

Содержание

1	Виды сходимости случайных векторов и связи между ними	2
2	Закон больших чисел, усиленный закон больших чисел...	3
3	Вероятно-статистическая модель...	5
4	Эмпирическое распределение и эмпирическая функция распределения	6
5	Статистики и оценки	8
6	О наследовании состоятельности	9
7	Метод подстановки и метод моментов	9
8	Квантили и выборочные квантили	11
9	Сравнение оценок, функция потерь и функция риска	13
10	Понятие плотности в дискретном случае	13
11	Экспоненциальные семейства распределений	15
12	Достаточные статистики	16
13	Полные статистики, оптимальные оценки	18

1 Виды сходимости случайных векторов и связи между ними

Пусть $\xi, \{\xi_n\}_{n=1}^{\infty}$ – случайные векторы размерности m .

Определение 1.1. Сходимость почти наверное:

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

Определение 1.2. Сходимость по вероятности:

$$\xi_n \xrightarrow{P} \xi \Leftrightarrow \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(\|\xi_n - \xi\|_2 > \varepsilon) = 0$$

Определение 1.3. Сходимость в L_p (в среднем):

$$\xi_n \xrightarrow{L_p} \xi \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{E}\|\xi_n - \xi\|_p^p = 0$$

Определение 1.4. Сходимость по распределению:

$$\xi_n \xrightarrow{d} \xi \Leftrightarrow \forall f \in \text{BC}(\mathbb{R}^m) : \mathbb{E}f(\xi_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}f(\xi)$$

Утверждение 1.1. Связь между сходимостями:

1. $\text{п.н.} \Rightarrow P$
2. $L_p \Rightarrow P$
3. $P \Rightarrow d$

Утверждение 1.2. $\xi_n \xrightarrow{d} \text{const} \Rightarrow \xi_n \xrightarrow{P} \text{const}$

Утверждение 1.3. Связь между сходимостью векторов и сходимостью их компонент:

1.

$$\xi_n \xrightarrow{\text{п.н.}} \xi \Leftrightarrow \forall i : \xi_n^{(i)} \xrightarrow{\text{п.н.}} \xi^{(i)}$$

2.

$$\xi_n \xrightarrow{P} \xi \Leftrightarrow \forall i : \xi_n^{(i)} \xrightarrow{P} \xi^{(i)}$$

3.

$$\xi_n \xrightarrow{L_p} \xi \Leftrightarrow \forall i : \xi_n^{(i)} \xrightarrow{L_p} \xi^{(i)}$$

4.

$$\xi_n \xrightarrow{d} \xi \Rightarrow \forall i : \xi_n^{(i)} \xrightarrow{d} \xi^{(i)}$$

Доказательство. 1.

$$\bigcap_{i=1}^m \{\xi_n^{(i)} \rightarrow \xi^{(i)}\} = \{\xi_n \rightarrow \xi\} \subset \{\xi_n^{(i)} \rightarrow \xi^{(i)}\}$$

Тогда для \Rightarrow используем включение и свойство меры:

$$1 = P(\{\xi_n \rightarrow \xi\}) \leq P(\{\xi_n^{(i)} \rightarrow \xi^{(i)}\})$$

А для \Leftarrow :

$$1 = P(\bigcap_{i=1}^m \{\xi_n^{(i)} \rightarrow \xi^{(i)}\}) = P(\{\xi_n \rightarrow \xi\})$$

2. Для \Rightarrow :

$$\{|\xi_n^{(i)} - \xi^{(i)}| > \varepsilon\} \subset \{\|\xi_n - \xi\|_2 > \varepsilon\}$$

А для \Leftarrow :

$$\{\|\xi_n - \xi\|_2 > \varepsilon\} \subset \bigcup_{i=1}^m \left\{ |\xi_n^{(i)} - \xi^{(i)}| > \frac{\varepsilon}{\sqrt{m}} \right\}$$

3. Заметим, что

$$\forall i : \lim_{n \rightarrow \infty} \mathbb{E} |\xi_n^{(i)} - \xi^{(i)}|^p = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{E} \|\xi_n - \xi\|_p^p = \lim_{n \rightarrow \infty} \mathbb{E} \sum_{i=1}^n |\xi_n^{(i)} - \xi^{(i)}|^p = 0$$

4. Для \Rightarrow в качестве f возьмём функцию-проектор.

□

Теорема 1.1. *О наследовании сходимостей.*

Пусть $\xi, \{\xi_n\}_{n=1}^\infty$ – случайные векторы в \mathbb{R}^m , причём $\exists B \in \mathcal{B}(\mathbb{R}^m) : P(\xi \in B) = 1$ и $h : \mathbb{R}^m \rightarrow \mathbb{R}^k$ непрерывна в каждой точке множества B . Тогда

$$\xi_n \xrightarrow{n.n., P, d} \xi \Rightarrow h(\xi_n) \xrightarrow{n.n., P, d} h(\xi) \quad (1)$$

Доказательство. • Случай п.н.:

$$P(h(\xi_n) \rightarrow h(\xi)) \geq P(h(\xi_n) \rightarrow h(\xi), \xi \in B) \geq P(\xi_n \rightarrow \xi, \xi \in B) = 1$$

• Случай P :

Пусть $h(\xi_n) \not\xrightarrow{P} h(\xi) \Rightarrow$:

$$\exists \varepsilon_0, \delta_0, \{n_k\}_{k=1}^\infty : P(\|h(\xi_{n_k}) - h(\xi)\| > \varepsilon_0) \geq \delta_0$$

Но из неё мы можем выбрать $\{\xi_{n_{k_m}}\}_{m=1}^\infty$, сходящуюся почти всюду (по прошлому семестру), но тогда мы получили противоречие с предыдущим пунктом доказательства.

• Докажем для непрерывных h :

Тогда

$$\forall f \in BC(\mathbb{R}^k) : f(h(x)) \in BC(\mathbb{R}^m)$$

Значит мы можем взять $f \circ h$ в качестве функции из определения сходимости по распределению и получить требуемое.

□

2 Закон больших чисел, усиленный закон больших чисел...

Теорема 2.1. *ЗБЧ.*

Пусть $\{\xi_n\}_{n=1}^\infty$ – попарно некоррелированные вектора и $\sup_{n,i} \mathbb{V} \xi_n^{(i)} \leq C$. Тогда

$$\frac{s_n - \mathbb{E} s_n}{n} \xrightarrow{P} 0$$

где $\{s_n\}_{n=1}^\infty = \{\sum_{i=1}^n \xi_i\}_{n=1}^\infty$

Теорема 2.2. УЗБЧ.

Пусть $\{\xi_n\}_{n=1}^{\infty}$ – независимые одинаково распределённые, причём $\mathbb{E}\xi_1 < +\infty$. Тогда

$$\frac{s_n}{n} \xrightarrow{n.n.} \mathbb{E}\xi_1$$

Теорема 2.3. ЦПТ.

Пусть $\{\xi_n\}_{n=1}^{\infty}$ – независимые одинаково распределённые, причём \exists ковариационные матрица $\mathbb{V}\xi_1$. Тогда

$$\sqrt{n} \left(\frac{s_n}{n} - \mathbb{E}\xi \right) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}\xi_1)$$

Лемма 2.1. Лемма Служцкого.

Пусть $\xi_n \xrightarrow{d} \xi$ и $\eta_n \xrightarrow{d} c$ ($const$). Тогда

$$\xi_n + \eta_n \xrightarrow{d} \xi + \eta; \quad \xi_n \cdot \eta_n \xrightarrow{d} \xi \cdot c$$

Доказательство. По некому утверждению без доказательства, будет верно

$$\begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi \\ c \end{pmatrix}$$

Тогда, применив теорему о наследовании сходимостей с функциями $+$, \cdot всё получится. \square

Пример. Применение леммы Служцкого.

Пусть $\xi_n \xrightarrow{d} \xi$ – последовательность случайных величин и $H : \mathbb{R} \rightarrow \mathbb{R}$ – дифференцируемая в точке a и $b_n \rightarrow 0$, причём $b_n \neq 0$. Тогда

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

Доказательство. Введём

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

Тогда h непрерывна в 0.

По лемме Служцкого:

$$b_n \xi_n \xrightarrow{d} 0$$

По теореме о наследовании сходимости

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

\square

Теорема 2.4. Обобщение на многомерный случай.

Пусть $\xi_n \xrightarrow{d} \xi$ в \mathbb{R}^m , и $H : \mathbb{R}^m \rightarrow \mathbb{R}^s$, у которой в точке $a \in \mathbb{R}^m \exists$ матрица частных производных $H'(x) = \left(\frac{\partial H_i}{\partial x_j} \right)_{i=1, j=1}^{s, m}$, а также числовая последовательность $b_n \rightarrow 0, b_n \neq 0$. Тогда

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

3 Вероятно-статистическая модель...

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

Определение 3.1. Если $\xi : \Omega \rightarrow E$ такова, что

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

то ξ называется **случайным элементом**.

Если $(E, \mathcal{E}) = (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, то ξ называется **случайным вектором**.

Более того, если $m = 1$, то ξ называется **случайной величиной**.

Определение 3.2. Распределением случайного элемента ξ называется мера P_ξ на \mathcal{E} , такая что $P_\xi(B) = P(\xi \in B)$

Определение 3.3. Выборочное пространство \mathcal{X} – множество всевозможных исходов одного эксперимента (обычно \mathbb{R}^m).

$\mathcal{B}_\mathcal{X}$ – σ -алгебра на \mathcal{X} будем считать Барелевской.

Утверждение 3.1. Построим модель эксперимента, как случайной величины.

Пусть

$$\forall x \in \mathcal{X} : X(x) = x$$

получим отображение $X : \mathcal{X} \rightarrow \mathcal{X}$, которое является случайным элементом на вероятностном пространстве $(\mathcal{X}, \mathcal{B}_\mathcal{X}, P)$ и имеет распределение $P_X = P$

Доказательство. Проверим, что данная случайная величина действительно имеет необходимое нам распределение

$$P_X(B) = P(X \in B) = P(x : X(x) \in B) = P(x \in B) = P(B)$$

□

Утверждение 3.2. Построим модель n независимых повторений нашего эксперимента.

Рассмотрим $\mathcal{X}^n = \mathcal{X} \times \cdots \times \mathcal{X}$ и $\mathcal{B}_\mathcal{X}^n = \mathcal{B}(\mathcal{X}^n) = \sigma(B_1 \times \cdots \times B_n), B_i \in \mathcal{B}_\mathcal{X}$, а $P^n = P \otimes \cdots \otimes P$ – мера на $(\mathcal{X}^n, \mathcal{B}_\mathcal{X}^n)$, такая что $P^n(B_1 \times \cdots \times B_n) = P(B_1) * \cdots * P(B_n)$.

Для этого рассмотрим тождественное отображение $X : \mathcal{X}^n \rightarrow \mathcal{X}^n$. Его i -я компонента X_i (по сути i -й проектор) является случайным вектором с распределением P , причём X_1, \dots, X_n независимы в совокупности.

Доказательство. Фиксируем i , рассмотрим вероятность

$$P^n(X_i \in B_i) = P^n((x_1, \dots, x_n) \in \mathcal{X} : X_i(x_1, \dots, x_n) \in B_i) = P^n((x_1, \dots, x_n) \in \mathcal{X} : x_i \in B_i) = P^n(\mathcal{X} \times \cdots \times B_i \times \cdots \times \mathcal{X}) = 1 * \cdots * P(B_i) * \cdots * 1$$

Теперь докажем независимость:

$$P^n(X_1 \in B_1, \dots, X_n \in B_n) = P^n((x_1, \dots, x_n) \in \mathcal{X} : X_1(x_1, \dots, x_n) \in B_1, \dots) = P^n(B_1 \times \cdots \times B_n) = \prod_{i=1}^n P(B_i) = \prod_{i=1}^n P^n(X_i \in B_i)$$

□

Определение 3.4. Совокупность $X = (X_1, \dots, X_n)$ независимых одинаково распределённых случайных величин (или векторов) с распределением P называется **выборкой** размера n из распределения P .

Также выборку X иногда будем называть **наблюдением**.

Замечание. Для бесконечных выборок определим $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \dots$ и $\mathcal{B}_\mathcal{X}^\infty = \sigma(\{B_1 \times \dots \times B_n \times \mathcal{X} \times \mathcal{X} \times \mathcal{X} \times \dots\}_{n=1}^\infty)$, а меру $P^\infty(B_1 \times \dots \times B_n \times \mathcal{X} \times \dots) = P(B_1) * \dots * P(B_n)$, такая мера существует и единственна.

Аналогично предыдущим пунктам определяем **бесконечную серию экспериментов**.

Определение 3.5. Тройка $(\mathcal{X}, \mathcal{B}_\mathcal{X}, \mathcal{P})$ называется **вероятностно-статистической моделью**.

Замечание. Пусть X_1, \dots, X_n – случайные величины (или векторы), и $X_1(\omega) = x_1, \dots, X_n(\omega) = x_n$ – их значения, называются **реализацией выборки**.

Задачей статистики является сделать вывод о неизвестном распределении по реализации выборки.

Определение 3.6. Вероятностно-статистическая модель $(\mathcal{X}, \mathcal{B}_\mathcal{X}, \mathcal{P})$ называется **параметрической**, если семейство \mathcal{P} параметризовано, то есть

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

обычно $\Theta \subset \mathbb{R}^m$.

4 Эмпирическое распределение и эмпирическая функция распределения

Определение 4.1. Для $\forall B \in \mathcal{B}(\mathbb{R}^m)$ положим

$$P_n^*(B) := \frac{\sum_{i=1}^n \mathbb{I}\{X_i \in B\}}{n}$$

распределение P_n^* называется **эмпирическим распределением**, построенным по выборке X_1, \dots, X_n .

Это случайное распределение (зависит от ω)

Определение 4.2. Функция $F_n^*(x) = \frac{\sum_{i=1}^n \mathbb{I}\{X_i \leq x\}}{n}$ называется **эмпирической функцией распределения**.

Утверждение 4.1. Пусть X_1, \dots, X_n – выборка на вероятностном пространстве (Ω, \mathcal{F}, P) из распределения P_X . Пусть $B \in \mathcal{B}(\mathbb{R}^m)$. Тогда

$$P_n^*(B) \xrightarrow{n.н.} P_X(B), n \rightarrow +\infty$$

Доказательство. Заметим, что $\mathbb{I}\{X_i \in B\}$ – независимые, одинаково распределённые величины.

Тогда мы можем применить УЗБЧ:

$$P_n^*(B) = \frac{S_n}{n} \xrightarrow{п.н.} \mathbb{E}\mathbb{I}\{X_1 \in B\} = P_X(B)$$

□

Теорема 4.1. Гливенко-Кантелли.

Пусть X_1, \dots, X_n – независимые случайные величины с функцией распределения $F(x)$. Тогда

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow{н.н.} 0$$

Доказательство. Почему D_n – случайная величина?

F непрерывна справа, и $\forall \omega : F_n^*$ также непрерывна справа \Rightarrow

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

Значит D_n является случайной совокупностью случайных величин $\Rightarrow D_n$ – случайная величина.

Фиксируем $N \in \mathbb{N}$, тогда $\forall k \in \{1, \dots, N-1\}$ положим

$$X_{N,K} = \inf \left\{ x \in \mathbb{R} \mid F(x) \geq \frac{K}{N} \right\}$$

Заметим, что это число конечно, а также определим $X_{N,0} = -\infty, X_{N,N} = +\infty$.

Если $x \in [X_{N,K}, X_{N,K+1}) \Rightarrow$

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

Последний переход получили благодаря тому, что $F(X_{N,K+1} - 0)$ – отступ чуть влево, от нижней границы значения, где $F(x) \geq \frac{K+1}{N}$, значит там $\leq \frac{K+1}{N}$. Ну а $F(X_{N,K})$ по определению $\geq \frac{K}{N}$.

Аналогично $F_n^*(x) - F(x) \geq F_n^*(X_{N,K}) - F(X_{N,K}) - \frac{1}{N}$. Тогда

$$|F_n^*(x) - F(x)| \leq \max(|F_n^*(X_{N,K+1} - 0) - F(X_{N,K+1} - 0)|, |F_n^*(X_{N,K}) - F(X_{N,K})|) + \frac{1}{N}$$

Но тогда супремум по всей прямой

$$\sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \leq \max_{0 \leq K \leq N-1} \max(|F_n^*(X_{N,K+1} - 0) - F(X_{N,K+1} - 0)|, |F_n^*(X_{N,K}) - F(X_{N,K})|) + \frac{1}{N}$$

Из предыдущего утверждения следует, что $F_n^*(y - 0) = P_n^*((-\infty, y)) \rightarrow P_X((-\infty, y)) = F(y - 0)$.

Теперь для ε фиксируем $\frac{1}{N} < \varepsilon \Rightarrow$

$$\overline{\lim}_N \sup_{x \in \mathbb{R}} |F_N^*(x) - F(x)| \stackrel{н.н.}{<} \varepsilon$$

В силу произвольности ε получаем требуемое. □

5 Статистики и оценки

Определение 5.1. Пусть $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, \mathcal{P})$ – вероятно-статистическая модель, X – наблюдение, (E, \mathcal{E}) – измеримое пространство, и $S : \mathcal{X} \rightarrow E$ – измеримое отображение. Тогда $S(x)$ называется **статистикой**.

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

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

1. Если $g(x)$ – борелевская функция, то

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

называется выборочной характеристикой функции $g(x)$. Например $\overline{X} = \frac{\sum X_i}{n}$ – выборочное среднее. $\overline{X^k} = \frac{1}{n} \sum_{i=1}^n X_i^k$ – выборочный момент k -го порядка.

2. Функции от выборочных квантилей:

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

где h – борелевская.

Например, $s^2 = \overline{X^2} - (\overline{X})^2$ – выборочная дисперсия. $M_k = \frac{1}{n} \sum (X_i - \overline{X})^k$ – выборочный центральный момент k -го порядка.

3. Порядковые статистики:

$$X_{(1)} = \min(X_1, \dots, X_n)$$

$X_{(2)}$ – второй элемент в отсортированной выборке

$$X_{(n)} = \max(X_1, \dots, X_n)$$

вектор $(X_{(1)}, \dots, X_{(n)})$ называется **вариационным рядом**.

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

Определение 5.3. Оценка $\theta^*(X)$ называется **несмещённой** оценкой параметра θ , если

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

где \mathbb{E}_{θ} – матожидание в случае, когда элементы выборки имеют распределение P_{θ} .

Определение 5.4. Оценка $\theta_n^*(X_1, \dots, X_n)$ (а точнее последовательность оценок) называется **состоятельной**, если

$$\forall \theta \in \Theta : \theta^*(X) \xrightarrow{P_{\theta}} \theta$$

и называется **сильно состоятельной** если

$$\forall \theta \in \Theta : \theta^*(X) \xrightarrow{P_{\theta} \text{ п. н.}} \theta$$

Определение 5.5. Оценка $\theta^*(X_1, \dots, X_n)$ называется **асимптотически нормальной** оценкой θ , если

$$\forall \theta \in \Theta : \sqrt{n}(\theta_n^* - \theta) \xrightarrow{d_q} \mathcal{N}(0, \sigma^2(\theta))$$

Утверждение 5.1. Пусть $T(X)$ – асимптотически нормальная оценка для $\tau(\theta)$. Тогда $T(X)$ – состоятельная оценка для $\tau(\theta)$.

Доказательство. Используя лемму Слуцкого, получаем

$$\frac{1}{\sqrt{n}} \cdot \sqrt{n}(T_n - \tau(\theta)) \xrightarrow{d_q} 0$$

Но мы знаем, что из сходимости по распределению к константе следует сходимость по мере. \square

Утверждение 5.2. Из сильной состоятельности и асимптотической нормальности оценки следует её состоятельность.

Доказательство. Следствие из сильной состоятельности автоматически следует из связи сходимостей.

Следствие из асимптотической нормальности было доказано в предыдущем утверждении. \square

6 О наследовании состоятельств

Утверждение 6.1. Наследование состоятельности и сильной состоятельности при взятии непрерывной функции.

Пусть $\theta_n^*(X)$ – сильно состоятельная (состоятельная) оценка θ . Если $\tau : \mathbb{R}^k \rightarrow \mathbb{R}^s$ непрерывна на $\Theta \subset \mathbb{R}^k$, то $\tau(\theta_n^*)$ – сильно состоятельная (состоятельная) оценка $\tau(\theta)$.

Доказательство. Смотри доказательство теоремы о наследовании сходимости. \square

Лемма 6.1. О наследовании асимптотической нормальности.

Пусть $\theta_n^*(X)$ – асимптотически нормальная оценка $\theta \in \Theta$ с асимптотической дисперсией $\sigma^2(\theta)$ и числовая функция $T : \mathbb{R} \rightarrow \mathbb{R}$ дифференцируема в $\forall \theta \in \Theta$. Тогда $T(\theta_n^*)$ – асимптотически нормальная оценка $T(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)(T'(\theta))^2$

Доказательство. Фиксируем $\theta, \xi_n := \sqrt{n}(\theta_n^*(X) - \theta) \xrightarrow{d_q} \xi \sim \mathcal{N}(0, \sigma^2(\theta)), b_n := \frac{1}{\sqrt{n}} \rightarrow 0$.

Вспомним дельта метод, взяв

$$a = \theta, h = T \Rightarrow \frac{T(\theta + \xi_n b_n) - T(\theta)}{b_n} \xrightarrow{d_q} T'(\theta)\xi \Rightarrow \sqrt{n}(T(\theta_n^*) - T(\theta)) \xrightarrow{d_q} T'(\theta)\mathcal{N}(0, \sigma^2(\theta))$$

\square

7 Метод подстановки и метод моментов

Определение 7.1. Пусть в параметрическом семействе $\{P_\theta, \theta \in \Theta\}$ для некоторой функции G выполнено:

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

Тогда оценкой по **методу подстановки** называется $\theta^*(X_1, \dots, X_n) = G(P_n^*)$

Пусть X_1, \dots, X_n – выборка из $P \in \{P_\theta, \theta \in \Theta\}, \Theta \subset \mathbb{R}^k$. Рассмотрим барелевские функции $g_1(x), \dots, g_k(x)$ со значениями в \mathbb{R} .

Пусть $m_1(\theta) = \mathbb{E}_\theta g_1(X_1)$ конечно при $1 \leq i \leq k$.

Определение 7.2. Если $\exists!$ решение системы

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

Тогда оценкой по **методу моментов** называется $\theta^* = m^{-1}(\bar{g})$, где

$$m(\theta) := \begin{pmatrix} m_1(\theta) \\ \vdots \\ m_k(\theta) \end{pmatrix}; \quad \bar{g} = \begin{pmatrix} \frac{\sum_{i=1}^n g_1(X_i)}{n} \\ \vdots \\ \frac{\sum_{i=1}^n g_k(X_i)}{n} \end{pmatrix}$$

Стандартные **пробные функции**: $g_i(X) = X^i$ (i -й момент).

Замечание. О связи методов.

Заметим, что

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

Тогда по методу подстановки получим

$$\theta_n^* = m^{-1} \begin{pmatrix} \int_{\mathcal{X}} g_1(x) dP_n^*(x) \\ \vdots \\ \int_{\mathcal{X}} g_k(x) dP_n^*(x) \end{pmatrix} = G(P_n^*)$$

Таким образом, метод моментов – это частный случай метода подстановки.

Теорема 7.1. *Сильная состоятельность оценки методом моментов.*

Если m биективна и функцию m^{-1} можно доопределить до функции, заданной на всём \mathbb{R}^k и непрерывной в каждой точке множества $m(\Theta)$ тогда оценка по методу моментов является сильно состоятельной оценкой параметра θ .

Доказательство. Фиксируем θ , по УЗБЧ знаем, что

$$\bar{g} \xrightarrow{P_\theta \text{ п.н.}} m(\theta)$$

Используя теорему о наследовании сходимости, навесим m^{-1} :

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

□

Теорема 7.2. *Асимптотическая нормальность ОММ.*

Если в условиях предыдущей теоремы m^{-1} дифференцируема на $m(\Theta)$ и $\forall i \leq k : \mathbb{E}_\theta g_i^2(X_1) < +\infty$. Тогда ОММ θ_n^* является асимптотически нормальной оценкой параметра θ .

Доказательство. По ЦПТ:

$$\sqrt{n}(\bar{g} - m(\theta)) \xrightarrow{d_g} \mathcal{N}(0, \Sigma)$$

Применяем многомерный дельта-метод и получаем требуемое.

□

8 Квантили и выборочные квантили

Определение 8.1. Пусть P – распределение вероятности на \mathbb{R} . Пусть $p \in (0, 1)$. p -квантилью распределения P называют

$$z_p = \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$$

Определение 8.2. Пусть X_1, \dots, X_n – выборка, статистика

$$z_{n,p} = \begin{cases} X_{(\lceil np \rceil)}, & np \notin \mathbb{Z} \\ X_{(np)}, & np \in \mathbb{Z} \end{cases}$$

называется **выборочной p -квантилью**.

Теорема 8.1. О выборочной квантили.

Пусть X_1, \dots, X_n – выборка из распределения P с плотностью $f(x)$. Пусть z_p – это p -квантиль распределения P , причём $f(x)$ непрерывно дифференцируема в окрестности z_p , причём $f(z_p) > 0$. Тогда

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

Доказательство. Пусть $k := \lceil np \rceil$.

Из соображений комбинаторики, заметим, что

$$P(X_{(k)} \leq x) = \sum_{m=k}^n C_n^m F^m(x) (1 - F(x))^{n-m}$$

Засчёт свойств биномиальных коэффициентов, после дифференцирования выражения выше, получим

$$p_{X_{(k)}}(x) = n C_{n-1}^{k-1} F^{k-1}(x) (1 - F(x))^{n-k} f(x)$$

Введём

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

Плотность такого линейного преобразования легко считается

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

где $t_n(x) = z_p + \frac{x}{f(z_p)} \sqrt{\frac{p(1-p)}{n}}$

Откуда это взялось? Вспомним, как меняется плотность при линейном преобразовании:

$$p_{a\xi+b}(x) = F'_{a\xi+b}(x) = P'(a\xi + b \leq x) = P'(\xi \leq \frac{x-b}{a}) = F'_\xi(\frac{x-b}{a}) = \frac{1}{a} p'_\xi(\frac{x-b}{a})$$

Раскроем $p_{X_{(k)}}$ по формуле, которую получили в начале доказательства и разложим полученную плотность η_n в следующее произведение:

$$p_{\eta_n}(x) = A_1(n) A_2(n) A_3(n)$$

где

$$\begin{aligned} A_1(n) &= \sqrt{npq} C_{n-1}^{k-1} p^{k-1} q^{n-k} \\ A_2(n) &= \frac{f(t_n(x))}{f(z_p)} \\ A_3(n) &= \left(\frac{F(t_n(x))}{p} \right)^{k-1} \left(\frac{1 - F(t_n(x))}{q} \right)^{n-k} \end{aligned}$$

Осталось заметить, что

$$A_1(n) \rightarrow \frac{1}{\sqrt{2\pi}}; \quad A_2(n) \rightarrow 1;$$

Для $A_3(n)$ немного сложнее, разложим $F(t_n(x))$ в ряд Тейлора в окрестности z_p . (так как $t_n(x) \rightarrow z_p$):

$$F(t_n(x)) = F(z_p) + (t_n - z_p)F'(z_p) + \frac{1}{2}(t_n - z_p)^2 F''(z_p) + o(t_n - z_p)^2$$

Давайте упростим это выражение, раскрыв t_n и применив свойство квантиля $F(z_p) = p$:

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

Теперь должны расписать приближение $\ln\left(\frac{F(t_n(x))}{p}\right)$, используя формулу $\ln(1+x) = x - \frac{x^2}{2} + o(x^3)$, причём в квадрате нам нужен будет только $x\sqrt{\frac{pq}{n}}$:

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

Аналогично разложив для $\ln\left(\frac{1-F(t_n(x))}{q}\right)$, получим

$$\ln A_3(n) \rightarrow -\frac{x^2}{2}$$

Таким образом, $p_{\eta_n(x)} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ и эта сходимость равномерна на $\forall[-N, N]$.

Используя теорему из теории вероятностей,

$$\eta_n \xrightarrow{d} \mathcal{N}(0, 1)$$

□

Определение 8.3. Медианой распределения P называется $\frac{1}{2}$ квантиль.

Выборочной медианой называется

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

Теорема 8.2. О выборочной медиане.

В условиях теоремы о выборочной квантили:

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

9 Сравнение оценок, функция потерь и функция риска

Определение 9.1. Борелевская неотрицательная функция $g(x, y)$ называется **функцией потерь**.

Если $\theta^*(X)$ – оценка, то $g(\theta^*(X), \theta)$ называется величиной потерь.

Определение 9.2. Если задана функция потерь g , то **функцией риска** оценки θ^* называется $R(\theta^*, \theta) = \mathbb{E}_\theta g(\theta^*, \theta)$

Определение 9.3. Оценка $\theta^*(X)$ лучше оценки $\hat{\theta}(X)$ в **равномерном подходе**, если

$$\forall \theta \in \Theta : R(\theta^*(X), \theta) \leq R(\hat{\theta}(X), \theta)$$

и для некоторого θ неравенство строгое.

Определение 9.4. Оценка $\theta^*(X)$ называется наилучшей в **минимаксном подходе**, если

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

то есть у $\theta^*(X)$ наименьший максимум функции риска.

Определение 9.5. Предположим, что на Θ задано некоторое **априорное** распределение вероятности Q и θ выбирается случайно в соответствии с распределением Q .

Если $\hat{\theta}(X)$ – оценка θ и $R(\hat{\theta}, \theta)$ – её функция риска, тогда

$$R(\hat{\theta}(X)) = \mathbb{E}_\theta R(\hat{\theta}(X), \theta) = \int_{\Theta} R(\hat{\theta}(X), t) Q(dt)$$

Оценка $\theta^*(X)$ называется наилучшей в **байесовском подходе**, если

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

Определение 9.6. Пусть $\hat{\theta}_1, \hat{\theta}_2$ – две асимптотически нормальных оценки параметра θ с дисперсиями $\sigma_1^2(\theta), \sigma_2^2(\theta)$.

Оценка $\hat{\theta}_1$ лучше $\hat{\theta}_2$ в **асимптотическом подходе**, если

$$\forall \theta \in \Theta : \sigma_1^2(\theta) \leq \sigma_2^2(\theta)$$

10 Понятие плотности в дискретном случае

Определение 10.1. **Считающей мерой** μ на \mathbb{Z} называется функция $\mu : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{Z}_+ \cup \{+\infty\}$, определённая по правилу

$$\mu(B) = \sum_{k \in \mathbb{Z}} \mathbb{I}\{k \in B\}$$

Определение 10.2. Интегралом по считающей мере от функции $f(x)$ называется

$$\int_{\mathbb{R}} f(x) \mu(dx) = \sum_{k \in \mathbb{Z}} f(k)$$

если ряд в правой части сходится абсолютно.

Определение 10.3. Пусть ξ – дискретная случайная величина, принимающая значения в \mathbb{Z} . Её плотностью относительно считающей меры μ называется функция

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

Замечание. Всюду далее, когда говорим о плотности, считаем, что либо это обычная плотность в абсолютно непрерывном случае, либо это плотность в дискретном случае по считающей мере на \mathbb{Z}^n .

Определение 10.4. Пусть X – наблюдение из неизвестного распределения $P \in \{P_\theta, \theta \in \Theta\}$, причём $\forall \theta \in \Theta : p_\theta(x)$ имеет плотность $p_\theta(x)$ по одной и той же мере μ .

В этом случае семейство $\{P_\theta, \theta \in \Theta\}$ называется **доминируемым** относительно μ .

Определение 10.5. Случайная величина $u_\theta(x) = \frac{\partial}{\partial \theta} L_{p_\theta}(x)$ называется **вкладом** наблюдения X , и функция $I_X(\theta) = \mathbb{E}_\theta u_\theta^2(X)$ называется **количеством информации** о параметре θ содержащемся в X (информация по Фишеру).

Замечание. Будем считать, что выполнено условие **регулярности**:

1. $\Theta \subset \mathbb{R}$ – открытый интервал
2. Множество $A = \{x \in \mathcal{X} \mid p_\theta(x) > 0\}$ не зависит от θ .
3. Для \forall статистики $S(X)$ с условием $\mathbb{E}_\theta S^2(X) < +\infty$ выполнено $\forall \theta$ выполнено

$$\frac{\partial}{\partial \theta} \int_A S(x) p_\theta(x) \mu(dx) = \int_A S(x) \frac{\partial}{\partial \theta} p_\theta(x) \mu(dx)$$

Левая часть это $\frac{\partial}{\partial \theta} \mathbb{E}_\theta S(X)$, а правая часть

$$\int_A S(x) \frac{\partial}{\partial \theta} p_\theta(x) \frac{1}{p_\theta(x)} p_\theta(x) \mu(dx) = \mathbb{E}_\theta S(X) \frac{\partial}{\partial \theta} \ln p_\theta(X) = \mathbb{E}_\theta S(X) u_\theta(X)$$

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

Теорема 10.1. *Неравенство Рао-Крамера.*

Пусть выполнено условие регулярности и $\hat{\theta}(X)$ – несмещённая оценка $\tau(\theta)$ с условием

$$\forall \theta \in \Theta : \mathbb{E}_\theta (\hat{\theta}(X))^2 < +\infty$$

Тогда

$$\mathbb{V}_\theta \hat{\theta}(X) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}$$

Доказательство. В силу условия 3, при $S(X) = 1$ имеем

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

Также в силу условия 3, при $S(X) = \hat{\theta}(X)$ имеем

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

Умножим первое равенство на $-\tau(\theta)$ и сложим со вторым:

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

Возведём обе части в квадрат и применим КБШ:

$$(\tau'(\theta))^2 \leq \left(\mathbb{E}_\theta(\hat{\theta} - \tau(\theta))^2 \right) \cdot \left(\mathbb{E}_\theta u_\theta^2(X) \right) = \mathbb{V}_\theta \hat{\theta} \cdot I_X(\theta)$$

□

Определение 10.6. Если в неравенстве Рао-Крамера для оценки $\hat{\theta}(X)$ достигается равенство, то $\hat{\theta}(X)$ называется **эффе́ктивной**.

Теорема 10.2. *Критерий эффе́ктивности.*

В условиях регулярности $\hat{\theta}(X)$ эффе́ктивная для $\tau(\theta) \Leftrightarrow \hat{\theta}(X)$ – линейная функция от $u_\theta(X)$ вида $\hat{\theta} - \tau(\theta) = c(\theta)u_\theta(X)$.

Причём последнее равенство может быть выполнено $\Leftrightarrow c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$

Доказательство. Пусть $\hat{\theta}$ – эффе́ктивная для $\tau(\theta) \Rightarrow \tau'(\theta) = \mathbb{E}(\hat{\theta} - \tau(\theta))u_\theta(X)$. А мы знаем, что равенство в КБШ достигается $\Leftrightarrow (\hat{\theta} - \tau(\theta))$ и $u_\theta(X)$ линейно зависимы:

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

Матожидания рассматриваемых величин равны нулю $\Rightarrow \alpha(\theta) \equiv 0$.

Можем поделить обе части на $\gamma(\theta) \neq 0$, это верно ведь иначе

$$\mathbb{V}_\theta \beta(\theta)u_\theta(X) = 0 \Rightarrow \beta = 0 \Rightarrow \perp$$

То есть $\hat{\theta} - \tau(\theta) = r(\theta)u_\theta(X)$.

Обратно, пусть $\hat{\theta} - \tau(\theta) = c(\theta)u_\theta(X) \Rightarrow \hat{\theta} = \tau(\theta) + c(\theta)u_\theta(X) \Rightarrow \hat{\theta}$ – несмещённая оценка $\tau(\theta)$. Умножим обе части на $u_\theta(X)$ и берём матож:

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

□

Замечание. Эффе́ктивная оценка $\tau(\theta)$ – наилучшая оценка $\tau(\theta)$ в классе несмещённых L_2 оценок в равномерном подходе с квадратичной функцией потерь.

11 Экспоненциальные семейства распределений

Определение 11.1. Пусть $\theta = (\theta_1, \dots, \theta_k)$

Экспоненциальным семейством распределений называют все распределения, обобщённая плотность которых имеет вид

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

и где $a_0(\theta) \equiv 1, a_1(\theta), \dots, a_k(\theta)$ линейно независимы на Θ .

Замечание. Проверим, существует ли эффективная оценка, если семейство экспоненциальное:

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

Тогда распишем вклад

$$u_{\theta}(X) = \frac{\partial}{\partial \theta} \ln f_{\theta}(x) = \frac{\partial}{\partial \theta} (a(\theta) \sum_{i=1}^n T(x_i) + nV(\theta)) = a'(\theta) \sum_{i=1}^n T(x_i) + nV'(\theta)$$

Работаем в предположении $T \neq \text{const}$, так как иначе

$$p_{\theta}(x) = h(x) e^{b(\theta)} \Rightarrow \int_{\mathcal{X}} p_{\theta}(x) d\mu = 1 \Rightarrow b(\theta) = \text{const} \Rightarrow p_{\theta}(x) \text{ не зависит от } \theta$$

Пусть также $a'(\theta) \neq 0$, тогда

$$\frac{1}{na'(\theta)} u_{\theta}(x) = \frac{\sum_{i=1}^n T(x_i)}{n} - \frac{-V'(\theta)}{a'(\theta)}$$

По критерию эффективности получаем, что $T^*(X) = \frac{\sum_{i=1}^n T(x_i)}{n}$ является эффективной оценкой для $\tau(\theta) = \frac{-V'(\theta)}{a'(\theta)}$

Обратно, пусть \exists эффективная оценка T для $\tau(\theta)$, пусть $\forall \theta : \tau'(\theta) \neq 0$. Значит достигается равенство в Рау-Крамера:

$$\exists \tau'(\theta) < +\infty : \mathbb{V}_{\theta} \hat{\theta} = \frac{(\tau'(\theta))^2}{I_X(\theta)} < +\infty \Rightarrow \hat{\theta} \in L_2$$

Значит

$$\forall \theta : T(X) - \tau(\theta) = c(\theta) u_{\theta}(x) = \frac{\tau'(\theta)}{I_X(\theta)} u_{\theta}(X)$$

Выразив вклад, получим

$$\frac{\partial}{\partial \theta} \ln f_{\theta}(X) = \frac{T(X) - \tau(\theta)}{c(\theta)}$$

Проинтегрируем, предполагая корректность:

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

Возведём экспоненту в обе части равенства и получим, что правдоподобие имеет нужный нам вид. Но как перейти от произведения плотностей с плотности определённого X_i ? Зафиксируем остальные $X_j, j \neq i$ из носителя A и заметим, что вид остался экспоненциальным.

12 Достаточные статистики

Определение 12.1. Статистика $T(X)$ называется **достаточной** для параметра θ , если

$$P_{\theta}(X \in B \mid T(X) = t)$$

не зависит от θ .

Теорема 12.1. Критерий факторизации Неймана-Фишера.

Пусть $\{P_\theta, \theta \in \Theta\}$ – доминирующее семейство. Статистика T является достаточной для параметра $\theta \Leftrightarrow$ функция правдоподобия $f_\theta(X)$ представима в виде

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

где функции ψ, h неотрицательны, $\psi(t, \theta)$ измерима по t и h измерима по X .

Доказательство. Для дискретного случая.

То есть $f_\theta(x) = P_\theta(X = x)$. Пусть $f_\theta(X) = \psi(T(X), \theta)h(X) \Rightarrow$

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

После сокращения имеем

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

То есть получили что-то, независящее от θ , что подходит под определение достаточной статистики.

Обратно, пусть статистика T достаточная:

$$\begin{aligned} f_\theta(x) &= P_\theta(X = x) = P_\theta(X = x, T(X) = T(x)) = \\ &= P_\theta(T(X) = T(x)) \cdot P_\theta(X = x | T(X) = T(x)) = \psi(T(x), \theta)h(x) \end{aligned}$$

□

Лемма 12.1. Пусть $\eta \in L_1$, тогда $\mathbb{E}(\mathbb{E}(\eta | \xi) - \mathbb{E}\eta)^2 \leq \mathbb{V}\eta$.

Более того, если $\eta \in L_2$, то равенство в неравенстве выше достигается $\Leftrightarrow \eta = \mathbb{E}(\eta | \xi) \Leftrightarrow \eta$ является ξ -измеримой.

Доказательство. Докажем лишь для L_2 .

Пусть $\varphi = \mathbb{E}(\eta | \xi)$. Тогда по неравенству Йенсена

$$\varphi^2 = (\mathbb{E}(\eta | \xi))^2 \leq \mathbb{E}(\eta^2 | \xi)$$

Навесив матожидание, получим $\mathbb{E}\varphi^2 \leq \mathbb{E}\eta^2 < +\infty$. Далее,

$$\mathbb{V}\eta = \mathbb{E}(\eta - \mathbb{E}\eta)^2 = \mathbb{E}(\eta - \varphi + \varphi - \mathbb{E}\eta)^2 = \mathbb{E}(\eta - \varphi)^2 + \mathbb{E}(\varphi - \mathbb{E}\eta)^2 + 2\mathbb{E}(\eta - \varphi)(\varphi - \mathbb{E}\eta)$$

Распишем последнее слагаемое:

$$\mathbb{E}(\mathbb{E}((\eta - \varphi)(\varphi - \mathbb{E}\eta) | \xi)) = \mathbb{E}(\varphi - \mathbb{E}\eta)\mathbb{E}((\eta - \varphi) | \xi) = 0$$

Заметим, что мы всё доказали: оценим первое слагаемое нулём снизу и всё получится. □

Теорема 12.2. Колмогорова-Блэкуэлла-Рао.

Пусть $T(X)$ – достаточная статистика для θ и пусть $d(X)$ – несмещённая для $\tau(\theta)$, положим $\varphi(T) = \mathbb{E}_\theta(d(X) | T)$. Тогда $\varphi(T)$ зависит от выборки только через $T(X)$ (и не зависит от θ), причём

$$\mathbb{E}_\theta\varphi(T) = \tau(\theta); \quad \mathbb{V}_\theta\varphi(T) \leq \mathbb{V}_\theta d(X)$$

Доказательство. Рассмотрим $\varphi(T) := \mathbb{E}_\theta(d(X)|T)$. Распределение X (при фиксированном значении T) не зависит от $\theta \Rightarrow$ распределение $d(X)$ тоже не зависит $\Rightarrow \mathbb{E}_\theta(d(X) | T)$ является измеримой функцией только от T (и, как функция, не зависит от θ) $\Rightarrow \varphi(T)$ действительно статистика.

Очевидно, что $d(X)$ – несмещённая $\Rightarrow \varphi$ тоже (св-во УМО).

$$\mathbb{V}_\theta \varphi(T) = \mathbb{E}_\theta(\varphi - \mathbb{E}_\theta \varphi)^2 = \mathbb{E}_\theta(\mathbb{E}_\theta(d | T) - \mathbb{E}_\theta d)^2 \stackrel{\text{по лемме}}{\leq} \mathbb{V}_\theta d(X)$$

Если $d \in L_2 \Rightarrow$ неравенство переходит в равенство $\Leftrightarrow d = \varphi \Leftrightarrow d(X)$ – борелевская функция от T . \square

13 Полные статистики, оптимальные оценки

Определение 13.1. Наилучшая оценка $T(\theta)$ в классе несмещённых оценок в равномерном подходе с квадратичной функцией потерь называется **оптимальной** оценкой.

Определение 13.2. Статистика $S(X)$ называется **полной** для параметра θ , если из условия

$$\forall \theta \in \Theta : \mathbb{E}_\theta f(S(X)) = 0$$

следует, что

$$f(S(X)) \stackrel{P_\theta \text{ п.н.}}{=} 0$$