Московский физико-технический институт
Физтех-школа прикладной математики и информатики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

V СЕМЕСТР

Лектор: *Максим Павлович Савёлов*

h\nu

Автор: Максимов Даниил
Проект на Github

осень 2023

Содержание

1 Напоминание теории вероятностей

Замечание. В этом разделе мы живём в вероятностном пространстве $(\Omega, \mathcal{F}, P)$

Напоминание. Пусть $\xi, \{\xi_n\}_{n=1}^{\infty}$ — случайные векторы из $\mathbb{R}^m$. Тогда мы рассматриваем следующие сходимости:

1. Сходимость P-почти наверное (с вероятностью 1)

$$\xi_n \xrightarrow{P \text{ п.н.}} \xi \Longleftrightarrow P(\xi_n \to \xi) = 1$$

2. Сходимость по вероятности

$$\xi_n \xrightarrow{P} \xi \Longleftrightarrow \forall \varepsilon > 0 \;\; P(\|\xi_n - \xi\|_2 > \varepsilon) \xrightarrow[n\to\infty]{} 0$$

3. Сходимость в среднем порядка $p \geqslant 1$ (по норме L_p)

$$\xi_n \xrightarrow{L_p} \xi \Longleftrightarrow \mathbb{E}\|\xi_n - \xi\|_p^p \xrightarrow[n\to\infty]{} 0$$

4. Сходимость по распределению

$$\xi_n \xrightarrow{d} \xi \Longleftrightarrow \forall f\colon \mathbb{R}^m \to \mathbb{R} \text{ — ограниченная непрерывная } \; \mathbb{E}f(\xi_n) \xrightarrow[n\to\infty]{} \mathbb{E}f(\xi)$$

Напоминание. Для всех сходимостей из векторной сходимости следует покоординатная. В обратную сторону это неверно только для сходимости по распределению.

Доказательство.

1. Для сходимости с вероятностью 1 достаточно заметить соотношение:

$$\forall j \in \{1, \ldots, m\} \quad \bigcap_{i=1}^{m} \{\xi_{i,n} \to \xi_i\} = \{\xi_n \to \xi\} \subseteq \{\xi_{j,n} \to \xi_j\}$$

2. Для сходимости по вероятности всё же нужно 2 отдельных вложения (для любого $\varepsilon > 0$):

$\Rightarrow \; \{|\xi_{i,n} - \xi_i| > \varepsilon\} \subseteq \{\|\xi_n - \xi\|_2 > \varepsilon\}$

$\Leftarrow \; \bigcup_{i=1}^{m} \{|\xi_{i,n} - \xi_i| > \varepsilon\} \supseteq \{\|\xi_n - \xi\|_2 > \varepsilon\}$

3. Покомпонентная сходимость из векторной тривиальна, а в обратную сторону нужно разложить вектор на сумму векторов с лишь одной его компонентой и воспользоваться неравенством треугольника. Тогда всё следует из предполагаемого условия (покомпонентная сходимость):

$$\mathbb{E}\sum_{i=1}^{m} \|\xi_{i,n} - \xi_i\|_p^p \xrightarrow[n\to\infty]{} 0$$

4. Доказать нужно (и возможно) только в одну сторону. Зафиксируем $g\colon \mathbb{R} \to \mathbb{R}$ — непрерывную ограниченную функцию и рассмотрим $h_i(x_1, \ldots, x_m) = x_i$ — функция проектора. Тогда композиция $g \circ h$ является ограниченной непрерывной функцией $\mathbb{R}^m \to \mathbb{R}$, а значит можем воспользоваться предположением:

$$\mathbb{E}g(\xi_{i,n}) = \mathbb{E}g(h(\xi_n)) \xrightarrow[n\to\infty]{} \mathbb{E}g(h(\xi)) = \mathbb{E}g(\xi_i)$$

□

Напоминание. Имеют место следующие посылки:

- ▷ $(\xi_n \xrightarrow{P \text{ п.н.}} \xi) \Rightarrow (\xi_n \xrightarrow{P} \xi)$
- ▷ $(\xi_n \xrightarrow{L_p} \xi) \Rightarrow (\xi_n \xrightarrow{P} \xi)$
- ▷ $(\xi_n \xrightarrow{P} \xi) \Rightarrow (\xi_n \xrightarrow{d} \xi)$

Утверждение 1.1. *Если $\xi_n \to^d c$, где $c = const \in \mathbb{R}^m$, то $\xi_n \to^P c$*

Доказательство. Перейдём к сходимостям в координатах, а для них мы уже доказали эту лемму в курсе теории вероятностей:

$$(\xi_n \xrightarrow{d} c) \Rightarrow (\xi_{i,n} \xrightarrow{d} c_i) \Longrightarrow (\xi_{i,n} \xrightarrow{P} c_i) \Rightarrow (\xi_n \xrightarrow{P} c)$$

□

Теорема 1.1. *(О наследовании сходимостей) Пусть существует $B \in \mathfrak{B}(\mathbb{R}^m)$ такое, что $P(\xi \in B) = 1$ и $h\colon \mathbb{R}^m \to \mathbb{R}^k$ непрерывна в каждой точке множества B. Тогда верны импликации:*

1. $\xi_n \xrightarrow{P \text{ п.н.}} \xi \Longrightarrow h(\xi_n) \xrightarrow{P \text{ п.н.}} h(\xi)$
2. $\xi_n \xrightarrow{P} \xi \Longrightarrow h(\xi_n) \xrightarrow{P} h(\xi)$
3. $\xi_n \xrightarrow{d} \xi \Longrightarrow h(\xi_n) \xrightarrow{d} h(\xi)$

Доказательство.

1. $P(h(\xi_n) \to h(\xi)) \geqslant P(h(\xi_n) \to h(\xi) \wedge \xi \in B) \geqslant P(\xi_n \to \xi \wedge \xi \in B) = 1$

2. Предположим противное. Это означает следующее:

$$\exists \varepsilon_0 > 0 \; \exists \delta_0 > 0 \; \exists \{n_k\}_{k=1}^{\infty} \; \Big| \; P(\|h(\xi_{n_k}) - h(\xi)\| > \varepsilon_0) \geqslant \delta_0$$

   При этом $\xi_{n_k} \xrightarrow{P} \xi$. Как известно, из такой последовательности можно извлечь подпоследовательность, которая будет сходиться P-почти наверное:

$$\xi_{n_{k_j}} \xrightarrow{P \text{ п.н.}} \xi \Rightarrow h(\xi_{n_{k_j}}) \xrightarrow{P \text{ п.н.}} h(\xi) \Rightarrow h(\xi_{n_{k_j}}) \xrightarrow{P} h(\xi)$$

   Получили противоречие с предположением

3. Докажем случай лишь когда h просто непрерывна в $\mathbb{R}^m$. Зафиксируем $f\colon \mathbb{R}^k \to \mathbb{R}$ — непрерывная ограниченная функция. В силу условия, $f \circ h$ тоже непрерывна и ограничена на $\mathbb{R}^m$. Отсюда из сходимости по распределению:

$$\mathbb{E}f(h(\xi_n)) \xrightarrow[n\to\infty]{} \mathbb{E}f(h(\xi))$$

Всё доказанное вместе означает по определению, что $h(\xi_n) \xrightarrow{d} h(\xi)$.

□

Утверждение 1.2. *(без доказательства) Пусть $\{\eta_n\}_{n=1}^\infty$ — случайные вектора из $\mathbb{R}^s$, причём $\eta_n \xrightarrow{d} c = const \in \mathbb{R}^s$ и также $\xi_n \xrightarrow{d} \xi$. Тогда верна векторная сходимость по распределению:*

$$\begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix} \xrightarrow[n\to\infty]{d} \begin{pmatrix} \xi \\ c \end{pmatrix} \in \mathbb{R}^{m+s}$$

Следствие. (Лемма Слуцкого) Пусть $\{\eta_n\}_{n=1}^\infty$ — случайные величины, причём $\eta_n \xrightarrow{d} c \in \mathbb{R}$ и также $\xi_n \xrightarrow{d} \xi$. Тогда верны такие сходимости:

- ▷ $\xi_n + \eta_n \xrightarrow{d} \xi + c$
- ▷ $\xi_n \cdot \eta_n \xrightarrow{d} \xi \cdot c$

Доказательство. Просто комбинируем утверждение без доказательства и теорему о наследовании сходимости для таких $f(x, y)$:

- ▷ $f(x, y) = x + y$
- ▷ $f(x, y) = xy$

□

Теорема 1.2. *(Дельта-метод, одномерный случай) Пусть ξ_n, ξ — случайные величины, $H\colon \mathbb{R} \to \mathbb{R}$ и $\{b_n\}_{n=1}^\infty \subset \mathbb{R}$, на которые наложены следующие условия:*

- ▷ $\xi_n \xrightarrow{d} \xi$
- ▷ *$H \in D(a)$, где $a \in \mathbb{R}$ — фиксированная точка*
- ▷ $\lim_{n\to\infty} b_n = 0$
- ▷ $b_n \neq 0$

Тогда верна сходимость:

$$\frac{H(a + \xi_n b_n) - H(a)}{b_n} \xrightarrow[n\to\infty]{d} H'(a)\xi$$

Доказательство. Идея состоит в том, чтобы просто применить теорему о наследовании сходимостей. Итак, определим h следующим образом:

$$h(x) := \begin{cases} \dfrac{H(a + x) - H(a)}{x}, \, x \neq 0 \\ H'(a), \, x = 0 \end{cases}$$

Тогда h непрерывна на $\mathbb{R}$. По лемме Слуцкого $b_n\xi_n \xrightarrow{d} 0 \cdot \xi = 0$. Осталось применить уже упомянутую теорему о наследовании:

$$\frac{H(a + b_n\xi_n) - H(a)}{b_n\xi_n} = h(b_n\xi_n) \xrightarrow{d} h(0) = H'(a)$$

Повторно используем лемму Слуцкого с доказанной сходимостью и $\xi_n \xrightarrow{d} \xi$, это и даёт утверждение теоремы. $\square$

Теорема 1.3. *(Дельта-метод, многомерный случай) Пусть ξ_n, ξ — случайные вектора из $\mathbb{R}^m$, $H\colon \mathbb{R}^m \to \mathbb{R}^s$, $a \in \mathbb{R}^m$ и $\{b_n\}_{n=1}^{\infty} \subset \mathbb{R}$, на которые наложены следующие условия:*

- $\xi_n \xrightarrow{d} \xi$
- $H \in D(a)$
- $b_n \to 0$
- $b_n \neq 0$

Тогда верна сходимость:

$$\frac{H(a + \xi_n b_n) - H(a)}{b_n} \xrightarrow[n\to\infty]{d} H'(a)\xi$$

2 Основные определения

Замечание. Далее $(\Omega, \mathcal{F})$ и $(E, \mathcal{E})$ — измеримые пространства.

Определение 2.1. Отображение $\xi\colon \Omega \to E$ называется *случайным элементом*, если оно $\mathcal{F}/\mathcal{E}$-измеримо:

$$\forall B \in \mathcal{E} \;\; \xi^{-1}(B) \in \mathcal{F}$$

Определение 2.2. Случайный элемент ξ называется *случайным вектором*, если $(E, \mathcal{E}) = (\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$.

Определение 2.3. Случайный вектор ξ называется *случайной величиной*, если $(E, \mathcal{E}) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$.

Замечание. Далее мы считаем, что $(\Omega, \mathcal{F}, P)$ — вероятностное пространство

Определение 2.4. Пусть ξ — случайный элемент. Тогда его *распределением* называется мера P_ξ на $(E, \mathcal{E})$, заданная следующим образом:

$$\forall B \in \mathcal{E} \;\; P_\xi(B) := P(\xi \in B)$$

Замечание. Статистика, в отличие от теории вероятностей, напрямую связана с проводимыми экспериментами и их результатами. Так, пусть наблюдается некоторый эксперимент, который можно описать m-мерным случайным вектором с распределением P. Мы должны построить вероятностное пространство и измеримый на нём случайный элемент, соответствующий этому вектору.

Определение 2.5. Рассмотрим вероятностное пространство $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), P)$, определённое следующим образом:

- ▷ $\mathcal{X}$ — выборочное множество всевозможных исходов одного эксперимента (если рассматриваем m-мерный вектор, то можно взять $\mathbb{R}^m$). Требуем, что $\mathcal{X}$ является топологическим пространством.
- ▷ $\mathfrak{B}(\mathcal{X})$ — борелевская σ-алгебра подмножеств $\mathcal{X}$
- ▷ P — мера, соответствующая распределению исходов одного эксперимента

Такое пространство называется *вероятностно-статистической моделью.*

Замечание. В самом деле, если мы рассмотрим тождественный случайный элемент χ в таком пространстве, то $P_\chi = P$. Таким образом, этот случайный элемент соответствует проводимому эксперименту.

Определение 2.6. (Не по лектору) Любой случайный элемент в вероятностно-статистической модели, действующий в $\mathcal{X}$, называется *наблюдением.*

Замечание. Покажем, как построить математическую модель для n независимых повторений одного и того же эксперимента:

- ▷ $\mathcal{X}^n = \underbrace{\mathcal{X} \times \ldots \times \mathcal{X}}_{n}$
- ▷ $\mathfrak{B}^n(\mathcal{X}) = \mathfrak{B}(\mathcal{X}^n) = \sigma(\{B_1 \times \ldots \times B_n, B_i \in \mathfrak{B}(\mathcal{X})\})$
- ▷ $P^n = \underbrace{P \otimes \ldots \otimes P}_{n}$ — тензорное произведение n мер

Несложно показать, что P^n — действительно вероятностная мера на $(\mathcal{X}^n, \mathfrak{B}(\mathcal{X}^n))$.

Утверждение 2.1. *Рассмотрим пространство $(\mathcal{X}^n, \mathfrak{B}(\mathcal{X}^n), P^n)$ и тождественный случайный вектор $\chi\colon \mathcal{X}^n \to \mathcal{X}^n$. Тогда любая компонента этого вектора будет случайной величиной того же пространства, причём всем компоненты независимы в совокупности и имеют распределение P.*

Доказательство. Доказывать измеримость компоненты мы не будем, это тривиально из теории вероятностей. Распишем распределение χ_i по определению:

$$\forall B_i \in \mathfrak{B}(\mathcal{X}) \quad P^n(\chi_i \in B_i) = P^n\{(x_1, \ldots, x_n)\colon x_i \in B_i\} = \\ P^n(\mathcal{X} \times \ldots \times \mathcal{X} \times B_i \times X \times \ldots \times \mathcal{X}) = P(B_i)$$

Независимость тоже проверяем через эквивалентное определение:

$$\forall B_i \in \mathfrak{B}(\mathcal{X}) \quad P^n(\chi_1 \in B_1 \wedge \ldots \wedge \chi_n \in B_n) = \\ P^n\{(x_1, \ldots, x_n)\colon x_1 \in B_1 \wedge \ldots \wedge x_n \in B_n\} = \\ P^n(B_1 \times \ldots \times B_n) = P(B_1) \cdot \ldots \cdot P(B_n) = P^n(\chi_1 \in B_1) \cdot \ldots \cdot P^n(\chi_n \in B_n)$$

□

Определение 2.7. Совокупность $X = (X_1, \ldots, X_n)$ независимых одинаково распределённых наблюдений с распределением P называется *выборкой размера n из распределения P*. Число n также называют *объёмом выборки.*

Замечание. Для счётного числа экспериментов модель строится аналогичным образом. Там уже происходит работа с вероятностным пространством $(\mathcal{X}^\infty, \mathfrak{B}(\mathcal{X}^\infty), P^\infty)$. Тот факт, что P^∞ существует и согласовано со всеми возникающими P^n, гарантируется теоремой Колмогорова.

Замечание. Всюду далее для простоты опускаются индексы пространств $(\mathcal{X}^n, \mathfrak{B}(\mathcal{X}^n), P^n)$ и $(\mathcal{X}^\infty, \mathfrak{B}(\mathcal{X}^\infty), P^\infty)$, пишем просто $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), P)$ и фиксируем это обозначение для вероятностно-статистической модели

Определение 2.8. Значение выборки на конкретном исходе называется *реализацией выборки.*

Замечание. Основная задача статистики — это сделать вывод о неизвестном распределении выборки по её реализации.

Определение 2.9. Пусть $X_1, \ldots, X_n$ — выборка из распределения P_X на пространстве $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$, а также $B \in \mathfrak{B}(\mathbb{R}^m)$. Тогда $P_n^*(B) = \frac{1}{n}\left(\sum_{i=1}^n \chi\{X_i \in B\}\right)$ называется *эмпирическим распределением, построенным по выборке* $X_1, \ldots, X_n$.

Замечание. Внимательный читатель заметит, что это эмпирическое распределение является *случайным*, то есть неявно зависит от $\omega \in \Omega$. Более того, это распределение является *случайной величиной при фиксированном $B \in \mathfrak{B}(\mathbb{R}^m)$*, ну а при фиксированном $\omega \in \Omega$ это действительно является распределением (вероятностной мерой на $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$).

Утверждение 2.2. *Пусть $X_1, \ldots, X_n$ — выборка на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ из распределения P_X на пространстве $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$. Имеет место сходимость:*

$$\forall B \in \mathfrak{B}(\mathbb{R}^m) \quad P_n^*(B) \xrightarrow[n\to\infty]{P\ \text{п.н.}} P_X(B)$$

Доказательство. Доказательство опирается на УЗБЧ. Действительно, раз $\{X_i\}_{i=1}^n$ — это выборка, то $\chi\{X_i \in B\}$ при фиксированном B являются тоже независимыми одинаково распределёнными случайными величинами (ибо выражены из выборки через борелевские функции). Стало быть, верна сходимость:

$$\forall B \in \mathfrak{B}(\mathbb{R}^m) \quad \frac{\sum_{i=1}^n \chi\{X_i \in B\}}{n} = P_n^*(B) \xrightarrow[n\to\infty]{P\ \text{п.н.}} \mathbb{E}\chi\{X_1 \in B\} = P_X(B)$$

□

Замечание. До конца главы мы считаем, что $m = 1$ и $X_1, \ldots, X_n$ — случайные величины, образующие выборку из распределения P_X в пространстве $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$

Определение 2.10. Функция $F_n^*(x) = \frac{1}{n}\left(\sum_{i=1}^n \chi\{X_i \leqslant x\}\right)$ называется *эмпирической функцией распределения.*

Замечание. Опять же, отметим, что эмпирическая функция распределения зависит неявно от $\omega \in \Omega$. При каждом отдельном ω эта функция действительно является функцией распределения

Следствие. Имеет место сходимость:

$$\forall x \in \mathbb{R} \quad F_n^*(x) \xrightarrow[n\to\infty]{P\text{ п.н.}} F(x)$$

Доказательство. Следует из того, что $F_n^*(x) = P_n^*(-\infty; x]$ □

Теорема 2.1. *(Гливенко-Кантелли) Пусть $\{X_i\}_{i=1}^{\infty}$ — независимые одинаково распределённые случайные величины с функцией распределения $F(x)$. Тогда имеет место сходимость:*

$$D_n := \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow[n\to\infty]{P\text{ п.н.}} 0$$

Замечание. То есть P-почти наверное имеет место равномерная сходимость $F_n^*(x)$ к $F(x)$.

Доказательство. Для начала нужно установить, что D_n является случайной величиной. Это действительно так, ведь F просто непрерывна справа, а при любом $\omega \in \Omega$ и F_n^* тоже непрерывна справа. Значит, в силу всюду плотности $\mathbb{Q}$ в $\mathbb{R}$ можем записать супремум следующим образом:

$$D_n(\omega) = \sup_{x\in\mathbb{R}} |F_n^*(\omega, x) - F(x)| = \sup_{x\in\mathbb{Q}} |F_n^*(\omega, x) - F(x)|$$

Стало быть, D_n является случайной величиной, коль скоро это супремум по счётной совокупности случайных величин. Чтобы показать стремление D_n к нулю, сделаем зажимающие оценки сверху и снизу. Нам потребуется «порезать» $F(x)$ на дробные кусочки. Эти кусочки мы зададим соответствующими точками:

$$\forall N \in \mathbb{N}\ \forall K \in \{1, \ldots, N-1\}\ x_{N,K} = \inf\left\{x \in \mathbb{R} \colon F(x) \geqslant \frac{K}{N}\right\}$$

Естественно определим $x_{N,0} = -\infty$ и $x_{N,N} = +\infty$. Теперь мы готовы писать оценки:

$\leqslant$ Пусть $x \in [x_{N,K}; x_{N,K+1})$. Тогда верна такая цепочка неравенств:

$$\begin{aligned} F_n^*(x) - F(x) \leqslant F_n^*(x_{N,K+1} - 0) - F(x_{N,K}) = \\ (F_n^*(x_{N,K+1} - 0) - F(x_{N,K+1} - 0)) + (F(X_{N,K+1} - 0) - F(x_{N,K})) \leqslant \\ F_n^*(x_{N,K+1} - 0) - F(x_{N,K+1} - 0) + \frac{1}{N} \end{aligned}$$

$\geqslant$ Аналогичным методом получаем оценку $F_n^*(x) - F(x) \geqslant F_n^*(x_{N,K}) - F(x_{N,K}) - \frac{1}{N}$

Стало быть, верна такая общая оценка на модуль:

$$|F_n^*(x) - F(x)| \leqslant \max\left(|F_n^*(x_{N,K+1} - 0) - F(x_{N,K+1} - 0)|, |F_n^*(x_{N,K}) - F(x_{N,K})|\right) + \frac{1}{N}$$

Отсюда же получаем оценку на супремум:

$$\begin{aligned} \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \leqslant \\ \max_{0\leqslant k\leqslant N-1} \left(|F_n^*(x_{N,K+1} - 0) - F(x_{N,K+1} - 0)|, |F_n^*(x_{N,K}) - F(x_{N,K})|\right) + \frac{1}{N} \end{aligned}$$

Понятно, что при стремлении $n \to \infty$ второй элемент максимума точно устремится к нулю, но вот с первым это не совсем ясно. Проясним этот момент через определения и уже известные факты:

$$\forall y \in \mathbb{R} \quad F_n^*(y-0) = P_n^*(-\infty; y) \xrightarrow[n\to\infty]{P \text{ п.н.}} P_X(-\infty; y) = F(y-0)$$

Стало быть, и первый элемент стремится к нулю. Чтобы не оставить пробелов в формализме, зафиксируем $\varepsilon > 0$ и выберем такое N, что $\frac{1}{N} < \varepsilon$. В силу всего сказанного выше, остаётся сказать следующее:

$$\overline{\lim_{n\to\infty}} \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| <^{P \text{ п.н.}} \varepsilon$$

Ну а это уже напрямую означает, что $D_n \to^{P \text{ п.н.}} 0$ □

3 Параметрическая модель

Определение 3.1. Вероятностно-статистическая модель $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ называется *параметрической*, если $\mathcal{P}$ является не мерой, а параметризованным множеством мер на $(\mathcal{X}, \mathfrak{B}(\mathcal{X}))$, то есть

$$\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$$

Замечание. Мы считаем верной импликацию $(\theta_1 \neq \theta_2) \Rightarrow (P_{\theta_1} \neq P_{\theta_2})$

Замечание. Как правило, $\Theta \subseteq \mathbb{R}^k$. Можно вспомнить базовые примеры, такие как $Exp(\theta)$, $Cauchy(\sigma)$, $N(a, \sigma^2)$

Замечание автора. Далее, когда встречаются понятия, такие как наблюдение, в рамках параметрической модели, то их нужно понимать в вероятностно-статистической модели $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), P_\theta)$ для некоторого $\theta \in \Theta$

3.1 Статистики и оценки

Определение 3.2. Пусть $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ — вероятностно-статистическая модель, X — наблюдение, $(E, \mathcal{E})$ — измеримое пространство, $S \colon \mathcal{X} \to E$ — измеримое отображение. Тогда $S(X)$ называют *статистикой.*

Замечание. Как правило, $(E, \mathcal{E}) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ или $(\mathbb{R}^2, \mathfrak{B}(\mathbb{R}^2))$.

Пример. Пусть $\mathcal{X} = \mathbb{R}^n$. Тогда для выборки $X = (X_1, \ldots, X_n)$ можно рассмотреть статистику среднего значения:

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

Определение 3.3. Пусть X — наблюдение в параметрической модели $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$, а $S(X)$ — статистика, чьё множество значений вложено в Θ. Тогда $S(X)$ называется *оценкой неизвестного параметра θ.*

Замечание. Аналогичное определение можно ввести и для $\tau(\theta)$, где τ — некоторая произвольная функция.

Пример. Приведём несколько базовых классов статистик. Далее $X = (X_1, \ldots, X_n)$ — это выборка из некоторого распределения в $(\mathbb{R}^m, \mathfrak{B}(\mathbb{R}^m))$:

1. Если $g\colon \mathbb{R}^m \to \mathbb{R}^n$ — борелевская функция, то *выборочной характеристикой функции* $g(x)$ называется следующая статистика:

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

   - ▷ $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ — выборочное среднее
   - ▷ $(m = 1)$ $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ — выборочный момент k-го порядка

2. Если h — борелевская функция, то *функцией от выборочных квантилей* называется статистика следующего вида:

$$S(X) = h(\overline{g_1(X)}, \ldots, \overline{g_k(X)})$$

   - ▷ $(m = 1)$ $S^2 = \overline{X^2} - \overline{X}^2$ — выборочная дисперсия. Здесь $h(x, y) = x - y^2$
   - ▷ $M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$ — выборочный центральный момент k-го порядка

3. $(m = 1)$ *k-й порядковой статистикой* называется k-й по величине снизу элемент.

   - ▷ $X_{(1)} = \min(X_1, \ldots, X_n)$
   - ▷ $X_{(2)} =$ второй по порядку возрастания элемент выборки
   - ▷ $X_{(n)} = \max(X_1, \ldots, X_n)$

   При этом вектор $(X_{(1)}, \ldots, X_{(n)})$ называется *вариационным рядом*

Упражнение. *Доказать, что* $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = M_2$

Замечание. Отметим, что до этого мы смогли ввести понятие эмпирической функции распределения. Для неё верен факт $F_n^*(x) \xrightarrow{P_\theta \text{ п.н.}} F(x)$. Стало быть, имеет место и слабая сходимость этих функций, отсюда получаем связь между статистиками и их матожиданиями (в довольно частном случае, но его очевидно можно обобщать):

$\forall g\colon \mathbb{R} \to \mathbb{R}$ — непрерывная ограниченная

$$\overline{g(X)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i) = \int_{\mathbb{R}} g(x)dF_n^*(x) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} g(x)dF(x) = \mathbb{E}g(X_i)$$

Утверждение 3.1. *Для выборочной дисперсии S^2 выборки X верна формула:*

$$\mathbb{E}S^2 = \frac{n-1}{n}DX_1$$

Доказательство. Просто раскроем среднее и пересоберём формулу:

$$\mathbb{E}S^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2 - \frac{2}{n}\sum_{j=1}^{n}X_iX_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}X_jX_k\right) =$$
$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^2 - \frac{2}{n}X_i^2 - \frac{2}{n}\sum_{j\neq i}X_iX_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}X_jX_k + \frac{1}{n^2}\sum_{j=1}^{n}X_j^2\right)$$

Внесём в отдельные части матожидание и посчитаем их:

- ▷ $\mathbb{E}\left(\frac{n-2}{n}X_i^2\right) = \frac{n-2}{n}\mathbb{E}X_i^2$
- ▷ $\mathbb{E}\left(\frac{2}{n}\sum_{j\neq i}X_iX_j\right) = \frac{2}{n}\sum_{j\neq i}(\mathbb{E}X_i)\mathbb{E}X_j = \frac{2}{n}\mathbb{E}X_i\sum_{j\neq i}\mathbb{E}X_j = \frac{2(n-1)}{n}(\mathbb{E}X_1)^2$
- ▷ $\mathbb{E}\left(\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}X_jX_k\right) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}(\mathbb{E}X_j)\mathbb{E}X_k = \frac{n-1}{n^2}\sum_{j=1}^{n}(\mathbb{E}X_1)^2 = \frac{n-1}{n}(\mathbb{E}X_1)^2$
- ▷ $\mathbb{E}\left(\frac{1}{n^2}\sum_{j=1}^{n}X_j^2\right) = \frac{1}{n}\mathbb{E}X_1^2$

Итого:

$$\mathbb{E}S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{n-1}{n}\mathbb{E}X_1^2 - \frac{2(n-1)}{n}(\mathbb{E}X_1)^2 + \frac{n-1}{n}(\mathbb{E}X_1)^2\right) =$$
$$\frac{n-1}{n}(\mathbb{E}X_1^2 - (\mathbb{E}X_1)^2) = \frac{n-1}{n}DX_1$$

□

Замечание. Собственно, наблюдается неожиданная проблема: наша статистика немного смещена от реальной дисперсии, но это очень просто поправить!

$$S_0^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{ — \textit{исправленная выборочная дисперсия}}$$

3.2 Свойства оценок

Замечание. Далее, если не оговорено явно иного, мы живём в параметрической модели $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ и рассматриваем выборку $X = (X_1, \ldots, X_n)$ с неизвестным распределением $P = P_\theta \in \mathcal{P}$

Определение 3.4. Оценка $\theta^*(X)$ называется *несмещённой оценкой параметра* θ, если выполнено условие:

$$\forall \theta \in \Theta \ \ \mathbb{E}_\theta\theta^*(X) = \theta$$

где $\mathbb{E}_\theta$ — интеграл Лебега (матожидание) по мере P_θ

Замечание автора. При работе с вероятностно-статистическим пространством с параметром θ принято писать не $\mathbb{E}$, а $\mathbb{E}_\theta$, чтобы подчеркнуть зависимость интеграла от θ.

Пример. Рассмотрим $\mathcal{P} = \{N(\theta, 1) \colon \theta \in \mathbb{R}\}$. Тогда оценки X_1 и $\overline{X}$ — несмещённые

Замечание автора. Последовательность оценок $\{\theta_n^*(X)\}_{n=1}^{\infty}$ мы тоже будем называть оценкой $\theta_n^*(X)$ далее.

Определение 3.5. Пусть $\{X_k\}_{k=1}^{\infty}$ — выборка из неизвестного распределения $P_\theta \in \mathcal{P}$. Тогда оценка $\theta_n^*(X)$ называется *состоятельной оценкой* θ, если имеется сходимость:

$$\forall \theta \in \Theta \quad \theta_n^*(X) \xrightarrow[n\to\infty]{P_\theta} \theta$$

Замечание. Определение выше можно расписать так:

$$\forall \theta \in \Theta \ \forall \varepsilon > 0 \quad P_\theta(\|\theta_n^*(X) - \theta\|_2 \geqslant \varepsilon) \to 0,\ n \to \infty$$

Определение 3.6. Пусть $\{X_k\}_{k=1}^{\infty}$ — выборка из неизвестного распределения $P_\theta \in \mathcal{P}$. Тогда оценка $\theta_n^*(X)$ называется *сильно состоятельной оценкой* θ, если имеется сходимость:

$$\forall \theta \in \Theta \quad \theta_n^*(X) \xrightarrow[n\to\infty]{P_\theta \text{ п.н.}} \theta$$

Пример. Для $\mathcal{P} = \{N(\theta, 1)\}$ верно, что $\overline{X}$ — сильно состоятельная оценка (в силу УЗБЧ)

Определение 3.7. Пусть $\{X_k\}_{k=1}^{\infty}$ — выборка из случайных величин с распределением P_θ. Тогда оценка $\theta_n^*(X)$ называется *асимптотически нормальной оценкой* θ, если выполнена сходимость:

$$\forall \theta \in \Theta \quad \sqrt{n}(\theta_n^*(X) - \theta) \xrightarrow[n\to\infty]{d_\theta} N(0, \sigma^2(\theta))$$

где d_θ — сходимость по распределению P_θ. Величину $\sigma^2(\theta)$ называют *асипмтотической дисперсией оценки* θ_n^*

Замечание автора. По сути определение мотивировано одним из видов ЦПТ:

$$\sqrt{n}\left(\frac{S_n}{n} - \mathbb{E}X_1\right) \xrightarrow[n\to\infty]{d} N(0, DX_1)$$

Замечание. Иногда в определении выше требуют, что $\sigma(\theta) > 0$

Пример. Рассмотрим $\mathcal{P} = \{N(\theta, 1)\}$. Тогда $\overline{X}$ является асимптотически нормальной оценкой θ в силу ЦПТ

Замечание. Введённые определения состоятельной, сильно состоятельной и асимптотически нормальной выборки являются *асимптотическими свойствами*, то есть имеют смысл только в том случае, когда число элементов выборки растёт.

Замечание. Мы можем оценивать не только θ, но и $\tau(\theta)$ для некоторой функции τ. Определения меняются соответствующим образом.

Утверждение 3.2. *Если $\tau_n^*(X)$ — асимптотически нормальная оценка для $\tau(\theta)$, то $\tau_n^*(X)$ — состоятельная оценка для $\tau(\theta)$*

Доказательство. Зафиксируем $\theta \in \Theta$. Тогда, в силу определения асимптотической нормальности и леммы Слуцкого, верна сходимость:

$$\frac{1}{\sqrt{n}} \cdot \sqrt{n}(\tau_n^*(X) - \tau(\theta)) \xrightarrow[n\to\infty]{d_\theta} 0$$

Итак, $\tau_n^* - \tau(\theta) \to^{d_\theta} 0$. Так как $\tau(\theta)$ — константа, то имеет место эквивалентная сходимость $\tau_n^* \to^{P_\theta} \tau(\theta)$, что и требовалось доказать. □

Утверждение 3.3. *(Наследование состоятельности и сильно состоятельности) Пусть $\theta_n^*(X)$ — (сильно) состоятельная оценка $\theta \in \Theta \subseteq \mathbb{R}^k$, при этом $\tau\colon \mathbb{R}^k \to \mathbb{R}^s$ — непрерывная на Θ функция. Тогда $\tau(\theta_n^*(X))$ — (сильно) состоятельная оценка $\tau(\theta)$*

Доказательство. Фактически, это просто теорема о наследовании сходимости для случайных векторов. □

Лемма 3.1. *(О наследовании асимптотической нормальности) Если $\widehat{\theta}_n$ — асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}$ с асимптотической дисперсией $\sigma^2(\theta)$, а также $\tau\colon \mathbb{R} \to \mathbb{R}$ — дифференцируемая на Θ функция, тогда $\tau(\widehat{\theta}_n)$ — асимптотически нормальная оценка $\tau(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)(\tau'(\theta))^2$*

Доказательство. Зафиксируем $\theta \in \Theta$. Из условия мы знаем, что

$$\xi_n := \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d_\theta} \xi \sim N(0, \sigma^2(\theta))$$

Воспользуемся дельта-методом для τ. Тогда:

$$\frac{\tau(\theta + \xi_n b_n) - \tau(\theta)}{b_n} \xrightarrow[n\to\infty]{d_\theta} \tau'(\theta) \cdot \xi \sim N(0, \sigma^2(\theta)(\tau'(\theta))^2)$$

Несложно понять, что нам нужно взять $b_n = \frac{1}{\sqrt{n}}$, чтобы метод дал нужную сходимость. □

Утверждение 3.4. *(Многомерный случай леммы о наследовании асимптотической нормальности) Если $\widehat{\theta}_n$ — асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}^k$ с асимптотической матрицей ковариаций $\Sigma(\theta)$, а также $\tau\colon \mathbb{R}^k \to \mathbb{R}^s$ — дифференцируемая на Θ функция, тогда $\tau(\widehat{\theta}_n)$ — асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $J_\tau \Sigma J_\tau^T$, где $J_\tau = J_\tau(\theta)$ — матрица Якоби для τ*

Доказательство. Аналогично одномерному случаю, просто применяем многомерный дельта-метод □

3.3 Методы нахождения оценок

Замечание. Далее, как обычно, $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ — параметрическая модель, причём $\mathcal{P} = \{P_\theta\colon \theta \in \Theta\}$.

Метод подстановки (ОМП)

Определение 3.8. Пусть G — функционал из множества мер над $(\mathcal{X}, \mathfrak{B}(\mathcal{X}))$ в множество параметров Θ, а также выполнено условие:

$$\forall \theta \in \Theta \quad \theta = G(P_\theta)$$

Тогда, *оценкой по методу подстановки (ОМП)* называется оценка $\theta_n^*(X) = G(P_n^*)$, где P_n^* — эмпирическое распределение, построенное по выборке $\{X_k\}_{k=1}^n$

Пример. Рассмотрим $\mathcal{P} = \{Bern(\theta)\}$, $G(P) = \int_{\mathbb{R}} xdP(x)$. В силу определения распределения Бернулли, $\theta = G(P_\theta)$. При этом для эмпирического распределения имеем такую формулу:

$$G(P_n^*) = \int_{\mathbb{R}} xdP_n^*(x) = \frac{\sum_{k=1}^n X_k}{n} = \overline{X}$$

Замечание автора. Говоря простым языком, метод ОМП заключается в нахождении несмещённого функционала для P_θ, а затем мы просто пихаем в него эмпирическое распределение P_n^*.

Метод моментов (ОММ)

Замечание. Пусть $\mathcal{X} = \mathbb{R}^m$, $\Theta \subseteq \mathbb{R}^k$. Рассмотрим выборку $\{X_i\}_{i=1}^n$ из распределения $P \in \mathcal{P}$, а также борелевские функции $\forall i \in \{1, \ldots, k\}$ $g_i \colon \mathbb{R}^m \to \mathbb{R}$ (они ещё называются *пробными функциями*). Тогда далее мы принимаем следующие обозначения:

- $m_i(\theta) := \mathbb{E}_\theta g_i(X_1)$
- $m(\theta) := (m_1(\theta), \ldots, m_k(\theta))^T$
- $\overline{g}(X) := (\overline{g_1(X)}, \ldots \overline{g_k(X)})^T$

При этом мы требуем, что $m_i(\theta)$ конечны.

Определение 3.9. Если существует и единственно решение системы уравнений относительно θ:

$$\begin{cases} m_1(\theta) = \overline{g_1(X)} \\ \vdots \\ m_k(\theta) = \overline{g_k(X)} \end{cases} \Leftrightarrow m(\theta) = \overline{g}(X)$$

то значение $\theta^* = m^{-1}(\overline{g}(X))$ (где m^{-1} понимается покоординатно) называется *оценкой по методу моментов (ОММ)*

Замечание автора. Метод довольно интуитивен после того, как мы обсудили соотношение между $m_i(\theta)$ и $\overline{g_i(X)}$ выше.

Определение 3.10. *Стандартными пробными функциями* называются $g_i(x) = x^i$.

Теорема 3.1. *(о сильно состоятельности ОММ) Пусть выполнены следующие условия на $m \colon \Theta \to m(\Theta)$:*

- *m — биекция*
- *m^{-1} можно доопределить до функции, заданной на $\mathbb{R}^k$*
- *Доопределённая m^{-1} непрерывна в каждой точке $m(\Theta)$*

Тогда оценка по методу моментов θ_n^* является сильно состоятельной оценкой параметра θ.

Доказательство. Зафиксируем $\theta \in \Theta$. По УЗБЧ мы имеем сходимость $\overline{g}(X) \to^{P_\theta \text{ п.н.}} m(\theta)$, а из этого и условия теоремы, по теореме о наследовании сходимости, можем получить такую сходимость:

$$\theta_n^* = m^{-1}(\overline{g}(X)) \xrightarrow{P_\theta \text{ п.н.}} m^{-1}(m(\theta)) = \theta$$

□

Замечание. Несколько поясним, за что отвечает каждое требование к m в теореме:

- ▷ Биективность позволяет заявить, что $m^{-1}(m(\theta)) = \theta$
- ▷ Доопределение нужно из-за $m^{-1}(\overline{g}(X))$
- ▷ Непрерывность на $m(\Theta)$ требуется, чтобы мы могли воспользоваться теоремой о наследовании сходимости

Теорема 3.2. *(о асимптотической нормальности ОММ) Пусть выполнены следующие условия на* $m\colon \Theta \to m(\Theta)$*:*

- ▷ m *— биекция*
- ▷ m^{-1} *можно доопределить до функции, заданной на* $\mathbb{R}^k$
- ▷ *Доопределённая* m^{-1} *дифференцируема в каждой точке* $m(\Theta)$
- ▷ $\forall i \in \{1, \dots, k\}\ \ \mathbb{E}_\theta g_i^2(X_1) < \infty$

Тогда оценка по методу моментов θ_n^* *является асимптотически нормальной оценкой параметра* θ*.*

Доказательство. По Центральной Предельной Теореме:

$$\sqrt{n}(\overline{g}(X) - m(\theta)) \xrightarrow{d_\theta} N(0, \Sigma)$$

где Σ — матрица коваріаций для вектора $\overline{g}(X)$. Осталось применить многомерную лемму о наследовании асимптотической сходимости с функцией m^{-1}. Тогда $m^{-1}(\overline{g}(X)) = \theta_n^*$ является асимптотически нормальной оценкой $m^{-1}(m(\theta)) = \theta$ с матрицей ковариаций $J_m^{-1}\Sigma J_m^{-T}$. □

Замечание. Метод моментов на самом деле является частным случаем метода подстановки, ибо посмотрим на решение системы и реальный вектор θ в развёрнутом виде:

$$\theta^* = m^{-1}\begin{pmatrix} \int_{\mathcal{X}} g_1(x)dP_n^*(x) \\ \vdots \\ \int_{\mathcal{X}} g_k(x)dP_n^*(x) \end{pmatrix} =: G(P_n^*); \quad \theta = m^{-1}\begin{pmatrix} \int_{\mathcal{X}} g_1(x)dP_\theta(x) \\ \vdots \\ \int_{\mathcal{X}} g_k(x)dP_\theta(x) \end{pmatrix} = G(P_\theta)$$

Метод выборочных квантилей

Замечание. Здесь мы считаем, что $\mathcal{X} = \mathbb{R}$

Определение 3.11. Пусть P — распределение на $\mathbb{R}$, F — соответствующая функция распределения и $p \in (0; 1)$. Тогда *p-квантилью распределения* P называется следующая точка z_p:

$$z_p := \inf\{x\colon F(x) \geqslant p\}$$

Замечание. В силу правой непрерывности функции распределения F инфинум достигается. При этом $F^{-1}(p)$ совершенно не обязательно всегда подходит:

- ▷ Если F непрерывна, то существует решение уравнения $F(z_p) = p$, но оно не единственно в силу нестрогого возрастания
- ▷ Если F строго монотонна, то у нас есть гарантия на единственность решения $F(z_p) = p$, однако оно не обязательно существует (инфинум-то найдём, но вот в случае разрыва просто может не быть точки, в которой $F(x) = p$)

Определение 3.12. Пусть $\{X_i\}_{i=1}^n$ — выборка. Статистика $z_{n,p} = X_{(\lceil np \rceil)}$ называется *выборочной p-квантилью*

Замечание. Несложно понять, что выборочный p-квантиль — это просто p-квантиль для эмпирического распределения P_n^*

Теорема 3.3. *(О выборочных квантилях) Пусть $\{X_i\}_{i=1}^n$ — выборка из распределения P с плотностью $f(x)$. Пусть z_p — это p-квантиль этого распределения, причём f непрерывно дифференцируема в некоторой окрестности z_p и $f(z_p) > 0$. Тогда имеет место сходимость:*

$$\sqrt{n}(z_{n,p} - z_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{f^2(z_p)}\right)$$

Напоминание. Пусть $\{X_i\}_{i=1}^n$ — выборка из распределения P с плотностью $f(x)$, а F — функция распределения P, то верны следующие факты про порядковые статистики:

- ▷ $P(X_{(k)} \leqslant x) = \sum_{m=k}^n C_n^m F^m(x)(1 - F(x))^{n-m}$
- ▷ $p_{X_{(k)}} = nC_{n-1}^{k-1} F^{k-1}(x)(1 - F(x))^{n-k} f(x)$

Доказательство. Сходимость по распределению эквивалентна тому, что функции распределения сходятся во всех точках непрерывности своего предела. Мы пронормируем доказываемую сходимость так, чтобы при доказательстве получить справа $N(0, 1)$ (к результату теоремы же вернёмся просто при помощи теоремы о наследовании):

$$\eta_n = \frac{\sqrt{n}(z_{n,p} - z_p)}{\sqrt{\frac{p(1-p)}{f^2(z_p)}}}$$

Итак, обозначим $k = \lceil np \rceil$. Идея состоит в том, чтобы найти плотность p_{η_n} и показать, что она сходится к плотности нормального распределения, а дальше останется показать сходимость интегралов по этим плотностям.

1. (Разбираемся с плотностью) Запишем η_n в более удобном виде:

$$\eta_n = (z_{n,p} - z_n)\sqrt{\frac{nf^2(z_p)}{p(1-p)}}$$

Заметим, что η_n является линейной комбинацией от $X_{(k)}$, чью плотность мы знаем. Стало быть, если $t_n(x) = z_p + \frac{x}{f(z_p)}\sqrt{\frac{p(1-p)}{n}}$ (функция, переводящая x из $F_\eta(x)$ в соответствующий параметр $F_{X_{(k)}}$), то имеет место равенство:

$$p_{\eta_n}(x) = \sqrt{\frac{p(1-p)}{nf^2(z_p)}} \cdot p_{X_{(k)}}(t_n(x))$$

Разобьём p_{η_n} на три части (теперь считаем, что x фиксирован):

- ▷ $p_{\eta_n}(x) = A_1(n) \cdot A_2(n) \cdot A_3(n)$
- ▷ $A_1(n) = \sqrt{npq} C_{n-1}^{k-1} p^{k-1} q^{n-k}$, где $q = 1 - p$
- ▷ $A_2(n) = \frac{f(t_n(x))}{f(z_p)}$
- ▷ $A_3(n) = \left(\frac{F(t_n)}{p}\right)^{k-1} \left(\frac{1-F(t_n)}{q}\right)^{n-k}$

Покажем, что при стремлении $n \to \infty$ каждая из этих частей даст нам сомножитель из плотности $N(0, 1)$:

- ▷ $A_1(n) \to \frac{1}{\sqrt{2\pi}}$ — просто применение формулы Стирлинга
- ▷ $A_2(n) \to 1$. Действительно, ведь f непрерывна, а $\lim_{n\to\infty} t_n(x) = z_p$
- ▷ $A_3(n) \to e^{-x^2/2}$. Для получения этой сходимости, мы рассмотрим отдельно каждый сомножитель и получим какие-то формы для логарифмов от них. Так как это делается аналогично, то опишем только первый. Итак, $F(z_p) = p$, $\lim_{n\to\infty} t_n(x) = z_p$ и F как минимум дважды гладкая. Напишем формулу Тейлора до второй производной в точке z_p:

$$F(t_n) = F(z_p) + (t_n - z_p) \cdot F'(z_p) + \frac{1}{2}(t_n - z_p)^2 \cdot F''(z_p) + o\big((t_n - z_p)^2\big),\ t_n \to z_p$$

  Её можно расписать так:

$$F(t_n) = p + x\sqrt{\frac{p(1-p)}{n}} + \frac{1}{2} \cdot \frac{x^2 pq}{n} \cdot \frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right),\ n \to \infty$$

  Осталось поделить на p разложить логарифм (пользуемся методом $\ln x = \ln(1 + (x - 1)))$):

$$\ln \frac{F(t_n)}{p} = x\sqrt{\frac{q}{pn}} + \frac{1}{2} \cdot x^2 \frac{q}{n} \cdot \frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right) - \frac{x^2}{2} \cdot \frac{q}{pn}$$

  То же самое проделывается и для $\ln \frac{1-F(t_n)}{q}$. В итоге:

$$\ln A_3(n) = (k-1)\ln\frac{F(t_n)}{p} + (n-k)\ln\frac{1-F(t_n)}{q} \xrightarrow[n\to\infty]{} -\frac{x^2}{2}$$

Таким образом, $p_{\eta_n}(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Более того, если посмотреть ход наших рассуждений, то тривиально оказывается, что мы получили равномерную сходимость на любом отрезке $[-N; N]$.

2. Сходимость по распределению эквивалентна сходимости функций распределения в основном:

$$\forall x \in C(F) \quad F_n(x) = \int_{(-\infty;x]} p_{\eta_n}(x) d\mu(x) \xrightarrow[n\to\infty]{} \int_{(-\infty;x]} p(x) d\mu(x)$$

Зафиксируем $x \in C(F)$ и посмотрим модуль разности интегралов:

$$\left|\int_{(-\infty;x]} p(x)d\mu(x) - \int_{(-\infty;x]} p_{\eta_n}(x)d\mu(x)\right| = \left|\int_{(-\infty;x]} (p(x) - p_{\eta_n}(x))d\mu(x)\right|$$

Для фиксированного $\varepsilon > 0$ выберем такой компакт $[-N;N]$, что интегралы по $\mathbb{R}\setminus[-N;N]$ не больше ε. Тогда, либо $x < -N$ и всё тривиально, либо мы можем свести оценку модуля выше к исследованию интеграла на компакте, а там мы имеем равномерную оценку на подынтегральную функцию и, следовательно, можем избавиться от интеграла вообще:

$$\left|\int_{(-\infty;x]} (p(x) - p_{\eta_n}(x))d\mu(x)\right| \leqslant \\ \left|\int_{\mathbb{R}\setminus[-N;N]} (p(x) - p_{\eta_n}(x))d\mu(x)\right| + \left|\int_{[-N;x]} (p(x) - p_{\eta_n}(x))d\mu(x)\right| \leqslant \\ 2\varepsilon + \int_{[-N;x]} |p(x) - p_{\eta_n}(x)|d\mu(x)$$

В силу равномерной оценки, можем найти такой номер n_0, что при $n \geqslant n_0$ верна оценка $|p(x) - p_{\eta_n}(x)| \leqslant \frac{\varepsilon}{x-(-N)}$. Тогда приходим к нужному результату:

$$|F(x) - F_{\eta_n}(x)| \leqslant 2\varepsilon + \int_{[-N;x]} |p(x) - p_{\eta_n}(x)|d\mu(x) \leqslant 2\varepsilon + (x+N)\cdot\frac{\varepsilon}{x+N} = 3\varepsilon$$

□

Замечание. Так как z_p — константа, то при помощи леммы Слуцкого несложно показать, что есть сходимость $z_{n,p} \to^P z_p$.

Определение 3.13. *Медианой распределения* P называется его $\frac{1}{2}$-квантиль.

Определение 3.14. Пусть $\{X_i\}_{i=1}^n$ — выборка. *Выборочной медианой* $\hat{\mu}$ называется следующая оценка:

$$\hat{\mu} = \begin{cases} X_{(k+1)},\ n = 2k+1 \\ \dfrac{X_{(k)} + X_{(k+1)}}{2},\ n = 2k \end{cases} = \frac{X_{(\lfloor\frac{n}{2}\rfloor)} + X_{(\lceil\frac{n}{2}\rceil)}}{2}$$

Теорема 3.4. *(о выборочной медиане) Пусть $\{X_i\}_{i=1}^n$ — выборка из распределения с плотностью $f(x)$. Пусть $z_{1/2}$ — это $\frac{1}{2}$-квантиль этого распределения, причём f непрерывно дифференцируема в некоторой окрестности $z_{1/2}$ и $f(z_{1/2}) > 0$. Тогда выполнена сходимость:*

$$\sqrt{n}(\hat{\mu}(X) - z_{1/2}) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{4f^2(z_{1/2})}\right)$$

Доказательство. Типа тривиально, но не тривиально □

Замечание. Как и в случае теоремы о выборочных квантилях, имеет место сходимость $\hat{\mu} \to^P z_{1/2}$

3.4 Сравнение оценок

Определение 3.15. *Функцией потерь* называется любая борелевская неотрицательная функция $g\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$.

Определение 3.16. Пусть $\Theta \subseteq \mathbb{R}^m$ и g — функция потерь из $\mathbb{R}^m \times \mathbb{R}^m$. Тогда, если θ_n^* — оценка параметра θ, то функция $g(\theta^*(X), \theta)$ называется *величиной потерь*.

Пример. Покажем несколько стандартных функций потерь:

1. $g(x, y) = |x - y|$
2. $g(x, y) = (x - y)^2$ — *квадратичная функция потерь*
3. Пусть $A \in \mathbb{R}^{k\times k}$, $A \succeq 0$ и $\theta \in \mathbb{R}^k$. Тогда $g(x, y) = \langle A(x - y), x - y\rangle$

Определение 3.17. Пусть g — функция потерь из $\Theta \times \Theta$. Тогда *функцией риска оценки* θ^* называется следующий объект:

$$R(\theta^*, \theta) := \mathbb{E}_\theta g(\theta^*, \theta)$$

3.4.1 Равномерный подход к сравнению оценок

Определение 3.18. Пусть $\hat{\theta}(X)$ и $\theta^*(X)$ — оценки параметра θ. Тогда $\hat{\theta}(X)$ *лучше оценки* $\theta^*(X)$ *в равномерном подходе с функцией потерь* g, если выполнены утверждения:

- ▷ $\forall \theta \in \Theta \quad R(\hat{\theta}(X), \theta) \leqslant R(\theta^*(X), \theta)$
- ▷ Для некоторого $\theta_0 \in \Theta$ неравенство выше является строгим

Замечание. Далее мы отождествляем оценки T_1 и T_2, если выполнено утверждение:

$$\forall \theta \in \Theta \quad T_1(X) \overset{P_\theta \text{ п.н.}}{=} T_2(X)$$

Определение 3.19. Оценка $\hat{\theta}$ называется *наилучшей в классе оценок* $\mathcal{K}$, если она лучше любой другой оценки $\theta^* \in \mathcal{K}$.

Замечание. Наилучшая оценка не всегда существует.

Доказать, что в классе всех оценок нет наилучшей

3.4.2 Минимаксный подход

Определение 3.20. Оценка $\theta^*(X)$ называется *наилучшей в минимаксном подходе*, если выполнено равенство:

$$\sup_{\theta\in\Theta} R(\theta^*(X), \theta) = \inf_{\hat{\theta}\in\mathcal{K}} \sup_{\theta\in\Theta} R(\hat{\theta}, \theta)$$

Замечание. Иными словами, $\theta^*(X)$ наилучшая в минимаксном подходе, если она обладает наименьшим максимумом функции риска.

3.4.3 Байесовский подход

Замечание. До этого момента мы искали θ без какого-либо предположения, просто считая, что оно фиксировано. Альтернативный подход состоит в том, чтобы считать θ *случайной величиной с распределением* Q. Тогда задание распределения выборки есть задание условного распределения $P(x|\theta)$.

Замечание автора. А как выглядит вероятностное пространство в таком случае, почему это всё работает? Тройка будет иметь вид $(\Theta \times \mathcal{X}, \mathfrak{B}(\mathfrak{B}(\mathcal{F}) \times \mathfrak{B}(\mathcal{X})), P)$. Самым интересным, пожалуй, здесь будет определение P. А его можно задать следующим образом:

$$P(A \times B) = \int_A P(B|\theta) dQ(\theta)$$

По сути это формула полной вероятности в форме интеграла.

Определение 3.21. Пусть $\hat{\theta}(X)$ — оценка θ, $R(\hat{\theta}, \theta)$ — её функция риска. Тогда *риском оценки* не факт, что лекторское определение назовём следующую величину:

$$R(\hat{\theta}(X)) = \mathbb{E}_Q R(\hat{\theta}(X), \theta) = \int_\Theta R(\hat{\theta}(X), t) dQ(t)$$

Определение 3.22. Оценка $\theta^*(X)$ называется *наилучшей в байесовском подходе*, если выполнено равенство:

$$R(\theta^*(X)) = \min_{\hat{\theta} \in \mathcal{K}} R(\hat{\theta}(X))$$

3.4.4 Асимптотический подход

Определение 3.23. Пусть $\hat{\theta}_1$ и $\hat{\theta}_2$ — асимптотически нормальные оценки параметра θ с асимптотическими дисперсиями $\sigma_1^2(\theta)$ и $\sigma_2^2(\theta)$ соответственно. Тогда *оценка* $\hat{\theta}_1$ *лучше оценки* $\hat{\theta}_2$ *в асимптотическом подходе*, если выполнено утверждение:

$$\forall \theta \in \Theta \quad \sigma_1^2(\theta) \leqslant \sigma_2^2(\theta)$$

Пример. Рассмотрим выборку из $N(\theta, 1)$, а также оценки $\hat{\mu}$ и $\overline{X}$.

- ▷ С одной стороны, по ЦПТ $\sqrt{n}(\overline{X} - \theta) \xrightarrow{d_\theta} N(0, 1)$
- ▷ С другой стороны, по теореме о выборочной медиане $\sqrt{n}(\hat{\mu} - \theta) \xrightarrow{d_\theta} N(0, \pi/2)$

Сразу видно, что оценка средним лучше оценки медианой в асимптотическом подходе для семейства $N(\theta, 1)$.

Определение 3.24. Оценка $\hat{\theta}(X)$ называется *наилучшей в классе оценок* $\mathcal{K}$ *в асимптотически нормальном подходе*, если она лучше любой другой оценки.

Замечание. *Плотностью дискретного распределения* P *из* $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ мы называем плотность этого распределения по считающей мере μ множества $\mathbb{Z}^n$. Понятно, что $p(x) = P(\{x\})$ для любого $x \in \mathbb{Z}^n$ (остальные точки доопределяются, например, нулём)

Пример. Если $\xi \sim Bin(n, p)$, то $P(\xi = k) = C_n^k p^k q^{n-k} \chi\{k \in \{0, \ldots, n\}\} = p_\xi(k)$

Напоминание. Если ξ — дискретная случайная величина с плотностью $p(x)$, то по теореме о вычислении интеграла Лебега для меры с плотностью верна формула:

$$\mathbb{E}g(\xi) = \sum_{k\in\mathbb{Z}} g(k)P(\xi = k) = \int_{\mathbb{R}} g(x)p(x)d\mu(x)$$

Замечание. Всюду далее, когда говорим о плотности распределения, мы считаем, что либо это обычная плотность абсолютно непрерывного распределения, либо плотность дискретного распределения по считающей мере на $\mathbb{Z}^n$.

3.4.5 Среднеквадратический подход

Замечание. Среднеквадратический подход — это особый случай равномерного подхода, когда

- ▷ $\mathcal{K}$ — класс несмещённых оценок для $\tau(\theta)$
- ▷ g — квадратичная функция потерь

Утверждение 3.5. *Если $T_1, T_2 \in \mathcal{K}$, причём обе оценки не хуже любых из того же класса $\mathcal{K}$, то они эквивалентны в определённом ранее смысле:*

$$\forall\theta \in \Theta \;\; T_1(X) =^{P_\theta \text{ п.н.}} T_2(X)$$

Доказательство. Если оценки являются не хуже любых из того же класса, то выполнены равенства:

$$E_\theta(T_i - \tau(\theta))^2 = \inf_{T\in\mathcal{K}} \mathbb{E}_\theta(T - \tau(\theta))^2$$

Более того, мы рассматриваем класс несмещённых оценок. Стало быть:

$$\mathbb{E}_\theta(T - \tau(\theta))^2 = \mathbb{E}_\theta T^2 - 2\tau(\theta)\mathbb{E}_\theta T + \tau^2(\theta) = \mathbb{E}_\theta T^2 - \tau^2(\theta)$$

Таким образом, для T_1 и T_2 мы имеем равенство $\mathbb{E}_\theta T_1^2 = \mathbb{E}_\theta T_2^2$. Воспользуемся следующим алгебраическим свойством:

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

Положим $a = T_1$ и $b = T_2$ и обозначим $T_3 = \frac{T_1+T_2}{2}$. Если взять матожидание с обеих сторон равенства, то получим:

$$\mathbb{E}_\theta T_3^2 + \frac{1}{4}\mathbb{E}_\theta(T_1 - T_2)^2 = \mathbb{E}_\theta T_1^2$$

Если вычесть с каждой стороны по $\tau^2(\theta)$, то придём к функциям риска T_3 и T_1. Так как T_1 не хуже любой оценки из класса $\mathcal{K}$ и $T_3 \in \mathcal{K}$, то верно неравенство:

$$\frac{1}{4}\mathbb{E}_\theta(T_1 - T_2)^2 = \mathbb{E}_\theta T_1^2 - \mathbb{E}_\theta T_3^2 = \mathbb{E}_\theta(T_3 - \tau(\theta))^2 - \mathbb{E}_\theta(T_1 - \tau(\theta))^2 \leqslant 0$$

Так как слева написана неотрицательная величина, то $\mathbb{E}_\theta(T_1 - T_2)^2 = 0$ абсолютно точно. А такое возможно тогда и только тогда, когда $(T_1 - T_2)^2 =^{P_\theta \text{ п.н.}} 0$, что эквивалентно $T_1 =^{P_\theta \text{ п.н.}} T_2$ □

Определение 3.25. Если семейству распределений $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ соответствует семейство плотностей $\{p_\theta, \theta \in \Theta\}$ по одной и той же мере μ, то говорят, что $\mathcal{P}$ *доминируемо относительно* μ

Замечание. Далее мы считаем, что X — наблюдение с распределением из $\mathcal{P}$, причём семейство доминируемо относительно μ, и рассматриваем $\Theta \subseteq \mathbb{R}$

Замечание. Выборка (как вектор или кортеж) тоже является наблюдением. Её распределение согласовано с распределениями каждой компоненты, а отсюда плотность является тензорным произведением плотностей компонент.

Определение 3.26. Пусть X — наблюдение. Случайная величина $U_\theta(X) = \frac{\partial}{\partial\theta} \ln p_\theta(X)$ называется *вкладом наблюдения* X

Определение 3.27. Пусть X — наблюдение. Функция $I_X(\theta) = \mathbb{E}_\theta U_\theta^2(X)$ называется *количеством информации о параметра θ, содержащемся в X (информация по Фишеру)*

Утверждение 3.6. *Если $X = (X_1, \ldots, X_n)$ — выборка из n наблюдений, то, при обозначении информации Фишера одного наблюдения $i(\theta)$ верно равенство $I_X(\theta) = ni(\theta)$.*

Доказательство. Как уже было упомянуто ранее, верно равенство:

$$p_\theta(x) = p_\theta(x_1, \ldots, x_n) = p_{\theta,1}(x_1) \cdot \ldots \cdot p_{\theta,n}(x_n)$$

Стало быть, $U_\theta(x) = \frac{\partial}{\partial\theta}\left(\sum_{i=1}^n \ln p_{\theta,i}(x_i)\right) = \sum_{i=1}^n U_{\theta,i}(x_i)$. С учётом того, что X_i независимы, получаем равенство:

$$I_X(\theta) = \mathbb{E}_\theta\left(\sum_{i=1}^n U_{\theta,i}(X_i)\right)^2 = \sum_{i=1}^n \mathbb{E}_\theta U_{\theta,i}^2(X_i) = ni(\theta)$$

□

Определение 3.28. *Условиями регулярности* для вероятностно-статистической модели $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$ и наблюдения X называют следующие требования:

1. $\Theta \subseteq \mathbb{R}$ — открытый интервал (может быть бесконечным)
2. $A = \{x \in \mathcal{X} \colon p_\theta(x) > 0\}$ не зависит от θ. Это множество называют *носителем*
3. Для любой статистики $S(X)$ с условием $\forall\theta \in \Theta\ \mathbb{E}_\theta S^2(X) < \infty$ выполнено равенство:

   $$\frac{\partial}{\partial\theta}\mathbb{E}_\theta S(X) = \frac{\partial}{\partial\theta}\int_A S(x)p_\theta(x)d\mu(x) = \int_A S(x)\frac{\partial}{\partial\theta}p_\theta(x)d\mu(x) = \mathbb{E}_\theta\left(S(X)\frac{\partial}{\partial\theta}\ln p_\theta(X)\right)$$

   В частности требуем, чтобы для любого $\theta \in \Theta$ $\frac{\partial}{\partial\theta}\ln p_\theta(x)$ существовало на A и было конечно
4. $\forall\theta \in \Theta\quad 0 < I_X(\theta) < \infty$

Теорема 3.5. *(Неравенство Рао-Крамéра) Пусть выполнены условия регулярности, а также оценка $\hat{\theta}(X) \in \mathcal{K}$ обладает условием $\forall\theta \in \Theta\ \mathbb{E}_\theta\hat{\theta}^2(X) < \infty$. Тогда верно следующее неравенство:*

$$D_\theta\hat{\theta}(X) \geqslant \frac{(\tau'(\theta))^2}{I_X(\theta)}$$

Доказательство. Подставим $S(X) = 1$ в третье условие регулярности:

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta S(X) = \frac{\partial}{\partial\theta}1 = 0 = \mathbb{E}_\theta\frac{\partial}{\partial\theta}\ln p_\theta(X) = \mathbb{E}_\theta U_\theta(X)$$

Получили $E_\theta U_\theta(X) = 0$. Теперь, в то же условие подставим $\hat{\theta}(X)$:

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta\hat{\theta}(X) = \frac{\partial}{\partial\theta}\tau(\theta) = \tau'(\theta) = \mathbb{E}_\theta\hat{\theta}(X)U_\theta(X)$$

Умножим равенство $\mathbb{E}_\theta U_\theta(X) = 0$ на $\tau(\theta)$ и вычтем его из последнего равенства. Получим следующее:

$$\tau'(\theta) = \mathbb{E}_\theta(\hat{\theta}(X) - \tau(\theta))U_\theta(X)$$

Остаётся воспользоваться интегральным неравенством КБШ:

$$(\tau'(\theta))^2 \leqslant \left(\mathbb{E}_\theta(\hat{\theta} - \tau(\theta))^2\right)\left(\mathbb{E}_\theta U_\theta^2(X)\right) = D_\theta\hat{\theta}(X)\cdot I_X(\theta)$$

□

Следствие. Если $\tau(\theta) = \theta$, то в условиях теоремы имеем неравенство $D_\theta\hat{\theta}(X) \geqslant \frac{1}{I_X(\theta)}$

Замечание автора. Далее $\widehat{\mathcal{K}}$ — это подмножество класса $\mathcal{K}$, у оценок которого существует конечная дисперсия (то есть выполнено условие $\forall\theta\in\Theta\ \mathbb{E}_\theta\hat{\theta}^2(X) < \infty$)

Определение 3.29. Если в неравенстве Рао-Крамера для оценки $\hat{\theta}(X) \in \widehat{\mathcal{K}}$ достигается равенство, то оценка $\hat{\theta}(X)$ называется *эффективной оценкой* $\tau(\theta)$.

Теорема 3.6. *(Критерий эффективности) Пусть выполнены условия регулярности. Тогда оценка $\hat{\theta}(X)$ принадлежит классу $\widehat{\mathcal{K}}$ и является эффективной оценкой $\tau(\theta)$ тогда и только тогда, когда выполнено равенство*

$$\hat{\theta}(X) - \tau(\theta) = c(\theta)\cdot U_\theta(X)$$

где $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$.

Доказательство.

⇒ Если $\hat{\theta}$ — эффективная оценка, то в доказательстве неравенства Рао-Крамера у нас достигается равенство в КБШ. Это происходит тогда и только тогда, когда $\eta = \hat{\theta}(X) - \tau(\theta)$ и $\xi = U_\theta(X)$ являются линейно зависимыми случайными величинами, то есть $\alpha(\theta)+\beta(\theta)\xi+\gamma(\theta)\eta =^{P_\theta\text{ п.н.}} 0$. Заметим, что $\mathbb{E}_\theta\xi = \mathbb{E}_\theta\eta = 0$. Стало быть, если применить матожидание к имеющемуся равенству, то $\alpha(\theta) =^{P_\theta\text{ п.н.}} 0$. Этот факт также позволяет заявить, что $\gamma(\theta) \neq^{P_\theta\text{ п.н.}} 0$ (иначе линейная комбинация тривиальна). Таким образом:

$$\eta =^{P_\theta\text{ п.н.}} -\frac{\beta(\theta)}{\gamma(\theta)}\xi;\quad \hat{\theta}(X) - \tau(\theta) = c(\theta)\cdot U_\theta(X)$$

⇐ Выразим оценку из равенства:

$$\hat{\theta}(X) = \tau(\theta) + c(\theta)\cdot U_\theta(X)$$

Так как $E_\theta U_\theta(X) = 0$, то $\hat{\theta} \in \mathcal{K}$. Домножим исходное равенство на $U_\theta(X)$ и возьмём матожидание:

$$\mathbb{E}_\theta(\hat{\theta} - \tau(\theta))U_\theta(X) = c(\theta)\mathbb{E}_\theta U_\theta^2(X) = c(\theta)I_X(\theta)$$

В силу регулярности $0 < I_X(\theta) < +\infty$. Таким образом, $\hat{\theta} \in \widehat{\mathcal{K}}$. При этом, из доказательства неравенства Рао-Крамера мы знаем, что левая часть равна $\tau'(\theta)$. Отсюда выражение для $c(\theta)$. В силу линейной зависимости $\eta = \hat{\theta} - \tau(\theta)$ и $\xi = U_\theta(X)$, $\hat{\theta}$ является эффективной оценкой.

□

Следствие. Если $\theta^*(X) \in \widehat{\mathcal{K}}$ не хуже эффективной оценки $\hat{\theta} \in \widehat{\mathcal{K}}$, то по критерию эффективности $\theta^* =^{P_\theta \text{ п.н.}} \hat{\theta}$.

Замечание. Если есть эффективная оценка $\tau(\theta)$, то она наилучшая оценка $\tau(\theta)$ в классе $\widehat{\mathcal{K}}$. Обратное, при этом, неверно.

Теорема 3.7. *Если в условиях регулярности существует эффективная оценка для $\tau(\theta)$, $\tau \neq const$, то множество функций, для которых существует эффективная оценка, может быть выражено как $\{a\tau(\theta) + b\colon a, b \in \mathbb{R}\}$.*

Замечание автора. Теорема работает так же, как и неопределённый интеграл: для выражения множества всех первообразных достаточно знать одну.

Утверждение 3.7. *Пусть $A = \{x \in \mathcal{X}\colon p_\theta(x) > 0\}$ не зависит от θ. Тогда имеет место эквивалентность:*

$$\forall B \in \mathfrak{B}(\mathcal{X}) \quad \Big(\exists \theta_0 \in \Theta \ P_{\theta_0}(X \in B) = 1\Big) \Leftrightarrow \Big(\forall \theta \in \Theta \ P_\theta(X \in B) = 1\Big)$$

Доказательство. Доказывать утверждение нужно только в одну сторону, ибо в другую очевидно. Итак, $P_{\theta_0}(X \in B) = 1$. Вспомним, что всегда выполнено равенство:

$$\forall C, D \in \mathfrak{B}(\mathcal{X}) \ \ P_\theta(C) + P_\theta(D) = P_\theta(C \cup D) + P_\theta(C \cap D)$$

При всех θ верно, что $P_\theta(X \in A) = P_\theta(X \in A \cup B) = 1$. Стало быть, $P_\theta(X \in B) = P_\theta(X \in A \cap B)$. Появится ближе к сессии. Идейно надо от противного □

Доказательство. Доказывать равенство множеств будем двумя вложениями:

$\subseteq$ Пусть $\hat{\tau}(X)$ и $\hat{v}(X)$ — эффективные оценки для $\tau(\theta)$ и $v(\theta)$ соответственно. По критерию эффективности мы знаем, что:

$$\forall \theta \in \Theta \ \begin{cases} \hat{\tau}(X) =^{P_\theta \text{ п.н.}} \tau(\theta) + c(\theta)U_\theta(X) \\ \hat{v}(X) =^{P_\theta \text{ п.н.}} v(\theta) + d(\theta)U_\theta(X) \end{cases}$$

В условиях регулярности Θ является интервалом, причём мы знаем, что $\tau \neq const$. Стало быть, существует θ_0 такая, что $\tau'(\theta_0) \neq 0$. Тогда и $c(\theta_0) \neq 0$, а потому при θ_0 из первого равенства можно выразить $U_\theta(X)$ и подставить во второе:

$$\hat{v}(X) = v(\theta_0) + d(\theta_0)\left(\frac{\hat{\tau}(X) - \tau(\theta_0)}{c(\theta_0)}\right) = a(\theta_0)\hat{\tau}(X) + b(\theta_0)$$

$\supseteq$ Итак, $\hat{\tau}(X)$ — эффективная оценка для $\tau(\theta)$. Нужно проверить, что и $\hat{v}(X) = a\hat{\tau}(X)+b$ является эффективной оценкой для $v(\theta) = a\tau(\theta) + b$. Простыми манипуляциями преобразуем уже имеющийся критерий эффективности для $\hat{\tau}(X)$ так, чтобы он соответствовал $\hat{v}(X)$:

$$\hat{\tau}(X) = \tau(\theta) + c(\theta)U_\theta(X) \Rightarrow (a\tau(\hat{X}) + b) = (a\tau(\theta) + b) + (ac(\theta))U_\theta(X)$$

Получили верный критерий эффективности для $\hat{v}(X)$, что завершает доказательство.

$\square$

Определение 3.30. *Экспоненциальным семейством распределений* называют все распределения, чья обобщённая плотность имеет следующий вид:

$$p_\theta(x) = h(x)\exp\left(\sum_{i=1}^{k} a_i(\theta)T_i(x) + V(\theta)\right)$$

где $a_0(\theta) \equiv 1$, а оставшиеся $a_1(\theta), \ldots, a_k(\theta)$ являются линейно независимой системой на Θ.

Пример. Рассмотрим распределение $\Gamma(\alpha, \beta)$. Оно принадлежит к экспоненциальному семейству:

$$p_{\alpha,\beta}(x) = \frac{\alpha^\beta x^{\beta-1}}{\Gamma(\beta)}e^{-\alpha x}\chi\{x > 0\} = \frac{1}{x}\chi\{x > 0\} \cdot \exp\left(\beta \ln x - \alpha x + \ln\frac{\alpha\beta}{\Gamma(\beta)}\right)$$

Связь экспоненциальных семейств и условия существования эффективной оценки

Замечание. Хорошо, вот мы узнали неравенство Рао-Крамера, ввели эффективные оценки, выяснили критерий эффективности и даже нашли вид множества всех эффективных оценок, если знаем хотя бы одну. Закономерный вопрос: «А как по распределению с условиями регулярности понять, что в нём найдутся эффективные оценки вообще?» На это помогает ответить уже введённое *экспоненциальное семейство распределений*.

Замечание. В этом параграфе мы предполагаем, что выполнены условия регулярности.

Теорема 3.8. *Пусть $\{X_i\}_{i=1}^n$ — выборка из распределения $P_\theta \in \mathcal{P}$. Тогда для этого распределения существует $\tau(\theta) \neq const$ и соответствующая эффективная оценка $\hat{\tau}(X)$ тогда и только тогда, когда распределение относится к экспоненциальному семейству.*

Доказательство.

$\Leftarrow$ Итак, пусть f_θ — функция плотности наблюдения $X = (X_1, \ldots, X_n)$. Тогда её можно расписать так (для простоты рассматриваем случай, когда в экспоненте сумма из одного слагаемого):

$$f_\theta(x) = \prod_{i=1}^{n} p_\theta(x_i),\ p_\theta(x_i) = h(x_i)\exp(a(\theta)T(x_i) + V(\theta))$$

Стало быть, у нас есть корректно определённый вклад наблюдения:

$$U_\theta(X) = \frac{\partial}{\partial\theta}\ln f_\theta(X) = \frac{\partial}{\partial\theta}\left(\sum_{i=1}^n \ln(h(X_i)) + a(\theta)T(X_i) + V(\theta)\right) =$$
$$\frac{\partial}{\partial\theta}\left(a(\theta)\sum_{i=1}^n T(X_i) + nV(\theta)\right) = a'(\theta)\sum_{i=1}^n T(X_i) + nV'(\theta)$$

Если $T = const$, то $p_\theta(x) = h(x)e^{b(\theta)}$, а тогда $\int_{\mathbb{R}} p_\theta(x)d\mu(x) = 1$, то есть $b(\theta) = const$ и $p_\theta(x)$ не зависит от θ. Такие случаи мы не рассматриваем. Мы также считаем, что $a'(\theta) \neq 0$. Тогда мы можем переписать равенство в стиле критерия эффективности:

$$\frac{1}{na'(\theta)}U_\theta(X) = \frac{1}{n}\sum_{i=1}^n T(X_i) - \frac{-V'(\theta)}{a'(\theta)}$$

Итак, $\hat{\tau}(X) = \overline{T(X)}$ является эффективной оценкой для $\tau(\theta) = \frac{-V'(\theta)}{a'(\theta)}$ в случае, если это отношение не стало константой.

$\Rightarrow$ Пусть $\hat{\tau}(X)$ — эффективная оценка некоторой $\tau(\theta) \neq 0$. Потребуем, что $\forall\theta \in \Theta\ \tau'(\theta) \neq 0$. Так как оценка эффективна, то выполнено равенство Рао-Крамера:

$$D_\theta\hat{\tau}(X) = \frac{(\tau'(\theta))^2}{I_X(\theta)} < \infty$$

Отсюда автоматически следует, что $\hat{\tau}(X) \in L_2$. За счёт этого мы можем воспользоваться критерием эффективности:

$$\hat{\tau}(X) - \tau(\theta) =^{P_\theta \text{ п.н.}} c(\theta)U_\theta(X)$$

За счёт того, что $\tau'(\theta) \neq 0$, мы также имеем $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)} \neq 0$. Выразим $U_\theta(X)$ и подставим его в своё определение:

$$\frac{\partial}{\partial\theta}\ln f_\theta(X) =^{P_\theta \text{ п.н.}} \frac{\hat{\tau}(X) - \tau(\theta)}{c(\theta)}$$

От случайных величин перейдём к их значениям и, в предположении корректности операции, проинтегрируем равенство. Тогда:

$$\ln f_\theta(x) = \int \frac{\hat{\tau}(x) - \tau(\theta)}{c(\theta)}d\theta + g(x)$$
$$\prod_{i=1}^n p_\theta(x_i) = f_\theta(x) = \exp(\beta(\theta)T(x) + D(\theta) + g(x)) = H(x)\exp(\beta(\theta)T(x) + D(\theta))$$

Теперь, если мы зафиксируем $x_{2,0}, \ldots, x_{n,0} \in A$ (где A из условий регулярности), а x_1 оставим переменной, то можно получить формулу плотности одной случайной величины:

$$p_\theta(x_1) = \frac{H(x_1, x_{2,0}, \ldots, x_{n,0})}{\prod_{i=2}^n p_\theta(x_{i,0})}\exp(\beta(\theta)T(x_1, x_{2,0}, \ldots, x_{n,0}) + D(\theta))$$

Нужно ещё что-то сказать про независимость a_1 с 1. Дописать □

3.5 Достаточные статистики

Замечание. Далее мы находимся в вероятностно-статистическом пространстве $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P_\theta \in \mathcal{P}$.

Определение 3.31. Статистика $T(X)$ называется *достаточной для параметра* θ, если выполнено условие:

$$\forall t\ \forall B \in \mathcal{F}\quad P_\theta(X \in B | T(X) = t) \text{ — не зависит от } \theta$$

Замечание. Если существует биекция между статистиками S и T, причём T достаточная, то и S тоже достаточная. Таким образом, важна не сама статистика, а порождённое ей разбиение вероятностного пространства.

А определение функции правдоподобия кто давать будет? Видимо я

Теорема 3.9. *(Нейман, Фишер. Критерий факторизации) Пусть $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ — доминируемое семейство. Тогда статистика T является достаточной для параметра θ тогда и только тогда, когда функция правдоподобия $f_\theta(x)$ представима в следующем виде:*

$$f_\theta(x) = \psi(T(x), \theta) h(x)$$

где обе функции ψ, h неотрицательны, а также $\psi(t, \theta)$ измерима по t, h измерима по x

Замечание. Разложение функции правдоподобия неоднозначно, коль скоро можно h поделить на константу, а ψ на неё же домножить.

Доказательство. Проведём доказательство только в дискретном случае.

$\Rightarrow$ $T(X)$ — достаточная статистика. Тогда можно записать цепочку равенств:

$$f_\theta(x) = P_\theta(X = x) = P_\theta(X = x \wedge T(X) = T(x)) = \\ \underbrace{P_\theta(T(X) = T(x))}_{\psi(T(x), \theta)} \cdot \underbrace{P_\theta(X = x | T(X) = T(x))}_{h(x)}$$

Определение h корректно в силу достаточности $T(X)$.

$\Leftarrow$ Итак, $f_\theta(x) = \psi(T(x), \theta) h(x)$. В силу дискретности, мы можем заявить следующее:

$$f_\theta(x) = P_\theta(X = x) = \psi(T(x), \theta) h(x) \Longrightarrow P_\theta(X = x | T(X) = t) = \frac{P_\theta(X = x \wedge T(X) = t)}{P_\theta(T(X) = t)}$$

С учётом этого, распишем условную вероятность по определению и покажем явно, что зависимости от θ нет:

$$P_\theta(X = x | T(X) = t) = \frac{P_\theta(X = x \wedge T(X) = t)}{P_\theta(T(X) = t)} = \\ \begin{cases} 0, \ T(x) \neq t \\ \dfrac{P_\theta(X = x)}{P_\theta(T(X) = t)} = \dfrac{\psi(t, \theta) h(x)}{\sum_{y \colon T(y) = t} \psi(t, \theta) h(y)} = \dfrac{h(x)}{\sum_{y \colon T(y) = t} h(y)}, \ T(x) = t \end{cases}$$

В итоговых выражениях системы нет θ, что и требовалось.

□

3.5.1 Улучшение оценок с помощью достаточных оценок

Теорема 3.10. *(Колмогоров-Блэквелл-Рао, об улучшении несмещённой оценки) Пусть $T(X)$ — достаточная статистика для θ. Также пусть $d(X)$ — несмещённая оценка для $\tau(\theta)$. Положим $\varphi(T(X)) = \mathbb{E}_\theta(d(X)|T(X))$. Тогда:*

1. *$\varphi(T(X))$ не зависит от θ, то есть является тоже статистикой, причём $\mathbb{E}_\theta\varphi(T(X)) = \tau(\theta)$ и $D_\theta(\varphi(T(X))) \leqslant D_\theta(d(X))$.*
2. *Если дополнительно $\mathbb{E}_\theta d^2(X) < \infty$, то неравенство дисперсий обрщается в равенство тогда и только тогда, когда $\varphi(T(X)) =^{P_\theta \text{ п.н.}} d(X)$ при любом $\theta \in \Theta$.*

Замечание. Во втором пункте равенство почти всюду можно заменить на эквивалентное требование, что $d(X)$ является $T(X)$-измеримой (такой факт был в курсе теории вероятностей).

Лемма 3.2. *Пусть η, ξ — случайные величины в вероятностном пространстве $(\Omega, \mathcal{F}, P)$. Тогда верно два утверждения:*

1. *Если $\eta \in L_1$, то $\mathbb{E}(\mathbb{E}(\eta|\xi) - \mathbb{E}\eta)^2 \leqslant D\eta$*
2. *Если $\eta \in L_2$, то равенство в неравенстве выше достигается тогда и только тогда, когда $\eta =^{P \text{ п.н.}} \mathbb{E}(\eta|\xi)$*

Замечание автора. Отметим, что дисперсия случайной величины $\eta \in L_1$ существует всегда, ибо $D\eta = \mathbb{E}\eta^2 - (\mathbb{E}\eta)^2$, а интеграл Лебега для неотрицательной функции существует точно, просто может быть равен бесконечности.

Доказательство. Обозначим $\zeta = \mathbb{E}(\eta|\xi)$ и распишем дисперсию:

$$D\eta = \mathbb{E}(\eta - \mathbb{E}\eta)^2 = \mathbb{E}(\eta - \zeta + \zeta - \mathbb{E}\eta)^2 = \mathbb{E}(\eta - \zeta)^2 + \mathbb{E}(\zeta - \mathbb{E}\eta)^2 + 2\mathbb{E}(\eta - \zeta)(\zeta - \mathbb{E}\eta)$$

1. Первое слагаемое мы без проблем можем убрать в неравенстве, а вот с последним непонятно. Так как это константа, то ничто не мешает изучить её, скажем, в условном матожидании:

$$\mathbb{E}\Big(\mathbb{E}\big((\eta - \zeta)(\zeta - \mathbb{E}\eta)|\xi\big)\Big) = \mathbb{E}\Big((\zeta - \mathbb{E}\eta)\mathbb{E}(\eta - \zeta|\xi)\Big)$$

   Вынесение сомножителя законно, ибо $\zeta - \mathbb{E}\eta$ является ξ-измеримой случайной величиной. При этом $\zeta = \mathbb{E}(\eta|\xi)$ по определению, а стало быть оставшееся условное матожидание равно нулю, то есть и внешнее матожидание равно нулю, и $\mathbb{E}(\eta - \zeta)(\zeta - \mathbb{E}\xi) = 0$.

2. Так как $\eta \in L_2$, то $\mathbb{E}\eta^2 < \infty$. Аналогичное утверждение верно и про ζ (неравенство ниже написано за счёт использования неравенства Йенсена)

$$\zeta^2 = \big(\mathbb{E}(\eta|\xi)\big)^2 \leqslant \mathbb{E}(\eta^2|\xi) \Longrightarrow \mathbb{E}\zeta^2 \leqslant \mathbb{E}\eta^2 < \infty$$

   Осталось показать, что в равенстве с $D\eta$ выше первое и последнее слагаемые убираются тогда и только тогда, когда $\eta =^{P \text{ п.н.}} \zeta$, а это тривиально, ибо последнее также эквивалентно $\mathbb{E}(\xi - \eta)^2 = 0$.

□

Доказательство. (теоремы об улучшении несмещённой оценки) Если зафиксировать значение $T(X) = t$, то распределение X не зависит от θ, это гарантируется достаточностью оценки $T(X)$. Стало быть, то же самое верно и по $d(X)$, а значит $\varphi(T(X)) = \mathbb{E}_\theta(d(X)|T(X))$ является $T(X)$-измеримой, не зависящей от θ функцией, что напрямую говорит о том, что $\varphi(T(X))$ является статистикой. Так как $d(X)$ — несмещённая оценка, то $\varphi(T(X))$ наследует этот факт (свойство УМО). Остаётся применить лемму, с учётом которой мы имеем следующее:

$$D_\theta\varphi(T(X)) = \mathbb{E}_\theta\big(\varphi(T(X)) - \mathbb{E}_\theta\varphi(T(X))\big)^2 = \\ \mathbb{E}_\theta\big(E_\theta(d(X)|T(X)) - \mathbb{E}_\theta d(X)\big)^2 \leqslant [\text{применение леммы}] \leqslant D_\theta d(X)$$

При этом, если $d(X) \in L_2$, то неравенство становится равенством тогда и только тогда, когда $d(X) =^{P_\theta \text{ п.н.}} \varphi(T(X))$ □

Определение 3.32. Наилучшая оценка для $\tau(\theta)$ в классе $\mathcal{K}$ называется *оптимальной оценкой.*

Определение 3.33. Статистика $S(X)$ называется *полной для параметра* θ, если верна импликация:

$$\forall f \ \ \Big(\forall\theta \in \Theta \ \ \mathbb{E}_\theta f(S(X)) = 0\Big) \Rightarrow \Big(\forall\theta \in \Theta \ \ f(S(X)) =^{P_\theta \text{ п.н.}} 0\Big)$$

Теорема 3.11. *(Лемана-Шеффера об оптимальной оценке) Пусть T — полная достаточная статистика для параметра θ, $d(X)$ — несмещённая оценка $\tau(\theta)$. Тогда:*

1. *$\varphi(T(X)) = \mathbb{E}_\theta(d(X)|T(X))$ — несмещённая оценка с равномерно минимальной дисперсией для $\tau(\theta)$*

2. *Если дополнительно $D_\theta\varphi(T(X)) < \infty$, то $\varphi(T(X))$ является оптимальной оценкой*

Доказательство.

1. По теореме Колмогорова-Блэквелла-Рао мы уже знаем, что $\varphi(T(X))$ — несмещённая оценка, которая как минимум не хуже $d(X)$ по дисперсии. Пусть $\widetilde{d}(X)$ — тоже несмещённая оценка. По той же теореме мы можем построить $\widetilde{\varphi}(T(X)) = \mathbb{E}(\widetilde{d}(X)|T(X))$ с аналогичными свойствами. Теперь осталось заметить, что для $h = \varphi - \widetilde{\varphi}$ матожидание $h(T(X))$ нулевое:

$$\forall\theta \in \Theta \ \ \mathbb{E}_\theta h(T(X)) = \mathbb{E}_\theta(\varphi(T(X)) - \widetilde{\varphi}(T(X))) = \tau(\theta) - \tau(\theta) = 0$$

   Так как $T(X)$ — полная статистика, то это автоматически означает $\varphi(T(X)) =^{P_\theta \text{ п.н.}} \widetilde{\varphi}(T(X))$. Отсюда следует, что для любой несмещённой оценки мы умеем строить её улучшение, которое точно не лучше $\varphi(T(X))$. Это означает, что у $\varphi(T(X))$ равномерно минимальная дисперсия среди всего класса несмещённых оценок.

2. Теперь у нас ещё есть условие, что $D_\theta\varphi(T(X)) < \infty$. Если теперь какая-то оценка $\widetilde{d}(X)$ имеет ровно ту же дисперсию при всех $\theta \in \Theta$, то по теореме Колмогорова-Блэквелла-Рао это возможно тогда и только тогда, когда $\widetilde{d}(X) =^{P_\theta \text{ п.н.}} \varphi(T(X))$. С учётом предыдущего результата, оценка $\varphi(T(X))$ становится оптимальной.

$\square$

Следствие. Если $T(X)$ — полная достаточная оценка для параметра θ, а $\varphi(T(X))$ — несмещённая оценка $\tau(\theta)$, то

1. $\varphi(T(X))$ не хуже остальных оценок класса $\mathcal{K}$
2. Если $\varphi(T(X)) \in L_2$, то $\varphi(T(X))$ является оптимальной оценкой $\tau(\theta)$

Доказательство. В рамках доказанной теоремы, положим $d(X) = \varphi(T(X))$. Но тогда $\psi(T(X)) = \mathbb{E}(\varphi(T(X))|T(X)) = \varphi(T(X))$, коль скоро $\varphi(T(X))$ является $T(X)$-измеримой, а значит все свойства далее доказаны уже по теореме. $\square$

Теорема 3.12. *(об экспоненциальном семействе) Пусть $\{X_i\}_{i=1}^n$ — выборка из экспоненциального семейства с плотностью общего вида:*

$$p_\theta(x) = h(x)\exp\left(\sum_{i=1}^k a_i(\theta)T_i(x) + V(\theta)\right)$$

Тогда, если область значений векторной функции $\vec{a}(\theta) = (a_1(\theta), \ldots, a_k(\theta))^T, \theta \in \Theta$ содержит k-мерный параллелепипед, то внезапно, вектор стал оценкой скаляра

Замечание. Мы доказали достаточно много хороших фактов про наличие и выражение оптимальной статистики. Из них можно получить некоторый общий алгоритм поиск оптимальной статистики:

1. Найти достаточную статистику T
2. Проверить эту статистику на полноту
3. Если всё верно, то нужно решить уравнение вида $\mathbb{E}_\theta g(T(X)) = \tau(\theta)$, которое ещё называют *уравнением несмещённости*

3.6 Доверительные интервалы

3.6.1 Основные определения

Замечание. Далее мы находимся в вероятностно-статистическом пространстве $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$, $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ и рассматриваем наблюдение X с неизвестным распределением $P \in \mathcal{P}$.

Определение 3.34. Пара статистик $(T_1(X), T_2(X))$ называется *доверительным интервалом уровня доверия γ для параметра θ*, если выполнено неравенство:

$$\forall \theta \in \Theta \quad P(T_1(X) < \theta < T_2(X)) \geqslant \gamma$$

Причём, если при всех $\theta \in \Theta$ достигается равенство, то доверительный интервал называется *точным*.

Замечание. На практике интересуют уровни $\gamma \in \{0.9, 0.95, 0.99\}$. Также иногда удобно использовать односторонние доверительные интервалы вида $(-\infty, T_2(X))$ или $(T_1, +\infty)$.

Замечание. В многомерном случае $\theta \in \Theta \subseteq \mathbb{R}^k$ можно определить аналогичное понятие доверительного интервала для θ_i или скалярной функции $\tau(\theta)$.

Определение 3.35. Множество $S(X) \subseteq \Theta$ называется *доверительным множеством уровня доверия γ для параметра θ*, если выполнено неравенство:

$$\forall \theta \in \Theta \ \ P_\theta(\theta \in S(X)) \geqslant \gamma$$

Замечание автора. Да, множество зависит от выборки.

3.6.2 Метод центральных статистик

Определение 3.36. Пусть одномерная функция $G(x, \theta)$ такова, что распределение $G(X, \theta)$ не зависит от параметра θ (другими словами, при любом фиксированном $\theta \in \Theta$ распределение статистики $G(X, \theta)$ остаётся неизменным). Тогда $G(X, \theta)$ называется *центральной статистикой*.

Замечание автора. Название тут несколько противоречивое, как и в случае упорядоченного множества: центральная статистика сама по себе не является статистикой, однако при каждом фиксированном θ — да.

Пример. Пусть $X_i \sim N(\theta, 1)$, $\theta \in \mathbb{R}$. Построим доверительный интервал для θ с уровнем доверия γ. За счёт УЗБЧ и ЦПТ мы имеем 2 прекрасных факта:

- ▷ $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim N(\theta, 1/n)$
- ▷ $\sqrt{n}(\overline{X} - \theta) \sim N(0, 1)$ (Действительно, $\overline{X} \sim N(\theta, 1/n)$, тогда $\overline{X} - \theta \sim N(0, 1/n)$ ну и домножение нормирует дисперсию)

Во втором факте оценка не зависит от θ, этим мы и воспользуемся. Пусть u_p — p-квантиль $N(0, 1)$. Тогда мы можем сказать следующее про интервал $\sqrt{n}(\overline{X} - \theta)$:

$$\forall \theta \in \Theta \ \ P_\theta(u_{\frac{1-\gamma}{2}} < \sqrt{n}(\overline{X} - \theta) < u_{\frac{1+\gamma}{2}}) = P_\theta(|\sqrt{n}(\overline{X} - \theta)| < u_{\frac{1+\gamma}{2}}) = \gamma$$

Неравенство можно переписать так, чтобы θ оказалось посередине. Это же даст нам вид оценок:

$$\overline{X} - \frac{u_{\frac{1+\gamma}{2}}}{\sqrt{n}} < \theta < \overline{X} - \frac{u_{\frac{1-\gamma}{2}}}{\sqrt{n}}$$

Итак, доверительный интервал — это $\left(\overline{X} - \frac{u_{\frac{1+\gamma}{2}}}{\sqrt{n}}, \overline{X} - \frac{u_{\frac{1-\gamma}{2}}}{\sqrt{n}}\right)$

Замечание. Пример выше, по факту, показал применение центральной статистики, где можно отделить X от θ. В общем случае, пусть у нас есть центральная статистика $G(X, \theta)$, $\gamma_{1,2} \in (0; 1)\colon \gamma_2 - \gamma_1 = \gamma$, а также g_i — обозначение γ_i-квантиля для функции распределения $G(X, \theta)$ при фиксированном θ. Тогда:

$$\forall \theta \in \Theta \ \ P_\theta(g_1 \leqslant G(X, \theta) \leqslant g_2) \geqslant \gamma_2 - \gamma_1 = \gamma$$

Это, увы, не доверительный интервал. Однако, мы можем построить доверительное множество $S(X) = \{\theta \in \Theta\colon g_1 \leqslant G(X, \theta) \leqslant g_2\}$. Тогда сразу имеем требуемое:

$$\forall \theta \in \Theta \ \ P_\theta(\theta \in S(X)) = P_\theta(g_1 \leqslant G(X, \theta) \leqslant g_2) \geqslant \gamma$$

Замечание. Есть так называемые *статистические таблицы*. Когда нужны конкретные квантили g_1, g_2, их обычно можно найти там.

Замечание. Следующая лемма и её следствия будут посвящены понятному вопросу: «А как искать центральные статистики?»

Лемма 3.3. *Пусть* $\{X_i\}_{i=1}^n$ *— выборка со строго возрастающей функцией распределения* F. *Если* F *непрерывна, то*

$$G(X_1, \ldots, X_n) = -\sum_{i=1}^{n} \ln F(X_i) \sim \Gamma(1, n)$$

Доказательство. Во-первых, покажем, что $F(X_i) \sim U[0;1]$. Действительно:

$$\forall y \in (0;1) \quad P(F(X_i) \leqslant y) = P(X_i \leqslant F^{-1}(y)) = F(F^{-1}(y)) = y$$

Стало быть, $-\ln F(X_i) =^d -\ln U[0;1] \sim Exp(1)$, а в силу свойства сложения случайных величин с экспоненциальным распределением получаем требуемое. □

Следствие. Если $\{X_i\}_{i=1}^n$ — выборка из распределения P_θ, причём при любом $\theta \in \Theta$ функция распределения $F_\theta(x)$ непрерывна и строго возрастает по x, то

$$G(X_1, \ldots, X_n, \theta) = -\sum_{i=1}^{n} \ln F_\theta(X_i) \text{ — центральная статистика}$$

Определение 3.37. Пусть $\{X_n\}_{n=1}^\infty$ — выборка из распределения P_θ. Последовательность пар статистик $(T_{n,1}(X), T_{n,2}(X))$ называется *асимптотическим доверительным интервалом уровня доверия* γ *для параметра* θ, если выполнено неравенство:

$$\forall \theta \in \Theta \quad \varliminf_{n\to\infty} P\big(T_{n,1}(X) < \theta < T_{n,2}(X)\big) \geqslant \gamma$$

Причём если для всех $\theta \in \Theta$ достигается равенство, то асимптотический доверительный интервал называется *точным.*

3.6.3 Построение доверительных интервалов

Утверждение 3.8. *Если имеется асимптотически нормальная оценка* $\hat\theta_n(X)$ *с непрерывным асимптотическим среднеквадратичным отклонением* $\sigma(\theta) > 0$, *то мы можем построить асимптотический доверительный интервал уровня доверия* γ *для* θ *следующего вида:*

$$\left(\hat\theta_n(X) - u_{\frac{1+\gamma}{2}} \cdot \frac{\sigma(\hat\theta_n)}{\sqrt{n}}, \hat\theta_n(X) + u_{\frac{1+\gamma}{2}} \cdot \frac{\sigma(\hat\theta_n)}{\sqrt{n}}\right)$$

где $\sigma(\hat\theta_n)$ *— среднеквадратичное отклонение оценки* $\hat\theta_n(X)$, u_p *— p-квантиль распределения* $N(0,1)$.

Доказательство. Действительно, воспользуемся определением асимптотической нормальности:

$$\forall \theta \in \Theta \quad \sqrt{n}(\hat\theta_n(X) - \theta) \xrightarrow[n\to\infty]{d_\theta} N(0, \sigma^2(\theta))$$

Поделим на $\sigma(\theta)$, чтобы перейти к $N(0,1)$ в правой части, однако это создаст проблемы слева, ибо появится лишняя зависимость от θ:

$$\forall\theta\in\Theta\quad \frac{\sqrt{n}(\hat{\theta}_n(X)-\theta)}{\sigma(\theta)}\xrightarrow[n\to\infty]{d_\theta} N(0,1)$$

Эту проблему можно решить при помощи леммы Слуцкого. Действительно, в силу асимптотической нормальности, $\hat{\theta}_n(X)\to^{P_\theta}\theta$, причём $\sigma(\theta)$ непрерывна по условию. Стало быть, есть сходимость $\sigma(\hat{\theta}_n(X))\to^{P_\theta}\sigma(\theta)$. При помощи упомянутой леммы можем собрать это в следующий факт:

$$\forall\theta\in\Theta\quad \frac{\sqrt{n}(\hat{\theta}_n(X)-\theta)}{\sigma(\theta)}\cdot\frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)}\xrightarrow[n\to\infty]{d_\theta} N(0,1)\cdot 1 = N(0,1)$$

Далее мы классически рассматриваем вероятность попадания между $u_{\frac{1-\gamma}{2}}$ и $u_{\frac{1+\gamma}{2}}$, которая будет сходиться к γ. Её несложно свернуть в следующую форму:

$$\forall\theta\in\Theta\quad P_\theta\left(\sqrt{n}\left|\frac{\hat{\theta}_n(X)-\theta}{\sigma(\hat{\theta}_n)}\right|<u_{\frac{1+\gamma}{2}}\right)\xrightarrow[n\to\infty]{}\gamma$$

Таким образом, мы нашли нужный асимптотический доверительный интервал. □

3.7 Метод максимального правдоподобия

Замечание. Далее мы живём в вероятностно-статистическом пространстве $(\mathcal{X},\mathfrak{B}(\mathcal{X}),\mathcal{P})$, $\mathcal{P}=\{P_\theta,\theta\in\Theta\}$.

Определение 3.38. Пусть X — наблюдение с неизвестным распределением $P_\theta\in\mathcal{P}$, причём $\mathcal{P}$ доминируется относительно меры μ. *Функцией правдоподобия* называется функция $f_\theta(x)=p_\theta(x)$, где p_θ — плотность P_θ по мере μ.

Замечание автора. В физическом смысле, функция правдоподобия говорит статисту, насколько вероятен тот или иной исход.

Пример. Пусть $X=(X_1,\ldots,X_n)$ — выборка с плотностью $p_\theta(x)$. Тогда функция правдоподобия является плотностью X как наблюдения:

$$f_\theta(x)=p_\theta(x)=\prod_{i=1}^n p_\theta(x_i)$$

Определение 3.39. Пусть X — наблюдение с функцией правдоподобия f_θ. *Оценкой параметра θ по методу максимального правдоподобия (ОМП)* называется такая статистика $\hat{\theta}(X)$, что верно равенство:

$$\hat{\theta}(X)=\arg\max_{\theta\in\Theta} f_\theta(X)$$

Замечание автора. То есть из всех возможных параметров ОМП выбирает тот, при котором заданная выборка наиболее вероятна.

Пример. Рассмотрим дискретную модель трёх бросков монетки с вероятностью θ получения орла и исход $X=(1,1,0)$. Тогда, для $\theta_1=\frac{1}{9}$ мы имеем значение функции правдоподобия $\frac{8}{9^3}$, а для $\theta_2=\frac{7}{8}$ это будет $\frac{7^2}{8^3}$ и это больше предыдущей вероятности. В связи с этим мы верим, что вторая монетка должна лучше предсказывать реальность, нежели первая.

Пример. Найдём явно оценку ОМП в базовом случае $X_i \sim U[0;\theta]$. Тогда функция правдоподобия имеет вид:

$$f_\theta(X) = \frac{1}{\theta^n}\prod_{i=1}^{n}\chi\{0 \leqslant X_i \leqslant \theta\} = \frac{\chi\{0 \leqslant X_{(1)} \leqslant X_{(n)} \leqslant \theta\}}{\theta^n}$$

Так как мы считаем, что реализация выборки X фиксирована при выборе θ, то оценка должна быть очевидна: $\hat{\theta}(X) = X_{(n)}$. Исправить, тут не вероятность написана, а что-то другое

Определение 3.40. Функция $L_\theta(x) = \ln f_\theta(x)$ называется *логарифом функции правдоподобия.*

Замечание. Далее мы расширяем и нумеруем *условия регулярности*:

0. $\mathcal{P}$ - доминируемое семейство относительно меры μ.
1. Множество носителей $A = \{x \in \mathcal{X} \colon p_\theta(x) > 0\}$ не зависит от θ
2. Наблюдение X есть выборка из неизвестного распределения P_θ
3. $\Theta \subseteq \mathbb{R}$ — открытый интервал (возможно бесконечный)
4. Функция $p_\theta(x)$ непрерывно дифференцируема по θ при всех $x \in A$
5. Функция $p_\theta(x)$ трижды непрерывно дифференцируема по θ при всех $x \in A$
6. Интеграл $\int_A p_\theta(x)d\mu(x)$ трижды дифференцируем по θ под знаком интеграла
7. Имеет место конечность информации Фишера для одного наблюдения из выборки:
$$\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\ln p_\theta(X_1)\right)^2 = i(\theta) \in (0;+\infty)$$
8. Существует равномерная интегрируемая оценка сверху в некотором интервале вокруг любого параметра $\theta_0 \in \Theta$:
$$\forall\theta_0 \in \Theta\ \exists c > 0, H(x) \mid \mathbb{E}_\theta H(X) < \infty \wedge \forall\theta \in (\theta_0 - c;\theta_0 + c),\ x \in A \quad \left|\frac{\partial^3}{\partial\theta^3}\ln p_\theta(x)\right| < H(x)$$

Теорема 3.13. *(Экстремальное свойство правдоподобия) Пусть выполнены условия регулярности 0-2. Тогда*

$$\forall\theta_0, \theta \in \Theta \quad (\theta_0 \neq \theta) \Rightarrow \lim_{n\to\infty} P_{\theta_0}(f_{\theta_0}(X_1,\ldots,X_n) > f_\theta(X_1,\ldots,X_n)) = 1$$

Доказательство. Рассмотрим $x_i \in A$. Тогда есть цепочка эквивалентностей:

$$f_{\theta_0}(x_1,\ldots,x_n) > f_\theta(x_1,\ldots,x_n) \Leftrightarrow \frac{1}{n}\ln\frac{f_\theta(x_1,\ldots,x_n)}{f_{\theta_0}(x_1,\ldots,x_n)} < 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n}\ln\frac{f_\theta(x_i)}{f_{\theta_0}(x_i)} < 0$$

Получили выражение, которое является средним арифметическим от $\ln \frac{f_\theta(x_i)}{f_{\theta_0}(x_i)}$. По УЗБЧ, если мы заменим x_i на $X_i \sim P_{\theta_0}$, будет сходимость P_{θ_0}-почти наверное к матожиданию $\mathbb{E}_{\theta_0} \ln \frac{f_\theta(X_1)}{f_{\theta_0}(X_1)}$. Если оно меньше нуля, то требуемое доказано:

$$\mathbb{E}_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \int_A \ln\left(\frac{p_\theta(x)}{p_{\theta_0}(x)}\right) p_{\theta_0}(x) d\mu(x) = \int_A \ln\left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x) d\mu(x) \leqslant \int_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x) d\mu(x) = \int_A p_\theta(x) d\mu(x) - \int_A p_{\theta_0}(x) d\mu(x) = 1 - 1 = 0$$

Последнее равенство опирается на то, что A — это носитель положительной вероятности. Сейчас мы доказали нестрогое неравенство, а надо строгое. Если выполнено равенство, то в том числе

$$\int_A \ln\left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x) d\mu(x) = 0 = \int_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x) d\mu(x)$$

Так как $\ln(1 + p_\theta(x)/p_{\theta_0}(x) - 1) \leqslant p_\theta(x)/p_{\theta_0}(x) - 1$, то такое возможно тогда и только тогда, когда эти функции равны почти-наверное на A. Стало быть, $\mu\{x \in A \colon p_\theta(x) \neq p_{\theta_0}(x)\} = 0$, а это противоречит условию регулярности 0. Значит неравенство строгое, что и требовалось. □

Следствие. Если Θ — конечное множество, то ОМП существует и единственна с вероятностью, стремящейся к единице (то есть, есть множество, на котором ОМП принимает одно и то же истинное значение, и вероятность этого множества стремится к 1) и состоятельна.

Доказательство. ОМП имеет вид $\widehat{\theta}(X) = \arg\max_{l \in \{1, \dots, L\}} f_{\theta_l}(X)$. Если есть несколько кандидатов в максимум, то выбираем то θ_l, чей номер минимален. Существование в таком виде тривиально, но теперь надо установить измеримость оценки (что она является случайной величиной). Введём событие $C_{i<j} = \{x \colon f_{\theta_i}(x) < f_{\theta_j}(x)\}$, аналогично $C_{i \leqslant j}$. Тогда мы можем описать прообраз оценки следующим образом:

$$\{x \colon \widehat{\theta}(x) = \theta_i\} = \left(\bigcap_{l<i} C_{l<i}\right) \cap \left(\bigcap_{l>i} C_{l \leqslant i}\right)$$

Так как f_{θ_i} есть фактически плотность меры P_{θ_i}, то это измеримая функция. Стало быть, каждое множество $C_{l<i}$, $C_{l \leqslant i}$ в конечном пересечении измеримо, а значит и пересечение тоже измеримо. Причём, если $X_k \sim P_{\theta_i}$, то работает доказанная теорема и вероятность такого множества стремится к единице (этим установлена единственность и одновременно состоятельность). □

Определение 3.41. *Уравнением правдоподобия* называется одно из двух эквивалентных уравнений:

$$\frac{\partial \ln f_\theta}{\partial \theta} = 0 \Leftrightarrow \frac{\partial f_\theta}{\partial \theta} = 0$$

Теорема 3.14. *(Аналог состоятельности ОМП) Пусть выполнены условия регулярности* 0-4 *и элементы выборки* $X_1, \dots, X_n \sim P_{\theta_0}$. *Тогда, существует отображение* $\widehat{\theta}_n(X_1, \dots, X_n, \theta_0)$ *со значениями в* Θ, *для которого есть свойства:*

▷ $\lim_{n\to\infty} P^*_{\theta_0}\{\widehat{\theta}_n$ *— не решение уравнения правдоподобия*$\} = 0$

▷ $\forall\varepsilon > 0 \quad \lim_{n\to\infty} P^*_{\theta_0}\{|\widehat{\theta}_n - \theta_0| > \varepsilon\} = 0$

где $P^*_{\theta_0}$ *— внешняя мера* P_{θ_0}.

Замечание. Неформально, с ростом размера выборки $\widehat{\theta}_n$ мы, во-первых, получаем через $\widehat{\theta}_n$ корень уравнения правдоподобия, а во-вторых, он будет близок к θ_0.

Доказательство. Пусть $x_1, \ldots, x_n \in A$ — реализация выборки. Определим $\widehat{\theta}_n$ следующим образом:

▷ Если у уравнения правдоподобия $\frac{\partial \ln f_\theta}{\partial\theta}(x_1, \ldots, x_n, \theta) = 0$ есть хоть 1 корень, то возьмём ближайший корень к θ_0 (что делать, если таких корней несчётное число, почему ближайшая к θ_0 точка тоже будет корнем? Потому что $\frac{\partial f}{\partial\theta}$ непрерывна из-за условий регулярности).

▷ Если у уравнения правдоподобия нет корней, то доопределим $\widehat{\theta}_n := \theta_0$.

Перейдём к доказательству свойств, а для этого зафиксируем $\varepsilon > 0$ так, что $[\theta_0-\varepsilon; \theta_0+\varepsilon] \subset \Theta$. Рассмотрим событие следующего вида:

$$S_n(\theta_0, \varepsilon) := \{x\colon f_{\theta_0-\varepsilon} < f_{\theta_0}(x_1, \ldots, x_n) \wedge f_{\theta_0}(x_1, \ldots, x_n) > f_{\theta_0+\varepsilon}(x_1, \ldots, x_n)\}$$

Тогда, в силу уже доказанного экстремального свойства правдоподобия, $\lim_{n\to\infty} P_{\theta_0}(S_n) = 1$. Если $x \in S_n$, то в силу условия регулярности 4 существует точка $\widetilde{\theta} \in (\theta_0 - \varepsilon; \theta_0 + \varepsilon)$, являющаяся локальным максимумом $f_\theta(x_1, \ldots, x_n, \theta)$. Стало быть, в этой точке $\frac{\partial f_\theta}{\partial\theta}(\widetilde{\theta}) = 0$. Так как по определению $\widehat{\theta}_n(x)$ — ближайший к θ_0 корень, то $|\widehat{\theta}_n(x) - \theta_0| < \varepsilon$. Таким образом:

$$\{\widehat{\theta}_n \text{ — не решение уравнения правдоподобия}\} \subseteq \mathcal{X}^n \backslash S_n$$

Из рассуждений выше мы сразу получаем все утверждения теоремы. □

Замечание. Важно отметить следующие факты о $\widehat{\theta}_n$:

▷ Не факт, что $\widehat{\theta}_n$ измерима по x

▷ Это не оценка, ибо полученное отображение зависит явно от θ_0

▷ Если корней несколько, то может быть неясно, какой ближе к θ_0

▷ Не обязательно, что корень является глобальным максимумом

▷ Корень существует не всегда

Следствие. Пусть выполнены условия регуляности 0-4 и для любого $n \in \mathbb{N}$ и любого $x \in A$ существует единственное решение уравнения правдоподобия $\widehat{\theta}_n(x_1, \ldots, x_n)$. Дополнительно, пусть оно является измеримой функцией. Тогда

1. $\widehat{\theta}_n(X)$ — состоятельная оценка θ

2. С вероятностью, стремящейся к единице, $\widehat{\theta}_n$ является ОМП

Доказательство.

1. Следует из доказанной теоремы
2. Согласно предыдущим обозначениям, $x \in S_n$. Тогда $\widehat{\theta}_n(x) = \widetilde{\theta}_n$ — точка локального максимума. Предположим, что ОМП даёт лучшую точку, отличную от $\widetilde{\theta}_n$. Тогда строго между ними обязана существовать точка, в которой занулится производная (точка локального минимума). Получаем противоречие с единственностью корня уравнения правдоподобия, а значит $\widehat{\theta}_n$ — ОМП на S_n, причём $\lim_{n\to\infty} P_{\theta_0}(S_n) = 1$.

□

Теорема 3.15. *(без доказательства) Пусть выполнены условия регулярности 0-8. Тогда для любая состоятельная оценка $\widehat{\theta}_n$, являющаяся решением уравнения правдоподобия, удовлетворяет соотношению:*

$$\forall \theta \in \Theta \quad \sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d_\theta} N\left(0, \frac{1}{i(\theta)}\right)$$

Замечание. Идея доказательства состоит в том, чтобы разложить $\widehat{\theta}_n$ в ряд Тейлора.

Следствие. (асимптотическая нормальность ОМП) Пусть выполнены условия регулярности 0-8. Если при любом $n \in \mathbb{N}$ и $x \in A$ существует единственное решение уравнения правдоподобия $\widehat{\theta}_n(x)$. Дополнительно, пусть оно является измеримой функцией. Тогда $\widehat{\theta}_n(X)$ является асимптотически нормальной оценкой θ с асимптотической дисперсией $\sigma^2(\theta) = \frac{1}{i(\theta)}$ и с вероятностью, стремящейся к 1, $\widehat{\theta}_n(X)$ является ОМП.

Теорема 3.16. *(Бахадура, без доказательства) Пусть выполнены условия регулярности 0-8. Также $\widehat{\theta}_n(X)$ — асимптотически нормальная оценка θ, причём асимптотическое отклонение $\sigma(\theta)$ непрерывно по θ (асимптотическая дисперсия — $\sigma^2(\theta)$). Тогда:*

$$\forall \theta \in \Theta \quad \sigma^2(\theta) \geqslant \frac{1}{i(\theta)}$$

Замечание. Таким образом, в условиях следствия про асипмтотическую нормальность ОМП, эта оценка обладает наилучшей асимптотической дисперсией среди асимптотических оценок с непрерывной дисперсией.

Определение 3.42. Пусть $\hat{\theta}_n(X)$ — асимптотически нормальная оценка θ с непрерывной $\sigma(\theta)$. Если $\sigma^2(\theta) = \frac{1}{i(\theta)}$, то оценка $\hat{\theta}_n(X)$ называется *асимтотически эффективной оценкой* θ.

Утверждение 3.9. *Пусть выполнены условия регулярности неравенства Рао-Крамера, $\hat{\theta}(X)$ — эффективная оценка θ, причём*

$$\forall x \in A, \theta \in \Theta \quad \widehat{\theta}(x) - \theta = c(\theta)U_\theta(x)$$

Тогда $\widehat{\theta}_n(X)$ — ОМП.

Доказательство. Подставим $c(\theta)$ явно, ведь мы знаем из критерия эффективности формулу для него:

$$\widehat{\theta}_n(X) - \theta = \frac{1}{I_X(\theta)} \cdot \frac{\partial}{\partial\theta} \ln f_\theta(X)$$

Отсюда при $\theta < \widehat{\theta}_n(x)$ получим $\frac{\partial}{\partial\theta} \ln f_\theta(x) > 0$ и аналогично при $\theta > \widehat{\theta}_n(x)$ будет $\frac{\partial}{\partial\theta} \ln f_\theta(x) < 0$, а значит $\theta = \widehat{\theta}(X)$ — точка максимума для $\ln f_\theta(X)$.

Я не знаю, что даёт этот факт и почему отсюда $\widehat{\theta}_n(X)$ — ОМП. Это не похоже на правду □

4 Линейная регрессионная модель

Замечание. В линейной модели наблюдением является случайный вектор $X \in \mathbb{R}^n$, который представим в виде $X = l + \varepsilon$, где l — неизвестный фиксированный вектор, а ε — тоже случайный вектор. Вектор l называется *оцениваемой величиной*, а ε — *ошибка измерений.*

Мы полагаем, что $\mathbb{E}\varepsilon = 0$ и $D\varepsilon = \sigma^2 E_n$, $\sigma > 0$ и неизвестна. Про l мы знаем, что $l \in L \subseteq \mathbb{R}^n$ — некоторое подпространство.

Наша задача состоит в том, чтобы оценить неизвестные параметры l и σ^2.

Пусть L задано при помощи своего базиса $\{z_1, \ldots, z_k\}$. Составим матрицу $Z = (z_1, \ldots, z_k)^{\square}$, тогда $l = Z\theta$, где $\theta = (\theta_1, \ldots, \theta_k)^T$ — неизвестные координаты в базисе z. Таким образом, задача сводится к оценке (θ, σ^2). Итоговый вид модели такой:

$$X = Z\theta + \varepsilon, \mathbb{E}\varepsilon = 0, D\varepsilon = \sigma^2 E_n$$

4.1 Метод Наименьших Квадратов (МНК)

Определение 4.1. *Оценкой по методу наименьших квадратов (МНК) для θ* называется оценка следующего вида:

$$\widehat{\theta}(X) = \arg\min_{Z\theta \in L} \|X - Z\theta\|^2 = \arg\min_{Z\theta \in L} \langle X - Z\theta, X - Z\theta \rangle$$

Замечание. Геометрический смысл состоит в том, что $\widehat{\theta}$ соответствует координатам проекции X на подпространство L.

Замечание. Далее $\widehat{\theta}(X)$ обозначает оценку МНК, если явно не сказано иного.

Лемма 4.1. *(Явный вид МНК) Для МНК имеет место формула:*

$$\widehat{\theta}(X) = (Z^T Z)^{-1} Z^T X$$

Доказательство. Найдём минимум скалярного произведения через дифференцирование по θ:

$$d\langle X - Z\theta, X - Z\theta \rangle = 2\langle -Zd\theta, X - Z\theta \rangle = \left\langle -2Z^T(X - Z\theta), d\theta \right\rangle \Rightarrow \nabla_\theta = 2Z^T(Z\theta - X)$$

Приравняем $\nabla_\theta = 0$ и найдём значение θ:

$$2Z^T(Z\theta - X) = 0;\ \theta = (Z^T Z)^{-1} Z^T X$$

$\square$

Замечание. Матрица Z^TZ невырождена, коль скоро это матрица Грама.

Утверждение 4.1. *Для оценки МНК* $E\widehat{\theta}(X) = \theta$, $D\widehat{\theta}(X) = \sigma^2(Z^TZ)^{-1}$

Доказательство. Считаем при помощи подстановки явной формулы:

▷ $\mathbb{E}\widehat{\theta}(X) = \mathbb{E}(Z^TZ)^{-1}Z^TX = (Z^TZ)^{-1}Z^T\mathbb{E}X = (Z^TZ)^{-1}Z^TZ\theta = \theta$

▷ $D\widehat{\theta}(X) = (Z^TZ)^{-1}Z^TDX((Z^TZ)^{-1}Z^T)^T = \sigma^2(Z^TZ)^{-1}$

$\square$

Теорема 4.1. *(без доказательства) Пусть* $t = T\theta$ *— линейный оператор,* $T \in \mathcal{M}_{m\times k}$*. Тогда оценка* $\widehat{t}(X) = T\widehat{\theta}(X)$ *является оптимальной оценкой* t *в классе линейных несмещённых оценок (то есть оценок вида* $B \cdot X$*)*

Лемма 4.2. *Имеет место равенство* $\mathbb{E}\|X - Z\widehat{\theta}(X)\|^2 = \sigma^2(n-k)$

Доказательство. По доказанному, $\mathbb{E}(X - Z\widehat{\theta}(X)) = 0$. Стало быть, $\mathbb{E}\|X - Z\widehat{\theta}(X)\|^2 = \operatorname{tr} D(X - Z\widehat{\theta}(X))$. Осталось расписать матрицу ковариаций в явном виде:

$$D(X - Z\widehat{\theta}(X)) = D\big((E_n - \underbrace{Z(Z^TZ)^{-1}Z}_{A})X\big) = (E_n - A)DX(E_n - A)^T = [A^T = A] =$$
$$\sigma^2(E_n - A)^2 = \sigma^2(E_n - 2A + A^2) = [A^2 = A] = \sigma^2(E_n - A)$$

Осталось подставить полученное выражение в матожидание (при этом $A = Z(Z^TZ)^{-1}Z = ZZ^{-1}Z^{-T}Z = Z^{-T}Z$):

$$\mathbb{E}\|X - Z\widehat{\theta}(X)\|^2 = \sigma^2(n - \operatorname{tr} A) = \sigma^2(n-k)$$

$\square$

Следствие.

1. $X - Z\widehat{\theta}(X) = \Pi_{L^\perp}X$ — проекция X на ортогональное дополнение к L
2. $\frac{1}{n-k}\|X - Z\widehat{\theta}(X)\|^2$ — несмещённая оценка σ^2

Доказательство.

1. В силу гильбертовости $\mathbb{R}^n$, $L \oplus L^\perp = \mathbb{R}^n$. Стало быть, $X = \Pi_L X + \Pi_{L^\perp}X$, причём первый вектор равен $Z\widehat{\theta}(X)$.
2. Тривиально.

$\square$

4.2 Гауссовская линейная модель

Замечание. Гауссовская линейная модель отличается от обычной тем, что добавляется условие
$\varepsilon \sim N(0, \sigma^2 E_n)$.

Напоминание. Распределение хи-квадрат с k степенями свободы можно с одной стороны считать $\chi^2_k = \Gamma(\frac{1}{2}, \frac{k}{2})$, а с другой стороны, это соответствует случайной величине $\xi_1^2 + \ldots + \xi_k^2$, $\xi_i \sim N(0,1)$.

Теорема 4.2. *(об ортогональном разложении, без доказательства) Пусть $X_1, \ldots, X_n \sim N(l, \sigma^2 E_n)$. Также $L_1 \oplus \ldots \oplus L_r = \mathbb{R}^n$ — есть разложение через попарно ортогональные пространства. Обозначим $Y_i = \Pi_{L_i} X$. Тогда $Y_1, \ldots, Y_r$ — независимые в совокупности гауссовские вектора, причём $\mathbb{E}Y_i = \Pi_{L_i} l$ и $\frac{1}{\sigma^2}(Y_i - \mathbb{E}Y_i)^2 \sim \chi^2_{\dim L_i}$*

Утверждение 4.2. *Статистика $S(X) = (\Pi_L X, \|\Pi_{L^\perp} X\|^2)$ является достаточной статистикой для (l, σ^2).*

Доказательство. Воспользуемся теоремой о факторизации. Распишем плотность X по формуле для многомерного нормального распределения:

$$p(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - l_i)^2\right)$$

В силу теоремы Пифагора:

$$\sum_{i=1}^n (x_i - l_i)^2 = \|x - l\|^2 = \|\Pi_L(x-l)\|^2 + \|\Pi_{L^\perp}(x-l)\|^2 = \|\Pi_L x - l\|^2 + \|\Pi_{L^\perp} x\|^2$$

Стало быть:

$$p(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) \exp\left(-\frac{1}{2\sigma^2}\left(\|\Pi_L x - l\|^2 + \|\Pi_{L^\perp} x\|^2\right)\right)$$

Значит, $S(X)$ действительно достаточная статистика. □

Теорема 4.3. *(без доказательства) Статистика $S(X) = (\Pi_L X, \|\Pi_{L^\perp} X\|^2)$ является полной.*

Следствие.
- ▷ $\widehat{\theta}(X)$ — оптимальная оценка для θ
- ▷ $Z\widehat{\theta}(X)$ — оптимальная оценка для l
- ▷ $\frac{1}{n-k}\|X - Z\widehat{\theta}(X)\|^2$ — оптимальная оценка для σ^2

Доказательство. Достаточно доказать, что все вышеперечисленные оценки являются функциями от $S(X)$, которая уже является полной достаточной статистикой.

- ▷ Сразу имеем $Z\widehat{\theta}(X) = \Pi_L X$
- ▷ Выразим $\widehat{\theta}(X)$ из верхнего равенства:

$$Z^T Z\widehat{\theta}(X) = Z^T \Pi_L X \Rightarrow \widehat{\theta}(X) = (Z^T Z)^{-1} Z^T \Pi_L X$$

▷ $\frac{1}{n-k}\|X - Z\widehat{\theta}(X)\|^2 = \frac{1}{n-k}\|X - \Pi_L X\|^2 = \frac{1}{n-k}\|\Pi_{L^\perp} X\|^2$

□

Утверждение 4.3. *Статистики $\widehat{\theta}(X)$ и $X - Z\widehat{\theta}(X)$ независимы. Более того:*

▷ $\frac{1}{\sigma^2}\|Z\widehat{\theta}(X) - Z\theta\|^2 \sim \chi^2_k$

▷ $\frac{1}{\sigma^2}\|X - Z\widehat{\theta}(X)\|^2 \sim \chi^2_{n-k}$

▷ $\widehat{\theta}(X) \sim N(\theta, \sigma^2(Z^T Z)^{-1})$

Доказательство. Согласно теореме об ортогональном разложении, гауссовские векторы $Z\widehat{\theta}(X) = \Pi_L X$ и $X - Z\widehat{\theta}(X) = \Pi_{L^\perp} X$ являются независимыми, причём выполнены первые 2 пункта из утверждения. Для последнего заметим, что

$$\widehat{\theta}(X) = (Z^T Z)^{-1} Z^T Z\widehat{\theta}(X) \text{ — линейная функция от } Z\widehat{\theta}(X)$$

Стало быть, $\widehat{\theta}(X) \perp\!\!\!\perp X - Z\widehat{\theta}(X)$, независимость установлена. Так как $\widehat{\theta}(X)$ является линейной функцией от гауссовского вектора, то это тоже гауссовский вектор. Параметры мы уже находили ранее. □

Определение 4.2. Пусть $\xi \sim N(0,1)$, $\eta \sim \chi^2_k$ и $\xi \perp\!\!\!\perp \eta$. Тогда случайная величина $\zeta = \frac{\xi}{\sqrt{\eta/k}}$ обладает *распределением Стьюдента с k степенями свободы*. Обозначается как $\zeta \sim T_k$

Утверждение 4.4. *У распределения Стьюдента есть несколько базовых свойств:*

1. *Если $\zeta \sim T_k$, то и $-\zeta \sim T_k$*
2. *$T_1 \sim Cauchy(0,1)$. Плотность этого распределения равна $p(x) = \frac{1}{\pi(1+x^2)}$*
3. *Если $\zeta_k \sim T_k$, то $\zeta_k \xrightarrow{d} N(0,1)$*

Определение 4.3. Пусть $\xi \sim \chi^2_k$, $\eta \sim \chi^2_m$ и $\xi \perp\!\!\!\perp \eta$. Тогда случайная величина $\zeta = \frac{\xi/k}{\eta/m}$ обладает *распределением Фишера с параметрами k, m*. Обозначается как $\zeta \sim F_{k,m}$

Утверждение 4.5. *Распределение Фишера обладает следующими свойствами:*

1. *Если $\xi \sim T_m$, то $\xi^2 \sim F_{1,m}$*
2. *Если $\xi \sim F_{k,m}$, то $\frac{1}{\xi} \sim F_{m,k}$*
3. *Пусть k фиксировано, $\xi_m \sim F_{k,m}$. Тогда $k\xi_m \xrightarrow{d} \chi^2_k$*
4. *Пусть $\xi_{k,m} \sim F_{k,m}$. Тогда $\xi_{k,m} \xrightarrow[k,m\to\infty]{d} 1$*

4.2.1 Доверительные интервалы в гауссовской линейной модели

Далее γ — уровень доверия.

- Для σ^2: у нас есть статистика $\frac{1}{\sigma^2}\|X - Z\widehat{\theta}(X)\|^2 \sim \chi^2_{n-k}$. Возьмём $u_{1-\gamma}$ — соответствующий квантиль χ^2_{n-k}. Тогда

$$\gamma = P(\frac{1}{\sigma^2}\|X - Z\widehat{\theta}(X)\|^2 > u_{1-\gamma}) = P\left(\sigma^2 \in \left(0; \frac{\|X - Z\widehat{\theta}(X)\|^2}{u_{1-\gamma}}\right)\right)$$

- Для θ_i: воспользуемся тем фактом, что $\widehat{\theta}(X) \sim N(\theta, \sigma^2 A)$, где $A = (Z^T Z)^{-1}$. Тогда $\theta_i \sim N(\theta_i, \sigma^2 a_{ii})$, а значит $\frac{\widehat{\theta}_i - \theta_i}{\sqrt{\sigma^2 a_{ii}}} \sim N(0,1)$. Чтобы убрать σ^2 из знаменателя, вспомним, что $\widehat{\theta}(X) \perp\!\!\!\perp X - Z\widehat{\theta}(X)$. Стало быть, можем поделить оценку на корень из $\frac{1}{\sigma^2}\|X - Z\widehat{\theta}(X)\|^2 \sim \chi^2_{n-k}$ и получить распределение Стюдента:

$$\sqrt{\frac{n-k}{a_{ii}}} \cdot \frac{\widehat{\theta}_i - \theta_i}{\sqrt{\|X - Z\widehat{\theta}(X)\|^2}} \sim T_{n-k}$$

  Дальше уже понятно, как получить доверительный интервал.

- Для θ: доверительный интервал тут по определению не получишь, однако мы сможем получить хорошее доверительное множество. Воспользуемся независимыми статистиками $\frac{1}{\sigma^2}\|Z\widehat{\theta}(X) - Z\theta\|^2 \sim \chi^2_k$ и $\frac{1}{\sigma^2}\|X - Z\widehat{\theta}\|^2 \sim \chi^2_{n-k}$. Их комбинацией можно получить распределение Фишера:

$$\frac{n-k}{k} \cdot \frac{\|Z\widehat{\theta}(X) - Z\theta\|^2}{\|X - Z\widehat{\theta}(X)\|^2} \sim F_{k,n-k}$$

  Множество, которое из этой оценки задаётся как $\{\frac{n-k}{k} \cdot \frac{\|Z\widehat{\theta}(X) - Z\theta\|^2}{\|X - Z\widehat{\theta}(X)\|^2} < u_\gamma\}$, является доверительным эллипсоидом в $\mathbb{R}^k$.

5 Проверка статистических гипотез

Замечание. Далее мы живём в вероятностно-статистической модели $(\mathcal{X}, \mathfrak{B}(\mathcal{X}), \mathcal{P})$. Пусть наблюдение $X \sim P \in \mathcal{P}$.

Определение 5.1. *Статистической гипотезой* называется предположение вида $P \in \mathcal{P}_0$, где $\mathcal{P}_0 \subseteq \mathcal{P}$ — подмножество распределений. Обозначается как $H_0\colon P \in \mathcal{P}_0$ — гипотеза H_0.

Замечание. Какую-то зафиксированную, выделенную гипотезу мы будем называть *основной*.

Наша задача состоит в том, чтобы по наблюдению X либо *отвергнуть*, либо *не отвергнуть («принять»)* гипотезу H_0. Если гипотеза отвергнута, то мы считаем, что распределение $P \notin \mathcal{P}_0$, а значит ответ надо искать среди $P \in \mathcal{P} \setminus \mathcal{P}_0$ (при этом мы не сможем гарантировать, что мы не ошиблись в своих суждениях из-за неудачной реализации наблюдения).

Определение 5.2. Пусть $H_0 \colon P \in \mathcal{P}_0$ — основная гипотеза. Тогда $H_1 \colon P \in \mathcal{P}_1 \subseteq \mathcal{P} \backslash \mathcal{P}_0$ называется *альтернативой (альтернативной гипотезой)*.

Определение 5.3. Пусть X принимает значения в выборочном множестве $\mathcal{X}_1$, а $S \subseteq \mathcal{X}$ — некоторое подмножество. Если правило принятия гипотезы H_0 выглядит следующим образом:

$$H_0 \text{ отвергается} \Leftrightarrow X \in S$$

То S называется *критическим множеством, или же критерием для проверки гипотезы H_0 (против альтернативы H_1, если она есть)*.

Определение 5.4. *Ошибкой первого рода* называется ситуация, когда H_0 отвергли, но при этом гипотеза верна.

Определение 5.5. *Ошибкой второго рода* называется ситуация, когда H_0 не отвергли, но при этом гипотеза неверна.

Замечание автора. В первую очередь стремятся минимизировать ошибку первого рода, а потом вторую. Причину можно проиллюстрировать на жизненном примере: если гипотеза H_0 состоит в том, что человек болен раком, то мы бы сильно хотели, чтобы у него не было заболевания при отвержении гипотезы, а если мы его будем лечить от болезни, которой у него нет, то это не так страшно.

Определение 5.6. Пусть S — критерий для проверки гипотезы $H_0 \colon P \in \mathcal{P}_0$. Функция $\beta(Q, S) = Q(X \in S)$, где $Q \in \mathcal{P}$, называется *функцией мощности критерия S*.

Определение 5.7. Если при некотором $\varepsilon > 0$ для критерия S выполнено неравенство

$$\forall Q \in \mathcal{P}_0 \quad \beta(Q, S) \leqslant \varepsilon \quad (\text{это эквивалентно } \sup_{Q \in \mathcal{P}_0} \beta(Q, S) \leqslant \varepsilon)$$

то говорят, что *критерий S имеет уровень значимости ε*.

Замечание. Уровень значимости критерия даёт верхнюю оценку на вероятность ошибки первого рода.

Замечание автора. Важно не путать уровень *значимости* и уровень *доверия*. Сумма этих величин даёт единицу.

Определение 5.8. Минимальный уровень значимости $\alpha(S) = \sup_{Q \in \mathcal{P}_0} \beta(Q, S)$ называется *размером критерия S*

Замечание. Далее S — это критерий проверки гипотезы H_0 против альтернативы H_1.

Определение 5.9. Критерий S называется *несмещённым*, если выполнено условие (несмещённости):

$$\sup_{Q \in \mathcal{P}_0} \beta(Q, S) \leqslant \inf_{Q \in \mathcal{P}_1} \beta(Q, S)$$

Замечание автора. Говоря другими словами, критерий несмещён, если вероятность ошибки первого рода меньше, чем вероятность отвержения гипотезы при верности альтернативы.

Определение 5.10. Пусть S_n — последовательность критериев (или просто критерий), отвечающих выборке $X = (X_1, \ldots, X_n)$. Тогда критерий S_n называется *состоятельным*, если

$$\forall Q \in \mathcal{P}_1 \quad \lim_{n\to\infty} \beta(Q, S_n) = 1$$

Замечание автора. Другими словами, критерий состоятелен, если с увеличением выборки (следовательно, информации, которую мы знаем о распределении) вероятность отвергнуть гипотезу при верной альтернативе стремится к единице.

Определение 5.11. Критерий S называется более мощным, чем критерий R того же уровня значимости, если выполнено утверждение:

$$\forall Q \in \mathcal{P}_1 \quad \beta(Q, S) \geqslant \beta(Q, R)$$

Иными словами, вероятность ошибки второго рода у критерия S равномерно меньше $(1 - \beta(Q, S) \leqslant 1 - \beta(Q, R))$.

Определение 5.12. Критерий S называется *равномерно наиболее мощным критерием (РНМК) уровня значимости* ε, если его мощность $\alpha(S) \leqslant \varepsilon$ и S мощнее любого другого критерия R с уровнем значимости ε (эквивалентно $\alpha(R) \leqslant \varepsilon$).

Замечание. Хороший вопрос состоит в том, как искать РНМК. Для этого нужно конкретизировать гипотезы, с которыми мы работаем.

Определение 5.13. Гипотеза $H_0\colon P = P_0$, где P_0 — известное распределение, называется *простой*.

5.1 Проверка простых гипотез

Замечание. Далее мы полагаем, что у нас есть основная H_0 и альтернативная H_1 простые гипотезы, причём P_i имеет плотность $p_i(x)$ по одной и той же мере $\mu(x)$.

Также мы определим критерий S_λ следующим образом:

$$S_\lambda := \{x\colon p_1(x) - \lambda p_0(x) \geqslant 0\},\ \lambda \geqslant 0$$

Замечание автора. Если описывать словами критерий S_λ, то мы отвергаем основную гипотезу, если вероятность реализации выборки при H_1 в λ раз больше, чем та же вероятность при H_0.

Лемма 5.1. *(Неймана-Пирсона) Пусть критерий R удовлетворяет соотношению $\beta(P_0, R) \leqslant \beta(P_0, S_\lambda)$. Тогда*

1. $\beta(P_1, R) \leqslant \beta(P_1, S_\lambda)$ *(S_λ мощнее R)*
2. $\beta(P_0, S_\lambda) \leqslant \beta(P_1, S_\lambda)$ *(S_λ несмещён)*

Доказательство. Далее $I_A(x)$ — индикатор $x \in A$.

1. Что получится, если проинтегрировать функцию из определения S_λ по этому же множеству?

$$\int_{S_\lambda} (p_1(x) - \lambda p_0(x)) d\mu(x) = \int_{\mathcal{X}} I_{S_\lambda}(p_1(x) - \lambda p_0(x)) d\mu(x) = P_1(X \in S_\lambda) - \lambda P_0(X \in S_\lambda)$$

Осталось заметить, что этот же интеграл можно оценить снизу аналогичным, но по R, ибо все неотрицательные значения подынтегральной функции уже учтены (в силу определения S_λ):

$$P_1(X \in R) - \lambda P_0(X \in R) = \int_R (p_1(x) - \lambda p_0(x)) d\mu(x) \leqslant \int_{S_\lambda} (p_1(x) - \lambda p_0(x)) d\mu(x) = P_1(X \in S_\lambda) - \lambda P_0(X \in S_\lambda)$$

Это неравенство можно переписать так:

$$P_1(X \in S_\lambda) - P_1(X \in R) \geqslant \lambda(P_0(X \in S_\lambda) - P_0(X \in R)) \geqslant 0$$

2. Придётся разобрать несколько случаев

   ▷ $\lambda > 1$. Значит, $p_1(x) \geqslant p_0(x)$ при $x \in S_\lambda$:

   $$P_0(X \in S_\lambda) = \int_{S_\lambda} p_0(x) d\mu(x) \leqslant \int_{S_\lambda} p_1(x) d\mu(x) = P_1(X \in S_\lambda)$$

   ▷ $\lambda \in [0; 1]$. В этом случае наоборот, при $x \in \overline{S_\lambda}$ имеем $p_1(x) \leqslant p_0(x)$. Для нужной оценки обратимся к вероятности на дополнении $\overline{S_\lambda}$:

   $$1 - \beta(P_1, S_\lambda) = P_1(X \in \overline{S_\lambda}) = \int_{\overline{S_\lambda}} p_1(x) d\mu(x) \leqslant \int_{\overline{S_\lambda}} p_0(x) d\mu(x) = P_0(X \in \overline{S_\lambda}) = 1 - \beta(P_0, S_\lambda)$$

□

Следствие. Если $\lambda > 0$ и $P_0(X \in S_\lambda) = \varepsilon > 0$, то S_λ — РНМК уровня значимости ε.

Замечание. Стало быть, для нахождения РНМК нужно решить уравнение относительно λ:

$$\int_{S_\lambda} p_0(x) d\mu(x) = \varepsilon > 0$$

В случае абсолютно непрерывного распределения (μ — мера Лебега) решение, как правило, существует, а в дискретном случае решения может не быть для всех $\varepsilon > 0$. Поэтому приходится выбирать из тех ε, при которых уравнение разрешимо.

5.2 Монотонное отношение правдоподобия

Замечание. Пусть семейство $\mathcal{P}$ параметризовано параметром $\theta \in \mathbb{R}$, а также оно доминируемо относительно меры μ (этим мы гарантируем налачие функции правдоподобия).

Определение 5.14. Семейство $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ называется *монотонным относительно правдоподобия по статистике* $T(X)$, если

$$\forall \theta_0 < \theta_1 \quad \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} \text{ — монотонная функция от } T(x)$$

причём характер монотонности один и тот же.

Теорема 5.1. *(о монотонности относительно правдоподобия, без доказательства) Пусть даны гипотезы $H_0\colon \theta \leqslant \theta_0$ (или $\theta = \theta_0$), $H_1\colon \theta > \theta_0$, а семейство $\mathcal{P}$ монотонно относительно правдоподобия, причём характер монотонности — неубывание. Тогда критерий $S_\varepsilon = \{T(x) \geqslant c_\varepsilon\}$ с условием $P_{\theta_0}(S_\varepsilon) = \varepsilon$ является РНМК с уровнем значимости ε для проверки H_0 против H_1.*

5.3 Двойственность доверительного оценивания и проверки гипотез

$\Rightarrow$ Пусть $S(X)$ — доверительная область уровня доверия $1-\varepsilon$ для параметра $\theta \in \Theta$. Если я хочу проверить простую гипотезу $H_0\colon \theta = \theta_0$, то имеет смысл рассмотреть такой критерий:

$$\widetilde{S}(\theta) = \{x \in \mathcal{X}\colon \theta \notin S(x)\}$$

Тогда $\widetilde{S}(\theta_0)$ — критерий с уровнем значимости ε для проверки H_0. Проверим этот факт явно:

$$P_{\theta_0}(X \in \widetilde{S}(\theta_0)) = P_{\theta_0}(\theta_0 \notin S(X)) = 1 - P_{\theta_0}(\theta_0 \in S(X)) \leqslant \varepsilon$$

$\Leftarrow$ Наоборот, S_{θ_0} — критерий проверки гипотезы $H_0\colon \theta = \theta_0$ с уровнем значимости ε. Пусть известен критерий S_{θ_0} при любом $\theta_0 \in \Theta$. Тогда мы можем рассмотреть $S(X) = \{\theta \in \Theta\colon X \notin S_\theta\}$, это будет искомым доверительным множеством с уровнем доверия $1-\varepsilon$:

$$P_\theta(\theta \in S(X)) = P_\theta(X \notin S_\theta) = 1 - P_\theta(X \in S_\theta) \geqslant 1 - \varepsilon$$

5.4 Проверка гипотез в гауссовской линейной модели

Замечание. Рассмотрим гауссовскую линейную модель:

$$X = Z\theta + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 E_n),\ \theta \in \mathbb{R}^k,\ Z \in \mathcal{M}_{n\times k},\ \operatorname{rk} Z = k \leqslant n$$

Задача. *Проверить простую гипотезу $H_0\colon T\theta = t$, где $T \in \mathcal{M}_{m\times k}$, $t \in \mathbb{R}^m$ и $\operatorname{rk} T = m \leqslant k$. Уровень значимости α.*

Решение. (F-критерий или критерий Фишера) Идея состоит в том, что мы умеем по выборке оценивать $T\theta$ с использованием оценки $\widehat{\theta}(X)$. Мы будем строить критерий исходя из того, что надо проверить, насколько сильное отклонение в сравнении $T\widehat{\theta}(X)$ и t. Далее для вывода критерия мы предполагаем верность гипотезы $T\theta = t$.

Итак, $\widehat{\theta}(X) = (Z^TZ)^{-1}Z^TX$ — это ОНК для θ. В силу известных фактов, $\widehat{t}(X) = T\widehat{\theta}(X)$ — оптимальная оценка для $T\theta$. Так как распределение $\widehat{\theta}(X) \sim N(\theta, \sigma^2(Z^TZ)^{-1})$, то $\widehat{t}(X) \sim N(T\theta, T\sigma^2(Z^TZ)^{-1}T^T) =: N(T\theta, \sigma^2 B)$. Матрица B положительно определена и симметрична, а поэтому существует $\sqrt{B}$ — тоже симметричная матрица. Это позволяет оценку с независящим от параметров распределением:

$$\frac{1}{\sigma}(\sqrt{B})^{-1}(\widehat{t}(X) - T\theta) \sim N(0, E_m) \Rightarrow Q_T(X) := \left\|\frac{1}{\sigma}(\sqrt{B})^{-1}(\widehat{t}(X) - T\theta)\right\|^2 \sim \chi^2_m$$

Как и раньше, $Q_T(X)$ нам не подойдёт, ибо присутствует σ, но это мы решим применением ещё одной статистики. Сейчас же перепишем $Q_T(X)$ в более удобном виде (с учётом верности гипотезы):

$$Q_T(X) = \frac{1}{\sigma^2}\big((\sqrt{B})^{-1}(\widehat{t}(X) - t)\big)^T(\sqrt{B})^{-1}(\widehat{t}(X) - t) = \frac{1}{\sigma^2}\underbrace{(\widehat{t}(X) - t)^T B^{-1}(\widehat{t}(X) - t)}_{\widehat{Q}_T(X)}$$

Так как $\widehat{Q}_T(X)$ выражается через $\widehat{\theta}(X)$, то $\widehat{Q}_T(X) \perp\!\!\!\perp X - Z\widehat{\theta}(X)$, при этом $\frac{1}{\sigma^2}\|X - Z\widehat{\theta}(X)\|^2 \sim \chi^2_{n-k}$. Отсюда, в условиях гипотезы H_0, получаем следующую статистику:

$$F(X) := \frac{\sigma^2 Q_T(X)}{\|X - Z\widehat{\theta}(X)\|^2} \cdot \frac{n-k}{m} = \frac{\widehat{Q}_T(X)}{\|X - Z\widehat{\theta}(X)\|^2} \cdot \frac{n-k}{m} \sim F_{m,n-k}$$

Остаётся взять квантиль $u_{1-\alpha}$, тогда $\{x\colon F(x) > u_{1-\alpha}\}$ — искомый критерий, также называемый F*-критерием или критерием Фишера*

Пример. Рассмотрим следующий пример: есть два груза с неизвестными весами a_1, a_2. Производится n взвешиваний первого и m взвешиваний второго. Требуется проверить гипотезу $H_0\colon a_1 = a_2$.

Для начала, переформулируем условие задачи: $X_1, \ldots, X_n \sim N(a_1, \sigma^2)$, $Y_1, \ldots, Y_m \sim N(a_2, \sigma^2)$, $W := (X_1, \ldots, X_n, Y_1, \ldots, Y_m)^T$. Получается гауссовская линейная модель:

$$W = Z\theta + \varepsilon,\ \theta = (a_1, a_2)^T,\ \varepsilon \sim N(0, \sigma^2 E_{n+m})$$

Для краткости, описание Z опускается (задача читателю на дом). Гипотезу H_0 можно переформулировать в терминах, изложенных ранее, тогда $T = (1, -1)$ и $t = 0$.

- $\widehat{\theta}(W) = (Z^T Z)^{-1} Z^T W = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}$ (проверяется явным вычислением с Z)
- $\widehat{t}(W) = \overline{X} - \overline{Y}$
- $B = T(Z^T Z)^{-1} T^T = \frac{1}{n} + \frac{1}{m}$
- $\widehat{Q}_T(W) = \frac{nm}{n+m}(\overline{X} - \overline{Y})^2$
- $\|W - Z\widehat{\theta}(W)\|^2 = \sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{j=1}^m (Y_i - \overline{Y})^2 = nS_X^2 + mS_Y^2$

Из всего вышесказанного, F-статистика принимает вид:

$$F(W) = \frac{\frac{nm}{n+m}(\overline{X} - \overline{Y})^2}{nS_X^2 + mS_Y^2} \cdot \frac{n+m-2}{1} \sim F_{1,n+m-2}$$