

Содержание

1	Виды сходимости случайных векторов и связи между ними	3
2	Закон больших чисел, усиленный закон больших чисел...	4
3	Вероятно-статистическая модель...	6
4	Эмпирическое распределение и эмпирическая функция распределения	7
5	Статистики и оценки	9
6	О наследовании состоятельности	10
7	Метод подстановки и метод моментов	10
8	Квантили и выборочные квантили	12
9	Сравнение оценок, функция потерь и функция риска	14
10	Понятие плотности в дискретном случае	14
11	Экспоненциальные семейства распределений	17
12	Достаточные статистики	18
13	Полные статистики, оптимальные оценки	19
14	Доверительные интервалы	20
15	Метод максимального правдоподобия	22
16	Дополнительные свойства ОМП	24
17	Линейная регрессионная модель	25
18	Гауссовская линейная модель	27
19	Безумные распределения и их свойства	28
20	Гипотезы	30
21	Построение гипотез	31
22	Проверка гипотез в гауссовской линейной модели	32
23	Критерий Пирсона, Колмогорова	34
24	Доказательство теоремы Пирсона	35
25	Байесовские оценки	36

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

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

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, \cdots, X_n независимы в совокупности.

Доказательство. Фиксируем i , рассмотрим вероятность

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

Теперь докажем независимость:

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

□

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

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

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 $\frac{1}{N} < \varepsilon \Rightarrow$

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

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 Слущкого, 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 $m_i(\theta) = \mathbb{E}_\theta g_i(X_1)$ 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 $R(\theta^*, \theta) = \mathbb{E}_\theta g(\theta^*(X), \theta)$

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. Считаящей 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. Пусть X – наблюдение с неизвестным распределением $P \in \{P_\theta, \theta \in \Theta\}$, где семейство доминируемо относительно μ (значит, что либо все дискретные, либо все абсолютно непрерывные).

Функцией правдоподобия называют

$$f_\theta(X) := p_\theta(X)$$

где $p_\theta(X)$ – плотность p_θ по мере μ .

Пример. Пусть $X = (X_1, \dots, X_n)$ – выборка с плотностью $p_\theta(x)$, то $f_\theta(X) = \prod_{i=1}^n p_\theta(X_i)$

Определение 10.6. Определим

$$L_\theta(X) = \ln f_\theta(X)$$

называется **логарифмической функцией правдоподобия**.

Определение 10.7. 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 (в силу несмещённости нашей оценки)

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

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

Теорема 10.2. *Критерий эффективности.*

В условиях регулярности $\hat{\theta}(X)$ эффективная 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 $\hat{\theta}$ – эффективная 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:

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

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

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

$$\mathbb{V}_\theta \beta(\theta)(\hat{\theta} - \tau(\theta)) = 0 \Rightarrow \beta = 0 \Rightarrow \perp$$

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. Эффективная 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 к 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 $f_\theta(X) = \psi(S(X), \theta)h(X) \Rightarrow$

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

После сокращения имеем

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

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

Обратно, пусть статистика T достаточная:

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

□

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

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

Доказательство. Докажем лишь для L_2 .

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

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

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:

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

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

Теорема 12.2. Колмогорова-Блекуэлла-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

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

Теорема 13.1. Лемана-Шеффе.

Пусть T – полная достаточная статистика для $\{P_\theta, \theta \in \Theta\}$, $d(X)$ – несмещённая для $\tau(\theta)$. Тогда $\varphi = \mathbb{E}(d | T)$ – несмещённая оценка с равномерно минимальной дисперсией для $\tau(\theta)$.

Если $\mathbb{V}_\theta(\varphi) < +\infty$, то φ – оптимальная оценка.

Доказательство. Очевидно, что φ несмещённая по той же логике, что и в теореме Колмогорова-Блекуэлла-Рао (свойство УМО).

Пусть \tilde{d} – другая несмещённая оценка. Тогда улучшим её $\tilde{\varphi} = \mathbb{E}(\tilde{d} | T)$ не хуже d по (12) и несмещённая.

Имеем

$$\mathbb{E}_\theta(\varphi(T) - \tilde{\varphi}(T)) = \tau(\theta) - \tau(\theta) = 0$$

то есть для $h = \varphi - \tilde{\varphi}$ имеем $\forall \theta : \mathbb{E}_\theta h(T) = 0$. В силу полноты T получаем, что $\forall \alpha : h(T) \stackrel{P_\theta \text{ п.н.}}{=} 0$. То есть наши оценки на самом деле равны почти наверное.

То есть любая несмещённая оценка, пройдя процедуру улучшения с помощью T почти наверное превращается в φ .

То есть

$$\forall \tilde{d} : \mathbb{V}_\theta(\tilde{d}) \geq \mathbb{V}_\theta(\tilde{\varphi}) = \mathbb{V}_\theta(\varphi)$$

Пусть $\mathbb{V}_\theta(\varphi) < +\infty$, теперь предполагаем, что неравенство на самом деле равенство. (проверяем, что если какая-то другая оценка имеет такую же дисперсию, то это на самом деле $\tilde{\varphi}$)

Это по (12.1) означает, что \tilde{d} уже была T -измеримой, то есть $\tilde{d} = \tilde{\varphi}$ □

Теорема 13.2. *Об экспоненциальном семействе.*

Пусть X_i – выборка из экспоненциального семейства. Если область значений вектор-функции

$$\bar{a}(\theta) = (a_1(\theta), \dots, a_k(\theta)), \theta \in \Theta$$

содержит k -мерный параллелепипед, то

$$T(X) = \left(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

является полной и достаточной для θ .

Замечание. Алгоритм поиска оптимальной оценки:

1. Ищем достаточную статистику T
2. Проверяем на полноту
3. Если полная, то решаем уравнение

$$\forall \theta \in \Theta : \mathbb{E}_\theta g(T(X)) = \tau(\theta)$$

14 Доверительные интервалы

Определение 14.1. Пара статистик $(T_1(X), T_2(X))$ называются **доверительным интервалом** уровня доверия γ для параметра θ , если

$$\forall \theta \in \Theta : P_\theta(T_1(X) \leq \theta \leq T_2(X)) \geq \gamma$$

Если равенство достигается при всех $\theta \in \Theta$, то доверительный интервал называют **точным**.

Определение 14.2. Множество $S(X) \subset \Theta$ называют **доверительным множеством** уровня доверия γ для параметра θ , если

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

Замечание. Метод центральных статистик.

Пусть \exists известная одномерная функция $G(x, \theta)$, такая что её распределение не зависит от параметра θ . Такая функция G называется **центральной статистикой**.

Пусть $\gamma_1, \gamma_2 \in (0, 1)$ таковы, что $\gamma_2 - \gamma_1 = \gamma$ и при $i = 1, 2 : \exists g_i - \gamma_i$ -квантиль $G(X, \theta)$.

Тогда

$$\forall \theta \in \Theta : P_\theta(g_1 \leq G(X, \theta) \leq g_2) \geq \gamma_2 - \gamma_1 = \gamma$$

Введём обозначение

$$S(X) = \{\theta \in \Theta \mid g_1 \leq G(X, \theta) \leq g_2\}$$

для $\forall \theta \in \Theta$ имеем

$$P_\theta(\theta \in S(X)) \geq \gamma$$

то есть $S(X)$ – доверительное множество уровня доверия γ .

Докажем корректность данного метода, заметим, что

$$g_i = F^{-1}(\gamma_i) = \inf\{x : F(x) \geq \gamma_i\} \Rightarrow F(g_i) \geq \gamma_i$$

и при $t < g_i : F(t) < \gamma_i \Rightarrow$ устремляя $t \rightarrow g_i - 0$ получим $F(g_i - 0) \leq \gamma_i$. Тогда

$$P_\theta(g_1 \leq G(X, \theta) \leq g_2) = P_\theta(G(X, \theta) \leq g_2) - P_\theta(G(X, \theta) < g_1) \geq \gamma$$

что и требовалось.

Для поиска центральных статистик в общем случае можно пользоваться следующей леммой:

Лемма 14.1. X_1, \dots, X_n – независимые одинаково распределённые с непрерывной функцией распределения $F(x)$. Тогда

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

Доказательство. Заметим, что

$$P(F(y) \leq x) = P(y \leq F^{-1}(x)) = F(F^{-1}(x)) = x, x \in [0, 1] \Rightarrow F(x) \sim U[0, 1]$$

Несложным упражнением докажите, что $-\ln U[0, 1] \sim \exp(1)$ и тогда из свойства аддитивности экспоненциальных распределений, утверждение леммы станет очевидным. \square

Определение 14.3. Пусть $\{X_n\}_{n=1}^\infty$ – выборка неограниченного размера из неизвестного распределения $P \in \{P_\theta, \theta \in \Theta\}$. Последовательность пар статистик

$$(T_n^{(1)}(X_1, \dots, X_n), T_n^{(2)}(X_1, \dots, X_n))$$

называют **асимптотическим доверительным интервалом** уровня доверия γ для θ , если

$$\forall \theta \in \Theta : \lim_{n \rightarrow +\infty} P_\theta(T_n^{(1)}(X_1, \dots, X_n) \leq \theta \leq T_n^{(2)}(X_1, \dots, X_n)) \geq \gamma$$

Если неравенство выше заменить на равенство (при условии, что нижний предел равен верхнему), то асимптотический доверительный интервал называют **точным**.

Замечание. Построение асимптотических интервалов.

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

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

Тогда при условии $\sigma(\theta)$ – непрерывна, будем иметь, что $\hat{\theta}_n \xrightarrow{P_\theta} \theta$. Значит

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

по лемме Slutsky.

И теперь не сложно догадаться, как будет выглядеть доверительный интервал

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

15 Метод максимального правдоподобия

Определение 15.1. Пусть X – наблюдения с функцией правдоподобия $f_\theta(X)$, тогда оценкой параметра θ по методу максимального правдоподобия (ОМП) называется такая статистика $\hat{\theta}(X)$, что

$$\hat{\theta}(X) = \operatorname{argmax}_{\theta \in \Theta} f_\theta(X)$$

Замечание. Будем использовать новые условия регулярности:

1. $\{P_\theta, \theta \in \Theta\}$ – параметрическое семейство доминируемое относительно меры μ , причём $P_{\theta_1} \neq P_{\theta_2}$ при $\theta_1 \neq \theta_2$ и для $\forall \theta$ определена $P_\theta(X)$ – плотность P_θ относительно меры μ .
2. $A = \{x \in X : P_\theta(x) > 0\}$ не зависит от θ
3. Наблюдение X есть выборка из неизвестного распределения $p \in \{P_\theta, \theta \in \Theta\}$
4. Θ – открытый интервал в \mathbb{R} (возможно бесконечный)
5. Функция $p_\theta(x)$ непрерывно дифференцируемая по θ при всех $x \in A$.
6. $\forall x \in A : p_\theta(x)$ трижды непрерывно дифференцируема по θ .
7. Интеграл $\int_A p_\theta(x) \mu(dx)$ трижды дифференцируем по θ под знаком интеграла.
8. $\mathbb{E}(\frac{\partial}{\partial \theta} \ln p_\theta(X_1))^2 = i(\theta) \in (0, +\infty)$
9. Выполняется

$$\forall \theta_0 \in \Theta \exists c > 0 \exists H(x) \forall \theta \in (\theta_0 - c, \theta_0 + c) \forall x \in A : \left| \frac{\partial^3}{\partial \theta^3} \ln p_\theta(x) \right| < H(x)$$

причём $\mathbb{E}_{\theta_0} H(X_1) < +\infty$

Теорема 15.1. Экстремальное свойство правдоподобия.

Пусть выполнены условия регулярности 1-3. Тогда

$$\forall \theta_0, \theta \in \Theta \theta_0 \neq \theta : P_{\theta_0}(f_{\theta_0}(X_1, \dots, X_n) > f_\theta(X_1, \dots, X_n)) \xrightarrow{n \rightarrow +\infty} 1$$

Доказательство. Пусть $X_i \in A$. Заметим, что

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

Хотим применить ЗБЧ, а для этого нужно

$$\mathbb{E}_{\theta_0} \ln \frac{f_\theta(X)}{f_{\theta_0}(X)} < 0$$

то есть

$$\begin{aligned} \int_A \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) dx &= \int_A \ln \left(1 + \frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 \right) p_{\theta_0}(x) dx \leq \\ \int_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 \right) p_{\theta_0}(x) dx &= \int_A p_\theta(x) dx - \int_A p_{\theta_0}(x) dx = 1 - 1 = 0 \end{aligned}$$

Причём равенство выполняется тогда и только тогда, когда

$$\mu(x \in A : p_\theta(x) = p_{\theta_0}(x)) = 0$$

что противоречит первому условию регулярности. \square

Следствие. Если Θ конечно, то ОМП существует, единственная с вероятностью $\rightarrow 1$, и состоятельная

Доказательство. Максимум $f_\theta(X_1, \dots, X_n)$ с ростом n будет достигаться на истинном значении θ с вероятностью $\rightarrow 1$.

Почему вообще оценка измеримая?

$$\{x : \operatorname{argmax} \dots = \theta_2\} = \{f_{\theta_1}(x) < f_{\theta_2}(x)\} \cap \{f_{\theta_3}(x) < f_{\theta_2}(x)\} \cap \dots$$

то есть конечное пересечение измеримых множеств.

Как проверить, что \exists измеримая версия ОМП, если максимум может достигаться при разных θ ?

Если кандидатов несколько, то выберем с наименьшими номером. Тогда если введём $c_{i < j} = \{x : f_{\theta_i}(x) < f_{\theta_j}(x)\}$, то

$$\{\hat{\theta} = \theta_1\} = \cap_{j \neq 1} c_{j \leq 1}; \quad \{\hat{\theta} = \theta_2\} = c_{2 > 1} \cap \left(\bigcap_{j \geq 3} c_{j \leq 2} \right); \quad \dots$$

\square

Определение 15.2. Уравнением правдоподобия называют

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

Теорема 15.2. Аналог состоятельности ОМП.

Пусть выполняются условия регулярности 1-5 и пусть элементы выборки имеют распределение P_{θ_0} . Тогда \exists отображение $\hat{\theta}_n(X_1, \dots, X_n, \theta_0)$ со значениями в Θ :

$$(P_{\theta_0})^*(\{\hat{\theta}_n \text{ не решение уравнения правдоподобия}\}) \xrightarrow{n \rightarrow +\infty} 0$$

и

$$\forall \varepsilon > 0 : (P_{\theta_0})^*(\{|\hat{\theta}_n - \theta_0| > \varepsilon\}) \xrightarrow{n \rightarrow +\infty} 0$$

Доказательство. Определим $\hat{\theta}_n$, фиксируя X_1, \dots, X_n из множества A .

Если у уравнения $\frac{\partial \ln f}{\partial \theta} = 0$ есть хотя бы 1 корень, то возьмём ближайший корень к θ_0 (в силу непрерывности $\frac{\partial f}{\partial \theta}$ такое возможно, так как предел последовательности корней сам является корнем).

Если же у уравнения нет корней, то доопределим $\hat{\theta}_n := \theta_0$.

Фиксируем $\varepsilon > 0$, что $[\theta_0 - \varepsilon, \theta_0 + \varepsilon] \subset \Theta$. Рассмотрим

$$s_n(\theta_0, \varepsilon) = \{x : f_{\theta_0 - \varepsilon}(x_1, \dots, x_n) < f_{\theta_0}(x_1, \dots, x_n), f_{\theta_0}(\dots) > f_{\theta_0 + \varepsilon}(\dots)\}$$

Но по предыдущей теореме мы можем сказать, что $P_{\theta_0}(s_n) \xrightarrow{n \rightarrow +\infty} 1$.

Далее, $\forall x \in s_n \exists$ точка $\tilde{\theta}_n$, в которой f_θ имеет локальный максимум $\Rightarrow f'(\tilde{\theta}_n) = 0$, причём $\tilde{\theta}_n \in U_\varepsilon(\theta_0)$.

А так как $\hat{\theta}_n$ – ближайший к θ_0 корень, то $|\hat{\theta}_n - \theta_0| < \varepsilon$.

Тогда $\{\hat{\theta}_n - \text{не решение уравнения}\} \subset \{\mathcal{X}^n \setminus s_n\}$, навесив внешнюю меру, то получим первое неравенство из теоремы.

Теперь осталось заметить, что если у нас выполняется неравенство $|\hat{\theta}_n - \theta_0| > \varepsilon$, то мы точно не в s_n , а значит снова сможем оценить сверху мерой $(P_{\theta_0}^*)(\mathcal{X}^n \setminus s_n) \xrightarrow{n \rightarrow +\infty} 0$ \square

Замечание. Почему это почти состоятельность? Не проверяли измеримость $\hat{\theta}_n$ и всё равно $\hat{\theta}_n$ зависит от θ_0 .

Если корней несколько, то неясно, какой ближе к θ_0 .

Не факт, что корень – глобальный максимум.

Корень существует не всегда.

Следствие. Пусть выполняются условия регулярности 1-5 и $\forall n \forall X_1, \dots, X_n : \exists!$ решение $\hat{\theta}_n(X_1, \dots, X_n)$ уравнения правдоподобия и пусть оно является измеримой функцией от выборки.

Тогда $\hat{\theta}_n$ – состоятельная оценка θ и с вероятностью стремящейся к 1, $\hat{\theta}_n$ является ОМП.

Доказательство. Первая часть теоремы следует из предыдущей.

Как и ранее, с большой вероятностью выполняется

$$f_{\theta_0-\varepsilon}(x_1, \dots, x_n) < f_{\theta_0}(x_1, \dots, x_n), f_{\theta_0+\varepsilon}(x_1, \dots, x_n) < f_{\theta_0}(x_1, \dots, x_n)$$

На отрезке $[\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ достигается максимум, это следует из непрерывности плотности, а это следует из предположения регулярности.

Этот максимум достигается на внутренней точке отрезка, которую обозначим θ_n^* , и в ней $\frac{\partial}{\partial \theta} f_\theta = 0$, так как корень единственный, то $\hat{\theta}_n = \theta_n^*$.

То есть $\hat{\theta}_n$ – локальный максимум, но пусть существует $\tilde{\theta}_n$, в которой значение f_θ не меньше, чем в $\hat{\theta}_n$. Но тогда между $\hat{\theta}_n, \tilde{\theta}_n$ будет точка, в которой занулитсся производная (локального минимума), что противоречит с единственностью корня уравнения правдоподобия. \square

16 Дополнительные свойства ОМП

Теорема 16.1. (б/д)

В условиях регулярности 1-9 \forall состоятельной последовательности оценок $\hat{\theta}_n$, являющихся решениями уравнения правдоподобия, выполняется

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

Теорема 16.2. Бахадура. (б/д)

Пусть выполнены условия регулярности 1-9 и $\hat{\theta}_n(X_1, \dots, X_n)$ – асимптотически нормальная оценка θ , причём $\sigma(\theta)$ непрерывна по θ . Тогда

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

Определение 16.1. Если $\sqrt{n}(\hat{\theta}_n(X_1, \dots, X_n) - \theta) \xrightarrow{d_q} \mathcal{N}(0, \frac{1}{i(\theta)})$, то $\hat{\theta}_n$ называется **асимптотически эффективной** оценкой θ .

Утверждение 16.1. Пусть выполняются условия регулярности для неравенство Крамера-Рао, $\hat{\theta}(X)$ – эффективная оценка и равенство из критерия эффективности для $\hat{\theta}(X)$ выполняется $\forall x \forall \theta$. Тогда $\hat{\theta}(X)$ – ОМП.

Доказательство. Распишем это равенство

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

Тогда при $\theta < \hat{\theta}(X) : \frac{\partial}{\partial \theta} \ln f_\theta(X) > 0$, а при $\theta > \hat{\theta}(X) : \frac{\partial}{\partial \theta} \ln f_\theta(X) < 0 \Rightarrow \theta = \hat{\theta}(X)$ – точка максимума $\ln f_\theta(X) \Rightarrow$ ОМП. \square

17 Линейная регрессионная модель

В линейной модели наблюдения – случайный вектор $X \in \mathbb{R}^n$, который представляется в виде $X = l + \varepsilon$, где l неслучайный неизвестный вектор, а ε – случайный вектор (ошибка).

Про ε известно, что $\mathbb{E}\varepsilon = 0$ и $\mathbb{V}\varepsilon = \sigma^2 I_n$, где I_n – единичная матрица $n \times n$, $\sigma^2 > 0$.

Про l известно, что $l \in L$ – линейное подпространство \mathbb{R}^n .

Задача: оценить неизвестные параметры l, σ^2 .

L задано с помощью своего базиса $\{z_1, \dots, z_k\}$ из вектор-столбцов, $\dim L = k$. Составим $Z = (z_1, \dots, z_k)$, то есть $l = Z\theta$, где θ – неизвестные координаты в базисе z_1, \dots, z_k .

То есть задача свелась к оценке θ, σ^2 , где $\theta \in \mathbb{R}^k$.

Определение 17.1. $\hat{\theta}(X) = \operatorname{argmin}_\theta \|X - Z\theta\|^2$ называется **оценкой наименьших квадратов** для θ .

Лемма 17.1. Решением задачи выше является

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

Доказательство. Вначале раскроем скалярку

$$\|X - Z\theta\|^2 = \langle X - Z\theta, X - Z\theta \rangle = X^T X - 2X^T Z\theta + \theta^T Z^T Z\theta$$

Функция минимальна в точке, где частные производные равны нулю.

Дифференцируем по $\theta_i \Rightarrow$

$$-2(X^T Z)_i + 2(\theta^T Z^T Z)_i = 0$$

Это должно выполняться для всех координат, то есть

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

\square

Утверждение 17.1. Данная оценка обладает следующими свойствами:

$$\mathbb{E}\hat{\theta} = \theta; \quad \mathbb{V}\hat{\theta} = \sigma^2 (Z^T Z)^{-1}$$

Доказательство. Распишем матож:

$$\mathbb{E}\hat{\theta} = \mathbb{E}(Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathbb{E}X$$

Но мы знаем, что $X = Z\theta + \varepsilon \Rightarrow \mathbb{E}X = Z\theta \Rightarrow$ всё кроме θ сократится и доказали.
Теперь дисперсия:

$$\mathbb{V}\hat{\theta} = \mathbb{V}(Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathbb{V}X ((Z^T Z)^{-1} Z^T)^T = \sigma^2 I_n (Z^T Z)^{-1}$$

□

Теорема 17.1. (б/д)

Пусть $t = T\theta$ – линейная вектор-функция от θ , $T \in \text{Mat}_{m \times k}$. Тогда оценка $\hat{t} = T\hat{\theta}$ является оптимальной оценкой t в классе линейных несмещённых оценок.

Лемма 17.2.

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

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

$$\mathbb{E}\|X - Z\hat{\theta}\|^2 = \text{tr}\mathbb{V}(X - Z\hat{\theta})$$

Распишем ковариационную матрицу:

$$\begin{aligned} \mathbb{V}(X - Z\hat{\theta}) &= \mathbb{V}[(I_n - Z(Z^T Z)^{-1} Z^T)X] = (I_n - A)\mathbb{V}X(I_n - A)^T = \\ &= (I_n - A)\mathbb{V}X(I_n - A) = \sigma^2(I_n - 2A + A^2) = \sigma^2(I_n - A) \end{aligned}$$

Перейдём обратно к числам:

$$\begin{aligned} \mathbb{E}\|X - Z\hat{\theta}\|^2 &= \sigma^2 \text{tr}(I_n - A) = \sigma^2(\text{tr}I_n - \text{tr}A) = \\ &= \sigma^2(n - \text{tr}I_k) = \sigma^2(n - k) \end{aligned}$$

В предпоследнем переходе воспользовались цикличностью следа:

$$\text{tr}A = \text{tr}(Z(Z^T Z)^{-1} Z^T) = \text{tr}((Z^T Z)(Z^T Z)^{-1}) = \text{tr}(I_k)$$

□

Следствие. • $X - Z\hat{\theta} = \text{proj}_{L^\perp} X$

• $\frac{\|X - Z\hat{\theta}\|^2}{n - k} = \frac{\|\text{proj}_{L^\perp} X\|^2}{n - k}$ – несмещённая оценка σ^2 .

Доказательство. Из линала помним, что

$$X = \text{proj}_L X + \text{proj}_{L^\perp} X$$

Но по определению оценки $\text{proj}_L X = Z\hat{\theta}$.

Второй факт следует из предыдущей леммы.

□

18 Гауссовская линейная модель

Если в линейной регрессионной модели $\varepsilon \sim \mathcal{N}(0, \sigma^2 E) \Rightarrow$ модель называется **гауссовской линейной** моделью.

Замечание. χ -квадрат распределением с k степенями свободы называют

$$\chi_k^2 = \Gamma\left(\frac{1}{2}, \frac{k}{2}\right) \stackrel{d}{=} \xi_1^2 + \dots + \xi_k^2$$

где ξ_i – независимые стандартные нормальные.

Утверждение 18.1. Статистика $S(X) = (\text{proj}_L X, \|\text{proj}_{L^\perp} X\|^2)$ является достаточной для (l, σ^2)

Доказательство. Будем искать достаточную статистику с помощью критерия факторизации, для этого выпишем правдоподобие

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

Заметим, что $\sum_{i=1}^n (X_i - l_i)^2 = \|X - l\|^2$, применим теорему Пифагора:

$$\|X - l\|^2 = \|\text{proj}_L X - \text{proj}_L l\|^2 + \|\text{proj}_{L^\perp} X - \text{proj}_{L^\perp} l\|^2 \stackrel{l \in L}{=} \|\text{proj}_L X - l\|^2 + \|\text{proj}_{L^\perp} X\|^2$$

То есть

$$p(X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}(\|\text{proj}_L X - l\|^2 + \|\text{proj}_{L^\perp} X\|^2)\right)$$

Видно, что мы привели правдоподобие к виду, необходимому для применения критерия Неймана-Фишера \Rightarrow рассматриваемые статистики действительно достаточные. \square

Теорема 18.1. (δ/∂)

$S(X)$ – полная статистика.

Следствие. • $\hat{\theta}$ – оптимальная оценка для θ

- $Z\hat{\theta}$ – оптимальная оценка для l
- $\frac{1}{n-k}\|X - Z\hat{\theta}\|^2$ – оптимальная оценка для σ^2

Доказательство. Несмещённость всех оценок очевидна из предыдущих рассуждений.

Покажем, что все они являются функциями от полных достаточных статистик:

$$Z\hat{\theta} = \text{proj}_L X; \quad \hat{\theta} = (Z^T Z)^{-1} Z^T \text{proj}_L X; \quad \frac{1}{n-k}\|X - Z\hat{\theta}\|^2 = \frac{1}{n-k}\|\text{proj}_{L^\perp} X\|^2$$

\square

Утверждение 18.2. В гауссовской линейной модели $\hat{\theta}$ и $X - Z\hat{\theta}$ независимы, причём

$$\frac{1}{\sigma^2}\|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2; \quad \frac{1}{\sigma^2}\|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2; \quad \hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(Z^T Z)^{-1})$$

Доказательство. Знаем, что

$$\begin{cases} Z\hat{\theta} = \text{proj}_L X \\ X - Z\hat{\theta} = \text{proj}_{L^\perp} X \end{cases}$$

Согласно теореме об ортогональном разложении гауссовские вектора $Z\hat{\theta}$, $X - Z\hat{\theta}$ независимы, причём

$$\frac{1}{\sigma^2} \|Z\hat{\theta} - \mathbb{E}(Z\hat{\theta})\|^2 = \frac{1}{\sigma^2} \|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2$$

И для другого

$$\frac{1}{\sigma^2} \|X - Z\hat{\theta} - \mathbb{E}(X - Z\hat{\theta})\|^2 = \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2$$

Мы знаем, что $Z\hat{\theta}$, $X - Z\hat{\theta}$ независимы, но по выкладкам выше мы знаем, что $\hat{\theta}$ – линейная функция от $Z\hat{\theta}$, поэтому искомая независимость существует.

Осталось заметить, что $\hat{\theta}$ – гауссовская функция, как линейная функция от гауссовского X , причём её матожидание и дисперсия считались выше, поэтому её распределение, очевидно, $\mathcal{N}(\theta, \sigma^2(Z^T Z)^{-1})$ \square

19 Безумные распределения и их свойства

Определение 19.1. Пусть $\xi \sim \mathcal{N}(0, 1)$, $\eta \sim \chi_k^2$, причём ξ и η независимые. Тогда случайная величина

$$\zeta = \frac{\xi}{\sqrt{\frac{\eta}{k}}}$$

имеет **распределение Стьюдента** с k степенями свободы, обозначение $\zeta \sim T_k$.

Лемма 19.1. Свойства распределения Стьюдента:

1. $\zeta \sim T_k \Rightarrow -\zeta \sim T_k$
2. $T_1 \sim \text{Cauchy}, p(x) = \frac{1}{\pi(1+x^2)}$
3. $\zeta_k \sim T_k \Rightarrow \zeta_k \xrightarrow{d} \mathcal{N}(0, 1), k \rightarrow +\infty$

Определение 19.2. Пусть $\xi \sim \chi_k^2$, $\eta \sim \chi_m^2$, причём они независимы. Тогда случайная величина

$$\frac{\frac{\xi}{k}}{\frac{\eta}{m}} \sim F_{k,m}$$

имеет **распределение Фишера** с параметрами k, m .

Лемма 19.2. Свойства распределения Фишера:

1. $\xi \sim T_m \Rightarrow \xi^2 \sim F_{1,m}$
2. $\xi \sim F_{k,m} \Rightarrow \frac{1}{\xi} \sim F_{m,k}$
3. Если k фиксированно и $\xi_m \sim F_{k,m} \Rightarrow k\xi_m \xrightarrow{d} \chi_k^2; m \rightarrow +\infty$
4. $\xi_{k,m} \sim F_{k,m} \Rightarrow \xi_{k,m} \xrightarrow{d} \xi \equiv 1; k, m \rightarrow +\infty$

Теорема 19.1. *Об ортогональном разложении. (б/д)*

Пусть $(X_1, \dots, X_n) \sim \mathcal{N}(l, \sigma^2 I_n)$ и L_1, \dots, L_r — попарно ортогональные подпространства \mathbb{R}^n , причём

$$L_1 \oplus \dots \oplus L_r = \mathbb{R}^n$$

Тогда $Y_i := \text{proj}_{L_i} X, i = \overline{1, r}$ — независимые в совокупности нормальные случайные векторы, причём $\mathbb{E}Y_i = \text{proj}_{L_i} l$ и

$$\frac{1}{\sigma^2} \|Y_i - \mathbb{E}Y_i\|^2 \sim \chi_{\dim L_i}^2$$

Замечание. Доверительные интервалы для параметров гауссовской линейной модели:

1. Доверительный интервал для σ^2 :

$$\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2$$

и $u_{1-\gamma}$ — квантиль χ_{n-k}^2 . Тогда

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

2. Доверительный интервал для θ_i

Мы знаем распределение вектора $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(Z^T Z)^{-1})$, пусть $A := (Z^T Z)^{-1}$. Тогда компонента имеет распределение $\hat{\theta}_i \sim \mathcal{N}(\theta_i, \sigma^2 A_{ii})$.

Сейчас мы знаем, что

$$\begin{cases} \frac{\hat{\theta}_i - \theta_i}{\sqrt{\sigma^2 A_{ii}}} \sim \mathcal{N}(0, 1) \\ \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2 \end{cases}$$

Причём эти случайные величины независимы (доказали выше), значит

$$\sqrt{\frac{n-k}{A_{ii}}} \frac{\hat{\theta}_i - \theta_i}{\|X - Z\hat{\theta}\|} \sim T_{n-k}$$

Теперь мы можем написать доверительный интервал, используя табличку квантилей распределения Стюдента.

3. Доверительный интервал для θ :

Опять знаем распределения следующих независимых случайных величин:

$$\begin{cases} \frac{1}{\sigma^2} \|Z\hat{\theta} - Z\theta\|^2 \sim \chi_k^2 \\ \frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi_{n-k}^2 \end{cases}$$

Тогда

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

Используя эту случайную величину мы можем построить доверительный эллипсоид для θ (так как это многомерная величина).

20 Гипотезы

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

Определение 20.1. Статистическая гипотеза – это предположение вида $P \in \mathcal{P}_0, \mathcal{P}_0 \subset \mathcal{P}$ – подмножество распределений.

Обозначение $H_0 : P \in \mathcal{P}_0$ – гипотеза H_0 .

Текущая рассматриваемая гипотеза называется **основной**.

Задача: по наблюдению X либо принять H_0 (тогда мы сузим класс класс распределений с \mathcal{P} до \mathcal{P}_0), либо отвергнем.

В последнем случае мы переходим к рассмотрению альтернативы (если она есть): $H_1 : P \in \mathcal{P}_1, \mathcal{P}_1 = \mathcal{P} \setminus \mathcal{P}_0$

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:

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

то S называют **критическим множеством**, то есть критерием для проверки гипотезы H_0 .

Определение 20.3. Ошибка первого рода – отвергли H_0 , когда она верна.

Ошибка второго рода – приняли H_0 , когда она неверна.

Будем выбирать S так, что вероятность ошибки 1 рода была меньше заранее выбранного ε , а вероятность ошибки 2 рода сделаем как можно меньше

Определение 20.4. Пусть S – критерий для проверки гипотезы $H_0 : \mathcal{P}_0$. Функция $\beta(Q, S) = Q(X \in S), Q \in \mathcal{P}$ называется **функцией мощности критерия S** .

Определение 20.5. Если для критерия S выполнено неравенство

$$\forall Q \in \mathcal{P}_0 : \beta(Q, S) \leq \varepsilon$$

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

Определение 20.6. Минимальный уровень значимости

$$\alpha(S) = \sup_{Q \in \mathcal{P}_0} \beta(Q, S)$$

называется **мощностью критерия**.

Определение 20.7. Критерий S называется **несмещённым**, если

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

Определение 20.8. Если $X = (X_1, \dots, X_n)$ – выборка растущего размера, то критерий S_n (точнее, последовательность критериев) называется **состоятельной**, если

$$\forall Q \in \mathcal{P}_1 : \beta(Q, S_n) \xrightarrow{n \rightarrow +\infty} 1$$

то есть вероятность ошибки второго рода стремится к нулю.

Определение 20.9. Критерий S уровня значимости ε называется **более мощным**, чем критерий R того же уровня значимости, если

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

то есть вероятность ошибки 2го рода у S равномерно меньше.

Определение 20.10. Критерий S называется **равномерно наиболее мощным** критерием уровня значимости ε , если $\alpha(S) \leq \varepsilon$, и S мощнее \forall другого критерия R , который удовлетворяет условию $\alpha(R) \leq \varepsilon$

Определение 20.11. Гипотеза $H : P = P_0$, где P_0 – известное распределение, называется **простой**.

21 Построение гипотез

Лемма 21.1. *Неймана-Пирсона.*

Пусть для простой гипотезы мы выбрали критерий $S_\lambda := \{X : p_1(X) - \lambda p_0(X) \geq 0\}$ и пусть критерий R удовлетворяет условию $P_0(X \in R) \leq P_0(X \in S_\lambda)$. Тогда

$$\begin{aligned} P_1(X \in R) &\leq P_1(X \in S_\lambda) \\ P_0(X \in S_\lambda) &\leq P_1(X \in S_\lambda) \end{aligned}$$

Доказательство. Заметим, что

$$\begin{aligned} \mathbb{I}_R(p_1(X) - \lambda p_0(X)) &\leq \mathbb{I}_R(X)(p_1(X) - \lambda p_0(X)) \mathbb{I}_{\{x: p_1(X) - \lambda p_0(X) \geq 0\}}(X) = \\ &= \mathbb{I}_R(X) \mathbb{I}_{S_\lambda}(X)(p_1(X) - \lambda p_0(X)) \leq (p_1(X) - \lambda p_0(X)) \mathbb{I}_{S_\lambda}(X) \end{aligned}$$

Следовательно

$$\begin{aligned} P_1(X \in R) - \lambda P_0(X \in R) &= \mathbb{E}_1 \mathbb{I}_R(X) - \lambda \mathbb{E}_0 \mathbb{I}_R(X) = \\ &= \int_{\mathcal{X}} \mathbb{I}_R(X)(p_1 - \lambda p_0)(X) d\mu \leq \int_{\mathcal{X}} \mathbb{I}_{S_\lambda}(X)(p_1(X) - \lambda p_0(X)) d\mu = \\ &= P_1(X \in S_\lambda) - \lambda P_0(X \in S_\lambda) \end{aligned}$$

Значит

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

Для доказательства второго факта предположим $\lambda \geq 1$:

$$P_0(X \in S_\lambda) = \int_{\mathcal{X}} \mathbb{I}_{S_\lambda} p_0(X) d\mu \leq \int_{\mathcal{X}} \mathbb{I}_{S_\lambda} p_1(X) d\mu = P_1(X \in S_\lambda)$$

Пусть теперь $\lambda < 1$, тогда $\forall x \in \overline{S_\lambda} : p_1(x) \leq p_0(x)$:

$$P_1(X \in \overline{S_\lambda}) = \int_{\mathcal{X}} \mathbb{I}_{\overline{S_\lambda}} p_1(X) d\mu \leq \int_{\mathcal{X}} \mathbb{I}_{\overline{S_\lambda}} p_0(X) d\mu = P_0(X \in \overline{S_\lambda}) \Rightarrow P_0(X \in S_\lambda) \leq P_1(X \in S_\lambda)$$

□

Следствие. Если $\lambda > 0$ удовлетворяет условию $P_0(X \in S_\lambda) = \varepsilon \Rightarrow S_\lambda$ – равномерно наиболее мощный критерий размера ε .

Замечание. Для нахождения λ из этого критерия необходимо решить уравнение

$$\int_{\{x: p_1(X) - \lambda p_0(X) \geq 0\}} p_0(x) d\mu = \varepsilon$$

В случае абсолютно непрерывных распределений, решение, как правило, есть.

А в дискретном случае уравнение неразрешимо для многих $\varepsilon > 0$. Тогда можно взять ε_0 , для которого оно разрешимо.

Определение 21.1. Пусть семейство \mathcal{P} параметризовано параметром $\theta \in \mathbb{R}$. Пусть \mathcal{P} доминируемо относительно меры μ , то есть \exists функция правдоподобия $f_\theta(X)$.

Семейство $\{p_\theta, \theta \in \Theta\}$ называется **семейством с монотонным отношением правдоподобия** по статистике $T(X)$ если $\forall \theta_0 < \theta_1: \frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}$ является монотонной функцией от $T(X)$, причём тип монотонности один и тот же для всех $\theta_1 > \theta_0$

Теорема 21.1. О монотонном отношении правдоподобия. (б/д)

Пусть даны гипотезы

$$H_0 : \theta \leq \theta_0; \quad H_1 : \theta > \theta_0$$

а семейство $\{p_\theta, \theta \in \Theta\}$ – семейство с монотонным отношением правдоподобия, причём $\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}$ не убывает по $T(X)$ при $\theta_1 > \theta_2$.

Тогда критерий $S_\varepsilon = \{T(X) \geq C_\varepsilon\}$ с условием $P_{\theta_0}(S_\varepsilon) = \varepsilon$ является равномерным наиболее мощным критерием уровня значимости ε для проверки H_0 против H_1 .

Замечание. Двойственность доверительного оценивания и проверки гипотез.

1. 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 ε для проверки H_0 . Действительно:

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

2. Пусть, наоборот, нам дан критерий S_{θ_0} уровня значимости $1 - \varepsilon$ для проверки $H_0 : \theta = \theta_0$. И пусть S_{θ_0} известен для всех $\theta_0 \in \Theta$. Рассмотрим $S(X) = \{\theta \in \Theta : X \notin S_{\theta_0}\}$. Проверим, что это доверительное множество уровня доверия ε . Для

$$\forall \theta \in \Theta : P_\theta(\theta \in S(X)) = P_\theta(X \notin S_\theta) = 1 - P_\theta(X \in S_\theta) \geq \varepsilon$$

22 Проверка гипотез в гауссовской линейной модели

Цель: построить критерий для проверки линейной гипотезы $H_0 : T\theta = t$, где $T \in \mathbb{R}^{m \times k}$, $t \in \mathbb{R}^m$, $\text{rk} T = m \leq k$.

Знаем: $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(Z^T Z)^{-1})$ и $\hat{t} = T\hat{\theta} \sim \mathcal{N}(T\theta, T\sigma^2(Z^T Z)^{-1}T) =: \mathcal{N}(T\theta, \sigma^2 B)$.

Заметим, что матрица B положительно определена и симметрична, тогда

$$\exists \sqrt{B} : \sqrt{B}\sqrt{B} = B, (\sqrt{B})^T = \sqrt{B}$$

Тогда

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

А это значит

$$\chi_m^2 \sim \left\| \frac{1}{\sigma}(\sqrt{B})^{-1}(\hat{t} - T\theta) \right\|^2 = \left(\frac{1}{\sigma}(\sqrt{B})^{-1}(\hat{t} - T\theta) \right)^T \left(\frac{1}{\sigma}(\sqrt{B})^{-1}(\hat{t} - T\theta) \right) = \frac{1}{\sigma^2}(\hat{t} - T\theta)^T B^{-1}(\hat{t} - T\theta)$$

Обозначим последнее выражение за \hat{Q}_T , рассмотрим статистику

$$\hat{Q}_T = (\hat{t} - t)^T B^{-1}(\hat{t} - t)$$

так как \hat{Q}_T выражается через $\hat{\theta} \Rightarrow$ не зависит от $X - Z\hat{\theta}$.

Значит в условиях H_0 :

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

Теперь можем сформулировать **F-критерий**:

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

где $u_{1-\gamma}$ – квантиль $F_{m, n-k}$ распределения.

Пример. Пример с двумя гауссовскими выборками, отличающимися сдвигом: проверка гипотезы об их однородности.

Пусть $X_1, \dots, X_n \sim \mathcal{N}(a_1, \sigma^2)$; $Y_1, \dots, Y_m \sim \mathcal{N}(a_2, \sigma^2)$. Построим F-критерий для $H_0 : a = b$, тогда

$$W = \begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_1 \\ a_2 \\ \vdots \\ a_2 \end{pmatrix} + \vec{\varepsilon} = Z\theta + \vec{\varepsilon}$$

где

$$Z = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}, \theta = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Получили, что

$$\hat{\theta} = (Z^T Z)^{-1} Z^T W = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix}; T = \begin{pmatrix} 1 & -1 \end{pmatrix}; \hat{t} = \bar{X} - \bar{Y}; t = 0$$

Ну и

$$T(Z^T Z)^{-1} T^T = \frac{1}{n} + \frac{1}{m}; \hat{Q}_T = \left(\frac{1}{n} + \frac{1}{m} \right)^{-1} (\bar{X} - \bar{Y})^2$$

Дальше всё это подставляем в формулу и всё получается.....

23 Критерий Пирсона, Колмогорова

Замечание. Критерий согласия Пирсона (критерий хи-квадрат).

Пусть X_1, \dots, X_n – i.i.d., причём $P(X_1 = a_i) = p_i, i = \overline{1, m}; \sum_{i=1}^m p_i = 1$.

Положим $\nu_j = \sum_{i=1}^n \mathbb{I}\{X_i = a_j\}$ – количество осуществления исхода a_j . Очевидно, $\sum_{j=1}^m \nu_j = n$. Нам неизвестен

$$\vec{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix}$$

Хотим проверить гипотезу $H_0 : \vec{p} = \vec{p}_0$, где \vec{p}_0 принадлежит вероятностному симплексу.

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

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

Теорема 23.1. Пирсона. Доказательство в отдельном билете.

Если выполнена H_0 , то

$$\hat{\chi}_n^2 \xrightarrow{d} \chi_{m-1}^2, n \rightarrow +\infty$$

Пусть $u_{1-\varepsilon}$ – $(1 - \varepsilon)$ -квантиль распределения χ_{m-1}^2 . Если $\hat{\chi}_n^2 > u_{1-\varepsilon} \Rightarrow H_0$ отвергается.

Лемма 23.1. Критерия Пирсона состоятелен против альтернативы $\vec{p} \neq \vec{p}_0$

Доказательство. Пусть на самом деле $\vec{p} \neq \vec{p}_0$:

$$\hat{\chi}_n^2 = n \sum_{i=1}^m \left(\frac{\nu_i}{n} - p_i^0 \right)^2 \frac{1}{p_i^0}$$

По УЗБЧ имеем, что $\forall i : \frac{\nu_i}{n} = \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{X_j = a_i\} \xrightarrow{\text{п.н.}} P(X_j = a_i) = p_i \Rightarrow$ по теореме о наследовании сходимости:

$$\sum_{j=1}^m \left(\frac{\nu_j}{n} - p_j^0 \right) \frac{1}{p_j^0} \xrightarrow{\text{п.н.}} \sum_{j=1}^m (p_j - p_j^0)^2 \frac{1}{p_j^0} > 0$$

Значит $\hat{\chi}_n^2 \rightarrow +\infty$ почти наверное при $n \rightarrow +\infty : P(\hat{\chi}_n^2 > u_{1-\varepsilon}) \rightarrow 1, n \rightarrow +\infty$. То есть ошибка второго рода стремится к нулю. \square

Замечание. Критерий согласия Колмогорова.

Пусть $X = (X_1, \dots, X_n)$ – выборка из неизвестного распределения на \mathbb{R} с непрерывной функцией распределения F .

Вспомним, что

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

Где D_n – случайная величина.

Теорема 23.2. (б/д)

1. Распределение D_n не зависит от вида F .

2. $\sqrt{n}D_n \xrightarrow{d} \xi$, где ξ имеет распределение Колмогорова, то есть

$$P(\xi \leq z) = \sum_{j \in \mathbb{Z}} (-1)^j e^{-2j^2 z^2} \mathbb{I}\{z > 0\}$$

Определение 23.2. Рассмотрим $H_0 : F = F_0$ – непрерывная, критерием Колмогорова называют множество

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

где $k_{1-\alpha}$ – $(1 - \alpha)$ -квантиль распределения Колмогорова, применим при $n \geq 20$.

Замечание. Критерий Смирнова-фон Мизеса.

Рассмотрим статистику

$$\omega^2 = n \int_{\mathbb{R}} (F_n^*(x) - F_0(x))^2 dF_0(x)$$

при $H_0 : F = F_0$ – непрерывной, распределение ω^2 не зависит от вида F_0 и $\omega^2 \xrightarrow{d}$ известному распределению.

Определение 23.3. Критерий Пирсона, Колмогорова и ω^2 называют **критериями согласия**, так как они проверяют гипотезу вида $H_0 : P = P_0$.

24 Доказательство теоремы Пирсона

Теорема 24.1. Пирсона.

Если выполнена H_0 , то

$$\hat{\chi}_n^2 \xrightarrow{d} \chi_{m-1}^2, n \rightarrow +\infty$$

Доказательство. Введём

$$Y_j = \begin{pmatrix} \mathbb{I}\{X_j = a_1\} \\ \vdots \\ \mathbb{I}\{X_j = a_m\} \end{pmatrix}, 1 \leq j \leq n$$

Получается, $\{Y_j\}_{j=1}^n$ – независимые одинаково распределённые случайные векторы. Очевидно, что

$$\mathbb{E}Y_j = \begin{pmatrix} p_1 \\ \vdots \\ p_m \end{pmatrix} = \begin{pmatrix} p_1^0 \\ \vdots \\ p_m^0 \end{pmatrix}$$

Теперь посчитаем ковариацию:

$$\text{cov}(\mathbb{I}\{X_j = a_i\}, \mathbb{I}\{X_j = a_k\}) = \mathbb{E}\mathbb{I}\{X_j = a_i, X_j = a_k\} - p_i^0 p_k^0 = \begin{cases} p_i^0 - (p_i^0)^2, i = k \\ -p_i^0 p_j^0, i \neq k \end{cases}$$

Положим

$$B := \begin{pmatrix} p_1^0 & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & p_m^0 \end{pmatrix}$$

Тогда $\mathbb{V}Y_j = B - p_0 p_0^T$. По многомерной ЦПТ

$$\frac{Y_1 + \dots + Y_n}{\sqrt{n}} - \sqrt{n}p_0 \xrightarrow{d} \mathcal{N}(0, B - p_0 p_0^T)$$

Заметим, что сумма Y_i это

$$Y_1 + \dots + Y_n = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_m \end{pmatrix} = \vec{\nu}$$

Пусть $\xi_n := (\sqrt{B})^{-1} \sqrt{n} \left(\frac{\vec{\nu}}{n} - \vec{p}_0 \right)$ тогда по теореме о наследовании сходимости

$$\xi_n \xrightarrow{d} \mathcal{N}\left(0, (\sqrt{B})^{-1}(B - p_0 p_0^T)(\sqrt{B})^{-1}\right) = \mathcal{N}(0, I_m - ZZ^T); \quad Z = \begin{pmatrix} \sqrt{p_1^0} \\ \vdots \\ \sqrt{p_m^0} \end{pmatrix}$$

Рассмотрим ортогональную матрицу $V \in \mathbb{R}^{m \times m}$, такую что её первая строка – это $(\sqrt{p_1^0}, \dots, \sqrt{p_m^0})$
Тогда по теореме о наследовании сходимости:

$$V\xi_n \xrightarrow{d} V\mathcal{N}(0, I_m - ZZ^T) = \mathcal{N}(0, V(I_m - ZZ^T)V^T)$$

Заметим, что

$$VZ = \begin{pmatrix} \sqrt{p_1^0} & \dots & \sqrt{p_m^0} \\ \vdots & \ddots & \vdots \\ \vdots & \dots & \ddots \end{pmatrix} \begin{pmatrix} \sqrt{p_1^0} \\ \vdots \\ \sqrt{p_m^0} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}$$

Тогда

$$V(I_m - ZZ^T)V^T = I_m - (VZ)(VZ)^T = I_m - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \vdots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix} := \tilde{I}_m$$

По теореме о наследовании сходимости

$$\|V\xi_n\|^2 \xrightarrow{d} \|\mathcal{N}(0, \tilde{I}_m)\|^2 \stackrel{d}{=} \chi_{m-1}^2$$

Но V ортогональна, значит $\|V\xi_n\|^2 = \|\xi_n\|^2 = \|(\sqrt{B})^{-1} \sqrt{n} \left(\frac{\vec{\nu}}{n} - \vec{p}_0 \right)\|^2 = \hat{\chi}_n^2$ □

25 Байесовские оценки

Пусть $\Theta \subset \mathbb{R}^d$. В байесовском подходе параметр θ является случайной величиной (или вектором) с известным распределением Q на множестве Θ .

А именно, пусть Q – распределение на (Θ, β_Θ) с плотностью $q(t)$. Рассмотрим вероятностное пространство $(\Theta \times \mathcal{X}, \beta_\Theta \otimes \beta_\mathcal{X}, \tilde{p})$, где \tilde{p} – мера с плотностью $f(t, x) = q(t)p_t(x)$.

Тогда (θ, X) – случайный вектор на этом вероятностном пространстве с плотностью $f(t, x)$ при таком подходе $p_t(x)$ – условная плотность X при условии, что $\theta = t$ фиксированная.

Рассмотрим $\theta(t) = t$ – случайная величина на $(\Theta, \beta_\Theta, Q)$ и пусть X – случайная величина на $(\mathcal{X}, \beta_\mathcal{X}, p_t)$. Тогда (θ, X) – случайный вектор на $(\Theta \times \mathcal{X}, \beta_\Theta \otimes \beta_\mathcal{X}, \tilde{p})$ с плотностью f , причём $q(t)$ – плотность θ , а $p_t(x)$ – условная плотность X при $\theta = t$

Определение 25.1. Плотность $q(t)$ называется **априорной** плотностью параметра θ , а условной плотностью θ относительно x

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

называется **апостериорной** плотностью θ .

Определение 25.2. Оценка $\hat{\theta}(X) = \int_{\Theta} tq(t | x)dt = \mathbb{E}_{\hat{p}}(\theta | X)$ называется **байесовской** оценкой параметра θ .

Теорема 25.1. Байесовская оценка является наилучшей оценкой в байесовском подходе с квадратичной функцией потерь.

Доказательство. Пусть $\hat{\theta}(X)$ – байесовская оценка θ . Хотим минимизировать

$$\int_{\Theta} R(\hat{\theta}, \theta)q(\theta)d\theta = \int_{\Theta} \mathbb{E}_{\theta}(\hat{\theta} - \theta)^2 q(\theta)d\theta = \int_{\Theta} \int_{\mathcal{X}} (\hat{\theta} - \theta)^2 f(\theta, x)dx d\theta = \mathbb{E}_{\hat{p}}(\hat{\theta}(X) - \theta)^2$$

А по теореме о наилучшем квадратичном прогнозе данное матожидание будет минимальным при $\mathbb{E}_{\hat{p}}(\theta | X)$ □

26 Коэффициенты корреляции

Теперь будем рассматривать выборки $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_n)$ – две выборки одинакового распределения. Нас интересует гипотеза о независимости \vec{X}, \vec{Y} :

$$H_0 : F_{X,Y}(t, s) = F_X(t)F_Y(s)$$

Пусть $\mathbb{E}X_1^2 < +\infty$ и $\mathbb{E}Y_1^2 < +\infty$.

Определение 26.1. Величина

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

называется **коэффициентом корреляции Пирсона**.

Утверждение 26.1. Выполняется сходимость

$$\hat{\rho} \xrightarrow{n.n.} \rho = \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{V}X\mathbb{V}Y}}$$

Доказательство. Вспомним свойство выборочной дисперсии:

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \overline{X^2} - (\bar{X})^2$$

По УЗБЧ:

$$\begin{cases} \overline{X^2} \xrightarrow{n.n.} \mathbb{E}X_1^2 \\ \bar{X} \xrightarrow{n.n.} \mathbb{E}X_1 \end{cases} \Rightarrow S_X^2 \rightarrow \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \mathbb{V}X$$

То есть сходимость знаменателя доказали. Теперь числитель:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{\sum_{i=1}^n X_i Y_i}{n} - \bar{X} \frac{\sum_{i=1}^n Y_i}{n} - \bar{Y} \frac{\sum_{i=1}^n X_i}{n} + \bar{X} \cdot \bar{Y} \xrightarrow{n.n.} \mathbb{E}X_1 Y_1 - \mathbb{E}X_1 \mathbb{E}Y_1$$

□

Теорема 26.1. Б/Д.

Пусть $n > 2$. X и Y – две независимые выборки, имеющие нормальное распределение. Тогда

$$T := \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \sim T_{n-2}$$

Тогда критерием будет $S = \{|T| > const\}$

Замечание. Коэффициент корреляции Спирмана.

Пусть X_1, \dots, X_n – выборка из непрерывного распределения, упорядочим элементы выборки по возрастанию.

Определение 26.2. Номера, которые получили элементы выборки при таком упорядочивании, называются **рангами**

$R(X_i)$ – номер X_i в вариационном ряду

Замечание. Если $(r_1, \dots, r_n) \in s_n$ – перестановка, то

$$P(R(X_1) = r_1, \dots, R(X_n) = r_n) = \frac{1}{n!}$$

Обозначим $R_i = R(X_i)$.

Аналогично Y_1, \dots, Y_n – выборка из непрерывного распределения и S_1, \dots, S_n – соответствующие ранги.

Заметим, что

$$\bar{R} = \bar{S} = \frac{n+1}{2}$$

Определение 26.3. Величину

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

называют **коэффициентом корреляции Спирмана**.

Замечание. Заметим, что

$$\sum_{i=1}^n (S_i - \bar{S})^2 = \sum_{i=1}^n (R_i - \bar{R})^2 = \sum_{i=1}^n \left(R_i - \frac{n+1}{2}\right)^2 = \sum_{i=1}^n \left(i - \frac{n+1}{2}\right)^2 = \frac{n^3 - n}{12}$$

Определим T_i : переставим пары (R_i, S_i) в порядке возрастания первой координаты \Rightarrow получим

$$(1, T_1), \dots, (n, T_n)$$

причём $R_i = k \Leftrightarrow T_k = S_i$.

Тогда

$$\begin{aligned} \rho_S &= \frac{12}{n^3 - n} \sum_{i=1}^n \left(R_i - \frac{n+1}{2}\right) \left(S_i - \frac{n+1}{2}\right) = \frac{12}{n^3 - n} \sum_{i=1}^n \left(i - \frac{n+1}{2}\right) \left(T_i - \frac{n+1}{2}\right) = \\ &= \frac{12}{n^3 - n} \left(\sum_{i=1}^n iT_i - \frac{n+1}{2} \sum_{i=1}^n T_i - \frac{n+1}{2} \sum_{i=1}^n i + n \left(\frac{n+1}{2}\right)^2 \right) = \\ &= 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (R_i - S_i)^2 \end{aligned}$$

Знаем, что (T_1, \dots, T_n) – перестановка $(1, \dots, n)$, и при H_0 все $n!$ перестановок равновероятны:

$$\forall k : \mathbb{E}T_k = \sum_{j=1}^n jP(T_k = j) = \frac{n+1}{2}$$

Тогда очевидно, что

$$\mathbb{E}\rho_S = 0$$

Лемма 26.1. *Свойства ρ_S :*

1. $H_0 \Rightarrow \begin{cases} \mathbb{E}\rho_S = 0 \\ \mathbb{V}\rho_S = \frac{1}{n-1} \end{cases}$
2. По КБШ: $\rho_S \in [-1, 1]$, причём крайние значения достигаются (при $R_i = S_i$ и обратном порядке).
3. При H_0 распределение ρ_S известно, не зависит от F_X и F_Y и его квантили есть в таблицах.
4. $H_0 \Rightarrow \frac{\rho_S}{\sqrt{\mathbb{V}\rho_S}} \xrightarrow{d} \mathcal{N}(0, 1), n \rightarrow +\infty$, причём при $n \geq 50$ это приближение используется.

Получается, критерием будет $\{|\rho_S| > const\}$ – больше какой-то квантили.

Замечание. Коэффициент корреляции Кендала.

Опять есть выборки X_1, \dots, X_n и Y_1, \dots, Y_n .

Определение 26.4. Пары (X_i, Y_i) и $(X_j, Y_j), 1 \leq i < j \leq n$ **согласованы**, если

$$\text{sgn}(X_i - X_j)\text{sgn}(Y_i - Y_j) = 1$$

Пусть S – число согласованных пар, R – число несогласованных. Тогда

$$S + R = \frac{n(n-1)}{2}$$

Определим

$$T := S - R = \sum_{i < j} \text{sgn}(X_i - X_j)(Y_i - Y_j)$$

Очев $T \in [-\frac{n(n-1)}{2}, \frac{n(n-1)}{2}]$

Определение 26.5. Величину

$$\tau = \frac{T}{\frac{n(n-1)}{2}}$$

называют **коэффициентом корреляции Кендала**

Заметим,

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

Лемма 26.2. *Свойства коэффициента Кендала:*

$$1. H_0 \Rightarrow \begin{cases} \mathbb{E}T = 0 \\ \mathbb{V}T = \frac{2(2n+5)}{9n(n-1)} \end{cases}$$

2. $\tau \in [-1, 1]$, крайние значения достигаются при всех согласованных и всех несогласованных.

3.

$$\tau = 1 - \frac{4}{n^2 - n} \sum_{i < j} \mathbb{I}\{T_i > T_j\}; \quad \rho_S = 1 - \frac{12}{n^3 - n} \sum_{i < j} (j - i) \mathbb{I}\{T_i > T_j\}$$

то есть ρ_S сильнее реагирует на различия рангов в несогласованных парах.

$$4. H_0 \Rightarrow \rho(\tau, \rho_S) \xrightarrow{n \rightarrow +\infty} 1$$

Кроме того, распределение τ известно, не зависит от F_X, F_Y и его квантили известны.