# Математическая статистика
# 5 семестр
# Лектор: Савёлов М.П.
# осень 2022

**Авторы билетов (лучшие котики):**

Спицын Николай

Савичев Дмитрий

Подзорова Полина

Чубенко Полина

Климанова Ирина

Клячин Артемий

Сбродов Егор

Ксения Куринова

Special thanks to автору конспекта прошлого года, Хаймоненко Виктору

# Содержание

# 1 Виды сходимостей случайных векторов и связи между ними. Связь между сходимостью векторов и сходимостью их компонент. Теорема о наследовании сходимости.

Зафиксируем $\xi, \{\xi_n\}_{n=1}^{\infty}$−случайные векторы.

**Определение 1.1.** $\xi_n \xrightarrow{a.s.} \xi \overset{Def}{\Longleftrightarrow} P\left(\xi_n \to \xi\right) = 1 \Leftrightarrow P\left(\{\omega : \xi_n(\omega) \to \xi(\omega)\}\right) = 1.$

**Определение 1.2.** $\xi_n \xrightarrow{P} \xi \overset{Def}{\Longleftrightarrow} \forall \varepsilon > 0\ P(\|\xi_n - \xi\|_2 > \varepsilon) \to 0$, где $\|x\|_p = \left(\sum\limits_i |x_i|^p\right)^{1/p}$.

**Определение 1.3.** $\xi_n \xrightarrow{L_p} \xi \overset{Def}{\Longleftrightarrow} E(\|\xi_n - \xi\|_p)^p \xrightarrow[n\to\infty]{} 0.$

**Определение 1.4.** $\xi_n \xrightarrow{d} \xi \overset{Def}{\Longleftrightarrow} \forall f : \mathbb{R}^m \to \mathbb{R}$ – ограниченной и непрерывной выполняется $Ef(\xi_n) \xrightarrow[n\to\infty]{} Ef(\xi)$.

**Утверждение 1.5.** Пусть $\xi, \{\xi_n\} \in \mathbb{R}^m$

1. $\xi_n \xrightarrow{a.s.} \xi \Leftrightarrow \forall i \in \overline{1, m}\ \xi_n^{(i)} \xrightarrow{a.s.} \xi^{(i)}$
2. $\xi_n \xrightarrow{P} \xi \Leftrightarrow \forall i \in \overline{1, m}\ \xi_n^{(i)} \xrightarrow{P} \xi^{(i)}$
3. $\xi_n \xrightarrow{L_p} \xi \Leftrightarrow \forall i \in \overline{1, m}\ \xi_n^{(i)} \xrightarrow{L_p} \xi^{(i)}$
4. $\xi_n \xrightarrow{d} \xi \Rightarrow \forall i \in \overline{1, m}\ \xi_n^{(i)} \xrightarrow{d} \xi^{(i)}$

**Доказательство.**

1. $$\{\xi_n \to \xi\} = \bigcap_{i=1}^{m} \left\{\xi_n^{(i)} \to \xi^{(i)}\right\} \subset \left\{\xi_n^{(i)} \to \xi^{(i)}\right\} \Rightarrow P\left(\xi_n \to \xi\right) \leqslant P\left(\xi_n^{(i)} \to \xi^{(i)}\right)\ \forall i \in \overline{1, m}.$$

   Пусть $\xi_n \xrightarrow{a.s.} \xi$. Тогда

   $$1 = P\left(\xi_n \to \xi\right) \leqslant P\left(\xi_n^{(i)} \to \xi^{(i)}\right) \Rightarrow P\left(\xi_n^{(i)} \to \xi^{(i)}\right) = 1\ \forall i \in \overline{1, m}.$$

   Обратно, пусть $\forall i \in \overline{1, m}\ \xi_n^{(i)} \xrightarrow{a.s.} \xi^{(i)}$. Предположив противное, получаем

   $$\{\xi_n \nrightarrow \xi\} = \bigcup_{i=1}^{m} \overline{\left\{\xi_n^{(i)} \to \xi^{(i)}\right\}} = \bigcup_{i=1}^{m} \left\{\xi_n^{(i)} \nrightarrow \xi^{(i)}\right\} \Rightarrow P(\xi_n \nrightarrow \xi) \leqslant \sum_{i=1}^{m} P\left(\xi_n^{(i)} \nrightarrow \xi^{(i)}\right) = 0.$$

2. Пусть $\xi_n \xrightarrow{P} \xi$, то есть

   $$\forall \delta > 0\ \exists N : \forall n \geqslant N\ \delta > P(\|\xi_n - \xi\|_2 > \varepsilon).$$

   Тогда

   $$\delta > P(\|\xi_n - \xi\|_2 > \varepsilon) \geqslant P\left(\left|\xi_n^{(i)} - \xi^{(i)}\right| > \varepsilon\right) \geqslant 0 \Rightarrow P\left(\left|\xi_n^{(i)} - \xi^{(i)}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0.$$

Обратно, пусть $\forall i \in \overline{1,m}\ \xi_n^{(i)} \xrightarrow{P} \xi^{(i)}$. Тогда

$$\forall \varepsilon > 0\ \sqrt{\sum_{i=1}^{m}\left(\xi_n^{(i)} - \xi^{(i)}\right)^2} > \varepsilon \Leftrightarrow \sum_{i=1}^{m}\left(\xi_n^{(i)} - \xi^{(i)}\right)^2 > \varepsilon^2 \Rightarrow \exists i : \left(\xi_n^{(i)} - \xi^{(i)}\right)^2 > \frac{\varepsilon^2}{m} \Leftrightarrow \left|\xi_n^{(i)} - \xi^{(i)}\right| > \frac{\varepsilon}{\sqrt{m}}.$$

Получаем, что

$$P\left(\|\xi_n - \xi\|_2 > \varepsilon\right) \leqslant \sum_{i=1}^{m} P\left(\left|\xi_n^{(i)} - \xi^{(i)}\right| > \frac{\varepsilon}{\sqrt{m}}\right) \xrightarrow{n\to\infty} 0.$$

3.

$$E\sum_{i=1}^{m}\left|\xi_n^{(i)} - \xi^{(i)}\right|^p = \sum_{i=1}^{m} E\left(\left|\xi_n^{(i)} - \xi^{(i)}\right|^p\right)$$

Значит, если левая часть стремится к нулю, то $\forall\ i:\ E\left(\left|\xi_n^{(i)} - \xi^{(i)}\right|^p\right) \xrightarrow{n\to\infty} 0$.

Если $\forall\ iE\left(\left|\xi_n^{(i)} - \xi^{(i)}\right|^p\right) \xrightarrow{n\to\infty} 0$, то и левая часть стремится к 0

4. Пусть $g : \mathbb{R} \to \mathbb{R}$ – непрерывная ограниченная функция, $h : \mathbb{R}^m \to \mathbb{R},\ h(x) = x^{(i)}$. Тогда $g \circ h$ – непрерывная ограниченная функция и

$$E\left(g\left(\xi_n^{(i)}\right)\right) = E\left(g\left(h\left(\xi_n\right)\right)\right) \xrightarrow{n\to\infty} E\left(g\left(h\left(\xi\right)\right)\right) = E\left(g\left(\xi^{(i)}\right)\right).$$

□

**Замечание 1.6.** *Обратная импликация в п.4 неверна.* Достаточно взять $\xi, \eta$ - НОРСВ. $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} \eta$, где $\xi \overset{d}{=} \eta$. Но $(\xi_n, \eta_n) \overset{d}{\nrightarrow} (\xi, \xi)$, а $(\xi_n, \eta_n) \xrightarrow{d} (\xi, \eta)$

**Утверждение 1.7.** Пусть $\{\xi_n\}_{n=1}^{\infty}$ – случайные векторы, и $\xi_n \xrightarrow{d} c = const$, то $\xi_n \xrightarrow{P} c$.

**Доказательство.** Из п.4 предыдущего утверждения

$$\forall i \in \overline{1,m}\ \xi_n^{(i)} \xrightarrow{d} c^{(i)} \Rightarrow \xi_n^{(i)} \xrightarrow{P} c^{(i)} \xrightarrow{p.2} \xi_n \xrightarrow{P} c.$$

□

**Теорема 1.8.** (б/д) Пусть $\xi,\ \{\xi_n\}$ – случайные векторы. Тогда

$$\begin{array}{lll} \xi_n \xrightarrow{a.s.} \xi & & \\ & \searrow & \\ & & \xi_n \xrightarrow{P} \xi \quad \Rightarrow \quad \xi_n \xrightarrow{d} \xi \\ & \nearrow & \\ \xi_n \xrightarrow{L_p} \xi & & \end{array}$$

**Теорема 1.9.** (О наследовании сходимостей) Пусть $\xi,\ \{\xi_n\}$ – случайные векторы в $\mathbb{R}^m$, $\exists B \in \mathcal{B}\left(\mathbb{R}^m\right) : P(\xi \in B) = 1$ и $h : \mathbb{R}^m \to \mathbb{R}^k$ непрерывна в каждой точке множества $B$. Тогда

1. $\xi_n \xrightarrow{a.s.} \xi \Rightarrow h\left(\xi_n\right) \xrightarrow{a.s.} h\left(\xi\right)$
2. $\xi_n \xrightarrow{P} \xi \Rightarrow h\left(\xi_n\right) \xrightarrow{P} h\left(\xi\right)$
3. $\xi_n \xrightarrow{d} \xi \Rightarrow h\left(\xi_n\right) \xrightarrow{d} h\left(\xi\right)$

**Доказательство.**

1. Так как $h$ непрерывна в каждой точке множества $B$, то справедливо
$$P\left(h\left(\xi_n\right) \to h\left(\xi\right)\right) \geqslant P\left(h\left(\xi_n\right) \to h\left(\xi\right),\ \xi \in B\right) \geqslant P\left(\xi_n \to \xi,\ \xi \in B\right) = 1,$$
так как
$$P\left(\overline{\xi_n \to \xi,\ \xi \in B}\right) \leqslant P\left(\overline{\xi_n \to \xi}\right) + P\left(\overline{\xi \in B}\right) = 0.$$

2. От противного. Пусть $h\left(\xi_n\right)$ не сходится по вероятности к $h\left(\xi\right)$. Тогда
$$\exists \varepsilon_0,\ \delta_0,\ \{n_k\}_{k=1}^{\infty} : P(\|h(\xi_{n_k}) - h(\xi)\|_2 > \varepsilon_0) \geqslant \delta_0.$$
Выберем из $\xi_{n_k}$ подпоследовательность, сходящуюся к $\xi$ почти наверное. Такая подпоследовательность получается как
$$\{n_k^1\} \subset \{n_k\} : \xi_{n_k^1}^{(1)} \xrightarrow{a.s.} \xi^{(1)}, \{n_k^2\} \subset \{n_k^1\} : \xi_{n_k^2}^{(2)} \xrightarrow{a.s.} \xi^{(2)}, \ldots, \{n_k^m\} \subset \{n_k^{m-1}\} : \xi_{n_k^m}^{(m)} \xrightarrow{a.s.} \xi^{(m)}.$$
Тогда $\xi_{n_k^m} \xrightarrow{a.s.} \xi \Rightarrow h(\xi_{n_k^m}) \xrightarrow{a.s.} h(\xi) \Rightarrow h(\xi_{n_k^m}) \xrightarrow{P} h(\xi)$, но $\{n_k^m\} \subset \{n_k\}$ по построению. Противоречие.

3. Докажем только для всюду непрерывных функций $h$. Пусть $f : \mathbb{R}^m \to \mathbb{R}$ – непрерывная ограниченная функция. Тогда $f \circ h$ – непрерывная ограниченная функция, и $E\left(f\left(h\left(\xi_n\right)\right)\right) \to E\left(f\left(h\left(\xi\right)\right)\right)$.

$\square$

# 2 Закон больших чисел, усиленный закон больших чисел и многомерная центральная предельная теорема для случайных векторов (б/д). Лемма Слуцкого. Пример применения леммы Слуцкого и его обобщение на многомерный случай (доказательство для одномерного случая).

**Утверждение 2.1.** (б/д) Пусть $\xi_n \xrightarrow{d} \xi$ в $\mathbb{R}^m$, $\eta_n \xrightarrow{d} C = const$ в $\mathbb{R}^s$. Тогда
$$\begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi \\ C \end{pmatrix} \in \mathbb{R}^{m+s}$$

**Следствие 2.1.1.** (Лемма Слуцкого) Пусть $\xi_n \xrightarrow{d} \xi,\ \eta_n \xrightarrow{d} C$ (все в $\mathbb{R}^m$). Тогда
$$\xi_n + \eta_n \xrightarrow{d} \xi + C,\ \xi_n \cdot \eta_n \xrightarrow{d} \xi \cdot C.$$

**Доказательство.** Пусть $f : \mathbb{R}^{2m} \to \mathbb{R},\ f(x, y) = x + y$ Тогда по теореме о наследовании сходимостей получаем требуемое. Аналогично для произведения. $\square$

**Пример 2.2.** Пусть $\xi_n \xrightarrow{d} \xi$ – последовательность случайных величин, $H(x)$ – дифференцируемая в точке $a$ функция, $\{b_n\} : b_n \xrightarrow{n\to\infty} 0,\ b_n \neq 0$. Тогда
$$\frac{H\left(a + \xi_n b_n\right) - H(a)}{b_n} \xrightarrow{d} H'(a)\xi.$$

**Доказательство.** Пусть

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

Функция $h$ непрерывна в нуле, и по лемме Слуцкого $\xi_n b_n \xrightarrow{d} 0$. Тогда по теореме о наследовании сходимостей $h\left(\xi_n b_n\right) \xrightarrow{d} h\left(0\right) = H'(a)$, и

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

. $\square$

**Утверждение 2.3.** (б/д) Пусть $\xi_n \xrightarrow{d} \xi$ в $\mathbb{R}^m$, $H(x) : \mathbb{R}^m \to \mathbb{R}^s$ – вектор-функция, у которой в точке $a$ есть матрица частных производных $H'(x) = \left\| \dfrac{\partial H_i(x)}{\partial x_j} \right\|_{i,j}$,
$b_n \to 0,\ b_n \neq 0$. Тогда

$$\frac{H(a+\xi_n b_n) - H(a)}{b_n} \xrightarrow{d} H'(a)\xi.$$

**Предельные теоремы для векторов**

Пусть $\{\xi_n\}_{n=1}^{\infty}, \xi$ – случайные векторы в $\mathbb{R}^m$, $S_n = \xi_1 + \ldots + \xi_n$.

**Теорема 2.4.** (ЗБЧ, б/д) Пусть $\{\xi_n\}_{n=1}^{\infty}$ попарно некоррелированные случайные векторы, и $\forall n \geqslant 1 \sup\limits_{1 \leqslant i \leqslant m} D\xi_n^{(i)} \leqslant c = const$. Тогда

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0.$$

**Теорема 2.5.** (УЗБЧ, б/д) Пусть $\{\xi_n\}_{n=1}^{\infty}$ – н.о.р.с.в., и $\forall i \in \{1, \ldots, m\}\ E\left|\xi_1^{(i)}\right| < \infty$. Тогда

$$\frac{S_n}{n} \xrightarrow{a.s.} E\xi_1.$$

**Теорема 2.6.** (ЦПТ, б/д) Пусть $\{\xi_n\}_{n=1}^{\infty}$ – н.о.р.с.в., и существует ковариационная матрица $D\xi_1$. Тогда

$$\sqrt{n}\left(\frac{S_n}{n} - E\xi_1\right) \xrightarrow{d} N(0, D\xi_1).$$

# 3 Вероятностно–статистическая модель. Понятия наблюдения и выборки. Моделирование выборки из неизвестного распределения. Основная задача математической статистики. Параметрическая статистическая модель.

Пусть $(\Omega, \mathcal{F})$ и $(E, \mathcal{E})$ – измеримые пространства.

**Определение 3.1.** Если $\xi : \Omega \to E$ таково, что $\xi^{-1}(B) \in \mathcal{F}\ \forall B \in \mathcal{E}$. Тогда $\xi$ называется *случайным элементом*. Если $(E, \mathcal{E}) = (\mathbb{R}^m, \mathcal{B}\left(\mathbb{R}^m\right))$, то $\xi$ – *случайный вектор*.

Пусть $(\Omega, \mathcal{F}, P)$ – вероятностное пространство.

**Определение 3.2.** *Распределением* случайного элемента $\xi$ называют меру $P_\xi$ на $\mathcal{E}$, такую что $P_\xi(B) = P(\xi \in B)$.

Пусть наблюдается некоторый эксперимент над $m$-мерным случайным вектором с распределением $P$. Нужно построить математическую модель эксперимента, то есть построить вероятностное пространство и случайный вектор над этим вероятностым пространством с распределением $P$.

**Определение 3.3.** Множество $\mathcal{X}$ всевозможных исходов эксперимента называется *выборочным пространством.*

Определим $\mathcal{B}_\mathcal{X}$ – некоторую $\sigma$-алгебру над выборочным множеством (обычно будем считать ее борелевской) и вероятностную меру $P$ на измеримом пространстве $(\mathcal{X}, \mathcal{B}_\mathcal{X})$.

Также, определим функцию $X(x) = x, \ \forall x \in \mathcal{X}$.

**Утверждение 3.4.** Функция $X : \mathcal{X} \to \mathcal{X}$ является *случайным элементом* на вероятностном пространстве $(\mathcal{X}, \mathcal{B}_\mathcal{X}, P)$ и имеет распределение $P_X = P$.

**Доказательство.**

$$\forall B \in \mathcal{B}_\mathcal{X} \ X^{-1}(B) = \{x \in \mathcal{X} : X(x) \in B\} = \{x \in \mathcal{X} : x \in B\} = B.$$

Значит, $X$ – случайный элемент, и

$$\forall B \in \mathcal{B}_\mathcal{X} \ P_X(B) = P(X \in B) = P(x \in \mathcal{X} : X(x) \in B) = P(x \in \mathcal{X} : x \in B) = P(B).$$

□

**Определение 3.5.** Фукнция $X$, определенная выше, называется *наблюдением.*

Построим модель $n$ независимых экспериментов. Пусть

$$\mathcal{X}^n = \mathcal{X} \times \ldots \times \mathcal{X},$$
$$\mathcal{B}_\mathcal{X}^n = \mathcal{B}\left(\mathcal{X}^n\right) = \sigma(B_1 \times \ldots \times B_n, B_i \in \mathcal{B}_\mathcal{X}),$$
$$P^n(B_1 \times \ldots \times B_n) = P(B_1) \cdot \ldots \cdot P(B_n).$$

Снова рассмотрим тождественное отображение $X : \mathcal{X}^n \to \mathcal{X}^n$. Также, определим функцию $X_i : \mathcal{X}^n \to \mathcal{X}, \ X_i(x_1, \ldots, x_n) = x_i$.

**Утверждение 3.6.** $X_i$ – случайная величина на вероятностном пространстве $(\mathcal{X}^n, \mathcal{B}_\mathcal{X}^n, P^n)$ с распределением $P$, а $X_1, \ldots, X_n$ – независимы в совокупности.

**Доказательство.**

$$\forall B \in \mathcal{B}_\mathcal{X} \ X^{-1}(B) = \{\overline{x} \in \mathcal{X}^n : X(\overline{x}) \in B\} = \{\overline{x} \in \mathcal{X}^n : x_i \in B\} =$$
$$\mathcal{X} \times \ldots \times \mathcal{X} \times B \times \mathcal{X} \times \ldots \times \mathcal{X} \in \mathcal{X}^n.$$

Зафиксируем $i \in \{1, \ldots, n\}$. Тогда

$$P^n(X_i \in B_i) = P^n\left(\{\overline{x} \in \mathcal{X}^n : X_i(\overline{x}) \in B_i\}\right) = P^n\left(\{\overline{x} \in \mathcal{X}^n : x_i \in B_i\}\right) =$$
$$P^n(\mathcal{X} \times \ldots \times \mathcal{X} \times B_i \times \mathcal{X} \times \ldots \mathcal{X}) = \ P(\mathcal{X}) \cdot \ldots \cdot P(\mathcal{X}) \cdot P(B_i) \cdot P(\mathcal{X}) \cdot \ldots \cdot P(\mathcal{X}) = P(B_i).$$

Докажем, что случайные величины независимы в совокупности.

$$P^n(X_1 \in B_1, \ldots, \ X_n \in B_n) = P^n\left(\overline{x} \in \mathcal{X}^n : X_1(\overline{x}) \in B_1, \ldots, \ X_n(\overline{x}) \in B_n\right) =$$
$$P^n\left(\overline{x} \in \mathcal{X}^n : x_1 \in B_1, \ldots, \ x_n \in B_n\right) = P^n(B_1 \times \ldots \times B_n) = P(B_1) \ldots P(B_n) =$$
$$P^n(X_1 \in B_1) \ldots P^n(X_n \in B_n).$$

□

**Определение 3.7.** Совокупность $X = (X_1, \dots, X_n)$ независимых одинаково распределенных случайных векторов с распределением $P$ называется выборкой размера $n$ с распределением $P$.

Часто нам придется рассматривать выборку при $n \to \infty$. Поэтому, введем следующие объекты:

$$\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \dots = \{(x_1, x_2, \dots) : x_i \in \mathcal{X}\},$$
$$\mathcal{B}_\mathcal{X}^\infty = \sigma\left(\{(x_1, x_2, \dots) : \exists n \in \mathbb{N},\ \exists B \in \mathcal{B}_\mathcal{X}^n \hookrightarrow (x_1, \dots, x_n) \in B\}\right),$$

$P^\infty$ – вероятностная мера на $(\mathcal{X}^\infty, \mathcal{B}_\mathcal{X}^\infty)$, такая что

$$P^\infty(B_1 \times B_2 \times \dots \times B_n \times \mathcal{X} \times \mathcal{X} \times \dots) = P(B_1)\dots P(B_n).$$

Снова определим $X : \mathcal{X}^\infty \to \mathcal{X}^\infty$ – тождественное отображение, $X_i(x_1, x_2, \dots) = x_i$. Тогда $X_1,\ X_2, \dots$ – независимы и одинаково распределены с распределением $P$.
В дальнейшем для простоты обозначений будем писать $(\mathcal{X},\ \mathcal{B}_\mathcal{X},\ P)$ вместо

$$(\mathcal{X}^n,\ \mathcal{B}_\mathcal{X}^n,\ P^n)$$

и

$$(\mathcal{X}^\infty,\ \mathcal{B}_\mathcal{X}^\infty,\ P^\infty)$$

Пусть $\mathcal{P}$ – семейство вероятностных мер на измеримом пространстве $(\mathcal{X}, \mathcal{B}_\mathcal{X})$.

**Определение 3.8.** Тройка $(\mathcal{X},\ \mathcal{B}_\mathcal{X},\ \mathcal{P})$ называется *вероятностно-статистической моделью.*

**Определение 3.9.** $X_1(\omega) = x_1,\ \dots,\ X_n(\omega) = x_n$ будем называть *реализацией выборки.*

**Замечание 3.10.** Задача статистики - сделать вывод о неизвестном распределении по реализции выборки

**Параметрическая модель**

**Определение 3.11.** Вероятностно-статистическая модель называется параметрической, если семейство распределений $\mathcal{P}$ параметризовано, т.е. $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.

**Пример 3.12.** $\mathcal{P} = \{exp(\theta),\ \theta > 0\}$ – семейство экспоненциальных распределений, параметризованных $\theta$.

# 4 Эмпирическое распределение и эмпирическая функция распределения. Теорема Гливенко–Кантелли.

Пусть $X_1, \dots,\ X_n$ – выборка из неизвестного распределения $P_X$ на $(\mathbb{R}^m,\ \mathcal{B}(\mathbb{R}^m))$, и $B \in \mathcal{B}(\mathbb{R}^m)$.

**Определение 4.1.** $P_n^*(B) = \dfrac{\sum_{i=1}^n \mathbb{I}_{x_i \in B}}{n}$ называется эмпирическим распределением, построенным по выборке $X_1, \dots, X_n$.

Нетрудно проверить, что эмпирическое распределение действительно является распределением.

**Определение 4.2.** Функция $F_n^*(x) = \dfrac{\sum_{i=1}^n \mathbb{I}_{x_i \leqslant x}}{n}$ называется эмпирической функцией распределения.

**Утверждение 4.3.** Пусть $X_1, \dots, X_n$ – выборка на вероятностном пространстве $(\Omega, \mathcal{F}, P)$ из распределения $P_X$. Тогда $\forall B \in \mathcal{B}(\mathbb{R}^m) \hookrightarrow\ P_n^*(B) \xrightarrow{\text{п.н.}} P_X(B)$.

**Доказательство.** $\mathbb{I}_{x_i \in B}$ – независимые (потому что являются функциями от $X_i$) одинаково распределенные случайные величины с конечным математическим ожиданием. Поэтому, по УЗБЧ

$$P_n^*(B) \xrightarrow{\text{п.н.}} E(\mathbb{I}_{x_i \in B}) = P(X_1 \in B) = P_X(B).$$

□

В дальнейшем будем считать размерность $m = 1$. Для $m > 1$ доказательства аналогичные, но более трудоемкие.

**Следствие 4.3.1.** $F_n^*(x) \xrightarrow{\text{п.н.}} F_{X_1}(x)$.

**Доказательство.** Достаточно взять в утверждении $B = (-\infty, x]$. □

Следующая теорема усиливает только что приведенное следствие:

**Теорема 4.4.** (Гливенко-Кантелли) Пусть $X_1, \ldots, X_n$ – независимые одинаково распределенные случайные величины с функцией распределения $F(x)$. Тогда $D_n(\omega) := \sup\limits_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow{\text{п.н.}} 0$.

**Доказательство.** Так как обе функции $F_n^*$ и $F$ непрерывны справа, то

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

Этот переход верен, поскольку в силу непрерывности справа значение в любой точке можно приблизить рациональными точками из правой окрестности.

Так как объединение счетного числа измеримых функций измеримо, то $D_n$ – случайная величина. Зафиксируем $N \in \mathbb{N}$ и положим $x_{N,k} = \inf \left\{ x \in \mathbb{R} : F(x) \geqslant \dfrac{k}{N} \right\}$.

Заметим, что $x_{N,0} = -\infty, x_{N,N} = +\infty$. $x_{N,k}$ конечно $\forall k \in \{1, \ldots, N-1\}$.

Рассмотрим произвольный $x \in [x_{N,k}, x_{N,k+1})$ и оценим значения функции в точке $x$ через значения в концах:

$$F_n^*(x) - F(x) \underbrace{\leqslant}_{\text{ф.р. неубывает}} F_n^*(x_{N,k+1} - 0) - F(x_{N,k}) = F_n^*(x_{N,k+1} - 0) - F(x_{N,k+1} - 0) +$$

$$+ \underbrace{F(x_{N,k+1} - 0)}_{\leqslant \frac{k+1}{N}} - \underbrace{F(x_{N,k})}_{\geqslant \frac{k}{N}} \leqslant F_n^*(x_{N,k+1} - 0) - F(x_{N,k+1} - 0) + \frac{k+1}{N} - \frac{k}{N} =$$

$$F_n^*(x_{N,k+1} - 0) - F(x_{N,k+1} - 0) + \frac{1}{N}.$$

С другой стороны,

$$F_n^*(x) - F(x) \geqslant F_n^*(x_{N,k}) - F(x_{N,k+1} - 0) = F_n^*(x_{N,k}) - F(x_{N,k}) + F(x_{N,k}) -$$

$$F(x_{N,k+1} - 0) \geqslant F_n^*(x_{N,k}) - F(x_{N,k}) + \frac{k}{N} - \frac{k+1}{N} = F_n^*(x_{N,k}) - F(x_{N,k}) - \frac{1}{N}.$$

Получили, что

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

В итоге

$$\sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \leqslant$$

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

Осталось доказать стремление к нулю правой части.

Так как (простое следствие УЗБЧ) $F_n^*(x) \xrightarrow{\text{п.н.}} F(x)$, $F_n^*(x-0) = \underbrace{P^*((-\infty, x))}_{\text{живет на } \mathbb{R}} \underbrace{\xrightarrow{\text{п.н.}}}_{\text{по мере на } \Omega} \underbrace{P_X((-\infty, x))}_{\text{живет на } \mathbb{R}} = F(x-0)$, то для достаточно большого $N \in \mathbb{N}$ и почти всех $\omega \in \Omega$ выполняется

$$\overline{\lim_n} \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \overset{\text{п.н.}}{<} \varepsilon \Rightarrow D_n(\omega) \xrightarrow{\text{п.н.}} 0.$$

□

# 5 Статистики и оценки. Примеры статистик: выборочные характеристики, порядковые статистики. Основные свойства оценок (несмещенность, состоятельность, сильная состоятельность, асимптотическая нормальность) и взаимосвязи между ними.

Пусть $(\mathcal{X},\ \mathcal{B}_{\mathcal{X}},\ \mathcal{P})$ – вероятностно-статистическая модель, $X$ – наблюдение, $(E,\ \mathcal{E})$ – измеримое пространство.

**Определение 5.1.** Измеримое отображение $S : \mathcal{X} \to E$ называется статистикой от наблюдения $X$.

**Определение 5.2.** Пусть $X$ – наблюдение в параметрической модели $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \{\mathcal{P}_\theta\}_{\theta\in\Theta})$ и $S(X)$ – статистика со значением в $\Theta$ (или в $\tau(\Theta)$). Тогда $S(X)$ называется оценкой неизвестного параметра $\theta$ (или $\tau(\theta)$).

**Пример 5.3.** Пусть $X = (X_1, \dots, X_n)$ – выборка из распределения в $\mathbb{R}^m$.

1. Если $g(x)$ – борелевская функция, то $\overline{g(x)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} g(X_i)$ называется выборочной характеристикой функции $g$.
2. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ – выборочное среднее.
3. Для $m = 1$ $\overline{X^k} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^k$ – выборочный момент $k$-го порядка.

**Замечание 5.4.** $Eg(x) = \int g(x)dF(x) \int g(x)dF_n(x) = \frac{1}{n}\sum g(x_i) = \overline{g(x)}$

Теоретическому среднему соответствует эмпирическое среднее.

Композиция измеримых функций измерима, поэтому можно рассматривать *функции от выборочных характеристик*

$$S(X) = \ h\left(\overline{g_1(X)}, \dots,\ \overline{g_k(X)}\right),$$

где $h$ – борелевская функция.

**Пример 5.5.** Все для $m = 1$

1. $S^2 = \overline{X^2} - (\overline{X})^2$ – выборочная дисперсия

2. $M_k = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$ – выборочный центральный момент $k$-го порядка.

**Замечание 5.6.** Верны несколько утверждений:

1. $S^2 = \dfrac{1}{n}\sum(X_i - \overline{X})^2 = M_2$
2. $ES^2 = \frac{n-1}{n}DX_1$
3. $S_0^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$ – исправленная выборочная дисперсия. В отличие от обычной выборочной дисперии, является несмещенной оценкой. Собственно, поэтому и исправленная.

**Определение 5.7.** Для $m = 1$ определим

- $X_{(1)} = \min(X_1, \ldots,\ X_n)$,
- $X_{(2)} = \min(\{X_1, \ldots,\ X_n\} \backslash \{X_{(1)}\})$, то есть второй в порядке неубывания элемент выборки,
- $\ldots$,
- $X_{(n)} = \max(X_1, \ldots,\ X_n)$

Тогда $X_{(k)}$ называется $k$-ой порядковой статистикой выборки $X$, а $(X_{(1)}, \ldots X_{(n)})$ называется вариационным рядом.

**Свойства оценок**

Пусть $X = (X_1, \ldots, X_n)$ – выборка из неизвестного распределения $P \in \{P_\theta,\ \theta \in \Theta\}$, где $\Theta \subset \mathbb{R}^k$.

**Определение 5.8.** Оценка $\theta^*(X)$ называется несмещенной оценкой параметра $\theta$, если $\forall\theta \in \Theta \hookrightarrow E_\theta\theta^*(X) = \theta$, где $E_\theta$ – математическое ожидание в случае, когда элементы выборки имеют распределение $P_\theta$.

**Замечание 5.9.** В этом определении смысл квантора $\forall$ такой. Мы бы хотели, чтобы для нашего неизвестного $\theta$ условие выполнялось. Но мы его не знаем, поэтому требуем более строгое условие – чтобы выполнялось для всех. Тогда и для нашего точно будет.

**Пример 5.10.** $\{N(\theta, 1),\ \theta \in \mathbb{R}\}$, тогда $\overline{X}$ и $X_1$ являются несмещенными оценками параметра $\theta$.

К тому же, наш пример (а именно оценка $X_1$) показывает, что несмещенная оценка может быть довольно бестолковой. То есть это условие желательно, но его явно недостаточно.

Пусть $X_1, \ldots,\ X_n, \ldots$ – выборка неограниченного размера из распределения $P \in \{P_\theta,\ \theta \in \Theta\},\ \Theta \subset \mathbb{R}^k$.

**Определение 5.11.** Оценка $\theta_n^*(X_1, \ldots, X_n)$ (а точнее, последовательность оценок) называется состоятельной, если $\forall\theta \in \Theta \hookrightarrow \theta_n^* \xrightarrow{P_\theta} \theta$, т.е.

$$\forall\varepsilon > 0\ \lim_{n\to\infty} P_\theta\left(\|\theta_n^*(X) - \theta\|_2 > \varepsilon\right) = 0.$$

**Определение 5.12.** $\theta_n^*(X_1, \ldots, X_n)$ называется сильно состоятельной оценкой параметра $\theta$, если $\theta_n^* \xrightarrow{P_\theta-\text{п.н.}} \theta$, т.е.

$$\forall\theta \in \Theta\ P_\theta\left(\theta_n^*(X) \to \theta\right) = 1\ \forall\theta \in \Theta.$$

**Пример 5.13.** $\{N(\theta,\ 1),\ \theta \in \mathbb{R}\}$, $\overline{X}$ – сильно состоятельная оценка параметра $\theta$ по УЗБЧ.

Пусть $X_i$ – случайные величины образуют выборку.

**Определение 5.14.** Оценка $\theta_n^*(X_1, \ldots, X_n)$ называется асимптотически нормальной оценкой параметра $\theta$, если

$$\forall \theta \in \Theta \ \sqrt{n}\left(\theta_n^*(X) - \theta\right) \xrightarrow{d_\theta} N\left(0,\ \sigma^2(\theta)\right),$$

где $\sigma^2(\theta)$ называется асимптотической дисперсией.

**Замечание 5.15.** $d_\theta$ означает, что элементы выборки имеют распределение $P_\theta$

**Замечание 5.16.** В многомерном случае вместо асимптотической дисперсии появляется асимптотическая матрица ковариаций $\Sigma(\theta)$.

**Пример 5.17.** $\{N(\theta,\ 1),\ \theta \in \mathbb{R}\},\ \overline{X}$ является асимптотически нормальной оценкой параметра $\theta$ по ЦПТ.

**Замечание 5.18.** Свойства состоятельности, сильной состоятельности и асимптотической нормальности называются асимптотическими свойствами и имеют смысл только при растущем объеме выборки.

**Замечание 5.19.** Также, можно оценивать $\tau(\theta)$. Все свойства оценок определяются аналогично.

**Теорема 5.20** (Связь между свойствами оценок)**.**

1. $\theta_n^*$ – сильно состоятельная оценка параметра $\tau(\theta) \Rightarrow \theta_n^*$ – состоятельная оценка параметра $\tau(\theta)$.
2. $\theta_n^*$ – асимптотически нормальная оценка параметра $\tau(\theta) \Rightarrow \theta_n^*$ – состоятельная оценка параметра $\tau(\theta)$.
3. Никакие другие импликации неверны.

**Доказательство.**

1. $\theta_n^* \xrightarrow{P_\theta-\text{п.н.}} \theta \Rightarrow \theta_n^* \xrightarrow{P_\theta} \theta$.
2. Так как $\dfrac{1}{\sqrt{n}} \xrightarrow{P_\theta} 0$, то по лемме Слуцкого

$$\frac{1}{\sqrt{n}}\sqrt{n}\left(\theta_n^*(X) - \tau(\theta)\right) \xrightarrow{d_\theta} \xi \cdot 0,\ \forall \theta \in \Theta,$$

где $\xi \sim N\left(0,\ \sigma^2(\theta)\right)$. Тогда $\theta_n^*(X) - \tau(\theta) \xrightarrow{d_\theta} 0 \Rightarrow \theta_n^*(X) - \tau(\theta) \xrightarrow{P_\theta} 0 \Rightarrow \theta_n^*(X) \xrightarrow{P_\theta} \tau(\theta)$.

□

**Пример 5.21.** Из несмещенности не следуют остальные свойства, т.к. можно взять распределение $Bern$ и $X_{(1)}$.

Также можно привести пример, когда есть все, кроме несмещенности. Возьмем хорошую оценку, например, $\overline{X}$. Дальше ее нужно сместить. Например, добавить $\frac{1}{n}$. Тогда все свойства сохранятся, кроме несмещенности.

# 6 Наследование состоятельности и сильной состоятельности при взятии непрерывной функции. Лемма о наследовании асимптотической нормальности.

**Утверждение 6.1.** Пусть $\theta_n^*(X)$ – (сильно) состоятельная оценка $\theta$, $\tau(\theta) : \mathbb{R}^k \to \mathbb{R}^s$ – непрерывная функция. Тогда $\tau(\theta^*)$ – (сильно) состоятельная оценка $\tau(\theta)$.

**Доказательство.** $\theta_n^* \xrightarrow{P_\theta - a.s.} \theta$. По теореме о наследовании сходимостей $\tau(\theta_n^*) \xrightarrow{P_\theta - a.s.} \tau(\theta)$. $\theta_n^* \xrightarrow{P_\theta} \theta$. По теореме о наследовании сходимостей $\tau(\theta_n^*) \xrightarrow{P_\theta} \tau(\theta)$. $\square$

**Лемма 6.2.** (о наследовании асимптотической нормальности) Пусть $\theta_n^*$ – асимптотически нормальная оценка $\theta \in \Theta \subset \mathbb{R}$ с асимптотической дисперсией $\sigma^2(\theta)$, $\tau(\theta) : \mathbb{R} \to \mathbb{R}$ дифференцируема $\forall \theta \in \Theta$. Тогда $\tau(\theta_n^*(X))$ – асимптотически нормальная оценка для $\tau(\theta)$ с асимптотической дисперсией $\sigma^2(\theta) \cdot (\tau'(\theta))^2$.

**Доказательство.** Из определения асимптотической нормальности $\sqrt{n}\,(\theta_n^*(X) - \theta) \xrightarrow{d_\theta} \xi \sim N\left(0,\ \sigma^2(\theta)\right)$. Обозначим $\xi_n = \sqrt{n}\,(\theta_n^*(X) - \theta)\,,\ b_n = \dfrac{1}{\sqrt{n}}$. Тогда по следствию из леммы Слуцкого

$$\frac{\tau(\theta + \xi_n b_n) - \tau(\theta)}{b_n} = \sqrt{n}\,(\tau(\theta_n^*) - \tau(\theta)) \xrightarrow{d_\theta} \xi \cdot \tau'(\theta).$$

$\square$

**Упражнение 6.3.** Показать, что несмещенность оценки не наследуется.

**Утверждение 6.4.** (о наследовании асимптотической нормальности в многомерном случае, б/д) Пусть $\Theta \subset \mathbb{R}^k$, $\tau : \mathbb{R}^k \to \mathbb{R}^s$ – дифференцируема в точках $\Theta$, $\theta_n^*$ – асимптотически нормальная оценка для $\theta$ с асимптотической ковариационной матрицей $\Sigma(\theta)$. Тогда $\tau(\theta_n^*)$ – асимптотически нормальная оценка для $\tau(\theta)$ с асимптотической ковариационной матрицей $\left(\dfrac{\partial \tau}{\partial \theta}\right) \cdot \Sigma(\theta) \cdot \left(\dfrac{\partial \tau}{\partial \theta}\right)^T$.

# 7 Метод подстановки и метод моментов, их связь. Состоятельность и асимптотическая нормальность оценки метода моментов.

## Метод подстановки

**Определение 7.1.** Пусть в параметрическом семействе $\{P_\theta : \theta \in \Theta\}$ для некоторого функционала $G$ выполнено $\forall \theta \in \Theta \hookrightarrow \ \theta = G(P_\theta)$. Тогда оценкой по методу подстановки называется $\theta^*(X_1, \dots, X_n) = G(P_n^*)$, где $P_n^*$ – эмпирическое распределение по выборке.

**Пример 7.2.** $Bern(\theta)$, $N(\theta, 1)$. Пусть $G = \displaystyle\int_{\mathbb{R}} x dP$, $\theta^* = \displaystyle\int_{\mathbb{R}} x dF_n^* = \dfrac{\sum_{i=1}^n X_i}{n} = \overline{X}$.

## Метод моментов

Пусть $X_1, \dots, X_n$ – выборка из распределения $P \in \{P_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$. Рассмотрим борелевские функции $g_1(X),\ \dots,\ g_k(X)$ со значениями в $\mathbb{R}$. Положим $m_i(\theta) = E_\theta g_i(X_1)$, где $m_i(\theta)$ конечны $\forall i : 1 \leqslant i \leqslant k,\ \forall \theta \in \Theta$. Также, обозначим

$$m(\theta) = \begin{pmatrix} m_1(\theta) \\ \dots \\ m_k(\theta) \end{pmatrix},\ \overline{g} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n g_1(X_i) \\ \dots \\ \frac{1}{n}\sum_{i=1}^n g_k(X_i) \end{pmatrix}.$$

**Определение 7.3.** Если существует и притом единственное решение системы

$$\begin{cases} m_1(\theta) = \overline{g_1(X)} \\ \dots \\ m_k(\theta) = \overline{g_k(X)} \end{cases},$$

то его решение $\theta^* = m^{-1}(\overline{g})$ называется оценкой по методу моментов.

**Определение 7.4.** Функции $g_i(x) = X^i$ называются пробными.

**Замечание 7.5.** В случае, когда $\overline{g} \notin m(\Theta)$, можно для нахождения $m^{-1}(g)$ взять ближайшую к $\overline{g}$ точку из $m(\Theta)$.

**Теорема 7.6.** (сильная состоятельности оценки по методу моментов) Пусть $m : \Theta \to m(\Theta)$ – биекция, и функцию $m^{-1}$ можно доопределить до функции, заданной на всем $\mathbb{R}^k$, и непрерывной в каждой точке множества $m(\Theta)$. Также, $E_\theta g_i(X_1) < \infty$ $\forall i : 1 \leqslant i \leqslant k$, $\forall \theta \in \Theta$. Тогда оценка по методу моментов является сильно состоятельной оценкой параметра $\theta$.

**Доказательство.** Фиксируем $\theta$. По УЗБЧ $\overline{g} \xrightarrow{P_\theta - a.s.} m(\theta)$. Тогда, по теореме о наследовании сходимостей $\theta_n^* = m^{-1}(\overline{g}) \xrightarrow{P_\theta - a.s.} m^{-1}(m(\theta)) = \theta$. □

**Теорема 7.7.** (асимптотическая нормальность оценки по методу моментов) Если в условиях предыдущей теоремы функция $m^{-1}$, доопределенная на $\mathbb{R}^k$, дифференцируема на $m(\Theta)$, и $E_\theta g_i^2(X_1) < \infty$ $\forall i : 1 \leqslant i \leqslant k$, то оценка, полученная по методу моментов, асимптотически нормальна.

**Доказательство.** $\sqrt{n}(\overline{g} - m(\theta)) \xrightarrow{d_\theta} N(0,\, \Sigma(\theta))$, где $\Sigma(\theta)$ – матрица ковариаций вектора $(g_1(X_1), \dots, g_k(X_1))^T$. Далее применяем утверждение о наследовании асимптотической нормальности. □

**Замечание 7.8.** Метод моментов – частный случай метода подстановки:

$$\theta = m^{-1} \begin{pmatrix} \int_{\mathcal{X}^n} g_1(x) dP_\theta(x) \\ \dots \\ \int_{\mathcal{X}^n} g_k(x) dP_\theta(x) \end{pmatrix} \Rightarrow \theta_n^* = m^{-1} \begin{pmatrix} \int_{\mathcal{X}^n} g_1(x) dP^*(x) \\ \dots \\ \int_{\mathcal{X}^n} g_k(x) dP^*(x) \end{pmatrix} = G\left(P_n^*\right).$$

# 8 Квантили и выборочные квантили. Теорема об асимптотической нормальности выборочной квантили. Теорема о выборочной медиане (б/д).

## Метод выборочных квантилей

Пусть $P$ – распределение вероятностей на $\mathbb{R}$, $F(x)$ – его функция распределения.

**Определение 8.1.** $p$-квантилью распределения $P$ называется $z_p = \inf\{x : F(x) \geqslant p\}$, $p \in (0, 1)$.

**Замечание 8.2.** Если $F$ непрерывна, то существует точное решение $F(z_p) = p$. Если к тому же $F$ строго монотонна, то решение единственно.

**Определение 8.3.** Пусть $X_1, \dots, X_n$ – выборка. Статистика $z_{n,p} = \begin{cases} X_{([np]+1)}, & np \notin \mathbb{Z} \\ X_{(np)}, & np \in \mathbb{Z} \end{cases}$ называется выборочным $p$-квантилем.

**Замечание 8.4.** $z_{n,p}$ – это, по сути, $p$-квантиль эмпирического распределения $P_n^*$.

**Теорема 8.5.** (о выборочном квантиле) Пусть $X_1, \ldots,\ X_n$ – выборка из распределения $P$ с плотностью $f(x)$. Пусть $z_p$ – $p$-квантиль распределения $P$, причем $f(x)$ непрерывно дифференцируема в окрестности $z_p$ и $f(z_p) > 0$. Тогда

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

**Доказательство.**

$$F_{X_{(k)}}(x) = P(X_{(k)} \leqslant x) = \sum_{m=k}^{n} C_n^m F^m(x)(1-F(x))^{n-m} \Rightarrow f_{X_{(k)}}(x) =$$
$$nC_{n-1}^{k-1}F^{k-1}(x)(1-F(x))^{n-k}f(x).$$

Формально эта плотность получается из свойств сочетаний. Интуитивно: это вероятность попасть в какую-то маленькую окрестность $x$ - $n$ способов выбрать точку, которая попадет, потом выбираем $k-1$ точку, которые будут до нее, а остальные после. Ну и $f(x)$ собственно вероятность конкретной точки попасть в эту окрестность

Обозначим $\eta_n = \sqrt{n}(z_{n,p} - z_p)\cdot\sqrt{\dfrac{f^2(z_p)}{p(1-p)}} = (z_{n,p} - z_p)\cdot\sqrt{\dfrac{nf^2(z_p)}{p(1-p)}},\ k = \lceil np \rceil$. Нужно показать, что $\eta_n \xrightarrow{d_\theta} N(0,1)$.

$$F_{\eta_n}(x) = P(\eta_n \leqslant x) = P\left((z_{n,p} - z_p)\cdot\sqrt{\frac{nf^2(z_p)}{p(1-p)}} \leqslant x\right) =$$
$$P\left((X_{(k)} - z_p)\cdot\sqrt{\frac{nf^2(z_p)}{p(1-p)}} \leqslant x\right) \overset{f(z_p)\,>\,0}{=} P\left(X_{(k)} \leqslant x\sqrt{\frac{p(1-p)}{nf^2(z_p)}} + z_p\right) =$$
$$F_{X_{(k)}}\left(x\sqrt{\frac{p(1-p)}{nf^2(z_p)}} + z_p\right).$$

Тогда

$$f_{\eta_n}(x) = \sqrt{\frac{p(1-p)}{nf^2(z_p)}} \cdot f_{X_{(k)}}\left(x\sqrt{\frac{p(1-p)}{nf^2(z_p)}} + z_p\right).$$

Положим

$$t_n(x) := x\sqrt{\frac{p(1-p)}{nf^2(z_p)}} + z_p,\ q := 1-p.$$

Перепишем формулу плотности для $\eta_n$ в виде

$$f_{\eta_n} = \sqrt{\frac{p(1-p)}{nf^2(z_p)}} f_{X_{(k)}}(t_n) = \sqrt{\frac{p(1-p)}{nf^2(z_p)}}\cdot nC_{n-1}^{k-1}F^{k-1}(t_n)(1-F(t_n))^{n-k}f(t_n) =$$
$$\left[\sqrt{npq}\cdot C_{n-1}^{k-1}p^{k-1}q^{n-k}\right]\cdot\left[\frac{f(t_n)}{f(z_p)}\right]\cdot\left[\left(\frac{F(t_n)}{p}\right)^{k-1}\left(\frac{1-F(t_n)}{q}\right)^{n-k}\right] := A_1(n)\cdot A_2(n)\cdot A_3(n).$$

Тогда $A_1 \to \frac{1}{\sqrt{2\pi}}$ по формуле Стирлинга (на лекции было дано как упражнение)

$$A_1(n) = \sqrt{npq}\cdot C_{n-1}^{k-1}p^{k-1}q^{n-k} \;= \sqrt{npq}\frac{(n-1)!}{(k-1)!\,(n-k)!}p^{k-1}q^{n-k} =$$

$$\sqrt{npq}\frac{\sqrt{2\pi\cdot(n-1)}\left(\frac{n-1}{e}\right)^{n-1}(1+o(1))}{\sqrt{2\pi\cdot(k-1)}\left(\frac{k-1}{e}\right)^{k-1}(1+o(1))\sqrt{2\pi\cdot(n-k)}\left(\frac{n-k}{e}\right)^{n-k}(1+o(1))}p^{k-1}q^{n-k} \sim$$

$$\sqrt{\frac{npq}{2\pi}}\sqrt{\frac{n-1}{(k-1)(n-k)}}\frac{(n-1)^{n-1}}{(k-1)^{k-1}(n-k)^{n-k}}p^{k-1}q^{n-k} \sim$$

$$\sqrt{\frac{npq}{2\pi}}\sqrt{\frac{n-1}{(np-1)(n-np)}}\frac{(n-1)^{n-1}}{(np-1)^{np-1}(n-np)^{n-np}}p^{np-1}q^{n-np} =$$

$$\sqrt{\frac{npq}{2\pi}}\sqrt{\frac{n-1}{(np-1)(n-np)}}\frac{(n-1)^{n-1}}{(np-1)^{np-1}(n-np)^{n-np}}p^{np-1}q^{n-np} =$$

$$\sqrt{\frac{npq}{2\pi}}\sqrt{\frac{n\left(1-\frac{1}{n}\right)}{n^2\left(p-\frac{1}{n}\right)(1-p)}}\frac{n^{n-1}\left(1-\frac{1}{n}\right)^{n-1}}{(np)^{np-1}\left(1-\frac{1}{np}\right)^{np-1}n^{n-np}q^{n-np}}p^{np-1}q^{n-np} =$$

$$\sqrt{\frac{pq}{2\pi}}\sqrt{\frac{\left(1-\frac{1}{n}\right)}{\left(p-\frac{1}{n}\right)q}}\frac{\left(1-\frac{1}{n}\right)^{n-1}}{\left(1-\frac{1}{np}\right)^{np-1}} \;\sim\; \sqrt{\frac{pq}{2\pi}}\sqrt{\frac{1}{pq}}\frac{\exp\left\{-\frac{1}{n}(n-1)\right\}}{\exp\left\{-\frac{1}{np}(np-1)\right\}} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}.$$

$t_n \xrightarrow[n\to\infty]{} z_p$, и так как $f(x)$ непрерывна в точке $z_p$, то $A_2(n) = \dfrac{f(t_n)}{f(z_p)} \xrightarrow[n\to\infty]{} 1$. Разложим $F$ в ряд Тейлора в окрестности точки $z_p$:

$$F(t_n) = F(z_p) + \frac{\partial F}{\partial t_n}(z_p)(t_n - z_p) + \frac{1}{2}\frac{\partial^2 F}{\partial t_n^2}(z_p)(t_n - z_p)^2 + o\left((t_n - z_p)^2\right) =$$

$$p + \frac{x}{f(z_p)}\sqrt{\frac{pq}{n}}f(z_p) + \frac{1}{2}\cdot\frac{x^2pq}{n}\cdot\frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right) \Rightarrow$$

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

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

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

$$(n-k)\cdot\left[-x\sqrt{\frac{p}{nq}} - \frac{1}{2}\cdot\frac{x^2p}{n}\cdot\frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right) - \frac{x^2p}{2qn}\right] \sim$$

$$np\left[x\sqrt{\frac{q}{np}} + \frac{1}{2}\cdot\frac{x^2q}{n}\cdot\frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right) - \frac{x^2q}{2pn}\right] + nq\left[-x\sqrt{\frac{p}{nq}} - \frac{1}{2}\cdot\frac{x^2p}{n}\cdot\frac{f'(z_p)}{f^2(z_p)} + o\left(\frac{1}{n}\right) - \frac{x^2p}{2qn}\right] =$$

$$o(1) - \frac{x^2q}{2} - \frac{x^2p}{2} = o(1) - \frac{x^2}{2} \xrightarrow[n\to\infty]{} -\frac{x^2}{2}.$$

Таким образом, $f_{\eta_n}(x) \xrightarrow[n\to\infty]{} \dfrac{1}{\sqrt{2\pi}}\exp\left\{-\dfrac{x^2}{2}\right\}$ равномерно на каждом отрезке $[-N, N]$, т.к. каждую из предыдущих сходимостей можно мажорировать сходящейся последовательностью, не зависящей от $x$. Пусть $\varepsilon > 0$, $N_\varepsilon : P(|\xi| > N_\varepsilon) \leqslant \varepsilon$, где $\xi \sim N(0,1)$, $g : \mathbb{R} \to \mathbb{R}$ – непрерывная ограниченная функция. Тогда $|g(x)|\leqslant C\ \forall x \in \mathbb{R}$, и

$$|Eg(\eta_n) - Eg(\xi)|\leqslant \left|\int_{-N_\varepsilon}^{N_\varepsilon} g(x)(p_{\eta_n}(x) - p_\xi(x))dx\right| + (P(|\eta_n|> N_\varepsilon) + P(|\xi|> N_\varepsilon))\cdot C \xrightarrow{n\to\infty}$$
$$2C\cdot P(|\xi|> N_\varepsilon) \leqslant 2C\cdot \varepsilon \Rightarrow Eg(\eta_n) \xrightarrow{n\to\infty} Eg(\xi).$$

□

**Определение 8.6.** Медианой распределения $P$ называется его 0.5-квантиль.

**Определение 8.7.** Выборочной медианой называется

$$\hat{\mu} = \begin{cases} X_{(k+1)}, & n = 2k+1 \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}, & n = 2k \end{cases}.$$

**Теорема 8.8.** (о выборочной медиане, б/д) В условиях теоремы о выборочном квантиле

$$\sqrt{n}(\hat{\mu} - z_{0.5}) \xrightarrow{d_\theta} N\left(0,\ \frac{1}{4f^2(z_{0.5})}\right).$$

**Замечание 8.9.** В условиях соответствующих теорем $z_{n,p} \xrightarrow{P} z_p$ и $\hat{\mu} \xrightarrow{P} z_{0.5}$.

# 9 Сравнение оценок, функция потерь и функция риска. Подходы к сравнению оценок: равномерный, байесовский, минимаксный, асимптотический.

**Определение 9.1.** Борелевская неотрицательная функция двух переменных $g(x,\ y)$ называется *функцией потерь*. Если $\theta^*$ – оценка $\theta$, то $g\left(\theta^*, \theta\right)$ называется *величиной потерь*.

**Пример 9.2.**

1. $\rho(x,\ y) = |x-y|$,
2. $\rho(x,\ y) = (x-y)^2$ – квадратичная функция потерь,
3. если $A$ – неотрицательно определенная матрица, $\theta$ – многомерный параметр, $\theta \in \Theta \subset \mathbb{R}^k$, то $g(x,\ y) = \langle A(x-y)\ , x-y\rangle$.

**Определение 9.3.** Если задана функция потерь $g$, то *функцией риска* оценки $\theta^*$ называется величина $R\left(\theta^*,\ \theta\right) = E_\theta g\left(\theta^*(X),\ \theta\right)$.

## Равномерный подход

**Определение 9.4.** Пусть $\theta^*,\ \hat{\theta}$ – оценки параметра $\theta$. Говорят, что оценка $\theta^*$ *лучше* оценки $\hat{\theta}$ (в равномерном подходе), если $\forall \theta \in \Theta \hookrightarrow R\left(\theta^*, \theta\right) \leqslant R(\hat{\theta}, \theta)$, и $\exists \theta \in \Theta : R\left(\theta^*, \theta\right) < R(\hat{\theta}, \theta)$.

**Замечание 9.5.** Будем придерживаться более слабого определения, т.е. оценка $\theta^*$ называется лучше оценки $\hat{\theta}$, если $\forall \theta \in \Theta \hookrightarrow R\left(\theta^*, \theta\right) \leqslant R(\hat{\theta}, \theta)$.

**Определение 9.6.** Оценка $\theta^*$ называется наилучшей оценкой параметра $\theta$ в классе оценок $K$, если она лучше любой другой оценки из $K$.

**Замечание 9.7.** Наилучшая оценка существует не всегда.

**Пример 9.8.** Пусть $g(x, y) = (x - y)^2$, $K$ – класс всевозможных оценок параметра $\theta$. Тогда $\hat{\theta}_0(X) = \theta_0 \in \Theta$ удовлетворяет условию $R(\hat{\theta}_0, \theta_0) = 0$, следовательно, наилучшая оценка $\theta^*$ должна удовлетворять соотношению $R(\theta^*, \theta_0) \leqslant R(\hat{\theta}_0, \theta_0)$, т.е. $R(\theta^*, \theta) = 0\ \forall \theta \in \Theta$. В общем случае это приводит к противоречию.

**Пример 9.9.** Пусть $U\left[\theta, \theta + \frac{1}{2}\right]$, $\theta \in \mathbb{N}$. Тогда оценка $[X_1]$ является наилучшей.

## Минимаксный подход

**Определение 9.10.** Оценка $\theta^*$ параметра $\theta$ называется наилучшей в минимаксном подходе, если $\sup\limits_{\theta \in \Theta} R(\theta^*, \theta) = \inf\limits_{\hat{\theta}} \sup\limits_{\theta \in \Theta} R(\hat{\theta}, \theta)$, т.е. $\theta^*$ достигает наименьший максимум функции риска.

## Байесовский подход

Предположим, что на множестве $\Theta$ задано априорное распределение вероятностей $Q$, и $\theta$ выбирается в соответствии с распределением $Q$ из $\Theta$. Если $\hat{\theta}(X)$ – оценка $\theta$ и $R(\hat{\theta}, \theta)$ – ее функция риска, то положим $R(\hat{\theta}(X)) = E_Q(R(\hat{\theta}(X), \theta)) = \int_\Theta R(\hat{\theta}(X), t) Q(dt)$.

**Определение 9.11.** Оценка $\theta^*(X)$ называется наилучшей в байесовском подходе, если $R(\theta^*(X)) = \min\limits_{\hat{\theta}} R(\hat{\theta}(X))$.

## Асимптотический подход

Пусть $\hat{\theta}_1$, $\hat{\theta}_2$ – асимптотически нормальные оценки с асимптотическими дисперсиями $\sigma_1^2(\theta)$, $\sigma_2^2(\theta)$ соответственно.

**Определение 9.12.** Оценка $\hat{\theta}_1$ лучше оценки $\hat{\theta}_2$ в асимптотическом подходе, если $\forall \theta \in \Theta \hookrightarrow \sigma_1^2(\theta) \leqslant \sigma_2^2(\theta)$ и существует $\theta \in \Theta$, для которого неравенство строгое.

**Пример 9.13.** Пусть $X_1, \ldots, X_n$ – выборка из распределения $N(\theta, 1)$. Тогда по ЦПТ

$$\sqrt{n}(\overline{X} - \theta) \xrightarrow{d_\theta} N(0, 1), \tag{1}$$

а по теореме о выборочной медиане

$$\sqrt{n}(\hat{\mu} - \theta) \xrightarrow{d_\theta} N\left(0, \frac{\pi}{2}\right). \tag{2}$$

Следовательно, оценка $\overline{X}$ лучше оценки $\hat{\mu}$ в асимптотическом подходе.

**Определение 9.14.** Оценка $\hat{\theta}$ называется наилучшей оценкой в классе оценок $K$ (в асимптотическом подходе), если она лучше любой другой оценки из $K$.

# 10 Понятие плотности в дискретном случае. Неравенство Рао–Крамера и эффективные оценки. Критерий эффективности оценки.

**Определение 10.1.** Считающей мерой на $\mathbb{Z}$ называется функция $\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{Z}_+ \cup \{+\infty\}$, определяемая по правилу $\mu(B) = \sum\limits_{k \in \mathbb{Z}} I_{k \in B}$.

**Определение 10.2.** Интегралом от функции $f$ по считающей мере $\mu$ называется $\int_{\mathbb{R}} f(x)\mu(dx) = \sum\limits_{k\in\mathbb{Z}} f(k)$, если ряд сходится абсолютно.

Аналогично определяется интеграл по любому борелевскому множеству $B$: $\int_B f(x)\mu(dx) = \sum\limits_{k\in\mathbb{Z}\cap B} f(k)$.

**Замечание 10.3.** Это интеграл Лебега, поэтому все свойства интеграла Лебега переносятся на интеграл по считающей мере.

**Замечание 10.4.** Подобный интеграл можно определить в $\mathbb{R}^n$, используя считающую меру в $\mathbb{Z}^n$.

**Определение 10.5.** Пусть $\xi$ – дискретная случайная величина, принимающая значения в $\mathbb{Z}$. Ее плотностью относительно считающей меры называется

$$p(x) = P(\xi = x),\ x \in \mathbb{Z}.$$

**Пример 10.6.** $\xi \ \sim\ Bin(n,\, p)$.

$$p(k) = P(\xi = k) = C_n^k p^k (1-p)^{n-k} \mathbb{I}_{k\in\{0,1,\dots,n\}}.$$

Если $\xi$ дискретная случайная величина с плотностью $p(x)$, то для любой борелевсккой функции $g$ выполняется

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

**Замечание 10.7.** Всюду далее, говоря о плотности, будем подразумевать либо привычную плотность распределения, либо плотность дискретной случайной величины относительно считающей меры на $\mathbb{Z}$ (или на $\mathbb{Z}^n$).

**Определение 10.8.** Среднеквадратичный подход – это равномерный подход с квадратичной функцией потерь.

**Упражнение 10.9.** Пусть $K$ – класс несмещенных оценок параметра $\tau(\theta),\ \tau : \Theta \to \mathbb{R}$, и $T_1,\, T_2 \in K$. Если $\forall\theta \in \Theta \hookrightarrow E_\theta(T_i - \tau(\theta))^2 = \inf\limits_{T\in K} E_\theta(T - \tau(\theta))^2 < \infty$, то $T_1 = T_2$ $P_\theta$-п.н. $\forall\theta \in \Theta$.

**Определение 10.10.** Пусть $X$ – наблюдение с неизвестным распределением $P \in \{P_\theta,\ \theta \in \Theta\}$, и $\forall\theta \in \Theta$ $P_\theta$ имеет плотность $p_\theta(x)$ по одной и той же мере $\mu$ (мера Лебега, либо считающая). В этом случае семейство $\{P_\theta,\ \theta \in \Theta\}$ называется доминируемым относительно $\mu$.

**Определение 10.11.** Случайная величина $u_\theta(X) = \dfrac{\partial}{\partial\theta} \ln p_\theta(X)$ называется вкладом наблюдения $X$, а функция $I_X(\theta) = E_\theta u_\theta^2(X)$ называется количеством информации о параметре $\theta$ (информация по Фишеру).

**Определение 10.12.** Следующие условия называются условиями регулярности

1. $\Theta \subset \mathbb{R}$ – открытый интервал, возможно, бесконечный.

2. $A = \{x \in \mathcal{X} : p(x) > 0\}$ – носитель, не зависит от $\theta$.

3. Для любой статистики $S(X)$ с условием $\forall\theta\in\Theta \hookrightarrow\ E_\theta S^2(X)<\infty$ выполняется
$$\frac{\partial}{\partial\theta}\int_{\mathcal{X}} S(x)p_\theta(x)\mu(dx)=\int_{\mathcal{X}} S(x)\frac{\partial}{\partial\theta}p_\theta(x)\mu(dx),$$
откуда
$$\frac{\partial}{\partial\theta}E_\theta S(X)=\int_A S(x)\left(\frac{\partial}{\partial\theta}p_\theta(x)\right)\frac{p_\theta(x)}{p_\theta(x)}\mu(dx)=E_\theta\left[S(X)\frac{\partial}{\partial\theta}\ln p_\theta(x)\right]=E_\theta S(X)u_\theta(X).$$
В частности, будем считать, что $\forall\theta\in\Theta$ величина $\dfrac{\partial}{\partial\theta}\ln p_\theta(x)$ существует на $A$ и конечна.

4. $0<I_X(\theta)<\infty,\ \forall\theta\in\Theta.$

**Упражнение 10.13.** Проверить, что $\mathcal{N}\left(a,\,\sigma^2\right)$ регулярно.

**Теорема 10.14.** (Неравенство Рао-Крамера) Пусть выполняется условие регулярности, и $\hat\theta(X)$ – несмещенная оценка $\tau(\theta)$ с условием, что $E_\theta(\hat\theta(X))^2<\infty\ \forall\theta\in\Theta$. Тогда
$$D_\theta\hat\theta(X)\geqslant\frac{(\tau'(\theta))^2}{I_X(\theta)}.$$

**Доказательство.** Положим $S(X)\equiv1$. Тогда по третьему условию регулярности имеем $\dfrac{\partial}{\partial\theta}E_\theta S(X)=\dfrac{\partial}{\partial\theta}1=0$. С другой стороны,
$$\frac{\partial}{\partial\theta}E_\theta S(X)=E_\theta\frac{\partial}{\partial\theta}\ln p_\theta(x)\Rightarrow E_\theta u_\theta(X)=0.$$
Пусть теперь $S(X)=\hat\theta(X),\ \tau'(\theta)=E_\theta S(X)u_\theta(X)$. С учетом предыдущих неравенств получаем
$$\tau'(\theta)=E_\theta S(X)u_\theta(X)-\tau(\theta)E_\theta u_\theta(X)=E_\theta(S(X)-\tau(\theta))u_\theta(X).$$
Используя неравенство Коши-Буняковского получаем
$$(\tau'(\theta))^2\leqslant E_\theta(S(X)-\tau(\theta))^2E_\theta u_\theta^2(X)=D_\theta S(X)\cdot I_X(\theta)\Rightarrow$$
$$D_\theta S(X)\geqslant\frac{(\tau'(\theta))^2}{I_X(\theta)}.$$

□

**Следствие 10.14.1.** Если $\tau(\theta)=\theta$, то в условиях предыдущей теоремы $D_\theta\hat\theta(X)\geqslant\dfrac{1}{I_X(\theta)}$, где $\hat\theta$ – несмещенная оценка $\tau(\theta)$.

**Упражнение 10.15.** $I_X(\theta)=n\cdot i(\theta)$, где $i(\theta)$ – информация одного наблюдения.

**Определение 10.16.** Если в неравенстве Рао-Крамера для несмещенной оценки $\hat\theta$ достигается равенство, то $\hat\theta$ называется эффективной оценкой $\tau(\theta)$.

**Теорема 10.17.** (Критерий эффективности) В условиях регулярности $\hat\theta$ – эффективная оценка для $\tau(\theta)\Leftrightarrow\hat\theta$ – линейная функция от $u_\theta(X)$, т.е.
$$\hat\theta(X)-\tau(\theta)=c(\theta)\cdot u_\theta(X)$$
для некоторого $c(\theta)$. Причем равенство выполняется тогда и только тогда, когда $c(\theta)=\dfrac{\tau'(\theta)}{I_X(\theta)}$.

**Доказательство.** Пусть $\hat{\theta}$ – эффективная оценка для $\tau(\theta)$. Тогда

$$\tau'(\theta) = E_\theta(\hat{\theta}(X) - \tau(\theta))u_\theta(X).$$

Так как оценка $\hat{\theta}$ эффективна, то выполнилось равенство в КБШ. Тогда величины $\xi := (\hat{\theta}(X) - \tau(\theta))$ и $\eta := u_\theta(X)$ линейно зависимы, т.е. $\alpha(\theta) + \beta(\theta)\eta + \gamma(\theta)\xi = 0$. Беря математическое ожидание от обеих частей, получаем

$$E\alpha(\theta) + \beta(\theta)Eu_\theta(X) + \gamma(\theta)E(\hat{\theta} - \tau(\theta)) = \alpha(\theta) = 0.$$

Тогда

$$\beta(\theta)u_\theta(X) + \gamma(\theta)(\hat{\theta} - \tau(\theta)) = 0.$$

Предположим, что $\gamma(\theta) = 0$. Тогда $D_\theta\beta(\theta)u_\theta(X) = \beta^2(\theta)D_\theta u_\theta(X) = 0$. То, что $\hat{\theta}$ эффективная оценка, подразумевает существование обеих частей в выражении

$$D_\theta\hat{\theta} = \frac{(\tau'(\theta))^2}{I_X(\theta)} \Rightarrow I_X(\theta) > 0 \Rightarrow I_X(\theta) = E_\theta u_\theta^2(X) = D_\theta u_\theta(X) + (E_\theta u_\theta(X))^2 = D_\theta u_\theta(X) > 0.$$

Из этого заключаем, что и $\beta(\theta) \equiv 0$. Пришли к противоречию с тем, что существует нетривиальная линейная зависимость. Следовательно, $\gamma(\theta) \neq 0$, и

$$\hat{\theta} - \tau(\theta) = -\frac{\beta(\theta)}{\gamma(\theta)}u_\theta(X) = c(\theta)u_\theta(X).$$

Обратно, пусть

$$\hat{\theta} - \tau(\theta) = c(\theta)u_\theta(X).$$

Тогда, взяв математическое ожидание от обеих частей, получаем

$$E_\theta\hat{\theta} = \tau(\theta),$$
$$E_\theta(\hat{\theta} - \tau(\theta))^2 = c^2(\theta)Eu_\theta^2(X) < \infty \Rightarrow E_\theta\hat{\theta}^2 < \infty.$$

Домножив обе части на $u_\theta(X)$, выходит, что

$$\tau'(\theta) = E_\theta(\hat{\theta} - \tau(\theta))u_\theta(X) = E_\theta c(\theta)u_\theta^2(X) = c(\theta)I_X(\theta) \Leftrightarrow$$
$$c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)},$$

и из линейной зависимости $\xi, \eta$ следует эффективность. □

**Следствие 10.17.1.** Если $\theta^*$ не хуже $\hat{\theta}$ из предыдущей теоремы, то из критерия эффективности следует, что $\theta^* = \hat{\theta}$ $P_\theta$-п.н.

**Замечание 10.18.** Эффективная оценка – наилучшая оценка в классе несмещенных в равномерном подходе с квадратичной функцией потерь (обратное неверно).

**Теорема 10.19.** Если в условиях регулярности существует эффективная оценка для $\tau(\theta) \not\equiv const$, то множество функций, для которых существует эффективная оценка – это

$$\{a + b\tau(\theta) : a,\, b - const\}.$$

**Доказательство.** Если $T(X)$ – эффективная оценка для $\tau(\theta)$, $V(X)$ – эффективная оценка $v(\theta)$. Производная функции $\tau$ существует в любой точке в силу условий регулярности. Из того, что $\tau \not\equiv const$ следует, что $\exists \theta_0 \in \Theta : \tau'(\theta_0) \neq 0$ (нужно чтобы делить на $c(\theta)$). Так как $0 < I_X(\theta) < \infty$, то $c(\theta) \neq 0$. По критерию эффективности

$$T(X) = \tau(\theta) + c(\theta) u_\theta(x)$$

$$V(X) = v(\theta) + d(\theta) u_\theta(x)$$

Тогда

$$u_{\theta_0}(X) \xlongequal{P_{\theta_0}-a.s.} \frac{T(X) - \tau(\theta_0)}{c(\theta_0)} \Rightarrow V(X) \xlongequal{P_{\theta_0}-a.s.} v(\theta_0) + \frac{d(\theta_0)}{c(\theta_0)}(T(X) - \tau(\theta_0)).$$

**Упражнение 10.20.** Рассмотрим $A = \{x : p_\theta(x) > 0\}$ не зависит от $\theta$. Тогда

$$(\exists \theta : \ P_\theta(x \in B) = 1) \Leftrightarrow (\forall \theta : P_\theta(x \in B) = 1).$$

**Доказательство.** Пусть $C = X \backslash B$. Тогда

$$P_\theta(x \in C) = P_\theta(x \in C \cap A) = \int_{C \cap A} p_\theta(x) \mu(dx).$$

Если $P_{\theta_0}(x \in B) = 1$, то

$$P_{\theta_0}(x \in C) = \int_{C \cap A} p_{\theta_0}(x) \mu(dx) = 0 \Rightarrow \mu(C \cap A) = 0 \Rightarrow P_\theta(x \in C) = 0.$$

□

Получили, что $\forall \theta \hookrightarrow V(X) \xlongequal{P_\theta-a.s.} a + bT(X)$. Взяв математическое ожидание, получаем

$$v(\theta) = a + b\tau(\theta).$$

□

# 11 Экспоненциальные семейства распределений. Их связь с условием существования эффективной оценки.

**Определение 11.1.** Пусть $\theta = (\theta_1, \ldots, \theta_k)$. Экспоненциальным семейством распределений называются все распределения, обобщенная плотность которых имеет вид

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

и где функции $a_0(\theta) \equiv 1$, $a_1(\theta), \ldots,$ $a_k(\theta)$ линейно независимы.

**Пример 11.2.** Рассмотрим распределение $\Gamma(\alpha, \lambda)$. Его плотность распределения имеет вид

$$p_{\alpha,\lambda}(x) = \frac{\alpha^\lambda x^{\lambda-1}}{\Gamma(\alpha)} e^{-\alpha x} \mathbb{I}_{x>0} = \frac{1}{x} \exp \left( \lambda \ln x - \alpha x + \ln \frac{\alpha^\lambda}{\Gamma(\alpha)} \right) \mathbb{I}_{x>0}. \tag{3}$$

**Утверждение 11.3.** Пусть выполнено условие регулярности. Тогда существует эффективная оценка для некоторой функции $\tau(\theta) \Leftrightarrow$ наблюдения принадлежат экспоненциальному семейству.

**Доказательство.** Пусть семейство экспоненциальное, то есть

$$f_\theta(X) = \prod_{i=1}^{n} p_\theta(X_i),\ p_\theta(X_i) = h(X_i)e^{a(\theta)T(X_i)+V(\theta)}.$$

Тогда

$$\exists u_\theta(X) = \frac{\partial}{\partial\theta}\ln f_\theta(X) = \frac{\partial}{\partial\theta}\left(a(\theta)\sum_{i=1}^{n}T(X_i) + nV(\theta)\right).$$

Заметим, что, если $T(X_i) \equiv const$, то

$$p_\theta(x) = h(x)b(\theta) \Rightarrow \int_{\mathcal{X}} h(x)b(\theta)dx = 1 \Rightarrow b(\theta) \equiv const.$$

В этом случае распределение не зависит от $\theta$, и такой случай не рассматриваем.

Заметим, что если функция $f(x)h(y) + g(x)$ дифференцируема по $x$ при любом $y$, и $h(y) \neq const$, то функции $f(x)$ и $g(x)$ дифференцируемы по $x$, т.к. $\exists y_1,\, y_2 : h(y_1) \neq h(y_2)$, и функция $[f(x)h(y_1) + g(x)] - [f(x)h(y_2) + g(x)] = f(x)(h(y_1) - h(y_2))$ дифференцируема по $x$ как разность двух дифференцируемых функций. Следовательно, и функции $f(x)$, а вместе с ней и $g(x)$ дифференцируемы. Из этих соображений получаем, что $a(\theta)$ и $V(\theta)$ дифференцируемы, и

$$u_\theta(X) = a'(\theta)\sum_{i=1}^{n}T(X_i) + nV'(\theta).$$

Из предположения, что $\forall\theta \hookrightarrow a'(\theta) \neq 0$, получаем

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

Обозначив $\tau(\theta) = \dfrac{-V'(\theta)}{a'(\theta)}$, $\tau^*(X) = \dfrac{\sum_{i=1}^{n}T(X_i)}{n}$, $c(\theta) = \dfrac{1}{na'(\theta)}$, по критерию эффективности получаем требуемое. Обратно, пусть существует эффективная оценка $T(X)$ для $\tau(\theta)$. Пусть $\tau'(\theta) \neq 0$. Тогда

$$\forall\theta \in \Theta \hookrightarrow T(X) - \tau(\theta) = c(\theta)u_\theta(X)\ P_\theta - a.s. \Leftrightarrow$$
$$\frac{T(X) - \tau(\theta)}{c(\theta)} = \frac{\partial}{\partial\theta}\ln f_\theta(X).$$

Предполагая корректность, интегрируем по $\theta$.

$$\ln f_\theta(X) = \int \frac{T(X) - \tau(\theta)}{c(\theta)}d\theta + g(X),$$

$g(X)$ - для каждого $X$ своя константа интегрирования. Тогда

$$f_\theta(X) = \prod_{i=1}^{n} p_\theta(X_i) = H(X)\exp\{B(\theta)T(X) + D(\theta)\}\cdot\mathbb{I}_A(X).$$

Фиксируя $x_2^0,\, \ldots,\, x_n^0 \in A$, получаем

$$p_\theta(x_1) = \frac{H(x_1,\, x_2^0, \ldots,\, x_n^0)\exp\{B(\theta)T(x_1,\, x_2^0, \ldots,\, x_n^0) + D(\theta)\}}{\prod_{i=2}^{n} p_\theta(x_i^0)}\cdot\mathbb{I}_A(X).$$

Из предположения, что функция распределения зависит от $\theta$, получаем, что $a_1(\theta)$ независимо с $a_0 \equiv 1$. Следовательно, семейство экспоненциальное. $\square$

# 12 Достаточные статистики. Критерий факторизации Неймана–Фишера (доказательство для дискретного случая). Теорема Колмогорова–Блекуэлла–Рао об улучшении несмещенной оценки.

**Определение 12.1.** Пусть $X$ – выборка из неизвестного распределения $P \in \{P_\theta,\, \theta \in \Theta\}$. Статистика $T(X)$ называется *достаточной* для параметра $\theta$, если для любого борелевского множества $B$ и любого $t$ выполняется, что $P_\theta(X \in B \,|\, T(X) = t)$ не зависит от $\theta$.

**Замечание 12.2.** Так как условная вероятность не однозначно определенный объект, то в определении подразумевается, что существует такой вариант условного распределения, что оно не зависит от $\theta$.

**Замечание 12.3.** Если статистики $S$ и $T$ находятся во взаимно однозначном соответствии, и $T$ достаточная, тогда и $S$ достаточная.

**Теорема 12.4.** (Неймана-Фишера, критерий факторизации). Пусть $\{P_\theta,\, \theta \in \Theta\}$ – доминируемое семейство. Тогда статистика $T$ является достаточной для параметра $\theta$ тогда и только тогда, когда функция правдоподобия $f_\theta(x)$ представляется в виде $f_\theta(x) = \psi(T(x),\, \theta)h(x)$, где функции $\psi$, $h$ неотрицательны, $\psi(t,\, \theta)$ измерима по $t$, и $h(x)$ измерима по $x$.

**Доказательство.** (Для дискретного случая, т.е. $f_\theta(x) = P_\theta(X = x)$). Пусть $f_\theta(x) = \psi(T(X),\, \theta)h(x)$. Тогда

$$P_\theta(X = x \,|\, T(X) = t) = \frac{P_\theta(X = x,\, T(X) = t)}{P_\theta(T(X) = t)}.$$

Если $T(X) \neq t$, то эта вероятность нулевая. Иначе,

$$P_\theta(X = x \,|\, T(X) = t) = \frac{P_\theta(X = x)}{\sum_{y\colon T(y)=t} P(X = y)} =$$
$$\frac{\psi(T(X),\, \theta)h(x)}{\sum_{y\colon T(y)=t} \psi(T(y),\, \theta)h(y)} = \frac{h(x)}{\sum_{y\colon T(y)=t} h(y)}.$$

Следовательно, $T$ – достаточная.

Обратно, пусть $T$ – достаточная статистика. Тогда

$$f_\theta(X) = P_\theta(X = x) = P_\theta(X = x,\, T(X) = T(x)) =$$
$$P_\theta(T(X) = T(x))P_\theta(X = x \,|\, T(X) = T(x)).$$

Обозначив $\psi(T(X),\, \theta) = P_\theta(T(X) = T(x)),\ h(x) = P_\theta(X = x \,|\, T(X) = T(x))$, получаем требуемое. □

**Теорема 12.5.** (Колмогоров-Блэкуэлл-Рао, об улучшении несмещенных оценок). Пусть $d(X)$ – несмещенная оценка $\tau(\theta)$, $E_\theta d^2(X) < \infty\ \forall\theta \in \Theta$, и $T(X)$ – достаточная статистика для $\theta$. Пусть $\varphi(T) = E_\theta(d(X) \,|\, T)$. Тогда $\varphi(T)$ зависит от выборки только через $T(X)$ (и не зависит от $\theta$), т.е. $\varphi(T)$ – статистика. Более того, $E_\theta\varphi(T) = \tau(\theta),\ D_\theta\varphi(T) \leqslant D_\theta(d(X))$, и равенство достигается тогда и только тогда, когда $\varphi(X) = d(X)$ ($\forall\theta\ P_\theta - a.s.$), и в этом случае $d$ является $T$-измеримой, т.е. борелевской функцией от $T$.

**Доказательство.** Рассмотрим $E_\theta(d(X) \,|\, T)$. При фиксированном значении $T$ распределение $X$ не зависит от $\theta$ (так как $T$ - достаточная). Тогда распределение $d(X)$ также не зависит от $\theta$, следовательно, $E_\theta(d(X)\,|\,T)$ является функцией только от $T$ (и как функция не зависит от $\theta$). Поэтому $\varphi(T)$ – статистика.

Несмещенность очевидна в силу свойств условного математического ожидания.

По неравенству Йенсена, если $h(x)$ – выпуклая функция, и $E(h(\xi)) < \infty$, то $h(E(\xi\,|\,\mathcal{G})) \leqslant E(h(\xi)\,|\,\mathcal{G})$. Тогда, $\varphi^2(T) = E^2(d(X)\,|\,T(X)) \leqslant E(d^2(X)\,|\,T(X)) < \infty$. Тогда $E_\theta\varphi^2(T) \leqslant E_\theta(E(d^2(X)\,|\,T(X))) = E_\theta d^2(X) < \infty$. Далее,

$$D_\theta d(X) = E_\theta(d(X)-\tau(\theta))^2 = E_\theta(d(X)-\varphi(T)+\varphi(T)-\tau(\theta))^2 =$$
$$E_\theta(d(X)-\varphi(T))^2 + E_\theta(\varphi(T)-\tau(\theta))^2 + 2E_\theta(d(X)-\varphi(T))(\varphi(T)-\tau(\theta)) =$$
$$E_\theta(d(X)-\varphi(T))^2 + D_\theta\varphi,$$

т.к. $E_\theta(d-\varphi)(\varphi-\tau) = E_\theta(E_\theta((d-\varphi)(\varphi-\tau)\,|\,T)) = E_\theta((\varphi-\tau)E_\theta(d-\varphi\,|\,T)) = 0$. При этом, неравенство переходит в равенство тогда и только тогда, когда $E_\theta(d(X)-\varphi(T))^2 = 0$, т.е. $d = \varphi\ \forall\theta\ P_\theta - a.s.$ □
Введем расширенное определение оценки.

**Определение 12.6.** *Оценка* – измеримая функция выборки.

**Определение 12.7.** Наилучшая оценка $\tau(\theta)$ в классе несмещенных оценок в равномерном подходе с квадратичной функцией потерь называется *оптимальной* оценкой.

**Следствие 12.7.1.** Если $T(X)$ – достаточная статистика для $\theta$, и $d \in L_2$, то

1. Если $d(X)$ – несмещенная для $\tau(\theta)$, то $E_\theta(d(X)\,|\,T(X))$ не хуже $d(X)$, а даже лучше, если $d(X)$ не является $T$- измеримой.

2. Если $d$ – единственная несмещенная $T$-измеримая оценка $\tau(\theta)$, то $d(X)$ – оптимальная оценка $\tau(\theta)$ ($\forall\theta\ P_\theta - a.s.$).

**Доказательство.**

1. следует из теоремы Колмогорова-Блэкуэлла-Рао.

2. для $\widetilde{d} \in L_2$ выполнено $D_\theta\widetilde{\varphi} \leqslant D_\theta\widetilde{d}$, т.е. $E_\theta\left(E_\theta\left(\widetilde{d}\,|\,T\right) - E_\theta\widetilde{d}\right)^2 \leqslant E_\theta\left(\widetilde{d} - E_\theta\widetilde{d}\right)^2$. На самом деле, это выполнено для любой несмещенной оценки $\widetilde{d}$ по свойству условного математического ожидания, если допустить значения $+\infty$ (б/д).

   Для любой несмещенной $\widetilde{d}$ получаем $\widetilde{\varphi}$ – несмещенная оценка, которая не хуже $\widetilde{d}$. Но т.к. $\widetilde{\varphi}$ – несмещенная $T$-измеримая оценка $\tau(\theta)$, то $\widetilde{\varphi} = d$ в силу единственности. Поэтому, $d$ не хуже $\widetilde{d}$.

   Если $\forall\theta D_\theta\widetilde{d} = D_\theta d$, то из условия $d \in L_2$ следует $D_\theta\widetilde{d} = D_\theta d = D_\theta\widetilde{\varphi} < \infty$. Получили, что $\tilde{d} \in L_2,\ D_\theta\tilde{d} = D_\theta\left(E_\theta\left(\tilde{d}\,|\,T\right)\right)$. Значит, по теореме Колмогорова-Блэкуэлла-Рао $\tilde{d} = E\left(\tilde{d}\mid T\right) = \tilde{\varphi} = d$. Следовательно, $d$ – оптимальна.

□

# 13 Полные статистики. Теорема Лемана-Шеффе об оптимальной оценке. Теорема о полной достаточной статистике в экспоненциальном семействе (б/д). Нахождение оптимальных оценок с помощью полных достаточных статистик.

**Определение 13.1.** Статистика $S(X)$ называется *полной* для $\theta$, если из условия $E_\theta f(S(X)) = 0\ \forall\theta \in \Theta$ следует, что $f(S(X)) = 0\ \forall\theta \in \Theta\ P_\theta - a.s.$

**Теорема 13.2.** (Лемана-Шеффе, об оптимальной оценке) Пусть $T(X)$ – полная достаточная статистика для $\{P_\theta,\ \theta \in \Theta\}$, $\varphi(T(X))$ – несмещенная оценка для $\tau(\theta)$. Тогда $\varphi(T(X))$ не хуже любой несмещенной оценки $\tau(\theta)$ в среднеквадратичном подходе, и если $\varphi(T) \in L_2$, то $\varphi(T)$ – оптимальная оценка.

**Доказательство.** Пусть $\tilde{d}$ – несмещенная оценка $\tau(\theta)$. Тогда $\tilde{\varphi}(T) = E\left(\tilde{d} \mid T\right)$ не хуже $\tilde{d}$ и является несмещенной. Значит, $E_\theta\left(\varphi(T) - \tilde{\varphi}(T)\right) = 0$. Возьмем $h = \varphi - \tilde{\varphi}$. Тогда $\forall\theta\ E_\theta h(T) = 0$. По определению полной статистики $h(T) = 0\ P_\theta - a.s.\ \forall\theta,$ а следовательно, $\varphi(T) = \tilde{\varphi}(T)\ (P_\theta - a.s.)$.

Пусть теперь $\varphi(T) \in L_2$. В силу следствия достаточно доказать, что $\varphi(T)$ – единственная несмещенная $T$-измеримая оценка.

Пусть $\psi(T)$ – несмещенная $T$-измеримая оценка. Тогда $E_\theta(\psi(T) - \varphi(T)) = 0 \Rightarrow \psi(T) = \varphi(T)$. □

**Следствие 13.2.1.** Пусть $T(X)$ – полная достаточная статистика, $d(X)$ – несмещенная оценка для $\tau(\theta)$ $(d \in L_1)$. Тогда $\varphi = E_\theta(d \mid T)$ не хуже любой другой оценки в среднеквадратичном подходе. Если $\varphi \in L_2$, то $\varphi$ – оптимальна.

**Доказательство.** $E_\theta(d \mid T(X))$ – несмещенная оценка и является функцией от $T$. □

**Теорема 13.3.** (об экспоненциальном семействе, б/д) Пусть $X_1,\ \ldots,\ X_n$ – выборка из экспоненциального семейства распределений. Если множество значений вектор-функции $\forall\theta\ \overline{a}(\theta) = \begin{pmatrix} a_0(\theta) & \ldots & a_k(\theta) \end{pmatrix}$ из определения экспоненциального семейства содержит $k$-мерный параллелепипед в $\mathbb{R}^k$, то $T(X) = \begin{pmatrix} T_1(X) & \ldots & T_k(X) \end{pmatrix}$ – полная достаточная статистика.

**Замечание 13.4.** Для того, чтобы вектор-функция $\overline{a}(\theta)$ содержала $k$-мерный параллелепипед, можно потребовать, чтобы $\Theta$ содержало открытое множество, и чтобы $a_1(\theta), \ldots, a_k(\theta)$ были гладкими.

**Утверждение 13.5.** (Алгоритм поиска оптимальной оценки)

1. Ищем достаточную статистику
2. Проверяем полноту (по определению или по предыдущей теореме)
3. Если полна, то решаем уравнение несмещенности: $E_\theta g(T(X)) = \tau(\theta)\ \forall\theta$.

# 14 Доверительные интервалы. Метод центральной статистики. Асимптотические доверительные интервалы. Построение асимптотических доверительных интервалов с помощью асимптотически нормальных оценок.

Пусть $X$ – наблюдение с неизвестным распределением $\{P_\theta,\ \theta \in \Theta\},\ \Theta \subset \mathbb{R}$.

**Определение 14.1.** Пара статистик $(T_1(X), T_2(X))$ называется доверительным интервалом уровня доверия $\gamma$ для параметра $\theta$, если $\forall\theta \in \Theta$ выполняется

$$P_\theta(T_1(X) < \theta < T_2(X)) \geqslant \gamma.$$

Если равенство достигается при всех $\theta$, то доверительный интервал называется точным.

**Замечание 14.2.** На практике используют $\gamma = 0,9;\ 0,95;\ 0,99$.

Иногда удобно использовать односторонние доверительные интервалы вида $(-\infty,\ T_2(X))$ или $(T_1(X),\ +\infty)$.

В случае многомерного $\theta \in \Theta \subset \mathbb{R}^k$ можно определить понятие доверительного интервала для компонент вектора $\theta_i$ вектора $\theta = \begin{pmatrix} \theta_1 & \ldots & \theta_k \end{pmatrix}$ и для скалярных функций $\tau(\theta)$.

**Определение 14.3.** Множество $S \subset \Theta$ называется доверительным множеством уровня доверия $\gamma$, если $\forall \theta \in \Theta \hookrightarrow P_\theta(\theta \in S(X)) \geqslant \gamma$.

**Метод центральной статистики**

**Определение 14.4.** Предположим, что существует известная одномерная функция $G(X, \theta)$, такая что ее распределение не зависит от параметра $\theta$. Тогда такая функция называется центральной статистикой.

Пусть $\gamma_1,\ \gamma_2 \in (0,\ 1)$ таковы, что $\gamma_2 - \gamma_1 = \gamma$, и при $i = 1, 2$ существует $g_i$ – $\gamma_i$-квантиль $G(X,\ \Theta)$. Тогда $\forall \theta \in \Theta$

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

Тогда $S(X) := \{\theta \in \Theta :\ g_1 \leqslant G(X,\ \theta) \leqslant g_2\} \Rightarrow P_\theta(\theta \in S(X)) = P_\theta(g_1 \leqslant G(X,\ \theta) \leqslant g_2) \geqslant \gamma$, т.е. $S(X)$ – доверительное множество уровня $\gamma$.

**Пример 14.5.** $X_i \ \sim\ \mathcal{N}(\theta,\ 1),\ \theta \in \mathbb{R}$. Строим доверительный интервал для $\theta$. Так как $\overline{X} \ \sim \ \mathcal{N}\left(\theta,\ \dfrac{1}{n}\right)$, то $\sqrt{n}(\overline{X} - \theta) \ \sim\ \mathcal{N}(0,\ 1)$. Тогда $G(X,\ \theta) = \sqrt{n}(\overline{X} - \theta)$ – центральная статистика. Пусть $u_p$ – $p$-квантиль стандартного нормального распределения. Тогда

$$\forall \theta \hookrightarrow P_\theta\left(u_{\frac{1-\gamma}{2}} \leqslant \sqrt{n}(\overline{X} - \theta) \leqslant u_{\frac{1+\gamma}{2}}\right) = P\left(\sqrt{n}|\overline{X}\ -\ \theta| \leqslant u_{\frac{1+\gamma}{2}}\right).$$

Тогда $\left(\overline{X} - \dfrac{1}{\sqrt{n}} u_{\frac{1-\gamma}{2}},\ \overline{X} + \dfrac{1}{\sqrt{n}} u_{\frac{1+\gamma}{2}}\right)$ – доверительный интервал для $\theta$.

**Лемма 14.6.** $X_1,\ \ldots,\ X_n$ – независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$. Тогда $G(X_1,\ \ldots,\ X_n) = -\sum\limits_{i=1}^{n} \ln F(x_i) \ \sim\ \Gamma(1,\ n)$.

**Доказательство.** Покажем, что $F(x_i) \sim \mathcal{U}[0,\ 1]$.

$$P(F(x_i) \leqslant y) = P\left(x_i \leqslant F^{-1}(y)\right),$$

где $y \in (0,\ 1),\ F^{-1}(y) = \min(x :\ F(x) = y)$. Следовательно, $F\left(F^{-1}(y)\right) = y$. Тогда

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

□

**Следствие 14.6.1.** Если $X_1,\ \ldots,\ X_n$ – выборка из распределения $P \in \{P_\theta,\ \theta \in \Theta\}$, причем $\forall \theta$ функция распределения $F_\theta(x)$ непрерывна по $x$, тогда $G(X_1,\ \ldots,\ X_n,\ \theta) = -\sum\limits_{i=1}^{n} \ln F_\theta(X_i)$ – центральная статистика.

**Определение 14.7.** Пусть $\{X_n\}_{n \geqslant 1}$ – выборка неограниченного размера из неизвестного распределения $P \in \{P_\theta,\ \theta \in \Theta\}$. Последовательности статистик $T_n^{(1)}(X_1,\ \ldots,\ X_n),\ T_n^{(2)}(X_1,\ \ldots,\ X_n)$ называются асимптотическим доверительным интервалом уровня доверия $\gamma$ для параметра $\theta$, если

$$\forall \theta \in \Theta \hookrightarrow \varliminf_{n \to \infty} P_\theta\left(T_n^{(1)}(X_1,\ \ldots,\ X_n) < \theta < T_n^{(2)}(X_1,\ \ldots,\ X_n)\right) \geqslant \gamma.$$

Если $\forall\theta \in \Theta \hookrightarrow \exists \lim\limits_{n\to\infty} P_\theta\left(T_n^{(1)}(X_1,\ \ldots,\ X_n) < \theta < T_n^{(2)}(X_1,\ \ldots,\ X_n)\right) = \gamma$, то асимптотический интервал называется точным.

**Построение асимптотических доверительных интервалов**

Пусть $\hat{\theta}_n(X_1,\ \ldots,\ X_n)$ – асимптотически нормальная оценка $\theta$ с асимптотической дисперсией $\sigma^2(\theta) > 0$, то есть

$$\forall\theta \in \Theta \hookrightarrow \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \sigma^2(\theta)\right) \Leftrightarrow$$
$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} \xrightarrow{d_\theta} \mathcal{N}(0,\ 1).$$

Пусть функция $\sigma(\theta)$ непрерывна по $\theta$. Из того, что оценка $\hat{\theta}_n$ асимптотически нормальна следует, что она состоятельная, т.е. $\hat{\theta} \xrightarrow{P_\theta} \theta$. Тогда, по теореме о наследовании сходимостей и лемме Слуцкого

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} = \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} \cdot \frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)} \xrightarrow{d_\theta} \mathcal{N}(0,\ 1).$$

Следовательно,

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

где $u_{\frac{1+\gamma}{2}}$ – квантиль уровня $\dfrac{1+\gamma}{2}$ стандартного нормального распределения. Следовательно, асимптотический доверительный интервал имеет вид

$$\left(\hat{\theta}_n - u_{\frac{1+\gamma}{2}}\frac{\sigma(\hat{\theta}_n)}{\sqrt{n}},\ \hat{\theta}_n + u_{\frac{1+\gamma}{2}}\frac{\sigma(\hat{\theta}_n)}{\sqrt{n}}\right).$$

# 15 Метод максимального правдоподобия. Экстремальное свойство функции правдоподобия. Состоятельность оценки максимального правдоподобия.

**Определение 15.1.** Пусть $X$ – наблюдение с неизвестным распределением $P \in \{P_\theta,\ \theta \in \Theta\}$, где $\{P_\theta,\ \theta \in \Theta\}$ – доминируемое семейство относительно меры $\mu$. Функцией правдоподобия называют $f_\theta(X) = p_\theta(X)$.

**Пример 15.2.** Пусть $X = (X_1,\ \ldots,\ X_n)$ – выборка из распределения с плотностью $p(x)$ и обычной мерой Лебега. Тогда $f_\theta(X) = p_\theta(X) = \prod\limits_{i=1}^{n} p_\theta(X_i)$ – функция правдоподобия.

**Определение 15.3.** Пусть $X$ – наблюдение с функцией правдоподобия $f_\theta(X)$. Оценкой параметра $\theta$ по методу максимального правдоподобия (ОМП) называется такая оценка $\hat{\theta}$, что $\hat{\theta}(X) = \arg\max\limits_{\theta\in\Theta} f_\theta(X)$.

**Замечание 15.4.** ▷ Важно, что это оценки может и **не быть** вовсе.

▷ Может быть не единственной, если максимумов несколько

**Пример 15.5.** Пусть $X$ – выборка из распределения $\mathcal{U}[0,\ \theta]$. Тогда

$$f_\theta(X) = \frac{1}{\theta^n} I(X_{(1)} \geqslant 0,\ X_{(n)} \leqslant \theta),$$

и ОМП – $\hat{\theta}(X) = X_{(n)}$.

**Определение 15.6.** Логарифмом функции правдоподобия называется функция $L_\theta(X) = \ln f_\theta(X)$.

## Условия регулярности

Перечислим следующие условия регулярности по книге "Теория точечного оценивания" Лемана.

- ▷ (R0) $\{P_\theta,\ \theta \in \Theta\}$ – параметрическое семейство, доминируемое относительно меры $\mu$, при $\theta_1 \neq \theta_2 \hookrightarrow P_{\theta_1} \neq P_{\theta_2}$, и $\forall \theta \in \Theta$ определена плотность $p_\theta(X)$ меры $P_\theta$.
- ▷ (R1) $A = \{x \in \mathcal{X} :\ p_\theta(x) > 0\}$ не зависит от $\theta$.
- ▷ (R2) Наблюдение $X$ есть выборка $X = (X_1,\ \ldots,\ X_n)$ из неизвестного распределения $P \in \{P_\theta,\ \theta \in \Theta\}$.
- ▷ (R3) $\Theta$ – открытый интервал в $\mathbb{R}$ (возможно, бесконечный).
- ▷ (R4) Функция $p_\theta(x)$ непрерывно дифференцируема по $\theta$ при всех $x \in A$.
- ▷ (R5) $p_\theta(x)$ трижды непрерывно дифференцируема по $\theta$ для всех $x \in A$.
- ▷ (R6) интеграл $\displaystyle\int_A p_\theta(x)\mu(dx)$ трижды дифференцируем по $\theta$ под знаком интеграла.
- ▷ (R7) $\forall \theta \in \Theta \hookrightarrow E_\theta \left(\dfrac{\partial}{\partial \theta} \ln p_\theta(X)\right)^2 = i(\theta) \in (0, +\infty)$.
- ▷ (R8) $\forall \theta_0 \in \Theta\ \exists c > 0\ \exists H(x) :\ \forall \theta \in (\theta_0 - c, \theta_0 + c)\, \forall x \in A \hookrightarrow \left|\dfrac{\partial^3}{\partial \theta^3} \ln p_\theta(x)\right| \leqslant H(x)$, и $E_{\theta_0} H(x) < \infty$.

**Лемма 15.7.** Если $\xi_1,\ \ldots,\ \xi_n$ – независимые одинаково распределенные случайные величины, $E\xi_1 \in \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. Тогда $\dfrac{S_n}{n} \xrightarrow{a.s.} E\xi_1$.

**Доказательство.** Если $E\xi_1 \in \mathbb{R}$, то работает ЗБЧ. Пусть $E\xi_1 = +\infty$. Представим $E\xi_1$ в виде $E\xi_1 = E^+\xi_1 - E^-\xi_1 \Rightarrow E^-\xi_1 < \infty$. Зафиксируем $M > 0$. Определим $\xi_i^M := \min\{\xi_i,\ M\} \Rightarrow E\xi_1^M < \infty$. Тогда

$$\varliminf_{n\to\infty} \frac{S_n}{n} \geqslant \varliminf_{n\to\infty} \frac{S_n^M}{n} = \lim_{n\to\infty} \frac{S_n^M}{n} = E\xi_1^M$$

на множестве единичной меры по УЗБЧ. Значит, $\forall M \hookrightarrow E\xi_1^M \leqslant \varliminf\limits_{n\to\infty} \dfrac{S_n}{n}$, $M_1 \leqslant M_2 \Rightarrow \xi_1^{M_1} \leqslant \xi_1^{M_2}$, и по теореме Леви о монотонной сходимости $\lim\limits_{M\to\infty} E\xi_1^M = E \lim\limits_{M\to\infty} \xi_1^M = E\xi_1 = +\infty \Rightarrow \varliminf\limits_{n\to\infty} \dfrac{S_n}{n} = +\infty$. Аналогично доказывается случай $E\xi_1 = -\infty$. □

**Теорема 15.8.** (экстремальное свойство правдоподобия) Пусть выполняются условия регулярности (R0)-(R2). Тогда $\forall \theta_0,\ \theta \in \Theta,\ \theta_0 \neq \theta$ выполняется

$$P_{\theta_0}(f_{\theta_0}(X_1,\ \ldots,\ X_n) > f_\theta(X_1,\ \ldots X_n)) \xrightarrow[n\to\infty]{} 1.$$

**Доказательство.** Пусть $X_1, \ \dots, \ X_n \in A$. Тогда

$$f_{\theta_0}(X_1, \ \dots, \ X_n) > f_\theta(X_1, \ \dots, \ X_n) \Leftrightarrow \frac{1}{n} \ln \left( \frac{f_\theta(X_1, \ \dots, \ X_n)}{f_{\theta_0}(X_1, \ \dots, \ X_n)} \right) < 0 \Leftrightarrow$$
$$\frac{1}{n} \sum_{i=1}^{n} \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} < 0.$$

По предыдущей лемме $\frac{1}{n} \sum_{i=1}^{n} \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} \xrightarrow{P_{\theta_0}-a.s.} E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)}$. Осталось доказать, что $E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} < 0$.

$$E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \int_A \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x)\mu(dx) = \int_A \ln \left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x)\mu(dx).$$

Так как $\forall x > -1 \hookrightarrow \ln(1+x) \leqslant x$, то

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

Равенство нулю достигается только в том случае, когда

$$\int_A \ln \left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x)\mu(dx) = \int_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}(x)\mu(dx) \Rightarrow$$
$$\ln \left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) \overset{P_\theta-a.s.}{=\!=\!=} \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 \Rightarrow$$
$$\frac{p_\theta(x)}{p_{\theta_0}(x)} \overset{P_\theta-a.s.}{=\!=\!=} 1 \Rightarrow p_\theta(x) \overset{P_\theta-a.s.}{=\!=\!=} p_{\theta_0}(x).$$

Но это противоречие с условием (R0), что $P_\theta \neq P_{\theta_0}$ при $\theta \neq \theta_0$.
Следовательно, $E_{\theta_0} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} < 0$. □

**Следствие 15.8.1.** Если $\Theta$ – конечно, то ОМП существует, единственна с вероятностью близкой к единице и состоятельна.

> **Доказательство.** Например, пусть
>
> $$\theta_1, \ \dots, \ \theta_n \in \Theta, \ A := \{x : \ \arg\max_{\theta \in \Theta} f_\theta(x) = \theta_2 \neq \theta_i \ \forall i \neq 2\} =$$
> $$\{x : \ f_{\theta_1}(x) < f_{\theta_2}(x)\} \cap \{x : \ f_{\theta_3}(x) < f_{\theta_2}(x)\} \cap \ \dots \ \cap \{x : \ f_{\theta_n}(x) < f_{\theta_2}(x)\}.$$
>
> Эти множества по теореме имеют меру стремящуюся к единице относительно $P_{\theta_2}$. Из конечности $\Theta$ следует, что и мера множества $A$ близка к единице, то есть с вероятностью близкой к единице ОМП единственна.
>
> Почему существует измеримая ОМП? Введем множества
>
> $$c_{i<j} := \{x : \ f_{\theta_i}(x) < f_{\theta_j}(x)\},$$
> $$c_{i \leqslant j} := \{x : \ f_{\theta_i}(x) \leqslant f_{\theta_j}(x)\}.$$

Если при каком-то $x$ максимум достигается при нескольких $\theta_j$, то выберем ту, у которой номер меньше. Тогда, условие равенства ОМП $\theta_j$ можно записать в виде

$$\{x:\ \hat{\theta}=\theta_1\}=\bigcap_{j=1}^{n} c_{j\leqslant 1},\ \{x:\ \hat{\theta}=\theta_2\}=c_{1<2}\cap\left(\bigcap_{j=2}^{n} c_{j\leqslant 2}\right),\ \ldots,$$
$$\{x:\ \hat{\theta}=\theta_n\}=\bigcap_{i=1}^{n-1} c_{i<n}.$$

Так как плотность распределения интегрируема, то она измерима. Разность измеримых функций – измеримая функция. Тогда

$$c_{i<j}=\{x:\ f_{\theta_i}(x)-f_{\theta_j}(x)<0\},$$
$$c_{i\leqslant j}=\{x:\ f_{\theta_i}(x)-f_{\theta_j}(x)\leqslant 0\}$$

измеримые множества. □

**Определение 15.9.** Уравнением правдоподобия называется

$$\frac{\partial \ln f_\theta}{\partial\theta}=0\Leftrightarrow\frac{\partial f_\theta}{\partial\theta}=0. \tag{4}$$

**Теорема 15.10.** (Аналог состоятельности ОМП) Пусть выполняются условия регулярности (R0)-(R4). Пусть элементы выборки имеют распределение $P_{\theta_0}$. Тогда существует отображение $\hat{\theta}_n(X_1,\ \ldots,\ X_n,\ \theta_0)$ со значениями в $\Theta$, такое что $P^*_\theta(\{\hat{\theta}_n$ – не решение уравнения (4)$\})\to 0$, где $P^*_{\theta_0}$ – внешняя мера, и

$$\forall\varepsilon>0\hookrightarrow P^*_{\theta_0}(\{|\hat{\theta}_n-\theta_0|>\varepsilon\})\xrightarrow[n\to\infty]{}0.$$

**Доказательство.** Определим $\hat{\theta}_n$. Фиксируем $X_1,\ \ldots,\ X_n$ из носителя $A$. Если у уравнения правдоподобия есть хотя бы один корень, то возьмем ближайший к $\theta_0$. Если у (4) нет корней, то $\hat{\theta}_n=\theta_0$. Зафиксируем $\varepsilon>0$, что $[\theta_0-\varepsilon,\ \theta_0+\varepsilon]\subset\Theta$. Рассмотрим

$$S_n(\theta_0,\ \varepsilon):=\{x:\ f_{\theta_0-\varepsilon}(X_1,\ \ldots,\ X_n)<f_{\theta_0}(X_1,\ \ldots,\ X_n),$$
$$f_{\theta_0}(X_1,\ \ldots,\ X_n)>f_{\theta_0+\varepsilon}(X_1,\ \ldots,\ X_n)\}.$$

В силу предыдущей теоремы $P_{\theta_0}(S_n)\xrightarrow[n\to\infty]{}1$. $\forall x\in S_n\ \exists\tilde{\theta}_n$, в которой $f_{\tilde{\theta}_n}$ имеет локальный максимум, т.е. $f'_{\tilde{\theta}_n}(x)=0$ при некотором $\tilde{\theta}_n\in U_\varepsilon(\theta_0)$ (так как в $\theta_0$ значение больше чем на концах окрестности, имеем «холмик» и значит на этом промежутке производная где-то обнулится). А так как $\hat{\theta}_n$ – ближайший к $\theta_0$ корень уравнения правдоподобия, когда он существует, то $\hat{\theta}_n$ является корнем на $S_n$, и $|\hat{\theta}-\theta_0|\leqslant|\tilde{\theta}-\theta_0|<\varepsilon$. Таким образом, $\left\{\hat{\theta}_n\text{ – не корень (4)}\right\}\subset\mathcal{X}^n\backslash S_n$, и в силу монотонности внешней меры выполняется

$$P^*_{\theta_0}\left(\left\{\hat{\theta}_n\text{ – не корень}\right\}\right)\leqslant P_{\theta_0}\left(\mathcal{X}^n\backslash S_n\right)\xrightarrow[n\to\infty]{}0.$$

□

**Замечание 15.11.**

1. $\hat{\theta}_n$ зависит от истинного значения параметра т.е. от $\theta_0$, а значит, не является оценкой.
2. Когда корней несколько непонятно какой из корней ближе.
3. Корень уравнения правдоподобия может не быть ОМП, то есть глобальным максимумом

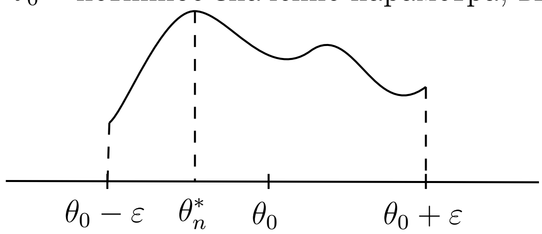
4. Корень уравнения правдоподобия существует не всегда, а на множестве $S_n$, у которого большая вероятность

**Следствие 15.11.1.** Пусть выполняются условия регулярности (R0)-(R4), и $\forall n \in \mathbb{N}\ \forall X_1,\ \ldots,\ X_n$ существует и единственно решение $\hat{\theta}_n(X_1,\ \ldots,\ X_n)$ уравнения правдоподобия, которое является статистикой (т.е. измеримой функцией выборки). Тогда $\hat{\theta}_n$ – состоятельная оценка $\theta$ и с вероятностью стремящейся к единице $\hat{\theta}$ является ОМП.

**Доказательство.** Первая часть следует из доказательства предыдущей теоремы, т.к. $\hat{\theta}_n$ является измеримой функцией, которая не зависит от $\theta_0$. Также, из доказательства предыдущей теоремы событие

$$S_n = \{x:\ f_{\theta_0-\varepsilon}(X_1,\ \ldots,\ X_n) < f_{\theta_0}(X_1,\ \ldots,\ X_n),\\ f_{\theta_0}(X_1,\ \ldots,\ X_n) > f_{\theta_0+\varepsilon}(X_1,\ \ldots,\ X_n)\},$$

где $\theta_0$ – истинное значение параметра, выполняется с вероятностью, стремящейся к единице.




На отрезке $[\theta_0 - \varepsilon,\ \theta_0 + \varepsilon]$ достигается максимум во внутренней точке $\theta_n^*$, и в ней $\dfrac{\partial f}{\partial \theta} = 0$. Так как корень $\hat{\theta}_n$ единственный, то $\hat{\theta}_n = \theta_n^*$.

Докажем, что $\hat{\theta}_n$ – глобальный максимум. Предположим, что существует точка $\tilde{\theta}_n$, в которой значение функции $f$ не меньше, чем в $\hat{\theta}_n$. Тогда, между $\tilde{\theta}_n$ и $\hat{\theta}_n$ будет точка локального минимума, в которой $\dfrac{\partial f}{\partial \theta} = 0$. Но это противоречие с условием единственности корня уравнения правдоподобия.

Таким образом, $\hat{\theta}_n$ – ОМП с большой вероятностью, т.е. на $S_n$. □

# 16 Асимптотическая нормальность оценки максимального правдоподобия в регулярном случае для одномерного параметра (б/д). Теорема Бахадура (б/д). Асимптотическая эффективность оценки максимального правдоподобия. Условия, при которых эффективная оценка параметра является оценкой максимального правдоподобия.

**Теорема 16.1.** (б/д) В условиях (R0)-(R8) любая состоятельная последовательность оценок $\hat{\theta}_n(X_1,\ \ldots,\ X$ являющихся решениями уравнения правдоподобия, удовлетворяют соотношению

$$\forall \theta \in \Theta \hookrightarrow \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \frac{1}{i(\theta)}\right),$$

где $i(\theta)$ – количество информации Фишера в одном наблюдении.

**Замечание 16.2.** Формально перенеся $\sqrt{n}$ в правую часть, получим, что асимптотическая дисперсия будет равняться информации Фишера всей выборки. Таким образом, эту теорему можно воспринимать, как асимптотическую версию неравенства Рао-Крамера.

**Следствие 16.2.1.** (асимптотическая нормальность ОМП) В условиях (R0)-(R8), если $\forall n \in \mathbb{N}\ \forall X_1,\ \ldots,\ X_n$ существует единственное решение $\hat{\theta}_n$ уравнения правдоподобия, и если оно является статистикой (т.е. измеримой функцией выборки), то $\hat{\theta}_n$ – асимптотически нормальная оценка $\theta$ с асимптотической дисперсией $\dfrac{1}{i(\theta)}$ и с вероятностью, стремящейся к единице, является ОМП.

**Теорема 16.3.** (Бахадур, б/д) Пусть выполнены условия регулярности (R0)-(R8), и $\hat{\theta}_n(X_1,\ \ldots,\ X_n)$ асимптотически нормальная оценка $\theta$, т.е. $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \sigma^2(\theta)\right)$, причем $\sigma(\theta)$ непрерывна. Тогда, $\forall \theta \in \Theta \hookrightarrow \sigma^2(\theta) \geqslant \dfrac{1}{i(\theta)}$.

**Замечание 16.4.** Эта теорема также является некоторой асимптотической версией неравенства Крамера-Рао.

**Следствие 16.4.1.** В условиях следствия про асимптотическую нормальность ОМП имеем, что ОМП не хуже любой другой оценки в асимптотическом подходе в классе оценок с непрерывной дисперсией.

**Определение 16.5.** Если $\sqrt{n}(\hat{\theta}_n(X_1,\ \ldots,\ X_n) - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \dfrac{1}{i(\theta)}\right)$, то $\hat{\theta}_n$ называется асимптотически эффективной оценкой $\theta$.

**Утверждение 16.6.** Пусть выполнены условия регулярности для неравенства Крамера-Рао, $\hat{\theta}(X)$ – эффективная оценка $\theta$, и равенство из критерия эффективности (для $\hat{\theta}(X)$) выполнено для любого $X$ и любого $\theta$. Тогда $\hat{\theta}$ – ОМП.

**Доказательство.** По неравенству Крамера-Рао $\hat{\theta}(X) - \theta = \dfrac{1}{I_X(\theta)}\left(\dfrac{\partial}{\partial\theta}\ln f_\theta(X)\right)$. Тогда, если $\theta < \hat{\theta}(X)$, то $\dfrac{\partial}{\partial\theta}\ln f_\theta(X) > 0 \Rightarrow f_\theta(X)$ возрастает. Если же $\theta > \hat{\theta}(X)$, то $f$ убывает. Следовательно, максимум функции правдоподобия достигается в точке $\hat{\theta}(X)$. $\square$

# 17 Линейная регрессионная модель. Оценка наименьших квадратов, ее основные свойства. Теорема о наилучшей оценке в классе линейных оценок (б/д). Несмещенная оценка для дисперсии ошибки измерений $\sigma^2$

В линейной модели наблюдение – случайный вектор $X \in \mathbb{R}^n$, который представляется в виде $X = l + \varepsilon$, где $l$ – неслучайный неизвестный вектор, а $\varepsilon$ – случайный вектор. Тогда, $l$ является оцениваемой величиной, а $\varepsilon$ трактуется как вектор ошибок. Будем считать, что $E\varepsilon = 0,\ D\varepsilon = \sigma^2 I_n,\ \sigma^2 > 0$, при этом $\sigma^2$ неизвестен. Про $l$ известно, что $l \in L$, где $L$ – линейное подпространство в $\mathbb{R}^n$. Задача состоит в оценивании $l$ и $\sigma^2$.

Пусть $L$ задано с помощью базиса $(Z_1,\ \ldots,\ Z_k)$ из вектор-столбцов. Следовательно, $\dim L = k$. Составим матрицу $Z = \begin{pmatrix} Z_1 & \ldots & Z_k \end{pmatrix}$. Тогда, $l = Z\theta$, где $\theta$ – вектор неизвестных координат в базисе $(Z_1,\ \ldots,\ Z_k)$.

Таким образом, задача сведена к оцениванию $\left(\theta,\ \sigma^2\right),\ \theta \in \mathbb{R}^k$.

**Определение 17.1.** Оценка по методу наименьших квадратов для линейной регрессионной модели для параметра $\theta$ – это оценка

$$\hat{\theta} = \arg\min_\theta \|X - Z\theta\|_2^2 = \arg\min_\theta \langle X - Z\theta,\ X - Z\theta\rangle.$$

**Замечание 17.2.** $Z\hat{\theta}$ является проекцией вектора $X$ на подпространство $L$, т.е. $Z\hat{\theta} = proj_L X$.

**Лемма 17.3.** $\hat{\theta} = \left(Z^T Z\right)^{-1} Z^T X$, где $\hat{\theta}$ – оценка МНК параметра $\theta$.

**Доказательство.** Так как функция $\|X - Z\theta\|_2^2$ выпукла, то в минимум достигается в ее точке стационарности. Путем дифференцирования выражения $(X - Z\theta)^T(X - Z\theta) = 0$ и, выражая $\theta$, получаем требуемое. □

**Замечание 17.4.** $Z^T Z$ является матрицей Грама, поэтому она обратима.

**Утверждение 17.5.** $E\hat{\theta} = \theta,\ D\hat{\theta} = \sigma^2 \left(Z^T Z\right)^{-1}$.

**Доказательство.**

$$E\hat{\theta} = E\left(Z^T Z\right)^{-1} Z^T X = \left(Z^T Z\right)^{-1} Z^T EX = \left(Z^T Z\right)^{-1} Z^T E(Z\theta + \varepsilon) = \\ \left(Z^T Z\right)^{-1} Z^T Z\theta = \theta,$$

$$D\hat{\theta} = D\left(Z^T Z\right)^{-1} Z^T X = \left(Z^T Z\right)^{-1} Z^T \cdot DX \cdot Z\left(Z^T Z\right)^{-1} = \\ \left(Z^T Z\right)^{-1} Z^T \cdot \sigma^2 I_n \cdot Z\left(Z^T Z\right)^{-1} = \sigma^2 \left(Z^T Z\right)^{-1}.$$

□

**Замечание 17.6.** Таким образом, оценка МНК является несмещенной.

**Теорема 17.7.** (б/д) Пусть $t = T\theta$ – линейная вектор-функция от $\theta$, $T \in M_{n\times k}$. Тогда оценка $\hat{t} = T\hat{\theta}$ является оптимальной оценкой параметра $t$ в классе линейных несмещенных оценок, т.е. оценок вида $B\cdot X$.

**Лемма 17.8.** $E\|X - Z\hat{\theta}\|_2^2 = \sigma^2(n-k)$.

**Доказательство.** Так как $E(X - Z\hat{\theta}) = 0$, то

$$E\|X - Z\hat{\theta}\|_2^2 = E(X - Z\hat{\theta}, X - Z\hat{\theta}) = E(tr(X - Z\hat{\theta})(X - Z\hat{\theta})^T) = trE(X - Z\hat{\theta})(X - Z\hat{\theta})^T = trD(X - Z\hat{\theta})$$

Рассмотрим

$$D(X - Z\hat{\theta}) = D\left[\left(I_n - Z\left(Z^T Z\right)^{-1} Z^T\right) X\right].$$

Обозначим $A := Z\left(Z^T Z\right)^{-1} Z^T$. Тогда

$$D(X - Z\hat{\theta}) = D[(I_n - A)X] = (I_n - A)\cdot DX \cdot (I_n - A)^T = \\ (I_n - A)\cdot DX \cdot (I_n - A) = \sigma^2(I_n - A)^2 = \sigma^2 I_n - 2\sigma^2 A + \sigma^2 A^2.$$

Заметим, что $A^2 = Z\left(Z^T Z\right)^{-1} Z^T Z\left(Z^T Z\right)^{-1} Z^T = A$, т.е.

$$D(X - Z\hat{\theta}) = \sigma^2(I_n - A).$$

Тогда

$$E\|X - Z\hat{\theta}\|_2^2 = \sigma^2 tr(I_n - A) = \sigma^2\left(trI_n - tr\left(Z\left(Z^T Z\right)^{-1} Z^T\right)\right) = \\ \sigma^2\left(n - tr\left(\left(Z^T Z\right)^{-1} Z^T Z\right)\right) = \sigma^2(n-k).$$

□

**Следствие 17.8.1.**

1. $X - Z\hat{\theta} = proj_{L^\perp} X$
2. $\dfrac{\|X - Z\hat{\theta}\|_2^2}{n-k} = \dfrac{\|proj_{L^\perp} X\|_2^2}{n-k}$ – несмещенная оценка $\sigma^2$.

**Доказательство.**

1. $X = proj_L X + proj_{L^\perp} X \Rightarrow proj_{L^\perp} X = X - Z\hat{\theta}$.
2. $E\dfrac{\|X - Z\hat{\theta}\|_2^2}{n-k} = \dfrac{1}{n-k} E\|X - Z\hat{\theta}\|_2^2 = \dfrac{1}{n-k}\sigma^2(n-k) = \sigma^2$.

□

# 18 Гауссовская линейная модель. Достаточные статистики в гауссовской линейной модели. Наилучшие несмещенные оценки параметров в гауссовской линейной модели, их распределения.

**Определение 18.1.** Если в линейной регрессионной модели $\varepsilon \sim \mathcal{N}\left(0,\ \sigma^2 I_n\right)$, то модель называется *гауссовской линейной моделью.*

**Утверждение 18.2.** $S(X) = \left(proj_L X,\ \|proj_{L^\perp} X\|^2\right)$ – достаточная статистика для $\left(l,\ \sigma^2\right)$.

**Доказательство.** Так как $X$ – гауссовский вектор с некоррелированными компонентами, то его компоненты независимы. Тогда

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

Рассмотрим $\sum\limits_{i=1}^n (X_i - l_i)^2 = \|X - l\|_2^2$. По теореме Пифагора

$$\|X - l\|_2^2 = \|proj_L(X - l)\|_2^2 + \|proj_{L^\perp}(X - l)\|_2^2 = \\ \|proj_L X - proj_L l\|_2^2 + \|proj_{L^\perp} X - proj_{L^\perp} l\|_2^2.$$

Так как $l \in L$, то $\|X - l\|_2^2 = \|proj_L X - l\|_2^2 + \|proj_{L^\perp} X\|_2^2$, и

$$p(X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\|proj_L X - l\|_2^2 + \|proj_{L^\perp} X\|_2^2}{2\sigma^2}\right).$$

Следовательно, $S(X)$ достаточная статистика по критерию факторизации. □

**Теорема 18.3.** (б/д) $S(X)$ – полная статистика.

**Следствие 18.3.1.** $\hat{\theta}$ – оптимальная оценка для $\theta$, $Z\hat{\theta}$ – оптимальная оценка для $l$ и $\dfrac{1}{n-k}\|X - Z\hat{\theta}\|^2$ – оптимальная оценка для $\sigma^2$.

**Доказательство.** Все эти оценки несмещенные и являются функциями от полной достаточной статистики. Так,

$$Z\hat{\theta} = proj_L X \Rightarrow \hat{\theta} = \left(Z^T Z\right)^{-1} Z^T proj_L X.$$

Наконец,

$$\frac{1}{n-k}\|X - Z\hat{\theta}\|^2 = \frac{1}{n-k}\|proj_{L^\perp} X\|^2.$$

□

**Утверждение 18.4.** В гауссовской линейной модели $\hat{\theta}$ не зависит от $X - Z\hat{\theta}$, и

$$\frac{1}{\sigma^2}\|Z\hat{\theta} - Z\theta\|^2 \sim \chi^2_k,$$
$$\frac{1}{\sigma^2}\|X - Z\hat{\theta}\|^2 \sim \chi^2_{n-k}.$$

**Доказательство.** Так как $L \oplus L^{\perp} = \mathbb{R}^n$, то по теореме об ортогональном разложении (19.1) векторы $Z\hat{\theta}$ и $X - Z\hat{\theta}$ являются гауссовскими и независимыми. Причем,

$$\frac{1}{\sigma^2}\|Z\hat{\theta} - E(Z\hat{\theta})\|^2 \sim \chi^2_k,$$
$$\frac{1}{\sigma^2}\|(X - Z\hat{\theta}) - E(X - Z\hat{\theta})\| \sim \chi^2_{n-k}.$$

Из того, что $E(Z\hat{\theta}) = Z\theta$, а $E(X - Z\hat{\theta}) = 0$, получили требуемое.

Докажем независимость. Так как $\hat{\theta} = \left(Z^T Z\right)^{-1} Z^T Z\hat{\theta}$, то $\hat{\theta}$ является функцией от $Z\hat{\theta}$. Следовательно, $\hat{\theta}$ и $X - Z\hat{\theta}$ независимы. $\square$

# 19 Распределения хи-квадрат, Стьюдента и Фишера, их свойства. Теорема об ортогональном разложении гауссовского вектора (б/д). Доверительные интервалы для параметров гауссовской линейной модели.

**Теорема 19.1.** (об ортогональном разложении, б/д) Пусть $(X_1, \ldots, X_n) \sim \mathcal{N}\left(l, \sigma^2 I_n\right)$ и $L_1, \ldots, L_r$ – попарно ортогональные подпространства $\mathbb{R}^n$, причем $L_1 \oplus \ldots \oplus L_r = \mathbb{R}^n$. Обозначим $Y_i = proj_{L_i} X$ – проекция $X$ на $L_i$. Тогда $Y_1, \ldots, Y_r$ – независимые в совокупности нормальные случайные векторы, причем $EY_i = proj_{L_i} l$, и

$$\frac{1}{\sigma^2}\|Y_i - EY_i\|^2 \sim \chi^2_{dim L_i}.$$

**Определение 19.2.** Распределение суммы $k$ квадратов независимых стандартных нормальных случайных величин $\chi^2_k = \xi_1^2 + \ldots \xi_k^2$ называют распределением *хи-квадрат* с $k$ степенями свободы.

**Утверждение 19.3.**

1. $\xi \sim \chi^2_k$, $\eta \sim \chi^2_m$ – независимые случайные величины. Тогда $\xi + \eta \sim \chi^2_{k+m}$
2. Если $\xi \sim \chi^2_k$, то $\mathsf{E}\,\xi = k$ и $\mathsf{D}\,\xi = 2k$.

**Определение 19.4.** Пусть $\xi \sim \mathcal{N}(0, 1)$, $\eta \sim \chi^2_k$ независимые случайные величины. Тогда случайная величина $\zeta = \dfrac{\xi}{\sqrt{\eta/k}}$ имеет *распредение Стьюдента с $k$ степенями* свободы (обозначение $\zeta \sim T_k$).

**Утверждение 19.5.**

1. $\zeta \sim T_k \Rightarrow (-\zeta) \sim T_k$.
2. $T_1$ совпадает с распределением Коши с плотностью $p(x) = \dfrac{1}{\pi\left(1 + x^2\right)}$.
3. $\zeta_k \sim T_k$. Тогда $\zeta_k \xrightarrow[k\to\infty]{d} \mathcal{N}(0, 1)$.

**Определение 19.6.** Пусть $\xi \sim \chi^2_k$, $\eta \sim \chi^2_m$ – независимые случайные величины. Тогда случайная величина $\zeta = \dfrac{\xi/k}{\eta/m}$ *имеет распределние Фишера с параметрами* $k, m$ (обозначение $\zeta \sim F_{k,m}$).

**Утверждение 19.7.**

1. $\xi \sim T_m \Rightarrow \xi^2 \sim F_{1,m}$.
2. $\xi \sim F_{k,m} \Rightarrow \dfrac{1}{\xi} \sim F_{m,k}$.
3. $\xi_m \sim F_{k,m} \Rightarrow k\xi_m \xrightarrow[m\to\infty]{d} \chi^2_k$, где $k$ - фиксированное
4. $\xi_{k,m} \sim F_{k,m} \Rightarrow \xi_{k,m} \xrightarrow[k,m\to\infty]{d} \xi \equiv 1$.

## Доверительные интервалы для параметров гауссовской линейной модели

### Доверительный интервал для $\sigma^2$.

Из того, что $\dfrac{1}{\sigma^2}\|X - Z\hat\theta\|^2 \sim \chi^2_{n-k}$, и $u_{1-\gamma}$ – $(1-\gamma)$-квантиль распределения $\chi^2_{n-k}$, то выполнено

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

### Доверительный интервал для $\theta_j$.

Так как выполнено, что $\hat\theta \sim \mathcal{N}\left(\theta,\ \sigma^2\left(Z^TZ\right)^{-1}\right)$ (так как это линейная функция от гауссовского вектора $X$), то, обозначив $A := \left(Z^TZ\right)^{-1} = (a_{ij})$, то $\hat\theta_j \sim \mathcal{N}\left(\theta_j,\ \sigma^2 a_{jj}\right) \Rightarrow \dfrac{\hat\theta_j - \theta_j}{\sqrt{\sigma^2 a_{jj}}} \sim \mathcal{N}(0,\ 1)$, и $\dfrac{1}{\sigma^2}\|X - Z\hat\theta\|^2 \sim \chi^2_{n-k}$. Тогда $\dfrac{\hat\theta_j - \theta_j}{\sqrt{\sigma^2 a_{jj}}}$ независимо от $\dfrac{1}{\sigma^2}\|X - Z\hat\theta\|^2$ как функция от $\hat\theta$. Следовательно,

$$\sqrt{\frac{n-k}{a_{jj}}}\cdot\frac{\hat\theta_j - \theta_j}{\sqrt{\|X - Z\hat\theta\|^2}} \sim T_{n-k}.$$

Отсюда ищется доверительный интервал для $\theta_j$.

### Доверительная область для $\theta$.

Так как $\dfrac{1}{\sigma^2}\|Z\hat\theta - Z\theta\|^2 \sim \chi^2_k$, $\dfrac{1}{\sigma^2}\|X - Z\hat\theta\|^2 \sim \chi^2_{n-k}$, причем эти случайные величины независимы, и

$$\frac{n-k}{k}\cdot\frac{\|Z\hat\theta - Z\theta\|^2}{\|X - Z\hat\theta\|^2} \sim F_{k,n-k}.$$

Следовательно, доверительная область имеет вид $\left\{\dfrac{n-k}{k}\cdot\dfrac{\|Z\hat\theta - Z\theta\|^2}{\|X - Z\hat\theta\|^2} < u_\gamma\right\}$ – эллипсоид в $\mathbb{R}^k$ (Важно, что $u_\gamma$ – квантиль $F_{k,n-k}$).

# 20 Проверка статистических гипотез: общие принципы и основные понятия (критическое множество, уровень значимости, альтернативы, ошибки первого и второго рода, функция мощности). Наиболее мощные и равномерно наиболее мощные критерии. Несмещенность и состоятельность статистического критерия.

Пусть наблюдение $X$ имеет неизвестное распределение $P \in \mathcal{P}$, где $\mathcal{P}$ – некоторое семейство распределений.

**Определение 20.1.** *Статистическая гипотеза* – это предположение вида $P \in \mathcal{P}_0$, где $\mathcal{P}_0 \subset \mathcal{P}$ – подмножество распределений (обозначение $H_0: \ P \in \mathcal{P}_0$, $H_0$ называется *основной гипотезой*).

Задача заключается по наблюдению $X$ либо принять $H_0$, либо отвергнуть $H_0$. В последнем случае переходим к рассмотрению альтернативы (если она есть). *Альтернативная гипотеза* имеет вид $H_1: \ P \in \mathcal{P}_1, \ \mathcal{P}_1 \subset \mathcal{P} \backslash \mathcal{P}_0$. Если приняли $H_0$, тогда мы сузили класс распределения $\mathcal{P}$ до $\mathcal{P}_0$

**Определение 20.2.** Пусть $X$ принимает значения в выборочном пространстве $\mathcal{X}$, а $S \subset \mathcal{X}$ – некоторое подмножество. Если правило принятия $H_0$ выглядит следующим образом: $H_0$ отвергается тогда и только тогда, когда $X \in S$, то $S$ называется *критерием* для проверки $H_0$ против альтернативы $H_1$, если она есть.

**Определение 20.3.** *Ошибкой первого рода* называется ситуация, когда отвергается гипотеза $H_0$, когда она верна.

**Определение 20.4.** *Ошибкой второго рода* называется ситуация, когда принимается гипотеза $H_0$, когда она неверна.

**Замечание 20.5.** На практике считается, что ошибка первого рода менее желательна, чем ошибка второго рода. Поэтому, $S$ выбирается так, чтобы вероятность ошибки первого рода была меньше заранее выбранной величины, а вероятность ошибки второго рода была как можно меньше.

**Определение 20.6.** Пусть $S$ – критерий для проверки гипотезы $H_0: \ P \in \mathcal{P}_0$. Функция $B(Q, \ S) = Q(X \in S), \ Q \in \mathcal{P}$ называется *функцией мощности критерия.*

**Определение 20.7.** Если для критерия $S$ выполнено неравенство $B(Q, \ S) \leqslant \varepsilon, \ \forall Q \in \mathcal{P}_0$, то говорят, что $S$ имеет *уровень значимости* $\varepsilon$.

**Определение 20.8.** Минимальный уровень значимости $\alpha(S) = \sup\limits_{Q \in \mathcal{P}_0} B(Q, \ S)$ называется *размером критерия.*

**Определение 20.9.** Пусть $S$ – критерий для проверки $H_0: \ P \in \mathcal{P}_0$ против альтернативы $H_1: P \in \mathcal{P}_1$. Критерий $S$ называется *несмещенным*, если $\sup\limits_{Q \in \mathcal{P}_0} B(Q, \ S) \leqslant \inf\limits_{Q \in \mathcal{P}_1} B(Q, \ S)$.

**Определение 20.10.** Если $X = (X_1, \ \ldots, \ X_n)$ – выборка растущего размера, то последовательность критериев $S_n$ называется *состоятельной*, если $B(Q, \ S_n) \xrightarrow[n \to \infty]{} 1, \ \forall Q \in \mathcal{P}_1$.

**Определение 20.11.** Критерий $S$ уровня значимости $\varepsilon$ называется *более мощным*, чем критерий $R$ того же уровня значимости $\varepsilon$, если $B(Q, \ S) \geqslant B(Q, \ R) \ \forall Q \in \mathcal{P}_1$, т.е. вероятность ошибки второго рода у $S$ равномерно меньше.

**Определение 20.12.** Критерий $S$ называют равномерно наиболее мощным (р.н.м.к.) уровня значимости $\varepsilon$, если $\alpha(S) \leqslant \varepsilon$, и $S$ мощнее любого другого критерия $R$, который удовлетворяет условию $\alpha(R) \leqslant \varepsilon$.

# 21 Лемма Неймана–Пирсона. Построение с ее помощью наиболее мощных критериев. Теорема о монотонном отношении правдоподобия (б/д). Построение равномерно наиболее мощных критериев для односторонних альтернатив. Двойственность доверительного оценивания и проверки гипотез.

**Определение 21.1.** Гипотеза $H:\ P = P_0$, где $P_0$ – известное распределение, называется *простой*.

Пусть заданы простые гипотезы $H_0:\ P = P_0$ и альтернатива $H_1:\ P = P_1$, причем $P_i$ имеют плотность $p_i(x)$ по одной и той же мере $\mu$. Рассмотрим критерий для $\lambda > 0$:

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

**Лемма 21.2.** (Неймана-Пирсона) Пусть критерий $R$ удовлетворяет соотношению $P_0(X \in R) \leqslant P_0(X \in S_\lambda)$. Тогда $P_1(X \in R) \leqslant P_1(X \in S_\lambda)$, т.е. $S_\lambda$ мощнее $R$, и $P_0(X \in S_\lambda) \leqslant P_1(X \in S_\lambda)$, т.е. $S_\lambda$ – несмещенный.

**Доказательство.** Рассмотрим следующее соотношение

$$I_R(x)(p_1(x) - \lambda p_0(x)) \leqslant I_R(x)(p_1(x) - \lambda p_0(x))\cdot I(\{x:\ p_1(x) - \lambda p_0(x) \geqslant 0\}) =$$
$$I_R(x)(p_1(x) - \lambda p_0(x))I_{S_\lambda}(x) \leqslant (p_1(x) - \lambda p_0(x))I_{S_\lambda}(x).$$

Рассмотрим первое неравенство. Если $p_1(x) - \lambda p_0(x) \geqslant 0$, то обе части равны. Если $p_1(x) - \lambda p_0(x) < 0$, то левая часть меньше нуля, а правая равна нулю.

Рассмотрим второе неравенство. Если $x \in R$, то обе части равны. Иначе, левая часть равна нулю, а правая часть неотрицательна в силу выбора критерия $S_\lambda$.

Тогда

$$P_1(X \in R) - \lambda P_0(X \in R) = E_1 I_R(X) - \lambda E_0 I_R(X) =$$
$$\int_{\mathcal{X}} I_R(x)p_1(x)d\mu - \lambda \int_{\mathcal{X}} I_R(x)p_0(x)d\mu = \int_{\mathcal{X}} I_R(x)(p_1(x) - \lambda p_0(x))d\mu \leqslant$$
$$\int_{\mathcal{X}} I_{S_\lambda}(p_1(x) - \lambda p_0(x))d\mu = P_1(X \in S_\lambda) - \lambda P_0(X \in S_\lambda) \Rightarrow$$
$$P_1(X \in S_\lambda) - P_1(X \in R) \geqslant \lambda(P_0(X \in S_\lambda) - P_0(X \in R)) \geqslant 0.$$

Докажем несмещенность. Если $\lambda \geqslant 1$, то $p_0(x) \leqslant p_1(x)\ \forall x \in S_\lambda$. Тогда

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

Если $\lambda < 1$, то $p_1(x) \leqslant p_0(x)\ \forall x \in \overline{S}_\lambda$. Тогда

$$1 - P_1(X \in S_\lambda) = P_1(X \in \overline{S}_\lambda) = \int_{\mathcal{X}} I_{\overline{S}_\lambda}(x)p_1(x)d\mu \leqslant \int_{\mathcal{X}} I_{\overline{S}_\lambda}(x)p_0(x)d\mu =$$
$$P_0(X \in \overline{S}_\lambda) = 1 - P_0(X \in S_\lambda) \Rightarrow P_0(X \in S_\lambda) \leqslant P_1(X \in S_\lambda).$$

□

**Следствие 21.2.1.** Если $\lambda > 0$ удовлетворяет условию $P_0(X \in S_\lambda) = \varepsilon$, то $S_\lambda$ – равномерно наиболее мощный критерий размера $\varepsilon$. Таким образом, для нахождения равномерно наиболее мощного критерия необходимо решить уравнение

$$\int_A p_0(x)d\mu = \varepsilon,$$

где $A := \{x:\ p_1(x) - \lambda p_0(x) \geqslant 0\}$.

**Замечание 21.3.** В абсолютно непрерывном случае, как правило, решение есть.

**Монотонное отношение правдоподобия**

Пусть семейство распределений параметризовано параметром $\Theta \subset \mathbb{R}: \ \mathcal{P} = \{P_\theta, \ \theta \in \Theta\}$. Пусть $\mathcal{P}$ доминируемое семейство, т.е. существует функция правдоподобия $f_\theta(x)$.

**Определение 21.4.** Семейство $\{P_\theta, \ \theta \in \Theta\}$ называется *семейством с монотонным отношением правдоподобия* по статистике $T(X)$, если $\forall \theta_0 < \theta_1$ функция $\dfrac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$ является монотонной функцией от $T(X)$, причем тип монотонности один и тот же для всех $\theta_1 > \theta_0$.

**Теорема 21.5.** (о монотонном отношении правдоподобия, б/д) Пусть даны гипотезы $H_0 : \theta \leqslant \theta_0 \ (\theta = \theta_0), \ H_1 : \theta > \theta_0$, и семейство $\{P_\theta, \ \theta \in \Theta\}$ с монотонным отношением правдоподобия, причем $\dfrac{f_{\theta_1}(x)}{f_{\theta_2}(x)}$ не убывает по $T(X)$ при $\theta_1 > \theta_2$. Тогда критерий $S_\varepsilon = \{T(X) \geqslant c_\varepsilon\}$ с условием $P_{\theta_0}(X \in S_\varepsilon) = \varepsilon$ является равномерно наиболее мощным критерием уровня значимости $\varepsilon$ для проверки $H_0$ против $H_1$.

**Двойственность доверительного оценивания и проверки гипотез**

**Утверждение 21.6.** Пусть $S(X)$ – доверительная область уровня доверия $1 - \varepsilon$ для параметра $\theta \in \Theta$. Хотим проверить простую гипотезу $H_0 : \ \theta = \theta_0$. Рассмотрим $\tilde{S}(\theta) = \{X \in \mathcal{X} : \ \theta \notin S(X)\}$. Тогда $\tilde{S}(\theta_0)$ – критерий уровня значимости $\varepsilon$ для проверки $H_0$.

**Доказательство.**

$$P_{\theta_0}\left(X \in \tilde{S}(\theta_0)\right) = P_{\theta_0}(\theta_0 \notin S(X)) = 1 - P_{\theta_0}(\theta_0 \in S(X)) \leqslant \varepsilon.$$

□

**Утверждение 21.7.** Пусть $S_{\theta_0}$ – критерий уровня значимости $\varepsilon$ для проверки $H_0 : \ \theta = \theta_0$, который известен $\forall \theta_0 \in \Theta$. Рассмотрим $S(X) = \{\theta \in \Theta : \ X \notin S_\theta\}$. Тогда $S(X)$ – доверительное множество уровня доверия $1 - \varepsilon$.

**Доказательство.**

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

□

# 22 F-критерий для проверки линейных гипотез в гауссовской линейной модели. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

**Проверка гипотез в гауссовской линейной модели**

Цель этого параграфа – построить критерий для проверки линейной гипотезы $H_0 : \ T\theta = t$, где $T \in M_{m \times k}, \ t \in \mathbb{R}^m, \ rk(T) = m \leqslant k$.

**Утверждение 22.1.** Пусть $H_0$ верна. Тогда выполнено

$$\frac{(T\hat{\theta} - t)^T \left(T \left(Z^T Z\right)^{-1} T^T\right)^{-1} (T\hat{\theta} - t)}{\|X - Z\hat{\theta}\|^2} \cdot \frac{n-k}{m} \ \sim \ F_{m,n-k}.$$

**Доказательство.** Так как $\hat{\theta} \sim \mathcal{N}\left(\theta,\ \sigma^2\left(Z^TZ\right)^{-1}\right)$, где $\hat{\theta} = \left(Z^TZ\right)^{-1}Z^TX$ – ОНК для $\theta$, $\hat{t} = T\hat{\theta}$ – оптимальная оценка для $T\theta$, то

$$\hat{t} = T\hat{\theta} \sim \mathcal{N}\left(T\theta,\ T\sigma^2\left(Z^TZ\right)^{-1}T^T\right) = \mathcal{N}\left(T\theta,\ \sigma^2B\right),$$

где $B = T\left(Z^TZ\right)^{-1}T^T$. Матрица $B$ положительно определена и симметрична. Следовательно, $\exists\sqrt{B}:\ B = \sqrt{B}\cdot\sqrt{B},\ \left(\sqrt{B}\right)^T = \sqrt{B}$. Тогда

$$\frac{1}{\sigma}\left(\sqrt{B}\right)^{-1}(\hat{t} - T\theta) \sim \mathcal{N}(0,\ I_m) \Rightarrow \left\|\frac{1}{\sigma}\left(\sqrt{B}\right)^{-1}(\hat{t} - T\theta)\right\|^2 =$$

$$\frac{1}{\sigma^2}(\hat{t} - T\theta)^TB^{-1}(\hat{t} - T\theta) \sim \chi^2_m.$$

Рассмотрим статистику $\hat{Q}_T := \frac{1}{\sigma^2}(\hat{t} - t)^TB^{-1}(\hat{t} - t)$. Тогда в условиях гипотезы $H_0$ выполнено $\frac{1}{\sigma^2}\hat{Q}_T \sim \chi^2_m$. Так как $\hat{Q}_T$ выражается через $\hat{\theta}$, то она не зависит от $X - Z\hat{\theta}$. В силу того, что $\frac{1}{\sigma^2}\|X - Z\hat{\theta}\|^2 \sim \chi^2_{n-k}$, верно

$$\frac{\hat{Q}_T}{\|X - Z\hat{\theta}\|^2}\cdot\frac{n-k}{m} \sim F_{m,n-k}.$$

□

**Определение 22.2.** Пусть $u_{1-\alpha}$ – $(1-\alpha)$-квантиль распределения $F_{m,n-k}$. Тогда *F-критерием* называется

$$\left\{\frac{(T\hat{\theta} - t)^T\left(T\left(Z^TZ\right)^{-1}T^T\right)^{-1}(T\hat{\theta} - t)}{\|X - Z\hat{\theta}\|^2}\cdot\frac{n-k}{m} > u_{1-\alpha}\right\}.$$

**Пример 22.3.** Пусть $X_1,\ \ldots,\ X_n \sim \mathcal{N}\left(a_1,\ \sigma^2\right),\ Y_1,\ \ldots,\ Y_m \sim \mathcal{N}\left(a_2,\ \sigma^2\right)$. Построим F-критерий для проверки гипотезы $H_0:\ a_1 = a_2$. Сведем задачу к гауссовской модели линейной регрессии:

$$W := \begin{pmatrix} X_1 \\ \ldots \\ X_n \\ Y_1 \\ \ldots \\ Y_m \end{pmatrix} = Z\cdot\theta + \varepsilon,$$

где

$$Z = \begin{pmatrix} 1 & 0 \\ \ldots & \ldots \\ 1 & 0 \\ 0 & 1 \\ \ldots & \ldots \\ 0 & 1 \end{pmatrix},\ \theta = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \ldots \\ \varepsilon_n \\ \varepsilon_{n+1} \\ \ldots \\ \varepsilon_{n+m} \end{pmatrix},\ T = \begin{pmatrix} 1 & -1 \end{pmatrix},\ t = 0,$$

и $\varepsilon_j \sim \mathcal{N}\left(0,\ \sigma^2\right),\ cov(\varepsilon_i,\ \varepsilon_j) = 0,\ i \neq j$. Тогда $\hat{\theta} = \left(Z^TZ\right)^{-1}Z^TW = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix}$, и $\hat{t} = T\hat{\theta} = \overline{X} - \overline{Y}$. Следовательно, $T\left(Z^TZ\right)^{-1}T^T = \frac{1}{n} + \frac{1}{m}$. Таким образом, $\hat{Q} = \left(\frac{1}{n} + \frac{1}{m}\right)^{-1}(\overline{X} - \overline{Y})^2$, и $\|W - Z\hat{\theta}\|^2 =$

$\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{j=1}^{m}(Y_j-\overline{Y})^2=nS_X^2+mS_Y^2$. Тогда F-критерий имеет вид

$$\left\{\frac{nm}{n+m}\cdot\frac{(\overline{X}-\overline{Y})^2}{nS_X^2+mS_Y^2}\cdot\frac{n+m-2}{1}>u_{1-\varepsilon}\right\},$$

где $u_{1-\varepsilon}$ – $(1-\varepsilon)$-квантиль распределения $F_{1,n+m-2}$.

# 23 Критерий хи-квадрат Пирсона. Теорема Пирсона (формулировка). Состоятельность критерия хи-квадрат. Критерий Колмогорова. Критерий фон Мизеса-Смирнова.

**Критерий согласия Пирсона (критерий хи-квадрат)**

Пусть выборка $X_1,\ \ldots,\ X_n$ с распределением $P(X_1=a_i)=p_i,\ i=\overline{1,m},\ \sum_{i=1}^{m}p_i=1$ (полиномиальная схема). Положим $\nu_j=\sum_{i=1}^{n}I_{X_i=a_j},\ j=\overline{1,m}$ – количество осуществлений исхода $a_j$. Тогда $\sum_{i=1}^{m}\nu_i=n$. Вектор $\overline{p}=\begin{pmatrix}p_1\\ \ldots\\ p_m\end{pmatrix}$ считается неизвестным. Хотим проверить простую гипотезу $H_0:\ \overline{p}=\overline{p}_0$, где $\overline{p}_0=\begin{pmatrix}p_1^0\\ \ldots\\ p_m^0\end{pmatrix}$ – заданный вектор с положительными координатами, и $\sum_{i=1}^{m}p_i^0=1$.

**Определение 23.1.** *Статистикой хи-квадрат Пирсона* называют величину

$$\hat{\chi}_n^2=\sum_{j=1}^{m}\frac{\left(\nu_j-np_j^0\right)^2}{np_j^0}.$$

**Теорема 23.2.** (Пирсона) Если $H_0$ верна, то $\hat{\chi}_n^2\xrightarrow[n\to\infty]{d}\chi_{m-1}^2$.

**Определение 23.3.** Пусть $u_{1-\varepsilon}$ – $(1-\varepsilon)$-квантиль $\chi_{m-1}^2$. Тогда *критерием хи-кварат называется* $\left\{\hat{\chi}_n^2>u_{1-\varepsilon}\right\}$.

**Замечание 23.4.** На практике критерий применяется при $np_j^0\geqslant 5,\ 1\leqslant j\leqslant m$.

**Замечание 23.5.** Критерий хи-квадрат является асимптотическим, т.е. вероятность ошибки только стремится к $\varepsilon$ при $n\to\infty$.

**Утверждение 23.6.** Критерий хи-квадрат состоятелен.

**Доказательство.** Пусть на самом деле $\overline{p}\neq\overline{p}_0$. Тогда $\hat{\chi}_n^2=n\sum_{i=1}^{m}\left(\frac{\nu_i}{n}-p_i^0\right)^2\frac{1}{p_i^0}$. В силу УЗБЧ для любого фиксированного $i\in\{1,\ \ldots,\ m\}$ выполнено $\frac{\nu_i}{n}=\frac{1}{n}\sum_{j=1}^{n}I_{X_j=a_i}\xrightarrow{a.s.}p_i$. По теореме о наследовании сходимости $\sum_{i=1}^{m}\left(\frac{\nu_i}{n}-p_i^0\right)^2\cdot\frac{1}{p_i^0}\xrightarrow{a.s.}\sum_{i=1}^{m}\frac{\left(p_i-p_i^0\right)^2}{p_i^0}>0$. Значит, $\hat{\chi}_n^2\xrightarrow[n\to\infty]{}\infty$, а, следовательно, и $P\left(\hat{\chi}_n^2>u_{1-\varepsilon}\right)\xrightarrow[n\to\infty]{}1$. Таким образом, вероятность ошибки второго рода стремится к нулю. □

**Критерий согласия Колмогорова (Колмогорова-Смирнова)**

Пусть $X=(X_1,\ \ldots,\ X_n)$ – выборка из неизвестного распределения с непрерывной функцией распределения $F$. Тогда $D_n=\sup_{x\in\mathbb{R}}|F_n^*(x)-F(x)|$.

**Теорема 23.7.** (Колмогорова, б/д) Пусть $F$ – непрерывна. Тогда

1) Распределение $\sqrt{n}D_n$ не зависит от вида $F$

2) $\sqrt{n}D_n \xrightarrow{d} \xi$, где $\xi$ имеет распределение Колмогорова, т.е. $P(\xi \leqslant z) = \sum\limits_{j\in\mathbb{Z}}(-1)^j e^{-2j^2z^2} I_{z>0}$.

Рассмотрим гипотезу $H_0: \ F = F_0$. - непрерывна

**Определение 23.8.** *Критерием Колмогорова* называется

$$S = \left\{\sqrt{n}\sup_{x\in\mathbb{R}} |F_n^*(x) - F_0(x)| > K_{1-\alpha}\right\},$$

где $K_{1-\alpha}$ – $(1-\alpha)$-квантиль распределения Колмогорова.

**Замечание 23.9.** На практике критерий применяется при $n \geqslant 20$.

**Критерий омега-квадрат (или критерий фон Мизеса-Смирнова)**

Пусть гипотеза $H_0: \ F = F_0$, где $F_0$ – непрерывная функция распределения.

**Определение 23.10.** *Статистикой омега-квадрат* называется

$$\omega^2 := n\int |F_n^*(x) - F_0(x)|^2\, dF_0(x).$$

**Утверждение 23.11.** (б/д) Если гипотеза $H_0$ верна, то статистика $\omega^2$ слабо сходится к известному распределению.

**Определение 23.12.** *Критерием омега-квадрат* называется $S := \left\{\omega^2 > u_{1-\alpha}\right\}$, где $u_{1-\alpha}$ – $(1-\alpha)$-квантиль распределения из предыдущего утверждения.

**Определение 23.13.** Критерии Пирсона, Колмогорова и омега-квадрат называют *критериями согласия*, так как они проверяют гипотезу $H_0: \ P = P_0$ против альтернативы $H_1: \ P \neq P_0$.

# 24 Доказательство теоремы Пирсона.

**Теорема 24.1.** (Пирсона) Если $H_0$ верна, то $\hat{\chi}_n^2 \xrightarrow[n\to\infty]{d} \chi^2_{m-1}$.

**Доказательство.** Положим $Y_j = \begin{pmatrix} I_{X_j=a_1} \\ \dots \\ I_{X_j=a_m}\end{pmatrix}$, $j = \overline{1,n}$ – независимые одинаково распределенные случайные векторы, $EY_j = \begin{pmatrix} p_1 \\ \dots \\ p_m\end{pmatrix} = \begin{pmatrix} p_1^0 \\ \dots \\ p_m^0\end{pmatrix}$. Найдем матрицу ковариаций вектора $Y_j$:

$$cov(I_{X_j=a_i},\ I_{X_j=a_k}) = EI_{X_j=a_i,\ X_j=a_k} - EI_{X_j=a_i}\cdot EI_{X_j=a_k} = \begin{cases} p_i^0 - p_i^0p_k^0 & ,\ i=k \\ -p_i^0p_k^0 & ,\ i\neq k\end{cases}.$$

Положим $B := \begin{pmatrix} p_1^0 & 0 & \dots & 0 \\ 0 & p_2^0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & p_m^0\end{pmatrix} \Rightarrow DY_j = B - \overline{p}_0\overline{p}_0^T$. По многомерной ЦПТ

$$\frac{Y_1 + \ \dots\ + Y_n}{\sqrt{n}} - \sqrt{n}\overline{p}_0 \xrightarrow{d} \mathcal{N}\left(0,\ B - \overline{p}_0\overline{p}_0^T\right).$$

Заметим, что $Y_1 + \ \dots \ + Y_n = \begin{pmatrix} \nu_1 \\ \dots \\ \nu_m \end{pmatrix} =: \overline{\nu}$, и матрица $B$ симметрична и положительно определена в силу положительности $p_j^0 \ \forall j \in \{1, \ \dots, \ m\}$. Пусть $\xi_n = \left(\sqrt{B}\right)^{-1} \sqrt{n} \left(\frac{\overline{\nu}}{n} - \overline{p}_0\right)$. По теореме о наследовании сходимостей

$$\xi_n \xrightarrow{d} \mathcal{N}\left(0, \ \left(\sqrt{B}\right)^{-1} \left(B - \overline{p}_0 \overline{p}_0^T\right) \left(\sqrt{B}\right)^{-1}\right) = \mathcal{N}\left(0, \ I_m - zz^T\right),$$

где $z = \left(\sqrt{B}\right)^{-1} \overline{p}_0 = \begin{pmatrix} \sqrt{p_1^0} \\ \dots \\ \sqrt{p_m^0} \end{pmatrix}$. Заметим, что $\|z\|^2 = \sum\limits_{i=1}^{m} p_i^0 = 1 \Rightarrow \|z\| = 1$. Рассмотрим ортогональную матрицу $V \in M_{m \times m}$ такую, что ее первая строка – это вектор $z$. Тогда по теореме о наследовании сходимости

$$V\xi_n \xrightarrow{d} V \cdot \mathcal{N}\left(0, \ I_m - zz^T\right) = \mathcal{N}\left(0, \ V\left(I_m - zz^T\right) V^T\right).$$

Рассмотрим вектор $Vz$. В силу ортогональности матрицы $V$ выполнено

$$Vz = \begin{pmatrix} \sqrt{p_1^0} & \sqrt{p_2^0} & \dots & \sqrt{p_m^0} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & \dots & \dots & \dots \\ v_{m1} & v_{m2} & \dots & v_{mm} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{p_1^0} \\ \sqrt{p_2^0} \\ \dots \\ \sqrt{p_m^0} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}.$$

Таким образом,

$$V\left(I_m - zz^T\right) V^T = I_m - \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} =: \tilde{I}_m.$$

По теореме о наследовании сходимости

$$\|V\xi_n\|^2 \xrightarrow{d} \left\|\mathcal{N}\left(0, \ \tilde{I}_m\right)\right\|^2 = \chi^2_{m-1}.$$

Но в силу ортогональности матрицы $V$ выполнено

$$\|V\xi_n\|^2 = \|\xi_n\|^2 = \left\|\left(\sqrt{B}\right)^{-1} \sqrt{n} \left(\frac{\overline{\nu}}{n} - \overline{p}_0\right)\right\|^2 = \sum_{i=1}^{m} \frac{\left(\nu_i - np_i^0\right)^2}{np_i^0} = \hat{\chi}^2_n.$$

$\square$

# 25 Байесовские оценки. Теорема о наилучшей оценке в байесовском подходе.

Пусть $\Theta \subset \mathbb{R}^d$, $\{P_\theta, \ \theta \in \Theta\}$ – доминируемое семейство распределений относительно меры $\mu$, т.е. $P_\theta$ имеет плотность $p_\theta(x)$ относительно меры $\mu$. В байесовском подходе параметр $\theta$ является случайной величиной (вектором) с известным априорным распределением $Q$ на множестве $\Theta$.

Пусть $Q$ – распределение на $(\Theta, \ \mathcal{B}_\Theta)$ с плотностью $q(t)$. Рассматриваем вероятностное пространство $\left(\Theta \times \mathcal{X}, \ \mathcal{B}_\Theta \times B_\mathcal{X}, \ \tilde{P}\right)$, где мера $\tilde{P}$ – вероятностная мера с плотностью $f(t, \ x) = q(t)p_t(x)$. Таким образом, $(\theta, \ X)$ – случайный вектор на вероятностном пространстве с плотностью $f(t, \ x)$. При таком подходе $p_t(x)$ – условная плотность $X$ при условии, что $\theta = t$.

## 25.1 Математическая модель эксперимента

Рассмотрим $\theta(t) = t$ – случайная величина (вектор) на $(\Theta,\ \mathcal{B}_\Theta,\ Q)$, и пусть $X$ – случайная величина, определенная на $(\mathcal{X},\ \mathcal{B}_\mathcal{X},\ P_t)$, определенная как тождественное отображение $X(x) = x$. Тогда $(\theta,\ X)$ – случайный вектор на $\left(\Theta \times \mathcal{X},\ \mathcal{B}_\Theta \times \mathcal{B}_\mathcal{X},\ \tilde{P}\right)$ с плотностью $f(t,\ x)$, причем $q(t)$ – плотность $\theta$, а $p_t(x)$ – условная плотность $X$ при условии $\theta = t$.

**Определение 25.1.** Плотность $q(t)$ называется *априорной плотностью* параметра $\theta$.

**Определение 25.2.** Условная плотность $\theta$ относительно $X$

$$q(t \mid x) = \frac{q(t)p_t(x)}{\int_\Theta q(u)p_u(x)du} = \frac{f(t,\ x)}{\int_\Theta f(u,\ x)du}$$

называется *апостериорной плотностью.*

**Определение 25.3.** Оценка $\hat{\theta}(X) = \displaystyle\int_\Theta tq(t \mid x)dt = E_{\tilde{P}}(\theta \mid x)$ называется *байесовской оценкой* $\theta$.

**Теорема 25.4.** (о наилучшем среднеквадратичном прогнозе, б/д) Пусть $\mathcal{G} \subset \mathcal{F}$, $\xi$ – случайная величина на $(\Omega, \mathcal{F}, P)$, $\mathcal{L}$ – множество $\mathcal{G}$-измеримых случайных величин с конечным математическим ожиданием. Тогда выполнено

$$\arg\min_{\eta \in \mathcal{L}} E(\xi - \eta)^2 = E(\xi \,|\, \mathcal{G}) \text{ п.н.}$$

**Теорема 25.5.** Байесовская оценка является наилучшей оценкой в байесовском подходе с квадратичной функцией потерь.

**Доказательство.** Пусть $\hat{\theta}$ – байесовская оценка параметра $\theta$. Тогда

$$\int_\Theta R(\hat{\theta},\ \theta)q(\theta)d\theta = \int_\Theta E_\theta(\hat{\theta} - \theta)^2 q(t)dt = \int_\Theta \int_\mathcal{X} (\hat{\theta} - \theta)^2 f(t,\ x)dxdt = E_{\tilde{P}}(\hat{\theta}(X) - \theta)^2.$$

По теореме о наилучшем среднеквадратичном прогнозе

$$\arg\min_{\hat{\theta}} E_{\tilde{P}}(\hat{\theta}(X) - \theta)^2 = E_{\tilde{P}}(\theta \,|\, x).$$

□

# 26 Коэффициенты корреляции Пирсона, Спирмена и Кендалла. Их свойства

## 26.1 Коэффициент корреляции Пирсона

Пусть у нас есть две выборки X, Y одинаковых размеров. Нас интересует гипотеза о независимости X, Y

$$H_0 : F_{XY}(s,t) = F_X(s)F_Y(t)\ \forall t, s$$

Пусть $EX_i^2, EY_i^2 < \infty$

**Определение 26.1.** Величина

$$\hat{\rho} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2 \sum_{i=1}^n (Y_i - \overline{Y})^2}}$$

называется *коэффициентом корреляции Пирсона.*

**Утверждение 26.2.** $\hat{\rho} \xrightarrow{a.s.} \rho = \dfrac{cov(X_1,\ Y_1)}{\sqrt{DX_1 DY_1}}$.

**Доказательство.** $S_X^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2 = \overline{X^2} - (\overline{X})^2$. Из УЗБЧ $\begin{cases} \overline{X^2} \xrightarrow{a.s.} EX_1^2 \\ \overline{X} \xrightarrow{a.s.} EX_1 \end{cases}$. Значит, по теореме о наследовании сходимости $S_X^2 \xrightarrow{a.s.} EX_1^2 - (EX_1)^2 = DX_1$. Аналогично, $S_Y^2 \xrightarrow{a.s.} DY_1$. Тогда

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

$$\frac{\sum_{i=1}^{n} X_i Y_i}{n} - \overline{X}\overline{Y}.$$

Следовательно, по УЗБЧ

$$\frac{\sum_{i=1}^{n} X_i Y_i}{n} - \overline{X}\overline{Y} \xrightarrow{a.s.} E(X_1 Y_1) - EX_1 EY_1 = cov(X_1,\ Y_1).$$

□

**Теорема 26.3.** (б/д) Пусть $n > 2$, $X,\ Y$ – независимые выборки, имеющие нормальное распределение. Тогда $T := \hat{\rho}\sqrt{\dfrac{n-2}{1-\hat{\rho}^2}} \sim T_{n-2}$.

В условиях предыдущей теоремы критерий для проверки $H_0$ имеет вид $S := \{|T| > const\}$. const - это квантиль

## 26.2 Коэффициент корреляции Спирмена

Пусть $X_1,\ \ldots,\ X_n$ – выборка из непрерывного распределения. Тогда $P(X_i \neq X_j) = 1$. Упорядочим элементы выборки по возрастанию.

**Определение 26.4.** Номера, которые получили элементы выборки после упорядовачивания, называются *рангами, т.е.* $R(X_i)$ – номер $X_i$ в вариационном ряду.

Если $(r_1,\ \ldots,\ r_n) \in S_n$, то $P(R(X_1) = r_1,\ \ldots,\ R(X_n) = r_n) = \dfrac{1}{n!}$. Положим $R_i = R(X_i)$. Аналогично рассмотрим $Y_1,\ \ldots,\ Y_n$ – выборка из непрерывного распределения и $S_1,\ \ldots,\ S_n$ – соответствующие ранги. Положим

$$\overline{R} = \frac{\sum_{i=1}^{n} R_i}{n} = \frac{\sum_{i=1}^{n} i}{n} = \frac{n+1}{2},\ \overline{S} = \frac{\sum_{i=1}^{n} S_i}{n} = \frac{n+1}{2}.$$

**Определение 26.5.** *Коэффициентом корреляции Спирмена* называется величина

$$\rho_S = \frac{\sum_{i=1}^{n}\left(R_i - \overline{R}\right)\left(S_i - \overline{S}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_i - \overline{R}\right)^2 \sum_{i=1}^{n}\left(S_i - \overline{S}\right)^2}}.$$

**Замечание 26.6.**

$$\sum_{i=1}^{n}\left(R_i - \overline{R}\right)^2 = \sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)^2 = \sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^2 =$$

$$\sum_{i=1}^{n} i^2 - (n+1)\sum_{i=1}^{n} i + \frac{n(n+1)^2}{4} = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{2} + \frac{n(n+1)^2}{4} =$$

$$\frac{n^3 - n}{12} = \sum_{i=1}^{n}\left(S_i - \overline{S}\right)^2.$$

**Утверждение 26.7.**
$$\rho_S = 1 - \frac{6}{n^3-n}\sum_{i=1}^{n}(R_i - S_i)^2.$$

**Доказательство.** Определим $T_i$: переставим пары $(R_i,\ S_i)$ в порядке возрастания первой компоненты. В итоге получим пары $(1,\ T_1),\ (2,\ T_2),\ \ldots,\ (n,\ T_n)$. Таким образом из пары $(R_i,\ S_i)$ образовалась пара $(k,\ T_k)$, т.е. $(R_i = k) \Leftrightarrow (T_k = S_i)$. В силу замечания

$$\rho_S = \frac{12}{n^3-n}\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right) = \frac{12}{n^3-n}\sum_{k=1}^{n}\left(k - \frac{n+1}{2}\right)\left(T_k - \frac{n+1}{2}\right) =$$
$$\frac{12}{n^3-n}\left(\sum_{k=1}^{n}kT_k - \frac{n+1}{2}\sum_{k=1}^{n}T_k - \frac{n+1}{2}\sum_{k=1}^{n}k + n\left(\frac{n+1}{2}\right)^2\right) =$$
$$\frac{6}{n^3-n}\left(2\sum_{k=1}^{n}kT_k - \frac{n(n+1)^2}{2}\right) =$$
$$\frac{6}{n^3-n}\left(-\sum_{k=1}^{n}k^2 + 2\sum_{k=1}^{n}kT_k - \sum_{k=1}^{n}T_k^2 + 2\frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{2}\right) =$$
$$1 - \frac{6}{n^3-n}\sum_{k=1}^{n}(k - T_k)^2 = 1 - \frac{6}{n^3-n}\sum_{i=1}^{n}(R_i - S_i)^2.$$

□

**Замечание 26.8.** $(T_1,\ \ldots,\ T_n) \in S_n$. Причем, если гипотеза $H_0$ верна, то все перестановки равновероятны.

**Утверждение 26.9.** Если $H_0$ верна, то $E\rho_S = 0$.

**Доказательство.** Так как $ET_k = \sum_{j=1}^{n} jP(T_k = j) = \sum_{j=1}^{n} j\frac{1}{n} = \frac{n+1}{2}$, то

$$E\rho_S = \frac{12}{n^3-n}\sum_{k=1}^{n}\left(k - \frac{n+1}{2}\right)E\left(T_k - \frac{n+1}{2}\right) = 0.$$

□

**Упражнение 26.10.** $D\rho_S = \frac{1}{n-1}$.

**Утверждение 26.11.** $\rho_s \in [-1,\ 1]$, причем крайние значения достигаются.

**Доказательство.** Ограниченность следует из неравенста Коши-Буняковского. $\rho_s = 1$, когда $R_i = S_i\ \forall i \in \{1,\ \ldots,\ n\}$, $\rho_S = -1$, если $R_i = n + 1 - S_i$. □

**Утверждение 26.12.** При $H_0$ распределение $\rho_S$ известно, не зависит от конкретных функций распределений $F_X$ и $F_Y$, и его квантиль есть в таблицах.

**Доказательство.** Следует из того, что распределение $\rho_S$ зависит только от распределения $T_k$, которое известно. □

**Утверждение 26.13.** (б/д) Если $H_0$ верна, то $\frac{\rho_S}{\sqrt{D\rho_S}} \xrightarrow{d} \mathcal{N}(0,\ 1)$.

**Замечание 26.14.** Нормальное приближение применяют при $n \geqslant 50$.

Критерий для проверки $H_0$ имеет вид $\{|\rho_S| > c\}$, где $c$ является квантилью распределения $\rho_S$.

## 26.3 Коэффициент корреляции Кендалла

Пусть $X = (X_1,\ \dots,\ X_n)$, $Y = (Y_1,\ \dots,\ Y_n)$ – выборки.

**Определение 26.15.** Пары $\begin{pmatrix} X_i \\ Y_i \end{pmatrix}$ и $\begin{pmatrix} X_j \\ Y_j \end{pmatrix}$ при $1 \leqslant i < j \leqslant n$ называются *согласованными*, если $\operatorname{sgn}(X_i - X_j) \operatorname{sgn}(Y_i - Y_j) = \operatorname{sgn}(X_i - X_j)(Y_i - Y_j) = 1$.

Пусть $S$ – число согласованных пар, $R$ – число несогласованных пар. $S + R = \dfrac{n(n-1)}{2}$ – число всех (неупорядоченных) пар, $T = S - R = \sum\limits_{i<j} \operatorname{sgn}(X_i - X_j)(Y_i - Y_j)$. Следовательно, $-\dfrac{n(n-1)}{2} \leqslant T \leqslant \dfrac{n(n-1)}{2}$.

**Определение 26.16.** *Коэффициентом корреляции Кендалла* называется величина $\tau = \dfrac{T}{\frac{n(n-1)}{2}}$.

**Утверждение 26.17.** Если $H_0$ верна, то $E\tau = 0$.

**Доказательство.** Аналогично $\rho_S$. □

**Утверждение 26.18.** (б/д) Если $H_0$ верна, то $D\tau = \dfrac{2(2n+5)}{9n(n-1)}$.

**Утверждение 26.19.** $\tau \in [-1,\ 1]$, и границы достигаются.

**Доказательство.** Ограниченность очевидна. Границы достигаются, если все пары согласованы, и, если все пары не согласованы. □

**Утверждение 26.20.** $\tau = 1 - \dfrac{4}{n(n-1)} \sum\limits_{i<j} I_{T_i > T_j}$.

**Доказательство.** Пусть $\begin{pmatrix} R_a \\ S_a \end{pmatrix} = \begin{pmatrix} R(X_a) \\ R(Y_a) \end{pmatrix} = \begin{pmatrix} i \\ T_i \end{pmatrix}$, $\begin{pmatrix} R_b \\ S_b \end{pmatrix} = \begin{pmatrix} j \\ T_j \end{pmatrix}$, $i < j$. Тогда для согласованности пары $\begin{pmatrix} R_a \\ S_a \end{pmatrix}$ и $\begin{pmatrix} R_b \\ S_b \end{pmatrix}$ необходимо и достаточно, чтобы $T_i < T_j$. Следовательно,

$$\tau = \frac{2}{n(n-1)}(S-R) = \frac{2(S+R) - 4R}{n(n-1)} = 1 - \frac{4R}{n(n-1)} = 1 - \frac{4}{n(n-1)} \sum_{i<j} I_{T_i > T_j}.$$

□

**Упражнение 26.21.** $\rho_S = 1 - \dfrac{12}{n^3 - n} \sum\limits_{i<j} (j-i) I_{T_i > T_j}$.

**Замечание 26.22.** Другими словами, $\rho_S$ сильнее реагирует на различия рангов.

**Утверждение 26.23.** (б/д) Если гипотеза $H_0$ верна, то $\rho(\tau,\ \rho_S) \xrightarrow[n\to\infty]{} 1$.

**Утверждение 26.24.** Если $H_0$ верна, то распределение $\tau$ считается известным. Оно не зависит от выбора функций $F_X$ и $F_Y$ и квантили этого распределения известны.

**Доказательство.** Аналогично $\rho_S$. □

**Утверждение 26.25.** Если $H_0$ верна, то $\dfrac{\tau}{\sqrt{D\tau}} \xrightarrow{d} \mathcal{N}(0,\ 1)$.

Критерий для проверки гипотезы $H_0$ имеет вид $\{|\tau| > c\}$, где $c$ – квантиль из известного распределения $\tau$ при верности $H_0$.