0{,}922$.

ЧАСТЬ III

Математическое программирование

- Классические задачи линейного программирования
- Основные понятия задач линейного программирования и графический метод их решения
- Симплексный метод. Теория двойственности
- Транспортная задача

Математическое программирование является составной частью предмета исследования операций. Его содержание включает задачи нахождения оптимальных решений при анализе сложных экономических систем. При этом экономические процессы описываются с помощью математических моделей. Линейное программирование выделяется линейными связями между параметрами объектов. Отыскание условных экстремумов проводится на множествах, заданных линейными уравнениями или неравенствами.

Модели линейного программирования отличаются наглядностью и относительной простотой, они получили довольно широкое распространение, начиная с середины прошлого века. К числу классических задач линейного программирования (ЗЛП) относятся:

1) задача оптимального планирования;
2) задача об оптимальной корзине продуктов;
3) задача оптимального раскроя материалов;
4) транспортная задача;
5) задача о назначениях.

Наряду с линейным программированием раздел математического программирования включает и другие задачи. Среди них можно отметить целочисленное, нелинейное, выпуклое, динамическое, стохастическое программирование и др. [5], [8].

Усвоение отмеченных разделов математического программирования требует значительной математической подготовки и затрат времени. По этой причине в III части авторы ограничились основными разделами линейного программирования, уделяя больше внимания примерам решения задач. Приводятся графический и симплекс-метод решения задач. Даются основы теории двойственности и анализа решения задач. Подробно описана процедура решения и проведения анализа на базе компьютерных технологий с использованием MS Excel. Решение транспортной задачи проводится методом потенциалов.

13

КЛАССИЧЕСКИЕ ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В этой главе приведены классические примеры задач линейного программирования. Методы их решения будут рассмотрены позднее (в тексте даны лишь ответы).

13.1. Задача о планировании производственной программы предприятия

Предприятие может выпускать n видов продукции $P_1, P_2, ..., P_n$, располагая для этого m различными ресурсами $R_1, R_2, ..., R_m$ в количествах $b_1, b_2, ..., b_m$ соответственно. Известно, что для выпуска единицы продукции P_j необходимо затратить a_{ij} единиц ресурса R_i, $i = 1, 2, ..., m$, $j = 1, 2, ..., n$. Кроме того, известен доход от продажи единицы каждого вида продукции — $c_1, c_2, ..., c_n$ соответственно, где c_j — стоимость единицы продукта P_j, например штуки, тонны и т.д.

Требуется спланировать производственную программу — объемы выпуска каждого вида продукции (в штуках, тоннах и т.п.), чтобы максимизировать доход предприятия. Для удобства дальнейших выводов и рассуждений сведем исходную информацию в единую табл. 13.1, где через x_j обозначаются объемы продукции P_j, выпускаемой предприятием. Тогда набор переменных $(x_1, x_2, ..., x_n)$ представляет собой не что иное, как производственную программу предприятия. Доход, полученный предприятием при производстве продукта P_j в количествах x_j, составит $c_j x_j$, а при реализации производственной программы $(x_1, x_2, ..., x_n)$ будет равен величине

$$F = c_1x_1 + c_2x_2 + ... + c_nx_n, \qquad (13.1)$$

которая называется *целевой функцией*.

Подсчитаем, какое количество ресурсов будет израсходовано, если выбрать некоторый план $(x_1, x_2, ..., x_n)$. Ресурса R_1 потребуется $a_{11}x_1 + a_{21}x_2 + ... + a_{1n}x_n$, в то время как в наличии имеется b_1 и т.д.

Таблица 13.1

Необходимые ресурсы	Запасы ресурсов	Продукция			
		P_1	P_2	...	P_n
		Объемы выпуска			
		x_1	x_2	...	x_n
R_1	b_1	a_{11}	a_{12}	...	a_{1n}
R_2	b_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...	...
R_m	b_m	a_{m1}	a_{m2}	...	a_{mn}
Доход от реализации единицы продукции		c_1	c_2	...	c_n

Очевидно, что производственная программа может быть выполнена только в том случае, если имеющихся ресурсов окажется достаточно, т.е. при выполнении условий:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ ..... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m. \end{cases} \qquad (13.2)$$

Кроме того, понятно, что переменные плана (x_1, x_2, ..., x_n) должны быть неотрицательными числами, т.е.

$$x_1 \ge 0, \quad x_2 \ge 0, \quad ..., \quad x_n \ge 0. \qquad (13.3)$$

Объединяя полученные результаты, получаем следующую ЗЛП.

Требуется найти совокупность значений (x_1, x_2, ..., x_n), обращающих в максимум целевую функцию (13.1), при условии, что переменные (x_1, x_2, ..., x_n) удовлетворяют системам ограничений (13.2) и (13.3).

Приведем конкретный пример такой задачи.

Пример 13.1. Фирма выпускает два вида древесностружечных плит — обычные и улучшенные. При этом производятся две основные операции — прессование и отделка. По технологии производства плиты изготавливаются партиями по 100 штук в каждой. Требуется указать, какое количество плит каждого типа можно изготовить в течение месяца так, чтобы обеспечить максимальную прибыль при следующих ограничениях на ресурсы (материал, время, затраты) (табл. 13.2).

Таблица 13.2

Ресурсы	Количество ресурсов на 1 мес.	Партия из 100 плит	
		обычных	улучшенных
Материал, м^2	4000	20	40
Время на прессование, чел.-ч	900	4	6
Время на отделку, чел.-ч	600	4	4
Необходимые денежные затраты, ден. ед.	6000	30	50
Доход от реализации единицы продукции, ден. ед.		80	100

Построим математическую модель задачи. Введем обозначения. Пусть x_1 — количество партий в 100 плит обычного вида, изготавливаемых в течение месяца, а x_2 — количество партий в 100 плит улучшенного вида. Тогда доход от реализации плит при производственной программе (x_1, x_2) будет равен $80x_1 + 100x_2$.

Подсчитаем расходы ресурсов, необходимые для выполнения производственной программы (x_1, x_2).

Материала будет израсходовано $20x_1 + 40x_2$ (m^2) при запасе, равном $4000m^2$. Следовательно, для того чтобы реализовать программу, необходимо выполнить условие

$$2x_1 + 40x_2 \le 4000.$$

Расход ресурса «время на прессование» составит $4x_1 + 6x_2$ (чел.-ч) при запасе, равном 900 чел.-ч. Следовательно, для того чтобы реализовать программу, необходимо выполнить условие

$$4x_1 + 6x_2 \le 900.$$

Расход ресурса «время на отделку» составит $4x_1 + 4x_2$ (чел.-ч) при запасе, равном 600 чел.-ч. Следовательно, для того чтобы реализовать программу, необходимо выполнить условие

$$4x_1 + 4x_2 \le 600.$$

Расход ресурса «денежные затраты» составит $30x_1 + 50x_2$ (ден. ед.) при запасе, равном 6000 ден. ед. Следовательно, для того чтобы реализовать программу, необходимо выполнить условие

$$30x_1 + 50x_2 \le 6000.$$

Кроме этого, переменные x_1 и x_2 не могут принимать отрицательных значений, т.е. $x_1 \ge 0$, $x_2 \ge 0$.

Подводя итог, получаем следующую ЗЛП:

$$F = 80x_1 + 100x_2(\max),$$

$$\begin{cases} 20x_1 + 40x_2 \le 4000, \\ 4x_1 + 6x_2 \le 900, \\ 4x_1 + 4x_2 \le 600, \\ 30x_1 + 50x_2 \le 6000, \end{cases}$$

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Методы решения ЗЛП будут даны ниже. Рассмотренную задачу можно также решить графически.

Приведем ответ: для получения максимального дохода в размере 13 000 ден. ед. необходимо изготавливать 100 партий плит обычного вида и 50 партий плит улучшенного вида.

13.2. Задача об оптимальной корзине продуктов (задача о диете)

Медициной установлена суточная потребность организма в питательных веществах T_1, T_2, ..., T_m (белки, жиры, углеводы и т.д.). Пусть эта потребность выражается в количествах не менее чем b_1, b_2, ..., b_m некоторых условных единиц (например, мг). Пусть нужные питательные вещества содержатся в продуктах питания $П_1$, $П_2$, ..., $П_n$, стоимость которых известна и равна c_1, c_2, ..., c_n соответственно, где c_j — стоимость единицы продукта $П_j$, например литра, килограмма, упаковки и т.д. Пусть известно также, что питательное вещество T_i содержится в единице продукта $П_j$ в количестве a_{ij} единиц ($i = 1, 2, ..., m; j = 1, 2, ..., n$).

Требуется составить продуктовую корзину с минимальной стоимостью, т.е. определить количество продуктов каждого вида x_1, x_2, ..., x_n, которые следует приобрести для того, чтобы, с одной стороны, удовлетворить суточную потребность организма в питательных веществах, с другой — израсходовать минимум средств.

Обозначим стоимость искомой корзины через F. Тогда для некоторого продуктового набора (x_1, x_2, ..., x_n)

$$F = c_1x_1 + c_2x_2 + ... + c_nx_n.$$

Подсчитаем количество каждого питательного вещества T_i, $i = 1, 2, ..., m$, содержащегося в корзине $(x_1, x_2, ..., x_n)$.

Содержание вещества T_1 равно $a_{11}x_1 + a_{21}x_2 + ... + a_{1n}x_n$, в то время, как его должно быть не менее чем b_1. В общем, содержание вещества T_i равно $a_{i1}x_1 + a_{i1}x_2 + ... + a_{in}x_n$, в то время, как его должно быть не менее чем b_i.

Кроме того, существует потребность организма в калориях. Пусть a_j количество ккал, содержащихся в единице продукта T_j. Тогда ограничение по калорийности имеет вид

$$a_1x_1 + a_2x_2 + ... + a_nx_n \geq b_{m+1}.$$

Из смысла задачи следует также, что решения $(x_1, x_2, ..., x_n)$ не могут быть отрицательными числами, т.е. $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_n \geq 0$.

В результате приходим к следующей ЗЛП: необходимо найти такой набор значений для переменных $(x_1, x_2, ..., x_n)$, которые обращают в минимум целевую функцию (13.4) и одновременно удовлетворяют системе ограничений (13.5), (13.6):

$$F = c_1x_1 + c_2x_2 + ... + c_nx_n(\min), \tag{13.4}$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \geq b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \geq b_2, \\ ..... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \geq b_m, \\ a_1x_1 + a_2x_2 + ... + a_nx_n \geq b_{m+1}, \end{cases} \tag{13.5}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad ..., \quad x_n \geq 0. \tag{13.6}$$

На переменные $x_1, ..., x_n$ могут также накладываться условия, записанные линейными неравенствами или равенствами.

Пример 13.2. Для похудения диетологи советуют перейти на рациональное питание, состоящее из двух продуктов P и Q. Дневное питание этими новинками должно давать не более 14 ед. жира (чтобы похудеть), но и не менее 300 ккал (чтобы не сойти с дистанции раньше). На банке с продуктом P написано, что в одном килограмме этого продукта содержится 15 ед. жира и 150 ккал, а на банке с продуктом Q — 4 ед. жира и 200 ккал. При этом цена 1 кг продукта P равна 15 усл. ед., а 1 кг продукта Q — 25 усл. ед. С целью оптимизации средств возникает вопрос: в какой пропорции надо брать эти продукты P и Q для того, чтобы выдержать условия диеты и истратить как можно меньше денег?

Обозначим через x_1 количество продукта P и через x_2 количество продукта Q, требуемые для выполнения условий диеты.

Количество единиц жира, содержащегося в x_1 кг продукта P и в x_2 кг продукта Q, равно $15x_1 + 4x_2$ и по условию диеты не должно превосходить 14:

$$15x_1 + 4x_2 \le 14.$$

В свою очередь, количество ккал, содержащихся в x_1 кг продукта P и в x_2 кг продукта Q, равно $150x_1 + 200x_2$ и по условию диеты должно быть не менее 300:

$$150x_2 + 200x_2 \ge 300.$$

Теперь о стоимости F продуктов. Она равна

$$F = 15x_1 + 25x_2$$

и в соответствии с высказанными пожеланиями должна быть минимальной.

Кроме этого, переменные x_1 и x_2 не могут принимать отрицательных значений, т.е. $x_1 \ge 0$, $x_2 \ge 0$.

Объединяя результаты, получаем следующую ЗЛП:

$$F = 15x_1 + 25x_2(\min),$$

$$\begin{cases} 15x_1 + 4x_2 \le 14, \\ 150x_1 + 200x_2 \ge 300, \end{cases}$$

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Рассмотренную задачу можно решить графическим методом.

Приведем ответ: для того, чтобы выдержать условия диеты и истратить как можно меньше денег, нужно брать рекомендованные продукты в отношении 2:3.

13.3. Задача оптимального раскроя материалов

Часто на практике приходится вести раскрой как рулонного, так и листового материала. Исходные параметры задачи следующие:

- x_j — количество рулонов, раскраиваемых по j-му варианту;
- P_{ij} — количество полос i-го размера, выкраиваемых из рулона по j-му варианту раскроя;
- c_j — размер краевого отхода при раскрое рулона по j-му варианту;
- a_i — общее количество полос i-го размера, которое надо получить при раскрое.

В общем виде по минимуму отходов модель задачи может быть записана в виде:

$$F = \sum_{j=1}^{n} c_j x_j (\min),$$

$$\sum_{j=1}^{n} P_{ij} x_j = a_i, \quad i = \overline{1, m},$$

$$x_j \geq 0, \quad j = \overline{1, n}.$$

Пример 13.3. Пусть имеются отрезы полотна длиной 200 м. Из них необходимо выкроить скатерти (салфетки) трех видов: квадратные, прямоугольные и овальные, ширина которых равна ширине отреза, а длина соответственно — 57, 82 и 110 см. Из скатертей составляются комплекты из 4 квадратных, 4 прямоугольных и 5 овальных скатертей. Необходимо изготовить 50 комплектов. Известны 5 видов раскроя одного отреза полотна. Количество скатертей, выкраиваемых из одного отреза полотна при каждом способе раскроя, дано в табл. 13.3.

Таблица 13.3

Способ раскроя	Вид скатертей		
	Квадратные	Прямоугольные	Овальные
1	3	—	—
2	2	1	—
3	1	—	1
4	—	2	—
5	—	1	1

Определить, какое количество отрезов полотна необходимо разрезать каждым способом для изготовления 50 комплектов, чтобы отходы, получаемые при раскрое полотна, были наименьшими.

Для производства 50 комплектов необходимо $4 \cdot 50 = 200$ квадратных скатертей, $4 \cdot 50 = 200$ прямоугольных и $5 \cdot 50 = 250$ овальных скатертей. Общее количество изготовления квадратных скатертей при условии, что первым способом раскроено x_1 отрезов полотна, вторым способом — x_2 отрезов и т.д., можно выразить суммой $3x_1 + 2x_2 + x_3 + 0x_4 + 0x_5$. По условию эта сумма должна равняться 200. Аналогично записываем общее количество прямоугольных и овальных скатертей.

$$\begin{cases} 3x_1 + 2x_2 + x_3 = 200, \\ x_2 + 2x_4 + x_5 = 200, \\ x_3 + x_5 = 250. \end{cases}$$

Кроме записанных условий, на переменные x_1, ..., x_5 накладывается обязательное условие неотрицательности $x_i \geq 0$, $i = 1, 2, 3, 4, 5$.

Нашей целью является нахождение таких значений переменных, при которых отходы производства будут минимальными. Составим функцию, выражающую суммарную величину отходов. При первом способе отходы от каждой полосы составят $200 - 57 \cdot 3 = 29$, при втором — $200 - (57 \cdot 2 + 82) = 4$, при третьем — $200 - (57 + 101) = 42$, при четвертом — $200 - 82 \cdot 2 = 36$, при пятом — $200 - (82 + 101) = 17$. Суммарную величину отходов можно выразить в виде

$$F(x) = 29x_1 + 4x_2 + 42x_3 + 36x_4 + 17x_5(\min).$$

Решив пример, получили наличие единственного оптимального плана. При этом значение целевой функции равно 6950, которое характеризует суммарную величину отходов в размере 6950 см. Чтобы получить такие минимальные отходы, необходимо раскроить первым способом 50 отрезов, третьим способом 50 отрезов и пятым способом 200 отрезов, а вторым и четвертым не раскраивать. По условию примера компоненты оптимального плана должны выражаться целыми числами, которые и получили при решении.

13.4. Транспортная задача

Транспортная модель — это частный случай модели линейного программирования. Транспортная задача (ТЗ) включает в себя задание некоторого множества пунктов производства или нескольких торговых складов, которые осуществляют поставки в некоторое множество пунктов назначения — несколько магазинов. Цель состоит в минимизации общей стоимости транспортировки в рамках ограничений на потребности и предложения. Решение этой задачи может быть найдено с помощью традиционных методов решения ЗЛП. Однако относительно простая структура задачи позволяет ее решать более простыми методами.

Пример 13.4. Компания имеет два товарных склада и трех оптовых покупателей. Известно, что общий объем запасов на складах составляет 300 тыс. ед. продукции и совпадает с общим объемом заказов покупателей. Конкретные данные о загруженности каждого из складов (в тыс. ед.), потребности каждого покупателя (в тыс. ед.) и стоимости перевозки (ден. ед. за 1 тыс. ед.) приведены в табл. 13.4.

Таблица 13.4

		Стоимость перевозок к потребителям, ден. ед. за 1 тыс. ед.			Запас, тыс. ед.
		B_1	B_2	B_3	
Склады	A_1	8	5	6	120
	A_2	4	9	7	180
Спрос, тыс. ед.		70	140	90	300

Обозначим через x_{ij} количество товара, поставляемого со склада A_i покупателю B_j. Тогда эта задача может быть сформулирована следующим образом.

Минимизировать общую стоимость перевозок:

$$F = 8x_{11} + 5x_{12} + 6x_{13} + 4x_{21} + 9x_{22} + 7x_{23}(\min)$$

при условии, что потребности в грузах и их запасы будут одинаковыми

$$\begin{cases} x_{11} + x_{12} + x_{13} = 120, \\ x_{11} + x_{22} + x_{23} = 180, \\ x_{11} + x_{21} = 70, \\ x_{12} + x_{22} = 140, \\ x_{13} + x_{23} = 90, \end{cases}$$

$$x_{11} \geq 0, \quad x_{12} \geq 0, \quad x_{13} \geq 0, \quad x_{21} \geq 0, \quad x_{22} \geq 0, \quad x_{23} \geq 0.$$

Получаем ЗЛП, в которой основные ограничения задаются равенствами вследствие того, что ТЗ сбалансирована. Такая ТЗ называется транспортной задачей *закрытого* типа.

Приведем ответ: для того чтобы общая стоимость была минимальной $F_{\min} = 1690$ ден. ед., надо взять $x_{11} = 0$, $x_{12} = 120$, $x_{13} = 0$, $x_{21} = 70$, $x_{22} = 20$, $x_{23} = 90$.

13.5. Задача о назначениях

Задача о назначениях — это распределительная задача, в которой для выполнения каждой работы требуется один и только один ресурс (один человек, одна автомашина и т.п.) и каждый ресурс может быть использован на одной и только одной работе, т.е. ресурсы неделимы между работами, а работы — между ресурсами.

Задача о назначениях имеет место при распределении людей на должности или работы, автомашин на маршруты, водителей на машины, групп по аудиториям, научных тем по научно-исследовательским лабораториям и т.п.

Исходные параметры задачи о назначениях:

- m — количество ресурсов;
- n — количество работ;
- $a_i = 1$ — единичное количество ресурса A_i, $i = 1, 2, ..., m$ (например, один работник, одно транспортное средство, одна научная тема);
- $a_j = 1$ — единичное количество работы B_j, $j = 1, 2, ..., n$ (например, одна должность, один маршрут, одна лаборатория);
- c_{ij} — характеристика качества выполнения работы B_j с помощью ресурса A_i (например, компетентность работника i при работе на должности j; время, за которое транспортное средство i перевезет груз по маршруту j; степень квалификации лаборатории i при работе над научной темой j).

Необходимо определить x_{ij} — факт назначения или неназначения ресурса A_i на работу B_j:

$$x_{ij} = \begin{cases} 0, \text{ если ресурс } i \text{ не назначен на работу } j, \\ 1, \text{ если ресурс } i \text{ назначен на работу } j, \end{cases}$$

где $F(x)$ — общая (суммарная) характеристика качества распределения ресурсов по работам.

Запишем следующую задачу линейного программирования:

$$F(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \, (\min),$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, ..., m,$$

$$\sum_{i=1}^{m} x_{ij} = 1, \quad j = 1, 2, ..., n,$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n.$$

14

ОСНОВНЫЕ ПОНЯТИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ И ГРАФИЧЕСКИЙ МЕТОД ИХ РЕШЕНИЯ

В этой главе приведены общая постановка ЗЛП, симметричная и каноническая формы ее записи, введены понятия базисных и свободных переменных и опорного плана. Для случая задач, сводимых к двум переменным, на конкретном примере дан графический метод решения ЗЛП.

14.1. Постановка общей задачи линейного программирования

В самом общем виде ЗЛП формулируется следующим образом.

Необходимо найти значения неизвестных $(x_1, x_2, ..., x_n)$, для которых *целевая функция* достигает экстремума:

$$F = \sum_{j=1}^{n} c_j x_j \rightarrow \max \ (\min),$$

при условии, что неизвестные удовлетворяют соотношениям

$$\sum_{j=1}^{n} a_{ij} x_j \{\leq, =, \geq\} b_i, \quad i = \overline{1, m},$$

$x_j \geq 0$, $j = \overline{1, n_1}$, а неизвестные $x_{n_1+1}, ..., x_n$ могут быть и отрицательными.

Обычно набор неизвестных $(x_1, x_2, ..., x_n)$ называется *планом*, коэффициенты a_{ij} называются *нормами затрат* i-го ресурса на производство единицы j-го вида продукции, c_j — *стоимостью* единицы j-го вида продукции, b_i — *запасом* i-го вида ресурса.

Замечание. Во многих ЗЛП переменные принимают целые значения. Для решения таких задач используются специальные методы целочисленного программирования [8].

14.2. Формы записи задач линейного программирования

Если в модели задачи линейного программирования целевая функция максимизируется (минимизируется), ограничения имеют знак ≤ (≥) и все неизвестные неотрицательны, то запись задачи называется *симметричной* (нормальной):

$$F = \sum_{j=1}^{n} c_j x_j (\max),$$
$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i, \; i = 1, \; 2, \; ..., \; m, \\ x_j \ge 0, \; j = i = 1, \; 2, \; ..., \; n. \end{cases} \qquad (14.1)$$

В случае когда ищется минимум целевой функции, можно перейти к задаче на отыскание максимума, изменив знак целевой функции.

Если в модели задачи ограничения имеют вид строгих равенств, то запись модели задачи называется *канонической*:

$$F = \sum_{j=1}^{n} c_j x_j \; (\max),$$
$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i, \; i = 1, \; 2, \; ..., \; m, \\ x_j \ge 0, \; j = 1, \; 2, \; ..., \; n. \end{cases} \qquad (14.2)$$

Будем считать, что $b_i \ge 0$, в противном случае можно поменять знак уравнения.

Для решения задачи удобнее пользоваться каноническим видом записи. Для перехода от симметричной к канонической форме записи вводятся *балансовые* (*фиктивные*) *переменные*. Например, неравенство вида

$$a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1$$

введением новой дополнительной (балансовой) переменной $x_{n+1} \ge 0$ можно заменить уравнением

$$a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + x_{n+1} = b_1,$$

а неравенство вида

$$a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \geq b_2$$

заменить уравнением

$$a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n - x_{n+2} = b_2$$

введением переменной $x_{n+2} \geq 0$.

В экономических моделях балансовые переменные имеют важный экономический смысл, который будет пояснен в дальнейшем.

14.3. Базисные и свободные переменные. Опорный план

В разделе линейной алгебры рассматриваются неопределенные системы линейных уравнений, которые имеют бесконечное множество решений и выделяются тем, что ранг матрицы системы меньше числа неизвестных. При решении ЗЛП система ограничений, записанная в каноническом виде, допускает множество решений. Иначе, если она имеет единственное решение или несовместна, то выбор оптимального решения отсутствует, целевая функция или однозначно определена, или ЗЛП не имеет решения.

Для неопределенных систем, имеющих множество решений, неизвестные разделяются на две части: базисные и свободные.

Базисные неизвестные системы линейных уравнений — это такие неизвестные, для которых столбцы матрицы системы, соответствующие этим неизвестным, образуют матрицу, ранг которой равен рангу всей матрицы системы вместе со столбцом свободных членов. Оставшиеся неизвестные называются *свободными*. Количество базисных переменных равно рангу системы, или числу уравнений (после исключения зависимых уравнений). Базисные неизвестные характеризуются также тем, что их можно однозначно выразить через свободные. Придавая свободным неизвестным произвольные значения, можно получить все решения системы.

Пример 14.1. Пусть система ограничений ЗЛП имеет вид:

$$\begin{cases} x_1 + 2x_2 + x_3 = 4, \\ 3x_1 + 6x_2 + x_4 = 5, \end{cases}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0.$$

Пара неизвестных (x_1, x_2) не является базисной, так как соответствующая матрица $\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$ имеет ранг 1, ее определитель равен нулю. Пара неизвестных (x_3, x_4) является базисной, соответствующая матрица $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ имеет ранг 2. Из данной системы базисные неизвестные (x_3, x_4) можно выразить через свободные неизвестные (x_1, x_2): $x_3 = 4 - x_1 - 2x_2$, $x_4 = 5 - 3x_1 - 6x_2$. Все решения системы можно получить, придавая свободным неизвестным произвольные значения. В частности, при $x_1 = 0$, $x_2 = 0$ имеем $x_3 = 4$, $x_4 = 5$. Полученное решение (0, 0, 4, 5) будет неотрицательным. Легко проверить, что базисными являются также все оставшиеся пары неизвестных: (x_1, x_3), (x_1, x_4), (x_2, x_3), (x_2, x_4).

При решении системы линейных уравнений, являющихся ограничениями ЗЛП, записанной в каноническом виде, возникают дополнительные трудности. Они вызваны тем, что нам нужны *неотрицательные* решения. Более того, эти решения при подстановке в целевую функцию должны давать *максимум* (минимум) этой функции. Эти трудности можно преодолеть, используя преобразования системы ограничений с помощью симплекс-таблиц. С помощью этих таблиц мы будем находить такие базисные неизвестные, значения которых при соответствующих нулевых свободных неизвестных составляют неотрицательное оптимальное решение. Введем одно из основных определений.

Опорным (*базисным*) *планом* (опорным решением) ЗЛП называется неотрицательное решение системы ограничений ЗЛП, составленное из значений набора базисных неизвестных при нулевых значениях соответствующих свободных неизвестных. В приведенном выше примере решение (0, 0, 4, 5) является опорным планом.

14.4. Графическое решение задачи линейного программирования, сводимой к двум переменным

Если ЗЛП содержит две неизвестные, то систему ее ограничений можно графически изобразить на плоскости и проиллюстрировать нахождение оптимального решения.

Пример 14.2. Предприятие изготавливает два вида продукции — $П_1$ и $П_2$, которая поступает в оптовую продажу. Для производства продукции

используются два вида сырья — C_1 и C_2. Максимально возможные запасы сырья в сутки составляют 9 и 13 единиц соответственно. Расход сырья на единицу продукции дан в табл. 14.1.

Таблица 14.1

Сырье	Расход сырья на единицу продукции		Запас сырья
	$П_1$	$П_2$	
C_1	2	3	9
C_2	3	2	13

Опыт работы показал, что суточный спрос на продукцию $П_1$ никогда не превышает спроса на продукцию $П_2$ более чем на одну единицу. Кроме того, известно, что спрос на продукцию $П_2$ никогда не превышает двух единиц в сутки.

Оптовые цены единицы продукции равны: 3 ден. ед. для $П_1$ и 4 ден. ед. для $П_2$. Какое количество продукции каждого вида должно производить предприятие, чтобы доход от реализации продукции был максимальным?

Составляем математическую модель.

Пусть x_1 — количество продукции $П_1$, выпускаемой предприятием в сутки, x_2 — количество продукции $П_2$.

$$F = 3x_1 + 4x_2(\max),$$

$$\begin{cases} 2x_1 + 3x_2 \le 9, \\ 3x_1 + 2x_2 \le 13, \\ x_1 - x_2 \le 1, \\ x_2 \le 2, \end{cases}$$

$$x_1 \ge 0, \quad x_2 \ge 0.$$

Для нахождения решения изобразим множество пар неизвестных (x_1, x_2) точками плоскости. Тогда множеством значений линейного неравенства $2x_1 + 3x_2 \le 9$ является полуплоскость с границей $2x_1 + 3x_2 = 9$. Для выбора нужной полуплоскости с этой границей можно взять некоторую точку, не лежащую на прямой $2x_1 + 3x_2 = 9$ (например, точку (0, 0)), и проверить, удовлетворяют ли ее координаты исходному неравенству. Если удовлетворяют, то нужная полуплоскость содержит выбранную точку, если нет — то это другая полуплоскость. Аналогично поступаем с остальными неравенствами.

На рис. 14.1 изображены прямые $2x_1 + 3x_2 = 9$, $3x_1 + 2x_2 = 13$, $x_1 - x_2 = 1$, $x_2 = 2$. Условия $x_1 \ge 0$, $x_2 \ge 0$ задают первый квадрант плоскости x_1Ox_2. Пересечением нужных полуплоскостей является пятиугольник $OABCD$.

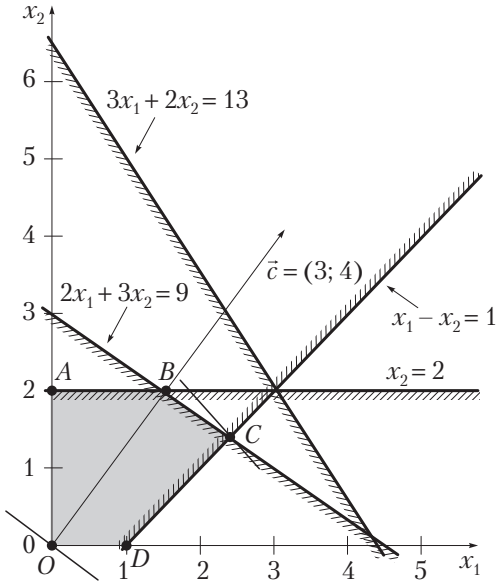


Рис. 14.1. Многоугольник решений ЗЛП

Координаты его вершин находятся как пересечение прямых. Например, точка C является пересечением прямых $2x_1 + 3x_2 = 9$ и $x_1 - x_2 = 1$. Ее координаты удовлетворяют системе

$$\begin{cases} 2x_1 + 3x_2 = 9, \\ x_1 - x_2 = 1, \end{cases}$$

решением которой являются числа $x_1 = 2{,}4$, $x_2 = 1{,}4$.

Аналогично, остальные точки имеют координаты: B ($x_1 = 1{,}5$, $x_2 = 2$); A ($x_1 = 0{,}4$, $x_2 = 2$); D ($x_1 = 1$, $x_2 = 0$).

Из всех решений нам нужно выбрать то, для которого целевая функция имеет максимальное значение. Значения целевой функции являются постоянными, равными α, на прямых $3x_1 + 4x_2 = \alpha$. Все эти прямые перпендикулярны вектору-градиенту $\vec{C}$ с координатами $\vec{C}(3, 4)$. Перемещая такую прямую в направлении вектора-градиента до пересечения с последней точкой многоугольника решений, мы найдем его вершину, которая и служит решением задачи. В нашем случае это будет точка $C(2{,}4; 1{,}4)$. В этой точке

$$F_{\max} = 3 \cdot 2{,}4 + 4 \cdot 1{,}4 = 12{,}8.$$

Приведем ответ: максимальный доход в размере 12,8 ден. ед. можно получить при производстве 2,4 ед. продукции $П_1$ и 1,4 ед. продукции $П_2$.

Геометрическая интерпретация решения ЗЛП. Графический метод решения ЗЛП позволяет дать идею решения ЗЛП, записанной в симметричном виде.

Множеством решений системы ограничений ЗЛП является многогранник в пространстве, размерность которого равна числу n неизвестных задачи. Вершины этого многогранника являются опорными решениями ЗЛП. Значения целевой функции будут постоянными на гиперплоскостях, перпендикулярных вектору-градиенту $\overline{C}(c_1, c_2, ..., c_n)$. Передвигая такую гиперплоскость в направлении вектора $\overline{C}$, мы увеличиваем значения целевой функции, а в обратном направлении — уменьшаем. Для нахождения экстремального значения целевой функции мы передвигаем указанную гиперплоскость в нужном направлении до пересечения с последней вершиной многогранника решений. Полученная вершина многогранника и дает решение ЗЛП. Идея приведенного метода используется для решения ЗЛП симплекс-методом.

Задачи с решениями

Задача 14.1. Решить графическим методом:

$$F = 2x_1 + x_2(\max),$$

$$\begin{cases} 2x_1 + 4x_2 \le 16, \\ -4x_1 + 2x_2 \le 8, \\ x_1 + 3x_2 \ge 9, \end{cases}$$

$$x_1 \ge 0,\ x_2 \ge 0.$$

Решение. В условии задачи дана математическая модель. Для нахождения решения изобразим множество пар неизвестных (x_1, x_2) точками плоскости.

Множеством значений линейного неравенства $2x_1 + 4x_2 \le 16$ является полуплоскость прямой (1). Для выбора нужной полуплоскости с этой границей возьмем точку (0,0) и проверим, что ее координаты удовлетворяют исходному неравенству [на рисунке указано стрелочкой, идущей от прямой (1) к точке (0,0)]. Аналогично поступаем с неравенством $-4x_1 + 2x_2 \le 8$ — прямая (2) и неравен-

ством $7x_1 + 5x_2 \geq 32$ — прямая (3). Пересечением нужных полуплоскостей является закрашенный треугольник.

Выбираем решение, для которого целевая функция имеет максимальное значение. Строим вектор-градиент $\bar{C}$ (2, 1). Прямые $2x_1 + x_2 = \alpha$ перпендикулярны вектору-градиенту $\bar{C}$. Перемещая такую прямую в направлении вектора $\bar{C}$ до пересечения с последней точкой треугольника решений, найдем точку max, которая и служит решением задачи.

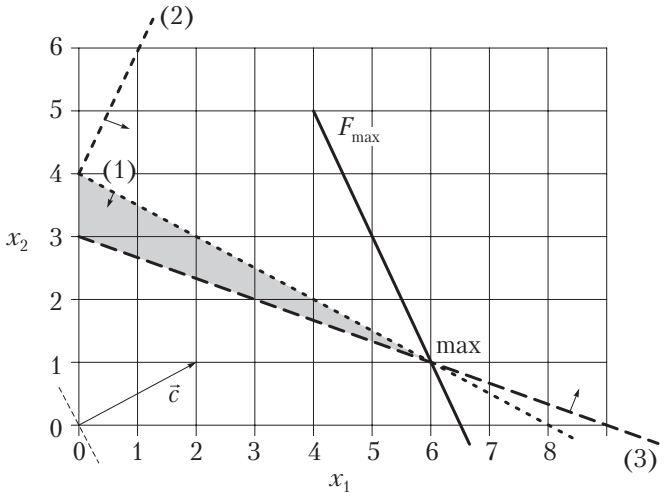
Для нахождения координат точки решим систему

$$\begin{cases} 2x_1 + 4x_2 = 16, \\ x_1 + 3x_2 = 9, \end{cases}$$

решением которой являются числа $x_1 = 6$; $x_2 = 1$.

Подставим координаты точки в целевую функцию и получим оптимальное решение задачи

$$F_{\max} = 2 \cdot 6 + 1 \cdot 1 = 13.$$



Ответ: $X^*(6, 1)$, $F_{\max} = 13$.

Задача 14.2. Решить графическим методом ЗЛП со следующей моделью:

$$F = 2x_1 - 4x_2(\min),$$
$$\begin{cases} 7x_1 + 5x_2 \geq 32, \\ 2x_1 + 7x_2 \geq 14, \\ -x_1 + 6x_2 \geq 0, \\ x_2 \leq 5, \\ x_1 \leq 8, \end{cases}$$
$$x_1 \geq 0,\ x_2 \geq 0.$$

Решение. В условии задачи дана математическая модель. Для нахождения оптимального решения изобразим множество решений в виде многоугольника планов.

Множеством значений линейного неравенства $7x_1 + 5x_2 \geq 32$ является полуплоскость прямой (1). Для выбора полуплоскости подставим точку O с координатами (0, 0) и проверим, выполняется ли условие неравенства. Условие $0 < 32$ не выполнено, поэтому нужная полуплоскость направлена в противоположную сторону. Аналогично поступаем с неравенствами $2x_1 + 7x_2 \geq 14$ (2) и $-x_1 + 6x_2 \geq 0$ (3).

Для построения прямой $x_2 \leq 5$ (4) проводим прямую, параллельную оси Ox_1 через точку (0, 5), а для построения прямой $x_1 \leq 8$ (5) — прямую, параллельную оси Ox_2, через точку (8, 0). Для прямых (4) и (5) нужные нам полуплоскости будут направлены в сторону начала координат. Получили заштрихованный многоугольник, который и является множеством решений.

Строим вектор-градиент $\vec{C}$ (2,–4). Для определения минимального значения целевой функции строим вектор-антиградиент $-\vec{C}$ (–2, 4). Перемещая перпендикулярную прямую в направлении вектора $-\vec{C}$ до пересечения с последней точкой четырехугольника решений найдем точку min, которая и служит решением задачи.

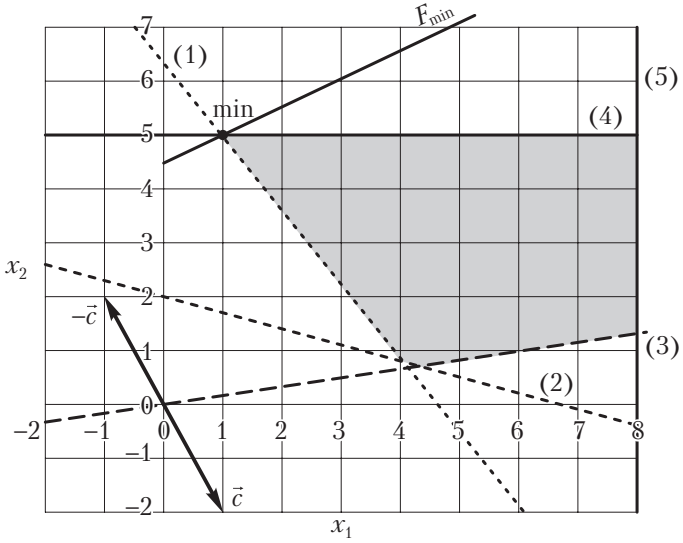
Для нахождения координат точки min решим систему из двух уравнений

$$\begin{cases} 7x_1 + 5x_2 = 32, \\ x_2 = 5, \end{cases}$$

решением которой являются числа $x_1 = 1$, $x_2 = 5$.

Подставим координаты точки в целевую функцию и получим оптимальное решение задачи

$$F_{\min} = 2 \cdot 1 - 4 \cdot 5 = -18.$$



Ответ: $X^*(1, 5)$, $F_{\min} = -18$.

Замечание. Графический метод решения ЗЛП применим также в случае, когда система ограничений содержит n неизвестных, а ранг системы равен $n - 2$. Тогда $n - 2$ базисные переменные можно выразить через две свободные переменные. Подставляя полученные выражения переменных через свободные в систему ограничений, мы получим систему с двумя переменными, которую можно решать графическим методом.

Задачи для самостоятельного решения

14.1. Привести к канонической форме записи следующие ЗЛП:

а) $F = 2x_1 + x_2(\max)$,
$\begin{cases} x_1 + x_2 \geq 1, \\ -2x_1 + 3x_2 \leq 16, \\ 4x_1 + x_2 \leq 12, \end{cases}$
$x_1 \geq 0,\ x_2 \geq 0.$

б) $F = x_1 - 3x_2 + 2x_3(\min)$,
$\begin{cases} 2x_1 - x_3 \leq 2, \\ x_1 + 2x_2 + x_3 \geq 2, \end{cases}$
$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0.$

Указание. При решении задачи обратитесь к § 14.2.

14.2. Привести к симметричной форме записи ЗЛП:

$$F = 7x_1 + x_2 - x_3 + x_4 - 9(\max),$$
$$\begin{cases} -x_1 + 2x_2 + x_3 = 6, \\ x_1 + x_2 + x_4 = 9, \\ 3x_1 - x_2 + x_5 = 12, \end{cases}$$
$$x_j \geq 0,\ j = 1,\ \ldots,\ 5.$$

Указание. При решении задачи обратитесь к § 14.2.

14.3. Решить графически следующие задачи:

а) $F = 2x_1 + 3x_2(\max)$,
$\begin{cases} x_1 + 4x_2 \geq 8, \\ x_2 \leq 5, \\ x_1 \leq 4, \end{cases}$
$x_j \geq 0,\ j = 1,\ 2.$

в) $F = 2x_1 + x_2(\min)$,
$\begin{cases} -x_1 + 4x_2 \leq 8, \\ 2x_1 - x_2 \leq 4, \\ x_1 + x_2 \geq 1, \end{cases}$
$x_1 \geq 0,\ x_2 \geq 0.$

б) $F = x_1 + 3x_2(\max)$,
$\begin{cases} 4x_1 + 7x_2 \leq 55, \\ 11x_1 + 12x_2 \leq 132, \\ -2x_1 + 3x_2 \leq 5, \end{cases}$
$x_j \geq 0,\ j = 1,\ 2.$

14.4. Торговое предприятие реализует две группы товаров — А и Б. Нормы затрат ресурсов на каждый тип товаров, лимиты ресурсов, а также доход на единицу каждой продукции заданы в табл. 14.2.

Таблица 14.2

Виды ресурсов	Норма затрат ресурсов на единицу товара		Лимит ресурсов
	Товары группы А	Товары группы Б	
Рабочее время продавцов, чел.-ч	1	3	24
Площадь торговых залов, $м^2$	0,5	0,5	8
Площадь складских помещений, $м^2$	3	1	32
Накладные расходы, ден. ед.	5	4	75
Доход на единицу продукции, ден. ед.	4	3	

Графическим методом определить плановый объем продаж и структуру товарооборота так, чтобы доход торгового предприятия был максимален.

14.5. Цех изготавливает изделия А и Б. Расход сырья, его запас и прибыль от реализации каждого изделия указаны в табл. 14.3.

Таблица 14.3

Вид сырья	Расход на изделие		Запас
	А	Б	
C_1	16	4	200
C_2	8	7	280
C_3	5	9	450
Прибыль	4	6	

Определить план производства изделий, обеспечивающий предприятию максимальную прибыль от их реализации.

14.6. Цех изготавливает изделия А и Б. Расход сырья, его запас и прибыль от реализации каждого изделия указаны в табл. 14.4.

Таблица 14.4

Вид сырья	Расход на изделие		Запас
	А	Б	
C_1	1	1	6
C_2	0	1	4
C_3	2	1	10
Прибыль	1	2	

Определить план производства изделий, обеспечивающий предприятию максимальную прибыль от их реализации.

14.7. Намечается выпуск двух видов костюмов — мужских и женских. На женский костюм требуется 1 м шерсти, 2 м лавсана и 1 чел.-дн. трудозатрат, на мужской — 3,5 м шерсти, 0,5 м лавсана и 1 чел.-дн. трудозатрат. Всего имеется 350 м шерсти, 240 м лавсана 150 чел.-дн. трудозатрат.

Определить, сколько костюмов каждого вида необходимо сшить, чтобы обеспечить максимальную прибыль, если прибыль от реализации женского костюма составляет 10 ден. ед., а от мужского — 20 ден. ед. При этом следует иметь ввиду, что необходимо сшить не менее 60 мужских костюмов.

Ответы

14.3. а) $X^*(4, 5)$, $f_{max} = 23$; б) $X^*(5, 5)$, $f_{max} = 20$; в) $X^*(0, 1)$, $f_{min} = 1$.
14.4. Объем продаж товара А — 9 ед., товара Б — 5 ед., при этом максимальный доход — 51 ден. ед.
14.5. Изделие А невыгодно производить, а изделие В необходимо производить в количестве 40 ед., при этом максимальная прибыль — 240 ден. ед.
14.6. Для получения максимальной прибыли — 10 ден. ед. необходимо произвести 2 единицы изделия А и 4 единицы изделия Б.
14.7. Для получения максимальной прибыли — 2300 ден. ед. необходимо сшить 70 женских и 80 мужских костюмов.

15

СИМПЛЕКСНЫЙ МЕТОД. ТЕОРИЯ ДВОЙСТВЕННОСТИ

Для решения ЗЛП создан поэтапный метод, называемый симплексным. Приведем в общем виде описание этапов этого метода, а затем на конкретных примерах подробно рассмотрим все этапы и особенности решения.

15.1. Симплексный метод решения ЗЛП

В § 14.3 введено важное понятие опорного плана. При его построении мы пока не учитывали соответствующее значение целевой функции. Основываясь на геометрических свойствах множества допустимых решений ЗЛП, можно показать, что целевая функция достигает оптимального значения при тех значениях неизвестных, которые составляют опорный план. Таким образом, оптимальное решение ЗЛП можно получить перебором всех опорных планов задачи. Эта задача, как правило, трудоемка. Для ее упрощения создан *симплекс-метод* — это метод целенаправленного перебора опорных планов для нахождения среди них оптимального. Алгоритм этого метода включает следующие этапы.

1. Модель задачи записывается в каноническом виде с нахождением максимума целевой функции.
2. Находится какой-либо начальный опорный план.
3. Опорный план проверяется на оптимальность.
4. Если опорный план не оптимальный, осуществляется переход к следующему опорному плану, не уменьшающему значение целевой функции, и опять проводится проверка на оптимальность.

Процедура продолжается до получения оптимального опорного решения ЗЛП или условий отсутствия такового.

Пример 15.1. Найти опорный план ЗЛП и выразить целевую функцию через свободные переменные.

$$F = -x_1 + 2x_2 \ (\max),$$

$$\begin{cases} 2x_1 - x_2 \geq 1, \\ -x_1 + 4x_2 \leq 24, \\ x_1 - x_2 \leq 3, \end{cases}$$

$$x_1 \geq 0, \quad x_2 \geq 0.$$

1. Запись системы в каноническом виде.

Для этого вводим балансовые переменные. При этом свободные члены (СЧ) уравнений должны быть неотрицательными. Если это не так, умножаем соответствующее уравнение на –1.

В приведенном примере вводим балансовые переменные (x_3, x_4, x_5) и записываем ограничения в каноническом виде:

$$F = -x_1 + 2x_2 \ (\max),$$

$$\begin{cases} 2x_1 - x_2 - x_3 = 1, \\ -x_1 + 4x_2 + x_4 = 24, \\ x_1 - x_2 + x_5 = 3, \end{cases}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0.$$

2. Запись ЗЛП в виде симплексной таблицы.

Записываем ЗЛП в виде таблицы (табл. 15.1), где каждая строка соответствует одному из уравнений системы, а последняя строка — целевой функции F.

Таблица 15.1

		↓			
БП	1 (СЧ)	$-x_1$	$-x_2$	$-x_3$	q
$\rightarrow 0 =$	1	$\boxed{2}$	–1	–1	$\frac{1}{2}$
$x_4 =$	24	–1	4	0	—
$x_5 =$	3	1	–1	0	3
$F =$	0	1	–2	0	

Если какая-либо переменная входит только в одно из уравнений системы, а ее выражение через остальные переменные имеет только положительный свободный член (СЧ), то такую переменную включают в число базисных (БП). В данном примере такими переменными будут x_4 и x_5. Если же в каком-либо уравнении системы такой переменной нет, то в столбце БП вместо переменной пишут 0, и выражают ее через остальные переменные с положительным свободным членом. В данном примере таким будет первое уравнение.

Чтобы получить из таблицы первое уравнение системы, переписанное в виде $0 = 1 - 2x_1 + x_2 + x_3$, нужно элементы первой строки умножить на элементы заглавной строки и просуммировать. Так же получают остальные уравнения, включая целевую функцию. В последнем столбце таблицы выписаны симплексные отношения, которые будут пояснены ниже.

Теперь будем преобразовывать таблицу по «правилу прямоугольника»: из произведения преобразуемого и разрешающего элементов вычитается произведение двух элементов, один из которых расположен в разрешающей строке и столбце преобразуемого элемента, второй — в разрешающем столбце и строке преобразуемого элемента. Полученная разность делится на разрешающий элемент и записывается вместо преобразуемого элемента. Эти преобразования соответствуют элементарным преобразованиям уравнений системы и не изменяют ее решений. Преобразованиями таблицы будем заменять нули столбца БП переменными. Выбор соответствующей переменной производится на основе выбора разрешающего элемента.

3. Выбор разрешающего элемента.

В число БП можно перевести любую переменную, в столбце которой имеются положительные элементы (без учета строки целевой функции). Выбранный столбец назовем *разрешающим*. Если в этом столбце имеется несколько положительных элементов, то необходим выбор разрешающей строки. Для ее выбора составляем *симплексные отношения* элементов столбца СЧ к соответствующим *положительным* элементам разрешающего столбца. В качестве разрешающей выбирается строка с *наименьшим* симплексным отношением. Разрешающий элемент расположен на пересечении разрешающих столбца и строки.

В нашем примере в качестве разрешающего столбца можно выбрать первый. Этот выбор отмечен стрелочкой над столбцом. Поскольку симплексное отношение, равное $\frac{1}{2}$, будет меньшим для первой строки, выбираем ее в качестве разрешающей. Отмечаем разрешающий элемент квадратиком.

4. Преобразование симплекс-таблицы.

Переменную из разрешающего столбца записываем в разрешающую строку столбца БП вместо нуля. Элементы разрешающего столбца исключаем из рассмотрения. Элементы разрешающей строки делим на разрешающий элемент. Все оставшиеся элементы, включая элементы столбца СЧ и целевой строки, преобразуются по «правилу прямоугольника».

В нашем примере в результате преобразований получаем табл. 15.2.

Таблица 15.2

↓

БП	1 (СЧ)	$-x_2$	$-x_3$	q
$x_1 =$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	—
$\rightarrow x_4 =$	$\frac{49}{2}$	$\boxed{\frac{7}{2}}$	$-\frac{1}{2}$	7
$x_5 =$	$\frac{5}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	—
$F =$	$-\frac{1}{2}$	$-\frac{3}{2}$	$\frac{1}{2}$	

Таким образом, мы получили следующие выражения базисных неизвестных (x_1, x_4, x_5) и целевой функции F через свободные неизвестные (x_2, x_3):

$$F = -\frac{1}{2} - \frac{3}{2}x_2 + \frac{1}{2}x_3,$$

$$\begin{cases} x_1 = \frac{1}{2} + \frac{1}{2}x_2 + \frac{1}{2}x_3, \\ x_4 = \frac{49}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3, \\ x_5 = \frac{5}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3. \end{cases}$$

Соответствующее начальное опорное решение получается при нулевых значениях свободных переменных $x_2 = 0$, $x_3 = 0$ и равно $\left(\frac{1}{2}, 0, 0, \frac{49}{2}, \frac{5}{2}\right)$. Значения базисных переменных и целевой функции появляются в столбце СЧ. В общем случае на первом этапе заменяем все нули столбца БП переменными и получаем разбиение переменных на базисные, которые будут записаны в столбце БП, и свободные. При этом появляется начальное опорное решение.

При построении начального опорного решения, выбирая разрешающий элемент мы не заботились о значении целевой функции. На следующем этапе после проверки опорного решения на оптимальность находится новое разбиение переменных на базисные и свободные, для которого соответствующее опорное решение увеличивает (не уменьшает) значение целевой функции.

5. Проверка опорного решения на оптимальность целевой функции и переход к другому опорному решению, увеличивающему значение целевой функции.

Сначала заметим, что если в целевую функцию какая-либо свободная переменная входит с положительным коэффициентом (в целевой строке таблицы этот коэффициент отрицателен), то при увеличении значений этой переменной целевая функция увеличивается. Поэтому нулевое значение этой свободной переменной, входящее в опорный план, не дает максимума целевой функции.

Таким образом, если в строке целевой функции симплекс-таблицы (без учета столбца СЧ) имеются отрицательные элементы, то опорный план неоптимален. Поэтому мы будем переводить в число базисных те свободные переменные, у которых в строке целевой функции симплекс-таблицы коэффициенты отрицательны.

В нашем примере в последней симплекс-таблице в *F*-строке имеется отрицательный элемент $-\frac{3}{2}$. Поэтому в число базисных переменных будем переводить переменную x_2. Это можно сделать, если в столбце x_2 имеются положительные элементы. В нашем случае только одно число положительное.

Итак, при переходе от одного опорного плана к более предпочтительному надо снова делать выбор разрешающего элемента в симплекс-таблице. В качестве разрешающего столбца выбирается тот, в котором элемент *F*-строки отрицателен и имеются положительные коэффициенты. Разрешающая строка выбирается по минимальному значению симплексного отношения элементов столбца СЧ к положительным элементам разрешающего столбца. В этом случае преобразования симплекс-таблицы отличаются только тем, что разрешающий столбец не исключается из рассмотрения, а преобразуется по следующему правилу: разрешающий элемент заменяется обратным ему числом, а все остальные элементы столбца делятся на разрешающий элемент и меняют знак на противоположны.

В качестве следующего разрешающего элемента берем элемент $\frac{7}{2}$, обведенный квадратиком. Дополним правила преобразования симплекс-таблицы двумя пунктами, после которых они будут иметь следующий вид:

1) сам разрешающий элемент заменяется обратным ему числом;

2) остальные элементы разрешающей строки делятся на разрешающий элемент;

3) остальные элементы разрешающего столбца делятся на разрешающий элемент и меняют знак на противоположный;

4) все остальные элементы симплексной таблицы преобразуются по «правилу прямоугольника».

По указанным правилам преобразуем симплекс-таблицу 15.2 в следующую таблицу (табл. 15.3):

Таблица 15.3

БП	1 (СЧ)	$-x_4$	$-x_3$
$x_1 =$	4	$\frac{1}{7}$	$-\frac{4}{7}$
$x_2 =$	7	$\frac{2}{7}$	$-\frac{1}{7}$
$x_5 =$	6	$\frac{1}{7}$	$\frac{3}{7}$
$F =$	10	$\frac{3}{7}$	$\frac{2}{7}$

В строке целевой функции все элементы — положительные, значит, полученное опорное решение (4, 7, 0, 0, 6) оптимально, для него значение $F_{\max} = 10$.

Пример 15.2. Найти оптимальное решение задачи

$$F = x_1 - 3x_2 - 2x_3 - x_4 + 5 \ (\max),$$

$$\begin{cases} 4x_1 - 2x_2 - 2x_3 + 3x_4 = 2, \\ 2x_1 - x_2 + x_3 - x_4 = 3, \\ 2x_1 - x_2 + 5x_3 - 6x_4 = 1, \end{cases}$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0.$$

Модель задачи записана в каноническом виде. Ее симплекс-таблица имеет следующий вид (табл. 15.4).

При построении начального опорного решения в столбец БП записываем нули. Переменные, которые можно принять в качестве базисных, вписываем в этот столбец с учетом коэффициентов при них в уравнениях.

Таблица 15.4

		↓				
БП	1 (СЧ)	$-x_1$	$-x_2$	$-x_3$	$-x_4$	q
0 =	2	4	–2	–2	3	$\frac{1}{2}$
0 =	3	2	–1	1	–1	$\frac{3}{2}$
→ 0 =	1	$\boxed{3}$	–1	5	–6	$\frac{1}{3}$
F =	5	–1	3	2	1	–

Преобразованиями таблицы будем переводить переменные из разрешающих столбцов в столбец БП. Переводим в число базисных переменную x_1. Получим табл. 15.5.

Таблица 15.5

БП	1 (СЧ)	$-x_2$	$-x_3$	$-x_4$	q
0 =	0	0	–12	$\boxed{15}$	0
0 =	2	0	–4	5	$\frac{2}{5}$
x_1 =	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{5}{2}$	–3	–
F =	$\frac{11}{2}$	$\frac{5}{2}$	$\frac{9}{2}$	–2	–

Продолжаем преобразовывать нашу симплекс-таблицу, вводя в базис переменную x_4. Получаем табл. 15.6.

Таблица 15.6

БП	1 (СЧ)	$-x_2$	$-x_3$
x_4 =	0	0	$-\frac{4}{5}$
0 =	2	0	0
x_1 =	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{10}$
F =	$\frac{11}{2}$	$\frac{5}{2}$	$\frac{29}{10}$

Во второй строке получили противоречие $0 = 2 - 0x_2 - 0x_3$. Следовательно, система ограничений несовместна и задача не имеет решения.

Замечание. Если после преобразования в симплекс-таблице появится противоречивое равенство, то задача не имеет решения. Если над отрицательным элементом целевой строки все элементы столбца отрицательны, задача также не имеет решения.

Пример 15.3. Предприятие производит три вида продукции $П_1$, $П_2$, $П_3$, используя два вида сырья $С_1$ и $С_2$. Объемы затрат сырья на производство единицы продукции $П_1$ составляют 5 и 3 ед., продукции $П_2$ — 7 и 5 ед., продукции $П_3$ — по 4 ед. каждого вида. Запас сырья $С_1$ составляет 376 ед., а запас сырья $С_2$ — 280 ед. Цена на продукцию $П_1$, $П_2$ и $П_3$ составляет соответственно — 36 ден. ед., 20 ден. ед. и 40 ден. ед. Определить оптимальную производственную программу.

Составим модель задачи, для этого введем обозначения. Пусть количество продукции $П_1$ — x_1, $П_2$ — x_2 и $П_3$ — x_3.

$$F = 36x_1 + 20x_2 + 40x_3 \ (\max),$$

$$\begin{cases} 5x_1 + 7x_2 + 4x_3 \le 376, \\ 3x_1 + 5x_2 + 4x_3 \le 280, \end{cases}$$

$$x_j \ge 0, \quad j = 1, 2, 3.$$

Итак, задача заключается в нахождении оптимального плана выпуска продукции при данных ограниченных ресурсах, обеспечивающего максимум стоимости продукции. Для построения начального опорного плана необходимо преобразовать модель задачи из симметричной формы записи в каноническую. Для этого введем в рассмотрение балансовые переменные, которые отражают возможные остатки сырья. Пусть остаток сырья первого вида будет x_4, а остаток второго вида сырья — x_5. Получим модель задачи в каноническом виде:

$$F = 36x_1 + 20x_2 + 40x_3 \ (\max),$$

$$\begin{cases} 5x_1 + 7x_2 + 4x_3 + x_4 = 376, \\ 3x_1 + 5x_2 + 4x_3 + x_5 = 280, \end{cases}$$

$$x_j \ge 0, \quad j = 1, 2, ..., 5.$$

Из приведенной записи легко видеть, что балансовые переменные x_4 и x_5 могут служить базисными переменными. Остальные переменные будут свободными. Базисные переменные (БП) однозначно выражаются через

свободные (СП). При этом правые части ограничений неотрицательны. Таким образом, план ($x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 376$, $x_5 = 280$) может служить начальным опорным планом.

Для проверки этого плана на оптимальность и перехода к лучшему плану запишем данные в следующей симплекс-таблице 15.7.

Таблица 15.7

БП	1	СП		
		$-x_1$	$-x_2$	$-x_3$
$x_4 =$	376	5	7	4
$x_5 =$	280	3	5	[4]
$F =$	0	–36	–20	–40

В табл. 15.7 каждая строка, кроме заглавной, соответствует уравнению системы ограничений. Последняя строка задает целевую функцию. Для проверки полученного опорного плана на оптимальность просматриваем последнюю целевую строку. Если в этой строке в столбцах свободных переменных все числа неотрицательны, то план оптимален. В противном случае необходимо переходить к улучшенному плану.

В результате получаем две симплексные таблицы (табл. 15.8 и 15.9). Таблица 15.9 является завершающей, поскольку в ее *F*-строке отсутствуют отрицательные числа.

Таблица 15.8

БП	1	СП		
		$-x_1$	$-x_2$	$-x_5$
$x_4 =$	96	[2]	2	–1
$x_3 =$	70	$\frac{3}{4}$	$\frac{5}{4}$	$\frac{1}{4}$
$F =$	2800	–6	30	10

Из табл. 15.9, полагая свободные переменные равными нулю, выпишем оптимальный производственный план

$$X^* = (48, 0, 34),$$

согласно которому следует производить 48 ед. продукции $П_1$ и 34 ед. продукции $П_3$, а продукцию $П_2$ производить невыгодно; при этом предприятие будет получать максимальный доход $F_{max} = 3088$ ден. ед.

Таблица 15.9

БП	1	СП		
		$-x_4$	$-x_2$	$-x_5$
$x_1 =$	48	$\frac{1}{2}$	1	$-\frac{1}{2}$
$x_3 =$	34	$-\frac{3}{8}$	$\frac{1}{2}$	$\frac{5}{8}$
$F =$	3088	3	36	7

Замечание. В приведенном примере введенные балансовые переменные могут служить базисными переменными. Более того, если в системе ограничений имеется переменная, входящая только в одно из уравнений с положительным коэффициентом при неотрицательной правой части уравнения, то ее сразу можно включить в число базисных переменных, разделив уравнение на коэффициент при данной переменной. В общем случае, когда балансовые переменные входят с отрицательным знаком, первоначально их нельзя включить в число базисных.

15.2. Двойственность в линейном программировании

На основе теории двойственности проводится анализ оптимального решения. Изменения системы ограничений ЗЛП и коэффициентов целевой функции могут приводить к качественному изменению оптимального решения. Устойчивость и чувствительность оптимального решения к таким изменениям проверяются с использованием двойственной задачи и теорем теории двойственности.

15.2.1. Понятия анализа устойчивости и чувствительности

Получение оптимального решения ЗЛП не всегда является ее завершающим этапом. В реальных задачах экономики ограничения ЗЛП и целевая функция могут изменяться вместе с изменениями запасов и потребностей. Изменения цен на сырье и продукцию влекут также изменения целевой функции. Поэтому, естественно, возникает задача исследования поведения оптимального решения при указанных изменениях. Выделяют два аспекта этой проблемы:

анализ устойчивости и анализ чувствительности оптимального решения. В линейном программировании под ***устойчивостью*** понимают неизменность оптимального решения при изменениях коэффициентов целевой функции (цен на продукцию), а под ***чувствительностью*** — степень влияния изменения правых частей ограничений (запасов ресурсов) на изменение значений целевой функции (доходность производства) [5, 8]. Мы проиллюстрируем эти понятия графически.

На рис. 15.1 изображены многоугольник *ABCDE* допустимых решений ЗЛП, линия l_1 максимального уровня целевой функции $Z = c_1x_1 + c_2x_2$ и оптимальное решение *A*. При изменении коэффициентов c_1, c_2 целевой функции соответствующая оптимальная линия уровня будет поворачиваться вокруг точки *A*, если эта точка остается оптимальной. Когда же эта линия достигнет точки *B* (прямая l_3) (или точки *E*), то весь отрезок *AB* прямой l_2 (или отрезок *AE*) состоит из оптимальных решений. При дальнейшем повороте линии уровня оптимальное решение сместится в точку *B* (оптимальная линия уровня — l_3) (или *E*). Таким образом, оптимальное решение остается в точке *A* при изменении коэффициентов c_1, c_2 в определенных пределах. Это и есть *пределы устойчивости* оптимального решения.

Иллюстрацию границ чувствительности проведем на рис. 15.2. Пусть четырехугольник *ABCD* — ОДР (область допустимых решений). Стороны этого многоугольника соответствуют неравенствам системы ограничений ЗЛП. В свою очередь, эти неравенства соот-

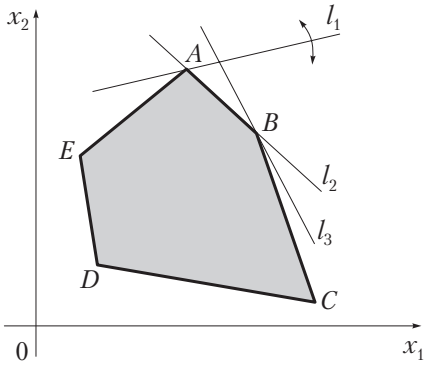


Рис. 15.1. Предел устойчивости

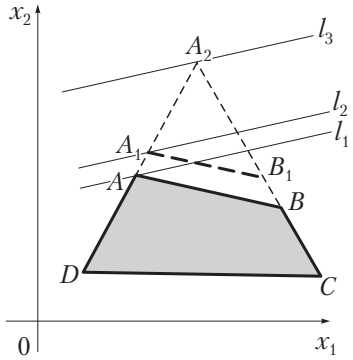


Рис. 15.2. Граница чувствительности

ветствуют ограничениям по ресурсам. Изменение запасов какого-либо ресурса влечет изменение ОДР параллельным сдвигом соответствующей стороны. Например, сторона AB на рис. 15.2 при увеличении соответствующего ресурса R занимает положение A_1B_1. При этом оптимальное решение, лежащее на прямой l_1, из точки A перемещается в точку A_1, лежащую на прямой l_2. Этот процесс можно продолжать, пока точка A не совпадет с точкой A_2, лежащей на прямой l_3. Ресурс R, изменение которого вызывает изменение значения целевой функции (уровень дохода), называется *дефицитным*. Дальнейшее увеличение запаса ресурса R выводит линию уровня l целевой функции за пределы ОДР и не влияет на уровень дохода, а этот ресурс становится бездефицитным. На рис. 15.2 видно, что незначительные изменения ресурсов, соответствующих сторонам BC и CD, не влияют на оптимальное значение целевой функции. Эти ресурсы являются бездефицитными.

Возникает вопрос: как в общем случае провести анализ решения ЗЛП и определить пределы устойчивости и границы чувствительности? Итог решения задачи симплексным методом позволяет определить дефицитные ресурсы и использовать его для расчета указанных пределов. Эта задача довольно трудоемка, ее решение основано на использовании двойственных задач линейного программирования. Для решения ЗЛП лучше использовать стандартные пакеты программ линейной оптимизации. Например, в надстройке Excel *Поиск решения* приводятся отчеты по устойчивости и по границам чувствительности.

15.2.2. Понятие двойственности

Для анализа решения ЗЛП вводится другая задача, которая называется *двойственной*, или *сопряженной*, по отношению к данной. Дадим определение двойственной задачи по отношению к исходной задаче линейного программирования, имеющей симметричную форму и состоящую в нахождении *максимального* значения функции

$$Z = c_1x_1 + c_2x_2 + ... + c_nx_n \tag{15.1}$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2, \\ ..... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m, \end{cases} \tag{15.2}$$

$$x_j \ge 0, \quad j = 1, 2, ..., n. \tag{15.3}$$

Двойственной задачей по отношению к задаче (15.1)–(15.3) называется задача (15.4)–(15.6), состоящая в нахождении минимального значения функции

$$Z' = b_1y_1 + b_2y_2 + ... + b_my_m \tag{15.4}$$

при ограничениях

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + ... + a_{m1}y_m \ge c_1, \\ a_{12}y_1 + a_{22}y_2 + ... + a_{m2}y_m \ge c_2, \\ ... \quad ... \quad ... \quad ... \\ a_{1n}y_1 + a_{2n}y_2 + ... + a_{mn}y_m \ge c_n, \end{cases} \tag{15.5}$$

$$y_j \ge 0, \quad i = 1, 2, ..., m. \tag{15.6}$$

Неизвестными переменными в этой задаче являются переменные $y_1, ..., y_m$, количество которых равно количеству используемых ресурсов и которые служат ценами единицы соответствующего ресурса. Эти переменные в экономической литературе получили различные названия: учетные, неявные, теневые, объективно обусловленные оценки, двойственные оценки или «цены» ресурсов.

В отличие от цены c_j единицы продукции, заданной заранее, цена y_i единицы ресурса находится при решении двойственной задачи.

Рассмотренная пара взаимно двойственных задач может быть экономически интерпретирована, например так.

Прямая задача: сколько и какой продукции $\bar{X} = (x_1, x_2, ..., x_n)$ надо произвести, чтобы при заданных стоимостях единицы продукции c_j, $j = 1, 2, ..., n$, объемах имеющихся ресурсов b_i, $i = 1, 2, ..., m$ и нормах расходов a_{ij} максимизировать выпуск продукции в стоимостном выражении?

Двойственная задача: какова должна быть оценка y_i, $i = 1, 2, ..., m$, единицы каждого из ресурсов, чтобы при заданных b_i, c_j и a_{ij}, $j = 1, 2, ..., n$, минимизировать общую оценку всех использованных ресурсов? Каждая из пары двойственных задач фактически является самостоятельной ЗЛП и может быть решена независимо от другой задачи.

Правила составления двойственной задачи:

1. Коэффициентами целевой функции двойственной задачи являются свободные члены ограничений прямой задачи, а свободными членами ограничений двойственной задачи — коэффициенты целевой функции прямой задачи. В двойственной задаче будет столько переменных, сколько ограничений в прямой, и столько ограничений, сколько переменных в прямой задаче. Таким образом, каждому ограничению задачи отвечает соответствующая переменная двойственной задачи, и наоборот.

2. Матрицы коэффициентов при переменных в двойственных задачах взаимно транспонированы.

3. Каждому ограничению-неравенству в двойственной задаче отвечает неотрицательная переменная, а каждому ограничению-равенству — переменная произвольного знака, и наоборот: каждой неотрицательной переменной — ограничение-неравенство, а каждой переменной произвольного знака — ограничение-равенство. При этом в задаче максимизации ограничения имеют вид $\leq$, а в задаче минимизации — вид $\geq$.

4. Если в прямой задаче целевая функция максимизируется, то в двойственной задаче — минимизируется, и наоборот.

Замечание. Для ЗЛП, запись которой не является симметричной, правила составления двойственной задачи несколько отличаются от перечисленных [5, 8].

При решении ЗЛП (15.1)–(15.3) симплексным методом мы переходим к канонической форме записи введением балансовых переменных x_{n+1}, ..., x_{n+m}. Эти переменные являются базисными, а переменные x_1, ..., x_n — свободными. Соответственно, для двойственной задачи вводятся переменные y_{m+1}, ..., y_{m+n}. Эти переменные являются базисными, а переменные y_1, ..., y_m — свободными. Таким образом, возникает взаимно-однозначное соответствие между переменными прямой и двойственной задач, которое мы запишем в виде табл. 15.10:

Таблица 15.10

Переменные исходной задачи											
Свободные						Базисные					
x_1	x_2	...	x_j	...	x_n	x_{n+1}	x_{n+2}	...	x_{n+i}	...	x_{n+m}
↕	↕		↕		↕	↕	↕		↕		↕
y_{m+1}	y_{m+2}	...	y_{m+j}	...	y_{m+n}	y_1	y_2	...	y_i	...	y_m
Базисные						Свободные					
Переменные двойственной задачи											

Пример 15.4. Сформулируем модель задачи, двойственной к задаче о производстве продукции, рассмотренной в качестве примера 15.3, и приведем оптимальные значения переменных двойственной задачи.

В исходной прямой задаче три переменные:
- x_1 — количество продукции $П_1$;
- x_2 — количество продукции $П_2$;
- x_3 — количество продукции $П_3$.

Целевая функция —

$$F(\bar{X}) = 36x_1 + 20x_2 + 40x_3 \ (\max),$$

ограничения —

$$\begin{cases} 5x_1 + 7x_2 + 4x_3 \le 376, \\ 3x_1 + 5x_2 + 4x_3 \le 280, \end{cases}$$

$$x_j \ge 0, \quad j = 1, 2, 3.$$

В двойственной задаче две переменные:
- y_1 — двойственная оценка ресурса «сырье $С_1$»;
- y_2 — двойственная оценка ресурса «сырье $С_2$».

Целевая функция —

$$Z(\bar{Y}) = 376y_1 + 280y_2 \ (\min),$$

Ограничения —

$$\begin{cases} 5y_1 + 3y_2 \geq 36, \\ 7y_1 + 5y_2 \geq 20, \\ 4y_1 + 4y_2 \geq 40, \end{cases}$$

$$y_1 \geq 0, \quad y_2 \geq 0.$$

Из решения прямой задачи с помощью симплексных таблиц можно выписать решение двойственной задачи. Укажем сначала соответствие между переменными согласно табл. 15.10:

$$\begin{array}{ccccc} \overbrace{\qquad\qquad}^{\text{СП}} & & & \overbrace{\qquad}^{\text{БП}} & \\ x_1 & x_2 & x_3 & x_4 & x_5 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ \underbrace{y_3 \quad y_4 \quad y_5}_{\text{БП}} & & & \underbrace{y_1 \quad y_2}_{\text{СП}} & \end{array}$$

Для записи оптимального решения двойственной задачи воспользуемся симплексной таблицей (см. табл. 15.3) решения прямой задачи.

Значения переменных оптимального решения двойственной задачи находятся в соответствующих клетках последней строки этой таблицы:

$$\begin{cases} y_1 \leftrightarrow x_4 \Rightarrow y_1^* = 3, \\ y_2 \leftrightarrow x_5 \Rightarrow y_2^* = 7, \end{cases}$$

$$Z(\bar{Y}) = 376y_1 + 280y_2 = 376 \cdot 3 + 280 \cdot 7 = 3088.$$

Оптимальное решение канонической формы двойственной задачи имеет вид

$$\begin{cases} y_1 \leftrightarrow x_4 \Rightarrow y_1^* = 3, \\ y_2 \leftrightarrow x_5 \Rightarrow y_2^* = 7, \end{cases} \qquad \begin{array}{l} y_3 \leftrightarrow x_1 \Rightarrow y_3^* = 0, \\ y_4 \leftrightarrow x_2 \Rightarrow y_4^* = 36, \\ y_5 \leftrightarrow x_3 \Rightarrow y_5^* = 0, \end{array}$$

или

$$\bar{Y} = (y_1 = 3, y_2 = 7, y_3 = 0, y_4 = 36, y_5 = 0),$$

$$Z(\bar{Y})_{\min} = 3088 \text{ ден. ед.}$$

Значения балансовых переменных ($y_3 = 0$, $y_4 = 36$, $y_5 = 0$) этого решения имеют экономический смысл (подробнее см. ниже).

15.3. Основные теоремы теории двойственности и их экономический смысл

Связь между решениями прямой и двойственной задач устанавливается в следующей теореме.

Первая теорема двойственности:

1. Если одна из двойственных задач имеет оптимальный план, то и другая разрешима, т.е. имеет оптимальный план. При этом экстремальные значения целевых функций совпадают:

$$F_{\max} = Z_{\min}.$$

2. Если в одной из задач целевая функция не ограничена на множестве планов, то в другой задаче система ограничений противоречива.

Экономический смысл первой (основной) теоремы двойственности: план производства $\bar{X} = (x_1, x_2, ..., x_n)$ и набор цен (оценок) ресурсов $\bar{Y} = (y_1, y_2, ..., y_m)$ оказываются оптимальными тогда и только тогда, когда прибыль (выручка) от продукции, найденная при «внешних» (известных заранее) ценах c_1, c_2, ..., c_n, равна затратам на ресурсы по «внутренним» (определяемым только из решения задачи) ценам y_1, y_2, ..., y_m.

Кроме этого, для всех других планов $\bar{X}$ и $\bar{Y}$ обеих задач прибыль (выручка) от продукции всегда меньше (или равна) затрат на ресурсы $F(\bar{X}) \le Z(\bar{Y})$.

Экономическое содержание первой теоремы двойственности состоит также и в следующем: если задача определения оптимального плана, максимизирующего выпуск продукции, разрешима, то разрешима и задача определения оценок ресурсов. При этом цена продукта, полученного в результате реализации оптимального плана, совпадает с суммарной оценкой ресурсов. Совпадения значений целевых функций для соответствующих решений пары двойственных задач достаточно для того, чтобы эти решения были оптимальными.

Свойство I: Оценки y_i^* выступают как инструмент балансирования затрат и результатов.

Двойственные оценки обладают тем свойством, что они гарантируют рентабельность оптимального плана, т.е. равенство общей стоимости продукции и оценок ресурсов обусловливает убыточность всякого другого плана, отличного от оптимального. Двойственные оценки позволяют сопоставлять и балансировать затраты и результаты производства.

Пример 15.5. Продолжим рассмотрение примера 15.4. Для него пара двойственных задач имеет вид

$$F(\bar{X}) = 36x_1 + 20x_2 + 40x_3 \ (\max),$$

$$\begin{cases} 5x_1 + 7x_2 + 4x_3 + x_4 = 376, \\ 3x_1 + 5x_2 + 4x_3 + x_5 = 280, \end{cases}$$

$$x_j \geq 0, \quad j = 1, ..., 5;$$

$$Z(\bar{Y}) = 376y_1 + 280y_2 \ (\min),$$

$$\begin{cases} 5y_1 + 3y_2 - y_3 = 36, \\ 7y_1 + 5y_2 - y_4 = 20, \\ 4y_1 + 4y_2 - y_5 = 40, \end{cases}$$

$$y_i \geq 0, \quad i = 1, ..., 5,$$

где стоимость готовой продукции в прямой задаче

$$F(\bar{X}) = 36x_1 + 20x_2 + 30x_3 = 36 \cdot 48 + 20 \cdot 0 + 30 \cdot 34 = 3088,$$

в двойственной — затраты

$$Z(\bar{Y}) = 376y_1 + 280y_2 = 376 \cdot 3 + 280 \cdot 7 = 3088,$$

т.е. имеется совпадение оптимальных значений целевых функций.

С экономической точки зрения первая теорема означает, что по оптимальному плану выпуска продукции все затраты внутри производства совпадают с оценкой готовой продукции, произведенной по этому плану, т.е. при оптимальном плане вся стоимость затрат внутри производства поглощается в стоимости готовой продукции.

Вторая теорема двойственности (теорема о дополняющей нежесткости): пусть $\bar{X} = (x_1, x_2, ..., x_n)$ — допустимое решение прямой задачи, а $\bar{Y} = (y_1, y_2, ..., y_m)$ — допустимое решение двойственной задачи. Для того чтобы они были оптимальными решениями

соответственно прямой и двойственной задач, необходимо и достаточно, чтобы выполнялись следующие соотношения:

$$x_j^*\left(\sum_{i=1}^{m} a_{ij}y_i^* - c_j\right) = 0,\ j = 1, 2, ..., n, \qquad (W)$$

$$y_i^*\left(\sum_{j=1}^{n} a_{ij}x_j^* - b_i\right) = 0,\ i = 1, 2, ..., m, \qquad (V)$$

которые называются *условиями дополняющей нежесткости*. Из них следует, что если какое-либо ограничение одной из задач ее оптимальным планом обращается в строгое неравенство, то соответствующая компонента оптимального плана двойственной задачи должна равняться нулю; если же какая-либо компонента оптимального плана одной из задач положительна, то соответствующее ограничение в двойственной задаче ее оптимальным планом должно обращаться в строгое равенство.

Таким образом,

- если $\sum_{i=1}^{m} a_{ij}y_i^* > c_j$, то $x_j^* = 0$,
- если $x_j^* > 0$, то $\sum_{i=1}^{m} a_{ij}y_i^* = c_j$.

Точно так же,

- если $\sum_{j=1}^{n} a_{ij}x_j^* < b_i$, то $y_i^* = 0$,
- если $y_i^* > 0$, то $\sum_{j=1}^{n} a_{ij}x_j^* = b_i$.

Экономически это означает, что если по некоторому оптимальному плану производства расход i-го ресурса строго меньше его запаса b_i, то в оптимальном плане соответствующая двойственная оценка единицы этого ресурса равна нулю. Если же в некотором оптимальном плане оценок его i-я компонента строго больше нуля, то в оптимальном плане производства расход соответствующего ресурса равен его запасу. Отсюда следует *вывод*: двойственные

оценки могут служить мерой дефицитности ресурсов. Дефицитный ресурс (полностью используемый по оптимальному плану производства) имеет положительную оценку, а ресурс избыточный (используемый не полностью) имеет нулевую оценку.

Свойство II: оценки позволяют установить целесообразность выпуска тех или иных видов продукции, т.е. являются мерой убыточности при производстве невыгодных видов продукции.

Пример 15.6. Рассмотрим свойство II для примера 15.5.

1. Если $x_j^* > 0$, то $\sum_{i=1}^{m} a_{ij} y_i^* = c_j$

$$X^* = (48;\ 0;\ 34;\ 0;\ 0),\quad \begin{matrix} x_1^* = 48 > 0, \\ x_2^* = 0 = 0, \\ x_3^* = 34 > 0; \end{matrix}$$

$$\begin{cases} 5y_1^* + 3y_2^* = 5 \cdot 3 + 3 \cdot 7 = 36 = 36, \\ 7y_1^* + 5y_2^* = 7 \cdot 3 + 5 \cdot 7 = 56 > 20, \\ 4y_1^* + 4y_2^* = 4 \cdot 3 + 4 \cdot 7 = 40 = 40, \end{cases}$$

т.е. выпуск по первому и третьему видам продукции оправдан.

$\sum_{i=1}^{m} a_{ij} y_i^* > c_j \Rightarrow x_j^* = 0$, $(W) \Rightarrow$ затраты выше, чем стоимость $\Rightarrow$

$\Rightarrow$ второй вид продукции невыгодно выпускать.

В условиях нашего примера соотношения (V), (W) интерпретируются так: в оптимальный план войдут только те виды продукции, затраты на которые внутри производства совпадут со стоимостью готовой продукции, и не войдут те виды, затраты на которые превышают стоимость готовой продукции.

Свойство III: оценки позволяют судить о степени дефицитности ресурсов: чем выше оценка, тем дефицитней ресурс.

Пример 15.7. Рассмотрим свойство III для примера 15.5.

2. Если $y_i^* > 0$, то $\sum_{j=1}^{n} a_{ij} x_j^* = b_i$

$$Y^* = (3; 7; 0; 36; 0),\quad \begin{matrix} y_1^* = 3 > 0, \\ y_2^* = 7 > 0; \end{matrix}$$

$$\begin{cases} 5x_1 + 7x_2 + 4x_3 = 5 \cdot 48 + 7 \cdot 0 + 4 \cdot 34 = 376, \\ 3x_1 + 5x_2 + 4x_3 = 3 \cdot 48 + 5 \cdot 0 + 4 \cdot 34 = 280, \end{cases}$$

т.е. оба ресурса израсходованы полностью.

В условиях нашего примера соотношения (V), (W) интерпретируются следующим образом: если ресурс получил положительную оценку, то он считается дефицитным и весь будет израсходован при реализации оптимального плана.

Теорема об оценках: значения переменных y_i^* в оптимальном решении двойственной задачи представляют собой оценки влияния свободных членов b_i системы ограничений (неравенств) прямой задачи на величину изменения значения целевой функции

$$\Delta F(\bar{X}) = \Delta b_i y_i^*.$$

Она справедлива для канонической формы записи.

Экономически оценки показывают, как изменяется экстремальное значение функции в зависимости от изменения правых частей ограничений задачи. Если запас дефицитного ресурса изменится на единицу, т.е. $\Delta b_i = 1$, то оценки показывают, насколько изменится максимальная выручка предприятия $y_i^* \approx \Delta F\left(\bar{X}\right)$.

Свойство IV: Оценка y_i^* является мерой влияния i-го ограничения задачи на экстремальное значение целевой функции.

Пример 15.8. Рассмотрим свойство IV для примера 15.5.

Так как $y_1^* = 3 < y_2^* = 7$, то второй ресурс более дефицитный.

Если $y_i^* = 0$, то $\sum_{j=1}^{n} a_{ij} x_j^* < b_i$, поэтому ресурс израсходован не полностью и находится в остатке.

Замечание. Это свойство, как и все предыдущие, остается справедливым до тех пор, пока правые части ограничений b_i меняются в определенных пределах (пределах чувствительности).

15.4. Информационные технологии в решении задач математического программирования

Покажем на примере применение информационных технологий при решении и анализе задач линейного программирования, а также их экономической интерпретации.

Пример 15.9. Предприятие выпускает 4 вида продукции $П_1$, $П_2$, $П_3$, $П_4$. Для производства продукции оно располагает 3 видами ресурсов, запасы которых ограничены величинами 35, 30 и 40 ед. Цена единицы готовой продукции равна 14, 10, 14 и 11 ден. ед. соответственно. Удельные затраты на единицу продукции заданы в виде табл. 15.11.

Таблица 15.11

Ресурс	Расход ресурсов на единицу продукции			
	$П_1$	$П_2$	$П_3$	$П_4$
$Р_1$	4	2	2	3
$Р_2$	1	1	2	3
$Р_3$	3	1	2	1

Необходимо выполнить следующие задания.

I. Сформулировать экономико-математическую модель задачи для определения производственной программы предприятия на максимум общей стоимости продукции, используя данные таблицы.

II. Используя надстройку Excel *Поиск решения*, найти оптимальный план выпуска продукции, при котором общая стоимость продукции будет максимальной.

III. Сформулировать экономико-математическую модель двойственной задачи к задаче о продукции.

IV. Используя надстройку Excel *Поиск решения*, найти оптимальный план двойственной задачи.

V. Используя протоколы *Поиска решения*, выполнить анализ полученного оптимального решения исходной задачи.

VI. Пояснить равенство нулю переменных x_1 и x_4.

I. Сформулируем экономико-математическую модель задачи.

Переменные: x_1, x_2, x_3, x_4 — количество продукции каждого вида.

Целевая функция:

$$F(\bar{X}) = 14x_1 + 10x_2 + 14x_3 + 11x_4 \ (\max).$$

Ограничения по ресурсам:

$$\begin{cases} 4x_1 + 2x_2 + 2x_3 + 3x_4 \le 35, \\ x_1 + x_2 + 2x_3 + 3x_4 \le 30, \\ 3x_1 + x_2 + 2x_3 + x_4 \le 40, \end{cases}$$

$$x_j \ge 0, \quad j = 1, 2, 3, 4.$$

II. Найдем оптимальный план выпуска продукции. Решение задачи выполняем с помощью надстройки Excel *Поиск решения*.

IIа. Введем условия задачи. Ввод условий задачи состоит из следующих основных шагов:

- создание формы для ввода условий задачи;
- ввод исходных данных;
- ввод зависимостей из математической модели;
- назначение целевой функции;
- ввод ограничений и граничных условий.

Для нашей задачи сделаем форму для ввода условий задачи и введем исходные данные в форму с помощью режима формул (рис. 15.3).

Вводим зависимости для целевой функции:

1) выделяем ячейку F5;

2) щелкаем левой кнопкой мыши по кнопке *Мастер функций*, появится диалоговое окно *Мастер функций — шаг 1 из 2* (рис. 15.4);

3) в окне *Категория* выбираем *Математические*;

4) в окне *Функции* выбираем *суммпроизв*;

5) *ОК*;

6) в диалоговое окно *Аргументы функции* (рис. 15.5) вводим массив 1 — B$3:E$3, в массив 2 — B5:E5.

Заметим, что во все диалоговые окна адреса ячеек удобно вводить не с клавиатуры, а выделяя мышью требуемые ячейки.

Ввод зависимости для левых частей ограничений производим по аналогии.

На этом ввод данных в таблицы закончен.

	A	B	C	D	E	F	G	H
1		Переменные						
2	Имя	X1	X2	X3	X4			
3	Значение							
4	Нижн. гр.	0	0	0	0	ЦФ	Экстр.	
5	Коэф. в ЦФ	14	10	14	11	=СУММПРОИЗВ(B$3:E$3;B5:E5)	max	
6								
7		Ограничения				Левая часть	Знак	Правая часть
8	1-ое огран.	4	2	2	3	=СУММПРОИЗВ(B$3:E$3;B8:E8)	<=	35
9	2-ое огран.	1	1	2	3	=СУММПРОИЗВ(B$3:E$3;B9:E9)	<=	30
10	3-е огран.	3	1	2	1	=СУММПРОИЗВ(B$3:E$3;B10:E10)	<=	40

Рис. 15.3. Ввод исходных данных

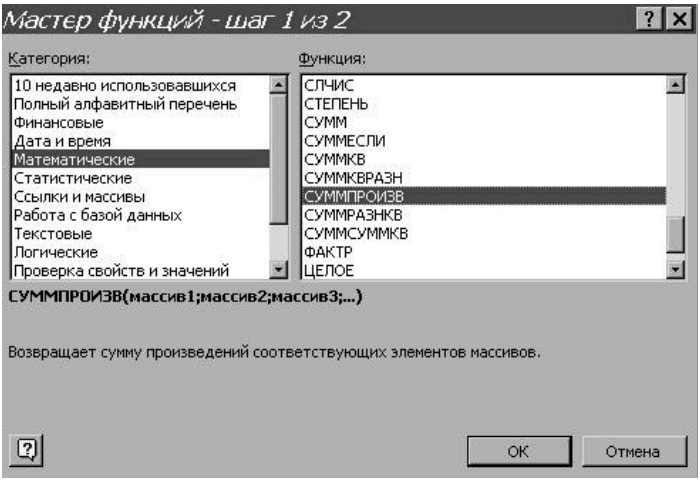


Рис. 15.4. Диалоговое окно *Мастер функций*

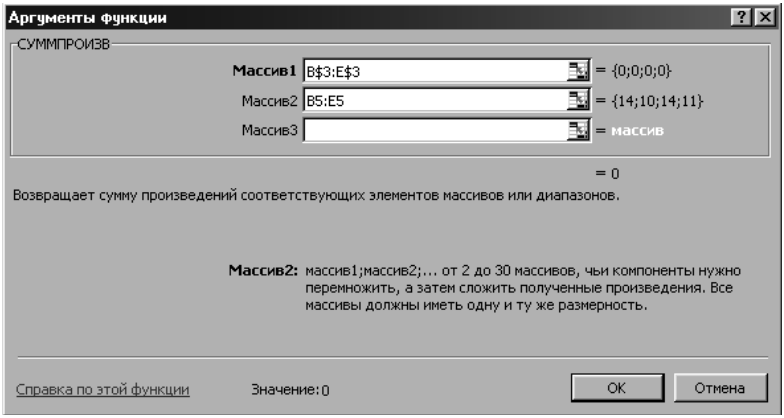


Рис. 15.5. Диалоговое окно *Аргументы функции*

Далее работаем с диалоговым окном *Поиск решения* (рис. 15.6).

1. Назначаем целевую функцию:

1) в поле *Установить целевую ячейку* вводим адрес: F$5;

2) вводим направление целевой функции: *максимальному значению*.

2. Вводим адреса искомых переменных в поле *Изменяя ячейки*: B3:E3.

3. Вводим граничные условия в поле *Ограничения*.

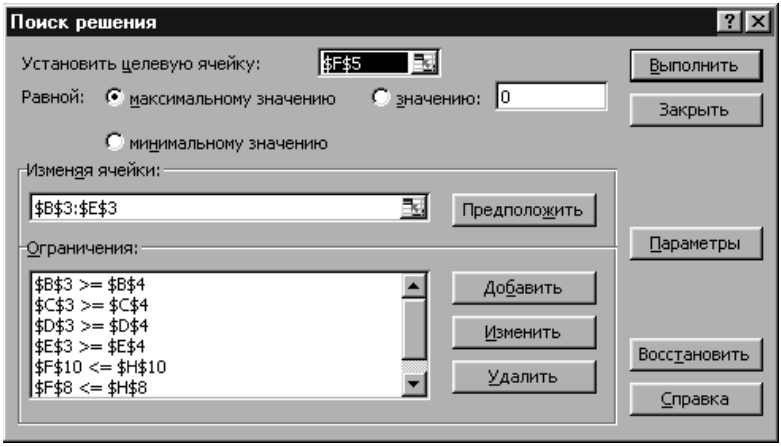


Рис. 15.6. Диалоговое окно *Поиск решения*

Для этого нажимаем кнопку *Добавить* и в появившемся диалоговом окне *Добавление ограничения* (рис. 15.7) вводим граничные условия на переменные (x_1 и x_2) ≥ 0 (B3 ≥ B4, C3 ≥ C4):

1) в поле *Ссылка на ячейку* вводим B3;

2) выбираем необходимый знак ограничения: >=;

3) в поле *Ограничение* вводим B4.

Аналогично вводим граничное условие для второй переменной.

Аналогично вводим ограничения: D9 ≤ F9, D10 ≥ F10.

После ввода последнего ограничения вместо *Добавить* нажимаем *ОК*.

На этом ввод условий задачи заканчивается.

IIб. Решение задачи.

Решение задачи производится сразу же после ввода данных в диалоговом окне *Поиск решения*. Щелчком по кнопке *Параметры* открывается диалоговое окно *Параметры поиска решения* (рис. 15.8).

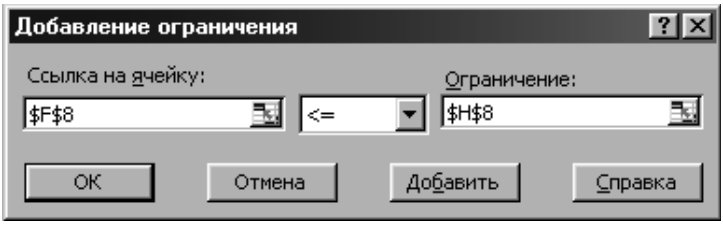


Рис. 15.7. Диалоговое окно *Добавление ограничения*

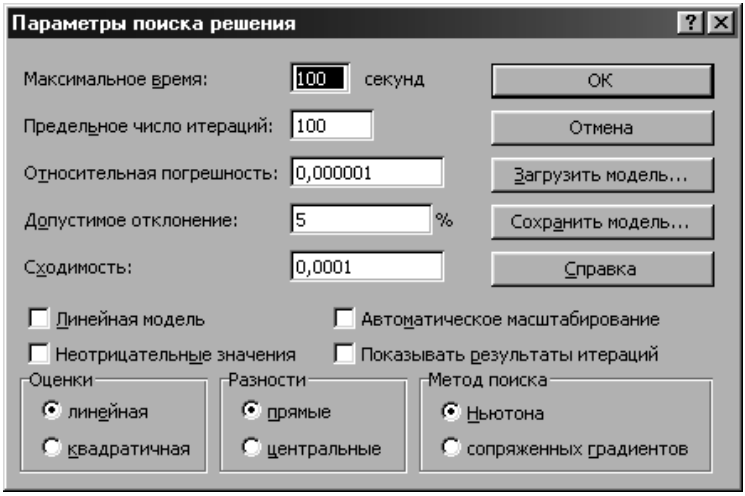


Рис. 15.8. Диалоговое окно *Параметры поиска решения*

С помощью находящихся в нем команд можно вводить условия для решения задач оптимизации всех классов. Значения, используемые по умолчанию, подходят для решения большей части практических задач.

Устанавливаем флажок *Линейная модель*, что обеспечивает применение симплексного метода.

После нажатия кнопки *Выполнить* окна *Поиск решения* открывается диалоговое окно *Результаты поиска решения* (рис. 15.9). Если решение не найдено, будет выведено соответствующее сообщение.

Далее выделяем необходимый тип отчета для анализа полученного оптимального решения (в Excel предусмотрены три типа отчетов: отчет по результатам, устойчивости, пределам); в нашем примере следует выбрать

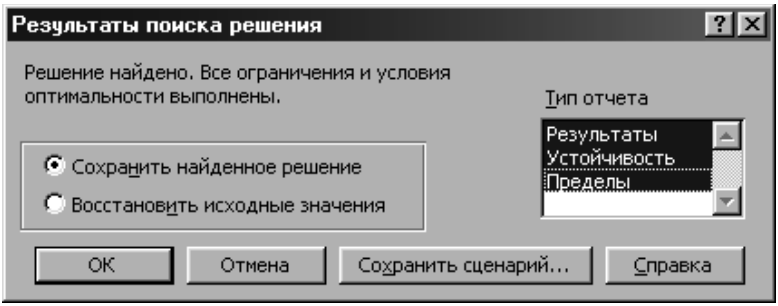


Рис. 15.9. Диалоговое окно *Результаты поиска решения*

все три типа. Нажимаем *OK* и на экране появляется окно *Результаты решения задачи* (рис. 15.10): в нем приводится оптимальное решение $X_1 = 0$ (B3), $X_2 = 5$ (C3), $X_3 = 12{,}5$ (D3), $X_4 = 0$ (E3).

Полученное решение означает, что максимальный доход 225 ден. ед. (F5) предприятие может получить при выпуске 5 ед. второго вида продукции и 12,5 ед. третьего вида продукции. При этом ресурсы P_1 и P_2 будут израсходованы полностью, из 40 ед. P_3 будет использовано 30 ед.

В *Отчете по результатам* (рис. 15.11) приведены сведения о целевой функции, значениях искомых переменных и результаты оптималь-

	A	B	C	D	E	F	G	H
1		Переменные						
2	Имя	X1	X2	X3	X4			
3	Значение	0	5	12,5	0			
4	Нижн. гр.	0	0	0	0	ЦФ	Экстр.	
5	Коэф. в ЦФ	14	10	14	11	225	max	
6								
7		Ограничения				Левая часть	Знак	Правая часть
8	1-ое огран.	4	2	2	3	35	<=	35
9	2-ое огран.	1	1	2	3	30	<=	30
10	3-е огран.	3	1	2	1	30	<=	40

Рис. 15.10. *Результаты решения задачи*

	A	B	C	D	E	F	G
5							
6	Целевая ячейка (Максимум)						
7		**Ячейка**	**Имя**	**Исходное значение**	**Результат**		
8		$F$5	Коэф. в ЦФ ЦФ	0	225		
9							
10	Изменяемые ячейки						
11		**Ячейка**	**Имя**	**Исходное значение**	**Результат**		
12		$B$3	Значение X1	0	0		
13		$C$3	Значение X2	0	5		
14		$D$3	Значение X3	0	12,5		
15		$E$3	Значение X4	0	0		
16							
17	Ограничения						
18		**Ячейка**	**Имя**	**Значение**	**Формула**	**Статус**	**Разница**
19		$F$8	1-ое огран. Левая часть	35	$F$8<=$H$8	связанное	0
20		$F$9	2-ое огран. Левая часть	30	$F$9<=$H$9	связанное	0
21		$F$10	3-е огран. Левая часть	30	$F$10<=$H$10	не связан.	10
22		$B$3	Значение X1	0	$B$3>=$B$4	связанное	0
23		$C$3	Значение X2	5	$C$3>=$C$4	не связан.	5
24		$D$3	Значение X3	12,5	$D$3>=$D$4	не связан.	12,5
25		$E$3	Значение X4	0	$E$3>=$E$4	связанное	0

Рис 15.11. *Отчет по результатам*

ного решения для ограничений. Для ограничений в столбце *Формула* приведены зависимости, которые были введены в диалоговое окно *Поиск решения*, в столбце *Значение* — величины использованного ресурса, в столбце *Разница* показано количество неиспользованного ресурса. Если ресурс используется полностью, то в столбце *Статус* указывается «связанное», если не полностью — «не связан.». Для переменных показывается разность между значениями переменных в найденном оптимальном решении и заданным для них граничным условием.

В *Отчете по устойчивости* (рис. 15.12) дан анализ по переменным и ограничениям. Исследования устойчивости оптимального решения — это изучение влияния изменений отдельно взятых параметров модели (оценок целевой функции, технико-экономических коэффициентов, объемов ограничений по ресурсам и продуктам, значений базисных переменных и др.) и ее структуры (введение новых ограничений и переменных или их сокращение) на показатели оптимального решения. Такой анализ позволяет судить о пределах допустимых изменений в оптимальном плане и его устойчивости.

В анализе переменных приведены следующие данные:

- результирующие значения переменных;
- нормированная стоимость, т.е. дополнительные двойственные переменные, которые показывают, насколько изменяется целевая функция при принудительном включении единицы этой переменной в оптимальное решение.

В зависимости от имеющейся версии могут присутствовать дополнительно и такие данные, как коэффициенты целевой функции и допустимые значения приращения коэффициентов целевой функции, при которых сохраняется набор переменных, входящих в оптимальное решение.

Изменяемые ячейки

Ячейка	Имя	Результ. значение	Нормир. стоимость	Целевой коэффициент	Допустимое увеличение	Допустимое уменьшение
$B$3	Значение X1	0	–2	14	2	1E+30
$C$3	Значение X2	5	0	10	4	0,666666667
$D$3	Значение X3	12,5	0	14	2	4
$E$3	Значение X4	0	–10	11	10	1E+30

Ограничения

Ячейка	Имя	Результ. значение	Теневая цена	Ограничение правая часть	Допуст. увелич.	Допуст. уменьш.
$F$8	1-е огран. Левая часть	35	3	35	25	5
$F$9	2-е огран. Левая часть	30	4	30	5	12,5
$F$10	3-е огран. Левая часть	30	0	40	1E+30	10

Рис 15.12. *Отчет по устойчивости*

В анализе ограничений приведены значения:

- величин использованных ресурсов;
- теневые цены, т.е. двойственные оценки, которые показывают, как изменится целевая функция при изменении ресурсов на единицу.

В зависимости от имеющейся версии могут присутствовать дополнительно и такие данные, как значения приращения ресурсов, при которых сохраняется оптимальный набор переменных, входящих в оптимальное решение.

В *Отчете по пределам* (рис. 15.13) показано, в каких пределах может изменяться выпуск продукции, вошедшей в оптимальное решение, при сохранении структуры оптимального решения.

III. Сформулируем двойственную задачу для задачи об оптимальном планировании производства.

Предположим, что некоторая организация может закупить все ресурсы, которыми располагает предприятие.

Необходимо определить оптимальные оценки на эти ресурсы, исходя из естественного условия, что покупающая организация стремится минимизировать общую оценку ресурсов. Однако следует учитывать и тот факт, что за ресурсы покупающая организация должна уплатить сумму не меньше, чем та, которую может выручить предприятие при организации собственного производства продукции.

На основе условий производства необходимо установить оценки используемых в производстве ресурсов с учетом их влияния на конечный результат производства.

Обозначим оценки единицы каждого вида ресурса:

- y_1 — двойственная оценка ресурса P_1, или «цена» P_1.
- y_2 — двойственная оценка ресурса P_2, или «цена» P_2.
- y_3 — двойственная оценка ресурса P_3, или «цена» P_3.

	Целевое					
Ячейка	Имя	Значение				
$F$5	Коэф. в ЦФ ЦФ	225				
	Изменяемое		Нижний	Целевой	Верхний	Целевой
Ячейка	Имя	Значение	предел	результат	предел	результат
$B$3	Значение X1	0	0	225	0	225
$C$3	Значение X2	5	0	175	5	225
$D$3	Значение X3	12,5	0	50	12,5	225
$E$3	Значение X4	0	0	225	0	225

Рис 15.13. *Отчет по пределам*

Для объективной оценки ресурсов необходимо, чтобы общая стоимость всех ресурсов, находящихся в распоряжении предприятия, была как можно меньше, т.е. *целевая функция*

$$Z(\bar{Y}) = 35y_1 + 30y_2 + 40y_3 \to \min.$$

Ограничения. В исходной задаче 4 переменные, следовательно, в двойственной задаче 4 ограничения. В правых частях ограничений двойственной задачи стоят коэффициенты при неизвестных в целевой функции исходной задачи. Левая часть ограничений определяет стоимость ресурсов, затраченных на производство единицы продукции. Каждое ограничение соответствует определенному виду продукции. Оценка ресурсов, затрачиваемых на выпуск единицы готовой продукции, должна быть не меньше оценки единицы готовой продукции:

$$\begin{cases} 4y_1 + y_2 + 3y_3 \geq 14, \\ 2y_1 + y_2 + y_3 \geq 10, \\ 2y_1 + 2y_2 + 2y_3 \geq 14, \\ 3y_1 + 3y_2 + y_3 \geq 11, \end{cases}$$

$$y_j \geq 0, \quad i = 1, 2, 3.$$

IV. Выпишем оптимальный план двойственной задачи.

Решение двойственной задачи содержится в *Отчете по устойчивости* (рис. 15.12). Основные переменные в столбце *Теневая цена*, а дополнительные — в столбце *Нормируемая стоимость*. Выпишем значения переменных:

$$y_1^* = 3, \quad y_2^* = 4, \quad y_3^* = 0.$$

Отсюда получаем значение целевой функции двойственной задачи при подстановке этих значений:

$$Z(\bar{Y}) = 35y_1 + 30y_2 + 40y_3 = 35 \cdot 3 + 30 \cdot 4 + 40 \cdot 0 = 225.$$

V. Используя протоколы *Поиска решения*, выполним анализ полученного оптимального решения исходной задачи.

Во второй части *Отчета по устойчивости* содержится информация по ограничениям.

1. *Величина использованных ресурсов* (графа *Результирующее значение*).

Сравнивая первоначальное количество каждого ресурса с результирующим значением, мы получим следующий результат:

- ресурс P_1 — было 35, израсходовано 35;
- ресурс P_2 — было 30, израсходовано 30;
- ресурс P_3 — было 40, а израсходовано 30, значит, 10 ед. P_3 остались неизрасходованными.

2. *Предельные значения приращения ресурсов* Δb_i. В графе *Допустимое уменьшение* показано, насколько можно уменьшить (устранить излишек) или увеличить (повысить минимально необходимое требование) ресурс, сохранив при этом структуру оптимального решения. Рассмотрим анализ дефицитных ресурсов. Анализируя отчет по результатам, мы установили, что существуют ограничения, не позволяющие предприятию выпускать больше продукции, чем в оптимальном решении, и получать более высокий доход. В рассматриваемом примере ограничениями являются дефицитные ресурсы P_1 и P_2. Поскольку знак ограничений этих запасов имеет вид ≤, то возникает вопрос, насколько максимально должен возрасти запас этих ресурсов, чтобы обеспечить увеличение выпуска продукции. Ответ на этот вопрос приведен в графе *Допустимое увеличение*. Ресурс P_1 имеет смысл увеличить самое большое на 25 ед., а ресурс P_2 — на 5 ед.

3. *Ценность дополнительной единицы ресурса i* (графа *Теневая цена*). Ресурсы P_1 и P_2 имеют отличные от нуля оценки $y_1^* = 3$ и $y_2^* = 4$. Эти ресурсы полностью используются в оптимальном плане и являются дефицитными, т.е. сдерживающими рост целевой функции.

VI. Поясним равенство нулю x_1 и x_4.

Если изделие вошло в оптимальный план ($x_j \geq 0$), то в двойственных оценках оно не убыточно, т.е. стоимость ресурсов, затраченных на производство единицы изделия, равна его цене. Такие изделия эффективны, выгодны с точки зрения принятого критерия оптимальности. В нашей задаче — это продукция второго и третьего видов.

Если стоимость ресурсов, затраченных на производство одного изделия, больше его цены, то это изделие не войдет в оптимальный план из-за своей убыточности. В нашей задаче в план выпуска продукции не вошла продукция первого и четвертого видов, потому что затраты по ним превышают цену на 2 ден. ед. (16 – 14 = 2) и 10 ден. ед. (21 – 11 = 10) соответственно. Этот факт можно подтвердить, подставив в ограничения двойственной задачи оптимальные значения Z:

$$4 \cdot 3 + 4 + 3 \cdot 0 = 16 > 14,$$

$$2 \cdot 3 + 4 + 0 = 10 = 10,$$

$$2 \cdot 3 + 2 \cdot 4 + 2 \cdot 0 = 14 = 14,$$

$$3 \cdot 3 + 3 \cdot 4 + 0 = 21 > 11.$$

Разница между правыми и левыми частями ограничений двойственной задачи можно найти в *Отчете по устойчивости* в столбце *Нормируемая стоимость* (с обратным знаком).

Задачи для самостоятельного решения

Симплексный метод

15.1. Издательский дом «Садовод» издает два журнала — «Пчеловод» и «Сад и огород», которые печатаются в трех типографиях: «Типография МК», «Полиграф» и «Труд», где общее количество часов, отведенное для печати, и производительность печати одной тыс. экз. ограничены и представлены в табл. 15.12.

Таблица 15.12

Типография	Время печати 1000 экз.		Ресурс времени, отведенный типографией, ч
	«Пчеловод»	«Сад и огород»	
«Типография МК»	6	8	80
«Полиграф»	4	6	120
«Труд»	4	5	70
Оптовая цена, ден. ед./шт.	22	25	

Спрос на журнал «Пчеловод» составляет не более 12 тыс. экз., а на журнал «Сад и огород» — не более 14 тыс. экз. в месяц.

Симплексным методом определить, какое оптимальное количество журналов надо издавать, чтобы обеспечить максимальную выручку от продажи.

15.2. Молочный комбинат освоил выпуск новых видов сыров — «Приятный» и «Столичный», спрос на которые составляет соответственно не более 15 и 12 тыс. ед. в месяц. По причине занятости 4 цехов выпуском традиционных видов молочных продуктов каждый цех может выделить только ограниченный ресурс времени в месяц. В силу специфики технологического оборудования затраты времени на производство сыров разные и представлены в табл. 15.13.

Таблица 15.13

Номер цеха	Время на производство сыра, ч		Время, отведенное цехами на производство, ч/мес.
	«Приятный»	«Столичный»	
1	2	7	66
2	3	5	45

Окончание табл. 15.13

Номер цеха	Время на производство сыра, ч		Время, отведенное цехами на производство, ч/мес.
	«Приятный»	«Столичный»	
3	2	4	58
4	1	6	72
Оптовая цена, ден. ед./т	7800	8400	

Определить оптимальный объем выпуска названных сыров, обеспечивающий максимальную выручку от их продажи.

15.3. По предписанию врача пациенту необходимо перейти на диету и за сезон употребить питательные вещества, содержащиеся в фруктах и ягодах, в количествах, указанных в табл. 15.14.

Таблица 15.14

Вещества	Содержание питательных веществ		Нормы потребления, г
	Фрукты	Ягоды	
P_1	3	1	18
P_2	1	2	20
P_3	2	5	40
P_4	0	2	14
P_5	2	4	32
Цена, ден. ед./кг	30	40	

Симплексным методом определить, какое количество питательных веществ должно быть во фруктах и ягодах, чтобы выполнить предписание врача с минимальными расходами.

15.4. Торговое предприятие реализует две группы товаров А и Б. Нормы затрат ресурсов на каждый тип товаров, лимиты ресурсов, а также доход на единицу каждой продукции заданы в табл. 15.15. Определить структуру товарооборота так, чтобы доход торгового предприятия был максимален.

Таблица 15.15

Виды ресурсов	Норма затрат ресурсов на ед. товара		Лимит ресурсов
	Товары группы А	Товары группы Б	
Рабочее время продавцов, чел.-ч	0,2	3	24
Площадь торговых залов, $м^2$	0,5	0,1	5

Окончание табл. 15.15

Виды ресурсов	Норма затрат ресурсов на ед. товара		Лимит ресурсов
	Товары группы А	Товары группы Б	
Площадь складских помещений, м^2	3	1	32
Накладные расходы, ден. ед.	5	4	75
Доход на ед. продукции, ден. ед.	4	3	

15.5. Цех может производить три вида продукции $П_1$, $П_2$ и $П_3$, используя для этого три вида сырья $С_1$, $С_2$ и $С_3$. Данные о норме затрат сырья на единицу продукции, запасах сырья и цене реализации единицы продукции приведены в табл. 15.16.

Таблица 15.16

Сырье	Затраты на единицу продукции			Запас сырья
	$П_1$	$П_2$	$П_3$	
$С_1$	1	5	4	340
$С_2$	2	1	3	230
$С_3$	1	2	0	100
Цена реализации	4	3	1	

Определить оптимальный план производства продукции, при котором доход максимален.

15.6. Цех может производить три вида продукции $П_1$, $П_2$ и $П_3$, используя для этого три вида сырья $С_1$, $С_2$ и $С_3$. Данные о норме затрат сырья на единицу продукции, запасах сырья и цене реализации единицы продукции приведены в табл. 15.17.

Таблица 15.17

Сырье	Затраты на единицу продукции			Запас сырья
	$П_1$	$П_2$	$П_3$	
$С_1$	3	1	2	270
$С_2$	0	5	7	370
$С_3$	6	7	0	420
Цена реализации	3	8	5	

Определить план производства продукции, при котором доход максимален.

15.7. Цех может производить два вида продукции $П_1$ и $П_2$, используя для этого три вида сырья $С_1$, $С_2$ и $С_3$. Данные о норме затрат сырья на единицу продукции, запасах сырья и цене реализации единицы продукции приведены в табл. 15.18.

Таблица 15.18

Сырье	Затраты на единицу продукции		Запас сырья
	$П_1$	$П_2$	
$С_1$	1	2	180
$С_2$	1	1	90
$С_3$	0	2	150
Цена реализации	5	4	

Определить оптимальный план производства продукции, при котором доход максимален.

Двойственность в линейном программировании

15.8. Составить двойственную задачу к задаче 15.1.

15.9. Составить двойственную задачу к задаче 15.3 и решить ее симплексным методом.

15.10. Составить двойственную задачу к задаче 15.4 и на основании решения прямой задачи, используя соответствия переменных, выписать ответ двойственной задачи.

15.11. В задаче 15.5 определить, как изменится доход, если запас сырья второго вида будет меньше на 60 ед., а третьего — больше на 30 ед.

15.12. В задаче 15.6 определить наиболее дефицитный вид сырья и целесообразность включения в план продукции $П_4$, затраты сырья $С_1$, $С_2$ и $С_3$ на единицу которой равны соответственно 4, 1 и 3 ед., а цена составляет 7 ден. ед.

Ответы

15.1. Журнал «Пчеловод» надо издавать в количестве 12 тыс. ед., а журнал «Сад и огород» — 1 тыс. ед. Максимальная выручка при этом будет составлять 289 ден. ед.
15.2. Сыр «Приятный» надо производить в количестве 15 тыс. ед. Максимальная выручка при этом будет составлять 117 000 ден. ед.
15.3. Питательных веществ во фруктах должно быть не менее 3,2 ед., а в ягодах — не менее 8,4 ед. При этом расходы будут составлять 432 ден. ед.
15.4. Товарооборот группы А должен быть 8,18, товарооборот группы Б — 7,45. При этом доход будет составлять 55,1 ден. ед.
15.5. Продукцию $П_1$ надо производить в количестве 100 ед., $П_3$ — 10 ед. Доход составит 410 ден. ед.
15.6. Продукцию $П_2$ надо производить в количестве 60 ед., $П_3$ — 10 ед. Доход будет составлять 530 ден. ед.
15.7. Продукцию $П_1$ надо производить в количестве 90 ед. Доход будет составлять 450 ден. ед.
15.9. Оптимальные затраты на вещество $Р_1$ — 4 ден. ед., $Р_2$ — 18 ден. ед.
15.10. Оптимальные оценки ресурсов: $Р_1 = 0{,}57$ ден. ед., $Р_3 = 1{,}3$ ден. ед.
15.11. При изменении запасов план производства изменится: $П_1$ — 70 ед., $П_2$ — 30 ед. Доход будет составлять 370 ден. ед.
15.12. Продукцию $П_4$ целесообразно включать в план производства.

16

ТРАНСПОРТНАЯ ЗАДАЧА

Транспортная задача является одной из широко используемых задач линейного программирования. Ее постановка имеет свои особенности, для учета которых разработан специальный метод потенциалов для ее решения.

16.1. Постановка транспортной задачи

Среди задач линейного программирования особо выделяется транспортная задача (ТЗ) и ее частный случай — задача о назначениях. Для этой задачи разработан специальный алгоритм ее решения.

Пусть имеется m поставщиков однородного груза с запасами $a_1, a_2, ..., a_m$ и n потребителей этого груза с потребностями $b_1, b_2, ..., b_n$. Требуется найти план поставок грузов, минимизирующий стоимость перевозок при заданных тарифах перевозок.

В общем случае модель ТЗ записывается в виде:

$$F = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}(\min), \tag{16.1}$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i, i = 1, 2, ..., m \text{ (ограничения по запасам)} \\ \sum_{i=1}^{m} x_{ij} = b_j, j = 1, 2, ..., n \text{ (ограничения по потребностям)} \end{cases} \tag{16.2}$$

$$x_{ij} \geq 0, \quad i = 1, 2, ..., m, \quad j = 1, 2, ..., n.$$

В условии (16.1) числа c_{ij} обозначают тарифы перевозок единицы груза от i-го поставщика к j-му потребителю, x_{ij} — объемы перевозок от i-го поставщика к j-му потребителю. Первое равенство условия (16.2) означает, что все запасы грузов вывезены. Второе равенство условия (16.2) означает, что все потребности потребителей удовлетворены.

Данные задачи удобно записывать в виде *распределительной таблицы*, или *таблицы поставок* (табл. 16.1).

Таблица 16.1

Потребности / Запасы	b_1	b_2	...	b_n
a_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1n} x_{1n}
a_2	c_{21} x_{21}	c_{22} x_{22}	...	c_{2n} x_{2n}
...	...	...	...	...
a_m	c_{m1} x_{m1}	c_{m2} x_{m2}	...	c_{mn} x_{mn}

В ее клетках расположены: в правом верхнем углу — тарифы c_{ij}, в левом нижнем углу — объемы перевозок x_{ij}. *Допустимым планом* (опорным планом) решения задачи называется неотрицательное решение системы ограничений. Если при этом плане целевая функция достигает минимального значения, то план называется *оптимальным*.

Система ограничений (16.2) ТЗ по сравнению с ограничениями общей ЗЛП имеет свои особенности, которые позволяют применять специальный метод решения ТЗ.

Теорема 1 (о существовании допустимого плана). Для того чтобы транспортная задача имела допустимый план, необходимо и достаточно выполнение равенства

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \qquad (16.3)$$

ТЗ, для которых выполняется условие (16.3) теоремы, называются транспортными задачами *закрытого типа*. В дальнейшем на примере мы покажем, как можно любую ТЗ перевести в задачу закрытого типа введением фиктивных поставщиков или потребителей.

Как и ранее при решении ЗЛП, опорный план строится на основе выражения базисных переменных через свободные. Для закрытой ТЗ количество базисных неизвестных равно рангу матрицы

системы ограничений: $r = m + n - 1$. Это свойство следует из вида системы ограничений, содержащей $m + n$ равенств с учетом условия (16.3) закрытости задачи.

Таким образом, любой опорный план закрытой ТЗ содержит ровно $m + n - 1$ неотрицательных базисных переменных и $mn - (m + n - 1)$ — свободных переменных, равных нулю.

Этапы решения ТЗ будем иллюстрировать на следующем примере.

Пример 16.1. Имеется три поставщика с запасами груза 40, 10 и 30 ед. соответственно и четыре потребителя с потребностями груза 15, 22, 8 и 35 ед. Тарифы перевозок единицы груза от i-го поставщика к j-му потребителю заданы в виде матрицы тарифов

$$(c_{ij}) = \begin{pmatrix} 8 & 10 & 9 & 1 \\ 7 & 14 & 1 & 2 \\ 6 & 2 & 3 & 5 \end{pmatrix}.$$

Требуется найти план x_{ij}, $i = 1, 2, ..., m$, $j = 1, 2, ..., n$, перевозок груза, минимизирующий стоимость перевозок

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}.$$

Данные задачи запишем в виде распределительной таблицы (табл. 16.2).

Таблица 16.2

a_i \ b_j	15	22	8	35
40	8 x_{11}	10 x_{12}	9 x_{13}	1 x_{14}
10	7 x_{21}	14 x_{22}	1 x_{23}	2 x_{24}
30	6 x_{31}	2 x_{32}	3 x_{33}	5 x_{34}

16.2. Построение начального опорного плана

Опорные решения ТЗ будем находить, пользуясь распределительной таблицей, при заполнении которой будем руководствоваться следующим правилом.

Если переменная x_{ij} является базисной, то в соответствующую клетку (i, j) будем вписывать ее значение, называемое *поставкой клетки*. При этом сама клетка будет называться *загруженной*, а если x_{ij} является свободной переменной, то клетка будет называется *свободной*. Таким образом, любой опорный план будет загружать ровно $m + n - 1$ клеток. Поскольку базисные переменные также могут принимать нулевые значения, то в загруженных клетках могут стоять нули. Решение ТЗ будем записывать в виде *таблицы поставок*, которая получается из распределительной таблицы заполнением клеток, соответствующих базисным переменным, их значениями. Левый нижний угол свободных клеток остается незаполненным. Построение решения ТЗ следует идее симплекс-метода. Сначала находится превоначальное опорное решение, затем это решение проверяется на оптимальность. При необходимости осуществляется переход к следующему опорному решению, не ухудшающему целевую функцию, и т.д. В отличие от симплекс-метода решения общей ЗЛП, когда этапы решения проводились с использованием симплекс-таблиц, этапы решения ТЗ будут проводиться на основе таблиц поставок.

Существует несколько методов построения начального опорного плана. Наиболее известными из них являются следующие: метод северо-западного угла, метод наименьшего элемента (тарифа), метод Фогеля, Венгерский метод. Мы примем за основу *метод наименьшего элемента* и будем его иллюстрировать в приведенных примерах.

Просматривая тарифы клеток таблицы, находим клетки с минимальным тарифом и загружаем их максимальным грузом. В нашем примере это клетки (1, 4) и (2, 3). Максимум поставки клетки (1, 4) равен $\min\{40, 35\} = 35$, максимум поставки клетки (2, 3) равен $\min\{10, 8\} = 8$. Поскольку $35 > 8$, загружаем клетку (1, 4)

поставкой, равной 35, и записываем ее в левый нижний угол клетки. В результате четвертый потребитель (столбец) полностью загружен и исключается из рассмотрения. Таблица поставок принимает вид

a_i \ b_j	15	22	8	35
40	8	10	9	1 35
10	7	14	1	2
30	6	2	3	3

Дальнейший выбор поставки проводится на основе следующей теоремы.

Теорема 2. Пусть на каждом шаге заполнения таблицы поставок загружается только одна клетка, причем на каждом таком шаге (кроме последнего) оказывается или полностью загружен один потребитель (столбец), или полностью разгружен один поставщик (строка). Тогда переменные, соответствующие заполненным клеткам, можно принять за базисные. При таком заполнении таблицы будет выполнено основное условие на ранг матрицы: $r = m + n - 1$.

Отметим, что на последнем шаге оказываются полностью загруженными и строка, и столбец. При этом число загруженных клеток равно $m + n - 1$.

На следующем шаге разгружается поставщик (строка), содержащая клетку, загруженную на предыдущем шаге, если заполнен столбец. Если же заполнена строка, то недостающей поставкой закружается потребитель (столбец). При этом опять же выбирается минимальный тариф строки или столбца.

В нашем примере клеткой (1, 4) заполнен четвертый столбец. Поэтому будем разгружать первую строку от оставшегося в ней груза, равного $40 - 35 = 5$. Минимальный тариф незаполненных клеток первой строки, равный $\min\{8, 10, 9\} = 8$, находится в первом столбце. Поэтому загружаем клетку (1, 1) поставкой, равной 5.

На этом шаге полностью загружена первая строка. Таблица поставок принимает вид

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	7	14	1	2
30	6	2	3	3

На следующем шаге загружаем первый столбец недостающей потребностью, равной 15 – 5 = 10. Минимальный тариф незагруженных клеток этого столбца равен min{7, 6} = 6 и расположен в клетке (3, 1). Загружаем эту клетку поставкой, равной 10, тогда первый столбец загружен. Таблица поставок принимает вид

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	7	14	1	2
30	6 10	2	3	3

Переходим к третьей строке. Ее поставка осталась равной 30 – 10 = 20, а ее минимальный тариф, равный min{2, 3} = 2, расположен в клетке (3, 2). Загружаем клетку (3, 2) поставкой, равной 20, и загружаем третью строку. Таблица принимает вид

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	7	14	1	2
30	6 10	2 20	3	5

На следующем шаге загружается второй столбец: в клетку (2, 2) заносим поставку, равную 2.

На последнем шаге в клетку (2, 3) заносим поставку, равную $10 - 2 = 8$.

Окончательно таблица поставок принимает вид

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	7	14 2	1 8	2
30	6 10	2 20	3	5

Заполненными оказались $6 = 3 + 4 - 1$ клеток, и условие теоремы 2 выполнено. Таким образом, получено начальное опорное решение,

$$x_{11} = 5, \quad x_{12} = 0, \quad x_{13} = 0, \quad x_{14} = 35, \quad x_{21} = 0, \quad x_{22} = 2,$$
$$x_{23} = 8, \quad x_{24} = 0, \quad x_{31} = 10, \quad x_{32} = 20, \quad x_{33} = 0, \quad x_{34} = 0,$$

в котором базисные переменные соответствуют заполненным клеткам, а переменные, соответствующие свободным клеткам, равны нулю.

При построении начального опорного плана может быть нарушено условие теоремы 2. Может оказаться, что при загрузке клетки полностью загружены строка и столбец. Это означает, что одна из базисных переменных принимает нулевое значение. Это нулевое значение записывается в незаполненную клетку с минимальным тарифом клеток в строке или столбце, загруженных на последнем шаге. Тогда условие теоремы будет выполнено.

Проиллюстрируем этот случай следующим примером. Пусть условие задачи задано в виде таблицы:

a_i \ b_j	15	22	8	35
40	8 x_{11}	10 x_{12}	9 x_{13}	1 x_{14}
30	7 x_{21}	14 x_{22}	1 x_{23}	2 x_{24}
10	6 x_{31}	2 x_{32}	3 x_{33}	5 x_{34}

Действуем по описанному выше алгоритму. Выбираем начальную клетку (1, 4) с минимальным тарифом и максимальной загрузкой, загружаем ее поставкой 35. Четвертый столбец становится загруженным. Далее в клетку (1, 1) загружаем поставку, равную 5, затем клетку (3, 1) загружаем поставкой, равной 10. После этого оказываются загруженными первый столбец и четвертая строка. Занесем это в таблицу:

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
30	7	14	1	2
10	6 10	2	3	5

Среди тарифов незагруженных клеток первого столбца и третьей строки выбираем минимальный. Это тариф клетки (3, 2). Помещаем в эту клетку поставку, равную нулю. Тогда третья строка заполнена и согласно алгоритму мы должны переходить к загрузке второго столбца. Единственной незаполненной клеткой этого столбца является клетка (2, 2). Загружаем ее максимальной поставкой 22. Теперь загружаем вторую строку. Оставшуюся в ней клетку (2, 3) загружаем оставшимся грузом, равным 8. Полученный результат заносим в табл. 16.3.

Таблица 16.3

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
30	7	14 22	1 8	2
10	6 10	2 0	3	5

Соответствующее начальное опорное решение будет следующим:

$$x_{11} = 5, \quad x_{12} = 0, \quad x_{13} = 0, \quad x_{14} = 35, \quad x_{21} = 0, \quad x_{22} = 2,$$
$$x_{23} = 8, \quad x_{24} = 0, \quad x_{31} = 10, \quad x_{32} = 0, \quad x_{33} = 0, \quad x_{34} = 0.$$

16.3. Проверка опорного решения на оптимальность

Как и при решении общей задачи линейного программирования, оптимальность опорного решения проверяется по коэффициентам в выражении целевой функции через свободные переменные. Эти коэффициенты носят свое название.

Оценкой свободной клетки (i, j) называется коэффициент при свободной переменной x_{ij} в выражении целевой функции только через свободные переменные. В опорном решении значения свободных переменных равны нулю. Поэтому целевая функция принимает минимальное значение, соответствующее этому опорному решению, если оценки свободных клеток неотрицательны. Иначе: увеличивая значения свободных переменных, можно уменьшить значение целевой функции. Итак, опорное решение является *оптимальным*, если оценки свободных клеток *неотрицательны*. Теперь возникают вопросы: как найти оценки свободных клеток и как сделать их неотрицательными, переходя к другому опорному решению? Ответ на оба эти вопроса можно получить на основании таблицы поставок.

Выражение целевой функции через свободные переменные можно получить из таблицы поставок, в которой уже найдено начальное опорное решение, как в § 16.2. Согласно табл. 16.3 значение целевой функции равно сумме произведений поставок загруженных клеток на их тарифы. А поскольку загруженные клетки соответствуют базисным переменным, то мы имеем выражение целевой функции F только через базисные переменные. При этом в опорном решении сами свободные переменные равны нулю. Для перехода к выражению F через свободные переменные, т.е. к нахождению оценок свободных клеток, будем пользоваться свойством таблицы поставок, которое является результатом следующей теоремы.

Теорема 3 (о потенциалах). Оценки свободных клеток не изменятся, если к тарифам некоторой строки (или столбца) таблицы поставок прибавить (или вычесть) некоторое число.

На основании этой теоремы мы будем вычитать из тарифов c_{ij} строк и столбцов поставок такие числа u_i и v_j, называемые потенциалами, чтобы тарифы всех загруженных клеток обратились

в нуль. Тогда в свободных клетках матрицы поставок появятся их оценки

$$\delta_{ij} = c_{ij} - u_i - v_i, \tag{16.4}$$

где c_{ij} — тариф клетки (i, j), u_i — потенциал i-й строки, v_j — потенциал j-го столбца. Способ нахождения потенциалов мы опишем на примере таблицы поставок. Справа от табл. 16.4 запишем столбец потенциалов строк (u_1, u_2, u_3), а снизу от таблицы — строку потенциалов столбцов (v_1, v_2, v_3, v_4).

Таблица 16.4

b_j /	15	22	8	35	u_i
40	8 5	10	9	1 35	u_1
10	7	14 2	1 8	2	u_2
30	6 10	2 20	3	5	u_3
v_j	v_1	v_2	v_3	v_4	

Для значений потенциалов, обращающих в нуль тарифы загруженных клеток, имеем систему линейных уравнений

$$8 = u_1 + v_1, \quad 1 = u_1 + v_4, \quad 14 = u_2 + v_2,$$
$$1 = u_2 + v_3, \quad 6 = u_3 + v_1, \quad 2 = u_3 + v_2.$$

В общем случае эта система состоит из $m + n - 1$ независимых уравнений и содержит $m + n$ неизвестных. Поэтому один из потенциалов, например u_1, можно положить равным нулю. Тогда остальные потенциалы находятся однозначно. В нашем случае имеем решение системы

$$u_1 = 0, \quad u_2 = 10, \quad u_3 = -2, \quad v_1 = 8, \quad v_2 = 4, \quad v_3 = -9, \quad v_4 = 1.$$

Теперь по формуле (16.4) получаем оценки свободных клеток:

$$\delta_{12} = 10 - 0 - 4 = 6, \quad \delta_{13} = 9 - 0 + 9 = 18,$$

$$\delta_{21} = 7 - 10 - 8 = -11, \quad \delta_{24} = 2 - 10 - 1 = -9,$$

$$\delta_{33} = 3 + 2 + 9 = 14, \quad \delta_{34} = 5 + 2 - 1 = 6.$$

Поскольку среди этих оценок есть отрицательные, найденный опорный план поставок неоптимальный.

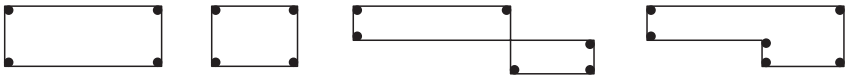
16.4. Переход к следующему распределению поставок

Если опорный план поставок, полученный после заполнения таблицы поставок, не является оптимальным, то необходим новый опорный план. Для его построения выбираем свободную клетку в таблице поставок с минимальной отрицательной оценкой и строим замкнутый цикл.

Циклом в таблице поставок называется упорядоченный набор клеток

$$(v) \to (i_1, j_2) \to (i_2, j_2) \to (i_2, j_3) \to \ldots \to (i_s, j_s) \to (i_s, j_1),$$

в котором среди наборов номеров (i_1, i_2, ..., i_s) и (j_1, j_2, ..., j_s) нет повторяющихся и все клетки, кроме первой, являются загруженными. Графически цикл можно изобразить связной ломаной линией, состоящей из горизонтальных и вертикальных отрезков с концами в клетках таблицы, одна из которых свободна, остальные загружены. При этом каждая строка и столбец таблицы содержат не более одного отрезка ломаной. Приведем графические изображения некоторых примеров цикла:



Цикл обладает следующим свойством: если в таблице поставок содержится опорный план, то для каждой свободной клетки можно построить, и притом только один, цикл с вершиной в этой клетке. Для наглядности в таблице поставок клетки цикла выделяются цветом.

В нашем примере минимальную отрицательную оценку имеет свободная клетка (2, 1). Ей соответствует цикл

$$(2, 1) \to (2, 2) \to (3, 2) \to (3, 1).$$

На основании выбранного цикла проводится перераспределение поставок клеток таблицы по следующей схеме. В клетках цикла расставляем чередующиеся знаки «+» и «–», которые означают увеличение или уменьшение загрузки клетки на величину, равную минимальному значению поставки из всех загруженных клеток цикла, отмеченных знаком «–». Свободная клетка цикла всегда наделяется знаком «+». В нашем примере знаки в таблице поставок распределяются следующим образом:

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	10 + 6	14 – 2	1 8	2
30	6 – 10	2 + 20	3	5

Минимальное значение загрузки клеток (2, 2) и (3, 1) равно $\min\{2, 10\} = 2$ и находится в клетке (2, 2). Далее проводится перераспределение поставок клеток цикла, в клетки со знаком «+» добавляется выбранное минимальное значение поставки, из клеток со знаком «–» вычитается это значение. В результате свободная клетка цикла (2, 1) становится загруженной, а клетка (2, 2) — свободной. Таблица поставок принимает вид:

a_i \ b_j	15	22	8	35
40	8 5	10	9	1 35
10	7 2	14	1 8	2
30	6 8	2 22	3	5

Таким образом, получен новый опорный план. Для проверки этого плана на оптимальность вновь составляем систему уравнений для нахождения потенциалов:

$$8 = u_1 + v_1, \quad 1 = u_1 + v_4, \quad 7 = u_2 + v_1,$$
$$1 = u_2 + v_3, \quad 6 = u_3 + v_1, \quad 2 = u_3 + v_2.$$

Решением этой системы являются следующие значения потенциалов:

$$u_1 = 0, \quad u_2 = -1, \quad u_3 = -2, \quad v_1 = 8, \quad v_2 = 4, \quad v_3 = 2, \quad v_4 = 1.$$

Теперь по формуле (16.4) получаем следующие оценки свободных клеток:

$$\delta_{12} = 10 - 0 - 4 = 6, \quad \delta_{13} = 9 - 0 - 2 = 7, \quad \delta_{22} = 14 + 1 - 4 = 11,$$
$$\delta_{24} = 2 + 1 - 1 = 2, \quad \delta_{33} = 3 + 2 - 2 = 3, \quad \delta_{34} = 5 + 2 - 1 = 6.$$

Поскольку все оценки положительны, найденный опорный план является оптимальным. Значениями плана перевозок будут следующие:

$$x_{11} = 5, \quad x_{12} = 0, \quad x_{13} = 0, \quad x_{14} = 35, \quad x_{21} = 2, \quad x_{22} = 0,$$
$$x_{23} = 8, \quad x_{24} = 0, \quad x_{31} = 8, \quad x_{32} = 22, \quad x_{33} = 0, \quad x_{34} = 0.$$

При этом суммарные затраты на перевозку составляют

$$F = 5 \cdot 8 + 35 \cdot 1 + 2 \cdot 7 + 8 \cdot 1 + 8 \cdot 6 + 22 \cdot 2 = 189 \text{ ден. ед.}$$

Приведем теперь алгоритм *метода потенциалов*:

1. На основании данных о запасах и потребностях перевозимых грузов и о тарифах перевозок составляем таблицу поставок. Проверяем условие разрешимости задачи.

2. Находим начальный опорный план (например, методом минимального элемента).

3. Проверяем полученный план на оптимальность. Если план оптимален, то решение получено.

4. Если план неоптимален, строим цикл перераспределения и находим новый опорный план.

5. Переходим к пункту 3.

Замечание. В алгоритме метода потенциалов иногда присутствует явление «зацикливания», когда последующий опорный план дает такое же значение целевой функции. Для устранения этого явления эффективным является правило, по которому нулевое значение присваивается потенциалу с одним и тем же номером.

Задача с решением

Готовая продукция заводов A_i, $i = 1, 2, 3$, направляется на склады B_j, $j = 1, 2, 3, 4$. Заводы A_i производят соответственно $a_1 = 250$, $a_2 = 150$, $a_3 = 400$ тыс. изделий. Пропускная способность складов B_j за это время характеризуется величинами $b_1 = 100$, $b_2 = 500$, $b_3 = 100$, $b_4 = 300$ тыс. изделий. Стоимость перевозки с завода A_i, $i = 1, 2, 3$, на склад B_j, $j = 1, 2, 3, 4$, одной тысячи изделий задается матрицей

$$C = \begin{bmatrix} 2 & 1 & 3 & 6 \\ 1 & 4 & 7 & 9 \\ 6 & 2 & 4 & 5 \end{bmatrix}.$$

Требуется:

1) составить экономико-математическую модель задачи, которая позволила бы найти план перевозки готовой продукции заводов на склады с минимальными затратами;

2) методом потенциалов найти оптимальный план перевозки готовой продукции на склады при дополнительном условии, что на складе B_4 созданы лучшие условия для хранения готовой продукции, а поэтому он должен быть загружен полностью.

Решение. Составим математическую модель транспортной задачи.

Обозначим через x_{ij}, $i = 1, 2, 3$, $j = 1, 2, 3, 4$, количество тыс. изделий, перевозимых с i-го завода на j-й склад.

Установим, сколько всего тыс. изделий производят заводы:

$$250 + 150 + 400 = 800 \text{ тыс. изделий.}$$

Пропускная способность складов:

$$100 + 500 + 100 + 300 = 1000 \text{ тыс. изделий.}$$

Поскольку пропускная способность складов превышает производство изделий заводами, то задача имеет открытый тип и для

дальнейшего ее решения введем фиктивный завод, который будет производить

$$a_4 = \sum_{j=1}^{4} b_j - \sum_{i=1}^{3} a_i = 1000 - 800 = 200 \text{ тыс. изделий.}$$

Стоимость перевозки одной тысячи изделий с фиктивного завода A_4 на склады B_j, $j = 1, 2, 3, 4$, равна $c_{4j} = 0$, $j = 1, 2, 3, 4$. После введения фиктивного завода открытая модель задачи преобразовалась в закрытую.

Таким образом, математическая постановка данной ТЗ состоит в нахождении такого неотрицательного решения системы линейных уравнений, при котором целевая функция принимает минимальное значение. Задачу решим методом потенциалов. Поскольку имеет место задача минимизации, то признаком оптимальности опорного плана будет отсутствие оценок свободных клеток с отрицательными оценками.

Для того чтобы учесть дополнительное условие, что на складе B_4 созданы лучшие условия для хранения готовой продукции, а значит, он должен быть загружен полностью, надо исключить перевозку изделий с фиктивного завода на этот склад B_4. Добиться этого можно искусственным завышением тарифа перевозки с завода A_4 на склад B_4, положив, например, $c_{44} = 100$.

Построим начальный опорный план ТЗ методом «минимального элемента» (табл. 16.5).

Таблица 16.5

Заводы	Склады				u_i
	B_1 (100)	B_2 (500)	B_3 (100)	B_4 (300)	
A_1(250)	2	1 250_1	3	6	$u_1 = -1$
A_2 (150)	1 100_2	4	7	9 50_6	$u_2 = 4$
A_3 (400)	6	2 250_3	4 — 100_4	5 + 50_5	$u_3 = 0$
A_4 (200)	0	0	0 +	100 — 200_7	$u_4 = 95$
v_j	$v_1 = -3$	$v_2 = 2$	$v_3 = 4$	$v_4 = 5$	

В результате распределения готовой продукции заводов на склады получили невырожденный план: условие для занятых клеток $m + n - 1 = 4 + 4 - 1 = 7$ выполняется. Полученный начальный опорный план имеет вид

$$X_0 = \begin{bmatrix} 0 & 250 & 0 & 0 \\ 100 & 0 & 0 & 50 \\ 0 & 250 & 100 & 50 \\ 0 & 0 & 0 & 200 \end{bmatrix}.$$

При данном плане перевозок их общая стоимость составляет:

$$F_{\min} = 1 \cdot 250 + 1 \cdot 100 + 9 \cdot 50 + 2 \cdot 250 + 4 \cdot 100 +$$

$$+ 5 \cdot 50 + 100 \cdot 200 = 21\,950.$$

Для исследования плана распределения продукции по складам на оптимальность необходимо найти оценки свободных клеток. Для этого сначала надо вычислить потенциалы u_i, $i = 1, 2, 3, 4$, и v_j, $j = 1, 2, 3, 4$, которые определяются в результате решения системы из семи уравнений, составленных по занятым клеткам, по формуле $u_i + v_j = c_{ij}$, $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$. Вычислим потенциалы и запишем их значения прямо в табл. 16.5.

Зная значения u_i, $i = 1, 2, 3, 4$, и v_j, $j = 1, 2, 3, 4$, можем найти оценки свободных клеток по формуле $\delta_{ij} = c_{ij} - (u_i + v_j)$:

$$\delta_{11} = 2 - (-1 - 3) = 6 \geq 0,$$

$$\delta_{13} = 3 - (-1 + 4) = 0 \geq 0,$$

$$\delta_{14} = 6 - (-1 + 5) = 2 \geq 0,$$

$$\delta_{22} = 4 - (4 + 2) = -2 < 0,$$

$$\delta_{23} = 7 - (4 + 4) = -1 < 0,$$

$$\delta_{31} = 6 - (0 - 3) = 9 \geq 0,$$

$$\delta_{41} = 0 - (95 - 3) = -92 < 0,$$

$$\delta_{42} = 0 - (95 + 2) = -97 < 0,$$

$$\delta_{43} = 0 - (95 + 5) = -100 < 0.$$

Поскольку среди оценок имеются отрицательные, то план X_0 неоптимальный и его можно улучшить, занимая клетку A_4B_3. Для этой клетки в табл. 16.5 построим замкнутый контур — цикл, по которому находим величину $\lambda = \min\{200, 100\} = 100$. Прибавляя величину λ к «положительным» клеткам и вычитая из «отрицательных», получаем новый опорный план, содержащийся в табл. 16.6.

Таблица 16.6

Заводы	Склады				u_i
	B_1 (100)	B_2 (500)	B_3 (100)	B_4 (300)	
A_1 (250)	2	1 250	3	6	$u_1 = -1$
A_2 (150)	1 100	4	7	9 50	$u_2 = 4$
A_3 (400)	6	2 — 250	4	5 + 150	$u_3 = 0$
A_4 (200)	0	0 +	0 100	100 — 100	$u_4 = 95$
v_j	$v_1 = -3$	$v_2 = 2$	$v_3 = -95$	$v_4 = 5$	

Опять найдем потенциалы и оценки свободных клеток:

$$\delta_{11} = 2 - (-1 - 3) = 6 \geq 0,$$

$$\delta_{13} = 3 - (-1 - 95) = 99 \geq 0,$$

$$\delta_{14} = 6 - (-1 + 5) = 2 \geq 0,$$

$$\delta_{22} = 4 - (4 + 2) = -2 < 0,$$

$$\delta_{23} = 7 - (4 - 95) = 98 \geq 0,$$

$$\delta_{31} = 6 - (0 - 3) = 9 \geq 0,$$

$$\delta_{33} = 4 - (0 - 95) = 99 \geq 0,$$

$$\delta_{41} = 0 - (95 - 3) = -92 < 0,$$

$$\delta_{42} = 0 - (95 + 2) = -97 < 0.$$

Поскольку среди оценок имеются отрицательные, то план неоптимален и его можно улучшить, занимая клетку A_4B_2. Для этой клетки в табл. 16.6 построим замкнутый контур — цикл, по которому находим величину $\lambda = \min\{250, 100\} = 100$. Прибавляя величину λ к «положительным» клеткам и вычитая из «отрицательных», получаем новый опорный план, содержащийся в табл. 16.7.

Таблица 16.7

Заводы	Склады				u_i
	B_1 (100)	B_2 (500)	B_3 (100)	B_4 (300)	
A_1 (250)	2	1 250	3	6	$u_1 = -1$
A_2 (150)	1 100	4 +	7	9 – 50	$u_2 = 4$
A_3 (400)	6	2 – 150	4	5 + 250	$u_3 = 0$
A_4 (200)	0	0 100	0 100	100	$u_4 = -2$
v_j	$v_1 = -3$	$v_2 = 2$	$v_3 = 2$	$v_4 = 5$	

Опять найдем потенциалы и оценки свободных клеток:

$$\delta_{11} = 2 - (-1 - 3) = 6 \geq 0,$$

$$\delta_{13} = 3 - (-1 + 2) = 2 \geq 0,$$

$$\delta_{14} = 6 - (-1 + 5) = 2 \geq 0,$$

$$\delta_{22} = 4 - (4 + 2) = -2 < 0,$$

$$\delta_{23} = 7 - (4 + 2) = 1 \geq 0,$$

$$\delta_{31} = 6 - (0 - 3) = 9 \geq 0,$$

$$\delta_{33} = 4 - (0 + 2) = 2 \geq 0,$$

$$\delta_{41} = 0 - (-2 - 3) = 5 \geq 0,$$

$$\delta_{44} = 100 - (5 - 2) = 97 \geq 0.$$

Поскольку среди оценок имеются отрицательные, то план неоптимален и его можно улучшить, занимая клетку A_2B_2. Для этой клетки в табл. 16.7 построим замкнутый контур — цикл, по которому находим величину $\lambda = \min\{150, 50\} = 50$. Прибавляя величину λ к «положительным» клеткам и вычитая из «отрицательных», получаем новый опорный план, содержащийся в табл. 16.8.

Таблица 16.8

Заводы	Склады				u_i
	B_1 (100)	B_2 (500)	B_3 (100)	B_4 (300)	
A_1 (250)	2 4 > 0	1 250	3 2 > 0	6 2 > 0	$u_1 = 1$
A_2 (150)	1 100	4 50	7 3 > 0	9 2 > 0	$u_2 = 4$
A_3 (400)	6 7 > 0	2 100	4 2 > 0	5 300	$u_3 = 2$
A_4 (200)	0 3 > 0	0 100	0 100	100 97 > 0	$u_4 = 0$
v_j	$v_1 = -3$	$v_2 = 0$	$v_3 = 0$	$v_4 = 3$	

Так как все оценки свободных клеток, которые записаны непосредственно в таблице, положительные, то в табл. 16.8 содержится единственный оптимальный план

$$X^* = \begin{bmatrix} 0 & 250 & 0 & 0 \\ 100 & 50 & 0 & 0 \\ 0 & 100 & 0 & 300 \\ 0 & 100 & 100 & 0 \end{bmatrix}.$$

При оптимальном плане перевозок общая стоимость перевозок составит

$$F_{\min} = 250 + 100 + 200 + 200 + 1500 = 2250 \text{ ден. ед.}$$

Итак, по оптимальному плану: с первого завода на второй склад надо отправить 250 тыс. изделий; со второго завода на первый склад — 100 тыс. изделий продукции и на второй — 50 тыс. изделий; с третьего завода надо отправить 100 тыс. изделий продукции на второй склад и 300 тыс. изделий на четвертый склад.

При этом будет выполнено дополнительное условие, т.е. первый склад будет загружен полностью. При этом продукция с заводов будет вывезена полностью, но пропускная способность третьего и четвертого складов будет не удовлетворена по 100 тыс. изделий. Общая стоимость перевозок при этом составит 2250 ден. ед.

Задачи для самостоятельного решения

16.1. Три молочных фермы с суточным производством 40, 25 и 35 тыс. л молока снабжают четыре молокозавода, спрос у которых 15, 40, 30 и 15 тыс. л молока в сутки. Молоко доставляется на заводы молоковозами, одинаковыми по вместимости. Стоимость провоза молока молоковозом на расстояние 1 км составляет 3 ден. ед. Расстояние от ферм до молокозаводов дано в таблице:

Молочные фермы	Молокозаводы			
	1	2	3	4
I	10	5	7	4
II	7	4	9	10
III	6	14	8	7

Найти оптимальный план поставки молока с ферм на молокозаводы с минимальными транспортными издержками. Рассчитать стоимость доставки молока от каждой фермы до молокозавода.

16.2. Имеется три участка пашни, на которых могут быть засеяны кукуруза, пшеница, ячмень и просо. Площадь каждого из участков равна соответственно 600, 180 и 220 га. Кукурузой, пшеницей, ячменем и просом следует засеять 290, 180, 110 и 420 га соответственно. Урожайность для каждого из участков различна и задается матрицей

$$\begin{pmatrix} 40 & 45 & 50 & 48 \\ 30 & 28 & 22 & 25 \\ 18 & 22 & 14 & 19 \end{pmatrix}.$$

Определить, сколько гектаров каждой культуры на каждом из участков следует засеять, чтобы общий сбор зерна был максимальным.

16.3. Четыре растворных узла поставляют раствор четырем строительным фирмам. Для перевозки раствора используются однотипные автомашины. Объем производства растворных узлов в день равен 30, 20, 40 и 50 т. Потребности строительных фирм в день составляют 35, 20, 55, 30 т. Расстояние в километрах от растворных узлов до строительных объектов указано в таблице:

Растворный узел	Строительные фирмы			
	1	2	3	4
I	2	4	1	3
II	5	6	3	4
III	3	6	7	5
IV	1	2	9	3

Определить, в каком объеме, с каких растворных узлов и куда должен доставляться раствор, чтобы транспортные издержки по его доставке были минимальными.

16.4. Мелькомбинаты №№ 1, 2, 3 ежедневно дают соответственно 60, 50 и 45 т муки. Методом потенциалов составить оптимальный план перевозки муки четырем хлебозаводам, потребности которых равны 35, 25, 75 и 15 т в день. Стоимость перевозки одной тонны муки с мелькомбината № 1 хлебозаводам соответственно равна 3, 2, 4 и 1 ден. ед., с мелькомбината № 2 — 1, 2, 2, 4 ден. ед. и с мелькомбината № 3 — 5, 2, 4, 3 ден. ед.

16.5. Имеются четыре овощехранилища, расположенные в разных районах города, в которых сосредоточено 10, 20, 35 и 45 т овощей соответственно. Овощи необходимо перевезти четырем потребителям соответственно в количестве 25, 30, 40 и 15 т. Расстояния от хранилищ до потребителей следующие:

Хранилище	Потребители			
	1	2	3	4
I	7	3	3	8
II	7	6	2	7
III	4	7	7	3
IV	5	2	4	5

Затраты на перевозку 1 т овощей на 1 км постоянны и равны 20 ден. ед.

Определить план перевозок продукта от хранилищ до потребителей из условия минимизации транспортных расходов.

16.6. В районе имеются три станции технического обслуживания (СТО) машин фирмы «МАЗ». Основными потребителями их ремонтных услуг являются четыре автопредприятия. Производственные мощности СТО, стоимость их услуг, затраты на транспортировку от автопредприятий на СТО и обратно, прогнозируемое количество ремонтов в планируемом периоде на каждом предприятии приведены в таблице:

СТО	Стоимость рем. часа, ден. ед.	Затраты на транспортировку				Производственные мощности
		$П_1$	$П_2$	$П_3$	$П_4$	
1	30	1	5	2	6	100
2	45	3	6	2	4	15
3	60	8	10	4	5	90
Количество ремонтов		30	40	55	80	

Требуется определить, какое количество автомашин из каждого автопредприятия необходимо отремонтировать на каждой СТО, чтобы суммарные расходы на ремонт и транспортировку были минимальными.

16.7. Три бензохранилища с суточным объемом хранения 60, 40 и 100 тыс. т авиационного бензина снабжают четыре аэропорта, спрос на бензин у которых составляет 30, 80, 65, 35 тыс. т бензина в сутки. Бензин транспортируется в аэропорты одинаковыми по вместимости бензозаправщиками. Стоимость провоза бензина бензозаправщиком на расстояние 1 км составляет 7 ден. ед. Расстояния от бензохранилищ до аэропортов следующие:

Бензохранилище	Аэропорты			
	1	2	3	4
I	8	12	4	9
II	7	5	15	3
III	9	4	6	12

Найти оптимальный план поставки бензина с минимальными транспортными издержками. Рассчитать стоимость доставки бензина от каждого аэропорта до хранилища.

Ответы

16.1. $X^* = \begin{bmatrix} 0 & 15 & 10 & 15 \\ 0 & 25 & 0 & 0 \\ 15 & 0 & 20 & 0 \end{bmatrix}$, $F_{\min} = 555$.

16.2. $X^* = \begin{bmatrix} 0 & 70 & 110 & 420 \\ 180 & 0 & 0 & 0 \\ 110 & 110 & 0 & 0 \end{bmatrix}$, $F_{\min} = 38\,610$.

16.3. $X^* = \begin{bmatrix} 0 & 0 & 30 & 0 \\ 0 & 0 & 20 & 0 \\ 5 & 0 & 5 & 30 \\ 30 & 20 & 0 & 0 \end{bmatrix}$, $F_{\min} = 360$.

16.4. $X^* = \begin{bmatrix} 0 & 25 & 15 & 15 \\ 35 & 0 & 15 & 0 \\ 0 & 0 & 45 & 0 \end{bmatrix}$, $F_{\min} = 370$.

16.5. $X^* = \begin{bmatrix} 0 & 0 & 10 & 0 \\ 0 & 0 & 20 & 0 \\ 20 & 0 & 0 & 15 \\ 5 & 30 & 10 & 0 \end{bmatrix}$, $F_{\min} = 6400$.

16.6. $X^* = \begin{bmatrix} 30 & 40 & 30 & 0 \\ 0 & 0 & 15 & 0 \\ 0 & 0 & 10 & 80 \end{bmatrix}$, $\mathrm{F}_{\min} = 36\,450$.

16.7. $X^* = \begin{bmatrix} 0 & 0 & 60 & 0 \\ 5 & 0 & 0 & 35 \\ 15 & 80 & 5 & 0 \end{bmatrix}$, $F_{\min} = 6055$.

ПРИЛОЖЕНИЯ

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

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061

Окончание приложения

	0	1	2	3	4	5	6	7	8	9
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

2. Таблица значений функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,15	0,0596	0,30	0,1179	0,45	0,1736
0,01	0,0040	0,16	0,0636	0,31	0,1217	0,46	0,1772
0,02	0,0080	0,17	0,0675	0,32	0,1255	0,47	0,1808
0,03	0,0120	0,18	0,0714	0,33	0,1293	0,48	0,1844
0,04	0,0160	0,19	0,0753	0,34	0,1331	0,49	0,1879
0,05	0,0199	0,20	0,0793	0,35	0,1368	0,50	0,1915
0,06	0,0239	0,21	0,0832	0,36	0,1406	0,51	0,1950
0,07	0,0279	0,22	0,0871	0,37	0,1443	0,52	0,1985
0,08	0,0319	0,23	0,0910	0,38	0,1480	0,53	0,2019
0,09	0,0359	0,24	0,0948	0,39	0,1517	0,54	0,2054
0,10	0,0398	0,25	0,0987	0,40	0,1554	0,55	0,2088
0,11	0,0438	0,26	0,1026	0,41	0,1591	0,56	0,2123
0,12	0,0478	0,27	0,1064	0,42	0,1628	0,57	0,2157
0,13	0,0517	0,28	0,1103	0,43	0,1664	0,58	0,2190
0,14	0,0557	0,29	0,1141	0,44	0,1700	0,59	0,2224

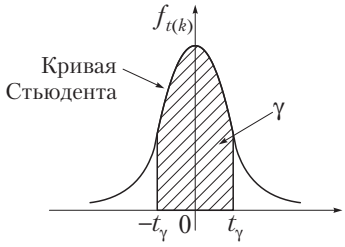
Продолжение приложения

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,60	0,2257	0,94	0,3264	1,28	0,3997	1,62	0,4474
0,61	0,2291	0,95	0,3289	1,29	0,4015	1,63	0,4484
0,62	0,2324	0,96	0,3315	1,30	0,4032	1,64	0,4495
0,63	0,2357	0,97	0,3340	1,31	0,4049	1,65	0,4505
0,64	0,2389	0,98	0,3365	1,32	0,4066	1,66	0,4515
0,65	0,2422	0,99	0,3389	1,33	0,4082	1,67	0,4525
0,66	0,2454	1,00	0,3413	1,34	0,4099	1,68	0,4535
0,67	0,2486	1,01	0,3438	1,35	0,4115	1,69	0,4545
0,68	0,2517	1,02	0,3461	1,36	0,4131	1,70	0,4554
0,69	0,2549	1,03	0,3485	1,37	0,4147	1,71	0,4564
0,70	0,2580	1,04	0,3508	1,38	0,4162	1,72	0,4573
0,71	0,2611	1,05	0,3531	1,39	0,4177	1,73	0,4582
0,72	0,2642	1,06	0,3554	1,40	0,4192	1,74	0,4591
0,73	0,2673	1,07	0,3577	1,41	0,4207	1,75	0,4599
0,74	0,2703	1,08	0,3599	1,42	0,4222	1,76	0,4608
0,75	0,2734	1,09	0,3621	1,43	0,4236	1,77	0,4616
0,76	0,2764	1,10	0,3643	1,44	0,4251	1,78	0,4625
0,77	0,2794	1,11	0,3665	1,45	0,4265	1,79	0,4633
0,78	0,2823	1,12	0,3686	1,46	0,4279	1,80	0,4641
0,79	0,2852	1,13	0,3708	1,47	0,4292	1,81	0,4649
0,80	0,2881	1,14	0,3729	1,48	0,4306	1,82	0,4656
0,81	0,2910	1,15	0,3749	1,49	0,4319	1,83	0,4664
0,82	0,2939	1,16	0,3770	1,50	0,4332	1,84	0,4671
0,83	0,2967	1,17	0,3790	1,51	0,4345	1,85	0,4678
0,84	0,2995	1,18	0,3810	1,52	0,4357	1,86	0,4686
0,85	0,3023	1,19	0,3830	1,53	0,4370	1,87	0,4693
0,86	0,3051	1,20	0,3849	1,54	0,4382	1,88	0,4699
0,87	0,3078	1,21	0,3869	1,55	0,4394	1,89	0,4706
0,88	0,3106	1,22	0,3883	1,56	0,4406	1,90	0,4713
0,89	0,3133	1,23	0,3907	1,57	0,4418	1,91	0,4719
0,90	0,3159	1,24	0,3925	1,58	0,4429	1,92	0,4726
0,91	0,3186	1,25	0,3944	1,59	0,4441	1,93	0,4732
0,92	0,3212	1,26	0,3962	1,60	0,4452	1,94	0,4738
0,93	0,3238	1,27	0,3980	1,61	0,4463	1,95	0,4744

Окончание приложения

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,96	0,4750	2,24	0,4875	2,56	0,4948	2,88	0,4980
1,97	0,4756	2,26	0,4881	2,58	0,4951	2,90	0,4981
1,98	0,4761	2,28	0,4887	2,60	0,4953	2,92	0,4982
1,99	0,4767	2,30	0,4893	2,62	0,4956	2,94	0,4984
2,00	0,4772	2,32	0,4898	2,64	0,4959	2,96	0,4985
2,02	0,4783	2,34	0,4904	2,66	0,4961	2,98	0,4986
2,04	0,4793	2,36	0,4909	2,68	0,4963	3,00	0,49865
2,06	0,4803	2,38	0,4913	2,70	0,4965	3,20	0,49931
2,08	0,4812	2,40	0,4918	2,72	0,4967	3,40	0,49966
2,10	0,4821	2,42	0,4922	2,74	0,4969	3,60	0,499841
2,12	0,4830	2,44	0,4927	2,76	0,4971	3,80	0,499928
2,14	0,4838	2,46	0,4931	2,78	0,4673	4,00	0,499968
2,16	0,4846	2,48	0,4934	2,80	0,4974	4,50	0,499997
2,18	0,4854	2,50	0,4938	2,82	0,4976	5,00	0,499997
2,20	0,4861	2,52	0,4941	2,84	0,4977		
2,22	0,4868	2,54	0,4945	2,86	0,4979		

3. Значения t_γ, определяемые уравнением $P = \left(\left|t(k) < t_\gamma\right|\right) = \gamma$



k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598

Окончание приложения

k \ γ	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,487
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,103	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,685	0,887	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Литература

1. *Белько И.В.* Теория вероятностей и математическая статистика. Примеры и задачи : учеб. пособие / И.В. Белько, Г.П. Свирид. Минск : Новое знание, 2007. 250 с.

2. *Белько И.В.* Высшая математика для экономистов. III семестр. Экспресс-курс / И.В. Белько, К.К. Кузьмич. Минск : Новое знание, 2007. 144 с.

3. *Белько И.В.* Эконометрика. Практикум / И.В. Белько, Е.А. Криштапович. Минск : Изд-во Гревцова, 2011. 221 с.

4. Высшая математика для экономистов : учебник. В 3 т. Т. 3 / С.А. Самаль [и др.]. Минск : Элайда, 2009. 226 с.

5. *Карманов В.Г.* Математическое программирование : учеб. пособие / В.Г. Карманов. 5-е изд. М. : ФИЗМАТЛИТ, 2001. 264 с.

6. *Кремер Н.Ш.* Теория вероятностей и математическая статистика : учебник для вузов / Н.Ш. Кремер. М. : ЮНИТИ ДАНА, 2001. 543 с.

7. *Кузнецов А.В.* Высшая математика. Математическое программирование : учеб. пособие / А.В. Кузнецов, В.А. Сакович, Н.И. Холод; под общ. ред. А.В. Кузнецова. 2-е изд. Минск : Высш. шк., 2001. 351 с.

8. *Кузнецов А.В.* Руководство к решению задач по математическому программированию : учеб. пособие / А.В. Кузнецов, Н.И. Холод, Л.С. Костевич; под общ. ред. А.В. Кузнецова. 2-е изд. Минск : Вышэйш. шк., 2001. 448 с.

9. *Орлова И.В.* Экономико-математические методы и модели : учеб. пособие / И.В. Орлова, В.А. Половников. М. : Вузовский учебник, 2008. 365 с.

10. *Письменный Д.Т.* Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д.Т. Письменный. 3-е изд. М. : Айрис-пресс, 2008. 288 с.

11. *Солодовников А.С.* Математика в экономике : учебник. В 3 ч. Ч. 3. Теория вероятностей и математическая статистика / А.С. Солодовников, В.А. Бабайцев, А.В. Браилов. М. : Финансы и статистика, 2008. 464 с.

12. *Урубков А.Р.* Курс MBA по оптимизации управленческих решений. Практическое руководство по использованию моделей линейного программирования / А.Р. Урубков. М. : Альпина Бизнес Букс, 2006. 176 с.

По вопросам приобретения книг обращайтесь:

Республика Беларусь

ООО «Новое знание»
220050, а/я 79, Минск.
Тел./факс: (10-375-17) 211-50-38.
E-mail: nk@wnk.biz
http:// wnk.biz

Российская Федерация

Отдел оптовых продаж «ИНФРА-М»:
127282, Москва, ул. Полярная, д. 31в, стр. 1.
Тел. (495) 280-15-96; факс (495) 280-36-29.
E-mail: books@infra-m.ru

Отдел «Книга–почтой»:
Тел. (495) 280-15-96 (доб. 246)

Учебное издание

Высшее образование: Бакалавриат

Белько Иван Васильевич
Морозова Инна Михайловна
Криштапович Елена Александровна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, МАТЕМАТИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Учебное пособие

Ответственный за выпуск *С.В. Исаенко*

Оригинал-макет подготовлен ООО «Новое знание»

Подписано в печать 28.12.2015.
Формат 60×90 $^{1}/_{16}$. Бумага офсетная. Гарнитура Петербург.
Печать офсетная. Усл. печ. л. 18,69. Уч.-изд. л. 15,03.
Тираж 400 экз. Заказ №

ТК 488800-542521-281215

Общество с ограниченной ответственностью «Новое знание».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/276 от 04.04.2014.
Пр. Пушкина, д. 15, ком. 16, Минск, Республика Беларусь.
Почтовый адрес: а/я 79, 220050, Минск, Республика Беларусь.
Телефон/факс: (10-375-17) 211-50-38
E-mail: nk@wnk.biz http://wnk.biz

ООО «Научно-издательский центр ИНФРА-М»
127282, Москва, ул. Полярная, д. 31В, стр. 1
Тел.: (495) 280-33-86. Факс: (495) 280-36-29
E-mail: books@infra-m.ru http://www.infra-m.ru

Отпечатано в ОАО «Можайский полиграфический комбинат».
143200, г. Можайск, ул. Мира, 93
www.oaompk.ru, www.oaompk.рф, тел.: 8-495-745-84-28, 8-49638-20-685

ИЗДАТЕЛЬСТВО “НОВОЕ ЗНАНИЕ”

Для записей

Наши координаты:
в Минске: (+375-17) 211-50-38, e-mail: sale@wnk.biz

http://wnk.biz

ИЗДАТЕЛЬСТВО “НОВОЕ ЗНАНИЕ”

Для записей

Наши координаты:
в Минске: (+375-17) 211-50-38, e-mail: sale@wnk.biz

http://wnk.biz

ИЗДАТЕЛЬСТВО “НОВОЕ ЗНАНИЕ”

Для записей

Наши координаты:
в Минске: (+375-17) 211-50-38, e-mail: sale@wnk.biz

http://wnk.biz

ИЗДАТЕЛЬСТВО “НОВОЕ ЗНАНИЕ”

Для записей

Наши координаты:
в Минске: (+375-17) 211-50-38, e-mail: sale@wnk.biz

http://wnk.biz

ИЗДАТЕЛЬСТВО “НОВОЕ ЗНАНИЕ”

Для записей

Наши координаты:
в Минске: (+375-17) 211-50-38, e-mail: sale@wnk.biz

http://wnk.biz