МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Механико-математический факультет
Кафедра теории вероятностей и математической статистики

Н. И. Чернова

Математическая статистика

Учебное пособие

Второе издание, исправленное и дополненное

Новосибирск
2014

УДК 519.21
ББК В17я73-2
Ч 493

Рецензент
канд. физ.-мат. наук, доцент А. П. Ковалевский

Чернова, Н. И.
Ч 493 Математическая статистика : учеб. пособие / Н. И. Чернова ; Новосиб. гос. ун-т — 2-е изд., испр. и доп. — Новосибирск : РИЦ НГУ, 2014. — 150 с.

ISBN 978-5-4437-0304-6

Учебное пособие содержит полный курс лекций по математической статистике, много лет читаемый автором на втором курсе отделения экономики экономического факультета НГУ. Подбор материала является традиционным для курса теоретической статистики, излагаемого студентам экономических специальностей университетов, и включает точечное оценивание параметров, проверку гипотез, элементы регрессионного и дисперсионного анализа.

Пособие соответствует требованиям ФГОС ВПО к профессиональным образовательным программам по специальности 080100 «Экономика».

Предназначено для студентов экономических специальностей.

УДК 519.21
ББК В17я73-2

ISBN 978-5-4437-0304-6

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Учебное пособие содержит полный курс лекций по математической статистике, читаемый автором с 1996 г. студентам второго курса отделения экономики экономического факультета НГУ. Курс математической статистики опирается на семестровый курс теории вероятностей и является основой для годового курса эконометрики. В результате изучения предмета студенты должны овладеть математическими методами исследования различных моделей математической статистики.

Особенностью пособия является его теоретическая направленность. От читателя требуется хорошее знание теории вероятностей, включая совместные распределения, предельные теоремы, соотношения между различными видами сходимости последовательностей случайных величин.

Курс состоит из 11 глав. Первая глава является главной для понимания предмета. Она знакомит читателя с основными понятиями математической статистики. Материал глав II—VI и VII в известной степени независим. Читателю следует быть готовым к тому, что в курсе лекций материал седьмой главы будет излагаться раньше второй.

Вторая глава знакомит читателя с методами точечного оценивания неизвестных параметров распределения: моментов и максимального правдоподобия. Байесовское оценивание рассмотрено в главе одиннадцатой.

Третья глава посвящена сравнению оценок в среднеквадратическом и асимптотическом смыслах. Эта тема продолжается в четвёртой главе, где изучается неравенство Рао — Крамера как средство проверки эффективности оценок. Точка в этой теме будет поставлена лишь в одиннадцатой главе, где наилучшие в среднеквадратическом смысле оценки будут построены как функции от достаточных и полных статистик.

В пятой главе начинается изучение интервального оценивания параметров, которое завершается в следующей главе построением интервалов для параметров нормального распределения. Для этого вводятся специальные статистические распределения, которые затем используются в критериях согласия в восьмой главе. Глава седьмая стоит несколько особняком, будучи связанной с не очень жизненной задачей математической статистики: проверкой двух простых гипотез. Однако здесь даны необходимые основные понятия теории проверки гипотез, поэтому изучить её читателю следует весьма тщательно.

Наконец, главы с девятой по одиннадцатую представляют совсем специальные разделы математической статистики и могут считаться предисловием к курсу эконометрики. В девятой главе рассмотрены простые модели и методы регрессионного анализа, доказаны основные свойства полученных оценок. В десятой главе подробно изучены свойства многомерного нормального распределения и приведён пример использования многомерной центральной предельной теоремы.

Последняя глава содержит необходимый теоретический материал об условных распределениях и условных математических ожиданиях. Показано использование условных математических ожиданий для построения байесовских и эффективных оценок.

Практически каждая глава завершается списком упражнений по тексту главы. Приложение содержит таблицы с перечнем основных характеристик дискретных и абсолютно непрерывных распределений, таблицы основных статистических распределений.

В конце книги приведён подробный предметный указатель. В списке литературы перечислены учебники, которые можно использовать в дополнение к курсу, и сборники задач для практических занятий.

Нумерация параграфов в каждой главе отдельная. Формулы, примеры, утверждения и т. п. имеют сквозную нумерацию. При ссылке на объект из другой главы для удобства читателя указан номер страницы, на которой содержится объект. При ссылке на объект из той же главы приводится только номер формулы, примера, утверждения. Окончание доказательств отмечено значком $\square$.

Автор глубоко признателен своим коллегам по кафедре теории вероятностей и математической статистики ММФ НГУ и студентам ЭФ НГУ разных лет, немало содействовавшим улучшению курса.

ГЛАВА I

ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика опирается на методы теории вероятностей, но решает иные задачи. В теории вероятностей рассматриваются случайные величины с *заданным* распределением. Но откуда берутся знания о распределениях и вероятностях в практических экспериментах? С какой вероятностью выпадает герб на данной монете? Основная проблема состоит в том, что выводы приходится делать по результатам конечного числа наблюдений. Так, наблюдая 5 035 гербов после 10 000 бросаний монеты, нельзя сделать точный вывод о вероятности выпадения герба: даже если эта монета несимметрична, герб может выпасть 5035 раз. Точные выводы о распределении можно делать лишь тогда, когда проведено бесконечное число испытаний, что неосуществимо. Математическая статистика позволяет по результатам конечного числа экспериментов делать более-менее точные выводы о распределениях случайных величин, наблюдаемых в этих экспериментах.

§ 1. Основные понятия выборочного метода

Предположим, что мы повторяем один и тот же случайный эксперимент в одинаковых условиях и получаем некоторый набор данных (числовых или каких-то иных). При этом возникают следующие вопросы.

1. Если мы наблюдаем одну случайную величину — как по набору её значений в нескольких опытах сделать как можно более точный вывод о распределении этой случайной величины?

2. Если мы наблюдаем одновременно проявление двух или более признаков, т. е. имеем набор значений нескольких случайных величин,— что можно сказать об их зависимости или о совместном распределении этих случайных величин?

Часто бывает возможно высказать некие предположения о наблюдаемом распределении или о его свойствах. В этом случае по опытным данным требуется подтвердить или опровергнуть эти предположения («гипотезы»). При этом надо помнить, что ответ «да» или «нет» может быть дан лишь с определённой степенью достоверности, и чем дольше мы можем продолжать эксперимент, тем точнее могут быть выводы. Часто оказываются заранее известными некоторые свойства наблюдаемого эксперимента и можно сформулировать какие-то априорные выводы о распределении: о наличии функциональной зависимости между наблюдаемыми величинами, о нормальности распределения, о его симметричности, о наличии у распределения плотности или о его дискретном характере и т. д. Наличие таких знаний помогает на основании результатов эксперимента делать выводы о прочих, неизвестных, свойствах распределения.

Итак, математическая статистика работает там, где есть случайный эксперимент, свойства которого частично или полностью неизвестны и который мы умеем воспроизводить в одних и тех же условиях некоторое (а лучше — неограниченное) число раз.

Пусть $\xi : \Omega \to \mathbb{R}$ — случайная величина, наблюдаемая в случайном эксперименте. Предполагается, что вероятностное пространство задано и не будет нас интересовать.

Проведя n раз этот эксперимент в одинаковых условиях, получим числа X_1, X_2, $\ldots$, X_n — значения наблюдаемой случайной величины в первом, втором и т. д. экспериментах. Случайная величина ξ имеет некоторое распределение $\mathcal{F}$, которое нам *частично или полностью неизвестно*. Рассмотрим подробнее набор $\vec{X} = (X_1, \ldots, X_n)$, называемый *выборкой*.

В серии уже проведённых экспериментов выборка — это набор чисел. До того, как эксперимент проведён, имеет смысл считать выборку набором случайных величин (независимых и распределённых так же, как ξ). Действительно, до проведения опытов мы не можем сказать, какие значения примут элементы выборки: это будут какие-то из значений случайной величины ξ. Поэтому имеет смысл считать, что до опыта X_i — случайная величина, одинаково распределённая с ξ, а после опыта — число, которое мы наблюдаем в i-м по счёту эксперименте, т. е. одно из возможных значений *случайной величины* X_i.

Определение 1. *Выборкой* $\vec{X} = (X_1, \ldots, X_n)$ объёма n из распределения $\mathcal{F}$ называется набор из n независимых и одинаково распределённых случайных величин, имеющих распределение $\mathcal{F}$.

Что значит «по выборке сделать вывод о распределении»? Распределение характеризуется функцией распределения, плотностью или таблицей, набором числовых характеристик — $\mathsf{E}\xi$, $\mathsf{D}\xi$, $\mathsf{E}\xi^k$ и т. д. По выборке нужно уметь строить приближения для всех этих характеристик.

§ 2. Выборочные характеристики

Выборочное распределение. Рассмотрим *реализацию* выборки на одном элементарном исходе — числа $X_1 = X_1(\omega_0), \ldots, X_n = X_n(\omega_0)$. Разыграем новую случайную величину ξ^*, которая принимает значения $X_1, \ldots, X_n$ с одинаковыми вероятностями (например, с помощью правильного n-гранного кубика). Эта случайная величина определена на совсем ином вероятностном пространстве, чем изначальные случайные величины (на пространстве, связанном с бросанием кубика), поэтому будем вероятностную меру на нём обозначать $\tilde{\mathsf{P}}$ (соответственно, математическое ожидание — $\tilde{\mathsf{E}}$ и т. п.).

Запишем таблицу и функцию распределения случайной величины ξ^*:

$$\begin{array}{c|c|c|c} \xi^* & X_1 & \ldots & X_n \\ \hline \tilde{\mathsf{P}} & \frac{1}{n} & \ldots & \frac{1}{n} \end{array}, \quad F_n^*(y) = \sum_{X_i < y} \frac{1}{n} = \frac{\text{количество } X_i \in (-\infty, y)}{n}.$$

Распределение величины ξ^* называют *эмпирическим*, или *выборочным* распределением. Его можно описать как вероятностную меру на борелевской сигма-алгебре

$$P_n^*(B) = \tilde{\mathsf{P}}(\xi^* \in B) = \frac{\text{количество } X_i \in B}{n}.$$

Введём обозначения для числовых характеристик величины ξ^*. Её математическое ожидание равно

$$\tilde{\mathsf{E}}\xi^* = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

Дисперсия этой случайной величины равна

$$\tilde{\mathsf{D}}\xi^* = \sum_{i=1}^{n} \frac{1}{n} (X_i - \tilde{\mathsf{E}}\xi^*)^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = S^2.$$

Точно так же вычислим и момент порядка k

$$\tilde{\mathsf{E}}\,(\xi^*)^k = \sum_{i=1}^{n} \frac{1}{n}\, X_i^k = \frac{1}{n} \sum_{i=1}^{n} X_i^k = \overline{X^k}.$$

В общем случае обозначим через $\overline{g(X)}$ число

$$\tilde{\mathsf{E}} g(\xi^*) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) = \overline{g(X)}.$$

Если теперь мы позволим элементарному исходу ω_0 меняться, то все перечисленные выше характеристики $P_n^*(B)$, $F_n^*(y)$, $\overline{X}$, S^2, $\overline{X^k}$, $\overline{g(X)}$ станут величинами случайными, поскольку каждая из них будет функцией от n случайных величин $X_1, \ldots, X_n$.

Эти характеристики выборочного распределения используют для оценки (приближения) соответствующих неизвестных характеристик истинного распределения. Причина использования характеристик распределения ξ^* для оценки характеристик истинного распределения ξ (или X_1) — законы больших чисел. Все построенные нами выборочные характеристики являются средними арифметическими независимых и одинаково распределённых случайных величин и с ростом объёма выборки сходятся по вероятности к истинным характеристикам: математическому ожиданию, моментам, дисперсиям, вероятностям и т. п.

Познакомимся подробно с каждой из введённых выше характеристик и исследуем её свойства, в том числе поведение с ростом объёма выборки.

Эмпирическая функция распределения. Неизвестное истинное распределение $\mathcal{F}$ можно полностью описать с помощью его функции распределения $F(y) = \mathsf{P}(X_1 < y)$. Рассмотрим оценку для этой функции.

Определение 2. *Эмпирической функцией распределения*, построенной по выборке $\vec{X} = (X_1, \ldots, X_n)$ объёма n, называется случайная функция $F_n^* : \mathbb{R} \times \Omega \to [0, 1]$, при каждом $y \in \mathbb{R}$ равная

$$F_n^*(y) = \frac{\text{количество } X_i \in (-\infty,\, y)}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(X_i < y).$$

Напомним, что случайная функция переменной y

$$\mathrm{I}(X_i < y) = \begin{cases} 1, & \text{если } X_i < y, \\ 0 & \text{иначе} \end{cases}$$

называется *индикатором* события $\{X_i < y\}$. При каждом y этот индикатор является случайной величиной из распределения Бернулли с параметром $p = \mathsf{P}(X_i < y) = F(y)$ (*почему*?).

Если элементы выборки $X_1, \ldots, X_n$ упорядочить по возрастанию на каждом элементарном исходе, получится новый набор случайных величин, называемый *вариационным рядом*:

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n-1)} \leqslant X_{(n)}.$$

Здесь

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

Элемент $X_{(k)}$ называется k-м членом вариационного ряда или k-й *порядковой статистикой.*

Пример 1. Пусть дана числовая выборка

$$\vec{X} = (0;\ 2;\ 1;\ 2{,}6;\ 3{,}1;\ 4{,}6;\ 1;\ 4{,}6;\ 6;\ 2{,}6;\ 6;\ 7;\ 9;\ 9;\ 2{,}6).$$

Построим по ней вариационный ряд

$$(0;\ 1;\ 1;\ 2;\ 2{,}6;\ 2{,}6;\ 2{,}6;\ 3{,}1;\ 4{,}6;\ 4{,}6;\ 6;\ 6;\ 7;\ 9;\ 9)$$

и эмпирическую функцию распределения (рис. 1).

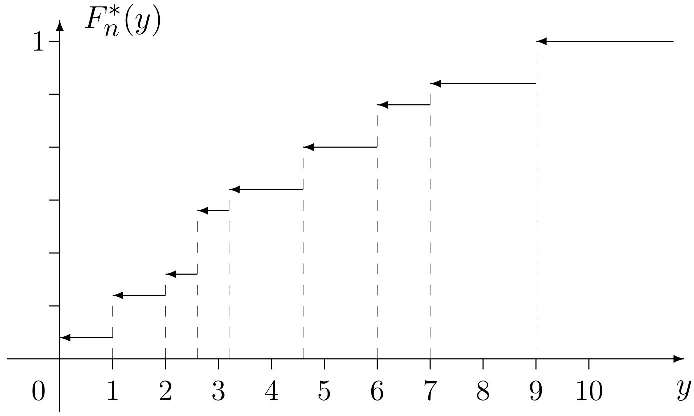

Рис. 1. Эмпирическая функция распределения

Эта функция является функцией распределения случайной величины, принимающей значение 0 с вероятностью $\frac{1}{15}$, значение 1 с вероятностью $\frac{2}{15}$, значение 2 с вероятностью $\frac{1}{15}$ и т. д.

Эмпирическая функция распределения имеет скачки в точках выборки (вариационного ряда), величина скачка в точке X_i равна $\frac{m}{n}$, где m — количество элементов выборки, совпадающих с X_i. Эмпирическая функция распределения по вариационному ряду строится так:

$$F_n^*(y) = \begin{cases} 0, & \text{если } y \leqslant X_{(1)}, \\ \dfrac{k}{n}, & \text{если } X_{(k)} < y \leqslant X_{(k+1)}, \\ 1 & \text{при } y > X_{(n)}. \end{cases}$$

Гистограмма. Другой характеристикой распределения является таблица для дискретных распределений или плотность — для абсолютно непрерывных. Эмпирическим аналогом таблицы или плотности является так называемая *гистограмма*.

Гистограмма строится по *группированным* данным. Предполагаемую область значений случайной величины ξ (или область выборочных данных) делят на некоторое количество не обязательно одинаковых интервалов. Пусть $A_1, \ldots, A_k$ — интервалы на прямой, называемые *интервалами группировки*. Обозначим для $j = 1, \ldots, k$ через ν_j число элементов выборки, попавших в интервал A_j:

$$\nu_j = \{\text{число } X_i \in A_j\} = \sum_{i=1}^{n} \mathrm{I}(X_i \in A_j), \quad n = \sum_{j=1}^{k} \nu_j. \tag{1}$$

На каждом из интервалов A_j строят прямоугольник, площадь которого пропорциональна ν_j. Общая площадь всех прямоугольников должна равняться единице. Если l_j — длина интервала A_j, то высота f_j прямоугольника над этим интервалом равна $f_j = \dfrac{\nu_j}{n\,l_j}$. Полученная фигура, состоящая из объединения прямоугольников, называется гистограммой.

Пример 2. Имеется вариационный ряд из примера 1:

$$(0;\ 1;\ 1;\ 2;\ 2{,}6;\ 2{,}6;\ 2{,}6;\ 3{,}1;\ 4{,}6;\ 4{,}6;\ 6;\ 6;\ 7;\ 9;\ 9).$$

Разобьём отрезок $[0,\ 10]$ на четыре равных отрезка. Отрезку $[0,\ 2{,}5)$ принадлежат четыре элемента выборки, отрезку $[2{,}5,\ 5)$ — шесть, отрезку $[5,\ 7{,}5)$ — три, и отрезку $[7{,}5,\ 10]$ — два элемента выборки. Строим гистограмму (рис. 2). На рис. 3 — гистограмма для той же выборки, но при разбиении области на пять равных отрезков.

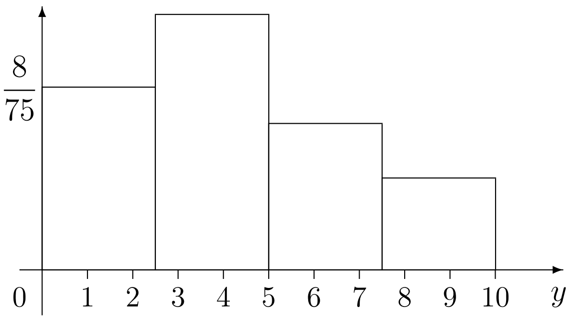

Рис. 2. Гистограмма при $k = 4$

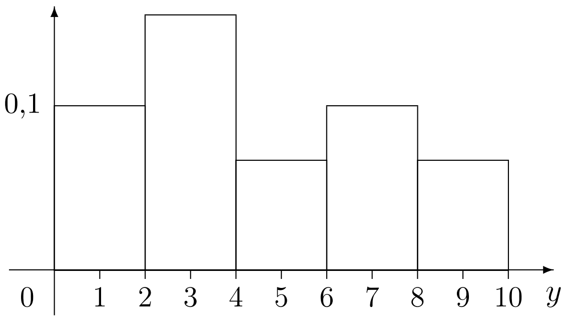

Рис. 3. Гистограмма при $k = 5$

Чем больше интервалов группировки, тем лучше: фигура, состоящая из более узких прямоугольников, точнее приближает истинную плотность распределения. С другой стороны, бессмысленно брать число интервалов

$k(n)$ порядка n: тогда в каждый интервал попадёт в среднем по одной точке и гистограмма не будет приближаться к плотности с ростом n.

Замечание 1. Справедливо следующее утверждение. *Пусть плотность распределения элементов выборки является непрерывной функцией. Если количество интервалов группировки стремится к бесконечности таким образом, что* $\frac{k(n)}{n} \to 0$, *то имеет место сходимость по вероятности гистограммы к плотности в каждой точке* y.

Обычно берут число интервалов порядка $\sqrt[3]{n}$ (или длину интервала порядка $c/\sqrt[3]{n}$).

Кроме гистограммы, для оценивания плотности используют так называемые *ядерные оценки* плотности, или *оценки Розенблата — Парзена*. Читатель может познакомиться с ними в учебнике [1, глава 1, §10]).

Выборочные моменты. Знание моментов распределения также многое может сказать о его виде и свойствах. Рассмотрим выборочные аналоги неизвестных истинных моментов распределения.

Пусть $\mathsf{E}\xi = \mathsf{E}X_1 = a$, $\mathsf{D}\xi = \mathsf{D}X_1 = \sigma^2$, $\mathsf{E}\xi^k = \mathsf{E}X_1^k = m_k$ — теоретические среднее, дисперсия, k-й момент. В качестве их оценок используем среднее, дисперсию и моменты выборочного распределения.

Истинные моменты	**Оценки для истинных моментов**
$\mathsf{E}\xi = \mathsf{E}X_1 = a$	$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — *выборочное среднее*
$\mathsf{D}\xi = \mathsf{D}X_1 = \sigma^2$	$S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — *выборочная дисперсия*, либо $S_0^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — *несмещённая выборочная дисперсия*
$\mathsf{E}\xi^k = \mathsf{E}X_1^k = m_k$	$\overline{X^k} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ — *выборочный k-й момент*
$\mathsf{E}g(\xi)$	$\overline{g(X)} = \frac{1}{n}\sum\limits_{i=1}^{n} g(X_i)$

Ещё раз напомним, что все оценки в правом столбце таблицы являются случайными величинами, если $X_1, \ldots, X_n$ — набор случайных величин, а не их реализаций на одном элементарном исходе.

§ 3. Свойства выборочных характеристик

Мы ввели три вида эмпирических характеристик, предназначенных для оценивания неизвестных теоретических характеристик распределения: эмпирическую функцию распределения, гистограмму, выборочные моменты. Если наши оценки удачны, разница между ними и истинными характеристиками должна стремиться к нулю (например, по вероятности) с ростом объёма выборки. Такое свойство выборочных характеристик называют *состоятельностью*. Убедимся, что введённые нами характеристики этим свойством обладают.

Свойства эмпирической функции распределения. Следующие четыре утверждения описывают поведение случайной функции $F_n^*(y)$.

Теорема 1. *Пусть $X_1, \ldots, X_n$ — выборка из распределения $\mathcal{F}$ с функцией распределения F и пусть F_n^* — эмпирическая функция распределения, построенная по этой выборке. Тогда $F_n^*(y) \xrightarrow{\text{p}} F(y)$ при $n \to \infty$ для любого $y \in \mathbb{R}$.*

Доказательство. По определению 2

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

Случайные величины $\mathrm{I}(X_1 < y)$, $\mathrm{I}(X_2 < y)$, $\ldots$ независимы и одинаково распределены, их математическое ожидание конечно:

$$\mathsf{E}\mathrm{I}(X_1 < y) = 1 \cdot \mathsf{P}(X_1 < y) + 0 \cdot \mathsf{P}(X_1 \geqslant y) = \mathsf{P}(X_1 < y) = F(y) < \infty,$$

поэтому можно применить ЗБЧ Хинчина (*а что это такое*?):

$$F_n^*(y) = \frac{\sum\limits_{i=1}^{n} \mathrm{I}(X_i < y)}{n} \xrightarrow{\text{p}} \mathsf{E}\mathrm{I}(X_1 < y) = F(y). \qquad \square$$

Таким образом, с ростом объёма выборки эмпирическая функция распределения сходится по вероятности к неизвестной теоретической функции распределения при любом фиксированном $y \in \mathbb{R}$. На самом деле, как показывает следующее утверждение, эта сходимость имеет даже «равномерный» характер. *Наибольшее* из расхождений между эмпирической и теоретической функциями распределения стремится к нулю.

Теорема 2 (Гливенко — Кантелли). *В условиях теоремы* 1

$$\sup_{y\in\mathbb{R}}\bigl|F_n^*(y)-F(y)\bigr| \xrightarrow{\text{p}} 0 \quad \text{при} \quad n\to\infty.$$

Более того, в теоремах 1 и Гливенко — Кантелли имеет место сходимость не только по вероятности, но и *почти наверное*.

Если функция распределения F непрерывна, то, как показывает следующая теорема, скорость сходимости к нулю в теореме Гливенко — Кантелли имеет порядок $1/\sqrt{n}$.

Теорема 3 (Колмогорова). *Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения $\mathscr{F}$ с непрерывной функцией распределения F, а F_n^* — эмпирическая функция распределения. Тогда*

$$\sqrt{n}\cdot\sup_{y\in\mathbb{R}}\bigl|F_n^*(y)-F(y)\bigr| \Rightarrow \eta \quad \text{при} \quad n\to\infty,$$

где случайная величина η имеет распределение Колмогорова с непрерывной функцией распределения

$$K(x)=\sum_{j=-\infty}^{\infty}(-1)^j e^{-2j^2x^2} \quad \text{при } x\geqslant 0, \quad K(x)=0 \quad \text{при } x<0.$$

Теоремы Гливенко — Кантелли и Колмогорова мы доказывать не будем. Доказательство первой читатель может прочесть в учебнике [1].

Следующие свойства эмпирической функции распределения — это хорошо знакомые нам свойства среднего арифметического n независимых слагаемых, имеющих распределение Бернулли.

Теорема 4. *Для любого $y\in\mathbb{R}$*

1) $\mathsf{E}F_n^*(y)=F(y)$, *т. е.* $F_n^*(y)$ — *несмещённая оценка для* $F(y)$;

2) $\mathsf{D}F_n^*(y)=\dfrac{F(y)(1-F(y))}{n}$;

3) $\sqrt{n}(F_n^*(y)-F(y))\Rightarrow \mathrm{N}_{0,\,F(y)(1-F(y))}$ *при* $F(y)\neq 0,\,1$, *т. е.* $F_n^*(y)$ — *асимптотически нормальная оценка для* $F(y)$;

4) *величина* $nF_n^*(y)$ *имеет биномиальное распределение* $\mathrm{B}_{n,F(y)}$.

Доказательство. Заметим снова, что $\mathrm{I}(X_1<y)$ имеет распределение Бернулли $\mathrm{B}_{F(y)}$ (*почему?*), поэтому

$$\mathsf{E}\mathrm{I}(X_1<y)=F(y) \quad \text{и} \quad \mathsf{D}\mathrm{I}(X_1<y)=F(y)(1-F(y)).$$

Докажем свойство (1). Случайные величины $\mathrm{I}(X_i < y)$ одинаково распределены, поэтому

$$\mathsf{E}F_n^*(y) = \mathsf{E}\,\frac{\sum\limits_{i=1}^{n}\mathrm{I}(X_i < y)}{n} = \frac{\sum\limits_{i=1}^{n}\mathsf{E}\mathrm{I}(X_i < y)}{n} = \frac{n\mathsf{E}\mathrm{I}(X_1 < y)}{n} = F(y)$$

(*где использована одинаковая распределённость*?).

Докажем свойство (2). Случайные величины $\mathrm{I}(X_i < y)$ независимы и одинаково распределены, поэтому

$$\mathsf{D}F_n^*(y) = \mathsf{D}\,\frac{\sum\limits_{i=1}^{n}\mathrm{I}(X_i < y)}{n} = \frac{\sum\limits_{i=1}^{n}\mathsf{D}\mathrm{I}(X_i < y)}{n^2} = \frac{n\mathsf{D}\mathrm{I}(X_1 < y)}{n^2} = \frac{F(y)(1-F(y))}{n}$$

(*где используется независимость*?).

Для доказательства свойства (3) используем ЦПТ (*а что это*?):

$$\sqrt{n}(F_n^*(y) - F(y)) = \sqrt{n}\left(\frac{\sum \mathrm{I}(X_i < y)}{n} - F(y)\right) = \frac{\sum\limits_{i=1}^{n}\mathrm{I}(X_i < y) - nF(y)}{\sqrt{n}} =$$

$$= \frac{\sum\limits_{i=1}^{n}\mathrm{I}(X_i < y) - n\mathsf{E}\mathrm{I}(X_1 < y)}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\,\mathsf{D}\mathrm{I}(X_1<y)} = \mathrm{N}_{0,\,F(y)(1-F(y))}.$$

Наконец, свойство (4) выполнено из-за устойчивости по суммированию биномиального распределения (*сформулировать*!). Поскольку $\mathrm{I}(X_i < y)$ независимы и имеют распределение Бернулли $\mathrm{B}_{F(y)}$, то их сумма

$$nF_n^*(y) = \mathrm{I}(X_1 < y) + \ldots + \mathrm{I}(X_n < y)$$

имеет биномиальное распределение $\mathrm{B}_{n,F(y)}$. □

Замечание 2. Все определения терминов «оценка», «несмещённость», «состоятельность», «асимптотическая нормальность» будут даны в главе II. Но смысл этих терминов должен быть понятен уже сейчас.

Упражнение. Сформулировать и доказать утверждения, аналогичные теоремам 1 и 4, для выборочного распределения

$$P_n^*(B) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in B), \quad B \in \mathfrak{B}(\mathbb{R}).$$

Свойства гистограммы. Пусть распределение $\mathcal{F}$ абсолютно непрерывно, f — его истинная плотность. Пусть, кроме того, число k интервалов группировки *не зависит от* n. Случай, когда $k = k(n)$, отмечен в замечании 1. Следующая теорема утверждает, что площадь столбца гистограммы, построенного над произвольным интервалом группировки, с ростом объёма выборки сближается с площадью области под графиком плотности над этим же интервалом.

Теорема 5. *При* $n \to \infty$ *для любого* $j = 1, \ldots, k$

$$l_j \cdot f_j = \frac{\nu_j}{n} \xrightarrow{\text{p}} \mathsf{P}(X_1 \in A_j) = \int\limits_{A_j} f(x)\,dx.$$

Упражнение. Доказать теорему 5, используя (1) и ЗБЧ Бернулли для слагаемых $\mathrm{I}(X_1 \in A_j)$, …, $\mathrm{I}(X_n \in A_j)$.

Свойства выборочных моментов. Выборочное среднее $\overline{X}$ является несмещённой, состоятельной и асимптотически нормальной оценкой для теоретического среднего (математического ожидания).

Теорема 6. 1. *Если* $\mathsf{E}|X_1| < \infty$, *то* $\mathsf{E}\overline{X} = \mathsf{E}X_1 = a$.

2. *Если* $\mathsf{E}|X_1| < \infty$, *то* $\overline{X} \xrightarrow{\text{p}} \mathsf{E}X_1 = a$ *при* $n \to \infty$.

3. *Если* $\mathsf{D}X_1 < \infty$, $\mathsf{D}X_1 \neq 0$, *то* $\sqrt{n}\left(\overline{X} - \mathsf{E}X_1\right) \Rightarrow \mathrm{N}_{0,\,\mathsf{D}X_1}$.

Доказательство. Первое утверждение следует из свойств математического ожидания:

$$\mathsf{E}\overline{X} = \frac{1}{n}(\mathsf{E}X_1 + \ldots + \mathsf{E}X_n) = \frac{1}{n} \cdot n\,\mathsf{E}X_1 = \mathsf{E}X_1 = a.$$

Из ЗБЧ в форме Хинчина получаем второе утверждение:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{p}} \mathsf{E}X_1 = a.$$

Третье утверждение есть прямое следствие ЦПТ:

$$\sqrt{n}\left(\overline{X} - \mathsf{E}X_1\right) = \frac{\sum\limits_{i=1}^{n} X_i - n\mathsf{E}X_1}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\mathsf{D}X_1}. \qquad \square$$

Замечание 3. УЗБЧ позволяет утверждать также, что при $\mathsf{E}|X_1| < \infty$ имеет место сходимость п. н. $\overline{X}$ к $\mathsf{E}X_1$. Такое свойство оценок называют *сильной состоятельностью*.

Выборочный k-й момент $\overline{X^k}$ является несмещённой, состоятельной и асимптотически нормальной оценкой для теоретического k-го момента.

Теорема 7. 1. *Если* $\mathsf{E}|X_1|^k < \infty$, *то* $\mathsf{E}\overline{X^k} = \mathsf{E}X_1^k = m_k$.

2. *Если* $\mathsf{E}|X_1|^k < \infty$, *то* $\overline{X^k} \xrightarrow{\text{p}} \mathsf{E}X_1^k = m_k$ *при* $n \to \infty$.

3. *Если* $\mathsf{D}X_1^k < \infty$, $\mathsf{D}X_1^k \neq 0$, *то* $\sqrt{n}\left(\overline{X^k} - \mathsf{E}X_1^k\right) \Rightarrow \mathrm{N}_{0,\mathsf{D}X_1^k}$.

Выборочные дисперсии обладают следующими свойствами.

Теорема 8. *Пусть* $\mathsf{D}X_1 < \infty$. 1. *Выборочные дисперсии* S^2 *и* S_0^2 *являются состоятельными оценками для истинной дисперсии:*

$$S^2 \xrightarrow{\text{p}} \mathsf{D}X_1 = \sigma^2, \quad S_0^2 \xrightarrow{\text{p}} \mathsf{D}X_1 = \sigma^2.$$

2. *Величина* S^2 *— смещённая оценка дисперсии, а* S_0^2 *— несмещённая:*

$$\mathsf{E}S^2 = \frac{n-1}{n}\mathsf{D}X_1 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad \mathsf{E}S_0^2 = \mathsf{D}X_1 = \sigma^2.$$

3. *Если* $0 \neq \mathsf{D}(X_1 - \mathsf{E}X_1)^2 < \infty$, *то* S^2 *и* S_0^2 *являются асимптотически нормальными оценками истинной дисперсии:*

$$\sqrt{n}\left(S^2 - \mathsf{D}X_1\right) \Rightarrow \mathrm{N}_{0,\mathsf{D}(X_1-\mathsf{E}X_1)^2}.$$

Доказательство. Докажем первое утверждение теоремы. Раскрыв скобки, полезно убедиться в том, что

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

Используя состоятельность первого и второго выборочных моментов и свойства сходимости по вероятности, получаем

$$S^2 = \overline{X^2} - (\overline{X})^2 \xrightarrow{\text{p}} \mathsf{E}X_1^2 - (\mathsf{E}X_1)^2 = \sigma^2.$$

Далее, $\dfrac{n}{n-1} \to 1$, поэтому $S_0^2 = \dfrac{n}{n-1}\, S^2 \xrightarrow{\text{p}} \sigma^2$.

Для доказательства второго утверждения теоремы воспользуемся формулой (2) и несмещённостью первого и второго выборочных моментов:

$$\begin{aligned}
\mathsf{E}S^2 &= \mathsf{E}\left(\overline{X^2} - (\overline{X})^2\right) = \mathsf{E}\overline{X^2} - \mathsf{E}(\overline{X})^2 = \mathsf{E}X_1^2 - \mathsf{E}(\overline{X})^2 = \\
&= \mathsf{E}X_1^2 - \left(\left(\mathsf{E}\overline{X}\right)^2 + \mathsf{D}\overline{X}\right) = \mathsf{E}X_1^2 - (\mathsf{E}X_1)^2 - \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \\
&= \sigma^2 - \frac{1}{n^2}\, n\mathsf{D}X_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2,
\end{aligned}$$

откуда сразу следует $\mathsf{E}S_0^2 = \dfrac{n}{n-1}\,\mathsf{E}S^2 = \sigma^2$.

Проверим третье утверждение теоремы. Введём случайные величины $Y_i = X_i - a$ с нулевым математическим ожиданием и такой же дисперсией $\mathsf{D}Y_1 = \mathsf{D}X_1 = \sigma^2$. Выборочную дисперсию можно представить в следующем виде:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - a - (\overline{X} - a)\right)^2 = \overline{Y^2} - \left(\overline{Y}\right)^2.$$

Тогда

$$\begin{aligned}\sqrt{n}\left(S^2-\sigma^2\right) &= \sqrt{n}\left(\overline{Y^2}-(\overline{Y})^2-\sigma^2\right)=\sqrt{n}\left(\overline{Y^2}-\mathsf{E}Y_1^2\right)-\sqrt{n}(\overline{Y})^2= \\ &= \frac{\sum\limits_{i=1}^{n}Y_i^2-n\mathsf{E}Y_1^2}{\sqrt{n}} - \overline{Y}\cdot\sqrt{n}\,\overline{Y}\Rightarrow \mathrm{N}_{0,\,\mathsf{D}(X_1-a)^2},\end{aligned}$$

поскольку первое слагаемое слабо сходится к $\mathrm{N}_{0,\,\mathsf{D}Y_1^2}$ по ЦПТ, а второе слагаемое $\overline{Y}\cdot\sqrt{n}\,\overline{Y}$ сходится к нулю слабо как произведение двух последовательностей: последовательности $\overline{Y}$, сходящейся к нулю по вероятности при $n\to\infty$ (*почему*?), и последовательности $\sqrt{n}\,\overline{Y}$, слабо сходящейся к $\mathrm{N}_{0,\,\mathsf{D}X_1}$ (*почему*?).

Слабая же сходимость к нулю влечёт сходимость по вероятности. □

Упражнение. При доказательстве дважды использовано одно из свойств слабой сходимости. Какое именно? Что такое слабая сходимость? Что такое сходимость по вероятности?

§ 4. Свойства выборочных квантилей

Квантилью заданного уровня δ распределения $\mathcal{F}$ называют такое число τ_δ, что $F(\tau_\delta)=\delta$, где $F(x)$ — функция распределения распределения $\mathcal{F}$. Это очень грубое понимание квантилей, поскольку функция распределения может ни в одной точке не равняться δ, а может равняться δ на целом отрезке точек. Дадим строгое определение.

Определение 3. *Квантилью* уровня δ распределения $\mathcal{F}$ называется любое из *чисел* τ_δ таких, что

$$F(\tau_\delta)\leqslant\delta,\quad F(\tau_\delta+0)\geqslant\delta.$$

Для случайной величины ξ из распределения $\mathcal{F}$ эти неравенства равносильны следующим:

$$\mathsf{P}(\xi<\tau_\delta)\leqslant\delta,\quad \mathsf{P}(\xi\leqslant\tau_\delta)\geqslant\delta.$$

Таким образом, квантиль уровня δ распределения $\mathcal{F}$ — любая из точек, в которой функция распределения этого распределения переходит через уровень δ.

Квантиль уровня 0,5 называется *медианой* распределения. Квантили уровней, кратных 0,01, называют *процентилями*, кратных 0,1, — *децилями*, кратных 0,25, — *квартилями*.

Выборочными квантилями называют квантили выборочного распределения. Поскольку эмпирическая функция распределения является случайной функцией, выборочные квантили суть случайные величины. Более того, эмпирическая функция распределения кусочно-постоянна, поэтому для δ, кратного $\dfrac{1}{n}$, выборочная квантиль уровня δ не будет единственной. Если $\delta = \dfrac{k}{n}$, то $F_n^*(y)$ может равняться δ на целом интервале $(X_{(k)}, X_{(k+1)}]$ (если, конечно, этот интервал не пуст). В этом случае любая точка этого интервала, включая его концы, является квантилью уровня δ. Часто в таких случаях для практических расчётов в качестве квантили берут одну точку из отрезка — левый конец, правый, середину или какую-либо иную точку. Например, $X_{([n\delta]+1)}$ — правая точка отрезка квантилей уровня δ. Она же является единственной выборочной квантилью при δ не кратном $\dfrac{1}{n}$.

Отметим одно полезное свойство выборочных квантилей уровней, близких к половине.

Теорема 9. *Пусть медиана* μ *распределения* $\mathcal{F}$ *единственна, и функция распределения* F *имеет отличную от нуля производную в точке* μ. *Тогда для любой последовательности* $c(n) = o(\sqrt{n})$ *имеет место слабая сходимость*

$$\sqrt{n}\Big(X_{\left(\left[\frac{n}{2}+c(n)\right]\right)} - \mu\Big) \Rightarrow \mathrm{N}_{0,\,\frac{1}{4(F'(\mu))^2}}.$$

Из этой теоремы следует, в частности, что *выборочная медиана*, т.е. выборочная квантиль уровня $0{,}5$, при определённых условиях является асимптотически нормальной оценкой теоретической медианы.

Доказательство. Чтобы проверить слабую сходимость к нормальному распределению, выпишем функцию распределения $F_n(y)$ левой части и покажем, что при любом значении y она сходится к функции распределения

$$\Phi_{0,\,\frac{1}{4(F'(\mu))^2}}(y) = \Phi_{0,1}(2yF'(\mu)).$$

Итак,

$$F_n(y) = \mathsf{P}\Big(\sqrt{n}\big(X_{\left(\left[\frac{n}{2}+c(n)\right]\right)} - \mu\big) < y\Big) = \mathsf{P}\Big(X_{\left(\left[\frac{n}{2}+c(n)\right]\right)} < \mu + \frac{y}{\sqrt{n}}\Big) = \mathsf{P}(A),$$

где событие A означает, что не менее чем $k = \left[\frac{n}{2} + c(n)\right]$ элементов выборки оказались левее точки $\mu + \dfrac{y}{\sqrt{n}}$. Введём независимые в совокупности случайные величины $\xi_i(n) = I\Big(X_i < \mu + \dfrac{y}{\sqrt{n}}\Big)$, имеющие распределение Бернулли с параметром $F\Big(\mu + \dfrac{y}{\sqrt{n}}\Big)$. Заметим, что это распределение зависит от n. Это означает, что у величин $\xi_1(n), \ldots, \xi_n(n)$ одно и то же распределение, но при изменении n и их количество, и распределение каждой меняется. Такую ситуацию принято называть *схемой серий*.

Далее

$$F_n(y) = \mathsf{P}\left(\sum_{i=1}^{n} \xi_i(n) \geqslant k\right) =$$

$$= 1 - \mathsf{P}\left(\frac{\sum_{i=1}^{n}\xi_i(n) - n\mathsf{E}\xi_1(n)}{\sqrt{n\mathsf{D}\xi_1(n)}} < \frac{[\frac{n}{2} + c(n)] - nF\Big(\mu + \frac{y}{\sqrt{n}}\Big)}{\sqrt{nF\Big(\mu + \frac{y}{\sqrt{n}}\Big)\Big(1 - F\Big(\mu + \frac{y}{\sqrt{n}}\Big)\Big)}}\right).$$

Величина в левой части неравенства под знаком вероятности слабо сходится к стандартному нормальному распределению. Конечно, из обычной ЦПТ это не следует, поскольку распределения слагаемых зависят от n. Предлагаем читателю внести изменения в доказательство ЦПТ и убедиться, что для данных слагаемых ЦПТ остаётся верной. В правой части неравенства стоит последовательность, зависящая от n. Чтобы найти предел вероятности, достаточно выяснить, каков предел этой последовательности, и затем воспользоваться «теоремой о двойном пределе». Рассмотрим правую часть отдельно. В силу непрерывности функции F в точке μ величина $F\Big(\mu + \dfrac{y}{\sqrt{n}}\Big)$ в знаменателе стремится к $F(\mu) = 0{,}5$. Величина $\big[\dfrac{n}{2} + c(n)\big]$ отличается от $\dfrac{n}{2} = nF(\mu)$ на величину $o(\sqrt{n})$, поэтому

$$\frac{[\frac{n}{2} + c(n)] - nF\Big(\mu + \frac{y}{\sqrt{n}}\Big)}{\sqrt{nF\Big(\mu + \frac{y}{\sqrt{n}}\Big)(1 - F\Big(\mu + \frac{y}{\sqrt{n}}\Big)}} \sim \frac{\frac{n}{2} - nF\Big(\mu + \frac{y}{\sqrt{n}}\Big)}{\sqrt{\frac{n}{4}}} = 2y\,\frac{F(\mu) - F\Big(\mu + \frac{y}{\sqrt{n}}\Big)}{\frac{y}{\sqrt{n}}}.$$

Последнее выражение по определению производной сходится к $-2yF'(\mu)$. Таким образом, $F_n(y) \to 1 - \Phi_{0,1}(-2yF'(\mu)) = \Phi_{0,1}(2yF'(\mu))$, что и требовалось доказать. □

Упражнение. Сформулировать и доказать утверждение об асимптотической нормальности порядковых статистик $X_{([n\delta + c(n)])}$ при $0 < \delta < 1$ как оценок теоретической квантили уровня δ.

§ 5. Вопросы и упражнения

1. Проверить, выполнено ли утверждение теоремы Колмогорова для выборки объёма n из распределения Бернулли B_p. Найти предельное распределение.

2. Вспомнить, как найти по функции распределения величины X_1 функцию распределения первой и последней порядковой статистики $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Выписать выражения для плотности этих порядковых статистик через функцию и плотность распределения величины X_1.

3. Из курса эконометрики: доказать, что среднее степенное

$$\left(\overline{X^k}\right)^{\frac{1}{k}} = \sqrt[k]{\frac{1}{n}\sum_{i=1}^{n} X_i^k}$$

а) сходится к $X_{(1)}$ при $k \to -\infty$; б) сходится к $X_{(n)}$ при $k \to +\infty$. Имеется в виду сходимость для любого набора чисел $X_1, \ldots, X_n$, такого, что среднее степенное определено, т. е. сходимость п. н.

Указание. Вынести $X_{(1)}$ или $X_{(n)}$ из-под корня, применить лемму о двух милиционерах и свойства $\sqrt[k]{k} \to 1$, $\sqrt[k]{1} \to 1$ при $k \to \infty$.

4. Пусть $x_1 \geqslant 0, \ldots, x_n \geqslant 0$ — произвольные неотрицательные числа. Доказать, что в этом случае числовая последовательность

$$\sqrt[k]{\overline{x^k}} = \sqrt[k]{\frac{x_1^k + \ldots + x_n^k}{n}}, \quad k = 1, 2, 3, \ldots$$

не убывает по k. Воспользоваться неравенством Йенсена.

5. Пусть дана выборка $X_1, \ldots, X_n$ такая, что $X_1 \geqslant 0$ п. н. Доказать, что в этом случае последовательность случайных величин

$$\sqrt[k]{\overline{X^k}}, \quad k = 1, 2, 3, \ldots$$

почти наверное не убывает по k. Воспользоваться результатом предыдущей задачи.

6. Доказать теорему 7.

7. Объяснить термины: «оценка», «несмещённость», «состоятельность», «асимптотическая нормальность» оценки.

ГЛАВА II

ТОЧЕЧНОЕ ОЦЕНИВАНИЕ

Ситуация, когда о распределении наблюдений не известно совсем ничего, встречается довольно редко. Проводя эксперимент, мы можем предполагать или утверждать что-либо о распределении его результатов. Например, может оказаться, что это распределение нам известно с точностью до значений одного или нескольких числовых параметров. Так, в широких предположениях рост юношей одного возраста имеет нормальное распределение с неизвестными средним и дисперсией, а число покупателей в магазине в течение часа — распределение Пуассона с неизвестной «интенсивностью» λ. Рассмотрим задачу оценивания по выборке неизвестных параметров распределения. Оказывается, различными способами бывает возможно построить даже не одну, а множество оценок для одного и того же неизвестного параметра.

§ 1. Точечные оценки и их свойства

Параметрические семейства распределений. Пусть имеется выборка $X_1, \ldots, X_n$ объёма n, извлечённая из распределения $\mathcal{F}_\theta$, которое известным образом зависит от неизвестного параметра θ.

Здесь $\mathcal{F}_\theta$ — некий класс распределений, целиком определяющихся значением скалярного или векторного параметра θ. Параметр θ принимает значения из некоторого множества Θ, которое мы будем называть множеством возможных значений параметра.

Примерами параметрических семейств распределений могут служить все известные нам распределения: распределение Пуассона Π_λ, где $\lambda > 0$; распределение Бернулли B_p, где $p \in (0, 1)$; равномерное распределение $\mathrm{U}_{a,b}$, где $a < b$; равномерное распределение $\mathrm{U}_{0,\theta}$, где $\theta > 0$; нормальное распределение N_{a,σ^2}, где $a \in \mathbb{R}$, $\sigma > 0$ и т. д.

Точечные оценки. Итак, пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta$.

Определение 4. *Статистикой* называется произвольная борелевская функция $\theta^* = \theta^*(X_1, \ldots, X_n)$ от элементов выборки.

Замечание 4. Статистика есть функция от эмпирических данных, но никак не от параметра θ. Статистика, как правило, предназначена именно для оценивания неизвестного параметра θ (поэтому её иначе называют *оценкой*) и уже поэтому от него зависеть не может.

Статистика есть не *любая*, а *измеримая* функция от выборки (*борелевская*, для которой прообраз любого борелевского множества из $\mathbb{R}$ есть снова борелевское множество в $\mathbb{R}^n$), иначе оценка θ^* не будет случайной величиной. Далее мы всюду будем иметь дело только с измеримыми функциями, и отдельно это оговаривать не будем.

Свойства оценок. Дадим три определения хороших свойств оценок.

Определение 5. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется *несмещённой* оценкой параметра θ, если для любого $\theta \in \Theta$ выполнено равенство $\mathsf{E}\theta^* = \theta$.

Определение 6. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется *асимптотически несмещённой* оценкой параметра θ, если для любого $\theta \in \Theta$ имеет место сходимость $\mathsf{E}\theta^* \to \theta$ при $n \to \infty$.

Определение 7. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется *состоятельной* оценкой параметра θ, если для любого $\theta \in \Theta$ имеет место сходимость $\theta^* \xrightarrow{\mathrm{p}} \theta$ при $n \to \infty$.

Несмещённость — свойство оценок при фиксированном n. Означает это свойство отсутствие ошибки «в среднем», т. е. при систематическом использовании данной оценки. Несмещённость является желательным, но не обязательным свойством оценок. Достаточно, чтобы смещение оценки (разница между её средним значением и истинным параметром) уменьшалось с ростом объёма выборки. Поэтому асимптотическая несмещённость является весьма желательным свойством оценок. Свойство состоятельности означает, что последовательность оценок приближается к неизвестному параметру при увеличении количества наблюдений. В отсутствие этого свойства оценка совершенно «несостоятельна» как оценка.

Пример 3. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения N_{a,σ^2}, где $a \in \mathbb{R}$, $\sigma > 0$. Как найти оценки для параметров a и σ^2, если оба эти параметра (можно считать это и одним двумерным параметром) неизвестны?

Мы уже знаем хорошие оценки для математического ожидания и дисперсии любого распределения. Оценкой для истинного среднего $a = \mathsf{E}X_1$ может служить выборочное среднее $a^* = \overline{X}$. Теорема 6 (с. 17) утверждает, что эта оценка несмещённая и состоятельная.

Для дисперсии $\sigma^2 = \mathsf{D}X_1$ у нас есть сразу две оценки:

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

Как показано в теореме 8 (с. 18), обе эти оценки состоятельны, и одна из них — несмещённая (*которая?*), а другая — асимптотически несмещённая.

§ 2. Метод моментов

Рассмотрим некоторые стандартные методы получения точечных оценок. Метод моментов предлагает для нахождения оценки неизвестного параметра использовать выборочные моменты вместо истинных. Этот метод заключается в следующем: любой момент случайной величины X_1 (например, k-й) является функцией от параметра θ. Но тогда и параметр θ может оказаться функцией от теоретического k-го момента. Подставив в эту функцию вместо неизвестного теоретического k-го момента его выборочный аналог, получим вместо параметра θ его оценку θ^*.

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta \subseteq \mathbb{R}$. Выберем некоторую функцию $g(y) : \mathbb{R} \to \mathbb{R}$ так, чтобы существовал момент

$$\mathsf{E}g(X_1) = h(\theta) \tag{3}$$

и функция h была обратима в области Θ.

Решим уравнение (3) относительно θ, а затем *вместо истинного момента возьмём выборочный*:

$$\theta = h^{-1}\left(\mathsf{E}g(X_1)\right), \qquad \theta^* = h^{-1}\left(\overline{g(X)}\right) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right).$$

Полученная таким образом оценка θ^* называется *оценкой метода моментов* (ОММ) для параметра θ.

Чаще всего в качестве функции $g(y)$ берут $g(y) = y^k$. В этом случае

$$\mathsf{E}X_1^k = h(\theta), \quad \theta = h^{-1}\left(\mathsf{E}X_1^k\right), \quad \theta^* = h^{-1}\left(\overline{X^k}\right) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right),$$

если, конечно, функция h обратима в области Θ.

Можно сказать, что при построении оценки метода моментов мы берём в качестве оценки такое (случайное) значение параметра θ, при котором истинный момент совпадает с выборочным.

Замечание 5. Если параметр $\theta = (\theta_1, \ldots, \theta_k)$ — вектор, а не скаляр и принимает значения в множестве $\Theta \subseteq \mathbb{R}^k$, то в качестве функции g берут вектор-функцию $g(y) = (g_1(y), \ldots, g_k(y))$. Тогда равенство $\mathsf{E}g(X_1) = h(\theta)$ представляет из себя систему из k уравнений, которая должна быть однозначно разрешима относительно $\theta_1, \ldots, \theta_k$. Решая эту систему и подставляя вместо истинных моментов $\mathsf{E}g_i(X_1)$ выборочные моменты $\overline{g_i(X)}$, получают ОММ для $\theta_1, \ldots, \theta_k$.

Пример 4. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного на отрезке $[0, \theta]$ распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$.

Найдём оценку метода моментов θ_1^* по первому моменту, т. е. с помощью функции $g(y) = y$:

$$\mathsf{E}X_1 = \frac{\theta}{2}, \qquad \theta = 2\mathsf{E}X_1, \qquad \theta_1^* = 2\overline{X}.$$

Найдём оценку метода моментов θ_k^* по k-му моменту:

$$\mathsf{E}X_1^k = \int\limits_0^\theta y^k \, \frac{1}{\theta} \, dy = \frac{\theta^k}{k+1}, \qquad \theta = \sqrt[k]{(k+1)\mathsf{E}X_1^k},$$

тогда

$$\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}. \tag{4}$$

Пример 5. Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения Пуассона Π_λ с параметром $\lambda > 0$. Введём новый параметр

$$\theta = \theta(\lambda) = \mathsf{P}(X_1 = 1) = \lambda \, e^{-\lambda}$$

и найдём ОММ для параметра θ с помощью функции $g(y) = \mathrm{I}(y = 1)$:

$$\mathsf{E}g(X_1) = \mathsf{E}\mathrm{I}(X_1 = 1) = \mathsf{P}(X_1 = 1) = \lambda \, e^{-\lambda} = \theta,$$

$$\theta^* = \overline{I(X = 1)} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(X_i = 1).$$

Заметим, что найти оценку для параметра λ с помощью функции $g(y) = \mathrm{I}(y = 1)$ нельзя: равенство $\mathsf{E}g(X_1) = \lambda \, e^{-\lambda}$ не является однозначно разрешимым относительно λ в области $\lambda > 0$. Оценку для параметра λ можно найти по первому моменту: $\mathsf{E}X_1 = \lambda$, поэтому $\lambda^* = \overline{X}$.

Замечание 6. Может случиться так, что $\theta^* = h^{-1}\big(\overline{g(X)}\big) \notin \Theta$. В этом случае оценку корректируют. Например, в качестве ОММ берут ближайшую к $h^{-1}\big(\overline{g(X)}\big)$ точку из Θ или из замыкания Θ.

Пример 6. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения $\mathrm{N}_{a,1}$ с неотрицательным средним $a \geqslant 0$.

Ищем оценку для a по первому моменту: $\mathsf{E}X_1 = a$, поэтому $a^* = \overline{X}$. Однако по условию $a \geqslant 0$, тогда как $\overline{X}$ может быть и отрицательно. Если $\overline{X} < 0$, то в качестве оценки для a более подойдет 0. Если же $\overline{X} > 0$, в качестве оценки нужно брать $\overline{X}$. Итого: $a^* = \max\{0, \overline{X}\}$ — «исправленная» оценка метода моментов.

§ 3. Свойства оценок метода моментов

Теорема 10. *Пусть $\theta^* = h^{-1}\bigl(\overline{g(X)}\bigr)$ — оценка параметра θ, полученная по методу моментов, причём функция h^{-1} непрерывна. Тогда оценка θ^* состоятельна.*

Доказательство. По ЗБЧ Хинчина имеем

$$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{\mathrm{p}} \mathsf{E}g(X_1) = h(\theta).$$

Поскольку функция h^{-1} непрерывна, то и

$$\theta^* = h^{-1}\bigl(\overline{g(X)}\bigr) \xrightarrow{\mathrm{p}} h^{-1}(\mathsf{E}g(X_1)) = h^{-1}(h(\theta)) = \theta. \qquad \square$$

Напомним, что для обратимой функции $h : \mathbb{R} \to \mathbb{R}$ непрерывность h и непрерывность h^{-1} равносильны.

Если полученные разумным путём оценки обязаны быть состоятельными, то свойство несмещённости — скорее исключение, нежели правило.

Действительно, несмещённость ОММ вида $\theta^* = h^{-1}\bigl(\overline{g(X)}\bigr)$ означала бы, что при всех $\theta \in \Theta$ выполнено равенство

$$\mathsf{E}h^{-1}\bigl(\overline{g(X)}\bigr) = \theta = h^{-1}\left(h(\theta)\right) = h^{-1}\bigl(\mathsf{E}\overline{g(X)}\bigr). \tag{5}$$

Но функция h^{-1} очень часто оказывается выпуклой или вогнутой. В этом случае из доказательства неравенства Йенсена можно сделать вывод (*сделайте его*!): между левой и правой частью в (5) равенство возможно, лишь если случайная величина $\overline{g(X)}$ вырождена либо если функция h^{-1} линейна на множестве значений этой случайной величины и её математического ожидания.

Пример 7. Рассмотрим последовательность оценок для неизвестного параметра θ равномерного на отрезке $[0, \theta]$ распределения, полученную в примере 4, и исследуем напрямую их свойства.

Проверим состоятельность всех оценок. По ЗБЧ Хинчина

$$\overline{X^k} \xrightarrow{\text{p}} \mathsf{E}X_1^k = \frac{\theta^k}{k+1} \quad \text{при} \ \ n \to \infty.$$

Функция $\sqrt[k]{(k+1)y}$ непрерывна для всех $y > 0$, поэтому при $n \to \infty$

$$\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}} \xrightarrow{\text{p}} \sqrt[k]{(k+1)\frac{\theta^k}{k+1}} = \theta.$$

Упражнение. Зачем нужна непрерывность функции $\sqrt[k]{(k+1)y}$?

Проверим несмещённость полученных оценок. По определению

$$\mathsf{E}\theta_1^* = \mathsf{E}2\overline{X} = 2\mathsf{E}\overline{X} = 2\theta/2 = \theta,$$

т. е. оценка $\theta_1^* = 2\overline{X}$ несмещённая. Рассмотрим оценку θ_2^*. Её математическое ожидание равно

$$\mathsf{E}\theta_2^* = \mathsf{E}\sqrt{3\overline{X^2}}.$$

Чтобы внести знак математического ожидания под корень, воспользуемся неравенством Йенсена. Функция $g(y) = \sqrt{y}$ строго вогнута в области $y > 0$, а случайная величина $3\overline{X^2}$ имеет невырожденное распределение. Поэтому (*обратите внимание на знак*!)

$$\mathsf{E}\theta_2^* = \mathsf{E}\sqrt{3\overline{X^2}} < \sqrt{3\mathsf{E}\overline{X^2}} = \sqrt{3\mathsf{E}X_1^2} = \theta.$$

Итак, оценка $\theta_2^* = \sqrt{3\overline{X^2}}$ — смещённая. Такими же смещёнными будут и оценки θ_k^* при всех $k > 2$ (*докажите*!).

§ 4. Метод максимального правдоподобия

Метод максимального правдоподобия — ещё один разумный способ построения оценки неизвестного параметра. Состоит он в том, что в качестве «наиболее правдоподобного» значения параметра берут значение θ, максимизирующее вероятность получить при n опытах данную выборку $\vec{X} = (X_1, \ldots, X_n)$. Это значение параметра θ зависит от выборки и является искомой оценкой.

Выясним сначала, что такое «вероятность получить данную выборку», т. е. что́ именно нужно максимизировать. Вспомним, что для абсолютно непрерывных распределений $\mathcal{F}_\theta$ их плотность $f_\theta(y)$ — «почти» (с точностью до dy) вероятность попадания в точку y:

$$\mathsf{P}(X_1 \in (y,\, y+dy)) = f_\theta(y)\,dy.$$

А для дискретных распределений $\mathcal{F}_\theta$ вероятность попасть в точку y равна $\mathsf{P}_\theta(X_1 = y)$. В зависимости от типа распределения $\mathcal{F}_\theta$ обозначим через $f_\theta(y)$ одну из следующих двух функций:

$$f_\theta(y) = \begin{cases} \text{плотность } f_\theta(y), & \text{если } \mathcal{F}_\theta \text{ абсолютно непрерывно,} \\ \mathsf{P}_\theta(X_1 = y), & \text{если } \mathcal{F}_\theta \text{ дискретно.} \end{cases} \tag{6}$$

В дальнейшем функцию $f_\theta(y)$, определённую в (6), мы будем называть *плотностью* распределения $\mathcal{F}_\theta$ независимо от того, является ли это распределение дискретным или абсолютно непрерывным.

Определение 8. Функция

$$f(\vec{X};\, \theta) = f_\theta(X_1) \cdot f_\theta(X_2) \cdot \ldots \cdot f_\theta(X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

называется *функцией правдоподобия*. При фиксированном θ эта функция является случайной величиной. Функция (тоже случайная)

$$L(\vec{X};\, \theta) = \ln f(\vec{X};\, \theta) = \sum_{i=1}^{n} \ln f_\theta(X_i)$$

называется *логарифмической* функцией правдоподобия.

В дискретном случае при фиксированных $x_1, \ldots, x_n$ значение функции правдоподобия $f(x_1, \ldots, x_n, \theta)$ равно вероятности, с которой выборка $X_1, \ldots, X_n$ в данной серии экспериментов принимает значения $x_1, \ldots, x_n$. Эта вероятность меняется в зависимости от θ:

$$\begin{aligned} f(\vec{x};\, \theta) = \prod_{i=1}^{n} f_\theta(x_i) &= \mathsf{P}_\theta(X_1 = x_1) \cdot \ldots \cdot \mathsf{P}_\theta(X_n = x_n) = \\ &= \mathsf{P}_\theta(X_1 = x_1, \ldots, X_n = x_n). \end{aligned}$$

В абсолютно непрерывном случае эта функция пропорциональна вероятности попасть «почти» в точку $x_1, \ldots, x_n$, а именно в «кубик» со сторонами $dx_1, \ldots, dx_n$ вокруг точки $x_1, \ldots, x_n$.

Определение 9. *Оценкой максимального правдоподобия* (ОМП) $\hat{\theta}$ для неизвестного параметра θ называют такое значение θ, при котором достигается максимум функции $f(\vec{X};\, \theta)$.

Замечание 7. Поскольку функция $\ln y$ монотонна, то точки максимума функций $f(\vec{X};\, \theta)$ и $L(\vec{X};\, \theta)$ совпадают (*обоснуйте*!). Поэтому оценкой максимального правдоподобия можно называть точку максимума (по переменной θ) функции $L(\vec{X};\, \theta)$.

Напомним, что точки экстремума функции — это либо точки, в которых производная обращается в нуль, либо точки разрыва функции или её производной, либо крайние точки области определения функции.

Смысл метода максимального правдоподобия состоит в следующем. Вероятность получить в n экспериментах выборку $X_1, \ldots, X_n$, описываемая функцией правдоподобия, может быть больше или меньше в зависимости от θ. Но выборка дана. Какое значение параметра следует выбрать в качестве оценки? Видимо, то, при котором вероятность получить эту выборку оказывается наибольшей. Поэтому в качестве оценки максимального правдоподобия и выбирается значение параметра θ, при котором максимальна функция правдоподобия.

Пример 8. Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения Пуассона Π_λ, где $\lambda > 0$. Найдём ОМП $\hat\lambda$ для неизвестного параметра λ. Здесь

$$f_\lambda(y) = \mathsf{P}(X_1 = y) = \frac{\lambda^y}{y!}\, e^{-\lambda}, \qquad y = 0,\, 1,\, 2,\, \ldots,$$

поэтому функция правдоподобия равна

$$f(\vec{X};\, \lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!}\, e^{-\lambda} = \frac{\lambda^{\Sigma X_i}}{\prod X_i!}\, e^{-n\lambda} = \frac{\lambda^{n\overline{X}}}{\prod X_i!}\, e^{-n\lambda}.$$

Поскольку эта функция при всех $\lambda > 0$ непрерывно дифференцируема по λ, можно искать точки экстремума, приравняв к нулю частную производную по λ. Но удобнее это делать для логарифмической функции правдоподобия:

$$L(\vec{X};\, \lambda) = \ln f(\vec{X};\, \lambda) = \ln\left(\frac{\lambda^{n\overline{X}}}{\prod X_i!}\, e^{-n\lambda}\right) = n\overline{X}\ln\lambda - \ln\prod_{i=1}^{n} X_i! - n\lambda.$$

Тогда

$$\frac{\partial}{\partial\lambda} L(\vec{X}, \lambda) = \frac{n\overline{X}}{\lambda} - n.$$

Точку экстремума $\hat\lambda = \overline{X}$ находим как решение уравнения $\dfrac{n\overline{X}}{\lambda} - n = 0$.

Упражнение. Проверить, что $\hat\lambda = \overline{X}$ — точка максимума, а не минимума. Убедиться, что $\hat\lambda = \overline{X}$ совпадает с одной из оценок метода моментов (*полученной по какому моменту*?), т. е. в данном случае новой оценки мы не получили.

Пример 9. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения N_{a,σ^2}, где $a \in \mathbb{R}$, $\sigma > 0$ — два неизвестных параметра.

Это распределение имеет плотность

$$f_{(a,\sigma^2)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(y-a)^2/2\sigma^2}.$$

Перемножив плотности в точках $X_1, \ldots, X_n$, получим функцию правдоподобия

$$f(\vec{X};\, a,\, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(X_i-a)^2/2\sigma^2} = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}}\, e^{-\sum(X_i-a)^2/2\sigma^2},$$

а затем логарифмическую функцию правдоподобия

$$L(\vec{X};\, a,\, \sigma^2) = \ln f(\vec{X};\, a,\, \sigma^2) = -\ln(2\pi)^{n/2} - \frac{n}{2}\,\ln\sigma^2 - \frac{\sum\limits_{i=1}^{n}(X_i-a)^2}{2\sigma^2}.$$

В точке экстремума (по a и σ^2) гладкой функции L обращаются в нуль обе частные производные

$$\begin{cases} \dfrac{\partial}{\partial a}\, L(\vec{X};\, a,\, \sigma^2) & = \dfrac{2\sum(X_i-a)}{2\sigma^2} = \dfrac{n\overline{X}-na}{\sigma^2}, \\ \dfrac{\partial}{\partial \sigma^2}\, L(\vec{X};\, a,\, \sigma^2) & = -\dfrac{n}{2\sigma^2} + \dfrac{\sum(X_i-a)^2}{2\sigma^4}. \end{cases}$$

Оценка максимального правдоподобия для (a, σ^2) является решением системы уравнений

$$\frac{n\overline{X}-na}{\sigma^2} = 0, \qquad -\frac{n}{2\sigma^2} + \frac{\sum(X_i-a)^2}{2(\sigma^2)^2} = 0.$$

Решая, получаем хорошо знакомые оценки

$$\hat{a} = \overline{X}, \qquad \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2.$$

Упражнение. Проверить, что $(\overline{X},\, S^2)$ — точка максимума, а не минимума. Убедиться, что эти оценки совпадают с некоторыми оценками метода моментов.

Пример 10. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Тогда $\hat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$ (см. [5, пример 4.4] или [1, пример 5, с. 91]).

Пример 11. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{\theta,\,\theta+5}$, где $\theta \in \mathbb{R}$ (см. также [1, пример 4, с. 91]).

Плотность этого распределения равна

$$f_\theta(y) = \begin{cases} \dfrac{1}{5}, & \text{если } y \in [\theta,\, \theta+5], \\ 0 & \text{иначе.} \end{cases}$$

Запишем функцию правдоподобия

$$\begin{aligned} f(\vec{X};\, \theta) &= \begin{cases} \dfrac{1}{5^n}, & \text{если } \theta \leqslant X_i \leqslant \theta+5 \text{ для любого } i, \\ 0 & \text{иначе} \end{cases} = \\ &= \begin{cases} \dfrac{1}{5^n}, & \theta \leqslant X_{(1)} \leqslant X_{(n)} \leqslant \theta+5, \\ 0 & \text{иначе.} \end{cases} \end{aligned}$$

Представим функцию $f(\vec{X};\, \theta)$ как функцию переменной θ:

$$f(\vec{X};\, \theta) = \begin{cases} \dfrac{1}{5^n}, & \text{если } X_{(n)} - 5 \leqslant \theta \leqslant X_{(1)}, \\ 0 & \text{иначе.} \end{cases}$$

Функция правдоподобия (рис. 4) постоянна на целом отрезке. Поэтому она достигает своего максимального значения $1/5^n$ во всех точках θ, принадлежащих отрезку $[X_{(n)} - 5,\, X_{(1)}]$.

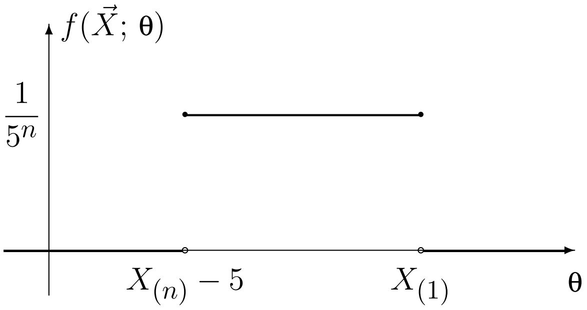

Рис. 4. Функция правдоподобия в примере 11

Любая точка $\hat{\theta} \in [X_{(n)} - 5,\, X_{(1)}]$ может служить оценкой максимального правдоподобия. Например, оценками максимального правдоподобия являются линейные комбинации концов отрезка:

$$\hat{\theta}_\alpha = (1-\alpha)(X_{(n)} - 5) + \alpha X_{(1)}, \quad \text{где } \alpha \in [0,\, 1],$$

в том числе $\hat{\theta}_0 = X_{(n)} - 5$ и $\hat{\theta}_1 = X_{(1)}$ — концы отрезка.

Упражнение. Проверить, что отрезок $[X_{(n)} - 5,\, X_{(1)}]$ не пуст. Найти оценку метода моментов (по первому моменту). Найти ОМП параметра θ равномерного распределения $\mathrm{U}_{\theta,\,2\theta}$.

Пример 12. Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли B_p с параметром $p \in (0,1)$. Построим ОМП для параметра p.

Можно получить ОМП для неизвестного параметра распределения Бернулли как частный случай ОМП для параметра p биномиального распределения $\mathrm{B}_{m,p}$ с известным m. Но мы попробуем записать явным образом функцию правдоподобия для распределения Бернулли.

Функция $f_p(y) = \mathsf{P}(X_1 = y)$ принимает два значения: p и $1-p$ в зависимости от того, равно y единице или нулю. Соответственно, каждый сомножитель $f_p(X_i)$ в функции правдоподобия равен либо p (если $X_i = 1$), либо $1 - p$ (если $X_i = 0$). Количество сомножителей, равных p, равно числу элементов выборки, равных единице, и равно $X_1 + \ldots + X_n = n\overline{X}$. Поэтому

$$f(\vec{X};\, p) = p^{n\overline{X}}(1-p)^{n-n\overline{X}}.$$

Далее находим логарифмическую функцию правдоподобия и точку экстремума этой функции по переменной p. Получаем $\hat{p} = \overline{X}$.

Упражнение. Довести до конца вычисления ОМП.

§ 5. Вопросы и упражнения

1. Дана выборка $X_1, \ldots, X_n$ из распределения Бернулли B_p с параметром $p \in (0,1)$. Проверить, что X_1, X_1X_2, $X_1(1 - X_2)$ являются несмещёнными оценками соответственно для p, p^2, $p(1-p)$. Являются ли эти оценки состоятельными?

2. Дана выборка $X_1, \ldots, X_n$ из распределения Пуассона Π_λ с параметром $\lambda > 0$. Проверить, что X_1 и $\mathrm{I}(X_1 = k)$ являются несмещёнными оценками соответственно для λ и $\dfrac{\lambda^k}{k!}\, e^{-\lambda}$. Являются ли эти оценки состоятельными?

3. Дана выборка $X_1, \ldots, X_n$ из распределения Пуассона Π_λ с параметром $\lambda > 0$. Проверить, что

$$\theta^* = \frac{n}{n-1}\overline{X}\left(1 - \frac{1}{n}\right)^{n\overline{X}}$$

является несмещённой оценкой для параметра $\theta = \lambda e^{-\lambda} = \mathsf{P}(X_1 = 1)$.

4. Дана выборка $X_1, \ldots, X_n$ из равномерного распределения $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Проверить состоятельность и несмещённость оценок $\theta^* = X_{(n)}$, $\theta^{**} = X_{(n)} + X_{(1)}$ для параметра θ.

5. Построить оценки неизвестных параметров по методу моментов для неизвестных параметров следующих семейств распределений: B_p — по первому моменту, Π_λ — по первому и второму моменту, $\mathrm{U}_{a,b}$ — по первому и второму моменту, E_α — по всем моментам, $\mathrm{E}_{1/\alpha}$ — по первому моменту, $\mathrm{U}_{-\theta,\theta}$ — как получится, $\Gamma_{\alpha,\lambda}$ — по первому и второму моменту, N_{a,σ^2} (для σ^2 при a известном и при a неизвестном).

6. Построить оценки неизвестных параметров по методу максимального правдоподобия для следующих семейств распределений: $\mathrm{B}_{m,p}$ при известном значении $m \in \mathbb{N}$, $\Pi_{\lambda+1}$, $\mathrm{U}_{0,2\theta}$, $\mathrm{E}_{2\alpha+3}$, $\mathrm{U}_{-\theta,\theta}$, N_{a,σ^2} при известном a.

7. Какие из оценок в задачах 5 и 6 несмещённые? Какие из них состоятельны?

8. Эмпирическая функция распределения $F_n^*(y)$ строится по выборке из равномерного распределения на отрезке $[0, a]$, где $a > 1$. Для какого параметра $\theta = \theta(a)$ статистика $F_n^*(1)$ является несмещённой оценкой? Является ли она состоятельной оценкой того же параметра?

9. Пусть элементы выборки $X_1, \ldots, X_n$ имеют распределение с плотностью

$$f_\theta(y) = \begin{cases} 3\theta y^2 e^{-\theta y^3}, & \text{если } y > 0, \\ 0, & \text{если } y \leqslant 0, \end{cases}$$

где $\theta > 0$ — неизвестный параметр. Проверить, является ли оценка максимального правдоподобия асимптотически несмещённой оценкой для параметра θ.

10. Дана выборка из равномерного распределения на отрезке $[0, \theta]$. Найти предел (в смысле сходимости п. н.) при $k \to \infty$ последовательности оценок параметра θ, полученных методом моментов по k-му моменту.

ГЛАВА III

СРАВНЕНИЕ ОЦЕНОК

Используя метод моментов и метод максимального правдоподобия, мы получили для каждого параметра достаточно много различных оценок. Каким же образом их сравнивать? Что должно быть показателем «хорошести» оценки? Понятно, что чем дальше оценка отклоняется от параметра, тем она хуже. Но величина $|\theta^* - \theta|$ для сравнения непригодна: во-первых, параметр θ неизвестен, во-вторых, θ^* — случайная величина, поэтому при разных значениях выборки эти расстояния будут, вообще говоря, различны. Для сравнения оценок используют обычно усреднённые характеристики рассеяния. Это может быть либо $\mathsf{E}(\theta^* - \theta)^2$, либо некий «предельный» разброс последовательности оценок относительно параметра. В зависимости от этого различают среднеквадратический и асимптотический подходы к сравнению оценок.

§ 1. Среднеквадратический подход к сравнению оценок

Среднеквадратический подход использует в качестве «расстояния» от оценки до параметра величину $\mathsf{E}(\theta^* - \theta)^2$. Это среднее является функцией от параметра θ, поэтому сравнивают такие отклонения при каждом возможном значении θ.

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta$.

Определение 10. Говорят, что оценка θ_1^* *лучше* оценки θ_2^* в смысле *среднеквадратического подхода* (или *в среднеквадратичном*), если для любого $\theta \in \Theta$

$$\mathsf{E}(\theta_1^* - \theta)^2 \leqslant \mathsf{E}(\theta_2^* - \theta)^2,$$

и хотя бы при одном θ это неравенство строгое.

Существует ли среди всех оценок наилучшая в смысле среднеквадратического подхода? Здравый смысл подсказывает, что ответ на этот вопрос скорее отрицателен, равно как и на любой вопрос о существовании слишком глобального экстремума. Предположим, что мы имеем дело с *невырожденной задачей*: ни для какой статистики θ^* невозможно, чтобы оценка θ^* равнялась θ п. н. при любых $\theta \in \Theta$.

Теорема 11. *В классе всех возможных оценок наилучшей в смысле среднеквадратического подхода оценки не существует.*

Доказательство. Пусть, напротив, θ^* — наилучшая, т. е. для любой другой оценки θ_1^*, при любом $\theta \in \Theta$ выполнено

$$\mathsf{E}(\theta^* - \theta)^2 \leqslant \mathsf{E}(\theta_1^* - \theta)^2.$$

Пусть θ_1 — произвольная точка области Θ. Рассмотрим статистику $\theta_1^* \equiv \theta_1$. Тогда $\mathsf{E}(\theta^* - \theta)^2 \leqslant \mathsf{E}(\theta_1 - \theta)^2$ при любом $\theta \in \Theta$. В частности, при $\theta = \theta_1$ получим $\mathsf{E}(\theta^* - \theta_1)^2 \leqslant \mathsf{E}(\theta_1 - \theta_1)^2 = 0$. Поэтому $\mathsf{E}(\theta^* - \theta_1)^2 = 0$. Поскольку θ_1 произвольно, равенство $\mathsf{E}(\theta^* - \theta)^2 = 0$ имеет место при любом $\theta \in \Theta$.

Но величина $(\theta^* - \theta)^2$ неотрицательна, поэтому равенство нулю её математического ожидания возможно только в случае, когда $\theta^* = \theta$ п. н. (оценка в точности отгадывает неизвестный параметр), т. е. для вырожденной с точки зрения математической статистики задачи. □

Приведём примеры вырожденных задач, когда параметр однозначно определяется по выборке. Так, для выборки из вырожденного распределения I_θ параметр совпадает с любым элементом выборки: $\theta = X_1$ п. н.; для выборки из равномерного распределения $\mathrm{U}_{\theta,\,\theta+1}$ при $\Theta = \mathbb{Z}$ имеет место равенство $\theta = \left[X_1\right]$ п. н.

Если в классе *всех* оценок наилучшей не существует, то, возможно, следует разбить класс всех оценок на отдельные подклассы и в каждом искать наилучшую. Обычно рассматривают оценки, имеющие одинаковое *смещение* $b(\theta) = \mathsf{E}\theta^* - \theta$.

Обозначим через $K_b = K_{b(\theta)}$ класс всех оценок со смещением, равным заданной функции $b(\theta)$:

$$K_b = \{\theta^* \mid \mathsf{E}\theta^* = \theta + b(\theta)\}, \quad K_0 = \{\theta^* \mid \mathsf{E}\theta^* = \theta\}.$$

Здесь K_0 — класс несмещённых оценок.

Определение 11. Оценка $\theta^* \in K_b$ называется *эффективной* оценкой в классе K_b, если она лучше (не хуже) всех других оценок класса K_b в смысле среднеквадратического подхода, т. е. для любой $\theta_1^* \in K_b$, для любого $\theta \in \Theta$

$$\mathsf{E}(\theta^* - \theta)^2 \leqslant \mathsf{E}(\theta_1^* - \theta)^2.$$

Определение 12. Эффективная оценка в классе K_0 называется просто *эффективной*.

Замечание 8. Для оценки $\theta^* \in K_0$ по определению дисперсии

$$\mathsf{E}(\theta^* - \theta)^2 = \mathsf{E}(\theta^* - \mathsf{E}\theta^*)^2 = \mathsf{D}\theta^*,$$

т. е. сравнение в среднеквадратичном несмещённых оценок есть просто сравнение их дисперсий. Поэтому эффективную оценку ещё называют «несмещённой оценкой с равномерно минимальной дисперсией» («н. о. р. м. д.»). Равномерность имеется в виду по всем $\theta \in \Theta$.

Для смещённых оценок $\theta^* \in K_b$

$$\mathsf{E}(\theta^* - \theta)^2 = \mathsf{D}(\theta^* - \theta) + (\mathsf{E}\theta^* - \theta)^2 = \mathsf{D}\theta^* + b^2(\theta),$$

т. е. сравнение в среднеквадратичном оценок с одинаковым смещением также приводит к сравнению их дисперсий.

Упражнение. Мы хотим найти наилучшую оценку в классе K_b. Объясните, почему доказательство теоремы 11 не пройдет в классе K_b.

Следующее утверждение показывает, что в классе оценок с одинаковым смещением не может существовать двух различных эффективных оценок: если эффективная оценка существует, она единственна.

Теорема 12. *Если* $\theta_1^* \in K_b$ *и* $\theta_2^* \in K_b$ *— две эффективные оценки в классе* K_b, *то с вероятностью* 1 *они совпадают*: $\theta_1^* = \theta_2^*$ *п. н.*

Доказательство. Заметим сначала, что $\mathsf{E}(\theta_1^* - \theta)^2 = \mathsf{E}(\theta_2^* - \theta)^2$. Действительно, так как θ_1^* эффективна в классе K_b, то она не хуже оценки θ_2^*, т. е. $\mathsf{E}(\theta_1^* - \theta)^2 \leqslant \mathsf{E}(\theta_2^* - \theta)^2$ и наоборот.

Рассмотрим оценку $\theta^* = \dfrac{\theta_1^* + \theta_2^*}{2}$. Она также принадлежит классу K_b (*доказать*). Вычислим её среднеквадратическое отклонение. Запишем

$$\left(\frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2 = \frac{a^2+b^2}{2} \tag{7}$$

и положим в этом равенстве $a = \theta_1^* - \theta$, $b = \theta_2^* - \theta$. Тогда

$$\frac{a+b}{2} = \theta^* - \theta, \qquad a - b = \theta_1^* - \theta_2^*.$$

Подставим эти выражения в (7) и вычислим математические ожидания обеих частей:

$$\mathsf{E}(\theta^* - \theta)^2 + \mathsf{E}\left(\frac{\theta_1^* - \theta_2^*}{2}\right)^2 = \mathsf{E}\,\frac{(\theta_1^* - \theta)^2 + (\theta_2^* - \theta)^2}{2} = \mathsf{E}(\theta_1^* - \theta)^2. \tag{8}$$

Но оценка θ^* принадлежит K_b, т. е. она не лучше, например, эффективной оценки θ_1^*. Поэтому

$$\mathsf{E}(\theta^* - \theta)^2 \geqslant \mathsf{E}(\theta_1^* - \theta)^2.$$

Сравнивая это неравенство с равенством (8), видим, что

$$\mathsf{E}\left(\frac{\theta_1^* - \theta_2^*}{2}\right)^2 = \frac{1}{4}\,\mathsf{E}(\theta_1^* - \theta_2^*)^2 \leqslant 0 \quad \text{и, следовательно,} \quad \mathsf{E}(\theta_1^* - \theta_2^*)^2 = 0.$$

Тогда (*почему*?) $\theta_1^* = \theta_2^*$ п. н., что и требовалось доказать. □

Для примера рассмотрим сравнение двух оценок. Разумеется, сравнивая оценки попарно между собой, наилучшей оценки в целом классе не найти, но выбрать лучшую из двух тоже полезно. Поиску наилучшей оценки в целом классе посвящена следующая глава.

Пример 13. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. В примерах 4 и 10 мы нашли ОМП $\hat\theta = X_{(n)} = \max\{X_1, \ldots, X_n\}$ и ОММ по первому моменту $\theta^* = 2\overline{X}$. Сравним их в среднеквадратичном.

Оценка $\theta^* = 2\overline{X}$ несмещённая, поэтому

$$\mathsf{E}(\theta^* - \theta)^2 = \mathsf{D}\theta^* = \mathsf{D}2\overline{X} = 4 \cdot \mathsf{D}\overline{X} = 4\,\frac{\mathsf{D}X_1}{n} = 4 \cdot \frac{\theta^2}{12n} = \frac{\theta^2}{3n}.$$

Для $\hat\theta = X_{(n)} = \max\{X_1, \ldots, X_n\}$ имеем $\mathsf{E}(\hat\theta - \theta)^2 = \mathsf{E}\hat\theta^2 - 2\theta\,\mathsf{E}\hat\theta + \theta^2$. Найдём функцию и плотность распределения случайной величины $\hat\theta$:

$$\mathsf{P}(X_{(n)} < y) = \mathsf{P}(X_1 < y, \ldots, X_n < y) = \mathsf{P}^n(X_1 < y) = \begin{cases} 0, & y < 0, \\ \dfrac{y^n}{\theta^n}, & y \in [0, \theta], \\ 1, & y > \theta, \end{cases}$$

$$f_{X_{(n)}}(y) = \begin{cases} 0, & \text{если } y \notin [0, \theta], \\ n\dfrac{y^{n-1}}{\theta^n}, & \text{если } y \in [0, \theta]. \end{cases}$$

Посчитаем первый и второй моменты случайной величины $\hat\theta = X_{(n)}$:

$$\mathsf{E}X_{(n)} = \int\limits_0^\theta yn\,\frac{y^{n-1}}{\theta^n}\,dy = \frac{n}{n+1}\,\theta, \quad \mathsf{E}X_{(n)}^2 = \int\limits_0^\theta y^2 n\,\frac{y^{n-1}}{\theta^n}\,dy = \frac{n}{n+2}\,\theta^2.$$

Поэтому

$$\mathsf{E}(X_{(n)} - \theta)^2 = \frac{n}{n+2}\,\theta^2 - 2\,\frac{n}{n+1}\,\theta^2 + \theta^2 = \frac{2}{(n+1)(n+2)}\,\theta^2.$$

При $n = 1,\ 2$ квадратические отклонения оценок θ^* и $\hat\theta$ равны: ни одна из этих оценок не лучше другой в среднеквадратическом смысле, а при

$n > 2$ оценка $X_{(n)}$ оказывается лучше, чем $2\overline{X}$:

$$\mathsf{E}(X_{(n)} - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)} < \frac{\theta^2}{3n} = \mathsf{E}(2\overline{X} - \theta)^2.$$

При этом $\mathsf{E}(X_{(n)} - \theta)^2$ стремится к нулю со скоростью n^{-2}, тогда как $\mathsf{E}(2\overline{X} - \theta)^2$ — всего лишь со скоростью n^{-1}.

§ 2. Асимптотический подход к сравнению оценок

Асимптотически нормальные оценки. Среднеквадратический подход к сравнению оценок имеет существенный недостаток: не всегда возможно вычислить моменты величины $\theta^* - \theta$. Например, не удастся сравнить в среднеквадратичном смысле оценки $\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}$ из примера 4 (с. 26). Оценки такого вида (нелинейные функции от сумм) можно сравнивать с помощью асимптотического подхода. Этот подход применим к *асимптотически нормальным* оценкам.

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, $\theta \in \Theta$.

Определение 13. Оценка θ^* называется *асимптотически нормальной оценкой* (АНО) параметра θ с коэффициентом $\sigma^2(\theta)$, если для любого $\theta \in \Theta$ имеет место слабая сходимость при $n \to \infty$

$$\sqrt{n}(\theta^* - \theta) \Rightarrow \mathrm{N}_{0,\,\sigma^2(\theta)} \quad \text{или, что то же самое,} \quad \sqrt{n}\,\frac{\theta^* - \theta}{\sigma(\theta)} \Rightarrow \mathrm{N}_{0,1}.$$

Асимптотическая нормальность оценок является важным свойством последовательностей оценок. В дальнейшем мы увидим, что это свойство используется не только для сравнения оценок между собой, но и при построении доверительных интервалов для неизвестных параметров и в задачах проверки гипотез о значениях этих параметров.

Пример 14. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Проверим, являются ли оценки $\theta^* = 2\overline{X}$ и $\hat{\theta} = X_{(n)}$ асимптотически нормальными. По ЦПТ,

$$\begin{aligned}\sqrt{n}(\theta^* - \theta) &= \sqrt{n}(2\overline{X} - \theta) = \sqrt{n}\left(2\frac{\sum X_i}{n} - \theta\right) = \frac{\sum\limits_{i=1}^{n}(2X_i) - n\theta}{\sqrt{n}} = \\ &= \frac{\sum\limits_{i=1}^{n}(2X_i) - n\mathsf{E}(2X_1)}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\,\mathsf{D}(2X_1)} = \mathrm{N}_{0,\,4\mathsf{D}X_1}.\end{aligned}$$

Итак, оценка $\theta^* = 2\overline{X}$ асимптотически нормальна с коэффициентом

$$\sigma^2(\theta) = 4\mathsf{D}X_1 = 4 \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3}.$$

Для проверки асимптотической нормальности оценки $\hat\theta = X_{(n)}$ воспользуемся определением слабой сходимости. По определению, $\xi_n \Rightarrow F$, *если для любой точки x, являющейся точкой непрерывности функции распределения F, имеет место сходимость*

$$\mathsf{P}(\xi_n < x) \to F(x) \ \ \textit{при} \ \ n \to \infty.$$

Слабая сходимость $\sqrt{n}\,(X_{(n)} - \theta)$ к N_{0,σ^2} имеет место, если для любого $x \in \mathbb{R}$ (*почему для любого?*)

$$\mathsf{P}(\sqrt{n}(X_{(n)} - \theta) < x) \to \Phi_{0,\sigma^2}(x) \ \ \text{при} \ \ n \to \infty.$$

Но в точке $x = 0$ функция распределения величины $\sqrt{n}\,(X_{(n)} - \theta)$ равна единице. Действительно,

$$\sqrt{n}(\hat\theta - \theta) = \sqrt{n}\,(X_{(n)} - \theta) < 0 \ \ \text{п. н.,} \tag{9}$$

поэтому $\mathsf{P}_\theta(\sqrt{n}\,(X_{(n)} - \theta) < 0) = 1$. А для нормального распределения $\mathrm{N}_{0,\sigma^2(\theta)}$ функция распределения в нуле равна $\Phi_{0,\sigma^2(\theta)}(0) = 0{,}5$. Но 1 не сходится к 0,5 при $n \to \infty$, поэтому слабая сходимость $\sqrt{n}\,(X_{(n)} - \theta)$ к $\mathrm{N}_{0,\sigma^2(\theta)}$ места не имеет и оценка $\hat\theta = X_{(n)}$ асимптотически нормальной не является.

Осталось ответить на напрашивающиеся вопросы.

Вопрос 1. Куда всё же сходится по распределению $\sqrt{n}(X_{(n)} - \theta)$?

Упражнение. Доказать, что $\sqrt{n}(X_{(n)} - \theta) \Rightarrow 0$.

Порядок действий: Выписать определение слабой сходимости. Нарисовать функцию распределения нуля. Найти по определению функцию распределения $\sqrt{n}(X_{(n)} - \theta)$. Убедиться, что она сходится к функции распределения нуля во всех точках непрерывности последней. Не забыть о существовании *замечательных пределов*, *логарифмов* и *ряда Тейлора.*

Вопрос 2. Если $\sqrt{n}(X_{(n)} - \theta) \Rightarrow 0$, то на какую степень n нужно попробовать умножить $X_{(n)} - \theta$, чтобы получить сходимость к величине, отличной от нуля и от бесконечности?

Упражнение. Доказать, что $-n(X_{(n)} - \theta) \Rightarrow \eta$, где случайная величина η имеет показательное распределение $\mathrm{E}_{1/\theta}$.

Порядок действий: прежний.

Вопрос 3. Для оценки $\frac{n+1}{n}X_{(n)}$ свойство (9) не выполнено. Может ли эта оценка быть АНО?

Упражнение. Модифицировать рассуждения и доказать, что эта оценка тоже не является асимптотически нормальной.

Вопрос последний. Плохо ли, что оценка $\hat{\theta} = X_{(n)}$ не асимптотически нормальна? Может быть, сходимость $n(X_{(n)} - \theta) \Rightarrow -\eta$ ещё лучше (особенно из-за множителя n при разности ($X_{(n)}$ и θ)?

«Скорость» сходимости оценки к параметру. Попробуем ответить на последний вопрос и заодно объясним себе, о чём говорит множитель $\sqrt{n}$ в определении асимптотической нормальности оценок.

Теорема 13. *Если θ^* — асимптотически нормальная оценка для параметра θ, то она состоятельна.*

Доказательство. Вспомним свойство слабой сходимости: произведение двух последовательностей, одна из которых сходится по вероятности к постоянной, а другая слабо сходится к некоторой случайной величине, слабо сходится к произведению пределов. Поэтому

$$\theta^* - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\theta^* - \theta) \Rightarrow 0 \cdot \xi = 0,$$

где ξ имеет нормальное распределение $\mathrm{N}_{0,\,\sigma^2(\theta)}$. Но слабая сходимость к нулю влечет сходимость к нулю по вероятности. □

Упражнение. Верно ли утверждение теоремы 13, если предельная величина ξ имеет распределение, отличное от нормального?

Итак, если θ^* асимптотически нормальна, то $\theta^* \xrightarrow{p} \theta$, т. е. $\theta^* - \theta \xrightarrow{p} 0$. Свойство асимптотической нормальности показывает, в частности, что скорость этой сходимости имеет порядок $\frac{1}{\sqrt{n}}$. Действительно, расстояние между θ^* и θ сходится к нулю, но перестаёт это делать после умножения на $\sqrt{n}$:

$$\theta^* - \theta \xrightarrow{p} 0, \quad \text{но} \quad \sqrt{n}(\theta^* - \theta) \Rightarrow \mathrm{N}_{0,\sigma^2(\theta)}.$$

Взглянем с этой точки зрения на оценку $\hat{\theta} = X_{(n)}$ в примере 14. Для этой оценки (и для тех, кто справился с упражнениями)

$$n(X_{(n)} - \theta) \Rightarrow \xi, \tag{10}$$

где ξ — некоторая случайная величина. Здесь расстояние между $\hat{\theta}$ и θ ведёт себя как $\frac{1}{n}$. Оценка *быстрее* сходится к параметру.

Упражнение. Лучше это или хуже?

Сделаем вывод из предыдущих рассуждений. Асимптотическая нормальность представляет собой типичное, ничем не выдающееся качество оценок. Асимптотически нормальные оценки сближаются с параметром со скоростью $\dfrac{1}{\sqrt{n}}$. Отсутствие такого качества *может* означать, что оценка *быстрее* сходится к параметру, нежели любая асимптотически нормальная оценка. Примером такой «выдающейся» оценки может служить ОМП для параметра θ равномерного распределения на отрезке $[0,\theta]$, не являющаяся асимптотически нормальной. Она *так быстро* сходится к параметру, как не умеет ни одна АНО. Таким образом, отсутствие свойства АНО не делает эту оценку хуже. Скорее наоборот.

Асимптотическая нормальность ОММ. Продолжим рассмотрение асимптотически нормальных оценок. В примере 14 мы видели, что для оценки $2\overline{X}$ свойство асимптотической нормальности сразу следует из ЦПТ. Установим асимптотическую нормальность оценок более сложного вида, какими обычно оказываются оценки метода моментов.

Лемма 1. *Пусть функция* $g(y)$ *такова, что* $0 \neq \mathsf{D}g(X_1) < \infty$. *Тогда статистика* $\overline{g(X)}$ *является асимптотически нормальной оценкой для* $\mathsf{E}g(X_1)$ *с коэффициентом* $\sigma^2(\theta) = \mathsf{D}g(X_1)$:

$$\sqrt{n}\,\frac{\overline{g(X)} - \mathsf{E}g(X_1)}{\sqrt{\mathsf{D}g(X_1)}} \Rightarrow \mathrm{N}_{0,1}.$$

Упражнение. Вспомнить ЦПТ и доказать лемму 1.

Следующая теорема утверждает асимптотическую нормальность оценок вида

$$\theta^* = H\Bigl(\overline{g(X)}\Bigr) = H\left(\frac{g(X_1) + \ldots + g(X_n)}{n}\right).$$

Такие оценки получаются обычно (*найти примеры*) при использовании метода моментов, при этом всегда $\theta = H\left(\mathsf{E}g(X_1)\right)$.

Теорема 14. *Пусть функция* $g(y)$ *такова, что* $0 \neq \mathsf{D}g(X_1) < \infty$, *функция* $H(y)$ *дифференцируема в точке* $a = \mathsf{E}g(X_1)$ *и её производная в этой точке* $H'(a) = H'(y)\big|_{y=a}$ *отлична от нуля.*

Тогда оценка $\theta^* = H\bigl(\overline{g(X)}\bigr)$ *является асимптотически нормальной оценкой для параметра* $\theta = H\left(\mathsf{E}g(X_1)\right) = H(a)$ *с коэффициентом асимптотической нормальности*

$$\sigma^2(\theta) = \bigl(H'(a)\bigr)^2 \cdot \mathsf{D}g(X_1).$$

Доказательство. Согласно ЗБЧ последовательность $\overline{g(X)}$ стремится к $a = \mathsf{E}g(X_1)$ по вероятности с ростом n. Функция

$$G(y) = \begin{cases} \dfrac{H(y) - H(a)}{y - a}, & y \neq a, \\ H'(a), & y = a \end{cases}$$

по условию непрерывна в точке a. Поскольку сходимость по вероятности сохраняется под действием непрерывной функции, получим, что $G(\overline{g(X)}) \xrightarrow{\text{p}} G(a) = H'(a)$.

Заметим также, что по лемме 1 величина $\sqrt{n}\bigl(\overline{g(X)} - a\bigr)$ слабо сходится к нормальному распределению $\mathrm{N}_{0,\,\mathsf{D}g(X_1)}$. Пусть ξ — случайная величина из этого распределения. Тогда

$$\sqrt{n}\left(H\left(\overline{g(X)}\right) - H(a)\right) = \sqrt{n}\left(\overline{g(X)} - a\right) \cdot G\left(\overline{g(X)}\right) \Rightarrow \xi \cdot H'(a).$$

Мы использовали (*в который раз*?) следующее свойство слабой сходимости: если $\xi_n \Rightarrow \xi$ и $\eta_n \xrightarrow{\text{p}} c = \text{const}$, то $\xi_n\eta_n \Rightarrow c\xi$. Но распределение случайной величины $\xi \cdot H'(a)$ как раз и есть $\mathrm{N}_{0,\,(H'(a))^2 \cdot \mathsf{D}g(X_1)}$. Поэтому

$$\sigma^2(\theta) = \bigl(H'(a)\bigr)^2 \cdot \mathsf{D}g(X_1). \qquad \square$$

Пример 15. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Проверим, являются ли асимптотически нормальными оценки

$$\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}, \quad k = 1, 2, \ldots,$$

полученные методом моментов в примере 4 (с. 26).

Пусть $g(y) = (k+1)y^k$, $H(y) = \sqrt[k]{y}$. Тогда

$$\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}} = \sqrt[k]{\frac{\sum(k+1)X_i^k}{n}} = H\left(\frac{\sum g(X_i)}{n}\right).$$

При этом

$$\theta = H\left(\mathsf{E}g(X_1)\right) = \sqrt[k]{\mathsf{E}(k+1)X_1^k} = \sqrt[k]{(k+1)\ \frac{\theta^k}{k+1}}.$$

Впрочем, иначе быть не могло по определению метода моментов (*верно*?). Проверим другие условия теоремы 14:

$$a = \mathsf{E}g(X_1) = (k+1)\ \frac{\theta^k}{k+1} = \theta^k,$$

дисперсия

$$\mathsf{D}g(X_1) = \mathsf{E}(k+1)^2 X_1^{2k} - a^2 = (k+1)^2 \frac{\theta^{2k}}{2k+1} - \theta^{2k} = \frac{k^2}{2k+1}\,\theta^{2k}$$

конечна и отлична от нуля. Функция $H(y)$ дифференцируема в точке a:

$$H'(y) = \frac{1}{k}\,y^{\frac{1-k}{k}}, \qquad H'(a) = H'(\theta^k) = \frac{1}{k}\,\theta^{1-k} \neq 0.$$

По теореме 14, оценка θ_k^* является АНО для θ с коэффициентом

$$\sigma_k^2(\theta) = \big(H'(a)\big)^2\,\mathsf{D}g(X_1) = \frac{1}{k^2}\,\theta^{2-2k}\cdot\frac{k^2}{2k+1}\,\theta^{2k} = \frac{\theta^2}{2k+1}.$$

Например, для $\theta_1^* = 2\overline{X}$ имеем коэффициент $\sigma_1^2(\theta) = \dfrac{\theta^2}{3}$. Это в точности совпадает с коэффициентом, полученным нами в примере 14 (с. 39).

Осталось понять, как сравнивать асимптотически нормальные оценки и что показывает коэффициент асимптотической нормальности.

Асимптотический подход к сравнению оценок. Возьмём две случайные величины: $\xi \in \mathrm{N}_{0,1}$ и $10\,\xi \in \mathrm{N}_{0,100}$. Разброс значений у величины $10\,\xi$ гораздо больший:

$$0,9973 = \mathsf{P}(|\xi| < 3) = \mathsf{P}(|10\,\xi| < 30),$$

и дисперсия (показатель этого рассеяния) соответственно больше.

То же самое показывает и коэффициент асимптотической нормальности. Возьмём две АНО с коэффициентами 1 и 100:

$$\sqrt{n}(\theta_1^* - \theta^*) \Rightarrow \mathrm{N}_{0,1} \quad \text{и} \quad \sqrt{n}(\theta_2^* - \theta^*) \Rightarrow \mathrm{N}_{0,100}.$$

При больших n разброс значений величины $\sqrt{n}(\theta_2^* - \theta^*)$ около нуля гораздо больше, чем у величины $\sqrt{n}(\theta_1^* - \theta^*)$, поскольку больше предельная дисперсия (она же коэффициент асимптотической нормальности).

Но чем меньше отклонение оценки от параметра, тем лучше. Получаем естественный способ сравнения асимптотически нормальных оценок.

Определение 14. Пусть θ_1^* — АНО с коэффициентом $\sigma_1^2(\theta)$, θ_2^* — АНО с коэффициентом $\sigma_2^2(\theta)$. Говорят, что θ_1^* *лучше*, чем θ_2^* в смысле асимптотического подхода, если для любого $\theta \in \Theta$

$$\sigma_1^2(\theta) \leqslant \sigma_2^2(\theta),$$

и хотя бы при одном θ это неравенство строгое.

Пример 16. Сравним между собой в асимптотическом смысле оценки в последовательности $\theta_1^*, \theta_2^*, \ldots$ из примера 15. Для θ_k^* коэффициент асимптотической нормальности имеет вид $\sigma_k^2(\theta) = \dfrac{\theta^2}{2k+1}$. Коэффициент тем меньше, чем больше k, т. е. каждая следующая оценка в этой последовательности лучше предыдущей.

Оценка θ_∞^*, являющаяся «последней» оценкой в этой последовательности, могла бы быть лучше всех оценок в этой последовательности в смысле асимптотического подхода, если бы являлась асимптотически нормальной. Но если читатель решил задачу 7 к главе I или задачу 10 к главе II, он знает, что этой «последней» оценкой является $X_{(n)}$, а она не асимптотически нормальна.

Ещё раз напомним, что оценка $\hat\theta = X_{(n)}$ оказывается *лучше* любой асимптотически нормальной оценки: «скорость» её сходимости к параметру, как показывает (10), равна n^{-1} в отличие от скорости $n^{-1/2}$, которая наблюдается у любой АНО.

§ 3. Вопросы и упражнения

1. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{\theta,\,\theta+5}$, где $\theta \in \mathbb{R}$. Сравнить оценки $\hat\theta_0 = X_{(n)} - 5$ и $\hat\theta_1 = X_{(1)}$ из примера 11 (с. 31) в среднеквадратическом смысле. Сравнить с этими оценками оценку метода моментов $\theta^* = \overline{X} - 2{,}5$.

2. Для показательного распределения с параметром α оценка, полученная методом моментов по k-му моменту, имеет вид: $\alpha_k^* = \sqrt[k]{\dfrac{k!}{\overline{X^k}}}$. Сравнить оценки α_k^*, $k = 1, 2, \ldots$ в смысле асимптотического подхода. Доказать, что оценка α_1^* наилучшая.

3. Выполнить все упражнения в тексте главы III.

4. Получить утверждение теоремы Гливенко — Кантелли из утверждения и в условиях теоремы Колмогорова аналогично доказательству теоремы 13.

5. Является ли оценка $\overline{X}+1$ асимптотически нормальной оценкой для параметра λ распределения Пуассона Π_λ?

6. Привести пример состоятельной оценки для параметра λ распределения Пуассона, которая не являлась бы АНО.

7. Дана выборка из показательного распределения с неизвестным параметром $\alpha > 0$. Проверить асимптотическую нормальность оценки параметра α, полученной методом моментов по первому моменту.

8. Дана выборка объема n из распределения Пуассона с параметром $\lambda > 0$. Для какого параметра $\theta = \theta(\lambda)$ оценка $\theta^* = \overline{X}e^{-\overline{X}}$ является состоятельной оценкой? Проверить, является ли эта оценка асимптотически нормальной оценкой для того же параметра.

9. Дана выборка $X_1, \ldots, X_n$ из распределения Пуассона с параметром $\lambda > 0$. Построить оценку метода моментов по первому моменту для параметра $\theta = \mathsf{P}(X_1 = 0)$. Является ли эта оценка асимптотически нормальной?

10. Доказать асимптотическую нормальность выборочной медианы, построенной по выборке из распределения Коши с параметрами a и 1. Найти коэффициент асимптотической нормальности.

11. Пусть $X_1, \ldots, X_n$ — выборка объёма n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Доказать, что $X_{(n)} \in K_b$, где $b = b(\theta) = -\dfrac{\theta}{n+1}$. Доказать, что $\dfrac{n+1}{n}\,X_{(n)} \in K_0$. Сравнить эти оценки в среднеквадратичном смысле.

ГЛАВА IV

ЭФФЕКТИВНЫЕ ОЦЕНКИ

В классе одинаково смещённых оценок эффективной мы назвали оценку с наименьшим среднеквадратическим отклонением. Но попарное сравнение оценок — далеко не лучший способ отыскания эффективной оценки. Сегодня мы познакомимся с утверждением, позволяющим во многих случаях *доказать* эффективность оценки (если, конечно, она на самом деле эффективна). Это утверждение называется неравенством Рао — Крамéра и говорит о том, что в любом классе $K_{b(\theta)}$ существует нижняя граница для среднеквадратического отклонения любой оценки. Таким образом, если найдётся оценка, отклонение которой в точности равно этой нижней границе (самое маленькое), то данная оценка — эффективна, поскольку у всех остальных оценок отклонение меньшим быть не может. К сожалению, данное неравенство верно лишь для так называемых «регулярных» семейств распределений, к которым *не* относится, например, большинство равномерных.

§ 1. Регулярность семейства распределений

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta$, а область $\Theta \subset \mathbb{R}$ представляет собой конечный или бесконечный интервал. Пусть, как в главе II,

$$f_\theta(y) = \begin{cases} \text{плотность } f_\theta(y), & \text{если распределение абсолютно непрерывно,} \\ \mathsf{P}_\theta(X_1 = y), & \text{если распределение дискретно.} \end{cases}$$

Введём понятие *носителя* семейства распределений $\{\mathcal{F}_\theta,\ \theta \in \Theta\}$.

Определение 15. *Носителем* параметрического семейства распределений $\mathcal{F}_\theta$ будем называть любое множество $C \subseteq \mathbb{R}$ такое, что при всех $\theta \in \Theta$ выполняется равенство $\mathsf{P}(X_1 \in C) = 1$ для $X_1 \Subset \mathcal{F}_\theta$.

Замечание 9. Мы ввели понятие носителя семейства мер в $\mathbb{R}$, отличное от общепринятого (*найти общепринятое*!). Так, носитель в смысле данного нами определения не единствен, но все эти носители отличаются на множество нулевой вероятности.

Следующие два условия принято называть *условиями регулярности.*

(R) Существует такой носитель C семейства распределений $\mathcal{F}_\theta$, что при каждом $y \in C$ функция $\sqrt{f_\theta(y)}$ непрерывно дифференцируема по θ *всюду* в области Θ.

(RR) *Информация Фишера* $I(\theta) = \mathsf{E}\left(\dfrac{\partial}{\partial\theta}\ln f_\theta(X_1)\right)^2$ существует, положительна и непрерывна по θ во всех точках $\theta \in \Theta$.

Пример 17 (регулярное семейство). Рассмотрим показательное распределение E_α с параметром $\alpha > 0$. Плотность этого распределения имеет вид

$$f_\alpha(y) = \begin{cases} \alpha e^{-\alpha y}, & \text{если } y > 0,\\ 0, & \text{если } y \leqslant 0, \end{cases} \qquad \sqrt{f_\alpha(y)} = \begin{cases} \sqrt{\alpha} e^{-\alpha y/2}, & \text{если } y > 0,\\ 0, & \text{если } y \leqslant 0. \end{cases}$$

В качестве множества C можно взять полупрямую $(0, +\infty)$, поскольку $\mathsf{P}(X_1 > 0) = 1$. При любом $y \in C$, т. е. при $y > 0$, существует производная функции $\sqrt{f_\alpha(y)}$ по α, и эта производная непрерывна во всех точках $\alpha > 0$:

$$\frac{\partial}{\partial\alpha}\sqrt{f_\alpha(y)} = \frac{1}{2\sqrt{\alpha}}e^{-\alpha y/2} - \sqrt{\alpha}\,\frac{y}{2}\,e^{-\alpha y/2}.$$

Условие (RR) проверим непосредственным вычислением $I(\alpha)$:

$$f_\alpha(X_1) = \alpha\, e^{-\alpha X_1}, \quad \ln f_\alpha(X_1) = \ln\alpha - \alpha X_1, \quad \frac{\partial}{\partial\alpha}\ln f_\alpha(X_1) = \frac{1}{\alpha} - X_1,$$

$$I(\alpha) = \mathsf{E}\left(\frac{\partial}{\partial\alpha}\ln f_\alpha(X_1)\right)^2 = \mathsf{E}\left(X_1 - \frac{1}{\alpha}\right)^2 = \mathsf{D}X_1 = \frac{1}{\alpha^2}.$$

Итак, информация Фишера $I(\alpha) = \dfrac{1}{\alpha^2}$ существует, положительна и непрерывна по α при всех $\alpha > 0$, т. е. условие (RR) выполнено.

Пример 18 (нерегулярное семейство). Рассмотрим равномерное распределение $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Плотность этого распределения имеет вид

$$f_\theta(y) = \begin{cases} \dfrac{1}{\theta}, & \text{если } 0 \leqslant y \leqslant \theta,\\ 0, & \text{если } y \notin [0,\theta] \end{cases} \quad = \quad \begin{cases} \dfrac{1}{\theta}, & \text{если } \theta \geqslant y \text{ и } y > 0,\\ 0 & \text{иначе.} \end{cases}$$

Поскольку параметр θ может принимать любые положительные значения, никакой ограниченный интервал $(0, x)$ не может быть носителем этого семейства распределений: $\mathsf{P}(X_1 \in (0, x)) < 1$ при $X_1 \Subset \mathrm{U}_{0,\theta}$ с параметром $\theta > x$. Возьмём в качестве носителя луч $C = (0, +\infty)$ — он при любом $\theta > 0$ обладает свойством $\mathsf{P}(X_1 \in C) = 1$. Уменьшить существенно этот носитель не удастся — из него можно исключать лишь множества нулевой лебеговой меры.

Покажем, что условие (R) не выполнено: множество тех $y \in C$, при каждом из которых функция $\sqrt{f_\theta(y)}$ дифференцируема по θ, абсолютно пусто. При фиксированном $y > 0$ изобразим функцию $f_\theta(y)$ как функцию переменной θ (рис. 5).

Видим, что при любом $y \in C$ функция $f_\theta(y)$, равно как и корень из неё, даже не является непрерывной по θ, а тем более дифференцируемой. Следовательно, условие (R) не выполнено.

Пример 19 (ещё одно нерегулярное семейство). Рассмотрим смещённое показательное распределение с параметром сдвига $\theta \in \mathbb{R}$ и плотностью

$$f_\theta(y) = \begin{cases} e^{\theta - y}, & \text{если } y > \theta, \\ 0, & \text{если } y \leqslant \theta \end{cases} \quad = \quad \begin{cases} e^{\theta - y}, & \text{если } \theta < y, \\ 0, & \text{если } \theta \geqslant y. \end{cases}$$

Поскольку при любом θ распределение сосредоточено на $(\theta, +\infty)$, а параметр θ может принимать любые вещественные значения, то только $C = \mathbb{R}$ (плюс-минус множество меры нуль) таково, что при любом $\theta > 0$ выполнено $\mathsf{P}(X_1 \in C) = 1$. Покажем, что условие (R) опять не выполняется: множество тех $y \in C$, при каждом из которых функция $\sqrt{f_\theta(y)}$ дифференцируема по θ, столь же пусто, как и в примере 18.

При фиксированном $y \in \mathbb{R}$ на рис. 6 приведён график функции $f_\theta(y)$ (или корня из неё) как функции переменной θ. Независимо от выбора y функция $f_\theta(y)$ не является непрерывной по θ. Тем более она не является дифференцируемой.

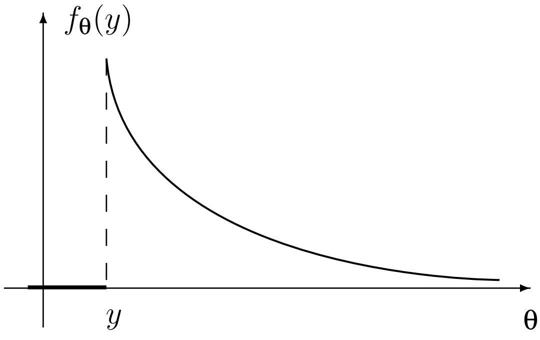

Рис. 5. Плотность в примере 18

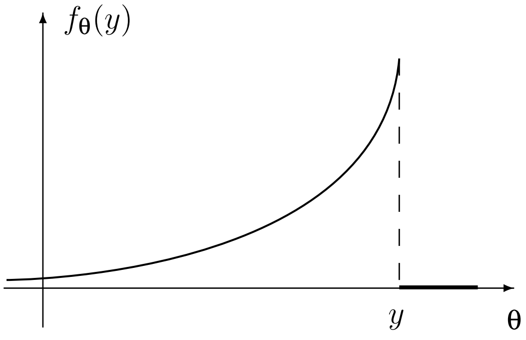

Рис. 6. Плотность в примере 19

Замечание 10. Вместо непрерывной дифференцируемости $\sqrt{f_\theta(y)}$ можно требовать того же от $\ln f_\theta(y)$.

§ 2. Неравенство Рао — Крамера

Пусть $X_1, \ldots, X_n$ — выборка объёма n из параметрического семейства распределений $\mathcal{F}_\theta$, и семейство $\{\mathcal{F}_\theta,\ \theta \in \Theta\}$ удовлетворяет условиям регулярности (R) и (RR).

Справедливо следующее утверждение.

Теорема 15 (неравенство Рао — Крамера). *Пусть семейство распределений $\mathcal{F}_\theta$ удовлетворяет условиям* (R) *и* (RR). *Тогда для любой несмещённой оценки $\theta^* \in K_0$, дисперсия которой $\mathsf{D}\theta^*$ ограничена на любом компакте в области Θ, справедливо неравенство*

$$\mathsf{D}\theta^* = \mathsf{E}\,(\theta^* - \theta)^2 \geqslant \frac{1}{nI(\theta)}.$$

Упражнение. Проверить, что для показательного семейства распределений E_α с параметром $\alpha > 0$ дисперсия $\mathsf{D}X_1$ не ограничена глобально при $\alpha > 0$, но ограничена на любом компакте $\alpha \in K \subset (0, +\infty)$.

Неравенство сформулировано для класса несмещённых оценок. В классе оценок с произвольным смещением $b(\theta)$ неравенство Рао — Крамера выглядит следующим образом.

Теорема 16 (неравенство Рао — Крамера). *Пусть семейство распределений $\mathcal{F}_\theta$ удовлетворяет условиям* (R) *и* (RR). *Тогда для любой оценки $\theta^* \in K_{b(\theta)}$, дисперсия которой $\mathsf{D}\theta^*$ ограничена на любом компакте в области Θ, справедливо неравенство*

$$\mathsf{E}\,(\theta^* - \theta)^2 \geqslant \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta), \qquad \textit{т. е.} \qquad \mathsf{D}\theta^* \geqslant \frac{(1 + b'(\theta))^2}{nI(\theta)}.$$

Для доказательства нам понадобится следующее утверждение.

Лемма 2. *При выполнении условий* (R) *и* (RR) *для любой статистики $T = T(\vec{X})$, дисперсия которой ограничена на компактах, имеет место равенство*

$$\frac{\partial}{\partial\theta}\mathsf{E}T = \mathsf{E}\left(T \cdot \frac{\partial}{\partial\theta} L(\vec{X};\, \theta)\right),$$

где $L(\vec{X};\, \theta)$ — логарифмическая функция правдоподобия.

Упражнение. Вспомнить, что такое функция правдоподобия $f(\vec{X};\, \theta)$, логарифмическая функция правдоподобия $L(\vec{X};\, \theta)$ (определение 8, с. 29), как они связаны друг с другом, с плотностью распределения случайной величины X_1 и плотностью совместного распределения элементов выборки.

Доказательство. Напомним, что математическое ожидание функции от нескольких случайных величин есть (многомерный) интеграл от этой функции, помноженной на совместную плотность распределения этих случайных величин. Поэтому

$$\mathsf{E}T(X_1, \ldots, X_n) = \int\limits_{\mathbb{R}^n} T(y_1, \ldots, y_n) \cdot f(y_1, \ldots, y_n, \theta)\, dy_1 \ldots\, dy_n.$$

В следующей цепочке преобразований равенство, помеченное звёздочкой, мы доказывать не будем, поскольку его доказательство требует знания условий дифференцируемости интеграла по параметру. Это равенство — смена порядка дифференцирования и интегрирования — то *единственное*, ради чего введены условия регулярности (см. пример ниже).

$$\begin{aligned}\frac{\partial}{\partial\theta}\mathsf{E}T(\vec{X}) &= \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} T(\vec{y})\, f(\vec{y};\,\theta)\, d\vec{y} \overset{*}{=} \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta}\left(T(\vec{y})\, f(\vec{y};\,\theta)\right)\, d\vec{y} = \\ &= \int\limits_{\mathbb{R}^n} T(\vec{y}) \cdot \frac{\partial}{\partial\theta} f(\vec{y};\,\theta)\, d\vec{y} = \int\limits_{\mathbb{R}^n} T(\vec{y}) \cdot \left(\frac{\frac{\partial}{\partial\theta} f(\vec{y};\,\theta)}{f(\vec{y};\,\theta)}\right) \cdot f(\vec{y};\,\theta)\, d\vec{y} = \\ &= \int\limits_{\mathbb{R}^n} T(\vec{y}) \cdot \left(\frac{\partial}{\partial\theta} L(\vec{y};\,\theta)\right) \cdot f(\vec{y};\,\theta)\, d\vec{y} = \mathsf{E}\left(T(\vec{X}) \cdot \frac{\partial}{\partial\theta} L(\vec{X};\,\theta)\right).\end{aligned}$$

Через $\vec{y}$ в интегралах обозначен вектор $(y_1, \ldots, y_n)$. □

Доказательство неравенства Рао — Крамера. Мы докажем только неравенство для класса K_0 — теорему 15. Необходимые изменения в доказательство для класса K_b читатель внесёт самостоятельно.

Воспользуемся леммой 2. Будем брать в качестве $T(\vec{X})$ разные функции и получать забавные формулы, которые потом соберём вместе и используем в неравенстве Коши — Буняковского (*в котором*?).

Пусть $T(\vec{X}) \equiv 1$. Тогда математическое ожидание производной от логарифмической функции правдоподобия равно нулю:

$$0 = \frac{\partial}{\partial\theta} 1 = \mathsf{E}\,\frac{\partial}{\partial\theta} L(\vec{X};\,\theta).$$

Далее, поскольку $f(\vec{X};\,\theta) = \prod f_\theta(X_i)$, то $L(\vec{X};\,\theta) = \sum \ln f_\theta(X_i)$ и

$$0 = \mathsf{E}\,\frac{\partial}{\partial\theta} L(\vec{X};\,\theta) = \mathsf{E}\sum \frac{\partial}{\partial\theta} \ln f_\theta(X_i) = n \cdot \mathsf{E}\,\frac{\partial}{\partial\theta} \ln f_\theta(X_1). \qquad (11)$$

Поэтому $\mathsf{E}\,\frac{\partial}{\partial\theta} \ln f_\theta(X_1) = 0$.

Пусть теперь $T(\vec{X}) = \theta^* \in K_0$, т. е. $\mathsf{E}\theta^* = \theta$. Тогда

$$\frac{\partial}{\partial\theta}\mathsf{E}\theta^* = \frac{\partial}{\partial\theta}\theta = 1 = \mathsf{E}\theta^* \cdot \frac{\partial}{\partial\theta}L(\vec{X};\,\theta). \tag{12}$$

Вспомним свойство коэффициента корреляции

$$|\mathrm{cov}(\xi,\eta)| = |\mathsf{E}\xi\eta - \mathsf{E}\xi\mathsf{E}\eta| \leqslant \sqrt{\mathsf{D}\xi\mathsf{D}\eta}.$$

Используя формулы (11) и (12), получаем

$$\mathrm{cov}\left(\theta^*, \frac{\partial}{\partial\theta}L(\vec{X};\,\theta)\right) = \mathsf{E}\left(\theta^*\cdot\frac{\partial}{\partial\theta}L(\vec{X};\,\theta)\right) - \mathsf{E}\theta^*\mathsf{E}\,\frac{\partial}{\partial\theta}L(\vec{X};\,\theta) =$$
$$= \mathsf{E}\left(\theta^*\cdot\frac{\partial}{\partial\theta}L(\vec{X};\,\theta)\right) = 1 \leqslant \sqrt{\mathsf{D}\theta^*\cdot\mathsf{D}\,\frac{\partial}{\partial\theta}L(\vec{X};\,\theta)}. \tag{13}$$

Найдём $\mathsf{D}\,\frac{\partial}{\partial\theta}L(\vec{X};\,\theta)$ с помощью равенства (11):

$$\begin{aligned}\mathsf{D}\frac{\partial}{\partial\theta}L(\vec{X};\,\theta) &= \mathsf{D}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f_\theta(X_i) = n\mathsf{D}\frac{\partial}{\partial\theta}\ln f_\theta(X_1) = \\ &= n\mathsf{E}\left(\frac{\partial}{\partial\theta}\ln f_\theta(X_1)\right)^2 = nI(\theta).\end{aligned}$$

Подставив дисперсию в неравенство (13), получим окончательно

$$1 \leqslant \mathsf{D}\theta^*\cdot nI(\theta) \quad \text{или} \quad \mathsf{D}\theta^* \geqslant \frac{1}{nI(\theta)}. \qquad \square$$

Следующий пример показывает, что условие регулярности является существенным для выполнения равенства, помеченного $(*)$ в лемме 2.

Пример 20. Рассмотрим равномерное распределение $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Выпишем при $n = 1$ какой-нибудь интеграл и сравним производную от него по параметру и интеграл от производной: скажем, для $T(X_1) = 1$,

$$\frac{\partial}{\partial\theta}\mathsf{E}T(X_1) = \frac{\partial}{\partial\theta}\int\limits_0^\theta \frac{1}{\theta}\,dy = \frac{\partial}{\partial\theta}1 = 0; \quad \text{но} \quad \int\limits_0^\theta \frac{\partial}{\partial\theta}\,\frac{1}{\theta}\,dy = -\frac{1}{\theta} \neq 0.$$

Заметим, что и само утверждение неравенства Рао — Крамера для данного семейства распределений не выполнено: найдётся оценка, дисперсия которой ведёт себя как $1/n^2$, а не как $1/n$ в неравенстве Рао — Крамера.

Упражнение. Проверить, что в качестве этой «выдающейся» из неравенства Рао — Крамера оценки можно брать, скажем, смещённую оценку $X_{(n)}$ или несмещённую оценку $\frac{n+1}{n}X_{(n)}$.

§ 3. Проверка эффективности оценок

Сформулируем очевидное следствие из неравенства Рао — Крамера. Пусть семейство распределений $\mathcal{F}_\theta$ удовлетворяет условиям регулярности (R) и (RR).

Следствие 1. *Если для оценки* $\theta^* \in K_{b(\theta)}$ *достигается равенство в неравенстве Рао — Крамера*

$$\mathsf{E}\,(\theta^* - \theta)^2 = \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta) \quad \textit{или} \quad \mathsf{D}\theta^* = \frac{(1 + b'(\theta))^2}{nI(\theta)},$$

то оценка θ^* *эффективна в классе* $K_{b(\theta)}$.

Оценку для параметра θ регулярного семейства, для которой достигается равенство в неравенстве Рао — Крамера, иногда называют *R-эффективной оценкой*. Следствие 1 можно сформулировать так: если оценка R-эффективна, то она эффективна в соответствующем классе.

Пример 21. Для выборки $X_1, \ldots, X_n$ из распределения Бернулли B_p несмещённая оценка $p^* = \overline{X}$ эффективна, так как для неё достигается равенство в неравенстве Рао — Крамера (см. [5, пример 13.20]).

Пример 22. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения N_{a,σ^2}, где $a \in \mathbb{R}$, $\sigma > 0$. Проверим, является ли оценка $a^* = \overline{X} \in K_0$ эффективной (см. также [5, пример 13.6]).

Найдём информацию Фишера относительно параметра a (считая, что имеется один неизвестный параметр a). Плотность распределения равна

$$f_{(a,\sigma^2)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(y-a)^2/(2\sigma^2)}, \quad \ln f_{(a,\sigma^2)}(y) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(y-a)^2}{2\sigma^2}.$$

Соответственно, $\dfrac{\partial}{\partial a}\ln f_{(a,\sigma^2)}(y) = \dfrac{y-a}{\sigma^2}$. Найдя второй момент этого выражения при $y = X_1$, получим информацию Фишера

$$I(a) = \mathsf{E}\left(\frac{\partial}{\partial a}\ln f_{(a,\sigma^2)}(X_1)\right)^2 = \frac{\mathsf{E}(X_1 - a)^2}{\sigma^4} = \frac{\mathsf{D}X_1}{\sigma^4} = \frac{1}{\sigma^2}.$$

Найдём дисперсию оценки $\overline{X}$: $\mathsf{D}\overline{X} = \frac{1}{n}\mathsf{D}X_1 = \frac{\sigma^2}{n}$.

Сравнивая левую и правую части в неравенстве Рао — Крамера, получаем равенство

$$\mathsf{D}\overline{X} = \frac{\sigma^2}{n} = \frac{1}{nI(a)}.$$

Итак, оценка $a^* = \overline{X}$ эффективна (т. е. обладает наименьшей дисперсией среди несмещённых оценок).

Пример 23. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения N_{0,σ^2}, где $\sigma > 0$. Проверим, является ли эффективной оценка

$$\sigma^{2*} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2} \in K_0.$$

Упражнение. Получить эту оценку методом моментов и методом максимального правдоподобия.

Найдём информацию Фишера относительно параметра σ^2. Плотность распределения равна

$$f_{\sigma^2}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-y^2/(2\sigma^2)}, \quad \ln f_{\sigma^2}(y) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{y^2}{2\sigma^2}.$$

Продифференцируем это выражение по параметру σ^2:

$$\frac{\partial}{\partial\sigma^2}\ln f_{\sigma^2}(y) = -\frac{1}{2\sigma^2} + \frac{y^2}{2\sigma^4}.$$

Вычислим информацию Фишера

$$I(\sigma^2) = \mathsf{E}\left(\frac{X_1^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right)^2 = \frac{1}{4\sigma^8}\mathsf{E}(X_1^2 - \sigma^2)^2 = \frac{1}{4\sigma^8}\mathsf{D}X_1^2.$$

Осталось найти $\mathsf{D}X_1^2 = \mathsf{E}X_1^4 - (\mathsf{E}X_1^2)^2 = \mathsf{E}X_1^4 - \sigma^4$. Можно вспомнить некоторые формулы вероятности: величина $\xi = X_1/\sigma$ имеет стандартное нормальное распределение, и для неё

$$\mathsf{E}\xi^{2k} = (2k-1)!! = (2k-1)(2k-3)\cdot\ldots\cdot 3\cdot 1,$$

Тогда $\mathsf{E}\xi^4 = 3,\ X_1 = \xi\cdot\sigma$,

$$\mathsf{E}X_1^4 = \mathsf{E}\xi^4\cdot\sigma^4 = 3\sigma^4.$$

Если вспомнить не удалось, посчитаем заново. Воспользуемся свойствами характеристических функций и вычислим четвёртую производную характеристической функции $\varphi_\xi(t) = e^{-t^2/2}$ в нуле. Делать это удобнее через разложение в ряд:

$$e^{-t^2/2} = \sum_{k=0}^{\infty}\frac{1}{k!}\left(-\frac{t^2}{2}\right)^k = 1 - \frac{t^2}{2} + \frac{t^4}{8} - \ldots = 1 - \frac{t^2}{2} + \frac{3t^4}{4!} - \ldots$$

Производная четвёртого порядка в нуле равна коэффициенту при $\dfrac{t^4}{4!}$ ряда Тейлора: $\mathsf{E}\xi^4 = i^4\mathsf{E}\xi^4 = \varphi_\xi^{(4)}(0) = 3$.

Можно вычислить интеграл и напрямую. Интегрированием по частям получаем

$$\begin{aligned}\mathsf{E}\xi^4 &= \int\limits_{-\infty}^{\infty} y^4 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = -\frac{2}{\sqrt{2\pi}} \int\limits_{0}^{\infty} y^3 \, de^{-y^2/2} = \\ &= -\frac{2}{\sqrt{2\pi}} \left(y^3 \, e^{-y^2/2} \Big|_0^{\infty} - \int\limits_0^{\infty} e^{-y^2/2} \, dy^3 \right) = \\ &= \frac{2}{\sqrt{2\pi}} \cdot 3 \int\limits_0^{\infty} y^2 \, e^{-y^2/2} \, dy = 3 \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} y^2 \, e^{-y^2/2} \, dy = 3\mathsf{D}\xi = 3.\end{aligned}$$

Итак, $\mathsf{D}X_1^2 = \mathsf{E}X_1^4 - \sigma^4 = 2\sigma^4$,

$$I(\sigma^2) = \frac{1}{4\sigma^8} \mathsf{D}X_1^2 = \frac{1}{4\sigma^8} 2\sigma^4 = \frac{1}{2\sigma^4}.$$

Найдём дисперсию оценки $\sigma^{2*} = \overline{X^2}$ и сравним её с правой частью неравенства Рао — Крамера:

$$\mathsf{D}\overline{X^2} = \frac{1}{n^2} \mathsf{D} \sum_1^n X_i^2 = \frac{1}{n} \mathsf{D}X_1^2 = \frac{2\sigma^4}{n} = \frac{1}{nI(\sigma^2)},$$

Итак, оценка $\sigma^{2*} = \overline{X^2}$ R-эффективна и, следовательно, эффективна.

Упражнение. Пусть $X_1, \ldots, X_n$ — выборка объёма n из нормального распределения N_{a,σ^2}, где оба параметра a и σ^2 неизвестны. Проверить, является ли R-эффективной оценкой для σ^2 несмещённая выборочная дисперсия S_0^2, используя равенство: $\mathsf{D}\left((n-1)S_0^2/\sigma^2\right) = 2(n-1)$. Это равенство читатель сможет доказать несколькими главами позднее, когда познакомится с χ^2-распределением. При некотором терпении его можно доказать и непосредственным вычислением.

Пример 24. Пусть $X_1, \ldots, X_n$ — выборка объёма n из показательного распределения $\mathsf{E}_{1/\alpha}$ с параметром $1/\alpha$, где $\alpha > 0$. Проверим, является ли несмещённая (*почему*?) оценка $\alpha^* = \overline{X}$ эффективной оценкой для параметра α.

Найдём информацию Фишера относительно параметра α

$$I(\alpha) = \mathsf{E}_\alpha \left(\frac{\partial}{\partial\alpha} \ln f_\alpha(X_1) \right)^2.$$

Плотность данного показательного распределения имеет вид

$$f_\alpha(y) = \begin{cases} \dfrac{1}{\alpha}\, e^{-\frac{y}{\alpha}}, & \text{если } y > 0, \\ 0, & \text{если } y \leqslant 0. \end{cases}$$

Тогда $f_\alpha(X_1) = \dfrac{1}{\alpha}\, e^{-X_1/\alpha}$ п. н., $\ln f_\alpha(X_1) = -\ln\alpha - \dfrac{X_1}{\alpha}$,

$$\frac{\partial}{\partial\alpha} \ln f_\alpha(X_1) = -\frac{1}{\alpha} + \frac{X_1}{\alpha^2} = \frac{1}{\alpha^2}\,(X_1 - \alpha)$$

и информация Фишера равна

$$I(\alpha) = \frac{\mathsf{E}(X_1 - \alpha)^2}{\alpha^4} = \frac{\mathsf{D}X_1}{\alpha^4} = \frac{\alpha^2}{\alpha^4} = \frac{1}{\alpha^2}\,.$$

Найдём дисперсию оценки $\overline{X}$ и сравним с правой частью в неравенстве Рао — Крамера для несмещённых оценок:

$$\mathsf{D}\overline{X} = \frac{1}{n}\mathsf{D}X_1 = \frac{\alpha^2}{n} = \frac{1}{nI(\alpha)}\,.$$

Следовательно, оценка $\alpha^* = \overline{X}$ R-эффективна и поэтому является эффективной оценкой для параметра α.

Упражнение. Получить эту оценку методом моментов и методом максимального правдоподобия. Она действительно несмещённая? А ещё какими свойствами обладает?

Упражнение. Проверить, что для несмещённой оценки $\alpha^{**} = X_1$ равенство в неравенстве Рао — Крамера не достигается. Объяснить, почему, *исходя только из этого*, нельзя сделать вывод о её неэффективности в классе K_0. Сделать этот вывод на основании того, что оценки $\alpha^* = \overline{X}$ и $\alpha^{**} = X_1$ принадлежат классу оценок с одинаковым смещением, и одна из них эффективна. Сформулировать теорему о единственности эффективной оценки в классе оценок с фиксированным смещением.

Отсутствие равенства в неравенстве Рао — Крамера вовсе не означает, что оценка не является эффективной. Для некоторых семейств распределений это неравенство не точнó в том смысле, что самая маленькая из дисперсий несмещённых оценок всё же оказывается строго большей, чем правая часть неравенства. Приведём пример оценки, которая является эффективной (в этом мы убедимся много позже), но не R-эффективной, т. е. для неё не достигается равенство в неравенстве Рао — Крамера.

Пример 25. Пусть $X_1, \ldots, X_n$ — выборка объёма n из показательного распределения E_α с параметром α, где $\alpha > 0$. Возьмём чуть поправленную оценку метода моментов

$$\alpha^* = \frac{n-1}{n} \cdot \frac{1}{\overline{X}} = \frac{n-1}{X_1 + \ldots + X_n}\,.$$

Убедимся, что это несмещённая оценка. Согласно свойству устойчивости по суммированию для гамма-распределения, сумма $X_1 + \ldots + X_n$ независимых случайных величин с распределением $\mathsf{E} = \Gamma_{\alpha,1}$ имеет распределение $\Gamma_{\alpha,n}$ с плотностью распределения

$$\gamma_{\alpha,n}(y) = \begin{cases} \dfrac{\alpha^n}{(n-1)!}\, y^{n-1}\, e^{-\alpha y}, & y > 0, \\ 0, & y \leqslant 0. \end{cases}$$

Вычислим математическое ожидание

$$\begin{aligned} \mathsf{E}\alpha^* &= \mathsf{E}\left(\frac{n-1}{X_1 + \ldots + X_n}\right) = (n-1)\int\limits_0^\infty \frac{1}{y}\,\frac{\alpha^n}{(n-1)!}\, y^{n-1}\, e^{-\alpha y}\, dy = \\ &= \alpha\,\frac{(n-1)}{(n-1)!}\int\limits_0^\infty (\alpha y)^{n-2}\, e^{-\alpha y}\, d(\alpha y) = \frac{\alpha}{(n-2)!}\cdot(n-2)! = \alpha. \end{aligned}$$

Итак, оценка α^* принадлежит классу K_0. Информацию Фишера относительно параметра α мы вычислили в примере 17 (с. 48): $I(\alpha) = \dfrac{1}{\alpha^2}$. Найдём второй момент и дисперсию оценки α^*:

$$\begin{aligned} \mathsf{E}(\alpha^*)^2 &= \mathsf{E}\frac{(n-1)^2}{(\sum X_i)^2} = (n-1)^2\int\limits_0^\infty \frac{1}{y^2}\,\frac{\alpha^n}{(n-1)!}\, y^{n-1}\, e^{-\alpha y}\, dy = \\ &= \alpha^2\,\frac{(n-1)^2}{(n-1)!}\int\limits_0^\infty (\alpha y)^{n-3}\, e^{-\alpha y}\, d(\alpha y) = \alpha^2\,\frac{(n-1)}{(n-2)!}\cdot(n-3)! = \frac{n-1}{n-2}\,\alpha^2, \end{aligned}$$

тогда

$$\mathsf{D}\alpha^* = \mathsf{E}(\alpha^*)^2 - (\mathsf{E}\alpha^*)^2 = \frac{n-1}{n-2}\,\alpha^2 - \alpha^2 = \frac{\alpha^2}{n-2}\,.$$

Подставив дисперсию и информацию Фишера в неравенство Рао — Крамера, получим, что при любом n есть *строгое неравенство*

$$\mathsf{D}\alpha^* = \frac{\alpha^2}{n-2} > \frac{\alpha^2}{n} = \frac{1}{nI(\alpha)}\,.$$

В главе XI мы докажем эффективность оценки α^*.

§ 4. Вопросы и упражнения

1. Проверить эффективность оценок максимального правдоподобия для неизвестных параметров следующих семейств распределений: B_p, Π_λ, N_{a,σ^2} при известном a, $\mathrm{B}_{m,p}$ при известном m.

2. Выполнить все упражнения, содержащиеся в тексте главы IV.

3. Пусть $X_1, \ldots, X_n$ — выборка из параметрического семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta$. Доказать, что если оценка θ^* является R-эффективной оценкой для θ в классе оценок со смещением $b(\theta) = \theta/n$, то она состоятельна.

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

5. Выполнены ли условия регулярности для семейства распределений $\mathcal{F}_\theta$ с плотностью распределения $4(\theta - y)^3/\theta^4$ на отрезке $[0, \theta]$?

6. Семейство распределений $\{\mathcal{F}_\theta;\ \theta \in \Theta\}$ называется экспоненциальным, если функция правдоподобия $f(\vec{X};\,\theta)$ допускает представление

$$f(\vec{X};\,\theta) = e^{A(\theta)T(\vec{X})+B(\theta)}h(\vec{X}).$$

Проверить, являются ли экспоненциальными семейства распределений: N_{a,σ^2} при известном σ^2, N_{a,σ^2} при известном a, $\Gamma_{\alpha,\lambda}$ при известном λ, $\Gamma_{\alpha,\lambda}$ при известном α, Π_λ.

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

ГЛАВА V

ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ

Пусть, как обычно, имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из распределения $\mathcal{F}_\theta$ с неизвестным параметром $\theta \in \Theta \subseteq \mathbb{R}$. До сих пор мы занимались «точечным оцениванием» неизвестного параметра — находили оценку (для каждой реализации выборки — число), способную в некотором смысле заменить параметр. Существует другой подход к оцениванию, при котором мы указываем интервал, накрывающий параметр с заданной наперед вероятностью. Такой подход называется интервальным оцениванием. Сразу заметим: чем больше уверенность в том, что параметр лежит в интервале, тем шире интервал. Поэтому бессмысленно искать диапазон, внутри которого θ содержится гарантированно,— это вся область Θ.

§ 1. Доверительные интервалы

Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка объёма n из распределения $\mathcal{F}_\theta$ с параметром $\theta \in \Theta \subseteq \mathbb{R}$.

Определение 16. Пусть $0 < \varepsilon < 1$. Интервал со случайными концами $(\theta^-, \theta^+) = \bigl(\theta^-(\vec{X}, \varepsilon), \theta^+(\vec{X}, \varepsilon)\bigr)$ называется *доверительным интервалом* для параметра θ *уровня доверия* $1 - \varepsilon$, если для любого $\theta \in \Theta$

$$\mathsf{P}\left(\theta^- < \theta < \theta^+\right) \geqslant 1 - \varepsilon.$$

Определение 17. Пусть $0 < \varepsilon < 1$. Интервал со случайными концами $(\theta^-, \theta^+) = \bigl(\theta^-(\vec{X}, \varepsilon), \theta^+(\vec{X}, \varepsilon)\bigr)$ называется *асимптотическим* доверительным интервалом для параметра θ (асимптотического) уровня доверия $1 - \varepsilon$, если для любого $\theta \in \Theta$

$$\liminf_{n\to\infty} \mathsf{P}\left(\theta^- < \theta < \theta^+\right) \geqslant 1 - \varepsilon.$$

На самом деле в определении 17 речь идёт, конечно, не об одном интервале, но о *последовательности* интервалов, зависящих от n.

Замечание 11. Случайны здесь границы интервала $(\theta^-,\,\theta^+)$, поэтому читают формулу $\mathsf{P}\left(\theta^- < \theta < \theta^+\right)$ как «интервал $(\theta^-,\,\theta^+)$ накрывает параметр θ», а не как «θ лежит в интервале…».

Замечание 12. Неравенство «$\geqslant 1-\varepsilon$» обычно соответствует дискретным распределениям, когда нельзя обязаться добиться равенства: например, для $\xi \Subset \mathrm{B}_{1/2}$ при любом x равенство $\mathsf{P}(\xi < x) = 0{,}25$ невозможно, а неравенство имеет смысл:

$$\mathsf{P}(\xi < x) \geqslant 0{,}25 \quad \text{для} \quad x > 0.$$

Если вероятность доверительному интервалу накрыть параметр *равна* $1-\varepsilon$ (или стремится к $1-\varepsilon$), интервал называют *точным* (или *асимптотически точным*) доверительным интервалом.

Прежде чем рассматривать какие-то регулярные способы построения точных и асимптотических доверительных интервалов, разберем два примера, предлагающих очень похожие способы, и затем попробуем извлечь из этих примеров некоторую общую философию построения точных и асимптотически точных доверительных интервалов.

Пример 26. Пусть $X_1,\,\ldots,\,X_n$ — выборка объёма n из нормального распределения $\mathrm{N}_{a,\,\sigma^2}$, где $a \in \mathbb{R}$ — неизвестный параметр, а значение $\sigma > 0$ известно. Требуется при произвольном n построить точный доверительный интервал для параметра a уровня доверия $1-\varepsilon$.

Знаем, что нормальное распределение устойчиво по суммированию.

Лемма 3. *Пусть случайные величины* ξ_i, *где* $i = 1,\,2$, *имеют нормальные распределения* $\mathrm{N}_{a_i,\,\sigma_i^2}$ *и независимы. Тогда случайная величина* $\eta = b\xi_1 + c\xi_2 + d$ *имеет нормальное распределение с параметрами*

$$\mathsf{E}\eta = b\,a_1 + c\,a_2 + d, \qquad \mathsf{D}\eta = b^2\sigma_1^2 + c^2\sigma_2^2.$$

Упражнение. Доказать лемму 3.

Поэтому распределение суммы элементов выборки при любом её объёме n нормально: $n\overline{X} = X_1 + \ldots + X_n \Subset \mathrm{N}_{na,\,n\sigma^2}$, а центрированная и нормированная величина

$$\eta = \frac{n\overline{X} - na}{\sigma\sqrt{n}} = \sqrt{n}\,\frac{\overline{X} - a}{\sigma}$$

имеет стандартное нормальное распределение.

По заданному $\varepsilon \in (0,\,1)$ найдём число $c > 0$ такое, что

$$\mathsf{P}(-c < \eta < c) = 1 - \varepsilon.$$

Число c является квантилью уровня $1-\dfrac{\varepsilon}{2}$ стандартного нормального распределения (рис. 7):

$$\mathsf{P}(-c<\eta<c)=\Phi_{0,1}(c)-\Phi_{0,1}(-c)=2\Phi_{0,1}(c)-1=1-\varepsilon,\ \ \Phi_{0,1}(c)=1-\frac{\varepsilon}{2}.$$

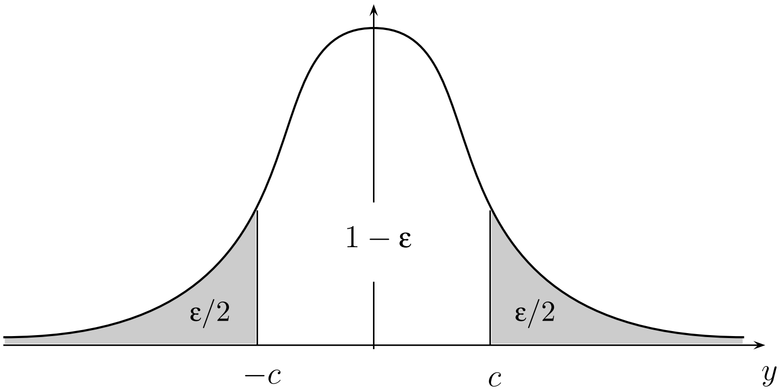

Рис. 7. Квантили стандартного нормального распределения

По заданному ε в таблице значений функции $\Phi_{0,1}(x)$ найдём квантили $c=\tau_{1-\varepsilon/2}$ или $-c=\tau_{\varepsilon/2}$. Разрешив затем неравенство $-c<\eta<c$ относительно a, получим точный доверительный интервал:

$$\begin{aligned}1-\varepsilon &= \mathsf{P}(-c<\eta<c)=\mathsf{P}\left(-c<\sqrt{n}\,\frac{\overline{X}-a}{\sigma}<c\right)=\\ &= \mathsf{P}\left(\overline{X}-\frac{c\sigma}{\sqrt{n}}<a<\overline{X}+\frac{c\sigma}{\sqrt{n}}\right).\end{aligned}$$

Можно подставить $c=\tau_{1-\varepsilon/2}$:

$$\mathsf{P}\left(\overline{X}-\frac{\sigma\tau_{1-\varepsilon/2}}{\sqrt{n}}<a<\overline{X}+\frac{\sigma\tau_{1-\varepsilon/2}}{\sqrt{n}}\right)=1-\varepsilon.$$

Итак, искомый точный доверительный интервал уровня доверия $1-\varepsilon$ имеет вид

$$\left(\overline{X}-\frac{\sigma\tau_{1-\varepsilon/2}}{\sqrt{n}},\ \overline{X}+\frac{\sigma\tau_{1-\varepsilon/2}}{\sqrt{n}}\right). \qquad (14)$$

Упражнение. Имеет смысл ответить на несколько вопросов.

1. Зачем мы брали симметричные квантили? Почему не брать границы для η вида $\mathsf{P}(\tau_{\varepsilon/3}<\eta<\tau_{1-2\varepsilon/3})=1-\varepsilon$? Изобразить эти квантили на графике плотности. Как изменилось расстояние между квантилями? Как изменится длина доверительного интервала?

2. Какой из двух доверительных интервалов одного уровня доверия и разной длины следует предпочесть?

3. Какова середина полученного в примере 26 доверительного интервала? Какова его длина? Что происходит с границами доверительного интервала при $n \to \infty$? Как быстро это с ними происходит?

Пример 27. Пусть $X_1, \ldots, X_n$ — выборка объёма n из показательного распределения E_α, где $\alpha > 0$. Требуется построить асимптотический (асимптотически точный) доверительный интервал для параметра α уровня доверия $1-\varepsilon$.

Вспомним ЦПТ:

$$\frac{\sum X_i - n\,\mathsf{E}X_1}{\sqrt{n\,\mathsf{D}X_1}} = \sqrt{n}\,\frac{\overline{X}-1/\alpha}{1/\alpha} = \sqrt{n}\,\left(\alpha\overline{X}-1\right) \Rightarrow \eta \in \mathrm{N}_{0,1}.$$

Возьмём $c = \tau_{1-\varepsilon/2}$ — квантиль стандартного нормального распределения. По определению слабой сходимости, при $n \to \infty$

$$\mathsf{P}\left(-c < \sqrt{n}\,\left(\alpha\overline{X}-1\right) < c\right) \to \mathsf{P}(-c < \eta < c) = 1-\varepsilon.$$

Разрешив относительно α неравенство $-\tau_{1-\varepsilon/2} < \sqrt{n}\,\left(\alpha\overline{X}-1\right) < \tau_{1-\varepsilon/2}$, получим асимптотический доверительный интервал:

$$\mathsf{P}\left(\frac{1}{\overline{X}} - \frac{\tau_{1-\varepsilon/2}}{\sqrt{n}\,\overline{X}} < \alpha < \frac{1}{\overline{X}} + \frac{\tau_{1-\varepsilon/2}}{\sqrt{n}\,\overline{X}}\right) \to 1-\varepsilon \quad \text{при } n \to \infty.$$

§ 2. Принципы построения доверительных интервалов

Общий принцип построения точных доверительных интервалов. Чтобы построить точный доверительный интервал, необходимо реализовать следующие шаги.

1. Найти функцию $G(\vec{X}, \theta)$, распределение которой $\mathcal{G}$ не зависит от параметра θ. Необходимо, чтобы $G(\vec{X}, \theta)$ была обратима по θ при любом фиксированном $\vec{X}$.

2. Найти числа g_1 и g_2 — квантили распределения $\mathcal{G}$, для которых

$$1-\varepsilon = \mathsf{P}(g_1 < G(\vec{X}, \theta) < g_2).$$

3. Разрешив неравенство $g_1 < G(\vec{X}, \theta) < g_2$ относительно θ, получить точный доверительный интервал.

Замечание 13. Часто в качестве g_1 и g_2 берут квантили распределения $\mathcal{G}$ уровней $\varepsilon/2$ и $1-\varepsilon/2$. Но, вообще говоря, квантили следует выбирать так, чтобы получить самый короткий доверительный интервал.

Пример 28. Попробуем, пользуясь приведённой выше схемой, построить точный доверительный интервал для параметра $\theta > 0$ равномерного на отрезке $[\theta,\, 2\theta]$ распределения.

Мы знаем, что если X_i имеют распределение $\mathrm{U}_{\theta,\,2\theta}$, то $Y_i = \dfrac{X_i}{\theta} - 1$ имеют распределение $\mathrm{U}_{0,\,1}$. Тогда величина

$$Y_{(n)} = \max\{Y_1,\,\ldots,\,Y_n\} = \frac{\max\{X_1,\,\ldots,\,X_n\}}{\theta} - 1 = \frac{X_{(n)}}{\theta} - 1 = G(\vec{X},\,\theta)$$

распределена так же, как максимум из n независимых равномерно распределённых на отрезке $[0,\,1]$ случайных величин, т. е. имеет не зависящую от параметра θ функцию распределения

$$F_{Y_{(n)}}(y) = \mathsf{P}(\eta < y) = \begin{cases} 0, & \text{если } y < 0 \\ y^n, & \text{если } y \in [0,\,1] \\ 1, & \text{если } y > 1. \end{cases}$$

Для любых положительных g_1 и g_2

$$\mathsf{P}\left(g_1 < G(\vec{X},\,\theta) < g_2\right) = \mathsf{P}\left(\frac{X_{(n)}}{g_2+1} < \theta < \frac{X_{(n)}}{g_1+1}\right). \tag{15}$$

Длина доверительного интервала 15 равна

$$X_{(n)} \cdot \frac{g_2 - g_1}{(g_1+1)(g_2+1)}$$

и уменьшается с ростом g_1 и g_2 и с их сближением. Выберем квантили g_1 и g_2 так, чтобы длина интервала стала наименьшей.

Плотность распределения $Y_{(n)}$ на отрезке $[0,\,1]$ равна ny^{n-1} и монотонно возрастает. Поэтому самые большие значения g_1 и g_2 при самом маленьком расстоянии между ними и при фиксированной площади $1-\varepsilon$ под графиком плотности возможны при $g_2 = 1$, а g_1 таком, что

$$1 - \varepsilon = \mathsf{P}(g_1 < Y_{(n)} < 1) = F_{Y_{(n)}}(1) - F_{Y_{(n)}}(g_1) = 1 - g_1^n,$$

т. е. при $g_1 = \sqrt[n]{\varepsilon}$. Подставим найденные квантили в (15):

$$1 - \varepsilon = \mathsf{P}\left(\sqrt[n]{\varepsilon} < Y_{(n)} < 1\right) = \mathsf{P}\left(\frac{X_{(n)}}{2} < \theta < \frac{X_{(n)}}{1+\sqrt[n]{\varepsilon}}\right).$$

Упражнение. Можно ли, пользуясь схемой примера 26, построить точный доверительный интервал для параметра σ при известном значении a, если разрешить неравенство

$$-c < \sqrt{n}\,\frac{\overline{X} - a}{\sigma} < c$$

относительно σ? Чем плох интервал бесконечной длины? *А получился ли* интервал бесконечной длины?

Можно построить точный доверительный интервал для параметра σ при известном значении a (*сделайте это*!) по функции

$$G(\vec{X}, \sigma^2, a) = \sqrt{n}\,\frac{|\overline{X} - a|}{\sigma}.$$

Но при неизвестном a по этой функции нельзя построить точный доверительный интервал для параметра σ: границы интервала не могут зависеть от неизвестного параметра a. В следующей главе мы займёмся поиском подходящих для решения этой задачи функций.

Общий принцип построения асимптотических доверительных интервалов. Необходимо проделать следующее.

1. Найти функцию $G(\vec{X}, \theta)$, *слабо сходящуюся* к распределению $\mathcal{G}$, не зависящему от параметра θ. Необходимо, чтобы $G(\vec{X}, \theta)$ была обратима по θ при любом фиксированном $\vec{X}$.

2. Найти числа g_1 и g_2 — квантили распределения $\mathcal{G}$, для которых

$$\mathsf{P}(g_1 < G(\vec{X}, \theta) < g_2) \to \mathsf{P}(g_1 < \eta < g_2) = 1 - \varepsilon, \quad \text{где } \eta \in \mathcal{G}.$$

3. Разрешив неравенство $g_1 < G(\vec{X}, \theta) < g_2$ относительно θ, получить асимптотически точный доверительный интервал.

Следующий пример (как и пример 27) показывает, что ЦПТ дает универсальный вид функции G для построения *асимптотических* доверительных интервалов.

Пример 29. Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения Пуассона Π_λ, где $\lambda > 0$. Требуется построить асимптотический доверительный интервал для параметра λ уровня доверия $1 - \varepsilon$.

Согласно ЦПТ

$$G(\vec{X}, \lambda) = \frac{X_1 + \ldots + X_n - n\mathsf{E}X_1}{\sqrt{n\mathsf{D}X_1}} = \sqrt{n}\,\frac{\overline{X} - \lambda}{\sqrt{\lambda}} \Rightarrow \eta,$$

где случайная величина η имеет стандартное нормальное распределение. Пусть $c = \tau_{1-\varepsilon/2}$ — квантиль уровня $1 - \varepsilon/2$ стандартного нормального распределения. По определению слабой сходимости, при $n \to \infty$

$$\mathsf{P}\left(-c < \sqrt{n}\,\frac{\overline{X} - \lambda}{\sqrt{\lambda}} < c\right) \to \mathsf{P}(-c < \eta < c) = 1 - \varepsilon.$$

Но разрешить неравенство под знаком вероятности относительно λ непросто: мешает корень в знаменателе. Попробуем от него избавиться. Не испортится ли сходимость, если мы заменим $\sqrt{\lambda}$, например, на $\sqrt{\overline{X}}$, т. е. умножим функцию $G(\vec{X}, \lambda)$ на $\sqrt{\lambda}$ и поделим на $\sqrt{\overline{X}}$?

Воспользуемся (*в который раз*?) следующим свойством слабой сходимости: если $\xi_n \xrightarrow{\text{p}} 1$ и $\eta_n \Rightarrow \eta$, то $\xi_n \cdot \eta_n \Rightarrow \eta$. Оценка $\lambda^* = \overline{X}$ состоятельна, поэтому

$$\sqrt{\frac{\lambda}{\overline{X}}} \xrightarrow{\text{p}} 1 \quad \text{при} \quad n \to \infty.$$

Тогда

$$\sqrt{\frac{\lambda}{\overline{X}}} \cdot \sqrt{n}\,\frac{\overline{X} - \lambda}{\sqrt{\lambda}} = \sqrt{n}\,\frac{\overline{X} - \lambda}{\sqrt{\overline{X}}} \Rightarrow \eta \in \mathrm{N}_{0,1}.$$

Поэтому при $c = \tau_{1-\varepsilon/2}$

$$\mathsf{P}\left(-c < \sqrt{n}\,\frac{\overline{X} - \lambda}{\sqrt{\overline{X}}} < c\right) \to \mathsf{P}(-c < \eta < c) = 1 - \varepsilon.$$

Разрешив неравенство под знаком вероятности относительно λ, получим

$$\mathsf{P}\left(\overline{X} - \frac{c\sqrt{\overline{X}}}{\sqrt{n}} < \lambda < \overline{X} + \frac{\sqrt{\overline{X}}}{\sqrt{n}}\right) \to 1 - \varepsilon \text{ при } n \to \infty.$$

Итак, искомый асимптотический доверительный интервал уровня доверия $1 - \varepsilon$ имеет вид

$$\left(\overline{X} - \frac{\tau_{1-\varepsilon/2}\sqrt{\overline{X}}}{\sqrt{n}},\ \overline{X} + \frac{\tau_{1-\varepsilon/2}\sqrt{\overline{X}}}{\sqrt{n}}\right).$$

Для построения асимптотических доверительных интервалов можно использовать асимптотически нормальные оценки (это тоже ЦПТ).

Теорема 17. *Пусть θ^* — АНО для параметра θ с коэффициентом $\sigma^2(\theta)$, и функция $\sigma(\theta)$ непрерывна в области Θ. Тогда интервал*

$$\left(\theta^* - \frac{\tau_{1-\varepsilon/2}\,\sigma(\theta^*)}{\sqrt{n}},\ \theta^* + \frac{\tau_{1-\varepsilon/2}\,\sigma(\theta^*)}{\sqrt{n}}\right)$$

является асимптотическим доверительным интервалом для параметра θ уровня доверия $1 - \varepsilon$.

Доказательство. По определению АНО

$$\sqrt{n}\,\frac{\theta^* - \theta}{\sigma(\theta)} \Rightarrow \eta \in \mathrm{N}_{0,1}.$$

Оценка θ^* асимптотически нормальна и, следовательно, состоятельна. Применяя к ней непрерывную функцию $\sigma(\theta)$, получаем

$$\sigma(\theta^*) \xrightarrow{\text{p}} \sigma(\theta), \qquad \frac{\sigma(\theta)}{\sigma(\theta^*)} \xrightarrow{\text{p}} 1.$$

В качестве функции $G(\vec{X}, \theta)$ возьмём

$$G(\vec{X}, \theta) = \sqrt{n}\, \frac{\theta^* - \theta}{\sigma(\theta^*)} = \frac{\sigma(\theta)}{\sigma(\theta^*)} \cdot \sqrt{n}\, \frac{\theta^* - \theta}{\sigma(\theta)} \Rightarrow \eta \in \mathrm{N}_{0,1}.$$

Пусть $c = \tau_{1-\varepsilon/2}$ — квантиль стандартного нормального распределения. Разрешив неравенство

$$-c < \sqrt{n}\, \frac{\theta^* - \theta}{\sigma(\theta^*)} < c$$

относительно θ, получим асимптотический доверительный интервал

$$\left(\theta^* - \frac{c\,\sigma(\theta^*)}{\sqrt{n}},\ \ \theta^* + \frac{c\,\sigma(\theta^*)}{\sqrt{n}}\right).$$

Полезно отметить, что длина этого интервала ведёт себя как $1/\sqrt{n}$, потому что именно с такой скоростью асимптотически нормальная оценка сближается с параметром. □

§ 3. Вопросы и упражнения

1. Что больше: квантиль стандартного нормального распределения уровня 0,05 или уровня 0,1? Почему? Нарисовать их на графике плотности этого распределения.

2. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$. Построить точные доверительные интервалы для параметра θ, используя статистики X_1 и $X_{(n)}$.

3. Пусть $X_1, \ldots, X_n$ — выборка из показательного распределения с параметром α. Построить точные доверительные интервалы для параметра α, используя статистики X_1 и $X_{(1)}$.

ГЛАВА VI

РАСПРЕДЕЛЕНИЯ, СВЯЗАННЫЕ С НОРМАЛЬНЫМ

В предыдущей главе мы построили в числе других точный доверительный интервал для параметра a нормального распределения при известном σ^2. Остался нерешённым вопрос: как построить точные доверительные интервалы для σ при известном и при неизвестном a, а также для a при неизвестном σ? Мы уже видели, что для решения этих задач требуется отыскать такие функции от выборки и неизвестных параметров, распределения которых не зависят от этих параметров. При этом сами искомые функции не должны зависеть ни от каких лишних параметров. Особый интерес к нормальному распределению связан, разумеется, с центральной предельной теоремой: почти всё в этом мире нормально (или близко к тому). В этой главе мы изучим новые распределения, связанные с нормальным, их свойства и свойства выборок из нормального распределения.

§ 1. Основные статистические распределения

Гамма-распределение. С гамма-распределением мы познакомились в курсе теории вероятностей. Нам понадобится свойство устойчивости по суммированию этого распределения.

Лемма 4. *Пусть* $X_1, \ldots, X_n$ *независимы, и* ξ_i *имеет гамма-распределение* $\Gamma_{\alpha,\lambda_i}$, $i = 1, \ldots, n$. *Тогда их сумма* $S_n = \xi_1 + \ldots + \xi_n$ *имеет гамма-распределение с параметрами* α *и* $\lambda_1 + \ldots + \lambda_n$.

Оказывается, что квадрат случайной величины со стандартным нормальным распределением имеет гамма-распределение.

Лемма 5. *Если* ξ *имеет стандартное нормальное распределение, то* ξ^2 *имеет гамма-распределение* $\Gamma_{1/2,\,1/2}$.

Доказательство. Найдём производную функции распределения величины ξ^2 и убедимся, что она является плотностью распределения. При $y \leqslant 0$

$$F_{\xi^2}(y) = \mathsf{P}(\xi^2 < y) = 0, \text{ поэтому } f_{\xi^2}(y) = 0.$$

При $y > 0$

$$F_{\xi^2}(y) = \mathsf{P}(\xi^2 < y) = \mathsf{P}(-\sqrt{y} < \xi < \sqrt{y}) = F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}).$$

Тогда

$$f_{\xi^2}(y) = \big(F_{\xi^2}(y)\big)' = F'_\xi(\sqrt{y})\cdot(\sqrt{y})' - F'_\xi(-\sqrt{y})\cdot(-\sqrt{y})' =$$

$$= \big(f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})\big)\cdot\frac{1}{2\sqrt{y}} = \frac{f_\xi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi y}}\,e^{-y/2}.$$

Но функция $f_{\xi^2}(y)$, равная 0 при $y \leqslant 0$ и равная

$$f_{\xi^2}(y) = \frac{1}{\sqrt{2\pi y}}\,e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)}\,y^{1/2-1}\,e^{-y/2}$$

при $y > 0$, является плотностью гамма-распределения $\Gamma_{1/2,\,1/2}$. □

Распределение χ^2 Пирсона. Из лемм 4 и 5 следует утверждение.

Лемма 6. *Если $\xi_1,\ \ldots,\ \xi_k$ независимы и имеют стандартное нормальное распределение, то случайная величина*

$$\chi^2 = \xi_1^2 + \ldots + \xi_k^2$$

имеет гамма-распределение $\Gamma_{1/2,\,k/2}$.

В статистике это распределение играет совершенно особую роль и имеет собственное название.

Определение 18. Распределение суммы k квадратов независимых случайных величин со стандартным нормальным распределением называется распределением χ^2 (*хи-квадрат*) с k степенями свободы и обозначается H_k.

Согласно лемме 6, распределение H_k совпадает с $\Gamma_{1/2,\,k/2}$. Поэтому плотность распределения H_k равна

$$f(y) = \begin{cases} \dfrac{1}{2^{k/2}\,\Gamma(k/2)}\,y^{\frac{k}{2}-1}e^{-y/2}, & \text{если } y > 0;\\ 0, & \text{если } y \leqslant 0.\end{cases}$$

Заметим ещё, что $\mathrm{H}_2 = \Gamma_{1/2,\,1} = \mathrm{E}_{1/2}$ — показательное распределение с параметром $\alpha = 1/2$.

Плотности распределений H_k при $k = 1,\ 2,\ 4,\ 8$ показаны на рис. 8.

Упражнение. Доказать, что при $k \geqslant 2$ максимум плотности распределения H_k достигается в точке $y = k - 2$.

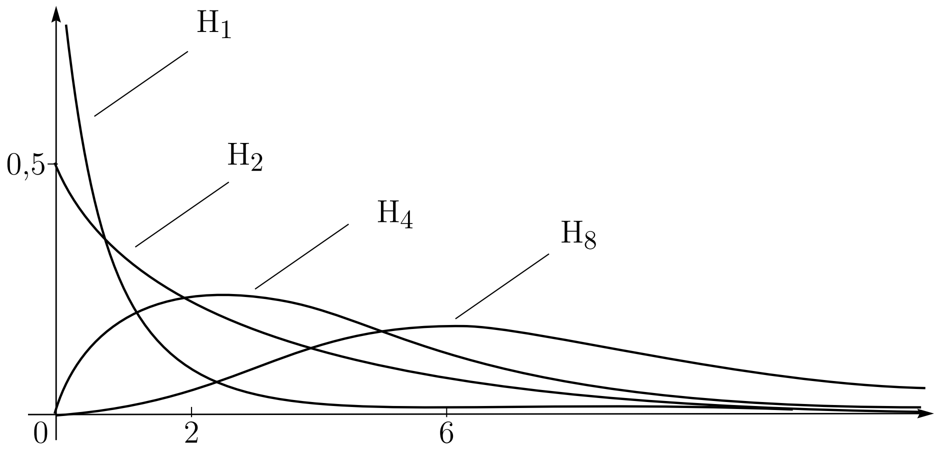

Рис. 8. Плотности χ^2-распределений с различным числом степеней свободы

Рассмотрим свойства χ^2-распределения. Устойчивость его относительно суммирования следует из устойчивости гамма-распределения.

Свойство 1. Если случайные величины $\chi^2 \in \mathrm{H}_k$ и $\psi^2 \in \mathrm{H}_m$ независимы, то их сумма $\chi^2 + \psi^2$ имеет распределение H_{k+m}.

Доказательство. Свойство устойчивости можно доказать и непосредственно. Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Тогда случайная величина χ^2 распределена так же, как $\xi_1^2 + \ldots + \xi_k^2$, величина ψ^2 распределена так же, как $\xi_{k+1}^2 + \ldots + \xi_{k+m}^2$, а их сумма — как $\xi_1^2 + \ldots + \xi_{k+m}^2$, т. е. имеет распределение H_{k+m}. □

Свойство 2. Если величина χ^2 имеет распределение H_k, то

$$\mathsf{E}\chi^2 = k \quad \text{и} \quad \mathsf{D}\chi^2 = 2k.$$

Доказательство. Пусть $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Тогда

$$\mathsf{E}\xi_1^2 = 1, \qquad \mathsf{D}\xi_1^2 = \mathsf{E}\xi_1^4 - (\mathsf{E}\xi_1^2)^2 = 3 - 1 = 2.$$

Четвёртый момент стандартного нормального распределения мы вычислили в примере 23 (с. 54). Поэтому

$$\mathsf{E}\chi^2 = \mathsf{E}(\xi_1^2 + \ldots + \xi_k^2) = k, \quad \mathsf{D}\chi^2 = \mathsf{D}(\xi_1^2 + \ldots + \xi_k^2) = 2k. \quad \square$$

Свойство 3. Пусть $\chi_n^2 \in \mathrm{H}_n$. Тогда при $n \to \infty$

$$\frac{\chi_n^2}{n} \xrightarrow{\mathrm{p}} 1, \qquad \frac{\chi_n^2 - n}{\sqrt{2n}} \Rightarrow \mathrm{N}_{0,1}.$$

Доказательство. При любом n случайная величина χ_n^2 распределена так же, как $\xi_1^2+\ldots+\xi_n^2$, где все случайные величины ξ_i независимы и имеют стандартное нормальное распределение. Применяя ЗБЧ и ЦПТ, получаем сходимости

$$\frac{\xi_1^2+\ldots+\xi_n^2}{n} \xrightarrow{\text{p}} \mathsf{E}\xi_1^2=1, \qquad \frac{\xi_1^2+\ldots+\xi_n^2-n}{\sqrt{2n}} = \frac{\xi_1^2+\ldots+\xi_n^2-n\mathsf{E}\xi_1^2}{\sqrt{n\mathsf{D}\xi_1^2}} \Rightarrow \mathrm{N}_{0,1},$$

равносильные утверждению свойства 3. □

Распределение H_n при небольших n табулировано. Однако при большом числе степеней свободы для вычисления функции этого распределения или, наборот, его квантилей пользуются различными аппроксимациями с помощью стандартного нормального распределения. Одно из приближений предлагается в следующем свойстве, более точную аппроксимацию Уилсона — Хилферти читатель найдёт в упражнениях в конце главы.

Свойство 4 (аппроксимация Фишера). Пусть $\chi_n^2 \in \mathrm{H}_n$. Тогда при $n\to\infty$ имеет место слабая сходимость

$$\sqrt{2\chi_n^2}-\sqrt{2n-1} \Rightarrow \mathrm{N}_{0,1},$$

поэтому при больших n можно пользоваться аппроксимацией для функции распределения $\mathrm{H}_n(x)=\mathsf{P}\left(\chi_n^2<x\right)$:

$$\mathrm{H}_n(x) \approx \Phi_{0,1}\left(\sqrt{2x}-\sqrt{2n-1}\right). \tag{16}$$

Доказательство. Заметим сначала, что $\sqrt{2n}-\sqrt{2n-1} \to 0$ при $n\to\infty$ (*проверить*!), поэтому достаточно обосновать сходимость

$$\sqrt{2\chi_n^2}-\sqrt{2n} \Rightarrow \mathrm{N}_{0,1}.$$

Домножим и поделим эту разность на сопряжённое, и представим результат в виде:

$$\sqrt{2\chi_n^2}-\sqrt{2n} = \frac{2}{1+\sqrt{\chi_n^2/n}} \cdot \frac{\chi_n^2-n}{\sqrt{2n}}.$$

Первый сомножитель по свойству 3 сходится по вероятности к единице, а второй слабо сходится к $\mathrm{N}_{0,1}$. По свойствам слабой сходимости, их произведение слабо сходится к $\mathrm{N}_{0,1}$.

Для доказательства (16) заметим, что

$$\mathsf{P}\left(\chi_n^2<x\right) = \mathsf{P}\left(\sqrt{2\chi_n^2}-\sqrt{2n-1} < \sqrt{2x}-\sqrt{2n-1}\right). \qquad \square$$

Свойство 5. Если случайные величины $\xi_1, \ldots, \xi_k$ независимы и имеют нормальное распределение N_{a,σ^2}, то

$$\chi_k^2 = \sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma} \right)^2 \in \mathrm{H}_k.$$

Упражнение. Доказать свойство 5.

Распределение Стью́дента. Английский статистик Госсет, публиковавший научные труды под псевдонимом Стьюдент, ввёл следующее распределение.

Определение 19. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$t_k = \frac{\xi_0}{\sqrt{\dfrac{\xi_1^2 + \ldots + \xi_k^2}{k}}}$$

называется *распределением Стью́дента* с k степенями свободы и обозначается T_k.

Распределение Стьюдента совпадает с распределением случайной величины $t_k = \dfrac{\xi}{\sqrt{\chi_k^2 / k}}$, где $\xi \in \mathrm{N}_{0,1}$ и $\chi_k^2 \in \mathrm{H}_k$ независимы.

Читатель может найти плотность распределения Стьюдента самостоятельно, либо посмотреть, как это делается в [1, § 2, гл. 2]. Плотность распределения Стьюдента с k степенями свободы равна

$$f_k(y) = \frac{\Gamma\big((k+1)/2\big)}{\sqrt{\pi k}\,\Gamma(k/2)} \left(1 + \frac{y^2}{k} \right)^{-(k+1)/2}. \tag{17}$$

Свойство 6. Распределение Стьюдента симметрично: если случайная величина t_k имеет распределение Стьюдента T_k с k степенями свободы, то и $-t_k$ имеет такое же распределение.

Упражнение. Доказать.

Свойство 7. Распределение Стьюдента T_n слабо сходится к стандартному нормальному распределению при $n \to \infty$.

Доказательство. По свойству 3, $\chi_n^2 / n \xrightarrow{\mathrm{p}} 1$ при $n \to \infty$. По свойствам слабой сходимости получаем

$$t_k = \frac{\xi}{\sqrt{\chi_k^2 / k}} \Rightarrow \xi \in \mathrm{N}_{0,1}. \qquad \square$$

Графики плотностей стандартного нормального распределения и распределения Стьюдента приведены для сравнения на рис. 9.

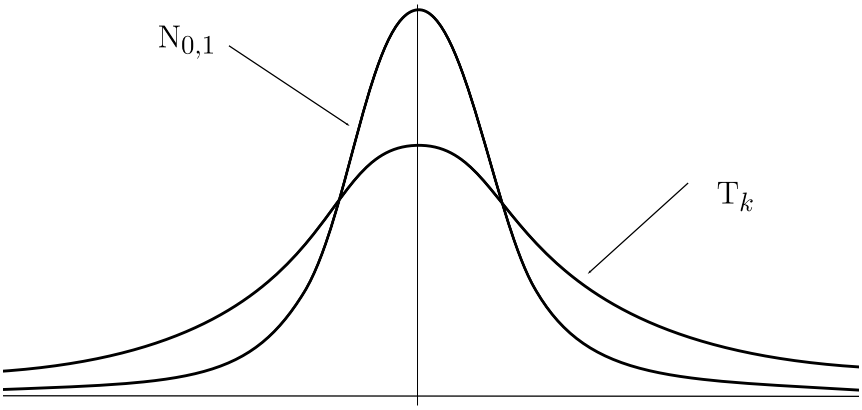

Рис. 9. Плотности распределений T_k и $N_{0,\,1}$

Отметим, что распределение Стьюдента табулировано: если в каких-то доверительных интервалах появятся квантили этого распределения, то мы найдём их по соответствующей таблице, либо, при больших n, используем нормальную аппроксимацию для распределения Стьюдента.

Распределение Стьюдента с одной степенью свободы есть стандартное распределение Коши. Действительно, если подставить $k = 1$ в плотность (17) и учесть $\Gamma(1/2) = \sqrt{\pi}$ и $\Gamma(1) = 1$, то получится плотность распределения Коши:

$$f_1(y) = \frac{1}{\pi}\,\left(1 + y^2\right)^{-1}.$$

Упражнение. Как получить случайную величину с распределением Коши, имея две независимые случайные величины со стандартным нормальным распределением?

Свойство 8. У распределения Стьюдента T_k существуют только моменты порядка $m < k$ и не существуют моменты порядка $m \geqslant k$. При этом все существующие моменты нечётного порядка равны нулю.

Упражнение. Рассмотрите плотность (17) и убедитесь в сходимости или расходимости на бесконечности при соответствующих m интегралов

$$C(k) \cdot \int\limits_{-\infty}^{\infty} |y|^m \cdot \frac{1}{(k + y^2)^{(k+1)/2}}\, dy.$$

Упражнение. Существует ли $\mathsf{D}t_2$, если $t_2 \in T_2$?

Распределение Фишера. Следущее распределение тоже тесно связано с нормальным распределением, но понадобится нам не при построении доверительных интервалов, а чуть позже — в задачах проверки гипотез. Там же мы поймём, почему его называют *распределением дисперсионного отношения*.

Определение 20. Пусть χ_k^2 имеет распределение H_k, а ψ_n^2 — распределение H_n, причём эти случайные величины независимы. Распределение случайной величины

$$f_{k,n} = \frac{\chi_k^2 \,/\, k}{\psi_n^2 \,/\, n} = \frac{n}{k} \cdot \frac{\chi_k^2}{\psi_n^2}$$

называется *распределением Фишера* с k и n степенями свободы и обозначается $\mathrm{F}_{k,n}$.

Свойства распределения Фишера (или *Фишера — Снедекора*):

Свойство 9. Если случайная величина $f_{k,n}$ имеет распределение Фишера $\mathrm{F}_{k,n}$, то $1/f_{k,n}$ имеет распределение Фишера $\mathrm{F}_{n,k}$.

Заметим, что распределения $\mathrm{F}_{k,n}$ и $\mathrm{F}_{n,k}$ различаются, но связаны соотношением: для любого $x > 0$

$$\mathrm{F}_{k,n}(x) = \mathsf{P}(f_{k,n} < x) = \mathsf{P}\left(\frac{1}{f_{k,n}} > \frac{1}{x}\right) = 1 - \mathrm{F}_{n,k}\left(\frac{1}{x}\right).$$

Распределение Фишера также табулировано при многих k, n, причём свойство 9 позволяет приводить таблицы распределений только в половине случаев: например, при $k \geqslant n$.

Свойство 10. Распределение Фишера $\mathrm{F}_{k,n}$ слабо сходится к вырожденному в точке $c = 1$ распределению при любом стремлении k и n к бесконечности.

Доказательство. Пусть $\xi_1, \xi_2, \ldots$ и $\eta_1, \eta_2, \ldots$ — две независимые последовательности, составленные из независимых случайных величин со стандартным нормальным распределением. Требуемое утверждение вытекает из того, что любая последовательность случайных величин $f_{k,n}$, распределение которой совпадает с распределением отношения двух средних арифметических

$$\frac{\xi_1^2 + \ldots + \xi_k^2}{k} \quad \text{и} \quad \frac{\eta_1^2 + \ldots + \eta_n^2}{n},$$

сходится к единице по вероятности при $k \to \infty$, $n \to \infty$ по ЗБЧ. □

Свойство 11. Пусть $t_k \Subset \mathrm{T}_k$ — случайная величина, имеющая распределение Стьюдента. Тогда $t_k^2 \Subset \mathrm{F}_{1,k}$.

§ 2. Преобразования нормальных выборок

Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка из $\mathrm{N}_{0,1}$, т. е. набор независимых случайных величин со стандартным нормальным распределением. Там, где нам понадобятся операции матричного умножения, будем считать $\vec{X}$ вектором-столбцом. Пусть C — ортогональная матрица $(n \times n)$, т. е.

$$CC^T = E = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix},$$

и $\vec{Y} = C\vec{X}$ — вектор с координатами $Y_i = C_{i1}X_1 + \ldots + C_{in}X_n$.

Координаты вектора $\vec{Y}$ имеют нормальные распределения как линейные комбинации независимых нормальных величин. Какие именно нормальные и с каким совместным распределением? Чтобы ответить на этот вопрос, выясним, как изменится плотность распределения вектора после умножения его на произвольную невырожденную матрицу.

Вспомним, как найти плотность распределения случайной величины $\eta = a\xi + b$ по плотности распределения ξ:

$$f_\eta(y) = |a|^{-1} \cdot f_\xi\left(a^{-1}(y-b)\right).$$

Сформулируем аналогичное утверждение в многомерном случае.

Теорема 18. *Пусть случайный вектор $\vec{X}$ имеет плотность распределения $f_{\vec{X}}(y_1, \ldots, y_n) = f_{\vec{X}}(\vec{y})$ и A — невырожденная матрица. Тогда вектор $\vec{Y} = A\vec{X} + \vec{b}$ имеет плотность распределения*

$$f_{\vec{Y}}(\vec{y}) = f_{A\vec{X}+\vec{b}}(\vec{y}) = |\det A|^{-1} \cdot f_{\vec{X}}\big(A^{-1}(\vec{y} - \vec{b})\big). \tag{18}$$

Доказательство. Если найдётся функция $h(\vec{y}) \geqslant 0$ такая, что для любого борелевского множества $B \subseteq \mathbb{R}^n$

$$\mathsf{P}\big(\vec{Y} \in B\big) = \int\!\!\int_B \!\ldots\! \int h(\vec{y})\, d\vec{y},$$

то функция $h(\vec{y})$ является плотностью распределения вектора $\vec{Y}$.

Вычислим

$$\mathsf{P}\big(\vec{Y} \in B\big) = \mathsf{P}\big(A\vec{X} + \vec{b} \in B\big) = \mathsf{P}\big(\vec{X} \in A^{-1}(B - \vec{b})\big) = \int\!\!\int_{A^{-1}(B-\vec{b})} \!\ldots\! \int f_{\vec{X}}(\vec{x})\, d\vec{x},$$

где $A^{-1}(B - \vec{b}) = \{\vec{x} = A^{-1}(\vec{y} - \vec{b}) \mid \vec{y} \in B\}$. Сделаем замену переменных $\vec{y} = A\vec{x} + \vec{b}$. При такой замене область интегрирования по множеству

$\vec{x} \in A^{-1}(B - \vec{b})$ превратится в область интегрирования по $\vec{y} \in B$, а дифференциал заменится на $d\vec{x} = |J|\,d\vec{y}$, где J — якобиан обратной замены $\vec{x} = A^{-1}(\vec{y} - \vec{b})$, т. е. определитель матрицы A^{-1}. Итак,

$$\mathsf{P}(\vec{Y} \in B) = \underset{B}{\iint \ldots \int} |\det A|^{-1} \cdot f_{\vec{X}}\bigl(A^{-1}(\vec{y} - \vec{b})\bigr)\,d\vec{y}. \quad \square$$

Докажем самое удивительное свойство нормального распределения.

Теорема 19. *Пусть вектор $\vec{X}$ состоит из независимых случайных величин со стандартным нормальным распределением, C — ортогональная матрица, $\vec{Y} = C\vec{X}$. Тогда и координаты вектора $\vec{Y}$ независимы и имеют стандартное нормальное распределение.*

Доказательство. Запишем плотность совместного распределения координат вектора $\vec{X}$. В силу независимости это есть произведение плотностей координат вектора (то же самое, что функция правдоподобия)

$$f_{\vec{X}}(\vec{y}) = \prod_{i=1}^{n} f_{X_i}(y_i) = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\,(y_1^2 + \ldots + y_n^2)} = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\|\vec{y}\|^2}.$$

Здесь для произвольного вектора $\vec{y}$ квадрат нормы $\|\vec{y}\|^2$ есть

$$\|\vec{y}\|^2 = \sum_{i=1}^{n} y_i^2 = \vec{y}^{\,T} \cdot \vec{y}.$$

Пользуясь формулой (18), вычислим плотность распределения вектора $\vec{Y} = C\vec{X}$. Матрица C ортогональна, поэтому $C^{-1} = C^T$ и $\det C = 1$. Получим

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}\left(C^T \cdot \vec{y}\right) = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\|C^T \cdot \vec{y}\,\|^2}.$$

Но умножение на ортогональную матрицу не меняет норму вектора:

$$\|C^T \cdot \vec{y}\,\|^2 = (C^T\vec{y})^T \cdot (C^T\vec{y}) = \vec{y}^{\,T} C\, C^T \vec{y} = \vec{y}^{\,T} \cdot E \cdot \vec{y} = \|\vec{y}\|^2. \qquad (19)$$

Окончательно имеем

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\|\vec{y}\|^2} = f_{\vec{X}}(\vec{y}) = \frac{1}{(2\pi)^{n/2}}\, e^{-\frac{1}{2}\,(y_1^2 + \ldots + y_n^2)}.$$

Итак, вектор $\vec{Y}$ распределен так же, как и вектор $\vec{X}$, т. е. состоит из независимых случайных величин со стандартным нормальным распределением. □

Упражнение. Пусть ξ и η независимы и имеют стандартное нормальное распределение. Зависимы ли случайные величины $\frac{1}{\sqrt{2}}(\xi-\eta)$ и $\frac{1}{\sqrt{2}}(\xi+\eta)$? Какое распределение имеют? Зависимы ли $\xi-\eta$ и $\xi+\eta$? Является ли ортогональной матрица

$$C=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}?$$

Следующее утверждение носит вспомогательный характер и по традиции называется леммой. Эта лемма является, пожалуй, самым главным вспомогательным утверждением во всех разделах теоретической статистики и эконометрики, связанных с нормальными наблюдениями.

Лемма 7 (лемма Фишера). *Пусть вектор* $\vec{X}$ *состоит из независимых случайных величин со стандартным нормальным распределением,* C *— ортогональная матрица,* $\vec{Y}=C\vec{X}$.

Тогда при любом $k=1,\ldots,n-1$ *случайная величина*

$$T(\vec{X})=\sum_{i=1}^{n}X_i^2-Y_1^2-\ldots-Y_k^2$$

не зависит от $Y_1,\ldots,Y_k$ *и имеет распределение* H_{n-k}.

Доказательство. Как мы видели в (19), нормы векторов $\vec{X}$ и $\vec{Y}=C\vec{X}$ совпадают: $X_1^2+\ldots+X_n^2=\|\vec{X}\|^2=\|C\vec{X}\|^2=Y_1^2+\ldots+Y_n^2$. Поэтому

$$T(\vec{X})=\sum_{i=1}^{n}Y_i^2-Y_1^2-\ldots-Y_k^2=Y_{k+1}^2+\ldots+Y_n^2.$$

Случайные величины $Y_1,\ \ldots,\ Y_n$ по теореме 19 независимы и имеют стандартное нормальное распределение, поэтому $T(\vec{X})=Y_{k+1}^2+\ldots+Y_n^2$ имеет распределение H_{n-k} и не зависит от $Y_1,\ldots,Y_k$. □

Второй и третий пункты следующего утверждения выглядят неправдоподобно, особенно если вспомнить обозначения:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S_0^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2.$$

Действительно, обе эти величины являются функциями от одних и тех же наблюдений. Более того, в определение S_0^2 явным образом входит $\overline{X}$.

Теорема 20 (основное следствие леммы Фишера). *Пусть* $X_1, \ldots, X_n$ *независимы и имеют нормальное распределение с параметрами* a *и* σ^2. *Тогда:*

1) $\sqrt{n}\,\dfrac{\overline{X}-a}{\sigma} \in \mathrm{N}_{0,1}$,

2) $\dfrac{(n-1)\,S_0^2}{\sigma^2} = \sum\limits_{i=1}^{n} \dfrac{\left(X_i-\overline{X}\right)^2}{\sigma^2} \in \mathrm{H}_{n-1}$,

3) *случайные величины* $\overline{X}$ *и* S_0^2 *независимы.*

Доказательство. Первое утверждение теоремы очевидно (*доказать, что очевидно*!). Докажем второе и третье. Убедимся сначала, что можно рассматривать выборку из стандартного нормального распределения вместо N_{a,σ^2}:

$$\frac{(n-1)S_0^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-a}{\sigma} - \frac{\overline{X}-a}{\sigma}\right)^2 = \sum_{i=1}^{n}(z_i-\overline{z})^2,$$

где $\overline{z} = \dfrac{\overline{X}-a}{\sigma}$ — среднее арифметическое величин $z_i = \dfrac{X_i-a}{\sigma} \in \mathrm{N}_{0,1}$. Итак, можно с самого начала считать, что X_i имеют стандартное нормальное распределение, $a=0$, $\sigma^2=1$.

Применим лемму Фишера. Представим величину $(n-1)S_0^2$ в виде

$$T(\vec{X}) = (n-1)S_0^2 = \sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - n(\overline{X})^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2.$$

Здесь через Y_1 мы обозначили

$$Y_1 = \sqrt{n}\,\overline{X} = \frac{X_1}{\sqrt{n}} + \ldots + \frac{X_n}{\sqrt{n}}.$$

Чтобы применить лемму Фишера, нужно найти ортогональную матрицу C такую, что Y_1 будет первой координатой вектора $\vec{Y} = C\vec{X}$.

Возьмём матрицу C с первой строкой $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$. Так как длина (норма) этого вектора равна единице, его можно дополнить до ортонормального базиса в $\mathbb{R}^n$. Иначе говоря, этот столбец можно дополнить до ортогональной матрицы. Тогда величина $Y_1 = \sqrt{n}\,\overline{X}$ и будет первой координатой вектора $\vec{Y} = C\vec{X}$. Осталось применить лемму Фишера и получить второе утверждение теоремы.

Из леммы Фишера следует также, что $(n-1)S_0^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2$ не зависит от $Y_1 = \sqrt{n}\,\overline{X}$, т. е. $\overline{X}$ и S_0^2 независимы. □

Вопреки определению χ^2-распределения с $n-1$ степенью свободы, величина $(n-1)S_0^2/\sigma^2$ есть сумма не $n-1$, а n слагаемых, причём эти слагаемые *зависимы* из-за присутствия в каждом $\overline{X}$. К тому же они хоть и одинаково распределены (*почему*?), но их распределение вовсе не является стандартным нормальным (*а какое оно*?).

Отметим без доказательства, что независимость величин $\overline{X}$ и S_0^2 — свойство, характерное *только* для нормального распределения. Так же, как и способность сохранять независимость координат после умножения на ортогональную матрицу.

Очередное следствие из леммы Фишера наконец позволит нам строить доверительные интервалы для параметров нормального распределения, ради чего мы и доказали уже так много утверждений. В каждом пункте указано, для какого параметра мы построим доверительный интервал с помощью данного утверждения.

Теорема 21 (полезное следствие леммы Фишера). *Пусть $X_1, \ldots, X_n$ независимы и имеют нормальное распределение с параметрами a и σ^2. Тогда*

1) $\sqrt{n}\,\dfrac{\overline{X}-a}{\sigma} \in \mathrm{N}_{0,1}$ (*для a при σ известном*),

2) $\sum\limits_{i=1}^{n}\left(\dfrac{X_i-a}{\sigma}\right)^2 \in \mathrm{H}_n$ (*для σ^2 при a известном*),

3) $\dfrac{(n-1)\,S_0^2}{\sigma^2} \in \mathrm{H}_{n-1}$ (*для σ^2 при a неизвестном*),

4) $\sqrt{n}\,\dfrac{\overline{X}-a}{S_0} \in \mathrm{T}_{n-1}$ (*для a при σ неизвестном*).

Доказательство. Утверждения (1) и (3) следуют из леммы Фишера, (2) — из теоремы 5. Осталось воспользоваться леммой Фишера и определением распределения Стьюдента, чтобы доказать (4). Запишем

$$\sqrt{n}\,\frac{\overline{X}-a}{S_0} = \sqrt{n}\,\frac{\overline{X}-a}{\sigma}\cdot\frac{1}{\sqrt{\dfrac{(n-1)S_0^2}{\sigma^2}\cdot\dfrac{1}{n-1}}} = \frac{\xi_0}{\sqrt{\dfrac{\chi_{n-1}^2}{n-1}}}, \tag{20}$$

где величины

$$\xi_0 = \sqrt{n}\,\frac{\overline{X}-a}{\sigma} \in \mathrm{N}_{0,1} \quad \text{и} \quad \chi_{n-1}^2 = (n-1)S_0^2/\sigma^2 \in \mathrm{H}_{n-1}$$

независимы по теореме 5. По определению 19, величина (20) имеет распределение Стьюдента T_{n-1}. □

§ 3. Точные доверительные интервалы для параметров нормального распределения

Пусть $X_1, \ldots, X_n$ — выборка объёма n из распределения N_{a,σ^2}. Построим точные доверительные интервалы (ДИ) с уровнем доверия $1-\varepsilon$ для параметров нормального распределения, используя соответствующие утверждения теоремы 21.

Пример 30 (ДИ для a при известном σ^2). Этот интервал мы построили в примере 26 (с. 60):

$$\mathsf{P}\left(\overline{X} - \frac{\tau\sigma}{\sqrt{n}} < a < \overline{X} + \frac{\tau\sigma}{\sqrt{n}}\right) = 1-\varepsilon, \quad \text{где} \quad \Phi_{0,1}(\tau) = 1-\varepsilon/2.$$

Пример 31 (ДИ для σ^2 при известном a). По теореме 21

$$\frac{nS_1^2}{\sigma^2} \Subset \mathrm{H}_n, \quad \text{где} \quad S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2.$$

Пусть g_1 и g_2 — квантили распределения H_n уровней $\varepsilon/2$ и $1-\varepsilon/2$ соответственно. Тогда

$$1-\varepsilon = \mathsf{P}\left(g_1 < \frac{nS_1^2}{\sigma^2} < g_2\right) = \mathsf{P}\left(\frac{nS_1^2}{g_2} < \sigma^2 < \frac{nS_1^2}{g_1}\right).$$

Пример 32 (ДИ для σ^2 при неизвестном a). По теореме 21

$$\frac{(n-1)S_0^2}{\sigma^2} \Subset \mathrm{H}_{n-1}, \quad \text{где} \quad S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

Пусть g_1 и g_2 — квантили распределения H_{n-1} уровней $\varepsilon/2$ и $1-\varepsilon/2$ соответственно. Тогда

$$1-\varepsilon = \mathsf{P}\left(g_1 < \frac{(n-1)S_0^2}{\sigma^2} < g_2\right) = \mathsf{P}\left(\frac{(n-1)S_0^2}{g_2} < \sigma^2 < \frac{(n-1)S_0^2}{g_1}\right).$$

Упражнение. Найти 17 отличий примера 31 от примера 32.

Пример 33 (ДИ для a при неизвестном σ). По теореме 21

$$\sqrt{n}\,\frac{\overline{X}-a}{S_0} \Subset \mathrm{T}_{n-1}.$$

Пусть c — квантиль распределения T_{n-1} уровня $1-\varepsilon/2$. Распределение Стьюдента симметрично. Поэтому

$$1-\varepsilon = \mathsf{P}\left(-c < \sqrt{n}\,\frac{\overline{X}-a}{S_0} < c\right) = \mathsf{P}\left(\overline{X} - \frac{c\,S_0}{\sqrt{n}} < a < \overline{X} + \frac{c\,S_0}{\sqrt{n}}\right).$$

Упражнение. Сравнить примеры 30 и 33.

Замечание 14. Доверительные интервалы, полученные в примерах 31 и 32, выглядят странно по сравнению с доверительными интервалами из примеров 30 и 33: они содержат n в числителе, а не в знаменателе. Но если квантили нормального распределения от n не зависят вовсе, квантили распределения Стьюдента асимптотически не зависят от n по свойству $\mathrm{T}_n \Rightarrow \mathrm{N}_{0,1}$, то квантили распределения H_n зависят от n существенно.

Действительно, пусть g_n таковы, что $\mathsf{P}(\chi^2_n < g_n) = \delta$ при всех n, и пусть τ_δ — квантиль уровня δ стандартного нормального распределения. Тогда по свойству 3 (с. 69)

$$\mathsf{P}(\chi^2_n < g_n) = \mathsf{P}\left(\frac{\chi^2_n - n}{\sqrt{2n}} < \frac{g_n - n}{\sqrt{2n}}\right) \to \Phi_{0,1}(\tau_\delta) = \delta.$$

Поэтому $\dfrac{g_n - n}{\sqrt{2n}} \to \tau_\delta$ при $n \to \infty$ и, следовательно,

$$g_n = n + \tau_\delta\sqrt{2n} + o(\sqrt{n}).$$

Упражнение. Подставить в границы доверительных интервалов из п. 2—3 асимптотические выражения для квантилей и выяснить, как ведёт себя длина этих интервалов с ростом n.

§ 4. Вопросы и упражнения

1. Величины ξ_1 и ξ_2 независимы и имеют нормальное распределение с параметрами $a = 0$, $\sigma^2 = 16$. Найти k, при котором величины $\xi_1 - 3\xi_2$ и $k\xi_1 + \xi_2$ независимы. Можно использовать теорему 19 (с. 75).

2. Как, пользуясь таблицей *стандартного нормального* распределения, найти квантиль заданного уровня для χ^2-распределения с *одной* степенью свободы?

3. Изобразить квантили уровней $\varepsilon/2$ и $1-\varepsilon/2$ на графиках плотностей распределений H_n и T_{n-1}.

4. Вычислить, зная распределение $(n-1)S_0^2/\sigma^2$ и пользуясь известным математическим ожиданием и дисперсией распределения χ^2, математическое ожидание и дисперсию длины доверительного интервала для дисперсии нормального распределения при неизвестном среднем.

5. Вычислить математическое ожидание и дисперсию длины доверительного интервала для среднего нормального распределения при неизвестной дисперсии.

ГЛАВА VII

ПРОВЕРКА ГИПОТЕЗ

Имея выборку, мы можем выдвинуть несколько взаимоисключающих гипотез о теоретическом распределении, одну из которых следует предпочесть остальным. Задача выбора одной из нескольких гипотез решается построением статистического критерия. Как правило, по выборке конечного объёма безошибочных выводов о распределении сделано быть не может, поэтому всегда есть опасность выбрать неверную гипотезу. Так, бросая монету, можно выдвигать предположения об истинной вероятности выпадения герба. Допустим, есть две гипотезы: вероятность либо находится в пределах 0,45—0,55, либо нет. Получив после ста бросков ровно 51 герб, мы наверняка выберем первую гипотезу. Однако есть ненулевые шансы на то, что и при $p = 0{,}3$ выпадет 51 герб: выбирая первую гипотезу, мы можем ошибиться. Напротив, получив 33 герба, мы скорее всего предпочтём вторую гипотезу. И опять не исключена возможность, что столь далёкое от половины число гербов есть просто результат случайности, а монета на самом деле симметрична.

§ 1. Гипотезы и критерии

Пусть дана выборка $X_1, \ldots, X_n$ из распределения $\mathcal{F}$. Мы будем считать выборку набором независимых случайных величин с одним и тем же распределением, хотя в ряде задач и эти предположения нуждаются в проверке. Тогда одинаковая распределённость или независимость наблюдений не предполагается.

Определение 21. *Гипотезой* (H) называется любое предположение о распределении наблюдений:

$$H = \{\mathcal{F} = \mathcal{F}_1\} \qquad \text{или} \qquad H = \{\, \mathcal{F} \in \mathbb{F}\,\},$$

где $\mathbb{F}$ — некоторое подмножество в множестве всех распределений. Гипотеза H называется *простой*, если она указывает на единственное распределение: $\mathcal{F} = \mathcal{F}_1$. Иначе H называется *сложной*: $\mathcal{F} \in \mathbb{F}$.

Если гипотез всего две, то одну из них принято называть *основной*, а другую — *альтернативой*, или отклонением от основной гипотезы.

Пример 34. Перечислим типичные задачи проверки гипотез.

1. Выбор из нескольких простых гипотез: есть $H_1 = \{\mathcal{F} = \mathcal{F}_1\}, \ldots, H_k = \{\mathcal{F} = \mathcal{F}_k\}$, и другие предположения невозможны.

2. Простая основная гипотеза и сложная альтернатива:

$$H_1 = \{\mathcal{F} = \mathcal{F}_1\}, \quad H_2 = \{\mathcal{F} \neq \mathcal{F}_1\}.$$

Например, дана выборка из семейства распределений B_p, где $p \leqslant 1/2$. Есть простая гипотеза $H_1 = \{p = 1/2\}$ и сложная односторонняя альтернатива $H_2 = \{p < 1/2\}$. Случай $p > 1/2$ исключен априори.

3. Сложная основная гипотеза и сложная альтернатива:

$$H_1 = \{\mathcal{F} \in \mathbb{F}\}, \quad H_2 = \{\mathcal{F} \notin \mathbb{F}\}.$$

Например, гипотеза о нормальности: $H_1 = \{$распределение $\mathcal{F}$ является нормальным$\}$ при альтернативе $H_2 = \{H_1$ неверна$\}$.

4. Гипотеза однородности: есть несколько выборок; основная гипотеза состоит в том, что эти выборки извлечены из одного распределения.

5. Гипотеза независимости: по выборке $(X_1, Y_1), \ldots, (X_n, Y_n)$ из n независимых наблюдений пары случайных величин проверяется гипотеза $H_1 = \{X_i$ и Y_i независимы$\}$ при альтернативе $H_2 = \{H_1$ неверна$\}$. Обе гипотезы являются сложными.

6. Гипотеза случайности. В эксперименте наблюдаются n случайных величин $X_1, \ldots, X_n$ и проверяется сложная гипотеза $H_1 = \{X_1, \ldots, X_n$ независимы и одинаково распределены$\}$.

Эту задачу ставят, например, при проверке качества генератора случайных чисел.

Пусть дана выборка $X_1, \ldots, X_n$, относительно распределения которой выдвинуты гипотезы $H_1, \ldots, H_k$.

Определение 22. *Критерием* $\delta = \delta(X_1, \ldots, X_n)$ называется измеримое отображение

$$\delta : \mathbb{R}^n \to \{H_1, \ldots, H_k\}$$

из множества всех возможных значений выборки в множество гипотез. Измеримость понимается в обычном смысле: $\{\omega \mid \delta(X_1, \ldots, X_n) = H_i\}$ есть событие при любом $i = 1, \ldots, k$.

Определение 23. Говорят, что произошла *ошибка i-го рода* критерия δ, если критерий отверг *верную* гипотезу H_i. *Вероятностью ошибки i-го рода* критерия δ называется число

$$\alpha_i(\delta) = \mathsf{P}_{H_i}(\delta(\vec{X}) \neq H_i).$$

Замечание 15. Говоря «H_i верна» и вычисляя $\mathsf{P}_{H_i}(\cdot)$, мы имеем в виду, что распределение выборки именно такое, как предполагает гипотеза H_i, и вычисляем вероятность в соответствии с этим распределением. Если гипотеза H_i простая, т. е. указывает ровно на одно возможное распределение выборки, то $\alpha_i(\delta)$ — число. Если же H_i — сложная гипотеза, то $\alpha_i(\delta)$ будет зависеть от того, при каком именно из распределений $\mathcal{F}$, отвечающих H_i, вычисляется вероятность:

$$\alpha_i(\delta) = \alpha_i(\delta, \mathcal{F}) = \mathsf{P}_{\mathcal{F}}\big(\delta(\vec{X}) \neq H_i\big) = \mathsf{P}\big(\delta(\vec{X}) \neq H_i \,|\, X_i \in \mathcal{F}\big).$$

Пример 35. Пусть любое изделие некоторого производства оказывается браком с вероятностью p. Контроль продукции допускает ошибки: годное изделие бракует с вероятностью γ, а бракованное пропускает (признаёт годным) с вероятностью ε.

Если ввести для проверяемого изделия гипотезы $H_1 = \{$изделие годное$\}$ и $H_2 = \{$изделие бракованное$\}$, а критерием выбора одной из них считать контроль продукции, то γ — вероятность ошибки первого рода этого критерия, а ε — второго рода:

$$\begin{aligned} \gamma &= \mathsf{P}_{H_1}(\delta = H_2) = \mathsf{P}(\text{контроль забраковал годное изделие}); \\ \varepsilon &= \mathsf{P}_{H_2}(\delta = H_1) = \mathsf{P}(\text{контроль пропустил бракованное изделие}); \end{aligned}$$

Упражнение. Вычислить вероятности ошибок первого и второго рода того же критерия, если гипотезы занумеровать иначе:

$$H_1 = \big\{\text{изделие бракованное}\big\}, \quad H_2 = \big\{\text{изделие годное}\big\}.$$

Надеемся, что читатель на основании своего опыта и воображения сделал для себя следующие выводы.

1. Статистический критерий не отвечает на вопрос, верна или нет проверяемая гипотеза. Он лишь решает, противоречат или не противоречат выдвинутой гипотезе выборочные данные, можно ли принять или следует отвергнуть данную гипотезу.

2. Вывод «данные противоречат гипотезе» всегда весомее и категоричнее, нежели вывод «данные не противоречат гипотезе».

3. Нам неизвестно, какая из гипотез верна в действительности, поэтому следует считаться с гипотетическими вероятностями ошибок критерия. Смысл этих ошибок в следующем: если много раз применять критерий к выборкам из распределения, для которого гипотеза H_i верна, то в среднем доля α_i таких выборок будет признана противоречащей гипотезе H_i.

§ 2. Подходы к сравнению критериев

Рассмотрим подробно случай, когда имеются две простые гипотезы о распределении наблюдений

$$H_1 = \{\mathcal{F} = \mathcal{F}_1\} \quad \text{и} \quad H_2 = \{\mathcal{F} = \mathcal{F}_2\}.$$

Тогда любой критерий $\delta(\vec{X})$ принимает не более двух значений. Это означает, что область $\mathbb{R}^n$ делится на две части $\mathbb{R}^n = S \cup (\mathbb{R}^n \backslash S)$ так, что

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \vec{X} \in \mathbb{R}^n \backslash S, \\ H_2, & \text{если } \vec{X} \in S. \end{cases}$$

Область S, в которой принимается вторая (альтернативная) гипотеза, называется *критической областью.*

Определение 24. Вероятность ошибки первого рода $\alpha_1 = \alpha_1(\delta)$ иначе называют *размером* или *критическим уровнем* критерия δ:

$$\alpha_1 = \alpha_1(\delta) = \mathsf{P}_{H_1}(\delta(\vec{X}) \neq H_1) = \mathsf{P}_{H_1}(\delta(\vec{X}) = H_2) = \mathsf{P}_{H_1}(\vec{X} \in S).$$

Мощностью критерия δ называют величину $1 - \alpha_2$, где $\alpha_2 = \alpha_2(\delta)$ — вероятность ошибки второго рода критерия δ. Мощность критерия равна

$$1 - \alpha_2(\delta) = 1 - \mathsf{P}_{H_2}(\delta(\vec{X}) \neq H_2) = \mathsf{P}_{H_2}(\delta(\vec{X}) = H_2) = \mathsf{P}_{H_2}(\vec{X} \in S).$$

Заметим, что вероятности ошибок первого и второго рода вычисляются при *разных* предположениях о распределении (верна H_1 либо верна H_2), поэтому никакими фиксированными соотношениями вида $\alpha_1 \equiv 1 - \alpha_2$ эти ошибки не связаны.

Как сравнивать критерии? Разумеется, критерий тем лучше, чем меньше вероятности его ошибок. Но если сравнивать критерии по двум вероятностям ошибок одновременно, чтобы

$$\alpha_i(\delta_1) \leqslant \alpha_i(\delta_2) \quad \text{при} \quad i = 1,\, 2,$$

то слишком многие критерии окажутся несравнимыми. Например, рассмотрим два крайних случая, когда критерий, независимо от выборки, всегда принимает одну и ту же гипотезу.

Пример 36. Пусть критерий $\delta(\vec{X}) \equiv H_1$ всегда выбирает первую гипотезу. Тогда $\alpha_1 = \mathsf{P}_{H_1}(\delta = H_2) = 0$, $\alpha_2 = \mathsf{P}_{H_2}(\delta = H_1) = 1$.

Наоборот: пусть критерий $\delta(\vec{X}) \equiv H_2$ всегда выбирает вторую гипотезу. Тогда $\alpha_1 = \mathsf{P}_{H_1}(\delta = H_2) = 1$, $\alpha_2 = \mathsf{P}_{H_2}(\delta = H_1) = 0$.

Пример 37. Имеется выборка объёма $n = 1$ из нормального распределения $\mathrm{N}_{a,1}$ и две простые гипотезы $H_1 = \{a = 0\}$ и $H_2 = \{a = 1\}$. Рассмотрим при некотором $b \in \mathbb{R}$ следующий критерий:

$$\delta(X_1) = \begin{cases} H_1, & \text{если } X_1 \leqslant b, \\ H_2, & \text{если } X_1 > b. \end{cases}$$

Изобразим на графике (рис. 10) соответствующие гипотезам плотности распределений и вероятности ошибок первого и второго рода критерия δ

$$\alpha_1 = \mathsf{P}_{H_1}(X_1 > b), \qquad \alpha_2 = \mathsf{P}_{H_2}(X_1 \leqslant b).$$

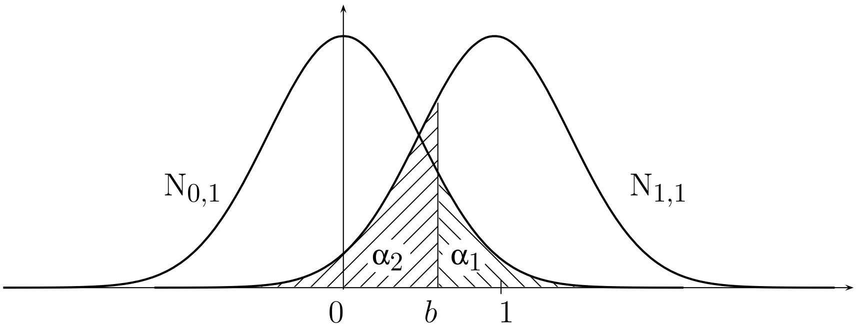

Рис. 10. Две простые гипотезы

Видим, что с ростом числа b вероятность ошибки первого рода α_1 уменьшается, но вероятность ошибки второго рода α_2 растёт.

Итак, примеры 36 и 37 показывают общую тенденцию: при попытке уменьшить одну из вероятностей ошибок другая, *как правило*, увеличивается. Так, если уменьшать $\alpha_1 = \mathsf{P}_{H_1}(\vec{X} \in S)$ за счёт сужения критической области S, то одновременно будет расти вероятность ошибки второго рода и уменьшаться мощность критерия $1 - \alpha_2 = \mathsf{P}_{H_2}(\vec{X} \in S)$.

Перечислим общепринятые подходы к сравнению критериев. Ограничимся для простоты задачей проверки двух простых гипотез. Пусть имеются критерии δ и ρ с вероятностями ошибок первого и второго рода $\alpha_1(\delta)$, $\alpha_2(\delta)$ и $\alpha_1(\rho)$, $\alpha_2(\rho)$.

Минимаксный подход. Говорят, что критерий δ не хуже критерия ρ в смысле минимаксного подхода, если

$$\max\{\alpha_1(\delta), \alpha_2(\delta)\} \leqslant \max\{\alpha_1(\rho), \alpha_2(\rho)\}.$$

Определение 25. Критерий δ называется *минимаксным*, если он не хуже всех других критериев в смысле минимаксного подхода.

Иначе говоря, минимаксный критерий имеет самую маленькую «наибольшую ошибку» $\max\{\alpha_1(\delta), \alpha_2(\delta)\}$ среди всех прочих критериев.

Упражнение. Убедиться, что в примере 37 критерий δ при $b = 1/2$ является минимаксным.

Байесовский подход. Этот подход применяют в следующих случаях: а) если известно априори, что с вероятностью r верна гипотеза H_1, а с вероятностью $s = 1 - r$ — гипотеза H_2; б) если задана линейная «функция потерь»: потери от ошибочного решения равны r, если происходит ошибка первого рода, и равны s, если второго. Здесь $r + s$ уже не обязательно равно 1, но потери можно свести к единице нормировкой $r' = r/(r+s)$ и $s' = s/(r+s)$.

Пусть априорные вероятности или потери r и s заданы. Говорят, что критерий δ не хуже критерия ρ в смысле байесовского подхода, если

$$r\alpha_1(\delta) + s\alpha_2(\delta) \leqslant r\alpha_1(\rho) + s\alpha_2(\rho).$$

Определение 26. Критерий δ называют *байесовским*, если он не хуже всех других критериев в смысле байесовского подхода.

Иначе говоря, байесовский критерий имеет самую маленькую «средневзвешенную ошибку» $r\alpha_1(\delta) + s\alpha_2(\delta)$ среди всех прочих критериев. По формуле полной вероятности это есть вероятность ошибки критерия в случае (а) или математическое ожидание потерь в случае (б).

Упражнение. Убедиться, что в примере 37 критерий δ при $b = 1/2$ является байесовским для $r = s$.

Выбор наиболее мощного критерия. Ошибки первого и второго рода обычно неравноправны. Поэтому возникает желание контролировать одну из ошибок. Например, зафиксировать вероятность ошибки первого рода на достаточно низком (безопасном) уровне и рассматривать только критерии с такой же или ещё меньшей вероятностью этой ошибки. Среди них наилучшим следует признать критерий с наименьшей вероятностью ошибки второго рода.

Введём при $\varepsilon \in [0, 1]$ класс критериев $K_\varepsilon = \{\,\delta(\vec{X}) \mid \alpha_1(\delta) \leqslant \varepsilon\}$.

Определение 27. Критерий $\delta_0 \in K_\varepsilon$ называют *наиболее мощным критерием* (НМК) размера ε, если $\alpha_2(\delta_0) \leqslant \alpha_2(\delta)$ для любого другого критерия $\delta \in K_\varepsilon$.

В следующем параграфе мы рассмотрим способы построения оптимальных критериев. Оказывается, все оптимальные критерии (минимаксные, байесовские, наиболее мощные) могут быть построены простым выбором различных констант в некотором универсальном критерии — *критерии отношения правдоподобия*.

§ 3. Построение оптимальных критериев

Критерий отношения правдоподобия. Выборка $\vec{X} = (X_1, \ldots, X_n)$ состоит из независимых и одинаково распределённых величин, про распределение которых возможны только две гипотезы:

$$H_1 = \{X_i \in \mathcal{F}_1\} \quad \text{и} \quad H_2 = \{X_i \in \mathcal{F}_2\}.$$

Пусть $f_1(y)$ и $f_2(y)$ — плотности распределений $\mathcal{F}_1$ и $\mathcal{F}_2$ соответственно. Термин «плотность» здесь понимается в смысле равенства (6) на с. 29. Построим функции правдоподобия для этих распределений:

$$f_1(\vec{X}) = \prod_{i=1}^{n} f_1(X_i) \quad \text{и} \quad f_2(\vec{X}) = \prod_{i=1}^{n} f_2(X_i).$$

Пусть выполнено предположение (I).

(I) Распределения $\mathcal{F}_1$ и $\mathcal{F}_2$ либо оба дискретны, либо оба абсолютно непрерывны.

Замечание 16. Если одно из распределений дискретно, а другое абсолютно непрерывно, то всегда существует критерий с нулевыми вероятностями ошибок. Смешанные распределения мы рассматривать не будем. Математики вместо (I) могут предполагать, что оба распределения абсолютно непрерывны относительно одной и той же σ-конечной меры и имеют относительно неё плотности $f_1(y)$ и $f_2(y)$.

Мы будем выбирать гипотезу в зависимости от отношения функций правдоподобия. Обратимся к примеру 37. Естественным кажется принимать вторую гипотезу, если X_1 лежит правее точки пересечения плотностей $b = 1/2$: *там, где вторая плотность больше, принимать вторую гипотезу, там, где первая — первую*. Такой критерий сравнивает отношение $f_2(x_1, \ldots, x_n)/f_1(x_1, \ldots, x_n)$ с единицей, относя к критической области ту часть $\mathbb{R}^n$, где это отношение больше единицы. Заметим, что при этом мы получим ровно один, не обязательно оптимальный, критерий с некоторым фиксированным размером и мощностью.

Если нужно получить критерий с заранее заданным размером $\alpha_1 = \varepsilon$, либо иметь возможность варьировать и размер, и мощность критерия, то следует рассмотреть класс похожим образом устроенных критериев, введя свободный параметр: *там, где вторая плотность в c раз превосходит первую, выбирать вторую гипотезу, иначе — первую*: сравнивать отношение плотностей $f_2(x_1, \ldots, x_n)/f_1(x_1, \ldots, x_n)$ не с единицей, а с некоторой постоянной c.

Назовём *отношением правдоподобия* частное

$$T(\vec{x}) = T(x_1, \ldots, x_n) = \frac{f_2(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)}\,, \qquad (21)$$

рассматривая его лишь при таких значениях $\vec{x}$, когда хотя бы одна из плотностей отлична от нуля. Имеется в виду, что $0/a = 0$, $a/0 = +\infty$.

Конструкция критерия, который мы описали выше, сильно усложнится в случае, когда распределение случайной величины $T(\vec{X})$ не является непрерывным, т. е. существует такое число c, вероятность попасть в которое $\Delta_c = \mathsf{P}_{H_1}\big(f_2(\vec{X})/f_1(\vec{X}) = c\big)$ отлична от нуля. Это означает, что на некотором «большом» множестве значений выборки обе гипотезы «равноправны»: отношение правдоподобия постоянно. Относя это множество целиком к критическому множеству или целиком исключая из него, мы меняем вероятность ошибки первого рода (размер) критерия сразу на положительную величину Δ_c:

$$\mathsf{P}_{H_1}(T(\vec{X}) \geqslant c) = \mathsf{P}_{H_1}(T(\vec{X}) > c) + \mathsf{P}_{H_1}(T(\vec{X}) = c) = \mathsf{P}_{H_1}(T(\vec{X}) > c) + \Delta_c.$$

И если вдруг мы захотим приравнять размер критерия заранее выбранному числу ε, может случиться так, что у критерия с критическим множеством $S = \{T(\vec{x}) \geqslant c\}$ размер превысит ε, а у критерия с критическим множеством $S = \{T(\vec{x}) > c\}$ размер будет меньше, чем ε.

Чтобы избежать этой искусственной проблемы, предположим (II).

(II) Функция $R(c) = \mathsf{P}_{H_1}(T(\vec{X}) \geqslant c)$ непрерывна по c при $c > 0$.

Здесь $R(c)$ есть просто хвост функции распределения случайной величины $T(\vec{X})$, вычисленной при верной первой гипотезе:

$$R(c) = 1 - \mathsf{P}_{H_1}(T(\vec{X}) < c).$$

Её непрерывность означает, что величина $\Delta_c = \mathsf{P}_{H_1}(T(\vec{X}) = c)$ равна нулю для любого $c > 0$.

Определение 28. В условиях предположений (I), (II) критерий

$$\delta_c(\vec{X}) = \begin{cases} H_1, & \text{если } T(\vec{X}) < c, \\ H_2, & \text{если } T(\vec{X}) \geqslant c \end{cases} = \begin{cases} H_1, & \text{если } \dfrac{f_2(X_1, \ldots, X_n)}{f_1(X_1, \ldots, X_n)} < c, \\ H_2, & \text{если } \dfrac{f_2(X_1, \ldots, X_n)}{f_1(X_1, \ldots, X_n)} \geqslant c \end{cases}$$

назовём *критерием отношения правдоподобия* (КОП). Размер и вероятность ошибки второго рода этого критерия равны соответственно

$$\alpha_1(\delta_c) = \mathsf{P}_{H_1}(T(\vec{X}) \geqslant c) = R(c), \qquad \alpha_2(\delta_c) = \mathsf{P}_{H_2}(T(\vec{X}) < c).$$

Явный вид оптимальных критериев. Следующая теорема утверждает, что все оптимальные критерии суть критерии отношения правдоподобия. Третье утверждение теоремы называют *леммой Неймана — Пирсона*.

Теорема 22. *Пусть выполнены предположения* (I) *и* (II). *Тогда критерий отношения правдоподобия является*

1) *минимаксным критерием при* c *таком, что* $\alpha_1(\delta_c) = \alpha_2(\delta_c)$;

2) *байесовским критерием при заданных априорных вероятностях* r *и* s, *если* $c = r/s$;

3) *НМК размера* ε, *где* $0 < \varepsilon \leqslant \mathsf{P}_{H_1}(f_2(\vec{X}) > 0)$, *если* c *выбрано так, что* $\alpha_1(\delta_c) = \varepsilon$.

Упражнение. Прочитать доказательство теоремы 22 в [1, § 2, гл. 3].

Пример 38. Дана выборка $X_1, \ldots, X_n$ из нормального распределения со средним a и единичной дисперсией. Построим минимаксный, байесовский при $r = 1/3$, $s = 2/3$ и наиболее мощный критерии для проверки гипотезы $H_1 = \{a = a_1\}$ против альтернативы $H_2 = \{a = a_2\}$, где $a_1 < a_2$.

Отношение правдоподобия имеет абсолютно непрерывное распределение при любой из гипотез, поэтому условие (II) выполнено. Построим критерий отношения правдоподобия. Достаточно описать его критическую область $S = \{\delta(\vec{X}) = H_2\}$. Она определяется неравенством

$$T(\vec{X}) = \frac{f_2(\vec{X})}{f_1(\vec{X})} = \exp\left\{\frac{1}{2}\sum_{i=1}^{n}(X_i - a_1)^2 - \frac{1}{2}\sum_{i=1}^{n}(X_i - a_2)^2\right\} \geqslant c. \qquad (22)$$

Критерий будет байесовским при $c = r/s = 1/2$. Упростим неравенство (22). Получим

$$\delta(\vec{X}) = H_2 \quad \text{при} \quad \overline{X} \geqslant \frac{a_1 + a_2}{2} - \frac{\ln 2}{n(a_2 - a_1)}.$$

Чтобы построить минимаксный и наиболее мощный критерии, запишем неравенство (22) в эквивалентном виде $\overline{X} \geqslant c_1$, и искать будем c_1, а не c. Размер и вероятность ошибки второго рода равны соответственно

$$\begin{aligned}
\alpha_1(\delta) &= \mathsf{P}_{H_1}\big(\overline{X} \geqslant c_1\big) = \mathsf{P}_{H_1}\big(\sqrt{n}\,(\overline{X}-a_1) \geqslant \sqrt{n}\,(c_1-a_1)\big) = \\
&= 1 - \Phi_{0,1}\big(\sqrt{n}\,(c_1-a_1)\big), \\
\alpha_2(\delta) &= \mathsf{P}_{H_2}\big(\overline{X} < c_1\big) = \mathsf{P}_{H_2}\big(\sqrt{n}\,(\overline{X}-a_2) < \sqrt{n}\,(c_1-a_2)\big) = \\
&= \Phi_{0,1}\big(\sqrt{n}\,(c_1-a_2)\big).
\end{aligned}$$

Равенство $\alpha_1(\delta) = \varepsilon$ означает, что $\sqrt{n}(c_1 - a_1) = \tau_{1-\varepsilon}$, где $\tau_{1-\varepsilon}$ — квантиль уровня $1-\varepsilon$ стандартного нормального распределения. Тогда выразим $c_1 = a_1 + \tau_{1-\varepsilon}/\sqrt{n}$. Получим НМК размера ε

$$\delta(\vec{X}) = H_2 \quad \text{при} \quad \overline{X} \geqslant a_1 + \frac{\tau_{1-\varepsilon}}{\sqrt{n}}.$$

При $\alpha_1(\delta) = \alpha_2(\delta)$ получим минимаксный критерий. Пользуясь свойствами функции распределения стандартного нормального закона, запишем

$$1 - \Phi_{0,1}\left(\sqrt{n}\,(c_1 - a_1)\right) = \Phi_{0,1}\left(\sqrt{n}\,(c_1 - a_2)\right) = 1 - \Phi_{0,1}\left(\sqrt{n}\,(a_2 - c_1)\right),$$

откуда $c_1 - a_1 = a_2 - c_1$ и $c_1 = (a_1 + a_2)/2$. Минимаксный критерий имеет вид

$$\delta(\vec{X}) = H_2 \quad \text{при} \quad \overline{X} \geqslant \frac{a_1 + a_2}{2}.$$

Пример 39. Имеется выборка $X_1, \ldots, X_n$ из нормального распределения со средним $a = 0$ и дисперсией σ^2, $\sigma > 0$. Построим наиболее мощный критерий размера ε для проверки гипотезы $H_1 = \{\sigma = \sigma_1\}$ против альтернативы $H_2 = \{\sigma = \sigma_2\}$, где $\sigma_1 < \sigma_2$.

Отношение правдоподобия снова имеет абсолютно непрерывное распределение при любой из гипотез, поэтому условие (II) выполнено. Критическая область критерия отношения правдоподобия $S = \{\delta(\vec{X}) = H_2\}$ определяется неравенством

$$T(\vec{X}) = \frac{\sigma_1^n}{\sigma_2^n} \exp\left\{\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\sum_{i=1}^{n} X_i^2\right\} \geqslant c,$$

что равносильно неравенству $\overline{X^2} \geqslant c_1$. Найдём c_1, при котором размер критерия равен ε:

$$\alpha_1(\delta) = \mathsf{P}_{H_1}\left(\overline{X^2} \geqslant c_1\right) = \mathsf{P}_{H_1}\left(\frac{n\overline{X^2}}{\sigma_1^2} \geqslant \frac{nc_1}{\sigma_1^2}\right) = 1 - H_n\left(\frac{nc_1}{\sigma_1^2}\right) = \varepsilon.$$

Отсюда $nc_1/\sigma_1^2 = h_{1-\varepsilon}$, где $h_{1-\varepsilon}$ — квантиль χ^2-распределения с n степенями свободы. Тогда $c_1 = h_{1-\varepsilon}\sigma_1^2/n$ и НМК размера ε имеет вид

$$\delta(\vec{X}) = H_2 \quad \text{при} \quad \overline{X^2} \geqslant \frac{h_{1-\varepsilon}\sigma_1^2}{n}.$$

Следующее определение касается асимптотических свойств последовательности критериев, построенных по выборке растущего объёма n в задаче проверки двух простых гипотез.

Определение 29. Критерий $\delta_n = \delta_n(X_1, \ldots, X_n)$ называется критерием *асимптотического размера* ε, если $\alpha_1(\delta_n) \to \varepsilon$ при $n \to \infty$. Критерий $\delta_n = \delta_n(X_1, \ldots, X_n)$ называется *состоятельным*, если $\alpha_2(\delta_n) \to 0$ при $n \to \infty$.

Замечание 17. Отметим снова, что для сложной гипотезы H_i вероятность ошибки i-го рода $\alpha_i(\delta_n) = \alpha_i(\delta_n, \mathcal{F})$ зависит от конкретного распределения $\mathcal{F}$, удовлетворяющего этой гипотезе, по которому и вычисляется вероятность ошибки. Тогда сходимость в определении 29 должна иметь место для каждого такого распределения $\mathcal{F}$.

Пример 40. Является ли состоятельными НМК, построенные нами в двух предыдущих примерах? Проверим состоятельность критерия из примера 38:

$$\delta_n(\vec{X}) = H_2 \quad \text{при} \quad \overline{X} \geqslant a_1 + \frac{\tau_{1-\varepsilon}}{\sqrt{n}}.$$

Вероятность ошибки второго рода этого критерия равна

$$\alpha_2(\delta_n) = \mathsf{P}_{H_2}\Big(\overline{X} < a_1 + \frac{\tau_{1-\varepsilon}}{\sqrt{n}}\Big) = \mathsf{P}_{H_2}\Big(\overline{X} - \frac{\tau_{1-\varepsilon}}{\sqrt{n}} < a_1\Big).$$

При верной гипотезе H_2 по ЗБЧ

$$\xi_n = \overline{X} - \frac{\tau_{1-\varepsilon}}{\sqrt{n}} \xrightarrow{\text{p}} a_2 > a_1.$$

Из сходимости по вероятности следует слабая сходимость, т. е. сходимость функций распределения $F_{\xi_n}(x)$ во всех точках непрерывности предельной функции распределения $F_{a_2}(x)$. Функция $F_{a_2}(x) = \mathsf{P}(a_2 < x)$ непрерывна в точке a_1 (*а где разрывна?*) и равна в этой точке нулю. Поэтому

$$\alpha_2(\delta_n) = \mathsf{P}_{H_2}\Big(\overline{X} - \frac{\tau_{1-\varepsilon}}{\sqrt{n}} < a_1\Big) = F_{\xi_n}(a_1) \to F_{a_2}(a_1) = 0.$$

Проверим состоятельность критерия из примера 39:

$$\delta_n(\vec{X}) = H_2 \quad \text{при} \quad \overline{X^2} \geqslant \frac{h_{1-\varepsilon}\,\sigma_1^2}{n}.$$

Вероятность ошибки второго рода этого критерия равна

$$\alpha_2(\delta_n) = \mathsf{P}_{H_2}\Big(\overline{X^2} < \frac{h_{1-\varepsilon}\,\sigma_1^2}{n}\Big).$$

В замечании 14 (с. 80) мы выяснили, что квантили распределения χ^2 с n степенями свободы с ростом n ведут себя следующим образом:

$$h_{1-\varepsilon} = n + \tau_{1-\varepsilon}\sqrt{2n} + o(\sqrt{n}).$$

Тогда

$$\alpha_2(\delta_n) = \mathsf{P}_{H_2}\Big(\overline{X^2} < \sigma_1^2 + \sigma_1^2\tau_{1-\varepsilon}\frac{\sqrt{2}}{\sqrt{n}} + o\Big(\frac{1}{\sqrt{n}}\Big)\Big).$$

Осталось перенести в левую часть неравенства всё, что зависит от n, и применить ЗБЧ вместе с определением слабой сходимости: при верной гипотезе H_2

$$\overline{X^2} - \sigma_1^2\tau_{1-\varepsilon}\frac{\sqrt{2}}{\sqrt{n}} + o\Big(\frac{1}{\sqrt{n}}\Big) \xrightarrow{\mathrm{p}} \sigma_2^2 > \sigma_1^2.$$

В силу непрерывности предельной функции распределения $F_{\sigma_2^2}(x)$ в точке σ_1^2 имеем

$$\alpha_2(\delta_n) = \mathsf{P}_{H_2}\Big(\overline{X^2} - \sigma_1^2\tau_{1-\varepsilon}\frac{\sqrt{2}}{\sqrt{n}} + o\Big(\frac{1}{\sqrt{n}}\Big) < \sigma_1^2\Big) \to F_{\sigma_2^2}(\sigma_1^2) = 0.$$

§ 4. Вопросы и упражнения

1. Есть две гипотезы: основная состоит в том, что элементы выборки имеют нормальное распределение, а альтернатива — в том, что элементы выборки имеют распределение Пуассона. Построить критерий, обладающий нулевыми вероятностями ошибок первого и второго рода.

2. Говорят, что распределения $\mathcal{F}$ и $\mathcal{G}$ *взаимно сингулярны*, если существует борелевское множество B такое, что $\mathcal{F}(B) = 0$, $\mathcal{G}(B) = 1$. Есть две гипотезы: основная состоит в том, что элементы выборки имеют распределение $\mathcal{F}$, а альтернатива — в том, что элементы выборки имеют распределение $\mathcal{G}$, причём эти распределения взаимно сингулярны. Построить критерий, обладающий нулевыми вероятностями ошибок и первого, и второго рода.

3. Пусть $X_1, \ldots, X_n$ — выборка из биномиального распределения с параметрами m и p, где p может принимать лишь значения $1/3$ и $2/3$ с априорными вероятностями $1/5$ и $4/5$ соответственно, а параметр m известен и фиксирован. Построить байесовский критерий.

4. По выборке из показательного распределения с параметром α построить наиболее мощный критерий асимптотического размера ε, различающий гипотезу $\alpha = \alpha_1$ и альтернативу $\alpha = \alpha_2$, если $\alpha_1 < \alpha_2$. Вычислить предел мощности построенного критерия при $n \to \infty$.

ГЛАВА VIII

КРИТЕРИИ СОГЛАСИЯ

Критериями согласия обычно называют критерии, предназначенные для проверки простой гипотезы $H_1 = \{\mathcal{F} = \mathcal{F}_1\}$ при сложной альтернативе $H_2 = \{H_1 \text{ неверна}\}$. Мы рассмотрим более широкий класс основных гипотез, включающий в том числе и сложные гипотезы, а критериями согласия будем называть любые критерии, устроенные по одному и тому же принципу. А именно, пусть задана некоторая случайная величина, измеряющая отклонение эмпирического распределения от теоретического, распределение которой существенно разнится в зависимости от того, верна или нет основная гипотеза. Критерии согласия принимают или отвергают основную гипотезу исходя из величины этой функции отклонения.

§ 1. Общий вид критериев согласия

Мы опишем конструкцию критерия для случая простой основной гипотезы, а в дальнейшем будем её корректировать по мере изменения задачи.

Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка из распределения $\mathcal{F}$. Проверяется основная гипотеза $H_1 = \{\mathcal{F} = \mathcal{F}_1\}$ при альтернативе $H_2 = \{\mathcal{F} \neq \mathcal{F}_1\}$.

Определение 30. Пусть существует борелевская функция $\rho(\vec{X})$, обладающая следующими свойствами:

(K1) если гипотеза H_1 верна, т. е. если $X_i \in \mathcal{F}_1$, то $\rho(\vec{X}) \Rightarrow \mathcal{G}$, где $\mathcal{G}$ — полностью известное непрерывное распределение;

(K2) если гипотеза H_1 неверна, т. е. если X_i имеют какое-то распределение $\mathcal{F}_2 \neq \mathcal{F}_1$, то $|\rho(\vec{X})| \xrightarrow{\text{p}} \infty$ при $n \to \infty$ для любого такого $\mathcal{F}_2$.

Для случайной величины $\eta \in \mathcal{G}$ определим постоянную C из равенства $\varepsilon = \mathsf{P}(|\eta| \geqslant C)$. Построим критерий

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C, \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C. \end{cases} \tag{23}$$

Этот критерий называется *критерием согласия*.

Критерий согласия «работает» по принципу: если для данной выборки функция отклонения велика по абсолютному значению, то это свидетельствует в пользу альтернативы, и наоборот. При этом степень «великости» определяется исходя из того, как функция отклонения должна себя вести, если бы основная гипотеза была верна. Действительно, если H_1 верна, статистика $\rho(\vec{X})$ имеет почти распределение $\mathcal{G}$. Следовательно, она должна себя вести подобно типичной случайной величине η из этого распределения. Но для той попадание в область $\{|\eta| \geqslant C\}$ маловероятно: вероятность этого события равна малому числу ε. Поэтому попадание величины $\rho(\vec{X})$ в эту область заставляет подозревать, что гипотеза H_1 неверна. Тем более, что больших значений величины $|\rho(\vec{X})|$ следует ожидать именно при альтернативе H_2.

Убедимся в том, что этот критерий имеет (асимптотический) размер ε и является *состоятельным*. Повторим определение состоятельности критерия. Поскольку альтернатива H_2 всегда является сложной, то, как мы уже отмечали, вероятность ошибки второго рода любого критерия δ будет зависеть от конкретного распределения $\mathcal{F}_2$ из числа альтернатив.

Определение 31. Критерий δ для проверки гипотезы H_1 против сложной альтернативы H_2 называется *состоятельным*, если для любого распределения F_2, отвечающего альтернативе H_2, вероятность ошибки второго рода стремится к нулю с ростом объёма выборки:

$$\alpha_2(\delta, \mathcal{F}_2) = \mathsf{P}_{\mathcal{F}_2}\big(\delta(\vec{X}) = H_1\big) \to 0 \text{ при } \quad n \to \infty.$$

Теорема 23. *Критерия согласия* δ, *заданный в определении 30, имеет асимптотический размер* ε *и является состоятельным.*

Доказательство. Условие (K1) отвечает за размер критерия:

$$\alpha_1(\delta) = \mathsf{P}_{H_1}\big(|\rho(\vec{X})| \geqslant C\big) \to \mathsf{P}\left(|\eta| \geqslant C\right) = \varepsilon.$$

Расшифруем условие (K2), отвечающее за состоятельность критерия. По определению, запись $\xi_n \xrightarrow{\text{p}} \infty$ означает, что для любого $C > 0$

$$\mathsf{P}\left(\xi_n < C\right) \to 0 \text{ при } n \to \infty.$$

Согласно этому определению, для любого распределения $\mathcal{F}_2$ из числа альтернатив вероятность ошибки второго рода стремится к нулю:

$$\alpha_2(\delta,\, \mathcal{F}_2) = \mathsf{P}_{\mathcal{F}_2}\big(|\rho(\vec{X})| < C\big) \to 0. \qquad \square$$

Замечание 18. Если вместо слабой сходимости в (K1) выполняется $\rho(\vec{X}) \in G$, то критерий (23) будет иметь точный размер ε.

Проверяя гипотезу, мы задали ε, затем по точному или предельному распределению $\rho(\vec{X}) \Rightarrow \eta \Subset \mathcal{G}$ вычислили «барьер» C, с которым сравнили значение $|\rho(\vec{X})|$. На практике поступают иначе. Пусть по данной реализации выборки получено число $\rho^* = \rho\bigl(\vec{X}(\omega_0)\bigr)$. Число

$$\varepsilon^* = \mathsf{P}(|\eta\,| > |\rho^*|)$$

называют *реально достигнутым уровнем значимости* критерия. По величине ε^* можно судить о том, следует принять или отвергнуть основную гипотезу. Именно это число является результатом проверки гипотезы в любом статистическом пакете программ. Каков же смысл величины ε^*?

Легко проверить, что критерий (23) можно записать так:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если} \quad \varepsilon^* > \varepsilon, \\ H_2, & \text{если} \quad \varepsilon^* \leqslant \varepsilon. \end{cases}$$

При больших n вероятность

$$\mathsf{P}_{H_1}(|\rho(\vec{X})| > |\rho^*|) \tag{24}$$

стремится к ε^* или равна ей — в зависимости от того, является $\mathcal{G}$ точным или предельным распределением для $\rho(\vec{X})$. Поэтому ε^* есть почти то же самое, что (24). Вероятность (24) имеет следующий смысл: это вероятность, взяв выборку из распределения $\mathcal{F}_1$, получить по ней *большее* отклонение $|\rho(\vec{X})|$ эмпирического от истинного распределения, чем получено по проверяемой выборке. Бóльшие значения вероятности (24) или ε^* свидетельствуют в пользу основной гипотезы. Напротив, малые значения вероятности (24) или ε^* свидетельствуют в пользу альтернативы.

Если, например, вероятность (24) равна 0,2, следует ожидать, что в среднем 20 % «контрольных» выборок, удовлетворяющих основной гипотезе (каждая пятая), будут обладать бóльшим отклонением $|\rho(\vec{X})|$ по сравнению с тестируемой выборкой, в принадлежности которой распределению $\mathcal{F}_1$ мы не уверены. Можно отсюда сделать вывод, что тестируемая выборка ведёт себя не хуже, чем 20 % «правильных» выборок.

Но попадание в область вероятности 0,2 не является редким или «почти невозможным» событием. В статистике редкими обычно считают события с вероятностями $\varepsilon = 0{,}01$ или $\varepsilon = 0{,}05$ (это зависит от последствий ошибочного решения). Поэтому при $\varepsilon^* = 0{,}2 > 0{,}05$ основную гипотезу можно принять.

Реально достигнутый уровень ε^* равен точной верхней грани тех значений ε, при которых критерий (23) размера ε принимает гипотезу H_1.

§ 2. Критерии для проверки гипотезы о распределении

Критерий Колмогорова. Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из распределения $\mathcal{F}$. Проверяется простая гипотеза $H_1 = \{\mathcal{F} = \mathcal{F}_1\}$ против сложной альтернативы $H_2 = \{\mathcal{F} \neq \mathcal{F}_1\}$. В том случае, когда распределение $\mathcal{F}_1$ имеет *непрерывную* функцию распределения F_1, можно пользоваться критерием Колмогорова.

Пусть

$$\rho(\vec{X}) = \sqrt{n} \sup_y |F_n^*(y) - F_1(y)|.$$

Покажем, что $\rho(\vec{X})$ обладает свойствами (K1), (K2). Если H_1 верна, то X_i имеют распределение $\mathcal{F}_1$. По теореме Колмогорова $\rho(\vec{X}) \Rightarrow \eta$, где случайная величина η имеет распределение с функцией распределения Колмогорова $K(y)$ (рис. 11).

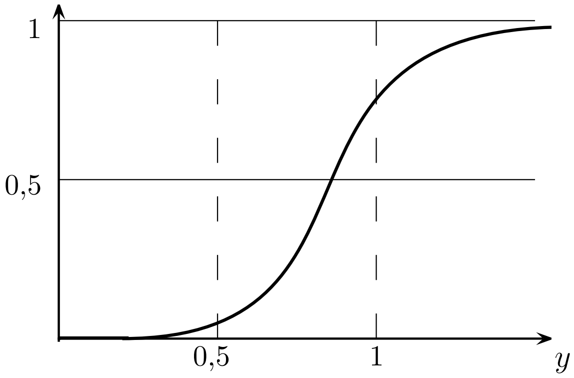

Рис. 11. График функции $K(y)$

Если H_1 неверна, то X_i имеют распределение $\mathcal{F}_2$, отличное от $\mathcal{F}_1$. По теореме Гливенко — Кантелли $F_n^*(y) \xrightarrow{\text{p}} F_2(y)$ для любого y при $n \to \infty$. Но $\mathcal{F}_1 \neq \mathcal{F}_2$, поэтому найдётся y_0 такое, что $|F_2(y_0) - F_1(y_0)| > 0$. Тогда

$$\sup_y |F_n^*(y) - F_1(y)| \geqslant |F_n^*(y_0) - F_1(y_0)| \xrightarrow{\text{p}} |F_2(y_0) - F_1(y_0)| > 0.$$

Умножив на $\sqrt{n}$, получим, что $\rho(\vec{X}) = \sqrt{n}\, \sup_y |F_n^*(y) - F_1(y)| \xrightarrow{\text{p}} \infty$.

Пусть случайная величина η имеет распределение с функцией распределения Колмогорова $K(y)$. Это распределение табулировано, т. е. по заданному ε легко найти C такое, что $\varepsilon = \mathsf{P}(\eta \geqslant C)$.

Критерий Колмогорова выглядит так:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}) < C, \\ H_2, & \text{если } \rho(\vec{X}) \geqslant C. \end{cases}$$

Критерий χ^2 Пирсона. Критерий χ^2 основывается на группированных данных. Область значений предполагаемого распределения $\mathcal{F}_1$ делят на некоторое число интервалов. После чего строят функцию отклонения ρ по разностям *теоретических вероятностей* попадания в интервалы группировки и *эмпирических частот*.

Дана выборка $\vec{X} = (X_1, \ldots, X_n)$ из распределения $\mathcal{F}$. Проверяется простая гипотеза $H_1 = \{\mathcal{F} = \mathcal{F}_1\}$ при альтернативе $H_2 = \{\mathcal{F} \neq \mathcal{F}_1\}$.

Пусть $A_1, \ldots, A_k$ — попарно непересекающиеся интервалы группировки, на которые разбита *вся* область значений случайной величины с распределением $\mathcal{F}_1$. Обозначим для $j = 1, \ldots, k$ через ν_j число элементов выборки, попавших в интервал A_j:

$$\nu_j = \{\text{число } X_i \in A_j\} = \sum_{i=1}^{n} \mathrm{I}(X_i \in A_j),$$

и через $p_j > 0$ — теоретическую вероятность $\mathsf{P}_{H_1}(X_1 \in A_j)$ попадания в интервал A_j случайной величины с распределением $\mathcal{F}_1$. По определению, $p_1 + \ldots + p_k = 1$. Как правило, длины интервалов выбирают так, чтобы $p_1 = \ldots = p_k = 1/k$. Пусть

$$\rho(\vec{X}) = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}. \tag{25}$$

Замечание 19. Свойство (K2) выполнено далеко не для всех альтернатив. Если распределение выборки $\mathcal{F}_2 \neq \mathcal{F}_1$ имеет такие же как у $\mathcal{F}_1$ вероятности p_j попадания в каждый из интервалов A_j, то по данной функции ρ эти распределения различить невозможно.

Поэтому на самом деле критерий, который мы построим по функции ρ из (25), решает совсем иную задачу. А именно, пусть задан набор «эталонных» вероятностей $p_1, \ldots, p_k$ такой, что $p_1 + \ldots + p_k = 1$. Критерий χ^2 предназначен для проверки сложной гипотезы о теоретическом распределении $\mathcal{F}$:

$$H_1' = \big\{\mathcal{F} \text{ таково, что } \mathsf{P}(X_1 \in A_j) = p_j \ \ \forall\, j = 1, \ldots, k\big\}$$

против сложной альтернативы $H_2' = \{H_1' \text{ неверна}\}$, т. е.

$$H_2' = \big\{\text{хотя бы для одного из интервалов } \mathsf{P}(X_1 \in A_j) \neq p_j\big\}.$$

Покажем, что $\rho(\vec{X})$ удовлетворяет условию (K1) независимо от того, проверяем ли мы гипотезу H_1 или H_1'.

Теорема 24 (Пирсона). *Если верна гипотеза H_1 или H_1', то при фиксированном k и при $n \to \infty$*

$$\rho(\vec{X}) = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j} \Rightarrow \mathrm{H}_{k-1},$$

где H_{k-1} есть χ^2-распределение с $k-1$ степенью свободы[1].

Доказательство. Докажем теорему Пирсона при $k = 2$. В этом случае $\nu_2 = n - \nu_1$, $p_2 = 1 - p_1$. Посмотрим на ρ и вспомним ЦПТ:

$$\rho(\vec{X}) = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(n - \nu_1 - n(1-p_1))^2}{n(1-p_1)} =$$

$$= \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(-\nu_1 + np_1)^2}{n(1-p_1)} = \frac{(\nu_1 - np_1)^2}{np_1(1-p_1)} = \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2.$$

Но величина ν_1 есть сумма n независимых случайных величин с распределением Бернулли B_{p_1}, и по ЦПТ

$$\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} \Rightarrow \xi \in \mathrm{N}_{0,1}, \qquad \rho(\vec{X}) = \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2 \Rightarrow \xi^2 \in \mathrm{H}_1.$$

Доказательство при произвольном k приведено в главе X. □

Функция $\rho(\vec{X})$ удовлетворяет условию (K2), если рассматривать гипотезы H_1' и H_2' вместо первоначальных.

Упражнение. Вспомнить закон больших чисел и доказать, что если H_1' неверна, то найдётся $j \in \{1, \ldots, k\}$ такое, что

$$\frac{(\nu_j - np_j)^2}{np_j} = \frac{n}{p_j}\left(\frac{\nu_j}{n} - p_j\right)^2 \xrightarrow{\mathrm{p}} \infty \text{ при } n \to \infty.$$

Осталось построить критерий согласия по определению 30. Пусть случайная величина η имеет распределение H_{k-1}. По таблице распределения H_{k-1} найдём C, равное квантили уровня $1 - \varepsilon$ этого распределения: $\varepsilon = \mathsf{P}(\eta \geqslant C)$. Критерий χ^2 устроен обычным образом:

$$\delta(\vec{X}) = \begin{cases} H_1', & \text{если } \rho(\vec{X}) < C, \\ H_2', & \text{если } \rho(\vec{X}) \geqslant C. \end{cases}$$

---

[1] Здесь следует остановиться и задать себе вопрос. Величина ρ есть сумма k слагаемых. Слагаемые, если мы не забыли теорему Муавра — Лапласа, имеют распределения, близкие к квадратам каких-то нормальных. Куда потерялась одна степень свободы? Причина кроется, конечно, в зависимости слагаемых: $\nu_k = n - \nu_1 - \ldots - \nu_{k-1}$.

Замечание 20. На самом деле критерий χ^2 применяют и для решения первоначальной задачи о проверке гипотезы $H_1 = \{\mathcal{F} = \mathcal{F}_1\}$. Необходимо только помнить, что этот критерий не отличит основную гипотезу от альтернативы, если вероятности попадания в интервалы разбиения у альтернативы такие же как у $\mathcal{F}_1$. Поэтому берут большое число интервалов разбиения — чем больше, тем лучше, чтобы «уменьшить» число альтернатив, неразличимых с предполагаемым распределением.

С другой стороны, следующее замечание предостерегает нас от чрезмерно большого числа интервалов.

Замечание 21. Сходимость по распределению $\rho(\vec{X}) \Rightarrow \mathrm{H}_{k-1}$ обеспечивается ЦПТ, поэтому разница допредельной и предельной вероятностей имеет тот же порядок, что и погрешность нормального приближения

$$|\,\mathsf{P}_{H_1}(\rho(\vec{X}) \geqslant C) - \mathsf{P}(\chi^2_{k-1} \geqslant C)| \leqslant \max\left\{\frac{b}{\sqrt{np_j(1-p_j)}}\right\},$$

где b — некоторая постоянная (неравенство Берри — Эссеена). Маленькие значения np_j в знаменателе приведут к тому, что распределение $\rho(\vec{X})$ будет существенно отличаться от H_{k-1}. Тогда и реальная вероятность $\mathsf{P}(\rho \geqslant C)$ — точный размер полученного критерия — будет сильно отличаться от ε. Поэтому число интервалов разбиения выбирают так, чтобы обеспечить нужную точность при замене распределения $\rho(\vec{X})$ на H_{k-1}. Обычно требуют, чтобы $np_1 = \ldots = np_k$ были не менее 5—6.

Критерий χ^2 для проверки параметрической гипотезы. Критерий χ^2 часто применяют для проверки гипотезы о принадлежности распределения выборки некоторому параметрическому семейству.

Пусть дана выборка $\vec{X} = (X_1, \ldots, X_n)$ из неизвестного распределения $\mathcal{F}$. Проверяется сложная гипотеза

$$H_1 = \bigl\{\mathcal{F} \in \{\mathcal{F}_\theta;\ \theta \in \Theta \subseteq \mathbb{R}^d\}\bigr\},$$

где θ — неизвестный параметр, d — его размерность.

Разобьём всю числовую ось на $k > d+1$ интервалов группировки $A_1, \ldots A_k$ и вычислим ν_j — число элементов выборки, попавших в интервал A_j. Но теперь вероятность $p_j = \mathsf{P}_{H_1}(X_1 \in A_j) = p_j(\theta)$ зависит от неизвестного параметра θ. Функция отклонения (25) также зависит от неизвестного параметра θ, и использовать её в критерии Пирсона нельзя:

$$\rho(\vec{X};\, \theta) = \sum_{j=1}^{k} \frac{(\nu_j - np_j(\theta))^2}{np_j(\theta)}. \qquad (26)$$

Пусть $\theta^* = \theta^*(\vec{X})$ — значение параметра θ, доставляющее минимум функции $\rho(\vec{X};\, \theta)$ при данной выборке $\vec{X}$. Подставив вместо истинных вероятностей p_j их оценки $p_j(\theta^*)$, получим функцию отклонения

$$\rho(\vec{X};\, \theta^*) = \sum_{j=1}^{k} \frac{(\nu_j - np_j(\theta^*))^2}{np_j(\theta^*)}. \tag{27}$$

Условие (K1) (при выполнении некоторых условий[2] относительно гладкости $p_j(\theta)$) обеспечивается теоремой, которую мы доказывать не будем.

Теорема 25 (Р. Фишер). *Если верна гипотеза H_1, d — размерность вектора параметров θ, то при фиксированном k и при $n \to \infty$*

$$\rho(\vec{X};\, \theta^*) = \sum_{j=1}^{k} \frac{(\nu_j - np_j(\theta^*))^2}{np_j(\theta^*)} \Rightarrow \mathrm{H}_{k-1-d},$$

где H_{k-1-d} есть χ^2-распределение с $k-1-d$ степенями свободы.

Построим критерий χ^2. Пусть случайная величина η имеет распределение H_{k-1-d}. По заданному ε найдём C такое, что $\varepsilon = \mathsf{P}(\eta \geqslant C)$.

Критерий согласия χ^2 устроен обычным образом:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X};\, \theta^*) < C, \\ H_2, & \text{если } \rho(\vec{X};\, \theta^*) \geqslant C. \end{cases}$$

Замечания 20, 21 о количестве интервалов разбиения остаются в силе.

Замечание 22. Вычисление точки минимума функции $\rho(\vec{X};\, \theta)$ в общем случае возможно лишь численно. Поэтому есть соблазн использовать вместо оценки θ^* оценку максимального правдоподобия $\hat{\theta}$, построенную по выборке $X_1, \ldots, X_n$. Однако при такой замене предельное распределение величины $\rho(\vec{X};\, \theta)$ уже не равно H_{k-1-d} и зависит от θ.

Попытаемся всё же использовать простую оценку $\hat{\theta}$ вместо сложно вычисляемой θ^*. По определению, $\rho(\vec{X};\, \theta^*) \leqslant \rho(\vec{X};\, \hat{\theta})$. И если верно неравенство $\rho(\vec{X};\, \hat{\theta}) < C$, то тем более $\rho(\vec{X};\, \theta^*) < C$. Таким образом, если гипотеза H_1 принимается из-за того, что $\rho(\vec{X};\, \hat{\theta}) < C$, она тем более будет приниматься по функции $\rho(\vec{X};\, \theta^*)$. Но для того чтобы отвергнуть основную гипотезу, придётся вычислять $\rho(\vec{X};\, \theta^*)$.

---

[2] Все $\partial^2 p_j(\theta)/\partial\theta_i\partial\theta_l$ непрерывны по θ; ранг матрицы $\|\partial p_j(\theta)/\partial\theta_i\|$ равен d.

§ 3. Критерии для проверки однородности

Двувыборочный критерий Колмогорова — Смирнова. Даны две независимые выборки $\vec{X} = (X_1, \ldots, X_n)$ и $\vec{Y} = (Y_1, \ldots, Y_m)$ из неизвестных распределений $\mathcal{F}$ и $\mathcal{G}$ соответственно. Проверяется сложная гипотеза $H_1 = \{\mathcal{F} = \mathcal{G}\}$ при альтернативе $H_2 = \{H_1 \text{ неверна}\}$.

Критерий Колмогорова — Смирнова используют, если $\mathcal{F}$ и $\mathcal{G}$ имеют *непрерывные функции распределения.*

Пусть $F_n^*(y)$ и $G_m^*(y)$ — эмпирические функции распределения, построенные по выборкам $\vec{X}$ и $\vec{Y}$,

$$\rho(\vec{X}, \vec{Y}) = \sqrt{\frac{mn}{m+n}} \sup_y \left| F_n^*(y) - G_m^*(y) \right|.$$

Теорема 26. *Если гипотеза H_1 верна, то $\rho(\vec{X}, \vec{Y}) \Rightarrow \eta$ при $n, m \to \infty$, где η имеет распределение Колмогорова.*

Упражнение. Доказать, что $\rho(\vec{X}, \vec{Y}) \xrightarrow{\text{p}} \infty$ при любом стремлении $n, m \to \infty$, если H_2 верна.

В таблице распределения Колмогорова по заданному ε найдём C такое, что $\varepsilon = \mathsf{P}(\eta \geqslant C)$, и построим критерий Колмогорова — Смирнова

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}) < C, \\ H_2, & \text{если } \rho(\vec{X}) \geqslant C. \end{cases}$$

Замечание 23. Если есть более двух выборок, и требуется проверить гипотезу однородности, часто пользуются одним из вариантов критерия χ^2 Пирсона. Этот критерий можно посмотреть в [3, § 3.4].

Ранговый критерий Вилкоксона, Манна и Уитни. Даны две независимые выборки $\vec{X} = (X_1, \ldots, X_n)$ и $\vec{Y} = (Y_1, \ldots, Y_m)$ из неизвестных распределений $\mathcal{F}$ и $\mathcal{G}$. Проверяется сложная гипотеза $H_1 = \{\mathcal{F} = \mathcal{G}\}$ при альтернативе $H_2 = \{H_1 \text{ неверна}\}$.

Критерий Вилкоксона, Манна и Уитни (Wilcoxon, Mann, Whitney) используют, если $\mathcal{F}$ и $\mathcal{G}$ имеют *непрерывные функции распределения.* Составим из выборок $\vec{X}$ и $\vec{Y}$ общий вариационный ряд и подсчитаем статистику Вилкоксона W, равную сумме рангов $r_1, \ldots, r_m$ (номеров мест) элементов выборки $\vec{Y}$ в общем вариационном ряду. Зададим функцию U так (статистика Манна — Уитни):

$$U = W - \frac{1}{2}\, m(m+1) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{I}(X_i < Y_j).$$

Если H_1 верна, то $\mathsf{P}(X_1 < Y_1) = \frac{1}{2}$ (для этого требуется непрерывность распределений выборок). В этом случае

$$\mathsf{E}U = \frac{nm}{2}, \qquad \mathsf{D}U = \frac{nm(n+m+1)}{12}.$$

Статистику критерия зададим, центрировав и нормировав статистику U:

$$\rho(\vec{X}, \vec{Y}) = \frac{U - nm/2}{\sqrt{nm(n+m+1)/12}}.$$

Мы не будем доказывать следующее утверждение.

Теорема 27. *Если непрерывные распределения $\mathcal{F}$ и $\mathcal{G}$ таковы, что $\mathsf{P}(X_1 < Y_1) = \frac{1}{2}$, то $\rho(\vec{X}, \vec{Y}) \Rightarrow \mathrm{N}_{0,1}$ при $n, m \to \infty$.*

Построим критерий асимптотического размера ε:

$$\delta(\vec{X}, \vec{Y}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C, \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C, \end{cases}$$

где C — квантиль уровня $1-\varepsilon/2$ распределения $\mathrm{N}_{0,1}$. Пользоваться этим критерием рекомендуют при $\min(n, m) > 25$.

Условие (K2) не выполнено. Если $\mathcal{F} \neq \mathcal{G}$, но $\mathsf{P}(X_1 < Y_1) = 1/2$, то по теореме 27 статистика $\rho(\vec{X}, \vec{Y})$ ведёт себя так же, как и при основной гипотезе, поэтому критерий не будет состоятельным. Он реагирует лишь на то, как часто $X_i < Y_j$, и принимает H_1, если это происходит примерно в половине случаев.

Например, если $\mathcal{F}$ и $\mathcal{G}$ — два нормальных распределения с одним и тем же средним, но разными дисперсиями, то разность $X_i - Y_j$ имеет нормальное распределение с нулевым средним, и условие теоремы 27 выполнено.

Итак, на самом деле построенный выше критерий проверяет гипотезу

$$H_1' = \left\{\text{распределения выборок таковы, что } \mathsf{P}(X_1 < Y_1) = \tfrac{1}{2}\right\}.$$

Используя его для проверки первоначальной гипотезы однородности, следует помнить, какие альтернативы он не отличает от основной гипотезы.

Существуют модификации этого критерия (критерий Вилкоксона), которые применяют, если заранее известно, каких альтернатив следует опасаться. В качестве W вместо суммы рангов $r_1, \ldots, r_m$ возьмём сумму $s(r_1), \ldots, s(r_m)$, где $s : (1, \ldots, n+m) \mapsto (s(1), \ldots, s(n+m))$ — заранее выбранная перестановка всех рангов. Статистика $U = W - m(m+1)/2$ уже не выражается через индикаторы, но её асимптотическая нормальность при верной гипотезе H_1 по-прежнему имеет место.

Например, при альтернативе с дисперсиями $\mathsf{D}Y_1 \gg \mathsf{D}X_1$ (см. выше), когда ранги Y_j могут оказаться очень большими и очень маленькими, берут перестановку вида

$$(1, 2, 3, 4, 5, 6, 7, 8, 9, 10) \mapsto (10, 8, 6, 4, 2, 1, 3, 5, 7, 9),$$

присваивающую бо́льшие значения крайним номерам.

Критерий Фишера. Критерий Фишера используют в качестве первого шага в задаче проверки однородности двух независимых нормальных выборок. Особенно часто возникает необходимость проверить равенство *средних* двух нормальных совокупностей: например, в медицине или биологии для выяснения наличия или отсутствия действия препарата. Эта задача решается с помощью критерия Стьюдента (с ним мы познакомимся на следующей странице), но только в случае, когда неизвестные дисперсии *равны*. Для проверки же равенства дисперсий пользуются сначала критерием Фишера. Самое печальное, если гипотеза равенства дисперсий отвергается критерием Фишера. Задачу о построении критерия точного размера ε (что особенно важно при маленьких выборках) для проверки равенства средних в этих условиях называют *проблемой Беренса — Фишера*. Её решение возможно лишь в частных случаях.

Пусть даны две независимые выборки из нормальных распределений: $\vec{X} = (X_1, \ldots, X_n)$ из $\mathrm{N}_{a_1, \sigma_1^2}$ и $\vec{Y} = (Y_1, \ldots, Y_m)$ из $\mathrm{N}_{a_2, \sigma_2^2}$, средние которых, вообще говоря, неизвестны. Критерий Фишера предназначен для проверки гипотезы $H_1 = \{\sigma_1 = \sigma_2\}$.

Обозначим через $S_0^2(\vec{X})$ и $S_0^2(\vec{Y})$ несмещённые выборочные дисперсии

$$S_0^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_0^2(\vec{Y}) = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \overline{Y})^2$$

и зададим функцию $\rho(\vec{X}, \vec{Y})$ как их отношение $\rho(\vec{X}, \vec{Y}) = S_0^2(\vec{X})/S_0^2(\vec{Y})$.

Удобно, если $\rho > 1$. С этой целью выборкой $\vec{X}$ называют ту из двух выборок, несмещённая дисперсия которой больше. Поэтому предположим, что $S_0^2(\vec{X}) > S_0^2(\vec{Y})$.

Теорема 28. *При верной гипотезе H_1 величина $\rho(\vec{X}, \vec{Y})$ имеет распределение Фишера $\mathrm{F}_{n-1, m-1}$ с $n-1$ и $m-1$ степенями свободы.*

Доказательство. По лемме Фишера, независимые случайные величины

$$\chi_{n-1}^2 = \frac{(n-1)\,S_0^2(\vec{X})}{\sigma_1^2} \quad \text{и} \quad \psi_{m-1}^2 = \frac{(m-1)\,S_0^2(\vec{Y})}{\sigma_2^2}$$

имеют распределения H_{m-1} и H_{n-1} соответственно. При $\sigma_1 = \sigma_2$, по определению распределения Фишера,

$$\rho(\vec{X}, \vec{Y}) = \frac{S_0^2(\vec{X})}{\sigma_1^2} \cdot \frac{\sigma_2^2}{S_0^2(\vec{Y})} = \frac{\chi_{n-1}^2 / (n-1)}{\psi_{m-1}^2 / (m-1)} \Subset \text{F}_{n-1,\, m-1}. \qquad \square$$

Возьмём квантиль $f_{1-\varepsilon}$ распределения Фишера $\text{F}_{n-1,\, m-1}$. Критерием Фишера называют критерий

$$\delta(\vec{X}, \vec{Y}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}, \vec{Y}) < f_{1-\varepsilon}, \\ H_2 & \text{если } \rho(\vec{X}, \vec{Y}) \geqslant f_{1-\varepsilon}. \end{cases}$$

Упражнение. Доказать, что для любой альтернативы $\sigma_1 \neq \sigma_2$

$$\rho(\vec{X}, \vec{Y}) \xrightarrow{\text{p}} \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \text{ при } n, m \to \infty. \tag{28}$$

Докажем состоятельность критерия Фишера. Достаточно в качестве альтернативы рассмотреть $\sigma_1 > \sigma_2$ (иначе при больших объёмах выборок будет, согласно (28), нарушаться предположение $S_0^2(\vec{X}) > S_0^2(\vec{Y})$).

Убедимся, что последовательность квантилей $f_\delta = f_\delta(n, m)$ распределения $\text{F}_{n,\, m}$ любого уровня $\delta \in (0,\, 1)$ сходится к единице при $n,\, m \to \infty$.

Пусть $\xi_{n,m} \Subset \text{F}_{n,\, m}$. По определению квантилей, вероятности $\mathsf{P}(\xi_{n,m} < f_\delta)$ и $\mathsf{P}(\xi_{n,m} > f_\delta)$ не зависят от n, m и равны фиксированным числам δ и $1 - \delta$ соответственно.

Знаем, что $\xi_{n,m} \xrightarrow{\text{p}} 1$. Поэтому для любого фиксированного $\varepsilon > 0$ вероятности $\mathsf{P}(\xi < 1 - \varepsilon)$ и $\mathsf{P}(\xi > 1 + \varepsilon)$ стремятся к нулю при $n,\, m \to \infty$, становясь *рано или поздно* меньше как δ, так и $1 - \delta$.

Следовательно, $1 - \varepsilon < f_\delta < 1 + \varepsilon$ при *достаточно больших* n, m. Это означает, что $f_\delta \to 1$ при $n,\, m \to \infty$.

Для доказательства состоятельности критерия Фишера осталось предположить, что гипотеза H_1 неверна, т. е. $\sigma_1 > \sigma_2$, и использовать сходимости (28) и $f_\delta \to 1$. Сходимость по вероятности влечёт слабую сходимость

$$\rho(\vec{X}, \vec{Y}) - f_{1-\varepsilon} \Rightarrow \frac{\sigma_1^2}{\sigma_2^2} - 1 > 0.$$

Предельная функция распределения $\mathsf{P}(\sigma_1^2 / \sigma_2^2 - 1 < x)$ непрерывна в точке $x = 0$ (*почему*?) и равна нулю в этой точке. Отсюда

$$\alpha_2(\delta) = \mathsf{P}_{H_2}(\rho < f_{1-\varepsilon}) = \mathsf{P}_{H_2}(\rho - f_{1-\varepsilon} < 0) \to \mathsf{P}\Big(\frac{\sigma_1^2}{\sigma_2^2} - 1 < 0\Big) = 0. \quad \square$$

Критерий Стьюдента. Пусть имеются две независимые выборки: выборка $\vec{X} = (X_1, \ldots, X_n)$ из $\mathrm{N}_{a_1, \sigma^2}$ и выборка $\vec{Y} = (Y_1, \ldots, Y_m)$ из $\mathrm{N}_{a_2, \sigma^2}$ с неизвестными средними и *одной и той же* неизвестной дисперсией σ^2. Проверяется сложная гипотеза $H_1 = \{a_1 = a_2\}$.

Построим критерий Стьюдента *точного* размера ε.

Теорема 29. *Случайная величина* t_{n+m-2}, *равная*

$$t_{n+m-2} = \sqrt{\frac{nm}{n+m}} \cdot \frac{(\overline{X} - a_1) - (\overline{Y} - a_2)}{\sqrt{\dfrac{(n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})}{n+m-2}}}$$

имеет распределение Стьюдента T_{n+m-2}.

Доказательство. Легко видеть (*убедитесь, что легко*!), что случайная величина $\overline{X} - a_1$ имеет распределение $\mathrm{N}_{0, \sigma^2/n}$, а случайная величина $\overline{Y} - a_2$ имеет распределение $\mathrm{N}_{0, \sigma^2/m}$. Тогда их разность распределена тоже нормально с нулевым средним и дисперсией

$$\mathsf{D}\big((\overline{X} - a_1) - (\overline{Y} - a_2)\big) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2 \cdot \frac{n+m}{nm}.$$

Нормируем эту разность:

$$\xi_0 = \frac{1}{\sigma} \sqrt{\frac{nm}{n+m}} \, \big((\overline{X} - a_1) - (\overline{Y} - a_2)\big) \in \mathrm{N}_{0,1}$$

Из леммы Фишера следует, что независимые случайные величины $(n-1)\, S_0^2(\vec{X})/\sigma^2$ и $(m-1)\, S_0^2(\vec{Y})/\sigma^2$ имеют распределения H_{n-1} и H_{m-1} соответственно, а их сумма

$$S^2 = \frac{1}{\sigma^2} \, \big((n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})\big)$$

имеет χ^2-распределение H_{n+m-2} с $n+m-2$ степенями свободы (*почему*?) и не зависит от $\overline{X}$ и от $\overline{Y}$ (*почему*?).

По определению 19 (с. 71), отношение $\dfrac{\xi_0}{\sqrt{S^2/(n+m-2)}}$ имеет распределение Стьюдента T_{n+m-2}. Осталось подставить в эту дробь ξ_0 и S^2 и убедиться, что σ сократится и получится t_{n+m-2} из теоремы 29. □

Введём функцию

$$\rho(\vec{X}, \vec{Y}) = \sqrt{\frac{nm}{n+m}} \cdot \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{(n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})}{n+m-2}}}.$$

Из теоремы 29 следует свойство (K1): если H_1 верна, т. е. если $a_1 = a_2$, то величина $\rho = t_{n+m-2}$ имеет распределение Стьюдента T_{n+m-2}.

Поэтому остаётся по ε найти $C = \tau_{1-\varepsilon/2}$ — квантиль распределения T_{n+m-2}. Критерий Стьюдента выглядит как все критерии согласия:

$$\delta(\vec{X}, \vec{Y}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X}, \vec{Y})| < C, \\ H_2, & \text{если } |\rho(\vec{X}, \vec{Y})| \geqslant C. \end{cases}$$

Упражнение. Доказать, что этот критерий имеет размер ε.

Упражнение. Доказать свойство (K2): если $a_1 \neq a_2$, величина $|\rho|$ неограниченно возрастает по вероятности с ростом n и m.

Указание. Воспользовавшись ЗБЧ для числителя и свойством 3 (с. 69) распределения χ^2 — для знаменателя, доказать, что числитель и знаменатель сходятся к постоянным:

$$\overline{X} - \overline{Y} \xrightarrow{\text{p}} \text{const} \neq 0, \qquad \frac{(n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})}{n+m-2} \xrightarrow{\text{p}} \text{const} \neq 0,$$

тогда как корень перед дробью неограниченно возрастает.

Однофакторный дисперсионный анализ. Предположим, что влияние некоторого «фактора» на наблюдаемые нормально распределённые величины может сказываться только на значениях их математических ожиданий. Мы наблюдаем несколько выборок при различных «уровнях» фактора. Требуется определить, влияет или нет изменение уровня фактора на математическое ожидание.

Говоря формальным языком, однофакторный дисперсионный анализ решает задачу проверки равенства средних нескольких независимых нормально распределённых выборок с одинаковыми дисперсиями. Для двух выборок эту задачу мы решили с помощью критерия Стьюдента.

Пусть даны k независимых выборок

$$X^{(1)} = (x_1^{(1)}, \ldots, x_{n_1}^{(1)}), \quad \ldots, \quad X^{(k)} = (x_1^{(k)}, \ldots, x_{n_k}^{(k)})$$

из нормальных распределений $x_i^{(j)} \in \mathrm{N}_{a_j, \sigma^2}$ с одной и той же дисперсией. Верхний индекс у наблюдений отвечает номеру выборки. Проверяется основная гипотеза $H_1 = \{a_1 = \ldots = a_k\}$.

Для каждой выборки вычислим выборочные среднее и дисперсию

$$\overline{X}^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i^{(j)}, \qquad S^{(j)} = \frac{1}{n_j}\sum_{i=1}^{n_j}\bigl(x_i^{(j)} - \overline{X}^{(j)}\bigr)^2.$$

Положим $n = n_1 + \ldots + n_k$. Определим также общее выборочное среднее и общую выборочную дисперсию

$$\overline{X} = \frac{1}{n}\sum_{i,j} x_i^{(j)} = \frac{1}{n}\sum_{j=1}^{k} n_j \overline{X}^{(j)}, \qquad S^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\bigl(x_i^{(j)} - \overline{X}\bigr)^2.$$

Критерий для проверки гипотезы H_1 основан на сравнении внутригрупповой и межгрупповой дисперсий. Определим эти характеристики. Вычислим так называемую *межгрупповую дисперсию*, или *дисперсию выборочных средних*

$$S_{\text{м}}^2 = \frac{1}{n}\sum_{j=1}^{k} n_j\bigl(\overline{X}^{(j)} - \overline{X}\bigr)^2.$$

Она показывает, насколько отличны друг от друга выборочные средние при разных уровнях фактора. Именно эта дисперсия отражает влияние фактора. При этом каждое выборочное среднее вносит в дисперсию вклад, пропорциональный объёму соответствующей выборки: выбросы средних могут быть вызваны малым числом наблюдений.

Вычислим так называемую *внутригрупповую дисперсию*

$$S_{\text{в}}^2 = \frac{1}{n}\sum_{j=1}^{k} n_j\, S^{(j)} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\bigl(x_i^{(j)} - \overline{X}^{(j)}\bigr)^2.$$

Она показывает, насколько велики разбросы внутри выборок относительно выборочных средних. Эти разбросы определяются случайностью внутри выборок. Вывод о том, что средние существенно различны, т. е. присутствует влияние фактора на среднее, может быть сделан, если межгрупповая дисперсия оказывается существенно больше внутригрупповой. Чтобы понять, насколько больше, следует рассмотреть распределения этих случайных величин при верной основной гипотезе.

По основному следствию из леммы Фишера при любом $j = 1, \ldots, k$ величина $n_j\, S^{(j)}/\sigma^2$ имеет распределение H_{n_j-1} и не зависит от $\overline{X}^{(j)}$. Из независимости выборок и устойчивости χ^2-распределения относительно суммирования получаем

$$\frac{nS_{\text{в}}^2}{\sigma^2} = \sum_{j=1}^{k}\frac{n_j\, S^{(j)}}{\sigma^2} \in \text{H}_{n-k}, \quad \text{где } n-k = n_1 - 1 + \ldots + n_k - 1.$$

Кроме того, величина $S^2_{\text{в}}$ не зависит от $\overline{X}^{(1)}, \ldots, \overline{X}^{(k)}$. Поэтому она не зависит и от их взвешенного среднего $\overline{X}$, а также (что уже совсем невероятно) от межгрупповой дисперсии $S^2_{\text{м}}$, поскольку последняя является функцией *только* от перечисленных средних. Эти свойства никак не связаны с проверяемой гипотезой и верны независимо от равенства или неравенства истинных средних.

Пусть гипотеза H_1 верна. Тогда выборки можно считать одной выборкой объёма n. По основному следствию леммы Фишера $nS^2/\sigma^2 \Subset \mathrm{H}_{n-1}$.

Величины S^2, $S^2_{\text{м}}$ и $S^2_{\text{в}}$ удовлетворяют легко проверяемому *основному дисперсионному соотношению*

$$\frac{nS^2}{\sigma^2} = \frac{nS^2_{\text{м}}}{\sigma^2} + \frac{nS^2_{\text{в}}}{\sigma^2}.$$

Величина в левой части имеет распределение H_{n-1}, справа — сумма двух независимых слагаемых, второе из которых имеет распределение H_{n-k}. Покажем, что тогда первое распределено по закону H_{k-1}.

Лемма 8. *Пусть случайные величины ξ и η независимы, причём $\xi \Subset \mathrm{H}_m$ и $\xi+\eta \Subset \mathrm{H}_s$. Тогда $\eta \Subset \mathrm{H}_{s-m}$.*

Доказательство. Используем известную нам характеристическую функцию $\varphi(t) = (1-2it)^{-k/2}$ гамма-распределения $\mathrm{H}_k = \Gamma_{1/2,\,k/2}$:

$$\varphi_\xi(t)\cdot\varphi_\eta(t) = (1-2it)^{-m/2}\cdot\varphi_\eta(t) = \varphi_{\xi+\eta}(t) = (1-2it)^{-s/2}.$$

Отсюда $\varphi_\eta(t) = (1-2it)^{-(s-m)/2}$ и $\eta \Subset \mathrm{H}_{s-m}$. □

Итак, при верной гипотезе H_1 мы получили два χ^2-распределения независимых случайных величин

$$\chi^2 = \frac{nS^2_{\text{м}}}{\sigma^2} \Subset \mathrm{H}_{k-1} \quad \text{и} \quad \psi^2 = \frac{nS^2_{\text{в}}}{\sigma^2} \Subset \mathrm{H}_{n-k}.$$

Построим по ним статистику из распределения Фишера $\mathrm{F}_{k-1,\,n-k}$

$$\rho = \frac{\chi^2}{k-1}\cdot\frac{n-k}{\psi^2} = \frac{n-k}{k-1}\cdot\frac{S^2_{\text{м}}}{S^2_{\text{в}}} \Subset \mathrm{F}_{k-1,\,n-k}.$$

По заданному ε найдём квантиль C уровня $1-\varepsilon$ распределения Фишера $\mathrm{F}_{k-1,\,n-k}$ и устроим следующий критерий точного размера ε:

$$\delta = \begin{cases} H_1, & \text{если } \rho < C, \\ H_2, & \text{если } \rho \geqslant C. \end{cases}$$

Замечание 24. Предположение о равенстве дисперсий проверяют, например, с помощью критерия Бартлетта (см. [6]).

§ 4. Критерий χ^2 для проверки независимости

Есть выборка (X_1, Y_1), …, (X_n, Y_n) значений двух наблюдаемых совместно случайных величин X и Y в n независимых экспериментах. Проверяется гипотеза $H_1 = \{X$ и Y независимы$\}$.

Введём k интервалов группировки $\Delta_1, \ldots, \Delta_k$ для значений X и m интервалов группировки $\nabla_1, \ldots, \nabla_m$ для значений Y:

$\vec{Y}$ $\vec{X}$	∇_1	∇_2	$\ldots$	∇_m	$\sum\limits_{j=1}^{m}$
Δ_1	ν_{11}	ν_{12}	$\ldots$	ν_{1m}	$\nu_{1\cdot}$
$\vdots$			$\ldots$		
Δ_k	ν_{k1}	ν_{k2}	$\ldots$	ν_{km}	$\nu_{k\cdot}$
$\sum\limits_{i=1}^{k}$	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$	$\ldots$	$\nu_{\cdot m}$	n

Посчитаем эмпирические частоты:

$$\nu_{ij} = \text{число пар } (X_l,\, Y_l), \text{ попавших в } \Delta_i \times \nabla_j,$$

$$\nu_{\cdot j} = \text{число } Y_l, \text{ попавших в } \nabla_j, \quad \nu_{i\cdot} = \text{число } X_l, \text{ попавших в } \Delta_i.$$

Если гипотеза H_1 верна, то теоретические вероятности попадания пары $(X,\, Y)$ в любую из областей $\Delta_i \times \nabla_j$ равны произведению вероятностей: для всех i и j

$$p_{ij} = \mathsf{P}\big((X,\, Y) \in \Delta_i \times \nabla_j\big) = \mathsf{P}\big(X \in \Delta_i\big) \cdot \mathsf{P}\big(Y \in \nabla_j\big) = p_i^x \cdot p_j^y$$

Именно эту гипотезу (назовём её H_1') мы в действительности и проверяем. По ЗБЧ при $n \to \infty$

$$\frac{\nu_{i\cdot}}{n} \xrightarrow{\mathrm{p}} p_i^x, \qquad \frac{\nu_{\cdot j}}{n} \xrightarrow{\mathrm{p}} p_i^y, \qquad \frac{\nu_{ij}}{n} \xrightarrow{\mathrm{p}} p_{ij}.$$

Поэтому большая разница между $\dfrac{\nu_{ij}}{n}$ и $\dfrac{\nu_{i\cdot}}{n} \times \dfrac{\nu_{\cdot j}}{n}$ (или между ν_{ij} и $\dfrac{\nu_{i\cdot}\,\nu_{\cdot j}}{n}$) служит основанием для отклонения гипотезы независимости. Пусть

$$\rho(\vec{X},\, \vec{Y}) = n \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\big(\nu_{ij} - (\nu_{i\cdot}\nu_{\cdot j})/n\big)^2}{\nu_{i\cdot}\nu_{\cdot j}}. \tag{29}$$

Теорема 30. *Если гипотеза H_1 верна, то при $n \to \infty$*

$$\rho(\vec{X},\, \vec{Y}) \Rightarrow \mathrm{H}_{(k-1)(m-1)}.$$

Критерий согласия асимптотического размера ε строится как обычно: по заданному ε найдём C, равное квантили уровня $1-\varepsilon$ распределения $\mathrm{H}_{(k-1)(m-1)}$. Тогда критерий имеет вид

$$\delta(\vec{X}, \vec{Y}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}, \vec{Y}) < C, \\ H_2, & \text{если } \rho(\vec{X}, \vec{Y}) \geqslant C. \end{cases}$$

Упражнение. Чтобы объяснить вид функции ρ и теорему 30, убедитесь, что гипотеза H_1' есть гипотеза о принадлежности распределения выборки параметрическому семейству распределений с вектором неизвестных параметров $(p_1^x, \ldots, p_{k-1}^x; p_1^y, \ldots, p_{m-1}^y)$, имеющим размерность $d = k + m - 2$. Подставив ОМП $\nu_{i\cdot}/n$ для p_i^x и $\nu_{\cdot j}/n$ для p_j^y в функцию

$$\rho = \sum_{i,j} \frac{(\nu_{ij} - np_i^x p_j^y)^2}{np_i^x p_j^y}$$

из равенства (26), получим (29). Всего есть km интервалов. По теореме 25 (с. 100) при верной H_1' предельное χ^2-распределение имеет число степеней свободы $km - 1 - d = (k-1)(m-1)$.

Замечания 20 и 21 по поводу числа km интервалов группировки остаются в силе.

§ 5. Проверка простых гипотез о параметрах

Проверка гипотезы о среднем нормального распределения с известной дисперсией. Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из нормального распределения N_{a,σ^2} с известной дисперсией σ^2. Проверяется простая гипотеза $H_1 = \{a = a_0\}$ против сложной альтернативы $H_2 = \{a \neq a_0\}$.

Построим критерий *точного* размера ε с помощью функции

$$\rho(\vec{X}) = \sqrt{n}\,\frac{\overline{X} - a_0}{\sigma}.$$

Очевидно свойство (K1): если H_1 верна, то $\rho(\vec{X}) \in \mathrm{N}_{0,1}$.

По ε выберем $C = \tau_{1-\varepsilon/2}$ — квантиль стандартного нормального распределения. Критерий выглядит как все критерии согласия:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C, \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C. \end{cases} \tag{30}$$

Упражнение. Доказать (K2): если $a \neq a_0$, то $|\rho(\vec{X})| \xrightarrow{\mathrm{p}} \infty$. Доказать, что критерий 30 имеет точный размер ε и является состоятельным.

Проверка гипотезы о среднем нормального распределения с неизвестной дисперсией. Проверяется та же гипотеза, что и в предыдущем разделе, но в случае, когда дисперсия σ^2 неизвестна. Критерий, который мы построим, называют одновыборочным критерием Стьюдента.

Введём функцию отклонения

$$\rho(\vec{X}) = \sqrt{n}\,\frac{\overline{X} - a_0}{\sqrt{S_0^2}}, \qquad \text{где } S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

По п. 4 *самого полезного следствия леммы Фишера* (с. 78) выполнено свойство (K1): если $a = a_0$, то ρ имеет распределение Стьюдента T_{n-1}.

Критерий строится в точности как в (30), но в качестве C следует брать квантиль распределения Стьюдента, а не стандартного нормального распределения (*почему*?).

Упражнение. Доказать свойство (K2). Записать критерий и доказать, что он имеет точный размер ε и является состоятельным.

Критерии, основанные на доверительных интервалах. Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из семейства распределений $\mathcal{F}_\theta$, где $\theta \in \Theta$. Проверяется простая гипотеза $H_1 = \{\theta = \theta_0\}$ против сложной альтернативы $H_2 = \{\theta \neq \theta_0\}$.

Пусть имеется точный доверительный интервал (θ^-, θ^+) для параметра θ уровня доверия $1 - \varepsilon$. Взяв произвольное θ', для выборки из распределения $\mathcal{F}_{\theta'}$ имеем

$$\mathsf{P}(\theta^- < \theta' < \theta^+) = 1 - \varepsilon$$

Тогда критерий

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если} \quad \theta_0 \in (\theta^-, \theta^+), \\ H_2, & \text{если} \quad \theta_0 \notin (\theta^-, \theta^+) \end{cases}$$

имеет точный размер ε:

$$\alpha_1(\delta) = \mathsf{P}_{H_1}(\delta = H_2) = \mathsf{P}_{H_1}(\theta_0 \notin (\theta^-, \theta^+)) = 1 - \mathsf{P}_{H_1}(\theta^- < \theta_0 < \theta^+) = \varepsilon.$$

Если доверительный интервал строится с помощью функции $G(\vec{X};\, \theta)$, то эта же функция годится и в качестве «функции отклонения» $\rho(\vec{X})$ для построения критерия согласия. Критерий заданного асимптотического размера по асимптотическому доверительному интервалу строится совершенно аналогично.

§ 6. Вопросы и упражнения

1. Построить критерий для проверки равенства дисперсий двух независимых нормальных выборок с *известными* средними, статистика которого имеет при верной основной гипотезе распределение Фишера с n и m степенями свободы.

2. Построить критерий для проверки гипотезы о равенстве средних двух независимых нормальных выборок с произвольными *известными* дисперсиями, статистика которого имеет при верной основной гипотезе стандартное нормальное распределение.

3. Построить критерий точного размера ε для различения трёх гипотез о среднем нормального распределения с неизвестной дисперсией: $H_1 = \{a = a_0\}$, $H_2 = \{a < a_0\}$ и $H_3 = \{a > a_0\}$.

4. Какие из приведённых в главе VIII критериев можно сформулировать, используя доверительные интервалы? Сделать это.

5. Проверяется простая гипотеза о параметре $H_1 = \{\theta = \theta_0\}$ против альтернативы $H_2 = \{\theta \neq \theta_0\}$. Какими свойствами должен обладать доверительный интервал, чтобы критерий, построенный с его помощью, был состоятелен?

6. Имеется выборка из распределения Бернулли. Построить критерий для проверки гипотезы $p = 1/2$ при альтернативе $p \neq 1/2$.

7. Подбросить игральную кость 300 раз и проверить её правильность с помощью подходящего критерия.

8. Подбросить симметричную монету 200 раз и проверить своё умение правильно её подбрасывать с помощью критерия χ^2.

9. Построить критерий асимптотического размера ε для проверки гипотезы однородности двух независимых выборок с разными объёмами из распределения Бернулли.

10. Показать, что при $k = 2$ критерий для решения задачи однофакторного дисперсионного анализа совпадает с критерием Стьюдента.

11. Доказать основное дисперсионное соотношение.

ГЛАВА IX

ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКОЙ ЗАВИСИМОСТИ

Часто требуется определить, как зависит наблюдаемая случайная величина от одной или нескольких других величин. Самый общий случай такой зависимости — зависимость статистическая: например, $X = \xi + \eta$ и $Z = \xi + \varphi$ зависимы, но эта зависимость не функциональная. Для зависимых случайных величин имеет смысл рассмотреть математическое ожидание одной из них при фиксированном значении другой и выяснить, как влияет на среднее значение первой величины изменение значений второй. Так, стоимость квартиры зависит от площади, этажа, района и других параметров, но не является функцией от них. Зато можно считать её среднее функцией от этих величин. Разумеется, наблюдать это среднее значение мы не можем — в нашей власти лишь наблюдать значения результирующей случайной величины при разных значениях остальных. Эту зависимость можно воображать как вход и выход некоторой машины — «ящика с шуршавчиком». Входные данные (*факторы*) известны. На выходе мы наблюдаем результат преобразования входных данных в ящике по каким-либо правилам.

§ 1. Математическая модель регрессии

Пусть наблюдаемая случайная величина X зависит от случайной величины или случайного вектора Z. Значения Z мы либо задаём, либо наблюдаем. Обозначим через $f(t)$ функцию, отражающую зависимость среднего значения X от значений Z:

$$\mathsf{E}(X \mid Z = t) = f(t). \tag{31}$$

Функция $f(t)$ называется *линией регрессии* X *на* Z, а уравнение $x = f(t)$ — уравнением регрессии. После n экспериментов, в которых Z последовательно принимает значения $Z = t_1, \ldots, Z = t_n$, получим значения наблюдаемой величины X, равные $X_1, \ldots, X_n$. Обозначим через ε_i разницу $X_i - \mathsf{E}(X \mid Z = t_i) = X_i - f(t_i)$ между наблюдаемой в i-м эксперименте случайной величиной и её математическим ожиданием.

Итак, $X_i = f(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, где ε_i — ошибки наблюдения, равные в точности разнице между реальным и усредненным значением случайной величины X при значении $Z = t_i$. Про совместное распределение $\varepsilon_1, \ldots, \varepsilon_n$ обычно что-либо известно или предполагается: *например*, что вектор ошибок $\vec{\varepsilon}$ состоит из независимых и одинаково *нормально* распределённых случайных величин с нулевым средним. Нулевое среднее тут необходимо:

$$\mathsf{E}\varepsilon_i = \mathsf{E}X_i - f(t_i) = \mathsf{E}(X \mid Z = t_i) - \mathsf{E}(X \mid Z = t_i) = 0.$$

Требуется по значениям $t_1, \ldots, t_n$ и $X_1, \ldots, X_n$ оценить как можно точнее функцию $f(t)$. Величины t_i не являются случайными, вся случайность сосредоточена в неизвестных ошибках ε_i и в наблюдаемых X_i. Но пытаться в классе всех возможных функций восстанавливать $f(t)$ по «наилучшим оценкам» для $f(t_i)$ довольно глупо: наиболее точными приближениями к $f(t_i)$ оказываются X_i, и функция $f(t)$ будет просто ломаной, построенной по точкам (t_i, X_i). Поэтому сначала определяют вид функции $f(t)$. Часто в качестве $f(t)$ берут полином небольшой степени с неизвестными коэффициентами.

Будем пока предполагать, что функция $f(t)$ полностью определяется неизвестными параметрами $\theta_1, \ldots, \theta_k$.

Метод максимального правдоподобия. Оценки неизвестных параметров находят с помощью метода максимального правдоподобия. Он предписывает выбирать неизвестные параметры так, чтобы максимизировать функцию правдоподобия случайного вектора $X_1, \ldots, X_n$.

Будем, для простоты, предполагать, что вектор ошибок $\vec{\varepsilon}$ состоит из независимых и одинаково распределённых случайных величин с плотностью распределения $h(x)$ из некоторого семейства распределений с нулевым средним и, вообще говоря, неизвестной дисперсией. Обычно полагают, что ε_i имеют симметричное распределение — нормальное N_{0,σ^2}, Стьюдента, Лапласа и т. п. Поскольку X_i от ε_i зависят линейно, то распределение X_i окажется таким же, как у ε_i, но с центром уже не в нуле, а в точке $f(t_i)$.

Поэтому X_i имеет плотность $h\big(x - f(t_i)\big)$. Функция правдоподобия вектора $X_1, \ldots, X_n$ в силу независимости координат равна

$$f(\vec{X};\, \theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} h\big(X_i - f(t_i)\big) = h(\varepsilon_1) \cdot \ldots \cdot h(\varepsilon_n). \qquad (32)$$

Если величины ε_i имеют разные распределения, то h следует заменить на соответствующие h_i. Для зависимых ε_i произведение плотностей в формуле (32) заменится плотностью их совместного распределения.

Метод максимального правдоподобия предписывает находить оценки неизвестных параметров θ_i функции $f(t)$ и оценки неизвестной дисперсии $\sigma^2 = \mathsf{D}\varepsilon_i$, максимизируя по этим параметрам функцию правдоподобия (32). Рассмотрим, во что превращается метод максимального правдоподобия в наиболее частых на практике предположениях.

Метод наименьших квадратов. Предположим, что вектор ошибок $\vec{\varepsilon}$ состоит из независимых случайных величин с нормальным распределением N_{0,σ^2}. Функция правдоподобия (32) имеет вид

$$f\left(\vec{X};\vec{\theta}\right) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(X_i - f(t_i))^2}{2\sigma^2}\right\} =$$
$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - f(t_i))^2\right\}.$$

Очевидно, что при любом фиксированном σ^2 максимум функции правдоподобия достигается при наименьшем значении суммы квадратов ошибок

$$\sum (X_i - f(t_i))^2 = \sum \varepsilon_i^2.$$

Определение 32. *Оценкой метода наименьших квадратов* (ОМНК) для неизвестных параметров $\theta_1, \ldots, \theta_k$ уравнения регрессии называется набор значений параметров, доставляющий минимум сумме квадратов отклонений

$$\sum_{i=1}^{n} (X_i - f(t_i))^2 = \sum_{i=1}^{n} \varepsilon_i^2.$$

Найдя оценки для θ_i, найдём тем самым оценку $\hat{f}(t)$ для $f(t)$. Обозначим через $\hat{f}(t_i)$ значения этой функции, и через $\hat{\varepsilon}_i = X_i - \hat{f}(t_i)$ соответствующие оценки ошибок. Оценка максимального правдоподобия для σ^2, она же точка максимума по σ^2 функции правдоподобия, равна

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{f}(t_i))^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2. \tag{33}$$

Мудрый читатель понял, что основная цель рассмотренного выше примера — показать, что метод наименьших квадратов не падает с неба, а есть в точности метод максимального правдоподобия в том, например, случае, когда вектор ошибок, а вместе с ним и вектор наблюдаемых откликов регрессии, состоит из независимых и одинаково распределённых случайных величин с *нормальным распределением*.

Пример 41. Пусть независимые случайные величины ε_i имеют распределение Лапласа с плотностью распределения

$$h(x) = \frac{1}{2\sigma}\ \exp\left\{-\frac{|x|}{\sigma}\right\}.$$

Тогда при любом фиксированном σ^2 максимум функции правдоподобия достигается при наименьшем значении суммы $\sum|X_i - f(t_i)|$ *абсолютных отклонений*. Оценка максимального правдоподобия (ОМП) для набора $\theta_1, \ldots, \theta_k$ уже не есть ОМНК. Даже для самой простой функции $f(t)$ эти методы приводят к разным оценкам.

Упражнение. Пусть функция $f(t) = \theta$ постоянна, а ошибки ε_i взяты из распределения Лапласа. Покажите, что оценкой максимального правдоподобия для θ, минимизирующей $\sum|X_i - \theta|$, является выборочная медиана

$$\hat\theta = \begin{cases} X_{(m)}, & \text{если } n = 2m{-}1 \text{ (нечётно)},\\ \frac{1}{2}\left(X_{(m)} + X_{(m+1)}\right), & \text{если } n = 2m \text{ (чётно)}. \end{cases}$$

Вместо полусуммы можно брать любую точку отрезка $[X_{(m)}, X_{(m+1)}]$. ОМП для дисперсии равна $\hat\sigma^2 = \frac{1}{n}\sum|X_i - \hat\theta|$. Покажите, что ОМНК для θ в той же ситуации равна $\overline{X}$, а оценка для σ^2 равна выборочной дисперсии S^2 (см. также пример 42 ниже).

Найдём ОМНК для функций $f(t)$ в ряде частных случаев.

Пример 42. Пусть функция $f(t) = \theta$ — постоянная, θ — неизвестный параметр. Тогда наблюдения равны $X_i = \theta + \varepsilon_i$, $i = 1, \ldots, n$. Легко узнать задачу оценивания неизвестного математического ожидания θ по выборке из независимых и одинаково распределённых случайных величин $X_1, \ldots, X_n$. Найдём ОМНК $\hat\theta$ для параметра θ:

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n}(X_i - \theta)^2 = -2\sum_{i=1}^{n}(X_i - \theta)\bigg|_{\theta=\hat\theta} = 0 \ \text{ при } \ \hat\theta = \overline{X}.$$

Трудно назвать этот ответ неожиданным. Соответственно, $\hat\sigma^2 = S^2$.

Пример 43 (линейная регрессия). Рассмотрим линейную регрессию $X_i = \theta_1 + t_i\theta_2 + \varepsilon_i$, $i = 1, \ldots, n$, где θ_1 и θ_2 — неизвестные параметры. Здесь $f(t) = \theta_1 + t\theta_2$ — прямая.

Найдём оценку метода наименьших квадратов $\hat{\theta}_1$, $\hat{\theta}_2$, на которой достигается минимум величины $\sum \varepsilon_i^2 = \sum (X_i - \theta_1 - t_i\theta_2)^2$. Приравняв к нулю частные производные этой суммы по параметрам, найдём точку экстремума.

Упражнение. Убедиться, что решением системы уравнений

$$\frac{\partial}{\partial\theta_1}\sum_{i=1}^{n}\varepsilon_i^2 = 0\,, \qquad \frac{\partial}{\partial\theta_2}\sum_{i=1}^{n}\varepsilon_i^2 = 0$$

является пара

$$\hat{\theta}_2 = \frac{\frac{1}{n}\sum X_i t_i - \overline{X}\cdot\overline{t}}{\frac{1}{n}\sum (t_i - \overline{t})^2}\,, \qquad \hat{\theta}_1 = \overline{X} - \overline{t}\,\hat{\theta}_2.$$

Определение 33. *Выборочным коэффициентом корреляции* называется величина

$$\rho^* = \frac{\frac{1}{n}\sum X_i t_i - \overline{X}\cdot\overline{t}}{\sqrt{\frac{1}{n}\sum (t_i - \overline{t})^2 \cdot \frac{1}{n}\sum (X_i - \overline{X})^2}}\,,$$

которая характеризует степень линейной зависимости между наборами чисел $X_1, \ldots, X_n$ и $t_1, \ldots, t_n$.

Пример 44. Термин «регрессия» ввёл Гальтон (*Francis Galton.* Regression towards mediocrity in hereditary stature // Journal of the Anthropological Institute. — 1886. — v. 15. — p. 246—265).

Гальтон исследовал, в частности, рост детей высоких родителей и установил, что он «регрессирует» в среднем, т. е. в среднем дети высоких родителей не так высоки, как их родители. Пусть X — рост сына, а Z_1 и Z_2 — рост отца и матери. Для линейной модели регрессии

$$\mathsf{E}(X \mid Z_1 = t, Z_2 = u) = f(t, u) = \theta_1 t + \theta_2 u + c$$

Гальтон нашел оценки параметров

$$\mathsf{E}(\text{роста сына} \mid Z_1 = t, Z_2 = u) = 0,27t + 0,2u + \mathrm{const},$$

а средний рост дочери ещё в 1,08 раз меньше. Независимо от добавочной постоянной суммарный вклад высокого роста родителей в рост детей не превышает половины. Остальное — неизменная добавка.

§ 2. Общая модель линейной регрессии

Введём вектор факторов регрессии $\vec{Z} = (Z_1, \ldots, Z_k)$ и вектор неизвестных параметров регрессии $\vec{\beta} = (\beta_1, \ldots, \beta_k)$. Каждый вектор есть вектор-столбец, а изображён по горизонтали для удобства. Рассматривается *простая* (линейная) регрессия

$$\mathsf{E}\big(X \mid \vec{Z} = \vec{t}\,\big) = f\big(\vec{t}\,\big) = \beta_1\, t_1 + \ldots + \beta_k\, t_k,$$

или, равносильно,

$$\mathsf{E}(X \mid \vec{Z}\,) = f(\vec{Z}\,) = \beta_1\, Z_1 + \ldots + \beta_k\, Z_k.$$

Пусть в i-м эксперименте факторы регрессии принимают заранее заданные значения

$$\vec{Z}^{(i)} = \big(Z_1^{(i)}, \ldots, Z_k^{(i)}\big), \quad \text{где} \quad i = 1, \ldots, n.$$

После $n \geqslant k$ экспериментов получен набор откликов $X_1, \ldots, X_n$:

$$\begin{cases} X_1 = \beta_1\, Z_1^{(1)} + \ldots + \beta_k\, Z_k^{(1)} + \varepsilon_1 \\ X_2 = \beta_1\, Z_1^{(2)} + \ldots + \beta_k\, Z_k^{(2)} + \varepsilon_2 \\ \ldots \\ X_n = \beta_1\, Z_1^{(n)} + \ldots + \beta_k\, Z_k^{(n)} + \varepsilon_n, \end{cases}$$

или, в матричной форме, $\vec{X} = Z^T\vec{\beta} + \vec{\varepsilon}$, с *матрицей плана* $Z(k \times n)$

$$Z = \begin{pmatrix} Z_1^{(1)} & \ldots & Z_1^{(n)} \\ \vdots & \ldots & \vdots \\ Z_k^{(1)} & \ldots & Z_k^{(n)} \end{pmatrix} = \big(\vec{Z}^{(1)} \; \ldots \; \vec{Z}^{(n)}\big).$$

Требуется по данным матрице плана Z и вектору $\vec{X}$ найти оценки для параметров регрессии $\vec{\beta}$ и параметров распределения вектора ошибок $\vec{\varepsilon}$.

МНК и нормальное уравнение. Будем считать в дальнейшем выполненным следующее условие.

(A1) Матрица Z имеет ранг k, т. е. все её строки линейно независимы.

Лемма 9. *Предположение* (A1) *означает, что симметричная матрица* $A = ZZ^T$ *положительно определена.*

Доказательство. Напомним, что матрица $A(k \times k)$ называется положительно определённой, если неотрицательна квадратичная форма $\vec{t}^{\,T}A\vec{t} \geqslant 0$ для любого вектора $\vec{t} = (t_1, \ldots, t_k)$, причём равенство $\vec{t}^{\,T}A\vec{t} = 0$ возможно только для $\vec{t} = \vec{0} = (0, \ldots, 0)$. Напомним также, что квадрат нормы вектора $\vec{u}$ равен

$$\|\vec{u}\|^2 = \vec{u}^T\vec{u} = \sum u_i^2 \geqslant 0.$$

Норма равна нулю, если и только если $\vec{u} = \vec{0}$.

Матрица A симметрична, поскольку $A = ZZ^T$ и $A^T = A$. Её неотрицательная определённость имеет место и без предположения (A1):

$$\vec{t}^{\,T} A \vec{t} = \vec{t}^{\,T} Z{\cdot}Z^T\vec{t} = (Z^T\vec{t}\,)^T \cdot (Z^T\vec{t}\,) = \|Z^T\vec{t}\|^2 \geqslant 0.$$

Равенство же $\|Z^T\vec{t}\| = 0$ возможно только если $Z^T\vec{t} = \vec{0}$. Но ранг Z равен k, поэтому $Z^T\vec{t} = \vec{0}$ влечёт $\vec{t} = \vec{0}$. □

Скоро нам пригодится корень из матрицы A, существование которого гарантирует следующее утверждение.

Лемма 10. *Положительная определённость и симметричность матрицы A влекут существование вещественной симметричной матрицы $\sqrt{A}$ такой, что $\sqrt{A}\sqrt{A} = A$.*

Существование матрицы $\sqrt{A}$ с нужными свойствами следует из возможности привести симметричную матрицу A ортогональными преобразованиями $A = Q^T D\, Q$ к диагональному виду с положительными, в силу положительной определённости, собственными значениями A на диагонали матрицы D. Тогда $\sqrt{A} = Q^T \sqrt{D}\, Q$. □

Найдём ОМНК $\hat{\beta}$, которая минимизирует функцию $S(\vec{\beta})$, равную

$$S(\vec{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \|\vec{\varepsilon}\,\|^2 = \|\vec{X} - Z^T\vec{\beta}\,\|^2 = (\vec{X} - Z^T\vec{\beta})^T \cdot (\vec{X} - Z^T\vec{\beta}).$$

Можно искать точку экстремума дифференцированием по β_i. Заметим вместо этого, что величина $S(\vec{\beta})$ есть квадрат расстояния от точки $\vec{X} \in \mathbb{R}^n$ до точки $Z^T\vec{\beta}$ — одной из точек линейного подпространства (гиперплоскости) в $\mathbb{R}^n$, в которой лежит любой вектор вида $Z^T\vec{t}$, где $\vec{t} \in \mathbb{R}^k$.

Минимальное расстояние $S(\hat{\beta})$ мы получим, когда вектор $\vec{X} - Z^T\hat{\beta}$ будет ортогонален всем векторам этого подпространства, т. е. когда для любого $\vec{t} \in \mathbb{R}^k$ скалярное произведение векторов $Z^T\vec{t}$ и $\vec{X} - Z^T\hat{\beta}$ обратится в нуль. Запишем это скалярное произведение в матричном виде

$$\left(Z^T\vec{t}, \vec{X} - Z^T\hat{\beta}\right) = \left(Z^T\vec{t}\right)^T \left(\vec{X} - Z^T\hat{\beta}\right) = \vec{t}^{\,T} \cdot \left(Z\vec{X} - ZZ^T\hat{\beta}\right) = 0.$$

Подставив в это равенство в качестве $\vec{t}$ поочерёдно базисные векторы $(0, \ldots, 0, 1, 0, \ldots, 0)$ из $\mathbb{R}^k$, сразу же получим, что все координаты вектора $Z\vec{X} - ZZ^T\hat{\beta}$ равны нулю. Итак, ОМНК $\hat{\beta}$ есть любое решение уравнения

$$ZZ^T\hat{\beta} = Z\vec{X} \quad \text{или} \quad A\hat{\beta} = Z\vec{X}. \tag{34}$$

По лемме 9, уравнение (34) имеет единственное решение

$$\hat{\beta} = A^{-1}Z\vec{X} \tag{35}$$

в том и только в том случае, когда матрица $Z(k \times n)$ имеет полный ранг k, где $k \leqslant n$. Уравнение (34) называется *нормальным уравнением*.

В предположении, что вектор ошибок $\vec{\varepsilon}$ состоит из независимых случайных величин с нормальным распределением N_{0,σ^2} с одной и той же дисперсией, ОМНК совпадает с оценкой максимального правдоподобия, а ОМП для σ^2, согласно (33), равна

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = \frac{1}{n}\|\vec{X} - Z^T\hat{\beta}\|^2 = \frac{1}{n}S(\hat{\beta}). \tag{36}$$

Свойства ОМНК. Сначала докажем несколько простых свойств, которые нам понадобятся в дальнейшем.

Свойство 12. Разность $\hat{\beta} - \vec{\beta}$ равна $A^{-1}Z\vec{\varepsilon}$.

Доказательство. Подставим в разность вместо $\hat{\beta}$ решение (35):

$$A^{-1}Z\vec{X} - \vec{\beta} = A^{-1}Z(Z^T\vec{\beta} + \vec{\varepsilon}) - \vec{\beta} = A^{-1}A\vec{\beta} + A^{-1}Z\vec{\varepsilon} - \vec{\beta} = A^{-1}Z\vec{\varepsilon}. \quad \square$$

Свойство 13. Если $\mathsf{E}\vec{\varepsilon} = 0$, то $\hat{\beta}$ — несмещённая оценка для $\vec{\beta}$.

Доказательство. Действительно, по предыдущему свойству

$$\mathsf{E}\hat{\beta} = \vec{\beta} + A^{-1}Z\,\mathsf{E}\vec{\varepsilon} = \vec{\beta}. \qquad \square$$

Дальнейшие свойства требуют знания распределения вектора ошибок. Пусть выполнены предположение (A1) и следующее предположение (A2).

(A2) Вектор $\vec{\varepsilon}$ состоит из независимых случайных величин с распределением N_{0,σ^2} с одной и той же дисперсией.

Напомним, что для произвольного случайного вектора $\vec{x}$, координаты которого имеют вторые моменты, матрицей ковариаций

$$\mathsf{D}\vec{x} = \mathsf{E}(\vec{x} - \mathsf{E}\vec{x})(\vec{x} - \mathsf{E}\vec{x})^T$$

называется матрица, (i, j)-й элемент которой равен

$$\operatorname{cov}(x_i,\, x_j) = \mathsf{E}(x_i - \mathsf{E}x_i)(x_j - \mathsf{E}x_j).$$

В частности, $\mathsf{D}\vec{\varepsilon} = \sigma^2 E_n$, где E_n — единичная $(n{\times}n)$-матрица.

Следующее очень важное свойство утверждает, что в предположениях (A1)—(A2) вектор $\sqrt{A}\hat{\beta}$ имеет диагональную матрицу ковариаций.

Свойство 14. Матрица ковариаций вектора $\sqrt{A}\hat{\beta}$ равна $\sigma^2 E_k$.

Доказательство. Воспользуемся свойством 12 и вычислим матрицу ковариаций вектора $\sqrt{A}\hat{\beta}$:

$$\begin{aligned}\mathsf{D}\sqrt{A}\hat{\beta} &= \mathsf{E}\left(\sqrt{A}\hat{\beta}-\mathsf{E}\sqrt{A}\hat{\beta}\right)\left(\sqrt{A}\hat{\beta}-\mathsf{E}\sqrt{A}\hat{\beta}\right)^T = \\ &= \mathsf{E}\left(\sqrt{A}(\hat{\beta}-\vec{\beta})\right)\left(\sqrt{A}(\hat{\beta}-\vec{\beta})\right)^T = \mathsf{E}\left(\sqrt{A}A^{-1}Z\vec{\varepsilon}\right)\left(\sqrt{A}A^{-1}Z\vec{\varepsilon}\right)^T = \\ &= \sqrt{A}A^{-1}Z\,\mathsf{E}\left(\vec{\varepsilon}\,\vec{\varepsilon}^T\right)Z^T(A^{-1})^T\sqrt{A}^T.\end{aligned}$$

И так как $A^T = A$, $\mathsf{E}\vec{\varepsilon}\,\vec{\varepsilon}^T = \sigma^2 E_n$, то

$$\mathsf{D}\sqrt{A}\hat{\beta} = \sigma^2\cdot\sqrt{A}A^{-1}ZZ^TA^{-1}\sqrt{A} = \sigma^2 E_k. \qquad \square$$

Свойство 14 означает, что координаты вектора $\sqrt{A}\hat{\beta}$ некоррелированы. Сформулируем дальнейшее следствие этого свойства первым пунктом следующей теоремы. С утверждениями второго и третьего пунктов читатель встретится в следующем семестре многократно.

Теорема 31. *Пусть выполнены предположения* (A1)—(A2). *Тогда*

1) *вектор* $\frac{1}{\sigma}\sqrt{A}(\hat{\beta}-\vec{\beta})$ *имеет* k*-мерное стандартное нормальное распределение, т. е. состоит из* k *независимых случайных величин со стандартным нормальным распределением;*

2) *величина* $\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\|\vec{X}-Z^T\hat{\beta}\|^2$ *имеет распределение* χ^2 *с* $n-k$ *степенями свободы и не зависит от* $\hat{\beta}$;

3) *исправленная оценка* $(\sigma^2)^* = \frac{n\hat{\sigma}^2}{n-k} = \frac{1}{n-k}\|\vec{X}-Z^T\hat{\beta}\|^2$ *является несмещённой оценкой для* σ^2.

Доказательство. Первое свойство вытекает из того, что вектор

$$\sqrt{A}(\hat{\beta}-\vec{\beta}) = \sqrt{A}A^{-1}Z\vec{\varepsilon} = (\sqrt{A})^{-1}Z\vec{\varepsilon}$$

является линейным преобразованием нормального вектора $\vec{\varepsilon}$ и поэтому имеет нормальное совместное распределение. По свойству 14, матрица ковариаций этого вектора есть $\sigma^2 E_k$, поэтому матрица ковариаций нормированного вектора $\sqrt{A}(\hat{\beta}-\vec{\beta})/\sigma$ есть просто E_k, а математическое ожидание равно нулю по свойству 13.

Координаты *многомерного нормального вектора* независимы тогда и только тогда, когда они некоррелированы. Подробнее этот факт обсуждается в следующей главе. Первое утверждение теоремы доказано.

Докажем второе. По построению ОМНК, вектор $X - Z^T\hat{\beta}$ ортогонален любому вектору вида Z^Tt. В частности, он ортогонален имеющему нужный вид вектору $Z^T(\hat{\beta} - \vec{\beta})$. По теореме Пифагора в треугольнике с катетами $X - Z^T\hat{\beta}$ и $Z^T(\hat{\beta} - \vec{\beta})$ сумма квадратов их длин равна квадрату длины гипотенузы:

$$\|X - Z^T\hat{\beta}\|^2 + \|Z^T(\hat{\beta} - \vec{\beta})\|^2 = \|X - Z^T\hat{\beta} + Z^T(\hat{\beta} - \vec{\beta})\|^2 = \|X - Z^T\vec{\beta}\|^2.$$

Поэтому

$$\|X - Z^T\hat{\beta}\|^2 = \|X - Z^T\vec{\beta}\|^2 - \|Z^T(\hat{\beta} - \vec{\beta})\|^2 = \|\vec{\varepsilon}\|^2 - \|Z^T(\hat{\beta} - \vec{\beta})\|^2. \quad (37)$$

Но квадрат нормы $\|Z^T(\hat{\beta} - \vec{\beta})\|^2$ равен квадрату нормы $\|\sqrt{A}(\hat{\beta} - \vec{\beta})\|^2$:

$$\|Z^T(\hat{\beta} - \vec{\beta})\|^2 = (\hat{\beta} - \vec{\beta})^T Z Z^T (\hat{\beta} - \vec{\beta}) = (\hat{\beta} - \vec{\beta})^T \sqrt{A}^T \sqrt{A}(\hat{\beta} - \vec{\beta}) =$$
$$= \|\sqrt{A}(\hat{\beta} - \vec{\beta})\|^2 = \|(\sqrt{A})^{-1} Z\vec{\varepsilon}\|^2.$$

Осталось заметить, что *строки* $(k{\times}n)$-матрицы $(\sqrt{A})^{-1}Z$ ортогональны:

$$\left((\sqrt{A})^{-1}Z\right)\left((\sqrt{A})^{-1}Z\right)^T = (\sqrt{A})^{-1} Z Z^T (\sqrt{A})^{-1} = E_k,$$

поэтому k её строк можно дополнить до некоторой ортогональной матрицы $C(n{\times}n)$. Первые k координат n-мерного вектора $\vec{Y} = C\,\vec{\varepsilon}/\sigma$ совпадают с вектором $(\sqrt{A})^{-1} Z\,\vec{\varepsilon}/\sigma$. В результате из (37) получим

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\|X - Z^T\hat{\beta}\|^2 = \|\vec{\varepsilon}/\sigma\|^2 - \|(\sqrt{A})^{-1} Z\,\vec{\varepsilon}/\sigma\|^2 =$$
$$= \sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2 - Y_1^2 - \ldots - Y_k^2. \quad (38)$$

Но вектор $\vec{\varepsilon}/\sigma$ имеет n-мерное стандартное нормальное распределение. Тогда вся разность (38) по лемме Фишера имеет распределение χ^2 с $n{-}k$ степенями свободы и не зависит от вычитаемого, т. е. от случайного вектора $\vec{\varepsilon}$ (и от $\hat{\beta}$ тоже, поскольку $\hat{\beta}$ есть линейная функция от $\vec{\varepsilon}$).

Третье утверждение теоремы сразу следует из второго. Напомним, что распределение χ^2 с $n - k$ степенями свободы имеет математическое ожидание $n - k$. Поэтому

$$\mathsf{E}(\sigma^2)^* = \mathsf{E}\left(\frac{n\hat{\sigma}^2}{n{-}k}\right) = \frac{\sigma^2}{n{-}k}\,\mathsf{E}\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) = \frac{\sigma^2}{n{-}k}\cdot(n-k) = \sigma^2,$$

что доказывает третье утверждение теоремы. □

ГЛАВА X

МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

В предыдущих главах мы неоднократно встречались с нормальными выборками. Но применения многомерного нормального распределения в математической статистике не ограничиваются лишь свойствами наборов независимых нормальных случайных величин. Скажем, зависимость между собой нормальных ошибок регрессии приводит к необходимости преобразований, устраняющих эту зависимость. Поэтому вернёмся к многомерному нормальному распределению и изучим его свойства подробнее, чем в теории вероятностей. Затем мы используем наши знания для доказательства теоремы Пирсона. В конце главы рассмотрим модель однофакторного дисперсионного анализа, для изучения которой вновь понадобится лемма Фишера.

§ 1. Свойства нормальных векторов

Напомним полное умолчаний определение, данное нами в прошлом семестре, и наведём в нём порядок.

Определение 34. Пусть случайный вектор $\vec{\xi} = (\xi_1, \ldots, \xi_m)$ имеет вектор средних $\vec{a} = \mathsf{E}\vec{\xi}$, пусть Σ — симметричная, невырожденная, положительно определённая матрица. Говорят, что вектор $\vec{\xi}$ имеет нормальное распределение $\mathrm{N}_{\vec{a},\Sigma}$ в $\mathbb{R}^m$, если плотность этого вектора при всех $\vec{x} \in \mathbb{R}^m$ равна

$$f_{\vec{\xi}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^m \sqrt{|\det\Sigma|}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{a})^T \Sigma^{-1} (\vec{x}-\vec{a})\right\}, \qquad (39)$$

Покажем, что матрица коварияций случайного вектора, имеющего плотность распределения (39), в точности равна Σ. Для вычисления «дисперсии» $\mathsf{D}\vec{\xi}$ поступим так же, как в одномерном случае: свяжем вектор $\vec{\xi}$ с вектором $\vec{\eta}$, имеющим многомерное стандартное нормальное распределение, а затем воспользуемся свойствами дисперсий.

Пусть вектор $\vec{\eta}$ состоит из m независимых стандартных нормальных случайных величин и имеет плотность распределения

$$f_{\vec{\eta}}(\vec{x}\,) = \frac{1}{(\sqrt{2\pi})^m} \exp\left\{-\frac{1}{2}\vec{x}^{\,T}\vec{x}\right\}.$$

Матрица Σ симметрична, невырождена и положительно определена. По лемме 10 (с. 119), существует симметричная матрица $B = \sqrt{\Sigma}$. Положим $\vec{\xi} = B\vec{\eta} + \vec{a}$ и найдём распределение этого вектора. По теореме 18 (с. 74), плотность распределения вектора $\vec{\xi}$ равна

$$f_{\vec{\xi}}(\vec{x}) = |\det B|^{-1} \cdot f_{\vec{\eta}}\left(B^{-1}(\vec{x}-\vec{a})\right) =$$
$$= \frac{1}{(\sqrt{2\pi})^m\sqrt{|\det\Sigma|}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{a})^T(B^{-1})^T B^{-1}(\vec{x}-\vec{a})\right\}. \quad (40)$$

В показателе экспоненты в равенстве (40) стоит матрица

$$(B^{-1})^T B^{-1} = B^{-1}B^{-1} = \Sigma^{-1}.$$

Итак, вектор $\vec{\xi}$ имеет плотность распределения, заданную формулой (39). Найдём матрицу коваріаций этого вектора. Вектор его математических ожиданий есть $\vec{a}$. Действительно, $\mathsf{E}(B\vec{\eta}+\vec{a}) = B\mathsf{E}\vec{\eta}+\vec{a} = \vec{a}$, поэтому сразу вычтем его:

$$\mathsf{D}\vec{\xi} = \mathsf{E}\left(B\vec{\eta}\,(B\vec{\eta}\,)^T\right) = B\cdot\mathsf{E}\left(\vec{\eta}\,\vec{\eta}^{\,T}\right)\cdot B^T = BB^T = \Sigma.$$

Мы воспользовались тем, что $\mathsf{E}\left(\vec{\eta}\,\vec{\eta}^{\,T}\right) = \mathsf{D}\vec{\eta} = E$ — единичная матрица.

Сформулируем результат предыдущих рассмотрений в виде теоремы.

Теорема 32. 1. *Пусть вектор* $\vec{\eta}$ *состоит из независимых стандартных нормальных случайных величин,* B *— невырожденная матрица. Тогда вектор* $\vec{\xi} = B\vec{\eta} + \vec{a}$ *имеет многомерное нормальное распределение* $\mathrm{N}_{\vec{a},\Sigma}$ *с вектором средних* $\vec{a}$ *и матрицей коваріаций* $\Sigma = BB^T$.

2. *Пусть вектор* $\vec{\xi}$ *имеет многомерное нормальное распределение* $\mathrm{N}_{\vec{a},\Sigma}$, *где* $\Sigma > 0$. *Тогда вектор* $\vec{\eta} = B^{-1}\left(\vec{\xi}-\vec{a}\right)$ *при* $B = \sqrt{\Sigma}$ *состоит из независимых стандартных нормальных случайных величин.*

Итак, имеет место замечательный факт: любой нормальный вектор в $\mathbb{R}^m$ со сколь угодно зависимыми координатами, имеющий плотность совместного распределения, может быть умножением на подходящую невырожденную матрицу превращён в вектор, состоящий из независимых нормальных случайных величин. И наоборот, стандартный нормальный случайный вектор можно линейным преобразованием превратить в вектор с заданным многомерным нормальным распределением.

Напомним, что по теореме 19 (с. 75) поворот, т. е. умножение на *ортогональную* матрицу, не меняет совместного распределения координат стандартного нормального вектора. А что можно с помощью поворота сделать с произвольным нормальным вектором? Проверьте справедливость следующей леммы.

Лемма 11. *Если вектор* $\vec{\xi}$ *имеет матрицу ковариаций* $\Sigma = \mathsf{D}\vec{\xi}$, *то матрица ковариаций вектора* $\vec{\eta} = B\vec{\xi}$ *равна* $\mathsf{D}\vec{\eta} = B\Sigma B^T$.

Однако симметричную положительно определённую матрицу Σ можно ортогональными преобразованиями привести к диагональному виду:

$$Q\Sigma Q^T = D = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_m^2).$$

Вот почему любой нормальный вектор $\vec{\xi}$ подходящим поворотом $\vec{\eta} = Q\vec{\xi}$ можно превратить в вектор с независимыми, но не обязательно одинаково распределёнными координатами.

Замечание 25. Бывает удобно считать нормальным распределение вектора с вырожденной матрицей ковариаций, если поворотом этот вектор можно превратить в нормальный вектор «меньшей размерности», т. е. если $Q\vec{\xi} = (\eta_1, \ldots, \eta_k, c_{k+1}, \ldots, c_m)$, где первые k координат имеют нормальное совместное распределение в $\mathbb{R}^k$.

Замечание 26. Вектор, составленный из нормальных случайных величин, не обязательно имеет многомерное нормальное распределение. Так, для $\xi \in \mathrm{N}_{0,1}$ вектор $(\xi, c\xi)$ имеет вырожденную матрицу ковариаций $\left(\begin{smallmatrix} 1 & c \\ c & c^2 \end{smallmatrix}\right)$ и не имеет плотности в $\mathbb{R}^2$.

Поэтому в условиях следующей теоремы существование совместной нормальной плотности координат вектора обязательно.

Теорема 33. *Пусть вектор* $\vec{\xi}$ *имеет многомерное нормальное распределение* $\mathrm{N}_{\vec{a},\Sigma}$ *с плотностью, заданной равенством* (39). *Координаты этого вектора независимы тогда и только тогда, когда они некоррелированы, т. е. когда матрица ковариаций* Σ *диагональна.*

Доказательство. Только в случае диагональной матрицы Σ с элементами $\Sigma_{ii} = \sigma_i^2 = \mathsf{D}\xi_i$ квадратичная форма $(\vec{x} - \vec{a})^T \Sigma^{-1} (\vec{x} - \vec{a})$ превращается в сумму квадратов

$$\sum_{i,j} (x_i - a_i) \cdot (\Sigma^{-1})_{ij} \cdot (x_j - a_j) = \sum_i \frac{(x_i - a_i)^2}{\sigma_i^2},$$

и плотность (39) распадается в произведение плотностей координат. □

Пример 45. Очень часто утверждение предыдущей теоремы трактуют следующим образом: если нормально распределённые случайные величины некоррелированы, то они независимы. Это утверждение неверно: некоррелированные нормальные величины могут быть зависимы, и даже функционально зависимы, если их совместное распределение не обладает плотностью (39).

Пусть $\xi \Subset \mathrm{N}_{0,1}$. Построим при некотором $c > 0$ случайную величину

$$\eta = \begin{cases} -\xi, & \text{если } |\xi| \leqslant c, \\ \xi & \text{иначе.} \end{cases}$$

Легко проверить, что $\eta \Subset \mathrm{N}_{0,1}$ для любого c. Ещё проще проверяется зависимость ξ и η. Например, события $\{\eta \in (-c, 0)\}$ и $\{\xi \in (-c, 0)\}$ несовместны. Но ковариация

$$\mathrm{cov}(\xi, \eta) = \frac{2}{\sqrt{2\pi}} \int\limits_{c}^{\infty} x^2 e^{-x^2/2} dx - \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{c} x^2 e^{-x^2/2} dx$$

выбором числа c может быть сделана нулевой.

Так же как и одномерное, многомерное нормальное распределение часто возникает как предел распределений центрированных и нормированных сумм. Только теперь это должны быть суммы независимых случайных *векторов*. Читатель поверит следующей теореме без доказательства, будучи в состоянии доказать её, как и в одномерном случае, с помощью характеристических функций.

Теорема 34 (многомерная ЦПТ). *Пусть* $\vec{\xi}^{(1)}, \vec{\xi}^{(2)}, \ldots$ — *последовательность независимых и одинаково распределённых случайных векторов, каждый из которых имеет среднее* $\mathsf{E}\,\vec{\xi}^{(1)} = \vec{a}$ *и невырожденную матрицу ковариаций* Σ. *Обозначим через* $S_n = \vec{\xi}^{(1)} + \cdots + \vec{\xi}^{(n)}$ *вектор частичных сумм. Тогда при* $n \to \infty$ *имеет место слабая сходимость распределений векторов*

$$\vec{\eta}^{(n)} = \frac{S_n - n\vec{a}}{\sqrt{n}} \Rightarrow \vec{\eta}, \text{ где } \vec{\eta} \text{ имеет распределение } \mathrm{N}_{\vec{0},\Sigma}.$$

В условиях многомерной ЦПТ распределение любых непрерывных функций $g\big(\vec{\eta}^{(n)}\big)$ слабо сходится к распределению $g(\vec{\eta}\,)$. В качестве $g(\vec{x}\,)$ нам в дальнейшем будет нужна только $g(\vec{x}\,) = \sum x_i^2 = \|\vec{x}\,\|^2$.

Следствие 2. *В условиях многомерной ЦПТ имеет место сходимость* $\|\vec{\eta}^{(n)}\|^2 \Rightarrow \|\vec{\eta}\,\|^2$.

§ 2. Доказательство теоремы Пирсона

План действий. 1. Сначала покажем, что статистика ρ критерия χ^2, участвующая в теореме Пирсона, есть квадрат нормы некоторого k-мерного вектора $\vec{\eta}^{(n)} = \dfrac{S_n - n\vec{a}}{\sqrt{n}}$. Затем убедимся в том, что матрица ковариаций типичного слагаемого $\vec{\xi}^{(1)}$ в сумме S_n вырождена, что мешает использовать ЦПТ.

2. Найдём поворот Q, приводящий $\vec{\xi}^{(1)}$ к виду $Q\,\vec{\xi}^{(1)} = (\hat{\xi}^{(1)}, 1)$, где вектор $\hat{\xi}^{(1)} \in \mathbb{R}^{k-1}$ уже имеет невырожденную и даже единичную матрицу ковариаций. При том же повороте центрированный вектор $\vec{\eta}^{(n)}$ перейдёт в вектор $Q\,\vec{\eta}^{(n)} = (\hat{\eta}^{(n)}, 0)$ с нулевой последней координатой. Но его норма не изменится из-за ортогональности матрицы Q.

3. К вектору сумм $\hat{\eta}^{(n)}$ примéним многомерную ЦПТ. В пределе получим $(k{-}1)$-мерный нормальный вектор с нулевым средним и единичной матрицей ковариаций, т. е. составленный из независимых величин со стандартным нормальным распределением. Воспользуемся следствием 2 и тем, что квадрат нормы этого вектора имеет распределение H_{k-1}.

Реализация. 1. Введём вспомогательный вектор-столбец из постоянных $P = (\sqrt{p_1}, \ldots, \sqrt{p_k}\,) \in \mathbb{R}^k$. С каждым элементом выборки X_i свяжем свой вектор-столбец $\vec{\xi}^{(i)}$, где $i = 1, \ldots, n$, так:

$$\vec{\xi}^{(i)} = (\xi_1, \ldots, \xi_k) = \left(\frac{\mathrm{I}(X_i \in A_1)}{\sqrt{p_1}}, \ldots, \frac{\mathrm{I}(X_i \in A_k)}{\sqrt{p_k}}\right).$$

Получим n независимых и одинаково распределённых векторов. Видим, что $\vec{a} = \mathsf{E}\,\vec{\xi}^{(1)} = P$, поскольку $\mathsf{E}\mathrm{I}(X_1 \in A_j) = p_j$ для любого $j = 1, \ldots, k$. Складывая векторы $\vec{\xi}^{(i)}$ по i от единицы до n, центрируя и нормируя, получаем величину из многомерной ЦПТ

$$\left(\frac{\nu_1 - np_1}{\sqrt{np_1}}, \ldots, \frac{\nu_k - np_k}{\sqrt{np_k}}\right) = \frac{\sum_{i=1}^n \vec{\xi}^{(i)} - n\vec{a}}{\sqrt{n}} = \frac{S_n - n\vec{a}}{\sqrt{n}}.$$

Найдём матрицу ковариаций вектора $\vec{\xi}^{(1)}$, составленную из элементов

$$\sigma_{ij} = \operatorname{cov}\left(\frac{\mathrm{I}(X_1 \in A_i)}{\sqrt{p_i}}, \frac{\mathrm{I}(X_1 \in A_j)}{\sqrt{p_j}}\right) = \frac{\mathsf{E}\mathrm{I}(X_1 \in A_i,\, X_1 \in A_j) - p_i p_j}{\sqrt{p_i p_j}} =$$

$$= \frac{1}{\sqrt{p_i p_j}} \cdot \begin{cases} p_i - p_i p_j, & \text{если } i = j, \\ 0 - p_i p_j, & \text{если } i \neq j \end{cases} = \begin{cases} 1 - p_i, & \text{если } i = j, \\ -\sqrt{p_i p_j}, & \text{если } i \neq j. \end{cases}$$

Удобно записать эту матрицу ковариаций следующим образом:

$$\Sigma = E - PP^T = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} \sqrt{p_1} \\ \dots \\ \sqrt{p_k} \end{pmatrix} \times \left(\sqrt{p_1} \ \dots \ \sqrt{p_k} \right).$$

Умножив её матрицу на вектор P справа, получим нулевой вектор. Действительно, $\Sigma P = EP - PP^T P = P - P = \vec{0}$, поскольку норма вектора P равна единице: $P^T P = \|P\|^2 = 1$. Значит, столбцы матрицы Σ линейно зависимы и она вырождена. Заметим также, что

$$P^T \vec{\xi}^{(1)} = \mathrm{I}(X_1 \in A_1) + \ldots + \mathrm{I}(X_1 \in A_k) = 1. \tag{41}$$

2. Из равенства (41) мораль: если последняя строка ортогональной матрицы Q будет равна P^T, то после умножения Q на $\vec{\xi}^{(1)}$ получим вектор с единичной последней координатой (а после центрирования — с нулевой).

Пусть Q — любая ортогональная матрица с последней строкой P^T. Вектор $Q\vec{\xi}^{(1)}$ имеет по лемме 10 матрицу ковариаций $B = Q\Sigma Q^T$. Убедимся, что B — почти единичная матрица (от единичной её отличает лишь последний диагональный элемент $b_{kk} = 0$).

Заметим сначала, что скалярное произведение i-й строки матрицы Q на вектор P, лежащий в её последней строке, равно нулю при $i \neq k$ и единице при $i = k$, поэтому $QP = (0, \ldots, 0, 1)$.

$$B = Q\Sigma Q^T = Q(E - PP^T)Q^T = E - QP \cdot (QP)^T.$$

Матрица $C = QP \cdot (QP)^T$ состоит из нулей, и только её элемент c_{kk} равен единице. Окончательно получаем

$$B = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \begin{pmatrix} E_{k-1} & 0 \\ 0 & 0 \end{pmatrix}. \tag{42}$$

3. Осталось повторить то, что мы уже описали в плане: умножение $Q\vec{\xi}^{(1)} = (\hat{\xi}^{(1)}, 1)$ приводит к вектору с постоянной последней координатой из равенства (41) и вектором средних $QP = (0, \ldots, 0, 1) = (\mathsf{E}\hat{\xi}^{(1)}, 1)$.

Равенствами (42) мы показали, что вектор $\hat{\xi}^{(1)} \in \mathbb{R}^{k-1}$ имеет невырожденную единичную матрицу ковариаций E_{k-1}.

Векторы $\hat{\xi}^{(1)}, \hat{\xi}^{(2)}, \ldots$ независимы и одинаково распределены. Все условия многомерной ЦПТ выполнены, поэтому

$$\hat{\eta}^{(n)} = \frac{\hat{\xi}^{(1)} + \ldots + \hat{\xi}^{(n)}}{\sqrt{n}} \Rightarrow \eta \in \mathrm{N}_{\vec{0}, E_{k-1}}.$$

По следствию 2, норма вектора $\hat{\eta}^{(n)}$ слабо сходится к норме вектора η, состоящего согласно теореме 33 из независимых случайных величин со стандартным нормальным распределением:

$$\|\hat{\eta}^{(n)}\|^2 \Rightarrow \|\eta\|^2 = \sum_{i=1}^{k-1} \eta_i^2 = \chi^2_{k-1} \in \ H_{k-1}. \qquad (43)$$

Осталось заметить, что у векторов $\vec{\eta}^{(n)}$, $Q\vec{\eta}^{(n)}$, $\hat{\eta}^{(n)}$, связанных равенствами

$$Q\vec{\eta}^{(n)} = \frac{Q\vec{\xi}^{(1)} + \ldots + Q\vec{\xi}^{(n)} - QP}{\sqrt{n}} = (\hat{\eta}^{(n)}, 0),$$

нормы одинаковы: $\|\vec{\eta}^{(n)}\|^2 = \|Q\vec{\eta}^{(n)}\|^2 = \|(\hat{\eta}^{(n)}, 0)\|^2 = \|\hat{\eta}^{(n)}\|^2$. И все эти нормы ведут себя согласно (43). □

Упражнение. Найти среди этих норм величину ρ из теоремы Пирсона и убедиться, что доказательство завершено.

§ 3. Вопросы и упражнения

1. Случайные величины $\xi \in \mathrm{N}_{a_1, \sigma_1^2}$ и $\eta \in \mathrm{N}_{a_2, \sigma_2^2}$ имеют коэффициент корреляции $-1 < \rho < 1$. Найти плотность совместного распределения случайных величин ξ и η.

2. Нормально распределённый случайный вектор имеет нулевые средние и следующую матрицу ковариаций:

$$\Sigma = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \times \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Найти линейное преобразование, переводящее исходный вектор в вектор со стандартным нормальным распределением.

3. Нормально распределённый случайный вектор имеет нулевые средние и следующую матрицу ковариаций:

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Найти поворот, делающий координаты вектора независимыми. Найти линейное преобразование, переводящее исходный вектор в вектор со стандартным нормальным распределением.

4. Случайные величины ξ и η имеют плотность совместного распределения

$$f(x,\,y) = C \cdot e^{-\frac{1}{2}\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)}.$$

Найти постоянную C. Найти преобразование, делающее координаты вектора независимыми случайными величинами со стандартным нормальным распределением.

5. Доказать, что случайная величина η из примера 45 (с. 125) имеет стандартное нормальное распределение.

6. Случайные величины ξ и η имеют двумерное стандартное нормальное распределение. Найти вероятность $\mathsf{P}(\xi^2+\eta^2<r^2)$.

7. Случайные величины ξ и η имеют плотность совместного распределения

$$f(x,\,y) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{20}\left(\frac{x^2}{5} - \frac{3xy}{5} + \frac{y^2}{2}\right)\right\}.$$

Найти вероятность попадания случайной точки $(\xi,\,\eta)$ в эллипс

$$\frac{x^2}{5} - \frac{3xy}{5} + \frac{y^2}{2} \leqslant \frac{9}{10}.$$

8. Случайные величины ξ и η имеют двумерное нормальное распределение с вектором средних $(a,\,b)$ и матрицей ковариаций $\Sigma = B^2$. Найти вероятность вектору $(\xi - a,\,\eta - b)$ принадлежать эллипсу, который получается из круга $x^2+y^2<r^2$ линейным преобразованием с помощью матрицы B всех его точек: $(x,\,y) \mapsto B(x,\,y)$.

ГЛАВА XI

ПОСТРОЕНИЕ ЭФФЕКТИВНЫХ ОЦЕНОК

В этом разделе мы научимся строить эффективные оценки. Напомним, что до сих пор мы иногда могли обосновать эффективность оценки с помощью неравенства Рао — Крамера. При этом для некоторых регулярных семейств мы просто не смогли обнаружить оценок, обращающих это неравенство в равенство. Для нерегулярных же семейств, подобных равномерному, и это средство не работало. Оказывается, эффективные оценки в любом классе строятся с помощью оператора ортогонального проектирования произвольной оценки этого класса на множество случайных величин, являющихся борелевскими функциями от *полной и достаточной статистики*. Такая ортопроекция носит название *условного математического ожидания*. Условные математические ожидания (УМО) суть жизненно важные объекты для экономистов: например, оценки метода наименьших квадратов в линейной регрессии возникают как раз при ортогональном проектировании некоторого случайного вектора на некоторое линейное подпространство в $\mathbb{R}^n$. Байесовские подходы к построению оценок при имеющейся априорной информации о параметре тоже сводятся к вычислению УМО. При изучении случайных процессов во второй половине курса эконометрики УМО тоже возникают.

§ 1. Условные математические ожидания

Определение УМО. Условное математическое ожидание имеет простой геометрический смысл, начнём с него.

Пусть ξ и η — две случайные величины на некотором вероятностном пространстве, причём $\mathsf{E}|\xi| < \infty$. Ничего принципиально не изменится, если они обе или только одна из них будет многомерной случайной величиной (случайным вектором).

Пусть $L = L(\eta)$ — множество, в котором собраны все случайные величины, имеющие вид $\zeta = g(\eta)$, где $g(x)$ — произвольная борелевская функция. Скалярным произведением двух случайных величин φ и ζ назовём $(\varphi, \zeta) = \mathsf{E}(\varphi\zeta)$, если это математическое ожидание существует.

Условное математическое ожидание $\mathsf{E}(\xi\,|\,\eta)$ *случайной величины* ξ *относительно* η можно представлять себе как результат ортогонального проектирования случайной величины ξ на пространство L.

Изобразим случайные величины векторами в $\mathbb{R}^3$, а пространство L — плоскостью, проходящей через начало координат, как на рис. 12.

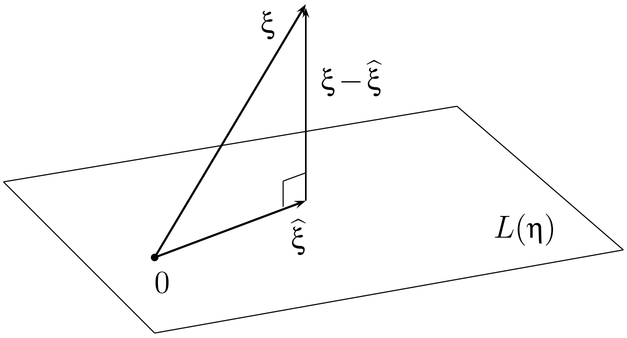

Рис. 12. Ортогональное проектирование

Результат проектирования — такая *случайная величина* $\mathsf{E}(\xi\,|\,\eta) = \widehat{\xi}$ из L, для которой выполнено основное и единственное свойство ортопроекции: её разность с ξ ортогональна всем элементам L. Ортогональность означает, что для любой $g(\eta) \in L$ обращается в нуль (если вообще существует) скалярное произведение $(\xi - \widehat{\xi},\, g(\eta))$, т. е.

$$\mathsf{E}\big((\xi - \widehat{\xi})g(\eta)\big) = 0 \quad \text{или} \quad \mathsf{E}\big(\xi g(\eta)\big) = \mathsf{E}\big(\widehat{\xi} g(\eta)\big).$$

Это свойство называют *тождеством ортопроекции*. Чтобы не иметь проблем с существованием матожидания произведения, достаточно в качестве $g(y)$ в тождество подставлять лишь ограниченные функции.

Определение 35. Пусть $\mathsf{E}|\xi| < \infty$, $L = L(\eta)$ — множество всех борелевских функций от случайной величины η. *Условным математическим ожиданием* $\mathsf{E}(\xi\,|\,\eta)$ называется случайная величина $\widehat{\xi} \in L$, удовлетворяющая тождеству ортопроекции:

$$\mathsf{E}\big(\xi g(\eta)\big) = \mathsf{E}\big(\widehat{\xi} g(\eta)\big) \quad \text{для любой ограниченной} \quad g(\eta) \in L. \tag{44}$$

Полезно заметить, что вместе с $\widehat{\xi}$ свойством (44) будет обладать любая $\tilde{\xi} = \widehat{\xi}$ п. н. Иначе говоря, условное математическое ожидание определяется не однозначно. Различные варианты УМО совпадают друг с другом п. н. Обратим внимание на то, что $\widehat{\xi}$ есть элемент множества $L(\eta)$, т. е. она является некоторой борелевской функцией $h(\eta)$ от величины η.

Данное выше определение не является конструктивным. Однако из него вытекают многие замечательные свойства, которых часто бывает достаточно для вычисления УМО. Видимо, самым важным и самым очевидным свойством является следующее. Мы не будем его доказывать, а читатель увидит его на рис. 12.

Свойство 15. Пусть $\mathsf{E}\xi^2 < \infty$. Тогда расстояние от ξ до её ортопроекции $\widehat{\xi} = \mathsf{E}(\xi \,|\, \eta)$ является наименьшим из расстояний от ξ до всех «точек» множества L:

$$\min \mathsf{E}\big(\xi - g(\eta)\big)^2 = \mathsf{E}\big(\xi - \widehat{\xi}\big)^2,$$

где минимум берётся по всем $g(\eta) \in L$.

УМО обладает обычными свойствами математических ожиданий, например, линейностью $\mathsf{E}(\xi_1 + \xi_2 \,|\, \eta) = \mathsf{E}(\xi_1 \,|\, \eta) + \mathsf{E}(\xi_2 \,|\, \eta)$ п. н. Но теперь борелевские функции от случайной величины η выносятся из-под знака математического ожидания как постоянные.

Свойство 16. Если $f(\eta) \in L$ такова, что $\mathsf{E}|f(\eta) \cdot \xi| < \infty$, то

$$\mathsf{E}\big(f(\eta) \cdot \xi \;\big|\; \eta\big) = f(\eta) \cdot \mathsf{E}(\xi \,|\, \eta) \text{ п. н.}$$

Доказательство. Рассмотрим только случай, когда $f(\eta)$ ограничена. Проверим, что $\zeta = f(\eta) \cdot \mathsf{E}(\xi \,|\, \eta)$ удовлетворяет тождеству ортопроекции: для любой ограниченной $g(\eta) \in L$

$$\mathsf{E}\big(f(\eta)\, \xi \cdot g(\eta)\big) = \mathsf{E}\big(\zeta \cdot g(\eta)\big).$$

Обозначим $h(\eta) = f(\eta) g(\eta) \in L$. Эта функция ограничена, поэтому

$$\mathsf{E}\big(\xi\, f(\eta) \cdot g(\eta)\big) = \mathsf{E}\big(\xi\, h(\eta)\big) = \mathsf{E}\big(\widehat{\xi}\, h(\eta)\big) = \mathsf{E}\big(\zeta \cdot g(\eta)\big).$$

Второе равенство верно по тождеству (44) для $\widehat{\xi}$. □

Свойство 17. Пусть $f(\eta) \in L$ и $\mathsf{E}|f(\eta)| < \infty$. Тогда

$$\mathsf{E}\big(f(\eta) \;\big|\; \eta\big) = f(\eta) \text{ п. н.}$$

Доказательство. В предыдущем свойстве возьмём $\xi = 1$. □

Полезной оказывается *формула последовательного усреднения* или *полной вероятности*, вытекающая из тождества (44) при $g(\eta) = 1$.

Свойство 18. $\mathsf{E}\xi = \mathsf{E}\big[\mathsf{E}(\xi \mid \eta)\big]$, т. е. $\mathsf{E}\xi = \mathsf{E}\widehat{\xi}$.

Более общий вариант этой формулы выглядит следующим образом.

Свойство 19. Если $\mathsf{E}|g(\xi, \eta)| < \infty$, то

$$\mathsf{E} g(\xi, \eta) = \mathsf{E}\Big[\mathsf{E}(g(\xi, y) \,|\, \eta)\big|_{y=\eta}\Big].$$

Свойство 20. Если ξ и η независимы, то $\mathsf{E}(\xi \,|\, \eta) = \mathsf{E}\xi$.

Вычисление УМО. Поскольку мы не особенно различаем случайные величины, совпадающие п. н., полезно явным образом предъявить хотя бы одну функцию $h(y)$ такую, что $\mathsf{E}(\xi \,|\, \eta) = h(\eta)$ п. н.

Можно в качестве такой функции взять $h(y) = \mathsf{E}(\xi \,|\, \eta = y)$. Что такое условное математическое ожидание относительно события $\{\eta = y\}$? Ответим на этот вопрос в двух практически значимых случаях: когда случайные величины ξ и η имеют либо дискретные распределения, либо их совместное распределение абсолютно непрерывно.

В первом случае пусть ξ принимает значения $a_1, a_2, \ldots,$ а η — значения $b_1, b_2, \ldots$ Тогда $h(\eta)$ может принимать только значения $h(b_1), h(b_2), \ldots,$ где

$$h(y) = \sum_i a_i \mathsf{P}(\xi = a_i \,|\, \eta = y).$$

Иначе говоря, при каждом фиксированном y значение $h(y)$ определяется как математическое ожидание дискретного распределения со значениями a_i и вероятностями $\mathsf{P}(\xi = a_i \,|\, \eta = y)$. Такое распределение называется *условным распределением* случайной величины ξ при условии $\eta = y$.

Во втором случае пусть $f_{\xi,\eta}(x, y)$ — плотность совместного распределения, $f_\eta(y)$ — плотность распределения величины η. Тогда положим

$$h(y) = \int\limits_{\mathbb{R}} x \, \frac{f_{\xi,\eta}(x, y)}{f_\eta(y)} \, dx. \tag{45}$$

При фиксированном y число $h(y)$ есть математическое ожидание абсолютно непрерывного распределения с плотностью распределения

$$f(x \,|\, y) = \frac{f_{\xi,\eta}(x, y)}{f_\eta(y)} = \frac{f_{\xi,\eta}(x, y)}{\int\limits_{\mathbb{R}} f_{\xi,\eta}(x, y)\, dx}.$$

Такое распределение называется *условным распределением* величины ξ при условии $\eta = y$, а функция $f(x \,|\, y)$ — *условной плотностью*.

Убедимся формально (скажем, в абсолютно непрерывном случае), что определённая выше $h(\eta)$, где $h(y)$ задаётся формулой (45), удовлетворяет тождеству ортопроекции (44) и, следовательно, является УМО $\mathsf{E}(\xi \,|\, \eta)$. Для любой $g(\eta) \in L$ (такой, что соответствующее математическое ожидание существует) левая часть тождества (44) равна

$$\mathsf{E}(\xi\, g(\eta)) = \iint\limits_{\mathbb{R}^2} x g(y) f_{\xi,\eta}(x, y)\, dx\, dy.$$

Правая часть равна

$$\mathsf{E}\big(h(\eta)g(\eta)\big) = \int\limits_{\mathbb{R}} h(y)g(y)f_\eta(y)\,dy = \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} x\,\frac{f_{\xi,\eta}(x,\,y)}{f_\eta(y)}\,dx\cdot g(y)f_\eta(y)\,dy.$$

Сокращая $f_\eta(y)$, получаем равенство левой и правой частей.

Пример 46. Пусть $X_1,\,\ldots,\,X_n$ — выборка из распределения Пуассона, $S_n = X_1+\ldots+X_n$. Вычислим по определению $\mathsf{E}(X_1\,|\,S_n)$.

Найдём условное распределение. При $k\leqslant m$

$$\mathsf{P}(X_1=k\,|\,S_n=m) = \frac{\mathsf{P}(X_1=k,\ S_n=m)}{\mathsf{P}(S_n=m)} = \frac{\mathsf{P}(X_1=k,\ X_2+\ldots+X_n=m-k)}{\mathsf{P}(S_n=m)}.$$

В числителе $X_2+\ldots+X_n \in \Pi_{\lambda(n-1)}$, поэтому он равен

$$\begin{aligned}\mathsf{P}(X_1=k,\ X_2+\ldots+X_n=m-k) &= \frac{\lambda^k}{k!}\,e^{-\lambda}\cdot\frac{((n-1)\lambda)^{m-k}}{(m-k)!}\,e^{-\lambda(n-1)} = \\ &= \frac{(n-1)^{m-k}}{k!\,(m-k)!}\,\lambda^m e^{-n\lambda}.\end{aligned}$$

Знаменатель равен $\mathsf{P}(S_n=m) = \dfrac{n^m}{m!}\,\lambda^m e^{-n\lambda}$. Поделив одно на другое, получим

$$\mathsf{P}(X_1=k\,|\,S_n=m) = C_m^k\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{m-k}.$$

Итак, условное распределение является биномиальным с параметрами m и $1/n$. Его математическое ожидание $\mathsf{E}(X_1\,|\,S_n=m) = h(m) = m/n$. Поэтому УМО равно $\mathsf{E}(X_1\,|\,S_n) = h(S_n) = S_n/n = \overline{X}$.

Заметим, что это УМО легко предъявляется и без вычислений: в силу независимости и одинаковой распределённости элементов выборки распределения векторов $(X_1,\,X_2,\,\ldots,\,X_n)$ и $(X_2,\,X_1,\,\ldots,\,X_n)$ совпадают, поэтому совпадают и

$$\mathsf{P}(X_1=k\,|\,S_n=y) = \mathsf{P}(X_2=k\,|\,S_n=y),$$

а значит, совпадают п. н. и УМО

$$\mathsf{E}(X_1\,|\,S_n) = \mathsf{E}(X_2\,|\,S_n) = \ldots = \mathsf{E}(X_n\,|\,S_n).$$

Складывая их, получаем по свойству 17

$$\mathsf{E}(X_1\,|\,S_n)+\ldots+\mathsf{E}(X_n\,|\,S_n) = \mathsf{E}(S_n\,|\,S_n) = S_n, \quad \mathsf{E}(X_1\,|\,S_n) = \frac{S_n}{n} = \overline{X}.$$

Упражнение. Изучить [1, пример 1, § 10, гл. 2] и исправить одну опечатку в плотности совместного распределения.

§ 2. Байесовский подход к оцениванию параметров

Пусть $X_1, \ldots, X_n$ — выборка из распределения $\mathcal{F}_\theta$, причём параметр θ сам является случайной величиной с некоторым априорным распределением на множестве Θ с плотностью распределения $q(t)$ с конечным математическим ожиданием $\mathsf{E}|\theta| < \infty$.

Определение 36. *Байесовской оценкой* для параметра θ называется $\theta^* = \mathsf{E}(\theta \mid X_1, \ldots, X_n)$.

Если второй момент случайной величины θ конечен, то по свойству 15 байесовская оценка θ^* обеспечивает самое маленькое среднеквадратичное отклонение

$$\min \mathsf{E}(\theta - g(X_1, \ldots, X_n))^2 = \mathsf{E}(\theta - \theta^*)^2.$$

Заметим, что это качество никак не связано с эффективностью оценок: там квадратичное отклонение вычисляется от любой *точки* $\theta \in \Theta$, здесь же при вычислении математического ожидания ведётся дополнительное усреднение с плотностью $q(t)$ по всем возможным $\theta = t \in \Theta$.

Пример 47. Рассмотрим классический пример вычисления байесовской оценки p^* по выборке из распределения Бернулли с параметром p. Пусть про параметр p *настолько* ничего не известно, что становится возможным предположить, будто любые его значения априори одинаково вероятны, т. е. $p \in \mathrm{U}_{0,1}$ с плотностью $q(t) = I(0 < t < 1)$.

По определению $p^* = \mathsf{E}(p \mid X_1, \ldots, X_n) = h(X_1, \ldots, X_n)$ п. н. По формуле (45) $h(\vec{y}\,)$ равно математическому ожиданию условного распределения с плотностью

$$f(t \mid \vec{y}\,) = \frac{f(\vec{y};\, t) \cdot q(t)}{f_{\vec{X}}(\vec{y}\,)}.$$

В числителе стоит плотность совместного распределения выборки и параметра. Она равна произведению плотности $q(t)$ распределения параметра на условную плотность распределения выборки, если параметр равен t, т. е. на функцию правдоподобия (см. пример 12, с. 33)

$$f(\vec{y};\, t) = t^{n\overline{y}}(1 - t)^{n - n\overline{y}}.$$

Для удобства заметим, что знаменатель в формуле условной плотности, равно как и в формуле (45), нам совершенно не интересен: он не зависит от переменной t и поэтому является для условной плотности просто нормирующей постоянной $C = \mathrm{const}$. Итак, условная плотность равна

$$f(t \mid \vec{y}\,) = \frac{t^{n\overline{y}}(1 - t)^{n - n\overline{y}} \cdot \mathrm{I}(0 < t < 1)}{C}.$$

Данная плотность является плотностью *бета-распределения* $\mathrm{B}_{\lambda_1,\lambda_2}$ с параметрами $\lambda_1 = n\overline{y}+1$, $\lambda_2 = n - n\overline{y}+1$. С помощью *бета-функции* Эйлера вычисляется его математическое ожидание (см. [1, п. 8, § 2, гл. 2])

$$h(\vec{y}) = \frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{n\overline{y}+1}{n\overline{y}+1+n-n\overline{y}+1} = \frac{n\overline{y}+1}{n+2}, \quad p^* = h(\vec{X}) = \frac{n\overline{X}+1}{n+2}.$$

Оценка имеет довольно экзотический вид. Никакими другими методами такой оценки мы не получали.

§ 3. Полные и достаточные статистики

Пусть есть выборка $X_1, \ldots, X_n \in \mathcal{F}_\theta$, $\theta \in \Theta$. Вспомним, какие задачи мы решали по выборке: задачи точечной оценки параметра θ, построения доверительных интервалов для него, проверки гипотез относительно него. Ситуация парадоксальна: неизвестное число одно (если параметр одномерный), но мы вынуждены хранить в памяти громадные объёмы данных — всю выборку. Нельзя ли сократить хранимую информацию так, чтобы при этом не потерялись никакие сведения о параметре, содержащиеся в выборке?

Определение 37. Статистика $S = S(X_1, \ldots, X_n)$ называется *достаточной* для параметра θ, если при любом s и $B \in \mathfrak{B}(\mathbb{R}^n)$ условное распределение $\mathsf{P}(X_1, \ldots, X_n \in B \mid S = s)$ не зависит от параметра θ.

Определение достаточной статистики говорит следующее: если значение статистики S известно и фиксировано, то выборка после этого бесполезна; даже знание её распределения (*разве не его мы искали до сих пор*?) не даёт более никакой информации о параметре! *Достаточно* по выборке вычислить S, и выборку можно выбросить. Следует ожидать, что наилучшие оценки, короткие доверительные интервалы, оптимальные критерии будут зависеть только от достаточных статистик.

Существует простой критерий достаточности статистик (доказательство см. в [1, § 12, гл. 2]).

Теорема 35 (факторизационная теорема Неймана — Фишера). *Статистика S является достаточной тогда и только тогда, когда функция правдоподобия представима в виде произведения двух функций*

$$f(X_1, \ldots, X_n; \theta) = h(\vec{X}) \cdot \Psi(S, \theta) \quad \text{п. н.},$$

каждая из которых зависит только от указанных аргументов.

Пример 48. Найдём достаточные статистики для параметров некоторых семейств распределений.

Если выборка взята из распределений B_p, Π_λ или E_α, то достаточной статистикой для соответствующего параметра будет $S = n\overline{X}$ или $S = \overline{X}$:

$$\begin{aligned} f(\vec{X};\, p) &= p^{n\overline{X}}(1-p)^{n-n\overline{X}} = \Psi(n\overline{X},\, p), \qquad h(\vec{X}) \equiv 1; \\ f(\vec{X};\, \lambda) &= \frac{\lambda^{n\overline{X}}}{\prod X_i!}\, e^{-n\lambda} = \frac{1}{\prod X_i!} \cdot \Psi(n\overline{X},\, \lambda); \\ f(\vec{X};\, \alpha) &= \alpha^n e^{-\alpha n\overline{X}} \cdot \mathrm{I}(X_{(1)} \geqslant 0) = \mathrm{I}(X_{(1)} \geqslant 0) \cdot \Psi(n\overline{X},\, \alpha). \end{aligned}$$

При $X_i \in \mathrm{U}_{0,\,\theta}$ достаточной статистикой для θ будет $S = X_{(n)}$:

$$f(\vec{X};\, \theta) = \frac{1}{\theta^n} \cdot \mathrm{I}(X_{(1)} \geqslant 0) \cdot \mathrm{I}(X_{(n)} \leqslant \theta) = \mathrm{I}(X_{(1)} \geqslant 0) \cdot \Psi(X_{(n)},\, \theta).$$

Пусть $X_i \in \mathrm{N}_{a,\sigma^2}$. Для двумерного параметра $(a,\, \sigma^2)$ достаточной статистикой будет $S = (n\overline{X^2},\, n\overline{X})$:

$$\begin{aligned} f(\vec{X};\, a,\, \sigma^2) &= \left(2\pi\sigma^2\right)^{-n/2} e^{-\sum(X_i-a)^2/2\sigma^2} = \\ &= \left(2\pi\sigma^2\right)^{-n/2} e^{-(n\overline{X^2}-2an\overline{X}+na)/2\sigma^2} = \Psi(S,\, a,\, \sigma^2). \end{aligned}$$

Если достаточная статистика является к тому же *полной*, то с её помощью можно строить эффективные оценки.

Определение 38. Пусть $X_i \in \mathcal{F}_\theta$, $\theta \in \Theta$. Статистика S называется *полной*, если равенство

$$\mathsf{E}g(S) = 0 \quad \text{для всех} \quad \theta \in \Theta$$

влечёт $g(S) = 0$ п. н. (здесь $g(x)$ — просто борелевская функция).

Свойство полноты статистики S необходимо только для того, чтобы в любом классе оценок K_b оценка, являющаяся функцией от S, была единственна (если таковая вообще существует). Действительно, если таких оценок две: $\theta_1^*(S) \in K_b$ и $\theta_2^*(S) \in K_b$, то $\mathsf{E}(\theta_1^*(S) - \theta_2^*(S)) = 0$ для всех $\theta \in \Theta$. Тогда $g(S) = \theta_1^*(S) - \theta_2^*(S) = 0$ п. н. из-за полноты S.

А если мы вспомним, что эффективная оценка в любом классе тоже не более чем одна, то дальнейшие шаги очевидны: будем в качестве эффективной оценки искать функцию от полной и достаточной статистики.

Теорема 36. *Пусть* $X_i \in \mathcal{F}_\theta$, $\theta \in \Theta$, S *— полная и достаточная статистика. Если оценка* $\theta_S^* \in K_b$ *является функцией от* S, *то она эффективна в классе* K_b.

Доказательство. Возьмём произвольную оценку $\theta^* \in K_b$. Докажем вспомогательное утверждение.

Лемма 12. $\mathsf{E}(\theta^* - \theta_S^*)(\theta_S^* - \theta) = 0$ *для любого* $\theta \in \Theta$.

Доказательство леммы 12. Вичисляя по формуле последовательного усреднения (свойство 18) сначала УМО относительно S и вынося по свойству 16 величину $(\theta^*_S - \theta)$ из-под знака УМО как борелевскую функцию от S, получаем

$$\mathsf{E}(\theta^* - \theta^*_S)(\theta^*_S - \theta) = \mathsf{E}\big[\mathsf{E}\big((\theta^* - \theta^*_S)(\theta^*_S - \theta)\,|\,S\big)\big] = \mathsf{E}\big[(\theta^*_S - \theta)\mathsf{E}\big(\theta^* - \theta^*_S\,|\,S\big)\big].$$

Заметим, что $\mathsf{E}\big(\theta^*\,|\,S\big) = \theta^*_S$ п. н. Действительно, это УМО есть функция от S, математическое ожидание которой по свойству 18 равно $\mathsf{E}\theta^*$. Следовательно, $\mathsf{E}\big(\theta^*\,|\,S\big)$ — оценка из класса K_b. Но из-за полноты S в классе K_b может быть только одна оценка, являющаяся функцией от S. Такая уже есть — это θ^*_S. Поэтому $\mathsf{E}\big(\theta^*\,|\,S\big) = \theta^*_S$, $\mathsf{E}\big(\theta^* - \theta^*_S\,|\,S\big) = 0$ п. н.

Утверждение леммы вытекает из равенств

$$\mathsf{E}(\theta^* - \theta^*_S)(\theta^*_S - \theta) = \mathsf{E}\big[(\theta^*_S - \theta)\mathsf{E}\big(\theta^* - \theta^*_S\,|\,S\big)\big] = \mathsf{E}\big[(\theta^*_S - \theta)\cdot 0\big] = 0.$$

Вернёмся к доказательству теоремы. Используя равенство нулю смешанного момента $\mathsf{E}(\theta^* - \theta^*_S)(\theta^*_S - \theta) = 0$ по лемме 12, сравним

$$\mathsf{E}(\theta^* - \theta)^2 = \mathsf{E}(\theta^* - \theta^*_S + \theta^*_S - \theta)^2 = \mathsf{E}(\theta^* - \theta^*_S)^2 + \mathsf{E}(\theta^*_S - \theta)^2 \geqslant \mathsf{E}(\theta^*_S - \theta)^2.$$

Среднеквадратичное отклонение произвольной оценки $\theta^* \in K_b$ оказалось не меньше, чем у θ^*_S. Поэтому θ^*_S эффективна в K_b. □

А бывают ли полными достаточные статистики?

Пример 49. Пусть $X_i \in \mathrm{U}_{0,\theta}$, $\theta > 0$, $S = X_{(n)}$. Проверим её полноту. Предположим, что для любого $\theta > 0$

$$\mathsf{E}g(S) = \int\limits_0^\theta g(y)\,\frac{ny^{n-1}}{\theta^n}\,dy = 0.$$

Покажем, что тогда $g(S) = 0$ п. н. Постоянные под интегралом в нуль не обращаются, поэтому достаточно доказать требуемый факт для функции $h(y) = g(y)\cdot y^{n-1}$. Положим для удобства $h(y) = 0$ при $y < 0$.

Вычитая друг из друга два нулевых интеграла, получаем

$$\int\limits_a^b h(y)\,dy = 0 \quad \text{для любых} \quad a < b.$$

Покажем, что тогда интеграл от функции h по любому борелевскому множеству B равен нулю. Пусть множество $\mathcal{A}$ состоит из всех B таких, что интеграл по B от функции h равен нулю:

$$\mathcal{A} = \Big\{B \mid \int\limits_B h(y)\,dy = 0\Big\}.$$

Множество $\mathcal{A}$ является σ-алгеброй (*проверьте*) и содержит все интервалы на прямой. Следовательно, $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{A}$.

Рассмотрим теперь два борелевских (*почему*?) множества

$$B_1 = \{x \mid h(x) > 0\}, \quad B_2 = \{x \mid h(x) < 0\}.$$

Интеграл от h по каждому из них должен быть равен нулю. Это возможно, только если мера Лебега каждого из этих множеств нулевая. Иначе первый интеграл строго положителен, второй строго отрицателен.

Окончательно имеем $\lambda\{x \mid h(x) \neq 0\} = 0$, т. е. $g(S) = 0$ п. н.

Итак, достаточная статистика $S = X_{(n)}$ полна. Воспользуемся теоремой 35 и получим: оценка $\theta^* = X_{(n)}$ эффективна в классе $K_{-\theta/(n+1)}$, несмещённая оценка $\theta^{**} = (n+1)X_{(n)}/n$ эффективна в K_0 и т. д.

Пример 50. Пусть $X_i \in \mathrm{E}_\alpha$, $\alpha > 0$, $S = \overline{X}$. Доказать, что статистика S является полной, можно как в [5, задача 11.2].

Тогда несмещённая оценка $\alpha^* = \dfrac{n-1}{n} \cdot \dfrac{1}{\overline{X}}$ из примера 25 (с. 56) является функцией от полной и достаточной статистики и, следовательно, эффективна в классе K_0.

§ 4. Вопросы и упражнения

1. Вычислить по определению $\mathsf{E}(X_1 \mid n\overline{X})$ по выборке из распределения Бернулли.

2. Исследовать свойства байесовской оценки p^* из примера 47. Найти *байесовский риск* $\mathsf{E}(p^* - p)^2$.

3. Доказать, что статистика $S = (X_{(1)}, X_{(n)})$ является достаточной, но не полной статистикой для параметра $\theta \in \mathbb{R}$ распределения $\mathrm{U}_{\theta,\,\theta+1}$.

4. Предполагая полноту достаточной статистики $S = (n\overline{X^2},\, n\overline{X})$ для двумерного параметра $(a,\, \sigma^2)$ нормального распределения, найти эффективную оценку для параметра $(a,\, \sigma^2)$.

ПРИЛОЖЕНИЕ

Таблица 1

Основные дискретные распределения

Название, обозначение, параметры	**Возможные значения k**	$\mathsf{P}(\xi = k)$	$\mathsf{E}\xi$	$\mathsf{D}\xi$
Вырожденное I_c, $c \in \mathbb{R}$	c	$\mathsf{P}(\xi = c) = 1$	c	0
Бернулли B_p $p \in (0,\ 1)$	$k = 0,\ 1$	$\mathsf{P}(\xi = 0) = 1{-}p$, $\mathsf{P}(\xi = 1) = p$	p	$p(1-p)$
Биномиальное $\mathrm{B}_{n,\,p}$ $p \in (0,\ 1)$ $n = 1, 2, \ldots$	$k = 0, \ldots, n$	$C_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$
Пуассона Π_λ $\lambda > 0$	$k = 0,\ 1,\ 2, \ldots$	$\dfrac{\lambda^k}{k!}\, e^{-\lambda}$	λ	λ
Геометрическое G_p $p \in (0,\ 1)$	$k = 1,\ 2, \ldots$	$p(1-p)^{k-1}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
Гипергеометрическое $n, K, N \in \mathbb{N}$ $0 \leqslant n, K \leqslant N$	целые от $\max(0,\ n{+}K{-}N)$ до $\min(n,\ K)$	$\dfrac{C_K^k C_{N_K}^{n-k}}{C_N^n}$	$n\dfrac{K}{N}$	$n\dfrac{K}{N}\left(1-\dfrac{K}{N}\right)\dfrac{N-n}{N-1}$

Таблица 2

Основные абсолютно непрерывные распределения

Название, обозначение, параметры	**Плотность распределения**	**Eξ**	**Dξ**	**Асимметрия**	**Эксцесс**
Равномерное на отрезке $[a,\, b]$ $\mathrm{U}_{a,\,b}$, $a < b$	$\begin{cases} \dfrac{1}{b-a}, & x \in [a,\, b], \\ 0, & x \notin [a,\, b] \end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	0	−1,2
Показательное (экспоненциальное) $\mathrm{E}_\alpha = \Gamma_{\alpha,\,1}$, $\alpha > 0$	$\begin{cases} \alpha\, e^{-\alpha x}, & x > 0, \\ 0, & x \leqslant 0 \end{cases}$	$\dfrac{1}{\alpha}$	$\dfrac{1}{\alpha^2}$	2	6
Нормальное (гауссовское) $\mathrm{N}_{a,\,\sigma^2}$, $a \in \mathbb{R}$, $\sigma > 0$	$\dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-a)^2/2\sigma^2}$, $-\infty < x < \infty$	a	σ^2	0	0
Коши $\mathrm{C}_{a,\,\sigma}$, $a \in \mathbb{R}$, $\sigma > 0$	$\dfrac{1}{\pi}\, \dfrac{\sigma}{\sigma^2 + (x-a)^2}$, $-\infty < x < \infty$	—	—	—	—
Гамма $\Gamma_{\alpha,\,\lambda}$, $\alpha > 0,\ \lambda > 0$	$\begin{cases} \dfrac{\alpha^\lambda}{\Gamma(\lambda)}\, x^{\lambda-1} e^{-\alpha x}, & x > 0, \\ 0, & x \leqslant 0 \end{cases}$	$\dfrac{\lambda}{\alpha}$	$\dfrac{\lambda}{\alpha^2}$	$\dfrac{2}{\sqrt{\lambda}}$	$\dfrac{6}{\lambda}$
Лапласа $\mathrm{L}_{\alpha,\,\mu}$, $\alpha > 0,\ \mu \in \mathbb{R}$	$\dfrac{\alpha}{2}\, e^{-\alpha\lvert x-\mu\rvert}$, $-\infty < x < \infty$	μ	$\dfrac{2}{\alpha^2}$	0	3
Парето, $\alpha > 0$	$\begin{cases} \dfrac{\alpha}{x^{\alpha+1}}, & x \geqslant 1, \\ 0, & x < 1 \end{cases}$	$\dfrac{\alpha}{\alpha - 1}$ $(\alpha > 1)$	$\dfrac{\alpha}{(\alpha-1)^2(\alpha-2)}$, $\alpha > 2$	$\sqrt{\dfrac{\alpha-2}{\alpha}}\, \dfrac{2(\alpha+1)}{\alpha-3}$, $\alpha > 3$	$\dfrac{6(\alpha^3+\alpha^2-6\alpha-2)}{\alpha(\alpha-3)(\alpha-4)}$, $\alpha > 4$

Таблица 3

Критические точки распределения χ^2

Приведены значения x, при которых $\mathsf{P}(\chi_k^2 > x) = \alpha$

α k	0,005	0,01	0,025	0,05	0,95	0,975	0,99	0,995
3	12, 84	11,34	9,35	7,81	0,35	0,22	0,12	0, 07
4	14, 86	13,28	11,14	9,49	0,71	0,48	0,30	0, 21
5	16, 75	15,09	12,83	11,07	1,15	0,83	0,55	0, 41
6	18, 55	16,81	14,45	12,59	1,64	1,24	0,87	0, 68
7	20, 28	18,48	16,01	14,07	2,17	1,69	1,24	0, 99
8	21, 95	20,09	17,53	15,51	2,73	2,18	1,65	1, 34
9	23, 59	21,67	19,02	16,92	3,33	2,70	2,09	1, 73
10	25, 19	23,21	20,48	18,31	3,94	3,25	2,56	2, 16
11	26, 76	24,73	21,92	19,68	4,57	3,82	3,05	2, 60
12	28, 30	26,22	23,34	21,03	5,23	4,40	3,57	3, 07
13	29, 82	27,69	24,74	22,36	5,89	5,01	4,11	3, 57
14	31, 32	29,14	26,12	23,68	6,57	5,63	4,66	4, 07
15	32, 80	30,58	27,49	25,00	7,26	6,26	5,23	4, 60
16	34, 27	32,00	28,85	26,30	7,96	6,91	5,81	5, 14
17	35, 72	33,41	30,19	27,59	8,67	7,56	6,41	5, 70
18	37, 16	34,81	31,53	28,87	9,39	8,23	7,01	6, 26
19	38, 58	36,19	32,85	30,14	10,12	8,91	7,63	6, 84
20	40, 00	37,57	34,17	31,41	10,85	9,59	8,26	7, 43
21	41, 40	38,93	35,48	32,67	11,59	10,28	8,90	8, 03
22	42, 80	40,29	36,78	33,92	12,34	10,98	9,54	8, 64
23	44, 18	41,64	38,08	35,17	13,09	11,69	10,20	9, 26
24	45, 56	42,98	39,36	36,42	13,85	12,40	10,86	9, 89
25	46, 93	44,31	40,65	37,65	14,61	13,12	11,52	10, 52
26	48, 29	45,64	41,92	38,89	15,38	13,84	12,20	11, 16
27	49, 65	46,96	43,19	40,11	16,15	14,57	12,88	11, 81
28	50, 99	48,28	44,46	41,34	16,93	15,31	13,56	12, 46
29	52, 34	49,59	45,72	42,56	17,71	16,05	14,26	13, 12
49	78, 23	74, 92	70, 22	66, 34	33, 93	31, 55	28, 94	27, 25
99	139, 0	134,6	128,4	123,2	77,05	73,36	69,23	66, 51
499	584, 1	575,4	562,8	552,1	448,2	439,0	428,5	421, 4
999	1117, 9	1105,9	1088,5	1073,6	926,6	913,3	898, 0	887, 6

Таблица 4

Критические точки распределения Стьюдента

Приведены значения x, при которых $\mathsf{P}(|t_k| > x) = \alpha$

α k	0,001	0,002	0,005	0,01	0,02	0,05	0,1	0,2
3	12,92	10,21	7,45	5,84	4,54	3,18	2,35	1,64
4	8,61	7,17	5,60	4,60	3,75	2,78	2,13	1,53
5	6,87	5,89	4,77	4,03	3,36	2,57	2,02	1,48
6	5,96	5,21	4,32	3,71	3,14	2,45	1,94	1,44
7	5,41	4,79	4,03	3,50	3,00	2,36	1,89	1,41
8	5,04	4,50	3,83	3,36	2,90	2,31	1,86	1,40
9	4,78	4,30	3,69	3,25	2,82	2,26	1,83	1,38
10	4,59	4,14	3,58	3,17	2,76	2,23	1,81	1,37
11	4,44	4,02	3,50	3,11	2,72	2,20	1,80	1,36
12	4,32	3,93	3,43	3,05	2,68	2,18	1,78	1,36
13	4,22	3,85	3,37	3,01	2,65	2,16	1,77	1,35
14	4,14	3,79	3,33	2,98	2,62	2,14	1,76	1,35
15	4,07	3,73	3,29	2,95	2,60	2,13	1,75	1,34
16	4,01	3,69	3,25	2,92	2,58	2,12	1,75	1,34
17	3,97	3,65	3,22	2,90	2,57	2,11	1,74	1,33
18	3,92	3,61	3,20	2,88	2,55	2,10	1,73	1,33
19	3,88	3,58	3,17	2,86	2,54	2,09	1,73	1,33
20	3,85	3,55	3,15	2,85	2,53	2,09	1,72	1,33
21	3,82	3,53	3,14	2,83	2,52	2,08	1,72	1,32
22	3,79	3,50	3,12	2,82	2,51	2,07	1,72	1,32
23	3,77	3,48	3,10	2,81	2,50	2,07	1,71	1,32
24	3,75	3,47	3,09	2,80	2,49	2,06	1,71	1,32
25	3,73	3,45	3,08	2,79	2,49	2,06	1,71	1,32
26	3,71	3,43	3,07	2,78	2,48	2,06	1,71	1,31
27	3,69	3,42	3,06	2,77	2,47	2,05	1,70	1,31
28	3,67	3,41	3,05	2,76	2,47	2,05	1,70	1,31
29	3,66	3,40	3,04	2,76	2,46	2,05	1,70	1,31
49	3,50	3,27	2,94	2,68	2,40	2,01	1,68	1,30
99	3,39	3,17	2,87	2,63	2,36	1,98	1,66	1,29
∞	3,29	3,09	2,81	2,58	2,33	1,96	1,64	1,28

Таблица 5

Критические точки распределения Фишера

Приведены значения x, при которых $\mathsf{P}(f_{k_1,k_2} > x) = 0{,}05$

$\boldsymbol{k_1}$ $\boldsymbol{k_2}$	1	2	3	4	5	6	7	8	9	10
1	161	199	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

СПИСОК ЛИТЕРАТУРЫ

1. **Боровков А. А.** Математическая статистика. М.: Наука, 1984. 472 с.

2. **Колемаев В. А., Калинина В. Н.** Теория вероятностей и математическая статистика. М.: Инфра-М, 1997. 302 с.

3. **Ивченко Г. И., Медведев Ю. И.** Математическая статистика. М.: Высш. шк., 1984. 248 с.

4. **Пугачев В. С.** Теория вероятностей и математическая статистика. М.: Физматлит, 2002. 496 с.

5. **Коршунов Д. А., Чернова Н. И.** Сборник задач и упражнений по математической статистике. Новосибирск: Изд-во Ин-та математики, 2001. 128 с.

6. **Большев Л. Н., Смирнов Н. В.** Таблицы математической статистики. М.: Наука, 1965.

Учебное издание

Чернова Наталья Исааковна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Второе издание, исправленное и дополненное

Подписано в печать
Формат $60 \times 84\ ^1/_{16}$. Офсетная печать.
Уч.-изд. л. 9,4. Усл. печ. л. 8,7. Тираж 200 экз.
Заказ №

Редакционно-издательский центр НГУ.
630090, Новосибирск-90, ул. Пирогова, 2.