Краткий конспект лекций по математической статистике

для студентов 2 курса ЭФ, отделение
«математические методы и исследование операций в экономике»
(1 семестр 1997-98 уч.года)

Чернова Н.И.
cher@nsu.ru, тел. (3832) 46-73-00

1 Основные понятия математической статистики

1.1 Задачи математической статистики

Математическая (или теоретическая) статистика опирается на методы и понятия теории вероятностей, но решает в каком-то смысле обратные задачи.

В теории вероятностей рассматриваются случайные величины с *заданным* распределением или случайные эксперименты, свойства которых *целиком известны*. Предмет теории вероятностей — свойства и взаимосвязи этих величин (распределений).

Но часто эксперимент представляет собой черный ящик, выдающий лишь некие результаты, по которым требуется сделать вывод о свойствах самого эксперимента. Наблюдатель имеет набор числовых (во всяком случае, их всегда можно сделать числовыми) результатов, полученных повторением одного и того же случайного эксперимента в одинаковых условиях. Примером такой серии экспериментов может служить социологический опрос, набор экономических показателей или, наконец, последовательность гербов и решек при тысячекратном подбрасывании монеты.

При этом возникают следующие вопросы:

1) Если мы наблюдаем одну случайную величину — как по набору ее значений в нескольких опытах сделать как можно более точный вывод о ее распределении?

2) Если мы наблюдаем одновременно проявление двух (или более) признаков, т.е. имеем набор значений нескольких случайных величин — что можно сказать об их зависимости? Есть она или нет? А если есть, то какова эта зависимость?

Часто бывает возможно высказать некие предположения о распределении, спрятанном в «черном ящике», 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 определенной степенью достоверности, и чем дольше мы можем продолжать эксперимент, тем точнее могут быть выводы (а это далеко не всегда возможно).

Итак, о (математической) статистике имеет смысл вспоминать, если

а) имеется случайный эксперимент, свойства которого частично или полностью незвестны,

б) мы умеем воспроизводить этот эксперимент в одних и тех же условиях некоторое (а лучше — какое угодно) число раз.

1.2 Основные понятия выборочного метода

Пусть $\xi : \mathbf{\Omega} \to \mathbb{R}$ — случайная величина, наблюдаемая в случайном эксперименте. Предполагается, что вероятностное пространство задано (и не будет нас интересовать). Будем считать, что проведя n раз этот эксперимент в одинаковых условиях, мы получили числа $X_1, X_2, \ldots, X_n$ — значения этой случайной величины в первом, втором, и т.д. экспериментах. Пусть случайная величина ξ имеет некоторое распределение $\mathcal{F}$, *которое нам частично или совсем неизвестно*.

Рассмотрим подробнее вектор $\vec{X} = (X_1, \ldots, X_n)$, называемый *выборкой (случайной выборкой)*. В конкретной серии экспериментов выборка — это набор чисел. Но стоит эту серию экспериментов повторить еще раз, и вместо этого набора мы получим *новый* набор чисел. Вместо числа X_1 появится другое число — одно из значений случайной величины ξ. То есть X_1 (и X_2, и X_3, и т.д.) — не какое-то конкретное, раз и навсегда заданное число, а *переменная величина*, которая может принимать те же значения, что и случайная величина ξ, и так же часто (с теми же вероятностями). То есть X_1 — случайная величина, одинаково распределенная с ξ, а число, которое мы наблюдаем в данном первом эксперименте — одно из возможных значений *случайной величины* X_1.

Итак, выборка $\vec{X} = (X_1, \ldots, X_n)$ объема n это:

1) *в конкретной* серии экспериментов — набор из n чисел, являющихся значениями («реализациями») случайной величины ξ в n независимых экспериментах;

2) *в математической модели* — набор из n *независимых и одинаково распределенных* случайных величин («копий ξ»), имеющих, как и ξ, 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.

1.3 Эмпирическая функция распределения, гистограмма

Поскольку неизвестное распределение $\mathcal{F}$ можно описать, например, его функцией распределения F, построим по выборке «приближение» для этой функции.

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

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

Напоминание: функция

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

называется индикатором события $\{X_i < y\}$. Это — случайная величина, имеющая распределение 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 $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Элемент $X_{(k)}$, $k = 1, \ldots, n$, называется k-м членом вариационного ряда или k-й порядковой статистикой.

Пример 1. Выборка, $n = 15$: $\vec{X} = (0;\ 2;\ 1;\ 2{,}6;\ 3{,}1;\ 4{,}6;\ 1;\ 4{,}6;\ 6;\ 2{,}6;\ 6;\ 7;\ 9;\ 9;\ 2{,}6)$.
Вариационный ряд: $(0;\ 1;\ 1;\ 2;\ 2{,}6;\ 2{,}6;\ 2{,}6;\ 3{,}1;\ 4{,}6;\ 4{,}6;\ 6;\ 6;\ 7;\ 9;\ 9)$.
Эмпирическая функция распределения имеет скачки в точках выборки, величина скачка в точке X_i равна m/n, где m — количество элементов выборки, совпадающих с X_i.

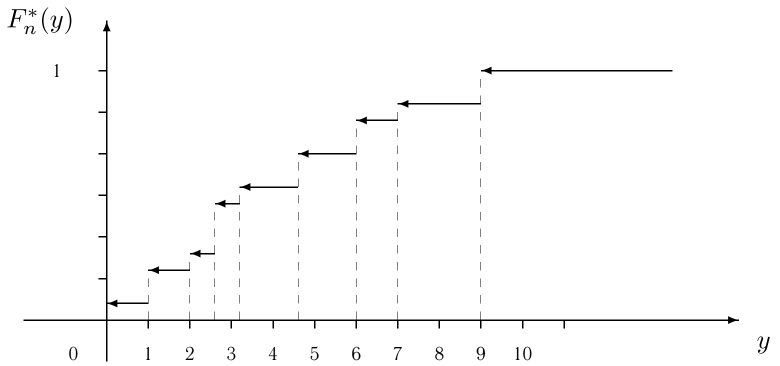


Рис. 1: Пример 1

Можно изобразить эмпирическую функцию распределения так:

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

Другой характеристикой распределения является таблица (для дискретных распределений) или плотность (для абсолютно непрерывных). Эмпирическим, или выборочным аналогом таблицы или плотности является так называемая *гистограмма*.

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

$$\nu_j = \{\text{число } X_i \in A_j\} = \sum_{i=1}^{n} \mathbf{I}(X_i \in A_j). \tag{1}$$

На каждом из интервалов A_j строят прямоугольник, площадь которого пропорциональна ν_j. Общая площадь всех прямоугольников должна равняться единице. Пусть l_j — длина интервала A_j. Высота прямоугольника над A_j равна

$$f_j = \frac{\nu_j}{n l_j}, \quad \text{здесь} \quad \sum_{j=1}^{k} \nu_j = n.$$

Полученная фигура называется гистограммой.

Пример 2. Имеется вариационный ряд (см. 1):

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

Разобьем отрезок $[0, 10]$ на 4 равных отрезка. В отрезок $A_1 = [0; 2{,}5)$ попали 4 элемента выборки, в $A_2 = [2{,}5; 5)$ — 6, в $A_3 = [5; 7{,}5)$ — 3, и в отрезок $A_4 = [7{,}5; 10]$ попали 2 элемента выборки. Строим гистограмму (слева). Справа — тоже гистограмма для той же выборки, но при разбиении области на 5 равных отрезков.

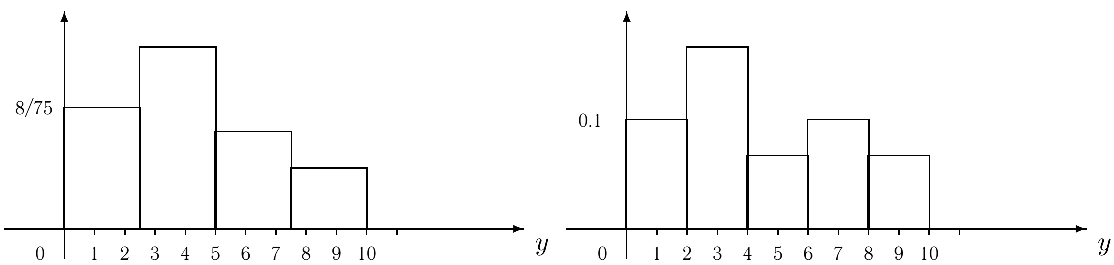


Рис. 2: Пример 2

Замечание 1. Как утверждается в курсе «Эконометрия», наилучшим числом интервалов группировки («формула Стерджесса») является

$$k = k(n) = 1 + [3.322 \lg n].$$

Здесь $\lg n$ — десятичный логарифм, поэтому $k = 1 + \log_2 10 \log_{10} n = 1 + \log_2 n$, т.е. при увеличении выборки в 2 раза число интервалов группировки увеличивается на 1.

Заметим, что чем больше интервалов группировки, тем лучше. Но это «чем больше» имеет свои границы: если брать число интервалов, скажем, порядка n, то с ростом n гистограмма, очевидно, не будет **поточечно** приближаться к плотности.

Справедливо следующее утверждение: *если плотность распределения элементов выборки является непрерывной функцией, то при $k(n) \to \infty$, так что $k(n)/n \to 0$, имеет место поточечная сходимость по вероятности гистограммы к плотности* (см. замечание 1).

1.4 Эмпирические моменты

Знание моментов распределения также многое может сказать о его виде и свойствах. Введем эмпирические (выборочные) аналоги неизвестных теоретических (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. Хорошо известны их выборочные «двойники»:

Теоретические характеристики	Эмпирические характеристики
$\mathsf{E}\xi = \mathsf{E}X_1 = a$	$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — выборочное среднее
$\mathsf{D}\xi = \mathsf{D}X_1 = \sigma^2$	$\sigma^{2*} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — выборочная дисперсия или $S_0^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — несмещенная 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

Коротко определить содержание правого и левого столбцов таблицы можно так: неизвестное «среднее по пространству» заменяется «средним по времени» (цитата, группа 476).

1.5 Сходимость эмпирических характеристик к теоретическим

Мы ввели три вида эмпирических характеристик, предназначенных для замены (оценивания) неизвестных теоретических характеристик распределения: эмпирическую функцию распределения, гистограмму, выборочные моменты. Понятно, что любое приближение хорошо, если с ростом объема выборки разница между истинной характеристикой и выборочной стремится к нулю. Такое свойство эмпирических характеристик («оценок») называют *состоятельностью*. Убедимся, что наши выборочные характеристики таким свойством обладают.

Свойства эмпирической функции распределения

Теорема 1. *Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка объема n из неизвестного распределения $\mathfrak{F}$ с функцией распределения F. Пусть F_n^* — эмпирическая функция распределения, построенная по этой выборке. Тогда для любого $y \in \mathbb{R}$*

$$F_n^*(y) \xrightarrow{\text{p}} F(y) \quad \textit{при} \quad n \to \infty.$$

Замечание 2. $F_n^*(y)$ — случайная величина, так как она является функцией от случайных величин $X_1, \ldots, X_n$. То же самое можно сказать про гистограмму и выборочные моменты.

***Доказательство теоремы* 1.** По определению, $F_n^*(y) = \frac{\sum_{i=1}^{n} \mathbf{I}(X_i < y)}{n}$. Случайные величины $\mathbf{I}(X_1 < y), \mathbf{I}(X_2 < y), \ldots$ независимы и одинаково распределены, их математическое ожидание конечно:

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

поэтому примени́м ЗБЧ Хинчина (*а что это такое?*), и

$$F_n^*(y) = \frac{\sum_{i=1}^{n} \mathbf{I}(X_i < y)}{n} \xrightarrow{\text{p}} \mathsf{E}\mathbf{I}(X_1 < y) = F(y).$$

□

Таким образом, с ростом объема выборки эмпирическая функция распределения сходится (по вероятности) к неизвестной теоретической.

На самом деле, верен более общий результат, показывающий, что сходимость эмпирической функции распределения к теоретической имеет «равномерный» характер.

Теорема 2 (Гливенко, Кантелли). *Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка объема n из неизвестного распределения $\mathcal{F}$ с функцией распределения F. Пусть F_n^* — эмпирическая функция распределения, построенная по этой выборке. Тогда*

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

Если функция распределения F непрерывна, то скорость сходимости к нулю в теореме Гливенко-Кантелли имеет порядок $\dfrac{1}{\sqrt{n}}$, как показывает

Теорема 3 (Колмогоров). *Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка объема n из неизвестного распределения $\mathcal{F}$ с непрерывной функцией распределения F. Пусть F_n^* — эмпирическая функция распределения. Тогда*

$$\sqrt{n} \sup_{y \in \mathbb{R}} |F_n^*(y) - F(y)| \Rightarrow \zeta \textit{ при} \quad n \to \infty,$$

где случайная величина ζ имеет распределение Колмогорова с функцией распределения

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

Выпишем еще ряд свойств эмпирической функции распределения, которые нам потребуются в дальнейшем. Это хорошо знакомые свойства среднего арифметического n независимых слагаемых, имеющих к тому же распределение Бернулли.

Свойство 1. *Для любого $y \in \mathbb{R}$*

1) $\mathsf{E}F_n^*(y) = F(y)$, *то есть величина $F_n^*(y)$ — «несмещенная» оценка для $F(y)$;*
2) $\mathsf{D}F_n^*(y) = \dfrac{F(y)(1 - F(y))}{n}$;
3) $\sqrt{n}(F_n^*(y) - F(y)) \Rightarrow \mathrm{N}_{0, F(y)(1-F(y))}$, *то есть величина $F_n^*(y)$ «асимптотически нормальна»;*
4) $n \cdot F_n^*(y)$ *имеет биномиальное распределение* $\mathrm{B}_{n, F(y)}$.

В первых трех пунктах утверждается, что случайная величина $F_n^*(y)$ имеет математическое ожидание $F(y)$, имеет убывающую со скоростью $1/n$ дисперсию $\dfrac{F(y)(1 - F(y))}{n}$ и, в дополнение к теореме Гливенко-Кантелли, сходится к $F(y)$ со скоростью $1/\sqrt{n}$.

Замечание 3. Полезно сравнить (3) с теоремой Колмогорова.

Замечание 4. Все определения, как то: «оценка», «несмещенность», «состоятельность», «асимптотическая нормальность» будут даны в главе 2. Но смысл этих терминов *должен быть* вполне понятен уже сейчас.

***Доказательство свойства* 1.**

1) Случайные величины $\mathbf{I}(X_1 < y), \mathbf{I}(X_2 < y), \ldots$ одинаково распределены, поэтому (*где используется одинаковая распределенность?*)

$$\mathsf{E}F_n^*(y) = \mathsf{E}\frac{\sum_{i=1}^n \mathbf{I}(X_i < y)}{n} = \frac{\sum_{i=1}^n \mathsf{E}\mathbf{I}(X_i < y)}{n} = \frac{n\mathsf{E}\mathbf{I}(X_1 < y)}{n} = F(y).$$

2) Случайные величины $\mathbf{I}(X_1 < y), \mathbf{I}(X_2 < y), \ldots$ независимы и одинаково распределены, поэтому (*где используется независимость?*)

$$\mathsf{D}F_n^*(y) = \mathsf{D}\frac{\sum_{i=1}^n \mathbf{I}(X_i < y)}{n} = \frac{\sum_{i=1}^n \mathsf{D}\mathbf{I}(X_i < y)}{n^2} = \frac{n\mathsf{D}\mathbf{I}(X_1 < y)}{n^2} = \frac{\mathsf{D}\mathbf{I}(X_1 < y)}{n}.$$

Но $\mathsf{D}\mathbf{I}(X_1 < y) = F(y)(1 - F(y))$, поскольку $\mathbf{I}(X_1 < y) \Subset \mathrm{B}_{F(y)}$.

3) Воспользуемся ЦПТ Ляпунова (*а что это такое?*).

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

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

4) Поскольку $\mathbf{I}(X_1 < y)$ (число успехов в одном испытании) имеет распределение Бернулли $\mathrm{B}_{F(y)}$ (*еще раз - почему?*), то $n \cdot F_n^*(y) = \sum_{i=1}^n \mathbf{I}(X_i < y)$ имеет биномиальное распределение $\mathrm{B}_{n,F(y)}$ (*почему? и при чем тут смысл биномиального распределения? а также при чем тут его устойчивость по суммированию?*). □

Свойства гистограммы Пусть f — истинная неизвестная плотность распределения $\mathcal{F}$ (если $\mathcal{F}$ абсолютно непрерывно). Пусть, кроме того, число k интервалов группировки *не зависит от* n. См. замечание 5 для случая, когда $k = k(n)$.

Справедлива

Теорема 4. *При $n \to \infty$ для любого $j = 1, \ldots, k$*

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

Если, к тому же, истинная плотность $f(x)$ непрерывна на интервале A_j, то интеграл справа равен $l_j \cdot f(u_j)$, где u_j — некоторая точка внутри интервала группировки A_j (найдется по теореме о среднем).

Упражнение. Доказать теорему 4, используя (1) и ЗБЧ.

Теорема утверждает, что (для непрерывной плотности) высота столбца гистограммы, построенного над интервалом группировки, с ростом объема выборки сближается со значением плотности распределения в одной из точек этого интервала. Либо (для произвольной плотности) площадь соответствующего столбца гистограммы сближается с площадью над тем же интервалом под графиком плотности.

Упражнение. Нарисовать утверждение теоремы 4 на графике плотности / гистограммы.

Замечание 5. Заметим, что чем больше интервалов группировки, тем лучше. Но это «чем больше» имеет свои границы: если брать число интервалов, скажем, порядка n, то с ростом n гистограмма не будет **поточечно** сходиться к плотности.

Справедливо следующее утверждение: *если плотность распределения элементов выборки является непрерывной функцией и $k(n)/n \to 0$, то имеет место поточечная сходимость гистограммы к плотности* (см. замечание 1).

Со своей стороны, могу предложить всегда брать число интервалов, скажем, равное целой части от корня пятой степени из n (помноженного на e^π, если объем выборки больше 413):

$$k(n) = 1 + [\sqrt[5]{n} \cdot e^\pi].$$

Свойства выборочных моментов

Лемма 1. *Выборочное среднее $\overline{X}$ является несмещенной и состоятельной оценкой для теоретического среднего (математического ожидания):*

1) $\mathsf{E}\overline{X} = \mathsf{E}X_1 = a$ — *несмещенность;*

2) $\overline{X} \xrightarrow{\mathrm{p}} \mathsf{E}X_1 = a$ — *состоятельность.*

Лемма 2. *Выборочные дисперсии σ^{2*} и S_0^2 являются состоятельными оценками для дисперсии. При этом σ^{2*} — смещенная, а S_0^2 — несмещенная оценка дисперсии:*

1) $\mathsf{E}\sigma^{2*} = \dfrac{n-1}{n}\mathsf{D}X_1 = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$ — *смещенная;*

2) $\mathsf{E}S_0^2 = \mathsf{D}X_1 = \sigma^2$ — *несмещенная;*

3) $\sigma^{2*} \xrightarrow{\text{p}} \mathsf{D}X_1 = \sigma^2, \quad S_0^2 \xrightarrow{\text{p}} \mathsf{D}X_1 = \sigma^2$ — *обе оценки состоятельны.*

Лемма 3. *Выборочный k-й момент $\overline{X^k}$ является несмещенной и состоятельной оценкой для теоретического k-го момента:*

1) $\mathsf{E}\overline{X^k} = \mathsf{E}X_1^k = m_k$ — *несмещенность;*

2) $\overline{X^k} \xrightarrow{\text{p}} \mathsf{E}X_1^k = m_k$ *при* $n \to \infty$ — *состоятельность.*

Доказательство леммы 1.

1) $\mathsf{E}\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\mathsf{E}X_i = \dfrac{1}{n}n\mathsf{E}X_1 = \mathsf{E}X_1 = a$; 2) По ЗБЧ, $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n}X_i \xrightarrow{\text{p}} \mathsf{E}X_1 = a$. □

Упражнение. Доказать лемму 3.

Доказательство леммы 2.

1) Во первых, раскрыв скобки, полезно убедиться в том что

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

Затем,

$$\begin{aligned} \mathsf{E}\sigma^{2*} &= \mathsf{E}\left[\overline{X^2} - (\overline{X})^2\right] = \mathsf{E}\overline{X^2} - \mathsf{E}(\overline{X})^2 = (\text{по лемме 3}) = \mathsf{E}X_1^2 - \mathsf{E}(\overline{X})^2 = \\ &= \mathsf{E}X_1^2 - \left[(\mathsf{E}\overline{X})^2 + \mathsf{D}(\overline{X})\right] = \left[\mathsf{E}X_1^2 - (\mathsf{E}X_1)^2\right] - \mathsf{D}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \\ &= \sigma^2 - \frac{1}{n^2}n\mathsf{D}X_1 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2. \end{aligned}$$

2) Второе утверждение следует из первого, так как $S_0^2 = \dfrac{n}{n-1}\sigma^{2*}$.

3) Из (2) и ЗБЧ, $\sigma^{2*} = \overline{X^2} - (\overline{X})^2 \xrightarrow{\text{p}} \mathsf{E}X_1^2 - (\mathsf{E}X_1)^2 = \sigma^2$.
Кроме того, $\dfrac{n}{n-1} \to 1$, так что $S_0^2 \xrightarrow{\text{p}} \sigma^2$.

□

1.6 Группированные данные (некоторые вводные понятия к эконометрии)

Если объем выборки очень велик, часто работают не с элементами выборки, а с *группированными* данными. Приведем ряд понятий, связанных с группировкой. Для простоты будем делить область выборочных данных на k *одинаковых* интервалов $A_1, \ldots, A_k$ длины Δ:

$$A_1 = [a_0, a_1), \ldots, A_k = [a_{k-1}, a_k), \quad a_j - a_{j-1} = \Delta.$$

Как прежде, пусть ν_j — число элементов выборки, попавших в интервал A_j и w_j — частота попадания в интервал A_j (оценка вероятности попадания в интервал):

$$\nu_j = \{\text{число } X_i \in A_j\} = \sum_{i=1}^{n}\mathbf{I}(X_i \in A_j), \quad w_j = \frac{\nu_j}{n}.$$

На каждом из интервалов A_j строят прямоугольник с высотой $f_j = \dfrac{w_j}{\Delta}$, и получают гистограмму.

Рассмотрим середины интервалов: $\overline{a}_j = a_{j-1} + \Delta/2$ — середина A_j. Набор

$$\underbrace{\overline{a}_1, \ldots, \overline{a}_1}_{\nu_1 \text{ раз}}, \quad \ldots, \quad \underbrace{\overline{a}_k, \ldots, \overline{a}_k}_{\nu_k \text{ раз}}$$

можно считать «огрубленной» выборкой, в которой все X_i, попадающие в интервал A_j, заменены на $\overline{a}_j$. По этой выборке можно построить такие же (но более грубые) выборочные характеристики, что и по исходной (обозначим их так же), например выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{k} \overline{a}_j \nu_j = \sum_{j=1}^{k} \overline{a}_j w_j$$

или выборочную дисперсию

$$\sigma^{2*} = \frac{1}{n}\sum_{j=1}^{k} (\overline{a}_j - \overline{X})^2 \nu_j = \sum_{j=1}^{k} (\overline{a}_j - \overline{X})^2 w_j.$$

Упражнение.

1) Найти математическое ожидание и дисперсию случайной величины, принимающей значения $\overline{a}_1, \ldots, \overline{a}_k$ с вероятностями, соответственно, $w_1, \ldots, w_k$.

2) Найти математическое ожидание и дисперсию случайной величины, принимающей значения $X_1, \ldots, X_n$ с равными вероятностями.

Упражнение.** Понять, к чему предыдущее упражнение.

Указание. Обосновать фразу: выборочные характеристики (выборочное среднее, выборочная дисперсия, выборочная функция распределения, выборочный k-й момент и др.) есть *обычные* характеристики (математическое ожидание, дисперсия, функция распределения, k-й момент и т.д.) *выборочной* случайной величины, принимающей значения $X_1, \ldots, X_n$ с равными вероятностями.

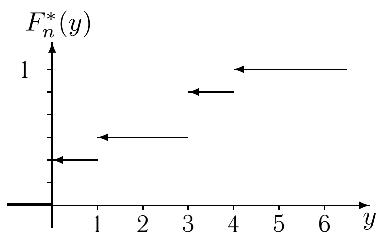
Кривая, соединяющая точки $(a_0, 0), (\overline{a}_1, f_1), \ldots, (\overline{a}_k, f_k), (a_k, 0)$ называется полигоном (частот). В отличие от гистограммы полигон — непрерывная функция (ломаная).

1.7 Вопросы и упражнения

1. Задачник [1], задачи 1.1 − 1.7, 1.11.

2. Можно ли по эмпирической функции распределения, приведенной на рис. 1, восстановить выборку $X_1, \ldots, X_n$, если n известно? А вариационный ряд? Как это сделать? А если n неизвестно?

3. Существует ли выборка $(X_1, \ldots, X_6)$ объёма 6 с нарисованной ниже эмпирической функцией распределения? А выборка $(X_1, \ldots, X_{12})$ объёма 12? Если «да», то записать её и нарисовать эмпирическую функцию распределения выборки $(2X_1, \ldots, 2X_{12})$.



4. Можно ли по гистограмме, приведенной на рис. 2, восстановить выборку $X_1, \ldots, X_n$?

5. Нарисовать эмпирическую функцию распределения, соответствующую выборке объема n из распределения Бернулли B_p. Использовать выборочное среднее $\overline{X}$. Доказать непосредственно, что выполнена теорема Гливенко-Кантелли:

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

6. Доказать, вспомнив ЦПТ, что выборочный k-й момент $\overline{X^k}$ является еще и асимптотически нормальной оценкой для теоретического k-го момента:

$$\sqrt{n}\frac{\overline{X^k}-\mathsf{E}X_1^k}{\sqrt{\mathsf{D}X_1^k}}=\frac{\sum_{i=1}^n X_i^k-n\mathsf{E}X_1^k}{\sqrt{n}\sqrt{\mathsf{D}X_1^k}}\Rightarrow \mathrm{N}_{0,1}.$$

Какой момент у случайной величины X_1 при этом должен быть конечен? Верна ли фраза: «выборочный k-й момент $\overline{X^k}$ стремится к теоретическому k-му моменту со скоростью $1/\sqrt{n}$»?

6. Вспомнить, как находить 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 распределения и плотность величины X_1.

7. Доказать (или вспомнить), что функция распределения k-й порядковой статистики $X_{(k)}$ имеет вид:

$$\mathsf{P}(X_{(k)}<y)=\mathsf{P}(\text{хотя бы } k \text{ элементов выборки } <y)=\sum_{i=k}^{n}C_n^iF(y)^i(1-F(y))^{n-i},$$

где $F(y)$ — функция распределения величины X_1.

8. Из курса «Эконометрии»: доказать, что среднее степенное

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

а) стремится к $X_{(1)}$ при $k\to-\infty$ б) стремится 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.
Указание. Вынести $X_{(1)}$ (или $X_{(n)}$) из-под корня, воспользоваться леммой о двух милиционерах и свойствами: $\sqrt[k]{k}\to 1$ при $k\to+\infty$, $\sqrt[k]{1}\to 1$ при $k\to+\infty$, и т.д.

2 Точечное оценивание

2.1 Параметрические семейства распределений

Рассматривается следующая задача. Имеется выборка объема n, элементы которой $X_1, \ldots, X_n$ независимы, одинаково распределены и имеют «известное» распределение $\mathcal{F}_\theta$ с некоторым неизвестным скалярным или векторным параметром θ. Пусть параметр θ принимает значения из некоторого множества $\Theta \subseteq \mathbb{R}^k$. Мы будем считать, что в классе распределений $\{\mathcal{F}_\theta, \theta \in \Theta\}$ каждое распределение целиком определяется значением параметра θ. То есть равенство $\theta_1 = \theta_2$ влечет равенство $\mathcal{F}_{\theta_1} = \mathcal{F}_{\theta_2}$.

Например, рассматривается задача следующего вида: для всех $i = 1, \ldots, n$

- знаем: $X_i \Subset \Pi_\lambda$, где $\lambda > 0$; не знаем: λ;
  здесь $\mathcal{F}_\theta = \Pi_\lambda$, $\theta = \lambda$, $\Theta = (0, \infty)$;
- знаем: $X_i \Subset \Pi_\lambda$, где $\lambda \in (3, 5)$; не знаем: λ;
  здесь $\mathcal{F}_\theta = \Pi_\lambda$, $\theta = \lambda$, $\Theta = (3, 5)$;
- знаем: $X_i \Subset \mathrm{B}_p$, где $p \in (0, 1)$; не знаем: p;
  здесь $\mathcal{F}_\theta = \mathrm{B}_p$, $\theta = p$, $\Theta = (0, 1)$;
- знаем: $X_i \Subset \mathrm{U}_{a,b}$, где $a < b$; не знаем: a, b;
  здесь $\mathcal{F}_\theta = \mathrm{U}_{a,b}$, $\theta = \{a, b\}$, $\Theta = \{\{a, b\} : a < b\}$;
  или одно знаем, другое — нет, например:
- знаем: $X_i \Subset \mathrm{U}_{0,\theta}$, где $\theta > 0$; не знаем: θ
  здесь $\mathcal{F}_\theta = \mathrm{U}_{0,\theta}$, $\Theta = (0, \infty)$;
- знаем: $X_i \Subset \mathrm{N}_{a,\sigma^2}$, где $a \in \mathbb{R}$, $\sigma > 0$; не знаем: a, σ^2
  здесь $\mathcal{F}_\theta = \mathrm{N}_{a,\sigma^2}$, $\theta = \{a, \sigma^2\}$, $\Theta = \mathbb{R} \times (0, \infty)$;
  (или одно знаем, другое — нет).

Такая постановка имеет смысл, поскольку редко о проводимом эксперименте совсем ничего нельзя сказать. Обычно тип распределения ясен заранее, и требуется лишь указать значения параметров этого распределения.

Так, в широких предположениях рост юношей одного возраста имеет нормальное распределение (с неизвестными средним и дисперсией), а число покупателей в магазине в течение часа (не часа пик) — распределение Пуассона, и опять-таки с неизвестной «интенсивностью» λ.

2.2 Точечные оценки. Несмещенность, состоятельность оценок

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

Заметим, что все характеристики случайных величин $X_1, \ldots, X_n$ зависят от параметра θ. Так, например, для $X_i \Subset \Pi_\lambda$:

$$\mathsf{E}X_1 = \lambda, \quad \mathsf{P}(X_1 = 2) = \frac{\lambda^2}{2}e^{-\lambda}, \quad \mathsf{D}X_1 = \lambda \quad \text{и т.д.}$$

Чтобы отразить эту зависимость, будем писать $\mathsf{E}_\theta X_1$ вместо $\mathsf{E}X_1$ и т.д. Так, $\mathsf{D}_{\theta_1} X_1$ означает дисперсию, вычисленную в предположении $\theta = \theta_1$.

Упражнение. Пусть $X_1 \Subset \mathrm{B}_p$, $p_1 = 0{,}5$, $p_2 = 0{,}1$. Вычислить $\mathsf{E}_{p_1} X_1$, $\mathsf{E}_{p_2} X_1$.

Определение 2. Статистикой называется любая (измеримая!) функция $\theta^* = \theta^*(X_1, \ldots, X_n)$.

Замечание 6. Статистика есть функция от эмпирических данных, т.е. от выборки $X_1, \ldots, X_n$, но никак не от параметра θ. Статистика, как правило, предназаначена именно для оценивания неизвестного параметра θ (в этом случае ее называют «оценкой»), и уже поэтому от него зависеть не должна.

Вообще говоря, статистика есть не «любая», а «измеримая» функция от выборки, но поскольку на практике иного не бывает :), мы на это обращать внимание не будем.

Определение 3. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется несмещенной оценкой параметра θ, если для любого $\theta \in \Theta$

$$\mathsf{E}_\theta \theta^* = \theta.$$

Несмещенность — свойство оценок при фиксированном n. Означает это свойство отсутствие ошибки «в среднем», т.е. при систематическом использовании данной оценки.

Определение 4. Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется состоятельной оценкой параметра θ, если для любого $\theta \in \Theta$

$$\theta^* \xrightarrow{\mathrm{p}} \theta \text{ при } n \to \infty.$$

Свойство состоятельности означает, что оценка приближается к неизвестному параметру при увеличении количества данных. Понятно, что при отсутствии этого свойства оценка совершенно «несостоятельна» как оценка.

Пример 3. Пусть $X_1, \ldots, X_n$ — выборка объема n, все $X_i \Subset \mathrm{N}_{a,\sigma^2}$, 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 теоретического среднего, дисперсии любого распределения. И поскольку $a = \mathsf{E}_{a,\sigma^2} X_1$, возьмем в качестве оценки для a выборочное среднее: $a^* = \overline{X}$. Лемма 1 утверждает, что эта оценка несмещенная и состоятельная.

Так как $\sigma^2 = \mathsf{D}_{a,\sigma^2} X_1$, то есть даже две оценки: $\sigma^{2*} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — выборочная дисперсия и $S_0^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — несмещенная выборочная дисперсия.

Как показано в лемме 2, обе эти оценки состоятельны, и одна из них — несмещенная (*какая?*).

Но такие «способы» получения оценок нуждаются в систематизации. Заметим сразу, что приведенные ниже методы получения оценок ни из каких математических аксиом не выводятся. Они просто разумны с практической точки зрения. *Именно на их разумность следует обратить внимание.*

2.3 Методы нахождения оценок: метод моментов

Метод моментов заключается в следующем: любой момент случайной величины X_1 (например, k-й) есть функция от параметра θ. В свою очередь, параметр θ есть функция (обратная, если существует) от теоретического k-го момента. Подставив в эту функцию вместо неизвестного теоретического k-го момента его выборочный аналог, получим вместо параметра θ оценку θ^*.

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

$$\mathsf{E}_\theta X_1^k = h(\theta), \tag{3}$$

причем функция h обратима. Тогда в качестве оценки θ^* для θ берем *решение* уравнения

$$\overline{X^k} = h(\theta^*).$$

Или (что то же самое), сначала решаем уравнение (3) относительно θ, а затем *вместо истинного момента берем выборочный*:

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

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

Пример 4. Пусть $X_1, \ldots, X_n$ — выборка объема n, все $X_i \Subset \mathrm{U}_{0,\theta}$, где $\theta > 0$.
Найдем оценку метода моментов (ОММ) по первому моменту:

$$\mathsf{E}_\theta X_1 = \frac{\theta}{2} \quad \models\!\!\Rightarrow \quad \theta = 2\mathsf{E}_\theta X_1 \quad \models\!\!\Rightarrow \quad \theta_1^* = 2\overline{X}.$$

Найдем оценку метода моментов (ОММ) по k-му моменту:

$$\mathsf{E}_\theta X_1^k = \int\limits_0^\theta y^k \frac{1}{\theta}\, dy = \frac{\theta^k}{k+1} \quad \models\!\!\Rightarrow \quad \theta = \sqrt[k]{(k+1)\mathsf{E}_\theta X_1^k} \quad \models\!\!\Rightarrow \quad \theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}.$$

Проверим свойства полученных оценок.
Несмещенность:

1. По определению, $\mathsf{E}_\theta \theta_1^* = \mathsf{E}_\theta 2\overline{X} = 2\mathsf{E}_\theta \overline{X} =$ (по лемме 1) $= 2\theta/2 = \theta \quad \models\!\!\Rightarrow \quad \theta_1^* = 2\overline{X}$ — несмещенная.
2. Рассмотрим оценку θ_2^*. Заметим, что $\mathsf{E}_\theta \theta_2^* = \mathsf{E}_\theta \sqrt{3\overline{X^2}}$, тогда как $\theta = \sqrt{3\mathsf{E}_\theta X_1^2} = \sqrt{3\mathsf{E}_\theta \overline{X^2}}$ (по лемме 1).

   Равенство $\mathsf{E}_\theta \theta_2^* = \theta$ означало бы, что для с.в. $\xi = 3\overline{X^2}$ выполнено $\mathsf{E}_\theta \sqrt{\xi} = \sqrt{\mathsf{E}_\theta \xi}$, а для величины $\eta = \sqrt{\xi}$ выполнено $\mathsf{E}_\theta \eta^2 = (\mathsf{E}_\theta \eta)^2$.

   Вспомните тот единственный случай, когда это возможно, чтобы согласиться с выводом: оценка $\theta_2^* = \sqrt{3\overline{X^2}}$ — смещенная (как, впрочем, и оценки θ_k^*, $k > 2$).

Состоятельность:

1. По ЗБЧ, $\theta_1^* = 2\overline{X} \xrightarrow{\mathrm{p}} 2\mathsf{E}_\theta X_1 = 2\theta/2 = \theta \quad \models\!\!\Rightarrow \quad \theta_1^* = 2\overline{X}$ — состоятельная.
2. Заметим, что по ЗБЧ (или по лемме 3 — только для тех, кто ее доказал) при $n \to \infty$

   $$\overline{X^k} \xrightarrow{\mathrm{p}} \mathsf{E}_\theta X_1^k = \frac{\theta^k}{k+1}.$$

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

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

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

То есть вся последовательность $\{\theta_k^*\}_{k=1}^\infty = \{\sqrt[k]{(k+1)\overline{X^k}}\}$ состоит из состоятельных оценок, при этом только оценка $\theta_1^* = 2\overline{X}$ — несмещенная.

Замечание 7. *Может случиться так, что оценка $\theta^* \notin \Theta$, тогда как $\theta \in \Theta$. В этом случае в качестве оценки берут ближайшую к θ^* точку из Θ, на худой конец — из замыкания Θ, если Θ не замкнуто.*

Пример 5. Пусть $X_1, \ldots, X_n$ — выборка объема n, все $X_i \Subset \mathrm{N}_{a,1}$, где по какой-то причине $a \geqslant 0$.
Ищем оценку для a по первому моменту:

$$\mathsf{E}_a X_1 = a \quad \models\!\!\Rightarrow \quad a^* = \overline{X}.$$

Однако по условию $a \geqslant 0$, тогда как $\overline{X}$ может быть и отрицательно. Понятно, что если $\overline{X} < 0$, то в качестве оценки для положительного параметра a более подойдет 0. Если же $\overline{X} > 0$, в качестве оценки нужно брать $\overline{X}$. Итого: $a^* = \max\{0, \overline{X}\}$ — оценка метода моментов.

2.4 Состоятельность оценок метода моментов

Теорема 5. *Пусть $\theta^* = h^{-1}(\overline{X^k})$ — оценка параметра θ, полученная по методу моментов, причем функция h^{-1} непрерывна. Тогда θ^* состоятельна.*

Доказательство теоремы 5.
По лемме 3 имеем: $\overline{X^k} \xrightarrow{\text{p}} \mathsf{E}_\theta X_1^k = h(\theta)$. Поскольку функция h^{-1} непрерывна, то и

$$\theta_k^* = h^{-1}(\overline{X^k}) \xrightarrow{\text{p}} h^{-1}(\mathsf{E}_\theta X_1^k) = h^{-1}(h(\theta)) = \theta.$$

$\square$

Замечание 8. Для обратимой, т.е. взаимнооднозначной функции $h : \mathbb{R} \to \mathbb{R}$ непрерывность h и непрерывность h^{-1} эквивалентны.

2.5 Методы нахождения оценок: метод максимального правдоподобия

Метод максимального правдоподобия — еще 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.

Решим сначала, что такое «вероятность получить данную выборку», т.е. что 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. А для дискретных распределений $\mathcal{F}_\theta$ вероятность попасть в точку y равна $\mathsf{P}_\theta(X_1 = y)$. И то, и другое мы будем называть плотностью распределения $\mathcal{F}_\theta$ (отступая при этом, для простоты, от терминологии теории вероятностей). Итак,

Определение 5. Функцию

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

мы будем называть *плотностью* распределения $\mathcal{F}_\theta$.

Определение 6. Функция (вообще говоря, случайная величина)

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

называется *функцией правдоподобия*. Функция (тоже случайная)

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

называется *логарифмической* функцией правдоподобия.

Если наша выборка $\vec{X}$ в данной серии экспериментов приняла значения $(x_1, \ldots, x_n)$, то (в дискретном случае) функция правдоподобия и есть вероятность этого события, разная при разных значениях параметра θ:

$$\Psi(\vec{x}, \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \mathsf{P}_\theta(X_1 = x_1) \cdot \ldots \cdot \mathsf{P}_\theta(X_n = x_n) \overset{\text{независ-ть}}{=} \mathsf{P}_\theta(X_1 = x_1, \ldots, X_n = x_n).$$

Определение 7. *Оценкой максимального правдоподобия* $\widehat{\theta}$ неизвестного параметра θ называют значение θ, при котором функция $\Psi(\vec{X}, \theta)$ достигает максимума (как функция от θ при фиксированных $X_1, \ldots, X_n$):

$$\widehat{\theta} = \text{ точка максимума (по переменной } \theta) \text{ функции } \Psi(\vec{X}, \theta).$$

Замечание 9. Поскольку функция $\ln y$ монотонна, то точки максимума $\Psi(\vec{X},\theta)$ и $L(\vec{X},\theta)$ совпадают. Поэтому оценкой максимального правдоподобия (ОМП) можно называть точку максимума (по θ) функции $L(\vec{x},\theta)$:

$$\widehat{\theta} = \text{ точка максимума (по переменной } \theta) \text{ функции } L(\vec{X},\theta).$$

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

Пример 6. Пусть $X_1,\ldots,X_n$ — выборка объема n из распределения Пуассона: $X_i \Subset \Pi_\lambda$, где $\lambda > 0$. Найдем ОМП $\widehat{\lambda}$ неизвестного параметра λ.

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

$$\Psi(\vec{X},\lambda) = \prod_{i=1}^n \frac{\lambda^{X_i}}{X_i!}e^{-\lambda} = \frac{\lambda^{\Sigma_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}e^{-n\lambda} = \frac{\lambda^{n\overline{X}}}{\prod_{i=1}^n X_i!}e^{-n\lambda}.$$

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:

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

$\dfrac{\partial}{\partial\lambda}L(\vec{X},\lambda) = \dfrac{n\overline{X}}{\lambda} - n$, и точка экстремума $\widehat{\lambda}$ — решение уравнения: $\dfrac{n\overline{X}}{\lambda} - n = 0$, то есть $\widehat{\lambda} = \overline{X}$.

Упражнение.

1) Убедиться, что $\widehat{\lambda} = \overline{X}$ — точка максимума, а не минимума.
2) Убедиться, что $\widehat{\lambda} = \overline{X}$ совпадает с одной из оценок метода моментов (*по какому моменту?*).

Пример 7. Пусть $X_1,\ldots,X_n$ — выборка объема n из нормального распределения N_{a,σ^2}, где $a \in \mathbb{R}$, $\sigma > 0$; и оба параметра a, σ^2 неизвестны.

Выпишем плотность, функцию правдоподобия и логарифмическую функцию правдоподобия. Плотность:

$$f_{(a,\sigma^2)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(y-a)^2}{2\sigma^2}\right),$$

функция правдоподобия:

$$\Psi(\vec{X},a,\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(X_i-a)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{\sum_{i=1}^n (X_i-a)^2}{2\sigma^2}\right),$$

логарифмическая функция правдоподобия:

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

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

$$\frac{\partial}{\partial a}L(\vec{X},a,\sigma^2) = \frac{2\sum_{i=1}^n (X_i-a)}{2\sigma^2} = \frac{n\overline{X}-na}{\sigma^2}; \quad \frac{\partial}{\partial\sigma^2}L(\vec{X},a,\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i-a)^2}{2\sigma^4}.$$

Оценка максимального правдоподобия $(\widehat{a},\widehat{\sigma^2})$ для (a,σ^2) — решение системы уравнений

$$\frac{n\overline{X}-na}{\sigma^2} = 0; \qquad -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (X_i-a)^2}{2(\sigma^2)^2} = 0.$$

Решая, получим хорошо знакомые оценки:

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

Упражнение.

1) Убедиться, что $\widehat{a} = \overline{X}, \widehat{\sigma^2} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — точка максимума, а не минимума.

2) Убедиться, что эти оценки совпадают с некоторыми оценками метода моментов.

Пример 8. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Тогда $\widehat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$ (см. [1], пример 2.5, с.10).

Пример 9. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{\theta,\theta+5}$, где $\theta \in \mathbb{R}$ (см. также [2], пример 4, с.48).

Выпишем плотность и функцию правдоподобия. Плотность:

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

функция правдоподобия:

$$\Psi(\vec{X}, \theta) = \begin{cases} (1/5)^n, & \text{если все } X_i \in [\theta, \theta+5] \\ 0 & \text{иначе} \end{cases} = \begin{cases} (1/5)^n, & \text{если } \theta \leqslant X_{(1)} \leqslant X_{(n)} \leqslant \theta+5 \\ 0 & \text{иначе} \end{cases} =$$

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

Функция правдоподобия достигает своего максимального значения $(1/5)^n$ во всех точках $\theta \in [X_{(n)} - 5, X_{(1)}]$. График этой функции изображен на рис. 9.

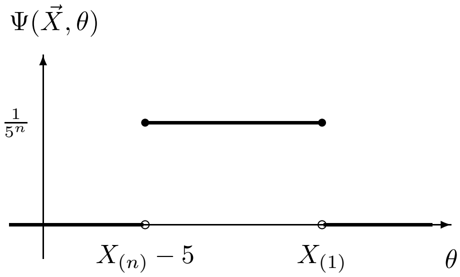


Рис. 3: Пример 9.

Любая точка $\widehat{\theta} \in [X_{(n)} - 5, X_{(1)}]$ может служить оценкой максимального правдоподобия. Получаем более чем счетное число оценок вида $\widehat{\theta}_\alpha = (1-\alpha)(X_{(n)} - 5) + \alpha X_{(1)}$, при разных $\alpha \in [0,1]$, в том числе $\widehat{\theta}_0 = X_{(n)} - 5$, $\widehat{\theta}_1 = X_{(1)}$ — концы отрезка.

Упражнение.

1) Убедиться, что отрезок $[X_{(n)} - 5, X_{(1)}]$ не пуст.
2) Найти оценку метода моментов (по первому моменту) и убедиться, что она отличается от ОМП.
3) Найти ОМП параметра θ равномерного распределения $\mathrm{U}_{\theta,\theta+3}$.

2.6 Вопросы и упражнения

1. Задачник [1], задачи $2.1 - 2.16$.

2. Дана выборка $X_1, \ldots, X_n, X_i \Subset \mathrm{B}_p$, $p \in (0,1)$ — неизвестный параметр. Проверить, что X_1, X_1X_2, $X_1(1-X_2)$ являются несмещенными оценками соответственно для p, p^2, $p(1-p)$. Являются ли эти оценки состоятельными?

3. Дана выборка $X_1, \ldots, X_n, X_i \Subset \Pi_\lambda$, $\lambda > 0$ — неизвестный параметр. Проверить, что X_1 и $\mathbf{I}(X_1 = k)$ являются несмещенными оценками соответственно для λ и $\dfrac{\lambda^k}{k!}e^{-\lambda}$. Являются ли эти оценки состоятельными?

4. Дана выборка $X_1, \ldots, X_n, X_i \Subset \mathrm{U}_{0,\theta}$, $\theta > 0$ — неизвестный параметр. Проверить состоятельность и несмещенность оценок: $\theta_1^* = X_{(n)}$, $\theta_2^* = 2\overline{X}$, $\theta_3^* = X_{(n)} + X_{(1)}$.

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

6. Какие из оценок в задаче 5 несмещенные? состоятельные?

7. Сравнить вид оценок для параметра α, полученных по первому моменту в задачах 5(г) и 5(д). Доказать, что среди них только одна несмещенная. **Указание.** Использовать свойство: если $\xi \geqslant 0$ п.н., то $\mathsf{E}\dfrac{1}{\xi} = \dfrac{1}{\mathsf{E}\xi} \iff \xi = \mathrm{const}$ п.н.

8*. Доказать свойство из задачи 7, используя неравенство Коши-Буняковского-Шварца: $\mathsf{E}|\xi\eta| \leqslant \sqrt{\mathsf{E}\xi^2\mathsf{E}\eta^2}$, которое обращается в равенство $\iff |\xi| = c|\eta|$ (см. доказательство свойства коэффициента корреляции $|\rho| \leqslant 1$ в курсе теории вероятностей).

9. Построить оценки неизвестных параметров по методу максимального правдоподобия для следующих распределений: а) B_p, б)$\Pi_{\lambda+1}$, в) $\mathrm{U}_{0,2\theta}$, г) $\mathrm{E}_{2\alpha+3}$, д) $\mathrm{U}_{-\theta,\theta}$, е) N_{a,σ^2} (a известно).

10. Какие из оценок в задаче 9 несмещенные? состоятельные?

3 Сравнение оценок

3.1 Способы сравнения оценок

Используя метод моментов и метод максимального правдоподобия, мы получили для каждого параметра уже 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, так что эти величины обычно сравнить нельзя. Как, например, сравнивать $|\overline{X} - \theta|$ и $|\overline{X^k} - \theta|$? Или, на одном элементарном исходе, $|2.15 - \theta|$ и $|3.1 - \theta|$?

Поэтому имеет смысл сравнивать не отклонения как таковые, а *средние значения* этих отклонений, то есть $\mathsf{E}_\theta|\theta^* - \theta|$.

Но математическое ожидание модуля с.в. считать обычно затруднительно, поэтому более удобной характеристикой для сравнения оценок считается $\mathsf{E}_\theta(\theta^* - \theta)^2$. Она удобна еще и тем, что очень чутко реагирует на маловероятные, но большие по абсолютному значению отклонения θ^* от θ (возводит их в квадрат).

Заметим еще, что $\mathsf{E}_\theta(\theta^* - \theta)^2$ есть функция от θ, так что сравнивать эти «среднеквадратические» отклонения нужно как функции от θ — поточечно. Такой подход к сравнению оценок называется *среднеквадратическим*.

Разумеется, в зависимости от потребностей исследователя можно пользоваться и другими характеристиками, например, $\mathsf{E}_\theta(\theta^* - \theta)^4$ или $\mathsf{E}_\theta|\theta^* - \theta|$.

Существует и так называемый *асимптотический* подход к сравнению оценок, при котором для сравнения оценок используется некая характеристика «разброса» оценки относительно параметра при больших n.

3.2 Среднеквадратический подход. Эффективность оценок

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

Определение 8. Говорят, что оценка θ_1^* *лучше* оценки θ_2^* в смысле среднеквадратического подхода, если для любого $\theta \in \Theta$

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

и хотя бы при одном θ это неравенство строгое.

Оценки могут быть несравнимы: например, при некоторых θ $\mathsf{E}_\theta(\theta_1^* - \theta)^2 \leqslant \mathsf{E}_\theta(\theta_2^* - \theta)^2$, при других θ — наоборот.

Существует ли среди всех оценок наилучшая в смысле среднеквадратического подхода? Скептик со склонностью к философии сразу ответит «нет». Докажем, что он прав.

Теорема 6. *В классе всех возможных оценок наилучшей оценки в смысле среднеквадратического подхода не существует.*

Доказательство теоремы 6. Пусть, напротив, θ^* — наилучшая, то есть для любой другой оценки θ_1^*, при любом $\theta \in \Theta$

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

Пусть θ_1 — произвольная точка Θ. Рассмотрим оценку (статистику) $\theta_1^* \equiv \theta_1$. Тогда

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

при любом $\theta \in \Theta$. Возьмем $\theta = \theta_1 \in \Theta$ и получим следующее неравенство:

$$\mathsf{E}_{\theta_1}(\theta^* - \theta_1)^2 \leqslant \mathsf{E}_{\theta_1}(\theta_1 - \theta_1)^2 = 0.$$

Поэтому $\mathsf{E}_{\theta_1}(\theta^* - \theta_1)^2 = 0$. В силу произвольности θ_1 это выполнено при любом $\theta \in \Theta$:

$$\mathsf{E}_\theta(\theta^* - \theta)^2 \equiv 0.$$

Но это возможно только если $\theta^* \equiv \theta$ (оценка в точности отгадывает неизвестный параметр). То есть θ^* даже не является статистикой. Такого типа примеры привести можно (например, по выборке из $\mathrm{U}_{\theta,\theta+1}, \theta \in \mathbb{Z}$ можно точно указать θ), но математической статистике здесь делать нечего. □

Упражнение. Объяснить словесно доказательство теоремы 6.

Если в очень широком классе всех оценок наилучшей не существует, то, возможно, следует сузить класс рассматриваемых оценок (или разбить класс всех оценок на отдельные подклассы и в каждом искать наилучшую).

Обычно рассматривают оценки, имеющие одинаковое *смещение* $b(\theta) = \mathsf{E}_\theta\theta^* - \theta$. Обозначим через K_b класс оценок, имеющих смещение $b(\theta)$:

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

Здесь K_0 — класс несмещенных оценок.

Определение 9. 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. То есть для любой $\theta_1^* \in K_b$, для любого $\theta \in \Theta$

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

Определение 10. Эффективная оценка в классе K_0 (несмещенных оценок) называется просто *эффективной*.

Замечание 10. Для $\theta^* \in K_0$, по определению дисперсии, $\mathsf{E}_\theta(\theta^* - \theta)^2 = \mathsf{E}_\theta(\theta^* - \mathsf{E}_\theta\theta^*)^2 = \mathsf{D}_\theta\theta^*$, так что сравнение в среднеквадратичном несмещенных оценок — это сравнение их дисперсий. Поэтому эффективную оценку (в классе K_0) часто называют «несмещенной оценкой с равномерно минимальной дисперсией». Равномерность подразумевается по всем $\theta \in \Theta$. Для $\theta^* \in K_b$

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

так что сравнение в среднеквадратичном оценок с одинаковым смещением — это также сравнение их дисперсий.

Упражнение. Мы собираемся искать наилучшую оценку в классе K_b. Объясните, почему доказательство теоремы 4 не пройдет в классе K_b.

3.3 Единственность эффективной оценки в классе с фиксированным смещением

Теорема 7. *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:* $\mathsf{P}_\theta(\theta_1^* = \theta_2^*) = 1$.

***Доказательство теоремы* 7.** Заметим сначала, что $\mathsf{E}_\theta(\theta_1^* - \theta)^2 = \mathsf{E}_\theta(\theta_2^* - \theta)^2$. Действительно, так как θ_1^* эффективна в классе K_b, то она не хуже оценки θ_2^*, то есть

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

и наоборот. Поэтому $\mathsf{E}_\theta(\theta_1^* - \theta)^2 = \mathsf{E}_\theta(\theta_2^* - \theta)^2$.

Рассмотрим оценку $\theta^* = \dfrac{\theta_1^* + \theta_2^*}{2} \in K_b$ (доказать!). Вычислим ее среднеквадратическое отклонение. Заметим, что

$$\left(\frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2 = \frac{a^2+b^2}{2}. \tag{4}$$

Положим $a = \theta_1^* - \theta$, $b = \theta_2^* - \theta$. Тогда $(a+b)/2 = \theta^* - \theta$, $a - b = \theta_1^* - \theta_2^*$. Подставим эти выражения в (4) и возьмем математические ожидания обеих частей:

$$\mathsf{E}_\theta(\theta^* - \theta)^2 + \mathsf{E}_\theta\left(\frac{\theta_1^* - \theta_2^*}{2}\right)^2 = \mathsf{E}_\theta\frac{(\theta_1^* - \theta)^2 + (\theta_2^* - \theta)^2}{2} = \mathsf{E}_\theta(\theta_1^* - \theta)^2 = \mathsf{E}_\theta(\theta_2^* - \theta)^2. \tag{5}$$

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

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

Сравнивая это неравенство с равенством (5), видим, что

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

Тогда (*почему?*) $\mathsf{P}_\theta(\theta_1^* = \theta_2^*) = 1$, 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. А способами поиска наилучшей в целом классе мы тоже скоро займемся.

Пример 10. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Тогда $\widehat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$ — оценка максимального правдоподобия, а $\theta^* = 2\overline{X}$ — оценка метода моментов, полученная по первому моменту. Сравним их в среднеквадратичном. Оценка $\theta^* = 2\overline{X}$ несмещенная (см. пример 4), поэтому

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

Для $\widehat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$ имеем

$$\mathsf{E}_\theta(\widehat{\theta} - \theta)^2 = \mathsf{E}_\theta\widehat{\theta}^2 - 2\theta\mathsf{E}_\theta\widehat{\theta} + \theta^2.$$

Посчитаем первый и второй момент случайной величины $\widehat{\theta} = X_{(n)}$. Найдем (полезно вспомнить, как это делалось в прошлом семестре!) функцию распределения и плотность $\widehat{\theta}$:

$$\mathsf{P}_\theta(X_{(n)} < y) = \mathsf{P}_\theta(X_1 < y)^n = \begin{cases} 0, & y < 0 \\ \dfrac{y^n}{\theta^n}, & y \in [0, \theta] \\ 1, & y > \theta, \end{cases} \qquad f_{X_{(n)}}(y) = \begin{cases} 0, & y \notin [0, \theta] \\ n\dfrac{y^{n-1}}{\theta^n}, & y \in [0, \theta] \end{cases}.$$

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

Поэтому

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

При $n = 1$ квадратические отклонения равны, а при $n > 1$

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

то есть $X_{(n)}$ лучше, чем $2\overline{X}$. При этом $\mathsf{E}_\theta(X_{(n)} - \theta)^2 \to 0$ со скоростью n^{-2}, тогда как $\mathsf{E}_\theta(2\overline{X} - \theta)^2 \to 0$ со скоростью n^{-1}.

Упражнение.

1. Доказать, что $X_{(n)} \in K_b$, где $b(\theta) = -\dfrac{\theta}{n+1}$.
2. Доказать, что $\dfrac{n+1}{n}X_{(n)} \in K_0$ (несмещенная).
3. Сравнить оценки $\dfrac{n+1}{n}X_{(n)}$ и $X_{(n)}$ в среднеквадратичном.

3.4 Асимптотически нормальные оценки (АНО)

Для того, чтобы уметь сравнивать оценки вида $\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}$ (см. пример 4), среднеквадратического подхода недостаточно: второй момент такой случайной величины посчитать вряд ли удастся. Оценки такого вида (функции от сумм) удается сравнивать с помощью асимптотического подхода. Более точно, этот подход применим к так называемым «асимптотически нормальным» оценкам.

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

Определение 11. Оценка θ^* называется *асимптотически нормальной оценкой* параметра θ с коэффициентом $\sigma^2(\theta)$, если

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

Пример 11. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Проверим, являются ли оценки $\theta^* = 2\overline{X}$ и $\widehat{\theta} = X_{(n)}$ асимптотически нормальными (АНО). По ЦПТ,

$$\sqrt{n}(\theta^* - \theta) = \sqrt{n}(2\overline{X} - \theta) = \sqrt{n}\left(2\frac{\sum\limits_{i=1}^n X_i}{n} - \theta\right) = \frac{\sum\limits_{i=1}^n 2X_i - n\theta}{\sqrt{n}} =$$

$$= \frac{\sum\limits_{i=1}^n 2X_i - n\mathsf{E}_\theta 2X_1}{\sqrt{n}} \Rightarrow \mathrm{N}_{0,\mathsf{D}_\theta 2X_1} = \mathrm{N}_{0,4\mathsf{D}_\theta X_1}.$$

То есть оценка $\theta^* = 2\overline{X}$ асимптотически нормальна с коэффициентом $\sigma^2(\theta) = 4\mathsf{D}_\theta X_1 = 4\theta^2/12 = \theta^2/3$.

Для оценки $\widehat{\theta} = X_{(n)}$ имеем:

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}(X_{(n)} - \theta) < 0 \text{ с вероятностью } 1. \tag{6}$$

По определению, $\xi_n \Rightarrow \xi$, если для любой точки x, являющейся точкой непрерывности функции распределения F_ξ, имеет место сходимость $F_{\xi_n}(x) = \mathsf{P}(\xi_n < x) \to F_\xi(x) = \mathsf{P}(\xi < x)$.

Но $\mathsf{P}_\theta(\sqrt{n}(X_{(n)} - \theta) < 0) = 1$, тогда как для $\xi \Subset \mathrm{N}_{0,\sigma^2(\theta)}$ функция распределения в нуле равна $\Phi_{0,\sigma^2(\theta)}(0) = \mathsf{P}(\xi < 0) = 0.5$, то есть слабая сходимость места не имеет.

Таким образом, оценка $\widehat{\theta} = X_{(n)}$ асимптотически нормальной не является.

Осталось ответить на напрашивающиеся вопросы:

1) Куда все 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 ф.р. нуля во всех точках непрерывности последней. Не забудьте о существовании *замечательных пределов, логарифмов и ряда Тейлора*.

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

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

Порядок действий: прежний.

3) Для оценки $\dfrac{n+1}{n}X_{(n)}$ свойство (6) не выполнено. Может ли эта оценка быть АНО?

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

4) Плохо ли, что оценка $\widehat{\theta} = X_{(n)}$ не асимптотически нормальна? Может быть, сходимость $n(X_{(n)} - \theta) \Rightarrow \eta$ еще лучше?

Попробуем ответить на последний вопрос.

3.5 «Скорость» сходимости оценки к параметру

Теорема 8. *Если θ^* — асимптотически нормальная оценка параметра θ, то θ^* состоятельна.*

Доказательство теоремы 8.
Заметим (доказать!), что если $\sqrt{n}(\theta^* - \theta) \Rightarrow \xi \Subset \mathrm{N}_{0,\sigma^2(\theta)}$, то $\sqrt{n}|\theta^* - \theta| \Rightarrow |\xi|$.
Поэтому при $n \to \infty$

$$\mathsf{P}_\theta(|\theta^* - \theta| < \varepsilon) = \mathsf{P}_\theta(\sqrt{n}|\theta^* - \theta| < \sqrt{n}\varepsilon) \sim \mathsf{P}(|\xi| < \sqrt{n}\varepsilon) \longrightarrow \mathsf{P}(|\xi| < \infty) = 1.$$

$\square$

Упражнение. Верно ли утверждение теоремы 8, если предельная величина ξ имеет распределение, отличное от нормального?

Таким образом, если θ^* асимптотически нормальна, то $\theta^* \xrightarrow{\mathrm{p}} \theta$, или $\theta^* - \theta \xrightarrow{\mathrm{p}} 0$. Свойство асимптотической нормальности показывает, в частности, что скорость этой сходимости имеет порядок $\dfrac{1}{\sqrt{n}}$ (расстояние между θ^* и θ ведет себя как $\dfrac{1}{\sqrt{n}}$):

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

Взглянем с этой точки зрения на оценку $\widehat{\theta} = X_{(n)}$ в примере 11. Для нее (и для тех, кто справился с упражнениями)

$$n(X_{(n)} - \theta) \Rightarrow \eta,$$

где η — некоторая случайная величина. Иначе говоря, расстояние между $\widehat{\theta}$ и θ ведет себя как $\dfrac{1}{n}$.

Упражнение. Лучше это или хуже?

3.6 Асимптотическая нормальность и ЦПТ

В примере 11 мы видели, что для оценок типа $2\overline{X}$ свойство асимптотической нормальности сразу следует из ЦПТ (см. также задачу 6 к лекции 1).

Установим асимптотическую нормальность оценок более сложного вида (функций от сумм X_i и сумм функций от X_i).

Лемма 4. *Пусть функция g такова, что $0 \neq \mathsf{D}_\theta g(X_1) < \infty$. Тогда оценка $\overline{g(X)} = \dfrac{\sum_1^n g(X_i)}{n}$ является асимптотически нормальной оценкой для $\mathsf{E}_\theta g(X_1)$ с коэффициентом $\sigma^2(\theta) = \mathsf{D}_\theta g(X_1)$:*

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

Упражнение. Вспомнить ЦПТ и доказать лемму 4.
Упражнение. Получить решение задачи 6 (после главы 1) в качестве следствия леммы 4.

3.7 Асимптотическая нормальность оценок вида $H(\overline{g(X)})$

Следующая теорема утверждает асимптотическую нормальность оценок вида

$$\theta^* = H(\overline{g(X)}) = H\left(\frac{\sum_1^n g(X_i)}{n}\right).$$

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

Теорема 9. *Пусть функция g такова, что $0 \neq \mathsf{D}_\theta g(X_1) < \infty$, функция H непрерывно дифференцируема в точке $a = \mathsf{E}_\theta g(X_1)$, и $H'(a) \neq 0$. Тогда оценка $\theta^* = H(\overline{g(X)})$ является асимптотически нормальной оценкой для $\theta = H(\mathsf{E}_\theta g(X_1)) = H(a)$ с коэффициентом $\sigma^2(\theta) = (H'(a))^2\, \mathsf{D}_\theta g(X_1)$.*

***Доказательство теоремы* 9.** Разложим $H(\overline{g(X)})$ в ряд Тейлора в точке a:

$$H(\overline{g(X)}) - H(a) = H'(a)(\overline{g(X)} - a) + \delta_n,$$

где $\delta_n = O((\overline{g(X)} - a)^2) \xrightarrow{\text{p}} 0$ при $n \to \infty$. Последнее верно, так как по ЗБЧ при $n \to \infty$

$$\overline{g(X)} = \frac{\sum_1^n g(X_i)}{n} \xrightarrow{\text{p}} \mathsf{E}_\theta g(X_1) = a.$$

Вспомним свойства слабой сходимости:

1) если $\xi_n \Rightarrow \xi$ и $\eta_n \xrightarrow{\text{p}} c =$const, то $\xi_n\eta_n \Rightarrow c\xi$;
2) если $\xi_n \Rightarrow \xi$ и $\eta_n \xrightarrow{\text{p}} c =$const, то $\xi_n + \eta_n \Rightarrow \xi + c$.

По лемме 4, $\sqrt{n}\big(\overline{g(X)} - a\big) \Rightarrow \xi \Subset \mathrm{N}_{0,\mathsf{D}_\theta g(X_1)}$, и по свойству (1) слабой сходимости

$$\sqrt{n}\delta_n = O(\sqrt{n}(\overline{g(X)} - a)(\overline{g(X)} - a)) \xrightarrow{\text{p}} 0 \cdot \xi = 0.$$

Отсюда (и по свойству (2) слабой сходимости)

$$\sqrt{n}\big(H(\overline{g(X)}) - H(a)\big) = \sqrt{n}H'(a)\big(\overline{g(X)} - a\big) + \sqrt{n}\delta_n \Rightarrow \mathrm{N}_{0,(H'(a))^2\,\mathsf{D}_\theta g(X_1)},$$

что и требовалось доказать. □

Пример 12. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{0,\theta}$, где $\theta > 0$. Проверим, являются ли оценки $\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}}$, $k = 1, 2, \ldots$, полученные методом моментов, асимптотически нормальными.

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

$$\theta_k^* = \sqrt[k]{(k+1)\overline{X^k}} = \sqrt[k]{\frac{\sum_1^n (k+1)X_i^k}{n}} = H\left(\frac{\sum_1^n g(X_i)}{n}\right).$$

При этом

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

Впрочем, иначе быть не могло по определению метода моментов (*верно?*). Проверим другие условия теоремы 9:

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

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

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

$$H'(y) = \frac{1}{k}y^{\frac{1-k}{k}},\ H'(a) = H'(\theta^k) = \frac{1}{k}\theta^{1-k} \text{ непрерывна для } \theta > 0.$$

По теореме 9, оценка θ_k^* — АНО для θ с коэффициентом

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

В том числе для $\theta_1^* = 2\overline{X}$ имеем коэффициент $\sigma_1^2(\theta) = \dfrac{\theta^2}{3}$ (см. пример 11).

Осталось понять, при чем тут сравнение оценок и что показывает коэффициент асимптотической нормальности.

3.8 Асимптотический подход к сравнению оценок

Возьмем две случайные величины: $\xi \Subset \mathrm{N}_{0,1}$ и $10\,\xi \Subset \mathrm{N}_{0,100}$. Если для ξ, например, $0,997 = \mathsf{P}(|\xi| < 3)$, то для $10\,\xi$ уже $0,997 = \mathsf{P}(|\xi| < 30)$. Разброс значений величины $10\,\xi$ гораздо больший, и дисперсия (показатель рассеяния) соответственно больше.

Что показывает коэффициент асимптотической нормальности? Возьмем две АНО с коэффициентами 1 и 100:

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

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. Отсюда — естественный способ сравнения асимптотически нормальных оценок:

Определение 12. 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 12 (продолжение). Сравним между собой оценки в последовательности $\theta_1^*, \theta_2^*, \ldots$. Для θ_k^* коэффициент асимптотической нормальности имеет вид

$$\sigma_k^2(\theta) = \frac{\theta^2}{2k+1} \to 0 \ \text{ при } \ k \to \infty.$$

Коэффициент тем меньше, чем больше k, то есть каждая следующая оценка в этой последовательности лучше предыдущей.

Оценка θ_∞^*, являющаяся «последней», могла бы быть лучше всех оценок в этой последовательности в смысле асимптотического подхода, если бы являлась асимптотически нормальной. Увы:

Упражнение. (См. задачу 8 к главе 1). Доказать, что $\theta_k^* \to X_{(n)}$ (поточечно), то есть для любого элементарного исхода ω при $k \to \infty$

$$\sqrt[k]{(k+1)\frac{\sum_1^n X_i^k(\omega)}{n}} \to \max\{X_1(\omega), \ldots, X_n(\omega)\}.$$

Упражнение.* Можно ли придать некий смысл фразе: «оценка $\widehat{\theta} = X_{(n)}$ асимптотически нормальна с коэффициентом 0»? Какой? И зачем?

3.9 Вопросы и упражнения

1. Задачник [1], задачи 7.5, 4.4 – 4.7, 4.9.

2. Пусть $X_1, \ldots, X_n$ — выборка объема n из равномерного распределения $\mathrm{U}_{\theta,\theta+5}$, где $\theta \in \mathbb{R}$. Сравнить оценки $\widehat{\theta}_0 = X_{(n)} - 5$, $\widehat{\theta}_1 = X_{(1)}$ (см. пример 9) в среднеквадратичном. Сравнить с этими оценками оценку метода моментов $\theta^* = \overline{X} - 2{,}5$.

3. Доказать, что в условиях предыдущей задачи оценка θ^* асимптотически нормальна. Вычислить коэффициент. Доказать, что $\widehat{\theta}_0$ и $\widehat{\theta}_1$ АНО не являются.

4. Для показательного распределения с параметром α оценка, полученная методом моментов по k-му моменту, имеет вид: $\alpha_k^* = \sqrt[k]{\dfrac{k!}{\overline{X^k}}}$ (задача 5(г) после главы 2). Сравнить оценки α_k^*, $k = 1, 2, \ldots$ в смысле асимптотического подхода. Доказать, что оценка α_1^* наилучшая.

4 Эффективные оценки

Вернемся к сравнению оценок в смысле среднеквадратического подхода. Мы договорились в классе одинаково смещенных оценок называть эффективной оценку с наименьшим среднеквадратическим отклонением (или наименьшей дисперсией). Но попарное сравнение оценок — далеко не лучший способ отыскания эффективной оценки. Сегодня мы познакомимся с утверждением, позволяющим *доказать* эффективность оценки (если, конечно, она на самом деле эффективна). Это утверждение называется неравенством Рао-Крамера и говорит о том, что в классе K_b существует нижняя граница для среднеквадратического отклонения $\mathsf{E}_\theta(\theta^* - \theta)^2$ любой оценки.

Таким образом, если найдется оценка, отклонение которой в точности равно этой нижней границе (самое маленькое), то данная оценка — эффективна, поскольку у прочих оценок отклонение меньше быть не может.

К сожалению, данное неравенство верно не для всех распределений, а лишь для так называемых «регулярных» семейств распределений, к которым **не** относится, например, большинство равномерных.

4.1 Условия регулярности

Пусть $X_1, \ldots, X_n$ — выборка объема n из параметрического семейства распределений $\mathcal{F}_\theta$, $\theta \in \Theta$. Пусть $f_\theta(y)$ — плотность $\mathcal{F}_\theta$ (в смысле определения 5).

Следующее условие назовем условием регулярности.

(R) *Для почти всех* $y \in \mathbb{R}$ *функция* $\sqrt{f_\theta(y)}$ *непрерывно дифференцируема по* θ *во всех точках* $\theta \in \Theta$.

Замечание 11. Термин «для почти всех y» означает «для всех y, за исключением (возможно) $y \in A$, где $\mathsf{P}_\theta(X_1 \in A) = 0$». В частности, для дискретного распределения Бернулли $\mathsf{P}_\theta(X_1 \in A) = 0$ для всех A, не содержащих 0 и 1. Термин «функция непрерывно дифференцируема по θ» означает, что функция непрерывна и дифференцируема по θ, и ее производная (частная) также непрерывна по θ.

4.2 «Регулярные» и «нерегулярные» семейства распределений

Пример 13. Регулярное семейство. Рассмотрим показательное распределение E_α с параметром $\alpha > 0$. Плотность этого распределения имеет вид

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

При любом $y \neq 0$ существует производная по α, и эта производная непрерывна во всех точках $\alpha > 0$:

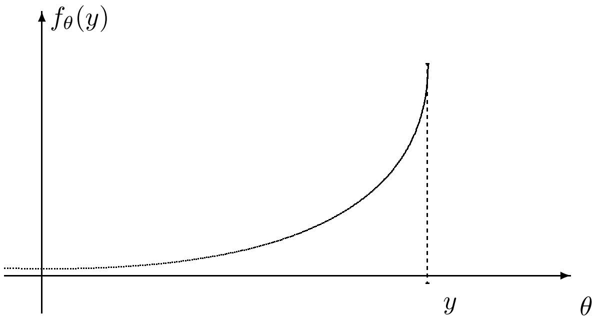
$$\frac{\partial}{\partial \alpha} \sqrt{f_\alpha(y)} = \begin{cases} \dfrac{1}{2\sqrt{\alpha}} e^{-\alpha y/2} - \sqrt{\alpha} \dfrac{y}{2} e^{-\alpha y/2}, & \text{если } y > 0, \\ 0, & \text{если } y < 0. \end{cases}$$

Пример 14. Нерегулярное семейство. Рассмотрим равномерное распределение $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Плотность этого распределения имеет вид

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

При фиксированном $y > 0$ изобразим функцию $f_\theta(y)$ (или ее корень — масштаб не соблюден) как функцию переменной θ.

Видим, что какое бы $y > 0$ ни было, $f_\theta(y)$ даже не является непрерывной по θ, а тем более дифференцируемой.



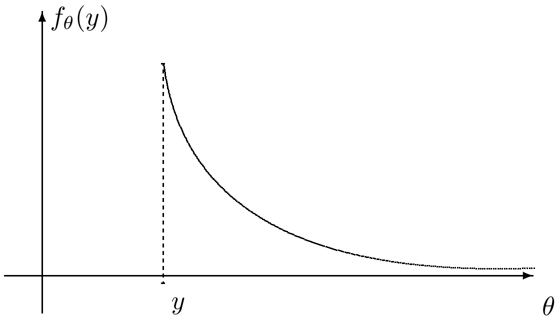


Рис. 4: Пример 15. Пример 14.

Пример 15. Нерегулярное семейство. Рассмотрим «сдвинутое» показательное распределение с параметром сдвига $\theta \in \mathbb{R}$ и плотностью

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

При фиксированном $y \in \mathbb{R}$ изобразим функцию $f_\theta(y)$ (или ее корень) как функцию переменной θ.

Видим, что какое бы y ни было, $f_\theta(y)$ даже не является непрерывной по θ, а тем более дифференцируемой.

Замечание 12. Вместо непрерывной дифференцируемости $\sqrt{f_\theta(y)}$ можно требовать того же от $\ln f_\theta(y)$.

4.3 Неравенство Рао-Крамера

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

Пусть, кроме того, выполнено условие

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

При выполнении условий (R) и (RR) справедливо следующее утверждение.

Теорема 10 («Неравенство Рао-Крамера»). *Пусть* $\theta^* \in K_0$, *и* $\mathsf{D}_\theta \theta^*$ *ограничена на компактах (то есть для любого компакта* $\Theta_0 \subset \Theta$ *найдется постоянная* C *такая, что* $\mathsf{D}_\theta \theta^* \leqslant C$ *для всех* $\theta \in \Theta_0$*). Тогда*

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

Упражнение. Проверить, что для показательного семейства E_α с параметром $\alpha > 0$ дисперсия $\mathsf{D}_\alpha X_1$ не ограничена глобально при $\alpha > 0$, но ограничена на компактах.

Неравенство сформулировано для класса несмещенных оценок. В классе оценок с ненулевым смещением $b(\theta)$ неравенство Рао-Крамера выглядит так:

Теорема 11 («Неравенство Рао-Крамера»). *Пусть* $\theta^* \in K_b$ *и* $\mathsf{D}_\theta \theta^*$ *ограничена на компактах. Тогда (при выполнении условий (R) и (RR))*

$$\mathsf{E}_\theta (\theta^* - \theta)^2 \geqslant \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta) \quad \text{или} \quad \mathsf{D}_\theta \theta^* \geqslant \frac{(1 + b'(\theta))^2}{nI(\theta)}.$$

Замечание 13. Доказательство неравенства Рао-Крамера при первом, втором и третьем прочтении можно опустить.

Доказательство неравенства Рао-Крамера. Мы докажем только неравенство для класса K_0 (теорему 10). Необходимые изменения в доказательство для класса K_b читатель может внести самостоятельно.

Нам понадобится следующее утверждение.

Лемма 5. *При выполнении условий (R) и (RR) для любой статистики $T = T(\vec{X})$, дисперсия которой ограничена на компактах,*

$$\frac{\partial}{\partial\theta}\mathsf{E}_\theta T = \mathsf{E}_\theta T \cdot \frac{\partial}{\partial\theta}L(\vec{X},\theta).$$

Упражнение. Вспомнить, что такое функция правдоподобия (Ψ) (определение 6), логарифмическая функция правдоподобия (L) (определение 7) и как они связаны друг с другом, с плотностью X_1 и совместной плотностью выборки.

***Доказательство леммы* 5.** Напоминание: математическое ожидание функции от нескольких случайных величин есть (многомерный) интеграл от этой функции помноженной на совместную плотность этих с.в.

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

В следующей цепочке равенств равенство, помеченное $(?)$, мы доказывать не будем, поскольку его доказательство требует несколько более основательных знаний математического анализа, нежели имеющиеся у студентов 2 курса ЭФ 1997 г. Это равенство — смена порядка дифференцирования и интегрирования — то единственное, ради которого и введены условия регулярности (см. пример ниже).

Напоминание: $\ln f(x)$ — сложная функция (см. правила дифференцирования суперпозиции).

$$\begin{aligned}
\frac{\partial}{\partial\theta}\mathsf{E}_\theta T(\vec{X}) &= \frac{\partial}{\partial\theta}\int T(\vec{y})\Psi(\vec{y},\theta)\,d\vec{y} \overset{?}{=} \int \frac{\partial}{\partial\theta}T(\vec{y})\Psi(\vec{y},\theta)\,d\vec{y} = \\
&= \int T(\vec{y})\frac{\frac{\partial}{\partial\theta}\Psi(\vec{y},\theta)}{\Psi(\vec{y},\theta)}\cdot\Psi(\vec{y},\theta)\,d\vec{y} = \\
&= \int T(\vec{y})\frac{\partial}{\partial\theta}L(\vec{y},\theta)\cdot\Psi(\vec{y},\theta)\,d\vec{y} = \mathsf{E}_\theta T(\vec{X})\cdot\frac{\partial}{\partial\theta}L(\vec{X},\theta).
\end{aligned}$$

$\square$

Воспользуемся леммой 5. Будем брать в качестве T разные функции и получать полезные формулы, которые потом соберем вместе.

1. Пусть $T(\vec{X}) \equiv 1$. Тогда

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

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

$$0 = \mathsf{E}_\theta\frac{\partial}{\partial\theta}L(\vec{X},\theta) = \mathsf{E}_\theta\sum \ln f_\theta(X_i) = n\mathsf{E}_\theta \ln f_\theta(X_1). \tag{7}$$

2. Пусть $T(\vec{X}) = \theta^* \in K_0$. Тогда

$$\frac{\partial}{\partial\theta}\mathsf{E}_\theta\theta^* = \frac{\partial}{\partial\theta}\theta = 1 = \mathsf{E}_\theta\theta^*\cdot\frac{\partial}{\partial\theta}L(\vec{X},\theta). \tag{8}$$

Вспомним свойство коэффициента корреляции:

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

Используя свойства (7) и (8), имеем

$$\operatorname{cov}(\theta^*, \frac{\partial}{\partial\theta}L(\vec{X},\theta)) = \mathsf{E}_\theta\theta^* \cdot \frac{\partial}{\partial\theta}L(\vec{X},\theta) - \mathsf{E}_\theta\theta^*\mathsf{E}_\theta\frac{\partial}{\partial\theta}L(\vec{X},\theta) =$$

$$= \mathsf{E}_\theta\theta^* \cdot \frac{\partial}{\partial\theta}L(\vec{X},\theta) = 1 \leqslant \sqrt{\mathsf{D}_\theta\theta^*\mathsf{D}_\theta\frac{\partial}{\partial\theta}L(\vec{X},\theta)}. \tag{9}$$

Найдем $\mathsf{D}_\theta\frac{\partial}{\partial\theta}L(\vec{X},\theta)$:

$$\mathsf{D}_\theta\frac{\partial}{\partial\theta}L(\vec{X},\theta) = \mathsf{D}_\theta\sum\frac{\partial}{\partial\theta}\ln f_\theta(X_i) = n\mathsf{D}_\theta\frac{\partial}{\partial\theta}\ln f_\theta(X_1) = n\mathsf{E}_\theta(\frac{\partial}{\partial\theta}\ln f_\theta(X_1))^2 = nI(\theta).$$

Подставляя дисперсию в неравенство (9), получим

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

что и требовалось доказать. □

Следующий пример показывает, что условие регулярности является существенным для выполнения равенства, помеченного (?) в лемме 5.

Пример 16. Нерегулярное семейство. Рассмотрим равномерное распределение $\mathrm{U}_{0,\theta}$ с параметром $\theta > 0$. Выпишем при $n = 1$ какой-нибудь интеграл и сравним производную от него и интеграл от производной:

$$\frac{\partial}{\partial\theta}\int\limits_0^\theta\frac{1}{\theta}\,dy = \frac{\partial}{\partial\theta}1 = 0; \qquad \int\limits_0^\theta\frac{\partial}{\partial\theta}\frac{1}{\theta}\,dy = -\frac{1}{\theta} \neq 0.$$

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

Упражнение. Проверить, что в качестве этой «выдающейся» из неравенства Рао-Крамера оценки можно брать $X_{(n)}$, или $\dfrac{n+1}{n}X_{(n)} \in K_0$.

4.4 Неравенство Рао-Крамера как способ проверки эффективности оценок

Сформулируем очевидное следствие из неравенства Рао-Крамера.

Следствие 1. *Если семейство распределений $\mathfrak{F}_\theta$ удовлетворяет условиям регулярности (R) и (RR), и оценка $\theta^* \in K_b$ такова, что в неравенстве Рао-Крамера достигается равенство:*

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

то оценка θ^* эффективна в классе K_b.

Пример 17. Для выборки $X_1, \ldots, X_n$ из распределения Пуассона Π_λ оценка $\lambda^* = \overline{X}$ эффективна (в классе K_0), так как для нее достигается равенство в неравенстве Рао-Крамера (см. [1], Пример на с. 42, решение 2).

Пример 18. Пусть $X_1, \ldots, X_n$ — выборка объема 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.

Найдем информацию Фишера относительно параметра a (считая, что имеется один неизвестный параметр — a).

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

Итак, $I(a) = 1/\sigma^2$. Найдем дисперсию оценки $\overline{X}$. По теореме о свойствах S_n/n из прошлого семестра, которую невредно вспомнить,

$$\mathsf{D}_{(a,\sigma^2)}\overline{X} \quad \overset{!}{=} \quad \frac{1}{n}\mathsf{D}_{(a,\sigma^2)}X_1 \quad = \quad \frac{\sigma^2}{n}.$$

Замечание 14. Тем, кто не желает обременять память воспоминаниями, предлагается воспользоваться свойствами дисперсии суммы независимых (и одинаково распределенных) случайных величин и доказать равенство

$$\boxed{\mathsf{D}\overline{X} \equiv \frac{1}{n}\mathsf{D}X_1}$$

Далее, сравнивая левую и правую части в неравенстве Рао-Крамера, получаем равенство:

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

То есть оценка $a^* = \overline{X}$ эффективна (обладает наименьшей дисперсией среди несмещенных оценок).

Пример 19. Пусть $X_1, \ldots, X_n$ — выборка объема n из нормального распределения N_{0,σ^2}, где $\sigma > 0$. Проверим, является ли оценка $\sigma^{2*} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 = \overline{X^2} \in K_0$ эффективной.

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

Найдем информацию Фишера относительно параметра σ^2.

$$f_{\sigma^2}(X_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{X_1^2}{2\sigma^2}\right)$$
$$\ln f_{\sigma^2}(X_1) = -\ln(2\pi)^{1/2} - \frac{1}{2}\ln\sigma^2 - \frac{X_1^2}{2\sigma^2}$$
$$\frac{\partial}{\partial\sigma^2}\ln f_{\sigma^2}(X_1) = -\frac{1}{2\sigma^2} + \frac{X_1^2}{2\sigma^4}$$
$$I(\sigma^2) = \mathsf{E}_{\sigma^2}\left(\frac{\partial}{\partial\sigma^2}\ln f_{\sigma^2}(X_1)\right)^2 = \mathsf{E}_{\sigma^2}\left(\frac{X_1^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right)^2 = \frac{1}{4\sigma^8}\mathsf{E}_{\sigma^2}(X_1^2-\sigma^2)^2 = \frac{1}{4\sigma^8}\mathsf{D}_{\sigma^2}X_1^2.$$

Осталось найти
$$\mathsf{D}_{\sigma^2}X_1^2 = \mathsf{E}_{\sigma^2}X_1^4 - (\mathsf{E}_{\sigma^2}X_1^2)^2 = \mathsf{E}_{\sigma^2}X_1^4 - \sigma^4.$$

Найдем четвертый момент X_1. Для тех, кто помнит некоторые формулы вероятности:

$$\mathsf{E}\xi^{2k} = (2k-1)!! = (2k-1)(2k-3)\cdot\ldots\cdot 3\cdot 1 \text{ при } \xi \in \mathrm{N}_{0,1},$$

$$X_1 = \xi \cdot \sigma \Subset \mathrm{N}_{0,\sigma^2} \qquad \Longrightarrow \qquad \mathsf{E} X_1^4 = \mathsf{E}\xi^4 \cdot \sigma^4 = 3\sigma^4.$$

Для тех, кто не помнит — краткое пособие по интегрированию:

$$\mathsf{E}_{\sigma^2} X_1^4 = \int\limits_{-\infty}^{\infty} y^4 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}\, dy = 2\sigma^4 \int\limits_0^{\infty} \left(\frac{y}{\sigma}\right)^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}\, d\left(\frac{y}{\sigma}\right) =$$

$$= 2\sigma^4 \int\limits_0^{\infty} t^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = -2\sigma^4 \frac{1}{\sqrt{2\pi}} \int\limits_0^{\infty} t^3\, de^{-\frac{t^2}{2}} = -2\sigma^4 \frac{1}{\sqrt{2\pi}} \left(t^3 e^{-\frac{t^2}{2}} \Big|_0^{\infty} - \int\limits_0^{\infty} e^{-\frac{t^2}{2}}\, dt^3 \right) =$$

$$= 2\sigma^4 \frac{1}{\sqrt{2\pi}} \cdot 3 \int\limits_0^{\infty} t^2 e^{-\frac{t^2}{2}}\, dt = 3\sigma^4 \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} t^2 e^{-\frac{t^2}{2}}\, dt = 3\sigma^4 \cdot \mathsf{D}\xi = 3\sigma^4 \cdot 1, \quad \text{где} \quad \xi \Subset \mathrm{N}_{0,1}.$$

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

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

Найдем дисперсию оценки $\sigma^{2*} = \overline{X^2}$.

$$\mathsf{D}_{\sigma^2} \overline{X^2} = \frac{1}{n^2} \mathsf{D}_{\sigma^2} \sum_1^n X_i^2 = \frac{1}{n} \mathsf{D}_{\sigma^2} X_1^2 = \frac{2\sigma^4}{n}.$$

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

$$\mathsf{D}_{\sigma^2} \overline{X^2} \quad = \quad \frac{2\sigma^4}{n} \quad = \quad \frac{1}{nI(\sigma^2)}.$$

Таким образом, оценка $\sigma^{2*} = \overline{X^2}$ эффективна.

Упражнение. Пусть $X_1, \ldots, X_n$ — выборка объема n из нормального распределения N_{a,σ^2}, где a известно, $\sigma > 0$. Проверить, является ли оценка $\sigma^{2*} = \dfrac{1}{n} \sum\limits_{i=1}^n (X_i - a)^2 = \overline{(X-a)^2}$ эффективной. Принадлежит ли эта оценка классу K_0? Какими методами получена? Является ли состоятельной и асимптотически нормальной?

Пример 20. Пусть $X_1, \ldots, X_n$ — выборка объема n из распределения Бернулли B_{17p} (а почему бы и нет?), где $p \in (0, 1/17)$. Проверим, является ли оценка $p^* = \overline{X}/17 \in K_0$ (оценка для параметра p!) эффективной.

Найдем информацию Фишера $I(p) = \mathsf{E}_p \left(\dfrac{\partial}{\partial p} \ln f_p(X_1) \right)^2$ относительно параметра p. Нам знаком только «табличный» вид «плотности» распределения Бернулли с вероятностью успеха $17p$:

$$f_p(y) = \mathsf{P}_p(X_1 = y) = \begin{cases} 17p, & \text{если } y = 1, \\ 1 - 17p, & \text{если } y = 0. \end{cases}$$

Заметим, что то же самое можно записать иначе:

$$f_p(y) = (17p)^y\, (1 - 17p)^{1-y}, \quad y = 0, 1.$$

Теперь уже удобно и логарифмировать, и дифференцировать.

$$f_p(X_1) = (17p)^{X_1} (1 - 17p)^{1 - X_1}$$

$$\ln f_p(X_1) = X_1 \ln 17 + X_1 \ln p + (1 - X_1) \ln (1 - 17p)$$

$$\frac{\partial}{\partial p} \ln f_p(X_1) = \frac{X_1}{p} - 17 \frac{1 - X_1}{1 - 17p} = \frac{X_1 - 17p}{p(1 - 17p)}$$

$$I(p) = \mathsf{E}_p \left(\frac{\partial}{\partial p} \ln f_p(X_1) \right)^2 = \frac{\mathsf{E}_p (X_1 - 17p)^2}{p^2 (1 - 17p)^2} = \frac{\mathsf{D}_p X_1}{p^2 (1 - 17p)^2} = \frac{17p(1 - 17p)}{p^2 (1 - 17p)^2} = \frac{17}{p(1 - 17p)}.$$

Итак, $I(p) = 17/p(1-17p)$. Найдем дисперсию оценки $\overline{X}/17$.

$$\mathsf{D}_p\frac{\overline{X}}{17} = \frac{\mathsf{D}_p\overline{X}}{17^2} = \frac{1}{17^2 n}\mathsf{D}_p X_1 = \frac{17p(1-17p)}{17^2 n} = \frac{p(1-17p)}{17n}.$$

Подставив дисперсию и информацию Фишера в неравенстве Рао-Крамера, получаем равенство:

$$\mathsf{D}_p\frac{\overline{X}}{17} \quad = \quad \frac{p(1-17p)}{17n} \quad = \quad \frac{1}{nI(p)}.$$

То есть оценка $p^* = \overline{X}/17$ эффективна (обладает наименьшей дисперсией среди несмещенных оценок).

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

4.5 Наилучшие линейные несмещенные оценки

В плане подготовки к курсу «Эконометрия» полезно заметить следующее: в практической статистике часто рассматривают оценки, являющиеся *линейными* (и по возможности несмещенными) функциями от выборки, то есть оценки вида $\theta^* = \sum_{i=1}^n a_i X_i$. В классе таких оценок наилучшая в смысле среднеквадратического подхода оценка обычно находится и без неравенства Рао-Крамера (что особенно полезно для нерегулярных семейств) — достаточно минимизировать $\sum a_i^2$ при заданной $\sum a_i$. Такую оценку принято называть «наилучшей линейной несмещенной оценкой», или по английски — BLUE (“best linear unbiased estimate").

Так, скажем, для распределения $\mathrm{U}_{0,\theta}$ оценка $\theta_0^* = 2\overline{X}$ является BLUE, так как ее дисперсия (*найти! или вспомнить пример 10*) не больше (*доказать!*) дисперсии любой оценки вида $\theta^* = \sum_{i=1}^n a_i X_i$, где $\sum_{i=1}^n a_i = 2$ (*почему это гарантирует несмещенность?*).

Справедливости ради следует добавить (см. пример 10), что оценка $\theta_0^* = 2\overline{X}$, хоть и является BLUE, не может конкурировать в среднеквадратичном смысле с нелинейной оценкой $\hat{\theta} = \frac{n+1}{n}X_{(n)}$ (которая является эффективной в классе несмещенных оценок, но это — страшная тайна).

4.6 Вопросы и упражнения

1. Проверить эффективность ОМП для следующих распределений:
   а) B_{p+1}, $-1 < p < 0$, б)$\Pi_{2\lambda}$, в) $\mathrm{E}_{1/\alpha}$, г) $\mathrm{N}_{a,1}$, д) $\mathrm{B}_{m,p}$ (биномиальное), $0 < p < 1$, при известном m.

2. Выполнить все упражнения, содержащиеся в тексте главы 4.

5 Доверительные интервалы

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 число («оценку»), способную (из каких-то разумных соображений) заменить параметр.

Существует другой подход к оцениванию, при котором мы указываем интервал, в котором с заданной наперед вероятностью лежит параметр. Такой подход называется «интервальным оцениванием». Сразу заметим, (хотя бы из философских соображений): чем больше уверенность в том, что параметр лежит в интервале, тем шире интервал. Так что мечтать найти диапазон, в котором θ лежит с вероятностью 1, бессмысленно — это вся область Θ.

5.1 Интервальное оценивание

Определение 13. Пусть $0 < \varepsilon < 1$. Интервал $(\theta^-, \theta^+) = \left(\theta^-(\vec{X}, \varepsilon), \theta^+(\vec{X}, \varepsilon)\right)$ называется *точным доверительным интервалом для параметра θ уровня доверия* $1 - \varepsilon$, если для любого $\theta \in \Theta$

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

Определение 14. Пусть $0 < \varepsilon < 1$. Интервал $(\theta^-, \theta^+) = \left(\theta^-(\vec{X}, \varepsilon), \theta^+(\vec{X}, \varepsilon)\right)$ называется *асимптотическим доверительным интервалом для параметра θ* (асимптотического) уровня доверия $1 - \varepsilon$, если для любого $\theta \in \Theta$

$$\underline{\lim}_{n\to\infty} \mathsf{P}_\theta\left(\theta^- < \theta < \theta^+\right) = \underline{\lim}_{n\to\infty} \mathsf{P}_\theta\left((\theta^-, \theta^+) \ni \theta\right) \geqslant 1 - \varepsilon.$$

Замечание 15. Случайны здесь границы интервала (θ^-, θ^+), поэтому читают формулу $\mathsf{P}_\theta\left(\theta^- < \theta < \theta^+\right)$ как «интервал (θ^-, θ^+) накрывает параметр θ», а не как «θ лежит в интервале...». Впрочем, это лишь фразеология (но точно отражающая суть дела).

Замечание 16. Знак «$\geqslant 1 - \varepsilon$» обычно соответствует дискретным распределениям, когда нельзя обязаться добиться равенства: например, для $\xi \in \mathrm{B}_{1/2}$ при любом x равенство $\mathsf{P}(\xi < x) = 0{,}25$ места не имеет, тогда как

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

Прежде чем рассматривать какие-то регулярные способы построения точных и асимптотических ДИ (доверительных интервалов), разберем два примера, предлагающих (очень похожие) способы.

Пример 21. Пусть $X_1, \ldots, X_n$ — выборка объема n из нормального распределения N_{a,σ^2}, где $a \in \mathbb{R}$, и $\sigma > 0$ известно. Требуется построить точный ДИ для параметра a уровня доверия $1 - \varepsilon$.

Вспомним, что нормальное распределение устойчиво по суммированию: (*доказать бы!*)

Свойство 2. *Пусть $\xi_1 \in \mathrm{N}_{a_1, \sigma_1^2}$, $\xi_2 \in \mathrm{N}_{a_2, \sigma_2^2}$ — независимые с.в. Тогда*

$$c\xi_1 + d\xi_2 + g \in \mathrm{N}_{ca_1 + da_2 + g,\, c^2\sigma_1^2 + d^2\sigma_2^2}.$$

Поэтому

$$\sum_1^n X_i \in \mathrm{N}_{na, n\sigma^2}, \quad \sum_1^n X_i - na \in \mathrm{N}_{0, n\sigma^2}, \quad \frac{\sum_1^n X_i - na}{\sqrt{n}\sigma} = \sqrt{n}\frac{\overline{X} - a}{\sigma} \in \mathrm{N}_{0,1}.$$

Обозначим $\eta = \sqrt{n}\dfrac{\overline{X} - a}{\sigma} \in \mathrm{N}_{0,1}$. По $\varepsilon \in (0, 1)$ найдем число $c > 0$, такое что $\mathsf{P}(-c < \eta < c) = 1 - \varepsilon$. Число c — квантиль уровня $1 - \varepsilon/2$ стандартного нормального распределения: (из области воспоминаний)

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

Напоминание:

Определение 15. Пусть распределение $\mathcal{F}$ с функцией распределения F абсолютно непрерывно. Число τ_δ называется *квантилью уровня* δ распределения $\mathcal{F}$, если $F(\tau_\delta) = \delta$.

Итак, $c = \tau_{1-\varepsilon/2}$, или $-c = \tau_{\varepsilon/2}$ (квантили стандартного нормального распределения).

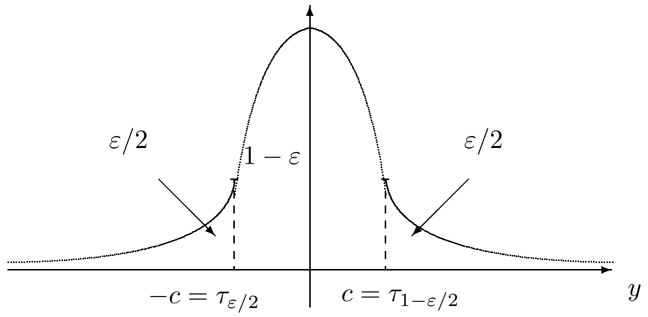


Рис. 5: Плотность стандартного нормального распределения и квантили.

Разрешив неравенство $-c < \eta < c$ относительно a, получим ДИ

$$1 - \varepsilon = \mathsf{P}(-c < \eta < c) = \mathsf{P}_a\left(-c < \sqrt{n}\frac{\overline{X} - a}{\sigma} < c\right) = \mathsf{P}_a\left(\overline{X} - \frac{c\sigma}{\sqrt{n}} < a < \overline{X} + \frac{c\sigma}{\sqrt{n}}\right). \tag{10}$$

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

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

Итак, искомый ДИ уровня доверия $1 - \varepsilon$ имеет вид $\left(\overline{X} - \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}}, \overline{X} + \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}}\right)$.

Вопросы, наводящие на размышления.
1. Зачем 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 ДИ?
2. Из двух ДИ одного уровня доверия и разной длины какой следует предпочесть?
3. Какова середина полученного в примере 21 ДИ? Какова его длина? Что происходит с границами ДИ при $n \to \infty$?

Пример 22. Пусть $X_1, \ldots, X_n$ — выборка объема n из показательного распределения E_α, где $\alpha > 0$. Требуется построить асимптотический ДИ для параметра α уровня доверия $1 - \varepsilon$.

Вспомним ЦПТ:

$$\frac{\sum_1^n X_i - n\mathsf{E}_\alpha X_1}{\sqrt{n\mathsf{D}_\alpha X_1}} = \sqrt{n}\frac{\overline{X} - \frac{1}{\alpha}}{\frac{1}{\alpha}} = \sqrt{n}(\alpha\overline{X} - 1) \Rightarrow \eta \Subset \mathrm{N}_{0,1}.$$

По определению слабой сходимости, при $n \to \infty$

$$\mathsf{P}_\alpha\left(-c < \sqrt{n}(\alpha\overline{X} - 1) < c\right) \to \mathsf{P}(-c < \eta < c) = 1 - \varepsilon \quad \text{при } c = \tau_{1-\varepsilon/2}.$$

То есть

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

Итак, искомый асимптотический ДИ уровня доверия $1 - \varepsilon$ имеет вид $\left(\frac{1}{\overline{X}} - \frac{\tau_{1-\varepsilon/2}}{\sqrt{n}\,\overline{X}}, \frac{1}{\overline{X}} + \frac{\tau_{1-\varepsilon/2}}{\sqrt{n}\,\overline{X}}\right)$.

Можно сформулировать общий принцип построения точных ДИ:

1. Найти функцию $G(\vec{X},\theta)$, имеющую известное (т.е. не зависящее от параметра θ) распределение $\mathcal{G}$: $G(\vec{X},\theta) \Subset \mathcal{G}$. Необходимо, чтобы $G(\vec{X},\theta)$ была обратима по θ при любом фиксированном $\vec{X}$.

2. Пусть g_1 и g_2 — квантили распределения $\mathcal{G}$, так что

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

3. Разрешив неравенство $g_1 < G(\vec{X},\theta) < g_2$ относительно θ (если это возможно), получим точный ДИ.

Замечание 17. Часто (но не всегда) в качестве g_1 и g_2 берут квантили уровня $\varepsilon/2$ и $1-\varepsilon/2$ распределения $\mathcal{G}$.

Совершенно аналогично выглядит общий принцип построения асимптотических ДИ:

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

2. Пусть g_1 и g_2 — квантили распределения $\mathcal{G}$, так что

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

3. Разрешив неравенство $g_1 < G(\vec{X},\theta) < g_2$ относительно θ, получим асимптотический ДИ.

Пример 23. Попробуем, пользуясь приведенной выше схемой, построить точный доверительный интервал для параметра θ равномерного на $[0,\theta]$ распределения (см. [1], задача 8.8, с очепяткой).

Мы знаем, что если $X_i \Subset \mathrm{U}_{0,\theta}$, то $\dfrac{X_i}{\theta} \Subset \mathrm{U}_{0,1}$. Тогда величина

$$\eta = \max\left\{\frac{X_1}{\theta},\ldots,\frac{X_n}{\theta}\right\} = \frac{\max\{X_1,\ldots,X_n\}}{\theta} = \frac{X_{(n)}}{\theta}$$

распределена так же, как максимум из n независимых равномерно распределенных на $[0,1]$ случайных величин, то есть имеет функцию распределения

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

Возьмем $g_2 = 1$. Найдем g_1 такое, что $1-\varepsilon = \mathsf{P}(g_1 < \eta < 1)$.

$$\mathsf{P}(g_1 < \eta < 1) = F_\eta(1) - F_\eta(g_1) = 1 - F_\eta(g_1) = 1 - g_1^n = 1-\varepsilon \quad \models\!\!\Rightarrow \quad g_1 = \sqrt[n]{\varepsilon}.$$

Тогда

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

Упражнение. Можно ли, пользуясь схемой примера 21, построить точный ДИ для σ при известном a, если разрешить неравенство $-c < \eta < c$ в (10) относительно σ? (Можно предположить, например, что $\overline{X} - a > 0$). Чем плох интервал бесконечной длины? Вопрос на засыпку: *а получился ли* у Вас интервал бесконечной длины?

Из упражнения видно, что функция G вида $\sqrt{n}\dfrac{\overline{X}-a}{\sigma}$ не годится для построения точного ДИ для σ при известном a, а тем более при неизвестном a. В следующей главе мы займемся поиском подходящих функций. Следующий пример (как и пример refex22) показывает, что ЦПТ дает универсальный вид функции G для построения асимптотических ДИ.

Пример 24. Пусть $X_1, \ldots, X_n$ — выборка объема n из распределения Пуассона Π_λ, где $\lambda > 0$. Требуется построить асимптотический ДИ для параметра λ уровня доверия $1 - \varepsilon$.

Вспомним ЦПТ:

$$\frac{\sum_1^n X_i - n\mathsf{E}_\lambda X_1}{\sqrt{n\mathsf{D}_\lambda X_1}} = \sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\lambda}} \Rightarrow \eta \in \mathrm{N}_{0,1}.$$

По определению слабой сходимости, при $n \to \infty$

$$\mathsf{P}_\lambda\left(-c < \sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\lambda}} < c\right) \to \mathsf{P}(-c < \eta < c) = 1 - \varepsilon \quad \text{при } c = \tau_{1-\varepsilon/2}.$$

Но разрешить неравенство под знаком вероятности относительно λ не просто — получается квадратное неравенство из-за корня в знаменателе. Заменим $\sqrt{\lambda}$ на $\sqrt{\overline{X}}$. Не испортится ли сходимость?

По свойствам слабой сходимости, если $\xi_n \xrightarrow{\mathrm{p}} 1$ и $\eta_n \Rightarrow \eta$, то $\xi_n\eta_n \Rightarrow \eta$. Воспользуемся этим.

Оценка $\lambda^* = \overline{X}$ состоятельна $\quad\Longrightarrow\quad \dfrac{\lambda}{\overline{X}} \xrightarrow{\mathrm{p}} 1 \quad\Longrightarrow\quad \sqrt{\dfrac{\lambda}{\overline{X}}} \xrightarrow{\mathrm{p}} 1$. Тогда

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

$$\mathsf{P}_\lambda\left(-\tau_{1-\varepsilon/2} < \sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\overline{X}}} < \tau_{1-\varepsilon/2}\right) \to \mathsf{P}(-\tau_{1-\varepsilon/2} < \eta < \tau_{1-\varepsilon/2}) = 1 - \varepsilon.$$

Разрешая неравенство под знаком вероятности относительно λ, получим

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

Итак, искомый асимптотический ДИ уровня доверия $1 - \varepsilon$ имеет вид $\left(\overline{X} - \dfrac{\tau_{1-\varepsilon/2}\sqrt{\overline{X}}}{\sqrt{n}}, \overline{X} + \dfrac{\tau_{1-\varepsilon/2}\sqrt{\overline{X}}}{\sqrt{n}}\right)$.

Вместо ЦПТ для построения асимптотических ДИ можно использовать асимптотически нормальные оценки (что по сути — та же ЦПТ):

Если θ^* — АНО параметра θ с коэффициентом $\sigma^2(\theta)$, то

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

Замечание 18. Если $\sigma(\theta)$ в знаменателе мешает, то (как в примере 24) ее можно заменить состоятельной оценкой $\sigma(\theta^*)$. Достаточно, чтобы функция $\sigma(\theta)$ была непрерывной. Требуется лишь ответить: почему θ^* — состоятельная оценка для θ?

5.2 Вопросы и упражнения

1. Задачник [1], т 8.2, 8.7 (а) — по ЦПТ.
2. Задачник [1], т 8.7 (б) (доказать, что $\dfrac{n}{\theta}(X_{(n)} - \theta) \Rightarrow \eta$, где $-\eta \in \mathrm{E}_1$, см. также упражнение к примеру 10).
3. Задачник [1], т 8.9. См. задачу 8.8 и пример 23. В п. (а) искать ДИ вида $X_{(1)} - \delta < \theta < X_{(1)}$, в п. (б) искать ДИ вида $X_{(1)} \cdot \delta < \theta < X_{(1)}$.
4. Объяснить, как додуматься до вида ДИ в предыдущей задаче. Исходить из вида распределений.

6 Распределения, связанные с нормальным

В предыдущей лекции мы построили (в числе других) точный доверительный интервал для параметра a нормального распределения при известном σ^2. Остались нерешенными следующие проблемы:

2) построить точный ДИ для σ при известном a,

3) построить точный ДИ для a при неизвестном σ,

4) построить точный ДИ для σ при неизвестном a.

Как мы уже видели, для решения этих задач требуется отыскать функции от выборки и параметров, распределение которых было бы известно. При этом в задаче (3) искомая функция не должна зависеть от σ, а в (4) — от a.

Такой особый интерес к нормальному распределению связан, разумеется, с центральной предельной теоремой — по большому счету все в этом мире нормально (или близко к тому).

Займемся поэтому распределениями, связанными с нормальным распределением, их свойствами и свойствами выборок из нормального распределения.

6.1 Гамма-распределение и его свойства

Определение 16. Случайная величина ξ имеет распределение $\Gamma_{\alpha,\lambda}$, где $\alpha > 0$, $\lambda > 0$, если ξ имеет плотность распределения

$$f_{\alpha,\lambda}(y) = \begin{cases} \dfrac{\alpha^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\alpha y}, & y > 0 \\ 0, & y \leqslant 0 \end{cases}$$

Здесь (см. справочник, «гамма-функция»)

$$\Gamma(\lambda) = \int\limits_0^\infty t^{\lambda-1} e^{-t}\, dt = (\lambda-1)\Gamma(\lambda-1),\ \Gamma(k) = (k-1)!,\ \Gamma(1/2) = \sqrt{\pi},\ \Gamma(1) = 1.$$

Найдем характеристическую функцию $\xi \in \Gamma_{\alpha,\lambda}$.

$$\varphi_\xi(t) = \mathsf{E} e^{it\xi} = \int\limits_0^\infty e^{ity} \frac{\alpha^\lambda}{\Gamma(\lambda)} y^{\lambda-1} e^{-\alpha y}\, dy = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int\limits_0^\infty y^{\lambda-1} e^{-(\alpha-it)y}\, dy =$$

$$= \frac{\alpha^\lambda}{\Gamma(\lambda)} \cdot \frac{1}{(\alpha-it)^\lambda} \underbrace{\int\limits_0^\infty ((\alpha-it)y)^{\lambda-1} e^{-(\alpha-it)y}\, d(\alpha-it)y}_{\Gamma(\lambda)} = \frac{\alpha^\lambda}{(\alpha-it)^\lambda} = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}.$$

Лемма 6. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ независимы и $\xi_i \in \Gamma_{\alpha,\lambda_i}$, $i = 1, \ldots, n$. Тогда*

$$S_n = \sum_{i=1}^n \xi_i \in \Gamma_{\alpha, \sum_1^n \lambda_i}.$$

Доказательство свойства устойчивости по суммированию (леммы 6).

$$\varphi_{S_n}(t) = \prod_{i=1}^n \varphi_{\xi_i}(t) = \prod_{i=1}^n \left(1 - \frac{it}{\alpha}\right)^{-\lambda_i} = \left(1 - \frac{it}{\alpha}\right)^{-\sum_1^n \lambda_i},$$

что и требовалось доказать. □

Следствие 2. *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ независимы, одинаково распределены и имеют показательное распределение E_α. Тогда $S_n = \sum_{i=1}^n \xi_i \in \Gamma_{\alpha,n}$.*

Доказательство. Достаточно заметить, что $\mathrm{E}_\alpha = \Gamma_{\alpha,1}$. ◻

Лемма 7. *Если* $\xi \in \mathrm{N}_{0,1}$, *то* $\xi^2 \in \Gamma_{1/2,1/2}$.

Доказательство. По определению, плотность величины ξ^2 связана с ее функцией распределения равенством:

$$\mathsf{P}(\xi^2 < y) = \int\limits_{-\infty}^{y} f_{\xi^2}(z)\,dz,$$

причем $f_{\xi^2}(z) = 0$ для $z < 0$ (*почему?*). При $y > 0$:

$$\mathsf{P}(\xi^2 < y) = \mathsf{P}(-\sqrt{y} < \xi < \sqrt{y}) = 2\int\limits_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt = \left[\begin{array}{l}\text{замена } t^2 = z,\ dt = \dfrac{1}{2\sqrt{z}}\,dz,\\ \text{границы:}\\ t=\sqrt{y}\mapsto z=y,\ t=0\mapsto z=0\end{array}\right]$$

$$= \int\limits_0^{y} \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2}\,dz = \int\limits_0^{y} \underbrace{\frac{(1/2)^{1/2}}{\Gamma(1/2)} z^{1/2-1} e^{-z/2}}_{f_{\xi^2}(z)}\,dz.$$

Плотность под интегралом есть плотность распределения $\Gamma_{1/2,1/2}$. ◻

Следствие 3. *Если* $\xi_1, \ldots, \xi_k$ *независимы и имеют стандартное нормальное распределение, то*

$$\chi_k^2 = \xi_1^2 + \ldots + \xi_k^2 \in \Gamma_{1/2,k/2}$$

Доказательство. Лемма 6 + лемма 7 $\implies$ следствие 3. ◻

6.2 Распределение «хи-квадрат» и его свойства

Определение 17. Распределение $\Gamma_{1/2,k/2}$ суммы k квадратов независимых стандартных нормальных случайных величин называют *распределением «хи-квадрат» с k степенями свободы* и обозначают H_k (или χ_k^2).

Для удобства мы часто будем обозначать через χ_k^2 случайную величину с распределением H_k.

Рис. 6: Плотности распределения H_k при разных k.

Свойства H_k:

1. Если $\phi \in \mathrm{H}_k$ и $\psi \in \mathrm{H}_m$ — независимы, то $\phi + \psi \in \mathrm{H}_{k+m}$.

Доказательство. Пусть $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Тогда ϕ распределено как $\xi_1^2 + \ldots + \xi_k^2$, ψ распределено как $\xi_{k+1}^2 + \ldots + \xi_{k+m}^2$, а их сумма — как $\xi_1^2 + \ldots + \xi_{k+m}^2 \in \mathrm{H}_{k+m}$.

2. Если $\chi_k^2 \Subset \mathrm{H}_k$, то $\mathsf{E}\chi_k^2 = k$, $\mathsf{D}\chi_k^2 = 2k$.

Доказательство. Пусть $\xi_1, \xi_2, \ldots$ независимы и имеют стандартное нормальное распределение. Тогда $\mathsf{E}\xi_1^2 = 1$, $\mathsf{D}\xi_1^2 = \mathsf{E}\xi_1^4 - (\mathsf{E}\xi_1^2)^2 = 3 - 1 = 2$ (см. пример 19).

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

Следствие 4. *Если $\xi_1, \ldots, \xi_k$ независимы и имеют распределение N_{a,σ^2}, то*

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

Упражнение. Доказать следствие 4.

6.3 Распределение Стьюдента и его свойства

Определение 18. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и имеют стандартное нормальное распределение. Распределение случайной величины

$$t_k = \frac{\xi_0}{\sqrt{\dfrac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}} = \frac{\xi_0}{\sqrt{\dfrac{\chi_k^2}{k}}}$$

называют *распределением Стьюдента с k степенями свободы* и обозначают T_k.

Рис. 7: Плотности распределения T_k и $\mathrm{N}_{0,1}$.

Свойства T_k:

1. Если $t_k \Subset \mathrm{T}_k$, то и $-t_k \Subset \mathrm{T}_k$.
Упражнение. Доказать.

2. Если $t_k \Subset \mathrm{T}_k$, то $t_k \Rightarrow \mathrm{N}_{0,1}$ при $k \to \infty$.

Доказательство. Пусть $\xi_0, \xi_1, \ldots, \xi_k$ независимы и имеют стандартное нормальное распределение. Тогда $\mathsf{E}\xi_1^2 = 1$, и по ЗБЧ

$$\frac{\chi_k^2}{k} = \frac{\xi_1^2 + \ldots + \xi_k^2}{k} \xrightarrow{\mathrm{p}} 1 \text{ при } k \to \infty.$$

Остается вспомнить свойства сходимости по распределению (*какие?*).

Отметим еще, что и распределение H_k, и распределение Стьюдента табулированы, так что если в каких-то доверительных интервалах появятся квантили этих распределений (вспомнить, к чему это!), то мы найдем их по таблице.

Упражнение. Как, пользуясь таблицей *стандартного нормального* распределения, найти квантиль заданного уровня распределения H_1 (с *одной* степенью свободы)?

6.4 Преобразования нормальных выборок

Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка из $\mathrm{N}_{0,1}$ (набор независимых и одинаково распределенных величин). Пусть C — ортогональная матрица $(n \times n)$, т.е. $CC^T = E = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$. Введем вектор $\vec{Y} = \vec{X} \cdot C$ с координатами $Y_i = \sum_{j=1}^n X_j C_{ji}$.

Лемма 8. *Координаты вектора $\vec{Y}$ независимы и имеют стандартное нормальное распределение.*

Доказательство. Воспользуемся свойством, утверждающим, что по характеристической функции однозначно восстанавливается распределение. Это верно и для характеристической функции вектора.

Определение 19. Характеристической функцией вектора $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ называется функция переменного $\vec{t} = (t_1, \ldots, t_n)$

$$\varphi_{\vec{\xi}}(\vec{t}) = \mathsf{E} e^{i\vec{t} \cdot \vec{\xi}^T} = \mathsf{E} e^{i\sum\limits_{j=1}^n t_j \xi_j}.$$

Характеристическая функция вектора $\vec{X}$ равна

$$\varphi_{\vec{X}}(\vec{t}) = \mathsf{E} e^{i\sum\limits_{j=1}^n t_j X_j} \overset{\text{независ-ть}}{=} \prod_{j=1}^n \mathsf{E} e^{it_j X_j} = \prod_{j=1}^n \varphi_{X_j}(t_j) =$$

$$= \prod_{j=1}^n \varphi_{X_j}(t_j) = \prod_{j=1}^n e^{-t_j^2/2} = e^{-\frac{1}{2}\sum\limits_{j=1}^n t_j^2}. \tag{11}$$

Вычислим х.ф. вектора $\vec{Y} = \vec{X} \cdot C$:

$$\varphi_{\vec{Y}}(\vec{t}) = \mathsf{E} e^{i\,\vec{t}\,\vec{Y}^T} = \mathsf{E} e^{i\,\vec{t}\,C^T\,\vec{X}^T} = \begin{bmatrix} \text{заменим} \\ \vec{t} = \vec{u} \cdot C \end{bmatrix} = \mathsf{E} e^{i\,\vec{u}\,C\,\cdot\,C^T\,\vec{X}^T} = \mathsf{E} e^{i\,\vec{u}\,\vec{X}^T} = e^{-\frac{1}{2}\sum\limits_{j=1}^n u_j^2} \quad (\text{по } (11)).$$

Так как C ортогональна, то $\sum_{j=1}^n u_j^2 = \sum_{j=1}^n t_j^2$ (доказать!). Поэтому

$$\varphi_{\vec{Y}}(\vec{t}) = e^{-\frac{1}{2}\sum\limits_{j=1}^n u_j^2} = e^{-\frac{1}{2}\sum\limits_{j=1}^n t_j^2} = \varphi_{\vec{X}}(\vec{t}),$$

и вектор $\vec{Y}$ распределен так же, как и вектор $\vec{X}$. □

Лемма 9. *Пусть $\vec{X} = (X_1, \ldots, X_n) \Subset \mathrm{N}_{0,1}$, независимы, C — ортогональная матрица и $\vec{Y} = \vec{X} \cdot C$. Тогда для любого $k = 1, \ldots, n$*

$$T(\vec{X}) = \sum_{i=1}^n X_i^2 - Y_1^2 - \ldots - Y_k^2 \Subset H_{n-k} \text{ и не зависит от } Y_1, \ldots, Y_k.$$

Доказательство. Поскольку $\vec{Y} = \vec{X} \cdot C$, то $\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$. Тогда

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

и не зависит от $Y_1, \ldots, Y_k$. □

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

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

Лемма 10 (Фишер). *Если $X_1, \ldots, X_k$ независимы и имеют распределение N_{a,σ^2}, то*

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

2. $\dfrac{(n-1)S_0^2}{\sigma^2} \in \mathrm{H}_{n-1}$;

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

Доказательство леммы Фишера.

1. Очевидно (доказать, что очевидно!).

2. Покажем 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 $z_i = \dfrac{X_i - a}{\sigma} \in \mathrm{N}_{0,1}$, $\overline{z} = \dfrac{\overline{X} - a}{\sigma}$.

То есть можно с самого начала считать, что $X_i \in \mathrm{N}_{0,1}$, и доказывать свойство 2 при $a = 0$, $\sigma^2 = 1$.

Применим лемму 9.

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

Мы обозначили

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

Чтобы применить лемму 9, нужно найти ортогональную матрицу C такую, что Y_1 — первая координата вектора $\vec{Y} = \vec{X} \cdot C$. Возьмем матрицу C с первым столбцом $\left(\dfrac{1}{\sqrt{n}}, \ldots, \dfrac{1}{\sqrt{n}}\right)$. Так как длина этого вектора — единица, его можно дополнить до ортонормального базиса (иначе — этот столбец можно дополнить до ортогональной матрицы). Тогда $Y_1 = \sqrt{n}\,\overline{X}$ и будет первой координатой вектора $\vec{Y} = \vec{X} \cdot C$.

Осталось применить лемму 9 (*непременно сделайте это!*).

3. Из леммы 9 сразу следует, что $T(\vec{X}) = (n-1)S_0^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2$ не зависит от $Y_1 = \sqrt{n}\,\overline{X}$, то есть $\overline{X}$ и S_0^2 независимы. □

Следующее следствие из леммы Фишера наконец позволит нам строить доверительные интервалы для параметров нормального распределения — то, ради чего мы и доказали уже так много утверждений. В каждом пункте указано, для какого параметра мы построим доверительный интервал с помощью данного утверждения.

Следствие 5. **1.** $\sqrt{n}\dfrac{\overline{X} - a}{\sigma} \in \mathrm{N}_{0,1}$ *— для a при σ известном;*

2. $\sum\limits_{i=1}^{n}\left(\dfrac{X_i - a}{\sigma}\right)^2 = \dfrac{n\sigma^{2*}}{\sigma^2} \in \mathrm{H}_n$ *— для σ^2 при a известном;*

3. $\sum\limits_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2 = \frac{(n-1)S_0^2}{\sigma^2} \Subset \mathrm{H}_{n-1}$ — *для σ^2 при a неизвестном;*

4. $\sqrt{n}\frac{\overline{X}-a}{\sqrt{S_0^2}} = \sqrt{n}\frac{\overline{X}-a}{S_0} \Subset \mathrm{T}_{n-1}$ — *для a при σ неизвестном.*

Доказательство следствия 5. 1) и 3) — из леммы Фишера, 2) — из следствия 4, осталось воспользоваться леммой Фишера и определением 18 распределения Стьюдента, чтобы доказать 4).

$$\sqrt{n}\frac{\overline{X}-a}{\sqrt{S_0^2}} = \underbrace{\sqrt{n}\frac{\overline{X}-a}{\sigma}}_{\mathrm{N}_{0,1}} \cdot \frac{1}{\sqrt{\underbrace{\frac{(n-1)S_0^2}{\sigma^2}}_{\mathrm{H}_{n-1}}\frac{1}{n-1}}} = \frac{\xi_0}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \Subset \mathrm{T}_{n-1}.$$

$\nwarrow$ независ. $\nearrow$

□

6.5 Построение точных доверительных интервалов для параметров нормального распределения

1. Для a при известном σ^2. Это мы уже построили:

$$\mathsf{P}_{a,\sigma^2}\left(\overline{X} - \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}} < a < \overline{X} + \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}}\right) = 1-\varepsilon, \text{ где } \Phi_{0,1}(\tau_{1-\varepsilon/2}) = 1-\varepsilon/2.$$

2. Для σ^2 при известном a: по п.2 из следствия 5, $\frac{n\sigma^{2*}}{\sigma^2} \Subset \mathrm{H}_n$, где $\sigma^{2*} = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-a)^2$. Пусть $g_1 = \chi^2_{n,\varepsilon/2}$ и $g_2 = \chi^2_{n,1-\varepsilon/2}$ — квантили распределения H_n уровня $\varepsilon/2$ и $1-\varepsilon/2$. Тогда

$$1-\varepsilon = \mathsf{P}_{a,\sigma^2}\left(g_1 < \frac{n\sigma^{2*}}{\sigma^2} < g_2\right) = \mathsf{P}_{a,\sigma^2}\left(\frac{n\sigma^{2*}}{g_2} < \sigma^2 < \frac{n\sigma^{2*}}{g_1}\right).$$

Искомый интервал построен.

3. Для σ^2 при неизвестном a: по п.3 из следствия 5, $\frac{(n-1)S_0^2}{\sigma^2} \Subset \mathrm{H}_{n-1}$, где $S_0^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$. Пусть $g_1 = \chi^2_{n-1,\varepsilon/2}$ и $g_2 = \chi^2_{n-1,1-\varepsilon/2}$ — квантили распределения H_{n-1} уровня $\varepsilon/2$ и $1-\varepsilon/2$. Тогда

$$1-\varepsilon = \mathsf{P}_{a,\sigma^2}\left(g_1 < \frac{(n-1)S_0^2}{\sigma^2} < g_2\right) = \mathsf{P}_{a,\sigma^2}\left(\frac{(n-1)S_0^2}{g_2} < \sigma^2 < \frac{(n-1)S_0^2}{g_1}\right).$$

Искомый интервал построен.

Упражнение. Найти 17 отличий п.2 от п.3.

4. Для a при неизвестном σ: по п.4 из следствия 5, $\sqrt{n}\frac{\overline{X}-a}{S_0} \Subset \mathrm{T}_{n-1}$, где $S_0^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$.

Пусть $a_1 = t_{n-1,\varepsilon/2}$ и $a_2 = t_{n-1,1-\varepsilon/2}$ — квантили распределения T_{n-1} уровня $\varepsilon/2$ и $1-\varepsilon/2$. Заметим, что в силу симметричности распределения Стьюдента $a_1 = -a_2 = -t_{n-1,1-\varepsilon/2}$. Тогда

$$1-\varepsilon = \mathsf{P}_{a,\sigma^2}\left(-t_{n-1,1-\varepsilon/2} < \sqrt{n}\frac{\overline{X}-a}{S_0} < t_{n-1,1-\varepsilon/2}\right) =$$

$$= \mathsf{P}_{a,\sigma^2}\left(\overline{X} - \frac{t_{n-1,1-\varepsilon/2}S_0}{\sqrt{n}} < a < \overline{X} + \frac{t_{n-1,1-\varepsilon/2}S_0}{\sqrt{n}}\right).$$

Искомый интервал построен.

6.6 Вопросы и упражнения

1. Величины ξ_1 и ξ_2 независимы и имеют нормальное распределение $N(0,16)$. Найти k, при котором величины $\xi_1 - 3\xi_2$ и $k\xi_1 + \xi_2$ независимы. Можно использовать лемму 8.

2. Доказать, что для величины $\chi_n^2 \Subset \mathrm{H}_n$ справедлива аппроксимация Фишера:

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

и вывести отсюда, что при больших n для вычисления квантилей распределения H_n можно пользоваться аппроксимацией

$$\mathrm{H}_n(x) = \mathsf{P}(\chi_n^2 < x) \sim \Phi_{0,1}(\sqrt{2x} - \sqrt{2n}).$$

Указание. Воспользоваться теоремой 9 (об асимптотической нормальности оценок вида $H\left(\overline{g(X)}\right)$). Вспомнить, что такое асимптотическая нормальность, в аппроксимации Фишера вынести за скобки $\sqrt{n}$ и заменить χ_n^2 суммой квадратов по определению 17.

3. Изобразить квантили уровня $\varepsilon/2$ и $1 - \varepsilon/2$ на графике плотности распределения H_n и T_{n-1}.

7 Проверка гипотез

Если возможно выдвинуть несколько взаимоисключающих «гипотез» о распределении элементов выборки, то возникает задача о выборе одной из этих гипотез на основании выборочных данных. Понятно, что по выборке конечного объема безошибочных выводов о распределении сделано быть не может, поэтому приходится считаться с возможностью выбрать неверную гипотезу.

7.1 Гипотезы и критерии

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathcal{F}$. Как правило, если не оговорено противное, считается, что все наблюдения имеют одно и то же распределение. В ряде случаев это предположение также нуждается в проверке (см., например, ниже: гипотеза об однородности или гипотеза о случайности) — в таких случаях одинаковая распределенность наблюдений не предполагается. То же касается и независимости наблюдений.

Определение 20. Гипотезой (H) называется любое предположение о распределении наблюдений:

$$H: \quad \mathcal{F} = \mathcal{F}_1 \quad \text{или} \quad H: \quad \mathcal{F} \in \{\widetilde{\mathcal{F}}\}.$$

Гипотеза H называется простой, если она однозначно определяет распределение $(H: \quad \mathcal{F} = \mathcal{F}_1)$. Иначе H называется сложной гипотезой $(H: \quad \mathcal{F} \in \{\widetilde{\mathcal{F}}\})$.

Если гипотез всего две, то одну из них принято называть основной, а другую — альтернативой или отклонением от основной гипотезы.

Пример 25. Обычные постановки задач.

1. Выбор из нескольких простых гипотез: $H_1 : \mathcal{F} = \mathcal{F}_1$, $H_2 : \mathcal{F} = \mathcal{F}_2$, ..., $H_k : \mathcal{F} = \mathcal{F}_k$ (и другие предположения невозможны).

2. Простая основная гипотеза и сложная альтернатива: $H_1 : \mathcal{F} = \mathcal{F}_1$, $H_2 : \mathcal{F} \neq \mathcal{F}_1$. Например: $H_1 : \mathcal{F} = \mathrm{U}_{0,1}$, $H_2 : \mathcal{F} \neq \mathrm{U}_{0,1}$.

3. Сложная основная гипотеза и сложная альтернатива: $H_1 : \mathcal{F} \in \{\widetilde{\mathcal{F}}\}$, $H_2 : \mathcal{F} \notin \{\widetilde{\mathcal{F}}\}$.
   Например (гипотеза о виде распределения):
   $H_1 : \mathcal{F} \in \{\mathrm{N}_{a,1}, a > 0\}$, $H_2 : H_1$ не верна.

4. Гипотеза однородности. Заданы несколько выборок
   $\vec{X}_1 = (X_{11}, \ldots, X_{1n_1}) \Subset \mathcal{F}_1$, ...
   $\vec{X}_k = (X_{k1}, \ldots, X_{kn_k}) \Subset \mathcal{F}_k$.
   Проверяется гипотеза $H_1 : \mathcal{F}_1 = \ldots = \mathcal{F}_k$ — сложная гипотеза — против (сложной) альтернативы $H_2 : H_1$ не верна.

5. Гипотеза независимости. Наблюдается пара случайных величин (ξ, η).
   По выборке $(\overrightarrow{X, Y}) = ((X_1, Y_1), \ldots, (X_n, Y_n))$ проверяется гипотеза
   $H_1 : \xi$ и η независимы — сложная гипотеза — против (сложной) альтернативы $H_2 : H_1$ не верна.

6. Гипотеза случайности. Наблюдается n случайных величин $(\xi_1, \ldots, \xi_n)$. По выборке $\vec{X} = (X_1, \ldots, X_n)$, в которой каждая с.в. представлена одним значением, проверяется гипотеза
   $H_1 : \xi_1, \ldots, \xi_n$ независимы и одинаково распределены — сложная гипотеза — против (сложной) альтернативы $H_2 : H_1$ не верна.

   Эту задачу ставят, например, если требуется проверить качество датчика случайных чисел.

Определение 21. Если имеются гипотезы $H_1, \ldots, H_k$, то критерием (нерандомизированным критерием) $\delta = \delta(\vec{X})$ называется отображение

$$\delta : \mathbb{R}^n \to \{H_1, \ldots, H_k\}.$$

О *рандомизированных* критериях, которые предписывают принимать каждую гипотезу с некоторой (зависящей от выборки) вероятностью, мы поговорим позднее.

Определение 22. Для заданного критерия $\delta : \mathbb{R}^n \to \{H_1, \ldots, H_k\}$ будем говорить, что произошла ошибка i-го рода, если гипотеза H_i отвергнута критерием, в то время как она верна. Вероятностью ошибки i-го рода называется число

$$\alpha_i(\delta) = \mathsf{P}(\delta(\vec{X}) \neq H_i | H_i \text{ верна}) = \mathsf{P}_{H_i}(\delta(\vec{X}) \neq H_i).$$

Для краткости мы часто будем говорить «вычислить ошибку 1-го рода», вместо «вычислить вероятность ошибки 1-го рода».

Пример 26. Контроль качества и ошибки.

Пусть любое изделие некоторого производства оказывается браком с вероятностью p. Контроль продукции допускает ошибки: годное изделие бракует с вероятностью γ, а бракованное признает годным с вероятностью ε.

Если ввести для данного изделия 2 гипотезы: H_1 : изделие годное и H_2 : изделие бракованное, и в качестве критерия выбора использовать контроль продукции, то γ — вероятность ошибки 1-го рода, а ε — 2-го рода данного критерия.

Упражнение. Вычислить ошибки 1-го и 2-го рода того же критерия для проверки гипотез H_1 : изделие бракованное и H_2 : изделие годное.

7.2 Две простые гипотезы

Рассмотрим подробно случай, когда имеется две простые гипотезы о распределении наблюдений:

$$\begin{cases} H_1 : \mathcal{F} = \mathcal{F}_1, \\ H_2 : \mathcal{F} = \mathcal{F}_2. \end{cases}$$

Каков бы ни был критерий $\delta(\vec{X}) : \mathbb{R}^n \to \{H_1, H_2\}$, он принимает не более двух значений. То есть 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 называют *критической областью*.

Определение 23. Ошибку 1-го рода критерия δ (в случае двух простых гипотез) обозначим α и назовем уровнем критерия:

$$\alpha_1(\delta) = \mathsf{P}_{H_1}(\delta(\vec{X}) \neq H_1) = \mathsf{P}_{H_1}(\vec{X} \in S) \stackrel{\text{def}}{=} \alpha.$$

Ошибку 2-го рода критерия δ (в случае двух простых гипотез) обозначим β. Величину $1 - \beta$ назовем мощностью критерия:

$$\alpha_2(\delta) = \mathsf{P}_{H_2}(\delta(\vec{X}) \neq H_2) = \mathsf{P}_{H_2}(\vec{X} \in \mathbb{R}^n \backslash S) \stackrel{\text{def}}{=} \beta.$$

Мощность критерия можно записать следующим образом:

$$1 - \beta = \mathsf{P}_{H_2}(\delta(\vec{X}) = H_2) = \mathsf{P}_{H_2}(\vec{X} \in S).$$

Заметим, что ошибки 1-го и 2-го рода вычисляются при *разных* предположениях о распределении (если верна H_1 и если верна H_2), так что никаких фиксированных соотношений (типа $\alpha = 1 - \beta$, независимо от вида гипотез и вида критерия) между ними нет.

Рассмотрим в качества примера «крайний случай», когда независимо ни от чего всегда принимается одна гипотеза.

Пример 27. Имеются гипотезы H_1, H_2 и критерий $\delta(\vec{X}) \equiv H_1$, то есть $S = \varnothing$. Тогда $\alpha = \mathsf{P}_{H_1}(\text{ принять } H_2) = 0$, $\beta = \mathsf{P}_{H_2}(\text{ принять } H_1) = 1$.

Наоборот: пусть имеются гипотезы H_1, H_2 и критерий $\delta(\vec{X}) \equiv H_2$, то есть $S = \mathbb{R}^n$. Тогда $\alpha = \mathsf{P}_{H_1}(\text{ принять } H_2) = 1$, $\beta = \mathsf{P}_{H_2}(\text{ принять } H_1) = 0$.

Примеры 27 и 28 показывают общую тенденцию: при попытке уменьшить одну из ошибок критерия другая, *как правило*, увеличивается. А в упражнении после примера 28 приведено отклонение от этой «тенденции»: для заданного критерия можно построить и «очень плохой» критерий, у которого обе ошибки больше, чем у заданного.

Пример 28. Имеется выборка объема 1 из нормального распределения $\mathrm{N}_{a,1}$. Проверяются простые гипотезы $\begin{cases} H_1 : a = 0, \\ H_2 : a = 1. \end{cases}$ Используется следующий критерий (при заданной постоянной c):

$$\delta(X_1) = \begin{cases} H_1, & \text{если } X_1 \leqslant c, \\ H_2, & \text{если } X_1 > c. \end{cases}$$

На графике изображены плотности, соответствующие гипотезам, и вероятности ошибок 1-го и 2-го рода критерия δ:

$$\alpha = \mathsf{P}_{H_1}(\delta(X_1) = H_2) = \mathsf{P}_{H_1}(X_1 > c), \quad \beta = \mathsf{P}_{H_2}(\delta(X_1) = H_1) = \mathsf{P}_{H_2}(X_1 \leqslant c).$$

Рис. 8: Две простые гипотезы.

Хорошо видно, что при $c \nearrow \infty$ вероятность ошибки 1-го рода $\alpha \searrow 0$, но вероятность ошибки 2-го рода $\beta \nearrow 1$.

Упражнение. Рассмотрим критерий следующего вида ($c_1 < c_2$ — постоянные):

$$\delta_1(X_1) = \begin{cases} H_1, & \text{если } X_1 \in [c_1, c_2], \\ H_2, & \text{если } X_1 \notin [c_1, c_2]. \end{cases}$$

Нарисовать на графике вероятности ошибок 1-го и 2-го рода критерия δ_1 и убедиться, что при одной и той же вероятности ошибки 1-го рода критерий δ_1 обладает большей вероятностью ошибки 2-го рода, чем критерий δ.

Итак, два главных вывода:

1. Критерий тем лучше, чем меньше вероятности ошибок.

2. Сравнивать критерии по паре ошибок $\alpha(\delta_1) \leqslant \alpha(\delta_2), \beta(\delta_1) \leqslant \beta(\delta_2)$ удается далеко не всегда.

7.3 Способы сравнения критериев

Ограничимся, для простоты, задачей сравнения двух простых гипотез.

Пусть имеются критерии δ_1 и δ_2 с ошибками 1-го и 2-го рода $\alpha(\delta_i)$, $\beta(\delta_i)$, $i = 1, 2$.

Перечислим общепринятые подходы к сравнению критериев:

1. Минимаксный подход.

Говорят, что критерий δ_1 не хуже, чем δ_2 (в смысле минимаксного подхода), если $\max\{\alpha(\delta_1), \beta(\delta_1)\} \leqslant \max\{\alpha(\delta_2), \beta(\delta_2)\}$.

Определение 24. Критерий δ называют минимаксным критерием, если он лучше (не хуже) всех других критериев в смысле минимаксного подхода.

Иначе говоря, минимаксный критерий имеет самую маленькую «наибольшую ошибку» $\max\{\alpha(\delta), \beta(\delta)\}$ среди всех прочих критериев.

Упражнение. Убедиться, что в примере 28 критерий δ является минимаксным, если $c = 1/2$.

2. Байесовский подход.

Этот подход применяют в двух случаях:
а) если известно априори, что с вероятностью p справедлива гипотеза H_1, а с вероятностью $q = 1 - p$ — гипотеза H_2,
б) если задана линейная «функция потерь»: потери от ошибочного решения равны p, если происходит ошибка 1-го рода, и равны q, если второго. Здесь $p + q$ уже не обязательно равно 1, но потери можно свести к единице нормировкой $p/(p+q)$ и $q/(p+q)$.

Говорят, что критерий δ_1 не хуже, чем δ_2 (в смысле байесовского подхода), если $p\alpha(\delta_1) + q\beta(\delta_1) \leqslant p\alpha(\delta_2) + q\beta(\delta_2)$.

Определение 25. Критерий δ называют байесовским критерием, если он лучше (не хуже) всех других критериев в смысле байесовского подхода.

Иначе говоря, байесовский критерий имеет самую маленькую «средневзвешенную ошибку» $p\alpha(\delta) + q\beta(\delta)$ 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 28 критерий δ является байесовским для $p = q$, если $c = 1/2$.

3. Выбор наиболее мощного критерия.

Ошибки 1-го и 2-го рода обычно неравноправны (см. пример 26, и пусть «изделие» = самолет или ядерный реактор). Поэтому возникает желание контролировать одну из ошибок (скажем, 1-го рода). Например, зафиксировать ее на достаточно низком (безопасном) уровне, и рассматривать только критерии с такой или еще меньшей вероятностью ошибки 1-го рода. Среди них наилучшим, очевидно, следует признать критерий, обладающий наименьшей вероятностью ошибки 2-го рода.

Введем при $\varepsilon \in [0, 1]$ класс критериев $K_\varepsilon = \{\delta(\vec{X}) : \alpha(\delta) \leqslant \varepsilon\}$

Определение 26. Критерий $\delta \in K_\varepsilon$ называют наиболее мощным критерием (НМК) уровня ε (в классе K_ε), если $\beta(\delta) \leqslant \beta(\delta_0)$ для любого критерия $\delta_0 \in K_\varepsilon$.

Если имеется более двух гипотез, то сравнивать критерии в смысле определения 26 можно, если зафиксировать все ошибки, кроме одной.

7.4 Построение НМК. Лемма Неймана - Пирсона

Мы рассмотрим подробно третий подход к сравнению критериев, и научимся строить наиболее мощный критерий заданного уровня.

Пусть $\vec{X} = (X_1, \ldots, X_n)$ — выборка (набор независимых, одинаково распределенных величин), и имеются две гипотезы о распределении X_i:

$$H_1 : X_i \Subset \mathcal{F}_1 \text{ с плотностью } f_1(y),\ \Psi_1(\vec{X}) = \prod_1^n f_1(X_i) - \text{ функция правдоподобия}$$

$$H_2 : X_i \Subset \mathcal{F}_2 \text{ с плотностью } f_2(y),\ \Psi_2(\vec{X}) = \prod_1^n f_2(X_i) - \text{ функция правдоподобия}$$

Предполагается, что распределения $\mathcal{F}_1$, $\mathcal{F}_2$ либо оба дискретны, либо оба абсолютно непрерывны.

Замечание 19. Если данное предположение не выполнено, то всегда существует критерий с нулевыми вероятностями ошибок (см. задачу 9.1 в задачнике [1]).

Мы собираемся выбирать гипотезу в зависимости от отношения функций правдоподобия. Напомню, что функция правдоподобия есть плотность выборки.

Если не ставить перед собой задачу получить определенное значение ошибки 1-го рода, этот подход выглядит так: «там, где вторая плотность больше, выбираем 2-ю гипотезу, там, где первая — первую».

Если же нужно получить критерий заданного уровня $\alpha = \varepsilon$, то данный подход выглядит так: «там, где вторая плотность в c больше первой, выбираем 2-ю гипотезу, иначе — первую». При этом c выбирают так, чтобы ошибка 1-го рода равнялась ε.

Проблема возникает лишь в том случае, когда $\Psi_2(\vec{X}) = \Psi_1(\vec{X})$ – в первом случае, или $\Psi_2(\vec{X}) = c\Psi_1(\vec{X})$ – во втором.

Если такое событие имеет нулевую вероятность, то можно выбирать любую гипотезу, и это никак не скажется на ошибках.

Если же это событие происходит с положительной вероятностью, то при попадании выборки в эту область мы не знаем какую из гипотез предпочесть — они «равновозможны». В этом случае нет более разумного подхода, чем подбросить монету и выбрать одну из гипотез наудачу (не обязательно с равными вероятностями – это зависит от требуемой ошибки 1-го рода).

Итак,

Лемма 11 (Нейман, Пирсон). *Для любого $\varepsilon \in [0,1]$ НМК уровня ε существует и совпадает с* **критерием отношения правдоподобия***:*

$$\begin{cases} \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} < c & \Longrightarrow \quad \delta(\vec{X}) = H_1 \\ \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c & \Longrightarrow \quad \delta(\vec{X}) = H_2 \\ \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c & \Longrightarrow \quad \begin{cases} \delta(\vec{X}) = H_2 & \text{с вероятностью } p, \\ \delta(\vec{X}) = H_1 & \text{с вероятностью } 1-p, \end{cases} \end{cases}$$

при этом c и p определяются из уравнения $\alpha(\delta) = \varepsilon$, или

$$\alpha(\delta) = \mathsf{P}_{H_1}(\delta(\vec{X}) = H_2) = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right) + p \cdot \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right) = \varepsilon. \tag{12}$$

Замечание 20. Отношение $\Psi_2(\vec{X})/\Psi_1(\vec{X})$ рассматривается только при $\vec{X}$ таких что $\Psi_1(\vec{X})+\Psi_2(\vec{X}) > 0$. Напоминаю, что $c/\infty = 0$, $c/0 = \infty$ (все постоянные и «нули» здесь положительны).

Следующие замечание и определение предназначены тем, кто заметил несоответствие формулировки леммы здесь и в любом учебнике (например, в [2]) и собирается пользоваться классической, а не упрощенной формой НМК.

Замечание 21. На самом деле критерий, предлагаемый в теореме, не является критерием в смысле определения 21. Это не просто функция из $\mathbb{R}^n$ в $\{H_1, H_2\}$, но *случайная* функция (т.е. зависящая еще и от результата некоторого случайного эксперимента).

Введем понятие *рандомизированного* критерия.

Определение 27. Пусть имеются гипотезы H_1, H_2. Рандомизированным критерием $\pi = \pi(\vec{X})$ называется функция

$$\pi : \mathbb{R}^n \to [0,1],$$

при каждом $\vec{X} \in \mathbb{R}^n$ равная **вероятности принять гипотезу** H_2.

Если использовать определение 27, критерий отношения правдоподобия в лемме Неймана - Пирсона примет вид:

$$\pi(\vec{X}) = \begin{cases} 0, & \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} < c \\ 1, & \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c \\ p, & \text{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c, \end{cases}$$

где c и p по-прежнему определяются из уравнения $\alpha(\pi) = \varepsilon$, которое можно записать так:

$$\alpha(\pi) = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right) + p \cdot \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right) = \mathsf{E}_{H_1}(\pi(\vec{X})) = \varepsilon.$$

7.5 Доказательство леммы Неймана - Пирсона

1. Докажем, что уравнение 12 разрешимо относительно c и p.

Рассмотрим невозрастающую (*почему?*) функцию $\phi(c) = \mathsf{P}_{H_1}\left(\dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right)$.

$$\phi(c) = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right) = \int\limits_{\frac{\Psi_2(\vec{y})}{\Psi_1(\vec{y})} > c} \Psi_1(\vec{y})\, d\vec{y}.$$

Поскольку интегрирование ведется по области $\frac{\Psi_2(\vec{y})}{\Psi_1(\vec{y})} > c$, то под интегралом $\Psi_1(\vec{y}) < \Psi_2(\vec{y})/c$. Поэтому

$$\phi(c) = \int\limits_{\frac{\Psi_2(\vec{y})}{\Psi_1(\vec{y})} > c} \Psi_1(\vec{y})\, d\vec{y} < \frac{1}{c} \int\limits_{\frac{\Psi_2(\vec{y})}{\Psi_1(\vec{y})} > c} \Psi_2(\vec{y})\, d\vec{y} = \frac{1}{c}\mathsf{P}_{H_2}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right) \to 0 \qquad (13)$$

при $c \to \infty$.

Рассмотрим $\phi(0)$:

$$\phi(0) = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > 0\right) = \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right)$$

Возможны два случая: (**а**) $\phi(0) = \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right) < \varepsilon$ и (**б**) $\phi(0) = \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right) \geqslant \varepsilon$.

В случае (а) положим $c = 0$, обозначим через Δ разницу $\Delta = \varepsilon - \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right) > 0$ и возьмем $p = \Delta/(1 + \Delta - \varepsilon)$. Тогда

$$\alpha(\delta) = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > 0\right) + p \cdot \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = 0\right) =$$

$$= \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right) + p \cdot \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) = 0\right) = (\varepsilon - \Delta) + \frac{\Delta}{1 + \Delta - \varepsilon}(1 - (\varepsilon - \Delta)) = \varepsilon.$$

Заметим сразу, что в случае (а) мы хоть и нашли критерий заданного уровня $\varepsilon > \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right)$, но для этого пришлось принимать (с вероятностью p) гипотезу H_2 там, где она быть верна не может — в области $\Psi_2(\vec{X}) = 0$.

Это означает лишь, что с самого начала наше желание найти критерий уровня ε, если $\varepsilon > \mathsf{P}_{H_1}\left(\Psi_2(\vec{X}) > 0\right)$ — абсурд: все разумные критерии имеют меньший уровень.

В случае (б) имеем: $\phi(0) \geqslant \varepsilon$, $\phi(c)$ не возрастает и стремится к нулю с ростом c. Тогда найдется c такое, что $\phi(c - 0) \geqslant \varepsilon$, $\phi(c) \leqslant \varepsilon$ (c может быть точкой разрыва).

Тогда (вспомнить свойство функций распределения $F_\xi(x+0) - F_\xi(x) = \mathsf{P}(\xi = x)$)

$$\Delta \stackrel{\text{def}}{=} \phi(c-0) - \phi(c) \equiv \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right) \geqslant 0.$$

Возьмем $p = \dfrac{\varepsilon - \phi(c)}{\Delta} \in [0,1]$. Для такого p

$$\varepsilon = \phi(c) + p\Delta = \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right) + p\cdot \mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right) = \alpha(\delta),$$

что и требовалось доказать.

2. Докажем, что критерий δ наиболее мощный. Нам потребуется следующее

Утверждение 1. *Обозначим* $\Psi(\vec{y}) = \min\{\Psi_2(\vec{y}), c\Psi_1(\vec{y})\}$. *Тогда* $\beta(\delta) + c\alpha(\delta) = \int\limits_{\mathbb{R}^n} \Psi(\vec{y})\, d\vec{y}$.

Действительно,

$$\beta(\delta)+c\alpha(\delta) = \mathsf{P}_{H_2}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} < c\right)+q\cdot\mathsf{P}_{H_2}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right)+c\mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c\right)+cp\cdot\mathsf{P}_{H_1}\left(\frac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c\right) =$$

$$= \int\limits_{\Psi_2(\vec{y})<c\Psi_1(\vec{y})} \Psi_2(\vec{y})\, d\vec{y} + q \int\limits_{\Psi_2(\vec{y})=c\Psi_1(\vec{y})} \Psi_2(\vec{y})\, d\vec{y} + \int\limits_{\Psi_2(\vec{y})>c\Psi_1(\vec{y})} c\Psi_1(\vec{y})\, d\vec{y} + p \int\limits_{\Psi_2(\vec{y})=c\Psi_1(\vec{y})} c\Psi_1(\vec{y})\, d\vec{y} =$$

$$= \int\limits_{\Psi_2(\vec{y})<c\Psi_1(\vec{y})} \Psi(\vec{y})\, d\vec{y} + q \int\limits_{\Psi_2(\vec{y})=c\Psi_1(\vec{y})} \Psi(\vec{y})\, d\vec{y} + \int\limits_{\Psi_2(\vec{y})>c\Psi_1(\vec{y})} \Psi(\vec{y})\, d\vec{y} + p \int\limits_{\Psi_2(\vec{y})=c\Psi_1(\vec{y})} \Psi(\vec{y})\, d\vec{y} = \int\limits_{\mathbb{R}^n} \Psi(\vec{y})\, dy.$$

Пусть $\delta_0 \in K_\varepsilon$ — другой критерий уровня ε. Докажем, что его ошибка 2-го рода не меньше, чем у критерия δ.

Используя определение функции Ψ и утверждение 1, имеем:

$$\beta(\delta_0)+c\alpha(\delta_0) = \int\limits_{\delta_0=H_1} \Psi_2(\vec{y})\, d\vec{y}+ \int\limits_{\delta_0=H_2} c\Psi_1(\vec{y})\, d\vec{y} \geqslant \int\limits_{\delta_0=H_1} \Psi(\vec{y})\, d\vec{y}+ \int\limits_{\delta_0=H_2} \Psi(\vec{y})\, d\vec{y} \equiv \int\limits_{\mathbb{R}^n} \Psi(\vec{y})\, d\vec{y} = \beta(\delta)+c\alpha(\delta).$$

Но $\alpha(\delta_0) = \alpha(\delta) \quad \Longrightarrow \quad \beta(\delta_0) \geqslant \beta(\delta)$. $\square$

Пример 29. Имеется выборка X_1 объема 1. Проверяются простые гипотезы $\begin{cases} H_1 : X_1 \Subset \mathrm{U}_{1,5}, \\ H_2 : X_1 \Subset \mathrm{U}_{0,2}. \end{cases}$ Требуется построить НМК уровня $\alpha = 1/3$ (объем выборки мал, так что ошибки не могут не быть большими). Воспользуемся леммой Неймана - Пирсона и выпишем в зависимости от X_1 отношение правдоподобия (см. рисунок).

Рис. 9: Две равномерные гипотезы.

Поскольку отношение правдоподобия постоянно на каждом из трех интервалов, для постоянной c есть лишь несколько возможностей, при которых критерии будут различаться. Перечислим *все* возможные критерии вида

$$\begin{cases} \textit{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} < c & \models\!\!\Rightarrow \quad \delta(\vec{X}) = H_1 \\ \textit{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} > c & \models\!\!\Rightarrow \quad \delta(\vec{X}) = H_2 \\ \textit{если } \dfrac{\Psi_2(\vec{X})}{\Psi_1(\vec{X})} = c & \models\!\!\Rightarrow \quad \begin{cases} \delta(\vec{X}) = H_2 & \text{с вероятностью } p, \\ \delta(\vec{X}) = H_1 & \text{с вероятностью } 1-p, \end{cases} \end{cases}$$

и убедимся, что лишь один из них имеет заданный уровень. Сначала рассмотрим нерандомизированные критерии.

1. $c < 0$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} > c$ при любом X_1, и критерий отношения правдоподобия имеет вид:

$$\delta_1(X_1) = H_2 \text{ при любом } X_1.$$

Его уровень $\alpha(\delta_1) = \mathsf{P}_{H_1}(\delta_1(X_1) = H_2) = 1$ — не то.

2. $0 < c < 2$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} > c$ при $X_1 \in [0,2]$, и $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} < c$ при $X_1 \in (2,5]$. Критерий отношения правдоподобия имеет вид:

$$\delta_2(X_1) = \begin{cases} H_2 & \text{при } X_1 \in [0,2] \\ H_1 & \text{при } X_1 \in (2,5]. \end{cases}$$

Его уровень $\alpha(\delta_2) = \mathsf{P}_{H_1}(\delta_2(X_1) = H_2) = \mathsf{P}_{H_1}(X_1 \in [0,2]) = \mathsf{P}_{H_1}(X_1 \in [1,2]) = 1/4$ — не то.

3. $2 < c < \infty$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} > c$ при $X_1 \in [0,1]$, и $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} < c$ при $X_1 \in (1,5]$. Критерий отношения правдоподобия имеет вид:

$$\delta_3(X_1) = \begin{cases} H_2 & \text{при } X_1 \in [0,1] \\ H_1 & \text{при } X_1 \in (1,5]. \end{cases}$$

Его уровень $\alpha(\delta_3) = \mathsf{P}_{H_1}(\delta_3(X_1) = H_2) = \mathsf{P}_{H_1}(X_1 \in [0,1]) = 0$ — не то.

Таким образом, ни один из нерандомизированных критериев не имеет нужного уровня. Посмотрим на рандомизированные.

4. $c = 0$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} > c$ при $X_1 \in [0,2]$, и $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} = c$ при $X_1 \in (2,5]$. Критерий отношения правдоподобия имеет вид:

$$\begin{cases} \textit{если } X_1 \in [0,2] & \models\!\!\Rightarrow \quad \delta_4(X_1) = H_2 \\ \textit{если } X_1 \in (2,5] & \models\!\!\Rightarrow \quad \begin{cases} \delta_4(X_1) = H_2 & \text{с вероятностью } p, \\ \delta_4(X_1) = H_1 & \text{с вероятностью } 1-p. \end{cases} \end{cases}$$

Его уровень $\alpha(\delta_4) = \mathsf{P}_{H_1}(\delta_4(X_1) = H_2) = \mathsf{P}_{H_1}(X_1 \in [0,2]) + p\mathsf{P}_{H_1}(X_1 \in [2,5]) = 1/4 + p3/4 \neq 1/5$ — не то.

5. $c = 2$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} > c$ при $X_1 \in [0,1]$, $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} < c$ при $X_1 \in [2,5]$, $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} = c$ при $X_1 \in (1,2)$. Критерий отношения правдоподобия имеет вид:

$$\begin{cases} \textit{если } X_1 \in [0,1] & \models\!\!\Rightarrow \quad \delta_5(X_1) = H_2 \\ \textit{если } X_1 \in [2,5] & \models\!\!\Rightarrow \quad \delta_5(X_1) = H_1 \\ \textit{если } X_1 \in (1,2) & \models\!\!\Rightarrow \quad \begin{cases} \delta_5(X_1) = H_2 & \text{с вероятностью } p, \\ \delta_5(X_1) = H_1 & \text{с вероятностью } 1-p. \end{cases} \end{cases}$$

Его уровень $\alpha(\delta_5) = \mathsf{P}_{H_1}(\delta_5(X_1) = H_2) = \mathsf{P}_{H_1}(X_1 \in [0,1]) + p\mathsf{P}_{H_1}(X_1 \in (1,2)) = 0 + p1/4 = 1/5$ при $p = 4/5$ — требуемый уровень.

6. $c = \infty$. В этом случае $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} < c$ при $X_1 \in [1,5]$, $\dfrac{\Psi_2(X_1)}{\Psi_1(X_1)} = c$ при $X_1 \in [0,1)$. Критерий отношения правдоподобия имеет вид:

$$\begin{cases} \text{\textit{если} } X_1 \in [1,5] & \models\!\Rightarrow \quad \delta_6(X_1) = H_1 \\ \text{\textit{если} } X_1 \in [0,1) & \models\!\Rightarrow \quad \begin{cases} \delta_6(X_1) = H_2 & \text{с вероятностью } p, \\ \delta_6(X_1) = H_1 & \text{с вероятностью } 1-p. \end{cases} \end{cases}$$

Его уровень $\alpha(\delta_6) = \mathsf{P}_{H_1}(\delta_6(X_1) = H_2) = p\mathsf{P}_{H_1}(X_1 \in [0,1)) = 0 \neq 1/5$ — не то.

Итак, среди всех критериев отношения правдоподобия только критерий δ_5 имеет уровень $\alpha = 1/5$, и является НМК: любые другие критерии того же уровня имеют меньшую мощность.

Критерий δ_5 предписывает на отрезке [0,1] принимать гипотезу H_2, на отрезке [2,5] — гипотезу H_1, и на отрезке (1,2) — гипотезы H_2 и H_1 с вероятностями $4/5$ и $1/5$ соответственно. Последнее равносильно разбиению отрезка (1,2) на две части в отношении 4:1. Но поскольку таких разбиений континуум, и ни одно из них не лучше другого, то разумнее выбирать гипотезы на отрезке (1,2) случайно, как и делает критерий.

7.6 Вопросы и упражнения

1. Задачник [1], задачи 9.1, 9.2, 9.4 − 9.6, 9.9, 9.12 − 9.14, 9.17.

8 Критерии согласия

Существует класс критериев, называемых *критериями согласия*, которые используются для проверки гипотез (простых или сложных) против сложных альтернатив. Все критерии согласия строятся по единому принципу: задается некоторая *функция отклонения* эмпирического распределения от теоретического. Требуют, чтобы эта функция сходилась к какому-то собственному распределению, если верна проверяемая гипотеза, и неограниченно возрастала, если гипотеза не верна. Гипотеза принимается или отвергается в зависимости от величины данной функции отклонения. Мы сформулируем ряд понятий для случая простой основной гипотезы, а в дальнейшем будем их корректировать по мере изменения задачи.

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathcal{F}$. Проверяется простая основная гипотеза против сложной альтернативы:

$$\begin{cases} H_1 : \mathcal{F} = \mathcal{F}_1, \\ H_2 : \mathcal{F} \neq \mathcal{F}_1 \end{cases}$$

Отметим сразу, что любой критерий для различения этих гипотез $\delta : \mathbb{R}^n \to \{H_1, H_2\}$ имеет вполне определенную ошибку 1-го рода $\alpha(\delta) = \mathsf{P}_{H_1}(\delta \neq H_1) = \mathsf{P}_{\mathcal{F}_1}(\delta \neq H_1)$. Но ошибка 2-го рода может быть вычислена только если известно конкретное распределение выборки $\mathcal{F}_2 \neq \mathcal{F}_1$ (одна из альтернатив).

Поэтому удобно представить альтернативу $H_2 : \mathcal{F} \neq \mathcal{F}_1$ в виде объединения (несчетного числа) простых альтернатив $H_2(F_2) : \mathcal{F} = \mathcal{F}_2 \neq \mathcal{F}_1$ по всем возможным $\mathcal{F}_2$. Будем рассматривать ошибку второго рода как функцию от $\mathcal{F}_2$: $\beta_{\mathcal{F}_2}(\delta) = \mathsf{P}_{\mathcal{F}_2}(\delta = H_1)$.

8.1 Состоятельность критерия

Определение 28. Критерий δ для проверки гипотезы $H_1 : \mathcal{F} = \mathcal{F}_1$ против простой альтернативы $H_2 : \mathcal{F} = \mathcal{F}_2$ называется <u>состоятельным</u>, если

$$\beta(\delta) = \mathsf{P}_{\mathcal{F}_2}(\delta(\vec{X}) = H_1) \to 0 \text{ при } n \to \infty.$$

Определение 29. Критерий δ для проверки гипотезы $H_1 : \mathcal{F} = \mathcal{F}_1$ против сложной альтернативы $H_2 : \mathcal{F} \neq \mathcal{F}_1$ называется <u>состоятельным</u>, если для любой простой альтернативы $H_2(\mathcal{F}_2) : \mathcal{F} = \mathcal{F}_2 \neq \mathcal{F}_1$

$$\beta_{\mathcal{F}_2}(\delta) = \mathsf{P}_{\mathcal{F}_2}(\delta(\vec{X}) = H_1) \to 0 \text{ при } n \to \infty.$$

Определение 30. Говорят, что критерий δ является критерием <u>асимптотического</u> уровня ε, если $\alpha(\delta) \to \varepsilon$ при $n \to \infty$.

8.2 Построение критериев согласия

K1. Требуется задать функцию $\rho(\vec{X}) = \rho(\vec{X}, \mathcal{F}_1)$, обладающую свойствами:

а) если гипотеза H_1 верна, то $\rho(\vec{X}) \Rightarrow \xi \Subset G$ (распределение G известно);

б) если гипотеза H_1 неверна, то $|\rho(\vec{X})| \xrightarrow{p} \infty$ при $n \to \infty$.

K2. Пусть такая функция $\rho(\vec{X})$ задана. Для $\xi \Subset G$ определим постоянную C из равенства $\varepsilon = \mathsf{P}(|\xi| \geqslant C)$ и построим критерий:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C \end{cases} \tag{14}$$

Мы построили критерий согласия. Он «работает» по принципу: если для данной выборки функция отклонения велика (по абсолютному значению), то это свидетельствует в пользу альтернативы, и наоборот. Убедимся в том, что этот критерий имеет (асимптотический) уровень ε и является состоятельным.

Свойство 3. *Для критерия δ, заданного в (14), при $n \to \infty$:*

1. $\alpha(\delta) = \mathsf{P}_{H_1}(|\rho(\vec{X})| \geqslant C) \to \mathsf{P}(|\xi| \geqslant C) = \varepsilon$;

2. $\beta_{\mathcal{F}_2}(\delta) = \mathsf{P}_{\mathcal{F}_2}(|\rho(\vec{X})| < C) \to 0$ *для любой альтернативы* $H_2(\mathcal{F}_2) : \mathcal{F} = \mathcal{F}_2 \neq \mathcal{F}_1$.

Замечание 22. По определению, $\xi_n \xrightarrow{\text{p}} \infty \iff$ для любого $C > 0 \quad \mathsf{P}(\xi_n < C) \to 0$ при $n \to \infty$.

Упражнение. Доказать свойство 3.

8.3 Критерии согласия: критерий Колмогорова

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathcal{F}$. 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 $\rho(\vec{X}) = \sqrt{n} \sup\limits_y |F_n^*(y) - F_1(y)|$. Покажем, что $\rho(\vec{X})$ удовлетворяет условиям **К1(а,б)**:

а) если гипотеза H_1 верна, т. е. $X_i \Subset \mathcal{F}_1$, то по теореме Колмогорова $\rho(\vec{X}) \Rightarrow \xi \Subset K$ — распределение с ф.р. Колмогорова;

б) если гипотеза H_1 неверна, т. е. $X_i \Subset \mathcal{F}_2 \neq \mathcal{F}_1$, то по теореме Гливенко-Кантелли $F_n^*(y) \xrightarrow{\text{p}} F_2(y)$ для любого y при $n \to \infty$. Поскольку $\mathcal{F}_1 \neq \mathcal{F}_2$, найдется y такое, что $|F_2(y) - F_1(y)| > 0$. Для таких y

$$|F_n^*(y) - F_1(y)| \xrightarrow{\text{p}} |F_2(y) - F_1(y)| > 0$$

Поэтому при $n \to \infty$

$$\rho(\vec{X}) = \sqrt{n} \sup_y |F_n^*(y) - F_1(y)| \xrightarrow{\text{p}} \infty.$$

Пусть случайная величина ξ имеет распределение с ф.р. $K(y) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 y^2}$. Это распределение табулировано, так что по заданному ε легко найти C такое, что $\varepsilon = \mathsf{P}(\xi \geqslant C)$.

Критерий Колмогорова выглядит так:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}) < C \\ H_2, & \text{если } \rho(\vec{X}) \geqslant C \end{cases}$$

8.4 Критерии согласия: критерий χ^2 (Пирсона)

Критерий χ^2 основывается на группированных данных. Предполагаемую область значений элементов выборки делят на некоторое число интервалов. После чего строят функцию отклонения ρ по разностям *теоретических вероятностей* попадания в интервалы группировки и *эмпирических частот*.

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathcal{F}$. Проверяется простая гипотеза $H_1 : \mathcal{F} = \mathcal{F}_1$ против сложной альтернативы $H_2 : \mathcal{F} \neq \mathcal{F}_1$.

Пусть, в соответствии с обозначениями 1-й лекции, $A_1, \ldots, A_k$ — интервалы группировки в области значений с.в. с распределением $\mathcal{F}_1$. Обозначим для $j = 1, \ldots, k$ через ν_j число элементов выборки, попавших в интервал A_j

$$\nu_j = \{\text{число } X_i \in A_j\} = \sum_{i=1}^{n} \mathbf{I}(X_i \in A_j),$$

и через p_j теоретическую вероятность $\mathsf{P}_{H_1}(X_1 \in A_j)$ попадания в интервал A_j случайной величины с распределением $\mathcal{F}_1$.

Пусть

$$\rho(\vec{X}) = \rho(\vec{X}, \mathcal{F}_1) = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j}. \tag{15}$$

Замечание 23. Свойство **К1(б)** выполнено далеко не для всех альтернатив. Если распределение выборки $\mathcal{F}_2 \neq \mathcal{F}_1$ имеет те же вероятности p_j попадания в интервалы A_j, $1 \leqslant j \leqslant k$, что и распределение $\mathcal{F}_1$, то по данной функции ρ эти распределения различить нельзя.

Поэтому, на самом деле, критерий χ^2, построенный по данной функции ρ (15), применим для проверки сложной гипотезы

$$\tilde{H}_1: \text{ распределение } X_1 \text{ обладает свойством: } \forall j = 1, \ldots, k \quad \mathsf{P}(X_1 \in A_j) = p_j$$

против сложной альтернативы $\tilde{H}_2 : \tilde{H}_1$ не верна, т.е. хотя бы для одного из интервалов $\mathsf{P}(X_1 \in A_j) \neq p_j$.

Для решения *этой* задачи мы и построим критерий χ^2. Покажем, что $\rho(\vec{X})$ удовлетворяет условиям **К1(а,б)**.

Теорема 12. (Пирсона). *Если верна гипотеза $\tilde{H}_1$, то при фиксированном k и при $n \to \infty$*

$$\rho(\vec{X}) = \sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j} \Rightarrow \chi^2 \Subset \chi^2_{k-1},$$

где χ^2_{k-1} есть χ^2-распределение с $k-1$ степенью свободы.

Доказательство теоремы Пирсона в случае $k = 2$. При бо́льших k см. упражнения ниже. Тогда $\nu_2 = n - \nu_1$, $p_2 = 1 - p_1$. 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 с.в. с распределением Бернулли B_{p_1}, и по ЦПТ

$$\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}} \Rightarrow \xi \Subset \mathrm{N}_{0,1}.$$

Поэтому

$$\rho(\vec{X}) = \left(\frac{\nu_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2 \Rightarrow \xi^2 \Subset \chi^2_1.$$

$\square$

Упражнение. Доказать, что для любого $j = 1, \ldots, k$ $\dfrac{\nu_j - np_j}{\sqrt{np_j}} = \sqrt{n}\dfrac{\frac{\nu_j}{n} - p_j}{\sqrt{p_j}} \Rightarrow \mathrm{N}_{0,(1-p_j)}$.

Вспомнить, как (по лемме Фишера) распределена величина $\sum_{i=1}^{k}(Y_i - \overline{Y})^2$ для выборки $Y_i \Subset \mathrm{N}_{0,1}$, $i = 1, \ldots, k$.

Доказать, что $Y_i - \overline{Y} \Subset \mathrm{N}_{0,1-1/k}$. Провести аналогии (только!) между утверждениями теоремы Пирсона и леммы Фишера.

Упражнение. Почему сумма квадратов k асимптотически нормальных случайных величин сходится к распределению χ^2_{k-1}, а не χ^2_k? Куда делась одна степень свободы? Наводящий вопрос: не связаны ли ν_j каким-либо уравнением?

Свойство **К1(б)**:

Упражнение. Вспомнить ЗБЧ и доказать, что если $\tilde{H}_1$ не верна, то найдется $j \in \{1, \ldots, k\}$ такое, что

$$\frac{(\nu_j - np_j)^2}{np_j} = n\frac{\left(\frac{\nu_j}{n} - p_j\right)^2}{p_j} \xrightarrow{\mathrm{p}} \infty.$$

Пусть случайная величина χ^2 имеет распределение χ^2_{k-1}. Это распределение табулировано, так что по заданному ε можно найти C такое, что $\varepsilon = \mathsf{P}(\chi^2 \geqslant C)$.

Осталось построить критерий согласия χ^2:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}) < C \\ H_2, & \text{если } \rho(\vec{X}) \geqslant C. \end{cases}$$

Замечание 24. 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 состоятелен для альтернатив с теми же вероятностями попадания в интервалы разбиения, что и у $\mathcal{F}_1$. Поэтому берут большое число интервалов разбиения — чем больше, тем лучше. **Но:**

Замечание 25. Сходимость по распределению $\rho(\vec{X}) \Rightarrow \chi^2$ обеспечивается ЦПТ, поэтому разница допредельной и предельной вероятностей ведет себя как

$$|\mathsf{P}(\rho(\vec{X}) \geqslant C) - \mathsf{P}(\chi^2 \geqslant C)| \sim \max\left\{\frac{b}{\sqrt{np_j(1-p_j)}}\right\}$$

(см. точность в ЦПТ, 1-й семестр), где b — некоторая постоянная. Поэтому для выборки объема n число интервалов разбиения выбирают так, чтобы обеспечить нужную точность при замене распределения $\rho(\vec{X})$ на χ^2_{k-1}. Обычно требуют, чтобы $np_1 = \ldots = np_k \approx 6 \div 9$.

8.5 Критерий χ^2 (Пирсона). Параметрическая гипотеза

Критерий χ^2 часто применяют для проверки гипотезы о виде распределения, или о принадлежности распределения выборки некоторому параметрическому семейству.

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathcal{F}$. Проверяется сложная гипотеза
$H_1 : \mathcal{F} \in \{\mathcal{F}_\theta, \theta \in \Theta\}$, где θ — неизвестный параметр (скалярный или векторный), против сложной же альтернативы
$H_2 : \mathcal{F} \notin \{\mathcal{F}_\theta\}$.

Пусть опять $A_1, \ldots, A_k$ — интервалы группировки, ν_j — число элементов выборки, попавших в A_j. Но теоретическая вероятность $p_j = \mathsf{P}_{H_1}(X_1 \in A_j) = p_j(\theta)$ зависит от неизвестного параметра θ.

Пусть $\widehat{\theta}$ — ОМП для параметра θ. Взяв в (15) вместо p_j оценки $\widehat{p}_j = p_j(\widehat{\theta})$, получим функцию отклонения

$$\rho(\vec{X}) = \sum_{j=1}^{k} \frac{(\nu_j - n\widehat{p}_j)^2}{n\widehat{p}_j}. \tag{16}$$

Замечание 26. Иначе говоря, в функции (16) вероятности $\widehat{p}_j$ попадания в интервалы A_j вычислены для распределения $\mathcal{F}_{\widehat{\theta}}$.

Отметим снова, что критерий χ^2, построенный по данной функции ρ (16), применим лишь для проверки гипотезы

$$\tilde{H}_1: \text{ распределение } X_1 \text{ обладает свойством: } \forall j = 1, \ldots, k \quad \mathsf{P}(X_1 \in A_j) = \widehat{p}_j$$

против сложной альтернативы $\tilde{H}_2 : \tilde{H}_1$ не верна.

Условие **K1(а)** (при выполнении некоторых условий относительно гладкости $p_j(\theta)$) обеспечивается теоремой:

Теорема 13. (Пирсона). *Если верна гипотеза $\tilde{H}_1$, и $\dim(\theta) = m$ — размерность параметра (вектора) θ, то при фиксированном k и при $n \to \infty$*

$$\rho(\vec{X}) = \sum_{j=1}^{k} \frac{(\nu_j - n\widehat{p}_j)^2}{n\widehat{p}_j} \Rightarrow \chi^2 \Subset \chi^2_{k-m-1},$$

где χ^2_{k-m-1} есть χ^2-распределение с $k-m-1$ степенью свободы.

Выполнение условия **K1(б)** очевидно.

Пусть теперь случайная величина χ^2 имеет распределение χ^2_{k-m-1}. По заданному ε найдем C такое, что $\varepsilon = \mathsf{P}(\chi^2 \geqslant C)$.

Критерий согласия χ^2 имеет вид:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \rho(\vec{X}) < C \\ H_2, & \text{если } \rho(\vec{X}) \geqslant C. \end{cases}$$

Замечание 27. Замечания 24, 25 остаются в силе.

8.6 Проверка гипотезы однородности: критерий Колмогорова - Смирнова

Есть две выборки $\vec{X} = (X_1, \ldots, X_n)$ и $\vec{Y} = (Y_1, \ldots, Y_m)$, причем $X_i \Subset \mathcal{F}_x$, $Y_i \Subset \mathcal{F}_y$, и распределения $\mathcal{F}_x, \mathcal{F}_y$, вообще говоря, неизвестны. Проверяется сложная гипотеза H_1: $\mathcal{F}_x = \mathcal{F}_y$ против (еще более сложной) альтернативы H_2: H_1 не верна.

Если $\mathcal{F}_x, \mathcal{F}_y$ имеют непрерывные функции распределения, применим критерий Колмогорова - Смирнова.

Пусть $F^*_{n,x}$ и $F^*_{m,y}$ — эмпирические функции распределения, построенные по выборкам $\vec{X}$ и $\vec{Y}$,

$$\rho(\vec{X}, \vec{Y}) = \sqrt{\frac{mn}{m+n}} \sup_t |F^*_{n,x}(t) - F^*_{m,y}(t)|.$$

Теорема 14. *Если гипотеза H_1 верна, то $\rho(\vec{X}, \vec{Y}) \Rightarrow \xi \Subset K$ (распределение с ф.р. Колмогорова) при $n, m \to \infty$.*

Упражнение. Доказать, что $\rho(\vec{X}, \vec{Y}) \xrightarrow{\mathsf{p}} \infty$ при $n, m \to \infty$, если гипотеза H_2 верна.

И снова: пусть случайная величина ξ имеет распределение с ф.р. $K(y)$. По заданному ε найдем C такое, что $\varepsilon = \mathsf{P}(\xi \geqslant C)$, и построим критерий согласия 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 28. Если есть более двух выборок, и требуется проверить гипотезу однородности, часто пользуются одним из вариантов критерия χ^2 Пирсона. Этот критерий (и ряд других критериев) рекомендую посмотреть в §3.4.2, с. 124 книги
Г. И. Ивченко, Ю. И. Медведев, Математическая статистика. Москва, 1984, 248 с.

8.7 Проверка гипотезы независимости: критерий χ^2 Пирсона

Есть выборка $(\vec{X}, \vec{Y}) = ((X_1, Y_1), \ldots, (X_n, Y_n))$ значений двух наблюдаемых совместно величин ξ и η в n экспериментах. Проверяется гипотеза H_1: ξ и η независимы.

Введем k интервалов группировки для значений ξ: $\Delta^x_1, \ldots, \Delta^x_k$ и m интервалов группировки для значений η: $\Delta^y_1, \ldots, \Delta^y_m$.

Посчитаем эмпирические частоты: для $i = 1, \ldots, k$, $j = 1, \ldots, m$
$\nu_{i,j} = \{\text{число пар } (X_l, Y_l) \in \Delta^x_i \times \Delta^y_j\}$,
$\nu_{\cdot,j} = \{\text{число пар } (X_l, Y_l) : Y_l \in \Delta^y_j\} = \{\text{число } Y_l \in \Delta^y_j\}$,
$\nu_{i,\cdot} = \{\text{число пар } (X_l, Y_l) : X_l \in \Delta^x_i\} = \{\text{число } X_l \in \Delta^x_i\}$.

$\vec{Y}$ / $\vec{X}$	Δ^y_1	Δ^y_2	$\ldots$	Δ^y_m	Σ
Δ^x_1	$\nu_{1,1}$	$\nu_{1,2}$	$\ldots$	$\nu_{1,m}$	$\nu_{1,\cdot}$
Δ^x_2	$\nu_{2,1}$	$\nu_{2,2}$	$\ldots$	$\nu_{2,m}$	$\nu_{2,\cdot}$
$\vdots$			$\ldots$		
Δ^x_k	$\nu_{k,1}$	$\nu_{k,2}$	$\ldots$	$\nu_{k,m}$	$\nu_{k,\cdot}$
Σ	$\nu_{\cdot,1}$	$\nu_{\cdot,2}$	$\ldots$	$\nu_{\cdot,m}$	n

Если гипотеза H_1 верна, то теоретические вероятности попадания пары (X_1, Y_1) в любую из областей $\Delta_i^x \times \Delta_j^y$ равны произведению вероятностей:

$$p_{i,j} = \mathsf{P}((X_1, Y_1) \in \Delta_i^x \times \Delta_j^y) = \mathsf{P}(X_1 \in \Delta_i^x)\mathsf{P}(Y_1 \in \Delta_j^y) = p_i^x p_j^y$$

По ЗБЧ,

$$\frac{\nu_{i,\cdot}}{n} \approx p_i^x, \quad \frac{\nu_{\cdot,j}}{n} \approx p_i^y, \quad \frac{\nu_{i,j}}{n} \approx p_{i,j}.$$

Поэтому значительная разница между $\dfrac{\nu_{i,j}}{n}$ и $\dfrac{\nu_{i,\cdot}}{n}\dfrac{\nu_{\cdot,j}}{n}$, или между $\nu_{i,j}$ и $\dfrac{\nu_{i,\cdot}\nu_{\cdot,j}}{n}$ может служить основанием для отклонения гипотезы независимости.

Пусть

$$\rho(\vec{X}, \vec{Y}) = n \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(\nu_{i,j} - (\nu_{i,\cdot}\nu_{\cdot,j})/n)^2}{\nu_{i,\cdot}\nu_{\cdot,j}} = n \left(\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\nu_{i,j}^2}{\nu_{i,\cdot}\nu_{\cdot,j}} - 1 \right).$$

Теорема 15. *Если гипотеза H_1 верна, то $\rho(\vec{X}, \vec{Y}) \Rightarrow \chi^2_{(k-1)(m-1)}$ при $n \to \infty$.*

Как обычно, строится критерий согласия асимптотического уровня ε.

Упражнение. Для каких альтернатив построенный критерий является состоятельным?

Замечание 29. Замечания 24, 25 остаются в силе.

8.8 Гипотеза о совпадении средних двух нормальных совокупностей с равными дисперсиями

Есть две независимые выборки $\vec{X} = (X_1, \ldots, X_n)$ и $\vec{Y} = (Y_1, \ldots, Y_m)$, причем $X_i \Subset \mathrm{N}_{a_1,\sigma^2}$, $Y_i \Subset \mathrm{N}_{a_2,\sigma^2}$, и дисперсия σ^2 одинакова для обоих распределений, но, вообще говоря, неизвестна. Проверяется сложная гипотеза H_1: $a_1 = a_2$ против сложной альтернативы H_2: H_1 не верна.

Разумеется, эта задача есть частный случай задачи об однородности, и можно построить критерий Колмогорова - Смирнова асимптотического уровня ε. Но мы построим другой критерий — критерий Стьюдента *точного* уровня ε.

Обозначим через $S_0^2(\vec{X}), S_0^2(\vec{Y})$ несмещенные выборочные дисперсии:

$$S_0^2(\vec{X}) = \frac{1}{n-1} \sum_{1}^{n} (X_i - \overline{X})^2, \quad S_0^2(\vec{Y}) = \frac{1}{m-1} \sum_{1}^{m} (Y_i - \overline{Y})^2.$$

Из леммы Фишера следует

Теорема 16. *Случайная величина t_{n+m-2} имеет распределение Стьюдента:*

$$t_{n+m-2} = \sqrt{\frac{nm}{n+m}} \frac{(\overline{X} - a_1) - (\overline{Y} - a_2)}{\sqrt{\frac{(n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})}{n+m-2}}} \Subset \mathrm{T}_{n+m-2}.$$

***Доказательство теоремы* 16.** Соглашение: $\mathrm{N}_{0,1} \equiv \mathrm{N}(0,1)$.

1. Легко видеть (**Упражнение:** убедиться, что легко):

$$\overline{X} - a_1 \Subset \mathrm{N}(0, \frac{\sigma^2}{n}), \overline{Y} - a_2 \Subset \mathrm{N}(0, \frac{\sigma^2}{m}) \quad \Longrightarrow$$

$$\Longrightarrow \quad (\overline{X} - a_1) - (\overline{Y} - a_2) \Subset \mathrm{N}(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}) = \mathrm{N}(0, \sigma^2 \frac{n+m}{nm}) \quad \Longrightarrow$$

$$\Longrightarrow \quad \xi_0 = \frac{1}{\sigma} \sqrt{\frac{nm}{n+m}} ((\overline{X} - a_1) - (\overline{Y} - a_2)) \Subset \mathrm{N}(0,1).$$

2. Из леммы Фишера следует, что

$$\frac{(n-1)}{\sigma^2}S_0^2(\vec{X}) \Subset \chi^2_{n-1}, \frac{(m-1)}{\sigma^2}S_0^2(\vec{Y}) \Subset \chi^2_{m-1}$$

$$\models\!\!\Rightarrow \qquad S \overset{\text{def}}{=} \frac{1}{\sigma^2}\Big((n-1)S_0^2(\vec{X}) + (m-1)S_0^2(\vec{Y})\Big) \Subset \chi^2_{n+m-2},$$

и не зависит от $\overline{X}, \overline{Y}$.

3. Тогда (см. распределение Стьюдента) $\dfrac{\xi_0}{\sqrt{S/(n+m-2)}} \Subset \mathrm{T}_{n+m-2}$. Осталось подставить в эту дробь ξ_0 и S и убедиться, что σ сократится и что получится в точности t_{n+m-2} из теоремы 16.

□

Введем функцию отклонения $\rho(\vec{X}, \vec{Y})$:

$$\rho = \sqrt{\frac{nm}{n+m}}\frac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{(n-1)S_0^2(\vec{X})+(m-1)S_0^2(\vec{Y})}{n+m-2}}}.$$

Из теоремы 16 следует свойство **K1(a)**:
Если H_1 верна ($a_1 = a_2$) $\quad \models\!\!\Rightarrow \quad \rho = t_{n+m-2} \Subset \mathrm{T}_{n+m-2}$.
Упражнение. Доказать свойство **K1(б)**:
Если H_2 верна ($a_1 \neq a_2$) $\quad \models\!\!\Rightarrow \quad |\rho| \xrightarrow{\text{p}} \infty$.

Указания. Воспользовавшись ЗБЧ или утверждениями лемм 1-3 из 1-й лекции, доказать, что числитель и знаменатель сходятся к постоянным: $\overline{X}-\overline{Y} \xrightarrow{\text{p}} \text{const} \neq 0$, $\dfrac{(n-1)S_0^2(\vec{X})+(m-1)S_0^2(\vec{Y})}{n+m-2} \xrightarrow{\text{p}}$ const $\neq 0$, тогда как корень перед дробью неограниченно возрастает.

Поэтому остается по ε выбрать C такое, что для величины $t_{n+m-2} \Subset \mathrm{T}_{n+m-2}$, в силу симметричности распределения Стьюдента,

$$\varepsilon = \mathsf{P}(|t_{n+m-2}| > C) = 2\mathsf{P}(t_{n+m-2} > C) \quad \models\!\!\Rightarrow \quad \mathsf{P}(t_{n+m-2} > C) = \varepsilon/2 \quad \models\!\!\Rightarrow \quad C = \tau_{1-\varepsilon/2},$$

где $\tau_{1-\varepsilon/2}$ — квантиль распределения T_{n+m-2}.
И критерий Стьюдента выглядит как все критерии согласия:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C. \end{cases}$$

Упражнение. Доказать, что этот критерий имеет точный уровень ε.

8.9 Гипотеза о среднем нормальной совокупности с известной дисперсией

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \Subset \mathrm{N}_{a,\sigma^2}$, и дисперсия σ^2 известна. Проверяется простая гипотеза H_1: $a = a_0$ против сложной альтернативы H_2: $a \neq a_0$.

Можно построить критерий Колмогорова (или χ^2) асимптотического уровня ε. Но мы (как и в предыдущей задаче) построим критерий *точного* уровня ε.

Введем функцию отклонения $\rho(\vec{X})$:

$$\rho = \sqrt{n}\frac{\overline{X}-a_0}{\sigma}.$$

Очевидно свойство **K1(a)**:
Если H_1 верна ($a = a_0$) $\quad \models\!\!\Rightarrow \quad \rho \Subset \mathrm{N}_{0,1}$.
Упражнение. Доказать свойство **K1(б)**:
Если H_2 верна ($a \neq a_0$) $\quad \models\!\!\Rightarrow \quad |\rho| \xrightarrow{\text{p}} \infty$.

Поэтому по ε выберем C такое, что для величины $\xi \in \mathrm{N}_{0,1}$, в силу симметричности стандартного нормального распределения,

$$\varepsilon = \mathsf{P}(|\xi| > C) = 2\mathsf{P}(\xi > C) \quad \models\!\!\Rightarrow \quad \mathsf{P}(\xi > C) = \varepsilon/2 \quad \models\!\!\Rightarrow \quad C = \tau_{1-\varepsilon/2},$$

где $\tau_{1-\varepsilon/2}$ — квантиль стандартного нормального распределения.

Критерий выглядит как все критерии согласия:

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } |\rho(\vec{X})| < C \\ H_2, & \text{если } |\rho(\vec{X})| \geqslant C. \end{cases} \tag{17}$$

Упражнение. Доказать, что этот критерий имеет точный уровень ε.

8.10 Гипотеза о среднем нормальной совокупности с неизвестной дисперсией

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \in \mathrm{N}_{a,\sigma^2}$, и дисперсия σ^2 неизвестна. Проверяется простая гипотеза H_1: $a = a_0$ против сложной альтернативы H_2: $a \neq a_0$.

Введем функцию отклонения $\rho(\vec{X})$:

$$\rho = \sqrt{n}\frac{\overline{X} - a_0}{S_0^2}, \quad \text{где } S_0^2 = \frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2.$$

Из леммы Фишера следует свойство **K1(а)**:

Если H_1 верна ($a = a_0$) $\quad \models\!\!\Rightarrow \quad \rho \in \mathrm{T}_{n-1}$.

Упражнение. Доказать свойство **K1(б)**:

Если H_2 верна ($a \neq a_0$) $\quad \models\!\!\Rightarrow \quad |\rho| \xrightarrow{\mathrm{p}} \infty$.

Поэтому по ε выберем C такое, что для величины $t_{n-1} \in \mathrm{T}_{n-1}$

$$\varepsilon = \mathsf{P}(|t_{n-1}| > C) = 2\mathsf{P}(t_{n-1} > C) \quad \models\!\!\Rightarrow \quad C = \tau_{1-\varepsilon/2},$$

где $\tau_{1-\varepsilon/2}$ — квантиль распределения T_{n-1}.

Критерий выглядит как все критерии согласия.

Упражнение. Нарисовать критерий и доказать, что этот критерий имеет точный уровень ε.

Упражнение. В самом ли деле три последних рассмотренных критерия состоятельны? **Напоминание.** А Вы доказали выполнение свойства **K1(б)** для этих критериев, чтобы говорить о состоятельности? **Примечание.** А что такое «состоятельность» критерия?

8.11 Гипотеза о параметрах распределения: критерии, основанные на доверительных интервалах

Имеется выборка $\vec{X} = (X_1, \ldots, X_n)$, $X_i \in \mathcal{F}_\theta$. Проверяется простая гипотеза $H_1 : \theta = \theta_0$ против сложной альтернативы $H_2 : \theta \neq \theta_0$.

Есть еще один разумный способ строить критерии согласия (помимо поиска функции отклонения ρ). Пусть имеется точный (асимптотический) доверительный интервал $(\theta^-, \theta^+) = (\theta^-(\varepsilon, \vec{X}), \theta^+(\varepsilon, \vec{X}))$ для параметра θ уровня доверия $1 - \varepsilon$: для любого θ

$$\mathsf{P}_\theta(\theta^- < \theta < \theta^+) = \mathsf{P}(\theta^- < \theta < \theta^+ \mid X_i \in \mathcal{F}_\theta) = 1 - \varepsilon \; (\to 1 - \varepsilon).$$

Тогда критерий

$$\delta(\vec{X}) = \begin{cases} H_1, & \text{если } \theta_0 \in (\theta^-, \theta^+) \\ H_2, & \text{если } \theta_0 \notin (\theta^-, \theta^+) \end{cases}$$

имеет точный (асимптотический) уровень ε. Действительно,

$$\alpha(\delta) = \mathsf{P}_{H_1}(\delta = H_2) = \mathsf{P}(\theta \notin (\theta^-, \theta^+) \mid X_i \in \mathcal{F}_{\theta_0}) = 1 - \mathsf{P}_{\theta_0}(\theta^- < \theta_0 < \theta^+) = \varepsilon \; (\to \varepsilon).$$

Если доверительный интервал строится с помощью «функции отклонения» $G(\vec{X}, \theta)$ (см. способы построения доверительных интервалов), то эта же функция годится и в качестве «функции отклонения» $\rho(\vec{X}, \mathcal{F})$ (см. способы построения критериев согласия).

Пример 30. Посмотрим на критерий (17) в задаче о параметре нормальной выборки с известной дисперсией.

$$\delta(\vec{X}) = H_1 \iff |\rho(\vec{X})| < C = \tau_{1-\varepsilon/2} \iff |\sqrt{n}\frac{\overline{X} - a_0}{\sigma}| < \tau_{1-\varepsilon/2}$$
$$\iff \overline{X} - \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}} < a_0 < \overline{X} + \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}}.$$

Осталось вспомнить, что точный доверительный интервал для параметра a нормального распределения с известной дисперсией как раз и есть $(\overline{X} - \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}}, \overline{X} + \frac{\tau_{1-\varepsilon/2}\sigma}{\sqrt{n}})$.

Упражнение. Какие из приведенных выше критериев можно сформулировать, используя доверительные интервалы? Сделать это.

9 Исследование статистической зависимости: линейная регрессия

Часто требуется определить зависимость наблюдаемой случайной величины от одной или нескольких других величин. Самый общий случай такой зависимости — зависимость статистическая: например, наблюдаемая с.в. X есть функция от двух с.в. ξ и η, а с.в. Z — от ξ и ϕ. Зависимость между X и Z есть, но она не является, вообще говоря, функциональной зависимостью. Наличие такой зависимости может быть проверено, скажем, по критерию χ^2, если имеется выборка из значений пары с.в. (X, Z).

Мы рассмотрим один из случаев этой задачи, когда имеет смысл предполагать наличие «почти функциональной» зависимости между двумя величинами. Часто эту зависимость можно воображать как вход и выход некоторой машины («ящика с шуршавчиком»). Входные данные («факторы»), как правило, известны. На выходе мы наблюдаем результат преобразования входных данных в ящике по каким-либо правилам. Мы можем получать даже значения случайных величин, которые (в среднем) функционально зависят от входных данных. При этом строгая функциональная зависимость входных и выходных данных редко имеет место, чаще на нее накладываются случайные «помехи»: ошибки наблюдения, воздействие неучтенных внешних факторов (случайность, наконец) и т.д.

9.1 Модель регрессии

Рассматривается модель, в которой наблюдаемая случайная величина X зависит от другой случайной величины Y (значения которой мы либо задаем, либо знаем). Пусть зависимость математического ожидания X от значений Y определяется формулой $\mathsf{E}(X|Y = t) = f(t)$, где f — неизвестная функция. После n экспериментов с входными данными $Y = t_1, \ldots, Y = t_n$ (какие-то заданные числа или вектора, природа которых чаще всего не имеет значения) получены значения $X_1, \ldots, X_n$.

Пусть $X_1 = f(t_1) + \varepsilon_1$, …, $X_n = f(t_n) + \varepsilon_n$, где ε_i — ошибки наблюдения, равные в точности разнице между реальным и усредненным значением случайной величины X при значении $Y = t_i$: $\varepsilon_i = X_i - f(t_i) = X_i - \mathsf{E}(X|Y = t_i)$.

Требуется по значениям $t_1, \ldots, t_n$ и $X_1, \ldots, X_n$ оценить как можно точнее функцию f.

9.2 Метод наименьших квадратов: примеры

Даже в отсутствие ошибок наблюдений функцию f можно восстановить лишь приближенно, в виде полинома. Поэтому обычно предполагают, что f есть полином (редко больше третьей - четвертой степени) с неизвестными коэффициентами. Метод наименьших квадратов состоит в выборе этих коэффициентов так, чтобы минимизировать сумму квадратов ошибок $\sum_{i=1}^n (X_i - f(t_i))^2$.

Пример 31. Пусть $X_i = \theta + \varepsilon_i$, $i = 1, \ldots, n$, где θ — неизвестный параметр. Здесь f — полином нулевой степени. Найдем оценку $\widehat{\theta}$ для параметра θ, на которой достигается минимум величины $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (X_i - \theta))^2$.

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n}(X_i - \theta))^2 = -2\sum_{i=1}^{n}(X_i - \theta))\Big|_{\theta=\widehat{\theta}} = 0 \quad \Longrightarrow \quad \widehat{\theta} = \overline{X}.$$

Определение 31. Оценка $\widehat{\theta}$ параметра θ, на которой достигается $\min\limits_{\theta} \sum_{i=1}^n \varepsilon_i^2$, называется оценкой <u>метода наименьших квадратов</u> (ОМНК).

Пример 32. **Линейная регрессия** Рассмотрим линейную регрессию $X_i = \theta_1 + t_i\theta_2 + \varepsilon_i$, $i = 1, \ldots, n$, где $\theta = (\theta_1, \theta_2)$ — неизвестный параметр. Здесь f — полином первой степени (прямая). Найдем оценку МНК $(\widehat{\theta}_1, \widehat{\theta}_1)$ для параметра θ, на которой достигается минимум величины $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (X_i - \theta_1 - t_i\theta_2))^2$.

Обозначим $\overline{t} = \frac{1}{n}\sum t_i$. Приравняв к нулю частные производные, найдем точку экстремума.

Упражнение. Убедиться, что решением системы уравнений

$$\frac{\partial}{\partial\theta_1}\sum_{i=1}^{n}\varepsilon_i^2\Big|_{\theta=\widehat{\theta}}=0; \qquad \frac{\partial}{\partial\theta_2}\sum_{i=1}^{n}\varepsilon_i^2\Big|_{\theta=\widehat{\theta}}=0$$

является пара

$$\widehat{\theta_2}=\frac{\frac{1}{n}\sum t_iX_i-\overline{X}\cdot\overline{t}}{\frac{1}{n}\sum(t_i-\overline{t})^2}; \qquad \widehat{\theta_1}=\overline{X}-\overline{t}\widehat{\theta_2}.$$

Линия $\mathsf{E}X=\widehat{\theta_1}+t\widehat{\theta_2}$ называется *линией регрессии* X на t.

Определение 32. Величина

$$\rho^*=\frac{\frac{1}{n}\sum t_iX_i-\overline{X}\cdot\overline{t}}{\sqrt{\frac{1}{n}\sum(t_i-\overline{t})^2\frac{1}{n}\sum(X_i-\overline{X})^2}}$$

называется «выборочным коэффициентом корреляции» и характеризует степень линейной зависимости между X и t.

Пример 33. Полиномиальная регрессия Модель регрессии имеет вид $\mathsf{E}X=\theta_0+\theta_1t+\theta_2t^2+\ldots+\theta_{k-1}t^{k-1}$. В следующем параграфе будет показано, как эта модель сводится к общей модели линейной регрессии.

Пример 34. Термин «регрессия» появился впервые в работе Francis Galton, "Regression towards mediocrity in hereditary stature" (Journal of the Anthropological Institute V. 15, p. 246–265, 1886).

Гальтон (в частности) исследовал рост детей высоких родителей, и установил, что он «регрессирует» в среднем, то есть в среднем дети высоких родителей не так высоки, как их родители. Для линейной модели регрессии $X=\theta_1t+\theta_2u+c$ Гальтон нашел оценки параметров:

$$\text{Рост сына}=0,27\ \text{Роста отца}+0,2\ \text{Роста матери}+\text{const},$$

а рост дочери еще в 1,08 раз меньше.

9.3 Общая модель линейной регрессии

Замечание 30. В отличие от обозначений Лекции 6, отныне и навеки все вектора есть вектора-столбцы. Благодарю А. Иванова за соответствующее замечание и надеюсь, что не только он заметил несоответствие обозначений в лемме Фишера курсу алгебры.

Введем два вектора: $\vec{Z}=(Z_1,\ldots,Z_k)$ — вектор факторов регрессии и $\vec{\beta}=(\beta_1,\ldots,\beta_k)$ — вектор неизвестных параметров регрессии. Рассматривается модель регрессии, которая в курсе «Эконометрия» называется «простой (линейной) регрессией»:

$$\mathsf{E}X=\beta_1Z_1+\ldots+\beta_kZ_k.$$

Пусть в i-м эксперименте факторы регрессии принимают (известные) значения $\overrightarrow{Z^{(i)}}=(Z_1^{(i)},\ldots,Z_k^{(i)})$, $i=1,\ldots,n$.

Пусть $\vec{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n)$, где ε_i — случайная ошибка в i-м эксперименте (неизвестна).

После n экспериментов ($n\geqslant k$) в данной модели получена выборка $\vec{X}=(X_1,\ldots,X_n)$, где

$$\begin{cases} X_1=\beta_1Z_1^{(1)}+\ldots+\beta_kZ_k^{(1)}+\varepsilon_1\\ X_2=\beta_1Z_1^{(2)}+\ldots+\beta_kZ_k^{(2)}+\varepsilon_2\\ \ldots\\ X_n=\beta_1Z_1^{(n)}+\ldots+\beta_kZ_k^{(n)}+\varepsilon_n,\end{cases}$$

или, в матричной форме, $\boxed{\vec{X} = Z^T\vec{\beta} + \vec{\varepsilon}}$, где матрица $Z(k\times n)$ называется «матрицей плана»:

$$Z = \begin{pmatrix} Z_1^{(1)} & \dots & Z_1^{(n)} \\ \vdots & \dots & \vdots \\ Z_k^{(1)} & \dots & Z_k^{(n)} \end{pmatrix} = (\overrightarrow{Z^{(1)}} \ \dots \ \overrightarrow{Z^{(n)}}).$$

Требуется по данным матрице плана Z и вектору результатов $\vec{X}$ найти оценки для параметров регрессии $\vec{\beta}$ и вектора ошибок $\vec{\varepsilon}$ (например, для его числовых характеристик $\mathsf{E}\varepsilon$, $\mathsf{D}\varepsilon$, матрицы ковариаций и т.д.). Мы рассмотрим только проблему оценивания $\vec{\beta}$.

Предположение 1. *Матрица Z имеет ранг k, т.е. все k ее строк линейно независимы.*

Лемма 12. *Предположение 1 $\iff$ матрица $A = Z{\cdot}Z^T$ положительно определена.*

Напоминание 1. *Матрица $A(k\times k)$ положительно определена, если $\vec{t}^{\,T}A\vec{t} \geqslant 0$ для любого $\vec{t} = (t_1,\dots,t_k)$, и $\vec{t}^{\,T}A\vec{t} = 0 \iff \vec{t} \equiv 0$.*

Напоминание 2. *Норма вектора (столбца) $\vec{u} = (u_1,\dots,u_k)$ есть $\vec{u}^{\,T}\vec{u} = \sum_{i=1}^k u_i^2 \geqslant 0$, и равна нулю $\iff \vec{u} \equiv 0$.*

Доказательство леммы 12. Благодаря напоминанию 2, $\vec{t}^{\,T}A\vec{t} = \vec{t}^{\,T}Z{\cdot}Z^T\vec{t} = (Z^T\vec{t}\,)^T{\cdot}(Z^T\vec{t}\,) \geqslant 0$, причем $(Z^T\vec{t}\,)^T\cdot(Z^T\vec{t}\,) = 0 \iff Z^T\vec{t} = 0$.

Но «ранг Z равен k» как раз и означает, по определению, что $Z^T\vec{t} = 0 \iff \vec{t} = 0$.

$\square$

9.4 Метод наименьших квадратов. Нормальное уравнение

Обозначим

$$S(\vec{\beta}\,) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(X_i - \sum_{j=1}^k \beta_j Z_j^{(i)} \right)^2 = (\vec{X} - Z^T\vec{\beta}\,)^T \cdot (\vec{X} - Z^T\vec{\beta}\,).$$

Определение 33. Если $S(\widehat{\beta}) = \min\limits_{\vec{\beta}} S(\vec{\beta}\,)$, то $\widehat{\beta}$ называется оценкой метода наименьших квадратов (ОМНК) вектора $\vec{\beta}$. Здесь $\widehat{\beta} = (\widehat{\beta}_1,\dots,\widehat{\beta}_k)$.

Найдем систему уравнений, определяющих точку экстремума функции $S(\vec{\beta}\,)$ (пока только точку экстремума!).

$$\frac{\partial S}{\partial \beta_m} = -2\sum_{i=1}^n Z_m^{(i)} \left(X_i - \sum_{j=1}^k \beta_j Z_j^{(i)} \right)\Bigg|_{\vec{\beta}=\widehat{\beta}} = 0, \ m = 1,\dots,k.$$

Раскрыв скобки, получим систему

$$\sum_{i=1}^n Z_m^{(i)} X_i = \sum_{i=1}^n Z_m^{(i)} \sum_{j=1}^k \beta_j Z_j^{(i)}, \ m = 1,\dots,k. \tag{18}$$

Слева стоит m-я координата вектора $Z\vec{X}$, справа — m-я координата вектора $ZZ^T\vec{\beta}$, так что систему уравнений 18 можно записать в виде

$$Z\vec{X} = Z{\cdot}Z^T\vec{\beta} = A\vec{\beta}.$$

Определение 34. Уравнение $Z\vec{X} = A\vec{\beta}$ называется нормальным уравнением.

Теорема 17. *1. Если $\widehat{\beta}$ — произвольное решение нормального уравнения, то $\widehat{\beta}$ — ОМНК.*
2. Если $\det A \neq 0 \quad \models\!\!\!\Rightarrow \quad \widehat{\beta} = A^{-1}Z\vec{X}$ — единственная ОМНК.

Стоп! Упражнение. *Что именно* нужно доказывать в п. 1 теоремы? Нужно ли доказывать п. 2?

Замечание 31. Предположение 1 и лемма 12 $\models\!\!\!\Rightarrow$ $\det A \neq 0$ (т.к. иначе система уравнений $A\vec{t} = \vec{0}$ имеет ненулевые решения (и много) $\models\!\!\!\Rightarrow$ A не положительно определена).

***Доказательство теоремы* 17.**

$$S(\vec{\beta}\,) = (\vec{X} - Z^T\vec{\beta}\,)^T \cdot (\vec{X} - Z^T\vec{\beta}\,) = \left(\vec{X} - Z^T\widehat{\beta} + Z^T(\widehat{\beta} - \vec{\beta}\,)\right)^T \cdot \left(\vec{X} - Z^T\widehat{\beta} + Z^T(\widehat{\beta} - \vec{\beta}\,)\right) =$$

$$= S(\widehat{\beta}) + (\widehat{\beta} - \vec{\beta}\,)^T(Z\vec{X} - A\widehat{\beta}) + \left((\widehat{\beta} - \vec{\beta}\,)^T\cdot(Z\vec{X} - A\widehat{\beta})\right)^T + (\widehat{\beta} - \vec{\beta}\,)^T Z\cdot Z^T(\widehat{\beta} - \vec{\beta}\,).$$

Поскольку $\widehat{\beta}$ — решение нормального уравнения, второе и третье слагаемое обращается в ноль. Далее,

$$S(\vec{\beta}\,) = S(\widehat{\beta}) + (\widehat{\beta} - \vec{\beta}\,)^T\, A\,(\widehat{\beta} - \vec{\beta}\,) \geqslant S(\widehat{\beta})$$

поскольку матрица A положительно определена. Итак, $\widehat{\beta}$ — ОМНК, т.к. она минимизирует $S(\vec{\beta}\,)$.

□

Пример 35. Полиномиальная регрессия (продолжение примера 33) Имеем n наблюдений

$$\begin{cases} X_1 = 1\cdot\theta_0 + t_1\theta_1 + t_1^2\theta_2 + \ldots + t_1^{k-1}\theta_{k-1} + \varepsilon_1 \\ \ldots \\ X_n = 1\cdot\theta_0 + t_n\theta_1 + t_n^2\theta_2 + \ldots + t_n^{k-1}\theta_{k-1} + \varepsilon_n \end{cases}$$

Эта модель сводится к линейной модели регрессии с матрицей плана

$$Z = \begin{pmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_n \\ t_1^2 & \ldots & t_n^2 \\ \vdots & \ldots & \vdots \\ t_1^{k-1} & \ldots & t_n^{k-1} \end{pmatrix}.$$

9.5 Свойства ОМНК

1. *Разница $\widehat{\beta}$ и $\vec{\beta}$:*

$$\widehat{\beta} = A^{-1}Z\vec{X} = A^{-1}Z(Z^T\vec{\beta} + \vec{\varepsilon}\,) = A^{-1}A\vec{\beta} + A^{-1}Z\vec{\varepsilon} = \vec{\beta} + A^{-1}Z\vec{\varepsilon}.$$

Предположение 2. *Ошибки $\varepsilon_1, \ldots, \varepsilon_n$ некоррелированы, все имеют нулевые математические ожидания и ненулевую дисперсию $\mathsf{D}\varepsilon_i = \mathsf{E}\varepsilon_i^2 = \sigma^2 < \infty$, $i = 1, \ldots, n$.*

Замечание 32. Напоминаю, что когда речь заходит о дисперсии, термины «ненулевая» и «положительная» означают одно и то же.

Замечание 33. Ниже символом $\mathsf{D}\vec{\varepsilon} = \mathsf{E}(\vec{\varepsilon}\,\vec{\varepsilon}^{\,T})$ обозначена матрица ковариаций вектора $\vec{\varepsilon}$, т.е. матрица, (i,j)-й элемент которой равен

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \mathsf{E}(\varepsilon_i - \mathsf{E}\varepsilon_i)(\varepsilon_j - \mathsf{E}\varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j \end{cases}.$$

Для произвольного случайного вектора $\vec{x}$, координаты которого имеют вторые моменты, $\mathsf{D}\vec{x} = \mathsf{E}\left((\vec{x} - \mathsf{E}\vec{x})(\vec{x} - \mathsf{E}\vec{x})^T\right)$ — матрица, (i,j)-й элемент которой равен

$$\operatorname{cov}(\vec{x}_i, \vec{x}_j) = \mathsf{E}(\vec{x}_i - \mathsf{E}\vec{x}_i)(\vec{x}_j - \mathsf{E}\vec{x}_j).$$

2. $\widehat{\beta}$ — *несмещенная оценка для* $\vec{\beta}$:

$$\mathsf{E}\widehat{\beta} = \vec{\beta} + A^{-1}Z\mathsf{E}\vec{\varepsilon} \equiv \vec{\beta}.$$

3. *Матрица ковариаций вектора* $\widehat{\beta}$:

$$\mathsf{D}\widehat{\beta} = \mathsf{E}(\widehat{\beta} - \mathsf{E}\widehat{\beta})(\widehat{\beta} - \mathsf{E}\widehat{\beta})^T = \mathsf{E}(\widehat{\beta} - \vec{\beta}\,)(\widehat{\beta} - \vec{\beta}\,)^T =$$

$$\mathsf{E}(A^{-1}Z\vec{\varepsilon}\,)(A^{-1}Z\vec{\varepsilon}\,)^T = \mathsf{E}(A^{-1}Z\vec{\varepsilon}\,\vec{\varepsilon}^{\,T}Z^TA^{-1^T})$$

И так как $A = ZZ^T$, $A^T = A$, $\mathsf{E}\vec{\varepsilon}\,\vec{\varepsilon}^{\,T} = \sigma^2 E$, где E — единичная матрица, имеем:

$$\mathsf{D}\widehat{\beta} = \sigma^2(A^{-1}ZZ^TA^{-1^T}) = \sigma^2A^{-1}.$$

9.6 Оптимальный выбор матрицы плана

Как и в теории точечного оценивания, о качестве несмещенной оценки $\widehat{\beta}$ для вектора $\vec{\beta}$ можно судить по величине ее «среднего квадратического отклонения», которое в многомерном случае описывают величиной $\mathsf{D}\widehat{\beta} = \mathsf{E}(\widehat{\beta} - \vec{\beta}\,)(\widehat{\beta} - \vec{\beta}\,)^T$.

Если мы управляем экспериментом, т.е. можем сами задавать значения факторов $Z_1, \ldots, Z_k$ (матрицу плана) и наблюдать затем результаты n экспериментов $X_1, \ldots, X_n$, то разумно задаться вопросом: *как зависит* $\mathsf{D}\widehat{\beta}$ *от выбора матрицы плана* Z?

Введем следующее ограничение:

Предположение 3. *Пусть $a_1, \ldots, a_k$ — ненулевые фиксированные числа. Будем рассматривать матрицы плана Z, у которых*

$$\sum_{i=1}^{n} \left(Z_j^{(i)}\right)^2 = a_j, \quad j = 1, \ldots, k.$$

Замечание 34. Мы доказали, что $\mathsf{D}\widehat{\beta} = \sigma^2A^{-1}$ — матрица ковариаций вектора $\widehat{\beta}$. Ее диагональные элементы равны $\mathsf{D}\widehat{\beta}_j = \sigma^2(A^{-1})_{jj}$. Если не требовать выполнения Предположения 3, то заменой Z на cZ можно добиться, что A^{-1} заменится на $\frac{1}{c^2}A^{-1}$ и вместо $\mathsf{D}\widehat{\beta}_j = \sigma^2(A^{-1})_{jj}$ получится $\mathsf{D}\widehat{\beta}_j = \frac{1}{c^2}\sigma^2(A^{-1})_{jj}$. (А получится ли? Что-то тут не так! :-))***

Следующая теорема может быть доказана из свойств матрицы A (рекомендую попытаться доказать):

Теорема 18. *Предположение 3* $\Longrightarrow$ $\mathsf{D}\widehat{\beta}_j \geqslant \dfrac{\sigma^2}{a_j^2}$ *для любого $j = 1, \ldots, k$. В неравенстве достигается равенство если (и только если) строки матрицы Z ортогональны, т.е.*

$$\sum_{i=1}^{n} Z_l^{(i)}Z_j^{(i)} = 0 \quad \forall \quad l \neq j.$$

Следствие 6. *Если строки матрицы Z ортогональны, то матрица A имеет диагональный вид, и $(A)_{jj} = a_j^2$. Это означает, что координаты вектора $\widehat{\beta}$ некоррелированы.*

9.7 Вопросы и упражнения

1. Выполнить третью часть расчетного задания.

Литература

[1] Сборник задач по математической статистике. Под редакцией А.А.Боровкова. Новосибирск: НГУ, 1989.

[2] Боровков А.А. Математическая статистика. Ч.I,II. Новосибирск: НГУ, 1983,1984.

[3] Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1965.

[4] Чистяков В.П. Курс теории вероятностей. М., 1982.