



СИБИРСКИЙ
ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ

SIBERIAN
FEDERAL
UNIVERSITY

Электронный учебно-методический комплекс

Математическая статистика

Учебная программа дисциплины

● **Курс лекций**

Практикум по решению задач

Методические указания по самостоятельной работе

Банк тестовых заданий в системе UniTest



Красноярск
ИПК СФУ
2009

УДК 519.22 (075)
ББК 22.172я73
К84

Электронный учебно-методический комплекс по дисциплине «Математическая статистика» подготовлен в рамках реализации Программы развития федерального государственного образовательного учреждения высшего профессионального образования «Сибирский федеральный университет» (СФУ) на 2007–2010 гг.

Рецензенты:

Красноярский краевой фонд науки;

Экспертная комиссия СФУ по подготовке учебно-методических комплексов дисциплин

Крупкина, Т. В.

К84 Математическая статистика [Электронный ресурс]: курс лекций / Т. В. Крупкина, А. К. Гречкосеев. – Электрон. дан. (3 Мб). – Красноярск : ИПК СФУ, 2009. – (Математическая статистика : УМКД № 1455/405–2008 / рук. творч. коллектива Т. В. Крупкина). – 1 электрон. опт. диск (DVD). – Систем. требования : *Intel Pentium* (или аналогичный процессор других производителей) 1 ГГц ; 512 Мб оперативной памяти ; 50 Мб свободного дискового пространства ; привод *DVD* ; операционная система *Microsoft Windows XP SP 2 / Vista* (32 бит) ; *Adobe Reader 7.0* (или аналогичный продукт для чтения файлов формата *pdf*).

ISBN 978-5-7638-1680-8 (комплекса)

ISBN 978-5-7638-1749-2 (курса лекций)

Номер гос. регистрации в ФГУП НТЦ «Информрегистр» 0320902499 (комплекса)

Настоящее издание является частью электронного учебно-методического комплекса по дисциплине «Математическая статистика», включающего учебную программу дисциплины, практикум по решению задач «Математическая статистика в примерах и задачах», методические указания по самостоятельной работе, контрольно-измерительные материалы «Математическая статистика. Банк тестовых заданий», наглядное пособие «Математическая статистика. Презентационные материалы».

Включает в себя теоретическую основу курса и контрольные вопросы.

Предназначен для студентов направлений подготовки бакалавров 010100.62 «Математика», 010500.62 «Прикладная математика и информатика», 010300.62 «Математика. Компьютерные науки» укрупненной группы 010000 «Физико-математические науки и фундаментальная информатика».

© Сибирский федеральный университет, 2009

Рекомендовано к изданию Инновационно-методическим управлением СФУ

Редактор Н. Ф. Ткачук

Разработка и оформление электронного образовательного ресурса: Центр технологий электронного обучения Информационно-телекоммуникационного комплекса СФУ; лаборатория по разработке мультимедийных электронных образовательных ресурсов при КрЦНИТ

Содержимое ресурса охраняется законом об авторском праве. Несанкционированное копирование и использование данного продукта запрещается. Встречающиеся названия программного обеспечения, изделий, устройств или систем могут являться зарегистрированными товарными знаками тех или иных фирм.

Подп. к использованию 30.11.2009

Объем 3 Мб

Красноярск: СФУ, 660041, Красноярск, пр. Свободный, 79

Предисловие

Данное издание предназначено для обеспечения аудиторной и самостоятельной работы студентов института математики Сибирского федерального университета по изучению теоретического курса «Математическая статистика».

Дисциплина размещена в учебном плане в 8-м семестре по 4 часа (2 часа лекций и 2 часа практических занятий) в неделю. Структура изложения определена учебным графиком и соответствует стандартному семестру (17 недель, 17 лекций); номер параграфа пособия соответствует номеру лекции. Дисциплина охватывает две большие темы (два модуля): выборочная теория, статистическое оценивание и проверка статистических гипотез. Первый модуль занимает 1–8-ю недели, второй 9–17-ю недели 8-го семестра. Модуль 1 «Выборочная теория» состоит из трех тем: 1) введение в математическую статистику (лекции 1, 2, 3); 2) распределения математической статистики (лекция 4); 3) статистическое оценивание (лекции 5, 6, 7, 8). Этот модуль является фундаментом следующего и посвящен введению в математическую статистику. Рассматриваются выборочные характеристики и методы статистической обработки случайной выборки, вводятся распределения Пирсона, Стьюдента, Фишера, изучаются распределения выборочных характеристик, рассматриваются простейшие методы решения одной из основных задач статистики — задачи оценивания. Вводятся понятия несмещенных, состоятельных, оптимальных и эффективных оценок, изучаются их свойства.

Второй модуль «Оценивание и проверка статистических гипотез» состоит из четырех тем: 4) достаточность и оптимальность (лекции 9, 10); 5) интервальное оценивание параметров (лекция 11); 6) проверка параметрических гипотез (лекции 12, 13, 14); 7) проверка гипотез о виде распределения и о связи (лекции 15, 16, 17). Во этом модуле изложение ведется на основе теории достаточных статистик. Помимо точечного и интервального оценивания параметров рассматривается вторая важнейшая задача статистики — проверка статистических гипотез. Излагаются как общие подходы к проверке статистической гипотезы и процедуры построения критериев, так и процедуры применения критериев для проверки гипотез, а также методы сравнения статистических критериев. В этом же модуле изучаются элементы регрессионного анализа, включая оценивание параметров уравнения регрессии.

Все разделы пособия включают широкий набор примеров и контрольных вопросов, которые позволят закрепить и углубить теоретические знания и получить навыки практического использования статистических методов.

Приложение содержит восемь таблиц: значения функций $\varphi(x)$, $\Phi_0(x)$, случайные числа, квантили распределения Стьюдента T_n , квантили распределения χ_n^2 , квантили распределения Фишера F порядков $\alpha = 0,01$ и $0,05$, критические значения критерия U Манна — Уитни при уровне значимости $\alpha = 0,05$, греческий алфавит.

Принятые обозначения и сокращения

$P(A)$ — вероятность события A

$E\xi$ — математическое ожидание случайной величины ξ

$D\xi$ — дисперсия случайной величины ξ

$X = (X_1, \dots, X_n)$ — выборка

$\langle \mathcal{F} \rangle$ — статистическая модель

$\langle \mathcal{F}_\theta \rangle$ — параметрическая модель

X_k^* — k -я порядковая статистика

$\nu_n(x)$ — эмпирическая частота

$F_n(x), F_n^*(x)$ — эмпирическая функция распределения

a_k — выборочный начальный момент k -го порядка

m_k — выборочный центральный момент k -го порядка

\bar{X} — выборочное среднее

S^2 — выборочная дисперсия

\bar{S}^2 — исправленная выборочная дисперсия

K_{XY} — выборочная ковариация

r_{XY} — выборочный коэффициент корреляции

\xrightarrow{p} — сходится по вероятности

\xrightarrow{d} — сходится по распределению

$\xrightarrow{\text{п.н.}}$ — сходится почти наверное

$\hat{\theta}$ — оценка θ

T_θ — класс несмещенных оценок параметра θ

$T_\tau(\theta)$ — класс несмещенных оценок параметрической функции $\tau(\theta)$

I — информационное количество Фишера

L — функция правдоподобия

о.м.п. — оценка максимального правдоподобия

о.м.м. — оценка метода моментов

о.н.к. — оценка методом наименьших квадратов

E — экспоненциальное семейство

I_θ — доверительный интервал параметра θ

$M(K)$ — мощность критерия

НКО — наилучшая критическая область

н.м.к. — наиболее мощный критерий

р.н.м.к. — равномерно наиболее мощный критерий

КОП — критерий отношения правдоподобия

$\pi(X)$ — рандомизированный статистический критерий

◀ — начало решения

▶ — конец решения

□ — конец доказательства

Лекция 1. Введение в математическую статистику

Я думаю, ничего нет прекраснее этой статистики.

О. Генри

План лекции: предмет математической статистики, статистические модели, вариационный ряд, эмпирическая функция распределения.

1.1. Предмет математической статистики

Математическая статистика — это раздел математики, который занимается разработкой методов сбора, описания и анализа экспериментальных результатов наблюдений массовых случайных явлений. Фундаментальными понятиями математической статистики являются **генеральная совокупность и выборка**. Определения этих понятий дадим несколько позже, пока же будем понимать под генеральной совокупностью общую совокупность объектов, подлежащих изучению, а под выборкой — часть генеральной совокупности, которую можно реально изучать. Основная задача математической статистики состоит в том, чтобы на основе анализа выборки сделать научно обоснованное заключение о генеральной совокупности.

Математическая статистика выделялась в отдельную науку постепенно. Уже в работах Бернулли¹ и, позднее, Лапласа² и Пуассона³ присутствовали идеи, которые легли в основу математической статистики. Первые работы собственно по математической статистике появились в XVIII веке. Они большей частью относились к описательной статистике и содержали данные о населении и армии. Но рассматриваемые в них вопросы страхования потребовали и разработки соответствующего математического аппарата. Одним из первых ученых – статистиков был бельгиец Адольф Кетле⁴. Он занимался метеорологией и астрономией и впервые применил современные методы сбора данных. Френсис Гальтон⁵, английский психолог и антрополог, разработал методы статистической обработки результатов исследований. Он, кстати, был родственником Чарльза Дарвина и изучал изменчивость видов и наследственность. Гальтон стал основоположником био-

¹ Якоб Бернулли (нем. *Jakob Bernoulli*; 1654–1705) — швейцарский математик, выходец из Голландии.

² Пьер-Симон Лаплас (фр. *Pierre-Simon Laplace*; 1749–1827) — французский математик, физик и астроном.

³ Симеон-Дени Пуассон (фр. *Siméon-Denis Poisson*; 1781–1840) — французский физик, математик.

⁴ Ламбер Адольф Кетлэ (фр. *Lambert Adolphe Jacques Quetelet*; 1796–1874) — бельгийский математик, астроном, метеоролог и статистик, один из создателей научной статистики.

⁵ Фрэнсис Гальтон (Голтон) (англ. *Francis Galton*; 1822–1911) — английский психолог и антрополог. В математике Гальтон разработал методы статистической обработки результатов исследований (в частности, метод исчисления корреляций между переменными); ввел коэффициент корреляции и понятие регрессии.

метрии (науки, применяющей статистические методы в биологии). Биологом был и англичанин Карл Пирсон⁶, разработавший теорию корреляции. Параллельно велись работы математиками русской классической школы (П. Л. Чебышев⁷, А. А. Марков⁸, А. М. Ляпунов⁹, С. Н. Бернштейн¹⁰). Они использовали другие подходы к разработке математической статистики, широко используя достижения теории вероятностей. Интерес к математической статистике быстро развивался, расширялась область применения, появлялись все новые и новые задачи, требующие новых методов решения. В XX веке в блестящих работах выдающегося математика Р. Фишера¹¹, а также в трудах Стьюдента¹², Э. Пирсона¹³ были выработаны основные принципы математической статистики. Большой вклад в это внесли глубокие и обширные исследования В. И. Гливенко¹⁴ и А. Н. Колмогорова¹⁵. В 1946 году ректор Стокгольмского университета Крамер¹⁶ издал книгу «Математические методы статистики», в которой ему удалось объединить результаты трудов англо — американской и русской школ. Эта книга представляет собой изложение математической статистики на основе теории вероятностей.

⁶ Карл (Чарлз) Пирсон (англ. *Karl (Charles) Pearson*; 1857–1936) — английский математик, биолог, философ. Основные труды по математической статистике (кривые Пирсона, распределение Пирсона). Разрабатывал тесты математической статистики и критерии согласия, составлял таблицы значений функций, необходимых для применения математической статистики.

⁷ Пафнутий Львович Чебышев (1821–1894) — величайший, наряду с Н. И. Лобачевским, русский математик XIX века. Математическая школа П. Л. Чебышева, получившая название петербургской, сыграла выдающуюся роль в прогрессе математики не только в России, но и в мире.

⁸ Андрей Андреевич Марков (1856–1922) — русский математик.

⁹ Александр Михайлович Ляпунов (1857–1918) — русский механик и математик. А. М. Ляпунов доказал центральную предельную теорему (теорему Ляпунова) при значительно более общих условиях; ввел метод характеристических функций.

¹⁰ Сергей Натанович Бернштейн (1880–1968) — русский математик. В теории вероятностей он разработал первую по времени (1917) аксиоматику, продолжил и в некотором отношении завершил исследования петербургской школы Чебышева — Маркова по предельным теоремам, разработал теорию слабозависимых величин, исследовал стохастические дифференциальные уравнения и указал на ряд применений вероятностных методов в физике, статистике и биологии.

¹¹ Роналд Эйлмер Фишер (англ. *Ronald Aylmer Fisher*; 1890–1962) — английский статистик и генетик, один из основателей математической статистики.

¹² Стьюдент (англ. *Student*) — псевдоним Уильяма Сили Госсета (англ. *William Sealy Gosset*; 1876–1937) — английский математик и статистик, один из основоположников теории статистических оценок и проверки гипотез.

¹³ Эгон Шарп Пирсон (англ. *Egon Sharpe Pearson*; 1895–1980) — английский математик. Основные труды по математической статистике. Один из основателей (совместно с Е. Нейманом) общей теории проверки статистических гипотез, занимался вопросами статистического контроля качества массовой продукции, большое внимание уделял составлению статистических таблиц.

¹⁴ Валерий Иванович Гливенко (1896–1940) — русский математик.

¹⁵ Андрей Николаевич Колмогоров (1903–1987) — великий русский математик XX века.

¹⁶ Карл Харальд Крамер (шв. *Karl Harald Cramer*; 1893–1985) — шведский математик. Основные труды по теории вероятностей, математической статистике, математической теории страхования.

1.2. Статистические модели

Итак, математическая статистика базируется на понятиях и методах теории вероятностей. Как и всякая математическая теория, она развивается в рамках некоторых моделей, описывающих определенный круг явлений.

В теории вероятностей математические модели случайных явлений основываются на понятии вероятностного пространства, состоящего из пространства элементарных событий Ω , системы событий \mathcal{F} и вероятности P , заданной на событиях $A \in \mathcal{F}$. При этом считается, что вероятность P полностью определена и задача заключается в нахождении вероятностей различных сложных событий для данной вероятностной модели (исходя из известных вероятностей более простых событий и правил вывода).

В математической статистике же предполагается, что вероятность P в модели наблюдаемого случайного явления не известна полностью. Известно только, что P из некоторого заданного класса вероятностей \mathcal{P} . Способы задания класса вероятностей \mathcal{P} могут быть различными. Если задан класс \mathcal{P} , то говорят, что задана **статистическая модель**. Таким образом, статистическая модель описывает ситуации, когда в вероятностной модели изучаемого эксперимента имеется неопределенность в задании вероятности P . Задача математической статистики — уменьшить эту неопределенность, используя информацию, извлеченную из наблюдаемых исходов эксперимента.

Исходным материалом всякого статистического исследования является совокупность результатов наблюдений. В большинстве случаев исходные статистические данные $X = (X_1, \dots, X_n)$ — результат наблюдения некоторой конечной совокупности случайных величин, характеризующей исход изучаемого эксперимента. Предполагается, что эксперимент состоит в проведении n испытаний и результат i -го эксперимента описывается случайной величиной X_i , $i = 1, \dots, n$.

Определение 1.1. Совокупность наблюдаемых случайных величин $X = (X_1, \dots, X_n)$ называется **выборкой**, сами величины X_i , $i = 1, \dots, n$, — **элементами выборки**, а их число n — ее **объемом**. Реализации выборки X будем обозначать строчными буквами $x = (x_1, \dots, x_n)$.

Пусть $X = \{x\}$ — множество всех возможных значений выборки X , которое называется выборочным пространством. Это пространство может совпадать с евклидовым пространством R^n или быть его частью, либо состоять из конечного или счетного числа точек из R^n . Известно, что распределение вероятностей случайной величины полностью определяется ее

функцией распределения и поэтому обычно статистическую модель задают в виде $\langle \mathcal{X}, \mathcal{F} \rangle$, где \mathcal{F} — семейство функций распределения, которому принадлежит неизвестная функция распределения выборки

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n).$$

Обычно рассматривают ситуации, когда компоненты выборки независимы и распределены так же, как некоторая случайная величина ξ с функцией распределения $F_\xi(x) = F(x)$. Мы рассматриваем в дальнейшем только такие модели экспериментов, в которых предполагается, что проводят повторные независимые наблюдения над случайной величиной ξ . Тогда функция распределения выборки полностью определяется функцией распределения $F = F_\xi(x)$ и

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F(x_1) \cdot \dots \cdot F(x_n).$$

Определение 1.2. Множество возможных значений ξ с распределением $F = F_\xi(x)$ называется **генеральной совокупностью**, из которой производят случайную выборку.

Таким образом, мы рассматриваем генеральную совокупность как случайную величину ξ , а выборку — как n -мерную случайную величину (X_1, \dots, X_n) , компоненты которой независимы и одинаково распределены (так же, как ξ). Такие выборки называются **простыми**, и в дальнейшем мы ограничимся их рассмотрением.

Статистическую модель $\langle \mathcal{X}, \mathcal{F} \rangle$ для повторных независимых наблюдений будем обозначать еще короче в виде $\langle \mathcal{F} \rangle$, то есть будем указывать только класс допустимых функций распределения исходной случайной величины (множество \mathcal{X} в данном случае указывать нет необходимости, поскольку оно полностью определяется функцией $F(x)$).

Если функции распределения из класса \mathcal{F} заданы с точностью до значений параметра θ (не обязательно скалярного) с множеством возможных значений Θ , то такая модель обозначается $\langle \mathcal{F}_\theta \rangle$ и называется **параметрической**.

Если модель $\langle \mathcal{F}_\theta \rangle$ такова, что можно дифференцировать по θ интегралы на \mathcal{X} , меняя порядок дифференцирования и интегрирования, то она называется **регулярной**.

Одно из наиболее существенных условий регулярности — то, что выборочное пространство \mathcal{X} не должно зависеть от параметра θ .

В табл. 1 приведены наиболее часто используемые параметрические статистические модели.

Наиболее известные параметрические статистические модели

Наименование модели	Обозначение модели	Функция плотности или распределение вероятности	Множество значений Θ
Нормальная θ_1	$N(\theta, \sigma)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2\sigma^2}}$	R
Нормальная θ_2	$N(a, \theta)$	$\frac{1}{\theta\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\theta^2}}$	R^+
Общая нормальная	$N(\theta_1, \theta_2)$	$\frac{1}{\theta_2\sqrt{2\pi}}e^{-\frac{(x-\theta_1)^2}{2\theta_2^2}}$	$\theta_1 \in R; \theta_2 \in R^+$
Гамма	$\Gamma_{\theta, \beta}$	$\frac{\theta^\beta x^{\beta-1}}{\Gamma(\beta)}e^{-\theta x}, \quad x \geq 0$	R^+
Равномерная	$R(0, \theta)$	$\frac{1}{\theta}, \quad 0 \leq x \leq \theta$	R^+
Общая равномерная	$R(\theta_1, \theta_2)$	$\frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq x \leq \theta_2$	$-\infty < \theta_1 < \theta_2 < +\infty$
Коши	K_θ	$\frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$	R
Биномиальная	$B(N, \theta)$	$C_N^x \theta^x (1-\theta)^{N-x}, \quad x = 0, 1, \dots, N$	$(0, 1)$
Пуассоновская	P_θ	$\frac{\theta^x e^{-\theta}}{x!}$	R^+

Первые семь моделей относятся к **непрерывным**, а две последние — к **дискретным**. Физическая природа перечисленных распределений рассматривалась в начале курса; подробнее о ней можно узнать из работ [11, 16, 33].

Заметим, что модели «Равномерная» и «Общая равномерная» не являются регулярными, так как выборочное пространство \mathcal{X} зависит от параметра θ и (θ_1, θ_2) соответственно.

1.3. Порядковые статистики и вариационный ряд выборки

Пусть $X = (X_1, \dots, X_n)$ — выборка объема n из распределения F и $x = (x_1, \dots, x_n)$ — наблюдавшееся значение X (или реализация вектора X). Каждой реализации x выборки X можно поставить в соответствие упорядоченную последовательность $x_1^*, x_2^*, \dots, x_n^*$,

$$x_1^* \leq x_2^* \leq \dots \leq x_n^*, \quad (1)$$

где $x_1^* = \min(x_1, x_2, \dots, x_n)$, x_2^* — второе по величине значение из (x_1, x_2, \dots, x_n) , \dots , $x_n^* = \max(x_1, x_2, \dots, x_n)$. Последовательность (1) называют **вариационным рядом** реализации. Если теперь через X_k^* обозначить случайную величину, которая для каждой реализации x выборки X принимает значение x_k^* , $k = 1, \dots, n$, из упорядоченной последовательности (1), то

X_k^* называется k -й **порядковой статистикой** выборки, а X_1^* и X_n^* — **экстремальными значениями** выборки. Очевидно, что порядковые статистики удовлетворяют неравенствам

$$X_1^* \leq X_2^* \leq \dots \leq X_n^*. \quad (2)$$

Последовательность (2) называют **вариационным рядом** выборки. Таким образом, порядковая статистика — случайная величина, порождаемая выборкой по правилу (1). Последовательность (1) представляет собой реализацию вариационного ряда (2).

1.4. Эмпирическая функция распределения

Пусть $X = (X_1, \dots, X_n)$ — выборка из генеральной совокупности наблюдаемой случайной величины с функцией распределения $F_\xi(x)$. Поставим задачу построить по выборке некоторое приближение функции $F_\xi(x)$. Введем следующее определение.

Определение 1.3. *Эмпирической частотой $\nu_n(x)$, соответствующей выборке X , называется случайная функция от x , равная числу элементов выборки $X = (X_1, \dots, X_n)$, значения которых меньше x .*

Теорема 1.1. *В модели $\langle \mathcal{F}_\xi \rangle$ эмпирическая частота имеет биномиальное распределение $B(n; F_\xi(x))$.*

Доказательство. Будем рассматривать испытание, успехом в котором считается осуществление события $\{\xi < x\}$. Число элементов выборки $X = (X_1, \dots, X_n)$, значения которых меньше x , равно числу успехов в n независимых испытаниях, связанных с n элементами выборки, то есть имеет биномиальное распределение. Первый параметр этого распределения равен числу испытаний, то есть числу элементов выборки n ; второй параметр равен вероятности успеха в одном испытании, то есть вероятности попадания значений ξ в интервал $(-\infty, x)$, равной $P(\xi < x) = F_\xi(x)$. Таким образом, эмпирическая частота имеет биномиальное распределение $B(n; F_\xi(x))$. \square

Замечание 1.1. *Величину $\nu_n(x)$ можно представить в виде суммы независимых случайных величин, имеющих распределение Бернулли $B(1, F_\xi(x))$; каждое слагаемое при этом является индикатором I события $\{X_i < x\}$.*

Теорема 1.1 позволяет найти распределение k -й порядковой статистики X_k^* .

Теорема 1.2. В модели $\langle \mathcal{F}_\xi \rangle$

$$F_{X_k^*}(x) = \sum_{i=k}^n C_n^i F_\xi^i(x) (1 - F_\xi(x))^{n-i}.$$

Доказательство.

$$\begin{aligned} F_{X_k^*}(x) &= P(X_k^* < x) = P(\nu_n(x) \geq k) = \\ &= \sum_{i=k}^n P(\nu_n(x) = i) = \sum_{i=k}^n C_n^i F_\xi^i(x) (1 - F_\xi(x))^{n-i}. \quad \square \end{aligned}$$

Эмпирическая частота равна абсолютному числу элементов выборки, меньших x . В качестве приближения функции распределения $F_\xi(x)$ целесообразно взять относительную частоту попадания случайных величин (X_1, \dots, X_n) в интервал $(-\infty, x)$.

Определение 1.4. Эмпирической функцией распределения $F_n(x)$, соответствующей выборке X , называется случайная функция от x , вычисляемая по формуле

$$F_n(x) = \frac{\nu_n}{n}, \quad (3)$$

где ν_n — число элементов выборки $X = (X_1, \dots, X_n)$, значения которых меньше x .

Следовательно, эмпирическая функция распределения — сжатая характеристика выборки. Для каждой реализации $x = (x_1, \dots, x_n)$ функция $F_n(x)$ однозначно определена и обладает всеми свойствами функции распределения: изменяется от 0 до 1; не убывает; непрерывна слева; $F_n(x) = 0$ при $x \leq x_1^*$ и $F_n^*(x) = 1$ при $x > x_n^*$. При этом она кусочно-постоянна и возрастает только в точках последовательности (2). Если все составляющие вектора x различны, то

$$F_n(x) = \begin{cases} 0 & \text{при } x \leq x_1^*, \\ \frac{k}{n} & \text{при } x_k^* < x \leq x_{k+1}^*, \quad k = 1, \dots, n-1, \\ 1 & \text{при } x > x_n^*. \end{cases} \quad (4)$$

Заметим, что эмпирическая функция распределения имеет скачки в точках выборки, величина скачка в точке x_i^* равна $1/n$. Если же элемент x_i^* встречается t раз, то величина скачка равна t/n .

График функции $F_n(x)$ имеет вид, изображенный на рис. 1.

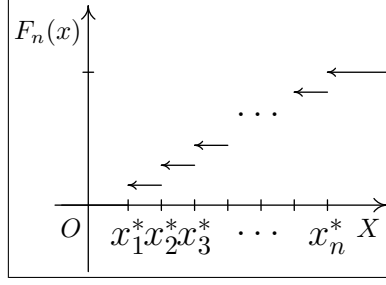


Рис. 1. Эмпирическая функция распределения

Эмпирическая функция распределения обладает важным свойством: при увеличении числа испытаний над случайной величиной ξ (увеличении объема выборки X) она сближается с теоретической функцией $F(x)$, то есть ее можно считать статистическим аналогом для $F(x)$. Это показывают следующие свойства.

1. $\mathbf{E} F_n(x) = F_\xi(x)$.

Доказательство. $\mathbf{E} F_n(x) = \mathbf{E} \left(\frac{\nu_n(x)}{n} \right) = \frac{\mathbf{E} \nu_n(x)}{n}$, но $\nu_n(x)$ имеет биномиальное распределение $B(n; F_\xi(x))$ с математическим ожиданием $nF_\xi(x)$. Поэтому $\mathbf{E} F_n(x) = \frac{\mathbf{E} \nu_n(x)}{n} = \frac{nF_\xi(x)}{n} = F_\xi(x)$. \square

2. $\mathbf{D} F_n(x) = \frac{F_\xi(x)(1 - F_\xi(x))}{n}$.

Доказательство. Аналогично $\mathbf{D} F_n(x) = \mathbf{D} \left(\frac{\nu_n(x)}{n} \right) = \frac{\mathbf{D} \nu_n(x)}{n^2}$, но $\nu_n(x)$ имеет биномиальное распределение $B(n; F_\xi(x))$ с дисперсией $nF_\xi(x)(1 - F_\xi(x))$, и $\mathbf{D} F_n(x) = \frac{\mathbf{D} \nu_n(x)}{n^2} = \frac{nF_\xi(x)(1 - F_\xi(x))}{n^2} = \frac{F_\xi(x)(1 - F_\xi(x))}{n}$. \square

3. $\sqrt{n}(F_n(x) - F_\xi(x)) \sim N \left(0; \sqrt{F_\xi(x)(1 - F_\xi(x))} \right)$.

Доказательство. Асимптотическая нормальность следует из ЦПТ, так как величину $nF_n(x) = \nu_n(x)$ можно представить в виде суммы независимых случайных величин, имеющих распределение Бернулли $B(1, F_\xi(x))$ (каждое слагаемое при этом является индикатором I события $\{X_i < x\}$). По ЦПТ для суммы одинаково распределенных независимых случайных величин справедливо: $\frac{n(F_n(x) - F_\xi(x))}{\sqrt{nF_\xi(x)(1 - F_\xi(x))}} \xrightarrow{d} u \in N(0, 1)$.

Отсюда следует, что $\frac{\sqrt{n}(F_n(x) - F_\xi(x))}{\sqrt{F_\xi(x)(1 - F_\xi(x))}} \xrightarrow{d} u \in N(0, 1)$, что и означает асимптотическую нормальность. \square

$$4. F_n(x) \xrightarrow{p} F_\xi(x).$$

Доказательство. Это следует, например, из ЗБЧ в форме Бернулли, так как эмпирическая функция распределения $F_n(x)$ — это частота события $\{X < x\}$, а $F_\xi(x)$ — вероятность этого события. Можно также сослаться на свойство среднего арифметического n независимых слагаемых, имеющих распределение Бернулли (ЗБЧ в форме Хинчина). \square

$$5. F_n(x) \xrightarrow{\text{п.н.}} F_\xi(x).$$

Доказательство. Следует из выполнимости УЗБЧ в форме Бореля. \square

Рассмотрим еще несколько важных свойств эмпирической функции распределения. Справедливы следующие теоремы [1, 2, 24].

Теорема 1.3 (теорема Гливенко). Пусть $F_n(x)$ — эмпирическая функция распределения, построенная по выборке $X = (X_1, \dots, X_n)$ из распределения F_ξ , и $F(x)$ — соответствующая теоретическая функция распределения. Тогда для любого $-\infty < x < \infty$

$$P(\lim_{n \rightarrow \infty} D_n = 0) = 1, \quad (5)$$

где

$$D_n = D_n(x) = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|. \quad (6)$$

Приведем еще две предельные теоремы для эмпирической функции распределения, имеющие важное значение для проверки различных предположений (гипотез) о выборках.

Теорема 1.4 (теорема Колмогорова). Если функция $F(x)$ непрерывна, то при любом фиксированном $t > 0$

$$\lim_{n \rightarrow \infty} P(|\sqrt{n}D_n| < t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 t^2}, \quad (7)$$

Функция $K(t)$ называется функцией Колмогорова, и ее можно с хорошим приближением использовать для практических расчетов уже при объеме выборки, равном или большим 20. Заметим, что теорема справедлива для любой непрерывной функции распределения и позволяет найти границы, в которых с заданной вероятностью $0 < \alpha < 1$ находится теоретическая функция $F(x)$ (она в нашей статистической модели неизвестна). Если задана вероятность α , то при больших n с вероятностью, близкой к α , функция $F(x)$ удовлетворяет неравенству

$$|F_n(x) - F(x)| \leq \frac{t_\alpha}{\sqrt{n}}, \quad (8)$$

где величина t_α вычисляется как корень уравнения $K(t) = \alpha$ с помощью таблиц для функции $K(t)$. Обоснование данного неравенства можно найти в [24].

Теорема 1.5 (теорема Смирнова). Пусть $F_n(x)$ и $F_m(x)$ — две эмпирические функции распределения, построенные на основе двух независимых выборок объемов n и m из одного и того же распределения F_ξ , и

$$D_{n,m} = \sup_{-\infty < x < \infty} |F_n(x) - F_m(x)|. \quad (9)$$

Тогда если теоретическая функция $F(x)$ непрерывна, то для любого фиксированного $t > 0$

$$\lim_{n,m \rightarrow \infty} P(\sqrt{nm/(n+m)} D_{n,m} \leq t) = K(t), \quad (10)$$

где функция $K(t)$ определена равенством (7).

Эта теорема позволяет ответить на важный практический вопрос, можно ли считать, что две выборки получены из одного и того же распределения.

Замечание 1.2. Поскольку эмпирическая функция распределения близка к теоретической, выборочные характеристики (выборочное среднее, выборочная дисперсия и другие) можно рассматривать как обычные характеристики (математическое ожидание, дисперсия и другие) случайной величины, принимающей значения X_1, \dots, X_n с равными вероятностями $\frac{1}{n}$. Таким образом, простую выборку можно рассматривать не только как n -мерную случайную величину (X_1, \dots, X_n) , но и как одномерную дискретную величину:

X	X_1	X_2	\dots	X_n
P	$1/n$	$1/n$	\dots	$1/n$

1.5. Контрольные вопросы

1. Какая статистическая модель применима для выборки, полученной следующим образом: 10 раз подсчитывалось число гербов при бросании трех монет?
2. Какая статистическая модель применима для выборки действующих цен на литр молока в 50 различных магазинах г. Красноярска?

3. *Дайте определение параметрической модели.*
4. *Является ли модель $N(a, \theta)$ параметрической?*
5. *Дайте определение регулярной модели.*
6. *Является ли модель $R[\theta, b]$ регулярной?*
7. *Дайте определение вариационного ряда реализации.*
8. *Дайте определение вариационного ряда выборки.*
9. *Можно ли восстановить по вариационному ряду выборку?*
10. *Дайте определение эмпирической функции распределения.*
11. *Какими свойствами обладает эмпирическая функция распределения?*
12. *Можно ли восстановить по эмпирической функции распределения вариационный ряд?*
13. *Верно ли, что все элементы простой выборки одинаково распределены?*
14. *Верно ли, что все элементы вариационного ряда одинаково распределены?*

Лекция 2. Основные выборочные характеристики

Дело не в числах, а в том, что вы с ними делаете.
К. Мэндервилл

План лекции: выборочные моменты, соответствие выборочных и теоретических характеристик, свойства \bar{X} и S^2 , другие характеристики выборки, сходимость выборочных моментов.

2.1. Выборочные моменты и функции от выборочных моментов

Пусть $X = (X_1, \dots, X_n)$ — выборка объема n из распределения F и $x = (x_1, \dots, x_n)$ — наблюдавшееся значение X . Любая функция от X представляет собой также случайную величину с распределением, однозначно определяемым распределением вектора X . В теории вероятностей введены определения различных моментов случайной величины. Так, начальным моментом порядка k случайной величины ξ называется число

$$\alpha_k = \mathbf{E}(\xi)^k, \quad (11)$$

а центральным моментом порядка k случайной величины ξ называется число

$$\mu_k = \mathbf{E}(\xi - \mathbf{E} \xi)^k. \quad (12)$$

Введем общее обозначение для моментов: $\mathbf{E} g(\xi)$ ($g(\xi) = (\xi)^k$ для начальных моментов и $(\xi - \mathbf{E} \xi)^k$ для центральных моментов k -го порядка). Каждому теоретическому моменту $g(\xi)$ можно поставить в соответствие его статистический аналог $G = G(X)$, вычисляемый по формуле

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

Случайную величину G называют **эмпирической или выборочной характеристикой**, соответствующей теоретической характеристике g .

Если $g(\xi) = \xi^k$, то G — **выборочный начальный момент** k -го порядка, который будем обозначать a_k .

Определение 2.1. *Выборочным начальным моментом k -го порядка называют случайную величину*

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k. \quad (13)$$

Определение 2.2. Если $k = 1$, то величину a_1 называют **выборочным средним** и обозначают символом \bar{X} ,

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

Значения случайной величины \bar{X} при конкретной реализации x выборки X будем обозначать строчной буквой (\bar{x}) . Напомним, что теоретические моменты случайной величины ξ обозначают греческими буквами, а соответствующие им выборочные — латинскими.

Определение 2.3. **Выборочным центральным моментом k -го порядка** называют случайную величину

$$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (14)$$

Определение 2.4. При $k = 2$ величину m_2 называют **выборочной дисперсией** и обозначают S^2 :

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

Выборочная дисперсия характеризует разброс значений.

Определение 2.5. **Выборочным среднеквадратическим отклонением** называют случайную величину

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

Среднеквадратическое отклонение — линейная мера разброса выборочных значений. При рассмотрении свойств плотности распределения непрерывных величин вводили такие характеристики, как коэффициенты асимметрии A и эксцесса E . Если график плотности распределения симметричен, то $A = 0$. По значению A судят о степени отклонения от симметрии, по знаку A — о правосторонней ($A > 0$) или левосторонней ($A < 0$) асимметрии.

Для нормального распределения $E = 0$, и поэтому о кривых плотности с $E = 0$ говорят, что у них нормальный эксцесс. Если $E > 0$ ($E < 0$), то эксцесс кривой положительный, то есть кривая плотности островершинна (отрицательный, то есть кривая плотности плосковершинна).

Пусть задана выборка X из распределения непрерывной случайной величины ξ . Тогда для нее можно определить **выборочные коэффициенты асимметрии \hat{A} и эксцесса \hat{E}** по следующим формулам.

Определение 2.6. *Выборочным коэффициентом асимметрии называется величина*

$$\hat{A} = \frac{m_3}{S^3}. \quad (15)$$

Определение 2.7. *Выборочным коэффициентом эксцесса называется величина*

$$\hat{E} = \frac{m_4}{S^4} - 3. \quad (16)$$

Аналогично можно ввести и другие характеристики, например, абсолютные моменты. Необходимо помнить, что выборочные характеристики вводятся соответственно теоретическим характеристикам случайной величины (табл. 2).

Таблица 2

Соответствие выборочных и теоретических характеристик

Теоретические характеристики	Выборочные характеристики
$a = \mathbf{E} \xi$ математическое ожидание	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ выборочное среднее
$\sigma^2 = \mathbf{D} \xi$ дисперсия	$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ выборочная дисперсия
$\alpha_k = \mathbf{E} \xi^k$ начальный k -й момент	$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ начальный выборочный k -й момент
$\mu_k = \mathbf{E}(\xi - \mathbf{E} \xi)^k$ центральный k -й момент	$m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$ центральный выборочный k -й момент
$A = \frac{\mu_3}{\sigma^3}$ коэффициент асимметрии	$\hat{A} = \frac{m_3}{S^3}$ выборочный коэффициент асимметрии
$E = \frac{\mu_4}{\sigma^4} - 3$ коэффициент эксцесса	$\hat{E} = \frac{m_4}{S^4} - 3$ выборочный коэффициент эксцесса

2.2. Свойства выборочного среднего \bar{X} и S^2

Выборочные характеристики являются случайными величинами, и поэтому можно говорить о их распределениях (выборочных распределениях) и изучать различные характеристики этих распределений.

Будем считать, что выборка взята из совокупности с математическим ожиданием $\mathbf{E} \xi = a$ и дисперсией $\mathbf{D} \xi = \sigma^2$; тогда $\mathbf{E} X_i = a$, $\mathbf{D} X_i = \sigma^2$ для любого элемента выборки X_i , при этом случайные величины X_i независимы.

Свойства $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

1. $\mathbf{E} \bar{X} = a$.

Доказательство.

$$\mathbf{E} \bar{X} = \mathbf{E} \left(\frac{1}{n} \sum X_i \right) = \frac{1}{n} \sum \mathbf{E} X_i = \frac{1}{n} \sum a = \frac{na}{n} = a. \square$$

2. $\mathbf{D} \bar{X} = \frac{\sigma^2}{n}$.

Доказательство.

Поскольку X_i независимы и распределены так же, как случайная величина ξ , то

$$\mathbf{D} \bar{X} = \mathbf{D} \left(\frac{1}{n} \sum X_i \right) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}. \square$$

3. $\bar{X} \xrightarrow{p} a$.

Доказательство.

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} a, \text{ по ЗБЧ в форме Хинчина. } \square$$

4. $\bar{X} \xrightarrow{\text{п.н.}} a$.

Доказательство. $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} a$, по критерию УЗБЧ Колмогорова. \square

5. $\frac{(\bar{X}-a)\sqrt{n}}{\sigma} \sim N(0, 1)$.

Доказательство.

Согласно обозначениям $\xi \sim N(a, \sigma) \Leftrightarrow \frac{\xi-a}{\sigma} \xrightarrow{d} u \in N(0, 1)$,

$$\frac{\bar{X} - a}{\sigma/\sqrt{n}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i - a}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - na}{\sigma/\sqrt{n}} \xrightarrow{d} u \in N(0, 1), \text{ по ЦПТ. } \square$$

Рассмотрим свойства выборочной дисперсии S^2 .

1.

$$S_{X+c}^2 = S_X^2.$$

Доказательство.

$$\begin{aligned} S_{X+c}^2 &= \frac{1}{n} \sum (X_i + c - \overline{X + c})^2 = [\overline{X + c} = \\ &= \frac{1}{n} \sum (X_i + c) = \frac{1}{n} \sum X_i + \frac{1}{n} \sum c = \overline{X} + c] = \frac{1}{n} \sum (X_i - \overline{X})^2 = S_X^2. \square \end{aligned}$$

2.

$$S_{kX}^2 = k^2 S_X^2.$$

Доказывается аналогично.

3.

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

Доказательство.

$$\begin{aligned} S^2 &= \frac{1}{n} \sum (X_i^2 - 2\overline{X}X_i + \overline{X}^2) = \frac{1}{n} \sum X_i^2 - \frac{2\overline{X}}{n} \sum X_i + \overline{X}^2 = \\ &= \frac{1}{n} \sum X_i^2 - 2\overline{X}^2 + \overline{X}^2 = \frac{1}{n} \sum X_i^2 - \overline{X}^2. \square \end{aligned}$$

4.

$$S^2 = \min_c \frac{1}{n} \sum (X_i - c)^2.$$

Доказательство.

$$\begin{aligned} \sum (X_i - c)^2 &= \sum (X_i - \overline{X} + \overline{X} - c)^2 = \\ &= \sum (X_i - \overline{X})^2 + n(\overline{X} - c)^2 + 2(\overline{X} - c) \sum (X_i - \overline{X}), \\ \sum (X_i - \overline{X}) &= \sum X_i - n\overline{X} = \sum X_i - n \frac{\sum X_i}{n} = 0, \\ \frac{1}{n} \sum (X_i - c)^2 &= \frac{1}{n} \sum (X_i - \overline{X})^2 + (\overline{X} - c)^2 = S^2 + (\overline{X} - c)^2 \geq S^2. \square \end{aligned}$$

5.

$$\mathbf{E} S^2 = \frac{(n-1)\mu_2}{n}. \quad (17)$$

Доказательство. Сделаем сдвиг величины X на $\mathbf{E} X$ (что не изменяет дисперсии); тогда $\mathbf{D} X = \mathbf{E} X^2$ и $\mathbf{D} \bar{X} = \mathbf{E} \bar{X}^2$.

$$S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = \frac{1}{n} \left(\sum X_i^2 \right) - \bar{X}^2.$$

$$\mathbf{E} S^2 = \frac{1}{n} \left(\sum \mathbf{E} X_i^2 \right) - \mathbf{E}(\bar{X}^2) = \mathbf{D} X - \mathbf{D} \bar{X} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \cdot \sigma^2. \quad \square$$

6.

$$\mathbf{D} S^2 = \frac{(n-1)^2}{n^3} \left(\mu_4 - \frac{(n-3)}{n-1} \mu_2^2 \right). \quad (18)$$

Доказательство.

По-прежнему будем делать сдвиг величины X на $\mathbf{E} X$; тогда все центральные моменты нечетных порядков равны 0. По свойствам выборочной дисперсии

$$S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = \frac{1}{n} \left(\sum X_i^2 - n\bar{X}^2 \right),$$

$$\mathbf{E} S^2 = \mu_2 - \frac{n\mu_2}{n^2} = \frac{n-1}{n} \mu_2.$$

Таким образом,

$$\mathbf{D} S^2 = \frac{1}{n^2} \mathbf{E} \left(\sum X_i^2 - n\bar{X}^2 \right)^2 - \left(\frac{n-1}{n} \right)^2 \mu_2^2. \quad (19)$$

Обозначим

$$\begin{aligned} J &= \mathbf{E} \left(\sum X_i^2 - n\bar{X}^2 \right)^2 = \mathbf{E} \left(\left(\sum X_i^2 \right)^2 - 2n\bar{X}^2 \sum X_i^2 + n^2\bar{X}^4 \right) = \\ &= \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3. \end{aligned} \quad (20)$$

$$1) \quad \mathbf{E}_1 = \mathbf{E} \left(\sum X_i^2 \right)^2 = \mathbf{E} \left(\sum X_i^4 + \sum_{\substack{i,j \\ i \neq j}} X_i^2 X_j^2 \right) = n\mu_4 + n(n-1)\mu_2^2.$$

$$\begin{aligned}
2) \quad \mathbf{E}_2 &= -2n \mathbf{E} \left(\bar{X}^2 \sum X_i^2 \right) = -\frac{2n}{n^2} \mathbf{E} \left(\left(\sum X_i \right)^2 \sum X_i^2 \right) = \\
&= -\frac{2}{n} \mathbf{E} \left(\sum X_i^2 \right)^2 = -\frac{2}{n} (n\mu_4 + n(n-1)\mu_2^2).
\end{aligned}$$

$$\begin{aligned}
3) \quad \mathbf{E}_3 &= n^2 \mathbf{E} \bar{X}^4 = \frac{n^2 \mathbf{E} (\sum X_i)^4}{n^4} = \frac{1}{n^2} \mathbf{E} \left(\sum X_i^4 + C_4^2 \sum_{i < j} X_i^2 X_j^2 \right) = \\
&= \frac{1}{n^2} \left(n\mu_4 + \frac{6n(n-1)}{2} \mu_2^2 \right).
\end{aligned}$$

Подставим найденные значения в (21).

$$J = n\mu_4 + n(n-1)\mu_2^2 - 2\mu_4 - 2(n-1)\mu_2^2 + \frac{\mu_4}{n} + \frac{3(n-1)}{n} \mu_2^2. \quad (21)$$

Подставляя (21) в (19), окончательно находим:

$$\begin{aligned}
\mathbf{D} S^2 &= \frac{J}{n^2} - \left(\frac{(n-1)}{n} \right)^2 \mu_2^2 = \\
&= \frac{\mu_4}{n} + \frac{n-1}{n} \mu_2^2 - \frac{2}{n^2} \mu_4 - \frac{2(n-1)}{n^2} \mu_2^2 + \\
&+ \frac{\mu_4}{n^3} + \frac{3(n-1)}{n^3} \mu_2^2 - \mu_2^2 + \frac{2}{n} \mu_2^2 - \frac{1}{n^2} \mu_2^2 = \\
&= \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3} = \frac{(n-1)^2}{n^3} \left(\mu_4 - \frac{(n-3)}{n-1} \mu_2^2 \right). \quad (22)
\end{aligned}$$

Таким образом,

$$\mathbf{D} S^2 = O \left(\frac{1}{n} \right). \quad \square$$

Следствие. Отсюда легко получить дисперсию S^2 для нормального распределения $N(a, \sigma)$:

$$\begin{aligned}
\mathbf{D} S^2 &= \frac{3\sigma^4 - \sigma^4}{n} - \frac{2(3\sigma^4 - 2\sigma^4)}{n^2} + \frac{3\sigma^4 - 3\sigma^4}{n^3} = \frac{2\sigma^4}{n} - \frac{2\sigma^4}{n^2} = \\
&= \frac{2\sigma^4(n-1)}{n^2} = \frac{2(n-1)}{n^2} \sigma^4.
\end{aligned}$$

Мы использовали найденные ранее значения моментов нормального распределения: $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$, (общая формула: $\mu_{2k} = (2k - 1)!! \cdot \sigma^{2k}$).

Часто используется также дисперсия исправленной выборочной дисперсии. Найдем ее:

$$\mathbf{D} \bar{S}^2 = \frac{n^2}{(n-1)^2} \mathbf{D} S^2 = \frac{2}{n-1} \sigma^4. \quad (23)$$

2.3. Другие характеристики выборки

Определение 2.8. Выборочной модой называется значение m_0 , чаще всего наблюдающееся:

$$n_i(m_0) = \max_i n_i.$$

Определение 2.9. Выборочной медианой называется значение m_e , равное среднему члену вариационного ряда:

$$m_e = X_{\left[\frac{n}{2}\right]+1}^*.$$

Определение 2.10. Выборочной квантилью порядка q , $0 < q < 1$ называется значение X_q , равное члену вариационного ряда с номером $[nq] + 1$.

Вычисление выборочных характеристик связи между величинами X и Y .

Выборочная ковариация K_{XY} :

$$K_{XY} = \overline{XY} - \bar{X} \cdot \bar{Y}, \text{ где } \overline{XY} = \frac{1}{n} \sum_{i=1}^n X_i Y_i.$$

Выборочный коэффициент корреляции r_{XY} :

$$r = \frac{K_{XY}}{s_X s_Y}.$$

Выборочное уравнение линейной регрессии Y на X :

$$\hat{Y} - \bar{Y} = r \frac{s_Y}{s_X} (X - \bar{X}).$$

Выборочное уравнение линейной регрессии X на Y :

$$\hat{X} - \bar{X} = r \frac{s_X}{s_Y} (Y - \bar{Y}).$$

2.4. Сходимость выборочных моментов и функций от них

Для практических приложений важное значение имеет вопрос о поведении выборочных характеристик при возрастании объема выборки n . По аналогии с эмпирической функцией распределения можно ожидать, что при неограниченном возрастании n выборочные моменты будут сходиться по вероятности к соответствующим теоретическим моментам. Действительно, пусть случайная величина ξ имеет все моменты α_k . Соответствующие им выборочные моменты обозначим a_{nk} , чтобы подчеркнуть зависимость от n . Все a_{nk} представляют собой случайные величины и при фиксированном k для каждого n имеют одинаковые математические ожидания α_k и дисперсии $\frac{\alpha_{2k} - \alpha_k^2}{n}$. Это следует из соотношений

$$\mathbf{E} a_{nk} = \frac{1}{n} \sum_{i=1}^n \mathbf{E} X_i^k = \mathbf{E} \xi^k = \alpha_k; \quad (24)$$

$$\mathbf{D} a_{nk} = \frac{1}{n} \sum_{i=1}^n \mathbf{D} X_i^k = \frac{1}{n} \mathbf{D} \xi^k = \frac{\mathbf{E} \xi^{2k} - (\mathbf{E} \xi^k)^2}{n} = \frac{\alpha_{2k} - \alpha_k^2}{n}. \quad (25)$$

Но тогда из закона больших чисел вытекает, что a_{nk} сходится по вероятности к α_k при $n \rightarrow \infty$. Поэтому выборочные моменты a_{nk} можно использовать в качестве приближенных значений для соответствующих теоретических моментов. Аналогичное утверждение справедливо вообще для любых непрерывных функций от конечного числа величин. Напомним формулировку соответствующей теоремы.

Теорема 2.1. Пусть $\varphi(x_1, \dots, x_m)$ — непрерывная функция. Тогда, если последовательность $\{\xi_{kn}\}$ сходится по вероятности к ξ_k , то и последовательность $\{\varphi(\xi_{1n}, \dots, \xi_{mn})\}$ сходится по вероятности к $\varphi(\xi_1, \dots, \xi_m)$:

$$\xi_{kn} \xrightarrow{p} \xi_k \Rightarrow \varphi(\xi_{1n}, \dots, \xi_{mn}) \xrightarrow{p} \varphi(\xi_1, \dots, \xi_m). \quad (26)$$

Теорема 2.1 имеет важное значение, поскольку из нее следует, в частности, что любые центральные выборочные моменты, коэффициенты асимметрии и эксцесса при неограниченном увеличении объема выборки сходятся по вероятности к соответствующим теоретическим характеристикам. Действительно, из (14), используя бином Ньютона, можно получить

$$m_k = \sum_{l=0}^k (-1)^l C_k^l \bar{X}^l a_{k-l}, \quad (27)$$

то есть любой центральный выборочный момент можно представить в виде непрерывной функции от выборочных моментов, которые сходятся по вероятности к теоретическим моментам. С другой стороны, центральные теоретические моменты выражаются через теоретические моменты также по аналогичной (27) формуле:

$$\mu_k = \mathbf{E}(\xi - \mathbf{E} \xi)^k = \sum_{l=0}^k (-1)^{k-l} C_k^l \alpha_1^l \alpha_{k-l}. \quad (28)$$

Таким образом, все условия теоремы 2.1 выполняются. Аналогичные рассуждения справедливы для коэффициентов асимметрии и эксцесса.

До сих пор мы рассматривали сходимость выборочных характеристик к соответствующим теоретическим. Однако важное значение имеет и вопрос о поведении распределений выборочных характеристик для больших выборок. Легко видеть, что каждый выборочный момент a_k представляет собой сумму n независимых и одинаково распределенных случайных величин, и поэтому к нему можно применить центральную предельную теорему. В этой связи приведем теорему.

Теорема 2.2. *Распределение выборочного момента a_k при $n \rightarrow \infty$ сходится к нормальному распределению.*

Теорема 2.2 означает, что выборочное распределение случайной величины a_k при $n \rightarrow \infty$ является приближенно нормальным. Последнее обстоятельство позволяет при больших выборках оценить вероятность, с которой значения выборочного момента отклонятся от теоретического на величину, не большую ε . Правда, при этом необходимы значения теоретических моментов. Действительно, для любого $\varepsilon > 0$ при $n \rightarrow \infty$ вероятность события

$$P \left(\sqrt{\frac{n}{\alpha_{2k} - \alpha_k^2}} |a_k - \alpha_k| < \varepsilon \right) \rightarrow 2\Phi(\varepsilon) - 1,$$

где $\Phi(t)$ — функция распределения закона $N(0, 1)$. Следовательно, вероятность события

$$\left\{ \sqrt{\frac{n}{\alpha_{2k} - \alpha_k^2}} |a_k - \alpha_k| < \varepsilon \right\}$$

при больших n примерно равна $2\Phi(\varepsilon) - 1$. Для центральных выборочных моментов m_k при $n \rightarrow \infty$ также справедливо утверждение о сходимости их выборочных распределений к нормальному (доказательство этого утверждения можно найти в работе [35]). Например, распределение выборочной дисперсии S_n^2 сходится к нормальному распределению $N(\mu_2, \sqrt{(\mu_4 - \mu_2^2)/n})$.

2.5. Контрольные вопросы

1. Дайте определение выборочного среднего.
2. Докажите, что в любом распределении $E \bar{X} = E \xi$.
3. Докажите, что $D \bar{X} = \frac{\sigma^2}{n}$.
4. Приведите формулу для вычисления выборочного среднего по статистическому ряду.
5. Приведите формулу для вычисления выборочной дисперсии по статистическому ряду.
6. Укажите два различных способа найти выборочную дисперсию по вариационному ряду.
7. Как изменятся выборочные дисперсия и среднее квадратическое отклонение, если прибавить к каждому элементу выборки постоянную c ?
8. Как изменятся выборочные дисперсия и среднее квадратическое отклонение, если умножить каждый элемент выборки на постоянную c ?
9. Что характеризует выборочный коэффициент асимметрии?
10. Выразите выборочный центральный момент 3-го порядка через выборочные начальные моменты.
11. Как называется средний член вариационного ряда?
12. Квантилью какого порядка является медиана?
13. Каково асимптотическое распределение выборочного среднего в модели Пуассона $\langle P_\lambda \rangle$?
14. Каково точное распределение выборочного среднего в модели Пуассона $\langle P_\lambda \rangle$?

Лекция 3. Группировка выборки

*Единица — ноль, единица — вздор,
голос единицы — тоньше писка.*
В. Маяковский

План лекции: группировка выборки, гистограмма и полигон частот, пример группировки выборки, поправки Шеппарда.

3.1. Гистограмма и полигон частот

Кроме эмпирической функции распределения существуют и другие способы наглядного представления и сжатия статистических данных. Одним из таких способов является построение гистограммы.

Пусть нам дана реализация (x_1, \dots, x_n) выборки X из F_ξ . Интервал $[x_1^*, x_n^*]$ разбивают на подынтервалы $[a_0, a_1], [a_1, a_2], \dots, [a_{k-1}, a_k]$, где $a_0 = x_1^*, a_k = x_n^*$. Затем подсчитывают число n_j координат реализации (элементов выборки), попавших в каждый подынтервал $[a_{j-1}, a_j]$, $j = 1, \dots, k$, то есть подсчитывают частоту попадания статистических данных в каждый подынтервал. Иначе говоря, данные группируют. На каждом подынтервале, как на основании, строят прямоугольники площади $\frac{n_j}{n}$, для чего берут высоту прямоугольника равную $\frac{n_j}{nh_j}$, где h_j — длина j -го подынтервала (на практике обычно длину каждого подынтервала полагают равной h). Полученную при этом фигуру называют **гистограммой** (рис. 2).

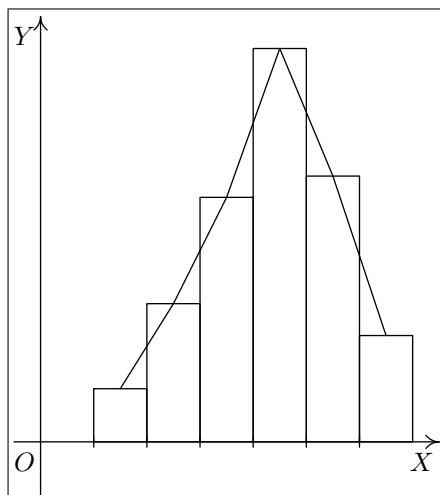


Рис. 2. Гистограмма и полигон частот

Поскольку площадь каждого прямоугольника есть не что иное, как относительная частота попадания выборочных значений в соответствующий

щий интервал, то по теореме Бернулли она будет сходиться по вероятности при $n \rightarrow \infty$ к вероятности попадания значения случайной величины ξ в соответствующий интервал. Если случайная величина ξ непрерывная и $f(x)$ ее плотность, то огибающая (верхняя граница) гистограммы является статистическим аналогом для теоретической плотности $f(x)$.

В методе гистограмм неизвестная плотность распределения приближается кусочно-постоянными функциями. Но точность приближения можно улучшить, если применить кусочно-линейные функции. Для этой цели строят **полигон частот**. Полигон частот — это ломаная линия, проходящая через середины верхних границ прямоугольников гистограммы. Построенная таким образом кусочно-линейная функция также является статистическим аналогом теоретической плотности (рис. 2).

Можно построить также кумуляту — ломаную с вершинами в точках $(a_i; \frac{1}{n} \sum_{j=1}^{i-1} n_j)$. Кумулята дает представление о графике функции распределения, а гистограмма и полигон — о графике плотности распределения.

Еще раз отметим, что интерпретация графических характеристик как статистических аналогов вероятностных характеристик имеет смысл и обоснование только для выборки X из непрерывного распределения F .

3.2. Пример группировки выборки

Рассмотрим на примере, как сгруппировать выборку, состоящую из 1000 элементов. Пусть для определенности минимальный элемент выборки равняется $x_{min} = 0,01$, а максимальный $x_{max} = 10,65$.

Определим сначала число интервалов k . Рекомендуются брать k такое, что $2^{k-1} \sim n$.

В данном примере $n = 1000$.

$$2^{10} = 1024 \sim 1000; \quad k - 1 = 10, \quad k = 11.$$

Выберем число интервалов $k = 11$.

2. Определим длину интервала h :

$$h = \frac{x_{max} - x_{min}}{k}.$$

Находим

$$h = \frac{10,65 - 0,01}{11} \approx 0,967.$$

3. Найдем границы интервалов группировки $x_0 = x_{min}$, $x_i = x_0 + hi$, $i = 1, \dots, k$ и составим таблицу группировки. Внесем границы интервалов в столбец «Интервал». Подсчитаем, сколько элементов выборки попало в каждый интервал и заполним в таблице столбец «Численность n_i ».

Остальные столбцы рассчитываются по столбцу численностей. Они пригодятся при построении графических характеристик.

№	Интервал	Численность n_i	$\frac{n_i}{n}$	$\frac{n_i}{nh}$	$\sum_{j=1}^i \frac{n_j}{n}$
1	0,01 – 0,98				
2	0,98 – 1,94				
3	1,94 – 2,91				
4	2,91 – 3,88				
5	3,88 – 4,85				
6	4,85 – 5,81				
7	5,81 – 6,78				
8	6,78 – 7,75				
9	7,75 – 8,72				
10	8,72 – 9,68				
11	9,68 – 10,65				

Ниже показана заполненная таблица. Обратите внимание, что сумма элементов столбца «Численность n_i » должна равняться n , а сумма элементов столбца « $\frac{n_i}{n}$ » должна равняться 1.

№	Интервал	Численность n_i	$\frac{n_i}{n}$	$\frac{n_i}{nh}$	$\sum_{j=1}^i \frac{n_j}{n}$
1	0,00 – 0,09	80	0,080	0,007	0,080
2	0,09 – 0,18	81	0,081	0,007	0,161
3	0,18 – 0,27	93	0,093	0,008	0,254
4	0,27 – 0,36	85	0,085	0,008	0,339
5	0,36 – 0,45	87	0,087	0,008	0,426
6	0,45 – 0,54	87	0,087	0,008	0,513
7	0,54 – 0,63	87	0,087	0,008	0,600
8	0,63 – 0,72	106	0,106	0,010	0,706
9	0,72 – 0,81	99	0,099	0,009	0,805
10	0,81 – 0,90	89	0,089	0,008	0,894
11	0,90 – 0,99	106	0,106	0,010	1,000

3.3. Поправки Шеппарда

Выборочные моменты можно приближенно найти по формулам

$$\begin{aligned}\bar{X} &\approx \frac{1}{n} \sum_{i=1}^k \bar{X}_i n_i, \\ m_s &\approx \frac{1}{n} \sum_{i=1}^k (\bar{X}_i - \bar{X})^s n_i, \quad \text{в частности } s^2 = \frac{1}{n} \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 n_i, \quad \text{где сумми-} \\ a_s &\approx \frac{1}{n} \sum_{i=1}^k \bar{X}_i^s n_i,\end{aligned}$$

рование ведется по всем интервалам. Попробуем оценить погрешность при использовании этих приближенных формул. Будем считать, что интервалы охватывают всю числовую ось (n_i могут быть равными 0).

Пусть δ — случайная точка из отрезка $[0, h]$, примем ее за середину некоторого интервала, тогда границы этого интервала будут $\delta - h/2$, $\delta + h/2$; а границы i -го интервала вправо от δ : $\delta + (i - 1/2)h$, $\delta + (i + 1/2)h$. Эти границы — случайные величины, как функции δ . Середины интервалов имеют координаты: $\bar{X}_i = \delta + ih$.

Имеем $\tilde{a}_s = \frac{1}{n} \sum_{i=-\infty}^{\infty} n_i (\delta + ih)^s$. Выражение в правой части имеет два источника случайности: n_i и δ .

Найдем $\mathbf{E} \tilde{a}_s$:

$$\begin{aligned}\mathbf{E} \tilde{a}_s &= \frac{1}{h} \int_0^h d\delta \sum_{-\infty}^{\infty} \mathbf{E} \left(\frac{n_i}{n} \right) (\delta + ih)^s / \delta = \\ &= \frac{1}{h} \int_0^h d\delta \sum_{-\infty}^{\infty} \int_{\delta+(i-1/2)h}^{\delta+(i+1/2)h} f(x) dx \cdot (\delta + ih)^s = \frac{1}{h} \int_0^h d\delta \sum_{-\infty}^{\infty} (\delta + ih)^s \int_{\delta+(i-1/2)h}^{\delta+(i+1/2)h} f(x) dx.\end{aligned}$$

Пусть $\delta + ih = y$, тогда

$$\begin{aligned}\mathbf{E} \tilde{a}_s &= \frac{1}{h} \sum_{-\infty}^{\infty} \int_{ih}^{(i+1)h} y^s dy \int_{y-\frac{h}{2}}^{y+\frac{h}{2}} f(x) dx = \frac{1}{h} \int_{-\infty}^{\infty} y^s dy \int_{y-\frac{h}{2}}^{y+\frac{h}{2}} f(x) dx = \\ &= \frac{1}{h} \int_{-\infty}^{\infty} f(x) dx \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} y^s dy = \frac{1}{h} \int_{-\infty}^{\infty} \frac{(x + \frac{h}{2})^{s+1} - (x - \frac{h}{2})^{s+1}}{s+1} \cdot f(x) dx.\end{aligned}$$

Вычислим $\mathbf{E} \tilde{a}_1$:

$$\mathbf{E} \tilde{a}_1 = \frac{1}{h} \int_{-\infty}^{\infty} \frac{(x + \frac{h}{2})^2 - (x - \frac{h}{2})^2}{2} \cdot f(x) dx = \int_{-\infty}^{\infty} x f(x) dx = \mathbf{E} \xi = a.$$

Таким образом, оценка является точной. Далее

$$\begin{aligned} \mathbf{E} \tilde{a}_2 &= \frac{1}{h} \int_{-\infty}^{\infty} \frac{(x + \frac{h}{2})^3 - (x - \frac{h}{2})^3}{3} \cdot f(x) dx = \\ &= \frac{1}{h} \int_{-\infty}^{\infty} \frac{x^3 + 3x^2 \frac{h}{2} + 3x \frac{h^2}{4} + \frac{h^3}{8} - (x^3 - 3x^2 \frac{h}{2} + 3x \frac{h^2}{4} - \frac{h^3}{8})}{3} \cdot f(x) dx = \\ &= \frac{1}{h} \int_{-\infty}^{\infty} \frac{3x^2 h + \frac{h^3}{4}}{3} \cdot f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx + \frac{h^2}{12} \int_{-\infty}^{\infty} f(x) dx = a_2 + \frac{h^2}{12}. \end{aligned}$$

Получили более точную оценку для \tilde{a}_2 :

$$\tilde{\tilde{a}}_2 = \frac{1}{n} \sum_{i=1}^k \bar{x}_i^2 n_i - \frac{h^2}{12}.$$

Выпишем еще некоторые формулы:

$$\tilde{\tilde{a}}_3 = \frac{1}{n} \sum_{i=1}^k \bar{x}_i^3 n_i - \frac{h^2}{4n} \sum \bar{x}_i n_i,$$

$$m_1 \equiv 0,$$

$$\tilde{\tilde{m}}_2 = \frac{1}{n} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 n_i - \frac{h^2}{12}, \quad (\text{поправка равна } -\frac{h^2}{12})$$

$$\tilde{\tilde{m}}_3 = \frac{1}{n} \sum_{i=1}^k (\bar{x}_i - \bar{x})^3 n_i, \quad (\text{поправки нет})$$

$$\tilde{\tilde{m}}_4 = \frac{1}{n} \sum_{i=1}^k (\bar{x}_i - \bar{x})^4 n_i - \frac{h^2}{2} \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 n_i + \frac{7h^4}{240}.$$

Применяются также интерполяционные формулы для нахождения выборочных медианы, квантилей и моды по группированной выборке.

Медианным называется интервал, в котором накопленная сумма частот впервые достигает $\frac{1}{2}$. **Выборочной группированной медианой** называется значение m_e^* :

$$m_e^* = x_e + h \cdot \frac{n/2 - (n_1 + \dots + n_{m_e-1})}{n_{m_e}},$$

где n — объем выборки, h — длина интервала группировки, x_e — левая граница медианного интервала, n_i — численность i -го интервала, n_{m_e} — численность медианного интервала.

Квантильным порядка q интервалом называется интервал, в котором сумма накопленных частот впервые достигает значения q . **Выборочной группированной квантилью** называется значение X_q^* :

$$X_q^* = X_{(q)} + h \cdot \frac{nq - (n_1 + \dots + n_{(q)-1})}{n_{(q)}},$$

где $X_{(q)}$ — левая граница квантильного интервала, $n_{(q)}$ — численность квантильного интервала, $n_1, \dots, n_{(q)-1}$ — численности интервалов, предшествующих квантильному.

Модальным называется интервал, имеющий наибольшую численность. **Выборочной группированной модой** называется значение m_0^* :

$$m_0^* = x_0 + h \cdot \frac{n_{m_0} - n_{m_0-1}}{2n_{m_0} - n_{m_0-1} - n_{m_0+1}},$$

где x_0 — левая граница модального интервала, n_{m_0} — численность модального интервала, n_{m_0-1} , n_{m_0+1} — численности интервалов слева и справа от модального.

Пример 3.1. Дан группированный статистический ряд величины X :

X	1 — 5	5 — 9	9 — 13	13 — 17
n_i	2	8	9	1

Найти приближенно квантиль порядка 0,4.

◀ Квантильным интервалом является второй, так как в этом интервале сумма накопленных частот впервые достигает значения 0,4. Его левая граница $X_{0,4} = 5$.

$$X_{0,4}^* = 5 + 4 \cdot \frac{20 \cdot 0,4 - 2}{8} = 8.$$



3.4. Контрольные вопросы

1. *Каким образом при группировке выборки можно найти оптимальное число интервалов?*
2. *Каким образом при группировке выборки рассчитывается длина интервала?*
3. *Какие графические характеристики выборки дают представление о графике плотности?*
4. *Какие графические характеристики выборки дают представление о графике функции распределения?*
5. *Чему равняется площадь под гистограммой частот?*
6. *Можно ли восстановить по гистограмме относительных частот вариационный ряд?*
7. *Можно ли восстановить по гистограмме частот группированный ряд?*
8. *Дайте определение медианного интервала.*
9. *Дайте определение модального интервала.*
10. *Дайте определение квантильного интервала порядка 0,3.*
11. *Что уточняют поправки Шеппарда?*
12. *Чему равна поправка Шеппарда для группированного первого начального момента?*
13. *Чему равна поправка Шеппарда для группированного второго начального момента?*
14. *Как изменится поправка Шеппарда для группированного второго начального момента при увеличении длины интервала группировки?*

Лекция 4. Важные распределения математической статистики

*Мы встретимся еще с одним значением слова «статистика»
– так называют функцию от элементов выборки из
некоторой генеральной совокупности.*

М. Кендалл, А. Стюарт «Теория распределений»

План лекции: распределение хи-квадрат, распределение Стьюдента, распределение Фишера, теорема Фишера, теоремы о распределении выборочных характеристик.

Рассмотрим некоторые функции от выборки $X = (X_1, \dots, X_n)$ из нормального распределения, которые играют важную роль в математической статистике. Для данных функций (статистик) удастся получить в явном виде их функции распределения и плотности вероятностей. Вообще говоря, задача о распределении вероятностей для случайной функции от выборки $X = (X_1, \dots, X_n)$ в конечном виде решается крайне редко и функции от выборки из нормального распределения составляют приятное исключение. Нормальное распределение часто используют в прикладных исследованиях при построении математической модели изучаемого явления. Теоретическим обоснованием данного факта может служить центральная предельная теорема. Любую характеристику изучаемого явления, которая представима в виде суммы значительного числа случайных одинаково распределенных факторов, можно приблизительно считать случайной величиной с нормальным законом распределения. О роли нормального распределения в статистике хорошо и подробно написано в работах [16, 33].

Прежде чем рассматривать распределения функций от выборки из нормальной генеральной совокупности, введем определения распределений для трех специальных функций от нормальных случайных величин.

4.1. Распределения хи-квадрат, Стьюдента, Фишера

1. Распределение хи-квадрат (χ^2).

Среди гамма-распределений $\Gamma_{\alpha, \beta}$ выделяют особый класс, который часто используют в статистических исследованиях.

Определение 4.1. Гамма-распределение с параметрами $\alpha = \frac{1}{2}$, $\beta = \frac{n}{2}$ называется **распределением хи-квадрат с n степенями свободы** и обозначается χ_n^2 .

Соответствующая случайная величина обозначается тем же символом χ_n^2 , а ее плотность имеет вид

$$f_{\chi^2}(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}}, \quad x > 0. \quad (29)$$

На рис. 3 изображены графики плотности распределения хи-квадрат при некоторых значениях n .

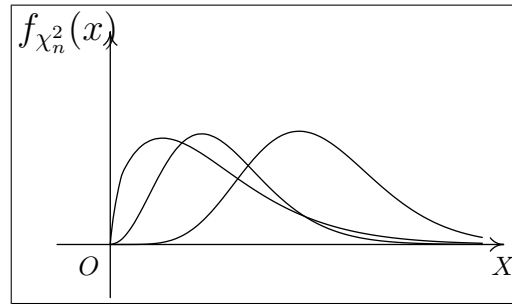


Рис. 3. Плотность распределения χ_n^2 при разных значениях параметра n

При увеличении числа степеней свободы график плотности становится все ближе к симметричному.

Свойства распределения хи-квадрат

1. Характеристическая функция $\varphi_{\chi_n^2}(t) = (1 - 2it)^{-n/2}$.

Доказательство. Для гамма-распределения $\Gamma_{\alpha, \beta}$ с функцией плотности

$$f_{\xi}(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)} e^{-\alpha x} x^{\beta-1} \text{ при } x \geq 0,$$

где $\Gamma(\beta)$ — гамма-функция, определяемая для всех $\beta > 0$ соотношением

$$\Gamma(\beta) = \int_0^{\infty} t^{\beta-1} e^{-t} dt,$$

характеристическая функция, как известно, равна $\varphi_{\xi}(t) = \left(\frac{\alpha}{\alpha - it}\right)^{\beta}$. Подставляя значения параметров $\alpha = \frac{1}{2}$, $\beta = \frac{n}{2}$, получим

$$\varphi_{\xi}(t) = \left(\frac{1}{1 - 2it}\right)^{n/2} = (1 - 2it)^{-n/2}. \square$$

2. $\mathbf{E} \chi_n^2 = n$, $\mathbf{D} \chi_n^2 = 2n$.

Доказательство. Из общих формул для распределения $\Gamma_{\alpha, \beta}$ ($\mathbf{E} \xi = \frac{\beta}{\alpha}$, $\mathbf{D} \xi = \frac{\beta}{\alpha^2}$) получаем $\mathbf{E} \chi_n^2 = n$, $\mathbf{D} \chi_n^2 = 2n$. \square

3. При $n \geq 2$ распределение χ_n^2 имеет моду $m_0 = n - 2$.

Доказательство. Достаточно проверить, что в этой точке достигается максимум плотности распределения. \square

4. Сумма квадратов независимых случайных величин, распределенных по закону $N(0, 1)$, распределена по закону хи-квадрат с числом степеней свободы, равным числу слагаемых, то есть

$$\sum_{i=1}^n \xi_i^2 = \chi_n^2.$$

Доказательство. Пусть ξ — случайная величина с распределением $N(0, 1)$. Найдем распределение случайной величины ξ^2 . Легко видеть, что для $x \leq 0$ функция распределения равна нулю, а для $x > 0$

$$F_{\xi^2}(x) = P(\xi^2 < x) = P(-\sqrt{x} \leq \xi < \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}), \quad (30)$$

где $\Phi(x)$ — функция стандартного нормального распределения.

$$f_{\xi^2}(x) = F'_{\xi^2}(x) = \frac{1}{2\sqrt{x}}(\varphi(\sqrt{x}) + \varphi(-\sqrt{x})) = \frac{1}{\sqrt{x}}\varphi(\sqrt{x}) = \frac{1}{\sqrt{\pi x}}e^{-\frac{x}{2}}. \quad (31)$$

Если рассмотреть распределение $\Gamma(\frac{1}{2}, \frac{1}{2})$, то получим следующую формулу для ее плотности:

$$f(x) = \frac{1}{\sqrt{\pi x}}e^{-\frac{x}{2}}, \quad x > 0. \quad (32)$$

Сравнивая (32) с плотностью в (31), получаем, что случайная величина ξ^2 имеет распределение $\Gamma(\frac{1}{2}, \frac{1}{2})$. Таким образом, случайная величина ξ^2 имеет распределение $\Gamma(\frac{1}{2}, \frac{1}{2})$ с характеристической функцией $(1 - 2it)^{-1/2}$, а сумма квадратов n независимых случайных величин, распределенных по нормальному закону $N(0, 1)$, имеет характеристическую функцию, равную произведению характеристических функций слагаемых¹⁷, то есть $(1 - 2it)^{-n/2}$, что соответствует распределению $\Gamma(\frac{1}{2}, \frac{n}{2}) = \chi_n^2$. \square

5. Сумма независимых случайных величин, распределенных по закону хи-квадрат, распределена также по закону хи-квадрат с числом степеней свободы, равным сумме степеней свободы слагаемых, то есть

$$\chi_n^2 + \chi_m^2 = \chi_{n+m}^2. \quad (33)$$

Доказательство. Характеристическая функция суммы

$$\varphi_{\chi_n^2 + \chi_m^2}(t) = \varphi_{\chi_n^2}(t) \cdot \varphi_{\chi_m^2}(t) = (1 - 2it)^{-n/2} \cdot (1 - 2it)^{-m/2} = (1 - 2it)^{-\frac{n+m}{2}},$$

а это характеристическая функция χ_{n+m}^2 . \square

¹⁷ Можно просто вспомнить о суммируемости гамма-распределения.

6.

$$\frac{\chi_n^2 - n}{\sqrt{2n}} \sim N(0, 1).$$

Доказательство.

$$\frac{\chi_n^2 - n}{\sqrt{2n}} = \frac{\sum_{i=1}^n \xi_i^2 - n}{\sqrt{2n}},$$

величины ξ_i^2 независимы и имеют одинаковое распределение $\Gamma(\frac{1}{2}, \frac{1}{2})$ с математическим ожиданием 1 и дисперсией 2. Поэтому $\mathbf{E}(\sum_{i=1}^n \xi_i^2) = n$, $\mathbf{D}(\sum_{i=1}^n \xi_i^2) = 2n$ и по ЦПТ

$$\frac{\sum_{i=1}^n \xi_i^2 - n}{\sqrt{2n}} \xrightarrow{d} u \in N(0, 1). \quad \square$$

7.

$$\frac{\chi_n^2}{n} \xrightarrow{p} \xi \equiv 1 \text{ (вырожденное в точке 1 распределение).}$$

Доказательство. Пусть ξ_1, \dots, ξ_n независимы и имеют стандартное нормальное распределение. Тогда $\mathbf{E} \xi_i^2 = \mathbf{D} \xi = 1$, и по ЗБЧ

$$\frac{\chi_n^2}{n} = \frac{\xi_1^2 + \dots + \xi_n^2}{n} \xrightarrow{p} 1. \quad \square$$

Пример 4.1. Используя свойства распределения χ_n^2 , найдем $\mathbf{D} S^2$ и $\mathbf{D} \bar{S}^2$ в $N(a, \sigma)$.



$$\frac{nS^2}{\sigma^2} = \chi_{n-1}^2, \quad S^2 = \frac{\chi_{n-1}^2 \cdot \sigma^2}{n}, \quad \mathbf{D} S^2 = \frac{\mathbf{D} \chi_{n-1}^2 \cdot \sigma^4}{n^2} = \frac{2(n-1)\sigma^4}{n^2}. \quad (34)$$

$$\bar{S}^2 = \frac{\chi_{n-1}^2 \cdot \sigma^2}{(n-1)}, \quad \mathbf{D} \bar{S}^2 = \frac{\mathbf{D} \chi_{n-1}^2 \cdot \sigma^4}{(n-1)^2} = \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2}{n-1}\sigma^4. \quad (35)$$



2. Распределение Стьюдента.

Рассмотрим последовательность независимых случайных величин ξ, ξ_1, \dots, ξ_n , распределенных по нормальному закону $N(0, 1)$.

Определение 4.2. *Распределением Стьюдента T_n с n степенями свободы называется распределение случайной величины*

$$t_n = \frac{\xi}{\sqrt{\frac{\sum_{i=1}^n \xi_i^2}{n}}}. \quad (36)$$

Можно случайную величину t определить как отношение

$$t_n = \frac{\xi}{\sqrt{\frac{\chi^2}{n}}},$$

где χ^2 — случайная величина, независимая от ξ и распределенная по закону χ_n^2 . Очевидно, что оба определения эквивалентны.

Плотность распределения Стьюдента $f_{T_n}(x)$ (рис. 4) нетрудно найти с помощью метода вычисления плотности распределения частного двух независимых величин (см. соответствующую задачу в § 4 практикума по решению задач).

Приведем окончательную формулу для плотности распределения Стьюдента:

$$f_{T_n}(x) = \frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \frac{1}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}, \quad x \in R. \quad (37)$$

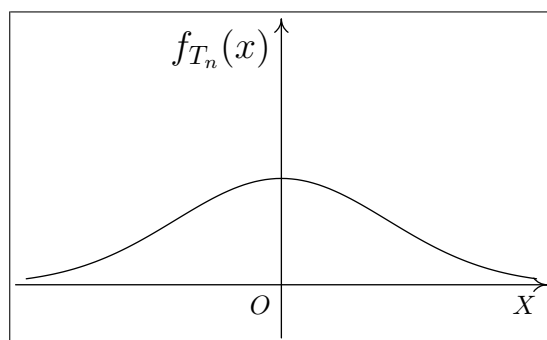


Рис. 4. Плотность распределения Стьюдента $f_{T_n}(x)$

Свойства распределения Стьюдента (см. задачи § 5 практикума по решению задач).

1. Распределение Стьюдента симметрично.
2. $\mathbf{E} T_n = 0$, $\mathbf{D} T_n = \begin{cases} \frac{n}{n-2} & \text{при } n > 2, \\ \infty & \text{при } n \leq 2. \end{cases}$
3. $t_n \xrightarrow{d} u \in N(0, 1)$
4. У распределения Стьюдента T_n существуют только моменты порядков меньших n , при этом все существующие моменты нечетного порядка $m = 2k + 1$ равны нулю.

Замечание 4.1. При $n = 1$ распределение Стьюдента совпадает с распределением Коши.

Замечание 4.2. Распределение Стьюдента можно рассматривать при $n \in \mathbb{R}^+$.

3. Распределение Фишера – Снедекора¹⁸.

Рассмотрим две независимые случайные величины χ_n^2 и χ_m^2 , имеющие хи-квадрат распределение с n и m степенями свободы соответственно.

Определение 4.3. *Распределением Фишера – Снедекора (F-распределением или распределением дисперсионного отношения Фишера) с n, m степенями свободы называется распределение случайной величины*

$$F_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}. \quad (38)$$

Функцию и плотность распределения $F_{n,m}(x)$ (рис. 5) можно найти, используя тот же способ, что и для распределения Стьюдента. Окончательно формулу для расчета плотности распределения Снедекора записывают в виде

$$f_{F_{n,m}}(x) = \left(\frac{n}{m}\right)^{\frac{n}{2}} \cdot \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \cdot \frac{x^{\frac{n}{2}-1}}{(1 + \frac{nx}{m})^{\frac{n+m}{2}}}, \quad x \in \mathbb{R}^+. \quad (39)$$

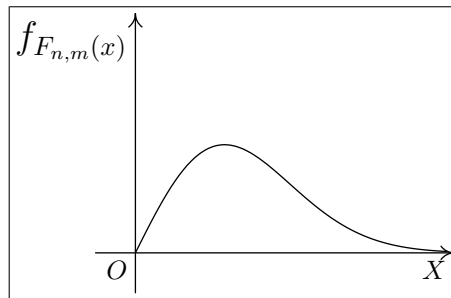


Рис. 5. Плотность распределения Фишера $f_{F_{n,m}}(x)$

Из свойств распределения Фишера отметим следующие (см. соответствующие задачи § 5 практикума по решению задач).

1. $\mathbf{E} F_{n,m} = \frac{m}{m-2}$ при $m > 2$.
2. $\mathbf{D} F_{n,m} = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ при $m > 4$.

¹⁸ Джордж Уоддел Снедекор (англ. *George Waddel Snedecor*; 1881–1974) – американский математик и статистик, ученик Рональда Фишера.

3. При $n \geq 2$ распределение $F_{n,m}$ имеет моду $m_0 = \frac{(n-m)m}{n(m+2)}$.
4. При $n \rightarrow \infty, m \rightarrow \infty$ $f_{n,m} \xrightarrow{d} \xi \equiv 1$ (вырожденное в точке 1 распределение).
5. Если $f_{n,m}$ имеет распределение Фишера $F_{n,m}$, то $1/f_{n,m}$ имеет распределение Фишера $F_{m,n}$.

Выясним распределение выборочных характеристик нормальной выборки.

4.2. Теорема Фишера

Легко видеть, какое распределение имеет выборочное среднее, найденное по выборке из совокупности с нормальным распределением $N(a, \sigma)$.

Теорема 4.1. Пусть X_1, \dots, X_n — выборка из распределения $N(a, \sigma)$. Тогда:

1. $\frac{X_i - a}{\sigma} \in N(0; 1)$;
2. $\sum_{i=1}^n \left(\frac{X_i - a}{\sigma}\right)^2 \in \chi_n^2$;
3. $\bar{X} \in N(a, \frac{\sigma}{\sqrt{n}})$;
4. $\frac{(\bar{X} - a)\sqrt{n}}{\sigma} \in N(0, 1)$.

Доказательство.

1. Нормальное распределение инвариантно относительно линейного преобразования, поэтому величина $u_i = \frac{X_i - a}{\sigma}$ нормально распределена; параметры этого распределения находятся как математическое ожидание и корень из дисперсии u_i , очевидно, они равны соответственно 0 и 1.
2. По свойству 4 распределения χ^2 сумма квадратов независимых случайных величин, распределенных по закону $N(0, 1)$, распределена по закону хи-квадрат с числом степеней свободы, равным числу слагаемых, то есть

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n \left(\frac{X_i - a}{\sigma}\right)^2 \in \chi_n^2.$$

3. Линейная комбинация нормальных величин есть нормальная величина. По свойствам среднего выборочного $\mathbf{E} \bar{X} = a$, $\mathbf{D} \bar{X} = \frac{\sigma^2}{n}$, это определяет параметры распределения.
4. Очевидно, $\mathbf{E} \left(\frac{(\bar{X}-a)\sqrt{n}}{\sigma} \right) = 0$, $\mathbf{D} \left(\frac{(\bar{X}-a)\sqrt{n}}{\sigma} \right) = 1$. \square

Какое распределение имеет выборочная дисперсия, найденная по выборке из совокупности с нормальным распределением $N(a, \sigma)$? На этот вопрос отвечает теорема, доказанная Р. Фишером.

Теорема 4.2 (теорема Фишера). Пусть X_1, \dots, X_n — выборка из распределения $N(a, \sigma)$. Тогда 1) величина $\frac{nS^2}{\sigma^2}$ имеет распределение χ_{n-1}^2 ; 2) \bar{X}, S^2 независимы.

Доказательство.

Нам потребуются две леммы.

Лемма 1. Пусть вектор $X = (X_1, \dots, X_n)^T$ состоит из n независимых случайных величин с нормальным распределением $N(0, \sigma)$, C — ортогональная матрица размерности $n \times n$ ($CC^T = C^TC = E$) и $Y = CX$. Тогда и координаты вектора $Y = (Y_1, \dots, Y_n)^T$ независимы и имеют нормальное распределение $N(0, \sigma)$, причем

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2.$$

Доказательство. Нормальное распределение инвариантно относительно линейного преобразования. $\mathbf{E} Y = \mathbf{E}(C \cdot X) = C \cdot \mathbf{E} X = 0$, где 0 , конечно, означает нулевой столбец размерности n .

Найдем ковариационную матрицу K_Y . Заметим сначала, что, поскольку величины X_1, \dots, X_n независимы и имеют одинаковую дисперсию σ^2 , ковариационная матрица $K_X = \sigma^2 E$.

$$\begin{aligned} K_Y &= \mathbf{E}(Y \cdot Y^T) = \mathbf{E}(CX \cdot (CX)^T) = \mathbf{E}(CX \cdot X^T C^T) = \\ &= C \mathbf{E}(X \cdot X^T) C^T = C K_X C^T = C \cdot \sigma^2 E \cdot C^T = \sigma^2 C C^T = \sigma^2 E. \end{aligned}$$

Отсюда следует, что величины Y_1, \dots, Y_n также независимы и имеют одинаковую дисперсию σ^2 .

Докажем последнее утверждение леммы 1.

$$\sum_{i=1}^n Y_i^2 = Y^T Y = (CX)^T CX = X^T C^T CX = X^T X = \sum_{i=1}^n X_i^2. \square$$

Лемма 2 (лемма Фишера). Пусть даны p линейных комбинаций, $p < n$:

[illegible]

коэффициенты которых удовлетворяют условию ортонормальности, а вектор $X = (X_1, \dots, X_n)^T$ состоит из n независимых случайных величин с нормальным распределением $N(0, \sigma)$. Тогда величина $Q(X) = \sum_{i=1}^n X_i^2 - \sum_{i=1}^p Y_i^2$ распределена как сумма $(n - p)$ квадратов независимых величин, имеющих распределение $N(0, \sigma)$.

Доказательство. Дополним матрицу $C'_{p \times n}$ коэффициентов линейных комбинаций до ортогональной матрицы $C_{n \times n}$, и пусть $Y = CX$. Тогда по лемме 1

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2,$$

И

$$Q(X) = \sum_{i=1}^n X_i^2 - \sum_{i=1}^p Y_i^2 = \sum_{i=1}^n Y_i^2 - \sum_{i=1}^p Y_i^2 = \sum_{i=p+1}^n Y_i^2. \square$$

Доказательство пункта 1 теоремы Фишера.

$$\begin{aligned} nS^2 &= \sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - \overline{X}^2 = \\ &= \sum_{i=1}^n X_i^2 - \left(\frac{X_1}{\sqrt{n}} + \frac{X_2}{\sqrt{n}} + \dots + \frac{X_n}{\sqrt{n}} \right)^2. \end{aligned} \quad (40)$$

Положим

$$Y_1 = \left(\frac{X_1}{\sqrt{n}} + \frac{X_2}{\sqrt{n}} + \dots + \frac{X_n}{\sqrt{n}} \right).$$

Статистика $nS^2 = \sum_{i=1}^n X_i^2 - Y_1^2$ и по лемме 2 распределена как сумма $(n-1)$ квадратов независимых величин, имеющих распределение $N(0, \sigma)$:

$$nS^2 = \sum_{i=2}^n Y_i^2, \quad Y_i \in N(0, \sigma).$$

Тогда статистика

$$\frac{nS^2}{\sigma^2} = \sum_{i=2}^n \left(\frac{Y_i}{\sigma} \right)^2 = \sum_{i=1}^{n-1} Z_i^2, \quad Z_i \in N(0, 1) = \chi_{n-1}^2.$$

Доказательство пункта 2 теоремы Фишера.

\bar{X} выражается через Y_1 , а S^2 через Y_2, \dots, Y_n , но по лемме 1 Y_i с разными индексами независимы, следовательно, независимы \bar{X} и S^2 как функции независимых величин. \square

Замечание 4.3. Обратите внимание, что в выражении для S^2 присутствует \bar{X} , то есть они функционально зависимы. Тем не менее, они, как только что доказано, являются независимыми случайными величинами. Это свойство характерно только для нормального распределения.

4.3. Теоремы о распределении выборочных характеристик

Теорема Фишера позволяет делать статистические выводы относительно дисперсии генеральной совокупности. На практике часто возникают ситуации, когда дисперсия генеральной совокупности неизвестна и необходимо делать выводы относительно математического ожидания. В этом случае оказывается полезной следующая теорема.

Теорема 4.3. Пусть X_1, \dots, X_n — выборка из распределения $N(a, \sigma)$ и функция от выборочных среднего и дисперсии t определена равенством

$$t = \sqrt{n-1} \frac{\bar{X} - a}{S}. \quad (41)$$

Тогда величина t , так называемое «студентово отношение»¹⁹, имеет распределение T_{n-1} .

Доказательство. По определению, распределением Стьюдента T_n с n степенями свободы называется распределение случайной величины

$$t = \frac{\xi}{\sqrt{\frac{\sum_{i=1}^n \xi_i^2}{n}}},$$

где ξ, ξ_1, \dots, ξ_n распределены по нормальному закону $N(0, 1)$ и независимы. Представим величину t в виде:

$$t = \frac{(\bar{X} - a)\sqrt{n}}{\sigma} \cdot \sqrt{\frac{n-1}{n}} \frac{\sigma}{S}.$$

¹⁹ Распределение студентова отношения впервые было получено Стьюдентом, затем эти результаты были строго доказаны Р. Фишером.

Первый сомножитель $\frac{(\bar{X} - a)}{\sqrt{n}}$ имеет распределение $N(0; 1)$; обозначим его ξ . Представим второй сомножитель в виде

$$\sqrt{\frac{n-1}{n}} \frac{\sigma}{S} = \sqrt{n-1} \sqrt{\frac{\sigma^2}{nS^2}}.$$

По теореме Фишера величина $\frac{nS^2}{\sigma^2}$ имеет распределение χ_{n-1}^2 . Таким образом, второй сомножитель равен

$$\sqrt{n-1} \sqrt{\frac{\sigma^2}{nS^2}} = \frac{1}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}.$$

Тогда

$$t = \frac{(\bar{X} - a)\sqrt{n}}{\sigma} \cdot \frac{1}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} = \frac{\xi}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}},$$

и очевидно, что эта величина имеет распределение Стьюдента T_{n-1} . \square

Замечание 4.4. Такое же распределение имеет тождественно равная t статистика $\sqrt{n} \frac{\bar{X} - a}{S}$.

Пусть нам необходимо выяснить, взяты ли две выборки из одного нормального распределения или из различных. Для ответа на этот вопрос важное значение имеет теорема (4.4).

Теорема 4.4. Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые выборки из распределения $N(a, \sigma)$, а $\bar{X}, \bar{Y}, S_X^2, S_Y^2$ — выборочные средние и дисперсии, и пусть

$$t = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{X} - \bar{Y}}{\sqrt{nS_X^2 + mS_Y^2}}. \quad (42)$$

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

Теорема доказывается аналогично теореме (4.3).

Если имеются две выборки из различных нормальных распределений, то представляет интерес следующая теорема 4.5.

Теорема 4.5. Пусть $X = (X_1, \dots, X_n) \in N(a_x, \sigma)$ и $Y = (Y_1, \dots, Y_m) \in N(a_y, \sigma)$ независимы. Тогда

$$Z = \frac{\bar{S}_X^2}{\bar{S}_Y^2} = f_{n-1, m-1},$$

то есть Z распределено по закону Фишера с $n - 1$, $m - 1$ степенями свободы.

Доказательство. По определению распределения Фишера

$$f_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}.$$

Из теоремы Фишера следует, что

$$\chi_{n-1}^2 = \frac{(n-1)\bar{S}^2}{\sigma^2} \Rightarrow \bar{S}^2 = \frac{\chi_{n-1}^2 \sigma^2}{n-1}.$$

Тогда

$$Z = \frac{\frac{\chi_{n-1}^2 \sigma^2}{n-1}}{\frac{\chi_{m-1}^2 \sigma^2}{m-1}} = \frac{\chi_{n-1}^2}{\chi_{m-1}^2} = f_{n-1,m-1}. \quad \square$$

Таким образом, когда дисперсии совпадают, величина $Z = \frac{\bar{S}_X^2}{\bar{S}_Y^2}$ не зависит от неизвестного параметра σ и имеет распределение $F_{n-1,m-1}$. Статистика $\frac{\bar{S}_X^2}{\bar{S}_Y^2}$ носит название «дисперсионное отношение» и связана с именем Фишера²⁰.

Эта теорема является частным случаем следующей:

Теорема 4.6. Пусть X_1, \dots, X_n и Y_1, \dots, Y_m — независимые выборки из распределений $N(a_1, \sigma_1)$, $N(a_2, \sigma_2)$, а S_X^2, S_Y^2 — выборочные дисперсии. Тогда случайная величина

$$F = \frac{n(m-1)\sigma_2^2 S_X^2}{m(n-1)\sigma_1^2 S_Y^2} \quad (43)$$

распределена по закону Фишера $F_{n-1,m-1}$.

Доказывается теорема 4.6 аналогично предыдущей.

4.4. Контрольные вопросы

1. Каково распределение суммы квадратов десяти независимых случайных величин, распределенных по нормальному закону $N(0, 1)$?

²⁰ Фишер использовал для дисперсионного отношения величину Z , связанную с F равенством $e^{2Z} = F$; Z -распределение было табулировано Фишером, F -распределение — Снедекором.

2. Каково распределение суммы пяти независимых случайных величин, каждая из которых распределена по закону χ_n^2 ?
3. Изобразите схематически на одном чертеже графики плотностей распределений плотности $\chi_{n_1}^2$, $\chi_{n_2}^2$ при $n_1 < n_2$.
4. Чему равняются математическое ожидание и дисперсия распределения χ_n^2 ?
5. Дайте определение распределения Стьюдента T_n .
6. Чему равняются математическое ожидание и дисперсия распределения Стьюдента?
7. Если нарисовать на одном чертеже графики плотности распределения $N(0, 1)$ и плотности распределения Стьюдента, чем они будут отличаться?
8. Каковы отличия графиков плотностей T_3 и T_4 ?
9. Как определяется распределение Фишера?
10. Пусть X_1, \dots, X_n — выборка из распределения $N(a, \sigma)$. Каково распределение величин \bar{X} , $\bar{X} - a$, $\frac{(\bar{X} - a)\sqrt{n}}{\sigma}$?
11. Сформулируйте теорему Фишера.

Лекция 5. Точечное статистическое оценивание

*Мораль здесь такова: позаботься о смысле,
а слова позаботятся о себе сами.*
Льюис Кэрролл ²¹

План лекции: оценки параметров, несмещенность, состоятельность.

5.1. Оценки параметров

Мы уже умеем по выборкам (X_1, \dots, X_n) вычислять выборочные статистические характеристики. По значениям этих характеристик можно делать вывод о соответствующих характеристиках генеральной совокупности (теоретических). Например, найденное по выборке \bar{X} дает представление о значении математического ожидания $\mathbf{E} \xi$; S^2 о значении дисперсии $\mathbf{D} \xi$ и т. д. При этом для каждой характеристики требуется своя формула. Возможен другой подход. Заметим, что все числовые характеристики случайных величин X_1, \dots, X_n зависят от параметра θ . Так, например, если $X \in E_\lambda$, то

$$\mathbf{E}(X) = \frac{1}{\lambda}, \quad \mathbf{D}(X) = \frac{1}{\lambda^2}.$$

Все другие числовые характеристики также выражаются через λ . Если мы найдем (приближенно) значение λ , то сумеем оценить значение любой теоретической характеристики.

Тогда возникает проблема оценки параметров генеральной совокупности. При решении этой проблемы мы будем использовать функции от выборки $g(X_1, \dots, X_n)$. Функции от выборки принято называть статистиками. Статистикой можно назвать любую функцию $\theta = \theta(X_1, \dots, X_n)$, которая не зависит от параметра θ . Условимся еще, что все упоминаемые функции измеримы.

Всякая статистика — функция от выборки $g(X_1, \dots, X_n)$ и сама является случайной величиной. Как случайная величина она имеет закон распределения, зависящий от закона распределения случайной величины ξ из генеральной совокупности.

В теории оценивания предполагается следующая модель: известно, что функция распределения случайной величины ξ принадлежит классу функций F_θ с неизвестным параметром (в общем случае векторным) $\theta \in R^k$. Задача заключается в выборе такой статистики $g(X_1, \dots, X_n)$, значения которой при заданной реализации (x_1, \dots, x_n) приближаются к значению па-

²¹ «Алиса в стране чудес».

параметра θ . Получение приближенного значения будем называть **оцениванием**.

Определение 5.1. Выборочная числовая характеристика (статистика) $\hat{\theta} = g(X_1, \dots, X_n)$, применяемая для оценивания неизвестного параметра θ генеральной совокупности, называется его **точечной оценкой**.

Таким образом, оценка — это статистика, предназначенная для оценивания неизвестного параметра θ .

Некоторые оценки мы уже рассматривали.

Пример 5.1. \bar{X} — оценка $E\xi$; S^2 — оценка $D\xi$.

Замечание 5.1. В дальнейшем для оценок параметра θ будут часто использоваться обозначения $\hat{\theta}$, $\tilde{\theta}$ и θ^* .

В общем случае для неизвестного параметра θ может существовать много оценок $\hat{\theta}$, которые близки к θ . Например, во многих распределениях математическое ожидание $E\xi$ является параметром (очевидные примеры — нормальное и пуассоновское распределения). Оценками для $E\xi$ могут быть такие статистики как выборочное среднее, выборочные мода и медиана (и не только они). Необходимо выбрать среди всех оценок данного параметра θ «наилучшую» в некотором смысле, то есть удовлетворяющую некоторым заданным условиям или обладающую некоторыми свойствами.

Обратите внимание, что в определении оценки (статистика, применяемая для оценивания неизвестного параметра) не оговорена близость оценки к оцениваемому параметру. Широкое определение, включающее в себя и совершенно неподходящие для использования в качестве оценок функции, — это характерный для математической статистики прием. Теперь надо выбрать из всего множества статистик самые лучшие.

«Чтобы что-то ненужное продать, надо сначала что-то ненужное купить» (Кот Матроскин).

И прежде всего надо выяснить, что нужно, а что ненужное, то есть выяснить критерии качества оценок.

Рассмотрим желательные для оценок свойства, чтобы иметь возможность сравнивать оценки и из всего множества статистик отбирать самые подходящие.

5.2. Несмещенность

Гайавата, как обычно,
Вверх свои направил стрелы.
Он так ловко это сделал,
Что остался несмещенным,
Но при этом, к сожалению,
В цель ни разу не попал.
М. Кендалл²²

Определение 5.2. Статистика $\hat{\theta} = g(X_1, \dots, X_n)$ называется **несмещенной оценкой** для параметра θ , если для любого $\theta \in \Theta$

$$\mathbf{E} \hat{\theta} = \theta.$$

Замечание 5.2. Математическое ожидание находится в предположении, что верна модель F_θ , то есть что параметр равен θ . Иногда это подчеркивают, записывая $\mathbf{E}_\theta \hat{\theta}$, но мы этот знак будем опускать.

Несмещенность — свойство оценок при фиксированном объеме выборки n . Это свойство означает, что в среднем оценка дает правильный результат, отсутствует смещение.

Пример 5.2. $\hat{\theta} = \bar{X}$ — несмещенная оценка для $\theta = \mathbf{E} \xi$, так как $\mathbf{E} \bar{X} = \mathbf{E} \xi$. Например, в $P_\lambda \bar{X}$ — несмещенная оценка λ .

Класс всех несмещенных оценок параметра θ будем обозначать T_θ .

Если $\mathbf{E} \hat{\theta} \neq \theta$, то оценка называется **смещенной** и ее **смещение** равно $\mathbf{E} \hat{\theta} - \theta$.

Определение 5.3. Статистика $\hat{\theta} = g(X_1, \dots, X_n)$ называется **асимптотически несмещенной** оценкой для параметра θ , если для любого $\theta \in \Theta$ при $n \rightarrow \infty$

$$\mathbf{E} \hat{\theta} \rightarrow \theta.$$

Пример 5.3. Пусть параметром распределения является его дисперсия. Исследовать на несмещенность оценку S^2 дисперсии σ^2 .

◀ По (17)

$$\mathbf{E} S^2 = \frac{(n-1)\mu_2}{n} = \frac{n-1}{n} \sigma^2,$$

следовательно, S^2 является асимптотически несмещенной оценкой для дисперсии. ▶

²² Отрывок из поэмы Кендалла про Гайавату и статистические оценки.

5.3. Состоятельность

— Состоятельность — это богатство?

— Нет, это гораздо лучше.

Консультация на экономическом факультете

Для несмещенных оценок возможна ситуация Гайаваты — все отдельные результаты неверны, так что само по себе это свойство еще не делает оценку пригодной.

При использовании той или иной оценки желательно, чтобы точность оценивания увеличивалась бы при возрастании объема выборки.

Определение 5.4. Статистика $\hat{\theta} = g(X_1, \dots, X_n)$ называется **состоятельной оценкой** θ , если для любого $\theta \in \Theta$

$$g(X_1, \dots, X_n) \xrightarrow{p} \theta. \quad (44)$$

Свойство состоятельности означает, что оценка делается точнее при увеличении количества данных.

Пример 5.4. $\hat{\theta} = \bar{X}$ — состоятельная оценка $\theta = \mathbf{E}\xi$, так как по закону больших чисел в форме Хинчина среднее независимых одинаково распределенных случайных величин сходится по вероятности при $n \rightarrow \infty$ к $\mathbf{E}\xi$. Например, в P_λ : \bar{X} — состоятельная оценка λ .

Таким образом, выборочное среднее является состоятельной оценкой математического ожидания в любом распределении.

Для исследования состоятельности оценок часто применяют следующие теоремы (см. задачи § 6 практикума по решению задач).

Теорема 5.1. Если $\mathbf{E}\hat{\theta} = \theta$ и $\mathbf{D}\hat{\theta} \rightarrow 0$ при $n \rightarrow \infty$, то $\hat{\theta}$ — состоятельная оценка θ .

Таким образом, несмещенная оценка, дисперсия которой стремится к 0 с ростом выборки, является состоятельной.

Часто применяется также теорема с ослабленными условиями:

Теорема 5.2. Если $\mathbf{E}\hat{\theta} \rightarrow \theta$ при $n \rightarrow \infty$ и $\mathbf{D}\hat{\theta} \rightarrow 0$ при $n \rightarrow \infty$, то $\hat{\theta}$ — состоятельная оценка θ .

Смысл ослабленного условия, очевидно, в том, что оценка может быть смещенной, но асимптотически несмещенной.

Следующая теорема утверждает, что непрерывные функции от состоятельных оценок тоже состоятельны.

Теорема 5.3. Если $\hat{\theta}$ — состоятельная оценка θ , а f — непрерывная функция, то $f(\hat{\theta})$ — состоятельная оценка $f(\theta)$.

Доказательство. Вспомним теорему о сходимости по вероятности функций случайных величин.

Пусть $f(x)$ — непрерывная функция. Тогда если последовательность $\{\xi_n\}$ сходится по вероятности к ξ , то и последовательность $\{f(\xi_n)\}$ сходится по вероятности к $f(\xi)$:

$$\xi_n \xrightarrow{p} \xi \Rightarrow f(\xi_n) \xrightarrow{p} f(\xi).$$

По условию $\hat{\theta} \xrightarrow{p} \theta$, следовательно, по теореме и $f(\hat{\theta}) \xrightarrow{p} f(\theta)$, что и означает состоятельность $f(\hat{\theta})$. \square

Пример 5.5. В P_λ статистика \overline{X}^3 — состоятельная оценка λ^3 , поскольку \overline{X} — состоятельная оценка λ .

Замечание 5.3. Если $\hat{\theta} \xrightarrow{n.n.} \theta$, то $\hat{\theta}$ называется **сильно состоятельной** оценкой θ .

Пример 5.6. Дана выборка (X_1, \dots, X_n) из распределения $B(N, p)$. Исследовать на несмещенность и состоятельность оценки параметра p : а) $\frac{X_1}{N}$; б) $\frac{2X_1 + 3X_2}{5N}$ в) $\frac{1}{5N} \sum_{i=1}^n X_i$.

◀ Все оценки являются несмещенными, так как

$$\mathbf{E} \frac{X_1}{N} = \mathbf{E} \frac{2X_1 + 3X_2}{5N} = \mathbf{E} \frac{1}{nN} \sum_{i=1}^n X_i = p.$$

В случаях а) и б) оценки, очевидно, не являются состоятельными (хотя бы потому, что они не зависят от объема выборки n). В случае в) оценка состоятельна по теореме 5.1, поскольку оценка несмещена и ее дисперсия стремится к 0 при $n \rightarrow \infty$. ▶

5.4. Контрольные вопросы

1. Дайте определение несмещенной оценки.
2. Дайте определение состоятельной оценки.

3. Дайте определение асимптотически несмещенной оценки.
4. Каким образом можно выяснить, является ли оценка несмещенной?
5. Каким образом можно выяснить, является ли оценка состоятельной?
6. При каком k оценка $\hat{p} = k \sum_{i=1}^n x_i$ является несмещенной в $B(N, p)$?
7. Приведите 2 примера несмещенных оценок параметра a в $N(a, \sigma)$.
8. Приведите пример смещенной оценки параметра a в $N(a, \sigma)$.
9. Приведите примеры смещенной и несмещенной оценок параметра σ в $N(a, \sigma)$.
10. Приведите 2 примера несмещенных оценок параметра p в $B(N, p)$.
11. Приведите 2 примера несмещенных оценок параметра λ в P_λ .
12. Верно ли, что \bar{X} в любом распределении является несмещенной оценкой математического ожидания?
13. Верно ли, что \bar{X} в любом распределении является состоятельной оценкой математического ожидания?
14. Чем отличается сильно состоятельная оценка от состоятельной?
15. Верно ли, что \bar{X} в любом распределении является сильно состоятельной оценкой математического ожидания?

Лекция 6. Неравенство Рао – Крамера

Хоть простота нужнее людям,
Но сложное понятней им.
Б. Пастернак

План лекции: оптимальность оценок, неравенство Рао – Крамера и его обобщения, различные формы информационного количества Фишера, эффективные оценки.

Может существовать несколько оценок, являющихся и несмещенными, и состоятельными. Как сравнить их и выбрать лучшую?

6.1. Оптимальность оценок

Определение 6.1. Пусть выбран критерий близости оценки к неизвестному параметру θ . Оценка $\hat{\theta}$ параметра θ называется **оптимальной по данному критерию** в рассматриваемом классе оценок, если она минимизирует выбранный критерий.

Пример 6.1. За критерий близости оценки к параметру θ можно взять $E(\hat{\theta} - \theta)^2$, где

$$\hat{\theta} = g(X_1, \dots, X_n) - \text{оценка } \theta. \quad (45)$$

Обратите внимание: если оценка $\hat{\theta}$ несмещена, то $E(\hat{\theta} - \theta)^2 = D\hat{\theta}$, где наименьшая дисперсия соответствует наиболее устойчивой оценке, которая меньше других меняется от выборки к выборке.

Несмещенные оценки с минимальной дисперсией оценки особенно важны; мы будем называть их просто оптимальными.

Определение 6.2. Несмещенная оценка $\hat{\theta}$ параметра θ называется **оптимальной** оценкой, если $D\hat{\theta} \leq D\tilde{\theta}$, $\forall \theta \in \Theta$, где $\tilde{\theta}$ – произвольная несмещенная оценка θ .

6.2. Неравенство Рао – Крамера

В случае регулярной статистической модели можно указать нижнюю границу дисперсий несмещенных оценок параметра θ .

Определение 6.3. Величина $\frac{\partial \ln f(x_1, x_2, \dots, x_n, \theta)}{\partial \theta}$ называется вкладом выборки, а величина $\frac{\partial \ln f(x, \theta)}{\partial \theta}$ называется вкладом одного элемента выборки.

Замечание 6.1. Плотность понимается в обобщенном смысле; для дискретных распределений вместо $f(x_1, x_2, \dots, x_n, \theta)$ используется $P(x_1, x_2, \dots, x_n, \theta)$.

Определение 6.4. Информационным количеством Фишера называется величина I , равная

$$I = \mathbf{E} \left(\frac{\partial \ln f(x_1, x_2, \dots, x_n, \theta)}{\partial \theta} \right)^2. \quad (46)$$

Для дисперсий несмещенных оценок параметра θ справедлива теорема:

Теорема 6.1 (неравенство Рао – Крамера). В регулярной модели $\langle F_\theta \rangle$ для любой оценки $\hat{\theta} \in T_\theta$ выполняется неравенство

$$\mathbf{D} \hat{\theta} \geq \frac{1}{I}. \quad (47)$$

Доказательство. Будем использовать упрощенные обозначения:

$$f(x_1, x_2, \dots, x_n, \theta) = f; \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} dx_1 \dots dx_n = \int dx.$$

1. По свойству плотности

$$\int f dx = 1. \quad (48)$$

2. Из несмещенности оценки

$$\int \hat{\theta} f dx = \mathbf{E} \hat{\theta} = \theta. \quad (49)$$

3. Продифференцируем (48) по параметру. Это возможно, так как модель регулярна.

$$\int \frac{\partial f}{\partial \theta} dx = 0. \quad (50)$$

4. Домножим (50) на θ :

$$\int \theta \frac{\partial f}{\partial \theta} dx = 0. \quad (51)$$

5. Продифференцируем (49) по параметру:

$$\int \hat{\theta} \frac{\partial f}{\partial \theta} dx = 1. \quad (52)$$

6. Вычтем (52) – (51):

$$\int (\hat{\theta} - \theta) \frac{\partial f}{\partial \theta} dx = 1. \quad (53)$$

$$\frac{\partial \ln f}{\partial \theta} = \frac{\partial f}{\partial \theta} \frac{1}{f} \Rightarrow \frac{\partial f}{\partial \theta} = f \frac{\partial \ln f}{\partial \theta}.$$

7. Подставим в (53) выражение для $\frac{\partial f}{\partial \theta}$:

$$\int (\hat{\theta} - \theta) \frac{\partial \ln f}{\partial \theta} f dx = 1. \quad (54)$$

Введем следующие обозначения:

$$\hat{\theta} = \varphi_1, \quad \theta = \mathbf{E} \varphi_1, \quad \frac{\partial \ln f}{\partial \theta} = \varphi_2.$$

8. Из курса теории вероятностей известно:

$$\mathbf{E}[(\varphi_1 - \mathbf{E} \varphi_1)(\varphi_2 - \mathbf{E} \varphi_2)] = \text{cov}(\varphi_1, \varphi_2). \quad (55)$$

Так как

$$\mathbf{E} \varphi_2 = \mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right) = \mathbf{E} \left(\frac{\partial f}{\partial \theta} \frac{1}{f} \right) = \int \frac{\partial f}{\partial \theta} \frac{1}{f} f dx = \int \frac{\partial f}{\partial \theta} dx \stackrel{(50)}{=} 0,$$

получим, что $\int (\hat{\theta} - \theta) \frac{\partial \ln f}{\partial \theta} f dx$ совпадает с ковариацией, то есть:

$$\text{cov} \left(\hat{\theta}, \frac{\partial \ln f}{\partial \theta} \right) = 1. \quad (56)$$

9. Из свойства коэффициента корреляции вытекает неравенство Коши – Буняковского:

$$|\text{cov}(\varphi_1, \varphi_2)| \leq \sqrt{\mathbf{D} \varphi_1 \mathbf{D} \varphi_2}. \quad (57)$$

Значит,

$$1 \leq \sqrt{\mathbf{D}(\hat{\theta}) \mathbf{D} \left(\frac{\partial \ln f}{\partial \theta} \right)}. \quad (58)$$

10. Преобразуем последнее выражение:

$$\mathbf{D}(\hat{\theta}) \geq \frac{1}{\mathbf{D} \left(\frac{\partial \ln f}{\partial \theta} \right)}. \quad (59)$$

$$\begin{aligned} \mathbf{D} \left(\frac{\partial \ln f}{\partial \theta} \right) &= \mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right)^2 - \mathbf{E}^2 \left(\frac{\partial \ln f}{\partial \theta} \right) = \mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right)^2 = I \\ &\Rightarrow (\mathbf{D} \hat{\theta}) \geq \frac{1}{I}. \quad \square \end{aligned}$$

Важное следствие: в неравенстве Рао – Крамера равенство достигается тогда и только тогда, когда $\hat{\theta}$ и $\frac{\partial \ln f}{\partial \theta}$ линейно зависимы (докажите это).

6.3. Формы информационного количества Фишера

Для проверки эффективности оценок удобно использовать следующие формулы информационного количества Фишера I , легко получаемые из (46):

$$I = n \mathbf{E} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2; \quad (60)$$

$$I = -n \mathbf{E} \left(\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} \right). \quad (61)$$

Обратите внимание, что в этих формулах $f(x, \theta)$ — одномерная плотность.

Докажем сначала две формулы для вычисления информационного количества Фишера с помощью многомерной плотности.

1.

$$I = \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial \theta} \right)^2 \frac{1}{f} dx. \quad (62)$$

Доказательство.

$$\begin{aligned} \int_{-\infty}^{\infty} \left(\frac{\partial f}{\partial \theta} \right)^2 \frac{1}{f} dx &= \left| \frac{\partial \ln f}{\partial \theta} = \frac{\partial f}{\partial \theta} \frac{1}{f} \right| = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f}{\partial \theta} f \right)^2 \frac{1}{f} dx = \\ &= \underbrace{\int_{-\infty}^{\infty} \left(\frac{\partial \ln f}{\partial \theta} \right)^2 f dx}_{\mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right)^2} = I \end{aligned}$$

2.

$$I = - \mathbf{E} \left(\frac{\partial^2 \ln f}{\partial \theta^2} \right). \quad (63)$$

Доказательство.

$$\begin{aligned} \frac{\partial \ln f}{\partial \theta} &= \frac{\partial f}{\partial \theta} \frac{1}{f}; \\ \frac{\partial^2 \ln f}{\partial \theta^2} &= -\frac{1}{f^2} \left(\frac{\partial f}{\partial \theta} \right)^2 + \frac{1}{f} \frac{\partial^2 f}{\partial \theta^2}; \\ \mathbf{E} \left(\frac{\partial^2 \ln f}{\partial \theta^2} \right) &= \int_{-\infty}^{\infty} \frac{\partial^2 \ln f}{\partial \theta^2} f dx = \int_{-\infty}^{\infty} \left[-\frac{1}{f^2} \left(\frac{\partial f}{\partial \theta} \right)^2 + \frac{1}{f} \frac{\partial^2 f}{\partial \theta^2} \right] f dx = \\ &= - \underbrace{\int_{-\infty}^{\infty} \frac{1}{f} \left(\frac{\partial f}{\partial \theta} \right)^2 dx}_I + \underbrace{\int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial \theta^2} dx}_{=0} = -I. \end{aligned}$$

По (50) $\int_{-\infty}^{\infty} \frac{\partial f}{\partial \theta} dx = 0$, а тогда и $\int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial \theta^2} dx = 0$. \square

Теперь можно доказать соотношение (60):

$$I = n \mathbf{E} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2.$$

Доказательство.

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_1, \dots, x_n, \theta)}{\partial \theta} \right)^2 f(x_1, \dots, x_n) dx_1 \dots dx_n; \\ f(x_1, \dots, x_n, \theta) &= \prod_{i=1}^n f(x_i, \theta); \\ \ln f(x_1, \dots, x_n, \theta) &= \ln \prod_{i=1}^n f(x_i, \theta) = \sum_{i=1}^n \ln f(x_i, \theta); \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\sum_{i=1}^n \partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 \prod_{i=1}^n f(x_i, \theta) dx_1 \dots dx_n &= \\ = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \sum_{i=1}^n \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 \prod_{i=1}^n f(x_i, \theta) dx_1 \dots dx_n}_{\Sigma_1} + \\ + 2 \underbrace{\sum_{i \neq j}^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta} \frac{\partial \ln f(x_j, \theta)}{\partial \theta} \right) \prod_{i=1}^n f(x_i, \theta) dx_1 \dots dx_n}_{\Sigma_2}. \end{aligned}$$

Каждое слагаемое Σ_1 имеет вид

$$\begin{aligned} \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_1, \theta)}{\partial \theta} \right)^2 f(x_1) dx_1 \int_{-\infty}^{\infty} f(x_2) dx_2 \dots \int_{-\infty}^{\infty} f(x_n) dx_n &= \\ \left[\text{так как по свойству плотности } \int_{-\infty}^{\infty} f(x) dx = 1 \right] &= \\ = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x) dx. \end{aligned}$$

Первое слагаемое Σ_2 равно произведению двух интегралов

$$\int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_1, \theta)}{\partial \theta} \right) f(x_1) dx_1 \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_2, \theta)}{\partial \theta} \right) f(x_2) dx_2$$

и $n - 2$ интегралов вида

$$\int_{-\infty}^{\infty} f(x_i) dx_i,$$

и поэтому первое слагаемое имеет вид

$$\int_{-\infty}^{\infty} \left(\frac{\partial f(x_1, \theta)}{\partial \theta} \right) dx_1 \int_{-\infty}^{\infty} \left(\frac{\partial f(x_2, \theta)}{\partial \theta} \right) dx_2 \cdot 1 \dots \cdot 1 = 0.$$

Получаем, что

$$\begin{aligned} & \sum_{i=1}^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta} \right)^2 \prod_{i=1}^n f(x_i, \theta) dx_1 \dots dx_n + \\ & + 2 \sum_{i \neq j}^n \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x_i, \theta)}{\partial \theta} \frac{\partial \ln f(x_j, \theta)}{\partial \theta} \right) \prod_{i=1}^n f(x_i, \theta) dx_1 \dots dx_n = \\ & \sum_{i=1}^n \underbrace{\int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x) dx}_{\mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right)^2} = n \mathbf{E} \left(\frac{\partial \ln f}{\partial \theta} \right)^2. \quad \square \end{aligned}$$

Аналогично доказывается соотношение (61):

$$I = -n \mathbf{E} \left(\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2} \right).$$

6.4. Эффективные оценки

Определение 6.5. Несмещенная оценка $\hat{\theta}$ параметра θ называется **эффективной** оценкой θ , если для любого $\theta \in \Theta$

$$\mathbf{D} \hat{\theta} = \frac{1}{I}. \quad (64)$$

Замечание 6.2. Если оценка является эффективной, она является оптимальной. Обратное, вообще говоря, не верно.

Пример 6.2. Исследовать на эффективность оценку $\hat{a} = \bar{X}$ в нормальной статистической модели, если второй параметр известен и равен σ .

◀ Вычислим I .

$$\begin{aligned} f(x) &= \varphi_{a, \sigma}(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \\ \ln f(x) &= \ln \left(\frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{(x-a)^2}{2\sigma^2}, \end{aligned}$$

$$\frac{\partial \ln f(x, \theta)}{\partial \theta} = \frac{x - a}{\sigma^2}.$$

$$I = n \mathbf{E} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 = n \frac{\mathbf{E}(X - a)^2}{\sigma^4} = \frac{n}{\sigma^2}.$$

Найдем нижнюю границу дисперсий:

$$\frac{1}{I} = \frac{\sigma^2}{n}.$$

Вычислим дисперсию \bar{X} .

$$\mathbf{D} \bar{X} = \mathbf{D} \left(\frac{\sum X_i}{n} \right) = \frac{\sum \mathbf{D} X_i}{n^2} = \frac{\sigma^2}{n}.$$

$\mathbf{D} \bar{X} = \frac{1}{I}$, следовательно, оценка эффективна. ►

Замечание 6.3. Для дискретной случайной величины вместо $f(x)$ используется $P(\xi = x)$.

Пример 6.3. Доказать, что оценка $\hat{\lambda} = \bar{X}$ в P_λ эффективна.

◀

$$\mathbf{D} \hat{\lambda} = \frac{\mathbf{D} X}{n} = \frac{\lambda}{n}.$$

$$\ln P_\lambda(X) = X \ln \lambda - \lambda - \ln X!$$

$$I = -n \mathbf{E} \left(\frac{X}{\lambda^2} \right) = \frac{n}{\lambda}.$$

Получили, что $\mathbf{D} \hat{\lambda} = \frac{1}{I}$, то есть оценка \bar{X} является эффективной. ►

В регулярной статистической модели для несмещенных оценок можно рассматривать показатель эффективности.

Определение 6.6. Показателем эффективности несмещенной оценки $\hat{\theta}$ параметра θ называется число

$$e(\hat{\theta}) = \frac{1}{I \mathbf{D} \hat{\theta}}. \quad (65)$$

Очевидно, $0 < e(\hat{\theta}) \leq 1$. Для эффективных оценок $e(\hat{\theta}) = 1$.

6.5. Байесовский и минимаксный подходы к сравнению оценок

Пока что мы сравнивали несмещенные оценки параметра по их дисперсии $\mathbf{D} \hat{\theta}$ для любого $\theta \in \Theta$, а произвольные оценки по величине $d(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \theta)^2$ (так называемый **среднеквадратический подход**). Оценку $\hat{\theta}_1$ в соответствии со среднеквадратическим подходом мы считаем лучшей, чем $\hat{\theta}_2$, если $\mathbf{E}(\hat{\theta}_1 - \theta)^2 < \mathbf{E}(\hat{\theta}_2 - \theta)^2 \forall \theta \in \Theta$. Таким образом, задача сравнения оценок приводит к вопросу о сравнении функций $d(\hat{\theta})$. Однако множество функций $d(\hat{\theta})$ в общем случае неупорядоченное. Как же сравнить оценки, у которых разность $d(\hat{\theta}_1) - d(\hat{\theta}_2)$ при различных значениях параметра меняет знак? Существуют два подхода, которые позволяют упорядочить множество всех оценок с помощью одной числовой характеристики — это байесовский и минимаксный подходы.

В качестве числовой характеристики используется среднее или максимальное значение $d\hat{\theta}$ по множеству Θ значений параметра θ . Первый из этих способов называется **байесовским**, второй — **минимаксным**.

В первом случае неизвестный параметр θ рассматривается как случайная величина с некоторой (априорной) плотностью распределения $q(t)$.

Определение 6.7. *Байесовской оценкой параметра θ , соответствующей априорному распределению с плотностью $q(t)$, называется оценка θ^* , определенная формулой (66):*

$$\theta^* = \mathbf{E}(\theta/X) = \int t q(t/X) dt. \quad (66)$$

По свойствам условного математического ожидания для байесовской оценки безусловное среднеквадратическое уклонение $\mathbf{E}(\theta^* - \theta)^2$ принимает наименьшее возможное значение. Это означает, что если параметр θ — случайная величина с плотностью распределения $q(t)$, то байесовская оценка является наилучшей в среднеквадратическом смысле.

Пример 6.4. Пусть в модели $\langle B(N, p) \rangle$ параметр N известен, а параметр p имеет априорное бета-распределение: $B(\alpha, \beta)$. Найдём по измеренному значению $X \in B(N, p)$ байесовскую оценку \hat{p} .

◀ Совместная плотность $f_{X,p}$ и равна произведению одномерной плотности p на условную плотность X/p при фиксированном значении p :

$$f_{X,p} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \cdot C_N^X p^X (1-p)^{N-X} =$$

$$= C_N^X \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{X+\alpha-1} (1-p)^{N-X+\beta-1}.$$

Одномерная плотность X равна

$$f_X(x) = \int_0^1 f_{X,p} dp = C_N^X \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + X)\Gamma(N - X + \beta)}{\Gamma(N + \alpha + \beta)},$$

это бета-биномиальное распределение. Найдем условную плотность p/X :

$$f_{p/X} = \frac{f_{X,p}}{f_X} = \frac{\Gamma(N + \alpha + \beta)}{\Gamma(X + \alpha)\Gamma(N - X + \beta)} p^{X+\alpha-1} (1-p)^{N-X+\beta-1}.$$

Это бета-распределение $B(X + \alpha, N - X + \beta)$. Байесовская оценка \hat{p} равна математическому ожиданию $\mathbf{E}(p/X)$. Вспомним, что для бета-распределения математическое ожидание равно отношению первого параметра к их сумме и получим

$$\hat{p} = \frac{X + \alpha}{\alpha + \beta + N}.$$

Заметим, что математическое ожидание априорного распределения, то есть априорная оценка до наблюдений, равна $\frac{\alpha}{\alpha + \beta}$, а оценка по наблюдению, игнорирующая априорное распределение, равна $\frac{X}{N}$. Байесовская оценка находится между ними. ►

Определение 6.8. *Минимаксной оценкой параметра θ называется оценка $\bar{\theta}^*$, если для любой оценки $\hat{\theta}$ справедливо*

$$\sup_{\theta \in \Theta} d(\bar{\theta}^*) \leq \sup_{\theta \in \Theta} d(\hat{\theta}). \quad (67)$$

При минимаксном оценивании добиваются наилучшего результата в наихудшем случае; поэтому минимаксная оценка будет байесовской при наихудшем априорном распределении (доставляющем самые большие средние потери).

6.6. Контрольные вопросы

1. Дайте определение оптимальной оценки.
2. Дайте определение эффективной оценки.
3. Какова связь между эффективностью и оптимальностью?
4. Введите определение асимптотически эффективной оценки.

5. *Каким образом можно выяснить, является ли оценка эффективной?*
6. *В любой ли модели существует эффективная оценка?*
7. *Каковы особенности регулярной статистической модели?*
8. *Является ли регулярной нормальная модель?*
9. *Существует ли эффективная оценка в равномерной модели $R[0, \theta]$?*
10. *В любой ли регулярной модели существует эффективная оценка?*
11. *Может ли смещенная оценка быть эффективной?*
12. *Может ли смещенная оценка быть асимптотически эффективной?*

Лекция 7. Методы получения оценок

Дело к тому же минутное, если с умом.

М. Щербаков

План лекции: метод максимального правдоподобия, примеры, некоторые свойства оценок максимального правдоподобия, метод моментов, некоторые свойства оценок метода моментов.

7.1. Метод максимального правдоподобия

Метод максимального правдоподобия для получения точечных оценок был предложен Р. Фишером. Пусть генеральная совокупность определяется случайной величиной ξ с функцией распределения $F(x, \theta)$ и задана выборка (X_1, \dots, X_n) . Суть метода состоит в нахождении такого значения θ^* неизвестного параметра θ , при котором вероятность реализации (x_1, \dots, x_n) вектора (X_1, \dots, X_n) была бы максимальна.

Пусть ξ — непрерывная случайная величина с плотностью $f(x, \theta)$, где θ — неизвестный параметр. Тогда

$$f(x_1, \dots, x_n, \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)$$

— плотность распределения вектора (X_1, \dots, X_n) .

Определение 7.1. Для непрерывной случайной величины функция

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta), \quad (68)$$

рассматриваемая при фиксированных (x_1, \dots, x_n) как функция параметра θ , называется **функцией правдоподобия**.

Для дискретной случайной величины закон распределения задается вероятностями

$$P(\xi = x_i) = p_i;$$

ξ	x_1	x_2	\dots	x_n
p	p_1	p_2	\dots	p_n

Пусть имеется выборка (x_1, \dots, x_n) . Тогда **функция правдоподобия** для дискретной случайной величины определяется в виде

$$L(x_1, \dots, x_n, \theta) = P(\xi = x_1) \cdot \dots \cdot P(\xi = x_n). \quad (69)$$

Определение 7.2. Оценка θ^* , обеспечивающая по параметру θ максимум функции правдоподобия, называется оценкой максимального правдоподобия параметра θ (о.м.п.)

Для дискретной случайной величины это условие означает максимум вероятности получения реализации (x_1, \dots, x_n) при $\theta = \theta^*$, а для непрерывной величины плотность пропорциональна вероятности. Пусть $L(x_1, \dots, x_n, \theta)$ дифференцируема по θ для любой реализации (x_1, \dots, x_n) и достигает максимума по θ во внутренней точке Θ .

$$\frac{\partial L}{\partial \theta} = 0$$

или

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, \dots, k, \quad (70)$$

если $\theta \in R^k$.

Функции L и $\ln L$ достигают максимума при одном и том же значении θ , поэтому вместо отыскания максимума функции L часто удобнее находить максимум функции $\ln L$ и решать **уравнение правдоподобия**

$$\frac{\partial \ln L}{\partial \theta} = 0. \quad (71)$$

В результате решения уравнения правдоподобия мы найдем критическую точку, необходимо еще убедиться, что это точка максимума.

Пример 7.1. Найдем о.м.п. параметра α в распределении $\Gamma_{\alpha, \beta}$ при известном β .

◀

$$\begin{aligned} L &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{\alpha^\beta}{\Gamma(\beta)} e^{-\alpha x_i} x_i^{\beta-1} = \\ &= \left(\frac{\alpha^\beta}{\Gamma(\beta)} \right)^n \cdot \left(\prod_{i=1}^n x_i^{\beta-1} \right) \cdot \exp \left\{ -\alpha \sum_{i=1}^n x_i \right\}. \end{aligned}$$

$$\ln L = n\beta \ln \alpha - n \ln \Gamma(\beta) + (\beta - 1) \sum_{i=1}^n \ln x_i - \alpha \sum_{i=1}^n x_i.$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n\beta}{\alpha} - \sum_{i=1}^n x_i = 0 \quad \implies \quad \alpha = \frac{\beta}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{\beta}{\bar{x}}.$$

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{n\beta}{\alpha^2} < 0 \quad \implies \quad \alpha = \frac{\beta}{\bar{x}} - \text{точка максимума}$$

$$\Rightarrow \hat{\alpha} = \frac{\beta}{\bar{x}} - \text{о.м.п.}$$



7.2. Примеры

Рассмотрим другие примеры использования метода максимального правдоподобия.

Пример 7.2 (дискретная модель). Найдем о.м.п. параметра распределения Пуассона.



$$L = \prod_{i=1}^n P_{\lambda}(x_i) = \frac{e^{-\lambda n} \lambda^{\sum x_i}}{\prod (x_i!)}.$$

$$\ln L(X, \lambda) = -\lambda n + \sum x_i \ln \lambda - \ln \prod (x_i!).$$

Найдем $\max \ln L(X, \lambda)$:

$$\frac{\partial \ln L(X, \lambda)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} = 0.$$

Получаем $\hat{\lambda} = \frac{\sum x_i}{n} = \bar{x}$. Очевидно, это точка максимума, так как

$$\frac{\partial^2 \ln L}{\partial \lambda^2} < 0 \quad \Rightarrow \quad \bar{x} - \text{о.м.п. } \lambda.$$



Пример 7.3 (двумерный параметр). Пусть величина ξ распределена по нормальному закону, то есть имеет распределение N_{θ} , где $\theta = (a, \sigma)$. Найдем о.м.п. параметра θ .



Функция правдоподобия для ξ имеет следующий вид:

$$L(x_1, \dots, x_n) = \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - a)^2}{2\sigma^2}}.$$

Найдем логарифмическую функцию правдоподобия:

$$\ln L(x, a, \sigma) = -n(\ln \sigma + 0,5 \ln 2\pi) - \sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma^2}.$$

Найдем частные производные по a и по σ и приравняем их к нулю:

$$\begin{cases} \frac{\partial \ln L(x, a, \sigma)}{\partial a} = \sum_{i=1}^n \frac{(x_i - a)}{\sigma^2} = 0, \\ \frac{\partial \ln L(x, a, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma^3} = 0. \end{cases}$$

Из первого уравнения выразим a : $\sum_{i=1}^n (x_i - a) = 0 \Rightarrow \sum_{i=1}^n x_i - n \cdot a = 0$,

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

Из второго уравнения выразим σ^2 : $n\sigma^2 = \sum_{i=1}^n (x_i - a)^2$,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2 = S^2, \quad \hat{\sigma} = S.$$

Остается убедиться, что точка (\bar{x}, S^2) — точка максимума. Для этого надо составить матрицу вторых производных

$$A(a, \sigma) = \begin{pmatrix} \frac{\partial^2 \ln L(x, a, \sigma)}{\partial a^2} & \frac{\partial^2 \ln L(x, a, \sigma)}{\partial a \partial \sigma} \\ \frac{\partial^2 \ln L(x, a, \sigma)}{\partial \sigma \partial a} & \frac{\partial^2 \ln L(x, a, \sigma)}{\partial \sigma^2} \end{pmatrix}$$

и проверить отрицательную определенность матрицы A в точке $(a, \sigma) = (\bar{x}, S)$. ►

Пример 7.4 (нерегулярная модель). Найдем о.м.п. параметра $\theta = (a, b)$ в распределении $R[a, b]$.

◀

$$L = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{1}{b - a} = \frac{1}{(b - a)^n}.$$

$\frac{\partial \ln L}{\partial \theta}$ не обращается в 0. Но функция L монотонна по a и по b . Поэтому она достигает своего наибольшего значения при минимально возможном значении b и максимально возможном значении a . Таким образом, о.м.п. будут

$$\hat{a} = y_{\min} = x_1^*, \quad \hat{b} = x_{\max} = x_n^*.$$

►

Следует заметить, что не всегда существует максимум функции L по параметру и не всегда возможно найти аналитическое решение уравнения правдоподобия (71).

7.3. Некоторые свойства оценок максимального правдоподобия

1. **Свойство инвариантности.** Если оценивается некоторая взаимно однозначная параметрическая функция $\tau(\theta)$, то ее о.м.п. $\widehat{\tau(\theta)} = \tau(\widehat{\theta})$. Это свойство вполне очевидно, так как точки максимума L , найденные по θ и по $\tau(\theta)$, совпадают. Из свойства инвариантности следует, что для нахождения о.м.п. можно выбирать наиболее удобную параметризацию, а о.м.п. получать затем с помощью соответствующих преобразований.

Пример 7.5. Найдем в условиях примера (7.1) о.м.п. α^3 .

◀ По свойству инвариантности

$$\widehat{\alpha^3} = (\widehat{\alpha})^3 = \left(\frac{\beta}{\bar{x}}\right)^3.$$

►

2. Оценки максимального правдоподобия **асимптотически несмещены, состоятельны** и при некоторых дополнительных предположениях о модели **асимптотически нормальны**. (Дополнительные предположения касаются мажорируемости третьей производной f по параметру и обычно выполняются в регулярных моделях.)
3. Если оценки максимального правдоподобия асимптотически нормальны, то они и **асимптотически эффективны**, то есть

$$\mathbf{D} \hat{\theta} \rightarrow \frac{1}{I}.$$

7.4. Метод моментов

Идея метода: выборочные моменты принимают в качестве оценок для моментов распределения случайной величины ξ , которые суть функции от неизвестного параметра θ (в общем случае многомерного). Рассмотрим случайную величину ξ с плотностью $f(x, \theta)$ и выборку объема n (x_1, \dots, x_n). У случайной величины ξ существуют моменты $\alpha_1, \dots, \alpha_r$, которые являются функциями от θ .

Выборочные моменты a_k вычисляют по формуле

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

Приравнивая выборочные и теоретические моменты, получаем уравнения относительно параметра θ . Пусть уравнения однозначно и непрерывно разрешимы относительно θ . Решая эти уравнения, получаем оценку $\hat{\theta}$ параметра θ . Эта оценка называется **оценкой метода моментов** и обозначается **о.м.м.**

Пример 7.6 (показательное распределение).

◀ Найдем о.м.м. параметра показательной модели. Плотность распределения ξ

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

где λ — неизвестный параметр. Как мы знаем,

$$\alpha_1 = \mathbf{E} \xi = \frac{1}{\lambda}, \quad a_1 = \bar{X}.$$

Приравняем $\alpha_1(\hat{\lambda})$ и a_1

$$\frac{1}{\hat{\lambda}} = \bar{X}$$

и получим оценку для λ :

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

►

Пример 7.7 (гамма-распределение).

◀ В распределении $\Gamma_{\alpha, \beta}$

$$\mathbf{E} \xi = \frac{\beta}{\alpha}, \quad \mathbf{D} \xi = \frac{\beta}{\alpha^2}.$$

Запишем систему уравнений

$$\begin{cases} \frac{\hat{\beta}}{\hat{\alpha}} = \bar{X} \\ \frac{\hat{\beta}}{(\hat{\alpha})^2} = S^2 \end{cases}$$

Решив эту систему, получим оценки

$$\hat{\alpha} = \frac{\bar{X}}{S^2}, \quad \hat{\beta} = \frac{\bar{X}^2}{S^2}.$$

►

Пример 7.8 (равномерное распределение).

◀ Рассмотрим равномерное распределение $R[a, b]$, где a и b — неизвестные параметры. Можно записать систему, разрешить ее относительно относительно параметров, а потом вместо теоретических моментов подставить выборочные.

$$\begin{aligned} \mathbf{E} \xi &= \frac{a+b}{2}, & \mathbf{D} \xi &= \frac{(b-a)^2}{12}, \\ a &= 2 \mathbf{E} \xi - b \implies \mathbf{D} \xi = \frac{(b - \mathbf{E} \xi)^2}{3}. \end{aligned}$$

Отсюда $b = \mathbf{E} \xi + \sigma\sqrt{3}$, $a = \mathbf{E} \xi - \sigma\sqrt{3}$. Окончательно:

$$\hat{a} = \bar{X} - s\sqrt{3}, \quad \hat{b} = \bar{X} + s\sqrt{3}.$$



Оценки метода моментов не единственны, можно брать различные моменты и получать различные о.м.м.

Пример 7.9 (равномерное распределение $R(0, \theta)$).

◀ Найдем о.м.м. по математическому ожиданию (первому моменту α_1):

$$\alpha_1 = \frac{\theta}{2} \implies a_1 = \frac{\hat{\theta}}{2}, \quad \hat{\theta} = 2a_1, \text{ или } \hat{\theta} = 2\bar{x}.$$

Найдем о.м.м. по k -му моменту α_k :

$$\alpha_k = \int_0^\theta x^k \frac{1}{\theta} dx = \frac{\theta^k}{k+1} \implies \hat{\theta} = \sqrt[k]{(k+1)a_k}.$$



7.5. Некоторые свойства оценок метода моментов

Из свойств оценок метода моментов отметим их **состоятельность**.

Теорема 7.1. Пусть $\hat{\theta} = g(a_1, \dots, a_k)$ — оценка параметра θ , полученная по методу моментов, причем функция g^{-1} непрерывна. Тогда $\hat{\theta}$ состоятельна.

Доказательство. Если $\hat{\theta} = g(a_1, \dots, a_k)$, то $\theta = g(\alpha_1, \dots, \alpha_k)$. По свойству выборочных моментов a_k сходятся по вероятности к α_k при $n \rightarrow \infty$. Тогда и $g(a_1, \dots, a_k) \xrightarrow{p} g(\alpha_1, \dots, \alpha_k)$ по теореме 2.1 о сходимости по вероятности:

$$\xi_{kn} \xrightarrow{p} \xi_k \implies \varphi(\xi_{1n}, \dots, \xi_{kn}) \xrightarrow{p} \varphi(\xi_1, \dots, \xi_k),$$

то есть $\hat{\theta} \xrightarrow{p} \theta$. \square

7.6. Контрольные вопросы

1. Дайте определение функции правдоподобия.
2. Чем отличается функция правдоподобия от n -мерной плотности в непрерывной модели?
3. Как найти функцию правдоподобия в дискретной модели?
4. В чем состоит идея метода максимального правдоподобия?
5. Запишите уравнение правдоподобия.
6. Существует ли о.м.п. в нерегулярной модели?
7. Верно ли, что о.м.п. единственна?
8. Каким образом можно найти о.м.м.?
9. Верно ли, что о.м.м. единственна?
10. Какими свойствами обладают о.м.п.?
11. Какими свойствами обладают о.м.м.?
12. Любая ли оценка, полученная методом моментов, является состоятельной?
13. Запишите известные вам о.м.п. для параметров различных моделей.
14. Запишите некоторые известные вам о.м.м. для параметров различных моделей.
15. Приведите пример совпадающих о.м.м. и о.м.п.
16. Приведите пример не совпадающих о.м.м. и о.м.п.

Лекция 8. Оценивание параметрической функции

Есть разные пути,
а истина одна.
Jay Thunder

План лекции: неравенство Рао — Крамера для параметрической функции, эффективная оценка параметрической функции, теоремы об эффективных оценках, экспоненциальное семейство.

8.1. Неравенство Рао — Крамера для параметрической функции

Определение эффективной оценки параметра 6.5 можно распространить и на параметрическую функцию. Рассмотрим задачу оценивания заданной параметрической функции $\tau(\theta)$ в модели $\langle F_\theta \rangle$. Предположим, что модель $\langle F_\theta \rangle$ регулярна, функция $\tau(\theta)$ дифференцируема и пусть $T_{\tau(\theta)}$ — класс всех несмещенных оценок $\tau(\theta)$. Тогда для дисперсий несмещенных оценок параметрической функции $\tau(\theta)$ справедливо **неравенство Рао — Крамера для параметрической функции**.

Теорема 8.1. Пусть $\langle F_\theta \rangle$ — регулярная модель, $T(x) \in T_{\tau(\theta)}$ — оценка параметрической функции $\tau(\theta)$. Тогда выполняется неравенство

$$\mathbf{D} T(x) \geq \frac{(\tau')^2}{I}. \quad (72)$$

Доказательство проводится как в теореме (6.1).

Доказательство. По-прежнему используем упрощенные обозначения:

$$f(x_1, x_2, \dots, x_n, \tau(\theta)) = f; \quad \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} dx_1 \dots dx_n = \int dx.$$

По свойству плотности

$$\int f dx = 1. \quad (73)$$

Из несмещенности оценки

$$\int T(x) f dx = \mathbf{E} T(x) = \tau(\theta). \quad (74)$$

Продифференцируем (73) по параметру. Это возможно, так как модель регулярна.

$$\int \frac{\partial f}{\partial \theta} dx = 0. \quad (75)$$

Домножим (75) на $\tau(\theta)$:

$$\int \tau(\theta) \frac{\partial f}{\partial \theta} dx = 0. \quad (76)$$

Продифференцируем (74) по параметру:

$$\int T(x) \frac{\partial f}{\partial \theta} dx = \tau'(\theta). \quad (77)$$

Вычтем (77) – (76):

$$\int (T(x) - \tau(\theta)) \frac{\partial f}{\partial \theta} dx = \tau'(\theta). \quad (78)$$

Подставим в (78) выражение для $\frac{\partial f}{\partial \theta} : \frac{\partial f}{\partial \theta} = f \frac{\partial \ln f}{\partial \theta}$.

$$\int (T(x) - \tau(\theta)) \frac{\partial \ln f}{\partial \theta} f dx = \tau'(\theta). \quad (79)$$

Полученный интеграл совпадает с ковариацией, то есть

$$\text{cov}\left(T(x), \frac{\partial \ln f}{\partial \theta}\right) = \tau'(\theta). \quad (80)$$

Из свойства коэффициента корреляции вытекает неравенство Коши – Буняковского:

$$\tau'(\theta) \leq \sqrt{\mathbf{D}(T(x)) \mathbf{D}\left(\frac{\partial \ln f}{\partial \theta}\right)}.$$

Преобразуем последнее выражение:

$$\mathbf{D} T(x) \geq \frac{(\tau'(\theta))^2}{\mathbf{D}\left(\frac{\partial \ln f}{\partial \theta}\right)}.$$

$$\begin{aligned} \mathbf{D}\left(\frac{\partial \ln f}{\partial \theta}\right) &= \mathbf{E}\left(\frac{\partial \ln f}{\partial \theta}\right)^2 - \mathbf{E}^2\left(\frac{\partial \ln f}{\partial \theta}\right) = \mathbf{E}\left(\frac{\partial \ln f}{\partial \theta}\right)^2 = I \\ \implies \mathbf{D} T(x) &\geq \frac{(\tau'(\theta))^2}{I}. \end{aligned}$$

Определение 8.1. Статистика $T(x) \in T_{\tau(\theta)}$ называется **эффективной оценкой** параметрической функции $\tau(\theta)$, если

$$\mathbf{D} T(x) = \frac{(\tau'(\theta))^2}{I}, \quad (81)$$

где

$$I = \mathbf{E}\left(\frac{\partial^2 \ln f(X)}{\partial \theta^2}\right).$$

Из доказательства неравенства Рао — Крамера вытекает важное следствие: равенство

$$\mathbf{D} T(x) = \frac{(\tau'(\theta))^2}{I} \quad (82)$$

имеет место тогда и только тогда, когда $T(x)$ — линейная функция вклада выборки, то есть

$$T(x) = a(\theta) \frac{\partial \ln f(X)}{\partial \theta} + \tau(\theta),$$

где $a(\theta)$ — некоторая функция от θ .

Это следствие можно переформулировать так:

Теорема 8.2. *Статистика $T(x) \in T_{\tau(\theta)}$ является эффективной оценкой параметрической функции $\tau(\theta)$ тогда и только тогда, когда $T(x)$ — линейная функция вклада выборки, то есть*

$$T(x) = a(\theta) \frac{\partial \ln f(X)}{\partial \theta} + \tau(\theta),$$

где $a(\theta)$ — некоторая функция от θ .

Мы уже рассматривали некоторые свойства оценок — несмещенность, состоятельность, оптимальность, эффективность. Одна из задач статистики — исследовать, обладает ли конкретная оценка этими свойствами. Подытожим, что нам известно о методах решения такой задачи.

Исследовать оценку на **несмещенность** можно напрямую, по определению.

Доказать, что оценка **состоятельна**, можно с помощью теорем 5.1—5.3. Эти теоремы дают достаточные условия состоятельности. Можно использовать и происхождение оценки: о.м.п. и о.м.м. состоятельны. Для доказательства того факта, что оценка не является состоятельной, надо показывать, что нарушается определение сходимости по вероятности.

Эффективность оценки мы умеем проверять по определению, вычисляя дисперсию оценки и информационное количество Фишера.

Про **оптимальность** нам известно меньше всего, а для проверки оптимальности единственный (пока) инструмент — утверждение: «Эффективная оценка является оптимальной. Обратное, вообще говоря, не верно.» Почему это так? Несмещенная оценка θ оптимальна, если она имеет наименьшую дисперсию среди всех несмещенных оценок θ , и эффективна, если она имеет минимальную возможную дисперсию, совпадающую с границей Рао — Крамера. Разница примерно как между наименьшим значением из

реально существующих и теоретическим минимумом. Теоретический минимум, соответствующий эффективной оценке, может и не достигаться в данной модели, оптимальной же оценке соответствует минимальное значение из реально существующих. Кроме этого, эффективные оценки рассматриваются только в регулярных моделях, а оптимальные существуют и в нерегулярных.

8.2. Другой подход к эффективным оценкам

Мы пойдем другим путем.
В.И. Ульянов

При исследовании эффективности, помимо прямого применения неравенства Рао — Крамера, можно использовать его следствие (теорема 8.2):

Статистика $T(x) \in T_{\tau(\theta)}$ является эффективной оценкой параметрической функции $\tau(\theta)$ тогда и только тогда, когда $T(x)$ — линейная функция вклада выборки, то есть

$$T(x) = a(\theta) \frac{\partial \ln f(X)}{\partial \theta} + \tau(\theta),^{23} \quad (83)$$

где $a(\theta)$ — некоторая функция от θ .

Посмотрим внимательно на (83). Оказывается, функциональная часть эффективной оценки, зависящая от X , равна

$$\frac{\partial \ln f(X)}{\partial \theta}.$$

Важно помнить, что X тут — многомерная величина.

Попробуем применить этот подход.

Пример 8.1. Пусть ξ распределена по нормальному закону $N(\theta, \sigma)$. Функция плотности

$$f(X) = f(x_1, \dots, x_n) = f(x_1) \cdot \dots \cdot f(x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}}.$$

◀ Найдем логарифм плотности

$$\ln f(X) = -n(\ln \sigma + 0,5 \ln 2\pi) - \sum_{i=1}^n \frac{(x_i - \theta)^2}{2\sigma^2}.$$

²³ $f(X)$ можно записывать как L (функция правдоподобия).

Найдем частную производную по θ :

$$\frac{\partial \ln f(X)}{\partial \theta} = \sum_{i=1}^n \frac{(x_i - \theta)}{\sigma^2}.$$

Функциональная часть представляет из себя

$$\sum_{i=1}^n x_i.$$

Таким образом, эффективная оценка любой функции параметра $\theta = a$ должна зависеть от $\sum_{i=1}^n x_i$ (или, что то же самое, от \bar{X} .) С учетом того, что эффективная оценка должна быть несмещенной, это означает, что эффективных оценок существует немного. ►

Теорема 8.3. В модели $\langle F_\theta \rangle$ эффективные оценки могут существовать только для одного класса функций, то есть может существовать $\tau(\theta)$ такая, что $\forall u(\theta) \in \{a\tau(\theta) + b \mid a, b - \text{const}\}$ существует эффективная оценка, и если $T(x)$ — эффективная оценка $\tau(\theta)$, то $aT(x) + b$ — эффективная оценка функции $u(\theta) = a\tau(\theta) + b$. Для всех других функций параметра θ эффективных оценок не существует.

Например, если θ имеет эффективную оценку, то у θ^2 уже не может существовать эффективная оценка. Теорему несложно доказать напрямую, к тому же она является следствием теоремы об эффективных оценках в экспоненциальном семействе (8.4), которую мы сейчас рассмотрим.

8.3. Экспоненциальное семейство

*Нормальные герои
всегда идут в обход.
Песня из к/ф «Айболит-66»*

Определение 8.2. Говорят, что распределение с плотностью $f(x)$ принадлежит экспоненциальному семейству (коротко будем записывать это $f \in E$), если $f(x)$ представима в виде

$$f(x, \theta) = e^{A(x) \cdot B(\theta) + C(x) + D(\theta)}. \quad (84)$$

Теорема 8.4. Для того чтобы в модели существовала эффективная оценка, необходимо и достаточно, чтобы модель принадлежала экспоненциальному семейству. При этом эффективной оценкой является статистика

$$T(x) = \frac{1}{n} \sum_{i=1}^n A(x_i),$$

и она оценивает параметрическую функцию

$$\tau(\theta) = -\frac{D'(\theta)}{B'(\theta)}.$$

Доказательство.

Достаточность. Пусть $f(x, \theta) \in E$. Тогда

$$\begin{aligned} \ln L &= \sum_{i=1}^n \ln f(x_i, \theta) = B(\theta) \sum_{i=1}^n A(x_i) + \sum_{i=1}^n C(x_i) + nD(\theta). \\ \frac{\partial \ln L}{\partial \theta} &= B'(\theta) \sum_{i=1}^n A(x_i) + nD'(\theta). \\ \frac{\sum_{i=1}^n A(x_i)}{n} &= \frac{1}{nB'(\theta)} \frac{\partial \ln L}{\partial \theta} - \frac{D'(\theta)}{B'(\theta)}. \end{aligned} \quad (85)$$

По (83) несмещенная оценка $T(x)$ параметрической функции $\tau(\theta)$ является ее эффективной оценкой тогда и только тогда, когда

$$T(x) = a(\theta) \frac{\partial \ln L}{\partial \theta} + \tau(\theta), \quad (86)$$

где $a(\theta)$ — некоторая функция от θ . Сравнивая (85) и (86), видим, что

$$T(x) = \frac{\sum_{i=1}^n A(x_i)}{n}, \quad \tau(\theta) = -\frac{D'(\theta)}{B'(\theta)}.$$

Следовательно, статистика

$$T(x) = \frac{1}{n} \sum_{i=1}^n A(x_i)$$

является эффективной оценкой параметрической функции

$$\tau(\theta) = -\frac{D'(\theta)}{B'(\theta)},$$

и других эффективных оценок в этой модели нет (с точностью до линейного преобразования).

Необходимость. Пусть в модели $\langle F_\theta \rangle$ существует эффективная оценка $T(x)$. Покажем, что модель принадлежит экспоненциальному семейству. Поскольку $T(x)$ эффективна,

$$T(x) = a(\theta) \frac{\partial \ln L(X)}{\partial \theta} + \tau(\theta),$$

отсюда

$$\frac{T(x) - \tau(\theta)}{a(\theta)} = \frac{\partial \ln L(X)}{\partial \theta}.$$

Выразим $L(X)$:

$$\begin{aligned} \ln L(X) &= \int \frac{T(x)}{a(\theta)} d\theta - \int \frac{\tau(\theta)}{a(\theta)} d\theta, \\ L(X) &= \exp \left\{ T(x) \int \frac{1}{a(\theta)} d\theta - \int \frac{\tau(\theta)}{a(\theta)} d\theta \right\}. \end{aligned} \quad (87)$$

Сравним (87) с определением экспоненциального семейства:

$$L(x) = \exp\{A(x) \cdot B(\theta) + C(x) + D(\theta)\}.$$

Видим, что

$$A(x) = T(x), \quad B(\theta) = \int \frac{d\theta}{a(\theta)}, \quad C(x) = 0, \quad D(\theta) = - \int \frac{\tau(\theta)}{a(\theta)} d\theta. \quad \square$$

Пример 8.2. Докажем, что $N(a, \sigma)$ принадлежит экспоненциальному семейству по каждому из параметров, взятых отдельно, и найдем эффективные оценки.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp^{-\frac{(x-a)^2}{2\sigma^2}} = \exp^{\ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2\sigma^2} + \frac{x \cdot a}{\sigma^2} - \frac{a^2}{2\sigma^2}}.$$

Пусть параметром будет a , тогда

$$A(x) = x, \quad B(a) = \frac{a}{\sigma^2}, \quad C(x) = \ln \frac{1}{\sigma\sqrt{2\pi}} - \frac{x^2}{2\sigma^2}, \quad D(a) = -\frac{a^2}{2\sigma^2}.$$

Следовательно, распределение $N(\theta, \sigma)$ принадлежит экспоненциальному семейству. Найдем эффективную оценку параметрической функции от параметра $\theta = a$:

$$\tau(a) = -\frac{D'(a)}{B'(a)} = \frac{a}{\sigma} : \frac{1}{\sigma} = a,$$

$$T(x) = \frac{1}{n} \sum_{i=1}^n A(x_i) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Получили, что \bar{x} является эффективной оценкой a .

Пусть параметром будет σ , тогда

$$A(x) = (x - a)^2, \quad B(\sigma) = -\frac{1}{2\sigma^2}, \quad C(x) = 0, \quad D(\sigma) = -\ln \sigma - \frac{1}{2} \ln 2\pi.$$

Следовательно, распределение $N(a, \theta)$ также принадлежит экспоненциальному семейству. Найдем эффективную оценку параметрической функции от параметра $\theta = \sigma$:

$$\tau(\sigma) = -\frac{D'(a)}{B'(a)} = \frac{1}{\sigma} : \frac{1}{\sigma^3} = \sigma^2,$$

$$T(x) = \frac{1}{n} \sum_{i=1}^n A(x_i) = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2.$$

По теореме 8.4 оценка $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2$ эффективна. ►

Определение 8.3. В случае многомерного параметра $\theta = (\theta_1, \dots, \theta_k)$ и непрерывной параметрической модели говорят, что распределение с плотностью $f(x)$ принадлежит экспоненциальному семейству, если $f(x)$ представима в виде

$$f(x) = \exp\left\{\sum_{i=1}^k A_i(x) \cdot B_i(\theta) + C(x) + D(\theta)\right\}.$$

Чтобы сделать представление по возможности более однозначным, мы будем предполагать, что функции линейно не зависимы на 0.

Примером могут служить семейства нормальных распределений $N(\theta_1, \theta_2)$, гамма-распределений $\Gamma_{\theta_1, \theta_2}$.

8.4. Контрольные вопросы

1. Укажите три способа поиска эффективной оценки.
2. Дайте определение вклада выборки.
3. Дайте определение экспоненциального семейства.
4. Принадлежат ли экспоненциальному семейству распределения $N(a, \theta)$, $N(\theta, \sigma)$, $N(\theta, \theta)$?
5. Принадлежит ли экспоненциальному семейству биномиальное распределение $B(N, \theta)$?
6. Принадлежит ли экспоненциальному семейству распределение Пуассона?
7. Запишите закон распределения Пуассона в экспоненциальном виде.

8. Существует ли эффективная оценка в биномиальном распределении?
9. Укажите необходимые и достаточные условия существования эффективной оценки.
10. Могут ли в одной модели статистики \bar{X} и $\sqrt{\bar{X}}$ являться эффективными оценками (возможно, различных параметрических функций)?
11. Могут ли в одной модели статистики \bar{X} и S^2 являться эффективными оценками (возможно, различных параметрических функций)?
12. Могут ли в одной модели статистики \bar{X} и $2\sum X_i$ являться эффективными оценками (возможно, различных параметрических функций)?
13. Запишите плотность показательного распределения в экспоненциальном виде.
14. Какая параметрическая функция имеет в показательной модели эффективную оценку?
15. Какая статистика является в показательной модели эффективной оценкой?

Лекция 9. Достаточные статистики

- Достаточная статистика – это статистика, которой достаточно для оценивания.
 - Смысл верный, теперь дайте определение.
- С устного экзамена

План лекции: определения и примеры, критерий достаточности, свойства достаточных статистик, свойства оценок максимального правдоподобия.

9.1. Определение и примеры

Напомним, что мы по-прежнему рассматриваем параметрическую модель $\langle F_\theta \rangle$.

Определение 9.1. Статистика $T = T(X)$ называется **достаточной** для параметра θ , если условное распределение (плотность или вероятность) случайной величины $X = (X_1, \dots, X_n)$ (выборки) при условии $T(X) = t$ не зависит от параметра θ .

Это означает, что в дискретном случае условная вероятность

$$P(X = x / T(X) = t),$$

а в непрерывном случае условная плотность $f_{X/T(X)=t}(x)$ не зависит от параметра θ , вся информация о θ , имеющаяся в выборке, содержится в $T(X)$. Рассмотрим подробнее дискретный случай. Поскольку

$$\begin{aligned} P(A/B) &= \frac{P(AB)}{P(B)}, \\ P(X = x / T(X) = t) &= \frac{P(X = x, T(X) = t)}{P(T(X) = t)} = \\ &= \begin{cases} \frac{P(X=x)}{P(T(X)=t)}, & x : T(x) = t \\ 0, & x : T(x) \neq t, \end{cases} \end{aligned}$$

так как для $x : T(x) = t$ событие $\{X = x, T(X) = t\}$ влечет событие $\{X = x\}$. Таким образом, в дискретной модели статистика $T(X)$ достаточна, если

$$\frac{P(X = x)}{P(T(X) = t)} \tag{88}$$

не зависит от параметра для x таких, что $T(x) = t$.

Пример 9.1. Исследуем на достаточность статистику

$$T = \frac{1}{n} \sum_{i=1}^n X_i \text{ в } P_\lambda.$$

◀ Надо показать, что $\frac{P(X=x)}{P(T(X)=t)}$ не зависит от параметра λ .

$$\begin{aligned} P(X=x) &= P(X_1=x_1, \dots, X_n=x_n) = \prod_{i=1}^n P(X_i=x_i) = \\ &= \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-\lambda n} \lambda^{\sum x_i}}{\prod (x_i!)}. \end{aligned}$$

Случайная величина $\sum_{i=1}^n X_i$ имеет распределение $P_{\lambda n}$, значит,

$$P(T=t) = P\left(\sum_{i=1}^n X_i = nt\right) = \frac{e^{-n\lambda} (n\lambda)^{tn}}{(tn)!}.$$

$$P(X/T=t) = \frac{P(X=x)}{P(T=t)} = \frac{e^{-\lambda n} \lambda^{\sum x_i} (tn)!}{\prod (x_i!) e^{-n\lambda} (n\lambda)^{tn}} = \frac{(\sum x_i)!}{\prod (x_i!) n^{\sum x_i}}.$$

Полученное выражение не зависит от параметра λ , следовательно, статистика $T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{x}$ является достаточной. ▶

Для непрерывных величин распределение задается плотностью. В этом случае также достаточно показать, что

$$\frac{f_X(x)}{f_T(t)} \tag{89}$$

не зависит от параметра θ .

Замечание 9.1. $\frac{f_{X,T}(x,t)}{f_T(t)}$ не равно $\frac{f_X(x)}{f_T(t)}$, но отличается множителем, не зависящим от θ (якобианом), следовательно, статистика $T(x)$ достаточна тогда и только тогда, когда $\frac{f_X(x)}{f_T(t)}$ не зависит от θ .

Пример 9.2. Пусть (X_1, \dots, X_n) — выборка из показательного распределения E_α с параметром α . Исследуем на достаточность статистику $T(x) = \sum_{i=1}^n X_i$.

◀ Надо показать, что $\frac{f_X(x)}{f_T(t)}$ не зависит от параметра α .

$$f_X(x) = f_X(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \alpha e^{-\alpha x_i} = \alpha^n e^{-\alpha \sum x_i}.$$

Чтобы найти $f_T(t)$, найдем распределение суммы n показательно распределенных случайных величин.

Характеристическая функция показательного распределения E_α

$$\varphi_{X_i}(t) = \frac{\alpha}{\alpha - it}.$$

Тогда характеристическая функция статистики $T(x) = \sum_{i=1}^n X_i$ равна произведению характеристических функций слагаемых,

$$\varphi_T(t) = \left(\frac{\alpha}{\alpha - it} \right)^n,$$

но это характеристическая функция гамма-распределения $\Gamma_{\alpha, n}$. Плотность гамма-распределения $\Gamma_{\alpha, \beta}$ равна

$$f_\xi(x) = \frac{\alpha^\beta}{\Gamma(\beta)} e^{-\alpha x} x^{\beta-1}, \quad x \geq 0,$$

где $\Gamma(\beta) = \int_0^{+\infty} x^{\beta-1} e^{-x} dx$. Соответственно плотность $T(t)$ равна

$$f_T(t) = \frac{\alpha^n}{\Gamma(n)} e^{-\alpha t} t^{n-1}, \quad t \geq 0,$$

или, поскольку $\Gamma(n) = (n-1)! \forall n \in \mathbb{N}$,

$$f_T(t) = \frac{\alpha^n}{(n-1)!} e^{-\alpha t} t^{n-1}, \quad t \geq 0.$$

Найдем отношение плотностей.

$$\frac{f_X(x)}{f_T(t)} = \frac{\alpha^n e^{-\alpha \sum x_i}}{\frac{\alpha^n}{(n-1)!} e^{-\alpha t} t^{n-1}} = \frac{(n-1)!}{t^{n-1}},$$

что не зависит от α . ►

9.2. Критерий достаточности

*С холодным вниманьем посмотришь вокруг —
какая-то скука, читатель и друг!
Когда же посмотришь с вниманьем горячим,
увидится все совершенно иначе.*
Т. Кибиров

Существует удобный критерий достаточности статистики — теорема Неймана — Фишера о факторизации.

Теорема 9.1 (теорема факторизации). В модели $\langle F_\theta \rangle$ статистика $T(X)$ является достаточной для параметра θ тогда и только тогда, когда

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

Замечание 9.2. Представление $L(x, \theta)$ в таком виде (множитель g может зависеть от θ , а от x зависит лишь через $T(x)$, а множитель h от параметра θ не зависит) называется факторизацией распределения. Факторизация не единственна. При $h \equiv 1$ говорят о тривиальной факторизации.

Докажем теорему для дискретной модели.

Доказательство.

1. Пусть $T(X)$ — достаточная статистика. Возьмем t такое, что $T(x) = t$.

$$L(x, \theta) = P(X = x),$$

и, поскольку

$$\{X = x\} \subseteq \{T(X) = t\},$$

$$L(x, \theta) = P(X = x) = P(X = x, T(X) = t).$$

По теореме умножения

$$P(X = x, T(X) = t) = P(T(X) = t) \cdot P(X = x/T(X) = t).$$

По определению достаточной статистики, $P(X = x/T(X) = t)$ не зависит от параметра, то есть

$$P(X = x/T(X) = t) = h_1(x, t) = h_1(x, T(x)) = h(x).$$

Функция $P(T(X) = t)$ зависит от параметра, поскольку вычисляется по закону распределения. Но от x она зависит только через $T(x)$, то есть $P(T(X) = t) = g(T(x), \theta)$. Тогда

$$L(x, \theta) = P(T(X) = t)P(X = x/T(X) = t) = g(T(x), \theta)h(x).$$

Получена факторизация распределения.

2. Пусть

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

$$P(X = x/T(X) = t) = \frac{P(X = x, T(X) = t)}{P(T(X) = t)} = \frac{P(X = x)}{P(T(X) = t)} =$$

$$\begin{aligned}
&= \frac{L(x, \theta)}{\sum_{x:T(x)=t} L(x, \theta)} = \\
&= \frac{g(T(x), \theta)h(x)}{\sum_{x:T(x)=t} g(T(x), \theta)h(x)} = \frac{g(t, \theta)h(x)}{g(t, \theta) \sum_{x:T(x)=t} h(x)} = \frac{h(x)}{\sum_{x:T(x)=t} h(x)},
\end{aligned}$$

что не зависит от θ . Следовательно, $T(x)$ — достаточная статистика. \square

Пример 9.3. Пусть (X_1, \dots, X_n) — выборка из нормального распределения $N(a, \sigma)$. Найдем достаточную статистику для двумерного параметра $\theta = (a, \sigma)$.

◀ Функция правдоподобия имеет следующий вид:

$$L = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum (x_i - a)^2}{2\sigma^2}} = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum x_i^2 - 2a \sum x_i + na^2}{2\sigma^2}} = g(T(x), \theta),$$

где

$$T(x) = \left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i \right).$$

►

9.3. Свойства достаточных статистик

Да как же это может быть, Сократ, чтобы ни одна из двух вещей не имела какого-то свойства, а затем чтобы это самое свойство, которого ни одна из них не имеет, оказалось в обеих?

Платон, «Диалог с Гиппием»

Свойство 1. Всякая эффективная оценка является достаточной статистикой.

Доказательство. Пусть $T(x)$ — эффективная оценка параметрической функции $\tau(\theta)$. Тогда

$$T(x) = a(\theta) \frac{\partial \ln L(X)}{\partial \theta} + \tau(\theta),$$

где $a(\theta)$ — некоторая функция от θ .

$$\frac{\partial \ln L(X)}{\partial \theta} = \frac{T(x) - \tau(\theta)}{a(\theta)}.$$

$$\ln L(x) = \int \frac{T(x) - \tau(\theta)}{a(\theta)} d\theta;$$

$$L(x) = e^{\int \frac{T(x) - \tau(\theta)}{a(\theta)} d\theta}$$

и при фиксированном $T(x) = t$

$$L(x) = g(t, \theta).$$

Получили тривиальную факторизацию. \square

Пример 9.4. В $N(a, \sigma)$ \bar{X} является эффективной оценкой a .

◀ Следовательно, \bar{X} является достаточной статистикой для a . ▶

Свойство 2. Любая взаимно однозначная функция от достаточной статистики T является достаточной статистикой.

Доказательство. Пусть $H = \varphi(T)$; $T = \varphi^{-1}(H)$.

$$L(x, \theta) = g(T(x), \theta)h(x) = g(\varphi^{-1}(H), \theta)h(x) = g_1(H, \theta)h(x).$$

По теореме факторизации $H = \varphi(T)$ является достаточной статистикой. \square

Пример 9.5.

◀ В примере 9.3 наряду с

$$T(x) = (T_1, T_2) = \left(\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i \right)$$

достаточной статистикой будет являться и статистика

$$H(x) = (S^2, \bar{x}),$$

поскольку $H(x)$ является взаимно однозначной функцией от $T(x)$. ▶

Определение 9.2. Достаточная статистика, являющаяся функцией любых других достаточных статистик, называется **минимальной**.

Это достаточная статистика наименьшей размерности, представляющая исходные данные в наиболее сжатом виде. Минимальная достаточная статистика является самой экономной. Вернемся к оценкам максимального правдоподобия и выясним еще некоторые их свойства.

9.4. Свойства оценок максимального правдоподобия

*Настоящие свойства обнаруживаются лишь тогда,
когда наступает время проявить их.*
Л. Фейербах

Свойство 1. Связь с эффективными оценками. Если для скалярного параметра θ существует эффективная оценка $T(x)$, то $T(x)$ совпадает с о.м.п. $\hat{\theta}$.

Доказательство. Пусть $T(x)$ — эффективная оценка параметра θ . Это справедливо тогда и только тогда, когда

$$T(x) = a(\theta) \frac{\partial \ln L(X)}{\partial \theta} + \theta.$$

Решим уравнение правдоподобия

$$\begin{aligned} \frac{\partial \ln L}{\partial \theta} &= 0. \\ \frac{\partial \ln L}{\partial \theta} &= \frac{T - \theta}{a(\theta)} = 0 \Rightarrow \hat{\theta} = T. \end{aligned}$$

Покажем, что это точка максимума. Поскольку для эффективной оценки

$$\rho_{T, \frac{\partial \ln L}{\partial \theta}} = +1,$$

коэффициент линейной зависимости $a(\theta) > 0$, и при переходе через критическую точку $\hat{\theta} = T$ производная меняет знак с плюса на минус, что является достаточным условием максимума. \square

Свойство 2. Связь с достаточными статистиками. Если $T(x)$ — достаточная статистика, а о.м.п. $\hat{\theta}$ существует и единственна, то $\hat{\theta}$ является функцией от $T(x)$.

Доказательство.

По факторизационному критерию достаточности статистика $T(x)$ является достаточной для параметра θ тогда и только тогда, когда

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

Запишем уравнение правдоподобия:

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln g(T, \theta)}{\partial \theta} = 0.$$

Очевидно, решение этого уравнения $\hat{\theta}$ должно быть функцией T :

$$\hat{\theta} = \hat{\theta}(T). \quad \square$$

Следствие. Если о.м.п. $\hat{\theta}$ существует и единственна, то $\hat{\theta}$ является функцией минимальной достаточной статистики.

Утверждение вытекает из того, что по свойству 2 о.м.п. $\hat{\theta}$ является функцией любой достаточной статистики).

9.5. Контрольные вопросы

1. Дайте определение достаточной статистики.
2. Укажите метод доказательства по определению достаточности статистики в непрерывной модели.
3. Укажите метод доказательства по определению достаточности статистики в дискретной модели.
4. Сформулируйте теорему факторизации.
5. Верно ли, что достаточная статистика единственна?
6. Верно ли, что всякая достаточная статистика является эффективной оценкой?
7. Верно ли, что всякая эффективная оценка является достаточной статистикой?
8. Существует ли достаточная статистика в нерегулярной модели?
9. В любой ли модели существует достаточная статистика?
10. Приведите пример достаточной статистики для параметра показательного распределения.
11. Приведите 2 примера достаточных статистик для параметра p биномиального распределения.
12. Приведите 3 примера достаточных статистик для параметра закона Пуассона.
13. Дайте определение минимальной достаточной статистики.

Лекция 10. Оптимальные оценки

Несмещенная оценка $\hat{\theta}$ параметра θ называется оптимальной оценкой, если $\mathbf{D}\hat{\theta} \leq \mathbf{D}\tilde{\theta}, \forall \theta \in \Theta, \tilde{\theta} \in T_{\theta}$.
Определение

План лекции: свойства оптимальных оценок, достаточные статистики и оптимальные оценки, полные статистики.

10.1. Свойства оптимальных оценок

И эта ее единственность делает ее более превосходной.
Леонардо да Винчи

Теорема 10.1 (теорема единственности). Если оптимальная оценка существует, то она единственна.

Доказательство. Пусть у функции $\tau(\theta)$ существуют две оптимальные оценки $T_1(x)$ и $T_2(x)$. Оптимальные оценки по определению являются несмещенными, поэтому $\mathbf{E}(T_1(x)) = \mathbf{E}(T_2(x)) = \tau(\theta)$. Рассмотрим новую оценку $T_3 = \frac{T_1 + T_2}{2}$:

$$\mathbf{E}(T_3) = \mathbf{E}\left(\frac{T_1 + T_2}{2}\right) = \tau(\theta).$$

$$\mathbf{D}(T_3(x)) = \mathbf{D}\left(\frac{T_1 + T_2}{2}\right) = \frac{\mathbf{D}T_1 + \mathbf{D}T_2 + 2\text{cov}(T_1, T_2)}{4}.$$

Поскольку $T_1(x)$ и $T_2(x)$ — оптимальные оценки, их дисперсии минимальны и равны между собой; пусть $\mathbf{D}T_1 = \mathbf{D}T_2 = d$. По свойству ковариации

$$\text{cov}(T_1, T_2) \leq \sqrt{\mathbf{D}T_1 \cdot \mathbf{D}T_2} = d$$

(неравенство Коши — Буняковского). Тогда

$$\mathbf{D}(T_3(x)) = \frac{d + d + 2\text{cov}(T_1, T_2)}{4} \leq d.$$

$T_3(x)$ — несмещенная оценка, и ее дисперсия не больше минимальной, следовательно, она тоже оптимальна. Тогда ее дисперсия в точности равна d , и в неравенстве достигается равенство. Но это возможно только, если $\text{cov}(T_1, T_2) = \sqrt{\mathbf{D}T_1 \cdot \mathbf{D}T_2}$, что означает линейную зависимость T_1, T_2 .

Пусть $T_1 = aT_2 + b$. $\mathbf{D}T_1 = a^2 \mathbf{D}T_2$, но $\mathbf{D}T_1 = \mathbf{D}T_2$, следовательно, $a^2 = 1$.

$\text{cov}(T_1, T_2) = d > 0$ (положительная связь), значит, $a > 0$ и $a = 1$. $\mathbf{E}T_1 = \mathbf{E}T_2 + b$, следовательно, $b = 0$. Окончательно получаем: $T_1 = aT_2 + b = T_2$ (если есть две оптимальные оценки, то они совпадают). \square

Теорема 10.2. Пусть $T(x)$ — оптимальная оценка некоторой параметрической функции $\tau(\theta)$, а $H(x)$ такая статистика, что $\mathbf{E}(H(x)) = 0$, $\forall \theta \in \Theta$. Тогда $\text{cov}(T(x), H(x)) = 0 \forall \theta \in \Theta$.

Доказательство. Рассмотрим $T'(x) = T(x) + \lambda H(x)$, $\lambda \in R$. $\mathbf{E} T'(x) = \mathbf{E} T(x) + \lambda \mathbf{E} H(x) = \mathbf{E} T(x) = \tau(\theta)$, то есть $T' \in T_\tau$.

$$\mathbf{D} T'(x) = \mathbf{D} T(x) + \lambda^2 \mathbf{D} H(x) + 2\lambda \text{cov}(T(x), H(x)) \geq \mathbf{D} T(x)$$

(поскольку $T'(x)$ несмещенная оценка, а $T(x)$ оптимальная).

$$\lambda^2 \mathbf{D} H(x) + 2\lambda \text{cov}(T(x), H(x)) \geq 0.$$

Это квадратное неравенство относительно λ , справедливое при всех значениях λ . Коэффициент при λ^2 положительный, значит, дискриминант не больше 0, то есть

$$\text{cov}^2(T(x), H(x)) \leq 0.$$

Следовательно,

$$\text{cov}(T(x), H(x)) = 0. \square$$

Теорема 10.3. Если $T_1(x)$ — оптимальная оценка параметрической функции τ_1 , а $T_2(x)$ — оптимальная оценка параметрической функции τ_2 , то $T = a_1 T_1(x) + a_2 T_2(x)$ — оптимальная оценка параметрической функции $\tau = a_1 \tau_1 + a_2 \tau_2$.

Доказательство. Очевидно, $\mathbf{E} T = a_1 \mathbf{E} T_1(x) + a_2 \mathbf{E} T_2(x) = a_1 \tau_1 + a_2 \tau_2 = \tau$. Пусть S — произвольная несмещенная оценка функции τ . Рассмотрим разность $S - T$:

$$\mathbf{E}(S - T) = \mathbf{E} S - \mathbf{E} T = \tau - \tau = 0.$$

Теперь рассмотрим ковариацию:

$$\text{cov}(T, S - T) = \text{cov}(a_1 T_1 + a_2 T_2, S - T) = a_1 \text{cov}(T_1, S - T) + a_2 \text{cov}(T_2, S - T).$$

Тогда по теореме 10.2

$$\text{cov}(T, S - T) = 0$$

(нулю равно каждое слагаемое в силу оптимальности T_1, T_2). С другой стороны,

$$\text{cov}(T, S - T) = \text{cov}(T, S) - \mathbf{D} T,$$

значит, $\mathbf{D} T = \text{cov}(T, S)$. Но

$$\text{cov}(T, S) \leq \sqrt{\mathbf{D} T \cdot \mathbf{D} S}.$$

Следовательно, $\mathbf{D} T \leq \sqrt{\mathbf{D} T \cdot \mathbf{D} S}$, или $\mathbf{D} S \geq \mathbf{D} T$. Это верно для любой несмещенной оценки S , значит, T — оптимальная оценка. \square

10.2. Достаточные статистики и оптимальные оценки

– Это верный подход, – Благозвон произнес,
Торопливо вмешавшись опять,
– Это правильный путь, если хочешь всерьез
Настоящего Снарка поймать!
Льюис Кэрролл, «Охота на Снарка»

Оказывается, что оптимальная оценка, если она существует, является функцией от достаточной статистики. Точнее говоря, для любой несмещенной оценки, не являющейся функцией от достаточной статистики, можно указать несмещенную оценку, зависящую от достаточной статистики, дисперсия которой меньше, чем дисперсия исходной оценки. Это доказывает следующая теорема.

Теорема 10.4 (теорема Рао – Блекуэлла – Колмогорова). Пусть $T(X)$ – достаточная статистика, $d(X)$ – несмещенная оценка θ , $\varphi(T) = \mathbf{E}(d(X)/T(X) = t)$. Тогда

$$\mathbf{D} d(X) \geq \mathbf{D} \varphi(T).$$

Доказательство.

Схема доказательства.

(1) Покажем, что $\varphi(T)$ не зависит от θ и, значит, может служить оценкой θ .

(2) Покажем, что $\varphi(T)$ – несмещенная оценка θ .

(3) Докажем, что $\mathbf{D} d(X) \geq \mathbf{D} \varphi(T)$.

Для простоты записи обозначим $f_{X/T(X)=t} = f_T$.

(1) $\varphi(T)$ не зависит от θ , так как

$$\varphi(T) = \mathbf{E}(d(X)/T(X) = t) = \int d(x) f_T(x) dx.$$

($f_T(x)$ не зависит от θ , так как $T(x)$ – достаточная статистика).

(2) $\varphi(T) \in T_\theta$, так как

$$\mathbf{E}[\varphi(T)] = \mathbf{E}_T[\mathbf{E}_X(d(X)/T(X) = t)] = \mathbf{E}(d(X)) = \theta.$$

Использована формула из курса теории вероятностей

$$\mathbf{E}_\eta[\mathbf{E}_\xi(\xi/\eta)] = \mathbf{E} \xi.$$

(3) Вычислим дисперсию $d(X)$:

$$\mathbf{D} d(X) = \mathbf{E}(d(X) - \theta)^2 = \mathbf{E}(d(X) - \varphi(T) + \varphi(T) - \theta)^2 =$$

$$= \mathbf{E}(d(X) - \varphi(T))^2 + \mathbf{E}(\varphi(T) - \theta)^2 + 2 \mathbf{E}[(d(X) - \varphi(T))(\varphi(T) - \theta)].$$

Рассмотрим последний член:

$$\begin{aligned} \mathbf{E}[(d(X) - \varphi(T))(\varphi(T) - \theta)] &= \mathbf{E}[(d(X) - \varphi(T))\varphi(T)] - \theta \mathbf{E}[(d(X) - \varphi(T))] = \\ &= \mathbf{E}[(d(X) - \varphi(T))\varphi(T)], \end{aligned}$$

так как $\mathbf{E} d(X) = \mathbf{E} \varphi(T)$, и последний член обращается в 0.

$$\mathbf{E}[(d(X) - \varphi(T))\varphi(T)] = \mathbf{E}_T(\mathbf{E}_X[(d(X) - \varphi(T))\varphi(T)/T(X) = t]).$$

$\varphi(T)$ можно вынести за знак математического ожидания по X :

$$\begin{aligned} \mathbf{E}_T(\mathbf{E}_X[(d(X) - \varphi(T))\varphi(T)/T(X) = t]) &= \\ = \mathbf{E}_T(\varphi(T) \mathbf{E}_X[(d(X) - \varphi(T))/T(X) = t]). \end{aligned}$$

Рассмотрим отдельно математическое ожидание по X :

$$\mathbf{E}_X[(d(X) - \varphi(T))/T(X) = t] = \mathbf{E}_X[d(X)/T(X) = t] - \mathbf{E}_X[\varphi(T)/T(X) = t]$$

По условию

$$\mathbf{E}_X[d(X)/T(X) = t] = \varphi(T).$$

Но и

$$\mathbf{E}_X[\varphi(T)/T(X) = t] = \varphi(t) = \varphi(T),$$

поэтому

$$\mathbf{E}_T(\mathbf{E}_X[(d(X) - \varphi(T))/T(X) = t]) = \mathbf{E}_T(\varphi(T) - \varphi(T)) = \mathbf{E}_T(0) = 0,$$

$$\mathbf{D} d(X) = \mathbf{E}(d(X) - \varphi(T))^2 + \mathbf{E}(\varphi(T) - \theta)^2.$$

$\mathbf{E}(\varphi(T) - \theta)^2$, очевидно, равно $\mathbf{D} \varphi(T)$, а $\mathbf{E}(d(X) - \varphi(T))^2 \geq 0$, поэтому $\mathbf{D} d(X) \geq \mathbf{D} \varphi(T)$. \square

Выделим очевидное следствие этой теоремы.

Теорема 10.5. *Оптимальная оценка, если она существует, necessarily является функцией от достаточной статистики.*

Доказательство. Если это не так, можно получить оценку, у которой дисперсия не больше и которая является функцией от достаточной статистики. Она тоже будет оптимальной, но оптимальная единственна. \square

Следовательно, оптимальную оценку надо искать среди функций от минимальной статистики.

Оказывается, свойство минимальности статистики тесно связано со свойством **полноты**.

10.3. Полные статистики

Да поразит тя пуще грома
Ужасна, сильна аксиома!²⁴

Определение 10.1. Статистика T называется **полной**, если для всякой (ограниченной) функции $\varphi(T)$ из того, что $\mathbf{E} \varphi(T) = 0$ следует, что $\varphi(T) \equiv 0$.

Из определения²⁵ следует теорема единственности.

Теорема 10.6. Если $T(x)$ полна, то для любой функции $\tau(\theta)$ существует единственная несмещенная оценка, зависящая от $T(x)$.

Доказательство. Пусть для некоторой $\tau(\theta)$ существуют две несмещенные оценки, зависящие от $T(x)$: φ_1 и φ_2 .

$$\mathbf{E} \varphi_1 = \mathbf{E} \varphi_2 = \tau(\theta).$$

Тогда

$$\mathbf{E}(\varphi_1 - \varphi_2) = 0 \Rightarrow \varphi_1 - \varphi_2 \equiv 0,$$

(по определению полной статистики), то есть $\varphi_1 \equiv \varphi_2$. \square

Следующая теорема указывает на связь между понятиями полноты и минимальности (доказательство можно посмотреть в (1)).

Теорема 10.7. Всякая полная достаточная статистика S является минимальной достаточной статистикой.

Это означает, что при поиске оптимальных оценок может быть плодотворным использование полных статистик. Действительно, справедлива теорема:

Теорема 10.8. Если существует полная достаточная статистика, то всякая функция от нее является оптимальной оценкой своего математического ожидания.

Доказательство. По теореме Рао — Блекуэлла — Колмогорова оптимальная оценка, если она существует, является функцией от достаточной статистики. Пусть T — полная достаточная статистика, $\varphi(T)$ — некоторая функция от нее; обозначим ее математическое ожидание $\tau(\theta) = \mathbf{E}[\varphi(T)]$. Тогда

²⁴ Из пародии на Ломоносова двухсотлетней давности.

²⁵ Это и есть то определение, которому эпиграф предлагает поразиться.

$\varphi(T) \in T_\tau$ (T_τ — множество несмещенных оценок $\tau(\theta)$). Оптимальная оценка по определению принадлежит T_τ . Обе оценки, $\varphi(T)$ и оптимальная, являются функциями T . Но по теореме 10.6 существует единственная несмещенная оценка, зависящая от T . Следовательно, они совпадают и $\varphi(T)$ является оптимальной оценкой своего математического ожидания $\tau(\theta)$. \square

Итак, если оценивается заданная параметрическая функция $\tau(\theta)$, то оптимальная несмещенная оценка — такая функция $H(T(x))$ от полной достаточной статистики $T(x)$, которая удовлетворяет уравнению несмещенности $\mathbf{E} H(T(x)) = \tau(\theta)$. Это уравнение либо имеет единственное решение, либо решений нет. В последнем случае класс T_τ несмещенных оценок $\tau(\theta)$ пуст. \square

10.4. Контрольные вопросы

1. Дайте определение оптимальной оценки.
2. Верно ли, что оптимальная оценка единственна?
3. Чему равна ковариация оптимальной оценки и любой статистики, имеющей нулевое математическое ожидание?
4. Сформулируйте линейное свойство оптимальных оценок.
5. Сформулируйте теорему Рао — Блекуэлла — Колмогорова.
6. Даны достаточные статистики T_1 и T_2 . Может ли существовать оптимальная оценка, которая является функцией от T_1 , но не является функцией от T_2 ?
7. Дайте определение полной статистики.
8. Укажите связь между полной и минимальной достаточными статистиками.
9. Для какой параметрической функции полная достаточная статистика является оптимальной оценкой?
10. Запишите уравнение несмещенности.

Лекция 11. Интервальные оценки

Винер сказал: «Чем более вероятно сообщение, тем меньше оно содержит информации».
Такие плакаты ты увидишь в любом зале ожидания.
А. и Б. Стругацкие «Дни кракена»

План лекции: понятие интервального оценивания параметров, построение доверительного интервала с помощью центральной статистики, доверительные интервалы для параметров нормального распределения, примеры расчетов, асимптотический метод.

11.1. Понятие интервального оценивания параметров

Доверительный интервал среднего балла экзамена:
с достоверностью 0,95 — $[72; 74]$,
с достоверностью 1 — $[0; 100]$.
Консультация на экономическом факультете

Мы рассмотрели точечные оценки неизвестных параметров распределения наблюдаемой случайной величины. Всякая оценка параметра представляет собой функцию от выборки, которая для данной реализации выборки равна конкретному значению оценки, принимаемому за приближенное значение параметра. Возникает вопрос, насколько близка данная оценка к неизвестному параметру распределения. Ответить на этот вопрос можно только в вероятностном смысле. Например, указать такой интервал, внутри которого с высокой вероятностью $1 - \alpha$ находится точное значение параметра. Построение таких интервалов называют интервальным или доверительным оцениванием, а соответствующий интервал — доверительным интервалом параметра или интервальной оценкой параметра.

Таким образом, $(1 - \alpha)$ -доверительный интервал — случайный интервал, который с вероятностью $1 - \alpha$ накрывает истинное значение параметра θ . Уровень значимости α обычно берут равным одному из чисел 0,001; 0,005; 0,01; 0,05; 0,1. Уровень значимости выражает ошибку доверительного интервала. Чем меньше α , тем больше доверительная вероятность и тем надежнее доверительный интервал, но более надежный интервал является более широким и менее информативным. Стандартный уровень значимости $\alpha = 0,05$.

Рассмотрим доверительное оценивание скалярного параметра θ . Пусть $X = (X_1, \dots, X_n)$ — выборка из генеральной совокупности наблюдаемой случайной величины ξ с функцией распределения $F(x, \theta)$.

Определение 11.1. Интервал $I = [I_1(X); I_2(X)]$, (где $I_1(X) < I_2(X)$) называют доверительным интервалом значимости α для параметра

θ ($0 < \alpha < 1$), если выполняется условие

$$P(I_1(X) \leq \theta \leq I_2(X)) = 1 - \alpha. \quad (90)$$

Число $1 - \alpha$ называется **доверительной вероятностью**, а $I_1(X)$, $I_2(X)$ — нижней и верхней доверительными границами²⁶.

Таким образом, $(1 - \alpha)$ -доверительный интервал — случайный интервал, который с вероятностью $1 - \alpha$ накрывает истинное значение параметра θ . Заметим, что границы интервала не зависят от параметра θ , а зависят только от выборки²⁷.

Введем сразу же понятие асимптотического доверительного интервала для параметра θ .

Определение 11.2. Интервал $I = [I_1(X); I_2(X)]$, (где $I_1(X) < I_2(X)$) называют **асимптотическим доверительным интервалом значимости α для параметра θ** ($0 < \alpha < 1$), если выполняется условие

$$\lim_{n \rightarrow \infty} P(I_1(X) \leq \theta \leq I_2(X)) = 1 - \alpha. \quad (91)$$

В этом случае говорят об асимптотической значимости α и об асимптотической доверительной вероятности $1 - \alpha$.

Определенные выше интервалы являются **двусторонними**. Можно рассматривать и односторонние интервалы:

$$P(I_1(X) \leq \theta) = 1 - \alpha \text{ или } P(\theta \leq I_2(X)) = 1 - \alpha.$$

Существуют различные подходы к построению доверительных интервалов. Рассмотрим метод нахождения доверительных интервалов, основанный на понятии центральной статистики.

Пусть распределение наблюдаемой случайной величины имеет плотность $f(x)$ и пусть существует такая функция от выборки и параметра $G(X, \theta)$, что:

- 1) распределение случайной величины $G(X, \theta)$ не зависит от θ ;
- 2) при каждом x функция $G(X, \theta)$ непрерывна и строго монотонна по параметру θ .

Такую случайную функцию называют **центральной статистикой**. (Обратите внимание, что центральная статистика зависит от параметра.)

²⁶ Если увидите в определении доверительного интервала строгое неравенство — вспомните про «пресловутую меру нуль».

²⁷ Границы интервала случайны, а параметр неслучаен, поэтому говорить «параметр θ попадает в интервал» было бы неверно. Аналогия: на колышек (неслучайный параметр) набрасывают кольцо (случайный интервал). Не колышек попадает в кольцо, а кольцо накрывает колышек.

11.2. Построение доверительного интервала

Красота нашей жизни – в теории.
М. М. Жванецкий

Обозначим через $g(X, \theta)$ плотность распределения центральной статистики. Функция $g(X, \theta)$ не зависит от параметра θ (по определению центральной статистики), и поэтому для любого заданного $0 < \alpha < 1$ можно подобрать такие числа g_1 и g_2 , $g_1 < g_2$, что

$$P(g_1 \leq G(X, \theta) \leq g_2) = 1 - \alpha. \quad (92)$$

Если определили g_1 и g_2 , то можно составить уравнения относительно θ :

$$G(X, \theta) = g_1, \quad G(X, \theta) = g_2.$$

Поскольку $G(X, \theta)$ — строго монотонная по θ функция, то каждое из уравнений имеет по одному решению $T_1(X), T_2(X)$. Положим

$$I_1(X) = \min\{T_1(X), T_2(X)\},$$

$$I_2(X) = \max\{T_1(X), T_2(X)\}.$$

Тогда события

$$\{g_1 \leq G(X, \theta) \leq g_2\} \text{ и } \{I_1 \leq \theta \leq I_2\}$$

совпадают и, следовательно, совпадают их вероятности, то есть

$$P(I_1 \leq \theta \leq I_2) = 1 - \alpha. \quad (93)$$

Поэтому $[I_1; I_2]$ — доверительный интервал для параметра θ значимости α (доверительной вероятности $1 - \alpha$).

Таким образом, для построения доверительного интервала параметра θ надо взять статистику $G(X, \theta)$, такую, что она сама монотонно зависит от параметра θ , а ее распределение от θ не зависит, записать уравнение $P(g_1 \leq G(X, \theta) \leq g_2) = 1 - \alpha$, и разрешить неравенство под знаком вероятности относительно параметра θ . Монотонность функции $G(X, \theta)$ дает теоретическую возможность разрешить уравнение (хотя и не гарантирует, что это легко сделать).

11.3. Доверительные интервалы для параметров $N(a, \sigma)$

*И теперь в лесу дремучем
Бродит грустный Гайавата,
Непрестанно размышляя,
Вспоминает он нормальный
Тот закон распределенья
Отклонений и ошибок...*
М. Кендалл

1. Рассмотрим построение доверительного интервала для математического ожидания $\theta_1 = a$ в модели $N(\theta_1, \theta_2) = N(a, \sigma)$. Пусть $X = (X_1, \dots, X_n)$ – выборка из данного распределения. Из рассмотренной ранее теоремы следует, что случайная величина

$$t = t(\bar{X}, a) = \sqrt{n-1} \cdot \frac{a - \bar{X}}{S}$$

имеет распределение Стьюдента с $n - 1$ степенью свободы T_{n-1} и эту величину t можно считать центральной статистикой для θ_1 .

Найдем такие t_1 и t_2 , ($t_1 < t_2$), что

$$P \left(t_1 \leq \sqrt{n-1} \cdot \frac{a - \bar{X}}{s} \leq t_2 \right) = 1 - \alpha. \quad (94)$$

Это можно сделать неоднозначно, решением является любая пара t_1, t_2 , такая, что $F_T(t_2) - F_T(t_1) = 1 - \alpha$, где $F_T(x)$ – функция распределения Стьюдента.

Разрешим левую часть (94) относительно a :

$$P \left(\bar{X} + \frac{s}{\sqrt{n-1}} \cdot t_1 \leq a \leq \bar{X} + \frac{s}{\sqrt{n-1}} \cdot t_2 \right) = 1 - \alpha.$$

Отсюда доверительный интервал для $a = \theta_1$:

$$I = \left[\bar{X} + \frac{s}{\sqrt{n-1}} \cdot t_1, \bar{X} + \frac{s}{\sqrt{n-1}} \cdot t_2 \right], \quad (95)$$

где $F_T(t_2) - F_T(t_1) = 1 - \alpha$. Минимизируем длину интервала $l(I) = \frac{s}{\sqrt{n-1}} \cdot (t_2 - t_1)$ при этом условии. По методу Лагранжа для нахождения условного экстремума рассмотрим функцию

$$\psi(\lambda) = \frac{s}{\sqrt{n-1}} \cdot (t_2 - t_1) + \lambda \cdot (F_T(t_2) - F_T(t_1) - 1 + \alpha)$$

и составим систему уравнений

$$\begin{cases} \frac{\partial \psi}{\partial t_1} = -\frac{s}{\sqrt{n-1}} - \lambda \cdot f_T(t_1) = 0, \\ \frac{\partial \psi}{\partial t_2} = \frac{s}{\sqrt{n-1}} + \lambda \cdot f_T(t_2) = 0, \\ \frac{\partial \psi}{\partial \lambda} = F_T(t_2) - F_T(t_1) - 1 + \alpha = 0, \end{cases}$$

где $f_T(x)$ — плотность распределения Стюдента. Из первых двух уравнений получаем, что $f_T(t_1) = f_T(t_2)$. Поскольку распределение Стюдента симметрично относительно 0, то $t_2 = -t_1 = t$, $F_T(t) - F_T(-t) = 2F_T(t) - 1 = 1 - \alpha$. Отсюда $F_T(t) = 1 - \frac{\alpha}{2}$, $t = t_{n-1, 1-\frac{\alpha}{2}}$ (квантиль распределения Стюдента T_{n-1} порядка $1 - \frac{\alpha}{2}$). Получили доверительный интервал минимальной длины

$$I_a = \left[\bar{X} - \frac{s}{\sqrt{n-1}} \cdot t_{n-1, 1-\frac{\alpha}{2}}, \bar{X} + \frac{s}{\sqrt{n-1}} \cdot t_{n-1, 1-\frac{\alpha}{2}} \right]. \quad (96)$$

Замечание 11.1. Вместо минимизации длины, которую не всегда удастся осуществить, можно использовать **центральный интервал**, в котором $g_1 = G_{\frac{\alpha}{2}}$, $g_2 = G_{1-\frac{\alpha}{2}}$. Целесообразно считать, что вероятность попадания случайной величины правее g_2 равна вероятности ее попадания левее g_1 . Обычно, если распределение G симметрично, центральный интервал совпадает с интервалом минимальной длины. Это относится, например, к доверительному интервалу (96).

2. Найдем доверительный интервал значимости α для параметра $\theta_1 = a$ нормального распределения $N(\theta_1, \sigma)$ при известном σ .

В качестве статистики G можно взять $G = \frac{(a-\bar{X})\sqrt{n}}{\sigma}$. Известно, что

$$\frac{(a - \bar{X})\sqrt{n}}{\sigma} = u \in N(0, 1).$$

Поэтому

$$P \left(u_{\frac{\alpha}{2}} \leq \sqrt{n} \cdot \frac{a - \bar{X}}{\sigma} \leq u_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Разрешим относительно a , учитывая, что $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$:

$$P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \leq a \leq \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

$$I_a = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \right]. \quad (97)$$

3. Построим теперь доверительный интервал для $\sigma = \theta_2$ в модели $N(\theta_1, \theta_2)$.

Из теоремы Фишера следует, что случайная величина

$$\frac{nS^2}{\sigma^2} = \frac{nS^2}{\theta_2^2}$$

имеет распределение χ_{n-1}^2 , то есть она зависит от θ_2 , а ее распределение от неизвестных параметров не зависит. Поэтому данную функцию от выборки с учетом ее монотонности можно считать центральной статистикой. Тогда

$$p\left(g_1 \leq \frac{nS^2}{\sigma^2} \leq g_2\right) = 1 - \alpha,$$

и g_1, g_2 находят, решая данное уравнение. Распределение χ^2 задано таблично, и значения квантилей можно взять из таблиц.

Для центрального доверительного интервала справедливо

$$g_1 = z_{\frac{\alpha}{2}}, \quad g_2 = z_{1-\frac{\alpha}{2}},$$

где z_q — квантиль порядка q распределения статистики G . Тогда g_1, g_2 находят по таблице как решения уравнений

$$F_{\chi_{n-1}^2}(x) = \frac{\alpha}{2}, \quad F_{\chi_{n-1}^2}(x) = 1 - \frac{\alpha}{2}.$$

Решениями являются квантили $g_1 = \chi_{n-1, \frac{\alpha}{2}}^2$, $g_2 = \chi_{n-1, 1-\frac{\alpha}{2}}^2$. Разрешая неравенство в левой части уравнения (3) относительно неизвестной дисперсии, получаем

$$P\left(\frac{nS^2}{g_2} \leq \sigma^2 \leq \frac{nS^2}{g_1}\right) = 1 - \alpha. \quad (98)$$

Отсюда доверительный интервал для дисперсии σ^2 значимости α :

$$I_{\sigma^2} = \left[\frac{nS^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right], \quad (99)$$

и доверительный интервал для среднеквадратичного отклонения σ значимости α :

$$I_{\sigma} = \left[\sqrt{\frac{nS^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}}, \sqrt{\frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2}} \right]. \quad (100)$$

4. Найдите самостоятельно доверительный интервал значимости α для параметра $\theta_2 = \sigma$ нормального распределения $N(a, \theta_2)$ при известном a .

Напоминание: в качестве статистики G надо взять такую статистику, которая монотонно зависит от параметра σ , причем ее распределение от σ не зависит. Подсказка: попробуйте поработать со статистикой $G = \sum_{i=1}^n \frac{X_i - a}{\sigma}$.

11.4. Примеры расчетов

*Имеется три разновидности людей:
те, которые умеют считать,
и те, которые этого не умеют²⁸.
Закон Уинкорна*

Пример 11.1. Федеральные центры оздоровительного питания изучают потребление населением биологически активных добавок (БАД). Было опрошено 4 группы по 100 покупателей каждая. В среднем из 100 опрошенных встречалось 60 человек употребляющих БАД, причем $\sigma = 10$. Каков доверительный интервал значимости $\alpha = 0,05$ для числа людей, употребляющих БАД, на 100 опрошенных?

◀ Требуется найти доверительный интервал значимости $\alpha = 0,05$ для параметра a нормального распределения $N(a, \sigma]$ при известном $\sigma = 10$; воспользуемся вышеприведенной формулой:

$$\begin{aligned} I_a &= \left[\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}; \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \right] = \\ &= \left[60 - \frac{10}{\sqrt{4}} \cdot u_{0,975}; 60 + \frac{10}{\sqrt{4}} \cdot u_{0,975} \right] = \\ &= \left[60 - \frac{10}{2} \cdot 1,96; 60 + \frac{10}{2} \cdot 1,96 \right] = \\ &= [60 - 9,8; 60 + 9,8] = [50,2; 69,8]. \end{aligned}$$

²⁸ На самом деле, конечно, это 10 разновидностей.

СПРАВКА. Квантили $N(0, 1)$: $u_{0,95} = 1,64$, $u_{0,975} = 1,96$, $u_{0,995} = 2,58$.

Пример 11.2. В условиях предыдущего примера как изменился бы доверительный интервал, если бы те же данные были получены по 25 группам численностью 100 человек каждая?

$$\begin{aligned} I_a &= \left[\bar{X} - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}; \bar{X} + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \right] = \\ &= \left[60 - \frac{10}{\sqrt{25}} \cdot u_{0,975}; 60 + \frac{10}{\sqrt{25}} \cdot u_{0,975} \right] = \\ &= \left[60 - \frac{10}{5} \cdot 1,96; 60 + \frac{10}{5} \cdot 1,96 \right] = \\ &= [60 - 3,92; 60 + 3,92] = [56,08; 63,92]. \end{aligned}$$

Длина интервала равна $3,92 \cdot 2 = 7,84$, а в предыдущем примере длина доверительного интервала равнялась $9,8 \cdot 2 = 19,6$, то есть новый интервал в 2,5 раза короче. ►

11.5. Асимптотические доверительные интервалы

*Надпись «Нет в жизни счастья» он смыл бесповоротно.
Счастье было! Вот оно было! Прямо перед нами!
Ю. Коваль, «Суер-Вьер»*

Параметр произвольного распределения можно оценить так же, как параметр нормального распределения, если известно распределение некоторой статистики, зависящей от параметра и его оценки (эту информацию можно найти в справочниках, а можно вывести самим). Другой способ связан с использованием **асимптотического метода** (это и есть заявленное счастье. Почти всегда можно обойтись нормальным распределением). Если $\hat{\theta}$ асимптотически нормальна и несмещена (например, является о.м.п.), то

$$\begin{aligned} \frac{\hat{\theta} - \theta}{\sqrt{\mathbf{D}\hat{\theta}}} &\xrightarrow{d} u \in N(0, 1). \\ 1 - \alpha &= p \left(u_{\frac{\alpha}{2}} \leq \frac{\hat{\theta} - \theta}{\sqrt{\mathbf{D}\hat{\theta}}} \leq u_{1-\frac{\alpha}{2}} \right). \end{aligned} \quad (101)$$

Разрешим относительно θ , получим доверительный интервал значимости α . Кстати, длина его пропорциональна $\sqrt{\mathbf{D}\hat{\theta}}$ и минимальна, когда минимальна $\mathbf{D}\hat{\theta}$, то есть при эффективной $\hat{\theta}$.

Пример 11.3. Построим асимптотический доверительный интервал для параметра распределения Пуассона λ уровня значимости α .

◀ Для этого рассмотрим статистику

$$S = \frac{\sqrt{n}(\bar{X} - \lambda)}{\sqrt{\lambda}} \xrightarrow{d} u \in N(0, 1)$$

и найдем из уравнения $P(I_1 \leq \lambda \leq I_2) = 1 - \alpha$ доверительный интервал $[I_1, I_2]$.

$$P(u_{\alpha/2} \leq S \leq u_{1-\alpha/2}) = 1 - \alpha;$$

$$u_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - \lambda)}{\sqrt{\lambda}} \leq u_{1-\alpha/2}.$$

Так как $-u_{\alpha/2} = u_{1-\alpha/2}$, то

$$\left| \frac{\sqrt{n}(\bar{X} - \lambda)}{\sqrt{\lambda}} \right| \leq u_{1-\alpha/2},$$

$$\frac{n(\bar{X} - \lambda)^2}{\lambda} \leq u_{1-\alpha/2}^2,$$

$$n\lambda^2 - \lambda(2n\bar{X} + u_{1-\alpha/2}^2) + n\bar{X}^2 \leq 0.$$

Решая относительно λ последнее неравенство, получаем интервал

$$\left[\frac{2n\bar{X} + u_{1-\alpha/2}^2 \pm u_{1-\alpha/2} \sqrt{4n\bar{X} + u_{1-\alpha/2}^2}}{2n} \right].$$

►

Замечание 11.2. Точный доверительный интервал для параметра распределения Пуассона λ можно найти, используя точное распределение \bar{X} .

Пример 11.4. Найдем асимптотический доверительный интервал для параметра λ показательного распределения E_λ уровня значимости α .

◀ Вспомним, что о.м.п. λ равна $\hat{\lambda} = \frac{1}{\bar{X}}$. Чтобы не искать дисперсию $\frac{1}{\bar{X}}$, изменим параметризацию, возьмем за параметр $\frac{1}{\lambda}$. В силу инвариантности о.м.п., оценка $\frac{1}{\lambda}$ равна \bar{X} . Тогда

$$\sqrt{n} \frac{\bar{X} - \frac{1}{\lambda}}{\frac{1}{\lambda}} = \sqrt{n}(\lambda\bar{X} - 1) \xrightarrow{d} u \in N(0, 1),$$

и дальше, как в предыдущих примерах (только вычисления проще).▶

11.6. Контрольные вопросы

1. Чем отличается интервальная оценка от точечной?
2. Дайте определение доверительного интервала.
3. Что выражает уровень значимости?
4. Что выражает доверительная вероятность?
5. Как связана доверительная вероятность с длиной доверительного интервала?
6. Дайте определение центральной статистики.
7. В чем состоят преимущества использования центральной статистики для построения доверительного интервала?
8. В чем отличия доверительных интервалов математического ожидания нормального распределения при известной и неизвестной дисперсии?
9. Какой информацией надо располагать, чтобы построить доверительный интервал для математического ожидания нормального распределения?

Лекция 12. Основные понятия теории проверки гипотез

*Чтобы правильно задать вопрос,
надо знать большую часть ответа.*
Роберт Шекли

План лекции: основные понятия, примеры математических формулировок гипотез, общая схема проверки параметрических гипотез, ошибки первого и второго рода, мощность критерия.

12.1. Основные понятия

Определение 12.1. *Статистической гипотезой (в дальнейшем просто гипотезой) называется любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин.*

Задача разработки рациональных методов проверки статистических гипотез — одна из основных в математической статистике.

Если для исследуемого явления (процесса, стратегии) сформулирована та или иная гипотеза H_0 (основная, нулевая), то задача состоит в том, чтобы сформулировать такое правило, которое позволяло бы по имеющимся результатам наблюдений отклонить или принять (точнее, не отклонить) эту гипотезу.

Нулевая гипотеза H_0 обычно соответствует естественному, наиболее ожидаемому состоянию (норме).

Определение 12.2. *Правило, согласно которому проверяют гипотезу H_0 (отвергают или не отвергают), называется **статистическим критерием** проверки гипотезы H_0 .*

Таким образом, разработка правил и их обоснование с точки зрения оптимальности и составляет предмет теории проверки статистических гипотез.

Определение 12.3. *Любая гипотеза о распределении наблюдаемой случайной величины, которая может оказаться истинной, но отличается от основной гипотезы, называется **альтернативной** гипотезой.*

Определение 12.4. *Статистическая гипотеза называется **простой**, если однозначно фиксирует распределение наблюдений. Иначе это **сложная** гипотеза.*

Основная идея проверки статистической гипотезы: чтобы построить критерий, возьмем статистику $T = T(X)$ (функцию от результатов наблюдений), такую, что известно ее распределение при условии, что проверяемая гипотеза H_0 верна. Если же гипотеза H_0 не верна, распределение T должно отличаться от распределения T/H_0 (распределения T при условии, что верна H_0). Эта функция T , являющаяся случайной величиной, называется статистикой критерия. Далее мы вычислим значение статистики T по выборке. Поскольку известно распределение T/H_0 , можно оценить вероятность наблюдаемого значения. И если эта вероятность слишком мала, мы делаем вывод, что столь маловероятное значение не должно было наблюдаться. Но оно наблюдается, это факт. Следовательно, распределение, из которого мы исходили, T/H_0 — неверно. Значит, неверна гипотеза H_0 и она отвергается.

Разумеется, принятое решение может быть ошибочным (например, случайно произошло маловероятное событие). Ошибки, которые могут возникать при проверке гипотез, будут подробно рассмотрены позднее.

12.2. Примеры математических формулировок гипотез

Количество разумных гипотез, способных объяснить любой феномен, бесконечно.

Постулат гипотез

1. **Гипотеза о виде распределения.** Пусть проводят n независимых наблюдений над некоторой случайной величиной ξ с неизвестной функцией распределения $F(x)$. Нулевая гипотеза может иметь вид

$$H_0 : F(x) = F_0(x),$$

где $F_0(x)$ полностью задана, или

$$H_0 : F(x) \in \{\mathcal{F}\},$$

где $\{\mathcal{F}\}$ — заданное семейство функций распределения. При этом обычно семейство $\{\mathcal{F}\}$ задают в параметрическом виде: $\{F_\theta\}$. Здесь $F_\theta = F(x, \theta)$.

2. **Гипотеза однородности.** Произведено k серий независимых наблюдений. Если можно считать, что закон распределения наблюдений от серии к серии не меняется, то говорят, что статистические данные однородны. Пусть $F_l(x)$ — функция распределения²⁹ наблюдений l -й се-

²⁹ Функции распределения наблюдений, как правило, неизвестны, а то мы бы просто их сравнили. Задача как раз в том, чтобы сравнить их не зная.

рии, $l = 1, \dots, k$. Гипотеза однородности имеет вид

$$H_0 : F_1(x) \equiv \dots \equiv F_k(x).$$

3. **Гипотеза независимости.** Наблюдается двумерная случайная величина (ξ, η) с неизвестной функцией распределения $F_{\xi, \eta}(x, y)$, проверяется предположение о независимости компонент (ξ, η) :

$$H_0 : F_{\xi, \eta}(x, y) = F_{\xi}(x)F_{\eta}(y).$$

4. **Гипотеза случайности.** Результат эксперимента описывают случайной n -мерной величиной $X = (X_1, \dots, X_n)$ с неизвестной функцией распределения $F(x)$. Для выяснения, можно ли рассматривать X как случайную выборку из распределения некоторой случайной величины ξ (то есть являются ли компоненты X_i независимыми и одинаково распределенными), проверяют гипотезу случайности:

$$H_0 : F = F_{X_1} = \dots = F_{X_n} = F_{\xi}.$$

5. **Параметрические гипотезы.** Это один из наиболее важных классов гипотез³⁰.

Определение 12.5. *Статистическая гипотеза называется **параметрической**, если она представляет собой предположение о том, что неизвестный параметр распределения имеет наперед заданное значение или множество значений.*

Под параметром θ , как обычно, понимается вектор $\theta = (\theta_1, \dots, \theta_r)$. В этом случае класс допустимых распределений наблюдаемой случайной величины ξ имеет вид $\{F(x, \theta_1), \dots, F(x, \theta_r)\}$ и гипотезы относятся к неизвестным скалярным параметрам $\theta_i, i = 1, \dots, r$ — ко всем или к части их.

Рассмотрим несколько вариантов параметрических гипотез.

Пример 12.1.



1. $H_0 : \theta_1 = \dots = \theta_r.$
2. $H_0 : \theta = \theta_0$, где θ_0 — фиксированное известное значение.
3. $H_0 : g(\theta) = g(\theta_0)$, где $g(t)$ — функция.

³⁰ Надо иметь в виду, что рассмотренные примеры не представляют из себя исчерпывающего перечня возможных гипотез.



Пример 12.2.

◀ $H_0 : \theta_0 \in \Theta_0$ (сложная гипотеза).

Альтернативная гипотеза : $H_1 : \theta = \theta_1$, где $\theta_1 \neq \theta_0$ — простая гипотеза, а варианты $\theta > \theta_0$ или $\theta < \theta_0$ или $\theta \in \Theta_1$ задают сложные альтернативные гипотезы.▶

Пример 12.3. Для нормального распределения $N(\theta_1, \theta_2)$:

◀ $H_0: \theta_1 = \theta_1^0, \theta_2 = \theta_2^0$ — простая гипотеза, а

$H_0: \theta_1 = \theta_1^0$ — сложная гипотеза, так как θ_2 — свободный параметр.▶

12.3. Общая схема проверки параметрических гипотез

*Природа в ответ на наши вопросы
отвечает «нет» громко, а «да» тихо.
Фольклор*

Рассмотрим процесс проверки гипотезы на примере параметрической гипотезы.

Пусть (X_1, \dots, X_n) — выборка объема n из распределения $F(x, \theta)$, $\theta \in \Theta$, а (x_1, \dots, x_n) — реализация. Сформулированы гипотезы (сложные в общем виде):

$H_0: \theta \in \Theta_0$ — нулевая, $H_1: \theta \in \Theta_1$ — альтернативная.

Пока будем для ясности рассматривать простые гипотезы, когда каждое из множеств Θ_0, Θ_1 содержит только одну точку:

$H_0: \theta = \theta_0$ — нулевая, $H_1: \theta = \theta_1$ — альтернативная.

Необходимо построить **критерий**, то есть правило, по которому для каждой реализации x можно было бы принять H_0 или отвергнуть H_0 (принять H_1). Тем самым выборочная область $X = \{(x_1, \dots, x_n)\}$ разбивается на две части \bar{V} и V :

$$\bar{V} \cap V = \emptyset, \quad \bar{V} \cup V = X,$$

где \bar{V} состоит из тех (x_1, \dots, x_n) , для которых H_0 принимают, а множество V состоит из тех (x_1, \dots, x_n) , для которых принимают H_1 . Множество \bar{V} называют областью принятия гипотезы H_0 , а V — областью ее отклонения или **критической областью**.

Тогда критерий формулируется следующим образом: пусть x — наблюдаемая реализация выборки, тогда при $x \in V$ гипотезу H_0 отвергают (в пользу альтернативной гипотезы H_1); если же $x \in \bar{V}$, то H_0 не отвергают.

Обратите внимание: если $x \in V$, то гипотезу H_0 отвергают (громкое «нет»). В противном случае ($x \notin V$) мы не можем утверждать, что гипотеза подтвердилась, тем более — что она доказана, можно сказать только, что гипотеза H_0 не отвергается (тихое «да»). С практической точки зрения это означает, что гипотезу H_0 можно использовать дальше и опираться на нее при получении различных выводов.

Для простоты мы будем иногда говорить о принятии гипотезы, вкладывая в это смысл «гипотеза не отвергается».

Критерий, определяемый критической областью V , называют **критерием** K_V .

Обозначим через H_d статистическое решение, тогда определение критерия 12.2 может быть конкретизировано следующим образом:

$$K_V(x) = H_d(x) = \begin{cases} H_1, & x \in V, \\ H_0, & x \in \bar{V}. \end{cases}$$

Обычно критическую область V задают с помощью статистики $T(x)$, для которой распределение известно точно или приближенно. Однако нет необходимости непременно связывать понятие критерия со статистикой, поскольку в конечном счете нас интересует соответствие между x и принимаемым статистическим решением. Это соответствие можно задать функцией $\varphi(x)$, которая принимает значение 1, когда следует отвергнуть H_0 , и значение 0, когда H_0 не отвергается:

$$K_V(x) = H_d(x) = \begin{cases} H_1, & \varphi(x) = 1, \\ H_0, & \varphi(x) = 0. \end{cases} \quad (102)$$

В процессе проверки гипотезы можно принять правильное решение или совершить ошибку.

12.4. Ошибки первого и второго рода

*Авторизованный пользователь классифицируется как нарушитель
— ошибка первого рода,
нарушитель классифицируется как авторизованный пользователь
— ошибка второго рода.*
Рольф Т. Моултон

Определение 12.6. Вероятностью **ошибки первого рода** называется вероятность отклонить H_0 , когда H_0 верна.

Эта вероятность обозначается α и называется еще **уровнем значимости**³¹ критерия или **размером** критерия. Очевидно,

$$\alpha = P(H_d = H_1/H_0) = P(X \in V/H_0),$$

(α равняется вероятности того, что X принадлежит критической области V при условии, что верна H_0).

Определение 12.7. Вероятностью **ошибки второго рода** называется вероятность принять H_0 , когда H_0 не верна.

Вероятность ошибки второго рода обозначается β . Очевидно,

$$\beta = P(H_d = H_0/H_1) = P(X \in \bar{V}/H_1),$$

(β равняется вероятности того, что X не принадлежит критической области V при условии, что верна H_1).

Пример 12.4. При контроле качества продукции возможны ошибки: изделие, удовлетворяющее стандарту, отбраковывают с вероятностью γ , а бракованное изделие принимают с вероятностью δ .

Если ввести для данного изделия две гипотезы: H_0 : {изделие годное} и H_1 : {изделие бракованное}, и в качестве критерия выбора использовать контроль продукции, то γ — вероятность ошибки 1-го рода, а δ — вероятность ошибки 2-го рода данного критерия.

Подумаем, что изменится, если гипотезы являются сложными. Простой гипотезе соответствует единственное распределение, а сложной — множество распределений. Пусть H_0 — сложная гипотеза и ей соответствуют распределения F_1, \dots, F_m (для параметрической гипотезы $F_i = F_{\theta_i}$). Тогда вероятность

$$\alpha = P(H_d = H_1/H_0) = P(X \in V/H_0)$$

будет зависеть от того, при каком именно из распределений F_1, \dots, F_m вычисляется вероятность.

Если гипотез больше двух, например, имеются гипотезы H_1, \dots, H_k , то говорят, что произошла ошибка i -го рода, если гипотеза H_i отвергнута

³¹ Некоторые математики, например, А.А. Боровков, называют уровнем значимости величину $1 - \alpha$, и это обоснованно, поскольку α — малое число, а $1 - \alpha$, как правило, близка к 1 (чтобы у хорошего критерия значимость была высокой). Но мы используем традиционную терминологию, в которой уровень значимости означает ошибку и должен быть низким у хорошего критерия. Из контекста всегда бывает ясно, какой смысл вкладывает автор в это понятие.

критерием, в то время как она верна. Вероятностью ошибки i -го рода называется число

$$\alpha_i(K_V) = P(H_d(X) \neq H_i/H_i \text{ верна}) = P_{H_i}(K_V(X) \neq H_i).$$

При рассмотрении простых параметрических гипотез $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$ используют обозначения

$$\alpha = P(X \in V/\theta_0),$$

$$\beta = P(X \in \bar{V}/\theta_1).$$

Иногда удобно использовать обозначения $\alpha_1 = \alpha$, $\alpha_2 = \beta$.

Ошибки первого и второго рода взаимосвязаны; при попытке уменьшить одну из ошибок критерия другая, как правило³², увеличивается (рис. 6).

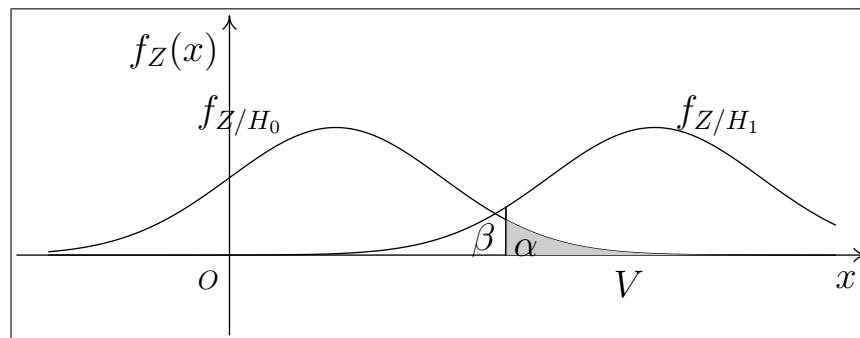


Рис. 6. Ошибки первого (α) и второго (β) рода

Из рис. 6 видно, что если передвигать границу критической области d вправо, вероятность ошибки первого рода α уменьшается, но зато вероятность ошибки второго рода β увеличивается.

Вероятности этих ошибок можно выразить через функцию мощности критерия.

Определение 12.8. *Функцией мощности критерия K_V называется функция*

$$W(V, \theta) = P(X \in V), \quad \theta \in \Theta. \quad (103)$$

В случае простых гипотез $\Theta = \{\theta_0, \theta_1\}$. Через $W(V, \theta)$ легко вычислить

$$\alpha = P(X \in V/\theta_0) = W(V, \theta_0) \quad (104)$$

³² Как правило, но не всегда.

— вероятность совершения ошибки первого рода и

$$\beta = P(X \in \bar{V}/\theta_1) = 1 - P(X \in V/\theta_1) = 1 - W(V, \theta_1) \quad (105)$$

— вероятность совершения ошибки второго рода. Таким образом,

$$W(V, \theta) = \begin{cases} \alpha, & \theta = \theta_0, \\ 1 - \beta, & \theta = \theta_1. \end{cases}$$

Величину $1 - \beta$ будем называть **мощностью критерия** K_V и обозначать $M(K_V)$ или просто $M(K)$. Тогда

$$W(V, \theta) = \begin{cases} \alpha, & \theta = \theta_0, \\ M(K), & \theta = \theta_1. \end{cases}$$

Замечание 12.1. Понятие мощности критерия введено для случая простых H_0, H_1 ; существенно, что множество Θ_1 состоит из единственной точки θ_1 .

12.5. Контрольные вопросы

1. Дайте определение статистической гипотезы.
2. Приведите примеры статистических гипотез.
3. Что такое нулевая и альтернативная гипотезы?
4. Что такое простая и сложная гипотезы?
5. Приведите примеры простых и сложных гипотез.
6. Что выражает уровень значимости при проверке гипотезы?
7. Что выражает доверительная вероятность?
8. Дайте определение ошибки первого рода.
9. Как связаны уровень значимости и вероятность ошибки первого рода?
10. Дайте определение ошибки второго рода.
11. Можно ли неограниченно уменьшать вероятность ошибки первого рода?

12. Можно ли неограниченно уменьшать вероятность ошибки второго рода?
13. Дайте определение мощности критерия.
14. Как связаны мощность критерия и вероятность ошибки второго рода?
15. Сформулируйте гипотезу о виде распределения.
16. Сформулируйте гипотезу однородности.
17. Сформулируйте гипотезу независимости.
18. Сформулируйте гипотезу случайности.
19. Какие гипотезы называются параметрическими?
20. Приведите примеры параметрических гипотез.

Лекция 13. Подходы к сравнению критериев

Следующее замечательное утверждение, по недоразумению называемое леммой, заявляет, что оптимальные во всех трех смыслах (минимаксные, байесовские, наиболее мощные) критерии могут быть построены в самом общем случае простым выбором различных констант в одном и том же критерии — критерии отношения правдоподобия.

Н. И. Чернова

План лекции: сравнение мощности критериев, критерий Неймана — Пирсона, состоятельность критерия, рандомизированные критерии, байесовские и минимаксные критерии, лемма Неймана — Пирсона.

13.1. Сравнение мощности критериев

Желательно выбрать критерий так, чтобы вероятности всех ошибок были по возможности малы³³ (речь идет об ошибках первого и второго рода в случае двух гипотез, ошибках i -го рода в общем случае). Напомним, что ошибка i -го рода состоит в отвержении критерием гипотезы H_i в то время как она верна. Ошибки i -го рода связаны между собой: при попытке уменьшить одну из ошибок критерия другие обычно (но не обязательно) увеличиваются. На множестве всех критериев для проверки гипотез H_1, \dots, H_k можно ввести частичный порядок следующим образом.

Определение 13.1. *Критерий K_1 мощнее критерия K_2 , если для всех $i = 1, \dots, k$ вероятности ошибок $\alpha_i(K_1) \leq \alpha_i(K_2)$ и хотя бы для одного i имеет место строгое неравенство.*

Данное понятие легко распространить на случай нескольких критериев.

Определение 13.2. *Пусть значения всех ошибок, кроме одной (i -й), зафиксированы и все критерии, имеющие данные ошибки, образуют класс S . Критерий K называется **наиболее мощным критерием (н.м.к.)** в классе S , если $\alpha_i(K) \leq \alpha_i(K') \forall K' \in S$.*

Для ясности далее будем рассматривать две простые гипотезы. В этом случае определение означает, что лучше тот критерий, у которого хотя бы одна из ошибок строго меньше, а вторая не больше соответствующих ошибок другого критерия. Однако не любые критерии сравнимы в этом смысле, например, типичной является ситуация: $\alpha_1(K_1) \leq \alpha_1(K_2)$, но $\alpha_2(K_1) \geq$

³³ Употребляется также выражение «малая ошибка», имеющее смысл «ошибка, имеющая малую вероятность».

$\alpha_2(K_2)$. Один из подходов к решению этой проблемы — зафиксировать одну из ошибок и минимизировать вторую.

Что для нас критичнее: совершить ошибку первого или второго рода? Вспомним пример 12.4, что опаснее, забраковать годное (ошибка первого рода), или пропустить негодное (ошибка второго рода)? Обычно (но не всегда) для практики важнее не совершить ошибку второго рода. Поэтому поступают таким образом: фиксируют ошибку первого рода на достаточно низком безопасном уровне и выбирают критическую область V так, чтобы ошибка второго рода была бы минимальна (что соответствует наибольшей вероятности отклонения гипотезы H_0 , когда она неверна), то есть была бы максимальной функция мощности критерия при H_1 .

В случае, когда альтернативная гипотеза H_1 простая, наилучшим будет критерий, который имеет наибольшую мощность среди всех других критериев с заданным уровнем значимости α . Если альтернативная гипотеза H_1 сложная, то мощность критерия будет функцией, определенной на классе простых альтернатив, составляющих H_1 .

Определение 13.3. *Равномерно наиболее мощным критерием (р.н.м.к.) размера α называется статистический критерий с заданным размером (уровнем значимости) α для проверки сложной гипотезы H_0 против сложной альтернативы H_1 , мощность которого не меньше мощности любого другого статистического критерия, предназначенного для проверки H_0 против H_1 и имеющего тот же размер α .*

Таким образом, р.н.м.к. — это критерий, имеющий наибольшую мощность при каждой альтернативной гипотезе из класса H_1 .

Такой критерий не обязательно существует и не обязательно единствен.

Наиболее мощные критерии существуют в случае двух простых гипотез³⁴.

Определение 13.4. *В случае двух простых гипотез наилучшей критической областью (НКО) называется область V , которая при заданном уровне значимости α обеспечивает $\min \beta$ — ошибки второго рода.*

Таким образом, НКО — критическая область, обеспечивающая максимальную мощность, и критерий, использующий НКО, является наиболее мощным критерием (н.м.к.).

³⁴ В случае когда альтернативная гипотеза простая, вместо термина р. н. м. к. используют термин «наиболее мощный критерий» (н.м.к.).

НКО можно найти с помощью следующей теоремы.

13.2. Критерий Неймана – Пирсона

Теорема 13.1 (критерий Неймана – Пирсона). Пусть F_{θ_i} абсолютно непрерывны и $f_{\theta_i} > 0$, $i = 1, 2$; $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$. Тогда НКО заданного уровня значимости α состоит из точек выборочного пространства, удовлетворяющих неравенству

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c_\alpha, \quad (106)$$

где c_α – константа, зависящая от α , L – функция правдоподобия.

Доказательство. Пусть V – критическая область, удовлетворяющая (106), то есть

$$V = \left\{ x : \frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c_\alpha \right\},$$

и пусть V_1 – некоторая другая критическая область того же уровня значимости α .

Рассмотрим

$$\begin{aligned} W(V_1, \theta_1) &= P(X \in V_1 / \theta_1) = \int_{V_1} L(x, \theta_1) dx = \\ &= \int_{V_1 \cap V} L(x, \theta_1) dx + \int_{V_1 \cap \bar{V}} L(x, \theta_1) dx. \end{aligned}$$

Теперь найдем $W(V, \theta_1)$:

$$\begin{aligned} W(V, \theta_1) &= P(X \in V / \theta_1) = \int_V L(x, \theta_1) dx = \\ &= \int_{V_1 \cap V} L(x, \theta_1) dx + \int_{V \cap \bar{V}_1} L(x, \theta_1) dx. \end{aligned}$$

Выразим одну функцию мощности через другую:

$$W(V_1, \theta_1) = W(V, \theta_1) + \int_{V_1 \cap \bar{V}} L(x, \theta_1) dx - \int_{V \cap \bar{V}_1} L(x, \theta_1) dx.$$

По условию точки области V удовлетворяют условию

$$L(x, \theta_1) \geq c_\alpha L(x, \theta_0).$$

Значит, это верно и для подмножества V области $V\overline{V}_1$. Соответственно, точки области \overline{V} (в частности, точки $V_1\overline{V}$) удовлетворяют противоположному условию:

$$L(x, \theta_1) < c_\alpha L(x, \theta_0).$$

Таким образом,

$$\begin{cases} -L(x, \theta_1) \leq -c_\alpha L(x, \theta_0), & x \in V\overline{V}_1, \\ L(x, \theta_1) < c_\alpha L(x, \theta_0), & x \in V_1\overline{V}. \end{cases}$$

Тогда

$$\begin{aligned} W(V_1, \theta_1) &\leq W(V, \theta_1) + c_\alpha \int_{V_1\overline{V}} L(x, \theta_0) dx - c_\alpha \int_{V\overline{V}_1} L(x, \theta_0) dx = \\ &= W(V, \theta_1) + c_\alpha \left[\int_{V_1\overline{V}} L(x, \theta_0) dx - \int_{V\overline{V}_1} L(x, \theta_0) dx \right]. \end{aligned} \quad (107)$$

Рассмотрим теперь два последних интеграла:

$$\begin{aligned} \int_{V_1\overline{V}} L(x, \theta_0) dx &= \int_{V_1} L(x, \theta_0) dx - \int_{V_1V} L(x, \theta_0) dx, \\ \int_{V\overline{V}_1} L(x, \theta_0) dx &= \int_V L(x, \theta_0) dx - \int_{V_1V} L(x, \theta_0) dx. \end{aligned}$$

Подставим в (107):

$$W(V_1, \theta_1) \leq W(V, \theta_1) + c_\alpha \left[\int_{V_1} L(x, \theta_0) dx - \int_V L(x, \theta_0) dx \right].$$

Но

$$\int_{V_1} L(x, \theta_0) dx - \int_V L(x, \theta_0) dx = \alpha,$$

так как это вероятность совершения ошибки первого рода, см. (104). Окончательно получаем

$$W(V_1, \theta_1) \leq W(V, \theta_1),$$

то есть критическая область V , заданная неравенством (106), обеспечивает максимальную мощность, или, что то же, минимальную ошибку второго рода. Следовательно, V — наилучшая критическая область (НКО).

Определение 13.5. Построенный критерий вида

$$H_d(X) = \begin{cases} H_0, & \frac{L(x, \theta_1)}{L(x, \theta_0)} < c, \\ H_1, & \frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c \end{cases} \quad (108)$$

называется критерием отношения правдоподобия (КОП) или критерием Неймана – Пирсона.

Смысл критерия вполне прозрачен: действительно, функция правдоподобия выражает вероятность получения данной выборки. В случае, когда $L(x, \theta_1)$ больше $L(x, \theta_0)$, альтернативная гипотеза более вероятна. По критерию мы должны принимать H_1 , если $L(x, \theta_1)$ больше $L(x, \theta_0)$ в c «раз», причем c находится по заданному уровню значимости. Наилучшая критическая область, получаемая в результате решения неравенства (106), обычно имеет вид

$$V = \{x : T(x) \leq C\}, \text{ или } \{x : T(x) \geq C\}, \text{ или } \{x : |T(x)| \geq C\},$$

где C – критическое значение статистики $T(x)$ (берется из таблиц распределения).

Пример 13.1. $\langle N(\theta, \sigma) \rangle$. $H_0 : \theta = a_0$, $H_1 : a = a_1$, $a_1 > a_0$. Найти НКО.



$$\frac{L(x, \theta_1)}{L(x, \theta_0)} = \frac{\frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - a_1)^2}{2\sigma^2}}}{\frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{\sum_{i=1}^n (x_i - a_0)^2}{2\sigma^2}}} = e^{-\frac{1}{2\sigma^2}(na_1^2 - na_0^2 - 2(a_1 - a_0) \sum x_i)} \geq c_\alpha.$$

$$-\frac{1}{2\sigma^2}(na_1^2 - na_0^2 - 2(a_1 - a_0) \sum x_i) \geq \ln c_\alpha.$$

Выразим член, содержащий статистику (функцию от x), и переобозначим для простоты константу:

$$-2(a_1 - a_0) \sum x_i \leq b.$$

$a_1 > a_0$, поэтому

$$\sum x_i \geq d$$

(где d – новая константа). Таким образом, мы нашли статистику

$$T(x) = \sum x_i$$

и форму НКО:

$$V = \{x : \sum x_i \geq d\}. \quad (109)$$

Чтобы найти границу критической области, перейдем к стандартной статистике (то есть к статистике, которая зависит от $T(x) = \sum x_i$ и распределение которой нам известно). Такой статистикой является

$$Z = \frac{\bar{X} - a_0}{\sigma} \sqrt{n}.$$

Ранее было показано, что $Z \in N(0, 1)$. Неравенство $\sum x_i \geq d$ равносильно неравенству $Z \geq u$, где u — новая константа. Поэтому НКО, заданная через Z , будет иметь форму

$$V = \{x : Z \geq u\}.$$

По заданному α из соотношения $\alpha = P(Z \in V/a_0) = P(Z \geq u)$ найдем u :

$$P(Z \geq u) = 1 - F_Z(u) = 1 - \Phi(u) = \alpha,$$

следовательно,

$$\Phi(u) = 1 - \alpha, \quad u = u_{1-\alpha},$$

а НКО:

$$\frac{\bar{X} - a_0}{\sigma} \sqrt{n} \geq u_{1-\alpha}.$$



Пример 13.2. В условиях предыдущего примера найти мощность критерия.

◄ В предыдущем примере была найдена НКО для проверки гипотез $H_0 : a = a_0$, $H_1 : a = a_1$, $a_1 > a_0$ в нормальной модели $\langle N(\theta, \sigma) \rangle$:

$$\frac{\bar{X} - a_0}{\sigma} \sqrt{n} \geq u_{1-\alpha}.$$

По определению, мощность критерия $M(K)$ равна $1 - \beta$, где

$$\beta = P(T(x) \in \bar{V}/\theta_1) = 1 - P(T(x) \in V/\theta_1).$$

Тогда

$$M(K) = P(T(x) \in V/\theta_1).$$

Статистика $T(x)$ в данном случае равна $\frac{\bar{X} - a_0}{\sigma} \sqrt{n}$, $\theta_1 = a_1$, а критическая область $V = [u_{1-\alpha}, \infty)$. Таким образом,

$$M(K) = P_{a_1} \left(\frac{\bar{X} - a_0}{\sigma} \sqrt{n} \geq u_{1-\alpha} \right).$$

(P_{a_1} означает, что вероятность рассчитывается на основе модели $N(a_1, \sigma)$.)

$$\begin{aligned} M(K) &= P_{a_1} \left(\frac{\bar{X} - a_0}{\sigma} \sqrt{n} \geq u_{1-\alpha} \right) = P_{a_1} \left(\bar{X} \geq \frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 \right) = \\ &= 1 - P_{a_1} \left(\bar{X} < \frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 \right). \end{aligned}$$

Вспомним, что в $N(a_1, \sigma)$ \bar{X} имеет распределение $N(a_1, \frac{\sigma}{\sqrt{n}})$. Тогда

$$P_{a_1} \left(\bar{X} < \frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 \right) = F_{\bar{X}} \left(\frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 \right) = \Phi \left(\frac{\frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 - a_1}{\frac{\sigma}{\sqrt{n}}} \right)$$

и

$$M(K) = 1 - \Phi \left(\frac{\frac{\sigma u_{1-\alpha}}{\sqrt{n}} + a_0 - a_1}{\frac{\sigma}{\sqrt{n}}} \right).$$

Можно слегка преобразовать последнее выражение, используя свойство $\Phi(x)$: $\Phi(-x) = 1 - \Phi(x)$, и свойство квантилей стандартного нормального распределения: $u_\alpha = -u_{1-\alpha}$. Окончательно получим

$$M(K) = \Phi \left(u_\alpha + \frac{(a_1 - a_0)\sqrt{n}}{\sigma} \right).$$

►

Пример 13.3. В условиях предыдущего примера найти мощность критерия $H_0 : a = 1$, $H_1 : a = 2$, в нормальной модели $\langle N(\theta, 2) \rangle$ по выборке объема $n = 100$ при уровне значимости $\alpha = 0,05$.

◀

$$M(K) = \Phi \left(u_{0,05} + \frac{(2 - 1)\sqrt{100}}{2} \right).$$

Из таблиц находим (или вспоминаем), что $u_{0,05} = -1,64$.

$$M(K) = \Phi \left(-1,64 + \frac{(2 - 1)\sqrt{100}}{2} \right) = \Phi(3,36) = 0,999610288.$$

Как видим, мощность данного критерия (то есть его способность правильно отвергнуть гипотезу) практически равна 1. ►

Замечание 13.1. Критерий Неймана – Пирсона применим и к простым гипотезам о виде распределения.

13.3. Состоятельность критерия

Важным желательным свойством критерия является увеличение его мощности при возрастании объема выборки. Пусть имеется выборка $X = (X_1, \dots, X_n)$, $X_i \in F$ и проверяется простая основная гипотеза H_0 против сложной альтернативы H_1 .

Любой критерий для различения этих гипотез K имеет вполне определенную ошибку 1-го рода $\alpha(K) = P(H_d \neq H_0/H_0) = P_{F_0}(H_d \neq H_0)$. Но ошибка 2-го рода может быть вычислена только если известно конкретное распределение выборки (одна из альтернатив).

Будем рассматривать ошибку второго рода как функцию от H_1 : $\beta_{H_1}(K) = P_{H_1}(K = H_0)$.

Сначала зафиксируем H_1 .

Определение 13.6. Критерий K для проверки гипотезы H_0 против простой альтернативы H_1 называется **состоятельным**, если

$$\beta(K) = P_{H_1}(H_d(X) = H_0) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Эквивалентная формулировка этого определения в терминах мощности $M(K)$:

критерий K для проверки гипотезы H_0 против простой альтернативы H_1 называется состоятельным, если

$$M(K) = P_{H_1}(H_d(X) = H_1) \rightarrow 1 \text{ при } n \rightarrow \infty.$$

Теперь представим сложную альтернативу как объединение простых альтернатив $H_1 \neq H_0$ по всем возможным H_1 .

Определение 13.7. Критерий K для проверки гипотезы H_0 против сложной альтернативы H_1 называется **состоятельным**, если для любой простой альтернативы H_1

$$\beta_{H_1}(K) = P_{H_1}(H_d(X) = H_0) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Пример 13.4. В условиях примеров 13.1–13.3 проверить состоятельность критерия.

◀ В примерах рассматривается критерий со статистикой

$$Z = \frac{\bar{X} - a_0}{\sigma} \sqrt{n},$$

для проверки гипотезы $H_0 : a = 1$ против простой альтернативы $H_1 : a = 2$, в нормальной модели $\langle N(\theta, 2) \rangle$ по выборке объема n при уровне значимости $\alpha = 0,05$. Как найдено выше,

$$M(K) = \Phi \left(u_\alpha + \frac{(a_1 - a_0)\sqrt{n}}{\sigma} \right) = \Phi \left(-1,64 + \frac{\sqrt{n}}{2} \right).$$

$$M(K) \rightarrow 1 \text{ при } n \rightarrow \infty.$$

Следовательно, критерий состоятелен. ►

13.4. Рандомизированные критерии

Критерий Неймана – Пирсона в случае дискретных распределений

Соответствующие рассуждения можно провести и для дискретных распределений. Однако в силу дискретности распределения выборки при попытке разрешить неравенство

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c_\alpha$$

мы можем столкнуться с тем, что нельзя получить точное значение α за счет выбора границы c_α в неравенстве

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c_\alpha$$

(поскольку случайная величина изменяется скачкообразно и возможно, что включив в V очередную точку, мы еще не достигнем уровня α , а включив следующую — превзойдем его). В этом случае можно несколько изменить уровень значимости, а если желательно иметь вероятность ошибки первого рода, равной точно α , то надо использовать на пограничном множестве **рандомизированный критерий**.

Определение 13.8. Пусть имеются гипотезы H_0, \dots, H_k . **Рандомизированным статистическим критерием** $\pi = \pi(X)$ называется измеримая функция $\pi(X)$, сопоставляющая выборке X набор вероятностей гипотез:

$$\pi(X) = \{\pi_0(X), \dots, \pi_k(X)\}, \pi_i(X) \in [0, 1], \sum_i \pi_i(X) = 1.$$

При каждом X функция $\pi(X)$ указывает вероятности $\pi_i(X)$ принятия гипотез H_i , $i = \overline{0, k}$. Далее надо промоделировать случайную величину со значениями, равными номерам гипотез, и вероятностями $\pi_i(X)$. Принимается гипотеза, номер которой получен в результате моделирования. В случае двух гипотез для рандомизированного критерия достаточно задать одну функцию (будем задавать вероятность $\pi_1(X)$ принятия альтернативы H_1 .) Обычный статистический критерий есть частный случай рандомизированного, когда все $\pi_i(X)$ равны 0 и лишь одно равно 1. Такие критерии называются **нерандомизированными**.

Пример 13.5. Пусть по одному наблюдению x надо проверить гипотезу H_0 о том, что выборка взята из равномерного распределения $R[0, 1]$ при альтернативной гипотезе $H_1 : R[1, 2]$. Критерий имеет вид

$$K(x) = \begin{cases} \alpha, & x \in [0, 1], \\ 1, & x \in [1, 2]; \end{cases}$$

◀ Это рандомизированный критерий, функция $\pi_1(x) = K(x)$, $\pi_0(x) = 1 - K(x)$. Размер этого критерия равен вероятности $\pi_1(X)$ принятия альтернативы H_1 в случае, когда верна H_0 , то есть когда $x \in [0, 1]$. По условию эта вероятность равна α . Мощность критерия равна вероятности $\pi_1(x)$ принятия альтернативы H_1 в случае, когда верна H_1 , то есть когда $x \in [1, 2]$, значит, мощность равна 1. ►

Пример 13.6. Рассмотрим второй критерий для проверки той же гипотезы:

$$G(x) = \begin{cases} 0, & x \in [\alpha, 1], \\ 1, & x \in [0, \alpha] \cup [1, 2]. \end{cases}$$

◀ По-прежнему функция $\pi_1(x) = K(x)$, $\pi_0(x) = 1 - K(x)$, но это нерандомизированный критерий, так как функции принимают лишь значения 0 и 1. Вероятность $\pi_1(x)$ в случае, когда верна H_0 , равна $P_{H_0}(x \in [0, \alpha]) = P_{x \in [0, 1]}(x \in [0, \alpha]) = \alpha$, это размер критерия. Мощность критерия равна 1. Таким образом, оба критерия имеют одинаковый размер α и наибольшую возможную мощность, то есть являются н.м.к. ►

Замечание 13.2. Поскольку критерий $K(x)$ полностью определяется функцией $\pi(x)$, он часто и обозначается $\pi(x)$. Напомним, что в случае двух гипотез $\pi(x)$ означает вероятность $\pi_1(X)$ принятия альтернативы H_1 .

Рандомизированный критерий отношения правдоподобия в теореме Неймана — Пирсона имеет вид:

$$\pi(X) = \begin{cases} 0, & \frac{L(x, \theta_1)}{L(x, \theta_0)} < c, \\ 1, & \frac{L(x, \theta_1)}{L(x, \theta_0)} > c, \\ p, & \frac{L(x, \theta_1)}{L(x, \theta_0)} = c. \end{cases} \quad (110)$$

Если отношение $\frac{L(x, \theta_1)}{L(x, \theta_0)}$ строго меньше или строго больше c , то принимаются гипотезы H_0 или H_1 соответственно. Если же это отношение равно c , то гипотеза H_1 принимается с вероятностью p .

Уровень значимости (размер критерия)

$$\alpha(\pi) = P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) + p \cdot P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right).$$

Вероятности, как обычно для ошибки первого рода, рассчитываются при распределении H_0 . Заметим, что данное выражение представляет из себя математическое ожидание $\pi(X)$:

$$\alpha(\pi) = 1 \cdot P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) + p \cdot P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right) = \mathbf{E}(\pi(X)).$$

Требуемый уровень значимости α , c и p связаны уравнением

$$P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) + p \cdot P\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right) = \alpha.$$

Пусть

$$P_{H_0}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) = \alpha_0; \quad P_{H_0}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right) = p_0.$$

Тогда

$$\alpha_0 + p \cdot p_0 = \alpha, \quad p = \frac{\alpha - \alpha_0}{p_0}. \quad (111)$$

Найдем мощность этого критерия (вероятности рассчитываются при распределении H_1):

$$M(K) = 1 - \beta = P_{H_1}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) + p \cdot P_{H_1}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right).$$

Если обозначить

$$P_{H_1}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} > c\right) = \alpha_1; \quad P_{H_1}\left(\frac{L(x, \theta_1)}{L(x, \theta_0)} = c\right) = p_1,$$

то

$$M(K) = \alpha_1 + p \cdot p_1,$$

или, с учетом (111),

$$M(K) = \alpha_1 + \frac{(\alpha - \alpha_0)p_1}{p_0}.$$

Этот критерий является наиболее мощным [2].

13.5. Байесовские и минимаксные критерии

В теории проверки гипотез так же, как и в теории оценок (6.5), применяются **байесовский** и **минимаксный** подходы. Напомним, что при первом подходе сравнивают средние некоторой величины, а при втором ее максимальные значения. Упомянутая величина в оценивании представляет собой среднеквадратическое отклонение от параметра, а в теории проверки гипотез — вероятность ошибки. Как известно, вероятностью ошибки i -го рода называется число

$$\alpha_i(K) = P(H_d(X) \neq H_i / H_i \text{ верна}) = P_{H_i}(K(X) \neq H_i).$$

Байесовский подход предполагает, что распределение, из которого извлечена выборка X , было выбрано случайно. Гипотезы H_1, \dots, H_k являются случайными событиями; обозначим вероятности (априорные) этих событий через $P(H_j)$. По формуле полной вероятности найдем вероятность ошибки:

$$\alpha_K = \sum_j P(H_j) \alpha_j(K).$$

Определение 13.9. Критерий K называют **байесовским** критерием, если он минимизирует вероятность ошибки α_K .

Таким образом, байесовский критерий имеет наименьшую среднюю ошибку среди всех критериев.

Этот подход можно применять также, если известны не вероятности гипотез, а потери от ошибочного решения при каждой гипотезе. Пусть величина R_j составляет потерю в случае $(K(X) \neq H_j) / (H_j \text{ верна})$. Тогда математическое ожидание риска

$$R_K = \sum_j R_j \alpha_j(K).$$

В данном случае критерий K называют байесовским, если он минимизирует средний риск R_K .

Пример 13.7. Пусть в нормальной модели $\langle N(\theta, 1) \rangle$ по одному наблюдению проверяется гипотеза $H_0 : \theta = a_0 = 1$, против гипотезы $H_1 : \theta = a_1 = 4$, причем априорные вероятности гипотез равны. Предлагается критерий

$$K(X) = \begin{cases} H_1, & X \geq d, \\ H_0, & X < d. \end{cases} \quad (112)$$

При каком d вероятность ошибки критерия будет минимальной?

◀ Найдем полную вероятность ошибки критерия.

$$\begin{aligned} \alpha_K &= \sum_j P(H_j) \alpha_j(K) = 1/2 \alpha + 1/2 \beta = \\ &= 1/2 (P_{H_0}(X \geq d) + P_{H_1}(X < d)) = 1/2 (1 - \Phi_{1,1}(d) + \Phi_{4,1}(d)) = \\ &= 1/2 \left(1 - \Phi \left(\frac{d-1}{1} \right) + \Phi \left(\frac{d-4}{1} \right) \right) = 1/2 \left(1 - \int_{d-4}^{d-1} \varphi(t) dt \right). \end{aligned}$$

Вычитаемый интеграл достигает максимального значения при симметричных относительно 0 пределах, то есть при $4 - d = d - 1$, откуда $d = 2,5$. ▶

Определение 13.10. Критерий K называют **минимаксным**, если $\max_i \{\alpha_i(K)\} \leq \max_i \{\alpha(K')\}$ для любого критерия K' .

Пример 13.8. В условиях примера 13.7 при каком d будет минимален максимум $\{\alpha, \beta\}$?

◀ Посмотрим на рис. 7, на котором изображены плотности, соответствующие гипотезам, и вероятности ошибок 1-го и 2-го рода критерия $K(X)$.

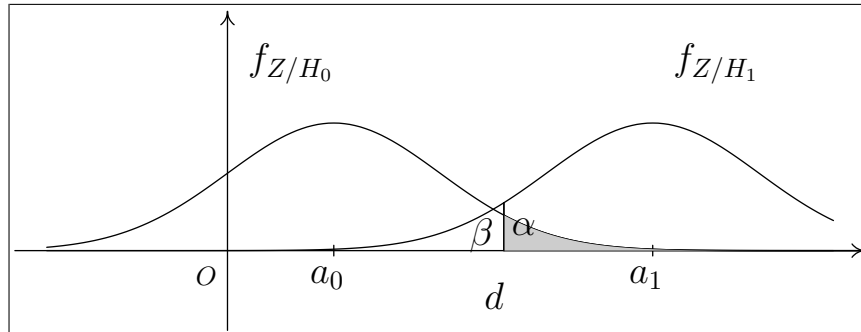


Рис. 7. Связь между ошибками первого и второго рода

$$\alpha = P_{H_0}(K(X) = H_1) = P_{H_0}(X \geq d), \quad \beta = P_{H_1}(K(X) = H_0) = P_{H_1}(X < d).$$

Будем передвигать границу критической области d . При этом α и β изменяются непропорционально, поскольку рассчитываются исходя из разных распределений, но тем не менее видно, что при $d \rightarrow \infty$ вероятность ошибки 1-го рода $\alpha \rightarrow 0$, но вероятность ошибки 2-го рода $\beta \rightarrow 1$. Если же $d \rightarrow -\infty$, то $\beta \rightarrow 0$, но $\alpha \rightarrow 1$. Максимум (α, β) будет минимален, когда они будут равны. Это достигается, когда граница критической области занимает среднее положение при $d = 2, 5$.►

Минимаксный критерий имеет самую маленькую максимальную ошибку среди всех критериев.

Справедлива следующая теорема³⁵.

Теорема 13.2 (лемма Неймана – Пирсона). Пусть проверяется простая гипотеза H_0 против простой гипотезы H_1 . Существуют постоянные s и r , при которых критерий отношения правдоподобия $\pi(X)$ (110) является

1) минимаксным критерием; числа s и r следует выбрать так, чтобы вероятности ошибок первого и второго рода были одинаковы;

2) байесовским критерием при заданных априорных вероятностях p_1, p_2 ; число r может быть любым, а s выбирается равным отношению p_1/p_2 ;

3) для любого $0 < \alpha_0 < 1$ наиболее мощным критерием уровня значимости α_0 ; числа s и r должны быть выбраны так, чтобы уровень значимости равнялся α_0 .

Пример 13.9. Рассмотрим критерий, используемый в примерах 13.7, 13.8:

$$\langle N(\theta, 1) \rangle, \quad H_0 : \theta = a_0 = 1, \quad H_1 : \theta = a_1 = 4, \quad n = 1;$$

$$K(X) = \begin{cases} H_1, & X \geq d, \\ H_0, & X < d. \end{cases}$$

◄ Согласно (109) данный критерий эквивалентен нерандомизированному КОП при $n = 1$ (конечно, с другой константой c):

$$K(x) = \begin{cases} 1, & x \geq d, \\ 0, & x < d. \end{cases} \quad \sim \quad \pi(x) = \begin{cases} 1, & \frac{L(x, \theta_1)}{L(x, \theta_0)} \geq c, \\ 0, & \frac{L(x, \theta_1)}{L(x, \theta_0)} < c. \end{cases}$$

³⁵ Именно этой теореме посвящен эпиграф к данной главе. С ее доказательством можно познакомиться в учебниках [1], [2].

Установим связь между константами c и d :

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - a_1)^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - a_0)^2}{2}}} = e^{-\frac{1}{2}(a_1^2 - a_0^2 - 2(a_1 - a_0)x)} \geq c.$$

$$-\frac{1}{2}(a_1^2 - a_0^2 - 2(a_1 - a_0)x) \geq \ln c.$$

$$x \geq \left(\ln c + \frac{a_1^2 - a_0^2}{2} \right) \frac{1}{a_1 - a_0} = d.$$

Значению $c = 1$ соответствует $d = \frac{a_1 + a_0}{2} = 2,5$. По лемме критерий будет байесовским при $c = p_1/p_2$. В условиях примера 13.7 априорные вероятности гипотез равны $1/2$ и $c = 1$. И действительно, мы получили в примере 13.7, что полная вероятность ошибки критерия минимальна при $d = 2,5$, что соответствует $c = 1$. Пример 13.4 подтверждает, что при $c = 1$ минимален максимум $\{\alpha, \beta\}$, то есть критерий будет минимаксным. ►

Итак, для двух простых гипотез $H_0: \theta = \theta_0$ против $H_1: \theta = \theta_1$ существует наиболее мощный критерий (при выполнении условий теоремы) и им является КОП.

Рассмотрим теперь случай, когда нулевая гипотеза простая, а альтернативная сложная односторонняя, например, $H_1: \theta > \theta_0$. Из примера 13.1 видно, что в нормальной модели критическая область не зависит от конкретного значения $a_1 = \theta_1$, то есть критерий максимизирует мощность при любом $\theta_1 \in \Theta_1$ и поэтому является р.н.м.к. Очевидно, то же будет верно при сложной альтернативе $H_1: \theta > \theta_0$. Это верно и для некоторых других моделей; выясним достаточные условия.

Если $T(x)$ — достаточная статистика, то

$$L(x, \theta) = g(T(x), \theta) \cdot h(x)$$

и отношение правдоподобия имеет вид

$$\frac{g(T(x), \theta_1)}{g(T(x), \theta_0)}.$$

Определение 13.11. Модель имеет **монотонное отношение правдоподобия**, если отношение правдоподобия $l(x) = g(T(x), \theta_1)/g(T(x), \theta_0)$ является монотонной функцией достаточной статистики $T(x)$.

Теорема 13.3. Для моделей с монотонным отношением правдоподобия и гипотезы $H_0: \theta = \theta_0$ против односторонней альтернативы $\theta_1 \in \Theta_1$ существует р.н.м.к., совпадающий с критерием Неймана – Пирсона для проверки гипотезы H_0 против произвольной фиксированной альтернативы из Θ_1 .

Доказательство. Пусть для определенности $H_1: \theta > \theta_0$, а отношение правдоподобия – возрастающая функция $T(x)$. Тогда неравенство $l(x) \geq c$ эквивалентно неравенству $T(x) \geq b_\alpha$, причем граница b_α определяется заданным уровнем значимости α и не зависит от конкретной альтернативы. Поэтому критерий Неймана – Пирсона одновременно является р.н.м.к. при сложной односторонней альтернативе.

Теорема 13.4. Если модель принадлежит экспоненциальному семейству с плотностью, представимой в виде $f(x, \theta) = e^{A(x) \cdot B(\theta) + C(x) + D(\theta)}$, и функция $B(\theta)$ строго монотонна, то критерий Неймана – Пирсона является р.н.м.к.

Доказательство. Статистика $T(x) = \sum_{i=1}^n A(x_i)$ является достаточной статистикой. Отношение правдоподобия

$$l(x) = e^{T(x) \cdot (B(\theta_1) - B(\theta_0)) + n(D(\theta_1) - D(\theta_0))},$$

если $B(\theta)$ монотонна, то это модель с монотонным отношением правдоподобия. \square

Иначе обстоит дело при двусторонней альтернативе $\theta \neq \theta_0$. Здесь теорема Неймана – Пирсона дает различные односторонние оптимальные критерии и р.н.м.к. не существует. Потребуем, чтобы критерий размера α имел мощность не меньше α ; такой критерий называется **несмещенным**. Задачу построения оптимального критерия иногда удается решить в классе несмещенных критериев. Объединим две соответствующие односторонние критические области размером $\alpha/2$ и таким образом получим критерий размера α .

Пример 13.10. Пусть в нормальной модели проверяется $H_0: a = \theta = a_0$ (против $\theta \neq a_0$).

◀ Статистикой критерия является

$$Z = \frac{\bar{X} - a_0}{\sigma} \sqrt{n},$$

а критическая область представляет из себя объединение левосторонней и правосторонней областей:.

$$V = \left\{ X : \frac{\bar{X} - a_0}{\sigma} \sqrt{n} \leq u_{\alpha/2} \right\} \cup \left\{ X : \frac{\bar{X} - a_0}{\sigma} \sqrt{n} \geq u_{1-\alpha/2} \right\}.$$

Этот критерий имеет мощность не меньше α , то есть является несмещенным. ► На рис. 8 показан график мощности такого критерия.

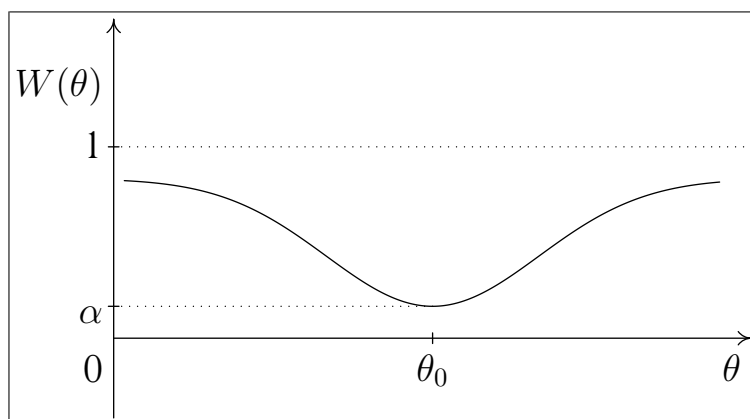


Рис. 8. Мощность двустороннего критерия

Критерий обладает наибольшей мощностью среди всех несмещенных критериев уровня значимости α , то есть является р.н.м. несмещенным критерием.

13.6. Контрольные вопросы

1. Дайте определение наиболее мощного (в некотором классе) критерия (н.м.к.).
2. Дайте определение р.н.м.к.
3. Дайте определение наилучшей критической области.
4. Сформулируйте теорему Неймана – Пирсона.
5. Запишите критерий отношения правдоподобия.
6. Каковы особенности применения теоремы Неймана – Пирсона к дискретным распределениям?

7. *Дайте определение рандомизированного статистического критерия.*
8. *Приведите примеры рандомизированного и нерандомизированного статистического критерия.*
9. *В чем состоит минимаксный подход в теории проверки гипотез?*
10. *В чем состоит байесовский подход в теории проверки гипотез?*
11. *Дайте определение минимаксного критерия.*
12. *Дайте определение байесовского критерия.*
13. *Сформулируйте лемму Неймана – Пирсона.*
14. *Какой критерий называется несмещенным?*
15. *Дайте определение модели с монотонным отношением правдоподобия.*
16. *Приведите достаточные условия существования р.н.м.к.*

Лекция 14. Критерии проверки гипотез о параметрах нормального распределения

*Каждый уверен, что все распределения — нормальные.
Экспериментаторы — потому, что они думают,
что это математическая теорема,
математики — потому, что они думают,
что это экспериментальный факт.*
Фольклор

План лекции: схема проверки параметрических гипотез, гипотезы о параметрах одного распределения, гипотезы о параметрах двух независимых распределений, гипотеза о средних для парных совокупностей, примеры.

14.1. Алгоритм проверки параметрических гипотез

Для проверки параметрических гипотез (не обязательно в нормальной модели) можно применять следующий алгоритм.

1. Сформулировать статистическую параметрическую модель, нулевую и альтернативную гипотезы, задать уровень значимости α .
2. Выбрать статистику $Z(x)$, такую, что она сама зависит от параметра θ , а ее распределение при верной H_0 от θ не зависит и различается при H_0 и при H_1 .
3. Найти критическую область V .
4. Рассчитать по выборке значение статистики Z_B .
5. Если Z_B попадает в критическую область V , то нулевая гипотеза отвергается (в пользу альтернативной). Если Z_B не попадает в критическую область V , то нулевая гипотеза не отвергается.
6. Сформулировать ответ в терминах вопроса.

Замечание 14.1. Гипотеза H_0 отвергается или не отвергается с уровнем значимости α .

Возникают вопросы: 1) как выбрать статистику Z и 2) как найти критическую область. Подчеркнем, что ответ на первый вопрос зависит от H_0 , а ответ на вопрос о форме критической области зависит от альтернативной гипотезы H_1 .

Есть по крайней мере три доступных способа получения ответов.

1. В [27] подробно расписано, какая статистика берется для каждой H_0 и какая критическая область берется для каждой H_1 . Эта информация вкратце приведена и ниже.

2. Если применить теорему Неймана — Пирсона и найти для данной H_1 НКО — получится доказательство этих результатов.

3. Наконец, можно выбрать статистику, просто исходя из здравого смысла — в качестве статистики берем функцию Z , зависящую от параметра и его оценки, распределение которой нам известно и чувствительно к тому, верна H_0 или нет. Критическая область должна быть смещена в направлении распределения Z при H_1 .

14.2. Гипотезы о параметрах одного распределения

Пусть наблюдаемая случайная величина принадлежит классу нормальных распределений $N(\theta_1, \theta_2)$. Перечислим критерии проверки гипотез о параметрах нормального распределения.

Замечание 14.2. В данной таблице индекс «0» внизу означает, что значение параметра известно. Например, $N(a, \sigma_0)$ — дисперсия известна, а математическое ожидание неизвестно. Индекс «0» вверху означает предполагаемое значение параметра. В последнем столбце указывается распределение Z при условии, что верна H_0 ($Z_{/H_0}$).

Пусть наблюдаемая случайная величина принадлежит классу нормальных распределений $N(\theta_1, \theta_2) \sim N(a, \sigma)$.

Рассмотрим выборку $X = \{(x_1, \dots, x_n)\}$, элементы которой независимы и имеют распределение $N(a, \sigma)$.

Гипотеза о дисперсии. $H_0 : \sigma = \sigma^0$.

Статистическая модель	Статистика Z	$Z_{/H_0}$
$\langle N(a_0, \sigma) \rangle$	$\frac{\sum (x_i - a_0)^2}{(\sigma^0)^2}$	χ_n^2
$\langle N(a, \sigma) \rangle$	$\frac{nS^2}{(\sigma^0)^2}$	χ_{n-1}^2

Гипотеза о среднем. $H_0 : a = a^0$

Статистическая модель	Статистика Z	$Z_{/H_0}$
$\langle N(a, \sigma_0) \rangle$	$\frac{(\bar{x}-a^0)\sqrt{n}}{\sigma_0}$	$N(0, 1)$
$\langle N(a, \sigma) \rangle$	$\frac{(\bar{x}-a^0)\sqrt{n}}{\bar{s}}$	T_{n-1}

14.3. Гипотезы о параметрах двух независимых распределений

Рассматриваются две независимые случайные величины и выборки наблюдений этих величин X, Y :

$$X = \{(x_1, \dots, x_{n_1})\} \in N(a_1, \sigma_1), Y = \{(y_1, \dots, y_{n_2})\} \in N(a_2, \sigma_2).$$

Гипотеза о дисперсии. $H_0 : \sigma_1 = \sigma_2$.

Статистическая модель	Статистика Z	$Z_{/H_0}$
$\langle N(a_0, \sigma) \rangle, \langle N(a, \sigma) \rangle$	$\frac{(\bar{s}_1)^2}{(\bar{s}_2)^2}, \quad s_1 > s_2.$	F_{n_1-1, n_2-1}

Замечание 14.3. Этот критерий называется **критерием Фишера**.

Гипотеза о средних. $H_0 : a_1 = a_2$

Модель	Статистика Z	$Z_{/H_0}$
$\langle N(a, \sigma_0) \rangle$ (известны σ_1, σ_2)	$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0, 1)$
$\langle N(a, \sigma) \rangle$ (σ_1, σ_2 неизвестны, но гипотеза $H_0: \sigma_1 = \sigma_2$ не отвергается)	$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{s}_1^2(n_1-1) + \bar{s}_2^2(n_2-1)}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \checkmark$	$T_{n_1+n_2-2}$
$\langle N(a, \sigma) \rangle$ (σ_1, σ_2 неизвестны, и гипотеза $H_0 : \sigma_1 = \sigma_2$ отвергается)	$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\bar{s}_1^2}{n_1} + \frac{\bar{s}_2^2}{n_2}}}$	$T_\nu,$ где $\nu \approx \frac{\left(\frac{\bar{s}_1^2}{n_1} + \frac{\bar{s}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\bar{s}_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{\bar{s}_2^2}{n_2}\right)^2}{n_2+1}} - 2$

Замечание 14.4. Критерий, использующий статистику, отмеченную галочкой (\checkmark), называется **критерием Стьюдента**.

14.4. Гипотеза о средних для парных совокупностей

Рассматриваются две случайные величины и выборки наблюдений этих величин X, Y : $X = \{(x_1, \dots, x_n)\} \in N(a_1, \sigma_1), Y = \{(y_1, \dots, y_n)\} \in N(a_2, \sigma_2)$, причем X_i и Y_i связаны, например, через объект наблюдения. Перейдем к разности случайных величин; тогда $d = X - Y, d \in N(a_d, \sigma_d)$. Очевидно, $a_d = a_1 - a_2$. Обычно σ_d неизвестно. Гипотеза $H_0 : a_1 = a_2 \sim H_0 : a_d = 0$.

Модель	Статистика Z	$Z_{/H_0}$
$\langle N(a, \sigma) \rangle$	$\frac{\bar{d}\sqrt{n}}{\bar{s}_d} = \frac{\sum d_i \sqrt{n-1}}{n \sum d_i^2 - (\sum d_i)^2}$	T_{n-1}

Замечание 14.5. Этот критерий называется **критерием Стьюдента для парных выборок**.

14.5. Примеры

Пример 14.1 (проверка гипотезы о математическом ожидании). По выборке объема $n = 113$ проверить гипотезу о том, что выборка взята из нормального распределения с математическим ожиданием 10. Дисперсия известна и равна 57,56. Рассчитанное по выборке среднее $\bar{x} = 9,279$.

◀ 1. Пусть $\alpha = 0,05$.

$$\langle N(a, \sigma_0) \rangle, \sigma_0 = \sqrt{57,56}.$$

$$H_0 : a = 10; H_1 : a < 10^{36}.$$

2. В нашем примере параметр a , оценка параметра — \bar{X} , известно распределение функции

$$Z(x) = \frac{(\bar{X} - a)\sqrt{n}}{\sigma}.$$

3. Если верна $H_0 : a = 10$, то $Z(x) \in N(0, 1)$. Если же верна $H_1 : a = a_1 < 10$, то распределение $X \in N(a_1, \sigma)$ смещено влево относительно распределения X при нулевой гипотезе $N(10, \sigma)$ и, значит, влево смещено и $Z(x) = \frac{(\bar{X} - a)\sqrt{n}}{\sigma}$. Поэтому критическая область левосторонняя.

³⁶ Альтернативная гипотеза формулируется как $H_1 : a < 10$, поскольку $\bar{x} = 9,279 < 10$, и у нас возникает вопрос — не будет ли математическое ожидание меньше 10.

Замечание 14.6. Нельзя рассуждать так: знак $<$ означает, что область левосторонняя. Можно было взять для критерия статистику $Z_1(x) = \frac{(a-\bar{X})\sqrt{n}}{\sigma}$, и при том же знаке $<$ область была бы правосторонняя. Лучше рассуждать так: при H_1 значения $Z(x)$ меньше, поэтому область левосторонняя.

Итак, критическая область

$$V = \{x : Z(x) \leq u_{0,05}\},$$

то есть

$$V = \{x : Z(x) \leq -1,64\}.$$

4.

$$Z(x) = \frac{(\bar{X} - a)\sqrt{n}}{\sigma}.$$

$$Z_B = \frac{(9,279 - 10)\sqrt{113}}{\sqrt{57,76}} \approx -1,01.$$

5. Z_B не попадает в критическую область V , поскольку $-1,01 > -1,64$. Следовательно, с уровнем значимости $\alpha = 0,05$ нулевая гипотеза H_0 не отвергается.

6. Вывод: данная выборка может быть взятой из совокупности с математическим ожиданием 10. ►

В примере (14.1) статистику и критическую область выбирали «на глаз» (третий способ). В следующем примере это будет сделано с помощью теоремы Неймана – Пирсона.

Пример 14.2 (проверка гипотезы о дисперсии). По выборке объема $n = 100$ проверить гипотезу о том, что выборка взята из нормального распределения с дисперсией 2. Математическое ожидание известно и равно -1 , выборочная дисперсия оказалась меньше 2.

◀ 1. $\langle N(a_0, \sigma) \rangle$, $a_0 = -1$, $\alpha = 0,05$.

Гипотезы: $H_0 : \sigma = \sigma_0 = \sqrt{2} \approx 1,4142$; $H_1 : \sigma = \sigma_1 < 1,4142$.

2-3. По теореме Неймана – Пирсона найдем НКО:

$$\frac{L(x, \sigma_1)}{L(x, \sigma_0)} \geq c_\alpha,$$

$$\frac{L(x, \sigma_1)}{L(x, \sigma_0)} = \frac{\left(\frac{1}{\sigma_1\sqrt{2\pi}}\right)^n \exp^{-\sum_{i=1}^{100} \frac{(x_i - a)^2}{2\sigma_1^2}}}{\left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp^{-\sum_{i=1}^{100} \frac{(x_i - a)^2}{2\sigma_0^2}}} \geq c_\alpha,$$

$$\exp^{-\sum_{i=1}^{100} \frac{(x_i - a)^2}{2\sigma_1^2} + \sum_{i=1}^{100} \frac{(x_i - a)^2}{2\sigma_0^2}} \geq c_1,$$

$$-\sum_{i=1}^{100} \left[(x_i - a)^2 \left(\frac{1}{2\sigma_1} - \frac{1}{2\sigma_0} \right) \right] \geq c_2.$$

Так как рассматривается случай $\sigma_1 < \sigma_0$, значит, $\frac{1}{2\sigma_1} - \frac{1}{2\sigma_0} > 0$. Тогда

$$-\sum_{i=1}^{100} (x_i - a)^2 \geq c_3,$$

$$\sum_{i=1}^{100} (x_i - a)^2 \leq c_4.$$

Таким образом, статистика должна включать функцию

$$\sum_{i=1}^{100} (x_i - a)^2$$

и область левосторонняя. Перейдем к стандартной статистике:

$$(x_i - a) \in N(0, \sigma_0),$$

$$\frac{(x_i - a)}{\sigma_0} \in N(0, 1),$$

$$Z = \frac{1}{\sigma_0^2} \sum_{i=1}^{100} (x_i - a)^2 \in \chi_{100}^2$$

(это справедливо при условии, что верна H_0).

$$V_k = \{Z < \chi_{0,05;100}^2\}.$$

Из таблиц находим

$$\chi_{0,05;100}^2 = 77,929,$$

значит,

$$V_k = \{Z < 77,929\}$$

4. Подсчитаем

$$Z_B = \frac{1}{\sigma_0^2} \sum_{i=1}^{100} (x_i - a)^2 = \frac{1}{2} \sum_{i=1}^{100} (x_i + 1)^2 = 88,2482255^{37}.$$

5. Следовательно, $z_B \notin V_k$, а значит, гипотеза H_0 не отвергается.

6. Можно считать, что дисперсия равна 2. ►

³⁷ Значение найдено по выборке, которая здесь не приводится.

14.6. Контрольные вопросы

1. Даны количества покупателей в день для первого и второго магазина $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_m)$, которые можно рассматривать как выборки из генеральных совокупностей числа покупателей. Сформулируйте нулевую и альтернативную гипотезы для получения ответа на вопрос: «Больше ли покупателей в первом магазине?»
2. Даны количества покупателей в день для первого и второго магазина $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_m)$, которые можно рассматривать как выборки из генеральных совокупностей числа покупателей. Сформулируйте нулевую и альтернативную гипотезы для получения ответа на вопрос: «Одинакова ли посещаемость магазинов покупателями?»
3. Даны количества покупателей в день для первого и второго магазина $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_m)$, которые можно рассматривать как выборки из генеральных совокупностей числа покупателей. Сформулируйте нулевую и альтернативную гипотезы для получения ответа на вопрос: «Можно ли считать, что колебания числа покупателей в этих двух магазинах примерно одинаковы?»
4. Какую гипотезу проверяет критерий Стьюдента?
5. Какую гипотезу проверяет критерий Фишера?
6. Какую гипотезу проверяет критерий Стьюдента для парных выборок?
7. Имеются данные о числе продаж товара и числе распространенных рекламных листовок в ряде городов. Сформулируйте нулевую и альтернативную гипотезы для проверки содержательной гипотезы: «Распространение листовок повышает продажи товара».

Лекция 15. Непараметрические критерии

*Есть правила для выбора решения,
но нет правила для выбора этих правил.
Правило о правилах*

План лекции: непараметрические критерии, критерий знаков, ранговый U -критерий.

15.1. Непараметрические критерии для проверки параметрических гипотез

Критерии, рассмотренные выше, получены с помощью теоремы Неймана — Пирсона и основаны на нормальном распределении генеральной совокупности (точном или приближенном). Для других статистических моделей также можно применить этот способ (исходя из других распределений). Но существуют критерии, которые не предполагают знания вида распределения генеральной совокупности и могут применяться в любой модели. Они называются **непараметрическими**. Непараметрические критерии основаны на статистиках, распределение которых не зависит от истинного распределения F (или не зависит от параметра θ , если $X \in F_\theta$). Для проверки гипотез о параметрах применяются, например, порядковые непараметрические критерии, которые строятся по статистикам вариационного ряда, не зависящим от конкретных значений членов вариационного ряда. Используется тот факт, что при H_0 все перестановки значений равновероятны, и с помощью комбинаторики подсчитывается вероятность конкретного расположения. Рассмотрим в качестве примера критерий знаков.

15.2. Критерий знаков

Пусть $X = \{(x_1, \dots, x_m)\}$, $Y = \{(y_1, \dots, y_m)\}$ — парные выборки. Гипотеза $H_0 : F_X = F_Y$ (выборки относятся к одной и той же генеральной совокупности). Содержательная интерпретация: некоторый показатель снимался с m объектов до (X) и после (Y) некоторого воздействия. H_0 : воздействие не повлияло на показатель.

Выбросим из выборки пары с равными значениями (x_i, y_i) ; пусть их осталось n . Тогда если H_0 справедлива, то

$$P(x_i - y_i > 0) = P(x_i - y_i < 0) = \frac{1}{2}.$$

Статистикой критерия является $Z = \min(n^-, n^+)$, где n^- — число отрицательных разностей $x_i - y_i$, n^+ — число положительных разностей $x_i - y_i$. Z имеет распределение $B(n, \frac{1}{2})$. Зададим α :

$$P(Z < z^*) = \sum_{k=1}^{z^*} C_n^k \left(\frac{1}{2}\right)^n = \alpha.$$

Можно решать это уравнение относительно z^* , а можно по данному Z_B найти $P(Z < Z_B)$. Если $P(Z < Z_B) \leq \alpha$, то H_0 отвергается с уровнем значимости α .

Пример 15.1. У группы студентов проверялись знания элементарной математики через полгода (X) и через полтора года (Y) после окончания школы (max = 60 баллов).

X_i	Y_i	Знак ($X_i - Y_i$)
40	30	+
35	30	+
41	38	+
55	21	+
46	49	—
60	53	+
51	37	+
43	21	+

◀Зададим $\alpha = 0,05$. H_0 : знания не изменились. Тогда

$$P(x_i - y_i < 0) = \frac{1}{2}.$$

$$Z_B = \min(1, 7) = 1.$$

$$P(Z < Z_B) = P(Z = 0 \cup Z = 1) = \left(\frac{1}{2}\right)^8 + C_8^1 \left(\frac{1}{2}\right)^8 \approx 0,035 < 0,05.$$

H_0 отвергается, знания ухудшились. ▶

Критерий знаков проверяет гипотезу однородности для парных выборок, являясь, таким образом, непараметрическим аналогом критерия Стьюдента для парных выборок. Это достаточно грубый критерий, поэтому его обычно применяют на стадии предварительного анализа. Пары (X_i, Y_i) , образующие выборки, могут быть составлены искусственно, но можно использовать только выборки X и Y одинакового объема.

Непараметрические аналоги имеются для каждого параметрического критерия.

Рассмотрим теперь непараметрический аналог критерия Стьюдента для независимых выборок.

15.3. Ранговый U -критерий

Ранговый U -критерий³⁸ был предложен в 1945 году Френком Уилкоксоном³⁹ (F. Wilcoxon). В 1947 году он был существенно переработан и расширен Х. Б. Манном (H. B. Mann) и Д. Р. Уитни (D. R. Whitney). Область применения теста хорошо отражена в названии статьи Манна и Уитни: «On a test whether one of two random variables is stochastically larger than the other»⁴⁰. Критерий Уилкоксона проверяет гипотезу H_0 о том, что две выборки извлечены из общей генеральной совокупности. В частности, он применим к проверке гипотезы о равенстве средних для независимых выборок. Данные должны быть числовыми или «полуколичественными» (так называют данные, которые не выражаются числами, но могут быть упорядочены, проранжированы.) Элементы первой выборки попарно сравниваются с элементами второй выборки и подсчитывается число инверсий. Алгоритм определяет, достаточно ли мала зона перекрещивающихся значений между двумя ранжированными рядами значений в первой и второй выборках. Чем меньше значение статистики, тем вероятнее, что различия между выборками достоверны. Распределение генеральной совокупности предполагается непрерывным (используется свойство рангов элементов выборки из непрерывного распределения — все возможные последовательности рангов равновероятны).

1. Алгоритм вычисления статистики U -критерия.

Составим единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака. Обозначим все элементы одной выборки буквой А, и все элементы другой выборки буквой В. Предположим, нам кажется, что в правой части объединенного ряда больше элементов А. Посчитаем для каждого элемента А число элементов В, которые расположены правее его, и сложим эти числа. Получим значение статистики $U = U_{n_1, n_2}$, которое будет сравниваться с табличным. При

³⁸ Другие названия: критерий Уилкоксона (англ. Wilcoxon test), U -критерий Манна — Уитни (англ. Mann — Whitney U-test), критерий Манна — Уитни — Уилкоксона (англ. Mann — Whitney — Wilcoxon, MWW), критерий суммы рангов Уилкоксона (англ. Wilcoxon rank-sum test) или критерий Уилкоксона — Манна — Уитни (англ. Wilcoxon — Mann — Whitney test). Часто пишут также Вилкоксон вместо Уилкоксона.

³⁹ Френк Уилкоксон (англ. Frank Wilcoxon; 1892–1965) — американский химик и статистик.

⁴⁰ Хорошо американцам — у них приняты простые выражения!

верной гипотезе H_0 случайная величина U_{n_1, n_2} распределена одинаково для всех непрерывных генеральных совокупностей. Гипотезу H_0 отвергают, если выборочное значение статистики U_{n_1, n_2} значительно отклоняется от ожидаемого. Критические значения содержатся в табл. 12.

Пример 15.2. Пусть объединенный ряд таков: ВАААААВВВВВА.

◀ В правой части ряда больше элементов В. Правее первого В расположены 6 элементов А, правее второго В — один элемент А, то же относится к остальным элементам В. Выборочное значение статистики U равно $6+1+1+1+1+1=11$. ▶

Для больших выборок удобнее другой метод нахождения значения статистики U .

2. Алгоритм вычисления статистики U -критерия.

Составим единый ранжированный ряд из обеих сопоставляемых выборок, в нем будет $N = n_1 + n_2$ элементов, где n_1 — количество элементов первой выборки, а n_2 — второй выборки. Припишем каждому значению в объединенном ряду ранги (по возрастанию). Подсчитаем отдельно сумму рангов, пришедшихся на долю элементов первой выборки, и отдельно — на долю элементов второй выборки. Определим бóльшую из двух ранговых сумм (S_x), объем выборки, которой соответствует эта сумма, обозначим n_x ($x = 1$ или 2). Найдем значение статистики U -критерия Манна — Уитни по формуле $U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - S_x$.

Пример 15.3. Вычислим значение статистики U для объединенного ряда из предыдущего примера: ВАААААВВВВВА.

◀ Расставим ранги 1(В), 2(А), 3(А), 4(А), 5(А), 6(А), 7(В), 8(В), 9(В), 10(В), 11(В), 12(А). Сумма рангов элементов выборки А равна $S_A = 2 + 3 + 4 + 5 + 6 + 12 = 32$. Сумма рангов элементов выборки В равна $S_B = 1 + 7 + 8 + 9 + 10 + 11 = 46$. Бóльшей из двух ранговых сумм является S_B , значит, $S_x = S_B = 46$. Выборка В содержит 6 элементов, поэтому $n_x = 6$.

$$U = 6 \cdot 6 + \frac{6 \cdot 7}{2} - 46 = 11.$$

▶

Значение статистики критерия, найденное любым из этих способов, сравнивается с критическим, которое определяется по таблице для данных n_1 и n_2 . Если полученное значение U меньше табличного или равно ему для избранного уровня статистической значимости, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза). Если же полученное значение

U больше табличного, принимается нулевая гипотеза. Достоверность различий тем выше, чем меньше значение U .

Пример 15.4. Проверим гипотезу о том, что выборки A и B с объединенным рядом из предыдущих примеров (ВАААААВВВВВВВВ) имеют одинаковое генеральное среднее (взяты из совокупностей с одинаковым математическим ожиданием).

◀ Критическое значение при $n_1 = 6$, $n_2 = 6$ равняется 5 для уровня значимости $\alpha = 0,05$. Таким образом, нулевая гипотеза отвергается. ▶

Распределение U для обоснования этого критерия может быть получено из соотношения между производящими функциями частот [20].

Ограничения применимости критерия: в каждой из выборок должно быть не менее трех значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти. Таблицы составлены только до $n = 60$, поэтому в каждой выборке должно быть не более 60 значений параметра. При больших n можно использовать нормальную аппроксимацию.

Недостатком всех непараметрических методов является меньшая (относительно параметрических) мощность. Однако критерий Манна — Уитни по сравнению с критерием Стьюдента оказывается лишь слегка менее мощным: если мы применяем эти два критерия к нормальным выборкам, асимптотическая эффективность критерия Манна — Уитни составляет относительно критерия Стьюдента $3/\pi \approx 0,95$ [30]. Преимуществом этого критерия (как и вообще непараметрических методов) является возможность применения к различным распределениям. Напомним, что критерий Стьюдента пригоден только для выборок из нормального распределения. Кроме того, ранговые критерии обладают устойчивостью к выбросам. Наличие небольшого числа выбросов (причем они могут быть очень большими) мало влияет на результат при использовании U -критерия, но весьма значительно при использовании критерия Стьюдента⁴¹.

15.4. Контрольные вопросы

1. Что такое непараметрический критерий?
2. Каким требованиям должны удовлетворять статистики, используемые в непараметрических критериях?

⁴¹ Свойство устойчивости, в частности, к выбросам, называют *робастностью* (от англ. robust — крепкий, грубый).

3. *Сформулируйте недостатки и преимущества непараметрических критериев относительно параметрических критериев.*
4. *Какую гипотезу проверяет критерий знаков?*
5. *Какую гипотезу проверяет критерий Уилкоксона?*
6. *Какую гипотезу проверяет критерий Уилкоксона для парных выборок?*
7. *Приведите пример парных выборок.*
8. *Приведите пример парных выборок с искусственно составленными парами.*
9. *Даны оценки одной и той же группы студентов за две контрольные работы. Какие критерии можно применить для получения ответа на вопрос: «Одинакова ли успеваемость?»*
10. *Даны оценки двух групп студентов за контрольную работу. Какие критерии можно применить для получения ответа на вопрос: «Одинакова ли успеваемость?»*
11. *Какие критерии называются робастными?*

Лекция 16. Критерии согласия

Нет ничего практичнее хорошей теории.

Роберт Кирхгоф

План лекции: построение критериев согласия, критерий согласия Колмогорова, критерий согласия χ^2 Пирсона, проверка гипотезы однородности: критерий Колмогорова — Смирнова, проверка гипотезы независимости: критерий χ^2 Пирсона.

Пусть сформулирована только одна гипотеза H_0 о виде распределения и требуется проверить, согласуются ли имеющиеся статистические данные с гипотезой или же они ее опровергают. Критерии, применяемые для проверки гипотез о виде распределения (простых или сложных) против сложных альтернатив с использованием статистик, характеризующих отклонение эмпирических данных от соответствующих гипотезе H_0 теоретических значений, называются **критериями согласия**. Как обычно, должны выполняться условия: распределение такой статистики в случае справедливости H_0 можно определить точно или приближенно и оно не должно зависеть от неизвестного нам распределения генеральной совокупности.

Пусть дана выборка $X = (X_1, \dots, X_n)$, описывающая результат изучаемого эксперимента. Предположим, $T = T(X)$ — статистика, удовлетворяющая вышеупомянутым условиям. Пусть G — множество всех значений статистики

$$G = \{t : t = T(x), x \in X\}. \quad (113)$$

Определим для малого $\alpha > 0$ подмножество $G \setminus V$, $V \subseteq G$ так, чтобы вероятность осуществления события $\{T(x) \in V\}$ в случае справедливости гипотезы H_0 удовлетворяла бы условию

$$P(T(x) \in V/H_0) = \alpha. \quad (114)$$

Если окажется, что $t \in V$, то в предположении справедливости гипотезы H_0 произошло маловероятное событие и эта гипотеза должна быть отвергнута как противоречащая статистическим данным. В противном случае нет основания отказываться от рассматриваемой гипотезы и следует считать, что наблюдения не противоречат гипотезе (согласуются с ней).

Как и при проверке параметрических гипотез, статистика $T(X)$, определенная выше, называется *статистикой критерия*, V — *критической областью критерия*, α — *уровнем значимости критерия* (вероятностью ошибочного отвержения гипотезы H_0 , когда она верна). Статистика $T(X)$ должна быть функцией отклонения эмпирического распределения от

теоретического. Требуется, чтобы эта функция сходилась к какому-то собственному распределению, если верна проверяемая гипотеза, и неограниченно возрастала, если гипотеза не верна. Гипотеза принимается или отвергается в зависимости от величины данной функции отклонения.

16.1. Построение критериев согласия

Критерий согласия K для проверки гипотезы $H_0 : F = F_0$ против сложной альтернативы $H_1 : F \neq F_0$ можно построить следующим образом:

$$K(X) = \begin{cases} H_0, & \text{если } |Z(X)| < C; \\ H_1, & \text{если } |Z(X)| \geq C. \end{cases}, \quad (115)$$

где $Z(X) = Z(X, F_0)$ — функция, характеризующая отклонение эмпирического распределения от предполагаемого F_0 . Если для данной выборки модуль функции отклонения принимает малые значения, то это свидетельствует в пользу нулевой гипотезы $H_0 : F = F_0$, если же отклонения большие, то в пользу альтернативной гипотезы $H_1 : F \neq F_0$.

Функция $Z(X)$, очевидно, является случайной величиной (как функция выборки X). Для нахождения вероятностей ошибок надо знать распределение этой функции в случаях, когда верны гипотезы H_0 и H_1 . Однако гипотеза H_1 является сложной и вероятность ошибки второго рода может быть вычислена только если известно конкретное распределение выборки $F_1 \neq F_0$ (одна из альтернатив). Выходом из положения будет использование статистики, сильно возрастающей при $n \rightarrow \infty$, когда гипотеза H_0 неверна. Если же гипотеза H_0 верна, то распределение $Z(X)$ должно быть известно точно или приближенно ($Z(X) \in G$ или $Z(X) \rightarrow \xi \in G$). Найдем ошибки этого критерия.

Ошибка первого рода $\alpha_K = P_{H_0}(|Z(X)| \geq C)$.

По условию эта вероятность равна или стремится к $P(|\xi| \geq C)$. Выберем постоянную C так, чтобы $P(|\xi| \geq C) = \alpha$ (мы можем это сделать), тогда ошибка первого рода равна α (или стремится к α).

Ошибка второго рода $\beta_K = P_{H_1}(|Z(X)| < C)$.

Поскольку мы выбрали $Z(X)$ так, что при верной гипотезе H_1 $Z(X) \xrightarrow{P} \infty$ при $n \rightarrow \infty$, то $\beta_K = P_{H_1}(|Z(X)| < C) \rightarrow 0$.

Следовательно, критерий согласия имеет уровень значимости α (точный или асимптотический⁴²) и является состоятельным.

⁴² Критерий K является критерием асимптотического уровня α , если $\alpha_K \rightarrow \alpha$ при $n \rightarrow \infty$.

16.2. Критерий согласия Колмогорова

Для построения этого критерия нам потребуются две теоремы.

Теорема 16.1 (теорема Колмогорова). *Если функция $F(x)$ непрерывна, то при любом фиксированном $t > 0$*

$$\lim_{n \rightarrow \infty} P(|\sqrt{n}D_n| \leq t) = K(t) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 t^2}, \quad (116)$$

где

$$D_n = D_n(X) = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|. \quad (117)$$

Функция $K(t)$ называется функцией Колмогорова.

Теорема 16.2 (теорема Гливенко – Кантелли). *Пусть $X = (X_1, \dots, X_n)$ – выборка объема n из неизвестного распределения с функцией распределения F , и F_n – эмпирическая функция распределения, построенная по этой выборке. Тогда*

$$\sup_{y \in R} |F_n(y) - F(y)| \xrightarrow{P} 0, \quad \text{при } n \rightarrow \infty.$$

Замечание 16.1. *Если функция распределения F непрерывна, то скорость сходимости к нулю в теореме Гливенко – Кантелли имеет порядок $\frac{1}{\sqrt{n}}$ (это видно из теоремы Колмогорова).*

Пусть $x = (x_1, \dots, x_n)$ – выборка из генеральной совокупности с неизвестной функцией непрерывного распределения $F(x)$. Выдвинута простая гипотеза $H_0 : F(x) = F_0(x)$, где $F_0(x)$ задана. Рассмотрим построение критерия согласия. В качестве статистики критерия выбирают следующую функцию отклонения эмпирического распределения от теоретического:

$$Z(X) = \sqrt{n}D_n = \sqrt{n} \sup_x |(F_n(x) - F(x))|. \quad (118)$$

Покажем, что $Z(X)$ удовлетворяет оговоренным ранее необходимым условиям: если гипотеза H_0 верна, то распределение $Z(X)$ должно быть известно точно или приближенно, а когда гипотеза H_0 неверна, $Z(X) \rightarrow \infty$ при $n \rightarrow \infty$.

Особенностью статистики D_n является тот факт, что ее распределение (при справедливости гипотезы H_0) не зависит от вида функции $F(x)$.

При $n \rightarrow \infty$, если H_0 — верная гипотеза, распределение статистики в соответствии с теоремой Колмогорова сходится к функции Колмогорова $K(t)$, таким образом, известно асимптотическое распределение.

Если гипотеза H_0 неверна, то есть $X_i \in F_1 \neq F_0$, то по теореме Гливенко — Кантелли $F_n(y) \xrightarrow{P} F_1(y)$ для любого y при $n \rightarrow \infty$. Поскольку $F_0 \neq F_1$, найдется y такое, что $|F_1(y) - F_0(y)| > 0$. Для таких y

$$|F_n(y) - F_0(y)| \xrightarrow{P} |F_1(y) - F_0(y)| > 0.$$

Поэтому при $n \rightarrow \infty$

$$Z(X) = \sqrt{n} \sup_y |F_n(y) - F_0(y)| \xrightarrow{P} \infty.$$

Итак, с помощью статистики $Z(X) = \sqrt{n}D_n$ можно построить критерий согласия

$$K(X) = \begin{cases} H_0, & \text{если } |Z(X)| < C; \\ H_1, & \text{если } |Z(X)| \geq C. \end{cases},$$

осталось рассчитать C .

Распределение Колмогорова табулировано, так что для заданного уровня значимости α с помощью таблиц для функции $K(t)$ можно найти $C = t^*$ такое, что $\alpha = P(\xi \geq t^*)$.

Критическая граница t^* находится из решения уравнения

$$K(t^*) = 1 - \alpha.$$

Так, $t^* = t_\alpha^* = 1,3581$ при $\alpha = 0,05$; $t_\alpha^* = 1,6276$ при $\alpha = 0,01$.

Мы построили критерий

$$K(X) = \begin{cases} H_0, & \text{если } \sqrt{n}D_n < t^*; \\ H_1, & \text{если } \sqrt{n}D_n \geq t^*. \end{cases}$$

Критическая область имеет вид $V = [t^*, \infty)$.

Таким образом, при заданном уровне значимости α правило проверки гипотезы H_0 сводится к следующему:

если значение статистики $\hat{t} = D_n(x) = \max_x |F_n^*(x) - F_0(x)|$ удовлетворяет неравенству

$$\sqrt{n} \cdot \hat{t} \geq t^*,$$

то H_0 отвергают, в противном случае делают вывод, что статистические данные не противоречат гипотезе.

При практических расчетах значения $K(t)$ можно применять уже при $n > 20$. Критерий согласия Колмогорова применяют для непрерывных функций распределения $F(x)$. В конкретных задачах величину α берут равной 0,01; 0,005; 0,001.

Пример 16.1. Рассмотрим использование критерия Колмогорова для проверки при $\alpha = 0,05$ гипотезы: выборка объема $n = 113$ взята из распределения $N(10; 76)$.

◀ Критерий применяется к негруппированной выборке. В первом столбце содержатся упорядоченные значения выборки x , во втором — эмпирическая функция распределения, равная $F_n^*(x) = \frac{\sum n_i}{n}$, в третьем — предполагаемая теоретическая функция распределения $F_0(x)$, равная в данном случае $\Phi_{10;76}(x) = \Phi(\frac{x-10}{\sqrt{76}})$. В четвертом столбце содержатся модули разностей теоретической и эмпирической функций распределения $|F_n^*(x) - F_0(x)|$, а в пятом — они же, упорядоченные по убыванию. Поэтому максимальное значение модуля разности, равное значению статистики \hat{t} — первый элемент пятого столбца. Приведена только часть таблицы, содержащей $n = 113$ строк.

x	$F_n^*(x)$	$F_0(x)$	$ F_n^*(x) - F_0(x) $	$\max F_n^*(x) - F_0(x) $
-175,2	1/113=0,00885	0,007408	0,001441481	0,076255
-158,002	2/113=0,017699	0,013534	0,004165511	0,076191
...
-17,7772	0,433628	0,357373	0,07625536	0,050072
...
162,7886	0,982301	0,977804	0,004496574	0,002554
163,7419	0,99115	0,978459	0,012691171	0,001441
177,3939	1	0,986187	0,013813193	0,000197

$$\hat{t} = \max |F_n^*(x) - F_0(x)| = 0,07625536$$

$$\sqrt{n} \cdot \hat{t} = 0,810605598.$$

Критическое значение $t^* = 1,358$. Поскольку $0,810605598 < 1,358$, гипотеза о распределении $N(10; 76)$ не отвергается. ►

16.3. Критерий согласия χ^2 Пирсона

Критерий согласия χ^2 Пирсона⁴³ также проверяет гипотезу $H_0 : F(x) = F_0(x)$, но его можно применять для любых распределений. Чтобы воспользоваться этим критерием, выборочные данные предварительно группируют. Пусть n_i — число значений, попавших в i -й интервал, $i =$

⁴³ Не запутайтесь в Пирсонах! Критерий согласия разработан Карлом Пирсоном, классиком математической статистики. А общую теорию проверки статистических гипотез совместно с Е.Нейманом создавал его сын, Эгон Пирсон.

$1, \dots, k, n$ — объем выборки, p_i — теоретическая вероятность попадания одного элемента выборки в i -й интервал. Однако в теоретическом распределении могут быть неизвестные параметры $(\theta_1, \dots, \theta_r)$, что обычно и встречается на практике. Тогда по выборке (x_1, \dots, x_n) первоначально находят оценки $(\theta_1^*, \dots, \theta_r^*)$ и затем по $F(x, \theta_1^*, \dots, \theta_r^*)$ вычисляют теоретические вероятности p_i . Если H_0 верна, то относительные частоты $\frac{n_i}{n}$ при больших n близки к p_i (теорема Бернулли). За меру расхождения между распределением выборки и теоретическим распределением целесообразно принять (в соответствии с методом наименьших квадратов) величину

$$\sum_{i=1}^k c_i \left(\frac{n_i}{n} - p_i \right)^2, \quad (119)$$

где c_i — коэффициенты. Пирсон показал [3], что если взять $c_i = \frac{n}{p_i}$ и если $(\theta_1^*, \dots, \theta_r^*)$ — оценки максимального правдоподобия, то полученная статистика критерия

$$Z = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \quad (120)$$

при $n \rightarrow \infty$ стремится к распределению χ^2 с $\nu = k - r - 1$ степенями свободы. На практике предельное распределение χ_{k-r-1}^2 можно использовать уже при $n > 50$ и $np_i \geq 4$. Критическую область задаем в виде $V = [t^*, \infty)$. Тогда

$$F_{\chi_{k-r-1}^2}(t^*) = 1 - \alpha. \quad (121)$$

Критическую границу находим из решения последнего уравнения. Таким образом, критерий согласия χ^2 имеет вид: если значение статистики $Z_B = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$ удовлетворяет неравенству

$$Z_B \geq t^*,$$

гипотезу H_0 отвергают, в противном случае гипотеза H_0 не противоречит условиям испытаний. Иначе говоря,

$$K(X) = \begin{cases} H_0, & \text{если } Z_B < t^*; \\ H_1, & \text{если } Z_B \geq t^*. \end{cases}$$

Пример 16.2. Рассмотрим использование критерия Пирсона для проверки гипотезы о нормальности на основе выборки объема $n = 113$ из примера (16.1) при $\alpha = 0,05$.

◀ Первый столбец табл. 16.3 содержит границы интервалов группировки; второй — нормированную переменную t_i , $t_i = \frac{y_i - \bar{x}}{s}$; третий — значение $\Phi(t_i)$; четвертый — вероятность p_i , $p_i = \Phi(t_{i+1}) - \Phi(t_i)$; пятый — теоретическую численность i -го интервала $n_i^T = np_i$ и т. д. В последнем столбце содержатся значения $Z_i = (n_i - n_i^T)^2 / n_i^T$. Выборочное значение статистики равно сумме элементов последнего столбца: $Z_B = \sum Z_i$.

1	2	3	4	5	6	7	8	9
y_i	t_i	$\Phi(t_i)$	p_i	n_i^T	n_i	$n_i - n_i^T$	$(n_i - n_i^T)^2$	Z_i
-175,2	-2,149	0,018	0,041	4,633	6	1,367	1,868689	0,403343
-131,126	-1,616	0,059	0,092	10,396	11	0,604	0,364816	0,035092
-87,052	-1,083	0,151	0,1576	17,809	18	0,1912	0,036557	0,002053
-42,977	-0,55	0,3086	0,2044	23,097	21	-2,0972	4,398248	0,190423
1,097	-0,017	0,513	0,201	22,713	24	1,287	1,656369	0,072926
45,171	0,516	0,714	0,149	16,837	12	-4,837	23,39657	1,389593
89,245	1,0495	0,863	0,085	9,605	13	3,395	11,52603	1,200003
133,320	1,583	0,948	0,037	4,181	8	3,819	14,58476	3,488343
177,394	2,1645	0,985						$\sum Z_i \approx 6,78$

Выборочное значение статистики $Z_B = \sum Z_i \approx 6,78$. Критическое значение: квантиль распределения хи-квадрат χ^2 с числом степеней свободы $\nu = k - r - 1$. При вычислении теоретических вероятностей p_i использовались оценки двух параметров (a и σ), поэтому $r = 2$ и $\nu = 8 - 2 - 1 = 5$. Квантиль χ_5^2 порядка 0,95 равна 11,07, то есть $t^* = 11,07$. Критическая область V — интервал $[11,07, +\infty)$. Выборочное значение статистики меньше критического, следовательно, гипотеза о нормальном распределении не отвергается. ▶

Пример 16.3. Число выпадений герба при 20 бросаниях двух монет распределилось следующим образом:

Количество гербов	0	1	2
Число подбрасываний n_i	4	8	8
p_i	0,25	0,5	0,25
np_i	5	10	5

◀ p_i — вероятность того, что при бросании двух монет выпало i гербов, $i = 0, 1, 2$.

$$\begin{aligned}
 Z_B &= \sum_{i=0}^2 \frac{(n_i - np_i)^2}{np_i} = \\
 &= \frac{(4 - 5)^2}{5} + \frac{(8 - 10)^2}{10} + \frac{(8 - 5)^2}{5} = 2,4.
 \end{aligned}$$

$\alpha = 0,05$.

$$t^* = \chi_{l-r-1;0,95}^2 = \chi_{3-0-1;0,95}^2 = \chi_{2;0,95}^2 = 5,99$$

(последнее значение нашли по таблице).

$$Z_B < t^*,$$

следовательно, гипотезу H_0 не отвергают.►

16.4. Проверка гипотезы однородности: критерий Колмогорова – Смирнова

Теорема 16.3 (теорема Смирнова). Пусть $F_n(x)$ и $F_m(x)$ – две эмпирические функции, построенные на основе двух независимых выборок объемов n и m из одного и того же распределения F_ξ , и

$$D_{n,m} = \sup_{-\infty < x < \infty} |F_n(x) - F_m(x)|. \quad (122)$$

Тогда если теоретическая функция $F(x)$ непрерывна, то для любого фиксированного $t > 0$

$$\lim_{n,m \rightarrow \infty} P(|\sqrt{nm/(n+m)} D_{n,m}| \leq t) = K(t), \quad (123)$$

где функция $K(t)$ определена равенством (116).

Эта теорема позволяет ответить на важный практический вопрос, можно ли считать, что две выборки получены из одного и того же распределения.

Есть две выборки $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$, причем $X_i \in F_x$, $Y_i \in F_y$, и распределения F_x, F_y , вообще говоря, неизвестны. Проверяется сложная гипотеза $H_0: F_x = F_y$ против альтернативы $H_1: F_x \neq F_y$.

Если F_x, F_y имеют непрерывные функции распределения, применим критерий Колмогорова – Смирнова.

Пусть $F_{n,x}$ и $F_{m,y}$ – эмпирические функции распределения, построенные по выборкам X и Y ,

$$Z(X, Y) = \sqrt{\frac{nm}{n+m}} \max_t |F_{n,x}(t) - F_{m,y}(t)|.$$

Если гипотеза H_0 верна, то $Z(X, Y) \Rightarrow \xi \in K$ (распределение, имеющее функцию распределения Колмогорова) при $n, m \rightarrow \infty$.

Замечание 16.2. Если верна гипотеза H_1 , то $Z(X, Y) \xrightarrow{P} \infty$ при $n, m \rightarrow \infty$.

Если случайная величина ξ имеет функцию распределения $K(y)$, то по заданному α найдем C такое, что $\alpha = P(\xi \geq C)$. Построен критерий согласия Колмогорова – Смирнова:

$$H_d = \begin{cases} H_0, & Z(X, Y) < C; \\ H_1, & Z(X, Y) \geq C. \end{cases}$$

Если число выборок больше двух, то для проверки гипотезы однородности можно пользоваться одним из вариантов критерия χ^2 Пирсона⁴⁴.

16.5. Проверка гипотезы независимости: критерий χ^2 Пирсона

*Не волнуйся, голова!
Теперь будет думать компьютер.
Гомер Симпсон*

По выборке $(X, Y) = ((X_1, Y_1), \dots, (X_n, Y_n))$ значений двух наблюдаемых совместно величин ξ и η в n экспериментах будем проверять гипотезу H_0 : ξ и η независимы.

Сгруппируем значения ξ в k интервалов, а значения η в m интервалов и подсчитаем эмпирические частоты для каждого интервала двумерной группировки ($i = 1, \dots, k, j = 1, \dots, m$).

(Здесь $\mu_{i,j}$ – число пар (X, Y) , попавших в клетку $\delta_i^x \times \delta_j^y$, и т. д.)

Y	δ_1^y	δ_2^y	\dots	δ_m^y	Σ
X					
δ_1^x	$\mu_{1,1}$	$\mu_{1,2}$	\dots	$\mu_{1,m}$	b_1
δ_2^x	$\mu_{2,1}$	$\mu_{2,2}$	\dots	$\mu_{2,m}$	b_2
\vdots			\dots		
δ_k^x	$\mu_{k,1}$	$\mu_{k,2}$	\dots	$\mu_{k,m}$	b_k
Σ	a_1	a_2	\dots	a_m	n

Если гипотеза H_0 верна, то теоретические вероятности попадания пары (X_1, Y_1) в любую из областей $\delta_i^x \times \delta_j^y$ равны произведению вероятностей:

$$p_{i,j} = P((X_1, Y_1) \in \delta_i^x \times \delta_j^y) = P(X_1 \in \delta_i^x)P(Y_1 \in \delta_j^y) = p_i^x p_j^y,$$

⁴⁴ Критерий Пирсона применяется для проверки разных гипотез – согласия, однородности, независимости.

и по ЗБЧ частоты должны быть близки к вероятностям:

$$\frac{b_i}{n} \approx p_i^x, \quad \frac{a_j}{n} \approx p_j^y, \quad \frac{\mu_{i,j}}{n} \approx p_{i,j}.$$

Статистикой критерия служит функция, учитывающая различия между $\mu_{i,j}$ и $\frac{b_i a_j}{n}$:

$$Z(X, Y) = n \sum_{i=1}^k \sum_{j=1}^m \frac{(\mu_{i,j} - (b_i a_j)/n)^2}{b_i a_j} = n \left(\sum_{i=1}^k \sum_{j=1}^m \frac{\mu_{i,j}^2}{b_i a_j} - 1 \right).$$

Поэтому значительная разница может служить основанием для отклонения гипотезы независимости. Если гипотеза H_0 верна, то $Z(X, Y) \Rightarrow \chi_{(k-1)(m-1)}^2$ при $n \rightarrow \infty$.

16.6. Контрольные вопросы

1. Что такое критерий согласия?
2. Изложите схему построения критерия согласия.
3. Можно ли проверить гипотезу о виде распределения с помощью параметрического критерия?
4. Дайте определение критерия асимптотического уровня α .
5. Какую гипотезу проверяет критерий согласия Пирсона?
6. Какая статистика используется в критерии согласия Пирсона?
7. Какую гипотезу проверяет критерий согласия Колмогорова?
8. Какая статистика используется в критерии согласия Колмогорова?
9. Каковы ограничения на статистическую модель и выборку для применения критериев согласия Пирсона и Колмогорова?
10. Сформулируйте теорему Колмогорова.
11. Сформулируйте теорему Гливенко — Кантелли.
12. Сформулируйте теорему Смирнова.
13. Какую гипотезу проверяет критерий Колмогорова — Смирнова?

14. Для каждого из трех текстов имеются данные о частотах, с которыми встречаются в тексте различные служебные слова и знаки препинания. Какой критерий можно применить для проверки содержательной гипотезы: «Эти три текста принадлежат одному автору»?
15. В условиях предыдущего вопроса какую статистику можно использовать для проверки содержательной гипотезы: «Эти три текста принадлежат одному автору»?
16. Датчик случайных чисел выдал N чисел $0, 1, \dots, 9$. Среди них число i встречалось n_i раз ($n_0 + \dots + n_9 = N$). Как проверить гипотезу о случайности чисел?

Лекция 17. Оценка параметров уравнения регрессии

*Все модели неправильны,
но некоторые полезны.*
Джордж Бокс

План лекции: метод наименьших квадратов, общая модель линейной регрессии, свойства оценок метода наименьших квадратов, нормальная регрессия.

17.1. Метод наименьших квадратов

Пусть Y — случайная величина, $X_i, i = 1, \dots, k$ — контролируемые (неслучайные) переменные. При этом значения величины Y зависят не только от значений X_i , но и от других факторов, в том числе таких, которые не поддаются контролю. Поэтому для фиксированного значения X^* величина Y подвержена некоторому разбросу (рис. 9).

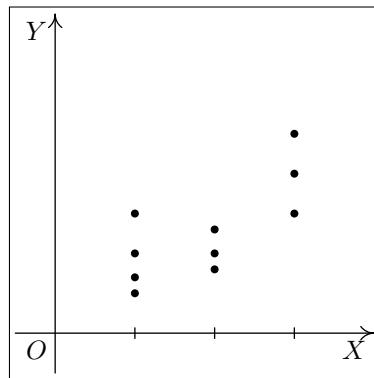


Рис. 9. Данные для отыскания регрессионной зависимости

Модель (функциональная зависимость) известна из предварительных соображений с точностью до параметров:

$$Y = f(X_1, \dots, X_k, \theta_1, \dots, \theta_s) + \varepsilon,$$

где $\theta_i, i = 1, \dots, s$ — параметры, ε — вектор ошибок.

Набор данных имеет вид

$$Y_1 = f(X_{11}, \dots, X_{1k}, \theta_1, \dots, \theta_s) + \varepsilon_1,$$

...

$$Y_n = f(X_{n1}, \dots, X_{nk}, \theta_1, \dots, \theta_s) + \varepsilon_n,$$

где X_{ij} — значение j -й переменной при i -м измерении. Будем считать, что $\mathbf{E} \varepsilon = 0$ и ошибки некоррелированы: $K_\varepsilon = \text{cov}_\varepsilon = \mathbf{E}(\varepsilon \varepsilon^T) = \sigma^2 E_n$.

По методу наименьших квадратов (МНК) оценки параметров выбираются так, чтобы минимизировать сумму квадратов ошибок, то есть оцениваются из условия

$$R = \sum_{i=1}^n (Y_i - f(X_{i1}, \dots, X_{ik}, \theta_1, \dots, \theta_s))^2 \rightarrow \min,$$

$$\frac{\partial R}{\partial \theta_j} = 0, \quad j = 1, \dots, s.$$

Линия $Y = f(X_1, \dots, X_k, \theta_1, \dots, \theta_s)$ называется *линией регрессии* Y на X ⁴⁵.

17.2. Общая модель линейной регрессии

Наиболее часто используются линейные по параметрам модели вида

$$Y = \theta_1 \cdot a_1(X_1) + \theta_2 \cdot a_2(X_2) + \dots + \theta_k \cdot a_k(X_k) + \varepsilon.$$

(Если $a_k(X_k) \equiv 1$, будет свободный член.)

Рассмотрим основной случай, когда $a_i(X_i) = X_i$. Одномерный случай ($k = 1$) представлен на рис. 10.

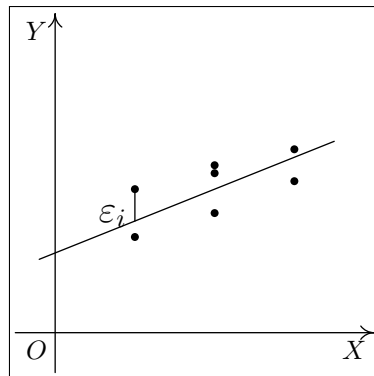


Рис. 10. Иллюстрация модели линейной регрессии

В общем случае исходные данные имеют вид

$$\begin{aligned} Y_1 &= X_{11}\theta_1 + \dots + X_{1k}\theta_k + \varepsilon_1 \\ &\dots \\ Y_n &= X_{n1}\theta_1 + \dots + X_{nk}\theta_k + \varepsilon_n \end{aligned} \quad \text{или} \quad Y = X \cdot \theta + \varepsilon,$$

⁴⁵ Термин «регрессия» введен Ф. Гальтоном (1886). Он установил, что особенности роста родителей проявляются и у их детей, но в среднем в меньшей степени. У родителей низкого или высокого роста рост детей оказывается несколько ближе к среднему, то есть (в среднем) у низких дети выше, а у высоких родителей — ниже. Гальтон назвал это явление «возвратом» (regression).

где

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_{11} & \dots & X_{1k} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nk} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Кроме того, $\mathbf{E} \varepsilon = 0$, $\mathbf{E}(\varepsilon \varepsilon^T) = \sigma^2 E_n$.

Сумма квадратов ошибок или остаточная дисперсия

$$R = K(\theta) = \sum (Y_i - \sum_{j=1}^k X_{ij} \theta_j)^2 \rightarrow \min,$$

что эквивалентно

$$R = (Y - X\theta)^T(Y - X\theta) \rightarrow \min.$$

Предположим, что матрица $X^T X$ не вырождена и найдем оценку θ из условия

$$\frac{\partial R}{\partial \theta_l} = -2 \sum_{i=1}^n X_{il} (Y_i - \sum X_{ij} \theta_j) = 0, \quad l = 1, \dots, k.$$

Имеем

$$X^T(Y - X\theta) = 0 \iff X^T Y = X^T X \theta \iff \hat{\theta} = (X^T X)^{-1} \cdot X^T Y.$$

Уравнение $X^T X \theta = X^T Y$ называется **нормальным уравнением** метода наименьших квадратов, а $\hat{\theta} = (X^T X)^{-1} \cdot X^T Y$ — оценкой параметров линейной регрессии методом наименьших квадратов (о.н.к.).

Теорема 17.1. Пусть $\hat{\theta}$ — любое решение нормального уравнения. Тогда $\min R(\theta) = R(\hat{\theta})$. Если $|X^T X| \neq 0$, то о. н. к. единственна и равна $\hat{\theta} = (X^T X)^{-1} \cdot X^T Y$.

Доказательство. Пусть $\bar{\theta}$ — произвольное фиксированное значение θ ; тогда

$$\begin{aligned} R(\theta) &= (Y - X\theta)^T(Y - X\theta) = [Y - X\bar{\theta} + X(\bar{\theta} - \theta)]^T[Y - X\bar{\theta} + X(\bar{\theta} - \theta)] = \\ &= R(\bar{\theta}) + 2(\bar{\theta} - \theta)^T(X^T Y - X^T X \bar{\theta}) + (\bar{\theta} - \theta)^T X^T X (\bar{\theta} - \theta). \end{aligned}$$

При $\bar{\theta} = \hat{\theta} = (X^T X)^{-1} \cdot X^T Y$ средний член обращается в 0 и

$$R(\theta) = R(\hat{\theta}) + (\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta) \geq R(\hat{\theta}), \quad (124)$$

так как матрица $X^T X$ неотрицательно определена. Равенство достигается при $\theta = \hat{\theta}$, следовательно, это и есть точка минимума. Единственность следует из однозначной разрешимости нормального уравнения при невырожденной матрице $X^T X$ \square .

На рис. 11 изображена расчетная линия регрессии $\hat{y} = a_0 + a_1x$, среднее значение зависимой переменной y и три точки, соответствующие заданному значению переменной x_i : наблюдаемое значение y_i , вычисленное по регрессии $\hat{y}_i = a_0 + a_1x_i$, и среднее значение \bar{y} .

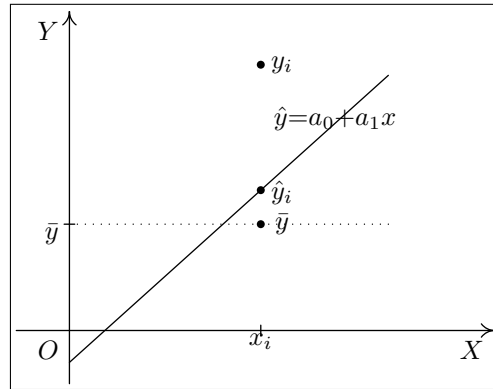


Рис. 11. Фактическое, расчетное и среднее значения зависимой переменной y

Пример 17.1. Найдём оценку параметра линейной модели $Y = X \cdot \theta + \varepsilon$ по данным:

X	0	1	2	3	4
Y	-1,2	0,1	3,0	4,9	7,3

◀ Напомним, что X и Y в формулах — это векторы-столбцы (как бы ни были они заданы в условиях).

$$\hat{\theta} = (X^T X)^{-1} \cdot X^T Y.$$

$$X^T X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = 30;$$

$$X^T Y = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} -1,2 \\ 0,1 \\ 3,0 \\ 4,9 \\ 7,3 \end{pmatrix} = 50;$$

$$\hat{\theta} = (X^T X)^{-1} \cdot X^T Y = 5/3; \hat{Y} = 5/3X \approx 1,7X.$$



Пример 17.2. Найдем в условиях предыдущего примера величину \hat{R} , равную сумме квадратов ошибок.



$$\hat{R} = \sum_i (Y_i - \hat{\theta} X_i)^2 = \sum_i (Y_i - \hat{Y}_i)^2,$$

где \hat{Y}_i — значения, вычисленные по уравнению регрессии $\hat{Y} = 1,7X$.

X	0	1	2	3	4
Y	-1,2	0,1	3,0	4,9	7,3
\hat{Y}	0	1,7	3,4	5,1	6,8
$ Y - \hat{Y} $	1,2	1,6	0,4	0,2	0,5

$$\hat{R} = \sum_i (Y_i - \hat{Y}_i)^2 = 4,45.$$



Пример 17.3. Найдем по данным примера 17.1 оценки параметров линейной модели

$$Y = X \cdot \theta + \theta_1 + \varepsilon.$$

◀ Будем считать, что переменная X двумерна, и рассматривать модель $Y = X_1 \cdot \theta + X_2 \theta_1 + \varepsilon$, где $X_2 \equiv 1$.

X_1	0	1	2	3	4
X_2	1	1	1	1	1
Y	-1,2	0,1	3,0	4,9	7,3

$$\hat{\theta} = (X^T X)^{-1} \cdot X^T Y.$$

$$X^T X = \begin{pmatrix} 30 & 10 \\ 10 & 5 \end{pmatrix}; (X^T X)^{-1} = 1/50 \cdot \begin{pmatrix} 5 & -10 \\ -10 & 30 \end{pmatrix} = \begin{pmatrix} 0,1 & -0,2 \\ -0,2 & 0,6 \end{pmatrix}.$$

$$X^T Y = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1,2 \\ 0,1 \\ 3,0 \\ 4,9 \\ 7,3 \end{pmatrix} = \begin{pmatrix} 50 \\ 14,1 \end{pmatrix}.$$

$$\hat{\theta} = (X^T X)^{-1} \cdot X^T Y = \begin{pmatrix} 0,1 & -0,2 \\ -0,2 & 0,6 \end{pmatrix} \cdot \begin{pmatrix} 50 \\ 14,1 \end{pmatrix} = \begin{pmatrix} 2,18 \\ -1,54 \end{pmatrix};$$

$$\hat{Y} = 2,18X - 1,54.$$

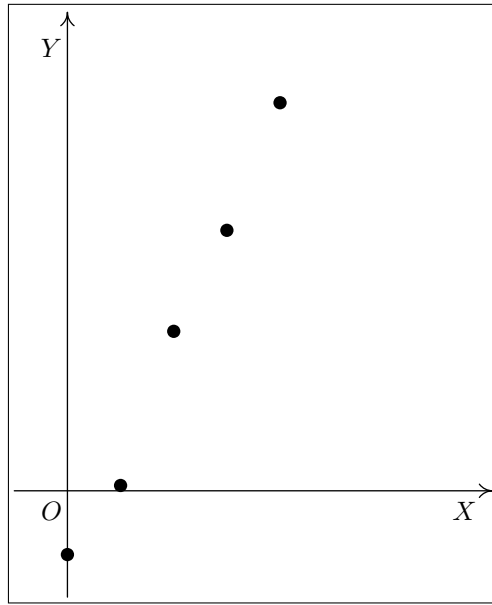


Рис. 12. Графическое изображение данных (примеры 17.1–17.3)

Найдем для этой модели величину \hat{R} , равную сумме квадратов ошибок:

$$\hat{R} = \sum_i (Y_i - \sum_{j=1}^2 X_{ij}\theta_j)^2 = \sum_i (Y_i - \hat{Y}_i)^2,$$

где \hat{Y}_i — значения, вычисленные по уравнению регрессии $\hat{Y} = 2,18X - 1,54$ (рис. 12).

X	0	1	2	3	4
Y	−1,2	0,1	3,0	4,9	7,3
\hat{Y}	−1,54	0,64	2,82	5	7,18
$ Y - \hat{Y} $	0,34	0,54	0,18	0,1	0,12

$$\hat{R} = \sum_i (Y_i - \hat{Y}_i)^2 = 0,4640.$$

Как видим, ошибка при использовании модели со свободным членом примерно в 10 раз меньше. Причина этого понятна из рис. 12. Очевидно, модель без свободного члена ($Y = X \cdot \theta + \varepsilon$) не является адекватной. ►

В одномерном случае для нахождения параметров линейной модели $Y = X \cdot \theta + \theta_1 + \varepsilon$ удобно использовать следующие формулы, получаемые из

соотношения $\hat{\theta} = (X^T X)^{-1} \cdot X^T Y$ (или непосредственно путем отыскания минимума функции $R = R(\theta, \theta_1)$ с помощью частных производных):

$$\hat{\theta} = \frac{SS_{xy}}{SS_{xx}}, \quad \hat{\theta}_1 = \bar{y} - \hat{\theta} \cdot \bar{x}, \quad (125)$$

где

$$SS_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n},$$

$$SS_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}.$$

Пример 17.4. Найдём оценки параметров линейной модели $Y = X \cdot \theta + \theta_1 + \varepsilon$ по данным примера 17.1, используя формулы (125).

◀

$$SS_{xy} = \sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n} = 50 - \frac{10 \cdot 14,1}{5} = 21,8.$$

$$SS_{xx} = \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n} = 30 - \frac{100}{5} = 10.$$

$$\hat{\theta} = \frac{SS_{xy}}{SS_{xx}} = \frac{21,8}{10} = 2,18.$$

$$\hat{\theta}_1 = \bar{y} - \hat{\theta} \cdot \bar{x} = \frac{14,1}{5} - 2,18 \cdot 2 = 2,82 - 4,36 = -1,54.$$

Полученный ответ $\hat{Y} = 2,18X - 1,54$ совпадает с ответом примера 17.3. ►

17.3. Свойства оценок МНК

1. Несмещенность.

$$\begin{aligned} \mathbf{E} \hat{\theta} &= \mathbf{E}[(X^T X)^{-1} X^T Y] = \mathbf{E}[(X^T X)^{-1} X^T (X\theta + \varepsilon)] = \\ &= (X^T X)^{-1} X^T X (\mathbf{E} \theta + \mathbf{E} \varepsilon) = \mathbf{E} \theta + \mathbf{E} \varepsilon = \theta. \end{aligned}$$

2. $\mathbf{K}_{\hat{\theta}} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$.

Имеем

$$K_{\hat{\theta}} = \mathbf{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T].$$

Так как

$$\begin{aligned}\hat{\theta} - \theta &= (X^T X)^{-1} X^T (X\theta + \varepsilon) - \theta = \\ &= (X^T X)^{-1} X^T X\theta + (X^T X)^{-1} X^T \varepsilon - \theta = \\ &= \theta + (X^T X)^{-1} X^T \varepsilon - \theta = (X^T X)^{-1} X^T \varepsilon,\end{aligned}$$

то

$$\begin{aligned}K_{\hat{\theta}} &= \mathbf{E}[(X^T X)^{-1} X^T \varepsilon \cdot \varepsilon^T X (X^T X)^{-1}] = \\ &= (X^T X)^{-1} X^T \cdot \mathbf{E}[\varepsilon \cdot \varepsilon^T] \cdot X (X^T X)^{-1} = \\ &= (X^T X)^{-1} X^T \cdot \sigma^2 E \cdot X (X^T X)^{-1} = \\ &= \sigma^2 E \cdot (X^T X)^{-1} (X^T X) (X^T X)^{-1} = \\ &= \sigma^2 \cdot (X^T X)^{-1}.\end{aligned}$$

Использовалось, что $X^T X$ и $(X^T X)^{-1}$ — симметрические матрицы, и $((X^T X)^{-1})^T = (X^T X)^{-1}$.

3. Оценивание остаточной дисперсии $R(\hat{\theta})$.

$$\mathbf{E} R(\hat{\theta}) = \mathbf{E}[(Y - X\hat{\theta})^T (Y - X\hat{\theta})] = \sum \mathbf{D} Y_i = n\sigma^2.$$

Из равенства (124) имеем

$$R(\theta) = R(\hat{\theta}) + (\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta),$$

и

$$\mathbf{E} R(\theta) = \mathbf{E} R(\hat{\theta}) + \mathbf{E}(\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta), \quad (126)$$

$$\begin{aligned}\mathbf{E}(\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta) &= \sum_{i,j=1}^k a_{ij} \mathbf{E}(\hat{\theta}_i - \theta_i)^T X^T X (\hat{\theta}_j - \theta_j) = \\ &= \sum_{i,j=1}^k a_{ij} \text{cov}(\hat{\theta}_i, \hat{\theta}_j) = \sum_{i,j=1}^k a_{ij} \sigma^2 a_{ij}^{(-1)} = \sigma^2 \text{tr}(E_k) = k\sigma^2.\end{aligned}$$

Подставляя $k\sigma^2$ в (126), получаем:

$$n\sigma^2 = \mathbf{E} R(\hat{\theta}) + k\sigma^2,$$

откуда

$$\mathbf{E} R(\hat{\theta}) = (n - k)\sigma^2.$$

Таким образом, несмещенной оценкой для остаточной дисперсии σ^2 является статистика

$$\hat{\sigma}^2 = \frac{R(\hat{\theta})}{n - k} = \frac{[(Y - X\hat{\theta})^T(Y - X\hat{\theta})]}{n - k}.$$

Вектор $U = Y - X\hat{\theta}$ называют **остаточным вектором**, а его компоненты — **остатками**. Остаточный вектор можно записать в виде

$$U = (E_n - X(X^T X)^{-1} X^T)Y,$$

а несмещенная оценка для остаточной дисперсии σ^2 тогда имеет вид

$$\hat{\sigma}^2 = \frac{Y(E_n - X(X^T X)^{-1} X^T)Y^T}{n - k}, \quad (127)$$

поскольку $(E_n - X(X^T X)^{-1} X^T)^T(E_n - X(X^T X)^{-1} X^T) = (E_n - X(X^T X)^{-1} X^T)$.

4. Теорема Гаусса — Маркова.

Теорема 17.2 (теорема Гаусса — Маркова). *В классе несмещенных линейных оценок оценки МНК имеют наименьшую дисперсию.*

Доказательство. Надо доказать, что в ковариационной матрице $K_{\hat{\theta}}$, где $\hat{\theta}$ — оценка МНК, по диагонали стоят элементы не больше чем в $K_{\bar{\theta}}$, где $\bar{\theta}$ — произвольная несмещенная линейная оценка.

Оценка, линейная от наблюдения, имеет вид $\bar{\theta} = Ty$, где T — матрица, элементы которой константы. Сравним $K_{\hat{\theta}}$ и $K_{\bar{\theta}}$:

$$K_{\hat{\theta}} = \sigma^2(X^T X)^{-1}, \quad K_{\bar{\theta}} = \mathbf{E}[(\bar{\theta} - \theta)(\bar{\theta} - \theta)^T].$$

Но $\mathbf{E} \bar{\theta} = \mathbf{E}[T(X\theta + \varepsilon)] = \mathbf{E} TX\theta + T \mathbf{E} \varepsilon = TX\theta$, значит, $TX = E$. Тогда

$$\bar{\theta} - \theta = T(\theta X + \varepsilon) - \theta = \theta TX + T\varepsilon - \theta = T\varepsilon.$$

Значит,

$$\mathbf{E}[(\bar{\theta} - \theta)(\bar{\theta} - \theta)^T] = \mathbf{E}[(T\varepsilon)(T\varepsilon)^T] = \mathbf{E}[(T\varepsilon \cdot \varepsilon^T T^T)] = \sigma^2 \cdot T \cdot T^T.$$

Сравним диагональные элементы матриц $(X^T X)^{-1}$ для $\hat{\theta}$ и $T \cdot T^T$ для $\bar{\theta}$.

Рассмотрим матрицу $A = T - (X^T X)^{-1} X^T$, тогда

$$A^T = T^T - X \cdot (X^T X)^{-1},$$

откуда

$$\begin{aligned}
A \cdot A^T &= (T - (X^T X)^{-1} X^T)(T^T - X \cdot (X^T X)^{-1}) = T \cdot T^T - \\
&- (X^T X)^{-1} X^T \cdot T^T - T \cdot X \cdot (X^T X)^{-1} + (X^T X)^{-1} X^T X \cdot (X^T X)^{-1} = \\
&= T \cdot T^T - (X^T X)^{-1} - (X^T X)^{-1} + X^T X^{-1} = \\
&= T \cdot T^T - (X^T X)^{-1}
\end{aligned}$$

(так как $TX = E$ и $X^T T^T = (TX)^T = E$).

Но в матрице AA^T диагональные элементы больше либо равны 0, так как это скалярные квадраты строк. Следовательно, диагональные элементы $T \cdot T^T$ больше либо равны диагональным элементам $(X^T X)^{-1}$, то есть дисперсии оценок $\hat{\theta}$ метода наименьших квадратов — наименьшие в классе линейных несмещенных оценок. Это означает, что о.н.к. оптимальны в данном классе. \square

17.4. Нормальная регрессия

Раньше мы предполагали, что ошибки $\varepsilon_1, \dots, \varepsilon_n$ некоррелированы, имеют нулевые математические ожидания и одинаковую положительную дисперсию: $\mathbf{E} \varepsilon = 0$, $K_\varepsilon = \text{cov}_\varepsilon = \mathbf{E}(\varepsilon \varepsilon^T) = \sigma^2 E_n$. Для находжений вероятностей отклонений о.н.к. от истинных значений рассматриваемых параметров, расчета доверительных интервалов, проверки гипотез необходимо сделать дополнительные предположения о виде распределения случайного вектора $Y = (Y_1 \dots Y_n)^T$. Поскольку $Y = X \cdot \theta + \varepsilon$, где X — неслучайная переменная, закон распределения Y определяется законом распределения ε . Будем считать, что ошибки подчиняются нормальному закону распределения $N(0, \sigma)$ (распределение Y будет тогда тоже нормальным, с математическим ожиданием $X\theta$). В данной параметрической модели можно найти оценку параметра θ с помощью метода максимального правдоподобия. Возникает вопрос, как связаны о.м.п. и о.н.к. Легко видеть, что максимум функции правдоподобия достигается при минимуме $R = \sum (Y_i - \sum_{j=1}^k X_{ij} \theta_j)^2 = \sum \varepsilon_i^2$, таким образом, оценки максимального правдоподобия совпадают с оценками наименьших квадратов.

Из теоремы Гаусса — Маркова известно, что о.н.к. оптимальны в классе несмещенных линейных оценок. В нормальной модели справедливо и более сильное утверждение: оценки МНК имеют наименьшую дисперсию в классе всех несмещенных оценок [2].

Выясним, каково распределение $\hat{\theta}$ и связанных с ней статистик. Спра-

ведлива теорема, напоминающая теорему Фишера и лежащая в основе теории нормальной регрессии.

Теорема 17.3. *В нормальной регрессионной модели*

- 1) величина $\hat{\theta}$ имеет многомерное нормальное распределение с вектором математических ожиданий θ и ковариационной матрицей $\sigma^2(X^T X)^{-1}$;
- 2) статистика $\frac{\hat{\theta}_l - \theta_l}{\sigma \sqrt{a_{ll}^{(-1)}}}$, где $a_{ll}^{(-1)}$ — соответствующий элемент матрицы $(X^T X)^{-1}$, имеет распределение $N(0; 1)$ для $l = 1, \dots, k$;
- 3) случайные величины $\hat{\theta}$ и $R(\hat{\theta})$ независимы;
- 4) случайные величины $R(\hat{\theta})$ и $R(\theta) - R(\hat{\theta})$ независимы;
- 5) статистика $\frac{R(\theta) - R(\hat{\theta})}{\sigma^2}$ имеет распределение χ_k^2 ;
- 6) статистика $\frac{R(\hat{\theta})}{\sigma^2}$ имеет распределение χ_{n-k}^2 .

Доказательство.

1) Так как $\hat{\theta}$ линейно выражается через нормально распределенный случайный вектор ε , то закон распределения $\hat{\theta}$ также нормален. Параметры этого распределения уже вычислены выше:

$$\mathbf{E} \hat{\theta} = \theta, \mathbf{K}_{\hat{\theta}} = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}.$$

Следовательно, величина $\hat{\theta}$ имеет многомерное нормальное распределение с вектором математических ожиданий θ и ковариационной матрицей $\sigma^2(X^T X)^{-1}$.

2) Тогда каждый из выборочных коэффициентов регрессии $\hat{\theta}_l, l = 1, \dots, k$, имеет распределение $N(\theta_l; \sigma \sqrt{a_{ll}^{(-1)}})$, где $a_{ll}^{(-1)}$ — соответствующий элемент матрицы $(X^T X)^{-1}$, а статистика

$$\frac{\hat{\theta}_l - \theta_l}{\sigma \sqrt{a_{ll}^{(-1)}}}$$

имеет распределение $N(0; 1)$.

3) Будем использовать следующую лемму [2].

Лемма.

Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $N(0, 1)$. Рассмотрим квадратичную форму $Q = X^T A X$ размерности n с действительной и симметричной матрицей коэффициентов A и линейную

форму $T = BX$, где B есть матрица порядка $m \times n$. Если

$$BA = O$$

(O — матрица с нулевыми элементами), то функции Q и T независимы.

По предположению, ошибки ε имеют дисперсию σ^2 ; будем рассматривать нормированные ошибки $\varepsilon^* = \varepsilon/\sigma$:

$$\begin{aligned}\hat{\theta} &= (X^T X)^{-1} \cdot X^T Y = \theta + (X^T X)^{-1} X^T \varepsilon = \\ &= \theta + \sigma (X^T X)^{-1} X^T \varepsilon / \sigma = \theta + \sigma (X^T X)^{-1} X^T \varepsilon^*.\end{aligned}$$

Это линейная форма с матрицей $B = (X^T X)^{-1} X^T$.

$$R(\hat{\theta}) = (Y - X\hat{\theta})^T (Y - X\hat{\theta}) = U^T U,$$

что равно, по (127),

$$U^T U = Y(E_n - X(X^T X)^{-1} X^T) Y^T.$$

Подставим сюда $Y = X\theta + \varepsilon$ и раскрыв скобки, получим, что

$$X^T (E_n - X(X^T X)^{-1} X^T) = (E_n - X(X^T X)^{-1} X^T) X = O.$$

Тогда

$$\frac{R(\hat{\theta})}{\sigma^2} = (\varepsilon/\sigma)^T (E_n - X A^{-1} X^T) \varepsilon / \sigma = (\varepsilon^*)^T (E_n - X A^{-1} X^T) \varepsilon^*;$$

это квадратичная форма с действительной и симметричной матрицей коэффициентов $A = E_n - X(X^T X)^{-1} X^T$. Условие $BA = O$ выполняется, так как $BA = (X^T X)^{-1} (E_n - X(X^T X)^{-1} X^T) = O$.

По лемме $\hat{\theta}$ и $R(\hat{\theta})$ независимы.

4) Из равенства (124) имеем

$$R(\theta) - R(\hat{\theta}) = (\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta).$$

Статистика $R(\theta) - R(\hat{\theta})$ зависит от выборки только через $\hat{\theta}$, но $\hat{\theta}$ и $R(\hat{\theta})$ независимы, значит, независимы и $R(\hat{\theta})$ и $R(\theta) - R(\hat{\theta})$.

5) Тот факт, что статистика $\frac{R(\theta) - R(\hat{\theta})}{\sigma^2}$ имеет распределение χ_k^2 , вытекает из представления

$$R(\theta) - R(\hat{\theta}) = (\hat{\theta} - \theta)^T X^T X (\hat{\theta} - \theta),$$

где $\hat{\theta}$ распределено по нормальному закону.

6) Статистика $\frac{R(\hat{\theta})}{\sigma^2}$ также имеет распределение χ^2 в силу представления

$$\frac{R(\hat{\theta})}{\sigma^2} = (\varepsilon^*)^T (E_n - X(X^T X)^{-1} X^T) \varepsilon^*,$$

однако надо выяснить число степеней свободы этого распределения. Известна теорема [2]. Пусть $Q = X^T A X$ и $\text{rang} A = r \leq n$. Если матрица идемпотентна ($A^2 = A$), то Q имеет распределение χ_r^2 и при этом $r = \text{tr} A$. В данном случае $A = E_n - X(X^T X)^{-1} X^T$, идемпотентность этой матрицы мы уже проверяли. Найдем ее след.

$$\text{tr} A = \text{tr} E_n - \text{tr}(X(X^T X)^{-1} X^T) = n - k = \text{tr} E_k = n - k.$$

Следовательно, по данной теореме $\frac{R(\hat{\theta})}{\sigma^2}$ имеет распределение χ_{n-k}^2 . \square

17.5. Интервальное оценивание

Рассмотренная теорема позволяет построить доверительный интервал для коэффициентов регрессии $\hat{\theta}_l, l = 1, \dots, k$ так же, как строили раньше интервал для среднего нормального закона с неизвестной дисперсией — на основе стьюдентова отношения:

$$t_l = \sqrt{\frac{n-k}{a_{ll}^{(-1)} R(\hat{\theta})}} (\hat{\theta}_l - \theta_l),$$

где $a_{ll}^{(-1)}$ — соответствующий элемент матрицы $(X^T X)^{-1}$. Очевидно, что статистика t_l (как частное независимых величин с распределением $N(0; 1)$ в числителе и χ_{n-k}^2 в знаменателе) имеет распределение Стьюдента T_{n-k} . Доверительный интервал для коэффициента регрессии θ_l имеет вид

$$I_{\theta_l} = \left[\hat{\theta}_l - t_{n-k, 1-\frac{\alpha}{2}} \sqrt{\frac{a_{ll}^{(-1)}}{n-k} R(\hat{\theta})}; \hat{\theta}_l + t_{n-k, 1-\frac{\alpha}{2}} \sqrt{\frac{a_{ll}^{(-1)}}{n-k} R(\hat{\theta})} \right].$$

Пример 17.5. Найдем доверительный интервал для коэффициента регрессии θ в линейной модели $Y = X \cdot \theta + \theta_1 + \varepsilon$.

◀ Данные имеют вид

$$x = \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix}.$$

$$X^T X = \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_1 & 1 \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix}.$$

Определитель матрицы $X^T X$ равен

$$|X^T X| = n \sum x_i^2 - (\sum x_i)^2 = n \sum (x_i - \bar{x})^2;$$

тогда обратная матрица $(X^T X)^{-1}$:

$$(X^T X)^{-1} = \frac{1}{|X^T X|} \begin{pmatrix} n & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}.$$

Нас интересует элемент $a_{ll}^{(-1)}$ обратной матрицы:

$$a_{ll}^{(-1)} = \frac{n}{n \sum (x_i - \bar{x})^2} = \frac{1}{\sum (x_i - \bar{x})^2}.$$

Таким образом, доверительный интервал для коэффициента регрессии θ равен

$$\left[\hat{\theta} - t_{n-2, 1-\frac{\alpha}{2}} \sqrt{\frac{R(\hat{\theta})}{(n-2) \sum (x_i - \bar{x})^2}}; \hat{\theta} + t_{n-2, 1-\frac{\alpha}{2}} \sqrt{\frac{R(\hat{\theta})}{(n-2) \sum (x_i - \bar{x})^2}} \right].$$

$R(\hat{\theta})$ — это сумма квадратов ошибок: $R(\hat{\theta}) = \sum_i (y_i - \hat{y}_i)^2$, окончательно получаем доверительный интервал для θ :

$$\hat{\theta} \pm t_{n-2, 1-\frac{\alpha}{2}} \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{(n-2) \sum (x_i - \bar{x})^2}}$$

или, в обозначениях регрессионного анализа,

$$\hat{\theta} \pm t_{n-2, 1-\frac{\alpha}{2}} S_{\hat{\theta}},$$

где

$$S_{\hat{\theta}} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{(n-2) \sum (x_i - \bar{x})^2}}.$$



Рассматривая доверительные области для линейных комбинаций параметров, можно получить для одномерной линейной модели доверительные интервалы для среднего и индивидуального значения [2].

Доверительный интервал для среднего значения y при фиксированном x_0 :

$$\hat{y} \pm t_{n-2, 1-\frac{\alpha}{2}} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}; \quad (128)$$

доверительный интервал для индивидуального значения y при фиксированном x_0 :

$$\hat{y} \pm t_{n-2, 1-\frac{\alpha}{2}} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}, \quad (129)$$

где

$$S = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{(n-2)}}.$$

Можно рассматривать интервалы как функции $x_0 = x$ (рис. 13).

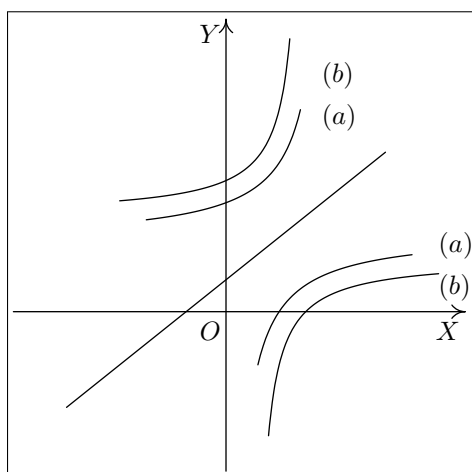


Рис. 13. Доверительные интервалы среднего (a) и индивидуального значения (b)

Чем дальше x_0 от \bar{x} , тем шире доверительные интервалы (128, 129). Очевидно, интервал для индивидуального значения y задает более широкую область.

17.6. Контрольные вопросы

1. Опишите общую модель регрессии.
2. Опишите общую модель линейной регрессии.
3. Опишите модель полиномиальной регрессии.

4. *Какая модель называется линейной по параметрам?*
5. *Как свести модель полиномиальной регрессии к общей модели линейной регрессии?*
6. *В чем суть метода наименьших квадратов?*
7. *Сформулируйте свойство несмещенности о.н.к. параметров уравнения регрессии.*
8. *Запишите ковариационную матрицу оценок метода наименьших квадратов.*
9. *Сформулируйте теорему Гаусса — Маркова.*

Библиографический список

- [1] *Боровков, А. А. Математическая статистика. Оценка параметров. Проверка гипотез* / А. А. Боровков. — М. : ФИЗМАТЛИТ, 2007. — 704 с.
- [2] *Ивченко, Г. И. Математическая статистика : учеб. пособие* / Г. И. Ивченко, Ю. И. Медведев. — М. : Высш. шк., 1984. — 248 с.
- [3] *Крамер, Г. Математические методы статистики* / Г. Крамер. — М. : Мир, 1975. — 648 с.
- [4] *Крупкина, Т. В. Математическая статистика в примерах и задачах : практикум по решению задач* / Т. В. Крупкина, А. К. Гречкосеев. — Красноярск: ИПК СФУ, 2009. — 104 с. — (Математическая статистика : УМКД № 1455-2008 / рук. творч. коллектива Т. В. Крупкина).
- [5] *Крупкина, Т. В. Математическая статистика : метод. указания* / Т. В. Крупкина, Г. А. Федоров. — Красноярск: ИПК СФУ, 2009. — 52 с. — (Математическая статистика : УМКД № 1455-2008 / рук. творч. коллектива Т. В. Крупкина).
- [6] *Чернова, Н. И. Математическая статистика : учеб. пособие* / Н. И. Чернова ; Новосибир. гос. ун-т. — Новосибирск, 2007. — 148 с.

Дополнительная литература

- [7] *Андерсон, Т. Введение в многомерный статистический анализ* / Т. Андерсон — М. : Гос. изд. физ.-мат. лит., 1963. — 500 с.
- [8] *Беклемишев, Д. В. Курс аналитической геометрии и линейной алгебры* / Д. В. Беклемишев. — М. : ФИЗМАТЛИТ, Наука, 2007. — 307 с.
- [9] *Бешелев, С. Д. Математико-статистические методы экспертных оценок* / С. Д. Бешелев, Ф. Г. Гурвич. — М. : Статистика, 1980. — 263 с.
- [10] *Большев, Л. Н. Таблицы математической статистики* / Л. Н. Большев, Н. В. Смирнов. — М. : Наука, 1983. — 415 с.
- [11] *Вентцель, Е. С. Теория вероятностей и ее инженерные приложения : учеб. пособие* / Е. С. Вентцель, Л. А. Овчаров. — 2-е изд., стереотип. — М. : Высш. шк., 2000. — 480 с.

- [12] *Вирт, Н. Алгоритмы и структуры данных* / Н. Вирт. — СПб. : Невский диалект, 2001. — 352 с.
- [13] *Володин, Б. Г. Сборник задач по теории вероятностей, математической статистике и теории случайных функций* : учеб. пособие / Б. Г. Володин, М. П. Ганин, И. Я. Динер, Л. Б. Комаров, А. А. Свешников, К. Б. Старобин ; под ред. А. А. Свешникова. — М. : Наука. Гл. ред. физ.-мат. лит., 1970. — 632 с.
- [14] *Вуколов, Э. Л. Основы статистического анализа. Практикум по статистическим методам и исследованию операции с использованием пакетов STATISTICA и EXCEL* : учеб. пособие / Э. Л. Вуколов. — 2-е изд., испр. и доп. — М. : ФОРУМ, 2008. — 464 с.
- [15] *Гихман, И. И. Теория вероятностей и математическая статистика* : учебник / И. И. Гихман, А. В. Скороход, М. И. Ядренко. — 2-е изд., перераб. и доп. — Киев. : Выща шк., 1988. — 439 с.
- [16] *Гнеденко, Б. В. Курс теории вероятностей* : учебник / Б. В. Гнеденко. — 8-е изд., испр. и доп. — М. : Едиториал УРСС, 2005. — 448 с.
- [17] *Емельянов, Г. В. Задачник по теории вероятностей и математической статистике* : учеб. пособие / Г. В. Емельянов, В. П. Скитович. — Л. : Изд-во Ленингр. ун-та, 1967. — 331 с.
- [18] *Зорич, В. А. Математический анализ. Т. 1, 2* / В. А. Зорич. — М. : МЦНМО, 2007. — 1458 с.
- [19] *Кендалл, М. Теория распределений* / М. Кендалл, А. Стюарт. — М. : Наука. Гл. ред. физ.-мат. лит., 1966. — 588 с.
- [20] *Кендалл, М. Статистические выводы и связи* / М. Кендалл, А. Стюарт. — М. : Наука. Гл. ред. физ.-мат. лит., 1973. — 899 с.
- [21] *Кибзун, А. И. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами* : учеб. пособие / А. И. Кибзун, Е. Р. Горяинова, А. В. Наумов, А. Н. Сиротин. — М. : ФИЗМАТЛИТ, 2002. — 224 с.
- [22] *Климов, Г. П. Теория вероятностей и математическая статистика* / Г. П. Климов. — М. : Изд-во Моск. ун-та, 1983. — 328 с.
- [23] *Кнут, Д. Искусство программирования. Т. 3. Сортировка и поиск* / Д. Кнут. — М. : Вильямс, 2000. — 822 с.
- [24] *Козлов, М. В. Введение в математическую статистику* : учеб. пособие / М. В. Козлов, А. В. Прохоров. — М. : МГУ, 1987. — 264 с.
- [25] *Коршунов, Д. А. Сборник задач и упражнений по математической статистике* : учеб. пособие / Д. А. Коршунов, Н. И. Чернова. — Новосибирск : Изд-во Ин-та математики, 2004. — 128 с.

- [26] *Крупкина, Т. В. Конспект лекций по теории вероятностей* : учеб. пособие / Т. В. Крупкина. — Красноярск : КрасГУ, 1996. — 87 с.
- [27] *Крупкина, Т. В. Теория вероятностей, математическая статистика и эконометрика* : учеб. пособие : в 2 кн. Кн. 1 / Т. В. Крупкина, А. К. Гречкосеев. — Красноярск : КрасГУ, 1999. — 216 с.
- [28] *Крупкина, Т. В. Теория вероятностей и математическая статистика* : учеб. пособие : в 2 ч. Ч. 1 / Т. В. Крупкина, В. П. Малый. — Красноярск : КрасГУ, 1991. — 80 с.
- [29] *Леман, Э. Теория точечного оценивания* / Э. Леман. — М. : Наука, 1991. — 448 с.
- [30] *Никитин, Я. Ю. Асимптотическая эффективность непараметрических критериев* / Я. Ю. Никитин. — М. : Наука, 1995. — 240 с.
- [31] *Орлов, А. И. Экспертные оценки* : учеб. пособие / А. И. Орлов. — М. : ИВСТЭ, 2002. — 31 с.
- [32] *Орлов, А. И. Прикладная статистика* : учебник / А. И. Орлов. — М. : Экзамен, 2004. — 656 с.
- [33] *Пугачев, В. С. Теория вероятностей и математическая статистика* : учеб. пособие / В. С. Пугачев. — М. : ФИЗМАТЛИТ, 2002. — 496 с.
- [34] *Севастьянов, Б. А. Курс теории вероятностей и математической статистики* : учебник / Б. А. Севастьянов. — М. : Наука. Гл. ред. физ.-мат. лит., 1982. — 255 с.
- [35] *Феллер, В. Введение в теорию вероятностей и ее приложения* : учебник : в 2 т. Т. 1 / В. Феллер. — 2-е изд., перераб. и доп. — М. : Мир, 1984. — 528 с.
- [36] *Ширяев, А. Н. Вероятность* : учеб. пособие / А. Н. Ширяев. — М. : Наука, 1989. — 610 с.
- [37] *Фихтенгольц, Г. М. Курс дифференциального и интегрального исчисления* : в 3 томах / Г. М. Фихтенгольц. — М. : ФИЗМАТЛИТ, 2001. (т. 1 — 616 с.; т. 2 — 810 с.; т. 3 — 662 с.)
- [38] **СТО 4.2-07-2008. Система менеджмента качества. Общие требования к построению, изложению и оформлению документов учебной и научной деятельности** [Текст] / разраб. : Т. В. Сильченко, Л. В. Белошапко, В. К. Младенцева, М. И. Губанова. — Введ. впервые 09.12.2008. — Красноярск : ИПК СФУ, 2008. — 47 с.
- [39] **Каталог лицензионных программных продуктов, используемых в СФУ** / сост. : А. В. Сарафанов, М. М. Торопов. — Красноярск : Сиб. федер. ун-т, 2008. — Вып. 1–4.
- [40] **Стандарт организации СТО СФУ 7.2.04-2007. Электронные образовательные ресурсы на базе гипертекстовых технологий со встроенной системой компьютерной проверки знаний тестированием. Требования к структуре, организации и интерфейсу** / разраб. : К. Н. Захарьин, А. В. Сарафанов, А. Г. Суковатый и др. — Красноярск : Сиб.

федер. ун-т, 2007. Утвержден и введен в действие приказом ректора СФУ № 659 от 15.11.2007 г.

- [41] *Захарьин, К. Н. Электронные презентационные материалы* / А. В. Сарафанов, К. Н. Захарьин, А. Г. Суковатый. — Красноярск, 2008.
- [42] **Учебно-методические комплексы дисциплин.** Основные компоненты : метод. рекомендации / сост. : Л. И. Вейсова, С. И. Почекутов, А. В. Сарафанов, А. Ю. Смолин. — Красноярск : ИПК СФУ, 2008. — 12 с.
- [43] **Каталог инновационных учебно-методических комплексов дисциплин и электронных ресурсов** / сост. : К. Н. Захарьин, А. В. Сарафанов, А. Г. Суковатый, А. С. Теремов, М. В. Шипова. — Красноярск : ИПК СФУ, 2008. — Вып. 1. — 298 с.

Электронные и интернет-ресурсы

- [44] **Унифицированная система компьютерной проверки знаний тестированием UniTest версии 3.0.0** : руководство пользователя / А. Н. Шниперов, Б. М. Бидус. — Красноярск, 2008.
- [45] **Математическая статистика. [Электронный ресурс]** : электрон. учеб.-метод. комплекс по дисциплине / Т. В. Крупкина, А. К. Гречкосеев. — Красноярск : ИПК СФУ, 2009. — (Математическая статистика : УМКД № 1455-2008 / рук. творч. коллектива Т. В. Крупкина).
- [46] **Математическая статистика. Банк тестовых заданий.** Версия 1.0 [Электронный ресурс] : контрольно-измерительные материалы / Т. В. Крупкина, Е. С. Кирик, Г. А. Федоров. — Красноярск : ИПК СФУ, 2009. — (Математическая статистика : УМКД № 1455-2008 / рук. творч. коллектива Т. В. Крупкина).
- [47] *Чернова, Н. И. Лекции по математической статистике* : учеб. пособие / Н. И. Чернова. — Новосибирск : НГУ, 2002 (Internet).

Перечень наглядных и других пособий, методических указаний и материалов по техническим средствам обучения

- [48] **Математическая статистика.** Презентационные материалы. Версия 1.0 [Электронный ресурс] : наглядное пособие / Т. В. Крупкина, А. К. Гречкосеев. — Красноярск : ИПК СФУ, 2009. — (Математическая статистика : УМКД № 1455-2008 / рук. творч. коллектива Т. В. Крупкина).
- [49] **Интерактивные технические средства обучения: практическое руководство** / сост. : А. Г. Суковатый, К. Н. Захарьин, А. В. Казанцев, А. В. Сарафанов. — Красноярск : ИПК СФУ, 2009. — 84 с.

Таблица 3

Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0, 0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0, 1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0, 2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0, 3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0, 4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0, 5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0, 6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0, 7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0, 8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0, 9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1, 0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1, 1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1, 2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1, 3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1, 4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1, 5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1, 6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1, 7	09405	09246	09089	08933	08780	08628	08478	08330	08183	08038
1, 8	07895	07754	07614	07477	07341	07207	07074	06943	06814	06687
1, 9	06562	06438	06316	06195	06077	05960	05844	05730	05618	05508
2, 0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04492
2, 1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2, 2	03548	03470	03394	03319	03246	03174	03103	03034	02966	02899
2, 3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2, 4	02240	02186	02134	02083	02033	01984	01936	01889	01842	01797
2, 5	01753	01710	01667	01625	01585	01545	01506	01468	01431	01394
2, 6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2, 7	01042	01014	00987	00961	00935	00910	00885	00861	00837	00814
2, 8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2, 9	00595	00578	00562	00545	00530	00514	00499	00485	00471	00457
3, 0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
x	Десятые доли x									
	0	1	2	3	4	5	6	7	8	9
3	00443	00327	00238	00172	00123	00084	00061	00043	00029	00020

Замечание. В таблице даны значения, округленные до пятого знака после запятой.

Указание. Пусть необходимо получить значение $\varphi(0,62)$. На пересечении столбца 2 («Сотые доли x ») и строки 0,6 (« x ») получаем значение 32 918, то есть

$$\varphi(0,62) = 0,32918.$$

Таблица 4

Значения функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05117	05567	05962	06356	06749	07142	07535
0,2	07920	08317	08700	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17365	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22241
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26731	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31328
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40148
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43447	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45819	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47671
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48499	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48839	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49491	49506	49520
2,6	49535	49547	49560	49573	49586	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49737
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49830	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49899
x	Десятые доли x									
	0	1	2	3	4	5	6	7	8	9
3	49865	49903	49931	49952	49966	49977	49984	49989	49993	49995

Замечание. В таблице даны значения, округленные до пятого знака после запятой.

Указание. Пусть необходимо получить значение $\Phi_0(1,57)$. На пересечении столбца 7 («Сотые доли x ») и строки 1,5 (« x ») получаем значение 44 179, то есть

$$\Phi_0(1,57) = 0,44179.$$

Таблица 5

Случайные числа

5686	4215	5470	4452	4841	0477	9567	7496	1297	3594	1020	3531	4296
3106	9375	4545	0447	0754	6377	1127	7126	1021	4070	4488	2365	9633
9359	1271	7562	0122	8112	4863	1022	0731	8446	2302	8433	3299	5987
2057	0762	1429	8535	9029	9745	3458	5023	3502	2436	6435	2646	0295
6177	2755	3080	3275	0521	6623	1133	3278	0500	7573	7426	3188	0187
7707	3047	4901	3519	7888	6411	1631	6981	1972	4269	0022	3860	1580
6751	4022	6540	7804	5528	4690	3586	9839	6641	0404	0735	0888	3504
2651	9051	5764	7155	6489	2660	3341	8784	0605	4640	8692	7712	9832
6607	0480	2557	3461	9755	4398	8857	0221	3844	1823	4407	5914	7545
2362	2428	7899	2623	9965	7366	0486	8185	5896	3985	3105	7210	5375
2213	8481	0919	2350	7310	7106	0046	1683	6269	1120	5436	8921	6457
8361	9849	9902	4244	2377	9213	4625	5978	5266	7521	8488	6854	9203
2598	2673	2399	5112	4318	5003	3532	6430	5679	5041	2108	1813	4235
3915	9380	3918	5957	3603	6553	6247	8907	5282	1106	9223	5629	6982
4138	2901	7592	1650	2580	5676	6470	0122	0820	2140	5291	8499	3653
1727	0453	3032	2902	4114	2462	2820	0414	7197	3854	2940	3500	8685
6131	0774	7788	5011	4971	0848	0748	7103	3262	5182	1185	1493	3425
0114	4662	0802	1125	8745	5513	9750	0695	5727	7577	8631	0759	5430
9953	1426	0405	2109	2304	5329	2475	8555	8172	1376	3459	6778	6917
0159	9635	7058	4886	2373	5937	9383	5763	8004	8602	2457	9134	0099
2200	2369	8140	4865	4874	4867	5206	0434	3845	0659	0499	3671	2771
2104	9275	2118	8024	1033	0528	3665	9721	6339	3377	3780	0366	4746

Таблица 6

Значения квантилей распределения Стьюдента T_n

n/α	0,95	0,975	0,99	0,995
2	2,92	4,30	6,97	9,52
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,37	4,03
6	1,94	2,45	3,14	3,71
7	1,90	2,37	3,00	3,50
8	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17
11	1,80	2,20	2,72	3,11
12	1,78	2,18	2,68	3,06
13	1,77	2,16	2,65	3,01
14	1,76	2,15	2,62	2,98
15	1,75	2,13	2,60	2,95
16	1,75	2,12	2,58	2,92
18	1,73	2,10	2,55	2,88
20	1,73	2,09	2,53	2,85
22	1,72	2,07	2,51	2,82
25	1,71	2,06	2,49	2,79
27	1,70	2,06	2,48	2,77
30	1,70	2,04	2,46	2,75
40	1,68	2,02	2,42	2,70
50	1,68	2,00	2,40	2,68
70	1,67	1,99	2,38	2,65
100	1,66	1,98	2,36	2,63
∞	1,65	1,96	2,33	2,58

Указание. Пусть необходимо получить значение квантили распределения Стьюдента T_9 порядка $\alpha = 0,95$. На пересечении столбца 0,95 (порядок « α ») и строки 9 (число степеней свободы « n ») находим значение 1,83, то есть

$$t_{9;0,95} = 1,83.$$

Замечание. В силу симметричности распределения Стьюдента $t_{n;1-\alpha} = -t_{n;\alpha}$.

Таблица 7

Значения квантилей распределения χ_n^2

n/α	0,01	0,05	0,10	0,90	0,95	0,99
1	0,000	0,004	0,016	2,71	3,84	6,63
2	0,020	0,103	0,211	4,61	5,99	9,21
3	0,115	0,352	0,584	6,25	7,81	11,34
4	0,297	0,711	1,064	7,78	9,49	13,28
5	0,554	1,15	1,61	9,24	11,07	15,09
6	0,872	1,64	2,20	10,64	12,59	16,81
7	1,24	2,17	2,83	12,02	14,07	18,48
8	1,65	2,73	3,49	13,36	15,51	20,09
9	2,09	3,33	4,17	14,68	16,92	21,67
10	2,56	3,94	4,87	15,99	18,31	23,21
11	3,05	4,57	5,58	17,28	19,68	24,73
12	3,57	5,23	6,30	18,55	21,03	26,22
13	4,11	5,89	7,04	19,81	22,36	27,69
14	4,66	6,57	7,79	21,06	23,68	29,14
15	5,23	7,26	8,55	22,31	25,00	30,58
16	5,81	7,96	9,31	23,54	26,30	32,00
17	6,41	8,67	10,09	24,77	27,59	33,41
18	7,01	9,39	10,86	25,99	28,87	34,81
19	7,637	10,12	11,65	27,20	30,14	36,19
20	8,26	10,85	12,44	28,41	31,41	37,57
25	11,52	14,62	14,62	34,38	37,65	44,31
30	14,95	10,85	18,49	40,26	43,77	50,89
40	22,16	26,51	29,05	51,81	55,76	63,69
60	37,48	43,19	46,46	74,40	79,08	88,38
80	53,54	60,39	64,28	96,58	101,88	112,33
100	70,06	95,70	100,62	140,23	146,57	158,95

Указание. Пусть необходимо получить значение квантили распределения χ_{10}^2 порядка $\alpha = 0,95$. На пересечении столбца 0,95 (порядок « α ») и строки 10 (число степеней свободы « n ») находим значение 18,31, то есть

$$\chi_{10;0,95}^2 = 18,31.$$

Таблица 8

Значения квантилей распределения F порядка $\alpha = 0,01$

n_2/n_1	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6054
2	98,5	99,0	99,2	99,25	99,3	99,33	99,34	99,36	99,38	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,8	11,0	9,8	9,2	8,8	8,5	8,3	8,1	8,0	7,9
7	12,3	9,6	8,5	7,9	7,5	7,2	7,0	6,8	6,7	6,6
8	11,3	8,7	7,6	7,0	6,6	6,4	6,2	6,0	5,9	5,8
9	10,6	8,0	7,0	6,4	6,1	5,8	5,6	5,5	5,4	5,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,2	5,1	5,0	4,9
11	9,9	7,2	6,2	5,7	5,3	5,1	4,9	4,7	4,6	4,5
12	9,3	6,9	6,0	5,4	5,1	4,8	4,7	4,5	4,4	4,3

Таблица 9

Значения квантилей распределения F порядка $\alpha = 0,05$

n_2/n_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,25	19,3	19,33	19,36	19,37	19,38	19,38
3	10,1	9,6	9,3	9,1	9,0	8,94	8,88	8,84	8,81	8,78
4	7,7	7,0	6,6	6,4	6,3	6,2	6,1	6,04	6,0	5,96
5	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,78	4,74
6	6,0	5,1	4,8	4,5	,4	4,3	4,2	4,15	4,1	4,06
7	5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7	3,68	3,63
8	5,3	4,5	4,1	3,8	3,7	3,6	3,5	3,44	3,4	3,34
9	5,1	4,3	3,9	3,6	3,5	3,4	3,3	3,2	3,18	3,13
10	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,07	3,02	2,97
11	4,8	4,0	3,6	3,4	3,2	3,1	3,0	2,95	2,9	2,86
12	4,75	3,9	3,5	3,3	3,1	3,0	2,9	2,85	2,8	2,76

Указание. Пусть необходимо получить значение квантили распределения $F_{10,12}$ порядка $\alpha = 0,05$. На пересечении столбца 10 (число степеней свободы « n_1 ») и строки 12 (число степеней свободы « n_2 ») находим значение 2,76, то есть

$$F_{10,12;0,95} = 2,76.$$

Таблица 10

Критические значения критерия U Манна-Уитни при уровне значимости $\alpha = 0,05$

n_1/n_2	4	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60
4	1																			
6	3	5																		
9	6	12	21																	
12	9	17	30	42																
15	12	23	39	55	72															
18	16	28	48	68	88	109														
21	19	34	57	81	105	130	154													
24	22	39	66	94	122	150	179	207												
27	25	45	75	107	139	171	203	236	268											
30	28	50	85	120	156	192	228	265	301	338										
33	31	56	94	133	173	213	253	293	334	374	415									
36	35	61	103	146	189	233	278	322	367	411	456	501								
39	38	67	112	159	206	254	302	351	399	448	497	546	595							
42	41	72	121	172	223	275	327	380	432	485	538	591	644	697						
45	44	78	131	185	240	296	352	408	465	522	579	636	693	750	808					
48	47	83	140	198	257	317	377	437	498	559	620	681	742	804	865	927				
51	50	89	149	211	274	338	402	466	531	596	661	726	791	857	922	988	1054			
54	53	94	158	224	291	359	427	495	564	633	702	771	841	910	980	1050	1119	1189		
57	57	100	167	237	308	379	451	524	597	670	743	816	890	964	1037	1111	1185	1259	1333	
60	60	105	177	250	325	400	476	553	630	707	784	862	939	1017	1095	1199	1251	1329	1407	1486

Греческий алфавит

Буква		Название
Α	α	альфа
Β	β	бета
Γ	γ	гамма
Δ	δ	дельта
Ε	ε	эпсилон
Ζ	ζ	дзета
Η	η	эта
Θ	ϑ	тета
Ι	ι	йота
Κ	κ	каппа
Λ	λ	лямбда
Μ	μ	мю
Ν	ν	ню
Ξ	ξ	кси
Ο	\omicron	омикрон
Π	π	пи
Ρ	ρ	ро
Σ	σ	сигма
Τ	τ	тау
Υ	υ	ипсилон
Φ	φ	фи
Χ	χ	хи
Ψ	ψ	пси
Ω	ω	омега

Предметный указатель

- байесовская
 - оценка, 60
- байесовское
 - сравнение оценок, 60
- вероятность
 - доверительная, 95
- вклад
 - выборки, 53
- выборка, 5, 7
 - простая, 8
- гипотеза
 - двусторонняя, 128
 - о дисперсии, 132
 - о дисперсиях, 133
 - о математическом ожидании, 133
 - о среднем, 133
 - о средних, 133
 - о средних для парных совокупностей, 134
 - односторонняя, 127
 - статистическая, 104
 - альтернативная, 104
 - независимости, 106
 - о виде распределения, 105
 - однородности, 105
 - параметрическая, 106
 - простая, 104
 - сложная, 104
 - случайности, 106
- гистограмма выборки, 27
- граница
 - доверительная, 95
- дисперсия
 - выборочная, 17
- значение
 - среднее выборочное, 17
- значения
 - выборки экстремальные, 10
- интервал
 - двусторонний, 95
 - доверительный, 94
 - асимптотический, 95
 - квантильный, 32
 - медианный, 32
 - модальный, 32
 - центральный, 98
- квантиль
 - выборочная
 - порядка q , 23
 - группированная выборочная, 32
- ковариация
 - выборочная, 23
- количество
 - Фишера информационное, 54
- коэффициент
 - асимметрии
 - выборочный, 18
 - корреляции
 - выборочный, 23
 - экссесса
 - выборочный, 18
- критериев
 - сравнение, 113
- критерий
 - асимптотического уровня α , 145
 - байесовский, 124
 - знаков, 138

Колмогорова — Смирнова, 151
 Манна — Уитни, 140
 Манна-Уилкоксона — Уитни, 140
 минимаксный, 125
 н.м.к., 113
 непараметрический, 138
 знаков, 138
 нерандомизированный, 122
 несмещенный, 128
 оптимальный, 128
 отношения правдоподобия, 117
 рандомизированный, 123
 р.н.м. несмещенный, 129
 р.н.м.к., 114
 ранговый, 140
 рандомизированный, 122
 Стьюдента, 133
 для парных выборок, 134
 согласия, 144
 Колмогорова, 146
 Пирсона, 148
 состоятельный, 120
 статистический, 104
 суммы рангов, 140
 Уилкоксона-Манна — Уитни, 140
 Фишера, 133
 критерия
 ошибка, 125
 размер, 109, 123
 уровень значимости, 109
 лемма
 Фишера, 42
 МНК, *см.* метод наименьших квадратов
 медиана
 выборочная, 23
 группированная
 выборочная, 32
 метод
 наименьших квадратов, 155, 156
 минимаксное
 сравнение
 оценок, 60
 мода
 выборочная, 23
 группированная
 выборочная, 32
 модель
 биномиальная, 9
 Гамма, 9
 дискретная, 9
 Коши, 9
 линейной регрессии, 156
 непрерывная, 9
 нормальная, 9
 общая, 9
 Пуассоновская, 9
 параметрическая, 8
 равномерная, 9
 общая, 9
 регулярная, 8
 статистическая, 7
 момент
 начальный
 выборочный, 16
 центральный
 выборочный, 17
 монотонное отношение правдоподобия, 127
 мощность
 критерия, 110, 111
 Неймана — Пирсона
 лемма, 126
 НКО, *см.* область критическая наилучшая
 неравенство
 Рао — Крамера, 54
 для параметрической функции, 71
 о.м.м., *см.* оценка метода моментов
 о.м.п, *см.* оценка максимального правдоподобия
 область
 критерия
 критическая, 107, 144
 критическая
 наилучшая (НКО), 114
 объем

- выборки, 7
- ошибка
 - i -го рода, 110
 - второго рода, 109
 - первого рода, 108
- оценивание, 48
- оценка
 - асимптотически несмещенная, 49
 - максимального
 - правдоподобия, 64
 - метода моментов, 68
 - несмещенная, 49
 - оптимальная, 53
 - оптимальная по критерию, 53
 - параметрической функции
 - эффективная, 72
 - эффективная, 58
 - сильно состоятельная, 51
 - смещенная, 49
 - состоятельная, 50
 - точечная, 48
- полигон частот выборки, 28
- распределение
 - Стьюдента, 37
 - Фишера, 39
 - Фишера — Снедекора, см. распределе-
ние Фишера
 - хи-квадрат, 34
- распределение
 - k -й порядковой статистики, 11
 - эмпирической частоты, 10
- регрессия
 - интервальное оценивание, 167
 - нормальная, 164
- риск
 - математическое ожидание, 124
- робастность, 142
- ряд
 - вариационный, 9
 - выборки
 - вариационный, 10
 - реализации
 - вариационный, 9
- смещение
 - оценки, 49
- совокупность
 - генеральная, 5, 8
- среднеквадратический
 - подход, 60
- среднеквадратическое отклонение
 - выборочное, 17
- статистика
 - выборочная, 47
 - достаточная, 80
 - критерия, 144
 - минимальная, 85
 - полная, 92
 - порядковая, 10
 - центральная, 95
- студентово отношение, 43
- теорема
 - Гаусса — Маркова, 163
 - Гливенко, 13
 - Гливенко — Кантелли, 146
 - Колмогорова, 13
 - Неймана — Пирсона, 115
 - Неймана — Фишера, см. факторизации
нормальной регрессии, 165
 - о выборочном среднем, 40
 - Рао — Блекуэлл — Колмогорова, 90
 - Смирнова, 14
 - Фишера, 41
 - факторизации, 83
- теория
 - оценивания, 47
- уравнение
 - правдоподобия, 64
 - регрессии
 - выборочное, 23
- уровень
 - значимости критерия, 144
- функция
 - Колмогорова, 13
 - мощности критерия, 110
 - правдоподобия, 63

дискретной величины, [63](#)
распределения
эмпирическая, [11](#)

характеристика
выборочная, [16](#)
эмпирическая, *см.* выборочная

частота
эмпирическая, [10](#)

экспоненциальное семейство, [128](#)

элемент
выборки, [7](#)

Оглавление

Предисловие	3
Принятые обозначения и сокращения	4
1. Введение в математическую статистику	5
1.1. Предмет математической статистики	5
1.2. Статистические модели	7
1.3. Порядковые статистики и вариационный ряд выборки	9
1.4. Эмпирическая функция распределения	10
1.5. Контрольные вопросы	14
2. Основные выборочные характеристики	16
2.1. Выборочные моменты и функции от выборочных моментов	16
2.2. Свойства выборочного среднего \bar{X} и S^2	18
2.3. Другие характеристики выборки	23
2.4. Сходимость выборочных моментов и функций от них	24
2.5. Контрольные вопросы	26
3. Группировка выборки	27
3.1. Гистограмма и полигон частот	27
3.2. Пример группировки выборки	28
3.3. Поправки Шеппарда	30
3.4. Контрольные вопросы	33
4. Важные распределения математической статистики	34
4.1. Распределения хи-квадрат, Стьюдента, Фишера	34
4.2. Теорема Фишера	40
4.3. Теоремы о распределении выборочных характеристик	43
4.4. Контрольные вопросы	45
5. Точечное статистическое оценивание	47
5.1. Оценки параметров	47
5.2. Несмещенность	49
5.3. Состоятельность	50
5.4. Контрольные вопросы	51
6. Неравенство Рао — Крамера	53
6.1. Оптимальность оценок	53
6.2. Неравенство Рао — Крамера	53
6.3. Формы информационного количества Фишера	56
6.4. Эффективные оценки	58
6.5. Байесовский и минимаксный подходы к сравнению оценок	60

6.6.	Контрольные вопросы	61
7.	Методы получения оценок	63
7.1.	Метод максимального правдоподобия	63
7.2.	Примеры	65
7.3.	Некоторые свойства оценок максимального правдоподобия	67
7.4.	Метод моментов	67
7.5.	Некоторые свойства оценок метода моментов	69
7.6.	Контрольные вопросы	70
8.	Оценивание параметрической функции	71
8.1.	Неравенство Рао — Крамера для параметрической функции	71
8.2.	Другой подход к эффективным оценкам	74
8.3.	Экспоненциальное семейство	75
8.4.	Контрольные вопросы	78
9.	Достаточные статистики	80
9.1.	Определение и примеры	80
9.2.	Критерий достаточности	82
9.3.	Свойства достаточных статистик	84
9.4.	Свойства оценок максимального правдоподобия	86
9.5.	Контрольные вопросы	87
10.	Оптимальные оценки	88
10.1.	Свойства оптимальных оценок	88
10.2.	Достаточные статистики и оптимальные оценки	90
10.3.	Полные статистики	92
10.4.	Контрольные вопросы	93
11.	Интервальные оценки	94
11.1.	Понятие интервального оценивания параметров	94
11.2.	Построение доверительного интервала	96
11.3.	Доверительные интервалы для параметров $N(a, \sigma)$	97
11.4.	Примеры расчетов	100
11.5.	Асимптотические доверительные интервалы	101
11.6.	Контрольные вопросы	103
12.	Основные понятия теории проверки гипотез	104
12.1.	Основные понятия	104
12.2.	Примеры математических формулировок гипотез	105
12.3.	Общая схема проверки параметрических гипотез	107
12.4.	Ошибки первого и второго рода	108
12.5.	Контрольные вопросы	111
13.	Подходы к сравнению критериев	113
13.1.	Сравнение мощности критериев	113
13.2.	Критерий Неймана — Пирсона	115
13.3.	Состоятельность критерия	120
13.4.	Рандомизированные критерии	121
13.5.	Байесовские и минимаксные критерии	124
13.6.	Контрольные вопросы	129

14.	Критерии проверки гипотез о параметрах нормального распределения	131
14.1.	Алгоритм проверки параметрических гипотез	131
14.2.	Гипотезы о параметрах одного распределения	132
14.3.	Гипотезы о параметрах двух независимых распределений	133
14.4.	Гипотеза о средних для парных совокупностей	134
14.5.	Примеры	134
14.6.	Контрольные вопросы	137
15.	Непараметрические критерии	138
15.1.	Непараметрические критерии для проверки параметрических гипотез	138
15.2.	Критерий знаков	138
15.3.	Ранговый U -критерий	140
15.4.	Контрольные вопросы	142
16.	Критерии согласия	144
16.1.	Построение критериев согласия	145
16.2.	Критерий согласия Колмогорова	146
16.3.	Критерий согласия χ^2 Пирсона	148
16.4.	Проверка гипотезы однородности: критерий Колмогорова – Смирнова	151
16.5.	Проверка гипотезы независимости: критерий χ^2 Пирсона	152
16.6.	Контрольные вопросы	153
17.	Оценка параметров уравнения регрессии	155
17.1.	Метод наименьших квадратов	155
17.2.	Общая модель линейной регрессии	156
17.3.	Свойства оценок МНК	162
17.4.	Нормальная регрессия	164
17.5.	Интервальное оценивание	167
17.6.	Контрольные вопросы	169
	Список литературы	171
	Приложение	175
	Оглавление	187