Математическая статистика. Прикладной поток.

Лектор: Никита Волков

Конспект набирали: Никита Павличенко, Артем Ямалутдинов

18 января 2020 г.

Оглавление

1 Глава 2. Точечные оценки параметров

Лекция 2 (от 9.09)

2.1. Статистики и оценки

Пусть $(\mathscr{X}, \mathcal{B}_{\mathscr{X}}, \mathcal{P})$ — вероятностно-статистическая модель, $\mathcal{P} = \{P_\theta \,|\theta \in \Theta\}$ — параметрическое семейство распределений.

Задача: оценить θ.
Пусть $X = (X_1, \ldots, X_n)$ — выборка из неизвестного распределения $P \in \mathcal{P}$.
Определение 1. Пусть $(E, \mathcal{E})$ — измеримое пространство. Тогда измеримая функция $S : \mathscr{X}^n \to E$ называется *статистикой*. Если $E = \Theta$, то $S(X)$ называется *оценкой* θ.

Примеры статистик:
Пусть $X = (X_1, \ldots, X_n)$ — действительная выборка, т. е. $\mathscr{X} = \mathbb{R}$.

1. Выборочные характеристики:
   - $\overline{g(X)} = \frac{1}{n}\sum\limits_{i=1}^{n} g(X_i)$ — *выборочная характеристика* функции g (g борелевская).
   - $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — *выборочное среднее.*
   - $\overline{X^k} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ — *выборочный k-ый момент.*
2. Функции от выборочных характеристик (т.е $h(\overline{g_1(X)}, \ldots, \overline{g_k(X)})$; h, g_i — борелевские):
   - $g_1(x) = x^2, g_2(x) = x, h(x, y) = x - y^2$
     $h(\overline{g_1(X)}, \overline{g_2(X)}) = \overline{X}^2 - \overline{X}^2 = S^2$ — *выборочная дисперсия.*

   Утверждение 1. $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$.
3. Порядковые статистики:
   Упорядочим выборку по возрастанию: $(X_{(1)}, \ldots, X_{(n)})$ — *вариационный ряд.*
   $X_{(k)}$ — *k-я порядковая статистика.*

Пример 1. $(X_1, X_2, X_3) = (2, 5, 1)$.
$\overline{X} = 8/3$
$\overline{X^2} = 10$
$S^2 = 10 - 64/9 = 26/9$.
Вариационный ряд: $(X_{(1)}, X_{(2)}, X_{(3)}) = (1, 2, 5)$.

2.2. Свойства оценок

Замечание 1. для распределения P_θ будем обозначать: E_θ — матожидание, D_θ — дисперсия, P_θ-п.н., d_θ.

Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения
$P \in \{P_\theta \mid \theta \in \Theta\}, \Theta \in \mathbb{R}^d$.
Определение 2. оценка $\hat{\theta}$ называется *несмещенной оценкой* $\tau(\theta)$, если $E_\theta \hat{\theta}(X) = \tau(\theta) \ \ \forall \theta \in \Theta$.
Пример 2.

- $\hat{\theta}_1 = X_1, \ \hat{\theta}_2 = \overline{X}$ — несмещенные оценки для $\tau(\theta) = E_\theta X_1$.
- $\mathcal{P} = \{Bern(\theta) \mid \theta \in (0,1)\} : \overline{X}, X_1$ — несмещенные оценки θ.
- $\mathcal{P} = \{Exp(\theta) \mid \theta > 0\} : \overline{X}, X_1$ — несмещенные оценки $\frac{1}{\theta}$.

Асимптотические свойства

Пусть $X = (X_1, \dots)$ — выборка неограниченного размера из $P \in \{P_\theta \mid \theta \in \Theta\}, \Theta \in \mathbb{R}^d$.
Определение 3.

1. Оценка $\hat{\theta_n}(X_1, \dots, X_n)$ называется *состоятельной оценкой* θ, если
$$\hat{\theta_n}(X_1, \dots, X_n) \xrightarrow{P_\theta} \theta \quad \forall \theta \in \Theta.$$
2. Оценка $\hat{\theta_n}(X_1, \dots, X_n)$ называется *сильно состоятельной оценкой* θ, если
$$\hat{\theta_n}(X_1, \dots, X_n) \xrightarrow{P_\theta - \text{п.н.}} \theta \quad \forall \theta \in \Theta.$$
3. Оценка $\hat{\theta_n}(X_1, \dots, X_n)$ называется *асимптотически нормальной оценкой* θ, если
$$\sqrt{n}(\hat{\theta_n}(X_1, \dots, X_n) - X) \xrightarrow{d_\theta} \mathcal{N}(0, \Sigma(\theta)) \quad \forall \theta \in \Theta,$$
где $\Sigma(\theta)$ — *асимптотическая матрица ковариаций.* Если $d = 1$, то $\Sigma(\theta) = \sigma^2(\theta)$ — *асимптотическая дисперсия.*

Смысл 1.

1. *Состоятельность:* при больших n вероятность большого отклонения оценки $\hat{\theta_n}$ от θ мала, но нет численной характеристики степени отклонения.
2. *асимптотическая нормальность:* дает численную характеристику степени отклонения

   Пусть $\hat{\theta_n}$ — а.н.о. θ с а.д. $\sigma^2(\theta)$. Тогда при больших n $\quad \hat{\theta_n} \sim_{\text{прибл.}} \mathcal{N}\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$.
3. *Сильная состоятельность* важна тогда, когда данные поступают последовательно.

Пример 3. Пусть $X_1, \dots, X_n$ – выборка из распределения Лапласа со сдвигом θ.

$p_\theta(x) = \dfrac{1}{2} e^{-|x-\theta|}. \quad E_\theta X_1 = \theta, D_\theta X_1 = 2.$

УЗБЧ: $\overline{X} \xrightarrow{P_\theta - \text{п.н.}} \theta \implies \overline{X}$ — (сильно) состоятельная оценка θ.

ЦПТ: $\sqrt{n}(\overline{X} - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, 2) \implies \overline{X} \sim_{\text{прибл.}} \mathcal{N}(0, \frac{2}{n})$. По свойствам нормального распределения, с вероятностью > 0.99:

$$\theta - 3\sqrt{\frac{2}{n}} < \overline{X} < \theta + 3\sqrt{\frac{2}{n}} \overline{X} - 3\sqrt{\frac{2}{n}} < \theta < \overline{X} + 3\sqrt{\frac{2}{n}}$$

(доверительный интервал).

Пусть $n = 200, \overline{X} = 1$. Тогда неравенство имеет вид

$$0.7 < \theta < 1.3$$

(реализация доверительного интервала).
Утверждение 2.

$$\begin{array}{ccc} \text{\textit{Сильная состоятельность}} & & \\ & \searrow & \\ & & \text{\textit{Состоятельность}} \\ & \nearrow & \\ \text{\textit{Асимпт. нормальность}} & & \end{array}$$

Других следствий нет.
Утверждение 3. *Пусть $X_1, \ldots, X_n$ — выборка, т. ч. $E_\theta |X_1|^{2k} < +\infty$. Тогда $\overline{X^k}$ — несмещенная сильно состоятельная асимптотически нормальная оценка $E_\theta X^k$.*

2.3 Наследование свойств

Цель: получить оценку для $\tau(\theta)$, обладающие некоторым свойством, если имеется оценка для $\psi(\theta)$ с тем же свойством.
Теорема 1 (о наследовании сходимостей)**.** *Пусть $\{\xi_n, n \in \mathbb{N}\}, \xi$ — случайные векторы размерности d. Тогда:*

1. *Если $\xi_n \xrightarrow{P} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^k$, т. ч. h непрерывна на $B : P(\xi \in B) = 1$. Тогда $h(\xi_n) \xrightarrow{P} h(\xi)$.*
2. *Аналогично для сходимости п. н.*
3. *Если $\xi_n \xrightarrow{d} \xi$ и $h : \mathbb{R}^d \to \mathbb{R}^k$ непрерывна, то $h(\xi_n) \xrightarrow{d} h(\xi)$.*

Пример 4. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — н.о.р.с.в., т.ч. $\mathbb{E}\xi_1 = a \neq 0$, $\mathbb{D}\xi_n$ ограничена.

Из ЗБЧ: $\dfrac{S_n}{n} \xrightarrow{P} a, \quad S_n = \sum \xi_i$. Рассмотрим $h(x) = 1/x$ и применим теорему:

$$h\left(\frac{S_n}{n}\right) = \frac{n}{S_n} \xrightarrow{P} h(a) = \frac{1}{a}.$$

Утверждение 4. *Пусть $\hat{\theta}$ — (сильно) состоятельная оценка θ. Пусть τ непрерывна на Θ. Тогда $\tau(\hat{\theta})$ — (сильно) состоятельная оценка $\tau(\theta)$.*
Замечание 2. Условие непрерывности на Θ нельзя ослабить.
Теорема 2 (лемма Слуцкого)**.** *Пусть $\{\xi_n, n \in \mathbb{N}\}$, $\{\eta_n, n \in \mathbb{N}\}$, ξ — случайные величины, $C \in \mathbb{R}$. Пусть $\xi_n \xrightarrow{d} \xi$, $\eta_n \xrightarrow{d} C$. Тогда $\xi_n + \eta_n \xrightarrow{d} \xi + C$,*
$\xi_n \cdot \eta_n \xrightarrow{d} \xi C$.
Теорема 3 (о производной)**.** *Пусть $\{\xi_n, n \in \mathbb{N}\}$, ξ — случайные векторы размерности d, т.ч. $\xi_n \xrightarrow{d} \xi, h : \mathbb{R}^d \to \mathbb{R}^k$ непрерывно дифференцируема в точке $a \in \mathbb{R}^d$,*
$\{b_n\} : b_n > 0, b_n \to 0$ — числовая последовательность. Тогда

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} \frac{\partial h}{\partial x}\Big|_a \cdot \xi,$$

где $\dfrac{\partial h}{\partial x}\Big|_a$ — матрица Якоби функции h в точке a.

Доказательство: $(d = 1)$:

Определим функцию

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

Функция H непрерывна в нуле. Тогда по лемме Слуцкого $\xi_n b_n \xrightarrow{d} \xi \cdot 0 = 0 \implies$
$\implies \xi_n b_n \xrightarrow{p} 0$. Применим теорему о наследовании сходимостей:

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

Применим еще раз лемму Слуцкого:

$$\xi_n H(\xi_n b_n) \xrightarrow{d} h'(a)\xi.$$

Следовательно, $\dfrac{h(\xi_n b_n + a) - h(a)}{b_n} \xrightarrow{d} h'(a)$. $\square$

Пример 5. Пусть $\{\xi_n, n \in \mathbb{N}\}$ — н.о.р.с.в, т.ч. $\mathbb{E}\xi_1 = a \neq 0, \ , \mathbb{D}\xi_1 = \sigma^2$.

$\sqrt{n}\left(\dfrac{n}{S_n} - \dfrac{1}{a}\right) \xrightarrow{d} ?$

$\triangle$ ЦПТ: $\sqrt{n}(\frac{S_n}{n} - a) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Воспользуемся теоремой о производной с $\xi_n = \sqrt{n}(\frac{S_n}{n} - a)$,
$\xi \sim \mathcal{N}(0, \sigma^2), \ h(x) = \frac{1}{x}, \ b_n = \frac{1}{\sqrt{n}}$:

$$\frac{h(\xi_n b_n + a) - h(a)}{b_n} = \sqrt{n}\left[h\left(a + \left(\frac{S_n}{n} - a\right)\right) - h(a)\right] =$$

$$= \sqrt{n}\left(\frac{n}{S_n} - \frac{1}{a}\right) \xrightarrow{d} \xrightarrow{d} \xi \cdot \left(\frac{1}{x}\right)\Bigg|_a = -\xi \cdot \frac{1}{a^2} \sim \mathcal{N}\left(0, \frac{\sigma^2}{a^4}\right) \qquad \square.$$

Замечание 3. Если мы рассмотрим ξ_n как выборку $(X_1, X_2, \dots)$, то $1/\overline{X}$ — а. н. о. для $1/a$ с асимптотической дисперсией σ^2/a^4.

Лекция 3

Теорема 4 (дельта-метод). *Пусть $\hat{\theta}_n$ — асимптотически нормальная оценка $\theta \in \Theta \subseteq \mathbb{R}^d$ с асимптотической матрицей ковариаций $\Sigma(\theta)$ и $\tau : \mathbb{R}^d \to \mathbb{R}^k$ — непрерывно дифференцируемая функция. Тогда $\tau(\hat{\theta}_n)$ — асимптотически нормальная оценка $\tau(\theta)$ с асимптотической матрицей ковариаций $D(\theta)\Sigma(\theta)D^T(\theta)$, где $D(\theta) = \dfrac{\partial \tau(\theta)}{\partial \theta}$.*

Доказательство: Применим теорему о производной:

$$a = \theta, \ h(x) = \tau(x), \ \xi_n = \sqrt{n}(\hat{\theta}_n - \theta), \ \xi \sim \mathcal{N}(0, \Sigma(\theta)), \ b_n = \frac{1}{\sqrt{n}}$$

$$\frac{h(a+\xi_n b_n)-h(a)}{b_n}=\frac{\tau\left(\theta+\frac{1}{\sqrt{n}}\sqrt{n}(\hat\theta-\theta)\right)-\tau(\theta)}{1/\sqrt{n}}=$$

$$=\sqrt{n}(\tau(\hat\theta)-\tau(\theta))\xrightarrow{d}\underbrace{\left.\frac{\partial h}{\partial x}\right|_\theta}_{D(\theta)}\xi\sim\mathcal{N}(0,D(\theta)\Sigma(\theta)D^T(\theta)).$$

$\square$

Пример 6. $X_1,\dots X_n\sim Exp(\theta)$, $\theta>0$. ЦПТ:

$$\sqrt{n}\left(\overline{X}-\frac{1}{\theta}\right)\xrightarrow{d_\theta}\mathcal{N}(0,1/\theta^2)\Rightarrow\overline{X}\text{— а.н.о. }\frac{1}{\theta}\text{ с асимптотической дисперсией }1/\theta^2.$$

Примерим дельта-метод с функцией $\tau(x)=1/x$: $\tau(\overline{X})=\dfrac{1}{\overline{X}}$ — а.н.о. $\tau\left(\dfrac{1}{\theta}\right)$. с асимптотической дисперсией $\dfrac{1}{\theta^2}\cdot\left(\left.\dfrac{\partial\theta}{\partial x}\right|_{1/\theta}\right)^2=\dfrac{1}{\theta^2}\left(-\dfrac{1}{x^2}\right)^2=\theta^2$.

Доказательство теоремы о наследовании сходимостей:

2. $\xi_n\xrightarrow{\text{п.н.}}\xi$, $h:\mathbb{R}^d\to\mathbb{R}^k$ непрерывна на множестве $B: P(\xi\in B)=1$. $\xi_n\xrightarrow{\text{п.н.}}\xi\Leftrightarrow P(\lim\limits_{n\to\infty}\xi_n=\xi)=1$. Хотим доказать, что $P(\lim\limits_{n\to\infty}h(\xi_n)=h(\xi))=1$. $P(\lim\limits_{n\to\infty}h(\xi_n)=h(\xi))=1\geqslant P(\lim\limits_{n\to\infty}\xi_n=\xi,\xi\in B)=1$, так как вероятность этого события равна 1.

1. $\xi_n\xrightarrow{P}\xi$ и $h:\mathbb{R}^d\to\mathbb{R}^k$ непрерывна на B таком, что $P(\xi\in B)=1$.

$$h(\xi_n)\xrightarrow{P}h(\xi)\Leftrightarrow\forall\varepsilon>0\quad\underbrace{P(\|h(\xi_n)-h(\xi)\|)>\varepsilon}_{\forall\delta>0\exists N:\forall n>N P(\|h(\xi_n)-h(\xi)\|)>\varepsilon<\delta}\to 0.$$

$$h(\xi_n)\overset{P}{\nrightarrow}h(\xi)\Rightarrow\exists\varepsilon,\delta,\{\xi_n\}_{k=1}^\infty:P(\|h(\xi_n)-h(\xi)\|>\varepsilon)>\delta.$$

Заметим, что $\xi_n\to\xi\Rightarrow$ существует последовательность $\{\xi_{n_{k_s}}\}_{s=1}^\infty$ такая, что $\xi_{n_{k_s}}\xrightarrow{\text{п.н.}}\xi$, $s\to\infty$.

3. $\xi_n\xrightarrow{d}\xi$ и h непрерывна. Возьмем $f:\mathbb{R}^k\to\mathbb{R}$ — непрерывная ограниченная. Тогда $f(h(x))$ непрерывная ограниченная на $\mathbb{R}^d$, и, поскольку $\xi_n\xrightarrow{d}\xi$, то $\mathbb{E}(h(\xi_n))\to\mathbb{E}f(h(\xi))\Rightarrow h(\xi_n)\xrightarrow{d}h(\xi)$ по определению.

$\square$

Доказательство леммы Слуцкого для суммы: $\xi_n\xrightarrow{d}\xi$, $\eta_n\xrightarrow{d}c\Rightarrow\xi_n+\eta_n\xrightarrow{d}\xi+c$.

$\xi_n\xrightarrow{d}\xi\Leftrightarrow F_{\xi_n}(x)\to F_\xi(x)$ в точках непрерывности F_ξ. $F_{\xi+c}(x)=F_\xi(x-c)$. $\xi_n\to\xi\Rightarrow\xi_n+c\to\xi+c$, так как есть сходимость в точках непрерывности $F_{\xi+c}(x)$. Пусть t — точка непрерывности $F_{\xi+c}$, $\varepsilon>0: t\pm\varepsilon$ тоже точка непрерывности.

$$F_{\xi_n+\eta_n}(t)=P(\xi_n+\eta_n\leqslant t)=P(\xi_n+\eta_n\leqslant t,\ \eta_n<c-\varepsilon)+P(\xi_n+\eta_n\leqslant t,\ \eta_n\geqslant c-\varepsilon)\boxed{\leqslant}$$

1. $$\{\xi_n+\eta_n\leqslant t,\ \eta_n<c-\varepsilon\}\subseteq\{\eta_n<c-\varepsilon\}\subseteq\{|\eta_n-c|>\varepsilon\}.$$

2.

$$\{\xi_n + \eta_n \leqslant t,\ \eta_n \geqslant c - \varepsilon\} \subseteq \{\xi_n + c - \varepsilon \leqslant t,\ \eta_n \geqslant c - \varepsilon\} \subseteq \{\xi_n + c \leqslant t + \varepsilon\}.$$

$$\boxed{\leqslant} P(|\eta_n - c| > \varepsilon) + P(\xi_n + c \leqslant t + \varepsilon).$$

$$\lim_{n\to\infty} \sup F_{\xi_n+\eta_n}(t) \leqslant \underbrace{\lim_{n\to\infty} P(|\eta_n - c| > \varepsilon)}_{=0 \text{ т.к. } \eta_n \xrightarrow{d} c \Rightarrow \eta_n \xrightarrow{P} c} + \underbrace{\lim_{n\to\infty} F_{\xi_n+c}(t+\varepsilon)}_{=F_{\xi+c}(t+\varepsilon), \text{ т.к. } \xi_n + c \xrightarrow{d} \xi + c \text{ и } t+c \text{ — т.непр.}}$$

То есть $\lim\limits_{n\to\infty} \sup F_{\xi_n+\eta_n}(t) \leqslant F_{\xi+c}(t+\varepsilon)$. Аналогично $\lim\limits_{n\to\infty} \inf F_{\xi_n+\eta_n}(t) \geqslant F_{\xi+c}(t-\varepsilon)$, следовательно $F_{\xi+c}(t-\varepsilon) \Rightarrow F_{\xi+c}(t-\varepsilon) \leqslant \lim\limits_{n\to\infty} \inf F_{\xi_n+\eta_n}(t) \leqslant \lim\limits_{n\to\infty} \sup F_{\xi_n+\eta_n}(t) \leqslant F_{\xi+c}(t+\varepsilon)$. В силу произвольности $\varepsilon > 0$ и непрерывности $F_{\xi+c}(t)$, получаем, что существует $\lim\limits_{n\to\infty} F_{\xi_n+\eta_n}(t) = F_{\xi+c}(t) \Rightarrow \xi_n + \eta_n \xrightarrow{d} \xi + c$.

□

2.4. Методы нахождения оценок

(1) Метод моментов

Идея: приравняем друг к другу теоретические и выборочные моменты.

Пусть $X = (X_1, \ldots, X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta | \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^d$. Составим систему:

$$\begin{cases} E_\theta X_1 = \overline{X} \\ E_\theta X_1^2 = \overline{X^2} \\ \ldots \\ E_\theta X_1^d = \overline{X^d} \end{cases}$$

Решение этой системы называется оценкой θ по методу моментов.

Обобщенный метод моментов

Пусть $g_1(x), \ldots, g_d(x)$ — борелевские функции, такие, что $|E_\theta g_j(x_j)| < +\infty$. Составим систему:

$$\begin{cases} E_\theta g_1(X_1) = \overline{g_1(X)} \\ E_\theta g_2(X_1) = \overline{g_2(X)} \\ \ldots \\ E_\theta g_d(X_1) = \overline{g_d(X)} \end{cases}$$

Пример 7. $X_1, \ldots, X_n \sim Exp(\theta)$. Найти оценку стандартным методом моментов и обобщенным с функцией $g(x) = I\{x > 1\}$.

Решение:

1. Стандартный метод моментов дает уравнение $E_\theta X_1 = \overline{X} \Rightarrow \hat{\theta}_1 = \dfrac{1}{\overline{X}}$. Ранее мы получали, что $\hat{\theta}_1$ — асимптотически нормальная оценка θ с асимптотической дисперсией $\theta^2 = \sigma^2(\theta)$.

2. Получаем систему из одного уравнения: $E_\theta I(X_1 > 1) = \overline{I(X > 1)}$. $E_\theta I\{X_1 > 1\} = \int_1^{+\infty} \theta e^{-\theta x} dx = e^{-\theta}$, $\overline{I\{X > 1\}} = \frac{1}{n}\sum_{i=1}^{n} I\{X_i > 1\}$. Получаем $\hat{\theta}_2 = \ln \overline{I\{X > 1\}}$. ЦПТ: $\overline{I\{X > 1\}}$ — асимптотически нормальная оценка $e^{-\theta}$ с асимптотической дисперсией $D_\theta I\{X_1 > 1\} = e^{-\theta} - e^{-2\theta}$. Применим дельта-метод с функцией $\tau(x) = -\ln x$. Отсюда $\hat{\theta}_2$ — асимптотически нормальная оценка θ с асимптотической дисперсией $(e^{-\theta} - e^{-2\theta}) \cdot ((-\ln x)')^2 \Big|_{e^{-\theta}} = (e^{-\theta} - e^{-2\theta}) \cdot \frac{1}{x^2}\Big|_{e^{-\theta}} = e^{\theta} - 1 = \sigma_2^2(\theta)$.

Вывод: нужен метод сравнения оценок. Видимо, $\hat{\theta}_1$ лучше $\hat{\theta}_2$, так как $\sigma_1^2(\theta) < \sigma_2^2(\theta)$.

Распишем оценку по методу моментов: пусть $g_1(x), \dots, g_d(x)$ — борелевские функции такие, что $|E_\theta g_i(x_i)| < +\infty$.

$$m(\theta) = \begin{pmatrix} E_\theta g_1(X_1) \\ \dots \\ E_\theta g_d(x_1) \end{pmatrix} = \begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_d(x)} \end{pmatrix} = \overline{g(X)} \Rightarrow \hat{\theta} = m^{-1}(g(\overline{X})).$$

Утверждение 5.

1. *Если m^{-1} непрерывна, то $\hat{\theta}$ — сильно состоятельная оценка θ.*
2. *Если m^{-1} непрерывно дифференцируема и $E_\theta g_i^2(X_i) < +\infty$, то $\hat{\theta}$ — асимптотически нормальная оценка θ.*

Доказательство:

1. В силу выбора $g_i : |E_\theta g_i(X_i)| < +\infty$ по УЗБЧ: $\overline{g(X)} \xrightarrow{P_\theta-\text{п.н.}} m(\theta) = E_\theta g(X_1)$. Поскольку m^{-1} непрерывна, то по теореме о наследовании сходимостей $\hat{\theta} = m^{-1}(\overline{g(X)})$ — сильно состоятельная оценка $m^{-1}(m(\theta)) = \theta$.
2. ЦПТ: $\sqrt{n}(\overline{g(X)} - m(\theta)) \xrightarrow{d_\theta} \mathcal{N}(0, \Sigma(\theta)) \Rightarrow \overline{g(X)}$ — асимптотически нормальная оценка $m(\theta)$. Применяем дельта-метод с функцией m^{-1}: $\hat{\theta}$ — асимптотически нормальная оценка θ. □

□

(2) Метод максимального правдоподобия

Пусть $X = (X_1, \dots X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta | \theta \in \Theta\}$, где

1. Либо все P_θ абсолютно непрерывные и $p_\theta(x)$ — плотность P_θ.
2. Либо все P_θ дискретные и $p_\theta(x) = P_\theta(X_1 = x)$ — дискретная плотность.

Определение 4. $L_X(\theta) = p_\theta(X) = \prod_{i=1}^{n} p_\theta(X_i)$ — функция правдоподобия (как функция от θ).

Определение 5. $l_X(\theta) = L_X(\theta)$ — логарифмическая функция правдоподобия.
Замечание 4. При фиксированном θ функция правдоподобия равна плотности выборки, в которую в качестве аргумента подставлена сама выборка.

Смысл 2. "вероятность"выборки в зависимости от значения параметра. Степень доверия к конкретному значению параметра. Интересует только относительное значение.
Пример 8. пусть x_1 — наблюдение.

Рисунок

Видимо θ_2 более правдоподобно, чем θ_1 и θ_3.
Определение 6. $\hat{\theta} = \arg\max_{\theta\in\Theta} L_X(\theta)$ называется оценкой максимального правдоподобия.
Утверждение 6. *ОМП не зависит от параметризации. Пусть $\hat{\theta}$ — ОМП для θ. $\tau : \Theta \to \Psi$ — биекция. Тогда $\tau(\hat{\theta})$ — ОМП для $\tau(\theta)$.*

Утверждение 7. *Пусть $\forall n,\ \forall x_1, \dots, x_n$ уравнение правдоподобия $\sum\limits_{i=1}^{n} \frac{\partial}{\partial\theta} \ln p_\theta(x_i) = 0$ имеет только одно решение. Тогда*

1. *$[L1 - L5] \Rightarrow$ ОМП состоятельна;*
2. *$[L1 - L9] \Rightarrow$ ОМП является асимптотически нормальной оценкой θ с асимптотической матрицей ковариаций $i(\theta)^{-1}$, где $i(\theta)_{jk} = E_\theta \frac{\partial l_{X_1}(\theta)}{\partial\theta_j} \frac{\partial l_{X_1}(\theta)}{\partial\theta_k}$.*
3. *$[L1 - L9] \Rightarrow$ решение уравнения и есть ОМП.*

Задача 1. $X_1, \dots, X_n \sim Exp(\theta)$. Найти ОМП для θ и $1/\theta$.

Решение: $p_\theta(x) = \theta e^{-\theta x} \cdot I\{x > 0\}$. Отсюда

$$L_X(\theta) = \prod_{i=1}^{n} \theta e^{-\theta X_i} \cdot I\{X_i > 0\} = \theta^n e^{-\theta \sum X_i} \cdot I\{\forall i\ X_i > 0\}.$$

Прологарифмируем:

$$l_X(\theta) = n \ln\theta - \theta \sum_{i=1}^{n} X_i.$$

$$\frac{\partial l_X(\theta)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} X_i = 0 \Rightarrow \hat{\theta} = \frac{1}{\overline{X}}.$$

По утверждению о независимости от способа параметризации $\overline{X}$ — ОМП для $1/\theta$, $i(\theta) = E_\theta \left(\frac{\partial l_{X_1}(\theta)}{\partial}\right)^2 = E_\theta \left(\frac{1}{\theta} - X_1\right)^2 = D_\theta X_1 = \frac{1}{\theta^2} \Rightarrow \hat{\theta} = \frac{1}{\overline{X}}$ — асимптотически нормальная оценка θ с асимптотической дисперсией $i(\theta)^{-1} = \theta^2$. $\square$

Лекция 4

Пример 9. $X_1, \dots, X_n \sim Bern(\theta)$. Найти ОМП для θ и $\ln \frac{\theta}{1-\theta}$.

Решение: $p_\theta(x) = P_\theta(X_1 = x) = \begin{cases} \theta, & x = 1 \\ 1-\theta, & x = 0 \end{cases} = \theta^x (1-\theta)^{1-x}$.

$$L_X(\theta) = \prod_{i=1}^{n} p_\theta(X_i) = \prod_{i=1}^{n} \theta^{X_i} (1-\theta)^{1-X_i} = \theta^{\sum X_i} (1-\theta)^{n - \sum X_i}$$

$$l_X(\theta) = \ln L_X(\theta) = \sum X_i \ln\theta + (n - \sum X_i)\ln(1-\theta)$$
$$\frac{\partial l_X(\theta)}{\partial\theta} = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta} = 0$$
$$(1-\theta)\sum X_i = \theta(n - \sum X_i)$$
$$\sum X_i = n\theta \Rightarrow \theta = \overline{X}.$$

По свойству независимости от способа параметризации ОМП для $\ln\dfrac{\theta}{1-\theta}$ это $\ln\dfrac{\overline{X}}{1-\overline{X}}$. Посчитаем асимптотическую для $\hat\theta = \overline{X}$. $i(\theta) = E_\theta\left(\dfrac{\partial l_{X_1}(\theta)}{\partial\theta}\right)^2 = E_\theta\left(\dfrac{X_1}{\theta} - \dfrac{1-X_1}{1-\theta}\right)^2 = \dfrac{1}{\theta^2(1-\theta)^2}E_\theta((1-\theta)X_1 - \theta(1-X_1))^2 = \dfrac{1}{\theta^2(1-\theta)^2}E_\theta(X_1-\theta)^2 = \dfrac{1}{\theta^2(1-\theta)^2}D_\theta X_1 = \dfrac{\theta(1-\theta)}{\theta^2(1-\theta)^2} = \dfrac{1}{\theta(1-\theta)}$. $\sigma^2(\theta) = 1/i(\theta) = \theta(1-\theta)$. $\square$

Задача 2. На высоте 1м от поверхности находится γ-излучатель. Регистрируются точки пересечения с горизонтальной осью. Направление равномерно распределено по полуокружности. Оценить θ.

Решение: x — точка пересечения с осью, α_x — угол, который образует точка x. Найдем распределение x. Заметим, что оно симметрично относительно θ. При $x \geqslant 0$: $F_\theta(x) = P_\theta(X \leqslant x) = P_\theta(X \leqslant \theta) + P_\theta(\theta \leqslant X \leqslant x) = \dfrac{1}{2} + \dfrac{\alpha_x}{\pi} = \dfrac{1}{2} + \dfrac{\arctan(x-\theta)}{\pi}$.

$$p_\theta(x) = F'_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)} \text{ — распределение Коши.}$$

1. Метод моментов неприменим, т. к. несуществует $E_\theta X_1$.
2. Метод максимизации правдоподобия:
$$L_X(\theta) = \prod_{i=1}^n \frac{1}{\pi(1+(X_i-\theta)^2)};$$
$$l_X(\theta) = -\sum_{i=1}^n \ln(1+(X_i-\theta)^2);$$
$$\frac{\partial l_X(\theta)}{\partial\theta} = 2\sum_{i=1}^n \frac{X_i-\theta}{1+(X_i-\theta)^2} = 0.$$
Дальше решать это грустно.
3. Почему бы не взять $\hat\theta = \overline{X}$? Посчитаем распределение $\overline{X}$: $\varphi_X(t) = \mathbb{E}e^{itX}$. Для Коши $\varphi_{X_1} = e^{-|t|}$ $(\theta = 0)$.
$$\varphi_{\overline{X}}(t) = Ee^{it\overline{X}} = Ee^{it\frac{1}{n}\sum X_i} = \mathbb{E}\prod_{i=1}^n e^{i(t/n)X_i} = /\text{незав.}/ = \prod_{i=1}^n Ee^{i(t/n)X_i} = |X_i \overset{d}{=} X_1| =$$
$$= \left(Ee^{i(t/n)X_1}\right)^n = e^{-|t|} = \varphi_{X_1}(t) \Rightarrow \text{ по теореме о единственности } \overline{X} \overset{d}{=} X_1.$$
Вывод: усреднение ничего не дает.

4. Медиана - рассмотрим далее.

$\square$

Выборочные квантили

Определение 7. Пусть P — распределение на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ с функцией распределение $F(X)$. Пусть $p \in (0,1)$. Тогда p-квантилью распределения P называется $u_p = \min\{x | F(x) \geqslant p\}$; 1/2-квантиль называется медианой.
Пример 10. $Exp(1)$, $F(x) = 1 - e^{-x}$, $u_p = -\ln(1-p)$ — p-квантиль $Exp(1)$.
Определение 8. Пусть $X = (X_1, \dots X_n)$ — выборка. Выборочной p-квантилью называется $\hat{u}_p = X_{(\lceil np \rceil)}$. Выборочной медианой

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

Пример 11. $X = (7, 9, 15, 8, 12, 1, 8, 5, 17, 21)$. Найти выборочные квантили уровней 0.01, 0.1, 0.25 и медиану.

Решение: Сортируем: $(1, 5, 7, 8, 8, 9, 12, 15, 17, 21)$. $\hat{\mu} = \dfrac{8+9}{2} = 8.5$. $\hat{u}_{0.01} = X_{(\lceil 10 \cdot 0.01 \rceil)} = X_{(1)} = 1$. $\hat{u}_{0.1} = X_{(1)} = 1$. $\hat{u}_{0.25} = X_{(\lceil 10 \cdot 0.25 \rceil)} = X_{(3)} = 7$. $\square$

Теорема 5. *Пусть $(X_n, n \in \mathbb{N})$ — выборка неограниченного размера из распределения P с плотностью $f(x)$. Число $p \in (0,1)$, такое что $f(x)$ непрерывна в окрестности u_p и $f(u_p) > 0$. Тогда*

$$\sqrt{n}(\hat{u}_p - u_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(u_p)}\right).$$

Аналогично для выборочной медианы

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

Вспомним про γ-котиков. $\hat{\mu}$ — а.н.о. θ с асимптотической дисперсией $\dfrac{1}{4\frac{1}{\pi^2(1-\theta)^2}} = \dfrac{\pi^2}{4} \approx 2.47$. При этом $i(\theta) = 1/2 \Rightarrow 1/i(\theta) = 2$ — асимптотическая дисперсия ОМП.

2.5. Достаточные статистики

Определение 9. Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P \in \mathcal{P}$, где $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$. Статистика $S(X)$ называется *достаточной для семейства $\mathcal{P}$*, если условное распределение $P_\theta(X \in B | S(X))$ не зависит от θ $\forall B$.
Смысл 3. вся информация о θ, которая содержится содержится в выборке, содержится в достаточной статистике.
Следствие 1. *если данные поступают последовательно, можно только пересчитывать $S(X)$.*
Пример 12. $X_1, \dots, X_n \sim Bern(\theta)$. Какая информация есть в выборке?

1. $S(X) = \sum X_i$ — количество единиц.

2. Порядок нулей и единиц — бесполезная информация, так как выборка.

Покажем, что $S(X)$ — достаточная статистика.

$$\frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, \sum X_i = s)}{P_\theta(\sum X_i = s)} = \frac{\theta^{\sum X_i}(1-\theta)^{n-\sum X_i} \cdot I\{\sum X_i = s\}}{C_n^s \theta^s (1-\theta)^{n-s}} =$$

$$= \frac{1}{C_n^s} I\{\sum X_i = s\} \text{ — не зависит от } \theta \Rightarrow S(X) \text{ - достаточная статистика}$$

Теорема 6 (критерий факторизации Неймана-Фишера)**.** *Пусть $X = (X_1, \ldots X_n)$ — выборка из распределение $P \in \mathcal{P} = \{P_\theta | \theta \in \Theta\}$, причем $\mathcal{P}$ — доминируемое семейство с плотностью $p_\theta(x)$. Тогда $S(X)$ — достаточная статистика для $\mathcal{P} \Leftrightarrow$ справедлива факторизация:*

$$p_\theta(x) = \psi(S(x), \theta) \cdot h(x),$$

$h(x)$ не зависит от θ.

Доказательство: (для дискретного случая):

($\Rightarrow$) Пусть $S(X)$ — достаточная статистика.

$$p_\theta(x) = P_\theta(X = x) = P_\theta(X = x, S(X) = S(x)) =$$

$$= \underbrace{P_\theta(X = x | S(X) = S(x))}_{\text{не зависит от} \theta} \cdot \underbrace{P_\theta(S(X) = S(x))}_{\text{зависит только от } S(x)} = h(x) \cdot \psi(S(x), \theta).$$

($\Leftarrow$) Пусть имеет место факторизация. Покажем, что $P_\theta(X = x | S(X) = s)$ не зависит от θ. Если $S(x) \neq s$, то вероятность $= 0$.

$$P_\theta(X = x | S(X) = S(x)) = \frac{P_\theta(X = x, S(X) = S(x))}{P_\theta(S(X) = S(x))} = \frac{P_\theta(X = x)}{\sum\limits_{y: S(y) = S(x)} P_\theta(X = y)} =$$

$$= \frac{p_\theta(x)}{\sum\limits_{y: S(y) = S(x)} p_\theta(y)} = \frac{\psi(S(x), \theta) h(x)}{\sum\limits_{y: S(y) = S(x)} \psi(S(y), \theta) h(y)} = \frac{h(x)}{\sum\limits_{y: S(y) = S(x)} h(y)} \text{ — не зависит от } \theta.$$

$\square$

Пример 13. $X_1, \ldots, X_n \sim \Gamma(\alpha, \beta)$. Найти достаточные статистики.

Решение:

$$p_\theta(x) = \frac{\alpha^\beta}{\Gamma(\beta)} x^{\beta - 1} e^{-\alpha x},\ x > 0$$

$$p_\theta(x_1, \ldots, x_n) = \frac{\alpha^{n\beta}}{\Gamma^n(\beta)} \left(\prod_{i=1}^n x_i \right)^{\beta - 1} e^{-\alpha \sum x_i}.$$

Вывод: $(\sum X_i, \prod X_i)$ — достаточная статистика.

Лучше $(\sum X_i, \sum \ln X_i)$. $\square$

Лекция 5 (от 30.09)

2.6. Экспоненциальный класс распределений

Определение 10. Семейство распределений $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ принадлежит *экспоненциальному классу*, если плотность $p_\theta(x)$ имеет вид

$$p_\theta(x) = \frac{g(x)}{h(\theta)} e^{a(\theta)^T u(x)},$$

где $g(x) > 0,\ u(x)$ — произвольные борелевские функции, $h(\theta) = \int\limits_{\mathscr{X}} g(x) e^{a(\theta)^T u(x)} dx$ — нормировочная константа. Если $a(\theta) = \theta$, будем говорить что *параметризация естественная.*

Пример 14. $\mathcal{P} = \{\mathcal{N}(a, \sigma^2 | a \in \mathbb{R}, \sigma > 0\}$. Перейдем к естественным параметрам:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{xa}{\sigma^2} - \frac{a^2}{2\sigma^2}\right).$$

Введем параметры $\theta = (\theta_1, \theta_2)$: $\theta_1 = -\dfrac{1}{2\sigma^2},\ \theta_2 = \dfrac{a}{\sigma^2}$.

$$p(x) = \sqrt{-\frac{\theta_1}{\pi}} e^{\theta_1 x^2 + \theta_2 x + \frac{\theta_2^2}{4\theta_1}}.$$

$u(x) = \begin{pmatrix} x^2 \\ x \end{pmatrix}, a(\theta) = \theta,\ g(x) = 1,\ h(\theta) = \sqrt{-\dfrac{\theta_1}{\pi}} e^{\frac{\theta_2^2}{4\theta_1}}$. Найдем достаточные статистики для семейства $\mathcal{P}$:

$$p_\theta(x_1, \ldots, x_n) = h^{-n}(\theta) \prod_{i=1}^n g(x_i) e^{a(\theta)^T \sum\limits_{i=1}^n u(x_i)}.$$

По критерию факторизации Неймана-Фишера $S(X) = \sum u(X_i)$ — достаточная статистика.

Замечание 5. $S(X)$ — статистика фиксированной размерности.

Теорема 7. *Пусть $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ — семейство распределений т.ч. плотность $p_\theta(x)$ непрерывно дифференцируема по x и носитель не зависит от θ. Пусть также $S(X)$ — достаточная статистика фиксированной размерности m. Тогда семейство $\mathcal{P}$ принадлежит экспоненциальному классу.*

Следствие 2. *Если плотность достаточно хорошая, то только семейства из экспоненциального класса допускают сжатие данных с помощю достаточных статистик.*

Пример 15.

1. $\mathcal{P} = \{\text{Коши со сдвигом}\}$ не лежит в экспоненциальном классе $\Longrightarrow$ нет достаточных статистик фиксированного размера.
2. $\mathcal{P} = \{U[0, \theta]\}$ — носитель зависит от θ. Однако достаточная статистика фикс. размера существует: $S(X) = X_{(n)}$.

Далее потребуем некоторые условия:

1. Параметризация естественная
2. $g(x),\ u(x)$ непрерывны

3. Условие равномерной сходимости интеграла по параметру:

$$\forall s\ \forall j \leqslant k\ \exists \varphi(x) : \forall \theta \in \Theta\ |g(x)u_s^j(x)e^{\theta u(x)}| \leqslant \varphi(x),$$

и при этом $\int\limits_{\mathscr{X}} \varphi(x)dx$ сходится.

Следствие 3.

1. *$h(\theta)$ непрерывно дифференцируема k раз*
2. *$p_\theta(x)$ непрерывно дифференцируема k раз по θ*
3. *Можно менять местами $\frac{\partial}{\partial\theta}$ и $\int$*

Утверждение 8.

1.
$$E_\theta u(X_1) = \nabla \ln h(\theta) = \left(\frac{\partial}{\partial\theta}\ln h(\theta)\right)_j$$

2.
$$D_\theta u(X_1) = \nabla^2 \ln h(\theta) = \left(\frac{\partial^2}{\partial\theta^2}\ln h(\theta)\right)_{jk}$$

Доказательство:

$$\frac{\partial h(\theta)}{\partial \theta_j} = \frac{\partial}{\partial\theta}\int\limits_{\mathscr{X}} g(x)e^{\theta^T u(x)}dx = \{\text{следствие 3}\} = \int\limits_{\mathscr{X}} u_j(x)g(x)e^{\theta^T u(x)}dx =$$

$$= h(\theta)\int\limits_{\mathscr{X}} \frac{u_j(x)}{h(\theta)})g(x)e^{\theta^T u(x)}dx = h(\theta)E_\theta u_j(X_1).$$

$$E_\theta u_j(X_1) = \frac{\partial h(\theta)/\partial\theta_j}{h(\theta)} = \frac{\partial \ln h(\theta)}{\partial\theta}$$

□

Утверждение 9. *Если Θ — выпуклое множество, то ОМП существует и единственна.*

Доказательство: $\nabla\nabla \ln h(\theta) = D_\theta u(X_1) \geqslant 0 \implies \ln h(\theta)$ выпукла.

$$l_X(\theta) = \underbrace{\sum \ln g(X_i)}_{\text{не зависит от } \theta}\underbrace{-n\overbrace{\ln h(\theta)}^{\text{выпукла}}}_{\text{вогнута}} + \underbrace{\theta\sum u(X_i)}_{\text{линейна по } \theta} \implies l_X(\theta) \text{ вогнута.}$$

Значит, максимум существует и единственный. □

Утверждение 10. *Если Θ — выпуклое открытое множество, то выполнены условия L5-L9.*

Доказательство: L5-L7 выполнены из следствий 1-3

L8: $\frac{\partial \ln p_\theta(x)}{\partial\theta} = \frac{\partial}{\partial\theta}(\ln g(x) - n\ln h(\theta) + \theta u(x)) = \frac{\partial h(\theta)}{h(\theta)} + u(x)$

$i(\theta) = E_\theta(\frac{\partial \ln p_\theta(X_1)}{\partial\theta})^2$ по утверждению 1 существует и конечна

L9 следует из того, что $\frac{\partial^2 \ln p_\theta(X_1)}{\partial\theta^2}$ не зависит от θ. □

2.7. Сравнение оценок

Ранее было: $X_1, \ldots, X_n \sim Exp(\theta)$.

$\widehat{\theta}_1 = 1/\overline{X}$, $\widehat{\theta}_2 = -\ln \overline{I\{X > 1\}}$ — (сильно) состоятельная, а. н. оценка θ. Хотим построить оценку для $\tau(\theta) \in \mathbb{R}^d$.
Определение 11. Функция $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$, которая характеризует степень отклонения оценки от $\tau(\theta)$, называется *функцией потерь (loss function)*.
Пример 16.

1. $L(x, y) = (x - y)^2$ — квадратичная функция потерь
2. $L(x, y) = |x - y|$ — абсолютная функция потерь
3. $L(x, y) = \log(1 + |x - y|)$ многомерный случай:
4. $L(x, y) = (x - y)^T A(x - y)$, A — симметричная, положительно определенная матрица если $A = I_d : L(x, y) = \sum\limits_{j=1}^{d}(x_j - y_j)^2$.

Пусть $\widehat{\theta}$ — оценка $\tau(\theta)$, θ — истинное значение параметра. Тогда $L(\widehat{\theta}, \theta)$ — штраф при оценивании $\tau(\theta)$ оценкой $\widehat{\theta}$. Проблема: штраф случаен
Определение 12. Функция риска

$$R_{\widehat{\theta},\tau}(\theta) = E_\theta L(\widehat{\theta}, \tau(\theta)).$$

Пример 17.

- $\mathrm{MSE}_{\widehat{\theta},\tau}(\theta) = E_\theta(\widehat{\theta} - \tau(\theta))^2$ — *среднеквадратичная ошибка.*
- $\mathrm{MAE}_{\widehat{\theta},\tau}(\theta) = E_\theta|\widehat{\theta} - \tau(\theta)|$ — *средняя абсолютная ошибка.*

Замечание 6. если $\tau(\theta) = \theta$, то индекс τ опускаем.
Задача 3. $X_1, \ldots, X_n$ — выборка. $\widehat{\theta}_1 = X_1$, $\widehat{\theta}_2 = \overline{X}$ — оценки $\tau(\theta) = E_\theta(X_1)$. Посчитать MSE.

Решение: $\mathrm{MSE}_{\widehat{\theta}_1,\tau}(\theta) = E_\theta(X_1 - E_\theta X_1)^2 = D_\theta X_1$ $\qquad \mathrm{MSE}_{\widehat{\theta}_2,\tau}(\theta) = E_\theta(\overline{X} - E_\theta\overline{X})^2 = D_\theta\overline{X} = \frac{1}{n}D_\theta X_1$. $\square$

Вывод: усреднение уменьшает среднеквадратичный риск в n раз.

Подходы к сравнению оценок

1. Равномерный

- $\widehat{\theta}_1$ *не хуже* $\widehat{\theta}_2$, если $\forall\theta R_{\widehat{\theta}_1,\tau(\theta)} \leqslant R_{\widehat{\theta}_2,\tau(\theta)}$.
- $\widehat{\theta}_1$ *лучше* $\widehat{\theta}_2$, если, кроме того, $\exists\theta : R_{\widehat{\theta}_1,\tau(\theta)} < R_{\widehat{\theta}_2,\tau(\theta)}$.
- Пусть $\mathscr{K}$ — множество оценок. $\widehat{\theta}$ — *наилучшая в* $\mathscr{K}$, если она лучше всех оценок из $\mathscr{K}$.
- Если $L(x, y) = (x - y)^2$, то подход называется *среднеквадратичным.*

Утверждение 11. *Наилучшей оценки может не существовать.*

Доказательство: $\mathscr{K} = \{\widehat{\theta}_1 \equiv 1, \widehat{\theta}_2 \equiv 2\}$ $\qquad \mathrm{MSE}_{\widehat{\theta}_1}(\theta) = E_\theta(\theta - 1)^2 = (\theta - 1)^2$ $\qquad \mathrm{MSE}_{\widehat{\theta}_2}(\theta) = E_\theta(\theta - 2)^2 = (\theta - 2)^2$ Если $\theta < 1.5$, то $\mathrm{MSE}_{\widehat{\theta}_1}(\theta) < \mathrm{MSE}_{\widehat{\theta}_2}(\theta)$; если $\theta > 1.5$, то $\mathrm{MSE}_{\widehat{\theta}_2}(\theta) < \mathrm{MSE}_{\widehat{\theta}_1}(\theta)$
$\square$

Утверждение 12. *Справедливо bias-variance разложение:*

$$\underbrace{\text{MSE}_{\widehat{\theta},\tau}(\theta)}_{error} = \underbrace{D_\theta\widehat{\theta}}_{variance} + \underbrace{(E_\theta\widehat{\theta} - \theta)^2}_{bias^2}.$$

Доказательство: $\text{MSE}_{\widehat{\theta},\tau}(\theta) = E_\theta(\widehat{\theta} - \tau(\theta))^2 = E_\theta((\widehat{\theta} - E_\theta\widehat{\theta}) + (E_\theta\widehat{\theta} - \tau(\theta))^2 = E_\theta(\widehat{\theta} - E_\theta(\widehat{\theta}))^2 + 2E_\theta(\widehat{\theta} - E_\theta\widehat{\theta})(E_\theta\widehat{\theta} - \tau(\theta)) + (E_\theta(\widehat{\theta}) - \tau(\theta))^2$ Второе слагаемое равно нулю, следовательно, получаем требуемое. □

Следствие 4. *Среди все несмещенных оценок наилучшей будет та, у которой меньше дисперсия.*

2. Байесовский

Пусть Q — некоторое распределение на Θ. Тогда $\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $E_Q R_{\widehat{\theta}_1}(\theta) \leqslant E_Q R_{\widehat{\theta}_2}(\theta)$.

3. Минимаксный

$\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $\sup\limits_{\theta\in\Theta} R_{\widehat{\theta}_1}(\theta) \leqslant \sup\limits_{\theta\in\Theta} R_{\widehat{\theta}_2}(\theta)$.

4. Асимптотический (для а.н.о)

Пусть $\widehat{\theta}_1, \widehat{\theta}_2$ — а.н.о. $\tau(\theta)$ с асимпт. дисперсией σ_1^2 и σ_2^2. Тогда

- $\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $\sigma_1(\theta) \leqslant \sigma_2(\theta)\ \forall\theta \in \Theta$.
- $\widehat{\theta}_1$ лучше $\widehat{\theta}_2$, если, кроме того $\exists\theta \in \Theta:\ \sigma_1(\theta) < \sigma_2(\theta)$.
- *Относительная асимптотическая эффективность*: $\text{ARE}^\tau_{\widehat{\theta}_1,\widehat{\theta}_2}(\theta) = \dfrac{\sigma_2^2}{\sigma_1^2}$ показывает, насколько $\widehat{\theta}_1$ лучше $\widehat{\theta}_2$.

$\widehat{\theta}_1$ не хуже $\widehat{\theta}_2$, если $\text{ARE}^\tau_{\widehat{\theta}_1,\widehat{\theta}_2}(\theta) \geqslant 1 \forall\theta \in \Theta$.
Определение 13. Оценка $\widehat{\theta}$ называется *асимптотически эффективной оценкой* $\tau(\theta)$, если она имеет наименьшую асимптотическую дисперсию среди всех а.н.о. $\tau(\theta)$ с непрерывной а. д.
Утверждение 13. *Если выполнены условия L1-L9, то ОМП асимптотически эффективна.*
Пример 18. $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$.

- ОМП: $\widehat{\theta}_1 = \overline{X}$ — а.н.о θ с а.д. $\sigma_1^2 = 1$.
- Теор. о выборочной медиане: $\widehat{\theta}_2 = \widehat{\mu}$ — а.н.о θ с а.д. $\sigma_2^2 = \frac{2\pi}{4} = \frac{\pi}{2}$.

$\text{ARE}_{\overline{X},\widehat{\mu}}(\theta) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)} = \frac{\pi}{2} \approx 1.57$.

Лекция 6 (от 7.10)

2.8. Приближенный поиск ОМП

Метод Ньютона:

Пусть $f : \mathbb{R} \to \mathbb{R}$ — функция. Нужно решить уравнение $f(x) = 0$.

x_0 — начальное приближение Формула касательной в точке x_k : $y = f(x_k) + f'(x_k)(x - x_k)$. Получим соотношение

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta \mid \theta \in \Theta\}, \Theta \subset \mathbb{R}^d$. Пусть θ^* — ОМП. Хотим приблизить оценку θ^*.

Уравнение правдоподобия: $\dfrac{\partial l_X(\theta)}{\partial \theta} = 0$. Применим метод Ньютона для функции $l'_X(\theta)$. $\widehat{\theta}_0$ — начальное приближение. Шаг метода:

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k - \underbrace{(l''_X(\widehat{\theta}_k))^{-1}}_{\text{матрица}} \cdot \underbrace{l'_X(\widehat{\theta}_k)}_{\text{вектор}}.$$

Теорема 8. *В условиях регулярности $L1 - L9$, если $\widehat{\theta}_0$ — а.н.о, то*

1. *$\widehat{\theta}_1$ — а.н.о с асимт. дисперсией $(i(\theta))^{-1}$.*
2. *$\widehat{\theta}_1$ асимптотически эквивалентна ОМП θ^*, т.е*

$$\sqrt{n}(\widehat{\theta}_1 - \theta^*) \xrightarrow{P_\theta} 0.$$

Доказательство: (для $d = 1$, идея)
Утверждение 14 (б/д)**.** $\widehat{\theta}_1 - \theta^* = (\widehat{\theta}_0 - \theta^*)\varepsilon_n(\theta)$, *где* $\varepsilon_n(\theta) \xrightarrow{P_\theta} 0$.

(2). $\sqrt{n}(\widehat{\theta}_1 - \theta^*) = \sqrt{n}(\widehat{\theta}_0 - \theta^*)\varepsilon_n(\theta) =$

$= \underbrace{\sqrt{n}(\widehat{\theta}_0 - \theta)}_{\xrightarrow{d_\theta} \mathcal{N}(0,\dots)} \underbrace{\varepsilon_n(\theta)}_{\xrightarrow{d_\theta} 0} + \underbrace{\sqrt{n}(\theta - \theta^*)}_{\xrightarrow{d_\theta} \mathcal{N}(0,\dots)} \underbrace{\varepsilon_n(\theta)}_{\xrightarrow{d_\theta} 0}$. По лемме Слуцкого первое слагаемое $\xrightarrow{d_\theta} 0$, второе слагаемое $\xrightarrow{d_\theta} 0$. Применяя еще раз лемму Слуцкого для их суммы, получим $\sqrt{n}(\widehat{\theta}_1 - \theta^*) \xrightarrow{d_\theta (\iff P_\theta, \text{т.к const})} 0$.

(1). $\sqrt{n}(\widehat{\theta}_1 - \theta) = \underbrace{\sqrt{n}(\widehat{\theta}_1 - \theta^*)}_{\xrightarrow{P_\theta} 0 (\text{из } (2))} - \underbrace{\sqrt{n}(\widehat{\theta}_0 - \theta)}_{\xrightarrow{d_\theta} \mathcal{N}(0, \frac{1}{i(\theta)}) \text{ (ОМП)}}$. По лемме Слуцкого

$$\sqrt{n}(\widehat{\theta}_1 - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0, \frac{1}{i(\theta)}\right)$$

□

Замечание 7. Утверждение теоремы не изменится, если заменить $l''_X(\theta)$ на $E_\theta l''_X(\theta) = -ni(\theta)$, т.е.

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k + \frac{i(\widehat{\theta}_k)^{-1}}{n} l'_X(\widehat{\theta}_k).$$

Определение 14. Оценка $\widehat{\theta}_1$ называется *одношаговой оценкой.*
Смысл 4. Отклонение $\widehat{\theta}_1$ от θ^* на порядок меньше, чем отклонение θ^* от θ. Значит отклонение $\widehat{\theta}_1$ от θ тоже имеет порядок $\sqrt{\frac{1/i(\theta)}{n}}$.
Пример 19 (γ-котики). $\widehat{\mu}$ — а.н.о. с асимпт. дисперсией $\pi^2/4 \approx 2.47$. При этом $i(\theta) = 1/2$, т.е наименьшая возможная асимпт. дисперсия равна 2. Запишем одношаговую оценку:

$$\widehat{\theta}_1 = \widehat{\mu} + \frac{\sum\limits_{i=1}^{n} \frac{X_i - \widehat{\mu}}{1+(X_i-\widehat{\mu})^2}}{\sum\limits_{i=1}^{n} \frac{1-(X_i-\widehat{\mu})^2}{(1+(X_i-\widehat{\mu})^2)^2}}.$$

$\widehat{\theta}_1$ — наиболее асимптотически эффективная оценка.

2.9. Робастность и симметричные распределения

Пусть $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{N}(\theta, \sigma^2)$, σ известна. Оценка $\widehat{\theta} = \overline{X}$ обладает всеми хорошими свойствами (сильная состоятельность, асимптотическая нормальность, ОМП и т. д.). Однако если в данных есть выбросы, то все свойства теряются. Для того, чтобы визуализировать выбросы в данных, можно использовать *ящик с усами (box plot).*

Будем рассматривать только одномерный случай.
Определение 15. *Робастная оценка* — оценка, допускающая отклонение от заданной модели.
Определение 16. Пусть оценка имеет вид $\widehat{\theta} = f(X_{(1)}, \dots, X_{(n)})$. Пусть k_n^* — наименьшее число k, т. ч. выполнено одно из условий:

1. Если $x_1, \dots, x_{k+1} \to -\infty$, а $x_{k+2}, \dots, x_n$ фиксированы,то $f(x_1, \dots, x_n) \to -\infty$.
2. Если $x_{n-k}, \dots, x_n \to +\infty$, а $x_1, \dots, x_{n-k+1}$ фиксированы, то $f(x_1, \dots, x_n) \to +\infty$.

Тогда число $\tau_{\widehat{\theta}} = \lim\limits_{n\to\infty} \dfrac{k_n^*}{n}$ называется *асимптотической толерантностью оценки $\widehat{\theta}$.*
Смысл 5. $\tau(\theta)$ — наибольшая доля выбросов, которые способна выдержать оценка, не смещаясь на $\pm\infty$.
Пример 20.
- $\overline{X} : k_n^* = 0, \tau_{\overline{X}} = 0$
- $\widehat{\mu} : k_n^* = \lceil n/2 \rceil - 1, \tau_{\widehat{\mu}} = 1/2.$

Далее будем рассматривать класс распределений $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, т. ч.

- P_0 имеет плотность $p_0(x)$ — симметричная, непрерывная, носитель плотности имеет вид $(-c, c),\ 0 < c \leqslant +\infty$.
- θ — параметр сдвига, т. е. $p_\theta(x) = p_0(x - \theta$.

Будем искать оценки, которые:

1. Достаточно эффективные в классе $\mathcal{P}$ (в асимптотическом подходе).
2. Робастные — допускают отклонение от $\mathcal{P}$.

α	0	1/20	1/8	1/4	3/8	1/2
$\mathrm{ARE}_{\overline{X}_\alpha,\overline{X}}$	1	0.99	0.94	0.84	0.74	0.64

α	0	1/20	1/8	1/4	3/8	1/2
$(1-2\alpha)^2$	1	0.81	0.5	0.25	0.06	0

1. Усеченное среднее

Определение 17. Пусть $\alpha \in (0, 1/2)$, $k = \lceil \alpha n \rceil$. Тогда *усеченным средним по выборке* $X_1, \ldots, X_n$ называется оценка

$$\overline{X}_\alpha = \frac{1}{n-2k}(X_{(k-1)} + \cdots + X_{(n-k)}).$$

- $\alpha = 0$: $\overline{X}_\alpha = \overline{X}$
- $\alpha = 1/2 : \overline{X}_\alpha = \widehat{\mu}$.

Асимптотическая толерантность: $\tau_{\overline{X}_\alpha} = \alpha$.
Теорема 9 (б/д). *Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения $P \in \mathcal{P}$. Тогда*

$$\sqrt{n}(\overline{X}_\alpha - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma_\alpha^2), \text{ где}$$

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left(\int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \alpha u_{1-\alpha}^2\right),$$

$u_{1-\alpha}$ — $(1-\alpha)$*-квантиль распределения* P_0.
Пример 21. для $\mathcal{N}(0,1)$

При $\alpha = 1/8$ достигается защита от 12.5% загрязнения выборки, но эффективность теряется на 6%.
Утверждение 15. *Если $D_\theta X_1 < +\infty$, то* $\mathrm{ARE}_{\overline{X}_\alpha,\overline{X}} \geqslant (1-2\alpha)^2$.

Доказательство: $\overline{X}_\alpha$ — а.н.о θ с асимпт. дисперсией σ_α^2. Из ЦПТ: $\overline{X}$ — а.н.о θ с асимпт. дисперсией $D_\theta X_1$. Так как дисперсия не зависит от сдвига, посчитаем дисперсию при $\theta = 0$:

$$\frac{1}{2}D_\theta X_1 = \frac{1}{2}\int\limits_{\mathbb{R}} x^2 p_0(x)dx = \int\limits_0^{+\infty} x^2 p_0(x)dx =$$

$$= \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \int\limits_{u_{1-\alpha}}^{+\infty} x^2 p_0(x)dx \geqslant$$

$$\geqslant \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + u_{1-\alpha}^2 \underbrace{\int\limits_{u_{1-\alpha}}^{+\infty} p_0(x)dx}_{=\alpha} = \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \alpha u_{1-\alpha}^2 = \frac{\sigma_\alpha^2(1-2\alpha)^2}{2}.$$

Отсюда $\mathrm{ARE}_{\overline{X}_\alpha,\overline{X}} = \dfrac{D_\theta X_1}{\sigma_\alpha^2} \geqslant (1-2\alpha)^2$ $\square$

При $\alpha = 1/8$ возможна потеря эффективности до 44%.

2. Медиана средних Уолша

Определение 18. $Y_{ij} = \dfrac{X_i + X_j}{2}$ — *среднее Уолша.*

$W = \mathrm{med}\{Y_{ij},\ 1 \leqslant i \leqslant j \leqslant n\}$ — *медиана средних Уолша.*
Теорема 10. *Пусть* $X = (X_1, \ldots, X_n)$ — *выборка из распределения* $P \in \mathcal{P}$. *Тогда*

$$\sqrt{n}(W - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2), \text{ где}$$

$$\sigma^2 = \frac{1}{12\left(\int\limits_{\mathbb{R}} p_0^2(x)dx\right)^2}.$$

Пример 22. $\mathcal{N}(0,1) : \mathrm{ARE}_{W,\overline{X}} \approx 0.955$ (потеря эффективности на 4.5%).
Утверждение 16. *Для* $P_\theta \in \mathcal{P}$ $\mathrm{ARE}_{W,\overline{X}} \geqslant \frac{108}{125} = 0.864$ *(в худшем случае теряем* 14% *эффективности). Равенство достигается при*

$$p_0(x) = \frac{3\sqrt{5}}{100}(5 - x^2)I\{|x| < \sqrt{5}\}.$$

Утверждение 17. $\tau_W \approx 0.293$ *(доказательство см. в ДЗ).*

2 | Глава 3. Сложные оценки параметров

3.1. Доверительные интервалы

Определение 19. Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta \mid \theta \in \Theta\}$.

- Если $\Theta \subset \mathbb{R}$, то пара статистик $(T_1(X), T_2(X))$ называется *доверительным интервалом для θ уровня доверия α,* если

$$\forall \theta \in \Theta \quad P_\theta(T_1(X) \leqslant \theta \leqslant T_2(X)) \geqslant \alpha.$$

- Если $\Theta \subset \mathbb{R}^d$, то статистика $S(X) \subset \Theta$ называется *доверительной областью для θ уровня доверия α,* если

$$\forall \theta \in \Theta \quad P_\theta(\theta \in S(X)) \geqslant \alpha.$$

- Если равенство точное, то интервал называтся *точным.*

Замечание 8. 1. Если $X = (X_1, \dots, X_n)$ — выборка, то утверждение $P_\theta(T_1(X) \leqslant \theta \leqslant T_2(X)) = \alpha$ имеет смысл ($(T_1(X), T_2(X))$ — доверительный интервал).

2. Если $x = (x_1, \dots, x_n)$ — реализация выборки, то утверждение $P_\theta(T_1(x) \leqslant \theta \leqslant T_2(x)) = \alpha$ некорректно.

$(T_1(x), T_2(x))$ — *реализация доверительного интервала.*

Первая магическая константа статистики: $\alpha = 0.95$ (она же 0.05).

Лекция 7 (от 14.10)

Методы поиска доверительных интервалов

1. Метод центральной функции

Пусть $G(X, \theta)$ — функция, распределение которой известно и не зависит от θ *(центральная функция).* Возьмем $\alpha_1, \alpha_2 \in (0, 1)$ т. ч. $\alpha_2 - \alpha_1 = \alpha$ и g_j — α_j-квантиль распределения $G(X, \theta)$. Тогда $S(X) = \{\theta \in \Theta | g_1 \leqslant G(X, \theta) \leqslant g_2\}$ — доверительная область уровня доверия α. Действительно, $P_\theta(\theta \in S(X)) = P_\theta(g_1 \leqslant G(X, \theta) \leqslant g_2) = \alpha_2 - \alpha_1 = \alpha$.

Пример 23. $X_1, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, σ известно. Построить точные доверительные интервалы для θ.

Решение: аметим, что $X_i - \theta \sim \mathcal{N}(0, \sigma^2)$, следовательно, $\overline{X} - \theta \sim \mathcal{N}(0, \frac{\sigma^2}{n})$. $G(X, \theta) = \sqrt{n}\dfrac{\overline{X} - \theta}{\sigma} \sim \mathcal{N}(0,1)$ — центральная функция. Будем обозначать через z_p p-квантили распределения $\mathcal{N}(0,1)$. Тогда

$$P_\theta\left(-z_{\frac{1+\alpha}{2}} \leqslant \sqrt{n}\frac{\overline{X} - \theta}{\sigma} \leqslant z_{\frac{1+\alpha}{2}}\right) = \alpha \implies P_\theta\left(\overline{X} - \frac{z_{\frac{1+\alpha}{2}}\sigma}{\sqrt{n}} \leqslant \theta \leqslant \overline{X} + \frac{z_{\frac{1+\alpha}{2}}\sigma}{\sqrt{n}}\right) = \alpha.$$

□

Ответ: $\left(\overline{X} \pm \dfrac{z_{\frac{1+\alpha}{2}}\sigma}{\sqrt{n}}\right)$.

Пусть $\alpha = 0.95 \implies z_{\frac{1+\alpha}{2}} = z_{0.975} \approx 1.96 \approx 2$. $n = 100, \overline{x} = 5, \sigma = 1$. Тогда реализация интервала $(5 \pm 2/10) = (4.8, 5.2)$.

2. Асимптотические доверительные интервалы

Определение 20. Пусть $X = (X_1, X_2, \dots)$ — выборка неограниченного размера из распределения $P \in \{P_\theta | \theta \in \Theta\}$. Последовательность пар статистик $(T_1^{(n)}(X_1, \dots, X_n), T_2^{(n)}(X_1, \dots, X_n))$ называется *асимптотическим доверительным интервалом* уровня доверия α, если

$$\forall \theta \in \Theta \liminf_{n\to\infty} P_\theta(T_1^{(n)}(X_1, \dots, X_n) \leqslant \theta \leqslant T_2^{(n)}(X_1, \dots, X_n)) \geqslant \alpha.$$

Он называется *точным*, если

$$\forall \theta \in \Theta \lim_{n\to\infty} P_\theta(T_1^{(n)} \leqslant \theta \leqslant T_2^{(n)}) = \alpha.$$

Метод построения асимптотического доверительного интервала:

1. Пусть $\widehat{\theta}$ — а.н.о θ с асимпт. дисперсией $\sigma^2(\theta)$.

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2(\theta)).$$

2. Поделим все на $\sigma(\theta)$:

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

   Из теоремы Александрова

$$P_\theta\left(\frac{\sqrt{n}|\widehat{\theta} - \theta|}{\sigma(\theta)} \leqslant z_{\frac{1+\alpha}{2}}\right) \to \alpha.$$

   Проблема: $\sigma(\theta)$ может плохо зависеть от θ.

3. Пусть $\widehat{\sigma}$ — состоятельная оценка $\sigma(\theta)$. Тогда

$$\sqrt{n}\frac{\widehat{\theta} - \theta}{\widehat{\sigma}} = \underbrace{\sqrt{n}\frac{\widehat{\theta} - \theta}{\sigma(\theta)}}_{\xrightarrow{d_\theta}\mathcal{N}(0,1)} \cdot \underbrace{\frac{\sigma(\theta)}{\widehat{\sigma}}}_{\xrightarrow{P_\theta} 1 \text{ (th о насл. сх-тей)}}.$$

   По лемме Слуцкого $\sqrt{n}\dfrac{\widehat{\theta} - \theta}{\widehat{\sigma}} \xrightarrow{d_\theta} \mathcal{N}(0,1)$.

4. $P_\theta\left(\frac{\sqrt{n}|\widehat{\theta}-\theta|}{\widehat{\sigma}} \leqslant z_{\frac{1+\alpha}{2}}\right) \to \alpha$. Получаем интервал $\left(\widehat{\theta} \pm \frac{z_{\frac{1+\alpha}{2}}\widehat{\sigma}}{\sqrt{n}}\right)$ — точный асимптотический доверительный интервал уровная доверия α.

5. Откуда взять $\widehat{\sigma}$? Если $\sigma(\theta)$ непрерывна, то по теореме о наследовании сходимостей $\widehat{\sigma} = \sigma(\widehat{\theta})$ — состоятельная оценка $\sigma(\theta)$.

Пример 24.

1. $X_1, \ldots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, σ неизвестна. Построить асимптотический доверительный интервал уровня доверия α для θ. △ $\overline{X}$ — а.н.о θ с асимпт. дисперсией σ^2. S — состоятельная оценка θ. Получаем интервал $\left(\overline{X} \pm z_{\frac{1+\alpha}{2}}\frac{S}{\sqrt{n}}\right)$. □

2. $X_1, \ldots, X_n \sim Pois(\theta)$. Построить асимптотический доверительный интервал уровня доверия α для θ. △ $\overline{X}$ — а.н.о θ с асимпт. дисперсией $\sigma^2(\theta) = \theta$. $\sqrt{\overline{X}}$ — состоятельная оценка $\sigma(\theta) = \sqrt{\theta}$. Получаем интервал $\left(\overline{X} \pm z_{\frac{1+\alpha}{2}}\sqrt{\frac{\overline{X}}{n}}\right)$. □

Замечание 9. При $n = 30$ условие ЦПТ применимо с хорошей точностью. Поэтому при $n \geqslant 30$ имеет смысл пользоваться асимптотическими доверительными интервалами.

3.2. Точные доверительные интервалы в нормальной модели

Пусть $X = (X_1, \ldots, X_n) \sim \mathcal{N}(a, \sigma^2)$.

1. Интервал для a, если σ известна

Уже получили: $\left(\overline{X} \pm z_{\frac{1+\alpha}{2}}\frac{S}{\sqrt{n}}\right)$.

2. Интервал для σ, если a известно

$\frac{X_i - \theta}{\sigma} \sim \mathcal{N}(0,1)$

$G(X,\theta) = \sum\limits_{i=1}^{n}\left(\frac{X_i - a}{\sigma}\right)^2 \sim \chi^2_n$ — центральная функция *(распределение хи-квадрат с n степенями свободы)*

$$P_\theta\left(\chi^2_{n,\frac{1-\alpha}{2}} \leqslant \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - a)^2 \leqslant \chi^2_{n,\frac{1+\alpha}{2}}\right) = \alpha$$

Получаем интервал $\left(\sqrt{\frac{\sum(X_i - a)^2}{\chi^2_{n,\frac{1+\alpha}{2}}}}, \sqrt{\frac{\sum(X_i - a)^2}{\chi^2_{n,\frac{1-\alpha}{2}}}}\right)$.

3. Интервал для a, если σ неизвестна

Теорема 11. *Пусть $X = (X_1, \ldots, X_n) \sim \mathcal{N}(a, \sigma^2)$. Тогда:*

1. *Статистики $\overline{X}$ и S^2 независимы*

2. $\dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

3. $\sqrt{n-1}\dfrac{\overline{X}-a}{S} \sim T_{n-1}$ — распределение Стьюдента с $n-1$ степенями свободы.

Доказательство: 1), 2) — позже 3) $\sqrt{n}\dfrac{\overline{X}-a}{\sigma} \sim \mathcal{N}(0,1)$; $\dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

Свойство распределения Стьюдента: если $\xi \sim \mathcal{N}(0,1), \eta \sim \chi^2_k$ — независимые с.в., то $\zeta = \frac{\xi}{\sqrt{\eta/k}} \sim T_k$. Следовательно:

$$\frac{\sqrt{n}\dfrac{\overline{X}-a}{\sigma}}{\sqrt{\dfrac{nS^2}{\sigma^2}\cdot\dfrac{1}{n-1}}} = \sqrt{n-1}\frac{\overline{X}-a}{S} \sim T_{n-1}. \quad \square$$

$\square$

$G(X,\theta) = \sqrt{n-1}\dfrac{\overline{X}-\theta}{S}$ — центральная функция.

Получаем интервал $\left(\overline{X} \pm T_{n-1,\frac{1+\alpha}{2}}\dfrac{S}{\sqrt{n-1}}\right)$.

Замечание 10. При больших n интервал почти совпадает с интервалом из пункта 1.

4. Интервал для σ, если a неизвестно

$G(X,\sigma) = \dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$ — центральная функция.

Аналогично п.2 получаем интервал $\left(\sqrt{\dfrac{nS^2}{\chi^2_{n,\frac{1+\alpha}{2}}}}, \sqrt{\dfrac{nS^2}{\chi^2_{n,\frac{1-\alpha}{2}}}}\right)$.

Теорема 12 (о разложении гауссовского вектора). *Пусть $\xi = (\xi_1,\ldots,\xi_n) \sim \mathcal{N}(a,\sigma^2 I_n)$, $\mathbb{R}^n = \mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_k$ — разложение в прямую сумму ортогональных подпространств, $\eta_j = \operatorname{proj}_{\mathcal{L}_j}\xi$ — проекция на $\mathcal{L}_j$. Тогда:*

1. *$\eta_1,\ldots,\eta_k$ независимы в совокупности;*

2. *$\mathbb{E}\eta_j = \operatorname{proj}_{\mathcal{L}_j} a$;*

3. *$\frac{1}{\sigma^2}\|\eta_j - \mathbb{E}\eta_j\|^2 \sim \chi^2_{d_j}$, где $d_j = \dim\mathcal{L}_j$.*

Доказательство: Выберем ортонормированный базис в $\mathbb{R}^n$ следуюзим образом:

$$\underbrace{e_1, e_2, \ldots}_{\text{базис в } \mathcal{L}_1} \underbrace{\ldots\ldots}_{\text{базис в } \mathcal{L}_2} \cdots \underbrace{\ldots e_n}_{\text{базис в } \mathcal{L}_k}.$$

Обозначим:

- I_j — набор индексов, соответствующий базису в $\mathcal{L}_j$;
- $B = (e_1,\ldots,e_n) \in \mathbb{R}^{n\times n}$ — ортогональная матрица;
- $\zeta_i = \langle \xi, e_i\rangle = e^T\xi$ — проекция на e_i.

Получаем:

$$\zeta = \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_n \end{pmatrix} = \begin{pmatrix} e_1^T \xi \\ \vdots \\ e_n^T \xi \end{pmatrix} = B^T \xi$$

$$\xi = \sum_{i=1}^{n} \langle \xi, e_i \rangle \cdot e_i = \sum_{i=1}^{n} \zeta_i e_i = (e_1 \dots e_n) \cdot \zeta$$

$\xi = B\zeta$

- $\mathbb{E}\zeta = \mathbb{E}B^T\xi = B^T\mathbb{E}\xi = B^T a$
- $\mathbb{D}\zeta = \mathbb{D}B^T\xi = B\mathbb{D}\xi B^T = B\sigma^2 I_n B^T = \sigma^2 \underbrace{BB^T}_{=I_n} = \sigma^2 I_n$

Вывод: ζ — гауссовский вектор с независимыми компонентами.

$$\eta_j = \operatorname{proj}_{\mathcal{L}_j} \xi = \sum_{i \in I_j} \langle \xi, e_i \rangle e_i = \sum_{i \in I_j} \zeta_i e_i.$$

Компоненты вектора ζ в разных η_j не пересекаются, следовательно, $\eta_1, \dots, \eta_k$ независимы в совокупности — утв. 1 доказано;

$\mathbb{E}\eta_j = \sum_{i \in I_j} \langle \mathbb{E}\xi, e_i \rangle e_i = \sum_{i \in I_j} \langle a, e_i \rangle e_i = \operatorname{proj}_{\mathcal{L}_j} a$ — утв. 2 доказано;

$$\frac{1}{\sigma^2} \|\eta_j - \mathbb{E}\eta_j\|^2 = \frac{1}{\sigma^2} \left\| \sum_{i \in I_j} \langle \xi - a, e_i \rangle e_i \right\|^2 = \sum_{i \in I_j} \underbrace{\left(\frac{\zeta_i - \mathbb{E}\zeta_i}{\sigma} \right)^2}_{\sim \mathcal{N}(0,1) \text{ и незав.}} \sim \chi^2_{\dim \mathcal{L}_j}. \quad \square$$

$\square$

Доказательство пп. 1-2 из предыдущей теоремы:

1. $$\mathbb{R}^n = \mathcal{L} \oplus \mathcal{L}^\perp, \text{ где } \mathcal{L} = \left\langle \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right\rangle.$$

$$\operatorname{proj}_{\mathcal{L}} X = \arg\min_{c \in \mathbb{R}} \left\| X - \begin{pmatrix} c \\ c \\ \vdots \\ c \end{pmatrix} \right\|^2 = \arg\min_{c \in \mathbb{R}} \sum_{i=1}^{n} (X_i - c)^2 = \begin{pmatrix} \overline{X} \\ \overline{X} \\ \vdots \\ \overline{X} \end{pmatrix}.$$

$$\operatorname{proj}_{\mathcal{L}^\perp} X = X - \operatorname{proj}_{\mathcal{L}} X = \begin{pmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}.$$

По теореме о разложении гауссовского вектора $\overline{X}$ и $(X_1 - \overline{X}, \dots, X_n - \overline{X})$ независимы, а S^2 зависит только от $(X_1 - \overline{X}, \dots, X_n - \overline{X})$. Вывод: $\overline{X}$ и S^2 независимы.

2. Докажем, что $\dfrac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$:
$$\frac{1}{\sigma^2}\|\operatorname{proj}_{\mathcal{L}^\perp} X - \mathbb{E}\operatorname{proj}_{\mathcal{L}^\perp} X\|^2 = \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$
по теореме о разложении гауссовского вектора. $\square$

Лекция 8

3.3. Байесовский подход

Пусть $(\Omega, \mathcal{F}, P)$ — вероятностное пространство. $\Omega = \bigsqcup\limits_{n=1}^{\infty} D_n$, то есть $\{D_n\}$ — разбиение. Событие $A \in \mathcal{F}$.

Теорема Байеса:
$$P(D_n|A) = \frac{P(A|D_n)P(D_n)}{\sum\limits_{n=1}^{\infty} P(A|D_n)P(D_n)}.$$

Терминология:

1. A — результат эксперимента;
2. $P(D_n)$ — априорная вероятность D_n *(a priori)*;
3. $P(D_n|A)$ — апостериорная вероятность D_n *(a posteriori)*.

Теорема 13 (общий случай теоремы Байеса)**.** *Пусть ξ, η — случайные векторы. Тогда*
$$p_{\xi|\eta}(x|y) = \frac{p_{\eta|\xi}(y|x)p_\xi(x)}{\int p_{\eta|\xi}(y|x)p_\xi(x)dx}.$$

Математическое описание байесовского подхода к статистике

θ — случайный вектор, принимающий значения в $\Theta \subset \mathbb{R}^n$, имеющий распределение Q с плотностью $q(t)$.

- θ — параметр;
- t — значение параметра (реализация).

При $\theta = t$: $X = (X_1, \dots, X_n)$ — выборка из распределения $P \in \{P_t | t \in \Theta\}$, причем P_t имеет плотность $p_t(x)$. Плотность пары (X, θ) имеет вид:
$$f(x_1, \dots, x_n, t) = q(t)p_t(x_1) \cdot \ldots \cdot p_t(x_n).$$

Способ генерации выборки:

1. Выбрать значение θ из плотности $q(t)$;
2. Сгенерировать выборку X из распределения P_t, где t — выбранное значение параметра.

Замечание 11. Q — априорное распределение θ.

Способы оценки параметра

1. Апостериорное распределение, которое имеет плотность
$$q(t|x) = \frac{q(t) \cdot p_t(x_1) \cdot \ldots \cdot p_t(x_n)}{\int\limits_{\Theta} q(t) \cdot p_t(x_1) \cdot \ldots \cdot p_t(x_n) dt}.$$
2. Доверительный интервал $(u_{\frac{1-\alpha}{2}}, u_{\frac{1+\alpha}{2}})$, где u_p — p-квантиль апостериорного распределения.
3. Точечные оценки
   (a) $\hat{\theta}_1 = \mathbb{E}(\theta|X)$ — математическое ожидание апостериорного распределения;
   (b) $\hat{\theta}_2 = \arg\max\limits_{t \in \Theta} q(t|X)$ — мода апостериорного распределения;
   (c) $\hat{\theta}_3$ — медиана апостериорного распределения.

Пример 25. $X_1, \ldots, X_n \sim U[0, \theta+1]$, причем $\forall i X_i \leqslant 2$, $\theta \sim Bern(1/2)$. Найти апостериорное распределение θ.

Решение:
$$p_t(x_1, \ldots, x_n) = \frac{1}{(t+1)^n} I\{X_{(n)} \leqslant t+1\}$$
$$q(t) = \frac{1}{2} \text{ при } t \in \{0, 1\}$$
$$q(0|X) = \frac{1}{z} \cdot \frac{1}{2} I\{X_{(n)} \leqslant 1\},$$
z — знаменатель в формуле Байеса.
$$q(1|X) = \frac{1}{z} \cdot \frac{1}{2} \cdot \frac{1}{2^n} = \frac{1}{z} \cdot \frac{1}{2^{n+1}}.$$
$$z = \frac{1}{2} I\{X_{(n)} \leqslant 1\} + \frac{1}{2^{n+1}}.$$
Ответ: "$\theta|X$" $\sim Bern(q(1/X))$. $\square$

Теорема 14. *Пусть $q(t)$ интегрируема по Риману, $p_t(x)$ дифференцируема по t, $\sqrt{i(t)}$ интегрируем на любом конечном отрезке. Пусть $\hat{\theta} = \mathbb{E}(\theta|X)$, θ^* — ОМП для θ. Тогда*
$$\mathbb{E} n(\theta^* - \hat{\theta})^2 \to 0 \text{ и } \sqrt{n}(\theta^* - \hat{\theta}) \xrightarrow{P} 0.$$
(при большой выборке подходы почти эквивалентны).

Теорема 15. *Байесовская оценка $\hat{\theta}_1 = \mathbb{E}(\theta|X)$ — наилучшая в байесковском подходе с квадратичной функцией потерь (MSE). Аналогично $\hat{\theta}_3$ — медиана апостериорного распределения — наилучшая оценка в байесковском подходе с MAE.*

Доказательство: Теорема о наилучшем средневадратичном приближении, X — случайный вектор:
$$\mathop{\arg\min}_{\eta - X \text{ измерима}} \mathbb{E}(\xi - \eta)^2 = \mathbb{E}(\xi|X),$$
$$\int\limits_{\Theta} MSE_{\hat{\theta}}(t) q(t) dt = \int\limits_{\Theta} \int\limits_{\mathscr{X}} (\hat{\theta}(x) - t)^2 f(x,t) dt dx = \mathbb{E}(\hat{\theta} - \theta)^2 \to \min$$

По теореме о наилучшем средневадратичном приближении $\hat{\theta}(X) = \mathbb{E}(\theta|X)$. $\square$

3.4. Сопряженные распределения в байесовском подходе

Недостатки байесовского подхода:

1. Предполагается, что априорное распределение задано и не предлагается конструктивный способ по его выбору.

2. Требует больших вычислительных затрат.

Пример 26. $X_1, \dots, X_n \sim \mathcal{N}(\theta, 1)$. θ имеет априорное распределение Коши.

Вычислим знаменатель в формуле Байеса:

$$\int\limits_{-\infty}^{+\infty} \frac{1}{\pi(1+t^2)} \cdot \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\sum\limits_{i=1}^{n}(X_i - t)^2} dt \text{ — не берется.}$$

Определение 21. Пусть $X_1, \dots, X_n$ — выборка из неизвестного распределения $P \in \mathcal{P}$, где $\mathcal{P} = \{P_t | t \in \Theta\}$ — семейство распределений на $\mathscr{X}$. Пусть также на Θ задано семейство распределений $\mathcal{Q} = \{Q_\alpha | \alpha \in \mathcal{A}\}$. Семейство распределений $\mathcal{Q}$ называется *сопряженным к семейтву* $\mathcal{P}$, если взятии априорного распределения из $\mathcal{Q}$, соответствующее апостериорное распределение тоже лежит в $\mathcal{Q}$. Иными словами, если "$X|\theta = t \sim P_t$ и $\theta \sim Q_\alpha$, то "$\theta|X$" $\sim Q_{\alpha'}$.

Пример 27. $X_1, \dots, X_n \sim Exp(\theta)$ — подобрать сопряженное распределение и найти байесовскую оценку.

Решение: Плотность выборки $p_t(x) = t^n e^{-t\sum X_i}$ — зависит от выборки, в том числе от ее размера, и связана с t. Выпишем плотность по t пропорционально этому выражению, где вместо n и $\sum X_i$ подставим новые параметры из $\mathcal{A}$.

$$q(t) \propto t^{\beta-1} e^{-\alpha t} \text{ — это распределение } \Gamma(\alpha, \beta).$$

То есть $\mathcal{Q} = \{\Gamma(\alpha, \beta)\}$ — кандидат на сопряженное. Докажем, что $\mathcal{Q}$ — сопряженное к $\{Exp(\theta)\}$. Для этого найдем апостериорное распределение.

$$q(t|x) =\propto q(t) p_t(x) \propto t^{\beta-1} e^{-\alpha t} \cdot t^n \cdot e^{-t\sum X_i} = t^{\beta+n-1} e^{-t(\alpha + \sum X_i)}.$$

Это $\Gamma(\alpha + \sum X_i, \beta + n)$.

Ответ: "$\theta|X$" $\sim \Gamma(\alpha + \sum X_i, \beta + n)$, $\hat{\theta}_1 = \mathbb{E}(\theta|X) = \dfrac{\beta + n}{\alpha + \sum X_i}$. $\square$

3 | Глава 4. Непараметрический подход

4.1. Эмпирическое распределение

Пусть $X_1, \ldots, X_n$ — выборка из распределения P, рассматриваем $\mathcal{P} = \{$все распределения на $\mathscr{X}\}$.
Определение 22. Эмпирическим распределением, построенном по выборке, называется вероятностная мера $\hat{P}_n$, определенная по правилу:

$$\forall B \in \mathcal{B}_{\mathscr{X}} \ \ \hat{P}_n(B) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \in B\}.$$

Свойства:

1. $\hat{P}_n(B)$ — случайная величина, равная доле элементов выборки, попавших в B.
2. $\hat{P}_n$ — случайная дискретная вероятностная мера.
3. $n\hat{P}_n(B) \sim Bin(n, P(B))$, $\mathbb{E}(\hat{P}_n)(B) = P(B)$, $D\hat{P}_n(B) = \dfrac{P(B)(1-P(B))}{n}$.
4. УЗБЧ: $\hat{P}_n(B) \xrightarrow{P\text{ - п.н.}} P(B)$.

Рассмотрим случай $(\mathscr{X}, \mathscr{B}_{\mathscr{X}}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. В таком случае для $\hat{P}_n$ есть эмпирическая функция распределения.

$$\hat{F}_n(x) = \hat{P}_n((-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \leqslant x\}.$$

Утверждение 18. $\hat{F}_n \xrightarrow{P\text{ - }п.н.} F(X)$.
Теорема 16 (Гливенко-Кантелли)**.**

$$D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{P\text{ - }п.н.} 0.$$

Заметим, что

$$D_n = \sup_{B\in\mathscr{A}} |\hat{P}_n(B) - P(B)|, \text{ где } \mathscr{A} = \{(-\infty, x] | x \in \mathbb{R}\}.$$

Теорема 17 (Вапника-Червоненкиса)**.** $\sup\limits_{B\in\mathscr{A}} |\hat{P}_n(B) - P(B)| \xrightarrow{P\text{ - }п.н.} 0$ *тогда и только тогда, когда конечна размерность Вапника-Червоненкиса при разбиении* $\mathbb{R}^d$ *множествами из* $\mathscr{A}$.
Теорема 18 (Колмогорова-Смирнова)**.**

$$\sqrt{n}D_n = \sqrt{n}\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{d} \xi,$$

где ξ *имеет распределение Колмогорова:*

$$F_\xi(x) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2} I\{x \geqslant 0\}.$$

4.2. Метод подстановки

Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения P с функцией распределения F. Пусть $\theta = G(P)$ — функционал, значение которого нужно оценить. Тогда $\hat{\theta} = G(\hat{P}_n)$ — оценка θ по методу подстановки.

Пример 28.

1. $\theta = G(P) = \int\limits_{\mathscr{X}} f(x)dF(x) = \mathbb{E}_P f(X_1)$ — линейный функционал.

$$G(aP_1 + bP_2) = aG(P_1) + bG(P_2)$$

$$\hat{\theta} = G(\hat{P}_n) = \int\limits_{\mathscr{X}} f(x)d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \overline{f(X)}.$$

   Если $f(x) = x$, то $\theta = \mathbb{E}_P X_1$ и $\hat{\theta} = \overline{X}$.

2. $\theta = G(P) = D_P(X_1) = \int\limits_{\mathscr{X}} x^2 dF(x) - \left(\int\limits_{\mathscr{X}} xdF(x)\right)^2$

   $\hat{\theta} = \overline{X^2} - \overline{X}^2$.

3. $\theta = G(P) = \min\{x | F(x) \geqslant \alpha\}$ — α-квантиль.

   $\hat{\theta} = G(\hat{P}_n) = \min\{x | \hat{F}_n(x) \geqslant \alpha\} = X_{(\lceil n\alpha \rceil)}$ — выборочная квантиль.

Замечание 12. Метод моментов — частный случай метода подстановки. (Какой функционал $G(P)$ взять?)

4 Глава 5. Гипотезы и критерии

Лекция 10

Пусть S — критерий для проверки H_0 vs. H_1.

	H_0 верна	H_0 не верна
H_0 не отвергается	:)	Ошибка II рода: $P(II_S) = \sup\limits_{P\in\mathscr{P}_1} P(x \notin S)$
H_0 отвергается	Ошибка I рода: $P(I_S) = \sup\limits_{P\in\mathscr{P}_0} P(x \in S)$	:)

Минимизировать обе сразу не получится, поэтому решаем такую задачу:

$$\begin{cases} P(I_S) \leqslant \alpha \\ P(II_S) \to \min\limits_S \end{cases}$$

Определение 23. α — уровень значимости критерия S, то есть число $\alpha \in (0,1)$ называется уровнем значимости критерия S, если $P(I_S) \leqslant \alpha$.

Определение 24. Число $\alpha_0 = P(I_S)$ — реальный уровень значимости.

Первая магическая константа статистики $\alpha = 0.05$.

Как правило, альтернативная гипотеза сложная:

$$H_0 :\ \theta = \theta_0 \quad H_1 :\ \theta > \theta_0$$

$$H_0 :\ X_i \text{ имеет нормальное распределение} \quad H_1 :\ X_i \text{ имеет распр., отличающееся от норм.}$$

Определение 25. Для сравнения критериев определим *мощность критерия* S:

$$\beta_S(P) = P(X \in S), \text{ где } P \in \mathscr{P}_1$$

Пример 29. $X \sim Exp(\theta)$ — выборка из одного наблюдения. H_0: $\theta = \theta_0$ vs. H_1: $\theta > \theta_0$.

Решение: Заметим, что $\mathbb{E}_0 X = \dfrac{1}{\theta} \Rightarrow$ при больших θ стоит ожидать меньшее значение X. Логично взять критерий $S = \{x \in \mathscr{X} | x < c\}$, где c подберем из условия:

$$P(I_S) = P_{\theta_0}(X < c) = 1 - e^{-\theta_0 c} \leqslant \alpha \Rightarrow c \leqslant -\frac{1}{\theta_0}\ln(1-\alpha).$$

Мощность критерия:

$$\beta_S(\theta) = P_\theta(X < c) = 1 - e^{-\theta c} \to \max \text{ при } c \leqslant -\frac{1}{\theta_0}\ln(1-\alpha).$$

Следовательно, получаем критерий $S = \{x \in \mathscr{X} | x < -\dfrac{1}{\theta_0}\ln(1-\alpha)\}$. $\beta_S(\theta) = 1$ и $\alpha = 0.05 \Rightarrow \ln(1-\alpha) \approx -0.051$.

Критерий: $S = \{x \in \mathscr{X} | x < 0.051\}$.

Выводы:

1. $x < 0.051 \Rightarrow H_0$ отвергается. Результат статистически значим. "$x < 0.051$" — статистическое доказательство против H_0.

2. $x \geqslant 0.051 \Rightarrow H_0$ не отвергается. Результат статистически не значим.

□

5.2. Критерий Вальда

Определение 26. Критерий S называется *асимптотическим критерием уровня значимости* α, если

$$\lim_{n\to\infty} \sup P(I_S) \leqslant \alpha.$$

Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $P \in \{P_\theta | \theta \in \Theta\}$, $\Theta \subset \mathbb{R}$. $\hat{\theta}$ — асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$. $\hat{(\sigma)}$ — состоятельная оценка $\sigma(\theta)$. Рассмотрим гипотезы H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$ и статистику $W(X) = \sqrt{n}\dfrac{\hat{\theta}-\theta_0}{\hat{\sigma}}$.

При справедливости H_0 $W(X) \xrightarrow{d_{\theta_0}} \mathcal{N}(0,1)$.

$S = \{|W(X)| > z_{1-\alpha/2}\}$.

$$P(I_S) = P_{\theta_0}(|W| > z_{1-\alpha/2}) = P_{\theta_0}(W > z_{1-\alpha/2}) + P_{\theta_0}(W < -z_{1-\alpha/2}) \to$$

$$\to 1 - \Phi(z_{1-\alpha/2}) + \Phi(-z_{1-\alpha/2}) = 1 - (1-\alpha/2) + \alpha/2 = \alpha.$$

$$\beta_S(\theta) = P_\theta(|W| > z_{1-\alpha/2}) = P(W > z_{1-\alpha/2}) + P(W < -z_{1-\alpha/2}) =$$

$$= P_\theta\left(\underbrace{\sqrt{n}\frac{\hat{\theta}-\theta}{\hat{\sigma}}}_{\xrightarrow{d_\theta}\mathcal{N}(0,1)} > z_{1-\alpha/2} - \underbrace{\sqrt{n}\frac{\theta-\theta_0}{\hat{\sigma}}}_{w(\theta)}\right) + P_\theta\left(\underbrace{\sqrt{n}\frac{\hat{\theta}-\theta}{\hat{\sigma}}}_{\xrightarrow{d_\theta}\mathcal{N}(0,1)} < -z_{1-\alpha/2} - \underbrace{\sqrt{n}\frac{\theta-\theta_0}{\hat{\sigma}}}_{w(\theta)}\right) \approx$$

$$\approx 1 - \Phi(z_{1-\alpha/2} - w(\theta)) + \Phi(-z_{1-\alpha/2} - w(\theta)).$$

Заметим, что при $|w(\theta)| \to +\infty$: $\beta_S(\theta) \to 1$.

Вывод: мощность велика, если

1. выборка достаточно большая;

2. θ далека от θ_0.

Замечание 13.

1. Критерий Вальда можно получить для случая односторонней альтернативы:

   - H_0: $\theta = \theta_0$ vs. H_1: $\theta > \theta_0 \Rightarrow S_1 = \{W > z_{1-\alpha}\}$;

- H_0: $\theta = \theta_0$ vs. H_1: $\theta < \theta_0 \Rightarrow S_1 = \{W < z_\alpha\}$.

2. Если при односторонней альтернативе у H_0 поставить неравенство, ничего не изменится.
3. Рассмотрим H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$.

$$P_\theta\left(\sqrt{n}\frac{\hat\theta - \theta}{\hat\sigma} < z_{1-\alpha/2}\right) \to 1 - \alpha \Rightarrow c = \left(\hat\theta \pm \frac{z_{1-\alpha/2}\hat\sigma}{\sqrt{n}}\right).$$

H_0 отвергается $\Leftrightarrow \theta_0 \notin c$.

Пример 30. $X_1, \ldots, X_n \sim Cauchy(\theta)$. H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$.

Решение: $\hat\mu$ — а.н.о. θ с асимптотической дисперсией $\pi^2/4$.

$$W(X) = \sqrt{n}\frac{\hat\mu - \theta_0}{\pi/2} \xrightarrow{d_{\theta_0}} \mathcal{N}(0,1).$$

Критерий $\{|W(X)| > z_{1-\alpha/2}\}$.

$$z_{1-\alpha_2} = \text{sps.norm.ppf}(1-\alpha/2).$$

$$\beta_S(\theta) = \text{sps.norm.sf}(z_{1-\alpha_2} - w(\theta)) + \text{sps.norm.cdf}(-z_{1-\alpha_2} - w(\theta)).$$

□

5.3. Критерии отношения правдоподобия

Пусть $X = (X_1, \ldots, X_n)$ — выборка из неизвестного распределения $P \in \mathscr{P}$, где $\mathscr{P} = \{P_\theta | \theta \in \Theta\}$ — доминируемое семейство. $L_X(\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i)$ — функция правдоподобия.

Гипотезы: H_0: $\theta \in \Theta_0$ vs. H_1: $\theta \in \Theta_1$, $\hat\theta_j$ — ОМП на множестве Θ_j, $j \in \{0,1\}$. Статистика отношения правдоподобия:

$$\lambda(X) = 2\ln\left(\frac{L_X(\hat\theta_1)}{L_X(\hat\theta_0)}\right) = 2\ln\left(\frac{\sup\limits_{\theta\in\Theta_1} L_X(\theta)}{\sup\limits_{\theta\in\Theta_0} L_X(\theta)}\right)$$

Замечание 14. На практике $\Theta \subset \mathbb{R}^D$ и $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta \backslash \Theta_0$, $\dim \Theta_0 = d < D$, тогда $\hat\theta_1 = \hat\theta$ — глобальная ОМП на Θ.

$$\lambda(X) = 2\ln\frac{L_X(\hat\theta)}{L_X(\hat\theta_0)}.$$

Теорема 19. *Пусть* $\Theta_0 = \{\theta \in \Theta_0 | \theta_{d+1} = \theta^0_{d+1}, \ldots, \theta_D = \theta^0_D\}$. *Тогда при справедливости* H_0*:* $\theta \in \Theta_0$*:* $\lambda(X) \to \chi^2_{D-d}$.

Пример 31. H_0: $\theta_4 = \theta_5 = 0$. Тогда $\lambda(X) \to \chi^2_{5-3} = \chi^2_2$, $\Theta = \mathbb{R}^5$.

Критерий: $S = \{\lambda(X) > \chi^2_{D-d,1-\alpha}\}$, α — уровень значимости, $\chi^2_{k,p}$ — p-квантиль χ^2_k.

В некоторых случаях статистика $\lambda(X)$ позволяет построить неасимптотический критерий, в точности решающий заданную задачу

$$\begin{cases} P(I_S) \leqslant \alpha \\ \beta_S(P) \to \max\limits_S \forall P \in \mathscr{P}_1 \end{cases}.$$

(1) Простые гипотезы: H_0: $\theta = \theta_0$ **vs.** H_1: $\theta = \theta_1$

Рассмотрим статистику $\Lambda = \dfrac{L_X(\theta_1)}{L_X(\theta_0)}$.

Теорема 20 (лемма Неймара-Пирсона)**.** *Если существует C_α такая, что $P_{\theta_0}(\Lambda(X) > C_\alpha) = \alpha$, то $S = \{\Lambda(X) > C_\alpha\}$ — критерий уровня значимости α, который имеет максимальную мощность.*

(2) Сложные гипотезы

Определение 27. Критерий S уровня значимости α называется равномерно наиболее мощным критерием (РНМК), если для любого критерия R уровня значимости α: $\beta_S(P) \geqslant \beta_R(P)\ \forall P \in \mathscr{P}_1$.

Теорема 21 (о монотонном отношении правдоподобия)**.** *Пусть при $\theta_1 > \theta_2$ отношение правдоподобия представимо в виде $\dfrac{L_X(\theta_1)}{L_X(\theta_2)} = f_{\theta_1,\theta_2}(T(X))$, где $T(X)$ — статистика, $f_{\theta_1,\theta_2}(t)$ возрастает по t. Тогда критерий $S = \{T(X) > C_\alpha\}$ — РНМК уровня значимости α для H_0: $\theta = \theta_0$ vs. H_1: $\theta > \theta_0$, где C_α подберем из условия $P_\theta(T(X) > C_\alpha) = \alpha$.*

Замечание 15.

1. Пусть $\theta_1 > \theta_2 \Rightarrow \theta_1$ из альтернативы. $L_X(\theta_1)/L_X(\theta_2)$ возрастает при возрастании $T(X)$, следовательно, большие значение $T(X)$ более экстремальны.
2. В дискретном случае берем $\alpha_0 < \alpha$, такое что $P_{\theta_0}(T(X) > C_\alpha) = \alpha_0$.
3. Утверждение не изменится, если вместо H_0: $\theta = \theta_0$ поставить H_0: $\theta \leqslant \theta_0$.
4. H_0: $\theta = \theta_0$ vs. H_1: $\theta < \theta_0 \Rightarrow S = \{T(X) < C_\alpha\}$.
5. Если f_{θ_1,θ_2} убывает, то меняем знак в S.

Пример 32. $X_1, \dots, X_n \sim Exp(\theta)$, H_0: $\theta \leqslant \theta_0$ vs. H_1: $\theta > \theta_0$.

Решение: Рассчитываем отношение правдоподобия при $\theta_1 > \theta_2$:

$$\frac{L_X(\theta_1)}{L_X(\theta_2)} = \frac{\theta_1^n e^{-\theta_1 \sum X_i}}{\theta_2^n e^{-\theta_2 \sum X_i}} = \left(\frac{\theta_1}{\theta_2}\right) e^{(\theta_2 - \theta_1)\sum X_i},$$

то есть убывает по $T(X) = \sum X_i$. Тогда критерий $S = \{\sum X_i < C_\alpha\}$, где C_α подбираем из условия $P_{\theta_0}(\sum X_i < C_\alpha) = \alpha$. Заметим, что $\sum X_i \sim \Gamma(\theta, n) \Rightarrow C_\alpha$ — α-квантиль $\Gamma(\theta_0, n)$.

$$C_\alpha = \text{sps.gamma(a=n, scale=}1/\theta_0\text{).ppf}(\alpha),$$

$$\beta_S(\theta) = \text{sps.gamma(a=n, scale=}1/\theta_0\text{).cdf}(C_\alpha).$$

$\square$

5 Глава 7. Линейная регерессия

Лекция 13 (от 25.11)

7.2. Метод наименьших квадратов

Предполагается зависимость $y(x) = x^T\theta$, $\theta \in \mathbb{R}^d$.

Наблюдения: $Y = X\theta + \varepsilon$, где $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times d}$, $\theta \in \mathbb{R}^d$, $\varepsilon \in \mathbb{R}^n$. Y случаен, у X строки — объекты, столбцы — признаки, θ неизвстен, ε случаен и неизвестен. $RSS(\theta) = \sum\limits_{i=1}^{n}(y_i - x_i^T\theta)^2 = \|Y - X\theta\|^2$ — остаточная сумма квадратов. $\hat{\theta} = \underset{\theta\in\mathbb{R}^d}{\arg\min} RSS(\theta)$ — МНК-оценка.

Утверждение 19. *Если X^TX невырождена, то*

$$\hat{\theta} = (X^TX)^{-1}X^TY.$$

Доказательство:

$$RSS(\theta) = \|Y - X\theta\|^2 = (Y - X\theta)^T(Y - X\theta) =$$
$$= Y^TY - \underbrace{Y^TX\theta}_{=} - \underbrace{\theta^TX^TY}_{=} + \theta^TX^TX\theta$$

$$\frac{\partial RSS(\theta)}{\partial\theta} = -2X^TY + 2X^TX\theta = 0 \Rightarrow \hat{\theta} = (X^TX)^{-1}X^TY.$$

□

Обучение: $\hat{\theta} = (X^TX)^{-1}X^TY$.

Предсказание отклика на одном объекте x: $\hat{y}(x) = x^T\hat{\theta}$.

Теорема 22. *Свойства:*

1. $\mathbb{E}\varepsilon = 0 \Rightarrow \mathbb{E}\hat{\theta} = \theta$, $\mathbb{E}\hat{y}(x) = y(x)$.
2. $\mathbb{D}\varepsilon = \sigma^2 I_n$, $\mathbb{E}\varepsilon = 0 \Rightarrow \mathbb{D}\hat{\theta} = \sigma^2(X^TX)^{-1}$, $\mathbb{D}\hat{y}(x) = \sigma^2x^T(X^TX)^{-1}x$.

Доказательство:

1. $\mathbb{E}(\hat{\theta} = \mathbb{E}(X^TX)^{-1}X^TY) = (X^TX)^{-1}X^T\mathbb{E}(X\theta + \varepsilon) = (X^TX)^{-1}X^TX\theta = \theta$.
2. $\mathbb{D}\hat{\theta} = \mathbb{D}(X^TX)^{-1}X^TY = (X^TX)^{-1}X^T \cdot \mathbb{D}Y \cdot X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma(X^TX)^{-1}$.

□

Замечание 16. Часто на практике матрица $X^T X$ вырождена или близка к вырожденной, следовательно, $\mathbb{D}\hat{\theta}$ очень большая.

Пусть $\lambda_{\min}, \lambda_{\max}$ — минимальное и максимальное собственные числа матрицы $X^T X$.

$$CI = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

— индекс обусловленности.

$CI > 30$ — плохо.

Геометрический смысл МНК:

$L(X) = \{X\theta | \theta \in \mathbb{R}^d\}$ — пространство, порожденное столбцами матрицы $X \Rightarrow X\hat{\theta} = \mathrm{proj}_{L(X)} Y$.
Утверждение 20. *Оценка на σ:*

- $\hat{\varepsilon}_i = y_i - x_i^T \hat{\theta}$ — *остатки модели,*
- $\|\hat{\varepsilon}\| = RSS(\hat{\theta})$,
- $\hat{\sigma}^2 = \dfrac{RSS(\hat{\theta})}{n-d}$ — *несмещенная оценка* σ^2, *если* $\mathbb{E}\varepsilon = 0$, $\mathbb{D}\varepsilon = \sigma^2 I_n$.

Доказательство:

$$\mathbb{E}RSS(\hat{\theta}) = \sum_{i=1}^{n} \mathbb{E}(y_i - x_i^T\hat{\theta})^2 = /\mathbb{E}y_i = x_i^T\theta,\ \mathbb{E}x_i^T\hat{\theta} = x_i^T\theta/ = \sum_{i=1}^{n} \mathbb{D}(y_i - x_i^T\hat{\theta}) = \operatorname{Tr}\mathbb{D}(Y - X\hat{\theta})$$

.

$$\mathbb{D}(Y - X\hat{\theta}) = \mathbb{D}(Y - X(X^TX)^{-1}X^TY) = \mathbb{D}((I_n - \underbrace{X(X^TX)^{-1}X^T}_{A})Y) =$$

$$= (I_n - A)\cdot \mathbb{D}Y \cdot (I_n - A)^T = \sigma^2(I_n - 2A + AA^T) = \sigma^2(I_n - a),$$

так как $AA^T = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = A$.

$$\mathbb{E}RSS(\hat{\theta}) = \operatorname{Tr}(\sigma^2(I_n - A)) = \sigma^2(\operatorname{Tr} I_n - \operatorname{Tr} A) = \sigma^2(n - \operatorname{Tr}(X(X^TX)^{-1}X^T)) =$$

$$= \sigma^2(n - \operatorname{Tr}(X^TX(X^TX)^{-1})) = \sigma^2(n - \operatorname{Tr} I_d) = \sigma^2(n-d).$$

□

7.3. Гауссовская линейная модель

Предполагается модель $Y = X\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, \sigma I_n)$ — нормальность, несмещенность, гомоскедастичность.
Утверждение 21.

1. $\hat{\theta}$ *и* $Y - X\hat{\theta}$ *независимы.*
2. $\dfrac{1}{\sigma^2}\|X\hat{\theta} - X\theta\|^2 \sim \chi^2_d$, $\dfrac{1}{\sigma^2}\|Y - X\hat{\theta}\|^2 \sim \chi^2_{n-d}$.

Доказательство: $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n) \Rightarrow \hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(X^TX)^{-1})$ — потом и $Y \sim \mathcal{N}(X\theta, \sigma^2 I_n)$.

$L(X) = \{X\theta | \theta \in \mathbb{R}^d\}$. Разбиение $\mathbb{R}^n = L(X) \oplus L^{\perp}(X)$.

$\operatorname{proj}_{L^{\perp}(X)} Y = Y - X\hat{\theta}$.

По теореме о разложении гауссовского вектора $X\hat{\theta}$ и $Y - X\hat{\theta}$ независимы.

1. $\hat{\theta} = (X^TX)^{-1}X^TX\hat{\theta} = [(X^TX)^{-1}X^T] \cdot X\hat{\theta} \Rightarrow \hat{\theta}$ — линейная комбинация $X\hat{\theta} \Rightarrow \hat{\theta}$ независима с $Y - X\hat{\theta}$.

2.
$$\frac{1}{\sigma^2}\|X\hat{\theta} - \mathbb{E}X\hat{\theta}\|^2 = \frac{1}{\sigma^2}\|X\hat{\theta} - X\theta\|^2 \sim \chi^2_d,\ d = \dim L(X)$$
$$\frac{1}{\sigma^2}\|Y - X\hat{\theta} - \underbrace{\mathbb{E}(Y - X\hat{\theta})}_{=0}\|^2 = \frac{1}{\sigma^2}\|Y - X\hat{\theta}\|^2 \sim \chi^2_{n-d}.$$

$\square$

1. Доверительный интервал на σ

$\hat{\sigma}^2 = \dfrac{RSS(\hat{(\theta)})}{n-d} = \dfrac{\|Y - X\hat{\theta}\|^2}{n-d}$ — несмещенная оценка.

$\underbrace{\dfrac{\hat{\sigma}^2(n-d)}{\sigma^2}}_{\text{центральная функция}} \sim \chi^2_{n-d}$ по утверждению.

$P\left(\dfrac{\hat{\sigma}(n-d)}{\sigma^2} > \chi^2_{n-d,\alpha}\right) = 1 - \alpha.$

Интервал: $\left(0, \dfrac{\sigma^2(\hat{n} - d)}{\chi^2_{n-d,\alpha}}\right)$.

2. Доверительный интервал для θ_j и гипотезы H_0: $\theta_j = 0$

Утверждение 22.
$$\forall c \in \mathbb{R}^n \to T(X, Y) = \frac{c^T(\hat{\theta} - \theta)}{\hat{\sigma}\sqrt{c^T(X^TX)^{-1}c}} \sim T_{n-d}.$$

Доказательство:
$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(X^TX)^{-1}).$$
$$\frac{c^T(\hat{\theta} - \theta)}{\sigma\sqrt{c^T(X^TX)^{-1}c}} \sim \mathcal{N}(0, 1)$$

— зависит только от $\hat{\theta} \Rightarrow$ назвисима с $Y - X\hat{\theta}$, то есть и с $\hat{\sigma}^2$.

$$T(X, Y) = \frac{c^T(\hat{\theta} - \theta)}{\sigma\sqrt{c^T(X^TX)^{-1}c}} \frac{1}{\sqrt{\frac{\hat{\sigma}^2(n-d)}{\sigma^2}/(n-d)}} \sim T_{n-d}.$$

$\square$

Возьмем $c = (0, \dots, \underbrace{1}_{j} \dots, 0)^T$. Тогда

$$T_j(X,Y) = \frac{\hat\theta_j - \theta_j}{\sigma\sqrt{\widehat{(X^TX)^{-1}_{jj}}}} \sim T_{n-d}.$$

1. $P(|T_j(X,Y)| < T_{n-d,1-\alpha/2}) = 1-\alpha \Rightarrow (\hat\theta_j \pm \hat\sigma^2\sqrt{(X^TX)^{-1}_{jj}} \cdot T_{n-d,1-\alpha/2})$ — довиретильный интервал для θ_j.
2. H_0: $\theta_j = 0$ — гипотеза о незначимости коэффициента. При справедливости H_0:

$$T^0_j(X,Y) = \frac{\hat\theta_j}{\hat\sigma^2\sqrt{(X^TX)^{-1}_{jj}}} \sim T_{n-d}.$$

   Критерий: $S = \{|T^0_j(X,Y)| > T_{n-d,1-\alpha/2}\}$.

3. Доверительная область для θ

Определение 28. Пусть $\xi \sim \chi^2_{k_1}$, $\eta \sim \chi^2_{k_2}$ — независимы, тогда случайная величина $\zeta = \dfrac{\xi k_2}{\eta k_1}$ имеет *распределение Фишера с k_1, k_2 степенями свободы.* Обозначение F_{k_1,k_2}.

Используем утверждение из начала 7.3:

$$F(X,Y) = \frac{\frac{1}{\sigma^2}\|X\hat\theta - X\theta\|^2}{\frac{1}{\sigma^2}\|Y - X\theta\|^2} \cdot \frac{n-d}{d} \sim F_{d,n-d}.$$

Доверительная область: $\{\theta \in \mathbb{R}^d | F(X,Y) \leqslant F_{d,n-d,1-\alpha}\}$.

4. Общий случай линейных гипотез

Линейная гипотеза: $H_0 : T\theta = \tau$, где $T \in \mathbb{R}^{k\times d}$, $\tau \in \mathbb{R}^k$, $k \leqslant d$, $\operatorname{rg} T = k$.

Пример 33. Пусть $H_0 : \begin{cases} \theta_1 = 0 \\ \theta_2 = \theta_3 \end{cases}$, $T = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & \dots \end{pmatrix}$, $\tau = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\hat\theta \sim \mathcal{N}(\theta, \sigma^2(X^TX)^{-1})$. Обозначим $\hat t = T\hat\theta \sim \mathcal{N}(\underbrace{T\theta}_{=\tau \text{ при } H_0}, \sigma^2 \underbrace{T(X^TX)^{-1}T^T}_{=B})$.

Тогда при справедливости H_0:

$$\frac{1}{\sigma}B^{-1/2}(\hat t - \tau) \sim \mathcal{N}(0, I_k).$$

Возьмем скалярный квадрат:

$$\frac{1}{\sigma^2}(\hat t - \tau)^T B^{-1}(\hat t - \tau) \overset{H_0}{\sim} \chi^2_k \Rightarrow$$

$$\text{/по утверждению из начала 7.3: } \frac{1}{\sigma^2}\|Y - X\hat\theta\|^2 \sim \chi^2_{n-d}/$$

$$\Rightarrow \text{зависит только от } \hat{\theta} \text{ и не зависит от } Y - X\hat{\theta}.$$

$$F(X,Y) = \frac{(\hat{t}-\tau)^T B^{-1} (\hat{t}-\tau)}{\|Y - X\hat{\theta}\|^2} \cdot \frac{n-d}{l} \overset{H_0}{\sim} F_{k,n-d}.$$

Критерий $S = \{F(X,Y) > F_{k,n-d,1-\alpha}\}$ — F-критерий.

6 Глава 8. Теория наилучших оценок

Лекция 14 (от 2.12)

8.1. Информация и расстояния

1. Вклад и информация Фишера

Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P \in \mathcal{P} = \{P_\theta | \ \theta \in \Theta\}$, $\mathcal{P}$ — доминируемое семейство распределений с плотностью $p_\theta(x)$.

- $L_X(\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i)$ — функция правдоподобия.
- $l_X(\theta) = \sum\limits_{i=11}^{n} \ln p_\theta(X_i)$ — логарифмическая функция правдоподобия.

Определение 29. $u_X(\theta) = \dfrac{\partial}{\partial\theta} l_X(\theta)$ *— вклад выборки X в параметр θ.*

Определение 30. $I_X(\theta) = D_\theta u_X(\theta)$ *— информация Фишера, содержащаяся в выборке X о параметре θ.*

Пример 34. $X_1, \dots, X_n \sim Bern(\theta)$

$L_X(\theta) = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i}$.

$l_x(\theta) = \sum X_i \cdot \ln\theta + (n - \sum X_i)\ln(1-\theta)$.

$u_X(\theta) = \dfrac{\partial}{\partial\theta} l_X(\theta) = \dfrac{\sum X_i}{\theta} - \dfrac{n - \sum X_i}{1-\theta} = \dfrac{(1-\theta)\sum X_i - \theta(n - \sum X_i)}{\theta(1-\theta)} = \dfrac{\sum X_i - n\theta}{\theta(1-\theta)}$.

$I_X(\theta) = D_\theta u_X(\theta) = \dfrac{1}{\theta^2(1-\theta)^2} D_\theta \sum X_i = \dfrac{n\theta(1-\theta)}{\theta^2(1-\theta)^2} = \dfrac{n}{\theta(1-\theta)}$.

Утверждение 23. *В условиях E1-E4 (см. условия регулярности)*

1. $E_\theta u_X(\theta) = 0$;
2. $I_X(\theta) = E_\theta u_X^2(\theta)$;
3. $I_X(\theta) = ni(\theta)$, *где* $i(\theta) = I_{X_1}(\theta)$ *(информация одного наблюдения);*
4. $I_X(\theta) = -E_\theta \dfrac{\partial^2 l_X(\theta)}{\partial\theta^2}$.

Доказательство:

1. $u_X(\theta) = \dfrac{\partial}{\partial\theta} l_X(\theta) = \dfrac{\partial}{\partial\theta} \sum\limits_{i=1}^{n} \ln p_\theta(X_i) = \sum\limits_{i=1}^{n} \dfrac{\partial \ln p_\theta(X_i)}{\partial\theta} = \sum\limits_{i=1}^{n} u_{X_i}(\theta)$

Посчитаем матожидание: $E_\theta u_{X_1}(\theta) = E_\theta \dfrac{\partial \ln p_\theta(X_1)}{\partial\theta} = \int\limits_{\mathscr{X}} \dfrac{\partial \ln p_\theta(x)}{\partial \theta} p_\theta(x)dx = \int\limits_{\mathscr{X}} \dfrac{\frac{\partial p_\theta(x)}{\partial\theta}}{p_\theta(x)} p_\theta(x)dx = \int\limits_{\mathscr{X}} \dfrac{\partial p_\theta(x)}{\partial\theta}dx \overset{E3}{=} \dfrac{\partial}{\partial\theta}\int\limits_{\mathscr{X}} p_\theta(x)dx = \dfrac{\partial(1)}{\partial\theta} = 0.$

2. очевидным образом следует из п. 1.
3. $I_X(\theta) = D_\theta u_X(\theta) = D_\theta \sum u_{X_i}(\theta) \overset{\text{н.о.р.с.в}}{=} \sum D_\theta u_{X_i}(\theta) = ni(\theta).$
4. $\dfrac{\partial^2 \ln p_\theta(x)}{\partial\theta^2} = \dfrac{\partial}{\partial\theta}\left(\dfrac{\frac{\partial p_\theta(x)}{\partial\theta}}{p_\theta(x)}\right) = \dfrac{\frac{\partial^2 p_\theta(x)}{\partial\theta^2}}{p_\theta(x)} - \dfrac{(\frac{\partial p_\theta(x)}{\partial\theta})^2}{p_\theta^2(x)}$

$$E_\theta \frac{\partial^2 \ln p_\theta(X)}{\partial\theta^2} = E_\theta \frac{\frac{\partial^2 p_\theta(X)}{\partial\theta^2}}{p_\theta(X)} - \underbrace{E_\theta\left(\frac{\frac{\partial p_\theta(X)}{\partial\theta}}{p_\theta(X)}\right)^2}_{=I_X(\theta)}.$$

Покажем, что первое слагаемое равно нулю:

$$E_\theta \frac{\frac{\partial^2 p_\theta(X)}{\partial\theta^2}}{p_\theta(X)} = \int\limits_{\mathscr{X}} \frac{\frac{\partial^2 p_\theta(x)}{\partial\theta^2}}{p_\theta(x)} p_\theta(x)dx = \int\limits_{\mathscr{X}} \frac{\partial^2 p_\theta(x)}{\partial\theta^2}dx = \frac{\partial^2}{\partial\theta^2}\int\limits_{\mathscr{X}} p_\theta(x)dx = 0$$

□

2. Энтропия в дискретном случае

Пусть P — распределение на $\{a_1, \dots a_k\}$ с вероятностями $p_1, \dots, p_k$.

Определение 31. $H(P) = -\sum\limits_{j=1}^{k} p_j \log p_j$ — *энтропия* (считаем, что $0 \cdot \log 0 = 0$).

Свойства:

1. $H(P) \geqslant 0,\ H(P) = 0 \iff \exists j:\ p_j = 1;$
2. $H(P) \leqslant \log k,\ H(P) = \log k \iff \forall j\ p_j = 1/k.$

Доказательство:

1. $p_j \in [0,1] \implies -\log p_j \geqslant 0.$
2. $H(P) = -E\log p(\xi)$, где $\xi \sim P$.

$$H(P) = E\log\frac{1}{P(\xi)} \leqslant |\text{ неравенство Йенсена }| \leqslant \log E\frac{1}{p(\xi)} = \log\sum_{j=1}^{k}\frac{1}{p_j}p_j = \log k.$$

□

3. Общий случай

Пусть P, Q — распределения по одной и той же мере (либо оба дискретные, либо оба абсолютно непрерывные) с плотностями $p(x)$ и $q(x)$ соответственно.

Определение 32. 1. $H(P) = -E \log p(\xi)$, где $\xi \sim P$ — *энтропия*;

2. $H(P, Q) = -E \log q(\xi)$, где $\xi \sim P$ — *кросс-энтропия*;

3. $KL(P, Q) = E \log \dfrac{p(\xi)}{q(\xi)}$, где $\xi \sim P$ — *дивергенция Кульбака-Лейблера.*

Замечание 17. В общем случае $H(P)$ может быть отрицательной:

$P = U[0, 1/2],\ p(x) = 2I\{x \in [0, 1/2]\}$. Тогда $H(P) = -E \log(p(\xi)) = -E \log 2 = -\log 2$.

Свойства KL:

1. $KL(P, Q) \geqslant 0;\ KL(P, Q) = 0 \iff P \overset{\text{п.в.}}{=} Q$

   Доказательство: $-KL(P, Q) = E \log \dfrac{q(\xi)}{p(\xi)} \leqslant |$ неравенство Йенсена $| \leqslant \log E \dfrac{q(\xi)}{p(\xi)} = \log \int\limits_{\mathscr{X}} \dfrac{q(x)}{p(x)} p(x) dx = \log \int\limits_{\mathscr{X}} q(x) dx = \log 1 = 0.$

   $-KL(P, Q) \leqslant 0 \iff KL(P, Q) \geqslant 0$, причем равенство в неравенстве Йенсена достигается тогда и только тогда, когда $P \overset{\text{п.в.}}{=} Q$. $\square$

2. $KL(P, Q) \neq KL(Q, P)$

3. Пусть $X = (X_1, \ldots, X_n)$ — выборка из дискретного распределения $P \in \{P_\theta |\ \theta \in \Theta\}$. Тогда

$$KL(\hat{P}_n, P_\theta) = E_{\hat{P}_n} \log \frac{\hat{P}_n(X_i)}{P_\theta(X_i)} = \frac{1}{n} \sum_{i=1}^n \log \frac{1/n}{p_\theta(X_i)} = \underbrace{-\frac{1}{n} \sum_{i=1}^n \log p_\theta(X_i)}_{H(\hat{P}_n, P_\theta)} - \underbrace{\log n}_{H(\hat{P}_n)}.$$

   $KL(\hat{P}_n, P_\theta) \to \min\limits_\theta \iff H(\hat{P}_n, P_\theta) \to \min\limits_\theta \iff l_X(\theta) \to \max\limits_\theta$, т.е. **ОМП**.

8.2. Свойства ОМП

Теорема 23 (Экстремальное свойство правдоподобия (L1-L3))**.**

$$\forall \theta_0, \theta_1 \in \Theta : \theta_0 \neq \theta_1\ P_{\theta_0}(L_X(\theta_0) > L_X(\theta_1)) \xrightarrow{n \to \infty} 1.$$

Доказательство: $L_X(\theta_0) > L_X(\theta_1) \iff \dfrac{1}{n} \log \dfrac{L_X(\theta_0)}{L_X(\theta_1)} > 0$

$\dfrac{1}{n} \log \dfrac{L_X(\theta_0)}{L_X(\theta_1)} = \dfrac{1}{n} \sum\limits_{i=1}^n \log \dfrac{p_{\theta_0}(X_i)}{p_{\theta_1}(X_i)} \overset{P_{\theta_0}\text{-п.н (УЗБЧ)}}{\to} E_{\theta_0} \log \dfrac{p_{\theta_0}(X_1)}{p_{\theta_1}(X_1)} = KL(P_{\theta_0}, P_{\theta_1}) > 0$, т.к. $\theta_0 \neq \theta_1$ и выполнены условия L1-L2. $\square$

Теорема 24 (Состоятельность ОМП (L1-L5))**.** *С вероятностью* $\to 1$ *уравнение правдоподобия* $\dfrac{\partial l_X(\theta)}{\partial\theta} = 0$ *имеет решение* $\tilde\theta$*, причем* $\tilde\theta$ *— состоятельная оценка* θ*.*

Доказательство: Пусть θ_0 — истинное значение. Тогда по свойству L4 $\exists\varepsilon > 0 : (\theta_0-\varepsilon,\ \theta_0+\varepsilon) \subset \Theta$. Из экстремального свойства правдоподобия получим, что

$$P_{\theta_0}(L_X(\theta_0) > L_X(\theta_0+\varepsilon), L_X(\theta_0) > L_X(\theta_0-\varepsilon)) \to 1. \tag{6.1}$$

Тогда из (6.1) и условия L5 следует, что на $(\theta_0-\varepsilon,\ \theta_0+\varepsilon)$ имеется корень уравнения правдоподобия. Пусть $\tilde\theta$ — ближайший к θ_0 корень. Из (6.1) следует, что $P_{\theta_0}(|\tilde\theta-\theta_0| > \varepsilon) \to 0$. В силу произвольности ε $\tilde\theta$ — состоятельная оценка θ. □

Следствие 5. *Если* $\forall n\ \forall X_1, \ldots, X_n$ *есть ровно одно решение уравнения правдоподобия* $\tilde\theta$*, то* $\tilde\theta$ *— состоятельная оценка* θ *и* $P_{\theta_0}(\tilde\theta = \hat\theta_{\text{ОМП}}) \to 1$ *и тогда ОМП также состоятельна.*

Теорема 25 (Асимптотическая нормальность ОМП (L1-L9), б/д)**.**

1. *Пусть* $\tilde\theta$ *— решение уравнения правдоподобия, т.ч.* $\tilde\theta$ *— состоятельная оценка* θ*. Тогда* $\tilde\theta$ *— а.н.о.* θ *с асимптотической дисперсией* $\frac{1}{i(\theta)}$*.*
2. *Пусть* $\hat\theta$ *— произвольная а.н.о. с асимптотической дисперсией* $\sigma^2(\theta)$*, т. ч.* $\sigma(\theta)$ *непрерывна. Тогда* $\sigma^2(\theta) \geqslant \dfrac{1}{i(\theta)}$*.*

Следствие 6.

1. *Если* $\forall n\ \forall X_1, \ldots, X_n$ *есть ровно один корень, то он является а.н.о.*
2. *ОМП асимптотически эффективная оценка (т.е. наилучшая среди всех а.н.о с непрерывной асимптотической дисперсией).*

Замечание 18. Если L* не выполнено, то может быть еще круче!

$X_1, \ldots, X_n \sim U[0,\theta]$; $\hat\theta = X_{(n)}$ — ОМП. Тогда

$$n(\theta - X_{(n)}) \xrightarrow{d_\theta} Exp(1),$$

т.е. скорость сходимости $\sim 1/n$.

8.3. Эффективные оценки

Пусть $X_1, \ldots, X_n$ — выборка из доминируемого семейства $P \in \{P_\theta |\ \theta \in \Theta\}$ с плотностью $p_\theta(x)$ и $\Theta \subset \mathbb{R}$. Рассмотрим семейство $\mathcal{K} = \{$все несмещенные оценки $\tau(\theta)\}$. **Задача:** Найти наилучшую в с/к подходе оценку, т.е. нужно минимизировать $MSE_{\hat\theta}(\theta) = D_\theta\hat\theta$ по всем θ сразу (такие оценки называются *оптимальными* в $\mathcal{K}$).

Теорема 26 (Неравенство Рао-Крамера (E1-E4))**.** *Для любой оценки из* $\mathcal{K}$

$$D_\theta\hat\theta \geqslant \frac{(\tau'(\theta))^2}{I_X(\theta)} \forall\theta \in \Theta.$$

Доказательство: $\tau'(\theta) = \dfrac{\partial}{\partial\theta}\mathbb{E}_\theta\hat\theta = \dfrac{\partial}{\partial\theta}\displaystyle\int_{\mathscr{X}} \hat\theta(x)p_\theta(x)dx = \int_{\mathscr{X}} \hat\theta(x)\frac{\partial p_\theta(x)}{\partial\theta}dx = \int_{\mathscr{X}} \hat\theta(x)\frac{\partial \ln p_\theta(x)}{\partial\theta}p_\theta(x)dx =$ $\displaystyle\int_{\mathscr{X}} \hat\theta(x)u_x(\theta)p_\theta(x)dx = \mathbb{E}_\theta\hat\theta u_X(\theta).$

$\mathbb{E}_\theta u_X(\theta) = 0 \Rightarrow \tau(\theta) = \mathbb{E}_\theta(\hat\theta - \tau(\theta))u_X(\theta)$.

Применим неравенство Коши-Буняковского: $(\tau'(\theta))^2 \leqslant \underbrace{\mathbb{E}_\theta(\hat\theta - \tau(\theta))^2}_{\mathbb{D}_\theta\hat\theta}\underbrace{\mathbb{E}_\theta u_X^2(\theta)}_{I_X(\theta)} \implies D_\theta\hat\theta \geqslant \dfrac{(\tau'(\theta))^2}{I_X(\theta)}$. □

Теорема 27. *(Критерий эффективности)* $\hat\theta$ *— эффективная оценка* $\tau(\theta) \iff \hat\theta$ *— линейная функция от вклада, т.е.*

$$\hat\theta - \tau(\theta) = c(\theta)u_X(\theta),$$

где $c(\theta) = \dfrac{\tau'(\theta)}{I_X(\theta)}$ *— линейная по* X *функция при фиксированном* θ.

Доказательство: Равенство в неравенстве Коши-Буняковского достигается, когда величины линейно зависимы, т.е.

$$\hat\theta - \tau(\theta) = c(\theta)u_X(\theta) + a(\theta).$$

1. $\underbrace{\mathbb{E}_\theta(\hat\theta - \tau(\theta))}_{0} = \underbrace{\mathbb{E}_\theta c(\theta)u_X(\theta)}_{0} + \underbrace{\mathbb{E}_\theta a(\theta)}_{a(\theta)} \implies a(\theta) \equiv 0$

2. Домножим на $u_X(\theta)$ и возьмем матожидание:

$$E_\theta(\hat\theta - \tau(\theta))u_X(\theta) = c(\theta)\mathbb{E}_\theta u_X^2(\theta) = c(\theta)I_X(\theta) = \tau'(\theta) \implies c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}.$$

□

Лекция 15

8.4. Оптимальные оценки

Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $P \in \mathscr{P} = \{P_\theta | \theta \in \Theta\}$.

$\mathcal{K} = \{$все несмещенные оценки параметра $\theta\}$.
Определение 33. Оценка $\hat\theta \in \mathcal{K}$, которая для всех $\theta \in \Theta$ дает минимум величины

$$MSE_{\hat\theta}(\theta) = \mathbb{E}(\hat\theta - \theta)^2 = \mathbb{D}_\theta\hat\theta$$

называется *оптимальной*.
Теорема 28 (Колмогорова-Блекуэлла-Рао). *Пусть* $\hat\theta$ *— несмещенная оценка* $\tau(\theta)$*, причем* $\mathbb{E}_\theta\hat\theta^2 < +\infty$*;* $S(X)$ *— достаточная статистика. Тогда*

1. $\theta^* = \mathbb{E}_\theta(\hat\theta | S(X))$ *тоже является несмещенной оценкой* $\tau(\theta)$.
2. $\mathbb{D}_\theta\theta^* \leqslant \mathbb{D}_\theta\hat\theta \ \forall\theta \in \Theta$

   Равенство возможно $\Leftrightarrow \theta^* = \hat\theta$ *-* P_θ*-п.н.* $\forall\theta \in \Theta$*, то есть* $\hat\theta$ *изначально является* $S(X)$*-измеримой.*

Доказательство:

1. $S(X)$ — достаточная, следовательно, $P_\theta(X \in B | S(X))$ не зависит от θ, значит $\mathbb{E}_\theta(\hat{\theta}|S(X))$ тоже не зависит от θ (как матожидание условного распределения), поэтому θ^* — действительно оценка.

$$\mathbb{E}_\theta(\mathbb{E}_\theta)(\hat{\theta}|S(X)) = \mathbb{E}_\theta\hat{\theta} = \tau(\theta) \Rightarrow \theta^* \text{ — несмещенная оценка } \tau(\theta)$$

.

2. (для $\tau(\theta) \in \mathbb{R}$):

$$\mathbb{D}_\theta\hat{\theta} = \mathbb{E}_\theta(\hat{\theta}-\tau(\theta))^2 = \mathbb{E}_\theta(\hat{\theta}-\theta^*+\theta^*-\tau(\theta))^2 = \underbrace{\mathbb{E}_\theta(\hat{\theta}-\theta^*)^2}_{\geqslant 0} + \mathbb{D}_\theta\theta^* + 2\underbrace{\mathbb{E}_\theta(\hat{\theta}-\theta^*)(\theta^*-\tau(\theta))}_{=0} \geqslant \mathbb{D}_\theta\theta^*.$$

$$\mathbb{E}_\theta(\hat{\theta}-\theta^*)(\theta^*-\tau(\theta)) = E_\theta(\mathbb{E}_\theta((\hat{\theta}-\theta^*)(\theta^*-\tau(\theta))|S(X))) =$$
$$= E_\theta((\theta^*-\tau(\theta))\mathbb{E}_\theta(\hat{\theta}-\theta^*|S(X))) = \mathbb{E}_\theta((\theta^*-\tau(\theta))\cdot 0) = 0.$$

Равенство возможно $\Leftrightarrow \mathbb{E}_\theta(\hat{\theta}-\theta^*)^2 = 0 \forall\theta \in \Theta \Leftrightarrow \hat{\theta} = \theta^*$ P_θ-п.н. $\forall\theta \in \Theta \Leftrightarrow$

$\Leftrightarrow \hat{\theta} = \mathbb{E}_\theta(\hat{\theta}|S(X))$ P_θ-п.н. $\forall\theta \in \Theta \Leftrightarrow \hat{\theta}$ является $S(X)$-измеримой.

□

Следствие 7.

1. *θ^* не хуже $\hat{\theta}$ в среднеквадратичном подходе;*
2. *Если $\hat{\theta}$ не является $S(X)$-измеримой, то θ^* лучше в средневадратичном подходе;*
3. *Если θ^* —* ***единственная*** *несмещенная $S(X)$-измеримая оценка $\tau(\theta)$, то она и является оптимальной.*

Доказательство: Если есть не $S(X)$ измеримая оценка, то возьмем УМО, получим лучше и $S(X)$-измеримую и несмещенную, а она одна. Противоречие. □

Единственность гарантирует свойство полноты.
Определение 34. Статистика $S(X)$ называется *полной*, если для семейства распределений $\{P_\theta|\theta \in \Theta\}$, если выполнение свойства $\forall\theta \in \Theta$ $\mathbb{E}_\theta f(S(X)) = 0$ возможно только в случае $\forall\theta \in \Theta$ $f(S(X)) \overset{P_\theta\text{-п.н.}}{=} 0$.
Смысл 6. несмещенной $S(X)$-измеримой оценкой нуля может быть только ноль.
Теорема 29 (об оптимальной оценке). *Пусть $S(X)$ — полная и достаточная статистика для $\{P_\theta|\theta \in \Theta\}$. Оценка $\theta^* = \varphi(S(X))$ — несмещенная $S(X)$-измеримая оценка $\tau(\theta)$. Тогда θ^* — оптимальная оценка $\tau(\theta)$.*

Доказательство: Согласно предыдущему следствию достаточно проверить, что θ^* — единственная несмещенная $S(X)$-измеримая оценка $\tau(\theta)$.

Пусть $\psi(S(X))$ — тоже несмещенная оценка $\tau(\theta)$. Обозначим $f(x) = \varphi(x) - \psi(x)$. Тогда

$$\mathbb{E}_\theta f(S(X)) = \mathbb{E}_\theta\varphi(S(X)) - \mathbb{E}_\theta\psi(S(X)) = 0 \ \ \forall\theta \in \Theta.$$

Но $S(X)$ — полная, следовательно, P_θ-п.н. $\forall\theta \in \Theta$ $f(S(X)) = 0 = \varphi(S(X)) - \psi(S(X))$. □

Следствие 8. $S(X)$ *— полная и достаточная статистика для* $\{P_\theta|\theta \in \Theta\}$.

1. *Если θ^* — несмещенная оценка $\tau(\theta)$, то $\mathbb{E}_\theta(\theta^*|S(X))$ — оптимальная оценка $\tau(\theta)$.*

2. *Если θ_1^*, θ_2^* — оптимальные оценки $\tau_1(\theta)$, $\tau_2(\theta)$, то $a\theta_1^* + b\theta_2^*$ — оптимальная оценка $a\tau_1(\theta) + b\tau_2(\theta)$.*

3. *Если $\tau(\theta) = (\tau_1(\theta), \ldots, \tau_k(\theta)) \in \mathbb{R}^k$ и θ_j^* — оптимальная оценка $\tau_j(\theta)$, то $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ — оптимальная оценка вектора $\tau(\theta)$.*

Алгоритм поиска оптимальных оценок

1. Найти $S(X)$ — полную и достаточную статистику в данной модели;

2. Решить уравнение несмещенности $\mathbb{E}_\theta\varphi(S(X)) = \tau(\theta)$ относительно φ. Оценка $\theta^* = \varphi(S(X))$ будет оптимальной согласно теореме об оптимальной оценке.

Оптимальные оценки в экспоненциальном семействе

Пусть $\mathscr{P} = \{P_\theta|\theta \in \Theta\}$, причем $p_\theta(x) = \dfrac{g(x)}{h(\theta)}e^{a(\theta)^T u(x)}$.

Теорема 30. *Если множетсво Θ телесно (то есть содержит все внутренние точки), а функция $a(\theta)$ непрерывна и содержит линейно независимые компоненты, то статистика $S(X) = \sum\limits_{i=1}^{n} u(X_i)$ является полной и достаточной для семейства $\mathscr{P}$.*

Оптимальные оценки в гауссовской линейной модели

Гауссовская линейная модель $Y = X\theta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

$L(X) = \{X\theta|\theta \in \mathbb{R}^d\}$.

Утверждение 24. *$S(Y) = (\operatorname{proj}_{L(X)} Y, \|\operatorname{proj}_{L^T(X)} Y\|^2)$ — достаточная статистика.*

Доказательство: Запишем плотность $Y \sim \mathcal{N}(X\theta, \sigma^2 I_n)$, $c = (2\pi\sigma^2)^{-n/2}$:

$$p(y) = c \cdot \exp\left(-\frac{1}{2\sigma^2}\sum(Y_i - x_i^T\theta)^2\right) = c \cdot \exp\left(-\frac{1}{2\sigma^2}\|Y - X\theta\|^2\right) =$$

$$= c \cdot \exp\left(-\frac{1}{2\sigma^2}\left(\|\operatorname{proj}_{L(X)}(Y - X\theta)\|^2 + \|\operatorname{proj}_{L^T(X)}(Y - X\theta)\|^2\right)\right) =$$

$$= c \cdot \exp\left(-\frac{1}{2\sigma^2}\left(\|\operatorname{proj}_{L(X)} Y - X\theta\|^2 + \|\operatorname{proj}_{L^T(X)} Y\|^2\right)\right).$$

□

Утверждение 25.

1. *$S(Y)$ — полная статистика (б/д);*

2. *$\hat\theta = (X^T X)^{-1} X^T Y$ — оптимальная оценка θ;*

3. *$\hat\sigma^2 = \dfrac{1}{n-d}\|Y - X\hat\theta\|^2$ — оптимальная оценка σ^2.*

Доказательство: Обе несмещенные и являются функциями от $S(Y)$. $\square$

Утверждение 26. *Если не предполагать нормальность ошибки, то* $\hat{\theta}$ *— наилучшая в среднеквадратичном подходе среди всех несмещенных оценок, линейных по* Y.

7 | Глава 9. Доказательства теорем

9.1. Теорема Гливенко-Кантелли

$X = (X_1, X_2, \ldots)$ — выборка из распределения P с функцией распределения F. Тогда

$$D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{P\text{ - п.н.}} 0.$$

Доказательство: Замечание: $D_n = \max(2n \text{ точек}) \Rightarrow D_n$ — случайная величина. Обозначим u_p — p-квантиль распределения P. Выберем $N \in \mathbb{N}$, $k \in \{1, \ldots, N-1\}$. Пусть $x \in [u_{\frac{k}{N}}, u_{\frac{k+1}{N}})$. Тогда

$$\hat{F}_n(x) - F(x) \leqslant \hat{F}_n(u_{\frac{k+1}{N}} - 0) - F(u_{\frac{k}{N}}) = \hat{F}_n(u_{\frac{k+1}{N}} - 0) - F(u_{\frac{k+1}{N}} - 0) + \underbrace{F(u_{\frac{k+1}{N}} - 0)}_{\leqslant \frac{k+1}{N}} - \underbrace{F(u_{\frac{k}{N}})}_{\geqslant k/N} \leqslant$$

$$\leqslant \hat{F}_n(u_{\frac{k+1}{N}} - 0) - F(u_{\frac{k+1}{N}} - 0) + \frac{1}{N}.$$

Аналогично, $\hat{F}_n(x) - F(x) \geqslant \hat{F}_n(u_{k/N}) - F(u_{k/N}) - \dfrac{1}{N}$.

Пусть x произвольный

$$|\hat{F}_n(x) - F(x)| \leqslant \max_{k\in\{1,\ldots,N-1\}} \left\{ \hat{F}_n\left(u_{\frac{k+1}{N}} - 0\right) - F\left(u_{\frac{k+1}{N}} - 0\right),\ \hat{F}_n\left(u_{\frac{k}{N}}\right) - F\left(u_{\frac{k}{N}}\right) \right\} + \frac{1}{N}.$$

Правая часть не зависит от x, следовательно, слева ставим sup.

УЗБЧ: $\hat{F}_n\left(u_{\frac{k+1}{N}} - 0\right) \xrightarrow{P\text{ - п.н.}} F\left(u_{\frac{k+1}{N}} - 0\right)$; $\hat{F}_n\left(u_{\frac{k}{N}}\right) \xrightarrow{P\text{ - п.н.}} F\left(u_{\frac{k}{N}}\right)$.

По теореме о наследовании сходимостей

$$\lim_{n\to\infty} \sup \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \leqslant \frac{1}{N},$$

В силу произвольности N существует $\lim\limits_{n\to\infty} \sup\limits_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|$ P - п.н. $\square$

9.2. Лемма Неймана-Пирсона

Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения P, $H_0 : P = P_0$ vs. H_1: $P = P_1$, p_0, p_1 — плотности. Если

$$\exists C_\alpha : P_0\left(\frac{p_1(X)}{p_0(X)} \geqslant C_\alpha\right) = \alpha,$$

то $S = \left\{ \dfrac{p_1(x)}{p_0(x)} \geqslant C_\alpha \right\}$ — наиболее мощный критерий уровня значимости α для проверки H_0 vs. H_1.

Доказательство: Пусть R — произвольный критерий уровня значимости α: $P_0(X \in R) \leqslant \alpha = P_0(X \in S)$

$(p_1(x) - C_\alpha p_0(x)) I\{x \in R\} \leqslant (p_1(x) - C_\alpha p_0(x)) I\{x \in R\} I\{p_1(x) \geqslant C_\alpha p_0(x)\} \leqslant (p_1(x) - C_\alpha p_0(x)) I\{x \in S\}$.

Берем интеграл от левой и правой части

$$\underbrace{P_1(X \in R)}_{\beta_R} - C_\alpha P_0(X \in R) \leqslant \underbrace{P_1(X \in S)}_{\beta_S} - C_\alpha P_0(X \in S),$$

$$\beta_S - \beta_R \geqslant C_\alpha(\underbrace{P_0(X \in S)}_{=\alpha} - \underbrace{P_0(X \in R)}_{=\alpha}) \geqslant 0.$$

□

Утверждение 27. *Для критерия Неймана-Пирсона* $P(I_S) \leqslant \beta_S$.

Доказательство: $S = \left\{ \dfrac{p_1(X)}{p_0(X)} \right\}$.

1. $C_\alpha \geqslant 1 \Rightarrow \forall x \in S p_1(x) \geqslant p_0(x)$

   $\beta_S = P_1(X \in S) = \int\limits_S p_1(x) dx \geqslant \int\limits_S p_0(x) dx = P_0(X \in S) = P(I_S).$

2. $C_\alpha < 1 \Rightarrow \forall x \in \overline{S} p_1(x) < p_0(x)$

   Интегрируем по $\overline{S}$: $\underbrace{P_1(X \notin S)}_{=1-\beta_S} < \underbrace{P_0(X \notin S)}_{=1-P(I_S)}$.

□

9.3. Критерий хи-квадрат

Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения P, $H_0 : P = P_0$ vs. H_1: $P \neq P_0$. Разбиение $\mathscr{X} = \bigsqcup\limits_{j=1}^{k} B_j$, $\mu_j = \#\{i | X_i \in B_j\}$, $p_j^0 = P_0(X_1 \in B_j)$. Статистика критерия

$$\chi(x) = \sum_{j=1}^{k} \frac{(\mu_j - np_j^0)^2}{np_j^0}.$$

Теорема: $\chi(X) \xrightarrow{d_0} \chi^2_{k-1}$.

Доказательство: Рассмотрим вектор $Y_i = \begin{pmatrix} I\{X_i \in B_1\} \\ \ldots \\ I\{X_i \in B_k\} \end{pmatrix}$.

$$\mathbb{E}Y_i = p_0 = \begin{pmatrix} p_1^0 \\ \dots \\ p_k^0 \end{pmatrix},$$

$$\operatorname{cov}_0(I\{X_i \in B_j\}, I\{X_i \in B_l\}) = \mathbb{E}_0 I\{X_i \in B_j \cap B_l\} - \mathbb{E}_0 I\{X_i \in B_j\}\mathbb{E}_0 I\{X_i \in B_l\} =$$
$$= \begin{cases} p_j^0 - (p_j^0)^2, & j = l \\ -p_j^0 p_l^0, & j \neq l \end{cases}.$$

$$\mathbb{D}_0 Y_i = A - p_0 p_0^T, \text{ где } A = \operatorname{diag}(p_1^0, \dots, p_k^0).$$

ЦПТ:

$$\sqrt{n}(Y - p_0) \xrightarrow{d_0} \mathcal{N}(0, A - p_0 p_0^T),$$
$$A^{-1/2} = \operatorname{diag}\left(\frac{1}{\sqrt{p_1^0}}, \dots, \frac{1}{\sqrt{p_k^0}}\right).$$

По теореме о наследовании сходимостей

$$\xi = A^{-1/2}\sqrt{n}(Y - p_0) \xrightarrow{d_0} \mathcal{N}(0, A^{-1/2}(A - p_0 p_0^T)A^{-1/2}) = \mathcal{N}(0, I_k - \sqrt{p_0} \cdot \sqrt{p_0}^T),$$

где $\sqrt{p_0} = \begin{pmatrix} \sqrt{p_1^0} \\ \vdots \\ \sqrt{p_k^0} \end{pmatrix}$. Возьмем $B \in \mathbb{R}^{k \times k} = \begin{pmatrix} \sqrt{p_1^0} & \dots & \sqrt{p_k^0} \\ & \text{что-то} & \end{pmatrix}$ — ортонормированная. По теореме о наследовании сходимостей

$$B\xi \xrightarrow{d_0} \mathcal{N}(0, \underbrace{B I_k B^T}_{I_k} - \underbrace{B\sqrt{p_0}\sqrt{p_0}^T B^T}_{(B\sqrt{p_0})(B\sqrt{p_0})^T}) =$$

$$/B\sqrt{p_0} = \begin{pmatrix} \sqrt{p_1^0} & \dots & \sqrt{p_k^0} \\ & \text{что -то} & \end{pmatrix} \begin{pmatrix} \sqrt{p_1^0} \\ \vdots \\ \sqrt{p_k^0} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix}, \text{ т.к. } \sum_{j=1}^k p_j^0 = 1 \text{ и } B \text{ ортогональна}/$$

$$= \mathcal{N}\left(0, I_k - \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots \\ \vdots & \dots & \dots & \\ 0 & \dots & \dots & 0 \end{pmatrix}\right) = \mathcal{N}\left(0, \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots \\ \vdots & \dots & \dots & \\ 0 & \dots & \dots & 1 \end{pmatrix}\right) = \mathcal{N}(0, I'_k).$$

По теореме о наследовании сходимостей

$$\underbrace{\|B\xi\|^2}_{=\|\xi\|^2,\text{ т.к. } B \text{ орт.}} \xrightarrow{d_0} \|\mathcal{N}(0, I'_k)\| = \chi^2_{k-1}.$$

$$\|\xi\|^2 = \|A^{-1/2}\sqrt{n}(Y - p_0)\|^2 = \sum_{j=1}^k \left[\frac{1}{\sqrt{p_j^0}} \cdot \sqrt{n}\left(\frac{\mu_j}{n} - p_j^0\right)\right]^2 = \sum_{i=1}^k \frac{(\mu_j - np_j^0)^2}{np_j^0} \sim \chi^2_{k-1}.$$

□

Задача 4. $X_1, \dots, X_n \sim U[0, \theta]$.

1. Найти полную статистику. Возьмем $S(X) = X_{(n)}$ — достаточная статистика.

$$p_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n}, \quad x \in [0, \theta],$$

$$\mathbb{E}_\theta f(X_{(n)}) = \int\limits_0^\theta f(x) \frac{nx^{n-1}}{\theta^n} dx = 0 \Leftrightarrow \forall\theta \int\limits_0^\theta f(x) n x^{n-1} dx = 0 \Leftrightarrow f(\theta) \cdot \theta^{n-1} = 0 \Leftrightarrow f(x) = 0.$$

2. Найти оптимальную оценку.

   - $X_{(n)}$ — полная и достаточная статистика;
   - $\mathbb{E}X_{(n)} = \dfrac{n}{n+1}\theta$, берем $\varphi(x) = \dfrac{n+1}{n}x$

     $E_\theta \varphi(X_{(n)}) = \theta$.

     $\theta^* = \dfrac{n+1}{n} X_{(n)}$.