РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ МАТЕМАТИКИ ИМ. С. Л. СОБОЛЕВА

Д. А. КОРШУНОВ, Н. И. ЧЕРНОВА

СБОРНИК ЗАДАЧ И УПРАЖНЕНИЙ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

ИЗДАНИЕ ВТОРОЕ, ИСПРАВЛЕННОЕ

Учебное пособие

НОВОСИБИРСК
ИЗДАТЕЛЬСТВО ИНСТИТУТА МАТЕМАТИКИ
2004

УДК 519.2
ББК 22.172
К66

Коршунов Д. А., Чернова Н. И.

Сборник задач и упражнений по математической статистике: Учебное пособие. — 2-е изд., испр. — Новосибирск: Изд-во Института математики, 2004. — 128 с.

ISBN 5–86134–121–4.

Сборник содержит 461 задач и упражнений, относящихся к основным разделам учебного курса математической статистики. Весьма широко представлены теоретические задачи на эмпирическое распределение, построение и свойства оценок, интервальное оценивание параметров и проверку статистических гипотез. Приведены решения типовых задач. Все задачи снабжены ответами. В приложение включены таблицы наиболее важных распределений.

Данное учебное пособие предназначено для студентов и аспирантов математических, физических, естественных, технических и экономических специальностей.

Табл. 6. Библиогр. 26 назв.

Адрес авторов: 630090 Новосибирск, Университетский пр., 4.
Институт математики им. С. Л. Соболева СО РАН
E-mail: korshunov@math.nsc.ru, cher@nsu.ru

$\text{К}\dfrac{1602090000-01}{\text{Я}82(03)-04}$ Без объявл.

ISBN 5–86134–121–4

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Второе издание задачника отличается от первого незначительно. Все задачи сохранили прежние номера, что позволяет использовать на практических занятиях как первое, так и второе издания одновременно. Немногочисленные изменения вызваны, в основном, опечатками и неточностями, которые не удалось избежать в первом издании. Мы глубоко признательны своим коллегам и студентам, сообщавшим нам о них.

Новосибирск,
февраль – март 2004 г.

Д. А. Коршунов
Н. И. Чернова

Более развитая, более рефлектированная мера есть необходимость; судьба, Немезида, ограничивается в общем определённостью меры [в том смысле], что всё чрезмерное, всё, что делает себя слишком великим, слишком высоким, приводится ею к другой крайности, умаляется, уничижается и тем самым восстанавливается средняя мера — посредственность.

Г. В. Фр. Гегель. *Наука логики*

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Настоящий сборник призван обеспечить достаточным количеством материала семинарские занятия по курсу «Математическая статистика» на математических факультетах университетов. Свою цель авторы видели в том, чтобы собрать по возможности более широкий набор задач и упражнений, который освещал бы основные разделы стандартного университетского курса математической (теоретической) статистики, преимущественно теорию оценок параметров и теорию проверки гипотез.

В сборник включены в основном теоретические задачи, общим числом более четырёхсот. Источником задач послужили многочисленные книги и сборники задач по статистике. Часть задач и упражнений заимствована из опыта преподавания авторами и их коллегами статистических курсов на различных факультетах Новосибирского государственного университета. При этом авторы стремились унифицировать формулировки задач из разных источников, которые изначально были весьма разнородными.

С целью использования задачника для самостоятельной работы приведены решения основных типовых задач. Идя навстречу многочисленным пожеланиям студентов, мы включили в сборник ответы ко всем задачам.

Мы искренне признательны коллективу кафедры теории вероятностей и математической статистики Новосибирского университета, в составе которого имеем честь работать. Совместная работа

с А. А. Боровковым, И. С. Борисовым, В. И. Лотовым, А. И. Саханенко, С. Г. Фоссом и В. В. Юринским оказала решающее влияние на формирование наших взглядов на преподавание статистики.

Мы хотели бы особо поблагодарить А. Д. Коршунова, В. И. Лотова и С. Г. Фосса, взявших на себя труд просмотреть рукопись, за их замечания и предложения по форме и существу изложения материала, способствовавшие устранению ряда неточностей и неясных мест. Мы будем весьма признательны за любые критические замечания и предложения как по тексту и составу задач, так и по структуре сборника в целом.

Новосибирск,
декабрь 2000 г. – январь 2001 г.

Д. А. Коршунов
Н. И. Чернова

ОТДЕЛ I

ЭМПИРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ

§ 1. Выборка и вариационный ряд

Пусть F — некоторое распределение на действительной прямой. *Выборкой объёма n из распределения F* называется последовательность независимых случайных величин $X_1, \ldots, X_n$ с общим распределением F.

Статистикой называется любая измеримая функция выборки, т. е. любая случайная величина вида $S(X_1, \ldots, X_n)$, где S — измеримая по Борелю функция из $\mathbf{R}^n$ в $\mathbf{R}$.

Важными примерами статистик являются выборочные моменты. Для выборочного среднего значения используется обозначение

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

а для выборочного момента порядка k —

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

Вообще, для произвольной функции $g : \mathbf{R} \to \mathbf{R}$ полагается

$$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

Для выборочной дисперсии используются обозначения

$$S^2 = \overline{X^2} - (\overline{X})^2 \quad \text{и} \quad S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Другие важные примеры статистик связаны с понятием вариационного ряда. Если все n элементов выборки $X_1, \ldots, X_n$ расположены в порядке неубывания их величины и члены такой неубывающей последовательности обозначены $X_{(k)}$: $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$, то каждое из $X_{(k)}$ называется *порядковой статистикой*, а соответствующая неубывающая последовательность — *вариационным рядом*, построенным по выборке $X_1, \ldots, X_n$ объёма n.

Значение $X_{(k)}$, стоящее на k-м месте вариационного ряда, называется *k-й порядковой статистикой.* Случайная величина $X_{(1)}$ называется *минимальным членом вариационного ряда*, а $X_{(n)}$ — *максимальным.*

Выборочной медианой называется статистика

$$\zeta^* = \begin{cases} X_{(m)}, & \text{если } n = 2m-1 \text{ (нечётно)}, \\ \dfrac{X_{(m)} + X_{(m+1)}}{2}, & \text{если } n = 2m \text{ (чётно)}. \end{cases}$$

Выборочной квантилью ζ^*_δ уровня $\delta \in (0,1)$ называется порядковая статистика $X_{([n\delta]+1)}$; здесь $[x]$ — целая часть числа x.

1.1. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[a, b]$, $a < b$, причём значение параметра a известно. Какие из перечисленных ниже функций являются статистиками?

а) $2\overline{X}$; г) $\overline{X}$; ж) 199;

б) $X_{(n)} - a/n$; д) $X_1/(b-a)$; з) $X_1 + X_3 + 1$;

в) $(a+b)/2$; е) $\sum\limits_{i=1}^{n} X_i$; и) $X_{(1)}$.

Решение. а), б) Функции являются статистиками, поскольку зависят лишь от элементов выборки; в) не является статистикой, поскольку зависит от неизвестного параметра b.

1.2. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром $\lambda > 0$. Какие из перечисленных ниже функций являются статистиками?

а) $\dfrac{\lambda^{\Sigma X_i}}{\prod X_i!} e^{-n\lambda}$; г) $X_1 - \lambda$; ж) $\prod\limits_{i=1}^{n} X_i^2$;

б) 201; д) $\sum\limits_{i=1}^{n} (X_i - \lambda)^2$; з) $\lambda^2 + \lambda$;

в) $\overline{X}$; е) $\sum\limits_{i=1}^{n} X_i$; и) $X_{(n)}$.

1.3. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами a и σ^2.

а) Вычислить среднее значение и дисперсию статистики $\overline{X}$. Какое распределение имеет $\overline{X}$?

б) Вычислить среднее значение выборочной медианы.

в) Вычислить среднее значение статистик S^2 и S_0^2.

1.4. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Вычислить среднее значение и дисперсию статистики $\overline{X}$. Имеет ли статистика $\overline{X}$ распределение Пуассона? Нормальное распределение?

1.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[a,b]$. Вычислить среднее значение и дисперсию статистики $\overline{X}$. Имеет ли статистика $\overline{X}$ равномерное распределение? Нормальное распределение?

1.6. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром 3. Найти распределение выборки $Y_1, \ldots, Y_n$, где $Y_i = 1 - e^{-3X_i}$.

Решение. Значение функции распределения случайной величины Y_1 в точке $y \in [0,1)$ равно

$$\begin{aligned}\mathbf{P}\{Y_1 < y\} &= \mathbf{P}\{1 - e^{-3X_1} < y\} \\ &= \mathbf{P}\Big\{X_1 < -\frac{\ln(1-y)}{3}\Big\} = 1 - e^{\ln(1-y)} = y.\end{aligned}$$

Следовательно, $Y_1, \ldots, Y_n$— выборка из равномерного распределения на отрезке $[0,1]$.

1.7. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью

$$f(y) = \begin{cases} 2y & \text{при } y \in [0,1], \\ 0 & \text{при } y \notin [0,1]. \end{cases}$$

Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = X_i^2$?

1.8. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью

$$f(y) = \begin{cases} 2/y^3 & \text{при } y \geqslant 1, \\ 0 & \text{при } y < 1. \end{cases}$$

Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = 1 - 1/X_i^2$?

1.9. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,1]$. Найти распределение выборки $Y_1, \ldots, Y_n$, где $Y_i = -\ln X_i$.

1.10. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения F, у которого функция распределения $F(y)$ непрерывна и строго возрастает. Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = F(X_i)$?

1.11. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения F с непрерывной функцией распределения $F(y)$. Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = F(X_i)$?

1.12. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = F(X_i)$, а $F(y)$ — функция распределения Бернулли?

1.13. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Какое распределение имеет выборка $Y_1, \ldots, Y_n$, где $Y_i = F(X_i)$, а $F(y)$ — функция распределения Пуассона?

1.14. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения F с плотностью f. Найти совместную плотность всех порядковых статистик $(X_{(1)}, \ldots, X_{(n)})$.

1.15. В терминах общей функции распределения элементов выборки найти функцию распределения

а) максимального члена вариационного ряда $X_{(n)}$;

б) минимального члена вариационного ряда $X_{(1)}$.

Решение. а) Поскольку событие $\{X_{(n)} < y\}$ совпадает с событием $\{X_1 < y, \ldots, X_n < y\}$ и случайные величины $X_1, \ldots, X_n$ независимы, имеем равенства

$$\begin{aligned} \mathbf{P}\{X_{(n)} < y\} &= \mathbf{P}\{X_1 < y, \ldots, X_n < y\} \\ &= \mathbf{P}\{X_1 < y\} \cdot \ldots \cdot \mathbf{P}\{X_n < y\} = F^n(y). \end{aligned}$$

1.16. Найти вероятность $\mathbf{P}\{X_{(k)} < y,\ X_{(k+1)} \geqslant y\}$ в терминах общей функции распределения элементов выборки.

1.17. Найти функцию распределения k-й порядковой статистики $X_{(k)}$ в терминах общей функции распределения $F(y)$.

1.18. Для выборки из равномерного распределения на отрезке $[0, \theta]$ найти плотность распределения

а) минимального члена вариационного ряда $X_{(1)}$;

б) максимального члена вариационного ряда $X_{(n)}$;

в) k-й порядковой статистики $X_{(k)}$.

Решение. в) Воспользуемся ответом к задаче 1.17 и вычислим плотность как производную функции распределения величины $X_{(k)}$:

$$\begin{aligned} f(y) &= \frac{d}{dy}\sum_{i=k}^{n} C_n^i y^i (\theta - y)^{n-i}/\theta^n \\ &= \frac{1}{\theta^n}\left(\sum_{i=k}^{n} iC_n^i y^{i-1}(\theta-y)^{n-i} - \sum_{i=k}^{n-1}(n-i)C_n^i y^i(\theta-y)^{n-i-1}\right) \\ &= nC_{n-1}^{k-1} y^{k-1}(\theta-y)^{n-k}/\theta^n; \end{aligned}$$

здесь использованы равенства $iC_n^i = nC_{n-1}^{i-1}$ и $(n-i)C_n^i = nC_{n-1}^i$.

Вычисление плотности можно провести и непосредственно:

$$f(y)\,dy = \mathbf{P}_\theta\{X_{(k)} \in (y,\, y+dy)\}.$$

Событие $\{X_{(k)} \in (y,\, y+dy)\}$ означает, что из n элементов выборки один принимает значения из множества dy, $k-1$ элемент — левее y и $n-k$ элементов — правее y. Вероятность этого события вычислим в соответствии с полиномиальным распределением:

$$\begin{aligned} \mathbf{P}_\theta\{X_{(k)} \in (y,\, y+dy)\} &= \frac{n!}{(k-1)!\,1!\,(n-k)!}\left(\frac{y}{\theta}\right)^{k-1}\frac{dy}{\theta}\left(\frac{\theta-y}{\theta}\right)^{n-k} \\ &= \left(nC_{n-1}^{k-1}y^{k-1}(\theta-y)^{n-k}/\theta^n\right)dy. \end{aligned}$$

Следовательно, плотность распределения случайной величины $X_{(k)}$ равна $nC_{n-1}^{k-1}y^{k-1}(\theta-y)^{n-k}/\theta^n$. Полезно заметить, что величина $X_{(k)}/\theta$ имеет *бета*-распределение с параметрами k и $n-k+1$.

1.19. Для выборки из распределения F с плотностью f найти плотность распределения

а) минимального члена вариационного ряда $X_{(1)}$;

б) максимального члена вариационного ряда $X_{(n)}$;

в) k-й порядковой статистики $X_{(k)}$.

1.20. Для выборки из равномерного распределения на отрезке $[0,\theta]$ найти среднее значение, второй момент и дисперсию

а) минимального члена вариационного ряда $X_{(1)}$;

б) максимального члена вариационного ряда $X_{(n)}$;

в) k-й порядковой статистики $X_{(k)}$.

1.21. Пусть $X_1, \ldots, X_n$ — выборка из дискретного распределения с вероятностями $\mathbf{P}\{X_1 = m\} = p_m$, где $\sum_{m=0}^{N} p_m = 1$. Найти распределение k-й порядковой статистики $X_{(k)}$.

1.22. Найти совместную функцию распределения минимального и максимального членов вариационного ряда для выборки из некоторого распределения F.

1.23. Для выборки из равномерного распределения на отрезке $[0, \theta]$ найти

а) совместную плотность минимального $X_{(1)}$ и максимального $X_{(n)}$ членов вариационного ряда;

б) ковариацию минимального $X_{(1)}$ и максимального $X_{(n)}$ членов вариационного ряда;

в) совместную плотность $X_{(k)}$ и $X_{(j)}$, $1 \leqslant k < j \leqslant n$;

г) ковариацию $X_{(k)}$ и $X_{(j)}$, $1 \leqslant k \leqslant j \leqslant n$.

1.24. Пусть дана выборка из показательного распределения с параметром α.

а) Доказать, что случайные величины $X_{(1)}$, $X_{(2)} - X_{(1)}$, ..., $X_{(n)} - X_{(n-1)}$ независимы.

б) Каково распределение минимального члена вариационного ряда $X_{(1)}$?

в) Каково распределение разности соседних порядковых статистик $X_{(k+1)}$ и $X_{(k)}$?

г) Доказать справедливость равенства

$$\mathbf{E}X_{(k)} = a^{-1}((n-k+1)^{-1} + \cdots + n^{-1}).$$

1.25. Найти ошибку в следующем рассуждении: «Пусть дана выборка $X_1, \ldots, X_n$ из некоторого распределения. Поскольку при каждом элементарном исходе случайная величина $X_{(n)}$ совпадает с одним из элементов выборки, то $X_{(n)}$ имеет такое же распределение, как и X_1».

1.26. Пусть дана выборка из распределения F такого, что

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

Доказать, что $X_{(1)}/n \to 0$ и $X_{(n)}/n \to 0$ по вероятности.

1.27. Пусть дана выборка из распределения с конечным средним значением. Доказать, что $X_{(1)}/n \to 0$ и $X_{(n)}/n \to 0$ почти наверное.

1.28. Для выборки из равномерного распределения на отрезке $[0, \theta]$ найти предельное при $n \to \infty$ распределение случайной величины

а) $nX_{(1)}/\theta$; б) $n(\theta - X_{(n)})/\theta$.

1.29. Пусть дана выборка из равномерного распределения на отрезке $[0, \theta]$. Зафиксируем натуральное число k. Доказать слабую сходимость при $n \to \infty$ распределения случайной величины

а) $nX_{(k)}/\theta$; б) $n(\theta - X_{(n-k+1)})/\theta$

к Г-распределению с параметрами 1 и k.

Решение. а) Воспользовавшись решением задачи 1.18в), выпишем плотность $f_n(y)$ распределения величины $nX_{(k)}/\theta$ при $y < n$:

$$f_n(y) = C_{n-1}^{k-1}(y/n)^{k-1}(1 - y/n)^{n-k}.$$

Последнее выражение равно вероятности получить ровно $k-1$ успех в $n-1$ испытании схемы Бернулли с вероятностью успеха $p_{n-1} = y/n$. Воспользовавшись оценкой скорости сходимости в теореме Пуассона, для любого $y \in [0, n]$ получим оценку

$$\left| f_n(y) - \frac{y^{k-1}}{(k-1)!}\, e^{-y} \right| \leqslant (n-1)p_{n-1}^2 \leqslant y^2/n.$$

Следовательно, имеет место равномерная по $y \geqslant 0$ из любого компакта сходимость плотности $f_n(y)$ к плотности Г-распределения

$$\lim_{n\to\infty} f_n(y) = \frac{y^{k-1}}{(k-1)!}\, e^{-y} = \frac{y^{k-1}}{\Gamma(k)}\, e^{-y}.$$

Равномерная сходимость плотностей влечёт слабую сходимость распределения величины $nX_{(k)}/\theta$ к Г-распределению с параметрами 1 и k.

1.30. Пусть дана выборка из распределения с непрерывной функцией распределения $F(y)$. Для любого фиксированного $k \geqslant 1$ найти слабый предел при $n \to \infty$ распределения случайной величины

а) $nF(X_{(k)})$; б) $n(1 - F(X_{(n-k+1)}))$.

1.31. Пусть дана выборка из равномерного распределения на отрезке $[0, \theta]$. Для любых фиксированных $k \geqslant 1$ и $j \geqslant 1$ найти совместное предельное при $n \to \infty$ распределение случайного вектора

$$(nX_{(k)}/\theta,\ n(\theta - X_{(n-j+1)})/\theta).$$

1.32. Пусть дана выборка из равномерного распределения на отрезке $[0, \theta]$. Для любых фиксированных $k \geqslant 1$ и $j \geqslant 1$, $k < j$, найти совместное предельное при $n \to \infty$ распределение случайного вектора

$$(nX_{(k)}/\theta,\ nX_{(j)}/\theta).$$

1.33. Пусть дана выборка из равномерного распределения на отрезке [0,1]. Показать, что если k и n растут таким образом, что $k/n \to p$, то распределение случайной величины $\sqrt{n}\,(X_{(k)} - k/n)$ слабо сходится к нормальному закону с нулевым средним и дисперсией $p(1-p)$.

1.34. Пусть дана выборка из равномерного распределения на отрезке [0,1]. Показать, что если k, j и n растут таким образом, что $k/n \to p$ и $j/n \to s$, $p < s$, то распределение случайного вектора

$$(\sqrt{n}\,(X_{(k)} - k/n),\ \sqrt{n}\,(X_{(j)} - j/n))$$

слабо сходится к двумерному нормальному закону. Найти параметры предельного распределения.

1.35. Пусть дана выборка из показательного распределения с параметром α. Найти слабый предел распределения разности $\alpha X_{(n)} - \ln n$.

§ 2. Эмпирическая функция распределения

Эмпирическим распределением P_n^*, построенным по выборке $X_1, \ldots, X_n$, называется распределение, определяемое для любого борелевского множества $B \subseteq \mathbf{R}$ равенством

$$P_n^*(B) = \frac{\nu_n(B)}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{X_i \in B\},$$

где $\nu_n(B)$ — число элементов выборки, попавших в множество B. Для каждого фиксированного элементарного исхода P_n^* есть распределение на $\mathbf{R}$. Для каждого фиксированного борелевского множества B отображение $P_n^*(B)$: $\Omega \to \mathbf{R}$ есть случайная величина.

Эмпирической функцией распределения $F_n^*(y)$, построенной по выборке $X_1, \ldots, X_n$, называется функция

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{X_i < y\}.$$

В силу определения справедливо равенство $F_n^*(y) = P_n^*((-\infty, y))$. Для каждого фиксированного элементарного исхода функция F_n^* есть функция распределения на $\mathbf{R}$. Для каждого фиксированного числа y отображение $F_n^*(y) : \Omega \to \mathbf{R}$ есть случайная величина.

Справедлива следующая

Теорема Гливенко – Кантелли. *Пусть $F(y)$ — общая функция распределения элементов выборки. Тогда почти наверное при $n \to \infty$ имеет место сходимость*

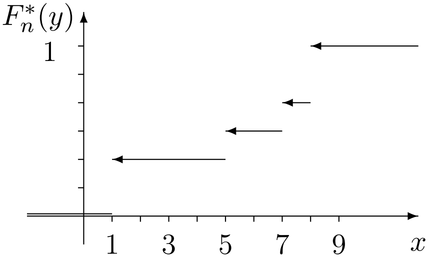
$$\sup_{y \in \mathbf{R}} |F_n^*(y) - F(y)| \to 0.$$

2.1. Пусть $(-0{,}8;\ 2{,}9;\ 4{,}3;\ -5{,}7;\ 1{,}1;\ -3{,}2)$ — наблюдавшиеся значения выборки. Построить эмпирическую функцию распределения и проверить, что $F_6^*(-5) = 1/6$, $F_6^*(0) = 1/2$ и $F_6^*(4) = 5/6$.

2.2. Пусть $(3, 0, 4, 3, 6, 0, 3, 1)$ — наблюдавшиеся значения выборки. Построить эмпирическую функцию распределения и проверить, что $F_8^*(1) = 1/4$, $F_8^*(3) = 3/8$ и $F_8^*(5) = 7/8$.

2.3. По выборке объёма n из распределения Бернулли с параметром p построить график эмпирической функции распределения $F_n^*(y)$.

2.4. Найти по крайней мере 2 выборки различных объёмов, которым соответствует следующая эмпирическая функция распределения:



Р е ш е н и е. Можно взять следующие выборки: (1, 1, 5, 7, 8, 8), или (1, 5, 1, 7, 8, 8), или (1, 1, 1, 1, 8, 8, 8, 8, 7, 7, 5, 5).

2.5. Можно ли по функции $F_n^*(y)$ из задачи 2.4 восстановить исходную выборку, если объём выборки известен? Можно ли восстановить вариационный ряд? А если объём выборки неизвестен?

2.6. Пусть $F_n^*(y)$ — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$ объёма n. Пусть a — положительное вещественное число. Является ли эмпирической функцией распределения функция $F_n^*(ay)$? Если «да», то какой выборке она соответствует?

2.7. Пусть $a > 0$ и b — два фиксированных действительных числа. Пусть F_n^* — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$, а G_n^* — эмпирическая функция распределения, построенная по выборке $Y_1, \ldots, Y_n$, где $Y_i = aX_i + b$. Доказать, что при всех y имеет место равенство

$$G_n^*(y) = F_n^*\Big(\frac{y-b}{a}\Big).$$

2.8. Пусть $F_n^*(y)$ — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$ объёма n. Является ли эмпирической функцией распределения функция

а) $F_n^*(y^3)$; б) $(F_n^*(y))^3$?

Если «да», то какой выборке она соответствует?

2.9. Пусть $F_n^*(y)$ — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$ объёма n, а $G_n^*(y)$ — по выборке $Y_1, \ldots, Y_n$ того же объёма n. Является ли эмпирической функцией распределения функция $(F_n^*(y) + G_n^*(y))/2$? Если «да», то какой выборке она соответствует?

2.10. Пусть F_n^* — эмпирическая функция распределения, построенная по выборке $X_1, \ldots, X_n$, а G_n^* — эмпирическая функция распределения, построенная по выборке $Y_1, \ldots, Y_n$, где $Y_i = G(X_i)$ и G — монотонно возрастающая непрерывная функция. Доказать, что при всех y и v справедливо равенство

$$\mathbf{P}\{F_n^*(y) < v\} = \mathbf{P}\{G_n^*(G(y)) < v\}.$$

2.11. Пусть $X_1, \ldots, X_n$ — выборка из распределения F. Доказать, что для любых $y \in \mathbf{R}$ и $k \in \{0, 1, \ldots, n\}$ справедливо равенство (ср. с задачей 1.16)

$$\mathbf{P}\{F_n^*(y) = k/n\} = C_n^k F^k(y)(1 - F(y))^{n-k}.$$

2.12. Для выборки из распределения F найти

а) $\mathbf{E}F_n^*(y)$; б) $\mathbf{D}F_n^*(y)$; в) $\mathbf{D}(F_n^*(z) - F_n^*(y))$.

2.13. Для выборки из биномиального распределения с параметрами p и m найти (ср. с задачей 1.21)

$$\mathbf{P}\{F_n^*(y+0) - F_n^*(y) = k/n\}.$$

2.14. Чему равна вероятность $\mathbf{P}\{F_n^*(y) < F_n^*(z)\}$?

2.15. Какова вероятность наличия у эмпирической функции распределения хотя бы одного скачка размера $2/n$, если функция распределения выборки непрерывна?

2.16. Доказать, что для выборки с непрерывной функцией распределения $F(y)$ при любом $t \in [0,1]$ справедливо равенство

$$\mathbf{P}\Big\{\sup_y |F_n^*(y) - F(y)| > t\Big\} = \mathbf{P}\Big\{\sup_{0 \leqslant y \leqslant 1} |G_n^*(y) - y| > t\Big\},$$

где $G_n^*(y)$ — эмпирическая функция распределения, построенная по выборке из равномерного распределения на отрезке $[0,1]$ (это свойство используется при построении критерия Колмогорова и называется непараметричностью этого критерия).

2.17. Доказать, что для выборки из общего распределения F при любом $t \in [0,1]$ справедливо неравенство

$$\mathbf{P}\Big\{\sup_y |F_n^*(y) - F(y)| > t\Big\} \leqslant \mathbf{P}\Big\{\sup_{0 \leqslant y \leqslant 1} |G_n^*(y) - y| > t\Big\},$$

где $G_n^*(y)$ — эмпирическая функция распределения, построенная по выборке из равномерного распределения на отрезке $[0,1]$.

2.18. Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ — две независимые выборки одинакового объёма n из одного и того же непрерывного распределения, а F_n^* и G_n^* — эмпирические функции распределения, построенные по этим выборкам. Доказать, что для любого $t \in (0,1]$ справедливо равенство

$$\mathbf{P}\Big\{\sup_{y \in \mathbf{R}} |F_n^*(y) - G_n^*(y)| < t\Big\} = \mathbf{P}\Big\{\sup_{1 \leqslant k \leqslant 2n} |S_k| < tn \Big| S_{2n} = 0\Big\},$$

где $S_k = \xi_1 + \cdots + \xi_k$, причём случайные слагаемые ξ_i независимы и $\mathbf{P}\{\xi_i = 1\} = \mathbf{P}\{\xi_i = -1\} = 1/2$.

2.19. Пусть $X_1, \ldots, X_n$ — выборка из распределения на множестве целых чисел; положим $p_k = \mathbf{P}\{X_1 = k\}$. Обозначим через $\nu_k(n)$ число элементов выборки, равных k. Доказать, что

$$\sup_{A \subseteq \mathbf{Z}} |P_n^*(A) - \mathbf{P}\{X_1 \in A\}| = \frac{1}{2} \sum_{k=-\infty}^{\infty} \left| \frac{\nu_k(n)}{n} - p_k \right|.$$

2.20. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром $1/4$, $F_n^*(y)$ — эмпирическая функция распределения и $F(y)$ — функция распределения выборки. Доказать, что при $n \to \infty$ с вероятностью 1 имеет место сходимость

$$\sup_{y \in \mathbf{R}} |F_n^*(y) - F(y)| \to 0.$$

2.21. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром 1. При каждом $n \geqslant 1$ через ν_n обозначим количество элементов выборки, не превышающих 2. Указать по крайней мере две различные последовательности чисел c_n такие, что последовательность $(\nu_n - c_n)/\sqrt{n}$ слабо сходится при $n \to \infty$ к некоторому нормальному распределению. Найти параметры этого распределения.

2.22. Доказать, что для любого фиксированного $\lambda \in \mathbf{R}$ значение выборочной характеристической функции

$$\varphi_n^*(\lambda) = \int\limits_{\mathbf{R}} e^{i\lambda y} F_n^*(dy)$$

сходится при $n \to \infty$ почти наверное к значению истинной характеристической функции $\varphi(\lambda) = \mathbf{E} e^{i\lambda X_1}$.

2.23. Доказать, что для любого компакта $K \subset \mathbf{R}$ при $n \to \infty$ имеет место равномерная по $\lambda \in K$ сходимость почти наверное

$$\sup_{\lambda \in K} |\varphi_n^*(\lambda) - \varphi(\lambda)| \to 0.$$

2.24. Пусть распределение F сосредоточено на решётке целых чисел. Доказать, что при $n \to \infty$ имеет место равномерная по $\lambda \in \mathbf{R}$ сходимость почти наверное

$$\sup_{\lambda \in \mathbf{R}} |\varphi_n^*(\lambda) - \varphi(\lambda)| \to 0.$$

ОТДЕЛ II
МЕТОДЫ ПОСТРОЕНИЯ ОЦЕНОК

§ 3. Метод моментов

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений, причём Θ есть подмножество d-мерного евклидова пространства $\mathbf{R}^d$. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ.

В случае одномерного параметра θ ($d = 1$) оценка этого параметра по методу моментов строится следующим образом. Выбирается пробная функция $g : \mathbf{R} \to \mathbf{R}$ такая, что функция

$$m(\theta) = \mathbf{E}_\theta g(X_1) = \int\limits_{\mathbf{R}} g(y) F_\theta(dy)$$

является непрерывной и монотонной. *Оценкой по методу моментов* называется оценка $\theta_n^* \in \Theta$ такая, что

$$m(\theta_n^*) = \overline{g(X)} = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

Ясно, что оценка, построенная по методу моментов, не единственна и зависит от выбора пробной функции g. В качестве g чаще всего выбирают функции вида $g(y) = y^k$; в этом случае $m(\theta)$ есть k-й момент распределения выборки, что и дало название методу.

В случае многомерного параметра θ ($d \geqslant 2$) для построения оценки этого параметра по методу моментов выбираются d функций $g_j : \mathbf{R} \to \mathbf{R}$, $j = 1, \ldots, d$, и рассматриваются функции

$$m_j(\theta) = \mathbf{E}_\theta g_j(X_1).$$

Пробные функции g_j выбираются таким образом, чтобы система уравнений

$$m_j(\theta) = z_j, \quad j = 1, \ldots, d,$$

была однозначно и непрерывно разрешима относительно d-мерного параметра θ. *Оценкой по методу моментов* называется оценка $\theta_n^* \in \Theta$ такая, что

$$m_j(\theta_n^*) = \overline{g_j(X)}, \quad j = 1, \ldots, d.$$

В качестве пробных функций g_j чаще всего выбираются степенные функции.

3.1. Пусть дана выборка из нормального распределения с параметрами a и σ^2. Используя метод моментов, построить оценку

а) неизвестного среднего значения a;

б) неизвестной дисперсии σ^2, если среднее значение a известно;

в) двумерного параметра (a, σ^2).

Решение. а) Возьмём пробную функцию $g(y) = y$. Имеем равенства

$$m(a) = \mathbf{E}_{a,\sigma^2} g(X_1) = \mathbf{E}_{a,\sigma^2} X_1 = a.$$

Поэтому $m^{-1}(y) = y$ и искомая оценка a_n^* метода моментов равна $\overline{X}$.

б) Для пробной функции $g(y) = y^2$ справедливы равенства

$$m(\sigma^2) = \mathbf{E}_{a,\sigma^2} g(X_1) = \mathbf{E}_{a,\sigma^2} X_1^2 = \sigma^2 + a^2.$$

Поэтому $m^{-1}(y) = y - a^2$ и искомая оценка $(\sigma^2)_n^*$ метода моментов равна $\overline{X^2} - a^2$.

в) Используя пробные функции $g_1(y) = y$ и $g_2(y) = y^2$, получаем

$$m_1(a, \sigma^2) = \mathbf{E}_{a,\sigma^2} g_1(X_1) = \mathbf{E}_{a,\sigma^2} X_1 = a,$$
$$m_2(a, \sigma^2) = \mathbf{E}_{a,\sigma^2} g_2(X_1) = \mathbf{E}_{a,\sigma^2} X_1^2 = a^2 + \sigma^2.$$

Решая систему

$$\begin{cases} a_n^* = \overline{X} \\ (a_n^*)^2 + (\sigma^2)_n^* = \overline{X^2}, \end{cases}$$

находим искомые оценки метода моментов: $a_n^* = \overline{X}$ и $(\sigma^2)_n^* = \overline{X^2} - (\overline{X})^2 \equiv S^2$.

3.2. Используя метод моментов с пробной функцией

а) $g(y) = |y - a|$; б) $g(y) = (y - a)^2$,

оценить неизвестную дисперсию $\sigma^2 > 0$ нормального распределения с известным средним значением a.

3.3. Используя пробные функции $g_1(y) = y$ и $g_2(y) = y^2$, оценить неизвестный параметр $\theta > 0$ нормального распределения со средним θ и дисперсией

а) 2θ; б) θ^2.

3.4. Используя пробные функции y^{2k}, $k = 1, 2, \ldots$, оценить неизвестную дисперсию σ^2 нормального распределения с нулевым средним.

3.5. Используя метод моментов, оценить параметр θ равномерного распределения на отрезке

а) $[0, \theta]$, $\theta > 0$; в) $[0, 2\theta]$, $\theta > 0$;

б) $[\theta - 1, \theta + 1]$, $\theta \in \mathbf{R}$; г) $[-\theta, \theta]$, $\theta > 0$.

3.6. Используя пробные функции $g_1(y) = y$ и $g_2(y) = y^2$, построить оценку векторного параметра (a, b) равномерного распределения на отрезке

а) $[a, b]$, $a < b$; б) $[a, a+b]$, $b > 0$.

3.7. Используя пробные функции $g(y) = y^k$, $k = 1, 2, \ldots$, оценить параметр $\theta > 0$ равномерного распределения на отрезке $[0, \theta]$.

3.8. Используя метод моментов с пробной функцией $g(y) = y$, оценить параметр $\alpha > 0$ показательного распределения.

3.9. Используя метод моментов с пробной функцией $g(y) = y$, оценить параметр сдвига $\beta \in \mathbf{R}$ показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

3.10. Пусть дана выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\ e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Используя метод моментов, оценить параметры масштаба $\alpha > 0$ и сдвига $\beta \in \mathbf{R}$.

3.11. Используя пробные функции $g(y) = y^k$, $k \in \mathbf{N}$, по выборке из показательного распределения с параметром

а) α; б) $1/\alpha$

оценить неизвестное значение $\alpha > 0$.

3.12. Используя метод моментов с подходящей пробной функцией $g(y)$, оценить параметр $\alpha > 0$ распределения Лапласа.

3.13. Используя метод моментов, оценить значение α по выборке из показательного распределения с параметром $1/\sqrt{\alpha}$.

3.14. Пусть дана выборка из показательного распределения с параметром α. Используя метод моментов с пробной функцией $g(y) = y$, оценить параметр $\theta(\alpha) = \mathbf{P}_\alpha\{X_1 \geqslant 1\}$.

3.15. Пусть дана выборка из Г-распределения с параметрами $\alpha > 0$ и $\beta > 0$. Построить оценки по методу моментов

а) параметра α, если значение β известно;

б) параметра β, если значение α известно;

в) векторного параметра (α, β).

3.16. Пусть имеется выборка из распределения Парето с параметрами β и θ. Построить оценки по методу моментов

а) параметра $\beta > 1$, если значение $\theta > 0$ известно;

б) параметра $\theta > 0$, если значение $\beta > 1$ известно;

в) векторного параметра (β, θ), где $\beta > 2$ и $\theta > 0$.

3.17. Пусть дана выборка из распределения Вейбулла с параметрами $\alpha > 1$ и $\theta > 0$, причём значение α известно. С помощью пробной функции $g(y) = y^\alpha$ построить оценку параметра θ.

3.18. Пусть дана выборка из распределения Вейбулла с параметрами $\alpha > 0$ и 1. Показать, что с помощью пробной функции $g(y) = y$ нельзя построить оценку параметра α.

3.19. Пусть дана выборка из распределения с плотностью

$$f_\alpha(y) = \begin{cases} 3y^2\alpha^{-3}\ e^{-(y/\alpha)^3} & \text{при } y \geqslant 0, \\ 0 & \text{при } y < 0. \end{cases}$$

Построить оценку параметра $\alpha > 0$ с помощью пробной функции $g(y) = y^k$.

3.20. Используя пробную функцию $g(y) = y$, построить оценку параметра $\theta > 0$, если распределение выборки имеет плотность

а) $\theta y^{\theta-1}$ при $y \in [0, 1]$; б) $2y/\theta^2$ при $y \in [0, \theta]$.

3.21. Можно ли построить оценку по методу моментов с помощью одной из пробных функций y, y^2, y^3, $\ldots$ для параметра сдвига распределения Коши?

3.22. Используя метод моментов с пробной функцией $g(y) = y$, оценить параметр p распределения Бернулли.

3.23. Можно ли методом моментов с помощью какой-нибудь пробной функции $g(y)$ получить оценку параметра p распределения Бернулли, отличную от $\overline{X}$?

3.24. Пусть дана выборка из биномиального распределения с параметрами m и p. Используя метод моментов, построить оценку

а) параметра p, если значение m известно;

б) параметра m, если значение p известно;

в) векторного параметра (m, p).

3.25. Пусть дана выборка из биномиального распределения с параметрами 2 и p. Используя метод моментов с пробной функцией $g(y) = y$, найти оценку для параметра $\theta = e^{2p}$.

3.26. Используя метод моментов с пробной функцией

а) $g(y) = y$; б) $g(y) = y^2$,

оценить параметр $\lambda > 0$ распределения Пуассона.

3.27. Используя метод моментов, оценить значение $\lambda > 1$ по выборке из распределения Пуассона с параметром $\ln \lambda$.

3.28. Пусть дана выборка из распределения Пуассона с параметром λ. Используя метод моментов с пробной функцией $g(y) = \mathbf{I}\{y = 1\}$, оценить параметр $\theta = \theta(\lambda) = \lambda\, e^{-\lambda}$.

3.29. При нейтронной бомбардировке ядер урана начинается расщепление ядра, при котором ядро урана распадается на две части различного типа; в камере Вильсона это явление обнаруживается в виде двух траекторий, исходящих из одной точки. Эти траектории вскоре разделяются на несколько ветвей, получающихся от столкновения частиц с молекулами газа в камере. Можно показать, что число ветвей в одной траектории имеет так называемое «двойное» распределение Пуассона, т. е.

$$\mathbf{P}\{X_1 = k\} = \frac{1}{2}\Big(\frac{\lambda_1^k}{k!}e^{-\lambda_1} + \frac{\lambda_2^k}{k!}e^{-\lambda_2}\Big), \quad k = 0, 1, 2, \ldots,$$

где $\lambda_1 < \lambda_2$ — некоторые положительные постоянные. Используя метод моментов, построить оценку векторного параметра (λ_1, λ_2).

3.30. Используя метод моментов, оценить параметр $p \in (0, 1)$ геометрического распределения.

3.31. Пусть P и Q — два распределения с известными математическими ожиданиями a и b соответственно, причём $a < b$. Пусть P_θ является смесью распределений P и Q:

$$P_\theta = \theta P + (1 - \theta) Q, \quad 0 \leqslant \theta \leqslant 1.$$

Используя метод моментов, оценить параметр θ по выборке из распределения P_θ.

3.32. Привести пример, когда нельзя построить оценку по методу моментов с помощью пробной функции $g(y) = y$.

§ 4. Метод максимального правдоподобия

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть выполнено условие доминирования относительно некоторой меры μ на $\mathbf{R}$, т. е. это параметрическое семейство состоит из распределений, абсолютно непрерывных относительно μ. Плотность распределения F_θ относительно меры μ обозначим через f_θ:

$$f_\theta(y) = \frac{dF_\theta}{d\mu}(y).$$

Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ. *Функцией правдоподобия* называется функция

$$f_\theta(X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i).$$

Оценка $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ параметра θ называется *оценкой максимального правдоподобия*, если в точке θ_n^* достигается максимум функции правдоподобия (при фиксированном значении выборки $X_1, \ldots, X_n$).

При нахождении оценки максимального правдоподобия часто удобно перейти к *логарифмической функции правдоподобия*

$$L_\theta(X_1, \ldots, X_n) = \ln f_\theta(X_1, \ldots, X_n) = \sum_{i=1}^{n} \ln f_\theta(X_i).$$

Ясно, что максимум функции L также достигается в точке θ_n^*.

4.1. Найти оценку максимального правдоподобия векторного параметра (a, σ^2) нормального распределения.

Р е ш е н и е. Логарифмическая функция правдоподобия равна

$$L_{a,\sigma^2}(X_1, \ldots, X_n) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - a)^2.$$

Точку, в которой достигается максимум (проверьте!) функции L, находим из следующей системы уравнений:

$$\frac{\partial L_{a,\sigma^2}(X_1, \ldots, X_n)}{\partial a} = 0, \qquad \frac{\partial L_{a,\sigma^2}(X_1, \ldots, X_n)}{\partial (\sigma^2)} = 0,$$

т. е.

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - a) = 0, \qquad -\frac{n}{\sigma^2} + \frac{1}{\sigma^4} \sum_{i=1}^{n} (X_i - a)^2 = 0.$$

Решение этой системы есть $a_n^* = \overline{X}$ и $(\sigma^2)_n^* = \overline{X^2} - (\overline{X})^2$.

4.2. Найти оценку максимального правдоподобия дисперсии σ^2 нормального распределения, если среднее значение a известно.

4.3. Найти оценку максимального правдоподобия параметра $\theta > 0$, если распределение выборки имеет нормальную плотность со средним θ и дисперсией

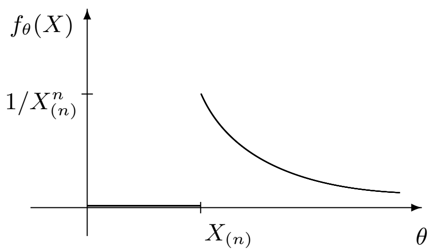
а) 2θ; б) θ^2.

4.4. Найти оценку максимального правдоподобия параметра $\theta > 0$ равномерного распределения на отрезке $[0, \theta]$.

Решение. Функция правдоподобия выборки равна

$$f_\theta(X_1, \ldots, X_n) = \begin{cases} \theta^{-n}, & \text{если все } X_j \in [0, \theta], \\ 0, & \text{если хотя бы одно } X_j \notin [0, \theta] \end{cases} = \begin{cases} \theta^{-n}, & \text{если } X_{(n)} \leqslant \theta, \\ 0, & \text{если } X_{(n)} > \theta. \end{cases}$$

При *фиксированных* значениях выборки (и, следовательно, при фиксированном значении $X_{(n)}$) зависимость $f_\theta(X_1, \ldots, X_n)$ от переменной θ имеет вид



Максимум функции правдоподобия достигается в точке $\theta = X_{(n)}$. Поэтому искомая оценка максимального правдоподобия есть $\theta_n^* = X_{(n)}$.

4.5. Найти оценку максимального правдоподобия параметра θ равномерного распределения на отрезке

а) $[-\theta, 0]$, $\theta > 0$; в) $[\theta, \theta + 2]$, $\theta \in \mathbf{R}$;

б) $[-\theta, \theta]$, $\theta > 0$; г) $[\theta, 2\theta]$, $\theta > 0$.

4.6. Найти оценку максимального правдоподобия двумерного параметра (a, b) для равномерного распределения на отрезке $[a, b]$.

4.7. Найти оценку максимального правдоподобия параметра α показательного распределения.

4.8. Найти оценку максимального правдоподобия параметра сдвига $\beta \in \mathbf{R}$ показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta-y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

4.9. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$. Найти оценку максимального правдоподобия двумерного параметра (α, β).

4.10. Найти оценку максимального правдоподобия параметра сдвига $\mu \in \mathbf{R}$ распределения Лапласа с плотностью

$$f_\mu(y) = e^{-|y-\mu|}/2.$$

4.11. Пусть дана выборка из двухпараметрического распределения Лапласа с плотностью

$$f_{\mu,\sigma}(y) = e^{-|y-\mu|/\sigma}/2\sigma,$$

где $\mu \in \mathbf{R}$ и $\sigma > 0$. Найти оценку максимального правдоподобия для двумерного параметра (μ, σ).

4.12. Построить оценку максимального правдоподобия параметра $\alpha > 0$ для Г-распределения, если значение β известно.

4.13. Пусть дана выборка из распределения Парето с параметрами $\beta > 0$ и $\theta > 0$. Найти оценку максимального правдоподобия

а) параметра β, если значение θ известно;

б) параметра θ, если значение β известно;

в) векторного параметра (β, θ).

4.14. Пусть дана выборка из распределения Вейбулла с параметрами α и θ. Построить оценку максимального правдоподобия параметра $\theta > 0$, если значение $\alpha > 1$ известно.

4.15. Найти оценку максимального правдоподобия параметра $\alpha > 0$ распределения с плотностью

$$f_\alpha(y) = \begin{cases} 3y^2\alpha^{-3}\, e^{-(y/\alpha)^3} & \text{при } y \geqslant 0, \\ 0 & \text{при } y < 0. \end{cases}$$

4.16. Распределение Кэптейна определяется плотностью

$$f_\theta(y) = \frac{g'(y)}{\sqrt{2\pi}}\, e^{-(\theta - g(y))^2/2},$$

где $g(y)$ — неубывающая дифференцируемая функция. Найти оценку максимального правдоподобия параметра θ.

4.17. Найти оценки максимального правдоподобия параметра $\theta > 0$, если распределение выборки имеет плотность

а) $\theta y^{\theta-1}$ при $y \in [0, 1]$;

б) $2y/\theta^2$ при $y \in [0, \theta]$;

в) $\theta e^{-\theta^2/2y}/\sqrt{2\pi y^3}$ при $y \geqslant 0$;

г) $\theta\left(\ln^{\theta-1} y\right)/y$ при $y \in [1,\, e]$;

д) $e^{-|y|}/2(1 - e^{-\theta})$ при $|y| \leqslant \theta$.

4.18. Пусть дана выборка из распределения Коши с параметром сдвига θ. Построить оценку максимального правдоподобия параметра θ по выборке

а) объёма 1; б) объёма 2.

4.19. На плоской фольге в неизвестной точке находится источник радиоактивного излучения, посылающий лучи равномерно по всем направлениям пространства. Параллельно фольге на расстоянии 1 имеется экран, на котором наблюдаются вспышки, вызываемые радиоактивным излучением. Требуется по местам вспышек на экране определить положение источника излучения на фольге. *Указание*: плоскость экрана принять за координатную плоскость xoy, ось oz направить к фольге и рассмотреть оценку какой-нибудь одной координаты точки; используя метод максимального правдоподобия, выписать уравнение правдоподобия и рассмотреть случай $n = 2$.

4.20. Построить оценку максимального правдоподобия параметра p распределения Бернулли.

Решение. Плотность $f_p(y)$ распределения Бернулли относительно считающей (на множестве $\{0,\, 1\}$) меры равна

$$f_p(y) = p^y (1-p)^{1-y}.$$

Поэтому логарифмическая функция правдоподобия равна

$$L_p(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i \ln p + \Big(n - \sum_{i=1}^{n} X_i\Big) \ln(1-p).$$

Точка, в которой достигается максимум функции L, находится из уравнения

$$\frac{\partial L_p(X_1, \ldots, X_n)}{\partial p} = \frac{1}{p} \sum_{i=1}^{n} X_i - \frac{1}{1-p}\Big(n - \sum_{i=1}^{n} X_i\Big) = 0.$$

Решение этой системы есть $p_n^* = \overline{X}$.

4.21. Пусть дана выборка из биномиального распределения с параметрами $p \in (0,1)$ и m. Найти оценку максимального правдоподобия параметра

а) p, если значение параметра m известно;

б) m по выборке объёма $n = 1$, если значение p известно.

4.22. Найти оценку максимального правдоподобия параметра $\lambda > 0$ распределения Пуассона.

4.23. Найти оценку максимального правдоподобия параметра $p \in (0,1)$ геометрического распределения.

4.24. Пусть $X_1, \ldots, X_n$ — выборка из усечённого на заданном уровне m геометрического распределения с параметром $p \in (0,1)$:

$$\mathbf{P}\{X_1 = k\} = p(1-p)^k, \quad k = 0, \ldots, m-1,$$
$$\mathbf{P}\{X_1 = m\} = 1 - \mathbf{P}\{X_1 \leqslant m-1\} = (1-p)^m.$$

Найти оценку максимального правдоподобия для p.

4.25. Найти оценку максимального правдоподобия параметра θ равномерного распределения на конечном множестве $\{1, \ldots, \theta\}$, θ — целый положительный параметр.

4.26. Пусть дана выборка из нормального распределения со средним a и единичной дисперсией, где a может принимать лишь два значения: 1 и 2. Найти оценку максимального правдоподобия параметра a.

Решение. Поскольку множество $\Theta = \{1,\ 2\}$ двухточечное, то оценка максимального правдоподобия $a_n^*(X_1, \ldots, X_n)$ принимает значение 1, если $f_1(X_1, \ldots, X_n) > f_2(X_1, \ldots, X_n)$, что эквивалентно неравенству

$$\frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum (X_i - 1)^2} > \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum (X_i - 2)^2}.$$

Решив последнее неравенство, получим

$$a_n^* = \begin{cases} 1, & \text{если } \overline{X} < 3/2, \\ 2, & \text{если } \overline{X} \geqslant 3/2. \end{cases}$$

4.27. Пусть дана выборка из показательного распределения с параметром α, где α может принимать лишь значения 1, 2 и 3. Построить оценку максимального правдоподобия параметра α.

4.28. Пусть $X_1, \ldots, X_n$ — выборка из следующего трёхточечного распределения, зависящего от параметра $\theta \in (0, 1/3)$:

$$\mathbf{P}_\theta\{X_1 = 1\} = \theta, \quad \mathbf{P}_\theta\{X_1 = 2\} = 2\theta, \quad \mathbf{P}_\theta\{X_1 = 3\} = 1 - 3\theta.$$

Найти оценку максимального правдоподобия параметра θ.

4.29. Пусть распределение выборки имеет плотность $f_\theta(y) = f(y-\theta)$, где функция $f(y)$ имеет единственный максимум в точке $y = 0$. Построить оценку максимального правдоподобия θ_1^* параметра сдвига θ по одному наблюдению X_1.

4.30. Пусть в условиях предыдущей задачи функция $f(y)$ убывает с ростом $|y|$. Доказать, что оценка максимального правдоподобия θ_n^*, построенная по выборке объёма n, лежит в интервале $[X_{(1)}, X_{(n)}]$.

4.31. Привести пример параметрического семейства распределений, для которого оценка максимального правдоподобия не единственна.

4.32. Привести пример, когда оценка максимального правдоподобия не совпадает с оценкой по методу моментов, полученной с помощью функции $g(y) = y$.

§ 5. Байесовские оценки

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть выполнено условие доминирования относительно некоторой меры μ на $\mathbf{R}$, т. е. это параметрическое семейство состоит из распределений, абсолютно непрерывных относительно μ. Обозначим через f_θ плотность распределения F_θ относительно меры μ.

Пусть параметр θ является случайной величиной с плотностью $q(t)$ относительно некоторой меры λ. Функция

$$f(t, x_1, \ldots, x_n) = f_t(x_1, \ldots, x_n) q(t)$$

является плотностью некоторого распределения в $\mathbf{R}^n \times \Theta$ относительно меры $\mu^n \times \lambda$. *Байесовской оценкой* параметра θ, построенной по выборке $X_1, \ldots, X_n$, называется

$$\theta_n^* = \int\limits_{\Theta} tq(t|X_1, \ldots, X_n)\lambda(dt),$$

где *апостериорная плотность* $q(t|x_1, \ldots, x_n)$ параметра θ вычисляется по формуле

$$q(t|x_1, \ldots, x_n) = \frac{f_t(x_1 \ldots, x_n)q(t)}{\int\limits_{\Theta} f_s(x_1, \ldots, x_n)q(s)\lambda(ds)}.$$

5.1. Пусть дана выборка из нормального распределения со средним a и единичной дисперсией, причём параметр a имеет нормальное распределение с нулевым средним и известной дисперсией σ^2. Построить байесовскую оценку параметра a.

Решение. Так как

$$q(t) = (2\pi\sigma^2)^{-1/2} e^{-t^2/2\sigma^2},$$
$$f_t(x_1, \ldots, x_n) = (2\pi)^{-n/2} e^{-\sum_{i=1}^n (x_i - t)^2/2},$$

то плотность $q(t|x_1, \ldots, x_n)$ пропорциональна (как функция от t) произведению $q(t)f_t(x_1, \ldots, x_n)$ или, что то же, пропорциональна

$$e^{-t^2/2\sigma^2 - \sum_{i=1}^n (x_i - t)^2/2} = e^{-t^2(1/\sigma^2+n)/2 + \overline{x}nt - n\overline{x^2}/2}.$$

Из равенства

$$-\frac{t^2}{2}\Big(\frac{1}{\sigma^2} + n\Big) + \overline{X}nt = -\frac{1}{2}\Big(\frac{1}{\sigma^2} + n\Big)\Big(t - \frac{\overline{X}n}{1/\sigma^2 + n}\Big)^2 + \frac{(\overline{X}n)^2}{2(1/\sigma^2 + n)}$$

следует что плотность $q(t|x_1, \ldots, x_n)$ отвечает нормальному распределению со средним $\overline{X}n\sigma^2/(1 + n\sigma^2)$ и дисперсией $\sigma^2/(1 + n\sigma^2)$. Поэтому искомая оценка имеет вид

$$a_n^* = \int\limits_{\Theta} tq(t|X_1, \ldots, X_n)\, dt = \frac{\overline{X}n\sigma^2}{1 + n\sigma^2}.$$

5.2. Пусть дана выборка из нормального распределения со средним a и единичной дисперсией, причём параметр a имеет нормальное распределение с известным средним b и известной дисперсией σ^2. Построить байесовскую оценку параметра a.

5.3. Пусть дана выборка из нормального распределения со средним a и единичной дисперсией, причём параметр a имеет распределение Бернулли с параметром $1/2$. Построить байесовскую оценку параметра a.

5.4. Построить байесовскую оценку параметра θ равномерного распределения на отрезке $[0, \theta]$, если параметр θ имеет

а) плотность $q(t) = 1/t^2$ при $t \geqslant 1$;

б) равномерное распределение на отрезке $[0, 1]$.

5.5. Пусть дана выборка из равномерного распределения на отрезке $[0, \theta]$, причём θ принимает значения 1 и 2 с равными вероятностями. Построить байесовскую оценку параметра θ.

5.6. Пусть дана выборка из показательного распределения с параметром α, причём α имеет показательное распределение с параметром β. Построить байесовскую оценку параметра α.

5.7. Пусть дана выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

причём β равномерно распределено на отрезке $[0, 1]$. Построить байесовскую оценку параметра β.

5.8. Построить байесовскую оценку параметра θ распределения с плотностью $2y/\theta^2$ на отрезке $[0, \theta]$, если параметр θ имеет

а) равномерное распределение на отрезке $[0, 1]$;

б) плотность $3\theta^2$ на отрезке $[0, 1]$;

в) распределение Парето с параметрами β и 1, где значение $\beta > 0$ известно.

5.9. Пусть дана выборка из распределения Бернулли с параметром p, причём p равномерно распределено на отрезке $[0, 1]$. Построить байесовскую оценку параметра p.

5.10. Пусть дана выборка из распределения Бернулли с параметром p, причём p принимает значения $1/2$ и $1/3$ с одинаковыми вероятностями. Построить байесовскую оценку параметра p.

5.11. Пусть дана выборка из распределения Бернулли с параметром p, причём p имеет плотность $q(t) = \lambda t^{\lambda - 1}$ на отрезке $[0, 1]$, где $\lambda > 0$ известно. Построить байесовскую оценку параметра p.

5.12. Пусть дана выборка из распределения Пуассона с параметром λ, причём λ имеет показательное распределение с параметром 1. Построить байесовскую оценку параметра λ.

5.13. Пусть дана выборка из распределения Пуассона, причём параметр λ принимает значения 1 и 2 с вероятностями 1/3 и 2/3 соответственно. Построить байесовскую оценку параметра λ.

5.14. Пусть дана выборка из геометрического распределения с параметром p, причём p равномерно распределено на множестве $\{1/4,\ 1/2,\ 3/4\}$. Построить байесовскую оценку параметра p.

ОТДЕЛ III

СВОЙСТВА ОЦЕНОК

§ 6. Несмещённость и состоятельность

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть $X_1, X_2, \ldots$ — выборка из распределения F_θ и $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ — некоторая оценка параметра θ, построенная по данной выборке.

Оценка θ_n^* называется *состоятельной оценкой параметра* θ, если при любом $\theta \in \Theta$ величина θ_n^* сходится при $n \to \infty$ по вероятности к θ.

Оценка θ_n^* называется *сильно состоятельной оценкой параметра* θ, если при любом $\theta \in \Theta$ величина θ_n^* почти наверное сходится при $n \to \infty$ к θ.

Смещением оценки θ_n^* называется величина $b_n(\theta) = \mathbf{E}_\theta \theta_n^* - \theta$.

Оценка θ_n^* называется *несмещённой оценкой параметра* θ, если $b_n(\theta) = 0$ при любом $\theta \in \Theta$.

6.1. Для выборки из равномерного распределения на отрезке $[0, \theta]$ проверить состоятельность и несмещённость оценки $X_{(n)}$ параметра θ.

Решение. Плотность распределения величины $X_{(n)}$ равна ny^{n-1}/θ^n при $y \in [0, \theta]$. Поэтому

$$\mathbf{E}X_{(n)} = \int\limits_0^\theta y \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta.$$

Следовательно, смещение оценки $X_{(n)}$ равно $-\theta/(n+1)$ и она является смещённой, но асимптотически несмещённой. Проверим состоятельность: для любого фиксированного $\varepsilon \in (0, \theta)$

$$\mathbf{P}\{|\theta - X_{(n)}| \geqslant \varepsilon\} = \mathbf{P}\{X_{(n)} \leqslant \theta - \varepsilon\} = \left(\frac{\theta - \varepsilon}{\theta}\right)^n \to 0$$

при $n \to \infty$. Следовательно, оценка $X_{(n)}$ состоятельна. Более того, она сильно состоятельна, так как последовательность случайных величин $X_{(n)}$ не убывает с вероятностью 1.

6.2. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Проверить состоятельность и несмещённость следующих оценок параметра θ:

а) $2\overline{X}$;

б) $\overline{X} + X_{(n)}/2$;

в) $(n+1)X_{(1)}$;

г) $X_{(1)} + X_{(n)}$;

д) $\frac{n+1}{n} X_{(n)}$.

6.3. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. С помощью неравенства Чебышёва доказать состоятельность следующих оценок параметра θ:

а) $2\overline{X}$;

б) $X_{(n)}$.

6.4. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[a, b]$. Является ли оценка $\theta_n^* = X_{(n)} - X_{(1)}$ несмещённой оценкой длины отрезка $b - a$? Состоятельной?

6.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[-\theta, \theta]$. Является ли оценка максимального правдоподобия несмещённой оценкой параметра θ? Состоятельной?

6.6. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[-3\theta, \theta]$. Является ли оценка $\theta_n^* = 4X_{(n)} + X_{(1)}$ несмещённой оценкой параметра θ? Состоятельной?

6.7. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Доказать, что оценка метода моментов $\theta_{k,n}^* = \sqrt[k]{(k+1)\overline{X^k}}$ является

а) сильно состоятельной оценкой θ при любом $k \geqslant 1$;

б) смещённой оценкой θ при любом $k \geqslant 2$.

Решение. б) Заметим, что $\theta = \sqrt[k]{(k+1)\mathbf{E}_\theta \overline{X^k}}$. Поскольку в области $y \geqslant 0$ функция $g(y) = -\sqrt[k]{(k+1)y}$ строго выпуклая при $k \geqslant 2$, то по неравенству Йенсена

$$\mathbf{E}_\theta \theta_{k,n}^* = -\mathbf{E}_\theta g(\overline{X^k}) < -g(\mathbf{E}_\theta \overline{X^k}) = \theta,$$

причём неравенство строгое, так как распределение случайной величины $\overline{X^k}$ невырождено.

6.8. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения с конечным математическим ожиданием. Доказать, что $\overline{X}$ является несмещённой и состоятельной (сильно состоятельной) оценкой параметра $m_1 = \mathbf{E}X_1$.

6.9. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения с конечным моментом k-го порядка $m_k = \mathbf{E}X_1^k$. Является ли выборочный момент $\overline{X^k}$ порядка k несмещённой и состоятельной (сильно состоятельной) оценкой параметра m_k?

6.10. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения с конечной дисперсией. Доказать, что статистика

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

является состоятельной (сильно состоятельной) оценкой параметра $\sigma^2 = \mathbf{D}X_1$. Является ли S^2 несмещённой оценкой дисперсии σ^2? Построить оценку, являющуюся одновременно сильно состоятельной и несмещённой оценкой параметра σ^2.

6.11. Пусть $X_1, \ldots, X_n$ — выборка из распределения с конечным вторым моментом и значение $a = \mathbf{E}X_1$ известно. Проверить на несмещённость и состоятельность следующие оценки неизвестной дисперсии:

а) $\frac{1}{n-1}\sum\limits_{i=1}^{n}\big(X_i - \overline{X}\big)^2$;

б) $\overline{X^2} - a^2$;

в) $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - a)^2$;

г) $\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - a)^2$.

6.12. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с известным средним значением a и с неизвестной дисперсией σ^2. Проверить, является ли оценка $\sigma_n^* = \sqrt{\pi/2}\cdot\overline{|X-a|}$ несмещённой и состоятельной оценкой неизвестного параметра σ.

6.13. Пусть $X_1, \ldots, X_{2n}$ — выборка объёма $2n$ из некоторого распределения с конечным вторым моментом. Проверить на несмещённость и состоятельность оценку дисперсии σ^2

$$(\sigma^2)_n^* = \frac{1}{2n}\sum_{i=1}^{n}(X_{2i} - X_{2i-1})^2.$$

6.14. Пусть $(X_1, Y_1), \ldots, (X_n, Y_n)$ — выборка, соответствующая случайному вектору (ξ, η), т. е. $\mathbf{P}\{X_1 < x,\ Y_1 < y\} = \mathbf{P}\{\xi < x,\ \eta < y\}$. Доказать, что величина

$$m_{1,1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

является несмещённой и состоятельной оценкой $\mathbf{Cov}(\xi, \eta)$.

6.15. Даны результаты 8 независимых измерений одной и той же величины прибором, не имеющим систематических ошибок: 369, 378, 315, 420, 385, 401, 372, 383 *м*. Определить несмещённую оценку дисперсии ошибок измерений, если истинная длина

а) известна и равна 375 *м*; б) неизвестна.

6.16. Производится n измерений неизвестного диаметра d круга. В первом приближении считается, что измерения $X_i = d + \xi_i$ производятся с независимыми случайными ошибками ξ_i, имеющими одинаковое нормальное распределение с нулевым средним и неизвестной дисперсией σ^2. Проверить несмещённость и состоятельность следующей оценки площади круга:

$$s^* = \frac{\pi}{4}\left((\overline{X})^2 - S_0^2/n\right).$$

6.17. Производится n измерений неизвестной длины диагонали a квадрата. В первом приближении считается, что измерения $X_i = a + \xi_i$ производятся с независимыми случайными ошибками ξ_i, имеющими одинаковое нормальное распределение с нулевым средним и неизвестной дисперсией σ^2. Проверить несмещённость и состоятельность следующей оценки площади квадрата:

$$s^* = \left((\overline{X})^2 - S_0^2/n\right)/2.$$

6.18. Пусть $X_1, \ldots, X_{3n}$ — выборка объёма $3n$ из нормального распределения со средним a и единичной дисперсией. Проверить несмещённость и состоятельность следующих оценок параметра a:

а) $\displaystyle \frac{1}{n}\sum_{i=n+1}^{2n} X_i$; б) $\displaystyle \frac{1}{n}\sum_{i=1}^{n} X_{3i}$.

6.19. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Будет ли оценка $\alpha_n^* = 1/\overline{X}$ несмещённой? Если «нет», найти смещение. Является ли оценка состоятельной?

6.20. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Для какого параметра $\theta = \theta(\alpha)$ статистика $\theta_n^* = e^{\overline{X}}$ будет состоятельной оценкой? Является ли θ_n^* несмещённой оценкой того же параметра?

6.21. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром $1/\sqrt{\alpha}$. Является ли оценка $\alpha_n^* = (\overline{X})^2$ несмещённой оценкой параметра α? Состоятельной?

6.22. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Проверить оценки метода моментов

$$\alpha_k^* = \sqrt[k]{k!/\overline{X^k}}, \qquad k = 1, 2, \ldots,$$

на несмещённость. Будут ли эти оценки состоятельными?

6.23. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Выяснить, являются ли несмещёнными и состоятельными следующие оценки параметра сдвига β:

а) $X_{(1)}$; б) $\overline{X} - 1$.

6.24. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1} e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

Проверить состоятельность оценок параметров масштаба $\alpha > 0$ и сдвига $\beta \in \mathbf{R}$, построенных по методу моментов и по методу максимального правдоподобия.

6.25. Проверить состоятельность оценок метода моментов

$$\beta_n^* = 1 + \sqrt{1 + \overline{X}^2/S^2} \quad \text{и} \quad \theta_n^* = \overline{X}(1 - 1/\beta^*)$$

параметров $\beta > 2$ и $\theta > 0$ распределения Парето.

6.26. Проверить на несмещённость и состоятельность оценки максимального правдоподобия $\beta_n^* = 1/(\overline{\ln X} - \ln X_{(1)})$ и $\theta_n^* = X_{(1)}$, $n \geqslant 2$, параметров β и θ распределения Парето.

Р е ш е н и е. Заметим, что $\ln X_1$ имеет двухпараметрическое показательное распределение с плотностью

$$f(y) = \begin{cases} \beta\, e^{-\beta(y - \ln\theta)} & \text{при } y \geqslant \ln\theta, \\ 0 & \text{при } y < \ln\theta. \end{cases}$$

Поэтому статистика $\overline{\ln X} - \ln X_{(1)}$ распределена так же, как $\overline{Y} - Y_{(1)}$, где Y_i имеют показательное распределение с параметром β. Найдем распределение $n\overline{Y} - nY_{(1)}$, пользуясь результатом задачи 1.24. Величина $Y_{(k+1)} - Y_{(k)}$ имеет показательное распределение с параметром $(n-k)\beta$, поэтому $\xi_k = (n-k)(Y_{(k+1)} - Y_{(k)})$ имеет показательное распределение с параметром β, и при разных k эти величины независимы. Поскольку $\sum_{i=1}^n Y_i = \sum_{i=1}^n Y_{(i)}$, то

$$n\overline{Y} - nY_{(1)} = \sum_{k=1}^{n-1}(n-k)(Y_{(k+1)} - Y_{(k)}) = \sum_{k=1}^{n-1}\xi_k.$$

Смещение величины $(n-1)/(\xi_1 + \cdots + \xi_{n-1})$ вычислено в задаче 6.19 и равно $\alpha/(n-2)$. Поэтому смещение оценки β_n^* равно $2\alpha/(n-2)$.

6.27. Пусть дана выборка из распределения Вейбулла с параметрами α и θ, причём значение параметра α известно. Проверить на несмещённость и состоятельность оценку $1/\overline{X^\alpha}$ параметра θ.

6.28. Будет ли статистика $\overline{X}$ состоятельной оценкой параметра сдвига θ распределения Коши?

6.29. В партии из n изделий оказалось m бракованных. Неизвестная вероятность p появления бракованного изделия оценивается величиной m/n. Проверить состоятельность и несмещённость этой оценки.

6.30. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром $\sqrt{p}$. Является ли статистика $p_n^* = (\overline{X})^2$ несмещённой оценкой параметра p? Состоятельной?

6.31. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Показать, что для параметра $\tau(p) = 1/p$ не существует несмещённых оценок.

6.32. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Проверить, что статистики X_n, $X_1(1-X_n)$ и X_1X_n являются несмещёнными оценками для p, $p(1-p)$ и p^2 соответственно. Являются ли эти оценки состоятельными?

6.33. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Рассматривается класс оценок вида

$$p_n^* = \frac{n\overline{X} + \alpha}{n + \beta}, \quad \alpha \geqslant 0,\ \beta \geqslant 0.$$

Вычислить смещение и среднеквадратическую ошибку оценки p_n^*. Показать, что при $\alpha = \sqrt{n}/2$ и $\beta = \sqrt{n}$ ошибка не зависит от p.

6.34. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами 2 и p. Для какого параметра $\theta = \theta(p)$ статистика $\theta_n^* = e^{\overline{X}}$ будет состоятельной оценкой? Является ли θ_n^* несмещённой оценкой того же параметра?

6.35. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Проверить, что статистики $(X_1+X_n)/2$, $\overline{\mathbf{I}\{X = k\}}$ и X_n являются несмещёнными оценками для λ, $\lambda^k e^{-\lambda}/k!$ и λ соответственно. Являются ли эти оценки состоятельными?

6.36. Пусть $X_1, \ldots, X_n$ — выборка объёма $n \geqslant 5$ из распределения Пуассона с параметром λ. Для какого параметра $\theta = \theta(\lambda)$ статистика $\theta_n^* = X_1 \cdot \ldots \cdot X_5$ будет несмещённой оценкой? Является ли θ_n^* состоятельной оценкой для того же параметра θ?

6.37. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Для какого параметра $\theta = \theta(\lambda)$ статистика $\theta_n^* = \overline{X}e^{-\overline{X}}$ будет состоятельной оценкой? Является ли θ_n^* несмещённой оценкой того же параметра?

Решение. Так как $\overline{X}$ является состоятельной оценкой параметра λ, то $\theta_n^* = \overline{X}e^{-\overline{X}}$ сходится при $n \to \infty$ по вероятности к $\theta(\lambda) = \lambda e^{-\lambda}$. Случайная величина $n\overline{X}$ имеет распределение Пуассона с параметром $n\lambda$, поэтому

$$\mathbf{E}\,\theta_n^* = \mathbf{E}\,\overline{X}e^{-\overline{X}} = \sum_{k=0}^{\infty} \frac{k}{n}\, e^{-k/n}\, \frac{(n\lambda)^k}{k!}\, e^{-n\lambda}$$

$$= \frac{e^{-n\lambda}}{n} \sum_{k=1}^{\infty} \frac{\left(n\lambda\, e^{-1/n}\right)^k}{(k-1)!} = \lambda\, e^{n\lambda(e^{-1/n}-1)-1/n},$$

т. е. θ_n^* является смещённой (но асимптотически несмещённой) оценкой параметра $\theta = \theta(\lambda) = \lambda\, e^{-\lambda}$.

6.38. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Для какого параметра $\theta = \theta(\lambda)$ статистика $\theta_n^* = \overline{\mathbf{I}\{X = 1\}}$ будет состоятельной оценкой? Является ли θ_n^* несмещённой оценкой того же параметра?

6.39. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром $\ln\lambda$. Является ли оценка $\lambda_n^* = e^{\overline{X}}$ несмещённой оценкой параметра λ? Состоятельной?

6.40. Имеется одно наблюдение X_1 с усечённым снизу распределением Пуассона:

$$\mathbf{P}\{X_1 = k\} = \frac{\lambda^k}{k!} \cdot \frac{e^{-\lambda}}{1 - e^{-\lambda}}, \quad k \geqslant 1.$$

Доказать, что единственная несмещённая оценка параметра $\theta = 1 - e^{-\lambda}$ имеет вид

$$\theta_1^* = \begin{cases} 0, & \text{если } X_1 \text{ нечетно}, \\ 2, & \text{если } X_1 \text{ четно}. \end{cases}$$

6.41. Построить оценку параметра λ распределения Пуассона, которая одновременно является

а) состоятельной и смещённой;

б) несостоятельной и несмещённой.

6.42. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Показать, что для параметра $\tau(\lambda) = 1/\lambda$ не существует несмещённых оценок.

6.43. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на конечном множестве $\{1, \ldots, \theta\}$, где θ — целый положительный параметр. Проверить оценку максимального правдоподобия параметра θ на несмещённость и состоятельность.

6.44. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p. Будет ли оценка $p_n^* = 1/(1 + \overline{X})$ несмещённой? Состоятельной?

6.45. Пусть дана выборка из распределения P_q, $q \in (0, 1/2)$:

$$P_q\{X_1 = k\} = \begin{cases} q^5 & \text{при } k = 1, \\ 1 - q - q^5 & \text{при } k = 2, \\ q & \text{при } k = 3. \end{cases}$$

Пусть ν_n — число элементов выборки, равных 1. Является ли оценка $q_n^* = \sqrt[5]{\nu_n/n}$ несмещённой оценкой параметра q? Состоятельной?

6.46. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2. Проверить, является ли выборочная медиана ζ^* состоятельной и несмещённой оценкой параметра a.

6.47. Пусть $X_1, \ldots, X_n$ — выборка с функцией распределения F, причем производная $F'(y)$ всюду положительна. Доказать, что выборочная квантиль ζ_δ^* уровня $\delta \in (0,1)$ является сильно состоятельной оценкой истинной квантили $\zeta_\delta = F^{-1}(\delta)$.

6.48. Привести пример функции распределения F, для которой выборочная квантиль ζ_δ^* не является сильно состоятельной оценкой квантили $\zeta_\delta = \sup\{y\colon F(y) \leqslant \delta\}$.

6.49. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, 2\theta]$. Проверить, является ли выборочная медиана несмещённой и состоятельной оценкой параметра θ.

6.50. Пусть дана выборка $X_1, \ldots, X_n$ из распределения F_θ с параметром $\theta \in \{1, 2, \ldots, N\}$, причем $F_{\theta_1} \neq F_{\theta_2}$ при $\theta_1 \neq \theta_2$. Пусть $\rho(F, G) = \sup\limits_y |F(y) - G(y)|$. Доказать, что оценка θ_n^*, выбираемая по правилу

$$\rho(F_n^*, F_{\theta_n^*}) = \min_\theta \rho(F_n^*, F_\theta),$$

является состоятельной.

6.51. Пусть θ^* — оценка параметра θ со смещением $b(\theta) = 2\theta$. Построить несмещённую оценку параметра θ.

6.52. Пусть θ_n^* — асимптотически несмещённая оценка для θ и $\mathbf{D}_\theta \theta_n^* \to 0$ при $n \to \infty$ для любого $\theta \in \Theta$. Доказать, что оценка θ_n^* состоятельна.

6.53. Пусть имеется выборка из распределения F_θ, $\theta \in \Theta \subseteq \mathbf{R}$, и $\alpha = f(\theta)$, где f — выпуклая вещественнозначная функция. Пусть θ^* — несмещённая оценка параметра θ. Доказать, что оценка $\alpha^* = f(\theta^*)$ параметра α имеет неотрицательное смещение. При каких условиях смещение будет строго положительным?

6.54. Привести пример, когда оценка

а) метода моментов является смещённой;

б) максимального правдоподобия является несмещённой;

в) является несмещённой и не состоятельной;

г) является смещённой и состоятельной.

6.55. Доказать, что выборочное среднее $\overline{X}$ и выборочная дисперсия S^2 некоррелированы, если третий момент выборки равен нулю. *Указание*: доказать, что $\mathbf{Cov}(\overline{X}, S^2) = \frac{n-1}{n^2}\,\mathbf{E}X_1^3$.

6.56. Доказать, что при любом фиксированном y значение эмпирической функции распределения $F_n^*(y)$ является (сильно) состоятельной и несмещённой оценкой значения функции распределения выборки $F(y)$.

6.57. Пусть $X_1, \ldots, X_n$ — выборка с функцией распределения F и ν_n — число элементов выборки, попавших в полуинтервал $[a, b)$, где $a < b$ — фиксированные числа. Доказать, что статистика ν_n/n является состоятельной и несмещённой оценкой разности $F(b) - F(a)$.

6.58. Доказать, что для любого фиксированного $\lambda \in \mathbf{R}$ значение выборочной характеристической функции

$$\varphi_n^*(\lambda) = \int\limits_{\mathbf{R}} e^{i\lambda y} F_n^*(dy)$$

является (сильно) состоятельной и несмещённой оценкой истинного значения характеристической функции $\varphi(\lambda) = \mathbf{E}e^{i\lambda X_1}$.

§ 7. Асимптотическая нормальность

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть $X_1, X_2, \ldots$ — выборка из распределения F_θ и $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ — некоторая оценка параметра θ, построенная по данной выборке.

Статистика θ_n^* называется *асимптотически нормальной оценкой параметра θ с коэффициентом* $\sigma^2(\theta) > 0$, если при любом $\theta \in \Theta$ распределение случайной величины $(\theta_n^* - \theta)\sqrt{n}$ слабо сходится при $n \to \infty$ к нормальному закону с нулевым средним и дисперсией $\sigma^2(\theta)$.

7.1. Пусть дана выборка из распределения с конечной дисперсией. Доказать, что статистика $\overline{X}$ является асимптотически нормальной оценкой для $\theta = \mathbf{E}X_1$. Найти коэффициент асимптотической нормальности.

Решение. Имеем равенство

$$\sqrt{n}(\overline{X} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbf{E}X_1).$$

Поэтому в силу центральной предельной теоремы распределение отношения

$$\frac{\sqrt{n}(\overline{X}-\theta)}{\sqrt{\mathbf{D}X_1}}$$

слабо сходится к стандартному нормальному закону. Следовательно, оценка $\overline{X}$ асимптотически нормальна с коэффициентом асимптотической нормальности $\sigma^2 = \mathbf{D}X_1$.

7.2. Пусть $\mathbf{E}g^2(X_1) < \infty$. Доказать, что статистика $\overline{g(X)}$ является асимптотически нормальной оценкой для параметра $\theta = \mathbf{E}g(X_1)$. Найти коэффициент асимптотической нормальности.

7.3. Доказать, что выборочная дисперсия S^2 при условии конечности $\mathbf{E}X_1^4$ является асимптотически нормальной оценкой дисперсии. Вычислить коэффициент асимптотической нормальности.

Решение. Положим $a = \mathbf{E}X_1$ и $\sigma^2 = \mathbf{D}X_1$. Представим выборочную дисперсию $S^2 = \frac{1}{n}\sum(X_i-\overline{X})^2$ в виде $S^2 = \overline{(X-a)^2} - (\overline{X}-a)^2$. Заметим, что по центральной предельной теореме величина

$$\sqrt{n}(\overline{X}-a)^2 = \frac{1}{\sqrt{n}}(\sqrt{n}(\overline{X}-a))^2$$

сходится при $n \to \infty$ по вероятности к нулю, а распределение случайной величины

$$\sqrt{n}(\overline{(X-a)^2} - \sigma^2) = \frac{\sum_1^n (X_i-a)^2 - n\mathbf{E}(X_1-a)^2}{\sqrt{n}}$$

слабо сходится к нормальному закону с нулевым средним и дисперсией равной $\mathbf{D}(X_1-a)^2$. Сложив слабо сходящуюся последовательность с последовательностью, сходящейся по вероятности к нулю, получим слабую сходимость

$$\sqrt{n}(S^2-\sigma^2) = \sqrt{n}(\overline{(X-a)^2}-\sigma^2) - \sqrt{n}(\overline{X}-a)^2$$

также к нормальному закону с нулевым средним и дисперсией $\mathbf{D}(X_1-a)^2$. Таким образом, S^2 — асимптотически нормальная оценка параметра σ^2 с коэффициентом $\mathbf{D}(X_1-a)^2 = \mathbf{E}(X_1-a)^4 - \sigma^4$.

7.4. Доказать, что любая асимптотически нормальная оценка является состоятельной.

7.5. Пусть θ_n^* — асимптотически нормальная оценка для θ с коэффициентом σ^2, причем $\mathbf{E}_\theta(\theta_n^*-\theta)^4 < C/n^2$ для любого θ. Доказать, что при $n \to \infty$ имеет место соотношение

$$\mathbf{E}_\theta(\theta_n^*-\theta)^2 = \sigma^2 n^{-1}(1+o(1)).$$

7.6. Пусть θ_n^* — асимптотически нормальная оценка для параметра θ с коэффициентом σ^2. Пусть $\theta \neq 0$. Доказать, что $(\theta_n^*)^2$ — асимптотически нормальная оценка для θ^2. Найти коэффициент асимптотической нормальности.

Решение. Имеем равенство

$$\sqrt{n}\,\left((\theta_n^*)^2 - \theta^2\right) = \sqrt{n}(\theta_n^* - \theta)(\theta_n^* + \theta).$$

Поскольку θ_n^* состоятельна (см. задачу 7.4), $\theta_n^* + \theta \xrightarrow{\text{p}} 2\theta$. Умножая слабо сходящуюся к нормальному закону с нулевым средним и дисперсией σ^2 последовательность $\sqrt{n}(\theta_n^* - \theta)$ на последовательность, сходящуюся по вероятности к постоянной 2θ, получим слабую сходимость распределения $\sqrt{n}((\theta_n^*)^2 - \theta^2)$ к нормальному закону с нулевым средним и дисперсией $4\sigma^2\theta^2$.

7.7. Пусть θ^* — асимптотически нормальная оценка для θ. Будет ли $|\theta^*|$ асимптотически нормальной оценкой для $|\theta|$?

7.8. Пусть θ_n^* — асимптотически нормальная оценка для параметра $\theta \in \Theta$ с коэффициентом $\sigma^2(\theta)$, функция $H(y)$ непрерывно дифференцируема в области Θ и $H'(\theta) \neq 0$. Доказать, что $H(\theta_n^*)$ — асимптотически нормальная оценка для $H(\theta)$ с коэффициентом $\widetilde{\sigma}^2(\theta) = \left(H'(\theta)\right)^2 \sigma^2(\theta)$.

7.9. Пусть $X_1, \ldots, X_n$ — выборка со средним $\mathbf{E}X_1 = a$ и дисперсией $\mathbf{D}X_1 = \sigma^2 > 0$. Пусть функция $H(t)$ дважды непрерывно дифференцируема в точке $t = a$ и $H'(a) = 0$. Показать, что

а) величина $\sqrt{n}(H(\overline{X}) - H(a))$ сходится при $n \to \infty$ по вероятности к нулю;

б) распределение случайной величины $n(H(\overline{X}) - H(a))$ слабо сходится при $n \to \infty$ к распределению квадрата случайной величины, распределённой по нормальному закону с нулевым средним и дисперсией $H''(a)\sigma^2/2$.

7.10. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами a и σ^2, причём значение параметра a известно. Является ли асимптотически нормальной для параметра σ оценка $\sigma_n^* = \sqrt{\pi/2} \cdot \overline{|X - a|}$?

7.11. Пусть $X_1, \ldots, X_{2n}$ — выборка объёма $2n$ из нормального распределения. Является ли оценка

$$\frac{1}{2n}\sum_{i=1}^{n}(X_{2i}-X_{2i-1})^2.$$

асимптотически нормальной для неизвестной дисперсии σ^2?

7.12. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,\theta]$ и $k \geqslant 1$. Доказать асимптотическую нормальность оценки $\sqrt[k]{(k+1)\overline{X^k}}$ параметра θ и найти коэффициент асимптотической нормальности.

7.13. В условиях предыдущей задачи доказать, что почти наверное $\theta^*_{k,n} \to X_{(n)}$ при $k \to \infty$. Является ли $X_{(n)}$ асимптотически нормальной оценкой параметра θ?

7.14. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,\theta]$. Является ли статистика $\frac{n+1}{n}X_{(n)}$ асимптотически нормальной оценкой параметра θ?

7.15. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta/2,\theta]$. Показать, что статистика $\ln(4\overline{X}/3)$ является асимптотически нормальной оценкой параметра $\tau = \ln\theta$. Найти коэффициент асимптотической нормальности.

7.16. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,a]$. Для какого параметра $\theta = \theta(a)$ статистика $\theta^*_n = \ln\overline{X}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

7.17. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Доказать, что для любого натурального k статистика $\sqrt[k]{k!/\overline{X^k}}$ является асимптотически нормальной оценкой параметра α. Найти коэффициент асимптотической нормальности.

7.18. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Показать, что статистика $-\ln\overline{X}$ является асимптотически нормальной оценкой параметра $\tau = \ln\alpha$. Найти коэффициент асимптотической нормальности.

7.19. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Для какого параметра $\theta = \theta(\alpha)$ статистика $\theta_n^* = e^{-\overline{X^2}}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

7.20. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta-y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Проверить, являются ли асимптотически нормальными следующие оценки параметра сдвига β:

а) $X_{(1)}$; б) $\overline{X} - 1$?

Если «да», найти коэффициент асимптотической нормальности.

7.21. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1} e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

а) Доказать асимптотическую нормальность оценок метода моментов $\alpha_n^* = \sqrt{S^2}$ и $\beta_n^* = \overline{X} - \sqrt{S^2}$ параметров $\alpha > 0$ и $\beta \in \mathbf{R}$. Найти коэффициенты асимптотической нормальности.

б) Являются ли асимптотически нормальными оценки максимального правдоподобия $\alpha_n^* = \overline{X} - X_{(1)}$ и $\beta_n^* = X_{(1)}$ параметров α и β? Если «да», найти коэффициенты асимптотической нормальности.

Решение. а) Рассмотрим случайные величины $Y_i = X_i - \beta$, которые имеют показательное распределение с параметром $1/\alpha$. Тогда

$$\sqrt{n}\,(\beta_n^* - \beta) = \sqrt{n}\Big(\overline{X} - \sqrt{\overline{X^2} - (\overline{X})^2} - \beta\Big) \equiv \sqrt{n}\Big(\overline{Y} - \sqrt{\overline{Y^2} - (\overline{Y})^2}\,\Big).$$

Положим $h(t_1, t_2) = t_1 - \sqrt{t_2 - t_1^2}$,

$$G(F) = h(\mathbf{E}_F Y_1, \mathbf{E}_F Y_1^2), \quad G(F_n^*) = h(\overline{Y}, \overline{Y^2}) = \overline{Y} - \sqrt{\overline{Y^2} - (\overline{Y})^2}.$$

Функция $h(t)$ дифференцируема в точке $a = (\mathbf{E}Y_1, \mathbf{E}Y_1^2) = (\alpha, 2\alpha^2)$. Частные производные в этой точке равны $(2, -1/2\alpha)$, и $h(a) = \alpha - \sqrt{2\alpha^2 - \alpha^2} = 0$. Матрица ковариаций σ^2 конечна:

$$\begin{aligned} \sigma_{1,1} &= \mathbf{D}Y_1 = \alpha^2, \\ \sigma_{2,2} &= \mathbf{D}Y_1^2 = 20\alpha^4, \\ \sigma_{1,2} &= \sigma_{2,1} = \mathbf{Cov}(Y_1, Y_1^2) = 4\alpha^3. \end{aligned}$$

Поэтому по теореме 1А из [4, гл. 1, § 7] получаем, что случайная величина

$$\sqrt{n}\Big(\overline{Y} - \sqrt{\overline{Y^2} - (\overline{Y})^2}\Big) = \sqrt{n}\Big(h(\overline{Y}, \overline{Y^2}) - h(a)\Big)$$

слабо сходится к величине $\eta = \frac{\partial h}{\partial t_1}(a) \cdot \xi_1 + \frac{\partial h}{\partial t_2}(a) \cdot \xi_2 = 2\xi_1 - \xi_2/2\alpha$, где вектор (ξ_1, ξ_2) имеет нормальное распределение с нулевым вектором средних и матрицей ковариаций σ^2. Величина η имеет нормальное распределение с нулевым средним и дисперсией

$$\mathbf{D}\eta = 4\sigma_{1,1} + \sigma_{2,2}/4\alpha^2 - 2 \cdot 2 \cdot \sigma_{1,2}/2\alpha = \alpha^2.$$

Итак, оценка β_n^* асимптотически нормальна с коэффициентом α^2.

7.22. Пусть дана выборка из распределения Парето с параметрами β и θ. Являются ли асимптотически нормальными оценки максимального правдоподобия $\beta_n^* = 1/(\overline{\ln X} - \ln X_{(1)})$ и $\theta_n^* = X_{(1)}$ параметров β и θ? Если «да», найти коэффициенты асимптотической нормальности. *Указание*: воспользоваться решением задачи 6.26.

7.23. Пусть дана выборка из распределения Вейбулла с известным параметром α и с неизвестным параметром $\theta > 0$. Является ли асимптотически нормальной оценка $1/\overline{X^\alpha}$ неизвестного значения $\theta > 0$?

7.24. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Для какого параметра $\theta = \theta(p)$ статистика $\theta_n^* = \arcsin\sqrt{\overline{X}}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

Решение. Статистика $\theta_n^* = \arcsin\sqrt{\overline{X}}$ имеет вид $\theta_n^* = H(\overline{g(X)})$, где $H(t) = \arcsin\sqrt{t}$, $g(y) = y$. Функция $H(t)$ непрерывно дифференцируема в точке $\mathbf{E}_p g(X_1) = p$,

$$H'(t) = \frac{1}{2\sqrt{(1-t)t}}, \qquad H'(t)\big|_{t=\mathbf{E}_p g(X_1)} = \frac{1}{2\sqrt{(1-p)p}}.$$

Следовательно, θ_n^* является асимптотически нормальной оценкой параметра $\theta = \arcsin\sqrt{\mathbf{E}_p g(X_1)} = \arcsin\sqrt{p}$ с коэффициентом

$$\sigma^2(p) = \big(H'(\mathbf{E}_p g(X_1))\big)^2 \cdot \mathbf{D}_p g(X_1) = 1/4.$$

7.25. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Показать, что статистика $\arcsin(2\overline{X} - 1)$ является асимптотически нормальной оценкой параметра $\tau = \arcsin(2p-1)$.

Найти коэффициент асимптотической нормальности.

7.26. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами m и p. Для какого параметра $\theta = \theta(m,p)$ статистика $\theta_n^* = e^{\overline{X}}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

7.27. Показать, что статистика $\overline{X}$ является асимптотически нормальной оценкой параметра λ распределения Пуассона. Найти коэффициент асимптотической нормальности.

7.28. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Показать, что статистика $\sqrt{\overline{X}}$ является асимптотически нормальной оценкой параметра $\tau = \sqrt{\lambda}$. Найти коэффициент асимптотической нормальности.

7.29. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ, $\lambda \neq 1$. Для какого параметра $\theta = \theta(\lambda)$ статистика $\theta_n^* = \overline{X}\, e^{-\overline{X}}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

7.30. Построить оценку параметра λ распределения Пуассона, которая одновременно является состоятельной и не асимптотически нормальной.

7.31. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p. Будет ли статистика $p_n^* = 1/(1+\overline{X})$ асимптотически нормальной оценкой параметра p? Найти коэффициент асимптотической нормальности.

7.32. Пусть $X_1, \ldots, X_n$ — выборка из распределения со средним $\mathbf{E}X_1 = 1$ и дисперсией $\mathbf{D}X_1 = \sigma^2 > 0$; обозначим $S_n = X_1 + \cdots + X_n$. Найти слабый предел последовательности распределений случайных величин

$$\psi_n \equiv S_n^3/n^{5/2} - \sqrt{n}.$$

7.33. Доказать, что для случайной величины χ_n^2, имеющей χ^2-распределение с n степенями свободы, справедлива «аппроксимация Фишера»: распределение разности $\sqrt{2\chi_n^2} - \sqrt{2n}$ слабо сходится при $n \to \infty$ к стандартному нормальному закону.

7.34. Доказать, что для случайной величины χ_n^2, имеющей χ^2-распределение с n степенями свободы, справедлива «аппроксимация Уилсона – Хилферти»: распределение величины

$$\sqrt{9n/2}\left(\sqrt[3]{\chi_n^2/n}-1+2/9n\right)$$

слабо сходится при $n\to\infty$ к стандартному нормальному закону.

7.35. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,2\theta]$. Доказать, что выборочная медиана ζ^* — асимптотически нормальная оценка для θ. Найти коэффициент асимптотической нормальности.

7.36. Пусть $X_1, \ldots, X_n$ — выборка из распределения Коши с параметром сдвига a. Доказать, что выборочная медиана ζ^* — асимптотически нормальная оценка для a. Найти коэффициент асимптотической нормальности.

7.37. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Доказать, что выборочная медиана ζ^* — асимптотически нормальная оценка для параметра $\tau=(\ln 2)/\alpha$. Найти коэффициент асимптотической нормальности.

7.38. Пусть $X_1, \ldots, X_n$ — выборка из некоторого распределения с абсолютно непрерывной функцией распределения F, для которой плотность $f(x)$ непрерывно дифференцируема в некоторой окрестности медианы ζ распределения F. Доказать, что выборочная медиана является асимптотически нормальной оценкой медианы ζ распределения F. Найти коэффициент асимптотической нормальности.

7.39. Пусть $X_1, \ldots, X_n$ — выборка с абсолютно непрерывной функцией распределения F, причем плотность $f(x)$ всюду непрерывно дифференцируема. Доказать, что выборочная квантиль ζ_δ^* уровня $\delta\in(0,1)$ является асимптотически нормальной оценкой истинной квантили $\zeta_\delta=F^{-1}(\delta)$. Найти коэффициент асимптотической нормальности.

7.40. Доказать, что при любом фиксированном y таком, что $0<F(y)<1$, значение эмпирической функции распределения $F_n^*(y)$ является асимптотически нормальной оценкой значения общей функции распределения выборки $F(y)$. Найти коэффициент асимптотической нормальности.

7.41. Доказать, что для любого фиксированного $\lambda \in \mathbf{R}$ значение выборочной характеристической функции

$$\varphi_n^*(\lambda) = \int\limits_{\mathbf{R}} e^{i\lambda y} F_n^*(dy)$$

является асимптотически нормальной оценкой истинного значения характеристической функции $\varphi(\lambda) = \mathbf{E} e^{i\lambda X_1}$.

Решение. В силу центральной предельной теоремы распределение комплекснозначной случайной величины $\sqrt{n}(\overline{e^{i\lambda X}} - \varphi(\lambda))$ слабо сходится к распределению $\xi + i\eta$, где вектор (ξ, η) имеет нормальное распределение на плоскости с нулевым вектором средних и (возможно, в зависимости от значения λ, вырожденной) ковариационной матрицей

$$\sigma^2 = \begin{pmatrix} \mathbf{D}\cos(\lambda X_1) & \mathbf{Cov}(\cos(\lambda X_1), \sin(\lambda X_1)) \\ \mathbf{Cov}(\cos(\lambda X_1), \sin(\lambda X_1)) & \mathbf{D}\sin(\lambda X_1) \end{pmatrix}.$$

ОТДЕЛ IV

СРАВНЕНИЕ ОЦЕНОК

§ 8. Среднеквадратический подход

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть $X_1, X_2, \ldots$ — выборка из распределения F_θ и $\theta_n^* = \theta^*(X_1, \ldots, X_n)$ — некоторая оценка, построенная по данной выборке.

Среднеквадратическим отклонением оценки θ_n^* параметра θ называется величина $\mathbf{E}_\theta(\theta_n^* - \theta)^2$. Среднеквадратическое отклонение оценки связано с дисперсией и смещением оценки следующим равенством:

$$\mathbf{E}_\theta(\theta_n^* - \theta)^2 = \mathbf{D}_\theta\theta_n^* + b^2(\theta).$$

В соответствии со среднеквадратическим подходом оценка θ_n^* *не хуже* оценки θ_n^{**}, если при любом $\theta \in \Theta$ выполняется неравенство

$$\mathbf{E}_\theta(\theta_n^* - \theta)^2 \leqslant \mathbf{E}_\theta(\theta_n^{**} - \theta)^2.$$

В соответствии со среднеквадратическим подходом оценка θ_n^* *лучше* оценки θ_n^{**}, если θ_n^* не хуже оценки θ_n^{**} и хотя бы для одного $\theta \in \Theta$ выполняется неравенство

$$\mathbf{E}_\theta(\theta_n^* - \theta)^2 < \mathbf{E}_\theta(\theta_n^{**} - \theta)^2.$$

8.1. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним $a > 0$ и единичной дисперсией. Сравнить оценки $\overline{X}$ и $\max(0, \overline{X})$ параметра a в среднеквадратичном смысле.

8.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с известным средним a и с неизвестной дисперсией σ^2. Сравнить в среднеквадратичном смысле оценки параметра σ^2

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{и} \quad S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2.$$

8.3. Пусть $X_1, \ldots, X_{2n}$ — выборка объёма $2n$ из нормального распределения со средним a и дисперсией σ^2. Сравнить в среднеквадратичном смысле оценки параметра σ^2

$$S_0^2 = \frac{1}{2n-1}\sum_{i=1}^{2n}(X_i - \overline{X})^2 \quad \text{и} \quad (\sigma^2)^*_{2n} = \frac{1}{2n}\sum_{i=1}^{n}(X_{2i} - X_{2i-1})^2.$$

8.4. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и с дисперсией σ^2. Найти оценку дисперсии, наилучшую в среднеквадратичном смысле в классе оценок вида

$$c_n\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Найти её смещение.

8.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Сравнить оценки $2\overline{X}$, $X_{(n)}$, $\frac{n+1}{n}X_{(n)}$ и $X_{(1)} + X_{(n)}$ параметра θ в среднеквадратичном смысле.

8.6. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Сравнить оценки $\theta^*_{k,n} = \frac{n+k}{n}X_{(n)}$, $k = 0, 1, 2, \ldots,$ параметра θ в среднеквадратичном смысле.

8.7. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найти оценку параметра θ, наилучшую в среднеквадратичном смысле в классе оценок вида $c_nX_{(n)}$. Найти её смещение.

8.8. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta, 2\theta]$. Рассматривается класс несмещённых оценок для θ вида $aX_{(1)} + bX_{(n)}$. Сравнить оценки из этого класса в среднеквадратичном смысле.

8.9. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta, \theta+1]$.

а) Сравнить в среднеквадратичном смысле оценки $\overline{X} - 1/2$, $X_{(1)}$ и $X_{(n)} - 1$ параметра θ.

б) Найти оценку параметра θ, наилучшую в среднеквадратичном смысле в подклассе оценок максимального правдоподобия вида $a(X_{(n)} - 1) + (1 - a)X_{(1)}$, $a \in [0, 1]$.

8.10. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta-y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Сравнить в среднеквадратичном смысле оценки $\overline{X} - 1$, $X_{(1)}$ и $X_{(1)} - 1/n$ параметра сдвига β.

8.11. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Построить любые две различные оценки параметра λ и сравнить их в среднеквадратичном смысле.

8.12. Привести примеры оценок θ_1^* и θ_2^* с дисперсиями $\sigma_1^2(\theta)$ и $\sigma_2^2(\theta)$ соответственно такими, что $\sigma_1^2(\theta) \leqslant \sigma_2^2(\theta)$ при всех $\theta \in \Theta$, но $\mathbf{E}_\theta(\theta_1^* - \theta)^2 > \mathbf{E}_\theta(\theta_2^* - \theta)^2$.

8.13. Доказать, что если оценки θ_1^* и θ_2^* имеют одинаковое смещение, то $\mathbf{D}_\theta\theta_1^* \leqslant \mathbf{D}_\theta\theta_2^*$ для любого $\theta \in \Theta$ тогда и только тогда, когда $\mathbf{E}_\theta(\theta_1^* - \theta)^2 \leqslant \mathbf{E}_\theta(\theta_2^* - \theta)^2$.

§ 9. Асимптотический подход

Наряду со среднеквадратическим часто применяется *асимптотический подход* к сравнению оценок. Он удобен для сравнения асимптотически нормальных оценок в случае, когда объём выборки очень велик. Согласно асимптотическому подходу, асимптотически нормальная оценка θ_1^* с коэффициентом асимптотической нормальности $\sigma_1^2(\theta)$ *лучше* асимптотически нормальной оценки θ_2^* с коэффициентом $\sigma_2^2(\theta)$, если при любом $\theta \in \Theta$ выполняется неравенство $\sigma_1^2(\theta) \leqslant \sigma_2^2(\theta)$ и хотя бы для одного $\theta \in \Theta$ выполняется неравенство $\sigma_1^2(\theta) < \sigma_2^2(\theta)$.

9.1. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с известным средним a и с неизвестной дисперсией σ^2. При помощи асимптотического подхода сравнить следующие оценки параметра σ^2:

$$\frac{\pi}{2}\Big(\frac{1}{n}\sum_{i=1}^{n}|X_i - a|\Big)^2 \quad \text{и} \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2.$$

9.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией $\sigma^2 > 0$. При помощи асимптотического подхода сравнить выборочное среднее и выборочную медиану как оценки параметра a.

9.3. Пусть $X_1, \ldots, X_n$ — выборка из распределения, являющегося смесью двух нормальных распределений, а именно, 92% составляет нормальное распределение со средним a и дисперсией 1, а 8% составляет нормальное распределение с тем же средним a и дисперсией 16. При помощи асимптотического подхода сравнить выборочное среднее и выборочную медиану как оценки параметра a.

9.4. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Существует ли наилучшая асимптотически нормальная оценка среди оценок $\theta^*_{k,n} = \sqrt[k]{(k+1)\overline{X^k}}$?

9.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, 2\theta]$. При помощи асимптотического подхода сравнить выборочное среднее и выборочную медиану как оценки параметра θ.

9.6. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Существует ли наилучшая асимптотически нормальная оценка среди оценок $\alpha^*_{k,n} = \sqrt[k]{k!/\overline{X^k}}$?

§ 10. Достаточные статистики

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений и $X_1, X_2, \ldots$ — выборка из распределения F_θ.

Статистика $S(X_1, \ldots, X_n)$, построенная по выборке $X_1, \ldots, X_n$, называется *достаточной* для параметра θ, если условное распределение выборки при фиксированном значении статистики S

$$\mathbf{P}_\theta\{(X_1, \ldots, X_n) \in B \mid S = s\}, \quad B \subseteq \mathbf{R}^n,$$

не зависит от параметра θ[1].

Пусть выполнено условие доминирования относительно некоторой меры μ на $\mathbf{R}$, т. е. это параметрическое семейство состоит из распределений, абсолютно непрерывных относительно μ. Обозначим через f_θ плотность распределения F_θ относительно меры μ.

[1]Поскольку условное распределение определяется с точностью до эквивалентности, корректнее было бы сказать, что найдётся вариант условного распределения, не зависящий от параметра θ.

Тогда справедлив следующий критерий достаточности статистики.

Теорема Неймана – Фишера о факторизации. *Статистика S является достаточной для параметра θ тогда и только тогда, когда совместная плотность выборки может быть представлена в виде*

$$f_\theta(x_1,\ldots,x_n) \equiv \prod_{i=1}^{n} f_\theta(x_i) = \psi(S(x_1,\ldots,x_n),\theta)\cdot h(x_1,\ldots,x_n).$$

10.1. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ. Найти условное совместное распределение выборки при условии $X_1 = x_1, \ldots, X_n = x_n$. Является ли достаточной статистикой для параметра θ

а) выборка; б) вариационный ряд?

10.2. Распределение F_θ задано плотностью f_θ относительно некоторой меры μ. Пользуясь теоремой Неймана – Фишера о факторизации, доказать, что вариационный ряд, построенный по выборке $X_1, \ldots, X_n$, является достаточной статистикой для параметра θ.

10.3. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией. Найти условное совместное распределение выборки при условии $X_1 + \cdots + X_n = y$. Является ли $\overline{X}$ достаточной статистикой для параметра a?

10.4. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с нулевым средним и дисперсией σ^2. Найти достаточную статистику для параметра σ^2 со значениями в $\mathbf{R}$.

10.5. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2. Будет ли статистика $\overline{X^2}$ достаточной для

а) двумерного параметра (a, σ^2);

б) параметра σ^2, если $a = 0$;

в) параметра σ^2, если $a = 3$?

10.6. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2. Найти достаточную для двумерного параметра (a, σ^2) статистику со значениями в $\mathbf{R}^2$.

10.7. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найти достаточную для параметра θ статистику со значениями в $\mathbf{R}$.

10.8. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[a, b]$. Достаточна ли для двумерного параметра (a, b) статистика $\overline{X}$? Статистика $X_{(n)}$? Двумерная статистика $(X_{(1)}, X_{(n)})$?

10.9. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке

а) $[\theta, \theta + 1]$; б) $[\theta, 2\theta]$.

Найти достаточную для параметра θ статистику со значениями в $\mathbf{R}^2$.

10.10. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[-\theta, \theta]$. Найти достаточную для параметра θ статистику со значениями в $\mathbf{R}$.

10.11. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Найти достаточную для параметра α статистику $S(X_1, \ldots, X_n)$ со значениями в $\mathbf{R}$ такую, что любая другая достаточная статистика есть неслучайная функция от S (такая достаточная статистика S называется *минимальной*).

10.12. Найти достаточную статистику для параметра сдвига $\beta \in \mathbf{R}$ смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Является ли $\overline{X}$ достаточной статистикой?

10.13. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$. Найти достаточную статистику для

а) параметра β, если значение α известно;

б) параметра α, если значение β известно;

в) для двумерного параметра $\theta = (\alpha, \beta)$.

10.14. Пусть $X_1, \ldots, X_n$ — выборка из Г-распределения с параметрами $\alpha = 1/\theta$ и β, причём β известно. Существует ли достаточная для параметра $\theta > 0$ статистика со значениями в $\mathbf{R}$? Найти распределение статистики $\overline{X}/\beta$. Достаточная ли это статистика?

10.15. Пусть $X_1, \ldots, X_n$ — выборка из Г-распределения с параметрами α и β. Существует ли достаточная для двумерного параметра (α, β) статистика со значениями в $\mathbf{R}^2$?

10.16. Пусть дана выборка из распределения Парето с параметрами $\beta > 0$ и $\theta > 0$. Найти достаточную статистику для

а) параметра β, если значение θ известно;

б) параметра θ, если значение β известно;

в) векторного параметра (β, θ).

10.17. Пусть дана выборка из распределения Вейбулла с параметрами $\alpha > 0$ и $\theta > 0$. Найти достаточную статистику для θ, если значение α известно.

10.18. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью $\theta x^{\theta-1}$ при $x \in (0,1)$, где $\theta > 0$. Найти достаточную статистику для параметра θ.

10.19. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Является ли $\overline{X}$ достаточной статистикой параметра p? Доказать, что статистика вида $S = g(n\overline{X})$ не является достаточной, если отображение $g : \{0, \ldots, n\} \to \mathbf{R}$ не взаимно однозначно.

Решение. Пусть $k \in \{0, 1, \ldots, n\}$ и числа $k_1, \ldots, k_n \in \{0, 1\}$ таковы, что $k_1 + \cdots + k_n = k$. Имеем равенства

$$\mathbf{P}\{X_1 = k_1, \ldots, X_n = k_n | n\overline{X} = k\} = \frac{\mathbf{P}\{X_1 = k_1, \ldots, X_n = k_n\}}{\mathbf{P}\{n\overline{X} = k\}} = \frac{p^k(1-p)^{n-k}}{C_n^k p^k (1-p)^{n-k}} = \frac{1}{C_n^k}\,.$$

Таким образом, при условии $n\overline{X} = k$ любой набор из k единиц и $n-k$ нулей имеет одну и ту же вероятность $1/C_n^k$ и не зависит от параметра p.

10.20. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами m и p. Найти условное совместное распределение выборки при условии $X_1 + \cdots + X_n = k$. Является ли $\overline{X}$ достаточной статистикой для параметра p при известном m?

10.21. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Найти условное совместное распределение выборки при условии $X_1 + \cdots + X_n = k$. Является ли $\overline{X}$ достаточной статистикой для параметра λ? Являются ли достаточными статистики $(\overline{X})^2$, $\overline{X^2}$ и $\sin\overline{X}$?

10.22. Является ли статистика $S = n\overline{X} - 5$ достаточной для параметра λ распределения Пуассона? Будут ли достаточными следующие статистики:

а) $2S$; г) $\sin S$;

б) S^2; д) e^S;

в) S/n^2; е) $-S$?

10.23. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Доказать, что статистика вида $S = g(n\overline{X})$ достаточна для параметра λ только в случае, когда отображение $g : \mathbf{Z}_+ \to \mathbf{R}$ взаимно однозначно.

10.24. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p. Найти условное совместное распределение выборки при условии $X_1 + \cdots + X_n = k$. Является ли $\overline{X}$ достаточной статистикой для параметра p?

10.25. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ на множестве целых чисел $\{1, \ldots, m\}$. Доказать, что набор из m статистик

$$\nu(k) = \sum_{i=1}^{n} \mathbf{I}\{X_i = k\}, \quad k = 1, \ldots, m,$$

составляет достаточную для параметра θ статистику.

Решение. Рассмотрим случай $m = 2$. Пусть $\nu(1) = n_1$ и $\nu(2) = n - n_1$. При выполнении этого условия выборка $X_1, \ldots, X_n$ имеет равномерное дискретное распределение на множестве последовательностей из единиц и двоек, содержащих ровно n_1 единиц и $n - n_2$ двоек; это равномерное распределение не зависит от θ.

10.26. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на конечном множестве $\{1, \ldots, \theta\}$, где θ — целый положительный параметр. Найти достаточную для параметра θ статистику.

10.27. Привести пример распределения, зависящего от параметра θ произвольной природы, для которого все достаточные статистики получаются взаимно однозначными преобразованиями вариационного ряда.

10.28. Пусть F_θ, $\theta \in \Theta$, — параметрическое семейство распределений на решётке целых чисел такое, что для некоторого $k \in \mathbf{Z}$

$$\inf_{\theta\in\Theta} F_\theta(k) > 0.$$

Пусть S — достаточная для параметра θ статистика, и статистики S и T независимы при всех значениях θ. Доказать, что распределение статистики T не зависит от θ.

§ 11. Полные статистики

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений и $X_1, \ldots, X_n$ — выборка из распределения F_θ. Статистика $S(X_1, \ldots, X_n)$, построенная по данной выборке, называется *полной*, если функция $\mathbf{E}_\theta g(S)$ переменной θ равна тождественно нулю в том и только в том случае, когда $\mathbf{P}_\theta\{g(S) = 0\} = 1$ при любом значении параметра $\theta \in \Theta$.

11.1. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ, $\theta \in \Theta \subseteq \mathbf{R}^k$, и $S(X_1, \ldots, X_n)$ — некоторая статистика со значениями в $\mathbf{R}^m$. Пусть борелевские функции g_1 и g_2, действующие из $\mathbf{R}^m$ в $\mathbf{R}^k$, таковы, что $g_1(S)$ и $g_2(S)$ имеют одинаковое смещение. Доказать, что если статистика S полна, то $\mathbf{P}_\theta\{g_1(S) = g_2(S)\} = 1$.

11.2. Доказать полноту статистики $\overline{X}$ для выборки из нормального распределения со средним a и дисперсией 1.

Решение. Поскольку статистика $\overline{X}$ имеет нормальное распределение со средним a и дисперсией $1/n$, предположение $\mathbf{E}_a g(\overline{X}) = 0$ для любого действительного числа a означает тождественное равенство нулю интеграла

$$\frac{1}{\sqrt{2\pi/n}} \int\limits_{-\infty}^{\infty} g(x)\, e^{-(x-a)^2 n/2}\, dx \equiv 0,$$

причём интеграл сходится абсолютно. Следовательно, абсолютно сходится и тождественно равен нулю интеграл

$$H(a) \equiv \int\limits_{-\infty}^{\infty} h(x)\, e^{ax}\, dx,$$

где $h(x) = g(x)\, e^{-x^2 n/2}$. Представим функцию h в виде разности положительной и отрицательной частей: $h(x) = h^+(x) - h^-(x)$, где

$$h^+(x) = h(x) \cdot \mathbf{I}\{h(x) > 0\} \geqslant 0 \quad \text{и} \quad h^-(x) = -h(x) \cdot \mathbf{I}\{h(x) < 0\} \geqslant 0.$$

Ввиду равенства $H(a) \equiv 0$ при всех a совпадают значения интегралов

$$H^+(a) = \int\limits_{-\infty}^{\infty} h^+(x)\, e^{ax}\, dx = \int\limits_{-\infty}^{\infty} h^-(x)\, e^{ax}\, dx = H^-(a).$$

Отсюда при $a = 0$ вытекает равенство

$$c = \int\limits_{-\infty}^{\infty} h^+(x)\, dx = \int\limits_{-\infty}^{\infty} h^-(x)\, dx.$$

Поэтому функции $f^+(x) = h^+(x)/c$ и $f^-(x) = h^-(x)/c$ являются плотностями некоторых абсолютно непрерывных распределений F^+ и F^- в $\mathbf{R}$. Таким образом, совпадают следующие преобразования Лапласа:

$$\varphi^+(a) = \int\limits_{-\infty}^{\infty} e^{ax} f^+(x) dx \equiv \int\limits_{-\infty}^{\infty} e^{ax} f^-(x) dx = \varphi^-(a).$$

Рассмотрим аналитическое продолжение $\varphi^+(a)$ и $\varphi^-(a)$ на плоскость комплексного переменного. Функции

$$\varphi^+(a+ib) = \int\limits_{-\infty}^{\infty} e^{(a+ib)x} f^+(x) dx \quad \text{и} \quad \varphi^-(a+ib) = \int\limits_{-\infty}^{\infty} e^{(a+ib)x} f^-(x) dx$$

являются аналитическими на всей комплексной плоскости и совпадают на вещественной прямой. По *внутренней теореме единственности* эти функции совпадают на всей комплексной плоскости и, в частности, на мнимой прямой $a = 0$. Осталось заметить, что $\varphi^+(ib)$ и $\varphi^-(ib)$ суть характеристические функции в точке b распределений F^+ и F^- соответственно. Из совпадения характеристических функций следует равенство плотностей $f^+(x) = f^-(x)$ почти всюду. Следовательно, $h^+(x) = h^-(x)$ почти всюду относительно меры Лебега, и, соответственно, $h(x) = 0$. Поэтому $g(x) = 0$ почти всюду относительно меры Лебега, и статистика $\overline{X}$ является полной.

11.3. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с нулевым средним и дисперсией σ^2. Доказать полноту статистики $\overline{X^2}$ для параметра σ^2.

11.4. Доказать полноту статистики $\overline{X}$ для выборки из показательного распределения с параметром α.

11.5. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\beta \in \mathbf{R}$. Доказать полноту статистики $X_{(1)}$.

11.6. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$.

а) Доказать, что $X_{(1)}$ — полная статистика для β при известном значении α.

б) Доказать, что $\overline{X}$ — полная статистика для α при известном значении β.

в) Привести пример полной статистики для двумерного параметра $\theta = (\alpha, \beta)$.

11.7. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta \in \Theta$. Доказать, что статистика $X_{(n)}$ является полной для параметра θ при $\Theta = (0, \infty)$. Является ли $X_{(n)}$ полной статистикой для θ при $\Theta = (1, \infty)$?

11.8. Доказать полноту статистики $S = \max\{|X_1|, \ldots, |X_n|\}$ для параметра θ равномерного распределения на отрезке $[-\theta, \theta]$.

11.9. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta, \theta{+}1]$. Доказать, что двумерная статистика $(X_{(1)}, X_{(n)})$ не является полной.

Решение. При фиксированном n укажем функцию $g_n : \mathbf{R}^2 \to \mathbf{R}$ такую, что $\mathbf{E}_\theta g_n(X_{(1)}, X_{(n)}) = 0$ для любого $\theta \in \mathbf{R}$, но $\mathbf{P}_\theta\{g_n(X_{(1)}, X_{(n)}) = 0\} \neq 1$ (даже равно нулю).

Для этого найдём $\mathbf{E}_\theta X_{(1)} = \theta + 1/(n+1)$ и $\mathbf{E}_\theta X_{(n)} = \theta + 1 - 1/(n+1)$. Искомой функцией может служить $g_n(x, y) = y - x + 1 + 2/(n+1)$.

11.10. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta, 2\theta]$. Доказать, что двумерная статистика $(X_{(1)}, X_{(n)})$ не является полной.

11.11. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью $\theta y^{\theta-1}$ при $y \in (0, 1)$, где $\theta > 0$. Доказать, что статистика $\overline{\ln X}$ является полной для параметра θ.

11.12. Доказать полноту статистики $\overline{X}$ для выборки из распределения Бернулли с параметром p.

Р е ш е н и е. Величина $n\overline{X}$ имеет биномиальное распределение с параметрами n и p. Поэтому

$$\mathbf{E}_p g(\overline{X}) = \sum_{k=0}^{n} g(k/n) C_n^k p^k (1-p)^{n-k} = (1-p)^n \sum_{k=0}^{n} g(k/n) C_n^k \Big(\frac{p}{1-p}\Big)^k.$$

Сумма $\sum_{k=0}^{n} g(k/n) C_n^k x^k$ является полиномом степени не выше n по переменной $x = p/(1-p)$. Предположение $\mathbf{E}_p g(\overline{X}) = 0$ для любого $p \in (0,1)$ означает, что любая точка $x \in (0, \infty)$ является корнем этого полинома. Следовательно, все коэффициенты $g(k/n) C_n^k$ полинома равны нулю. Таким образом $g(k/n) = 0$ при $k = 0, 1, \ldots, n$, и, тем самым, $\overline{X}$ — полная статистика.

11.13. Доказать полноту статистики $\overline{X}$ для выборки из биномиального распределения с параметрами m и p, если значение m известно.

11.14. Доказать полноту статистики $\overline{X}$ для выборки из распределения Пуассона с параметром λ.

11.15. Доказать полноту статистики $\overline{X}$ для выборки из геометрического распределения с параметром p.

11.16. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на конечном множестве $\{1, \ldots, \theta\}$, где θ — целый положительный параметр. Доказать, что статистика $X_{(n)}$ является полной для параметра θ.

11.17. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ, $\theta \in \Theta$. Пусть $\mathbf{E}_\theta X_1 = \theta$. Доказать, что выборка $X_1, \ldots, X_n$ не является полной статистикой для параметра θ.

§ 12. Эффективные оценки

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ и $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ — некоторая оценка параметра θ со смещением $b_n(\theta) = \mathbf{E}_\theta \theta_n^* - \theta$.

Оценка θ_n^* называется *эффективной в классе оценок со смещением* $b_n(\theta)$, если она не хуже в среднеквадратическом смысле любой другой оценки с тем же смещением $b_n(\theta)$. Справедлива следующая

Теорема. *Пусть $S = S(X_1 \ldots, X_n)$ — достаточная полная статистика для параметра θ. Тогда оценка $\mathbf{E}\{\theta_n^* | S\}$ является единственной эффективной оценкой в классе оценок со смещением $b_n(\theta)$.*

12.1. Пусть θ^* — эффективная оценка в классе оценок со смещением равным $\alpha\theta$, α — постоянная. Построить эффективную оценку в классе несмещённых оценок.

12.2. Пусть $X_1, \ldots, X_n$ — выборка из распределения с конечным первым моментом. Найти условное математическое ожидание $\mathbf{E}(X_1 \,|\, \overline{X})$.

Решение. Элементы выборки независимы и одинаково распределены. Поэтому распределение пары $(X_i, \overline{X})$ не зависит от $i \in \{1, \ldots, n\}$. Следовательно, $\mathbf{E}(X_1 \,|\, \overline{X}) = \mathbf{E}(X_2 \,|\, \overline{X}) = \cdots = \mathbf{E}(X_n \,|\, \overline{X})$. Суммируя, получим

$$\mathbf{E}\{X_1 \,|\, \overline{X}\} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\{X_i \,|\, \overline{X}\} = \mathbf{E}\{\overline{X} \,|\, \overline{X}\} = \overline{X}.$$

12.3. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией. Улучшить оценку $a^* = X_1$ усреднением при фиксированном значении достаточной статистики $\overline{X}$. Найти распределение, математическое ожидание, смещение и дисперсию улучшенной оценки. Является ли улучшенная оценка эффективной?

12.4. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найти эффективную несмещённую оценку неизвестного параметра θ усреднением оценки $\frac{n+1}{n} X_{(n)}$ по статистике $X_{(n)}$.

12.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найти смещение и дисперсию оценки $\theta^* = 2X_1$ неизвестного параметра θ. Улучшить эту оценку усреднением при фиксированном значении полной и достаточной статистики $X_{(n)}$. Найти смещение и дисперсию улучшенной оценки. Является ли улучшенная оценка эффективной?

Решение. Оценка $2X_1$ является несмещённой оценкой параметра θ, а статистика $X_{(n)}$ — достаточной и полной. При условии $X_{(n)} = u$ величина X_1 с вероятностью $1/n$ совпадает с $X_{(n)}$ и, следовательно, равна u. С вероятностью же $(n-1)/n$ величина X_1 не совпадает с $X_{(n)}$ и имеет равномерное распределение на отрезке $[0, u]$. Поэтому среднее значение X_1 при условии $X_{(n)} = u$ равно

$$\frac{u}{n} + \frac{n-1}{n} \frac{u}{2} = \frac{n+1}{2n} u.$$

Таким образом, оценка

$$\mathbf{E}\{2X_1|X_{(n)}\} = \frac{n+1}{n}X_{(n)}$$

является эффективной в классе несмещённых оценок.

12.6. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Найти эффективную несмещённую оценку неизвестного параметра $\tau(\theta, y) = \mathbf{P}_\theta\{X_1 \geqslant y\}$.

12.7. Найти эффективную несмещённую оценку для параметра α показательного распределения по выборке объёма $n \geqslant 2$.

12.8. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

Найти эффективную несмещённую оценку для параметра $\beta \in \mathbf{R}$.

12.9. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$. Найти эффективную несмещённую оценку для

а) параметра β, если значение α известно;

б) параметра α, если значение β известно;

в) двумерного параметра $\theta = (\alpha, \beta)$.

12.10. Пусть $X_1, \ldots, X_n$ — выборка из распределения Парето с параметрами β и θ, причём значение β известно. Найти эффективную несмещённую оценку параметра θ.

12.11. Пусть $X_1, \ldots, X_n$ — выборка из распределения Вейбулла с параметрами α и θ, причём значение α известно. Проверить, что $\overline{X^\alpha}$ является полной и достаточной статистикой для параметра θ. Построить эффективную оценку параметра $\tau(\theta) = 1/\theta$.

12.12. Распределение Кэптейна определяется плотностью

$$f_\theta(y) = \frac{g'(y)}{\sigma\sqrt{2\pi}}\; e^{-(\theta - g(y))^2/2\sigma^2},$$

где $g(y)$ — неубывающая дифференцируемая функция. Найти эффективную несмещённую оценку для

а) параметра θ, если значение σ известно;

б) параметра σ^2, если значение θ известно.

12.13. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью $\theta y^{\theta-1}$ при $y \in (0,1)$, где $\theta > 0$. Найти эффективную несмещённую оценку параметра $\tau(\theta) = 1/\theta$.

12.14. Найти эффективную оценку параметра p распределения Бернулли усреднением какой-либо несмещённой оценки по статистике $\overline{X}$.

Решение. Возьмём несмещённую оценку $p^* = X_1$ и вычислим $\mathbf{E}\{p^* \,|\, \overline{X}\}$. Из задачи 12.2 следует, что $p^{**} = \mathbf{E}\{p^* \,|\, \overline{X}\} = \mathbf{E}\{X_1 \,|\, \overline{X}\} = \overline{X}$. Полученная оценка является единственной эффективной оценкой в классе несмещённых оценок, поскольку статистика $\overline{X}$ является полной и достаточной для параметра p распределения Бернулли.

12.15. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами m и p при известном m. Найти смещение и дисперсию оценки

а) $p_n^* = X_1$; б) $p_n^* = X_1/m$

неизвестного параметра p. Улучшить эту оценку усреднением при фиксированном значении достаточной статистики $\overline{X}$. Найти смещение и дисперсию улучшенной оценки. Является ли улучшенная оценка эффективной?

12.16. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Улучшить оценку $\lambda_n^* = X_1$ усреднением при фиксированном значении достаточной статистики $\overline{X}$. Найти смещение и дисперсию улучшенной оценки. Является ли улучшенная оценка эффективной?

12.17. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. В качестве оценки параметра $\theta = e^{-\lambda} = \mathbf{P}_\lambda\{X_1 = 0\}$ рассматривается $\theta_n^* = \mathbf{I}\{X_1 = 0\}$. Вычислить смещение $b_n(\theta) = \mathbf{E}_\lambda \theta_n^* - \theta$ этой оценки и построить эффективную оценку в классе оценок со смещением $b_n(\theta)$ усреднением по полной и достаточной для параметра θ статистике.

Решение. Имеем $b_n(\theta) = 0$. Статистика $n\overline{X}$ является полной и достаточной. Заметим, что θ_n^* принимает значения 0 и 1. Поэтому

$$\begin{aligned}\mathbf{E}_\lambda\{\theta_n^* \,|\, n\overline{X}=k\} &= 0\cdot\mathbf{P}_\lambda\{\theta_n^*=0\,|\,n\overline{X}=k\}+1\cdot\mathbf{P}_\lambda\{\theta_n^*=1\,|\,n\overline{X}=k\}\\ &= \mathbf{P}_\lambda\{X_1=0\,|\,n\overline{X}=k\}.\end{aligned}$$

Вычислив последнюю вероятность по определению условной вероятности, получим $\mathbf{E}_\lambda\{\theta_n^*\,|\,n\overline{X}=k\}=(1-1/n)^k$. Оценка $\theta_n^{**}=(1-1/n)^{n\overline{X}}$ эффективна в классе несмещённых оценок.

12.18. Пусть $X_1, \ldots, X_n$, $n \geqslant 2$, — выборка из геометрического распределения с параметром $p \in (0,1)$.

а) Доказать, что статистика $S = n\overline{X}$ имеет распределение $\mathbf{P}_p\{S=k\}=C_{n+k-1}^k p^n(1-p)^k$ при $k=0,\,1,\,\ldots$.

б) Доказать, что статистика $S=n\overline{X}$ является достаточной и полной статистикой.

в) Найти смещение оценки $p_n^*=\mathbf{I}\{X_1=0\}$. Используя а) и б), построить эффективную оценку в классе с таким смещением.

12.19. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на конечном множестве $\{1,\ldots,\theta\}$, где θ — целый положительный параметр. Доказать, что статистика

$$\frac{X_{(n)}^{n+1}-(X_{(n)}-1)^{n+1}}{X_{(n)}^n-(X_{(n)}-1)^n}$$

является эффективной оценкой параметра θ в классе несмещённых оценок.

12.20. Пусть θ_1^* и θ_2^* — две несмещённые эффективные оценки параметра θ. Доказать, что $\theta_1^*=\theta_2^*$ с вероятностью 1. *Указание*: рассмотреть оценку $(\theta_1^*+\theta_2^*)/2$.

§ 13. Неравенство Рао – Крамера

Пусть $\{F_\theta,\ \theta\in\Theta\}$ — некоторое параметрическое семейство распределений, удовлетворяющее условию доминирования относительно некоторой меры μ на $\mathbf{R}$, т. е. это параметрическое семейство состоит из распределений, абсолютно непрерывных относительно μ. Плотность распределения F_θ относительно меры μ обозначим через

$$f_\theta(x) = \frac{dF_\theta}{d\mu}(x).$$

Пусть X_1, X_2, …— выборка из распределения F_θ и $\theta_n^* = \theta_n^*(X_1, \ldots, X_n)$ — некоторая оценка параметра θ со смещением $b_n(\theta) = \mathbf{E}_\theta \theta_n^* - \theta$. Справедлива

Теорема (неравенство Рао — Крамера). *Пусть выполнены следующие условия регулярности: для почти всех (по мере μ) значений y функция $\sqrt{f_\theta(y)}$ непрерывно дифференцируема по θ и информация Фишера*

$$I(\theta) \equiv \mathbf{E}_\theta \Big(\frac{\partial \ln f_\theta(X_1)}{\partial \theta} \Big)^2$$

положительна и непрерывна по θ. Тогда для любых $\theta \in \Theta$ и $n \geqslant 1$ справедливо неравенство

$$\mathbf{E}_\theta (\theta_n^* - \theta)^2 \geqslant \frac{(1 + b_n'(\theta))^2}{nI(\theta)} + b_n^2(\theta).$$

Оценка θ_n^* называется *R-эффективной в классе оценок со смещением* $b_n(\theta)$, если для неё достигается нижняя граница в неравенстве Рао — Крамера. R-эффективная оценка в классе оценок со смещением $b_n(\theta)$ с необходимостью является эффективной в этом же классе.

13.1. Объяснить на качественном уровне присутствие выражения $(1 + b'(\theta))^2$ в общей форме неравенства Рао – Крамера. В процессе этого:

а) объяснить, почему появляется $b'(\cdot)$, а не $b(\cdot)$;

б) объяснить, почему граница должна обращаться в нуль, когда $b'(\cdot) = -1$;

в) объяснить, почему упомянутое выше выражение возводится в квадрат, а не в первую степень.

13.2. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ, $\theta \in \Theta$. Доказать, что если оценка θ_n^* является R-эффективной оценкой для θ в классе оценок со смещением $b_n(\theta) = \theta/n$, то она состоятельна. Построить эффективную оценку в классе несмещённых оценок.

13.3. Привести пример состоятельной оценки, которая не является R-эффективной.

13.4. Для всякого ли параметрического семейства распределений найдется $c > 0$ такое, что для любой несмещённой оценки θ_n^* неизвестного параметра θ выполняется неравенство $\mathbf{D}\theta_n^* \geqslant c/n$?

13.5. Выполнены ли условия регулярности для следующих семейств распределений, зависящих от параметра θ:

а) нормальное со средним θ и дисперсией θ^2, $\theta > 0$;

б) равномерное на отрезке $[\theta, \theta + 1]$;

в) равномерное на отрезке $[-\theta, 0]$, $\theta > 0$;

г) с плотностью $\theta e^{-\theta y}$ при $y > 0$;

д) с плотностью $e^{\theta + y}$ при $y < -\theta$;

е) биномиальное с параметрами 5 и θ, $0 < \theta < 1$;

ж) Пуассона с параметром θ, $\theta > 0$;

з) с функцией распределения $F_\theta(y) = 1 - \theta/y$ при $y \geqslant \theta$, $\theta > 1$;

и) с плотностью $4(\theta - y)^3/\theta^4$ на отрезке $[0, \theta]$?

13.6. Проверить, является ли R-эффективной оценка максимального правдоподобия среднего значения a нормального распределения.

Решение. Среднеквадратическое отклонение несмещённой оценки $\overline{X}$ от параметра a равно σ^2/n. Вычислим информацию Фишера:

$$I(a) = \mathbf{E}_a\Big(\frac{\partial}{\partial a}\ln\Big(\frac{1}{\sqrt{2\pi}}\, e^{-(X_1-a)^2/2\sigma^2}\Big)\Big)^2$$

$$= \mathbf{E}_a\Big(\frac{\partial}{\partial a}(X_1 - a)^2/2\sigma^2\Big)^2 = \mathbf{E}_a(X_1 - a)^2/\sigma^4 = 1/\sigma^2.$$

Следовательно, правая часть неравенства Рао – Крамера имеет вид σ^2/n и совпадает со среднеквадратическим отклонением оценки $\overline{X}$. Оценка $\overline{X}$ является R-эффективной.

13.7. Проверить, является ли R-эффективной

а) оценка максимального правдоподобия;

б) оценка S_0^2

дисперсии σ^2 нормального распределения с нулевым средним.

13.8. Пусть $X_1, \ldots, X_{3n}$ — выборка объёма $3n$ из нормального распределения со средним a и единичной дисперсией. Являются ли R-эффективными (эффективными) следующие оценки параметра a:

а) $\displaystyle\frac{1}{n}\sum_{i=n+1}^{2n} X_i$; б) $\displaystyle\frac{1}{n}\sum_{i=1}^{n} X_{3i}$; в) $\displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i$?

13.9. Пусть $X_1, \ldots, X_n$ — выборка из распределения, являющегося смесью двух нормальных распределений, а именно, 92%

составляет нормальное распределение со средним a и дисперсией 1, а 8% составляет нормальное распределение с тем же средним a и дисперсией 16. Является ли выборочное среднее R-эффективной оценкой параметра a?

13.10. Пусть $X_1, \ldots, X_n$, $n \geqslant 2$, — выборка из показательного распределения с параметром α. Будет ли R-эффективной оценка

$$\alpha_n^* = \frac{n-1}{n\overline{X}}?$$

Будет ли эта оценка эффективной?

13.11. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$. Пусть β известно. Является ли R-эффективной оценка метода моментов для параметра α? Эффективной?

13.12. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью из задачи 13.11. Пусть α известно. Является ли R-эффективной оценка метода моментов для параметра β? Эффективной? Является ли R-эффективной оценка максимального правдоподобия для параметра β? Эффективной?

13.13. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,\ \theta]$. Будет ли R-эффективной оценка

$$\theta^* = \frac{n+1}{n}\, X_{(n)}?$$

Будет ли эта оценка эффективной?

13.14. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[\theta,\ \theta+1]$. Будет ли R-эффективной оценка

$$\theta^* = X_{(1)} - (n+1)^{-1}?$$

13.15. Пусть $X_1, \ldots, X_n$ — выборка из логистического распределения с плотностью

$$f_\theta(y) = \frac{e^{\theta-y}}{(1+e^{\theta-y})^2}, \quad y \in \mathbf{R}.$$

а) Проверить, что $\overline{X}$ — несмещённая оценка параметра θ.

б) Найти среднеквадратическое отклонение оценки $\overline{X}$ от параметра θ. *Указание*: использовать равенство

$$\int_0^\infty \frac{y}{1+e^y}\,dy = \frac{\pi^2}{12}\,.$$

в) Найти информацию Фишера.

г) Проверить R-эффективность оценки $\overline{X}$.

13.16. Пусть $X_1, \ldots, X_n$ — выборка из распределения с плотностью $\theta y^{\theta-1}$, $y \in [0,1]$, где $\theta > 0$. Доказать, что оценка $-\overline{\ln X}$ является R-эффективной для $\tau = 1/\theta$ в классе несмещённых оценок.

13.17. Пусть $X_1, \ldots, X_n$ — выборка из распределения Вейбулла с параметрами α и θ, причём значение α известно. Доказать, что оценка $\overline{X^\alpha}$ является R-эффективной для $\tau = 1/\theta$ в классе несмещённых оценок.

13.18. Пусть $X_1, \ldots, X_n$ — выборка из распределения Коши с параметром сдвига a. Исследовать R-эффективность выборочной медианы как оценки параметра a.

Р е ш е н и е. Из задачи 7.36 следует, что выборочная медиана является асимптотически нормальной оценкой для медианы a распределения Коши с коэффициентом асимптотической нормальности $\pi^2/4$. По лемме Фату

$$\liminf_{n\to\infty}\ n\mathbf{D}\zeta^* \geqslant \pi^2/4.$$

Информация Фишера равна

$$I(a) = \frac{8}{\pi}\int_0^\infty \frac{t^2}{(1+t^2)^3}\,dt = 1/2.$$

Так как $\pi^2/4 > 2$, то ζ^* не является R-эффективной оценкой, по крайней мере, при достаточно больших значениях n.

13.19. Пусть F — распределение с нулевым средним значением и плотностью $f(y)$. Пусть $f(y)$ — дифференцируемая чётная функция. Рассматривается распределение F_θ с плотностью $f(y-\theta)$, $\theta \in \mathbf{R}$. Доказать, что выборочная медиана не может быть R-эффективной оценкой параметра сдвига θ.

13.20. Проверить, является ли R-эффективной оценка максимального правдоподобия параметра p распределения Бернулли.

Решение. Среднеквадратическое отклонение несмещённой оценки $\overline{X}$ от параметра p равно $p(1-p)/n$. Вычислим информацию Фишера

$$\begin{aligned}
I(p) &= \mathbf{E}_p\Big(\frac{\partial}{\partial p}\ln p^{X_1}(1-p)^{1-X_1}\Big)^2 \\
&= \mathbf{E}_p\Big(\frac{\partial}{\partial p}\big(X_1\ln p + (1-X_1)\ln(1-p)\big)\Big)^2 \\
&= \frac{1}{p^2}\mathbf{E}_p X_1 + \frac{1}{(1-p)^2}\mathbf{E}_p(1-X_1) = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.
\end{aligned}$$

Правая часть неравенства Рао – Крамера имеет вид $p(1-p)/n$ и совпадает со среднеквадратическим отклонением оценки $\overline{X}$. Следовательно, оценка $\overline{X}$ является R-эффективной.

13.21. Проверить, является ли R-эффективной оценка максимального правдоподобия параметра p биномиального распределения с параметрами m и p, если значение m известно.

13.22. Проверить, является ли R-эффективной оценка максимального правдоподобия параметра λ распределения Пуассона.

13.23. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. В качестве оценки параметра $\theta = e^{-\lambda}$ рассматривается статистика $\theta_n^* = \overline{\mathbf{I}\{X = 0\}}$. Вычислить смещение $b_n(\theta) = \mathbf{E}\theta_n^* - \theta$ этой оценки и проверить, является ли она R-эффективной.

13.24. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p. Является ли R-эффективной оценкой параметра $\tau = 1/p$ оценка $\tau_n^* = 1 + \overline{X}$?

13.25. Пусть $X_1, \ldots, X_n$ — выборка из следующего трёхточечного распределения, зависящего от параметра $\theta \in (0, 1/3)$:

$$\mathbf{P}_\theta\{X_1 = 1\} = \theta, \quad \mathbf{P}_\theta\{X_1 = 2\} = 2\theta, \quad \mathbf{P}_\theta\{X_1 = 3\} = 1 - 3\theta.$$

Проверить R-эффективность оценки максимального правдоподобия параметра θ.

Решение. В задаче 4.28 найдена оценка максимального правдоподобия $\theta_n^* = \overline{Y}/3$, где

$$Y_i = \begin{cases} 1, & \text{если } X_i \neq 3, \\ 0, & \text{если } X_i = 3. \end{cases}$$

Плотность $f_\theta(y)$ относительно считающей (на множестве $\{1,2,3\}$) меры равна

$$f_\theta(y) = \begin{cases} \theta & \text{при } y=1, \\ 2\theta & \text{при } y=2, \\ 1-3\theta & \text{при } y=3. \end{cases}$$

Условия регулярности для этого семейства выполнены. Вычислим информацию Фишера

$$\begin{aligned} I(\theta) &= \mathbf{E}\Big(\frac{\partial}{\partial\theta}\ln f_\theta(X_1)\Big)^2 \\ &= \theta\Big(\frac{\partial}{\partial\theta}\ln\theta\Big)^2 + 2\theta\Big(\frac{\partial}{\partial\theta}\ln 2\theta\Big)^2 + (1-3\theta)\Big(\frac{\partial}{\partial\theta}\ln(1-3\theta)\Big)^2 \\ &= \frac{3}{\theta(1-3\theta)}. \end{aligned}$$

Дисперсия несмещённой оценки θ_n^* равна $\mathbf{D}\theta_n^* = 3\theta(1-3\theta)/9n$. В неравенстве Рао – Крамера достигается равенство. Поэтому оценка θ^* является R-эффективной и, следовательно, эффективной.

13.26. Семейство распределений $\{F_\theta, \theta\in\Theta\}$ называется экспоненциальным, если функция правдоподобия $f_\theta(X_1,\ldots,X_n)$ допускает представление

$$f_\theta(X_1,\ldots,X_n) = e^{A(\theta)T(X_1,\ldots,X_n)+B(\theta)}h(X_1,\ldots,X_n).$$

Являются ли экспоненциальными семейства

а) нормальных распределений с параметрами a и σ^2, если значение σ^2 известно;

б) нормальных распределений с параметрами a и σ^2, если значение a известно;

в) Γ-распределений с параметрами α и λ, если значение λ известно;

г) Γ-распределений с параметрами α и λ, если значение α известно;

д) распределений Бернулли с параметром p;

е) распределений Пуассона с параметром λ;

ж) равномерных распределений на отрезке $[a,b]$?

13.27. Пусть $X_1, \ldots, X_n$ — выборка из экспоненциального семейства, причём функции $A(\theta)$ и $B(\theta)$ непрерывно дифференцируемы. Доказать, что для оценки $\theta_n^* = T(X_1,\ldots,X_n)$ в неравенстве Рао – Крамера достигается равенство.

ОТДЕЛ V

ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ

§ 14. Доверительные интервалы

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений, $\Theta \subseteq \mathbf{R}$, и $X_1, \ldots, X_n$ — выборка из распределения F_θ.

Пусть $\theta_n^- = \theta_n^-(X_1, \ldots, X_n)$ и $\theta_n^+ = \theta_n^+(X_1, \ldots, X_n)$ — некоторые статистики. Случайный интервал (θ_n^-, θ_n^+) называется *доверительным интервалом уровня* $1-\varepsilon$, если

$$\mathbf{P}_\theta\{\theta \in (\theta_n^-, \theta_n^+)\} \geqslant 1-\varepsilon.$$

Случайный интервал (θ_n^-, θ_n^+) называется *точным доверительным интервалом уровня* $1-\varepsilon$, если при всех θ

$$\mathbf{P}_\theta\{\theta \in (\theta_n^-, \theta_n^+)\} = 1-\varepsilon.$$

Для построения точного доверительного интервала обычно используется следующий подход. Выбирается функция $G(x_1, \ldots, x_n, \theta)$ такая, что распределение $\mathbf{P}_\theta\{G(X_1, \ldots, X_n, \theta) \in \cdot\}$ не зависит от параметра θ (распределение *свободно* от параметра θ). Функция G должна быть монотонной и обратимой функцией аргумента θ при любых фиксированных значениях выборки X_1, $\ldots$, X_n. Пусть, для определённости, функция G возрастает. Обозначим через $t(X_1, \ldots, X_n, y)$ функцию, обратную к функции $G(X_1, \ldots, X_n, \theta)$ по параметру θ. Тогда доверительный интервал уровня $1-\varepsilon$ имеет вид

$$\big(t(X_1, \ldots, X_n, y^-),\ t(X_1, \ldots, X_n, y^+)\big),$$

где числа y^- и y^+ находятся (вообще говоря, неоднозначно) из уравнения

$$\mathbf{P}_\theta\{y^- < G(X_1, \ldots, X_n, \theta) < y^+\} = 1-\varepsilon.$$

14.1. Пусть X_1, $\ldots$, X_n — выборка из нормального распределения со средним a и дисперсией σ^2, причём значение σ^2 известно. Построить точный доверительный интервал для a.

14.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2, причём значение a известно. Построить точный доверительный интервал для σ^2, используя статистику $S_1^2 = \overline{(X-a)^2}$.

14.3. В условиях предыдущей задачи построить точный доверительный интервал для σ^2, используя статистику $|\overline{X} - a|$. Какой из полученных доверительных интервалов следует предпочесть?

Решение. Случайная величина $\sqrt{n}\,|\overline{X}-a|/\sqrt{\sigma^2}$ распределена как $|\xi|$, где ξ имеет стандартное нормальное распределение. Пусть ζ_δ — квантиль уровня δ стандартного нормального распределения. Тогда

$$\mathbf{P}\left\{\zeta_{0,5+\varepsilon/4} < |\xi| < \zeta_{1-\varepsilon/4}\right\} = 1 - \varepsilon$$

и искомый точный доверительный интервал уровня $1-\varepsilon$ находится из соотношений

$$\mathbf{P}\left\{\zeta_{0,5+\varepsilon/4} < \sqrt{n}\,\frac{|\overline{X}-a|}{\sqrt{\sigma^2}} < \zeta_{1-\varepsilon/4}\right\} = \mathbf{P}\left\{\frac{n(\overline{X}-a)^2}{\zeta^2_{1-\varepsilon/4}} < \sigma^2 < \frac{n(\overline{X}-a)^2}{\zeta^2_{0,5+\varepsilon/4}}\right\}.$$

Распределение левой и правой границ полученного интервала

$$\left(\frac{n(\overline{X}-a)^2}{\zeta^2_{1-\varepsilon/4}},\ \frac{n(\overline{X}-a)^2}{\zeta^2_{0,5+\varepsilon/4}}\right) = \left(\frac{\sigma^2\xi^2}{\zeta^2_{1-\varepsilon/4}},\ \frac{\sigma^2\xi^2}{\zeta^2_{0,5+\varepsilon/4}}\right)$$

не зависит от n. Поэтому точный доверительный интервал, полученный в предыдущей задаче, предпочтительнее — его длина почти наверное стремится к нулю с ростом n.

Действительно, пусть λ_δ есть квантиль уровня δ распределения χ^2 с n степенями свободы. Тогда интервал $\left(nS_1^2/\lambda_{1-\varepsilon/2},\, nS_1^2/\lambda_{\varepsilon/2}\right)$ является точным доверительным интервалом для σ^2 уровня доверия $1-\varepsilon$. Согласно центральной предельной теореме $\lambda_\delta = n + \zeta_\delta\sqrt{n} + o(\sqrt{n})$ и обе границы интервала стремятся к σ^2 с ростом n.

14.4. Пусть X_1, X_2 — выборка объёма 2 из нормального распределения со средним 2 и дисперсией 3. Указать число c такое, что случайные величины $X_1 - cX_2$ и $X_1 + X_2$ независимы.

14.5. Пусть X_1, X_2 — выборка объёма 2 из нормального распределения со средним 1 и дисперсией 2. Обозначим $S_1 = X_1 + X_2$ и $S_2 = X_1^2 + X_2^2$. Указать число c такое, что случайные величины S_1 и $cS_2 - S_1^2$ независимы.

14.6. Пусть X_1, X_2 — выборка объёма 2 из нормального распределения со средним 0 и дисперсией 5. Указать число c такое, что величины $|X_1 - 2X_2|$ и $(cX_1 + X_2)^3$ независимы.

14.7. По выборке из нормального распределения построить точные доверительные интервалы для среднего a и дисперсии σ^2.

14.8. По выборке из нормального распределения со средним $\theta > 0$ и дисперсией θ^2 построить точный доверительный интервал для параметра θ уровня доверия $1 - \varepsilon$.

Решение. Величина $\sqrt{n}|\overline{X}|/\theta$ распределена как $|\xi|$, где ξ имеет нормальное распределение со средним $\sqrt{n}$ и единичной дисперсией. Пусть ζ_δ — квантиль уровня δ распределения случайной величины $|\xi|$, т. е. $\mathbf{P}\{|\xi| < \zeta_\delta\} = \delta$. Тогда искомый доверительный интервал равен $(\sqrt{n}|\overline{X}|/\zeta_{1-\varepsilon/2},\ \sqrt{n}|\overline{X}|/\zeta_{\varepsilon/2})$.

14.9. Имеется выборка объёма 3 из нормального распределения с неизвестными математическим ожиданием и дисперсией. Несмещённая выборочная дисперсия равна 1. Не пользуясь таблицами, построить точный доверительный интервал для неизвестной дисперсии уровня 0,9.

14.10. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$, где $\theta \in (0, 1]$. Используя неравенство Чебышёва, построить доверительный интервал для θ с помощью

а) оценки $2\overline{X}$; б) оценки $X_{(n)}$.

14.11. С помощью статистики X_1 по выборке объёма 1 из равномерного распределения на отрезке $[0, \theta]$ построить точный доверительный интервал уровня $1 - \varepsilon$ для параметра θ.

14.12. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. С помощью статистики $X_{(n)}$ построить точный доверительный интервал уровня $1 - \varepsilon$ для параметра θ.

Решение. Пусть $Y_i = X_i/\theta$, $i = 1, \ldots, n$, — элементы выборки объёма n из равномерного распределения на отрезке $[0, 1]$. Распределение случайной величины $Y_{(n)} = X_{(n)}/\theta$ не зависит от θ. Найдём $\psi \in (0, 1)$ такое, что $\mathbf{P}\{\psi < Y_{(n)} < 1\} = 1 - \varepsilon$. Функция распределения максимальной порядковой статистики $Y_{(n)}$ равна $F(y) = y^n$ для $0 < y < 1$. Поэтому $1 - \psi^n = 1 - \varepsilon$ и, соответственно, $\psi = \sqrt[n]{\varepsilon}$.

Доверительный интервал для θ получим из соотношений

$$1 - \varepsilon = \mathbf{P}\{\psi < X_{(n)}/\theta < 1\} = \mathbf{P}\{X_{(n)} < \theta < X_{(n)}/\psi\}.$$

Искомый доверительный интервал равен $(X_{(n)},\ X_{(n)}/\sqrt[n]{\varepsilon})$.

14.13. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения на отрезке $[0, \theta]$. Показать, что в качестве точного доверительного интервала уровня $1 - \varepsilon$ можно взять интервал $(X_{(n-1)},\ X_{(n-1)}/\psi)$, где ψ находится из уравнения

$$\psi^{n-1}(n - (n-1)\psi) = \varepsilon.$$

14.14. С помощью оценки $X_{(1)}$ построить точный доверительный интервал для параметра θ по выборке объёма n из

а) равномерного распределения на отрезке $[\theta, \theta + 1]$;

б) равномерного распределения на отрезке $[\theta, 2\theta]$.

14.15. С помощью оценки $X_{(1)}$ по выборке объёма n из смещённого показательного распределения с параметром сдвига β построить точный доверительный интервал для параметра β.

14.16. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Построить точные доверительные интервалы для параметра α, используя статистики $S_1(\vec{X}) = X_1$ и $S_2(\vec{X}) = X_{(1)}$.

§ 15. Асимптотические доверительные интервалы

Случайный интервал (θ_n^-, θ_n^+) называется *асимптотическим доверительным интервалом уровня* $1 - \varepsilon$, если при всех θ

$$\liminf_{n\to\infty} \mathbf{P}_\theta\{\theta \in (\theta_n^-, \theta_n^+)\} \geqslant 1 - \varepsilon.$$

Случайный интервал (θ_n^-, θ_n^+) называется *асимптотически точным доверительным интервалом уровня* $1 - \varepsilon$, если при всех θ

$$\lim_{n\to\infty} \mathbf{P}_\theta\{\theta \in (\theta_n^-, \theta_n^+)\} = 1 - \varepsilon.$$

15.1. С помощью оценки $\overline{X}$ построить асимптотический доверительный интервал уровня $1 - \varepsilon$ для неизвестного параметра p распределения Бернулли.

Р е ш е н и е. По центральной предельной теореме распределение случайной величины

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}$$

слабо сходится к стандартному нормальному закону, а $\overline{X}$ сходится по вероятности к p. Поэтому

$$\frac{\sqrt{n}(\overline{X}-p)}{\sqrt{\overline{X}(1-\overline{X})}}$$

слабо сходится также к стандартному нормальному закону. Следовательно, случайный интервал

$$\left(\overline{X}-\frac{\zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}},\ \overline{X}+\frac{\zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}}\right)$$

является асимптотическим доверительным интервалом уровня $1-\varepsilon$, если $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ стандартного нормального распределения.

15.2. В результате проверки 400 электрических лампочек 40 штук оказалось бракованными. Найти доверительный интервал уровня 0,99 для вероятности брака.

15.3. С помощью оценки $\overline{X}$ построить асимптотический доверительный интервал уровня $1-\varepsilon$ для неизвестного параметра p биномиального распределения (значение параметра m известно).

15.4. С помощью статистики $\overline{X}$ построить асимптотический доверительный интервал уровня $1-\varepsilon$ для параметра λ распределения Пуассона.

15.5. С помощью статистики $\overline{X}$ построить асимптотический доверительный интервал уровня $1-\varepsilon$ для параметра p геометрического распределения.

15.6. Пусть $X_1, \ldots, X_n$ — выборка из распределения F_θ с конечной дисперсией, $\mathbf{E}_\theta X_1=\theta$ и $\mathbf{D}_\theta X_1=\sigma^2(\theta)$, где $\sigma(\theta)$ — непрерывная по θ функция. С помощью оценки $\overline{X}$ построить асимптотический доверительный интервал для θ уровня $1-\varepsilon$.

15.7. Пусть θ_n^* — асимптотически нормальная оценка параметра θ с коэффициентом $\sigma^2(\theta)$, где $\sigma(\theta)$ — непрерывная по θ функция. С помощью оценки θ_n^* построить асимптотический доверительный интервал для θ уровня $1-\varepsilon$.

15.8. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0,\theta]$. Используя результат задачи 1.28, построить асимптотический доверительный интервал для θ с помощью оценки $X_{(n)}$.

15.9. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. С помощью асимптотически нормальных оценок $\theta_1^* = 2\overline{X}$ и $\theta_2^* = \sqrt{3\overline{X^2}}$ построить асимптотические доверительные интервалы для параметра θ уровня $1-\varepsilon$ и показать, что второй интервал асимптотически короче первого.

15.10. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. С помощью асимптотически нормальных оценок $\alpha_1^* = 1/\overline{X}$ и $\alpha_2^* = \sqrt{2/\overline{X^2}}$ построить асимптотические доверительные интервалы для параметра α уровня $1-\varepsilon$ и показать, что первый интервал короче второго.

15.11. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с параметром сдвига β. С помощью статистики $\overline{X}$ построить асимптотический доверительный интервал для параметра β уровня $1-\varepsilon$. Сравнить его с точным доверительным интервалом из задачи 14.15. Какой из интервалов следует предпочесть?

15.12. Пусть дана выборка из распределения Парето с параметрами β и θ. Пользуясь результатами задачи 7.22, построить асимптотический доверительный интервал для β.

15.13. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2, причём значение σ^2 известно. Построить асимптотический доверительный интервал для a, используя выборочную медиану. Сравнить полученный интервал с точным доверительным интервалом, построенным по выборочному среднему значению.

15.14. Пусть $X_1, \ldots, X_n$ — выборка из распределения Коши с параметром сдвига a. Построить асимптотический доверительный интервал для a, используя выборочную медиану.

15.15. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и дисперсией σ^2, причём значение a известно. Построить асимптотический доверительный интервал для σ^2, используя статистику $\sqrt{\pi/2} \cdot \overline{|X-a|}$. Сравнить полученный интервал с точным доверительным интервалом, построенным по выборочной дисперсии.

Р е ш е н и е. Оценка $\sigma_n^* = \sqrt{\pi/2} \cdot \overline{|X-a|}$ является асимптотически нормальной оценкой для σ с коэффициентом $\sigma^2(\pi/2-1)$ (см. задачу 7.10). Поэтому оценка $(\sigma_n^*)^2 = (\pi/2)(\overline{|X-a|})^2$ является асимптотически нормальной оценкой для σ^2 с коэффициентом $4\sigma^4(\pi/2-1)$. Отсюда получаем следующий асимптотический доверительный интервал для σ^2:

$$\left(\sigma_n^* - \frac{2(\sigma_n^*)^2\zeta_{1-\varepsilon/2}\sqrt{\pi/2-1}}{\sqrt{n}},\ \frac{\sigma_n^* + 2(\sigma_n^*)^2\zeta_{1-\varepsilon/2}\sqrt{\pi/2-1}}{\sqrt{n}}\right),$$

где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ стандартного нормального распределения. Его длина имеет порядок $\frac{4\sigma^2\zeta_{1-\varepsilon/2}\sqrt{\pi/2-1}}{\sqrt{n}}$. Длина точного доверительного интервала равна (см. решение задачи 14.3)

$$\frac{nS_1^2}{n-\sqrt{n}\zeta_{1-\varepsilon/2}+o(\sqrt{n})} - \frac{nS_1^2}{n+\sqrt{n}\zeta_{1-\varepsilon/2}+o(\sqrt{n})},$$

что есть величина порядка $\frac{2\sigma^2\zeta_{1-\varepsilon/2}}{\sqrt{n}}$. Таким образом, точный доверительный интервал асимптотически короче.

О Т Д Е Л VI

ПРОВЕРКА ГИПОТЕЗ

§ 16. Различение двух простых гипотез: основные понятия

Пусть имеется выборка $X_1, \ldots, X_n$. Кроме того, пусть имеются два распределения F_1 и F_2, а также две простые гипотезы H_1 и H_2 о распределении выборки: гипотеза H_j состоит в том, что выборка взята из распределения F_j, $j = 1, 2$. Гипотеза H_1 называется основной, а H_2 — альтернативной.

Нерандомизированным критерием называется произвольное измеримое по Борелю отображение $\delta : \mathbf{R}^n \to \{0, 1\}$. Если $\delta(X_1, \ldots, X_n) = 1$, то основная гипотеза отвергается и принимается альтернатива; если $\delta(X_1, \ldots, X_n) = 0$, то принимается основная гипотеза.

Рандомизированным критерием называется произвольное измеримое по Борелю отображение $\delta : \mathbf{R}^n \to [0, 1]$. Величина $\delta(X_1, \ldots, X_n)$ интерпретируется как вероятность отвергнуть основную гипотезу.

Вероятностью ошибки j-го рода нерандомизированного критерия δ называется вероятность

$$\alpha_j = \mathbf{P}_{H_j}\{\text{гипотеза } H_j \text{ отвергается}\},$$

где вероятность $\mathbf{P}_{H_j}$ вычисляется в предположении, что выборка $X_1, \ldots, X_n$ взята из распределения F_j.

Вероятностью ошибки первого рода рандомизированного критерия δ называется математическое ожидание

$$\alpha_1 = \mathbf{E}_{H_1}\delta(X_1, \ldots, X_n),$$

а *вероятностью ошибки второго рода* — величина

$$\alpha_2 = 1 - \mathbf{E}_{H_2}\delta(X_1, \ldots, X_n).$$

Вероятность ошибки первого рода называется также *размером* критерия и обозначается $\alpha(\delta)$, а $1 - \alpha_2(\delta)$ — *мощностью* критерия и обозначается $\beta(\delta)$.

Критерий называется *состоятельным*, если с ростом объёма выборки мощность критерия стремится к 1.

16.1. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией. Для проверки основной гипотезы $a = 0$ против альтернативы $a = 1$ используется следующий критерий: основная гипотеза принимается, если $X_{(n)} < 3$, и отвергается в противном случае. Найти вероятности ошибок первого и второго рода.

Решение. Имеем равенства

$$\begin{aligned} \alpha_1 &= \mathbf{P}_{H_1}\{\text{гипотеза } H_1 \text{ отвергается}\} = \mathbf{P}_{H_1}\{X_{(n)} \geqslant 3\} \\ &= 1 - \mathbf{P}_{H_1}\{X_{(n)} < 3\} = 1 - \big(\mathbf{P}_{H_1}\{X_1 < 3\}\big)^n = 1 - \big(1 - \overline{\Phi}(3)\big)^n \end{aligned}$$

и

$$\begin{aligned} \alpha_2 &= \mathbf{P}_{H_2}\{\text{гипотеза } H_1 \text{ принимается}\} = \mathbf{P}_{H_2}\{X_{(n)} < 3\} \\ &= \big(\mathbf{P}_{H_2}\{X_1 < 3\}\big)^n = \big(\mathbf{P}_{H_2}\{X_1 - 1 < 2\}\big)^n = \big(1 - \overline{\Phi}(2)\big)^n. \end{aligned}$$

16.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией. Рассматриваются две простые гипотезы: основная $a = -1$ и альтернативная $a = 0$. Предлагается следующий статистический критерий для проверки этих гипотез: основная гипотеза принимается, если $\overline{X} < -n^{\gamma}$; в противном случае принимается альтернативная гипотеза. Здесь γ — заранее выбранное вещественное число. Определить все числа γ, при которых критерий является состоятельным.

16.3. Есть две гипотезы: основная состоит в том, что элементы выборки имеют нормальное распределение, а альтернатива — в том, что элементы выборки имеют распределение Пуассона. Построить критерий, обладающий нулевыми вероятностями ошибок первого и второго рода.

16.4. Пусть $X_1, \ldots, X_n$ — выборка, о плотности распределения которой высказаны две гипотезы: гипотеза H_1 о том, что X_i имеют распределение с плотностью

$$f_1(y) = \begin{cases} e^{-(y-6)} & \text{при } y \geqslant 6, \\ 0 & \text{при } y < 6, \end{cases}$$

и альтернатива H_2, состоящая в том, что X_i имеют плотность

$$f_2(y) = \begin{cases} 2e^{-2(y-3)} & \text{при } y \geqslant 3, \\ 0 & \text{при } y < 3. \end{cases}$$

Найти пределы при $n \to \infty$ вероятностей ошибок первого и второго рода следующего критерия: гипотеза H_1 принимается тогда и только тогда, когда

а) $\overline{X} > 3{,}5 + 1/\sqrt{n}$; б) $\overline{X} > 3{,}5 + 1/n$; в) $\overline{X} > 3{,}5$.

16.5. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Рассматриваются две простые гипотезы: $\lambda = 1$ и $\lambda = 3$. Критерий δ предписывает принимать первую гипотезу, если $X_{(n)} \leqslant 1$, и альтернативу в противном случае. Найти минимальный размер выборки, при котором мощность этого критерия превышает заданное значение γ.

16.6. Основная гипотеза состоит в том, что данный человек лишён телепатических способностей и угадывает мысли на расстоянии в каждом единичном эксперименте с вероятностью $1/2$. Гипотеза же о наличии телепатических способностей у данного человека принимается, если в 100 независимых однотипных экспериментах по угадыванию мыслей на расстоянии не менее 70 заканчиваются успехом. Чему равна вероятность признать телепатом человека без телепатических способностей?

§ 17. Байесовские и минимаксные критерии

Пусть имеется выборка $X_1, \ldots, X_n$. Кроме того, пусть имеются k распределений $F_1, \ldots, F_k$ и k простых гипотез $H_1, \ldots, H_k$ о распределении выборки: гипотеза H_j состоит в том, что выборка взята из распределения F_j, $j = 1, 2, \ldots, k$.

Нерандомизированным критерием называется произвольное отображение $\delta = (\delta_1, \ldots, \delta_k) : \mathbf{R}^n \to \{0, 1\}^k$ такое, что для любого $(x_1, \ldots, x_n) \in \mathbf{R}^n$ лишь одно (по j) из значений $\delta_j(x_1, \ldots, x_n)$ равно 1, а остальные равны 0. Если $\delta_j = 1$, то принимается гипотеза H_j.

Рандомизированным критерием называется произвольное отображение $\delta = (\delta_1, \ldots, \delta_k) : \mathbf{R}^n \to [0, 1]^k$ такое, что для любого $(x_1, \ldots, x_n) \in \mathbf{R}^n$ имеет место равенство

$$\sum_{i=1}^{k} \delta_i(x_1, \ldots, x_n) = 1.$$

Значение δ_j интерпретируется как вероятность принять гипотезу H_j.

Вероятностью ошибки j-го рода нерандомизированного критерия δ называется вероятность

$$\alpha_j(\delta) = \mathbf{P}_{H_j}\{\delta_j(X_1,\ldots,X_n)=0\}.$$

Вероятностью ошибки j-го рода рандомизированного критерия δ называется математическое ожидание

$$\alpha_j(\delta) = 1-\mathbf{E}_{H_j}\delta_j(X_1,\ldots,X_n).$$

Байесовский подход. Этот подход предполагает, что распределение F_j, из которого извлечена выборка, было выбрано случайно. В этом случае гипотезы H_j становятся случайными событиями; известные вероятности этих событий обозначим через

$$\mathbf{P}\{H_j\}=q_j,$$

так что $q=(q_1,\ldots,q_k)$ есть априорное распределение на множестве гипотез. Определим среднюю вероятность ошибки критерия δ:

$$\alpha(\delta)=\sum_{j=1}^{k} q_j\alpha_j(\delta).$$

Критерий δ_q назавается *байесовским*, если он имеет минимальную среднюю вероятность ошибки.

Пусть выполнено условие доминирования относительно некоторой меры μ на $\mathbf{R}$. Обозначим через $f_j(y)$ плотность распределения F_j относительно меры μ. Тогда справедлива следующая

Теорема. *Байесовский критерий $\delta_q=(\delta_{q,1},\ldots,\delta_{q,k})$ существует при любом априорном распределении q. Он имеет вид: критерий δ_q принимает гипотезу H_j, т. е. $\delta_{q,j}(x_1,\ldots,x_n)=1$, если*

$$q_j f_j(x_1,\ldots,x_n)=\max_{i=1,\ldots,k} q_i f_i(x_1,\ldots,x_n).$$

Минимаксный подход. При этом подходе сравниваются максимальные значения

$$\alpha(\delta)=\max_j\ \alpha_j(\delta).$$

Критерий δ назавается *минимаксным*, если он имеет минимальную ошибку $\alpha(\delta)$.

17.1. По выборке объёма n из нормального распределения с неизвестным средним a и неизвестной дисперсией построить байесовский критерий для различения двух простых гипотез о параметре a, если априорные вероятности гипотез равны.

17.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией, где параметр a может принимать лишь значения

а) 1 и 2; б) 1, 2 и 3

с равными априорными вероятностями. Построить байесовский критерий $\delta = \delta(\overline{X})$. Вычислить $\delta(3)$.

17.3. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения со средним α, где параметр α может принимать лишь значения 1, 2 и 3 с равными априорными вероятностями. Построить байесовский критерий.

17.4. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p, где p может принимать лишь значения $1/2$ и $1/4$ с априорными вероятностями $1/3$ и $2/3$ соответственно. Построить байесовский критерий.

17.5. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами m и p, где p может принимать лишь значения $1/3$ и $2/3$ с априорными вероятностями $1/5$ и $4/5$ соответственно, а параметр m известен и фиксирован. Построить байесовский критерий.

17.6. По выборке объёма 1 из нормального распределения с неизвестным средним a и неизвестной дисперсией построить минимаксный критерий для различения двух простых гипотез о параметре a.

§ 18. Наиболее мощные критерии

Пусть имеются два распределения F_1 и F_2, а также две простые гипотезы H_1 и H_2 о распределении выборки: гипотеза H_j состоит в том, что выборка взята из распределения F_j, $j = 1, 2$.

Наиболее мощным критерием размера ε, различающим гипотезы H_1 и H_2, называется такой критерий δ (вообще говоря, рандомизированный), что $\alpha(\delta) \leqslant \varepsilon$ и любой другой критерий с размером, не превосходящим ε, обладает меньшей мощностью, нежели δ.

Пусть выполнено условие доминирования относительно некоторой меры μ на $\mathbf{R}$. Обозначим плотность распределения F_1 относительно меры μ через $f_1(x)$, а распределения F_2 через $f_2(x)$. Пусть $f_1(x_1, \ldots, x_n)$ — функция правдоподобия при основной гипотезе и $f_2(x_1, \ldots, x_n)$ — при альтернативе. В этом случае справедлива следующая

Лемма Неймана – Пирсона. *Наиболее мощный критерий размера ε существует при любом $\varepsilon > 0$ и определяется равенством*

$$\delta(x_1,\ldots,x_n) = \begin{cases} 1, & \textit{если } \dfrac{f_2(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)} > c, \\ 0, & \textit{если } \dfrac{f_2(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)} < c, \\ \rho, & \textit{если } \dfrac{f_2(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)} = c, \end{cases}$$

где константы c и ρ однозначным образом находятся из уравнения

$$\mathbf{E}_{H_1}\delta(X_1,\ldots,X_n) = \mathbf{P}_{H_1}\left\{\frac{f_2(X_1,\ldots,X_n)}{f_1(X_1,\ldots,X_n)} > c\right\} + \rho\mathbf{P}_{H_1}\left\{\frac{f_2(X_1,\ldots,X_n)}{f_1(X_1,\ldots,X_n)} = c\right\} = \varepsilon.$$

18.1. Пусть имеется некоторая выборка. Основная гипотеза состоит в том, что элементы выборки имеют стандартное нормальное распределение. Альтернатива — в том, что элементы выборки имеют распределение Бернулли с параметром $1/2$. Построить наиболее мощный критерий, различающий эти две гипотезы с вероятностью ошибки первого рода, равной $1/2$.

18.2. По выборке X_1 объёма 1 проверяются гипотезы о плотности f распределения наблюдения X_1: гипотеза $H_1 = \{f = f_1\}$ против альтернативы $H_2 = \{f = f_2\}$. Здесь

$$f_1(y) = \begin{cases} 2y & \text{при } y \in [0,1], \\ 0 & \text{при } y \notin [0,1], \end{cases} \qquad f_2(y) = \begin{cases} 2(1-y) & \text{при } y \in [0,1], \\ 0 & \text{при } y \notin [0,1]. \end{cases}$$

Построить наиболее мощный критерий размера ε и вычислить его мощность.

Р е ш е н и е. Отношение правдоподобия при $n = 1$ равно $1/x_1 - 1$. Поэтому наиболее мощный критерий отвергает основную гипотезу, если $1/x_1 - 1 > c$, что равносильно неравенству $x_1 < c_1$. Число c_1 определяется из равенства

$$\alpha(\delta) = \mathbf{P}_{H_1}\{X_1 < c_1\} = c_1^2 = \varepsilon.$$

Следовательно, $c_1 = \sqrt{\varepsilon}$ и основная гипотеза отвергается, если $X_1 < \sqrt{\varepsilon}$. Мощность этого критерия равна

$$\beta(\delta) = \mathbf{P}_{H_2}\{X_1 < c_1\} = 1 - (1 - c_1)^2 = 1 - (1 - \sqrt{\varepsilon})^2.$$

18.3. Проверяются гипотезы о плотности f распределения наблюдений $X_1, \ldots, X_n$: гипотеза $H_1 = \{f = f_1\}$ против альтернативы $H_2 = \{f = f_2\}$. Здесь

$$f_1(y) = \begin{cases} 1 & \text{при } y \in [0,1], \\ 0 & \text{при } y \notin [0,1], \end{cases} \qquad f_2(y) = \begin{cases} 2y & \text{при } y \in [0,1], \\ 0 & \text{при } y \notin [0,1]. \end{cases}$$

Построить наиболее мощный критерий размера ε

а) при $n = 1$; б) при $n = 2$.

18.4. По выборке X_1 объёма 1 проверяется гипотеза о том, что X_1 распределено равномерно на отрезке $[0,1]$, против альтернативы о том, что X_1 имеет распределение с плотностью

$$f(y) = \begin{cases} 3/2 & \text{при } y \in [0,1/2], \\ 1/2 & \text{при } y \in (1/2,1], \\ 0 & \text{при } y \notin [0,1]. \end{cases}$$

Построить наиболее мощный критерий размера $1/4$.

Решение. Отношение правдоподобия равняется $3/2$, если $x_1 \in [0,\ 1/2]$, и $1/2$, если $x_1 \in (1/2,\ 1]$. Заметим, что для любого $1/2 \leqslant c < 3/2$

$$\mathbf{P}_{H_1}\left\{\frac{f_2(X_1)}{f_1(X_1)} > c\right\} = \mathbf{P}_{H_1}\{0 \leqslant X_1 \leqslant 1/2\} = 1/2 > 1/4.$$

Как только $c \geqslant 3/2$, вероятность $\mathbf{P}_{H_1}\{f_2(X_1)/f_1(X_1) > c\}$ становится равной нулю, что меньше $1/4$. Поэтому следует взять $c = 3/2$ и найти ρ из условия

$$\frac{1}{4} = \mathbf{P}_{H_1}\left\{\frac{f_2(X_1)}{f_1(X_1)} > \frac{3}{2}\right\} + \rho\mathbf{P}_{H_1}\left\{\frac{f_2(X_1)}{f_1(X_1)} = \frac{3}{2}\right\} = \rho\mathbf{P}_{H_1}\{0 \leqslant X_1 \leqslant 1/2\} = \frac{\rho}{2}.$$

Отсюда $\rho = 1/2$. Поэтому наиболее мощный критерий размера $1/4$ имеет вид: $\delta(X_1) = 0$ (гипотеза H_1 принимается) при $X_1 \in (1/2,\ 1]$ и $\delta(X_1) = 1/2$ (гипотеза H_2 принимается с вероятностью $1/2$) при $X_1 \in [0,\ 1/2]$.

18.5. Пусть X_1 — выборка объёма 1. Основная гипотеза состоит в том, что элементы выборки распределены равномерно на отрезке $[0,1]$. Альтернатива — в том, что элементы выборки имеют показательное распределение с параметром 1. Построить наиболее мощный критерий размера ε для различения этих гипотез и вычислить его мощность.

18.6. Пусть X_1 — выборка объёма 1 из распределения Пуассона с параметром λ. Рассматриваются две простые гипотезы: $\lambda = 1$ и $\lambda = 2$. Построить наиболее мощный критерий $\delta = \delta(X_1)$ с вероятностью ошибки первого рода $\alpha = 1 - e^{-1}$. Найти мощность этого критерия.

18.7. Пусть X_1 — выборка объёма 1. Гипотеза состоит в том, что X_1 имеет показательное распределение с параметром $\alpha = 2$. Альтернатива состоит в том, что X_1 имеет плотность

$$f_2(y) = \begin{cases} 1/2 & \text{при } y \in [0,1], \\ 1 & \text{при } y \in [3/2, 2], \\ 0 & \text{при } y \notin [0,1] \cup [3/2, 2]. \end{cases}$$

Построить наиболее мощный критерий размера $1/3$.

18.8. Пусть X_1 — выборка объёма 1. Гипотеза состоит в том, что X_1 имеет равномерное распределение на отрезке $[1,2]$. Альтернатива состоит в том, что X_1 имеет плотность

$$f_2(y) = \begin{cases} 1/2 & \text{при } y \in [0,1), \\ 1 & \text{при } y \in [1, 3/2], \\ 0 & \text{при } y \notin [0, 3/2]. \end{cases}$$

Построить наиболее мощный критерий размера $1/6$.

18.9. Пусть X_1 — выборка объёма 1. Гипотеза состоит в том, что X_1 имеет распределение Бернулли с параметром $p = 1/2$. Альтернатива состоит в том, что X_1 имеет биномиальное распределение с параметрами $m = 2$ и $p = 1/2$. Построить наиболее мощный критерий размера $1/5$.

18.10. В последовательности независимых испытаний вероятности положительных исходов одинаковы и равны p. Построить критерий для проверки гипотезы $p = 0$ против альтернативы $p = 0,01$ и определить наименьший объём выборки, при котором вероятности ошибок первого и второго рода не превосходят 0,01.

18.11. У игрока, наблюдавшего за игрой в кости, создалось впечатление, что шестёрка выпадает в 18% бросаний, пятёрка — в 14%, а остальные четыре грани выпадают равновероятно (т. е. с вероятностью 0,17). Получив приглашение принять участие в игре, игрок попросил разрешения предварительно проверить свою гипотезу на n производимых подряд бросаниях кости. Единственная рассматриваемая им альтернатива состоит в том, что игральная кость сделана «честно». При $n = 2$ найти наиболее мощный критерий размера 0,0196.

18.12. Пусть X_1 — выборка объёма 1. Проверяются гипотезы о распределении F наблюдения X_1: гипотеза $H_1 = \{F = F_1\}$ против альтернативы $H_2 = \{F = F_2\}$. Распределение F_1 есть смесь в равной пропорции вырожденного в нуле распределения и равномерного на отрезке $[0, 1]$. Распределение F_2 есть также смесь в равной пропорции вырожденного в нуле распределения и распределения с плотностью $2y$ на отрезке $[0, 1]$. Построить наиболее мощный критерий размера $1/2$. Найти все $\varepsilon \in [0, 1]$, при которых наиболее мощный критерий с ошибкой первого рода равной ε будет рандомизированным.

18.13. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и известной дисперсией σ^2. Построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{a = a_1\}$ против альтернативы $H_2 = \{a = a_2\}$, где $a_1 < a_2$. Будет ли этот критерий состоятельным?

Решение. Отношение правдоподобия имеет абсолютно непрерывное распределение при гипотезе H_1, поэтому наиболее мощный критерий будет нерандомизированным. Критическое множество определяется неравенством

$$\frac{f_2(X_1, \ldots, X_n)}{f_1(X_1, \ldots, X_n)} \equiv \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((X_i - a_1)^2 - (X_i - a_2)^2\right)\right\} \geqslant c,$$

что эквивалентно соотношению $\overline{X} \geqslant c_1$, где c_1 определяется по заданному размеру ε следующим образом:

$$\begin{aligned}\alpha(\delta) &= \mathbf{P}_{H_1}\left\{\overline{X} \geqslant c_1\right\} \\ &= \mathbf{P}_{H_1}\left\{\sqrt{n}\,\frac{\overline{X} - a_1}{\sigma} \geqslant \sqrt{n}\,\frac{c_1 - a_1}{\sigma}\right\} = \overline{\Phi}\left(\sqrt{n}\,\frac{c_1 - a_1}{\sigma}\right) = \varepsilon.\end{aligned}$$

Следовательно, $\sqrt{n}(c_1 - a_1)/\sigma = \zeta_{1-\varepsilon}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1 - \varepsilon$ стандартного нормального распределения. Поэтому $c_1 = a_1 + \sigma\zeta_{1-\varepsilon}/\sqrt{n}$ и наиболее мощный критерий размера ε имеет вид

$$\delta = \begin{cases} 1, & \text{если } \overline{X} \geqslant a_1 + \sigma\zeta_{1-\varepsilon}/\sqrt{n}, \\ 0, & \text{если } \overline{X} < a_1 + \sigma\zeta_{1-\varepsilon}/\sqrt{n}. \end{cases}$$

Мощность этого критерия равна

$$\begin{aligned}\beta(\delta) &= \mathbf{P}_{H_2}\left\{\overline{X} \geqslant a_1 + \sigma\zeta_{1-\varepsilon}/\sqrt{n}\right\} \\ &= \mathbf{P}_{H_2}\left\{\sqrt{n}\,\frac{\overline{X} - a_2}{\sigma} \geqslant \zeta_{1-\varepsilon} - \sqrt{n}\,\frac{a_2 - a_1}{\sigma}\right\} = \overline{\Phi}\left(\zeta_{1-\varepsilon} - \sqrt{n}\,\frac{a_2 - a_1}{\sigma}\right).\end{aligned}$$

Мощность критерия стремится к 1 с ростом n при любом фиксированном ε, так как $\zeta_{1-\varepsilon} - \sqrt{n}(a_2 - a_1)/\sigma \to -\infty$. Поэтому критерий состоятелен.

18.14. По выборке объёма n при заданной вероятности ошибки первого рода построить наиболее мощный критерий для различения двух простых гипотез относительно неизвестной дисперсии нормального распределения, если математическое ожидание известно и равно нулю.

18.15. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами a и σ^2. Построить наиболее мощный критерий для проверки гипотезы $H_1 = \{a = a_1, \ \sigma^2 = \sigma_1^2\}$ против альтернативы $H_2 = \{a = a_2, \ \sigma^2 = \sigma_2^2\}$.

18.16. По выборке из показательного распределения с параметром α построить наиболее мощный критерий асимптотического размера ε, различающий гипотезу $\alpha = \alpha_1$ и альтернативу $\alpha = \alpha_2$, если $\alpha_1 < \alpha_2$. Вычислить предел мощности построенного критерия при $n \to \infty$.

18.17. По выборке из распределения Пуассона с параметром λ построить наиболее мощный критерий асимптотического размера ε, различающий гипотезу $\lambda = \lambda_1$ и альтернативу $\lambda = \lambda_2$, если $\lambda_1 < \lambda_2$. Вычислить предел мощности построенного критерия при $n \to \infty$.

18.18. По выборке из биномиального распределения с параметрами m и p построить наиболее мощный критерий асимптотического размера ε, различающий гипотезу $p = p_1$ и альтернативу $p = p_2$, если $p_1 < p_2$. Вычислить предел мощности построенного критерия при $n \to \infty$.

18.19. По выборке из геометрического распределения с параметром p построить наиболее мощный критерий асимптотического размера ε, различающий гипотезу $p = p_1$ и альтернативу $p = p_2$, если $p_1 < p_2$. Вычислить предел мощности построенного критерия при $n \to \infty$.

18.20. Вероятность успеха p в схеме Бернулли неизвестна. Для проверки гипотезы $p = p_1$ против альтернативы $p = p_2$, где $p_2 > p_1$, проведён эксперимент, в котором наблюдали число успехов, предшествующих первой неудаче. Построить наиболее мощный критерий размера p_1^s, где $s \geqslant 1$ — заданное целое число. Найти мощность этого критерия.

18.21. Для какой постоянной c, участвующей в определении наиболее мощного критерия (в лемме Неймана — Пирсона), этот критерий совпадает с байесовским, если предположить, что априорные вероятности гипотез H_1 и H_2 равны соответственно 1/3 и 2/3?

18.22. Доказать состоятельность наиболее мощного критерия.

18.23. Обозначим через $m(\varepsilon)$ мощность наиболее мощного критерия среди всех рандомизированных критериев размера ε. Доказать, что $m(\varepsilon) \geqslant \varepsilon$.

§ 19. Равномерно наиболее мощные критерии

Пусть $\{F_\theta,\ \theta \in \Theta\}$ — некоторое параметрическое семейство распределений и X_1, X_2, … — выборка из распределения F_θ. Пусть проверяется простая гипотеза $\theta = \theta_0$ против сложной альтернативы $\theta \in \Theta_0$, где θ_0 — фиксированная точка в Θ, а Θ_0 — некоторое подмножество в Θ, причём $\theta_0 \notin \Theta_0$. Обозначим через

$$\alpha(\delta) = \mathbf{E}_{\theta_0}\delta(X_1, \ldots, X_n)$$

размер критерия δ, а через

$$\beta_\theta(\delta) = 1 - \mathbf{E}_\theta\delta(X_1, \ldots, X_n), \quad \theta \in \Theta_0,$$

функцию мощности критерия δ.

Равномерно наиболее мощным критерием размера ε называется такой критерий δ (вообще говоря, рандомизированный), что $\alpha(\delta) \leqslant \varepsilon$ и любой другой критерий с размером, не превосходящим ε, при любом значении $\theta \in \Theta_0$ имеет мощность, не превосходящую $\beta_\theta(\delta)$.

19.1. Пусть X_1, …, X_n — выборка из нормального распределения со средним a и известной дисперсией σ^2. Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{a = a_1\}$ против альтернативы $H_2 = \{a > a_1\}$.

Решение. Критерий, построенный в задаче 18.13, не зависит от a_2, т. е. является наиболее мощным при любой простой альтернативе $a = a_2 > a_1$. Поэтому этот критерий является и равномерно наиболее мощным критерием для проверки простой гипотезы $a = a_1$ против сложной альтернативы $a > a_1$.

19.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с известным средним значением a и неизвестной дисперсией σ^2. Используя достаточную статистику $\overline{(X-a)^2}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\sigma^2 = \sigma_1^2\}$ против альтернативы $H_2 = \{\sigma^2 < \sigma_1^2\}$.

19.3. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\alpha = \alpha_1\}$ против альтернативы $H_2 = \{\alpha > \alpha_1\}$.

19.4. Пусть $X_1, \ldots, X_n$ — выборка из двухпараметрического показательного распределения с плотностью

$$f_{\alpha,\beta}(y) = \begin{cases} \alpha^{-1}\, e^{-(y-\beta)/\alpha} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta, \end{cases}$$

где $\alpha > 0$, $\beta \in \mathbf{R}$.

а) Пусть α известно. Используя достаточную статистику $X_{(1)}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\beta = \beta_1\}$ против альтернативы $H_2 = \{\beta \neq \beta_1\}$.

б) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\alpha = \alpha_1,\ \beta = \beta_1\}$ против альтернативы $H_2 = \{\alpha < \alpha_1,\ \beta < \beta_1\}$.

19.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Используя достаточную статистику $X_{(n)}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\theta = \theta_0\}$ против альтернативы $H_2 = \{\theta \neq \theta_0\}$.

19.6. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{p = p_1\}$ против альтернативы $H_2 = \{p > p_1\}$.

19.7. В условиях задачи 16.6 построить равномерно наиболее мощный критерий размера 0,1 по результатам 100 экспериментов.

19.8. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\lambda = \lambda_1\}$ против альтернативы $H_2 = \{\lambda > \lambda_1\}$.

19.9. Пусть $X_1, \ldots, X_n$ — выборка из геометрического распределения с параметром p. Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{p = p_1\}$ против альтернативы $H_2 = \{p > p_1\}$.

19.10. По выборке X_1 объёма 1 проверяется основная гипотеза о том, что X_1 имеет стандартное нормальное распределение, против альтернативы, состоящей в том, что распределение X_1 обладает свойством $\mathbf{P}\{X_1 \in [0, 1]\} = 0$. Построить критерий с единичной мощностью. Каков наименьший возможный размер такого критерия?

§ 20. Критерии согласия

Пусть имеется выборка $X_1, \ldots, X_n$ из неизвестного распределения F. Пусть F_1 — некоторое распределение. Критерии, предназначенные для проверки основной гипотезы $H_1 = \{F = F_1\}$, называются *критериями согласия*. Альтернативной гипотезой чаще всего является $H_2 = \{F \neq F_1\}$. Иногда в качестве H_1 выступает тоже сложная гипотеза.

Пусть задан некоторый функционал $d(P_n^*, F_1)$, обладающий следующим свойством: по заданному ε можно найти c такое, что

$$\mathbf{P}_{H_1}\{d(P_n^*, F_1) > c\} = \varepsilon \quad \text{или} \quad \lim_{n\to\infty} \mathbf{P}_{H_1}\{d(P_n^*, F_1) > c\} = \varepsilon.$$

Значение функционала $d(P_n^*, F_1)$ можно трактовать как «расстояние» между эмпирическим и предполагаемым теоретическим распределением.

Критерий согласия (асимптотического) размера ε, основанный на функционале d, строится следующим образом: критерий отвергает основную гипотезу, если для данной выборки значение $d(P_n^*, F_1)$ превосходит c.

Если $d(P_n^*, F_1)$ стремится по вероятности к бесконечности при $n \to \infty$, как только распределение F отлично от F_1, то данный критерий согласия состоятелен. А именно, при любом распределении F, отличном от F_1, вероятность ошибки второго рода стремится к нулю.

Критерий Колмогорова. Пусть имеется выборка $X_1, \ldots, X_n$ из неизвестного распределения F и $F_n^*(y)$ — эмпирическая функция распределения, построенная по этой выборке. Пусть F_1 — некоторое распределение с непрерывной функцией распределения $F_1(y)$. Для проверки простой гипотезы $H_1 = \{F = F_1\}$ используется статистика Колмогорова

$$d(X_1, \ldots, X_n) = \sqrt{n} \sup_{y \in \mathbf{R}} |F_n^*(y) - F_1(y)|.$$

Справедлива следующая

Теорема Колмогорова. *Если $F = F_1$, то при $n \to \infty$ распределение статистики Колмогорова слабо сходится к распределению Колмогорова с функцией распределения*

$$K(y) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 y^2}, \quad y > 0.$$

Критерий Колмогорова асимптотического размера ε отвергает основную гипотезу, если значение статистики Колмогорова $d(X_1, \ldots, X_n)$ превосходит квантиль $\zeta_{1-\varepsilon}$ уровня $1 - \varepsilon$ распределения Колмогорова.

Критерий Пирсона хи-квадрат. Пусть имеется выборка $X_1, \ldots, X_n$ из неизвестного распределения F и F_1 — некоторое распределение. Пусть задан конечный набор из k непересекающихся интервалов $\Delta_1, \ldots, \Delta_k$, покрывающих $\mathbf{R}$. Обозначим через $p_j = F_1(\Delta_j)$ вероятности попадания в эти интервалы для распределения F_1 и через ν_j — число элементов выборки, попавших в интервал Δ_j.

Для проверки гипотезы H_1 о совпадении вектора неизвестных истинных вероятностей $(F(\Delta_1), \ldots, F(\Delta_k))$ с вектором $(p_1, \ldots, p_k)$ используется статистика хи-квадрат

$$\chi^2(X_1, \ldots, X_n) = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}.$$

Справедлива следующая

Теорема Пирсона. *Если гипотеза H_1 верна, то при $n \to \infty$ распределение статистики хи-квадрат слабо сходится к χ^2-распределению с $k - 1$ степенью свободы.*

Критерий Пирсона асимптотического размера ε отвергает основную гипотезу, если значение статистики хи-квадрат $\chi^2(X_1, \ldots, X_n)$ превосходит квантиль $\zeta_{1-\varepsilon}$ уровня $1 - \varepsilon$ χ^2-распределения с $k - 1$ степенью свободы.

20.1. Имеется выборка X_1, X_2, X_3 объёма 3. Для проверки гипотезы о том, что выборка взята из равномерного на отрезке $[0, 1]$ распределения, используется критерий Колмогорова: гипотеза о равномерности отвергается, если $\sup\limits_{y \in [0,1]} |F_3^*(y) - y| > 1/3$. Сформулировать этот критерий в явном виде в терминах порядковых статистик. Чему равен размер этого критерия?

20.2. Доказать состоятельность критерия Колмогорова.

20.3. Для проверки гипотезы о том, что выборка взята из равномерного на отрезке $[0,1]$ распределения, используется статистика омега-квадрат

$$\omega^2 = \int\limits_0^1 (F_n^*(y) - y)^2\, dy.$$

Гипотеза о равномерности отвергается, если $\omega^2 \geqslant \gamma$, где число $\gamma > 0$ выбирается заранее. Доказать, что для выборки из равномерного на отрезке $[0,1]$ распределения справедливо равенство

$$\mathbf{E}\omega^2 = 1/6n.$$

С помощью неравенства Чебышёва указать значение γ, при котором размер критерия не превосходит ε.

20.4. Доказать, что при условии $0 \leqslant X_{(1)} \leqslant X_{(n)} \leqslant 1$ справедливо равенство

$$\int\limits_0^1 (F_n^*(y) - y)^2\, dy = \frac{1}{12n^2} + \frac{1}{n}\sum_{k=1}^{n}\left(X_{(k)} - \frac{2k-1}{n}\right)^2$$

(с помощью этого представления часто вычисляется значение статистики ω^2).

20.5. Для проверки гипотезы о том, что выборка взята из распределения с непрерывной функцией распределения F, используется статистика

$$\omega^2 = \int\limits_0^1 (F_n^*(y) - F(y))^2\, dF(y).$$

Доказать, что при выполнении основной гипотезы распределение статистики ω^2 не зависит от непрерывного распределения F.

20.6. При $n = 4040$ бросаниях монеты Бюффон получил 2048 выпадений герба и 1992 выпадений решётки. Совместимо ли это с гипотезой о том, что существует постоянная вероятность $p = 1/2$ выпадения герба?

20.7. В ходе $n = 4000$ независимых испытаний события A_1, A_2 и A_3, составляющие полную группу событий, появились 1905, 1015 и 1080 раз соответственно. Проверить, согласуются ли эти данные на уровне 0,05 с гипотезой $H = \{p_1 = 1/2,\ p_2 = p_3 = 1/4\}$, где $p_j = \mathbf{P}\{A_j\}$.

20.8. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и единичной дисперсией. Построить такой критерий δ с вероятностью ошибки первого рода $\alpha_1(\delta) = \varepsilon$ для различения гипотез $H_1 = \{a = a_0\}$, $H_2 = \{a < a_0\}$ и $H_3 = \{a > a_0\}$, что $\alpha_2(\delta) \to 0$ при любом $a < a_0$ и $\alpha_3(\delta) \to 0$ при любом $a > a_0$.

Решение. Построим критерий с помощью статистики $\sqrt{n}\left(\overline{X} - a_0\right)$, имеющей при гипотезе H_1 стандартное нормальное распределение. Если $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1 - \varepsilon/2$ стандартного нормального распределения, то критерий

$$\delta(X_1, \ldots, X_n) = \begin{cases} H_2, & \text{если } \sqrt{n}\left(\overline{X} - a_0\right) < -\zeta_{1-\varepsilon/2}, \\ H_1, & \text{если } -\zeta_{1-\varepsilon/2} \leqslant \sqrt{n}\left(\overline{X} - a_0\right) \leqslant \zeta_{1-\varepsilon/2}, \\ H_3, & \text{если } \sqrt{n}\left(\overline{X} - a_0\right) > \zeta_{1-\varepsilon/2} \end{cases}$$

имеет вероятность ошибки первого рода ε. Вероятность ошибки второго рода стремится к нулю:

$$\mathbf{P}_{H_2}\{\delta \neq H_2\} = \mathbf{P}_{H_2}\left\{\sqrt{n}\left(\overline{X} - a_0\right) > -\zeta_{1-\varepsilon/2}\right\} \to 0,$$

поскольку $\overline{X} - a_0 \xrightarrow{\text{p}} a - a_0 < 0$ для любого $a < a_0$, так что величина $\sqrt{n}\left(\overline{X} - a_0\right)$ с ростом n стремится по вероятности к минус бесконечности. В силу симметрии вероятность ошибки третьего рода также стремится к нулю.

20.9. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p. Построить какой-либо состоятельный критерий асимптотического размера ε для проверки гипотезы $p = p_0$ против альтернативы $p \neq p_0$.

20.10. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ. Построить какой-либо состоятельный критерий асимптотического размера ε для проверки гипотезы $\lambda = \lambda_0$ против альтернативы $\lambda \neq \lambda_0$.

20.11. Используя конструкции доверительного интервала, построить критерий с (точной или асимптотической) ошибкой первого рода ε для проверки гипотезы $\theta = 1$ по выборке из а) нормального распределения со средним θ и дисперсией 1;

б) нормального распределения со средним 1 и дисперсией θ;

в) показательного распределения с параметром θ;

г) распределения Бернулли с параметром $\theta/2$;

д) распределения Пуассона с параметром θ.

20.12. Пусть выборка $X_1, \ldots, X_n$ имеет нормальное распределение со средним a и дисперсией σ^2. Для проверки гипотезы о том, что $a = a_0$, используется статистика

$$\frac{\sqrt{n}\,\left|\overline{X} - a_0\right|}{\sqrt{S_0^2}}\,.$$

Доказать, что соответствующий критерий состоятелен.

20.13. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения со средним a и единичной дисперсией, а $Y_1, \ldots, Y_m$ — выборка объёма m из нормального распределения со средним b и единичной дисперсией; выборки X и Y независимы. Проверяется гипотеза о близости математических ожиданий a и b. Основная гипотеза $H_1 = \{a = b\}$ принимается, если $\left|\overline{X} - \overline{Y}\right| \leqslant 1$. Иначе принимается альтернатива $H_2 = \{|a - b| > 1\}$. Является ли данный критерий состоятельным (при n, $m \to \infty$)? Найти вероятность ошибки первого рода этого критерия.

20.14. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения со средним a и единичной дисперсией, а $Y_1, \ldots, Y_m$ — выборка объёма m из нормального распределения со средним b и единичной дисперсией; выборки X и Y независимы. Известно, что $a \geqslant b$. Проверяется гипотеза о равенстве математических ожиданий a и b. Основная гипотеза $H_1 = \{a = b\}$ принимается, если

$$\sqrt{\frac{nm}{n+m}}\left(\overline{X} - \overline{Y}\right) \leqslant c.$$

В противном случае принимается альтернатива $H_2 = \{a > b\}$. Здесь $c > 0$ — заранее выбранное число. Найти размер данного критерия в зависимости от c. Проверить состоятельность этого критерия.

20.15. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения со средним a и известной дисперсией σ_1^2, а $Y_1, \ldots, Y_m$ — выборка объёма m из нормального распределения со средним b и известной дисперсией σ_2^2; выборки X и Y независимы. Рассматривается основная гипотеза $H_1 = \{a = b\}$ против альтернативы $H_2 = \{a > b\}$. Построить какой-нибудь состоятельный критерий размера ε, основываясь на статистике

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}.$$

20.16. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения со средним a и дисперсией σ^2, а выборка $Y_1, \ldots, Y_m$ объёма m — из нормального распределения со средним b и той же дисперсией σ^2. Для проверки гипотезы о том, что $a = b$, используется статистика

$$\frac{\sqrt{\dfrac{nm}{n+m}}\,|\overline{X} - \overline{Y}|}{\sqrt{\dfrac{1}{n+m-2}\left(\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum\limits_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2\right)}}.$$

Доказать, что соответствующий критерий состоятелен.

20.17. Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ — две независимые выборки из непрерывных распределений. Для проверки гипотезы о совпадении распределений используют набор разностей $Z_i = X_i - Y_i$, $i = 1, \ldots, n$. Гипотеза о совпадении распределений отклоняется, если число положительных членов в последовательности $Z_1, \ldots, Z_n$ отличается от $n/2$ более чем на γ, где число $\gamma > 0$ заранее выбирается подходящим образом (критерий знаков). Найти вероятность ошибки первого рода в точном виде и оценить её значение при больших n с помощью нормального приближения. Каким нужно выбрать число γ, чтобы вероятность ошибки первого рода равнялась ε? Является ли критерий состоятельным?

20.18. При переписи населения Англии и Уэльса в 1901 г. было зарегистрировано (с точностью до тысячи) 15 729 000 мужчин и 16 799 000 женщин; 3 497 мужчин и 3 072 женщины были зарегистрированы как глухонемые от рождения. Проверить гипотезу о том, что глухонемота не связана с полом.

Решение. Можно использовать критерий, основанный на том, что при верной основной гипотезе о равенстве параметров двух распределений Бернулли статистика

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{p^*(1-p^*)(1/n+1/m)}},$$

(нормированное расстояние между выборочными средними) слабо сходится к стандартному нормальному распределению. Здесь через n и m обозначены объёмы независимых выборок $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ соответственно, а $p^* = \frac{\sum X_i + \sum Y_j}{n+m}$ — оценка для параметра p, полученная по объединённой выборке в предположении справедливости основной гипотезы. Подставив данные задачи, получим значение $T = 7{,}91489$. Реально достигнутый уровень значимости равен (практически точно, так как объёмы выборок очень велики)

$$\varepsilon^* = \mathbf{P}\{|\xi| > T\} = 2\overline{\Phi}(7{,}91),$$

где ξ имеет стандартное нормальное распределение. Правая часть равна 0 с большой степенью точности. Тем самым следует отвергнуть основную гипотезу, так как статистика критерия даёт такое громадное отклонение, какое при верной основной гипотезе может получиться лишь с почти нулевой вероятностью.

20.19. Пользуясь интегральной теоремой Муавра – Лапласа, доказать теорему Пирсона при $k = 2$.

20.20. Пользуясь законом больших чисел Бернулли, доказать состоятельность критерия «хи-квадрат».

20.21. Цифры 0, 1, 2, ..., 9 среди 800 первых десятичных знаков числа π появились 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно. Проверить гипотезу о согласии этих данных с законом равномерного на множестве $\{0, 1, \ldots, 9\}$ распределения.

20.22. По официальным данным в Швеции в 1935 г. родилось 88 273 ребенка, причем в январе родилось 7280 детей, в феврале — 6957, марте — 7883, апреле — 7884, мае — 7892, июне — 7609, июле — 7585, августе — 7393, сентябре — 7203, октябре — 6903, ноябре — 6552 и в декабре — 7132 ребенка. Совместимы ли эти данные с гипотезой, что день рождения наудачу выбранного человека с равной вероятностью приходится на любой из 365 дней года?

20.23. Ниже приводятся результаты 4096 опытов, состоящих в одновременном подбрасывании 12 костей (данные Уэлдона). В каждом из опытов подсчитывалось число костей, выпавших кверху шестёркой (гранью с шестью очками). Проверить гипотезу правильности костей.

Число шестёрок	0	1	2	3	4	5	6	$\geqslant 7$	Всего
Число случаев	447	1145	1181	796	380	115	24	8	4096

20.24. В [13] приведены данные, собранные доктором Э. Бэрром из Оксфордского окружного госпитального совета, о моментах поступления пациентов в отделение интенсивной терапии с понедельника 4 февраля 1963 г. по среду 18 марта 1964 г. Рассмотрим три различных способа группировки этих событий.

А. Изменчивость по месяцам

Месяц и год	Число дней	Число пациентов	Месяц и год	Число дней	Число пациентов
Февраль 63	25	13	Сентябрь 63	30	17
Март 63	31	16	Октябрь 63	31	17
Апрель 63	30	12	Ноябрь 63	30	28
Май 63	31	18	Декабрь 63	31	32
Июнь 63	30	23	Январь 64	31	23
Июль 63	31	16	Февраль 64	29	17
Август 63	31	15	Март 64	18	7

Согласуются ли эти данные с гипотезой, что пациенты попадают в отделение с равной вероятностью в любой из дней? Исследовать тот же вопрос при исключении двух последних месяцев в году — ноября и декабря.

В. Изменчивость по дням недели

День недели	Пн	Вт	Ср	Чт	Пт	Сб	Вс
Число пациентов	37	53	35	27	30	44	28

Согласуются ли эти данные с гипотезой, что пациенты попадают в отделение с равной вероятностью в любой из семи дней недели? В любой из дней недели кроме вторника?

С. Изменчивость по времени суток в часах

Интервал времени	Число пациентов	Интервал времени	Число пациентов
0 – 2	14	12 – 14	31
2 – 4	17	14 – 16	30
4 – 6	5	16 – 18	26
6 – 8	8	18 – 20	29
8 – 10	5	20 – 22	31
10 – 12	25	22 – 24	23

Согласуются ли эти данные с гипотезой, что вероятность попасть в отделение интенсивной терапии не зависит от времени суток? Проверить, так ли это хотя бы в дневное время, т. е. с 10.00 до 24.00.

ОТДЕЛ VII

ЗАДАЧИ НА ПОВТОРЕНИЕ

§ 21. Оценка параметров

21.1. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с параметрами a и σ^2.

а) Найти смещение и дисперсию оценок $S^2 = \overline{X^2} - (\overline{X})^2$ и $S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ для параметра σ^2.

б) Используя метод моментов, оценить параметры a и σ^2. Проверить полученные оценки на несмещённость, состоятельность и асимптотическую нормальность.

в) Найти оценку максимального правдоподобия двумерного параметра $\theta = (a, \sigma^2)$. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

г) Является ли выборочная медиана ζ^* несмещённой, состоятельной и асимптотически нормальной оценкой параметра a?

д) Считая, что параметр a имеет показательное распределение с параметром α, найти байесовскую оценку параметра a.

е) Сравнить оценки дисперсии S^2 и S_0^2 с помощью среднеквадратического подхода.

ж) Является ли двумерная статистика $(\overline{X},\ S_0^2)$ достаточной для двумерного параметра (a, σ^2)?

з) Является ли двумерная статистика $(\overline{X},\ S_0^2)$ полной?

и) Найти эффективную несмещённую оценку двумерного параметра (a, σ^2).

к) Построить точные доверительные интервалы уровня $1 - \varepsilon$ для параметров a и σ^2.

21.2. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$.

а) Найти смещение и дисперсию оценок $2\overline{X}$, $X_{(n)}$ и $\frac{n+1}{n}X_{(n)}$.

б) Используя метод моментов, оценить параметр θ. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

в) Найти оценку максимального правдоподобия параметра θ и проверить её на несмещённость, состоятельность и асимптотическую нормальность.

г) Считая, что параметр θ имеет распределение Парето с параметрами 1 и 2, найти байесовскую оценку параметра θ.

д) Сравнить оценки $2\overline{X}$, $X_{(n)}$ и $\frac{n+1}{n}X_{(n)}$ с помощью среднеквадратического подхода.

е) Какие из статистик $\overline{X}$ и $X_{(n)}$ являются достаточными?

ж) Является ли статистика $X_{(n)}$ полной?

з) Является ли R-эффективной оценка $2\overline{X}$?

и) Найти эффективную несмещённую оценку параметра θ.

к) Используя статистику $2\overline{X}$, построить асимптотический доверительный интервал уровня $1 - \varepsilon$ для параметра θ.

л) Используя статистику $X_{(n)}$, построить точный доверительный интервал уровня $1 - \varepsilon$ для параметра θ.

21.3. Пусть дана выборка $X_1, \ldots, X_n$ из равномерного распределения в некоторой области $G \subset \mathbf{R}^d$. Для оценки значения интеграла

$$a = \int \cdots \int_G f(x_1, \ldots, x_d) dx_1 \ldots dx_d$$

по методу Монте-Карло используется статистика

$$a_n^* = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$

а) Найти $\mathbf{E}a_n^*$ и $\mathbf{D}a_n^*$.

б) Построить несмещённую оценку дисперсии a_n^*.

в) Предполагая конечность интеграла

$$\int_G \cdots \int f^4(x_1, \ldots, x_d) dx_1 \ldots dx_d,$$

построить асимптотический доверительный интервал для параметра a уровня $1 - \varepsilon$.

21.4. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α.

а) Используя метод моментов, оценить параметр α. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

б) Найти несмещённую оценку параметра α.

в) Проверить оценку $\overline{X}$ на несмещённость, состоятельность и асимптотическую нормальность для параметра $\tau = 1/\alpha$.

г) Является ли статистика $\overline{X}$ достаточной для параметра α?

д) Является ли статистика $\overline{X}$ полной?

е) Найти эффективную несмещённую оценку параметра α.

ж) Построить асимптотический доверительный интервал уровня $1 - \varepsilon$ для параметра α.

21.5. Пусть $X_1, \ldots, X_n$ — выборка из смещённого показательного распределения с плотностью

$$f_\beta(y) = \begin{cases} e^{\beta - y} & \text{при } y \geqslant \beta, \\ 0 & \text{при } y < \beta. \end{cases}$$

а) Используя метод моментов, оценить параметр сдвига β. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

б) Найти оценку максимального правдоподобия для параметра сдвига β и проверить её на несмещённость, состоятельность и асимптотическую нормальность.

в) Сравнить оценки $\overline{X} - 1$ и $X_{(1)}$ с помощью среднеквадратического подхода.

г) Является ли статистика $X_{(1)}$ достаточной для параметра β?

д) Является ли статистика $X_{(1)}$ полной?

е) Найти эффективную несмещённую оценку параметра β.

ж) Построить точный доверительный интервал уровня $1-\varepsilon$ для параметра β.

21.6. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p.

а) Найти смещение и дисперсию оценок $\overline{X}$, $X_{(n)}$ и X_1.

б) Используя метод моментов, оценить параметр p. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

в) Найти оценку максимального правдоподобия параметра p и проверить её на несмещённость, состоятельность и асимптотическую нормальность.

г) Считая, что параметр p принимает значения $1/4$ и $3/4$ с вероятностями $1/4$ и $3/4$ соответственно, найти байесовскую оценку параметра p.

д) Сравнить оценки $\overline{X}$ и X_1 с помощью среднеквадратического подхода.

е) Какие из статистик $\overline{X}$, $X_{(n)}$ и $2\overline{X}$ являются достаточными?

ж) Какие из статистик $\overline{X}$, $X_{(n)}$ и $2\overline{X}$ являются полными?

з) Является ли R-эффективной оценка $\overline{X}$?

и) Найти эффективную несмещённую оценку параметра p.

к) Используя статистику $\overline{X}$, построить асимптотический доверительный интервал уровня $1-\varepsilon$ для параметра p.

21.7. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ.

а) Найти смещение и дисперсию оценок $\overline{X}$ и X_1.

б) Используя метод моментов, оценить параметр λ. Проверить полученную оценку на несмещённость, состоятельность и асимптотическую нормальность.

в) Найти оценку максимального правдоподобия параметра λ.

г) Считая, что параметр λ принимает значения 1 и 2 с равными вероятностями, найти байесовскую оценку параметра λ.

д) Сравнить оценки $\overline{X}$ и X_1 с помощью среднеквадратического подхода.

е) Какие из статистик $\overline{X}$, $X_{(n)}$ и $2\overline{X}$ являются достаточными?

ж) Какие из статистик $\overline{X}$, $X_{(n)}$ и $2\overline{X}$ являются полными?

з) Является ли R-эффективной оценка $\overline{X}$?

и) Найти эффективную несмещённую оценку параметра λ.

к) Используя статистику $\overline{X}$, построить асимптотический доверительный интервал уровня $1-\varepsilon$ для параметра λ.

§ 22. Проверка гипотез

22.1. Дана выборка $X_1, \ldots, X_n$. Основная гипотеза H_1 состоит в том, что элементы выборки имеют распределение с плотностью

$$f_1(y) = \begin{cases} 2^y \ln 2 & \text{при } y \leqslant 0, \\ 0 & \text{при } y > 0. \end{cases}$$

Альтернатива H_2 состоит в том, что элементы выборки имеют распределение с плотностью

$$f_2(y) = \begin{cases} 3^y \ln 3 & \text{при } y \leqslant 0, \\ 0 & \text{при } y > 0. \end{cases}$$

а) Критерий δ_1 предписывает принимать гипотезу H_1, если $\overline{X} \geqslant -1/\ln 2$; альтернативу H_2, если $\overline{X} < -1/\ln 2$. Найти пределы вероятностей ошибок первого и второго рода этого критерия при $n \to \infty$.

б) Построить наиболее мощный критерий асимптотического размера $\varepsilon = 0{,}05$ и проверить его состоятельность.

в) Критерий δ_2 предписывает принимать гипотезу H_1, если $X_{(n)} \leqslant -1/4$ и альтернативу H_2, если $X_{(n)} > -1/4$. Найти вероятности ошибок первого и второго рода критерия δ_2.

22.2. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним a и известной дисперсией σ^2.

а) Критерий δ_1 предписывает принимать гипотезу $a = 1$, если $\overline{X} < 1 + 1/\sqrt{n}$; иначе принимается альтернатива $a = 2$. Найти вероятности ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $a = 1$, если $\overline{X} < 3/2$; иначе принимается альтернатива $a = 2$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

в) Используя достаточную статистику $\overline{X}$, построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 =$

$\{a = a_1\}$ против альтернативы $H_2 = \{a = a_2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{a = a_1\}$ против альтернативы $H_2 = \{a > a_1\}$. Проверить состоятельность этого критерия.

д) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{a = a_1\}$ против альтернативы $H_2 = \{a < a_1\}$. Проверить состоятельность этого критерия.

22.3. Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения с известным средним значением a и неизвестной дисперсией σ^2.

а) Критерий δ_1 предписывает принимать гипотезу $\sigma^2 = 1$, если $\overline{(X-a)^2} \leqslant 1$; иначе принимается альтернатива $\sigma^2 = 2$. Найти вероятности ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $\sigma^2 = 1$, если $\overline{X} < 4/3$; иначе принимается альтернатива $\sigma^2 = 2$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

в) При помощи достаточной статистики $\overline{(X-a)^2}$ построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\sigma^2 = \sigma_1^2\}$ против альтернативы $H_2 = \{\sigma^2 = \sigma_2^2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $\overline{(X-a)^2}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\sigma^2 = \sigma_1^2\}$ против альтернативы $H_2 = \{\sigma^2 > \sigma_1^2\}$. Проверить состоятельность этого критерия.

д) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\sigma^2 = \sigma_1^2\}$ против альтернативы $H_2 = \{\sigma^2 < \sigma_1^2\}$. Проверить состоятельность этого критерия.

22.4. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α.

а) Критерий δ_1 предписывает принимать гипотезу $\alpha = 2$, если $\overline{X} \leqslant 1/2$; иначе принимается альтернатива $\alpha = 4$. Найти вероятности ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $\alpha = 2$, если

$\overline{X} > 1/3$; иначе принимается альтернатива $\alpha = 4$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

в) Используя достаточную статистику $\overline{X}$, построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\alpha = \alpha_1\}$ против альтернативы $H_2 = \{\alpha = \alpha_2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\alpha = \alpha_1\}$ против альтернативы $H_2 = \{\alpha > \alpha_1\}$. Проверить состоятельность этого критерия.

д) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\alpha = \alpha_1\}$ против альтернативы $H_2 = \{\alpha < \alpha_1\}$. Проверить состоятельность этого критерия.

22.5. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$.

а) Критерий δ_1 предписывает принимать гипотезу $\theta = 2$, если $\overline{X} \leqslant 3$; иначе принимается альтернатива $\theta = 4$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $\theta = 2$, если $X_{(n)} < 3$; иначе принимается альтернатива $\theta = 4$. Найти вероятности ошибок первого и второго рода этого критерия.

в) Используя достаточную статистику $X_{(n)}$, построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\theta = \theta_1\}$ против альтернативы $H_2 = \{\theta = \theta_2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $X_{(n)}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\theta = \theta_1\}$ против альтернативы $H_2 = \{\theta \neq \theta_1\}$. Проверить состоятельность этого критерия.

22.6. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с параметром p.

а) Критерий δ_1 предписывает принимать гипотезу $p = 1/2$, если $\overline{X} \leqslant 1/2$; иначе принимается альтернатива $p = 3/4$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $p = 1/2$, если $\overline{X} < 1/3$; иначе принимается альтернатива $p = 3/4$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

в) Используя достаточную статистику $\overline{X}$, построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{p = p_1\}$ против альтернативы $H_2 = \{p = p_2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{p = p_1\}$ против альтернативы $H_2 = \{p > p_1\}$. Проверить состоятельность этого критерия.

д) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{p = p_1\}$ против альтернативы $H_2 = \{p < p_1\}$. Проверить состоятельность этого критерия.

22.7. Пусть $X_1, \ldots, X_n$ — выборка из распределения Пуассона с параметром λ.

а) Критерий δ_1 предписывает принимать гипотезу $\lambda = 10$, если $\overline{X} \leqslant 11$; иначе принимается альтернатива $\lambda = 12$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

б) Критерий δ_2 предписывает принимать гипотезу $\lambda = 10$, если $X_{(n)} < 9$; иначе принимается альтернатива $\lambda = 12$. Найти пределы вероятностей ошибок первого и второго рода этого критерия.

в) Используя достаточную статистику $\overline{X}$, построить наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\lambda = \lambda_1\}$ против альтернативы $H_2 = \{\lambda = \lambda_2\}$. Проверить состоятельность этого критерия.

г) Используя достаточную статистику $\overline{X}$, построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\lambda = \lambda_1\}$ против альтернативы $H_2 = \{\lambda > \lambda_1\}$. Проверить состоятельность этого критерия.

д) Построить равномерно наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\lambda = \lambda_1\}$ против альтернативы $H_2 = \{\lambda < \lambda_1\}$. Проверить состоятельность этого критерия.

ПРИЛОЖЕНИЯ

1. Важнейшие дискретные распределения

Тип распределения и обозначение	Параметры	Возможные значения k	Вероятность $\mathbf{P}\{\xi = k\}$
Бернулли, B_p	$p \in [0, 1]$	$k = 0,\, 1$	$\mathbf{P}\{\xi = 0\} = 1 - p$ $\mathbf{P}\{\xi = 1\} = p$
Биномиальное, $B_{m,p}$	$m \in \{1, 2, \ldots\}$, $p \in [0, 1]$	$k = 0,\, \ldots,\, m$	$C_m^k p^k (1-p)^{m-k}$
Отрицательное биномиальное, $\overline{B}_{m,p}$	$m \in \{1, 2, \ldots\}$, $p \in (0, 1]$	$k = 0,\, 1,\, 2,\, \ldots$	$C_{m+k-1}^k (1-p)^k p^m$
Геометрическое, G_p	$p \in (0, 1]$	$k = 0,\, 1,\, 2,\, \ldots$	$p(1-p)^k$
Пуассона, Π_λ	$\lambda \in (0, \infty)$	$k = 0,\, 1,\, 2,\, \ldots$	$\dfrac{\lambda^k}{k!} e^{-\lambda}$

2. Важнейшие плотности распределения

Тип распределения и обозначение	Параметры	Область изменения y	Плотность в точке y
Стандартное нормальное, $N_{0,1}$		$y \in \mathbf{R}$	$\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
Невырожденное нормальное, N_{a,σ^2}	$a \in \mathbf{R}$, $\sigma^2 > 0$	$y \in \mathbf{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-a)^2/2\sigma^2}$
Равномерное на отрезке $[a, b]$, $U_{a,b}$	$a, b \in \mathbf{R}$, $a < b$	$y \in [a, b]$ $y \notin [a, b]$	$(b-a)^{-1}$ 0
Бета-распределение, $B_{\alpha,\beta}$	$\alpha,\ \beta > 0$	$y \in [0, 1]$ $y \notin [0, 1]$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$ 0
Показательное (экспоненциальное), E_α	$\alpha > 0$	$y \geqslant 0$ $y < 0$	$\alpha e^{-\alpha y}$ 0
Лапласа, L_α	$\alpha > 0$	$y \in \mathbf{R}$	$(\alpha/2) e^{-\alpha\|y\|}$
Гамма, $\Gamma_{\alpha,\beta}$	$\alpha > 0,\ \beta > 0$	$y \geqslant 0$ $y < 0$	$\frac{\alpha^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\alpha y}$ 0
Коши, C_{a,σ^2}	$a \in \mathbf{R}$, $\sigma > 0$	$y \in \mathbf{R}$	$\frac{\sigma}{\pi(\sigma^2 + (y-a)^2)}$
Хи-квадрат с n степенями свободы, χ^2_n	$n \in \{1, 2, \ldots\}$	$y \geqslant 0$ $y < 0$	$\frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}$ 0
Стьюдента с n степенями свободы, t_n	$n \in \{1, 2, \ldots\}$	$y \in \mathbf{R}$	$c_n (1 + y^2/n)^{-(n+1)/2}$, $c_n = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}$
Вейбулла, $W_{\alpha,\theta}$	$\alpha > 0,\ \theta > 0$	$y \geqslant 0$ $y < 0$	$\theta\alpha y^{\alpha-1} e^{-\theta y^\alpha}$ 0
Парето, $P_{\beta,\theta}$	$\beta > 0,\ \theta > 0$	$y \geqslant \theta$ $y < \theta$	$\beta\theta^\beta y^{-(\beta+1)}$ 0

3. Таблица нормального распределения

В таблице приведены значения функции $\overline{\Phi}(y) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{y}^{\infty} e^{-z^2/2} dz.$

y	0	1	2	3	4	5	6	7	8	9
0,0	,500	,496	,492	,488	,484	,480	,476	,472	,468	,464
0,1	,460	,456	,452	,448	,444	,440	,436	,433	,429	,425
0,2	,421	,417	,413	,409	,405	,401	,397	,394	,340	,386
0,3	,382	,378	,374	,371	,370	,363	,359	,356	,352	,348
0,4	,345	,341	,337	,334	,330	,326	,323	,319	,316	,312
0,5	,309	,305	,302	,298	,295	,291	,288	,284	,281	,278
0,6	,274	,271	,268	,264	,261	,258	,255	,251	,248	,245
0,7	,242	,239	,236	,233	,230	,227	,224	,221	,218	,215
0,8	,212	,209	,206	,203	,200	,198	,195	,192	,189	,187
0,9	,184	,181	,179	,176	,174	,171	,169	,166	,164	,161
1,0	,159	,156	,154	,152	,149	,147	,145	,142	,140	,138
1,1	,136	,134	,131	,129	,127	,125	,123	,121	,119	,117
1,2	,115	,113	,111	,109	,107	,106	,104	,102	,100	,099
1,3	,097	,095	,093	,092	,090	,089	,087	,085	,084	,082
1,4	,081	,079	,078	,076	,075	,074	,072	,071	,069	,068
1,5	,067	,066	,064	,063	,062	,061	,059	,058	,057	,056
1,6	,055	,054	,053	,052	,051	,049	,048	,047	,046	,046
1,7	,045	,044	,043	,042	,041	,040	,039	,038	,038	,037
1,8	,036	,035	,034	,034	,033	,032	,031	,031	,030	,029
1,9	,029	,028	,027	,027	,026	,026	,025	,024	,024	,023
2,0	,023	,022	,022	,021	,021	,020	,020	,019	,019	,018
2,1	,018	,017	,017	,017	,016	,016	,015	,015	,015	,014
2,2	,014	,014	,013	,013	,013	,012	,012	,012	,011	,011
2,3	,011	,010	,010	,010	,010	,009	,009	,009	,009	,008
2,4	,008	,008	,008	,008	,007	,007	,007	,007	,007	,006
2,5	,006	,006	,006	,006	,006	,005	,005	,005	,005	,005
2,6	,005	,005	,004	,004	,004	,004	,004	,004	,004	,004
2,7	,003	,003	,003	,003	,003	,003	,003	,003	,003	,003
2,8	,003	,002	,002	,002	,002	,002	,002	,002	,002	,002
2,9	,002	,002	,002	,002	,002	,002	,002	,001	,001	,001

$\overline{\Phi}(3) = 0,00135$; $\overline{\Phi}(4) = 0,00003167$; $\overline{\Phi}(5) = 0,0000002867$;
$\overline{\Phi}(6) = 0,00000000099$

4. Таблица χ^2-распределения

В таблице приведены значения квантилей $z_n(p)$ уровня p распределения χ^2 с n степенями свободы, т. е. значения $z_n(p)$, для которых

$$\mathbf{P}\{\chi_n^2 < z_n(p)\} = p, \quad p \in [0, 1].$$

n \ p	0,01	0,02	0,05	0,1	0,2	0,3	0,5	0,7	0,8	0,9	0,95	0,98	0,99	,995	,999
1	,000	,001	,004	,016	,064	,148	,455	1,07	1,64	2,71	3,84	5,41	6,64	7,88	10,8
2	,020	,040	,103	,211	,446	,713	1,39	2,41	3,22	4,61	5,99	7,82	9,21	10,6	13,8
3	,115	,185	,352	,584	1,01	1,42	2,37	3,67	4,64	6,25	7,82	9,84	11,3	12,8	16,3
4	,297	,429	,711	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	11,7	13,3	14,9	18,5
5	,554	,752	1,15	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,1	13,4	15,1	16,8	20,5
6	,872	1,13	1,64	2,20	3,07	3,83	5,35	7,23	8,56	10,6	12,6	15,0	16,8	18,5	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,0	14,1	16,6	18,5	20,3	24,3
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,0	13,4	15,5	18,2	20,1	22,0	26,1
9	2,09	2,53	3,33	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7	23,6	27,9
10	2,56	3,06	3,94	4,87	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	29,6
11	3,05	3,61	4,58	5,58	6,99	8,15	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	31,3
12	3,57	4,18	5,23	6,30	7,81	9,03	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	32,9
13	4,11	4,77	5,89	7,04	8,63	9,93	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	34,5
14	4,66	5,37	6,57	7,79	9,47	10,8	13,3	16,2	18,2	21,1	23,7	26,9	29,1	31,3	36,1
15	5,23	5,99	7,26	8,55	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,8	37,7
16	5,81	6,61	7,96	9,31	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,3	39,3
17	6,41	7,26	8,67	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,7	40,8
18	7,02	7,91	9,39	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,2	42,3
19	7,63	8,57	10,1	11,7	13,7	15,4	18,3	21,7	23,9	27,2	30,1	33,7	36,2	38,6	43,8
20	8,26	9,24	10,9	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6	40,0	45,3
21	8,90	9,92	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9	41,4	46,8
22	9,54	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3	42,8	48,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6	44,2	49,7
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,6	51,2
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	46,9	52,6
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	48,3	54,1
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	49,6	55,5
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51,0	56,9
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	52,3	58,3
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	53,7	59,7
31	15,7	17,0	19,3	21,4	24,3	26,4	30,3	34,6	37,4	41,4	45,0	49,2	52,2	55,0	61,1
32	16,4	18,2	20,1	22,3	25,1	27,4	31,3	35,7	38,5	42,6	46,2	50,5	53,5	56,3	62,5

5. Таблица распределения Стьюдента

В таблице приведены значения точек $z_n(p)$ для величины t_n с распределением Стьюдента с n степенями свободы такие, что

$$\mathbf{P}\{|t_n| > z_n(p)\} = p, \quad p \in [0, 1].$$

$n \backslash p$	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	,158	,325	,510	,727	1,00	1,38	1,96	3,08	6,31	12,7	31,8	63,7	637
2	,142	,289	,445	,617	,816	1,06	1,39	1,89	2,92	4,30	6,96	9,92	31,6
3	,137	,277	,424	,584	,765	,978	1,25	1,64	2,35	3,18	4,54	5,84	12,9
4	,134	,271	,414	,569	,741	,941	1,19	1,53	2,13	2,78	3,75	4,60	8,61
5	,132	,267	,408	,559	,727	,920	1,16	1,48	2,02	2,57	3,36	4,03	6,87
6	,131	,265	,404	,553	,718	,906	1,13	1,44	1,94	2,45	3,14	3,71	5,96
7	,130	,263	,402	,549	,711	,896	1,12	1,41	1,89	2,36	3,00	3,50	5,41
8	,130	,262	,399	,546	,706	,889	1,11	1,40	1,86	2,31	2,90	3,36	5,04
9	,129	,261	,398	,543	,703	,883	1,10	1,38	1,83	2,26	2,82	3,25	4,78
10	,129	,260	,397	,542	,700	,879	1,09	1,37	1,81	2,23	2,76	3,17	4,59
11	,129	,260	,396	,540	,697	,876	1,09	1,36	1,80	2,20	2,72	3,11	4,44
12	,128	,259	,395	,539	,695	,873	1,08	1,36	1,78	2,18	2,68	3,05	4,32
13	,128	,259	,394	,538	,694	,870	1,08	1,35	1,77	2,16	2,65	3,01	4,22
14	,128	,258	,393	,537	,692	,868	1,08	1,35	1,76	2,14	2,62	2,98	4,14
15	,128	,258	,393	,536	,691	,866	1,07	1,34	1,75	2,13	2,60	2,95	4,07
16	,128	,258	,392	,535	,690	,865	1,07	1,34	1,75	2,12	2,58	2,92	4,02
17	,128	,257	,392	,534	,689	,863	1,07	1,33	1,74	2,11	2,57	2,90	3,97
18	,127	,257	,392	,534	,688	,862	1,07	1,33	1,73	2,10	2,55	2,88	3,92
19	,127	,257	,391	,533	,688	,861	1,07	1,33	1,73	2,09	2,54	2,86	3,88
20	,127	,257	,391	,533	,687	,860	1,06	1,33	1,72	2,09	2,53	2,85	3,85
21	,127	,257	,391	,532	,686	,859	1,06	1,32	1,72	2,08	2,52	2,83	3,82
22	,127	,256	,390	,532	,686	,858	1,06	1,32	1,72	2,07	2,51	2,82	3,79
23	,127	,256	,390	,532	,685	,858	1,06	1,32	1,71	2,07	2,50	2,81	3,77
24	,127	,256	,390	,531	,685	,857	1,06	1,32	1,71	2,06	2,49	2,80	3,75
25	,127	,256	,390	,531	,684	,856	1,06	1,32	1,71	2,06	2,49	2,79	3,73
26	,127	,256	,390	,531	,684	,856	1,06	1,32	1,71	2,06	2,48	2,78	3,71
27	,127	,256	,389	,531	,684	,855	1,06	1,31	1,70	2,05	2,47	2,77	3,69
28	,127	,256	,389	,530	,683	,855	1,06	1,31	1,70	2,05	2,47	2,76	3,67
29	,127	,256	,389	,530	,683	,854	1,06	1,31	1,70	2,05	2,46	2,76	3,66
30	,127	,256	,389	,530	,683	,854	1,06	1,31	1,70	2,04	2,46	2,75	3,65
40	,126	,255	,388	,529	,681	,851	1,05	1,30	1,68	2,02	2,42	2,70	3,55
60	,126	,254	,387	,527	,679	,848	1,05	1,30	1,67	2,00	2,39	2,66	3,46
120	,126	,254	,386	,526	,677	,845	1,04	1,29	1,66	1,98	2,36	2,62	3,37
∞	,126	,253	,385	,524	,674	,842	1,04	1,28	1,65	1,96	2,33	2,58	3,29

6. Таблица распределения Колмогорова

В таблице приведены значения функции

$$K(y) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2y^2}, \quad y > 0.$$

y	0	1	2	3	4	5	6	7	8	9
0,3	,0000	,0000	,0000	,0001	,0002	,0003	,0005	,0008	,0013	,0019
0,4	,0028	,0040	,0055	,0074	,0097	,0126	,0160	,0200	,0247	,0300
0,5	,0361	,0428	,0503	,0585	,0675	,0772	,0876	,0987	,1104	,1228
0,6	,1357	,1492	,1632	,1778	,1927	,2080	,2236	,2396	,2558	,2722
0,7	,2888	,3055	,3223	,3391	,3560	,3728	,3896	,4064	,4230	,4395
0,8	,4559	,4720	,4880	,5038	,5194	,5347	,5497	,5645	,5791	,5933
0,9	,6073	,6209	,6343	,6473	,6601	,6725	,6846	,6964	,7079	,7191
1,0	,7300	,7406	,7508	,7608	,7704	,7798	,7889	,7976	,8061	,8143
1,1	,8223	,8300	,8374	,8445	,8514	,8580	,8644	,8706	,8765	,8823
1,2	,8878	,8930	,8981	,9030	,9076	,9121	,9164	,9206	,9245	,9283
1,3	,9319	,9354	,9387	,9418	,9449	,9478	,9505	,9531	,9557	,9580
1,4	,9603	,9625	,9646	,9665	,9684	,9702	,9718	,9734	,9750	,9764
1,5	,9778	,9791	,9803	,9815	,9826	,9836	,9846	,9855	,9864	,9873
1,6	,9880	,9888	,9895	,9902	,9908	,9914	,9919	,9924	,9929	,9934
1,7	,9938	,9942	,9946	,9950	,9953	,9956	,9959	,9962	,9965	,9967
1,8	,9969	,9971	,9973	,9975	,9977	,9979	,9980	,9981	,9983	,9984
1,9	,9985	,9986	,9987	,9988	,9989	,9990	,9991	,9991	,9992	,9992
2,0	,9993	,9994	,9994	,9995	,9995	,9996	,9996	,9996	,9997	,9997
2,1	,9997	,9997	,9998	,9998	,9998	,9998	,9998	,9998	,9999	,9999

$K(2,2) = 0,999874$; $K(2,25) = 0,999920$;
$K(2,3) = 0,999949$; $K(2,35) = 0,999968$;
$K(2,4) = 0,999980$; $K(2,45) = 0,999988$;
$K(2,49) = 0,999992$

СПИСОК ЛИТЕРАТУРЫ

1. Беляев Ю. К., Носко В. П. *Основные понятия и задачи математической статистики.* М.: Изд-во Московского ун-та, 1998.

2. Бикел П., Доксам К. *Математическая статистика.* Выпуск 1, 2. М.: Финансы и статистика, 1983.

3. Большев Л. Н., Смирнов Н. В. *Таблицы математической статистики.* М.: Наука, 1965.

4. Боровков А. А. *Математическая статистика.* Новосибирск: Наука; Изд-во Института математики, 1997.

5. Ван дер Варден Б. *Математическая статистика.* М.: Иностр. лит., 1960.

6. *Введение в теорию порядковых статистик.* Под редакцией Е. Сархана и Б. Гринберга. М.: Статистика, 1970.

7. Дэйвид Г. *Порядковые статистики.* М.: Наука, 1979.

8. Емельянов Г. В., Скитович В. П. *Задачник по теории вероятностей и математической статистике.* Л.: Изд-во Ленинградского ун-та, 1967.

9. Закс Ш. *Теория статистических выводов.* М.: Мир, 1975.

10. Зубков А. М., Севастьянов Б. А., Чистяков В. П. *Сборник задач по теории вероятностей.* М.: Наука, 1989.

11. Ивченко Г. И., Медведев Ю. И. *Математическая статистика.* М.: Высшая школа, 1984.

12. Ивченко Г. И., Медведев Ю. И., Чистяков А. В. *Сборник задач по математической статистике.* М.: Высшая школа, 1989.

13. Кокс Д., Снелл Э. *Прикладная статистика. Принципы и примеры.* М.: Мир, 1984.

14. Кокс Д., Хинкли Д. *Задачи по теоретической статистике с решениями.* М.: Мир, 1981.

15. Коршунов Д. А., Фосс С. Г. *Сборник задач и упражнений по теории вероятностей.* Новосибирск: Изд-во НИИ МИОО НГУ, 1997. (2-е изд., испр. Новосибирск: Новосиб. гос. ун-т, 2003).
16. Крамер Г. *Математические методы статистики.* М.: Мир, 1975.
17. Леман Э. *Проверка статистических гипотез.* М.: Наука, 1964.
18. Мешалкин Л. Д. *Сборник задач по теории вероятностей.* М.: Изд-во Московского ун-та, 1963.
19. *Сборник задач по математической статистике.* Учебное пособие под редакцией А. А. Боровкова. Новосибирск: Новосибирский государственный ун-т, 1989.
20. *Сборник задач по теории вероятностей, математической статистике и теории случайных функций.* Под редакцией А. А. Свешникова. М.: Наука, 1965.
21. Секей Г. *Парадоксы в теории вероятностей и математической статистике.* М.: Мир, 1990.
22. Уилкс С. *Математическая статистика.* М.: Наука, 1967.
23. Феллер В. *Введение в теорию вероятностей и её приложения. Т.* 2. М.: Мир, 1984.
24. Чибисов Д. М., Пагурова В. И. *Задачи по математической статистике.* М.: Изд-во Московского ун-та, 1990.
25. Dacunha-Castelle D., Duflo M. *Exercices de probabilités et statistiques. Tome* 1*. Problèmes à temps fixe.* Paris e.a.: Masson, 1982.
26. Garthwaite P. H., Jolliffe I. H., Jones B. *Statistical Inference.* Prentice Hall, 1995.

ОТВЕТЫ

§ 1. Выборка и вариационный ряд

1.1. а), б), г), е–и) Да; в), д) нет. **1.2.** б), в), е), ж), и) Да; а), г), д), з) нет. **1.3.** а) a, σ^2/n, $N_{a,\sigma^2/n}$; б) a; в) $\sigma^2(n-1)/n$, σ^2. **1.4.** λ, λ/n, нет, нет. **1.5.** $(a+b)/2$, $(b-a)^2/12n$, нет, нет. **1.6.** $U_{0,1}$. **1.7.** $U_{0,1}$. **1.8.** $U_{0,1}$. **1.9.** E_1. **1.10.** $U_{0,1}$. **1.11.** $U_{0,1}$. **1.12.** $\mathbf{P}\{Y_1 = 1-p\} = 1 - \mathbf{P}\{Y_1 = 0\} = p$. **1.13.** $\mathbf{P}\{Y_1 = 0\} = e^{-\lambda}$; $\mathbf{P}\Big\{Y_1 = \sum_{i=0}^{k-1} \lambda^i e^{-\lambda}/i!\Big\} = \lambda^k e^{-\lambda}/k!$, $k \geqslant 1$. **1.14.** Если $y_1 < \dots < y_n$, то $n!f(y_1)\cdot\ldots\cdot f(y_n)$, иначе 0. **1.15.** а) $F^n(y)$; б) $1-(1-F(y))^n$. **1.16.** $C_n^k F^k(y)(1-F(y))^{n-k}$. **1.17.** $\sum_{i=k}^n C_n^i F^i(y)(1-F(y))^{n-i}$. **1.18.** а) $n(\theta-y)^{n-1}/\theta^n$; б) ny^{n-1}/θ^n; в) $nC_{n-1}^{k-1}y^{k-1}(\theta-y)^{n-k}/\theta^n$. **1.19.** а) $n(1-F(y))^{n-1}f(y)$; б) $nF^{n-1}(y)f(y)$; в) $nC_{n-1}^{k-1}F^{k-1}(y)(1-F(y))^{n-k}f(y)$. **1.20.** а) $\theta/(n+1)$, $2\theta^2/(n+1)(n+2)$, $n\theta^2/(n+1)^2(n+2)$; б) $n\theta/(n+1)$, $n\theta^2/(n+2)$, $n\theta^2/(n+1)^2(n+2)$; в) $k\theta/(n+1)$, $k(k+1)\theta^2/(n+1)(n+2)$, $k(n-k+1)\theta^2/(n+1)^2(n+2)$. **1.21.** $\mathbf{P}\{X_{(k)} > l\} = \sum_{i=0}^{k-1} C_n^i \Big(\sum_{m=0}^{l} p_m\Big)^i \Big(\sum_{m=l+1}^{N} p_m\Big)^{n-i}$. **1.22.** $\mathbf{P}\{X_{(1)} < y, X_{(n)} < z\} = F^n(z) - (F(z)-F(y))^n$ в случае $y < z$ и $\mathbf{P}\{X_{(1)} < y, X_{(n)} < z\} = F^n(z)$ иначе. **1.23.** а) $n(n-1)(z-y)^{n-2}/\theta^n$ при $0 \leqslant y < z \leqslant \theta$; б) $\theta^2/(n+1)^2(n+2)$; в) $n(n-1)C_{n-2}^{k-1}C_{n-k-1}^{j-k-1}y^{k-1}(z-y)^{j-k-1}(\theta-z)^{n-j}/\theta^n$ при $0 \leqslant y < z \leqslant \theta$; г) $k(n-j+1)\theta^2/(n+1)^2(n+2)$. **1.24.** б) $E_{n\alpha}$; в) $E_{(n-k)\alpha}$. **1.28.** а), б) E_1. **1.30.** а), б) $\Gamma_{1,k}$. **1.31.** Вектор с независимыми координатами, первая имеет распределение $\Gamma_{1,k}$, вторая — $\Gamma_{1,j}$. **1.32.** Вектор $(\xi_1,\ \xi_1+\xi_2)$, где ξ_1 и ξ_2 независимы и имеют распределения $\Gamma_{1,k}$ и $\Gamma_{1,j-k}$. **1.34.** Нулевой вектор средних значений; дисперсии $p(1-p)$ и $s(1-s)$; ковариация $p(1-s)$. **1.35.** Предельная функция распределения равна $e^{-e^{-x}}$, $x \in \mathbf{R}$.

§ 2. Эмпирическая функция распределения

2.3. $F_n^*(y) = 0$ при $y \leqslant 0$, $F_n^*(y) = 1 - \overline{X}$ при $0 < y \leqslant 1$ и $F_n^*(y) = 1$ при $y > 1$. **2.4.** (1, 1, 5, 7, 8, 8), (1, 5, 1, 7, 8, 8). **2.5.** Нет, да, нет, нет. **2.6.** Да; $(X_1/a, \dots, X_n/a)$. **2.8.** а) Да, $(\sqrt[3]{X_1}, \dots, \sqrt[3]{X_n})$;

б) да, выборка объёма n^3, в которой $X_{(k)}$ повторяется $k^3-(k-1)^3$ раз. **2.9.** Да; объединённой выборке $(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$. **2.12.** а) $F(y)$; б) $F(y)(1-F(y))/n$; в) $(F(z)-F(y))(1-F(z)+F(y))/n$, если $y<z$. **2.13.** $C_n^k(C_m^y p^y(1-p)^{m-y})^k(1-C_m^y p^y(1-p)^{m-y})^{n-k}$ при $y\in\{0,\ldots,m\}$; 0 — иначе. **2.14.** $1-(1-F(z)+F(y))^n$, если $y<z$. **2.15.** 0. **2.21.** $c_n=n(1-1/e^2)$, $N_{0,1/e^2-1/e^4}$; $c_n=n(1-1/e^2)+13\sqrt{n}$, $N_{-13,1/e^2-1/e^4}$.

§ 3. Метод моментов

3.1. а) $\overline{X}$; б) $\overline{X^2}-a^2$; в) $\overline{X}$, S^2. **3.2.** а) $(\pi/2)(\overline{|X-a|})^2$; б) $\overline{(X-a)^2}$. **3.3.** а) $\max(0,\overline{X})$, $\sqrt{1+\overline{X^2}}-1$; б) $\max(0,\overline{X})$; $\sqrt{\overline{X^2}/2}$. **3.4.** $\sqrt[k]{\overline{X^{2k}}/(2k-1)!!}$. **3.5.** а) $2\overline{X}$; б), в) $\overline{X}$; г) $\sqrt{3\overline{X^2}}$. **3.6.** а) $a^*=\overline{X}-\sqrt{3S^2}$, $b^*=\overline{X}+\sqrt{3S^2}$; б) $a^*=\overline{X}-\sqrt{3S^2}$, $b^*=2\sqrt{3S^2}$. **3.7.** $\sqrt[k]{(k+1)\overline{X^k}}$. **3.8.** $1/\overline{X}$. **3.9.** $\overline{X}-1$. **3.10.** $\alpha^*=\sqrt{S^2}$, $\beta^*=\overline{X}-\sqrt{S^2}$. **3.11.** а) $\sqrt[k]{k!/\overline{X^k}}$; б) $\sqrt[k]{\overline{X^k}/k!}$. **3.12.** y^2, $\sqrt{2/\overline{X^2}}$. **3.13.** $(\overline{X})^2$. **3.14.** $e^{-1/\overline{X}}$. **3.15.** а) $\alpha^*=\beta/\overline{X}$; б) $\beta^*=\alpha\overline{X}$; в) $\alpha^*=\overline{X}/S^2$, $\beta^*=(\overline{X})^2/S^2$. **3.16.** а) $\overline{X}/(\overline{X}-\theta)$; б) $\overline{X}(1-1/\beta)$; в) $\beta^*=1+\sqrt{1+(\overline{X})^2/S^2}$, $\theta^*=\overline{X}(1-1/\beta^*)$. **3.17.** $1/\overline{X^\alpha}$. **3.19.** $\sqrt[k]{\overline{X^k}/\Gamma(1+k/3)}$. **3.20.** а) $\overline{X}/(1-\overline{X})$; б) $3\overline{X}/2$. **3.21.** Нет. **3.22.** $\overline{X}$. **3.23.** Нет. **3.24.** а) $\overline{X}/m$; б) ближайшее целое к числу $\overline{X}/p$; в) $p_n^*=1-S^2/\overline{X}$, ближайшее целое к числу $m_n^*=(\overline{X})^2/(\overline{X}-S^2)$. **3.25.** $e^{\overline{X}}$. **3.26.** а) $\overline{X}$; б) $\sqrt{\overline{X^2}+1/4}-1/2$. **3.27.** $e^{\overline{X}}$. **3.28.** $\theta_n^*=\overline{\mathbf{I}\{X=1\}}$. **3.29.** $\lambda_1^*=\overline{X}-\sqrt{S^2-\overline{X}}$, $\lambda_2^*=\overline{X}+\sqrt{S^2-\overline{X}}$. **3.30.** $1/(\overline{X}+1)$. **3.31.** $\theta_n^*=(b-\overline{X})/(b-a)$, если $a\leqslant\overline{X}\leqslant b$; 0, если $\overline{X}<a$; 1, если $\overline{X}>b$. **3.32.** Для параметра α распределения Лапласа.

§ 4. Метод максимального правдоподобия

4.1. $\overline{X}$, S^2. **4.2.** $\overline{(X-a)^2}$. **4.3.** а) $\sqrt{1+\overline{X^2}}-1$; б) $\frac{1}{2}\left(\sqrt{(\overline{X})^2+4\overline{X^2}}-\overline{X}\right)$. **4.4.** $X_{(n)}$. **4.5.** а) $-X_{(1)}$; б) $\max\{-X_{(1)},X_{(n)}\}=\max\{|X_i|\}$; в) любая точка отрезка $[X_{(n)}-2,\ X_{(1)}]$; г) $X_{(n)}/2$. **4.6.** $a_n^*=X_{(1)}$, $b_n^*=X_{(n)}$. **4.7.** $1/\overline{X}$. **4.8.** $X_{(1)}$. **4.9.** $\alpha_n^*=\overline{X}-X_{(1)}$, $\beta_n^*=X_{(1)}$. **4.10.** Выборочная медиана. **4.11.** $\mu^*=$ выборочная медиана, $\sigma^*=n^{-1}\sum_{i=1}^n|X_i-\mu^*|$. **4.12.** $\alpha^*=\beta/\overline{X}$. **4.13.** а) $1/(\overline{\ln X}-\ln\theta)$; б) $X_{(1)}$; в) $(1/(\overline{\ln X}-\ln X_{(1)}),\ X_{(1)})$. **4.14.** $1/\overline{X^\alpha}$. **4.15.** $\sqrt[3]{\overline{X^3}}$. **4.16.** $\overline{g(X)}$. **4.17.** а) $-1/\overline{\ln X}$; б) $X_{(n)}$; в)

$1/\sqrt{\overline{X^{-1}}}$; г) $-1/\overline{\ln\ln X}$; д) $\max\limits_{1\leqslant i\leqslant n}\{|X_i|\}$. **4.18.** а) X_1; б) $(X_1+X_2)/2$, если $|X_1-X_2|\leqslant 2$; $(X_1+X_2)/2\pm\sqrt{(X_1-X_2)^2/4-1}$ иначе (если $|X_1-X_2|>2$). **4.19.** См. ответ к задаче 4.18б). **4.20.** $\overline{X}$. **4.21.** а) $\overline{X}/m$; б) $[X_1/p]$, если X_1/p не целое, $X_1/p-1$ или X_1/p, если X_1/p целое. **4.22.** $\overline{X}$. **4.23.** $1/(\overline{X}+1)$. **4.24.** $p_n^*=\nu_n/(n\overline{X}+\nu_n)$, где ν_n — количество элементов выборки, отличных от m. **4.25.** $X_{(n)}$. **4.26.** $a_n^*=1$, если $\overline{X}<3/2$; $a_n^*=2$, если $\overline{X}\geqslant 3/2$. **4.27.** $\theta^*=1$, если $\overline{X}>\ln 2$; $\theta^*=2$, если $\ln 3/2<\overline{X}<\ln 2$; $\theta^*=3$, если $\overline{X}<\ln 3/2$. **4.28.** $\overline{\mathbf{I}\{X\neq 3\}}/3$. **4.29.** X_1. **4.31.** $U_{\theta,\theta+7}$. **4.32.** $U_{\theta,\theta+7}$.

§ 5. Байесовские оценки

5.2. $\frac{n\overline{X}\sigma^2+b}{n\sigma^2+1}$. **5.3.** $\left(1+e^{n/2-n\overline{X}}\right)^{-1}$. **5.4.** а) $\frac{n+1}{n}\cdot\max(X_{(n)},1)$; б) $\frac{n-1}{n-2}\frac{X_{(n)}-X_{(n)}^{n-1}}{1-X_{(n)}^{n-1}}$. **5.5.** $\theta_n^*=1+(1+2^n)^{-1}$, если $X_{(n)}\leqslant 1$; $\theta_n^*=2$, если $1<X_{(n)}\leqslant 2$. **5.6.** $(n+1)/(n\overline{X}+\beta)$. **5.7.** $ae^{na}/(e^{na}-1)-1/n$, где $a=\min(1,X_{(1)})$. **5.8.** а) $\frac{2n-1}{2n-2}\frac{X_{(n)}^{2n-1}-X_{(n)}}{X_{(n)}^{2n-1}-1}$; б) $\frac{2n-3}{2n-4}\frac{X_{(n)}^{2n-3}-X_{(n)}}{X_{(n)}^{2n-3}-1}$; в) $\max(X_{(n)},1)\cdot\frac{\beta+2n}{\beta+2n-1}$. **5.9.** $(n\overline{X}+1)/(n+2)$. **5.10.** $(2^{2n+1-n\overline{X}}+3^{n+1})/6(2^{2n-n\overline{X}}+3^n)$. **5.11.** $(n\overline{X}+\lambda)/(n+1+\lambda)$. **5.12.** $(n\overline{X}+1)/(n+1)$. **5.13.** $(e^n+2^{n\overline{X}+2})/(e^n+2^{n\overline{X}+1})$. **5.14.** $(3^{n\overline{X}}+4\cdot 2^{n\overline{X}}+9)/4(3^{n\overline{X}}+2\cdot 2^{n\overline{X}}+3)$.

§ 6. Несмещённость и состоятельность

6.1. Смещённая и состоятельная. **6.2.** а), г), д) Несмещённая и состоятельная; б) смещённая и состоятельная; в) несмещённая и несостоятельная. **6.4.** Нет; да. **6.5.** Нет; да. **6.6.** Нет; да. **6.9.** Да; да (да). **6.10.** Нет; S_0^2. **6.11.** а), б), в) Несмещённая и состоятельная; г) смещённая и состоятельная. **6.12.** Несмещённая и состоятельная. **6.13.** Несмещённая и состоятельная. **6.15.** а) 814,86 *м*^{2}; б) 921,84 *м*^{2}. **6.16.** Несмещённая, состоятельная. **6.17.** Несмещённая, состоятельная. **6.18.** а), б) Да, да. **6.19.** Нет, $\alpha/(n-1)$; да. **6.20.** $\theta=e^{1/\alpha}$; нет. **6.21.** Нет; да. **6.22.** Смещённые и состоятельные при любом k. **6.23.** а) Смещённая, состоятельная; б) несмещённая, состоятельная. **6.24.** Все четыре оценки состоятельные и смещённые. **6.25.** Состоятельные. **6.26.** Обе смещённые и состоятельные. **6.27.** Смещённая и состоятельная. **6.28.** Нет. **6.29.** Состоятельная и несмещённая. **6.30.** Нет; да. **6.32.** Нет. **6.33.** $b_n(p)=(\alpha-p\beta)/(n+\beta)$, $\mathbf{E}(p_n^*-p)^2=(np(1-p)+(\alpha-p\beta)^2)/(n+\beta)^2$. **6.34.** $\theta=e^{2p}$; нет. **6.35.** Вторая — да, первая и третья — нет. **6.36.** $\theta=\lambda^5$; нет.

6.37. $\theta = \lambda e^{-\lambda}$, нет. **6.38.** $\theta = \lambda e^{-\lambda}$; да. **6.39.** Нет; да. **6.41.** а) $\frac{n+3}{n+4}\overline{X}$; б) $(X_1 + X_3)/2$. **6.43.** Нет; да. **6.44.** Нет; да. **6.45.** Смещённая и состоятельная. **6.46.** Несмещённая и состоятельная. **6.48.** $B_{1/2}$, $\delta = 1/2$. **6.49.** Несмещённая, состоятельная. **6.51.** $\theta^*/3$. **6.53.** Распределение θ^* невырождено, а функция f не является линейной на множестве Θ. **6.54.** а) $U_{0,\theta}$, $g(y) = y^9$; б) B_p; в) Π_λ, $\lambda^* = X_7$; г) Π_λ, $\lambda^*_n = \overline{X} + 1/n$.

§ 7. Асимптотическая нормальность

7.1. $\mathbf{D}X_1$. **7.2.** $\mathbf{D}g(X_1)$. **7.3.** $\mathbf{D}(X_1 - a)^2$. **7.6.** $4\sigma^2\theta^2$. **7.7.** Только при $\theta \neq 0$. **7.10.** Да; $\sigma^2(\pi/2 - 1)$. **7.11.** Да; $4\sigma^4$. **7.12.** $\theta^2/(2k+1)$. **7.13.** Нет. **7.14.** Нет. **7.15.** $1/27$. **7.16.** $\theta = \ln(a/2)$, $\sigma^2(a) = 1/3$. **7.17.** $\dfrac{(2k)! - (k!)^2}{k^2(k!)^2}\,\alpha^2$. **7.18.** 1. **7.19.** $\theta = e^{-2/\alpha^2}$, $\sigma^2(\alpha) = 20\,e^{-4/\alpha^2}/\alpha^4$. **7.20.** а) Нет; б) да, 1. **7.21.** а) α^* — да, $2\alpha^2$; б) α^* — да, α^2, β^* — нет. **7.22.** β^*_n — да, $\sigma^2(\beta) = \beta^2$; θ^*_n — нет. **7.23.** Да, θ^2. **7.24.** $1/4$. **7.25.** 1. **7.26.** $\theta = e^{mp}$, $\sigma^2(m,p) = mp(1-p)e^{2mp}$. **7.27.** λ. **7.28.** $1/4$. **7.29.** $\theta = \lambda e^{-\lambda}$, $\sigma^2(\lambda) = \lambda(1-\lambda)^2 e^{-2\lambda}$. **7.30.** $\overline{X} + 5/\sqrt[5]{n}$. **7.31.** Да, $p^2(1-p)$. **7.32.** $N_{0,9\sigma^2}$. **7.35.** θ^2. **7.36.** $\pi^2/4$. **7.37.** $1/\alpha^2$. **7.38.** $1/4f^2(\zeta)$. **7.39.** $\delta(1-\delta)/f^2(\zeta_\delta)$. **7.40.** $F(y)(1 - F(y))$.

§ 8. Среднеквадратический подход

8.1. Вторая оценка лучше. **8.2.** $\mathbf{D}S_0^2 = 2\sigma^4/(n-1)$, $\mathbf{D}S_1^2 = 2\sigma^4/n$. **8.3.** $\mathbf{D}S_0^2 = 2\sigma^4/(2n-1)$, $\mathbf{D}(\sigma^2)^*_{2n} = 2\sigma^4/n$. **8.4.** $c_n = 1/(n+1)$; смещение$= -2\sigma^2/(n+1)$. **8.5.** Среднеквадратические отклонения: $\theta^2/3n$, $2\theta^2/(n+1)(n+2)$, $\theta^2/n(n+2)$, $2\theta^2/(n+1)(n+2)$. **8.6.** $\theta^*_{1,n}$ — наилучшая; $\theta^*_{0,n}$ лучше, чем $\theta^*_{2,n}$; $\theta^*_{k,n}$ лучше, чем $\theta^*_{k+1,n}$ при $k \geqslant 2$. **8.7.** $c_n = (n+2)/(n+1)$; смещение$= -1/(n+1)^2$. **8.8.** Для наилучшей в среднеквадратичном оценки $a = (n+1)/(5n+4)$, $b = 2a$, $\mathbf{E}_\theta(\theta^* - \theta)^2 = 1/(n+2)(5n+4)$. **8.9.** а) Вторая и третья эквивалентны в среднеквадратичном смысле и лучше, чем первая; б) $(X_{(1)} + X_{(n)} - 1)/2$. **8.10.** $\mathbf{E}(\overline{X} - 1 - \theta)^2 = 1/n$, $\mathbf{E}(X_{(1)} - \theta)^2 = 2/n^2$, $\mathbf{E}(X_{(1)} - 1/n - \theta)^2 = 1/n^2$. **8.11.** Например, $\lambda^*_1 = (X_1 + X_2)/2$ лучше в среднеквадратичном, чем $\lambda^*_2 = X_1$. **8.12.** Распределение Бернулли с параметром $\theta \in (0,1)$, $\theta^*_1 = \overline{X} + 33$, $\theta^*_2 = X_1$.

§ 9. Асимптотический подход

9.1. Вторая оценка лучше. **9.2.** Среднее лучше. **9.3.** Выборочная медиана лучше. **9.4.** Нет. **9.5.** Среднее лучше. **9.6.** Да, при $k = 1$.

§ 10. Достаточные статистики

10.1. Вырожденное в точке $(x_1, \ldots, x_n)$; а), б) да. **10.3.** Условное распределение $\mathbf{P}\{X_1 \in B_1, \ldots, X_n \in B_n \mid n\overline{X} = y\}$ есть (обобщённое) нормальное распределение с вырожденной матрицей ковариаций σ^2, диагональные элементы которой $\sigma^2_{ii} = (n-1)/n$, внедиагональные элементы $\sigma^2_{ij} = -1/n$, $i \neq j$, и вектором средних $(y/n, \ldots, y/n)$. Корень из σ^2 совпадает с σ^2, так что данное условное распределение совпадает с распределением вектора $(\xi_1, \ldots, \xi_n) \cdot \sigma^2 + (y/n, \ldots, y/n) = (\xi_1 - \overline{\xi} + y/n, \ldots, \xi_n - \overline{\xi} + y/n)$, где ξ_i — независимые в совокупности случайные величины со стандартным нормальным распределением; да. **10.4.** $\overline{X^2}$ **10.5.** а) Нет; б) да; в) нет. **10.6.** $(\overline{X}, \overline{X^2})$. **10.7.** $X_{(n)}$. **10.8.** Нет, нет, да. **10.9.** а), б) $(X_{(1)}, X_{(n)})$. **10.10.** $\max\{-X_{(1)}, X_{(n)}\} = \max|X_i|$. **10.11.** $S = \overline{X}$. **10.12.** $2X_{(1)}$; нет. **10.13.** а) $X_{(1)}$; б) $\overline{X}$; в) $(X_{(1)}, \overline{X})$. **10.14.** Да; Г-распределение с параметрами $n\beta/\theta$ и $n\beta$; да. **10.15.** $(\overline{X}, \overline{\ln X})$. **10.16.** а) $\overline{\ln X}$; б) $X_{(1)}$; в) $(\overline{\ln X}, X_{(1)})$. **10.17.** $(\overline{\ln X}, \overline{X^\alpha})$. **10.18.** $\overline{\ln X}$. **10.19.** Да. **10.20.** $\mathbf{P}\{\vec{X} = (k_1, \ldots, k_n) \mid n\overline{X} = k\} = \prod_{i=1}^n C_n^{k_i}/C_{nm}^k$, если $k_1 + \cdots + k_n = k$; да. **10.21.** Полиномиальное распределение: $\mathbf{P}(\vec{X} = (k_1, \ldots, k_n) \mid n\overline{X} = k) = \frac{k!}{k_1! \cdot \ldots \cdot k_n!} \frac{1}{n^k}$, если $k_1 + \cdots + k_n = k$; $\overline{X}$, $(\overline{X})^2$, $\sin \overline{X}$ — достаточные (так как число 2π иррационально), $\overline{X^2}$ — нет. **10.22.** а), в), г), д), е) Да; б) нет. **10.24.** $\mathbf{P}\{\vec{X} = (k_1, \ldots, k_n) \mid n\overline{X} = k\} = 1/C^k_{n+k-1}$, если $k_1 + \cdots + k_n = k$, — равновероятное распределение на множестве наборов натуральных чисел $\{(k_1, \ldots, k_n) : \sum k_i = k\}$; да. **10.26.** $X_{(n)}$. **10.27.** Распределение Коши с параметром сдвига a и параметром масштаба 1.

§ 11. Полные статистики

11.6. в) $(X_{(1)}, \overline{X})$. **11.7.** Нет.

§ 12. Эффективные оценки

12.1. $\theta^*/(\alpha + 1)$. **12.3.** $\overline{X}$, $N(a, 1/n)$; да. **12.4.** $\frac{n+1}{n} X_{(n)}$. **12.5.** Смещение 0, дисперсия $\theta^2/12$; улучшенная оценка $\frac{n+1}{n} X_{(n)}$, смещение 0, дисперсия $\theta^2/n(n+2)$, да. **12.6.** $(1 - (n-1)y/nX_{(n)})\mathbf{I}\{X_{(n)} \geqslant y\}$. **12.7.** $(n-1)/n\overline{X}$. **12.8.** $X_{(1)} - 1/n$. **12.9.** а) $X_{(1)} - \alpha/n$; б) $\overline{X} - \beta$; в) $\alpha^* = (n-1)(\overline{X} - X_{(1)})/n$, $\beta^* = (nX_{(1)} - \overline{X})/(n-1)$. **12.10.** $(1 - 1/n\beta)X_{(1)}$. **12.11.** $\overline{X^\alpha}$. **12.12.** а) $\overline{g(X)}$; б) $\overline{(g(X) - \theta)^2}$. **12.13.** $-\overline{\ln X}$. **12.14.** $\overline{X}$. **12.15.** а) Смещение $p(m-1)$, дисперсия $mp(1-p)$; $\overline{X}$, смещение $p(m-1)$, дисперсия $mp(1-p)/n$; эффективна в классе оценок со смещением

$p(m-1)$. б) Смещение 0, дисперсия $p(1-p)/m$; $\overline{X}/m$, смещение 0, дисперсия $p(1-p)/nm$; эффективна в классе несмещённых оценок. **12.16.** $\overline{X}$, смещение 0, дисперсия λ/n; да. **12.17.** $b_n(\theta) = 0$, $\theta_n^{**} = (1-1/n)^{n\overline{X}}$. **12.18.** в) 0; $(n-1)/(n\overline{X}+n-1)$.

§ 13. Неравенство Рао – Крамера

13.1. а) Добавление произвольной постоянной к оценке: оставляет дисперсию без изменения, не изменяет $b'(\cdot)$ и произвольно меняет $b(\cdot)$; б) оценка, принимающая постоянное значение, имеет нулевую дисперсию и для неё $b'(\cdot) = -1$; в) граница должна быть неотрицательной. **13.2.** $n\theta_n^*/(n+1)$. **13.3.** Оценка максимального правдоподобия для параметра θ равномерного распределения на отрезке $[3, \theta+5]$. **13.4.** Нет. В предыдущем примере $\mathbf{D}\theta_n^* \sim c/n^2$. **13.5.** а), г), е), ж), и) Да; б), в), д), з) нет. **13.6.** R-эффективна. **13.7.** а) R-эффективна; б) нет. **13.8.** а–в) Нет; нет. **13.9.** Нет. **13.10.** Нет; да. **13.11.** Да; да. **13.12.** Нет; нет; нет; да. **13.13.** Нет; да. **13.14.** Нет. **13.15.** б) $\pi^2/3n$; в) 1/3; г) не является. **13.20.** R-эффективна. **13.21.** R-эффективна. **13.22.** R-эффективна. **13.23.** 0; нет. **13.24.** Да. **13.26.** а–е) Да; ж) нет.

§ 14. Доверительные интервалы

14.1. $(\overline{X} - \sigma\zeta_{1-\varepsilon/2}/\sqrt{n},\ \overline{X} + \sigma\zeta_{1-\varepsilon/2}/\sqrt{n})$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **14.2.** $(nS_1^2/\zeta_{1-\varepsilon/2},\ nS_1^2/\zeta_{\varepsilon/2})$, где ζ_δ — квантиль уровня δ распределения χ^2 с n степенями свободы. **14.3.** $(n(\overline{X}-a)^2/\zeta^2_{1-\varepsilon/4},\ n(\overline{X}-a)^2/\zeta^2_{0,5+\varepsilon/4})$, где ζ_δ — квантиль уровня δ стандартного нормального распределения; первый. **14.4.** $c = 1$. **14.5.** $c = 2$. **14.6.** $c = 2$. **14.7.** Для a: $(\overline{X} - S_0\zeta_{1-\varepsilon/2}/\sqrt{n},\ \overline{X} + S_0\zeta_{1-\varepsilon/2}/\sqrt{n})$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения Стьюдента с $n-1$ степенью свободы, $S_0 = \sqrt{S_0^2}$. Для σ^2: $(nS^2/\zeta_{1-\varepsilon/2},\ nS^2/\zeta_{\varepsilon/2})$, где ζ_δ — квантиль уровня δ распределения χ^2 с $n-1$ степенью свободы. **14.9.** $(1/\ln 20,\ 1/(\ln 20 - \ln 19))$. **14.10.** а) $(2\overline{X} - \sqrt{1/3n\varepsilon},\ 2\overline{X} + \sqrt{1/3n\varepsilon})$; б) $(X_{(n)},\ X_{(n)} + /(n+1)\varepsilon)$. **14.11.** $(X_1,\ X_1/\varepsilon)$. **14.12.** $(X_{(n)},\ X_{(n)}/\sqrt[n]{\varepsilon})$. **14.14.** а) $(X_{(1)} - 1 + \sqrt[n]{\varepsilon},\ X_{(1)})$; б) $(X_{(1)}/(2-\sqrt[n]{\varepsilon}),\ X_{(1)})$. **14.15.** $(X_{(1)} + (\ln\varepsilon)/n,\ X_{(1)})$. **14.16.** $(0,\ -(\ln\varepsilon)/X_1)$; $(0,\ -(\ln\varepsilon)/nX_{(1)})$.

§ 15. Асимптотические доверительные интервалы

15.1. $\left(\overline{X} - \zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X})}/\sqrt{n},\ \overline{X} + \zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X})}/\sqrt{n}\right)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$.

15.2. (0,061; 0,139). **15.3.** $\big(\overline{X}/m-\zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X}/m)}/\sqrt{n},\ \overline{X}/m+\zeta_{1-\varepsilon/2}\sqrt{\overline{X}(1-\overline{X}/m)}/\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.4.** $\big(\overline{X}-\zeta_{1-\varepsilon/2}\sqrt{\overline{X}}/\sqrt{n},\ \overline{X}+\zeta_{1-\varepsilon/2}\sqrt{\overline{X}}/\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.5.** $\left(p_n^*-\frac{\zeta_{1-\varepsilon/2}p_n^*\sqrt{1-p_n^*}}{\sqrt{n}},\ p_n^*+\frac{\zeta_{1-\varepsilon/2}p_n^*\sqrt{1-p_n^*}}{\sqrt{n}}\right)$, где $p_n^*=1/(1+\overline{X})$ и $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.6.** $(\overline{X}-\zeta_{1-\varepsilon/2}\sigma(\overline{X})/\sqrt{n},\overline{X}+\zeta_{1-\varepsilon/2}\sigma(\overline{X})/\sqrt{n})$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.7.** $(\theta_n^*-\zeta_{1-\varepsilon/2}\sigma(\theta_n^*)/\sqrt{n},\theta_n^*+\zeta_{1-\varepsilon/2}\sigma(\theta_n^*)/\sqrt{n})$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.8.** $(X_{(n)},nX_{(n)}/(n+\ln\varepsilon))$. **15.9.** $\big(\theta_1^*-\theta_1^*\zeta_{1-\varepsilon/2}/\sqrt{3n},\theta_1^*+\theta_1^*\zeta_{1-\varepsilon/2}/\sqrt{3n}\big)$ и $\big(\theta_2^*-\theta_2^*\zeta_{1-\varepsilon/2}/\sqrt{5n},\theta_2^*+\theta_2^*\zeta_{1-\varepsilon/2}/\sqrt{5n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.10.** $\big(\alpha_1^*-\alpha_1^*\zeta_{1-\varepsilon/2}/\sqrt{n},\alpha_1^*+\alpha_1^*\zeta_{1-\varepsilon/2}/\sqrt{n}\big)$ и $\big(\alpha_2^*-\sqrt{5/4}\alpha_2^*\zeta_{1-\varepsilon/2}/\sqrt{n},\alpha_2^*+\sqrt{5/4}\alpha_2^*\zeta_{1-\varepsilon/2}/\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.11.** $\big(\overline{X}-1-\zeta_{1-\varepsilon/2}/\sqrt{n},\ \overline{X}-1+\zeta_{1-\varepsilon/2}/\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. Предпочесть следует точный, так как его длина есть величина порядка $1/n$, а не $1/\sqrt{n}$. **15.12.** $\big(\beta^*-\beta^*\zeta_{1-\varepsilon/2}/\sqrt{n},\beta^*+\beta^*\zeta_{1-\varepsilon/2}/\sqrt{n}\big)$, где $\beta^*=1/(\overline{\ln X}-\ln X_{(1)})$. **15.13.** $\big(\zeta^*-\sigma\sqrt{\pi}\zeta_{1-\varepsilon/2}/\sqrt{2n}\,,\zeta^*+\sigma\sqrt{\pi}\zeta_{1-\varepsilon/2}/\sqrt{2n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$; интервал, построенный по $\overline{X}$, короче. **15.14.** $\big(\zeta^*-\pi\zeta_{1-\varepsilon/2}/2\sqrt{n},\zeta^*+\pi\zeta_{1-\varepsilon/2}/2\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **15.15.** $\big(\sigma_n^*-2(\sigma_n^*)^2\zeta_{1-\varepsilon/2}\sqrt{\pi/2-1}/\sqrt{n},\ \sigma_n^*+2(\sigma_n^*)^2\zeta_{1-\varepsilon/2}\sqrt{\pi/2-1}/\sqrt{n}\big)$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$.

§ 16. Различение двух простых гипотез: основные понятия

16.1. $1-\big(1-\overline{\Phi}(3)\big)^n\approx 1-0{,}99865^n$; $\big(1-\overline{\Phi}(2)\big)^n\approx 0{,}977^n$. **16.2.** $\gamma>-1/2$. **16.3.** Основная гипотеза отвергается, если значение хотя бы одного элемента выборки целое. **16.4.** а) 0, $\overline{\Phi}(2)$; б), в) 0, 1/2. **16.5.** $\beta(\delta)>\gamma$ при $n>(\ln(1-\gamma))/(\ln 4-3)$. **16.6.** $\overline{\Phi}(4)\approx 0{,}000032$.

§ 17. Байесовские и минимаксные критерии

17.1. Гипотеза $a=a_1$ принимается, если $\overline{X}<(a_1+a_2)/2$, иначе принимается альтернатива $a=a_2$. **17.2.** а) $\delta(\overline{X})=(1,0)$, если $\overline{X}<3/2$;

иначе $\delta(\overline{X}) = (0,1)$; $\delta(3) = (0,1)$; б) $\delta(\overline{X}) = (1,0,0)$, если $\overline{X} < 3/2$; $\delta(\overline{X}) = (0,1,0)$, если $3/2 \leqslant \overline{X} < 5/2$; $\delta(\overline{X}) = (0,0,1)$, если $5/2 \leqslant \overline{X}$; $\delta(3) = (0,0,1)$. **17.3.** $\delta(\overline{X}) = (1,0,0)$, если $\overline{X} > \ln 2$; $\delta(\overline{X}) = (0,1,0)$, если $\ln 3/2 < \overline{X} \leqslant \ln 2$; $\delta(\overline{X}) = (0,0,1)$, если $\overline{X} \leqslant \ln 3/2$. **17.4.** $\delta(\overline{X}) = (1,0)$, если $\overline{X} < (n-1)(\ln 2)/n(\ln 3 - \ln 2)$. **17.5.** $\delta(\overline{X}) = (1,0)$, если $\overline{X} < m/2 - 1/n$. **17.6.** Гипотеза $\{a = a_1\}$ принимается, если $X_1 < (a_1 + a_2)/2$, иначе принимается альтернатива $\{a = a_2\}$.

§ 18. Наиболее мощные критерии

18.1. Например, критерий, принимающий основную гипотезу при $X_1 < 0$ и альтернативу — при $X_1 \geqslant 0$. **18.2.** $\delta = 1$, если $X_1 < \sqrt{\varepsilon}$; $\beta(\delta) = (1 - \sqrt{\varepsilon})^2$. **18.3.** а) $\delta = 1$, если $X_1 > 1 - \varepsilon$; б) $\delta = 1$, если $X_1X_2 > t$, где t — решение уравнения $t(1 - \ln t) = 1 - \varepsilon$. **18.4.** $\delta = 0$, если $X_1 > 1/2$; $\delta = 1/2$, если $X_1 \leqslant 1/2$. **18.5.** $\delta = 0$, если $X_1 \in (\varepsilon, 1)$; $\delta = 1$ иначе; $\beta(\delta) = 1 + 1/e - 1/e^{\varepsilon}$. **18.6.** $\delta = 0$, если $X_1 = 0$; $\delta = 1$ иначе; $\beta(\delta) = 1 - e^{-2}$. **18.7.** $\delta = 1$, если $X_1 \in (a,\, 1] \cup [3/2,\, 2]$; $\delta = 0$ иначе, где $a = -\frac{1}{2}\ln(1/3 - e^{-3} + e^{-4} + e^{-2}) \approx 0{,}413685$. **18.8.** $\delta = 0$, если $X_1 \in (3/2, 2]$; $\delta = 1/3$, если $X_1 \in [1, 3/2]$; $\delta = 1$, если $X_1 \in [0,1]$. **18.9.** $\delta = 0$, если $X_1 = 0$; $\delta = 2/5$, если $X_1 = 1$; $\delta = 1$, если $X_1 = 2$. **18.10.** $\delta = 0$, если $\overline{X} = 0$; $n = 459$; наиболее мощный критерий: $\delta = 1$, если $\overline{X} > 0$; $\delta = 0{,}01$ иначе; $n = 458$. **18.11.** Гипотеза отвергается, если выпадают две пятёрки. **18.12.** $\delta = 1$, если $X_1 \in [1/2, 1]$; $\delta = 1/2$, если $X_1 = 0$; $\delta = 0$, если $X_1 \in (0, 1/2)$; $\varepsilon \in (1/4,\, 3/4)$. **18.13.** $\delta = 1$, если $\overline{X} \geqslant a_1 + \sigma\zeta_{1-\varepsilon}/\sqrt{n}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1 - \varepsilon$ распределения $N_{0,1}$; состоятельный. **18.14.** $H_1 = \{\sigma^2 = \sigma_1^2\}$, $H_2 = \{\sigma^2 = \sigma_2^2\}$, $\sigma_2^2 < \sigma_1^2$; $\delta = 1$, если $\overline{X^2} < \sigma_1^2\zeta_\varepsilon/n$, где ζ_ε — квантиль уровня ε χ^2-распределения с n степенями свободы. **18.15.** При $\sigma_1^2 < \sigma_2^2$ критическая область имеет вид $\sum_{i=1}^n \bigl(X_i + (a_2\sigma_1^2 - a_1\sigma_2^2)/(\sigma_2^2 - \sigma_1^2)\bigr)^2 > c$. **18.16.** $\delta = 1$, если $\overline{X} < 1/\alpha_1 + \zeta_\varepsilon/\alpha_1\sqrt{n}$, где ζ_ε — квантиль уровня ε распределения $N_{0,1}$; 1. **18.17.** $\delta = 1$, если $\overline{X} > \lambda_1 + \sqrt{\lambda_1}\zeta_{1-\varepsilon}/\sqrt{n}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1 - \varepsilon$ распределения $N_{0,1}$; 1. **18.18.** $\delta = 1$, если $\overline{X} > mp_1 + \sqrt{mp_1(1-p_1)}\zeta_{1-\varepsilon}/\sqrt{n}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1 - \varepsilon$ распределения $N_{0,1}$; 1. **18.19.** $\delta = 1$, если $\overline{X} < (1-p_1)/p_1 + (1-p_1)\zeta_\varepsilon/p_1^2\sqrt{n}$, где ζ_ε — квантиль уровня ε распределения $N_{0,1}$; 1. **18.20.** $\delta = 1$, если $X_1 \geqslant s$; $\delta = 0$ иначе; p_2^s. **18.21.** $1/2$.

§ 19. Равномерно наиболее мощные критерии

19.2. $\delta = 1$, если $\overline{(X - a)^2} < \sigma_1^2\zeta_\varepsilon/n$; $\delta = 0$ иначе, где ζ_ε — квантиль уровня ε χ^2-распределения с n степенями свободы. **19.3.** Гипотеза принимается, если $\overline{X} > 1/\alpha_1 + \zeta_\varepsilon/\alpha_1\sqrt{n}$, где ζ_ε — квантиль уровня ε распределения $N_{0,1}$. **19.4.** а) Гипотеза принимается, если $X_{(1)} \in [\beta_1,\ \beta_1 - (\ln\varepsilon)/n]$; б) гипотеза принимается, если $\beta_1 \leqslant X_{(1)} \leqslant \overline{X} \leqslant \beta_1 + \alpha_1\zeta_{1-\varepsilon}/n$, где $\zeta_{1-\varepsilon}$ — квантиль $\Gamma_{1,n}$ распределения уровня $1-\varepsilon$. **19.5.** Гипотеза принимается, если $X_{(n)} \in [\sqrt[n]{\varepsilon}\theta_0,\ \theta_0]$. **19.6.** Гипотеза принимается, если $\overline{X} < p_1 + \sqrt{p_1(1-p_1)}\zeta_{1-\varepsilon}/\sqrt{n}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1-\varepsilon$ распределения $N_{0,1}$. **19.7.** Гипотеза $H_1 = \{p = 1/2\}$, альтернатива $H_1 = \{p > 1/2\}$; гипотеза «обычности» человека принимается, если угадано не более 56 мыслей. **19.8.** Гипотеза принимается, если $\overline{X} < \lambda_1 + \sqrt{\lambda_1}\zeta_{1-\varepsilon}/\sqrt{n}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1-\varepsilon$ распределения $N_{0,1}$. **19.9.** Гипотеза принимается, если $\overline{X} > (1-p_1)/p_1 + (1-p_1)\zeta_\varepsilon/p_1^2\sqrt{n}$, где ζ_ε — квантиль уровня ε распределения $N_{0,1}$; 1. **19.10.** Например, критерий, принимающий основную гипотезу при $X_1 \in [1/3,\ 1/2]$ и альтернативу — в противном случае; $0{,}5 + \overline{\Phi}(1)$.

§ 20. Критерии согласия

20.1. $\{X_{(1)} > 1/3\} \cup \{X_{(2)} < 1/3\} \cup \{X_{(2)} > 2/3\} \cup \{X_{(3)} < 2/3\}$; $7/9$. **20.3.** $\gamma = 1/6n\varepsilon$. **20.6.** Вероятность получить такое же или ещё большее число гербов (реально достигнутый уровень значимости) при верной основной гипотезе равна 0,189. **20.7.** Нет. **20.9.** Гипотеза $p = p_0$ принимается, если $\sqrt{n}\,|\overline{X} - p_0|/\sqrt{p_0(1-p_0)} < \zeta_{1-\varepsilon/2}$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **20.10.** Гипотеза $\lambda = \lambda_0$ принимается, если $\sqrt{n}\,|\overline{X} - \lambda_0|/\sqrt{\lambda_0} < \zeta_{1-\varepsilon/2}$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **20.11.** Гипотеза принимается, если: а) $|\overline{X} - 1| < \zeta_{1-\varepsilon/2}/\sqrt{n}$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$; б) $\zeta_{\varepsilon/2} < n\overline{(X-1)^2} < \zeta_{1-\varepsilon/2}$, где ζ_δ — квантиль уровня δ χ^2-распределения с n степенями свободы; в) $X_{(1)} < -(\ln\varepsilon)/n$; г) $|2\overline{X} - 1| < \zeta_{1-\varepsilon/2}\sqrt{\overline{X}(2-\overline{X})}/\sqrt{n}$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$; д) $|\overline{X} - 1 < \zeta_{1-\varepsilon/2}\sqrt{\overline{X}}/\sqrt{n}$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **20.13.** Да; $\alpha_1(\delta) = 2\overline{\Phi}\big(\sqrt{nm/(n+m)}\big)$. **20.14.** $\overline{\Phi}(c)$; состоятелен. **20.15.** Основная гипотеза принимается, если $T < \zeta_{1-\varepsilon}$, где $\zeta_{1-\varepsilon}$ — квантиль уровня $1-\varepsilon$ распределения $N_{0,1}$. **20.17.** $2\sum_{i=0}^{[n/2-\gamma]} C_n^i/2^n$; $2\overline{\Phi}(2\gamma/\sqrt{n})$; $\sqrt{n}\zeta_{1-\varepsilon/2}/2$, где $\zeta_{1-\varepsilon/2}$ — квантиль уровня $1-\varepsilon/2$ распределения $N_{0,1}$. **20.21.** Вероятность получить такое же или ещё большее отклонение (реально достигнутый уровень значимости) при верной основной гипотезе равна 0,823. **20.22.** Нет. Реально достигнутый уровень значимости равен $2{,}7 \cdot 10^{-49}$. **20.23.** Вероятность получить такое же или ещё большее отклонение (реально достигнутый уровень значимости) при верной основной гипотезе равна 0,654. **20.24.** A. Нет, реально достигнутый уровень значимости равен 0,058; да, реально достигнутый уровень значимости равен 0,79. B. Нет, реально достигнутый уровень значимости равен 0,022; да, но плохо: реально достигнутый уровень значимости равен 0,28. C. Нет, реально достигнутый уровень значимости равен $1{,}3 \cdot 10^{-7}$; да, реально достигнутый уровень значимости равен 0,9.

Учебное издание

Дмитрий Алексеевич Коршунов
Наталья Исааковна Чернова

СБОРНИК ЗАДАЧ И УПРАЖНЕНИЙ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Учебное пособие

Подписано в печать 15.03.04. Формат $60 \times 84\ ^1/_{16}$. Печать офсетная.
Усл. печ. л. 7,7. Уч.-изд. л. 5,1. Тираж 300 экз. Заказ № 17.

Издательство Института математики,
пр. Коптюга, 4, 630090 Новосибирск, Россия.

Отпечатано на полиграфическом участке ИМ СО РАН,
пр. Коптюга, 4, 630090 Новосибирск, Россия.