МИФИ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

(часть 4)

Москва 2008

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

МОСКОВСКИЙ ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации

(часть 4)

Москва 2008

УДК 519.2(07)
ББК 22.17я7
М 54

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА: Методические рекомендации (часть 4). В 4-х частях. М.: МИФИ, 2008. 108 с.

Данная, 4-я часть методических рекомендаций завершает исправленные издания (части 1, 2, 3) [8 — 10] ранее опубликованных работ [4 — 7]. Настоящая работа является исправленным изданием работы [7] и содержит решения задач из § 6 и § 7 сборника задач [3]. В теоретической части раздела 7 использован материал, взятый из работы [2]. Необходимые ссылки на теоретический материал (определения, теоремы и т.д.) приведены по учебнику: Чистяков В.П. Курс теории вероятностей. М.: Агар, 2000.

Методические рекомендации предназначены для студентов, изучающих курс теории вероятностей и математической статистики в течение одного семестра, и посвящены разбору задач из книги: Полякова Е.И., Постникова Л.П. Сборник задач по теории вероятностей и математической статистике. М.: МИФИ, 2004. Указанные методические рекомендации будут полезны также преподавателям, ведущим практические занятия по теории вероятностей и математической статистике.

Составители: Е.И. Полякова, Л.П. Постникова, Е.В. Сумин

Рекомендовано к изданию редсоветом МИФИ

Редактор и технический редактор М.В. Макарова
Оригинал-макет изготовлен М.В. Макаровой

Подписано в печать 30.06.2008. Формат 60x84 1/16.
Псч.л. 6,75. Уч.-изд.л. 6,75. Тираж 100 экз.
Изд. № 005-1 Заказ № 217

Московский инженерно-физический институт
(государственный университет).
Типография МИФИ.
115409, Москва, Каширское ш., 31

6. ЗАКОН БОЛЬШИХ ЧИСЕЛ. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА. ИНТЕГРАЛЬНАЯ ТЕОРЕМА МУАВРА-ЛАПЛАСА. ТЕОРЕМА ПУАССОНА

Задача 1

Найти $P\{|\xi - M\xi| \ge 3\sqrt{D\xi}\}$, если $\xi(\omega)$ имеет: а) нормальное распределение; б) равномерное распределение на отрезке $[-1, 1]$; в) распределение Пуассона с параметром $\lambda = 0{,}09$.

Решение

В [1], гл. 6, § 3, примеры 1 — 3, с. 111 — 112, найдены средние значения и дисперсии случайных величин, рассматриваемых в задаче.

а) Плотность нормального распределения равна

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

$a = M\xi$, $\sigma^2 = D\xi$. Тогда

$$P\{|\xi - M\xi| \ge 3\sqrt{D\xi}\} = P\{|\xi - a| \ge 3\sigma\} = 1 - P\{|\xi - a| < 3\sigma\}$$

и согласно равенству (3.1) (см. [1], гл. 5, § 3, с. 81)

$$P\{|\xi - a| < 3\sigma\} = P\{a - 3\sigma < \xi < a + 3\sigma\} =$$

$$= \int_{a-3\sigma}^{a+3\sigma} p_\xi(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{a-3\sigma}^{a+3\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-3}^{3} e^{-\frac{t^2}{2}}\,dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{3} e^{-\frac{t^2}{2}}\,dt = 2\Phi_0(3).$$

Из табл. 2 (см. [1], с. 236) находим $\Phi_0(3) = 0{,}4987$. Следовательно,

$$P\{|\xi - M\xi| \geq 3\sqrt{D\xi}\} = 1 - 2 \cdot 0{,}4987 = 0{,}0026.$$

б) Так как $M\xi = 0$, $D\xi = \dfrac{1}{3}$, то

$$P\{|\xi - M\xi| \geq 3\sqrt{D\xi}\} = P\{|\xi| \geq \sqrt{3}\} = 0$$

(ибо все значения случайной величины по модулю меньше единицы).

в) Для случайной величины, распределенной по закону Пуассона $M\xi = D\xi = \lambda$, где λ — параметр распределения и $\lambda = 0{,}09$. Следовательно,

$$P\{|\xi - M\xi| \geq 3\sqrt{D\xi}\} = P\{|\xi - 0{,}09| \geq 0{,}9\} =$$

$$= P\{\xi \geq 1\} = 1 - P\{\xi = 0\} = 1 - \mathrm{e}^{-0{,}09} \approx 0{,}08607.$$

Ответ

а) 0,0026; б) 0; в) 0,08607.

Задача 2

Пусть ξ_1, ξ_2, ..., ξ_n, ... — последовательность случайных величин; C_1, C_2, ..., C_n, ... — ограниченная числовая последовательность. Верно ли утверждение: если $\xi_n \xrightarrow{P} 0$, то $C_n\xi_n \xrightarrow{P} 0$ при $n \to \infty$?

Решение

Последовательность случайных величин ξ_n, $n = 1, 2, \ldots$, сходится по вероятности к 0, если при любом $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|\xi_n| < \varepsilon\} = 1,$$

следовательно, $\lim\limits_{n\to\infty} P\{|\xi_n| \geq \varepsilon\} = 0$.

Обозначим $C = \sup\limits_{1 \leq k < \infty} |C_k|$, $C > 0$. Для любого $\varepsilon > 0$ справедливы следующие соотношения между событиями

$$\{|\xi_n C_n| \geq \varepsilon\} \subset \{|\xi_n C| \geq \varepsilon\}, \quad n = 1, 2, \ldots.$$

Тогда

$$0 \le P\{|\xi_n C_n| \ge \varepsilon\} \le P\{|\xi_n| \ge \varepsilon / C\}$$

(см. [1], гл. 1, § 3, с. 26) и при $n \to \infty$ получаем

$$0 \le \lim_{n\to\infty} P\{|\xi_n C_n| \ge \varepsilon\} \le 0,$$

откуда следует, что $C_n \xi_n \xrightarrow{P} 0$ при $n \to \infty$.

Ответ

Да.

Задача 3

Случайные величины ξ_1, ξ_2, …, ξ_n, … попарно независимы, причем ξ_k принимает значения $\pm\sqrt{\ln k}$ ($k = 2, 3, \ldots, n, \ldots$) с вероятностью 1/2. Подчиняется ли последовательность ξ_2, ξ_3, …, ξ_n, … закону больших чисел?

Решение

Для того чтобы для последовательности случайных величин ξ_2, ξ_3, …, ξ_n, … выполнялся закон больших чисел (ЗБЧ), достаточно, чтобы

$$\lim_{n\to\infty} \frac{1}{n^2} D\left(\sum_{k=2}^{n+1} \xi_k\right) = 0$$

(см. [1], гл. 6, § 5, с. 120). Найдем $M\xi_k$ и $D\xi_k$:

$$M\xi_k = \sqrt{\ln k}\,\frac{1}{2} + (\sqrt{\ln k})\frac{1}{2} = 0;$$

$$D\xi_k = M(\xi_k - M\xi_k)^2 = M\xi_k^2 = (\ln k)\frac{1}{2} + (\ln k)\frac{1}{2} = \ln k,$$

$$k = 2, 3, \ldots, n+1, \ldots .$$

Так как ξ_2, ξ_3, …, ξ_{n+1} попарно независимы, то

$$\frac{1}{n^2} D\left(\sum_{k=2}^{n+1} \xi_k\right) = \frac{1}{n^2} \sum_{k=2}^{n+1} D\xi_k = \frac{1}{n^2} \sum_{k=2}^{n+2} \ln k \le \frac{1}{n^2} \sum_{k=2}^{n+1} \ln(n+1) =$$

$$= \frac{n \ln(n+1)}{n^2} = \frac{\ln(n+1)}{n}.$$

Значит, $\lim\limits_{n\to\infty} \frac{1}{n^2} D\left(\sum\limits_{k=2}^{n+1} \xi_k\right) = 0$. Следовательно, ЗБЧ выполняется, т.е. при $n \to \infty$ $\quad \frac{1}{n}\sum\limits_{k=2}^{n+1} \xi_k \xrightarrow{P} 0$.

Ответ

Да.

Задача 4

Найти характеристическую функцию дискретной случайной величины с распределением:

$$P\{\xi = -b_k\} = P\{\xi = b_k\} = p_k;$$

$$P\{\xi = 0\} = p_0, \quad k = 1, 2, \ldots, n.$$

Решение

Согласно определению (см. [1], гл. 7, § 2, равенство (2.3), с. 140) характеристическая функция случайной величины $\xi(\omega)$ равна

$$f_\xi(t) = M e^{i\xi t} = \sum_{k=-\pi}^{\pi} p_k e^{ib_k t} =$$

$$= p_0 + \sum_{k=1}^{n} (e^{ib_k t} + e^{-ib_k t}) p_k = p_0 + 2\sum_{k=1}^{n} p_k \cos b_k t.$$

Ответ

$$f_\xi(t) = p_0 + 2\sum_{k=1}^{n} p_k \cos b_k t.$$

Задача 5

Какому условию должны удовлетворять неотрицательные числа $a_0, a_1, \ldots, a_n, \ldots$ для того, чтобы функция $f(t) = \sum\limits_{n=0}^{\infty} a_n \cos nt$ была характеристической? Какой случайной величины?

Решение

Из теоремы 2.1 (см. [1], гл. 7, § 2, с. 140) следует, что $f_\xi(0) = Me^{i\xi 0} = M1 = 1$. Следовательно, $\sum_{n=0}^{\infty} a_n = 1$. Так как $\cos nt = \dfrac{e^{int} + e^{-int}}{2}$, то

$$f_\xi(t) = a_0 + \frac{1}{2}\sum_{n=1}^{\infty} a_n(e^{int} + e^{-int}),$$

$$f_\xi(t) = \sum_{k=1}^{\infty} e^{inx_1} P\{\xi = x_k\}.$$

Значит, ξ — дискретная случайная величина, принимающая целые значения 0, ±1, ±2, ... и $P\{\xi = 0\} = a_0$, $P\{\xi = n\} = P\{\xi = -n\} = \dfrac{1}{2}a_n$, $n = 1, 2, \ldots$.

Ответ

$\sum_{n=0}^{\infty} a_n = 1$, $P\{\xi = 0\} = a_0$,

$P\{\xi = n\} = P\{\xi = -n\} = \dfrac{1}{2}a_n$, $n = 1, 2, \ldots$.

Задача 6

Вероятность наступления события A в одном испытании равна 0,5. Какова вероятность того, что при проведении 1000 независимых испытаний количество наступлений события A будет заключено в интервале от 450 до 550?

Решение

Обозначим через μ_{1000} число наступлений события A в 1000 независимых испытаниях, μ_{1000} имеет биномиальное распределение с параметрами $n = 1000$, $p = 0{,}5$ и

$$P\{\mu_{1000} = m\} = C_{1000}^{m}(0{,}5)^{m}(0{,}5)^{1000-m},\ m = 0,\ 1,\ \ldots,\ 1000.$$

Нас интересует вероятность события: «μ_{1000} заключено в интервале (450, 550)». Эта вероятность равна

$$\sum_{450<m<550} P\{\mu_{1000} = m\} = \sum_{450<m<550} C_{1000}^{m}(0{,}5)^{1000}.$$

Так как $n = 1000$ «велико» и $npq = 250 > 20$, то для нахождения вышенаписанной суммы применим интегральную теорему Муавра-Лапласа (см. [1], гл. 4, § 3, с. 64, теорема 3.3), утверждающую, что нормированное число наступлений события A при больших n имеет распределение, близкое к нормальному

$$P\left\{\frac{\mu_n - np}{\sqrt{npq}} < x\right\} \underset{n\to\infty}{\to} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt.$$

Осталось найти границы, в которых заключено нормированное число наступлений события A: $\overline{\mu}_n = \dfrac{\mu_n - np}{\sqrt{npq}}$. Из неравенств $450 < \mu_{1000} < 550$ получаем

$$-\sqrt{10} < \frac{\mu_{1000} - 500}{\sqrt{250}} < \sqrt{10}.$$

Таким образом,

$$P\{450 < \mu_{1000} < 550\} = P\left\{-\sqrt{10} < \frac{\mu_{1000} - 500}{\sqrt{25}} < \sqrt{10}\right\} =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-3,16}^{3,16} e^{-\frac{t^2}{2}}\,dt = 2\Phi_0(3{,}16) = 0{,}998$$

(см. [1], табл. 2, с. 237).

Ответ

$P = 2\Phi_0(3{,}16) = 0{,}998$.

Задача 7

Вероятность наступления события A в одном испытании равна 0,4. Сколько надо провести независимых испытаний, чтобы было гарантировано неравенство:

$$P\left\{\left|\frac{n_A}{n} - p\right| < 0{,}01\right\} > 0{,}995$$

(n_A — число наступлений события A в n испытаниях).

Решение

Из интегральной теоремы Муавра-Лапласа можно получить

$$P\left\{\left|\frac{n_A}{n} - p\right| < \Delta\right\} = 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right)$$

(см. [1], гл. 4, § 3, формула (3.15), с. 68), $p = 0{,}4$, $q = 0{,}6$, $\Delta = 0{,}1$. Следовательно, число испытаний n удовлетворяет следующему неравенству:

$$2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right) = 2\Phi_0\left(0{,}01\sqrt{\frac{n}{0{,}24}}\right) > 0{,}995 .$$

Из табл. 2 (см. [1], с. 237) по значению функции $\Phi_0(x) = 0{,}4975$ находим значение аргумента $x = 2{,}81$. Тогда из равенства $2{,}81 = 0{,}01\sqrt{\frac{n}{0{,}24}}$ получаем $n = \frac{(2{,}81)^2}{(0{,}01)^2}0{,}24 \approx 1{,}9 \cdot 10^4$.

Ответ

$n \geq 1{,}9 \cdot 10^4$.

Задача 8

Книга в 500 с. содержит 50 опечаток. Оценить вероятность того, что на случайно выбранной странице не менее двух опечаток.

Решение

Обозначим $\nu_n(\omega)$ — число опечаток на случайно выбранной странице, $\nu_n(\omega)$ распределена по биномиальному закону с параметрами $n = 50$, $p = 0{,}002$ и

$$P\{\nu_{50} = m\} = C_{50}^m (0{,}002)^m (0{,}998)^{50-m}, \ m = \overline{0, 50}.$$

Так как $p = 0{,}002$ — «малое» число и $np = 0{,}1 < 10$, то применим теорему Пуассона с $\lambda = 0{,}1$ (см. [1], гл. 4, § 3, теорема 3.1, с. 61):

$$P\{\nu_{50} = m\} \approx \mathrm{e}^{-0{,}1} \frac{(0{,}1)^m}{m!}, m = 0, 1, \ldots .$$

Таким образом, искомая вероятность равна

$$P\{\nu_{50} \geq 2\} = \sum_{m=2}^{\infty} P\{\nu_{50} = m\} = \sum_{m=2}^{\infty} \mathrm{e}^{-0{,}1} \frac{(0{,}1)^m}{m!}.$$

Из табл. П1 на с. 96 – 97 данных методических указаний находим при $x = 2$

$$P\{\nu_{50} \geq 2\} \approx 0{,}005 .$$

Ответ

0,005.

Задача 9

Случайные величины ξ_k при $k = 1, 2, \ldots, 4500$ взаимно независимы и имеют одинаковые дисперсии $D\xi_k = 5$. Вычислить (приближенно) вероятность того, что среднее арифметическое этих случайных величин отклонится от своего математического ожидания не более, чем на 0,04.

Решение

Рассмотрим случайную величину $\eta = \frac{1}{4500}\sum_{k=1}^{4500}(\xi_k - M\xi_k)$, $M\eta = \frac{1}{4500}\sum_{k=1}^{4500} M(\xi_k - M\xi_k) = 0$, и в силу независимости случайных величин $\xi_1, \xi_2, \ldots, \xi_{4500}$

$$D\eta = \frac{1}{(4500)^2}\sum_{k=1}^{4500} D\xi_k = \frac{4500}{(4500)^2}\cdot 5 = \frac{1}{900}.$$

Введенная случайная величина η и есть отклонение среднего арифметического случайных величин $\xi_1, \xi_2, \ldots, \xi_{4500}$ от среднего арифметического их математических ожиданий, причем $M\eta = 0$ и $D\eta = \frac{1}{900}$, следовательно, можем для нахождения вероятности

$$P\left\{\left|\frac{1}{4500}\sum_{k=1}^{4500}\xi_k - \frac{1}{4500}\sum_{k=1}^{4500} M\xi_k\right| < 0{,}04\right\}$$

применить неравенство Чебышева: для любого $\varepsilon > 0$

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

(см. [1], гл. 6, § 5, теорема 5.2, с. 117). При $\varepsilon = 0{,}04$ получаем

$$P\{|\eta - M\eta| < 0{,}04\} = 1 - P\{|\eta - M\eta| \geq 0{,}04\} \geq$$

$$\geq 1 - \frac{1}{1{,}44} \approx 0{,}305.$$

Ответ

$$P\left\{\left|\frac{1}{4500}\sum_{k=1}^{4500}(\xi_k - M\xi_k)\right| < 0{,}04\right\} \geq 0{,}305.$$

Задача 10

Случайные величины ξ_1, ξ_2, …, ξ_k, … независимы, причем $\xi_k(\omega)$ принимает два значения

$$P\{\xi_k = \sqrt{k}\} = P\{\xi_k = -\sqrt{k}\} = \frac{1}{2}, \quad k = 1, 2, \ldots \ .$$

Методом характеристических функций найти предельное распределение средних арифметических $\frac{1}{n}(\xi_1 + \xi_2 + \ldots + \xi_n)$ при $n \to \infty$. Доказать, что последовательность ξ_1, ξ_2, …, ξ_n, … не подчиняется закону больших чисел.

Решение

Обозначим $\eta_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$. Найдем $f_{\eta_n}(t)$ — характеристическую функцию случайной величины η_n.

Согласно равенству 2.1 (см. [1], гл. 7, § 2, с. 139)

$$f_{\xi_k}(t) = Me^{it\xi_k} = e^{it\sqrt{k}}\,\frac{1}{2} + e^{-it\sqrt{k}}\,\frac{1}{2} = \cos\sqrt{k}t\,.$$

Так как случайные величины ξ_1, ξ_2, …, ξ_n независимы, то теорема 2.3 (см. [1], гл. 7, § 2, с. 145) в этом случае дает следующий результат:

$$f_{\sum\limits_{k=1}^{n}\xi_k}(t) = \prod_{k=1}^{n} f_{\xi_k}(t) = \prod_{k=1}^{n}\cos\sqrt{k}t\,,$$

переходя к случайной величине η_n и принимая во внимание свойство 3 из теоремы 2.1 (см. [1], гл. 7, § 2, с. 140), получаем

$$f_{\eta_n}(t) = \prod_{k=1}^{n} f_{\frac{\xi_k}{n}}(t) = \prod_{k=1}^{n} f_{\xi_k}\left(\frac{t}{n}\right) = \prod_{k=1}^{n}\cos\frac{\sqrt{k}}{n}t\,.$$

Рассмотрим $\ln f_{\eta_n}(t)$:

$$\ln f_{\eta_n}(t) = \sum_{k=1}^{n}\ln\cos\frac{\sqrt{k}}{n}t\,.$$

Так как при $n \to \infty$ $\sum_{k=1}^{n} \frac{k}{n^2} = \frac{n+1}{2n} - \frac{1}{2}$, а

$$\sum_{k=1}^{n} \frac{k^2}{n^4} = \frac{1}{n^4} \frac{(2n+1)(n+1)n}{6} = \frac{(n+1)(2n+1)}{6n^3} \underset{n\to\infty}{\to} 0,$$

то получаем, используя разложение функций $\cos x$ и $\ln(1+x)$ в ряд Тейлора по степеням x:

$$\lim_{n\to\infty} \ln f_{\eta_n}(t) = -\frac{t^2}{2}\frac{1}{2}.$$

Откуда

$$\lim_{n\to\infty} f_{\eta_n}(t) = \mathrm{e}^{-\frac{t^2}{2}\frac{1}{2}} = f(t).$$

В примере 4 (см. [1], гл. 7, § 2, с. 142) показано, что для случайной величины ν, распределенной по нормальному закону с параметрами a и σ^2, характеристическая функция равна

$$f_\nu(t) = \mathrm{e}^{ita - \frac{t^2}{2}\sigma^2}.$$

Учитывая этот результат, можем сказать, что последовательность характеристических функций $f_{\eta_n}(t)$, $n = 1, 2, \ldots$, случайных величин $\eta_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$ сходится для любого t к непрерывной при $t = 0$ функции $f(t) = \mathrm{e}^{-\frac{t^2}{2}\frac{1}{2}}$, являющейся характеристической функцией случайной величины θ, имеющей нормальное распределение с $M\theta = 0$ и $M\theta = 1/2$. Тогда согласно теореме 2.5 (см. [1], гл. 7, § 2, с. 147) последовательность функций распределения $F_{\eta_n}(x)$, $n = 1, 2, \ldots$, слабо сходится к функции распределения

$$F_\theta(x) = \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2\cdot\frac{1}{2}}\, dt.$$

Напомним определение (см. [1], гл. 7, § 2, с. 146): последовательность $F_{\eta_n}(x)$, $n = 1, 2, \ldots$, слабо сходится к функции распределения $F_\theta(x)$, если

$$\lim_{n\to\infty} F_{\eta_n}(x) = F_\theta(x)$$

при любом x, являющемся точкой непрерывности $F_\theta(x)$. Следовательно, для любого x ($F_\theta(x)$ — непрерывная на $(-\infty, +\infty)$ функция)

$$\lim_{n\to\infty} F_{\eta_n}(x) = \frac{1}{\sqrt{2\pi}\sqrt{1/2}} \int_{-\infty}^{x} e^{-t^2}\, dt$$

и

$$p_\theta(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}, \quad -\infty < x < +\infty .$$

Рассмотрим теперь вопрос, выполняется ли закон больших чисел для последовательности $\xi_1, \xi_2, \ldots, \xi_n, \ldots$. Так как $M\xi_k = 0$, $D\xi_k = M\xi_k^2 = k$, $k = 1, 2, \ldots$, и

$$\frac{1}{n^2} D\left(\sum_{k=1}^{n} \xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} k = \frac{n+1}{2n} \underset{n\to\infty}{\to} \frac{1}{2} \neq 0,$$

то достаточное условие не выполняется. Но найдено предельное распределение для η_n. Тогда для $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\{|\eta_n| < \varepsilon\} = \lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k\right| < \varepsilon\right\} = \frac{1}{\sqrt{\pi}} \int_{-z}^{z} e^{-t^2}\, dt \neq 1 .$$

Следовательно, закон больших чисел не выполняется.

Ответ

Нормальное распределение с параметрами 0 и 1/2.

Задача 11

Найти $P\{|\xi - M\xi| \geq 3\sqrt{D\xi}\}$, если $\xi(\omega)$ имеет:

а) $P\{\xi = 1\} = P\{\xi = -1\} = \dfrac{1}{18}$, $P\{\xi = 0\} = \dfrac{8}{9}$;

б) показательное распределение.

Решение

а) Найдем $M\xi$ и $D\xi$:

$$M\xi = 1\cdot\frac{1}{18} + 0\cdot\frac{8}{9} + (-1)\frac{1}{18} = 0;$$

$$D\xi = M\xi^2 = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}.$$

Значит,

$$P\{|\xi - M\xi| \ge 3\sqrt{D\xi}\} = P\{|\xi| \ge 1\} =$$

$$= P\{|\xi| = 1\} = P\{\xi = 1\} + P\{\xi = -1\} = \frac{1}{9} \approx 0{,}111.$$

б) Плотность распределения случайной величины ξ равна

$$p_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{если } x > 0; \\ 0, & \text{если } x \le 0. \end{cases}$$

Найдем $M\xi$ и $D\xi$:

$$M\xi = \int_0^\infty \lambda x e^{-\lambda x}\, dx = 1/\lambda;$$

$$M\xi^2 = \int_0^\infty \lambda x^2 e^{-\lambda x}\, dx = -x^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty x e^{-\lambda x}\, dx = \frac{2}{\lambda^2};$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{1}{\lambda^2}.$$

Тогда

$$P\{|\xi - M\xi| \ge 3\sqrt{D\xi}\} = P\left\{\left|\xi - \frac{1}{\lambda}\right| \ge \frac{3}{\lambda}\right\} =$$

$$= P\left\{\xi \ge \frac{4}{\lambda}\right\} = \int_{4/\lambda}^\infty \lambda e^{-\lambda x}\, dx = e^{-4} \approx 0{,}01832$$

(см. [1], табл. 1, с. 235).

Сравним эти результаты с оценкой вероятности, получаемой по неравенству Чебышева (см. [1], гл. 6, § 5, теорема 5.2, с. 117) с $\varepsilon = 3\sqrt{D\xi}$:

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

Видим, что оценка в первом случае оказалась достижимой.

Ответ

а) $1/9 = 0{,}111$; б) $0{,}01832$.

Задача 12

Пусть случайная величина $\xi(\omega)$ имеет нормальное распределение с параметрами (a, σ^2). Оценить по неравенству Чебышева $P\{|\xi - a| \geq 2\sigma\}$. Сравнить с точным значением этой вероятности.

Решение

Согласно неравенству Чебышева с $\varepsilon = 2\sigma$, получаем

$$P\{|\xi - M\xi| \geq 2\sigma\} \leq \frac{\sigma^2}{4\sigma^2} = 0{,}25.$$

Найдем точное значение вероятности события $(|\xi - M\xi| \geq 2\sigma\}$, используя информацию о распределении ξ:

$$P\{|\xi - M\xi| \geq 2\sqrt{D\xi}\} = 1 - P\{|\xi - a| < 2\sigma\} =$$

$$= 1 - \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{|x-a|<2a} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx = 1 - \frac{1}{\sqrt{2\pi}} \int\limits_{-2}^{2} e^{-\frac{t^2}{2}}\, dt =$$

$$= 1 - 2\Phi_0(2) = 1 - 2 \cdot 0{,}4772 = 0{,}0456$$

(см. [1], табл. 2, с. 237).

Ответ

1) 0,25; 2) 0,0456.

Задача 13

Пусть ξ_1, ξ_2, …, ξ_n — центрированные случайные величины с дисперсиями $D\xi_k = \sigma_k^2$. Коэффициенты корреляции заданы $\rho(\xi_k, \xi_j) = a_{kj}$ при $k \neq j$. Рассмотрим сумму $\eta_n = \xi_1 + \xi_2 + \ldots + \xi_n$. Найти $\rho(\xi_k, \eta_n)$. Рассмотреть частный случай, когда все случайные величины некоррелированы и одинаково распределены.

Решение

$M\xi_k = 0$, так как ξ_k — центрированные случайные величины, $k = 1, 2, \ldots, n$. Так как $\rho(\xi_k, \xi_j) = \dfrac{\operatorname{cov}(\xi_k, \xi_j)}{\sqrt{D\xi_k}\sqrt{D\xi_j}}$, то

$$\operatorname{cov}(\xi_k, \xi_j) = a_{kj}\sigma_k\sigma_j$$

и

$$D\eta_n = D\left(\sum_{l=1}^{n}\xi_l\right) = \sum_{l=1}^{n} D\xi_l + 2\sum\sum_{1\le l<j\le n} \operatorname{cov}(\xi_l, \xi_j)$$

(см. [1], гл. 6, § 4, теорема 4.1,с. 114)

$$D\eta_n = \sum_{l=1}^{n}\sigma_l^2 + 2\sum\sum_{1\le l<j\le n} a_{lj}\sigma_l\sigma_j$$

и

$$\operatorname{cov}(\xi_k, \eta_n) = \operatorname{cov}\left(\xi_k, \sum_{j=1}^{n}\xi_j\right) = M\left(\xi_k\sum_{j=1}^{n}\xi_j\right) =$$

$$= M\left(\sum_{j=1}^{n}\xi_k\xi_j\right) = \sigma_k^2 + \sum_{j=1} a_{kj}\sigma_k\sigma_j .$$

Обозначим $a_{kk} = 1$, тогда

$$\rho(\xi_k, \eta_n) = \sum_{j=1}^{n} a_{kj}\sigma_k\sigma_j\left[\sigma_k\sqrt{\sum_{l=1}^{n}\sum_{j=1}^{n} a_{lj}\sigma_l\sigma_j}\right]^{-1} =$$

$$= \sum_{j=1}^{n} a_{kj}\sigma_j \left[\sum_{l=1}^{n}\sum_{j=1}^{n} a_{lj}\sigma_l\sigma_j \right]^{-1/2}.$$

Если ξ_1, $\xi_2, \ldots, \xi_n$ некорректированы и одинаково распределены, то в последнем соотношении при $l \neq j$ $a_{lj} = 0$ и $a_{lj} = 1$, а $\sigma_l = \sigma$ для всех l = 1, 2, …, n. Тогда $\rho(\xi_k, \eta_n) = 1/\sqrt{n}$.

Ответ

1) $\rho(\xi_k, \eta_n) = \sum_{j=1}^{n} a_{kj}\sigma_j \left[\sum_{l=1}^{n}\sum_{j=1}^{n} a_{lj}\sigma_l\sigma_j \right]^{-1/2}$, где $a_{kk} = 1$;

2) $\rho(\xi_k, \eta_n) = 1/\sqrt{n}$.

Задача 14

Последовательность случайных величин ξ_1, $\xi_2, \ldots$, $\xi_k, \ldots$ удовлетворяет трем условиям: все $M\xi_k = a$ и $D\xi_k = C$, все $\rho(\xi_k, \xi_j) \le 0$ при $k \neq j$. Доказать, что эта последовательность удовлетворяет закону больших чисел.

Решение

Рассмотрим $D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2} D\left(\sum_{k=1}^{n}\xi_k\right)$. Так как при $k \neq j$

$$\operatorname{cov}(\xi_k, \xi_j) = \rho(\xi_k, \xi_j)\sqrt{D\xi_k}\sqrt{D\xi_j} = C\rho(\xi_k, \xi_j),$$

то применяя равенство (4.4) (см. [1], гл. 6, § 4, с. 114), получаем

$$\frac{1}{n^2} D\left(\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\left[\sum_{k=1}^{n} D\xi_k + 2\sum_{1\le k<j\le n}\sum C\rho(\xi_k, \xi_j)\right] \le \frac{Cn}{n^2} = \frac{C}{n} \underset{n\to\infty}{\to} 0,$$

следовательно, закон больших чисел выполняется для данной последовательности.

Задача 15

В практически неограниченной совокупности некоторых предметов половина из них обладает свойством A и пятая часть —

свойством B. Свойства A и B распределены между собой независимо. Из совокупности случайно выбрали 1000 предметов. Какова вероятность, что в этой выборке частоты, с которыми встречаются свойства A и B, отличаются от их вероятностей не более, чем на 5 %?

Решение

Пусть $\mu_{1000}(A)$ и $\mu_{1000}(B)$ — число предметов в выборке, обладающих соответственно свойствами A и B. Случайная величина $\mu_{1000}(A)$ имеет биномиальное распределение с параметрами $n = 1000$ и $p_1 = 0{,}5$, $q_1 = 0{,}5$; $\mu_{1000}(B)$ также распределена по биномиальному закону с параметрами $n = 1000$, $p_2 = 0{,}2$ и $q_2 = 0{,}8$. Обозначим события:

$$C_1 = \left\{ \left| \frac{\mu_{1000}(A)}{1000} - p_1 \right| < 0{,}025 \right\};$$

$$C_2 = \left\{ \left| \frac{\mu_{1000}(B)}{1000} - p_2 \right| < 0{,}01 \right\}.$$

Перейдем к вычислению $P(C_k)$, $k = 1, 2$. Так как $n = 1000$ — большое число, $np_1q_1 = 250 > 20$ и $np_2q_2 = 160 > 20$, то применим интегральную теорему Муавра-Лапласа, позволяющую оценить близость частоты и вероятности

$$P\left\{ \left| \frac{\mu_n}{n} - p \right| < \Delta \right\} = 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}} \right)$$

(см. [1], гл. 4, § 3, равенство (3.15), с. 68). Применяя это равенство дважды, получаем

$$P(C_1) = 2\Phi_0\left(0{,}025\,\frac{10\sqrt{10}}{0{,}5} \right) = 2\Phi_0(1{,}58) = 0{,}8858;$$

$$P(C_2) = 2\Phi_0\left(0{,}01\,\frac{10\sqrt{10}}{0{,}4} \right) = 2\Phi_0(0{,}79) = 0{,}5704$$

(см. [1], табл. 2, с. 236).

Ответ

Вероятность равна 0,8858 для свойства A и 0,5704 для свойства B.

Задача 16

В таблице случайных чисел цифры сгруппированы по две. Найти приближенное значение вероятности того, что среди 200 пар пара 01 встретится не менее трех раз.

Решение

Обозначим μ_{200} — число появлений пары 01 среди 200 пар, взятых из таблицы. Вероятность появления пары 01 равна 0,01 — малое число, следовательно, для приближенного нахождения вероятности

$$P\{\mu_{200} \geq 3\} = \sum_{k=3}^{200} C_{200}^{k} (0{,}01)^k (0{,}99)^{200-k} =$$

$$= 1 - \sum_{k=0}^{2} C_{200}^{k} (0{,}01)^k (0{,}99)^{200-k}$$

воспользуемся теоремой Пуассона (см. [1], гл. 4, § 3, теорема 3.1, с. 61) с $\lambda = 200 \cdot 0{,}01 = 2$. Тогда

$$P\{\mu_{200} \geq 3\} = 1 - \sum_{k=0}^{2} \mathrm{e}^{-2} \frac{2^k}{k!} = 1 - \mathrm{e}^{-2k} =$$

$$= 1 - 0{,}13534 \cdot 5 = 1 - 0{,}6767 = 0{,}3233$$

(из табл. 1 (см. [1], с. 235) нашли $\mathrm{e}^{-2} = 0{,}13534$).

Ответ

0,3233.

Задача 17

Найти приближенное значение вероятности того, что число выпадений «1» при 12000 бросаний игральной кости заключено между 1900 и 2150.

Решение

Пусть случайная величина μ_{12000} — число появлений «1» при 12000 бросаний игральной кости. Величина μ_{12000} имеет биномиальное распределение с параметрами $n = 12000$ и $p = 1/6$. Так как n — большое число, $npq = \frac{500}{3} > 20$, то для нахождения вероятности попадания μ_{12000} на интервал (1900, 2150) применим интегральную теорему Муавра-Лапласа (см. [1], гл. 4, § 3, теорема 3.3, с. 64)

$$P\{1900 < \mu_{12000} < 2150\} = P\left\{a < \frac{\mu_{12000} - 2000}{100/\sqrt{6}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}}\, dt\,,$$

где a и b соответственно равны

$$a = \frac{1900 - 2000}{100/\sqrt{6}} = -\sqrt{6} = -2{,}45\,; \quad b = \frac{2150 - 2000}{100/\sqrt{6}} = \frac{3}{2}\sqrt{6} = 3{,}67\,.$$

Итак,

$$P\{1900 < \mu_{12000} < 2150\} = \frac{1}{\sqrt{2\pi}} \int_{-2{,}45}^{3{,}67} e^{-\frac{t^2}{2}}\, dt = \Phi_0(2{,}45) + \Phi_0(3{,}67)\,.$$

$\Phi_0(2{,}45)$ находим из табл. 2 (см. [1], с. 236), а $\Phi_0(3{,}67)$ из табл. П3 на с. 99 данных методических указаний.

$$P\{1900 < \mu_{1200} < 2150\} = 0{,}4929 + 0{,}4999 = 0{,}9928\,.$$

Ответ

0,9928.

Задача 18

Театр, вмещающий 1000 человек, имеет два разных входа. Около каждого из входов имеется свой гардероб. Сколько мест должно быть в каждом из гардеробов для того, чтобы в среднем в 99 случаях из 100 все зрители могли раздеться в гардеробе того входа, через который они вошли? Рассмотреть два случая: а) зрители приходят парами; б) зрители приходят поодиночке. Предположить, что входы зрители выбирают с равными вероятностями.

Решение

Обозначим через μ_n число зрителей (или пар в первом случае), решивших идти налево. Тогда $n-\mu_n$ — число зрителей (или пар), идущих направо (в другой вход). Пусть x_0 — число мест в каждом из гардеробов. Рассмотрим оба случая отдельно.

а) Зрители приходят парами, тогда μ_{500} имеет биномиальное распределение с параметрами $n = 500$, $p = q = 1/2$. Пусть событие означает, что все зрители смогли раздеться в гардеробе того входа, через который они вошли. Тогда

$$P(A) = P\left\{\mu_{500} \le \frac{x_0}{2},\, 500-\mu_{500} \le \frac{x_0}{2}\right\} = P\left\{500-\frac{x_0}{2} \le \mu_{500} \le \frac{x_0}{2}\right\}.$$

Так как $n = 500$ — большое число и $npq = 500\cdot\frac{1}{4} = 125 > 20$, то для вычисления $P(A)$ используем интегральную теорему Муавра-Лапласа (см. [1], гл. 4, § 3, теорема 3.3, с. 64). Итак,

$$P(A) = P\left\{\frac{250-x_0/2}{\sqrt{125}} \le \frac{\mu_{500}-250}{\sqrt{125}} \le \frac{\frac{x_0}{2}-250}{\sqrt{125}}\right\} =$$

$$= 2\Phi_0\left(\frac{x_0/2-250}{\sqrt{125}}\right) = 0{,}99.$$

Из табл. 2 (см. [1], с. 236) находим значение y, при котором

$$\Phi_0(y) = \frac{1}{\sqrt{2\pi}}\int_0^y e^{-t^2/2}\,dt = 0{,}495, \quad \Phi_0(2{,}575) = 0{,}495,$$

откуда $\frac{x_0}{2} - 250 = \sqrt{125}\cdot 2{,}575$; $x_0 = 500 + 58 = 558$.

Итак, в случае, когда зрители приходят парами, $x_0 = 558$.

б) Случайная величина μ_{1000} имеет биномиальное распределение с параметрами $n = 1000$, $p = q = 1/2$. В этом случае

$$P(A) = P\{\mu_{1000} \le x_0, 1000 - \mu_{1000} \le x_0\} = P\{1000 - x_0 \le \mu_{1000} \le x_0\}.$$

Дальнейшие рассуждения аналогичны приведенным раньше в случае а):

$$P\left\{\frac{500 - x_0}{\sqrt{250}} \le \frac{\mu_{1000} - 500}{\sqrt{250}} \le \frac{x_0 - 500}{\sqrt{250}}\right\} = 2\Phi_0\left(\frac{x_0 - 500}{\sqrt{250}}\right) = 0{,}99,$$

$$x_0 = 500 + 2{,}575\sqrt{250} = 500 + 41 = 541$$

(x_0 — число мест в гардеробе — целое число!).

Ответ

а) 558; б) 541.

Задача 19

$\xi_1, \xi_2, \ldots, \xi_k, \ldots$ — последовательность взаимно независимых случайных величин.

$$P\{\xi_k = 2^k\} = P\{\xi_k = -2^k\} = 2^{-2k-1},\ P\{\xi_k = 0\} = 1 - 2^{-2k},\ k = 1, 2, \ldots .$$

Выполняется ли для данной последовательности закон больших чисел?

Решение

Найдем средние значения и дисперсии случайных величин $M\xi_k$, $D\xi_k$:

$$M\xi_k = 2^k \cdot 2^{-2k-1} + (-2^k)2^{-2k-1} = 0,$$

$$D\xi_k = M\xi_k^2 = 2^{2k}2^{-2k-1} = 1.$$

В этой задаче $\frac{1}{n^2} D\left(\sum\limits_{k=1}^{n} \xi_k\right) = \frac{1}{n^2}\sum\limits_{k=1}^{n} D\xi_k = \frac{n}{n^2} = \frac{1}{n} \to 0$ при $n \to \infty$.

Следовательно, для данной последовательности выполняется закон больших чисел (можно было сослаться на теорему 5.4 — теорему Чебышева из [1], с. 121).

Ответ

Да.

Задача 20

Случайные величины ξ_1, ξ_2, ..., ξ_n, ... независимы и имеют стандартное нормальное распределение,

$$\chi_n^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2.$$

А. Доказать, что $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{\chi_n^2}{n} - 1\right| \geq \varepsilon\right\} = 0.$$

Б. Найти

$$\lim_{n\to\infty} P\left\{\frac{\chi_n^2 - M\chi_n^2}{\sqrt{D\chi_n^2}} < 3\right\}.$$

Решение

Известно, что для случайной величины ξ, имеющей стандартное нормальное распределение $M\xi^2 = 1$, в задаче 4.4 (см. [5], с. 37) было найдено $M\xi^4 = 3!! = 3$. Тогда

$$D\xi^2 = 3 - 1 = 2, \; M\chi_n^2 = M\left(\sum_{k=1}^{n} \xi_k^2\right) = n.$$

А. Так как ξ_1, ξ_2, ..., ξ_n, ... независимы, то и ξ_1^2, ξ_2^2, ..., ξ_n^2, ... тоже независимы (функции от независимых случайных величин); χ_n^2 есть сумма независимых одинаково распределенных случайных величин с дисперсией, равной двум, тогда согласно теореме 5.5 (см. [1], гл. 6, § 5, с. 121) для этой последовательности выполняется закон больших чисел, т.е. для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{\chi_n^2}{n} - 1\right| < \varepsilon\right\} = 1.$$

Следовательно,

$$\lim_{n\to\infty} P\left\{\left|\frac{\chi_n^2}{n}-1\right|\geq\varepsilon\right\}=1-\lim_{n\to\infty} P\left\{\left|\frac{\chi_n^2}{n}-1\right|<\varepsilon\right\}=0\,.$$

Можно было сослаться и на теорему 3.1 из [1], гл. 7, § 3, с. 150, где закон больших чисел доказывается при более слабом требовании на последовательность независимых одинаково распределенных случайных величин, а именно — достаточно в этом случае требования существования математического ожидания случайных величин.

Б. Для последовательности независимых одинаково распределенных случайных величин, имеющих конечную дисперсию ($\sigma^2>0$), справедлива центральная предельная теорема (см. [1], гл. 7, § 4, теорема 4.1, с. 152). Последовательность ξ_1^2, ξ_2^2, ..., ξ_n^2, ... удовлетворяет условиям теоремы 4.1: $D\xi_k^2=2>0$, $k=1$, 2, ..., n, ..., следовательно, нормированная сумма квадратов случайных величин, имеющих стандартное нормальное распределение, при большом n имеет распределение, близкое к нормальному распределению с $a=0$, $\sigma=1$. Итак,

$$\lim_{n\to\infty} P\left\{\frac{\chi_n^2-M\chi_n^2}{\sqrt{D\chi_n^2}}<3\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{3}e^{-t^2/2}\,dt=0{,}5+\Phi_0(3)=0{,}9987$$

(см. [1], табл. 2, с. 236).

Ответ

а) 0; б) 0,9987.

Задача 21

Стрелок попадает при каждом выстреле по мишени в десятку с вероятностью 0,45; в девятку с вероятностью 0,35; в восьмерку — 0,1; в семерку — 0,05; в шестерку — 0,05. Стрелок сделал 100 выстрелов. Найти вероятность того, что сумма полученных очков заключена в интервале [880, 940].

Решение

Обозначим через ξ_k число очков, полученных при k-м выстреле, $k = 1, 2, \dots, 100$. Найдем $M\xi_k$ и $D\xi_k$:

$$M\xi_k = \sum_{l=6}^{10} lP\{\xi_k = l\} = 10 \cdot 0{,}45 + 9 \cdot 0{,}35 + 8 \cdot 0{,}1 + 7 \cdot 0{,}05 + 6 \cdot 0{,}05 = 9{,}1,$$

$$M\xi_k^2 = \sum_{l=6}^{10} l^2 P\{\xi_k = l\} = 84,$$

$$D\xi_k = M\xi_k^2 - (M\xi_k)^2 = 1{,}19, \quad \sigma = \sqrt{D\xi_k} = 1{,}0908.$$

Так как ξ_1, $\xi_2, \dots,$ ξ_{100} независимы, одинаково распределены и имеют конечную дисперсию $\sigma^2 > 0$, то нормированная сумма очков, полученных при большом числе выстрелов, имеет распределение, близкое к стандартному нормальному распределению (см. [1], гл. 7, § 4, теорема 4.1, с. 152). Следовательно,

$$P\left\{880 < \sum_{k=1}^{100} \xi_k < 940\right\} = P\left\{\frac{880 - 910}{10{,}908} < \frac{\sum_{k=1}^{100} \xi_k - 910}{10{,}908} < \frac{940 - 910}{10{,}908}\right\} =$$

$$= P\left\{-2{,}75 < \frac{\sum_{k=1}^{100} \xi_k - 910}{10{,}908} < 2{,}75\right\} = \frac{1}{\sqrt{2\pi}} \int_{-2,75}^{2,75} e^{-\frac{t^2}{2}} dt =$$

$$= 2\Phi_0(2{,}75) = 2 \cdot 0{,}4970 = 0{,}9940$$

(см. табл. 2 в [1], с. 237).

Ответ

0,9940.

Задача 22

Случайные величины ξ_1, ξ_2, ..., ξ_{300} независимы и равномерно распределены на отрезке [1, 7]. Найти $P\left\{1140 < \sum_{k=1}^{300}\xi_k < 1290\right\}$.

Решение

В [1] на с. 106 и 112 найдены соответственно среднее значение и дисперсия для случайной величины ξ, равномерно распределенной на отрезке $[a, b]$:

$$M\xi = \frac{a+b}{2}, \quad D\xi = \frac{(b-a)^2}{12}.$$

Тогда для нашей задачи $M\xi_k = 4$ и $D\xi_k = 3$, $n = 300$. Согласно центральной предельной теореме (ЦПТ) (см. [1], теорема 4.1, с. 152) нормированная сумма независимых одинаково распределенных случайных величин, имеющих конечную дисперсию ($\sigma^2 > 0$), при большом числе слагаемых имеет распределение, близкое к нормальному распределению с $a = 0$, $\sigma = 1$. Следовательно,

$$P\left\{1140 < \sum_{k=1}^{300}\xi_k < 1290\right\} = P\left\{-2 < \frac{\sum_{k=1}^{300}\xi_k - 1200}{30} < 3\right\} = \frac{1}{\sqrt{2\pi}}\int_{-2}^{3} e^{-\frac{t^2}{2}}\,dt =$$

$$= \Phi_0(2) + \Phi_0(3) = 0{,}4772 + 0{,}4987 = 0{,}9759$$

(см. [1], табл. 2, с. 236).

Ответ

0,9759.

Задача 23

Случайные величины ξ_1, ξ_2, ..., ξ_{100} независимы и распределены по закону Пуассона с параметром $\lambda = 0{,}25$. Найти

$$P\left\{11 < \sum_{k=1}^{100}\xi_k < 39\right\}.$$

Решение

В [1] на с. 106 и 112 были найдены математическое ожидание и дисперсия случайной величины, имеющей распределение Пуассона с параметром λ: $M\xi_k = D\xi_k = \lambda = \sigma^2$. Так как ξ_1, ξ_2, ..., ξ_{100} независимы, одинаково распределены и имеют конечную дисперсию $\sigma^2 > 0$, то распределение их нормированной суммы $\frac{1}{10\sqrt{\lambda}}\left(\sum_{k=1}^{100}\xi_k - 100\lambda\right)$ приближенно можно считать стандартным нормальным распределением и, значит,

$$P\left\{11 < \sum_{k=1}^{100}\xi_k < 39\right\} = P\left\{\frac{11-25}{5} < \frac{\sum_{k=1}^{100}\xi_k - 25}{5} < \frac{39-25}{5}\right\} =$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-2,8}^{2,8} e^{-\frac{t^2}{2}}\,dt = 2\Phi_0(2,8) = 2\cdot 0,4974 = 0,9948$$

(см. [1], табл. 2, с. 236).

Ответ

0,9948.

Задача 24

Величина $\xi_k(\omega)$ распределена по закону Пуассона с параметром λ. Найти

$$\lim_{\lambda\to\infty} P\left\{\frac{\xi_k - \lambda}{\sqrt{\lambda}} < x\right\}.$$

Решение

Используя метод характеристических функций, найдем предельное распределение случайной величины $\eta_\lambda = \frac{\xi_\lambda - \lambda}{\sqrt{\lambda}}$. Так как

$$P\{\xi_\lambda = m\} = e^{-\lambda}\frac{\lambda^m}{m!},\quad m = 0, 1, 2, \ldots,$$

то

$$f_{\xi_\lambda}(t)=\sum_{m=0}^{\infty}\mathrm{e}^{itm}\mathrm{e}^{-\lambda}\frac{\lambda^m}{m!}$$

(см. [1], равенство (2.3), с. 140). Далее

$$f_{\xi_\lambda}(t)=\mathrm{e}^{-\lambda}\sum_{m=0}^{\infty}(\lambda\mathrm{e}^{it})^{ot}/m!=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda\mathrm{e}^{it}}=\mathrm{e}^{\lambda(\mathrm{e}^{it}-1)}.$$

Случайная величина η_1 есть линейная функция от случайной величины ξ_λ: $\eta_\lambda=\frac{1}{\sqrt{\lambda}}\xi_\lambda-\sqrt{\lambda}$, тогда из свойства 3 теоремы 2.1 (см. [1], с. 140) следует

$$f_{\eta_\lambda}(t)=\mathrm{e}^{-it\sqrt{\lambda}}f_{\xi_\lambda}\left(\frac{t}{\sqrt{\lambda}}\right)=\mathrm{e}^{-it\sqrt{\lambda}}\mathrm{e}^{\lambda\left(\mathrm{e}^{i\frac{t}{\sqrt{\lambda}}}-1\right)}.$$

Разложим экспоненту $\mathrm{e}^{i\frac{t}{\sqrt{\lambda}}}-1$ в показателе степени второго множителя в ряд Тейлора по степеням $i\frac{t}{\sqrt{\lambda}}$ с точностью до членов второго порядка малости:

$$e^{i\frac{t}{\sqrt{\lambda}}}-1=i\frac{t}{\sqrt{\lambda}}-\frac{t^2}{2\lambda}+o\left(\frac{t^2}{\lambda}\right).$$

Подставляя это разложение в выражение для $f_{\eta_\lambda}(t)$, получаем

$$f_{\eta_\lambda}(t)=\mathrm{e}^{-it\sqrt{\lambda}}\mathrm{e}^{\lambda\left(i\frac{t}{\sqrt{\lambda}}-\frac{t^2}{2\lambda}+o\left(\frac{t^2}{\lambda}\right)\right)}=\mathrm{e}^{-\frac{t^2}{2}}\mathrm{e}^{\lambda o\left(\frac{t^2}{\lambda}\right)}\underset{\lambda\to\infty}{\to}\mathrm{e}^{-\frac{t^2}{2}},$$

что соответствует характеристической функции стандартного нормального распределения. Отсюда вытекает, что при больших значениях параметра λ пуассоновское распределение можно приближенно аппроксимировать нормальным распределением. Следовательно,

$$\lim_{\lambda\to\infty}P\left\{\frac{\xi_\lambda-\lambda}{\sqrt{\lambda}}<x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}dt=\Phi(x).$$

Ответ

$\Phi(x)$.

Задача 25

ξ_1, ξ_2, …, ξ_n, … — последовательность независимых, одинаково распределенных случайных величин

$$P\{\xi_k = 1\} = P\{\xi_k = -1\} = \frac{1}{2}, \quad k = 1, 2, \ldots, n, \ldots .$$

Обозначим $\eta_n = \sum_{k=1}^{n} \xi_k / 2^k$. Найти закон распределения $\eta = \lim_{n\to\infty} \eta_n$.

Решение

Предельное распределение случайной величины η_n будем находить методом характеристических функций.

Характеристическая функция случайной величины ξ_k равна

$$f_{\xi_k}(t) = e^{it}\frac{1}{2} + e^{-it}\frac{1}{2} = \cos t .$$

Случайные величины ξ_1, ξ_2, …, ξ_n независимы, следовательно, независимы и случайные величины ξ_1, $\xi_2 / 2$, …, $\xi_n / 2^n$, являющиеся линейными функциями от независимых случайных величин, поэтому характеристическая функция случайной величины η_n равна произведению характеристических функций слагаемых $\xi_k / 2^k$, $k = 1, 2, \ldots, n$, (см. [1], гл. 7, с. 145)

$$f_{\eta_n}(t) = \prod_{k=1}^{n} f_{\frac{\xi_k}{2^k}}(t) = \prod_{k=1}^{n} f_{\xi_k}\left(\frac{t}{2^k}\right) =$$

$$= \prod_{k=1}^{n} \cos\frac{t}{2^k} = \cos\frac{t}{2}\cos\frac{t}{2^2}\ldots\cos\frac{t}{2^n}\frac{\sin\frac{t}{2^n}}{\sin\frac{t}{2^n}} .$$

Так как $\cos\frac{t}{2^k}\sin\frac{t}{2^k} = \frac{1}{2}\sin\frac{t}{2^{k-1}}$, то

$$f_{\eta_n}(t) = \frac{1}{\sin t / 2^n}\frac{1}{2}\cos\frac{t}{2}\cos\frac{t}{2^2}\ldots\cos\frac{t}{2^{n-1}}\sin\frac{t}{2^{n-1}} =$$

$$= \frac{1}{\sin t/2^n}\frac{1}{2^2}\cos\frac{t}{2}\cos\frac{t}{2^2}\ldots\cos\frac{t}{2^{n-2}}\sin\frac{t}{2^{n-2}} =$$

$$= \ldots = \frac{\sin t}{\sin t/2^n}\frac{t}{2^n t} = \frac{\sin t}{t}\frac{t/2^n}{\sin t/2^n}$$

и

$$\lim_{n\to\infty} f_{\eta_n}(t) = \frac{\sin t}{t}.$$

Известно, что $\frac{1}{2}\int_{-1}^{1} e^{ixt}dx = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}$. Следовательно, $\frac{\sin t}{t}$ есть характеристическая функция случайной величины η, имеющей равномерное распределение на отрезке $[-1, 1]$, т.е.

$$p_\eta(x) = \begin{cases} \frac{1}{2}, & \text{если } x \in [-1, 1]; \\ 0, & \text{если } x \notin [-1, 1], \end{cases}$$

и согласно теореме 2.5 (см. [1], гл. 7, § 2, с. 147) последовательность функций распределения $F_{\eta_n}(x)$, $n = 1, 2, \ldots$, слабо сходится к функции распределения $F_\eta(x)$, являющейся функцией распределения случайной величины, равномерно распределенной на отрезке $[-1, 1]$. Итак,

$$\lim_{n\to\infty} P\{\eta_n < x\} = \begin{cases} 0, & \text{если } x \le -1; \\ \frac{x+1}{2}, & \text{если } -1 < x \le 1; \\ 1, & \text{если } x > 1. \end{cases}$$

Ответ

$$p_\eta(x) = \begin{cases} \frac{1}{2}, & \text{если } x \in [-1, 1]; \\ 0, & \text{если } x \notin [-1, 1]. \end{cases}$$

Задача 26

Вероятность изделию некоторого производства, оказаться бракованным, равна 0,005. Какова вероятность того, что из 10000 наудачу взятых изделий бракованных окажется ровно 40? Найти наиболее вероятное число бракованных изделий, моду числа бракованных изделий и вероятность такого числа изделий в партии из 10000 изделий.

Решение

Случайная величина μ_{10000} — число бракованных изделий в партии из 10000, имеет биномиальное распределение

$$P\{\mu_{10000}=m\}=C_{10000}^{m}\,(0{,}005)^m\,(0{,}995)^{10000-m},\quad m=\overline{0,10000}.$$

Так как n = 10000 — большое число, и $\sqrt{npq}=49{,}75>20$, то для приближенного вычисления вероятности $P\{\mu_{10000}=40\}$ применим теорему 3.2 (локальная теорема Муавра-Лапласа) (см. [1], гл. 4, § 3, с. 63)

$$P\{\mu_n=m\}=\frac{1}{\sqrt{npq}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x_m^2}{2}},$$

где $x_m=\dfrac{m-np}{\sqrt{npq}}$.

В нашем случае $x_{40}=\dfrac{40-50}{\sqrt{49{,}75}}=-1{,}42$. По табл. П2 (см. с. 98 данных методических указаний) находим $\varphi(1{,}42)=0{,}1456$. Следовательно, $P\{\mu_{10000}=40\}=0{,}1456/\sqrt{49{,}75}=0{,}02065$.

Найдем моду случайной величины μ_{10000}. Вероятность $P\{\mu_n=m\}$ принимает максимальное значение при минимальном значении модуля разности $|m-50|$. Таким образом,

$$\max_{0\le m\le 10000} P\{\mu_{1000}=m\}=P\{\mu_{10000}=50\}=\frac{1}{\sqrt{49{,}75}}0{,}3989=0{,}0566.$$

Ответ

0,02065; $P\{\mu_{10000}=50\}\approx 0{,}0566$.

Замечание. Рассматривая вероятность $P\{\mu_n=m\}=P_n(m)$ как функцию от m, $m=\overline{0,n}$, найдем, что максимальное значение $P_n(m)$ достигает при $m_0=[(n+1)p]$ (в случае же, если $(n+1)p$ — целое число, то

$$\max_{0\le m\le n} P_n(m)=P_n((n+1)p)=P_n(np-q)\,).$$

7. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ: ВЫБОРКА, ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ, ВЫБОРОЧНЫЕ МОМЕНТЫ. ТОЧЕЧНАЯ ОЦЕНКА НЕИЗВЕСТНЫХ ПАРАМЕТРОВ; СОСТОЯТЕЛЬНЫЕ, НЕСМЕЩЕННЫЕ ОЦЕНКИ. МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРА. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Выборка, оценки

Задачи, относящиеся к сфере «компетенции» математической статистики, являются по своей сути обратными к задачам теории вероятностей. Если при постановке вероятностных задач задаются такие характеристики случайных величин как функции распределения, плотность распределения, математическое ожидание, дисперсия и т.д., то в задачах математической статистики, напротив, по результатам, как правило, независимых экспериментов, связанных со случайной величиной X, возникают задачи восстановления неизвестного закона распределения $F(x)$ (или его отдельных характеристик).

Исследователь, желающий воспользоваться методами математической статистики, должен иметь в своем распоряжении несколько, скажем n, независимых измерений той или иной величины, проводившихся в одинаковых условиях. Математической моделью такого набора измерений является выборка объема n — последовательность $(X_1, X_2, \ldots, X_n)$ независимых одинаково распределенных случайных величин.

Используя выборку, будем строить оценки интересующих нас параметров. Понятно, что в зависимости от целей или вкусов экспериментатора приближенные значения (оценки) одного и того же параметра могут иметь разный вид. Поэтому в математической статистике оценкой некоторого параметра, например θ, связанного со случайной выборкой (X_1, X_2, ..., X_n), называется произвольная функция от выборки $\hat{\theta} = \hat{\theta}_n = \hat{\theta}(X_1, X_2, ..., X_n)$.

Несмещенность, состоятельность

Поскольку величины X_1, X_2, ..., X_n случайны, то и оценка параметра $\hat{\theta}_n$ также является случайной величиной, и основная задача математической статистики состоит в том, чтобы дать ответ на вопрос: какими свойствами обладают те или иные оценки неизвестного параметра θ. Понятно, что предлагать в качестве оценки неизвестного параметра произвольную оценку было бы неразумно. Мы должны стремиться использовать «хорошие» оценки, а это, в свою очередь, предполагает наличие способов проверки качества оценок. Наиболее распространенными критериями качества являются несмещенность и состоятельность.

Оценка $\hat{\theta}_n$ параметра θ называется несмещенной, если $M\hat{\theta}_n = \theta$.

Задача 1

Пусть (X_1, X_2, ..., X_n) — выборка объема n. Показать, что оценки $\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$, $\hat{\theta} = \frac{1}{3}(X_1 + X_2 + X_n)$ являются несмещенными оценками параметра $a = MX_1$.

Решение

Используя свойства математического ожидания, находим

$$M\overline{X} = \frac{1}{n}(MX_1 + MX_2 + ... + MX_n) = \frac{1}{n}na = a; \qquad (1)$$

$$M\hat{\theta} = \frac{1}{3}(MX_1 + MX_2 + MX_n) = \frac{1}{3}3a = a. \qquad (2)$$

Хотя свойство несмещенности и является «приятным» свойством, но, как видно при решении задачи 1, этим свойством по отношению к математическому ожиданию обладает как разумная оценка $\overline{X}$, так и вряд ли приемлемая оценка $\widehat{\theta}$. Одной из причин возникновения такой ситуации является тот факт, что близость математических ожиданий к оцениваемому параметру еще не означает близость самой оценки. Для того чтобы это стало совсем очевидно, рассмотрим следующий пример.

Пусть случайные величины $X_1, X_2, \ldots, X_n$ независимы и одинаково распределены, причем

$$P\{X_i = 3000\} = P\{X_i = -3000\} = \frac{1}{2}.$$

Ясно, что $a = MX_i = 0 = M\widehat{\theta} = \frac{1}{3}M(X_1 + X_2 + X_n)$, в то время как $\left|\widehat{\theta}\right| = \frac{1}{3}\left|X_1 + X_2 + X_3\right| \geq 1000$ при любых значениях $X_i = \pm 3000$.

Суть в том, что требования несмещенности не запрещает оценке $\widehat{\theta}_n$ принимать с большой вероятностью сколь угодно далекие от параметра θ значения.

Упомянутого выше недостатка лишены в определенной степени состоятельные оценки (они не обязаны быть несмещенными).

Оценка $\widehat{\theta}$ параметра θ называется состоятельной, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right\} = 1.$$

Состоятельная оценка обладает естественным и нужным экспериментатору свойством: при увеличении объема выборки (числа экспериментов) вероятность больших отклонений оценки от оцениваемого параметра стремится к нулю.

Задача 2

Пусть $\sigma^2 = DX < \infty$. Показать, что $\overline{X}$ является состоятельной оценкой параметра $a = MX$.

Решение

Поскольку $M\overline{X} = a$ и

$$D\overline{X} = D\left\{\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right\} = \frac{1}{n^2}(DX_1 + DX_2 + \ldots + DX_n) = \frac{\sigma^2}{n},$$

можно воспользоваться неравенством Чебышева

$$P\left\{\left|\overline{X}-a\right|\geq\varepsilon\right\}\leq\frac{D\overline{X}}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}$$

при $n\to\infty$ для любого $\varepsilon>0$, что, очевидно, эквивалентно состоятельности $\overline{X}$.

При проверке состоятельности оценок полезна следующая теорема.

Теорема

Если $M\widehat{\theta}_n=\theta+k_n$, $D\widehat{\theta}_n=\nu_n$ и $\lim\limits_{n\to\infty}k_n=\lim\limits_{n\to\infty}\nu_n=0$, то оценка параметра θ состоятельна.

Доказательство

Используя соотношение:

$$\left|\widehat{\theta}_n-M\widehat{\theta}_n+k_n\right|\leq\left|\widehat{\theta}_n-M\widehat{\theta}_n\right|+\left|k_n\right|$$

и вспоминая неравенство Чебышева, находим при $\left|k_n\right|<\varepsilon$:

$$P\left\{\left|\widehat{\theta}_n-\theta\right|<\varepsilon\right\}=P\left\{\left|\widehat{\theta}_n-M\widehat{\theta}_n-k_n\right|<\varepsilon\right\}\geq P\left\{\left|\widehat{\theta}_n-M\widehat{\theta}_n\right|+\left|k_n\right|<\varepsilon\right\}=$$

$$=1-P\left\{\left|\widehat{\theta}_n-M\widehat{\theta}_n\right|\geq\varepsilon-\left|k_n\right|\right\}\geq1-\nu_n/(\varepsilon-\left|k_n\right|)^2-1$$

при $n\to\infty$.

Эмпирическая функция распределения

Одной из определяющих характеристик случайной величины является ее функция распределения $F(x)$. Если функция распределения неизвестна и в нашем распоряжении имеется выборка (X_1, X_2, …, X_n) объема n (т.е. набор независимых случайных величин, каждая из которых имеет распределение $F(x)$), то в качестве оценки $F(x)$ используют эмпирическую функцию распределения $F_n^*(x)$ (это случайная величина), определяемую равенством:

$$F_n^*(x)=\frac{1}{n}\sum_{k=1}^{n}\chi\{X_k<x\}=\frac{1}{n}\nu_n(x),$$

где $\nu_n(x)=$ (число тех X_k среди $X_1, X_2, \ldots, X_n$, которые меньше x).

Здесь $\chi(A)$ — индикатор события (см. [1], с. 91). В силу соотношений:

$$MF_n^*(x)=\frac{1}{n}\sum_{k=1}^{n}M\chi\{X_k<x\}=\frac{1}{n}nF(x)=F(x)$$

и

$$DF_n^*(x)=\frac{1}{n^2}\sum_{D}\chi\{X_k<x\}=\frac{1}{n}F(x)(1-F(x))$$

и теоремы предыдущего раздела эмпирическая функция распределения является несмещенной и состоятельной оценкой $F(x)$ в каждой точке $x\in(-\infty,\infty)$.

В практических расчетах имеют дело не с выборками $X_1, \ldots, X_n$ (наборами независимых случайных величин), а с их реализациями $x_1, x_2, \ldots, x_n$ (числами), полученными при конкретных измерениях. По реализациям выборки можно строить реализации эмпирической функции распределения. В качестве примера такого построения рассмотрим построение реализации эмпирической функции распределения по реализации выборки объема 10, заданной табл. 1.

Таблица 1

i	x_1	i	x_1
1	–2	6	1,7
2	2,3	7	0,8
3	0,5	8	2,7
4	–1,5	9	1,2
5	2,3	10	–1

Упорядочивая эту реализацию в соответствии с возрастанием ее элементов, приходим к вариационному ряду

$$-2;\ -1{,}5;\ -1;\ 0{,}5;\ 0{,}8;\ 1{,}2;\ 1{,}7;\ 2{,}3;\ 2{,}3;\ 2{,}7.$$

Отсюда, в свою очередь, для данной реализации $F_n^*(x)$ получаем график (рис. 1).

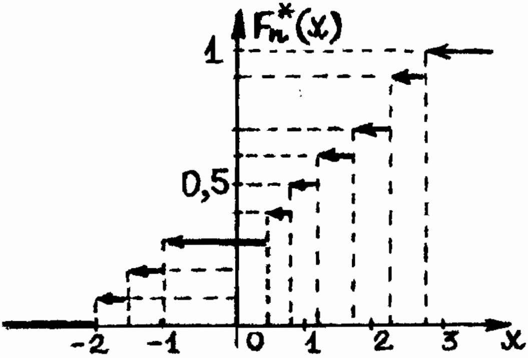

Рис. 1

Метод моментов

Пример, рассмотренный в конце предыдущего раздела, показывает, что построение реализаций эмпирической функции распределения при большом числе экспериментов — дело кропотливое и для практических задач не всегда нужное. Обычно ограничиваются оценками математического ожидания и дисперсии.

В основе таких оценок чаще всего лежит метод моментов. По методу моментов, если требуется оценить величину $a_j = MX_1^j$ — момент порядка j, то в качестве оценки следует использовать выборочный момент:

$$\hat{a}_j = \frac{1}{n}\sum_{k=1}^{n} X_k^j$$

порядка j. Если же нужно оценить величину $b_j = M(X_1 - MX_1)^j$ — центральный момент порядка j, то оценкой служит центральный выборочный момент порядка j, задаваемый равенством:

$$\widehat{b}_j = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^j .$$

Поскольку

$$M\widehat{a}_j = \frac{1}{n}\sum_{k=1}^{n} MX_k^j = a_j$$

и при $n \to \infty$

$$D\widehat{a}_j = \frac{1}{n}\sum_{k=1}^{n} DX_k^j = \frac{1}{n}(a_{2j} - a_j^2) \to 0,$$

то выборочный момент $\widehat{a}_j$ является несмещенной и, в силу теоремы, доказанной на с. 36 данных методических указаний, состоятельной оценкой параметра a_j.

Центральный выборочный момент при $j \geq 2$ не является несмещенной оценкой центрального момента. В частности, полагая $\sigma^2 = b_2$ и $\widehat{\sigma}^2 = \widehat{b}_2$, имеем ([1], с. 180)

$$M\widehat{b}_2 = M\widehat{\sigma}^2 = \frac{n-1}{n}\sigma^2$$

и, следовательно, $\sigma^2 = \widehat{b}_2$ — смещенная оценка дисперсии $\sigma^2 = b_2 = M(X_1 - MX_1)^2$. Хотя при больших значениях n смещение невелико, при малых n для оценки дисперсии обычно используют оценку

$$S^2 = \frac{n}{n-1}\widehat{b}_2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 . \qquad (3)$$

В уже упомянутой книге [1] на с. 181 показано, что оценка $\widehat{b}_j$ является состоятельной оценкой b_j (хотя и смещена). Поэтому разумность применения такого рода оценок с точки зрения критериев качества, обсуждавшихся на с. 34 — 36 данных методических указаний, очевидна.

Задача 3

По реализации выборки, заданной табл. 1 на с. 37 данных методических указаний, вычислить методом моментов оценки $\hat{\sigma}^2$, S^2 величины $\sigma^2 = M(X_1 - MX_1)^2$.

Решение

Имеем

$$\overline{X} = \frac{1}{10}\sum_{k=1}^{10} x_k = 0{,}7\ ;\ \hat{\sigma}^2 = \frac{1}{10}\sum_{k=1}^{10}(x_k - \overline{X})^2 = 2{,}424\ ;\ S^2 = \frac{10}{9}\hat{\sigma}^2 = 2{,}693.$$

Доверительные интервалы

Поскольку оценка параметра является функцией выборки, а элементы выборки являются случайными величинами, то две реализации выборки могут дать различные числовые значения для оцениваемого параметра. Так, например, если было проведено 20 независимых измерений некоторой величины, значения которой могут рассматриваться как случайные, то не исключено, что среднее значение первых десяти измерений будет значительно отличаться от среднего значения последних десяти измерений. И совершенно неясно, что выбрать в качестве оценки математического ожидания измеряемой величины. Конечно, на основе всех 20 измерений можно получить объединенную оценку, которая, как будто бы должна быть «ближе» к неизвестному математическому ожиданию элементов рассматриваемой выборки. Убежденность в том, что это должно быть так, обеспечивает нам как закон больших чисел, гласящий, что для последовательности X_1, X_2, …, X_n независимых одинаково распределенных случайных величин при любом $\varepsilon > 0$

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - MX_1\right| < \varepsilon\right\} \to 1, \quad n \to \infty, \tag{4}$$

так и центральная предельная теорема, согласно которой для любых $-\infty \le y_1 < y_2 \le \infty$ при $n \to \infty$

$$P\left\{ y_1 \le \frac{\sum\limits_{k=1}^{n} X_k - nMX_1}{\sqrt{nDX_1}} \le y_2 \right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{y_1}^{y_2} e^{-u^2/2}\, du = \Phi(y_2) - \Phi(y_1). \tag{5}$$

Однако ни соотношение (4), ни утверждение (5) не дают права ни написать с абсолютной уверенностью, используя конкретную реализацию выборки $x_1, x_2, \ldots, x_n$ объема n, что, например,

$$\theta = MX_1 = \overline{X} \pm 10000\sqrt{DX_1} \cdot \sqrt{n}\,, \tag{6}$$

ни утверждать, что с возрастанием n модуль разности $\left|\overline{X} - MX_1\right|$ будет убывать. Такая ситуация наводит на мысль о том, что следует несколько уменьшить требования и попытаться придать соотношениям вида (6) не детерминистский, а вероятностный смысл.

Пусть имеются выборка $(X_1, X_2, \ldots, X_n)$, параметр θ, связанный с распределением элементов данной выборки (математическое ожидание, дисперсия и т.п.) и две случайные величины $\underline{\theta} = \underline{\theta}(X_1, ..., X_n)$ и $\overline{\theta} = \overline{\theta}(X_1, ..., X_n)$, зависящие от выборки, причем $\underline{\theta} \le \overline{\theta}$ для любых $X_1, X_2, \ldots, X_n$.

Случайный интервал $(\underline{\theta}, \overline{\theta})$ называется <u>доверительным интервалом</u> для параметра θ <u>с уровнем доверия</u> $1 - 2\alpha$, если

$$P\{\underline{\theta} \le \theta \le \overline{\theta}\} = 1 - 2\alpha\,.$$

Сразу же возникает вопрос: как строить доверительные интервалы? Прежде чем дать один довольно общий рецепт такого построения, рассмотрим ситуацию, когда элементы выборки имеют нормальное распределение с параметрами $a = MX_i$ и $\sigma^2 = DX_i$. В этом случае величина $\overline{X}$ имеет нормальное распределение с параметрами a и σ^2 / n, и асимптотическое соотношение (5) превращается в равенство:

$$P\left\{ y_1 \le \frac{\overline{X} - a}{\sigma\sqrt{n}} \le y_2 \right\} = \Phi(y_2) - \Phi(y_1)\,. \tag{7}$$

Пусть теперь величина σ известна, а нам требуется построить доверительный интервал для a с уровнем доверия $1 - 2\alpha$. Преобразовав левую часть (7), находим

$$P\left\{ \overline{X} - y_2 \frac{\sigma}{\sqrt{n}} \le a \le \overline{X} - y_1 \frac{\sigma}{\sqrt{n}} \right\} = \frac{1}{\sqrt{2\pi}} \int_{y_1}^{y_2} e^{-u^2/2}\, du = 1 - 2\alpha\,. \tag{8}$$

Ясно, что интервал $\left(\overline{X} - y_2 \frac{\sigma}{\sqrt{n}}; \overline{X} - y_1 \frac{\sigma}{\sqrt{n}}\right)$ является доверительным интервалом для a с уровнем доверия $1-2\alpha$, и вопрос лишь в том, как выбрать значения y_1 и y_2. Такой выбор можно осуществить многими способами. Например, положить $y_2 = +\infty$, а y_1 найти из соотношения:

$$\frac{1}{\sqrt{2\pi}} \int_{y_1}^{\infty} e^{-u^2/2}\, du = 1 - 2\alpha .$$

Однако доверительный интервал $(y_1, +\infty)$, с точки зрения возможных приложений, мало информативен. Естественно выбирать значения x_1 и x_2 так, чтобы ширина доверительного интервала была как можно меньшей. Несложные прикидки с использованием симметрий подынтегральной функции в (7) подсказывают нужные значения, а именно: $y_2 = -y_1 = u_\alpha$, где u_α — квантиль уровня α, т.е. решение уравнения

$$\frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{\infty} e^{-u^2/2}\, du = \alpha . \qquad (9)$$

Итак, искомый доверительный интервал имеет вид: $\left(\overline{X} - u_\alpha \frac{\sigma}{\sqrt{n}}; \overline{X} + u_\alpha \frac{\sigma}{\sqrt{n}}\right)$. Проанализируем, что привело нас в данном случае к успеху. Во-первых, удалось найти случайную величину $\eta = (\overline{X} - a)\cdot\sqrt{n}/\sigma$, зависящую от выборки и единственного неизвестного параметра a, распределение которой не зависит от этого параметра и определяется равенством (7), а, во-вторых, сумели преобразовать систему неравенства под знаком вероятности в (7) в эквивалентную ей систему неравенств (8), разрешенную относительно неизвестного параметра. Эти соображения оправдывают введение понятия центральной статистики.

Случайная величина $G(X_1, X_2, ..., X_n; \theta)$, зависящая от выборки и единственного неизвестного параметра θ, называется центральной статистикой для параметра θ, если:

1) распределение $G(X_1, X_2, ..., X_n; \theta)$ не зависит от θ;

2) при фиксированных x_1, x_2, …, x_n функция $G(x_1, ..., x_n, \theta)$ монотонна по θ.

Покажем, как, зная центральную статистику, можно строить доверительные интервалы. Поскольку распределение $G(X_1, X_2, ..., X_n; \theta)$ не зависит от θ, то для любого $\alpha \in \left(0, \frac{1}{2}\right)$ можно выбрать такие y_1 и y_2, для которых при всех θ

$$P\{y_1 \le G(X_1, X_2, ..., X_n; \theta) \le y_2\} = \int_{y_1}^{y_2} p_G(u)\, du = 1 - 2\alpha, \quad (10)$$

где $p_G(u)$ — плотность распределения случайной величины G. В силу монотонности $G(x_1, ..., x_n; \theta)$ по θ систему неравенств под знаком вероятности в (10) можно разрешить относительно θ, если найти такие функции $\theta_i = \theta_i(X_1, X_2, ..., X_n)$, для которых

$$G(X_1, X_2, ..., X_n; \theta_i(X_1, X_2, ..., X_n)) = y_i, \quad i = 1, 2.$$

В частности, если функция G возрастает по θ, то из (10) следует неравенство $\theta_1 \le \theta_2$, и значит

$$P\{\theta_1(X_1, X_2, ..., X_n; y_1) \le \theta \le \theta_2(X_1, X_2, ..., X_n; y_2)\} = 1 - 2\alpha.$$

Если же функция G монотонно убывает по θ, то $\theta_2 \le \theta_1$ и

$$P\{\theta_2(X_1, X_2, ..., X_n; y_2) \le \theta \le \theta_1(X_1, X_2, ..., X_n; y_1)\} = 1 - 2\alpha.$$

Для завершения построения доверительного интервала осталось выбрать y_1^* и y_2^* так (если это, конечно, возможно), чтобы для каждой реализации выборки

$$\left| \theta_2(x_1, ..., x_n; y_2^*) - \theta_1(x_1, ..., x_n; y_1^*) \right| =$$

$$= \inf_{y_1, y_2} \left| \theta_2(x_1, ..., x_n; y_2) - \theta_1(x_1, ..., x_n; y_1) \right|$$

и в качестве концов искомого доверительного интервала взять величины

$$\theta_1(X_1, ..., X_n; y_1^*) \quad \text{и} \quad \theta_2(X_1, ..., X_n; y_2^*).$$

Задача 4

Пусть X_1, …, X_2 — выборка, элементы которой имеют нормальное распределение с неизвестными параметрами a и σ^2. Построить для доверительный интервал с уровнем доверия $1-2\alpha$.

Решение

Известно (см. [1], с. 185), что случайная величина

$$G(X_1,\ldots,X_n;a)=\frac{\overline{X}-a}{S}\sqrt{n},$$

где $S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2$ имеет распределение Стьюдента с $n-1$ степенями свободы. Ясно также, что $G(X_1,X_2,\ldots,X_n;a)$ монотонно убывает при возрастании a. Таким образом, случайная величина G является центральной статистикой для параметра a, причем

$$P\left\{y_1\le\frac{\overline{X}-a}{S}\sqrt{n}\le y_2\right\}=\int_{y_1}^{y_2}\tau_{n-1}(u)\,du, \tag{11}$$

где $\tau_{n-1}(u)$ — плотность распределения Стьюдента с $n-1$ степенями свободы. Решая уравнения ($i=1, 2$)

$$G(X_1,\ldots,X_n;a_i(X_1,\ldots,X_n;y_i))=y_i=\frac{\overline{X}-a_i(X_1,\ldots,X_n;y_i)}{S}\sqrt{n},$$

находим

$$a_i(X_1,\ldots,X_n;y_i)=\overline{X}-\frac{Sy_i}{\sqrt{n}},$$

и осталось из всех пар y_1, y_2, удовлетворяющих условию

$$\int_{y_2}^{y_1}\tau_{n-1}(u)\,du=1-2\alpha,$$

выбрать такие, для которых длина интервала $|a_1 - a_2| = \frac{S}{\sqrt{n}}|y_1 - y_2|$ была бы минимальной при каждом фиксированном наборе $x_1, \ldots, x_n$. Симметрия плотности распределения $\tau_{n-1}(u)$ приводит к заключению, что нужно взять $y_2 = t_{n-1,\,\alpha} = -y_1$, где $t_{n-1,\,\alpha}$ — квантиль уровня α для распределения Стьюдента с $n-1$ степенями свободы, т.е.

$$\int\limits_{t_{n-1,\,\alpha}}^{\infty} \tau_{n-1}(u)\,du = \alpha\,.$$

Итак,

$$P\left\{\overline{X} - \frac{St_{n-1,\,\alpha}}{\sqrt{n}} \le a \le \overline{X} + \frac{St_{n-1,\,\alpha}}{\sqrt{n}}\right\} = 1 - 2\alpha\,. \qquad (12)$$

В частности, рассматривая выборку из табл. 1, как выборку нормально распределенных случайных величин с неизвестными параметрами a и σ^2 с учетом результатов задачи 3, находим при $\alpha = 0{,}05$, привлекая табл. 4 из [1], с. 238: $t_{9;\,0{,}05} = 1{,}833$

$$\overline{X} \pm \frac{St_{n-1,\,\alpha}}{\sqrt{n}} = 0{,}7 \pm \frac{1{,}833\sqrt{2{,}693}}{\sqrt{10}} = 0{,}7 \pm 0{,}951\,.$$

Таким образом, в данном случае доверительный интервал для имеет вид [–0,251; 1,651].

Попытаемся в заключение осмыслить, какую же информацию о параметре a мы получили, решив поставленную задачу. Соотношение (12) говорит, что при $\alpha = 0{,}05$ случайный интервал $\left(\overline{X} - \frac{1{,}833}{\sqrt{10}}S;\, \overline{X} + \frac{1{,}833}{\sqrt{10}}S\right)$ накрывает неизвестный параметр a с вероятностью $1 - 2\alpha = 0{,}9$. Другими словами, в среднем лишь для одной из 10 реализаций выборки указанный интервал не накроет неизвестный параметр a. Однако ни для какой конкретной реализации мы не можем твердо сказать, что построенный по этой реализации интервал накрыл интересующий нас параметр. Поэтому выписывая для a неравенства $-0{,}251 \le a \le 1{,}651$, можем лишь уповать на то, что нам повезло, так как вероятность неудачи в данном случае мала.

Проверка статистических гипотез

Необходимость построения доверительных интервалов для истинного значения того или иного параметра θ возникает в целом ряде задач. Например, таких, как определение расстояния до какого-либо объекта, оценки скорости движущегося объекта, массы частиц и т.п.

Наряду с такого рода построениями весьма часто исследователя интересует некоторое конкретное значение параметра $\theta = \theta_0$. Типичную ситуацию можно описать при помощи следующего примера. При производстве деталей определенного вида происходит периодическая наладка оборудования, после чего проводится выборочный контроль качества изделий. В соответствии с технологическими требованиями номинальное (среднее) значение θ некоторой характеристики качества деталей (например, длины, веса, процентного содержания определенного вещества в сплаве и т.д.) должно равняться θ_0 со среднеквадратичным отклонением σ^2. Необходимо по выборке из n деталей, подвергшихся контролю, определить, соблюдается ли технологический режим. Для того чтобы сформулировать поставленную задачу на вероятностном языке, придется сделать ряд естественных допущений.

Предположим, что контролируемый параметр выдерживается (технологически) не абсолютно точно, а с некоторой погрешностью. Точнее, значения $X_1, X_2, \ldots, X_n$ этого параметра для контролируемых изделий являются независимыми нормально распределенными случайными величинами с параметрами a и σ^2, причем в промежутке между наладками значения a не изменяются, а величина σ^2 неизменна в ходе всего процесса производства. Наша цель — имея в своем распоряжении значения $X_1, X_2, \ldots, X_n$, указать критерий (правило) принятия одной из двух гипотез

$$H_0:\quad a = a_0 \quad \text{и} \quad H_1: a \neq a_0.$$

Первую из этих гипотез называют <u>основной</u>, а вторую — <u>конкурирующей</u>. Правило принятия либо гипотезы H_0, либо гипотезы H_1 таково. Выбирается область D (критическая область) в n-мерном пространстве, и если вектор $(X_1, ..., X_n) \in D$, то гипотеза H_0 отвергается (подвергается критике) и принимается конкурирующая гипотеза H_1. Если же $(X_1, X_2, ..., X_n) \notin D$, то гипотеза H_0 принимается. Поскольку элементы выборки случайны, то какова бы ни была область D, мы не сможем с абсолютной уверенностью различать гипотезы H_0 и H_1 и будем допускать ошибки двух видов. Вероятность α отклонения гипотезы H_0, если она верна, называется ошибкой первого рода или уровнем значимости. Ясно, что $\alpha = P\{(X_1, X_2, ..., X_n) \in D \mid a = a_0\}$. Заметим, что гипотеза H_1 является сложной, т.е. допускает неоднозначность (в зависимости от наладки) параметра a. В связи с этим вероятность β отклонения гипотезы H_1, если она верна, является функцией от параметра a, причем

$$\beta = \beta(a_1) = P\{(X_1, X_2, ..., X_n) \notin D \mid a = a_1\}, \quad a_1 \neq a_0 .$$

Для каждой конкретной альтернативы H_1: $a = a_1$ величина β называется ошибкой второго рода. Займемся теперь построением области D. Вид области D будет зависеть от того, известно или нет значение σ^2. Случай известного значения σ^2 разобран в книге [1], с. 204. Будем предполагать, что σ^2 неизвестно. Как упоминалось раньше на с. 44, если $X_i \sim N(a, \sigma^2)$, то случайная величина $G = \dfrac{(\overline{X} - a)\sqrt{n}}{S}$ имеет распределение Стьюдента с $(n-1)$ степенями свободы. Пусть α задано. Если верна гипотеза H_0, то (сравните с (12))

$$P\{\left|\overline{X} - a_0\right|\sqrt{n} / S > t_{n-1,\alpha/2}\} = \alpha .$$

Поэтому, полагая $\overline{u} = (u_1 + u_2 + ... + u_n)/n$ и выбирая в n-мерном пространстве область

$$D = \left\{(u_1, ..., u_n) : \left|\overline{u} - a_0\right|\sqrt{n} > t_{n-1,\alpha/2}\sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(u_k - \overline{u})^2}\right\},$$

получаем, что $P\{(X_1, ..., X_n) \in D / a = a_0\} = \alpha$.

Вычислим для каждого $a_1 \neq a_0$ ошибку второго рода. После очевидных преобразований, находим

$$\beta(a_1) = P\{(X_1, X_2, ..., X_n) \notin D \mid a = a_1\} = \tag{13}$$

$$= P\left\{\frac{|\overline{X} - a_0|}{S}\sqrt{n} \leq t_{n-1,\alpha/2} \mid a = a_1\right\} =$$

$$= P\left\{a_0 - \frac{St_{n-1,\alpha/2}}{\sqrt{n}} \leq \overline{X} \leq a_0 + \frac{St_{n-1,\alpha/2}}{\sqrt{n}} \mid a = a_1\right\} =$$

$$= P\left\{-t_{n-1,\alpha/2} + \frac{a_0 - a_1}{S}\sqrt{n} \leq \frac{\overline{X} - a_1}{S}\sqrt{n} \leq t_{n-1,\alpha/2} + \frac{a_0 - a_0}{S}\sqrt{n} \mid a = a_1\right\}.$$

Поскольку при условии $a = a_1$ случайная величина $(\overline{X} - a_1)\sqrt{n} / S$ имеет распределение Стьюдента с $n-1$ степенями свободы, то равенство (13) в сочетании с (11) дает

$$\beta(a_1) = \int_{y_-}^{y_+} \tau_{n-1}(u)\, du\,, \tag{14}$$

где

$$y_{\pm} = (a_0 - a_1)\sqrt{n} / S \pm t_{n-1,\alpha/2}.$$

<u>Задача 5</u>

Рассматривая элементы X_i табл. 1 на с. 37 как реализацию выборки нормально распределенных случайных величин с параметрами a и σ^2, σ^2 неизвестно, проверить с ошибкой первого рода $\alpha = 0,1$ гипотезу $a = 0,5$ против альтернативы $a \neq 0,5$.

<u>Решение</u>

Критическая область с учетом результатов задачи 4 имеет вид

$$D = \left\{(u_1, ..., u_{10}) : \frac{|\overline{u} - 0,5|}{S}\sqrt{10} > t_{9;\,0,05}\right\}.$$

Поскольку $\overline{X} = 0,7$, $S = \sqrt{2,693}$, то $\dfrac{|\overline{X} - 0,5|}{\sqrt{2,693}}\sqrt{10} = 0,3854 < 0,951$.

Гипотеза не отвергается.

Нетрудно, используя равенство (14), для любой альтернативы $\overline{H}_1: \ a = a_1$ вычислить значение ошибки второго рода β. Однако при этом сразу же возникает вопрос, а является ли для данного выбор области D оптимальным, если условием оптимальности считать минимальность ошибки второго рода? Подробный ответ на этот вопрос выходит за рамки данного пособия. Интересующийся читатель может обратиться к разд. 7.а, § 5 книги [11]. Упомянем лишь, что предложенный нами критерий действительно в определенном смысле оптимален.

До сих пор имели дело с элементами одной выборки. Предположим теперь, что в нашем распоряжении имеются две независимые выборки $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$, причем $X_i \sim N(a_1, \sigma^2)$, а $Y_j \sim N(a_2, \sigma^2)$, величина σ^2 известна, а величины a_i, $i = 1, 2$, неизвестны. Требуется построить критерий, различающий гипотезы

$$H_0: \quad a_1 = a_2 \quad \text{и} \quad H_1: \quad a_2 > a_1 \tag{15}$$

с данной ошибкой первого рода α.

Задачи такого типа могут возникать, например, при сравнении долговечности двух партий приборов, при выяснении преимуществ различных технологий и т.д.

Идея построения требуемого критерия проста: нужно воспользоваться тем, что линейная комбинация независимых нормально распределенных случайных величин также является нормально распределенной случайной величиной.

Пусть

$$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i\,; \quad \overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i\,.$$

Ясно, что $M\overline{X} = \dfrac{1}{n_1}\sum\limits_{i=1}^{n_1} MX_i = a_1$; $M\overline{Y} = \dfrac{1}{n_2}\sum\limits_{i=1}^{n_2} MY_i = a_2$ и $D\overline{X} = \sigma^2 / n_1$, $D\overline{Y} = \sigma^2 / n_2$. Поэтому

$$M(\overline{Y} - \overline{X}) = a_2 - a_1\,; \quad D(\overline{Y} - \overline{X}) = D\overline{Y} + D\overline{X} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

и, следовательно, $\bar{Y}-\bar{X}-N\left(a_2-a_1,\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$. Полагая $\Delta=a_2-a_1$, переформулируем гипотезы (15) в виде

$$H_0':\quad \Delta=0 \quad \text{и} \quad H_1':\quad \Delta>0.$$

Очевидно, что

$$P\left\{(\bar{Y}-\bar{X})/\sigma\sqrt{1/n_1+1/n_2}>x \,\middle|\, H_0'\right\}=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-u^2/2}\,du.$$

Отсюда, полагая $x=u_\alpha$, где u_α — квантиль уровня α (9), получаем, что нужное критическое множество D имеет вид

$$D=\left\{(x_1,\dots,x_{n_1}),(y_1,\dots,y_{n_2}):\bar{y}-\bar{x}>u_\alpha\sigma\sqrt{1/n_1+1/n_2}\right\}.$$

Задача 6

Рассматривая первые 5 членов табл. 1 на с. 37 как реализацию выборки из нормального распределения с параметрами a_1 и $\sigma^2=4$, а последние 5 — как реализацию выборки из нормального распределения с параметрами a_2 и $\sigma^2=4$, проверить гипотезы H_0: $a_1=a_2$ и H_1: $a_2>a_1$ с ошибкой первого рода $\alpha=0{,}01$.

Решение

В данном случае $\sigma=2$, $n_1=n_2=5$, $u_\alpha=2{,}33$ и область D имеет вид

$$D=\left\{(x_1,\dots,x_5),(y_1,\dots,y_5):\bar{y}-\bar{x}>2{,}33\cdot 2\sqrt{\frac{2}{5}}=2{,}95\right\}.$$

Поскольку

$$\bar{x}=\frac{1}{5}(-2+2{,}3+0{,}5-1{,}5+2{,}3)=0{,}32,$$

$$\bar{y}=\frac{1}{5}(1{,}7+0{,}8+2{,}7+1{,}2-1)=1{,}14$$

и $\bar{y}-\bar{x}=1{,}14-0{,}32=0{,}82<2{,}95$, гипотеза H_0 принимается.

ЗАДАЧИ К РАЗДЕЛУ 7

Задача 1

Имеется выборка объема 8:

$$2,\ 4,\ 2,\ 6,\ 4,\ 8,\ 4,\ 10. \tag{16}$$

Построить по этой выборке эмпирическую функцию распределения. Найти оценку математического ожидания (выборочное среднее) и несмещенную оценку для дисперсии.

Решение

Построим вариационный ряд для выборки (16) (см. [1], с. 177)

$$2,\ 2,\ 4,\ 4,\ 4,\ 6,\ 8,\ 10. \tag{17}$$

Тогда эмпирическая функция распределения $F_8^*(x)$ определяется формулой:

$$F_8^*(x) = \frac{\mu_8(x)}{8},$$

где $\mu_8(x)$ — число значений в выборке (16) (или в ряде (17)), меньших x (см. [1], равенство (2.1), с. 177). Следовательно,

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le 2; \\ 0{,}25, & \text{если } 2 < x \le 4; \\ 0{,}625, & \text{если } 4 < x \le 6; \\ 0{,}75, & \text{если } 6 < x \le 8; \\ 0{,}875, & \text{если } 8 < x \le 10; \\ 1, & \text{если } x > 10. \end{cases}$$

Равенства (3.4) и (3.8) (см. [1], с. 177 — 178) определяют соответственно выборочное среднее и несмещенную оценку для дисперсии

$$\overline{X} = \frac{1}{8}\sum_{k=1}^{8} x_k = \frac{2\cdot 2 + 3\cdot 4 + 6 + 8 + 10}{8} = 5,$$

$$S^2 = \frac{1}{8-1}\sum_{k=1}^{8}(x_k - \overline{X})^2 = \frac{1}{7}(2\cdot 3^2 + 3 + 1 + 3^2 + 5^2) = 8.$$

Ответ

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le 2; \\ 0{,}25, & \text{если } 2 < x \le 4; \\ 0{,}625, & \text{если } 4 < x \le 6; \\ 0{,}75, & \text{если } 6 < x \le 8; \\ 0{,}875, & \text{если } 8 < x \le 10; \\ 1, & \text{если } x > 10, \end{cases} \qquad \overline{X} = 5, \quad S^2 = 8.$$

Задача 2

Предполагая, что выборка (16) является выборкой из элементов экспоненциального распределения с неизвестным параметром λ, найти оценку максимального правдоподобия для λ.

Решение

Так как выборка взята из значений случайной величины, имеющей плотность распределения $p(x_k, \lambda)$, то функция правдоподобия $L(x_1, x_2, ..., x_n; \lambda)$ равна

$$L(x_1, x_2, ..., x_n; \lambda) = \prod_{k=1}^{n} p(x_k, \lambda)$$

(см. [1], равенство (4.14), с. 193). В задаче $\lambda = \theta$ — неизвестный параметр распределения, $n = 8$,

$$p(x_k, \lambda) = \lambda\, e^{-\lambda x_k}, \quad x_k > 0.$$

Итак,

$$L(x_1, ..., x_8; \lambda) = \lambda^8 e^{-\lambda \sum\limits_{k=1}^{8} x_k}.$$

Найдем значение λ, при котором функция правдоподобия $L(x_1, ..., x_8; \lambda)$ (или $\ln L(x_1, ..., x_8; \lambda)$) достигает максимального значения:

$$\ln L(x_1, ..., x_8; \lambda) = 8 \ln \lambda - \lambda \sum_{k=1}^{8} x_k = 8 \ln \lambda - 40\lambda,$$

$$\frac{d \ln L}{d\lambda} = \frac{8}{\lambda} - 40.$$

Нетрудно показать, что при $\hat{\lambda} = 0{,}2$ функция $L(x_1, ..., x_8; \lambda)$ принимает максимальное значение. Значит, оценкой максимального правдоподобия для λ является $\hat{\lambda} = 0{,}2$.

Ответ

$\hat{\lambda} = 0{,}2$.

Задача 3

Предполагая, что элементы выборки (16) являются элементами выборки из совокупности пуассоновских случайных величин с неизвестным параметром λ, найти оценку максимального правдоподобия для λ.

Решение

В данной задаче закон распределения элементов в выборке дискретный и согласно равенству (1.5) (см. [1], с. 193)

$$L(x_1, ..., x_8; \lambda) = \prod_{k=1}^{8} p_{x_k}(\lambda),$$

где

$$p_{x_k}(\lambda) = e^{-\lambda} \frac{\lambda^{x_k}}{x_k!}, \quad \lambda > 0.$$

Следовательно,

$$L(x_1, ..., x_8; \lambda) = e^{-8\lambda} \lambda^{\sum_{k=1}^{8} x_k} \prod_{k=1}^{8} (x_k!)^{-1} .$$

Рассмотрим $\ln L(x_1, ..., x_8; \lambda) = -8\lambda + \sum_{k=1}^{8} x_k \ln \lambda + \ln \prod_{k=1}^{8} (x_k!)^{-1}$. Из уравнения $\frac{d \ln L}{d\lambda} = -8 + \frac{40}{\lambda} = 0$ находим $\hat{\lambda} = 5$. Так как $\frac{d^2 \ln L}{d\lambda^2} = -\frac{40}{\lambda^2} < 0$, то при $\hat{\lambda} = 5$ функция правдоподобия принимает максимальное значение.

Ответ

$\hat{\lambda} = 5$.

Задача 4

Проведено 10 испытаний Бернулли с неизвестной вероятностью успеха θ. Результаты испытаний таковы:

$$1, 0, 0, 1, 1, 0, 1, 0, 0, 0, \qquad (18)$$

где 1 означает успех, а 0 — неуспех в соответствующем испытании. Построить по выборке (18) эмпирическую функцию распределения. Найти по выборке (18) выборочное среднее и несмещенную оценку для дисперсии (выборочную дисперсию). Получить оценку максимального правдоподобия для параметра θ.

Решение

Как и в задаче 1 данного разд. 7, построим для данной выборки вариационный ряд

$$0, 0, 0, 0, 0, 0, 1, 1, 1, 1.$$

Тогда эмпирическая функция распределения равна

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le 0; \\ 0{,}6, & \text{если } 0 < x \le 1; \\ 1, & \text{если } x > 1. \end{cases}$$

Выборочное среднее $\overline{X}$ и несмещенная оценка для дисперсии S^2 (выборочная дисперсия) соответственно равны

$$\overline{X} = \frac{1}{10}\sum_{k=1}^{10} x_k = 0{,}4$$

и

$$S^2 = \frac{1}{10-1}\sum_{k=1}^{10}(x_k - \overline{X})^2 = \frac{1}{9}(6\cdot 0{,}16 + 4\cdot 0{,}36) \approx 0{,}267 .$$

Так как

$$p_{x_k}(\theta) = \theta^{x_k}(1-\theta)^{1-x_k} = \begin{cases} \theta, & \text{если } x_k = 1; \\ 1-\theta, & \text{если } x_k = 0, \end{cases} \qquad 0 < \theta < 1,$$

то равенство (4.15) (см. [1], с. 193) позволяет получить функцию правдоподобия

$$L(x_1, \ldots, x_{10}, \theta) = \prod_{k=1}^{10} p_{x_k}(\theta) = \theta^{\sum\limits_{k=1}^{10} x_k}(1-\theta)^{10-\sum\limits_{k=1}^{10} x_k}$$

и

$$\ln L(x_1, \ldots, x_{10}, \theta) = \sum_{k=1}^{10} x_k \ln\theta + \left(10 - \sum_{k=1}^{10} x_k\right)\ln(1-\theta) =$$

$$= 4\ln\theta - 6\ln(1-\theta),$$

$$\frac{d\ln L}{d\theta} = \frac{4}{\theta} - \frac{6}{1-\theta} = \frac{4-10\theta}{\theta(1-\theta)},$$

откуда $\hat{\theta} = 0{,}4$, легко показать, что при $\hat{\theta} = 0{,}4$ функция правдоподобия $L(x_1, \ldots, x_{10}; \theta)$ принимает максимальное значение и, следовательно, $\hat{\theta} = 0{,}4$ — оценка максимального правдоподобия.

Ответ

$$F^*_{10}(x) = \begin{cases} 0, & \text{если } x \le 0; \\ 0{,}6, & \text{если } 0 < x \le 1; \\ 1, & \text{если } x > 1; \end{cases} \qquad \overline{X} = 0{,}4; \quad S^2 = 2{,}67; \quad \hat{\theta} = 0{,}4 .$$

Задача 5

Предполагая, что выборка (16) является выборкой из совокупности нормально распределенных случайных величин с параметрами $\theta_1 = a$ и $\theta_2 = \sigma^2$, получить оценку максимального правдоподобия для θ_2 при: 1) $a = 5$; 2) неизвестном a. Какая из этих оценок является несмещенной?

Решение

1) Согласно равенству (4.14) (см. [1], с. 193) функция правдоподобия $L(x_1, ..., x_8; \theta_2)$ равна

$$L(x_1, ..., x_8; \theta_2) = \left(\frac{1}{\sqrt{2\pi}\sqrt{\theta_2}}\right)^8 e^{-\frac{1}{2\theta_2}\sum_{k=1}^{8}(x_k - a)^2}.$$

Рассмотрим

$$\ln L = -\frac{8}{2}\ln\theta_2 - \frac{1}{2\theta_2}\sum_{k=1}^{8}(x_k - 5)^2 - 4\ln(2\pi),$$

$$\frac{d\ln L}{d\theta_2} = -\frac{4}{\theta_2} + \frac{1}{2\theta_2^2}\sum_{k=1}^{8}(x_k - 5)^2.$$

Из уравнения правдоподобия находим

$$\hat{\theta}_2 = \frac{1}{8}\sum_{k=1}^{8}(x_k - 5)^2 = \frac{1}{8}\sum_{k=1}^{8}(x_k - 5)^2 = 7.$$

Решением уравнения правдоподобия является $\hat{\theta}_2 = 7$. Так как при $\theta_2 = 7$ $\frac{d^2\ln L}{d\theta^2} < 0$, то $\hat{\theta}_2 = 7$ — оценка максимального правдоподобия.

Ввиду равенства:

$$M\hat{\theta}_2 = M\left(\frac{1}{8}\sum_{k=1}^{8}(x_k - 5)^2\right) = \frac{1}{8}\sum_{k=1}^{8}M(x_k - 5)^2 = \theta_2,$$

$\hat{\theta}_2 = 7$ есть несмещенная оценка для параметра θ.

2) Во втором случае функция правдоподобия

$$L(x_1, ..., x_8; \theta_1, \theta_2)$$

зависит от двух параметров θ_1 и θ_2. Используя решение примера 6 (см. [1], с. 194), получаем

$$\hat{\theta}_1 = \frac{1}{8}\sum_{k=1}^{8} x_k = \overline{X} = 5,$$

$$\hat{\theta}_2 = \frac{1}{8}\sum_{k=1}^{8}(x_k - \overline{X})^2 = 7,$$

где

$$M\hat{\theta}_2 = \frac{1}{8}\sum_{k=1}^{8} M(x_k - \overline{X})^2 = \frac{7}{8}\theta_2.$$

Следовательно, в этом случае оценка $\hat{\theta}_2$ — смещенная.

Ответ

1) $\hat{\theta}_2 = 7$, $\hat{\theta}_2$ — несмещенная оценка;

2) $\hat{\theta}_2 = 7$, $\hat{\theta}_2$ — смещенная оценка.

Задача 6

Предполагая, что выборка (16) является выборкой из совокупности нормально распределенных случайных величин с параметрами $\theta_1 = a$ и $\theta_2 = \sigma^2$, построить доверительный интервал:

1) для θ_1 с уровнем доверия $1 - 2\alpha = 0{,}9$, если $\sigma^2 = 4$;

2) для θ_1 с уровнем доверия $1 - 2\alpha = 0{,}95$, если θ_2 неизвестно;

3) для $\theta_2 = \sigma^2$ с уровнем доверия $1 - 2\alpha = 0{,}9$, если $a = 5$;

4) для $\theta_2 = \sigma^2$ с уровнем доверия $1 - 2\alpha = 0{,}95$, если $\theta_1 = a$ неизвестно.

Решение

1) Из теоремы 3.4 (см. [1], с. 184) следует, что $\dfrac{\overline{X}-\theta_1}{2/\sqrt{8}}$ нормально распределена с параметрами 0 и 1. Из табл. 3 (см. [1], с. 237) находим квантиль нормального распределения $u_{0,05}=1{,}6449$. Следовательно,

$$P\left\{\frac{\left|\overline{X}-\theta_1\right|}{2/\sqrt{8}}<1{,}6449\right\}=0{,}9,$$

откуда получаем, учитывая, что $\overline{X}=5$,

$$P\left\{5-\frac{1{,}6449}{\sqrt{2}}<\theta_1<5+\frac{1{,}6449}{\sqrt{2}}\right\}=0{,}9.$$

Окончательно

$$P\{3{,}83<\theta_1<6{,}17\}=0{,}9.$$

2) Если дисперсия неизвестна, то из теоремы 3.5 (см. [1], с. 185) следует, что величина $\tau_{n-1}=\dfrac{\overline{X}-\theta_1}{\sqrt{m_2}}\sqrt{n-1}$ имеет распределение Стьюдента с $(n-1)$ степенью свободы. $m_2=\dfrac{1}{n}\sum\limits_{k=1}^{n}(x_k-\overline{X})^2$. Для выборки (16) $\overline{X}=5$, $m_2=7$, $n=8$. Из табл. 4 (см. [1], с. 238) находим квантиль распределения τ_7 : $t_{0,025;\,7}=2{,}365$

$$P\{\left|\tau_7\right|<2{,}365\}=0{,}95.$$

Следовательно,

$$P\{\left|5-\theta_1\right|<2{,}365\}=0{,}95$$

и доверительный интервал для θ_1, если $\theta_2=\sigma^2$ неизвестно, такой

$$P\{2{,}635<\theta_1<7{,}365\}=0{,}95.$$

3) Перейдем к построению доверительных интервалов для дисперсии. Известно, что $a=5$, тогда $\sum\limits_{k=1}^{8}(x_k-a)^2/\sigma^2=\chi_8^2$ (имеет

распределение χ^2 с 8 степенями свободы). Из табл. П5 на с. 101 — 102 данных методических указаний находим $\chi^2_{0,05}(8) = 2,73$; $\chi^2_{0,95}(8) = 15,5$, т.е.

$$P\{\chi^2_8 < 2,73\} = 0,05,$$

$$P\{\chi^2_8 < 15,5\} = 0,95,$$

откуда получаем $P\{2,73 < \chi^2_8 < 15,5\} = 0,9$,

$$P\left\{2,73 < \frac{56}{\theta_2} < 15,5\right\} = 0,9,$$

$$P\left\{\frac{56}{15,5} < \theta_2 < \frac{56}{2,73}\right\} = 0,9,$$

$$P\{3,61 < \theta_2 < 20,51\} = 0,9.$$

4) Если $\theta_1 = a$ неизвестно, то из теоремы 3.4 (см. [1], с. 184) следует, что величина $\frac{1}{\theta_2}\sum_{k=1}^{8}(x_k - \overline{X})^2$ имеет распределение χ^2 с $8 - 1 = 7$ степенью свободы.

Из табл. П5 на с. 101 — 102 данных методических указаний находим квантили распределения χ^2_7:

$$\chi^2_{0,025}(7) = 1,69 \quad \text{и} \quad \chi^2_{0,975}(7) = 16,00,$$

т.е. $P\{\chi^2_7 < 1,69\} = 0,025$ и $P\{\chi^2_7 < 16,00\} = 0,975$. Следовательно,

$$P\{1,69 < \chi^2_7 < 16\} = 0,95$$

и тогда доверительный интервал для θ_2 имеет вид

$$P\left\{\frac{56}{16} < \theta_2 < \frac{56}{1,69}\right\} = P\{3,5 < \theta_2 < 33,13\} = 0,95.$$

Ответ

1) (3,83; 6,17), 2) (2,635; 7,365), 3) (3,61; 20,51), 4) (3,5; 33,13).

Задача 7

Предполагая, что выборка (16) является выборкой из совокупности нормально распределенных случайных величин с параметрами $\theta_1 = a$ и $\sigma^2 = 9$, проверить с уровнем значимости (ошибкой 1-го рода) $\alpha = 0{,}05$ гипотезы H_0 против альтернативы H_1 в следующих случаях:

1) $H_0: a = a_0 = 4$; 2) $H_0: a = a_0 = 6$; 3) $H_0: a = a_0 = 4$
$H_1: a = a_1 > 4$; $H_1: a = a_1 < 6$; $H_1: a = a_1 \neq 4$.

В первых двух случаях найти ошибку 2-го рода.

Решение

При любой из гипотез (H_0 или H_1) $\overline{X}$ имеет нормальное распределение с $D\overline{X} = \sigma^2 / n = 9/8$, $M\overline{X} = a$ зависит от гипотезы, следовательно, при различных гипотезах различаются и распределения. Поэтому вероятности событий, вычисляемые при гипотезах H_0 и H_1, будем обозначать соответственно P_0 и P_1.

1) В [1], с. 204, пример 1, разобран первый случай: $a_0 < a_1$. Выберем некоторое C: $a_0 < C < a_1$. Критерий можно сформулировать следующим образом: если $\overline{X} > C$, принимается гипотеза H_1, а если $\overline{X} < C$, то гипотеза H_0 не отвергается.

Если верна гипотеза H_0 и произошло событие $\overline{X} > C$, то принимается гипотеза H_1. Вероятность отвергнуть H_0, когда она верна (ошибка первого рода) равна $\alpha = 0{,}05 = P_0\{\overline{X} > C\}$.

Если верна гипотеза H_1 и произошло событие $\overline{X} < C$, то принимается H_0. Вероятность принять H_0, когда верна H_1 (ошибка второго рода), равна $\beta = P_1\{\overline{X} < C\}$.

Так как $\alpha = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{u_\alpha}^{\infty} e^{-t^2/2} dt$, то при из табл. 3 (см. [1], с. 237) находим $u_{0,05} = 1{,}6449$. Следовательно,

$$\alpha = 0{,}05 = P_0\{\overline{X} > C\} = P_0 \left\{ \frac{\overline{X} - a_0}{\sigma/\sqrt{n}} > \frac{C - a_0}{\sigma/\sqrt{n}} \right\},$$

и

$$(C - a_0)\frac{\sqrt{n}}{\sigma} = (C - 4)\frac{\sqrt{8}}{3} = u_{0,05} = 1{,}6449.$$

Тогда

$$C = a_0 + u_{0,05} \frac{\sigma}{\sqrt{n}} = 4 + 1{,}6449 \frac{3}{\sqrt{8}} = 5{,}74 .$$

Таким образом, критическая область D есть правосторонний интервал (C, ∞), т.е. интервал $(5{,}74; +\infty)$, лежащий правее точки 5,74 (рис. 2).

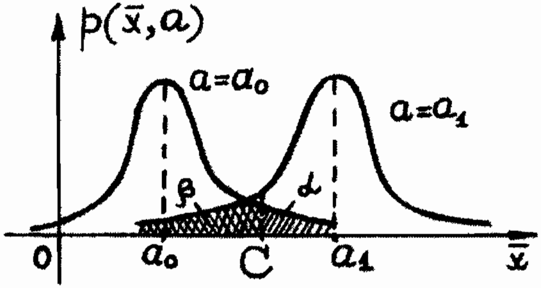

Рис. 2

В выборке (16) $\overline{X} = 5$. Так как $\overline{X} = 5 < 5{,}74 = C$, то гипотеза H_0: $a = a_0 = 4$ не отвергается.

Ошибка 2-го рода удовлетворяет соотношению:

$$\beta = P_1\{\overline{X} < C\} = P_1 \left\{ \frac{\overline{X} - a_1}{\sigma / \sqrt{n}} < \frac{C - a_1}{\sigma / \sqrt{n}} \right\},$$

поэтому $(C - a_1)\dfrac{\sqrt{n}}{\sigma} = u_{1-\beta} = -u_\beta$ и получаем

$$a_0 + u_\alpha \frac{\sigma}{\sqrt{n}} = -\frac{\sigma}{\sqrt{n}} u_\beta + a_1,$$

$$-u_\beta = -\frac{(a_1 - a_0)}{\sigma}\sqrt{n} + u_{0,05},$$

откуда

$$-u_\beta = 1{,}6449 - \frac{a_1 - 4}{3}\sqrt{8},$$

а так как $u_\beta < 0$, то $\beta = 1 - \Phi\left(\frac{a_1 - 4}{3}\sqrt{8} - 1{,}6449\right)$ зависит от значения a_1 в гипотезе H_1.

2) Рассмотрим второй случай. H_0: $a = a_0 = 6$; H_1: $a = a_1 < 6$, $\alpha = 0{,}05$. В этом случае критическая область есть левосторонний интервал $(-\infty, C)$, $u_{1-\alpha} = -u_\alpha = 1{,}6449$ и $C = a_0 - u_{0,05}\frac{3}{\sqrt{8}} = 4{,}26$.

Критерий в этом случае формулируется так: если $\overline{X} > C$, то гипотеза H_0 не отвергается, если же $\overline{X} < C$, то принимается гипотеза H_1. В выборке (16) $\overline{X} = 5$ и ввиду неравенства $\overline{X} = 5 > 4{,}26$, гипотеза H_0 не отвергается.

Ошибка 2-го рода

$$\beta = P_1\{\overline{X} > C\} = P_1\left\{\frac{\overline{X} - a_1}{\sigma/\sqrt{n}} > \frac{C - a_1}{\sigma/\sqrt{n}}\right\},$$

$$u_\beta = \frac{C - a_1}{3/\sqrt{8}}, \quad C = \frac{3}{\sqrt{8}}u_\beta + a_1 = a_0 - \frac{3}{\sqrt{8}}u_\alpha,$$

$$u_\beta = \frac{a_0 - a_1}{3/\sqrt{8}} - u_\alpha = \frac{6 - a_1}{3}\sqrt{8} - 1{,}6449.$$

Таким образом, ошибка 2-го рода — вероятность принять гипотезу H_1, когда верна H_0, зависит от значения a_1 в гипотезе H_1 и $\beta = 1 - \Phi\left(\frac{6 - a_1}{3}\sqrt{8} - 1{,}6449\right)$.

3) Рассмотрим гипотезы H_0: $a = a_0 = 4$, H_1: $a = a_1 \neq 4$. Уровень значимости $\alpha = 0{,}05$; $\sigma^2 = 9$, $n = 8$, $\overline{X} = 5$. Из табл. 3 (см. [1], с. 237) находим $u_{\alpha/2} = u_{0,025} = 1{,}96$ и $u_{1-\alpha/2} = -u_{\alpha/2} = -1{,}96$ и кри-

тическая область в данном случае двусторонняя: ее образуют интервалы $(-\infty; -1{,}96)$ и $(1{,}96; +\infty)$. Так как

$$\frac{\bar{X}-a_0}{\sigma/\sqrt{n}}=\frac{5-4}{3}\sqrt{8}=\frac{\sqrt{8}}{3}\approx 0{,}9428$$

и полученное число лежит в интервале $(-u_{\alpha/2}, u_{\alpha/2})$, то гипотеза H_0 не отвергается.

Ответ

Гипотеза H_0 не отвергается во всех случаях.

1) $\beta=1-\Phi\left(\dfrac{a_1-4}{3}\sqrt{8}-1{,}6449\right)$,

2) $\beta=1-\Phi\left(\dfrac{6-a_1}{3}\sqrt{8}-1{,}6449\right)$.

Задача 8

По выборке из задачи 4 проверить гипотезу H_0: $p=\dfrac{1}{2}$ против альтернативы H_1: $p>\dfrac{1}{2}$ с уровнем значимости $\alpha=0{,}025$.

Решение

При большом числе наблюдений можно применить свойство

$$(\bar{X}-p)\Big/\sqrt{\frac{p(1-p)}{n}}\sim N(0,1),$$

при малых n это утверждение не выполняется.

В случае малого числа испытаний проверка проводится следующим образом. Полагаем $\gamma=1-2\alpha=0{,}95$. При $n-m=6$ и $m=4$ по табл. П4 на с. 100 данных методических указаний находим $p_1=0{,}12$ (это нижнее число) и из неравенства $p=0{,}5>p_1$ делаем вывод, что гипотеза H_0 принимается.

Ответ

H_0 принимается.

Задача 9

Дана выборка из некоторой генеральной совокупности:

03	99	11	04	61	93	71	68	94	08	32	46	53	84	60	95
82	32	88	61	81	91	61	38	55	59	55	54	32	88	80	08
35	56	60	04	73	54	77	62	71	29	92	38	53	17	29	13

Проверить гипотезу (H_0) о равномерном распределении на интервале (0, 100) генеральной совокупности.

Решение

Плотность равномерно распределенной на отрезке $[a, b]$ величины

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \notin [a, b] \end{cases}$$

зависит от двух параметров a и b. В нашей задаче $a = 0$, $b = 100$ — параметры известны и

$$p(x) = \begin{cases} \dfrac{1}{100}, & \text{если } x \in [0, 100]; \\ 0, & \text{если } x \notin [0, 100]. \end{cases}$$

Для проверки гипотезы H_0 используем критерий Пирсона (хи-квадрат критерий). Согласно этому критерию, случайная величина $\chi_m^2 = \sum_{l=1}^{r} \dfrac{(\nu_i - np_i)^2}{np_i}$ при справедливости гипотезы H_0 будет иметь хи-квадрат распределение с $m = r - 1 - s$ степенями свободы, где r — число интервалов после группировки данных, s — число параметров проверяемого распределения, вычисляемых по выборке (в нашем случае $s = 0$); p_i — теоретические вероятности попадания генеральной совокупности в i-й интервал, а именно:

$$p_i = P\{\omega : \xi(\omega) \in \Delta_i\} = \int_{\Delta_i} p(x)\,dx = \int_{\Delta_i} \frac{1}{100}\,dx = \frac{\Delta_i}{100}$$

в нашем случае, ν_i — число вариант выборки, попавших в i-й интервал. Объем выборки $n = 48$, разбиваем отрезок [0, 100] на интервалы, учитывая, что в каждом интервале должно быть не менее 7 — 10 вариант. Оптимальное разбиение соответствует вероятностям $p_r = \frac{1}{r}$: $p_1 = p_2 = p_3 = p_5 = p_6 = \frac{1}{6}$. Следовательно, возьмем интервалы Δ_i длины $\frac{100}{6} = 16{,}6$; $i = 1, 2, 3, 4, 5, 6$.

Подсчитаем число вариант, попавших в эти интервалы (табл. 2).

Таблица 2

Интервалы	Число вариант ν_i
$\left(0, \frac{100}{6}\right)$	7
$\left(\frac{100}{6}, 2\frac{100}{6}\right)$	6
$\left(2\frac{100}{6}, 3\frac{100}{6}\right)$	4
$\left(3\frac{100}{6}, 4\frac{100}{6}\right)$	14
$\left(4\frac{100}{6}, 5\frac{100}{6}\right)$	8
$\left(5\frac{100}{6}, 100\right)$	9

Найдем границы критической области S_c, исходя из определения $P_{H_0}\left\{\chi_m^2 \in S_c\right\} = \alpha$. Ошибка первого рода $\alpha = 0{,}05$ считается заданной, а число степеней свободы равно $m = 6 - 1 = 5$. Используя таблицу χ-квадрат (см. табл. П5 на с. 101 — 102 данных методических указаний), находим $S_c = (11{,}1; +\infty)$, т.е. $P\left\{\chi_5^2(\omega) > 11{,}1\right\} = 0{,}05$. Найдем теперь наблюдаемое значение критерия

$$\chi_5^2(\text{набл.}) = \sum_{i=1}^{6} \frac{(\nu_i - np_i)^2}{4p_i} = \frac{\left(7 - n\cdot\frac{1}{6}\right)^2}{n\cdot 1/6} + \frac{\left(6 - n\cdot\frac{1}{6}\right)^2}{n\cdot 1/6} + \frac{\left(4 - n\cdot\frac{1}{6}\right)^2}{n\cdot 1/6} +$$

$$+\frac{\left(14-n\cdot\frac{1}{6}\right)^2}{n\cdot\frac{1}{6}}+\frac{\left(8-n\cdot\frac{1}{6}\right)^2}{n\cdot\frac{1}{6}}+\frac{\left(9-n\cdot\frac{1}{6}\right)^2}{n\cdot\frac{1}{6}}=$$

$$=\frac{(7-8)^2+(6-8)^2+(4-8)^2+(14-8)^2+(8-8)^2+(9-8)^2}{8}=\frac{58}{8}=7{,}25$$

($n=48$ — объем выборки).

Так как $7{,}25<11{,}1$, следовательно, наблюдаемое значение критерия попадает в область допустимых значений $\bar{S}_c$, т.е. $\chi_5^2(\text{набл.})\in\bar{S}_c$, поэтому гипотеза H_0 не отвергается на уровне значимости $\alpha=0{,}05$ (при $\alpha=0{,}1$: $S_c^1=(9{,}2;+\infty)-H_0$ — не отвергается, и при $\alpha=0{,}2$ $S_c^2=(7{,}3;+\infty)$ H_0 — также не отвергается).

Ответ

Гипотеза H_0 о равномерном распределении не отвергается при $\alpha=0{,}05$; $0{,}01$; $0{,}2$.

Задача 10

Дана выборка ξ_1, ξ_2, ..., ξ_n из генеральной совокупности, подчиняющейся закону распределения с плотностью $p_\xi(x)=\frac{\sqrt{\alpha}}{\pi}\mathrm{e}^{-\alpha x^2}$. Найти оценку параметра $\hat{\alpha}$ и дисперсию оценки.

Решение

Составим функцию правдоподобия

$$L(x_1,x_2,...,x_n,\alpha)=\prod_{i=1}^{n}p(x_i,\alpha)=\prod_{i=1}^{n}\left(\frac{\alpha}{\pi}\right)^{1/2}\mathrm{e}^{-\alpha x_i^2}=\left(\frac{\alpha}{\pi}\right)^{\frac{n}{2}}\mathrm{e}^{-\alpha\sum\limits_{i=1}^{n}x_i^2}.$$

Найдем значение параметра α, при котором функция $L(x_1,...,x_n,\alpha)$ достигает максимума:

$$\ln L(x_1,...,x_n,\alpha)=\frac{n}{2}\ln\left(\frac{\alpha}{\pi}\right)-\alpha\sum_{i=1}^{n}x_i^2,$$

$$\frac{\partial \ln L(x_1, ..., x_n, \alpha)}{\partial \alpha} = \frac{n}{2\alpha} - \sum_{i=1}^{n} x_i^2 = 0 \quad \to \quad \alpha = \frac{n}{2\sum_{i=1}^{n} x_i^2}.$$

Следовательно, в качестве оценки параметра α берем статистику $\hat{\alpha} = n\left(2\sum_{i=1}^{n} \xi_i^2\right)^{-1}$. Чтобы найти дисперсию оценки, воспользуемся неравенством Крамера-Рао (см. [1], с. 189):

$$D\hat{\alpha} \geq \frac{1}{n\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(x,\alpha)}{\partial \alpha}\right]^2 p(x,\alpha)\,dx},$$

где $p(x,\alpha)$ — плотность распределения генеральной совокупности, а знак равенства достигается на оценках максимального правдоподобия. Следовательно, в нашем случае

$$D\hat{\alpha} = D\frac{n}{2\sum_{i=1}^{n} x_i^2} = \frac{1}{n\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(x,\alpha)}{\partial \alpha}\right]^2 p(x,\alpha)\,dx}. \tag{19}$$

Вычислим интеграл J, стоящий в знаменателе равенства (19)

$$\ln p(x,\alpha) = \ln\left(\sqrt{\alpha/\pi}\; e^{-\alpha x^2}\right) = \frac{1}{2}\ln(\alpha/\pi) - \alpha x^2.$$

Следовательно, $\frac{\partial \ln p(x,\alpha)}{\partial \alpha} = \frac{1}{2\alpha} - x^2$. Подставляя полученное выражение в интеграл J, получаем

$$J = \int_{-\infty}^{\infty}\left(x^2 - \frac{1}{2\alpha}\right)^2 \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha x^2}\,dx = M\left(x^2 - \frac{1}{2\alpha}\right)^2.$$

Выясним смысл параметра α. Так как распределение симметрично относительно нуля и интеграл $\int_{-\infty}^{\infty} x^k p_\xi(x)\,dx$ сходится абсолютно при любом $k \geq 1$, следовательно, $M\xi$ существует и $M\xi = 0$,

$$D\xi = M(\xi - M\xi)^2 = M\xi^2 = \int_{-\infty}^{\infty} x^2\left(\frac{\alpha}{\pi}\right)^{1/2} e^{-\alpha x^2}\,dx =$$

$$=\sqrt{\frac{\alpha}{\pi}}\left[-x\frac{1}{2\alpha}e^{-\alpha x^2}\Big|_{-\infty}^{\infty}+\frac{1}{2\alpha}\int_{-\infty}^{\infty}e^{-\alpha x^2}dx\right]=\frac{1}{2\alpha}.$$

Таким образом, $J=D\xi^2=M(\xi^2-M\xi^2)^2=M\xi^4-(M\xi^2)^2$. Найдем

$$M\xi^4=\int_{-\infty}^{\infty}x^4\sqrt{\frac{\alpha}{\pi}}e^{-\alpha x^2}dx=\sqrt{\frac{\alpha}{\pi}}\left[-x^3\frac{1}{2\alpha}e^{-\alpha x^2}\Big|_{-\infty}^{\infty}+\frac{1}{2\alpha}\int_{-\infty}^{\infty}3x^2e^{-\alpha x^2}dx\right]=$$

$$=\sqrt{\frac{\alpha}{\pi}}\frac{3}{2\alpha}\int_{-\infty}^{\infty}x^2e^{-\alpha x^2}dx=\frac{3}{2\alpha}\int_{-\infty}^{\infty}\sqrt{\frac{\alpha}{\pi}}x^2e^{-\alpha x^2}dx=\frac{3}{2\alpha}\frac{1}{2\alpha}=\frac{3}{4\alpha^2},$$

$$J=\frac{3}{4\alpha^2}-\left(\frac{1}{2\alpha}\right)^2=\frac{2}{4\alpha^2}=\frac{1}{2\alpha^2}.$$

Следовательно, $D\hat{\alpha}=\frac{1}{n}\cdot\frac{1}{2\alpha^2}=\frac{2\alpha^2}{n}$, учитывая, что значение параметра α неизвестно, можем найти только оценку $D\hat{\alpha}$, т.е. $\hat{D}\hat{\alpha}=\frac{2\hat{\alpha}^2}{n}$. Окончательно получаем

$$\hat{D}\hat{\alpha}=\frac{2}{n}n^2\left(2\sum_{i=1}^{n}\xi_i^2\right)^{-2}=\frac{n}{2}\left(\sum_{i=1}^{n}\xi_i^2\right)^{-2}.$$

Ответ

$$\hat{\alpha}=\frac{n}{2}\left(\sum_{i=1}^{n}\xi_i^2\right)^{-1};\ \hat{D}\hat{\alpha}=\frac{n}{2}\left(\sum_{i=1}^{n}\xi_i^2\right)^{-2}.$$

Задача 11

Измерены значения x_1, x_2, ..., x_n случайной величины $\xi(\omega)$, подчиняющейся равномерному распределению на отрезке $[0,\beta]$

$$p(x,\beta)=\begin{cases}0, & \text{если } x\notin[0,\beta];\\ \frac{1}{\beta}, & \text{если } x\in[0,\beta].\end{cases}$$

Найти оценку параметра $\hat{\beta}$ методом максимального правдоподобия. Рассмотреть также случай, когда неизвестны оба параметра равномерного распределения, т.е.

$$p(x,\alpha,\beta)=\begin{cases} 0, & \text{если } x\notin[\alpha,\beta]; \\ \dfrac{1}{\beta-\alpha}, & \text{если } x\in[\alpha,\beta]. \end{cases}$$

Найти оценки $\hat{\alpha}$, $\hat{\beta}$.

Решение

В этой задаче поиск максимума функции правдоподобия не связан с решением уравнения правдоподобия

$$\frac{\partial \ln L(x,\ldots,x_n,\beta)}{\partial\beta}=0.$$

Очевидно, функция правдоподобия здесь $L(x_1,\ldots,x_n,\beta)=\left(\dfrac{1}{\beta}\right)^n$ (или $L(x_1,\ldots,x_n,\alpha,\beta)=\dfrac{1}{(\beta-\alpha)^n}$ — во втором случае, при неизвестных α и β). Точку максимума функции $L(x_1,\ldots,x_n,\beta)$ необходимо искать среди точек β (или (α,β)) из области допустимых значений параметра. Из элементов выборки составим вариационный ряд: $\xi_1^*\le\xi_2^*\le\ldots\le\xi_n^*$. Нетрудно видеть, что для всех $k=1$, 2, …, n имеет место следующее неравенство: $\alpha\le\xi_k^*\le\beta$. Учитывая этот факт, можно утверждать, что $\xi_n^*\le\beta<\infty$, а $\xi_n^*-\xi_1^*\le\beta-\alpha$. Эти неравенства и определяют область допустимых значений параметров распределения. Поэтому максимум функции правдоподобия $L(x_1,\ldots,x_n,\beta)=1/\beta^n$ достигается в точке $\hat{\beta}=\xi_n^*$ (а функции $L(x_1,\ldots,x_n,\alpha,\beta)=\dfrac{1}{(\beta-\alpha)^n}$ в точке $(\hat{\alpha},\hat{\beta})=(\xi_1^*,\xi_n^*)$). Указанные статистики и будут оценками максимального правдоподобия для равномерного распределения.

Ответ

$\hat{\beta} = \xi_n^*$, $(\hat{\alpha}, \hat{\beta}) = (\xi_1^*, \xi_n^*)$ — оценки параметров равномерного распределения, полученные методом максимального правдоподобия.

Задача 12

Дана выборка $(\xi_1, \xi_2, ..., \xi_n)$ из некоторой генеральной совокупности с функцией распределения $F_\xi(x)$. Найти функцию распределения k-й порядковой статистики ξ_k^*.

Решение

Чтобы получить функцию распределения $F_{\xi_k^*}(x)$ k-й порядковой статистики ξ_k^* (для $k = 2, \ldots, n-1$) предположим, что интервал $(-\infty, x)$ разбит на частичные интервалы $(-\infty, S_1)$, $[S_1, S_2)$, ..., $[S_{N-1}, S_N)$ и пусть $\Delta F_\xi(S_k) = F_\xi(S_k) - F_\xi(S_{k-1})$, $S_0 = -\infty$, $S_N = x$. Тогда $P\{\omega : \xi_k^* \in [S_{j-1}, S_j)\}$ равна, с точностью до бесконечно малых величин высшего порядка по сравнению с $\Delta F(S_j)$, вероятности того, что в интервал $(-\infty, S_{j-1})$ попадает $(k-1)$ выборочных значений, в интервал $[S_{j-1}, S_j)$ — одно, а в интервал $[S_j, +\infty)$ — $(n-k)$ выборочных значений. Эта вероятность равна

$$\frac{n!}{(k-1)!\,1!\,(n-k)!}[F_\xi(S_{j-1})]^{k-1}\Delta F_\xi(S_j)[1 - F_\xi(S_j)]^{n-k},$$

(где $\dfrac{n!}{(k-1)!\,1!\,(n-k)!}$ — число способов, с помощью которых n точек могут быть размещены по трем интервалам так, чтобы в первый попало $(k-1)$ точка, во второй — одна, а в третий остальные $(n-k)$ точек). Далее, суммируя эти вероятности по j от 0 до N, и переходя к пределу при измельчении разбиения отрезка интервала $(-\infty, x)$, при котором $\max\limits_j \Delta F(S_j) \to 0$, получим

$$F_{\xi_k^*}(x) = P\{\omega : \xi_k^*(\omega) < x\} =$$

$$= \frac{n!}{(k-1)!\,1!\,(n-k)!} \int\limits_{-\infty}^{x} [F_\xi(S)]^{k-1} [1 - F_\xi(S)]^{n-k} \, dF_\xi(S),$$

или, после замены переменного $F(S) = t$

$$F_{\xi_k^*}(x) = \frac{n!}{(k-1)!\,1!\,(n-k)!} \int\limits_{0}^{F_\xi(x)} t^{k-1}(1-t)^{n-k} \, dt.$$

Замечание. Нетрудно видеть, что для минимального и максимального членов вариационного ряда функцию распределения можно найти, используя свойство независимости элементов выборки и определение функции распределения. Действительно,

$$F_{\xi_n^*}(x) = P\{\omega : \xi_n^*(\omega) < x\} = P\{\omega : \max(\xi_1, \xi_2, ..., \xi_n) < x\} =$$

$$= P\{\omega : (\xi_1 < x) \cap (\xi_2 < x) \cap ... \cap (\xi_n < x)\} = \prod_{i=1}^{n} P\{\omega : (\xi_i(\omega) < x)\} =$$

$$= \prod_{i=1}^{n} F_\xi(x) = [F_\xi(x)]^n.$$

Аналогично для $\xi_1^*(\omega)$:

$$F_{\xi_1^*}(x) = P\{\omega : \xi_1^*(\omega) < x\} = 1 - P\{\xi_1^*(\omega) \ge x\} = 1 - [1 - F_\xi(x)]^n.$$

Ответ

$$F_{\xi_k^*}(x) = \frac{n!}{(k-1)!\,1!\,(n-k)!} \int\limits_{0}^{F_\xi(x)} t^{k-1}(1-t)^{n-k} \, dt, \quad k = \overline{1, n}.$$

Задача 13

В условиях задачи 12 доказать, что вероятность

$$P\{\xi_{k-1}^* \le \xi < \xi_k^*\} = \frac{1}{n+1}, \quad k = 1, 2, \ldots, n$$

(принимая $\xi_0^* = -\infty$, $\xi_{n+1}^* = +\infty$).

Решение

Предположим, что $\xi(\omega)$, $\xi_1(\omega)$, …, $\xi_n(\omega)$ — независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F_\xi(x)$, а $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ — вариационный ряд, построенный по последовательности $\xi_1, \xi_2, \ldots, \xi_n$.

Пусть $G_k(x)$ — функция распределения величины $\xi_k^*(\omega)$. Так как $\xi(\omega)$ не зависит от величин $\xi_k^*, \ldots, \xi_n^*$, то

$$P\{\xi(\omega) < \xi_k^*\} = \iint\limits_{y<x} dG_k(x)\, dF_\xi(y) = \int_{-\infty}^{\infty} dG_k(x) \int_{-\infty}^{x} dF_\xi(y) =$$

$$= \int_{-\infty}^{\infty} dG_k(x)[F_\xi(x) - F_\xi(-\infty)] = \int_{-\infty}^{\infty} F_\xi(x)\, dG_k(x) =$$

$$= \frac{n!}{(k-1)!\,1!\,(n-k)!} \int_{-\infty}^{\infty} F_\xi(x)[F_\xi(x)]^{k-1}[1 - F_\xi(x)]^{n-k}\, dF_\xi(x) =$$

$$= \frac{n!}{(k-1)!\,1!\,(n-k)!} \int_{-\infty}^{\infty} [F_\xi(x)]^{k}[1 - F_\xi(x)]^{n-k}\, dF_\xi(x).$$

После замены переменного $F_\xi(x) = t$ получаем

$$P\{\xi(\omega) < \xi_k^*\} = \frac{n!}{(k-1)!\,1!\,(n-k)!} \int_0^1 t^k (1-t)^{n-k}\, dt.$$

Используя свойства бета-функции $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ (или интегрируя k раз по частям), приходим к следующему результату:

$$P\{\xi(\omega) < \xi_k^*\} = \frac{n!}{(k-1)!\,1!\,(n-k)!} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(k+1+n-k+1)} =$$

$$= \frac{n!\,k!\,(n-k)!}{(k-1)!\,(n-k)!\,(n+1)!} = \frac{k}{n+1}.$$

Следовательно,

$$P\{\xi_{k-1}^* \le \xi(\omega) < \xi_k^*\} = \frac{k}{n+1} - \frac{k-1}{n+1} = \frac{1}{n+1}, \quad k = \overline{1, n},$$

что и требовалось доказать.

Замечание. Важным обстоятельством, вытекающим из этого результата, является то, что найденные вероятности не зависят от функции распределения величины $\xi(\omega)$, т.е. от $F_{\xi}(x)$.

Задача 14

В условиях задачи 12 найти распределение случайного вектора (ξ_1^*, ξ_n^*).

Решение

Для того, чтобы найти совместную функцию распределения вектора (ξ_1^*, ξ_n^*), т.е. $F(x, y) = P\{\omega : (\xi_1^* < x) \bigcap (\xi_n^* < y)\}$ для произвольных $x \in (-\infty, +\infty)$, $y \in (-\infty, +\infty)$ рассмотрим

$$F_{\xi_n^*}(y) = [F_{\xi}(y)]^n = P\{\xi_n^*(\omega) < y\} = P\{\omega : [(\xi_n^*(\omega) < y) \bigcap \Omega]\} =$$

$$= P\{\omega : (\xi_n^* < y) \bigcap [(\xi_1^* < x) \bigcup (\xi_1^* \geq x)]\} =$$

$$= P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* < x)] \bigcup [(\xi_n^* < y) \bigcap (\xi_1^* \geq x)]\} =$$

$$= P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* < x)]\} + P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* \geq x)]\}.$$

Отсюда для любых x и y получаем

$$P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* < x)]\} = [F_{\xi}(y)]^n - P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* \geq x)]\}.$$

Нетрудно видеть, что если $y \leq x$, то

$$P\{\omega [(\xi_n^* < y) \bigcap (\xi_1^* \geq x)]\} = P\{\varnothing\} = 0,$$

а если $y > x$, то

$$P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* \geq x)]\} =$$

$$= P\{\omega : (x \leq \xi_1 < y) \bigcap (x \leq \xi_2 < y) \bigcap \ldots \bigcap (x \leq \xi_n < y)\},$$

т.е. вероятность того, что все элементы выборки находятся в интервале $[x, y)$, а так как $\xi_1, \ldots, \xi_n$ — независимые случайные величины, то эта вероятность равна произведению вероятностей попадания каждой случайной величины $\xi_k(\omega)$ в интервал $[x, y)$, т.е.

$P\{x \le \xi_k(\omega) < y\} = F_\xi(y) - F_\xi(x)$. Таким образом, получаем при $y > x$

$$P\{\omega : [(\xi_n^* < y) \bigcap (\xi_1^* \ge x)]\} = [F_\xi(y) - F_\xi(x)]^n .$$

Окончательно имеем

$$F(x, y) = P\{\omega : (\xi_1^* < x) \bigcap (\xi_n^* < y)\} =$$

$$= \begin{cases} [F_\xi(y)]^n & \text{при } y \le x; \\ [F_\xi(y)]^n - [F_\xi(y) - F_\xi(x)]^n & \text{при } y > x. \end{cases}$$

Ответ

$$F(x, y) = \begin{cases} [F_\xi(y)]^n & \text{при } y \le x; \\ [F_\xi(y)]^n - [F_\xi(y) - F_\xi(x)]^n & \text{при } y > x. \end{cases}$$

Задача 15

В условиях задачи 11 найти методом максимального правдоподобия оценку $M\xi$. Является ли эта оценка несмещенной?

Решение

В задаче 11 данного раздела методом максимального правдоподобия были найдены оценки для параметров α и β, а именно $\hat{\alpha} = \xi_1^*$, $\hat{\beta} = \xi_n^*$, где ξ_i^* — i-я порядковая статистика. Так как для равномерного распределения на отрезке $[\alpha, \beta]$ математическое ожидание случайной величины $\xi(\omega)$ будет равно $M\xi = m = \dfrac{\alpha + \beta}{2}$, где α и β — параметры распределения, то оценкой для $M\xi$ в этом случае будет статистика $\hat{m} = \dfrac{\hat{\alpha} + \hat{\beta}}{2} = \dfrac{\xi_1^* + \xi_n^*}{2}$.

Оценка $\hat{\theta}$ называется несмещенной, если $M\hat{\theta} = 0$, следовательно, для решения задачи необходимо найти

$$M\hat{m} = M(\xi_1^* + \xi_n^*)/2 = \frac{1}{2}(M\xi_1^* + M\xi_n^*).$$

Найдем плотность распределения минимального и максимального членов вариационного ряда $p_1(x)$ и $p_2(x)$, используя результаты, полученные при решении задачи 12:

$$p_1(x) = \frac{dF_{\xi_1^*}(x)}{dx} = \frac{d[1-(1-F_\xi(x))^n]}{dx} = n[1-F_\xi(x)]^{n-1} p_\xi(x),$$

$$p_n(y) = \frac{dF_{\xi_n^*}(y)}{dy} = \frac{d[F_\xi(y)]^n}{dy^n} = n[F_\xi(y)]^n p_\xi(y),$$

где $p_\xi(x)$ — плотность распределения случайной величины $\xi(\omega)$. В нашем случае

$$p_\xi(t) = \begin{cases} \dfrac{1}{\beta-\alpha}, & \text{если } t \in [\alpha, \beta]; \\ 0, & \text{если } t \notin [\alpha, \beta] \end{cases} \quad \text{и} \quad F_\xi(t) = \begin{cases} 0, & \text{если } t \le \alpha; \\ \dfrac{t-\alpha}{\beta-\alpha}, & \text{если } \alpha < t \le \beta; \\ 1, & \text{если } t > \beta; \end{cases}$$

$$M\xi_1^* = \int_{-\infty}^{\infty} x p_1(x)\,dx, \quad M\xi_n^* = \int_{-\infty}^{\infty} y p_n(y)\,dy,$$

$$M\xi_1^* = \int_{-\infty}^{\infty} xn[1-F_\xi(x)]^{n-1} p_\xi(x)\,dx =$$

$$= \int_\alpha^\beta xn\left(1-\frac{x-\alpha}{\beta-\alpha}\right)^{n-1} \frac{dx}{\beta-\alpha} = \frac{n}{(\beta-\alpha)^n}\int_\alpha^\beta x(\beta-x)^{n-1}\,dx.$$

После замены $\beta - x = t$ получаем

$$M\xi_1^* = \frac{n}{(\beta-\alpha)^n}\int_0^{\beta-\alpha} t^{n-1}(\beta-t)\,dt = \frac{n}{(\beta-\alpha)^n}\int_0^{\beta-\alpha}(\beta t^{n-1} - t^n)\,dt = \frac{\beta+n\alpha}{n+1},$$

$$M\xi_n^* = \int_\alpha^\beta y\left(\frac{y-\alpha}{\beta-\alpha}\right)^{n-1} n\frac{dy}{\beta-\alpha} = \frac{n}{(\beta-\alpha)^n}\int_\alpha^\beta y(y-\alpha)^{n-1}\,dy.$$

После замены $y-\alpha=t$ получаем

$$M\xi_n^* = \frac{n}{(\beta-\alpha)^n}\int_0^{\beta-\alpha}(t+\alpha)t^{n-1}dt =$$

$$= \frac{n}{(\beta-\alpha)^n}\int_0^{\beta-\alpha}(t^n+\alpha t^{n-1})\,dt = \frac{\alpha+n\beta}{n+1},$$

$$M\hat{m} = \frac{1}{2}(M\xi_1^* + M\xi_n^*) = \frac{1}{2}\left(\frac{\beta+n\alpha}{n+1} + \frac{\alpha+n\beta}{n+1}\right) = \frac{\alpha+\beta}{2} = M\xi,$$

т.е. оценка математического ожидания $\hat{m}$ — несмещенная.

Замечание. В ходе решения было выяснено, что оценки $\hat{\alpha}=\xi_1^*$ и $\hat{\beta}=\xi_n^*$ являются смещенными, так как $M\hat{\alpha} = M\xi_1^* = \dfrac{\beta+n\alpha}{n+1} \neq \alpha$ и $M\beta_n = M\xi_n^* = \dfrac{\alpha+n\beta}{n+1} \neq \beta$.

Ответ

$\hat{M\xi} = \hat{m} = \dfrac{1}{2}(\xi_1^* + \xi_n^*)$. Оценка $\hat{m}$ является несмещенной, так как $M\hat{m} = \dfrac{1}{2}(\alpha+\beta) = M\xi$.

Задача 16

В условиях задачи 11 найти методом моментов оценки параметров $\hat{\alpha}$ и $\hat{\beta}$.

Решение

Для нахождения оценок двух параметров равномерного распределения необходимо выбрать две линейно независимые функции $g_1(x)$ и $g_2(x)$. В качестве этих функций возьмем $g_1(x)=x$ и $g_2(x)=x^2$.

Тогда, применяя метод моментов, найдем теоретические моменты:

$$m_1 = Mg_1(\xi) = M\xi = \int_{-\infty}^{\infty} x p_\xi(x)\,dx = \int_\alpha^\beta x\frac{1}{\beta-\alpha}\,dx = \frac{\alpha+\beta}{2},$$

$$m_2 = Mg_2(\xi) = M\xi^2 = \int_{-\infty}^{\infty} x^2 p_\xi(x)\,dx = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta-\alpha}\,dx = \frac{\beta^2+\alpha\beta+\alpha^2}{3}.$$

Метод подстановки дает следующие оценки для этих моментов $\hat{m}_1 = \frac{1}{n}\sum_{k=1}^{n}\xi_k$, $\hat{m}_2 = \frac{1}{n}\sum_{k=1}^{n}\xi_k^2$ (так как m_l является функционалом от функции распределения $F_\xi(x)$, $m_l = \int_{-\infty}^{\infty} x^l dF_\xi(x)$, то в качестве оценки для $\hat{m}_l$ выбираем такой же функционал от эмпирической функции распределения, т.е. $\hat{m}_l = \int_{-\infty}^{\infty} x^l d\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n}\xi_k^l$). Приравнивая теоретические моменты, зависящие от параметров, эмпирическим моментам, полученным методом подстановки, имеем систему:

$$\begin{cases} \hat{m}_1 = \frac{1}{n}\sum_{k=1}^{n}\xi_k = \frac{\alpha+\beta}{2}; \\ \hat{m}_2 = \frac{1}{n}\sum_{k=1}^{n}\xi_k^2 = \frac{\alpha^2+\alpha\beta+\beta^2}{3}, \end{cases}$$

α и β, найденные как решение данной системы, будут функциями от $\hat{m}_1$ и $\hat{m}_2$, т.е. статистиками, которые и являются оценками параметров $\hat{\alpha}$ и $\hat{\beta}$, полученными методом моментов.

Решим систему:

$$\begin{cases} \alpha = 2\hat{m}_1 - \beta; \\ \beta^2 - 2\hat{m}_1\beta + 4\hat{m}_1^2 - 3m_2 = 0, \end{cases}$$

$\beta = \hat{m}_1 \pm \sqrt{3}\sqrt{\hat{m}_2 - (\hat{m}_1)^2}$, а так как $\alpha < \beta$, то решение единственное

$$\hat{\alpha} = \hat{m}_1 - \sqrt{3}\sqrt{\hat{m}_2 - (\hat{m}_1)^2} ,$$

$$\hat{\beta} = \hat{m}_1 + \sqrt{3}\sqrt{\hat{m}_2 - (\hat{m})^2} .$$

Замечание. Если один из параметров известен (например, $\alpha = 0$), то для применения метода моментов достаточно взять одну функцию $g_1(x) = x$.

Тогда $\hat{m}_1 = \frac{1}{n}\sum_{k=1}^{n}\xi_k = \frac{\beta}{2}$ и оценкой параметра β будет $\hat{\beta} = 2\hat{m}_1$.

Ответ

$\hat{\alpha} = \hat{m}_1 - \sqrt{3(\hat{m}_2 - (\hat{m}_1)^2)}$, $\qquad \hat{\beta} = \hat{m}_1 + \sqrt{3(\hat{m}_2 - (\hat{m}_1)^2)}$, где $\hat{m}_1 = \frac{1}{n}\sum_{k=1}^{n}\xi_k$, $\hat{m}_2 = \frac{1}{n}\sum_{k=1}^{n}\xi_k^2$.

Задача 17

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — выборка из нормально распределенной генеральной совокупности с параметрами (a, σ^2). Доказать, что оценки $\hat{a} = \frac{1}{n}\sum_{k=1}^{n}\xi_k$ и $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \hat{a})^2$ независимы, а случайная величина $n\hat{\sigma}_n^2 / \sigma^2$ имеет χ-квадрат распределение с $(n-1)$ степенью свободы.

Решение

См. [1], с. 184.

Задача 18

Дана выборка объема $n = 100$ из некоторой генеральной совокупности. После группировки данных получена табл. 3.

Таблица 3

Интервал	(–15, –10)	(–10, –5)	(–5, 0)	(0, 5)	(5, 10)	(10, 15)
Число вариант, попавших в интервал	3	12	35	37	11	2

Проверить гипотезу о: а) нормальном распределении генеральной совокупности; б) не превышении измеряемой случайной величины (генеральной совокупности) по абсолютной величине 5.

Решение

а) Плотность нормально распределенной величины

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}, \quad -\infty<x<+\infty,$$

зависит от двух параметров. Используя метод максимального правдоподобия для отыскания оценок этих параметров, получаем $\hat{a}=\frac{1}{n}\sum_{k=1}^{n}\xi_k$, $\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(\xi_k-\hat{a})^2$, где ξ_k — варианты (элементы выборки). Так как в условии задачи приведена таблица сгруппированных данных выборки объема $n = 100$, то для вычисления оценок $\hat{a}$ и $\hat{\sigma}^2$ используем следующие формулы:

$$\hat{a}=\frac{1}{n}\sum_{k=1}^{r}\eta_k\nu_k, \quad \hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{r}(\eta_k-\hat{a})^2\nu_k,$$

где r — число интервалов после группировки; ν_k — число вариант, попавших в k-й интервал; η_k — варианта, равная середине k-го интервала. В нашем случае $r = 6$.

Таблица 4

η_k	–12,5	–7,5	–2,5	2,5	7,5	12,5
ν_k	3	12	35	37	11	2

После вычислений находим $\hat{a}=-0{,}15$; $\hat{\sigma}^2=25{,}2275$. Учитывая статистическую ошибку, возьмем для проверки гипотезу H_0 : генеральная совокупность имеет нормальное распределение с параметрами $a = 0$ и $\sigma^2 = 25$.

Для проверки этой гипотезы используем критерий Пирсона (хи-квадрат критерий) (см. [1], с. 207). Согласно этому критерию случайная величина $\chi_m^2 = \sum_{k=1}^{r} \frac{(\nu_k - np_k)^2}{np_k}$ при справедливости гипотезы H_0 будет иметь хи-квадрат распределение с m степенями свободы, где $m = r - 1 - S$, где r — число интервалов после группировки; S — число параметров проверяемого распределения, вычисленных по выборке (в нашем случае $r = 6$, $S = 2$, а $m = 6 - 1 - 2 = 3$); p_k — теоретические вероятности попадания генеральной совокупности в k-й интервал, а именно

$$p_k = P\{\omega : \xi(\omega) \in \Delta_k\} = \int_{\Delta_k} p(x)\, dx$$

(в нашем случае $p_k = \int_{\Delta_k} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}5} \int_{\Delta_k} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$),

ν_k — число вариант выборки, попавших в k-й интервал. Проверим гипотезу H_0 на уровне значимости $\alpha = 0{,}05$ (т.е. с заданной ошибкой первого рода α). Найдем границы критической области S_c, исходя из определения $P_{H_0}\{\chi_m^2 \in S_c\} = \alpha$ (вероятность попадания критерия χ_m^2 в критическую область, при условии, что гипотеза H_0 верна, равна ошибке 1-го рода α).

Используя таблицу χ^2-распределения (см. табл. П5 на с. 101 — 102 данных методических указаний), находим $S_c = (7{,}81; +\infty)$, $(P\{\chi_3^2 > 7{,}81\} = 0{,}05)$.

Найдем теперь наблюдаемое значение критерия χ_m^2(набл.). Для этого подсчитаем теоретические вероятности p_k, используя свойства нормально распределенных случайных величин и таблицу значений функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$. Так как

$$p_k = P\{\omega : \xi(\omega) \in \Delta_k\} = P\{a_k \le \xi(\omega) < b_k\} =$$

$$= P\left\{\frac{a_k - 0}{5} < \frac{\xi(\omega) - 0}{5} < \frac{b_k - 0}{5}\right\} = P\left\{\frac{a_k}{5} \le \xi_0 < \frac{b_k}{5}\right\} = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{a_k}{5}}^{\frac{b_k}{5}} \mathrm{e}^{-\frac{t^2}{2}} dt,$$

$$p_1 = P\{\xi(\omega) \in [-15, -10)\} = P\left\{\frac{-15}{5} \le \xi_0 < \frac{-10}{5}\right\} = \Phi_0(3) - \Phi_0(2) =$$

$$= 0{,}4986 - 0{,}4772 = 0{,}0214,$$

$$p_2 = P\{\xi(\omega) \in [-10, -5)\} = P\left\{\frac{-10}{5} \le \xi_0 < \frac{-5}{5}\right\} = \Phi_0(2) - \Phi_0(1) =$$

$$= 0{,}4772 - 0{,}3413 = 0{,}1359,$$

$$p_3 = P\{\xi(\omega) \in [-5, 0)\} = P\left\{\frac{-5}{5} \le \xi_0 < \frac{0}{5}\right\} = \Phi_0(1) = 0{,}3413,$$

$$p_4 = P\{\xi(\omega) \in [0, 5)\} = P\left\{\frac{0}{5} \le \xi_0 < \frac{5}{5}\right\} = \Phi_0(1) = 0{,}3413,$$

$$p_5 = P\{\xi(\omega) \in [5, 10)\} = P\left\{\frac{5}{5} \le \xi_0 < \frac{10}{5}\right\} = \Phi_0(2) - \Phi_0(1) =$$

$$= 0{,}4772 - 0{,}3413 = 0{,}1359,$$

$$p_6 = P\{\xi(\omega) \in [10, 15)\} = P\left\{\frac{10}{5} \le \xi_0 < \frac{15}{5}\right\} = \Phi_0(3)\Phi_0(2) =$$

$$= 0{,}4986 - 0{,}4772 = 0{,}0214,$$

$$\chi_m^2(\text{набл.}) = \sum_{k=1}^{6} \frac{(\nu_k - np_k)^2}{np_k} = \frac{(3 - n \cdot 0{,}0214)^2}{n \cdot 0{,}0214} +$$

$$+ \frac{(12 - n \cdot 0{,}1359)^2}{n \cdot 0{,}1359} + \frac{(35 - n \cdot 0{,}3413)^2}{n \cdot 0{,}3413} + \frac{(37 - n \cdot 0{,}3413)^2}{n \cdot 0{,}3413} +$$

$$+ \frac{(11 - n \cdot 0{,}1359)^2}{n \cdot 0{,}1359} + \frac{(2 - n \cdot 0{,}0214)^2}{n \cdot 0{,}0214} = 1{,}2979$$

(n =100 — объем выборки).

Поскольку $1,2979 < 7,81$, т.е. наблюдаемое значение критерия попадет в $\overline{S}_c$ — область допустимых значений критерия, гипотеза H_0 не отвергается (заметим, что гипотеза H_0 не будет отвергнута и на уровнях значимости $\alpha_1 = 0,1$ и $\alpha_2 = 0,2$; в этих случаях критическая область $S_c^1 = (6,25; +\infty)$ и $S_c^2 = (4,64; +\infty)$ соответственно, и наблюдаемое значение критерия в обоих случаях будет находиться в области допустимых значений).

б) Проверим гипотезу H_0: значения случайной величины не превышают по абсолютной величине число 5 (т.е. $P\{\omega : |\xi(\omega)| \le 5\} = 1$).

Для проверки этой гипотезы критерий Пирсона не применим, хотя бы потому, что для проверяемой случайной величины $P\{|\xi| \le 5\} = 1$, $P\{|\xi| > 5\} = 0$, и теоретические частоты p_k равны нулю в этом случае для Δ_1, Δ_2, Δ_5 и Δ_6.

Используем в этом случае критерий Колмогорова — Смирнова. Согласно этому критерию случайная величина

$$D_n = \sqrt{n} \sup_{-\infty<x<\infty} \left| F_n^*(x) - F(x) \right|$$

в случае справедливости гипотезы H_0 имеет распределение Колмогорова (здесь $F(x)$ — теоретическая (предполагаемая) функция распределения, а $F_n^*(x)$ — эмпирическая функция распределения).

Так как в условиях задачи данные уже сгруппированы, то не возможно найти $F_n^*(x)$, используя весь массив данных, т.е. определяя

$$F_n^*(x) = \frac{1}{n} \sum_{k=1}^{n} \eta_k(x), \quad \text{где} \quad \eta_k(x) = \begin{cases} 1, & \text{если } \xi_k(\omega) < x; \\ 0, & \text{если } \xi_k(\omega) \ge x; \end{cases}$$

$\xi_k(\omega)$ — элементы не сгруппированной выборки. Но согласно выбранному критерию, надо найти наблюдаемое значение критерия D_n(набл.), т.е. найти максимум отклонения $F_n^*(x)$ и $F(x)$, или хотя бы оценить этот максимум. Для этих целей и найдем оценку функции распределения генеральной совокупности $\hat{F}_n^*(x)$, используя оценку плотности распределения $\hat{p}_n(x)$ (гистограмму распределения). Иными словами, сначала построим гистограмму распределения, а затем восстановим по ней функцию распределения. В качестве оценки плотности возьмем

$$\hat{p}_n(x) = \frac{\nu_k}{n}\frac{1}{|\Delta_k|} \quad \text{при} \quad x \in \Delta_k$$

(где $n = 100$ — объем выборки; ν_k — число вариант, попавших в k-й интервал Δ_k, а $|\Delta_k|$ — длина этого интервала, $x \in (-\infty, \infty)$), $\hat{F}_n^*(x) = \int\limits_{-\infty}^{x} \hat{p}_n(x)\,dx$ — оценка функции распределения. Используя условия задачи, получаем

$$\hat{F}_n^*(x) = \begin{cases} 0 & \text{при } x \le -15; \\ 0{,}03 \cdot \dfrac{x+15}{5} & \text{при } -15 < x \le -10; \\ 0{,}03 + 0{,}12 \cdot \dfrac{x+10}{5} & \text{при } -10 < x \le -5; \\ 0{,}03 + 0{,}12 + 0{,}35 \cdot \dfrac{x+5}{5} = 0{,}15 + 0{,}35 \cdot \dfrac{x+5}{5} & \text{при } -5 < x \le 0; \\ 0{,}15 + 0{,}35 + 0{,}37 \cdot \dfrac{x-0}{5} = 0{,}5 + 0{,}37 \cdot \dfrac{x}{5} & \text{при } 0 < x \le 5; \\ 0{,}5 + 0{,}37 + 0{,}11 \cdot \dfrac{x-5}{5} = 0{,}87 + 0{,}11 \cdot \dfrac{x-5}{5} & \text{при } 5 < x \le 10; \\ 0{,}87 + 0{,}11 + 0{,}02 \cdot \dfrac{x-10}{5} = 0{,}98 + 0{,}02 \cdot \dfrac{x-10}{5} & \text{при } 10 < x \le 15; \\ 1 & \text{при } x > 15. \end{cases}$$

Выбираем уровень значимости критерия $\alpha = 0{,}05$ (ошибка 1-го рода) и находим критическую область S_c таким образом, чтобы $P_{H_0}(D_n \in S_c) = \alpha$. Используя таблицу распределения Колмогорова — Смирнова (см. табл. П7 на с. 103 данных методических указаний), находим $S_c = (1{,}34; +\infty)$, т.е. если гипотеза H_0 верна, то

$$P\left\{\sqrt{n}\sup_{-\infty<x<\infty}\left|F_n^*(x)-F(x)\right|>1{,}34\right\}=0{,}05$$

(это событие является маловероятным, поэтому, отвергнув верную гипотезу H_0, можем совершить ошибку, вероятность которой не более 0,05). Проверим гипотезу H_0. Для этого найдем наблюдаемое значение критерия D_n(набл.). Мы не знаем истинное распределение случайной величины, так как в гипотезе H_0 утверждается только тот факт, что $P\{\omega:|\xi(\omega)\leq 5|\}=1$. Поэтому можем точно записать только то, что теоретическая функция распределения $F(x)=0$ для любого $x\leq -5$ и $F(x)=1$ для любого $x>5$. Сравнивая на этих интервалах значения функции $F(x)$ с $\hat{F}_n^*(x)$, получаем

$$\left|\hat{F}_n^*(-5)-F(-5)\right|=\left|0{,}03+0{,}12\cdot\frac{-5+10}{5}-0\right|=0{,}15,$$

для любых $x<-5$ отклонение меньше;

$$\left|\hat{F}_n^*(5+0)-F(5+0)\right|=\left|0{,}87+0{,}11\cdot\frac{(5+0)-5}{5}-1\right|=0{,}13,$$

для любых $x>5$ отклонение меньше, т.е. можем гарантировать, что

$$\sup_{-\infty<x<\infty}\left|F_n^*(x)-F(x)\right|\geq\sup_{-\infty<x<+\infty}\left|\hat{F}_n^*(x)-F(x)\right|\geq 0{,}15.$$

Таким образом, наблюдаемое значение критерия

$$D_n(\text{набл.})=\sqrt{n}\sup_{-\infty<x<\infty}\left|F_n^*(x)-F(x)\right|\geq\sqrt{100}\cdot 0{,}15=1{,}5$$

(это минимальное возможное значение D_n(набл.)), а так как уже 1,5 > 1,34, следовательно, D(набл.)$\in S_c$, и гипотеза H_0 отвергается на уровне значимости $\alpha = 0{,}05$.

Ответ

а) Гипотеза H_0 не отвергается при $\alpha = 0{,}05$, $\alpha = 0{,}1$, $\alpha = 0{,}2$;

б) Гипотеза H_0 отвергается при $\alpha = 0{,}05$.

ДОПОЛНИТЕЛЬНЫЕ ЗАДАЧИ

1. Найти законы распределения, которым соответствуют следующие производящие функции:

а) $\frac{1}{4}(1+z)^2$; б) $e^{\lambda(z-1)}$;

в) $\frac{1}{2}\left(1-\frac{1}{2}z\right)^{-1}$; г) $\left(\frac{1}{3}+\frac{2}{3}z\right)^n$.

2. Пусть ξ — целочисленная случайная величина с производящей функцией $\varphi(z)$. Найти производящие функции: величины $\xi+1$; 2ξ, последовательностей $P\{\xi \le n\}$, $P\{\xi = 2n\}$.

3. Найти характеристическую функцию случайной величины ξ, если:

а) ξ равномерно распределена на отрезке $[-a, a]$, $a > 0$;

б) ξ нормально распределена с параметрами $M\xi = 8$ и $D\xi = 4$;

в) ξ имеет биномиальное распределение.

4. Найти распределение случайной величины ξ, если ее характеристическая функция равна:

а) $f_\xi(t) = \left(\frac{\sin 2t}{2t}\right)^2$;

б) $f_\xi(t) = \frac{1}{2}\cos 2t + \frac{1}{4}\cos 4t + \frac{1}{8}\cos t + \frac{1}{8}$.

5. Для некоторого автопарка среднее число автобусов, отправляемых в ремонт после месяца эксплуатации на городских линиях, равно 5. Оценить вероятность события A = {по истечению месяца в данном автобусном парке будет отправлено в ремонт меньше 15 автобусов}, если информация о дисперсии отсутствует.

Оценить вероятность события A, если дисперсия равна 4.

6. Устройство состоит из 10 независимо работающих элементов. Вероятность отказа каждого элемента за время t равна 0,05. Используя неравенство Чебышева, оценить вероятность того, что абсолютная величина разности между числом отказов элементов и средним числом отказов за время окажется меньше 2.

7. Длина изготавливаемых изделий представляет случайную величину, среднее значение которой равно 90 см, дисперсия этой величины равна 0,0225 см^2. Используя неравенство Чебышева, оценить вероятность того, что:

а) отклонение длины изготавливаемого изделия от ее среднего значения по абсолютной величине не превзойдет 0,4;

б) длина изделия выразится числом, заключенным между 89,7 и 90,3.

8. Последовательность $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ состоит из независимых, одинаково распределенных случайных величин $M\xi_n = 0$, $D\xi_n = \sigma^2 < \infty$, $n = 1, 2, \ldots$. Пусть $\eta_n = \xi_n + \xi_{n+1}$, $n = 1, 2, \ldots$. Удовлетворяет ли закону больших чисел последовательность случайных величин $\eta_1, \eta_2, \ldots$?

9. $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — последовательность независимых случайных величин, $M\xi_n = a$, $D\xi_n = \sigma^2 < \infty$, $n = 1, 2, \ldots$. Пусть $\eta_n = \sum\limits_{1 \le i < j \le n} \xi_i \xi_j$.

Показать, что последовательность $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ удовлетворяет закону больших чисел: для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{ \left| \frac{\eta_n}{C_n^2} - a^2 \right| \ge \varepsilon \right\} = 0$$

10. Пусть $\xi_n \xrightarrow{P} a$, $\eta_n \xrightarrow{P} b$ при $n \to \infty$, где a и b — постоянные числа, а функция $f(x, y)$ непрерывна в точке $x = a$, $y = b$. Доказать, что $f(\xi_n, \eta_n) \xrightarrow{P} f(a, b)$, $n \to \infty$.

11. Игральная кость подбрасывается до тех пор, пока сумма выпавших очков не превысит 700. Оценить вероятность того, что для этого потребуется более 210 бросаний, менее 180 бросаний, от 190 до 210 бросаний.

12. На улице стоит человек и продает газеты. Предположим, что каждый из проходящих мимо людей покупает газету с вероятностью 1/3. Пусть случайная величина ξ — число людей, прошедших мимо продавца за время, пока он продавал первые 100 экземпляров газеты. Найти (приближенно) распределение ξ.

13. Игрок выигрывает 7 копеек при появлении 6 очков на игральной кости и платит одну копейку в других случаях. Найти вероятность того, что его выигрыш будет составлять от 16 до 30 руб. после 8000 бросаний игральной кости.

14. В одном из экспериментов Пирсона по моделированию на вычислительной машине опытов с «подбрасыванием» монеты герб выпал 12012 раз, число «подбрасываний» равно 24000.

Какова вероятность получить: а) данный результат; б) при повторении эксперимента такое же или большее отклонение относительной частоты успехов от вероятности успеха в одном опыте?

15. Известно, что вероятность рождения мальчика приблизительно равна 0,515. Какова вероятность того, что среди 10 тысяч новорожденных мальчиков будет не более, чем девочек?

16. Работница обслуживает 800 веретен. Вероятность обрыва пряжи на каждом из веретен в течение некоторого промежутка времени t равна 0,005. Найти: а) наиболее вероятное число обрывов и его вероятность; б) вероятность того, что в течение времени t произойдет не более 5 обрывов.

17. Игральная кость брошена 105 раз. Найти вероятность того, что сумма выпавших очков заключена на интервале (315, 420).

18. Среди семян пшеницы 0,5 % семян сорняков. Какова вероятность при случайном отборе 1000 семян обнаружить: а) не менее 3 семян сорняков; б) не более 16 семян сорняков; в) ровно 6 семян сорняков? г) какое число семян наиболее вероятно и найти вероятность такого числа семян.

19. Вероятность наступления события A равна p при каждом из n испытаний.

а) Определить вероятность того, что относительная частота наступления события A при $n = 1500$ отклонится от $p = 0{,}4$ в ту или другую сторону меньше, чем на 0,02.

б) Определить вероятность того, что при $n = 1500$, $p = 0{,}4$ событие A наступит число раз, заключенное между 1) 570 и 630, 2) 600 и 660, 3) 620 и 680, 4) 580 и 640.

в) В каких границах находится та относительная частота события A при $n = 1200$, вероятность отклонения которой от $p = 2/3$ равна 0,985?

г) Сколько необходимо произвести испытаний, чтобы вероятность того, что отклонение относительной частоты от $p = 3/8$ в ту или другую сторону будет меньше, чем 0,001, была равна 0,995?

20. Отдел технического контроля проверяет качество наудачу отобранных 900 деталей. Вероятность того, что каждая деталь стандартна равна 0,9. Случайная величина ξ — число стандартных деталей в партии. Найти наименьший интервал, симметричный относительно $M\xi$, в который с вероятностью не меньшей 0,9544, будет заключено число стандартных деталей.

21. 500 раз брошена игральная кость. Какова вероятность того, что частота выпадения шестерки окажется в интервале $\left(\frac{1}{6} - 0{,}05; \frac{1}{6} + 0{,}05\right)$?

22. Игральная кость брошена 315 раз. Какое отклонение средней арифметической суммы выпавших очков от 3,5 следует ожидать с вероятностью, равной 0,998?

23. Есть две игральные кости, одна — «правильная» с $p_k = 1/6$, $k = \overline{1,6}$; другая — «неправильная» с $p_1 = 2/3$, $p_k = 1/15$, $k = \overline{2,6}$. Взята одна кость. Используя неравенство Чебышева, определить, сколько раз надо бросить взятую кость, чтобы с вероятностью, большей 0,9, понять правильная она или неправильная?

24. Сколько раз надо бросить игральную кость, чтобы с вероятностью 0,98 средняя арифметическая сумма выпавших очков отклонилась от 3,5 на $\Delta = 0{,}08$?

25. Имеется выборка объема $n = 9$

$$7,\ 6,\ 5,\ 11,\ 4,\ 7,\ 8,\ 6,\ 9. \qquad (20)$$

Построить по этой выборке эмпирическую функцию распределения. Найти оценку математического ожидания (выборочное среднее) и несмещенную оценку для дисперсии.

26. Предполагая, что выборка (20) есть выборка из элементов геометрического распределения с неизвестным параметром $\theta = p$, найти оценку максимального правдоподобия для θ.

27. Отказ прибора произошел при k-м испытании. Найти оценку максимального правдоподобия для вероятности отказа при каждом испытании и вычислить ее среднее значение.

28. Имеется выборка x_1, $x_2, \ldots,$ x_n из элементов гамма-распределения

$$p(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x > 0$$

($p(x) = 0$, $x \le 0$) с параметром $a > 0$ и неизвестном параметре $\lambda > 0$. Методом максимального правдоподобия найти оценку параметра λ. Показать, что эта оценка имеет положительное смещение.

В задачах 29 — 30 предполагается, что выборка объема получена из генеральной совокупности, имеющей распределение $N(a, \sigma^2)$.

29. Найти с доверительной вероятностью $\wp_1 = 0{,}9$, $\wp_2 = 0{,}99$ доверительные интервалы для математического ожидания a для следующих характеристик:

1) емкости конденсатора, если $\overline{X} = 20$ мкФ, $n = 16$, σ^2 известно, $\sigma = 4$ мкФ;

2) времени безотказной работы электронной лампы, если $\overline{X} = 500$ ч, $n = 100$, σ^2 известно, $\sigma = 10$ ч;

3) диаметра вала, если $\overline{X} = 30$, $n = 9$, $S^2 = 9$ мм^2;

4) содержание углерода в единице продукта, если $\overline{X} = 18$ г, $n = 25$, $S^2 = 16$ г^2.

30. Найти с доверительной вероятностью $\wp_1$ и $\wp_2$ доверительные интервалы для дисперсии следующих характеристик:

1) диаметра вала, если $n = 16$, $\wp_1 = 0{,}9$ и $\wp_2 = 0{,}95$

$$\frac{1}{4}\sum_{k=1}^{n}(x_k - a)^2 = 4{,}5 \text{ мм}^2,$$

a известно;

2) содержание углерода в единице продукта, если a известно,

$$\frac{1}{4}\sum_{k=1}^{n}(x_k - a)^2 = 20 \text{ г}^2,\ n = 9,\ \wp_1 = 0{,}9\ ,\ \wp_2 = 0{,}99\ .$$

31. Полагая, что выборка (20) является выборкой из совокупности нормально распределенных случайных величин с параметрами $\theta_1 = a$ и $\theta_2 = \sigma^2$, построить доверительный интервал для $\theta_2 = \sigma^2$ при неизвестном параметре $\theta_1 = a$ с уровнем доверия $\wp_1 = 0{,}9$ и $\wp_2 = 0{,}99$.

32. Среднее число сбоев в сутки для 100 ЭВМ одного типа равно 2,3. В предположении, что число сбоев имеет распределение Пуассона с параметром $\theta = \lambda$, приближенно найти доверительный интервал для θ с уровнем доверия 0,95.

33. Из урны, содержащей неотличимые на ощупь черные и белые шары в неизвестной пропорции, случайным образом извлекается 100 шаров (с возвращением). Найти с уровнем доверия $\wp_1 = 0{,}9$ и $\wp_2 = 0{,}95$ доверительный интервал для p — доли черных шаров, если среди вынутых оказалось 30 черных шаров.

34. В семи испытаниях Бернулли событие A произошло 3 раза. Используя табл. П4 на с. 100 данных методических указаний, найти с вероятностью $\wp = 0{,}95$ доверительный интервал для вероятности p наступления события A в каждом испытании.

35. В десяти испытаниях Бернулли событие A произошло 3 раза. Используя табл. П4 на с. 100 данных методических указаний, проверить гипотезу $H_0:\ p_0 = 0{,}3$ против альтернативы H_1 с уровнем значимости α:

а) $H_1:\ p > 0{,}3\ ,\ \alpha = 0{,}025$;

б) $H_1:\ p < 0{,}3\ ,\ \alpha = 0{,}025$;

в) $H_1:\ p \neq 0{,}3\ ,\ \alpha = 0{,}05$.

36. В условиях задачи 33 проверить гипотезу H_0 : в урне содержится $m = 20$ черных шаров при альтернативной гипотезе H_1: $m \neq 20$, уровень значимости гипотезы $\alpha_1 = 0{,}05$; $\alpha_2 = 0{,}01$.

37. Рассматривая выборку в задаче 25 как реализацию выборки нормально распределенных величин с параметрами a и σ^2, σ^2 неизвестно, проверить с ошибкой первого рода α гипотезу H_0 против альтернативы H_1 в следующих случаях:

1) H_0: $a = a_0 = 6$; H_1: $a = a_1 > 6$; $\alpha = 0{,}01$;
2) H_0: $a = a_0 = 8$; H_1: $a = a_1 < 8$; $\alpha = 0{,}01$;
3) H_0: $a = a_0 = 7$; H_1: $a = a_1 \neq 7$; $\alpha = 0{,}005$.

38. Пусть $x_1^* \le x_2^* \le \ldots \le x_n^*$ — вариационный ряд, построенный по выборке $x_1, x_2, \ldots, x_n$.

1. Положим

$$\hat{\theta}_1 = \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad \hat{\theta}_2 = \frac{x_1^* + x_n^*}{n}.$$

Найти $M\hat{\theta}_l$, $D\hat{\theta}_l$, $l = 1, 2$, если: а) случайная величина ξ равномерно распределена на отрезке $[a, b]$; б) $p_\xi(x) = \alpha \mathrm{e}^{-\alpha x}$, $x > 0$, $\alpha > 0$.

2. Пусть $p_\xi(x) = \mathrm{e}^{C-x}$, $x \ge C > 0$. Положим $\hat{C} = x_1^* - \frac{1}{n}$. Найти $M\hat{C}$ и $D\hat{C}$.

39. Имеются экспериментально измеренные значения $x_1, \ldots, x_n$ случайной величины ξ, которая подчиняется нормальному закону распределения:

$$p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Найти a и σ методом моментов.

40. Даны результаты измерений: 1,45; 0,16; -0,29; 0,86; 0,45. Можно ли утверждать с доверительной вероятностью 0,95, что среднее значение измеряемой величины отлично от 0? Предполагается, что результаты измерений подчиняются нормальному закону.

41. Экспериментально измерена зависимость y от x, т.е. в точках $x_1, \ldots, x_n$ измерены значения $y_1 + \delta_1, \ldots, y_n + \delta_n$. Зависимость аппроксимируется прямой линией: $y = \alpha_1 + \alpha_2 x$. Найти параметры

α_1 и α_2 методом наименьших квадратов и погрешность величины y, соответствующую произвольно выбранной точке x.

42. Популяция клеток облучается ионизирующим облучением (например, α-частицами). При попадании α-частицы в клетку, клетка с некоторой вероятностью гибнет. Можно показать, что в этом случае плотность распределения вероятности времени жизни клеток подчиняется экспоненциальному закону:

$$p(t,\lambda)=\begin{cases} 0, & \text{если } t<0; \\ \lambda e^{-\lambda t}, & \text{если } t \geq 0. \end{cases}$$

Параметр λ дает информацию о механизме взаимодействия излучения с клеткой. Проводились наблюдения за n клетками. Измерены времена их жизни: $\tau_1, \ldots, \tau_n$. Найти λ методом максимального правдоподобия и найти погрешность ее определения.

43. Экспериментально получены следующие значения (7 значений): 23,8; 24,1; 22,6; 25,4; 24,7; 23,1; 22,9. О них известно, что они подчиняются двойному экспоненциальному распределению (распределению Лапласа):

$$p(x,\alpha,\sigma)=\frac{1}{2\sigma}\mathrm{e}^{-|x-\alpha|/\sigma}, \quad -\infty<x<+\infty.$$

Найти α и σ методом максимального правдоподобия.

44. Измеренные случайные величины x_i, $i = 1, 2, \ldots, n$, независимы и имеют погрешности σ_i. Компоненты вектора $\overline{Y}$ определяются соотношениями:

а) $y_1 = x_1 - x_2$,
$y_2 = \sum_{i=1}^{n} x_i$;

б) $y_1 = x_1$,
$y_2 = \sum_{i=1}^{n} x_i$.

Найти ковариационную матрицу вектора $\overline{Y}$.

45. Для каких-то экспериментальных целей образец раздроблен на мелкие осколки так, что масса кусочков гораздо меньше массы исходного образца. Наугад выбираются n кусочков и определяются их массы x_i, $i = 1, 2, \ldots, n$. Требуется оценить среднюю массу осколков после дробления.

Замечание. При многократном делении отрезка на две части случайной длины, длина отрезка подчиняется логарифмически нормальному распределению:

$$p(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{(\ln x - a)^2}{2\sigma^2}\right\}, \quad x > 0 .$$

46. Экспериментально измерена (табл. 5) зависимость y_i от x_i (переменная x_i измерена без погрешности, переменная y_i с погрешностью $\delta_i = \delta = 0{,}2$).

Таблица 5

x_i	1	2	3	4	5	6	7	8	9	10
y_i	1,2	2,2	2,9	3,7	4,6	5,8	7,1	8,3	9,5	10,2

Можно ли утверждать с доверительной вероятностью 0,95, что измеренная зависимость отличается от зависимости $y = x$?

47. Сохраняется условие задачи 46. Можно ли утверждать с доверительной вероятностью 0,95, что измеряемая зависимость отличается от линейной $(y = \alpha_1 + \alpha_2 x)$?

Замечание. Сначала находятся оценки параметров линейной зависимости α_1 и α_2 с помощью метода наименьших квадратов.

48. Исследователь желает оценить вероятность изменения в организмах в результате некоторого воздействия. Он взял популяцию из n организмов и подверг их воздействию. Оказалось, что ни в одном из организмов никаких изменений не произошло. Это, разумеется, не означает, что вероятность изменения равна нулю, так как процесс имеет стохастический характер. Нужно оценить верхнюю границу вероятности измерений для доверительного уровня ε. Какова должна быть численность популяции, чтобы на доверительном уровне ε = 5 % можно было утверждать, что вероятность изменения $< 10^{-3}$?

49. Имеются числа:

0,12	0,88	0,39	0,73	0,43	0,65	0,02
0,78	0,11	0,84	0,04	0,28	0,50	0,13
0,92	0,17	0,97	0,41	0,50	0,77	0,90
0,71	0,22	0,63	0,69	0,21	0,83	0,09
0,76	0,38	0,80	0,73	0,68	0,61	0,31

Проверить гипотезу о том, что эти числа являются реализациями случайной величины равномерно распределенной на интервале (0, 1).

50. На основании данных задачи 49 проверить гипотезу о том, что данные подчиняются экспоненциальному распределению с плотностью вероятности $p(x) = 2e^{-2x}$, $x \geq 0$. Это распределение имеет среднее значение 1/2, т.е. то же самое, что и равномерное распределение на отрезке [0, 1], но существенно отличается по форме, являясь асимметричным. Расчеты провести для двух доверительных вероятностей $P_1 = 0{,}90$ и $P_2 = 0{,}98$.

51. Исследователь располагает прибором, показания на выходе которого флуктуируют (шум прибора). Флуктуации таковы, что они распределены по нормальному закону, единица измерения выбрана таким образом, что дисперсия нормального закона равна 1. Величины флуктуации в различающиеся моменты времени независимы друг от друга (так называемый «белый шум»). Исследователь ожидает появление сигнала в виде отклонения показания прибора в положительную сторону. С этой целью он через равные промежутки времени (t_i = 1, 2, …, 18) снимает показания прибора x_i. Получены результаты, которые показаны в табл. 6.

Таблица 6

t_1	1	2	3	4	5	6	7	8	9
x_1	–1,19	0,19	–1,56	0,46	0,14	2,46	–0,32	0,83	2,1
t_1	10	11	12	13	14	15	16	17	18
x_1	2,41	3,10	1,14	–0,96	0,06	–2,53	–0,53	–0,19	0,54

Можно ли утверждать с доверительной вероятностью 0,999, что в этих данных присутствует сигнал?

Замечание. Отсутствие сигнала означает, что все данные являются реализациями случайной величины со средним нуль и дисперсией единица.

52. Задана выборка объема n из генеральной совокупности, имеющей распределение Коши с параметром a

$$p(x, a) = \frac{1}{\pi}\frac{1}{1+(x-a)^2}, \quad -\infty < x < +\infty.$$

Найти оценку параметра a, погрешность оценки при большом n. Является ли эта оценка асимптотически эффективной?

53. Задана выборка объема n из генеральной совокупности, имеющей равномерное распределение на отрезке [0, 1]. Найти совместную плотность распределения n порядковых статистик $(\xi_1^*, \xi_2^*, ..., \xi_n^*)$.

54. Решить задачу 53 при условии, что генеральная совокупность имеет равномерное распределение на отрезке $[a, b]$.

ПРИЛОЖЕНИЯ

Таблица П1

Распределение Пуассона. Значения функции $\sum_{k=x}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$

x	λ					
	0,1	0,2	0,3	0,4	0,5	0,6
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,095	0,181	0,259	0,330	0,394	0,451
2	0,005	0,018	0,037	0,062	0,090	0,122
3		0,001	0,003	0,008	0,014	0,023
4				0,001	0,002	0,003

Продолжение 1 табл. П1

x	λ					
	0,7	0,8	0,9	1,0	2,0	3,0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,503	0,551	0,593	0,632	0,865	0,950
2	0,156	0,191	0,228	0,264	0,594	0,801
3	0,034	0,047	0,063	0,080	0,323	0,577
4	0,006	0,009	0,014	0,019	0,143	0,353
5	0,001	0,001	0,002	0,004	0,053	0,185
6				0,001	0,018	0,084
7					0,005	0,034
8					0,001	0,012

Продолжение 2 табл. П1

x	λ					
	4,0	5,0	6,0	7,0	8,0	9,0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,982	0,993	0,998	0,999	1,000	1,000
2	0,908	0,960	0,983	0,924	0,997	0,999
3	0,762	0,875	0,938	0,970	0,986	0,994
4	0,567	0,735	0,849	0,918	0,958	0,979
5	0,371	0,560	0,715	0,827	0,900	0,945
6	0,215	0,384	0,554	0,699	0,809	0,884
7	0,111	0,238	0,394	0,550	0,687	0,793
8	0,051	0,133	0,256	0,401	0,547	0,676
9	0,021	0,068	0,153	0,271	0,408	0,544
10	0,008	0,032	0,084	0,170	0,283	0,413
11	0,003	0,014	0,043	0,99	0,184	0,294
12	0,001	0,005	0,020	0,053	0,112	0,197
13		0,002	0,008	0,027	0,068	0,124
14		0,001	0,004	0,013	0,034	0,074
15			0,001	0,006	0,017	0,042
16			0,001	0,002	0,008	0,022
17				0,001	0,004	0,011
18					0,002	0,005
19					0,001	0,002
20						0,001

Таблица П2

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1801	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1530	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001

Таблица П3

Нормальное распределение. Квантили распределения: $p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_p} e^{-x^2/2} dx$

p	u_p	p	u_p	p	u_p
0,50	0,000	0,68	0,468	0,86	1,080
0,51	0,025	0,69	0,496	0,87	1,126
0,52	0,050	0,70	0,524	0,88	1,175
0,53	0,075	0,71	0,553	0,89	1,227
0,54	0,100	0,72	0,583	0,90	1,282
0,55	0,126	0,73	0,613	0,91	1,341
0,56	0,151	0,74	0,643	0,92	1,405
0,57	0,176	0,75	0,674	0,92	1,476
0,58	0,202	0,76	0,706	0,93	1,555
0,59	0,228	0,77	0,739	0,94	1,645
0,60	0,253	0,78	0,772	0,95	1,751
0,61	0,279	0,79	0,806	0,96	1,881
0,62	0,305	0,80	0,842	0,97	2,054
0,63	0,332	0,81	0,878	0,98	2,326
0,64	0,358	0,82	0,915	0,99	3,090
0,65	0,385	0,83	0,954	0,999	3,720
0,66	0,412	0,84	0,991	0,9999	4,265
0,67	0,440	0,85	1,036		

Биномиальное распределение.
95%-ные доверительные пределы (θ_1, θ_2)
для параметра θ: $P_\theta(\theta_1 < \theta < \theta_2) = 0{,}95$

k	$n-k$												
	0	1	2	3	4	5	6	7	8	9	10	11	12
0	–	0,98	0,84	0,71	0,60	0,52	0,46	0,41	0,37	0,34	0,31	0,29	0,27
	–	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1	1,00	0,99	0,91	0,81	0,72	0,64	0,58	0,53	0,48	0,45	0,41	0,39	0,36
	0,03	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	1,00	0,99	0,93	0,85	0,78	0,71	0,65	0,60	0,56	0,52	0,48	0,45	0,43
	0,16	0,09	0,07	0,05	0,04	0,04	0,03	0,03	0,03	0,03	0,02	0,02	0,02
3	1,00	0,99	0,95	0,88	0,82	0,76	0,70	0,65	0,61	0,57	0,54	0,51	0,48
	0,29	0,19	0,15	0,12	0,10	0,09	0,08	0,07	0,06	0,06	0,05	0,05	0,04
4	1,00	1,00	0,96	0,90	0,84	0,79	0,74	0,69	0,65	0,61	0,58	0,55	0,52
	0,40	0,29	0,22	0,18	0,16	0,14	0,12	0,11	0,10	0,09	0,08	0,08	0,07
5	1,00	1,00	0,96	0,92	0,86	0,81	0,77	0,72	0,68	0,65	0,62	0,59	0,56
	0,48	0,36	0,29	0,25	0,21	0,19	0,17	0,15	0,14	0,13	0,12	0,11	0,10
6	1,00	1,00	0,97	0,93	0,88	0,83	0,79	0,75	0,71	0,68	0,65	0,62	0,50
	0,54	0,42	0,35	0,30	0,26	0,23	0,21	0,19	0,18	0,16	0,15	0,14	0,13
7	1,00	1,00	0,97	0,93	0,89	0,85	0,81	0,77	0,73	0,70	0,67	0,64	0,62
	0,59	0,47	0,40	0,35	0,31	0,28	0,25	0,23	0,21	0,20	0,18	0,17	0,16
8	1,00	1,00	0,98	0,94	0,90	0,86	0,82	0,79	0,75	0,72	0,69	0,67	0,64
	0,63	0,52	0,44	0,39	0,35	0,32	0,29	0,27	0,25	0,23	0,22	0,20	0,19
9	1,00	1,00	0,98	0,95	0,91	0,87	0,84	0,80	0,77	0,74	0,71	0,69	0,66
	0,66	0,56	0,48	0,43	0,39	0,35	0,32	0,30	0,28	0,26	0,24	0,23	0,22
10	1,00	1,00	0,98	0,95	0,92	0,88	0,85	0,82	0,79	0,76	0,73	0,70	0,68
	0,69	0,59	0,52	0,46	0,42	0,38	0,35	0,33	0,31	0,29	0,27	0,26	0,24
11	1,00	1,00	0,98	0,95	0,92	0,89	0,86	0,83	0,80	0,77	0,74	0,72	0,69
	0,72	0,62	0,57	0,49	0,45	0,41	0,38	0,36	0,34	0,32	0,30	0,28	0,27
12	1,00	1,00	0,98	0,96	0,93	0,90	0,87	0,84	0,81	0,78	0,76	0,73	0,71
	0,74	0,64	0,57	0,52	0,48	0,44	0,41	0,38	0,36	0,34	0,32	0,31	0,29

Примечание. Значения θ_2 набраны в нервых строках, значения θ_1 — во вторых.

Таблица П5

Квантили хи-квадрат распределения $\chi_p^2(k)$, $P\{\chi_k^2 < \chi_p^2(k)\} = p$

k	p														
	0,005	0,010	0,025	0,05	0,10	0,20	0,30	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999
1	$0,0^4393$	$0,0^3157$	$0,0^3982$	$0,0^2393$	0,0158	0,0642	0,148	1,07	1,64	2,71	3,84	5,02	6,63	7,88	10,8
2	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	2,41	3,22	4,61	5,99	7,38	9,21	10,6	13,8
3	0,0717	0,115	0,216	0,352	0,584	1,00	1,42	3,67	4,64	6,25	7,81	9,35	11,3	12,8	16,3
4	0,207	0,297	0,484	0,711	1,06	1,65	2,19	4,88	5,99	7,78	9,49	11,1	13,3	14,9	18,5
5	0,412	0,554	0,831	1,15	1,61	2,34	3,00	6,06	7,29	9,24	11,1	12,8	15,1	16,7	20,5
6	0,676	0,872	1,34	1,64	2,20	3,07	3,83	7,23	8,56	10,6	12,6	14,4	16,8	18,5	22,5
7	0,989	1,24	1,69	2,17	2,83	3,82	4,67	8,38	9,80	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	4,59	5,53	9,52	11,0	13,4	15,5	17,5	20,1	22,0	26,1
9	1,73	2,09	2,70	3,33	4,17	5,38	6,39	10,7	12,2	14,7	16,9	19,0	21,7	23,6	27,9
10	2,16	2,56	3,25	3,94	4,87	6,18	7,27	11,8	13,4	16,0	18,3	20,5	23,2	25,2	29,6
11	2,60	3,05	3,82	4,57	5,58	6,99	8,15	12,9	14,6	17,3	19,7	21,9	24,7	26,8	31,3
12	3,07	3,57	4,40	5,23	6,30	7,81	9,03	14,0	15,8	18,5	21,0	23,3	26,2	28,3	32,9
13	3,57	4,11	5,01	5,89	7,04	8,63	9,93	15,1	17,0	19,8	22,4	24,7	27,7	29,8	34,5
14	4,07	4,66	5,63	6,57	7,79	9,47	10,8	16,2	18,2	21,1	23,7	26,1	29,1	31,3	36,1
15	4,60	5,23	6,26	7,26	8,55	10,3	11,7	17,3	19,3	22,3	25,0	27,5	30,6	32,8	37,7
16	5,14	5,81	6,91	7,96	9,31	11,2	12,6	18,4	20,5	23,5	26,3	28,8	32,0	34,3	39,3
17	5,70	6,41	7,56	8,67	10,1	12,0	13,5	19,5	21,6	24,8	27,6	30,2	33,4	35,7	40,8
18	6,26	7,01	8,23	9,39	10,9	12,9	14,4	20,6	22,8	26,0	28,9	31,5	34,8	37,2	42,3

Таблица П5 (продолжение)

k	p														
	0,005	0,010	0,025	0,05	0,10	0,20	0,30	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999
19	6,84	7,63	8,91	10,1	11,7	13,7	15,4	21,7	23,9	27,2	30,1	32,9	36,2	38,6	43,8
20	7,43	8,26	9,59	10,9	12,4	14,6	16,3	22,8	25,0	28,4	31,4	34,2	37,6	40,0	45,3
21	8,03	8,90	10,3	11,6	13,2	15,4	17,2	23,9	26,9	29,6	32,7	35,5	38,9	41,4	46,8
22	8,64	9,54	11,0	12,3	14,0	16,3	18,1	24,9	27,3	30,8	33,9	36,8	40,3	42,8	48,3
23	9,26	10,2	11,7	13,1	14,8	17,2	19,0	26,0	28,4	32,0	35,2	38,1	41,6	44,2	49,7
24	9,89	10,9	12,4	13,8	15,7	18,1	19,9	27,1	29,6	33,2	36,4	39,4	43,0	45,6	51,2
25	10,5	11,5	13,1	14,6	16,5	18,9	20,9	28,2	30,7	34,4	37,7	40,6	44,3	46,9	52,6
26	11,2	12,2	13,8	15,4	17,3	19,8	21,8	29,2	31,8	35,6	38,9	41,9	45,6	48,3	54,1
27	11,8	12,9	14,6	16,2	18,1	20,7	22,7	30,3	32,9	36,7	40,1	43,2	47,0	49,6	55,5
28	12,5	13,6	15,3	16,9	18,9	21,6	23,6	31,4	34,0	37,9	41,3	44,5	48,3	51,0	56,9
29	13,1	14,3	16,0	17,7	19,8	22,5	24,6	32,5	35,1	39,1	42,6	45,7	49,6	52,3	58,3
30	13,8	15,0	16,8	18,5	20,6	23,4	25,5	33,5	36,3	40,3	43,8	47,0	50,9	53,7	59,7
35	17,2	18,5	20,6	22,5	24,8	27,8	30,2	38,9	41,8	46,1	49,8	53,2	57,3	60,3	66,6
40	20,7	22,2	24,4	26,5	29,1	32,3	34,9	44,2	47,3	51,8	55,8	59,3	63,7	66,8	73,4
45	24,3	25,9	28,4	30,6	33,4	36,9	39,6	49,5	52,7	57,5	61,7	65,4	70,0	73,2	80,1
50	28,0	29,7	32,4	34,8	37,7	41,4	44,3	54,7	58,2	63,2	67,5	71,4	76,2	79,5	86,7
75	47,2	49,5	52,9	56,1	59,8	64,5	68,1	80,9	85,1	91,1	96,2	100,8	106,4	110,3	118,6
100	67,3	70,1	74,2	77,9	82,4	87,9	92,1	106,9	111,7	118,5	124,3	129,6	135,8	146,2	149,4

Таблица П6

Значения функции $P(\lambda) = 1 - \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2\lambda^2}$

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
0,00	1,0000	0,45	0,9874	0,90	0,3927	1,70	0,0062
0,05	1,0000	0,50	0,9639	0,95	0,3275	1,80	0,0032
0,10	1,0000	0,55	0,9228	1,00	0,2700	1,90	0,0015
0,15	1,0000	0,60	0,8643	1,10	0,1777	2,00	0,0007
0,20	1,0000	0,65	0,7920	1,20	0,1122	2,10	0,0003
0,25	1,0000	0,70	0,7112	1,30	0,0681	2,20	0,0001
0,30	1,0000	0,75	0,6272	1,40	0,0397	2,30	0,0001
0,35	0,9997	0,80	0,5441	1,50	0,0222	2,40	0,0000
0,40	0,9972	0,85	0,4653	1,60	0,0120	2,50	0,0000

Таблица П7

Критерий Колмогорова.

Значения функции $\lambda_p : \; p = P \;\; (Д_n = \sup\left|F_n^*(x) - F(x)\right| > \lambda_p)$

n	p			n	p		
	0,10	0,05	0,01		0,10	0,05	0,01
1	0,950	0,975	0,995	19	0,271	0,301	0,0361
2	0,776	0,842	0,929	20	0,265	0,294	0,352
3	0,636	0,708	0,829	25	0,238	0,264	0,317
4	0,565	0,624	0,734	30	0,218	0,242	0,290
5	0,509	0,563	0,669	35	0,202	0,224	0,269
6	0,468	0,519	0,617	40	0,189	0,210	0,252
7	0,436	0,483	0,576	45	0,179	0,198	0,238
8	0,410	0,454	0,542	50	0,170	0,188	0,226
9	0,387	0,430	0,513	55	0,162	0,180	0,216
10	0,369	0,409	0,489	60	0,155	0,172	0,207
11	0,352	0,391	0,468	65	0,149	0,166	0,199
12	0,338	0,375	0,449	70	0,144	0,160	0,192
13	0,325	0,361	0,432	75	0,139	0,154	0,185
14	0,314	0,349	0,418	80	0,135	0,150	0,179
15	0,304	0,338	0,404	85	0,131	0,145	0,174
16	0,295	0,327	0,392	90	0,127	0,141	0,169
17	0,286	0,318	0,381	95	0,124	0,137	0,165
18	0,279	0,309	0,371	100	0,121	0,134	0,161

ОТВЕТЫ К ДОПОЛНИТЕЛЬНЫМ ЗАДАЧАМ

1. а) $P\{\xi = 0\} = P\{\xi = 2\} = \dfrac{1}{4}$, $P\{\xi = 1\} = \dfrac{1}{2}$;

б) $P\{\xi = m\} = \mathrm{e}^{-\lambda} \dfrac{\lambda^m}{m!}$, $m = 0, 1, \ldots$; в) $P\{\xi = m\} = \left(\dfrac{1}{2}\right)^{m-1}$, $m = 0,$ $1, \ldots$; г) $P\{\xi = m\} = C_n^m \left(\dfrac{2}{3}\right)^m \left(\dfrac{1}{3}\right)^{n-m}$, $m = \overline{0, n}$.

2. $z\varphi(z)$; $\varphi(z^2)$; $\dfrac{\varphi(z)}{1-z}$, $\dfrac{\varphi(\sqrt{z}) + \varphi(-\sqrt{z})}{2}$.

3. а) $\dfrac{\sin at}{at}$; б) e^{8it-2t^2}; в) $(q + p\mathrm{e}^{it})^n$.

4. а) $p_\xi(x) = \begin{cases} \dfrac{4-|x|}{16}, & x \in [-4, 4]; \\ 0, & x \notin [-4, 4]. \end{cases}$

б)

K	4	2	1	0	–1	–2	–4
$P\{\xi = k\}$	1/8	1/4	1/16	1/8	1/16	1/4	1/8

5. $P(A) \geq 0{,}666$, $P(A) \geq 0{,}96$. **6.** $P \geq 0{,}88$.

7. а) $P \geq 0{,}86$; б) $P \geq 0{,}75$.

8. Да. **11.** а) 0,08; б) 0,0013; в) 0,89.

12. $\dfrac{\xi - 300}{10\sqrt{6}} - N(0, 1)$. **13.** 0,8943. **14.** а) $5{,}088 \cdot 10^{-3}$; б) 0,8668.

15. 0,0013. **16.** а) 3 и 4 0,19537; б) 0,78514.

17. 0,9974. **18.** а) 0,87535; б) 0,99998; в) 0,14622; г) 4 и 5 0,17547.

19. а) 0,8858; б) 0,8858; 0,499; 0,146; 0,835;

в) 0,033; г) n = 1850700 (с избытком).

20. (792, 828). **21.** = 0,9974. **22.** 0,298.

23. $n = 51$; $\mu_n = \{$число появлений «1» при n бросаниях кости$\}$. Если $\{0 \le \mu_n \le 16\}$, то с вероятностью, бóльшей 0,9, была взята «правильная» кость; в остальных случаях — «неправильная».
24. 2470.

$$\textbf{25.}\ F_n^*(x) = \begin{cases} 0, & \text{если } x \le 4; \\ 1/9, & \text{если } 4 < x \le 5; \\ 2/9, & \text{если } 5 < x \le 6; \\ 4/9, & \text{если } 6 < x \le 7; \\ 6/9, & \text{если } 7 < x \le 8; \\ 7/9, & \text{если } 8 < x \le 9; \\ 8/9, & \text{если } 9 < x \le 11; \\ 1, & \text{если } x > 11; \end{cases} \qquad \overline{x} = 7, \quad S^2 = 4{,}5.$$

26. $\hat{\theta} = \dfrac{1}{x} = \dfrac{1}{7}$. **27.** $\hat{p} = \dfrac{1}{k}$, $M\hat{p} = \dfrac{p \ln p}{p-1}$.

28. $\hat{\lambda} = \dfrac{a}{x}$, $M\hat{\lambda} = \lambda + \dfrac{\lambda}{an-1}$.

29. 1. (18,35; 21,64), (17,42; 22,58). 2. (498,35; 501,64), (497,42; 502,58). 3. (28,14; 31,86), (26,64; 33,36). 4. (16,63; 19,37), (15,76; 20,24).
30. 1. (2,74; 9), (2,5; 10,2). 2. (10,65; 54,55), (7,64; 104,1).
31. а) (2,61; 14,84), б) (1,84; 30,3). **32.** (2,00; 2,6).
33. 1. (0,225; 0,375). 2. (0,210; 0,390). **34.** (0,10; 0,82).
35. Гипотеза H_0 принимается во всех случаях.
36. 1. Гипотеза H_0 не принимается. 2. Гипотеза H_0 принимается.
37. Гипотеза H_0 принимается во всех случаях.

38. 1. а) $M\hat{\theta}_1 = M\hat{\theta}_2 = \dfrac{b+a}{2}$, $D\hat{\theta}_1 = \dfrac{(b-a)^2}{12n}$, $D\hat{\theta}_2 = \dfrac{(b-a)^2}{2(n+1)(n+2)}$.

б) $M\hat{\theta}_1 = \dfrac{1}{\alpha}$, $D\hat{\theta}_1 = \dfrac{1}{\alpha^2 n}$, $M\hat{\theta}_2 = \dfrac{1}{2\alpha}\left(1 + \dfrac{1}{2} + \ldots + \dfrac{1}{n-1} + \dfrac{2}{n}\right)$,

$D\hat{\theta}_2 = \dfrac{1}{4\alpha^2}\left(1 + \dfrac{1}{2^2} + \ldots + \dfrac{1}{(n-1)^2} + \dfrac{4}{n^2}\right)$. 2. $M\hat{c} = c$, $D\hat{c} = 1/n^2$.

39. $\hat{a} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{a})^2$. **40.** $P\{a > -0{,}12\} = 0{,}95$.

42. $\hat{\lambda} = n\left(\sum_{i=1}^{n} t_i\right)^{-1}$; $\widehat{D\hat{\lambda}} = n\left(\sum_{i=1}^{n} t_i\right)^{-2}$; $\Delta = \sqrt{\widehat{D\hat{\lambda}}} = \sqrt{n}\left(\sum_{i=1}^{n} t_i\right)^{-1}$.

43. $\hat{\alpha} = 23{,}8$; $\hat{\sigma} = 0{,}8$.

44. Указание. Воспользоваться формулой переноса ошибок $B_y = AB_xA^T$, где $B_x = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \dots & \\ 0 & & & \sigma_n^2 \end{bmatrix}$ — ковариационная матрица вектора $\bar{x}$.

45. $\hat{m} = \sqrt[n]{\prod_{i=1}^{n} x_i}$ $(e^{\sigma^2/2} = 1)$.

46. Да, можно (хотя с вероятностью 0,97 уже нельзя).

47. Нет, нельзя (с вероятностью 0,95 можно утверждать, что зависимость нелинейна).

48. $P_{\text{верх}} = 1 - \varepsilon^{1/n}$, если $\varepsilon = 0{,}05$, то $n = \ln e / \ln(1 - P_{\text{верх}}) = 30000$.

49. Гипотеза не отвергается для $\alpha = 0{,}1$.

50. С доверительной вероятностью 0,9 гипотеза отвергается, а с 0,98 — не отвергается.

51. Гипотеза об отсутствии сигнала при $\alpha = 0{,}001$ не отвергается, т.е. нельзя утверждать, что сигнал присутствует.

52. $\hat{a} = \xi^*_{n/2}$, $D\hat{a} = \frac{\pi^2}{4n}$, так как $\frac{\pi^2}{4n} > \frac{2}{n}$, оценка $\hat{a}$ не является асимптотически эффективной.

53. $p^*(x_1, x_2, ..., x_n) = \begin{cases} n!, & \text{если } 0 \le x_1 \le x_2 \le x_3 \le x_n \le 1; \\ 0, & \text{в противном случае.} \end{cases}$

54. $p^*(x_1, x_2, ..., x_n) = \begin{cases} \frac{n!}{(b-a)^n}, & \text{если } a \le x_1 \le x_2 \le x_3 \le \ldots \le x_n \le b; \\ 0, & \text{в противном случае.} \end{cases}$

ЛИТЕРАТУРА

1. *Чистяков В.П.* Курс теории вероятностей. М.: Агар, 2000.

2. *Ватутин В.А., Полякова Е.И., Постникова Л.П., Шерстнев В.И.* Методические указания по теме «Случайные величины. Элементы математической статистики». М.: МИФИ, 1990.

3. *Полякова Е.И., Постникова Л.П.* Сборник задач по теории вероятностей и математической статистике. М.: МИФИ, 2004.

4. Методические указания к решению задач по курсу «Теория вероятностей и математическая статистика» (часть 1). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 1999.

5. Методические указания к решению задач по курсу «Теория вероятностей и математическая статистика» (часть 2). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 1999.

6. Методические указания к решению задач по курсу «Теория вероятностей и математическая статистика» (часть 3). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 1999.

7. Методические указания к решению задач по курсу «Теория вероятностей и математическая статистика» (часть 4). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 2000.

8. Теория вероятностей и математическая статистика. Методические рекомендации (часть 1). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 2008.

9. Теория вероятностей и математическая статистика. Методические рекомендации (часть 2). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 2008.

10. Теория вероятностей и математическая статистика. Методические рекомендации (часть 3). В 4-х частях / Сост. *Е.И. Полякова, Л.П. Постникова, Е.В. Сумин.* М.: МИФИ, 2008.

11. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. М.: Высшая школа, 1984.

СОДЕРЖАНИЕ