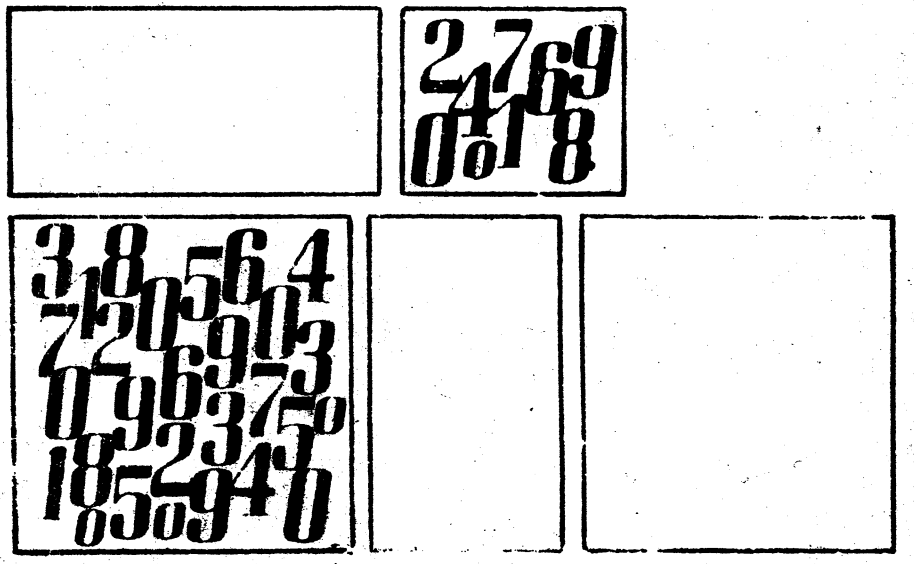
Е.Л.Александров

СБОРНИК ЗАДАЧ по математической СТАТИСТИКЕ



ОБСКИЙ ордена Трудового Красного Знамени
ственный университет имени Н.Г.Чернышевского

Е.Л.Александров

СБОРНИК ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Учебное пособие
для студентов механико-математического
и физического факультетов

Издательство Саратовского университета
1992

УДК 519.2 (076.1)
А 46

Рецензенты: кафедра теории вероятностей и математического моделирования Саратовского университета;
профессор В.Н.Иванов

А $\frac{1602090000 - 5}{176(02)-92}$ 57 -92

ISBN 5-292-01489-3

ПРЕДИСЛОВИЕ

Настоящее учебное пособие содержит задачи по основным разделам курса теории вероятностей и математической статистики, читаемого на механико-математическом факультете СГУ.

Пособие состоит из 6 параграфов, ответов и приложения. В начале каждого параграфа приводятся краткие теоретические сведения и формулы, необходимые для решения задач, а также наиболее типичные задачи с решениями. Задачи расположены в порядке возрастания их трудности: в начале даны такие, которые закрепляют материал лекционного курса, а наиболее трудные задачи, требующие творческого подхода или большей вычислительной работы, помещены в конце. Почти все задачи снабжены ответами, указаниями или решениями. В приложении помещены нужные в процессе работы (табл.I-II).

Необходимость выхода такого пособия продиктована тем, что в настоящее время не существует сборников задач по математической статистике, содержащих всех тем читаемого курса. Кроме того, испытывается нехватка задачников по отдельным темам.

Данное пособие частично (в пределах университета) сможет заполнить этот пробел и послужит улучшению преподавания теории вероятностей и математической статистики на механико-математическом и физическом факультетах Саратовского университета.

ВВЕДЕНИЕ

Дискретные распределения

1. Биномиальное распределение $Bi(n,p)$ – это распределение числа "успехов" в n независимых испытаниях с двумя исходами и с вероятностью "успеха" $p \in (0,1)$ (схема Бернулли).

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

$$M\xi = np; \quad D\xi = npq.$$

2. Полиномиальное распределение $M(n; p_1, \ldots, p_N)$, $p_i \geq 0$, $p_1 + \ldots + p_N = 1$ – это распределение случайного вектора, $\bar{\nu} = (\nu_1, \ldots, \nu_N)$, $\nu_1 + \ldots + \nu_N = n$, с целочисленными компонентами (ν_i – число реализаций i-го исхода в n независимых испытаниях с N исходами).

$$P(\bar{\nu} = \bar{k}) = \frac{n!}{k_1! \ldots k_N!} p_1^{k_1} \ldots p_N^{k_N}, \quad \bar{k} = (k_1, \ldots, k_N), \; k_1 + \ldots + k_N = n$$

$$M\nu_i = np_i, \quad cov(\nu_i, \nu_j) = \begin{cases} np_i(1-p_i) & \text{при } i = j, \\ -np_i p_j & \text{при } i = j. \end{cases}$$

3. Распределение Пуассона $\Pi(\lambda)$, $\lambda > 0$ задается вероятностями

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots, \qquad M\xi = D\xi = \lambda.$$

4. Отрицательное биномиальное распределение $\overline{Bi}(r, p)$, $p \in (0,1)$, $r \in N$ задается вероятностями

$$P(\xi=k)=C^{k}_{r+k-1}p^{k}q^{r},\quad k=0,1,2,\dots,\qquad q=1-p.$$

Это распределение числа "успехов", предшествующих r-му "неуспеху" в последовательности испытаний Бернулли.

$$M\xi=rp/q;\qquad D\xi=rp/q^{2}.$$

5. Гипергеометрическое распределение $H(r,N,n)$ задается вероятностями

$$P(\xi=k)=C^{k}_{r}C^{n-k}_{N-r}/C^{n}_{N},\quad \max(0,n+r-N)\le k\le \min(n,r).$$

Если в урне содержится N шаров, r из которых - белые и $N-r$ - черные, и из урны извлекается случайная выборка без возвращения объема n, то случайная величина ξ - число белых шаров в выборке - имеет указанное распределение:

$$M\xi=nr/N,\quad D\xi=\frac{nr}{N}\left(1-\frac{r}{N}\right)\frac{N-n}{N-1}.$$

Абсолютно непрерывные распределения

6. Равномерное распределение $R(a,b)$, $-\infty<a<b<\infty$, имеет плотность

$$f(x)=(b-a)^{-1},\quad x\in[a,b],\ f(x)=0,\quad x\notin[a,b].$$

$$M\xi=\frac{1}{2}(a+b),\quad D\xi=\frac{1}{12}(b-a)^{2}.$$

7. Нормальное распределение $N(a,\sigma^{2})$ имеет плотность

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^{2}}{2\sigma^{2}}}\qquad(\sigma>0).$$

$M\xi=a$; $D\xi=\sigma^{2}$; центральные моменты:

$$\mu_{k}=M(\xi-a)^{k};\quad \mu_{2k+1}=0;\quad \mu_{2k}=\frac{(2k)!}{2^{k}\cdot k!}\sigma^{2k}=1\cdot3\cdot\ldots\cdot(2k-1)\sigma^{2k}.$$

Распределение $N(0,1)$ называется стандартным нормальным, ее функцию распределения обозначают $\Phi(x)$.

Случайный вектор $\xi=(\xi_1,\dots,\xi_k)$ имеет нормальное (k-мерное) распределение: $N(\bar{a},\Sigma)$, $\bar{a}=M\xi=(a_1,\dots,a_k)$,

$$\Sigma=[\sigma_{ij}]_1^k=D\xi=M[(\xi-\bar{a})(\xi-\bar{a})^T]=[cov(\xi_i,\xi_j)]_1^k.$$

Если определитель $|\Sigma|\neq 0$, то распределение $N(\bar{a},\Sigma)$ называется невырожденным и имеет плотность

$$f(x)=(2\pi)^{-k/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(x-\bar{a})^T\Sigma^{-1}(x-\bar{a})\right\},$$

$$x=(x_1,\dots,x_k)\in R^k.$$

Важным свойством нормального распределения является то, что при линейном преобразовании $\eta=A\xi$ (A – заданная матрица) снова получается нормальный случайный вектор, при этом $L(\eta)=N(A\bar{a},A\Sigma A^T)$. Всегда существует линейное преобразование, переводящее невырожденный нормальный вектор в вектор с независимыми стандартными нормальными компонентами.

8. Гамма-распределение $\Gamma(a,\lambda)$, $a>0$, $\lambda>0$ задается плотностью

$$f(x)=\frac{x^{\lambda-1}e^{-x/a}}{\Gamma(\lambda)\,a^{\lambda}},\quad x\geq 0,\quad f(x)=0,\quad x<0,$$

где $\Gamma(\lambda)=\int_0^\infty t^{\lambda-1}e^{-t}dt$, $\lambda>0$ – гамма-функция.

$M\xi=a\lambda$, $D\xi=a^2\lambda$; моменты: $M\xi^b=a^b\Gamma(\lambda+b)/\Gamma(\lambda)$, $b>-\lambda$.

Свойства функции $\Gamma(\lambda)$: $\Gamma(a+1)=a\Gamma(a)$, $\Gamma(n+1)=n!$

$\Gamma(a+n)=(a+n-1)\dots(a+1)a\Gamma(a)$, $\Gamma(1/2)=\sqrt{\pi}$,

$\Gamma(a)\Gamma(1-a)=\pi/\sin a\pi$.

Формула Стирлинга для гамма-функции:

$$\Gamma(z)\sim\sqrt{2\pi z}\,z^{z-1}e^{z},\quad z\to\infty.$$

9. χ^2 – распределение со степенями свободы n $\chi^2(n)$. Распределение случайной величины $\xi=\xi_1^2+\dots+\xi_n^2$,

где $\xi_1, \ldots, \xi_n$ – независимые N (0,I) величины. Плотность распределения $\chi^2(n)$ есть

$$f(x)=\frac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)\,2^{n/2}}, \quad x\geq 0, \; f(x)=0, \quad x<0;$$

$$M\xi=n; \quad D\xi=2n.$$

Если $X=(X_1,\ldots,X_n)$ – выборка из распределения N (0,I) – и $Q=X^TAX$, где A – идемпотентная матрица: $A^2=A$, тогда $L(Q)=\chi^2(r)$, где $r=rang\,A=tr\,A$ – след матрицы A.

10. Распределение Вейбула, $W(a,\alpha,b)$, задается функцией распределения

$$F(x)=1-exp\left[-\left(\frac{x-a}{b}\right)^{\alpha}\right], \quad x\geq a,$$

где $a\in\mathbb{R}$ – параметр сдвига, $\alpha>0$ – параметр формы, $b>0$ – параметр масштаба.

$$M\xi=a+b\Gamma(1+1/\alpha), \quad D\xi=b^2\left[\Gamma(1-2/\alpha)-\Gamma^2(1+1/\alpha)\right].$$

В частности, $W(a,2,b)$ – распределение Релея.

11. Бета-распределение $\beta(a,b)$, $a,b>0$ задается плотностью

$$f(x)=\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, \quad 0\leq x\leq 1,$$

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}=\int_a^b x^{a-1}(1-x)^{b-1}dx$$ – бета-функция.

12. Распределение Коши $K(a)$, $a\in\mathbb{R}$ задается плотностью

$$f(x)=\frac{1}{\pi}\cdot\frac{1}{1+(x-a)^2}, \quad x\in\mathbb{R}.$$

Моменты не существуют. Свойство: если случайные величины $\xi_1,\ldots,\xi_n$ независимы и $L(\xi_i)=K(a_i)$, $i=1,\ldots,n$, то

$$L\left(\frac{\xi_1+\ldots+\xi_n}{n}\right)=K\left(\frac{a_1+\ldots+a_n}{n}\right).$$

13. Распределение Стьюдента $S(n)$ (или t - распределение) с n степенями свободы есть распределение случайной величины

$$t=\eta/\sqrt{\chi^2/n}\ ,$$

где $L(\eta)=N(0,1)$; $L(\chi^2_n)=\chi^2(n)$ и случайные величины, η и χ^2_n, независимы. Плотность этого распределения

$$S_n=\frac{1}{\sqrt{\pi n}}\ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)(1+x^2/n)^{(n+1)/2}}\ ,\ x\in R\ .$$

14. Экспоненциальное, или показательное, распределение задается плотностью

$$f(x)=\lambda e^{-\lambda x},\quad x\geq 0,\ f(x)=0,\quad x<0.$$

$$M\xi=1/\lambda\ ,\quad D\xi=1/\lambda^2.$$

15. Распределение Снедекора-Фишера $S(n_1,n_2)$ с n_1 и n_2 степенями свободы есть распределение случайной величины

$$F=\frac{\chi^2_{n_1}}{n_1}\Big/\frac{\chi^2_{n_2}}{n_2}\ ,$$

где $\chi^2_{n_1}$ и $\chi^2_{n_2}$ независимы. Плотность этого распределения имеет вид

$$f_{n_1n_2}(x)=\left(\frac{n_1}{n_2}\right)^{n_1/2}\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma(n_1/2)\,\Gamma(n_2/2)}\cdot\frac{x^{n_1/2-1}}{\left(1+\frac{n_1}{n_2}x\right)^{(n_1+n_2)/2}},\ x>0.$$

§ 1. ВЫБОРОЧНЫЙ МЕТОД В СТАТИСТИКЕ. ВЫБОРОЧНЫЕ ХАРАКТЕРИСТИКИ И ИХ РАСПРЕДЕЛЕНИЯ

Выборка. Вариационный ряд. Пусть эксперимент состоит в проведении n испытаний, в которых результат i-го испытания описывается случайной величиной X_i, $i=1,\ldots,n$. Выборкой на-

зывается совокупность наблюдаемых случайных величин $X = (X_1, \ldots, X_n)$, величины X_i называются элементами выборки, n – ее объемом, множество $\{(x_1, \ldots, x_n)\}$ всех возможных реализаций – выборочным пространством. Функция распределения

$$F_X(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n)$$

бывает неизвестна полностью или неизвестна частично и указан лишь класс допустимых распределений $\mathcal{F} = \{F(x_1, \ldots, x_n)\}$, которому принадлежит F_X. Пару $(X, \mathcal{F})$ называют статистической моделью.

Если эксперимент состоит в проведении n независимых испытаний случайной величины ξ, то выборка $X = (X_1, \ldots, X_n)$ в таком случае представляет собой совокупность независимых одинаково распределенных случайных величин с функцией распределения $F_\xi(x)$. Говорят, что X есть выборка из распределения $L(\xi)$. Мы будем также под выборкой из распределения $L(\xi)$ понимать реализацию $x_1, \ldots, x_n$ n независимых измерений случайной величины ξ. Каждой реализации $x_1, \ldots, x_n$ выборки $X = (X_1, \ldots, X_n)$ поставим в соответствие упорядоченную последовательность $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$, где $x_{(1)} = \min\{x_1, \ldots x_n\}$; $x_{(2)}$ – второе по величине значение и т.д., $x_{(n)} = \max\{x_1, \ldots, x_n\}$. Обозначим через $X_{(k)}$ случайную величину, которая для каждой реализации выборки X принимает значение $x_{(k)}$. $X_{(k)}$ называют k-й порядковой статистикой,

$$X_{(1)} = \min\{X_1, \ldots, X_n\}, \quad X_{(n)} = \max\{X_1, \ldots, X_n\} -$$

экстремальными значениями выборки, последовательность $X_{(1)}, \ldots, X_{(n)}$ – вариационным рядом выборки.

<u>Эмпирическая функция распределения.</u> Пусть $X_1, \ldots, X_n$ – выборка из распределения $L(\xi) = F(x)$, $F(x)$ называют теоретической функцией распределения, а функцию

$$F_n(x) = \frac{\mu_n(x)}{n} = \frac{1}{n} \sum_{i=1}^{n} e(x - X_i)$$

– эмпирической, или выборочной, функцией распределения. Здесь $\mu_n(x)$ – число элементов выборки, не превосходящих x, а

$$e(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}$$ – функция Хевисайда.

Для оценки точности равенства $F_n(x) \approx F(x)$ при больших n ис-

пользуют теорему А.Н.Колмогорова: если $F(x)$ – непрерывная функция распределения, то

$$\lim_{n\to\infty} P(\sqrt{n}\ \sup|F_n(x)-F(x)|\geq\varepsilon)=K(\varepsilon),$$

где

$$K(\varepsilon)=\sum_{n=-\infty}^{\infty}(-1)^{n-1}e^{-2n^2\varepsilon^2}.$$

Грубо говоря, функция $F_n(x)$ дает равномерную оценку функции $F(x)$ с точностью до величины порядка $1/\sqrt{n}$.

Выборочные характеристики. Каждой теоретической характеристике, $g=\int g(x)\ dF(x)$, соответствует ее статистический аналог

$$G=\int g(x)\,dF_n(x)=\frac{1}{n}\sum_{i=1}^{n}g(X_i),$$

называемый эмпирической или выборочной характеристикой. В частности, выборочный момент k-го порядка есть величина

$$A_{nk}=\frac{1}{n}\sum_{i=1}^{n}X_i^k.$$

При $k=1$ величина A_{n1} – выборочное среднее – обозначается $\bar{X}$: $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. Выборочный центральный момент k-го порядка есть

$$M_{nk}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k.$$

При $k=2$ выборочную дисперсию M_{n2} обозначают S^2:

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

S – среднеквадратическое отклонение, $\tilde{A}=M_{n3}/S^3$ и $\tilde{E}=M_{n4}/S^4-3$ – соответственно выборочные коэффициенты асимметрии и эксцесса.

Из равенств

$$MA_{nk}=\frac{1}{n}\sum_{i=1}^{n}MX_i^k=M\xi^k=a_k,$$

$DA_{nk} = D\xi^k/n = (a_{2k} - a_k^2)/n$ и из центральной предельной теоремы следует следующая теорема.

Т е о р е м а I.I. Выборочный момент A_{nk} асимптотически нормален, $N(a_k, (a_{2k} - a_k^2)/n)$.

Т е о р е м а I.2. (Фишер). Пусть $X_1, \ldots, X_n$ – выборка из распределения $N(a, \sigma^2)$. Выборочные среднее $\bar{X}$ и дисперсия S^2 независимы и

$$L(\sqrt{n}(\bar{X}-a)/\sigma) = N(0,1); \; L(nS^2/\sigma^2) = \chi^2(n-1).$$

Т е о р е м а I.3. Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ – независимые выборки из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$; $S^2(X)$ и $S^2(Y)$ – соответствующие выборочные дисперсии. Тогда отношение

$$F = \frac{n(m-1)\sigma_2^2 S^2(X)}{m(n-1)\sigma_1^2 S^2(Y)}$$

при любых $a_1, a_2, \sigma_1^2, \sigma_2^2$ распределено по закону Снедекора-Фишера с n – I и m – I степенями свободы.

Для выборки из нормального распределения $N(a, \sigma^2)$ важную роль играет следующее свойство: если $X = (X_1, \ldots, X_n)$ – выборка из распределения $N(a, \sigma^2)$ и $t = BX$, $Q_i = X^T A_i X$, i = I,2, то для независимости t и Q_i достаточно выполнения условия $BA_i = 0$, а для независимости Q_1 и Q_2 – условия $A_1A_2 = A_2A_1 = 0$.

Для $p \in (0,1)$ p-квантилью ζ_p непрерывного распределения $L(\xi) = F(x)$ называют корень уравнения $F(\zeta) = p$.

Выборочной p-квантилью Z_{np} называют порядковую статистику $X_{([np]+1)}$, где $[a]$ – целая часть числа a:

$$Z_{np} = \begin{cases} X_{([np]+1)}, & \text{при } np \text{ дробном,} \\ X_{np}, & \text{при } np \text{ целом.} \end{cases}$$

Величина $Z_{n\frac{1}{2}}$ называется выборочной медианой.

Теорема I.4. Если в некоторой окрестности точки ξ_p плотность $f(x)$ непрерывно дифференцируема и $f(\xi_p)>0$, то при $n \to \infty$

$$\mathcal{L}(Z_{np}) \sim N(\xi_p, pq/nf^2(\xi_p)), \quad q=1-p,$$

т.е. Z_{np} асимптотически нормальная.

Гистограмма и полигон частот. Выборку объема n часто для удобства записывают в виде таблицы:

x_i	$x_1 \dots x_k$
n_i	$n_1 \dots n_k$

,

где n_i – частота x_i, $n_1+\dots+n_k=n$.

Для наглядного представления статистических данных строят полигон или гистограмму частот.

Полигоном частот (относительных частот) выборки объема n из дискретного распределения $\mathcal{L}(\xi)$ называют ломаную с вершинами в точках $(x_1, n_1); \dots; (x_k, n_k)$ $((x_1, n_1/n); \dots; (x_k, n_k/n))$.

Для выборки абсолютно непрерывного распределения $\mathcal{L}(\xi)$ строят гистограмму частот. Для этого интервал, в котором заключены все значения выборки x_i, разобьём на частичные интервалы длиной h. Гистограммой частот (относительных частот) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высоты равны n_i/h (n_i/nh), где n_i – сумма частот вариант, попавших в i-й интервал. При большом значении объема выборки n и малом h гистограмма является статистическим аналогом плотности распределения наблюдаемой случайной величины. Если плотность $f(x)$ достаточно гладкая, то для ее оценки строят полигон частот следующим образом: если гистограмма уже имеется, то ординаты, соответствующие средним точкам интервалов, соединяют отрезками прямых. Полученная ломаная и является полигоном частот. Она даёт приближенное представление о теоретической плотности.

Примеры

I. По выборке

x_i	-2	-I	0	3	5	6
n_i	3	I2	20	40	I8	7

построить полигон относительных частот и найти эмпирическую функцию распределения.

Р е ш е н и е . Объем выборки n = 100. Составим таблицу:

x_i	-2	-1	0	3	5	6
$0{,}01 n_i$	0,03	0,12	0,20	0,40	0,18	0,07

Построим полигон относительных частот (рис.1).

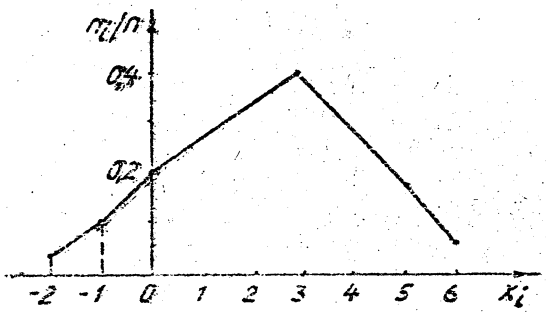


Рис.1

Найдём $F_n(x)$:

$$F_n(x) = \begin{cases} 0{,}00, & x < -2 \\ 0{,}03, & -2 \leqslant x < -1 \\ 0{,}15, & -1 \leqslant x < 0 \\ 0{,}35, & 0 \leqslant x < 3 \\ 0{,}70, & 3 \leqslant x < 5 \\ 0{,}88, & 5 \leqslant x < 6 \\ 1{,}00, & x \geqslant 6. \end{cases}$$

2. Результаты измерения роста (x_i , см) случайно отобранных 100 студентов приведены в таблице (n_i - число студентов):

x_i	154-158	158-162	162-166	166-170	170-174	174-178	178-182
n_i	8	16	22	30	14	8	2

Найти выборочное среднее и дисперсию роста студентов, построить гистограмму и полигон частот.

Р е ш е н и е . Найдём середины интервалов и примем их в качестве значений выборки.

Имеем $\overline{X} = 0{,}01(8\cdot156 + 16\cdot160 + 22\cdot164 + 30\cdot168 + 14\cdot172 + 8\cdot176 + 2\cdot180) = 166{,}32$.

$S^2 = 0{,}01(8\cdot156^2 + 16\cdot160^2 + 22\cdot164^2 + 30\cdot168^2 + 14\cdot172^2 + 8\cdot176^2 + 2\cdot180^2) = 27755$. По опытным данным строим полигон и гистограмму (рис.2).

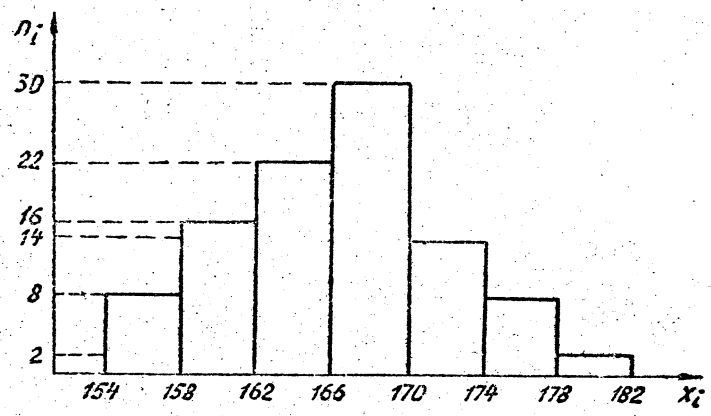


Рис.2

3. При сверлении 80 отверстий одним и тем же сверлом и последующем измерении диаметров отверстий получены следующие данные в мм :

40,26; 40,29; 40,37; 40,33; 40,28; 40,41; 40,35; 40,28; 40,29; 40,27; 40,55; 40,35; 40,41; 40,30; 40,33; 40,40; 40,34; 40,44; 40,39; 40,38; 40,44; 40,44; 40,35; 40,40; 40,31; 40,33; 40,34; 40,32; 40,39; 49,37; 40,35; 40,35; 40,30; 40,36; 40,33; 40,37; 40,31; 40,34; 40,37; 40,37; 40,39; 40,30; 40,33; 40,32; 40,36; 40,34; 40,43; 40,31; 40,38; 40,35; 40,42; 40,31; 40,33; 40,42; 40,30; 40,43; 40,34; 40,34; 40,36; 40,32; 40,40; 40,34; 40,38; 40,32; 40,34; 40,30; 40,36; 40,31; 40,37; 40,36; 40,32; 40,32; 40,33; 40,35; 40,30; 40,34; 40,34; 40,34; 40,41; 40,36.

Составить интервальные статистические ряды распределения частот. Найти эмпирическую функцию распределения и построить ее график, построить гистограмму и полигон относительных частот диаметров отверстий. Вычислить выборочные среднее $\overline{X}$, дисперсию S^2, среднеквадратическое отклонение S, коэффициенты асимметрии,

эксцесса и вариации, A, E и V. По виду гистограммы и полигона, а также по значениям A и E убедиться, что случайная величина ξ, равная значению диаметра отверстия, имеет нормальный закон распределения.

Р е ш е н и е . Разобьём весь диапазон значений на 5 интервалов. Длина частичного интервала

$$h = \frac{x_{max} - x_{min}}{5} = \frac{40,44 - 40,26}{5} = 0,04.$$

Составим статистический ряд распределения относительных частот.

Интервалы значений	40,24 - 40,28	40,28 - 40,32	40,32 - 40,36	40,36 - 40,40	40,40 - 40,44
Частости, n_i/n	0,05	0,24	0,40	0,19	0,12
$F_n(x)$	0,05	0,29	0,69	0,88	1,00

По этой таблице строим гистограмму и полигон, записываем выборочную функцию распределения $F_n(x)$ (рис. 3; 4, пунктирной линией изображен полигон относительных частот).

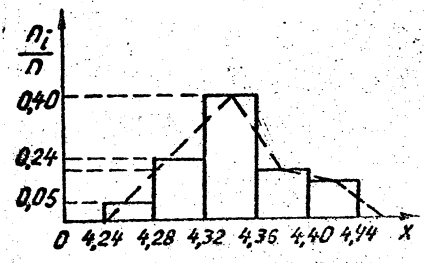


Рис. 3

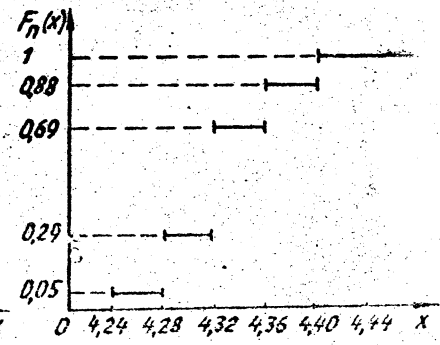


Рис. 4

$$F_n(x) = \begin{cases} 0{,}00, & x < 40{,}24 \\ 0{,}05, & 40{,}24 \leq x < 40{,}28 \\ 0{,}29, & 40{,}28 \leq x < 40{,}32 \\ 0{,}69, & 40{,}32 \leq x < 40{,}36 \\ 0{,}88, & 40{,}36 \leq x < 40{,}40 \\ 1{,}00, & x \geq 40{,}40. \end{cases}$$

Для вычисления числовых характеристик удобно пользоваться таблицей:

Интервалы значений, мм	Середины интервалов	n_i	$(x_i-\bar{x})n_i$	$(x_i-\bar{x})^2 n_i$	$(x_i-\bar{x})^3 n_i$	$(x_i-\bar{x})^4 n_i$
40,24-40,28	40,26	4	-0,336	0,028	-0,002	0,0002
40,28-40,32	40,30	19	-0,836	0,037	-0,002	0,0001
40,32-40,36	40,34	32	-0,128	0,001	0,000	0,0000
40,36-40,40	40,38	15	0,540	0,019	0,001	0,0000
40,40-40,44	40,42	10	0,760	0,058	0,004	0,0003
Суммы		80		0,143	0,001	0,0006

Используя таблицу, находим $\bar{x} = \frac{1}{80}\Sigma x_i n_i = 40{,}344$.

$$S^2 = \frac{0{,}143}{80} = 0{,}0018; \quad S = \sqrt{0{,}0018} = 0{,}042;$$

$$\tilde{A} = \frac{\Sigma(x_i-\bar{x})^3 n_i}{nS^3} = \frac{0{,}001}{80(0{,}042)^3} = 0{,}17; \quad \tilde{E} = \frac{\Sigma(x_i-\bar{x})^4 n_i}{nS^4} -$$

$$- 3 = \frac{0{,}0006}{80(0{,}042)^4} - 3 = -0{,}68. \quad V = \frac{S}{\bar{x}} 100 = \frac{0{,}042}{40{,}344} 100 =$$

$= 0{,}104\%$. Вычислим средние квадратические ошибки определения асимметрии и эксцесса:

$$S_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} = \sqrt{\frac{6 \cdot 79}{81 \cdot 83}} = 0{,}26,$$

$$S_E = \sqrt{\frac{24\,n\,(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} = \sqrt{\frac{24 \cdot 80 \cdot 78 \cdot 77}{79^2 \cdot 83 \cdot 85}} = 0{,}513.$$

Критерием распределения диаметров отверстий по нормальному закону является равенство нулю асимметрии и эксцесса. Мы видим, что вычисленные $\tilde{A}$ и $\tilde{E}$ отличаются от нуля не более чем на удвоенные среднеквадратические ошибки их определения, что соответствует нор-

мальному закону распределения.

4. Пусть $X_1, \ldots, X_n$ – выборка из абсолютно непрерывного распределения $L(\xi) = F(x)$, $F'(x) = f(x)$.
Найти: а) функцию распределения $F_{X_{(k)}}(x)$ порядковых статистик $X_{(k)}$; б) плотности экстремальных статистик.

Р е ш е н и е . Пусть $\mu_n(x)$ – число элементов выборки, не превосходящих x. События $\{X_{(k)} \le x\}$ и $\{\mu_n(x) \ge x\}$ $(1 \le k \le n)$ – эквивалентны. Учитывая, что

$$P(X_i \le x) = F(x) \quad \text{и} \quad P(\mu_n(x) = k) = C_n^k F^k_{(x)} [1 - F(x)]^{n-k},$$

получаем

а) $$F_{X_{(k)}}(x) = P(\mu_n(x) \ge k) = \sum_{m=k}^{n} C_n^m F^m(x) [1 - F(x)]^{n-m},$$

б) в частности, для $k = n$ $F_{X_{(n)}}(x) = F^n(x)$ и

$$f_{X_{(n)}}(x) = nF^{n-1}(x) f(x),$$

для $k = 1$

$$F_{X_{(1)}}(x) = \sum_{m=0}^{n} C_n^m F^m(x) [1 - F(x)]^{n-m} - [1 - F(x)]^n.$$

Так как

$$\sum_{m=0}^{n} C_n^m F^m(x) [1 - F(x)]^{n-m} = [F(x) + 1 - F(x)]^n = 1,$$

то

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n \quad \text{и} \quad f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} \cdot f(x).$$

З а д а ч и

I.I. Пусть –0,3; –0,8; 3,I; –6,2; I,2; 4,5; –3,4 – наблюдавшиеся значения случайной величины.
а) Найти выборочные среднее и дисперсию.
б) Построить эмпирическую функцию распределения $\tilde{F}(x)$ и найти значения $\tilde{F}(-6)$, $\tilde{F}(-3)$, $\tilde{F}(-I)$, $\tilde{F}(0)$, $\tilde{F}(3)$.

I.2. В эксперименте наблюдалась целочисленная случайная величина ξ. Соответствующие выборочные значения оказались равными 3, 0, 4, 3, 6, 0, 3, I. Найти выборочные среднее, дисперсию и медиану. Построить эмпирическую функцию распределения $\tilde{F}(x)$. Определить значения $\tilde{F}(0)$, $\tilde{F}(1)$, $\tilde{F}(4)$.

I.3. Нижеследующая таблица дает распределение продолжительности работы электрических ламп, в часах:

Продолжительность работы t_i	950 – I000	I000 – I050	I050 – II00	II00 – II50	II50 – I200	I200 – I250
Число n_i	2	2	3	6	7	I2

t_i	I250 – I300	I300 – I350	I350 – I400	I400 – I450	I450 – I500	I500 – I550	I550 – I600	I600 – I650
n_i	I6	20	24	27	29	29	28	25

t_i	I650 – I700	I700 – I750	I750 – I800	I800 – I850	I850 – I900	I900 – I950	I950 – 2000	2000 – 2050
n_i	2I	I6	I2	8	6	3	2	I

t_i	2050 – 2I00	
n_i	I	Всего $n = 300$

Найти среднюю продолжительность работы лампы и квадратическое отклонение.

I.4. При проведении $n = 2608$ опытов по наблюдению числа α-частиц (случайная величина ξ), излучаемых радиоактивным веществом за определенный период времени (7,5 с), получены следующие данные (n_i – число опытов, для которых число частиц $\xi = i$, $i = 0, 1, 2, \ldots$):

i	0	I	2	3	4	5	6	7	8	9	I0	II	I2	
n_i	57	203	383	525	532	408	273	I39	45	27	I0	4	2	$n = 2608$

Построить полигон относительных частот и вычислить выборочные среднее и дисперсию.

I.5. В опытах наблюдалась неотрицательная непрерывная случайная величина ξ. Ее значения, упорядоченные по величине для $n = 50$ опытов, оказались равными: 0,0I; 0,0I; 0,04; 0,I7; 0,I8; 0,22; 0,22; 0,25; 0,25; 0,29; 0,42; 0,46; 0,47; 0,47; 0,56; 0,59; 0,67; 0,68; 0,70; 0,72; 0,76; 0,78; 0,83; 0,85; 0,87; 0,93; I,00; I,0I; I,02; I,03; I,05; I,32; I,34; I,37; I,47; I,50; I,52; I,54; I,59; I,7I; I,90; 2,I0; 2,35; 2,46; 2,46; 2,50; 3,73; 4,07; 6,00.

Построить эмпирическую функцию распределения и гистограмму; сравнить гистограмму с графиком функции $ce^{-x/2}$, $x > 0$. Вычислить выборочные среднее и дисперсию. Указание: разбить интервал значений [0,6] на 6 подинтервалов с шагом $h = 1$.

1.6. Получена выборка объема $n = 100$:
0,144; 0,937; 1,787; 1,052; -0,192; 0,169; 2,623; 2,135; 1,759; 0,811; 0,724; -0,110; 1,752; -0,378; 0,417; 1,360; 1,365; 2,587; 1,621; 2,344; 1,379; 0,560; 1,858; 2,453; -0,356; 1,503; -0,134; 2,590; -0,816; 0,717; 2,468; 1,131; 1,047; 1,355; 1,162; -0,491; 0,261; -0,183; 0,467; 0,502; -0,805; 0,228; 2,286; 0,364; -0,312; -0,045; 2,559; 0,129; 0,898; 0,877; 3,285; 1,554; 1,418; 0,423; -0,489; -0,255; 1,092; 0,402; -0,051; 0,020; 0,398; 1,399; 2,121; -0,026; 1,087; 2,018; -0,437; 1,661; 1,091; 0,363; 1,229; 0,416; 1,705; 1,124; 1,341; 2,320; 0,176; -0,541; 0,837; 3,329; 2,382; -0,454; 2,537; -0,299; 1,363; 0,644; 0,975; 1,294; 3,1944; 0,605; 1,978; 1,109; 2,434; -0,094; 0,735; 0,143; -0,421; -0,773; 1,570; 0,947.

Построить ЭФР (эмпирическую функцию распределения) и гистограмму. Вычислить выборочные среднее, дисперсию, коэффициенты асимметрии и эксцесса.

Указание: разбить интервал значений на десять подинтервалов.

В задачах 1.7 – 1.17 приведены результаты наблюдений некоторой физической величины ξ, сгруппированные в интервальный статистический (вариационный) ряд частот. Требуется:

а) Найти ЭФР и построить ее график;

б) Построить гистограмму и полигон относительных частот;

в) Вычислить числовые характеристики выборки: среднее $\bar{x}$; выборочное среднее квадратическое отклонение S; выборочные коэффициенты асимметрии и эксцесса A и E; выборочный коэффициент вариации $V = 100\, S/\bar{x}$, %;

г) По виду гистограммы и полигона относительных частот, а также по значениям выборочных коэффициентов асимметрии и эксцесса убедиться, что закон распределения ξ близок к нормальному.

1.7. Даны значения пористости x_i напыленного покрытия шейки распределительного вала автомобиля ГАЗ-53:

x_i, %	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20
n_i	4	25	45	20	6

$\sum n_i = 100$.

I.8. Даны значения температуры масла (x_i) в двигателе БелАЗ при средних скоростях:

x_i, град.	44-46	46-48	48-52	52-54	54-56	56-58	
n_i	4	I3	34	32	I2	5	$\sum n_i$ = I00

I.9. Даны значения внутреннего диаметра гайки (x_i):

x_i, мм	4,00 - 4,20	4,20 - 4,40	4,40 - 4,60	4,60 - 4,80	4,80 - 5,00	
n_i	6	20	46	23	5	$\sum n_i$ = I00

I.I0. Даны отклонения диаметров валиков (x_i), обработанных на станке, от заданного размера:

x_i, мм	0 - 5	5 - I0	I0 - I5	I5 - 20	20 - 25	
n_i	I5	75	I00	50	I0	$\sum n_i$ = 250

I.II. Дана трудоемкость операции "проверка привода подачи топлива автомобиля БелАЗ" (x_i):

x_i, ч	8 - I0	I0 - I2	I2 - I4	I4 - I6	I6 - I8	
n_i	7	I6	54	I5	8	$\sum n_i$ = I00

I.I2. Даны размеры 200 деталей (x_i) после шлифовки:

x_i, мм	3,45 - 3,65	3,65 - 3,85	3,85 - 4,05	4,05 - 4,25	4,25 - 4,45
n_i	9	24	II9	43	5

I.I3. Даны размеры I00 отверстий (x_i), просверленных одним и тем же сверлом:

x_i, мм	4,I0 - 4,20	4,20 - 4,30	4,30 - 4,40	4,40 - 4,50	4,50 - 4,60
n_i	7	24	34	26	9

I.I4. Даны значения температуры масла (x_i) в двигателе автомобиля при средних скоростях:

x_i, град.	40 - 42	42 - 44	44 - 46	46 - 48	48 - 50
n_i	8	25	35	22	I0

I.I5. Даны результаты наблюдений (x_i) за сроком службы I00 однотипных станков до выхода за пределы норм точности (в месяцах двухсменной работы):

x_i	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45
n_i	9	24	35	22	10

I.16. Даны квартальные данные о среднесуточном пробеге 80 автомобилей (x_i):

x_i, км	150 - 170	170 - 190	190 - 210	210 - 230	230 - 250
n_i	4	19	32	15	10

I.17. Даны сведения о расходе воды (x_i), используемой заводом для технических нужд в течение 100 дней:

x_i, м³	8 - 10	10 - 12	12 - 14	14 - 16	16 - 18
n_i	9	24	36	23	9

I.18. Следующая последовательность дает 58 значений величины e, полученных Милликеном при определении величины заряда электрона, равной $e \cdot 10^{-10}$ ед.:

4,74; 4,75; 4,75; 4,76; 4,76; 4,76; 4,76; 4,76; 4,77; 4,77; 4,77; 4,77; 4,77; 4,77; 4,77; 4,77; 4,77; 4,77; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,78; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,79; 4,80; 4,80; 4,80; 4,80; 4,80; 4,81; 4,81; 4,81; 4,81; 4,81.

Вычислить среднее арифметическое значение $\bar{x}$ величины e, среднюю точку $(x_1+x_{58})/2$, выборочную медиану, выборочную дисперсию S^2, размах $r=x_{58}-x_1$ и выборочное среднее отклонение $d=(1/58)\sum_{i=1}^{58}|x_i-\bar{x}|$. Сравнить между собой полученные характеристики, учитывая, что для нормального распределения медиана и среднее совпадают, а между дисперсией, $\sigma^2=D\xi$, и средним отклонением, $\delta=M|\xi-M\xi|$, имеет место соотношение $\delta=\sigma\sqrt{2\pi}$.

I.19. Проводились опыты с бросанием одновременно 12 игральных костей. Наблюдалась случайная величина ξ, равная числу костей с 6 очками. Таблица наблюдавшихся данных имеет вид

i	0	1	2	3	4	5	6	7	
n_i	447	1145	1181	796	380	115	24	8	n = 4096

а) Построить полигон относительных частот n_i/n и сравнить его с графиком функции $c e^{-x^2/2}$.

б) Вычислить выборочные среднее и дисперсию, а также выборочные коэффициенты асимметрии и эксцесса.

в) Найти $P(|\bar{x}-a| \le \varepsilon)$, где $a = M\xi$, ε – вычисленное по указанным данным отклонение выборочного среднего от теоретического.

I.20. Наблюдались показания 500 наугад выбранных часов, выставленных в витринах. Пусть i – номер промежутка от i-го часа до (i + I)-го, i = 0,I,...,II, а n_i – число часов, показания которых принадлежат i-му промежутку. Результаты таким образом сгруппированных наблюдений оказались следующими:

i	0	I	2	3	4	5	6	7	8	9	I0	II	
n_i	4I	34	54	39	49	45	4I	33	37	4I	37	39	n = 500

а) Построить полигон частот и сравнить его с графиком функции $f(x) = c$, $0 \le x \le$ I2.

б) Рассматривая эти данные как независимые наблюдения над дискретной случайной величиной ξ, принимающей значения, совпадающие с серединами соответствующих интервалов: 0,5; I,5; ... ...;II,5, вычислить выборочные среднее и дисперсию.

в) Принимая, что случайная величина ξ имеет равномерное распределение, найти δ из условия $P(|\bar{x}-a| \le \delta)$ = 0,98, где $a = M\xi$, и сравнить с ним наблюдавшееся значение отклонения $|\bar{x}-a|$.

I.2I Проводились опыты с бросанием одновременно I2 игральных костей. Наблюдаемую случайную величину ξ брали равной числу костей, на которых выпало 4, 5 или 6 очков. Пусть n_i – число опытов, в которых наблюдалось значение $\xi = i$ (i = 0,I,2,...,I2). Данные для n = 4096 опытов приведены в следующей таблице:

i	0	I	2	3	4	5	6	7	8	9
n_i	0	7	60	I98	430	73I	948	847	536	257

i	I0	II	I2	
n_i	7I	II	0	n = 4096.

Взяв в качестве интервалов $\Delta = (i - 1/2,\ i + 1/2)$, $i = 0,1,\ldots,12$, построить гистограмму и полигон частот, сравнить его с графиком функции $c e^{-x^2/2}$, вычислить выборочные среднее и дисперсию, а также выборочные коэффициенты асимметрии и эксцесса.

1.22. В задаче 1.21 оценить вероятность $P(|\bar{x} - a| \le \varepsilon)$, где a – теоретическое среднее, ε – вычисленное по указанным данным отклонение выборочного среднего от теоретического.

1.23. Пусть ν_n – число успехов в n испытаниях Бернулли с вероятностью успеха $p \in (0,1)$. При больших n и заданной вероятности γ вычислить границу δ_γ, чтобы

$$P(|\nu_n/n - p| \le \delta_\gamma) = \gamma.$$

Укладываются ли в эти границы при $\gamma = 0{,}98$:

а) результаты эксперимента Бюффона: при $n = 4040$ бросаний монеты герб появился $k = 2048$ раз;

б) результаты эксперимента: из чисел 0,1,...,9 "случайно" выбиралось число. Испытания проводились $n = 10000$ раз. Числа не превосходящие 4 встретились $k = 5089$ раз?

1.24. Пусть $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ – вариационный ряд выборки из распределения $L(\xi)$ с функцией распределения $F(x)$. Для произвольных порядковых статистик $X_{(r)}$ и $X_{(s)}$, $1 \le r < s \le n$, найти совместное распределение $F_{r,s}(x, y)$. Для случая, когда $F(x)$ абсолютно непрерывна, найти совместную плотность распределения $X_{(r)}$ и $X_{(s)}$.

1.25. Для выборки объема n из абсолютно непрерывного распределения $L(\xi)$ с плотностью $f(x)$ доказать, что совместная плотность всех порядковых статистик, $X_{(1)}, \ldots, X_{(n)}$, будет

$$g(x_1, \ldots x_n) = \begin{cases} n!\, f(x_1) \ldots f(x_n) & \text{при } x_1 < x_2 < \ldots < x_n, \\ 0 & \text{в остальных случаях.} \end{cases}$$

1.26. Пусть $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ – вариационный ряд выборки из экспоненциального распределения с плотностью $f(x) = e^{-x}$, $x > 0$. Доказать, что случайные величины $Y_r = (n - r + 1)(X_{(r)} - X_{(r-1)})$, $r = 1, \ldots, n$, $X_{(0)} = 0$, независимы и одинаково распределены с плотностью $f(x)$. Найти $MX_{(k)}$ и $DX_{(k)}$

и исследовать асимптотическое поведение $MX_{(n)}$ и $DX_{(n)}$ при $n \to \infty$.

I.27. Пусть $X_1, \ldots, X_n$ – выборка из равномерного на отрезке $[a, b]$ распределения. Найти:

а) совместную плотность $\varphi(x, y)$ экстремальных значений выборки $X_{(1)}$ и $X_{(n)}$;

б) $MX_{(1)}$, $MX_{(n)}$, $DX_{(1)}$, $DX_{(n)}$, $Cov(X_{(1)}, X_{(n)})$.

I.28. Для выборки объема n из распределения $L(\xi)$ с известными моментами $a_k = M\xi^k$ найти $Cov(A_{nk}, A_{ns})$, где A_{nk} – выборочный момент k-го порядка.

I.29. Для случая $L(\xi) = R(0,1)$ найти распределения порядковых статистик $X_{(k)}$, их средние и дисперсии.

I.30. Пусть $L(\xi)$ есть распределение Вейбулла $W(a, \alpha, b)$. Найти распределение минимального значения выборки $X_{(1)}$. Вычислить $MX_{(1)}$ и $DX_{(1)}$.

I.31. Пусть $X = (X_1, \ldots, X_n)$ – выборка из $N(a, \sigma^2)$. Доказать, что $\bar{X}$ и $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ независимы. Здесь $\bar{X}$ – выборочное среднее. Получить отсюда независимость $\bar{X}$ и дисперсии S^2.

I.32. Доказать, что при $n \to \infty$

$$L\left(\frac{X^2(n) - n}{\sqrt{2n}}\right) \to N(0,1).$$

Найти $M(X_n^2)^k$.

I.33. Для выборки объема n из распределения $L(\xi)$ с известным центральным моментом 3-го порядка, $\mu_3 = M(\xi - M\xi)^3$, найти $Cov(\bar{X}, S^2)$.

I.34. Пусть $X_1, \ldots, X_n$ – независимые наблюдения над случайной величиной ξ с $M\xi = a$, $D\xi = \sigma^2 > 0$, $M\xi^4 < \infty$ и $\bar{X}$, S^2 – соответственно выборочные среднее и дисперсия. Доказать, что при $n \to \infty$

$$\sqrt{n}(\bar{X} - a)/S \to N(0,1).$$

I.35. Доказать:

а) если $L(t_n) = S(n)$, то моменты Mt_n^k существуют лишь при $k < n$ и

$$Mt_n^{2m} = \frac{1 \cdot 3 \ldots (2m-1) n^m}{(n-2)(n-4)\ldots(n-2m)}, \quad 2m < n, \quad Mt_n^{2m+1} = 0, \quad 2m+1 < n;$$

б) $S(n) \to N(0,1)$ при $n \to \infty$ и, более того, плотность $S_n(x) \to (2\pi)^{-1/2} e^{-x^2/2}$;

в) $L((1+t^2/n)^{-1}) = \beta(n/2, 1/2)$.

I.36. Доказать, что для распределения Стьюдента $S(n)$ с функцией распределения вероятностей $S_n(x)$

$$\lim_{n\to\infty} S_n(x) = \Phi(x) = 2\pi^{-1/2}\int_{-\infty}^{x} e^{-t^2/2}\,dt,$$

$$\lim_{n\to\infty} S_n(x) = \lim_{n\to\infty} S_n'(x) = (2\pi)^{-1/2} e^{-x^2/2}.$$

I.37. Доказать, что для любых фиксированных $r \geq 2$, $1 \leq k_1 < \ldots < k_r$ совместное распределение выборочных моментов $A_{nk_1}, \ldots A_{nk_r}$ при $n \to \infty$ асимптотически нормально: $N(\bar{a}, \Sigma/n)$, $\bar{a} = (a_{k_1}, \ldots, a_{k_r})$, где $\Sigma = [\sigma_{ij} = a_{k_i+k_j} - a_{k_i k_j}]_1^r$, a_k – теоретические моменты, т.е.

$$L(\sqrt{n}(A_{nk_i} - a_{ki})) \to N(0, \Sigma), \quad i = 1, \ldots, r.$$

I.38. Для заданной точки x_0, такой, что $0 < F(x_0) < 1$, и заданного числа t оценить при больших n вероятность

$$P(|F_n(x_0) - F(x_0)| \leq t/\sqrt{n}),$$

где $F_n(x)$ – ЭФР вероятностей выборки из распределения $L(\xi)$ с функцией распределения $F(x)$.

Указание. Использовать теорему Муавра-Лапласа.

I.39. Пусть $x_1 < x_2 < \ldots < x_{N-1}$ – заданные числа, такие, что $0 < F(x_1) < F(x_2) < \ldots < F(x_{N-1}) < 1$, $\mu_n(x)$ – число значений выборки из распределения $L(\xi)$ с функцией распределения $F(x)$, не превосходящих x, $\nu_i = \mu_n(x_i) - \mu_n(x_{i-1})$, $\mu_n(x_0) = 0$, $\mu_n(x_N) = n$. Показать, что случайный вектор, $\bar{\nu} = (\nu_1, \ldots, \nu_N)$, имеет полиномиальное распределение $M(n; p_1, \ldots, p_N)$, где $p_i = F(x_i) - F(x_{i-1})$, $i = 1, \ldots, N$.

I.40. Пусть $X_i = (X_{i1}, X_{i2})$, $i = 1, \ldots, n$ – независимые наблюдения над двумерной случайной величиной $\xi = (\xi_1, \xi_2)$

с функцией распределения $F(x_1, x_2)$. ЭФР определяется формулой

$$F_n(x_1, x_2) = \frac{1}{n}\sum_{i=1}^{n} e(x_1 - X_{i1})\, e(x_2 - X_{i2}),$$

где $e(x)$ – функция Хевисайда. Вычислить $MF_n(x_1, x_2)$ и $DF_n(x_1, x_2)$ и показать, что $F_n(x_1, x_2) \xrightarrow{P} F(x_1, x_2)$ при $n \to \infty$. Построить выборочный коэффициент корреляции ρ_n и показать, что

$$\rho_n \xrightarrow{P} \rho = Cov(\xi_1, \xi_2)/\sqrt{D\xi_1 D\xi_2}, \text{ если } M(\xi_1^2 \xi_2^2) < \infty,\ D\xi_i > 0,\ i = 1, 2.$$

§ 2. СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ. ТОЧЕЧНЫЕ ОЦЕНКИ

I. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения $L(\xi)$ с заданной статистической моделью $\{F\}$. Всякая функция $T(X)$ от выборки называется статистикой. Часто требуется оценить истинное значение некоторой неизвестной теоретической характеристики $g = g(F)$, т.е. построить такую статистику $T(X)$, которая являлась бы в определенном смысле приближением (оценкой) g. Наиболее распространенной мерой точности является среднеквадратическая ошибка $M[T(X) - g]^2$. Если для схемы повторных независимых наблюдений статистическая модель $\{F(x, \theta), \theta \in \Theta\}$ задаётся распределением какого-нибудь стандартного типа при неизвестном определяющем его векторном параметре $\theta = (\theta_1, \ldots, \theta_k)$, то такая модель называется параметрической, а множество Θ – параметрическим множеством. В этом случае любая теоретическая характеристика g есть функция от θ : $g = \tau(\theta)$. Оценку параметра θ (или функции $\tau(\theta)$) по выборке X будем обозначать $\tilde{\theta}_n = \tilde{\theta}(X_1, \ldots, X_n) = \tilde{\theta}(X)$ (или $\tilde{\tau}(X)$).

Оценка $\tilde{\theta}_n$ параметра θ называется: а) несмещенной, если $M\tilde{\theta}_n = \theta$; б) состоятельной, если для любого $\varepsilon > 0$

$\lim\limits_{n \to \infty} P(|\tilde{\theta}_n - \theta| > \varepsilon = 0$. Здесь под $|y|$, $y \in R^k$ понимается длина вектора y в R^k. При установлении свойства состоятельности полезен следующий критерий: если для оценки $\tilde{\tau}_n(X)$ параметрической функции $\tau(\theta)$ $M\tilde{\tau}_n(X) = \tau(\theta) + \varepsilon_n$, $D\tilde{\tau}_n(X) = \delta_n$,

где $\varepsilon_n = \varepsilon_n(\theta) \to 0$, $\delta_n = \delta_n(\theta) \to 0$ при $n \to \infty$ для всех $\theta \in \Theta$, то $\tilde{\tau}_n(X)$ - состоятельная оценка $\tau(\theta)$.

Оценка $\tilde{\theta}_n$ называется сильно состоятельной, если $\tilde{\theta}_n \to \theta$, $n \to \infty$ с вероятностью I, т.е. $P(\lim_{n\to\infty} \tilde{\theta}_n = \theta) = 1$.

Оценка $\tilde{\theta}_n$ называется асимптотически несмещенной, если

$$\lim_{n\to\infty} M\tilde{\theta}_n = \theta.$$

Для двух несмещенных оценок, $\tilde{\theta}$ и $\tilde{\theta}'$, скалярного параметра θ оценка $\tilde{\theta}$ считается эффективнее оценки $\tilde{\theta}'$, если $D\tilde{\theta} < D\tilde{\theta}'$. Несмещенная оценка $\tilde{\theta}^*$ называется оптимальной в классе несмещенных оценок параметра θ (наилучшей несмещенной оценкой), если для любой несмещенной оценки $\tilde{\theta}$ $D\tilde{\theta}^* \le D\tilde{\theta}$. Для установления эффективности оценки векторного параметра θ используют понятие дисперсионной матрицы оценки $\tilde{\theta}$: $D\tilde{\theta} = M[(\tilde{\theta}-\theta)(\tilde{\theta}-\theta)^T]$ ($(\tilde{\theta}-\theta)^T$ - строчка). Из двух несмещенных оценок, $\tilde{\theta}$ и $\tilde{\theta}'$, оценка $\tilde{\theta}$ считается эффективнее, если $D\tilde{\theta} < D\tilde{\theta}'$, т.е. для произвольного вектора $\bar{z} \in R^k$ $\bar{z}^T D\tilde{\theta}\bar{z} \le \bar{z}^T D\tilde{\theta}'\bar{z}$.

2. Пусть $x = (x_1, \dots, x_n)$ - реализация выборки $X = (X_1, \dots, X_n)$ из распределения $L(\xi)$, θ - скалярный параметр. Функция $L(x,\theta) = f(x_1,\theta)\dots f(x_n,\theta)$, рассматриваемая при фиксированном x как функция параметра $\theta \in \Theta$, называется функцией правдоподобия. Здесь $f(x_i,\theta)$ - плотность случайной величины ξ, если ξ - абсолютно непрерывная, и $f(x_i,\theta) = P(\xi = x_i)$, если ξ - дискретная. Будем предполагать, что статистическая модель регулярна. Это означает: $L(x,\theta) > 0$ при всех $x \in X$ и $\theta \in \Theta$ и дифференцируема по θ; для любой статистики $T(X)$, для которой записанные интегралы существуют и абсолютно сходятся для всех $\theta \in \Theta$,

$$\frac{\partial}{\partial\theta}\int T(x)L(x,\theta)dx = \int T(x)\frac{\partial}{\partial\theta}L(x,\theta)dx \quad (dx = dx_1,\dots dx_n),$$

интегрирование ведётся по всему выборочному пространству X. Везде для дискретных моделей интегрирование заменяется суммированием. Для функции

$$U(x,\theta) = \frac{\partial \ln L(x,\theta)}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial \ln f(x_i,\theta)}{\partial\theta},$$

называемой вкладом выборки, выполняется условие $0 < MU^2(X,\theta) < \infty$ для всех $\theta \in \Theta$.

Для регулярной модели $MU(X,\theta) = 0$ для всех $\theta \in \Theta$, а функция $i(\theta) = DU(X,\theta) = MU^2(X,\theta)$ существует и называется функцией информации (функцией Фишера).

Введенные понятия обобщаются на случай векторного параметра $\theta = (\theta_1, \ldots, \theta_k)$. В этом случае под вкладом выборки понимается случайный вектор $U = (U_1(X,\theta), \ldots, U_k(X,\theta))$, где $U_j(X,\theta) = \frac{\partial}{\partial \theta_j} \ln L(X,\theta)$, а под информационной матрицей выборки понимается $J_n = J_n(\theta) = DU = M(UU^T)$.

Т е о р е м а 2.I (Рао-Крамера). Если статистическая модель регулярна, то для любой несмещенной оценки $\tilde{\tau}(X)$ с конечной дисперсией параметрической функции $\tau(\theta)$, дифференцируемой по θ, справедливо неравенство

$$D\tilde{\tau}(X) = M|\tilde{\tau}(X) - \tilde{\tau}(\theta)|^2 \geqslant \frac{|\tau'(\theta)|^2}{i(\theta)}, \text{ если } \theta \text{ - скаляр,}$$

$$D\tilde{\tau}(X) = b^T(\theta) J_n^{-1}(\theta) b(\theta), \quad b^T(\theta) = \left(\frac{\partial \tau(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \tau(\theta)}{\partial \theta_k}\right),$$

если $\theta = (\theta_1, \ldots, \theta_k)$.

Если достигается равенство, то оценка $\tilde{\tau}(X)$ называется эффективной. Равенство достигается тогда и только тогда, когда имеет место представление

$$\tilde{\tau}(X) - \tau(\theta) = C(\theta) U(X,\theta), \text{ если } \theta \text{ = скаляр;}$$

$$\tilde{\tau}(X) - \tau(\theta) = C^T(\theta) U(X,\theta), \text{ если } \theta \text{ - вектор,}$$

где $C(\theta)$ - некоторая функция или вектор-функция θ.

Эффективная оценка будет и оптимальной оценкой.

Статистика $T(X)$ называется достаточной, если существуют положительные функции g и h такие, что функция правдоподобия $L(X,\theta)$ допускает представление

$$L(X,\theta) = g(T(\theta), \theta) h(X).$$

Оптимальная оценка, если она существует, является функцией от достаточной статистики. Эффективная статистика является достаточной статистикой. Достаточная статистика $T(X)$ называется полной,

если для всякой функции $\varphi(T(X))$ из условия $M\varphi(T)=0$ для любого θ следует $\varphi(t)\equiv 0$ на множестве значений статистики T.

Т е о р е м а 2.2. Если существует полная достаточная статистика $T=T(X)$, то всякая функция от нее, $H(T)$, является оптимальной оценкой своего математического ожидания,

$$MH(T)=\tau(\theta).$$

3. Приведём некоторые методы нахождения статистик для оценки параметров.

Метод наибольшего правдоподобия. По этому методу оценкой максимального правдоподобия $\hat{\theta}_n$ параметра θ по выборке $X=(X_1,\dots,X_n)$ является такая точка множества Θ, в которой функция правдоподобия $L(x,\theta)$ при каждом $X=x$ достигает максимума. Если для каждого $x\in X$ $L(x,\theta)$ дифференцируема по θ, то $\hat{\theta}_n$ удовлетворяет уравнению правдоподобия:

$$\frac{\partial \ln L(x,\theta)}{\partial\theta}=0 \quad \text{или} \quad \frac{\partial \ln L(x,\theta)}{\partial\theta_j}=0,\ j=1,\dots,k.$$

В случае, когда уравнение правдоподобия нельзя решить точно, то $\hat{\theta}_n$ находят приближенно, используя рекурентный метод Фишера, согласно которому ($m+1$) приближение для $\hat{\theta}_n$ вычисляется по формуле $\theta_{m+1}=\theta_m+U(x,\theta_m)/n\cdot i(\theta_m)$, $m=0,1,2,\dots$, где θ_0 – произвольная состоятельная оценка θ.

Для регулярной модели, если $\hat{\theta}_n$ существует, единственна и лежит внутри Θ, то она является состоятельной оценкой θ, причем ее распределение асимптотически нормальное:

$$L(\sqrt{n}(\hat{\theta}_n-\theta))\to N(0,J^{-1}(\theta)),\quad n\to\infty.$$

Метод моментов состоит в приравнивании теоретических моментов соответствующим выборочным моментам. Например, для оценки неизвестных параметров $\theta=(\theta_1,\dots,\theta_k)$ по выборочной функции распределения $F_n(t)$ находят первые k момента и приравнивают их соответствующим моментам, найденным по функции $F(t;\theta)$.

Примеры

1. Пусть $X_1, \ldots, X_n$ – выборка из распределения $\mathcal{L}(\xi)$ с функцией распределения $F(x)$ и $F_n(x) = \mu(x)/n$ – выборочная функция распределения, где $\mu(x)$ – число X_i меньших x. Показать, что $F_n(x)$ есть несмещенная и состоятельная оценка $F(x)$.

Решение. Для произвольного вещественного x введём события $A_k(x) = \{X_k \le x\}$. Из независимости случайных величин $X_1, \ldots, X_n$ следует независимость событий $A_1(x), \ldots, A_n(x)$. Обозначим $1_k(x)$ – индикаторную функцию множества $A_k(x)$. Случайные величины $1_1(x), \ldots, 1_n(x)$ независимы и одинаково распределены. Очевидно,

$$\mu(x) = \sum_{k=1}^{n} 1_k(x) \qquad \text{и} \qquad F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1_k(x).$$

Случайная величина $1_k(x)$ принимает лишь два значения, 0 и 1, при этом $P(1_k(x)=1) = P(A_k(x)) = F(x)$. Поэтому $M1_k(x) = F(x)$ и

$$MF_n(x) = \frac{1}{n}\sum_{k=1}^{n} M1_k(x) = F(x).$$

Далее, $D1_k(x) = F(x)[1-F(x)]$ и

$$DF_n(x) = \frac{1}{n^2}\sum_{k=1}^{n} D1_k(x) = \frac{1}{n}F(x)[1-F(x)].$$

Используя неравенство Чебышева, получаем

$$P(|F_n(x) - F(x)| > \varepsilon) \le \frac{DF_n(x)}{\varepsilon^2} = \frac{F(x)[1-F(x)]}{\varepsilon^2 \cdot n} \to 0$$

при $n \to \infty$.

2. Даны результаты независимых измерений, $x_1, \ldots, x_n$, известной постоянной величины a. Ошибки измерений подчиняются нормальному закону с нулевым математическим ожиданием. Обработка результатов ведётся по формуле

$$\tilde{\sigma} = k\sum_{j=1}^{n} |x_j - a|.$$

Каким должно быть k, чтобы была несмещенной оценка

а) среднего квадратического отклонения ошибок;

б) дисперсии ошибок?

Р е ш е н и е .

а) Введем новые случайные величины $y_j = x_j - a$; $My_j = 0$; $Dy_j = My_j^2 = \sigma^2$. Величины y_j нормальные $N(0,\sigma^2)$. Найдём $M|y_j|$.

$$M|y_j| = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}|x|e^{-x^2/2\sigma^2}dx = \frac{1}{\sqrt{2\pi}\sigma}\int_0^{\infty}e^{-x^2/2\sigma^2}dx^2 = \sqrt{\frac{2}{\pi}}\sigma.$$

Тогда $M\tilde{\sigma} = k\sum_{j=1}^{n}M|y_j| = k\sqrt{2}\sigma n/\sqrt{\pi}$. Из равенства $kn\sqrt{2}\sigma/\sqrt{\pi} = \sigma$ находим $k = \sqrt{\frac{\pi}{2}}/n$.

б) $\tilde{\sigma}^2 = k^2\left(\sum_{j=1}^{n}|x_j - a|\right)^2 = k^2\left(\sum_{j=1}^{n}|y_j|\right)^2 = k^2\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i||y_j|$.

$$M\tilde{\sigma}^2 = k^2\left[\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}M|y_i|M|y_j| + \sum_{i=1}^{n}My_i^2\right] = k^2\left[n(n-1)\frac{2}{\pi}\sigma^2 + n\sigma^2\right].$$

Из равенства $M\tilde{\sigma}^2 = \sigma^2$ находим $k = \left[\left(\frac{2}{\pi}(n-1)+1\right)n\right]^{-1/2}$.

3. По выборке $X_1,\ldots,X_n$ из распределения $\Gamma(a,\lambda)$ найти методом моментов оценки неизвестных параметров a и λ.

Р е ш е н и е . Для нахождения оценок a и λ найдём два выборочных момента, $\bar{X}$ и S^2, и приравняем их соответствующим теоретическим моментам. Имеем: $M\xi = a\lambda$, $D\xi = a^2\lambda$. Таким образом, $a\lambda = \bar{X}$; $S^2 = a^2\lambda$. Отсюда $\tilde{a} = S^2/\bar{X}$, $\tilde{\lambda} = \bar{X}^2/S^2$.

4. Пусть $x_1,\ldots,x_n$ – выборка из распределения $R(a,b)$. Оценить параметры a и b методом максимального правдоподобия. Будут ли эти оценки: а) несмещенные; б) асимптотически несмещенные?

Р е ш е н и е . Запишем функцию правдоподобия распределения $R(a,b)$:

$$L(x;a,b) = L(x_1,\ldots;a,b) = (b-a)^{-n}.$$

Оценки $\tilde{a}$ и $\tilde{b}$ найдём из условия

$$\ln L(x;\tilde{a},\tilde{b}) = \max_{a<b}[-n\ln(b-a)] = -n\ln[\min_{a<b}(b-a)].$$

Все значения $x_1,\dots,x_n$ лежат в интервале $[a,b]$, поэтому наименьшая длина интервала $[a,b]$, содержащего все x_i, будет когда

$$\tilde{a} = \min\{x_1,\dots,x_n\} \qquad \text{и} \quad \tilde{b} = \max\{x_1,\dots,x_n\}.$$

Так как случайные величины $x_1,\dots,x_n$ независимы, то

$$F_{\tilde{b}}(t) = P(\tilde{b}\le t) = P(x_1\le t,\dots,x_n\le t) = F^n(t),$$

где

$$F^n(t) = \begin{cases} 0, & t<a \\ \left(\dfrac{t-a}{b-a}\right)^n, & a\le t<b, \\ 1, & t\ge b. \end{cases}$$

Учитывая, что $\min_i\{x_i\} = -\max_i\{-x_i\}$, имеем $F_{\tilde{a}}(t) = 1-[1-F(t)]^n$. Поэтому

$$M\tilde{b} = \frac{n}{(b-a)^n}\int_a^b t(t-a)^{n-1}dt = b-\frac{b-a}{n+1}, \quad M\tilde{a} = a+\frac{b-a}{n+1}.$$

Таким образом, оценки $\tilde{a}$ и $\tilde{b}$ не являются несмещенными, но они асимптотически несмещенные.

5. Пусть $X = (x_1,\dots,x_n)$ – выборка из распределения $Bi(1,\theta)$, где $\theta\in(0,1)$ – неизвестный параметр. Показать, что статистика $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ является достаточной и полной, а также несмещенной эффективной оценкой θ.

Решение. Пусть $k = n\bar{x}$ – число "успехов" в выборке X. Запишем функцию правдоподобия, $L(X,\theta) = \theta^k(1-\theta)^{n-k}$, и вклад выборки; $U(X,\theta) = \frac{\partial}{\partial\theta}L(X,\theta) = \frac{n}{\theta(1-\theta)}(k/n-\theta)$.

Отсюда заключаем, что $\bar{X} = k/n$ есть эффективная несмещенная оценка θ, а также достаточная статистика. Случайная величина $\bar{X}$ принимает значения k/n с вероятностями $C_n^k \theta^k (1-\theta)^{n-k}$.

Пусть φ – произвольная функция, заданная на множестве значений $\bar{X}$ и для любого $\theta \in (0,1)$,

$$M\varphi(\bar{X}) = \sum_{k=0}^{n} \varphi(k/n) C_n^k \theta^k (1-\theta)^{n-k} = 0.$$

Полагая $x = \theta(1-\theta)^{-1}$, будем иметь для всех $x > 0$

$$\sum_{k=0}^{n} \varphi(k/n) C_n^k x^k = 0.$$

Откуда следует равенство нулю всех коэффициентов многочлена: $\varphi(k/n) C_n^k = 0$ или $\varphi(k/n) = 0$, $k = 0, 1, \ldots n$. Это доказывает полноту статистики $\bar{X}$.

Задачи

2.1. Пусть $x_1, \ldots, x_n$ – выборка из распределения $L(\xi)$ с дисперсией $\sigma^2 = D\xi$;

а) показать, что оценка

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

неизвестного математического ожидания $a = M\xi$ является несмещенной и состоятельной;

б) найти значение k, при котором

$$\tilde{\sigma}^2 = k \sum_{i=1}^{n} (x_i - \bar{x})^2$$

является несмещенной оценкой дисперсии σ^2.

2.2. Пусть $x_1, \ldots, x_n$ – выборка из распределения $L(\xi)$ с известным математическим ожиданием $M\xi = a$. Будет ли несмещенной оценка

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - a)^2$$

неизвестной дисперсии σ^2 ?

2.3. При обработке данных 15 испытаний спортивного самолёта были получены следующие значения его максимальной скорости: 422,2; 418,7; 425,6; 420,3; 425,8; 423,1; 431,5; 428,2; 438,3; 434,0; 411,3; 417,2; 413,6; 441,3; 423,0 м/с. Определить несмещенные оценки для математического ожидания и среднего квадратического отклонения максимальной скорости самолёта.

2.4. При обработке данных 6 испытаний спортивного катера были получены следующие значения его максимальной скорости: 27; 38; 30; 37; 35; 31 м/с. Определить несмещенные оценки для математического ожидания и среднего квадратического отклонения максимальной скорости катера.

2.5. Для определения точности измерительного прибора, систематическая ошибка которого практически равна нулю, было произведено 5 независимых измерений. Результаты измерений даны в таблице:

№	1	2	3	4	5
x_i, м	2781	2836	2807	2763	2858

Определить несмещенную оценку для дисперсии ошибок измерительного прибора, если:

а) значение измеряемой величины известно: a = 2800 м;

б) значение измеряемой величины неизвестно.

2.6. Определить несмещенную оценку для средней квадратической ошибки измерительного прибора, если при 12 независимых измерениях этим прибором базы длиной 232,38 м были получены следующие результаты: 232,5; 232,48; 232,15; 232,53; 232,45; 232,30; 232,48; 232,05; 232,45; 232,60; 232,47; 232,30 м.

2.7. Даны результаты 8 независимых измерений одной и той же величины прибором, не имеющим систематической ошибки: 369; 378; 315; 420; 385; 401; 372; 383 м. Определить несмещенную оценку дисперсии ошибок измерительного прибора, если:

а) длина измеряемой прибором базы известна: a = 375 м;

б) длина базы неизвестна.

2.8. Случайная величина ξ (ошибка измерения дальности радиодальномером) подчинена закону $R(a, b)$ с неизвестными параметрами a и b. Получено эмпирическое распределение средней ошибки n = 200 измерений дальности (x_i – средняя ошибка, n_i – количество измерений, имеющих среднюю ошибку x_i):

x_i	3	5	7	9	II	I3	I5	I7	I9	2I
n_i	2I	I6	I5	26	22	I4	2I	22	I8	25

Найти методом моментов оценки a и b.

2.9. По выборке $X = (X_1, \ldots, X_n)$ из распределения $N(a, \sigma^2)$ найти методом наибольшего правдоподобия неизвестные параметры a и σ^2.

2.I0. Устройство состоит из элементов, время безотказной работы которых подчинено закону $\Gamma(\theta; 2{,}I2)$. Испытания пяти элементов дали следующие наработки (время работы элемента в часах до отказа): 50, 75, I25, 250, 300. Методом наибольшего правдоподобия оценить неизвестный параметр θ.

2.II. При $n = I00$ испытаниям по схеме Бернулли было получено $m = 65$ "успехов". Пользуясь методом наибольшего правдоподобия, оценить θ – вероятность "успеха" в одном испытании.

2.I2. Пусть $X_1, \ldots, X_n$ – выборка из распределения Лапласа с плотностью $f(x, \theta) = 0{,}5 e^{-|x-\theta|}$, $x \in \mathbb{R}$. Методом максимального правдоподобия найти оценку $\hat{\theta}$ неизвестного параметра θ.

2.I3. Для модели $N(\theta_1, \theta_2^2)$ составить уравнения правдоподобия и найти их решения.

2.I4. Случайная величина ξ, характеризующая срок службы элементов электронной аппаратуры, имеет плотность $f(x, \theta) = (2x/\theta) e^{-x^2/\theta}$, $x \geqslant 0$. По выборке $X_1, \ldots, X_n$ найти оценку максимального правдоподобия неизвестного параметра θ.

2.I5. Найти методом моментов по выборке $X_1, \ldots, X_n$ из распределения $L(\xi)$ точечные оценки неизвестных параметров a и σ нормального распределения $N(a, \sigma^2)$.

2.I6. Устройство состоит из элементов, время безотказной работы которых подчинено распределению $\Gamma(a, \lambda)$. Испытания пяти элементов дали следующие наработки (время работы элемента в часах до отказа): 50, 75, I25, 250, 300. Найти методом моментов неизвестные параметры a и λ.

2.I7. Случайная величина ξ (время работы элемента) имеет показательное распределение с плотностью $f(x, \lambda) = \lambda e^{-\lambda x}$, $x \geqslant 0$. Ниже приведено эмпирическое распределение среднего времени работы $n = 200$ элементов:

x_i	2,5	7,5	I2,5	I7,5	22,5	27,5
n_i	I33	45	I5	4	2	I

Найти методом моментов оценку неизвестного параметра λ.

2.18. Случайная величина ξ (число нестандартных изделий в партии изделий) распределена по закону Пуассона. При проверке 200 партий получено следующее распределение нестандартных изделий (x_i – число нестандартных изделий, n_i – число партий, содержащих x_i):

x_i	0	I	2	3	4
n_i	I32	43	20	3	2

Найти методом моментов оценку неизвестного параметра распределения Пуассона.

2.I9. Случайная величина ξ (число семян сорняков в пробе зерна) распределена по закону Пуассона. Ниже приведено распределение семян сорняков в n = I000 пробах, где x_i – число семян сорняков, n_i – число проб, содержащих x_i семян сорняков:

x_i	0	I	2	3	4	5	6
n_i	405	366	I75	40	8	4	2

Найти методом моментов оценку неизвестного параметра распределения Пуассона.

2.20. Найти методом моментов оценки параметров "двойного" распределения Пуассона, задаваемого вероятностями

$$P(\xi = k) = \frac{1}{2}\left(e^{-\theta_1}\frac{\theta_1^k}{k!} - e^{-\theta_2}\frac{\theta_2^k}{k!}\right), \quad k = 1, 2, \ldots,$$

$$\theta = (\theta_1, \theta_2), \quad 0 < \theta_1 < \theta_2 .$$

Такое распределение описывает, например, число столкновений с молекулами газа в камере Вильсона частиц, получающихся при распаде ядра урана в результате бомбардировки его нейтронами.

Вычислить значения полученных оценок для следующих данных, полученных при n = 327 наблюдений над случайной величиной ξ :

i	0	I	2	3	4
n_i	28	47	8I	67	53

i	5	6	7	8	9	I0
n_i	24	I3	8	3	2	I

2.2I. Каким должно быть k , чтобы оценка

$$\tilde{\sigma}^2 = k\sum_{j=1}^{n}(x_{j+1} - x_j)^2,$$

определенная по выборке $x_1, \ldots, x_n$ из распределения $\mathcal{L}(\xi)$, была несмещенной?

2.22. Произведено n независимых измерений одной и той же неизвестной постоянной величины. Ошибки измерения подчиняются нормальному закону с математическим ожиданием, равным нулю. Дисперсию оценивают по формуле

$$\tilde{\sigma}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Найти $\mathcal{L}\tilde{\sigma}^2$. Будет ли оценка дисперсии состоятельной?

2.23. Пусть $x_1, \ldots, x_n$ – выборка из распределения $N(a, \sigma^2)$. Среднее квадратическое отклонение определяется по формуле

$$\tilde{\sigma} = k\sum_{j=1}^{n}|x_j - \bar{x}|, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

При каком k $\tilde{\sigma}$ является несмещенной оценкой σ?

2.24. Пусть $x_1, \ldots, x_n$ – выборка из равномерного на $[a,b]$ распределения $R(a,b)$, где a и b – неизвестные параметры. Центр $\theta = 0{,}5(a+b)$ оценивается по формулам

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad \theta_2 = 0{,}5\left[\max_i(x_i) + \min_i(x_i)\right].$$

а) Будут ли эти оценки несмещенными?

б) Какая из этих оценок эффективнее при больших n?

2.25. Пусть случайная величина ξ имеет равномерное распределение на отрезке $[c, c+1]$, причем c неизвестно. Было получено n независимых измерений ξ: $x_1, \ldots, x_n$. Параметр можно оценивать по формулам:

$$\tilde{c}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i - k_1, \quad \tilde{c}_2 = \max_i(x_i) - k_2 .$$

а) Найти k_1 и k_2 так, чтобы эти оценки были несмещенные.

б) Какая из этих оценок эффективнее при больших n?

2.26. Пусть $X = (x_1, \ldots, x_n)$ – выборка из распределения $R(\theta, 2\theta)$.

а) Методом максимального правдоподобия найти оценку для θ.

б) Среди оценок вида $\theta = \alpha X_{(n)} + \beta X_{(1)}$, $\alpha, \beta > 0$,

$$X_{(1)} = \min_i(X_i), \quad X_{(n)} = \max_i(X_i)$$

найти оптимальную (несмещенную с минимальной дисперсией).

2.27. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения Коши $K(\theta)$. Будет ли статистика $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ состоятельной оценкой параметра θ ?

2.28. Пусть $x_1 < x_2$ – два числа такие, что $0 < F(x_1) < F(x_2) < 1$, $F_n(x)$ – выборочная функция распределения выборки из распределения $L(\xi)$ с функцией распределения вероятностей $F(x)$. Найти

$$Cov(F_n(x_1), F_n(x_2)).$$

2.29. Для определения места судна в море используют среднее арифметическое значение измерений дальности до маяка. Какой из двух приборов, измеряющих дальность, рационально взять, если время для измерений задано, а второй прибор при вдвое большей средней квадратической ошибке требует для выполнения одного измерения втрое меньше времени?

2.30. По выборке $x_1, \ldots, x_n$ из нормального распределения $N(a, \sigma^2)$ с неизвестным параметром σ найти несмещенную и эффективную оценку дисперсии σ^2.

2.31. Для указанных в таблице регулярных моделей найти эффективные оценки τ^* соответствующих этим моделям параметрических функций $\tau(\theta)$ и дисперсии оценок $D\tau^*$.

Модель	$N(\theta,\sigma^2)$	$N(a,\theta^2)$	$\Gamma(\theta,\lambda)$	$Bi(k,\theta)$	$\Pi(\theta)$	$Bi(m,\theta)$
$\tau(\theta)$	θ	θ^2	θ	θ	θ	$\theta(1-\theta)^{-1}$

2.32. Для каждой из следующих моделей:

$$N(\theta,\sigma^2),\ N(a,\theta^2),\ \Gamma(\theta,\lambda),\ K(\theta),\ Bi(k,\theta),\ \Pi(\theta),\ \overline{Bi}(m,\theta),$$

найти функцию информации $i(\theta)$.

2.33. По выборке $x_1, \ldots, x_n$ из биномиального распределения $Bi(1,p)$ найти несмещенную и эффективную оценку неизвестного параметра p. Будет ли эта оценка состоятельной?

2.34. Пусть $L(\xi) = Bi(k\theta)$ и $n = 1$. Рассмотрим функцию вида $\tau(\theta) = \theta^r$ и $\tau_1(\theta) = (1-\theta)^r$, r – целое неотрицательное. Показать, что несмещенные оценки для $\tau(\theta)$ и $\tau_1(\theta)$ существуют лишь при $r \le k$, найти эти оценки.

2.35. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения

$B(I, \theta)$ и $\tau(\theta) = \theta^{-1}$. Показать, что несмещенных оценок не существует. Существует ли асимптотически несмещенная оценка?

2.36. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения $\Pi(\theta)$, где θ – неизвестный параметр. Показать, что статистика $n\bar{X} = \sum_{i=1}^{n} X_i$ является достаточной и полной, а статистика $\bar{X}$ – эффективной оценкой θ.

2.37. Для выборки $X = (X_1, \ldots, X_n)$ из распределения $\Gamma(a, \theta)$ показать, что статистика

$$T(X) = \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \ln a$$

является эффективной оценкой $\tau(\theta) = \Gamma'(\theta)/\Gamma(\theta)$. Найти DT.

2.38. Доказать, что в модели $N(a, \theta^2)$ не существует эффективной оценки для $\tau(\theta) = \theta$; статистика

$$T(X) = \frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n} |X_i - a|$$

является несмещенной оценкой θ. Найти DT.

2.39. Для выборки $X = (X_1, \ldots, X_n)$ из логистического распределения с плотностью $f(x, \theta) = e^{-x+\theta}(1+e^{-x+\theta})$, $x \in \mathbb{R}$ показать, что статистика $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ является несмещенной, но не эффективной оценкой θ. Найти $D\bar{X}$.

2.40. Показать, что для выборки $X = (X_1, \ldots, X_n)$ из распределения $N(\theta, \sigma^2)$ статистика $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ является полной и достаточной.

2.41. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения $\Gamma(\theta, \lambda)$. Показать, что $T(X) = \sum_{i=1}^{n} X_i$ – полная достаточная статистика для θ.

2.42. Продолжительность горения электрических ламп имеет распределение $\Gamma(\theta, I)$. Чтобы оценить θ, берется выборка из n ламп и наблюдается:

а) "время жизни" первой перегоревшей лампы $X_{(1)}$; найти λ, при которой оценка $T = \lambda X_{(1)}$ будет оптимальной;

б) "времена жизни" первых k перегоревших ламп $X_{(1)} < X_{(2)} < \ldots < X_{(k)}$. Построить оптимальную оценку вида

$$T(X) = \sum_{i=1}^{k} \lambda_i X_{(i)} .$$

2.43. Для выборки $X_1, \ldots, X_n$ из распределения $\Gamma(\theta, \lambda)$ найти оценку максимального правдоподобия $\tilde{\tau}_n$ параметрической функции $\tau(\theta) = \theta^{-1}$. Проверить состоятельность этой оценки и найти ее предельный при $n \to \infty$ закон распределения.

2.44. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения $L(\xi)$ Вейбулла с плотностью

$$f(x, \theta) = \frac{\lambda x^{\lambda-1}}{\theta^{\lambda}} e^{-(x/\theta)^{\lambda}}, \quad \lambda > 0, \quad x \geq 0.$$

а) Показать, что статистика $T(X) = \sum_{i=1}^{n} X_i^{\lambda}$ – полная и достаточная.

б) Найти оптимальную оценку для функции $\tau(\theta) = M\varphi(\xi)$, где φ – заданная функция.

2.45. Пусть $X = (X_1, \ldots, X_n)$ – выборка из двухпараметрического экспоненциального распределения с плотностью

$$f(x, \theta) = \theta_2^{-1} e^{-(x-\theta_1)/\theta_2}, \quad x \geq \theta_1,$$

где $\theta = (\theta_1, \theta_2)$ – неизвестный векторный параметр. Показать, что достаточной статистикой является пара

$$(X_{(1)}, \bar{X}), \quad X_{(1)} = \min_i X_i, \quad \bar{X} = n^{-1} \sum_{i=1}^{n} X_i .$$

2.46. Пусть $X = ((X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2}))$ – выборка из распределения двумерной случайной величины $\xi = (\xi_1, \xi_2)$. Доказать, что несмещенной оценкой для $\mu_{11} = cov(\xi_1, \xi_2)$ является статистика $T(X) = \frac{n}{n-1} S_{12}$, где S_{12} – выборочная ковариация.

§ 3. СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ

I. Доверительным интервалом скалярного параметра θ называется интервал (T_1, T_2), который с заданной доверительной вероятностью накрывает оцениваемый параметр θ, т.е.

$P(\theta \in (T_1, T_2)) \geq \gamma$, где $T_i = T_i(X)$, $i = 1,2$; $T_1 < T_2$; T_i – две статистики.

Аналогично определяют доверительный интервал для параметрической функции $\tau(\theta)$ в случае векторного параметра θ :

$$P(\tau(\theta)\in(T_1,T_2))\geq\gamma.$$

Длина интервала T_2-T_1 характеризует точность оцениваемого параметра θ ; число $\gamma\in(0,1)$ называется надёжностью. На практике величину γ выбирают близкой к I (γ = 0,90; 0,95; 0,99).

Симметричный доверительный интервал определяется условием $P(|\theta-\tilde{\theta}|\leq\varepsilon)=\gamma$, где $\tilde{\theta}=\tilde{\theta}(X)$ - оценка параметра θ.

2. Пусть $X=(X_1,\ldots,X_n)$ - выборка из распределения $N(a,\sigma^2)$,

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\ ,\quad \tilde{\sigma}^2=\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

а) Доверительный интервал для неизвестного математического ожидания a при известном σ определяется из равенства

$$\gamma=P(|\bar{X}-a|\leq\varepsilon)=2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)-1$$

($\Phi(x)$ - функция Лапласа), при неизвестном σ - из равенств

$$\gamma=\int_{-t_\gamma}^{t_\gamma}S_n(t)\,dt\ ,\quad t_\gamma=2\sqrt{n}/\tilde{\sigma},$$

где $S_n(t)$ - плотность распределения Стьюдента, ее значения даны в табл. 9.

б) Доверительный интервал для неизвестного параметра σ определяется из равенств

$$\gamma=P(|\tilde{\sigma}-\sigma|\leq\varepsilon)=\int_{\sqrt{k}/(1+q)}^{\sqrt{k}/(1-q)}P_k(t)\,dt\ ,\ k=n-1\ ,\ q=\varepsilon/\tilde{\sigma},$$

где $P_k(t)$ - плотность распределения $\chi_k=\sqrt{\chi_k^2}$. Значения интеграла от плотности $P_k(t)$ по известным k и q приведены в табл. 5.

Доверительный интервал для σ $(\gamma_1\tilde{\sigma}\ ,\ \gamma_2\tilde{\sigma})$, вероятности выхода σ за правую и левую границы которого одинаковы и равны δ = = 0,5(I - γ), определяется формулой

$$P(\sigma/\tilde{\sigma}\leq\gamma_1)=P(\sigma/\tilde{\sigma}\geq\gamma_2)=P(\chi^2\geq(n-1)/\gamma_1^2)=0{,}5(1-\gamma).$$

Для определения γ_1 и γ_2 по доверительной вероятности γ и числу степеней свободы $k = n - 1$ используют табл. 9.

3. Пусть $X_1, \ldots, X_n$ – выборка из экспоненциального распределения с неизвестным параметром λ. Доверительный интервал для математического ожидания $a = \lambda^{-1}$ $(\nu_1 \bar{X}, \nu_2 \bar{X})$ определяется выражением

$$P(a/\bar{X} \leq \nu_1) = P(a/\bar{X} \geq \nu_2) = P(\chi^2 \leq \chi^2_{1-\delta}) = P(\chi^2 \geq \chi^2_{\delta}) = \delta,$$

$$\delta = 0{,}5(1-\gamma), \quad \nu_1 = 2n/\chi^2_{\delta}, \quad \nu_2 = 2n/\chi^2_{1-\delta}.$$

Значения χ^2_{δ} и $\chi^2_{1-\delta}$ находят из табл. 7 для вероятностей δ и $1 - \delta$ и числа степеней свободы $k = 2n$. При $n > 15$ границы доверительного интервала для a приближенно рассчитываются по формулам

$$4\bar{X}/(\sqrt{4n-1} + \varepsilon_0)^2, \quad 4\bar{X}/(\sqrt{4n-1} - \varepsilon_0)^2, \quad \varepsilon_0 = \Phi^{-1}(\gamma).$$

4. Пусть $X_1, \ldots, X_n$ – выборка из распределения $\Pi(\lambda)$, где λ – неизвестный параметр, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Если $\bar{X} > 0$, то доверительный интервал (λ_1, λ_2) с надёжностью γ определяется из соотношения

$$P(2n\bar{X} > \chi^2_{\delta}) = P(2n\bar{X} < \chi^2_{1-\delta}) = \delta = 0{,}5(1-\gamma),$$

т.е. λ_1 и λ_2 равны соответственно наименьшему и наибольшему из чисел $\chi^2_{\delta}/2n\bar{X}$ и $\chi^2_{1-\delta}/2n\bar{X}$. Числа $\chi^2_{1-\delta}$ и χ^2_{δ} при заданном δ находятся из табл. 7, причем $\chi^2_{1-\delta}$ для числа степеней свободы $k = 2n\bar{X}$, а χ^2_{δ} – для $k = 2(n\bar{X} + 1)$.

Если $\bar{X} = 0$, то доверительный интервал будет $(0, \chi^2_{\delta}/2n)$, где χ^2_{δ} находится для $k = 2$ и $P(2n\bar{X} > \chi^2_{\delta}) = 2\delta$.

5. Пусть $X_1, \ldots, X_n$ – выборка из распределения $Bi(1, p)$, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Если n велико, то доверительный интервал параметра p с доверительной вероятностью γ есть $(\bar{X} - \varepsilon, \bar{X} + \varepsilon)$, где ε является решением уравнения $\gamma = \Phi(\varepsilon\sqrt{n}/\sqrt{\bar{X}(1-\bar{X})})$.

6. Если известна функция распределения $F(t, \theta)$ оценки $\tilde{\theta}$ неизвестного параметра θ и она непрерывна и монотонна по θ, то центральный γ - доверительный интервал (θ_1, θ_2) для θ определяется из уравнений

$$F(\tilde{\theta}, \theta) = 0{,}5(1-\gamma), \quad 1-F(\tilde{\theta}-0, \theta) = 0{,}5(1-\gamma).$$

Для больших выборок если параметрическая функция $\tau(\theta)$, $\theta = (\theta_1, \ldots, \theta_k)$, непрерывно дифференцируемая и $\hat{\tau}_n = \tau(\hat{\theta}_n)$ - ее оценка максимального правдоподобия, в случае регулярной модели асимптотическим γ -доверительным интервалом для $\tau(\theta)$ является интервал $(\hat{\tau}_n - C_\gamma \sigma_\tau(\hat{\theta}_n)/\sqrt{n}, \hat{\tau}_n + C_\gamma \sigma_\tau(\hat{\theta}_n)/\sqrt{n})$, где

$$\sigma_\tau^2(\theta) = b^T(\theta) J^{-1}(\theta) b(\theta), \quad b(\theta) = \left(\frac{\partial \tau(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \tau(\theta)}{\partial \theta_k}\right), C_\gamma = \Phi^{-1}(0{,}5(1+\gamma)).$$

В частности, асимптотическим γ -доверительным интервалом для скалярного параметра θ является интервал

$$(\hat{\theta}_n - C_\gamma/\sqrt{n \cdot i(\hat{\theta}_n)}, \quad \hat{\theta}_n + C_\gamma/\sqrt{n \cdot i(\hat{\theta}_n)}).$$

Примеры

1. Определение скорости самолёта было проведено на 9 испытаниях, в результате которых вычислено подходящее значение скорости: $\tilde{V}$ = 870,3 м/с. Найти 95% доверительный интервал скорости самолёта, если известно, что рассеивание скорости подчинено нормальному закону с среднеквадратическим отклонением σ = 2 м/с.

Решение. По табл. 3 из уравнения $2\Phi_1(\varepsilon\sqrt{n}/\sigma) = 0{,}95$ находим $\varepsilon\sqrt{n}/\sigma = 1{,}96$, откуда $\varepsilon = 1{,}96 \cdot 2/3 = 1{,}30$ и искомый доверительный интервал есть $[\tilde{V} - \varepsilon, \tilde{V} + \varepsilon] = [869{,}0; 871{,}6]$, м/с.

2. Даны результаты испытания на разрыв 100 образцов дюралюминия (x_i - предел прочности на разрыв, кг/мм^2):

x_i	40 - 42	42 - 44	44 - 46	46 - 48	48 - 50
ν_i	7	25	37	23	8

Считая, что предел прочности на разрыв есть нормальная случайная величина $N(a, \sigma^2)$, найти:

а) доверительные интервалы для математического ожидания

и среднеквадратического отклонения σ, которые с вероятностью $\gamma = 0{,}95$ покрывают истинные значения определяемых параметров;

б) с какой вероятностью можно утверждать, что абсолютное значение ошибки определения a не превзойдёт 0,5 кг/мм^2, а ошибка в определении σ будет меньше 0,3 кг/мм^2?

Р е ш е н и е . Найдём несмещенные оценки параметров a и σ, приняв за значения x_i середины интервалов:

$$\bar{x} = \frac{1}{100}(7\cdot 41 + 25\cdot 43 + 37\cdot 45 + 23\cdot 47 + 8\cdot 49) = 44;$$

$$\tilde{\sigma}^2 = \frac{1}{99}(7\cdot 9 + 25\cdot 1 + 37\cdot 1 + 23\cdot 9 + 8\cdot 25) \approx 5{,}37; \quad \tilde{\sigma} = 2{,}32.$$

а) Из уравнения

$$0{,}95 = \int_{-t_\gamma}^{t_\gamma} S_n(t)\,dt,$$

используя табл. 9, находим $t_\gamma = \varepsilon\sqrt{n}/\tilde{\sigma} = 1{,}98$. Откуда $\varepsilon = 1{,}98\cdot 2{,}32/10 = 0{,}46$ и доверительный интервал для a: $[\bar{x}-\varepsilon, \bar{x}+\varepsilon] = [43,54;\ 44,46]$. Установим доверительный интервал для σ. По табл. 9 по $k = n - 1 = 99$ и $\gamma = 0{,}95$ находим $\gamma_1 = 0{,}90$; $\gamma_2 = 1{,}135$. Таким образом, $[\gamma_1\tilde{\sigma}, \gamma_2\tilde{\sigma}] = [2{,}09;\ 2{,}62]$.

б) Так как $t_\gamma = \varepsilon\sqrt{n}/\tilde{\sigma} = 0{,}5\cdot 10/2{,}32 = 2{,}1$, то из табл. 9 находим ($n = 100$):

$$P(|\bar{x}-a| < 0{,}5) = \int_{-2{,}1}^{2{,}1} S_n(t)\,dt = 0{,}96.$$

Найдём вероятность

$$P(|\tilde{\sigma}-\sigma| < 0{,}3) = \int_{\sqrt{k}/(1+q)}^{\sqrt{k}/(1-q)} P_k(t)\,dt = P(\sqrt{k}/(1+q) < \chi < \sqrt{k}/(1-q)),$$

где $q = \varepsilon/\tilde{\sigma} = 0{,}13$. Из табл. 5 находим $P(|\tilde{\sigma}-\sigma| < 0{,}3) \geq 0{,}8$.

3. Глубина моря измеряется прибором, систематическая ошибка которого равна нулю, а случайные ошибки распределены нормально со среднеквадратическим отклонением $\sigma = 30$ м. Сколько надо сделать независимых измерений, чтобы определить глубину с ошибкой не более 15 м при доверительной вероятности $\gamma = 0{,}9$?

Р е ш е н и е . Пусть a – истинная глубина, n – количество измерений. Случайная величина $\xi_n = \bar{x}_n - a = n^{-1}\sum_{i=1}^{n} x_i - a$ нормальная, $M\xi_n = 0$, $D\xi_n = \sigma^2/n$. Поэтому

$$P(|X-a|<15)=2\Phi\left(\frac{15\sqrt{n}}{\sigma}\right)-1=0,9.$$

Из табл. 3 находим I5 $\sqrt{n}$ /30 = I,65. Отсюда $\sqrt{n} \geqslant 2\cdot$I,65 $\geqslant$ 3,5, или $n >$ I2.

4. Пусть $\vec{X} = (X_1, \ldots, X_n)$ – выборка из распределения $N(\theta, \sigma^2)$, $\bar{X}$ – выборочное среднее.

а) Показать, что для любых чисел $t_1 < t_2$, удовлетворяющих условию $\Phi(t_2) - \Phi(t_1) = \gamma$, интервал $\Delta_\gamma(\vec{X}) = (\bar{X} - \sigma t_2/\sqrt{n}, \bar{X} - \sigma t_1/\sqrt{n})$ является γ-доверительным для параметра θ. Найти среди этих интервалов интервал с наименьшей длиной.

б) Сколько надо произвести наблюдений $n = n(\ell, \gamma)$, чтобы длина наикратчайшего доверительного интервала при доверительной вероятности γ была равна ℓ? Вычислить $n(\ell, \gamma)$ при γ = = 0,99, ℓ = 0,5 и ℓ= 0,I; σ = I.

Р е ш е н и е .

а) Так как $M\bar{X} = \theta$ и $D\bar{X} = \sigma^2/n$, то статистика $\sqrt{n}(\bar{X}-\theta)/\sigma$ имеет распределение $N(0,I)$.

$$P(t_1 < \sqrt{n}(\bar{X}-\theta)/\sigma < t_2) = \Phi(t_2) - \Phi(t_1) = \gamma.$$

Но $P(t_1 < \sqrt{n}(\bar{X}-\theta)/\sigma < t_2) = P(\bar{X} - \sigma t_2/\sqrt{n} < \theta < \bar{X} - \sigma t_1/\sqrt{n})$.

Длина интервала $\Delta_\gamma(\vec{X})$ равна $\sigma(t_2 - t_1)/\sqrt{n} = \ell$. Чтобы получить интервал с наименьшей длиной, нужно найти такие числа $t_1 < t_2$, удовлетворяющие условию $\Phi(t_2) - \Phi(t_1) = \gamma$, при которых разность $t_2 - t_1$ была бы наименьшей. Для нахождения условного экстремума составим функцию Лагранжа:

$$H(x, y, \lambda) = y - x + \lambda[\Phi(y) - \Phi(x) - \gamma].$$

Приравнивая нулю все ее частные производные, получим систему $\Phi'(x) = \Phi'(y)$ и $\Phi(y) - \Phi(x) = \gamma$. Так как $\Phi'(x) = (\sqrt{2\pi})^{-I} e^{-x^2/2}$, то первое уравнение имеет вид $e^{-x^2/2} = e^{-y^2/2}$, откуда $y = -x$. В силу того, что $\Phi(-y) = I - \Phi(x)$, то из второго уравнения получаем $2\Phi(y) - I = \gamma$ или $\Phi(y) = 0,5(I + \gamma)$, откуда $y = t_\gamma$. Таким образом, дове-

рительный интервал с наименьшей длиной будет ($\bar{x} - \sigma C_\gamma / \sqrt{n}$, $\bar{x} + \sigma C_\gamma / \sqrt{n}$).

б) $\ell = 2\sigma C_\gamma / \sqrt{n}$ – длина доверительного интервала с наименьшей длиной. Отсюда $\sqrt{n} = 2\sigma C_\gamma / \ell$ и $n = [4\sigma^2 C_\gamma^2 / \ell^2]$ ($[a]$ – целая часть числа a). В частности $C_{0,99} = 2,576$ и для $\ell = 0,5$ (при $\sigma = 1$) $n = 106$, а для $\ell = 0,1$ $n = 2653$.

З а д а ч и

3.1. Постоянная величина измерена 25 раз с помощью прибора, систематическая ошибка которого равна нулю, а случайные ошибки измерения распределены нормально со средним квадратическим отклонением $\sigma = 10$ м. Определить доверительный интервал для значения измеряемой величины при доверительной вероятности 0,99, если $\bar{x} = 100$ м.

3.2. Среднее значение дальности до ориентира, полученное по четырем независимым измерениям, $\bar{x} = 2250$ м. Средняя квадратическая ошибка прибора $\sigma = 40$ м. Найти с надёжностью $\gamma = 0,95$ доверительный интервал для оценки подходящего значения измеряемой величины.

3.3. В качестве оценки расстояния до навигационного знака принимают среднее арифметическое результатов независимых измерений расстояния n дальномерами. Измерения не содержат систематической ошибки, а случайные ошибки распределены нормально со среднеквадратическим отклонением $\sigma = 10$ м. Сколько надо иметь дальномеров, чтобы абсолютная величина ошибки при определении дальности до навигационного знака с вероятностью 0,9 не превышала 15 м?

3.4. На контрольных испытаниях 16 осветительных ламп были определены подходящие значения для математического ожидания и среднего квадратического отклонения их срока службы: $\tilde{a} = 3000$, $\tilde{\sigma} = 20$ ч. Считая, что срок службы каждой лампы является нормальной случайной величиной, определить:

а) доверительный интервал для математического ожидания a и среднеквадратического отклонения σ, которые с вероятностью $\gamma = 0,9$ накрывают истинное значение определяемых параметров;

б) с какой вероятностью можно утверждать, что абсолютное значение ошибки определения a не превзойдёт 10 ч, а ошибка в определении σ будет меньше 2 ч ?

3.5. На основании 100 опытов было определено, что в среднем

для производства детали требуется $\tilde{t}$ = 5,5 с, а $\tilde{\sigma}$ = 1,7 с. Предположив, что время для производства детали есть нормальная случайная величина, определить границы, в которых лежат истинные значения для математического ожидания t и σ с доверительной вероятностью 85% и 90% соответственно.

3.6. Произведено 5 независимых равноточных измерений для определения заряда электрона. Опыты дали следующие результаты (в абсолютных электростатических единицах): $4{,}781\ 10^{-10}$; $4{,}795\ 10^{-10}$; $4{,}769\ 10^{-10}$; $4{,}792\ 10^{-10}$; $4{,}799\ 10^{-10}$. Определить оценку величины заряда электрона и найти доверительные интервалы при доверительной вероятности γ = 0,99, считая, что ошибки нормальны и измерения не имеют систематических ошибок.

3.7. Результаты 11 измерений постоянной величины: 9,9; 12,5; 10,3; 9,2; 6,0; 10,9; 10,3; 11,8; 11,6; 9,8; 14,0. Ошибки измерений распределены по нормальному закону, систематические ошибки отсутствуют. Определить:

а) вероятность того, что абсолютное значение ошибки в определении истинного значения измеряемой величины меньше 2% от среднего значения измерений $\bar{x}$;

б) вероятность того, что абсолютное значение ошибки определения среднего квадратического отклонения меньше 1 % от выборочного значения $\hat{\sigma}$.

3.8. По 15 независимым измерениям были рассчитаны оценки математического ожидания и среднего квадратического отклонения максимальной скорости самолёта $\tilde{V}$ = 424,7 м/с и $\tilde{\sigma}$ = 8,7 м/с. Определить:

а) доверительные интервалы для математического ожидания и среднего квадратического отклонения при доверительной вероятности 0,9;

б) вероятности, с которыми можно утверждать, что абсолютное значение ошибки в определении V и σ не превзойдёт 2 м/с (считать выборку нормальной).

В задачах 3.9 – 3.13 приведены результаты наблюдений некоторой случайной физической величины ξ, сгруппированные в интервальный статистический ряд частот. Считая, что ξ – нормальная случайная величина $N(a, \sigma^2)$, найти:

а) доверительные интервалы для математического ожидания a и среднеквадратического отклонения σ, которые с вероятностью γ = 0,90 покрывают истинные значения определяемых параметров;

б) с какой вероятностью можно утверждать, что абсолютное значение ошибки определения a не превзойдёт ε_a, а ошибка в

определении σ будет меньше ε_σ ? Указание. В качестве значений x_i взять середины интервалов.

3.9. Даны результаты исследований грануляции партии порошка (x_i , мкм):

x_i	0-40	40-80	80-120	120-160	160-200
n_i	3	12	18	13	4

ε_a = 10 мкм; ε_σ = 4 мкм.

3.10. Даны результаты исследования 49 определенных образцов на прочность напыленного слоя (x_i , кг/мм^2):

x_i	2,0-2,2	2,2-2,4	2,4-2,6	2,6-2,8	2,8-3,0
n_i	3	12	19	13	2

ε_a = 0,06 кг/мм^2; ε_σ = 0,03 кг/мм^2.

3.11. Даны результаты измерения диаметров валиков (x_i, мм):

x_i	9,74-9,76	9,76-9,78	9,78-9,80	9,80-9,82	9,82-9,84
n_i	3	8	17	16	6

ε_a = 0,005; ε_σ = 0,003.

3.12. Даны статистические данные о среднесуточном пробеге (x_i, сотни км) 49 автомобилей ЗИЛ-130:

x_i	1,2-1,6	1,6-2,0	2,0-2,4	2,4-2,8	2,8-3,2
n_i	5	12	24	6	2

ε_a = 0,1; ε_σ = 0,09.

3.13. Даны результаты измерения твердости по шкале HRC сверл (x_i):

x_i	20-30	30-40	40-50	50-60	60-70
n_i	3	6	23	14	4

ε_a = 2,3; ε_σ = 1,5.

3.14. Произведено 40 измерений базы постоянной длины a . По результатам опыта получены оценки измеряемой величины и среднеквадратического отклонения: $\bar{x}$ = 10400 м и $\tilde{\sigma}$ = 85 м. Ошибки измерения подчиняются закону нормального распределения. Найти вероятности того, что доверительные интервалы со случайными границами (0,999 $\bar{x}$; 1,001 $\bar{x}$) и (0,95 $\tilde{\sigma}$; 1,056 $\tilde{\sigma}$) соответствен-

но накроют неизвестные параметры a и σ.

3.15. Средняя квадратическая ошибка одного высотомера σ = 15 м. Сколько надо иметь таких приборов на самолёте, чтобы с надёжностью γ = 0,99 ошибка средней высоты $\bar{x}$ была бы больше ~30 м, если ошибки высотомеров нормальны и не имеют систематической ошибки?

3.16. Известно, что измерительный прибор не имеет систематических ошибок, а случайные ошибки каждого измерения подчиняются одному и тому же закону нормального распределения. Сколько надо произвести измерений для определения оценки среднего квадратического отклонения прибора, чтобы с доверительной вероятностью γ = = 0,7 абсолютное значение ошибки в определении этой величины было не более 20 % от выборочного $\bar{\sigma}$?

3.17. Найти минимальный объем выборки из распределения $N(a,\sigma^2)$, при котором с надёжностью 0,925 точность оценки a по выборочной средней равна 0,2, если известно, что σ = 1,5.

3.18. Найти минимальный объем выборки из распределения $N(a,\sigma^2)$, при котором с надёжностью 0,975 точность оценки a по выборочной средней ε = 0,3, если известно, что σ = 1,2.

3.19. В двух лабораториях определялась концентрация серы (в %) в стандартном образце дизельного топлива. Шесть независимых равноточных измерений в первой лаборатории были следующие: 0,869; 0,874; 0,867; 0,875; 0,870; 0,869. В результате аналогичных пяти равноточных измерений во второй лаборатории были получены такие значения: 0,865; 0,870; 0,866; 0,871; 0,868. Предполагая справедливым нормальный закон ошибок измерений, построить доверительный интервал с надёжностью 0,95 для отношения дисперсий измерений в 1-й и 2-й лабораториях. Если есть основания считать эти дисперсии одинаковыми, то рассчитать аналогичный интервал для разности систематических ошибок, допускаемых в обеих лабораториях.

3.20. В процессе независимых испытаний 30 образцов изделия у 10 из них наблюдался отказ. Определить доверительный интервал для вероятности отказа при доверительной вероятности 0,95, если число отказов имеет биномиальное распределение.

3.21. Стрелок А при 10 выстрелах попал в цель пять раз, а стрелок В после 100 выстрелов по той же цели имел пятьдесят попаданий. Найти доверительные интервалы для вероятности попадания в цель каждым стрелком одним выстрелом при доверительной вероятности 0,99, если число попаданий в цель имеет биномиальное распределение.

3.22. При испытаниях каждого из 10 приборов не наблюдалось ни одного отказа. Найти доверительный интервал для вероятности отказа

при доверительной вероятности 0,8; 0,9, если число отказов имеет биномиальное распределение.

3.23. Произведено 100 независимых опытов, в результате которых событие А наблюдалось 40 раз. Определить доверительный интервал для вероятности появления этого события в одном опыте при доверительной вероятности γ =0,95, если число появлений события А имеет биномиальное распределение.

3.24. При исследовании содержания повилики в семенах клевера было установлено, что выборка весом 100 г не содержит семян повилики. Найти доверительный интервал для среднего числа семян повилики в выборке весом 100 г при доверительной вероятности γ = 0,99, если число семян повилики, содержащихся в семенах клевера, имеет распределение Пуассона.

3.25. Производятся независимые испытания с одинаковой, но неизвестной вероятностью p появления события А в каждом испытании. Найти доверительный интервал для оценок p с надёжностью 0,99, если в 100 испытаниях событие А появилось 60 раз.

3.26. При испытаниях 1000 элементов зарегистрировано 100 отказов. Найти доверительный интервал, покрывающий неизвестную вероятность p отказа элемента с надёжностью: а) 0,95; б) 0,99.

3.27. Изготовлен экспериментальный игровой автомат, который должен обеспечить появление выигрыша в одном случае из 100 бросаний монеты в автомат. Для проверки пригодности автомата произведено 400 испытаний, причем выигрыш появился 5 раз. Найти доверительный интервал, покрывающий неизвестную вероятность появления выигрыша с надёжностью γ = 0,999.

3.28. Среди 250 деталей, изготовленных станком-автоматом, оказалось 32 нестандартных. Найти доверительный интервал, покрывающий с надёжностью 0,99 неизвестную вероятность p изготовления станком нестандартной детали.

3.29. За время испытаний, равное Т часов, в шести однотипных приборах было отмечено 12 отказов. Найти доверительный интервал для математического ожидания числа отказов за Т часов работы такого прибора при доверительной вероятности 0,9, если число отказов для проверяемых приборов имеет закон распределения Пуассона.

3.30. В результате 50 независимых испытаний трех типов приборов А, В и С в течение определенного промежутка времени фиксировалось число отказов (x_i). Найти доверительные интервалы для математических ожиданий числа отказов каждого типа приборов за

выбранный промежуток времени при доверительной вероятности γ = = 0,9, если число отказов для каждого типа приборов подчиняется закону распределения Пуассона.

	x_i	0	I	2	3	4
A	n_i	38	I2	0	0	0
B	n_i	4	I6	20	6	4
C	n_i	50	0	0	0	0

3.31. Случайно отобранная партия из 8 приборов была подвергнута испытаниям на срок безотказной работы. Количество часов, проработанных каждым прибором до выхода его из строя, оказалось равным I00, I70, 400, 250, 520, 680, I500, I200. Определить доверительный интервал при доверительной вероятности 0,8 для средней продолжительности работы прибора, если время безотказной работы прибора имеет экспоненциальное распределение.

3.32. При испытании I0 однотипных приборов зарегистрирован момент выхода каждого прибора из строя. Результаты наблюдений даны в таблице:

№ прибора	I	2	3	4	5
t_i, ч	200	350	600	450	400

№ прибора	6	7	8	9	I0
t_i, ч	400	500	350	450	550

Определить доверительный интервал математического ожидания T-времени безотказной работы прибора при доверительной вероятности 0,9, если T имеет экспоненциальное распределение.

3.33. Плотность вероятности для времени T между последовательными отказами радиоэлектронной аппаратуры задана формулой $f(t) = a^{-1} e^{-t/a}$. При определении оценки параметра a наблюдалось 25 отказов, а общая продолжительность безотказной работы с начала испытаний до последнего отказа $\sum_{j=1}^{25} t_j$ = I600 часов. Определить доверительный интервал для параметра a при доверительной вероятности γ = 0,8.

3.34. Пусть $x_1, \ldots, x_n$ – n независимых измерений экспоненциально распределенной случайной величины ξ с плотностью вероятности $f(t) = \lambda e^{-\lambda t}$, $t \geq 0$. Оценка параметра λ определяется по формуле $\bar{\lambda} = \bar{x}^{-1}$, где $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$. Выразить через $\bar{\lambda}$ границы доверительного интервала для λ^{-1}, удовлетворяющие условию $P(\lambda \nu_1 \geq \bar{\lambda}) = P(\lambda \nu_2 \leq \bar{\lambda}) = 0{,}5(1-\gamma)$, для доверительной вероятности $\gamma = 0{,}9$ при

а) $n = 10$;

б) $n = 20$.

3.35. Показать, что доверительный интервал с надёжностью γ для параметра θ модели $N(\theta, \theta^2)$, $\theta > 0$, по выборке $X_1, \ldots, X_n$ имеет вид:

$$[\bar{X}/(1+c_\gamma/\sqrt{n}),\ \bar{X}/(1-c_\gamma/\sqrt{n})],\ \text{где}\ c_\gamma = \Phi^{-1}(0{,}5(1+\gamma)),$$

$\bar{X}$ – выборочное среднее.

3.36. Пусть $X = (X_1, \ldots, X_n)$ – выборка из распределения $N(a, \theta^2)$, $T^2 = \sum_{i=1}^{n} (X_i - a)^2$. Доказать, что:

а) γ – доверительным интервалом для θ является любой интервал $\delta_\gamma(X) = (T/a_2,\ T/a_1)$, где $a_1 < a_2$ выбираются из условия

$$\int_{a_1}^{a_2} x k_n(x^2)\,dx = \gamma/2,$$

здесь $k_n(x)$ – плотность распределения $\chi^2(n)$; определить наикратчайший в этом классе интервал $\delta_\gamma(X)$;

б) центральный γ – доверительный интервал для θ^2 имеет вид

$$\Delta_\gamma(X) = (T^2/g_2,\ T^2/g_1),\ g_1 = \chi^2_{\frac{1-\gamma}{2}},\ g_2 = \chi^2_{\frac{1+\gamma}{2}}.$$

3.37. По выборке $X = (X_1, \ldots, X_n)$ из распределения $N(\theta_1, \theta_2^2)$ построить односторонние и двусторонние γ – доверительные интервалы а) для θ_1; б) для θ_2^2.

3.38. Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ – две независимые выборки из распределений соответственно $N(\theta_1, \sigma_1^2)$

и $N(\theta_2, \sigma_2^2)$. Построить γ -доверительный интервал для разности средних $\tau = \theta_1 - \theta_2$.

3.39. Пусть $X = (X_1, \ldots, X_n)$ - выборка из экспоненциального распределения с плотностью

$$f(x, \theta) = e^{-(x-\theta)}, \quad x \geq 0, \quad X_{(1)} = \min_i (X_i).$$

Доказать, что $(X_{(1)} + n^{-1} \ln(1-\gamma), X_{(n)})$ есть доверительный интервал для параметра θ.

3.40. По выборке $X = (X_1, \ldots, X_n)$ из распределения $Bi(1, \theta)$ построить асимптотический (при $n \to \infty$) γ - доверительный интервал для θ, основываясь на нормальной аппроксимации

$$L(\sqrt{n}(\bar{X} - \theta)/\sqrt{\theta(1-\theta)}) \sim N(0,1) \quad \text{(теорема Муавра-Лапласа).}$$

§ 4. ПРОВЕРКА ГИПОТЕЗ

I. Статистической гипотезой называют любое предположение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин. Статистическим критерием согласия для основной гипотезы H_0 называется такое правило, которое позволяло бы по результатам соответствующих наблюдений принять или отклонить H_0. Если гипотеза H_0 однозначно фиксирует распределение наблюдений, то ее называют простой, в противном случае - сложной.

Пусть $X = (X_1, \ldots, X_n)$ - выборка из распределения $L(\xi)$, H_0 -некоторая гипотеза о распределении ξ, $T = T(X)$ - статистика, характеризующая отклонение эмпирических данных от соответствующих гипотетических значений. Для каждого малого числа $\alpha > 0$ определим подмножество $T_{1\alpha} = \{t : t = T(x), x \in X\}$, удовлетворяющее условию $P(T \in T_{1\alpha} | H_0) \leq \alpha$.

Подмножество $T_{1\alpha}$ порождает критерий согласия для гипотезы H_0: если $t = T(x)$ - наблюдавшееся значение статистики $T(x)$, то при $t \in T_{1\alpha}$ гипотеза H_0 отвергается; в противном случае считается, что данные не противоречат H_0 (заметим, что факт $t \notin T_{1\alpha}$ не доказывает истинности H_0). Число α называется уровнем значимости критерия, множество $T_{1\alpha}$ - критическим

множеством для гипотезы H_0, статистика T – статистикой критерия, сам критерий – критерием $T_{1\alpha}$.

Чтобы сравнивать различные критерии вводят понятие альтернативной гипотезы и мощности критерия. Любое допустимое распределение $F_x = F$, отличающееся от распределения при гипотезе H_0, называется альтернативным распределением или альтернативой. Совокупность всех альтернатив называется альтернативной гипотезой и обозначается H_1. Функцией мощности критерия $T_{1\alpha}$ называется следующий функционал на множестве допустимых распределений $\{F\}$:

$$W(F) = P(T \in T_{1\alpha} \mid F) = W(T_{1\alpha}; F).$$

Если $F \in H_1$, то $W(F)$ называется мощностью критерия при альтернативе F; оно характеризует вероятность принятия правильного решения, когда H_0 ложна. Из двух критериев с одним и тем же уровнем значимости α лучшим считается тот, мощность которого при альтернативах больше.

Несмещенность критерия $T_{1\alpha}$ означает, что одновременно с условием $W(T_{1\alpha}; F) \leq \alpha$

для всех $F \in H_0$

выполняется условие $W(T_{1\alpha}; F) \geq \alpha$

для всех $F \in H_1$.

Состоятельность критерия означает, что для всех $F \in H_1$

$$\lim_{n \to \infty} W_n(F) = 1.$$

2. Критерий Колмогорова

Он применяется, когда функция распределения $F(x)$ непрерывна. Статистикой критерия является величина

$$D_n = D_n(X) = \sup_{x \in R} |F_n(x) - F(x)|,$$

где $F_n(x)$ – эмпирическая функция распределения. При $n \geq 20$

$$P(\sqrt{n}\, D_n \leq t) \approx K(t) = \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2 t^2).$$

Критическая область критерия определяется неравенством $\sqrt{n}\, D_n \geq t_\alpha$, где $K(t_\alpha) = 1 - \alpha$. Значение находится из табл.4.

3. Критерий хи-квадрат К.Пирсона

Пусть выборка $X = (X_1, \ldots, X_n)$ из распределения $L(\xi)$ сгруппирована в интервальный статистический ряд, ν_i – число X_i, попавших в i-й интервал Δ_i, $\nu_1 + \ldots + \nu_m = n$, H_0 – простая гипотеза, состоящая в том, что $F_\xi(x) = F(x)$. Статистикой критерия χ^2 является величина, предложенная К.Пирсоном:

$$T_n = \chi_q^2 = \sum_{j=1}^{m} \frac{(\nu_j - np_j)^2}{np_j},$$

где $p_i = P(\xi \in \Delta_i)$, $0 < p_i < 1$. Критическая область определяется неравенством $\{T_n \geqslant \chi^2_{1-\alpha, m-1}\}$, где $\chi^2_{p,r}$ – p – квантиль распределения $\chi^2(r)$.

Критерий χ^2 позволяет также проверить гипотезу о независимости двух случайных величин ξ и η. В этом случае

$$\chi_q^2 = \sum_{i=1}^{\ell} \sum_{j=1}^{m} (\nu_{ij} - m_{ij})^2 / m_{ij},$$

где ν_{ij} – число случаев, когда одновременно наблюдались значения $\xi = x_i$, $\eta = y_j$, $m_{ij} = h_{io} h_{oj} / n$; h_{io} – число случаев, когда наблюдалось значение $\xi = x_i$; h_{oj} – число случаев, когда наблюдалось значение $\eta = y_j$; ℓ и m – числа значений, которые принимали величины ξ и η; n – объем выборки.

Число степеней свободы k, необходимое для вычисления $P(\chi^2 \geqslant \chi_q^2)$, определяется формулой $k = (\ell - 1)(m - 1)$.

4. Критерий Смирнова проверки однородности статистического материала.

Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ – две независимые выборки из распределений $L(\xi_1)$ и $L(\xi_2)$ соответственно с непрерывными функциями распределения $F_{\xi_1}(x) = F_1(x)$, $F_{\xi_2}(x) = F_2(x)$. Гипотеза $H_0: F_1(x) = F_2(x)$. Критерий Смирнова основан на статистике

$$Д_{nm} = \sup_{x \in R} |F_{1n}(x) - F_{2m}(x)|,$$ где $F_{1n}(x)$ и $F_{2m}(x)$ –

– выборочные функции распределения. При больших n и m

$$P(\sqrt{nm/(n+m)}\, D_{nm} \leq t\ ) \approx K(t).$$

Критическая область определяется неравенством

$$\sqrt{mn/(n+m)}\ D_{nm} > t_\alpha \quad , \text{ где } K(t_\alpha) = 1-\alpha .$$

5. Критерий однородности χ^2

Пусть осуществлено K серий независимых наблюдений, объемы которых $n_1, \ldots, n_k$, и в каждой серии наблюдался некий переменный признак, принимающий одно из S возможных исходов. Пусть ν_{ij} – число реализаций i-го исхода в j-й серии $\left(\sum_{i=1}^{S} \nu_{ij} = n_j ,\right.$ $\left. j = 1, \ldots, k\right)$. Гипотеза H_0: все наблюдения проводились над одной и той же случайной величиной. Статистикой критерия χ^2 является

$$\chi_n^2 = n\left(\sum_{i=1}^{S} \sum_{j=1}^{k} \frac{\nu_{ij}^2}{n_i\, \nu_i} - 1\right),$$

$$\nu_i = \sum_{j=1}^{k} \nu_{ij} , \quad i = 1, \ldots S, \quad n = n_1 + \ldots + n_k .$$

Критическая область задаётся неравенством $\chi_n^2 \geqslant \chi^2_{1-\alpha,(S-1)(k-1)}$. Границу критерия определяют из таблиц квантилей распределения χ^2.

6. Проверка гипотез о средних нормальных распределениях

Пусть $X = (X_1, \ldots, X_n)$ – выборка из $L(\xi) = N(a, \sigma^2)$, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, $S^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Для проверки гипотезы $H_0 : M\xi = a$ используют статистику $t = (\bar{X} - a)/S(n-1)^{-1/2}$. Величина t имеет распределение Стьюдента с $n-1$ степенями свободы, и ее распределение не зависит от a и $D\xi$.

Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ - две независимые выборки из распределений $N(\theta_1, \sigma^2)$ и $N(\theta_2, \sigma^2)$. Нужно проверить гипотезу H_0: $\theta_1 = \theta_2$. Статистика

$$t = \sqrt{(n+m-2)(n^{-1}+m^{-1})}\,(\bar{X}-\bar{Y})/\sqrt{nS_X^2+mS_Y^2}$$

имеет t - распределение с $m+n-2$ степенями свободы.

Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ - две независимые выборки из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$,

$$\tilde{S}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2, \quad \tilde{S}_Y^2 = \frac{1}{m-1}\sum_{j=1}^{n}(Y_j-\bar{Y})^2,$$

$$S_Б^2 = max(\tilde{S}_X^2, \tilde{S}_Y^2), \quad S_М^2 = min\{\tilde{S}_X^2\ \tilde{S}_Y^2\}.$$

Нужно проверить гипотезу $H_0: \sigma_1^2 = \sigma_2^2$. Найдём $F_Н = S_Б^2/S_М^2$ (см. теорему I.3) и $F_{kp}(\alpha/2, k_1, k_2)$ по табл. IO критических точек распределения Фишера-Снедекора, где α - уровень значимости, $k = max(n_1, m_1)$, $n_1 = n - I$, $m_1 = m - I$. Если $F_Н < F_{kp}$, то H_0 принимается, в противном случае H_0 отвергается.

7. Параметрические гипотезы

Гипотеза H_0 для выборки $X = (X_1, \ldots, X_n)$ из параметрической модели $\{F(x,\theta),\ \theta=(\theta_1,\ldots,\theta_r)\in\Theta\}$ называется параметрической, если она задаётся в терминах неизвестного параметра θ в виде $H_0: \theta\in\Theta$, при некотором подмножестве $\Theta_0 \subset \Theta$. Альтернативная гипотеза имеет вид $H_1: \theta\in\Theta_1 = \Theta\setminus\Theta_0$. Критерий проверки гипотезы H_0 при уровне значимости α задается выбором подмножества $X_{1\alpha} \subset X$, для которого при любом $\theta\in\Theta$ выполняется условие $P(X\in X_{1\alpha}) \le \alpha$. Критерий формулируется так: если x - наблюдавшаяся реализация выборки X, то при $x\in X_{1\alpha}$ гипотезу H_0 отвергают, если же $x\in\bar{X}_{1\alpha} = X_{0\alpha}$, то гипотезу H_0 принимают. Функцию мощности обозначают $W(\theta) = W(X_{1\alpha}, \theta) = P(X\in X_{1\alpha}), \theta\in\Theta$. Вероятность ошибки первого рода: $W(\theta) = P(H_1/H_0), \theta\in\Theta$; вероятность ошибки второго

рода: $1-W(\theta)=P(H_0/H_1)$, $\theta\in\Theta_1$.

Пусть $X_{1\alpha}$ и $X^*_{1\alpha}$ – два критерия одного и того же уровня значимости α для гипотезы H_0. Если

$$W(X^*_{1\alpha},\theta)\leq W(X_{1\alpha},\theta) \text{ при } \theta\in\Theta_0 \text{ и } W(X^*_{1\alpha},\theta)\geq W(X_{1\alpha},\theta)$$

при $\theta\in\Theta_1$ со строгим неравенством, по крайней мере, при одном $\theta\in\Theta$, то говорят, что критерий $X^*_{1\alpha}$ равномерно мощнее критерия $X_{1\alpha}$. Если критерий $X^*_{1\alpha}$ равномерно мощнее любого критерия $X_{1\alpha}$, то $X^*_{1\alpha}$ называется равномерно наиболее мощным.

8. Критерий Неймана-Пирсона.

Если $\Theta=\{\theta_0,\theta_1\}$, то при любом уровне значимости α наиболее мощный критерий проверки гипотезы $H_0:\theta=\theta_0$ при альтернативе $H_1:\theta=\theta_1$ существует и задаётся критической областью

$$X^*_{1\alpha}=\{x:\ell(x)=L(x,\theta_1)/L(x,\theta_0)\geq C_\alpha\},$$

где $L(x_1,\theta)$ – функция правдоподобия.

9. Критерий отношения правдоподобия

Общий вид критерия отношения правдоподобия для проверки гипотезы $H_0:\theta\in\Theta$ таков:

$$X_{1\alpha}=\{x:\lambda_n(x)=\sup_{\theta\in\Theta_0}L(x,\theta)/\sup_{\theta\in\Theta}L(x,\theta)\leq C_\alpha\},$$

где C_α выбирается из условия: для любого $\theta\in\Theta_0$

$$W(\theta)=P(\lambda_n(X)\leq C_\alpha)\leq\alpha.$$

Если выполняются условия регулярности, обеспечивающие единственность и асимптотическую нормальность оценки максимального правдоподобия $\hat\theta_n$, то при простой гипотезе $H_0:\theta=\theta_0$ для больших объемов выборки в критерий отношения правдоподобия задается критической областью

$$X_{1\alpha}=\{x:-2\ln\lambda_n(x)\geq\chi^2_{1-\alpha,r}\},$$

где $\chi^2_{p,r}$ – p – квантиль $\chi^2(r)$.

Примеры

I. При сверлении 80 отверстий одним и тем же сверлом и последующим измерением диаметров отверстий получены следующие данные:

(x_i, x_{i+1}), мм	40,24 – – 40,26	40,28 – – 40,32	40,32 – – 40,36	40,36 – – 40,40	40,40 – – 40,44
n_i	4	I9	32	I5	I0

Проверить, используя критерии χ^2 и Колмогорова, гипотезу H_0 о согласии полученных данных с нормальным законом. Уровень значимости α принять равным 0,05.

Решение. Находим $\bar{x} = 40{,}344$, $S = 0{,}42$. Используя критерий χ^2, проверим гипотезу H_0 о согласии полученных данных с законом $N(\bar{x}, S^2)$. Введём величины $u_i = (x_i - \bar{x})/S$. Теоретические вероятности p_i вычисляем по формуле

$$p_i = P(x_i < \xi \leq x_{i+1}) = \Phi(u_{i+1}) - \Phi(u_i).$$

В частности, $p_1 = P(40{,}24 < \xi \leq 40{,}28) = \Phi(2{,}38) - \Phi(I{,}42) = 0{,}064$. Результаты вычислений приведём в таблице.

Интервалы наблюдаемых значений (x_i, x_{i+1})	n_i	Интервалы (u_i, u_{i+1})	p_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
(40,24–40,28)	4	(–2,48; –I,52)	0,064	5,I	I,2I	0,24
(40,28–40,32)	I9	(–I,52; –0,57)	0,220	I7,6	I,96	0,II
(40,32–40,36)	32	(–0,57; 0,38)	0,364	29,I	8,4I	0,29
(40,36–40,40)	I5	(0,38; I,33)	0,260	20,8	I7,64	0,85
(40,40–40,44)	I0	(I,33; 2,28)	0,092	7,4	6,76	0,9I

В результате вычислений получим $\chi^2_H = 2{,}40$. По таблице квантилей по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы

$k - I = 5 - I = 4$ найдём $\chi^2_{0{,}95;4} = 9{,}49$. Так как $2{,}40 < 9{,}49$, то нет основания для отклонения гипотезы.

Проверим гипотезу H_0 с помощью критерия Колмогорова. Все вспомогательные расчеты, необходимые для вычисления

$\lambda_{набл.} = D\sqrt{n} = \sqrt{n}\max\limits_{x}|F_n(x) - F(x)|$, сведём в таблицу.

Интервалы (x_i, x_{i+1})	n_i	$n_{нк}$	p_i	$F_n(x_i) = \frac{n_{нк}}{n}$	$F(x_i) = P_{пк}$	$\lvert F_n(x_i) - F(x_i)\rvert$
(40,24;40,28)	4	4	0,064	0,050	0,064	0,014
(40,28;40,32)	19	23	0,226	0,288	0,284	0,004
(40,32;40,36)	32	55	0,364	0,688	0,848	0,040
(40,36;40,40)	15	70	0,260	0,875	0,908	0,033
(40,40;40,44)	10	80	0,092	1,000	1,000	0,000

Здесь $n_{нк}$ - накопленные эмпирические частоты, $P_{пк}$ - накопленные теоретические вероятности. $D = \max\limits_{x}|F_n(x) - F(x)| = 0,040$. $\lambda_{набл.} = D\sqrt{n} = 0,04\sqrt{80} = 0,36$. По табл. 4 квантилей распределения Колмогорова по уровню значимости $\alpha = 0,05$ находим $\lambda_{кр} = 1,358$. Так как $0,36 < 1,358$, то нет основания отклонять гипотезу H_0.

2. Радиоактивное вещество наблюдалось в течение 2608 равных интервалов времени (по 7,5 с). Для каждого из этих интервалов регистрировалось число частиц, попавших в счетчик. В таблице приведены числа m_i интервалов времени, в течение которых в счетчик попало ровно i - частиц.

i	0	1	2	3	4	5	6	7	8	9	10
m_i	57	203	383	525	532	408	273	139	45	27	16

Проверить, используя критерий χ^2, гипотезу H_0 о согласии наблюдаемых данных с законом распределения Пуассона. Уровень значимости α принять равным 0,05.

Р е ш е н и е . Находим $\bar{a} = n^{-1}\sum\limits_{i=0}^{10} i m_i = 3,870$. Результаты вычислений теоретических вероятностей p_i и $\chi^2_q = \sum\limits_{i=0}^{10}(m_i - np_i)^2/np_i$ поместим в таблицу:

i	p_i	np_i	$(m_i-np_i)^2$	$(m_i-np_i)^2/np_i$
0	0,021	54,8	4,84	0,088
1	0,081	211,2	67,24	0,318
2	0,156	406,8	566,44	1,392
3	0,201	524,2	0,64	0,001
4	0,195	508,6	547,56	1,007
5	0,151	393,2	201,64	0,512
6	0,097	253,0	400,00	1,581
7	0,054	140,8	3,24	0,023
8	0,026	67,8	519,84	7,667
9	0,011	28,7	2,89	0,101
10	0,007	18,3	5,29	0,289
				$\chi_q^2 = 13,049$

Мы получили $\chi_q^2 = \sum_{i=0}^{10} (m_i - np_i)^2 / np_i = 13,05$. Из табл. 7 по степеням свободы $k = 11 - 1 - 1 = 9$ и $\chi_q^2 = 13,05$ находим $\alpha_k = P(\chi^2 \geqslant \chi_q^2) = 0,166$. Так как $\alpha_k > \alpha = 0,05$, то отклонение от закона Пуассона не значимо, т.е. гипотеза H_0 принимается.

3. По двум независимым выборкам, объемы которых $n_1 = 11$ и $n_2 = 14$, из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$ найдены несмещенные оценки дисперсии $\tilde{\sigma}_1^2 = 0,76$ и $\tilde{\sigma}_2^2 = 0,38$. При уровне значимости $\alpha = 0,05$ проверить гипотезу $H_0: \sigma_1^2 = \sigma_2^2$.

Р е ш е н и е. Имеем $F_H = 0,76/0,38 = 2$. По табл. 7 по $\alpha = 0,05$, $k_1 = 10$ и $k_2 = 13$ находим $F_{кр}(0,05; 10; 13) = 2,67$. Так как $F_H < F_{кр}$, то нет оснований отвергать H_0.

З а д а ч и

4.1. В процессе производства электрические счетчики с вращающимся диском были отрегулированы для того, чтобы синхронизировать их работу со стандартным счетчиком. Проверка 10 счетчиков, заключающаяся в определении постоянной счетчика, показала следующие результаты: 0,983; 1,002; 0,998; 0,996; 1,002; 0,983; 0,994; 0,991;

I,005; 0,986, причем постоянная, характеризующая стандартный счетчик, взята равной I,000. Можно ли отклонения от стандарта рассматривать как случайные? Ответить на вопрос, проверив гипотезу о том, что результаты измерений соответствуют нормальному закону со средним a = I,000.

4.2. Результаты измерения I000 деталей представлены в виде группированной выборки распределения:

x_i, мм	98,0	98,5	99,0	99,5	I00,0
n_i	2I	47	87	I58	I8I

x_i, мм	I00,5	I0I,0	I0I,5	I02,0	I02,5
n_i	20I	I42	97	4I	25

Пользуясь критерием Колмогорова, проверить гипотезу H_0 о согласии полученных наблюдений с предположением, что величина ξ подчинена нормальному закону с математическим ожиданием $\bar{x}$ = I00,25 мм и среднеквадратическим отклонением σ = I мм, при уровне значимости α = 0,05.

4.3. Ниже приводятся результаты опыта (данные Уэлдона) с подбрасыванием костей: выпадение 6 очков на одной грани при 4096 подбрасываниях I2 костей.

i	0	I	2	3	4	5	6	7
n_i	447	II45	II8I	796	380	II5	24	8

Здесь i - число выпадений 6 очков. Проверить с помощью критерия χ^2 гипотезу H_0 правильности костей.

В задачах 4.4 - 4.8 приведены результаты наблюдений некоторой случайной физической величины ξ, сгруппированные в интервальный статистический ряд частот. С помощью критериев χ^2 и Колмогорова проверить справедливость гипотезы H_0 о согласии полученных данных с нормальным законом. Уровень значимости α принять равным 0,05.

4.4. Даны результаты испытания стойкости (x_i) I00 фрез (в часах):

x_i	32 - 36	36 - 40	40 - 44	44 - 48	48 - 52
n_i	7	22	44	2I	6

4.5. Даны результаты измерения толщины (x_i) 50 слюдяных прокладок:

x_i, см	0,24–0,28	0,28–0,32	0,32–0,36	0,36–0,40	0,40–0,44
n_i	5	8	22	9	6

4.6. Даны результаты определения содержания фосфора в 100 чугунных образцах (в %):

x_i	0,10–0,20	0,20–0,30	0,30–0,40	0,40–0,50	0,50–0,60
n_i	6	23	38	25	8

4.7. В таблице приведены статистические данные о трудоёмкости операции (x_i) "ремонт валика водяного насоса автомобиля ЗИЛ-130" (в минутах):

x_i	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	16	48	70	47	19

4.8. Даны результаты испытания стойкости (x_i) 100 удлиненных сверл диаметром 4 мм:

x_i, ч	3,0–3,2	3,2–3,4	3,4–3,6	3,6–3,8	3,8–4,0
n_i	17	49	70	46	18

4.9. В экспериментах с селекцией гороха Мендель наблюдал частоты различных видов семян, подчиненных при скрещивании растений с круглыми желтыми (КЖ) семенами растений и растений с морщинистыми зелеными (МЗ) семенами. Эти данные и значения теоретических вероятностей по теории наследственности приведены в следующей таблице:

Семена	Частота	Вероятность
К и Ж	315	9/16
М и З	101	3/16
К и З	108	3/16
М и З	32	1/16
Σ	n = 556	1

Используя критерий χ^2, проверить гипотезу H_0 о согласовании частотных данных с теоретическими вероятностями (на уровне значимости α = 0,1).

4.10. Через равные промежутки времени в тонком слое раствора золота регистрировалось число частиц золота, попадавших в поле зрения микроскопа. Результаты наблюдений приведены в таблице:

i	0	1	2	3	4	5	6	7
n_i	112	168	130	68	32	5	1	1

Используя критерий χ^2, проверить гипотезу H_0 о согласии опытных данных с законом распределения Пуассона, приняв за уровень значимости $\alpha = 0{,}05$.

4.11. При $n = 4000$ независимых испытаний события A_1, A_2, A_3, составляющие полную группу, осуществились соответственно 1905, 1015 и 1080 раз. Проверить, согласуются ли эти данные при уровне значимости 0,05 с гипотезой $H_0: P_1 = 1/2,\ P_2 = P_3 = 1{,}4$, где $P_i = P(A_i)$.

4.12. По каждой из 100 мишеней произведено из спортивного пистолета по 10 независимых выстрелов. Фиксировалось число попаданий в мишень. Результаты приведены в таблице:

i	0	1	2	3	4	5	6	7	8	9	10
n_i	0	2	4	10	22	26	18	12	4	2	0

Используя критерий χ^2, проверить гипотезу H_0 о согласии результатов стрельбы с биномиальным законом распределения. Уровень значимости принять равным 0,10.

4.13. Среди 2020 семей, имеющих двух детей, 527 семей, в которых два мальчика, в 476 – две девочки, в остальных – дети разного пола. Можно ли с уровнем значимости 0,05 считать, что количество мальчиков в семье с двумя детьми – биномиальная случайная величина?

4.14. Цифры 0,1,2,...,9 среди 800 первых десятичных знаков появлялись 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно. Проверить с помощью критерия χ^2 гипотезу о согласии этих данных с законом равномерного распределения при уровне значимости $\alpha = 0{,}10$.

4.15. Отсчет по шкале измерительного прибора оценивается приблизительно в долях деления шкалы. В таблице приводится 200 результатов отсчета последней цифры между делениями шкалы. Установить, используя критерий χ^2, согласуются ли наблюдения с законом равномерного распределения, при котором вероятность появления любой цифры i, $P_i = 0{,}1$, при уровне значимости $\alpha = 0{,}05$.

i	0	I	2	3	4	5	6	7	8	9
n_i	35	I6	I5	I7	I7	I9	II	I6	30	24

4.I6. В десятичной записи числа π среди первых I002 знаков после запятой цифры 0,I,...,9 встречаются соответственно 968, I026, I02I, 974, I0I4, I046, I02I, 970, 948, I0I4 раз. Можно ли на уровне значимости 0,05 считать эти цифры случайными? При каком уровне значимости эта гипотеза отвергается?

4.I7. Испытания 200 ламп на продолжительность работы T (в часах) приведены в таблице:

Границы интервала (t_i, t_{i+1})	0, 300	300, 600	600, 900	900, I200	I200, I500	I500, I800
n_i	53	4I	30	22	I6	I2

Границы интервала (t_i, t_{i+1})	I800, 2I00	2I00, 2400	2400, 2700	2700, 3000	3000
n_i	9	7	5	3	2

Проверить, используя критерий χ^2, согласие опытных данных с экспоненциальным законом распределения при уровне значимости $\alpha = 0,05$.

4.I8. С помощью контрольного прибора было измерено расстояние x (в микронах) центра тяжести детали от оси ее наружной цилиндрической поверхности для 602 деталей. Результаты измерений представлены в таблице:

Границы интервала (x_i, x_{i+1})	0, I6	I6, 32	32, 48	48, 64	64, 80
n_i	40	I29	I40	I26	9I

Границы интервала (x_i, x_{i+1})	80, 96	96, II2	II2, I28	I28, I44
n_i	45	I9	8	4

Проверить, используя критерий χ^2, согласуются ли наблюдаемые

данные с законом распределения Рэлея:

$$f(x) = a^{-2} x e^{-x^2/2a^2}.$$

4.19. В таблице приведены числа n_i участков равной площади (0,25 км^2) южной части Лондона, на каждый из которых приходилось по i попаданий самолётов-снарядов во время второй мировой войны. Проверить с помощью критерия χ^2 гипотезу H_0 о согласии опытных данных с законом распределения Пуассона, приняв за уровень значимости α = 0,05.

i	0	I	2	3	4	5
n_i	229	2II	93	35	7	I

4.20. При переписи населения Англии и Уэльса в I90I г. было зарегистрировано (с точностью до тысячи) I5 729 000 мужчин и I6 799 000 женщин; 3497 мужчин и 3072 женщины были зарегистрированы как глухонемые от рождения. Проверить гипотезу H_0 о том, что глухота не связана с полом.

4.21. По двум независимым выборкам, объемы которых n_1 = I4 и n_2 = I0 из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$, найдены несмещенные оценки дисперсий: $\tilde{\sigma}_1^2 = n_1 S_x^2/(n_1-1)$ = 0,84 и $\tilde{\sigma}_2^2 = n_2 S_y^2/(n_2-1)$ = 2,52. При уровне значимости α = 0,I проверить гипотезу $H_0: \sigma_1^2 = \sigma_2^2$.

4.22. По двум независимым выборкам, объемы которых n_1 = 9 и n_2 = 6 из распределений $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$, найдены выборочные дисперсии: S_x^2 = I4,4 и S_y^2 = 20,5. При уровне значимости 0,I проверить гипотезу $H_0: \sigma_1^2 = \sigma_2^2$.

4.23. Крупная партия товаров может содержать долю дефектных изделий. Поставщик полагает, что эта доля составляет 3 %, а покупатель – I0 %. Условия поставки: если при проверке 20 случайным образом отобранных товаров будет обнаружено не более одного дефектного, то партия принимается на условиях поставщика, в противном случае – на условиях покупателя. Требуется определить:

I) каковы статистические гипотезы, статистика критерия, область ее значений, критическая область;

2) какое распределение имеет статистика критерия, в чем состоят ошибки первого и второго рода и каковы их вероятности.

4.24. Двумя методами проведены независимые измерения одной и той же физической величины. Получены следующие результаты:

а) в первом случае: 9,6; 10,0; 9,8; 10,2; 10,6;

б) во втором случае: 10,4; 9,7; 10,0; 10,3.

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, т.е. $\sigma_1^2 = \sigma_2^2$, если уровень значимости α = 0,1 ? Предполагается, что результаты измерений распределены нормально.

4.25. Понадобилось сравнить два метода определения содержания крахмала в картофеле. Были взяты 16 клубней картофеля и к каждому клубню были применены два метода измерения. Результаты измерения разности между процентным содержанием крахмала в каждом клубне, определенного I и II методами, следующие: 0,2; 0,0; 0,0; 0,1; 0,2; 0,2; 0,3; -0,3; 0,1; 0,2; 0,3; 0,0; -0,1; 0,1; -0,2; 0,1. Проверить гипотезу H_0 о том, что оба метода не дают статистического различия в определении процентного содержания крахмала.

4.26. Поступающие в институт абитуриенты разбиты на два потока по 300 человек в каждом. Итоги экзамена по одному и тому же предмету на каждом потоке оказались следующими: на I-м потоке баллы 2, 3, 4 и 5 получили соответственно 33, 43, 80 и 144 человека; соответствующие же данные для 2-го потока таковы: 39, 35, 72 и 15. Можно ли при уровне значимости 0,05 считать оба потока однородными.

4.27. Проверить гипотезу независимости для следующей таблицы сопряженности двух признаков (уровень значимости принять равным 0,05).

ξ_1 \ ξ_2	b_1	b_2	b_3	Σ
a_1	3009	2832	3008	8849
a_2	3047	3051	2997	9095
a_3	2974	3038	3018	9030
Σ	9030	8921	9023	26974

4.28. Из 300 абитуриентов, поступивших в институт, 97 человек имели балл 5 в школе и 48 получили 5 на вступительных экзаменах по тому же предмету, причем только 18 человек имели 5 в школе и на экзамене. С уровнем значимости 0,1 проверить гипотезу о независимости оценок 5 в школе и на экзамене.

4.29. При 8000 независимых испытаний A, B и C, составляющих полную группу, осуществились 2014, 5012 и 974 раза соответственно. Верна ли на уровне значимости 0,05 гипотеза: $P(A) = 0{,}5 - 2\theta$, $P(B) = 0{,}5 + \theta$, $P(C) = \theta$ $(0 < \theta < 0{,}25)$?

4.30. Чтобы проверить гипотезу $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta = \theta_1$ $(0 < \theta_0 < \theta < 1)$ в схеме Бернулли с неизвестной вероятностью "успеха", осуществлен эксперимент, в котором наблюдалось число "успехов", предшествующих первому "неуспеху". Построить наиболее мощный критерий при уровне значимости $\alpha = \theta_0^s$, где $s \geqslant 1$ – заданное число, и убедиться, что вероятность ошибки второго рода этого критерия $\beta = 1 - \theta_1^s$.

4.31. Пусть для распределения Коши $K(\theta)$ проверяется гипотеза $H_0: \theta = 0$ против альтернативы $H_1: \theta = 1$. Показать, что при уровне значимости $\alpha = 0{,}5 - \pi^{-1} arctg\, 0{,}5 \approx 0{,}352$ наиболее мощный критерий по одному наблюдению имеет вид $X_{1\alpha}^* = \{X \geqslant 0{,}5\}$ и его мощность равна $0{,}5 + \pi^{-1} arctg\, 0{,}5 \approx 0{,}648$. Если же $\alpha = \pi^{-1}(arctg\, 3 - arctg\, 1) \approx 0{,}148$, то критерий имеет вид $X_{1\alpha}^* = \{1 \leqslant X \leqslant 3\}$, а мощность равна $\pi^{-1} arctg\, 2 \approx 0{,}352$.

4.32. (выборочный контроль). Пусть партия из N изделий содержит неизвестное число θ дефектных изделий, $\theta \in \{0, 1, \ldots, N\}$. Чтобы проверить гипотезу $H_0: \theta \leqslant \theta_0$ против альтернативы $H_1: \theta > \theta_0$ берут на контроль n изделий и каждое из них проверяют. Основываясь на статистике T – обнаруженное в выборке число дефектных изделий, построить равномерно наиболее мощный критерий.

§ 5. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННОГО И РЕГРЕССИОННОГО АНАЛИЗА

1. Метод наименьших квадратов. Линейная регрессионная модель

Предположим, что между величинами y и x, $x = (x^{(1)}, \ldots, x^{(m)})$, существует функциональная зависимость $y = \varphi(x, \theta)$, где φ известна с точностью до неизвестных параметров $\theta = (\theta_1, \ldots, \theta_m)$. Например, $y = \theta_1 x^{(1)} + \ldots + \theta_m x^{(m)}$ (линейная регрессия). Проводятся N независимых измерений, при i-м измерении для известного значения x_i вычисляется с ошибкой значение y_i. Модель таких измерений можно определить равенствами $y_i = \varphi(x_i, \theta) + \varepsilon_i$, $i = 1, 2, \ldots, N$, где $\varepsilon_1, \ldots, \varepsilon_N$ – независимые случай-

ные величины $M\varepsilon_i = 0$, $D\varepsilon_i = \sigma^2$. Метод наименьших квадратов состоит в том, что по значениям (x_i, y_i), $i = 1, \ldots, N$ ищется $\hat{\theta}$, минимизирующий сумму квадратов отклонений измеренных y_i от соответствующих вычисленных значений $\varphi(x_i, \theta)$:

$$Q(\theta) = \sum_{i=1}^{N} [y_i - \varphi(x_i, \theta)]^2.$$

Если функция $\varphi(x, \theta)$ линейна, то модель называется линейной регрессионной моделью. В векторной форме она записывается так: $y = A\theta + \varepsilon$. Здесь случайный вектор $y^T = (y_1, \ldots, y_N)$ – вектор результатов измерений, $\varepsilon^T = (\varepsilon_1, \ldots, \varepsilon_N)$ – вектор ошибок измерений, $A = [x_i^{ij}]_{i=1, j=1}^{N, m}$ – матрица плана эксперимента, $\theta^T = (\theta_1, \ldots, \theta_m)$ – неизвестный векторный параметр. Оценка параметра $\hat{\theta}$ называется оценкой метода наименьших квадратов (МНК – оценкой), если

$$\min_{\theta \in R^m} (y - A\theta)^T (y - A\theta) = \min_{\theta \in R^m} \varepsilon^T \varepsilon = (y - A\hat{\theta})^T (y - A\hat{\theta}).$$

Если $(A^T A)^{-1}$ существует, то $\hat{\theta} = (A^T A)^{-1} A^T y$ является несмещенной МНК-оценкой параметра θ. Для нормальной регрессии, когда $\hat{\theta}$ имеет распределение $N(0, \sigma^2 J)$, МНК-оценка $\hat{\theta}$ совпадает с оценками максимального правдоподобия параметров θ. В этом случае γ –доверительный интервал для θ_j имеет вид

$$(\hat{\theta}_j - \Lambda(\hat{\theta}), \hat{\theta}_j + \Lambda(\hat{\theta})),$$

где

$$\Lambda(\hat{\theta}) = t_{\frac{1+\gamma}{2}, N-m} \sqrt{\frac{a^{jj}}{N-m} S(\hat{\theta})}, \quad S(\hat{\theta}) = \|y - A\hat{\theta}\|^2,$$

$\|z\|^2$ – квадрат нормы вектора z, a^{jj} – j-й диагональный элемент матрицы $(A^T A)^{-1}$. Для остаточной дисперсии

$$S(\hat{\theta}) / \chi^2_{\frac{1+\gamma}{2}, N-m} < \sigma^2 < S(\hat{\theta}) / \chi^2_{\frac{1-\gamma}{2}, N-m}.$$

2. Линия регрессии

Пусть $\xi_1, \ldots, \xi_n$ – независимые случайные величины с $M\xi_i = 0$, $0 < D\xi_i = \sigma_i^2$, $i = 1, \ldots, n$, а η – случайная величина с нулевым средним и конечной дисперсией $D\eta = \sigma^2$.

Линейная оценка $\hat{\eta} = \sum_{i=1}^{n} \rho_i \sigma \xi_i / \sigma_i$, где $\rho_i = M(\xi_i \eta)/\sigma_i \sigma$ – коэффициент корреляции величин ξ_i и η, является наилучшей линейной оценкой, минимизирующей среднеквадратическую ошибку $M(\eta - \hat{\eta})^2$ (МНК-оценка). В частности, для двух случайных величин ξ и η МНК-оценкой $\hat{\eta}$ величины η по величине ξ будет

$$\hat{\eta} = M\eta + \rho \frac{\sigma_\xi}{\sigma_\eta} \xi, \text{ где } D\xi = \sigma_\xi^2,\ D\eta = \sigma_\eta^2,\ \rho = M(\xi\eta)/\sigma_\xi \sigma_\eta.$$

Пусть $(x_1, y_1), \ldots, (x_n, y_n)$ – выборка из распределения $L(\xi, \eta)$ двумерного случайного вектора (ξ, η),

$\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$, $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$ – выборочные средние координат,

$$s_\xi^2 = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad s_\eta^2 = n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

$\rho_{\xi\eta}/s_\xi s_\eta$ – выборочный коэффициент корреляции, здесь

$$\mu_{\xi\eta} = n^{-1} \sum x_i y_i - \bar{x}\bar{y}.$$

Выборочное уравнение прямой линии регрессии η на ξ имеет вид

$$\bar{y}_x - \bar{y} = \rho \frac{s_\xi}{s_\eta} (x - \bar{x}),$$

где $\bar{y}_x$ – условная средняя; $\bar{y}_x = M(\eta | \xi = x)$.

Для определения значимости коэффициента корреляции находят статистику $t_н = \rho\sqrt{n-2}/\sqrt{1-\rho^2}$ и квантиль $t_{0,5\alpha; n-2}$ распределения Стьюдента (α – уровень значимости). Если $t_н > t_{0,5\alpha; n-2}$, то коэффициент корреляции ρ значимо отличается от нуля.

Примеры

1. Пусть $y = A\theta + \varepsilon$ – линейная регрессионная модель. Здесь $\theta^T = (\theta_1, \dots, \theta_m)$ – неизвестный вектор, $\bar{\varepsilon}^T = (\varepsilon_1, \dots \varepsilon_N)$ – случайный вектор с независимыми координатами, $M\varepsilon_i = 0$ и $D\varepsilon_i = \sigma^2$, A – матрица размерности $N \times m$ (матрица плана). Показать, что если матрица $(A^TA)^{-1}$ существует, то оценка $\hat{\theta} = (A^TA)^{-1}A^Ty$ является несмещенной МНК-оценкой.

Решение. Покажем несмещенность этой оценки. Имеем $M\hat{\theta} = (A^TA)^{-1}A^TMy = (A^TA)^{-1}A^TA\theta = \theta$. Положим $u = y - A\hat{\theta}$, $v = A(\hat{\theta} - \theta)$. Будем обозначать: $(x, y) = x^Ty = \sum_{i=1}^{n} x_i y_i$ – скалярное произведение векторов $x^T = (x_1, \dots x_n)$ и $y^T = (y_1, \dots y_n)$ и $\|x^2\| = (x, x) = x^Tx$ – квадрат нормы вектора x. В этих обозначениях для любого вектора θ имеем

$$\|y - A\theta\|^2 = \|y - A\hat{\theta} + A(\hat{\theta} - \theta)\|^2 = \|u - v\|^2 =$$

$$= \|u\|^2 + \|v\|^2 + 2(u, v) = \|u\|^2 + \|v\|^2 \geqslant \|u\|^2 = \|y - A\hat{\theta}\|^2,$$

так как

$$(u, v) = (u, A(\hat{\theta} - \theta)) = u^TA(\hat{\theta} - \theta) = (A^Tu)^T(\hat{\theta} - \theta) =$$
$$= (A^Ty - A^TA\hat{\theta})^T(\hat{\theta} - \theta) = 0.$$

Таким образом, для любого θ $\|y - A\hat{\theta}\|^2 \leq \|y - A\theta\|^2$,

или $\|y - A\hat{\theta}\|^2 = \min_{\theta} \|y - A\theta\|^2$.

2. Результаты измерений предела прочности (x, кг/мм^2) и предела текучести (y, кг/мм^2) у 50 марок стали приведены в таблице:

y	77	96	86	92	98	53	63	80	64	66
x	81	77	76	86	53	47	36	40	49	60
y	81	57	86	80	87	163	153	133	159	134
x	54	40	61	68	88	145	136	129	126	96
y	129	145	142	120	95	107	133	140	149	147
x	100	95	106	118	109	107	120	114	113	123
y	104	108	93	127	112	113	95	112	116	93
x	94	84	73	107	94	107	99	100	104	88
y	96	112	136	104	103	115	123	111	127	129
x	84	94	152	98	77	88	94	76	84	73

По приведенным экспериментальным данным требуется:

а) составить корреляционную таблицу;

б) по корреляционной таблице найти числовые характеристики $\bar{x}, \bar{y}, S_x, S_y, \mu_{xy}, \rho$;

в) проверить значимость коэффициента корреляции ρ (уровень значимости $\alpha = 0{,}05$);

г) найти эмпирические функции регрессии y на x и x на y и построить графики.

Решение.

а) Чтобы составить корреляционную таблицу, примем для величины x следующие границы интервалов: (30-50), (50-70), ..., (130-150), а для величины y – (50-70), (70-90), ..., (150-170). Длины интервалов $h_x = h_y = 20$ кг/мм^2. Подсчитаем количество точек (x, y), попадающих в соответствующие прямоугольники. Получим корреляционную таблицу:

y \ x	40	60	80	100	120	140	n_y
160					1	3	4
140				3	5		8
120			4	8	1		13
100		2	7	4			13
80	1	3	3				7
60	4	1					5
n_x	5	6	14	15	7	3	$n = 50$

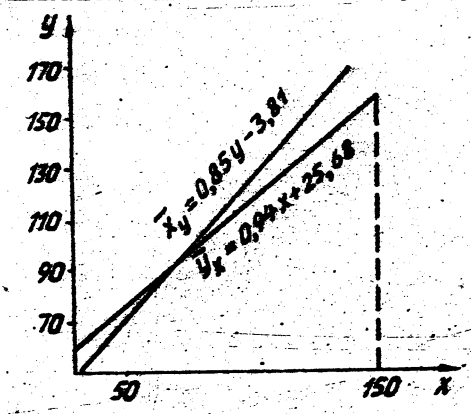
б) Найдём числовые характеристики $\bar{x} = 88{,}80$ кг/мм^2; $\bar{y} = 109{,}60$ кг/мм^2, $S_x^2 = 690{,}56$; $S_y^2 = 771{,}84$; $S_x = 26{,}28$; $S_y = 27{,}78$; $\mu_{xy} = n^{-1}\Sigma x_i y_i - \bar{x}\bar{y} = 651{,}52$; $\rho = \mu_{xy}/S_x S_y = 0{,}89$.

в) Проверим значимость выборочного коэффициента корреляции. Для этого вычислим статистику $t_н = \rho\sqrt{n-2}/\sqrt{1-\rho^2} = 13{,}47$. По таблицам квантилей распределения Стьюдента по $\alpha = 0{,}05$ и числу

степеней свободы $k = n - 2 = 48$ находим $t_{\frac{\alpha}{2}, n-2} = t_{0,25;48} = 2,02$.

Так как 13,47 > 2,02, то выборочный коэффициент корреляции значимо отличается от нуля.

г) Запишем эмпирическое линейное уравнение регрессии y на x $\left[\bar{y}_x = \bar{y} + \rho S_y (x - \bar{x})/S_x;\ \bar{y}_x = 0,94x + 25,68\right]$ и уравнение регрессии x на y $\left[\bar{x}_y = 0,85y - 3,81\right]$. Построим графики.



Задачи

5.1. Пусть $y = A\theta + \varepsilon$ – линейная регрессионная модель, $M\varepsilon = 0$, $D\varepsilon = M\varepsilon\varepsilon^T = \sigma^2 J_N$, $\theta^T = (\theta_1, \ldots, \theta_m)$ – неизвестный параметр, A – матрица плана размерности $N \times m$. Показать, что

а) любая МНК-оценка $\hat{\theta}$ параметра θ удовлетворяет системе нормальных уравнений $A^T A \hat{\theta} = A^T y$;

б) оценка $\hat{\theta} = Cy$, где C – матрица размерности $m \times N$ – является несмещенной тогда и только тогда, когда $CA = J_m$;

в) если существует матрица $(A^T A)^{-1}$, то оценка $\hat{\theta} = (A^T A)^{-1} A^T y$ является наилучшей в классе несмещенных оценок.

5.2. Для линейной регрессионной модели $y = A\theta + \varepsilon$,

$$M\varepsilon = 0, \quad D\varepsilon = \sigma^2 J_N, \quad y^T = (y_1, \ldots, y_N)$$

с матрицей плана

$$A^T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_N \end{bmatrix}$$

получить МНК-оценку $\hat{\theta}$ параметра $\theta^T = (\theta_1, \theta_2)$.

5.3. Результат измерения переменной y в точке x равен $y(x) = x^2$; измерены значения y в точках $x_{(1)} = 0$, $x_{(2)} = 0{,}5$, $x_{(3)} = 1$, аппроксимируется кривая $y(x)$ прямой $\eta(x) = \theta_1 + \theta_2 x$. Методом наименьших квадратов оценить неизвестные параметры θ_1 и θ_2.

5.4. Пусть $y = A\theta + \varepsilon$, $M\varepsilon = 0$, $D\varepsilon = M\varepsilon\varepsilon^T = \sigma^2 J_N$ – линейная регрессионная модель, где A – матрица плана размерности $N \times m$ такая, что $(A^T A)^{-1}$ существует.

а) Для МНК-оценки $\hat{\theta} = (A^T A)^{-1} A^T y$ параметра θ найти дисперсионную матрицу $D\hat{\theta}$.

б) Показать, что оценка

$$\hat{\sigma}^2 = (N-m)^{-1} \| y - A\hat{\theta} \|^2$$

является несмещенной оценкой параметра σ^2 (здесь $\|z\|^2$ – квадрат длины вектора z).

5.5. Для линейной регрессионной модели $y = A\theta + \varepsilon$, $M\varepsilon = 0$, $D\varepsilon = \sigma^2 J$ с матрицей плана

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -0{,}6 & -0{,}2 & 0{,}2 & 0{,}6 & 1 \end{bmatrix}$$

измерен вектор $y^T = (-1{,}49;\ 1{,}48;\ 0{,}54;\ 0{,}17;\ 1{,}71;\ 3{,}86)$. Найти МНК-оценки $\hat{\theta}$ и $\hat{\sigma}^2$ неизвестных параметров θ и σ^2.

5.6. Пусть η, $\xi_1, \ldots, \xi_n$ – случайные величины с нулевым средним и с невырожденной ковариационной матрицей

$$[M(\xi_i \xi_j)] = M\xi\xi^T, \quad \text{где } \xi^T = (\xi_1, \ldots, \xi_n).$$

Найти наилучшую линейную оценку $\hat{\eta} = \sum_{i=1}^{n} \alpha_i \xi_i$ по величинам ξ_i. Какой вид имеет эта оценка, если случайные величины $\xi_1, \ldots, \xi_n$ независимы?

5.7. Значения $y(x) = \theta_1 + \theta_2 x + \theta_3 x^2$ измерены в симметричных точках x_i (i = I,..., n) так, что $\sum_{i=1}^{n} x_i = 0$,

$$\sum_{i=1}^{n} x_i^3 = 0. \quad y_i = \theta_1 + \theta_2 x_i + \theta_3 x_i^2 + \varepsilon_i, \quad M\varepsilon_i = 0.$$

Найти:

а) МНК-оценку $\hat{\theta}$ параметра: $\theta^T = (\theta_1, \theta_2, \theta_3)$;

б) $D\hat{\theta}_i$, $Cov(\hat{\theta}_i, \hat{\theta}_j)$, i, j = I, 2, 3.

5.8. Для линейной регрессионной модели

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i, \quad i = 1, \ldots, N, \quad M\varepsilon_i = 0, \quad D\varepsilon_i = \sigma^2.$$

Найти:

а) МНК-оценку неизвестных параметров (θ_1, θ_2), проверить их несмещенность и найти условие их состоятельности;

б) несмещенную оценку $\hat{\sigma}^2$ параметра σ^2. Указать достаточное условие ее состоятельности.

5.9. Предполагая, что в задаче 5.8 ε_i - нормальные $N(0, \sigma^2)$, построить γ -доверительные интервалы для параметров θ_1, θ_2 и σ^2 (γ = 0,95).

5.10. По данным независимых равноточных измерений (y_i, x_i), i = I,..., n, значений некоторой линейной функции $y(x) = \theta_1 + \theta_2 x$ (погрешности измерений подчиняются нормальному распределению $N(0, \sigma^2)$ с неизвестной дисперсией) построить доверительный интервал для интеграла от этой функции на отрезке $[-a, a]$ (a задано). Произвести вычисления для следующих данных: (2,96; -2), (3,20; -I), (3,4I; 0), (3,63; I), (3,79; 2) при a = 2 и доверительной вероятности γ = 0,95.

5.II. Пусть в линейной регрессионной модели $y = A\theta + \varepsilon$, $M\varepsilon = 0$, $D\varepsilon = \sigma^2 J_N$, матрица плана A обладает свойством, что ее столбцы ортогональны. Как выглядят в этом случае МНК-оценки $\hat{\theta}_1, \ldots, \hat{\theta}_m$ параметров $\theta^T = (\theta_1, \ldots, \theta_m)$ и их вторые моменты?

В задачах 5.I2 - 5.24 приводятся результаты наблюдений над некоторой двумерной случайной физической величиной. Используя эти экспериментальные данные, требуется

а) найти числовые характеристики выборки - средние арифметические $\bar{x}$, $\bar{y}$; средние квадратические отклонения S_x, S_y; коэффициент корреляции ρ;

б) проверить значимость коэффициента корреляции;

в) найти эмпирическую функцию регрессии y на x и построить ее график.

5.12. В результате экспериментального исследования зависимости между отклонениями размеров валиков от номинала при чистовой обработке (y , мкм) от отклонения размеров валиков от номинала при черновой обработке (x , мкм) составлена корреляционная таблица:

y \ x	-30	-20	-10	0	n_y
-8	1	-	-	-	1
-4	4	1	-	-	5
0	1	15	1	-	17
4	-	2	13	-	15
8	-	-	2	1	3
12	-	-	-	9	9
n_x	6	18	16	10	$n = 50$

5.13. Результаты экспериментального исследования зависимости коэффициента обрабатываемости (y) от ударной вязкости (x , кг/мм^2) инструментальных быстрорежущих сталей представлены в виде корреляционной таблицы:

y \ x	0,7	0,9	1,1	1,3	1,5	n_y
0,5	3	2	-	-	-	5
0,6	6	14	3	1	-	24
0,7	1	6	27	12	2	48
0,8	-	-	5	4	8	17
0,9	-	-	-	2	4	6
n_x	10	22	35	19	14	$n = 100$

5.14. Результаты исследования зависимости между успеваемостью на I курсе вуза и успеваемостью в средней школе представлены в виде корреляционной таблицы (x - средний балл аттестата, y - средний балл результатов первых двух семестров):

y \ x	3,4	3,8	4,2	4,6	n_y
3,2	6	4	–	–	I0
3,6	2	8	3	–	I3
4,0	–	5	I0	I	I6
4,4	–	–	6	5	II
n_x	8	I7	I9	6	n = 50

5.I5. Результаты экспериментального исследования зависимости годового расхода топлива на один трактор (y , т) от годовой выработки трактора (x , сотни га) представлены в виде корреляционной таблицы:

y \ x	3	5	7	9	II	n_y
4	4	–	–	–	–	4
8	I	6	2	–	–	9
I2	–	4	I0	2	–	I6
I6	–	–	6	9	2	I7
20	–	–	–	I	3	4
n_x	5	I0	I8	I2	5	n = 50

5.I6. Исследовалась зависимость стойкости сверл (y , мин) от толщины сердцевины (x , мм). Результаты измерений представлены в виде корреляционной таблицы:

y \ x	0,75	0,80	0,85	0,90	n_y
I0	6	–	–	–	6
20	4	2	–	–	6
30	–	I2	–	–	I2
40	–	6	4	I	II
50	–	–	6	9	I5
n_x	I0	20	I0	I0	n = 50

5.I7. В результате исследования зависимости времени, затрачи-

ваемого на закрепление детали на токарном станке, от веса детали составлена корреляционная таблица (x – вес детали, кг; y – время закрепления детали, с):

y \ x	8	I0	I2	I4	I6	n_y
2,2	3	2	–	–	–	5
2,4	6	I4	3	I	–	24
2,6	I	6	27	I2	2	48
2,8	–	–	5	4	8	I7
3,0	–	–	–	2	4	6
n_x	I0	22	35	I9	I4	n = I00

5.I8. В результате исследования зависимости успеваемости на 4-м курсе от успеваемости на I-м курсе составлена корреляционная таблица (x – средний балл на I-м курсе, y – средний балл на 4-м курсе):

y \ x	3,2	3,4	3,6	3,8	4,2	4,4	n_y
3,5	3	I	–	–	–	–	4
3,7	I	5	2	–	–	–	8
3,9	–	4	I8	5	–	–	27
4,I	–	–	8	I8	7	–	33
4,3	–	–	–	5	I0	3	I8
4,5	–	–	–	–	I	9	I0
n_x	4	I0	28	28	I8	I2	n = I00

5.I9. В результате экспериментального исследования зависимости температуры смазочного масла в двигателе автомобиля (y , град) от скорости движения (x , км/ч) составлена корреляционная таблица:

y \ x	20	30	40	50	60	n_y
43	I	5	2	–	–	8
45	I	9	4	–	–	I4
47	–	4	40	8	–	52
49	–	–	I	I2	2	I5
51	–	–	I	3	7	II
n_x	2	I8	48	23	9	n = I00

5.20. В результате экспериментального исследования зависимости расхода топлива двигателя (y в сотнях литров) от длительности непрерывной работы (x в сотнях часов) составлена корреляционная таблица:

y \ x	0,5	I,0	I,5	2,0	2,5	3,0	3,5	4,0	n_y
0,3	I	7	2	–	–	–	–	–	I0
0,8	–	2	I5	3	–	–	–	–	20
I,3	–	–	I	30	4	–	–	–	35
I,8	–	–	–	4	I0	8	–	–	22
2,3	–	–	–	–	I	3	6	3	I3
n_x	I	9	I8	37	I5	II	6	3	n = I00

5.21. В результате экспериментального исследования зависимости успеваемости I00 студентов I-го курса (y – средний балл, вычисленный по результатам двух семестров) от числа пропущенных занятий (x, ч) составлена корреляционная таблица:

y \ x	5	I0	I5	20	25	n_y
3,2	–	–	2	I0	I	I3
3,6	–	6	I2	2	–	20
4,0	6	8	I5	I	–	30
4,4	6	II	5	–	–	22
4,8	I0	4	I	–	–	I5
n_x	22	29	35	I3	I	n = I00

5.22. В результате исследования зависимости мощности 50 двигателей (y, кВт) от числа оборотов (yx, сотни об/мин) составлена корреляционная таблица:

y \ x	I25	I26	I27	I28	I29	I30	I3I	n_y
25	3	2	I	–	–	–	–	6
26	–	4	3	3	–	–	–	I0
27	–	2	5	4	2	–	–	I3
28	–	–	I	5	5	2	–	I3
29	–	–	–	I	2	3	–	6
30	–	–	–	–	–	–	2	2
n_x	3	8	I0	I3	9	5	2	n =50

5.23. В результате экспериментального исследования зависимости предела прочности на сжатие 50 брикетов из карбонильного железного порошка (y , сотни кг/мм^2) от плотности брикетов (x , %) составлена корреляционная таблица:

y \ x	75	85	90	95	n_y
2	3	–	–	–	3
3	2	4	I	–	7
4	I	I8	9	–	28
5	–	I	8	I	I0
6	–	–	–	2	2
n_x	6	23	I8	3	n = 50

5.24. В результате исследования зависимости предела выносливости (y , кг/мм^2) от предела прочности (x , кг/мм^2) деформированных алюминиевых сплавов составлена корреляционная таблица:

y \ x	30	35	40	45	50	55	n_y
I2	3	–	–	–	–	–	3
I4	3	5	–	–	–	–	8
I6	–	2	I0	I2	–	–	24
I8	–	–	–	2	3	5	I0
20	–	–	–	–	I	4	5
n_x	6	7	I0	I4	4	9	n = 50

§ 6. МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

Для изучения эффективности различных статистических процедур удобно использовать статистическое моделирование, реализуемое с помощью последовательности псевдослучайных чисел, обладающих свойствами последовательности случайных чисел. Для получения реализации последовательности независимых случайных чисел с произвольным распределением используют реализацию последовательности независимых случайных чисел, равномерно распределенных на [0,I].

Алгоритм получения последовательности равномерно распределенных псевдослучайных чисел, предложенный Д.Лемером, реализуется в

ЭВМ серии ЕС следующим образом. Определяется последовательность целых чисел m_k, $k \geq 0$: задается $m_0 = 1$, а все последующие вычисляются по рекуррентной формуле $m_{k+1} = Mm_k \pmod{2^{31}}$, где $M = 2^{16} + 3 = 65539$. Выражение $Mm_k \pmod{2^{31}}$ означает остаток от деления Mm_k на 2^{31}. По числам m_k вычисляются псевдослучайные числа:

$$\gamma_k = 2^{-31} m_k. \qquad (*)$$

Так устроен датчик RANDU, включенный в математическое обеспечение ЭВМ серии ЕС.

Примеры

1. Указать способ моделирования независимых испытаний в полиномиальной схеме с исходами $1, 2, \ldots, N$ и вероятностями исходов соответственно $P_1, \ldots, P_N$.

Решение. Отрезок $[0,1]$ разобьём на N отрезков: $\Delta_1, \Delta_2, \ldots, \Delta_N$, где $\Delta_1 = [0, P_1]$, $\Delta_k = [P_1 + \ldots + P_{k-1}, P_1 + \ldots + P_k]$, $k = 1, \ldots, N$. Положим $X_r = k$, если $\gamma_r \in \Delta_k$, где $\{\gamma_r\}$ – последовательность псевдослучайных чисел. Очевидно, для полученной последовательности

$$X_1 X_2, \ldots, X_n, \ldots \qquad P(X_r = k) = P(\gamma_r \in \Delta_k) = P_k.$$

2. Предложить способ моделирования случайной последовательности $X_1, X_2, \ldots, X_n, \ldots$, где

$$P(X_n \leq t) = 1 - e^{-t/a}, \quad t \geq 0 \quad (a > 0).$$

Решение. Заметим сначала, что если ξ – равномерно распределенная на $[0,1]$ случайная величина и $F(t)$ – непрерывная функция распределения, имеющая обратную $F^{-1}(x)$, то случайная величина $\eta = F^{-1}(\xi)$ имеет функцию распределения $F(t)$. Действительно,

$$P(\eta \leq t) = P(F^{-1}(\xi) \leq t) = P(\xi \leq F(t)) = F(t).$$

Пусть $\{\gamma_n\}$ - последовательность псевдослучайных чисел. Из уравнения $1 - e^{-t/a} = x$ находим $t = -a \ln(1 - x)$. Положим $x_n = -a \ln(1 - \gamma_n)$. Из сделанного замечания следует, что функция распределения x_n есть $1 - e^{-t/a}$, $t \geq 0$.

Задачи

6.1. Смоделировать выборку $x_1, x_2, x_3, x_4, x_5, x_6$ биномиального распределения $Bi(4; 0,5)$.

6.2. Заданы вероятности трех событий A_1, A_2, A_3, образующих полную группу, $P(A_1) = 0,22$; $P(A_2) = 0,31$; $P(A_3) = 0,47$. Смоделировать эксперимент, в результате которого появляется одно из трех событий, получить пять результатов таких независимых экспериментов.

6.3. Заданы вероятности четырех событий, образующих полную группу: $P_1 = 0,15$; $P_2 = 0,64$; $P_3 = 0,05$; $P_4 = 0,16$. Смоделировать эксперимент, в результате которого появляется одно из этих событий, получить результаты 10-ти испытаний данного эксперимента.

6.4. События A и B независимы и совместны. Смоделировать:

а) 4 независимые испытания, в каждом из которых событие A появляется с вероятностью 0,7, а событие B - с вероятностью 0,4;

б) 5 независимых испытаний, в каждом из которых $P(A) = 0,5$; $P(B) = 0,6$; $P(AB) = 0,2$.

6.5. Указать способ моделирования выборки из биномиального распределения $Bi(k, p)$.

6.6. Смоделировать выборку объема $n = 1000$ из распределения Бернулли:

а) $Bi(1; 0,8)$;

б) $Bi(4; 1/3)$.

6.7. Предложить способ моделирования последовательности испытаний Бернулли $x_1, x_2, \ldots, x_n, \ldots$ с вероятностью "успеха" p: $P(x_i = 1) = p$, $P(x_i = 0) = 1 - p$.

Смоделировать последовательность испытаний Бернулли с $p = 0,3$ и $n = 500$. Вычислить частоты μ_k/k, где $\mu_k = x_1 + \ldots + x_k$ при $k = 100, 200, 300, 400, 500$; построить график на плоскости, соединив точки $(k, \mu_k/k)$.

6.8. Пусть частица движется по целым точкам прямой, переходя в дискретные моменты времени в соседнюю точку слева или справа с вероятностями I/2. Указать способ моделирования такого блуждания, если в начальный момент частица находилась в 0.

6.9. Частица, "стартуя" в момент $t = 0$ из точки k $(0<k<N)$, блуждает по целым точкам отрезка $[0,N]$. Если в момент t частица находилась в точке m, то в момент $t+1$ она находится в точке $m+1$ с вероятностью p или в точке $m-1$ с вероятностью $q = 1-p$. В точках 0 и N частица поглощается и случайное блуждание прекращается.

Смоделировать I00 реализаций описанного случайного блуждания при $N = 7$, $k = 3$, $p = 0{,}6$ и $p = 0{,}5$.

6.I0. Смоделировать выборку из полиномиального распределения M (500; 0,2; 0,2; 0,2; 0,2; 0,2).

6.II. Смоделировать последовательность $\{X_i\}$ полиномиальных псевдослучайных величин, принимающих значения I,..., N. Образовать из этой последовательности две выборки: $\{X_{2i}, i = 1,\dots,n\}$ и $\{X_{2i+1}, i = 0,\dots,n-1\}$. С помощью критерия χ^2 проверить гипотезу независимости величин, соответствующих данным выборкам. Расчеты провести для $N = 2, 4, 10$; $n = 100$.

6.I2. Смоделировать независимые равномерно распределенные на $[a,b]$ величины $X_1,\dots,X_n$. Получить четыре значения X_1, X_2, X_3, X_4, если $[a,b] = [4, 14]$.

6.I3. Указать способ моделирования выборки из распределения с плотностью $f(x) = b(1+ax^2)^{-1}$ в интервале $[0, (b-a)^{-1}]$. Получить пять значений при I) $a = 1$, $b = 1$; 2) $a = 2$, $b = 10$.

6.I4. Смоделировать независимые показательно распределенные величины $X_1, X_2, \dots, X_n$ с $MX_i = 1$, $n = 100$. Построить график эмпирической функции распределения и гистограмму; вычислить первый и второй выборочные моменты.

6.I5. Используя центральную предельную теорему, указать способ моделирования приближенно нормально распределенных случайных чисел X_n, $n = 1, 2, \dots$.

6.I6. Указать способ моделирования выборки из распределения Вейбулла, заданного плотностью вероятности

$$f(x) = \frac{n}{x_0} x^{n-1} e^{-x^n/x_0}$$

при $x \geqslant 0$; $f(x) = 0$ при $x < 0$.

6.17. Указать способ моделирования выборки из распределения Релея, заданного плотностью вероятности

$$f(x) = x\sigma^{-2}\exp(-x^2/2\sigma^2), \quad x \geq 0, \quad f(x) = 0, \quad x < 0.$$

6.18. Указать способ моделирования выборки из распределения с плотностью вероятности

$$f(x) = 0{,}5\sin x, \quad x \in (0, \pi), \quad f(x) = 0, \quad x \notin (0, \pi).$$

6.19. Указать способ моделирования эрлангового случайной последовательности $X_1, X_2, \ldots, X_n, \ldots, \quad L(X_j) = \Gamma(a, m)$, $j = 1, 2, \ldots$.

В задачах 6.20 – 6.22 смоделировать случайный вектор (X, Y), составляющие которого независимы.

6.20. Случайный вектор задан законом:

а)

$y \backslash x$	x_1	x_2	x_3
y_1	0,18	0,30	0,12
y_2	0,12	0,20	0,08

б)

$y \backslash x$	x_1	x_2	x_3
y_1	0,20	0,08	0,12
y_2	0,30	0,12	0,18

Получить пять пар значений.

6.21. Непрерывный случайный вектор задан плотностью вероятности $f(x, y) = 0{,}75\,xy^2$ в области, ограниченной $x = 0, y = 0$, $x = 1$, $y = 2$.

6.22. Непрерывный случайный вектор задан плотностью вероятности $f(x, y) = 4(x, y)$ в квадрате $[0{,}1 \times 0{,}1]$.

6.23. Смоделировать наблюдения $y_i = \theta_1 + \theta_2 x_i + \varepsilon_i$, $i = 1, \ldots, n$, если $n = 100$, $x_i = 2i/n$, $\theta_1 = 2$, $\theta_2 = 1$,

а) ε_i – независимые равномерно распределенные случайные величины на отрезке $-1{,}386;\ 1{,}386$;

б) ε_i – нормальные с $M\varepsilon_i = 0$, $D\varepsilon_i = 0{,}16$.

По полученным данным найти МНК-оценки $\hat{\theta}_1, \hat{\theta}_2$, построить графики функций $y(x) = 2 + x$, $\hat{y}(x) = \hat{\theta}_1 + \hat{\theta}_2 x$ на отрезке $[0,2]$.

6.24. Смоделировать наблюдения $y_i = \theta_1 + \theta_2 x_i + \theta_3 x_i^2 + \varepsilon_i$, $i = 1, \ldots, n$, $\theta_1 = -8$, $\theta_2 = 10$, $\theta_3 = -2$, $n = 100$, $x_i = 1 + 2i/n$,

а) ε_i - независимые равномерно распределенные случайные величины на отрезке (-I,386; I,386);

б) ε_i - нормальные с $M\varepsilon_i = 0$, $D\varepsilon_i = 0{,}16$.

По полученным данным найти МНК-оценки $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, построить графики функций $y(x) = \theta_1 + x\theta_2 + \theta_3 x^2$, $\hat{y}(x) = \hat{\theta}_1 + \hat{\theta}_2 x + \hat{\theta}_3 x^2$ на отрезке [I,3].

6.25. Смоделировать наблюдения $y_i = x_i^2 + \varepsilon_i$, $i = 1, \ldots, n$, если $n = 100$, $x_i = 2 + 0{,}1(i - 1)$, ε_i - независимые равномерно распределенные случайные величины на отрезке [0; 0,7].

а) Построить по этим данным интерполяционный многочлен

$$\varphi_k(x, \hat{\theta}) = \sum_{j=1}^{k} \hat{\theta}_j a_j(t)$$

для $k = 2, 3, 4$, где $a_j(t)$, $j = 1, \ldots, 4$ - многочлены Чебышева. Оценки $\hat{\theta}_j$ находятся по формулам:

$$\hat{\theta}_j = \frac{1}{a_j^2} \sum_{i=1}^{n} a_j(x_i) y_i, \quad a_j^2 = \sum_{i=1}^{n} a_j^2(x_i), \quad j = 1, 2, \ldots .$$

б) Построить графики функций $y = x^2$, $y = \varphi_k(x, \hat{\theta})$, $k = 2, 3, 4$.

в) Решить задачу при ε_i нормально распределенных с $M\varepsilon_i = 0$, $D\varepsilon_i = 0{,}04$.

г) Решить задачу с $y_i = e^{x_i} + \varepsilon_i$.

§ I

I.3. $\bar{t} = 1502{,}5$; $s^2 = 40720$. I.19. $a = 2$, $\mu_2 = 5/6$, $\delta = 0{,}044$. $P(|\bar{x} - 2| = 0{,}003) = 0{,}999$. I.20. $\delta = \sqrt{11{,}9167/500} \cdot 2{,}326 = 0{,}359$; наблюдавшееся значение отклонения $|x - a| = 5{,}942 - 6 = 0{,}058$, т.е. согласие хорошее.

I.22. Решение. Принимая во внимание, что $\mathcal{L}(\xi) = Bi(12; 0{,}5)$, находим теоретическое среднее $a_1 = 12 \cdot 1/2 = 6$ и выборочное среднее $\bar{x} = 6{,}1$. Следовательно, $\varepsilon = 6{,}1 - 6 = 0{,}1$. Из теоремы I.I следует, что при больших n $\bar{x} \sim N(a_1, (a_2 - a_1^2)/n) = N(6, 3/n)$ (так как $a_2 - a_1^2 = D\xi = 12 \cdot 0{,}25 = 3$). Величина $\eta_n = \sqrt{n/3} \times (\bar{x} - 6) \sim N(0,1)$. При $n = 4096$ получаем $\sqrt{n/3} \doteq 37$. Поэтому $P(|\bar{x} - 6| < 0{,}1) = P(|\eta_n| < 3{,}7) = 2\Phi(3{,}7) - 1 = 0{,}9998$.

I.23. Решение. а) При больших n по формуле Муавра-Лапласа

$$P(|\nu_n/n - p| < t\sqrt{pq/n}) = P(|\nu_n - np|/\sqrt{npq} \le t) = 2\Phi(t) - 1 = \gamma.$$

Из уравнения $\Phi(t) = 0{,}5(1 + \gamma)$ находим $t_\gamma = 0{,}5(1 + \gamma)$ – квантиль, затем $\delta_\gamma = t_\gamma\sqrt{pq/n}$. При $\gamma = 0{,}98$ имеем $\delta_\gamma = 0{,}0183$. Для эксперимента Бюффона $|k/n - 0{,}5| = 0{,}007$. Укладываются.

б) $|k/n - 0{,}5| = 0{,}0089 < \delta_\gamma = 0{,}0183$. Укладываются.

I.24. Решение. События $(X_{(r)} \le x)$ и $(\mu_n(x) \ge r)$ эквивалентны (здесь $\mu_n(x)$ – число значений выборки, не превосходящих x). Поэтому

$$F_{rs}(x, y) = P(X_{(r)} \le x;\ X_{(s)} \le y) = P(\mu_n(x) \ge r, \mu_n(y) \ge s).$$

Пусть $x < y$. Для произвольной статистики X_k имеем

$$P(X_k \le x) = F(x), \quad P(x < X_k \le y) = F(y) - F(x),$$

$P(X_k>y)=1-F(x)$. Совместное событие $(\mu_n(x)\geqslant r, \mu_n(y)\geqslant s)$ будет выполняться, если в промежуток $(-\infty, x)$ попадёт не менее r значений выборки, а в $(-\infty, y]$ – не менее s значений. Таким образом, для $x<y$ находим (полиномиальное распределение):

$$F_{rs}(x,y)=\sum_{i=r}^{n}\sum_{j=max(0,s-i)}^{n-i}\frac{n!}{i!j!(n-i-j)!}F^{i}(x)[F(y)-F(x)]^{j}[1-F(y)]^{n-i-j}.$$

Очевидно, для $x\geqslant y$

$$F_{rs}(x,y)=F_{X_{(s)}}(y)=\sum_{m=s}^{n}C_n^m F^m(y)\left[1-F(y)\right]^{n-m}.$$

Для $x<y$ событие $\{X_{(r)}\in(x, x+h_1), X_{(s)}\in(y, y+h_2)\}$ осуществляется тогда и только тогда, когда $r-1$ из всех наблюдений меньше x, одно наблюдение попадёт в интервал $(x, x+h_1)$, $s-r-1$ наблюдений – в интервал между $x+h_1$ и y, одно наблюдение – в интервал $(y, y+h_2)$, остальные $n-s$ наблюдений больше $y+h_2$. В силу независимости наблюдений вероятность указанного события при малых h_1 и h_2 с точностью до членов, имеющих более высокий порядок малости, будет

$$C_n^{r-1}F^{r-1}(x)\,(n-r+1)f(x)h_1C_{n-r}^{s-r-1}\left[F(y)-F(x)\right]^{s-r-1}\times$$
$$\times(n-s+1)f(y)h_2\left[1-F(y)\right]^{n-s}.$$

Разделив на h_1h_2 и устремляя h_1 и h_2 к нулю, получим:

$$f_{rs}(x,y)=\frac{n!}{(r-1)!(s-r-1)!(n-s)!}F^{r-1}(x)f(x)\left[F(y)-F(x)\right]^{s-r-1}\times$$
$$\times f(y)\left[1-F(y)\right]^{n-s},\quad x<y.$$

I.26.

$$MX_{(k)}=\sum_{j=n-k+1}^{n}1/j,\quad DX_{(k)}=\sum_{j=n-k+1}^{n}1/j^2.$$

Р е ш е н и е . Якобиан преобразования $y = nx_1$, $y_2 = (n-1)\times(x_2 - x_1), \ldots, y_n = x_n - x_{n-1}$ равен $J(x_1, \ldots, x_n) = n!$ Отсюда, учитывая, что совместную плотность распределения величин $X_{(k)}$ $(k = 1, \ldots, n)$ $n!\exp(-\sum_{k=1}^{n} x_k)$, $0 \le x_1 \le x_2 \le \ldots \le x_n$ (задача I.25) можно записать в виде $n!\exp(-\sum_{k=1}^{n}(n-k+1)(x_k - x_{k-1}))$, получаем, что совместная плотность распределения $y_1, \ldots, y_n$ есть $\exp(-y_1 - \ldots - y_n)$. Далее $X_{(k)} = \sum_{j=1}^{n} \sqrt{y_j}/(n-j+1)$, поэтому $MX_{(k)} = \sum_{j=1}^{n}(n-j+1)^{-1} My_j$,

$DX_{(k)} = \sum_{j=1}^{n}(n-j+1)^{-2} Dy_j$. Следовательно,

$$MX_{(k)} = \sum_{j=n-k+1}^{n} 1/j, \quad DX_{(k)} = \sum_{j=n-k+1}^{n} 1/j^2. \qquad \text{При } n \to \infty$$

$$MX_{(n)} = \sum_{j=1}^{n} 1/j = \ln n + C + O(1), \quad DX_{(n)} = \sum_{j=1}^{n} 1/j^2 = \pi^2/6 + O(1),$$

где $C = 0{,}5772\ldots$ – константа Эйлера.

I.27.

$$\varphi(x,y) = n(n-1)(b-a)^{-n}(y-x)^{n-2}, \quad a \le x \le y \le b.$$

$$MX_{(1)} = (na+b)(n+1)^{-1}, \quad MX_{(n)} = (nb+a)(n+1)^{-1},$$

$$DX_{(1)} = DX_{(n)} = n(b-a)^2(n+1)^{-2}(n+2)^{-1},$$

$$Cov(X_{(1)}, X_{(n)}) = (b-a)^2(n+1)^{-2}(n+2)^{-1}.$$

I.28. $Cov(A_{nk}, A_{ns}) = n^{-1}(a_{k+s} - a_k a_s)$.

I.29. $L(X_{(k)}) = \beta(k, n-k+1)$, $MX_{(k)} = k(n+1)^{-1}$,

$DX_{(k)} = k(n-k+1)(n+1)^{-2}(n+2)^{-1}$, n – объем выборки.

I.30. $P(X_{(1)} > x) = [1 - F(x)]^n = \exp\left[-n\left(\frac{x-a}{b}\right)^{\alpha}\right]$, $x \ge a$,

отсюда $P(X_{(1)} \le x) = 1 - \exp\left[-n\left(\frac{x-a}{b}\right)^{\alpha}\right]$, $x \ge a$, а также

$$P(n^{1/\alpha}(X_{(1)}-a)b^{-1}\le t)=1-\exp(-t^{\alpha}),\ t\ge 0.$$

Таким образом, случайная величина $n^{1/\alpha}(X_{(1)}-a)b^{-1}$ имеет распределение $W(0,\alpha,1)$. Отсюда

$$MX_{(1)}=a+b\Gamma(1+1/\alpha)n^{-1/\alpha},\ DX_{(1)}=b^2[\Gamma(1+2/\alpha)-\Gamma^2(1+1/\alpha)]n^{-2/\alpha}.$$

I.31. Вектор $(\bar{X},\ X_1-\bar{X},\ \ldots,\ X_n-\bar{X})$ распределен по нормальному закону, поскольку представляет собой линейное преобразование нормального вектора X. $Cov(\bar{X},X_i-\bar{X})=Cov(\bar{X},X_i)-D\bar{X}=$

$=n^{-1}DX_i-D\bar{X}=\sigma^2/n-\sigma^2/n=0,\ i=1,\ldots,n$. Следовательно, первая компонента не зависит от остальных.

I.32.

$$M(\chi_n^2)^k=n(n+2)\ldots(n+2(k-1)).$$

Указание. По центральной предельной теореме $(\chi_n^2-n)/\sqrt{2n}$ при $n\to\infty$ асимптотически нормальна $N(0,1)$.

I.33. $Cov(\bar{X},S^2)=(n-1)n^{-2}\mu_3$.

I.35. Указание. Использовать формулу Стирлинга для гамма-функции и применить закон больших чисел к χ^2/n.

I.37. Решение. Рассмотрим r-мерные векторы

$\xi_s=(X_s^{k_1},\ldots,X_s^{k_r})$, $s=1,2,\ldots,n$. Они независимы, одинаково распределены и $M\xi_1=a$, $D\xi_1=[Cov(X_1^{k_i},X_1^{k_j})]_1^r=\Sigma$. По центральной предельной теореме при $n\to\infty$ имеем

$$L(n^{-1/2}(\xi_1+\ldots+\xi_n-na))\to N(0,\Sigma).$$

Остается заметить, что

$$n^{-1/2}(\xi_1+\ldots+\xi_n-na)=\sqrt{n}\,(A_{nk_1}-a_{k_1},\ldots,A_{nk_r}-a_{k_r}).$$

I.38. $P(|F_n(x)-p_0|\le t/\sqrt{n})=2\Phi(t/\sqrt{p_0q_0})-1,$

$p_0=F(x_0),\ q_0=1-p_0.$

Р е ш е н и е . $F_n(x_0) = n^{-1}\mu(x_0)$, где $\mu_n(x_0)$ - число значений выборки, не превосходящих x_0. Так как $L(\mu_n(x_0)) = Bi(n, P_0)$, где $P_0 = F(x_0)$, то по теореме Муавра-Лапласа

$$P\left(\frac{\mu_n(x_0) - nP_0}{\sqrt{nP_0 q_0}} \le z\right) \to \Phi(z), \quad n \to \infty.$$

Отсюда $P(|F_n(x_0) - P_0| \le t/\sqrt{n}) \to 2\Phi(t/\sqrt{P_0 q_0}) - 1.$

I.39. Р е ш е н и е . Рассмотрим полную группу из N событий: $A_1 = (\xi \le x_1)$, $A_2 = (x_1 < \xi \le x_2)$, ..., $A_{N-1} = (x_{N-2} < \xi \le x_{N-1})$, $A_N = (\xi > x_{N-1})$. Их вероятности соответственно равны $P_1, \ldots, P_N$. Тогда ν_i, очевидно, есть число реализаций событий A_i в n независимых и однородных испытаний, $i = 1, 2, \ldots, N$. Следовательно,

$$L(\bar{\nu}) = M(n; P_1, \ldots, P_N).$$

I.40. Р е ш е н и е . Слагаемые в $F_n(x_1, x_2)$ независимы и имеют такое же распределение, что и величина $\eta = e(x_1 - \xi_1)e(x_2 - \xi_2)$, поэтому

$$MF_n(x_1, x_2) = M\eta = P(\eta = 1) = F(x_1, x_2),$$

$$DF_n(x_1, x_2) = n^{-1}D\eta = n^{-1}F(x_1, x_2)[1 - F(x_1, x_2)].$$

Отсюда по неравенству Чебышева

$$P(|F_n(x_1, x_2) - F(x_1, x_2)| > \varepsilon) \le \varepsilon^{-2} DF_n(x_1, x_2) \to 0, \; n \to \infty.$$

Обозначим через $X_j = (X_{1j}, \ldots, X_{nj})$, $j = 1, 2$, $\bar{X}_j$, $S_j^2 = S^2(X_j)$ соответственно выборочные средние и дисперсии, через

$$S_{12} = n^{-1}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) = n^{-1}\sum_{i=1}^{n} X_{i1}X_{i2} - \bar{X}_1\bar{X}_2 -$$

выборочную ковариацию. Тогда $\rho = s_{12}/s_1 s_2$. Если $M(\xi_1^2 \xi_2^2) < \infty$, то существует

$$Д\left(n^{-1}\sum_{i=1}^{n} x_{i1} x_{i2}\right) = n^{-1} Д(\xi_1 \xi_2),$$

а из неравенства Чебышева следует, что

$$n^{-1}\sum_{i=1}^{n} x_{i1} x_{i2} \xrightarrow{P} M(\xi_1 \xi_2),\ n \to \infty.$$

Также $\overline{X}_j \xrightarrow{P} M\xi_j$, $s^2(X_j) \xrightarrow{P} Д\xi_j$, $j = 1,2$, поэтому $\rho_n \xrightarrow{P} \rho$.

§ 2

2.1. Решение. а) $M\overline{X} = n^{-1}\sum_{i=1}^{n} Mx_i = a$; состоятельность оценки $Д\overline{X} = n^{-2}\sum_{i=1}^{n} Дx_i = \sigma^2/n$. По неравенству Чебышева, имеем

$$P(|\overline{x} - a| \geq \varepsilon) \leq \varepsilon^{-2} Д\overline{x} = \sigma^2/\varepsilon^2 n \to 0,\ n \to \infty.$$

б) Введём случайные величины: $y_j = x_j - a$, $My_j = 0$, $Дy_j = My_j^2 = \sigma^2$. Очевидно, $\tilde{\sigma}^2 = k\sum_{j=1}^{n}(y_j - \overline{y})^2$, где $\overline{y} = n^{-1}\sum_{i=1}^{n} y_i$. $M(y_i - \overline{y})^2 = M\left(y_i - n^{-1}\sum_{k=1}^{n} y_k\right)^2 =$

$$M\left(y_j^2 - 2n^{-1}\sum_{k=1}^{n} y_k + n^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n} y_i y_k\right) = My_j^2 - 2n^{-1}\sum_{k=1}^{n} M(y_j y_k) +$$

$$+ n^{-2}\sum_{i=1}^{n}\sum_{k=1}^{n} M(y_i y_k) = \sigma^2 - 2\sigma^2/n + \sigma^2/n = (n-1)n^{-1}\sigma^2.$$

Поэтому $M\tilde{\sigma}^2 = k(n-1)\sigma^2$. Из равенства $M\tilde{\sigma}^2 = \sigma^2$ находим $k = (n-1)^{-1}$. 2.2. Да. 2.3. $\overline{V} = 424{,}73$ м/с; $\tilde{\sigma} = 8{,}84$ м/с. 2.5. а) 1287,8 м^2; б) 1508,5 м^2. 2.6. 10,58 м. 2.7. а) 814,87м^2; б) 921,86 м^2. 2.8. $\tilde{a} = 2{,}24$; $\tilde{\sigma} = 22{,}38$. 2.9. $\tilde{a} = \overline{x}$, $\tilde{\sigma}^2 = s^2$. 2.10. $\tilde{\theta} = 75{,}47$. 2.11. $\tilde{\theta} = 0{,}65$. 2.12. $\hat{\theta} = z_{n/2}$ – выборочная медиана. 2.14. $\tilde{\theta} = n^{-1}\sum_{i=1}^{n} x_i$. 2.15. $a^* = \overline{x}$, $\sigma^* = s$.

2.16. $\tilde{a} = 75{,}4$; $\tilde{\lambda} = 2{,}12$. 2.17. 0,2. 2.18. 0,5 . 2.19. 0,9. 2.20. $\tilde{\theta}_1 = 2{,}17$; $\tilde{\theta}_2 = 3{,}57$. 2.21. Р е ш е н и е . Пусть $M\xi = a$, $D\xi = \sigma^2$, можно считать, что $a = 0$. Тогда

$$M(x_{j+1} - x_j)^2 = Mx_{j+1}^2 + Mx_j^2 - 2Mx_j x_{j+1} = 2\sigma^2,$$

$$M\tilde{\sigma}^2 = k\sum_{j=1}^{n} M(x_{j+1} - x_j)^2 = 2k(n-1)\sigma^2.$$

Из равенства $M\tilde{\sigma}^2 = \sigma^2$ находим $k = 1/2(n-1)$. 2.22. $D\tilde{\sigma}^2 = 2\sigma^4/(n-1)$. Да. 2.23. Р е ш е н и е . Положим

$$y_j = x_j - \bar{x}, \quad My_j = 0, \quad Dy_j = D\Big[(1-n^{-1})x_j - n^{-1}\sum_{i\neq j} x_i\Big] =$$

$$= \left(\frac{n-1}{n}\right)^2\sigma^2 + \frac{n-1}{n^2}\sigma^2 = \frac{n-1}{n}\sigma^2$$

. Следовательно, случайные величины y_j нормальные $N(0, (n-1)n^{-1}\sigma^2)$ с плотностью

$$f(x) = \frac{1}{\sigma}\sqrt{\frac{n}{2\pi(n-1)}}\, e^{-\frac{x^2 n}{2(n-1)\sigma^2}}.$$

Поэтому

$$M|y_j| = 2\int_0^\infty x f(x)\,dx = \frac{1}{\sigma}\sqrt{\frac{n}{2\pi(n-1)}}\int_0^\infty e^{-\frac{x^2 n}{2(n-1)\sigma^2}}\,dx^2 = \sqrt{\frac{2(n-1)}{\pi n}}\,\sigma.$$

$M\tilde{\sigma} = k\sqrt{2(n-1)/\pi n}\; n\sigma$. Из равенства $M\tilde{\sigma} = \sigma$ находим $k = \sqrt{\pi/2n(n-1)}$. Ответ. $k = \sqrt{\pi/2n(n-1)}$. 2.24. а) Да; б) $D\tilde{\theta}_2 < D\tilde{\theta}_1$, т.е. $\tilde{\theta}_2$ эффективнее. Р е ш е н и е . а) $Mx_i = \theta$, следовательно, $M\tilde{\theta}_1 = n^{-1}\sum_{i=1}^{n} Mx_i = \theta$. Пусть $a = \min(x_i)$, $\tilde{b} = \max(x_i)$, известно (пример 4), что $M\tilde{a} = a + (b-a)(n+1)^{-1}$,

$$M\tilde{b} = b - (b-a)(n+1)^{-1}.$$

Поэтому $M\tilde{\theta}_2 = 0{,}5(a + b) = \theta$.

б) $D\tilde{\theta}_1 = n^{-2}\sum_{i=1}^{n} Dx_i = (b-a)^2/12n$. Так как (пример 4)

$$f_{\tilde a}(x)=n(b-a)^{-n}(b-x)^{n-1},\ x\in[a,b],\ f_{\tilde b}(x)=n(b-a)^{-n}(x-a)^{n-1},\ x\in[a,b],$$

то

$$D\tilde a=D\tilde b=n(b-a)^2(n+1)^{-2}(n+2)^{-1},\ D\tilde\theta_2=0{,}25(\tilde a+\tilde b)\le D\tilde a.$$

Для всех $n\geqslant 8$ $D\tilde\theta_2<D\tilde\theta_1$. 2.25. Р е ш е н и е . а) k_1 = I/2. Так как плотность $\max\limits_i(x_i)$ будет (пример 4) $f(x)=n(x-a)^{n-1}$, $x\in[a,b]$, то $M\tilde a_2=n\int\limits_a^{a+1}x(x-a)^{n-1}dx-k_2=$

$$=n(n+1)+a-k_2 .$$

Из уравнения $M\tilde a_2=a$ находим $k=n(n+1)^{-1}$.

б) $D\tilde a_1$ = I/I2 n .

$$D\tilde a_2=n\int\limits_a^{a+1}(x-a-n(n+1)^{-1})^2(x-a)^{n-1}dx=n(n+1)^{-2}(n+2)^{-1}.$$

Очевидно, при $n\geqslant 8$ $D\tilde a_2\le D\tilde a_1$.

2.26. а) Любая $\tilde\theta\in[0{,}5X_{(n)},X_{(n)}]$; б) $\alpha=\dfrac{(n+1)(n+2)(2n-1)}{5n+4}$,

$\beta=-\dfrac{2(n+1)(2n^3+4n^2-3n-3}{(n+2)(5n+4)}$. Указание. Воспользоваться результатами задачи I.27. 2.27. Нет, так как характеристическая функция, а следовательно, и распределение $\bar X_n$, не зависят от n, поэтому $P(|\bar X_n-\theta|\geqslant\varepsilon)$ не зависит от n.

2.28. $\operatorname{cov}(F_n(x_1),F_n(x_2))=n^{-1}F(x_1)[1-F(x_2)]$.

Р е ш е н и е . Введём обозначения: $A_k(x)=(X_k\le x)$,

$A_k([x_1,x_2])=(x_1<X_k\le x_2)$, $1_k(x)$ – индикаторная функция $A_k(x)$. В этих обозначениях

$$F_n(x_1)=n^{-1}\sum_{k=1}^n 1_k(x_1),\ F_n(x_2)=F_n(x_1)+\zeta\ ,\ \text{где}\ \zeta=\frac{1}{n}\sum_{k=1}^n 1_k([x_1,x_2]).$$

Положим $F(x_1) = P_1$, $F(x_2) = P_2$. Тогда

$$cov(F_n(x_1), F_n(x_2)) = M[(F_n(x_1)-P_1)(F_n(x_2)-P_2)] =$$

$$= M[(F_n(x_1)-P_1)(F_n(x_1)-P_1+\zeta-P_2+P_1)] =$$

$$= DF_n(x_1) - M[\zeta(F_n(x_1)-P_1)].$$

Вычислим второе слагаемое $P_1 M\zeta = F(x_1)[F(x_2)-F(x_1)]$.

$$M[\zeta F_n(x_1)] = n^{-2}\sum_{k=1}^{n}\sum_{j=1}^{n} M 1_k(x_1)\, 1_j([x_1, x_2]) =$$

$$= n^{-2}\sum_{k=1}^{n}\sum_{j\neq k} F(x_1)[F(x_2)-F(x_1)] = n^{-1}(n-1)[F(x_2)-F(x_1)]F(x_1),$$

так как $M 1_k(x_1) 1_k([x_1 x_2]) = 0$. Окончательно получаем

$$cov(F_n(x_1), F_n(x_2)) = n^{-1}F(x_1)[1-F(x_1)] - n^{-1}(n-1)F(x_1)[F(x_2)-$$

$$-F(x_1)] + F(x_1)[F(x_2)-F(x_1)] = n^{-1}F(x_1)[1-F(x_2)].$$

2.29. Первый прибор. Р е ш е н и е . Пусть a – расстояние до цели, x_i' – показания первого прибора, x_i'' – показания второго прибора,

$$Mx_i' = Mx_i'' = a, \quad Dx_i' = \sigma^2, \quad Dx_i'' = 4\sigma^2, \quad \bar{x}' = n^{-1}\sum_{i=1}^{n} x_i',$$

$$D\bar{x}' = \sigma^2/n, \quad \bar{x}'' = (3n)^{-2}\sum_{i=1}^{3n} x_i'', \quad D\bar{x}'' = 4\sigma^2/3n.$$

Очевидно, что $D\bar{x}' < D\bar{x}''$. 2.30. $\tilde{\sigma}^2 = n^{-1}\sum_{k=1}^{n}(x_k-a)^2$.

2.31.

τ^*	$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$	$n^{-1}\sum_{i=1}^{n}(x_i-a)^2$	$\bar{x}/\lambda$	$\bar{x}/k$	$\bar{x}$	$\bar{x}/m$
$D\tau^*$	σ^2/n	$2\theta^4/n$	$\theta^2/\lambda n$	$\theta(1-\theta)/kn$	θ/n	$\theta/[mn(1-\theta^2)]$

σ^{-1}, $2\theta^{-2}$, $2\theta^{-2}$, $0{,}5$, $k/[\theta(1-\theta)]$, θ^{-1}, $m/[\theta(1-\theta^2)]$.

2.33. $\tilde{p} = k/n$, k – число "успехов" в выборке. Оценка $\tilde{p}$ состоятельна. 2.34. $\tilde{\tau} = (X)_r/(k)_r$, $\tilde{\tau}_1 = (k-X)_r/(k)_r$, X – выборка из распределения $L(\xi)$, $(a)_r = a(a-1)\ldots(a-r+1)$, $r \geq 1$, $(a)_0 = 1$. Для любой статистики $T(X)$ из условия несмещенности $MT(X) = \tau(\theta)$ следует $\sum_{j=0}^{k} T(j) C_k^j \theta^j (1-\theta)^{k-j} = \theta^r$ для любого $\theta \in (0,1)$. Слева стоит многочлен степени не выше k, поэтому равенство возможно лишь при $r \leq k$. Производящая функция распределения $Bi(k, \theta)$ есть $\psi(z,\theta) = [1+(z-1)\theta]^k$. Известно, что $\frac{\partial^r}{\partial z^r}\psi(z,\theta)\Big|_{z=1} = M(X)_r = (k)_r \theta^r$. Утверждение для $\tau_1(\theta)$ следует из того, что величина $(k - \xi)$ имеет распределение $Bi(k, 1-\theta)$. 2.35. Да. Например,

$$\tilde{\tau}(X) = \sum_{k=1}^{n} \prod_{i=1}^{k} (1-X_i), \; M\tilde{\tau}(X) = \theta^{-1}\left[1+(1-\theta)^{n+1}\right].$$

2.37. $D\tau = \tau'(\theta)/n$. 2.38. $D\tau = (\pi - 2)\theta^2/2n$.

2.39. $D\bar{X} = \pi^2/3n$. 2.40. Воспользоваться свойством преобразования Фурье: если для любого $t \in \mathbb{R}$ $\int_{-\infty}^{\infty} \varphi(x) \exp(ixt)\,dx = 0$, то $\varphi(x) = 0$.

2.42. а) $\lambda = n$; б) $T^* = k^{-1}(X_{(1)} + \ldots + X_{(k)}) + k^{-1}(n-k)X_{(k)}$.

2.43. $\hat{\tau}_n = \lambda/\bar{X}$, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, $L(\hat{\tau}_n) \sim N(\theta^{-1}, 1/n\lambda\theta^2)$.

2.44. $\tau^* = (n-1)\int_0^1 \varphi([tT(X)]^{1/\lambda})(1-t)^{n-2}\,dt$.

2.46. Рассмотреть случайную величину $\xi_1 + \xi_2$.

§ 3

3.1. (94,9; 105,1). 3.2. (2191,8; 2308,2). 3.3. $n \geq 2$.

3.4. а) $(\tilde{a} - \varepsilon, \tilde{a} + \varepsilon) = (2991,2; 3008,8)$. $(\gamma_1\tilde{\sigma}, \gamma_2\tilde{\sigma}) = (15,5; 28,74)$; б) $P(|\tilde{a} - a| < 10) = 0,93$; $P(|\tilde{\sigma} - \sigma| < 2) = 0,41$.

3.5. (5,249 с; 5,751 с). 3.6. ($4,761\ 10^{-10}$; $4,805\ 10^{-10}$).

3.7. а) 0,27; б) 0,035. 3.8. а) (420,75; 428,65);(6,69;I2,70); б) 0,6I; 0,76. 3.I4. 0,55; 0,34. 3.I5. $n > I$. 3.I6. $n \geqslant I5$. 3.I7. $n = I79$. 3.I8. $n = 8I$. 3.20. (0,I66; 0,526). 3.2I. А: (0,I28; 0,872), В: (0,369; 0,63I). 3.22. (0; 0,I49); (0; 0,206). 3.23. (0,303; 0,503). 3.24. (0; 4,6). 3.25. 0,47; 0,7I. 3.26. а)(0,083; 0,II9); б) (0,076; 0,I24). 3.27. (0; 0,0308). 3.28. (0,07; 0,I8). 3.29. (I,I5; 3,24). 3.30. Для А (0,I38; 0,389), для В (I,50; 2,I2), для С (0;0,046). 3.3I. (4I0,2I; I036,56). 3.32. (270,70; 779,82). 3.33. (50,75; 85,I4). 3.34. а) Доверительный интервал ($\nu_1/\bar{\lambda}$, $\nu_2/\bar{\lambda}$), $\nu_1 = 0,64$; $\nu_2 = I,83$; б) $\nu_1 = 0,72$; $\nu_2 = I,53$. 3.35. Использовать факт, что случайная величина $(\bar{X} - \theta)/\theta\sqrt{n}$ имеет распределение $N(0,I)$. 3.36. $\delta_\gamma^*(X) = (T/\chi_{1-a_1^*,n}, T/\chi_{a_2^*,n})$, где a_1^* и a_2^* есть решение системы

$$\chi^2_{1-a_2,n}/\chi^2_{a_1,n} = exp\{n^{-1}(\chi^2_{1-a_2,n} - \chi^2_{a_1,n})\}, \quad a_1 + a_2 = 1-\gamma.$$

3.37. а) $(\bar{X} - t_{\gamma,n-1}\, S/\sqrt{n-1} < \theta_1)$, $(\theta_1 < \bar{X} + t_{\gamma,n-1}\, S/\sqrt{n-1})$,

$$(\bar{X} - t_{\frac{1+\gamma}{2},n-1}\, S/\sqrt{n-1},\ \bar{X} + t_{\frac{1+\gamma}{2},n-1}\, S/\sqrt{n-1}).$$

б) $(nS^2/\chi^2_{\gamma,n-1} < \theta_2^2)$, $(\theta_2^2 < nS^2/\chi^2_{1-\gamma,n-1})$, $(nS^2/\chi^2_{\frac{1+\gamma}{2},n-1}$, $nS^2/\chi^2_{\frac{1-\gamma}{2},n-1})$.

Указание. Использовать факт, что $L(\sqrt{n-1}(\bar{X}-\theta)/S) = S(n-1)$.

3.38. $(\bar{X} - \bar{Y} - c_\gamma\sqrt{\sigma_1^2/n + \sigma_2^2/m}\ ,\ \bar{X} - \bar{Y} + c_\gamma\sqrt{\sigma_1^2/n + \sigma_2^2/m}\ )$,

где $c_\gamma = \Phi^{-1}(0,5(I + \gamma))$. Указание. Проверить, что

$$L((\bar{X} - \bar{Y} - \tau)/\sigma) = N(0,1),\ \sigma^2 = \sigma_1^2/n + \sigma_2^2/m.$$

3.39. Использовать, что $P(X_{(1)} > x) = exp\{-n(x-\theta)\}$.

3.40. (T_1, T_2), где $T_{1,2} = \left(\bar{X} + C_\gamma^2/2n \mp \sqrt{\bar{X}(1-\bar{X})C_\gamma^2/n + C_\gamma^4/n^2}\right)/(1+C_\gamma^2/n)$.

§ 4

4.1. Гипотеза H_0 верна. 4.9. Гипотеза H_0 не отвергается. 4.10. $\chi_q^2 = 7{,}326$; $k = 4$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}122$. Гипотеза H_0 принимается. 4.11. $\chi_q^2 = 11{,}23 > \chi^2_{0{,}95;2} = 5{,}99$. Гипотеза H_0 отвергается. 4.12. $\bar{X} = 5$; $\tilde{p} = 0{,}5$; $\chi_q^2 = 3{,}156$; $k = 9$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}944$. Гипотеза H_0 не опровергается. 4.13. Гипотеза не отвергается. Указание. Сначала методом наибольшего правдоподобия оценить параметр θ в распределении $Bi(2, \theta)$. 4.14. Гипотеза принимается. $\chi_q^2 = 5{,}012$; $k = 9$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}831$. 4.15. $\chi_q^2 = 24{,}9$; $k = 9$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}003$. Гипотеза отвергается. 4.17. $\tilde{t} = 871{,}5$ ч; $\tilde{\lambda} = 0{,}001$; $k = 8$; $\chi_q^2 = 4{,}495$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}808$. Гипотеза не отвергается. 4.18. Гипотеза не отвергается. 4.19. $\chi_q^2 = 0{,}997$; $k = 3$; $P(\chi^2 \geqslant \chi_q^2) = 0{,}802$. Гипотеза H_0 не опровергается. 4.20. Гипотеза H_0 не подтверждается. 4.21. Гипотеза H_0 отвергается. 4.22. Гипотеза H_0 не отвергается. 4.24. Гипотеза H_0 принимается. 4.25. Гипотеза H_0 верна. 4.26. Применим критерий однородности χ^2.

$$\chi_n^2 = n_1 n_2 \sum_{i=1}^{s} \frac{1}{\nu_{i1} + \nu_{i2}} \left(\nu_{i1}/n_1 - \nu_{i2}/n_2\right)^2 = 2{,}18 \quad (s=4, k=2).$$

$\chi^2_{1-\alpha,(s-1)(k-1)} = 6{,}25$, $\chi_n^2 < \chi^2_{0{,}95;3}$, согласие хорошее.

4.27. Гипотеза не отвергается. 4.29. Сначала методом наибольшего правдоподобия оценить параметр θ. 4.30. Критерий Неймана-Пирсона имеет вид $X_{1\alpha}^* = (X \geqslant S)$ и его мощность

$$1-\beta = W(\theta_1) = P(X \geqslant S) = \sum_{m=S}^{\infty} \theta_1^m (1-\theta_1) = \theta_1^S.$$

4.32. $P(T=x)=f(x,\theta)=C_\theta^x C_{n-\theta}^{n-x}/C_N^n$, $x=0,1,\ldots,\theta$,

поэтому функция

$$\ell(x)=\frac{f(x,\theta+1)}{f(x,\theta)}=\frac{\theta+1}{N-\theta}\cdot\frac{N-\theta-n+x}{\theta+1-x}$$

монотонно возрастает по x. Следовательно, равномерно наиболее мощный критерий имеет вид ($T \geqslant t$), т.е. H_0 отвергается, когда T велико.

§ 5

5.2. $\hat\theta=(\hat\theta_1,\hat\theta_2)$, $\hat\theta_1=n^{-1}\sum_{i=1}^N y_i$, $\hat\theta_2=\sum_{i=1}^N x_i y_i/\sum_{i=1}^N x_i^2$.

5.3. θ_1 = -I/I2; θ = I. 5.4. а) $D\hat\theta=\sigma^2(A^TA)^{-1}$.

5.5. $\hat\theta^T$ = (I,07; 5,4I); $\hat\sigma^2$ = I,38.

5.6. $\hat\eta=M(\eta\xi^T)[M(\xi\xi^T)]^{-1}\eta$. Если $\xi_1,\ldots,\xi_n$ независимы, то $\alpha_i=M(\xi_i\eta)/\sigma_i^2$, $\sigma_i^2=D\xi_i$.

5.8. а) $\hat\theta_1=\bar y-x\hat\theta_2$, $\hat\theta_2=\sum_{i=1}^N(x_i-\bar x)(y_i-\bar y)/\sum_{i=1}^N(x_i-\bar x)^2$,

$M\hat\theta_i=\theta_i$, i = I,2 (предполагается, что не все x_i одинаковы).

$$D\hat\theta_1=\frac{\sigma^2}{n}\sum_{i=1}^N x_i^2/\sum_{i=1}^N(x_i-\bar x)^2,\ D\hat\theta_2=\sigma^2/\sum_{i=1}^N(x_i-\bar x)^2,$$

поэтому если $\sum_{i=1}^N(x_i-\bar x)^2\to\infty$ при $N\to\infty$, то оценки $\hat\theta_i$ состоятельны.

б) $\tilde\sigma^2=S(\hat\theta)/(N-2)$, где $S(\theta)=\sum_{i=1}^N(y_i-\bar y)^2-\hat\theta_2^2\sum_{i=1}^N(x_i-\bar x)^2$.

Если $DS(\hat\theta)=O(N^2)$ при $n\to\infty$, то $\hat\sigma^2$ – состоятельная оценка σ^2.

5.I0. Так как $\int_{-a}^{a}y(x)dx=2a\theta_1$,

то нужно построить доверительный интервал для θ_1. 5.II. Матрица $Z=A^TA$ – диагональная, ее диагональные элементы рав-

ны $a_j^T a_j$, $j = 1,\dots,m$, где a_j – j -й столбец матрицы A, поэтому

$$\hat{\theta}_j = (a_j^T y)(a_j^T a_j)^{-1},\ D\hat{\theta}_j = \sigma^2 (a_j^T a_j)^{-1},\ cov(\hat{\theta}_j, \hat{\theta}_i) = 0,\ i \neq j.$$

§ 6

6.5. Если $\xi_1, \dots, \xi_k$ – независимые бернулевские случайные величины с параметром p , то величина $\eta = \xi_1 + \dots + \xi_k$ имеет распределение $Bi(k, p)$. 6.7. Положить $X_n = 1$, если $\gamma_n \leq p$, и $X_n = 0$, если $\gamma_n > p$, где $\{\gamma_n\}$ – последовательность псевдослучайных чисел (ж). 6.8. Если S_n – координата частицы в момент времени t_n , то $S_0 = 0$, $S_n = S_{n-1} + X_n$, $n \geq 1$, где $X_n = 1$, если $\gamma_n \leq 0{,}5$; $X_n = -1$; если $\gamma_n > 0{,}5$, $\{\gamma_n\}$ – последовательность псевдослучайных чисел (ж). 6.9. Если S_n – координата частицы в момент t_n , то $S_0 = k$, $S_n = S_{n-1} + X_n$, $n \geq 1$, $S_{n-1} \neq 0$, $S_{n-1} \neq N$, где $X_n = 1$, если $\gamma_n \leq p$, $X_n = -1$; если $\gamma_n > p$, $\{\gamma_n\}$ – последовательность (ж).

6.13. $X_i = \gamma_i (b - a\gamma_i)$, γ_i – последовательность (ж).

6.15. $$X_n = \frac{1}{\sqrt{N/12}}\left(\gamma_{(n-1)N+1} + \gamma_{(n-1)N+2} + \dots + \gamma_{nN} - \frac{N}{2}\right),$$

где γ_n – последовательность (ж). 6.16. $X_i = \sqrt{-x_0 \ln \gamma_i}$, γ_i – последовательность (ж). 6.17. $X_i = \sigma\sqrt{-2\ln \gamma_i}$, γ_i – последовательность (ж). 6.18. $X_i = \arccos(1 - 2\gamma_i)$, γ_i – последовательность (ж). 6.19. Указание. Если $\xi_1, \dots, \xi_m$ – независимые случайные величины, имеющие показательное распределение с параметром a , то случайная величина $\eta = \xi_1 + \dots + \xi_m$ имеет распределение Эрланга с параметрами (a , m).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Володин В. Г., Ганин М. П., Динер И. Я. и др./ Под ред. А.А.Свешникова. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. М.: Наука, 1970.

Герасимович А. И., Матвеева Я. И. Теория вероятностей и математическая статистика. Минск: Белорусский политехнический ин-т, 1977.

Гихман И. И., Скороход А. В., Ядренко М. И. Теория вероятностей и математическая статистика. Киев: Выща школа, 1979.

Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 1979.

Ивченко Г. И., Медведев Ю. И. Математическая статистика. М.: Высшая школа, 1984.

Ивченко Г. И., Медведев Ю. И., Чистяков А. В. Сборник задач по математической статистике. М.: Высшая школа, 1989.

Мешалкин Л. Д. Сборник задач по теории вероятностей. М.: Изд-во Моск. ун-та, 1963.

П Р И Л О Ж Е Н И Е

Таблица 1

1. Распределение Пуассона. Значение функции $\sum_{k=n} \frac{\lambda^k}{k!} e^{-\lambda}$

n \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
1	0,095	181	259	330	394	451	503	551	593
2	005	018	037	062	090	122	156	191	228
3		001	003	008	014	023	034	047	063
4				001	002	003	006	009	014
5							001	001	002

n \ λ	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0
1	0,632	865	950	982	993	998	999	1,000	1,000
2	264	594	801	908	960	983	924	997	999
3	080	323	577	762	875	938	970	986	994
4	019	143	353	567	735	849	918	958	979
5	004	053	185	371	560	715	827	900	945
6	001	018	084	215	384	554	699	809	884
7		005	034	111	238	394	550	687	793
8		001	012	051	138	256	401	547	676
9				021	068	153	271	408	544
10				008	032	084	170	283	413
11				003	014	043	099	184	294
12				001	005	020	053	112	197
13					002	008	027	068	124
14					001	004	013	034	074
15						001	006	017	042

Таблица 2

Биномиальное распределение

(0,95 – доверительные пределы (θ_1, θ_2) для параметра θ р ($\theta_1 < \theta < \theta_2$) = 0,95)

k \ n-k	0	1	2	3	4	5	6	7	8	9	10	11	12
0		0,98 0,00	84 00	71 00	60 00	52 00	46 00	41 00	37 00	34 00	31 00	29 00	27 00
1	1,00 0,03	0,99 01	91 01	81 01	72 01	64 00	59 00	53 00	48 00	45 00	41 00	39 00	36 00
2	1,00 0,16	0,99 09	93 07	85 05	78 04	71 04	65 03	60 03	56 03	52 02	48 02	45 02	43 02
3	1,00 0,29	0,99 19	95 15	88 12	82 10	76 09	70 08	65 07	61 06	57 06	54 05	51 05	48 04
4	1,00 0,40	1,00 29	96 22	90 18	84 16	79 14	74 12	69 11	65 10	61 09	58 08	55 08	52 07
5	1,00 0,48	1,00 36	96 29	92 25	86 21	81 19	77 17	72 15	68 14	65 13	62 12	59 11	56 10
6	1,00 0,54	1,00 42	97 35	93 30	88 26	83 23	79 21	75 19	71 18	68 16	65 15	62 14	59 13
7	1,00 0,59	1,00 47	97 40	93 35	89 31	85 28	81 25	77 23	73 21	70 20	67 18	64 17	62 16
8	1,00 0,63	1,00 52	98 44	94 39	90 35	86 32	82 29	79 27	75 25	72 23	69 22	67 20	64 19
9	1,00 0,66	1,00 56	98 48	95 43	91 39	87 35	84 32	80 30	77 28	74 26	71 24	69 23	66 22
10	1,00 0,69	1,00 59	98 52	95 46	92 42	88 38	85 35	82 33	79 31	76 29	73 27	70 26	68 24
11	1,00 0,72	1,00 62	98 57	95 49	92 45	89 41	86 38	83 36	80 34	77 32	74 30	72 28	69 27
12	1,00 0,74	1,00 64	98 57	96 52	93 48	90 44	87 41	84 38	81 36	78 34	75 32	73 31	71 29

Примечание.
Значения θ_2 помещены в первых строках, значения θ_1 – во вторых.

Таблица 3

Функция Лапласа $2\Phi_1(x) = \frac{2}{\sqrt{2\pi}} = \int_0^x e^{-t^2/2}\,dt$

x	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,	0,0000	0797	1585	2358	3108	3829	4515	5161	5763	6319
1,	6827	7287	7699	8064	8385	8664	8904	9109	9231	9426
2,	9545	9643	9722	9786	9836	9876	9907	9931	9949	9963
3,	9973	9981	9986	9990	9993	9995	9997	9998	9999	9999

$2\Phi_1(x)$	0,80	0,90	0,95	0,99	0,999
x	1,2816	1,6449	1,9600	2,5758	3,2905

Таблица 4

Критические значения λ_γ для распределения Колмогорова

γ	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
λ_γ	0,828	0,895	0,974	1,073	1,224	1,358	1,520	1,627	1,950

Таблица 5

Распределение χ^2. $P\left(\sqrt{k}/(1+q) < \chi < \sqrt{k}/(1-q)\right)$

k \ q	0,01	0,02	0,06	0,10	0,15	0,20	0,25
1	0,010	019	058	097	145	193	241
2	015	029	088	147	219	290	358
4	022	043	129	214	317	414	504
6	027	054	160	264	388	501	599
8	031	062	186	305	444	567	669
10	035	070	208	340	491	620	722
15	044	086	254	412	583	717	812
20	050	100	293	470	652	784	868
30	061	122	356	559	749	867	930
40	072	142	408	628	815	913	957

Таблица 6

Распределение $\chi^2(n)$. Квантили распределения

$$p = \int_0^{\chi^2_{p,n}} k_n(x)\,dx = 2^{-n/2}\Gamma^{-1}(n/2)\int_0^{\chi^2_{p_n}} x^{n/2-1}\,e^{-x/2}dx$$

n \ p	0,1	0,3	0,5	0,7	0,9	0,95	0,999	0,9999
1	0,016	148	455	1,07	2,71	3,84	6,63	10,8
2	211	713	1,39	2,41	4,61	5,99	9,21	13,8
3	584	1,42	2,37	3,67	6,25	7,82	11,3	16,3
4	1,06	2,20	3,36	4,86	7,88	9,49	13,3	18,5
5	1,61	3,00	4,35	6,06	9,24	11,1	15,1	20,5
6	2,20	3,83	5,35	7,23	10,6	12,6	16,8	22,5
7	2,83	4,67	6,35	8,38	12,0	14,1	18,5	24,3
8	3,49	5,53	7,34	9,52	13,4	15,5	20,1	26,1
9	4,17	6,39	8,34	10,7	14,7	16,9	21,7	27,9
10	4,87	7,27	9,34	11,8	16,0	18,3	23,2	29,6
11	5,58	8,15	10,3	12,9	17,3	19,7	24,7	31,3
12	6,30	9,03	11,3	14,0	18,5	21,0	26,2	32,9
13	7,04	9,93	12,3	15,1	19,8	22,4	25,7	34,5
14	7,99	10,0	13,3	16,2	21,1	23,7	29,1	36,1
15	8,55	11,7	14,3	17,3	22,3	25,0	30,6	37,7
16	9,31	12,6	15,3	18,4	23,5	26,3	32,0	39,3
17	10,09	13,5	16,3	19,5	24,8	27,6	33,4	40,8
18	10,9	14,4	17,3	20,6	26,0	28,9	34,8	42,3
19	11,7	15,4	18,3	21,7	27,2	30,1	36,2	43,8
20	12,4	16,3	19,3	22,8	28,4	31,4	37,6	45,3
21	13,2	17,2	20,3	23,9	29,6	32,7	38,9	46,8
22	14,0	18,1	21,3	24,9	30,8	33,9	40,3	48,3
23	14,8	19,0	22,3	26,0	32,0	35,2	41,6	49,7
24	15,7	19,9	23,3	27,1	33,2	36,4	43,0	51,2
25	16,5	20,9	24,3	28,2	34,3	37,7	44,3	52,6
26	17,3	21,8	25,3	29,2	35,6	38,9	45,6	54,1
27	18,1	22,7	26,3	30,3	36,7	40,1	47,0	55,5
28	18,9	23,6	27,3	31,4	37,9	41,3	48,3	56,9
29	19,8	24,6	28,3	32,5	39,1	42,6	49,6	58,3
30	20,6	25,5	29,3	33,5	40,3	43,8	50,9	59,7

Таблица 7

Распределение $\chi^2(n)$. $P\left(\chi^2 \geq \chi^2_q\right)$

χ^2_q \ n	2	4	6	8	10	12	15	20	25
1	0,606	910	986	999	999	999	999	999	999
2	368	736	920	981	999	999	999	999	999
3	223	558	809	934	981	995	999	999	999
4	135	406	677	857	933	998	999	999	999
5	082	287	544	758	891	958	992	999	999
6	050	199	423	647	815	916	980	999	999
7	030	136	321	537	725	858	958	997	999
8	018	092	238	433	629	785	924	992	999
9	011	061	174	342	532	703	877	983	999
10	007	040	125	265	440	616	820	968	997
12	002	017	062	151	285	446	679	918	987
14	001	007	030	082	173	301	525	830	962
16	000	003	014	042	100	191	382	717	915
18	000	001	006	021	055	116	263	587	842
20	000	000	003	010	029	067	172	458	747
25	000	000	000	002	005	015	050	201	462
30	000	000	000	000	001	003	012	070	224

Таблица 8

Закон распределения Стьюдента

$$p(t,k) = \int_{-\infty}^{t} s_n(x)dx = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(k/2)\sqrt{\pi k}} \int_{-\infty}^{t} (1+x^2/k)^{-(k+1)/2} dx$$

t \ k	1	2	3	4	6	8	10	12	14	16	18	
0,0	0,500	500	500	500	500	500	500	500	500	500	500	500
0,1	532	535	537	537	538	539	539	539	539	539	539	540
0,2	563	570	573	574	576	577	578	578	578	578	578	579
0,3	593	604	608	610	613	614	615	615	616	616	616	618
0,4	621	636	642	645	648	650	651	652	652	653	653	655
0,5	648	667	674	678	683	685	686	687	688	688	688	691
0,6	672	695	705	710	715	717	719	720	721	722	722	726
0,7	694	722	733	739	745	748	750	751	752	753	754	758
0,8	715	746	759	766	773	777	779	780	781	782	783	788
0,9	733	768	783	790	799	803	805	807	808	809	810	816
1,0	750	789	804	813	822	827	830	832	833	834	835	841
1,2	779	823	842	852	862	868	871	873	875	876	877	885
1,4	803	852	872	883	894	900	904	907	908	910	911	919
1,6	822	875	896	908	920	926	930	932	934	935	936	945
1,8	839	893	915	927	939	945	949	951	953	955	956	964
2,0	852	908	930	942	954	960	963	966	967	969	970	977
2,4	874	931	952	963	973	978	981	983	985	986	986	992
2,8	891	946	966	976	984	988	991	992	993	994	994	997
3,2	904	957	975	984	991	993	995	996	997	997	997	999
3,6	914	965	982	989	993	996	998	998	999	999	999	999
4,0	922	971	986	992	996	998	999	999	999	999	999	999

Таблица 9

Значения t_α доверительного интервала $-t_\alpha < t < t_\alpha$, где t имеет распределение Стьюдента; значения γ_1 и γ_2 доверительного интервала для среднеквадратического отклонения σ: $\gamma_1 \tilde{\sigma} < \sigma < \gamma_2 \tilde{\sigma}$ в зависимости от доверительной вероятности α и числа степеней свободы k

k \ α	t_α			γ_1			γ_2		
	0,90	0,95	0,99	0,90	0,95	0,99	0,90	0,95	0,99
1	6,31	12,71	63,7	0,510	0,446	0,356	15,9	31,9	15,9
2	2,92	4,30	9,92	578	521	434	4,40	6,28	14,1
3	2,35	3,18	5,84	620	566	483	2,92	3,73	6,47
4	2,13	2,77	4,60	649	599	519	2,37	2,87	4,39
5	2,02	2,57	4,03	672	624	546	2,09	2,45	3,48
6	1,943	2,45	3,71	690	644	569	1,916	2,20	2,98
7	895	2,36	3,50	705	661	588	797	2,04	2,66
8	860	2,31	3,36	718	675	604	711	1,91	2,44
9	833	2,26	3,25	729	688	618	645	82	2,28
10	812	2,23	3,17	739	699	630	593	75	2,15
11	796	2,20	3,11	748	708	641	550	69	2,06
12	782	2,18	3,06	755	717	651	515	65	1,976
13	771	2,16	3,01	762	725	660	485	61	910
14	761	2,14	2,98	769	732	669	460	57	854
15	753	2,13	2,95	775	739	676	437	54	806
16	746	2,12	2,92	780	745	683	418	52	764
17	740	2,11	2,90	785	750	690	400	49	727
18	734	2,10	2,88	790	756	696	385	47	695
19	729	2,09	2,86	794	760	702	370	46	666
20	725	2,08	2,84	798	765	707	358	44	640
22	717	2,07	2,82	805	773	717	335	41	595
24	711	2,06	2,80	812	781	726	316	39	558
26	706	2,06	2,78	818	788	734	300	37	526
28	701	2,05	2,76	823	794	741	286	35	499
30	697	2,04	2,75	828	799	748	274	33	475
40	684	2,02	2,70	847	821	774	228	27	390
60	671	2,00	2,66	871	849	808	179	21	290
120	658	1,98	2,62	925	912	887	106	11	150
	645	1,96	2,58	-	-	-	-	-	-

Таблица 10

Распределение Снедекора $S(n_1, n_2)$

Значения функции F_{p,n_1,n_2} при $p = 0,95$ (в первых строках) и $p = 0,99$ (во вторых).

$$p = \int_0^{F_{p,n_1,n_2}} f_{n_1,n_2}(x)\,dx.$$

"Левые" границы доверительных интервалов находятся из равенства $F_{1-p,n_1,n_2} = F^{-1}_{p,n_2,n_1}$

n_2 \ n_1	1	2	3	4	6	8
1	161 4052	200 4999	216 5403	225 5625	234 5859	239 5981
2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,33 99,33	19,37 99,36
3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	8,94 27,91	8,84 27,49
4	6,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,16 15,21	6,04 14,80
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	4,95 10,67	4,82 10,27
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,28 8,47	4,15 8,10
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,58 6,37	3,44 6,03
10	4,96 10,04	4,10 7,56	4,71 6,55	3,48 5,99	3,22 5,39	3,07 5,06
12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,00 4,82	2,85 4,50
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,60 3,87	2,45 3,56
30	4,17 7,56	3,32 5,39	2,92 4,51	2,69 4,02	2,42 3,47	2,27 3,17
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,29 3,18	2,13 2,88
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,19 2,99	2,03 2,69
200	3,89 6,76	3,04 4,71	2,65 3,88	2,41 3,41	2,14 2,90	1,98 2,60
	3,85 6,66	3,00 4,62	2,61 3,80	2,38 3,34	2,10 2,82	1,95 2,53

n_2 \ n_1	10	12	20	50	100
1	242 6056	244 6106	248 6208	252 6302	253 6334
2	19,39 99,40	19,41 99,42	19,44 99,45	19,47 99,48	19,49 99,49
3	8,78 27,23	8,74 27,05	8,66 26,69	8,58 26,35	8,56 26,23
4	5,96 14,54	5,91 14,37	5,80 14,02	5,70 13,69	5,66 13,57
5	4,74 10,05	4,68 9,89	4,56 9,55	4,44 9,24	4,40 9,13
6	4,06 7,87	4,00 7,72	3,87 7,39	3,75 7,09	3,71 6,99
8	3,34 5,82	3,28 5,67	3,15 5,36	3,03 5,06	2,98 4,96
10	2,97 4,85	2,91 4,71	2,77 4,41	2,64 4,12	2,59 4,01
12	2,76 4,30	2,69 4,16	2,54 3,86	2,40 3,56	2,35 3,46
20	2,35 3,37	2,28 3,23	2,12 2,94	1,96 2,63	1,90 2,53
30	2,16 2,98	2,09 2,84	1,93 2,55	1,76 2,24	1,69 2,13
50	2,02 2,70	1,95 2,56	1,78 2,26	1,60 1,94	1,52 1,82
100	1,92 2,51	1,85 2,36	1,68 2,06	1,48 1,73	1,39 1,59
200	1,87 2,41	1,80 2,28	1,62 1,97	1,42 1,62	1,32 1,48
1000	1,84 2,34	1,76 2,20	1,58 1,89	1,36 1,54	1,26 1,38

Таблица 11

Критерий Смирнова

Значение $p\left[D_{nn} \leq k/n\right]$, где $D_{nn} = \sup_x \left|F_{1n}(x) - F_{2n}(x)\right|$

k \ n	1	2	3	4	5	6	7	8	9	10	11
1	1,000										
2	0.667	1,000									
3	40	0,900	1,000								
4	229	771	0,971	1.000							
5	127	643	921	0,992	1.000						
6	069	526	857	974	0.998	1,000					
7	037	425	788	947	992	0.999	1.000				
8	020	340	717	913	981	998	1,000	1.000			
9	011	270	648	874	966	993	0.999	1.000			
10	006	213	582	832	948	988	998	1,000			
11	003	167	521	769	925	979	996	0.999	1.000		
12	002	131	464	744	900	969	992	999	1.000		
13	001	102	412	700	874	956	987	997	1,000		
14	000	079	365	657	845	941	981	995	0.999	1.000	
15	000	062	322	614	818	925	974	992	998	1.000	
16	000	048	284	574	785	907	965	989	997	1.000	1.000
17	000	037	249	535	755	888	955	984	995	999	999
18	000	028	219	497	725	868	944	979	993	998	999
19	000	022	192	462	494	847	932	973	991	997	999
20	000	017	168	429	664	825	919	966	988	996	999
21	000	013	147	397	635	804	905	959	984	995	998
22	000	010	128	368	606	782	891	951	980	993	998
23	000	008	112	340	578	759	876	942	975	991	997
24	000	006	098	314	551	737	860	932	970	988	996

ОГЛАВЛЕНИЕ

Учебное издание

АЛЕКСАНДРОВ Евгений Леонидович.

СБОРНИК ЗАДАЧ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Редактор Л.А.Горюнова
Художник Е.И.Бочаров
Технический редактор Т.Н.Крючина
Корректор Р.Э.Арбитман

Н/К

Подписано к печати 30.03.92 г.
Формат 60x84 1/16. Бум. типограф. №2. Печать офсетная.
Усл.печ.л. 6,51(7,0) Уч.-изд.л. 5,7 Тираж 400 экз.
Заказ 313 С 77

Издательство Саратовского университета. 410601, Саратов, Университетская,42.
Типография издательства СГУ. 410601, Саратов, Астраханская, 42.

ИЗДАТЕЛЬСТВО
САРАТОВСКОГО
УНИВЕРСИТЕТА