



ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*основные
понятия,
примеры,
задачи*



В.Н. ТУРЧИН

В.Н. ТУРЧИН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*ОСНОВНЫЕ
ПОНЯТИЯ,
ПРИМЕРЫ,
ЗАДАЧИ*

**Учебник для студентов
высших учебных заведений**

*Утверждено
Министерством образования и науки,
молодежи и спорта Украины*

Днепропетровск. «ИМА-ПРЕСС». 2012

УДК 519.2 (075.8)
ББК 22.17я73
Т89

Рецензенты:

А. В. Скороход, д-р физ.-мат. наук, проф., акад. НАН Украины (Институт математики НАН Украины),
М. И. Ядренко, д-р физ.-мат. наук, проф., чл.-кор. НАН Украины (Киевский национальный университет имени Тараса Шевченко).

Утверждено Министерством образования и науки, молодежи и спорта Украины в качестве учебника для студентов высших учебных заведений (письмо № 1/11 – 7270 от 04.08.11).

ТУРЧИН В. Н.

Т89 Теория вероятностей и математическая статистика. Основные понятия, примеры и задачи. Учебник для студентов высших учебных заведений. — Днепропетровск, ИМА-ПРЕСС. — 2012. — 576 с.

ISBN 978-966-331-437-2

Учебник охватывает программный материал курса "Теория вероятностей и математическая статистика" (или соответствующие разделы курса "Высшая математика").

Изложены основные понятия и факты теории вероятностей и математической статистики. Теоретические положения проиллюстрированы многочисленными примерами из всевозможных сфер деятельности человека (физики, химии, биологии, генетики, медицины, психологии, сельского хозяйства, космонавтики, военного дела, машиностроения, строительства, геологии, металлургии, экономики, лингвистики, социологии, психологии, спорта и т. д.).

К каждой главе приведен набор задач для самостоятельной работы, в отдельной главе даны ответы и краткие решения к задачам.

Для студентов высших учебных заведений.

ISBN 978-966-331-437-2

УДК 519.2 (075.8)
ББК 22.17я73
© В. Н. Турчин, 2012

Предисловие

Учебник предназначен для тех, кто приступает к изучению теории вероятностей и математической статистики. В каждой главе приведены основные понятия, утверждения, факты, задачи, примеры использования вероятностно-статистических методов и моделей в конкретных ситуациях. Учебник поможет овладеть основами теории, сформировать вероятностно-статистическое мышление и интуицию, овладеть навыками решения прикладных задач.

При первоначальном знакомстве с курсом необходимо рассмотреть как можно больше примеров и решить как можно больше задач. Их богатая подборка приведена в учебнике, при этом для иллюстрации накопленного на протяжении веков опыта применения теории вероятностей и математической статистики при исследовании реальных явлений и процессов в большинстве случаев задачи формулируются не в формально математических, а в естественно-научных терминах. Такие задачи естественным образом возникают в разнообразных сферах деятельности человека: физике, химии, биологии, генетике, медицине; в психологии, социологии, экологии, сельском хозяйстве; в астрономии, космонавтике, военном деле, машиностроении, строительстве, геологии, металлургии; в экономике, лингвистике, педагогике, спорте и т. д. Решение приведенных в книге задач требует неформального овладения материалом: необходимо предложить ту или иную математическую модель, обосновать этот выбор, выбрать метод решения задачи, дать интерпретацию полученным результатам.

Задачи классифицированы по степени трудности (хотя сложных задач в книге нет). При этом элементарные

задачи обозначены значком $^{\circ}$, более сложные — звездочкой, задачи средней сложности не выделены.

В отдельной главе приведены указания, решения задач и ответы.

Доказательство утверждений (если они не приведены) можно найти, например, в [11].

Учебником могут пользоваться как студенты механико-математических факультетов, факультетов прикладной математики и кибернетики университетов, так и студенты технических, педагогических, экономических высших учебных заведений. Поэтому одна глава посвящена комбинаторике, которая занимает важное место в курсе теории вероятностей этих учебных заведений.

В учебнике теоремы, примеры, формулы, таблицы, рисунки имеют тройную, а задачи — двойную нумерацию. Например, запись “формула (3.1.2)” обозначает формулу 2 из параграфа 1 главы 3, а запись “задача 5.12” — задачу 12 из главы 5.

Автор благодарен дочери Наталье Олерих и сыну Евгению Турчину за тщательное прочтение рукописи книги и ее содержательное обсуждение.

Автор будет благодарен всем, кто в той или иной форме выскажет свои соображения и замечания относительно содержания книги и стиля изложения материала. Замечания, предложения и пожелания просим посылать по адресу: В. Н. Турчину, кафедра статистики и теории вероятностей, механико-математический факультет, Днепропетровский национальный университет, пр. Гагарина, 72, Днепропетровск, 49010, Украина.

Автор

Глава 1

Элементы комбинаторики

1.1 Основной принцип комбинаторики

Комбинаторика изучает конечные множества. Множества будем обозначать большими латинскими буквами $A, B, C, \dots$, их элементы — малыми. Число элементов конечного множества A будем обозначать $n(A)$.

Вычисляя число элементов множества A , удобно пользоваться таким фактом: если между множествами A и B установлено взаимно однозначное соответствие, то

$$n(A) = n(B)$$

(часто число элементов множества B подсчитать проще, чем число элементов множества A).

Прямое произведение множеств. Пусть A и B — произвольные множества. Каждые два элемента $a \in A$ и $b \in B$ задают упорядоченную пару (a, b) . Множество всех упорядоченных пар (a, b) , $a \in A$, $b \in B$, будем называть *прямым (декартовым) произведением* множеств A и B и будем обозначать $A \times B$.

Пример 1.1.1. Найти прямые произведения $A \times B$ и $B \times A$, где $A = \{1, 2\}$ и $B = \{3, 4, 5\}$.

Решение. $A \times B = \{(1; 3), (1; 4), (1; 5), (2; 3), (2; 4), (2; 5)\}$, $B \times A = \{(3; 1), (3; 2), (4; 1), (4; 2), (5; 1), (5; 2)\}$.

Пусть дано k множеств $A_1, A_2, \dots, A_k$. Множество упорядоченных наборов $(a_1, a_2, \dots, a_k)$, где $a_1 \in A_1, a_2 \in A_2, \dots, a_k \in A_k$, будем называть прямым (декартовым) произведением множеств $A_1, A_2, \dots, A_k$ и будем обозначать $A_1 \times A_2 \times \dots \times A_k$.

Пример 1.1.2. Если $A_1 = R^1, A_2 = R^1, A_3 = R^1$, то прямое произведение $A_1 \times A_2 = R^1 \times R^1 = R^2$ является плоскостью, а $A_1 \times A_2 \times A_3 = R^1 \times R^1 \times R^1 = R^3$ — трехмерным пространством.

Правило умножения (основной принцип комбинаторики). Число $n(A \times B)$ элементов декартова произведения $A \times B$ конечных множеств A и B равно произведению $n(A)n(B)$ числа $n(A)$ элементов множества A и числа $n(B)$ элементов множества B :

$$n(A \times B) = n(A)n(B).$$

В самом деле, для каждого элемента $a \in A$ существует $n(B)$ элементов (a, b) , $b \in B$, декартова произведения $A \times B$. И поскольку множество A содержит $n(A)$ элементов, то число $n(A \times B)$ элементов декартова произведения $A \times B$ равно $n(A)n(B)$ (в левой части $n(A)$ слагаемых — по числу элементов в множестве A).

Правило умножения для декартова произведения k множеств: число $n(A_1 \times A_2 \times \dots \times A_k)$ элементов декартова произведения $A_1 \times A_2 \times \dots \times A_k$ конечных множеств $A_1, A_2, \dots, A_k$ равно произведению $n(A_1)n(A_2) \dots n(A_k)$ числа элементов $n(A_1), n(A_2), \dots, n(A_k)$ этих множеств:

$$n(A_1 \times A_2 \times \dots \times A_k) = n(A_1)n(A_2) \dots n(A_k).$$

Пример 1.1.3. Пусть $A = \{1, 2, 3\}$, $B = \{4, 5, 6, 7\}$. Найдите $n(A \times B)$.

Решение. $n(A \times B) = n(A)n(B) = 3 \cdot 4 = 12$.

Правило умножения в терминах действий. Часто правило умножения формулируют в терминах действий.

Пусть необходимо выполнить одно за другим k действий. Если первое действие можно выполнить n_1 числом способов, второе — n_2 числом способов и так до k -го

действия, которое можно выполнить n_k числом способов, то все k действие вместе могут быть выполнены $n_1 n_2 \dots n_k$ числом способов.

В самом деле, обозначим через A_1 множество способов выполнить первое действие, A_2 — второе, и т. д., A_k — k -е действие. Тогда элемент $(a_1, a_2, \dots, a_k)$ декартова произведения $A_1 \times A_2 \times \dots \times A_k$ задает способ выполнить все k действий вместе. Поэтому число всех способов выполнить k действий равно числу элементов декартова произведения $A_1 \times A_2 \times \dots \times A_k$. Следовательно,

$$n(A_1 \times A_2 \times \dots \times A_k) = n(A_1)n(A_2) \dots n(A_k) = n_1 n_2 \dots n_k.$$

Пример 1.1.4. *Сколько четырехзначных чисел можно записать цифрами 0, 1, 2, 3, 4, 5, если ни одна цифра не повторяется больше одного раза?*

Решение. Записывая четырехзначное число, мы выполняем четыре действия: записываем слева направо первую, вторую, третью, четвертую цифры. Первое действие можно выполнить пятью способами (ноль на первом месте не пишут), второе — пятью способами (одну цифру уже использовано при записи первой слева цифры, но, начиная со второго места, можно использовать ноль), третье действие — четырьмя способами, четвертое — тремя. Поэтому согласно правилу умножения все четыре действия вместе можно выполнить $5 \cdot 5 \cdot 4 \cdot 3 = 300$ способами. Следовательно, цифрами от 0 до 5 можно записать 300 различных четырехзначных чисел, в записи которых цифры не повторяются.

Упорядоченные множества. Множество, состоящее из n элементов, будем называть *n -элементным*.

Определение. *n -Элементное множество Ω будем называть упорядоченным, если каждому его элементу поставлено в соответствие число (номер элемента) от 1 до n , причем так, что различным элементам поставлены в соответствие различные номера (другими словами, установлено взаимно однозначное соответствие между множеством Ω и подмножеством $1, 2, \dots, n$ натуральных чисел).*

Упорядоченные множества различны, если они отличаются или своими элементами, или их порядком.

Каждое конечное множество можно упорядочить так: записать все его элементы в список $a, b, c, \dots, f$, а затем каждому элементу приписать номер места, на котором он стоит в списке. Как правило, так и будем поступать.

Перестановки. Упорядоченные множества, отличающиеся только порядком элементов, но не самими элементами, будем называть *перестановками*.

Определение. *Перестановкой n -элементного множества будем называть его n -элементное упорядоченное подмножество.*

Пример 1.1.5. *Выписать все перестановки множества $\Omega = \{a, b, c\}$.*

Решение.

$(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)$.

Число перестановок. Число P_n всех перестановок n -элементного множества (число способов упорядочить n -элементное множество) равно $n!$, т. е.

$$P_n = n!$$

Пример 1.1.6. *Сколькими способами можно упорядочить множество чисел $1, 2, \dots, 2n$ так, чтобы четные числа получили четные номера?*

Решение. Чтобы упорядочить множество $1, 2, \dots, 2n$, расположим $2n$ чисел на $2n$ местах, причем так, чтобы четные числа оказались на местах с четными номерами (а следовательно, нечетные — на местах с нечетными номерами). Выполним это в два действия.

Действие первое — расположить n четных чисел на n четных местах (упорядочить n -элементное множество) — можно выполнить $n!$ способами, действие второе — расположить n нечетных чисел на n нечетных местах — $n!$ способами.

Два действия вместе (расположить четные числа на четных местах, нечетные на нечетных) согласно правилу умножения можно выполнить $(n!)^2$ способами.

Размещения из n элементов по k . *Размещением из n элементов по k будем называть упорядоченное k -элементное подмножество n -элементного множества.*

Размещения из n элементов по k различны, если они отличаются или своими элементами, или их порядком.

Пример 1.1.7. Пусть $\Omega = \{a, b, c\}$. Выписать все размещения из 3 элементов по 2.

Решение. $(a, b), (b, a), (a, c), (c, a), (b, c), (c, b)$.

Число размещений. Число A_n^k всех упорядоченных k -элементных подмножеств n -элементного множества (число размещений из n элементов по k) равно $n(n-1) \dots (n-(k-1))$, т. е.

$$A_n^k = n(n-1) \dots (n-(k-1)).$$

Пример 1.1.8. Сколько трехзначных телефонных номеров можно составить из цифр от 0 до 9 так, чтобы в записи номера все цифры были разные?

Решение. Трехзначный телефонный номер из различных цифр — это 3-элементное упорядоченное подмножество множества $0, 1, \dots, 9$. А количество A_{10}^3 3-элементных упорядоченных подмножеств, которые можно составить из элементов 10-элементного множества, равно $10 \cdot 9 \cdot 8$, т. е.

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

Сочетания из n элементов по k . Сочетанием из n элементов по k будем называть k -элементное подмножество n -элементного множества.

Сочетания из n элементов по k различны, если они отличаются своими элементами (по меньшей мере одним). Порядок элементов в сочетании не существенный — сочетания, состоящие из одних и тех же элементов, неразличимы.

Пример 1.1.9. Пусть $\Omega = \{a, b, c\}$. Выписать все сочетания из 3 элементов по 1 и из 3 по 2.

Решение. $\{a\}, \{b\}, \{c\}$ — все сочетания из 3 элементов по 1, $\{a, b\}, \{a, c\}, \{b, c\}$ — из 3 по 2.

Заметим, что как сочетания $\{a, b\}$ и $\{b, a\}$; $\{b, c\}$ и $\{c, b\}$; $\{a, c\}$ и $\{c, a\}$ совпадают.

Число сочетаний. Число C_n^k всех k -элементных подмножеств n -элементного множества (число сочетаний из n элементов по k) равно $n!/(k!(n-k)!)$, т. е.

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

Пример 1.1.10 (шахматный город). Рассмотрим прямоугольную сетку квадратов — “шахматный город”, состоящий из $m \times n$ квадратных кварталов, разделенных $n - 1$ “горизонтальными” и $m - 1$ “вертикальными” улицами (рис. 1.1.1). Сколько на этой сетке различных кратчайших путей, ведущих из левого нижнего угла (точки $(0, 0)$) в правый верхний угол (в точку (m, n))?

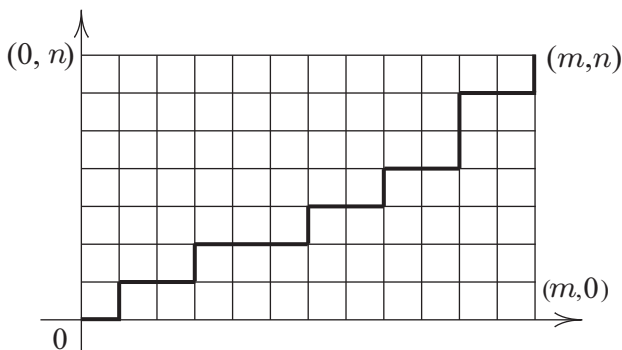


Рис. 1.1.1: “Шахматный город”

Решение. Обозначим буквой Γ горизонтальный отрезок пути, буквой $\mathbf{B}$ — вертикальный. Каждый кратчайший путь из $(0, 0)$ в (m, n) имеет n вертикальных отрезков и m горизонтальных. Он полностью задается упорядоченной последовательностью длиной $n + m$, составленной из m букв Γ и n букв $\mathbf{B}$ и наоборот. Поэтому число кратчайших путей равно числу последовательностей длиной $n + m$, составленных из m букв Γ и n букв $\mathbf{B}$. Каждая такая последовательность однозначно задается выбором m мест из $n + m$ для буквы Γ (оставшиеся места заполняются буквами $\mathbf{B}$), поэтому их число равно C_{n+m}^m .

Разбиение множеств. Разбиением n -элементного множества Ω на m попарно непересекающихся подмножеств, состоящих соответственно из $k_1, k_2, \dots, k_m$ элементов ($k_1 + k_2 + \dots + k_m = n$), будем называть набор

$$\{A, B, C, \dots, S\}$$

из m непересекающихся подмножеств множества Ω , состоящих соответственно из $k_1, k_2, \dots, k_m$ элементов.

Два разбиения на m попарно непересекающихся подмножеств, состоящих соответственно из $k_1, k_2, \dots, k_m$ элементов, различны, если хотя бы в одной паре соответствующих k_j -элементных подмножеств ($j = 1, 2, \dots, m$) имеются различные элементы.

Пример 1.1.11. Привести все возможные разбиения множества $\Omega = \{a, b, c, d\}$ на 3 непересекающихся подмножества: $\{A, B, C\}$, состоящие соответственно из $k_1 = 1$ элементов (множество A); $k_2 = 2$ элементов (множество B); $k_3 = 1$ элементов (множество C).

Решение.

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

Заметим, что например, $\{\{a\}, \{b, c\}, \{d\}\}$ и $\{\{d\}, \{b, c\}, \{a\}\}$ являются различными разбиениями множества $\Omega = \{a, b, c, d\}$.

Число разбиений множества на подмножества.

Число $C_n(k_1, k_2, \dots, k_m)$ разбиений n -элементного множества Ω на m непересекающихся подмножеств $B_1, B_2, \dots, B_m$, состоящих соответственно из $k_1, k_2, \dots, k_m$ элементов ($k_1 + k_2 + \dots + k_m = n$), равно

$$n! / (k_1! k_2! \dots k_m!),$$

т. е.

$$C_n(k_1, k_2, \dots, k_m) = \frac{n!}{k_1! k_2! \dots k_m!}.$$

Перестановки с повторениями. Перестановкой с повторениями (словом) длиной n , образованной k_1 элементами (буквами) a_1 , k_2 элементами (буквами) a_2 , и т. д., k_m элементами (буквами) a_m ($k_1 + k_2 + \dots + k_m = n$) будем называть упорядоченную последовательность длиной n , образованную k_1 элементами (буквами) a_1 , k_2 элементами (буквами) a_2 , и т. д., k_m элементами (буквами) a_m .

Два слова длиной n , образованные k_1 буквами a_1 , k_2 буквами a_2 , и т. д., k_m буквами a_m различны, если они отличаются порядком букв.

Число перестановок с повторениями. Число перестановок с повторениями (слов) длиной n , которые можно составить из k_1 элементов (букв) a_1 , k_2 элементов (букв) a_2 , и т. д., k_m элементов (букв) a_m ($k_1 + k_2 + \dots + k_m = n$), равно $C_n(k_1, k_2, \dots, k_m)$.

Пример 1.1.12. Сколько существует способов разместить n различных частиц по m ячейкам так, чтобы в первую ячейку попало k_1 частиц, во вторую — k_2 , и т. д., в m -ю — k_m ?

Решение. Разобьем n -элементное множество различных частиц на m непересекающихся подмножеств: k_1 -элементное подмножество частиц, которые попадут в первую ячейку, k_2 -элементное — во вторую, и т. д., k_m -элементное подмножество частиц, которые попадут в m -ю ячейку ($k_1 + k_2 + \dots + k_m = n$). А число способов разбить n -элементное множество, как это описано выше, равно $C_n(k_1, k_2, \dots, k_m)$.

Сочетания с повторениями. Сочетанием из m элементов по n с повторениями будем называть набор (множество) из n элементов, каждый из которых принадлежит одному из m типов.

Поскольку сочетание из m элементов по n с повторениями однозначно задается числом x_1 элементов первого типа, числом x_2 элементов второго типа, и т. д., числом x_m элементов m -го типа, в него входящих, т. е. задается последовательностью $(x_1, x_2, \dots, x_m)$ неотрицательных целых чисел, таких, что $x_1 + x_2 + \dots + x_m = n$, и наоборот, каждая такая последовательность задает набор из n элементов, x_1 из которых первого типа, x_2 — второго, и т. д., x_m — m -го типа, то можно дать такое определение сочетания из m элементов по n с повторениями.

Сочетанием из m элементов по n с повторениями будем называть последовательность $(x_1, x_2, \dots, x_m)$ неотрицательных целых чисел таких, что

$$x_1 + x_2 + \dots + x_m = n.$$

Два сочетания с повторениями из m элементов по n различны, если они отличаются количеством элементов

хотя бы одного типа. Порядок элементов в сочетании с повторениями не существенный.

Пример 1.1.13. *Выписать все сочетания с повторениями из 4 элементов a, b, c, d по 2.*

Решение. $aa, bb, cc, dd, ab, ac, ad, bc, bd, dc$.

Число сочетаний с повторениями. Число f_m^n сочетаний из m элементов по n с повторениями равно C_{n+m-1}^{m-1} , т. е.

$$f_m^n = C_{n+m-1}^{m-1}.$$

Если $n > m$, то число таких сочетаний из m элементов по n с повторениями, в которых каждый элемент встречается хотя бы один раз, равно C_{n-1}^{m-1} .

Пример 1.1.14. *Сколько неотрицательных решений в целых числах имеет уравнение $x_1 + x_2 + \dots + x_m = n$?*

Решение. Решением уравнения

$$x_1 + x_2 + \dots + x_m = n$$

в целых неотрицательных числах является последовательность $(x_1, x_2, \dots, x_m)$ целых неотрицательных чисел такая, что $x_1 + x_2 + \dots + x_m = n$, т. е. является сочетанием из m элементов по n с повторениями, а число таких сочетаний равно f_m^n .

1.2 Задачи

АЗ: 1.3°, 1.10, 1.14, 1.16°, 1.18, 1.19°, 1.22, 1.23, 1.25.

СЗ: 1.4°, 1.5°, 1.11°, 1.15, 1.17°, 1.20, 1.24, 1.27, 1.30, 1.32.

В предлагаемых далее задачах, прежде чем подсчитывать число элементов того или иного множества, необходимо четко определить: что именно представляют собой эти элементы (т. е., что подсчитывать). Прежде чем отвечать на вопрос “сколько?”, необходимо ответить на вопрос “что?” (что будем подсчитывать?).

1.1°. Из города A в город B ведет n дорог, а из B в C — m дорог. Сколькими способами можно осуществить путешествие по маршруту $A - B - C$?

1.2°. На вершину горы ведет семь тропинок. Сколькими способами турист может подняться на гору и спуститься с нее? Ответить на этот вопрос при условии, что восхождение и спуск осуществляются различными путями.

1.3°. В розыгрыше первенства страны по футболу берут участие 17 команд. Сколькими способами могут быть распределены между ними золотая, серебряная и бронзовая медали?

1.4°. Сколько трехзначных чисел можно записать цифрами 0, 1, 2, 3, 4?

1.4°. $4 \cdot 5 \cdot 5$.

1.5°. Сколько трехзначных чисел можно записать цифрами 0, 1, 2, 3, 4, если каждую из них использовать не более одного раза?

1.6°. Сколькими способами семь человек могут стать в очередь к кассе?

1.7°. Ученики изучают 10 предметов. В понедельник по расписанию 6 уроков, причем все разные. Сколькими способами можно составить расписание на понедельник?

1.8°. Сколько существует пятизначных чисел, делящихся на 5?

1.9°. Автомобильный номер состоит из двух букв и четырех цифр. Какое число различных номеров можно составить, используя 26 букв латинского алфавита?

1.10. В розыгрыше первенства страны по футболу принимают участие 16 команд. Команды, занявшие первое, второе и третье места, награждают соответственно золотой, серебряной и бронзовой медалями, а команды, оказавшиеся на двух последних местах, покинут высшую лигу. Сколько различных результатов первенства может быть?

1.11°. Сколькими способами можно из 9 человек выбрать комиссию в составе 4 человек?

1.12°. Сколькими способами читатель может выбрать три книги из пяти?

1.13°. Сколькими способами можно разместить на полке 4 разные книги?

1.14. Пусть $p_1, p_2, \dots, p_n$ — различные простые числа. Сколько делителей имеет число

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_n^{\alpha_n},$$

где $\alpha_1, \alpha_2, \dots, \alpha_n$ — некоторые натуральные числа?

1.15. Сколько существует перестановок из n элементов, у которых два данные стоят рядом?

1.16°. Сколькими способами можно рассадить 4 учеников на 25 местах?

1.17°. Студенту необходимо на протяжении 8 дней сдать 4 экзамена. Сколькими способами это можно сделать?

1.18. Сколькими способами можно упорядочить множество $\{1, 2, 3, \dots, n\}$ так, чтобы числа 1, 2, 3 стояли рядом и в порядке возрастания?

1.19. Сколько существует четырехзначных чисел, у которых каждая последующая цифра больше предыдущей?

1.20. Сколько существует четырехзначных чисел, у которых каждая последующая цифра меньше предыдущей?

1.21*. В прямоугольную таблицу из m строк и n столбцов необходимо записать числа $+1$ и -1 так, чтобы произведение чисел в каждой строке и каждом столбце было равно 1. Сколькими способами это можно сделать?

1.22. Имеется p белых и q черных шаров ($p > q$). Сколькими способами можно расположить в ряд все шары так, чтобы никакие 2 черных шара не лежали рядом?

1.23. На плоскости проведено n прямых так, что никакие две из них не параллельны и никакие три не пересекаются в одной точке.

1) Найдите количество точек пересечения прямых.

2) Сколько треугольников образуют прямые?

3) На сколько частей делят плоскость прямые?

4) Сколько среди частей, на которые делится плоскость прямыми, ограниченных и сколько неограниченных?

1.24. Сколько диагоналей имеет выпуклый n -угольник?

1.25. В скольких точках пересекаются диагонали выпуклого n -угольника, если никакие три из них не пересекаются в одной точке?

1.26*. В выпуклом n -угольнике проведено все диагонали. Известно, что никакие три из них не пересекаются в одной точке. На сколько частей при этом разделится n -угольник?

1.27. Доказать, что

$$n(A_1 \times A_2 \times \dots \times A_k) = n(A_1)n(A_2) \dots n(A_k).$$

1.28. Доказать, что число способов упорядочить n -элементное множество равно $n!$

1.29. Доказать, что число A_n^k размещений из n элементов по k равно $n(n-1) \dots (n-(k-1))$.

1.30. Доказать, что C_n^k — число k -элементных подмножеств n -элементного множества равно $n!/(k!(n-k)!)$.

1.31. Доказать, что

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

1.32. Доказать, что число $C_n(k_1, k_2, \dots, k_m)$ способов разбить n -элементное множество на m непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \dots, k_m$ элементов, равно $n!/(k_1!k_2! \dots k_m!)$.

1.33. Доказать, что существует $C_n(k_1, k_2, \dots, k_m)$ слов длиной n из k_1 букв a_1, k_2 букв a_2 , и т. д., k_m букв a_m .

1.34. Сколькими способами можно разделить $m+n+s$ предметов на три группы так, чтобы в одной группе было m предметов, в другой — n , в третьей — s ?

1.35. Сколькими способами можно разделить $3n$ предметов между тремя лицами так, чтобы каждый получил n предметов?

1.36 Доказать, что выражение $(a_1 + a_2 + \dots + a_m)^n$ равно сумме всех возможных слагаемых вида

$$\frac{n!}{k_1!k_2! \dots k_m!} a_1^{k_1} a_2^{k_2} \dots a_m^{k_m},$$

где $k_1 + k_2 + \dots + k_m = n$, т. е.

$$\begin{aligned} & (a_1 + a_2 + \dots + a_m)^n = \\ = & \sum_{k_1 \geq 0, k_2 \geq 0, \dots, k_m \geq 0; k_1 + k_2 + \dots + k_m = n} \frac{n!}{k_1!k_2! \dots k_m!} a_1^{k_1} a_2^{k_2} \dots a_m^{k_m}. \end{aligned}$$

1.37. Доказать, что число f_m^n сочетаний из m элементов по n с повторениями равно C_{n+m-1}^{m-1} .

Глава 2

Стохастический эксперимент

2.1 Пространство элементарных событий, алгебра событий

Теория вероятностей изучает *стохастические эксперименты*, исследуя их математические модели. Под стохастическим экспериментом мы понимаем эксперимент, результат которого невозможно предсказать заранее (до проведения эксперимента), но который можно повторить независимым образом (результаты предыдущих экспериментов не влияют на последующие) в принципе неограниченное число раз.

Примеры стохастических экспериментов

1. Последовательно подбрасывают две монеты и регистрируют стороны, какими легла каждая монета. Результат эксперимента — “герб, герб”, “герб, решка”, “решка, герб”, “решка, решка” — предсказать заранее невозможно.

2. Подбрасывают игральную кость и регистрируют число выпавших очков. Результат эксперимента — появление $1, 2, \dots, 6$ — предсказать заранее невозможно.

3. Подбрасывают монету и регистрируют число подбрасываний до первого появления герба. Результат эксперимента — число подбрасываний: $0, 1, 2, \dots$ — заранее предсказать невозможно.

4. Регистрируют интервал времени до выхода из строя прибора. Результат — продолжительность безотказной работы прибора — заранее предсказать невозможно.

5. Частица участвует в броуновском движении — движется под воздействием ударов молекул жидкости (молекулы постоянно пребывают в хаотическом тепловом движении). Результат эксперимента — траекторию движения частицы — заранее предсказать невозможно.

Заметим, что каждый из перечисленных выше экспериментов можно независимым образом повторить неограниченное число раз.

Пространство элементарных событий. Каждому стохастическому эксперименту соответствует *пространство* (множество) его исходов. Пространство исходов будем обозначать через Ω , а сами исходы, как правило (но не обязательно), будем обозначать через ω (возможно с индексами). Исходы стохастического эксперимента еще называют элементарными событиями.

В примере 1 в качестве пространства исходов естественно рассматривать множество $\Omega = \{ГГ, ГР, РГ, РР\}$, в примерах 2, 3, 4 соответственно множества $\Omega = \{1, 2, \dots, 6\}$, $\Omega = \{0, 1, \dots\}$, $\Omega = [0, \infty)$, в примере 5 в качестве Ω естественно рассматривать множество траекторий частицы.

Реализацию стохастического эксперимента мы интерпретируем как случайный выбор точки ω из пространства Ω .

Пространство элементарных событий Ω будем называть *дискретным*, если множество Ω конечно или счетно, т. е. его элементы можно занумеровать числами $1, 2, \dots$

Алгебра событий. В каждом стохастическом эксперименте можно наблюдать определенную совокупность событий, мы их будем называть *случайными событиями* и будем обозначать $A, B, C, \dots$, совокупность всех событий, наблюдаемых в данном стохастическом эксперименте, будем обозначать через $\mathfrak{A}$. В примере 1 наблюдаемыми событиями являются A — “выпал хотя бы один герб”, B — “монеты легли одной стороной”, $\dots$

Каждое событие A стохастического эксперимента можно описать некоторым подмножеством пространства эле-

ментарных событий Ω , а именно:

$$A = \{\omega : \omega \in \Omega, \text{ которые влекут событие } A\}.$$

Поэтому в дальнейшем случайное событие мы будем отождествлять с подмножеством элементарных событий (исходов), которыми это событие описывается и будем обозначать их одной и той же буквой.

Если в результате проведения стохастического эксперимента точка ω попала в множество A , которое описывает событие A , будем говорить, что событие A произошло, в противном случае — не произошло.

В примере 1 событие “монеты легли разными сторонами” описывается подмножеством $\{ГР, РГ\}$ пространства $\Omega = \{ГГ, ГР, РГ, РР\}$, в примере 2 событие “выпало четное число очков” описывается подмножеством $\{2, 4, 6\}$ пространства $\Omega = \{1, 2, \dots, 6\}$, в примере 3 событие “эксперимент закончился до четвертого подбрасывания” описывается подмножеством $\{0, 1, 2, 3\}$ пространства $\Omega = \{0, 1, \dots\}$, в примере 4 событие “продолжительность безотказной работы прибора превысила 100 единиц времени” описывается подмножеством $(100, \infty)$ пространства $\Omega = [0, \infty)$.

Среди наблюдаемых событий стохастического эксперимента выделяются два — событие, которое происходит при каждой реализации стохастического эксперимента (оно называется *достоверным* и описывается множеством Ω), и событие, которое не происходит ни при одной реализации эксперимента (оно называется *невозможным* и описывается множеством $\emptyset$).

Пусть A и B — случайные события стохастического эксперимента, которые описываются соответственно подмножествами A и B пространства элементарных событий Ω .

Если каждый раз, когда происходит событие A , происходит и событие B , то будем говорить, что событие A влечет событие B и будем обозначать это так: $A \subset B$ (в терминах подмножеств, которыми описываются события, A является подмножеством B).

События A и B , такие что $A \subset B$ и $B \subset A$ (т. е. A и B происходят вместе), в теории вероятностей неразличимы. Их отождествляют и обозначают это так: $A = B$

(в терминах подмножеств, описывающих события, подмножества A и B равны).

Событие, состоящее в том, что происходит хотя бы одно из событий A или B , называется *суммой* (*объединением*)¹ событий A и B и обозначается так: $A \cup B$ (сумма событий A и B описывается объединением $A \cup B$ множеств A и B).

Событие, состоящее в том, что происходит как событие A , так и событие B , называется *произведением* (*пересечением*) событий A и B и обозначается так: $A \cap B$ (произведение событий A и B описывается пересечением $A \cap B$ множеств A и B).

Если события A и B такие, что $A \cap B = \emptyset$, то они называются *несовместными*.

Событие, состоящее в том, что A происходит, а B не происходит, называется *разностью* событий A и B и обозначается так: $A \setminus B$ (разность событий A и B описывается разностью $A \setminus B$ множеств A и B).

Событие, состоящее в том, что A не происходит, называется *противоположным* к событию A и обозначается $\bar{A}$ (событие $\bar{A}$ описывается дополнением $\bar{A}$ множества A до Ω).

Класс $\mathfrak{A}$ событий, который с каждым событием A содержит событие $\bar{A}$, с каждым двумя событиями A и B содержит событие $A \cup B$, будем называть *алгеброй событий*. Так что совокупность событий, которые наблюдаются в стохастическом эксперименте является алгеброй (подробнее относительно алгебры событий см. гл. 7).

Пусть в эксперименте 1 событие B — “монеты легли разными сторонами”, A — “в результате первого подбрасывания выпал герб”. События B и A как подмножества пространства элементарных событий $\Omega = \{\Gamma\Gamma, \Gamma P, P\Gamma, PP\}$ опишутся так: $B = \{\Gamma P, P\Gamma\}$, $A = \{\Gamma\Gamma, \Gamma P\}$. Тогда, согласно приведенным определениям $A \cap B = \{\Gamma P\}$ — “в результате первого подбрасывания выпал герб, а в результате второго — решетка”, $A \cup B = \{\Gamma\Gamma, \Gamma P, P\Gamma\}$ — “хотя бы один раз выпал герб”, $A \setminus B = \{\Gamma\Gamma\}$ — “выпало два герба”, $\bar{B} = \{\Gamma\Gamma, PP\}$ — “монеты легли одной стороной”.

¹Относительно операций над событиями см. разд. 7.1. в гл. 7.

Приведем еще одну иллюстрацию операций над событиями.

Пример 2.1.1 (диаграмма Венна). В квадрат с прямоугольниками (рис. 2.1.1) наудачу бросают точку. Если точка попала в “вертикальный” прямоугольник, говорим, что произошло событие A , а если в “горизонтальный” — событие B . События A , B , $\bar{A}$, $A \cup B$, $A \cap B$, $B \setminus A$ происходят, если точка попадает в соответствующую фигуру, изображенную на рис. 2.1.1.

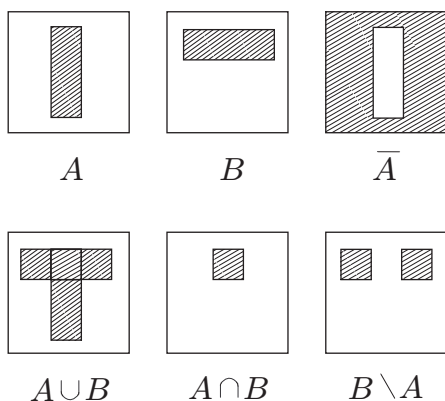


Рис. 2.1.1: Диаграмма Венна

2.2 Задачи

АЗ: 2.2°, 2.5, 2.8°, 2.11, 2.12, 2.15, 2.16, 2.17.

СЗ: 2.3°, 2.4, 2.6°, 2.7°, 2.10°, 2.13, 2.14, 2.18, 2.19.

З а м е ч а н и е. Термины “элементарное событие стохастического эксперимента”, “исход стохастического эксперимента”, “результат стохастического эксперимента” обозначают одно и то же.

2.1°. Указать события, противоположные к таким:

а) A — “выпадение герба в результате двух подбрасываний монеты”; б) B — “три попадания в результате трех

выстрелов по мишени”; в) C — “хотя бы одно попадание в результате трех выстрелов по мишени”.

2.2°. Сделано три выстрела по мишени. Пусть событие A_i состоит в том, что в результате i -го выстрела произошло попадание, $i = 1, 2, 3$. Выразить через A_i такие события: а) A — “произошло три попадания”; б) B — “не было ни одного попадания”; в) C — “произошло только одно попадание”; г) D — “произошло не менее двух попаданий”.

2.3°. Пусть A, B, C — случайные события. Записать события, состоящие в том, что не произошло ни одно из событий A, B, C ; из событий A, B, C произошло: а) только событие A ; б) события A и B и не произошло событие C ; в) все три события; г) хотя бы одно событие; д) одно и только одно событие; е) не более двух событий.

2.4. Радиолокационная станция ведет наблюдение за космическим объектом, при этом проведено n циклов осмотра. Обнаружение объекта в i -м цикле — случайное событие, обозначим его через $A_i, i = 1, 2, \dots, n$.

Выразить через $A_i, i = 1, 2, \dots, n$, такие события: A — “объект не будет обнаружен (ни в одном цикле)”; B — “объект будет обнаружен (хотя бы в одном цикле)”; C — “объект будет обнаружен только в одном цикле”.

2.5. Каждая из m радиолокационных станций, следящих за космическим объектом, делает n циклов наблюдений. Обозначим через A_{ji} событие “объект будет обнаружен j -й станцией в i -м цикле”, $j = 1, 2, \dots, m, i = 1, 2, \dots, n$.

Выразить через A_{ji} такие события: A — “объект будет обнаружен (хотя бы одной станцией)”; B — “объект будет обнаружен каждой станцией”.

2.6°. Дважды подбрасывают монету; предложить пространство элементарных событий Ω этого стохастического эксперимента. Описать как подмножества Ω события: A — “хотя бы один раз выпадет герб”, B — “при втором подбрасывании выпадет герб”. Найти число всех элементарных событий; число элементарных событий, принадлежащих A, B .

2.7°. Игральную кость подбрасывают дважды. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “сумма выпавших оч-

ков равна 8"; B — “хотя бы один раз выпала 6”; C — “при первом подбрасывании выпало четное число очков”; D — “при обоих подбрасываниях выпадет нечетное число очков”. Найти число элементарных событий в Ω , A , B , C , D .

2.8°. Подбрасывают монету, а затем игральную кость. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “выпадет герб”; B — “выпадет цифра 5”. Найти число элементарных событий в Ω , A , B .

2.9°. Пусть эксперимент состоит в измерении двух величин, принимающих значения из отрезка $[0; 1]$. Описать пространство элементарных событий.

2.10°. Из цифр 1, 2, 3, 4, 5 сначала выбирают одну, а затем из четырех оставшихся — другую. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “на первом шаге будет выбрана четная цифра”, B — “на втором шаге будет выбрана четная цифра”. Найти число элементарных событий в Ω , A , B .

2.11. Из партии, содержащей N изделий, среди которых M бракованных, наудачу выбирают n изделий. Предложить пространство элементарных событий Ω . Описать как подмножество Ω событие A — “среди выбранных n изделий имеется в точности m бракованных” ($n \leq N$, $m \leq M$, $m \leq n$). Найти число элементарных событий в Ω , A .

2.12. В лифте 7 пассажиров, лифт останавливается на 10 этажах. Для каждого пассажира фиксируется номер этажа, на котором он выходит. Предложить пространство элементарных событий Ω . Описать как подмножество Ω событие A — “все пассажиры выходят на разных этажах”. Найти число элементов в Ω , A .

2.13. Два игрока по очереди подбрасывают монету. Выигрывает тот, у кого впервые выпадет герб. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “игра закончится на k -м подбрасывании”, B — “игра закончится до k -го подбрасывания”, C — “выиграет начавший игру первым”, D — “выиграет начавший игру вторым”.

2.14. В урне содержится один шар, о котором известно, что он белый или черный. В урну положили белый

шар, а затем после тщательного перемешивания извлекли один за другим оба шара. Предложить пространство элементарных событий Ω этого стохастического эксперимента. Описать как подмножества Ω события: A — “первый раз извлечен белый шар”, B — “второй раз извлечен белый шар”.

2.15. Игральную кость последовательно подбрасывают n раз. Предложить пространство элементарных событий Ω . Описать как подмножества Ω событие A — “выпадет n_1 единиц, n_2 двоек, и т. д., n_6 шестерок” ($n_1 + n_2 + \dots + n_6 = n$). Вычислить количество элементов в Ω, A .

2.16. В турнире участвуют $2n$ человек, которые по жребию разделены на две группы по n человек в каждой. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “двое наиболее сильных игроков будут играть в разных группах”, B — “четверо наиболее сильных игроков будут играть по двое в разных группах” (все игроки различаются по силе). Вычислить число элементов в Ω, A, B .

2.17. В наудачу выбранной группе, насчитывающей r ($r \leq 12$) студентов, интересуемся месяцами их рождения. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “хотя бы два студента родились в одном месяце”, B — “только один студент родился в сентябре”, C — “хотя бы один студент родился в сентябре”, D — “ни один студент не родился в сентябре”. Найти число элементов в Ω, A, B, C, D .

2.18. Числа $1, 2, \dots, n$ располагают наудачу. Каждому из них приписывают номер места, на котором расположено число. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “число 1 получит номер 1”, B — “число 1 получит номер 1, а число n — номер n ”. Найти число элементов в Ω, A, B .

2.19. Подбрасывают три игральные кости. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “единица выпадет только на одной кости”, B — “шестерка выпадет только на двух костях”, C — “на всех костях выпадут разные грани”, D — “на костях выпадет одинаковое число очков”. Найти число элементов в Ω, A, B, C, D .

2.20. Монету и игральную кость подбрасывают по очереди неограниченное число раз. Предложить простран-

ство элементарных событий Ω . Описать как подмножество Ω события: A — “герб выпадет раньше шестерки”, B — “пятерка выпадет раньше герба”.

2.21*. Каждая из n различных частиц попадает в одну из m ячеек. Предложить пространство элементарных событий Ω . Описать как подмножество Ω событие: A — “в первую ячейку попало k_1 частиц, во вторую — k_2 , и т. д., в m -ю — k_m ”. Найти число элементов в Ω , A .

2.22*. Каждая из n неразличимых частиц попадает в одну из m ($n \geq m$) ячеек. Предложить пространство элементарных событий Ω . Описать как подмножество Ω событие: A — “каждая из ячеек содержит по меньшей мере одну частицу”. Найти число элементов в Ω , A .

2.23. В поезде m вагонов. Каждый из n ($n > m$) пассажиров выбирает себе вагон. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события:

A — “в каждом вагоне будет по меньшей мере один пассажир”,

B — “будет занято ровно r вагонов” (вагон занят, если в нем находится хотя бы один пассажир).

2.24. Девять пассажиров садятся в три вагона. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “в каждый вагон сядет по три пассажира”, B — “в один вагон сядут четыре, в другой — три, в третий — два пассажира”.

2.25. Студент пришел на экзамен, зная 35 из 40 вопросов программы. Экзаменатор задает студенту 4 вопроса. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “студент знает ответы на все вопросы”, C — “студент не знает ответа ни на один вопрос”, B_i — “студент знает ответы на i вопросов” ($i = 1, 2, 3$), D — “студент сдаст экзамен” (чтобы сдать экзамен, необходимо ответить не менее чем на 3 вопроса).

2.26. Участник лотереи “Спортлото” из 49 названий видов спорта (обозначенных числами 1–49) должен назвать 6. Предложить пространство элементарных событий Ω . Описать как подмножества Ω события: A — “участник правильно укажет все шесть названий”, B — “участник получит выигрыш (для получения выигрыша необходимо правильно указать не менее чем 3 вида)”.

2.27. Два лица, договорившись встретиться в течение часа, решили, что каждый независимо от другого приходит на место встречи в наудачу выбранный момент указанного часа. Предложить пространство элементарных событий Ω . Описать как подмножество Ω события: A — “встреча состоится, если каждый будет ждать другого на протяжении времени t ($t < 1$), после чего уходит с места встречи”, B — “данное лицо придет на место встречи раньше другого”, C — “данное лицо придет на место встречи раньше другого на время, не меньшее, чем q ($q < 1$)”.

2.28. Наудачу берут три отрезка, длина которых не превышает 1. Предложить пространство элементарных событий Ω . Описать как подмножества Ω событие: A — “из отрезков можно построить треугольник”.

2.29. Числа $1, 2, \dots, n$ расположены наудачу. Предложить пространство элементарных событий Ω . Описать как подмножество Ω событие A — числа 1 и 2 расположены рядом. Найти число элементов в Ω и A .

Глава 3

Дискретное вероятностное пространство

3.1 Вероятность, классическая модель

Частота события. Вероятность. Опыт свидетельствует, что в стохастическом эксперименте события различаются частотой своего появления (одни наблюдаются чаще, другие — реже). Например, при подбрасывании игральной кости событие B — “шестерка не выпала” происходит чаще, чем событие C — “шестерка выпала”. При подбрасывании симметричной монеты до первого появления герба событие B — “герб впервые выпадет при первом подбрасывании” происходит чаще, чем событие C “герб впервые выпадет при одиннадцатом подбрасывании”.

Количественной мерой частоты появления события A является *частота* $\nu_n(A)$ события A в последовательности n экспериментов. Частота определяется так. Проведем стохастический эксперимент независимым образом n раз; пусть $k_n(A)$ — число тех экспериментов, в которых произошло (наблюдалось) событие A , тогда

$$\nu_n(A) = k_n(A)/n.$$

Частота $\nu_n(A)$ обладает следующими свойствами.

1. Для каждого события A

$$\nu_n(A) \geq 0. \quad (3.1.1)$$

2. Для несовместных событий A и B

$$\nu_n(A \cup B) = \nu_n(A) + \nu_n(B). \quad (3.1.2)$$

3. Для достоверного события Ω

$$\nu_n(\Omega) = 1. \quad (3.1.3)$$

Частота $\nu_n(A)$ события A в последовательности экспериментов является устойчивой — колеблется около некоторого числа, причем с ростом n значительные отклонения $\nu_n(A)$ от этого числа встречаются все реже — частота $\nu_n(A)$ стабилизируется.

Математической моделью частоты является вероятность.

Вероятность — это функция

$$P : A \rightarrow P(A),$$

заданная на классе событий, такая, что

1. Для каждого события A

$$P(A) \geq 0. \quad (3.1.4)$$

2. Для непересекающихся событий A_i , $i = 1, 2, \dots$,
($A_i \cap A_j$, $i \neq j$)

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (3.1.5)$$

3. Для достоверного события Ω

$$P(\Omega) = 1. \quad (3.1.6)$$

Значение $P(A)$ функции P на A называют вероятностью события A .

Вероятность на дискретном пространстве. Пару $\{\Omega, P\}$, где Ω — дискретное пространство элементарных событий стохастического эксперимента, а P — вероятность на классе событий (подмножеств Ω), будем называть *дискретным вероятностным пространством*.

В дискретном пространстве $\{\Omega, P\}$ каждое событие A можно представить в виде объединения (не более чем счетного числа) исходов ω :

$$A = \bigcup_{\omega \in A} \{\omega\},$$

поэтому

$$P(A) = \sum_{\omega \in A} P(\omega), \quad (3.1.7)$$

в частности

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\omega).$$

Равенство (3.1.7) означает, что в дискретном вероятностном пространстве вероятность события A равна сумме вероятностей $P(\omega)$ исходов ω , описывающих событие A . И, следовательно, чтобы задать вероятность на дискретном пространстве Ω , достаточно задать вероятности исходов $P(\omega)$, $\omega \in \Omega$; при этом имеет место равенство (3.1.7), в общем случае (не дискретного Ω) это не так.

Дискретное вероятностное пространство $\{\Omega, P\}$ является математической моделью стохастического эксперимента с конечным или счетным множеством исходов.

Примечание. Вероятность $p = P(A)$ события A можно интерпретировать как количественную меру прогноза его появления: частота появления события A в длинной серии независимых экспериментов близка к ее вероятности p .

Пример 3.1.1. Подбрасывают игральную кость, масса которой распределена так, что частота появления определенной грани пропорциональна ее номеру (числу очков на ней). Построить вероятностное пространство

этого стохастического эксперимента. Описать как подмножества Ω события: A — “выпадет число очков, кратное 3”, B — “выпадет четное число очков”. Вычислить их вероятности.

Подбрасывают симметричную игральную кость (все грани выпадают одинаково часто). Предложить вероятностное пространство этого стохастического эксперимента. Вычислить вероятности событий A и B .

Решение. В качестве пространства элементарных событий первого из описанных стохастических экспериментов естественно рассматривать $\Omega = \{1, 2, \dots, 6\}$. Вероятность появления единицы обозначим p , тогда вероятность появления грани с номером j равна jp , и поскольку

$$\sum_{\omega \in \Omega} P(\omega) = 1,$$

то $p + 2p + \dots + 6p = 1$. Отсюда $p = 1/21$, $P(j) = j/21$, $j = 1, 2, \dots, 6$. Так что вероятностное пространство $\{\Omega, P\}$ стохастического эксперимента построено.

События A и B опишутся соответственно подмножествами $A = \{3; 6\}$ и $B = \{2; 4; 6\}$ пространства Ω .

Вероятности событий A и B (как вероятности событий в дискретном вероятностном пространстве, см. (3.1.7)) вычисляются так:

$$P(A) = P(\{3; 6\}) = P(3) + P(6) = \frac{3}{21} + \frac{6}{21} = \frac{3}{7};$$

$$P(B) = P(\{2; 4; 6\}) = P(2) + P(4) + P(6) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{4}{7}.$$

При подбрасывании симметричной игральной кости пространство элементарных событий будет таким же: $\Omega = \{1, 2, \dots, 6\}$, а вероятности элементарных событий естественно задать следующим образом:

$$P(i) = 1/6, i = 1, 2, \dots, 6.$$

Тогда

$$P(A) = P(\{3; 6\}) = P(3) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3};$$

$$P(B) = P(\{2; 4; 6\}) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

З а м е ч а н и е. Заметим, что здесь фактически решаются две задачи (и это стандартная ситуация).

Задача 1. Строится математическая модель стохастического эксперимента (вероятностное пространство), т. е. выбирается пространство элементарных событий Ω и задается вероятность каждого элементарного события:

$$P(i) = i/21, i = 1, 2, \dots, 6,$$

(для несимметричной кости) и

$$P(i) = 1/6, i = 1, 2, \dots, 6,$$

(для симметричной кости). При этом необходимо заметить, что теория вероятностей никаких указаний и рекомендаций относительно задания вероятностей элементарных событий не дает — ни из каких теорем не следует, что, скажем, для симметричной игральной кости

$$P(i) = 1/6, i = 1, 2, \dots, 6.$$

И каждый раз, когда (из тех или иных соображений) вероятностное пространство выбрано, вопрос о его адекватности стохастическому эксперименту остается открытым (ответить на него помогает, в частности, математическая статистика).

Заметим, что построение (выбор) вероятностного пространства не является задачей теории вероятностей.

Задача 2. После выбора вероятностного пространства $\{\Omega, P\}$ вычисляют вероятности тех или иных событий (пользуясь теоремами теории вероятностей).

Классическая модель. Дискретное вероятностное пространство $\{\Omega, P\}$, все элементарные события ω_i которого равновероятны, т. е. $P(\omega_i) = P(\omega_j)$ для всех i, j , будем называть *классической моделью*.

В классической модели $\{\Omega, P\}$ пространство элементарных событий Ω конечно, вероятность $P(\omega_i)$ элементарного события $\omega_i \in \Omega$ равна $1/n(\Omega)$, т. е.

$$P(\omega_i) = \frac{1}{n(\Omega)}, i = 1, 2, \dots, n(\Omega),$$

и для любого события A

$$P(A) = \frac{n(A)}{n(\Omega)}. \quad (3.1.8)$$

Последняя формула называется *формулой классической вероятности*.

Пример 3.1.2. *Симметричную игральную кость подбрасывают шесть раз. Построить математическую модель этого стохастического эксперимента. Описать событие A — “выпадут все шесть граней” и вычислить его вероятность.*

Решение. В качестве пространства элементарных событий стохастического эксперимента естественно рассматривать множество упорядоченных последовательностей, составленных из шести чисел (от 1 до 6). Например, последовательность $(6, 1, 6, 3, 2, 4)$ описывает исход стохастического эксперимента, состоящий в том, что при первом подбрасывании игральной кости выпала 6, при втором — 1, и т. д. при шестом — 4.

Событие A — “выпали все грани” опишется подмножеством множества исходов Ω , состоящим из последовательностей, в записи которых встречается каждая из цифр $1, 2, \dots, 6$.

Далее, поскольку каждый из исходов стохастического эксперимента ничем не лучше и не хуже другого (кость симметричная), то естественно считать, что все элементарные события равновероятны (ни из одной теоремы это утверждение не следует). Другими словами, в качестве математической модели данного стохастического эксперимента будем рассматривать классическую модель. Тем самым вероятностное пространство (модель стохастического эксперимента) построено — предложено Ω и для каждого $\omega \in \Omega$ задана его вероятность $P(\omega)$:

$$P(\omega) = 1/n(\Omega).$$

Теперь вычислим вероятность события A — “выпадут все шесть граней”. Поскольку модель классическая, то согласно формуле классической вероятности (3.1.8), вероятность события A равна отношению числа $n(A)$ эле-

ментарных событий, входящих в событие A , к числу $n(\Omega)$ всех элементарных событий Ω .

Элементарных событий в Ω столько, сколько последовательностей длиной 6, которые можно построить из шести чисел (от 1 до 6). Согласно правилу умножения их 6^6 . Элементарных событий, входящих в событие A , столько, сколько имеется последовательностей, которые можно построить из различных чисел от 1 до 6 (сколько существует перестановок из чисел $1, 2, \dots, 6$) — их $6!$

Следовательно, искомая вероятность

$$P(A) = \frac{6!}{6^6}.$$

3.2 Задачи

АЗ: 3.1, 3.5, 3.6, 3.7, 3.14, 3.16, 3.21, 3.27, 3.28*.

СЗ: 3.2, 3.9, 3.10, 3.11, 3.18, 3.19, 3.20, 3.21, 3.22.

В предложенных далее задачах прежде чем вычислять вероятность того или иного случайного события A , необходимо построить математическую модель стохастического эксперимента, т. е. задать пространство Ω элементарных событий ω и их вероятности $P(\omega)$, $\omega \in \Omega$. Затем описать событие A как подмножество Ω и вычислить его вероятность в дискретном вероятностном пространстве $\{\Omega, P\}$.

Примечание. Задачи с подбрасыванием игральных костей, монет следует решать в предположении, что кости и монеты симметричные (если не оговорено иное).

3.1. В наудачу выбранной группе, насчитывающей r студентов (например, академическая группа), интересуемся их датами рождения. Вычислить вероятность события A , состоящего в совпадении дат рождения хотя бы у двух студентов.

3.2. В наудачу выбранной группе, насчитывающей r студентов, интересуемся месяцами их рождения. Вычислить вероятность события A , состоящего в том, что хотя бы два студента родились в одном и том же месяце.

3.3. Для уменьшения общего количества игр $2n$ команд (разных по силе) по жребии делят на две подгруппы по n команд каждая.

1. Вычислить вероятность того, что две самые сильные команды окажутся: а) в разных подгруппах; б) в одной подгруппе?

2. Вычислить вероятность того, что четыре самые сильные команды окажутся: а) в разных подгруппах (по две в каждой); б) в одной подгруппе; в) в разных подгруппах, причем в одной подгруппе 3 команды, в другой — 1?

3.4°. В коробке содержится пять одинаковых пронумерованных кубиков. Наудачу по одному вынимают все кубики. Вычислить вероятность того, что номера кубиков появятся в порядке возрастания.

3.5. Числа $1, 2, \dots, n$ упорядочивают наудачу. Вычислить вероятность того, что 1 и 2 будут расположены рядом.

3.6. Студент пришел на экзамен, зная 20 из 25 вопросов программы. Экзаменатор задает студенту три вопроса. Вычислить вероятность того, что студент: 1) знает ответы на все вопросы; 2) знает ответы на два вопроса; 3) сдаст экзамен (если для этого необходимо ответить не менее чем на два вопроса); 4) не сдаст экзамен.

3.7. Игральную кость подбрасывают 6 раз. Вычислить вероятность того, что выпадут только четные грани.

3.8. Из карточек разрезной азбуки составлено слово “статистика”. Потом из этих десяти карточек наудачу выбирают семь. Вычислить вероятность того, что из выбранных карточек можно составить слово “статика”.

3.9. Ребенок играет десятью буквами разрезной азбуки А, А, А, Е, И, К, М, М, Т, Т. Вычислить вероятность того, что при случайном расположении букв в ряд образуется слово “математика”.

3.10. Пусть n человек, среди которых присутствуют A и B , строятся в ряд в произвольном порядке. Вычислить вероятность того, что между A и B будет стоять ровно r человек.

3.11. В шахматном турнире участвуют 20 человек (разных по силе), которых по жребии делят на две группы по 10 человек каждая. Вычислить вероятность того,

что: а) двое наиболее сильных участников будут играть в разных группах; б) четверо наиболее сильных участников будут играть по два в разных группах?

3.12. Найти вероятность того, что среди двух чисел, выбранных наудачу из последовательности $1, 2, \dots, n$, одно окажется меньшим, а другое большим заданного числа k ($1 < k < n$).

3.13. Девять пассажиров располагаются в трех вагонах. Вычислить вероятность того, что:

а) в каждый вагон сядет по три пассажира;

б) в один вагон сядут четыре, в другой — три, в третий — два пассажира?

3.14. В лифте 7 пассажиров. Лифт останавливается на десяти этажах. Найти вероятность того, что никакие два пассажира не выйдут на одном и том же этаже.

3.15°. Вычислить вероятность того, что даты рождения 12 человек придутся на разные месяцы года.

3.16. Среди N изделий M бракованных. Наудачу берут n изделий. Найти вероятность того, что среди них m бракованных ($m < M$)? Какова вероятность того, что среди них более чем m бракованных?

3.17. В лотерее n билетов, среди них m выигрышных. Найти вероятность выигрыша для владельца r билетов.

3.18. На экзамене предлагается N вопросов. Студент знает ответы на n вопросов. Экзаменатор задает студенту k вопросов, а для того чтобы сдать экзамен, необходимо ответить не менее чем на r ($r < k$) вопросов. Найти вероятность того, что студент сдаст экзамен?

3.19. Участник лотереи “Спортлото” из 49 названий видов спорта (обозначенных числами $1, 2, \dots, 49$) должен назвать 6. Полный выигрыш получает тот, кто правильно укажет все шесть названий. Выигрыш получают и те, кто угадает не менее трех названий. Найти вероятность полного выигрыша в “Спортлото”. Вычислить вероятность того, что участник “Спортлото” угадает 5, 4, 3 названий. Какова вероятность получить выигрыш в “Спортлото”?

3.20. Подбрасывают 12 игральные кости. Какова вероятность того, что каждое из чисел 1, 2, 3, 4, 5, 6 выпадет дважды?

3.21. Подбрасывают n игральные кости. Какова вероятность того, что выпадут n_1 единиц, n_2 двоек, $\dots$ n_6 шестерок ($n_1 + n_2 + \dots + n_6 = n$)?

3.22°. Четырехтомное собрание сочинений расставляют на полке наудачу. Вычислить вероятность того, что тома расположатся в порядке возрастания их номеров.

3.23°. Дважды подбрасывают симметричную монету. Предложить вероятностное пространство этого стохастического эксперимента. Описать события: A — “при первом подбрасывании выпадет герб”, B — “при втором подбрасывании выпадет герб”. Вычислить вероятности $P(A)$, $P(B)$, $P(A \cap B)$.

3.24°. Трижды подбрасывают симметричную монету. Предложить вероятностное пространство этого стохастического эксперимента. Описать события: A — “дважды выпадет герб”, B — “хотя бы один раз выпадет герб”. Вычислить вероятности $P(A)$, $P(B)$, $P(A \cap B)$, $P(B/A)$.

3.25. Подбрасывают n игральных костей. Вычислить вероятность того, что на всех игральные кубиках появится одинаковое число очков.

3.26°. Подбрасывают шесть игральных костей. Вычислить вероятность того, что суммарное число очков на костях будет равно 7.

3.27. Пусть n лиц садятся в ряд наудачу. Какова вероятность того, что два определенных лица окажутся рядом? Вычислить вероятность того же события, если лица садятся за круглый стол.

3.28*. Доказать, что вероятнее получить хотя бы одну единицу при подбрасывании четырех игральных костей, чем в случае 24 подбрасываний двух костей получить хотя бы один раз две единицы.

Примечание. Задача известна как парадокс “де Мере”. Придворный кавалер и азартный игрок шевалье де Мере, современник Блеза Паскаля, считал, что вероятности этих событий одинаковы, и обвинял математиков в своих проигрышах.

3.29. Доказать, что в классической модели $\{\Omega, P\}$ пространство элементарных событий Ω конечно и для каждого элементарного события $\omega_i \in \Omega$ значение вероятности $P(\omega_i) = 1/n(\Omega)$, $i = 1, 2, \dots, n(\Omega)$, где $n(\Omega)$ — число всех элементарных событий.

3.30. Доказать, что в классической модели $\{\Omega, P\}$ для любого события A (подмножества A пространства Ω)

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

где $n(\Omega)$ — число всех элементарных событий, а $n(A)$ — число элементарных событий, входящих в состав A .

3.31 (статистика Максвелла-Больцмана). Каждая из n различных частиц попадает в одну из m ячеек.

Найти вероятность того, что в первой, второй, и т. д., m -й ячейке будет соответственно $k_1, k_2, \dots, k_m$ частиц?

3.32. За круглым столом располагаются n лиц. Вычислить вероятность того, что три определенные лица окажутся рядом.

3.33. Подбрасывают три монеты. Найти вероятность событий: A — “на первой монете выпадет герб”, B — “выпадет ровно два герба”, C — “выпадет не более двух гербов”.

3.34. Из множества всех последовательностей длиной n , составленных из цифр $0, 1, 2$, наудачу выбирается одна. Найти вероятности событий: A — “последовательность начинается с 0”, B — “последовательность содержит $m + 2$ нулей, причем два из них — на концах последовательности”, C — “последовательность содержит m единиц”, D — “последовательность составлена из m_0 нулей, m_1 единиц, m_2 двоек”.

3.35. Подбрасывают две игральные кости. Найти вероятности событий: A — “сумма очков кратна 6”, B — “сумма очков кратна 2”, C — “произведение очков четное число”.

3.36. Вычислить вероятность того, что четырехзначный номер наудачу выбранного в большом городе автомобиля: а) состоит из разных цифр; б) имеет только две одинаковые цифры; в) имеет две пары одинаковых цифр; г) имеет только три одинаковые цифры; д) состоит из одинаковых цифр.

3.37. Из урны, содержащей M_1 шаров с номером 1, M_2 шаров с номером 2, и т. д., M_N шаров с номером N , наудачу извлекают n шаров. Вычислить вероятность события A — “появится m_1 шаров с номером 1, m_2 шаров

с номером 2, и т. д., появится m_N шаров с номером N ”, B — “каждый из N номеров встретится хотя бы один раз”.

3.38. Из множества $\{1, 2, \dots, N\}$ последовательно наудачу без возвращения выбирают два числа ξ_1 и ξ_2 . Вычислить вероятность события $\{\xi_1 < \xi_2\}$.

Из множества $\{1, 2, \dots, N\}$ последовательно наудачу без возвращения выбирают три числа ξ_1, ξ_2, ξ_3 . Вычислить вероятность того, что второе число будет расположено между первым и третьим, т. е. вычислить вероятность события $\{\xi_1 < \xi_2 < \xi_3\}$.

3.39. Куб, все грани которого покрашены, распилили на тысячу кубиков одинакового размера. Кубики перемешали. Найти вероятность того, что наудачу выбранный кубик: а) имеет хотя бы одну окрашенную грань; б) имеет две окрашенных грани; в) имеет три окрашенных грани.

3.40 (призовая игра). Вы участвуете в призовой игре (приз — автомобиль). Перед вами три закрытых двери. За одной из них автомобиль. Ведущий предлагает вам выбрать дверь. Вы выбираете. Прежде чем открыть выбранную вами дверь ведущий открывает одну из оставшихся дверей — за ней автомобиля нет. После этого ведущий предлагает вам два варианта последующих действий: открыть выбранную ранее вами дверь или открыть оставшуюся дверь. Если за открытой дверью автомобиль — он ваш.

Какой выбор сделать — открыть первоначально выбранную дверь или открыть оставшуюся?

Естественно, что предпочтительнее тот выбор, для которого вероятность выигрыша автомобиля больше.

Вычислите вероятность выигрыша автомобиля, если: 1) вы открываете первоначально выбранную дверь; 2) вы открываете оставшуюся дверь.

Задача и ее решение становятся особенно “прозрачными”, если автомобиль находится за одной из n дверей (пусть для наглядности $n = 1000$).

Вы выбираете дверь, после чего ведущий демонстрирует, что за $n - 2$ дверями из $n - 1$ оставшихся автомобиля нет. Теперь выбор за вами — открыть первоначально выбранную дверь или открыть оставшуюся дверь.

Глава 4

Условная вероятность

4.1 Формула полной вероятности. Формулы Байеса

Условная вероятность. Пусть $\{\Omega, P\}$ — вероятностное пространство. *Условной вероятностью* $P(A/B)$ события A относительно события B ($P(B) > 0$) будем называть величину $\frac{P(A \cap B)}{P(B)}$, т. е.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}. \quad (4.1.1)$$

Из (4.1.1) имеем

$$P(A \cap B) = P(A/B)P(B), \quad (P(B) > 0)$$

Это равенство называют *формулой умножения*.

В классической модели $\{\Omega, P\}$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{n(A \cap B)/n(\Omega)}{n(B)/n(\Omega)} = \frac{n(A \cap B)}{n(B)},$$

т. е. условная вероятность $P(A/B)$ события A относительно события B равна отношению числа $n(A \cap B)$ элементарных событий, входящих в $A \cap B$, к числу $n(B)$ элементарных событий, входящих в B .

Примечание. Условную вероятность $p = P(A/B)$ события A относительно события B можно интерпретировать как количественную меру прогноза появления A , когда дополнительно известно, что произошло событие B .

Пример 4.1.1. *Подбрасывают три симметричные игральные кости. Найти вероятность того, что хотя бы на одной выпадет единица, если известно, что на трех выпали разные грани.*

Решение. Кости будем считать различимыми. В качестве пространства элементарных событий Ω рассмотрим упорядоченные тройки, составленные из чисел $1, 2, \dots, 6$. Поскольку кости симметричны, то естественно считать все элементарные исходы равновероятными. Тем самым вероятностное пространство стохастического эксперимента построено.

Пусть B — событие “на трех костях выпали разные грани”, A — “хотя бы на одной кости выпала единица”. В задаче необходимо вычислить условную вероятность $P(A/B)$.

Имеем

$$P(A/B) = \frac{P(A \cap B)}{P(B)},$$

$$P(B) = \frac{6 \cdot 5 \cdot 4}{6^3}; \quad P(A \cap B) = \frac{C_3^1 \cdot 5 \cdot 4}{6^3};$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{C_3^1 \cdot 5 \cdot 4}{6^3} / \frac{6 \cdot 5 \cdot 4}{6^3} = \frac{1}{2}.$$

Вероятность $P(A/B)$ также можно вычислить, воспользовавшись тем, что в классической модели $\{\Omega, P\}$ условная вероятность $P(A/B)$ равна отношению числа элементарных событий, принадлежащих $A \cap B$, к числу элементарных событий, принадлежащих B , т. е.

$$P(A/B) = \frac{C_3^1 \cdot 5 \cdot 4}{6 \cdot 5 \cdot 4} = \frac{1}{2}.$$

Пример 4.1.2 (задача Льюиса Кэррола). *В урне содержится один шар, о котором известно, что он или белый (с вероятностью $1/2$), или черный. В урну положили белый шар, и затем, тщательно перемешав шары,*

извлекли наудачу один, который оказался белым. Какова вероятность того, что после этого из урны извлекут белый шар?

Решение 1. Эксперимент состоит в последовательном извлечении из урны двух шаров. Обозначим белый шар через W , черный — B . Множеством всех исходов стохастического эксперимента является $\Omega = \{WW, WB, BW\}$. Например, пара WB описывает исход: “первым извлечен белый шар, вторым — черный”. Пусть A_1 — событие “белый шар извлечен первым”, A_2 — “белый шар извлечен вторым”. Необходимо вычислить $P(A_2/A_1)$. События $A_1, A_2, A_1 \cap A_2$ как подмножества Ω опишутся так: $A_1 = \{WW, WB\}$, $A_2 = \{WW, BW\}$, $A_1 \cap A_2 = \{WW\}$. Поскольку в Ω входят три элементарные события, в A_1 — два, в $A_1 \cap A_2$ — одно, то

$$P(A_2/A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{1/3}{2/3} = \frac{1}{2}.$$

Решение 2. Белый шар, который кладут в урну, пометим, например, звездочкой и обозначим его через W^* . Тогда множеством всех исходов стохастического эксперимента является $\Omega^* = \{WW^*, W^*W, BW^*, W^*B\}$. События $A_1, A_2, A_1 \cap A_2$ как подмножества Ω^* опишутся так: $A_1 = \{WW^*, W^*W, W^*B\}$, $A_2 = \{WW^*, W^*W, BW^*\}$, $A_1 \cap A_2 = \{WW^*, W^*W\}$. Поэтому

$$P(A_2/A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{2/4}{3/4} = \frac{2}{3}.$$

Два решения — два разных ответа. Не хорошо.

Вопрос 1. Какое из решений неверное?

Вопрос 2. Где допущена ошибка?

Попробуйте сначала ответить на вопросы 1 и 2, не читая дальше.

Ответ к примеру 4.1.2.

Решение 1 неверное. Модель не классическая, а вероятности событий вычисляются как в классической модели. Вероятности элементарным событиям необходимо приписать так:

$$P(WW) = 1/2, P(WB) = 1/4, P(BW) = 1/4.$$

Поскольку сначала в урне находился белый шар (с вероятностью $1/2$) или черный (с вероятностью $1/2$), то после того, как в урну положили один белый шар, в урне находятся с вероятностью $1/2$ шары одного цвета и с вероятностью $1/2$ — разного. В этой модели имеем:

$$\begin{aligned} P(A_1 \cap A_2) &= P(WW) = 1/2, \\ P(A_1) &= P(\{WW, WB\}) = \\ &= P(WW) + P(WB) = 1/2 + 1/4 = 3/4, \\ P(A_2/A_1) &= \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{1/2}{3/4} = \frac{2}{3}. \end{aligned}$$

В решении 2 каждому исходу приписываем вероятность $1/4$. Заметим, что в урне, после того, как в нее положили один белый шар и до того, как шары начали извлекать, находилось: два белых шара (с вероятностью $1/2$), или один белый, другой черный (с вероятностью $1/2$).

Комментарий к примеру. Пример, в частности, иллюстрирует тот факт, что для одного и того же стохастического эксперимента можно предложить разные пространства элементарных событий, а вместе с ними и разные вероятностные пространства, которые адекватно описывают стохастический эксперимент. В примере было предложено два пространства элементарных событий: $\Omega = \{WW, WB, BW\}$ и $\Omega^* = \{WW^*, W^*W, BW^*, W^*B\}$. Распределение вероятностей на пространстве элементарных событий Ω такое: $P(WW) = 1/2$, $P(WB) = 1/4$, $P(BW) = 1/4$, а на пространстве Ω^* такое: $P(WW^*) = 1/4$, $P(W^*W) = 1/4$, $P(W^*B) = 1/4$, $P(BW^*) = 1/4$. Обе модели адекватно описывают стохастический эксперимент. Заметим, что одна из них классическая, другая — нет.

Формула полной вероятности. События $B_i \subset \Omega$, $i = 1, 2, \dots, n$, будем называть *полной группой событий*, если они попарно не пересекаются ($B_i \cap B_j = \emptyset, i \neq j$) и в объединении дают достоверное событие ($\bigcup_{i=1}^n B_i = \Omega$).

Теорема. Пусть $B_1, B_2, \dots, B_n$ — полная группа событий и $P(B_i) > 0, i = 1, 2, \dots, n$. Тогда для любого события A имеет место формула полной вероятности

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

Формула полной вероятности справедлива и для счетной полной группы событий.

Пример 4.1.3 (“счастливые” билеты). Среди N экзампнационных билетов имеется n “счастливых” (все студенты их знают). Студенты один за другим подходят за билетами. У кого больше вероятность вытянуть “счастливый” билет: у того, кто подошел первым, или у того, кто подошел вторым?

Решение. Обозначим через A_i^l событие “ i -й студент вытянул счастливый билет”, $i = 1, 2$, а через A_1^u — “1-й студент вытянул несчастливый билет”. Далее вычисляем $P(A_1^l)$ и $P(A_2^l)$.

Согласно формуле классической вероятности

$$P(A_1^l) = \frac{n}{N}.$$

Вероятность события $P(A_2^l)$ вычисляем по формуле полной вероятности, учитывая, что A_1^l и A_1^u образуют полную группу событий:

$$\begin{aligned} P(A_2^l) &= P(A_2^l/A_1^l)P(A_1^l) + P(A_2^l/A_1^u)P(A_1^u) = \\ &= \frac{n-1}{N-1} \frac{n}{N} + \frac{n}{N-1} \frac{N-n}{N} = \frac{n}{N}. \end{aligned}$$

Следовательно, вероятность вытянуть “счастливый” билет для обоих студентов одинакова.

Примечание. Такой ответ получается, если первый студент не говорит, какой он вытянул билет. В противном случае вероятность того, что “счастливый” билет вытянул второй студент, $P(A_2^l/A_1^l) = (n-1)/(N-1)$, если первый вытянул “счастливый” билет (и объявил об этом),

и $P(A_2^l/A_1^u) = n/(N-1)$ — если первый вытянул “несчастливый” билет.

Формулы Байеса. Пусть $B_1, B_2, \dots, B_n$ — полная группа событий и $P(B_i) > 0, i = 1, 2, \dots, n$. Тогда для любого события A ($P(A) > 0$)

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{k=1}^n P(A/B_k)P(B_k)}, \quad i = 1, 2, \dots, n.$$

Пример 4.1.4 (сигнал на фоне шума). На фоне шума на вход радиолокационного устройства с вероятностью p ($0 < p < 1$) поступает сигнал. Если поступил сигнал (с шумом), то устройство регистрирует наличие сигнала с вероятностью p_1 , если только шум, то устройство регистрирует наличие сигнала с вероятностью p_2 . Известно, что устройство зарегистрировало сигнал. Какова вероятность того, что на вход радиолокационного устройства поступил сигнал?

Решение. Введем такие обозначения: A_c — на вход пришел сигнал (с шумом), A — на вход пришел только шум, B_c — устройство зарегистрировало наличие сигнала, B — устройство зарегистрировало шум. Необходимо вычислить $P(A_c/B_c)$. Воспользовавшись формулой Байеса (события A_c и A образуют полную группу событий), получим

$$P(A_c/B_c) = \frac{P(B_c/A_c)P(A_c)}{P(B_c/A_c)P(A_c) + P(B_c/A)P(A)}.$$

По условию задачи $P(A_c) = p$, $P(A) = 1 - p$, $P(B_c/A_c) = p_1$, $P(B_c/A) = p_2$. Следовательно, искомая вероятность

$$P(A_c/B_c) = \frac{p_1 p}{p_1 p + p_2 (1 - p)}.$$

4.2 Независимые события

Интуитивно, случайные события независимы, если информация об одном событии (произошло, не произошло) не меняет прогноз появления другого. (Например, при подбрасывании монеты и игральной кости информация о выпадении шестерки на кости не меняет прогноз появления герба на монете.) Поэтому формально независимые события A и B можно определить как такие, для которых имеет место равенство

$$P(A/B) = P(A)$$

или, что то же,

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

Последнее равенство можно переписать и так

$$P(A \cap B) = P(A)P(B).$$

О п р е д е л е н и е. Пусть $\{\Omega, P\}$ — вероятностное пространство. События A и B ($A \subset \Omega, B \subset \Omega$) будем называть *независимыми*, если

$$P(A \cap B) = P(A)P(B).$$

События $A_1, A_2, \dots, A_n$ будем называть *независимыми в совокупности*, если для любых $k = 2, 3, \dots, n$ и $1 \leq i_1 < i_2 < \dots < i_k \leq n$

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

События $A_1, A_2, \dots, A_n$ будем называть *попарно независимыми*, если для любых двух различных индексов s и k

$$P(A_s \cap A_k) = P(A_s)P(A_k).$$

Если события A и B независимы, то события A и $\overline{B}$; $\overline{A}$ и $\overline{B}$ также независимы.

Пример 4.2.1 (С. Н. Бернштейн). *Подбрасывают правильный тетраэдр, три грани которого окрашены соответственно в красный, синий и зеленый цвета, а в окраске четвертой грани присутствуют все три цвета. События: R — “красный”, B — “синий”, G — “зеленый” означают, что в окраске грани, соприкасающейся с поверхностью, присутствуют соответствующие цвета. Убедиться, что события R , B , G попарно независимы, но не являются независимыми в совокупности.*

Решение. Исход эксперимента — тетраэдр данной гранью соприкасается с поверхностью. Поскольку тетраэдр правильный, то в качестве математической модели стохастического эксперимента естественно рассматривать классическую модель.

Каждый цвет присутствует в окраске двух граней, поэтому

$$P(R) = 2/4 = 1/2, \quad P(B) = 1/2, \quad P(G) = 1/2.$$

В окраске только одной грани имеются три цвета, поэтому

$$P(R \cap B \cap G) = 1/4, \quad P(R \cap B) = 1/4,$$

$$P(R \cap G) = 1/4, \quad P(B \cap G) = 1/4.$$

Так что

$$P(R \cap B) = P(R)P(B), \quad P(R \cap G) = P(R)P(G),$$

$$P(B \cap G) = P(B)P(G).$$

Следовательно, события R, B, G — попарно независимы. Но

$$P(R \cap B \cap G) = 1/4 \neq 1/2 \cdot 1/2 \cdot 1/2 = P(R)P(B)P(G).$$

Последнее означает, что события R, B, G не являются независимыми в совокупности.

4.3 Задачи

АЗ: 4.11°, 4.12, 4.13, 4.22, 4.23, 4.24.

СЗ: 4.3°, 4.5, 4.7°, 4.9, 4.10°, 4.13, 4.16, 4.19, 4.24, 4.26.

4.1°. События A и B — независимы. Доказать, что события A и $\overline{B}$, $\overline{A}$ и $\overline{B}$ также независимы.

4.2°. События $A_1, A_2, \dots, A_n$ независимы в совокупности и $P(A_k) = p_k$, $k = 1, 2, \dots, n$. Какова вероятность того, что:

- а) не произойдет ни одно из событий $A_1, A_2, \dots, A_n$;
- б) произойдет по меньшей мере одно из событий $A_1, A_2, \dots, A_n$;
- в) произойдет одно и только одно из событий $A_1, A_2, \dots, A_n$?

4.3°. За один цикл осмотра радиолокационной станции, наблюдающей за космическим объектом, он будет обнаружен с вероятностью p . Объект в каждом цикле обнаруживается независимо от других. Проведено n циклов наблюдения. Какова вероятность обнаружения объекта?

4.4°. Прибор, состоящий из n блоков, выходит из строя, если выходит из строя хотя бы один блок. Блоки выходят из строя независимо один от одного. Надежность каждого блока равна p . Вычислить надежность прибора.

Под надежностью прибора или блока приборов будем понимать вероятность его безотказной работы на определенном фиксированном интервале времени.

4.5. Для повышения надежности прибора (см. задачу 4.4°) он дублируется таким же прибором; надежность каждого прибора равна p . В случае выхода из строя первого прибора происходит мгновенное переключение на дублирующий прибор. Приборы выходят из строя независимо один от другого. Найти надежность:

- 1) этой системы приборов;
- 2) системы приборов, если устройство переключения срабатывает с надежностью p_1 .

4.6. Для повышения надежности прибора (см. задачу 4.4°) он дублируется $n - 1$ такими же приборами; надежность каждого прибора равна p . Приборы выходят из строя независимо один от другого. Найти надежность:

- 1) этой системы приборов;

2) системы приборов, если каждый из приборов, включающий дублирующий прибор, имеет надежность p_1 .

Сколько необходимо приборов, чтобы надежность системы приборов была не меньше P ?

4.7°. Подбрасывают две игральные кости: красного и розового цветов. Какова вероятность того, что сумма очков на костях больше или равна 10, если известно, что:

а) на красной кости выпало 5 очков;

б) на одной из костей выпало 5 очков (возможность выпадения 5 очков на обоих костях не исключена)?

4.8°. В урне содержится n шаров. Все возможные предположения относительно количества белых шаров в урне равновероятны. Наудачу из урны извлекают один шар. Какова вероятность того, что шар белый?

4.9°. В N урнах содержится $n_1, n_2, \dots, n_N$ шаров, среди них белых соответственно $m_1, m_2, \dots, m_N$. Наудачу выбирают урну, а из нее — шар. Какова вероятность того, что этот шар окажется белым?

4.10°. В двух урнах содержится n_1 и n_2 шаров, из них белых шаров соответственно m_1 и m_2 . Из первой урны переложили во вторую один шар, цвет которого неизвестен. Затем из второй урны извлекают один шар. Какова вероятность того, что этот шар белый?

4.11°. Подбрасывают две игральные кости. Какова вероятность того, что сумма очков не менее девяти, если на одной из костей выпало четыре очка?

4.12. Радиолокационная станция ведет наблюдение за космическим объектом, который может создавать противолокационные помехи. Если объект не создает помехи, то за один цикл осмотра станция обнаруживает его с вероятностью p_0 , если создает, то с вероятностью p_1 ($p_1 < p_0$). Вероятность того, что во время цикла осмотра будут созданы помехи, составляет p и не зависит от того, как и когда создавались помехи в других циклах. Найти вероятность обнаружения объекта по меньшей мере один раз за n циклов осмотра.

4.13. Прибор состоит из n блоков, дублирующих друг друга, и может работать в благоприятных условиях или в неблагоприятных. В неблагоприятных условиях надежность работы каждого блока равна p_1 , а в благоприятных — p_2 . Вероятность того, что прибор будет работать

в благоприятных условиях, равна p , а в неблагоприятных — $(1 - p)$. Вычислить надежность прибора.

4.14°. В урну, которая содержит n шаров, положили белый шар. Какова вероятность того, что извлеченный из урны шар окажется белым, если все предположения относительно первоначального количества белых шаров в урне равновероятны?

4.15°. В урне находится n шаров, часть из них белые. Все предположения относительно количества белых шаров в урне равновероятны. Наудачу извлеченный из урны шар оказался белым. Вычислить вероятность всех предположений относительно количества белых шаров в урне. Какое предположение наиболее вероятно?

4.16°. В каждой из k_1 урн (первая группа) содержится m_1 белых и n_1 черных шаров, а в каждой из k_2 урн (вторая группа) содержится m_2 белых и n_2 черных шаров. Из наудачу выбранной урны извлекли шар, который оказался белым. Какова вероятность того, что его выбрали из первой группы урн?

4.17. Стрелок A попадает в цель с вероятностью $p_1 = 0,6$, стрелок B — с вероятностью $p_2 = 0,5$, стрелок C — с вероятностью $p_3 = 0,4$. Стрелки сделали залп по мишени. Известно, что имеется два попадания. Что вероятнее: попал стрелок C в мишень или нет?

4.18. В урне содержится 12 белых, 8 черных и 10 красных шаров. Наудачу извлекают два шара. Какова вероятность того, что эти шары разного цвета, если известно, что красный шар не выбран?

4.19. Из урны, содержащей 3 белых и 2 черных шара, переложили два шара в урну, содержащую 4 белых и 4 черных шара. Какова вероятность теперь извлечь белый шар из второй урны?

4.20. Детали изготавливают на двух заводах. Объем продукции второго завода в n раз превышает объем продукции первого завода. Доля брака на первом заводе p_1 , на втором — p_2 . Наудачу выбранное изделие оказалось бракованным. Какова вероятность того, что оно изготовлено на втором заводе?

4.21* (задача Банаха). Математик носит с собой две коробки спичек. Каждый раз, когда ему нужна спичка, он наудачу берет одну из коробок. Когда то наступит такой момент, что вынутая коробка окажется пустой.

Найти вероятность того, что другая коробка содержит r спичек, в предположении, что сначала каждая коробка содержит по N ($N \geq r$) спичек.

4.22. Трём радиостанциям разрешена работа на одной из трех заданных частот. Вероятность выбора радиостанцией каждой из частот равна $1/3$ и не зависит от того, какие частоты выбраны другими радиостанциями. Найти вероятности событий: “все радиостанции работают на одной частоте”, “все радиостанции работают на разных частотах”.

4.23. При рентгеновском обследовании вероятность обнаружить заболевание у больного туберкулезом составляет $1 - \beta$, а вероятность признать здорового человека больным равна α . Пусть доля больных туберкулезом относительно всего населения составляет γ . Найти вероятность того, что человек здоров, если он был признан больным при обследовании.

4.24. Вероятность того, что изделие предприятия удовлетворяет стандарту, равна 0,96. Предлагается упрощенная схема контроля, которая классифицирует стандартное изделие как стандартное с вероятностью 0,98, а нестандартное как стандартное с вероятностью 0,05. Какова вероятность того, что изделие, выдержавшее контроль, удовлетворяет стандарту?

4.25°. Подбрасывают две игральные кости. Какова вероятность того, что выпадет по меньшей мере одна пятерка, если известно, что сумма очков равна восьми?

4.26. Два стрелка стреляют в цель. Один из них попадает в цель в среднем в 5 случаях из 10, а второй — в 8 случаях из 10. Перед выстрелом они подбрасывают симметричную монету для определения очередности. Посторонний наблюдатель знает условия стрельбы, но не знает, кто в данный момент стреляет. Он видит, что стрелок попал в цель. Какова вероятность того, что стрелял первый стрелок?

4.27. Имеются три урны: в первой находится N_1 белых и M_1 черных шаров, во второй — N_2 белых и M_2 черных, в третьей — N_3 белых и M_3 черных шаров. Наудачу выбирают одну из урн и из нее без возвращения

два шара. Один из них оказался белым, другой — черным. Найти вероятность того, что шары были извлечены а) из первой, б) из второй, с) из третьей урны.

4.28 (задача о разделе ставки). Задача о разделе ставки имеет давние корни. Впервые ее опубликовал в Венеции в 1494 г. известный математик Фра Лука Пачоли. Имеются основания считать, что задача имеет арабское происхождение. На ее решение было потрачено немало усилий знаменитых математиков. Сам Пачоли не видел связи этой задачи с теорией вероятностей. Неправильное решение дал Никколо Тарталья (1499–1557), хотя он был достаточно гениальным, чтобы в математической дуэли за одну ночь найти формулу корней кубического уравнения. Правильное решение независимо друг от друга получили Паскаль и Ферма в 1654 г. Это открытие выглядело настолько важным, что многие считают 1654 г. годом рождения теории вероятностей.

Итак, задача о разделе ставки. Два игрока играют в справедливую игру (у обоих шансы выиграть одинаковы). (Например, подбрасывают симметричную монету, если при этом монета легла гербом, партию выиграл первый игрок, если решеткой — второй.) Игроки договорились, что тот, кто первым выигрывает 6 партий, получает весь приз. На самом деле игра остановилась до того, как один из игроков выиграл приз (скажем, первый выиграл 5 партий, а второй — 3). Как справедливо разделить приз?

Хотя на самом деле эта задача не является парадоксом, безуспешные попытки выдающихся ученых решить ее (а также неправильные решения) создали легенду о парадоксе.

4.29. Два игрока играют в справедливую игру. Тот, кто первым выигрывает 7 партий, получает весь приз. Игра прекратилась, когда первый выиграл 6 партий, а второй — 3. Как справедливо разделить приз?

4.30 (обобщение задачи Льюиса Кэрролла). В урне находится один шар, о котором известно, что он либо белый (с вероятностью $1/2$), либо черный. В урну положили n белых шаров, а затем после тщательного перемешивания последовательно извлекли n шаров, которые оказались белыми.

Какова вероятность того, что последний шар, извлеченный из урны, окажется белым?

4.31. Из урны, содержащей 5 белых и 2 черных шара, потеряли один шар. Для того, чтобы определить состав урны, из нее извлекли 2 шара, которые оказались белыми. Вычислить вероятность того, что был потерян белый шар.

4.32. Работает m радиолокационных станций, каждая из которых за один цикл осмотра обнаруживает объект с вероятностью p (независимо от других циклов и других станций). За время T каждая станция успевает сделать n циклов. Найти вероятности следующих событий: A — “за время T объект будет обнаружен по меньшей мере одной станцией”, B — “за время T объект будет обнаружен каждой станцией”.

4.33. Вы участвуете в игре, описанной в задаче 3.40, но после выбора вами двери ведущий из оставшихся дверей одну выбирает наудачу и открывает. Если за ней оказался автомобиль, то игра прекращается (при этом вы остаетесь без приза), если автомобиля нет — игра продолжается по правилам, описанным в задаче 3.40. Найти вероятность выигрыша автомобиля.

Мы рассмотрим более общую задачу: автомобиль находится за одной из n дверей. После выбора вами двери ведущий из $n - 1$ оставшихся дверей наудачу выбирает $n - 2$ и открывает их. Если за ними автомобиля нет, вам предоставляется право открыть одну из двух оставшихся дверей.

Какой выбор сделать? Какова вероятность, что вы выиграете автомобиль?

Глава 5

Дискретная случайная величина и ее распределение

5.1 Вычисление распределения функции от случайной величины

Случайная величина на дискретном вероятностном пространстве. Интуитивно, случайная величина — это величина, которая принимает те или иные значения в зависимости от исхода стохастического эксперимента (число космических частиц, достигших счетчика, число мутаций в клетках при радиоактивном облучении, ...). Исходы стохастического эксперимента описываются точками ω множества исходов Ω вероятностного пространства $\{\Omega, P\}$, поэтому формально случайную величину мы определяем как функцию $\xi = \xi(\omega)$ на множестве Ω .

Определение. *Случайной величиной* на дискретном вероятностном пространстве $\{\Omega, P\}$ будем называть функцию $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ со значениями в $\mathbb{R}^n$, заданную на пространстве элементарных событий Ω .

Если $n > 1$, то случайную величину $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ называют *многомерной или случайной величиной со значениями в $\mathbb{R}^n$ (случайным вектором)*, в частности, если $n = 2$ — *двумерной*, если $n = 1$

— *одномерной* (скалярной) или просто случайной величиной.

Случайная величина на дискретном вероятностном пространстве принимает не более чем счетное число значений. Такая случайная величина называется *дискретной*.

Распределение дискретной случайной величины. Дискретная случайная величина описывается своими значениями и вероятностями, с которыми она эти значения принимает (коротко, своим распределением).

Точку $x \in \mathbb{R}^n$ будем называть возможным значением случайной величины $\xi = \xi(\omega)$ — со значениями в $\mathbb{R}^n$, заданной на дискретном пространстве $\{\Omega, P\}$, если $P\{\xi = x\} > 0$; множество возможных значений ξ будем обозначать через $X (X \subset \mathbb{R}^n)$.

О п р е д е л е н и е. *Распределением* дискретной случайной величины $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^n$ будем называть функцию

$$P_\xi : x \rightarrow P_\xi(x), x \in X,$$

определенную на множестве X различных возможных значений случайной величины ξ , и ставящую в соответствие каждому возможному значению $x \in X$ вероятность $P_\xi(x) = P\{\xi = x\}$, с которой ξ принимает это значение.

Распределение

$$P_\xi(x_i, y_j, \dots, z_k) = P\{\xi_1 = x_i, \xi_2 = y_j, \dots, \xi_n = z_k\}$$

случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ со значениями в $\mathbb{R}^n$ ($n > 1$) еще называют *совместным* распределением случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

Распределение

$$P_\zeta : (x_i, y_j) \rightarrow P_\zeta(x_i, y_j), (x_i, y_j) \in X \subset \mathbb{R}^2,$$

случайной величины $\zeta = (\xi, \eta)$ со значениями в $\mathbb{R}^2$ удобно записывать в виде табл. 5.1.1.

Таблица 5.1.1. Распределение $\zeta = (\xi, \eta)$

Значение ξ	Значение η				
	y_1	y_2	$\dots$	y_m	$\dots$
x_1	$P_\zeta(x_1, y_1)$	$P_\zeta(x_1, y_2)$	$\dots$	$P_\zeta(x_1, y_m)$	$\dots$
x_2	$P_\zeta(x_2, y_1)$	$P_\zeta(x_2, y_2)$	$\dots$	$P_\zeta(x_2, y_m)$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	$P_\zeta(x_n, y_1)$	$P_\zeta(x_n, y_2)$	$\dots$	$P_\zeta(x_n, y_m)$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Распределение дискретной случайной величины $\xi = \xi(\omega)$ со значениями в $\mathbb{R}^1$ часто записывают в виде таблицы, в верхней строке которой указывают различные возможные значения случайной величины, а в нижнем — вероятности, с которыми эти значения принимаются:

x_i	x_1	x_2	$\dots$	x_n	$\dots$
$P_\xi(x_i)$	$P_\xi(x_1)$	$P_\xi(x_2)$	$\dots$	$P_\xi(x_n)$	$\dots$

Пример 5.1.1. Пусть ξ — число гербов, выпавших в результате подбрасывания двух симметричных монет. Найти распределение ξ .

Решение. Случайная величина ξ — функция $\xi = \xi(\omega)$ на дискретном вероятностном пространстве $\{\Omega, P\}$, где $\Omega = \{\Gamma\Gamma, \Gamma P, P\Gamma, PP\}$, а вероятность каждого элементарного события равна $1/4$ (монеты симметричны); ξ принимает значения $0, 1, 2$, причем $P\{\xi = 0\} = P\{\omega : \xi(\omega) = 0\} = P(PP) = 1/4$, $P\{\xi = 1\} = P(\Gamma P, P\Gamma) = P(\Gamma P) + P(P\Gamma) = 1/4 + 1/4 = 1/2$, $P\{\xi = 2\} = P(\Gamma\Gamma) = 1/4$. Следовательно, распределением ξ является

x_i	0	1	2
$P_\xi(x_i)$	1/4	1/2	1/4

Вычисление распределения функции от случайной величины. По распределению P_ζ случайной величины ζ всегда можно найти распределение любой функции $g(\zeta)$ от ζ .

Теорема. Пусть ζ — случайная величина со значениями в $\mathbb{R}^n$ на дискретном вероятностном пространстве $\{\Omega, P\}$, $P_\zeta : x \rightarrow P_\zeta(x)$ — ее распределение и $g = g(x)$ — функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^l$.

Для произвольного множества B из $\mathbb{R}^l$

$$P\{g(\zeta) \in B\} = \sum_{x: g(x) \in B} P_\zeta(x). \quad (5.1.1)$$

В частности, если $\zeta = (\xi, \eta)$, $g = g(x, y)$ и $P_\zeta(x_i, y_j)$ — распределение $\zeta = (\xi, \eta)$, то

$$P\{g(\zeta) \in B\} = P\{g(\xi, \eta) \in B\} = \sum_{(x_i, y_j): g(x_i, y_j) \in B} P_\zeta(x_i, y_j). \quad (5.1.2)$$

Пример 5.1.2. Дважды подбрасывают пару симметричных монет, ξ — число гербов при подбрасывании пары монет первый раз, η — второй.

Найти распределение случайной величины $\min\{\xi, \eta\}$.

Решение. Сначала найдем распределение $\zeta = (\xi, \eta)$. Случайная величина $\zeta = \zeta(\omega) = (\xi(\omega), \eta(\omega))$ — функция на дискретном вероятностном пространстве $\{\Omega, P\}$. Пространство элементарных событий Ω образуют последовательности “длиной четыре” из букв Г и Р; например, элементарное событие $\omega = (\text{ГРГГ})$ означает, что при первом подбрасывании пары монет на первой монете выпал герб, на второй — решетка, а при втором — на обеих монетах выпали гербы. Элементарные события равновероятны (монеты симметричны), вероятность каждого из них равна $1/16$.

Вычислив $P_\zeta(i, j)$ для каждой пары (i, j) , получим распределение $\zeta = (\xi, \eta)$:

Значения ξ	Значения η		
	0	1	2
0	1/16	1/8	1/16
1	1/8	1/4	1/8
2	1/16	1/8	1/16

Например,

$$\begin{aligned}
 P_{\zeta}(1, 1) &= P\{\zeta = (1, 1)\} = P\{(\xi, \eta) = (1, 1)\} = \\
 &= P\{\omega : (\xi(\omega), \eta(\omega)) = (1, 1)\} = \\
 &= P\{\text{ГРГР, ГРРГ, РГГР, РГРГ}\} = \\
 &= P(\text{ГРГР}) + P(\text{ГРРГ}) + P(\text{РГГР}) + P(\text{РГРГ}) = \\
 &= 1/16 + 1/16 + 1/16 + 1/16 = 1/4.
 \end{aligned}$$

По распределению $P_{\zeta}(i, j)$, $i, j = 0, 1, 2$, случайной величины $\zeta = (\xi, \eta)$ всегда можно найти распределение любой функции от нее (см. (5.1.2)), в частности, распределение $g(\zeta) = g(\xi, \eta) = \min\{\xi, \eta\}$. Для $B = \{k\}, k = 0, 1, 2$, имеем:

$$P\{\min\{\xi, \eta\} = k\} = \sum_{(i,j): \min(i,j)=k} P_{\zeta}(i, j).$$

Например, если $k = 0$ имеем

$$\begin{aligned}
 P\{\min\{\xi, \eta\} = 0\} &= \sum_{(i,j): \min(i,j)=0} P_{\zeta}(i, j) = \\
 &= P_{\zeta}(0, 0) + P_{\zeta}(0, 1) + P_{\zeta}(0, 2) + P_{\zeta}(1, 0) + P_{\zeta}(2, 0) = \\
 &= 1/16 + 1/8 + 1/16 + 1/8 + 1/16 = 7/16.
 \end{aligned}$$

Аналогично

$$P\{\min\{\xi, \eta\} = 1\} = 1/2, P\{\min\{\xi, \eta\} = 2\} = 1/16.$$

Следовательно, распределением случайной величины $\min\{\xi, \eta\}$ является

x_i	0	1	2
$P_{g(\zeta)}(x_i)$	7/16	1/2	1/16

Независимые случайные величины. Случайные величины ξ и η будем называть *независимыми*, если для всех возможных значений x_i, y_j случайных величин ξ и η

$$P_{\xi\eta}(x_i, y_j) = P_{\xi}(x_i)P_{\eta}(y_j). \quad (5.1.3)$$

(Распределение $P_{\xi\eta}$, для которого имеет место равенство (5.1.3), называется *произведением распределений* P_{ξ} и P_{η} .) Другими словами, случайные величины ξ и η независимы, если их совместное распределение равно произведению распределений ξ и η .

Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ будем называть *независимыми*, если их совместное распределение равно произведению распределений $\xi_1, \xi_2, \dots, \xi_n$.

Функции от независимых случайных величин являются независимыми случайными величинами.

Примечание. Интуитивно, случайные величины независимы, если информация о значении, принятом одной случайной величиной, не меняет прогноз появления значений, принимаемых другой.

Пример 5.1.3. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет распределение

$$P\{\xi_l = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0, \quad k = 0, 1, \dots \quad (l = 1, 2, \dots, n).$$

Выписать совместное распределение случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

Решение. По распределениям $P_{\xi_1}, P_{\xi_2}, \dots, P_{\xi_n}$ независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$ их совместное распределение

$$P_{\xi}(k_1, k_2, \dots, k_n) = P\{\xi_1 = k_1, \xi_2 = k_2, \dots, \xi_n = k_n\}$$

получаем как произведение распределений этих случайных величин:

$$P_{\xi}(k_1, k_2, \dots, k_n) = \prod_{i=1}^n P_{\xi_i}(k_i) =$$

$$= \prod_{i=1}^n \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n k_i}}{k_1! k_2! \dots k_n!} e^{-n\lambda};$$

$$k_1 = 0, 1, \dots; k_2 = 0, 1, \dots; \dots; k_n = 0, 1, \dots$$

5.2 Дискретные распределения на $\mathbb{R}^1$

Испытания Бернулли. Биномиальное распределение, пуассоновское распределение. Последовательность независимых испытаний (экспериментов) с двумя исходами (“успех”, “неудача”) и вероятностью успеха, не меняющейся от испытания к испытанию, называется *испытаниями Бернулли*.

Пространство Ω исходов в n испытаниях Бернулли образуют последовательности $\omega = (1, 0, 1, \dots, 1)$ длиной n , составленные из 1 и 0 (1 интерпретируем как успех, 0 — как неудачу). Вероятность исхода ω определяется равенством

$$P(\omega) = p^{\mu(\omega)}(1-p)^{n-\mu(\omega)} \quad (0 < p < 1),$$

где $\mu(\omega)$ — число единиц в последовательности ω , p — вероятность успеха в одном испытании. Пара $\{\Omega, P\}$ является вероятностным пространством n испытаний Бернулли.

Определение. Случайная величина ξ имеет *биномиальное распределение* с параметрами $(n; p)$, если

$$P_{\xi}(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (5.2.1)$$

Далее $C_n^k p^k (1-p)^{n-k}$ будем обозначать через $B_{n;p}(k)$ (для $k \geq n+1$ будем считать, что $B_{n;p}(k) = 0$).

Число ξ успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании имеет биномиальное распределение с параметрами $(n; p)$ (см. (5.2.1)).

Определение. Случайная величина ξ имеет пуассоновское распределение (распределение Пуассона) с параметром λ ($\lambda > 0$), если

$$P_{\xi}(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots \quad (5.2.2)$$

Во многочисленных задачах возникает необходимость вычислять $B_{n;p}(k)$, когда n и k большие (сотни и тысячи). При этом, как правило, пользуются теоремами Пуассона (когда n большое, а p малое) и Муавра—Лапласа.

Теорема Пуассона. Если $np \rightarrow \lambda$ ($\lambda > 0$) при $n \rightarrow \infty$, то для каждого фиксированного k , $k = 0, 1, 2, \dots$

$$\lim_{n \rightarrow \infty} B_{n;p}(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (5.2.3)$$

Из теоремы Пуассона следует, что для больших n и малых p пуассоновское распределение является хорошей аппроксимацией биномиального распределения.

В условиях теоремы Пуассона имеет место неравенство

$$\sum_{k=0}^{\infty} \left| B_{n;p}(k) - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leq \frac{2\lambda \min\{2, \lambda\}}{n} \quad (5.2.4)$$

(Ю. В. Прохоров).

Теорема Муавра—Лапласа. Если $npq \rightarrow \infty$ при $n \rightarrow \infty$, то для m , удовлетворяющих условию

$$\left| \frac{m - np}{\sqrt{npq}} \right| \leq C$$

(C — произвольная, но фиксированная константа),

$$\lim_{n \rightarrow \infty} B_{n;p}(m) / \frac{1}{\sqrt{2\pi npq}} \exp \left\{ -\frac{1}{2} \left(\frac{m - np}{\sqrt{npq}} \right)^2 \right\} = 1. \quad (5.2.5)$$

Интегральная предельная теорема Муавра—Лапласа. Пусть m — число успехов в n испытаниях Бернулли с вероятностью успеха p в одном испытании. Для произвольных $x_1 < x_2$

$$\lim_{n \rightarrow \infty} P \left\{ x_1 < \frac{m - np}{\sqrt{npq}} < x_2 \right\} = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} \exp \left\{ -\frac{t^2}{2} \right\} dt. \quad (5.2.6)$$

Геометрическое распределение. Случайная величина ξ имеет *геометрическое распределение* с параметром p ($p > 0$), если

$$P_\xi(k) = (1 - p)^k p, \quad k = 0, 1, \dots \quad (5.2.7)$$

Число испытаний до первого успеха в последовательности независимых испытаний с вероятностью успеха p в каждом имеет геометрическое распределение с параметром p ($p > 0$).

Отрицательное биномиальное распределение. Случайная величина ξ имеет *отрицательное биномиальное распределение* с параметрами $(r; p)$, если

$$P_\xi(k) = C_{k+r-1}^{r-1} p^r (1 - p)^k, \quad k = 0, 1, \dots \quad (5.2.8)$$

Число неудач до r -го успеха в последовательности независимых испытаний с вероятностью успеха p в каждом имеет отрицательное биномиальное распределение с параметрами $(r; p)$.

Заметим, что если $r = 1$, отрицательное биномиальное распределение совпадает с геометрическим.

Гипергеометрическое распределение. Случайная величина ξ имеет *гипергеометрическое распределение* с параметрами (N, M, n) , если

$$P_\xi(m) = P_{N, M, n}(m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad (5.2.9)$$

$m = 0, 1, 2, \dots, \min\{n, M\}; n \leq N - M$.

Пусть в урне находится N шаров, из них M — белые, остальные $(N - M)$ — черные. Число белых шаров среди n

($n \leq N - M$) наудачу выбранных из урны, имеет гипергеометрическое распределение с параметрами (N, M, n) .

Если $N, M \rightarrow \infty$ так, что $M/N \rightarrow p$ ($0 < p < 1$), то

$$P_{N,M,n}(m) \rightarrow C_n^m p^m (1-p)^{n-m}, \quad m = 1, 2, \dots, n.$$

Пример 5.2.1. Известно, что вероятность p рождения мальчика равна 0,515. Какова вероятность того, что мальчиков среди $n = 10\,000$ новорожденных будет меньше, чем девочек?

Решение. Обозначим через ξ количество мальчиков среди $n = 10\,000$ новорожденных, тогда девочек будет $10\,000 - \xi$. Необходимо вычислить

$$P\{\xi < 10\,000 - \xi\} = P\{\xi < 5000\}.$$

Количество ξ мальчиков можно представить в виде $\sum_{i=1}^n \mu_i$,

где μ_i принимает значение 1, если i -й новорожденный является мальчиком и 0 — если девочкой. Случайные величины μ_i , $i = 1, 2, \dots, n$, — независимы, каждая с распределением

$$P\{\mu_i = 1\} = p, \quad P\{\mu_i = 0\} = 1 - p = q.$$

Далее воспользуемся интегральной теоремой Муавра—Лапласа. Поскольку в данном случае n большое ($n = 10\,000$), то можно считать, что

$$P\left\{x_1 < \frac{\xi - np}{\sqrt{npq}} < x_2\right\} = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

И следовательно,

$$\begin{aligned} & P\{\xi < 5000\} = \\ & = P\left\{\frac{\xi - 5150}{\sqrt{10000 \cdot 0,515 \cdot 0,485}} < \frac{-150}{\sqrt{10000 \cdot 0,515 \cdot 0,485}}\right\} = \end{aligned}$$

$$= P\left\{\frac{\xi - 5150}{49,98} < -3, 0\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-3,0} \exp\left\{-\frac{t^2}{2}\right\} dt = 0,001$$

(значение интеграла найдено по таблице нормального распределения, см. табл. 22.1.1).

Пример 5.2.2. Пусть ξ и η — независимые целочисленные случайные величины с распределениями

$$P\{\xi = k\} = p_k, \quad k = 0, 1, \dots; \quad P\{\eta = l\} = q_l, \quad l = 0, 1, \dots$$

Найти распределение случайной величины $\zeta = \xi + \eta$.

Решение. Поскольку случайные величины ξ и η независимы, то их совместным распределением является

$$P\{(\xi, \eta) = (k, l)\} = p_k q_l, \quad k = 0, 1, \dots; \quad l = 0, 1, \dots$$

Далее, согласно (5.1.2) имеем

$$\begin{aligned} P\{\zeta = m\} &= P\{\xi + \eta = m\} = \\ &= \sum_{(k,l): k+l=m} p_k q_l = \sum_{k=0}^m p_k q_{m-k}, \quad m = 0, 1, \dots \end{aligned}$$

Распределение

$$f_m = \sum_{k=0}^m p_k q_{m-k} = p_0 q_m + p_1 q_{m-1} + \dots + p_m q_0, \quad m = 0, 1, \dots$$

называют сверткой распределений $\{p_k\}$ и $\{q_l\}$.

Так что распределение суммы независимых целочисленных случайных величин равно свертке распределений слагаемых.

5.3 Задачи

АЗ: 5.3°, 5.8(1в), 5.8(2е), 5.8(3б), 5.13, 5.29, 5.30.

СЗ: 5.1°, 5.4°, 5.8(1б), 5.8(2з), 5.14, 5.18, 5.29, 5.31.

5.1°. Пусть ξ — число очков, выпавших при подбрасывании симметричной игральной кости. Найти распределение случайной величины $\eta = \sin \frac{\pi}{3} \xi$.

5.2°. Пусть ξ — число очков, выпавших при подбрасывании игральной кости, изготовленной так, что вероятность выпадения грани пропорциональна числу очков на ней. Найти распределение случайной величины $\eta = \text{sign} \left(\cos \frac{\pi}{3} \xi \right)$.

5.3°. Из урны, содержащей 3 белых и 2 черных шара, переложили два шара в урну, в которой содержится 1 белый и 2 черных шара. Затем из второй урны извлекают 2 шара. Пусть ξ — число белых шаров среди них. Найти распределение ξ .

5.4°. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет геометрическое распределение с параметром p :

$$P\{\xi_l = k\} = P_{\xi_l}(k) = (1-p)^k p, \quad k = 0, 1, 2, \dots;$$

$l = 1, 2, \dots, n$.

Найти совместное распределение случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

5.5°. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет биномиальное распределение с параметрами (m, p) :

$$P\{\xi_l = k\} = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m;$$

$l = 1, 2, \dots, n$.

Найти совместное распределение случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

5.6*. Пусть $\xi_j = \xi_j(\omega)$, $j = 1, 2, \dots, n$, — независимые случайные величины, для которых

$$P\{\xi_j > 0\} = p, P\{\xi_j < 0\} = q, P\{\xi_j = 0\} = f, \quad p+q+f = 1,$$

$j = 1, 2, \dots, n$; s — количество случайных величин среди ξ_j , $j = 1, 2, \dots, n$, отличных от нуля, а μ — число положительных среди ξ_j , $j = 1, 2, \dots, n$.

Найти совместное распределение случайных величин μ и s .

5.7. Стохастический эксперимент состоит в подбрасывании пары симметричных игральных костей. Построить вероятностное пространство $\{\Omega, P\}$ этого стохастического эксперимента. Пусть $\zeta = (\xi, \eta)$ — случайная величина на $\{\Omega, P\}$ со значениями в $\mathbb{R}^2$, где ξ — число очков на первой игральной кости, η — на второй. Найти совместное распределение случайных величин ξ и η (распределение векторной случайной величины $\zeta = (\xi, \eta)$). Доказать, что случайные величины ξ и η — независимы.

5.8. Подбрасывают пару симметричных игральных костей и рассматривают случайную величину $\zeta = (\xi, \eta)$, где ξ — число очков на первой игральной кости, η — на второй.

1. Найти распределения:

а) $\max\{\xi, \eta\}$;

б) $\min\{\xi, \eta\}$;

в) $\xi + \eta$.

2. Вычислить:

а) $P\{\xi \leq 2, \max\{\xi, \eta\} \geq 4\}$;

б) $P\{\max\{\xi, \eta\} \geq 4\}$;

в) $P\{|\eta - \xi| \geq 3\}$;

г) $P\{4 \leq \xi + \eta \leq 6\}$;

д) $P\{\xi \leq 1, \max\{\xi, \eta\} \geq 3\}$;

е) $P\{\max\{\xi, \eta\} \leq 4\}$;

ё) $P\{\min\{\xi, \eta\} \leq 1, \max\{\xi, \eta\} \geq 5\}$;

ж) $P\{\max\{\xi, \eta\} \geq 3\}$;

з) $P\{\xi \geq 2, \max\{\xi, \eta\} \geq 3\}$.

3. Найти совместное распределение случайных величин:

а) ξ и $\xi + \eta$; являются ли случайные величины ξ и $\xi + \eta$ независимыми?

б) ξ и $\max\{\xi, \eta\}$; являются ли случайные величины ξ и $\max\{\xi, \eta\}$ независимыми?

в) ξ и $\min\{\xi, \eta\}$; являются ли случайные величины ξ и $\min\{\xi, \eta\}$ независимыми?

5.9. Подбрасывают пару симметричных монет и симметричную игральную кость.

Предложить математическую модель (вероятностное пространство $\{\Omega, P\}$) этого стохастического эксперимента.

Пусть $\zeta = (\xi, \eta)$ — случайная величина на $\{\Omega, P\}$ со значениями в $\mathbb{R}^2$, где ξ — число гербов, выпавших на паре монет, η — число очков на грани игральной кости.

Найти распределения случайных величин, перечисленных в задаче 5.8, вычислить вероятности, ответить на вопросы.

5.10. Совместное распределение случайных величин ξ и η задано таблицей 5.3.2.

Найти распределения случайных величин, перечисленных в задаче 5.8, вычислить вероятности, ответить на вопросы.

Таблица 5.3.2. Совместное распределение ξ и η

Значения ξ	Значения η							
	1	2	3	4	5	6	7	8
1	1/16	1/32	0	1/32	1/32	1/32	1/32	1/32
2	1/32	1/16	1/32	0	1/32	1/32	1/32	1/32
3	1/32	1/32	1/16	1/32	0	1/32	1/32	1/32
4	1/32	1/32	1/32	1/16	1/32	0	1/32	1/32

5.11. Пусть ξ и η — независимые случайные величины с распределениями

$$P\{\xi = x_k\} = a_k, \quad P\{\eta = x_k\} = b_k, \quad k = 1, 2, \dots, n.$$

Вычислить $P\{\xi = \eta\}$.

5.12. Пусть ξ и η — независимые случайные величины с распределениями

$$P\{\xi = i\} = 1/(n+1), \quad P\{\eta = i\} = 1/(n+1), \quad i = 0, 1, \dots, n.$$

Найти распределение случайной величины $\zeta = \xi + \eta$.

5.13. Пусть ξ и η — независимые случайные величины, имеющие распределение Пуассона соответственно с параметрами λ и μ .

Найти распределение случайной величины $\zeta = \xi + \eta$.

5.14. Пусть ξ и η — независимые случайные величины, имеющие распределение Пуассона соответственно с параметрами λ и μ .

Доказать, что условным распределением случайной величины ξ при условии $\xi + \eta = n$ является биномиальное распределение с параметрами $(n, \lambda/(\lambda + \mu))$, т. е.

$$P\{\xi = k | \xi + \eta = n\} = C_n^k \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(1 - \frac{\lambda}{\lambda + \mu} \right)^{n-k},$$

$k = 0, 1, \dots, n$.

5.15. Пусть ξ и η — независимые геометрически распределенные с параметром p случайные величины:

$$P\{\xi = k\} = p(1-p)^k, \quad P\{\eta = k\} = p(1-p)^k, \quad k = 0, 1, \dots$$

Найти совместное распределение случайной величины ξ и $\max\{\xi, \eta\}$; распределение случайной величины $\max\{\xi, \eta\}$.

5.16. Подбрасывают две игральные кости. Пусть ξ — случайная величина, равная нулю, если на первой кости выпадет число очков меньше чем 3, и единице — в противном случае; η — случайная величина, равная числу очков на второй игральной кости.

Найти распределение случайной величины $\zeta = \xi + \eta$.

5.17. Подбрасывают симметричные монеты: одна достоинством 25 коп. и две достоинством 50 коп.; ξ_1 — число гербов, выпавших на монете достоинством 25 коп., а ξ_2 — на двух монетах достоинством 50 коп. Найти распределение случайной величины $\eta = (\xi_1, \xi_2)$.

5.18*. Доказать, что если каждая из независимых случайных величин ξ_1 и ξ_2 имеет геометрическое распределение, то и случайная величина $\eta = \min\{\xi_1, \xi_2\}$ имеет геометрическое распределение. Найти параметр этого распределения, если параметры распределений случайных величин ξ_1 и ξ_2 равны соответственно p_1 и p_2 .

5.19°. Симметричную игральную кость подбрасывают 5 раз. Найти вероятность того, что: 1) дважды появится число очков, кратное трем; 2) больше двух раз появится число очков, кратное трем.

5.20°. Пару симметричных игральных костей подбрасывают 6 раз. Вычислить вероятность того, что на обоих костях трижды выпадет одинаковое число очков.

5.21°. Симметричную игральную кость подбрасывают 6 раз. Вычислить вероятность того, что трижды выпадет четное число очков.

5.22. При техническом контроле изделий, каждое из них, независимо от остальных, может с вероятностью p оказаться дефектным.

1. Какова вероятность того, что из 10 проверенных изделий только одно окажется дефектным?

2. Найти вероятность того, что первым дефектным изделием окажется k -е проверенное изделий.

3. Найти вероятность того, что последующие 10 изделий окажутся стандартными, при условии, что предыдущие l изделий были стандартными. Зависит ли эта вероятность от l ?

5.23°. Трижды подбрасывают монету, вероятность появления герба которой составляет $1/3$; ξ — число выпавших гербов. Найти распределение случайной величины ξ .

5.24. Найти вероятность выпадения хотя бы:

а) одной шестерки при подбрасывании 6 игральных костей;

б) двух шестерок при подбрасывании 12 игральных костей;

в) трех шестерок при подбрасывании 18 игральных костей.

5.25°. Симметричную монету подбрасывают 10 раз. Найти распределение случайной величины ξ — числа выпавших гербов.

5.26°. Симметричную игральную кость подбрасывают 10 раз. Найти распределение случайной величины ξ — числа выпавших троек.

5.27°. Пару симметричных монет подбрасывают трижды, ξ — число тех подбрасываний, при которых на обеих монетах выпал герб. Найти распределение случайной величины ξ .

5.28. Что вероятнее, выиграть у игрока, равного по силе (игра ведется без ничьих):

- 1) 4 партии из 8 или 3 из 5?
- 2) 3 партии из 6 или 2 из 4?
- 3) 3 партии из 4 или 5 из 8?
- 4) не менее чем 3 партии из 4 или не менее чем 5 из 8?
- 5) не более чем n партий из $2n$ или более чем n партий из $2n$?
- 6) не более чем n партий из $2n + 1$ или более чем n партий из $2n$?

5.29. Найти распределение суммы $\eta = \xi_1 + \xi_2 + \dots + \xi_r$ независимых случайных величин, каждая из которых имеет геометрическое распределение с параметром p .

5.30. Вероятность того, что в справочное бюро в течение часа обратится k человек, равна

$$\frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

($\lambda > 0$). Для каждого из них вероятность отказа равна p . Найти вероятность того, что в течение часа s человек не получат ответ на свой вопрос.

5.31. Предположим, что некоторое насекомое с вероятностью $\frac{\lambda^k}{k!} e^{-\lambda}$ откладывает k ($k = 0, 1, \dots$) яиц. Вероятность появления насекомого из яйца равна p . Предполагая взаимную независимость появления насекомых из яиц, найти вероятность того, что потомство насекомого будет составлять i особей.

5.32. В экспериментах на встречных электронно-позитронных пучках вероятность того, что за единицу времени произойдет j столкновений, сопровождающихся рождением новых элементарных частиц, составляет

$$p_j = \frac{\lambda^j}{j!} e^{-\lambda}, \quad j = 0, 1, \dots,$$

λ — положительный параметр. При каждом столкновении в результате взаимодействия могут образоваться различные группы элементарных частиц, вероятность появления каждой группы постоянная и не зависит от результатов взаимодействий при других столкновениях. В качестве одной из таких групп рассмотрим пару μ -мезонов

и обозначим через p вероятность их появления при одном столкновении. Найти распределение случайной величины — количества пар рожденных μ -мезонов.

5.33. Автоматическую телефонную станцию A на 500 абонентов необходимо соединить с телефонной станцией B . Известно, что количество требований связи абонентов телефонной станции имеет распределение Пуассона.

Если параметр пуассоновского распределения: а) $\lambda = 5$; б) $\lambda = 8$; в) $\lambda = 10$, то каким должно быть минимальное число линий связи станции A со станцией B , чтобы вероятность отказа не превышала 0,01 (0,02; 0,05)?

5.34. Театр, вмещающий 1000 человек, имеет два разных входа. Возле каждого из них имеется гардероб. Количество мест в гардеробах одинаковое.

Какое минимальное количество мест должно быть в гардеробах, чтобы в среднем в 9 случаях из 10 все зрители могли раздеться в гардеробе того входа, через который они вошли? Рассмотрите два случая: а) зрители приходят парами; б) зрители приходят по одному. Предположите, что входы зрители выбирают с одинаковыми вероятностями.

5.35. Пусть ξ — дискретная случайная величина с распределением

x_i	x_1	x_2	$\dots$	x_n
$P_\xi(x_i)$	p_1	p_2	$\dots$	p_n

Для каждого $x \in (-\infty, \infty)$ вычислить $P\{\xi < x\}$. Построить график функции $F(x) = P\{\xi < x\}$, $x \in (-\infty, \infty)$. Какие значения принимает $F(x)$? Исследовать функцию $F(x)$ на монотонность и непрерывность.

5.36*. Доказать, что случайная величина ξ имеет геометрическое распределение тогда и только тогда, когда для каждого n , $n = 1, 2, \dots$, выполняется равенство

$$P\{\xi = n + m / \xi \geq n\} = P\{\xi = m\}, \quad m = 0, 1, \dots$$

(свойство отсутствия последействия геометрического распределения).

5.37. Два игрока подбрасывают по очереди монету. Выигрывает тот, у кого впервые выпадет герб. Найти вероятность p_k того, что игра закончится на k -м подбрасывании. Во сколько раз больше вероятность выигрыша для игрока, который начинает первым?

5.38. Двое одинаково метких стрелков по очереди стреляют в мишень. Каждый должен сделать не более двух выстрелов. Тот, кто первым попадет в мишень, получает приз.

1. Если вероятность попадания $p = 1/5$, то что вероятнее: получают стрелки приз или нет?

2. Чему равно отношение вероятности получить приз первым стрелком к вероятности получить приз вторым, если вероятность попадания составляет $p = 1/3$ и количество выстрелов не ограничено?

5.39. Для экспериментальной проверки свойства устойчивости частоты в разное время были проведены следующие опыты.

1° Монету подбросили 4040 раз, при этом герб выпал 2043 раз (Бюффон).

2° Монету подбросили 12 000 раз, при этом частота выпадения герба составила 0,5016; в другом эксперименте при подбрасывании монеты 24 000 раз частота выпадения герба составила 0,5005 (К. Пирсон).

Для каждого из экспериментов:

а) найти вероятность того, что в случае повторения эксперимента отклонение частоты от $1/2$ по абсолютной величине не превышает отклонение частоты от $1/2$ в экспериментах, проведенных Бюффоном и Пирсоном;

б) считая, что событие, вероятность появления которого составляет 0,9999, практически достоверное, найти такое минимальное ε , что событие “частота по абсолютной величине отклоняется от $1/2$ меньше чем на ε ” является практически достоверным событием.

5.40. Грани симметричной игральной кости занумерованы двумя единицами, двумя двойками и двумя тройками.

В квадратном уравнении

$$\xi x^2 + \eta x + \zeta = 0$$

ξ, η, ζ определяются как результаты трех последовательных независимых подбрасываний кости. Найти вероят-

ность того, что уравнение: 1) имеет действительные корни; 2) имеет рациональные корни; 3) не имеет действительных корней.

5.41. По распределению

$$P_{\zeta} : (x_k, y_l) \rightarrow P_{\zeta}(x_k, y_l)$$

случайной величины $\zeta = (\xi, \eta)$ со значениями в $\mathbb{R}^2$ найти распределение ξ .

5.42. Привести пример дискретного вероятностного пространства $\{\Omega, P\}$ и случайных величин $\xi = \xi(\omega)$, $\eta = \eta(\omega)$ на нем, таких, что распределения P_{ξ} и P_{η} случайных величин ξ и η совпадают, но $\xi \neq \eta$ в каждой точке $\omega \in \Omega$.

5.43. Пусть ξ — пуассоновская случайная величина с параметром λ . Для каждого x вычислить $P\{\xi < x\}$. Построить график функции

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

5.44. Симметричную монету подбрасывают 10 раз. Как известно, вероятность того, что герб выпадет 4 раза, равна

$$B_{10;1/2}(4) = C_{10}^4(1/2)^{10}.$$

А чему равна вероятность того, что при последовательном подбрасывании монеты для появления 4-х гербов необходимо сделать 10 подбрасываний?

Глава 6

Математическое ожидаие дискретной случайной величины

6.1 Определения, свойства, вычисления

Случайная величина описывается своим распределением

$$P_{\xi} : x_i \rightarrow P_{\xi}(x_i) = P\{\xi = x_i\},$$

задающим значения, принимаемые случайной величиной, и вероятности, с которыми эти значения принимаются. Информацию о распределении случайной величины мы получаем из эксперимента (независимых наблюдений случайной величины). Но зачастую получить распределение случайной величины непосредственно из эксперимента не удастся. Поэтому возникает необходимость описывать распределение случайной величины характеристиками, которые, возможно, и не всегда полностью описывают распределение, но дают достаточно хорошее представление о нем и которые можно оценить по экспериментальным данным. Такими характеристиками, в частности, является математическое ожидание и дисперсия случайной величины. (Некоторые распределения математическим ожиданием и дисперсией описываются полностью.)

Определение. Математическим ожиданием $M\xi$ случайной величины $\xi = \xi(\omega)$, заданной на дискретном вероятностном пространстве $\{\Omega, P\}$, будем называть сумму ряда

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega)P(\omega),$$

если он абсолютно сходится.

Свойства математического ожидания.

1. Математическое ожидание константы равно этой же константе:

$$Mc = c \text{ (} c \text{ — константа)}.$$

2. Математическое ожидание суммы случайных величин равно сумме математических ожиданий этих случайных величин:

$$M(\xi + \eta) = M\xi + M\eta.$$

3. Константа выносится из под знака математического ожидания:

$$Ma\xi = aM\xi.$$

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

$$M\xi\eta = M\xi M\eta.$$

Вычисление математического ожидания случайной величины по ее распределению. По распределению случайной величины всегда можно вычислить математическое ожидание функции от нее (если только оно существует).

Теорема 6.1.1 Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$; $P_\xi : x_i \rightarrow P_\xi(x_i)$ — ее распределение; g — функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$.

Если ряд $\sum_{x_i} g(x_i)P_\xi(x_i)$ сходится абсолютно, то

$$Mg(\xi) = \sum_{x_i} g(x_i)P_\xi(x_i), \quad (6.1.1)$$

в частности, если ряд $\sum_{x_i} x_i P_\xi(x_i)$ сходится абсолютно, то

$$M\xi = \sum_{x_i} x_i P_\xi(x_i). \quad (6.1.2)$$

Теорема о вычислении математического ожидания функции от случайной величины по ее распределению имеет место и для случайных величин со значениями в $\mathbb{R}^n$.

Если ряд

$$\sum_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n) P_\xi(x_1, x_2, \dots, x_n)$$

сходится абсолютно, то

$$\begin{aligned} & Mg(\xi_1, \xi_2, \dots, \xi_n) = \\ &= \sum_{x_1, x_2, \dots, x_n} g(x_1, x_2, \dots, x_n) P_\xi(x_1, x_2, \dots, x_n), \end{aligned} \quad (6.1.3)$$

где $P_\xi : (x_1, x_2, \dots, x_n) \rightarrow P_\xi(x_1, x_2, \dots, x_n)$ — распределение случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_n)$.

Примечание. Часто математическое ожидание (среднее) случайной величины определяют равенством (6.1.2).

На равенство (6.1.2) можно смотреть как на определение среднего вероятностного распределения.

Пример 6.1.1. Пусть ξ — число выпавших очков при подбрасывании игральной кости. Кость изготовлена так, что распределением случайной величины ξ является

x_i	1	2	3	4	5	6
$P_\xi(x_i)$	1/10	1/10	1/10	1/10	1/10	1/2

Вычислить математическое ожидание ξ .

Решение. Согласно (6.1.2)

$$M\xi = 1 \cdot \frac{1}{10} + 2 \cdot \frac{1}{10} + \dots + 5 \cdot \frac{1}{10} + 6 \cdot \frac{1}{2} = 4,5.$$

Заметим, что если игральная кость симметрична, т. е. распределением ξ является

x_i	1	2	3	4	5	6
$P_\xi(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

то

$$M\xi = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5.$$

Пример 6.1.2. *Случайная величина ξ имеет распределение Пуассона с параметром λ . Вычислить*

$$M \frac{1}{\xi + 1}.$$

Решение. По известному распределению P_ξ случайной величины ξ всегда можно вычислить математическое ожидание $Mg(\xi)$ функции $g(\xi)$ от ξ , если только оно существует, (см. (6.1.1)). В нашей задаче $g(t) = 1/(1+t)$, а случайная величина ξ имеет пуассоновское распределение с параметром λ :

$$P_\xi(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Поэтому

$$\begin{aligned} M \frac{1}{1+\xi} &= \sum_{k=0}^{\infty} \frac{1}{1+k} P_\xi(k) = \sum_{k=0}^{\infty} \frac{1}{1+k} \frac{\lambda^k}{k!} e^{-\lambda} = \\ &= \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(1+k)!} = \frac{e^{-\lambda}}{\lambda} (e^\lambda - 1) = \frac{1 - e^{-\lambda}}{\lambda}. \end{aligned}$$

Дисперсия случайной величины. Дисперсией $D\xi$ случайной величины ξ будем называть $M(\xi - M\xi)^2$ (если $M(\xi - M\xi)^2 < \infty$), т. е.

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

Непосредственно из определения получаем

$$D\xi = M\xi^2 - (M\xi)^2. \quad (6.1.4)$$

Приведенные далее свойства дисперсии непосредственно следуют из ее определения.

1. Дисперсия константы равна нулю:

$$Dc = 0 \quad (c - \text{константа}).$$

2. Константа выносится из-под знака дисперсии с квадратом:

$$Da\xi = a^2 D\xi.$$

3. Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$D(\xi + \eta) = D\xi + D\eta.$$

Пример 6.1.3. *Вычислить математическое ожидание и дисперсию биномиально распределенной с параметрами $(n; p)$ случайной величины ξ .*

Решение. Распределением ξ является

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

Согласно (6.1.2)

$$\begin{aligned} M\xi &= \sum_{k=0}^n k P_\xi(k) = \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = \\ &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}. \end{aligned}$$

Обозначив $k-1 = s$, последнюю сумму перепишем так:

$$np \sum_{s=0}^{n-1} \frac{(n-1)!}{s!((n-1)-s)!} p^s (1-p)^{(n-1)-s} =$$

$$= np(p + (1 - p))^{n-1} = np.$$

Следовательно,

$$M\xi = np.$$

Аналогично

$$M\xi^2 = \sum_{k=0}^n k^2 C_n^k p^k (1-p)^{n-k} = np(np - p + 1).$$

Отсюда

$$D\xi = M\xi^2 - (M\xi)^2 = np(np - p + 1) - (np)^2 = np(1 - p).$$

Пример 6.1.4. Подбрасывают две симметричные игральные кости. Вычислить математическое ожидание и дисперсию суммы выпавших очков.

Решение. Пусть ξ — число очков, выпавших на первой игральной кости, η — на второй, тогда $\zeta = \xi + \eta$ — сумма очков, выпавших на костях. Кости симметричны, поэтому распределением каждой из случайных величин ξ и η является

x_i	1	2	3	4	5	6
$P(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

Отсюда

$$M\zeta = M(\xi + \eta) = M\xi + M\eta = 7/2 + 7/2 = 7,$$

$$\begin{aligned} D\xi &= M\xi^2 - (M\xi)^2 = \\ &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} - \left(\frac{7}{2}\right)^2 = 35/12. \end{aligned}$$

А поскольку ξ и η — независимые случайные величины, то

$$D\zeta = D(\xi + \eta) = D\xi + D\eta = 35/12 + 35/12 = 35/6.$$

Заметим, что непосредственное вычисление математического ожидания и дисперсии суммы очков было бы громоздким.

Определение. *Ковариацией* случайных величин ξ и η будем называть

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

Определение. *Коэффициентом корреляции* случайных величин ξ и η будем называть

$$r(\xi, \eta) = \frac{M(\xi - M\xi)(\eta - M\eta)}{\sqrt{D\xi}\sqrt{D\eta}}.$$

Если случайные величины ξ и η таковы, что $r(\xi, \eta) = 0$, то они называются *некоррелированными*.

Теорема 6.1.2 (неравенство Чебышёва). *Если $D\xi < \infty$, то для произвольного $\varepsilon > 0$*

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}$$

Из неравенства Чебышёва, в частности, следует, что если дисперсия $D\xi$ мала, большие отклонения (больше чем ε) случайной величины от своего среднего встречаются изредка.

Теорема 6.1.3 (закон больших чисел). *Пусть $\xi_1, \xi_2, \dots$ — независимые одинаково распределенные случайные величины со средним $M\xi_i = a$ и конечными дисперсиями $D\xi_i < \infty$, $i = 1, 2, \dots$. Тогда для произвольного $\varepsilon > 0$*

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right\} \rightarrow 0,$$

при $n \rightarrow \infty$.

6.2 Задачи

АЗ: $6.1^\circ(1), 6.2, 6.4^\circ, 6.5, 6.16, 6.18, 6.21, 6.27^*, 6.31^*, 6.32;$
 СЗ: $6.3^\circ, 6.6^\circ, 6.13^\circ, 6.15, 6.17, 6.20, 6.24^\circ, 6.30, 6.33.$

6.1°. Подбрасывают симметричную монету и правильную игральную кость, ξ — число выпавших при этом гербов, η — очков.

Вычислить:

$$1) M \frac{1}{\xi + 1} \cos \frac{\pi}{6} \eta; \quad 2) M e^{\xi} \sin \frac{\pi}{6} \eta.$$

6.2. В урне содержится 2 белых и 8 черных шаров. Из урны наудачу извлекают три шара. Найти распределение случайной величины ξ — числа выбранных белых шаров; вычислить $M\xi$ и $D\xi$.

6.3°. Пусть ξ и η — независимые случайные величины соответственно с распределениями

x_i	0	1	2	3
$P_{\xi}(x_i)$	1/2	1/6	1/6	1/6

y_j	0	1	2
$P_{\eta}(y_j)$	1/2	1/4	1/4

Вычислить

$$M \frac{\xi}{\eta^2 + 1}.$$

6.4°. Дважды подбрасывают симметричную монету, ξ — число выпавших при этом гербов. Вычислить

$$M(-1)^{\xi} \sin \frac{\pi}{3} \xi.$$

6.5. Случайная величина ξ принимает целые неотрицательные значения. Доказать, что

$$M\xi = \sum_{m=1}^{\infty} P\{\xi \geq m\}.$$

6.6°. Пусть ξ и η — независимые случайные величины соответственно с распределениями

x_i	1	2	3	4
$P_\xi(x_i)$	1/4	1/4	1/4	1/4

y_j	-1	0	1
$P_\eta(y_j)$	1/3	1/3	1/3

Вычислить

$$M\eta^2 \sin\left(\frac{\pi}{2}\xi\right).$$

6.7°. Случайная величина ξ имеет распределение

x_i	-2	-1	1	2
$P_\xi(x_i)$	1/4	1/4	1/4	1/4

Найти математические ожидания случайных величин η :

$$1) \eta = \sin \frac{\pi}{\xi}; \quad 2) \eta = \xi 2^{|\xi|}; \quad 3) \eta = \xi \sin \frac{\pi}{3}\xi.$$

6.8°. Случайная величина ξ имеет распределение

x_i	-2	-1	1
$P_\xi(x_i)$	1/2	1/4	1/4

Вычислить

$$M\xi^2 2^{-\xi}.$$

6.9°. Случайная величина ξ имеет распределение

x_i	-1	1	2
$P_\xi(x_i)$	1/4	1/4	1/2

Вычислить:

$$1) M\xi^2 \sin \frac{\pi}{3}\xi; \quad 2) M2^{|\xi|} \cos^2 \frac{\pi}{12}\xi.$$

6.10. Случайная величина ξ имеет распределение

x_i	-2	-1	0	1
$P_\xi(x_i)$	1/4	1/4	1/6	1/3

Найти:

1) распределение случайной величины $\eta = |\xi|$;

2) математическое ожидание и дисперсию η .

6.11°. Случайная величина ξ имеет распределение

x_i	-2	-1	1	2
$P_\xi(x_i)$	1/6	1/6	1/2	1/6

Вычислить

$$M\xi \sin^2 \frac{\pi}{12} \xi.$$

6.12°. Случайная величина ξ имеет распределение

x_i	-2	-1	2
$P_\xi(x_i)$	1/2	1/4	1/4

Вычислить:

$$1) M2^{|\xi|} \cos^2 \frac{\pi}{12} \xi; \quad 2) M\xi 2^{|\xi|}.$$

6.13°. Пусть ξ и η — независимые случайные величины соответственно с распределениями

x_i	0	1	2	3
$P_\xi(x_i)$	1/2	1/6	1/6	1/6

y_j	-1	0	1
$P_\eta(y_j)$	1/4	1/2	1/4

Вычислить

$$M \frac{\xi + 1}{\eta^4 + 2}.$$

6.14°. Пусть ξ и η — независимые случайные величины соответственно с распределениями

x_i	0	1	2
$P_\xi(x_i)$	2/3	1/6	1/6

y_j	-1	0	1
$P_\eta(y_j)$	1/3	1/3	1/3

Вычислить

$$M \frac{\eta^4 - \eta}{\xi + 1}.$$

6.15. Пусть случайная величина ξ принимает значения $-n, -(n-1), \dots, -1, 0, 1, \dots, n-1, n$ с вероятностями $1/(2n+1)$. Вычислить: 1) $M\xi$; 2) $M|\xi|$.

6.16. Случайная величина ξ имеет распределение Пуассона с параметром λ . Вычислить: $M\xi$, $M\xi^2$, $D\xi$.

6.17. Пусть ξ и η — независимые случайные величины, каждая из которых имеет распределение Пуассона с параметром λ . Вычислить:

$$1) M \frac{\xi}{1 + \eta}; \quad 2) M\xi\eta; \quad 3) D\xi\eta; \quad 4) D(\xi + \eta).$$

6.18. Случайная величина ξ имеет геометрическое распределение с параметром p . Вычислить: $M\xi$, $M\xi^2$, $D\xi$.

6.19. Симметричную игральную кость подбрасывают до первого появления шестерки.

1. Сколько раз в среднем необходимо подбросить кость?

2. Какова вероятность того, что будет выполнено не более двух подбрасываний?

6.20. Случайная величина ξ имеет геометрическое распределение с параметром p . Вычислить:

$$1) Mx^\xi, |x| < 1; \quad 2) Me^{it\xi}.$$

6.21. Случайная величина ξ имеет распределение Пуассона с параметром λ . Вычислить:

$$1) Mx^\xi, |x| < 1; \quad 2) Me^{it\xi}.$$

6.22°. Вычислить математическое ожидание и дисперсию случайной величины ξ — числа появлений события в 100 независимых испытаниях, в каждом из которых вероятность события равна 0,7.

6.23°. Трижды подбрасывают монету, вероятность выпадения герба которой составляет $1/3$. Пусть ξ — число выпавших при этом гербов.

Найти распределение случайной величины ξ . Вычислить $M\xi$ и $D\xi$.

6.24°. Симметричную монету подбрасывают 10 раз. Найти распределение случайной величины ξ — числа выпавших гербов. Вычислить $M\xi$ и $D\xi$.

6.25°. Вычислить математическое ожидание и дисперсию случайной величины ξ из задачи 5.26.

6.26°. Вычислить математическое ожидание и дисперсию случайной величины ξ из задачи 5.27.

6.27*. Рассматривается последовательность независимых испытаний с вероятностью успеха p ($0 < p < 1$) в одном испытании. Пусть ξ — случайная величина, равная числу неудач до r -го успеха. Случайная величина ξ имеет отрицательное биномиальное распределение с параметрами (r, p) :

$$P\{\xi = k\} = C_{r+k-1}^k p^r (1-p)^k, \quad k = 0, 1, \dots$$

Доказать, что

$$M\xi = r \frac{1-p}{p}, \quad D\xi = r \frac{1-p}{p^2}.$$

6.28.¹ Для случайной величины $\zeta = (\xi, \eta)$ из задачи 5.8 вычислить

1) $M \max\{\xi, \eta\}$; 2) $M \sin(\pi \max\{\xi, \eta\}/6)$; 3) $M \min\{\xi, \eta\}$.

6.29. По совместному распределению случайных величин ξ и η из задачи 5.10 вычислить:

1) $M \min\{\xi, \eta\}$; 2) $M \cos(\pi \max\{\xi, \eta\})$

3) $M \max\{\xi, \eta - \xi\}$; 4) $M \sin(\pi \min\{\eta, \eta - \xi\}/4)$.

¹Для решения задач 6.28 — 6.31 следует воспользоваться формулой (6.1.3).

6.30. Случайный вектор $\mu = (\mu_1, \mu_2, \dots, \mu_r)$ имеет распределение

$$P\{(\mu_1, \mu_2, \dots, \mu_k, \dots, \mu_r) = (0, 0, \dots, 0, 1, 0, \dots, 0)\} = p_k,$$

у последовательности $(0, 0, \dots, 0, 1, 0, \dots, 0)$ на k -м месте находится единица, на остальных — нули, $k = 1, 2, \dots, r$.

Вычислить

$$M \exp\{i(t, \mu)\} = M \exp\{i(t_1\mu_1 + t_2\mu_2 + \dots + t_r\mu_r)\}.$$

6.31*. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины со значениями в $\mathbb{R}^1$;

$$R^1 = \bigcup_{j=1}^r X_j, X_s \cap X_l = \emptyset, s \neq l,$$

$$P\{\xi_k \in X_j\} = p_j, j = 1, 2, \dots, r;$$

$\nu = (\nu_1, \nu_2, \dots, \nu_r)$ — случайный вектор, j -я компонента ν_j которого равна количеству случайных величин $\xi_1, \xi_2, \dots, \xi_n$, которые попали в X_j , $j = 1, 2, \dots, r$.

Вычислить

$$M \exp\{i(t, \nu)\} = M \exp\{i(t_1\nu_1 + t_2\nu_2 + \dots + t_r\nu_r)\},$$

$t_i \in R^1, i = 1, 2, \dots, r$.

6.32. У большого числа k людей исследуют кровь. Исследование можно организовать двумя способами.

Способ 1. Кровь каждого человека исследуют отдельно. В этом случае необходимо выполнить k анализов.

Способ 2. Кровь k людей смешивают и анализируют полученную смесь. Если результат анализа отрицательный (диагноз не подтверждается), то этого одного анализа достаточно для k людей. Если результат положительный, то кровь каждого человека анализируют отдельно и в целом на k человек необходимо выполнить $k + 1$ анализов. Предположим, что вероятность p положительного результата анализа одна и та же для всех людей и что результаты анализов независимы.

1. Найти вероятность того, что анализ смешанной крови k людей положительный.

2. Вычислить математическое ожидание числа анализов ξ , необходимых во втором способе обследования.

6.33. В урне содержится 2 белых, 3 черных, 5 красных шаров. Из урны наудачу берут три шара. Найти распределения случайных величин: ξ — числа выбранных белых шаров, η — черных, ζ — красных. Вычислить математическое ожидание и дисперсию каждой из случайных величин.

6.34. Пусть ξ — неотрицательная целочисленная случайная величина с распределением

$$P\{\xi = k\} = p_k, \quad k = 0, 1, \dots$$

Функцию $P(t)$, определенную равенством

$$P(t) = Mt^\xi = \sum_{k=0}^{\infty} t^k p_k, \quad |t| \leq 1,$$

будем называть производящей функцией случайной величины ξ (распределения $\{p_k\}$).

Вычислить производящую функцию случайной величины ξ , имеющей: 1° пуассоновское распределение; 2° геометрическое распределение; 3° биномиальное распределение.

6.35. Доказать, что различные дискретные вероятностные распределения имеют разные производящие функции (теорема единственности).

6.36. Доказать, что производящая функция суммы независимых случайных величин равна произведению производящих функций слагаемых (мультипликативное свойство производящих функций).

6.37. Доказать, что сумма независимых пуассоновских случайных величин ξ и η с параметрами θ_1 и θ_2 соответственно является пуассоновской случайной величиной с параметром $\theta_1 + \theta_2$.

6.38. Доказать, что сумма независимых биномиально распределенных случайных величин ξ и η соответственно с параметрами $(n; \theta)$ и $(m; \theta)$ является биномиально распределенной случайной величиной с параметрами $(n + m; \theta)$.

6.39. Пусть $\Omega = \{0, 1, 2, \dots\}$, $P(\omega) = ar^\omega$, $\omega = 0, 1, \dots$, r — произвольное фиксированное число из $(0; 1)$. При каких значениях параметра a функция $P(\omega) = ar^\omega$, $\omega \in \Omega$, задает вероятность на подмножествах Ω ?

Пусть $\xi = \xi(\omega)$ — случайная величина на $\{\Omega, P\}$, заданная равенством $\xi = \xi(\omega) = 2^{-\omega}$. Вычислить $M\xi$, $M\xi^2$.

6.40. Пусть $\Omega = \{0, 1, 2, \dots\}$, $P(\omega) = a\frac{1}{\omega!}$, $\omega \in \Omega$. При каких значениях a пара $\{\Omega, P\}$ является вероятностным пространством?

Пусть $\xi = \xi(\omega)$ — случайная величина на $\{\Omega, P\}$, заданная равенством $\xi = \xi(\omega) = 2^\omega$. Вычислить $M\xi$, $D\xi$.

6.41. Пусть $\Omega = \{0, 1, \dots\}$, $P(\omega) = a\frac{\lambda^\omega}{\omega!}$, $\omega \in \Omega$, λ — произвольное фиксированное положительное число. При каких значениях a пара $\{\Omega, P\}$ является вероятностным пространством?

Пусть $\xi = \xi(\omega)$ — случайная величина на $\{\Omega, P\}$, заданная равенством $\xi = \xi(\omega) = 2^\omega$. Вычислить $M\xi$, $D\xi$.

6.42. Доказать неравенство Чебышёва (теорема 6.1.2).

6.43. Доказать закон больших чисел (теорема 6.1.3).

6.44. Пусть $P_\zeta(x_i, y_j)$ — распределение случайного вектора $\zeta = (\xi, \eta)$. Вычислить $M\xi$.

6.45. Подбрасывают три симметричные игральные кости. У первой три грани занумерованы единицей, три грани — двойкой, у второй три грани занумерованы тройкой, три — четверкой, а у третьей кости три грани занумерованы пятеркой, три — шестеркой. Найти математическое ожидание, дисперсию и распределение суммы выпавших очков.

6.46. Подбрасывают три симметричные игральные кости. Грани первой кости занумерованы числами от единицы до шести. У второй кости две грани занумерованы единицей, две — двойкой, две — тройкой. А у третьей кости три грани занумерованы единицей, три — двойкой. Найти математическое ожидание, дисперсию суммы выпавших очков.

6.47. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины, с конечными дисперсиями, $S_n = \xi_1 + \xi_2 + \dots + \xi_n$. Доказать,

что

$$DS_n = \sum_{k=1}^n D\xi_k + 2 \sum_{i < j} \text{cov}(\xi_i, \xi_j),$$

а если случайные величины $\xi_1, \xi_2, \dots, \xi_n$ попарно независимы, то

$$DS_n = \sum_{k=1}^n D\xi_k.$$

6.48. Пусть ζ_1 и ζ_2 — независимые случайные величины с одинаковыми распределениями и $\xi = \zeta_1 + \zeta_2$, $\eta = \zeta_1 - \zeta_2$. Доказать, что случайные величины ξ и η некоррелированы, т. е. $r(\xi, \eta) = 0$.

6.49. Пусть ξ и η — соответственно сумма и разность очков, выпавших при подбрасывании двух игральных костей. Доказать, что 1) случайные величины ξ и η некоррелированы; 2) ξ и η не являются независимыми.

6.50. Пусть случайная величина ξ принимает значения $-1, +1, -2, +2$ каждое с вероятностью $1/4$, а $\eta = \xi^2$. Найти совместное распределение ξ и η . Доказать, что: 1) случайные величины ξ и η некоррелированы; 2) ξ и η — не являются независимыми случайными величинами.

6.51. Вычислить математическое ожидание случайной величины ξ из задачи 5.3.

Глава 7

Аксиоматика теории вероятностей

7.1 Алгебры, σ -алгебры

Операции над множествами. Далее мы будем иметь дело с множеством (пространством) Ω элементов ω и его подмножествами Ω (подмножества Ω будем обозначать прописными латинскими буквами $A, B, C, \dots$).

Множество полностью задается своими элементами. Если ω является элементом множества A , то это будем записывать так: $\omega \in A$ (и будем читать: “ ω принадлежит A ”); если ω не является элементом множества A , то это будем записывать так: $\omega \notin A$ (и будем читать: “ ω не принадлежит A ”).

Множество, не содержащее ни одного элемента, будем называть *пустым* и будем обозначать символом $\emptyset$.

Если каждый элемент множества A является элементом множества B , будем это обозначать так: $A \subset B$ (и будем читать: “ A — подмножество B ”).

Если каждый элемент множества A является элементом множества B ($A \subset B$) и каждый элемент множества B является элементом множества A ($B \subset A$) (другими словами, совокупность элементов множеств A и B одна и та же), то множества A и B будем отождествлять и будем это обозначать так: $A = B$. Для того чтобы установить равенство $A = B$, достаточно проверить, что $A \subset B$ и

$B \subset A$, т. е. каждый элемент ω множества A является элементом множества B , и каждый элемент ω множества B является элементом множества A .

Из элементов множеств A и B можно строить новые множества. Множество, состоящее из элементов, каждый из которых принадлежит хотя бы одному из множеств A или B , будем называть *объединением* множеств A и B и будем обозначать $A \cup B$.

Множество, состоящее из элементов, принадлежащих как множеству A , так и B , будем называть *пересечением* множеств A и B и будем обозначать $A \cap B$.

Множество, состоящее из элементов, принадлежащих множеству A , но не принадлежащих B , будем называть *разностью* множеств A и B и будем обозначать $A \setminus B$.

Множество, состоящее из элементов Ω , не принадлежащих множеству A , будем называть *дополнением* множества A и будем обозначать $\bar{A}$.

Будем говорить, что множества A и B *не пересекаются* (*непересекающиеся*), если $A \cap B = \emptyset$.

Пример 7.1.1. Пусть $\{A_n\}$ — последовательность непересекающихся множеств ($A_i \cap A_j = \emptyset, i \neq j$) и

$$B_n = \bigcup_{i=n}^{\infty} A_i, \quad n = 1, 2, \dots$$

Доказать, что $\bigcap_{n=1}^{\infty} B_n = \emptyset$.

Решение. Предположим, что $\bigcap_{n=1}^{\infty} B_n \neq \emptyset$, и пусть

$\omega \in \bigcap_{n=1}^{\infty} B_n$. Тогда $\omega \in B_1 = \bigcup_{i=1}^{\infty} A_i$ и, следовательно, ω принадлежит хотя бы одному из множеств $A_i, i = 1, 2, \dots$, а поскольку $A_i \cap A_j = \emptyset, i \neq j$, то только одному множеству, обозначим его через A_{n_0} . Поэтому

$$\omega \notin \bigcup_{i=n_0+1}^{\infty} A_i = B_{n_0+1}.$$

А последнее противоречит предположению $\omega \in \bigcap_{n=1}^{\infty} B_n$.

Алгебра множеств. Непустой класс $\mathfrak{A}$ подмножеств пространства Ω будем называть *алгеброй множеств*, если:

1° вместе с любым множеством $A \in \mathfrak{A}$ его дополнение $\bar{A}$ принадлежит $\mathfrak{A}$;

2° вместе с любыми двумя множествами $A, B \in \mathfrak{A}$ их объединение $A \cup B$ принадлежит $\mathfrak{A}$.

Определение. Пусть $\mathfrak{K}$ — некоторый класс подмножеств Ω . *Наименьшей алгеброй, содержащей класс $\mathfrak{K}$ (алгеброй, порожденной классом $\mathfrak{K}$)*, будем называть алгебру $\mathfrak{A}(\mathfrak{K})$, которая:

1° содержит класс $\mathfrak{K}$, т. е. $\mathfrak{K} \subset \mathfrak{A}(\mathfrak{K})$;

2° содержится в любой другой алгебре $\mathfrak{A}$, содержащей класс $\mathfrak{K}$, т. е. $\mathfrak{A}(\mathfrak{K}) \subset \mathfrak{A}$.

σ -Алгебра множеств. Непустой класс $\mathfrak{F}$ подмножеств пространства Ω будем называть *σ -алгеброй множеств*, если:

1° вместе с каждым множеством $A \in \mathfrak{F}$ его дополнение $\bar{A}$ принадлежит $\mathfrak{F}$;

2° вместе с любой счетной последовательностью множеств $A_i \in \mathfrak{F}$, $i = 1, 2, \dots$, их объединение $\bigcup_{i=1}^{\infty} A_i$ принадлежит $\mathfrak{F}$.

Множества из σ -алгебры $\mathfrak{F}$ еще называют *событиями*.

Определение. Пусть $\mathfrak{K}$ — некоторый класс подмножеств Ω . *Наименьшей σ -алгеброй, содержащей класс $\mathfrak{K}$ (σ -алгеброй, порожденной классом $\mathfrak{K}$)*, будем называть σ -алгебру $\sigma(\mathfrak{K})$, которая:

1° содержит класс $\mathfrak{K}$, т. е. $\mathfrak{K} \subset \sigma(\mathfrak{K})$;

2° содержится в любой σ -алгебре $\mathfrak{F}$, содержащей класс $\mathfrak{K}$, т. е. $\sigma(\mathfrak{K}) \subset \mathfrak{F}$.

Пример 7.1.2. Пусть $\mathfrak{K}_1$ и $\mathfrak{K}_2$ — два класса подмножеств Ω , причем $\mathfrak{K}_1 \subset \mathfrak{K}_2$. Доказать, что

$$\sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}_2).$$

Решение. По определению $\mathfrak{K}_2 \subset \sigma(\mathfrak{K}_2)$, а по условию $\mathfrak{K}_1 \subset \mathfrak{K}_2$, поэтому $\mathfrak{K}_1 \subset \sigma(\mathfrak{K}_2)$. И поскольку $\sigma(\mathfrak{K}_1)$ — наименьшая σ -алгебра, порожденная классом $\mathfrak{K}_1$, а $\sigma(\mathfrak{K}_2)$ —

одна из σ -алгебр, содержащих класс $\mathfrak{K}_1$, то

$$\sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}_2).$$

Борелевские множества на $\mathbb{R}^1$. σ -Алгеброй борелевских множеств на $\mathbb{R}^1$ будем называть σ -алгебру $\mathfrak{B}^1$, порожденную классом $\mathfrak{K}$ промежутков вида $[a, b)$. Множества из σ -алгебры $\mathfrak{B}^1$ будем называть борелевскими множествами на $\mathbb{R}^1$.

Пример 7.1.3. Доказать, что борелевским множеством на $\mathbb{R}^1$ является:

1. Множество $\{a\}$ для каждого $a \in \mathbb{R}^1$.
2. Счетное множество на числовой прямой.
3. Интервал (a, b) .
4. Открытое множество на числовой прямой.
5. Замкнутое множество на числовой прямой.

Решение. Заметим, что σ -алгебра множеств замкнута относительно операции пересечения в счетном числе, поскольку

$$\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}}.$$

1. Очевидно, $\{a\} = \bigcap_{n=1}^{\infty} [a, a + 1/n)$. А поскольку каждый из промежутков $[a, a + 1/n)$ принадлежит $\mathfrak{B}^1$, а $\mathfrak{B}^1$ замкнута относительно операции пересечения в счетном числе, то

$$\{a\} = \bigcap_{n=1}^{\infty} [a, a + 1/n) \in \mathfrak{B}^1.$$

2. Счетное множество A можно представить в виде $\bigcup_{i=1}^{\infty} \{a_i\}$, где $\{a_i\} \in \mathfrak{B}^1$ для всех $i = 1, 2, \dots$ (см. пункт 1), поэтому $A \in \mathfrak{B}^1$.

3. Поскольку $(a, b) = [a, b) \cap \overline{\{a\}}$, $[a, b) \in \mathfrak{B}^1$, $\{a\} \in \mathfrak{B}^1$ и $\mathfrak{B}^1$ замкнута относительно операции пересечения, то $(a, b) \in \mathfrak{B}^1$.

4. Каждое открытое множество O на $\mathbb{R}^1$ можно представить в виде объединения конечного или счетного множества открытых непересекающихся интервалов: $O = \bigcup_i (a_i, b_i)$. Открытые интервалы (a_i, b_i) принадлежат $\mathfrak{B}^1$, поэтому $O = \bigcup_i (a_i, b_i) \in \mathfrak{B}^1$.

5. Замкнутое множество принадлежит $\mathfrak{B}^1$ как дополнение открытого.

Пример иллюстрирует тот факт, что класс борелевских множеств достаточно “широкий”. В книге мы будем иметь дело только с борелевскими множествами.

Борелевские множества на $\mathbb{R}^2$. σ -Алгеброй борелевских множеств на $\mathbb{R}^2$ будем называть σ -алгебру $\mathfrak{B}^2$, порожденную классом $\mathfrak{K}$ прямоугольников вида $[a_1, b_1) \times [a_2, b_2)$.

Борелевские множества на $\mathbb{R}^n$. σ -Алгеброй борелевских множеств на $\mathbb{R}^n$ будем называть σ -алгебру $\mathfrak{B}^n$, порожденную классом $\mathfrak{K}$ параллелепипедов вида $[a_1, b_1) \times [a_2, b_2) \times \dots \times [a_n, b_n)$.

Борелевские множества в метрическом пространстве. σ -Алгеброй борелевских множеств в метрическом пространстве (X, ρ) будем называть σ -алгебру $\mathfrak{B}(X)$, порожденную классом открытых в (X, ρ) подмножеств.

7.2 Аксиомы Колмогорова

Пусть Ω — пространство элементарных событий и $\mathfrak{F}$ — σ -алгебра подмножеств Ω .

Вероятность. Вероятностью, заданной на σ -алгебре $\mathfrak{F}$ подмножеств пространства Ω , будем называть неотрицательную счетноаддитивную нормированную функцию.

Другими словами, вероятность P — это функция множества, заданная на σ -алгебре $\mathfrak{F}$ подмножеств Ω , такая, что:

1° для каждого $A \in \mathfrak{F}$

$$P(A) \geq 0;$$

2° для произвольных $A_i \in \mathfrak{F}$, $i = 1, 2, \dots$, таких, что $A_i \cap A_j = \emptyset, i \neq j$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i);$$

3° $P(\Omega) = 1$.

В частности, если $\Omega = \mathbb{R}^n$, а $\mathfrak{F} = \mathfrak{B}^n$, то вероятность называют *вероятностным распределением на $\mathbb{R}^n$* .

Аксиомы 1°, 2°, 3° из определения вероятности образуют систему аксиом теории вероятностей. В таком виде они были предложены А. Н. Колмогоровым.

Аксиомы 1°, 2°, 3° вероятности являются аналогами свойств неотрицательности, аддитивности и нормированности частоты, см. (3.1.1), (3.1.2), (3.1.3).

Определение. *Вероятностным пространством* будем называть тройку $\{\Omega, \mathfrak{F}, P\}$, где Ω — пространство элементарных событий, $\mathfrak{F}$ — σ -алгебра подмножеств пространства Ω , P — вероятность на σ -алгебре $\mathfrak{F}$.

Вероятностное пространство $\{\Omega, \mathfrak{F}, P\}$ является математической моделью стохастического эксперимента. При этом Ω — пространство элементарных событий — математическая модель множества исходов стохастического эксперимента, $\mathfrak{F}$ — σ -алгебра подмножеств пространства Ω — математическая модель алгебры случайных событий, наблюдаемых в эксперименте, P — вероятность на $\mathfrak{F}$ — математическая модель частоты событий в последовательности экспериментов.

Свойства вероятности. Все перечисленные далее свойства являются следствиями аксиом 1°, 2°, 3° вероятности.

1. Вероятность невозможного события равна нулю:

$$P(\emptyset) = 0.$$

2. Вероятность любого события $A \in \mathfrak{F}$ не превышает 1:

$$P(A) \leq 1, \quad A \in \mathfrak{F}.$$

3. Если $A \subset B$ ($A, B \in \mathfrak{F}$), то вероятность разности событий B и A равна разности вероятностей этих событий:

$$P(B \setminus A) = P(B) - P(A).$$

4. $P(\overline{A}) = 1 - P(A)$.
 5. Для произвольных $A, B \in \mathfrak{F}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (7.2.1)$$

(теорема сложения).

Пример 7.2.1 Теорема сложения допускает следующее обобщение:

$$\begin{aligned} & P(A_1 \cup A_2 \cup \dots \cup A_n) = \\ &= \sum_{1 \leq i_1 \leq n} P(A_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\ & \dots + (-1)^{n-1} \sum_{1 \leq i_1 < i_2 < \dots < i_n \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}). \end{aligned} \quad (7.2.2)$$

Решение. Воспользуемся методом математической индукции. Для двух множеств равенство (7.2.2) справедливо (см. (7.2.1)). Предположим, что равенство (7.2.2) имеет место для $n-1$ множеств, и докажем, что оно справедливо для n множеств. Имеем:

$$\begin{aligned} & P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2 \cup A_3 \cup \dots \cup A_n) - \\ & - P((A_1 \cap A_2) \cup (A_1 \cap A_3) \cup \dots \cup (A_1 \cap A_n)) = P(A_1) + \\ & + \sum_{2 \leq i_1 \leq n} P(A_{i_1}) - \sum_{2 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\ & \dots + (-1)^{n-2} \sum_{2 \leq i_1 < i_2 < \dots < i_n \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) - \\ & - P((A_1 \cap A_2) \cup (A_1 \cap A_3) \cup \dots \cup (A_1 \cap A_n)). \end{aligned}$$

Для завершения доказательства, остается заметить, что

$$\begin{aligned} & \sum_{2 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) + \\ & + \sum_{2 \leq i_1 < i_2 < \dots < i_{k-1} \leq n} P((A_1 \cap A_{i_1}) \cap (A_1 \cap A_{i_2}) \cap \dots \end{aligned}$$

$$\begin{aligned} & \dots \cap (A_1 \cap A_{i_{k-1}}) = \\ & = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}), \end{aligned}$$

$k = 2, 3, \dots, n$.

Пример 7.2.2. Числа $1, 2, \dots, n$ расположены наудачу. Найти вероятность того, что по меньшей мере одно число совпадает с номером своего места. Найти предельное значение этой вероятности при $n \rightarrow \infty$.

Решение. Элементарное событие стохастического эксперимента, состоящего в расположении наудачу n чисел на n местах, — упорядоченная последовательность этих чисел. Естественно считать, что все элементарные события равновероятны, поэтому стохастический эксперимент описывается классической моделью. Заметим, что количество всех элементарных событий равно $n!$

Обозначим через A_{i_s} событие, состоящее в том, что число i_s находится на месте с номером i_s , тогда пересечение $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$, $k = 1, 2, \dots, n$, описывает событие “число i_1 находится на месте с номером i_1 , число i_2 — на месте с номером i_2 , и т. д., число i_k — на месте с номером i_k ” (остальные $n - k$ чисел на $n - k$ местах упорядочены произвольно). Задача сводится к вычислению вероятности события $A_1 \cup A_2 \cup \dots \cup A_n$. Воспользуемся теоремой сложения (см. (7.2.2)):

$$\begin{aligned} & P(A_1 \cup A_2 \cup \dots \cup A_n) = \\ & = \sum_{1 \leq i_1 \leq n} P(A_{i_1}) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \dots \\ & \dots + (-1)^{n-1} \sum_{1 \leq i_1 < i_2 < \dots < i_n \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}). \end{aligned}$$

Вычислим каждую сумму в правой части. Начнем с суммы $\sum_{1 \leq i_1 \leq n} P(A_{i_1})$. Сначала вычислим $P(A_{i_1})$. В событие A_{i_1} входят все последовательности длиной n , у которых

на месте с номером i_1 находится число i_1 . Таких последовательностей, очевидно, $(n-1)!$ Согласно формуле классической вероятности

$$P(A_{i_1}) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Поэтому

$$\sum_{1 \leq i_1 \leq n} P(A_{i_1}) = \sum_{1 \leq i_1 \leq n} \frac{1}{n} = 1.$$

Далее рассмотрим $\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$.

Вычислим $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$. В $A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}$ входят последовательности длиной n , у которых на месте с номером i_1 находится число i_1 , на месте с номером i_2 — число i_2 , и т. д., на месте с номером i_k — число i_k , а остальные $n-k$ чисел на $n-k$ оставшихся местах, упорядочены произвольно. Таких последовательностей, очевидно, $(n-k)!$ По формуле классической вероятности

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

Поэтому

$$\begin{aligned} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) &= \\ &= \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \frac{(n-k)!}{n!}. \end{aligned}$$

Количество одинаковых слагаемых в последней сумме равно числу способов выбрать k индексов $i_1, i_2, \dots, i_k$ из промежутка $[1, n]$ так, чтобы $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Очевидно, это число равно C_n^k . Так что,

$$\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = C_n^k \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Таким образом,

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} + \dots + (-1)^{n-1} \frac{1}{n!}.$$

Заметим, что при $n \rightarrow \infty$ последняя сумма сходится к $1 - 1/e$.

7.3 Задачи

АЗ: 7.1, 7.12, 7.13, 7.14, 7.15(1, 2), 7.16(1), 7.19.

СЗ: 7.2, 7.9, 7.10, 7.15(3, 4, 5), 7.16(2), 7.20.

7.1°. Доказать, что: а) $A \cup B = A \cup (B \setminus (A \cap B))$, причем $A \cap (B \setminus (A \cap B)) = \emptyset$; б) $A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$, причем, если $B_i \cap B_j = \emptyset, i \neq j$, то $(A \cap B_i) \cap (A \cap B_j) = \emptyset, i \neq j$.

7.2°. Последовательность подмножеств $\{A_n\}$ плоскости определена так:

$$A_n = \{(x, y) : x^2 + y^2 \leq 1/n^2\}, \quad n = 1, 2, \dots$$

Описать подмножества

$$A = \bigcup_{n=1}^{\infty} A_n, \quad B = \bigcap_{n=1}^{\infty} A_n.$$

7.3°. Доказать равенства:

а) $A \setminus B = A \setminus (A \cap B) = (A \cup B) \setminus B$;

б) $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$;

в) $(A \setminus C) \cap (B \setminus C) = (A \cap B) \setminus C$.

7.4°. Пусть $A_n = \left[\frac{1}{2n}, \frac{1}{n}\right)$, $n = 1, 2, \dots$. Описать множества

$$A = \bigcup_{n=1}^{\infty} A_n, \quad B = \bigcap_{n=1}^{\infty} A_n.$$

7.5. 1° Пусть $\{A_n\}$ — произвольная последовательность множеств. Доказать, что

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n,$$

где $B_1 = A_1$, $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$, $n = 1, 2, \dots$, причем множества B_n попарно не пересекаются.

2° Пусть $\{A_i\}$ — возрастающая последовательность множеств, т. е.

$$A_i \subset A_{i+1}, \quad i = 1, 2, \dots$$

Доказать, что

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1}),$$

где $A_0 = \emptyset$, причем

$$(A_i \setminus A_{i-1}) \cap (A_j \setminus A_{j-1}) = \emptyset, \quad i \neq j.$$

7.6. Пусть I — некоторое множество индексов. Доказать, что

$$\overline{\bigcup_{\alpha \in I} A_{\alpha}} = \bigcap_{\alpha \in I} \overline{A_{\alpha}}; \quad \overline{\bigcap_{\alpha \in I} A_{\alpha}} = \bigcup_{\alpha \in I} \overline{A_{\alpha}}.$$

7.7°. Пусть $\mathfrak{A}$ — алгебра множеств, $A, B \in \mathfrak{A}$. Доказать, что $A \cap B \in \mathfrak{A}$ (алгебра замкнута относительно операции пересечения).

7.8°. Пусть $\mathfrak{A}$ — алгебра множеств. Доказать:

а) если $A_k \in \mathfrak{A}$, $k = 1, 2, \dots, n$, то $\bigcup_{k=1}^n A_k \in \mathfrak{A}$;

б) если $A_k \in \mathfrak{A}$, $k = 1, 2, \dots, n$, то $\bigcap_{k=1}^n A_k \in \mathfrak{A}$;

в) $\Omega \in \mathfrak{A}$; $\emptyset \in \mathfrak{A}$.

7.9. Доказать, что:

а) класс всех подмножеств множества Ω является алгеброй;

б) для каждого класса $\mathfrak{K}$ подмножеств Ω существует по меньшей мере одна алгебра, содержащая класс $\mathfrak{K}$;

в) пересечение любой совокупности алгебр является алгеброй.

7.10. Доказать, что для каждого класса $\mathfrak{K}$ существует наименьшая алгебра, содержащая класс $\mathfrak{K}$.

7.11°. Пусть $\mathfrak{F}$ — σ -алгебра множеств, $A_n \in \mathfrak{F}$, $n = 1, 2, \dots$. Доказать, что $\bigcap_{n=1}^{\infty} A_n \in \mathfrak{F}$.

7.12. Доказать, что:

а) класс всех подмножеств множества Ω является σ -алгеброй;

б) для каждого класса $\mathfrak{K}$ подмножеств Ω существует по меньшей мере одна σ -алгебра, содержащая класс $\mathfrak{K}$;

в) пересечение любой совокупности σ -алгебр является σ -алгеброй.

7.13. Доказать, что для каждого класса $\mathfrak{K}$ существует наименьшая σ -алгебра $\sigma(\mathfrak{K})$, содержащая класс $\mathfrak{K}$.

7.14. Пусть $\mathfrak{K}$ — класс открытых интервалов (a, b) на числовой прямой. Доказать, что наименьшая σ -алгебра $\sigma(\mathfrak{K})$, содержащая класс $\mathfrak{K}$, является σ -алгеброй борелевских множеств на $\mathbb{R}^1$.

7.15. Доказать, что σ -алгебры, порожденные каждым из перечисленных далее классов, являются σ -алгебрами борелевских множеств на $\mathbb{R}^1$:

1) $\mathfrak{K}_1 = \{[a, b]; a, b \in \mathbb{R}^1\}$;

2) $\mathfrak{K}_2 = \{(-\infty, x]; x \in \mathbb{R}^1\}$;

3) $\mathfrak{K}_3 = \{(-\infty, x); x \in \mathbb{R}^1\}$;

4) $\mathfrak{K}_4 = \{(x, +\infty); x \in \mathbb{R}^1\}$;

5) $\mathfrak{K}_5 = \{[x, +\infty); x \in \mathbb{R}^1\}$.

7.16. Доказать, что σ -алгебра борелевских множеств порождается:

1) классом открытых множеств на числовой прямой,

2) классом замкнутых множеств на числовой прямой.

7.17. Пусть $\mathfrak{A}$ — алгебра множеств на $\mathbb{R}^1$, элементами которой являются конечные объединения промежутков

вида $[a, b)$, где $a, b \in \mathbb{R}^1$ (a и b могут быть равны $-\infty$ и $+\infty$). Доказать, что алгебра $\mathfrak{A}$ порождает σ -алгебру борелевских множеств на $\mathbb{R}^1$.

7.18. Пусть $\mathfrak{F}$ — σ -алгебра множеств Ω , B — произвольное, но фиксированное множество из $\mathfrak{F}$. Доказать, что класс множеств $\mathfrak{F}_B$ вида $A \cap B$, где $A \in \mathfrak{F}$, является σ -алгеброй.

Примечание. Дополнение берется до множества B .

7.19. Пусть $\mathfrak{B}^1$ — σ -алгебра борелевских множеств на $\mathbb{R}^1$, B — произвольное, но фиксированное множество из $\mathfrak{B}^1$. Пусть $\mathfrak{B}_B^1$ — класс множеств вида $B \cap A$, где $A \in \mathfrak{B}^1$. Доказать, что $\mathfrak{B}_B^1$ является σ -алгеброй.

Примечание. Дополнение берется до множества B .

7.20. $\mathfrak{B}^1$ — σ -алгебра борелевских множеств на $\mathbb{R}^1$. Пусть $\mathfrak{B}_{[a,b)}^1$ — класс множеств вида $B \cap [a, b)$, $B \in \mathfrak{B}^1$, $[a, b)$ — фиксированный промежуток. Доказать, что $\mathfrak{B}_{[a,b)}^1$ является σ -алгеброй.

Примечание. Дополнение берется до множества $[a, b)$.

7.21. Пусть класс $\mathfrak{K}$ образуют множества $A_i \subset \Omega$, $i = 1, 2, \dots, n$, причем $A_i \cap A_j = \emptyset, i \neq j$, и $\bigcup_{i=1}^n A_i = \Omega$.

Описать σ -алгебру, порожденную классом $\mathfrak{K}$.

7.22. Пусть класс $\mathfrak{K}$ образуют множества $A_i \subset \Omega$, $i = 1, 2, \dots$, причем $A_i \cap A_j = \emptyset, i \neq j$, и $\bigcup_{i=1}^{\infty} A_i = \Omega$.

Описать алгебру, порожденную классом $\mathfrak{K}$. Описать σ -алгебру, порожденную классом $\mathfrak{K}$.

7.23*. Доказать, что в $\mathbb{R}^n$ борелевским множеством является:

- а) множество $\{a\}$, для каждой точки $a \in \mathbb{R}^n$;
- б) счетное множество;
- в) открытое множество;

7.24. Доказать, что класс бесконечных интервалов $I_{a_1, a_2, \dots, a_n} =$

$$= \{(x_1, x_2, \dots, x_n) : x_1 < a_1, x_2 < a_2, \dots, x_n < a_n\},$$

$a_i \in \mathbb{R}^1, i = 1, 2, \dots, n$, порождает σ -алгебру $\mathfrak{B}^n$.

7.25. Доказать, что класс $\mathfrak{K}$ открытых множеств в $\mathbb{R}^n$ порождает σ -алгебру борелевских множеств $\mathfrak{B}^n$.

7.26. Доказать, что класс $\mathfrak{K}$ замкнутых множеств в $\mathbb{R}^n$ порождает σ -алгебру борелевских множеств $\mathfrak{B}^n$.

7.27. Пусть (X, ρ) — сепарабельное метрическое пространство. Доказать, что:

- 1) одноточечное множество является борелевским;
- 2) счетное множество является борелевским;
- 3) замкнутое множество является борелевским.

7.28. Пусть (X, ρ) — сепарабельное метрическое пространство, $\mathfrak{F}$ — класс замкнутых множеств, $\mathfrak{V}$ — класс множеств вида $B(x, r) = \{y : \rho(x, y) < r\}$, $\mathfrak{W}$ — класс множеств вида $\overline{B}(x, r) = \{y : \rho(x, y) \leq r\}$.

Доказать, что каждый из перечисленных классов порождает σ -алгебру борелевских множеств в метрическом пространстве (X, ρ) .

7.29. Пусть $\{A_n\}$ — последовательность подмножеств множества Ω . Обозначим через $\overline{\lim} A_n$ множество элементов ω , принадлежащих бесконечному числу множеств A_n ; через $\underline{\lim} A_n$ — множество элементов, принадлежащих всем множествам A_n , исключая, быть может, конечное их число.

Доказать, что

$$\underline{\lim} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m; \quad \overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

7.30. Пусть $\{A_n\}$ — монотонно убывающая последовательность подмножеств Ω , т. е. для всех n имеет место включение

$$A_{n+1} \subset A_n.$$

Доказать, что

$$\overline{\lim} A_n = \underline{\lim} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Глава 8

Геометрические вероятности

8.1 Определения. Примеры

Стохастический эксперимент состоит в бросании наудачу точки в множество B из $\mathbb{R}^n$ (например, на отрезок $[a, b]$, в прямоугольник $[a_1, b_1] \times [a_2, b_2]$, параллелепипед $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ и т. д.). В качестве математической модели этого стохастического эксперимента естественно рассматривать вероятностное пространство $\{B, \mathfrak{B}_B^n, P\}$, где B — борелевское¹ множество из $\mathbb{R}^n$, $\mathfrak{B}_B^n$ — класс борелевских подмножеств B , P — вероятность на классе $\mathfrak{B}_B^n$, определяемая для каждого A из $\mathfrak{B}_B^n$ равенством

$$P(A) = \frac{L(A)}{L(B)}, \quad (8.1.1)$$

L — мера Лебега на $\mathbb{R}^n$ (значение L на параллелепипедах $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$ равно $\prod_{i=1}^n (b_i - a_i)$, в частности, $L([a_1, b_1] \times [a_2, b_2]) = (b_1 - a_1) \times (b_2 - a_2)$, $L([a, b]) = b - a$). Определенную таким образом вероятность будем называть *геометрической* (разумеется, множество B должно удовлетворять условию $0 < L(B) < \infty$).

¹Относительно борелевских множеств см. гл. 7.

Далее в слова “точку наудачу бросают в множество B ” мы будем вкладывать вполне конкретное содержание: брошенная точка может попасть в любое (борелевское) подмножество B' множества B , и вероятность того, что точка окажется в B' , пропорциональна мере Лебега $L(B')$ множества B' . Формально это обозначает, что в качестве математической модели стохастического эксперимента, состоящего в бросании наудачу точки в множество B , мы будем рассматривать $\{B, \mathfrak{B}_B^n, P\}$, где P — геометрическая вероятность на $\mathfrak{B}_B^n$.

Пример 8.1.1 *На отрезок $[0; 1]$ наудачу бросают пару точек. Вычислить вероятность того, что расстояние между ними меньше $1/2$.*

Решение. Точки будем считать различимыми (например, они окрашены в разные цвета). Пусть x и y — координаты этих точек.

Упорядоченной паре точек на отрезке $[0; 1]$ с координатами x и y соответствует одна точка в квадрате $[0; 1] \times [0; 1]$ с координатами (x, y) и наоборот. Поэтому бросание наудачу пары точек на отрезок $[0; 1]$ равносильно бросанию наудачу одной точки в квадрат $[0; 1] \times [0; 1]$. При этом парам точек на отрезке $[0; 1]$, для которых расстояние меньше $1/2$ (т. е. $|y - x| < 1/2$), соответствует множество точек (x, y) квадрата $[0; 1] \times [0; 1]$, координаты которых удовлетворяют соотношению $|y - x| < 1/2$, и наоборот. Обозначим это множество точек квадрата через A :

$$\begin{aligned} A &= \{(x, y) \in [0; 1] \times [0; 1] : |y - x| < 1/2\} = \\ &= \{(x, y) \in [0, 1] \times [0, 1] : x - 1/2 < y < x + 1/2\} \end{aligned}$$

(см. рис. 8.1.1, множество A заштриховано).

Вероятность того, что расстояние между брошенными наудачу на отрезок $[0; 1]$ точками будет меньше $1/2$, равна вероятности попадания брошенной наудачу в квадрат $[0; 1] \times [0; 1]$ точки в множество A .

Математической моделью стохастического эксперимента, состоящего в бросании наудачу точки в квадрат $[0; 1] \times [0; 1]$, является вероятностное пространство $\{B, \mathfrak{B}_B^2, P\}$, где $B = [0; 1] \times [0; 1]$, $\mathfrak{B}_B^2$ — класс (σ -алгебра) борелевских

подмножеств квадрата $[0; 1] \times [0; 1]$, P — геометрическая вероятность (см. (8.1.1)) на $\mathfrak{B}_B^2$. Поэтому

$$P(A) = \frac{L(A)}{L(B)} = \frac{3/4}{1} = \frac{3}{4}.$$

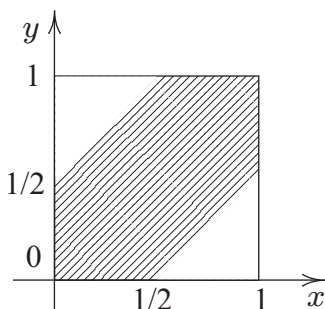


Рис. 8.1.1: К примеру 8.1.1

Пример 8.1.2. *Стержень длиной 1 наудачу разламывают на три части. Вычислить вероятность того, что из образовавшихся при этом частей можно построить треугольник.*

Решение. Стержень длиной 1 удобно интерпретировать как отрезок $[0; 1]$ на координатной прямой. Разламывать отрезок $[0; 1]$ на три части будем следующим образом: наудачу бросим на него две точки и разломаем отрезок в этих точках (точки будем считать различимыми, их координаты обозначим x, y). При этом образуются три отрезка: крайний левый (длиной $\min\{x, y\}$), средний (длиной $|x - y|$), крайний правый (длиной $1 - \max\{x, y\}$). Из этих отрезков можно построить треугольник, если длина каждого из них меньше суммы длин двух других:

$$\begin{cases} \min\{x, y\} < |x - y| + (1 - \max\{x, y\}); \\ |x - y| < \min\{x, y\} + (1 - \max\{x, y\}); \\ 1 - \max\{x, y\} < \min\{x, y\} + |x - y|. \end{cases} \quad (8.1.2)$$

Бросание двух точек на отрезок $[0; 1]$ равносильно бросанию одной точки в квадрат $[0; 1] \times [0; 1]$ (см. также при-

мер 8.1.1). При этом парам точек на отрезке $[0; 1]$, координаты которых удовлетворяют неравенствам (8.1.2), соответствуют точки (x, y) квадрата $[0; 1] \times [0; 1]$, координаты которых удовлетворяют тем же неравенствам. Обозначим это множество точек квадрата через A . Его удобно представить в виде

$$\begin{aligned} A &= A \cap (\{(x, y) : x \leq y\} \cup \{(x, y) : x > y\}) = \\ &= (A \cap \{(x, y) : x \leq y\}) \cup (A \cap \{(x, y) : x > y\}) = \\ &= \{(x, y) : x < 1/2, x + 1/2 > y, y > 1/2\} \cup \\ &\cup \{(x, y) : y < 1/2, x - 1/2 < y, x > 1/2\} \end{aligned}$$

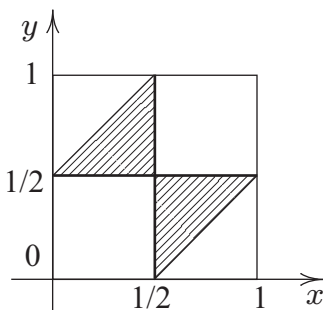


Рис. 8.1.2: К примеру 8.1.2

Так что вычисление вероятности построения треугольника из частей отрезка сводится к вычислению вероятности попадания брошенной в квадрат $[0; 1] \times [0; 1]$ точки в множество A (см. рис. 8.1.2, множество A заштриховано). А поскольку точку бросают наудачу, то искомая вероятность вычисляется как геометрическая:

$$P(A) = \frac{L(A)}{L([0; 1] \times [0; 1])} = \frac{1/4}{1} = \frac{1}{4}.$$

Пример 8.1.3 (задача Бюффона). *Плоскость разграфлена параллельными прямыми, находящимися на расстоянии $2a$ одна от другой. На плоскость наудачу² бросают иглу длиной $2l$ ($2l < 2a$). Найти вероятность того, что игла пересечет одну из прямых.*

Решение. Обозначим через x расстояние от середины иглы до ближайшей прямой, через φ — угол между иглой и прямой (φ будем откладывать против часовой стрелки от направления иглы до направления прямой (рис. 8.1.3)).

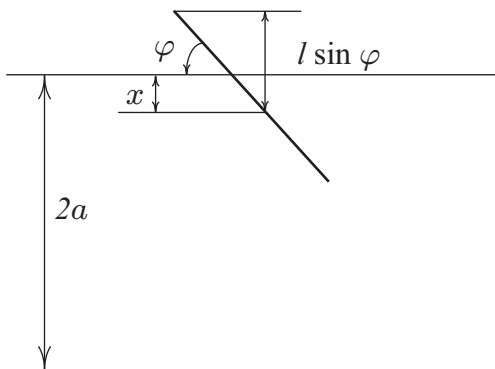


Рис. 8.1.3: Игла на разграфленной плоскости

Тогда игла на плоскости, разграфленной параллельными прямыми, задает упорядоченную пару чисел (φ, x) , $\varphi \in [0, \pi]$, $x \in [0, a]$. Та же пара чисел (φ, x) задает на координатной плоскости φ, x точку с координатами (φ, x) , принадлежащую прямоугольнику $B = [0, \pi] \times [0, a]$ (см. рис. 8.1.4). И обратно: каждая точка из прямоугольника B на плоскости φ, x задает пару чисел (φ, x) , а вместе с ней и положение иглы относительно параллельных пря-

²В задаче оборот “наудачу бросают иглу” обозначает следующее: 1° середина иглы (точка) равномерно распределена на отрезке длиной $2a$, перпендикулярном параллельным прямым; 2° угол φ , образованный иглой с параллельными прямыми, распределен равномерно на промежутке $[0, \pi]$; 3° случайные величины x — расстояние от центра иглы до ближайшей прямой — и φ являются независимыми.

мых на плоскости. Поэтому бросание иглы на плоскость, разграфленную параллельными прямыми, и регистрация ее положения относительно прямых (см. рис. 8.1.3) равносильны бросанию наудачу точки в прямоугольник $B = [0, \pi] \times [0, a]$ (см. рис. 8.1.4). При этом игла пересекает с прямую тогда и только тогда, когда выполняется неравенство $x \leq l \sin \varphi$ (см. рис. 8.1.3) или, что то же, когда брошенная в прямоугольник $B = [0, \pi] \times [0, a]$ точка попадает в область A , ограниченную кривой $x = l \sin \varphi$ и осью $O\varphi$ (см. рис. 8.1.4). А поскольку точку бросают наудачу, то вероятность p попадания точки в множество A (вероятность пересечения иглой прямой), вычисляется как геометрическая вероятность (см. (8.1.1)). Следовательно.

$$p = \frac{1}{a\pi} \int_0^{\pi} l \sin \varphi d\varphi = \frac{2l}{a\pi}.$$

Интересно, что полученное соотношение можно использовать для экспериментального определения числа π .

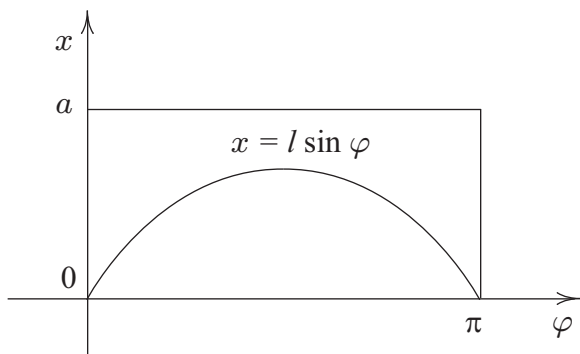


Рис. 8.1.4: К задаче Бюффона

Представим, что игла брошена на плоскость n раз (n — большое), при этом игла m раз пересекла прямую. Если построенная модель адекватно описывает эксперимент, то при больших n частота m/n числа пересечений

иглой прямой должна быть близка к вероятности p , т. е.
 $\frac{2l}{a\pi} \approx \frac{m}{n}$. Отсюда

$$\pi \approx \frac{2l}{a} \frac{n}{m}.$$

8.2 Задачи

АЗ: 8.1°(1, 2, 3, 10), 8.2°(1, 2, 3), 8.4(1, 7, 11, 14), 8.12°, 8.18, 8.20, 8.29, 8.31.

СЗ: 8.3°(3, 6, 8), 8.4(2, 10, 15), 8.11°, 8.16, 8.17°, 8.19, 8.25, 8.28, 8.30.

8.1°. В квадрат $[0;1] \times [0;1]$ наудачу бросают точку. Вычислить вероятность того, что ее координаты (x, y) связаны соотношениями:

- | | |
|---|--------------------------|
| 1) $xy \leq 1/2$; | 6) $x^2 + y^2 < 1/4$; |
| 2) $\min\{e^{-x}, \sqrt{y}\} \geq e^{-1/2}$; | 7) $x + y \leq 1/3$; |
| 3) $\min\{y - x^2, x - y^2\} \geq 0$; | 8) $y + 1/2 \leq 1/x$; |
| 4) $\max\{y - e^{-x}, y - 3/4\} \geq 0$; | 9) $y < x^2/2$; |
| 5) $y > -x^2 + 1/9$; | 10) $ y - x \geq 2/3$. |

8.2°. В квадрат $[0;1] \times [0;1]$ наудачу бросают точку. Обозначим ее координаты через (x, y) .

Найти вероятности следующих событий:

- 1) площадь прямоугольника со сторонами x и y больше $1/2$;
- 2) расстояние от брошенной точки до начала координат меньше $1/4$;
- 3) расстояние от брошенной точки до точки $(1,1)$ больше 1;
- 4) расстояние от брошенной точки до точки $(0,1)$ меньше 1.

8.3°. На отрезок $[0;1]$ наудачу бросают пару точек. Пусть x — координата одной точки, y — другой. Найти вероятности следующих событий:

- 1) $\max\{x, y\} < 1/2$; 8) $\max\{x^2, y\} < a, 0 < a < 1$;
- 2) $\max\{x, y\} > 1/3$; 9) $\max\{x, y^2\} > a, 0 < a < 1$;
- 3) $\min\{x, y\} < 1/4$; 10) $ye^x \leq 1$;
- 4) $\min\{x, y\} > 1/2$; 11) $y + \sqrt{1 - x^2} - 1 \geq 0$;
- 5) $x + \sqrt{1 - y^2} \leq 1$; 12) $y + \sqrt{x - x^2} \leq 1$.
- 6) $y + \sqrt{2x - x^2} \leq 1$; 13) $x^2 - y^2 - x + y \geq 0$.
- 7) $x + \sqrt{y - y^2} \geq 1$; 14) $(y - x)(y - 1/2) \geq 0$.

8.4. На отрезок $[-2; 2]$ наудачу бросают пару точек. Пусть x — координата одной точки, y — другой. Найти вероятности следующих событий:

- 1) $x + |x| = y + |y|$; 8) $(y - 2x)(y + 2x) \geq 0$;
- 2) $x - |x| = y - |y|$; 9) $(|x| + |y| - 1)(|x| + |y| - 2) \leq 0$;
- 3) $[y] = [x]$; 10) $|x - 1| + |y - 1| \leq 1$;
- 4) $[y] = -[x]$; 11) $(y - x)(y + x - 2) \geq 0$;
- 5) $[y] = [x - 1]$; 12) $([y] - [x])(\{y\} - \{x\}) \geq 0$;
- 6) $\{y\} \leq \{x\}$; 13) $(x - \operatorname{sign} x)^2 + (y - \operatorname{sign} y)^2 \leq 1$;
- 7) $|x| + |y| \leq 1$; 14) $|x - \operatorname{sign} x| + |y - \operatorname{sign} y| \leq 1$;
- 15) $(|x - 1| + |y - 1| - 1)(|x - 1| + |y - 1| - 2) \leq 0$;
- 16) $((x - \operatorname{sign} x)^2 + (y - \operatorname{sign} y)^2 - 1)(|x - \operatorname{sign} x| + |y - \operatorname{sign} y| - 1) \leq 0$;

(Здесь $[a]$ — целая часть числа a , $\{a\} = a - [a]$ — дробная часть числа a .)

8.5°. На паркетный пол бросают монету радиуса r . Паркет имеет форму прямоугольников со сторонами a и b ($a < b$, $2r < a$).

Найти вероятность того, что монета пересечет меньшую сторону какого-нибудь прямоугольника, если известно, что она пересекла одну из сторон.

8.6. На отрезке длиной l наудачу выбраны две точки. Какова вероятность того, что расстояние между ними не превосходит kl ($0 < k < 1$)?

8.7. Два судна должны подойти к одному причалу. Моменты прихода судов к причалу — независимые случайные события, равновозможные на протяжении суток. Найти вероятность того, что одно из судов будет вынуждено ждать освобождения причала, если время стоянки первого судна (судна номер 1) — 1 час, а второго — 2 часа.

8.8°. На паркетный пол наудачу бросают монету диаметра d . Паркет имеет форму квадратов со стороной a ($a > d$). Найти вероятность того, что монета не пересечет ни одну из сторон квадратов паркета.

8.9. В круг радиуса R наудачу бросают точку. Найти вероятность того, что точка окажется внутри правильного n -угольника, вписанного в круг.

8.10. На окружности единичного радиуса с центром в начале координат наудачу выбирают точку. Найти вероятность того, что: а) проекция точки на ось Ox находится от центра окружности на расстоянии, не превышающем r ($r < 1$); б) расстояние от выбранной точки до точки с координатами $(1; 0)$ не превышает r .

8.11°. В круг радиуса R наудачу бросают точку. Найти вероятность того, что расстояние от этой точки до центра круга: 1) не превышает r ; 2) больше r .

8.12°. В круг радиуса R с центром в точке O наудачу независимо одна от другой бросают N точек.

Вычислить вероятность того, что:

1) расстояние от центра круга до самой удаленной точки не превышает r ($r < R$);

2) расстояние от центра круга до ближайшей точки превышает r ($r < R$);

3) все точки попадут в круг радиуса r ($r < R$) с центром в точке O ;

4) ни одна точка не попадет в круг радиуса r ($r < R$) с центром в точке O ;

5) все точки попадут в круг радиуса r ($r < R$), содержащийся в данном круге;

6) ни одна точка не попадет в круг радиуса r ($r < R$), содержащийся в данном круге;

7) все точки попадут в квадрат со стороной a , содержащийся в данном круге;

8) ни одна точка не попадет в квадрат со стороной a , содержащийся в данном круге.

8.13. На отрезок AB длиной a наудачу независимо одна от другой бросают пять точек. Найти вероятность того, что две точки окажутся на расстоянии, меньшем b ($b < a$) от точки A , а три — на расстоянии, большем b .

8.14. Из наблюдений (с телескопом и спектрометром) можно заключить, что ряд химических элементов, най-

денных в земной коре, представлен также и на Солнце. Это заключение основано на физическом законе, открытом Густавом Кирхгофом, согласно которому пары элементов поглощают в точности те же лучи света, которые они излучают (при высокой температуре химических элементы превращаются в светящиеся пары). Пользуясь призмой, можно обнаружить в излучении Солнца (в солнечном спектре) некоторую последовательность параллельных линий (фраунгоферовы линии). Мы можем обнаружить последовательность параллельных спектральных линий также в излучении других элементов, в частности, железа. (В качестве модели фраунгоферовых линий (и спектральных линий излучения химических элементов) будем рассматривать последовательность параллельных линий на плоскости.)

В наблюдениях Г. Кирхгофа 60 спектральных линий излучения железа совпали³ с фраунгоферовыми линиями излучения Солнца. Оцените вероятность того, что эти совпадения случайны, если на шкале, использованной Г. Кирхгофом, линии на расстоянии меньшем 0,5 мм уже неразличимы, а среднее расстояние между соседними солнечными фраунгоферовыми линиями равно 2 мм.

8.15. На отрезок $[0; 1]$ наудачу бросают три точки, ξ, η, ζ — их координаты.

1. Вычислить вероятность того, что из отрезков длиной ξ, η, ζ можно построить треугольник.

2. Вычислить вероятность того, что $\xi + \eta + \zeta \leq 3/2$.

3. Вычислить вероятность того, что $\max\{\xi, \eta, \zeta\} \leq t$, $0 < t < 1$.

8.16. Два лица, решив встретиться в течение часа, договорились, что каждый независимо от другого приходит на место встречи в наудачу выбранный момент указанного часа.

1. Если лица договорились, что каждый ожидает в течение времени t ($t < 1$), после чего уходит с места встречи, то какова вероятность того, что встреча состоится?

³Заметим, что никакое физическое наблюдение не является абсолютно точным. Поэтому две линии, которые мы считаем совпадающими, в действительности могут быть различными, но в пределах точности наших наблюдений мы не можем их различить.

2. Какова вероятность того, что данное лицо придет на место встречи: а) раньше другого; б) раньше другого на время, не меньшее чем q ($q < 1$)?

8.17°. В сферу радиуса R , с центром в точке O наудачу независимо одну от другой бросили N точек. Вычислить вероятность того, что:

1) расстояние от центра до самой удаленной точки не превышает r ($r < R$);

2) расстояние от центра до ближайшей точки превышает r ($r < R$);

3) все точки попадут в сферу радиуса r ($r < R$) с центром в точке O ;

4) ни одна точка не попадет в сферу радиуса r ($r < R$) с центром в точке O ;

5) все точки попадут в сферу радиуса r ($r < R$), содержащуюся в данной сфере;

6) ни одна точка не попадет в сферу радиуса r ($r < R$), содержащуюся в данной сфере;

7) все точки попадут в куб со стороной a , содержащийся в данной сфере;

8) ни одна точка не попадет в прямоугольный параллелепипед со сторонами a, b, c , содержащийся в данной сфере.

8.19. На отрезке $[-1; 1]$ наудачу выбирают две точки. Пусть p и q — координаты этих точек. Найти вероятность того, что квадратное уравнение $x^2 + px + q = 0$: 1) имеет действительные корни; 2) не имеет действительных корней.

8.20. На отрезок наудачу одну за одной бросают три точки. Какова вероятность того, что третья по счету точка попадет между двумя первыми?

8.21°. В круг вписан квадрат. Точку наудачу бросают в круг. Найти вероятность того, что она попадет в квадрат.

8.22*. Пусть $x = (x_1, x_2)$ — координаты точки, наудачу брошенной в квадрат $[0; 1] \times [0; 1]$. При каких значениях r события $A_r = \{|x_1 - x_2| \geq r\}$ и $B_r = \{x_1 + x_2 \leq 3r\}$ независимы?

8.23. Пусть $x = (x_1, x_2)$ — координаты точки, наудачу брошенной в квадрат $[0; 1] \times [0; 1]$,

$$A_1 = \{(x_1, x_2) : x_1 \leq 1/2\}, \quad A_2 = \{(x_1, x_2) : x_2 \leq 1/2\},$$

$$A_3 = \{(x_1, x_2) : (x_1 - 1/2)(x_2 - 1/2) \leq 0\}.$$

Доказать, что любые два из событий A_1, A_2, A_3 независимы, но все три не являются независимыми.

8.24. Плоскость разбита сеткой прямых: 1) на квадраты со стороной 1; 2) на правильные треугольники со стороной 1.

На плоскость наудачу бросают монету диаметра 1. Какова вероятность того, что монета накроет одну из вершин сетки?

8.25. На отрезке $[0; 1]$ наудачу последовательно выбирают три числа. Какова вероятность того, что:

- 1) число, выбранное последним, наибольшее;
- 2) числа выбраны в порядке возрастания?

8.26. Отрезок разделили на три равные части. Какова вероятность того, что три точки, наудачу брошенные на отрезок, попадут в три разные части?

8.27*. Монету, радиус которой r и высота h , наудачу бросают на горизонтальную плоскость. Найти вероятность того, что монета станет на ребро.

8.28. На отрезок $[0; 1]$, разделенный на три равные части, наудачу бросают три точки. Вычислить вероятности следующих событий:

- 1) все три точки попадут на одну часть;
- 2) занятыми окажутся не менее двух частей;
- 3) на среднюю часть не попадет ни одна точка;
- 4) на крайние части не попадет ни одна точка.

8.29. На окружности наудачу выбирают три точки. Какова вероятность того, что треугольник с вершинами в этих точках остроугольный?

8.30. На окружности наудачу выбраны три точки. Вычислить вероятность того, что в треугольнике с вершинами в этих точках:

- 1) имеется угол меньший 30° ;
- 2) все углы больше 30° ;
- 3) имеется угол 90° .

8.31⁴. На окружности наудачу выбраны точки A, B, C, D . Вычислить вероятность того, что отрезки AC и BD пересекаются.

8.32*. На плоской горизонтально размещенной фольге находится точечный источник радиоактивного излучения, посылающий лучи равномерно во всех направлениях пространства. Если параллельно фольге на единичном расстоянии от нее поставить экран, то на нем можно наблюдать точечные вспышки, вызванные радиоактивным излучением. Найти вероятность того, что очередная вспышка будет наблюдаться на экране внутри круга радиуса R , центр которого находится над источником излучения.

8.33*. Шарикоподшипник можно собрать, если радиус R внешнего кольца, радиус r внутреннего кольца и диаметр d шарика удовлетворяют условиям

$$0 \leq R - r - d \leq \delta.$$

Предположим, что R, r, d — независимые и равномерно распределенные соответственно на отрезках $[50, 0; 51, 0]$, $[40, 0; 41, 0]$, $[9, 5; 10, 0]$ случайные величины.

Найти вероятность того, что шарикоподшипник будет собран, если $\delta = 0, 5$.

8.34. Прямая разбита точками с шагом d . Пусть для определенности $nd, n = \dots, -2, -1, 0, 1, 2, \dots$, — координаты точек деления. На прямую наудачу бросают отрезок случайной длины ξ , где ξ — координата наудачу выбранной на отрезке $[0, d]$ точки.

Вычислить вероятность того, что отрезок “накроет” одну из точек деления.

8.35°. В единичный квадрат наудачу бросают точку. Какова вероятность того, что точка будет находиться от центра квадрата на расстоянии, меньшем $1/3$, если известно, что от каждой из сторон квадрата она удалена более чем на $1/6$?

8.36°. На плоскости проведены параллельные прямые, расстояние между которыми составляет $2a$. На плоскость наудачу бросают монету радиуса r ($r < a$). Какова

⁴ *Васильев Н. Б.* Геометрические вероятности // Квант. — 1991. — N1. — С.47 — 53.

вероятность того, что монета не пересечет ни одной из прямых?

8.37°. На отрезок длиной L наудачу бросают точку. Вычислить вероятность того, что она упадет не далее, чем на l ($l < L/2$) от середины отрезка.

8.38°. Случайная точка A равномерно распределена в квадрате со стороной 1. Найти вероятности таких событий:

а) расстояние от точки A до фиксированной стороны квадрата не превышает x ;

б) расстояние от точки A до ближайшей стороны квадрата не превышает x ;

в) расстояние от точки A до центра квадрата не превышает x ;

г) расстояние от точки A до фиксированной вершины квадрата не превышает x .

8.39°. Случайная точка A равномерно распределена в прямоугольнике со сторонами 1 и 2. Найти вероятности таких событий:

а) расстояние от точки A до ближайшей стороны прямоугольника не превышает x ;

б) расстояние от точки A до обеих диагоналей прямоугольника не превышает x .

8.40°. Случайная точка A равномерно распределена в квадрате со стороной a . Найти вероятность того, что расстояние от точки A до ближайшей стороны квадрата меньше расстояния от A до ближайшей диагонали квадрата.

8.41°. Случайная точка X равномерно распределена в квадрате $A = \{(x, y) : |x| \leq a, |y| \leq a\}$. Найти вероятность того, что квадрат с центром в точке X и сторонами длиной b , параллельными осям координат, полностью содержится в квадрате A .

Глава 9

Распределение случайной величины

9.1 Функция и плотность распределения

Случайная величина — функция исхода стохастического эксперимента (см. гл. 5) может принимать значения не только из конечного или счетного множества, ее значения могут “заполнять” промежутки. Например, время безотказной работы прибора, результат измерения, ... Формально случайную величину определяют следующим образом.

Определение. Пусть $\{\Omega, \mathfrak{F}, P\}$ — вероятностное пространство. *Случайной величиной* будем называть функцию $\xi = \xi(\omega)$ на Ω со значениями в $\mathbb{R}^1$, такую, что для каждого $x \in \mathbb{R}^1$

$$\{\omega : \xi(\omega) < x\} \in \mathfrak{F}.$$

Свойства случайных величин. Если ξ и η — случайные величины, то случайными величинами являются

$$\xi + \eta, \xi - \eta, \xi\eta, \xi/\eta$$

(последняя при условии, что $\eta \neq 0$).

Функция от случайной величины является случайной величиной. Подробнее, если ξ — случайная величина, g —

борелевская¹ функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$, то $g(\xi)$ — случайная величина.

Если $\xi_1, \xi_2, \dots, \xi_n \dots$ — случайные величины, то случайными величинами являются

$$\sup_n \xi_n, \inf_n \xi_n, \overline{\lim} \xi_n, \underline{\lim} \xi_n,$$

в частности, предел $\lim \xi_n$ (если он существует) является случайной величиной.

Функция распределения случайной величины.

Дискретная случайная величина ξ описывается своим распределением

$$P_\xi : x_i \rightarrow P_\xi(x_i) = P\{\xi = x_i\}.$$

В общей ситуации случайная величина ξ описывается так называемой функцией распределения, по которой всегда можно вычислить вероятность попадания ξ в любой заданный промежуток $[a, b)$.

Определение. Функцию

$$F_\xi(x) = P\{\omega : \xi(\omega) < x\}, \quad x \in \mathbb{R}^1,$$

называют *функцией распределения* случайной величины ξ .

Свойства функции распределения $F_\xi(x)$ случайной величины ξ .

1. $0 \leq F_\xi(x) \leq 1$.
2. $F_\xi(x)$ — неубывающая: если $x_1 < x_2$, то

$$F_\xi(x_1) \leq F_\xi(x_2).$$

3. $F_\xi(x)$ непрерывна слева.
4. $\lim_{x \rightarrow -\infty} F_\xi(x) = 0, \quad \lim_{x \rightarrow +\infty} F_\xi(x) = 1$.
5. Для произвольных a и b ($a < b$)

$$P\{\xi \in [a, b)\} = F_\xi(b) - F_\xi(a).$$

¹Функция $g : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ борелевская, если для произвольного борелевского множества B его прообраз $g^{-1}(B)$ — борелевское множество. Непрерывные функции являются борелевскими, но класс борелевских функций значительно шире.

6. Для произвольного x_0

$$P\{\xi = x_0\} = F_\xi(x_0 + 0) - F_\xi(x_0 - 0).$$

Функцию

$$P_\xi(B) = P\{\omega : \xi(\omega) \in B\}, \quad B \in \mathfrak{B}^1,$$

определенную на классе $\mathfrak{B}^1$ борелевских множеств прямой, будем называть *распределением* случайной величины ξ .

Плотность распределения случайной величины. Если функцию распределения $F(x)$ случайной величины ξ можно представить в виде

$$F(x) = \int_{-\infty}^x p(t)dt, \quad x \in \mathbb{R}^1, \quad (9.1.1)$$

то говорят, что случайная величина ξ имеет абсолютно непрерывное распределение (*абсолютно непрерывна*), а функцию $p(x)$ называют *плотностью* распределения случайной величины ξ .

По плотности распределения $p(x)$ случайной величины ξ всегда можно вычислить $P\{\xi \in [a, b]\}$:

$$P\{\xi \in [a, b]\} = \int_a^b p(x)dx. \quad (9.1.2)$$

Непосредственно из равенства (9.1.1) следует, что почти для всех x

$$\frac{d}{dx}F(x) = p(x),$$

плотность распределения $p(x)$ случайной величины неотрицательна и

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

Вычисление распределения $g(\zeta)$ по плотности распределения ζ . Пусть ζ — абсолютно непрерывная случайная величина со значениями в $\mathbb{R}^1$ и $p(x)$ — ее плотность распределения, $g(x)$ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$. Тогда

$$P\{g(\zeta) \in B\} = \int_{x:g(x) \in B} p(x)dx, \quad (9.1.3)$$

в частности

$$P\{g(\zeta) < t\} = \int_{x:g(x) < t} p(x)dx, \quad (9.1.4)$$

$$P\{\zeta \in B\} = \int_B p(x)dx. \quad (9.1.5)$$

9.2 Функция и плотность распределения случайного вектора

Определение. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины на вероятностном пространстве $\{\Omega, \mathfrak{F}, P\}$. Функцию $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ будем называть *случайным вектором*, или *случайной величиной со значениями в $\mathbb{R}^n$* .

Если $\xi_1, \xi_2, \dots, \xi_n$ — случайные величины, то

$$\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \dots, \xi_n(\omega) < x_n\} \in \mathfrak{F}$$

(для произвольных $x_1, x_2, \dots, x_n$).

Определение. Функцию

$$F(x_1, x_2, \dots, x_n) =$$

$$= P\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \dots, \xi_n(\omega) < x_n\},$$

определенную на $\mathbb{R}^n$, будем называть *функцией распределения* случайного вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$.

О п р е д е л е н и е. Функцию

$$P_\zeta : B \rightarrow P_\zeta(B) = P\{\omega : \zeta(\omega) \in B\} = \\ = P\{\omega : (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)) \in B\}, \quad B \in \mathfrak{B}^n,$$

определенную на классе $\mathfrak{B}^n$ борелевских множеств пространства $\mathbb{R}^n$, будем называть *распределением* случайного вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$.

О п р е д е л е н и е. Если функцию распределения $F(x_1, x_2, \dots, x_n)$ случайного вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ можно представить в виде

$$F(x_1, x_2, \dots, x_n) = \\ = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} p(t_1, t_2, \dots, t_n) dt_n dt_{n-1} \dots dt_1, \quad (9.2.1)$$

то будем говорить, что вектор $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ имеет абсолютно непрерывное распределение (вектор *абсолютно непрерывный*), а функцию $p(x_1, x_2, \dots, x_n)$ называют его плотностью распределения, или совместной плотностью распределения случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

Плотность распределения $p(x_1, x_2, \dots, x_n)$ случайного вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ неотрицательна и

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p(t_1, t_2, \dots, t_n) dt_n dt_{n-1} \dots dt_1 = 1.$$

Вычисление распределения $g(\zeta)$ по плотности распределения ζ . Пусть $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ — абсолютно непрерывная случайная величина (вектор) со значениями в $\mathbb{R}^n$ и $p(x_1, x_2, \dots, x_n)$ — ее плотность распределения, $g(x_1, x_2, \dots, x_n)$ — борелевская функция на $\mathbb{R}^n$ со значениями в $\mathbb{R}^l$, B — борелевское множество в $\mathbb{R}^l$.

Тогда

$$P\{g(\zeta) \in B\} = \int_{x: g(x) \in B} p(x) dx, \quad x \in \mathbb{R}^n, \quad (9.2.2)$$

в частности

$$P\{\zeta \in B\} = \int_B p(x) dx, \quad x \in \mathbb{R}^n. \quad (9.2.3)$$

Если распределение $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ дискретное, то распределение $g(\zeta)$ вычисляется по формулам (5.1.1) и (5.1.2).

Независимые случайные величины. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ со значениями в $\mathbb{R}^1$ будем называть *независимыми*, если для произвольных $x_1, x_2, \dots, x_n \in \mathbb{R}^1$

$$\begin{aligned} P\{\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n\} = \\ = P\{\xi_1 < x_1\} P\{\xi_2 < x_2\} \dots P\{\xi_n < x_n\}. \end{aligned}$$

Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые абсолютно непрерывные случайные величины соответственно с плотностями распределений $p_1(x_1), p_2(x_2), \dots, p_n(x_n)$, то существует совместная плотность $p(x_1, x_2, \dots, x_n)$ распределения этих случайных величин, и она равна произведению плотностей распределений $p_1(x_1), p_2(x_2), \dots, p_n(x_n)$:

$$p(x_1, x_2, \dots, x_n) = p_1(x_1) p_2(x_2) \dots p_n(x_n).$$

Пример 9.2.1. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины, каждая с плотностью распределения $p(x)$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Выписать совместное распределение случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

Решение. Поскольку случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы и абсолютно непрерывные, то существует совместная плотность $p(x_1, x_2, \dots, x_n)$ распределения, и она равна произведению плотностей распределений $p(x_1), p(x_2), \dots, p(x_n)$ случайных величин $\xi_1, \xi_2, \dots, \xi_n$:

$$\begin{aligned} p(x_1, x_2, \dots, x_n) &= \prod_{i=1}^n p(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - a)^2}{2\sigma^2}\right\} = \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2\right\}. \end{aligned}$$

Пример 9.2.2. Пусть ξ и η — независимые случайные величины соответственно с плотностями $p_\xi(s)$ и $p_\eta(t)$. Найти плотность распределения случайной величины $\zeta = \xi + \eta$.

Решение. Сначала найдем функцию распределения $F_\zeta(z)$ суммы $\zeta = \xi + \eta$. По определению

$$F_\zeta(z) = P\{\zeta < z\} = P\{\xi + \eta < z\}.$$

Вычислим $P\{\xi + \eta < z\}$. Поскольку ξ и η — независимые, то существует совместная плотность распределения $p(s, t)$ случайных величин ξ и η , при этом $p(s, t) = p_\xi(s)p_\eta(t)$. Далее воспользуемся (9.2.2), рассмотрев в качестве x пару $(s, t) \in \mathbb{R}^2$, в качестве $g(x)$ — функцию $g(s, t) = s + t$, а в качестве множества B — промежуток $(-\infty, x)$. Имеем

$$\begin{aligned} F_\zeta(z) &= P\{\xi + \eta < z\} = \\ &= \int \int_{(s,t): s+t < z} p(s, t) ds dt = \int \int_{(s,t): s+t < z} p_\xi(s) p_\eta(t) ds dt = \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-t} p_\xi(s) ds \right) p_\eta(t) dt = \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^z p_{\xi}(u-t) du \right) p_{\eta}(t) dt = \\
&= \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} p_{\xi}(u-t) p_{\eta}(t) dt \right) du.
\end{aligned}$$

Таким образом,

$$F_{\zeta}(z) = \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} p_{\xi}(u-t) p_{\eta}(t) dt \right) du.$$

Поэтому по определению (см. (9.1.1)) функция

$$p(u) = \int_{-\infty}^{+\infty} p_{\xi}(u-t) p_{\eta}(t) dt \quad (9.2.4)$$

является плотностью распределения случайной величины $\zeta = \xi + \eta$.

Плотность $p(u)$, определенную равенством (9.2.4), называют *сверткой* плотностей $p_{\xi}(t)$ и $p_{\eta}(t)$.

Пример 9.2.3. Пусть ξ_i , $i = 1, 2, \dots, n$, — независимые случайные величины, каждая с функцией распределения $F(x)$. Найти функцию распределения случайной величины $\min\{\xi_1, \xi_2, \dots, \xi_n\}$.

Решение. Обозначим $\eta = \min\{\xi_1, \xi_2, \dots, \xi_n\}$. По определению функции распределения случайной величины

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

Отсюда

$$\begin{aligned}
1 - F_{\eta}(x) &= 1 - P\{\eta < x\} = P\{\eta \geq x\} = \\
&= P\{\min\{\xi_1, \xi_2, \dots, \xi_n\} \geq x\} =
\end{aligned}$$

$$\begin{aligned}
&= P\{\xi_1 \geq x, \xi_2 \geq x, \dots, \xi_n \geq x\} = \prod_{i=1}^n P\{\xi_i \geq x\} = \\
&= \prod_{i=1}^n (1 - F(x)) = (1 - F(x))^n
\end{aligned}$$

(мы воспользовались независимостью случайных величин $\xi_1, \xi_2, \dots, \xi_n$). Следовательно,

$$F_\eta(x) = 1 - (1 - F(x))^n.$$

9.3 Абсолютно непрерывные распределения на $\mathbb{R}^1$

Нормальное распределение. Случайная величина ξ имеет *нормальное распределение с параметрами* $(a; \sigma^2)$ (*гауссовское распределение, распределение* $N_{a; \sigma^2}$), если ее плотность распределения

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Равномерное распределение. Случайная величина ξ имеет *равномерное распределение на отрезке* $[a; b]$, если ее плотность распределения

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

Гамма-распределение. Случайная величина ξ имеет *гамма-распределение* с параметрами $(\nu; \theta)$, $\theta > 0, \nu > 0$, если ее плотность распределения

$$p(x) = \begin{cases} \frac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Показательное распределение. Случайная величина ξ имеет *показательное распределение* с параметром θ , если ее плотность распределения

$$p(x) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Показательное распределение является гамма-распределением с параметрами $(1; \theta)$.

Распределение Эрланга. Случайная величина ξ имеет *распределение Эрланга* с параметрами $(m; \theta)$, если ее плотность распределения

$$p(x) = \begin{cases} \frac{\theta^m}{(m-1)!} x^{m-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Распределение Эрланга является гамма-распределением с параметрами $(m; \theta)$, $m = 1, 2, \dots$

Распределение χ^2 . Случайная величина ξ имеет *распределение χ^2* с n степенями свободы, если ее плотность распределения

$$p(x) = \begin{cases} \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} \exp\left\{-\frac{1}{2}x\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Распределение χ^2 с n степенями свободы является гамма-распределением с параметрами $(n/2; 1/2)$.

Распределение Коши. Случайная величина ξ имеет *распределение Коши* с параметром a , если ее плотность распределения

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

Логарифмически нормальное распределение. Случайная величина ξ имеет *логарифмически нормальное распределение* с параметрами $(\mu; \sigma^2)$, если ее плот-

ность распределения

$$p(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Распределение Парето. Случайная величина ξ имеет *распределение Парето* с параметрами $(\lambda; \theta)$, если ее плотность распределения

$$p(x) = \begin{cases} \frac{\theta\lambda^\theta}{x^{\theta+1}}, & \text{если } x > \lambda; \\ 0, & \text{если } x \leq \lambda, \end{cases}$$

$\lambda > 0, \theta > 2$.

9.4 Задачи

АЗ: 9.2°, 9.4°(5), 9.5, 9.6, 9.12*, 9.14, 9.19, 9.23, 9.28.

СЗ: 9.1°, 9.4°(1–4, 6), 9.7, 9.10, 9.16, 9.24, 9.25, 9.29, 9.38, 9.42.

9.1°. Пусть ξ — случайная величина с плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \notin [-1; 1]; \\ 1 - |x|, & \text{если } x \in [-1; 1]. \end{cases}$$

Вычислить $P\{\xi^2 > 1/4\}$.

9.2°. Случайная величина ξ равномерно распределена на отрезке $[-1; 3]$. Вычислить $P\{|\xi| \geq 1/2\}$.

9.3°. Распределение случайной величины ξ абсолютно непрерывно с плотностью

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

Вычислить: 1) $P\{-\sqrt{3} \leq \xi \leq 1\}$; 2) $P\{|\xi| \geq \sqrt{3}\}$.

9.4°. На отрезок $[0; 1]$ наудачу бросают пару точек. Пусть ξ — координата одной точки, η — другой. Найти функции распределения случайных величин:

- 1) $\zeta = \max\{\xi; \eta\}$; 4) $\zeta = \xi + \eta$;
2) $\zeta = \min\{\xi; \eta\}$; 5) $\zeta = \max\{\xi^2; \eta\}$;
3) $\zeta = \xi\eta$; 6) $\zeta = |\eta - \xi|$.

9.5°. На отрезок $[0; l]$ наудачу бросают точку, ξ — ее координата. Найти функцию и плотность распределения случайной величины ξ .

9.6. Случайная величина ξ равномерно распределена на отрезке $[-1; 3]$. Найти функцию и плотность распределения случайной величины $\eta = \xi^2$.

9.7. Случайная величина ξ равномерно распределена на отрезке $[-2; 2]$. Найти плотность распределения случайной величины $\eta = |\xi|$.

9.8. Случайная величина ξ равномерно распределена на отрезке $[0; 1]$. Найти плотность распределения случайной величины $\eta = 1/\xi$ (если $\xi = 0$, то по определению $\eta = 0$).

9.9. Случайная величина ξ равномерно распределена на отрезке $[-2; 1]$. Найти плотность распределения случайной величины $\eta = 1/\xi^2$.

9.10. Случайная величина ξ равномерно распределена на отрезке $[0; 2]$. Найти плотность распределения случайной величины $\eta = |\xi - 1|$.

9.11. Случайная величина ξ равномерно распределена на отрезке $[a, b]$. Найти плотность распределения случайной величины $\eta = e^\xi$.

9.12*. Случайная величина ξ распределена показательно с параметром λ . Найти плотность распределения случайной величины $\eta = 1/(1 - \xi)$.

9.13. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функцию распределения случайной величины $\eta = -\xi$.

9.14. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функцию распределения случайной величины $\eta = \text{sign } \xi$.

9.15. Случайная величина ξ распределена показательно с параметром 1. Найти функцию распределения случайной величины $\eta = 1 - e^{-\xi}$.

9.16. Пусть $p(x)$ — плотность распределения случайной величины ξ . Найти плотности распределений случайных величин: 1) $\eta = |\xi|$, 2) $\eta = a\xi$, $a \neq 0$.

9.17. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функцию распределения случайной величины $\eta = \xi^2$.

9.18. Случайная величина ξ распределена показатель-но с параметром λ . Найти плотности распределений случайных величин: 1) $\eta = |\xi - 1|$; 2) $\eta = (\xi - 1)^3$.

9.19. Функция распределения $F(x)$ случайной величины ξ строго монотонна и непрерывна. Найти функцию распределения случайной величины $\eta = F(\xi)$.

9.20. Пусть $p(x)$ — плотность распределения случайной величины ξ . Найти плотности распределений случайных величин: 1) $\eta = -2\xi + 1$; 2) $\eta = \xi^2$.

9.21. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функции распределений случайных величин: 1) $\eta = e^\xi$; 2) $\eta = |\xi|$.

9.22. На отрезок $[0; 1]$ наудачу бросают пару точек. Пусть ξ — координата одной, η — другой. Для $0 \leq x \leq 1$ найти: 1) $P\{|\eta - \xi| < x\}$; 2) $P\{\xi\eta < x\}$.

9.23. Пусть $\xi_i, i = 1, 2, \dots, n$, — независимые случайные величины, каждая с функцией распределения $F(x)$. Найти функцию распределения случайной величины $\max\{\xi_1, \xi_2, \dots, \xi_n\}$.

9.24. Пусть $\xi_i, i = 1, 2, \dots, n$, — независимые случайные величины соответственно с функциями распределения $F_i(x)$, $i = 1, 2, \dots, n$. Найти функции распределений случайных величин:

$$1) \max\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 2) \min\{\xi_1, \xi_2, \dots, \xi_n\}.$$

9.25. Пусть $\xi_i, i = 1, 2, \dots, n$, — независимые случайные величины соответственно с плотностями распределения $p_i(x)$, $i = 1, 2, \dots, n$. Найти плотности распределений случайных величин:

$$1) \max\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 2) \min\{\xi_1, \xi_2, \dots, \xi_n\}.$$

9.26. Пусть $\xi_i, i = 1, 2, \dots, n$, — независимые случайные величины, каждая с плотностью распределения $p(x)$. Найти плотности распределений случайных величин:

$$1) \max\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 2) \min\{\xi_1, \xi_2, \dots, \xi_n\}.$$

9.27. Случайная величина имеет своей плотностью распределения функцию $p(x) = ae^{-\lambda|x|}$, $\lambda > 0$. Найти: 1) коэффициент a ; 2) функцию распределения случайной величины.

9.28. На отрезок $[0; 1]$ наудачу бросают точку, которая делит его на две части. Найти функцию распределения длины меньшей части.

9.29. На отрезок $[0; 1]$ наудачу бросают точку, которая делит его на две части. Найти функцию распределения длины большей части.

9.30. На отрезок $[0; T]$ наудачу бросают две точки. Найти функцию и плотность распределения расстояния между ними.

9.31. Пусть $\xi_i, i = 1, 2, \dots, n$, — независимые случайные величины, каждая с плотностью распределения $p(x)$:

$$1) p(x) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]; \end{cases}$$

$$2) p(x) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0; \end{cases}$$

$$3) p(x) = \frac{1}{2a} \exp\left\{-\frac{1}{a}|x-b|\right\};$$

$$4) p(x) = \begin{cases} \frac{1}{a} \exp\left\{-\frac{1}{a}(x-b)\right\}, & \text{если } x > b; \\ 0, & \text{если } x \leq b. \end{cases}$$

Выписать совместную плотность распределения случайных величин $\xi_1, \xi_2, \dots, \xi_n$.

9.32. Пусть случайная величина ξ распределена нормально с параметрами $(0; 1)$. Найти функцию распределения $\eta = 1/\xi$.

9.33. Пусть случайная величина ξ распределена нормально с параметрами $(0; 1)$. Найти функцию распределения случайной величины $\eta = 1/\xi^2$.

9.34. Случайная величина ξ называется симметрично распределенной, если распределения случайных величин ξ и $-\xi$ совпадают.

Доказать, что $N_{0;\sigma^2}$ -распределенная случайная величина является симметрично распределенной.

Сформулировать условие симметричной распределен-

ности случайной величины в терминах: а) функции распределения; б) плотности распределения.

9.35. Случайная величина ξ равномерно распределена на отрезке $[0; 1]$. Найти распределения случайных величин: 1) $\eta = 1 - \xi$; 2) $\eta = \ln \xi$.

9.36. Случайная величина η распределена $N_{a;\sigma^2}$. Доказать, что случайная величина $\xi = (\eta - a)/\sigma$ распределена $N_{0;1}$.

9.37. Случайная величина ξ распределена $N_{0;1}$. Найти распределение случайной величины $\eta = a + \sigma\xi$ ($\sigma > 0$).

9.38. Случайная величина η распределена $N_{0;1}$. Найти распределение случайной величины $\eta^+ = \max\{0, \eta\}$.

9.39. Случайная величина η распределена $N_{0;\sigma^2}$. Найти распределение случайной величины $\eta^+ = \max\{0, \eta\}$.

9.40. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функцию распределения $F_\eta(x)$ случайной величины $\eta = (\xi - a)^+ = \max\{0, \xi - a\}$ (a — константа).

По графику функции $F(x)$ построить график функции $F_\eta(x)$.

9.41. Случайная величина ξ абсолютно непрерывна с плотностью $p(x)$. Найти функцию распределения случайной величины $\eta = (\xi - a)^+ = \max\{0, \xi - a\}$ (a — константа).

Является ли η абсолютно непрерывной случайной величиной?

9.42. Пусть $F(x)$ — функция распределения случайной величины ξ . Найти функцию распределения случайной величины $\eta = \min\{\xi, L\}$ (L — константа).

9.43. Случайная величина ξ абсолютно непрерывна с плотностью $p(x)$. Найти функцию распределения случайной величины $\eta = \min\{\xi, L\}$ (L — константа).

Является ли η абсолютно непрерывной случайной величиной?

9.44°. По плотности распределения случайной величины найти ее функцию распределения, если случайная величина:

- 1) нормально распределена с параметрами $(a; \sigma^2)$;
- 2) равномерно распределена на промежутке $[a; b]$;
- 3) имеет гамма-распределение;

- 4) имеет распределение Коши;
- 5) имеет распределение Эрланга;
- 6) имеет распределение Парето.

9.45°. Построить графики плотностей распределений и функций распределений случайных величин, перечисленных в задаче 9.44°.

9.46°. Пусть ξ распределена $N_{a;\sigma^2}$. Вычислить:

- 1) $P\{a - \sigma \leq \xi \leq a + \sigma\}$; 2) $P\{a - 2\sigma \leq \xi \leq a + 2\sigma\}$;
- 3) $P\{a - 3\sigma \leq \xi \leq a + 3\sigma\}$; 4) $P\{a - 4\sigma \leq \xi \leq a + 4\sigma\}$.

Примечание. Воспользоваться таблицей нормального распределения (см. табл. 22.1.1).

9.47. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет показательное распределение с параметром λ .

Доказать, что случайная величина

$$\eta = \min\{\xi_1, \xi_2, \dots, \xi_n\}$$

имеет показательное распределение с параметром $n\lambda$.

9.48. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет показательное распределение с параметром λ .

Найти распределение случайной величины

$$\zeta = \max\{\xi_1, \xi_2, \dots, \xi_n\}.$$

9.49. Система состоит из n блоков. Пусть ξ_i — время безотказной работы i -го блока. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ — независимы и каждая из них имеет показательное распределение с параметром λ . Система выйдет из строя, если выйдет из строя хотя бы один блок. Найти функцию распределения времени безотказной работы системы.

9.50. Случайные величины $\xi_i = \xi_i(\omega)$, $i = 1, 2, \dots$, заданы на вероятностном пространстве $\{\Omega, \mathfrak{F}, P\}$, где $\Omega = [0; 1]$, $\mathfrak{F} = \mathfrak{B}_{[0;1]}$, $P = L$, равенствами:

$$1) \xi_1 = \xi_1(\omega) = \begin{cases} \omega, & \text{если } \omega \in [0; 1/3); \\ \omega + 1/3, & \text{если } \omega \in [1/3; 2/3); \\ \omega - 1/3, & \text{если } \omega \in [2/3; 1); \end{cases}$$

$$\begin{aligned}
 2) \quad \xi_2 = \xi_2(\omega) &= \begin{cases} \omega + 2/3, & \text{если } \omega \in [0; 1/3]; \\ \omega, & \text{если } \omega \in [1/3; 2/3]; \\ \omega - 2/3, & \text{если } \omega \in [2/3; 1]; \end{cases} \\
 3) \quad \xi_3 = \xi_3(\omega) &= \begin{cases} \omega, & \text{если } \omega \in [0; 1/4]; \\ 1/4, & \text{если } \omega \in [1/4; 2/4]; \\ \omega - 1/4, & \text{если } \omega \in [2/4; 3/4]; \\ 1/2, & \text{если } \omega \in [3/4; 1]. \end{cases}
 \end{aligned}$$

Найти их функции распределений.

9.51. Случайные величины

$$\xi_i = \xi_i(\omega), \quad i = 1, 2, \quad \eta_j = \eta_j(\omega), \quad j = 1, 2,$$

заданы на одном и том же вероятностном пространстве $\{\Omega, \mathfrak{F}, P\}$.

Распределения ξ_1 и ξ_2 совпадают, совпадают и распределения η_1 и η_2 . Будут ли совпадать распределения случайных величин: 1) $\xi_1\eta_1$ и $\xi_2\eta_2$; 2) $\xi_1 + \eta_1$ и $\xi_2 + \eta_2$?

9.52. Привести пример случайных величин $\xi = \xi(\omega)$ и $\eta = \eta(\omega)$, заданных на одном и том же вероятностном пространстве $\{\Omega, \mathfrak{F}, P\}$, распределения которых совпадают, но

$$P\{\omega : \xi(\omega) \neq \eta(\omega)\} = 1.$$

9.53. Пусть $\zeta = (\xi, \eta)$ — абсолютно непрерывная случайная величина со значениями в $\mathbb{R}^2$ и $f_\zeta(x, y)$ — ее плотность распределения.

1) Доказать, что ξ и η — абсолютно непрерывные случайные величины соответственно с плотностями

$$f_\xi(u) = \int_{\mathbb{R}^1} f_\zeta(u, v) dv, \quad f_\eta(v) = \int_{\mathbb{R}^1} f_\zeta(u, v) du.$$

2) Доказать, что при каждом фиксированном y функция

$$f_{\xi|\eta}(x|y) = \begin{cases} f_\zeta(x, y)/f_\eta(y), & \text{если } f_\eta(y) > 0; \\ 0, & \text{если } f_\eta(y) = 0; \end{cases}$$

является плотностью распределения вероятностей на $\mathbb{R}^1$; при каждом фиксированном x функция

$$f_{\eta|\xi}(y|x) = \begin{cases} f_\zeta(x, y)/f_\xi(x), & \text{если } f_\xi(x) > 0; \\ 0, & \text{если } f_\xi(x) = 0; \end{cases}$$

является плотностью распределения вероятностей на $\mathbb{R}^1$.

9.54. Пусть $f(t)$ — плотность распределения случайной величины ξ .

Является ли функция $f(t - a)$ (a — константа) плотностью распределения некоторой случайной величины и если да, то какой.

9.55*. Пусть ξ и η — независимые случайные величины соответственно с функциями распределения $F(x)$ и $G(x)$. Доказать, что функция распределения $Q(x)$ суммы $\xi + \eta$ есть

$$Q(x) = P\{\xi + \eta < x\} = \int_{-\infty}^{+\infty} F(x - y)G(dy) = \int_{-\infty}^{+\infty} G(x - y)F(dy).$$

Глава 10

Математическое ожидаие

10.1 Определения, свойства, вычисление

Распределения абсолютно непрерывных случайных величин часто можно описать несколькими числовыми характеристиками. Важнейшими из них являются математическое ожидание и дисперсия (см. также п. 6.1 в гл. 6).

Пусть $\xi = \xi(\omega)$ — случайная величина на вероятностном пространстве $\{\Omega, \mathfrak{F}, P\}$ со значениями в $\mathbb{R}^1$.

О п р е д е л е н и е. *Математическим ожиданием* $M\xi$ неотрицательной случайной величины ξ будем называть

$$\begin{aligned} M\xi &= \int_{\Omega} \xi(\omega) P(d\omega) = \\ &= \lim_{n \rightarrow \infty} \left(\sum_{j=1}^{2^n} \frac{j-1}{2^n} P \left\{ \omega : \frac{j-1}{2^n} \leq \xi(\omega) < \frac{j}{2^n} \right\} + \right. \\ &\quad \left. + n P \{ \omega : \xi(\omega) \geq n \} \right). \end{aligned}$$

Математическое ожидание $M\xi$ случайной величины ξ , принимающей значения обоих знаков, определяется равенством

$$M\xi = M\xi^+ - M\xi^-,$$

если только $M\xi^+$ и $M\xi^-$ одновременно не равны $+\infty$, где $\xi^+ = \max\{0, \xi\}$, $\xi^- = -\min\{0, \xi\}$.

Свойства математического ожидания:

1. Математическое ожидание константы равно этой же константе:

$$Mc = c \quad (c - \text{константа}).$$

2. Математическое ожидание суммы случайных величин равно сумме математических ожиданий этих случайных величин:

$$M(\xi + \eta) = M\xi + M\eta.$$

3. Константа выносится из-под знака математического ожидания:

$$Ma\xi = aM\xi.$$

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

$$M\xi\eta = M\xi M\eta.$$

Вычисление математического ожидания случайной величины. Математическое ожидание случайной величины (и функции от нее), если только оно существует, всегда можно вычислить по распределению случайной величины.

Теорема 10.1.1. Пусть $\xi = \xi(\omega)$ — случайная величина со значениями в $\mathbb{R}^1$, g — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$.

Если случайная величина ξ абсолютно непрерывна и $p(x)$ — ее плотность распределения, то при условии существования интеграла $\int_{-\infty}^{+\infty} g(x)p(x)dx$

$$Mg(\xi) = \int_{-\infty}^{+\infty} g(x)p(x)dx, \quad (10.1.1)$$

в частности, при условии существования интеграла

$$\int_{-\infty}^{+\infty} xp(x)dx$$

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx. \quad (10.1.2)$$

Если случайная величина ξ дискретна и

$$P_\xi : x_i \rightarrow P_\xi(x_i), x_i \in X,$$

ее распределение, то при условии абсолютной сходимости ряда $\sum_{x_i} g(x_i)P_\xi(x_i)$

$$Mg(\xi) = \sum_{x_i} g(x_i)P_\xi(x_i),$$

в частности, при условии абсолютной сходимости ряда $\sum_{x_i} x_i P_\xi(x_i)$

$$M\xi = \sum_{x_i} x_i P_\xi(x_i).$$

Дисперсия. Дисперсией $D\xi$ случайной величины ξ будем называть $M(\xi - M\xi)^2$ (если $M(\xi - M\xi)^2 < \infty$), т. е.

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

Свойства дисперсии:

1. Дисперсия константы равна нулю:

$$Dc = 0 \quad (c - \text{константа}).$$

2. Константа выносится из-под знака дисперсии с квадратом:

$$Da\xi = a^2 D\xi.$$

3. Дисперсия суммы независимых случайных величин равна сумме их дисперсий:

$$D(\xi + \eta) = D\xi + D\eta.$$

Пример 10.1.1. На отрезок $[0; 1]$ наудачу бросают точку. Она делит отрезок на две части. Найти распределение и вычислить математическое ожидание длины большей части отрезка, радиус которой равен длине большей части отрезка.

Решение. Пусть ξ — координата наудачу брошенной на отрезок $[0; 1]$ точки, тогда $\eta = \max\{\xi, 1 - \xi\}$ — длина большей части отрезка, а $\zeta = 2\pi\eta$ — длина окружности, радиус которой равен η .

Сначала найдем функцию распределения случайной величины $\eta = \max\{\xi, 1 - \xi\}$. Очевидно, при $x < 1/2$

$$P\{\eta < x\} = 0,$$

при $x > 1$

$$P\{\eta < x\} = 1.$$

А при $1/2 < x \leq 1$ имеем

$$\begin{aligned} P\{\eta < x\} &= P\{\max\{\xi, 1 - \xi\} < x\} = P\{\xi < x, 1 - \xi < x\} = \\ &= P\{1 - x < \xi < x\} = (x - (1 - x))/(1 - 0) = 2x - 1 \end{aligned}$$

($P\{1 - x < \xi < x\}$ вычислена как геометрическая вероятность, поскольку точку бросают на отрезок $[0; 1]$ наудачу). Таким образом, функцией распределения η является

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 1/2; \\ 2x - 1, & \text{если } 1/2 < x \leq 1; \\ 1, & \text{если } x > 1 \end{cases}$$

(т. е. η распределена равномерно на отрезке $[1/2; 1]$). Отсюда получаем функцию распределения ζ :

$$\begin{aligned} F_{\zeta}(x) &= P\{\zeta < x\} = P\{2\pi\eta < x\} = F_{\eta}\left(\frac{x}{2\pi}\right) = \\ &= \begin{cases} 0, & \text{если } x/(2\pi) \leq 1/2; \\ 2\left(\frac{x}{2\pi}\right) - 1, & \text{если } 1/2 < x/(2\pi) \leq 1; \\ 1, & \text{если } x/(2\pi) > 1. \end{cases} \end{aligned}$$

Или

$$F_{\zeta}(x) = \begin{cases} 0, & \text{если } x \leq \pi; \\ \frac{x}{\pi} - 1, & \text{если } \pi < x \leq 2\pi; \\ 1, & \text{если } x > 2\pi. \end{cases}$$

Далее, поскольку η равномерно распределена на отрезке $[1/2; 1]$, то

$$M\zeta = M2\pi\eta = 2\pi M\eta = 3\pi/2.$$

Заметим, что $M\zeta$ можно вычислить как математическое ожидание функции $\zeta = 2\pi \max\{\xi, 1 - \xi\}$ случайной величины ξ по ее распределению (см. формулу (10.1.1)). Распределение ξ равномерное на отрезке $[0; 1]$, поэтому

$$\begin{aligned} M\zeta &= M2\pi \max\{\xi, 1 - \xi\} = 2\pi \int_{-\infty}^{+\infty} \max\{x, 1 - x\} p_{\xi}(x) dx = \\ &= 2\pi \int_0^1 \max\{x, 1 - x\} dx = 3\pi/2. \end{aligned}$$

Пример 10.1.2. Пусть случайная величина ξ распределена показательным с параметром λ .

Найти распределение случайной величины $\eta = [\xi]$, вычислить $M\eta$ ($[x]$ — целая часть x).

Решение. Случайная величина $\eta = [\xi]$ принимает значения $0, 1, 2, \dots$ (является дискретной). Найдем ее распределение:

$$\begin{aligned} P_{\eta}(k) &= P\{\eta = k\} = P\{[\xi] = k\} = P\{k \leq \xi < k + 1\} = \\ &= \int_k^{k+1} \lambda e^{-\lambda x} dx = e^{-\lambda k} (1 - e^{-\lambda}) = p(1 - p)^k, \end{aligned}$$

где $p = 1 - e^{-\lambda}$. Следовательно, $\eta = [\xi]$ имеет геометрическое распределение с параметром $p = 1 - e^{-\lambda}$.

По известному распределению случайной величины η можем вычислить ее математическое ожидание (см. формулу (6.1.2) и задачу 6.18):

$$M\eta = \sum_{k=0}^{\infty} kP_{\eta}(k) = \sum_{k=0}^{\infty} k(1-p)^k p = \frac{1-p}{p} = \frac{e^{-\lambda}}{1-e^{-\lambda}}.$$

Пример 10.1.3. *Вычислить математическое ожидание и дисперсию случайной величины, распределенной нормально с параметрами $(a; \sigma^2)$.*

Решение. Плотность нормально распределенной с параметрами $(a; \sigma^2)$ случайной величины ξ

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}.$$

Вычислим $M(\xi - a)$, воспользовавшись формулой (10.1.1) (при этом выполним замену $(x - a)/\sigma = t$):

$$\begin{aligned} M(\xi - a) &= \int_{-\infty}^{+\infty} (x - a)p(x)dx = \\ &= \int_{-\infty}^{+\infty} (x - a) \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - a)^2}{2\sigma^2} \right\} dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{x - a}{\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x - a}{\sigma} \right)^2 \right\} dx = \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t \exp \left\{ -\frac{t^2}{2} \right\} dt = \\ &= \frac{\sigma}{\sqrt{2\pi}} \lim_n \int_{[-n, n]} t \exp \left\{ -\frac{t^2}{2} \right\} dt = 0 \end{aligned}$$

(последний интеграл равен нулю как интеграл от нечетной функции по симметричному промежутку). Таким образом, $M(\xi - a) = 0$, а следовательно,

$$M\xi = a.$$

Аналогично

$$\begin{aligned} D\xi &= M(\xi - M\xi)^2 = M(\xi - a)^2 = \\ &= \int_{-\infty}^{+\infty} (x - a)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - a)^2}{2\sigma^2}\right\} dx. \end{aligned}$$

После замены переменной $(x - a)/\sigma = t$ имеем

$$\begin{aligned} \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 \exp\left\{-\frac{t^2}{2}\right\} dt &= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t d \exp\left\{-\frac{t^2}{2}\right\} = \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{t^2}{2}\right\} dt = \sigma^2 \end{aligned}$$

(последний интеграл как интеграл Эйлера—Пуассона равен $\sqrt{2\pi}$).

Пример 10.1.4. *Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью*

$$p(x) = \begin{cases} \frac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($p(x)$ — плотность гамма-распределения с параметрами $(\nu; \theta)$).

Решение. По известной плотности $p(x)$ случайной величины ξ согласно формуле (10.1.2) получим

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx = \int_0^{+\infty} x \frac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\theta x\} dx =$$

$$\begin{aligned}
&= \frac{\Gamma(\nu+1)}{\theta\Gamma(\nu)} \int_0^{+\infty} \frac{\theta^{\nu+1}}{\Gamma(\nu+1)} x^\nu \exp\{-\theta x\} dx = \\
&= \frac{\Gamma(\nu+1)}{\theta\Gamma(\nu)} \cdot 1 = \frac{\nu\Gamma(\nu)}{\theta\Gamma(\nu)} = \frac{\nu}{\theta}.
\end{aligned}$$

Мы воспользовались тем, что интеграл

$$\int_0^{+\infty} \frac{\theta^{\nu+1}}{\Gamma(\nu+1)} x^\nu \exp\{-\theta x\} dx$$

равен единице как интеграл от плотности гамма-распределения с параметрами $(\nu+1, \theta)$.

Аналогично

$$M\xi^2 = \frac{\nu(\nu+1)}{\theta^2}.$$

Отсюда

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{\nu}{\theta^2}.$$

Пример 10.1.5. Пусть ξ — абсолютно непрерывная случайная величина с конечным математическим ожиданием $M\xi$ и плотностью $f(x)$, $x \in \mathbb{R}^1$, график которой симметричен относительно прямой $x = a$. Найти $M\xi$.

Решение. Поскольку график функции $f(t)$ симметричен относительно прямой $x = a$, то $f(a+t) = f(a-t)$ (там, где $f(a+t)$ и $f(a-t)$ определены). Последнее равенство означает, что функция $f(a+t)$ является четной.

Далее,

$$M\xi = \int_{-\infty}^{+\infty} xf(x)dx.$$

В последнем интеграле сделаем замену $x = t + a$. Имеем

$$\int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} (t+a)f(t+a)dt =$$

$$= a \int_{-\infty}^{+\infty} f(t+a)dt + \int_{-\infty}^{+\infty} tf(t+a)dt = a \cdot 1 + 0 = a.$$

Интеграл $\int_{-\infty}^{+\infty} tf(t+a)dt$ равен нулю как интеграл от нечетной функции по симметричному промежутку.

10.2 Задачи

АЗ: 10.1°(2), 10.2°(1), 10.6°(1), 10.7, 10.12, 10.13, 10.16(2), 10.19(1а), 10.19(3б), 10.21, 10.20(5).

СЗ: 10.1°(1), 10.2°(2), 10.6°(2, 3), 10.8, 10.10(1), 10.16(1), 10.14, 10.17(2), 10.19(1б), 10.19(3а), 10.20(1, 2), 10.22(2, 3), 10.26.

10.1°. Пусть ξ — случайная величина, равномерно распределенная на промежутке: 1) $[-a; a]$; 2) $[a; b]$. Вычислить $M\xi$ и $D\xi$.

10.2°. Случайная величина ξ распределена равномерно на промежутке $[0; 1]$. Вычислить математическое ожидание случайной величины η :

1) $\eta = \ln(1/\xi)$; 2) $\eta = \sin^2 \pi \xi$; 3) $\eta = e^\xi$.

10.3°. Случайная величина ξ распределена равномерно на промежутке $[a; b]$. Вычислить:

- 1) $M\xi^2$, если $a = 0$, $b = 3$;
- 2) $M\xi e^{-\xi}$, если $a = 0$, $b = 1$;
- 3) $M(\xi - 1)^2$, если $a = 1$, $b = 4$;
- 4) $M\xi e^{|\xi|}$, если $a = -1$, $b = 1$;
- 5) $Me^{2\xi}$, если $a = 0$, $b = 1/2$.

10.4. Пусть ξ — случайная величина с плотностью распределения

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

Вычислить 1) $M \min\{|\xi|, 1\}$, 2) $M \min\{|\xi|, \sqrt{3}\}$.

10.5. Пусть ξ — нормально распределенная случайная величина с параметрами $(0; \sigma^2)$. Вычислить Me^ξ .

10.6°. Пусть ξ — случайная величина с плотностью распределения

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

Вычислить математическое ожидание случайной величины η :

$$\begin{aligned} 1) \eta &= (\xi^2 + 1)I_{[0;\sqrt{3}]}(\xi); & 2) \eta &= \xi^2 I_{[1;\sqrt{3}]}(\xi); \\ 3) \eta &= I_{[1/3;3]}(\xi^2); & 4) \eta &= I_{[-1;1]}(\xi) \end{aligned}$$

($I_A(x)$ — индикатор множества A — функция, принимающая на A значение 1, а на $\bar{A}$ — значение 0).

10.7. Пусть ξ имеет показательное распределение с параметром λ . Вычислить:

$$1) M\xi; \quad 2) D\xi; \quad 3) P\{\xi > 1\}; \quad 4) M\xi^k.$$

10.8. Длительность работы электронной лампы (единица измерения времени — сутки) — случайная величина, имеющая показательное распределение с параметром $\lambda = 0,003$. Через год лампу меняют, даже если она не вышла из строя. Найти математическое ожидание времени работы лампы.

10.9. Найти математическое ожидание и дисперсию произведения $\zeta = \xi\eta$ независимых случайных величин ξ и η с равномерными распределениями: ξ — на отрезке $[0; 1]$, η — на отрезке $[1; 3]$.

10.10. Пусть ξ — случайная величина с плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \notin [a, a+2); \\ x-a, & \text{если } x \in [a, a+1); \\ -x+a+2, & \text{если } x \in [a+1, a+2). \end{cases}$$

Вычислить: 1) $M\xi$; 2) $M\xi^2$.

10.11. Пусть ξ — случайная величина с плотностью распределения

$$p(x) = \frac{1}{2}e^{-|x|}.$$

Вычислить $MI_{[0;4]}(\xi^2)$.

10.12. Плотность распределения случайной величины ξ

$$p(x) = \frac{\lambda}{2}e^{-\lambda|x|}, \quad \lambda > 0$$

(двустороннее показательное распределение с параметром λ). Вычислить $M\xi$ и $D\xi$.

10.13. Из точки A , равномерно распределенной на окружности радиуса R с центром в начале координат, проведена касательная к окружности. Найти функцию распределения и плотность распределения длины ξ отрезка касательной между точкой A и точкой пересечения этой касательной с осью Ox . Что можно сказать о существовании $M\xi$?

10.14. Точка P равномерно распределена в круге радиуса R . Пусть η — расстояние от точки P до центра круга. Найти функцию распределения $F(x)$ и плотность распределения $p(x)$ случайной величины η . Построить графики функций $F(x)$ и $p(x)$. Вычислить $M\eta$ и $D\eta$.

10.15. Точка A равномерно распределена на окружности единичного радиуса с центром в начале координат. Пусть ξ — проекция точки A на ось Ox . Найти: 1) функцию распределения $|\xi|$; 2) плотность распределения $|\xi|$; 3) $M|\xi|$; 4) $P\{|\xi| > 1/2\}$.

10.16. Длина стороны квадрата ξ — случайная величина,

- 1) распределенная равномерно на промежутке $[a; b]$, $a > 0$, $a < b$;
- 2) имеющая гамма-распределение с параметрами $(\nu; \theta)$;
- 3) распределенная показательным с параметром λ .

Найти распределение случайной величины η — площади квадрата со стороной ξ — и вычислить ее математическое ожидание.

10.17. Радиус окружности ξ — случайная величина,

- 1) распределенная равномерно на промежутке $[a; b]$, $a, b > 0$, $a < b$;
- 2) имеющая гамма-распределение с параметрами $(\nu; \theta)$;

3) распределенная показательно с параметром λ .

Найти распределение случайной величины η — длины окружности радиуса ξ — и вычислить ее математическое ожидание.

10.18. Длина ребра куба ξ — случайная величина, распределенная:

1) равномерно на промежутке $[a; b]$, $a > 0$, $a < b$;

2) показательно с параметром λ .

Найти распределение случайной величины η — объема куба с ребром ξ , вычислить математическое ожидание η .

10.19. На отрезок $[0; 1]$ наудачу бросают точку. Она делит его на две части. Найти распределение и вычислить математическое ожидание случайной величины η .

1. Длины окружности, радиус которой равен длине:

а) меньшей части отрезка; б) большей части отрезка.

2. Площади круга, радиус которого равен длине:

а) меньшей части отрезка; б) большей части отрезка.

3. Площади квадрата со стороной, равной длине:

а) меньшей части отрезка; б) большей части отрезка.

4. Объема куба с ребром, равным длине:

а) меньшей части отрезка; б) большей части отрезка.

10.20. Пусть ξ и η — независимые случайные величины. Вычислить математическое ожидание случайных величин: 1° $\min\{\xi, \eta\}$; 2° $\max\{\xi, \eta\}$; 3° $\xi\eta$; 4° $\eta/(\xi + 1)$; 5° $\eta \exp\{\xi\}$; 6° $\exp\{-\min\{\xi, \eta\}\}$, если: а) ξ и η распределены равномерно на промежутке $[0; 1]$; б) ξ распределена равномерно на промежутке $[0; 1]$, η — на промежутке $[0; 2]$; в) ξ распределена равномерно на промежутке $[0; 1]$, η имеет показательное распределение с параметром λ ; г) ξ и η имеют показательное распределение с параметром λ (для 1°, 2°, 3°, 6°).

10.21. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет своей плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \leq \alpha; \\ \exp\{\alpha - x\}, & \text{если } x > \alpha. \end{cases}$$

Вычислить $M \min\{\xi_1, \xi_2, \dots, \xi_n\}$.

10.22. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых распределена равномерно

на отрезке $[a; b]$. Вычислить математические ожидания случайных величин:

$$1) \min\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 2) \max\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 3) \frac{1}{n} \sum_{i=1}^n \xi_i.$$

10.23. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет своей плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \notin [\theta - h; \theta + h]; \\ 1/2h, & \text{если } x \in [\theta - h; \theta + h]. \end{cases}$$

Вычислить математические ожидания случайных величин:

$$1) \min\{\xi_1, \xi_2, \dots, \xi_n\}; \quad 2) \max\{\xi_1, \xi_2, \dots, \xi_n\};$$

$$3) (\max\{\xi_1, \xi_2, \dots, \xi_n\} - \min\{\xi_1, \xi_2, \dots, \xi_n\})/2.$$

10.24. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых имеет своей плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x \leq \theta; \\ \frac{1}{\alpha} \exp\left\{-\frac{1}{\alpha}(x - \theta)\right\}, & \text{если } x > \theta, \alpha > 0. \end{cases}$$

Вычислить математические ожидания случайных величин:

$$1) \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad 2) \min\{\xi_i\};$$

$$3) \hat{\theta}_1 = \min\{\xi_i\} - \frac{\bar{\xi} - \min\{\xi_i\}}{n}; \quad 4) \hat{\theta}_2 = \bar{\xi} - \hat{\theta}_1.$$

10.25. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ распределены показательно с параметром $1/\theta$. Вычислить математическое ожидание случайной величины $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$.

10.26. На отрезок $[0; T]$ наудачу бросают две точки. Пусть ξ — расстояние между ними. Найти функцию распределения и плотность распределения ξ , вычислить $M\xi$, $D\xi$, $M\xi^n$.

10.27. Точка P равномерно распределена на окружности

$$x^2 + y^2 = 1.$$

Пусть η — длина проекции радиус-вектора $\overline{OP}$ точки P на ось Ox . Найти плотность распределения и математическое ожидание случайной величины η .

10.28*. На окружности радиуса R наудачу выбирают две точки. Найти функцию распределения расстояния η между ними и вычислить $M\eta$.

10.29. На отрезок оси ординат с концами $(0; 0)$ и $(0; R)$ наудачу брошена точка. Через нее проведена хорда окружности $x^2 + y^2 = R^2$ перпендикулярно к оси Oy . Найти функцию распределения длины этой хорды.

10.30*. Пусть $A_x = \{(u, v) : u + v < x\}$ — множество из $\mathbb{R}^2$, x — произвольное, но фиксированное число.

Вычислить $MI_{A_x}(\xi, \eta)$, если известно:

1) распределение Q вектора $\zeta = (\xi, \eta)$;

2) распределения F и G независимых случайных величин ξ и η ;

3) плотности f и g абсолютно непрерывных независимых случайных величин ξ и η .

10.31. Случайная величина η распределена $N_{0;1}$. Вычислить математическое ожидание случайной величины $\eta^+ = \max\{0, \eta\}$.

10.32. Случайная величина η распределена $N_{0;\sigma^2}$. Вычислить математическое ожидание случайной величины $\eta^+ = \max\{0, \eta\}$.

10.33. Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью

$$p(x) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\theta > 0$ ($p(x)$ — плотность показательного распределения с параметром θ).

10.34. Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью

$$p(x) = \begin{cases} \frac{\theta^m}{(m-1)!} x^{m-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\theta > 0$ ($p(x)$ — плотность распределения Эрланга с параметрами $(m; \theta)$, $m = 1, 2, \dots$).

10.35. Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью

$$p(x) = \begin{cases} \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} \exp\left\{-\frac{1}{2}x\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($p(x)$ — плотность χ^2 -распределения с n степенями свободы).

10.36. Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью

$$p(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\sigma > 0$ ($p(x)$ — плотность логарифмически нормального распределения с параметрами $(\mu; \sigma^2)$).

10.37. Вычислить математическое ожидание, второй момент и дисперсию случайной величины ξ с плотностью

$$p(x) = \begin{cases} \frac{\theta\lambda^\theta}{x^{\theta+1}}, & \text{если } x > \lambda > 0; \\ 0, & \text{если } x \leq \lambda, \end{cases}$$

параметр $\theta > 2$ ($p(x)$ — плотность распределения Парето с параметрами $(\lambda; \theta)$).

10.38. Пусть ξ и η — независимые случайные величины, причем ξ распределена равномерно на промежутке $[1; 2]$, η имеет показательное распределение с параметром θ . Найти математическое ожидание и дисперсию случайных величин: 1) $\zeta_1 = \xi\eta$, 2) $\zeta_2 = \xi + \eta$, 3) $\zeta_3 = \eta/\xi$.

10.39. Точка A равномерно распределена на окружности единичного радиуса с центром в начале координат. Пусть ξ — проекция точки A на ось Ox .

Найти: 1) функцию распределения ξ , 2) плотность распределения ξ .

Вычислить: 1) $M\xi$; 2) $P\{\xi > 1/2\}$.

10.40. Пусть $\zeta = (\xi, \eta)$ — абсолютно непрерывный вектор с плотностью $f_\zeta(x, y)$. Вычислить $M\xi$, $M\eta$.

10.41. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные (каждая с распределением F) случайные

величины со значениями в $\mathbb{R}^1$; $\mathbb{R}^1 = \bigcup_{i=1}^r X_i$, ($X_i \cap X_j = \emptyset$, $i \neq j$); ν_i — число случайных величин $\xi_1, \xi_2, \dots, \xi_n$, попавших в X_i , $p_i = F(X_i)$ — вероятность, того, что ξ_k попадет в X_i , $i = 1, 2, \dots, r$. Вычислить

$$M\nu_i, D\nu_i, i = 1, 2, \dots, r.$$

Глава 11

Свертка

11.1 Свертка вероятностных распределений

Определение. *Вероятностным распределением на $\mathbb{R}^1$ будем называть неотрицательную, нормированную, счетноаддитивную функцию F , заданную на σ -алгебре $\mathfrak{B}^1$ борелевских множеств $\mathbb{R}^1$.*

Другими словами, вероятностное распределение на $\mathbb{R}^1$ — это вероятность на σ -алгебре $\mathfrak{B}^1$ (см. разд. 7.2 в гл. 7).

Определение. Пусть F — вероятностное распределение на $\mathbb{R}^1$. Функцию точки $F(x)$, определенную равенством

$$F(x) = F((-\infty, x)),$$

будем называть *функцией распределения F* .

Распределение F однозначно задается своей функцией распределения $F(x)$.

Определение. Если функцию распределения $F(x)$ можно представить в виде

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

то распределение F называют *абсолютно непрерывным*, а функцию f — *плотностью распределения F* .

Вероятностное распределение F будем называть *дискретным*, если существует не более чем счетное множество $X \subset \mathbb{R}^1$ точек x_i , таких, что $F(\{x_i\}) > 0$ и

$$\sum_{x_i \in X} F(\{x_i\}) = 1$$

(при этом точки x_i будем называть *атомами распределения* F и будем говорить, что распределение F сосредоточено на множестве X).

Напомним, что интеграл Лебега $\int_X \varphi(y) F(dy)$ от функции $\varphi(y)$ по распределению F равен $\int_X \varphi(y) f(y) dy$, если F — абсолютно непрерывное распределение с плотностью f , и $\sum_{x_i \in X} \varphi(x_i) F(\{x_i\})$, если F — дискретное распределение, сосредоточенное на множестве X .

Свертки. Пусть φ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$ и F — вероятностное распределение на $\mathbb{R}^1$.

Определение. *Сверткой* функции φ с вероятностным распределением F будем называть функцию $u(x)$, определенную для каждого $x \in \mathbb{R}^1$ равенством

$$u(x) = \int_{\mathbb{R}^1} \varphi(x - y) F(dy).$$

Обозначать свертку φ с F будем так: $u = F * \varphi$.

Определение. *Сверткой вероятностных распределений* G и F будем называть вероятностное распределение Q , функция распределения $Q(x)$ которого является сверткой функции распределения $G(x)$ с распределением F :

$$Q(x) = \int_{\mathbb{R}^1} G(x - y) F(dy).$$

Обозначать свертку вероятностного распределения G с вероятностным распределением F будем символом $F * G$.

Заметим, что если распределение F абсолютно непрерывно и f — его плотность, то

$$Q(x) = \int_{\mathbb{R}^1} G(x-y)F(dy) = \int_{\mathbb{R}^1} G(x-y)f(y)dy. \quad (11.1.1)$$

В классе вероятностных распределений операция свертки коммутативная и ассоциативная. Другими словами, если F, G, Q — вероятностные распределения, то

$$F * G = G * F, \\ (F * G) * Q = F * (G * Q).$$

Если хотя бы одно из вероятностных распределений абсолютно непрерывно, то их свертка так же абсолютно непрерывна. Более того, имеет место такое утверждение.

*Теорема. Свертка $V = F * G$ абсолютно непрерывного вероятностного распределения G с вероятностным распределением F является абсолютно непрерывным распределением и его плотность v равна свертке плотности g абсолютно непрерывного распределения G с распределением F :*

$$v(x) = \int_{\mathbb{R}^1} g(x-t)F(dt). \quad (11.1.2)$$

В частности, если F и G — абсолютно непрерывные, то теорему можно сформулировать так:

Свертка абсолютно непрерывных вероятностных распределений G и F соответственно с плотностями g и f является абсолютно непрерывным распределением, и его плотность v равна свертке плотностей g и f :

$$v(x) = \int_{\mathbb{R}^1} g(x-y)f(y)dy = \int_{\mathbb{R}^1} f(x-y)g(y)dy \quad (11.1.3)$$

(будем обозначать $v = f * g = g * f$).

Пример 11.1.1. Доказать, что класс гамма-распределений замкнут относительно операции свертки, а именно:

$$f_{\nu;\theta} * f_{\mu;\theta} = f_{\mu;\theta} * f_{\nu;\theta} = f_{\nu+\mu;\theta}.$$

Решение. Плотностью гамма-распределения с параметрами $(\nu; \theta)$ является

$$f_{\nu;\theta}(x) = \begin{cases} \frac{\theta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

$\theta > 0, \nu > 0$.

При $x \leq 0$ равенство

$$f_{\mu;\theta} * f_{\nu;\theta}(x) = f_{\nu+\mu;\theta}(x)$$

очевидно.

При $x > 0$ имеем

$$\begin{aligned} f_{\mu;\theta} * f_{\nu;\theta}(x) &= \int_0^{+\infty} f_{\nu;\theta}(x-y) f_{\mu;\theta}(y) dy = \\ &= \int_0^x f_{\nu;\theta}(x-y) f_{\mu;\theta}(y) dy = \\ &= \int_0^x \frac{\theta^\nu}{\Gamma(\nu)} (x-y)^{\nu-1} e^{-\theta(x-y)} \frac{\theta^\mu}{\Gamma(\mu)} y^{\mu-1} e^{-\theta y} dy = \\ &= \frac{\theta^{\nu+\mu}}{\Gamma(\nu)\Gamma(\mu)} e^{-\theta x} \int_0^x (x-y)^{\nu-1} y^{\mu-1} dy. \end{aligned}$$

Выполнив в последнем интеграле замену $y = xt$, получим:

$$\frac{\theta^{\nu+\mu}}{\Gamma(\nu)\Gamma(\mu)} e^{-\theta x} \int_0^x (x-y)^{\nu-1} y^{\mu-1} dy =$$

$$\begin{aligned}
&= \frac{\theta^{\nu+\mu}}{\Gamma(\nu+\mu)} x^{\nu+\mu-1} e^{-\theta x} \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt = \\
&= f_{\nu+\mu;\theta}(x) \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt.
\end{aligned}$$

Итак,

$$f_{\mu;\theta} * f_{\nu;\theta}(x) = f_{\nu+\mu;\theta}(x) \frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt.$$

Поскольку $f_{\mu;\theta} * f_{\nu;\theta}$ и $f_{\nu+\mu;\theta}$ — плотности, то, проинтегрировав последнее равенство по прямой, получаем, что

$$\frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)} \int_0^1 t^{\mu-1} (1-t)^{\nu-1} dt = 1.$$

Следовательно,

$$f_{\mu;\theta} * f_{\nu;\theta} = f_{\nu+\mu;\theta}.$$

11.2 Распределение суммы независимых случайных величин

Следующее утверждение устанавливает связь между сверткой вероятностных распределений и распределением суммы независимых случайных величин.

Теорема. *Функция распределения суммы независимых случайных величин равна свертке функций распределений слагаемых.*

Другими словами, если ξ и η — независимые случайные величины соответственно с распределениями F и G ,

а $Q(x)$ — функция распределения суммы, то

$$Q(x) = \int_{\mathbb{R}^1} G(x-y)F(dy).$$

Отсюда и из формул (11.1.2), (11.1.3) имеем.

Сумма независимых случайных величин, хотя бы одна из которых абсолютно непрерывна, является абсолютно непрерывной случайной величиной, и ее плотность распределения равна свертке плотности распределения абсолютно непрерывной случайной величины с распределением другой.

Плотность распределения суммы независимых абсолютно непрерывных случайных величин равна свертке плотностей распределений слагаемых. Другими словами, если ξ и η — независимые абсолютно непрерывные случайные величины соответственно с плотностями $p_\xi(t)$ и $p_\eta(t)$, то плотность распределения $u(x)$ суммы $\xi + \eta$ равна свертке плотностей $p_\xi(t)$ и $p_\eta(t)$, т. е.

$$u(x) = \int_{\mathbb{R}^1} p_\xi(x-y)p_\eta(y)dy = \int_{\mathbb{R}^1} p_\eta(x-y)p_\xi(y)dy. \quad (11.2.1)$$

В тех же предположениях плотность распределения $v(x)$ разности $\zeta = \xi - \eta$ равна

$$v(x) = \int_{\mathbb{R}^1} p_\xi(x+y)p_\eta(y)dy. \quad (11.2.2)$$

Пример 11.2.1. Пусть ξ и η — независимые случайные величины, каждая из которых равномерно распределена на $[0; 1]$.

Найти:

1° плотность распределения суммы $\zeta = \xi + \eta$;

2° функцию распределения суммы $\zeta = \xi + \eta$;

3° вероятность $P\{|\xi + \eta - 1/2| < 1\}$.

Решение. 1° По известным плотностям распределений случайных величин ξ и η :

$$f_\xi(y) = \begin{cases} 1, & \text{если } y \in [0; 1]; \\ 0, & \text{если } y \notin [0; 1], \end{cases}$$

$$f_{\eta}(y) = \begin{cases} 1, & \text{если } y \in [0; 1]; \\ 0, & \text{если } y \notin [0; 1], \end{cases}$$

плотность распределения суммы $\zeta = \xi + \eta$ находим как свертку плотностей слагаемых (см. (11.2.1)):

$$f_{\zeta}(x) = \int_{\mathbb{R}^1} f_{\xi}(x-y)f_{\eta}(y)dy = \int_0^1 f_{\xi}(x-y)dy = \int_{x-1}^x f_{\xi}(t)dt$$

(воспользовались заменой переменной $x-y=t$). Вычислив последний интеграл для каждого значения $x \in \mathbb{R}^1$, пробегающего $\mathbb{R}^1 = (-\infty, +\infty)$ от $-\infty$ до $+\infty$ (интегрирование ведется по промежутку $[x-1, x]$), получим:

$$\text{если } x < 0, \text{ то } \int_{x-1}^x f_{\xi}(t)dt = \int_{x-1}^x 0dt = 0;$$

$$\text{если } 0 \leq x < 1, \text{ то } \int_{x-1}^x f_{\xi}(t)dt = \int_{x-1}^0 0dt + \int_0^x 1dt = x;$$

$$\begin{aligned} \text{если } 0 \leq x-1 < 1, \text{ то } \int_{x-1}^x f_{\xi}(t)dt &= \int_{x-1}^1 1dt + \int_1^x 0dt = \\ &= 2-x; \end{aligned}$$

$$\text{если } x-1 \geq 1, \text{ то } \int_{x-1}^x f_{\xi}(t)dt = \int_{x-1}^x 0dt = 0.$$

Таким образом, плотность распределения случайной величины $\zeta = \xi + \eta$

$$f_{\zeta}(x) = \begin{cases} 0, & \text{если } x < 0; \\ x, & \text{если } 0 \leq x < 1; \\ 2-x, & \text{если } 1 \leq x < 2; \\ 0, & \text{если } x \geq 2. \end{cases}$$

2° Функцию распределения $F_{\zeta}(x)$ случайной величины ζ по ее плотности $f_{\zeta}(t)$ находим следующим образом:

$$F_{\zeta}(x) = \int_{-\infty}^x f_{\zeta}(t)dt =$$

$$= \begin{cases} \int_{-\infty}^x 0 dt = 0, & \text{если } x < 0; \\ \int_{-\infty}^0 0 dt + \int_0^x t dt = x^2/2, & \text{если } x \in [0, 1); \\ \int_0^1 t dt + \int_1^x (2-t) dt = -(x-2)^2/2 + 1, & \text{если } x \in [1, 2); \\ \int_{-\infty}^0 0 dt + \int_0^1 t dt + \int_1^2 (2-t) dt = 1, & \text{если } x \geq 2. \end{cases}$$

3° По известной плотности распределения $f_{\zeta}(t)$ случайной величины ζ вероятность попадания значения ζ в множество B , вычисляется так:

$$P\{\zeta \in B\} = \int_B f_{\zeta}(t) dt$$

(см. формулу (9.1.5)). В частности,

$$\begin{aligned} P\{|\xi + \eta - 1/2| < 1\} &= P\{|\zeta - 1/2| < 1\} = \\ &= P\{-1/2 < \zeta < 3/2\} = \\ &= \int_{-1/2}^{3/2} f_{\zeta}(t) dt = \int_{-1/2}^0 0 dt + \int_0^1 t dt + \int_1^{3/2} (2-t) dt = \frac{7}{8}. \end{aligned}$$

Пример 11.2.2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых распределена $N_{0;1}$. Найти распределение случайной величины

$$\eta = \sum_{i=1}^n \xi_i^2.$$

Решение. Сначала проверим, что если случайная величина ξ распределена $N_{0;1}$, то $\eta = \xi^2$ имеет гамма-распределение с параметрами $(1/2; 1/2)$. В самом деле,

$$F_{\eta}(x) = P\{\eta < x\} = P\{\xi^2 < x\}.$$

При $x \leq 0$ значение $F_\eta(x) = P\{\xi^2 < x\}$ равно нулю, а при $x > 0$

$$F_\eta(x) = P\{|\xi| < \sqrt{x}\} = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} \exp\{-t^2/2\} dt.$$

После замены $t^2 = s$ в последнем интеграле имеем

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}} \int_0^x s^{-1/2} e^{-s/2} ds = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \int_0^x s^{1/2-1} e^{-s/2} ds,$$

а это функция распределения гамма-распределения с параметрами $(1/2; 1/2)$ (воспользовались тем, что $\Gamma(1/2) = \sqrt{\pi}$).

Далее, поскольку гамма-распределение замкнуто относительно операции свертки (см. пример 11.1.1), то сумма независимых гамма-распределенных с параметрами $(1/2; 1/2)$ случайных величин является гамма-распределенной случайной величиной с параметрами $(n/2; 1/2)$.

Гамма-распределение с параметрами $(n/2; 1/2)$ еще называют χ^2 -распределением с n степенями свободы.

11.3 Задачи

АЗ: 11.2(1), 11.4(1, 3), 11.13(1), 11.17.

СЗ: 11.2(2, 3), 11.3, 11.4(2, 6), 11.13(2), 11.18.

11.1. Пусть ξ и η — независимые случайные величины; ξ распределена равномерно на промежутке $[0; 1]$, η имеет своим распределением

$$P_\eta(k) = P\{\eta = k\} = 1/2, \quad k = 0, 1.$$

Найти распределение случайной величины $\zeta = \xi + \eta$.

11.2. Случайные величины ξ и η независимы и распределены равномерно на промежутке: 1) $[a; b]$, $a < b$; 2) $[0; a]$, $a > 0$; 3) $[-a; a]$, $a > 0$; 4) $[-1/2; 1/2]$.

Найти плотность распределения суммы $\zeta = \xi + \eta$.

11.3. Пусть ξ и η — независимые случайные величины, равномерно распределенные соответственно на промежутках $[0; 1]$ и $[0; 2]$. Найти плотность распределения $p(x)$ суммы $\zeta = \xi + \eta$.

11.4. Пусть ξ и η — независимые случайные величины, ξ распределена равномерно на промежутке $[-1; 1]$, η — на промежутке $[0; 1]$.

Вычислить:

1) $P\{\xi^2 + \eta > 1/2\}$;

2) $P\{\xi + \eta > 1\}$;

3) $P\{|\xi + \eta| > 1/2\}$;

4) $P\{|\eta - \xi| < 1/2\}$;

5) $P\{\eta^2 - \xi > 0\}$;

6) $P\{|\xi| + \eta > 1\}$.

11.5. Случайные величины ξ и η — независимые и распределены равномерно на промежутке: 1) $[a; b]$, $a < b$; 2) $[0; a]$, $a > 0$; 3) $[-a; a]$, $a > 0$; 4) $[0; 1]$. Найти плотность распределения $p_\zeta(x)$ разности $\zeta = \xi - \eta$.

11.6. На отрезок $[0; 1]$ наудачу бросают две точки. Пусть ξ — координата одной точки, η — другой. Найти функцию распределения случайной величины $\zeta = (\xi + \eta)/2$.

11.7. Пусть ξ — координата точки, наудачу брошенной на отрезок $[0; 1]$, η — число очков, выпавших в результате подбрасывания симметричной игральной кости. Найти распределение случайной величины $\zeta = \xi + \eta$.

11.8. На отрезок $[0; 1]$ наудачу бросают две точки, ξ — координата одной точки, η — другой. Для $0 < x < 1$ найти $P\{|\eta - \xi| < x\}$.

11.9. Координаты каждой из двух случайно выбранных на отрезке $[0; 1]$ точек распределены равномерно. Найти математическое ожидание расстояния между ними.

11.10. Пусть ξ_1 и ξ_2 — независимые случайные величины с плотностями

$$p_i(x) = \begin{cases} \lambda_i e^{-\lambda_i x}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0; \end{cases}$$

$\lambda_i > 0$, $i = 1, 2$; $\lambda_1 \neq \lambda_2$.

Найти плотности распределения: 1) суммы $\xi_1 + \xi_2$; 2) разности $\xi_2 - \xi_1$.

11.11. Пусть ξ и η — независимые случайные величины, распределенные показательно с одним и тем же параметром λ . Найти плотности распределения случайных величин: 1) $\xi + \eta$; 2) $\xi - \eta$; 3) $|\xi - \eta|$.

11.12. Пусть ξ и η — независимые случайные величины, распределенные показательно соответственно с параметрами 1 и 2. Найти плотность распределения случайной величины $\xi_1 + \xi_2$.

11.13. Пусть ξ и η — независимые одинаково распределенные случайные величины с плотностями $p(x) = \exp\{-|x|\}/2$. Найти плотности распределения случайных величин: 1) $\xi + \eta$; 2) $\xi - \eta$.

11.14. Пусть ξ и η — независимые случайные величины, ξ распределена равномерно на промежутке $[-a; a]$, η имеет показательное распределение с параметром λ . Найти плотность распределения случайной величины $\zeta = \xi + \eta$.

11.15. Пусть ξ и η — независимые случайные величины, ξ_1 распределена равномерно на промежутке $[-1; 1]$, ξ_2 имеет показательное распределение с параметром $\lambda = 1$. Найти плотность распределения случайной величины $\eta = \xi_1 + \xi_2$.

11.16. Случайная величина η равномерно распределена на промежутке $[-h; h]$, ξ имеет своей функцией распределения $F(x)$, ξ и η — независимы. Найти функцию распределения и плотность распределения (если она существует) суммы $\zeta = \xi + \eta$.

11.17. Пусть ξ и η — независимые случайные величины. Случайная величина ξ имеет распределение

$$P\{\xi = (-1)^k\} = G\left(\{(-1)^k\}\right) = 1/2, \quad k = 0, 1,$$

а η — распределение Q . Найти функцию распределения случайной величины $\zeta = \xi + \eta$.

11.18. Пусть ξ и η — независимые случайные величины; ξ распределена равномерно на промежутке $[0; 1]$, η имеет своим распределением

$$P\{\eta = k\} = (1/4)^k (3/4)^{1-k}, \quad k = 0, 1.$$

Найти распределение случайной величины $\zeta = \xi + \eta$.

11.19. Найти распределение суммы n независимых случайных величин, каждая из которых имеет показательное распределение с параметром λ . Сравните ее с плотностью распределения Эрланга (см. разд. 9.3 в гл. 9).

11.20. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые равномерно распределенные на $[0; 1]$ случайные величины. Найти распределение случайной величины

$$\eta = \xi_1 \xi_2 \dots \xi_n.$$

Вычислить $M\eta$.

11.21. Независимые случайные величины ξ и η имеют гамма-распределения соответственно с параметрами (ν, θ) и (μ, θ) . Найти плотность распределения случайной величины $\xi + \eta$.

11.22. Пусть независимые случайные величины ξ и η имеют χ^2 -распределения соответственно с n и m степенями свободы. Доказать, что случайная величина $\xi + \eta$ имеет χ^2 -распределение с $n + m$ степенями свободы.

Глава 12

Сходимость распределений

12.1 Определения. Примеры

Мы будем рассматривать вероятностные распределения на $\mathbb{R}^1$ (см. разд. 11.1 в гл. 11).

Напомним, что для распределения F и его функции распределения

$$F(x) = F((-\infty, x))$$

имеет место соотношение

$$F([a; b)) = F(b) - F(a)$$

(при $a < b$), а значение $F(\{x_0\})$ распределения F на одноточечном множестве $\{x_0\}$ через функцию распределения $F(x)$ вычисляется так:

$$F(\{x_0\}) = F(x_0 + 0) - F(x_0 - 0).$$

О п р е д е л е н и е. Распределение F на $\mathbb{R}^1$ будем называть *собственным вероятностным распределением* (вероятностным распределением), если $F(\mathbb{R}^1) = 1$, и *несобственным*, если $F(\mathbb{R}^1) < 1$. В терминах функции распределения: F — собственное вероятностное распределение, если $F(+\infty) = 1$; F — несобственное вероятностное распределение, если $F(+\infty) < 1$.

Несобственные вероятностные распределения естественным образом возникают как пределы последовательностей вероятностных распределений.

Определение. Точку x_0 будем называть *атомом* распределения F , если $F(\{x_0\}) > 0$. Распределение F будем называть *атомическим*, если оно сосредоточено на множестве своих атомов.

Определение. Интервал I вида $[a; b)$ (конечный или бесконечный) будем называть *интервалом непрерывности* распределения F , если точки a и b не являются атомами распределения F .

Определение. Последовательность распределений $\{F_n\}$ *сходится к распределению F* при $n \rightarrow \infty$, если $F_n(I) \rightarrow F(I)$ для каждого конечного интервала непрерывности I распределения F (обозначение: $F_n \rightarrow F$ или $\lim_n F_n = F$). Если при этом F — собственное вероятностное распределение, то будем говорить, что последовательность $\{F_n\}$ *сходится к F собственно*, если F — несобственное, то — *несобственно*.

Часто полезными являются следующие достаточные условия сходимости распределений.

1° Если последовательность функций распределений $\{F_n(x)\}$ сходится к функции распределения $F(x)$ в каждой точке непрерывности последней, то последовательность распределений $\{F_n\}$ сходится к распределению F (в собственном или несобственном смысле).

2° Собственные вероятностные распределения F_σ с одним и тем же средним a и дисперсией σ^2 , стремящейся к 0, сходятся к собственному атомическому распределению, сосредоточенному в точке a .

Пример 12.1.1. Пусть $\{F_n\}$ — последовательность распределений с функциями распределения

$$F_n(x) = \begin{cases} 0, & \text{если } x < 0; \\ nx, & \text{если } 0 \leq x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Исследовать последовательность $\{F_n\}$ на сходимость.

Решение. Докажем, что при $n \rightarrow \infty$ последовательность функций $\{F_n(x)\}$ сходится к функции

$$F(x) = \begin{cases} 1, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

атомического распределения, сосредоточенного в точке $x = 0$, в каждой точке, кроме точки $x = 0$ (точка $x = 0$ является точкой разрыва $F(x)$).

В самом деле, по условию для произвольной, но фиксированной точки x , лежащей слева от нуля, при каждом n имеет место равенство $F_n(x) = 0$, поэтому

$$\lim_n F_n(x) = 0 = F(x).$$

Для произвольной, но фиксированной точки x , лежащей справа от нуля, начиная с некоторого N ($n \geq N$) имеет место неравенство $1/n < x$. Поэтому для $n \geq N$ значение $F_n(x)$ равно 1, и следовательно, $\lim_n F_n(x) = 1 = F(x)$.

Таким образом, при $n \rightarrow \infty$

$$F_n(x) \rightarrow F(x)$$

в каждой точке $x \in \mathbb{R}^1$, исключая, быть может, точку 0. Отсюда, согласно достаточному условию 1° сходимости распределений, при $n \rightarrow \infty$

$$F_n \rightarrow F.$$

Следовательно, последовательность распределений $\{F_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

Пример 12.1.2. Пусть $\{Q_n\}$ — последовательность распределений с плотностями

$$q_n(x) = \frac{n}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - (-1)^n)^2 n^2}{2} \right\}, \quad n = 1, 2, \dots$$

Исследовать последовательность $\{Q_n\}$ на сходимость.

Решение. Согласно достаточному условию сходимости 2° для четных n последовательность распределений $\{Q_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 1, а для нечетных — в точке -1 . Поэтому последовательность $\{Q_n\}$ не является сходящейся.

Пример 12.1.3. Пусть F_h — вероятностные распределения с плотностями

$$f_h(x) = \frac{1}{\sqrt{2\pi h}} \exp \left\{ -\frac{(x-a)^2}{2h^2} \right\}.$$

Исследовать F_h на сходимость 1° при $h \rightarrow \infty$, 2° при $h \rightarrow 0$.

Решение. 1° Заметим, что функция $F(x) = c$, $x \in \mathbb{R}^1$, где c — константа из промежутка $[0; 1]$, является функцией несобственного вероятностного распределения, тождественно равного нулю. В самом деле, значение $F([a; b])$ распределения F на промежутке $[a; b]$ равно $F(b) - F(a) = 0$, поэтому $F(A) = 0$ и на множествах A из алгебры $\mathfrak{A}$ конечных объединений непересекающихся промежутков вида $[a; b]$ (a и b не обязательно должны быть конечными), а следовательно, и на борелевских множествах, поскольку $\sigma(\mathfrak{A}) = \mathfrak{B}^1$. Поэтому, если последовательность $\{F_n(x)\}$ функций распределения сходится к функции $F(x) = c$ (c — константа из промежутка $[0; 1]$), то F_h при $h \rightarrow \infty$ сходится к распределению F , тождественно равному нулю.

Далее,

$$\begin{aligned} F_h(x) &= \int_{-\infty}^x f_h(t) dt = \frac{1}{\sqrt{2\pi h}} \int_{-\infty}^x \exp \left\{ -\frac{(t-a)^2}{2h^2} \right\} dt = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a)/h} \exp \left\{ -\frac{u^2}{2} \right\} du \end{aligned}$$

(мы воспользовались заменой $(t-a)/h = u$). Для каждого

фиксированного x

$$\lim_{h \rightarrow \infty} F_h(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \exp \left\{ -\frac{u^2}{2} \right\} du = \frac{1}{2}.$$

Поэтому F_h при $h \rightarrow \infty$ сходится к распределению, тождественно равному нулю.

2° При $h \rightarrow 0$ семейство распределений F_h сходится к собственному атомическому распределению, сосредоточенному в точке a (согласно достаточному условию сходимости к атомическому распределению).

О п р е д е л е н и е. Последовательность распределений $\{F_n\}$ слабо сходится к распределению F при $n \rightarrow \infty$ относительно класса функций U , если для каждой функции $u \in U$

$$\int_{\mathbb{R}^1} u(x) F_n(dx) \rightarrow \int_{\mathbb{R}^1} u(x) F(dx) \text{ при } n \rightarrow \infty.$$

Пусть $C(-\infty; +\infty)$ — класс непрерывных ограниченных функций на $\mathbb{R}^1$; $C_0[-\infty; +\infty]$ — класс непрерывных ограниченных функций, для которых $\lim_{x \rightarrow +\infty} u(x) = 0$, $\lim_{x \rightarrow -\infty} u(x) = 0$.

Теорема 12.1.1. Из сходимости (собственной или несобственной) последовательности вероятностных распределений $\{F_n\}$ к распределению F следует слабая сходимость $\{F_n\}$ к F относительно класса $C_0[-\infty; +\infty]$ и наоборот.

Из собственной сходимости вероятностных распределений $\{F_n\}$ к распределению F следует слабая сходимость $\{F_n\}$ к F относительно класса $C(-\infty; +\infty)$.

Пример 12.1.4. Пусть $\{F_n\}$ — последовательность вероятностных распределений с плотностями

$$f_n(x) = \frac{n}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2 n^2}{2} \right\}, \quad n = 1, 2, \dots$$

Вычислить

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^1} e^{itx} F_n(dx).$$

Решение. Последовательность распределений $\{F_n\}$ сходится к собственному атомическому распределению F , сосредоточенному в точке 0 (согласно достаточному условию 2°). Поэтому согласно теореме 12.1.1 последовательность $\{F_n\}$ слабо сходится к F относительно класса $C(-\infty; +\infty)$. И поскольку $e^{itx} \in C(-\infty; +\infty)$, то

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^1} e^{itx} F_n(dx) = \int_{\mathbb{R}^1} e^{itx} F(dx) = e^{it0} F(\{0\}) = 1.$$

Пример 12.1.5. Пусть F — вероятностное распределение со средним m и дисперсией σ^2 . Доказать, что для $a > 0$

$$F\{x : |x - m| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

Решение.

$$\begin{aligned} \sigma^2 &= \int_{\mathbb{R}^1} (x - m)^2 F(dx) \geq \int_{x: |x-m| \geq a} (x - m)^2 F(dx) \geq \\ &\geq \int_{x: |x-m| \geq a} a^2 F(dx) = a^2 \int_{x: |x-m| \geq a} F(dx) = a^2 F\{x : |x - m| \geq a\}. \end{aligned}$$

Пример 12.1.6. Пусть $\{F_n\}$ — последовательность собственных распределений соответственно со средними a_n , сходящимися к a , и дисперсиями σ_n^2 , сходящимися к 0.

Доказать, что последовательность $\{F_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке a .

Решение. Достаточно доказать, что последовательность $\{F_n(x)\}$ функций распределения сходится к функции $F_a(x)$ атомического распределения, сосредоточенного в точке a , т. е. к

$$F_a(x) = \begin{cases} 0, & \text{если } x \leq a; \\ 1, & \text{если } x > a. \end{cases}$$

Пусть $t > 0$, $x = a - 2t$. Поскольку $a_n \rightarrow a$, при $n \rightarrow \infty$, то при достаточно больших n

$$\begin{aligned} F_n(x) &= F_n(a-2t) = F_n((-\infty, a-2t)) \leq F_n((-\infty, a_n-t)) = \\ &= F_n\{y : y < a_n - t\} \leq F_n\{y : |y - a_n| \geq t\} \leq \frac{\sigma_n^2}{a^2}, \end{aligned}$$

а поэтому $F_n(a-2t) \rightarrow 0$, при $n \rightarrow \infty$ (воспользовались примером 12.1.5).

Аналогично убеждаемся, что $F_n(a+2t) \rightarrow 1$, при $n \rightarrow \infty$.

12.2 Задачи

АЗ: 12.1, 12.3, 12.5, 12.7(2, 5), 12.9, 12.11, 12.12, 12.15*.

СЗ: 12.2, 12.4, 12.6, 12.7(1, 3, 4), 12.10, 12.16*.

12.1. Пусть $\{N_{a;\sigma_n^2}\}$ — последовательность нормальных распределений соответственно со средним a и дисперсиями σ_n^2 . Доказать, что если σ_n^2 сходится к нулю, то $\{N_{a;\sigma_n^2}\}$ сходится к собственному атомическому распределению, сосредоточенному в точке a .

12.2. Пусть F_h — семейство вероятностных распределений с плотностями

$$f_h(x) = \frac{1}{\sqrt{2\pi}h} \exp \left\{ -\frac{(x-a)^2}{2h^2} \right\}, \quad h > 0.$$

Исследовать на сходимость F_h при $h \rightarrow 0$.

12.3. Пусть $F(x)$ — монотонно возрастающая функция распределения. Исследовать на сходимость последовательность распределений $\{F_n\}$, заданных своими функциями распределений,

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq -1/n; \\ \frac{F(x) - F(-1/n)}{F(1/n) - F(-1/n)}, & \text{если } -1/n < x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

12.4. Пусть $\{F_n\}$ — последовательность распределений с функциями

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq -1/n; \\ n(x + 1/n)/2, & \text{если } -1/n < x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Исследовать последовательность распределений $\{F_n\}$ на сходимость.

12.5. Пусть $\{F_n\}$ — последовательность распределений с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [-1/n; 1/n]; \\ 0, & \text{если } x \notin [-1/n; 1/n]. \end{cases}$$

Исследовать последовательность распределений $\{F_n\}$ на сходимость.

12.6. Пусть $\{F_n\}$ — последовательность распределений соответственно с плотностями

$$f_n(x) = \begin{cases} 0, & \text{если } x \leq -1/n; \\ n^2(x + 1/n), & \text{если } -1/n < x \leq 0; \\ -n^2(x - 1/n), & \text{если } 0 < x \leq 1/n; \\ 0, & \text{если } x > 1/n. \end{cases}$$

Исследовать последовательность распределений $\{F_n\}$ на сходимость при $n \rightarrow \infty$.

12.7. Пусть $F(x)$ — функция распределения непрерывного собственного вероятностного распределения.

Исследовать на сходимость каждую из приведенных далее последовательностей распределений, заданных своими функциями распределения:

- 1) $F_n(x) = F(x + 1/n)$, $n = 1, 2, \dots$;
- 2) $G_n(x) = F(x + n)$, $n = 1, 2, \dots$;
- 3) $S_n(x) = F(x - n)$, $n = 1, 2, \dots$;
- 4) $P_n(x) = F(x/n)$, $n = 1, 2, \dots$;
- 5) $Q_n(x) = F(x + (-1)^n n)$, $n = 1, 2, \dots$.

12.8. Исследовать на сходимость последовательность распределений $\{P_n\}$, заданных плотностями

$$p_n(x) = \frac{n}{\sqrt{2\pi}} \exp \left\{ -\frac{(x-1)^2 n^2}{2} \right\}, \quad n = 1, 2, \dots$$

12.9. Исследовать на сходимость последовательность распределений:

- 1) $F_n : \begin{pmatrix} -n & n \\ 1/2 & 1/2 \end{pmatrix}$, $n = 1, 2, \dots$;
- 2) $G_n : \begin{pmatrix} -1/n & 1/n \\ 1/2 & 1/2 \end{pmatrix}$, $n = 1, 2, \dots$.

12.10. Пусть $\{F_n\}$ — последовательность распределений соответственно с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [(-1)^n - 1/n; (-1)^n + 1/n]; \\ 0, & \text{если } x \notin [(-1)^n - 1/n; (-1)^n + 1/n]. \end{cases}$$

Исследовать $\{F_n\}$ на сходимость при $n \rightarrow \infty$.

12.11*. Пусть

$$N_{x;\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^y \exp \left\{ -\frac{(t-x)^2}{2\sigma^2} \right\} dt; \quad x, y \in \mathbb{R}^1, \sigma > 0.$$

Вычислить

$$\lim_{\sigma \rightarrow 0} N_{x;\sigma^2}(y).$$

12.12. Пусть $\{F_n\}$ — последовательность вероятностных распределений с плотностями

$$f_n(x) = \frac{n}{\sqrt{2\pi}} \exp \left\{ -\frac{(x-1)^2 n^2}{2} \right\}, \quad n = 1, 2, \dots$$

Вычислить

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^1} \sin x F_n(dx).$$

12.13. Пусть $\{F_n\}$ — последовательность вероятностных распределений с функциями распределения из задачи 12.12.

Вычислить

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^1} \cos x F_n(dx).$$

12.14*. Пусть F — собственное вероятностное распределение,

$$N_{x;\sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^y \exp \left\{ -\frac{(t-x)^2}{2\sigma^2} \right\} dt; \quad x, y \in \mathbb{R}^1, \sigma > 0.$$

Вычислить

$$\lim_{\sigma \rightarrow 0} \int_{-\infty}^{+\infty} N_{x;\sigma^2}(y) F(dx),$$

если y — точка непрерывности F .

12.15*. Пусть

$$F_\lambda(y) = \sum_{k: 0 \leq k < y} \frac{\lambda^k}{k!} e^{-\lambda}, \quad y \in \mathbb{R}^1, \lambda > 0.$$

Вычислить

$$\lim_{\lambda \rightarrow 0} F_\lambda(y).$$

12.16*. Пусть

$$F_\lambda(y) = \sum_{k: 0 \leq k < y} \frac{\lambda^k}{k!} e^{-\lambda}, \quad y \in \mathbb{R}^1, \lambda > 0.$$

Вычислить

$$\lim_{\lambda \rightarrow 0} \int_{-\infty}^{+\infty} e^{-y^2/2} F_{\lambda}(dy).$$

12.17. Пусть $\{F_n\}$ — последовательность распределений с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [(-1)^n/n - 1/n; (-1)^n/n + 1/n]; \\ 0, & \text{если } x \notin [(-1)^n/n - 1/n; (-1)^n/n + 1/n] \end{cases}$$

соответственно.

Исследовать последовательность $\{F_n\}$ на сходимость при $n \rightarrow \infty$.

12.18. Пусть последовательность $\{F_n\}$ вероятностных распределений и распределение F сосредоточены на множестве натуральных чисел и при $n \rightarrow \infty$

$$F_n(\{k\}) \rightarrow F(\{k\})$$

для каждого натурального k . Доказать, что $F_n \rightarrow F$ при $n \rightarrow \infty$.

12.19. Пусть $\{F_k\}$ — последовательность собственных вероятностных распределений, $\{p_k\}$ — последовательность неотрицательных чисел, для которых $\sum_{k=0}^{\infty} p_k = 1$.

Доказать, что последовательность распределений

$$U_n = \sum_{k=0}^n p_k F_k, \quad n = 1, 2, \dots,$$

сходится в собственном смысле к распределению

$$U = \sum_{k=0}^{\infty} p_k F_k.$$

12.20. Пусть F_h — семейство вероятностных распределений с плотностями

$$f_h(x) = \begin{cases} \frac{1}{h}, & \text{если } x \in [a, a+h]; \\ 0, & \text{если } x \notin [a, a+h] \end{cases}$$

(a — фиксировано). Исследовать на сходимость F_h :
1) при $h \rightarrow \infty$, 2) при $h \rightarrow 0$.

12.21. Пусть G_h — семейство вероятностных распределений с плотностями

$$g_h(x) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [-h, h]; \\ 0, & \text{если } x \notin [-h, h]. \end{cases}$$

Исследовать на сходимость G_h : 1) при $h \rightarrow \infty$,
2) при $h \rightarrow 0$.

Глава 13

Характеристическая функция

13.1 Определения, свойства, вычисление

Определение. Пусть ξ — случайная величина, F — ее распределение. *Характеристической функцией случайной величины ξ (распределения F)* будем называть комплекснозначную функцию $\varphi(t)$, определенную для всех $t \in \mathbb{R}^1$ равенством

$$\varphi(t) = Me^{it\xi} = \int_{\mathbb{R}^1} e^{itx} F(dx).$$

Если распределение F (распределение случайной величины ξ) имеет плотность f , то

$$\varphi(t) = Me^{it\xi} = \int_{\mathbb{R}^1} e^{itx} f(x) dx;$$

если распределение F (распределение случайной величины ξ) дискретно т. е.

$$F: x_k \rightarrow F(\{x_k\}) > 0, \quad k = 1, 2, \dots; \quad \sum_{x_k} F(\{x_k\}) = 1,$$

то

$$\varphi(t) = M e^{it\xi} = \sum_{x_k} \exp\{itx_k\} F(\{x_k\}).$$

Многочисленные применения характеристических функций используют приведенное далее свойство.

Мультипликативное свойство характеристических функций. *Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых. В терминах свертки характеристическая функция свертки вероятностных распределений равна произведению их характеристических функций.*

Теорема (о дифференцируемости характеристической функции). *Если n -й абсолютный момент распределения F конечный, то существует n -я производная характеристической функции*

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} F(dx)$$

распределения F , и ее можно получить дифференцированием под знаком интеграла:

$$\varphi^{(n)}(t) = i^n \int_{-\infty}^{+\infty} e^{itx} x^n F(dx).$$

Следствие. *В условиях теоремы*

$$\varphi^{(n)}(0) = i^n \int_{-\infty}^{+\infty} x^n F(dx) = i^n M \xi^n,$$

$$\varphi(t) = 1 + \frac{t}{1!} \varphi^{(1)}(0) + \frac{t^2}{2!} \varphi^{(2)}(0) + \dots + \frac{t^n}{n!} \varphi^{(n)}(0) + o(t^n),$$

при $t \rightarrow 0$.

Пример 13.1.1. Вычислить характеристическую функцию случайной величины ξ , имеющей распределение Пуассона с параметром λ :

$$P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Решение. Согласно определению

$$\begin{aligned} \varphi(t) &= M e^{it\xi} = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = \\ &= \exp\{-\lambda\} \exp\{\lambda \exp\{it\}\} = \exp\{\lambda(\exp\{it\} - 1)\}. \end{aligned}$$

Пример 13.1.2. Найти характеристическую функцию нормально распределенной с параметрами $(a; \sigma^2)$ случайной величины.

Решение. Сначала найдем характеристическую функцию $\varphi(t)$ случайной величины ξ , распределенной нормально с параметрами $(0; 1)$.

Согласно определению

$$\begin{aligned} \varphi(t) &= M e^{it\xi} = \int_{-\infty}^{+\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} \cos txdx + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} \sin txdx. \end{aligned}$$

Последний интеграл равен нулю, поскольку

$$\int_{-\infty}^{+\infty} e^{-x^2/2} \sin txdx = \int_0^{+\infty} e^{-x^2/2} \sin txdx - \int_0^{+\infty} e^{-u^2/2} \sin tudu.$$

Продифференцируем $\varphi(t)$, воспользовавшись теоремой о дифференцируемости характеристической функции (первый момент нормального распределения конечный). Имеем

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(-\sin tx) e^{-x^2/2} dx =$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sin tx de^{-x^2/2} = \\
&= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t \cos tx \cdot e^{-x^2/2} dx = -t\varphi(t).
\end{aligned}$$

Отсюда

$$\varphi'(t)/\varphi(t) = -t.$$

Решая это уравнение, получаем

$$\varphi(t) = e^{-t^2/2}$$

(мы учли, что $\varphi(0) = 1$).

Далее, если ξ распределена $N_{0;1}$, то $\eta = \sigma\xi + a$ распределена нормально с параметрами $(a; \sigma^2)$ (проверяется непосредственно), поэтому характеристической функцией $\psi(t)$ случайной величины η , распределенной $N_{a; \sigma^2}$, является

$$\begin{aligned}
\psi(t) &= Me^{it\eta} = Me^{it(\sigma\xi+a)} = e^{ita} Me^{i(\sigma t)\xi} = e^{ita} \varphi(\sigma t) = \\
&= \exp\{ita\} \exp\{-\sigma^2 t^2/2\} = \exp\{ita - \sigma^2 t^2/2\}.
\end{aligned}$$

13.2 Теоремы единственности и непрерывности

Теорема единственности. *Различным вероятностным распределениям соответствуют различные характеристические функции.*

В задачах исследования сходимости распределений часто используется теорема непрерывности.

Теорема непрерывности. *Для того чтобы последовательность $\{F_n\}$ вероятностных распределений собственно сходилась, необходимо и достаточно, чтобы*

соответствующая последовательность характеристических функций $\{\varphi_n(t)\}$ сходилась к непрерывной функции $\varphi(t)$ в каждой точке $t \in \mathbb{R}^1$. Предел $\varphi(t)$ последовательности характеристических функций $\{\varphi_n(t)\}$ является характеристической функцией предельного распределения F и $\{\varphi_n(t)\}$ сходится к $\varphi(t)$ равномерно на каждом конечном промежутке.

Пример 13.2.1 Пусть F и Q — нормальные распределения соответственно с параметрами $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$. Найдите свертку $F * Q$ распределений Q и F .

Решение. Воспользовавшись мультипликативным свойством характеристических функций (характеристическая функция свертки вероятностных распределений равна произведению характеристических функций распределений), по характеристическим функциям

$$\varphi_1(t) = \exp \{ ita_1 - t^2 \sigma_1^2 / 2 \} \text{ и } \varphi_2(t) = \exp \{ ita_2 - t^2 \sigma_2^2 / 2 \}$$

нормальных распределений F и Q с параметрами $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$ соответственно (см. пример 13.1.2) находим характеристическую функцию свертки $F * Q$:

$$\varphi(t) = \varphi_1(t) \varphi_2(t) = \exp \{ it(a_1 + a_2) - t^2(\sigma_1^2 + \sigma_2^2)/2 \}.$$

Далее, с одной стороны, $\varphi(t)$ — характеристическая функция свертки $F * Q$, с другой,

$$\varphi(t) = \exp \{ it(a_1 + a_2) - t^2(\sigma_1^2 + \sigma_2^2)/2 \}$$

является характеристической функцией нормального распределения с параметрами $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$, т. е. характеристическая функция распределения $F * Q$ и нормального распределения с параметрами $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$ совпадают. А поэтому согласно теореме единственности совпадают и сами распределения. Следовательно, *свертка нормальных распределений с параметрами $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$ является нормальным распределением с параметрами $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$ (класс нормальных распределений замкнут относительно операции свертки).*

Пример 13.2.2. Пусть F и Q — пуассоновские распределения соответственно с параметрами λ_1 и λ_2 . Найдите свертку $F * Q$ распределений Q и F .

Решение. Характеристическими функциями пуассоновских распределений соответственно с параметрами λ_1 и λ_2 являются

$$\varphi_1(t) = \exp\{\lambda_1(\exp\{it\} - 1)\} \text{ и } \varphi_2(t) = \exp\{\lambda_2(\exp\{it\} - 1)\}$$

(см. пример 13.1.1), поэтому характеристическая функция $\varphi(t)$ свертки $F * Q$, согласно мультипликативному свойству, равна $\varphi_1(t)\varphi_2(t)$, т. е.

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = \exp\{(\lambda_1 + \lambda_2)(\exp\{it\} - 1)\}.$$

Но $\exp\{(\lambda_1 + \lambda_2)(\exp\{it\} - 1)\}$ — характеристическая функция пуассоновского распределения с параметром $\lambda_1 + \lambda_2$. Поэтому согласно теореме единственности распределение $F * Q$ совпадает с пуассоновским распределением с параметром $\lambda_1 + \lambda_2$.

Следовательно, свертка пуассоновских распределений с параметрами λ_1 и λ_2 является пуассоновским распределением с параметром $\lambda_1 + \lambda_2$ (класс пуассоновских распределений замкнут относительно операции свертки).

Пример 13.2.3. Пусть случайная величина ξ_{n,p_n} распределена биномиально с параметрами $(n; p_n)$. Доказать, что если $np_n \rightarrow \lambda$ при $n \rightarrow \infty$, то распределение случайной величины ξ_{n,p_n} сходится к пуассоновскому распределению с параметром λ (теорема Пуассона).

Решение. Вычислим характеристическую функцию случайной величины ξ_{n,p_n} и покажем, что при $n \rightarrow \infty$, она сходится к характеристической функции пуассоновского распределения с параметром λ .

Характеристическая функция $\varphi_n(t)$ биномиально распределенной с параметрами $(n; p_n)$ случайной величины ξ_{n,p_n} равна $(1 + p_n(e^{it} - 1))^n$, т. е.

$$\varphi_n(t) = (1 + p_n(e^{it} - 1))^n$$

(см. решение задачи 13.7). Далее, поскольку $np_n \rightarrow \lambda$, при $n \rightarrow \infty$, то $p_n \rightarrow 0$ и для $\ln \varphi_n(t) = n \ln(1 + p_n(e^{it} - 1))$

при каждом фиксированном t имеем

$$\ln \varphi_n(t) \sim np_n (e^{it} - 1), \quad \text{при } n \rightarrow \infty.$$

Отсюда,

$$\begin{aligned} \ln \varphi_n(t) &\rightarrow \lambda (e^{it} - 1), \\ \varphi_n(t) &\rightarrow \exp\{\lambda(e^{it} - 1)\}. \end{aligned}$$

Но $\exp\{\lambda(e^{it} - 1)\}$ — характеристическая функция пуассоновского распределения с параметром λ , поэтому согласно теореме непрерывности распределение случайной величины ξ_{n,p_n} сходится к пуассоновскому распределению с параметром λ .

13.3 Задачи

АЗ: 13.1, 13.2, 13.5, 13.6, 13.7, 13.9, 13.10, 13.16, 13.17, 13.27.

СЗ: 13.2, 13.4, 13.8, 13.14, 13.18, 13.19, 13.20, 13.21, 13.23, 13.25.

13.1. Пусть случайная величина ξ принимает значения 1 и -1 , каждое с вероятностью $1/2$. Вычислить характеристическую функцию случайной величины ξ .

13.2. Пусть случайная величина ξ принимает значения $-1, 0, 1$, каждое с вероятностью $1/3$. Вычислить характеристическую функцию ξ .

13.3. Доказать, что $\varphi(z) = \cos^2 z$ является характеристической функцией; найти соответствующее распределение вероятностей.

13.4. Доказать, что

$$\varphi_1(z) = \sum_{k=0}^{\infty} a_k \cos kz, \quad \varphi_2(z) = \sum_{k=0}^{\infty} a_k e^{i\lambda_k z}$$

($a_k \geq 0$, $\sum_{k=0}^{\infty} a_k = 1$) являются характеристическими функциями. Найти соответствующие распределения вероятностей.

13.5. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых принимает значения 1 и -1 с вероятностью $1/2$. Вычислить характеристическую функцию случайной величины

$$S_n = \xi_1 + \xi_2 + \dots + \xi_n.$$

13.6. Доказать, что $\varphi(z) = \cos^n z$ является характеристической функцией (для каждого натурального n).

13.7. Вычислить характеристическую функцию случайной величины ξ , распределенной биномиально с параметрами $(n; p)$:

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

13.8. Вычислить характеристическую функцию случайной величины ξ , геометрически распределенной с параметром p :

$$P\{\xi = k\} = p(1-p)^k, \quad k = 0, 1, \dots$$

13.9. Пусть ξ_λ — случайная величина, имеющая распределение Пуассона с параметром λ .

Вычислить характеристическую функцию случайной величины $(\xi_\lambda - \lambda)/\sqrt{\lambda}$.

13.10. Найти характеристическую функцию случайной величины, равномерно распределенной на промежутке $[-a; a]$.

13.11. Вычислить характеристическую функцию случайной величины, равномерно распределенной на промежутке $[a; b]$.

13.12. Вычислить характеристическую функцию случайной величины с плотностью $p(x) = e^{-|x|}/2$.

13.13. Вычислить характеристическую функцию показательного распределения с параметром a .

13.14. Вычислить характеристическую функцию двустороннего показательного распределения с параметром a . Плотность двустороннего показательного распределения

$$f(x) = \frac{a}{2} e^{-a|x|}, \quad a > 0.$$

13.15. Вычислить характеристическую функцию треугольного распределения с параметром a . Плотность треугольного распределения

$$p(x) = \begin{cases} 0, & \text{если } |x| \geq a; \\ (a - |x|)/a^2, & \text{если } |x| < a. \end{cases}$$

13.16*. Вычислить характеристическую функцию распределения Коши с параметром a . Плотность распределения Коши

$$p(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}, \quad x \in \mathbb{R}^1.$$

13.17. Пусть F и Q — распределения Коши соответственно с параметрами a и b . Найти свертку $F * Q$ распределений Q и F .

13.18. Пусть ξ и η — независимые нормально распределенные случайные величины соответственно с параметрами $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$. Найти распределение случайной величины $\xi + \eta$.

13.19. Пусть ξ и η — независимые случайные величины, имеющие распределения Коши с параметрами a и b соответственно. Найти распределение случайной величины $\xi + \eta$.

13.20. Пусть ξ и η — независимые случайные величины, имеющие пуассоновские распределения соответственно с параметрами λ_1 и λ_2 . Найти распределение случайной величины $\zeta = \xi + \eta$.

13.21. Пусть $\varphi(t)$ — характеристическая функция случайной величины ξ , для которой $M\xi = a$ и $D\xi = \sigma^2 < \infty$. Выписать разложение Тейлора для $\varphi(t)$ в окрестности точки $t = 0$, содержащее t^2 .

13.22. Пусть $\{\xi_n\}$ — последовательность нормально распределенных случайных величин соответственно с параметрами $(a_n; \sigma_n^2)$, причем $a_n \rightarrow a$, $\sigma_n^2 \rightarrow 0$, при $n \rightarrow \infty$.

Доказать, что при $n \rightarrow \infty$, распределение случайной величины ξ_n сходится к собственному атомическому распределению, сосредоточенному в точке a .

13.23. Пусть ξ_λ — пуассоновская случайная величина с параметром $\lambda < \infty$.

Доказать, что если $\lambda \rightarrow \infty$, распределение случайной величины $(\xi_\lambda - \lambda)/\sqrt{\lambda}$ сходится к нормальному распределению с параметрами $(0; 1)$.

13.24. В каждой точке $x_k = k/n, k = 0, 1, \dots, n$, отрезка $[0; 1]$ сосредоточена масса $1/(n+1)$. Обозначим через F_n полученное таким образом вероятностное распределение.

1° Найти характеристическую функцию $\varphi_n(t)$ распределения F_n .

2° Вычислить предел $\varphi_n(t)$, при $n \rightarrow \infty$.

3° Доказать, что при $n \rightarrow \infty$, последовательность $\{F_n\}$ сходится к равномерному на промежутке $[0; 1]$ распределению.

13.25 (центральная предельная теорема). Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин со средним 0 и дисперсией σ^2 ($D\xi_k = \sigma^2 < \infty$). Доказать, что при $n \rightarrow \infty$, распределение случайной величины

$$S_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n \xi_i$$

сходится к нормальному распределению с параметрами $(0; 1)$.

Указание. Обозначим через $\varphi(t)$ характеристическую функцию случайной величины ξ_i ($i = 1, 2, \dots$), а через $\psi_n(t)$ — случайной величины $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n \xi_i$.

Последовательно решите такие задачи:

1° Выписать характеристическую функцию $\psi_n(t)$ через характеристическую функцию $\varphi(t)$.

2° Выписать разложение Тейлора для $\varphi(t)$ в окрестности точки $t = 0$, содержащее t^2 .

3° Воспользовавшись результатами пунктов 1° и 2°, найти предел характеристической функции $\psi_n(t)$, при $n \rightarrow \infty$.

4° Воспользовавшись результатом пункта 3° и теоремой непрерывности, доказать, что при $n \rightarrow \infty$, распре-

деление случайной величины $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n \xi_i$ сходится к нормальному распределению с параметрами $(0; 1)$.

13.26. Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин с конечными дисперсиями $D\xi_k = \sigma^2 < \infty$. Тогда из неравенства Чебышёва

$$P\{|\zeta - M\zeta| \geq \varepsilon\} \leq D\zeta/\varepsilon^2$$

следует, что при $n \rightarrow \infty$, последовательность $S_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ сходится по вероятности к $a = M\xi_k$, т. е. для любого $\varepsilon > 0$

$$P\{|S_n - a| \geq \varepsilon\} \rightarrow 0.$$

Не будем требовать, чтобы $\sigma^2 < \infty$, но пусть $M\xi_k = a \neq \infty$. Доказать, что при $n \rightarrow \infty$, последовательность S_n сходится по вероятности к a (закон больших чисел в форме Хинчина).

13.27. Пусть $\xi_1, \xi_2, \dots$ — независимые одинаково распределенные случайные величины, каждая из которых имеет распределение Коши с параметром a .

Доказать, что $S_n = \frac{1}{n} \sum_{i=1}^n \xi_i$ не сходится по вероятности к константе (ни к одной).

У к а з а н и е.

1° Найти характеристическую функцию S_n .

2° Вычислить при $n \rightarrow \infty$ предел характеристической функции случайной величины S_n .

Заметим, что у случайной величины, имеющей распределение Коши, математическое ожидание не существует.

13.28 (центральная предельная теорема). Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин со средним a ($M\xi_i = a \neq \infty$) и дисперсией σ^2 ($D\xi_i = \sigma^2 < \infty$). Доказать, что при $n \rightarrow \infty$ распределение случайной величины

$$S_n = \frac{1}{\sigma\sqrt{n}} \left(\sum_{i=1}^n \xi_i - na \right)$$

сходится к нормальному с параметрами $(0; 1)$.

У к а з а н и е. См. задачу 13.25.

13.29. Пусть $\xi_{1,n}, \xi_{2,n}, \dots, \xi_{k,n}, \dots, \xi_{n,n}$, — независимые одинаково распределенные случайные величины, каждая с распределением

$$P\{\xi_{k,n} = j\} = p_n^j (1 - p_n)^{1-j}, \quad j = 0, 1,$$

$k = 1, 2, \dots, n$, причем

$$np_n \rightarrow \lambda,$$

при $n \rightarrow \infty$.

Доказать, что если $n \rightarrow \infty$ распределение суммы

$\sum_{k=1}^n \xi_{k,n}$ сходится к пуассоновскому распределению с параметром λ :

$$P\left\{\sum_{k=1}^n \xi_{k,n} = m\right\} \rightarrow \frac{\lambda^m}{m!} e^{-\lambda}, \quad m = 0, 1, \dots$$

13.30. Пусть ν_i — случайная величина, определенная в задаче 10.41.

Доказать, что при $n \rightarrow \infty$, распределение

$$\frac{\nu_i - np_i}{\sqrt{np_i(1 - p_i)}}$$

сходится к нормальному распределению с параметрами $(0; 1)$.

13.31. Пусть $N_{(-1)^n; 1/n^2}$ — последовательность нормальных распределений с параметрами $((-1)^n; 1/n^2)$.

Исследовать последовательность вероятностных распределений $N_{(-1)^n; 1/n^2}$ на сходимость при $n \rightarrow \infty$.

Приведенные далее задачи можно решить, выписав характеристические функции соответствующих распределений и исследовав их сходимость при $n \rightarrow \infty$.

13.32. Пусть $\{F_n\}$ — последовательность вероятностных распределений с функциями распределения

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ nx, & \text{если } 0 < x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Доказать, что при $n \rightarrow \infty$, последовательность распределений $\{F_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

13.33. $\{F_n\}$ — последовательность вероятностных распределений с функциями распределения

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq -1/n; \\ n(x + 1/n)/2, & \text{если } -1/n < x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Доказать, что при $n \rightarrow \infty$ последовательность распределений $\{F_n\}$ сходится к атомическому распределению, сосредоточенному в точке 0.

13.34. $\{F_n\}$ — последовательность вероятностных распределений с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [-1/n, 1/n]; \\ 0, & \text{если } x \notin [-1/n, 1/n]. \end{cases}$$

Доказать, что при $n \rightarrow \infty$ последовательность распределений $\{F_n\}$ сходится к атомическому распределению, сосредоточенному в точке 0.

13.35. F_n — последовательность вероятностных распределений с функциями распределения

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq -1/n; \\ n(x + 1/n)/2, & \text{если } -1/n < x \leq 1/n; \\ 1, & \text{если } x > 1/n. \end{cases}$$

Доказать, что при $n \rightarrow \infty$ последовательность распределений F_n сходится к атомическому распределению, сосредоточенному в точке 0.

13.36. F_n — последовательность вероятностных распределений с плотностями

$$f_n(x) = \begin{cases} n/2, & \text{если } x \in [-1/n, 1/n]; \\ 0, & \text{если } x \notin [-1/n, 1/n]. \end{cases}$$

Доказать, что при $n \rightarrow \infty$ последовательность распределений F_n сходится к атомическому распределению, сосредоточенному в точке 0.

13.37. Пусть ξ_n , $n = 1, 2, \dots$, — последовательность нормально распределенных случайных величин соответственно с параметрами $(a; \sigma_n^2)$, для которых $\sigma_n^2 \rightarrow 0$, при $n \rightarrow \infty$.

Доказать, что распределение случайной величины ξ_n сходится к собственному атомическому распределению, сосредоточенному в точке a .

У к а з а н и е. Исследовать сходимость характеристических функций нормального распределения с параметрами $(a; \sigma_n^2)$ при $\sigma_n^2 \rightarrow 0$.

Глава 14

Оценивание параметров распределений

14.1 Точечные оценки

Типичная задача теории вероятностей — по математической модели стохастического эксперимента вычислить вероятности тех или иных событий и тем самым спрогнозировать те или иные его исходы.

Типичная задача математической статистики — по исходу стохастического эксперимента предложить (выбрать) математическую модель, которая бы адекватно его описывала. Во многих случаях решение последней задачи сводится к оцениванию параметров распределений и проверке статистических гипотез.

О п р е д е л е н и е. Случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ со значениями в пространстве $\mathbb{R}^n$ будем называть *выборкой* (*выборочным вектором*).

Выборку $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, образованную последовательностью независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n$, каждая из которых имеет распределение G , называют *выборкой из распределения (закона) G объемом n* .

Значение $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ будем называть ее реализацией.

Множество X всех возможных значений (реализаций) выборки будем называть *выборочным пространством* (выборочное пространство X — это $\mathbb{R}^n$ или его подмножество).

Мы будем иметь дело с выборками, распределения (функции распределения) которых зависят от параметра θ . В качестве множества возможных значений θ параметра θ мы, как правило, будем рассматривать подмножества конечномерного пространства R^s . (Распределениями, зависящими от параметров, например, являются нормальное, пуассоновское, показательное, равномерное, ...)

Постановка задачи оценивания параметров распределений. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ с распределением $F(\cdot; \theta)$. Распределение $F(\cdot; \theta)$ зависит от параметра θ , принимающего значения из множества Θ . Значение параметра θ неизвестно и его необходимо оценить (определить) по реализации $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$. В этом и состоит задача оценивания параметров распределений.

Единственное, что нам известно для оценивания неизвестного параметра θ , — это реализация $\xi(\omega)$ выборки ξ . Кроме реализации $\xi(\omega)$ выборки, мы не имеем ничего, что бы несло информацию о значении параметра θ . Поэтому “оценить (определить) θ по реализации $\xi(\omega)$ (точно или хотя бы приближенно)” обозначает поставить в соответствие реализации $\xi(\omega)$ выборки ξ значение $\hat{\theta}$ из Θ , которое мы будем использовать в качестве неизвестного параметра θ . Формально это обозначает, что для оценивания θ на выборочном пространстве X — множестве реализаций выборок — необходимо определить (построить, задать) функцию $h(\cdot)$ со значениями в Θ — множестве возможных значений параметра θ — такую, что $h(\xi(\omega))$ равно θ или $h(\xi(\omega))$ хотя бы приближенно равно θ . Значение

$$\hat{\theta} = h(\xi(\omega))$$

мы и будем использовать в качестве θ . Заметим, что для каждой реализации $\xi(\omega)$ значение $\hat{\theta} = h(\xi(\omega))$, свое.

Поэтому $\hat{\theta}$ как функция $\xi = \xi(\omega)$ является случайной величиной.

Определение. Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка со значениями в $\mathbb{R}^n$ и $F(\cdot, \theta)$ — ее распределение ($\theta \in \Theta \subset \mathbb{R}^n$). Борелевскую функцию $h(\cdot)$, заданную на выборочном пространстве $X \subset \mathbb{R}^n$, со значениями в Θ — множестве возможных значений параметра θ — будем называть *статистикой*, а случайную величину

$$\hat{\theta} = h(\xi) = h(\xi_1, \xi_2, \dots, \xi_n),$$

— борелевскую функцию от выборки — *оценкой*.

Строить статистики $h(\cdot)$, такие, что

$$\theta = h(\xi(\omega)) = \hat{\theta},$$

т.е. статистики, с помощью которых по $\xi(\omega)$ можно было бы точно определить θ , явно не удастся уже хотя бы потому, что θ является константой, а оценка $\hat{\theta} = h(\xi(\omega))$ как функция выборки (случайной величины) является случайной величиной. Поэтому нравится нам это или нет, для определения θ нам придется довольствоваться оценками $\hat{\theta} = h(\xi)$ как приближенными значениями θ .

Заметим, что для одного и того же параметра θ можно предложить много оценок. Например, для оценивания параметра θ равномерного на промежутке $[\theta - 1, \theta + 1]$ распределения по выборке $\xi_1, \xi_2, \dots, \xi_n$ из него можно предложить оценки

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad \hat{\theta}_2 = \frac{\xi_1 + \xi_n}{2}; \quad \hat{\theta}_3 = \frac{1}{2}(\max \{\xi_i\} + \min \{\xi_i\})$$

(при этом естественно возникают вопросы: какая из предложенных оценок лучше? Как построить наилучшую оценку? ...).

Погрешности оценивания параметров. В связи с постановкой задачи оценивания параметров распределений как задачи получения приближенных значений $\hat{\theta} = h(\xi)$ параметра θ необходимо уметь отвечать на вопрос: *насколько велика погрешность $\hat{\theta} - \theta$ при замене θ*

на $\hat{\theta}$, другими словами, как далеко могут уклоняться значения оценки

$$\hat{\theta} = h(\xi_1, \xi_2, \dots, \xi_n),$$

вычисленной по выборке $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, от оцениваемой величины θ ?

От оценки $\hat{\theta} = h(\xi)$, предлагаемой для оценивания того или иного параметра, естественно требовать малого разброса ее значений, другими словами, концентрации их в узкой окрестности. В качестве количественной меры разброса значений случайной величины $\hat{\theta} = h(\xi)$ будем рассматривать

$$D\hat{\theta} = M(\hat{\theta} - M\hat{\theta})^2$$

(для наглядности θ — одномерный параметр).

Количественно меру погрешности при замене θ на $\hat{\theta}$ (меру разброса $\hat{\theta}$ относительно θ) будем описывать величиной

$$M|\hat{\theta} - \theta|^2.$$

Среди всех оценок с одной и той же дисперсией $D\hat{\theta}$ (мерой разброса) минимальную меру разброса относительно θ имеют оценки, для которых $M\hat{\theta} = \theta$. Последнее следует из равенств

$$\begin{aligned} M|\hat{\theta} - \theta|^2 &= M|(\hat{\theta} - M\hat{\theta}) + (M\hat{\theta} - \theta)|^2 = M(\hat{\theta} - M\hat{\theta})^2 + \\ &+ 2M(\hat{\theta} - M\hat{\theta})(M\hat{\theta} - \theta) + M(M\hat{\theta} - \theta)^2 = D\hat{\theta} + (M\hat{\theta} - \theta)^2. \end{aligned}$$

О п р е д е л е н и е. Оценку $\hat{\theta}$ будем называть *несмещенной оценкой параметра θ* , если

$$M\hat{\theta} = \theta,$$

или, что то же, $M(\hat{\theta} - \theta) = 0$.

Наглядно несмещенность оценки $\hat{\theta}$ параметра θ можно трактовать так: при многократном использовании оценки $\hat{\theta}$ в качестве значения θ , т.е. при многократной замене θ на $\hat{\theta}$, среднее значение погрешности $\hat{\theta} - \theta$ равно нулю.

Часто рассматривают не одну оценку

$$\hat{\theta} = h(\xi) = h(\xi_1, \xi_2, \dots, \xi_n),$$

построенную по выборке $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, а последовательность оценок

$$\hat{\theta}_n = h_n(\xi_1, \xi_2, \dots, \xi_n), \quad n = 1, 2, \dots$$

В этой ситуации естественно говорить об асимптотическом поведении последовательности оценок.

О п р е д е л е н и е. Последовательность оценок $\{\hat{\theta}_n\}$ будем называть *состоятельной последовательностью оценок параметра θ* , если для каждого $\varepsilon > 0$

$$P\{|\hat{\theta}_n - \theta| > \varepsilon\} \rightarrow 0$$

при $n \rightarrow \infty$, или, что то же, $\{\hat{\theta}_n\}$ сходится по вероятности к θ при $n \rightarrow \infty$.

О п р е д е л е н и е. Последовательность оценок $\{\hat{\theta}_n\}$ будем называть *асимптотически несмещенной последовательностью оценок параметра θ* , если

$$M(\hat{\theta}_n - \theta) \rightarrow 0$$

при $n \rightarrow \infty$, или, что то же,

$$M\hat{\theta}_n \rightarrow \theta$$

при $n \rightarrow \infty$.

П р и м е р 14.1.1. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \min \{\xi_1, \xi_2, \dots, \xi_n\}, \quad \hat{\theta}_2 = \max \{\xi_1, \xi_2, \dots, \xi_n\},$$

$$\hat{\theta}_3 = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad \hat{\theta}_4 = \frac{\xi_{n-1} + \xi_n}{2}.$$

Какие из оценок $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ и для каких параметров (возможно, отличных от a и b) являются несмещенными оценками? Состоятельными оценками?

Решение. Рассмотрим каждую из оценок.

Оценка $\hat{\theta}_2 = \max \{\xi_1, \xi_2, \dots, \xi_n\}$. Найдем сначала распределение $\hat{\theta}_2$. По условию $\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного на отрезке $[a; b]$ распределения, т. е. $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая из которых распределена равномерно на отрезке $[a; b]$. Отсюда, учитывая, что распределения случайных величин $\xi_1, \xi_2, \dots, \xi_n$ абсолютно непрерывны с плотностями $f(x; a, b)$, имеем

$$\begin{aligned} F_{\hat{\theta}_2}(x) &= P\{\hat{\theta}_2 < x\} = P\{\max \{\xi_1, \xi_2, \dots, \xi_n\} < x\} = \\ &= P\{\xi_1 < x, \xi_2 < x, \dots, \xi_n < x\} = \prod_{i=1}^n P\{\xi_i < x\} = \\ &= \prod_{i=1}^n \int_{-\infty}^x f(y; a, b) dy = \left(\int_{-\infty}^x f(y; a, b) dy \right)^n. \end{aligned}$$

Поэтому $\hat{\theta}_2$ — абсолютно непрерывная случайная величина и ее плотность распределения

$$f_{\hat{\theta}_2}(x) = \frac{d}{dx} F_{\hat{\theta}_2}(x) = n f(x; a, b) \left(\int_{-\infty}^x f(y; a, b) dy \right)^{n-1}.$$

А поскольку

$$\int_{-\infty}^x f(y; a, b) dy = \begin{cases} 0, & \text{если } x < a; \\ \frac{x-a}{b-a}, & \text{если } a \leq x \leq b; \\ 1, & \text{если } x > b, \end{cases}$$

то

$$f_{\hat{\theta}_2}(x) = \begin{cases} 0, & \text{если } x \notin [a; b]; \\ \frac{n}{(b-a)^n}(x-a)^{n-1}, & \text{если } x \in [a; b]. \end{cases}$$

График плотности $f_{\hat{\theta}_2}(x)$ изображен на рис. 14.1.1.

По известной плотности распределения оценки $\hat{\theta}_2$ можем вычислить ее математическое ожидание:

$$\begin{aligned} M\hat{\theta}_2 &= \int_{-\infty}^{\infty} x f_{\hat{\theta}_2}(x) dx = \\ &= \int_a^b x \frac{n}{(b-a)^n} (x-a)^{n-1} dx = \frac{n}{n+1}b + \frac{a}{n+1}. \end{aligned}$$

Следовательно, оценка $\hat{\theta}_2$ не является несмещенной оценкой ни параметра a , ни параметра b ; но, при $n \rightarrow \infty$,

$$M\hat{\theta}_2 = \frac{n}{n+1}b + \frac{a}{n+1} \rightarrow b.$$

Последнее обозначает, что $\hat{\theta}_2$ является асимптотически несмещенной оценкой параметра b .

Выясним, является ли $\hat{\theta}_2$ состоятельной оценкой параметра b , другими словами, сходится ли $\hat{\theta}_2$ по вероятности к b . Для достаточно малых ε имеем

$$\begin{aligned} P\{|\hat{\theta}_2 - b| > \varepsilon\} &= P\{\hat{\theta}_2 \in (-\infty, b - \varepsilon) \cup (b + \varepsilon, +\infty)\} = \\ &= \int_{-\infty}^{b-\varepsilon} f_{\hat{\theta}_2}(x) dx + \int_{b+\varepsilon}^{+\infty} f_{\hat{\theta}_2}(x) dx = \int_a^{b-\varepsilon} f_{\hat{\theta}_2}(x) dx = \\ &= \int_a^{b-\varepsilon} \frac{n}{(b-a)^n} (x-a)^{n-1} dx = \left(1 - \frac{\varepsilon}{b-a}\right)^n, \end{aligned}$$

что сходится к нулю, при $n \rightarrow \infty$. Следовательно,

$$P\{|\hat{\theta}_2 - b| > \varepsilon\} \rightarrow 0,$$

при $n \rightarrow \infty$. Последнее обозначает, что $\hat{\theta}_2$ является состоятельной оценкой параметра b .

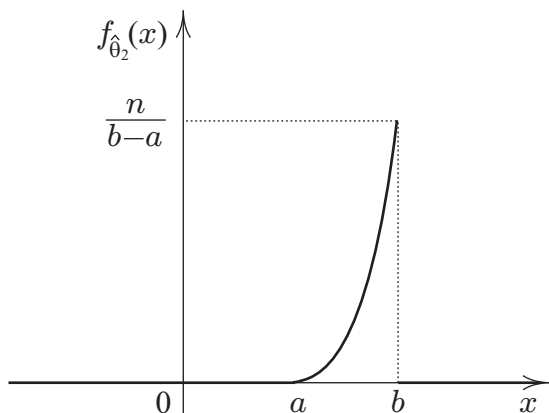


Рис. 14.1.1: Плотность распределения оценки $\hat{\theta}_2$

Оценка $\hat{\theta}_1 = \min\{\xi_1, \xi_2, \dots, \xi_n\}$. Оценка $\hat{\theta}_1$ исследуется аналогично оценке $\hat{\theta}_2$.

Распределение оценки $\hat{\theta}_1$ абсолютно непрерывно с плотностью

$$f_{\hat{\theta}_1}(x) = \begin{cases} 0, & \text{если } x \notin [a; b]; \\ \frac{n}{(b-a)^n}(b-x)^{n-1}, & \text{если } x \in [a; b]. \end{cases}$$

Отсюда

$$M\hat{\theta}_1 = \frac{n}{n+1}a + \frac{b}{n+1}.$$

Поэтому $\hat{\theta}_1$ не является несмещенной оценкой ни параметра a , ни параметра b , но она является асимптотически несмещенной оценкой параметра a .

Оценка $\hat{\theta}_1$ является состоятельной оценкой параметра a (см. исследование оценки $\hat{\theta}_2$).

Оценка $\hat{\theta}_3 = \frac{1}{n} \sum_{i=1}^n \xi_i$. Поскольку ξ_i , $i = 1, 2, \dots, n$, — равномерно распределенные на отрезке $[a; b]$ случайные величины, а для них математическое ожидание m_1 , как известно, равно $(a + b)/2$, то

$$M\hat{\theta}_3 = M\frac{1}{n} \sum_{i=1}^n \xi_i = m_1 = \frac{a + b}{2}.$$

Следовательно, $\hat{\theta}_3$ не является несмещенной оценкой ни параметра a , ни параметра b , но $\hat{\theta}_3$ — несмещенная оценка параметра $m_1 = (a + b)/2$ — математического ожидания равномерного на отрезке $[a; b]$ распределения.

Далее, согласно закону больших чисел $\hat{\theta}_3 = \frac{1}{n} \sum_{i=1}^n \xi_i$ сходится по вероятности к $m_1 = (a + b)/2$, т. е. $\hat{\theta}_3$ — состоятельная оценка параметра $m_1 = (a + b)/2$.

Оценка $\hat{\theta}_4 = (\xi_{n-1} + \xi_n)/2$. Для оценки $\hat{\theta}_4$

$$M\hat{\theta}_4 = M\frac{\xi_{n-1} + \xi_n}{2} = \frac{1}{2}(M\xi_{n-1} + M\xi_n) = \frac{a + b}{2}.$$

Следовательно, $\hat{\theta}_4$ — несмещенная оценка параметра $m_1 = (a + b)/2$.

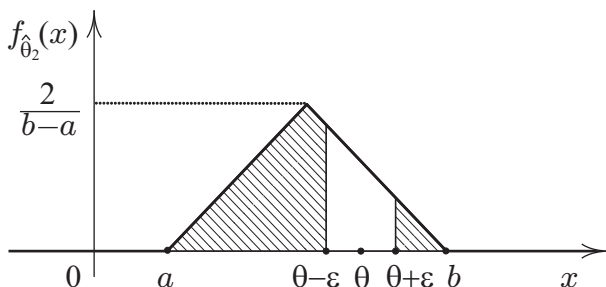


Рис. 14.1.2: Плотность распределения оценки $\hat{\theta}_4$

Оценка $\hat{\theta}_4$ не сходится по вероятности ни к одной константе (параметру). В самом деле, плотностью распределения случайной величины $\hat{\theta}_4 = (\xi_{n-1} + \xi_n)/2$ является функция

$$f_{\hat{\theta}_4}(x) = \begin{cases} 0, & \text{если } x \notin [a; b]; \\ \frac{4}{(b-a)^2}(x-a), & \text{если } x \in [a, \frac{a+b}{2}]; \\ \frac{4}{(b-a)^2}(b-x), & \text{если } x \in [\frac{a+b}{2}, b] \end{cases}$$

($f_{\hat{\theta}_4}(x)$ получена как свертка плотностей равномерных на отрезке $[a; b]$ распределений, график $f_{\hat{\theta}_4}(x)$ изображен на рис. 14.1.2). Поэтому для любого параметра θ , каждого достаточно малого $\varepsilon > 0$ вероятность

$$P\{|\hat{\theta}_4 - \theta| > \varepsilon\} = \int_{\{x: |x-\theta| > \varepsilon\}} f_{\hat{\theta}_4}(x) dx$$

не зависит от n (численно она равна площади заштрихованной фигуры (см. рис. 14.1.2)) и, следовательно, $\hat{\theta}_4$ не сходится к θ по вероятности при $n \rightarrow \infty$.

14.2 Доверительные интервалы

Помимо точечного оценивания параметров распределения, используют так называемое *интервальное оценивание*, суть которого состоит в том, что для неизвестного параметра находят интервал, содержащий с заданной вероятностью неизвестный параметр. Рассмотрим доверительные интервалы для параметров нормального распределения.

Доверительный интервал для a . Для выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из распределения $N_{a; \sigma^2}$ случайная величина $(\bar{\xi} - a) / \frac{s}{\sqrt{n}}$ имеет t -распределение с $n - 1$ степе-

ниями свободы. Поэтому

$$P \left\{ \left| \bar{\xi} - a \right| / \frac{s}{\sqrt{n}} < t_{\alpha; n-1} \right\} = 1 - 2\alpha$$

или

$$P \left\{ \bar{\xi} - \frac{s}{\sqrt{n}} t_{\alpha; n-1} < a < \bar{\xi} + \frac{s}{\sqrt{n}} t_{\alpha; n-1} \right\} = 1 - 2\alpha,$$

где $t_{\alpha; m}$ — верхняя α -граница t -распределения с m степенями свободы (см. гл. 22 и табл. 22.3.1). Другими словами,

$$\left(\bar{\xi} - \frac{s}{\sqrt{n}} t_{\alpha; n-1}, \bar{\xi} + \frac{s}{\sqrt{n}} t_{\alpha; n-1} \right) \quad (14.2.1)$$

содержит неизвестный параметр a с вероятностью $1 - 2\alpha$. Интервал (14.2.1) называют *доверительным интервалом* для параметра a с коэффициентом надежности $1 - 2\alpha$ (координата $\bar{\xi}$ середины доверительного интервала и его длина $2 \frac{s}{\sqrt{n}} t_{\alpha; n-1}$ — случайные величины). Как правило, для коэффициента надежности выбирают значения 0,99; 0,98; 0,95; 0,90.

Коэффициенту надежности $1 - 2\alpha$ можно дать следующую частотную интерпретацию: если провести достаточно длинную серию независимых экспериментов, скажем 100, то доля тех из них, в которых доверительный интервал (14.2.1) содержит неизвестный параметр a , близка к $100(1 - 2\alpha)$.

Доверительный интервал для σ^2 . Если $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из распределения $N_{a; \sigma^2}$, то случайная величина $(n - 1)s^2/\sigma^2$ имеет распределение χ^2 с $n - 1$ степенями свободы. Поэтому

$$P \left\{ \chi_{1-\alpha; n-1}^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_{\alpha; n-1}^2 \right\} = 1 - 2\alpha$$

или

$$P \left\{ \frac{(n-1)s^2}{\chi_{\alpha; n-1}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{1-\alpha; n-1}^2} \right\} = 1 - 2\alpha,$$

($\chi^2_{\beta;m}$ — верхний β -предел χ^2 -распределения с m степенями свободы, см. гл. 22 и табл. 22.2.1). То есть интервал

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha;n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}} \right) \quad (14.2.2)$$

содержит неизвестный параметр σ^2 с вероятностью $1 - 2\alpha$. Другими словами, интервал (14.2.2) является доверительным интервалом для параметра σ^2 (дисперсии нормального распределения) с коэффициентом надежности $1 - 2\alpha$.

Пример 14.2.1. *Найти доверительный интервал для показаний микроскопа I (см. пример 17.2.1) с коэффициентом надежности: 1) 0,90; 2) 0,99.*

Решение. Показания микроскопа естественно считать реализацией выборки из нормального распределения. Доверительный интервал для параметра a (математического ожидания) нормального распределения имеет вид (14.2.1). В рассматриваемом примере $n = 12$, $\bar{x} = 2,34$, $s^2 = 1,66$, $t_{0,05;11} = 1,796$, $t_{0,005;11} = 3,106$. Поэтому доверительным интервалом для параметра a с коэффициентом надежности 0,90 является интервал $(1,68; 3)$, а доверительным интервалом с коэффициентом надежности 0,99 — интервал $(1,19; 3,49)$ (фактически найдены реализации доверительных интервалов, хотя мы говорим о доверительных интервалах).

Обратим внимание на то, что за увеличение коэффициента надежности мы “расплачиваемся” расширением доверительного интервала, и наоборот — чем уже доверительный интервал, тем меньше коэффициент надежности.

Пример 14.2.2. *Найти доверительный интервал для дисперсии разности урожаев (см. пример 17.1.1) с коэффициентом надежности: 1) 0,90; 2) 0,98.*

Решение. Разность урожаев естественно считать реализацией выборки из нормального распределения. Доверительный интервал для параметра σ^2 (дисперсии) нормального распределения имеет вид (14.2.2). В рассматриваемом примере $n = 10$, $s^2 = 0,667$; $\chi^2_{0,95;9} = 3,32$,

$\chi^2_{0,05;9} = 16,92$; $\chi^2_{0,99;9} = 2,9$, $\chi^2_{0,01;9} = 21,67$. Поэтому доверительным интервалом для параметра σ^2 с коэффициентом надежности 0,90 (точнее, реализацией доверительного интервала, построенного по выборке) является интервал $(0,35; 1,81)$, а с коэффициентом надежности 0,99 — интервал $(0,28; 2,07)$.

14.3 Оценки с минимальной дисперсией

Основной вопрос задачи оценивания параметров распределений — насколько велика погрешность при замене параметра θ оценкой $\hat{\theta}$.

Оценки $\hat{\theta}$, предлагаемые для оценивания параметра θ , должны быть несмещенными, т. е. $M\hat{\theta} = \theta$. Такие оценки имеют меньшую меру разброса относительно θ по сравнению с оценками, для которых $M\hat{\theta} \neq \theta$. Для оценивания параметра θ можно предложить много несмещенных оценок. Из семейства таких оценок естественно выбрать те, которые имеют минимально возможную меру разброса (дисперсию).

О п р е д е л е н и е. Несмещенную оценку θ^* параметра θ будем называть его *наилучшей оценкой, оценкой минимальной дисперсии или эффективной оценкой*, если

$$D\theta^* = \inf_{\hat{\theta}: M\hat{\theta}=\theta} D\hat{\theta}.$$

В связи с этим определением естественно возникает вопрос: насколько малой может быть минимально возможная дисперсия оценки (насколько малыми могут быть отклонения $\hat{\theta}$ от θ)? Оказывается, что если семейство распределений $F(\cdot; \theta)$, $\theta \in \Theta$, выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ достаточно регулярно зависит от оцениваемого параметра θ , то можно указать нижнюю границу дисперсий всех несмещенных оценок параметра (неравенство Крамера—Рао). В некоторых случаях существуют оценки параметра, на которых нижняя граница достигается. Эти оценки

являются эффективными. Сравнивая дисперсию данной оценки с нижней границей дисперсий несмещенных оценок, можно выяснить, насколько оценка близка к наилучшей возможной.

Пусть выборка $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ фиксированного объема n имеет плотность распределения

$$f(x; \theta) = f(x_1, x_2, \dots, x_n; \theta).$$

Параметр θ будем считать одномерным, а относительно множества его возможных значений Θ предположим, что она является конечным интервалом числовой прямой.

В дальнейшем мы будем пользоваться приведенным далее утверждением.

Лемма 14.3.1. Если почти для всех $x \in \mathbb{R}^n$ существуют производные

$$\frac{\partial}{\partial \theta} f(x; \theta) \text{ и } \frac{\partial^2}{\partial \theta^2} f(x; \theta), \theta \in \Theta,$$

мажорируемые интегрируемыми функциями:

$$\left| \frac{\partial^i}{\partial \theta^i} f(x; \theta) \right| \leq u_i(x), \int_{\mathbb{R}^n} u_i(x) dx < \infty, i = 1, 2,$$

и выполняются условия

$$M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2 < \infty, M \left| \frac{\partial^2}{\partial \theta^2} \ln f(\xi; \theta) \right| < \infty, \theta \in \Theta,$$

то для всех $\theta \in \Theta$

$$M \frac{\partial}{\partial \theta} \ln f(\xi; \theta) = 0,$$

$$D \frac{\partial}{\partial \theta} \ln f(\xi; \theta) = M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2 = -M \frac{\partial^2}{\partial \theta^2} \ln f(\xi; \theta).$$

О п р е д е л е н и е. Функцию

$$I(\theta) = M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2$$

(когда она определена) называют *информацией по Фишера*.

В лемме 14.3.1 приведены достаточные условия, при которых информация $I(\theta)$ существует. Отметим, что

$$I(\theta) = M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2 = -M \frac{\partial^2}{\partial \theta^2} \ln f(\xi; \theta).$$

Т е о р е м а 14.3.1 (неравенство Крамера—Рао). Пусть выполняются условия леммы 14.3.1 и $\hat{\theta} = h(\xi)$ — несмещенная оценка параметра θ такая, что функция

$$h(x) \frac{\partial}{\partial \theta} f(x, \theta), \theta \in \Theta,$$

мажорируема интегрируемой функцией:

$$\left| h(x) \frac{\partial}{\partial \theta} f(x; \theta) \right| \leq U(x), \int_{\mathbb{R}^n} U(x) dx < \infty.$$

Тогда

$$D\hat{\theta} \geq \frac{1}{I(\theta)}, \quad (14.3.1)$$

причем равенство в (14.3.1) достигается тогда и только тогда, когда $\frac{\partial}{\partial \theta} \ln f(x; \theta)$ представима в виде

$$\frac{\partial}{\partial \theta} \ln f(x; \theta) = c(\theta)(h(x) - \theta).$$

С л е д с т в и е 1. Если оценка $\hat{\theta} = \theta^*$ удовлетворяет условиям теоремы и для нее неравенство Крамера—Рао

$$D\hat{\theta} \geq \frac{1}{I(\theta)}$$

обращается в равенство, то θ^* является эффективной оценкой параметра θ .

Следствие 2. Если оценка $\hat{\theta} = h(\xi)$ удовлетворяет условиям теоремы, а статистика $h(x)$ ($x \in \mathbb{R}^n$) — условию

$$\frac{\partial}{\partial \theta} \ln f(x; \theta) = c(\theta)(h(x) - \theta),$$

где $f(x; \theta)$ — плотность распределения выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, то $h(\xi)$ — несмещенная и эффективная оценка параметра θ .

Неравенство Крамера—Рао (распределение дискретное). Неравенство Крамера—Рао и утверждения, аналогичные приведенным выше, имеют место также тогда, когда распределение $P(\cdot; \theta)$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ дискретно, т. е. существует не более, чем счетное множество точек $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, для которых

$$\begin{aligned} P_\xi(x) &= P(x; \theta) = P(x_1, x_2, \dots, x_n; \theta) = \\ &= P(\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_n = x_n; \theta) > 0; \end{aligned}$$

$$\sum_{(x_1, x_2, \dots, x_n)} P(x_1, x_2, \dots, x_n; \theta) = 1.$$

Лемма 14.3.2 (распределение дискретное). Если для всех возможных значений $x = (x_1, x_2, \dots, x_n)$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ существуют производные

$$\frac{\partial}{\partial \theta} P(x; \theta) \quad \text{и} \quad \frac{\partial^2}{\partial \theta^2} P(x; \theta), \quad \theta \in \Theta,$$

ряды

$$\sum_x \frac{\partial}{\partial \theta} P(x; \theta) \quad \text{и} \quad \sum_x \frac{\partial^2}{\partial \theta^2} P(x; \theta)$$

сходятся абсолютно и равномерно относительно $\theta \in \Theta$ и выполняются условия

$$M \left(\frac{\partial}{\partial \theta} \ln P(\xi; \theta) \right)^2 < \infty, \quad M \left| \frac{\partial^2}{\partial \theta^2} \ln P(\xi; \theta) \right| < \infty, \quad \theta \in \Theta,$$

то для всех $\theta \in \Theta$

$$M \frac{\partial}{\partial \theta} \ln P(\xi; \theta) = 0,$$

$$D \frac{\partial}{\partial \theta} \ln P(\xi; \theta) = M \left(\frac{\partial}{\partial \theta} \ln P(\xi; \theta) \right)^2 = -M \frac{\partial^2}{\partial \theta^2} \ln P(\xi; \theta).$$

Теорема 14.3.2 (неравенство Крамера—Рао, распределение дискретное). Пусть выполняются условия леммы 14.3.2 и $\hat{\theta} = h(\xi)$ — несмещенная оценка параметра θ такая, что для всех возможных значений $x = (x_1, x_2, \dots, x_n)$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ ряд

$$\sum_x \frac{\partial}{\partial \theta} h(x) P(x; \theta)$$

сходится абсолютно и равномерно относительно $\theta \in \Theta$. Тогда

$$D\hat{\theta} \geq \frac{1}{I(\theta)},$$

причем равенство имеет место тогда и только тогда, когда $\frac{\partial}{\partial \theta} \ln P(x; \theta)$ представима в виде

$$\frac{\partial}{\partial \theta} \ln P(x; \theta) = c(\theta)(h(x) - \theta).$$

Следствие 1. Если оценка $\hat{\theta} = \theta^*$ удовлетворяет условиям теоремы и для нее неравенство Крамера—Рао

$$D\hat{\theta} \geq \frac{1}{I(\theta)}$$

обращается в равенство, то θ^* является эффективной оценкой параметра θ .

Следствие 2. Если оценка $\hat{\theta} = h(\theta)$ удовлетворяет условиям теоремы, а статистика $h(x)$ ($x \in \mathbb{R}^n$) — условию

$$\frac{\partial}{\partial \theta} \ln P(x; \theta) = c(\theta)(h(x) - \theta),$$

где $P(x; \theta)$ — распределение выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, то $h(\xi)$ — несмещенная и эффективная оценка параметра θ .

Пример 14.3.1 Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из нормального распределения с параметрами $(\theta; 1)$, θ принадлежит отрезку $[a; b]$. Является ли

$$\hat{\theta} = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

Решение. $\hat{\theta} = \bar{\xi}$ — несмещенная оценка математического ожидания θ . Чтобы выяснить, будет ли $\hat{\theta} = \bar{\xi}$ эффективной оценкой θ , воспользуемся неравенством Крамера—Рао (см. следствие 1 из теоремы 14.3.1): если

$$D\hat{\theta} = \frac{1}{I(\theta)},$$

то оценка $\hat{\theta}$ — эффективная.

Сначала убедимся, что для выборки из нормального распределения с параметрами $(\theta; 1)$ и функции

$$h(x) = \frac{1}{n} \sum_{k=1}^n x_k$$

выполняются условия теоремы 14.3.1.

Обозначим через $g(t; \theta)$ плотность нормального распределения с параметрами $(\theta; 1)$. Плотность

$$f(x; \theta) = \prod_{i=1}^n g(x_i; \theta)$$

распределения вектора $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ дважды дифференцируема по θ и

$$\frac{\partial}{\partial \theta} \ln f(x; \theta) = \frac{\partial}{\partial \theta} \ln \prod_{i=1}^n g(x_i; \theta) =$$

$$= \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln g(x_i; \theta) = \sum_{i=1}^n (x_i - \theta);$$

$$\frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) = \frac{\partial}{\partial \theta} \sum_{i=1}^n (x_i - \theta) = -n;$$

$$M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2 = M \left(\sum_{i=1}^n (\xi_i - \theta) \right)^2 = n;$$

$$M \left| \frac{\partial^2}{\partial \theta^2} \ln f(\xi; \theta) \right| = M | -n | = n.$$

Далее, функция $h(x)f(x; \theta)$ дифференцируема по θ , причем

$$\begin{aligned} \frac{\partial}{\partial \theta} h(x)f(x, \theta) &= \frac{\partial}{\partial \theta} h(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n; \theta) = \\ &= \frac{\partial}{\partial \theta} \bar{x} \prod_{j=1}^n g(x_j; \theta) = \sum_{i=1}^n \bar{x} \left(\frac{\partial}{\partial \theta} g(x_i; \theta) \right) \prod_{j=1, j \neq i}^n g(x_j; \theta) = \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^n x_k \left(\frac{\partial}{\partial \theta} g(x_i; \theta) \right) \prod_{j=1, j \neq i}^n g(x_j; \theta) = \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^n x_k (x_i - \theta) \prod_{j=1}^n g(x_j; \theta). \end{aligned}$$

Мы воспользовались тем, что

$$\begin{aligned} \frac{\partial}{\partial \theta} g(x_i; \theta) &= \frac{\partial}{\partial \theta} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x_i - \theta)^2}{2} \right\} = \\ &= (x_i - \theta) \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x_i - \theta)^2}{2} \right\} = (x_i - \theta) g(x_i; \theta). \end{aligned}$$

Чтобы убедиться в существовании интегрируемой мажоранты для функции $h(x) \frac{\partial}{\partial \theta} f(x; \theta)$, достаточно установить, что она существует для каждой из функций

$$x_k x_i \prod_{j=1}^n g(x_j; \theta), \quad k, i = 1, 2, \dots, n; \quad \theta \in [a; b].$$

Определим функцию $v(t)$:

$$v(t) = \begin{cases} g(t; b), & \text{если } t > b; \\ g(t; a), & \text{если } t < a; \\ c, & \text{если } a \leq t \leq b, \end{cases}$$

где $c = \max_{t, \theta \in [a; b]} g(t; \theta)$, $k = 1, 2, \dots, n$. Функция $|t|v(t)$ является интегрируемой мажорантой для функции $tg(t; \theta)$, $\theta \in [a; b]$. В последнем убеждаемся непосредственной проверкой. Поэтому функция

$$|x_k|v(x_k)|x_i|v(x_i) \prod_{j=1, j \neq i, j \neq k}^n v(x_j)$$

является интегрируемой мажорантой для функции

$$x_k x_i \prod_{j=1}^n g(x_j; \theta).$$

В этом легко убедиться, воспользовавшись теоремой Фубини.

Таким образом, условие теоремы 14.3.1 выполняется и, следовательно, неравенство Крамера—Рао имеет место:

$$D\hat{\theta} \geq \frac{1}{I(\theta)}.$$

А поскольку

$$I(\theta) = M \left(\frac{\partial}{\partial \theta} \ln f(\xi; \theta) \right)^2 = -M \left(\frac{\partial^2}{\partial \theta^2} \ln f(\xi; \theta) \right) = n;$$

$$D\hat{\theta} = D\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right) = \frac{1}{n^2}nD\xi_i = \frac{1}{n}1 = \frac{1}{n},$$

то неравенство Крамера—Рао обращается в равенство, что является достаточным условием эффективности оценки $\hat{\theta}$ параметра θ .

Другой способ решения примера состоит в проверке достаточного условия обращения неравенства Крамера—Рао в равенство, а именно:

$$\frac{\partial}{\partial\theta} \ln f(x; \theta) = c(\theta)(h(x) - \theta); \quad x \in \mathbb{R}^n.$$

Имеем

$$\begin{aligned} \frac{\partial}{\partial\theta} \ln f(x; \theta) &= \frac{\partial}{\partial\theta} \ln f(x_1, x_2, \dots, x_n; \theta) = \\ &= \frac{\partial}{\partial\theta} \ln \prod_{i=1}^n g(x_i; \theta) = \frac{\partial}{\partial\theta} \sum_{i=1}^n \ln g(x_i; \theta) = \\ &= \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x_i - \theta)^2}{2} \right\} = \\ &= \sum_{i=1}^n \frac{\partial}{\partial\theta} \left(-\frac{(x_i - \theta)^2}{2} \right) = \sum_{i=1}^n x_i - n\theta = \\ &= n \left(\frac{1}{n} \sum_{i=1}^n x_i - \theta \right) = n(h(x) - \theta). \end{aligned}$$

Отсюда следует, что $\hat{\theta} = \bar{\xi}$ — эффективная оценка параметра θ .

Пример 14.3.2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из пуассоновского распределения с параметром λ (λ принадлежит отрезку $[a; b]$, $a > 0$). Является ли $\hat{\lambda} = \bar{\xi}$ эффективной оценкой параметра λ ?

Решение. Отметим, что $\hat{\lambda} = \bar{\xi}$ является несмещенной оценкой λ . Далее воспользуемся неравенством Крамера—Рао. Убедимся, что для выборки из пуассоновского распределения и функции

$$h(x) = \frac{1}{n} \sum_{k=1}^n x_k$$

выполняются условия теоремы 14.3.2 (неравенство Крамера—Рао). Обозначим $\frac{\lambda^x}{x!} e^{-\lambda}$ через $g(x; \lambda)$, $x = 0, 1, \dots, n$. Распределением выборки $\xi_1, \xi_2, \dots, \xi_n$ является

$$P(x; \lambda) = P(x_1, x_2, \dots, x_n; \lambda) = \prod_{j=1}^n g(x_j; \lambda) = \prod_{j=1}^n \frac{\lambda^{x_j}}{x_j!} e^{-\lambda}.$$

Функция $P(x; \lambda)$ дважды дифференцируема по λ и

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln P(x; \lambda) &= \frac{\partial}{\partial \lambda} \ln \prod_{j=1}^n g(x_j; \lambda) = \\ &= \sum_{j=1}^n \frac{\partial}{\partial \lambda} \ln g(x_j; \lambda) = \frac{1}{\lambda} \sum_{j=1}^n (x_j - \lambda); \\ \frac{\partial^2}{\partial \lambda^2} \ln P(x; \lambda) &= \frac{\partial}{\partial \lambda} \left(\frac{1}{\lambda} \sum_{j=1}^n (x_j - \lambda) \right) = -\frac{1}{\lambda^2} \sum_{j=1}^n x_j. \end{aligned}$$

Функция

$$h(x_1, x_2, \dots, x_n) P(x_1, x_2, \dots, x_n; \lambda) = \bar{x} P(x_1, x_2, \dots, x_n; \lambda)$$

дифференцируема по λ , причем

$$\begin{aligned} h(x_1, x_2, \dots, x_n) \frac{\partial}{\partial \lambda} P(x_1, x_2, \dots, x_n; \lambda) &= \\ = \bar{x} \frac{\partial}{\partial \lambda} \prod_{j=1}^n g(x_j; \lambda) &= \bar{x} \sum_{i=1}^n \left(\frac{\partial}{\partial \lambda} g(x_i; \lambda) \right) \prod_{j=1, j \neq i}^n g(x_j; \lambda) = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n x_k \left(\frac{x_i}{\lambda} - 1 \right) \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \prod_{j=1, j \neq i}^n g(x_j; \lambda) = \\
&= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n x_k \left(\frac{x_i}{\lambda} - 1 \right) \prod_{j=1}^n g(x_j; \lambda),
\end{aligned}$$

ПОСКОЛЬКУ

$$\frac{\partial}{\partial \lambda} g(x_i; \lambda) = \frac{\partial}{\partial \lambda} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = \left(\frac{x_i}{\lambda} - 1 \right) g(x_i; \lambda).$$

Далее, при суммировании по $x = (x_1, x_2, \dots, x_n)$ величины x_i ($i = 1, 2, \dots, n$) принимают целые неотрицательные значения.

Ряд

$$\begin{aligned}
&\sum_x h(x) \frac{\partial}{\partial \lambda} P(x; \lambda) = \\
&= \sum_{(x_1, x_2, \dots, x_n)} \left(\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n x_k \left(\frac{x_i}{\lambda} - 1 \right) \prod_{j=1}^n g(x_j; \lambda) \right) = \\
&= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n \left(\sum_{(x_1, x_2, \dots, x_n)} x_k \left(\frac{x_i}{\lambda} - 1 \right) \prod_{j=1}^n g(x_j; \lambda) \right)
\end{aligned}$$

мажорируется рядом

$$\begin{aligned}
&\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n \left(\sum_{(x_1, x_2, \dots, x_n)} x_k \left| \frac{x_i}{\lambda} - 1 \right| \prod_{j=1}^n g(x_j; \lambda) \right) = \\
&= \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^n \left(\left(\sum_{x_1} g(x_1; \lambda) \right) \left(\sum_{x_2} g(x_2; \lambda) \right) \dots \right. \\
&\dots \left(\sum_{x_k} x_k g(x_k; \lambda) \right) \dots \left(\sum_{x_i} \left| \frac{x_i}{\lambda} - 1 \right| g(x_i; \lambda) \right) \dots
\end{aligned}$$

$$\dots \left(\sum_{x_n} g(x_n; \lambda) \right) \Bigg) \Bigg) .$$

Последний сходится абсолютно и равномерно относительно $\lambda \in [a; b]$, поскольку

$$\sum_x g(x; \lambda) = \sum_x \frac{\lambda^x}{x!} e^{-\lambda} \leq e^{-a} \sum_x \frac{b^x}{x!} = e^{b-a};$$

$$\begin{aligned} \sum_x x g(x; \lambda) &= \sum_x x \frac{\lambda^x}{x!} e^{-\lambda} \leq \\ &\leq e^{-a} \sum_x x \frac{b^x}{x!} \leq e^{b-a} \sum_x x \frac{b^x}{x!} e^{-b} = b e^{b-a}; \end{aligned}$$

$$\sum_x \left| \frac{x}{\lambda} - 1 \right| \frac{\lambda^x}{x!} e^{-\lambda} \leq e^{b-a} \left(\frac{b}{a} + 1 \right).$$

Так что для выборки из пуассоновского распределения и функции

$$h(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

выполняются условия теоремы 14.3.2 и следовательно имеет место неравенство Крамера—Рао

$$D\hat{\theta} \geq \frac{1}{I(\theta)}.$$

При этом

$$I(\lambda) = M \left(\frac{\partial}{\partial \lambda} \ln P_\lambda(\xi) \right)^2 = -M \frac{\partial^2}{\partial \lambda^2} \ln P_\lambda(\xi) = \frac{n}{\lambda};$$

$$D\hat{\lambda} = D \left(\frac{1}{n} \sum_{j=1}^n \xi_j \right) = \frac{1}{n^2} n D\xi_1 = \frac{1}{n^2} n \lambda = \frac{\lambda}{n}.$$

Поэтому для выборки из пуассоновского распределения и оценки $\hat{\lambda} = \bar{\xi}$ неравенство Крамера—Рао обращается в равенство, а следовательно, $\bar{\xi}$ — эффективная оценка параметра λ .

Другой способ решения примера состоит в проверке достаточного условия обращения неравенства Крамера—Рао в равенство, а именно:

$$\frac{\partial}{\partial \lambda} \ln P(x; \lambda) = c(\lambda)(h(x) - \lambda).$$

Имеем

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln P(x; \lambda) &= \frac{\partial}{\partial \lambda} \ln \prod_{j=1}^n g(x_j; \lambda) = \sum_{j=1}^n \frac{\partial}{\partial \lambda} \ln g(x_j; \lambda) = \\ &= \frac{1}{\lambda} \sum_{j=1}^n (x_j - \lambda) = \frac{1}{\lambda} \left(\sum_{j=1}^n x_j - n\lambda \right) = \frac{n}{\lambda} \left(\frac{1}{n} \sum_{j=1}^n x_j - \lambda \right). \end{aligned}$$

Поэтому $\hat{\lambda} = \bar{\xi}$ — эффективная оценка параметра λ .

14.4 Задачи

АЗ: 14.4, 14.6, 14.27, 14.43.

СЗ: 14.9, 14.12, 14.22, 14.47.

14.1. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta - h; \theta + h]; \\ 0, & \text{если } x \notin [\theta - h; \theta + h] \end{cases}$$

Обозначим через $(\hat{\theta}, \hat{h})$ точку, в которой функция $L(\theta, h) = \prod_{i=1}^n f(\xi_i; \theta, h)$ достигает наибольшего значения.

Являются ли $\hat{\theta}, \hat{h}$ оценками и, если да, то будут ли они несмещенными оценками параметров θ и h ? Состоятельными оценками этих параметров? Если нет, то, возможно, $\hat{\theta}$ и $\hat{h}$ являются несмещенными и состоятельными оценками других параметров распределения?

14.2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; r, \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta} \right)^r \left(\frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, 2, \dots,$$

где r — фиксированное и известное, параметр $\theta > 0$.

Обозначим через $\hat{\theta}$ решение уравнения

$$m_1(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \xi_i,$$

где $m_1(\theta) = \sum_{k=0}^{\infty} kP(k; r, \theta)$.

Является ли $\hat{\theta}$ оценкой и если да, то будет ли она несмещенной оценкой параметра θ ? Состоятельной оценкой θ ?

З а м е ч а н и е. Если обозначить $\frac{1}{1+\theta} = p$, то распределение

$$P(k; r, \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta} \right)^r \left(\frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, \dots,$$

принимает вид

$$P(k; r, p) = C_{r-1+k}^{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots,$$

хорошо известного отрицательного биномиального распределения (см. задачу 15.35).

14.3. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\},$$

σ^2 принадлежит конечному отрезку $[a; b]$, $a > 0$.
Является ли

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2$$

эффективной оценкой параметра σ^2 ?

14.4. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \min \{\xi_i\}; \quad \hat{\theta}_2 = \max \{\xi_i\}; \quad \hat{\theta}_3 = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\hat{\theta}_4 = \frac{\xi_{n-1} + \xi_n}{2}; \quad \hat{\theta}_5 = \max \{\xi_i\} - \min \{\xi_i\};$$

$$\hat{\theta}_6 = \frac{1}{2}(\max \{\xi_i\} + \min \{\xi_i\});$$

$$\hat{\theta}_7 = \min \{\xi_i\} - \frac{1}{n-1}(\max \{\xi_i\} - \min \{\xi_i\});$$

$$\hat{\theta}_8 = \max \{\xi_i\} + \frac{1}{n-1}(\max \{\xi_i\} - \min \{\xi_i\}).$$

Выяснить, какие из оценок $\hat{\theta}_i, i = 1, 2, \dots, 8$, являются несмещенными, состоятельными оценками параметров a и b . Возможно, среди них имеются несмещенные и состоятельные оценки параметров, отличных от a и b ?

14.5. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из пуассоновского распределения с параметром λ :

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Обозначим через $\hat{\lambda}$ точку, в которой функция

$$L(\lambda) = \prod_{i=1}^n P(\xi_i; \lambda)$$

достигает своего наибольшего значения.

Является ли $\hat{\lambda}$ оценкой и если да, то будет ли она несмещенной оценкой параметра λ ? Состоятельной оценкой λ ?

14.6. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$p(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left\{-\frac{(\ln x - \mu_0)^2}{2\sigma^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($p(x)$ — плотность логарифмически нормального распределения), μ_0 — известно, $\sigma^2 > 0$ и принадлежит конечному отрезку $[a; b]$, $a > 0$.

Является ли

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\ln \xi_i - \mu_0)^2$$

эффективной оценкой параметра σ^2 ?

14.7. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{2}, & \text{если } x \in [\theta - 1, \theta + 1]; \\ 0, & \text{если } x \notin [\theta - 1, \theta + 1]. \end{cases}$$

Обозначим через $\hat{\theta}$ точку, в которой функция

$$L(\theta) = \prod_{i=1}^n f(\xi_i; \theta)$$

достигает своего наибольшего значения.

Является ли $\hat{\theta}$ оценкой и если да, то будет ли она несмещенной оценкой параметра θ ? Состоятельной оценкой θ ?

14.8. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \alpha, \theta) = \begin{cases} \frac{1}{\alpha} \exp \left\{ -\frac{1}{\alpha}(x - \theta) \right\}, & \text{если } x \geq \theta; \\ 0, & \text{если } x < \theta. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \min \{\xi_i\} - \frac{\bar{\xi} - \min \{\xi_i\}}{n}; \quad \hat{\theta}_2 = \bar{\xi} - \hat{\theta}_1.$$

Несмещенными, состоятельными оценками каких параметров являются $\hat{\theta}_1$ и $\hat{\theta}_2$ (возможно, эти параметры отличны от α и θ)?

14.9. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left\{ -\frac{(x - \theta)^2}{2\sigma_0^2} \right\},$$

σ_0 — известно, θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

14.10. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} 0, & \text{если } x < \theta; \\ \exp \{\theta - x\}, & \text{если } x \geq \theta. \end{cases}$$

Является ли

$$\hat{\theta} = -\frac{1}{n} + \min \{\xi_i\}$$

несмещенной оценкой параметра θ ? Состоятельной оценкой θ ?

14.11. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}.$$

Обозначим через $(\hat{a}, \hat{\sigma}^2)$ точку, в которой функция

$$L(a, \sigma^2) = \prod_{i=1}^n f(\xi_i; a, \sigma^2)$$

достигает своего наибольшего значения.

Являются ли $\hat{a}$ и $\hat{\sigma}^2$ оценками и если да, то будут ли они несмещенными оценками соответственно параметров a и σ^2 ? Состоятельными оценками этих параметров?

14.12. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta} \right)^r \left(\frac{\theta}{1+\theta} \right)^k, k = 0, 1, \dots; \theta > 0,$$

r — фиксировано и известно, θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{nr} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

Указание. См. задачи 14.2 и 15.35.

14.13. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{2h_0}, & \text{если } x \in [\theta - h_0, \theta + h_0]; \\ 0, & \text{если } x \notin [\theta - h_0, \theta + h_0]. \end{cases}$$

Обозначим через $\hat{\theta}$ точку, в которой функция $L(\theta) = \prod_{i=1}^n f(\xi_i; \theta)$ достигает своего наибольшего значения.

Является ли $\hat{\theta}$ оценкой и если да, то будет ли она несмещенной оценкой параметра θ ? Состоятельной оценкой θ ?

14.14. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, \sigma^2) = \begin{cases} \frac{1}{2\sqrt{3}\sigma}, & \text{если } x \in [\theta - \sqrt{3}\sigma, \theta + \sqrt{3}\sigma]; \\ 0, & \text{если } x \notin [\theta - \sqrt{3}\sigma, \theta + \sqrt{3}\sigma]. \end{cases}$$

Обозначим через $\hat{\theta}, \hat{\sigma}^2$ решение системы уравнений

$$m_1(\hat{\theta}; \hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \xi_i,$$

$$m_2(\hat{\theta}; \hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \xi_i^2,$$

где

$$m_k(\theta; \sigma^2) = \int_{\mathbb{R}^1} x^k f(x; \theta, \sigma^2) dx, \quad k = 1, 2.$$

Являются ли $\hat{\theta}$ и $\hat{\sigma}^2$ оценками и, если да, являются ли они несмещенными оценками соответственно параметров θ и σ^2 ? Состоятельными оценками этих параметров?

14.15. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta} \right)^k, \quad k = 0, 1, \dots,$$

где $\theta > 0$ и принадлежит конечному отрезку $[a; b]$.

Является ли

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

У к а з а н и е. Если обозначить $\frac{1}{1+\theta} = p$, то распределение

$$P(k; \theta) = \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, \dots$$

принимает вид хорошо известного геометрического распределения

$$P(k; p) = p(1-p)^k, \quad k = 0, 1, \dots$$

14.16. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{2h_0}, & \text{если } x \in [\theta - h_0, \theta + h_0]; \\ 0, & \text{если } x \notin [\theta - h_0, \theta + h_0], \end{cases}$$

h_0 — известно ($f(x; \theta)$ — плотность равномерного на отрезке $[\theta - h_0, \theta + h_0]$ распределения).

Рассматриваются оценки

$$\hat{\theta}_1 = \min \{\xi_i\}; \quad \hat{\theta}_2 = \max \{\xi_i\}; \quad \hat{\theta}_3 = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\hat{\theta}_\lambda = \lambda(\max \{\xi_i\} - h_0) + (1 - \lambda)(\min \{\xi_i\} + h_0), \quad \lambda \in [0; 1].$$

Выяснить, имеются ли среди оценок $\hat{\theta}_i$ ($i = 1, 2, 3$), $\hat{\theta}_\lambda$ несмещенные, состоятельные оценки параметра θ . Возможно, среди них имеются несмещенные, состоятельные оценки других параметров? Каких именно?

14.17. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из нормального распределения с плотностью

$$f(x; a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}.$$

Обозначим через $\hat{a}, \hat{\sigma}^2$ решение системы уравнений

$$m_1(\hat{a}; \hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$m_2(\hat{a}; \hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n \xi_i^2,$$

где

$$m_k(a; \sigma^2) = \int_{\mathbb{R}^1} x^k f(x; a, \sigma^2) dx, \quad k = 1, 2.$$

Являются ли $\hat{a}$ и $\hat{\sigma}^2$ оценками? И если да, то являются ли $\hat{a}$ и $\hat{\sigma}^2$ несмещенными оценками соответственно параметров a и σ^2 ? Состоятельными оценками этих параметров?

14.18. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; m, \theta) = \begin{cases} \frac{1}{\theta^m (m-1)!} x^{m-1} \exp \left\{ -\frac{1}{\theta} x \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($f(x; m, \theta)$ — плотность распределения Эрланга с параметрами (m, θ)), m — известно; θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{mn} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

14.19. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

Обозначим через $(\hat{a}, \hat{b})$ точку, в которой функция

$$L(a, b) = \prod_{i=1}^n f(\xi_i; a, b)$$

достигает своего наибольшего значения (может существовать не одна такая точка).

Являются ли $\hat{a}$ и $\hat{b}$ оценками? Если да, то являются ли $\hat{a}$ и $\hat{b}$ несмещенными оценками соответственно параметров a и b ? Состоятельными оценками этих параметров?

14.20. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = \left(\frac{1}{1 + \theta} \right) \left(\frac{\theta}{1 + \theta} \right)^k, \quad k = 0, 1, \dots; \quad \theta > 0.$$

Обозначим через $\hat{\theta}$ решение уравнения

$$m_1(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \xi_i,$$

где

$$m_1(\theta) = \sum_{k=0}^{\infty} kP(k; \theta).$$

Является ли $\hat{\theta}$ оценкой? И если да, то является ли она несмещенной оценкой параметра θ ? Состоятельной оценкой θ ?

Указание. См. задачу 14.15.

14.21. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} \exp \left\{ -\frac{1}{\theta} x \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\theta > 0$ (θ принадлежит конечному отрезку $[a; b]$).

Является ли

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

14.22. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta - h, \theta + h]; \\ 0, & \text{если } x \notin [\theta - h, \theta + h]. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \frac{1}{2} (\max \{\xi_i\} - \min \{\xi_i\}); \quad \hat{\theta}_2 = \frac{1}{2} (\max \{\xi_i\} + \min \{\xi_i\});$$

$$\hat{\theta}_3 = \min \{\xi_i\} - \frac{1}{n-1} (\max \{\xi_i\} - \min \{\xi_i\});$$

$$\hat{\theta}_4 = \max \{\xi_i\} + \frac{1}{n-1} (\max \{\xi_i\} - \min \{\xi_i\}).$$

Выяснить, имеются ли среди $\hat{\theta}_i$, $i = 1, 2, 3, 4$, несмещенные, состоятельные оценки параметров θ и h . Возможно, среди них имеются несмещенные, состоятельные оценки других параметров? Каких именно?

14.23. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$Q(k; p) = p^k (1-p)^{1-k}, \quad k = 0; 1, \quad 0 < p < 1.$$

Обозначим через $\hat{p}$ точку, в которой функция

$$L(p) = \prod_{i=1}^n p^{\xi_i} (1-p)^{1-\xi_i}$$

достигает своего наибольшего значения.

Является ли $\hat{p}$ несмещенной оценкой параметра p ? Состоятельной оценкой этого параметра?

14.24. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} \exp \left\{ -\frac{1}{\theta} (x - m_0) \right\}, & \text{если } x > m_0; \\ 0, & \text{если } x \leq m_0, \end{cases}$$

параметр $\theta > 0$ (m_0 — известно, θ принадлежит конечному отрезку $[a; b]$).

Является ли $\hat{\theta} = \bar{\xi} - m_0$ эффективной оценкой параметра θ ?

14.25. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta_0 - h, \theta_0 + h]; \\ 0, & \text{если } x \notin [\theta_0 - h, \theta_0 + h]; \end{cases}$$

θ_0 — задано.

Рассматриваются оценки

$$\hat{\theta}_1 = \frac{1}{2} (\max \{\xi_i\} - \min \{\xi_i\}); \quad \hat{\theta}_2 = \theta_0 - \hat{\theta}_1;$$

$$\hat{\theta}_3 = \theta_0 + \hat{\theta}_1; \quad \hat{\theta}_4 = \max \{ \max \{\xi_i\} - \theta_0, \theta_0 - \min \{\xi_i\} \}.$$

Выяснить, имеются ли среди $\hat{\theta}_i$, $i = 1, 2, 3, 4$, несмещенные, состоятельные оценки параметра h . Возможно, среди них имеются несмещенные, состоятельные оценки других параметров? Каких именно?

14.26. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \alpha) = \begin{cases} \frac{1}{\alpha} \exp \left\{ -\frac{1}{\alpha} x \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Доказать, что

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i$$

является несмещенной и состоятельной оценкой параметра α .

Является ли

$$\hat{\theta}_2 = \frac{1}{2} (\xi_{n-1} + \xi_n)$$

несмещенной оценкой параметра α ? Состоятельной оценкой этого параметра?

14.27. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из биномиального распределения

$$P(k; \theta) = C_N^k \theta^k (1 - \theta)^{N-k}, \quad k = 0, 1, \dots, N,$$

где N — известное целое число; θ принадлежит конечно-му отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{nN} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

14.28. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, b) = \begin{cases} \frac{1}{\theta} \exp \left\{ -\frac{1}{\theta}(x - b) \right\}, & \text{если } x > b; \\ 0, & \text{если } x \leq b. \end{cases}$$

Рассматриваются оценки

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad \hat{\theta}_2 = \min \{ \xi_i \}; \quad \hat{\theta}_3 = \hat{\theta}_1 - \hat{\theta}_2.$$

Имеются ли среди $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ несмещенные, состоятельные оценки параметров θ и b ? Возможно, среди них имеются несмещенные и состоятельные оценки других параметров? Каких именно?

14.29. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из биномиального распределения с параметрами m и θ :

$$P(k; \theta) = C_m^k \theta^k (1 - \theta)^{m-k}, \quad k = 0, 1, \dots, m,$$

m — известно.

Обозначим через $\hat{\theta}$ точку, в которой функция

$$L(\theta) = \prod_{i=1}^n C_m^{\xi_i} \theta^{\xi_i} (1 - \theta)^{m-\xi_i}$$

достигает наибольшего значения (может существовать не одна такая точка). Несмещенной и состоятельной оценкой каких параметров является $\hat{\theta}$?

14.30. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из двустороннего показательного распределения с плотностью

$$f(x; \theta) = \frac{1}{2\theta} \exp \left\{ -\frac{1}{\theta} |x - m_0| \right\}, \quad \theta > 0,$$

где m_0 — известная константа, параметр θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n |\xi_i - m_0|$$

эффективной оценкой параметра θ ?

14.31. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta_0 - h, \theta_0 + h]; \\ 0, & \text{если } x \notin [\theta_0 - h, \theta_0 + h]. \end{cases}$$

Обозначим через $\hat{h}$ точку, в которой функция

$$L(h) = \prod_{i=1}^n f(\xi_i; h)$$

достигает своего наибольшего значения.

Является ли $\hat{h}$ оценкой и если да, то является ли она несмещенной оценкой параметра h ? Состоятельной оценкой этого параметра?

14.32. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \sigma^2) = \begin{cases} \sqrt{\frac{2}{\pi\sigma^2}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($f(x; \sigma^2)$ — плотность полунормального распределения).

Несмещенной состоятельной и эффективной оценкой какого параметра является

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i^2 ?$$

14.33. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{x}{\theta} \exp \left\{ -\frac{x^2}{2\theta} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($f(x; \theta)$ — плотность распределения Рэлея), параметр θ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\theta} = \frac{1}{2n} \sum_{i=1}^n \xi_i^2$$

эффективной оценкой параметра θ ?

14.34. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; b) = \begin{cases} \frac{1}{b - a_0}, & \text{если } x \in [a_0; b]; \\ 0, & \text{если } x \notin [a_0; b], \end{cases}$$

a_0 — известно.

Рассматриваются оценки

$$\hat{\theta}_1 = \max \{ \xi_i \}; \quad \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\hat{\theta}_3 = \max \{ \xi_i \} + \frac{1}{n-1} (\max \{ \xi_i \} - a_0).$$

Выяснить, имеются ли среди $\hat{\theta}_i$, $i = 1, 2, 3$, несмещенные, состоятельные оценки параметра b . Возможно, среди них имеются несмещенные и состоятельные оценки других параметров? Каких именно?

14.35. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из биномиального распределения с параметрами m и p :

$$P(k; p) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m.$$

Является ли

$$\hat{p} = \frac{1}{mn} \sum_{i=1}^n \xi_i$$

состоятельной оценкой параметра p ?

14.36. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \mu) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_0^2}x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma_0^2}\right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

где $\sigma_0 > 0$ ($f(x; \mu)$ — плотность логарифмически нормального распределения); σ_0 — известно; μ принадлежит конечному отрезку $[a; b]$.

Является ли

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln \xi_i$$

эффективной оценкой параметра μ ?

14.37. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, \lambda) = \begin{cases} \frac{\theta \lambda^\theta}{x^{\theta+1}}, & \text{если } x > \lambda; \\ 0, & \text{если } x \leq \lambda, \end{cases}$$

$\lambda > 0$, $\theta > 2$ (распределение Парето); θ — известно.

Рассматриваются оценки

$$\hat{\lambda}_1 = \frac{\theta - 1}{\theta n} \sum_{i=1}^n \xi_i; \quad \hat{\lambda}_2 = \frac{\theta - 2}{\theta n} \sum_{i=1}^n \xi_i^2.$$

Несмещенными и состоятельными оценками каких параметров являются $\hat{\lambda}_1$ и $\hat{\lambda}_2$?

14.38. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{\theta^\nu \Gamma(\nu)} x^{\nu-1} \exp \left\{ -\frac{1}{\theta} x \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

ν — известный параметр.

Является ли

$$\hat{\theta} = \frac{1}{n\nu} \sum_{i=1}^n \xi_i$$

эффективной оценкой параметра θ ?

14.39. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного на отрезке $[a; a+1]$ распределения. Рассматриваются две оценки параметра a :

$$\hat{a}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{2}; \quad \hat{a}_2 = \max \{ \xi_i \} - \frac{n}{n+1}.$$

Доказать, что $\hat{a}_1$ и $\hat{a}_2$ являются несмещенными оценками параметра a .

Найти $D\hat{a}_1$ и $D\hat{a}_2$. Доказать, что

$$D\hat{a}_2 = o(D\hat{a}_1),$$

при $n \rightarrow \infty$.

14.40. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \frac{1}{2\theta} \exp \left\{ -\frac{1}{\theta} |x| \right\}, \quad \theta > 0.$$

($f(x; \theta)$ — плотность двустороннего показательного распределения).

Является ли

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n |\xi_i|$$

эффективной оценкой параметра θ ?

14.41. Известно, что $\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного на отрезке $[a; b]$ распределения, но сам отрезок $[a; b]$ неизвестен. Можно предложить несколько оценок середины этого отрезка. В частности,

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad \hat{\theta}_2 = \frac{1}{2} (\max \{\xi_i\} + \min \{\xi_i\}).$$

Доказать, что $\hat{\theta}_1$ и $\hat{\theta}_2$ являются несмещенными оценками середины $(a + b)/2$ отрезка $[a; b]$ и

$$D\hat{\theta}_1 > D\hat{\theta}_2.$$

14.42. Найти доверительный интервал для:

1° границы прочности на разрыв материала (см. пример 17.4.1) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии границы прочности на разрыв материала с коэффициентом надежности: а) 0,90; б) 0,98 (для измерений, проведенных на стенде А).

14.43. Найти доверительный интервал для:

1° содержания золота на кубический метр породы, добываемой ударно-канатным бурением (см. задачу 17.8) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии содержания золота с коэффициентом надежности: а) 0,90; б) 0,98.

14.44. Найти доверительный интервал для:

1° содержания золота на кубический метр породы, добываемый из шурфов (см. задачу 17.8), с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии содержания золота с коэффициентом надежности: а) 0,90; б) 0,98.

14.45. Найти доверительный интервал для:

1° толщины пластикового покрытия провода (см. задачу 18.46) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии толщины пластикового покрытия провода с коэффициентом надежности: а) 0,90; б) 0,98.

14.46. Найти доверительный интервал для:

1° значения величины e , найденной Р. Милликеном при определении заряда электрона (см. задачу 18.34) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии измерений величины ϵ с коэффициентом надежности: а) 0,90; б) 0,98.

14.47. Найти доверительный интервал для:

1° содержания марганца в стали, выплавленной в конверторе I (см. задачу 17.17) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии содержания марганца в стали с коэффициентом надежности: а) 0,90; б) 0,98.

14.48. Найти доверительный интервал для:

1° отражательной способности краски A (см. задачу 17.2) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии отражательной способности краски B с коэффициентом надежности: а) 0,90; б) 0,98.

14.49. Найти доверительный интервал для:

1° содержания хрома в нержавеющей стали 18Cr10Ni2Mo (см. задачу 17.22) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии содержания хрома с коэффициентом надежности: а) 0,90; б) 0,98.

14.50. Найти доверительный интервал для:

1° сопротивления провода типа A (см. задачу 17.24) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии сопротивления провода с коэффициентом надежности: а) 0,90; б) 0,98.

14.51. Найти доверительный интервал для:

1° пористости конденсаторной бумаги партии I (см. задачу 19.36) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии пористости конденсаторной бумаги партии I с коэффициентом надежности: а) 0,90; б) 0,98.

14.52. Найти доверительный интервал для:

1° температуры левой шины автобуса (см. задачу 17.31) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии температуры левой шины автобуса с коэффициентом надежности: а) 0,90; б) 0,98.

14.53. Найти доверительный интервал для:

1° длины изделий после обжига (см. задачу 17.37) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии длины изделий после обжига с коэффициентом надежности: а) 0,90; б) 0,98.

14.54. Найти доверительный интервал для:

1° глубины канавки на шине, полученной контролером I (см. пример 19.2.1) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии глубины канавки с коэффициентом надежности: а) 0,90; б) 0,98.

14.55. Найти доверительный интервал для:

1° разности толщины плиток при износе для относительной концентрации наполнителя 1 к 1 (см. пример 19.2.1) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии разности толщины плиток при износе с коэффициентом надежности: а) 0,90; б) 0,98.

14.56. Найти доверительный интервал для:

1° объема хлебных булок, содержащих 1 мг бромистого калия (см. задачу 19.8) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии объема хлебных булок с коэффициентом надежности: а) 0,90; б) 0,98.

14.57. Найти доверительный интервал для:

1° температуры возгорания эмали, полученной оператором A (см. задачу 17.38) с коэффициентом надежности а) 0,90; б) 0,99;

2° дисперсии температуры возгорания эмали с коэффициентом надежности а) 0,90; б) 0,98.

14.58. Найти доверительный интервал для:

1° диаметра шейки рабочей части сверл (см. задачу 19.7) с коэффициентом надежности: а) 0,90; б) 0,99;

2° дисперсии диаметра шейки рабочей части сверл с коэффициентом надежности: а) 0,90; б) 0,98.

Глава 15

Методы построения оценок

15.1 Эмпирические оценки

Эмпирическая функция распределения. Пусть ξ — случайная величина, функция распределения $F(x)$, которой неизвестна. Мы имеем n независимых наблюдений $\xi_1, \xi_2, \dots, \xi_n$ случайной величины ξ — выборку из распределения F . По выборке $\xi_1, \xi_2, \dots, \xi_n$ можно построить “хорошую” оценку функции распределения.

О п р е д е л е н и е. Функцию $\hat{F}_n(x)$, заданную на $\mathbb{R}^1$ равенством

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n I_{(-\infty, x)}(\xi_k),$$

будем называть *эмпирической функцией распределения* ($I_A(t)$ — индикатор множества A). Эмпирическую функцию распределения еще можно определить так:

$$\hat{F}_n(x) = \frac{\mu(x)}{n},$$

где $\mu(x)$ — количество ξ_k , для которых $\xi_k < x$, поскольку

$$\mu(x) = \sum_{k=1}^n I_{(-\infty, x)}(\xi_k).$$

Для каждого фиксированного x эмпирическая функция распределения $\hat{F}_n(x)$ как функция случайного вектора $\xi_1, \xi_2, \dots, \xi_n$ является случайной величиной. Поэтому $\hat{F}_n(x)$ является функцией и ω , т.е.

$$\hat{F}_n(x) = \hat{F}_n(x, \omega), \quad \omega \in \Omega, \quad x \in \mathbb{R}^1.$$

Эмпирическая функция распределения $\hat{F}_n(x)$ для каждого фиксированного x является несмещенной и состоятельной оценкой значения $F(x)$ функции распределения.

По выборке $\xi_1, \xi_2, \dots, \xi_n$ определим случайные величины $\xi_1^*, \xi_2^*, \dots, \xi_n^*$: для каждого фиксированного $\omega \in \Omega$ упорядочим значения $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)$ в порядке возрастания:

$$\xi_{k_1}(\omega) \leq \xi_{k_2}(\omega) \leq \dots \leq \xi_{k_n}(\omega)$$

и положим

$$\xi_1^*(\omega) = \xi_{k_1}(\omega), \quad \xi_2^*(\omega) = \xi_{k_2}(\omega), \quad \dots, \quad \xi_n^*(\omega) = \xi_{k_n}(\omega).$$

О п р е д е л е н и е. Последовательность $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ называется *вариационным рядом* последовательности $\xi_1, \xi_2, \dots, \xi_n$, а случайные величины $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ называют — *порядковыми статистиками*.

В терминах порядковых статистик эмпирическую функцию распределения можно представить в виде

$$\hat{F}_n(x) = \frac{\mu^*(x)}{n}, \quad (15.1.1)$$

где $\mu^*(x)$ — количество ξ_k^* , для которых $\xi_k^* < x$. Непосредственно из равенства (15.1.1) получаем, что при фиксированном ω значение $\hat{F}_n(x)$ равно 0 в каждой точке x промежутка $(-\infty, \xi_1^*]$, поскольку количество тех ξ_k^* , для которых $\xi_k^* < x$, равно 0; $\hat{F}_n(x) = 1/n$ в каждой точке промежутка $(\xi_1^*, \xi_2^*]$, поскольку количество тех ξ_k^* , для которых $\xi_k^* < x$, равно 1, и т. д., и, наконец, $\hat{F}_n(x) = 1$ для каждого x из промежутка $(\xi_n^*, +\infty)$.

Из сказанного выше следует, что для каждого фиксированного ω функция $\hat{F}_n(x) = \hat{F}_n(x, \omega)$ неотрицательна; постоянна на каждом из промежутков $(-\infty, \xi_1^*], (\xi_k^*, \xi_{k+1}^*], k = 1, 2, \dots, n-1, (\xi_n^*, +\infty)$ (а следовательно, непрерывна слева) и неубывающая — растет в точках $\xi_k^*, k = 1, 2, \dots, n$, скачками $1/n$.

График эмпирической функции распределения $\hat{F}_n(x)$ (точнее, реализации $\hat{F}_n(x, \omega)$) изображен на рис. 15.1.1.

З а м е ч а н и е 1. Для выборки $\xi_1, \xi_2, \dots, \xi_n$ из непрерывных распределений вероятность того, что выборочные значения совпадут, равна нулю, но поскольку мы фиксируем результаты с заданной точностью (например, до третьего знака), некоторые выборочные значения могут совпадать. При этом скачок эмпирической функции распределения в точке ξ_k равен l/n , где l — количество выборочных значений, совпадающих с ξ_k , учитывая и ξ_k .

З а м е ч а н и е 2. Распределение, соответствующее эмпирической функции распределения $\hat{F}_n(x)$, будем называть *эмпирическим*. Для каждого фиксированного ω это дискретное распределение, ставящее в соответствие каждой точке $\xi_k, k = 1, 2, \dots, n$, “массу” $1/n$ (или l/n , если с ξ_k совпадают l выборочных значений, учитывая и ξ_k).

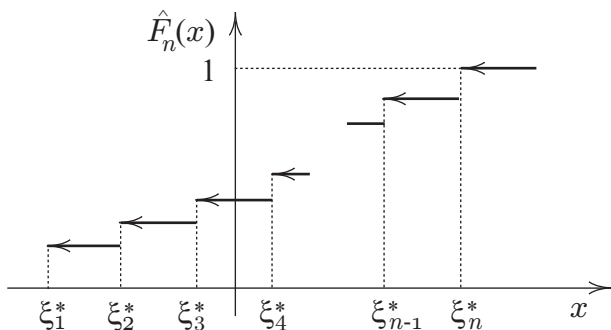


Рис. 15.1.1: График реализации эмпирической функции распределения $\hat{F}_n(x)$

Об уклонении эмпирической функции распределения $\hat{F}_n(x)$ от функции распределения $F(x)$. Эмпирическая функция распределения $\hat{F}_n(x)$, построенная по выборке $\xi_1, \xi_2, \dots, \xi_n$ из F , является несмещенной и состоятельной оценкой $F(x)$ (дает хорошее приближение для $F(x)$). Естественно поставить вопрос: насколько эмпирическая функция распределения $\hat{F}_n(x)$ уклоняется от $F(x)$? (Какими могут быть уклонения $\hat{F}_n(x)$ от $F(x)$?) Заметим, что уклонение $\hat{F}_n(x)$ от $F(x)$ является случайной величиной, и когда мы говорим о таком уклонении, то имеется в виду его распределение.

Меру уклонения эмпирической функции распределения $\hat{F}_n(x)$ от функции распределения $F(x)$ можно вводить различными способами. Рассмотрим уклонение $\hat{F}_n(x)$ от $F(x)$, предложенное А. Н. Колмогоровым (для непрерывной $F(x)$):

$$\sup_{-\infty < x < \infty} |F(x) - \hat{F}_n(x)|,$$

далее будем писать $\sup_x |F(x) - \hat{F}_n(x)|$. Эта мера уклонения обладает важным свойством: для достаточно больших n распределение $\sup_x |F(x) - \hat{F}_n(x)| / \frac{1}{\sqrt{n}}$, независимо от распределения F , из которого получена выборка $\xi_1, \xi_2, \dots, \xi_n$, близко к распределению А.Н. Колмогорова (см. 19.1).

Вычисление $\sup_x |F(x) - \hat{F}_n(x)|$. Пусть $\hat{F}_n(x)$ — эмпирическая функция распределения, построенная по выборке $\xi_1, \xi_2, \dots, \xi_n$, точнее — по реализации выборки; $F(x)$ — непрерывная функция распределения. Часто возникает необходимость вычислить

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

Заметим, что $\xi_1, \xi_2, \dots, \xi_n$ — не обязательно выборка из распределения F .

Поскольку $\hat{F}_n(x)$ постоянна на промежутках

$$(-\infty, \xi_1^*], (\xi_1^*, \xi_2^*], \dots, (\xi_{n-1}^*, \xi_n^*], (\xi_n^*, +\infty), \quad (15.1.2)$$

то для вычисления

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|,$$

когда x “пробегает” $\mathbb{R}^1$, значение $\sup_x \left| F(x) - \hat{F}_n(x) \right|$ достаточно вычислить на каждом из указанных промежутков, а затем из полученных чисел выбрать наибольшее.

Вычислим сначала

$$\sup_{x \in (\xi_k^*, \xi_{k+1}^*]} \left| F(x) - \hat{F}_n(x) \right|, \quad k = 1, 2, \dots, n-1.$$

Функция $\hat{F}_n(x)$ постоянна на каждом из промежутков (15.1.2), $F(x)$ — неубывающая функция, а следовательно и функция $F(x) - \hat{F}_n(x)$ на промежутках (15.1.2) неубывающая, и поэтому (см. рис. 15.1.2) значение

$$\sup_{x \in (\xi_k^*, \xi_{k+1}^*]} \left| F(x) - \hat{F}_n(x) \right|$$

равно одному из чисел

$$\left| \sup_{x \in (\xi_k^*, \xi_{k+1}^*]} \left(F(x) - \hat{F}_n(x) \right) \right|, \quad \left| \inf_{x \in (\xi_k^*, \xi_{k+1}^*]} \left(F(x) - \hat{F}_n(x) \right) \right|.$$

Далее,

$$\begin{aligned} & \sup_{x \in (\xi_k^*, \xi_{k+1}^*]} \left(F(x) - \hat{F}_n(x) \right) = \\ &= \lim_{x \uparrow \xi_{k+1}^*} \left(F(x) - \hat{F}_n(x) \right) = F(\xi_{k+1}^*) - \hat{F}_n(\xi_{k+1}^*); \\ & \inf_{x \in (\xi_k^*, \xi_{k+1}^*]} \left(F(x) - \hat{F}_n(x) \right) = \lim_{x \downarrow \xi_k^*} \left(F(x) - \hat{F}_n(x) \right) = \end{aligned}$$

$$= F(\xi_k^*) - \lim_{x \downarrow \xi_k^*} \hat{F}_n(x) = F(\xi_k^*) - \hat{F}_n(\xi_{k+1}^*).$$

Вычисляя пределы, мы учли, что функция $\hat{F}_n(x)$ постоянна на промежутках (15.1.2), а следовательно, непрерывна слева (см. рис. 15.1.2).

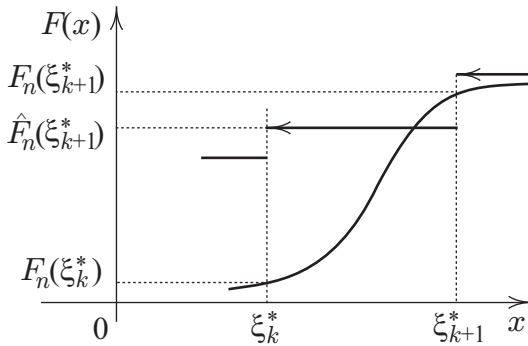


Рис. 15.1.2: К вычислению $\sup_x |F(x) - \hat{F}_n(x)|$

Таким образом,

$$\sup_{x \in (\xi_k^*, \xi_{k+1}^*]} |F(x) - \hat{F}_n(x)|$$

равен

$$\left| \sup_{x \in (\xi_k^*, \xi_{k+1}^*]} (F(x) - \hat{F}_n(x)) \right| = |F(\xi_{k+1}^*) - \hat{F}_n(\xi_{k+1}^*)|$$

или

$$\left| \inf_{x \in (\xi_k^*, \xi_{k+1}^*]} (F(x) - \hat{F}_n(x)) \right| = |F(\xi_k^*) - \hat{F}_n(\xi_{k+1}^*)|,$$

$$k = 1, 2, \dots, n-1.$$

Для промежутков $(-\infty, \xi_1^*]$ и $(\xi_n^*, +\infty)$

$$\sup_{x \in (-\infty, \xi_1^*]} |F(x) - \hat{F}_n(x)| = F(\xi_1^*) = |F(\xi_1^*) - \hat{F}_n(\xi_1^*)|;$$

$$\sup_{x \in (\xi_n^*, +\infty)} |F(x) - \hat{F}_n(x)| = |F(\xi_n^*) - 1| = 1 - F(\xi_n^*).$$

Если через ξ_{n+1}^* обозначить произвольную точку, лежащую правее ξ_n^* , то по определению функции $\hat{F}_n(x)$ ее значение $\hat{F}_n(\xi_{n+1}^*)$ равно единице и последнее равенство можно записать в той же форме, что и остальные:

$$\sup_{x \in (\xi_n^*, +\infty)} |F(x) - \hat{F}_n(x)| = |F(\xi_n^*) - \hat{F}_n(\xi_{n+1}^*)|.$$

Следовательно,

$$\sup_x |F(x) - \hat{F}_n(x)|$$

равен наибольшему из чисел

$$|F(\xi_k^*) - \hat{F}_n(\xi_k^*)|, |F(\xi_k^*) - \hat{F}_n(\xi_{k+1}^*)|, \quad k = 1, 2, \dots, n,$$

где $\hat{F}_n(\xi_{n+1}^*) = 1$.

Пример 15.1.1. Пусть 2,4; 1,0; 0,7; 0,0; 1,1; 1,6; 1,1; -0,4; 0,1; 0,7 — выборка из непрерывного распределения.

Построить график реализации эмпирической функции распределения $\hat{F}_n(x)$ и графики функции $F(x)$ нормального распределения с параметрами: 1) $a = 0,8$; $\sigma = 0,8$; 2) $a = 1$; $\sigma = 0,5$. Вычислить значение

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

Решение. Для построения графика реализации эмпирической функции распределения $\hat{F}_n(x)$, расположим

выборочные значения в вариационный ряд: $-0,4; 0,0; 0,1; 0,7; 0,7; 1,0; 1,1; 1,1; 1,6; 2,4$.

Поскольку выборка получена из непрерывного распределения одинаковых выборочных значений не должно быть. Но в рассматриваемой выборке они все-таки имеются: $0,7$ и $1,1$ встречаются дважды. Это связано с тем, что измерения производятся с ограниченным количеством знаков, в рассматриваемом примере — до десятых. Поэтому некоторые выборочные значения могут совпадать.

Для построения графика реализации эмпирической функции распределения $\hat{F}_{10}(x, \omega)$ нанесем на ось абсцисс порядковые статистики ξ_k^* , $k = 1, 2, \dots, 10$. Эмпирическая функция распределения $\hat{F}_{10}(x, \omega)$ постоянна на промежутках $(-\infty, \xi_1^*]$, $(\xi_k^*, \xi_{k+1}^*]$, $k = 1, 2, \dots, 9$, $(\xi_{10}^*, +\infty)$. На промежутке $(-\infty, \xi_1^*]$ она равна нулю и, “переходя через точки” ξ_k^* (кроме $0,7$ и $1,1$), растет скачками $1/10$, а при переходе через точки $0,7$ и $1,1$ — скачками $2/10$ (поскольку выборочные значения $0,7$ и $1,1$ в выборке встречаются дважды). График реализации $\hat{F}_{10}(x, \omega)$ эмпирической функции распределения $\hat{F}_{10}(x)$ изображен на рис. 15.1.3.

Для вычисления

$$\sup_x \left| F(x) - \hat{F}_n(x) \right| = \sup_x \left| N_{a; \sigma^2}(x) - \hat{F}_n(x) \right|,$$

кроме значений $\hat{F}_n(x)$, еще необходимо знать значения функции распределения

$$N_{a; \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp \left\{ -\frac{(t-a)^2}{2\sigma^2} \right\} dt$$

в точках $\xi_1^*, \xi_2^*, \dots, \xi_n^*$. Значения функции $N_{a; \sigma^2}(x)$ находим по значениям табулированной функции $N_{0;1}(x)$ нормального распределения с параметрами $(0;1)$ (см. табл. 22.1.1). После замены переменной $(t-a)/\sigma = y$ в послед-

нем интеграле получаем

$$\begin{aligned}
 N_{a;\sigma^2}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp \left\{ -\frac{(t-a)^2}{2\sigma^2} \right\} dt = \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a)/\sigma} \exp \left\{ -\frac{t^2}{2} \right\} dt = N_{0;1} \left(\frac{x-a}{\sigma} \right).
 \end{aligned}$$

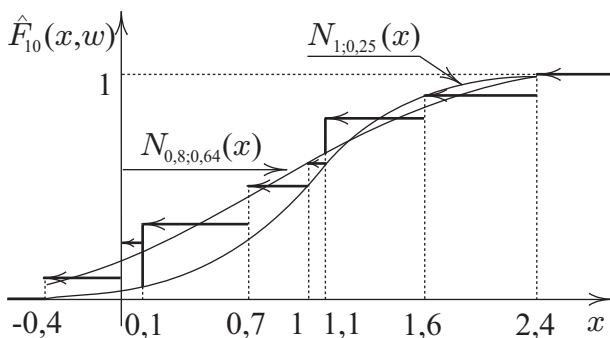


Рис. 15.1.3: К вычислению $\sup_x |N_{0,8;0,64}(x) - \hat{F}_{10}(x)|$ и $\sup_x |N_{1;0,25}(x) - \hat{F}_{10}(x)|$

Этапы вычисления

$$\delta = \sup_x |F(x) - \hat{F}_{10}(x)|,$$

когда $F(x) = N_{0,8;0,64}(x)$, сведены в табл. 15.1.1, согласно которой $\delta = 0,16$.

Аналогично получаем, что

$$\sup_x |N_{1;0,25}(x) - \hat{F}_{10}(x)| = 0,26.$$

Т а б л и ц а 15.1.1. К вычислению

$$\sup_x \left| F(x) - \hat{F}_{10}(x) \right| = \sup_x \left| N_{0,8;0,64}(x) - \hat{F}_{10}(x) \right|$$

ξ_k^*	$\hat{F}_{10}(\xi_k^*)$	$F(\xi_k^*)$	$ F(\xi_k^*) - \hat{F}_{10}(\xi_k^*) $	$ F(\xi_k^*) - \hat{F}_{10}(\xi_{k+1}^*) $
-0,4	0	0,07	0,07	0,03
0,0	0,1	0,16	0,06	0,04
0,1	0,2	0,19	0,01	0,11
(2) 0,7	0,3	0,45	0,15	0,05
1,0	0,5	0,58	0,08	0,02
(2) 1,1	0,6	0,64	0,04	0,16
1,6	0,8	0,84	0,04	0,06
2,4	0,9	0,98	0,08	0,02
ξ_{n+1}^*	1,0			

Эмпирические (выборочные) значения параметров. По эмпирической функции распределения можно строить интуитивно-наглядные оценки параметров распределения.

Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения $F(\cdot; \theta)$, зависящего от параметра θ . Параметр θ неизвестен, и его необходимо оценить по выборке.

Предположим, что параметр θ однозначно определяется распределением (функцией распределения $F(x; \theta)$), т.е.

$$\theta = \Phi(F(x; \theta)),$$

где Φ — функционал, заданный на некотором множестве функций распределения. Например, $a = M\xi$, $\sigma^2 = D\xi$ (когда они существуют) являются функционалами функции распределения $F(x)$ случайной величины:

$$a = \int_{\mathbb{R}^1} xF(dx); \quad \sigma^2 = \int_{\mathbb{R}^1} \left(x - \int_{\mathbb{R}^1} tF(dt) \right)^2 F(dx).$$

Выборка $\xi_1, \xi_2, \dots, \xi_n$ задает эмпирическую функцию распределения $\hat{F}_n(x)$. И поскольку последняя близка к функции распределения $F(x; \theta)$ ($\hat{F}_n(x)$ при каждом x является несмещенной и состоятельной оценкой $F(x; \theta)$), а $\theta =$

$= \Phi(F(x; \theta))$, то в качестве оценки параметра θ естественно рассмотреть

$$\hat{\theta}_n = \Phi(\hat{F}_n(x)).$$

Например, для параметра a формула $\hat{\theta}_n = \Phi(\hat{F}_n(x))$ дает оценку

$$\hat{a} = \int_{\mathbb{R}^1} x \hat{F}_n(dx) = \frac{1}{n} \sum_{i=1}^n \xi_i,$$

а для параметра σ^2 — оценку

$$\begin{aligned} \hat{\sigma}^2 &= \int_{\mathbb{R}^1} \left(x - \int_{\mathbb{R}^1} t \hat{F}_n(dt) \right)^2 \hat{F}_n(dx) = \\ &= \frac{1}{n} \sum_{i=1}^n \left(\xi_i - \frac{1}{n} \sum_{k=1}^n \xi_k \right)^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \hat{a})^2. \end{aligned}$$

Интегралы вычисляются как интегралы Лебега по дискретному распределению, сосредоточенному в точках $\xi_1, \xi_2, \dots, \xi_n$ (с “массой” $1/n$ в каждой).

О п р е д е л е н и е. Оценку $\hat{\theta}_n$ параметра $\theta = \Phi(F(x; \theta))$, построенную по формуле

$$\hat{\theta}_n = \Phi(\hat{F}_n(x)),$$

будем называть *эмпирическим (выборочным) значением параметра θ* .

В частности,

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

— эмпирическое (выборочное) среднее, а

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \hat{a})^2$$

— эмпирическая (выборочная) дисперсия.

Теорема 15.1.1. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения F и $g(x)$ — борелевская функция на $\mathbb{R}^1$ со значениями в $\mathbb{R}^1$. Если

$$G = \int_{\mathbb{R}^1} g(x) F(dx) \neq \infty,$$

то выборочное значение

$$\hat{G}_n = \int_{R_1} g(x) \hat{F}_n(dx) = \frac{1}{n} \sum_{k=1}^n g(\xi_k)$$

величины G является ее состоятельной и несмещенной оценкой.

Следствие. Пусть

$$G = \int_0^1 g(x) dx \neq \infty,$$

$\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного на промежутке $[0; 1]$ распределения, тогда

$$\hat{G}_n = \frac{1}{n} \sum_{k=1}^n g(\xi_k)$$

является состоятельной и несмещенной оценкой G .

Следствие из теоремы обозначает, что при больших n значительные отклонения суммы $\hat{G}_n$ от G встречаются изредка и поэтому в качестве приближенного значения интеграла G можно рассматривать сумму $\hat{G}_n$.

Описанный метод вычисления интегралов известен под названием метода Монте-Карло (статистических испытаний). Он особенно эффективен при вычислении интегралов большой кратности, когда другие методы приближенного анализа непригодны.

Пример 15.1.2. На отрезок $[0; 2]$ наудачу бросают точку, которая делит его на две части, и фиксируют координату ξ точки деления.

Эксперимент проведен 10 раз. При этом получено 10 значений независимых наблюдений случайной величины ξ : 0,00; 0,31; 0,70; 1,40; 0,08; 1,93; 0,79; 1,43; 1,42; 1,69.

Оценить по выборке математическое ожидание объема куба с ребром, равным большей части отрезка.

Решение. По условию задачи длина ребра куба равна $\max\{\xi, 2 - \xi\}$. Объем $g(\xi) = (\max\{\xi, 2 - \xi\})^3$ куба — функция случайной величины ξ . Значение $G = Mg(\xi)$ конечно, поскольку распределение случайной величины ξ сосредоточено на конечном промежутке $[0; 2]$, а функция $g(x)$ на этом промежутке ограничена. Поэтому несмещенной и состоятельной оценкой математического ожидания объема куба $G = Mg(\xi) = M(\max\{\xi, 2 - \xi\})^3$ является

$$\hat{G}_n = \frac{1}{n} \sum_{i=1}^n g(\xi_i) = \frac{1}{n} \sum_{i=1}^n (\max\{\xi_i, 2 - \xi_i\})^3.$$

В рассматриваемом примере $n = 10$, а выборочные значения $\xi_1, \xi_2, \dots, \xi_n$ приведены в условии. Значение $\hat{G}_{10} = 4,44$.

15.2 Метод моментов

Пусть $(\xi_1, \xi_2, \dots, \xi_n)$ — выборка из распределения $F(\cdot; \theta) = F(\cdot; \theta_1, \theta_2, \dots, \theta_s)$. Параметры $\theta_1, \theta_2, \dots, \theta_s$ неизвестны. Их необходимо оценить по выборке.

Первым общим методом построения оценок параметров по выборке был метод моментов, предложенный К. Пирсоном. Согласно этому методу определенное количество выборочных моментов $\hat{m}_k$ приравнивают соответствующим моментам

$$m_k(\theta_1, \theta_2, \dots, \theta_s) = \int_{R_1} x^k F(dx, \theta_1, \theta_2, \dots, \theta_s) \quad (15.2.1)$$

распределения $F(\cdot; \theta_1, \theta_2, \dots, \theta_s)$, вычисленным при значениях параметров $\theta_1, \theta_2, \dots, \theta_s$, равных соответственно оценкам $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s$ этих параметров (моменты $m_k = m_k(\theta_1, \theta_2, \dots, \theta_s)$ распределения $F(\cdot; \theta_1, \theta_2, \dots, \theta_s)$ являются функциями параметров $\theta_1, \theta_2, \dots, \theta_s$). Так что первый выборочный момент

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i$$

приравнивают первому теоретическому моменту

$$m_1(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s) = \int_{\mathbb{R}^1} x F(dx; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s),$$

второй выборочный момент

$$\hat{m}_2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2$$

— второму теоретическому моменту

$$m_2(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s) = \int_{\mathbb{R}^1} x^2 F(dx; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s)$$

и т. д. Рассматривая количество моментов, равное числу неизвестных параметров (подлежащих оцениванию), получают такое же количество уравнений для определения неизвестных параметров:

$$\hat{m}_1 = \int_{\mathbb{R}^1} x F(dx; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s);$$

$$\hat{m}_2 = \int_{\mathbb{R}^1} x^2 F(dx; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s);$$

$\vdots$

$$\hat{m}_s = \int_{\mathbb{R}^1} x^s F(dx; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s).$$

Решая эти уравнения относительно $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s$, находим искомые оценки.

Примечание. Некоторые теоретические моменты могут не зависеть от неизвестных параметров. В этом случае к выписанным уравнениям добавляют следующие, так чтобы количество уравнений было равно количеству неизвестных параметров.

Метод моментов можно использовать, если существуют все перечисленные моменты. На практике он приводит к сравнительно простым вычислениям.

Пример 15.2.1. *Найти оценки параметров a и σ^2 нормального распределения*

$$N_{a;\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left\{-\frac{(t-a)^2}{2\sigma^2}\right\} dt$$

методом моментов.

Решение. Для нормального распределения с параметрами a и σ^2

$$m_1(a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}^1} x \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = a;$$

$$m_2(a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}^1} x^2 \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx = a^2 + \sigma^2.$$

Моменты $m_1(a, \sigma^2)$ и $m_2(a, \sigma^2)$, вычисленные при значениях параметров a и σ^2 , равных соответственно $\hat{a}$ и $\hat{\sigma}^2$, равны

$$m_1(\hat{a}; \hat{\sigma}^2) = \hat{a}; \quad m_2(\hat{a}; \hat{\sigma}^2) = \hat{\sigma}^2 + \hat{a}^2.$$

Первым и вторым эмпирическими моментами являются

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2.$$

Приравнивая моменты $m_1(\hat{a}; \hat{\sigma}^2)$ и $m_2(\hat{a}, \hat{\sigma}^2)$ соответствующим эмпирическим моментам, получим уравнения

$$m_1(\hat{a}; \hat{\sigma}^2) = \hat{m}_1;$$

$$m_2(\hat{a}, \hat{\sigma}^2) = \hat{m}_2,$$

или подробнее

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\hat{\sigma}^2 + (\hat{a})^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2.$$

Решения этих уравнений относительно $\hat{a}$ и $\hat{\sigma}^2$ и являются оценками параметров a и σ^2 , найденными по методу моментов:

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - \hat{a}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

15.3 Метод максимального правдоподобия

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка с распределением $F(\cdot; \theta) = F(\cdot; \theta_1, \theta_2, \dots, \theta_s)$, зависящим от параметра $\theta = (\theta_1, \theta_2, \dots, \theta_s) \in \Theta \subset R^s$. Параметр $\theta \in \Theta$ неизвестный и его необходимо оценить по выборке $(\xi_1, \xi_2, \dots, \xi_n)$.

Общим (важным как с точки зрения теории, так и приложений) методом построения оценок является метод максимального правдоподобия, предложенный Р. Фишером.

О п р е д е л е н и е. *Функцией максимального правдоподобия выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ будем называть функцию $L(\theta) = L(\theta_1, \theta_2, \dots, \theta_s)$ параметра $\theta \in \Theta$, определяемую равенством*

$$L(\theta) = f(\xi; \theta), \quad \theta \in \Theta,$$

если выборочный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ абсолютно непрерывный с плотностью $f(x; \theta) = f(x_1, x_2, \dots, x_n; \theta)$ и равенством

$$L(\theta) = P(\xi; \theta), \quad \theta \in \Theta,$$

если выборочный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ дискретный с распределением $P(x; \theta) = P(x_1, x_2, \dots, x_n; \theta)$.

Метод максимального правдоподобия построения оценок состоит в том, что в качестве оценки параметра $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ выбирается точка $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s)$, в которой функция максимального правдоподобия $L(\theta)$ достигает наибольшего значения.

О п р е д е л е н и е. *Оценкой максимального правдоподобия* будем называть точку $\hat{\theta}$, в которой функция максимального правдоподобия достигает наибольшего значения.

Другими словами, оценкой максимального правдоподобия параметра θ будем называть отличные от константы решения уравнения

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta),$$

если такие решения существуют. Корни, не зависящие от выборки $\xi_1, \xi_2, \dots, \xi_n$, т. е. имеющие вид $\hat{\theta} = c$, где c — константа, следует отбросить (оценка — это функция выборки).

Логарифм $\ln L(\theta)$ от функции максимального правдоподобия $L(\theta)$ называют *логарифмической функцией максимального правдоподобия*.

Заметим, что функции $L(\theta)$ и $\ln L(\theta)$ достигают наибольшего значения в одной и той же точке. А найти точку, в которой функция $\ln L(\theta)$ достигает наибольшего значения, зачастую проще.

Поэтому если функция $L(\theta) = L(\theta_1, \theta_2, \dots, \theta_s)$ дифференцируема по $\theta_1, \theta_2, \dots, \theta_s$, то для решения уравнения

$$L(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s) = \max_{\theta_1, \theta_2, \dots, \theta_s \in \Theta} L(\theta_1, \theta_2, \dots, \theta_s) \quad (15.3.1)$$

достаточно найти стационарные точки функции

$$\ln L(\theta_1, \theta_2, \dots, \theta_s),$$

решая уравнение

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, \dots, \theta_s) = 0, \quad i = 1, 2, \dots, s,$$

и, сравнивая значения функции $\ln L(\theta_1, \theta_2, \dots, \theta_s)$ в стационарных точках и на границе множества Θ , выбрать точку $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_s)$, в которой функция $\ln L(\theta_1, \theta_2, \dots, \theta_s)$ достигает наибольшего значения. Эта точка и будет решением уравнения (15.3.1).

Уравнение

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \theta_2, \dots, \theta_s) = 0, \quad i = 1, 2, \dots, s,$$

называют *уравнением максимального правдоподобия*.

Пример 15.3.1. По выборке $\xi_1, \xi_2, \dots, \xi_n$ из распределения с плотностью

$$p(x; \theta) = \begin{cases} \theta \exp\{-\theta x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

(показательное распределение с параметром θ) найти оценку максимального правдоподобия параметра θ .

Решение. Совместной плотностью распределения независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$ является

$$\begin{aligned} p(x_1, x_2, \dots, x_n; \theta) &= p(x_1; \theta)p(x_2; \theta) \dots p(x_n; \theta) = \\ &= \theta \exp\{-\theta x_1\} \theta \exp\{-\theta x_2\} \dots \theta \exp\{-\theta x_n\} = \end{aligned}$$

$$= \theta^n \exp \left\{ -\theta \sum_{i=1}^n x_i \right\},$$

если все $x_i > 0$, $i = 1, 2, \dots, n$, а если хотя бы одно $x_i \leq 0$, то

$$p(x_1, x_2, \dots, x_n; \theta) = 0.$$

Функцией правдоподобия выборки $\xi_1, \xi_2, \dots, \xi_n$ является

$$L(\theta) = \theta^n \exp \left\{ -\theta \sum_{i=1}^n \xi_i \right\},$$

если все $\xi_i > 0$, $i = 1, 2, \dots, n$, и $L(\theta) = 0$ — в противном случае. Функция $L(\theta)$ дифференцируема по θ , а поэтому можем выписать уравнение правдоподобия:

$$\frac{\partial}{\partial \theta} \ln L(\theta) = 0,$$

или (после дифференцирования)

$$n \frac{1}{\theta} - \sum_{i=1}^n \xi_i = 0.$$

Решением уравнения правдоподобия является

$$\theta = \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right)^{-1} = (\bar{\xi})^{-1}.$$

На границе области допустимых значений параметра θ функция $L(\theta)$ принимает значение 0, поэтому в точке $\hat{\theta} = 1/\bar{\xi}$ функция максимального правдоподобия $L(\theta)$ достигает наибольшего значения, а следовательно, $\hat{\theta} = 1/\bar{\xi}$ является оценкой максимального правдоподобия параметра θ .

Пример 15.3.2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{a} \exp \left\{ -\frac{1}{a}(x - b) \right\}, & \text{если } x > b; \\ 0, & \text{если } x \leq b \end{cases}$$

($f(x; a, b)$ — плотность смещенного показательного распределения).

Найти оценки параметров a и b методом максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия параметров a и b их несмещенными и состоятельными оценками?

Решение. Функция максимального правдоподобия

$$L(a, b) = \frac{1}{a^n} \exp \left\{ -\frac{1}{a} \sum_{i=1}^n (\xi_i - b) \right\},$$

если все $\xi_i > b$ и $L(a, b) = 0$, если существует $\xi_j \leq b$.

Если $\hat{b}$ — точка, в которой функция $Q(b) = \sum_{i=1}^n (\xi_i - b)$ ($\xi_i > b$, $i = 1, 2, \dots, n$) достигает наименьшего значения, а $\hat{a}$ — точка, в которой функция $L(a, \hat{b})$ достигает наибольшего значения, то в точке $(\hat{a}, \hat{b})$ функция $L(a, b)$ достигает наибольшего значения (в этом убеждаемся непосредственной проверкой).

Функция $Q(b) = \sum_{i=1}^n (\xi_i - b)$ ($\xi_i > b$, $i = 1, 2, \dots, n$)

достигает наименьшего значения в точке $\hat{b} = \min\{\xi_i\}$, а функция $L(a, \hat{b})$ достигает наибольшего значения в точке $\hat{a} = Q(\hat{b})/n$.

Оценки $\hat{a}, \hat{b}$ — асимптотически несмещенные и состоятельные соответственно для параметров a, b .

15.4 Задачи

АЗ: 15.3, 15.18, 15.28, 15.32, 15.33.

СЗ: 15.13, 15.24, 15.31, 15.38, 15.51.

15.1. Пусть 1,31; 1,18; 1,50; 1,06; 1,01; 1,06; 1,33; 1,80; 1,30; 1,35 — выборка из распределения с плотностью

$$p(x; a, b) = \begin{cases} \frac{1}{a} \exp \left\{ -\frac{1}{a}(x - b) \right\}, & \text{если } x > b; \\ 0, & \text{если } x \leq b \end{cases}$$

(смещенное показательное распределение) при значениях параметров $a = 0,5$; $b = 1$.

По выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$. Построить график функции $F(x)$ смещенного показательного распределения с параметрами $a = 0,5$; $b = 1$. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

15.2. На отрезок $[0;2]$ наудачу бросают точку и фиксируют ее координату ξ . Эксперимент проведен 15 раз. При этом получено 15 значений независимых наблюдений случайной величины ξ : 0,14; 1,81; 1,58; 1,28; 1,39; 0,80; 0,31; 0,70; 1,40; 0,08; 1,93; 0,79; 1,43; 1,42; 1,68.

Оценить по выборке математическое ожидание объема куба, ребро которого равно меньшей части отрезка $[0;2]$ при его делении точкой ξ .

15.3. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$p(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} \exp \left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\sigma > 0$ ($p(x; \mu, \sigma^2)$ — плотность логарифмически нормального распределения).

Найти оценку параметров μ и σ^2 методом моментов.

З а м е ч а н и е. Вычисляя первый и второй моменты, целесообразно сделать замену $\frac{\ln x - \mu}{\sigma} = t$.

15.4. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{1}{2h_0}, & \text{если } x \in [\theta - h_0, \theta + h_0]; \\ 0, & \text{если } x \notin [\theta - h_0, \theta + h_0]. \end{cases}$$

Найти оценку $\hat{\theta}$ параметра θ методом максимального правдоподобия.

Выяснить, является ли $\hat{\theta}$ несмещенной, состоятельной оценкой параметра θ .

15.5. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1), получить выборку объемом $n = 20$ из нормального распределения со средним $a = 2$ и дисперсией $\sigma^2 = 4$ (см. также пример 19.1.1). По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$. Построить график функции $N_{2;4}(x)$. Вычислить

$$\sup_x \left| N_{2;4}(x) - \hat{F}_n(x) \right|.$$

15.6. С помощью таблицы случайных чисел (см. табл. 22.10.1) оценить значение интеграла Пуассона

$$\int_{-\infty}^{+\infty} \exp \left\{ -\frac{x^2}{2} \right\} dx.$$

У к а з а н и е 1. Сделать в интеграле замену

$$y = \frac{1}{\pi} \operatorname{arctg} x + \frac{1}{2}.$$

У к а з а н и е 2. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1), получить выборку из равномерного на отрезке $[0; 1]$ распределения и применить теорему 15.1.1.

15.7. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, b) = \frac{1}{2\sqrt{\theta}} \exp \left\{ -\frac{1}{\sqrt{\theta}} |x - b| \right\},$$

параметр $\theta > 0$ ($f(x; \theta, b)$ — плотность смещенного двустороннего показательного распределения).

Найти оценки $\hat{\theta}$ и $\hat{b}$ параметров θ и b методом моментов.

Выяснить, являются ли $\hat{\theta}$ и $\hat{b}$ несмещенными и состоятельными оценками θ и b .

15.8. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta} \right)^r \left(\frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, \dots;$$

параметр $\theta > 0$, r — известно.

Найти оценку параметра θ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия $\hat{\theta}$ несмещенной, состоятельной, эффективной оценкой параметра θ .

У к а з а н и е. См. задачу 14.2.

15.9. Воспользовавшись таблицей случайных чисел, получить выборку объемом $n = 20$ из равномерного на отрезке $[0;1]$ распределения (см. задачу 19.4). По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$, $n = 20$. Построить график функции $F(x)$ равномерного на отрезке $[0;1]$ распределения. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

15.10. На отрезок $[0;4]$ наудачу бросают точку и фиксируют ее координату ξ . Эксперимент проведен 15 раз. При этом получено 15 значений независимых наблюдений случайной величины ξ : 0,55; 2,94; 0,65; 1,49; 3,08; 1,01; 2,38; 1,18; 1,63; 3,18; 0,39; 1,71; 2,72; 0,95; 1,18.

Оценить по выборке математическое ожидание площади поверхности шара, радиус которого равен меньшей части отрезка $[0;4]$ при его делении точкой ξ .

15.11. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из геометрического распределения с параметром p :

$$P(k; p) = (1 - p)^k p, \quad k = 0, 1, \dots$$

Оценить параметр p методом моментов.

15.12. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \frac{1}{2a} \exp \left\{ -\frac{1}{a} |x - b| \right\}, \quad a > 0.$$

Найти оценки параметров a и b методом максимального правдоподобия.

15.13. Выпишите 50 случайных, на ваш взгляд, чисел из отрезка $[0; 1]$ (для определенности — с четырьмя знаками после запятой). Их можно рассматривать как выборку из некоторого распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$, $n = 10$. Построить график функции $F(x)$ равномерного на отрезке $[0; 1]$ распределения. Вычислить

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

15.14. Длительность работы элемента до первого выхода из строя является показательно распределенной случайной величиной. Наблюдалась работа 20 элементов и фиксировали длительность их работы до первого выхода из строя (в часах): 11, 149, 846, 563, 384, 950, 864, 63, 990, 77, 685, 158, 348, 318, 25, 278, 1803, 83, 1544, 380.

По выборке найти оценку математического ожидания длительности безотказной работы элемента, если его меняют после 200 часов непрерывной работы, даже если элемент не вышел из строя.

15.15. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, \theta) =$$

$$= \begin{cases} (x - (a - \sqrt{\theta}))/\theta, & \text{если } x \in [a - \sqrt{\theta}, a]; \\ -(x - (a + \sqrt{\theta}))/\theta, & \text{если } x \in [a, a + \sqrt{\theta}]; \\ 0, & \text{если } x \notin [a - \sqrt{\theta}, a + \sqrt{\theta}]. \end{cases}$$

Найти оценки $\hat{a}$ и $\hat{\theta}$ параметров a и θ методом моментов.

Выяснить, являются ли $\hat{a}$ и $\hat{\theta}$ несмещенными и состоятельными оценками параметров a и θ соответственно.

У к а з а н и е. Вычисляя второй момент, целесообразно воспользоваться соотношением $\sigma^2 = m_2 - m_1^2$.

15.16. Сделано n выстрелов из орудия по неподвижной точечной цели без изменения прицела, причем в k ($0 < k < n$) случаях наблюдался перелет, а в остальных $n - k$ случаях — недолет снарядов.

Предположим, что высоты выстрелов являются независимыми, нормально распределенными случайными величинами с единичной дисперсией и математическим ожиданием h . Какую поправку необходимо внести в установку орудия, чтобы среднее значение высот выстрелов было возможно ближе к цели?

П р и м е ч а н и е. Здесь “высота выстрела” обозначает высоту точки попадания снаряда в вертикальную плоскость, проходящую через цель перпендикулярно к плоскости направления выстрелов. Высота выстрела отсчитывается от горизонтальной плоскости, в которой расположено орудие.

15.17. Получить выборку объемом $n = 10$ из $N_{0;1}$ -распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$, $n = 10$. Построить график функции $N_{0;1}(x)$ нормального распределения с параметрами $(0;1)$. Вычислить

$$\sup_x \left| N_{0;1}(x) - \hat{F}_n(x) \right|.$$

У к а з а н и е. Если случайная величина ξ имеет своей функцией распределения $F(x)$, то случайная величина $\eta = F(\xi)$ распределена равномерно на промежутке $[0; 1]$. Отсюда, если случайная величина η распределена равномерно на промежутке $[0; 1]$, то $\xi = F^{-1}(\eta)$ имеет своей функцией распределения $F(x)$.

15.18. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1) оценить значение интеграла

$$\int_0^{\infty} \frac{\sin \beta x}{\alpha^2 + x^2} dx,$$

$\alpha > 0$, $\beta > 0$, если $\alpha = 1$, $\beta = 1$.

Указание 1. Сделать в интеграле замену

$$y = \frac{2}{\pi} \operatorname{arctg} x.$$

Указание 2. По таблице получить выборку из равнономерного на отрезке $[0;1]$ распределения и воспользоваться теоремой 15.1.1.

15.19. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \frac{1}{2\sqrt{\theta}} \exp \left\{ -\frac{1}{\sqrt{\theta}} |x| \right\}$$

($f(x; \theta)$ — плотность двустороннего показательного распределения).

Найти оценку дисперсии методом моментов и выяснить является ли она несмещенной и состоятельной оценкой дисперсии.

15.20. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta_0 - h, \theta_0 + h]; \\ 0, & \text{если } x \notin [\theta_0 - h, \theta_0 + h]. \end{cases}$$

Найти оценку параметра h методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра h его несмещенной и состоятельной оценкой.

15.21. Наблюдаются цифровые части номеров 10 автомобилей (в номере пять цифр), проезжающих мимо

вас. Пусть a_i, b_i, c_i, d_i, e_i — цифры, встречающиеся в номере i -го автомобиля, если его читать слева направо. Строится последовательность чисел

$$\xi_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i + 10^{-5}e_i,$$

$i = 1, 2, \dots, 10$. Их можно рассматривать как выборку из некоторого распределения.

По выборке $\xi_1, \xi_2, \dots, \xi_{10}$ построить график эмпирической функции распределения $\hat{F}_n(x)$ ($n = 10$). Построить график функции $F(x)$ равномерного на отрезке $[0; 1]$ распределения. Вычислить

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

З а м е ч а н и е. Если номер автомобиля содержит четыре цифры, то рассмотрим последовательность чисел

$$\zeta_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i,$$

$i = 1, 2, \dots, 10$.

15.22. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, p) = \begin{cases} \frac{p}{\theta} \left(\frac{x}{\theta}\right)^{p-1} \exp \left\{ -\left(\frac{x}{\theta}\right)^p \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\theta > 0$ и $p > 0$ ($f(x; \theta, p)$ — плотность распределения Вейбулла), p — известно.

Найти оценку параметра θ методом максимального правдоподобия.

15.23. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного на отрезке $[\theta - \sigma\sqrt{3}, \theta + \sigma\sqrt{3}]$ распределения, $\sigma > 0$.

Найти оценки $\hat{\theta}$ и $\hat{\sigma}^2$ параметров θ и σ^2 методом моментов.

Выяснить, являются ли оценки $\hat{\theta}$ и $\hat{\sigma}^2$ несмещенными, состоятельными оценками соответствующих параметров.

15.24. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{x}{\theta} \exp \left\{ -\frac{x^2}{2\theta} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

($f(x; \theta)$ — плотность распределения Рэлея).

Найти оценку параметра θ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра θ его несмещенной, состоятельной, эффективной оценкой?

15.25. Имеем выборку: 2,22; 0,67; -0,14; -0,33; -0,97; -1,81; -0,94; -0,91; 0,11; 3,94 из распределения с плотностью

$$f(x; a, c) = \frac{a}{\pi(a^2 + (x - c)^2)},$$

где $a = 0,5$ и $c = 1$ ($f(x; a, c)$ — плотность распределения Коши).

Построить график реализации эмпирической функции распределения $\hat{F}_n(x)$. Построить график функции $F(x)$ распределения Коши с параметрами $a = 0,5$; $c = 1$. Вычислить

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

15.26. На отрезок $[0;3]$ наудачу бросают точку и фиксируют ее координату ξ . Эксперимент проведен 20 раз. При этом получено 20 значений независимых наблюдений случайной величины ξ : 1,91; 1,08; 1,25; 0,87; 2,42; 1,49; 2,10; 2,21; 0,33; 0,09; 0,39; 0,54; 2,39; 0,51; 2,51; 1,44; 0,96; 1,85; 2,38; 0,67.

Оценить по выборке математическое ожидание объема шара, радиус которого равен большей части отрезка $[0;3]$ при его делении точкой ξ .

15.27. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения Эрланга с параметрами m и λ . Оценить параметры m и λ методом моментов.

Плотность распределения Эрланга с параметрами m и λ имеет вид

$$f(x; \lambda, m) = \begin{cases} \frac{\lambda^m}{(m-1)!} x^{m-1} \exp\{-\lambda x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

15.28. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из равномерного распределения на промежутке:

1) $[\theta - 1, \theta + 1]$; 2) $[\theta, \theta + 1]$; 3) $[\theta - h_0, \theta + h_0]$.

Найти оценку $\hat{\theta}$ параметра θ методом моментов.

Выяснить, является ли оценка $\hat{\theta}$ несмещенной, состоятельной оценкой параметра θ .

15.29. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1), получить 10 независимых выборок объемом 12 из равномерного на промежутке $[0;1]$ распределения (см. также задачу 19.4). Обозначим их

$$\xi_{k,1}, \xi_{k,2}, \dots, \xi_{k,12}, \quad k = 1, 2, \dots, 10.$$

Пусть

$$\eta_k = \sum_{i=1}^{12} \xi_{k,i} - 6, \quad k = 1, 2, \dots, 10.$$

Последовательность $\eta_1, \eta_2, \dots, \eta_{10}$ является выборкой из некоторого распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_{10}(x)$. Построить график функции $N_{0;1}(x)$ нормального распределения с параметрами $(0;1)$. Вычислить

$$\sup_x \left| N_{0;1}(x) - \hat{F}_{10}(x) \right|.$$

15.30. Интервалы безотказной эксплуатации (часы между последовательными отказами) аппаратуры кондиционирования воздуха на самолете “Боинг-720” приведены в таблице (читать по строкам):

6	23	261	87	7	120	14	62	32	24
47	225	71	246	21	42	20	5	97	18
12	120	11	3	14	71	11	14	100	7
11	16	90	1	16	52	95	97	98	5
51	11	4	141	18	142	68	77	85	91
80	1	16	106	206	82	54	31	3	230
216	46	111	39	63	18	191	18	3	131
163	24	50	44	102	72	22	39	18	95
3	15	197	188	79	88	46	5	62	74
5	36	22	139	210	97	30	23	81	48

Найти эмпирическую оценку математического ожидания длительности безотказной работы аппаратуры, если ее профилактический ремонт осуществляется после 100 часов эксплуатации.

15.31. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \nu, \alpha) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\alpha x\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0; \end{cases}$$

$\nu \geq 0, \alpha > 0$ ($f(x; \nu, \alpha)$ — плотность гамма-распределения с параметрами $(\nu; \alpha)$).

Найти оценку параметров ν и α методом моментов.

15.32. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta} \right)^k, \quad k = 0, 1, \dots, \quad \theta > 0.$$

Найти оценку параметра θ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра θ его несмещенной, состоятельной, эффективной оценкой.

15.33. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1), получить выборку объемом 20 из распределения арксинуса (см. также пример 19.1.1). По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$, $n = 20$. Построить график функции распределения арксинуса

$$A(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{2}{\pi} \arcsin \sqrt{x}, & \text{если } x \in (0; 1); \\ 1, & \text{если } x > 1. \end{cases}$$

Вычислить

$$\sup_x \left| A(x) - \hat{F}_n(x) \right|.$$

15.34. На отрезок $[0;1]$ наудачу бросают точку и фиксируют ее координату ξ . Эксперимент проведен 20 раз. При этом получено 20 значений независимых наблюдений случайной величины ξ : 0,33; 0,93; 0,62; 0,38; 0,65; 0,53; 0,31; 0,07; 0,61; 0,04; 0,70; 0,85; 0,83; 0,81; 0,40; 0,59; 0,93; 0,24; 0,55; 0,17.

Оценить по выборке математическое ожидание площади круга, радиус которого равен большей части отрезка $[0;1]$ при его делении точкой ξ .

15.35. Обозначим через ξ случайную величину — число неудач до появления r -го успеха в неограниченной последовательности независимых испытаний с вероятностью успеха p в одном испытании. Случайная величина ξ имеет распределение

$$P\{\xi = k\} = C_{r-1+k}^{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

Это распределение называют отрицательным биномиальным с параметрами $(r; p)$ или распределением Паскаля.

Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из отрицательного биномиального распределения с параметрами $(r; p)$, r — известно. Найти оценку параметра p методом моментов.

У к а з а н и е. Общее количество неудач до r -го успеха можно представить в виде суммы

$$\xi = \eta_1 + \eta_2 + \dots + \eta_r,$$

где $\eta_1, \eta_2, \dots, \eta_r$ — независимые случайные величины, каждая из которых имеет геометрическое распределение с параметром p .

15.36. С целью оценить количество N рыб в озере, выловили K рыб, поместили их и выпустили в озеро. Через некоторое время в озере выловили n рыб. При этом k из них оказались помеченными.

Какое наиболее вероятное количество рыб в озере?

У к а з а н и е. Задачу решить в предположении, что количество N рыб в озере большое по сравнению с n . В этом

предположении случайную величину ξ — количество меченых рыб среди n выловленных — можно считать биномиально распределенной.

15.37. Стохастический эксперимент состоит в последовательном подбрасывании монеты четыре раза и регистрации результатов следующим образом: выпадение герба (Г) обозначается единицей, решки (Р) — нулем. Скажем, результаты экспериментов ГРРГ, РГРГ, и т. д., регистрируются соответственно как 1001, 0101, и т. д. Полученным последовательностям из нулей и единиц ставятся в соответствие числа из промежутка $[0;1]$, которые в двоичной системе исчисления можно представить так: 0,1001; 0,0101; и т. д., т. е. 0,1001 — это запись в двоичной системе числа

$$1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2^2} + 0 \cdot \frac{1}{2^3} + 1 \cdot \frac{1}{2^4} = \frac{9}{16},$$

а 0,0101 — числа

$$0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2^2} + 0 \cdot \frac{1}{2^3} + 1 \cdot \frac{1}{2^4} = \frac{5}{16}.$$

Эксперимент проведен 16 раз. Получено 16 чисел. Их можно рассматривать как выборку из некоторого распределения. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$ ($n = 16$). Построить график функции $F(x)$ равномерного на отрезке $[0;1]$ распределения. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

15.38. Воспользовавшись таблицей случайных чисел (см. табл. 22.10.1), оценить значение интеграла

$$\int_0^1 \frac{\operatorname{arctg} x}{x} dx = G.$$

Указание. С помощью таблицы случайных чисел получить выборку из равномерного на отрезке $[0;1]$ распределения и воспользоваться теоремой 15.1.1.

Значение $G = 0,915965\dots$ известно под названием “постоянной Каталана”.

15.39. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{a} \exp \left\{ -\frac{1}{a}(x - b) \right\}, & \text{если } x > b; \\ 0, & \text{если } x \leq b. \end{cases}$$

Найти оценки параметров a и b методом моментов.

15.40. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; a, b) = \begin{cases} \frac{1}{b - a}, & \text{если } x \in [a; b]; \\ 0, & \text{если } x \notin [a; b]. \end{cases}$$

Найти оценки параметров a и b методом максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия $\hat{a}$ и $\hat{b}$ параметров a и b их несмещенными оценками? Состоятельными оценками?

15.41. Время, показываемое остановившимися механическими часами, естественно считать случайной величиной, распределенной равномерно на отрезке $[0; 12]$.

Ниже приведены показания 14 часов в витрине часового магазина: 9 ч 18 мин; 8 ч 35 мин; 10 ч 45 мин; 11 ч 30 мин; 3 ч 06 мин; 7 ч 50 мин; 4 ч 22 мин; 10 ч 12 мин; 7 ч 47 мин; 4 ч 28 мин; 11 ч 16 мин; 7 ч 08 мин; 5 ч 53 мин; 8 ч 00 мин. По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$. Построить график функции $F(x)$ равномерного на отрезке $[0; 12]$ распределения. Вычислить

$$\sup_x |F(x) - \hat{F}_n(x)|.$$

15.42. На отрезок $[0; 1]$ наудачу бросают точку и фиксируют ее координату ξ . Эксперимент проведен 10 раз. При этом получено 10 значений независимых наблюдений случайной величины ξ : 0,33; 0,93; 0,62; 0,38; 0,65; 0,53; 0,81; 0,07; 0,61; 0,04.

Оценить по выборке математическое ожидание площади квадрата со стороной, равной меньшей части отрезка $[0;1]$ при его делении точкой ξ .

15.43. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из биномиального распределения

$$P(k; m) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m,$$

m известно.

Найти оценку $\hat{p}$ параметра p методом моментов. Выяснить, является ли оценка $\hat{p}$ несмещенной, состоятельной и эффективной оценкой p .

15.44. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \frac{1}{2h}, & \text{если } x \in [\theta - h; \theta + h]; \\ 0, & \text{если } x \notin [\theta - h; \theta + h] \end{cases}$$

($f(x; \theta, h)$ — плотность равномерного на отрезке $[\theta - h, \theta + h]$ распределения).

Найти оценки параметров θ и h методом максимального правдоподобия.

Выяснить, являются ли оценки максимального правдоподобия $\hat{\theta}$ и $\hat{h}$ параметров θ и h несмещенными оценками? Состоятельными оценками?

15.45. Имеем выборку: 0,04; 1,95; 0,38; 0,20; 0,24; 1,14; 0,10; 1,96; 1,00; 0,07 из распределения с плотностью

$$f(x; a) = \begin{cases} \frac{1}{a} \exp \left\{ -\frac{x}{a} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

По этой выборке построить график реализации эмпирической функции распределения $\hat{F}_n(x)$. Построить график показательной функции распределения $F(x)$ с параметром 1. Вычислить

$$\sup_x \left| F(x) - \hat{F}_n(x) \right|.$$

15.46. Пользуясь таблицей случайных чисел (см. табл. 22.10.1), оценить значение интеграла

$$\int_0^{\infty} \frac{\cos \beta x}{\alpha^2 + x^2} dx,$$

если $\alpha = 1$, $\beta = 1$.

У к а з а н и е 1. Сделать в интеграле замену

$$y = \frac{2}{\pi} \operatorname{arctg} x.$$

У к а з а н и е 2. С помощью таблицы случайных чисел получить выборку из равномерного на отрезке $[0; 1]$ распределения и воспользоваться теоремой 15.1.1.

15.47. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, h) = \begin{cases} \frac{1}{2\sqrt{3h}}, & \text{если } x \in [\theta - \sqrt{3h}; \theta + \sqrt{3h}]; \\ 0, & \text{если } x \notin [\theta - \sqrt{3h}; \theta + \sqrt{3h}]. \end{cases}$$

Найти оценки $\hat{\theta}$ и $\hat{h}$ параметров θ и h методом моментов.

Выяснить, являются ли $\hat{\theta}$ и $\hat{h}$ несмещенными, состоятельными оценками параметров θ и h соответственно.

15.48. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} \exp \left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0, \end{cases}$$

параметр $\sigma > 0$ ($f(x; \mu, \sigma^2)$ — плотность логарифмически нормального распределения).

Найти оценки параметров μ и σ^2 методом максимального правдоподобия.

15.49. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = C_{r-1+k}^{r-1} (1/\theta)^r (1 - 1/\theta)^k, \quad k = 0, 1, \dots,$$

r известно (см. также задачу 15.35).

Найти оценку параметра θ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра θ его несмещенной, состоятельной, эффективной оценкой.

15.50. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; b) = \begin{cases} \frac{1}{b - a_0}, & \text{если } x \in [a_0; b]; \\ 0, & \text{если } x \notin [a_0; b]. \end{cases}$$

Найти оценку параметра b методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра b его несмещенной и состоятельной оценкой.

15.51. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения Пуассона

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Найти оценку параметра λ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра λ его несмещенной, состоятельной, эффективной оценкой.

15.52. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения

$$P(k; \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta} \right)^r \left(\frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, \dots;$$

параметр $\theta > 0$, r известно.

Найти оценку $\hat{\theta}$ параметра θ методом моментов.

Выяснить, является ли $\hat{\theta}$ несмещенной, состоятельной, эффективной оценкой параметра θ .

У к а з а н и е. См. задачу 14.2.

15.53. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из биномиального распределения

$$P(k; m) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m,$$

m известно.

Оценить параметр p методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра p его несмещенной, состоятельной и эффективной оценкой.

15.54. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} \frac{x}{\theta^2} \exp \left\{ -\frac{x^2}{2\theta^2} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Найти оценку $\hat{\theta}$ параметра θ методом моментов.

Выяснить, является ли $\hat{\theta}$ несмещенной и состоятельной оценкой параметра θ .

15.55. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \theta, \lambda) = \begin{cases} \frac{\theta \lambda^\theta}{x^{\theta+1}}, & \text{если } x > \lambda; \\ 0, & \text{если } x \leq \lambda, \end{cases}$$

параметр $\theta > 1$, параметр $\lambda > 0$ ($f(x; \theta, \lambda)$ — плотность распределения Парето).

Найти оценки параметров θ и λ методом максимального правдоподобия.

15.56. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения с плотностью

$$f(x; \sigma^2) = \begin{cases} \sqrt{\frac{2}{\pi \sigma^2}} \exp \left\{ -\frac{x^2}{2\sigma^2} \right\}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0 \end{cases}$$

($f(x; \sigma^2)$ — плотность полунормального распределения).

Найти оценку параметра $\theta = \sigma^2$ методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра σ^2 его несмещенной, состоятельной и эффективной оценкой.

15.57. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения y

$$Q(x) = p^x(1-p)^{1-x}, \quad x = 0; 1, \quad p \in (0; 1).$$

Найти оценку параметра p методом максимального правдоподобия.

Выяснить, является ли оценка максимального правдоподобия параметра p его несмещенной, состоятельной, эффективной оценкой.

15.58. Частота $\hat{p} = \mu/n$ события A в последовательности экспериментов является несмещенной и состоятельной оценкой его вероятности $p = P(A)$.

Выяснить, является ли частота $\hat{p} = \mu/n$ эффективной оценкой p .

Глава 16

Задача проверки статистических гипотез

16.1 Критерий, функция мощности критерия

Задача проверки статистических гипотез. Часто задачи математической статистики формулируют в следующем виде. Имеется результат $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ стохастического эксперимента, состоящего в наблюдении случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ со значениями в $\mathbb{R}^n$. Относительно распределения случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, или, что то же, выборки, в лучшем случае известно лишь то, что оно принадлежит некоторому классу распределений $\mathcal{P}$. Далее класс $\mathcal{P}$ будем считать параметрическим:

$$\mathcal{P} = \{P_\theta; \theta \in \Theta \subset \mathbb{R}^s\}.$$

По значению $\xi(\omega)$, принятому случайной величиной ξ , необходимо выбрать ее распределение из класса $\mathcal{P}$.

Будем поступать следующим образом. Из класса $\mathcal{P}$ возможных распределений случайной величины ξ выберем некоторое распределение G — “кандидатуру” на распределение ξ . Другими словами, относительно распределения случайной величины ξ выдвигаем гипотезу (предположение): распределением ξ является G . Затем проводим эксперимент — получаем реализацию $\xi(\omega)$ случайной

величины ξ и по значению $\xi(\omega)$, принятому ξ , делаем заключение: случайная величина ξ может иметь своим распределением G или G не может быть распределением ξ . При этом будем поступать так, как в свое время Сократ, когда хотел опровергнуть утверждение своего оппонента. А именно, если гипотеза: G является распределением ξ — противоречит результату эксперимента, то ее следует отклонить (G не может быть распределением ξ), в противном случае — гипотезу не отклоняем (G может быть распределением ξ).

Сформулированная задача относится к задачам проверки статистических гипотез.

Соглашения и определения. В теории проверки статистических гипотез приняты такие определения и соглашения.

Гипотезы относительно распределений случайных величин называют *статистическими*.

Выбор распределения (или класса распределений) из совокупности $\mathcal{P}$ будем называть *выбором основной (нулевой) гипотезы* относительно распределения случайной величины ξ со значениями в пространстве $\mathbb{R}^n$. После выбора нулевой гипотезы выбирают гипотезу альтернативную к нулевой (из множества возможных альтернативных гипотез). Альтернативная (конкурирующая) к H_0 гипотеза — это гипотеза, в пользу которой отклоняют H_0 (если H_0 отклоняют). В качестве множества альтернативных гипотез, как правило, будем рассматривать параметрическое семейство, обозначая его через $H_\theta, \theta \in \Theta$.

Проиллюстрируем введенные понятия на примере биномиального распределения.

Монету, вероятность выпадения герба θ которой неизвестна, независимым образом бросают n раз. Число выпавших гербов оказалось равным ξ . Можно ли считать, что монета симметрична?

Сформулируем эту задачу в терминах проверки статистических гипотез.

Распределением случайной величины ξ может быть любое распределение из класса $\mathcal{P} = \{P_\theta, \theta \in (0, 1)\}$, где

$$P_\theta(k) = C_n^k \theta^k (1 - \theta)^{n-k}, \quad k = 0, 1, \dots, n.$$

Поскольку мы интересуемся симметричностью монеты,

то в качестве кандидатуры на распределение случайной величины ξ из семейства распределений $\mathcal{P}$ выберем распределение $P_{0,5}(k)$, $k = 0, 1, \dots, n$. Тем самым выбрана нулевая гипотеза (H_0 : распределением ξ является $P_{0,5}(k)$, $k = 0, 1, \dots, n$). Конкурирующими к H_0 являются гипотезы H_θ : распределением ξ есть $P_\theta(k)$, $k = 0, 1, \dots, n$, $\theta \in (0, 1)$, $\theta \neq 0, 5$.

Ответ на вопрос о симметричности монеты сводится к проверке гипотезы H_0 .

Гипотеза относительно распределения, которая однозначно его определяет, называется *простой*; если гипотеза не определяет распределение однозначно, она называется *сложной*.

Например, гипотеза H_0 : случайная величина ξ имеет распределение

$$P_{\theta_0}(k) = C_n^k \theta_0^k (1 - \theta_0)^{n-k}, \quad k = 0, 1, \dots, n,$$

где θ_0 — фиксировано, простая. Гипотеза H : случайная величина ξ имеет распределение

$$P_\theta(k) = C_n^k \theta^k (1 - \theta)^{n-k}, \quad k = 0, 1, \dots, n,$$

где $\theta \in (0; 1/2)$, сложная.

Формулируя задачу проверки статистических гипотез, в качестве нулевой гипотезы (для наглядности и простоты) мы рассмотрели простую гипотезу: распределением случайной величины ξ является G , где G — полностью определенное распределение.

Заметим, что математическая статистика не дает рекомендаций относительно выбора нулевой гипотезы, этот выбор определяется поставленной задачей и исследователем.

Критерий. Пусть ξ — случайная величина со значениями в $\mathbb{R}^n$. Относительно неизвестного распределения случайной величины ξ выдвигается гипотеза H_0 : распределением ξ является G ; альтернативная гипотеза: G не является распределением ξ .

Необходимо проверить гипотезу H_0 , т. е. сделать вывод: G может быть распределением случайной величины ξ (будем говорить “гипотеза H_0 не отклоняется”) или G

не может быть распределением случайной величины ξ (будем говорить “гипотеза H_0 отклоняется”).

Согласно нашему общему подходу, проводим стохастический эксперимент — получаем реализацию $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ случайной величины (выборку) $\xi = (\xi_1, \xi_2, \dots, \xi_n)$. Далее, если гипотеза H_0 противоречит результату эксперимента, ее отклоняем, в противном случае — не отклоняем (отметим, что кроме реализации $\xi(\omega)$ выборки ξ , мы не имеем ничего, что бы несло информацию о распределении ξ).

Чтобы по результату эксперимента можно было выносить заключение об отклонении или неотклонении гипотезы H_0 , необходимо определить (задать) множество тех исходов $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ эксперимента, которым гипотеза противоречит. Пусть $S \subset \mathbb{R}^n$ — множество таких исходов. Далее проводим эксперимент. Если при этом его исход $\xi(\omega)$ оказался среди исходов из множества S (которым гипотеза H_0 противоречит) — гипотезу H_0 отклоняем, в противном случае — нет.

О п р е д е л е н и е. Борелевское множество S выборочного пространства $\mathbb{R}^n$ такое, что при $\xi \in S$ гипотеза H_0 отклоняется, а при $\xi \notin S$ — не отклоняется, будем называть *критическим множеством* (*критической областью*) или *критерием для проверки гипотезы H_0* .

Далее, естественно, возникает задача (это основная задача проверки статистических гипотез): как выбрать критическое множество $S \subset \mathbb{R}^n$ для проверки данной гипотезы (борелевских множеств S в пространстве $\mathbb{R}^n$ много и ясно, что не все они одинаково хороши как критерии для проверки гипотезы H_0).

Чтобы указать пути решения этой задачи, сначала рассмотрим так называемые ошибки первого и второго рода и вероятности этих ошибок.

Ошибки первого и второго рода. Пусть H_0 — нулевая гипотеза относительно распределения случайной величины ξ , $\xi(\omega)$ — реализация ξ , S — борелевское множество в пространстве $\mathbb{R}^n$. Будем проверять H_0 , пользуясь множеством S как критическим, а именно, если исход $\xi(\omega)$ эксперимента попадает в S , то гипотезу H_0 отклоняем, в противном случае — нет.

При этом возможны такие ситуации: гипотеза H_0 верна или неверна, реализация $\xi(\omega)$ случайной величины ξ попала в S или нет. Рассмотрим их подробнее.

1° Гипотеза H_0 верна. Реализация $\xi(\omega)$ не попала в S , поэтому согласно критерию S гипотезу H_0 не отклоняем.

2° Гипотеза H_0 неверна. Реализация $\xi(\omega)$ не попала в S ; и, следовательно, согласно критерию S гипотезу H_0 не отклоняем.

3° Гипотеза H_0 верна. Реализация $\xi(\omega)$ попала в S , и поэтому согласно критерию S гипотезу H_0 отклоняем.

4° Гипотеза H_0 неверна. Реализация $\xi(\omega)$ попала в S , и, следовательно, согласно критерию S гипотезу H_0 отклоняем.

Из четырех ситуаций две (1° и 4°) удовлетворительны, а остальные (2° и 3°) — неудовлетворительны. В ситуациях 2° и 3° мы говорим, что в результате использования критерия S допускается ошибка. При этом ошибки в ситуации 2° (гипотеза H_0 не отклоняется, когда она неверна) и в ситуации 3° (гипотеза H_0 отклоняется, когда она верна) качественно различны.

Проиллюстрируем отличие между ошибками в ситуациях 2° и 3° на примере проверки медицинского препарата на токсичность биологическими методами.

Исследуя препарат на токсичность, определенную его дозу вводят подопытным животным (кроликам) и регистрируют количество летальных исходов (заметим, что это число является случайной величиной). Необходимо по реализации $\xi(\omega)$ случайной величины ξ (количеству летальных исходов) сделать вывод относительно токсичности препарата — “препарат токсичен” или “препарат нетоксичен”. Ясно, что когда число $\xi(\omega)$ летальных исходов большое, препарат следует классифицировать как токсичный, в противном случае — нет.

Задачу исследования препарата на токсичность биологическими методами можно сформулировать в терминах проверки статистических гипотез. А именно, относительно токсичности препарата выдвигаются гипотезы: H_0 — препарат токсичен (альтернативная гипотеза H_1 — препарат нетоксичен). Необходимо проверить гипотезу H_0 , т. е. отклонить ее или не отклонить. Выбор между этими действиями осуществляют по реализации $\xi(\omega)$

случайной величины ξ — количеству летальных исходов: если $\xi(\omega) \in S = \{x : x \leq k\}$, то гипотезу отклоняют, в противном случае — нет (S — критерий для проверки гипотезы H_0).

Как и в каждой задаче проверки статистических гипотез, в рассматриваемой задаче возможны ошибки:

гипотеза H_0 неверна, но согласно критерию она не отклоняется;

гипотеза H_0 верна, но согласно критерию она отклоняется.

Посмотрим, какие последствия этих ошибок и какова их “цена”.

1. Пусть допущена ошибка: гипотеза H_0 неверна, но согласно критерию не отклоняется. Утверждение “гипотеза H_0 неверна” в рассматриваемой задаче обозначает, что препарат нетоксичный (не является опасным для здоровья пациентов), а утверждение “ H_0 не отклоняется” обозначает, что препарат классифицируется как токсичный. Таким образом, нетоксичный препарат согласно критерию классифицируется как токсичный и возвращается поставщику (для переработки или уничтожения). Последствия ошибки такого рода выражаются в финансовых убытках, увеличении стоимости препарата (“цена” ошибки — финансовые потери).

2. Пусть допущена ошибка: гипотеза H_0 верна, но согласно критерию отклоняется. Утверждение “гипотеза H_0 верна” обозначает, что препарат нетоксичен (опасен для здоровья пациентов). Утверждение “гипотеза H_0 отклоняется” обозначает, что препарат классифицируется как нетоксичный (безопасный для здоровья пациентов). Таким образом, токсичный препарат (опасный для здоровья пациентов) согласно критерию классифицируют как нетоксичный (безопасный для здоровья пациентов) и отправляют в продажу. Последствия ошибки такого рода могут привести к смерти пациентов, использующих этот препарат (“цена” ошибки — летальный исход для пациента).

Этот пример показывает, что описанные выше ошибки существенно отличаются своей “ценой”.

О п р е д е л е н и е. Ошибка, состоящая в том, что гипотеза H_0 отклоняется, когда она верна, называется *ошибкой первого рода*.

Ошибка, состоящая в том, что гипотеза H_0 не отклоняется, когда она неверна, называется *ошибкой второго рода*.

О выборе нулевой гипотезы. Как уже отмечалось, нулевую гипотезу из семейства всех возможных гипотез мы выбираем сами. Этот выбор тесно связан с ошибками, которые мы допускаем, проверяя гипотезы. (Об ошибках, связанных с проверкой гипотез, и состоящих в том, что данную гипотезу H_θ ($\theta \in \Theta$) мы отклоняем, когда она верна, можно говорить и не выбрав нулевую гипотезу.) Как правило, в качестве нулевой гипотезы выбираем ту, для которой важнее избежать ошибки, состоящей в ее отклонении, когда она верна.

Уровень значимости критерия, функция мощности критерия. Можно ли построить критическое множество (критерий) для проверки гипотезы, использование которого не приводило бы к ошибкам? Нет, нельзя. Поскольку каким бы ни было критическое множество $S \neq \emptyset$, значение $\xi(\omega)$ случайной величины ξ (результата стохастического эксперимента) может попасть в S , когда гипотеза H_0 верна, при этом будет допущена ошибка первого рода. Значение $\xi(\omega)$ может попасть и в $\bar{S}$, когда H_0 неверна (при этом будет допущена ошибка второго рода). И поскольку построить критерий для проверки гипотезы H_0 , который бы не приводил к ошибкам, невозможно в принципе, мы, естественно, будем стремиться строить такие критерии, которые бы гарантировали минимальную частоту ошибок при их использовании.

Пусть H_0 — основная гипотеза, H_1 — конкурирующая (для наглядности предположим, что H_0 и H_1 простые), S — критерий для проверки H_0 . Пользуясь критерием S , мы можем допустить ошибки двух типов: H_0 верна, но согласно критерию S отклоняется (ошибка первого рода); H_0 неверна, но согласно критерию S не отклоняется (ошибка второго рода). Вероятность ошибки первого рода равна вероятности значению ξ попасть в критическое множество S , когда гипотеза H_0 верна, т. е. $P\{\xi \in S | H_0\}$ ($P\{\xi \in S | H_0\}$ коротко будем записывать так: $P(S | H_0)$).

Вероятность ошибки второго рода равна вероятности выборочному значению попасть в множество $\bar{S}$, когда гипотеза H_1 верна, т. е. $P\{\xi \in \bar{S}|H_1\}$ ($P\{\xi \in \bar{S}|H_1\}$ коротко будем записывать так: $P(\bar{S}|H_1)$).

Заметим, что $P\{\xi \in S|H_0\}$ — это вероятность события $\{\xi \in S\}$, вычисленная в предположении, что гипотеза H_0 верна; она не имеет ничего общего с условной вероятностью.

Вероятности ошибок первого и второго рода однозначно определяются указанием критического множества S . Поэтому, выбирая S , скажем, из условия, что вероятность ошибки первого рода мала, мы одновременно получаем вероятность ошибки второго рода, которая определяется выбранным критическим множеством S и будет такой, какой получится. Можно выбрать S и из условия малости вероятности ошибки второго рода, но при этом вероятность ошибки первого рода однозначно определяется выбранным ранее множеством S . Следовательно, выбрать S так, чтобы можно было одновременно контролировать вероятности ошибок первого и второго рода, невозможно. Поэтому будем поступать следующим образом. Поскольку важнее избежать ошибки первого рода (ее “цена” выше), наше первое требование к критическому множеству S будет состоять в том, чтобы вероятность ошибки первого рода $P(S|H_0)$ была малой. (Это обозначает, что используя критерий S в длинной серии экспериментов, верную гипотезу H_0 будем отклонять изредка.) Формализуем это требование.

Фиксируем малое α и выберем критическое множество S так, чтобы вероятность ошибки первого рода не превосходила α :

$$P(S|H_0) \leq \alpha.$$

Если последнее неравенство удовлетворяет не одно множество S , то окончательно критерий выбираем так, чтобы вероятность $P(S|H_1)$ отклонения неверной гипотезы H_0 была максимальной. Для этого множество S выберем “пошире”.

О п р е д е л е н и е. Число α , ограничивающее сверху вероятность ошибки первого рода, называется *уровнем значимости*.

Если критическое множество S удовлетворяет условию

$$P(S|H_0) \leq \alpha,$$

то будем говорить, что S соответствует уровню значимости α .

О п р е д е л е н и е. Вероятность $P(S|H_1)$ отклонить основную гипотезу, когда верна альтернативная гипотеза H_1 , называется *мощностью критерия S* .

Если альтернативная гипотеза сложная, причем когда она верна, верна одна из простых гипотез $H_\theta, \theta \in \Theta$, то для каждого $\theta \in \Theta$ можно вычислить $P(S|H_\theta)$.

О п р е д е л е н и е. Функция

$$\beta(\theta) = P(S|H_\theta), \theta \in \Theta,$$

при каждом $\theta \in \Theta$ равная вероятности отклонить основную гипотезу H_0 , если верна гипотеза H_θ , называется *функцией мощности критерия*.

З а м е ч а н и е. Вопрос о том, каким должен быть уровень значимости, не является статистической задачей. Обычно уровень значимости полагают равным 0,10; 0,05; 0,01; 0,001. Чем серьезнее последствия ошибки первого рода, тем меньше должен быть уровень значимости.

П р и м е р 16.1.1 (выборочный контроль). *Предназначенная для продажи партия изделий в количестве $N = 10\,000$ штук проходит выборочный контроль на качество. Поставщик уверен, что доля дефектных изделий равна 1% (или меньше), и хочет, чтобы всякий раз, когда доля дефектных изделий составляет 1%, вероятность того, что партия выдерживает контроль, была равна 0,9. Покупатель считает, что партию целесообразно закупить даже в том случае, если доля дефектных изделий будет больше 1%, но 6% дефектных изделий он считает предельно допустимой долей и хочет, чтобы контроль обнаруживал партии с 6%-м содержанием дефектных изделий с вероятностью 0,95. Поставщик и покупатель договорились осуществлять контроль следующим образом: из партии в 10 000 изделий извлекают случайную выборку объемом n ; если при этом количество $\xi(\omega)$ дефектных изделий в выборке окажется малым (меньше некоторого l), то партия выдерживает контроль и закупается, в противном случае — нет.*

Каким должен быть объем n выборки из партии изделий и значение l ?

Ответить на эти вопросы, сформулировав и решив поставленную задачу как задачу проверки статистических гипотез.

Решение. Решение о закупке или отклонении партии изделий будем принимать по реализации $\xi(\omega)$ случайной величины ξ — количеству дефектных изделий из n выбранных.

Изделия в выборку объемом n извлекают последовательно. Поскольку объем N партии большой ($N = 10\,000$ изделий), а n по отношению к N мало, то можно считать, что случайная величина ξ имеет биномиальное распределение с параметрами $(n; p)$:

$$P\{\xi = k\} = P_p(k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

(параметр p — вероятность выбора дефектного изделия — неизвестен). Другими словами, семейством возможных распределений случайной величины ξ является

$$\mathcal{P} = \{P_p : p \in (0; 1)\}.$$

Отметим, что распределение из семейства $\mathcal{P}$ однозначно определяется значением параметра p и наоборот.

По реализации $\xi(\omega)$ случайной величины ξ необходимо сделать заключение о распределении ξ или, что то же, о значении параметра p . Представляют интерес такие значения параметра p : $p_1 = 0,01$ и $p_2 = 0,06$ (им соответствует 1 %-й и 6 %-е содержание дефектных изделий в партии).

Согласно изложенной методике из семейства возможных гипотез $\mathcal{P} = \{P_p, p \in (0; 1)\}$ необходимо выбрать нулевую, а затем проверить ее.

Поскольку между покупателем и поставщиком существует соглашение: “если количество $\xi(\omega)$ дефектных изделий в выборке мало, партия выдерживает контроль и ее закупают, в противном случае — нет”, то в качестве нулевой выбираем гипотезу “ ξ имеет биномиальное распределение с параметрами $(n; 0,06)$ ” (альтернативная гипотеза $P_p \in \{P_p : p \in (0; 0,06)\}$). Неотклонение нулевой гипотезы обозначает классификацию партии как такой, которая содержит 6 % (или больше) дефектных изделий, с

последующим отказом в закупке партии. (Если при этом какая-то из партий с малым процентом дефектных изделий будет забракована, то это уже проблема поставщика.) Отклонение нулевой гипотезы в пользу альтернативы $P_p \in \{P_p : p \in (0; 0,06)\}$ обозначает классификацию партии как такой, которая содержит меньше 6 % дефектных изделий, и последующую закупку партии. (Если при этом какая-то из партий с большим процентом дефектных изделий не будет забракована, и следовательно, будет закуплена, покупатель, разумеется, понесет убытки.)

Для проверки H_0 в качестве критического естественно выбрать множество вида $S = \{k : k \leq l\}$. Это множество определяется числом l и объемом выборки n .

По договоренности между покупателем и поставщиком множество $S = \{k : k \leq l\}$, а фактически числа n и l , должны быть такими, чтобы выполнялись приведенные далее условия, а именно: партия с 6 %-м содержанием дефектных изделий должна обнаруживаться с вероятностью 0,95, т. е.

$$P(\bar{S}|H_0) \geq 0,95,$$

или, что то же,

$$\begin{aligned} P(S|H_0) &= P\{\xi \in S|H_0\} = P_0\{\xi \leq l\} = \\ &= \sum_{k=0}^l C_n^k (0,06)^k (0,94)^{n-k} \leq \alpha = 0,05. \end{aligned}$$

(Последнее неравенство — формализованная запись требований покупателя, согласно которым партии с 6 %-м содержанием брака должны обнаруживаться с вероятностью, не меньшей чем 0,95.)

А значение $\beta(p) = P(S|H_p)$ функции мощности критерия S в точке $p = 0,01$ должно быть не меньшим чем 0,9:

$$\begin{aligned} \beta(0,01) &= P(S|H_{0,01}) = P\{\xi \in S|H_{0,01}\} = \\ &= P\{\xi \leq l|H_{0,01}\} = \sum_{k=0}^l C_n^k (0,01)^k (0,99)^{n-k} \geq 0,9. \end{aligned}$$

(Последнее неравенство является формализованной записью пожеланий поставщика, согласно которым партии с 1%-м содержанием брака должны выдерживать контроль с вероятностью, не меньшей 0,9.)

Таким образом, искомые n и l должны удовлетворять соотношениям

$$\sum_{k=0}^l C_n^k (0,06)^k (0,94)^{n-k} \leq 0,05; \quad (16.1.1)$$

$$\sum_{k=0}^l C_n^k (0,01)^k (0,99)^{n-k} \geq 0,9. \quad (16.1.2)$$

В рассматриваемой задаче биномиальное распределение

$$B_{n,p}(m) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, 1, \dots, n,$$

можно аппроксимировать пуассоновским, поскольку параметр p малый (вероятность p выбора дефектного изделия мала), а n — объем выборки — большой. Основанием для этого является теорема Пуассона. Напомним ее содержание.

Пусть $np \rightarrow \lambda > 0$, при $n \rightarrow \infty$. Тогда для каждого фиксированного m , $m = 0, 1, \dots$,

$$\lim_{n \rightarrow \infty} C_n^m p^m (1-p)^{n-m} = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Чтобы погрешность аппроксимации биномиального распределения пуассоновским была незначительной, n должно быть не меньше нескольких десятков (а лучше сотен), а произведение np должно удовлетворять неравенствам

$$1 \leq np \leq 10.$$

Собственно говоря, мы с самого начала в качестве распределения случайной величины ξ — количества дефектных изделий в выборке — могли бы рассматривать пуассоновское распределение (поскольку вероятность успеха

мала, а количество испытаний большое). Таким образом, будем исходить из того, что n и l должны удовлетворять неравенствам

$$\sum_{k=0}^l \frac{\lambda_0^k}{k!} e^{-\lambda_0} \leq 0,05, \quad \lambda_0 = n0,06; \quad (16.1.3)$$

$$\sum_{k=0}^l \frac{\lambda_1^k}{k!} e^{-\lambda_1} \geq 0,90, \quad \lambda_1 = n0,01 \quad (16.1.4)$$

(сравните с неравенствами (16.1.1) и (16.1.2)). Определяя n и l , мы, естественно, будем стремиться выбрать n по возможности меньшим.

Посмотрим, можно ли для $n = 50$ определить l так, чтобы выполнялись соотношения (16.1.3) и (16.1.4). Если нет, выбираем n большим, и так поступаем до тех пор, пока не найдем n и l , которые будут удовлетворять неравенствам (16.1.3) и (16.1.4).

При $n = 50$ имеем $\lambda_0 = 50 \cdot 0,06 = 3$; $\lambda_1 = 50 \cdot 0,01 = 0,5$. Пользуясь таблицей распределения Пуассона (см. табл. 22.6.1), находим, что неравенство (16.1.3) удовлетворяется при $l = 0$, а (16.1.4) — при $l \geq 1$. Поэтому для $n = 50$ не существует l , которое бы удовлетворяло неравенствам (16.1.3) и (16.1.4).

При $n = 100$ имеем $\lambda_0 = 100 \cdot 0,06 = 6$; $\lambda_1 = 100 \times 0,01 = 1$. Пользуясь таблицей распределения Пуассона находим, что неравенство (16.1.3) удовлетворяется, при $l \leq 1$, а (16.1.4) — при $l \geq 2$. Поэтому для $n = 100$ не существует l , которое бы удовлетворяло неравенствам (16.1.3) и (16.1.4).

При $n = 150$ имеем $\lambda_0 = 150 \cdot 0,06 = 9$; $\lambda_1 = 150 \times 0,01 = 1,5$. Для таких λ_0 и λ_1 неравенство (16.1.3) удовлетворяется, при $l \leq 4$, а неравенство (16.1.4) — при $l \geq 3$. Попытаемся уменьшить n . Пусть $n = 130$. Тогда $\lambda_0 = 130 \cdot 0,06 = 7,8$; $\lambda_1 = 130 \cdot 0,01 = 1,3$; неравенство (16.1.3) удовлетворяется, если $l \leq 3$, а неравенство (16.1.4) — если $l \geq 3$. Таким образом, для $n = 130$ и $l = 3$ оба неравенства (16.1.3) и (16.1.4) имеют место. И

следовательно, искомым критическим множеством (критерием) является

$$S = \{k: k \leq 3\}.$$

Для этого критерия вероятность ошибки первого рода

$$P(S|H_0) = P_0(S) = \sum_{k=0}^3 \frac{\lambda_0^k}{k!} e^{-\lambda_0} = \sum_{k=0}^3 \frac{7,8^k}{k!} e^{-7,8} = 0,048.$$

Значение функции мощности критерия $\beta(p) = P(S|H_p)$ в точке $p = 0,01$:

$$\begin{aligned} \beta(0,01) &= P(S|H_{0,01}) = P_{0,01}(S) = \\ &= \sum_{k=0}^3 \frac{\lambda_1^k}{k!} e^{-\lambda_1} = \sum_{k=0}^3 \frac{1,3^k}{k!} e^{-1,3} = 0,957. \end{aligned}$$

Вероятность ошибки первого рода, равная 0,048, обозначает следующее. Используя описанный критерий, на 100 партий изделий, содержащих 6 % брака, около 5 партий будут классифицироваться как содержащие меньше 6 % брака.

Значение мощности критерия $P(S|H_{0,01}) = 0,957$ интерпретируется так: на 100 партий изделий с 1 %-м содержанием брака, около 96 партий будут классифицироваться как содержащие 1 % брака.

Для решения задач можно пользоваться табл. 22.6.1.

16.2 Задачи

АЗ: 16.1, 16.15.

СЗ: 16.9, 16.11.

16.1 (о телепатах). Некоторые люди утверждают, что они могут читать мысли на расстоянии — являются телепатами. При этом телепат не претендует на безошибочное чтение мыслей, но утверждает, что иногда ошибаясь, он все-таки чаще читает мысли верно, чем неверно.

Вам необходимо проверить способности телепата читать мысли на расстоянии. С этой целью предлагается провести такой эксперимент. Вы задумываете наудачу число (для простоты нуль или единицу) и фиксируете его, например, на бумаге. Телепат читает вашу мысль — задуманное число и также фиксирует его, и так n раз. Если телепат действительно читает ваши мысли, то доля правильно прочитанных нулей (нуль читается как нуль) и единиц (единица читается как единица), т. е. доля успехов (успех — правильно прочитанный символ) будет большой.

Вы предлагаете телепату прочитать n символов, $\xi(\omega)$ из них были прочитаны правильно. Свидетельствует ли это о способности телепата читать мысли на расстоянии?

Предложите правило (критерий), с помощью которого можно обнаружить способности телепата читать мысли на расстоянии.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез, а именно:

- предложить нулевую (основную) и альтернативные гипотезы;

- выяснить, в чем состоят ошибки первого и второго рода;

- определить случайную величину, наблюдаемую в эксперименте (что означают выдвинутые гипотезы относительно распределения этой случайной величины?);

- предложить критерий для проверки нулевой гипотезы;

- выбрать уровень значимости критерия, обосновать свой выбор;

- вычислить вероятность ошибки первого рода;

- исследовать поведение функции мощности критерия в зависимости от уровня значимости и от n , построить графики функции мощности критерия;

- дать частотную интерпретацию полученных результатов.

16.2. Необходимо проверить монету на симметричность (она может быть как симметричной — вероятность выпадения герба составляет 0,5, так и несимметричной — вероятность выпадения герба отлична от 0,5; несимметричная и симметричная монеты внешне неразличимы).

С этой целью подбрасываем монету n раз ($n = 5, 10, \dots$) и регистрируем количество выпавших гербов.

Предложить правило (критерий) для проверки монеты на симметричность.

Какое минимальное число раз необходимо подбросить монету, чтобы симметричная монета классифицировалась как несимметричная с вероятностью, не большей чем 0,1, а несимметричная, вероятность выпадения герба которой равна 0,8, обнаруживалась с вероятностью 0,95?

Решить поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

Пусть теперь уровень значимости и число n подбрасываний монеты мы можем выбирать сами. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости и от n , построить графики функции мощности критерия.

16.3 (диагностика начальной формы туберкулёза). Рассмотрим задачу диагностики туберкулёза, связанную с открытием того факта, что метод рентгеновского анализа не является абсолютно надёжным при определении наличия или отсутствия заболевания: здоров индивидум или нет, результат рентгеновского исследования на туберкулёз может быть как положительным, так и отрицательным.

Из опыта известно, что:

а) если пациент страдает туберкулёзом, то вероятность того, что отдельный рентгеновский анализ выявит заболевание (анализ будет положительным), составляет p_0 ;

б) если у пациента нет никаких признаков туберкулёза, то вероятность того, что отдельный рентгеновский анализ будет положительным (пациенту будет поставлен диагноз: “болен туберкулёзом”), равна p_1 .

Предположим, что в некоторой клинике при обследовании состояния здоровья пациента делается одним и тем же способом n рентгеновских снимков с целью обнаружения возможных признаков туберкулёза. Допустим, что толкование снимков, принадлежащих одному и тому же пациенту, производится так, чтобы была обеспечена независимость диагноза по каждому снимку от выводов, сделанных по предыдущим снимкам.

Пусть $p_0 = 0,95$, $p_1 = 0,05$ и в условиях описанного обследования изготовили два снимка. При этом один из двух анализов оказался положительным. Вынести заключение, страдает пациент туберкулёзом или нет.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1). Исследуйте, как меняется функция мощности критерия в зависимости от его уровня значимости, постройте графики функции мощности критерия.

16.4 (о телепатах). Есть люди утверждающие, что они могут читать мысли на расстоянии — являются телепатами (утверждают, что читают, но не гарантируют, что правильно будут прочитаны все мысли, т. е. признают, что иногда ошибаются: ведь мысли на расстоянии читать непросто).

Вам предлагают проверить способности телепата читать мысли на расстоянии. С этой целью проведите такой эксперимент. Подбросьте правильную монету и зафиксируйте результат, например на бумаге. Телепат читает вашу мысль — герб или решетка — и также фиксирует результат. Далее подсчитываем количество правильно прочитанных гербов (герб читается как герб) и решеток (решетка читается как решетка). Если телепат в самом деле читает ваши мысли, то доля правильно прочитанных гербов и решеток, т. е. доля успехов, будет большой (успех — правильно прочитанный символ, неудача — неправильно прочитанный символ).

Монету подбросили n раз. При этом телепат правильно прочитал ξ символов. Свидетельствует ли это о способности телепата читать мысли на расстоянии?

Проверьте телепатические способности своего товарища. Предложите критерий, с помощью которого можно было бы обнаружить способности телепата читать мысли на расстоянии.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1).

Исследуйте, как меняется функция мощности критерия в зависимости от его уровня значимости и количества n подбрасываний монеты, постройте графики функции мощности критерия.

16.5 (о статистике и экспериментаторе). Два студента (одного будем называть статистиком, другого —

экспериментатором) имеют монеты: симметричную и несимметричную, вероятность выпадения герба которой равна p ($p \neq 1/2$). Экспериментатор и статистик договариваются о следующем: экспериментатор выходит в соседнюю комнату, выбирает одну из монет (какую — статистик не знает), подбрасывает ее n раз и результат (число выпадений герба) сообщает статистику. Последний утверждает, что может определить, какую из двух монет подбрасывал экспериментатор. Однако экспериментатор ставит под сомнение эти способности статистика и как аргумент выдвигает такие возражения: поскольку результат эксперимента (число выпавших гербов) предсказать невозможно, причем как для симметричной монеты, так и для несимметричной это число может быть любым $(0, 1, \dots, n)$, то определить, какая из монет подбрасывалась, невозможно.

Кто прав — статистик или экспериментатор?

Сформулируйте поставленную задачу с позиции статистика, т. е. как задачу проверки статистических гипотез (см. задачу 16.1).

Исследуйте, как меняется функция мощности критерия в зависимости от его уровня значимости и количества n подбрасываний монеты, постройте графики функции мощности критерия.

Что можно сказать относительно возможности различить симметричную и несимметричную монеты?

Пусть $n = 25$. Какими будут выводы, если герб выпал 21 раз?

16.6 (проверка препарата на токсичность). Процесс производства некоторого медицинского препарата достаточно сложный, так что несущественные на первый взгляд отклонения от технологии могут вызвать появление высокотоксичных побочных примесей. Токсичность последних может оказаться столь высокой, что даже незначительное их количество, которое не может быть обнаружено при обычном химическом анализе, опасно для человека, принимающего этот препарат. Поэтому до реализации партии препарата его исследуют на токсичность биологическими методами. Определенная доза препарата вводится подопытным животным и регистрируется результат (количество летальных исходов). Если препарат токсичный, то все или почти все животные гибнут — ко-

личество летальных исходов $\xi(\omega)$ большое ($\xi(\omega) \geq l$). В противном случае количество летальных исходов $\xi(\omega)$ малое ($\xi(\omega) < l$), или, что то же, количество выживших животных большое.

Известно, что если препарат токсичный, то вероятность летального исхода не меньше чем 0,9, а если препарат нетоксичный, то вероятность летального исхода не превышает 0,05. Какое минимальное число n животных необходимо injectировать, чтобы токсичный препарат обнаружился (классифицировался как токсичный) с вероятностью, не меньшей чем 0,999, а нетоксичный препарат выдерживал контроль (классифицировался как нетоксичный) с вероятностью 0,98?

Найти n , сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

16.7 (выборочный контроль). Качество партии изделий считается удовлетворительным, если дефектные изделия в ней составляют не более чем 2%. Какое минимальное количество изделий в партии необходимо испытать для того, чтобы партия, содержащая 8% дефектных изделий, обнаруживалась с вероятностью, не меньшей чем 0,95, а партия, содержащая 2% таких изделий, принималась с вероятностью, не меньшей чем 0,9?

Сформулировать и решить поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1 и пример 16.1.1).

Вычислить значение функции мощности критерия в точках 0,04; 0,06; 0,1 и вероятность отклонения партии, содержащей 4; 6; 10% дефектных изделий, построить график функции мощности критерия.

З а м е ч а н и е. Поскольку выборку для контроля можно взять большого объема, а процент дефектных изделий мал, то в качестве распределения числа ξ дефектных изделий можно рассматривать пуассоновское распределение.

16.8. Чтобы выяснить, является ли болезнь инфекционной, биолог прививает ее пяти мышам и размещает их в одной клетке с n непривитыми мышами. Если болезнь инфекционная, то вероятность того, что непривитая мышь заболеет в течение 10 дней, составляет 0,9; если же болезнь неинфекционная, то вероятность того,

что непривитая мышь заболеет на протяжении указанного времени, составляет 0,05. Через 10 дней фиксируется количество непривитых мышей с признаками заболевания.

Какое минимальное количество непривитых мышей необходимо использовать в эксперименте, чтобы с вероятностью, не меньшей чем 0,999, обнаружить инфекционное заболевание и с вероятностью, не большей чем 0,01, классифицировать неинфекционное заболевание как инфекционное.

Сформулировать и решить поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

16.9 (о контроле чистоты воды). Для нужд некоторого химического производства необходимо, чтобы используемая вода была чистой — содержала мало бактерий. Вода, содержащая в среднем менее одной бактерии на единицу объема, считается пригодной для данного химического производства. Если же среднее количество бактерий на единицу объема воды равно одной или больше, то ее использование недопустимо.

Обычная методика контроля воды на чистоту состоит в следующем. Берется 10 (в общем случае n) проб воды единичного объема. Затем каждая из этих проб добавляется в колбы с питательной средой, которые содержатся при температуре, благоприятной для роста бактерий. Если проба загрязнена, т. е. содержит по меньшей мере одну бактерию, то колония бактерий будет расти и первоначально прозрачный раствор помутнеет. О чистоте воды судят по количеству ξ загрязненных проб: если это количество не превышает определенного числа l , то вода считается чистой; в противном случае — загрязненной, тогда проводится дополнительная очистка воды, и она снова проходит контроль на чистоту. (Дополнительная очистка воды, естественно, требует труда и затрат, однако, если используемая вода содержит значительное количество бактерий, то вызванные этим потери существенно больше.)

Каким должно быть l , чтобы вода со средним содержанием бактерий, равным одной на единицу объема, обнаруживалась с вероятностью 0,99? Найти указанное l , сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

Пусть теперь уровень значимости и число n не фиксировано. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости и от n , построить графики функции мощности критерия.

З а м е ч а н и е. В качестве распределения числа бактерий в единичном объеме воды со средним содержанием λ бактерий на единицу объема естественно рассматривать пуассоновское распределение с параметром λ .

16.10 (о леди, пробующей чай). Некая леди утверждает, что, попробовав чашку чая с молоком, она может определить, что было сначала налито в чашку — молоко или чай (различает рецепты приготовления чая). При этом леди не претендует на безошибочное определение разницы во вкусе, но утверждает, что, пусть иногда ошибаясь, она чаще определяет правильно, нежели неправильно.

Прежде чем признать наличие у леди способностей различать рецепты приготовления чая, ей предлагают попробовать и классифицировать n пар чашек чая (по паре чашек за завтраком в течение n дней). В каждую пару входит по чашке чая, приготовленного по разным рецептам. Количество ξ правильно классифицированных пар регистрируется. Пусть леди из n пар чашек правильно классифицировала $\xi(\omega)$ пар. Свидетельствует ли это о способностях леди различать рецепты приготовления чая?

Вы — член доброжелательного жюри — не хотите понапрасну отвергнуть способности леди различать рецепты приготовления чая, если они у леди в самом деле имеются. Предложите правило (критерий), которым должно руководствоваться жюри, вынося заключение о способностях леди различать рецепты приготовления чая.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1).

Выясните, как меняется функция мощности критерия в зависимости от уровня значимости критерия и количества n пар чашек чая, предлагаемых для классификации, постройте графики функции мощности критерия.

З а м е ч а н и е. Как член жюри, вы, разумеется, не хотите признать за леди упомянутые способности, если их нет в самом деле; поэтому в качестве нулевой гипотезы предлагаете гипотезу H_0 : леди не имеет способностей.

Но как член доброжелательного жюри, не хотите понапрасну отвергать способности леди, если она их имеет, а поэтому назначаете не слишком малый уровень значимости критерия.

16.11 (о выборочном контроле). Потребитель приобретает крупные партии товара. Чтобы избежать значительной доли дефектных изделий, осуществляется выборочный контроль. Из каждой партии для контроля выбирается n изделий. Партия принимается, если число дефектных изделий среди n проверенных не превышает l .

Потребитель предъявляет следующие требования к контролю: если доля дефектных изделий в партии меньше 8%, то партия выдерживает контроль и закупается, но партии с 8% дефектных изделий должны обнаруживаться с вероятностью, не меньшей 0,9. У поставщика свои пожелания относительно контроля: он хочет, чтобы партия, доля дефектных изделий в которой составляет 1%, браковалась с вероятностью не большей, чем 0,04.

Каким должен быть объем n выборки из партии (разумеется, минимальный) и значение l ?

Сформулировать и решить поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1 и пример 16.1.1).

Какова вероятность забраковать партию товара, содержащую 1; 2; 4; 8; 10 % дефектных изделий? Построить графики функции мощности критерия.

16.12 (об игральные костях). Имеются две игральные кости: одна — симметричная (вероятность выпадения каждой грани равна $1/6$), другая — несимметричная, центр тяжести которой смещен так, что вероятность появления числа очков, большего 3, превышает $1/2$.

С целью обнаружения несимметричной кости, берем одну (какую именно, неизвестно — кости внешне различить невозможно), подбрасываем ее n раз и регистрируем количество выпавших очков. Какое минимальное количество подбрасываний необходимо выполнить, чтобы симметричная кость классифицировалась как несимметричная с вероятностью, не большей чем 0,1, а несимметричная, вероятность выпадения количества очков большего чем 3 которой равна 0,9, обнаруживалась с вероятностью 0,95?

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез (см. задачу 16.1).

Вычислить вероятность того, что несимметричная кость, вероятность выпадения числа очков большего чем 3 которой равна 0,6; 0,7; 0,8; 0,9, будет обнаружена; не будет обнаружена.

16.13. Решить задачу 16.9 в предположении, что в эксперименте используется n проб воды.

Выяснить, как изменяется функция мощности критерия в зависимости от уровня его значимости и количества n проб воды, используемых в эксперименте? Построить графики функции мощности критерия.

16.14. Имеются две игральные кости: одна — симметричная (вероятность выпадения каждой грани равна $1/6$), и другая — несимметричная, центр тяжести которой смещен так, что вероятность появления числа очков большего 3 не меньше 0,6. Выбираем одну из костей (какую именно, неизвестно — кости внешне неразличимы), подбрасываем ее n раз и регистрируем количество выпавших очков. Можно ли по результату эксперимента обнаружить несимметричную кость?

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1).

Исследуйте, как меняется функция мощности критерия в зависимости от его уровня значимости и от числа n подбрасываний кости, постройте графики функций мощности критерия.

Что можно сказать о вероятности обнаружения несимметричной кости?

16.15 (о булочках с изюмом). Государственным стандартом установлено, что при выпечке сладких булочек на 1000 изделий должно приходиться 10 000 изюмин (в среднем 10 на одну булочку). У нас, однако, имеются сомнения, что весь изюм использован по назначению (его могли, по крайней мере частично, использовать для других целей), и мы хотим проверить, так ли это. С этой целью покупаем одну булочку и подсчитываем количество изюмин в ней: оказалось, что в булочке $\xi(\omega)$ изюмин. По количеству $\xi(\omega)$ изюмин в булочке необходимо сделать вывод: весь изюм идет на выпечку булочек или

он расходится и по другим каналам (не предусмотренным стандартом).

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез (см. задачу 16.1).

Как изменятся выводы, если изюм подсчитывается в двух булочках?

Исследовать поведение функции мощности критерия в зависимости от его уровня значимости, построить графики функции мощности критерия.

16.16. Медицинский препарат проходит контроль на токсичность по методике, описанной в задаче 16.6.

Известно, что вероятность летального исхода при использовании токсичного препарата не меньше 0,6, а если препарат нетоксичный, то вероятность летального исхода не превосходит 0,02.

Можно ли, инъецировав 10 мышей, с вероятностью 0,999 выявить токсичный препарат (токсичный препарат классифицировать как токсичный)? Если да, то какова вероятность того, что нетоксичный препарат выдержит контроль (будет классифицирован как нетоксичный)?

Препарат был введен 10 мышам, при этом зафиксирована гибель двух мышей. Какой вывод относительно токсичности препарата необходимо сделать?

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез (см. задачу 16.1).

Пусть теперь уровень значимости и число n подопытных мышей мы можем выбирать сами. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости и от n , построить графики функции мощности критерия.

16.17. Чтобы выяснить, является ли болезнь инфекционной, биолог прививает ее пяти мышам и помещает их в одной клетке с тремя непривитыми мышами. Если болезнь инфекционная, то вероятность того, что непривитая мышь заболит в течение двух недель, равна 0,95; если же болезнь неинфекционная, то вероятность того, что непривитая мышь заболит, равна 0,04. Через две недели фиксируется количество ξ непривитых мышей с признаками болезни.

Можно ли, наблюдая три непривитые мыши, с вероятностью, не меньшей 0,99, выявить инфекционное заболевание и с вероятностью, не превышающей 0,05, класси-

фицировать неинфекционное заболевание как инфекционное? Если нет, то сколько непривитых мышей необходимо для этого? Если да, то можно ли обойтись меньшим числом мышей? Каким именно?

Сформулируйте и решите поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

Какой вывод вы сделаете, если у двух из трех непривитых мышей были обнаружены признаки заболевания?

Пусть теперь уровень значимости и число n подопытных мышей мы можем выбирать сами. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости и от n , построить графики функции мощности критерия.

16.18. Имеем пять внешне неразличимых монет, вероятность выпадания герба которых составляет соответственно 0,5; 0,6; 0,7; 0,8; 0,9.

Берем одну из монет (какую именно, неизвестно — монеты по внешнему виду неразличимы) и подбрасываем 20 раз. По результатам подбрасываний сделать вывод о симметричности монеты.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1), выбрав гипотезу “монета симметрична” как основную. Постройте критерий для проверки этой гипотезы с уровнем значимости $\alpha = 0,05$. Какова вероятность обнаружить несимметричную монету, вероятность выпадения герба которой составляет 0,6; 0,7; 0,8; 0,9?

Пусть теперь уровень значимости и число n подбрасываний монет мы можем выбирать сами. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости, построить графики функций мощности критерия.

16.19 (качество инсектицида). Как специалиста с математической подготовкой Вас приглашают проанализировать результаты исследовательской работы, связанной с производством инсектицидов. Качество инсектицида определяется процентом поражения — процентом погибших насекомых популяции, обработанной инсектицидом. Лучший из имеющихся инсектицидов имеет процент поражения 92. В лаборатории разработан новый инсектицид INC. Возникает вопрос: является ли он более качественным?

Чтобы выяснить это, предлагается обработать новым инсектицидом 100 насекомых и подсчитать количество выживших. Если количество выживших насекомых меньше определенного числа l , то INC качественнее, в противном случае — нет.

Авторы нового препарата утверждают, что он имеет процент поражения не меньший 98, и хотят, чтобы это его свойство проявлялось с вероятностью, не меньшей 0,96. Инсектицид INC дороже, чем имеющийся в продаже препарат с процентом поражения 92, и покупатель, естественно, отказывается приобретать INC, если его процент поражения 92, и настаивает, чтобы при тестировании инсектицида INC, препарат, процент поражения которого составляет 92, обнаруживался с вероятностью не меньшей 0,95.

Можно ли, используя в эксперименте 100 насекомых, сделать вывод относительно качества инсектицида INC, обеспечив перечисленные выше условия? Если 100 насекомых в эксперименте не достаточно, то каким должно быть минимальное количество насекомых в обрабатываемой группе, чтобы можно было ответить на поставленный вопрос?

Дайте ответ на перечисленные вопросы, сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

Пусть теперь уровень значимости и число n насекомых мы можем выбирать сами. Исследовать, как меняется функция мощности критерия в зависимости от его уровня значимости и от n , построить графики функции мощности критерия.

16.20. Вы занимаетесь исследованиями, связанными с производством инсектицидов. Качество инсектицида определяется процентом поражения — процентом насекомых, которые гибнут от препарата в исследуемой популяции. Лучший из имеющихся инсектицидов имеет процент поражения 90. В лаборатории получен новый инсектицид NIP. Возникает вопрос: является ли этот препарат лучшим?

Чтобы выяснить, отличаются ли инсектициды по качеству, популяция из n насекомых обрабатывается NIP и подсчитывается количество выживших насекомых. Если их меньше определенного числа l , то считается, что

инсектицид NIP качественнее, в противном случае — нет.

Авторы нового препарата утверждают, что он имеет процент поражения не меньший 99, и хотят, чтобы это его свойство проявлялось с вероятностью не меньшей 0,98. Инсектицид NIP дороже имеющихся в продаже препаратов, и покупатель, естественно, отказывается приобретать NIP, если его процент поражения 90, и настаивает, чтобы инсектицид NIP, процент поражения которого равен 90, обнаруживался с вероятностью 0,95.

Какими должны быть n (по возможности меньшим) и l , чтобы инсектицид, процент поражения которого 90, обнаруживался с вероятностью 0,95, а инсектицид, процент поражения которого 99, классифицировался как такой с вероятностью не меньшей чем 0,98?

Сформулировать и решить поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

16.21. Пусть в условиях эксперимента, описанного в задаче 16.3, значения $p_0 = 0,98$, $p_1 = 0,01$.

Какое минимальное количество n снимков необходимо сделать при обследовании каждого пациента, чтобы пациента с признаками туберкулеза можно было выявить с вероятностью 0,95, а пациент, не имеющий признаков туберкулеза, выдерживал проверку на туберкулез с вероятностью 0,90?

Определить n , сформулировав поставленную задачу как задачу проверки статистических гипотез (см. задачу 16.1).

16.22. Рассмотрим эксперимент, описанный в задаче 16.5. Пусть вероятность выпадания герба несимметричной монеты равна 0,8. Что вы можете сказать относительно возможности различить симметричную и несимметричную монеты? Свой ответ аргументируйте. Дайте количественные оценки.

Сформулируйте и решите эту задачу как задачу проверки статистических гипотез (см. задачу 16.1).

16.23. Некоторые люди претендуют на чтение мыслей на расстоянии, их называют телепатами. Как правило, телепаты не настаивают на безошибочном чтении мыслей, но утверждают, что чаще читают мысли верно, чем неверно.

Вам необходимо проверить способности телепата читать мысли на расстоянии. С этой целью предлагается

провести такой эксперимент. Вы задумываете наудачу цифру от 0 до 9 (при этом можно воспользоваться таблицей случайных чисел) и фиксируете результат. Телепат читает Вашу мысль — и фиксирует четность (четная, нечетная) задуманной Вами цифры и так n раз. Если эффект телепатии в самом деле имеет место, то доля правильно прочитанных четных и нечетных цифр будет большой.

Вы предлагаете телепату определить четность n цифр. При этом телепат правильно определил четность ξ цифр из n предложенных. Свидетельствует ли это о способности телепата читать мысли на расстоянии?

Проверьте телепатические способности своего товарища. Предложите критерий, с помощью которого можно было бы обнаружить способности телепата читать мысли на расстоянии.

Сформулируйте и решите поставленную задачу в терминах проверки статистических гипотез (см. задачу 16.1).

Исследуйте, как меняется функция мощности критерия в зависимости от его уровня значимости и количества n предлагаемых для чтения цифр, постройте графики функции мощности критерия.

Глава 17

Проверка гипотез о параметрах распределения $N_{a;\sigma^2}$

17.1 Проверка гипотезы $H_0: a = a_0$

Критерии для проверки гипотез о параметрах нормального распределения строятся на основании следующей теоремы:

Теорема 17.1.1. Если $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из нормального распределения $N_{a;\sigma^2}$, то оценки

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i \text{ и } s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$$

являются независимыми случайными величинами, причем $(n-1)s^2/\sigma^2$ имеет распределение χ^2 с $n-1$ степенями свободы, а $\bar{\xi}$ распределена $N_{a;\sigma^2/n}$.

С л е д с т в и е. Если $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из нормального распределения $N_{a;\sigma^2}$, то случайная величина

$$(\bar{\xi} - a) / \frac{s}{\sqrt{n}}$$

имеет распределение Стьюдента с $n-1$ степенями свободы

Относительно распределений χ^2 и Стьюдента, см. гл. 22.

Постановка задачи проверки гипотезы H_0 : $a = a_0$. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из нормального распределения $N_{a;\sigma^2}$. Параметры a и σ^2 неизвестны. Относительно значения неизвестного параметра a выдвигается гипотеза H_0 : $a = a_0$, или, что то же, $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из нормального распределения $N_{a_0;\sigma^2}$. Альтернатива к гипотезе H_0 может быть как односторонней: если $a \neq a_0$, то $a > a_0$ (она может быть и такой: $a < a_0$), так и двусторонней: если $a \neq a_0$, то $a > a_0$ или $a < a_0$. Альтернатива в каждой задаче своя.

Необходимо проверить нулевую гипотезу, т. е. по реализации выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ сделать вывод: отклонять H_0 или нет.

Выбор статистики для построения критерия. Независимо от того, верна или нет гипотеза H_0 : $a = a_0$ (и вообще, выдвигались какие-либо гипотезы или не выдвигались), оценка $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$, построенная по выборке

$\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из $N_{a;\sigma^2}$, является несмещенной и состоятельной оценкой параметра a , т. е. $\bar{\xi}$ дает хорошее приближение для a . Поэтому, если $a = a_0$ (гипотеза верна), отклонение $\bar{\xi} - a_0$ между оценкой $\bar{\xi}$ параметра a и его гипотетическим значением a_0 малое, если гипотеза H_0 : $a = a_0$ неверна, то $\bar{\xi}$, будучи близким к a , от a_0 отличается существенно — отклонение $\bar{\xi} - a_0$ большое. (Заметим, что отклонение $\bar{\xi} - a_0$ является случайной величиной.) И следовательно, для проверки гипотезы H_0 : $a = a_0$ естественно вычислить значение отклонения $\bar{\xi} - a_0$ и в зависимости от того, большим или малым оно окажется, отклонять гипотезу H_0 или не отклонять. (Нам будет удобно нормировать $\bar{\xi} - a_0$ величиной $\frac{s}{\sqrt{n}}$ и рассматривать нормированное отклонение $t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$, где

$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$.) Чтобы так можно было посту-

пать, необходимо определить границы, отделяющие большие значения отклонения t от малых, а для этого необходимо знать, какие значения принимает отклонение

$$t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$$

(в идеале найти распределение t), когда гипотеза H_0 верна и когда она неверна (или хотя бы, когда гипотеза H_0 верна).

Когда гипотеза $H_0: a = a_0$ верна, отклонение t , которое в этом случае будем обозначать через t_{n-1} :

$$t_{n-1} = (\bar{\xi} - a) / \frac{s}{\sqrt{n}},$$

имеет распределение Стьюдента с $(n - 1)$ степенями свободы (см. теорему 17.1.1). Отсюда, в частности, следует, что

$$Mt = Mt_{n-1} = 0$$

и значение $t = t_{n-1}$ “почти всегда” (с вероятностью $1 - 2\alpha$) принадлежат промежутку $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, где $t_{\alpha;n-1}$ — верхний α -предел t_{n-1} -распределения.

Если гипотеза $H_0: a = a_0$ неверна (пусть для определенности $a > a_0$), то отклонение t можно представить в виде

$$\begin{aligned} t &= (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}} = (\bar{\xi} - a) / \frac{s}{\sqrt{n}} + (a - a_0) / \frac{s}{\sqrt{n}} = \\ &= t_{n-1} + (a - a_0) / \frac{s}{\sqrt{n}}. \end{aligned} \quad (17.1.1)$$

Из равенства (17.1.1), в частности, следует, что при $n \rightarrow \infty$

$$Mt \rightarrow \infty$$

и

$$t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}} \xrightarrow{P} \infty$$

— когда гипотеза $H_0: a = a_0$ неверна, отклонение t принимает большие значения.

В качестве границ, отделяющих большие значения уклонения t от малых, естественно выбрать числа $-t_{\alpha;n-1}$, $t_{\alpha;n-1}$, при этом значения уклонения t , принадлежащие промежутку, $(-t_{\alpha;n-1}, t_{\alpha;n-1})$, будем классифицировать как малые, в противном случае — большие.

Теперь для проверки гипотезы $H_0: a = a_0$ вычисляем значение

$$t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$$

нормированного уклонения $\bar{\xi}$ от a_0 и выясняем, большое оно или малое, сравнивая t с границами $-t_{\alpha;n-1}$, $t_{\alpha;n-1}$ отделяющими большие значения уклонения t от малых. Если t приняло большое значение, т. е. $|t| > t_{\alpha;n-1}$, гипотезу $H_0: a = a_0$ отклоняем, в противном случае — нет.

Критерий Стьюдента для проверки гипотезы $H_0: a = a_0$. Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из $N_{a;\sigma^2}$ -распределения, $t_{\alpha;(n-1)}$ — верхний α -предел t -распределения с $(n-1)$ степенями свободы.

Если гипотезу $H_0: a = a_0$ отклонять при

$$|\bar{\xi} - a_0| / \frac{s}{\sqrt{n}} > t_{\alpha;(n-1)}$$

и не отклонять в противном случае, то с вероятностью 2α гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя: $a > a_0$ или $a < a_0$).

Если альтернатива односторонняя, например $a > a_0$, то $(\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$ сравниваем с $t_{\alpha;(n-1)}$: при

$$(\bar{\xi} - a_0) / \frac{s}{\sqrt{n}} > t_{\alpha;(n-1)}$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости одностороннего критерия равен α).

Ошибки при проверке гипотезы $H_0: a = a_0$. Независимо от того, верна или нет гипотеза $H_0: a = a_0$, случайная величина $(\bar{\xi} - a) / \frac{s}{\sqrt{n}}$ имеет распределение Стьюдента с $n - 1$ степенями свободы и, как следствие,

$\bar{\xi}$ “почти всегда” (с вероятностью $1 - 2\alpha$) принимает значения из окрестности

$$\left(a - \frac{s}{\sqrt{n}} t_{\alpha;(n-1)}, a + \frac{s}{\sqrt{n}} t_{\alpha;(n-1)} \right)$$

точки a , но может, хотя и изредка (с вероятностью 2α) принимать значения вне указанной окрестности точки a . Поэтому при верной гипотезе уклонение $t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$ может, хотя и изредка (с вероятностью 2α), принимать большие значения, т. е. $|\bar{\xi} - a_0| / \frac{s}{\sqrt{n}} > t_{\alpha;(n-1)}$, при этом верную гипотезу H_0 мы отклоняем и тем самым допускаем ошибку первого рода.

Далее, при неверной гипотезе H_0 случайная величина $\bar{\xi}$, “почти всегда” принимая значения из окрестности точки a , может (хотя и изредка) принять значение, мало отклоняющееся от a_0 (т. е. $|\bar{\xi} - a_0| / \frac{s}{\sqrt{n}} \leq t_{\alpha;(n-1)}$). При этом мы гипотезу $H_0: a = a_0$ не отклоним (хотя она и неверна) и тем самым допустим ошибку второго рода.

Пример 17.1.1 (об эффекте использования специальной сеялки). *С целью обнаружения эффекта использования специальной сеялки, 10 участков земли засеяли обыкновенной сеялкой и 10 — специальной, а затем сравнили полученные урожаи. 20 участков одинаковой площади были разделены на пары, причем в каждую пару входили смежные участки. Вопрос о том, какой из двух смежных участков должен обрабатываться специальной машиной, решался подбрасыванием монеты. В таблице приведены разности урожаев с пар смежных участков, засеянных специальной сеялкой и обыкновенной.*

Номер пары	Разность урожаев	Номер пары	Разность урожаев
1	2,4	6	1,6
2	1,0	7	-0,4
3	0,7	8	1,1
4	0,0	9	0,1
5	1,1	10	0,7

Свидетельствуют ли приведенные данные о наличии эффекта использования специальной сеялки, другими словами, дает ли использование специальной сеялки прибавку к урожаю?

Решение. В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеем 10 независимых наблюдений случайной величины (см. таблицу) — реализацию $\xi_1(\omega), \xi_2(\omega), \dots, \xi_{10}(\omega)$ выборки $\xi_1, \xi_2, \dots, \xi_{10}$ из нормального распределения $N_{a;\sigma^2}$ (предположение относительно нормального распределения результатов измерений в большинстве случаев оправдывается), параметры a и σ^2 этого распределения неизвестны. Относительно параметра a распределения $N_{a;\sigma^2}$ выдвигается гипотеза $H_0: a = 0$ (к сожалению, обидная для разработчиков новой сеялки). Альтернатива односторонняя: $a > 0$ (специальная сеялка конструировалась для повышения урожайности). Отклонение гипотезы $H_0: a = 0$ в пользу альтернативы $a > 0$ будем трактовать как наличие эффекта использования специальной сеялки, неотклонение — как отсутствие эффекта.

Необходимо проверить гипотезу H_0 . В соответствии с критерием Стьюдента для проверки гипотезы $H_0: a = a_0$ при альтернативе $a > a_0$ вычисляем значение

$$t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$$

и сравниваем его с $t_{\alpha;(n-1)}$ — верхним α -пределом $t_{(n-1)}$ -распределения. Если

$$t = (\bar{\xi} - a_0) / \frac{s}{\sqrt{n}} > t_{\alpha;(n-1)},$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости критерия равен α).

В рассматриваемом примере $n = 10$,

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i = \frac{1}{10} (2,4 + 1,0 + \dots + 0,7) = 0,83;$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 = \frac{1}{9}((2,4 - 0,83)^2 +$$

$$+ (1,0 - 0,83)^2 + \dots + (0,7 - 0,83)^2) = 0,667;$$

$$t = \bar{\xi} / \sqrt{\frac{s^2}{n}} = 0,83 / \sqrt{\frac{0,667}{10}} = 3,22.$$

Таким образом,

$$t = \bar{\xi} / \frac{s}{\sqrt{n}} = 3,22 > 2,26 = t_{0,025;9}.$$

Поэтому согласно критерию Стьюдента гипотезу $H_0: a = 0$ отклоняем в пользу альтернативы $a > 0$. Другими словами, гипотеза о том, что выборка получена из нормального распределения со средним нуль, противоречит имеющимся данным.

Полученный результат можно интерпретировать так. Предположение о том, что разность урожаев с участков, засеянных специальной сеялкой и обычной, несущественно отклоняется от нуля, противоречит имеющимся данным. Для сеялки, использование которой не дает эффекта, уклонения разности урожаев от нуля, приведенные в таблице, невозможны (точнее, возможны, но крайне редко). Следовательно, эксперимент дает основания утверждать: эффект использования специальной сеялки существует. (К радости разработчиков специальной сеялки.)

17.2 Проверка гипотезы $H_0: a_\xi = a_\eta$

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ и $\eta(\omega) = (\eta_1(\omega), \eta_2(\omega), \dots, \eta_m(\omega))$ — реализации независимых выборок $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ соответственно из распределений $N_{a_\xi; \sigma^2}$ и $N_{a_\eta; \sigma^2}$. Средние a_ξ и a_η и дисперсия σ^2 (одна и та же для обоих распределений) неизвестны.

Относительно параметров a_ξ и a_η выдвигается гипотеза $H_0: a_\xi = a_\eta$ или, что то же, $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$

— выборки из одного и того же нормального распределения. Альтернатива к гипотезе H_0 может быть как односторонней: если $a_\xi \neq a_\eta$, то $a_\xi > a_\eta$ (она может быть и такой: $a_\xi < a_\eta$), так и двусторонней: если $a_\xi \neq a_\eta$, то $a_\xi > a_\eta$ или $a_\xi < a_\eta$. Альтернатива в каждой задаче своя.

Необходимо по выборкам $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ соответственно из распределений $N_{a_\xi; \sigma^2}$ и $N_{a_\eta; \sigma^2}$ сделать вывод: отклонять гипотезу H_0 или нет.

Выбор статистики для построения критерия. Независимо от того, верна или нет гипотеза $H_0: a_\xi = a_\eta$, оценки $\bar{\xi}$ и $\bar{\eta}$ являются несмещенными и состоятельными оценками соответственно параметров a_ξ и a_η , а следовательно $\bar{\xi} - \bar{\eta}$ является несмещенной и состоятельной оценкой $a_\xi - a_\eta$. Поэтому, если $a_\xi = a_\eta$ (гипотеза H_0 верна), разность $\bar{\xi} - \bar{\eta}$ мало уклоняется от нуля, если гипотеза H_0 неверна, отклонение разности от нуля большое. И для проверки гипотезы $H_0: a_\xi = a_\eta$ естественно вычислить разность $\bar{\xi} - \bar{\eta}$ и в зависимости от того, какое значение она приняла: большое или малое — отклонять или не отклонять гипотезу H_0 . (Нам будет удобно нормировать $\bar{\xi} - \bar{\eta}$ величиной $s\sqrt{\frac{n+m}{nm}}$ и рассматривать нормированное отклонение $t = (\bar{\xi} - \bar{\eta}) / \left(s\sqrt{\frac{n+m}{nm}}\right)$, где s^2 определяется равенством (17.2.2).) Чтобы так можно было поступать, необходимо определить границы, отделяющие большие значения t от малых, а для этого необходимо знать, какие значения принимает отклонение

$$t = (\bar{\xi} - \bar{\eta}) / \left(s\sqrt{\frac{n+m}{nm}}\right)$$

(а в идеале найти распределение t), когда гипотеза H_0 верна и когда она неверна (или по меньшей мере, когда гипотеза H_0 верна).

Если гипотеза $H_0: a_\xi = a_\eta$ верна, нормированная раз-

ность t , которую в этом случае обозначим через t_{n+m-2} :

$$t_{n+m-2} = (\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right),$$

имеет распределение Стьюдента с $(m+n-2)$ степенями свободы. Отсюда, в частности, следует, что

$$Mt = Mt_{n+m-2} = 0$$

и значения уклонения $t = t_{n+m-2}$ “почти всегда” (с вероятностью $1 - 2\alpha$) принадлежат промежутку

$$(-t_{\alpha;(n+m-2)}, t_{\alpha;(n+m-2)}),$$

где $t_{\alpha;(n+m-2)}$ — верхний α -предел t -распределения с $n+m-2$ степенями свободы.

Если гипотеза $H_0: a_\xi - a_\eta = 0$ неверна (пусть для определенности $a_\xi > a_\eta$), разность можно представить в виде

$$\begin{aligned} t &= (\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right) = \\ &= ((\bar{\xi} - a_\xi) - (\bar{\eta} - a_\eta)) / \left(s \sqrt{\frac{n+m}{nm}} \right) + (a_\xi - a_\eta) / \left(s \sqrt{\frac{n+m}{nm}} \right) = \\ &= t_{n+m-2} + (a_\xi - a_\eta) / \left(s \sqrt{\frac{n+m}{nm}} \right). \end{aligned} \quad (17.2.1)$$

Из равенств (17.2.1) следует, что при $n, m \rightarrow \infty$

$$Mt \rightarrow \infty$$

и

$$t = t_{n+m-2} + (\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right) \xrightarrow{P} \infty$$

— когда гипотеза H_0 неверна, уклонение t принимает большие значения.

В качестве границ, отделяющих большие значения отклонения t от малых, естественно выбрать числа $t_{\alpha;n+m-2}$, $-t_{\alpha;n+m-2}$, при этом значения отклонения t , принадлежащие промежутку, $(-t_{\alpha;n+m-2}, t_{\alpha;n+m-2})$, будем классифицировать как малые, в противном случае — большие.

Теперь для проверки гипотезы $H_0: a_\xi - a_\eta = 0$ вычислим значение нормированной разности

$$t = (\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right)$$

и выясняем, большое оно или малое, сравнивая t с границами $-t_{\alpha;(n+m-2)}$ и $t_{\alpha;(n+m-2)}$, отделяющими большие значения отклонения t от малых. Если t приняло большое значение, т. е. $|t| > t_{\alpha;n+m-2}$, гипотезу отклоняем, в противном случае — нет.

Напомним, что

$$s^2 = \frac{1}{n+m-2} ((n-1)s_\xi^2 + (m-1)s_\eta^2), \quad (17.2.2)$$

где

$$s_\xi^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2, \quad s_\eta^2 = \frac{1}{m-1} \sum_{i=1}^m (\eta_i - \bar{\eta})^2.$$

Критерий Стьюдента для проверки гипотезы $H_0: a_\xi = a_\eta$ (о равенстве средних нормальных распределений). Пусть $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ — независимые выборки соответственно из распределений $N_{a_\xi; \sigma^2}$ и $N_{a_\eta; \sigma^2}$, $t_{\alpha;(n+m-2)}$ — верхний α -предел t -распределения с $(n+m-2)$ степенями свободы.

Если гипотезу $H_0: a_\xi = a_\eta$ отклонять при

$$|t| = |\bar{\xi} - \bar{\eta}| / \left(s \sqrt{\frac{n+m}{nm}} \right) \geq t_{\alpha;(n+m-2)}$$

и не отклонять в противном случае, то с вероятностью 2α гипотезу H_0 будем отклонять, когда она верна (альтернатива двусторонняя: $a_\xi < a_\eta$ или $a_\xi > a_\eta$).

Если альтернатива односторонняя, например $a_\xi > a_\eta$, то t сравниваем с $t_{\alpha; (n+m-2)}$: при

$$t = (\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right) \geq t_{\alpha; (n+m-2)}$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости одностороннего критерия равен α).

Пример 17.2.1. Ниже приведены данные измерения неровностей поверхности одной и той же чистоты обработки с помощью двух двойных микроскопов.

Можно ли считать, что между показаниями приборов нет систематического расхождения?

Микроскоп I: 0,8; 1,9; 3,0; 3,5; 3,8; 2,5; 1,7; 0,9; 1,0; 2,3; 3,3; 3,4.

Микроскоп II: 1,4; 2,1; 3,1; 3,6; 2,7; 1,7; 1,1; 0,2; 1,6; 2,8; 4,0; 4,7.

Решение. Для определенности обозначим через $\xi_1, \xi_2, \dots, \xi_n$ данные измерений, полученные с помощью первого микроскопа, через $\eta_1, \eta_2, \dots, \eta_m$ — второго.

В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеются реализации двух независимых выборок $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ соответственно из распределений $N_{a_\xi; \sigma^2}$ и $N_{a_\eta; \sigma^2}$ (предположение о нормальном распределении результатов измерений в большинстве случаев оправдывает себя). Относительно параметров a_ξ и a_η выдвигается гипотеза $H_0: a_\xi = a_\eta$. Это гипотеза об отсутствии систематического расхождения между показаниями приборов. Априори, если $a_\xi \neq a_\eta$, то может быть как $a_\xi < a_\eta$, так и $a_\xi > a_\eta$; поэтому альтернатива двусторонняя. Отклонение гипотезы H_0 в пользу этой альтернативы интерпретируется как наличие систематического расхождения показаний приборов, неотклонение — как отсутствие расхождений показаний.

Согласно критерию Стьюдента для проверки гипотезы $H_0: a_\xi = a_\eta$ при двусторонней альтернативе $a_\xi < a_\eta$ или $a_\xi > a_\eta$ необходимо сравнить значение

$$|t| = |\bar{\xi} - \bar{\eta}| / \left(s \sqrt{\frac{n+m}{nm}} \right)$$

с $t_{\alpha; (n+m-2)}$ — верхним α -пределом $t_{(n+m-2)}$ -распределения. Если

$$|t| = |\bar{\xi} - \bar{\eta}| / \left(s \sqrt{\frac{n+m}{nm}} \right) \geq t_{\alpha; (n+m-2)},$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости критерия 2α).

В рассматриваемом примере $n = 12$, $m = 12$,

$$\bar{\xi} = \frac{1}{12}(0,8 + 1,9 + \dots + 3,4) = 2,34;$$

$$\bar{\eta} = \frac{1}{12}(1,4 + 2,1 + \dots + 4,7) = 2,42;$$

$$s^2 = \frac{1}{n+m-2}((n-1)s_{\xi}^2 + (m-1)s_{\eta}^2) = 1,44,$$

где

$$s_{\xi}^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2, \quad s_{\eta}^2 = \frac{1}{m-1} \sum_{i=1}^m (\eta_i - \bar{\eta})^2;$$

$$\begin{aligned} |t| &= |\bar{\xi} - \bar{\eta}| / \left(s \sqrt{\frac{n+m}{nm}} \right) = \\ &= |2,42 - 2,34| / \sqrt{\frac{1,44(12+12)}{(12 \cdot 12)}} = 0,16. \end{aligned}$$

Таким образом,

$$|t| = 0,16 < 2,074 = t_{0,025; 22}.$$

Поэтому согласно критерию Стьюдента гипотезу H_0 о равенстве $a_{\xi} = a_{\eta}$ на 5%-м уровне значимости не отклоняем.

Этот результат можно трактовать так. Предположение об отсутствии систематического расхождения между

показаниями микроскопов не противоречит экспериментальным данным. Такие показания вполне могли быть получены при работе с одним и тем же прибором. Другими словами, эксперимент не дает оснований говорить о существовании систематического расхождения между показаниями микроскопов.

17.3 Проверка гипотезы $H_0: \sigma^2/\sigma_0^2 = 1$

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из $N_{a; \sigma^2}$. Параметры a и σ^2 неизвестны. Относительно значения параметра σ^2 выдвигается гипотеза

$$H_0: \sigma^2/\sigma_0^2 = 1.$$

Альтернатива к гипотезе $H_0: \sigma^2/\sigma_0^2 = 1$ может быть как односторонней: если $\sigma^2/\sigma_0^2 \neq 1$, то $\sigma^2/\sigma_0^2 > 1$ (она может быть и такой: $\sigma^2/\sigma_0^2 < 1$), так и двусторонней: если $\sigma^2/\sigma_0^2 \neq 1$, то $\sigma^2/\sigma_0^2 > 1$ или $\sigma^2/\sigma_0^2 < 1$. Альтернатива в каждой задаче своя.

Необходимо по реализации $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из $N_{a; \sigma^2}$ прийти к выводу: отклонять гипотезу H_0 или нет.

Выбор статистики для построения критерия. Независимо от того, верна гипотеза $H_0: \sigma^2/\sigma_0^2 = 1$ или нет,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$$

является несмещенной и состоятельной оценкой параметра σ^2 . Поэтому если гипотеза $H_0: \sigma^2/\sigma_0^2 = 1$ верна, отношение $s^2/\sigma_0^2 = s^2/\sigma^2$ мало уклоняется от 1 по сравнению с отношением s^2/σ_0^2 , когда гипотеза $H_0: \sigma^2/\sigma_0^2 = 1$ неверна. А именно, если $\sigma^2/\sigma_0^2 = 1$, то $M(s^2/\sigma_0^2) =$

$M(s^2/\sigma^2) = 1$ и $s^2/\sigma_0^2 \xrightarrow{P} \sigma^2/\sigma_0^2 = 1$. А если $\sigma^2/\sigma_0^2 \neq 1$, то $M(s^2/\sigma_0^2) = \sigma^2/\sigma_0^2 \neq 1$ и $s^2/\sigma_0^2 \xrightarrow{P} \sigma^2/\sigma_0^2 \neq 1$. Поэтому критерий для проверки гипотезы $H_0: \sigma^2/\sigma_0^2 = 1$ естественно строить, сравнивая отношение s^2/σ_0^2 с 1. Если отношение s^2/σ_0^2 существенно уклоняется от 1, то гипотезу $H_0: \sigma^2/\sigma_0^2 = 1$ естественно отклонять, в противном случае — нет. Границы, отделяющие значения s^2/σ_0^2 , существенно уклоняющиеся от 1, от значений s^2/σ_0^2 , мало (допустимо) уклоняющихся от 1, устанавливаются на основании того, что для выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из $N_{a;\sigma^2}$ случайная величина

$$(n-1)s^2/\sigma^2$$

имеет распределение χ^2 с $(n-1)$ степенями свободы (относительно распределения χ^2 см. гл. 22).

Критерий для проверки гипотезы $H_0: \frac{\sigma^2}{\sigma_0^2} = 1$.

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из $N_{a;\sigma^2}$, $\chi_{\beta;(n-1)}^2$ — верхний β -предел $\chi_{(n-1)}^2$ -распределения.

Если гипотезу $H_0: \sigma^2/\sigma_0^2 = 1$ отклонять при

$$\frac{s^2}{\sigma_0^2} \notin \left(\frac{1}{n-1} \chi_{(1-\alpha);(n-1)}^2, \frac{1}{n-1} \chi_{\alpha;(n-1)}^2 \right)$$

и не отклонять в противном случае, то с вероятностью 2α гипотезу H_0 будем отклонять, когда она верна (альтернатива двусторонняя: $\sigma^2/\sigma_0^2 < 1$ или $\sigma^2/\sigma_0^2 > 1$).

Если альтернатива односторонняя, например $\sigma^2/\sigma_0^2 > 1$, то при

$$\frac{s^2}{\sigma_0^2} > \frac{1}{n-1} \chi_{\alpha;(n-1)}^2$$

гипотезу H_0 отклоняем, в противном случае — нет (уровень значимости одностороннего критерия равен α).

17.4 Проверка гипотезы $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ и $\eta(\omega) = (\eta_1(\omega), \eta_2(\omega), \dots, \eta_m(\omega))$ — реализации независимых выборок $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ соответственно из распределений $N_{a_\xi; \sigma_\xi^2}$ и $N_{a_\eta; \sigma_\eta^2}$. Параметры a_ξ , a_η , σ_ξ^2 и σ_η^2 нормальных распределений неизвестны. Относительно значений σ_ξ^2 и σ_η^2 выдвигается гипотеза

$$H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$$

или, что то же, выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ получены из нормальных распределений с одной и той же дисперсией. Альтернатива к гипотезе $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ может быть как односторонней: если $\sigma_\xi^2/\sigma_\eta^2 \neq 1$, то $\sigma_\xi^2/\sigma_\eta^2 > 1$ (она может быть и такой: $\sigma_\xi^2/\sigma_\eta^2 < 1$), так и двусторонней: если $\sigma_\xi^2/\sigma_\eta^2 \neq 1$, то $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$. Выбор альтернативы определяется поставленной задачей.

Необходимо по реализациям выборок $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ из распределений $N_{a_\xi; \sigma_\xi^2}$ и $N_{a_\eta; \sigma_\eta^2}$ соответственно прийти к выводу: отклонять гипотезу H_0 или нет.

Выбор статистики для построения критерия. Независимо от того, верна гипотеза $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ или нет,

$$s_\xi^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 \quad \text{и} \quad s_\eta^2 = \frac{1}{m-1} \sum_{i=1}^m (\eta_i - \bar{\eta})^2$$

являются состоятельными оценками соответственно параметров σ_ξ^2 и σ_η^2 , а поэтому s_ξ^2/s_η^2 сходится по вероятности к $\sigma_\xi^2/\sigma_\eta^2$, когда $n \rightarrow \infty$, $m \rightarrow \infty$. Следовательно, если $\sigma_\xi^2/\sigma_\eta^2 = 1$, то отношение s_ξ^2/s_η^2 при больших n и

m мало отличается от $\sigma_\xi^2/\sigma_\eta^2 = 1$. Если же $\sigma_\xi^2/\sigma_\eta^2 \neq 1$, то отношение s_ξ^2/s_η^2 , будучи близким к $\sigma_\xi^2/\sigma_\eta^2$, от 1 отличается существенно. Поэтому критерий для проверки гипотезы $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ естественно строить, сравнивая отношение s_ξ^2/s_η^2 с 1: если это отношение существенно отличается от 1, то гипотезу H_0 естественно отклонять, в противном случае — нет. Границы, отделяющие значения s_ξ^2/s_η^2 , существенно отличающиеся от 1, от значений s_ξ^2/s_η^2 , которые мало (допустимо) отличаются от 1, устанавливаются на основании того, что при $\sigma_\xi^2/\sigma_\eta^2 = 1$ (т. е. когда гипотеза H_0 верна) случайная величина s_ξ^2/s_η^2 имеет F -распределение с $(n-1, m-1)$ степенями свободы (относительно F -распределения см. гл. 22).

Критерий для проверки гипотезы $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$.

Пусть $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ — независимые выборки соответственно из распределений $N_{a_\xi; \sigma_\xi^2}$ и $N_{a_\eta; \sigma_\eta^2}$; $F_{\beta; s; k}$ — верхний β -предел F -распределения с s, k степенями свободы.

Если гипотезу $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ отклонять при

$$\frac{s_\xi^2}{s_\eta^2} \notin \left(\frac{1}{F_{\alpha; (m-1); (n-1)}}, F_{\alpha; (n-1); (m-1)} \right)$$

и не отклонять в противном случае, то с вероятностью 2α гипотезу H_0 будем отклонять, когда она верна (этим критерием пользуемся, если альтернатива двусторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$).

Если альтернатива односторонняя, например $\sigma_\xi^2/\sigma_\eta^2 > 1$, то при

$$\frac{s_\xi^2}{s_\eta^2} > F_{\alpha; (n-1); (m-1)}$$

гипотезу H_0 будем отклонять, в противном случае — нет (уровень значимости этого одностороннего критерия равен α).

Пример 17.4.1. *Определяется предел прочности на разрыв материала на двух стендах А и В. Получены такие выборки значений предела прочности на разрыв.*

Стенд А: 1,32; 1,35; 1,32; 1,35; 1,30; 1,30; 1,37; 1,31; 1,39; 1,39.

Стенд В: 1,35; 1,31; 1,31; 1,41; 1,39; 1,37; 1,32; 1,34.

Выяснить, можно ли считать, что точность измерений предела прочности на разрыв на стендах А и В одинакова.

Решение. В терминах проверки статистических гипотез эту задачу можно сформулировать так. Имеем две реализации независимых выборок из нормальных распределений $N_{a\xi;\sigma_\xi^2}$ и $N_{a\eta;\sigma_\eta^2}$ (пусть для определенности $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка, полученная на стенде А, $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ — на стенде В). Относительно неизвестных параметров σ_ξ^2 и σ_η^2 выдвигается гипотеза $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ (это гипотеза об одинаковой точности измерений на стендах А и В). Альтернатива двусторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$, поскольку нет никакой априорной информации относительно точности работы станков А и В. Отклонение гипотезы H_0 в пользу этой альтернативы будем интерпретировать как наличие отличий в точности измерений на станках, неотклонение — как отсутствие отличий.

Согласно критерию для проверки гипотезы $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$ при двусторонней альтернативе гипотезу H_0 отклоняем, если

$$\frac{s_\xi^2}{s_\eta^2} \notin \left(\frac{1}{F_{\alpha;(m-1);(n-1)}}, F_{\alpha;(n-1);(m-1)} \right),$$

и не отклоняем в противном случае.

В рассматриваемом примере $n = 10$, $m = 8$, $\bar{\xi} = 1,34$; $\bar{\eta} = 1,35$,

$$s_{\xi}^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2 = 12,2 \cdot 10^{-4};$$

$$s_{\eta}^2 = \frac{1}{m-1} \sum_{j=1}^m (\eta_j - \bar{\eta})^2 = 14,0 \cdot 10^{-4};$$

$$\frac{s_{\xi}^2}{s_{\eta}^2} = \frac{12,2 \cdot 10^{-4}}{14,0 \cdot 10^{-4}} = 0,87;$$

$$F_{\alpha; (n-1); (m-1)} = F_{0,01; 9; 7} = 6,72;$$

$$\frac{1}{F_{\alpha; (m-1); (n-1)}} = \frac{1}{5,61} = 0,18.$$

Значение s_{ξ}^2/s_{η}^2 принадлежит промежутку

$$\left(\frac{1}{F_{\alpha; (m-1); (n-1)}}, F_{\alpha; (n-1); (m-1)} \right) = (0,18; 6,72),$$

поэтому гипотезу $H_0: \sigma_{\xi}^2/\sigma_{\eta}^2 = 1$ на 2 %-м уровне значимости не отклоняем.

Этот результат можно интерпретировать так. Предположение (гипотеза), что стенды A и B имеют одинаковую точность измерений предела прочности на разрыв, не противоречит экспериментальным данным. (Такие данные вполне могли быть получены при работе на одном и том же стенде.) Другими словами, эксперимент не дает оснований утверждать, что точность измерений предела прочности на разрыв на стендах A и B различна.

17.5 Задачи

АЗ: 17.1, 17.6, 17.8, 17.15.

СЗ: 17.4, 17.9, 17.14, 17.33.

Далее, если это не оговорено специально, мы будем предполагать, что результаты измерений являются выборками из нормальных распределений. В большинстве случаев это предположение оправдывает себя в опыте.

17.1. В течение летних каникул группа из 10 учеников находилась в спортивном лагере. В начале сезона и после его завершения у них измерялся объем легких (в миллилитрах). По результатам измерений необходимо сделать вывод, существенно ли изменился этот показатель под влиянием интенсивных физических упражнений.

Уче- ник	Объем легких		Уче- ник	Объем легких	
	В на- чале сезона	После окончания сезона		В на- чале сезона	После окончания сезона
1	3400	3800	6	3100	3200
2	3600	3700	7	3200	3200
3	3000	3300	8	3400	3300
4	3500	3600	9	3200	3500
5	2900	3100	10	3400	3600

17.2. С целью сравнения отражательной способности двух видов краски (A и B), провели следующий эксперимент: из 10 выбранных наудачу пробных образцов пять покрасили одной краской, остальные — другой и измерили их отражательную способность с помощью оптического прибора. Получили следующие результаты.

Краска A : 195; 150; 205; 120; 160.

Краска B : 200; 115; 220; 185; 170.

Свидетельствуют ли эти данные о различии отражательных способностей красок?

17.3. Один из методов количественного анализа степени износа шины состоит в измерении глубины проникновения щупа¹ в определенном месте шины. Имеется подозрение, что значительная часть дисперсии измерений

¹Здесь щуп — тонкая продолговатая металлическая пластинка прямоугольной формы.

связана с действиями контролеров. Чтобы выделить из общей дисперсии измерений указанную часть, двум контролерам (X и Y) предложили провести по 12 измерений в одной и той же точке шины. Получили следующие результаты.

Контролер X : 121; 121; 126; 130; 127; 131; 127; 124; 125; 119; 126; 123.

Контролер Y : 120; 129; 128; 136; 117; 138; 124; 119; 136; 136; 134; 132.

Существенно ли отличаются дисперсии измерений, проведенных контролерами X и Y ?

17.4. В процессе производства электрические счетчики с вращающимся диском были отрегулированы так, что их работа стала синхронной с работой стандартного счетчика (работа счетчика характеризуется постоянной, равной 1,000). Проверка 10 счетчиков, состоящая в определении их постоянной с помощью точных ваттметров и секундомеров, показала следующие результаты:

Номер счетчика	Значение постоянной	Номер счетчика	Значение постоянной
1	0,983	6	0,988
2	1,002	7	0,994
3	0,998	8	0,991
4	0,996	9	1,005
5	1,003	10	0,986

Можно ли отклонения от стандарта рассматривать как случайные, или, напротив, результаты указывают на то, что постоянная отрегулированных счетчиков систематически уклоняется от постоянной стандартного счетчика?

Дать ответ на этот вопрос, проверив гипотезу о том, что 10 измерений образуют выборку, полученную из нормального распределения со средним 1,000.

17.5. С целью сравнения двух сортов стали (I и II) на способность к глубокому отпуску, изготовили по семь образцов из каждого сорта и испытали их по методу Эриксона, согласно которому в образец вдавливается конус с шаровым наконечником. Глубину проникновения шарика измеряют в миллиметрах. Результаты испытаний приведены ниже.

Сталь сорта I: 10,85; 10,24; 10,48; 10,35; 11,07; 9,54; 11,18.

Сталь сорта II: 11,05; 10,07; 10,03; 10,57; 10,27; 9,97; 9,92.

Можно ли считать, что оба сорта стали имеют одинаковую способность к глубокому отпуску?

Решить поставленную задачу, сформулировав ее как задачу проверки статистических гипотез.

17.6. Масса упаковки с натуральным красителем должна составлять 50 кг, а дисперсия массы упаковки не должна превышать 0,01 кг². Взвешивание 10 упаковок дало следующие результаты: 50,10; 50,07; 49,93; 49,81; 49,70; 49,76; 50,08; 50,18; 49,79; 51,02 (масса в килограммах).

Свидетельствуют ли приведенные данные о том, что указанные требования

1° к массе упаковки,

2° к дисперсии массы упаковки
выполняются?

17.7. Приведенные ниже данные характеризуют твердость 14 образцов сплава в условных единицах: 12,1; 13,7; 11,0; 11,6; 11,9; 13,4; 12,2; 12,5; 11,9; 11,5; 12,9; 13,0; 10,5; 11,7.

Предположим, что твердость сплава распределена нормально. Можно ли считать, что среднее μ нормального распределения равно 12,0?

17.8. Исследовалась возможность снижения затрат на разведку залежей золота на одном из участков рассыпного месторождения. С этой целью вместо части запланированных для закладки шурфов пробурили скважины ударно-канатного бурения (затраты на бурение меньше). Получили такие результаты исследований на содержание золота в образцах пород из шурфов и скважин (содержание золота в миллиграммах на кубический метр породы).

Скважины: 322; 250; 225; 315; 399; 348; 192; 375; 381; 538; 198; 317; 293.

Шурфы: 478; 299; 541; 457; 251; 221; 548; 431; 397; 462; 457; 251; 221; 548.

Можно ли считать, что результаты исследований образцов на содержание золота из скважин и из шурфов отличаются существенно?

17.9. На станке-автомате производится один вид продукции. Критическим размером изделий является внешний диаметр. После настройки станка отобрали 20 изделий. При этом оказалось, что выборочная дисперсия размера внешнего диаметра составляет $0,84 \text{ мм}^2$. Через определенное время для контроля точности работы станка отобрали 15 изделий. Вычисленная по ним выборочная дисперсия оказалась равной $1,07 \text{ мм}^2$.

Свидетельствуют ли приведенные данные об изменении точности работы станка?

17.10. Проверялась скорость полимеризации на нескольких образцах полимера. Расчетная скорость полимеризации равна 24 %. В восьми экспериментах получены такие результаты (в %): 23,6; 22,8; 25,7; 24,8; 26,4; 24,3; 23,9; 25,0.

Имеются ли основания считать, что расчетная скорость полимеризации согласуется с этими данными?

17.11. С целью проверки влияния специального способа приготовления бетона на его прочность провели эксперимент. Из данной партии сырья взяли шесть порций. Затем порции случайным образом разделили на две группы по три в каждой. После этого сырье одной группы подвергли специальной обработке и из каждой порции изготовили пробный куб. После 28-дневной выдержки шести пробных кубов определили их сопротивление на сжатие. Регистрировали значения нагрузок, при которых происходит разрушение образцов (предел прочности на сжатие). Получили такие результаты.

Бетон (без обработки)	290	311	284
Бетон (с обработкой)	309	318	318

Свидетельствуют ли эти данные о наличии эффекта специальной обработки бетона?

17.12. С целью уменьшения выхода нежелательного побочного продукта использовали катализаторы A и B . Для каждого из них были получены выборки выхода нежелательного продукта (в %).

Катализатор A : 41; 52; 29; 43; 38; 45; 52; 42; 36.

Катализатор B : 40; 27; 59; 42; 25; 58; 42; 26; 37.

Можно ли считать, что разброс выхода нежелательного продукта (разброс характеризуется дисперсией) при использовании катализаторов A и B одинаковый?

17.13. На компьютере моделируется нормально распределенная случайная величина. Реализация выборки этой случайной величины: 2,41; 2,26; 2,10; 2,53; 1,96; 1,31; 2,32; 2,66.

Можно ли утверждать, что выборка получена из нормального распределения со средним 2,5?

17.14. В классе 20 детей, из них наудачу отобрали 10, которым ежедневно начали давать апельсиновый сок. Остальные 10 учеников ежедневно получали молоко. Через некоторое время зафиксировали следующее увеличение веса детей в фунтах (1 фунт = 453,6 г).

Сок	4,0	2,5	3,5	4,0	1,5	1,0	3,5	3,0	2,5	3,5
Молоко	1,5	3,5	2,5	3,0	2, 5	2,0	2,0	2,5	1,5	3,0

Среднее увеличение веса на одного ученика в группе, где давали апельсиновый сок, составило 2,9 фунта, а в группе, где давали молоко, — 2,4. Существенно ли отличается увеличение среднего веса детей в группах?

17.15. Были измерены неровности поверхности с регулярным профилем (одного и того же уровня чистоты) с помощью двойных микроскопов (I и II). При этом получены такие результаты.

Микроскоп I: 0,8; 2,0; 3,1; 3,5; 2,1; 1,7; 0,9; 1,1; 3,3.

Микроскоп II: 0,7; 2,1; 3,1; 0,9; 3,6; 2,7; 0,8; 4,7; 0,3; 4,1; 2,8; 0,6.

Выяснить, можно ли считать, что точность измерений каждым микроскопом одинакова? (Точность измерений характеризуется дисперсией.)

17.16. Номинальное сопротивление производимых резисторов 2000 Ом. С целью контроля отобрали партию из 12 резисторов. После измерения сопротивления каждого получили такие значения: 2130; 2090; 2030; 2080; 1920; 2020; 2015; 2000; 2045; 1940; 1980; 1970.

Можно ли отклонения от номинала (2000 Ом) рассматривать как случайные (допустимые), или, напротив, результаты измерений свидетельствуют о том, что сопротивление резисторов существенно отличается от номинала?

17.17. На сталелитейном заводе для контроля содержания марганца в одной из марок стали сделали 10 отливок из конвертора I и столько же из конвертора II. Ниже приведено содержание марганца (в %) в каждом отливке.

Конвертор I: 1,20; 1,17; 1,15; 1,21; 1,14; 1,17; 1,18; 1,21; 1,25; 1,14.

Конвертор II: 1,39; 1,29; 1,28; 1,34; 1,32; 1,30; 1,28; 1,35; 1,35; 1,30.

Можно ли считать содержание марганца в стали, выплавленной в этих конверторах, одинаковым?

17.18. В целях уменьшения дисперсии отражательной способности краски были внесены изменения в технологию ее изготовления. Чтобы убедиться, что такие изменения в самом деле дают эффект, изготовили 10 пробных образцов. С помощью специального оптического прибора определили отражательную способность краски, изготовленной по традиционной (A) и новой (B) технологии (в относительных единицах). Получили следующие результаты.

Технология A: 40; 45; 195; 65; 145.

Технология B: 110; 55; 120; 50; 80.

Свидетельствуют ли приведенные данные об изменении дисперсии отражательной способности краски?

17.19. С целью контроля напряжения в осветительной сети каждый час в течение суток регистрировали напряжение (в вольтах): 220; 222; 220; 220; 220; 222; 220; 218; 218; 220; 220; 222; 222; 220; 218; 220; 222; 216; 218; 214; 210; 218; 223; 215.

Можно ли отклонения от стандарта (стандарт равен 220 В) рассматривать как случайные? Или наоборот, приведенные данные указывают на систематическое отклонение напряжения от стандарта?

17.20. Исследования, проводимые на протяжении ряда лет после того, как в практике лыжного спорта стала применяться техника коньковых ходов, дают основания предполагать, что она в условиях специально подготовленной трассы, обеспечивает на дистанции 15 км у мужчин выигрыш по сравнению с традиционной техникой более чем на 1 мин. Ниже приведены результаты соревнований (в минутах) двух групп лыжников (дистанция 15 км:

одни проходили дистанцию традиционным ходом, другие — коньковым).

Традиционный ход: 37,02; 36,74; 37,82; 38,12; 36,91; 37,98; 38,21; 37,51; 37,56; 38,03.

Коньковый ход: 35,81; 35,61; 35,02; 35,53; 35,84; 35,12; 36,12; 36,49; 35,62; 36,28.

Можно ли на основании результатов этих соревнований сделать вывод, что эффект техники коньковых ходов на дистанции 15 км составляет более чем 1 мин?

17.21. На станках A и B одного класса точности изготавливают одинаковые изделия. Критическим размером изделия является его внешний диаметр. Ниже приведены значения несмещенных оценок внешнего диаметра и его дисперсии, вычисленные по выборкам объемов 15 и 10, полученным соответственно для станков A и B :

$$\bar{\xi}_A = 45,3; s_A^2 = 1,07; \bar{\xi}_B = 46,1; s_B^2 = 0,84.$$

Свидетельствуют ли приведенные данные о том, что внешний диаметр изделий, изготовленных на станках A и B , одинаковый?

17.22. Исследовалась коррозионная стойкость нержавеющей стали 18Cr10Ni2Mo (сталь, содержащая 18 % хрома, 10 % никеля, 2 % молибдена). Эксперимент проводили на 12 образцах. Прежде чем исследовать их на коррозионную стойкость, определяли содержание хрома, никеля и молибдена в стали, из которой были изготовлены образцы. В частности, содержание хрома (в процентах массы) в исследуемых образцах было таким: 17,4; 17,9; 17,6; 18,1; 17,6; 18,9; 16,9; 17,5; 17,8; 17,4; 24,6; 21,0.

Можно ли на основании приведенных данных утверждать, что содержание хрома в стали составляет 18 %?

17.23. В лаборатории, изучающей влияние окружающей среды на человека, с целью определения комнатной температуры, при которой наиболее комфортно чувствуют себя мужчины и женщины, были обследованы 10 мужчин и 10 женщин. Получили такие значения температуры наибольшей комфортности (по Фаренгейту).

Мужчины: 74; 71; 77; 76; 72; 75; 73; 74; 75; 72.

Женщины: 75; 77; 78; 79; 77; 73; 78; 78; 80; 76.

Свидетельствуют ли эти данные о том, что температура, при которой человек чувствует себя комфортно, для мужчин и женщин одинакова?

Примечание. $t^{\circ}\text{F} = (9/5)t^{\circ}\text{C} + 32$, где $t^{\circ}\text{F}$ — температура по Фаренгейту; $t^{\circ}\text{C}$ — температура по Цельсию.

17.24. Измеряя сопротивление провода двух типов (A и B), получили такие данные.

Провод A : 0,126; 0,131; 0,126; 0,127; 0,124; 0,130; 0,128; 0,124.

Провод B : 0,121; 0,121; 0,124; 0,122; 0,120; 0,124; 0,125; 0,120.

Утверждается, что между разбросом сопротивления провода типа A и типа B нет различий.

Не противоречит ли это утверждение имеющимся данным?

17.25. Расчетная прочность на изгиб изготовленной резины равна 11 условным единицам. Чтобы убедиться действительно ли это так, испытали на изгиб 16 образцов этой резины.

Можно ли по приведенным ниже данным прийти к выводу, что изготовлена резина с ожидаемой прочностью на изгиб?

Прочность на изгиб: 10,3; 11,1; 11,8; 12,0; 10,8; 13,6; 12,0; 12,5; 11,6; 12,2; 12,3; 12,5; 12,6; 13,7; 13,3; 10,5.

17.26. В задаче 17.21 утверждается, что станки, на которых изготавливаются изделия, имеют одинаковый класс точности. Так ли это в действительности?

17.27. Предлагается новая методика определения содержания марганца в одной из марок сталей. Ожидается, что эта методика дает меньший разброс результатов по сравнению с традиционной. Ниже приведено процентное содержание марганца в 10 отливках, определенное по традиционной методике, и в 8 отливках — по новой.

Традиционная методика: 1,22; 1,15; 1,17; 1,22; 1,26; 1,27; 1,19; 1,22; 1,20; 1,15.

Новая методика: 1,21; 1,24; 1,18; 1,17; 1,15; 1,18; 1,17; 1,17.

Имеются ли основания считать, что новая методика определения содержания марганца дает меньший разброс по сравнению с традиционной?

17.28. Для стальной проволоки, идущей на изготовление канатов, значение растягивающего усилия, при котором происходит разрыв проволоки (предел прочности на разрыв) должно быть равным 6720 кг/см^2 .

Проведенная серия наблюдений дает основания считать, что значение предела прочности на разрыв проволоки имеет нормальное распределение.

Для контроля качества из каждой партии проволоки, которая приходит на завод, 20 образцов испытывают на разрыв. Фиксируют значение усилия, при котором образец разрушается.

Результаты одного из таких испытаний приведены ниже.

Предел прочности на разрыв: 6300; 6870; 6720; 6980; 6780; 6780; 6780; 6720; 6630; 6660; 6900; 7130; 6690; 6750; 6560; 6700; 6930; 6720; 6950; 6960.

Можно ли на основании приведенных данных утверждать, что для этой партии проволоки предел прочности на разрыв не менее 6720 кг/см^2 ?

17.29. Два предприятия (*A* и *B*) производят кирпичную футеровку для кислородных конверторов. Потребитель хочет выяснить, отличается ли футеровка предприятий по своим характеристикам, чтобы в дальнейшем закупать продукцию с лучшими показателями. Для этого он регистрирует количество плавок, проведенных в конверторе до очередной замены футеровки. При этом получены такие результаты.

Предприятие *A*: 237; 224; 218; 227; 234; 215; 219; 225; 230.

Предприятие *B*: 216; 202; 205; 200; 207; 198; 222; 214; 226; 204.

Можно ли на основании этих данных заключить, что различие между числом плавок до замены футеровки, изготовленной на предприятиях *A* и *B*, существенно?

17.30. Дисперсия предела прочности на разрыв волокна составляет $35,63 \text{ фунта}^2$ (предел прочности — усилие, при котором происходит разрыв волокна). Ожидается, что внесенные в технологический процесс изменения уменьшат дисперсию. Зарегистрированы такие значения предела прочности на разрыв (в фунтах, $1 \text{ фунт} = 453,6 \text{ г}$): 151; 156; 147; 153; 155; 148; 160; 149; 156; 161; 154; 162; 163; 149; 150.

Уменьшилась ли дисперсия предела прочности на разрыв волокна с изменением технологического процесса?

17.31. В процессе разработки модификации автобусных шин, которые имели бы больший пробег, проводилось исследование ряда их параметров. В частности, регистрировалась температура, до которой нагревались передние шины во время движения автобуса. По приведенным в таблице значениям температуры (в условных единицах) левой и правой шин, которые использовались на 12-ти автобусах, выяснить, являются ли существенными отличия этих параметров.

Авто- бус	Шина		Авто- бус	Шина	
	Левая	Правая		Левая	Правая
1	36	27	7	41	60
2	42	45	8	40	34
3	55	84	9	100	117
4	59	84	10	58	78
5	79	70	11	38	56
6	108	99	12	73	88

17.32. На автоматическом станке обрабатывают втулки. После настройки станка получили выборку из 10 изделий. Оказалось, что выборочное среднее диаметра втулки $\bar{\xi} = 2,059$ мм, а значение несмещенной оценки дисперсии $s_{\xi}^2 = 4,4$ мкм². Через некоторое время для контроля настройки станка на заданный диаметр втулки снова получили выборку из 10 изделий, для которой выборочное среднее $\bar{\eta} = 2,063$ мм, а значение несмещенной оценки дисперсии $s_{\eta}^2 = 8,6$ мкм². Предположим, что в течение указанного интервала времени изменения в работе станка могут сказаться только на его настройке на заданный диаметр втулки, но точность работы станка не изменяется.

Свидетельствуют ли приведенные данные об изменении настройки станка за промежуток времени, отделяющий моменты получения выборок?

З а м е ч а н и е. Точность работы станка характеризуется дисперсией

17.33. При комнатной температуре исследовалась сила сцепления двух типов клейких веществ: изобутила 2-цианакрилата и MBR-4197. Ниже приведены значения силы сцепления для каждого образца, которая измерялась в фунтах на квадратный дюйм (1 фунт = 453,6 г, 1 дюйм = 0,0254 м).

Изобутил 2-цианакрилат: 365; 169; 210; 297; 228; 146; 163; 213; 300; 218.

MBR-4197: 518; 403; 473; 329; 457; 419; 437; 396; 424; 363.

Можно ли считать, что дисперсии силы сцепления изобутила 2-цианакрилата и MBR-4197 одинаковы?

17.34. В таблице приведены данные измерения длины образцов изделий (в миллиметрах) до и после отжига их в высокочастотной печи.

Образец	Длина, мм		Образец	Длина, мм	
	До отжига	После отжига		До отжига	После отжига
1	11,94	12,00	6	11,96	11,98
2	11,99	11,99	7	11,95	12,03
3	11,98	11,95	8	11,96	12,02
4	12,03	12,07	9	11,92	12,01
5	12,03	12,03	10	12,00	11,99

Привела ли термообработка к изменению размеров изделий?

17.35. С целью сравнения пределов прочности на растяжение резиновых смесей типов *A* и *B* изготовили партию из восьми образцов прямоугольной формы (по четыре из каждой смеси) и подвергли их продольному растяжению. (Один из образцов смеси *A*, который квалифицировали как дефектный, был исключен из партии до начала испытаний.) Фиксировался предел прочности на растяжение образцов двух резиновых смесей (значение растягивающего усилия, при котором происходит разрушение образца). Получили следующие результаты.

Предел прочности (смесь *A*): 3210; 3000; 3315.

Предел прочности (смесь *B*): 3225; 3320; 3365; 3145.

Можно ли считать, что состав резины не влияет на ее прочность?

17.36. В задаче 17.32 описывается процесс контроля за работой станка-автомата. При этом предполагается, что точность работы станка не изменяется, хотя значения несмещенных оценок дисперсий после настройки станка и через определенное время составляют соответственно $s_{\xi}^2 = 4,4 \text{ мкм}^2$ и $s_{\eta}^2 = 8,6 \text{ мкм}^2$.

Является ли обоснованным предположение о неизменности точности работы станка?

17.37 (длительность обезболивающего действия препарата). Сравняется действие обезболивающих препаратов A и B . (В некоторых случаях одним из “лекарств” может быть инертное плацебо (пустышка), которое используется для “контроля” при исследовании действия другого препарата.)

В группе больных, выразивших желание взять участие в эксперименте, насчитывается восемь человек. Вполне возможно (и естественно), что у этих больных возраст, общее состояние и т. д. далеко не одинаковы. Поэтому каждому больному дают оба препарата и каждый раз фиксируют продолжительность обезболивающего действия. Чтобы обеспечить чистоту эксперимента, приняты все разумные меры предосторожности: второй препарат дают не раньше чем закончится действие первого, четверо больных получают сначала препарат A , а другие четверо — сначала B , при этом ни один из них не знает, какой именно препарат он употребляет, и т. д. Результаты эксперимента приведены в таблице.

Боль- ной	Длительность действия препарата, час		Разность длительности действия препаратов B и A , час
	A	B	
1	3,2	3,8	0,6
2	1,6	1,0	−0,6
3	5,7	8,4	2,7
4	2,8	3,6	0,8
5	5,5	5,0	−0,5
6	1,2	3,5	2,3
7	6,1	7,3	1,2
8	2,9	4,8	1,9

Сделать вывод об эффективности действия препаратов в предположении, что:

1° отсутствует априорная информация об эффективности препаратов;

2° эффективность действия препарата B не уступает действию препарата A .

Ответ дать в терминах проверки статистических гипотез.

17.38. Два оператора (A и B) провели 14 независимых экспериментов, исследуя температуру возгорания эмали одного и того же состава. Каждый из операторов

испытал по семь образцов. Получены такие результаты экспериментов.

Оператор *A*: 1450; 1425; 1420; 1410; 1370; 1360; 1270.

Оператор *B*: 1430; 1420; 1380; 1320; 1320; 1290; 1280.

Является ли существенным отличие между результатами, полученными операторами?

17.39. В процессе производства синтетического волокна для уменьшения усадки продукции, которая будет изготавливаться из него, волокно, движущееся непрерывным потоком, проходит термическую обработку.

Далее приведены данные о величине усадки волокна (в %) после обработки при температурах 120°C и 140°C .

Усадка при 120°C : 3,45; 3,62; 3,60; 3,49; 4,64; 3,56; 3,52; 3,53; 3,57; 3,44; 3,56; 3,43.

Усадка при 140°C : 3,72; 4,01; 3,54; 3,67; 4,03; 3,40; 3,96; 3,60; 3,76; 3,91.

Необходимо выяснить, является ли усадка при температуре 140°C большей, чем при температуре 120°C .

17.40 (красное смещение). Астрономы М. Гумасон и Н. Майал определяли поправку на красное смещение (в километрах на секунду) галактик типа SO. Результаты, полученные в серии из 10 наблюдений (номера галактик приведены по каталогу NGC), оказались такими.

Номер галактики	М. Гумасон	Н. Майал	Номер галактики	М. Гумасон	Н. Майал
1332	1507	1471	5866	924	1033
3607	858	778	6661	4607	4430
3998	1205	1155	6703	2592	2670
4111	832	915	7625	1930	2050
5308	2206	2194	7679	5378	5278

Свидетельствуют ли приведенные данные о расхождении в результатах, полученных астрономами?

Сформулировать и решить задачу в терминах проверки статистических гипотез.

17.41. Значения несмещенной оценки дисперсии выборки *A* составляет $55,4 \text{ кг}^2$, а выборки *B* — $87,3 \text{ кг}^2$.

Можно ли утверждать, что выборки получены из распределений с одним и тем же значением дисперсии, если объемы выборок равны соответственно 15 и 12?

Какие при этом необходимо сделать предположения?

17.42. Исследовалась потеря массы десяти одинаковых резиновых стержней при испытаниях на износ.

Для проведения исследований от каждого стержня отрезали по два образца. Один из них прошел вулканизацию при 80°C , другой — при 150°C .

Стержень	Потеря массы		Стержень	Потеря массы	
	80°C	150°C		80°C	150°C
1	3,02	2,91	6	3,11	3,20
2	2,22	2,30	7	2,70	2,50
3	4,60	4,15	8	2,58	2,29
4	4,53	2,63	9	3,27	3,11
5	2,31	2,40	10	4,19	3,80

Можно ли на основании приведенных данных утверждать, что различие между средними потерями массы образцов, которые прошли различную вулканизацию, является существенным?

Глава 18

Критерий χ^2

18.1 Критерий χ^2 (гипотетическое распределение не зависит от параметров)

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ — реализация выборки из неизвестного распределения F , относительно которого выдвигается гипотеза $H_0: F = G$, где G принадлежит заданному классу распределений (в частности, G может быть полностью определенным распределением). Гипотезу H_0 можно сформулировать и так: $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ является выборкой из распределения G с заданными свойствами.

Необходимо по реализации выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ сделать вывод: отклонить гипотезу H_0 или нет.

Об уклонении эмпирического распределения от гипотетического. Независимо от того, верна гипотеза H_0 или нет, эмпирическое распределение $\hat{F}_n$, построенное по выборке $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из F , мало уклоняется от распределения F , а именно, для каждого фиксированного x значение эмпирической функции распределения $\hat{F}_n(x)$ является несмещенной и состоятельной оценкой $F(x)$. Поэтому, если ввести уклонение $D(\hat{F}_n, G)$ эмпирического $\hat{F}_n$ распределения от гипотетического G , причем так, чтобы оно принимало малые значения, когда гипотеза H_0 верна, и большие, когда гипотеза H_0 неверна,

то гипотезу H_0 естественно отклонять или не отклонять в зависимости от того, какое значение приняло отклонение $D(\hat{F}_n, G)$ — большое или малое.

Уклонение Пирсона эмпирического распределения от гипотетического. Два вероятностных распределения F и G , заданные на $X \subset \mathbb{R}^1$, совпадают, если для каждого борелевского множества B из X справедливо равенство $F(B) = G(B)$. Если F и G различны, то найдется борелевское множество $B' \subset X$ такое, что $F(B') \neq G(B')$. Поэтому в качестве уклонения между двумя распределениями: F и G естественно рассмотреть величину

$$\sum_{j=1}^r c_j (F(X_j) - G(X_j))^2,$$

где $\{X_j\}$ — разбиение X на непересекающиеся борелевские множества:

$$\bigcup_{j=1}^r X_j = X, \quad X_i \cap X_j = \emptyset, \quad i \neq j, \quad 2 \leq r < \infty;$$

c_j — неотрицательные константы. Уклонение Пирсона между эмпирическим распределением $\hat{F}_n$, построенным по выборке $\xi_1, \xi_2, \dots, \xi_n$, и гипотетическим G из этих соображений и строится. А именно, в качестве уклонения между $\hat{F}_n$ и G рассмотрим

$$D(\hat{F}_n, G) = \sum_{j=1}^r c_j (\hat{F}_n(X_j) - G(X_j))^2,$$

где $G(X_j)$ — вероятность попадания выборочного значения в множество X_j , вычисленная по гипотетическому распределению G (далее будем обозначать $G(X_j) = p_j$); $\hat{F}_n(X_j)$ — вероятность попадания выборочного значения в X_j , вычисленная по эмпирическому распределению $\hat{F}_n$ (она численно равна частоте ν_j/n попадания выборочного значения в множество X_j , вычисленной по выборке $\xi_1, \xi_2, \dots, \xi_n$, ν_j — число выборочных значений из

$\xi_1, \xi_2, \dots, \xi_n$, попавших в X_j); c_j — коэффициенты, которые выбираются более или менее произвольно. К. Пирсон в качестве c_j предложил рассматривать n/p_j , $j = 1, 2, \dots, r$. При этом уклонение $D(\hat{F}_n, G)$ принимает вид

$$D(\hat{F}_n, G) = \sum_{j=1}^r \frac{n}{p_j} \left(\hat{F}_n(X_j) - G(X_j) \right)^2 = \sum_{j=1}^r \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j \right)^2.$$

(Заметим, что $\hat{F}_n$ и G заданы на множестве X выборочных значений ξ_k ; X — это $\mathbb{R}^1$ или часть $\mathbb{R}^1$; если, например, ξ_k принимает значения 0 и 1, то $X = \{0, 1\}$, если ξ_k — нормально распределенная случайная величина, то $X = \mathbb{R}^1$.)

Далее, вне зависимости от того, верна гипотеза H_0 : $F = G$ или нет, частота ν_j/n попадания выборочных значений в X_j , вычисленная по выборке $\xi_1, \xi_2, \dots, \xi_n$ из F , является несмещенной и состоятельной оценкой $F(X_j)$, т. е.

$$M(\nu_j/n) = F(X_j)$$

и

$$\nu_i/n \xrightarrow{P} F(X_i)$$

при $n \rightarrow \infty$ (для всех $j = 1, 2, \dots, r$). Поэтому, если гипотеза H_0 : $F = G$ верна, то $F(X_j) = G(X_j) = p_j$ и, следовательно,

$$(\nu_j/n - p_j)^2 = (\nu_j/n - F(X_j))^2$$

сходится по вероятности к нулю при $n \rightarrow \infty$ (для всех $j = 1, 2, \dots, r$), при этом распределение уклонения К. Пирсона

$$D(\hat{F}_n, F) = \sum_{j=1}^r \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j \right)^2$$

сходится к χ^2 -распределению с $(r-1)$ степенями свободы. Если же гипотеза H_0 : $F = G$ неверна, т. е. $F \neq G$, то

$$D(\hat{F}_n, G) = \sum_{j=1}^r \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j \right)^2$$

сходится к $+\infty$ при $n \rightarrow \infty$, поскольку

$$\left(\frac{\nu_j}{n} - p_j\right)^2 \xrightarrow{P} (F(X_j) - p_j)^2 = (F(X_j) - G(X_j))^2,$$

а среди

$$(F(X_j) - G(X_j))^2, \quad j = 1, 2, \dots, r,$$

при достаточно “мелком” разбиении выборочного пространства X найдутся числа строго большие нуля.

Так что уклонение К. Пирсона

$$D(\hat{F}_n, G) = \sum_{j=1}^r \frac{n}{p_j} \left(\frac{\nu_j}{n} - p_j\right)^2 = \sum_{j=1}^r \frac{(\nu_j - np_j)^2}{np_j}$$

при верной гипотезе H_0 принимает малые значения (при этом распределение $D(\hat{F}_n, F)$ близко к χ_{r-1}^2), а когда гипотеза H_0 неверна — большие ($D(\hat{F}_n, G) \rightarrow +\infty$ при $n \rightarrow \infty$).

Таким образом, для проверки гипотезы H_0 : $\xi_1, \xi_2, \dots, \xi_n$ — выборка из распределения G , вычисляем уклонение $D(\hat{F}_n, G)$. Если при этом $D(\hat{F}_n, G)$ приняло малое значение, то гипотезу H_0 не отклоняем, в противном случае — отклоняем.

Границы, отделяющие большие значения уклонения $D(\hat{F}_n, G)$ от малых, устанавливаются на основании того, что для выборки $\xi_1, \xi_2, \dots, \xi_n$ из распределения F при больших n распределение $D(\hat{F}_n, F)$ мало отличается от распределения χ^2 с $(r-1)$ степенями свободы.

Критерий χ^2 (гипотетическое распределение не зависит от неизвестных параметров). Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из распределения F и $\chi_{\alpha; (r-1)}^2$ — верхний α -предел χ^2 -распределения с $(r-1)$ степенями свободы.

Если гипотезу H_0 : $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ является выборкой из распределения G отклонять при

$$D(\hat{F}_n, G) = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} > \chi_{\alpha; (r-1)}^2,$$

и не отклонять в противном случае, то с вероятностью α гипотеза H_0 будет отклоняться, когда она верна.

Рекомендации по применению критерия χ^2 . Критерием χ^2 можно пользоваться, когда объем n выборки и вероятности p_i , $i = 1, 2, \dots, r$, выборочным значениям попасть в множество X_i , вычисленные по гипотетическому распределению, являются такими, что

$$np_i \geq 10, \quad i = 1, 2, \dots, r. \quad (18.1.1)$$

По сути, это требования относительно объема выборки: n должно быть настолько большим, чтобы выполнялись условия (18.1.1).

Разбиение X — пространства выборочных значений — на непересекающиеся подмножества X_i :

$$X = \bigcup_{i=1}^r X_i, \quad X_i \cap X_j = \emptyset, \quad i \neq j, \quad i, j = 1, 2, \dots, r,$$

осуществляют, вообще говоря, произвольно, но количество r подмножеств стремятся выбрать как можно большим, при этом должны выполняться условия (18.1.1). Если при указанном делении для некоторых множеств X_i имеют место неравенства $np_i < 10$, то такие X_i следует объединить с другими X_j так, чтобы для полученных при этом новых множеств X'_i , выполнялись неравенства $np'_i \geq 10$ (где $p'_i = P\{\xi_k \in X'_i\} = G(X'_i)$, $i = 1, 2, \dots, r$) или чтобы они содержали по меньшей мере по 10 выборочных значений. Если выборочных значений несколько мало, что это сделать невозможно, пользоваться критерием χ^2 не следует.

Пример 18.1.1 (показания механических часов). В таблице приведены показания 500 наудачу выбранных механических часов, выставленных в витринах часовых магазинов.

Согласуется ли с приведенными данными гипотеза о равномерном распределении показаний часов на интервале $[0; 12)$?

i	n_i	i	n_i
0	41	6	41
1	34	7	33
2	54	8	37
3	39	9	41
4	49	10	47
5	45	11	39
		Всего	500

Здесь i — номер интервала от i -го часа до $(i + 1)$ -го, $i = 0, 1, \dots, 11$; n_i — количество часов, показания которых принадлежат i -му интервалу.

Решение. Показания 500 часов можно рассматривать как реализацию выборки объемом 500 из некоторого непрерывного на интервале $[0; 12)$ распределения F . Относительно распределения F случайной величины ξ (показаний часов) выдвигается гипотеза H_0 : F — равномерное на интервале $[0; 12)$ распределение, т. е. плотность $f(x)$ распределения F имеет вид

$$f(x) = \begin{cases} 1/12, & \text{если } x \in [0; 12); \\ 0, & \text{если } x \notin [0; 12). \end{cases}$$

Для проверки гипотезы H_0 воспользуемся критерием χ^2 . Разобьем множество возможных значений $X = [0; 12)$ случайной величины ξ на непересекающиеся подмножества $X_i = [i, i + 1)$, $i = 0, 1, \dots, 11$. Вероятность попадания ξ в X_i , вычисленная по гипотетическому распределению:

$$p_i = P\{\xi \in [i, i + 1)\} = \int_i^{i+1} \frac{1}{12} dx = \frac{1}{12}.$$

И поскольку

$$np_i = 500 \cdot \frac{1}{12} = 41,7 > 10; \quad i = 0, 1, \dots, 11,$$

то можно воспользоваться критерием χ^2 в приведенной выше форме.

Вычислим значение уклонения между эмпирическим распределением $\hat{F}_n$ и гипотетическим G . Имеем

$$\begin{aligned} D(\hat{F}_n, G) &= \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i} = \\ &= \frac{1}{41,7} \left((41 - 41,7)^2 + \dots + (39 - 41,7)^2 \right) = 9,99. \end{aligned}$$

Таким образом,

$$D(\hat{F}_n, G) = 9,99 < 19,7 = \chi_{0,05;11}^2 = \chi_{\alpha;(r-1)}^2.$$

Поэтому, согласно критерию χ^2 , гипотеза о равномерном на интервале $[0;12)$ распределении случайной величины ξ — показаниях механических часов в витринах магазинов — не отклоняется. Другими словами, предположение о том, что показания часов распределены равномерно на интервале $[0;12)$, не противоречит наблюдениям.

18.2 Критерий χ^2 (параметры неизвестны)

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из неизвестного распределения F , относительно которого выдвигается гипотеза

$$H_0: F(\cdot) = G(\cdot; \theta_1, \theta_2, \dots, \theta_k), \quad (\theta_1, \theta_2, \dots, \theta_k) = \theta \in \Theta.$$

Распределение $G(\cdot; \theta_1, \theta_2, \dots, \theta_k)$ зависит от неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$, причем единственным источником информации о значениях этих параметров является выборка $\xi = (\xi_1, \xi_2, \dots, \xi_n)$. Другими словами, гипотеза H_0 состоит в том, что $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ является выборкой из распределения, принадлежащего классу распределений

$$G(\cdot; \theta_1, \theta_2, \dots, \theta_k), \quad \theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta.$$

Необходимо по реализации выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ сделать вывод: отклонять гипотезу H_0 или нет.

Естественно поступать следующим образом: оценить параметры $\theta_1, \theta_2, \dots, \theta_k$ по выборке $\xi_1, \xi_2, \dots, \xi_n$ и в качестве гипотетического распределения рассмотреть распределение $G(\cdot; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$. Уклонение $D(\hat{F}_n, G)$ строим так же, как и ранее:

$$D(\hat{F}_n, G) = \sum_{i=1}^r \frac{(\nu_i - np_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k))^2}{np_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)},$$

где $p_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$ — вероятность попадания выборочного значения в множество X_i , $i = 1, 2, \dots, r$, вычисленная по гипотетическому распределению. Р. Фишер установил, что когда гипотеза H_0 верна и оценки $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ получены по методу максимального правдоподобия, то распределение уклонения $D(\hat{F}_n, G)$ между $\hat{F}_n$ и G при $n \rightarrow \infty$ сходится к распределению χ^2 с $(r - 1 - k)$ степенями свободы, где k — количество параметров, оцененных по выборке $\xi_1, \xi_2, \dots, \xi_n$.

Таким образом, если параметры оцениваются по выборке по методу максимального правдоподобия, можно пользоваться критерием χ^2 в следующей формулировке.

Если гипотезу H_0 отклонять при

$$D(\hat{F}_n, G) = \sum_{i=1}^r \frac{(\nu_i - np_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k))^2}{np_i(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)} > \chi_{\alpha; (r-1-k)}^2$$

и не отклонять в противном случае, то с вероятностью α гипотеза H_0 будет отклоняться, когда она верна.

Пример 18.2.1 Среди 2020 семей с двумя детьми 527 имеют двух мальчиков, 476 — двух девочек, у остальных (1017 семей) дети разного пола.

Можно ли считать, что количество мальчиков в семье, имеющей двух детей, является биномиально распределенной случайной величиной?

Решение. Обозначим через ξ количество мальчиков в семье с двумя детьми, ξ принимает значения 0, 1, 2. Относительно распределения P_ξ случайной величины ξ выдвигается гипотеза

$$H_0: P_\xi(k) = G(k; p) = C_2^k p^k (1-p)^{2-k}, \quad k = 0, 1, 2,$$

(гипотеза о биномиальном распределении ξ), которую необходимо проверить.

Гипотетическое распределение зависит от неизвестного параметра p . Оценкой максимального правдоподобия параметра p биномиального распределения

$$P(k; p) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m,$$

полученной по выборке $\xi_1, \xi_2, \dots, \xi_n$ является

$$\hat{p} = \frac{1}{mn} \sum_{k=1}^n \xi_k.$$

В рассматриваемом примере значение оценки максимального правдоподобия параметра p равно

$$\hat{p} = \frac{0 \cdot 476 + 1 \cdot 1017 + 2 \cdot 527}{2 \cdot 2020} = 0,513.$$

Поэтому гипотетическим распределением является

$$G(k; p) = C_2^k (0,513)^k (1 - 0,513)^{2-k}, \quad k = 0, 1, 2.$$

Его можно записать и в таком виде:

$$G = \begin{pmatrix} 0 & 1 & 2 \\ 0,237 & 0,500 & 0,263 \end{pmatrix}.$$

В качестве разбиения выборочного пространства $X = \{0, 1, 2\}$ рассмотрим $X_0 = \{0\}$, $X_1 = \{1\}$, $X_2 = \{2\}$. Для множеств X_i оценки $\hat{p}_i$ гипотетических вероятностей выборочному значению попасть в X_i , $i = 0, 1, 2$,

соответственно равны 0,237; 0,500; 0,263; объем выборки $n = 2020$, потому

$$n\hat{p}_i \geq 10, \quad i = 0, 1, 2,$$

и, следовательно, пользоваться критерием χ^2 можно.

Вычислим значение уклонения Пирсона между эмпирическим $\hat{F}_n$ и гипотетическим G распределениями:

$$\begin{aligned} D(\hat{F}_n, G) &= \frac{(\nu_0 - n\hat{p}_0)^2}{n\hat{p}_0} + \frac{(\nu_1 - n\hat{p}_1)^2}{n\hat{p}_1} + \frac{(\nu_2 - n\hat{p}_2)^2}{n\hat{p}_2} = \\ &= \frac{(476 - 478,74)^2}{478,74} + \frac{(1017 - 1010)^2}{1010} + \frac{(527 - 531,26)^2}{531,26} = \\ &= 0,098. \end{aligned}$$

Отсюда имеем

$$D(\hat{F}_n, G) = 0,098 < 3,84 = \chi_{0,05;1}^2 = \chi_{\alpha;(r-1-k)}^2$$

($r-1-k = 3-1-1$; $k = 1$ — количество параметров, оцененных по выборке; $r = 3$ — количество подмножеств, на которые разбивается выборочное пространство). Поэтому согласно критерию χ^2 гипотеза о биномиальном распределении случайной величины ξ не отклоняется. Предположение, что количество мальчиков в семьях с двумя детьми имеет биномиальное распределение, не противоречит имеющимся данным.

18.3 Критерий χ^2 как критерий независимости

Постановка задачи. Имеется n независимых наблюдений (a_i, b_j) , $i = 1, 2, \dots, s$, $j = 1, 2, \dots, k$ векторной случайной величины $\zeta = (\xi, \eta)$. При этом значение (a_i, b_j) случайная величина $\zeta = (\xi, \eta)$ приняла ν_{ij} раз, $i = 1, 2, \dots$,

$\dots, s, j = 1, 2, \dots, k$ (ясно, что $\sum_{i=1}^s \sum_{j=1}^k \nu_{ij} = n$). Результаты наблюдений удобно представлять в виде так называемой таблицы сопряженности признаков (см. табл. 18.3.1). Но ни распределение векторной случайной величины $\zeta = (\xi, \eta)$

$$P\{\zeta = (a_i, b_j)\} = p_{ij}, \quad i = 1, 2, \dots, s; \quad j = 1, 2, \dots, k$$

(см. также табл. (18.3.2)), ни распределения ее компонент ξ и η :

$$P\{\xi = a_i\} = p_{i\cdot}, \quad i = 1, 2, \dots, s;$$

$$P\{\eta = b_j\} = p_{\cdot j}, \quad j = 1, 2, \dots, k,$$

неизвестны.

Таблица 18.3.1. Таблица сопряженности признаков

Значения ξ	Значения η				Сумма
	b_1	b_2	$\dots$	b_k	
a_1	ν_{11}	ν_{12}	$\dots$	ν_{1k}	$\nu_{1\cdot}$
a_2	ν_{21}	ν_{22}	$\dots$	ν_{2k}	$\nu_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	ν_{s1}	ν_{s2}	$\dots$	ν_{sk}	$\nu_{s\cdot}$
Сумма	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	$\dots$	$\nu_{\cdot k}$	n

В таблице 18.3.1

$$\nu_{i\cdot} = \sum_{j=1}^k \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=1}^s \nu_{ij}, \quad i = 1, 2, \dots, s, \quad j = 1, 2, \dots, k.$$

Относительно совместного распределения случайных величин ξ и η (распределения вектора $\zeta = (\xi, \eta)$), см. табл. 18.3.2, выдвигается гипотеза

$$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}, \quad i = 1, 2, \dots, s, \quad j = 1, 2, \dots, k.$$

Другими словами, H_0 — это гипотеза о независимости случайных величин ξ и η (еще говорят, гипотеза о независимости признаков).

Т а б л и ц а 18.3.2. Совместное распределение случайных величин ξ и η

Значения ξ	Значения η				Сумма
	b_1	b_2	$\dots$	b_k	
a_1	p_{11}	p_{12}	$\dots$	p_{1k}	$p_{1\cdot}$
a_2	p_{21}	p_{22}	$\dots$	p_{2k}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_s	p_{s1}	p_{s2}	$\dots$	p_{sk}	$p_{s\cdot}$
Сумма	$p_{\cdot 1}$	$p_{\cdot 2}$	$\dots$	$p_{\cdot k}$	1

В табл. 18.3.2

$$p_{i\cdot} = \sum_{j=1}^k p_{ij}, \quad i = 1, 2, \dots, s; \quad p_{\cdot j} = \sum_{i=1}^s p_{ij}, \quad j = 1, 2, \dots, k.$$

$$\sum_{i=1}^s p_{i\cdot} = \sum_{j=1}^k p_{\cdot j} = \sum_{i=1}^s \sum_{j=1}^k p_{ij} = 1.$$

По результатам наблюдений $\zeta = (\xi, \eta)$ (см. табл. 18.3.1) необходимо проверить гипотезу H_0 о независимости компонент ξ и η векторной случайной величины $\zeta = (\xi, \eta)$.

Для проверки гипотезы H_0 воспользуемся критерием χ^2 , когда гипотетическое распределение зависит от неизвестных параметров. Неизвестными параметрами являются $p_{i\cdot}$ и $p_{\cdot j}$, $i = 1, 2, \dots, s$; $j = 1, 2, \dots, k$, причем

$$\sum_{i=1}^s p_{i\cdot} = 1; \quad \sum_{j=1}^k p_{\cdot j} = 1.$$

В качестве разбиения выборочного пространства

$$X = \{(a_i, b_j), i = 1, 2, \dots, s; j = 1, 2, \dots, k\}$$

рассмотрим подмножества

$$X_{ij} = \{(a_i, b_j)\}, i = 1, 2, \dots, s; j = 1, 2, \dots, k.$$

Уклонение эмпирического распределения

$$\hat{F}_n\{(a_i, b_j)\} = \frac{\nu_{ij}}{n}, i = 1, 2, \dots, s; j = 1, 2, \dots, k,$$

от гипотетического

$$G\{(a_i, b_j)\} = p_{i \cdot} p_{\cdot j}, i = 1, 2, \dots, s; j = 1, 2, \dots, k,$$

равно

$$D(\hat{F}_n, G) = \sum_{i=1}^s \sum_{j=1}^k \frac{(\nu_{ij} - n\hat{p}_{i \cdot} \hat{p}_{\cdot j})^2}{n\hat{p}_{i \cdot} \hat{p}_{\cdot j}},$$

где $\hat{p}_{i \cdot} = \nu_{i \cdot}/n$ ($i = 1, 2, \dots, s$) и $\hat{p}_{\cdot j} = \nu_{\cdot j}/n$ ($j = 1, 2, \dots, k$) — оценки максимального правдоподобия соответственно параметров $p_{i \cdot}$ ($i = 1, 2, \dots, s$) и $p_{\cdot j}$ ($j = 1, 2, \dots, k$). То есть

$$D(\hat{F}_n, G) = \sum_{i=1}^s \sum_{j=1}^k \frac{(\nu_{ij} - (\nu_{i \cdot} \nu_{\cdot j})/n)^2}{(\nu_{i \cdot} \nu_{\cdot j})/n}.$$

Количество параметров, оцененных по выборке, равно $(s-1)+(k-1)$. Количество подмножеств X_{ij} , $i = 1, 2, \dots, s$; $j = 1, 2, \dots, k$, на которые разбивается выборочное пространство X , равно sk . Поэтому число степеней свободы χ^2 -распределения, предельного для $D(\hat{F}_n, G)$, равно $sk - 1 - ((s-1) + (k-1)) = (s-1)(k-1)$.

Таким образом, гипотезу H_0 о независимости случайных величин ξ и η отклоняем, если

$$\sum_{i=1}^s \sum_{j=1}^k \frac{(\nu_{ij} - (\nu_{i \cdot} \nu_{\cdot j})/n)^2}{(\nu_{i \cdot} \nu_{\cdot j})/n} > \chi_{\alpha; (s-1)(k-1)}^2,$$

и не отклоняем в противном случае; при этом с вероятностью α гипотезу H_0 будем отклонять, когда она верна.

Значения случайных величин ξ и η еще называют признаками, и в этом случае говорят о критерии χ^2 для независимости признаков.

Разбиение выборочного пространства (X, Y) двумерной случайной величины $\zeta = (\xi, \eta)$ на непересекающиеся подмножества $X_i \times Y_j$ мы проводим сами. При этом необходимо следить, чтобы для гипотетических вероятностей $\hat{p}_{i \cdot} \hat{p}_{\cdot j}$ выборочным значениям (ξ, η) попасть в $X_i \times Y_j$ выполнялись неравенства

$$(\hat{p}_{i \cdot} \hat{p}_{\cdot j})n = \nu_{i \cdot} \nu_{\cdot j} / n \geq 10$$

(при всех возможных значениях i и j), где n — объем выборки.

Замечание. Следует отметить, что имея дело со статистическими зависимостями, необходимо быть очень осторожным. Из наличия статистической зависимости (отсутствия статистической независимости) не следует причинная зависимость. Идеи относительно причинных связей должны приходить “извне” статистики.

Пример 18.3.1 (прививка против холеры). В таблице (Гринвуд, Юл, 1915) приведены данные о 818 случаях, классифицированных по двум признакам: наличию прививки против холеры и отсутствию заболевания.

Можно ли на основании этих данных прийти к выводу о зависимости между отсутствием заболевания и наличием прививки?

Наличие прививки	Наличие заболевания		Всего
	Не заболели	Заболели	
Привитые	276	3	279
Непривитые	473	66	539
Всего	749	69	818

Решение. В терминах проверки статистических гипотез поставленная задача формулируется как задача проверки гипотезы о независимости признаков (наличие прививки и отсутствие заболевания).

Таблица сопряженности признаков (см. табл. 18.3.1) приведена в условии примера, заметим, что здесь $s = 2$, $k = 2$, $n = 818$.

Поскольку

$$(\hat{p}_i \cdot \hat{p}_{\cdot j})n = \nu_i \cdot \nu_{\cdot j} / n \geq 10, \quad i, j = 1, 2,$$

то можно воспользоваться критерием χ^2 .

Значение отклонения между эмпирическим распределением $\hat{F}_n$ и гипотетическим G :

$$\begin{aligned} D(\hat{F}_n, G) &= \sum_{i=1}^s \sum_{j=1}^k \frac{(\nu_{ij} - (\nu_i \cdot \nu_{\cdot j})/n)^2}{(\nu_i \cdot \nu_{\cdot j})/n} = \\ &= \frac{(276 - 279 \cdot 749/818)^2}{279 \cdot 749/818} + \frac{(3 - 279 \cdot 69/818)^2}{279 \cdot 69/818} + \\ &+ \frac{(473 - 539 \cdot 749/818)^2}{539 \cdot 749/818} + \frac{(66 - 539 \cdot 69/818)^2}{539 \cdot 69/818} = 29,61. \end{aligned}$$

Таким образом,

$$D(\hat{F}_n, G) = 29,61 > 3,84 = \chi_{0,05;1}^2 = \chi_{\alpha;(s-1)(k-1)}^2.$$

Поэтому согласно критерию χ^2 гипотеза о независимости признаков (отсутствие заболевания и наличие прививки) отклоняется; она противоречит имеющимся данным. Другими словами, приведенные экспериментальные данные дают основания утверждать, что эффект прививки против холеры существует.

18.4 Задачи

АЗ: 18.2, 18.7, 18.17, 18.20.

СЗ: 18.14, 18.23, 18.29, 18.52.

18.1 (адекватность модели). В ряде прикладных задач при использовании методов теории вероятностей,

прежде чем вычислять вероятность того или иного события, необходимо построить математическую модель стохастического эксперимента (вероятностное пространство); вероятность события можно вычислять только тогда, когда определено (явно или неявно) вероятностное пространство. Теория вероятностей дает правила, по которым в данной модели вычисляются вероятности событий, но ничего не говорит о выборе модели стохастического эксперимента. И принципиальный вопрос: “Адекватно ли описывает предложенная модель стохастический эксперимент?” остается открытым. Дать ответ на этот вопрос может математическая статистика. Ответ дается в терминах проверки статистических гипотез.

Например, рассматривается стохастический эксперимент, состоящий в подбрасывании игральной кости. Необходимо вычислить вероятность того, что выпадет четное число очков.

В качестве множества элементарных событий Ω выберем $\{1, 2, 3, 4, 5, 6\}$. Вероятность элементарных событий можно задать, скажем, так:

$$P(i) = \frac{1}{6}, \quad i = 1, 2, \dots, 6.$$

Тем самым вероятностное пространство определено. Почему именно $P(i) = 1/6, i = 1, 2, \dots, 6$, теория вероятностей не дает ответа. Вероятности мы задаем сами. Кто-то другой может задать эти вероятности, например, так:

$$P(i) = i/21, \quad i = 1, 2, \dots, 6,$$

— пропорционально числу очков на грани (или еще каким-нибудь способом), и это (как возможная гипотеза) будет ничуть не хуже.

По известной математической модели $\{\Omega, P\}$ вычислить вероятность случайного события просто. Например, вероятность события A — “выпало четное число очков”, вычисляется так:

$$P(A) = P(\{2, 4, 6\}) = P(2) + P(4) + P(6).$$

Для симметричной игральной кости имеем

$$P(A) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Таким образом, если предложенная модель адекватно описывает эксперимент, то $P(A) = 1/2$.

Однако, *адекватно ли описывается эксперимент предложенной моделью?*

Чтобы ответить на этот вопрос, необходимо провести эксперимент: подбросить кубик достаточно большое число раз, зафиксировать результат и далее проверить, согласуется ли модель с результатом эксперимента.

Результаты наблюдений удобно представить в таком виде:

i	1	2	3	4	5	6
n_i						

где i — количество выпавших очков; n_i — число подбрасываний, в которых было зафиксировано i очков.

18.2 (количество смертей от удара копытом).

Далее приведены классические данные фон Борткевича о количестве лиц, убитых ударом копыта коня в 10 прусских армейских корпусах за 20 лет (1875—1895):

i	0	1	2	3	4	5 и более	Всего
n_i	109	65	22	3	1	0	200

где i — количество смертей в одном корпусе за год; n_i — количество корпусов, в которых случилось i смертей.

Проверить гипотезу о пуассоновском распределении количества смертей от удара копытом коня в одном корпусе в течение года.

18.3. В таблице приведены данные К. Пирсона о возрасте кандидатов, которые сдали и не сдали вступительные экзамены в Лондонский университет в 1908 и 1909 гг. (для двух старших возрастных групп указаны оценки среднего возраста).

Возраст кандидата	Результат экзамена		Всего
	Сдали	Не сдали	
16	583	563	1146
17	666	980	1646
18	525	868	1393
19—21	383	814	1197
22—30 (средний возраст 25)	214	439	653
Свыше 30 (средний возраст 33)	40	81	121
Всего	2411	3745	6156

Свидетельствуют ли эти данные о наличии зависимости между возрастом кандидатов и успешно сданными экзаменами?

18.4. Стохастический эксперимент состоит в подбрасывании двух игральных костей и регистрации минимального числа очков, выпавших на них.

Предложить математическую модель этого стохастического эксперимента. Проверить ее адекватность эксперименту. Для этого провести эксперимент достаточное число раз и проверить, согласуется ли предложенная модель с полученными данными.

Сформулировать и решить поставленную задачу как задачу проверки статистических гипотез (воспользоваться критерием χ^2).

Результаты эксперимента удобно представить в виде таблицы:

i	1	2	3	4	5	6
n_i						

где i — минимальное число очков, выпавших на игральном костях; n_i — количество экспериментов, в которых минимальное число очков на игральном костях оказалось равным i , $i = 1, 2, \dots, 6$.

З а м е ч а н и е. О выборе модели эксперимента и проверке ее адекватности см. задачу 18.1.

У к а з а н и е. Чтобы предложить гипотетическое распределение (в предположении, что игральные кости симметричны), найдите распределение случайной величины $\min\{\xi, \eta\}$, где ξ, η — число очков на игральном костях. Распределение функции $g(\zeta) = \min\{\xi, \eta\}$ от случайной величины $\zeta = (\xi, \eta)$ по распределению P_ζ можно найти, воспользовавшись формулой

$$P\{g(\zeta) \in B\} = P\{g(\xi, \eta) \in B\} = \sum_{(x_i, y_j): g(x_i, y_j) \in B} P_\zeta(x_i, y_j)$$

(относительно распределения функции от случайных величин см. также гл. 5 и, в частности, пример 5.1.2).

18.5 (вероятность рождения мальчика). С 1871 по 1900 гг. в Швейцарии родились 1 359 671 мальчиков и 1 285 086 девочек.

Согласуется ли с этими данными гипотеза H_0 : вероятность рождения мальчика равна: а) 0,5? б) 0,515?

18.6 (распределение α -частиц). В эксперименте Резерфорда и Гейгера радиоактивное вещество наблюдалось в течение $N = 2612$ интервалов времени (каждый длительностью $1/8$ мин), для которых регистрировалось количество α -частиц, достигших счетчика.

В таблице приведено количество N_k интервалов времени, в течение которых наблюдалось k α -частиц. Общее количество частиц $T = \sum_k k N_k = 10\,132$. Среднее их количество, зарегистрированное за один интервал, равно $\lambda = T/N = 10\,132/2\,612 = 3,879$.

k	N_k	k	N_k
0	57	6	273
1	203	7	139
2	383	8	49
3	525	9	27
4	532	10	10
5	408	> 10	6
		Всего	2612

Воспользовавшись критерием χ^2 , проверить гипотезу о пуассоновском распределении числа частиц, достигших счетчика, за интервал времени длительностью $1/8$ мин.

18.7 (смышленость и материальные условия). В таблице приведены результаты обследования 697 школьников.

Мальчики были упорядочены согласно IQ и в соответствии с условиями их жизни дома. При этом использованы обозначения: A — очень способный, B — достаточно умный, C — имеет средние способности, D — недостаточно развит, E — умственно отсталый.

Материальная обеспеченность	Уровень смышлености					Всего
	A	B	C	D	E	
Хорошая	33	137	125	47	8	350
Плохая	21	127	129	61	9	347
Всего	54	264	254	108	17	697

Можно ли утверждать, что условия жизни (обеспеченность) детей влияют на их смышленость?

З а м е ч а н и е. IQ — Intellectual quality (умственные способности) — показатель умственных способностей школьников в баллах, используемый в американской педагогической практике.

18.8. Имеется три монеты: одна стоимостью 1 коп. и две — стоимостью 10 коп. Стохастический эксперимент состоит в выборе наудачу одной из монет (с возвращением) и регистрации количества шагов до первого появления монеты стоимостью 10 коп., т. е. в регистрации числа появлений монеты стоимостью 1 коп. до первого появления монеты стоимостью 10 коп.

Предложить математическую модель этого стохастического эксперимента. Проверить ее адекватность эксперименту. Для этого получить выборку (достаточно большого объема) и проверить, согласуется ли предложенная модель с результатами эксперимента — полученной выборкой.

Результаты эксперимента удобно представить в виде

i	0	1	2	3	4	5	6	7	8	9	...
n_i											

где i — количество шагов до первого появления монеты стоимостью 10 коп.; n_i — количество экспериментов, в которых до появления монеты стоимостью 10 коп. было сделано i шагов.

З а м е ч а н и е. О выборе модели стохастического эксперимента и проверке ее адекватности эксперименту см. задачу 18.1.

18.9. Четыре монеты были подброшены 20 160 раз, при этом комбинации: четыре герба, три герба и решетка, два герба и две решетки, один герб и три решетки, четыре решетки — появились соответственно такое число раз:

1181, 4909, 7583, 5085, 1402

(данные эксперимента, проведенного В. И. Романовским).

Свидетельствуют ли эти данные о том, что количество гербов, появившихся на четырех монетах, является биномиально распределенной случайной величиной?

Сформулировать и решить поставленную задачу как задачу проверки статистических гипотез.

18.10 (аромат сигарет). Менеджеры табачной фирмы хотели бы знать, можно ли отправлять заказчикам сигареты и трубочный табак в общей упаковке. (Если при этом качество сигарет не становится хуже, то можно существенно сократить затраты на транспортировку.) Для этого провели эксперимент. Изготовили 400 картонных коробок и в 250 из них положили табачные изделия обоих типов, а в остальные — только сигареты. Через месяц коробки открыли и все 400 упаковок сигарет разложили в случайном порядке. Несколько экспертов анализировали аромат сигарет, пытаясь обнаружить его отличие от начального. Результаты эксперимента приведены в таблице.

Вывод относительно аромата	Тип упаковки		Всего
	Общая	Раздельная	
Не изменился	72	119	191
Изменился	178	31	209
Всего	250	150	400

Можно ли на основании этих данных утверждать, что связь между ароматом сигарет и типом упаковки отсутствует?

18.11 (эксперимент Сведеборга). В начале XIX столетия было открыто замечательное явление, получившее название броуновского движения (по имени английского ботаника Р. Броуна, открывшего его). Это явление состоит в том, что мельчайшие частицы вещества, взвешенные в жидкости, пребывают в хаотическом движении без видимых на то причин.

Причины этого, казалось бы, самопроизвольного движения долго не могли объяснить, пока кинетическая теория газов не дала простого и исчерпывающего объяснения: движение взвешенных частиц является результатом ударов молекул жидкости по этим частицам. Кинетическая теория дает возможность вычислить вероятность того, что в данном объеме жидкости не будет ни одной частицы взвешенного вещества, таких частиц будет одна, две, три и т. д.

Для проверки результатов теории был проведен ряд экспериментов.

Далее приведены результаты эксперимента шведского физика Сведеборга.

Эксперимент состоял в наблюдении и регистрации числа мельчайших частиц золота, которые попали в поле зрения микроскопа через одинаковые промежутки времени.

Было проведено 518 таких наблюдений. При этом были получены следующие результаты.

i	0	1	2	3	4	5	6	7	8 и более
n_i	112	168	130	69	32	5	1	1	0

В таблице i — количество частиц золота в поле зрения микроскопа; n_i — количество наблюдений, когда в поле зрения микроскопа было зарегистрировано i частиц золота.

Проверить гипотезу о пуассоновском распределении числа частиц золота, наблюдавшихся в поле зрения микроскопа.

18.12. В таблице приведены официальные данные шведской статистики за 1935 г. о распределении новорожденных по месяцам.

Месяц	Девочки	Всего детей	Месяц	Девочки	Всего детей
Январь	3537	7280	Июль	3621	7585
Февраль	3407	6957	Август	3596	7393
Март	3866	7883	Сентябрь	3491	7203
Апрель	3711	7884	Октябрь	3391	6903
Май	3775	7892	Ноябрь	3160	6552
Июнь	3665	7609	Декабрь	3371	7132
			Всего	42 591	88 273

Свидетельствуют ли эти данные о том, что в каждом месяце в течение года дети рождаются одинаково часто? Девочки рождаются одинаково часто? Мальчики рождаются одинаково часто?

18.13 (непредвзятость экспериментатора). Каждую из 150 спичек разломайте наудачу на две части. Линейкой с миллиметровым делением измерьте с точностью до десятых долей миллиметра длину 300 полученных частей. При этом придется оценивать первый знак после

запятой в десятичной записи длины части спички в миллиметрах. Зафиксируйте этот знак (последнюю цифру записи).

Результаты измерений удобно представить в таком виде:

i	0	1	2	3	4	5	6	7	8	9
n_i										

где i — последняя цифра в записи результата измерений длины части спички, $i = 0, 1, \dots, 9$; n_i — число появлений последней цифры i в измерениях.

Проверьте с помощью критерия χ^2 непредвзятость ваших измерений.

З а м е ч а н и е. Предвзятость при снятии показаний встречается достаточно часто. Сомнительно, чтобы была какая-то объективная причина, обуславливающая более частое появление одних цифр по сравнению с другими; поэтому естественно предположить, что неодинаковая частота появления цифр указывает на предвзятость наблюдателя. Даже те наблюдатели, которые знают о возможности предвзятости и о необходимости осторожности при снятии показаний, все-таки не всегда могут избежать ее.

18.14 (ошибочное соединение телефонных номеров). В таблице приведены данные о количестве ошибочных соединений телефонных номеров. Всего наблюдалось 267 номеров.

k	n_k	k	n_k	k	n_k
0	0	6	22	12	18
1	1	7	43	13	12
2	0	8	31	14	7
3	5	9	40	15	6
4	11	10	35	16	2
5	14	11	20	Всего	267

Здесь k — количество ошибочных соединений; n_k — число номеров, для которых было зафиксировано точно k ошибочных соединений, $k = 0, 1, \dots, 16$.

Проверить гипотезу о пуассоновском распределении количества ошибочных соединений.

18.15 (цвет волос и цвет глаз). В таблице приведены данные о 147 наудачу выбранных студентах, которые были распределены согласно цвету их волос (светлые, темные) и глаз (голубые, карие).

Цвет волос	Цвет глаз		Всего
	Голубые	Карие	
Темные	31	41	72
Светлые	40	35	75
Всего	71	76	147

Можно ли на основании этих данных сделать вывод о том, что цвет глаз связан с цветом волос?

18.16 (предвзятость наблюдателя). Перед датчиком, состоящим из круглого диска, разделенного на 10 одинаковых секторов, занумерованных от 0 до 9, поставили наблюдателя. Диск вращался с большой скоростью. Время от времени перед ним вспыхивала электрическая лампочка на столь короткое время, что диск казался неподвижным. Наблюдатель должен был смотреть на диск и записывать номер того сектора, который в момент вспышки лампочки находился напротив фиксированного указателя.

Прибор предназначался для получения случайных чисел и в самом деле их выдавал, когда в эксперименте участвовал другой наблюдатель. Однако наблюдатель, о котором шла речь выше, снимал показания с заметной предвзятостью.

Частоты появления цифр в 10 000 выполненных им наблюдениях приведены в таблице (в таблице частота цифры — количество ее появлений в 10 000 наблюдениях).

Проверить, свидетельствуют ли эти данные о предвзятости наблюдателя.

Цифра	Частота	Цифра	Частота
0	1083	5	1007
1	865	6	1081
2	1053	7	997
3	884	8	1025
4	1057	9	948
		Всего	10 000

З а м е ч а н и е. Если бы наблюдатель был непредвзятым, то цифры появлялись бы приблизительно с одинаковой частотой. Из таблицы, однако, видно, что он имел пристрастие к четным цифрам и предубеждение против нечетных цифр 1, 3 и 9. Причина такого поведения и предвзятости непонятна и загадочна, поскольку наблюдатель не должен был давать оценку, а должен только записывать то, что видит, или думал, что видит. Объяснение, по-видимому, состоит в том, что наблюдатель отдавал значительное предпочтение некоторым цифрам, при этом фиксировал не те из них, которые в момент вспышки лампочки в самом деле видел напротив указателя. Здесь мы имеем дело с одной из чрезвычайно сильных форм психологического пристрастия.

18.17. Рассматривается стохастический эксперимент, состоящий в подбрасывании четырех монет и регистрации количества выпавших гербов. Предложите математическую модель этого эксперимента. Проверьте ее адекватность эксперименту. Для этого проведите эксперимент достаточное число раз и воспользуйтесь критерием χ^2 .

Результаты подбрасываний четырех монет удобно представить в виде таблицы:

i	0	1	2	3	4
n_i					

где i — количество гербов, выпавших на четырех монетах; n_i — количество экспериментов, в которых выпало i гербов.

З а м е ч а н и е. О выборе модели стохастического эксперимента и проверке ее адекватности эксперименту см. задачу 18.1.

18.18 (перестройка хромосом в клетке). Рентгеновское облучение вызывает в органических клетках определенные процессы, которые будем называть перестройкой хромосом. Число перестроек хромосом в клетке согласно теории должно подчиняться распределению Пуассона. В эксперименте подсчитывали количество перестроек хромосом под действием рентгеновских лучей. Результаты эксперимента оказались следующими.

i	0	1	2	3	4 и более	Всего
n_i	434	195	44	9	0	682

Здесь i — количество изменений в клетке; n_i — количество клеток, в которых произошло i изменений.

Согласуется ли с приведенными данными гипотеза о пуассоновском распределении числа перестроек?

18.19 (глухонемота и пол). При переписи населения Англии и Уэльса в 1901 г. было зарегистрировано (с точностью до тысяч) 15 729 000 мужчин и 16 799 000 женщин, из них 3497 мужчин и 3072 женщин глухонемые от рождения.

Проверить гипотезу о том, что глухонемота не связана с полом.

18.20 (случайно и произвольно). Случайно не означает произвольно: случайность подчиняется своим строгим законам. Например, на первый взгляд кажется, что предложить случайную последовательность цифр просто. Однако на самом деле это могут сделать немногие люди. Последовательность случайных цифр должна удовлетворять ряду требований случайности. В частности, от такой последовательности естественно потребовать, чтобы появление цифр $0, 1, \dots, 9$ было равновероятным.

Проверьте свои способности, а именно: выпишите случайную (с вашей точки зрения) последовательность: 1) из 150 “четырёхзначных” чисел (нуль может занимать первое место слева, как и любая другая цифра); 2) из 600 цифр. Проверьте с помощью критерия χ^2 , действительно ли в предложенных вами последовательностях цифры $0, 1, \dots, 9$ встречаются с вероятностью $1/10$.

Сравните результаты этих экспериментов.

Примечание. Выписывая последовательность, не просматривайте и не используйте каким либо иным способом уже выписанную часть последовательности: каждое следующее число (каждую следующую цифру) выписывайте так, как будто вы его (ее) пишете впервые, как будто до этого ничего не выписывалось.

Замечание. Встречаются люди, у которых психологические процессы настолько хорошо сбалансированы, что они могут получать случайные выборки. Обычно такие лица считают себя чрезвычайно одаренными. См. также замечание к задаче 18.17.

18.21. Стохастический эксперимент состоит в последовательном подбрасывании монеты четыре раза и ре-

гистрации результатов следующим образом: выпадение герба (Г) обозначается единицей, а решетки (Р) — нулем. Так что, выпадение, например, ГРРГ регистрируется как 1001. Рассмотрим эти последовательности из нулей и единиц как запись чисел в двоичной системе исчисления, т. е. 1001 — это запись в двоичной системе исчисления девятки.

Будем фиксировать только числа, меньшие 10.

Результаты эксперимента удобно представить в виде таблицы:

i	0	1	2	3	4	5	6	7	8	9
n_i										

где n_i — количество экспериментов, в которых было зарегистрировано число i .

Можно ли считать появление цифр 0, 1, 2, ..., 9 в описанном эксперименте равновероятным?

Дать ответ в терминах проверки статистических гипотез. Для этого предложить математическую модель стохастического эксперимента, провести его достаточное число раз и, воспользовавшись критерием χ^2 , проверить адекватность модели эксперименту.

18.22 (умственные способности и качество одежды). В таблице приведены данные (полученные Гилби) о 1725 учениках, классифицированных в соответствии с: 1) качеством их одежды; 2) умственными способностями. При этом для характеристики умственных способностей использована следующая градация: A — умственно отсталый; B — медлительный и недостаточно развитый; C — недостаточно развитый; D — медлительный, но умный; E — достаточно умный; F — явно способный; G — очень способный.

Умственные способности	Качество одежды				Всего
	Очень хорошее	Хорошее	Сносное	Очень плохое	
A и B	33	41	39	17	130
C	48	100	58	13	219
D	113	202	70	22	407
E	209	255	61	10	535
F	194	138	33	10	375
G	39	15	4	1	59
Всего	636	751	265	73	1725

Можно ли на основании этих данных сделать вывод, что качество одежды учеников и их умственные способности — независимые признаки?

18.23 (бактерии в чашке Петри). В чашке Петри наблюдаются колонии бактерий. Под микроскопом их видно как темные пятнышки. Чашка Петри (ее дно) разделена на маленькие квадраты, в каждом из которых подсчитывается количество колоний (пятнышек). Результаты наблюдений приведены в таблице:

k	0	1	2	3	4	5	6	7
n_k	5	19	26	26	21	13	8	0

где k — количество колоний в квадрате; n_k — количество квадратов с k колониями.

Проверить гипотезу о пуассоновском распределении количества колоний на квадрат.

18.24. В таблице приведены вероятности появления в тексте букв русского алфавита (включая пробел, он обозначен черточкой).

Вообще говоря, вероятности появления букв в тексте зависят от его характера, особенно это заметно в стихах. В качестве примера можно назвать стихотворение К. Д. Бальмонта “Камыши”:

Полночной порою в болотной глуши

Чуть слышно, бесшумно шуршат камыши ...

Все построено на обыгрывании шипящих звуков “ч” и “ш”, поэтому частота их появления заметно выше приведенных в таблице.

—	0,175	и	0,062	р	0,040	м	0,026
о	0,090	т	0,053	в	0,038	д	0,025
е, ё	0,072	н	0,053	л	0,035	п	0,023
а	0,062	с	0,045	к	0,028	у	0,021
я	0,018	б	0,014	х	0,009	ц	0,004
ы	0,016	г	0,013	ж	0,007	щ	0,003
з	0,016	ч	0,012	ю	0,006	э	0,003
ь, ъ	0,014	й	0,010	ш	0,006	ф	0,002

Рассмотрите отрывок из поэмы А. С. Пушкина “Руслан и Людмила” от “У лукоморья дуб зеленый ...” до “... сказку эту теперь поведаю я свету”.

Проверьте, согласуются ли в этом отрывке частоты появления букв с вероятностями, приведенными в таблице.

18.25. Стохастический эксперимент состоит в подсчете числа появлений пары (1;1) на 100 пар, извлеченных из таблицы случайных чисел, например из табл. 22.10.1.

Предложить математическую модель стохастического эксперимента. Проверить ее адекватность эксперименту. Для этого провести эксперимент достаточное число раз и воспользоваться критерием χ^2 .

Данные удобно представить в следующем виде:

i	0	1	2	3	4	5	6	7	8	9	...
n_i											

где i — число появлений пары (1;1) на 100 пар, извлеченных из таблицы случайных чисел; n_i — количество тех экспериментов, в которых пара (1;1) встречается i раз.

З а м е ч а н и е. О выборе модели стохастического эксперимента и проверке ее адекватности эксперименту см. задачу 18.1.

18.26 (селекция гороха Г. Менделем). В экспериментах по селекции гороха Г. Мендель наблюдал частоту появления различных видов семян (в этой задаче частота — количество семян определенного вида), полученных в результате скрещивания растений с круглыми желтыми и морщинистыми зелеными семенами. Эти данные и значения теоретических вероятностей, определенные согласно теории наследственности Менделя, приведены в таблице.

Вид семян	Частота	Вероятность
Круглые желтые	315	9/16
Морщинистые желтые	101	3/16
Круглые зеленые	108	3/16
Морщинистые зеленые	32	1/16
Всего	556	1

Согласуются ли вероятности, полученные согласно теории Менделя, с приведенными экспериментальными данными?

18.27 (смышленость и качество питания). В социальном обзоре (Пирсон и Моул, 1925) 618 мальчиков были классифицированы согласно уровню их смышлености и качеству питания.

Результаты обследований приведены в таблице. При этом использованы обозначения: A — очень способный, B — способный, C — смышленный, D — недостаточно смышленный, E — медлительный, F — очень медлительный.

Качество питания	Уровень смышлености						Всего
	A	B	C	D	E	F	
Хорошее	9	27	60	63	24	5	188
Посредственное	5	41	126	120	36	6	334
Плохое	5	12	38	32	8	1	96
Всего	19	80	224	215	68	12	618

Существует ли связь между качеством питания детей и их смышленостью?

18.28 (автомобильные номера). Стохастический эксперимент состоит в фиксировании цифр числовой части номеров автомобилей, проезжающих мимо наблюдателя. Относительно появления цифр выдвигается вполне естественная гипотеза: цифры в автомобильных номерах встречаются одинаково часто (вероятность появления цифр в автомобильных номерах одинакова).

С целью проверки выдвинутой гипотезы, зарегистрируйте числовую часть номеров 40—50 автомобилей, проезжающих мимо вас (почему этого будет достаточно?) и воспользуйтесь критерием χ^2 . Данные удобно представить в следующем виде:

i	0	1	2	3	4	5	6	7	8	9
n_i										

где n_i — количество появлений цифры i в числовых частях автомобильных номеров.

18.29. Во время Второй мировой войны на Лондон упало 537 самолетов-снарядов. Всю территорию Лондона разбили на 576 участков площадью 0,25 км². Ниже приведено количество ν_i участков, на которые упало i снарядов.

i	0	1	2	3	4	5	6	7	8 и более	Всего
ν_i	229	211	93	35	7	0	0	1	0	576

Воспользовавшись критерием χ^2 , проверить гипотезу о пуассоновском распределении количества самолетов-снарядов, упавших на участок.

18.30. Одновременно подбрасывают 12 игральных костей и регистрируют значение случайной величины ξ — количества костей, на которых выпало число очков большее 3. Данные 4096 экспериментов приведены в таблице.

i	n_i	i	n_i
0	0	7	847
1	7	8	536
2	60	9	257
3	198	10	71
4	430	11	11
5	731	12	0
6	948	Всего 4096	

Здесь i — значение случайной величины ξ ; n_i — количество экспериментов, в которых случайная величина ξ приняла значение i , $i = 0, 1, \dots, 12$.

Проверить гипотезу о биномиальном распределении случайной величины ξ .

18.31 (цвет волос и цвет бровей). В таблице приведено распределение цвета волос и цвета бровей у 46 542 шведских призывников.

Цвет бровей	Цвет волос		Всего
	Светлые или рыжие	Темные	
Светлые или рыжие	30 472	3238	33 710
Темные	3364	9468	12 832
Всего	33 836	12 706	46 542

Свидетельствуют ли эти данные о связи между цветом волос и цветом бровей?

18.32. В таблице приведены длины интервалов между последовательными импульсами вдоль нервного волокна (единица измерения $1/50$ с), полученные доктором П. Фетом и профессором Б.Кацу (Лондонский университет).

3,0	25,5	5,5	14,5	18,0	7,0	27,5	14,0
2,0	0,5	47,0	36,5	2,5	3,5	5,5	19,0
10,5	24,5	19,0	19,0	0,5	3,0	6,5	3,0
0,5	8,0	2,5	5,0	8,0	3,0	3,0	3,0
5,5	22,0	2,5	4,5	2,0	13,5	25,0	12,5
12,5	4,0	0,5	35,0	2,0	4,0	8,0	19,0
4,0	16,0	19,5	29,0	28,0	37,0	7,5	13,0
12,5	8,5	32,5	30,5	7,5	13,0	1,5	2,5
17,0	3,5	5,0	4,5	1,0	15,0		

Проверить гипотезу о: а) нормальном распределении; б) равномерном распределении длины интервала между импульсами.

18.33. Последовательно дважды подбрасывают пару монет и регистрируют число ξ выпавших гербов на паре монет при первом подбрасывании и η — при втором, т. е. регистрируют значение случайной величины $\zeta = (\xi, \eta)$.

Предложить математическую модель этого стохастического эксперимента. Проверить ее адекватность эксперименту. С этой целью провести эксперимент достаточное число раз и воспользоваться критерием χ^2 .

Результаты подбрасываний удобно представить в виде следующей таблицы:

Значения ξ	Значения η			Сумма
	0	1	2	
0	ν_{00}	ν_{01}	ν_{02}	$\nu_{0\cdot}$
1	ν_{10}	ν_{11}	ν_{12}	$\nu_{1\cdot}$
2	ν_{20}	ν_{21}	ν_{22}	$\nu_{2\cdot}$
Сумма	$\nu_{\cdot 0}$	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	n

Здесь ν_{ij} — количество экспериментов, в которых случайная величина $\zeta = (\xi, \eta)$ приняла значение (i, j) ,

$$\nu_{i\cdot} = \sum_{j=0}^2 \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=0}^2 \nu_{ij}, \quad i, j = 0, 1, 2, \quad n = \sum_{i=0}^2 \sum_{j=0}^2 \nu_{ij}.$$

З а м е ч а н и е. Относительно выбора модели стохастического эксперимента и проверки ее адекватности эксперименту см. задачу 18.1.

18.34. В таблице приведены данные о 1426 заключенных, классифицированных по отношению к алкогольной зависимости (алкоголик, не алкоголик) и характеру преступлений, за которые их осудили (данные Горинга, цитированные К. Пирсоном). Столбцы таблицы упорядочены в соответствии с “интеллектуальностью” вида преступления, хотя это упорядочение довольно условное.

Вид преступления	Алкоголики	Не алкоголики	Всего
Поджог	50	43	93
Изнасилование	88	62	150
Насильственные действия	155	110	265
Воровство	379	300	679
Изготовление фальшивых денег	18	14	32
Мошенничество	63	144	207
Всего	753	673	1426

Можно ли на основании этих данных сделать вывод о наличии связи между алкоголизмом и характером преступления?

18.35 (заряд электрона). В таблице приведены 58 значений величины e , найденных Р. Милликеном при определении заряда электрона, который равен $e \cdot 10^{-10}$ ед. СГС.

1	4,740	13	4,769	25	4,778	37	4,788	49	4,795
2	4,747	14	4,771	26	4,779	38	4,788	50	4,797
3	4,749	15	4,771	27	4,779	39	4,789	51	4,799
4	4,758	16	4,772	28	4,779	40	4,789	52	4,799
5	4,761	17	4,772	29	4,781	41	4,790	53	4,801
6	4,764	18	4,772	30	4,781	42	4,790	54	4,805
7	4,764	19	4,774	31	4,782	43	4,790	55	4,806
8	4,764	20	4,775	32	4,783	44	4,791	56	4,808
9	4,765	21	4,775	33	4,783	45	4,791	57	4,809
10	4,767	22	4,776	34	4,785	46	4,791	58	4,810
11	4,768	23	4,777	35	4,785	47	4,792		
12	4,769	24	4,777	36	4,785	48	4,792		

Проверить гипотезу о нормальном распределении результатов измерений величины e при определении заряда электрона.

18.36. Последовательно дважды подбрасывают пару монет и регистрируют число ξ выпавших гербов при первом подбрасывании и максимальное число η гербов выпавших при первом и вторым подбрасываниях, т. е. регистрируют значение случайной величины $\zeta = (\xi, \eta)$.

Предложить математическую модель этого стохастического эксперимента. Адекватно ли описывает его предложенная модель? Являются ли случайные величины ξ и η независимыми?

Дать ответ на вопрос об адекватности модели в терминах проверки статистических гипотез. Для этого провести стохастический эксперимент достаточное число раз и воспользоваться критерием χ^2 . Данные удобно представить в виде таблицы.

Значения ξ	Значения η			Сумма
	0	1	2	
0	ν_{00}	ν_{01}	ν_{02}	$\nu_{0\cdot}$
1	ν_{10}	ν_{11}	ν_{12}	$\nu_{1\cdot}$
2	ν_{20}	ν_{21}	ν_{22}	$\nu_{2\cdot}$
Сумма	$\nu_{\cdot 0}$	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	n

Здесь ν_{ij} — количество экспериментов, в которых случайная величина $\zeta = (\xi, \eta)$ приняла значение (i, j) , $i, j = 0, 1, 2$;

$$\nu_{i\cdot} = \sum_{j=0}^2 \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=0}^2 \nu_{ij}, \quad i, j = 0, 1, 2; \quad n = \sum_{i=0}^2 \sum_{j=0}^2 \nu_{ij}.$$

З а м е ч а н и е 1. О выборе математической модели стохастического эксперимента и проверке ее адекватности см. задачу 18.1.

З а м е ч а н и е 2. Распределение функции $g(\zeta)$ случайной величины $\zeta = (\xi, \eta)$ (в частности, функции $g(\zeta) = g(\xi, \eta) = (\xi, \max\{\xi, \eta\})$) по распределению ζ можно

найти, воспользовавшись соотношением

$$P\{g(\xi, \eta) = (x, y)\} = \sum_{\{(i, j): g(i, j) = (x, y)\}} P_{\zeta}(i, j).$$

18.37 (о телепатах). В Гарвардском университете, университете Дюка и других университетах при проверке способностей телепатов “читать” мысли, в частности, ставились эксперименты по чтению цифр $0, 1, \dots, 9$.

Проведите аналогичный эксперимент. Наудачу задумайте цифру: $0, 1, \dots, 9$ и зафиксируйте результат. Телепат читает задуманную вами цифру и также фиксирует результат. Эксперимент проводится достаточное количество раз. Если телепат в самом деле обладает способностями читать мысли, то доля правильно прочитанных цифр: нуль читается как нуль, единица — как единица, и т. д., девятка — как девятка, т. е. доля успехов будет большой (успех — правильно прочитанная цифра, неудача — неправильно прочитанная).

По результатам проведенного эксперимента сделать вывод относительно способностей телепата читать мысли на расстоянии.

Указание. Для чистоты эксперимента, задумывая цифры $0, 1, \dots, 9$, пользуйтесь таблицей случайных чисел.

18.38 (несчастные случаи с водителями автобусов). Частотное распределение 166 водителей лондонских автобусов в соответствии с количеством несчастных случаев, случившихся с ними в течение одного года, приведено в таблице.

i	n_i	i	n_i
0	45	6	3
1	36	7	2
2	40	8	1
3	19	9	0
4	12	10 и более	0
5	8	Всего	166

Здесь i — число несчастных случаев; n_i — количество водителей, с которыми случилось i несчастных случаев, $i = 0, 1, \dots$

Можно ли считать, что количество несчастных случаев, случившихся при участии водителя в течение года, подчиняется распределению Пуассона?

18.39 (пирожные и бактериальные колонии). В таблице приведены данные о колониях бактерий, содержащихся в трех видах пирожных (данные А. Абрахамсона).

Можно ли утверждать, что существует зависимость между видами пирожных и размерами бактериальных колоний, содержащихся в них?

Вид пирожных	Размеры колонии			Всего
	Малые	Средние	Большие	
Эклер	92	37	46	175
Наполеон	53	15	19	87
Ореховое	75	19	12	106
Всего	220	71	77	368

Ответ дать в терминах проверки статистических гипотез.

18.40 (момент последнего уравнивания). Симметричную монету подбрасывают четыре раза и регистрируют значение случайной величины — момент, когда в последний раз количество гербов сравняется с количеством решеток (момент последнего уравнивания), этими моментами в эксперименте являются 0; 2; 4.

Относительно момента последнего уравнивания выдвигается гипотеза: распределением момента последнего уравнивания является

$$\begin{pmatrix} 0 & 2 & 4 \\ 3/8 & 2/8 & 3/8 \end{pmatrix}$$

(дискретное распределение арксинуса порядка 2).

Проверить выдвинутую гипотезу. Для этого получить выборку момента последнего уравнивания (достаточного объема) и воспользоваться критерием χ^2 .

Результаты проведения экспериментов удобно представить в таком виде:

i	0	2	4
n_i			

где i — момент последнего уравнивания; n_i — количество наблюдений (экспериментов), в которых момент последнего уравнивания оказался равным i .

18.41 (показания механических часов). Механические часы, выставленные в витринах часовых магазинов, показывают случайное время. Выдвигается естественная гипотеза: показания часов распределены равномерно на интервале $[0;12)$. Результаты 1000 наблюдений приведены в таблице (интервал $[0;12)$ разбит на 12 равных частей: $[i, i+1)$, $i = 0, 1, \dots, 11$).

i	n_i	i	n_i
0	77	6	73
1	81	7	70
2	95	8	77
3	86	9	82
4	98	10	84
5	90	11	87
		Всего	1 000

Здесь i — номер интервала от i -го часа до $(i+1)$ -го, $i = 0, 1, \dots, 11$; n_i — количество часов, показания которых принадлежат i -му интервалу.

Согласуется ли выдвинутая гипотеза с этими данными?

18.42 (смысленность и сложение). В таблице приведены данные о степени смысленности учеников с атлетическим сложением и с неатлетическим.

Сложение ученика	Степень смысленности		Всего
	Высокая	Низкая	
Атлетическое	581	567	1 148
Неатлетическое	209	351	560
Всего	790	918	1 708

Что можно сказать относительно связи между смысленностью учеников и их сложением?

Ответ дать в терминах проверки статистических гипотез.

18.43. Данные о количестве трещин в стержне, обнаруженных при испытании 600 нейлоновых стержней, приведены в таблице

i	0	1	2	3	4	5	6 и более	Всего
n_i	275	207	81	23	8	6	0	600

где i — число трещин в стержне; n_i — количество стержней, в которых обнаружено i трещин.

Проверить гипотезу о пуассоновском распределении количества трещин в стержне.

18.44 (эксперимент К. Пирсона). В результате 24 000 подбрасываний монеты К. Пирсон зарегистрировал 12 012 случаев появления герба.

Согласуется ли гипотеза о симметричности монеты с этими данными?

18.45. В результате проверки 500 контейнеров со стеклянными изделиями получены такие данные о количестве поврежденных изделий:

i	n_i	i	n_i
0	199	5	3
1	169	6	1
2	87	7	1
3	31	8 и более	0
4	9	Всего	500

где i — число поврежденных изделий; n_i — количество контейнеров с i поврежденными изделиями, $i = 0, 1, \dots$

Можно ли считать, что количество поврежденных изделий, приходящееся на контейнер, подчиняется распределению Пуассона?

18.46. В таблице приведены значения толщины (в микрометрах) пластикового покрытия медной проволоки (в выборку вошли 225 бобин проволоки).

Толщина	Частота	Толщина	Частота
145	1	153	37
146	3	154	25
147	3	155	23
148	7	156	11
149	11	157	9
150	25	158	2
151	33	159	0
152	34	160	1

Можно ли на основании этих данных считать, что толщина покрытия проволоки имеет нормальное распределение?

З а м е ч а н и е. В этой задаче частота — это количество бобин проволоки, для которых была зарегистрирована данная толщина покрытия.

18.47 (возраст молодоженов и уровень их доходов). Проводилось обследование с целью обнаружения связи между возрастом вступления в первый брак и уровнем доходов молодоженов. Результаты обследования (количество семейных пар) приведены в таблице.

Уровень доходов	Возраст молодоженов		
	До 18	18—21	Старше 21
Низкий	45	25	15
Средний	35	60	25
Высокий	10	28	24

Свидетельствуют ли приведенные данные о наличии связи между возрастом вступления в первый брак и уровнем доходов молодоженов?

18.48. В эксперименте наблюдалась неотрицательная непрерывная случайная величина ξ . Для $n = 48$ наблюдений получены такие ее значения (упорядоченные и округленные до 0,01):

0,01	0,04	0,17	0,18	0,22	0,25	0,25	0,29
0,42	0,46	0,47	0,56	0,59	0,67	0,70	0,72
0,76	0,78	0,83	0,85	0,87	0,93	1,00	1,01
1,01	1,02	1,03	1,05	1,32	1,34	1,37	1,47
1,50	1,52	1,54	1,58	1,71	1,90	2,10	2,35
2,46	2,46	2,50	3,73	4,07	6,03	6,21	7,02

Проверить гипотезу о том, что эти данные являются реализацией выборки из показательного распределения с параметром $\lambda = 1$ (его функция распределения $F(x) = 1 - e^{-x}$, при $x > 0$).

18.49 (эксперимент Уэлдона). В одном из экспериментов с игральными костями Уэлдон подбросил кости 49 152 раз. При этом в 25 145 случаях выпали числа 4, 5 или 6.

Согласуется ли с этими данными гипотеза о симметричности костей?

18.50. При обследовании coat-colour (масти) 1000 пар чистокровных скаковых лошадей mother (матерей) и daughter (дочерей): black, brown, bay, chestnut, grey получены результаты, приведенные в таблице:

Coat-colour of daughter	Coat-colour of mother					Всего
	Black	Brown	Bay	Chestnut	Grey	
Black	7	8	11	11	5	42
Brown	7	40	75	20	9	151
Bay	13	95	230	101	42	481
Chestnut	6	23	113	82	17	241
Grey	5	7	18	16	39	85
Всего	38	173	447	230	112	1000

Свидетельствуют ли эти данные о связи между coat-colour лошади-матери и лошади-дочери?

18.51 (острота зрения). В таблице приведены данные об остроте зрения невооруженным глазом 3242 мужчин в возрасте 30—39 лет — служащих Королевских артиллерийских заводов Великобритании (1943 — 1946 гг.).

Степень остроты зрения (правый глаз)	Степень остроты зрения (левый глаз)				Всего
	Высшая	Вторая	Третья	Низшая	
Высшая	821	112	85	35	1053
Вторая	116	494	145	27	782
Третья	72	151	583	87	893
Низшая	43	34	106	331	514
Всего	1052	791	919	480	3242

Можно ли на основании этих данных сделать вывод о том, что острота зрения правого и левого глаз не связаны между собой?

18.52 (гипотеза Д’Аламбера). Стохастический эксперимент состоит в подбрасывании пары симметричных монет. Замечательный французский ученый и математик Д’Аламбер считал, что события A — “обе монеты выпали гербом”, B — “обе монеты выпали решеткой”, C — “монеты выпали разными сторонами”, равновероятны и, следовательно, вероятность каждого из них равна $1/3$. А впрочем, и до Д’Аламбера и после него был известен правильный подход к решению этой задачи: монеты необходимо различать. Для симметричных различимых монет каждому исходу ГГ, РР, ГР, РГ необходимо приписать вероятность $1/4$ (буква Г обозначает появление герба, буква Р — решетки). Тогда

$$P(A) = P(\text{ГГ}) = 1/4, \quad P(B) = P(\text{РР}) = 1/4,$$

$$P(C) = P(\{\Gamma P, P\Gamma\}) = P(\Gamma P) + P(P\Gamma) = 1/4 + 1/4 = 1/2.$$

Но, по-видимому, невозможно логически обосновать, почему не прав Д'Аламбер, считая, что события A, B, C равновероятны. В физике элементарных частиц встречаются ситуации, в которых Д'Аламбер скорее прав, чем неправ.

Рассмотрите две модели стохастического эксперимента, состоящего в подбрасывании двух монет.

Модель 1° (гипотеза Д'Аламбера).

$$\Omega^* = \{A, B, C\}, \quad P(A) = \frac{1}{3}, \quad P(B) = \frac{1}{3}, \quad P(C) = \frac{1}{3}$$

(события A, B, C определены выше).

Модель 2°.

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

$$P(\Gamma\Gamma) = \frac{1}{4}, \quad P(\Gamma P) = \frac{1}{4}, \quad P(P\Gamma) = \frac{1}{4}, \quad P(PP) = \frac{1}{4}.$$

С целью проверки адекватности описания этими моделями стохастического эксперимента, проведите его достаточное число раз и проверьте:

1. Адекватно ли описывается стохастический эксперимент моделью 1° (согласуется ли гипотеза Д'Аламбера с результатами эксперимента)?

2. Адекватно ли описывается стохастический эксперимент моделью 2°?

Заметим, что, вообще говоря, для описания одного и того же стохастического эксперимента можно предложить не одну модель, которая адекватно его описывает, но из предложенных выше моделей одна заведомо не будет адекватно описывать эксперимент, поскольку вероятности, например, события “монеты легли разными сторонами”, вычисленные в разных вероятностных пространствах, различны.

Для проверки адекватности модели 1° будем фиксировать число подбрасываний, в которых обе монеты легли гербом; обе — решеткой; монеты легли разными сторонами.

Для проверки адекватности модели 2° будем регистрировать число подбрасываний, в которых обе монеты легли гербом; обе — решеткой; первая монета легла гербом, а вторая — решеткой; первая монета легла решеткой, а вторая — гербом.

Результаты экспериментов удобно записывать в виде приведенных далее таблиц (первая таблица для проверки адекватности модели 1°, вторая — модели 2°), где, например, $N_{\text{гр}}$ — количество подбрасываний из N проведенных, в которых на первой монете выпал герб, на второй — решетка.

A	B	C
N_A	N_B	N_C

$\Gamma\Gamma$	PP	ΓP	$\text{P}\Gamma$
$N_{\Gamma\Gamma}$	N_{pp}	$N_{\Gamma\text{p}}$	$N_{\text{p}\Gamma}$

18.53 (эксперимент Ж. Бюффона). В результате $n = 4040$ подбрасываний монеты Ж. Бюффон зарегистрировал $\nu_1 = 2048$ случаев появления герба и $\nu_0 = 1992$ случаев появления решетки.

Согласуется ли с этими данными гипотеза: вероятность выпадения герба равна $1/2$?

18.54 (время ожидания). Рассматривается время ожидания четного числа в последовательности целых неотрицательных случайных чисел, меньших 100. Другими словами, рассматривается случайная величина, равная количеству нечетных чисел между двумя последовательными (соседними) четными числами, меньшими 100.

Предложить математическую модель (распределение) случайной величины — времени ожидания четного числа. Проверить адекватность модели стохастическому эксперименту. Воспользовавшись таблицей случайных чисел (табл. 22.10.1), получить выборку времени ожидания четного числа (достаточного объема) и проверить, согласуется ли предложенная модель с результатом эксперимента — полученной выборкой.

Результаты наблюдений удобно представить в виде:

[illegible]

где i — время ожидания четного числа; n_i — число наблюдений, в которых время ожидания равно i .

З а м е ч а н и е. О выборе модели стохастического эксперимента и проверке ее адекватности эксперименту см. задачу 18.1.

18.55 (катастрофы на угольных шахтах). В таблице приведены интервалы (в днях, читать по строкам) между катастрофами на угольных шахтах Великобритании с 1875 по 1951 г. (данные В. Мегью, Е. Пирсона и А. Винна). Maguire B. A., Pearson E. S., Wynn A. H. A. The time intervals between industrial accidents // Biometrika. 1952. Vol. 39. P. 168 – 180.

Проверить гипотезу о показательном распределении интервалов между катастрофами.

378	36	15	31	215	11	137	4
15	72	96	124	50	120	203	176
55	93	59	315	59	61	1	13
189	345	20	81	286	114	108	188
233	28	22	61	78	99	326	275
54	217	113	32	23	151	361	312
354	58	275	78	17	1205	644	467
871	48	123	457	498	49	131	182
255	195	224	566	390	72	228	271
208	517	1613	54	326	1312	348	745
217	120	275	20	66	291	4	369
338	336	19	329	330	312	171	145
75	364	37	19	156	47	129	1630
29	217	7	18	1357			

П р и м е ч а н и е. Катастрофой считается ситуация, влекущая за собой гибель 10 и более человек.

18.56 (предвзятость экспериментатора). Количество случаев появления последней цифры в результатах 1000 измерений, выполненных экспериментатором, приведено в таблице. Сомнительно, чтобы существовали объективные факторы, обуславливающие более частое появление одних цифр по сравнению с другими; поэтому вполне естественно предположить, что отклонение от одинаковой вероятности появления цифр свидетельствует о предвзятости экспериментатора.

Свидетельствуют ли приведенные данные о предвзятости экспериментатора?

П р и м е ч а н и е. О предвзятости при измерениях см. замечание к задаче 18.12.

k	n_k	k	n_k
0	158	5	71
1	97	6	90
2	125	7	56
3	73	8	125
4	76	9	129
		Всего	1 000

В таблице k — последняя цифра в результатах измерений; n_k — количество результатов измерений, в которых последней была цифра k , $k = 0, 1, \dots, 9$.

18.57. Стохастический эксперимент из известной задачи Льюиса Кэррола (см. пример 4.1.2) состоит в последовательном извлечении из урны двух шаров. Можно предложить по меньшей мере две модели (два вероятностных пространства) этого стохастического эксперимента (как и в примере 4.1.2, белый шар обозначим через W , черный — через B , белый шар, который положили в урну, пометим звездочкой и обозначим через W^*).

Модель 1°.

$$\Omega = \{WW, WB, BW\},$$

$$P(WW) = \frac{1}{3}, P(WB) = \frac{1}{3}, P(BW) = \frac{1}{3}.$$

Модель 2°.

$$\Omega = \{WW^*, W^*W, W^*B, BW^*\},$$

$$P(WW^*) = \frac{1}{4}, P(W^*W) = \frac{1}{4}, P(W^*B) = \frac{1}{4}, P(BW^*) = \frac{1}{4}.$$

Правдоподобные рассуждения (см. пример 4.1.2) склоняют нас к мысли, что адекватной моделью стохастического эксперимента является модель 2°. Но так ли это на самом деле?

Проведите стохастический эксперимент достаточное число раз и проверьте:

1. Адекватно ли описывается стохастический эксперимент моделью 1°?

2. Адекватно ли описывается стохастический эксперимент моделью 2°?

Заметим, что хотя для одного и того же эксперимента можно предложить не одну модель, которая адекватно его описывает, но для данного стохастического эксперимента по меньшей мере одна из моделей не будет адекватной эксперименту уже хотя бы потому, что вероятности одного и того же события, вычисленные в различных вероятностных пространствах, разные (см. пример 4.1.2).

Стохастический эксперимент будем проводить следующим образом. Возьмем две урны: вспомогательную и основную. Во вспомогательной находятся два шара: белый и черный. Из вспомогательной урны наудачу выберем один из шаров и переложим в основную, не регистрируя результат. Таким образом, в основной урне находится белый шар (с вероятностью $1/2$) или черный. Далее в основную урну кладем белый шар, его при проверке адекватности модели 2 пометим звездочкой. И наконец, из основной урны последовательно извлекаем оба шара, фиксируя результат эксперимента:

$$WW, WB, BW$$

— в случае проверки адекватности модели 1° и

$$WW^*, W^*W, W^*B, BW^*$$

— в случае проверки адекватности модели 2° .

Данные экспериментов удобно записать в виде таблиц (первая — для проверки адекватности модели 1° , вторая — модели 2°), где, например, N_{WW} — количество экспериментов из N проведенных, в которых исходом была пара WW .

WW	WB	BW
N_{WW}	N_{WB}	N_{BW}

WW^*	W^*W	W^*B	BW^*
N_{WW^*}	N_{W^*W}	N_{W^*B}	N_{BW^*}

18.58. Подбрасывают пару монет и игральную кость, при этом ξ — число выпавших гербов на паре монет, η — число выпавших очков на игральной кости, $\zeta = (\xi, \theta)$ — случайная величина со значениями в $\mathbb{R}^2$, где $\theta = I_{\{2,4,6\}}(\eta)$.

Проверить гипотезу о независимости компонент случайной величины $\zeta = (\xi, \theta)$. С этой целью подбросить монеты и игральную кость достаточное число раз и воспользоваться критерием χ^2 .

Результаты наблюдений удобно представить в виде таблицы:

Значения ξ	Значения θ		Сумма
	0	1	
0	ν_{00}	ν_{01}	$\nu_{0\cdot}$
1	ν_{10}	ν_{11}	$\nu_{1\cdot}$
2	ν_{20}	ν_{21}	$\nu_{2\cdot}$
Сумма	$\nu_{\cdot 0}$	$\nu_{\cdot 1}$	n

где ν_{ij} — количество подбрасываний, в которых случайная величина $\zeta = (\xi, \theta)$ приняла значение (i, j) , $i = 0, 1, 2$; $j = 0, 1$;

$$\nu_{i\cdot} = \sum_{j=0}^1 \nu_{ij}, \quad \nu_{\cdot j} = \sum_{i=0}^2 \nu_{ij}, \quad i = 0, 1, 2; \quad j = 0, 1.$$

З а м е ч а н и е. О выборе модели стохастического эксперимента и проверке ее адекватности см. задачу 18.1.

18.59 (телепатия и карты Зеннера). Начиная с двадцатых годов XX столетия в Гарвардском университете, университете Дюка и других университетах финансируются экспериментальные исследования по телепатии (“чтении” мыслей на расстоянии), в которых, в частности, используются так называемые карты Зеннера, на которых изображены пять символов: окружность, квадрат, плус, три волнистые линии, звезда (рис. 18.4.1).

С помощью карт Зеннера проверить способности телепата “читать” мысли на расстоянии. С этой целью провести следующий эксперимент. Из колоды карт Зеннера вы наудачу выбираете одну и фиксируете результат. Телепат читает выбранную вами карту и также фиксирует результат. Эксперимент проводится достаточное число раз.

Если телепат в самом деле читает мысли, то частота правильно прочитанных символов (окружность читается как окружность, квадрат — как квадрат, и т. д.), т. е. частота успехов будет большой (успех — правильно прочитанный символ, неудача — неправильно прочитанный символ).

По результатам проведенного эксперимента сделать вывод относительно способности телепата читать мысли на расстоянии.

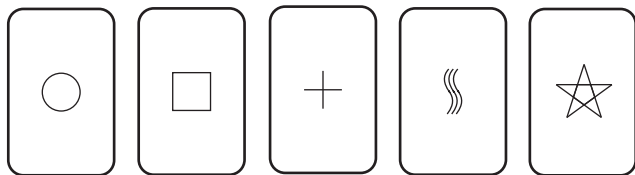


Рис. 18.4.1: Карты Зеннера

У к а з а н и е 1. Сформулировать задачу проверки способностей телепата читать мысли на расстоянии как задачу проверки статистических гипотез. В качестве нулевой рассмотреть гипотезу: телепат мысли не читает. Воспользоваться критерием χ^2 .

У к а з а н и е 2. Проверить телепатические способности своего товарища, знакомого (попросить прочесть ваши мысли).

18.60. Симметричную монету подбрасывают до первого появления герба и регистрируют количество μ решетонок, которые при этом выпали (количество шагов до первого появления герба).

Предложите математическую модель этого стохастического эксперимента. Проверьте ее адекватность эксперименту. Для этого проведите эксперимент достаточное число раз и воспользуйтесь критерием χ^2 .

18.61. Подбрасывают пару монет и регистрируют событие: “на обеих монетах выпал герб”.

Предложите математическую модель этого стохастического эксперимента. Проверьте ее адекватность эксперименту. Для этого проведите эксперимент достаточное число раз и воспользуйтесь критерием χ^2 .

18.62. Ниже приведены моменты прибытия пациентов на пункт скорой помощи (Оксфорд, данные А. Барроу) в феврале, марте, апреле, мае 1963 г.

4 февраля	11.00	17 марта	11.05	29 апреля	18.45
	17.00	20 марта	16.00	4 мая	16.30
8 февраля	23.15	22 марта	19.00	6 мая	22.00
11 февраля	10.00	24 марта	17.45	7 мая	8.45
16 февраля	12.00		20.20	11 мая	19.15
18 февраля	8.45		21.00	13 мая	15.30
	16.00	28 марта	12.00	14 мая	12.00
20 февраля	10.00		12.00		18.15
	15.30	30 марта	18.00	16 мая	14.00
21 февраля	20.20	2 апреля	22.00	18 мая	13.00
25 февраля	4.00		22.00	19 мая	23.00
	12.00	6 апреля	22.05	20 мая	19.05
28 февраля	2.20	9 апреля	12.45	22 мая	22.00
1 марта	12.00		19.30	23 мая	10.05
3 марта	5.30	10 апреля	18.45		12.30
7 марта	7.30	11 апреля	16.15	24 мая	18.15
	12.00	15 апреля	16.00	25 мая	21.05
9 марта	16.00	16 апреля	20.30	28 мая	21.00
15 марта	16.00	23 апреля	23.40	30 мая	0.30
16 марта	1.30	28 апреля	20.20		

Можно ли на основании этих данных заключить, что момент прибытия пациента распределен равномерно в течение суток? Если нет, предложите свои варианты распределений момента прибытия пациента.

18.63. Проверить гипотезу о показательном распределении длины интервала времени между прибытием пациентов на пункт скорой помощи (см. задачу 18.62).

Глава 19

Непараметрические критерии

19.1 Критерий Колмогорова

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \dots, \xi_n(\omega))$ — реализация выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из неизвестного непрерывного распределения F . Относительно F выдвигается гипотеза $H_0: F = G$ или, что то же, $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из G , где G — полностью определенное непрерывное распределение.

Необходимо проверить гипотезу H_0 , т. е. по реализации выборки $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ сделать вывод: отклонять гипотезу H_0 или не отклонять.

Статистика Колмогорова. Для проверки гипотезы $H_0: F = G$ естественно ввести отклонение $D(\hat{F}_n, G)$ эмпирического распределения $\hat{F}_n$, построенного по выборке $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, от гипотетического G (см. гл. 18). А. Н. Колмогоров в качестве отклонения эмпирического распределения $\hat{F}_n$, построенного по выборке $\xi = (\xi_1, \xi_2, \dots, \dots, \xi_n)$, от гипотетического G предложил рассматривать

$$D(\hat{F}_n, G) = \sup_x |G(x) - \hat{F}_n(x)|.$$

Так введенное отклонение при верной гипотезе H_0 , т. е. ко-

гда $F = G$, принимает значения

$$D_n = D(\hat{F}_n, F) = \sup_x \left| F(x) - \hat{F}_n(x) \right|$$

и является малым по сравнению с уклонением $D(\hat{F}_n, G)$, когда распределение G отлично от F (более того, $D_n = D(\hat{F}_n, F)$ — минимально возможное уклонение). Последнее вполне естественно, поскольку для каждого фиксированного x эмпирическая функция распределения $\hat{F}_n(x)$ является несмещенной и состоятельной оценкой $F(x)$. К тому же распределение минимально возможного уклонения D_n эмпирического распределения от гипотетического, во-первых, не зависит от F (одно и то же для всех непрерывных F) и, во-вторых, при больших n в качестве распределения нормированного минимально возможного уклонения

$$D_n / \frac{1}{\sqrt{n}} = \sup_x \left| F(x) - \hat{F}_n(x) \right| / \frac{1}{\sqrt{n}}$$

можно рассматривать распределение Колмогорова, его функция распределения

$$K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp\{-2k^2\lambda^2\}.$$

Последнее следует из теоремы Колмогорова:

Теорема. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — выборка из непрерывного распределения F , $\hat{F}_n(x)$ — эмпирическая функция распределения, построенная по выборке $\xi_1, \xi_2, \dots, \xi_n$, тогда

$$\lim_n P \left\{ \sup_x \left| F(x) - \hat{F}_n(x) \right| / \frac{1}{\sqrt{n}} < \lambda \right\} = K(\lambda), \quad \lambda > 0.$$

Так что при верной гипотезе $H_0: F = G$ уклонение

$$D(\hat{F}_n, G) = \sup_x \left| G(x) - \hat{F}_n(x) \right| = \sup_x \left| F(x) - \hat{F}_n(x) \right|$$

малое, в противном случае — большое. Поэтому, проверяя гипотезу H_0 , ее естественно отклонять, если $D(\hat{F}_n, G)$ приняло большое значение, и не отклонять в противном случае.

Границы $\varepsilon_{\alpha;n}$, отделяющие большие значения отклонения $D(\hat{F}_n, G)$ от малых, находят по известному распределению минимально возможного отклонения $D_n = D(\hat{F}_n, F)$ эмпирического распределения от гипотетического. А именно, $\varepsilon_{\alpha;n}$ находится как верхний α -предел (критическое значение) распределения отклонения

$$D(\hat{F}_n, F) = \sup_x |F(x) - \hat{F}_n(x)|,$$

т. е. как наименьшее ε , для которого

$$P \left\{ \sup_x |F(x) - \hat{F}_n(x)| \geq \varepsilon \right\} \leq \alpha.$$

Значения $\varepsilon_{\alpha;n}$ для заданных α (уровня значимости) и n (объема выборки) приведены в табл. 22.7.1. При больших n ($n \geq 100$), в качестве $\varepsilon_{\alpha;n}$ рассматривается $\lambda_\alpha/\sqrt{n}$, т. е.

$$\varepsilon_{\alpha;n} = \frac{\lambda_\alpha}{\sqrt{n}},$$

где λ_α — верхний α -предел распределения Колмогорова, т. е. λ_α — корень уравнения $K([\lambda_\alpha, +\infty)) = \alpha$.

Критерий Колмогорова. Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ — выборка из непрерывного распределения F . Если гипотезу H_0 : $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ является выборкой из распределения G отклонять при

$$D(\hat{F}_n, G) = \sup_x |G(x) - \hat{F}_n(x)| \geq \varepsilon_{\alpha;n}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α , гипотезу H_0 будем отклонять, когда она верна.

З а м е ч а н и е. Критерием Колмогорова можно пользоваться только тогда, когда гипотетическое распределение G полностью определено и непрерывно.

Пример 19.1.1 Воспользовавшись таблицей случайных чисел (табл. 22.10.1), получить выборку объемом 10 из стандартного нормального распределения (нормального распределения с параметрами $(0; 1)$).

С помощью критерия Колмогорова проверить, действительно ли эта выборка получена из указанного распределения.

Решение. Если η — случайная величина, равномерно распределенная на промежутке $[0; 1]$ и $F(x)$ — возрастающая непрерывная функция распределения, то случайная величина ξ , определенная равенством

$$\xi = F^{-1}(\eta),$$

имеет своей функцией распределения $F(x)$. В самом деле,

$$\begin{aligned} F_{\xi}(x) &= P\{\xi < x\} = P\{F^{-1}(\eta) < x\} = \\ &= P\{\eta < F(x)\} = F_{\eta}(F(x)) = F(x). \end{aligned}$$

Это утверждение дает возможность по выборке $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ из равномерного на промежутке $[0; 1]$ распределения строить выборку $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ из данного распределения F , а именно:

$$\xi_i = F^{-1}(\eta_i), \quad i = 1, 2, \dots, n,$$

— выборка из распределения F . В частности, если в качестве $F(x)$ рассмотреть функцию стандартного нормального распределения

$$N_{0;1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{s^2}{2}\right\} ds,$$

то

$$\xi_i = N_{0;1}^{-1}(\eta_i), \quad i = 1, 2, \dots, n,$$

будет выборкой из $N_{0;1}$.

Выборку из равномерного на промежутке $[0; 1]$ распределения можно получить, воспользовавшись таблицей случайных чисел (табл. 22.10.1) (см. также задачу 19.4).

Из табл. 22.10.1 выберем 10 чисел (для определенности — четырехзначных). Выбор можно начинать с любого места таблицы (скажем, с верхнего правого угла) и продолжать любым оговоренным наперед способом. Например, двигаясь по диагонали, получим

1009	5420	2689	2529	7080
3407	5718	1656	7048	7835

Числа

0,1009	0,5420	0,2689	0,2529	0,7080
0,3407	0,5718	0,1656	0,7048	0,7835

из промежутка $[0;1]$ можно рассматривать как реализацию выборки $\eta_1, \eta_2, \dots, \eta_{10}$ из равномерного на промежутке $[0;1]$ распределения. По значениям η_i , $i = 1, 2, \dots, \dots, 10$, пользуясь тем, что функция $N_{0;1}(x)$ табулирована (табл. 22.1.1), получим реализацию выборки из стандартного нормального распределения как значения

$$\xi_i = N_{0;1}^{-1}(\eta_i) :$$

-1,27	0,11	-0,62	-0,67	0,55
-0,41	0,18	-0,97	0,54	0,78

Последний знак получен методом линейной интерполяции.

Далее, пользуясь критерием Колмогорова, проверим гипотезу H_0 : выборка получена из распределения $N_{0;1}$. Для этого вычислим

$$D(\hat{F}_n, G) = \sup_x \left| G(x) - \hat{F}_n(x) \right| ,$$

где $G(x) = N_{0;1}(x)$ — функция нормального распределения с параметрами $(0;1)$; $\hat{F}_n(x)$ — реализация эмпирической функции распределения, построенная по выборке, и сравним его с критическим значением $\varepsilon_{\alpha;n}$. Значение

$$D(\hat{F}_n, N_{0;1}) = \sup_x \left| N_{0;1}(x) - \hat{F}_n(x) \right| = 0,21$$

(методика вычисления $\sup_x |G(x) - \hat{F}_n(x)|$ описана в гл. 15); критическое значение $\varepsilon_{0,05;10} = 0,4087$ найдено по табл. 22.7.1);

$$D(\hat{F}_n, N_{0;1}) = 0,21 < 0,4087 = \varepsilon_{0,05;10} = \varepsilon_{\alpha;n}$$

поэтому согласно критерию Колмогорова гипотеза: выборка получена из нормального распределения с параметрами $(0;1)$ не отклоняется.

Для выборки объемом 10 из нормального распределения с параметрами $(0;1)$ максимальное отклонение между эмпирической функцией распределения $\hat{F}_{10}(x)$ и функцией распределения $N_{0;1}(x)$, равное 0,21, не является большим, такое отклонение естественно и допустимо.

19.2 Критерий знаков

Постановка задачи (наблюдения повторные).

Имеем $(\xi_1(\omega), \eta_1(\omega)), (\xi_2(\omega), \eta_2(\omega)), \dots, (\xi_n(\omega), \eta_n(\omega))$ — результат наблюдения n пар случайных величин $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$, относительно которых известно, что разности $\zeta_j = \eta_j - \xi_j$ представимы в виде

$$\zeta_j = \theta + e_j, \quad j = 1, 2, \dots, n,$$

где θ — константа, а случайные величины $e_1, e_2, \dots, e_n$:
1° независимыми (сами η_j и ξ_j могут быть зависимыми);
2° симметрично распределены относительно нуля (распределения e_j и $-e_j$ совпадают) и абсолютно непрерывны (далее это ограничение будет снято).

Относительно неизвестного параметра θ выдвигается гипотеза

$$H_0: \theta = 0.$$

Альтернатива к гипотезе $H_0: \theta = 0$ может быть как односторонней — если $\theta \neq 0$, то $\theta > 0$ (она может быть и такой: $\theta < 0$), так и двусторонней — если $\theta \neq 0$, то $\theta > 0$ или $\theta < 0$. В каждой задаче альтернатива своя. Необходимо построить критерий для проверки гипотезы H_0 .

З а м е ч а н и е. Пары $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$ можно интерпретировать как $2n$ наблюдений — по два наблюдения на каждые n объектов, пациентов, приборов, и т. д., при этом ξ_j называют наблюдениями до обработки, а η_j — наблюдениями после обработки, $j = 1, 2, \dots, n$. Параметр θ , называют эффектом обработки. Отклонение гипотезы H_0 свидетельствует в пользу наличия эффекта обработки, неотклонение — в пользу отсутствия.

Статистика для построения критерия. Верна гипотеза $H_0: \theta = 0$ или нет, случайные величины e_j , $j = 1, 2, \dots, n$, абсолютно непрерывно, симметрично распределены относительно нуля и независимы. Отсюда следует, что случайная величина, равная количеству e_j , $j = 1, 2, \dots, n$, которые приняли положительные значения, имеет биномиальное распределение с параметрами $(n; 1/2)$ и поэтому количество положительных величин среди e_j , $j = 1, 2, \dots, n$, близко к половине имеющихся, т. е. к $n/2$.

Обозначим через μ количество положительных разностей среди

$$\zeta_j = \eta_j - \xi_j = \theta + e_j, \quad j = 1, 2, \dots, n. \quad (19.2.1)$$

Если гипотеза H_0 верна, т. е. $\theta = 0$, то количество положительных разностей среди (19.2.1) совпадает с количеством случайных величин e_j , $j = 1, 2, \dots, n$, принявших положительные значения, а следовательно, мало отличается от $n/2$ — половины имеющихся разностей. Если же гипотеза $H_0: \theta = 0$ неверна, то количество положительных разностей среди ζ_j , $j = 1, 2, \dots, n$, существенно отличается от $n/2$ — будет существенно большим или меньшим $n/2$ (в зависимости от знака θ).

Таким образом, если гипотеза H_0 верна, количество μ положительных разностей среди ζ_j , $j = 1, 2, \dots, n$ (обозначим его через μ_0) мало отличается от $n/2$ по сравнению с отклонением μ от $n/2$, когда гипотеза H_0 неверна. Поэтому, проверяя гипотезу $H_0: \theta = 0$, ее естественно отклонять, если количество положительных разностей μ существенно отличается от $n/2$, и не отклонять в противном случае.

Границы $m_{\alpha;n}$, отделяющие большие значения отклонений μ от $n/2$ от малых, строятся по известному распре-

делению количества μ_0 положительных разностей

$$\zeta_j = \eta_j - \xi_j = e_j, \quad j = 1, 2, \dots, n,$$

которое минимально отклоняется от $n/2$. А именно, $m_{\alpha;n}$ находится как минимальное число m , для которого

$$P\{\mu_0 > m\} \leq \alpha, \quad (19.2.2)$$

где μ_0 — биномиально распределенная случайная величина с параметрами $(n; 1/2)$.

Значения $m_{\alpha;n}$ по данному уровню значимости α и объему выборки n табулированы (см. табл. 22.9.1).

Критерий знаков. Пусть $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$ — n пар случайных величин, для которых разности $\zeta_j = \eta_j - \xi_j$ представимы в виде

$$\zeta_j = \theta + e_j, \quad j = 1, 2, \dots, n,$$

где случайные величины e_j : 1° независимы, 2° симметрично распределены относительно нуля и абсолютно непрерывны; число $m_{\alpha;n}$ определяется из (19.2.2).

Если гипотезу $H_0: \theta = 0$ отклонять при

$$\mu > m_{\alpha;n}$$

и не отклонять при

$$\mu \leq m_{\alpha;n},$$

то с вероятностью, не превосходящей α , гипотезу H_0 будем отклонять, когда она верна (односторонний критерий знаков, альтернатива: $\theta > 0$).

Если гипотезу H_0 отклонять при

$$\mu \notin [n - m_{\alpha;n}; m_{\alpha;n}]$$

и не отклонять при

$$\mu \in [n - m_{\alpha;n}; m_{\alpha;n}],$$

то с вероятностью, не превосходящей 2α , гипотезу H_0 будем отклонять, когда она верна (двусторонний критерий знаков, альтернатива: $\theta > 0$ или $\theta < 0$).

Ошибки первого и второго рода. При верной гипотезе $H_0: \theta = 0$ случайная величина μ — количество положительных разностей $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \dots, n$, имеет биномиальное распределение с параметрами $(n; 1/2)$. Поэтому μ , “почти всегда” мало уклоняясь от $n/2$, может (хотя и изредка) принимать значения, которые существенно отличаются от $n/2$. При этом гипотезу $H_0: \theta = 0$ отклоняем и тем самым допускаем ошибку первого рода (ее вероятность не превышает выбранного уровня значимости).

Если гипотеза H_0 неверна, например, $\theta > 0$, то, несмотря на то, что случайная величина μ (количество положительных разностей) почти всегда принимает большие значения (существенно большие, чем $n/2$), она может, хотя и изредка, принимать значения, которые мало отличаются от $n/2$. При этом мы гипотезу H_0 не отклоняем и тем самым допускаем ошибку второго рода.

Связи. Если снять требование об абсолютной непрерывности распределений случайных величин η_j и ξ_j , то разности $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \dots, n$, могут принимать нулевые значения с ненулевой вероятностью (говорят, что имеются связи). Можно ли в этом случае пользоваться критерием знаков для проверки гипотезы $H_0: \theta = 0$? Оказывается — да, причем в той же формулировке, что и раньше, отбросив равные нулю разности и применяя критерий знаков к оставшимся отличным от нуля разностям (разности, равные нулю, не учитываются, как будто их вообще не было).

Критерий знаков при наличии связей. Пусть $(\xi_1, \eta_1), (\xi_2, \eta_2), \dots, (\xi_n, \eta_n)$ — n пар случайных величин, для которых разности $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \dots, n$, представимы в виде

$$\zeta_j = \theta + e_j, \quad j = 1, 2, \dots, n,$$

где случайные величины e_j : 1° независимы; 2° симметрично распределены относительно нуля.

Обозначим через s количество разностей $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \dots, n$, отличных от нуля, а через μ — количество положительных среди них. Для каждого фиксированного s ($s = 1, 2, \dots, n$) определим число $m_{\alpha; s}$ как минималь-

ное m , для которого

$$P\{\mu > m\} \leq \alpha,$$

где μ — биномиально распределенная случайная величина с параметрами $(s; 1/2)$.

Если гипотезу H_0 отклонять при

$$\mu > m_{\alpha;s},$$

и не отклонять при

$$\mu \leq m_{\alpha;s},$$

то с вероятностью, не превышающей α , гипотезу H_0 будем отклонять, когда она верна (альтернатива односторонняя: $\theta > 0$).

Если гипотезу H_0 отклонять при

$$\mu \notin [s - m_{\alpha;s}, m_{\alpha;s}]$$

и не отклонять при

$$\mu \in [s - m_{\alpha;s}, m_{\alpha;s}],$$

то с вероятностью, не превышающей 2α , гипотезу H_0 будем отклонять, когда она верна (альтернатива двусторонняя: $\theta > 0$ или $\theta < 0$).

Двухвыборочный критерий знаков. Пусть $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_n$ — независимые выборки соответственно из распределений F и G . Распределения F и G связаны соотношением $G(x) = F(x - \theta)$. Параметр θ (эффект обработки) неизвестен. Необходимо выяснить, отлично значение θ от нуля или нет. Будем решать поставленную задачу, проверяя гипотезу $H_0: \theta = 0$ (нулевую гипотезу можно сформулировать и так $H_0: F = G$). Альтернатива к гипотезе $H_0: \theta = 0$ может быть как односторонней, так и двусторонней.

Гипотезу $H_0: \theta = 0$ можно проверять, пользуясь критерием знаков в той же формулировке, что и для парных наблюдений: подсчитать число μ положительных разностей $\zeta_j = \eta_j - \xi_j$ среди $\zeta_1, \zeta_2, \dots, \zeta_n$ и сравнивать его с $m_{\alpha;s}$.

Пример 19.2.1 (смещенность результатов измерений у контролеров). Один из методов количественного анализа степени износа шины состоит в измерении глубины проникновения щупа¹ в канавку протектора в определенном месте шины.

В рамках дорожно-эксплуатационных исследований два контролера измеряют глубину канавок на шинах после каждого эксперимента. Основная сложность проведения измерений состоит в том, что каждый контролер имеет свое, присущее только ему смещение результатов измерений, связанное с силой давления на щуп (от силы давления зависит глубина проникновения щупа в канавку протектора).

Результаты 10 измерений, выполненных контролерами в 10 фиксированных точках шины, приведены в таблице.

Точка	Первый контролер	Второй контролер	Точка	Первый контролер	Второй контролер
1	126	125	6	159	152
2	128	120	7	152	150
3	157	163	8	138	136
4	131	118	9	138	140
5	142	129	10	142	136

Имеются основания предполагать, что первый контролер получает более высокие результаты, чем второй.

Согласуется ли это предположение с результатами эксперимента?

Решение. Результаты измерений, выполненных первым контролером, обозначим через η_j , а вторым — через ξ_j , $j = 1, 2, \dots, 10$. Разности $\zeta_j = \eta_j - \xi_j$, $j = 1, 2, \dots, 10$, естественно считать представимыми в виде

$$\zeta_j = \theta + e_j, \quad j = 1, 2, \dots, 10,$$

где e_j — независимые случайные величины. Параметр θ характеризует различия (если они имеются) в силе давления на щуп первого и второго контролеров (если $\theta = 0$, то отличий нет, а если $\theta \neq 0$, они имеются).

¹Здесь щуп — тонкая продолговатая металлическая пластинка прямоугольной формы.

Необходимо по результатам измерений сделать вывод о значении параметра θ , а именно: $\theta = 0$ или $\theta \neq 0$. Сформулируем поставленную задачу в терминах проверки статистических гипотез.

Относительно параметра θ выдвигаем гипотезу $H_0: \theta = 0$, т. е. контролеры получают одинаковые результаты. Поскольку имеется подозрение, что первый контролер получает более высокие результаты, в качестве альтернативы к основной гипотезе выберем одностороннюю альтернативу: $\theta > 0$.

Отклонение гипотезы будем трактовать в пользу более высоких результатов первого контролера по сравнению с результатами второго, неотклонение — как отсутствие различий в измерениях контролеров.

Для проверки гипотезы H_0 воспользуемся критерием знаков. Имеем такие знаки разностей:

$$+ + - + + + + - + .$$

Всего 10 разностей (все они отличны от нуля), количество μ положительных среди них равно 8. При этом

$$\mu = 8 \leq 8 = m_{0,025;10} = m_{\alpha;n}.$$

Поэтому согласно критерию знаков для проверки гипотезы $H_0: \theta = 0$ против альтернативы $\theta > 0$ гипотеза H_0 на уровне значимости 0,025 не отклоняется.

Этот результат можно трактовать так. Эксперимент не дает оснований утверждать, что результаты измерений (см. таблицу), полученные первым контролером, существенно выше по сравнению с результатами, полученными вторым контролером. (Такие результаты вполне могли быть получены одним и тем же контролером.)

19.3 Критерий Вилкоксона

Постановка задачи. Пусть $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ и $\eta(\omega) = (\eta_1(\omega), \eta_2(\omega), \dots, \eta_m(\omega))$ — реализации независимых выборок $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ соответственно из непрерывных распределений F

и G . Относительно распределений F и G известно лишь то, что их функции распределений связаны соотношением

$$G(x) = F(x - \theta).$$

Параметр θ неизвестный. Относительно него выдвигается гипотеза

$$H_0: \theta = 0.$$

Альтернатива к нулевой гипотезе H_0 может быть как двусторонней: если $\theta \neq 0$, то $\theta > 0$ или $\theta < 0$, так и односторонней: если $\theta \neq 0$, то $\theta > 0$ (односторонняя альтернатива может быть и такой: $\theta < 0$). Выбор альтернативы определяется решаемой задачей.

Далее рассматривается критерий Вилкоксона для проверки гипотезы $H_0: \theta = 0$.

Статистика W (статистика Вилкоксона). Расположим выборочные значения $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ в общий вариационный ряд. Индексы, как правило, будем опускать. Получим перестановку из n букв ξ и m букв η , например

$$\omega = (\xi\eta\xi\eta\eta\xi\xi\dots\xi\eta). \quad (19.3.1)$$

Далее для определенности через $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ будем обозначать выборку меньшего объема.

Определим ранг каждого выборочного значения как номер места, на котором оно находится в общем вариационном ряду (19.3.1). Поскольку $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ — выборки из непрерывных распределений, вероятность того, что выборочные значения совпадут, равна нулю, но на практике выборочные значения регистрируются с конечным числом знаков и поэтому они могут совпадать с ненулевой вероятностью. В этом случае совпадающим выборочным значениям мы приписываем один и тот же средний ранг. Например, если $\xi_i = 7,1$; $\eta_j = 7,1$, причем имеются четыре выборочные значения, меньших 7,1, то каждому из выборочных значений ξ_i и η_j , равных 7,1 (они находятся на 5-м и 6-м местах) приписывается средний ранг $(5+6)/2 = 5,5$.

Далее определим W как сумму рангов выборки меньшего объема (т. е. выборки $\xi_1, \xi_2, \dots, \xi_n$), а именно:

$$W = r_1 + r_2 + \dots + r_n,$$

где $r_1, r_2, \dots, r_n$ — ранги выборочных значений $\xi_1, \xi_2, \dots, \dots, \xi_n$ в перестановке (19.3.1).

Величиной W можно описывать меру перемешанности букв ξ и η в общем вариационном ряду: если сумма рангов W большая (большинство выборочных значений ξ_i располагается справа от выборочных значений η_j) или малая (большинство выборочных значений ξ_i располагаются слева от выборочных значений η_j), то буквы ξ и η в общем вариационном ряду перемешаны плохо, в противном случае — хорошо.

Заметим, что минимально возможное значение суммы рангов W равно

$$1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2},$$

максимально возможное —

$$(m+1) + (m+2) + \dots + (m+n) = mn + \frac{n(n+1)}{2},$$

среднее —

$$\frac{1}{2} \left(\frac{n(n+1)}{2} + \left(mn + \frac{n(n+1)}{2} \right) \right) = \frac{n(n+m+1)}{2}.$$

Математическое ожидание W , когда гипотеза $H_0: \theta = 0$ верна, равно $n(n+m+1)/2$ (заметим, что $n(n+m+1)/2$ — координата середины отрезка с концами $n(n+1)/2$ и $nm + n(n+1)/2$).

Величину W называют *статистикой Вилкоксона*. На ее основе строится критерий для проверки гипотезы $H_0: F = G$.

Если гипотеза $H_0: F = G$ верна, то $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ и $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ являются независимыми выборками из одного и того же распределения; поэтому буквы ξ и η

в перестановке (19.3.1) перемешаны хорошо и W принимает значения, близкие к своему среднему значению

$$n(n + m + 1)/2$$

(статистику W , когда гипотеза верна, будем обозначать W_0). Если же гипотеза H_0 неверна, W принимает значения, которые существенно отличаются от

$$n(n + m + 1)/2$$

(при $\theta > 0$ статистика W принимает малые значения, а при $\theta < 0$ — большие). Следовательно, если гипотеза H_0 верна, случайная величина $W = W_0$ мало уклоняется от $n(n + m + 1)/2$ по сравнению с уклонением величины W от $n(n + m + 1)/2$, когда гипотеза H_0 неверна. Поэтому, проверяя гипотезу H_0 , ее естественно не отклонять, когда W приняло значение, мало уклоняющееся от $n(n + m + 1)/2$, и отклонять в противном случае. Границы $W_{\alpha;n;m}$ и $n(n + m + 1) - W_{\alpha;n;m}$, отделяющие значения W , мало уклоняющиеся от $n(n + m + 1)/2$, от значений W , отличающихся от $n(n + m + 1)/2$ существенно, строятся по известному распределению случайной величины W_0 .

По данным α (уровню значимости) и n, m (объемам выборок) число $W_{\alpha;n;m}$ находим как наибольшее целое t , для которого

$$P\{W \leq t\} \leq \alpha.$$

Значения $W_{\alpha;n;m}$ табулированы (см. табл. 22.8.1).

Критерий Вилкоксона. Пусть $\xi_1, \xi_2, \dots, \xi_n$ и $\eta_1, \eta_2, \dots, \eta_m$ — независимые выборки соответственно из непрерывных распределений F и G , причем $G(x) = F(x - \theta)$.

Если гипотезу

$$H_0: \theta = 0$$

отклонять при

$$W \leq W_{\alpha;n;m}$$

и не отклонять при

$$W > W_{\alpha;n;m},$$

то с вероятностью, не большей α , гипотезу H_0 будем отклонять, когда она верна (альтернатива односторонняя : $\theta > 0$).

Если гипотезу H_0 отклонять при

$$W \geq n(n + t + 1) - W_{\alpha;n;t}$$

и не отклонять в противном случае, то с вероятностью, не превосходящей α , гипотезу H_0 будем отклонять, когда она верна (альтернатива односторонняя : $\theta < 0$).

Если гипотезу H_0 отклонять при

$$W \notin (W_{\alpha;n;t}; n(n + t + 1) - W_{\alpha;n;t})$$

и не отклонять в противном случае, то с вероятностью, не большей чем 2α , гипотезу H_0 будем отклонять, когда она верна (альтернатива двусторонняя : $\theta < 0$ или $\theta > 0$).

Если объемы выборок n и t большие (в критерии Вилкоксона n и t большие, когда $\min\{n, t\} \geq 6$, $n+t \geq 20$), то в качестве $W_{\alpha;n;t}$ используется

$$W_{\alpha;n;t} = \frac{1}{2}n(n + t + 1) + z_\alpha \sqrt{\frac{1}{12}nt(n + t + 1)},$$

где z_α — α -квантиль нормального распределения с параметрами $(0;1)$, т. е. решение уравнения $N_{0;1}(z_\alpha) = \alpha$ (см. табл. 22.1.1). Это следует из того, что при $n, t \rightarrow \infty$, статистика W асимптотически нормальна со средним $n(n + t + 1)/2$ и дисперсией $nt(n + t + 1)/12$.

Ошибки первого и второго рода. При верной гипотезе H_0 : $F = G$ случайная величина W — сумма рангов выборки меньшего объема, принимая значения из промежутка

$$[n(n + 1)/2, nt + n(n + 1)/2]$$

и почти всегда мало уклоняясь от среднего значения $n(n + t + 1)/2$, т. е.

$$W \in (W_{\alpha;n;t}; n(n + t + 1) - W_{\alpha;n;t}),$$

может, хотя и изредка, принять значение, которое существенно отличается от $n(n + m + 1)/2$, т. е.

$$W \leq W_{\alpha;n;m} \quad \text{или} \quad W \geq n(n + m + 1) - W_{\alpha;n;m}.$$

При этом мы гипотезу H_0 отклоняем и тем самым допускаем ошибку первого рода.

Если же гипотеза H_0 неверна, то W , почти всегда принимаемая значения, существенно отличающиеся от среднего $n(n + m + 1)/2$ (малые или большие) может, хотя и изредка, принять значение, близкое к $n(n + m + 1)/2$, т. е.

$$W \in (W_{\alpha;n;m}; n(n + m + 1) - W_{\alpha;n;m}).$$

При этом мы гипотезу H_0 не отклоняем и тем самым допускаем ошибку второго рода.

Пример 19.3.1. Для исследования устойчивости к истиранию эпоксидной пластмассы с использованием как наполнителя окиси алюминия в разной концентрации изготовили две партии плиток.

Рассматривались два значения относительной концентрации наполнителя в смоле, выражаемые отношениями $1/2$ к 1 и 1 к 1 . Изготовили по 18 образцов плиток, при этом две из них с концентрацией $1/2$ к 1 были забракованы и не испытывались. Затем каждую из плиток приводили в возвратно-поступательное движение по абразивному материалу (по 10 000 циклов для плитки).

Измерялась разность толщины плиток до и после испытаний; точность измерений составляла 10^{-4} дюйма ($1 \text{ дюйм} = 0,0254 \text{ м.}$).

Концентрация $1/2$ к 1 : 7,0; 6,4; 7,3; 5,1; 5,7; 6,6; 5,0; 7,7; 6,8; 5,1; 4,6; 5,5; 5,8; 6,2; 4,8; 5,8.

Концентрация 1 к 1 : 7,1; 5,0; 6,4; 6,9; 5,7; 6,5; 6,5; 4,0; 6,2; 6,8; 4,0; 7,5; 7,2; 5,2; 7,8; 4,8; 4,4; 6,0.

Различаются ли плитки из эпоксидной пластмассы (по отношению к истиранию), в которых наполнителем была окись алюминия в разных концентрациях? Другими словами, влияет ли концентрация наполнителя на устойчивость плитки к истиранию?

Решение. Сформулируем поставленную задачу как задачу проверки статистических гипотез.

Обозначим через F и G распределения разности толщины плиток (до и после испытаний) соответственно с концентрацией наполнителя 1/2 к 1 и 1 к 1. Распределения F и G непрерывны. Предположим, что $G(x) = F(x - \theta)$ (θ — параметр, характеризующий различия в истирании, если они имеются).

Относительно различий плиток с разной концентрацией наполнителя выдвигается гипотеза H_0 : концентрация наполнителя не влияет на устойчивость плитки к истиранию. Эту гипотезу можно сформулировать так: H_0 : $F = G$ или, что то же, H_0 : $\theta = 0$. Поскольку относительно влияния концентрации наполнителя на истирание ничего не известно, рассматриваем двустороннюю альтернативу: $\theta < 0$ или $\theta > 0$.

Необходимо проверить гипотезу H_0 , т. е. по реализации выборки сделать вывод — отклонять гипотезу H_0 или не отклонять. Отклонение гипотезы H_0 будем интерпретировать как наличие влияния концентрации наполнителя на истирание плиток. Если же гипотеза H_0 не отклоняется, то можно утверждать, что в эксперименте влияние концентрации наполнителя на истирание не обнаружено.

Для проверки гипотезы H_0 воспользуемся критерием Вилкоксона. Гипотезу H_0 отклоняем, при

$$W \notin (W_{\alpha;n;m}; n(n+m+1) - W_{\alpha;n;m}),$$

и не отклоняем в противном случае.

Сначала вычислим W — сумму рангов выборки меньшего объема в общем вариационном ряду: 4,0; 4,0; 4,4; 4,6; 4,8; 4,8; 5,0; 5,0; 5,1; 5,1; 5,2; 5,5; 5,7; 5,7; 5,8; 5,8; 6,0; 6,2; 6,2; 6,4; 6,4; 6,5; 6,5; 6,6; 6,8; 6,8; 6,9; 7,0; 7,1; 7,2; 7,3; 7,5; 7,7; 7,8 (подчеркнуты выборочные значения выборки меньшего объема).

$$W = 4 + (5 + 6)/2 + (7 + 8)/2 + (9 + 10)/2 + (9 + 10)/2 + 12 + (13 + 14)/2 + (15 + 16)/2 + (15 + 16)/2 + (18 + 19)/2 + (20 + 21)/2 + 24 + (25 + 26)/2 + 28 + 31 + 33 = 273.$$

Далее, учитывая, что объемы выборок большие, а именно:

$$n = 16 > 6, \quad m = 18 > 6, \quad n + m = 34 > 20,$$

в качестве $W_{\alpha;n;m}$ можно рассмотреть

$$\frac{1}{2}n(n+m+1) + z_{\alpha}\sqrt{\frac{1}{12}nm(n+m+1)},$$

где z_{α} — α -квантиль нормального распределения с параметрами $(0;1)$.

По таблице нормального распределения находим $z_{\alpha} = z_{0,025} = -1,96$ (см. табл. 22.1.1). Таким образом

$$W_{\alpha;n;m} = W_{0,025;16;18} = \frac{1}{2}16(16+18+1) -$$

$$-1,96\sqrt{\frac{1}{12}16 \cdot 18(16+18+1)} = 223,2;$$

$$n(n+m+1) - W_{\alpha;n;m} =$$

$$= 16(16+18+1) - W_{0,025;16;18} = 336,8.$$

Для суммы рангов $W = 273$ имеем

$$W_{\alpha;n;m} = W_{0,025;16;18} = 223,2 < 273 < 336,8 =$$

$$= 16 \cdot 35 - W_{0,025;16;18} = n(n+m+1) - W_{\alpha;n;m}.$$

Поэтому согласно критерию Вилкоксона гипотеза $H_0: \theta = 0$ не отклоняется. Она не противоречит результатам эксперимента. (Сумма рангов $W = 273$ естественна (типична) для выборок объема $n = 16$ и $m = 18$ из одного и того же распределения.)

Следовательно, экспериментальные данные не дают оснований утверждать, что плитки с разной концентрацией наполнителя отличаются устойчивостью к истиранию.

19.4 Задачи

АЗ: 19.7, 19.11, 19.12.

СЗ: 19.1, 19.13, 19.30.

19.1 (анкетирование “Преподаватель глазами студентов” — эффект экзаменационной оценки). Вопрос о непредвзятости оценивания тех или иных качеств, величин, параметров весьма интересен. При этом предвзятость в оценивании встречается заметно чаще, чем непредвзятость.

В некоторых ситуациях причины предвзятости можно объяснить, в других они абсолютно загадочны и непонятны (см., например, задачу 18.17 из гл. 18, где приведен пример сильного пристрастия при считывании цифр с вращающегося с большой скоростью круга).

При обработке результатов анкетирования “Преподаватель глазами студентов”, как и при любом другом оценивании, возникают вопросы: 1) о непредвзятости оценивания (которую едва ли следует ожидать) и 2) о причинах предвзятости (в этом оценивании вполне понятных).

Мы, исследуя возможные причины предвзятого оценивания, рассмотрим:

1° эффект экзаменационной оценки (задача 19.1);

2° эффект практического занятия — зависимость оценки преподавателя студентами от того, проводит преподаватель в группе (параллельно с чтением лекций) практические занятия или нет (задача 19.5);

3° эффект экзамена (задача 19.17).

Можно исследовать и другие факторы как возможные причины предвзятого оценивания.

Эффект экзаменационной оценки. Еще до начала анкетирования высказывались соображения, что оценка характеристик преподавателя студентами не может быть непредвзятой. Как на причину предвзятого оценивания указывалось на то, что студенту на экзамене преподаватель выставляет оценку, и поэтому, оценивая преподавателя, студент сознательно или подсознательно учитывает результат экзамена — полученную оценку, особенно если она неудовлетворительная.

В связи с выдвинутой гипотезой о наличии эффекта экзаменационной оценки возникает вопрос: согласуется ли она с экспериментом (результатами анкетирования)?

Таблица 19.4.1. Оценки студентами характеристик преподавателя

Характеристика преподавателя	Оценка					
1	2	3	4	5	6	7
1. Преподает ясно и доступно	8,4	7,3	7,9	7,9	7,1	8,1
2. Разъясняет сложные места	8,3	7,3	8,5	8,0	7,5	8,3
3. Выделяет главные моменты	8,5	7,5	8,5	8,4	7,8	8,5
4. Умеет вызвать и поддержать интерес аудитории к предмету	8,6	6,4	7,9	8,7	5,8	7,2
5. Следит за реакцией аудитории	8,4	8,3	7,9	7,3	6,1	7,6
6. Задает вопросы, побуждает к дискуссии	8,2	5,5	7,2	6,0	4,7	6,7
7. Соблюдает логическую последовательность изложения	8,5	7,7	8,9	8,6	8,4	8,8
8. Демонстрирует культуру речи, четкость дикции, нормальный темп изложения	7,0	3,9	6,5	6,4	5,5	6,5
9. Умеет снять напряжение	8,1	5,6	7,1	5,8	5,1	6,5
10. Ориентирует на использование материала в будущей профессиональной деятельности	6,9	6,1	7,9	5,1	3,9	6,5
11. Творческий подход и интерес к своему делу	8,5	7,2	7,9	7,9	7,1	8,1
12. Доброжелательность и такт по отношению к студентам	8,0	6,6	7,9	8,1	7,7	7,9
13. Терпение	8,2	5,8	8,5	8,0	7,1	8,3
14. Требовательность	8,8	8,3	8,9	8,7	8,8	8,8
15. Заинтересованность в успехах студентов	8,4	6,4	8,0	7,7	7,2	7,8
16. Объективность в оценивании знаний студентов	8,7	6,2	7,4	7,7	7,3	7,5
17. Уважительное отношение к студентам	8,6	6,5	8,5	8,3	7,9	8,4
18. Располагает к себе высокой эрудицией, манерой поведения	8,8	6,5	7,9	8,4	8,1	8,2

В таблице 19.4.1 (столбцы 2 и 3) приведены результаты анкетирования в студенческих группах: столбец 2 — в группе ПМ-84-4, столбец 3 — в группе ПМ-84-1. При этом обе студенческие группы находились в одинаковых условиях относительно преподавателя: в обеих группах

на протяжении года он читал лекции, вел практические и лабораторные занятия, принимал зачеты, экзамены. Однако результаты экзамена в этих группах заметно отличаются количеством полученных студентами двоек (см. табл. 19.4.2 и примечание 1).

Сформулировать задачу об эффекте экзаменационной оценки как задачу проверки статистических гипотез и решить ее:

выбрать основную гипотезу и альтернативные (что обозначает отклонение основной гипотезы? ее неотклонение?);

предложить уровень значимости и критерий для проверки основной гипотезы;

проверить основную гипотезу;

дать частотную интерпретацию полученным результатам.

Т а б л и ц а 19.4.2. Результаты экзамена

Оценка на экзамене	Количество оценок в группах			
	ПМ-84-4	ПМ-84-1	ПМ-85-1 ПМ-85-2	ПМ-85-3 ПМ-85-5
Отлично	1	4	3	4
Хорошо	7	2	10	6
Удовлетворительно	4	6	7	6
Неудовлетворительно	6	9	11	15

Примечание 1. В таблице 19.4.1 приведены результаты оценивания автора как преподавателя студентами групп ПМ-84-4, ПМ-84-1, ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4, а именно, оценки в баллах характеристик преподавателя — среднее в группе или в нескольких группах (максимально возможная оценка характеристики — 9 баллов):

столбец 2 — в группе ПМ-84-4; на экзамене шесть студентов получили неудовлетворительные оценки (четверо из них отказались отвечать после ознакомления с содержанием экзаменационного билета, т. е. фактически сами оценили свои знания как неудовлетворительные (см. табл. 19.4.2));

столбец 3 — в группе ПМ-84-1; на экзамене девять студентов получили неудовлетворительные оценки (см. табл. 19.4.2);

столбец 4 — в группах ПМ-85-1, ПМ-85-2, где лектор вел также практические занятия;

столбец 5 — в группах ПМ-85-3, ПМ-85-4, где лектор не вел практических занятий (их проводил другой преподаватель);

столбец 6 — в группах ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4 (анкетирование проведено до экзамена);

столбец 7 — в группах ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4 (анкетирование проведено после экзамена).

Анкета приведена в оригинальном виде.

Примечание 2. В таблице 19.4.2 приведены результаты первой сдачи экзамена: речь идет об экзамене по курсу “Теория вероятностей и математическая статистика” на факультете прикладной математики Днепропетровского государственного университета во время зимних сессий 1986/87 и 1987/88 учебных годов.

19.2. Стохастический эксперимент состоит в последовательном подбрасывании симметричной игральной кости трижды и регистрации результатов следующим образом: появление 1 или 2 очков обозначаем нулем, 3 или 4 — единицей, 5 или 6 — двойкой. Если, например, результат эксперимента: выпало 3, 5, 1 очков, то это регистрируется как 120, если результат: выпало 6, 1, 4 очков, то это регистрируется как 201 и т. д.

Вместе с этими последовательностями из нулей, единиц и двоек рассмотрим числа из промежутка $[0;1]$, которые в троичной системе исчисления записываются соответственно так: 0,120; 0,201; ... Так что 0,120 — это запись в троичной системе числа

$$1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3^2} + 0 \cdot \frac{1}{3^3} = \frac{5}{9},$$

а 0,201 — числа

$$2 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3^2} + 1 \cdot \frac{1}{3^3} = \frac{19}{27}.$$

Эксперимент проводится восемь раз.

Можно ли считать, что полученные таким образом восемь чисел являются реализацией выборки объемом 8 из равномерного на промежутке $[0;1]$ распределения?

Дать ответ на этот вопрос в терминах проверки статистических гипотез, воспользовавшись критерием Колмогорова.

19.3. Регистрируются интервалы между последовательными импульсами вдоль нервного волокна у индивидуумов A и B (единица измерения — $1/50$ с). Они оказались такими.

Индивидуум A : 2,5; 2,0; 7,0; 4,0; 10,5; 1,0; 31,5; 17,5; 0,5; 19,5; 21,5; 1,5; 19,5; 2,0; 8,5; 11,5; 39,0; 7,0; 4,0; 5,5; 3,5; 22,5; 23,0; 10,0; 9,5; 25,5; 4,5; 11,0; 14,5; 0,5.

Индивидуум B : 9,5; 3,0; 19,5; 4,0; 1,5; 14,0; 4,5; 8,5; 22,5; 20,0; 3,5; 15,0; 8,0; 12,0; 40,5; 67,5; 0,5; 1,0; 1,5; 3,0; 6,0; 16,5; 32,5; 4,0; 7,5; 34,0; 15,0; 3,0.

Свидетельствуют ли эти данные об отличиях в интервалах между импульсами у индивидуумов A и B ?

19.4. С помощью таблицы случайных чисел (см. табл. 22.10.1) можно получить (смоделировать) реализацию выборки из равномерного на промежутке $[0;1]$ распределения описанным ниже способом (кстати, по выборке из равномерного на промежутке $[0;1]$ распределения можно построить выборку из любого распределения F с монотонно возрастающей функцией $F(x)$ (см. пример 19.1.1)).

Из таблицы случайных чисел (см. табл. 22.10.1) выберем n чисел (пусть, например, $n = 10$). Выбор чисел можно начинать с любого места любым наперед оговоренным способом. Начиная, скажем, с левого верхнего угла таблицы и двигаясь книзу (для определенности будем выбирать четырехзначные числа), получим 1009, 3754, 0842, 9901, 1280, 6606, 3106, 8526, 6357, 7379.

Рассмотрим числа из промежутка $[0;1]$, полученные по выбранным случайными числами: 0,1009; 0,3754; 0,0842; 0,9901; 0,1280; 0,6606; 0,3106; 0,8526; 0,6357; 0,7379.

Воспользовавшись критерием Колмогорова, убедитесь, что эту последовательность чисел можно считать реализацией выборки из равномерного на промежутке $[0;1]$ распределения.

19.5 (анкетирование “Преподаватель глазами студентов” — эффект практического занятия). При оценивании студентами преподавателя выдвигается предположение о наличии эффекта практического занятия: если в группе параллельно с чтением лекций преподаватель ведет практические (лабораторные) занятия, то оценка его студентами выше.

В таблице 19.4.1 приведены оценки характеристик преподавателя студентами: столбец 4 — в группах ПМ-85-1 и ПМ-85-2, где преподаватель читал лекции и вел практические и лабораторные занятия, а столбец 5 — в группах ПМ-85-3 и ПМ-85-4, где практические и лабораторные занятия вел другой преподаватель (анкетирование проведено после экзамена).

Проверить, существует ли эффект практического (лабораторного) занятия.

Сформулировать задачу о наличии эффекта практического (лабораторного) занятия при оценивании преподавателя как задачу проверки статистических гипотез и решить ее, подробнее см. задачу 19.1 (выберите 5% уровень значимости).

19.6. Решить задачу 17.8, воспользовавшись критерием Вилкоксона.

19.7. Шейки рабочей части сверл обрабатываются на шлифовальном станке. Номинальный диаметр шейки составляет 9,8 мм с техническим допуском 0,05 мм.

Измерения рабочей части шейки 20 сверл дало такие результаты (в миллиметрах): 9,76; 9,77; 9,83; 9,79; 9,80; 9,81; 9,79; 9,75; 9,80; 9,78; 9,75; 9,81; 9,82; 9,76; 9,78; 9,77; 9,81; 9,79; 9,77; 9,79.

Можно ли считать, что номинальный диаметр шейки равен 9,8 мм с техническим допуском 0,05 мм при норме отхода 1 %, другими словами, что диаметр шейки сверла находится в пределах от 9,75 до 9,85 мм с вероятностью 0,99?

Сформулировать поставленную задачу как задачу проверки статистических гипотез. Воспользоваться критерием Колмогорова.

19.8 (влияние работы метронома на плавность речи). В таблице приведены результаты эксперимента,

в котором изучалось влияние работы метронома на речь людей, страдающих заиканием.

Номер участника	Число заиканий при условии			Номер участника	Число заиканий при условии		
	N	R	A		N	R	A
1	15	3	5	7	10	0	2
2	11	3	3	8	8	0	3
3	18	1	3	9	13	0	2
4	21	5	4	10	4	1	0
5	6	2	2	11	11	2	4
6	17	0	2	12	17	2	1

Обследовалось 12 человек с тяжелой формой заболевания. Каждый из них импровизировал трехминутную речь при условиях N , R , A :

N — говорить без метронома;

R — говорить при регулярной (ритмичной) работе метронома (120 ударов за минуту), причем человек был предварительно проинструктирован о необходимости произносить один слог слова на каждый удар метронома;

A — говорить при неритмичной работе метронома, работающего со случайными интервалами между ударами (от 0,3 до 0,7 с), совершая при этом в среднем те же 120 ударов в минуту (при условии A , как и при условии R , человек должен произносить один слог на каждый удар метронома).

Приведенные данные, безусловно, свидетельствуют о том, что работа метронома (как ритмичная так и неритмичная) уменьшает количество заиканий (почему?). А существуют ли отличия во влиянии на заикание ритмично и неритмично работающих метрономов?

19.9. Пробы очень чистого железа, полученные по методам A и B , имели следующие точки плавления (в градусах Цельсия).

Метод A : 1439, 1519, 1518, 1512, 1514, 1489, 1508, 1503.

Метод B : 1509, 1494, 1512, 1483, 1507, 1491.

Различаются ли точки плавления железа, полученного разными методами?

19.10 (случайно и произвольно). Случайно — не означает произвольно. Случайность подчиняется своим

строгим законам (см. задачи 18.12, 18.17, 18.20). Выписать случайную последовательность чисел не так просто, как это может показаться на первый взгляд.

Проверьте свои способности: выпишите 50 случайных, на ваш взгляд, чисел из отрезка $[0;1]$ (для определенности — с четырьмя знаками после запятой). Если предложенные числа в самом деле случайные, то они должны быть равномерно распределены на промежутке $[0;1]$.

Воспользовавшись критерием Колмогорова, выясните, можно ли предложенную последовательность чисел считать случайной.

Примечание. Выписывая случайную последовательность чисел из промежутка $[0;1]$, не просматривайте и не используйте каким-либо другим способом уже выписанную часть последовательности.

19.11. Астрономы A и B измеряли в минутах большой диаметр наиболее ярких галактик. Каждый измерил диаметры 18 галактик. Результаты измерений приведены в таблице.

Номер галактики	Астроном		Номер галактики	Астроном	
	A	B		A	B
224	2,3	2,4	441	1,1	1,1
272	1,5	1,5	557	5,5	5,6
313	1,5	1,3	563	2,0	2,1
354	4,5	4,1	773	1,7	1,7
377	3,5	4,5	604	1,9	2,2
391	1,5	1,3	638	1,4	1,3
407	1,0	1,2	796	2,3	2,8
427	3,0	2,0	805	2,3	2,6
436	1,2	1,1	808	0,8	1,0

Указывают ли эти данные на существенное отличие результатов измерений?

19.12. Для исследования устойчивости к истиранию эпоксидной пластмассы с разными наполнителями изготовили две партии плиток с использованием как наполнителя железных и медных опилок. Изготовили по 27 образцов плиток с каждым наполнителем. Затем каждую плитку приводили в возвратно-поступательное движение по абразивному материалу (выполняя при этом 10 000 циклов для каждой плитки). Измеряли разницу в толщине плиток в дюймах (1 дюйм = 0,0254 м.) до и после

испытаний; точность измерений составляла 10^{-4} дюйма. Получены следующие результаты.

Железные опилки: 3,6; 1,6; 2,3; 3,8; 1,0; 2,5; 2,9; 1,6; 2,4; 2,1; 1,1; 1,7; 2,6; 1,1; 1,0; 1,6; 0,9; 1,4; 2,3; 1,1; 1,8; 2,6; 1,2; 2,0; 2,1; 0,7; 2,4.

Медные опилки: 2,8; 1,5; 2,4; 2,6; 1,1; 2,1; 1,9; 1,4; 2,0; 2,1; 1,0; 1,5; 1,0; 1,6; 1,5; 1,2; 1,7; 2,6; 1,3; 2,6; 2,9; 1,5; 3,3; 2,8; 1,2; 2,7; 1,8.

Указывают ли эти данные на существенное различие устойчивости к истиранию образцов эпоксидных пластмасс, изготовленных с разными наполнителями?

19.13 (момент последнего уравнивания). Эксперимент состоит в последовательном подбрасывании монеты $2n = 40$ раз и регистрации момента (номера испытания), когда в последний раз количество выпавших “гербов” и “решеток” было одинаково. Будем говорить: наблюдается момент последнего уравнивания количества “гербов” и “решеток”. В эксперименте этими моментами могут быть $0, 2, 4, \dots, 40$.

Относительно распределения момента последнего уравнивания выдвигается гипотеза: момент $2k$ последнего уравнивания имеет распределение арксинуса; подробнее — случайная величина $2k/2n = k/n$ имеет распределение арксинуса, т. е. ее функцией распределения является

$$A(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{2}{\pi} \arcsin \sqrt{x}, & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}$$

Провести эксперимент пять раз, фиксируя момент последнего уравнивания количества “гербов” и “решеток”.

Можно ли на основании полученных данных считать, что момент последнего уравнивания имеет распределение арксинуса?

Ответ дать в терминах проверки статистических гипотез. Воспользоваться критерием Колмогорова.

З а м е ч а н и е 1. Эксперимент удобно проводить так. Подбрасываем монету и выпадение “герба” обозначаем через $+1$, а “решетки” через -1 . После 40 подбрасываний получаем последовательность, образованную из чисел $+1$ и -1 . Последовательно складывая члены последовательности, начиная с первого, зафиксируем момент,

когда сумма будет равна нулю — это будет момент второго уравнивания (момент первого уравнивания равен нулю); и так продолжаем, пока не найдем значение момента последнего уравнивания (см. также задачу 19.22).

З а м е ч а н и е 2. Ниже приведены значения функции распределения арксинуса $A(x)$ в точках $0,05 \cdot i$, $i = 0, 1, 2, \dots, 10$:

x	$A(x)$	x	$A(x)$
0,00	0,000	0,30	0,369
0,05	0,144	0,35	0,403
0,10	0,205	0,40	0,436
0,15	0,253	0,45	0,468
0,20	0,295	0,50	0,500
0,25	0,333		

Для вычисления значений $A(x)$, если $0,5 < x < 1,0$, можно воспользоваться соотношением

$$A(x) = 1 - A(1 - x).$$

19.14. При испытании на устойчивость к истиранию люминесцентной краски типов А-102 и А-108 измеряли потерю массы (в граммах) через определенный интервал времени, причем известно, что устойчивость к истиранию краски А-108 не меньше, чем краски А-102. Краску каждого из двух типов наносили на восемь панелей. Получили такие результаты:

Потеря массы		Потеря массы	
А-102	А-108	А-102	А-108
17	19	33	35
10	17	20	25
19	16	22	23
17	12	28	21

Можно ли на основании этих данных утверждать, что краски существенно различаются по устойчивости к истиранию?

19.15. Из большой группы рабочих, которые работают на сборочном конвейере, наудачу выбирают двух и несколько раз хронометрируют продолжительность сборки ими определенного изделия (в минутах). Получили такие результаты.

Первый рабочий: 19,4; 21,1; 16,2; 21,2; 21,6; 17,8; 19,6.

Второй рабочий: 19,9; 15,7; 15,2; 19,8; 18,9; 16,1; 16,2; 18,5; 17,3; 20,4.

Существенно ли различается время сборки изделия этими рабочими?

19.16 (показания механических часов). Механические часы, выставленные в витринах часовых магазинов, через некоторое время останавливаются. Момент остановки часов — случайная величина. Естественно предположить, что показания часов распределены равномерно на промежутке $[0;12]$. Наблюдения 20 часов дали следующую выборку:

10 час 57 мин, 3 час 28 мин, 5 час 18 мин, 1 час 32 мин, 3 час 21 мин, 4 час 13 мин, 4 час 06 мин, 10 час 10 мин, 6 час 48 мин, 2 час 51 мин, 0 час 44 мин, 3 час 57 мин, 4 час 16 мин, 4 час 58 мин, 7 час 47 мин, 11 час 04 мин, 7 час 54 мин, 10 час 30 мин, 11 час 06 мин, 11 час 52 мин.

Согласуется ли с этими данными гипотеза о равномерном распределении на промежутке $[0;12]$ показаний часов?

19.17 (анкетирование “Преподаватель глазами студентов” — эффект экзамена). Анкетирование “Преподаватель глазами студентов” должно проводиться после экзамена, но в группах ПМ-85-1, ПМ-85-2, ПМ-85-3, ПМ-85-4 его ошибочно провели до экзамена (результаты анкетирования приведены в столбце 5 табл. 19.4.1). Потом анкетирование провели и после экзамена (результаты приведены в столбце 6 табл. 19.4.1). Эта ошибка дала возможность говорить о существовании эффекта экзамена (а может, это эффект еще какого-нибудь другого фактора?), а именно, влияет ли на оценивание преподавателя студентами проведение экзамена? В связи с этой гипотезой выяснить, согласуется ли она с данными анкетирования. Речь идет об экзамене по годовому курсу “Теория вероятностей и математическая статистика” (результаты экзамена приведены в табл. 19.4.2).

Сформулировать задачу о наличии эффекта экзамена при оценивании преподавателя как задачу проверки статистических гипотез и решить ее (подробнее см. задачу 19.1).

Примечание. Эффект экзамена оказался довольно неожиданным: несмотря на высокую требовательность преподавателя (см. табл. 19.4.2, а также строку “Требовательность” в табл. 19.4.1), данные анкетирования после экзамена заметно выше, чем до экзамена (см. столбцы 5 и 6 табл. 19.4.1).

19.18. Средний объем стока воды в реке (в кубических футах за секунду, 1 фут = 0,3048 м) фиксируется ежемесячно в течение двух лет (обозначим их I и II). Сравнивая сток за соответствующие месяцы (сток подчиняется годовому циклам), сделать вывод относительно отсутствия (или наличия) изменений объема стока воды в разные годы.

Месяц	Год		Месяц	Год	
	I	II		I	II
Январь	14,1	14,2	Июль	92,8	88,1
Февраль	12,2	10,5	Август	74,4	80,0
Март	104,0	123,0	Сентябрь	75,4	75,6
Апрель	220,0	190,0	Октябрь	51,7	48,8
Май	110,0	138,0	Ноябрь	29,3	27,1
Июнь	86,0	98,1	Декабрь	16,0	15,7

Проверить нулевую гипотезу об отсутствии систематических изменений объема стока за указанные годы.

19.19. Получили 10 чисел способом, описанным в задаче 15.37. Можно ли считать, что полученные таким образом 10 чисел являются реализацией выборки из равномерного на промежутке $[0;1]$ распределения?

Ответ дать в терминах проверки статистических гипотез, воспользоваться критерием Колмогорова.

19.20. Регистрировались интервалы безотказной работы (в часах) кондиционеров на двух самолетах “Боинг-720”. Получены такие результаты.

Самолет I: 74, 57, 48, 29, 502, 12, 70, 21, 386, 59, 22, 153, 26, 326.

Самолет II: 55, 320, 56, 104, 220, 239, 47, 246, 176, 182, 33, 15, 104, 35.

Можно ли на основании приведенных данных сделать вывод, что время безотказной работы кондиционеров на этих самолетах одинаково или данные свидетельствуют

о наличии существенной разницы в длительности безотказной работы кондиционеров?

19.21. Для изготовления корда с одинаковыми номинальными данными два завода (A и B) используют разные производственные процессы. Проведено по 20 испытаний продукции каждого завода на разрыв. Катушку корда для испытаний и место разрыва на ней выбирали наудачу.

Приведенные данные являются отклонениями прочности от 21,5 фунта (1 фунт = 453,6 г). Единица измерения составляет 0,1 фунта.

Завод A : $-1; -5; 1; 10; 2; -3; 6; 10; -1; 4; -8; -1; -10; -9; -2; -2; -8; 1; 2; 5$.

Завод B : $10; 8; 9; 12; 0; 8; 5; 9; -1; -1; 7; 16; -5; 1; 10; 9; -5; 6; 6; 15$.

Существенно ли отличается корд, изготовленный на заводах A и B ?

19.22 (время пребывания на “положительной” стороне). Эксперимент состоит в последовательном подбрасывании монеты $2n = 40$ раз (одно подбрасывание за единицу времени). Появление “герба” будем обозначать через $+1$, а появление “решетки” — через -1 . Таким образом, на k -м шаге, $k = 1, 2, \dots, 2n$, имеем символ ε_k , равный $+1$ или -1 в зависимости от того, какой стороной легла монета. Пусть

$$s_k = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_k, \quad s_0 = 0, \quad k = 1, 2, \dots, 2n,$$

где s_k — разность между количеством “плюсов” и “минусов” (между количеством “гербов” и “решеток”). Если воспользоваться геометрической терминологией и системой координат (t, x) , то результат эксперимента можно представить в виде ломаной с вершинами в точках (k, s_k) , $k = 1, 2, \dots, 2n$ (рис. 19.4.1).

Вычислим время, когда монета находилась на “положительной” стороне, т. е. когда разность между количеством “гербов” и “решеток” была положительной. Например, если в результате первых 10 подбрасываний монеты получена последовательность $+1, +1, -1, -1, -1, -1, -1, +1, +1, -1$ (см. также рис. 19.4.1), то время пребывания монеты на “положительной” стороне составляет 4 единицы.

Относительно времени пребывания на “положительной” стороне выдвигается гипотеза: время пребывания на “положительной” стороне имеет распределение арксинуса. Подробнее, распределение арксинуса (см. задачу 19.3) имеет случайная величина $2k/2n = k/n$, где через $2k$ обозначено количество единиц времени (из $2n$ единиц), когда монета пребывала на “положительной” стороне.

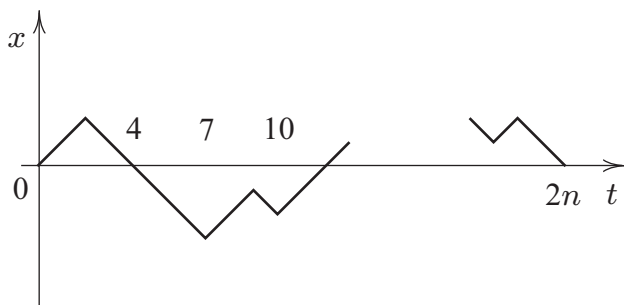


Рис. 19.4.1: Результат эксперимента — ломаная с вершинами в точках (k, s_k) , $k = 0, 1, \dots, 2n$

Провести эксперимент пять раз, фиксируя время пребывания монеты на “положительной” стороне.

Можно ли на основании полученных данных заключить, что время пребывания монеты на “положительной” стороне имеет распределение арксинуса?

Ответ дать в терминах проверки статистических гипотез, воспользоваться критерием Колмогорова и замечанием 1 к задаче 19.13.

19.23. Четное число мышей рассадили наудачу по одной в клетку. Затем клетки снова-таки наудачу объединили в две одинаковые по количеству мышей группы. Мыши первой группы (A) предназначались для контроля, а второй — подопытной (B) подвергались действию определенного препарата. После этого все мыши в случайном порядке инфицировались туберкулезом.

Заметим, что экспериментаторы, как правило, инфицируют сначала мышей контрольной группы, а затем — подопытной, что абсолютно неверно.

Дни гибели мышей после инфицирования приведены в таблице (данные об одной из мышей были утеряны).

Группа *A*: 5, 6, 6, 7, 8, 9, 11, 12.

Группа *B*: 7, 7, 8, 8, 9, 10, 13, 14, 15.

В предыдущих экспериментах установлено, что используемый препарат нетоксичный. Поэтому можно предположить, что мыши в подопытной группе гибнут не быстрее, чем в контрольной группе.

Свидетельствуют ли приведенные данные о наличии эффекта препарата?

У к а з а н и е. Проверить гипотезу H_0 : препарат не влияет на течение туберкулеза. Воспользоваться критерием Вилкоксона.

19.24. Для сравнения эффективности двух методов преподавания математики сформировали две группы учеников (*A* и *B*) одного уровня подготовки, в каждой из которых математику изучали с использованием своего метода. После окончания учебного года каждому ученику был предложен тест. Результаты теста (в баллах) оказались такими.

Группа *A*: 94, 92, 90, 86, 86, 84, 82, 90.

Группа *B*: 90, 86, 84, 82, 80, 82, 78.

Можно ли на основании этих данных сделать вывод, что оценки, полученные в группах *A* и *B*, существенно отличаются?

19.25. Провести серию из 10 экспериментов, каждый из которых состоит в подбрасывании 16 монет и регистрации количества s_i выпавших “гербов” (i — номер эксперимента, $i = 1, 2, \dots, 10$). Рассмотреть последовательность чисел

$$\xi_i = \frac{s_i - 8}{2}, \quad i = 1, 2, \dots, 10.$$

Можно ли считать, что $\xi_1, \xi_2, \dots, \xi_{10}$ является выборкой из стандартного нормального распределения?

Из каких соображений выдвинута сформулированная гипотеза?

П р и м е ч а н и е. Эксперимент, состоящий в подбрасывании 16 монет, удобно провести, например, так: поместить в коробку все 16 монет, а потом, хорошо встряхнув ее несколько раз, подсчитать количество выпавших “гербов”.

19.26. В таблице приведены значения объемов булок (в миллилитрах), выпеченных из 100-граммовых порций теста, изготовленного из 12 разных сортов муки, содержащей 1 и 2 мг бромистого калия (KBr).

Сорт муки	Объем булки, мл		Сорт муки	Объем булки, мл	
	KBr 1 мг	KBr 2 мг		KBr 1 мг	KBr 2 мг
1	1075	1055	7	900	905
2	980	955	8	860	870
3	850	820	9	940	1000
4	815	765	10	1000	1015
5	1040	1065	11	935	965
6	960	985	12	835	870

Свидетельствуют ли приведенные данные о влиянии содержания бромистого калия на объем булок?

Ответ дать в терминах проверки статистических гипотез.

19.27. Приборы типов A и B для измерения количества осадков размещены наудачу на некотором участке. За определенный промежуток времени над контролируемой областью пронеслось 14 ураганов. Среднее количество осадков, измеренное с помощью приборов A и B , приведено в таблице.

Ураган	Прибор типа		Ураган	Прибор типа	
	A	B		A	B
1	1,38	1,42	8	2,63	2,69
2	9,69	10,37	9	2,44	2,68
3	0,39	0,36	10	0,56	0,53
4	1,42	1,46	11	0,69	0,72
5	0,54	0,55	12	0,72	0,72
6	6,94	6,15	13	0,95	0,93
7	0,59	0,61	14	0,50	0,53

Можно ли заключить, что приборы A и B дают одинаковые результаты?

19.28. Стохастический эксперимент состоит в подбрасывании игральной кости пять раз и вычислении суммы выпавших очков. Для i -го эксперимента обозначим эту сумму через s_i . Проведем эксперимент 16 раз и рассмотрим последовательность

$$\xi_i = \frac{s_i - 17,5}{3,82}, \quad i = 1, 2, \dots, 16.$$

Можно ли считать, что последовательность чисел $\xi_1, \xi_2, \dots, \xi_{16}$ является реализацией выборки из нормального распределения, со средним $a = 0$ и дисперсией $\sigma^2 = 1$? Из каких соображений относительно распределения случайной величины ξ выдвинута именно эта гипотеза?

Ответ дать в терминах проверки статистических гипотез, воспользовавшись критерием Колмогорова.

19.29 (катастрофы на угольных шахтах). В таблицах приведены интервалы в днях между катастрофами (катастрофа влечет смерть 10 и более человек) на угольных шахтах Великобритании с 1875 по 1951 г. (данные В. Мегью, К. Пирсона, А. Винна, их следует читать по строкам).

Интервалы между катастрофами с 1875 по 1900 г.

378	36	15	31	215	11	137	4
15	72	96	124	50	120	203	176
55	93	59	315	59	61	1	13
189	345	20	81	286	114	108	188
233	28	22	61	78	99	326	275
54	217	113	32	23	151	361	312
354	58	275	78	17	1205	644	

Интервалы между катастрофами с 1901 по 1951 г.

467	871	48	123	457	498	49	131
182	255	195	224	566	390	72	228
271	208	517	1613	54	326	1312	348
745	217	120	275	20	66	291	4
369	338	336	19	329	330	312	171
145	75	364	37	19	156	47	129
1630	29	217	7	18	1357		

Свидетельствуют ли эти данные о наличии существенных отличий между интервалами времени от катастрофы к катастрофе за период с 1875 по 1900 г. и с 1901 по 1951 г.?

19.30. Исследовалась длительность эксплуатации авиационных пневматических шин марок U и V на самолетах, базирующихся на аэродромах. Для испытаний отобрали 10 самолетов. Приведенные в таблице данные — количество посадок до разрушения шин.

Можно ли по результатам этих испытаний сделать вывод о наличии отличий в длительности эксплуатации шин марок U и V ?

Само- лет	Количество посадок		Само- лет	Количество посадок	
	Марка U	Марка V		Марка U	Марка V
1	5	5	6	7	24
2	55	14	7	14	38
3	5	24	8	32	41
4	32	27	9	10	32
5	56	24	10	8	24

19.31 (автомобильные номера). Выпишите цифровые части номеров 20 автомобилей, проезжающих мимо (номер содержит пять цифр). Обозначим через a_i , b_i , c_i , d_i , e_i цифры, встретившиеся в номере, если его читать слева направо.

Рассмотрим последовательность чисел

$$\zeta_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i + 10^{-5}e_i,$$

где $i = 1, 2, \dots, 20$.

Можно ли считать, что последовательность ζ_i , $i = 1, 2, \dots, 20$, является реализацией выборки из равномерного на промежутке $[0; 1]$ распределения?

З а м е ч а н и е. Если номер автомобиля содержит четыре цифры, то рассмотрим последовательность

$$\zeta_i = 10^{-1}a_i + 10^{-2}b_i + 10^{-3}c_i + 10^{-4}d_i,$$

где $i = 1, 2, \dots, 20$.

19.32 (электрошок и аккомодация). При последовательном чтении вслух людьми, страдающими заиканием, одного и того же текста несколько раз подряд число заиканий имеет тенденцию к уменьшению. Электрошок как метод борьбы с заиканием претендует на ускорение этого процесса (аккомодацию). Цель описанного ниже эксперимента — ответить на вопрос: так ли это?

Восемнадцать человек студенческого возраста, страдающих заиканием, согласились участвовать в эксперименте — читали вслух фрагмент текста последовательно по пять раз (без действия электрошока). Потом они читали вслух пять раз другой фрагмент текста и в момент заикания подвергались воздействию электрошока.

В таблице приведены полученные по специальной методике оценки аккомодации в баллах. Высокий балл соответствует лучшей аккомодации. (Данные из работы: Daly

D. A. and Cooper E. B. Rate of stuttering adaptation under condition// Behav. Res. Theory. 1967. Vol. 5. P. 49–54.)

Свидетельствуют ли эти данные о том, что шок в момент заикания влияет на аккомодацию?

Номер участ- ника	Баллы		Номер участ- ника	Баллы	
	Без шока	С шо- ком		Без шока	С шо- ком
1	57	51	10	50	50
2	59	56	11	44	56
3	41	44	12	50	46
4	51	44	13	70	74
5	43	50	14	42	57
6	49	54	15	68	74
7	48	50	16	54	48
8	56	40	17	38	48
9	44	50	18	48	44

19.33. Токсичность двух препаратов (A и B) сравнивали на двух группах мышей (по 50 в каждой). Мышам первой группы ввели препарат A , а второй — такую же дозу препарата B . Количество мышей, которые остались живыми в каждой группе через определенное время T (в часах) после введения препаратов, приведено в таблице.

Время T	Количество живых мышей	
	После действия A	После действия B
1	29	35
3	21	23
8	16	16
16	12	13
24	10	10
48	9	8
72	7	8
96	6	7

Свидетельствуют ли приведенные данные о различной токсичности препаратов A и B ?

19.34. Шейки рабочей части сверл обрабатываются на шлифовальном станке. Номинальный диаметр шейки составляет 9,8 мм с техническим допуском 0,04 мм.

Были измерены диаметры рабочей части шейки 16 сверл. Их значения (в миллиметрах) оказались такими:

9,76; 9,78; 9,81; 9,77; 9,75; 9,78; 9,75; 9,77; 9,74; 9,78; 9,77; 9,83; 9,78; 9,81; 9,79; 9,80.

Можно ли считать, что номинальный диаметр шейки равен 9,8 мм с техническим допуском 0,04 мм при норме отхода 5 %, т. е. не выходит за пределы 9,76—9,84 мм с вероятностью 0,95?

Указание 1. Сформулировать поставленную задачу как задачу проверки статистических гипотез и воспользоваться критерием Колмогорова.

Указание 2. Решить задачу в предположении, что результаты измерений являются выборкой из нормального распределения (см. также решение к задаче 19.7).

19.35 (мантиcсы логарифмов чисел $n!$). В таблице приведены значения десятичных логарифмов чисел $n!$

n	$n!$	$\lg n!$
1	1	0,000000
2	2	0,301030
3	6	0,778151
4	24	1,380211
5	120	2,079181
6	720	2,857332
7	5040	3,702431
8	40320	4,605521
9	362880	5,559763
10	3628800	6,559763
11	39916800	7,601156
12	47900160 · 10	8,680337
13	62270208 · 10 ²	9,794280
14	87178291 · 10 ³	10,940408
15	13076744 · 10 ⁵	12,116499
16	20922790 · 10 ⁶	13,320619

Можно ли считать, что мантиcсы логарифмов чисел $n!$ равномерно распределены на промежутке $[0; 1]$?

Ответ дать в терминах проверки статистических гипотез, воспользоваться критерием Колмогорова.

19.36. Ниже приведены данные о пористости конденсаторной бумаги двух партий (из каждой партии рулоны, пористость бумаги в которых измерялась, брали наудачу).

Партия I : 1,5; 1,5; 2,7; 3,0; 1,6; 1,9; 2,4; 1,6; 1,7; 2,0.

Партия II : 1,9; 2,3; 1,8; 1,9; 1,5; 2,4; 2,9; 3,5; 4,7.

Можно ли по этим данным сделать вывод, что пористость конденсаторной бумаги в партиях I и II различна?

19.37. С помощью таблицы случайных чисел (см. табл. 22.10.1) получить выборку объемом 16 из нормального распределения со средним $a = 1$ и дисперсией $\sigma^2 = 0,01$.

Проверить, воспользовавшись критерием Колмогорова, действительно ли полученная последовательность чисел является реализацией выборки из указанного нормального распределения.

Глава 20

Линейная регрессия

20.1 Нормальная линейная регрессия

В математической статистике часто встречается следующая задача.

Предположим, что величины y и x связаны функциональной зависимостью

$$y = a + bx,$$

но коэффициенты (параметры) a и b неизвестны. (Зависимость между y и x может быть и более сложное, например, она может быть такой: $y = a_0 + a_1x + a_2x^2$ или такой: $y = a \cos x + b \sin x + cx$, y может зависеть от нескольких переменных: $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_dx_d$.) Мы имеем возможность в данных вполне определенных точках $x_1, x_2, \dots, x_n$ наблюдать значения ξ_i величины

$$y_i = a + bx_i, \quad i = 1, 2, \dots, n,$$

но не точно, а с некоторой погрешностью e_i , т. е. фактически наблюдаем

$$\xi_i = a + bx_i + e_i, \quad i = 1, 2, \dots, n.$$

Погрешности e_i , $i = 1, 2, \dots, n$, неизвестны, однако их естественно считать независимыми (наблюдения производятся независимо) нормально распределенными случайными величинами со средним 0 и некоторой (неизвестной) дисперсией σ^2 . По наблюдениям ξ_i , $i = 1, 2, \dots, n$,

необходимо определить (оценить) неизвестные параметры a, b, σ^2 . (Разумеется, если значения $y_i = a + bx_i$, $i = 1, 2, \dots, n$, наблюдаются с погрешностью, то параметры a, b, σ^2 также будут определяться с погрешностью.)

В строгой математической постановке эту задачу можно сформулировать следующим образом.

Дана последовательность независимых нормально распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n$ соответственно со средними $y_i = a + bx_i$ и дисперсией σ^2 . Параметры неизвестны, их необходимо оценить по $\xi_1, \xi_2, \dots, \xi_n$. Функцию

$$y = a + bx,$$

будем называть простой нормальной линейной регрессией, кратко — линейной регрессией (“простая” обозначает зависимость y от одной переменной x , “нормальная” — наблюдения нормально распределены).

Далее линейную регрессию будет удобно записывать в виде

$$y = \alpha + \beta(x - \bar{x}),$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $x_1, x_2, \dots, x_n$ — значения переменной x ;

α, β, σ^2 — неизвестные параметры. Для оценивания неизвестных параметров α, β, σ^2 воспользуемся методом максимального правдоподобия. Введем обозначения

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad S_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_1 = \sqrt{S_1^2},$$

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad S_2^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2, \quad S_2 = \sqrt{S_2^2},$$

$$R_{12} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(\xi_i - \bar{\xi}).$$

Теорема 20.1.1 (об оценках максимального правдоподобия параметров α, β, σ^2). Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые нормально распределенные случайные величины со средними

$$M\xi_i = \alpha + \beta(x_i - \bar{x}), \quad i = 1, 2, \dots, n,$$

и дисперсией σ^2 .

Оценками максимального правдоподобия параметров α, β, σ^2 линейной регрессии

$$y = \alpha + \beta(x - \bar{x})$$

являются

$$\hat{\alpha} = \bar{\xi}; \quad \hat{\beta} = \frac{R_{12}}{S_1^2}; \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - (\hat{\alpha} + \hat{\beta}(x_i - \bar{x})))^2.$$

З а м е ч а н и е. Для оценки σ^2 имеет место представление

$$\hat{\sigma}^2 = S_2^2 \left(1 - \frac{R_{1,2}^2}{S_1^2 S_2^2} \right).$$

Далее мы будем использовать обозначение $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$.

Теорема 20.1.2 (о распределении оценок $\alpha, \hat{\beta}, \hat{\sigma}^2$). Оценки максимального правдоподобия

$$\hat{\alpha} = \bar{\xi}; \quad \hat{\beta} = \frac{R_{12}}{S_1^2}; \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\xi_i - (\hat{\alpha} + \hat{\beta}(x_i - \bar{x})))^2$$

параметров α, β, σ^2 нормальной линейной регрессии

$$y = \alpha + \beta(x - \bar{x})$$

обладают следующими свойствами:

1) $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ — независимые в совокупности случайные величины;

2) $\hat{\alpha}, \hat{\beta}$ — несмещенные и состоятельные оценки соответственно α и β , $\hat{\sigma}^2$ — асимптотически несмещенная и состоятельная оценка σ^2 ;

3)

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\hat{\sigma}^2/(n-2)}} \sim t_{n-2}; \quad (20.1.1)$$

$$\frac{\hat{\beta} - \beta}{\sqrt{\hat{\sigma}^2/(S_1^2(n-2))}} \sim t_{n-2}; \quad (20.1.2)$$

$$n \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2; \quad (20.1.3)$$

$$\frac{((\hat{\alpha} - \alpha)^2 + S_1^2(\hat{\beta} - \beta)^2)/2}{\hat{\sigma}^2/(n-2)} \sim F_{2;n-2}, \quad (20.1.4)$$

(знак $\sim$ заменяет оборот “имеет распределение”).

По известным распределениям оценок $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ (см. (20.1.1), (20.1.2), (20.1.3), (20.1.4)) можно построить критерии проверки гипотез о параметрах α , β , σ^2 и доверительные интервалы для них.

Критерий для проверки гипотезы $H_0 : \beta = \beta_0$.
Если гипотезу $H_0 : \beta = \beta_0$ отклонять при

$$\left| \frac{\hat{\beta} - \beta_0}{\hat{\sigma}/(S_1\sqrt{n-2})} \right| > t_{\gamma;n-2} \quad (20.1.5)$$

и не отклонять в противном случае, то с вероятностью 2γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя).

В частности, гипотеза $H_0 : \beta = 0$ (гипотеза о значимости линии регрессии) проверяется при помощи критерия: если гипотезу $H_0 : \beta = 0$ отклонять при

$$\left| \frac{\hat{\beta}}{\hat{\sigma}/(S_1\sqrt{n-2})} \right| > t_{\gamma;n-2} \quad (20.1.6)$$

и не отклонять в противном случае, то с вероятностью 2γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя).

Критерий для проверки гипотезы $H_0 : \alpha = \alpha_0$. Если гипотезу $H_0 : \alpha = \alpha_0$ отклонять при

$$\left| \frac{\hat{\alpha} - \alpha_0}{\hat{\sigma}/\sqrt{n-2}} \right| > t_{\gamma; n-2} \quad (20.1.7)$$

и не отклонять в противном случае, то с вероятностью 2γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя).

В частности, гипотеза $H_0 : \alpha = 0$ проверяется с помощью критерия: если гипотезу $H_0 : \alpha = 0$ отклонять при

$$\left| \frac{\hat{\alpha}}{\hat{\sigma}/\sqrt{n-2}} \right| > t_{\gamma; n-2} \quad (20.1.8)$$

и не отклонять в противном случае, то с вероятностью 2γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива двусторонняя).

Критерий для проверки гипотезы $H_0 : \sigma^2 = \sigma_0^2$. Если гипотезу $H_0 : \sigma^2 = \sigma_0^2$ отклонять при

$$n \frac{\hat{\sigma}^2}{\sigma_0^2} > \chi_{\gamma; n-2}^2; \quad (20.1.9)$$

и не отклонять в противном случае, то с вероятностью γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива односторонняя: $\sigma^2 > \sigma_0^2$).

Критерий для проверки гипотезы $H_0 : \alpha = \alpha_0, \beta = \beta_0$. Если гипотезу $H_0 : \alpha = \alpha_0, \beta = \beta_0$ отклонять при

$$\frac{((\hat{\alpha} - \alpha_0)^2 + S_1^2(\hat{\beta} - \beta_0)^2)/2}{\hat{\sigma}^2/(n-2)} > F_{\gamma; 2; n-2} \quad (20.1.10)$$

и не отклонять в противном случае, то с вероятностью γ гипотеза H_0 будет отклоняться, когда она верна (альтернатива: $\alpha \neq \alpha_0$ или $\beta \neq \beta_0$).

Доверительные интервалы для параметров простой линейной регрессии. Доверительные интервалы с коэффициентом надежности $1 - 2\gamma$ для параметров α , β , σ^2 просто получить из (20.1.1), (20.1.2), (20.1.3):

$$\hat{\alpha} - \frac{\hat{\sigma}}{\sqrt{n-2}} t_{\gamma;n-2} \leq \alpha \leq \hat{\alpha} + \frac{\hat{\sigma}}{\sqrt{n-2}} t_{\gamma;n-2}, \quad (20.1.11)$$

$$\hat{\beta} - \frac{\hat{\sigma}}{S_1 \sqrt{n-2}} t_{\gamma;n-2} \leq \beta \leq \hat{\beta} + \frac{\hat{\sigma}}{S_1 \sqrt{n-2}} t_{\gamma;n-2}, \quad (20.1.12)$$

$$\frac{n\hat{\sigma}^2}{\chi_{\gamma;n-2}^2} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{\chi_{1-\gamma;n-2}^2}. \quad (20.1.13)$$

Доверительным интервалом с коэффициентом надежности $1 - 2\gamma$ для значения $y_0 = \alpha + \beta(x_0 - \bar{x})$ линии регрессии

$$y = \alpha + \beta(x - \bar{x}),$$

в точке x_0 является

$$\begin{aligned} \hat{\alpha} + \hat{\beta}(x_0 - \bar{x}) - \hat{\sigma} t_{\gamma;n-2} \sqrt{\frac{1}{n-2} \left(1 + \frac{(x_0 - \bar{x})^2}{S_1^2} \right)} &\leq \\ &\leq \alpha + \beta(x_0 - \bar{x}) \leq \\ &\leq \hat{\alpha} + \hat{\beta}(x_0 - \bar{x}) + \hat{\sigma} t_{\gamma;n-2} \sqrt{\frac{1}{n-2} \left(1 + \frac{(x_0 - \bar{x})^2}{S_1^2} \right)} \end{aligned} \quad (20.1.14)$$

Совместная доверительная область для α и β , доверительная область для регрессии. Из (20.1.4) можно получить доверительную область с коэффициентом надежности $1 - \gamma$ для пары (α, β) :

$$\frac{(\alpha - \hat{\alpha})^2}{2F_{\gamma;2;n-2} \hat{\sigma}^2 / (n-2)} + \frac{(\beta - \hat{\beta})^2}{2F_{\gamma;2;n-2} \hat{\sigma}^2 / (S_1^2 (n-2))} \leq 1. \quad (20.1.15)$$

Границей этой области является эллипс.

Доверительной областью с коэффициентом надежности $1 - \gamma$ для линии регрессии $y = \alpha + \beta(x - \bar{x})$ является

$$\begin{aligned} \hat{\alpha} + \hat{\beta}(x - \bar{x}) - \hat{\sigma} \sqrt{F_{\gamma; 2; n-2}} \sqrt{\frac{1}{n-2} \left(1 + \frac{(x - \bar{x})^2}{S_1^2} \right)} &\leq \\ &\leq \alpha + \beta(x - \bar{x}) \leq \\ &\leq \hat{\alpha} + \hat{\beta}(x - \bar{x}) + \hat{\sigma} \sqrt{F_{\gamma; 2; n-2}} \sqrt{\frac{1}{n-2} \left(1 + \frac{(x - \bar{x})^2}{S_1^2} \right)} \end{aligned} \quad (20.1.16)$$

(с вероятностью $1 - \gamma$ линия регрессии $y = \alpha + \beta(x - \bar{x})$ при всех x лежит в указанной области).

Проверка адекватности линейной регрессии. Если имеются повторные наблюдения, то можно ответить на вопрос: “Адекватно ли линия регрессии $y = \alpha + \beta(x - \bar{x})$ описывает данные?”

Под повторными наблюдениями в точке x_j значения $y_j = f(x_j)$ будем понимать независимые нормально распределенные случайные величины $\xi_{j,1}, \xi_{j,2}, \dots, \xi_{j,n_j}$ — со средним

$$y_j = f(x_j) : M\xi_{j,\nu} = f(x_j), \quad \nu = 1, 2, \dots, n_j, \quad (20.1.17)$$

и дисперсией σ^2 . Всего имеется k точек x_j , в которых проводились наблюдения, т. е. $j = 1, 2, \dots, k$. Число (количество) наблюдений, как и ранее, будем обозначать через n , так что

$$n = \sum_{j=1}^k n_j.$$

Для дисперсии σ^2 наблюдений (дисперсии случайных величин $\xi_{j,\nu}$, $\nu = 1, 2, \dots, n_j$, $j = 1, 2, \dots, k$) можно положить оценку

$$s_1^2 = \frac{1}{n-k} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (\xi_{j,\nu} - \bar{\xi}_j)^2,$$

где

$$\bar{\xi}_j = \frac{1}{n_j} \sum_{\nu=1}^{n_j} \xi_{j,\nu}, \quad j = 1, 2, \dots, k, \quad n = \sum_{j=1}^k n_j.$$

Оценка s_1^2 описывает разброс наблюдений $\xi_{j,\nu}$ вне зависимости от того, говорим мы о регрессии или нет.

Оценка

$$s_2^2 = \frac{1}{k-2} \sum_{j=1}^k n_j (\bar{\xi}_j - (\hat{\alpha} + \hat{\beta}(x_j - \bar{x}))^2$$

описывает разброс наблюдений $\xi_{j,\nu}$ относительно эмпирической линии регрессии

$$y = \hat{\alpha} + \hat{\beta}(x - \bar{x}).$$

Если выполнены предположения (20.1.17), и функция $f(x)$ имеет вид

$$f(x) = \alpha + \beta(x - \bar{x}),$$

то отношение s_2^2/s_1^2 имеет $F_{k-2;n-k}$ -распределение. Поэтому для проверки гипотезы

$$H_0 : M\xi_{j,\nu} = \alpha + \beta(x_j - \bar{x}), \quad \nu = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, k$$

(гипотезы об адекватности описания линейной регрессией данных) имеем следующий критерий.

Если гипотезу H_0 отклонять при

$$\frac{s_2^2}{s_1^2} > F_{\gamma; k-2; n-k}$$

и не отклонять в противном случае, то с вероятностью γ гипотезу будем отклонять, когда она верна.

Замечание 1. Выписывая оценки $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ соответственно параметров α , β , σ^2 и значение $\bar{x}$, в случае повторных наблюдений удобно пользоваться двойными индексами не только для наблюдений:

$$\xi_{j,\nu}, \quad \nu = 1, 2, \dots, n_j; \quad j = 1, 2, \dots, k,$$

но и для точек x_j , в которых проводились наблюдения, а именно

$$x_{j,\nu} = x_j, \quad \nu = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, k.$$

При этом для $\bar{x}$, S_1^2 , $\bar{\xi}$, S_2^2 , R_{12} (см. теорему 20.1.1 об оценках параметров) мы получим следующие выражения:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} x_{j,\nu} = \frac{1}{n} \sum_{j=1}^k n_j x_j; \\ S_1^2 &= \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (x_{j,\nu} - \bar{x})^2 = \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (x_j - \bar{x})^2 = \\ &= \frac{1}{n} \sum_{j=1}^k n_j (x_j - \bar{x})^2; \\ \bar{\xi} &= \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} \xi_{j,\nu}, \quad S_2^2 = \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (\xi_{j,\nu} - \bar{\xi})^2; \\ R_{12} &= \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (x_{j,\nu} - \bar{x})(\xi_{j,\nu} - \bar{\xi}) = \\ &= \frac{1}{n} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (x_j - \bar{x})(\xi_{j,\nu} - \bar{\xi}). \end{aligned}$$

З а м е ч а н и е 2. Если повторные наблюдения отсутствуют, т. е. $n_j = 1$, $j = 1, 2, \dots, k$, то

$$n = \sum_{j=1}^k n_j = \sum_{j=1}^k 1 = k, \quad \bar{\xi}_j = \xi_j, \quad j = 1, 2, \dots, k.$$

При этом оценку s_1^2 дисперсии σ^2 получить невозможно и, как следствие, невозможно получить и критерий для проверки адекватности линейной регрессии.

Проверка адекватности теории эксперименту.

Пусть $x_1, x_2, \dots, x_n$ — теоретически предсказанные значения некоторой измеряемой величины, а $\xi_1, \xi_2, \dots, \xi_n$ — фактические ее наблюдения. Возникает вопрос — можно ли с точностью до погрешности считать, что теория согласуется с наблюдениями. Другими словами, можно ли расхождения $\xi_1 - x_1, \xi_2 - x_2, \dots, \xi_n - x_n$ объяснить погрешностью наблюдений или они свидетельствуют о неадекватности теории?

Один из возможных подходов состоит в следующем. Предположим, что наблюдаемая величина y и теоретически предсказанная x связаны линейной зависимостью

$$y = \alpha + \beta(x - \bar{x})$$

($\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ введено для удобства). В эксперименте мы наблюдаем

$$\xi_i = \alpha + \beta(x_i - \bar{x}) + e_i,$$

$i = 1, 2, \dots, n$, где e_i — погрешности наблюдений. По наблюдениям $\xi_1, \xi_2, \dots, \xi_n$ можно оценить неизвестные параметры и проверить относительно них специальным образом выбранные гипотезы: 1) $H_{0;1}: \beta = 0$; 2) $H_{0;2}: \beta = 1$; 3) $H_{0;4}: \beta = \beta_0 = 1; \alpha = \alpha_0 = \bar{x}$. В зависимости от того, отклоняются эти гипотезы или нет, можно говорить о согласии или несогласии теории с экспериментом.

1. Гипотеза $H_{0;1}: \beta = 0$. Неотклонение гипотезы означает, что предсказанный теорией эффект вообще отсутствует.

2. Гипотеза $H_{0;2}: \beta = 1$. Неотклонение гипотезы свидетельствует в пользу согласия теории с экспериментом, если при этом $\alpha - \bar{x} \neq 0$, то присутствует систематическая ошибка.

3. Гипотеза $H_{0;4}: \beta = \beta_0 = 1; \alpha = \alpha_0 = \bar{x}$. Неотклонение гипотезы трактуется как согласие теории с экспериментом.

Пример 20.1.1 (эксперимент Эддингтона). *Рассмотрим пример проверки адекватности теории эксперименту. Речь идет о проверке общей теории относительности по отклонению луча света в поле тяготения. Данные для статистической обработки взяты из сборника*

статей Альберта Эйнштейна: “Физика и реальность” (М., “Наука”, 1965).

Схема опыта, проведенного под руководством Эддингтона, состоит в следующем (см. рис. 20.1.1). Пусть звезда лежит примерно в плоскости земной орбиты. Тогда в момент, когда Земля находится в положении 1, звезда видна в направлении 1.

Через полгода Земля окажется в положении 2, и если бы луч света в поле тяготения Солнца не отклонялся, звезда из положения 2 Земли в направлении 2 не наблюдалась бы. Но звезда из положения 2 тем не менее наблюдается (см. рис. 20.1.1), она как бы смещается, и это смещение можно вычислить. (Разумеется, наблюдать звезду из положения 2 Земли в направлении 2 можно лишь в момент полного солнечного затмения.)

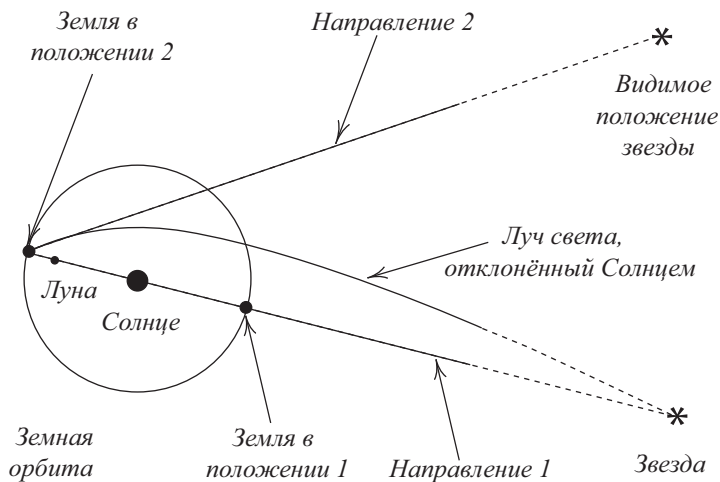


Рис. 20.1.1: Отклонение луча света звезды в поле тяготения Солнца

Для наблюдения было выбрано 7 звезд. Их видимые перемещения (векторы на небесной сфере, которые из-за малости можно считать векторами на плоскости) разлагались по двум осям координат. Полученные результаты (в угловых секундах) приведены в табл. 20.1.1,

20.1.2, где x_i — вычисленная координата вектора перемещения, ξ_i — наблюдаемая координата вектора перемещения.

Согласуются ли вычисленные значения координат векторов перемещений с наблюдаемыми?

Т а б л и ц а 20.1.1. Первая координата

x_i	-0,22	+0,31	+0,10	+0,12	+0,04	+0,09	+0,85
ξ_i	-0,19	+0,29	+0,11	+0,20	+0,10	-0,08	+0,95

Т а б л и ц а 20.1.2. Вторая координата

x_i	+0,02	-0,43	+0,74	+0,87	+0,40	+0,32	-0,09
ξ_i	+0,16	-0,46	+0,83	+1,00	+0,57	+0,35	-0,27

Решение. Рассмотрим вопрос о согласии вычисленных значений координат с наблюдаемыми для первой координаты (см. табл. 20.1.1).

Будем предполагать, что вычисленные (предсказанные теорией) значения x координат и наблюдаемые y связаны линейной зависимостью

$$y = \alpha + \beta(x - \bar{x}),$$

где

$$\bar{x} = \frac{1}{7} \sum_{i=1}^7 x_i = 0,184.$$

По результатам наблюдений (см. табл. 20.1.1) получим оценки неизвестных параметров регрессии:

$$\hat{\alpha} = \bar{\xi} = 0,197; \hat{\beta} = 1,081, \hat{\sigma}^2 = 0,006, S_1^2 = 0,0948$$

(см. теорему 20.1.1).

Сначала проверим гипотезу $H_{0;1} : \beta = 0$, неотклонение которой свидетельствует об отсутствии какой-либо связи между теорией и экспериментом.

Согласно (20.1.6),

$$\left| \frac{\hat{\beta}}{\hat{\sigma}/(S_1\sqrt{n-2})} \right| =$$

$$= \left| \frac{1,081}{\sqrt{0,006/(0,0948 \cdot 5)}} \right| = 9,61 > t_{0,01;5} = 3,365.$$

Так что обидная для Эйнштейна гипотеза об отсутствии связи теории с опытом отклоняется.

Далее проверим гипотезу $H_{0;2} : \beta = \beta_0 = 1$.

Согласно (20.1.5),

$$\begin{aligned} & \left| \frac{\hat{\beta} - \beta_0}{\hat{\sigma}/(S_1 \sqrt{n-2})} \right| = \\ &= \left| \frac{1,081 - 1}{\sqrt{0,006/(0,0948 \cdot 5)}} \right| = 0,72 < t_{0,01;5} = 3,365. \end{aligned}$$

И, следовательно, гипотеза $H_{0;2} : \beta = \beta_0 = 1$ не отклоняется. Последнее можно трактовать как согласие опыта с экспериментом, но при этом возможна систематическая ошибка (если $\alpha \neq \bar{x}$).

В согласии (или несогласии) теории с опытом можно убедиться, проверяя гипотезу

$$H_0 : \beta = 1; \alpha = 0,184.$$

Согласно (20.1.10)

$$\begin{aligned} & \frac{((\hat{\alpha} - \alpha)^2 + S_1^2(\hat{\beta} - \beta)^2)/2}{\hat{\sigma}^2/(n-2)} = \\ &= \frac{((0,197 - 0,184)^2 + 0,0948(1,081 - 1)^2)/2}{0,006/5} = \\ &= 0,33 < F_{0,01;2;5} = 13,3. \end{aligned}$$

Поэтому можно считать, что $\beta = 1; \alpha = 0,184$, а, следовательно, зависимость y от x имеет вид

$$y = x,$$

подтверждающий согласие теории с опытом.

Пример 20.1.2. В “Основах химии” Д. И. Менделеева приведены данные о растворимости азотно-кислого натрия NaNO_3 в зависимости от температуры воды. В таблице приведены данные о количестве условных частей NaNO_3 , которые растворяются в 100 частях воды при соответствующих температурах (x — температура в градусах, ξ — растворимость в условных частях на 100 частей воды).

x	ξ	x	ξ
0	66,7	29	92,9
4	71,0	36	99,4
10	76,3	51	113,6
15	80,6	68	125,1
21	85,7		

Теоретические соображения дают основания предполагать, что зависимость количества растворенного вещества от температуры является линейной:

$$\xi = \alpha + \beta(x - \bar{x}).$$

Найти оценки максимального правдоподобия параметров α , β и дисперсии σ^2 . Построить доверительные интервалы для α , β , σ^2 . Проверить гипотезу $H_0 : \beta = 0$ о значимости регрессии.

Решение.

$$\bar{x} = \frac{1}{9} \sum_{i=1}^9 x_i = \frac{234}{9} = 26; \quad \bar{\xi} = \frac{1}{9} \sum_{i=1}^9 \xi_i = \frac{811,3}{9} = 90,14,$$

$$S_1^2 = \frac{1}{9} \sum_{i=1}^9 (x_i - \bar{x})^2 = \frac{4060}{9} = 451,11,$$

$$S_2^2 = \frac{1}{9} \sum_{i=1}^9 (\xi_i - \bar{\xi})^2 = \frac{3083,98}{9} = 342,66,$$

$$R_{12} = \frac{1}{9} \sum_{i=1}^9 (x_i - \bar{x})(\xi_i - \bar{\xi}) = \frac{3534,8}{9} = 392,75.$$

Оценки максимального правдоподобия коэффициентов α , β простой линейной регрессии:

$$\hat{\alpha} = \bar{\xi} = 90,14; \quad \hat{\beta} = \frac{R_{12}}{S_1^2} = \frac{392,75}{451,11} = 0,87.$$

Оценка максимального правдоподобия дисперсии σ^2 :

$$\begin{aligned}\hat{\sigma}^2 &= S_2^2 \left(1 - \frac{R_{12}^2}{S_1^2 S_2^2} \right) = \\ &= 342,66 \left(1 - \frac{392,75^2}{451,11 \cdot 342,66} \right) = 0,72.\end{aligned}$$

Уравнение эмпирической линии регрессии

$$y = 90,14 + 0,87(x - 26).$$

Проверим гипотезу $H_0 : \beta = 0$ о значимости линии регрессии. Согласно (20.1.6)

$$\begin{aligned}\left| \frac{\hat{\beta}}{\hat{\sigma} / \sqrt{S_1^2(n-2)}} \right| &= \left| \frac{0,87}{\sqrt{0,7158/451,11(9-2)}} \right| = \\ &= 57,83 > t_{0,05;7} = 1,895,\end{aligned}$$

поэтому гипотеза $H_0 : \beta = 0$ отклоняется, линейная регрессия значимая.

Согласно (20.1.11), (20.1.12), (20.1.13) доверительными интервалами с коэффициентами надежности 0,90 для параметров регрессии являются: (89,54; 90,75) — для α ; (0,84; 0,90) — для параметра β ; (0,46; 2,97) — для параметра σ^2 .

20.2 Задачи

20.1 (тормозной путь). При изучении движения уличного транспорта фиксировалось расстояние s , пройденное автомобилем по инерции после сигнала “остановиться” (тормозной путь) в зависимости от скорости v .

Наблюдения проводились для различных автомобилей, с разными водителями, различным поверхностным покрытием дороги и т. д. Результаты наблюдений приведены в таблице, где s — тормозной путь в метрах, v — скорость автомобиля в км/час.

В предположении, что зависимость между тормозным путем автомобиля и скоростью линейная:

$$s = \alpha + \beta(v - \bar{v}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 ; построить доверительные интервалы для параметров α, β, σ^2 ; проверить гипотезу о значимости линейной регрессии. Проверить гипотезу об адекватности линейной регрессии.

v	s	v	s	v	s	v	s
6	0,6	19	7,3	26	9,8	32	14,6
	3,0		8,5		12,2		15,6
11	1,2	21	7,9	27	9,8		17,1
	6,7		10,4		12,2		19,5
13	4,9		10,4		15,2	35	20,1
14	3,0		14,0	29	12,8	37	16,5
16	5,5	23	7,9		17,1	39	21,3
	7,9		11,0		23,2		28,0
	10,4		18,3		25,6		28,4
18	5,2		24,4	31	11,0		36,6
	8,5	24	6,1		14,0	40	25,9
19	4,3		7,9		20,8		
	6,1		16,5	32	9,8		

На плоскости (v, s) изобразить исходные данные и построить эмпирическую линию регрессии.

20.2. Исследуется содержание аскорбиновой кислоты, сохранившейся в овощах в процессе их сушки и хранения.

Далее приведено содержание сухого вещества в свежем шпинате и содержание аскорбиновой кислоты, сохранившейся после сушки шпината при температуре 90°C (x — содержание сухого вещества в свежем шпинате в процентах; y — содержание аскорбиновой кислоты после сушки в процентах).

Имеются основания предполагать, что зависимость содержания сохранившейся аскорбиновой кислоты от содержания сухого вещества линейная:

$$y = \alpha + \beta(x - \bar{x}).$$

Найти оценки максимального правдоподобия параметров α, β простой линейной регрессии и дисперсии σ^2 , построить доверительные интервалы для параметров α, β , и σ^2 , проверить гипотезу о значимости регрессии, гипотезу об адекватности линии регрессии.

x	y	x	y	x	y
7,8	69,4	10,0	66,7	11,2	89,6
8,2	50,9		88,9		83,8
8,9	74,0	10,1	76,0		67,9
	58,6	10,2	77,2	11,8	79,9
9,0	66,4	10,3	69,8	12,3	83,1
9,2	80,6	10,7	69,0	12,5	74,2
9,5	61,9	10,8	65,2	12,9	86,0
10,0	70,9	11,1	77,2	14,9	88,2

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.3. В таблице приведены результаты экспериментального исследования количества тепла, выделяемого в процессе затвердения портландского цемента в зависимости от состава клинкеров (x — содержание в клинкерах алюмината кальция $3CaO \cdot Al_2O_3$ в процентах от веса клинкеров; y — количество тепла, выделенного на протяжении 180 дней, в калориях на грамм цемента).

x	y	x	y	x	y
1	74,3	7	78,5	11	109,2
	72,5		95,9		113,3
	83,8	10	109,4	21	115,9
2	93,1	11	104,3		
3	102,7		87,6		

Предполагая, что количество тепла является линейной функцией содержания алюмината кальция:

$$y = \alpha + \beta(x - \bar{x}),$$

найти оценки максимального правдоподобия параметров α, β простой линейной регрессии и дисперсии σ^2 . Построить доверительные интервалы для параметров α, β, σ^2 , проверить гипотезу о значимости регрессии, гипотезу об адекватности линейной регрессии.

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.4. Образцы медно-никелевого сплава, каждый с определенным содержанием железа, были испытаны на коррозионную устойчивость в установке с колесом. Колесо вращалось в соленой морской воде со скоростью 30 футов в секунду на протяжении 60 дней. Коррозия определялась по потере веса сплава (x — содержание железа в сплаве; y — потеря веса сплава, мг/см^2).

Предполагая, что зависимость потери веса образца от содержания железа в сплаве линейная:

$$y = \alpha + \beta(x - \bar{x}),$$

а результаты измерений потерь веса — независимые нормально распределенные случайные величины с дисперсией σ^2 , найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для параметров α, β, σ^2 , проверить гипотезу о значимости регрессии, гипотезу об адекватности линейной регрессии.

x	y	x	y	x	y
0,01	127,6	0,71	110,8	1,41	91,4
	130,1		113,1		83,7
	128,0	0,95	103,9	1,96	86,2
0,48	124,0	1,19	101,5		
	122,0	1,44	92,3		

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.5 (потребление вина и смерть от сердечного приступа). Полезно ли вино для здоровья? Имеются данные, свидетельствующие о том, что употребление вина в умеренных количествах способствует предотвращению сердечных приступов.

В таблице приведены данные о годовом потреблении вина на человека (в литрах l алкоголя, выпитого с вином) и количестве n смертей в год от сердечных заболеваний (n — число смертей на 100 000 человек) в 19 развитых странах.

Предполагая, что зависимость между количеством n смертей от сердечных приступов и потреблением l вина линейная:

$$n = \alpha + \beta(l - \bar{l}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для α, β, σ^2 . Проверить гипотезу о значимости регрессии.

Страна	l	n	Страна	l	n
Австралия	2,5	211	Нидерланды	1,8	167
Австрия	3,9	167	Новая Зеландия	1,9	266
Бельгия	2,9	131	Норвегия	0,8	227
Канада	2,4	191	Испания	6,5	86
Дания	2,9	220	Швеция	1,6	207
Финляндия	0,8	297	Швейцария	5,8	115
Франция	9,1	71	Великобритания	1,3	285
Исландия	0,8	211	США	1,2	199
Ирландия	0,7	300	Германия	2,7	172
Италия	7,9	107			

На плоскости (l, n) изобразить исходные данные и построить эмпирическую линию регрессии.

20.6 (остеопения и количество выводимого из организма кальция). Потеря кальция в костной ткани человека — остеопения 1, 2, 3, 4 степени в своих тяжелых формах (4-я степень остеопении — остеопороз) ведет к разрушению костной ткани. В настоящее время остеопороз приобретает формы широко распространенного заболевания. Степень потери кальция в костной ткани человека определяют при помощи ультразвуковой денситометрии, которая мало доступна как по причине ее дороговизны, так и недостаточной обеспеченности соответствующей аппаратурой. Предлагается метод диагностирования степени остеопении — по содержанию кальция, выводимого с мочой. Обозначим через T результат денситометрического исследования, характеризующего плотность костной ткани и содержание кальция в ней. Если значение T принадлежит промежутку $[-i - 1; -i)$, то состояние костной ткани определяется как i -я степень остеопении, $i = 1, 2, 3, 4$. При значении T из промежутка $[-2; -1)$ степень остеопении классифицируется как средняя, при $T \leq -2$ — как тяжелая. Через Ca обозначим

количество кальция, выводимого с мочой. Исходя из данных Института гастроэнтерологии Академии медицинских наук Украины (см. таблицу), убедиться, что между величинами T и Ca существует линейная зависимость — постройте линейную регрессию T на Ca . Проверьте гипотезы о значимости и адекватности линейной регрессии. По известной зависимости T от Ca дифференцируйте тяжелую форму остеопении (остеопороз) по количеству выводимого Ca .

Ca	T	Ca	T	Ca	T	Ca	T
2,2	0,23	3,12	-1,63	3,84	-2,1	4,4	-1,77
	0,6	3,2	-0,21	4,1	-1,51	4,63	-1,61
2,24	-1,21		2,06	4,12	-1,86	4,8	-1,04
2,32	-0,04		-1,2		-0,65	5,8	-2,13
2,5	0,14		-1,49	4,2	-0,8	5,84	-2,94
2,6	0,49	3,35	-0,9		-0,6	5,9	-2,34
	0,07	3,5	-1,4		0,54	6,1	-2,56
2,62	-0,02		0,23		-2,23	6,7	-1,62
2,64	-0,27	3,6	-0,15	4,24	0,15	6,8	-2,67
2,67	-1,62		-0,2		0,0	6,82	-2,96
2,76	-0,05		-0,2	4,3	-1,43	7,82	-4,08
2,8	-1,84		1,0	4,32	-1,8		-3,61
	-2,04	3,64	-1,46	4,4	-1,61		-2,91

На плоскости (Ca, T) изобразить исходные данные и построить эмпирическую линию регрессии.

20.7. В таблице приведено содержание в процентах крахмала и протеина в зернах пшеницы, которым характеризуется качество пшеницы (y — содержание протеина, x — содержание крахмала). Определение протеина требует сложного химического анализа, тогда как содержание крахмала определять просто.

y	x	y	x	y	x
10,3	60	10,8	45	14,4	85
12,2	75	11,4	51	15,8	96
14,5	87	11,0	17	15,6	92
11,1	55	10,2	36	15,0	94
10,9	34	17,0	97	13,3	84
18,1	98	13,8	74	19,0	99
14,0	91	10,1	24		

Предполагая, что зависимость содержания протеина от содержания крахмала линейная:

$$y = \alpha + \beta(x - \bar{x}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для параметров α, β, σ^2 . Проверить гипотезу о значимости регрессии.

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.8. В таблице приведены результаты измерения скоростей v (в км/сек) и расстояний s до Земли (в миллионах парсек) для десяти внегалактических туманностей.

v	s	v	s
1,20	630	9,12	4820
1,82	890	10,97	5230
3,31	2350	14,45	7500
7,24	3810	22,91	11800
8,92	4630	36,31	19600

Предполагая, что скорость галактики линейно зависит от расстояния:

$$v = \alpha + \beta(s - \bar{s}),$$

найти оценки параметров α, β и дисперсии σ^2 ; построить доверительные интервалы для параметров α, β, σ^2 . Проверить гипотезу $H_0 : \beta = 0$ о значимости регрессии.

На плоскости (s, v) изобразить исходные данные и построить эмпирическую линию регрессии.

20.9 (стоимость эксплуатации самолета и его “возраст”). Дирекция авиакомпании с целью планирования расходов хочет установить количественную зависимость стоимости эксплуатации самолета от продолжительности его эксплуатации (от “возраста” самолета). Со временем из-за старения деталей и узлов самолета компания несет большие затраты на поддержание его в рабочем состоянии, в частности, необходимо чаще проводить ремонтно-профилактические работы, заменять отдельные узлы.

В таблице приведены данные о стоимости эксплуатации самолетов и их “возрасте” (x — “возраст самолета” в годах; y — стоимость эксплуатации самолета в течение полугода в долларах).

Предполагая, что зависимость стоимости эксплуатации самолета от его “возраста” линейная, найти оценки

максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для параметров α, β, σ^2 . Проверить гипотезу о значимости линейной регрессии, гипотезу об адекватности линейной регрессии.

x	y	x	y
4,5	619	5,5	987
	1049	0,5	163
	1033		182
4,0	495	6,0	764
	723		1373
	681	1,0	978
5,0	890		466
	1522		549
	1194		

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.10. Ниже приведены результаты эксперимента по изучению нового метода измерения скорости крови в организме человека (x — скорость крови, найденная стандартным методом, y — с помощью нового метода).

x	y	x	y	x	y
1190	1115	1900	1830	2720	2630
1455	1425	1920	1920	2710	2740
1550	1515	1960	1970	2530	2390
1730	1795	2295	2300	2900	2800
1745	1715	2335	2280	2760	2630
1770	1710	2490	2520	3010	2970

Предполагая, что зависимость скорости крови, измеренной с помощью нового метода и стандартного, линейна:

$$y = \alpha + \beta(x - \bar{x}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для α, β, σ^2 . Проверить гипотезу о значимости регрессии.

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.11. Функциональной и структурной единицей нервной системы является нейрон (нервная клетка), составляющей которого является дендрит (он по форме напоминает дерево). Большинство характеристик дендрита нейрона — случайные величины, в частности, диаметр сегмента дендрита нейрона.

В таблице приведены результаты измерений (в мкм) диаметра y начала дочернего сегмента и диаметра x конца материнского сегмента. Предполагая, что зависимость y от x линейная:

$$y = \alpha + \beta(x - \bar{x}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для параметров α, β, σ^2 . Проверить гипотезу о значимости регрессии, гипотезу об адекватности линейной регрессии.

x	y	x	y	x	y	x	y
0,40	0,40	0,60	0,60	0,75	0,60	1,00	0,45
0,45	0,45		0,60		0,70		0,55
	0,45		0,60		0,75		0,75
	0,45	0,65	0,35		0,75		0,85
	0,45		0,45	0,80	0,40		0,85
	0,45		0,45		0,50		1,00
0,5	0,40		0,45		0,50	1,05	0,60
	0,40		0,50		0,50		0,65
	0,45		0,50		0,60		0,75
	0,45		0,50		0,60	1,10	0,90
	0,50		0,60		0,80	1,15	0,60
	0,50		0,60	0,85	0,50		0,65
	0,70		0,65		0,55	1,20	0,65
0,55	0,35		0,65		0,55	1,25	0,85
	0,40		0,65		0,60		1,00
	0,45		0,65		0,65	1,30	0,85
	0,45		0,65		0,65	1,35	0,90
	0,55	0,70	0,45		0,65		0,90
	0,55		0,45		0,85		1,00
	0,55		0,50	0,90	0,90		1,60
	0,55		0,50		0,90	1,40	0,75
	0,55		0,55	0,95	0,60	1,45	0,75
	0,55		0,65		0,60	1,50	0,90
0,60	0,50		0,80		0,80	1,60	0,95
	0,55	0,75	0,50		0,95	1,65	0,65
	0,60		0,50		1,05		

На плоскости (x, y) изобразить исходные данные и построить эмпирическую линию регрессии.

20.12. Проверить, согласуется ли теория с экспериментом по второй координате (см. пример 20.1.1 и данные таблицы 20.1.2).

Найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 , построить доверительные интервалы для параметров α, β, σ^2 .

20.13 (барометрическое давление и точка кипения воды). Исследовалась зависимость между барометрическим давлением и точкой кипения воды. Цель исследования — оценить высоту над уровнем моря по температуре кипения воды. В горных условиях для определения барометрического давления удобнее измерять температуру кипения воды. В таблице приведены данные об измерениях барометрического давления и точки кипения воды в 17 экспериментах.

Номер	Дав- ление	Точка кипения	Номер	Дав- ление	Точка кипения
1	20,79	194,5	10	24,01	201,3
2	20,79	194,3	11	25,14	203,6
3	22,40	197,9	12	26,57	204,6
4	22,67	198,4	13	28,49	209,5
5	23,15	199,4	14	27,76	208,6
6	23,35	199,9	15	29,04	210,7
7	23,89	200,9	16	29,88	211,9
8	23,99	201,1	17	30,66	212,2
9	24,02	201,4			

Предполагая, что зависимость барометрического давления от точки кипения воды линейная:

$$y = \alpha + \beta(x - \bar{x}),$$

найти оценки максимального правдоподобия параметров α, β и дисперсии σ^2 . Построить доверительные интервалы для α, β, σ^2 . Проверить гипотезу о значимости регрессии.

Глава 21

Решения, указания, ответы

21.1 К главе 1

1.1°. mn . 1.2°. 49; 42. 1.3°. $17 \cdot 16 \cdot 15$.

1.4°. $4 \cdot 5 \cdot 5$. 1.5°. $4 \cdot 4 \cdot 3$. 1.6°. 7!

1.7°. A_{10}^6 . 1.8°. $9 \cdot 10 \cdot 10 \cdot 10 \cdot 2$. 1.9°. $26^2 \cdot 10^4$.

1.10. $16 \cdot 15 \cdot 14 \cdot C_{13}^2$. 1.11°. C_9^4 . 1.12°. C_5^3 .

1.13°. $4!$ 1.14. $(\alpha_1 + 1)(\alpha_2 + 1) \dots (\alpha_n + 1)$.

1.15. $2 \cdot (n-1)!$ 1.16°. A_{25}^4 . 1.17°. A_8^4 . 1.18. $(n-2)!$

1.19. C_9^4 , указанные числа однозначно задаются множеством своих цифр.

1.20. C_{10}^4 .

1.21*. Заполняем все клетки таблицы, кроме клеток последней строки и последнего столбца (правого), числами $+1$ и -1 (согласно правилу умножения, это можно сделать $2^{(n-1)(m-1)}$ способами). Заполняем $n-1$ клеток последней строки, кроме последней (правой) так, чтобы произведение чисел в каждом столбце равнялось $+1$. Заполняем клетки последнего столбца так, чтобы произведение чисел в каждой из $m-1$ первых строк и в последнем столбце равнялось $+1$. Следовательно, произведение чисел в каждом столбце будет равно $+1$ и в каждом из $m-1$ первых строк будет равно $+1$, но тогда произведение чисел и в последней строке будет равно $+1$, поскольку

в таблице произведение чисел по строкам равно произведению по столбцам.

Ответ: $2^{(n-1)(m-1)}$.

1.22. Из $p + 1$ промежутков между белыми шарами (учитывая два крайние бесконечные промежутка) выберем q , в которых расположим черные шары. Это можно сделать C_{p+1}^q способами.

1.23 (3). Последовательно проводим на плоскости прямые. Пусть проведено $n - 1$ прямых и A_{n-1} — число частей плоскости, которые при этом образовались. Проведем n -ю прямую. К A_{n-1} частям n -я прямая добавит еще n частей, поэтому

$$A_n = A_{n-1} + n.$$

Подробнее: $A_1 = 2$, $A_2 = A_1 + 2, \dots, A_n = A_{n-1} + n, \dots$. Складывая почленно первые n равенств, получим

$$A_n = \frac{n(n+1)}{2} + 1.$$

Ответы: 1) C_n^2 ; 2) C_n^3 ; 3) $n(n+1)/2+1$; 4) $2n$ — количество неограниченных частей, $n(n-3)/2+1$ — количество ограниченных.

1.24. $n(n-3)/2$.

1.25. Каждая точка пересечения диагоналей задает четыре вершины n -угольника и наоборот, поэтому число точек пересечения диагоналей равно C_n^4 .

1.26*. Всего диагоналей $N = n(n-3)/2$ (см. задачу 1.24). Пусть $k-1$ первых диагоналей делят многоугольник на A_{k-1} частей. Проведем k -ю диагональ. Обозначим через p_k число точек пересечения k -й диагонали с первыми $k-1$ диагоналями. Тогда

$$A_k = A_{k-1} + (p_k + 1).$$

Ответ: $1 + C_n^4 + n(n-3)/2$.

1.27. Воспользоваться методом математической индукции.

1.28. Записать элементы в список и приписать каждому из них номер места в списке.

1.29. Воспользоваться правилом умножения.

1.30. С одной стороны, n -элементное множество можно упорядочить $n!$ способами. С другой — $C_n^k k!(n-k)!$ способами — сначала разбить его на два непересекающихся подмножества (C_n^k способами), а затем упорядочить каждое из них (соответственно $k!$ и $(n-k)!$ способами). Поэтому $n! = C_n^k k!(n-k)!$

1.32. С одной стороны, n -элементное множество можно упорядочить $n!$ способами. С другой стороны —

$$C_n(k_1, k_2, \dots, k_m) k_1! k_2! \dots k_m!$$

способами — сначала разбить его на m непересекающихся подмножеств, содержащих соответственно $k_1, k_2, \dots, k_m$ элементов (это можно сделать $C_n(k_1, k_2, \dots, k_m)$ способами), а затем упорядочим каждое из них соответственно $k_1!, k_2!, \dots, k_m!$ числом способов. Поэтому

$$n! = C_n(k_1, k_2, \dots, k_m) k_1! k_2! \dots k_m!$$

1.33. Слово задается выбором мест под букву каждого типа (разбиением множества на подмножества).

1.34. $C_{m+n+s}(m, n, s)$. **1.35.** $C_{3n}(n, n, n)$.

1.36. Слову длиной n из r_1 букв a_1 , r_2 букв a_2 , ..., r_k букв a_k (а таких слов $n!/(r_1! r_2! \dots r_k!)$) соответствует произведение $a_1^{r_1} a_2^{r_2} \dots a_k^{r_k}$.

1.37. Каждому сочетанию с повторениями из m элементов по n можно поставить в соответствие последовательность из n нулей, разделенных $m-1$ единицами (причем число нулей между единицами равно числу элементов данного типа) и наоборот.

21.2 К главе 2

2.1°. а) $\bar{A}$ — герб не выпал; б) $\bar{B}$ — хотя бы один промах в трех выстрелах в) $\bar{C}$ — ни одного попадания в трех выстрелах.

2.2°. а) $A = A_1 \cap A_2 \cap A_3$; б) $B = \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3$; в) $C = (A_1 \cap \bar{A}_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap A_2 \cap \bar{A}_3) \cup (\bar{A}_1 \cap \bar{A}_2 \cap A_3)$;

г) $D = (A_1 \cap A_2 \cap \bar{A}_3) \cup (A_1 \cap \bar{A}_2 \cap A_3) \cup (\bar{A}_1 \cap A_2 \cap A_3) \cup (A_1 \cap A_2 \cap A_3)$.

2.3°. а) $A \cap \bar{B} \cap \bar{C}$; б) $A \cap B \cap \bar{C}$; в) $A \cap B \cap C$;
г) $A \cup B \cup C$; д) $(A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)$;
е) $\bar{A} \cap B \cap \bar{C}$.

2.4.

$$A = \bigcap_{i=1}^n \bar{A}_i, \quad B = \bigcup_{i=1}^n A_i,$$

$$C = (A_1 \cap \bar{A}_2 \cap \bar{A}_3 \cap \dots \cap \bar{A}_n) \cup (\bar{A}_1 \cap A_2 \cap \bar{A}_3 \cap \dots \cap \bar{A}_n) \cup \dots \cup (\bar{A}_1 \cap \bar{A}_2 \cap A_3 \cap \dots \cap A_n).$$

2.5.

$$A = \bigcup_{j=1}^m \bigcup_{i=1}^n A_{ji}, \quad B = \bigcap_{j=1}^m \bigcup_{i=1}^n A_{ji}.$$

2.7°. Пространство Ω — множество упорядоченных пар (i, j) , $i, j = 1, 2, \dots, 6$, согласно правилу умножения $n(\Omega) = 6 \cdot 6 = 36$; $A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$, $n(A) = 5$; B — множество пар (i, j) , в которых встречается 6, $n(B) = 11$; C — множество пар (i, j) , в которых i принимает значения 2, 4, 6, $n(C) = 18$; $D = \{(i, j) : i, j = 1, 3, 5\}$, $n(D) = 9$.

2.11. Пространство Ω состоит из n -элементных подмножеств N -элементного множества изделий; в событие A входят те из них, которые содержат ровно m бракованных изделий; $n(\Omega) = C_N^n$; $n(A) = C_M^m C_{N-M}^{n-m}$.

2.12. Пространство Ω состоит из последовательностей длиной 7, образованных из 10 чисел (номеров этажей, на которых останавливается лифт); в событие A входят последовательности, состоящие из различных чисел; $n(\Omega) = 10^7$; $n(A) = 10 \cdot 9 \dots 4$.

2.13. $\Omega = \{\Gamma, \text{РГ}, \text{РРГ}, \dots\}$; $A = \{\underbrace{\text{РР} \dots \text{РГ}}_k\}$, $B = \{\Gamma, \text{РГ}, \text{РРГ}, \dots, \underbrace{\text{РР} \dots \text{РГ}}_{k-1}\}$; $C = \{\Gamma, \text{РРГ}, \text{РРРРГ}, \dots\}$;
 $D = \{\text{РГ}, \text{РРРГ}, \text{РРРРРГ}, \dots\}$.

2.14. Можно предложить по меньшей мере два пространства элементарных событий этого стохастического эксперимента. Если обозначить белый шар через W , черный — через B , то множеством исходов стохастического эксперимента будет $\Omega = \{WW, WB, BW\}$.

Если белый шар, который кладут в урну, обозначить, например, через W^* , то $\Omega^* = \{WW^*, W^*W, BW^*, W^*B\}$.

2.15. Ω состоит из слов длиной n , образованных из “букв” $1, 2, \dots, 6$; $n(\Omega) = 6^n$. Событие A описывается словами длиной n , образованными из n_1 букв 1 , n_2 букв 2 , $\dots$, n_6 букв 6 ; $n(A) = n!/(n_1!n_2!\dots n_6!)$.

2.16. $n(\Omega) = C_{2n}^n$; $n(A) = C_{2n-2}^{n-1}C_2^1$; $n(B) = C_{2n-4}^{n-2}C_4^2$. См. также задачу 3.3. и 3.11.

2.17. $n(\Omega) = 12^r$; $n(A) = 12^r - 12 \cdot 11 \dots (12 - (r-1))$; $n(B) = r11^{r-1}$; $n(C) = 12^r - 11^r$; $n(D) = 11^r$.

2.18. $n(\Omega) = n!$; $n(A) = (n-1)!$; $n(B) = (n-2)!$

2.19. Ω — множество упорядоченных троек, составленных из чисел от 1 до 6; A — подмножество упорядоченных троек, в которых 1 встречается один раз; B — в которых 6 встречается ровно два раза; C — составлено из различных чисел; $D = \{(1; 1; 1), (2; 2; 2), \dots, (6; 6; 6)\}$; $n(\Omega) = 6^3$; $n(A) = C_3^1 5^2$; $n(B) = C_3^2 5$; $n(C) = 6 \cdot 5 \cdot 4$; $n(D) = 6$.

2.20. Ω — множество всех бесконечных последовательностей, образованных буквами Γ , P на нечетных местах и цифрами $1, 2, \dots, 6$ на четных. Событие A описывается последовательностями из Ω , которые записываются до первого появления буквы Γ буквами P на нечетных местах и цифрами $1, 2, \dots, 5$ — на четных. Событие B описывается последовательностями из Ω , которые записываются до первого появления цифры 5 буквами P на нечетных местах и цифрами $1, 2, 3, 4, 6$ — на четных.

2.21*. Каждой частице поставим в соответствие номер ячейки, в которой она оказалась. Ω — множество последовательностей, длиной n , составленных из чисел $1, 2, \dots, m$. Событие A описывается последовательностями длиной n , составленными из k_1 единиц, k_2 двоек, $\dots$, k_m чисел m .

2.22*. Ω — множество последовательностей $(x_1, x_2, \dots, x_m)$, составленных из неотрицательных чисел, таких,

что

$$x_1 + x_2 + \dots + x_m = n.$$

Событие A описывается последовательностями $(x_1, x_2, \dots, x_m)$, образованными положительными целыми числами.

21.3 К главе 3

3.1. $P(A) = 1 - P(\bar{A}) = 1 - A_{365}^r / 365^r$.

3.2. $P(A) = 1 - A_{12}^r / 12^r$, если $r \leq 12$, и $P(A) = 1$, если $r > 12$.

3.3. Мы будем различать подгруппы: подгруппа 1° и подгруппа 2° . Исход эксперимента — упорядоченная пара n -элементных подмножеств $2n$ -элементного множества (пространство элементарных событий — множество всех таких пар). При этом выбор одного из n -элементных подмножеств пары (одной из подгрупп, например подгруппы 1°), определяет состав другого подмножества (подгруппы 2°). Поэтому исходов всего C_{2n}^n . Модель классическая.

Событие A — “две наиболее сильные команды оказались в разных подгруппах” — описывается упорядоченными парами n -элементных подмножеств, в состав каждого из которых входит ровно одна из наиболее сильных команд. Поэтому

$$P(A) = C_2^1 C_{2n-2}^{n-1} / C_{2n}^n.$$

Событие B — “две наиболее сильные команды оказались в одной подгруппе” описывается упорядоченными парами n -элементных подмножеств такими, что в состав одного n -элементного подмножества входят две наиболее сильные команды, в состав другого — ни одной. Поэтому сначала выберем подгруппу (1° или 2°), в состав которой войдут две наиболее сильные команды (C_2^1 способами), а затем дополним ее до полного состава C_{2n-2}^{n-2} способами. Отсюда

$$P(B) = C_2^1 C_{2n-2}^{n-2} / C_{2n}^n.$$

3.4°. $1/5!$ **3.5.** $2/n$.

3.6. 1) C_{20}^3/C_{25}^3 ; 2) $C_{20}^2 C_5^1/C_{25}^3$; 3) $C_{20}^2 C_5^1/C_{25}^3 + C_{20}^3/C_{25}^3$;
4) $1 - C_{20}^3/C_{25}^3 - C_{20}^2 C_5^1/C_{25}^3$.

3.7. $1/2^6$.

3.9. Решение 1. Будем считать буквы различимыми, например, $A_1, A_2, A_3, M_1, M_2, T_1, T_2$. Исход эксперимента — перестановка из 10 различных букв. Модель классическая. Событие A — получилось слово “математика” — образуют перестановки, в которых буквы стоят на своих местах. Таких перестановок $3!2!2!$. Поэтому $P(A) = 3!2!2!/10!$

Решение 2. Исход эксперимента — слово длиной 10, составленное из букв $A, A, A, E, И, К, М, М, Т, Т$. Таких слов $10!/3!2!2!$. Событие A состоит из одного слова, поэтому $P(A) = 3!2!2!/10!$

3.10. Исход эксперимента — перестановка из n элементов. Событие “между A и B расположены r человек” описывается перестановками, в которых между элементами A и B размещено r элементов. Число таких перестановок равно $(n - r - 1)2!(n - 2)!$. Искомая вероятность равна $2(n - r - 1)/(n(n - 1))$.

3.11. а) $C_2^1 C_{18}^9/C_{20}^{10}$; б) $C_4^2 C_{16}^8/C_{20}^{10}$.

3.12. $(k - 1)(n - k)/C_n^2$.

3.13. Каждому пассажиру припишем номер вагона, в который он садится, ω — последовательность длиной 9, составленная из чисел 1, 2, 3 (слово длиной 9, образованное “буквами” 1, 2, 3); модель классическая.

Событие “в каждый вагон сядут по три пассажира” описывается словами длиной 9, образованными 3 единицами, 3 двойками, 3 тройками; таких слов $9!/(3!3!3!)$. Поэтому искомая вероятность равна $\frac{9!/(3!3!3!)}{3^9}$.

Событие “в один вагон сядут четыре, второй — три, третий — два пассажира” описывается последовательностями (словами) длиной 9, образованными цифрами (буквами) 1, 2, 3, причем одна цифра встречается четыре, другая три, третья два раза. Подсчитаем число таких слов. Первым действием выбираем букву (цифру), которая встречается четыре раза (тремя способами), затем ту, которая встречается три раза (двумя способами); встреча-

ующуюся два раза, может быть выбрана одним способом. Вторым действием из выбранных букв (цифр) образуем слово длиной 9 с заданным числом букв ($9!/(4!3!2!)$ способами). Поэтому описанных выше слов, согласно правилу умножения, $3! \cdot 9!/(4!3!2!)$. Следовательно, искомая вероятность равна

$$\frac{3!9!/(4!3!2!)}{3^9}.$$

3.14. $A_{10}^7/10^7$, что приближенно равно 0,06.

3.15. $12!/12^{12}$. **3.16.** $\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \sum_{k=m+1}^{\min\{M,n\}} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$.

3.17. $\sum_{s=1}^{\min\{m,r\}} \frac{C_m^s C_{n-m}^{r-s}}{C_n^r}$, если $r \leq n - m$, и 1, если $r > n - m$.

3.18. $\sum_{s=r}^{\min\{n,k\}} \frac{C_n^s C_{N-n}^{k-s}}{C_N^k}$.

3.19. $A_r = \{\text{участник угадал } r \text{ видов}\}$, $r = 3, 4, 5, 6$; $P(A_3) = 0,017650$, $P(A_4) = 0,000969$, $P(A_5) = 0,000018$, $P(A_6) = 0,00000007151$, $P\{\text{участник получил выигрыш}\} = P(A_3 \cup A_4 \cup A_5 \cup A_6) = P(A_3) + P(A_4) + P(A_5) + P(A_6) = 0,01863707115$.

3.20. $12!/(2^6 6^{12})$, что составляет приблизительно 0,0034.

3.21. $n!/(6^n n_1! n_2! \dots n_6!)$. **3.22°.** $1/4!$

3.24°. $P(A) = 3/8$, $P(B) = 7/8$, $P(A \cap B) = 3/8$, $P(B/A) = 1$.

3.25. $1/6^{n-1}$. **3.26°.** $1/6^5$.

3.27. Если лица садятся в ряд, искомая вероятность равна $2/n$.

Способы размещения лиц за круглым столом отличаются взаимным размещением лиц между собой.

Сначала размещаем за круглым столом первое из двух указанных лиц (не имеет значения на каком месте — стол круглый). Далее перенумеруем оставшиеся места, например против часовой стрелки, и расположим на них остальные $(n-1)$ лиц, $(n-1)!$ способами. При этом число способов, когда два указанные лица будут сидеть рядом, равно

$2!(n-2)!$ Поэтому искомая вероятность равна $2!(n-2)!/(n-1)! = 2/(n-1)$.

3.28*. Вероятность появления единицы хотя бы один раз в четырех подбрасываниях равна $1 - (5/6)^4$, а появление пары (1, 1) хотя бы один раз в 24 подбрасываниях пары кубиков равно $1 - (35/36)^{24}$. Необходимо доказать, что $1 - (5/6)^4 > 1 - (35/36)^{24}$. Для этого достаточно установить, что $(35/36)^{24} > (5/6)^4$ или $(35/36)^6 > 5/6$. Последнее следует, например, из неравенства $(1+x)^n > 1+nx$ (где $x > -1, n \geq 1$) для $n=6, x=-1/36$.

3.29. Воспользоваться тем, что

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

3.31. Каждой из n частиц припишем номер ячейки, в которую она попала. Исход ω стохастического эксперимента — последовательность (слово) длиной n , составленная из чисел $1, 2, \dots, m$. Согласно правилу умножения исходов всего — m^n . Поскольку относительно ячеек все частицы находятся в одинаковых условиях, то каждому исходу естественно приписать одну и ту же вероятность (модель классическая). Событие “в первой, второй, ..., m -й ячейках будет соответственно $k_1, k_2, \dots, k_m$ частиц” образуют слова длиной n , составленные из k_1 единиц, k_2 двоек, ..., k_m чисел m . Всего таких слов $n!/(k_1!k_2! \dots k_m!)$. И поскольку модель классическая, то искомая вероятность равна $\frac{n!}{k_1!k_2! \dots k_m! m^n}$.

3.32. $\frac{6}{(n-2)(n-1)}$ (см. также решение к задаче 3.27).

3.36. Вероятность того, что четырехзначный номер состоит из разных цифр, равна $\frac{10 \cdot 9 \cdot 8 \cdot 7}{10^4}$.

Вероятность того, что четырехзначный номер имеет только две одинаковые цифры, равна $\frac{10C_9^2 4!/(2!1!1!)}{10^4}$.

Четырехзначный автомобильный номер — упорядоченная последовательность длиной 4 (слово), образованное цифрами $0, 1, 2, \dots, 9$; всего таких последовательностей 10^4 . Номера, имеющие две пары одинаковых цифр,

— это последовательности длиной 4 (слова), образованные двумя разными цифрами, которые встречаются по два раза, их всего $C_{10}^2 4!/(2!2!)$. (Сначала из 10 цифр выбираем две (C_{10}^2 способами), а затем из них образуем слово длиной 4, в которое каждая буква входит дважды — $4!/(2!2!)$ способами.) Поэтому вероятность того, что четырехзначный номер имеет две пары одинаковых цифр, равна $\frac{C_{10}^2 4!/(2!2!)}{10^4}$ (см. также задачу 3.13).

Вероятность того, что четырехзначный номер имеет три одинаковые цифры, равна $\frac{10 \cdot 9 \cdot 4!/(3!1!)}{10^4}$.

Вероятность того, что четырехзначный номер состоит из одинаковых цифр, равна $1/10^3$.

3.38. Исход ω — упорядоченная пара чисел $(i, j), i \neq j; i, j \leq N; n(\Omega) = N(N-1)$. Событие $A = \{\xi_1 < \xi_2\}$ описывается множеством пар (i, j) , для которых $i < j; n(A)$ вычисляем так: если первое число равно k , то второе можно выбрать $(N - k)$ способами, поэтому

$$n(A) = \sum_{k=1}^{N-1} (N - k) = \frac{1}{2} N(N - 1).$$

Следовательно, $P\{\xi_1 < \xi_2\} = 1/2$.

Рассмотрим задачу для трех чисел. Пусть k — первое число ($1 \leq k \leq N - 2$), тогда два других необходимо выбрать из множества чисел $\{k + 1, \dots, N\}$ так, чтобы $\xi_2 < \xi_3$, а это рассмотрено выше.

3.40. Вероятность того, что автомобиль находится за данной (выбранной вами) дверью, равна $1/3$, а следовательно, вероятность того, что за данной дверью автомобиля нет, т. е. автомобиль находится за одной из двух других дверей, равна $2/3$. Поэтому следует отказаться от первоначального выбора двери и выбрать оставшуюся дверь (одну дверь ведущий уже открыл и за ней автомобиля не оказалось). (Заметим, что открывает ведущий дверь или нет, вероятность того, что автомобиль находится за оставшимися, не выбранными Вами, дверями равна $2/3$.)

Предложенное решение особенно “прозрачно”, если автомобиль находится за одной из n дверей (например, $n = 1000$).

Вероятность автомобилю находиться за выбранной вами дверью равна $1/n$, и следовательно, вероятность противоположного события — автомобиль находится за одной из $n - 1$ оставшихся дверей, равна $1 - 1/n$. При $n = 1000$ эта вероятность составляет 0,999. Так что открыв оставшуюся дверь, вы почти наверняка уедете домой на автомобиле.

21.4 К главе 4

4.2°.

а) $\prod_{i=1}^n (1 - p_i)$; б) $1 - \prod_{i=1}^n (1 - p_i)$; в) $\sum_{i=1}^n p_i \prod_{k=1, k \neq i}^n (1 - p_k)$.

4.3°. Обозначим через A_i событие “объект обнаружен в i -м цикле”, $i = 1, 2, \dots, n$; события A_i , $i = 1, 2, \dots, n$, независимы. Событие C — “объект обнаружен” — через события $A_1, A_2, \dots, A_n$ можно выразить так: $C = \bigcup_{i=1}^n A_i$;

$$P(C) = 1 - P(\overline{C}) = 1 - P\left(\bigcap_{i=1}^n \overline{A}_i\right) = 1 - (1 - p)^n.$$

4.4°. A_i — событие “ i -й блок работает”, $i = 1, 2, \dots, n$; события A_i независимы. $A = \bigcap_{i=1}^n A_i$ — событие “прибор работает”; $P(A) = P\left(\bigcap_{i=1}^n A_i\right) = p^n$ (см. так же задачу 4.6).

4.5. 1) $1 - (1 - p)^2$; 2) $1 - (1 - p)(1 - pp_1)$.

4.6. 1) $1 - (1 - p)^n$;

2) A_1 — “прибор работает”; A_i — “работает i -й дублирующий прибор”, $i = 2, 3, \dots, n$; B_i — “работает прибор включения i -го дублирующего прибора”, $i = 2, 3, \dots, n$ (все события независимы); $\overline{C}$ — “система не работает”,

$$\overline{C} = \overline{A}_1 \bigcap_{i=2}^n \left(\overline{B}_i \cup (B_i \cap \overline{A}_i) \right);$$

$$P(C) = 1 - P(\overline{C}) = 1 - (1 - p)(1 - pp_1)^{n-1};$$

Чтобы надежность системы была не меньшей P , число дублирующих приборов должно быть не меньше

$$\ln(1 - P) / \ln(1 - p).$$

$$4.7^\circ. \text{ а) } 1/3; \text{ б) } 3/11. \quad 4.8^\circ. 1/2.$$

$$4.9^\circ. \frac{1}{N} \sum_{i=1}^N \frac{m_i}{n_i}. \quad 4.10^\circ. \frac{m_1 + n_1 m_2}{n_1(n_2 + 1)}.$$

$$4.11^\circ. 4/11.$$

$$4.12. 1 - (1 - pp_1 - (1 - p)p_0)^n.$$

$$4.13. 1 - (1 - pp_2 - (1 - p)p_1)^n. \quad 4.14^\circ. \frac{n+2}{2(n+1)}.$$

4.15°. Вероятность предположения “в урне k белых шаров” равна $2k/(n(n+1))$.

4.17. Вероятнее, что стрелок C в мишень попал.

$$4.18. 48/95. \quad 4.19. 0,52. \quad 4.20. \frac{np_2}{p_1 + np_2}.$$

4.21*. Пометим одну из коробок, например, звездочкой. Поскольку коробки берут наудачу, то вероятность взять меченую коробку равна $1/2$.

Если одна из коробок оказалась пустой, а другая содержит r спичек, то коробки брали $2N - r$ раз.

Рассмотрим $2N - r$ испытаний Бернулли (коробки брали $2N - r$ раз) с вероятностью успеха в одном испытании $1/2$ (успех — взяли меченую коробку). Если пустой оказалась меченая коробка, то вероятность N успехов в $2N - r$ испытаниях равна $C_{2N-r}^N / 2^{2N-r}$, а если пустой оказалась не меченая коробка, то вероятность $N - r$ успехов в $2N - r$ испытаниях

$$C_{2N-r}^{N-r} / 2^{2N-r} = C_{2N-r}^N / 2^{2N-r}.$$

Следовательно, искомая вероятность равна

$$C_{2N-r}^N / 2^{2N-r}.$$

$$4.22. 1/9; 2/9.$$

4.23. Пусть событие S — “человек болен”, $\bar{S}$ — “человек здоров”, R_S — “результат рентгеновского анализа положительный” (человек согласно результату рентгеновского анализа признан больным). По условию задачи

$$P(R_S|S) = 1 - \beta, P(S) = \gamma, P(\bar{S}) = 1 - \gamma, P(R_S|\bar{S}) = \alpha.$$

Необходимо вычислить $P(\bar{S}|R_S)$.

$$P(\bar{S}|R_S) = \alpha(1 - \gamma)/(\alpha(1 - \gamma) + (1 - \beta)\gamma).$$

4.24. Пусть событие S — “изделие стандартное”, $\bar{S}$ — “изделие нестандартное”, D_S — “классификация изделия как стандартного”. По условию

$$P(S) = 0,96, P(D_S|S) = 0,98, P(D_S|\bar{S}) = 0,05.$$

Необходимо вычислить $P(S|D_S)$.

$$P(S|D_S) = 0,9978.$$

$$\mathbf{4.25^\circ.} \ 2/5. \quad \mathbf{4.26.} \ 5/13.$$

$$\mathbf{4.27.}$$

$$\frac{N_k M_k}{(N_k + M_k)(N_k + M_k - 1)} \left(\sum_{i=1}^3 \frac{N_i M_i}{(N_i + M_i)(N_i + M_i - 1)} \right)^{-1}.$$

4.28. Задачу о разделе ставки Паскаль и Ферма рассматривали как вероятностную.

Пусть вместо игроков, которые прервали игру, ее продолжат двое новых игроков (один за первого, другой — за второго, мы по-прежнему будем называть их первым и вторым игроками). Они играют до выигрыша одним из игроков шести партий. При этом 6 партий может выиграть как первый игрок, так и второй. Вот только с разными вероятностями. Вполне естественно считать справедливым раздел приза пропорционально вероятностям выиграть 6 партий соответственно первым и вторым при продолжении игры.

Игра будет продолжаться не более чем три партии. Пусть событие A_i — “первый игрок выиграл i -ю партию”,

B_i — “второй игрок выиграл i -ю партию”, $i = 1, 2, 3$;
 C — игру выиграл первый. Очевидно,

$$C = A_1 \cup (B_1 \cap A_2) \cup (B_1 \cap B_2 \cap A_3),$$

C — объединение непересекающихся событий, причем события A_i и B_j при разных значениях индексов независимы, поэтому

$$\begin{aligned} P(C) &= P(A_1) + P(B_1)P(A_2) + P(B_1)P(B_2)P(A_3) = \\ &= \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} = \frac{7}{8}. \end{aligned}$$

Вероятность выигрыша игры вторым игроком равна $1/8$. Поэтому приз необходимо разделить в отношении $7 : 1$.

4.29. Приз необходимо разделить в отношении $15 : 1$.

4.30. Шар, который изначально находится в урне, обозначим через W , если он белый, и через B , если он черный, а белые шары, которые кладут в урну обозначим через W_i , $i = 1, 2, \dots, n$ (будем считать их различимыми). Эксперимент состоит в последовательном извлечении из урны $(n + 1)$ шаров. Исход эксперимента — последовательность длины $(n + 1)$, составленная из букв $W_1, W_2, \dots, W_n, W$ или из букв $W_1, W_2, \dots, W_n, B$. Поскольку сначала в урне с вероятностью $1/2$ находится белый или черный шар, то естественно считать модель классической. Пусть A_n — событие “первые n извлеченных шаров белые”, A — “последний извлеченный шар белый”. Ясно, что необходимо вычислить вероятность

$$P(A/A_n) = \frac{P(A \cap A_n)}{P(A_n)}.$$

Вычислим вероятности $P(A_n)$ и $P(A \cap A_n)$ по формуле полной вероятности, рассматривая как полную группу события: W — “сначала в урне находится белый шар”, B — “сначала в урне находится черный шар”. Имеем

$$\begin{aligned} P(A_n) &= P(A_n/W)P(W) + P(A_n/B)P(B) = \\ &= 1 \cdot \frac{1}{2} + \frac{n!}{(n+1)!} \cdot \frac{1}{2} = \frac{1}{2} \left(1 + \frac{1}{n+1} \right); \end{aligned}$$

$$\begin{aligned}
 P(A \cap A_n) &= P((A \cap A_n)/W)P(W) + P((A \cap A_n)/B)P(B) = \\
 &= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.
 \end{aligned}$$

Так что

$$P(A/A_n) = \frac{1/2}{(1/2)(1 + 1/(n+1))} = \frac{n+1}{n+2}.$$

Интересно, что когда $n \rightarrow \infty$, то

$$P(A/A_n) \rightarrow 1.$$

Последнее вполне согласуется с нашей интуицией. При наличии в урне черного шара вероятность того, что первые n выбранных шаров будут белыми, равна

$$P(A_n/B) = \frac{n!}{(n+1)!} = \frac{1}{n+1},$$

а вероятность того, что среди первых n извлеченных шаров окажется черный —

$$P(\overline{A_n}/B) = \frac{n}{(n+1)},$$

последняя при больших n близка к 1. Если в урне сначала был белый шар, то $P(A_n/W)$ — вероятность того, что первые n извлеченных шаров будут белыми, равна 1. И поскольку первые n шаров оказались белыми, то скорее всего сначала в урне был белый шар, а следовательно, вероятность $P(A/A_n)$ близка к 1.

Заметим, что в частном случае, когда $n = 1$, мы имеем дело с задачей Льюиса Кэррола, при этом

$$P(A/A_1) = \frac{n+1}{n+2} = \frac{2}{3},$$

что совпадает с ответом к этой задаче.

4.32. Обозначим через A_{ji} событие “объект будет обнаружен j -й станцией в i -м цикле”, $j = 1, 2, \dots, m$;

$i = 1, 2, \dots, n$; через A_j событие — “объект будет обнаружен j -й станцией”, $j = 1, 2, \dots, m$. Очевидно,

$$A_j = \bigcup_{i=1}^n A_{ji};$$

$$A = \bigcup_{j=1}^m A_j = \bigcup_{j=1}^m \bigcup_{i=1}^n A_{ji};$$

$$B = \bigcap_{j=1}^m A_j = \bigcap_{j=1}^m \left(\bigcup_{i=1}^n A_{ji} \right).$$

Отсюда, учитывая независимость событий A_{ji} , $j = 1, 2, \dots, m$; $i = 1, 2, \dots, n$, имеем

$$P(A) = 1 - P(\overline{A}) = 1 - P\left(\overline{\bigcup_{j=1}^m \bigcup_{i=1}^n A_{ji}}\right) =$$

$$= 1 - P\left(\bigcap_{j=1}^m \bigcap_{i=1}^n \overline{A_{ji}}\right) = 1 - (1 - p)^{nm};$$

$$P(B) = P\left(\bigcap_{j=1}^m \left(\bigcup_{i=1}^n A_{ji}\right)\right) = \prod_{j=1}^m P\left(\bigcup_{i=1}^n A_{ji}\right) =$$

$$= \prod_{j=1}^m \left(1 - P\left(\overline{\bigcup_{i=1}^n A_{ji}}\right)\right) = \prod_{j=1}^m \left(1 - P\left(\bigcap_{i=1}^n \overline{A_{ji}}\right)\right) =$$

$$= \prod_{j=1}^m (1 - (1 - p)^n) = (1 - (1 - p)^n)^m.$$

4.33. Обозначим через A событие “за выбранной Вами дверью находится автомобиль”, через B — “за открытыми

ведущим $n - 2$ дверями находится автомобиль". Вычислим $P(B)$, рассмотрев A и $\bar{A}$ в качестве полной группы событий:

$$\begin{aligned} P(B) &= P(B/A)P(A) + P(B/\bar{A})P(\bar{A}) = \\ &= 0 \cdot \frac{1}{n} + \frac{n-2}{n-1} \cdot \frac{n-1}{n} = \frac{n-2}{n}. \end{aligned}$$

Заметим, что $P(\bar{B}) = 1 - (n-2)/n = 2/n$ — вероятность того, что за открытыми ведущим дверями автомобиля нет (но это не вероятность выигрыша автомобиля Вами).

Вычислим $P(A/\bar{B})$ и $P(\bar{A}/\bar{B})$:

$$\begin{aligned} P(A/\bar{B}) &= \frac{P(\bar{B}/A)P(A)}{P(\bar{B}/A)P(A) + P(\bar{B}/\bar{A})P(\bar{A})} = \\ &= \frac{1/n}{1/n + (1/(n-1))((n-1)/n)} = \frac{1}{2}, \end{aligned}$$

а значит, и $P(\bar{A}/\bar{B}) = 1/2$.

Следовательно, если автомобиля нет за выбранными ведущим дверями, то он с вероятностью $1/2$ находится за первоначально выбранной Вами дверью (и с такой же — за оставшейся).

Вычислим $P(\bar{B} \cap \bar{A})$ и $P(\bar{B} \cap A)$.

Очевидно, $A \subset \bar{B}$, поэтому $\bar{B} \cap A = A$, и, следовательно,

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

Отсюда и из равенства $P(\bar{B} \cap A) = P(A) = 1/n$ имеем

$$\begin{aligned} P(\bar{B}) &= P(A) + P(\bar{B} \cap \bar{A}), \\ \frac{2}{n} &= \frac{1}{n} + P(\bar{B} \cap \bar{A}). \end{aligned}$$

Так что

$$P(\bar{B} \cap \bar{A}) = \frac{1}{n}, \quad P(\bar{B} \cap A) = \frac{1}{n}.$$

Поэтому, какую бы дверь вы ни выбрали (после выбора ведущего), вероятность выигрыша для Вас составляет $1/n$.

21.5 К главе 5

5.1°.

x_i	$-\sqrt{3}/2$	0	$\sqrt{3}/2$
$P_\xi(x_i)$	1/3	1/3	1/3

5.2°.

y_j	-1	1
$P_\eta(y_j)$	3/7	4/7

5.3°.

x_i	0	1	2
$P_\xi(x_i)$	0,27	0,58	0,15

5.4°.

$$P\{\xi_1 = k_1, \xi_2 = k_2, \dots, \xi_n = k_n\} = \prod_{i=1}^n p(1-p)^{k_i};$$

$$k_1 = 0, 1, \dots; k_2 = 0, 1, \dots; \dots; k_n = 0, 1, \dots$$

5.6*. $(\mu, s) = (\mu(\zeta^*), s(\zeta^*))$ — функция случайного вектора $\zeta^* = (\text{sign } \xi_1, \text{sign } \xi_2, \dots, \text{sign } \xi_n)$.

Сначала найдем распределение ζ^* :

$$\begin{aligned} P\{\zeta^* = x\} &= \\ &= P\{(\text{sign } \xi_1, \text{sign } \xi_2, \dots, \text{sign } \xi_n) = (x_1, x_2, \dots, x_n)\} = \\ &= \prod_{i=1}^n P\{\text{sign } \xi_i = x_i\} = p^{\mu(x)} q^{s(x) - \mu(x)} f^{n-s(x)}, \end{aligned}$$

где x_i , $i = 1, 2, \dots, n$, принимает значения $-1, 0, 1$, $s(x)$ — количество компонент вектора x отличных от нуля, $\mu(x)$ — равных $+1$.

$$P\{\mu = k, s = j\} = P\{\mu(\zeta^*) = k, s(\zeta^*) = j\} =$$

$$\begin{aligned}
&= \sum_{x: \mu(x)=k, s(x)=j} P_{\zeta^*}(x) = \\
&= \sum_{x: \mu(x)=k, s(x)=j} p^{\mu(x)} q^{s(x)-\mu(x)} f^{n-s(x)} = \\
&= C_n^j C_j^k p^k q^{j-k} f^{n-j},
\end{aligned}$$

$k = 1, 2, \dots, j; j = 1, 2, \dots, n.$

5.7. См. 5.8.

5.8. Решение к пункту 2(ж). Сначала найдем распределение $\zeta = (\xi, \eta)$.

Случайная величина $\zeta = \zeta(\omega) = (\xi(\omega), \eta(\omega))$ является функцией на вероятностном пространстве $\{\Omega, P\}$. Пространство элементарных событий образуют пары (i, j) , $i = 1, 2, \dots, 6, j = 1, 2, \dots, 6$, а $P(i, j) = 1/36$ для всех (i, j) (выбрана классическая модель, поскольку кости симметричны). Поэтому распределением $\zeta = (\xi, \eta)$ является

$$P_{\zeta}(i, j) = P\{\zeta = (i, j)\} = 1/36, \quad i, j = 1, 2, \dots, 6.$$

Табличная форма записи распределения случайной величины $\zeta = (\xi, \eta)$ приведена в табл. 21.5.1.

Т а б л и ц а 21.5.1. Распределение $\zeta = (\xi, \eta)$

Значения ξ	Значения η					
	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

По известному распределению случайной величины $\zeta = (\xi, \eta)$ (см. табл. 21.5.1), пользуясь формулой (5.1.2) (где $g = g(s, t) = \max\{s, t\}$, $B = [3, +\infty)$), находим

$$\begin{aligned} P\{\max\{\xi, \eta\} \geq 3\} &= 1 - P\{\max\{\xi, \eta\} < 3\} = \\ &= 1 - \sum_{(i,j): \max\{i,j\} < 3} P_{\zeta}(i, j) = 1 - (P_{\zeta}(1, 1) + P_{\zeta}(1, 2) + \\ &\quad + P_{\zeta}(2, 1) + P_{\zeta}(2, 2)) = 1 - 4/36 = 8/9. \end{aligned}$$

Решение к пункту 3(в). Совместное распределение ξ и $\min\{\xi, \eta\}$ (распределение вектора $(\xi, \min\{\xi, \eta\})$) находим как распределение функции

$$\theta = g(\xi, \eta) = (\xi, \min\{\xi, \eta\})$$

от $\zeta = (\xi, \eta)$ по известному распределению $\zeta = (\xi, \eta)$ (см. табл. 21.5.1). Пользуясь формулой (5.1.2), вычислим несколько значений $P_{\theta}(i, j) = P\{(\xi, \min\{\xi, \eta\}) = (i, j)\}$. Имеем

$$\begin{aligned} P\{(\xi, \min\{\xi, \eta\}) = (1, 1)\} &= \sum_{(i,j): (i, \min\{i,j\}) = (1,1)} P_{\zeta}(i, j) = \\ &= P_{\zeta}(1, 1) + P_{\zeta}(1, 2) + \dots + P_{\zeta}(1, 6) = 6/36; \\ P\{(\xi, \min\{\xi, \eta\}) = (1, 2)\} &= \sum_{(i,j): (i, \min\{i,j\}) = (1,2)} P_{\zeta}(i, j) = 0; \\ P\{(\xi, \min\{\xi, \eta\}) = (1, k)\} &= \sum_{(i,j): (i, \min\{i,j\}) = (1,k)} P_{\zeta}(i, j) = 0; \end{aligned}$$

$k = 3, 4, 5, 6$.

Совместное распределение ξ и $\min\{\xi, \eta\}$ (распределение случайной величины $\theta = (\xi, \min\{\xi, \eta\})$) приведено в табл. 21.5.2.

Случайные величины ξ и $\min\{\xi, \eta\}$ не являются независимыми, поскольку их совместное распределение отличается от произведения распределений случайных величин ξ и $\min\{\xi, \eta\}$ (распределение $\min\{\xi, \eta\}$ приведено в ответе к пункту 1(б)).

Таблица 21.5.2. Распределение $\theta = (\xi, \min\{\xi, \eta\})$

Значения ξ	Значения $\min\{\xi, \eta\}$					
	1	2	3	4	5	6
1	6/36	0	0	0	0	0
2	1/36	5/36	0	0	0	0
3	1/36	1/36	4/36	0	0	0
4	1/36	1/36	1/36	3/36	0	0
5	1/36	1/36	1/36	1/36	2/36	0
6	1/36	1/36	1/36	1/36	1/36	1/36

Ответ к пункту 1(а):

1	2	3	4	5	6
1/36	3/36	5/36	7/36	9/36	11/36

Ответ к пункту 1(б):

1	2	3	4	5	6
11/36	9/36	7/36	5/36	3/36	1/36

Ответы: 2(а) $1/6$; 2(б) $3/4$; 2(в) $1/3$; 2(г) $1/3$; 2(д) $1/9$; 2(е) $4/9$; 2(ё) $1/9$; 2(ж) $8/9$; 2(з) $7/9$.

Ответ к пункту 3(б).

Распределение случайной величины $\theta = (\xi, \max\{\xi, \eta\})$ (совместное распределение ξ и $\max\{\xi, \eta\}$) приведено в табл. 21.5.3.

Случайные величины ξ и $\max\{\xi, \eta\}$ не являются независимыми, поскольку их совместное распределение (см. табл. 21.5.3), отлично от произведения распределений случайных величин ξ и $\max\{\xi, \eta\}$ (распределение $\max\{\xi, \eta\}$ приведено в ответе к пункту 1(а)).

Т а б л и ц а 21.5.3. Совместное распределение ξ и $\max\{\xi, \eta\}$

Значения ξ	Значения $\max\{\xi, \eta\}$					
	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	0	2/36	1/36	1/36	1/36	1/36
3	0	0	3/36	1/36	1/36	1/36
4	0	0	0	4/36	1/36	1/36
5	0	0	0	0	5/36	1/36
6	0	0	0	0	0	6/36

$$5.11. P\{\xi = \eta\} = \sum_{k=1}^n a_k b_k.$$

5.12.

$$P\{\xi + \eta = k\} = \begin{cases} \frac{k+1}{(n+1)^2}, & \text{если } 0 \leq k \leq n; \\ \frac{2n-k+1}{(n+1)^2}, & \text{если } n < k \leq 2n. \end{cases}$$

5.13. Случайная величина $\zeta = \xi + \eta$ имеет распределение Пуассона с параметром $\lambda + \mu$.

Воспользуйтесь формулой (5.1.2). При этом совместное распределение случайных величин ξ и η равно произведению пуассоновских распределений, а $g = g(x, y) = x + y$.

5.15.

$$P\{(\xi, \max\{\xi, \eta\}) = (i, j)\} =$$

$$= \begin{cases} p^2(1-p)^{i+j}, & \text{если } i < j; \\ p(1-p)^i(1-(1-p)^{i+1}), & \text{если } i = j; \\ 0, & \text{если } i > j; \end{cases}$$

$$i, j = 0, 1, \dots;$$

$$P\{\max\{\xi, \eta\} = k\} = p^2(1-p)^k (2(1-(1-p)^k)/p + (1-p)^k),$$

$$k = 0, 1, \dots$$

5.16.

1	2	3	4	5	6	7
1/18	3/18	3/18	3/18	3/18	3/18	2/18

5.17.

Распределение случайной величины $\eta = (\xi_1, \xi_2)$

Значения ξ_1	Значения ξ_2		
	0	1	2
0	1/8	1/4	1/8
1	1/8	1/4	1/8

5.18*. $p_1 + p_2 - p_1 p_2$.

5.19°. 1) 80/243. **5.20°**. 0,0536. **5.21°**. 5/16.

5.22. 1) C_{10}^1 ; 2) $p(1-p)^{k-1}$; 3) $(1-p)^{10}$, независимо от l .

5.23°.

$$P\{\xi = k\} = C_3^k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{3-k}, \quad k = 0, 1, 2, 3.$$

5.24.

а) $1 - \frac{5^6}{6^6}$; б) $1 - \frac{17 \cdot 5^{11}}{6^{12}}$; в) $1 - \frac{268 \cdot 5^{16}}{6^{18}}$.

5.25°. $P\{\xi = k\} = C_{10}^k / 2^{10}$; $k = 0, 1, \dots, 10$.

5.26°.

$$P\{\xi = k\} = C_{10}^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k}, \quad k = 0, 1, \dots, 10.$$

5.27°.

$$P\{\xi = k\} = C_3^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{3-k}, \quad k = 0, 1, 2, 3.$$

5.28. Вероятнее выиграть:

- 1) 3 партии из 5;
- 2) 2 партии из 4;
- 3) 3 партии из 4;
- 4) не менее 5 партий из 8;
- 5) не более чем n из $2n$ партий;
- 6) не более чем n из $2n + 1$ партий.

5.29. Сначала выпишем совместное распределение $\xi_1, \xi_2, \dots, \xi_r$, а затем распределение функции

$$\eta = \xi_1 + \xi_2 + \dots + \xi_r$$

от случайных величин $\xi_1, \xi_2, \dots, \xi_r$.

Совместным распределением $\xi_1, \xi_2, \dots, \xi_r$ является

$$P(k_1, k_2, \dots, k_r) = P\{\xi_1 = k_1, \xi_2 = k_2, \dots, \xi_r = k_r\} =$$

$$= \prod_{i=1}^r p(1-p)^{k_i} = p^r (1-p)^{k_1+k_2+\dots+k_r},$$

$$k_1 = 0, 1, \dots; k_2 = 0, 1, \dots; \dots; k_r = 0, 1, \dots$$

Для каждого k , $k = 0, 1, \dots$,

$$P_\eta(k) = P\{\eta = k\} = P\{\xi_1 + \xi_2 + \dots + \xi_r = k\} =$$

$$= \sum_{(k_1, k_2, \dots, k_r): k_1+k_2+\dots+k_r=k} p^r (1-p)^{k_1+k_2+\dots+k_r} =$$

$$= \sum_{(k_1, k_2, \dots, k_r): k_1+k_2+\dots+k_r=k} p^r (1-p)^k = C_{k+r-1}^{r-1} p^r (1-p)^k$$

(суммирование ведется по всем последовательностям неотрицательных чисел $k_1, k_2, \dots, k_r$, являющихся решениями уравнения $x_1 + x_2 + \dots + x_r = k$, всего таких решений — C_{k+r-1}^{r-1} , см. пример 1.1.14). Так что η имеет отрицательное биномиальное распределение с параметрами (r, p) .

5.30. Пусть ξ — количество лиц, обратившихся в справочное бюро, η — количество отказав. События $\{\xi = k\}$,

$k = 0, 1, 2, \dots$, образуют полную группу событий. Согласно формуле полной вероятности

$$\begin{aligned}
 P\{\eta = s\} &= \sum_{k=0}^{\infty} P\{\eta = s | \xi = k\} P\{\xi = k\} = \\
 &= \sum_{k=s}^{\infty} P\{\eta = s | \xi = k\} P\{\xi = k\} = \\
 &= \sum_{k=s}^{\infty} \frac{k!}{s!(k-s)!} p^s (1-p)^{k-s} \frac{\lambda^k}{k!} e^{-\lambda} = \\
 &= \frac{e^{-\lambda} p^s \lambda^s}{s!} \sum_{k=s}^{\infty} \frac{1}{(k-s)!} (1-p)^{k-s} \lambda^{k-s} = \\
 &= \frac{e^{-\lambda} p^s \lambda^s}{s!} e^{\lambda(1-p)} = \frac{(\lambda p)^s}{s!} e^{-\lambda p}.
 \end{aligned}$$

Следовательно, количество отказов имеет пуассоновское распределение с параметром λp .

5.31. Пуассоновское распределение с параметром λp (см. задачу 5.30).

5.32. Пуассоновское распределение с параметром λp (см. задачу 5.30).

5.34. Обозначим через k количество мест в каждом гардеробе. Пусть μ_i — случайная величина, принимающая значение 1, если i -й зритель выбрал вход 1, и 0, если выбрал вход 2, $i = 1, 2, \dots, 1000$; μ_i — независимые случайные величины, каждая с распределением

$$P\{\mu_i = x\} = (1/2)^x (1/2)^{1-x}, \quad x = 0, 1.$$

Тогда $\mu = \sum_{i=1}^{1000} \mu_i$ — количество зрителей, которые вошли через вход 1, а $1000 - \mu$ — через вход 2.

Событие “все зрители имеют возможность раздеться” в терминах случайной величины μ запишется так:

$$\{\mu < k, 1000 - \mu < k\} = \{1000 - k < \mu < k\}.$$

Количество мест k (в каждом гардеробе) находится как минимальное натуральное число, удовлетворяющее условию

$$P\{1000 - k < \mu < k\} \geq 0,90.$$

Чтобы найти k , воспользуемся интегральной теоремой Муавра—Лапласа (формула (5.2.6)). См. также пример 5.2.1 и задачу 5.39.

Ответ: $k = 527$.

5.35. Упорядочить $x_1, x_2, \dots, x_n$ по возрастанию.

5.37. $p_k = (1/2)^k$. Вероятность выигрыша для начинающего первым в два раза больше.

5.38. 1. Вероятнее, что стрелки получат приз, поскольку $P\{\text{стрелки не получают приз}\} = (1 - 1/5)^4 < 0,5$; 2. Отношение вероятностей равно $3/2$.

5.39. Решение (1°а). Частота ν_n равна μ/n , где μ — число тех экспериментов из n проведенных, в которых событие произошло.

Отклонение в опыте Бюффона составляет $2048/4040 - 0,5 = 0,007 = \varepsilon$. Необходимо вычислить

$$P\{|\nu_n - 0,5| < \varepsilon\}$$

для $n = 4040$.

Частоту $\nu_n = \mu/n$ представим в виде

$$\nu_n = \mu/n = \frac{1}{n} \sum_{i=1}^n \mu_i,$$

где μ_i принимает значение 1, если в эксперименте событие произошло и 0 — в противном случае. Случайные величины μ_i независимы и $q = P\{\mu_i = 0\} = 1/2$, $p = P\{\mu_i = 1\} = 1/2$. Поскольку число экспериментов n большое ($n = 4040$), то для вычисления $P\{|\nu_n - 0,5| < \varepsilon\}$ можно воспользоваться интегральной предельной теоремой Муавра—Лапласа (см. (5.2.6)):

$$\begin{aligned} P\{|\nu_n - 0,5| \leq \varepsilon\} &= P\left\{0,5 - \varepsilon \leq \frac{\mu}{n} \leq 0,5 + \varepsilon\right\} = \\ &= P\left\{\frac{n(0,5 - \varepsilon) - np}{\sqrt{npq}} \leq \frac{\mu - np}{\sqrt{npq}} \leq \frac{n(0,5 + \varepsilon) - np}{\sqrt{npq}}\right\} = \end{aligned}$$

$$= P \left\{ -0,88 \leq \frac{\mu - np}{\sqrt{npq}} \leq 0,88 \right\}.$$

А последняя вероятность с незначительной погрешностью равна

$$\frac{1}{\sqrt{2\pi}} \int_{-0,88}^{0,88} \exp \left\{ \frac{-t^2}{2} \right\} dt = 0,6212$$

(значение интеграла найдено по табл. 22.1.1 нормального распределения).

5.40. Непосредственно “перебираем” все варианты.

1) $4/27$; 2) $1/9$; 3) $23/27$.

5.41.

$$\begin{aligned} P_{\xi}(x_i) &= P\{\xi = x_i\} = P\{(\xi, \eta) \in \{x_i\} \times \mathbb{R}^1\} = \\ &= \sum_{(x_k, y_j) \in \{x_i\} \times \mathbb{R}^1} P_{\zeta}(x_k, y_j) = \sum_{y_j} P_{\zeta}(x_i, y_j). \end{aligned}$$

5.42. Пусть $\Omega = \{\omega_1, \omega_2\}$; $P(\omega_1) = 1/2$, $P(\omega_2) = 1/2$. Случайные величины $\xi = \xi(\omega)$, $\eta = \eta(\omega)$ зададим равенствами

$$\xi(\omega_1) = -1, \quad \xi(\omega_2) = 1;$$

$$\eta(\omega_1) = 1, \quad \eta(\omega_2) = -1.$$

Распределения

$$P_{\xi}((-1)^i) = 1/2, \quad i = 1, 2;$$

$$P_{\eta}((-1)^j) = 1/2, \quad j = 1, 2;$$

совпадают, при этом

$$\xi \neq \eta.$$

5.44. Число неудач ξ до r -го успеха в последовательности независимых испытаний с вероятностью успеха p в каждом испытании имеет отрицательное биномиальное распределение с параметрами $(r; p)$:

$$P_{\xi}(k) = C_{k+r-1}^{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

Решение задачи сводится к вычислению вероятности того, что число неудач до 4-го успеха равно 6 (вероятность успеха в каждом испытании $1/2$). Искомая вероятность

$$C_{6+4-1}^{4-1} (1/2)^4 (1 - 1/2)^6 = C_9^3 (1/2)^{10}.$$

21.6 К главе 6

6.1°. 1) $-1/8$; 2) $(1+e)(2+\sqrt{3})/12$.

6.2. $M\xi = 3/5$, $D\xi = 28/75$,

x_i	0	1	2
$P_\xi(x_i)$	7/15	7/15	1/15

6.3°. $27/40$. **6.4°.** $-\sqrt{3}/8$. **6.6°.** 0.

6.7°. 1) 0; 2) 0; 3) $3\sqrt{3}/4$. **6.8°.** $69/8$.

6.9°. 1) $\sqrt{3}$; 2) $(8+\sqrt{3})/4$.

6.10. $M\eta = 13/12$; $D\eta = 59/144$,

y_j	0	1	2
$P_\xi(y_j)$	1/6	7/12	1/4

6.11°. $(2-\sqrt{3})/12$. **6.12°.** 1) $(20+\sqrt{3})/8$; 2) $-5/2$.

6.13. $5/6$. **6.14.** $29/54$.

6.15. 1) $M\xi = 0$; 2) $M|\xi| = \frac{n(n+1)}{2n+1}$.

6.16. $M\xi = \lambda$, $M\xi^2 = \lambda + \lambda^2$, $D\xi = \lambda$.

6.17. Решение к пункту 3):

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

(см. также пример 6.1.2).

Ответы: 1) $1 - e^{-\lambda}$; 2) λ^2 ; 3) $\lambda^2(1 + 2\lambda)$; 4) 2λ .

6.18. Для $|x| < 1$

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Продифференцировав последнее равенство и умножив на x , получим

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}. \quad (21.6.1)$$

Продифференцировав равенство (21.6.1) и умножив его на x , получим

$$\sum_{k=0}^{\infty} k^2 x^k = \frac{x(x+1)}{(1-x)^3}. \quad (21.6.2)$$

Из равенств (21.6.1) и (21.6.2) при $x = 1 - p$ имеем

$$M\xi = \sum_{k=0}^{\infty} k(1-p)^k p = p \sum_{k=0}^{\infty} k(1-p)^k = p \frac{1-p}{p^2} = \frac{1-p}{p},$$

$$M\xi^2 = \sum_{k=0}^{\infty} k^2(1-p)^k p = p \sum_{k=0}^{\infty} k^2(1-p)^k = \frac{(1-p)(2-p)}{p^2},$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{1-p}{p^2}.$$

6.19. Число подбрасываний ξ кости до первого выпадения шестерки — геометрически распределенная случайная величина с параметром $p = 1/6$. Поэтому $M\xi = 5$; $P\{\xi \leq 2\} = 91/6^3$.

6.20.

$$1) Mx^\xi = \frac{p}{1 - (1-p)x}; \quad 2) Me^{it\xi} = \frac{p}{1 - (1-p)e^{it}}.$$

$$\mathbf{6.21.} \quad 1) Mx^\xi = e^{\lambda(x-1)}; \quad 2) Me^{it\xi} = \exp\{\lambda(e^{it} - 1)\}.$$

$$\mathbf{6.22^\circ.} \quad M\xi = 70; \quad D\xi = 21.$$

$$\mathbf{6.24^\circ.} \quad P\{\xi = k\} = C_{10}^k/2^{10}, \quad k = 0, 1, \dots, 10; \\ M\xi = 5; \quad D\xi = 2,5.$$

$$\mathbf{6.25^\circ.} \quad P\{\xi = k\} = C_{10}^k(1/6)^k(5/6)^{10-k}, \quad k = 0, 1, \dots, 10; \\ M\xi = 5/3; \quad D\xi = 25/18.$$

$$\mathbf{6.26^\circ.} \quad M\xi = 3/4; \quad D\xi = 9/16.$$

6.27*. Распределение случайной величины ξ совпадает с распределением суммы $\xi = \xi_1 + \xi_2 + \dots + \xi_r$, где ξ_i , $i = 1, 2, \dots, r$, независимые случайные величины, каждая из которых имеет геометрическое распределение с параметром p (см. также задачу 5.29).

Математическое ожидание случайной величины, имеющей геометрическое распределение с параметром p , равно $(1-p)/p$ (см. задачу 6.18).

$$\mathbf{6.30.} \quad \sum_{k=1}^r p_k \exp\{it_k\}.$$

6.31*. Выразите вектор $\nu = (\nu_1, \nu_2, \dots, \nu_r)$ через векторы

$$(\mu_1^k, \mu_2^k, \dots, \mu_r^k) = (I_{X_1}(\xi_k), I_{X_2}(\xi_k), \dots, I_{X_r}(\xi_k)),$$

$k = 1, 2, \dots, r$, где I_A — индикатор множества A .

$$\text{Ответ: } \left(\sum_{k=1}^r p_k \exp\{it_k\} \right)^n.$$

6.32. Найдём распределение случайной величины ξ — количества анализов, необходимых для обследования крови k человек, в случае проведения обследования способом 2.

Случайная величина ξ может принимать два значения: 1 и $k+1$. Пусть η — количество положительных анализов на k проведенных. Очевидно,

$$P\{\xi = 1\} = P\{\eta = 0\}, \quad P\{\xi = k+1\} = P\{\eta \geq 1\}.$$

Случайная величина η имеет биномиальное распределение с параметрами $(n; p)$, поэтому вероятность хотя бы одного положительного анализа из k определяется так:

$$P\{\eta \geq 1\} = 1 - P\{\eta = 0\} = 1 - C_k^0 p^0 (1-p)^k = 1 - (1-p)^k.$$

Вероятность того, что все k анализов отрицательны, равна $(1-p)^k$. Следовательно, распределением ξ является

$$P\{\xi = 1\} = (1-p)^k, \quad P\{\xi = k+1\} = 1 - (1-p)^k.$$

По распределению ξ находим

$$M\xi = 1(1-p)^k + (k+1)(1 - (1-p)^k) = k(1 - (1-p)^k) + 1.$$

6.34.

1° Пусть

$$P\{\xi = k\} = \frac{\theta^k}{k!} e^{-\theta}, \quad \theta > 0, \quad k = 0, 1, \dots$$

Тогда

$$P(t) = Mt^\xi = \sum_{k=0}^{\infty} t^k \frac{\theta^k}{k!} e^{-\theta} = \sum_{k=0}^{\infty} \frac{(t\theta)^k}{k!} e^{-\theta} = e^{\theta(t-1)}.$$

2° Пусть

$$P\{\xi = k\} = (1 - \theta)^k \theta, \quad \theta \in (0; 1), \quad k = 0, 1, \dots$$

Тогда

$$P(t) = Mt^\xi = \sum_{k=0}^{\infty} t^k (1 - \theta)^k \theta = \theta \sum_{k=0}^{\infty} (t(1 - \theta))^k = \frac{\theta}{1 - t(1 - \theta)}.$$

3° Пусть

$$P\{\xi = k\} = C_n^k \theta^k (1 - \theta)^{n-k}, \quad \theta \in (0; 1), \quad k = 0, 1, \dots, n.$$

Тогда

$$\begin{aligned} P(t) &= Mt^\xi = \sum_{k=0}^n t^k C_n^k \theta^k (1 - \theta)^{n-k} = \\ &= \sum_{k=0}^n C_n^k (t\theta)^k (1 - \theta)^{n-k} = (t\theta + (1 - \theta))^n. \end{aligned}$$

6.35. Пусть $p_k, k = 0, 1, \dots$, и $f_n, n = 0, 1, \dots$, — вероятностные распределения,

$$P(t) = \sum_{k=0}^{\infty} t^k p_k, \quad |t| \leq 1, \quad F(t) = \sum_{k=0}^{\infty} t^k f_k, \quad |t| \leq 1$$

их производящие функции. Предположим, что

$$P(t) = F(t), |t| \leq 1. \quad (21.6.3)$$

Отсюда при $t = 0$ имеем $P(0) = F(0)$, или $p_0 = q_0$. Степенные абсолютно сходящиеся ряды можно почленно дифференцировать, поэтому из (21.6.3)

$$\sum_{k=1}^{\infty} kt^{k-1}p_k = P'(t) = Q'(t) = \sum_{k=1}^{\infty} kt^{k-1}f_k.$$

Отсюда при $t = 0$ имеем $P'(0) = F'(0)$ или $p_1 = q_1$ и т. д. Так что

$$p_k = q_k, k = 0, 1, \dots$$

Тем самым утверждение доказано.

6.36. Если ξ и η — независимые случайные величины, то

$$P_{\xi+\eta}(t) = Mt^{\xi+\eta} = Mt^{\xi}t^{\eta} = Mt^{\xi}Mt^{\eta} = P_{\xi}(t)P_{\eta}(t).$$

6.37. Производящая функция суммы независимых пуассоновских случайных величин

$$P_{\xi+\eta}(t) = P_{\xi}(t)P_{\eta}(t) = e^{\theta_1(t-1)}e^{\theta_2(t-1)} = e^{(\theta_1+\theta_2)(t-1)}.$$

С другой стороны $e^{(\theta_1+\theta_2)(t-1)}$ — производящая функция пуассоновской случайной величины с параметром $\theta_1 + \theta_2$. Поэтому согласно теореме единственности (см. задачу 6.35) распределение случайной величины $\xi + \eta$ совпадает с пуассоновским распределением с параметром $\theta_1 + \theta_2$.

6.38. Производящая функция суммы независимых биномиально распределенных случайных величин

$$\begin{aligned} P_{\xi+\eta}(t) &= P_{\xi}(t)P_{\eta}(t) = \\ &= (t\theta + (1-\theta))^n(t\theta + (1-\theta))^m = (t\theta + (1-\theta))^{n+m}. \end{aligned}$$

С другой стороны $(t\theta + (1-\theta))^{n+m}$ — производящая функция биномиально распределенной случайной величины с

параметрами $(n + m, \theta)$. Поэтому согласно теореме единственности (см. задачу 6.35) распределение суммы $\xi + \eta$ совпадает с биномиальным распределением с параметрами $(n + m, \theta)$.

6.39.

$$1 = P(\Omega) = \sum_{\omega \in \Omega} ap^{\omega} = a \sum_{\omega \in \Omega} p^{\omega} = a \frac{1}{1-p} = 1.$$

Отсюда $a = 1 - p$, поэтому $P(\omega) = (1 - p)p^{\omega}$, $\omega \in \Omega$, или, что то же

$$P(k) = (1 - p)p^k, \quad k = 0, 1, \dots$$

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega)P(\omega) = \sum_{k=0}^{\infty} 2^{-k}(1-p)p^k = \frac{2(1-p)}{2-p}.$$

6.40.

$$1 = P(\Omega) = \sum_{\omega \in \Omega} a \frac{1}{\omega!} = \sum_{k=0}^{\infty} a \frac{1}{k!} = ae.$$

Отсюда $a = e^{-1}$, поэтому $P(\omega) = e^{-1} \frac{1}{\omega!}$, $\omega \in \Omega$, или, что то же

$$P(k) = \frac{1}{k!}e^{-1}, \quad k = 0, 1, \dots$$

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega)P(\omega) = \sum_{k=0}^{\infty} 2^k \frac{1}{k!}e^{-1} = e^{-1}e^2 = e.$$

$$M\xi = e, \quad D\xi = M(\xi)^2 - (M\xi)^2 = e^3 - e^2.$$

6.41. См. решение к задаче 6.39, 6.40.

$$a = e^{-\lambda}, \quad M\xi = e^{\lambda}, \quad D\xi = M(\xi)^2 - (M\xi)^2 = e^{3\lambda} - e^{2\lambda}.$$

6.42.

$$D\xi = M(\xi - M\xi)^2 = \sum_{\omega \in \Omega} (\xi(\omega) - M\xi(\omega))^2 P(\omega) \geq$$

$$\begin{aligned}
&\geq \sum_{\omega: |\xi(\omega) - M\xi(\omega)| \geq \varepsilon} (\xi(\omega) - M\xi(\omega))^2 P(\omega) \geq \\
&\geq \varepsilon^2 \sum_{\omega: |\xi(\omega) - M\xi(\omega)| \geq \varepsilon} P(\omega) = \varepsilon^2 P\{\omega : |\xi(\omega) - M\xi(\omega)| \geq \varepsilon\}.
\end{aligned}$$

6.43.

$$\begin{aligned}
&P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right\} = \\
&P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - M\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right)\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} D\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \\
&= \frac{1}{\varepsilon^2} \frac{1}{n^2} D\left(\sum_{i=1}^n \xi_i\right) = \frac{1}{\varepsilon^2} \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.
\end{aligned}$$

6.44. Случайная величина ξ является функцией от вектора $\zeta = (\xi, \eta)$:

$$\xi = g(\zeta) = g(\xi, \eta).$$

Поэтому

$$M\xi = Mg(\xi, \eta) = \sum_{x_i, y_j} x_i P_\zeta(x_i y_j).$$

6.45. Обозначим число очков, выпавших на i -й кости, через ξ_i , $i = 1, 2, 3$. Тогда $\xi = \xi_1 + \xi_2 + \xi_3$ — сумма очков выпавших на трех костях. Случайные величины ξ_1, ξ_2, ξ_3 — независимы, их распределения имеют вид

$$P_{\xi_1}(i) = 1/2, \quad i = 1, 2;$$

$$P_{\xi_2}(j) = 1/2, \quad j = 3, 4;$$

$$P_{\xi_3}(k) = 1/2, \quad k = 5, 6.$$

Поэтому

$$M\xi = M\xi_1 + M\xi_2 + M\xi_3;$$

$$D\xi = D\xi_1 + D\xi_2 + D\xi_3.$$

Распределение $\xi = \xi_1 + \xi_2 + \xi_3$ можно найти, например, так. Определим случайные величины η_1, η_2, η_3 следующим образом:

$$\xi_1 = (\xi_1 - 1) + 1 = \eta_1 + 1,$$

$$\xi_2 = (\xi_2 - 3) + 3 = \eta_2 + 3,$$

$$\xi_3 = (\xi_3 - 5) + 5 = \eta_3 + 5.$$

Случайные величины η_1, η_2, η_3 независимы как функции от независимых случайных величин ξ_1, ξ_2, ξ_3 и одинаково распределены, каждая с распределением

n	0	1
p_n	1/2	1/2

Тогда

$$\xi = \xi_1 + \xi_2 + \xi_3 = \eta_1 + \eta_2 + \eta_3 + 9 = \eta + 9,$$

где $\eta = \eta_1 + \eta_2 + \eta_3$. Очевидно η имеет биномиальное распределение с параметрами $(3, 1/2)$. Распределение функции $\xi = \eta + 9$ от случайной величины η находим согласно (5.1.1).

21.7 К главе 7

7.2. $A = \{(x, y) : x^2 + y^2 \leq 1\}$; $B = \{(0, 0)\}$.

7.10. Рассмотрите совокупность всех алгебр $\mathfrak{A}_\alpha$, содержащих класс $\mathfrak{K}$, и докажите, что $\bigcap_{\alpha} \mathfrak{A}_\alpha = \mathfrak{A}(\mathfrak{K})$.

7.13. Рассмотрите совокупность всех σ -алгебр $\mathfrak{F}_\alpha$, содержащих класс $\mathfrak{K}$, и докажите, что $\bigcap_{\alpha} \mathfrak{F}_\alpha = \sigma(\mathfrak{K})$.

7.14. Пусть $\mathfrak{K}_1$ — класс промежутков $\{[a, b) : a, b \in \mathbb{R}^1\}$ (напомним, что $\sigma(\mathfrak{K}_1) = \mathfrak{B}^1$). Имеет место включение

$$\mathfrak{K}_1 \subset \sigma(\mathfrak{K}).$$

В самом деле, $\sigma(\mathfrak{K})$ содержит $\{a\}$, поскольку

$$\{a\} = \bigcap_{n=1}^n (a - 1/n, a + 1/n),$$

а $[a, b] = (a, b) \cup \{a\}$.

Из определения σ -алгебры, порожденной классом $\mathfrak{K}_1$, имеем

$$\sigma(\mathfrak{K}_1) \subset \sigma(\mathfrak{K}).$$

А поскольку $\sigma(\mathfrak{K}_1) = \mathfrak{B}^1$, то $\mathfrak{B}^1 \subset \sigma(\mathfrak{K})$.

Далее, $\mathfrak{K} \subset \sigma(\mathfrak{K}_1) = \mathfrak{B}^1$ (это устанавливается аналогично тому, как устанавливается включение $\mathfrak{K}_1 \subset \sigma(\mathfrak{K})$). Потому согласно определению σ -алгебры, порожденной классом $\mathfrak{K}$, имеем

$$\sigma(\mathfrak{K}) \subset \mathfrak{B}^1.$$

Следовательно,

$$\sigma(\mathfrak{K}) = \mathfrak{B}^1.$$

7.15. Указание. См. решение задачи 7.14.

7.16. Указание. См. решение задачи 7.14.

21.8 К главе 8

8.1°. 1) $(1 + \ln 2)/2$; 2) $(1 - e^{-1})/2$; 3) $1/3$; 4) $1/4 + e^{-1} + 3(\ln 3 - \ln 4)/4$; 5) $1 - 2/3^4$; 6) $\pi/16$; 7) $1/18$; 8) $1/2 + \ln 3 - \ln 2$; 9) $1/6$; 10) $1/9$.

8.2°. 1) $(1 - \ln 2)/2$; 2) $\pi/64$; 3) $1 - \pi/4$; 4) $\pi/4$.

8.3°. 1) $1/4$; 2) $8/9$; 3) $7/16$; 4) $1/4$; 5) $1 - \pi/4$; 6) $1 - \pi/4$; 7) $\pi/8$; 8) $a\sqrt{a}$; 9) $1 - a\sqrt{a}$; 10) $1 - 1/e$; 11) $\pi/4$; 12) $1 - \pi/8$; 13) $1/2$; 14) $3/4$.

8.4. 1) $1/4$; 2) $1/4$; 3) $1/4$; 4) $3/16$; 5) $3/16$; 6) $1/2$; 7) $1/8$; 8) $1/4$; 9) $3/8$; 10) $1/8$; 11) $1/2$; 12) $5/8$; 13) $\pi/4$; 14) $1/2$; 15) $1/4$; 16) $(\pi - 2)/4$.

8.5°. $a/(a + b - 2r)$. **8.6.** $1 - (1 - k)^2$.

8.7. Если ξ — момент прихода к причалу первого судна, η — второго, то событие “одно из судов будет вынуждено ждать освобождения причала” можно представить в виде

$$\{(\xi, \eta) : \eta \in [\xi, \xi + 1]\} \cup \{(\xi, \eta) : \xi \in [\eta, \eta + 2]\}.$$

Ответ: 0,12065.

8.8°. $(a - d)^2/a^2$. **8.9.** $\frac{1}{2\pi} n \sin \frac{2\pi}{n}$.

8.10. а) $1 - \frac{2}{\pi} \arccos r$; б) $\begin{cases} \frac{2}{\pi} \arcsin \frac{r}{2}, & \text{если } r \leq 2; \\ 1, & \text{если } r > 2. \end{cases}$

8.11°. 1) r^2/R^2 . **8.12°.** 1) r^{2N}/R^{2N} .

8.13. $C_5^2(b/a)^2(1 - b/a)^3$. **8.14.** $1/2^{60}$.

8.15. 1) $1/2$; 2) $1/2$; 3) t^3 .

8.16. 1) $1 - (1 - t)^2$; 2) а) $1/2$; б) $(1 - q)^2/2$.

8.17°. 2) Пусть ξ_i — расстояние от i -й точки до центра сферы, $i = 1, 2, \dots, N$. Тогда искомая вероятность $P\{\min\{\xi_i\} > r\} = P\left(\bigcap_{i=1}^n \{\xi_i > r\}\right) = \prod_{i=1}^n P\{\xi_i > r\} = (1 - r^3/R^3)^N$.

8.18. $1/2$. **8.19.** $13/24$; $11/24$. **8.20.** $1/3$.

8.21°. $2/\pi$. **8.24.** 1) $\pi/4$; 2) $\pi/(2\sqrt{3})$.

8.25. 1) $1/3$; 2) $1/6$. **8.26.** $2/9$.

8.27*. Монету будем рассматривать как цилиндр, радиуса r и высоты h . Слова “наудачу бросают” естественно понимать как равномерную ориентацию цилиндра относительно вертикального направления. Поскольку цилиндр однородный, то центр описанной около него сферы совпадает с его центром масс. Цилиндр упадет на боковую поверхность, если луч, направленный вертикально вниз из центра масс, пересечет его боковую поверхность. Соответствующую вероятность вычисляем как отношение площади сферического пояса, описанного около цилиндра, к площади всей сферы.

8.28. 1) $1/9$; 2) $8/9$; 3) $8/27$.

8.29. При любом повороте окружности вероятность события “треугольник остроугольный” (и не только этого события) остается неизменной, поэтому можно считать,

что из трех вершин A, B, C одна, например C , фиксирована. Две другие выбирают наудачу. Будем задавать положения точек A и B относительно точки C величинами дуг $CA = \alpha$ и $CB = \beta$ (рис. 21.8.1), откладывая их против движения часовой стрелки, дуги будем измерять в радианах.

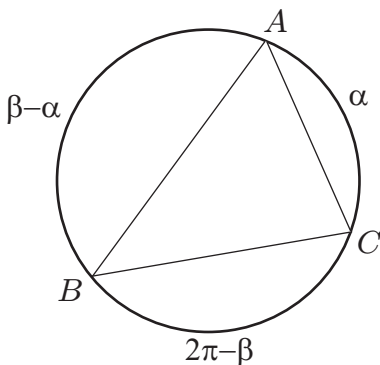


Рис. 21.8.1: К задаче 8.29

Треугольник ABC остроугольный, если каждая из дуг $CA = \alpha$, $AB = \beta - \alpha$, $BC = 2\pi - \beta$ меньше π , т. е.

$$\beta > \alpha, \alpha < \pi, \beta - \alpha < \pi, \beta > \pi, \quad (21.8.1)$$

или каждая из дуг $CB = \beta$, $BA = \alpha - \beta$, $CA = 2\pi - \alpha$ меньше π , т. е.

$$\beta < \alpha, \beta < \pi, \alpha - \beta < \pi, \alpha > \pi. \quad (21.8.2)$$

Далее, упорядоченной паре чисел (α, β) , $0 < \alpha < 2\pi$, $0 < \beta < 2\pi$ соответствует одна точка квадрата $[0, 2\pi] \times [0, 2\pi]$ на плоскости с осями $O\alpha$ и $O\beta$ и наоборот. А следовательно, множеству дуг α, β , которые удовлетворяют одному из условий: (21.8.1) или (21.8.2), соответствует множество точек квадрата $[0, 2\pi] \times [0, 2\pi]$, координаты которых удовлетворяют одному из условий: (21.8.1) или (21.8.2). Это множество точек квадрата имеет вид, аналогичный изображенному на рис. 8.1.2.

Искомая вероятность равна $\frac{2\pi^2/2}{(2\pi)^2} = \frac{1}{4}$.

8.30. 1) $3/4$; 2) $1/4$ (см. также задачу 8.29).

8.31. Зафиксируем одну из точек, например A . Тогда положение других можно описать дугами AB , AC , AD , которые откладываются против движения часовой стрелки (см. также задачу 8.29).

Ответ: $1/3$.

8.35°. $\pi/4$. **8.36°.** $(a-r)/a$. **8.37°.** $2l/L$.

8.38°. а) x , $0 \leq x \leq 1$.

21.9 К главе 9

9.1°. $1/4$. **9.2°.** $3/4$. **9.3°.** 1) $7/12$; 2) $1/3$.

9.4°. Решение к пункту (5). Бросание пары точек на отрезок $[0; 1]$ равносильно бросанию одной точки в квадрат $[0; 1] \times [0; 1]$ и наоборот (см. также примеры 8.1.1, 8.1.2, 8.1.3).

Обозначим $F_\zeta(x) = P\{\zeta < x\} = P\{\max\{\xi^2, \eta\} < x\}$. Очевидно, $F_\zeta(x) = 0$, если $x < 0$, и $F_\zeta(x) = 1$, если $x > 1$. Если $0 < x \leq 1$, то

$$\begin{aligned} F_\zeta(x) &= P\{\max\{\xi^2, \eta\} < x\} = \\ &= P\{\xi < \sqrt{x}, \eta < x\} = P\{(\xi, \eta) \in A\}, \end{aligned}$$

где $A = [0; \sqrt{x}] \times [0; x]$. Вероятность события $\{(\xi, \eta) \in A\}$ вычисляем как геометрическую вероятность:

$$F_\zeta(x) = P\{(\xi, \eta) \in A\} = \frac{L(A)}{L([0; 1] \times [0; 1])} = \frac{x\sqrt{x}}{1^2} = x^{3/2}.$$

Следовательно,

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

Ответы:

$$\begin{aligned}
1) F_{\zeta}(x) &= \begin{cases} 0, & \text{если } x < 0; \\ x^2, & \text{если } 0 \leq x \leq 1; \\ 1, & \text{если } x > 1; \end{cases} \\
2) F_{\zeta}(x) &= \begin{cases} 0, & \text{если } x < 0; \\ 1 - (1 - x)^2, & \text{если } 0 \leq x \leq 1; \\ 1, & \text{если } x > 1; \end{cases} \\
3) F_{\zeta}(x) &= \begin{cases} 0, & \text{если } x \leq 0; \\ x(1 - \ln x), & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1; \end{cases} \\
4) F_{\zeta}(x) &= \begin{cases} 0, & \text{если } x < 0; \\ x^2/2, & \text{если } 0 \leq x < 1; \\ 1 - (2 - x)^2/2, & \text{если } 1 \leq x < 2; \\ 1, & \text{если } x \geq 2; \end{cases} \\
6) F_{\zeta}(x) &= \begin{cases} 0, & \text{если } x < 0; \\ 1 - (1 - x)^2, & \text{если } 0 \leq x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}
\end{aligned}$$

9.5°. Функция распределения случайной величины ξ

$$F_{\xi}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x/l, & \text{если } 0 < x \leq l; \\ 1, & \text{если } x > l; \end{cases}$$

плотность распределения

$$p_{\xi}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1/l, & \text{если } 0 < x \leq l; \\ 0, & \text{если } x > l. \end{cases}$$

9.6.

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \sqrt{x}/2, & \text{если } 0 < x \leq 1; \\ (\sqrt{x} + 1)/4, & \text{если } 1 < x \leq 9; \\ 1, & \text{если } x > 9. \end{cases}$$

9.7.

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x \notin [0; 2]; \\ 1/2, & \text{если } x \in [0; 2]. \end{cases}$$

9.8.

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 1; \\ 1/x^2, & \text{если } x > 1. \end{cases}$$

9.9.

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 1/4; \\ x^{-\frac{3}{2}}/6, & \text{если } 1/4 < x \leq 1; \\ x^{-\frac{3}{2}}/3, & \text{если } x > 1. \end{cases}$$

9.10.

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x \notin [0; 1]; \\ 1, & \text{если } x \in [0; 1]. \end{cases}$$

9.11.

$$p_{\eta}(x) = \begin{cases} \frac{1}{(b-a)x}, & \text{если } x \in [e^a; e^b]; \\ 0, & \text{если } x \notin [e^a; e^b]. \end{cases}$$

9.12*. Воспользуйтесь (9.1.3).

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x \in [0; 1]; \\ \lambda \exp \{-\lambda(1 - 1/x)\} / x^2, & \text{если } x \notin [0; 1]. \end{cases}$$

9.13.

$$F_{\eta}(x) = 1 - F(-x + 0).$$

9.14. Сначала найти распределение дискретной случайной величины $\eta = \text{sign } \xi$, потом по распределению η выписать ее функцию распределения:

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq -1; \\ F(0), & \text{если } -1 < x \leq 0; \\ F(0+0), & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}$$

9.15.

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x, & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}$$

9.16.

$$1) p_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ p(x) + p(-x), & \text{если } x > 0, \end{cases}$$

$$2) p_{\eta}(x) = \frac{1}{|a|} p(x/a).$$

9.17.

$$F_{\eta}(x) = \begin{cases} F(\sqrt{x}) - F(-\sqrt{x} + 0), & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

9.18. Ответ к пункту (1):

$$p_{\eta}(x) = \begin{cases} 0, & \text{если } x < 0; \\ \lambda (e^{-\lambda(x+1)} + e^{\lambda(x-1)}), & \text{если } x \in [0; 1]; \\ \lambda e^{-\lambda(x+1)}, & \text{если } x > 1. \end{cases}$$

9.19. Случайная величина $\eta = F(\xi)$ распределена равномерно на отрезке $[0; 1]$.

9.20. 1) $p_\eta(x) = \frac{1}{2}p\left(\frac{1-x}{2}\right);$

2) $p_\eta(x) = \begin{cases} \frac{1}{2\sqrt{x}}(p(\sqrt{x}) + p(-\sqrt{x})), & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$

9.21.

1) $F_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ F(\ln x), & \text{если } x > 0; \end{cases}$

2) $F_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ F(x) - F(-x + 0), & \text{если } x > 0. \end{cases}$

9.22. 1) $2x - x^2$; 2) $x(1 - \ln x)$. **9.23.** $(F(x))^n$.

9.24. 1) $\prod_{i=1}^n F_i(x)$; 2) $1 - \prod_{i=1}^n (1 - F_i(x))$.

9.25.

1) $\sum_{i=1}^n p_i(x) \prod_{j=1, j \neq i}^n \left(\int_{-\infty}^x p_j(y) dy \right);$

2) $\sum_{i=1}^n p_i(x) \prod_{j=1, j \neq i}^n \left(\int_x^{+\infty} p_j(y) dy \right).$

9.26.

1) $np(x) \left(\int_{-\infty}^x p(y) dy \right)^{n-1};$ 2) $np(x) \left(\int_x^{+\infty} p(y) dy \right)^{n-1}.$

9.27. $a = \lambda/2, \quad F_\eta(x) = \begin{cases} e^{\lambda x}/2, & \text{если } x \leq 0; \\ 1 - e^{-\lambda x}/2, & \text{если } x > 0. \end{cases}$

9.28. Можно предложить по меньшей мере два подхода к решению этой задачи.

Пусть ξ — координата брошенной на отрезок $[0; 1]$ точки, тогда длина η меньшей части отрезка равна

$$\eta = g(\xi) = \min\{\xi, 1 - \xi\}.$$

По определению

$$F_{\eta}(x) = P\{\eta < x\} = P\{\min\{\xi, 1 - \xi\} < x\}.$$

Очевидно, $F_{\eta}(x) = 0$, если $x \leq 0$ и $F_{\eta}(x) = 1$, если $x \geq 1/2$. Если $0 < x \leq 1/2$, то

$$\begin{aligned} F_{\eta}(x) &= P\{\min\{\xi, 1 - \xi\} < x\} = 1 - P\{\min\{\xi, 1 - \xi\} \geq x\} = \\ &= 1 - P\{\xi \geq x, 1 - \xi \geq x\} = 1 - P\{x \leq \xi \leq 1 - x\}. \end{aligned}$$

Вероятность $P\{x \leq \xi \leq 1 - x\}$ вычисляем как геометрическую вероятность.

Другой подход.

$$\begin{aligned} F_{\eta}(x) &= P\{\eta < x\} = P\{g(\xi) < x\} = \\ &= \int_{t:g(t)<x} p_{\xi}(t)dt = \int_{t:\min\{t, 1-t\}<x} p_{\xi}(t)dt. \end{aligned}$$

Следовательно,

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

9.29. См. решение задачи 9.28.

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq l/2; \\ (2x - l)/l, & \text{если } l/2 < x \leq l; \\ 1, & \text{если } x > l. \end{cases}$$

9.30.

$$\begin{aligned} F(x) &= \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - (1 - x/T)^2, & \text{если } 0 < x \leq T; \\ 1, & \text{если } x > T; \end{cases} \\ p(x) &= \begin{cases} 0, & \text{если } x \leq 0; \\ 2(1 - x/T)/T, & \text{если } 0 < x \leq T; \\ 0, & \text{если } x > T. \end{cases} \end{aligned}$$

9.32. Воспользуемся формулой (9.1.3):

$$F(x) = P\left\{\frac{1}{\xi} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{t: 1/t < x} \exp\{-t^2/2\} dt =$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}} \int_{1/x}^0 \exp\{-t^2/2\} dt, & \text{если } x < 0; \\ 1/2, & \text{если } x = 0; \\ \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{1/x}^{+\infty} \exp\{-t^2/2\} dt, & \text{если } x > 0. \end{cases}$$

9.33. $F_\eta(x) = 0$, если $x \leq 0$, а если $x > 0$, то

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1/\sqrt{x}} \exp\{-t^2/2\} dt + \frac{1}{\sqrt{2\pi}} \int_{1/\sqrt{x}}^{+\infty} \exp\{-t^2/2\} dt.$$

9.34. Условие симметричной распределенности случайной величины в терминах ее функции распределения: $F(x) = 1 - F(-x + 0)$, а в терминах плотности распределения: $p(x) = p(-x)$.

9.35. Решение 1):

$$\begin{aligned} F_\eta(x) &= P\{\eta < x\} = P\{1 - \xi < x\} = \\ &= 1 - P\{\xi < 1 - x\} = 1 - F_\xi(1 - x), \end{aligned}$$

$$p_\eta(x) = \frac{d}{dx} F_\eta(x) = \frac{d}{dx} (1 - F_\xi(1 - x)) = p_\xi(1 - x) = p_\xi(x).$$

Ответ: $\eta = 1 - \xi$ равномерно распределена на промежутке $[0; 1]$.

9.36.

$$\begin{aligned} F_\xi(x) &= P\{\xi < x\} = P\left\{\frac{\eta - a}{\sigma} < x\right\} = P\{\eta < a + \sigma x\} = \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{a+\sigma x} \exp\left\{-\frac{(t-a)^2}{2\sigma^2}\right\} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\{-u^2/2\} du \end{aligned}$$

(воспользовались заменой $(t - a)/\sigma = u$).

9.37. Случайная величина η распределена $N_{a;\sigma^2}$.

9.38. Воспользуемся формулой (9.1.4):

$$\begin{aligned}
 F_{\eta^+}(x) &= P\{\max\{0, \eta\} < x\} = \\
 &= \int_{t: \max\{0, t\} < x} \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\} dt = \\
 &= \begin{cases} \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \exp\{-t^2/2\} dt, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}
 \end{aligned}$$

9.39. $F_{\eta^+}(x) =$

$$= \begin{cases} \frac{1}{2} + \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^x \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

9.40.

$$\begin{aligned}
 F_{\eta}(x) &= P\{(\xi - a)^+ < x\} = \\
 &= \int_{t: (t-a)^+ < x} P_{\xi}(dt) = \int_{t: \max\{0, t-a\} < x} P_{\xi}(dt) = \\
 &= \begin{cases} 0, & \text{если } x \leq 0; \\ P_{\xi}(-\infty, x+a), & \text{если } x > 0. \end{cases}
 \end{aligned}$$

Следовательно,

$$F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ F(x+a), & \text{если } x > 0. \end{cases}$$

9.41. Воспользуйтесь формулой (9.1.4) (см. также решение к задаче 9.40).

$$F_{\eta}(x) = \begin{cases} \int_{-\infty}^{x+a} p(t) dt, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

Случайная величина $\eta = (\xi - a)^+$ абсолютно непрерывна, если плотность $p(t)$ случайной величины ξ равна нулю для $t \leq a$.

9.42.

$$\begin{aligned} F_\eta(x) &= P\{\eta < x\} = P\{\min\{\xi, L\} < x\} = \\ &= \int_{t: \min\{t, L\} < x} P_\xi(dt) = \begin{cases} P_\xi(-\infty, x), & \text{если } x \leq L; \\ 1, & \text{если } x > L. \end{cases} \end{aligned}$$

Следовательно,

$$F_\eta(x) = \begin{cases} F(x), & \text{если } x \leq L; \\ 1, & \text{если } x > L. \end{cases}$$

9.43. См. решение задачи 9.42.

9.50.

$$\begin{aligned} 1) \quad F_{\xi_1}(x) &= \begin{cases} 0, & \text{если } x \in (-\infty; 0]; \\ x, & \text{если } x \in (0; 1]; \\ 1, & \text{если } x \in (1; \infty); \end{cases} \\ 2) \quad F_{\xi_2}(x) &= \begin{cases} 0, & \text{если } x \in (-\infty; 0]; \\ x, & \text{если } x \in (0; 1]; \\ 1, & \text{если } x \in (1; \infty); \end{cases} \\ 3) \quad F_{\xi_3}(x) &= \begin{cases} 0, & \text{если } x \in (-\infty; 0]; \\ x, & \text{если } x \in (0; 1/4]; \\ x + 1/4, & \text{если } x \in (1/4; 1/2]; \\ 1, & \text{если } x \in (1/2; \infty); \end{cases} \end{aligned}$$

9.51.

Пусть $\Omega = [0; 1]$, $\mathfrak{F} = \mathfrak{B}_{[0;1]}$, $P = L$,

$$\xi_1 = \xi_1(\omega) = \begin{cases} 1, & \text{если } \omega \in [0; 1/2); \\ 0, & \text{если } \omega \in [1/2; 1]; \end{cases}$$

$$\xi_2 = \xi_2(\omega) = \begin{cases} 0, & \text{если } \omega \in [0; 1/2); \\ 1, & \text{если } \omega \in [1/2; 1]; \end{cases}$$

$$\eta_1 = \eta_1(\omega) = \begin{cases} 1, & \text{если } \omega \in [0; 1/2); \\ 0, & \text{если } \omega \in [1/2; 1]; \end{cases}$$

$$\eta_2 = \eta_2(\omega) = \begin{cases} 1, & \text{если } \omega \in [0; 1/2); \\ 0, & \text{если } \omega \in [1/2; 1]; \end{cases}$$

Так определенные случайные величины удовлетворяют условиям задачи, но 1) распределения $\xi_1\eta_1$ и $\xi_2\eta_2$ различны; 2) распределения $\xi_1 + \eta_1$ и $\xi_2 + \eta_2$ различны.

9.52. Пусть $\Omega = [0; 1]$, $\mathfrak{F} = \mathfrak{B}_{[0;1]}$, $P = L$. Распределения случайных величин

$$\xi = \xi(\omega) = \begin{cases} 1, & \text{если } \omega \in [0; 1/2]; \\ -1, & \text{если } \omega \in [1/2; 1]; \end{cases}$$

и $\eta = \eta(\omega) = -\xi(\omega)$ совпадают, но $P\{\omega : \xi(\omega) \neq \eta(\omega)\} = 1$.

Примером абсолютно непрерывных случайных величин (на одном и том же вероятностном пространстве), удовлетворяющих условиям задачи являются

$$\zeta = \zeta(\omega) = \omega, \quad \chi = \chi(\omega) = 1 - \omega$$

(обе равномерно распределены на $[0; 1]$).

9.53.

$$\begin{aligned} F_\eta(y) &= P\{\eta < y\} = P\{(\xi, \eta) \in \mathbb{R}^1 \times (-\infty, y)\} = \\ &= \int_{\mathbb{R}^1 \times (-\infty, y)} f_\zeta(u, v) d(u, v) = \int_{-\infty}^y \left(\int_{\mathbb{R}^1} f_\zeta(u, v) du \right) dv, \end{aligned}$$

т.е.

$$F_\eta(y) = \int_{-\infty}^y \left(\int_{\mathbb{R}^1} f_\zeta(u, v) du \right) dv.$$

Последнее равенство означает, что случайная величина η абсолютно непрерывна и функция

$$f_\eta(v) = \int_{\mathbb{R}^1} f_\zeta(u, v) du$$

является ее плотностью.

9.54. Функция $f(t - a)$ является плотностью распределения случайной величины $\xi + a$.

21.10 К главе 10

10.1°. 1) $M\xi = 0$, $D\xi = a^2/3$; 2) $M\xi = (a + b)/2$, $D\xi = (b - a)^2/12$. **10.2°.** 1) 1; 2) $1/2$; 3) $e - 1$.

10.3°. 1) 3; 2) $1 - 2/e$; 3) 3; 4) 0; 5) $e - 1$.

10.4. 1) $(\ln 2)/\pi + 1/2$. **10.5.** $\exp\{\sigma^2/2\}$.

10.6°. Решение к пункту (3). Математическое ожидание функции $\eta = I_{[1/3, 3]}(\xi^2)$ от случайной величины ξ вычисляется по плотности распределения ξ согласно формуле (10.1.1):

$$\begin{aligned} M\eta &= MI_{[1/3, 3]}(\xi^2) = \int_{-\infty}^{+\infty} I_{[1/3, 3]}(x^2) \frac{1}{\pi(1+x^2)} dx = \\ &= 2 \int_0^{+\infty} I_{[1/3, 3]}(x^2) \frac{1}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_{1/\sqrt{3}}^{\sqrt{3}} \frac{1}{1+x^2} dx = \frac{1}{3}. \end{aligned}$$

Ответы: 1) $\sqrt{3}/\pi$; 2) $(\sqrt{3} - 1)/\pi - 1/12$; 3) $1/3$; 4) $1/2$.

10.7. 1) $M\xi = 1/\lambda$; 2) $D\xi = 1/\lambda^2$; 3) $P\{\xi > 1\} = e^{-\lambda}$; 4) $k!/\lambda^k$.

10.8.

$$\begin{aligned} M \min\{\xi, 365\} &= \int_0^{+\infty} \min\{x, 365\} \lambda \exp\{-\lambda x\} dx = \\ &= \lambda \int_0^{365} x \exp\{-\lambda x\} dx + 365 \lambda \int_{365}^{+\infty} \exp\{-\lambda x\} dx = \\ &= (1 - \exp\{-365\lambda\})/\lambda = 222. \end{aligned}$$

10.9. $M\zeta = 1$; $D\zeta = 4/9$.

10.10°. 1) $M\xi = a + 1$; 2) $M\xi^2 = (a + 1)^2 + 1/6$ (см. также пример 10.1.5).

10.11. $1 - 1/e^2$. **10.12.** $M\xi = 0$, $D\xi = 2/\lambda^2$.

10.13. Выражение “точка A равномерно распределена на окружности с центром в точке O ” означает, что угол φ между осью Ox и радиус-вектором $\overline{OA}$ равномерно распределен на отрезке $[0; 2\pi]$. Очевидно, $\xi = |R \operatorname{tg} \varphi|$.

Ответ:

$$p_{\xi}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{2R}{\pi(x^2 + R^2)}, & \text{если } x > 0. \end{cases}$$

Математическое ожидание случайной величины ξ равно $+\infty$.

10.14.

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x^2/R^2, & \text{если } 0 < x \leq R; \\ 1, & \text{если } x > R; \end{cases}$$

$$p(x) = \begin{cases} 0, & \text{если } x \notin (0; R]; \\ 2x/R^2, & \text{если } x \in (0; R]; \end{cases}$$

$$M\eta = 2R/3; \quad D\eta = R^2/18.$$

10.15. Условие “точка A равномерно распределена на окружности” означает, что угол φ между осью Ox и радиус-вектором $\overline{OA}$ равномерно распределен на отрезке $[0; 2\pi]$. Случайная величина ξ равна $\cos \varphi$, поэтому

$$F_{|\xi|}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - \frac{2}{\pi} \arccos x, & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1; \end{cases}$$

$$p_{|\xi|}(x) = \begin{cases} 0, & \text{если } x \notin (0; 1]; \\ \frac{2}{\pi\sqrt{1-x^2}}, & \text{если } x \in (0; 1]; \end{cases}$$

$$M|\xi| = 2/\pi.$$

10.16.

$$1) F(x) = \begin{cases} 0, & \text{если } x \leq a^2; \\ \frac{\sqrt{x} - a}{b - a}, & \text{если } a^2 < x \leq b^2; \\ 1, & \text{если } x > b^2; \end{cases}$$

$$M\eta = \frac{b^3 - a^3}{3(b - a)};$$

$$2) F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \int_0^{\sqrt{x}} \frac{\theta^\nu}{\Gamma(\nu)} t^{\nu-1} e^{-\theta t} dt, & \text{если } x > 0; \end{cases}$$

$$M\eta = \frac{\nu(\nu+1)}{\theta^2};$$

$$3) F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - e^{-\lambda\sqrt{x}}, & \text{если } x > 0; \end{cases}$$

$$M\eta = 2/\lambda^2.$$

10.17.

$$1) F(x) = \begin{cases} 0, & \text{если } x \leq 2\pi a; \\ \frac{x - 2\pi a}{2\pi(b - a)}, & \text{если } 2\pi a < x \leq 2\pi b; \\ 1, & \text{если } x > 2\pi b; \end{cases}$$

$$M\eta = \pi(a + b);$$

$$2) F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \int_0^{x/(2\pi)} \frac{\theta^\nu}{\Gamma(\nu)} t^{\nu-1} e^{-\theta t} dt, & \text{если } x > 0; \end{cases}$$

$$M\eta = 2\pi\nu/\theta;$$

$$3) F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - \exp\left\{-\frac{\lambda}{2\pi}x\right\}, & \text{если } x > 0; \end{cases}$$

$$M\eta = 2\pi/\lambda.$$

$$\mathbf{10.18.} \quad 1) M\eta = \frac{b^4 - a^4}{4(b - a)}; \quad 2) \frac{6}{\lambda^4}.$$

10.19.

$$1(a) \quad F_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x/\pi, & \text{если } 0 < x \leq \pi; \\ 1, & \text{если } x > \pi; \end{cases}$$

$$M\eta = \pi/2;$$

$$1(b) \quad F_\eta(x) = \begin{cases} 0, & \text{если } x \leq \pi; \\ x/\pi - 1, & \text{если } \pi < x \leq 2\pi; \\ 1, & \text{если } x > 2\pi; \end{cases}$$

$$M\eta = 3\pi/2;$$

$$2(a) \quad F_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 2\sqrt{x/\pi}, & \text{если } 0 < x \leq \pi/4; \\ 1, & \text{если } x > \pi/4; \end{cases}$$

$$M\eta = \pi/12;$$

$$2(б) F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq \pi/4; \\ 2\sqrt{x/\pi} - 1, & \text{если } \pi/4 < x \leq \pi; \\ 1, & \text{если } x > \pi; \end{cases}$$

$$M\eta = 7\pi/12;$$

$$3(а) F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 2\sqrt{x}, & \text{если } 0 < x \leq 1/4; \\ 1, & \text{если } x > 1/4; \end{cases}$$

$$M\eta = 1/12;$$

$$3(б) F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 1/4; \\ 2\sqrt{x} - 1, & \text{если } 1/4 < x \leq 1; \\ 1, & \text{если } x > 1; \end{cases}$$

$$M\eta = 7/12;$$

$$4(а) F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 2\sqrt[3]{x}, & \text{если } 0 < x \leq 1/8; \\ 1, & \text{если } x > 1/8; \end{cases}$$

$$M\eta = 1/32;$$

$$4(б) F_{\eta}(x) = \begin{cases} 0, & \text{если } x \leq 1/8; \\ 2\sqrt[3]{x} - 1, & \text{если } 1/8 < x \leq 1; \\ 1, & \text{если } x > 1; \end{cases}$$

$$M\eta = 15/32.$$

10.20. В задачах 1° , 2° , 6° целесообразно сначала найти распределения соответствующих случайных величин, в остальных задачах удобнее воспользоваться мультипликативным свойством математического ожидания.

$1^{\circ}а)$ $1/3$; $2^{\circ}а)$ $2/3$; $2^{\circ}б)$ $13/12$; $3^{\circ}а)$ $1/4$; $3^{\circ}б)$ $1/2$; $3^{\circ}в)$ $1/(2\lambda)$; $3^{\circ}г)$ $1/\lambda^2$; $5^{\circ}в)$ $(e-1)/\lambda$; $6^{\circ}г)$ $1/(\lambda^3(1+\lambda))$.

$$10.21. \alpha + 1/n.$$

$$10.22. 1) \frac{n}{n+1}a + \frac{1}{n+1}b; 2) \frac{n}{n+1}b + \frac{1}{n+1}a; 3) \frac{a+b}{2}.$$

$$10.23. 1) \theta - \frac{n-1}{n+1}h; 2) \theta + \frac{n-1}{n+1}h; 3) \frac{n-1}{n+1}h.$$

$$10.24. 1) M\bar{\xi} = \theta + \alpha; 2) M \min\{\xi_i\} = \theta + \alpha/n;$$

$$3) M\hat{\theta}_1 = \theta + \alpha/n^2; 4) M\hat{\theta}_2 = \alpha(1 - 1/n^2).$$

$$10.25. M\bar{\xi} = \theta.$$

10.26.

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - \frac{(T-x)^2}{T^2}, & \text{если } 0 < x \leq T; \\ 1, & \text{если } x > T. \end{cases}$$

$$p(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{2(T-x)}{T^2}, & \text{если } 0 < x \leq T; \\ 0, & \text{если } x > T. \end{cases}$$

$$M\xi = \frac{T}{3}; \quad D\xi = \frac{T^2}{18}; \quad M\xi^n = \frac{2T^n}{(n+1)(n+2)}.$$

10.27.

$$p(x) = \begin{cases} 0, & \text{если } x \notin (0; 1); \\ \frac{2}{\pi\sqrt{1-x^2}}, & \text{если } x \in (0; 1); \end{cases}$$

$$M\eta = 2/\pi.$$

10.28*. Поскольку при любом повороте окружности вероятность события, зависящая только от расстояния между двумя точками окружности, остается неизменной, то одну из точек можно считать фиксированной.

$$F_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{2}{\pi} \arcsin \frac{x}{2R}, & \text{если } 0 < x \leq 2R; \\ 1, & \text{если } x > 2R. \end{cases}$$

$$\mathbf{10.31.} \quad M\eta^+ = 1/\sqrt{2\pi}. \quad \mathbf{10.32.} \quad M\eta^+ = \sigma/\sqrt{2\pi}.$$

10.33. $M\xi = \frac{1}{\theta}$; $D\xi = \frac{1}{\theta^2}$; $M\xi^2 = \frac{2}{\theta^2}$ (см. также пример 10.1.4).

10.34. $M\xi = \frac{m}{\theta}$; $D\xi = \frac{m}{\theta^2}$; $M\xi^2 = \frac{m(m+1)}{\theta^2}$ (см. также пример 10.1.4).

10.35. $M\xi = n$; $D\xi = 2n$; $M\xi^2 = n(n+2)$ (см. также пример 10.1.4).

10.36. $M\xi = \exp\{\mu + \sigma^2/2\}$; $M\xi^2 = \exp\{2\mu + 2\sigma^2\}$; $D\xi = \exp\{2\mu + \sigma^2\}(\exp\{\sigma^2\} - 1)$.

$$\mathbf{10.37.} \quad M\xi = \frac{\theta\lambda}{1-\theta}; \quad M\xi^2 = \frac{\theta\lambda^2}{2-\theta}.$$

$$\mathbf{10.38.} \quad 1) M\zeta_1 = \frac{3}{2\theta}; \quad D\zeta_1 = \frac{29}{12\theta^2}; \quad 2) M\zeta_2 = \frac{3}{2} + \frac{1}{\theta}; \\ D\zeta_2 = \frac{1}{12} + \frac{1}{\theta^2}; \quad 3) M\zeta_3 = \frac{\ln 2}{\theta}; \quad D\zeta_3 = \frac{\ln^2 2 - 1}{\theta^2}.$$

10.40. Случайная величина ξ является функцией вектора $\zeta = (\xi, \eta)$:

$$\xi = g(\zeta) = g(\xi, \eta) = \xi,$$

поэтому

$$M\xi = Mg(\zeta) = \int_{\mathbb{R}^2} g(x, y) f_{\zeta}(x, y) d(x, y) = \int_{\mathbb{R}^2} x f_{\zeta}(x, y) d(x, y).$$

10.41. Очевидно,

$$\nu_i = \sum_{k=1}^n I_{X_i}(\xi_k), \quad i = 1, 2, \dots, r.$$

Отсюда

$$M\nu_i = np_i, \quad D\nu_i = np_i(1 - p_i), \quad i = 1, 2, \dots, r.$$

21.11 К главе 11

11.1. ξ — абсолютно непрерывная случайная величина, поэтому $\zeta = \xi + \eta$ также абсолютно непрерывная и

$$f_{\zeta}(x) = \int_{\mathbb{R}^1} f_{\xi}(x - y) P(dy) = \frac{1}{2} f_{\xi}(x) + \frac{1}{2} f_{\xi}(x - 1) = \\ = \begin{cases} 0, & \text{если } x < 0; \\ f_{\xi}(x)/2, & \text{если } 0 \leq x < 1; \\ f_{\xi}(x - 1)/2, & \text{если } 0 \leq x - 1 < 1; \\ 0, & \text{если } x - 1 \geq 1. \end{cases}$$

Или

$$f_{\zeta}(x) = \begin{cases} 0, & \text{если } x \notin [0; 2]; \\ 1/2, & \text{если } x \in [0; 2]. \end{cases}$$

Ответ:

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

11.2. Ответ к пункту 1):

$$p_{\zeta}(x) = \begin{cases} 0, & \text{если } x \leq 2a; \\ (x - 2a)/(b - a)^2, & \text{если } 2a < x \leq a + b; \\ (2b - x)/(b - a)^2, & \text{если } a + b < x \leq 2b; \\ 0, & \text{если } x > 2b. \end{cases}$$

11.3.

$$p(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ x/2, & \text{если } 0 < x \leq 1; \\ 1/2, & \text{если } 1 < x \leq 2; \\ (3 - x)/2, & \text{если } 2 < x \leq 3; \\ 0, & \text{если } x > 3. \end{cases}$$

11.4. Сначала найти распределение суммы (разности) соответствующих случайных величин. Другой подход — найти совместную плотность распределения ξ и η , а потом воспользоваться равенством (9.1.3).

Ответы: 1) $1 - 1/(3\sqrt{2})$; 2) $1/4$; 3) $9/16$; 4) $7/16$; 5) $2/3$; 6) $1/2$.

11.5. Ответ к пункту 1):

$$p_{\zeta}(x) = \begin{cases} 0, & \text{если } x \leq -(b - a); \\ (x + (b - a))/(b - a)^2, & \text{если } -(b - a) < x \leq 0; \\ ((b - a) - x)/(b - a)^2, & \text{если } 0 < x < b - a; \\ 0, & \text{если } x \geq b - a. \end{cases}$$

11.6.

$$F_{\zeta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 2x^2, & \text{если } 0 < x \leq 1/2; \\ 1 - 2(1 - x)^2, & \text{если } 1/2 < x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}$$

11.7. Случайная величина ζ распределена равномерно на промежутке $[1; 7]$.

11.8. $1 - (1 - x)^2$. **11.9.** $1/3$.

11.10.

$$1) p(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} (e^{-\lambda_2 x} - e^{-\lambda_1 x}), & \text{если } x > 0; \end{cases}$$

$$2) f(x) = \begin{cases} \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} e^{\lambda_1 x}, & \text{если } x \leq 0; \\ \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_2 x}, & \text{если } x > 0. \end{cases}$$

11.11.

$$1) p(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \lambda^2 x e^{-\lambda x}, & \text{если } x > 0; \end{cases}$$

$$2) p(x) = \begin{cases} \lambda e^{\lambda x} / 2, & \text{если } x \leq 0; \\ \lambda e^{-\lambda x} / 2, & \text{если } x > 0; \end{cases}$$

$$3) p(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ \lambda e^{-\lambda x}, & \text{если } x > 0. \end{cases}$$

11.12. Указание: см. задачу 11.10(1).

11.13. Плотность распределения $f(x)$ суммы $\xi + \eta$

$$\begin{aligned} f(x) &= \frac{1}{4} \int_{-\infty}^{+\infty} e^{-|x-y|} e^{-|y|} dy = \\ &= \frac{1}{4} \int_{-\infty}^0 e^{-|x-y|} e^y dy + \frac{1}{4} \int_0^{+\infty} e^{-|x-y|} e^{-y} dy. \end{aligned}$$

Сделаем в интегралах замену $x - y = t$. Получим

$$f(x) = \frac{1}{4} \left(e^x \int_x^{+\infty} e^{-|t|} e^{-t} dt + e^{-x} \int_{-\infty}^x e^{-|t|} e^t dt \right).$$

Рассмотрим два случая: $x < 0$ и $x > 0$.

При $x < 0$ имеем $f(x) =$

$$= \frac{1}{4} \left(e^x \left(\int_x^0 e^t e^{-t} dt + \int_0^{+\infty} e^{-2t} dt \right) + e^{-x} \int_{-\infty}^x e^{-|t|} e^t dt \right) =$$

$$= e^x(1-x)/4.$$

При $x > 0$

$$\begin{aligned} f(x) &= \frac{1}{4} \left(e^x \int_x^{+\infty} e^{-2t} dt + e^{-x} \left(\int_{-\infty}^0 e^{2t} dt + \int_0^x e^{-t} e^t dt \right) \right) = \\ &= e^{-x}(1+x)/4. \end{aligned}$$

Аналогично находим плотность $g(x)$ распределения разности $\xi - \eta$

Так что

$$1) f(x) = \begin{cases} e^x(1-x)/4, & \text{если } x \leq 0; \\ e^{-x}(1+x)/4, & \text{если } x > 0; \end{cases}$$

$$2) g(x) = \begin{cases} e^x(1-x)/4, & \text{если } x \leq 0; \\ e^{-x}(1+x)/4, & \text{если } x > 0. \end{cases}$$

11.14.

$$p_\zeta(x) = \begin{cases} 0, & \text{если } x \leq -a; \\ (1 - e^{-\lambda(x+a)})/2a, & \text{если } -a < x \leq a; \\ (e^{-\lambda(x-a)} - e^{-\lambda(x+a)})/2a, & \text{если } x > a. \end{cases}$$

11.16. Функция распределения суммы $\zeta = \xi + \eta$ независимых случайных величин ξ и η равна свертке функций распределения $F(x)$ и $H(x)$ слагаемых ξ и η ($H(x)$ — функция распределения случайной величины η), т. е.

$$F_\zeta(x) = \int_{\mathbb{R}^1} F(x-y)H(dy) = \int_{\mathbb{R}^1} H(x-y)F(dy).$$

Поскольку распределение H имеет плотность

$$h(y) = \begin{cases} 1/2h, & \text{если } y \in [-h; h]; \\ 0, & \text{если } y \notin [-h; h], \end{cases}$$

то

$$\begin{aligned} F_{\zeta}(x) &= \int_{\mathbb{R}^1} F(x-y)H(dy) = \int_{\mathbb{R}^1} F(x-y)h(y)dy = \\ &= \frac{1}{2h} \int_{-h}^h F(x-y)dy = \frac{1}{2h} \int_{x-h}^{x+h} F(t)dt \end{aligned}$$

(воспользовались заменой $x-y=t$).

Далее, распределение случайной величины η абсолютно непрерывно, поэтому распределение суммы $\zeta = \xi + \eta$ также абсолютно непрерывно и его плотность $f_{\zeta}(x)$ равна свертке плотности распределения $h = h(t)$ случайной величины η с распределением случайной величины ξ :

$$f_{\zeta}(x) = \int_{\mathbb{R}^1} h(x-y)F(dy).$$

По определению функции $h(t)$

$$h(x-y) = \begin{cases} 0, & \text{если } x-y < -h; \\ 1/(2h), & \text{если } -h \leq x-y \leq h; \\ 0, & \text{если } x-y > h, \end{cases}$$

или

$$h(x-y) = \begin{cases} 0, & \text{если } y < x-h; \\ 1/(2h), & \text{если } x-h \leq y \leq x+h; \\ 0, & \text{если } y > x+h. \end{cases}$$

Поэтому

$$f_{\zeta}(x) = \int_{\mathbb{R}^1} h(x-y)F(dy)$$

как интеграл от простой функции, принимающей значение $1/(2h)$ на промежутке $[x-h, x+h]$ и значение 0 вне

промежутка $[x - h; x + h]$, равен $F([x - h; x + h])/(2h)$. Следовательно,

$$f_{\zeta}(x) = \frac{1}{2h} F([x - h; x + h]).$$

11.17. Функция распределения суммы $\zeta = \xi + \eta$ независимых случайных величин ξ и η равна свертке функций распределения слагаемых, т. е.

$$F_{\zeta}(x) = \int_{\mathbb{R}^1} Q(x - y)G(dy).$$

Учитывая, что G — дискретное распределение, сосредоточенное в точках -1 и 1 (с “массой” $1/2$ в каждой), имеем

$$F_{\zeta}(x) = \int_{\mathbb{R}^1} Q(x - y)G(dy) = \frac{1}{2}Q(x - 1) + \frac{1}{2}Q(x + 1).$$

11.18. Сумма независимых случайных величин, одна из которых абсолютно непрерывна, является абсолютно непрерывной случайной величиной, и ее плотность равна свертке плотности абсолютно непрерывной случайной величины с распределением другой (см. также задачу 11.17):

$$p_{\zeta}(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 3/4, & \text{если } 0 < x \leq 1; \\ 1/4, & \text{если } 1 < x \leq 2; \\ 0, & \text{если } x > 2. \end{cases}$$

11.19.

$$f(x) = \begin{cases} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, & \text{если } x > 0; \\ 0, & \text{если } x \leq 0. \end{cases}$$

11.20. Чтобы найти распределение η , воспользуйтесь тем, что $\ln(1/\xi_i)$ имеет показательное распределение.

$$M\eta = M\xi_1\xi_2 \dots \xi_n = M\xi_1 M\xi_2 \dots M\xi_n = (1/2)^n$$

21.12 К главе 12

12.2. При $h \rightarrow 0$, распределение F_h сходится к собственному атомическому распределению, сосредоточенному в точке a .

12.3—12.6. Последовательность распределений $\{F_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

12.7. Последовательность распределений $\{F_n\}$ собственно сходится к F ; последовательности распределений $\{G_n\}$, $\{S_n\}$, $\{P_n\}$, $\{Q_n\}$ сходятся к несобственному распределению тождественно равному нулю (см. также пример 12.1.3).

12.8. Последовательность распределений $\{P_n\}$ собственно сходится к атомическому распределению, сосредоточенному в точке 1.

12.9. Воспользуйтесь достаточным условием 1° сходимости вероятностных распределений (см. также пример 12.1.3).

Ответы: 1) последовательность распределений $\{F_n\}$ сходится к несобственному распределению, тождественно равному нулю; 2) последовательность распределений $\{G_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0.

12.10. Последовательность распределений $\{F_n\}$ не является сходящейся.

12.11*. $I_{(-\infty; y)}(x)$ для каждой пары (x, y) , $x \neq y$.

12.12. $\sin 1$. **12.13.** $\cos 1$.

12.14*. Воспользоваться 12.11*.

Ответ: $F(-\infty; y)$.

12.15*. Достаточное условие 2° сходимости распределений можно сформулировать в более общем виде (см. пример 12.1.6).

Собственные вероятностные распределения F_{a, σ^2} со средним $a \rightarrow a_0$ и дисперсией $\sigma^2 \rightarrow 0$ сходятся к собственному атомическому распределению, сосредоточенному в точке a_0 .

Математическое ожидание и дисперсия пуассоновского распределения равны λ , поэтому, если $\lambda \rightarrow 0$, пуассоновское распределение сходится к собственному атоми-

ческому распределению, сосредоточенному в точке 0 (его функцией распределения является $I_{(0,\infty)}(y)$). Из сходимости последовательности распределений следует сходимость последовательности функций распределений, поэтому функция

$$F_\lambda(y) = \sum_{k:k < y} \frac{\lambda^k}{k!} e^{-\lambda}$$

пуассоновского распределения с параметром λ , при $\lambda \rightarrow 0$, сходится к $I_{(0,\infty)}(y)$.

12.16*. Ответ: 1. Воспользуйтесь решением задачи 12.15*.

12.17. Последовательность $\{F_n\}$ сходится к собственному атомическому распределению, сосредоточенному в точке 0, поскольку $a_n \rightarrow 0$ и $\sigma_n \rightarrow 0$ (см. пример 12.1.6).

21.13 К главе 13

13.1. $\cos t$. **13.2.** $(e^{it} + e^{-it} + 1)/3$.

13.3. Указание. См. задачи 13.1, 13.5.

13.4. См. задачу 13.1.

13.5. $\cos^n z$ (см. задачу 13.1).

13.6. См. задачу 13.5.

13.7. Как известно, случайную величину ξ , имеющую биномиальное распределение с параметрами $(n; p)$, можно представить в виде суммы

$$\xi = \sum_{k=1}^n \mu_k$$

независимых случайных величин $\mu_1, \mu_2, \dots, \mu_n$, каждая из которых имеет распределение

$$P\{\mu_k = s\} = p^s(1-p)^{1-s}, \quad s = 0, 1; \quad k = 1, 2, \dots, n.$$

Отсюда, пользуясь мультипликативным свойством характеристических функций, получаем

$$\varphi_{\xi}(t) = \prod_{k=1}^n \varphi_k(t),$$

где $\varphi_k(t) = Me^{it\mu_k} = e^{it \cdot 1}p + e^{it \cdot 0}(1-p) = pe^{it} + 1 - p = 1 + p(e^{it} - 1)$. Следовательно,

$$\varphi_{\xi}(t) = (1 + p(e^{it} - 1))^n.$$

$$13.8. \varphi(t) = \frac{p}{1 - e^{it}(1-p)}.$$

13.9. Найти характеристическую функцию $\varphi_{\xi}(t)$ случайной величины ξ . Воспользоваться тем, что характеристическая функция $\varphi_{\eta}(t)$ случайной величины $\eta = a + \sigma\xi$ равна $e^{ita}\varphi_{\xi}(\sigma t)$.

Ответ: $\exp\{-\lambda - it\sqrt{\lambda} + \lambda \exp\{it/\sqrt{\lambda}\}\}$.

$$13.10. \varphi(t) = \frac{e^{ita} - e^{-ita}}{2ita}. \quad 13.11. \varphi(t) = \frac{e^{itb} - e^{ita}}{i(b-a)t}.$$

$$13.12. \varphi(t) = \frac{1}{1+t^2}. \quad 13.13. \varphi(t) = \frac{a}{a-it}.$$

$$13.14. \varphi(z) = \frac{a^2}{a^2+z^2}. \quad 13.15. \varphi(z) = 2\frac{1-\cos az}{a^2z^2}.$$

13.16*. Сначала вычислим характеристическую функцию двустороннего показательного распределения с параметром a . Его плотность

$$f(x) = \frac{a}{2}e^{-a|x|}, \quad x \in \mathbb{R}^1,$$

а характеристическая функция

$$\varphi(t) = \int_{\mathbb{R}^1} e^{itx} f(x) dx = \int_{-\infty}^{+\infty} e^{itx} \frac{a}{2} e^{-a|x|} dx = \frac{a^2}{a^2 + t^2}.$$

Дальше воспользуемся утверждением, известным под названием теоремы обращения для характеристических функций.

Если характеристическая функция $\varphi(t)$ вероятностного распределения F интегрируема на $\mathbb{R}^1$, то распределение F абсолютно непрерывно, и его плотность $f(x)$ представима в виде

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}^1} e^{-ixt} \varphi(t) dt. \quad (21.13.1)$$

Характеристическая функция $\varphi(t) = \frac{a^2}{a^2 + t^2}$ двустороннего показательного распределения интегрируема на $\mathbb{R}^1$, поэтому имеет место равенство (21.13.1):

$$\frac{a}{2} e^{-a|x|} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixt} \frac{a^2}{a^2 + t^2} dt,$$

которое можно переписать так:

$$e^{-a|x|} = \int_{-\infty}^{+\infty} e^{ixt} \frac{1}{\pi} \frac{a}{a^2 + t^2} dt.$$

Остается заметить, что последний интеграл — не что иное, как характеристическая функция распределения Коши с параметром a .

13.17. Свертка $F * Q$ является распределением Коши с параметром $(a + b)$.

13.18. Сумма $\xi + \eta$ распределена нормально с параметрами $(a_1 + a_2; \sigma_1^2 + \sigma_2^2)$.

13.19. Сумма $\xi + \eta$ имеет распределение Коши с параметром $(a + b)$.

13.20. Сума $\xi + \eta$ имеет пуассоновское распределение с параметром $\lambda_1 + \lambda_2$.

13.21. $\varphi(t) = 1 + ita - \frac{1}{2}(\sigma^2 + a^2)t^2 + o(t^2)$.

13.23. Найти характеристическую функцию случайной величины $(\xi_\lambda - \lambda)/\sqrt{\lambda}$ и исследовать ее сходимость при $\lambda \rightarrow \infty$ (см. также задачу 13.9).

13.24. Воспользуйтесь теоремой непрерывности.

13.25.

$$1^\circ \psi_n(t) = \left(\varphi \left(\frac{t}{\sigma\sqrt{n}} \right) \right)^n;$$

$$2^\circ \varphi(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o(t^2);$$

$$3^\circ \lim_{n \rightarrow \infty} \psi_n(t) = \exp \{-t^2/2\}.$$

13.26. Обозначьте через $\varphi(t)$ характеристическую функцию случайной величины ξ_i ($i = 1, 2, \dots$), а через $\psi_n(t)$ — характеристическую функцию случайной величины $S_n = \frac{1}{n} \sum_{i=1}^n \xi_i$. Последовательно решите следующие задачи:

1° Выразите характеристическую функцию $\psi_n(t)$ через характеристическую функцию $\varphi(t)$.

2° Выпишите разложение Тейлора для $\varphi(t)$ в окрестности точки $t = 0$, включающее t .

3° Воспользовавшись решениями задачи из пунктов 1° и 2°, найдите предел характеристической функции $\psi_n(t)$ при $n \rightarrow \infty$.

4° Воспользовавшись решением задачи из пункта 3° и теоремой непрерывности, докажите, что если $n \rightarrow \infty$, то распределение случайной величины S_n сходится к собственному атомическому распределению, сосредоточенному в точке a .

5° Докажите, что из сходимости распределений случайных величин S_n к собственному атомическому распределению, сосредоточенному в точке a , следует сходимость по вероятности случайных величин S_n к a .

В самом деле, сходимость последовательности $\{F_n\}$ распределений случайных величин S_n/n к собственному атомическому распределению F_0 , сосредоточенному в точке a , равносильна сходимости функций распределений $F_n(x)$ к функции распределения $F_0(x)$, т. е.

$$F_n(x) = P\{|S_n/n < x|\} \rightarrow F_0(x) = \begin{cases} 0, & \text{если } x \leq a; \\ 1, & \text{если } x > a. \end{cases}$$

Далее несложно убедиться, что при $n \rightarrow \infty$ из сходимости $F_n(x) \rightarrow F_0(x)$ следует

$$P\{|S_n - a| \geq \varepsilon\} \rightarrow 0.$$

13.29. Найдите характеристическую функцию случайной величины $\sum_{k=1}^n \xi_{k,n}$ и воспользуйтесь теоремой непрерывности.

13.30. Представьте случайную величину ν_i в виде

$$\nu_i = \sum_{k=1}^n I_{X_i}(\xi_k), \quad i = 1, 2, \dots, r.$$

и воспользуйтесь центральной предельной теоремой.

13.31. Последовательность распределений $N_{(-1)^n; 1/n^2}$ не является сходящейся.

21.14 К главе 14

14.1. Функция $L(\theta, h)$ достигает наибольшего значения в точке

$$(\hat{\theta}, \hat{h}) = ((\max\{\xi_i\} + \min\{\xi_i\})/2, (\max\{\xi_i\} - \min\{\xi_i\})/2).$$

Учитывая, что случайные величины $\xi_1, \xi_2, \dots, \xi_n$ — независимы и каждая из них имеет своей плотностью $f(x; \theta, h)$, находим плотности случайных величин $\max\{\xi_i\}$, $\min\{\xi_i\}$; $M \max\{\xi_i\}$, $M \min\{\xi_i\}$, $P\text{-}\lim_{n \rightarrow \infty} \max\{\xi_i\}$, $P\text{-}\lim_{n \rightarrow \infty} \min\{\xi_i\}$

и убеждаемся, что $\hat{\theta}$ — несмещенная и состоятельная оценка θ , а $\hat{h}$ — асимптотически несмещенная и состоятельная оценка h .

14.2. $\hat{\theta} = \frac{1}{rn} \sum_{i=1}^n \xi_i$ — несмещенная и состоятельная оценка θ .

14.3. Да.

14.4. Относительно оценок $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ см. пример 14.1.1; $\hat{\theta}_5$ — асимптотически несмещенная и состоятельная оценка параметра $b-a$; $\hat{\theta}_6, \hat{\theta}_7, \hat{\theta}_8$ — несмещенные и состоятельные оценки соответственно параметров $(a+b)/2, a, b$.

14.5. $\hat{\lambda}$ — несмещенная и состоятельная оценка λ .

14.6. Да.

14.7. Функция

$$L(\theta) = \begin{cases} 1/2^n, & \text{если все } \xi_i \in [\theta - 1, \theta + 1]; \\ 0, & \text{если существует } \xi_j \notin [\theta - 1, \theta + 1] \end{cases}$$

принимает два значения: $1/2^n$ и 0. Наибольшее значение, равное $1/2^n$, функция $L(\theta)$ принимает, если

$$\theta - 1 \leq \xi_i \leq \theta + 1, i = 1, 2, \dots, n,$$

или, что то же, если

$$\max\{\xi_i\} - 1 \leq \theta \leq \min\{\xi_i\} + 1,$$

другими словами, — в точках

$$\hat{\theta}_\lambda = \lambda(\max\{\xi_i\} - 1) + (1 - \lambda)(\min\{\xi_i\} + 1), \lambda \in [0; 1],$$

промежутка $[\max\{\xi_i\} - 1, \min\{\xi_i\} + 1]$.

Относительно несмещенности и состоятельности $\hat{\theta}_\lambda$ см. пример 14.1.1.

14.8. $\hat{\theta}_1, \hat{\theta}_2$ — асимптотически несмещенные и состоятельные оценки соответственно параметров θ, α .

14.9. Да.

14.10. $\hat{\theta}$ — несмещенная и состоятельная оценка θ .

14.11. $\hat{a}$ — несмещенная и состоятельная оценка a , $\hat{\sigma}^2$ — асимптотически несмещенная и состоятельная оценка σ^2 .

14.12. Да.

14.13. $L(\theta)$ достигает наибольшего значения в точках

$$\hat{\theta}_\lambda = \lambda(\max\{\xi_i\} - h_0) + (1 - \lambda)(\min\{\xi_i\} + h_0), \lambda \in [0; 1],$$

промежутка $[\max\{\xi_i\} - h_0, \min\{\xi_i\} + h_0]$. См. также задачу 14.7.

14.14.

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right)^2.$$

$\hat{\theta}$ — несмещенная и состоятельная оценка θ ; $\hat{\sigma}^2$ — асимптотически несмещенная и состоятельная оценка σ^2 .

14.15. Да.

14.16. $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_\lambda$ — асимптотически несмещенные и состоятельные оценки параметров $\theta - h_0, \theta + h_0, \theta$ соответственно; $\hat{\theta}_3$ — несмещенная и состоятельная оценка θ .

14.17.

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \xi_i \right)^2.$$

$\hat{a}$ — несмещенная и состоятельная оценка a ; $\hat{\sigma}^2$ — асимптотически несмещенная и состоятельная оценка σ^2 .

14.18. Да.

14.19. Функция $L(a, b)$ достигает наибольшего значения в точке $(\hat{a}, \hat{b}) = (\min\{\xi_i\}, \max\{\xi_i\})$; $\hat{a}$ и $\hat{b}$ — асимптотически несмещенные и состоятельные оценки соответственно параметров a и b .

14.20. $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i$ — несмещенная и состоятельная оценка параметра θ .

14.21. Да.

14.22. $\hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ — несмещенные и состоятельные оценки соответственно параметров $\theta, \theta - h, \theta + h$; $\hat{\theta}_1$ — асимптотически несмещенная и состоятельная оценка h .

14.23. $\hat{p} = \frac{1}{n} \sum_{i=1}^n \xi_i$ — несмещенная и состоятельная оценка параметра $(1 - p)/p$.

14.24. Да.

14.25. $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ — асимптотически несмещенные и состоятельные оценки соответственно параметров $h, \theta - h_0, \theta + h_0, h$. Относительно оценок $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$ см. пример 14.1.1.

Для исследования оценки $\hat{\theta}_4$, сначала найдем ее функцию распределения $G(x)$.

Поскольку случайная величина

$$\hat{\theta}_4 = \max\{\theta_0 - \min\{\xi_i\}, \max\{\xi_i\} - \theta_0\}$$

неотрицательно, то при $x \leq 0$ значения

$$G(x) = P\{\hat{\theta}_4 < x\} = 0.$$

При $x > 0$, учитывая независимость $\xi_1, \xi_2, \dots, \xi_n$, имеем

$$\begin{aligned} G(x) &= P\{\hat{\theta}_4 < x\} = \\ &= P\{\max\{\theta_0 - \min\{\xi_i\}, \max\{\xi_i\} - \theta_0\} < x\} = \\ &= P\{\theta_0 - x < \min\{\xi_i\}, \max\{\xi_i\} < \theta_0 + x\} = \\ &= P\left\{\bigcap_{n=1}^n \{\theta_0 - x < \xi_i < \theta_0 + x\}\right\} = \\ &= \prod_{n=1}^n P\{\theta_0 - x < \xi_i < \theta_0 + x\} = \\ &= (P\{\theta_0 - x < \xi_i < \theta_0 + x\})^n = (F(\theta_0 + x) - F(\theta_0 - x))^n, \end{aligned}$$

где

$$F(t) = \begin{cases} 0, & \text{если } t \leq \theta_0 - h; \\ (t - (\theta_0 - h))/2h, & \text{если } \theta_0 - h < t \leq \theta_0 + h; \\ 1, & \text{если } t > \theta_0 + h \end{cases}$$

— функция равномерного на отрезке $[\theta_0 - h, \theta_0 + h]$ распределения. При $x > 0$

$$\begin{aligned} F(\theta_0 + x) &= \begin{cases} (x + h)/2h, & \text{если } 0 < x \leq h; \\ 1, & \text{если } x > h; \end{cases} \\ F(\theta_0 - x) &= \begin{cases} (h - x)/2h, & \text{если } 0 < x \leq h; \\ 0, & \text{если } x > h. \end{cases} \end{aligned}$$

После несложных преобразований получим

$$p(x) = \frac{d}{dx}G(x) = \begin{cases} nx^{n-1}/h^n, & \text{если } x \in [0, h]; \\ 0, & \text{если } x \notin [0, h]. \end{cases}$$

По известной плотности распределения оценки $\hat{\theta}_4$, найдем

$$M\hat{\theta}_4 = \frac{n}{n+1}h,$$

$$P\{|\theta_4 - h| > \varepsilon\} = \left(1 - \frac{\varepsilon}{h}\right)^n, \quad (0 < \varepsilon < h).$$

Поэтому $\hat{\theta}_4$ — асимптотически несмещенная и состоятельная оценка h .

14.26. $\hat{\theta}_2$ — несмещенная оценка параметра α , но не является его состоятельной оценкой.

14.27. Да.

14.28. $\hat{\theta}_1$ — несмещенная и состоятельная оценка $\theta + b$; $\hat{\theta}_2, \hat{\theta}_3$ — асимптотически несмещенные и состоятельные оценки соответственно параметров b и θ .

14.29. $\hat{\theta} = \frac{1}{mn} \sum_{i=1}^n \xi_i$ — несмещенная и состоятельная оценка θ .

14.30. Да.

14.31. $L(h)$ достигает наибольшего значения в точке

$$\hat{h} = \max\{\max\{\xi_i\} - \theta_0, \theta_0 - \min\{\xi_i\}\},$$

и $\hat{h}$ является асимптотически несмещенной и состоятельной оценкой h (см. также задачу 14.25).

14.32. $\hat{\theta}$ — несмещенная и состоятельная оценка σ^2 .

14.33. Да.

14.34. $\hat{\theta}_2$ — несмещенная и состоятельная оценка параметра $(a + b)/2$; $\hat{\theta}_1, \hat{\theta}_3$ — асимптотически несмещенные и состоятельные оценки b .

14.35. Да.

21.15 К главе 15

15.1. 0, 3. **15.2.** 0, 18. **15.3.** $\hat{\mu} = 2 \ln \hat{m}_1 - (\ln \hat{m}_2)/2$.

15.4. Оценкой максимального правдоподобия θ является

$$\hat{\theta}_\lambda = \lambda(\max\{\xi_i\} - h_0) + (1 - \lambda)(\min\{\xi_i\} + h_0), \quad \lambda \in [0; 1];$$

$\hat{\theta}_\lambda$ — асимптотически несмещенная и состоятельная оценка параметра θ ; $\hat{\theta}_{1/2}$ — несмещенная и состоятельная оценка θ .

15.7. $\hat{b} = \hat{m}_1$, $\hat{\theta} = (\hat{m}_2 + (\hat{m}_1)^2)/2$; $\hat{b}$ — несмещенная и состоятельная оценка параметра b ; $\hat{\theta}$ — асимптотически несмещенная и состоятельная оценка параметра θ .

15.8. $\hat{\theta} = \frac{1}{rn} \sum_{i=1}^n \xi_i$; $\hat{\theta}$ — несмещенная, состоятельная и эффективная оценка θ .

15.10. $5, 4\pi$. **15.11.** $\hat{p} = 1/(\bar{\xi} + 1)$.

15.12. Функцией максимального правдоподобия является

$$L(a, b) = \frac{1}{(2a)^n} \exp \left\{ -\frac{1}{a} \sum_{i=1}^n |\xi_i - b| \right\}, \quad a > 0.$$

Если $\hat{b}$ — точка, в которой $Q(b) = \sum_{i=1}^n |\xi_i - b|$ достигает наименьшего значения, а $\hat{a}$ — точка, в которой $L(a, \hat{b})$ достигает наибольшего значения, то в точке $(\hat{a}, \hat{b})$ функция $L(a, b)$ достигает наибольшего значения. Это утверждение проверяется непосредственно.

Далее, пусть $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ — вариационный ряд выборки $\xi_1, \xi_2, \dots, \xi_n$. При $n = 2k$, функция $Q(b)$ достигает наименьшего значения в каждой точке промежутка $[\xi_k^*, \xi_{k+1}^*]$, поэтому оценкой b является

$$\hat{b}_\lambda = \lambda \xi_k^* + (1 - \lambda) \xi_{k+1}^*,$$

$\lambda \in [0; 1]$. Если $n = 2k + 1$, функция $Q(b)$ достигает наименьшего значения в точке $\hat{b} = \xi_{k+1}^*$. Это хорошо видно из графиков функции $y = Q(b)$ (при четном и нечетном n).

Обозначим $Q^* = Q(\hat{b})$. Функция

$$L(a, \hat{b}) = \frac{1}{(2a)^n} \exp \left\{ -\frac{1}{a} Q^* \right\}$$

достигает наибольшего значения в точке $\hat{a} = Q^*/n$. Следовательно, оценкой максимального правдоподобия $(\hat{a}, \hat{b})$ параметра (a, b) является $(Q^*/n, \hat{b})$.

15.14. 158, 3.

15.15. Для случайной величины с плотностью распределения

$$f(x; a, \theta) =$$

$$= \begin{cases} (x - (a - \sqrt{\theta}))/\theta, & \text{если } x \in [a - \sqrt{\theta}, a]; \\ -(x - (a + \sqrt{\theta}))/\theta, & \text{если } x \in [a, a + \sqrt{\theta}]; \\ 0, & \text{если } x \notin [a - \sqrt{\theta}, a + \sqrt{\theta}] \end{cases}$$

имеем

$$m_1 = m_1(a, \theta) = a, \quad m_2 = m_2(a, \theta) = a^2 + \theta/6.$$

Согласно методу моментов оценки $\hat{a}$ и $\hat{\theta}$ получаем как решение системы

$$\begin{cases} \hat{m}_1 = \hat{a}; \\ \hat{m}_2 = \hat{a}^2 + \hat{\theta}/6. \end{cases}$$

Отсюда

$$\hat{a} = \hat{m}_1; \quad \hat{\theta} = 6(\hat{m}_2 - (\hat{m}_1)^2).$$

Из теоремы 15.1.1 следует, что $\hat{a}$ и $\hat{\theta}$ — состоятельные оценки соответственно параметров a и θ ; $\hat{a}$ — несмещенная оценка a . Поскольку выборочная дисперсия $\hat{\sigma}^2$ является асимптотически несмещенной оценкой σ^2 , то $\hat{\theta}$ является асимптотически несмещенной оценкой θ .

15.16. Пусть a — высота цели, ξ — высота выстрела — нормально распределенная с параметрами $(h; 1)$ случайная величина (h неизвестно). Событие $\{\xi \geq a\}$ будем считать успехом. Вероятность успеха

$$p(h) = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} \exp\left\{-\frac{(t-h)^2}{2}\right\} dt.$$

Наблюдается случайная величина η — число успехов в n испытаниях, при этом число таких наблюдений, т. е.

объем выборки, равен единице. Распределением η является

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

Функция максимального правдоподобия

$$L(p(h)) = C_n^k (p(h))^k (1 - p(h))^{n-k}.$$

Логарифмическое уравнение максимального правдоподобия

$$k \ln p(h) + (n - k) \ln(1 - p(h)) = 0.$$

Решая его относительно $p(h)$, получим

$$p(h) = k/n,$$

или

$$\frac{k}{n} = \frac{1}{\sqrt{2\pi}} \int_{a-h}^{+\infty} \exp\left\{-\frac{u^2}{2}\right\} du. \quad (21.15.1)$$

Условие к установке орудия: $|a - h|$ принимает минимальное значение, другими словами, $a - h = 0$. Поэтому из равенства (21.15.1) имеем

$$\frac{k}{n} = \frac{1}{2}.$$

Следовательно, орудие необходимо установить так, чтобы количество перелетов было равным количеству недолетов.

15.20. Оценкой максимального правдоподобия h является

$$\hat{h} = \max \{ \max\{\xi_i\} - \theta_0, \theta_0 - \min\{\xi_i\} \}.$$

Функция распределения $\hat{h}$

$$\begin{aligned} F(x) &= P\{\hat{h} < x\} = \\ &= P\{\max\{\max\{\xi_i\} - \theta_0, \theta_0 - \min\{\xi_i\}\} < x\} = \end{aligned}$$

$$\begin{aligned}
&= P\{\max\{\xi_i\} - \theta_0 < x, \theta_0 - \min\{\xi_i\} < x\} = \\
&= P\left(\left(\bigcap_{i=1}^n \{\xi_i < \theta_0 + x\}\right) \cap \left(\bigcap_{i=1}^n \{\xi_i > \theta_0 - x\}\right)\right) = \\
&= P\left(\bigcap_{i=1}^n \{\theta_0 - x < \xi_i < \theta_0 + x\}\right) = \\
&= \prod_{i=1}^n P\{\theta_0 - x < \xi_i < \theta_0 + x\} = (P\{\theta_0 - x < \xi_i < \theta_0 + x\})^n.
\end{aligned}$$

И поскольку $\xi_i, i = 1, 2, \dots, n$, распределены равномерно на $[\theta_0 - h, \theta_0 + h]$, то

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ (x/h)^n, & \text{если } 0 < x \leq h; \\ 1, & \text{если } x > h. \end{cases}$$

Отсюда

$$M\hat{h} = \frac{n}{n+1}h \rightarrow h,$$

при $n \rightarrow \infty$, т. е. $\hat{h}$ является асимптотически несмещенной оценкой h . Для достаточно малых $\varepsilon > 0$

$$\begin{aligned}
&P\{|\hat{h} - h| < \varepsilon\} = P\{h - \varepsilon < \hat{h} < h + \varepsilon\} = \\
&= P\{h - \varepsilon < \hat{h}\} = 1 - P\{\hat{h} < h - \varepsilon\} = 1 - \left(\frac{h - \varepsilon}{h}\right)^n \rightarrow 1,
\end{aligned}$$

при $n \rightarrow \infty$, т. е. $\hat{h}$ — состоятельная оценка h .

$$15.22. \hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^n \xi_i^p\right)^{1/p}.$$

15.23. $\hat{\theta} = \hat{m}_1$, $\hat{\sigma}^2 = \hat{m}_2 - (\hat{m}_1)^2$; $\hat{\theta}$ — несмещенная и состоятельная оценка θ ; $\hat{\sigma}^2$ — асимптотически несмещенная и состоятельная оценка σ^2 .

$$15.24. \hat{\theta} = \frac{1}{2n} \sum_{i=1}^n \xi_i^2; \hat{\theta} — несмещенная, состоятельная$$

и эффективная оценка параметра θ .

$$15.26. 15, 52 \cdot \pi.$$

$$15.27. \hat{m} = 1/(\hat{m}_2/(\hat{m}_1)^2 - 1); \hat{\lambda} = 1/(\hat{m}_2/\hat{m}_1 - m_1).$$

15.28.

1) $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i$; $\hat{\theta}$ — несмещенная, состоятельная, эффективная оценка параметра θ .

2) $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{2}$; $\hat{\theta}$ — несмещенная, состоятельная, эффективная оценка параметра θ .

3) $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \xi_i$; $\hat{\theta}$ — несмещенная, состоятельная, эффективная оценка параметра θ .

15.31. $\hat{\nu} = 1/(\hat{m}_2/\hat{m}_1^2 - 1)$; $\hat{\alpha} = 1/(\hat{m}_2/\hat{m}_1 - \hat{m}_1)$.

15.32. $\hat{\theta} = \hat{m}_1$ — несмещенная, состоятельная и эффективная оценка θ .

15.34. $0,56\pi$, **15.35.** $\hat{p} = \left(1 + \frac{1}{rn} \sum_{i=1}^n \xi_i\right)^{-1}$.

15.36. $\hat{N} = (nK/k)$.

15.39. $\hat{a} = \sqrt{\hat{m}_2 - (\hat{m}_1)^2}$; $\hat{b} = \hat{m}_1 - \hat{a}$.

15.40. Поскольку $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, то их совместная плотность распределения равна $\prod_{i=1}^n f(x_i; a, b)$, а функция максимального правдоподобия

$$L(a, b) = \prod_{i=1}^n f(\xi_i; a, b) =$$

$$= \begin{cases} 1/(b-a)^n, & \text{если все } \xi_i \in [a, b]; \\ 0, & \text{если существует } \xi_j \notin [a, b]. \end{cases}$$

Функция $L(a, b)$ достигает максимального значения в точке $(\hat{a}, \hat{b})$, где разность $(b-a)$ достигает минимального значения, последняя же достигает минимального значения при

$$a = \hat{a} = \min\{\xi_i\}; \quad b = \hat{b} = \max\{\xi_i\}.$$

$\hat{a}, \hat{b}$ — асимптотически несмещенные и состоятельные оценки соответственно параметров a, b (см. также пример 14.1.1).

15.41. Время указать в часах,

$$\sup_x \left| F(x) - \hat{F}_{14}(x) \right| = 0,306.$$

15.42. 0,09.

15.43. $\hat{p} = \frac{1}{mn} \sum_{i=1}^n \xi_i$; $\hat{p}$ — несмещенная, состоятельная и эффективная оценка p .

15.44. Оценкой параметра (θ, h) является

$$(\hat{\theta}, \hat{h}) = ((\max\{\xi_i\} + \min\{\xi_i\})/2; (\max\{\xi_i\} - \min\{\xi_i\})/2);$$

$\hat{\theta}, \hat{h}$ — асимптотически несмещенные и состоятельные оценки соответственно параметров a и b .

$$\mathbf{15.45.} \sup_x \left| F(x) - \hat{F}_n(x) \right| = 0,287.$$

15.47. $\hat{\theta} = \hat{m}_1$, $\hat{h} = \hat{m}_2 - (\hat{m}_1)^2$, $\hat{\theta}$ — несмещенная и состоятельная оценка θ ; $\hat{h}$ — асимптотически несмещенная и состоятельная оценка h .

$$\mathbf{15.48.} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln \xi_i; \quad \hat{\sigma}^2 = \frac{2}{n} \sum_{i=1}^n \left(\ln \xi_i - \sum_{i=1}^n \ln \xi_i \right)^2.$$

$$\mathbf{15.49.} \hat{\theta} = 1 + \frac{1}{rn} \sum_{i=1}^n \xi_i.$$

15.50. $\hat{b} = \max\{\xi_i\}$; $\hat{b}$ — асимптотически несмещенная и состоятельная оценка b .

15.51. $\hat{\lambda} = \bar{\xi}$; $\hat{\lambda}$ — несмещенная, состоятельная и эффективная оценка λ .

15.53. $\hat{p} = \frac{1}{mn} \sum_{i=1}^n \xi_i$; $\hat{p}$ — несмещенная, состоятельная и эффективная оценка p .

15.54. $\hat{\theta} = \sqrt{\frac{2}{\pi}} \hat{m}_1$; $\hat{\theta}$ — несмещенная и состоятельная оценка θ .

15.55. Функция максимального правдоподобия

$$L(\theta, \lambda) = \begin{cases} \frac{\theta^n \lambda^{n\theta}}{\left(\prod_{i=1}^n \xi_i \right)^{\theta+1}}, & \text{если все } \xi_i > \lambda; \\ 0, & \text{если хотя бы одно } \xi_i \leq \lambda. \end{cases}$$

Зафиксируем значение переменной θ , т. е. θ — произвольное, но фиксированное. Функция $L(\theta, \lambda)$ одной переменной λ (не дифференцируемая) достигает своего наибольшего значения в точке

$$\hat{\lambda} = \min\{\xi_i\}.$$

Функция $L(\theta, \hat{\lambda})$ дифференцируема и достигает наибольшего значения в точке

$$\hat{\theta} = 1 / \left(\frac{1}{n} \sum_{i=1}^n \ln \xi_i - \ln \min\{\xi_i\} \right).$$

И поскольку $L(\theta, \lambda) \leq L(\theta, \hat{\lambda}) \leq L(\hat{\theta}, \hat{\lambda})$, то $(\hat{\theta}, \hat{\lambda})$ — оценка максимального правдоподобия параметра (θ, λ) .

15.56. Оценка максимального правдоподобия параметра σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2;$$

$\hat{\sigma}^2$ — несмещенная, состоятельная, эффективная оценка параметра σ^2 .

15.57. $\hat{p} = \bar{\xi}$; $\hat{p}$ — несмещенная, состоятельная, эффективная оценка параметра p .

15.58. Частота $\hat{p} = \mu/n$ является эффективной оценкой p .

21.16 К главе 16

16.1. Нулевая гипотеза H_0 : телепат мысли не читает.

16.2. Нулевая гипотеза H_0 : монета симметричная, альтернатива — двусторонняя.

16.3. Нулевая гипотеза H_0 : больной страдает туберкулезом.

16.4. Нулевая гипотеза H_0 : телепат мысли не читает.

16.5. Нулевая гипотеза H_0 : монета симметричная, альтернатива — двусторонняя.

16.6. Нулевая гипотеза H_0 : препарат токсичный.

16.8. Нулевая гипотеза H_0 : болезнь инфекционная, альтернатива — болезнь неинфекционная.

16.9. Нулевая гипотеза H_0 : $\lambda = 1$ — среднее содержание бактерий на единицу объема воды составляет 1. Альтернативная гипотеза H_λ ($\lambda < 1$) — среднее содержание бактерий на единицу объема воды меньше чем 1.

Наблюдается случайная величина ξ — количество загрязненных проб; ξ имеет биномиальное распределение с параметрами $(10; p)$, где p — вероятность того, что в пробе будет по меньшей мере одна бактерия. Если гипотеза H_0 верна, т. е. среднее содержание λ бактерий на единицу объема равно 1, то $p = p_0 = 1 - e^{-1}$. Критическое множество для проверки гипотезы H_0 имеет вид $S = \{k : k \leq l\}$.

16.11. Нулевая гипотеза H_0 : доля дефектных изделий составляет 0,08; альтернативная гипотеза H_p : $p \in (0; 0,08)$ (сложная). Критическое множество $S = \{k : k < l\}$. Требование потребителя:

$$P(\bar{S}|H_0) \geq 0,90,$$

пожелания поставщика

$$P(S|H_{0,01}) \geq 0,96.$$

16.12. Нулевая гипотеза H_0 : игральная кость симметричная, альтернативная гипотеза H_p : $p \in (0,5; 1]$. Критическое множество $S = \{k : k > l\}$.

16.15. Нулевая гипотеза H_0 : весь изюм использован по назначению, альтернативная гипотеза H_λ : часть λ изюма разошлась не по назначению, $\lambda \in (0,1)$.

16.16. Регистрируем количество погибших мышей.

Нулевая гипотеза H_0 : препарат токсичный. Критическое множество будем искать в виде

$$S = \{k : k \leq l\}.$$

По условию

$$P\{S|H_0\} \leq 0,0001$$

или

$$\sum_{k=0}^l C_{10}^k (0,8)^k (0,2)^{10-k} \leq 0,0001.$$

Последнее неравенство удовлетворяют $l \in [0; 2]$, поэтому $S = \{k : k \leq 2\}$. Вероятность того, что при проверке на токсичность с помощью критерия S нетоксичный препарат выдержит контроль

$$P\{S|H_1\} = \sum_{k=0}^2 C_{10}^k (0,02)^k (0,98)^{10-k} \geq 0,996.$$

Результат эксперимента свидетельствует в пользу нетоксичности препарата.

16.18. Нулевая гипотеза H_0 : монета симметрична, альтернативная гипотеза H_p : p равно одному из чисел 0,6; 0,7; 0,8; 0,9 (альтернатива сложная).

16.19. Регистрируем количество k выживших насекомых. Нулевая гипотеза H_0 : процент поражения инсектицида $\theta = 98$, альтернативная гипотеза H_θ : $\theta \in [0; 98)$. Критическое множество $S = \{k : k \geq l\}$. Пожелание авторов нового инсектицида

$$P(\bar{S}|H_{98}) \geq 0,96,$$

Требование покупателя

$$P(S|H_{92}) \geq 0,95.$$

21.17 К главе 17

Далее, проверяя гипотезу $H_0: a = a_0$, через t будем обозначать $(\bar{\xi} - a_0) / \frac{s}{\sqrt{n}}$, а проверяя гипотезу $H_0: a_\xi = a_\eta$, через t будем обозначать $(\bar{\xi} - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right)$.

17.1. Рассмотрим разность $\zeta = \eta - \xi$ между объемом легких после и до сезона (в данном случае пользоваться критерием Стьюдента для проверки гипотезы $H_0: a_\xi = a_\eta$ нельзя, поскольку выборки объема легких после и до сезона не являются независимыми). Относительно

среднего разности ζ выдвигается гипотеза $H_0 : a = 0$ (гипотеза об отсутствии эффекта физических упражнений), в качестве альтернативной к гипотезе H_0 естественно рассматривать гипотезу $a > 0$. Значение $t = \bar{\zeta} / \frac{s}{\sqrt{n}} = 3,36 > 1,833 = t_{0,05;9} = t_{\alpha;n-1}$. Поэтому согласно критерию Стьюдента гипотеза H_0 отклоняется (на 5 %-м уровне значимости), что интерпретируется как существенное увеличение объема легких под влиянием физических упражнений.

17.2. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 0,51 < 2,306 = t_{0,025;8} = t_{\alpha;n+m-2}$. Поэтому согласно критерию Стьюдента гипотеза H_0 не отклоняется (на 5 %-м уровне значимости), что интерпретируем как отсутствие различий отражательных способностей красок.

17.3. Нулевая гипотеза $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$, альтернатива двусторонняя. Значение $s_\xi^2/s_\eta^2 = 0,24$. Гипотеза H_0 на 2 %-м уровне значимости отклоняется. Последнее обозначает, что точность работы контролеров необходимо квалифицировать как различную.

17.4. Нулевая гипотеза $H_0: a = 1$, альтернатива двусторонняя. Значение $|t| = 2,35 > 2,262 = t_{0,025;9} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 (на 5 %-м уровне значимости) отклоняется, что указывает на то, что счетчики отрегулированы неудовлетворительно.

17.5. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 0,891 < 2,262 = t_{0,025;9} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 на 5 %-м уровне значимости не отклоняется (не противоречит экспериментальным данным). Другими словами, эксперимент дает основания утверждать, что сталь I и сталь II не отличаются способностью к глубокому отпуску.

17.6(2°). Нулевая гипотеза $H_0: \sigma^2/0,01 = 1$, альтернатива односторонняя: $\sigma^2/0,01 > 1$ (нас интересует только превышение дисперсией заданного уровня). Значение $s^2/0,01 = 14,53$, $\chi_{\alpha;(n-1)}^2/(n-1) = \chi_{0,05;9}^2/9 = 1,88 < 14,53 = s^2/0,01$. Поэтому гипотеза H_0 на 5 %-м уровне значимости отклоняется. Взвешивание 10 упаковок да-

ет основания утверждать, что требования относительно дисперсии массы упаковок не выполняется — разброс массы упаковки превышает заданную норму.

17.7. Нулевая гипотеза $H_0: a = 12,0$, альтернатива двусторонняя. Значение $|t| = |\bar{\xi} - 12,0| / \frac{s}{\sqrt{n}} = 0,42$; $|t| = 0,42 < 1,77 = t_{0,025;13} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется.

17.8. Пусть a_η — среднее содержание золота, в образцах из шурфов, а a_ξ — из скважин, $H_0: a_\xi = a_\eta$, альтернатива односторонняя: $a_\xi < a_\eta$ (результат исследования образцов на содержание золота из скважин не выше, чем из шурфов). Мы не хотим пренебрегать возможным снижением расходов (за счет бурения скважин вместо закладки шурфов), поэтому назначаем не очень большой уровень значимости, скажем, $\alpha = 0,025$. Значение $t = (\bar{\eta} - \bar{\xi}) / \left(s \sqrt{\frac{n+m}{nm}} \right) = 1,82$; $t = 1,82 < 2,06 = t_{0,025;25} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 не отклоняется. Так что можно считать, что результаты исследования образцов из скважин и из шурфов существенно не отличаются. Следовательно, можно рекомендовать вести разведку залежей рассыпного месторождения золота по образцам, взятым из скважин, что уменьшает затраты.

17.9. Пусть $\sigma_\xi^2, \sigma_\eta^2$ — дисперсии размера внешнего диаметра изделия соответственно после настройки станка и через определенное время после этого. Нулевая гипотеза $H_0: \sigma_\xi^2 / \sigma_\eta^2 = 1$, альтернатива односторонняя: $\sigma_\xi^2 / \sigma_\eta^2 < 1$, поскольку со временем точность работы станка может только уменьшиться (дисперсия критического размера изделия может только увеличиться). Значение отношения $s_\xi^2 / s_\eta^2 = 0,77 > 0,44 = 1/F_{0,05;14;19}$. Поэтому гипотеза H_0 не отклоняется. Последнее обозначает, что приведенные данные не дают оснований говорить о снижении точности работы станка.

17.10. Нулевая гипотеза $H_0: a = 24$, альтернатива двусторонняя. Значение $|t| = |\bar{\xi} - 24| / \frac{s}{\sqrt{n}} = 1,363 < 2,365 = t_{0,025;7} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 не от-

клоняется (на 5 %-м уровне значимости). Следовательно, можно считать, что расчетная скорость полимеризации составляет 24 % за час.

17.11. Пусть a_ξ и a_η — средние предельного сопротивления на сжатие соответственно бетона, приготовленного без обработки, и бетона, приготовленного с обработкой. Относительно a_ξ и a_η выдвигается гипотеза H_0 : $a_\xi = a_\eta$ — гипотеза об отсутствии эффекта специального способа приготовления бетона (обидная для авторов нового способа). Альтернатива односторонняя: $a_\xi < a_\eta$ (специальный способ приготовления бетона ориентирован на повышение его прочности на сжатие). Вы — член доброжелательной комиссии по проверке эффекта специального способа приготовления бетона (не хотите несправедливо пренебрегать эффектом специального способа приготовления бетона, если он существует), поэтому назначаете не очень малый уровень значимости. Выберем, скажем, 5 %-й уровень. Значение

$$t = (\bar{\eta} - \bar{\xi}) / \left(s \sqrt{\frac{n+m}{nm}} \right) = 2,29 >$$

$$> 2,13 = t_{0,05;4} = t_{\alpha;n+m-2}.$$

Поэтому нулевая гипотеза H_0 отклоняется. Последнее обозначает, что эксперимент дает основания утверждать, что специальный способ приготовления бетона повышает его сопротивление на сжатие.

17.12. Пусть σ_ξ^2 , σ_η^2 — дисперсии выхода нежелательного продукта при использовании соответственно катализаторов A и B . Гипотеза H_0 : $\sigma_\xi^2/\sigma_\eta^2 = 1$, альтернатива двусторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$ или $\sigma_\xi^2/\sigma_\eta^2 > 1$, поскольку априори ничего не известно о стабильности выхода нежелательного продукта как в случае использования одного катализатора, так и другого. Значение $s_\xi^2/s_\eta^2 = 0,335$; $1/F_{\alpha;(m-1);(n-1)} = 1/F_{0,01;8;8} = 0,166 < s_\xi^2/s_\eta^2 < 6,03 = F_{0,01;8;8} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 не отклоняется. Эксперимент не дает оснований считать, что разброс выхода нежелательного продукта при использовании катализаторов A и B различен.

17.13. Нулевая гипотеза $H_0: a = 2,5$, альтернатива двусторонняя. Значение $|t| = |\bar{\xi} - 2,5| / \frac{s}{\sqrt{n}} = 2,08 < 2,365 = t_{0,025;7} = t_{\alpha;n-1}$, поэтому гипотеза H_0 не отклоняется (на 5 %-м уровне значимости). Следовательно, можно считать, что выборка получена из нормального распределения со средним 2,5.

17.14. Пусть a_ξ и a_η — среднее увеличение массы детей, которым выдавали соответственно сок и молоко. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя: $a_\xi < a_\eta$ или $a_\xi > a_\eta$. Значение $|t| = 1,30 < 2,01 = t_{0,025;18} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 не отклоняется, что мы интерпретируем как несущественное отличие в увеличении средней массы детей в группах.

17.15. Нулевая гипотеза $H_0: \sigma_\xi^2 / \sigma_\eta^2 = 1$, альтернатива двусторонняя: $\sigma_\xi^2 / \sigma_\eta^2 > 1$ или $\sigma_\xi^2 / \sigma_\eta^2 < 1$ (если H_0 неверна, то априори неизвестно, точность какого микроскопа выше, а какого ниже). Значение $s_\xi^2 / s_\eta^2 = 0,47; 1/F_{\alpha;(m-1);(n-1)} = 1/F_{0,01;11;8} = 0,17 < s_\xi^2 / s_\eta^2 < 4,47 = F_{0,01;8;11} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 на 2 %-м уровне значимости не отклоняется. Эксперимент не дает оснований ставить под сомнение одинаковую точность измерений микроскопами I и II.

17.16. Нулевая гипотеза $H_0: a = 2000$, альтернатива двусторонняя (сопротивление резисторов может быть как меньшим, так и большим 2000). Значение $|t| = 1,02 < 2,20 = t_{0,025;11} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется (на 5 %-м уровне значимости). Последнее обозначает, что такие отклонения значения сопротивления резисторов от номинала допустимы (они естественны и избежать их невозможно).

17.17. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 8,566 > 2,10 = t_{0,025;18} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 отклоняется (на 5 %-м уровне значимости).

17.18. Пусть $\sigma_\xi^2, \sigma_\eta^2$ — дисперсии отражательных способностей краски, изготовленной соответственно по технологиям A и B . Нулевая гипотеза $H_0: \sigma_\xi^2 / \sigma_\eta^2 = 1$, аль-

тернатива односторонняя: $\sigma_{\xi}^2/\sigma_{\eta}^2 > 1$ (изменения в технологии направлены на уменьшение дисперсии отражательных свойств краски). Отклонение в пользу этой альтернативы будет свидетельствовать об уменьшении дисперсии. Значение $s_{\xi}^2/s_{\eta}^2 = 4,74 < 6,39 = F_{0,05;4;4} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 на 5%-м уровне значимости не отклоняется. Эксперимент не дает оснований утверждать, что изменения в технологии, направленные на уменьшение дисперсии отражательных свойств краски, оказались эффективными.

17.19. Нулевая гипотеза $H_0: a = 220$, альтернатива двусторонняя: $a > 220$ или $a < 220$. Значение $|t| = |\bar{\xi} - 220|/\frac{s}{\sqrt{n}} = 1,48 < 2,07 = t_{0,025;23} = t_{\alpha;n-1}$.

Поэтому гипотеза H_0 не отклоняется (на 5%-м уровне значимости). Отклонения от стандарта можно считать допустимыми.

17.20. Пусть a_{ξ} и a_{η} — средние значения результатов прохождения дистанции в 15 км соответственно традиционным ходом и коньковым ходом. Нулевая гипотеза $H_0: a_{\xi} - 1 = a_{\eta}$, альтернатива односторонняя: $a_{\xi} - 1 > a_{\eta}$. Отклонение нулевой гипотезы в пользу альтернативной: $a_{\xi} - 1 > a_{\eta}$ будет свидетельствовать о наличии более чем одноминутного эффекта техники конькового хода, неотклонение гипотезы H_0 будет свидетельствовать об отсутствии такого эффекта. Значение

$$t = ((\bar{\xi} - 1) - \bar{\eta}) / \left(s \sqrt{\frac{n+m}{nm}} \right)$$

сравниваем с $t_{\alpha;n+m-2}$, если $t > t_{\alpha;n+m-2}$ гипотеза H_0 отклоняется, в противном случае — нет. Имеем $t = 3,774 > 2,10 = t_{0,05;18} = t_{\alpha;n+m-2}$. Поэтому согласно критерию Стьюдента гипотеза $H_0: a_{\xi} - 1 = a_{\eta}$ на 5%-м уровне значимости отклоняется, что интерпретируется как более чем одноминутный выигрыш во времени при использовании техники коньковых ходов по сравнению с традиционной техникой.

17.21. Нулевая гипотеза $H_0: a_{\xi} = a_{\eta}$, альтернатива двусторонняя: $a_{\xi} < a_{\eta}$ или $a_{\xi} > a_{\eta}$, поскольку нет апри-

орной информации относительно диаметров изделий, изготовленных на станках A и B . Значение $|t| = 1,98$; $|t| = 1,98 < 2,07 = t_{0,025;23} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется. Эксперимент не дает оснований утверждать, что диаметры изделий, изготовленных на разных станках, отличны.

17.22. Нулевая гипотеза $H_0: a = 18$, альтернатива двусторонняя. Значение $|t| = 0,89 < 2,20 = t_{0,025;11} = t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется (на 5 %-м уровне значимости). Последнее трактуется как выполнение требований стандарта относительно содержания хрома в стали 18Cr10Ni2Mo.

17.23. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 2,55 > 2,10 = t_{0,025;18} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 отклоняется (на 5 %-м уровне значимости). Эксперимент дает основания утверждать, что комфортная температура для женщин и мужчин различна.

17.24. Нулевая гипотеза $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$, альтернативная гипотеза двусторонняя. Значение s_ξ^2/s_η^2 равно 1,71; $1/F_{\alpha;(m-1);(n-1)} = 1/F_{0,01;7;7} = 0,14 < s_\xi^2/s_\eta^2 < 6,99 = F_{0,01;7;7} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 не отклоняется (на 2 %-м уровне значимости). Эксперимент свидетельствует в пользу отсутствия различий между разбросом сопротивлений провода типа A и B .

17.25. Нулевая гипотеза $H_0: a = 11$, альтернатива односторонняя: $a < 11$ (прочность на изгиб меньше 11). Значение $t = (\bar{\xi} - 11) / \frac{s}{\sqrt{n}} = 3,8 > -1,75 = -t_{0,05;15} = -t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется (на 5 %-м уровне значимости). Таким образом, можно считать, что прочность на изгиб не меньше чем 11.

17.26. Точность станка характеризуется величиной дисперсии: меньше дисперсия — больше точность, больше дисперсия — меньше точность.

Проверяем гипотезу $H_0: \sigma_A^2/\sigma_B^2 = 1$, альтернатива двусторонняя: $\sigma_A^2/\sigma_B^2 > 1$ или $\sigma_A^2/\sigma_B^2 < 1$, поскольку неизвестно, точность какого станка больше, а какого меньше. Значение $s_A^2/s_B^2 = 1,27$; $1/F_{\alpha;(m-1);(n-1)} = 1/F_{0,01;9;14} =$

$= 0,25 < 1,27 = s_A^2/s_B^2 < 5,00 = F_{0,01;14;9} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 не отклоняется (на 2%-м уровне значимости). Эксперимент не дает оснований утверждать, что станки принадлежат разным классам точности.

17.27. Разброс характеризуется дисперсией, больше дисперсия — больше разброс. Пусть σ_ξ^2 — дисперсия содержания марганца в случае его определения по традиционной методике, а σ_η^2 — по новой. Нулевая гипотеза H_0 : $\sigma_\xi^2/\sigma_\eta^2 = 1$, альтернатива односторонняя: $\sigma_\xi^2/\sigma_\eta^2 > 1$ (ожидается, что новая методика определения содержания марганца дает меньший разброс результатов). Отклонение H_0 в пользу этой альтернативы будет свидетельствовать об уменьшении дисперсии. Значение $s_\xi^2/s_\eta^2 = 1,55$; $s_\xi^2/s_\eta^2 = 1,55 < 3,68 = F_{0,05;9;7} = F_{\alpha;(n-1);(m-1)}$. Поэтому гипотеза H_0 не отклоняется. Эксперимент не дает оснований говорить об уменьшении разброса.

17.28. Нулевая гипотеза H_0 : $a = 6720$; альтернатива односторонняя: $a < 6720$ (главное — обнаружить проволоку, которая рвется, когда значение усилия на разрыв меньше предельного). Значение $t = (\bar{\xi} - 6720)/\frac{s}{\sqrt{n}} = -1,38$; $t > -1,729 = -t_{0,05;19} = -t_{\alpha;n-1}$. Поэтому гипотеза H_0 не отклоняется. Эксперимент дает основания утверждать, что требования к прочности партии проволоки, из которой изготавливают канаты, выполнены.

17.29. Нулевая гипотеза H_0 : $a_\xi = a_\eta$, альтернатива двусторонняя (нет априорной информации относительно характеристик футеровки предприятий A и B). Значения $|t| = 4,05 > 2,11 = t_{0,025;17} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 отклоняется (на 5%-м уровне значимости). Эксперимент дает основания утверждать, что футеровка, изготовленная на предприятиях A и B , отличается по своим характеристикам.

17.30. Пусть H_0 : $\sigma^2/35,63 = 1$, альтернатива односторонняя: $\sigma^2/35,63 < 1$. Отклонение H_0 в пользу этой альтернативы будет свидетельствовать об уменьшении дисперсии, в противном случае — нет. Значение $s^2/35,63 = 0,842 > 0,47 = \chi_{0,05;14}^2/14 = \chi_{\alpha;(n-1)}^2/(n-1)$. Поэтому гипотеза H_0 не отклоняется. Эксперимент не дает основа-

ний утверждать, что дисперсия за счет технологических изменений процесса уменьшилась (несмотря на ожидания их авторов).

17.31. Поскольку регистрируется температура правой и левой шин для одного и того же автобуса, то выборки не являются независимыми (а следовательно, нельзя пользоваться критерием для проверки гипотезы о равенстве средних), поэтому рассмотрим разность $\zeta = \eta - \xi$ температур, скажем правой и левой шин. Относительно среднего a разности ζ выдвигается гипотеза $H_0: a = 0$, альтернатива двусторонняя. Значение $|t| = |\bar{\xi}| / \frac{s}{\sqrt{n}} =$

$= 2,16$. Так что $|t| = 2,16 < 2,20 = t_{0,025;11} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 не отклоняется (на 5%-м уровне значимости), что свидетельствует в пользу несущественной разницы температуры, до которой нагреваются правая и левая шины автобуса во время его движения.

17.32. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 3,5 > 2,10 = t_{0,025;18} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 отклоняется. Эксперимент дает основания утверждать, что в течение заданного интервала времени произошли изменения в уровне настройки станка.

17.34. Рассмотрим разность $\zeta = \eta - \xi$ между длиной образцов после отжига и до него. (Пользоваться критерием Стьюдента для проверки гипотезы о равенстве средних нельзя, поскольку выборки не являются независимыми — измеряется длина одного и того же образца до отжига и после него.) Относительно среднего a разности ζ выдвигается гипотеза $H_0: a = 0$, альтернатива двусторонняя: $a > 0$ или $a < 0$. Значение $|t| = 2,14 < 2,262 = t_{0,025;9}$. Поэтому гипотеза H_0 не отклоняется. Эксперимент не дает оснований говорить об изменении длины образцов изделий после их отжига.

17.35. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя. Значение $|t| = 0,916 < 2,015 = t_{0,05;5} = t_{\alpha;n+m-2}$. Поэтому гипотеза H_0 не отклоняется. Предположение о том, что состав резины не влияет на ее прочность, не противоречит эксперименту.

17.36. Точность работы станка характеризуется дисперсией и со временем не возрастает (а дисперсия соот-

ветственно не уменьшается). В терминах проверки статистических гипотез предположение о неизменности точности работы станка формулируется как гипотеза $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$, альтернатива односторонняя: $\sigma_\xi^2/\sigma_\eta^2 < 1$. Значение $s_\xi^2/s_\eta^2 = 0,51 > 0,31 = 1/F_{0,05;9;9} = 1/F_{\alpha;(m-1);(n-1)}$. Поэтому нулевая гипотеза не отклоняется. Таким образом, можно считать, что точность работы станка не уменьшилась (хотя значение оценки дисперсии и возросло с $s_\xi^2 = 4,4 \text{ мкм}^2$ до $s_\eta^2 = 8,6 \text{ мкм}^2$, но это допустимое увеличение дисперсии).

17.37. Рассмотрим разность $\zeta = \eta - \xi$ между длительностью обезболивающего действия препаратов B и A . (В данном случае пользоваться критерием Стьюдента для проверки гипотезы $H_0: a_\xi = a_\eta$ нельзя, поскольку выборки не являются независимыми.) Относительно среднего a разности ζ выдвигается гипотеза $H_0: a = 0$, при этом в случае 1° (при отсутствии априорной информации об эффективности препаратов) альтернатива двусторонняя: $a < 0$ или $a > 0$; в случае 2°, когда известно, что B имеет не меньшую фармакологическую активность чем A , альтернатива односторонняя: $a > 0$. Далее, $t = \bar{\zeta} / \frac{s}{\sqrt{n}} = 2,43$; $|t| = 2,43 > 2,365 = t_{0,025;7} = t_{\alpha;n-1}$. Поэтому в случае 1° гипотеза H_0 отклоняется (на 5%-м уровне значимости), что мы трактуем как различную фармакологическую активность препаратов B и A .

Отклоняется гипотеза H_0 и при наличии априорной информации (фармакологическая активность B не меньше чем A), поскольку $t = 2,43 > 1,895 = t_{0,05;7}$. Это свидетельствует в пользу большей фармакологической активности препарата B по сравнению с A .

17.38. Нулевая гипотеза $H_0: a_\xi = a_\eta$, альтернатива двусторонняя: $a_\xi < a_\eta$ или $a_\xi > a_\eta$. Значение $|t| = 1,14 < 2,179 = t_{0,025;12}$. Поэтому гипотеза не отклоняется. Эксперимент не дает оснований утверждать, что результаты операторов существенно различаются.

21.18 К главе 18

Примеры интерпретации неотклонения нулевой гипотезы, отклонения нулевой гипотезы в пользу той или другой альтернативы можно найти в решениях и указаниях к задачам гл. 17.

18.7. Чтобы обеспечить выполнение условия $\nu_i \nu_j / n \geq 10$ для всех i, j , не исключено, что признаки D и E необходимо будет объединить в один признак S — недостаточно развитый или умственно отсталый.

18.8. Можно предложить две модели: в первой — монеты симметричны, в другой — информация о симметричности отсутствует. В последней модели гипотетическое распределение зависит от неизвестного параметра, который необходимо оценить по выборке методом максимального правдоподобия.

18.9. См. указание к задаче 18.8.

18.19. Формально гипотеза о независимости глухонемоты от пола формулируется как гипотеза о независимости признаков: пол и глухонемота.

Сначала необходимо построить таблицу сопряженности признаков. Наблюдается $15\,729\,000 + 16\,799\,000$ значений (по количеству жителей Англии и Уэльса) случайной величины (ξ, η) , где ξ — “индикатор” пола (принимает значение 0 для женского пола и 1 — для мужского), η — “индикатор” глухонемоты (принимает значение 1, если человек глухонемой и 0 — в противном случае).

18.22. Чтобы обеспечить выполнение условия

$$\nu_i \nu_j / n \geq 10$$

для всех i, j , не исключено, что два признака: F и G (или два признака: “сносное” и “очень плохое”) необходимо будет объединить в один признак.

18.25. Гипотеза H_0 : число появлений $(1, 1)$ на 100 пар имеет пуассоновское распределение.

18.27. Не исключено, что некоторые признаки необходимо будет объединить в один (см., например, указание к задаче 18.7).

18.30. См. указание к задаче 18.8.

18.37. Если телепат мысли не читает, то вероятность правильно прочесть задуманную цифру равна $1/10$.

Пусть ξ — случайная величина — число правильно прочитанных цифр в одном эксперименте, ξ принимает значения 1 и 0, распределение ξ имеет вид

$$P_{\xi}(k) = \theta^k(1 - \theta)^{1-k}; \quad k = 0, 1; \quad \theta \in (0; 1).$$

Относительно распределения P_{ξ} выдвигаем гипотезу

$$H_0 : P_{\xi}(k) = (1/10)^k(9/10)^{1-k}; \quad k = 0, 1.$$

Альтернативная гипотеза

$$H_{\theta} : P_{\xi}(k) = \theta^k(1 - \theta)^{1-k}; \quad k = 0, 1; \quad \theta > 1/10.$$

Отклонение нулевой гипотезы в пользу альтернативной свидетельствует о наличии способностей телепата читать мысли, не отклонение — об отсутствии таковых.

21.19 К главе 19

Примеры интерпретации неотклонения нулевой гипотезы, ее отклонения в пользу той или другой альтернативы можно найти в решениях и указаниях к задачам гл. 17.

19.7. Номинальный диаметр шейки рабочей части сверла составляет 9,8 мм, но в процессе производства получается то, что получается (см. выборку), по-другому в принципе не может быть — уж очень много случайных (неконтролируемых) факторов влияет на конечный результат. Поэтому измерения диаметров шеек 20 сверл, каждый из которых должен быть равен 9,8 мм, дают не одно число 9,8, а реализацию выборки некоторой случайной величины ζ , которую мы и называем диаметром шейки сверла.

Относительно распределения значений диаметра шейки сверла выдвигается вполне естественное предположение: ζ — нормально распределенная случайная величина

(теоретическим обоснованием этого предположения, которое в большинстве ситуаций оправдывается, является центральная предельная теорема). Нормальное распределение $N_{a;\sigma^2}$ определяется двумя параметрами: a и σ^2 ; его функция распределения

$$N_{a;\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x \exp\left\{-\frac{(t-a)^2}{2\sigma^2}\right\} dt.$$

Поскольку номинальный диаметр шейки сверла 9,8 мм, то $a = 9,8$. Требование “норма отхода при техническом допуске 0,05 мм должна составлять 1 %” обозначает, что

$$P\{9,75 \leq \zeta \leq 9,85\} = 0,99.$$

Последнее равенство, с учетом предположения о нормальности распределения ζ , можно переписать так:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{9,75}^{9,85} \exp\left\{-\frac{(t-9,8)^2}{2\sigma^2}\right\} dt = 0,99,$$

а после замены $(t-9,8)/\sigma = u$ так:

$$\frac{1}{\sqrt{2\pi}} \int_{-0,05/\sigma}^{0,05/\sigma} \exp\left\{-\frac{u^2}{2}\right\} du = 0,99,$$

или

$$\frac{1}{\sqrt{2\pi}} \int_0^{0,05/\sigma} \exp\left\{-\frac{u^2}{2}\right\} du = 0,495.$$

По таблице 22.1.1 нормального распределения находим $0,05/\sigma = 2,575$, $\sigma = 0,02$. Таким образом, функция гипотетического распределения диаметра шейки сверла имеет вид

$$N_{a;\sigma^2}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a)/\sigma} \exp\left\{-\frac{u^2}{2}\right\} du,$$

где $a = 9,8$, $\sigma^2 = 0,0004$.

Осталось, воспользовавшись критерием А. Н. Колмогорова, проверить гипотезу H_0 : диаметр шейки сверла имеет распределение $N_{a;\sigma^2}$; $a = 9,8$, $\sigma^2 = 0,0004$.

19.16. Показания часов записать в часах.

19.25. Случайную величину s_i можно представить в виде

$$s_i = \sum_{k=0}^{16} \varepsilon_k^{(i)},$$

где $\varepsilon_k^{(i)}$, $k = 0, 1, \dots, 16$, — независимые одинаково распределенные случайные величины, распределением каждой из которых является

$$P\{\varepsilon_k^{(i)} = j\} = \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{1-j}, \quad j = 0, 1.$$

Согласно центральной предельной теореме, нормированная сумма независимых одинаково распределенных случайных величин (с конечными дисперсиями) имеет распределение, близкое к $N_{0;1}$ -распределению. Заметим, что

$$Ms_i = M \sum_{k=0}^{16} \varepsilon_k^{(i)} = 8, \quad Ds_i = D \sum_{k=0}^{16} \varepsilon_k^{(i)} = 4,$$

поэтому

$$\frac{s_i - Ms_i}{\sqrt{Ds_i}} = \frac{s_i - 8}{2}.$$

19.28. См. задачу 19.25.

21.20 К главе 20

20.1. Оценки $\hat{\alpha} = 13,11$; $\hat{\beta} = 0,74$; $\hat{\sigma}^2 = 20,57$. Доверительными интервалами с коэффициентом надежности 0,90 являются: (12, 01; 14, 21) для параметра α ; (0, 61; 0, 87) для параметра β ; (15, 78; 31, 08) для параметра σ^2 .

Уравнение регрессии

$$s = 13,11 + 0,74(v - 24,78).$$

Проверка значимости регрессии: $|t| = 9,661 > 1,677 = t_{0,05;48}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости). Регрессия значима.

Для проверки адекватности линейной регрессии вычислим

$$s_1^2 = \frac{1}{n-k} \sum_{j=1}^k \sum_{\nu=1}^{n_j} (s_{j;\nu} - \bar{s}_j)^2,$$

$$s_2^2 = \frac{1}{k-2} \sum_{j=1}^k n_j (\bar{s}_j - (\hat{\alpha} + \hat{\beta}(v_j - \bar{v}))^2,$$

$$\bar{s}_j = \frac{1}{n_j} \sum_{\nu=1}^{n_j} s_{j;\nu}, j = 1, 2, \dots, k, n = \sum_{j=1}^k n_j, \bar{v} = \frac{1}{n} \sum_{i=1}^n v_i.$$

Имеем $n_1 = 2, n_2 = 2, n_3 = 1, \dots, n_{19} = 1$; $n = 50$, $k = 19$; $s_1^2 = 20,31$; $s_2^2 = 21,23$; $s_2^2/s_1^2 = 1,05 < 1,96 = F_{0,05;17;31}$ поэтому гипотеза об адекватности описания линейной моделью зависимости тормозного пути от его скорости не отклоняется.

20.2. Оценки $\hat{\alpha} = 73,98$; $\hat{\beta} = 3,85$; $\hat{\sigma}^2 = 59,44$. Доверительными интервалами с коэффициентом надежности 0,90 являются: (71, 15; 76, 80) для параметра α ; (2, 06; 5, 64) для параметра β ; (42, 06; 115, 6) для параметра σ^2 .

Уравнение эмпирической линии регрессии

$$y = 73,98 + 3,85(x - 10,53).$$

Проверка значимости регрессии: значение $|t| = 3,689 > 1,717 = t_{0,05;22}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости) — регрессия значима.

Проверка адекватности регрессии: имеем $n_1 = 1, n_2 = 1, n_3 = 2, \dots, n_{19} = 1; n = 24, k = 19; s_1^2 = 129,84; s_2^2 = 5,12$; значение $s_2^2/s_1^2 = 0,04 < 4,59 = F_{0,05;17;5}$ поэтому гипотеза об адекватности не отклоняется.

20.3. Оценки $\hat{\alpha} = 95,42; \hat{\beta} = 1,87; \hat{\sigma}^2 = 97,36$. Доверительными интервалами с коэффициентом надежности 0,90 являются: (90,08; 100,76) для параметра α ; (0,92; 2,82) для параметра β ; (64,31; 276,35) для параметра σ^2 .

Уравнение регрессии

$$y = 95,42 + 1,87(x - 7,46).$$

Проверка значимости регрессии: значение $|t| = 3,55 > 1,796 = t_{0,05;11}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости). Регрессия значима.

Проверка адекватности регрессии. Имеем $n_1 = 3, n_2 = 1, n_3 = 1, \dots, n_7 = 1; n = 13, k = 7; s_1^2 = 101,17; s_2^2 = 48,83; s_2^2/s_1^2 = 0,48 < 4,39 = F_{0,05;5;6}$ поэтому гипотеза об адекватности не отклоняется

20.4. Оценки $\hat{\alpha} = 108,82; \hat{\beta} = -24,07; \hat{\sigma}^2 = 8,37$. Доверительными интервалами с коэффициентами надежности 0,90 являются: (107,25; 110,38) для параметра α ; (-26,44; -21,70) для параметра β ; (5,53; 23,75) для параметра σ^2 .

Уравнение регрессии

$$y = 108,82 - 24,07(x - 0,87).$$

Проверка значимости регрессии: значение $|t| = 18,23 > 1,796 = t_{0,05;11}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости) — регрессия значима.

Проверка адекватности регрессии. Имеем $n_1 = 3, n_2 = 2, n_3 = 2, \dots, n_8 = 2; n = 13, k = 8; s_1^2 = 2,28; s_2^2 = 10,10; s_2^2/s_1^2 = 4,44 < 4,95 = F_{0,05;6;5}$ поэтому гипотеза об адекватности не отклоняется

20.6. Оценки $\hat{\alpha} = -1,10; \hat{\beta} = -0,59; \hat{\sigma}^2 = 0,75$. Доверительными интервалами с коэффициентом надежности

0,90 являются: $(-1, 31; -0, 90)$ для параметра α ; для параметра β — $(-0, 73; -0, 45)$; $(0, 58; 1, 12)$ для параметра σ^2 .

Уравнение регрессии

$$T = -1, 10 - 0, 59(Ca - 4, 09).$$

Проверка значимости регрессии: значение $|t| = 7, 19 > 1, 676 = t_{0,05;50}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости) — регрессия значима.

Проверка адекватности регрессии: имеем $n_1 = 2, n_2 = 1, n_3 = 1, \dots, n_{34} = 3; n = 52, k = 34; s_1^2 = 0, 87; s_2^2 = 0, 45; s_2^2/s_1^2 = 0, 51 < 2, 19 = F_{0,05;32;18}$ поэтому гипотеза об адекватности описания линейной моделью зависимости значения T денситометрического анализа от количества выводимого Ca не отклоняется.

В соответствии с уравнением регрессии значению $T = -2$ (границе между средней и тяжелой формой остеопении) соответствует значение $5, 7$ ммоль/л Ca . Поэтому, если $Ca > 5, 7$ (Ca выводится больше чем $5, 7$ единиц) степень остеопении естественно классифицировать как тяжелую.

20.7. Оценки $\hat{\alpha} = 13, 425; \hat{\beta} = 0, 078; \hat{\sigma}^2 = 1, 648$. Доверительными интервалами с коэффициентом надежности 0,90 являются: $(12, 90; 13, 95)$ для параметра α ; для параметра β — $(0, 061; 0, 096)$; $(1, 141; 3, 509)$ для параметра σ^2 .

Уравнение регрессии

$$\hat{y} = 13, 425 + 0, 078(x - 67).$$

Проверка значимости регрессии: значение $|t| = 7, 737 > 1, 734 = t_{0,05;18}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости). Регрессия значима.

20.9. Оценки $\hat{\alpha} = 804; \hat{\beta} = 131, 72; \hat{\sigma}^2 = 70906, 29$. Доверительными интервалами с коэффициентом надежности 0,90 являются: $(683, 47; 924, 52)$ для параметра α ; $(69, 29; 194, 14)$ для параметра β ; $(48235, 57; 166034, 03)$ для параметра σ^2 .

Уравнение регрессии

$$y = 804 + 131, 72(x - 3, 65).$$

Проверка значимости регрессии: значение $|t| = 3,69 > 1,753 = t_{0,05;15}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется на уровне значимости 0,05. Регрессия значима.

Проверка адекватности регрессии: имеем $n_1 = 3, n_2 = 3, n_3 = 3, \dots, n_7 = 3; n = 173, k = 7; s_1^2 = 68475; s_2^2 = 103385; s_2^2/s_1^2 = 1,51 < 3,33 = F_{0,05;5;10}$ поэтому гипотеза об адекватности не отклоняется.

20.10. Оценки $\hat{\alpha} = 2125,28; \hat{\beta} = 0,97; \hat{\sigma}^2 = 2826,22$. Доверительными интервалами с коэффициентом надежности 0,90 являются: (2102,07; 2148,48) для параметра α ; (0,93; 1,02) для параметра β ; (1934,29; 6390,95) для параметра σ^2 .

Уравнение регрессии

$$y = 2125,28 + 0,97(x - 2165).$$

Согласно (20.1.6) значение $|t| = 38,74 > 1,746 = t_{0,05;16}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости). Регрессия значима.

20.11. Оценки $\hat{\alpha} = 0,63; \hat{\beta} = 0,46; \hat{\sigma}^2 = 0,02$. Доверительными интервалами с коэффициентом надежности 0,90 являются: (0,606; 0,653) для параметра α ; для параметра β — (0,377; 0,543); (0,017; 0,026) для параметра σ^2 .

Уравнение регрессии

$$y = 0,63 + 0,46(x - 0,81).$$

Проверка значимости регрессии: значение $|t| = 9,184 > 1,660 = t_{0,05;101}$, поэтому гипотеза $H_0 : \beta = 0$ отклоняется (на 5% уровне значимости). Регрессия значима.

Проверка адекватности регрессии: имеем $n_1 = 1, n_2 = 5, n_3 = 7, \dots, n_{25} = 1; n = 103, k = 25; s_1^2 = 0,017; s_2^2 = 0,022; s_2^2/s_1^2 = 1,28 < 1,67 = F_{0,05;23;78}$ поэтому гипотеза об адекватности не отклоняется.

21.21 Задания для самостоятельной работы

Задание 1

**Стохастический эксперимент.
Дискретное вероятностное пространство**

Вариант	Задача							
1	1.10	2.26	2.11	3.14	3.36	4.2	4.17	
2	1.14	2.14	2.12	3.13	3.37	4.3	4.18	
3	1.15	2.13	2.15	3.12	3.36	4.4	4.19	
4	1.19	2.25	2.24	3.11	3.40	4.5	4.30	
5	1.20	2.10	2.23	3.9	3.32	4.6	4.21	
6	1.22	2.9	2.22	3.8	3.28	4.7	4.22	
7	1.24	2.8	2.21	3.7	3.27	4.8	4.23	
8	1.25	2.7	2.20	3.6	3.25	4.9	4.24	
9	1.23	2.5	2.19	3.5	3.21	4.10	4.28	
10	1.33	2.4	2.18	3.3	3.20	4.12	4.19	
11	1.21	2.3	2.17	3.2	3.16	4.13	4.23	
12	1.26	2.2	2.16	3.1	3.38	4.29	4.12	

Задание 2

**Дискретная случайная величина,
ее распределение и числовые характеристики**

Вариант	Задача							
1	5.41	5.8(1a)	5.10(3a)	5.28	6.34	6.32		
2	5.42	5.8(1б)	5.11	5.29	6.35	6.31		
3	5.44	5.8(1в)	5.12	5.30	6.37	6.28(1)		
4	5.6	5.8(2a)	5.13	5.31	6.39	6.27		
5	5.7	5.8(2б)	5.14	5.32	6.5	6.21		
6	5.19	5.8(2в)	5.15	5.34(a)	6.40	6.20		
7	5.20	5.8(2г)	5.16	5.34(б)	6.41	6.18		
8	5.21	5.8(2д)	5.18	5.38	6.45	6.17		
9	5.22	5.8(2е)	5.19	5.39	6.46	6.16		
10	5.23	5.8(3a)	5.8(2ё)	5.40(1)	6.49	6.30		
11	5.24	5.8(3б)	5.8(2ж)	5.40(2)	6.50	6.28(2)		
12	5.27	5.8(3в)	5.8(2з)	5.39	6.51	6.28(3)		

Задание 3
Аксиоматика теории вероятностей

Вариант	Задача				
1	7.1	7.7	7.15(1)	7.18	
2	7.2	7.8	7.17	7.19	
3	7.3	7.9	7.15(3)	7.20	
4	7.4	7.10	7.15(4)	7.29	
5	7.5	7.7	7.15(5)	7.30	
6	7.6	7.8	7.15(1)	7.20	
7	7.1	7.9	7.15(2)	7.14	
8	7.2	7.10	7.15(3)	7.21	
9	7.3	7.7	7.15(4)	7.22	
10	7.4	7.8	7.15(2)	7.30	
11	7.5	7.9	7.16	7.16	
12	7.6	7.10	7.17	7.14	

Задание 4
Геометрические вероятности

Вариант	Задача				
1	8.3(1)	8.4(1)	8.6	8.18	8.33
2	8.3(2)	8.4(2)	8.7	8.17	8.15(3)
3	8.3(3)	8.4(3)	8.32	8.19	8.13(2)
4	8.3(4)	8.4(6)	8.8	8.20	8.18
5	8.3(5)	8.4(7)	8.9	8.23	8.19
6	8.3(6)	8.4(8)	8.10	8.25	8.20
7	8.3(7)	8.4(9)	8.33	8.26	8.23
8	8.3(8)	8.4(11)	8.12	8.28	8.25
9	8.2(1)	8.4(13)	8.13(1)	8.24	8.28
10	8.2(2)	8.4(14)	8.14	8.30	8.16
11	8.2(3)	8.4(15)	8.15(1)	8.31	8.6
12	8.2(4)	8.4(16)	8.15(2)	8.34	8.7

Задание 5

Случайная величина и ее распределение

Вариант	Задача				
1	9.47	9.12	9.23	9.33	
2	9.48	9.11	9.24(1)	9.34	
3	9.49	9.13	9.24(2)	9.35(1)	
4	9.50	9.14	9.25(1)	9.35(2)	
5	9.51	9.15	9.25(2)	9.36	
6	9.52	9.16	9.26(1)	9.37	
7	9.53	9.17	9.26(2)	9.38	
8	9.6	9.18	9.27	9.39	
9	9.7	9.19	9.28	9.40	
10	9.8	9.20	9.29	9.41	
11	9.9	9.21	9.30	9.42	
12	9.10	9.22	9.32	9.43	

Задание 6

Математическое ожидание
случайной величины

Вариант	Задача				
1	10.1(1)	10.10	10.19(1a)	10.21	10.31
2	10.1(2)	10.11	10.19(1б)	10.22(1)	10.32
3	10.38	10.12	10.19(2a)	10.23(1)	10.33
4	10.39	10.13	10.19(2б)	10.23(2)	10.34
5	10.40	10.14	10.19(3a)	10.24(1)	10.35
6	10.4	10.15	10.19(3б)	10.24(2)	10.21
7	10.5	10.16(1)	10.19(4a)	10.24(3)	10.23(1)
8	10.41	10.16(2)	10.19(4б)	10.25	10.23(2)
9	10.6(3,4)	10.16(3)	10.20(1a)	10.26	10.31
10	10.7	10.17(1)	10.20(2б)	10.27	10.32
11	10.8	10.17(2)	10.20(3a)	10.28	10.22(1)
12	10.9	10.17(3)	10.20(4б)	10.29	10.22(2)

Задание 7
Свертка

Вари- ант	Задача				Вари- ант	Задача			
1	11.1	11.10(1)	11.16		7	11.4(1)	11.11(3)	11.18	
2	11.2(1)	11.10(2)	11.17		8	11.4(2)	11.12	11.5	
3	11.2(2)	11.8	11.16		9	11.4(3)	11.13(1)	11.18	
4	11.2(3)	11.6	11.19		10	11.4(4)	11.13(2)	11.19	
5	11.2(4)	11.11(1)	11.20		11	11.4(5)	11.14	11.22	
6	11.3	11.11(2)	11.9		12	11.4(6)	11.15	11.7	

Задание 8
**Сходимость распределений.
Характеристическая функция**

Вариант	Задача					
1	12.1	12.9(1)	13.1	13.16	13.23	
2	12.2	12.9(2)	13.2	13.15	13.24	
3	12.3	12.10	13.3	13.13	13.25	
4	12.4	12.11	13.4	13.17	13.26	
5	12.5	12.12	13.5	13.18	13.7	
6	12.6	12.13	13.6	13.19	13.9	
7	12.7(1)	12.14	13.7	13.20	13.26	
8	12.7(2)	12.16	13.8	13.27	13.25	
9	12.7(3)	12.17	13.9	13.28	13.24	
10	12.7(4)	12.18	13.10	13.29	13.23	
11	12.7(5)	12.19	13.11	13.30	13.19	
12	12.15	12.20	13.12	13.22	13.37	

Задание 9
Оценивание параметров распределений.

Вари- ант	Задача					Вари- ант	Задача				
1	14.1	14.2	14.3	14.40		7	14.19	14.20	14.21	14.46	
2	14.4	14.4	14.6	14.41		8	14.22	14.23	14.24	14.47	
3	14.7	14.8	14.9	14.42		9	14.25	14.26	14.27	14.48	
4	14.10	14.11	14.12	14.43		10	14.28	14.29	14.30	14.42	
5	14.13	14.14	14.15	14.44		11	14.31	14.32	14.33	14.40	
6	14.16	14.17	14.18	14.45		12	14.34	14.35	14.36	14.41	

Задание 10
Методы построения оценок

Вариант	Задача				Вариант	Задача			
1	15.1	15.2	15.3	15.4	7	15.25	15.26	15.27	15.28
2	15.5	15.6	15.7	15.8	8	15.29	15.30	15.31	15.32
3	15.9	15.10	15.11	15.12	9	15.33	15.34	15.35	15.56
4	15.13	15.14	15.15	15.16	10	15.37	15.38	15.39	15.40
5	15.17	15.18	15.19	15.20	11	15.41	15.42	15.43	15.44
6	15.21	15.22	15.23	15.24	12	15.45	15.46	15.47	15.48

Задание 11
**Задача проверки статистических гипотез.
 Критерий, функция мощности критерия**

Вариант	Задача		Вариант	Задача	
1	16.1	16.2	7	16.13	16.14
2	16.3	16.4	8	16.15	16.16
3	16.5	16.6	9	16.17	16.18
4	16.7	16.8	10	16.19	16.20
5	16.9	16.10	11	16.21	16.10
6	16.11	16.12	12	16.5	16.10

Задание 12
**Проверка гипотез о параметрах
 нормального распределения**

Вариант	Задача			Вариант	Задача		
1	17.1	17.2	17.3	7	17.19	17.20	17.26
2	17.4	17.5	17.6	8	17.22	17.23	17.24
3	17.7	17.8	17.9	9	17.25	17.21	17.27
4	17.10	17.11	17.12	10	17.28	17.29	17.30
5	17.13	17.14	17.15	11	17.31	17.32	17.33
6	17.16	17.17	17.18	12	17.34	17.35	17.36

Задание 13
Критерий χ^2

Вари- ант	Задача				Вари- ант	Задача			
1	18.1	18.2	18.3	18.4	8	18.29	18.30	18.31	18.32
2	18.5	18.6	18.7	18.8	9	18.33	18.34	18.35	18.36
3	18.9	18.10	18.11	18.12	10	18.37	18.38	18.39	18.40
4	18.13	18.14	18.15	18.16	11	18.41	18.42	18.43	18.48
5	18.17	18.18	18.19	18.20	12	18.45	18.46	18.47	18.48
6	18.21	18.22	18.23	18.24	13	18.49	18.50	18.51	18.52
7	18.25	18.26	18.27	18.28	14	18.54	18.53	18.56	18.55

Задание 14
**Непараметрические критерии: критерий
А. Н. Колмогорова, критерий Вилкоксона,
критерий знаков**

Вариант	Задача			Вариант	Задача		
1	19.2	19.1	19.3	7	19.19	19.20	19.21
2	19.4	19.5	19.6	8	19.22	19.23	19.24
3	19.7	19.8	19.9	9	19.25	19.26	19.27
4	19.10	19.11	19.12	10	19.28	19.29	19.30
5	19.13	19.14	19.15	11	19.31	19.32	19.21
6	19.16	19.17	19.18	12	19.34	19.35	19.36

Задание 15
Линейная регрессия

Вариант	Задача		Вариант	Задача	
1	20.1	20.5	7	20.5	20.7
2	20.2	20.6	8	20.2	20.8
3	20.3	20.12	9	20.9	20.11
4	20.4	20.11	10	20.5	20.7
5	20.7	20.8	11	20.10	20.11
6	20.9	20.10	12	20.9	20.12

Глава 22

Таблицы математической статистики

В эту главу включены все необходимые для работы с учебником таблицы. В частности, приведены квантили, верхние пределы (критические значения) основных распределений математической статистики: стандартного нормального, Стьюдента, Фишера, χ^2 -распределения.

Пусть F — абсолютно непрерывное распределение. Для каждого $\beta \in (0; 1)$ число x_β , являющееся решением уравнения

$$F(x_\beta) = \beta,$$

или, что то же, $F((-\infty, x_\beta)) = \beta$, называется β -квантилью распределения F .

Для каждого $\alpha \in (0; 1)$ число z_α , являющееся решением уравнения

$$1 - F(z_\alpha) = \alpha,$$

или, что то же, уравнения $F([z_\alpha, +\infty)) = \alpha$, будем называть верхним α -пределом (верхним 100α -процентным пределом, 100α -процентной точкой, 100α -критическим значением) распределения F .

22.1 Нормальное распределение

В таблице 22.1.1 приведены значения функции $\Phi(t)$ нормального распределения с параметрами $(0;1)$ (квантили нормального распределения): для заданных t табулированы значения функции

$$N_{0;1}(t) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left\{-\frac{s^2}{2}\right\} ds.$$

Для каждого t значение $N_{0;1}(t)$ численно равно площади заштрихованной на рисунке 22.1.1 фигуры.

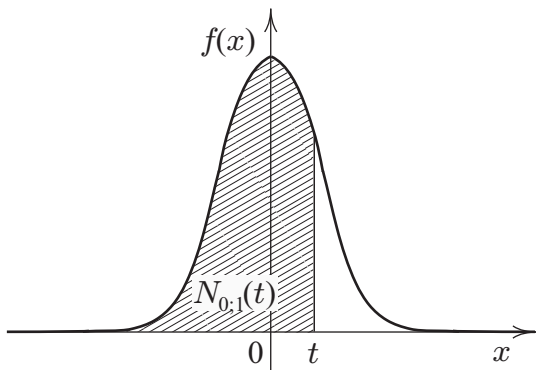


Рис. 22.1.1: К определению квантили нормального распределения; $f(x)$ — плотность распределения $N_{0;1}$

Значение $N_{a;\sigma^2}(x)$ — функции нормального распределения с параметрами $(a; \sigma^2)$ — вычисляется по значениям табулированной функции $N_{0;1}(x) = \Phi(x)$ нормального распределения с параметрами $(0; 1)$ следующим образом:

$$N_{a;\sigma^2}(x) = \Phi\left(\frac{x-a}{\sigma}\right).$$

Таблица 22.1.1 допускает линейную интерполяцию.

Таблица 22.1.1. Значения функции $\Phi(t)$

t	0	1	2	3	4	5	6	7	8	9
-0,0	,5000	,4960	,4920	,4880	,4840	,4801	,4761	,4721	,4681	,4641
-0,1	,4602	,4562	,4522	,4483	,4443	,4404	,4364	,4325	,4286	,4247
-0,2	,4207	,4168	,4129	,4090	,4052	,4013	,3974	,3936	,3897	,3859
-0,3	,3821	,3783	,3745	,3707	,3669	,3632	,3594	,3557	,3520	,3483
-0,4	,3446	,3409	,3372	,3336	,3300	,3264	,3228	,3192	,3156	,3121
-0,5	,3085	,3050	,3015	,2981	,2946	,2912	,2877	,2843	,2810	,2776
-0,6	,2743	,2709	,2676	,2643	,2611	,2578	,2546	,2514	,2483	,2451
-0,7	,2420	,2389	,2358	,2327	,2297	,2266	,2236	,2206	,2177	,2148
-0,8	,2119	,2090	,2061	,2033	,2005	,1977	,1949	,1922	,1894	,1867
-0,9	,1841	,1814	,1788	,1762	,1736	,1711	,1685	,1660	,1635	,1611
-1,0	,1587	,1562	,1539	,1515	,1492	,1469	,1446	,1423	,1401	,1379
-1,1	,1357	,1335	,1314	,1292	,1271	,1251	,1230	,1210	,1190	,1170
-1,2	,1151	,1131	,1112	,1093	,1075	,1056	,1038	,1020	,1003	,0985
-1,3	,0968	,0951	,0934	,0918	,0901	,0885	,0869	,0853	,0838	,0823
-1,4	,0808	,0793	,0778	,0764	,0749	,0735	,0721	,0708	,0694	,0681
-1,5	,0668	,0655	,0643	,0630	,0618	,0606	,0594	,0582	,0571	,0559
-1,6	,0548	,0537	,0526	,0516	,0505	,0495	,0485	,0475	,0465	,0455
-1,7	,0446	,0436	,0427	,0418	,0409	,0401	,0392	,0384	,0375	,0367
-1,8	,0359	,0351	,0344	,0336	,0339	,0322	,0314	,0307	,0301	,0294
-1,9	,0288	,0281	,0274	,0268	,0262	,0256	,0250	,0244	,0239	,0233
-2,0	,0228	,0222	,0217	,0212	,0207	,0202	,0197	,0192	,0188	,0183
-2,1	,0179	,0174	,0170	,0166	,0162	,0158	,0154	,0150	,0146	,0143
-2,2	,0139	,0136	,0132	,0129	,0125	,0122	,0119	,0116	,0113	,0110
-2,3	,0107	,0104	,0102	,0099	,0096	,0094	,0091	,0089	,0087	,0084
-2,4	,0082	,0080	,0078	,0075	,0073	,0071	,0069	,0068	,0066	,0064
-2,5	,0062	,0060	,0059	,0057	,0055	,0054	,0052	,0051	,0049	,0048
-2,6	,0047	,0045	,0044	,0043	,0041	,0040	,0039	,0038	,0037	,0036
-2,7	,0035	,0034	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026
-2,8	,0026	,0025	,0024	,0023	,0023	,0022	,0021	,0021	,0020	,0019
-2,9	,0019	,0018	,0018	,0017	,0016	,0016	,0015	,0015	,0014	,0014
t	-3,0	-3,1	-3,2	-3,3	-3,4	-3,5	-3,6	-3,7	-3,8	-3,9
$\Phi(t)$	,0013	,0010	,0007	,0005	,0003	,0002	,0002	,0001	,0001	,0000

Окончание табл. 22.1.1

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9900	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9923	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	,9987	,9990	,9993	,9995	,9997	,9998	,9998	,9999	,9999	,1000

22.2 χ^2 -распределение

χ^2 -распределением с n степенями свободы (коротко χ_n^2 -распределением) будем называть распределение случайной величины

$$\chi_n^2 = \sum_{i=1}^n \xi_i^2, \quad (22.2.1)$$

где $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, каждая с распределением $N_{0;1}$.

Для данных α и n значение $\chi_{\alpha;n}^2$ определяется как решение уравнения $\int_{\chi_{\alpha;n}^2}^{+\infty} f(x)dx = \alpha$, где $f(x)$ — плотность χ_n^2 -распределения (см. рис. 22.2.1).

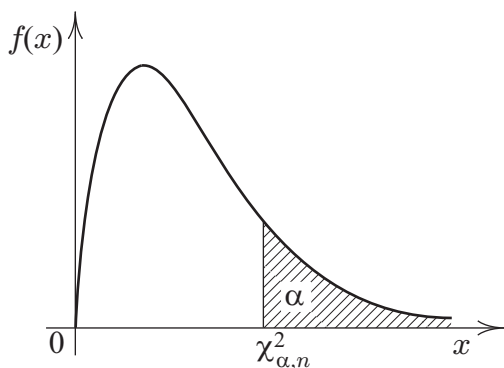


Рис. 22.2.1: К определению $\chi_{\alpha;n}^2$ — верхнего α -предела χ_n^2 -распределения, $f(x)$ — плотность χ_n^2 -распределения

В таблице 22.2.1 приведены значения $\chi_{\alpha;n}^2$, или, что то же, верхние α -пределы (100 α -критические значения) χ^2 -распределения.

Т а б л и ц а 22.2.1. Значения $\chi^2_{\alpha;n}$

n	Значения α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
1	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,64
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21
3	0,12	0,22	0,35	0,58	6,23	7,82	9,35	11,34
4	0,30	0,48	0,71	1,06	7,78	9,48	11,14	13,28
5	0,55	0,83	1,14	1,61	9,24	11,07	12,83	15,09
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09
9	2,90	2,70	3,32	4,17	14,68	16,92	19,02	21,67
10	2,56	3,25	3,94	4,86	15,99	18,31	20,48	23,21
11	3,05	3,82	4,58	5,58	17,28	19,68	21,92	24,72
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41
18	7,02	8,23	9,39	10,86	25,99	28,87	31,53	34,81
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,92
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,26
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89

Окончание табл. 22.2.1

n	Значения α							
	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010
40	22,16	24,43	26,51	29,05	51,80	55,76	59,34	63,69
50	29,71	32,36	34,76	37,69	63,17	67,50	71,42	76,15
60	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38
70	45,44	48,76	51,74	55,33	85,53	90,53	95,02	100,4
80	53,54	57,15	60,39	64,28	96,58	101,9	106,6	112,3
90	61,75	65,65	69,13	73,29	107,6	113,1	118,1	124,1
100	70,06	74,22	77,93	82,36	118,5	124,3	129,6	135,8

22.3 Распределение Стьюдента

Распределением Стьюдента или t -распределением с n степенями свободы (коротко t_n -распределением) будем называть распределение случайной величины

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\chi_n^2}}, \quad (22.3.1)$$

где ξ и χ_n^2 — независимые случайные величины, ξ распределена $N_{0;1}$, а χ_n^2 имеет χ^2 -распределение с n степенями свободы.

В таблице 22.3.1 приведены значения $t_{\alpha;n}$, или, что то же, верхние α -пределы (100 α -критические значения) распределения Стьюдента с n степенями свободы (t_n -распределения).

Для данных α и n значение $t_{\alpha;n}$ определяется как решение уравнения $\int_{t_{\alpha;n}}^{+\infty} f(x)dx = \alpha$, где $f(x)$ — плотность t_n -распределения, $t_{\alpha;n}$ — число, отсекающее правый “хвост” t_n -распределения, на который приходится “масса” α (см. рис. 22.3.1).

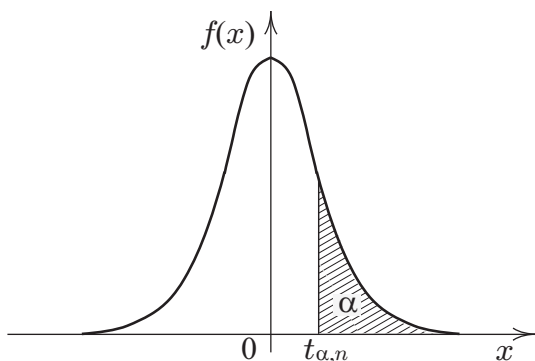


Рис. 22.3.1: К определению $t_{\alpha;n}$ — верхнего α -предела t_n -распределения, $f(x)$ — плотность t_n -распределения

Т а б л и ц а 22.3.1. Значения $t_{\alpha;n}$

n	Значения α				n	Значения α			
	0,050	0,025	0,010	0,005		0,050	0,025	0,010	0,005
1	6,314	12,706	31,821	63,657	18	1,734	2,101	2,552	2,878
2	2,920	4,303	6,965	9,925	19	1,729	2,093	2,539	2,861
3	2,353	3,182	4,541	5,841	20	1,725	2,086	2,528	2,845
4	2,132	2,776	3,747	4,604	21	1,721	2,080	2,518	2,831
5	2,015	2,571	3,365	4,032	22	1,717	2,074	2,508	2,819
6	1,943	2,447	3,143	3,707	23	1,714	2,069	2,500	2,807
7	1,895	2,365	2,998	3,499	24	1,711	2,064	2,492	2,797
8	1,860	2,306	2,896	3,355	25	1,708	2,060	2,485	2,787
9	1,833	2,262	2,821	3,250	26	1,706	2,056	2,479	2,779
10	1,812	2,228	2,764	3,169	27	1,703	2,052	2,473	2,771
11	1,796	2,201	2,718	3,106	28	1,701	2,048	2,467	2,763
12	1,782	2,179	2,681	3,055	29	1,699	2,045	2,462	2,756
13	1,771	2,160	2,650	3,012	30	1,697	2,042	2,457	2,750
14	1,761	2,145	2,624	2,977	40	1,684	2,021	2,423	2,704
15	1,753	2,131	2,602	2,947	60	1,671	2,000	2,390	2,660
16	1,746	2,120	2,583	2,921	120	1,658	1,980	2,358	2,617
17	1,740	2,110	2,567	2,898	∞	1,645	1,960	2,326	2,576

22.4 Распределение Фишера

Распределением Фишера или F -распределением с n, m степенями свободы (коротко $F_{n,m}$ -распределением) будем называть распределение случайной величины

$$F_{n;m} = \frac{\chi_n^2/n}{\chi_m^2/m}, \quad (22.4.1)$$

где χ_n^2 и χ_m^2 — независимые случайные величины, χ_n^2 имеет распределение Пирсона с n степенями свободы, χ_m^2 — с m степенями свободы.

Для данных α, n, m значение $F_{\alpha;n;m}$ определяется как решение уравнения $\int_{F_{\alpha;n;m}}^{+\infty} f(x)dx = \alpha$, где $f(x)$ — плотность $F_{n;m}$ -распределения (см. рис. 22.4.1).

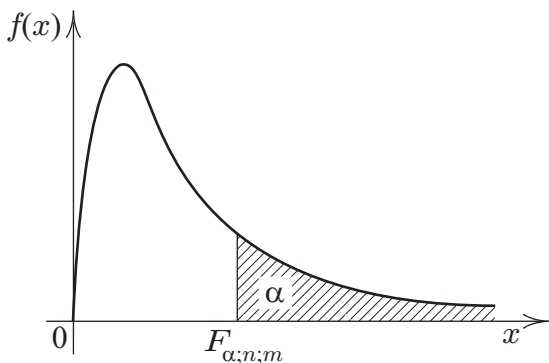


Рис. 22.4.1: К определению $F_{\alpha;n;m}$ — верхнего α -предела $F_{n;m}$ -распределения, $f(x)$ — плотность $F_{n;m}$ -распределения

В таблицах 22.4.1 и 22.4.2 приведены значения $F_{\alpha;n;m}$, или, что то же, верхние α -пределы (100α -критические значения) $F_{n;m}$ -распределения.

Таблица 22.4.1. Значения $F_{\alpha;n;m}$ для $\alpha = 0,05$

m	n (число степеней свободы числителя)									
	1	2	3	4	5	6	7	8	9	10
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93

Таблица 22.4.1 (окончание)

m	n (число степеней свободы числителя)									
	12	14	16	18	20	30	40	50	60	100
3	8,74	8,71	8,69	8,67	8,66	8,62	8,59	8,58	8,57	8,55
4	5,91	5,87	5,84	5,82	5,80	5,75	5,72	5,70	5,69	5,66
5	4,68	4,64	4,60	4,58	4,56	4,50	4,46	4,44	4,43	4,41
6	4,00	3,96	3,92	3,90	3,87	3,81	3,77	3,75	3,74	3,71
7	3,57	3,53	3,49	3,47	3,44	3,38	3,34	3,32	3,30	3,27
8	3,28	3,24	3,20	3,17	3,15	3,08	3,04	3,02	3,01	2,97
9	3,07	3,03	2,99	2,96	2,94	2,86	2,83	2,80	2,79	2,76
10	2,91	2,86	2,83	2,80	2,77	2,70	2,66	2,64	2,62	2,59
11	2,79	2,74	2,70	2,67	2,65	2,57	2,53	2,51	2,49	2,46
12	2,69	2,64	2,60	2,57	2,54	2,47	2,43	2,40	2,38	2,35
13	2,60	2,55	2,51	2,48	2,46	2,38	2,34	2,31	2,30	2,26
14	2,53	2,48	2,44	2,41	2,39	2,31	2,27	2,24	2,22	2,19
15	2,48	2,42	2,38	2,35	2,33	2,25	2,20	2,18	2,16	2,12
16	2,42	2,37	2,33	2,30	2,28	2,19	2,15	2,12	2,11	2,07
17	2,38	2,33	2,29	2,26	2,23	2,15	2,10	2,08	2,06	2,02
18	2,34	2,29	2,25	2,22	2,19	2,11	2,06	2,04	2,02	1,98
19	2,31	2,26	2,21	2,18	2,16	2,07	2,03	2,00	1,98	1,94
20	2,28	2,22	2,18	2,15	2,12	2,04	1,99	1,97	1,95	1,91
22	2,23	2,17	2,13	2,10	2,07	1,98	1,94	1,91	1,89	1,85
24	2,18	2,13	2,09	2,05	2,03	1,94	1,89	1,86	1,84	1,80
26	2,15	2,09	2,05	2,02	1,99	1,90	1,85	1,82	1,80	1,76
28	2,12	2,06	2,02	1,99	1,96	1,87	1,82	1,79	1,77	1,73
30	2,09	2,04	1,99	1,96	1,93	1,84	1,79	1,76	1,74	1,70
40	2,00	1,95	1,90	1,87	1,84	1,74	1,69	1,66	1,64	1,59
50	1,95	1,89	1,85	1,81	1,78	1,69	1,63	1,60	1,58	1,52
60	1,92	1,86	1,82	1,78	1,75	1,65	1,59	1,56	1,53	1,48
100	1,85	1,79	1,75	1,71	1,68	1,57	1,52	1,48	1,45	1,39

Таблица 22.4.2. Значения $F_{\alpha;n;m}$ для $\alpha = 0,01$

m	n (число степеней свободы числителя)									
	1	2	3	4	5	6	7	8	9	10
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50

Таблица 22.4.2 (окончание)

m	n (число степеней свободы числителя)									
	12	14	16	18	20	30	40	50	60	100
3	27,1	26,9	26,8	26,8	26,7	26,5	26,4	26,4	26,3	26,2
4	14,4	14,2	14,2	14,1	14,0	13,8	13,7	13,7	13,7	13,6
5	9,89	9,77	9,68	9,61	9,55	9,38	9,29	9,24	9,20	9,13
6	7,72	7,60	7,52	7,45	7,40	7,23	7,14	7,09	7,06	6,99
7	6,47	6,36	6,27	6,21	6,16	5,99	5,91	5,86	5,82	5,75
8	5,67	5,56	5,48	5,41	5,36	5,20	5,12	5,07	5,03	4,96
9	5,11	5,00	4,92	4,86	4,81	4,65	4,57	4,52	4,48	4,42
10	4,71	4,60	4,52	4,46	4,41	4,25	4,17	4,12	4,08	4,01
11	4,40	4,29	4,21	4,15	4,10	3,94	3,86	3,81	3,78	3,71
12	4,16	4,05	3,97	3,91	3,86	3,70	3,62	3,57	3,54	3,47
13	3,96	3,86	3,78	3,72	3,66	3,51	3,43	3,38	3,34	3,27
14	3,80	3,70	3,62	3,56	3,51	3,35	3,27	3,22	3,18	3,11
15	3,67	3,56	3,49	3,42	3,37	3,21	3,13	3,08	3,05	2,98
16	3,55	3,45	3,37	3,31	3,26	3,10	3,02	2,97	2,93	2,86
17	3,46	3,35	3,27	3,21	3,16	3,00	2,92	2,87	2,83	2,76
18	3,37	3,27	3,19	3,13	3,08	2,92	2,84	2,78	2,75	2,68
19	3,30	3,19	3,12	3,05	3,00	2,84	2,76	2,71	2,67	2,60
20	3,23	3,13	3,05	2,99	2,94	2,78	2,69	2,64	2,61	2,54
22	3,12	3,02	2,94	2,88	2,83	2,67	2,58	2,53	2,50	2,42
24	3,03	2,93	2,85	2,79	2,74	2,58	2,49	2,44	2,40	2,33
26	2,96	2,86	2,78	2,72	2,66	2,50	2,42	2,36	2,33	2,25
28	2,90	2,79	2,72	2,65	2,60	2,44	2,35	2,30	2,26	2,19
30	2,84	2,74	2,66	2,60	2,55	2,39	2,30	2,25	2,21	2,13
40	2,66	2,56	2,48	2,42	2,37	2,20	2,11	2,06	2,02	1,94
50	2,56	2,46	2,38	2,32	2,27	2,10	2,01	1,95	1,91	1,82
60	2,50	2,39	2,31	2,25	2,20	2,03	1,94	1,88	1,84	1,75
100	2,37	2,26	2,19	2,12	2,07	1,89	1,80	1,73	1,69	1,60

22.5 Биномиальное распределение

При получении значений $B_{n;p}(i)$ для $p \leq 0,5$ с помощью табл. 22.5.1 не пользуются двумя крайними правыми столбцами и двумя последними (нижними) строками, а при получении значений $B_{n;p}(i)$ для $p \geq 0,5$ не пользуются двумя крайними левыми столбцами и двумя первыми (верхними) строками.

Таблица 22.5.1. Значения $B_{n;p}(i) = C_n^i p^i (1-p)^{n-i}$

n	i	p						
		0,1	0,2	0,3	0,4	0,5		
10	0	0,34868	0,10737	0,02825	0,00605	0,00098	10	10
	1	,38742	,26844	,12106	,04031	,00977	9	
	2	,19371	,30199	,23347	,12093	,04395	8	
	3	,05740	,20133	,26683	,21499	,11719	7	
	4	,01116	,08808	,20012	,25082	,20508	6	
	5	0,00149	0,02642	0,10292	0,20066	0,24609	5	
	6	,00014	,00551	,03676	,11148	,20508	4	
	7	,00001	,00079	,00900	,04247	,11719	3	
	8		,00007	,00145	,01062	,04395	2	
	9			,00014	,00157	,00977	1	
	10			,00001	,00010	,00098	0	
15	0	0,20589	0,03518	0,00475	0,00047	0,00003	15	15
	1	,34315	,13194	,03052	,00470	,00046	14	
	2	,26690	,23090	,09156	,02194	,00320	13	
	3	,12851	,25014	,17004	,06339	,01389	12	
	4	,04284	,18760	,21862	,12678	,04166	11	
	5	0,01047	0,10318	0,20613	0,18594	0,09164	10	
	6	,00194	,04299	,14724	,20660	,15274	9	
	7	,00028	,01382	,08113	,17708	,19638	8	
	8	,00003	,00345	,03477	,11806	,19638	7	
	9		,00067	,01159	,06121	,15274	6	
		0,9	0,8	0,7	0,6	0,5	i	n
p								

Продолжение табл. 22.5.1

n	i	p						
		0,1	0,2	0,3	0,4	0,5		
15	10		0,00010 ,00001	0,00298 ,00058 ,00008 ,00001	0,02449 ,00742 ,00165 ,00025 ,00002	0,09164 ,04166 ,01389 ,00320 ,00046 ,00003	5	15
	11						4	
	12						3	
	13						2	
	14						1	
	15						0	
20	0	0,12158	0,01153	0,00080	0,00004		20	20
	1	,27017	,05765	,00684	,00049	,00002	19	
	2	,28518	,13691	,02785	,00309	,00018	18	
	3	,19012	,20536	,07160	,01235	,00109	17	
	4	,08978	,21820	,13042	,03499	,00462	16	
	5	0,03192	0,17456	0,17886	0,07465	0,01479	15	
	6	,00887	,10910	,19164	,12441	,03696	14	
	7	,00197	,05455	,16426	,16588	,07393	13	
	8	,00036	,02216	,11440	,17971	,12013	12	
	9	,00005	,00739	,06537	,15974	,16018	11	
	10	0,00001	0,00203	0,03082	0,11714	0,17620	10	
	11		,00046	,01201	,07099	,16018	9	
	12		,00009	,00386	,03550	,12013	8	
	13		,00001	,00102	,01456	,07393	7	
	14			,00022	,00485	,03696	6	
	15			0,00004	0,00129	0,01479	5	
	16			,00001	,00027	,00462	4	
	17				,00004	,00109	3	
	18					,00018	2	
	19					,00002	1	
20						0		
		0,9	0,8	0,7	0,6	0,5	i	n
		p						

Окончание табл. 22.5.1

n	i	p						
		0,1	0,2	0,3	0,4	0,5		
25	0	0,07179	0,00378	0,00013			25	25
	1	,19942	,02361	,00144	,00005		24	
	2	,26589	,07084	,00739	,00038	,00001	23	
	3	,22650	,13577	,02428	,00194	,00007	22	
	4	,13842	,18668	,05723	,00710	,00038	21	
	5	0,06459	0,19602	0,10302	0,01989	0,00158	20	
	6	,02392	,16335	,14717	,04420	,00528	19	
	7	,00722	,11084	,17119	,07999	,01433	18	
	8	,00180	,06235	,16508	,11998	,03223	17	
	9	,00038	,02944	,13364	,15109	,06089	16	
	10	0,00007	0,01178	0,09164	0,16116	0,09742	15	
	11	,00001	,00401	,05355	,14651	,13284	14	
	12		,00117	,02678	,11395	,15498	13	
	13		,00029	,01148	,07597	,15498	12	
	14		,00006	,00422	,04341	,13284	11	
	15		0,00001	0,00132	0,02122	0,09742	10	
	16			,00035	,00884	,06089	9	
	17			,00008	,00312	,03223	8	
	18			,00002	,00092	,01433	7	
	19				,00023	,00528	6	
	20				0,00005	0,00158	5	
	21				,00001	,00038	4	
	22					,00007	3	
	23					,00001	2	
		0,9	0,8	0,7	0,6	0,5	i	n
		p						

22.6 Распределение Пуассона

Т а б л и ц а 22.6.1. Значения $P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

k	Значения λ								
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	,9048	,8187	,7408	,6703	,6065	,5488	,4966	,4493	,4066
1	,0905	,1637	,2222	,2681	,3033	,3293	,3476	,3595	,3659
2	,0045	,0164	,0333	,0536	,0758	,0988	,1217	,1438	,1647
3	,0002	,0011	,0033	,0072	,0126	,0198	,0284	,0383	,0494
4		,0001	,0003	,0007	,0016	,0030	,0050	,0077	,0111
5				,0001	,0002	,0004	,0007	,0012	,0020
6							,0001	,0002	,0003

k	Значения λ								
	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
0	,3679	,1353	,0498	,0183	,0067	,0025	,0009	,0003	,0001
1	,3679	,2707	,1494	,0733	,0337	,0149	,0064	,0027	,0011
2	,1839	,2707	,2240	,1465	,0842	,0446	,0223	,0107	,0050
3	,0613	,1804	,2240	,1954	,1404	,0892	,0521	,0286	,0150
4	,0153	,0902	,1680	,1954	,1755	,1339	,0912	,0573	,0337
5	,0031	,0361	,1008	,1563	,1755	,1606	,1277	,0916	,0607
6	,0005	,0120	,0504	,1042	,1462	,1606	,1490	,1221	,0911
7	,0001	,0034	,0216	,0595	,1044	,1377	,1490	,1396	,1171
8		,0009	,0081	,0298	,0653	,1033	,1304	,1396	,1318
9		,0002	,0027	,0132	,0363	,0688	,1014	,1241	,1318
10			,0008	,0053	,0181	,0413	,0710	,0993	,1186
11			,0002	,0019	,0082	,0225	,0452	,0722	,0970
12			,0001	,0006	,0034	,0113	,0264	,0481	,0728
13				,0002	,0013	,0052	,0142	,0296	,0504
14				,0001	,0005	,0022	,0071	,0169	,0324
15					,0002	,0009	,0033	,0090	,0194
16						,0003	,0014	,0045	,0109
17						,0001	,0006	,0021	,0058

22.7 Критерий А. Н. Колмогорова. Критические значения

В таблице 22.7.1 приведены критические значения $\varepsilon_{\alpha;n}$ супремума модуля разности истинной и эмпирической функций распределений.

Значение $\varepsilon_{\alpha;n}$ для данных α и n определяется как минимальное ε , для которого

$$P\{\sup_x |F(x) - \hat{F}_n(x)| \geq \varepsilon\} \leq \alpha.$$

Т а б л и ц а 22.7.1. Критические значения $\varepsilon_{\alpha;n}$ для супремума модуля разности истинной и эмпирической функций распределения

n	Значения α			n	Значения α		
	0,05	0,02	0,01		0,05	0,02	0,01
1	0,9750	0,9900	0,9950	25	0,2640	0,2952	0,3166
2	0,8419	0,9000	0,9293	30	0,2417	0,2702	0,2899
3	0,7076	0,7846	0,8290	35	0,2243	0,2507	0,2690
4	0,6239	0,6889	0,7342	40	0,2101	0,2349	0,2520
5	0,5633	0,6272	0,6685	45	0,1984	0,2218	0,2380
6	0,5193	0,5774	0,6166	50	0,1884	0,2107	0,2260
7	0,4834	0,5384	0,5758	55	0,1798	0,2011	0,2157
8	0,4543	0,5065	0,5418	60	0,1723	0,1927	0,2067
9	0,4300	0,4796	0,5133	65	0,1657	0,1853	0,1988
10	0,4093	0,4566	0,4889	70	0,1598	0,1786	0,1917
11	0,3912	0,4367	0,4677	75	0,1544	0,1727	0,1853
12	0,3754	0,4192	0,4491	80	0,1496	0,1673	0,1795
13	0,3614	0,4036	0,4325	85	0,1452	0,1624	0,1742
14	0,3489	0,3897	0,4176	90	0,1412	0,1579	0,1694
15	0,3376	0,3771	0,4042	95	0,1375	0,1537	0,1649
20	0,2941	0,3287	0,3524	100	0,1340	0,1499	0,1608

При $n > 100$ можно считать, что

$$\varepsilon_{0,05;n} = \frac{1,36}{\sqrt{n}}; \quad \varepsilon_{0,01;n} = \frac{1,63}{\sqrt{n}}.$$

22.8 Критерий Вилкоксона.

Нижние критические значения

В таблице 22.8.1 приведены нижние критические значения $W_{\alpha;n;m}$ распределения W — суммы рангов выборки меньшего объема.

Значения $W_{\alpha;n;m}$ для данных α (уровня значимости), n и m — объемов выборок (n — объем меньшей выборки, m — большей) определяется как наибольшее целое t , для которого

$$P\{W \leq t\} \leq \alpha.$$

При значениях n и m больших приведенных в таблице (а фактически при n и m , удовлетворяющих неравенствам $\min\{n, m\} \geq 6, m + n \geq 20$), можно считать, что $W_{\alpha;n;m}$ равно

$$\frac{1}{2}n(n + m + 1) + z_{\alpha}\sqrt{\frac{1}{12}nm(n + m + 1)},$$

где z_{α} — это α -квантиль нормального распределения с параметрами $(0;1)$ (см. табл. 22.1.1).

Т а б л и ц а 22.8.1. Нижние критические значения $W_{\alpha;n;m}$ распределения W

Объ- емы		Значения α				Объ- емы		Значения α			
n	m	0,005	0,01	0,025	0,05	n	m	0,005	0,01	0,025	0,05
6	6	23	24	26	28	6	18	37	40	45	49
	7	24	25	27	29		19	38	41	46	51
	8	25	27	29	31	7	7	32	34	36	39
	9	26	28	31	33		8	34	35	38	41
	10	27	29	32	35		9	35	37	40	43
	11	28	30	34	37		10	37	39	42	45
	12	30	32	35	38		11	38	40	44	47
	13	31	33	37	40		12	40	42	46	49
	14	32	34	38	42		13	41	44	48	52
	15	33	36	40	44		14	43	45	50	54
	16	34	37	42	46		15	44	47	52	56
	17	36	39	43	47		16	46	49	54	58

Таблица 22.8.1 (окончание)

Объ- емы		Значения α				Объ- емы		Значения α			
n	m	0,005	0,01	0,025	0,05	n	m	0,005	0,01	0,025	0,05
7	17	47	51	56	61	9	13	65	68	73	78
	18	49	52	58	63		14	67	71	76	81
8	8	43	45	49	51		15	69	73	79	84
	9	45	47	51	54		16	72	76	82	87
	10	47	49	53	56	10	10	71	74	78	82
	11	49	51	55	59		11	73	77	81	86
	12	51	53	58	62		12	76	79	84	89
	13	53	56	60	64		13	79	82	88	92
	14	54	58	62	67		14	81	85	91	96
	15	56	60	65	69		15	84	88	94	99
	16	58	62	67	72	11	11	87	91	96	100
	17	60	64	70	75		12	90	94	99	104
9	9	56	59	62	66		13	93	97	103	108
	10	58	61	65	69		14	96	100	106	112
	11	61	63	68	72	12	12	105	109	115	120
	12	63	66	71	75		13	109	113	119	125

22.9 Критерий знаков. Границы критической области

В таблице 22.9.1 приведены левая ($n - m_{\alpha;n}$) (левый столбец) и правая $m_{\alpha;n}$ (правый столбец) границы критической области критерия знаков (области отклонения гипотезы $H_0: \theta = 0$).

Значения $m_{\alpha;n}$ для данных α (уровня значимости) и n (числа отличных от нуля разностей) определяется как минимальное целое m , для которого $P\{\mu > m\} \leq \alpha$, где μ — биномиально распределенная случайная величина с параметрами n и $1/2$.

Критические области критерия знаков:

$(m_{\alpha;n}; n]$ для односторонней альтернативы $\theta > 0$;

$[0; n - m_{\alpha;n})$ для односторонней альтернативы $\theta < 0$;

$[0; n - m_{\alpha;n}) \cup (m_{\alpha;n}; n]$ для двусторонней альтернативы $\theta < 0$ или $\theta > 0$;

Уровень значимости одностороннего критерия не превышает α , двустороннего — 2α .

Таблица 22.9.1. Границы критических областей критерия знаков

n	Значения α						n	Значения α					
	0,025		0,010		0,005			0,025		0,010		0,005	
5	0	5	0	5	0	5	25	8	17	7	18	6	19
6	1	5	0	6	0	6	26	8	18	7	19	7	19
7	1	6	1	6	0	7	27	8	19	8	19	7	20
8	1	7	1	7	1	7	28	9	19	8	20	7	21
9	2	7	1	8	1	8	29	9	20	8	21	8	21
10	2	8	1	9	1	9	30	10	20	9	21	8	22
11	2	9	2	9	1	10	31	10	21	9	22	8	23
12	3	9	2	10	2	10	32	10	22	9	23	9	23
13	3	10	2	11	2	11	33	11	22	10	23	9	24
14	3	11	3	11	2	12	34	11	23	10	24	10	24
15	4	11	3	12	3	12	35	12	23	11	24	10	25
16	4	12	3	13	3	13	36	12	24	11	25	10	26
17	5	12	4	13	3	14	37	13	24	11	26	11	26
18	5	13	4	14	4	14	38	13	25	12	26	11	27
19	5	14	5	14	4	15	39	13	26	12	27	12	27
20	6	14	5	15	4	16	40	14	26	13	27	12	28
21	6	15	5	16	5	16	41	14	27	13	28	12	29
22	6	16	6	16	5	17	42	15	27	14	28	13	29
23	7	16	6	17	5	18	43	15	28	14	29	13	30
24	7	17	6	18	6	18	44	16	28	14	30	14	30

Таблица 22.9.1 (окончание)

n	Значения α						n	Значения α					
	0,025	0,010	0,005					0,025	0,010	0,005			
45	16	29	15	30	14	31	73	28	45	27	46	26	47
46	16	30	15	31	14	32	74	29	45	27	47	26	48
47	17	30	16	32	15	32	75	29	46	27	48	26	49
48	17	31	16	32	15	33	76	29	47	28	48	27	49
49	18	31	16	33	16	33	77	30	47	28	49	27	50
50	18	32	17	33	16	34	78	30	48	29	49	28	50
51	19	32	17	34	16	35	79	31	48	29	50	28	51
52	19	33	18	34	17	35	80	31	49	30	50	29	51
53	19	34	18	35	17	36	81	32	49	30	51	29	52
54	20	34	19	35	18	36	82	32	50	31	51	29	53
55	20	35	19	36	18	37	83	33	50	31	52	30	53
56	21	35	19	37	18	38	84	33	51	31	53	30	54
57	21	36	20	37	19	38	85	33	52	32	53	31	54
58	22	36	20	38	19	39	86	34	52	32	54	31	55
59	22	37	21	38	20	39	87	34	53	33	54	32	55
60	22	38	21	39	20	40	88	35	53	33	55	32	56
61	23	38	21	40	21	40	89	35	54	34	55	32	57
62	23	39	22	40	21	41	90	36	54	34	56	33	57
63	24	39	22	41	21	42	91	36	55	34	57	33	58
64	24	40	23	41	22	42	92	37	55	35	57	34	58
65	25	40	23	42	22	43	93	37	56	35	58	34	59
66	25	41	24	42	23	43	94	38	56	36	58	35	59
67	26	41	24	43	23	44	95	38	57	36	59	35	60
68	26	42	24	44	23	45	96	38	58	37	59	35	61
69	26	43	25	44	24	45	97	39	58	37	60	36	61
70	27	43	25	45	24	46	98	39	59	38	60	36	62
71	27	44	26	45	25	46	99	40	59	38	61	37	62
72	28	44	26	46	25	47	100	40	60	38	62	37	63

22.10 Равномерно распределенные случайные числа

Приведенные в таблице 22.10.1 цифры можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 с одной и той же вероятностью 0,1.

Табулированные цифры сгруппированы по две. Пары можно рассматривать как реализации независимых и одинаково распределенных случайных величин, принимающих значения от 00 до 99 с одной и той же вероятностью 0,01.

Аналогичные утверждения можно сформулировать, если цифры группировать по три, четыре ...

Рассмотрим группы из k цифр как целые числа. Перемножим каждое из них на 10^{-k} . Полученные числа можно считать реализациями независимых случайных величин, равномерно распределенных на отрезке $[0; 1]$.

Т а б л и ц а 22.10.1. Равномерно распределенные случайные числа

10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68

Таблица 22.10.1 (продолжение)

65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56
98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27

Т а б л и ц а 22.10.1 (продолжение)

74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50
73 18 95 02 07	47 67 72 62 69	62 29 06 44 64	27 12 46 70 18
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56

Т а б л и ц а 22.10.1 (окончание)

07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51
59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31
24 63 73 87 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85
15 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79
03 99 11 04 61	93 71 61 68 94	66 08 32 46 53	84 60 95 82 32
38 55 59 55 54	32 88 65 97 80	08 35 56 08 60	29 73 54 77 62
17 54 67 37 04	92 05 24 62 15	55 12 12 92 81	59 07 60 79 36
32 64 35 28 61	95 81 90 68 31	00 91 19 89 36	76 35 59 37 79
69 57 26 87 77	39 51 03 59 05	14 06 04 06 19	29 54 96 96 16

Литература

- [1] *Большев Л. Н.* Таблицы математической статистики/ Л. Н. Большев, Н. В. Смирнов — 3-е изд. — М.: Наука, 1983. — 416 с.
- [2] *Ван дер Варден Б. Л.* Математическая статистика. — М.: Изд-во иностр. лит., 1960. — 434 с.
- [3] *Гихман И. И.* Теория вероятностей и математическая статистика/ И. И. Гихман, А. В. Скороход, М. И. Ядренко — К.: Вища шк., 1979. — 320 с.
- [4] *Джонсон Н.* Статистика и планирование эксперимента в науке и технике. Методы обработки данных/ Н. Джонсон, Ф. Лион — М.: Наука, 1980. — 600 с.
- [5] *Емельянов Г. В.* Задачник по теории вероятностей и математической статистике/ Г. В. Емельянов, В. П. Скитович — Л.: Изд-во Ленинград. ун-та, 1967. — 332 с.
- [6] *Зубков А. М.* Сборник задач по теории вероятностей/ А. М. Зубков, Б. А. Севастьянов, В. В. Чистяков — М.: Наука, 1989. — 320 с.
- [7] *Крамер Г.* Математические методы статистики. — 2-е изд., перераб. — М.: Мир, 1975. — 648 с.
- [8] *Кендалл М. Дж.* Теория распределений/ М. Дж. Кендалл, А. Стьюарт — М.: Наука, 1966. — 587 с.
- [9] *Мешалкин Л. Д.* Сборник задач по теории вероятностей. — М.: Изд-во Москов. ун-та, 1963. — 155 с.

- [10] *Прохоров А. В.* Задачи по теории вероятностей: Основные понятия. Предельные теоремы. Случайные процессы/ А. В. Прохоров, В. Г. Ушаков, Н. Г. Ушаков — М.: Наука, 1986. — 328 с.
- [11] *Турчин В. Н.* Теория вероятностей и математическая статистика: — Днепропетровск. — Д.: Изд-во Днепропетров. ун-та, 2008. — 656 с..
- [12] *Турчин В. М.* Математична статистика. — К.: Видавничий центр “Академія”, 1999. — 240 с.
- [13] *Турчин В. М.* Теорія ймовірностей. Основні поняття, приклади, задачі. — К.: А.С.К., 2003. — 208 с.
- [14] *Турчин В. М.* Теорія ймовірностей і математична статистика. Основні поняття, приклади, задачі. — Д.: Видавництво ДНУ, 2006. — 476 с.
- [15] *Тутубалин В. Н.* Теория вероятностей. — М.: Изд-во Москов. ун-та, 1972. — 230 с.
- [16] *Дороговцев А. Я.* Теорія ймовірностей. Зб. задач/ А. Я. Дороговцев, Д. С. Сільвестров, А. В. Скороход, М. Й. Ядренко — К.: Вища шк., 1976. — 384 с.
- [17] *Феллер В.* Введение в теорию вероятностей и ее приложения: В 2-х т.— М.: Мир, 1984. — Т. 1. — 527 с.; Т. 2. — 751 с.
- [18] *Хальд А.* Математическая статистика с техническими приложениями. — М.: ИЛ, 1956. — 664 с.
- [19] *Ядренко М. Й.* Дискретна математика. — К.: ВПЦ “Експрес”, 2003. — 244 с.
- [20] *David F. N.* Elementary Statistical Exercises, David F. N., Pearson E. S. — Cambridge: University Press, 1961 — 241 p.

Оглавление

1	Элементы комбинаторики	5
1.1	Основной принцип комбинаторики	5
1.2	Задачи	13
2	Стохастический эксперимент	17
2.1	Пространство элементарных событий, алгебра событий	17
2.2	Задачи	21
3	Дискретное вероятностное пространство	27
3.1	Вероятность, классическая модель	27
3.2	Задачи	33
4	Условная вероятность	39
4.1	Формула полной вероятности. Формулы Байеса	39
4.2	Независимые события	45
4.3	Задачи	47
5	Дискретная случайная величина и ее распределение	53
5.1	Вычисление распределения функции от случайной величины	53
5.2	Дискретные распределения на $\mathbb{R}^1$	59
5.3	Задачи	64

6	Математическое ожидание дискретной случайной величины	73
6.1	Определения, свойства, вычисления	73
6.2	Задачи	80
7	Аксиоматика теории вероятностей	89
7.1	Алгебры, σ -алгебры	89
7.2	Аксиомы Колмогорова	93
7.3	Задачи	98
8	Геометрические вероятности	103
8.1	Определения. Примеры	103
8.2	Задачи	109
9	Распределение случайной величины	117
9.1	Функция и плотность распределения	117
9.2	Функция и плотность распределения случайного вектора	120
9.3	Абсолютно непрерывные распределения на $\mathbb{R}^1$	125
9.4	Задачи	127
10	Математическое ожидание	135
10.1	Определения, свойства, вычисление	135
10.2	Задачи	143
11	Свертка	151
11.1	Свертка вероятностных распределений	151
11.2	Распределение суммы независимых случайных величин	155
11.3	Задачи	159
12	Сходимость распределений	163
12.1	Определения. Примеры	163
12.2	Задачи	169

13	Характеристическая функция	175
13.1	Определения, свойства, вычисление	175
13.2	Теоремы единственности и непрерывности	178
13.3	Задачи	181
14	Оценивание параметров распределений	189
14.1	Точечные оценки	189
14.2	Доверительные интервалы	198
14.3	Оценки с минимальной дисперсией	201
14.4	Задачи	213
15	Методы построения оценок	233
15.1	Эмпирические оценки	233
15.2	Метод моментов	245
15.3	Метод максимального правдоподобия	248
15.4	Задачи	253
16	Задача проверки статистических гипотез	271
16.1	Критерий, функция мощности критерия	271
16.2	Задачи	284
17	Проверка гипотез о параметрах распределения $N_{a;\sigma^2}$	299
17.1	Проверка гипотезы $H_0: a = a_0$	299
17.2	Проверка гипотезы $H_0: a_\xi = a_\eta$	305
17.3	Проверка гипотезы $H_0: \sigma^2/\sigma_0^2 = 1$	311
17.4	Проверка гипотезы $H_0: \sigma_\xi^2/\sigma_\eta^2 = 1$	313
17.5	Задачи	317

18	Критерий χ^2	331
18.1	Критерий χ^2 (гипотетическое распределение не зависит от параметров)	331
18.2	Критерий χ^2 (параметры неизвестны)	337
18.3	Критерий χ^2 как критерий независимости	340
18.4	Задачи	345
19	Непараметрические критерии	379
19.1	Критерий Колмогорова	379
19.2	Критерий знаков	384
19.3	Критерий Вилкоксона	390
19.4	Задачи	398
20	Линейная регрессия	419
20.1	Нормальная линейная регрессия	419
20.2	Задачи	433
21	Решения, указания, ответы	443
21.1	К главе 1	443
21.2	К главе 2	445
21.3	К главе 3	448
21.4	К главе 4	453
21.5	К главе 5	460
21.6	К главе 6	470
21.7	К главе 7	477
21.8	К главе 8	478
21.9	К главе 9	481
21.10	К главе 10	490
21.11	К главе 11	495
21.12	К главе 12	501
21.13	К главе 13	502
21.14	К главе 14	506
21.15	К главе 15	510
21.16	К главе 16	517
21.17	К главе 17	519
21.18	К главе 18	529
21.19	К главе 19	530
21.20	К главе 20	533

21.21	Задания для самостоятельной работы	537
22	Таблицы математической статистики	543
22.1	Нормальное распределение	544
22.2	χ^2 -распределение	547
22.3	Распределение Стьюдента	549
22.4	Распределение Фишера	551
22.5	Биномиальное распределение	556
22.6	Распределение Пуассона	559
22.7	Критерий А. Н. Колмогорова. Критические значения	560
22.8	Критерий Вилкоксона. Нижние критические значения	561
22.9	Критерий знаков. Границы критической области	562
22.10	Равномерно распределенные случайные числа	565
	Литература	569

Навчальне видання

ВАЛЕРІЙ МИКОЛАЙОВИЧ ТУРЧИН

**ТЕОРІЯ ЙМОВІРНОСТЕЙ
І МАТЕМАТИЧНА СТАТИСТИКА**
Основні поняття, приклади, задачі

**Підручник для студентів вищих навчальних
закладів**

(Російською мовою)

Редактор Козаченко Ю.В.
Відповідальний за випуск Турчин Є.В.
Художник Ткаченко К.Д.
Корректор Литвиненко М.Г.
Оригинал-макет Турчин В.М.

Підписано до друку 20.10.2011 р. Формат 84×108 $\frac{1}{32}$.

Друк офсетний. Папір офсетний.

Гарнітура Computer modern

Ум. др. арк. 30,24.

Зам. № 0-9339. Тираж 500 прим. 1-й завод —1—250 прим.

Видавництво ІМА-ПРЕС,

49051, м. Дніпропетровськ, вул. Журналістів, 7/215.

050-514-52-75, 067-910-61-01. E-mail: IMA@optima.com.ua

Свідоцтво про внесення до Державного реєстру суб'єктів
видавничої справи ДК № 244 від 16.11.2000 р.

Віддруковано у ПРАТ “Видавництво “Зоря”,

м.Дніпропетровськ, вул. Журналістів, 7

Свідоцтво про внесення до державного реєстру

ДК № 4127 від 28.07.2011 р.

ТУРЧИН Валерий Николаевич

Автор 95 научных
и научно- методических работ,
12 учебников и учебных пособий
по теории вероятностей
и математической статистике.
Заведующий кафедрой
статистики и теории вероятностей
Днепропетровского
национального университета.

