Вопросы:

1) Почерк, понятно расписывать

2) Документация

3) Прак N2 п.1 $X_i \sim Exp(\theta)$

$$y_i = \begin{cases} X_i, & X_i < c \\ c, & X_i \geq c \end{cases}$$

т. Байеса

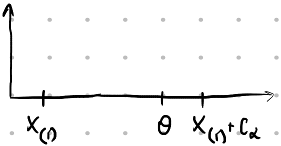
$$p_{y_i}(y_i \mid y_i < c) = \frac{p_{y_i}(y)\overbrace{P(y_i < c \mid y_i = y)}^{1}}{P(y_i < c)} = \frac{p_{y_i}(y)}{F(c)}$$

п. 2

$$E y_i = \int_0^c x\, p_\theta(x)\, dx + c\, P(X_i \geq c)$$

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

дов. инт для $X_{(1)}$



$$P(X_{(1)} \le \theta \le X_{(1)} + c_\alpha) = P(\theta \le X_{(1)} + c_\alpha) =$$

$$\left(X_{(1)}, \frac{X_{(1)}}{\alpha}\right) \quad \text{при } \alpha \approx 1 \sim (0, 0) \not\ni \theta$$

$\Rightarrow$ ошибка

$X_{(n)}$ — сил. сост. оц. θ

$X_{(1)}$ — сил. сост. оц. 0 $\Leftarrow$ $Y_i = \theta - X_i \sim U[0, \theta]$

$$= P(Y_{(n)} \le c_\alpha) = \left(\frac{c_\alpha}{\theta}\right)^n = \alpha \quad \Rightarrow \quad c_\alpha = \sqrt[n]{\alpha}\,\theta$$

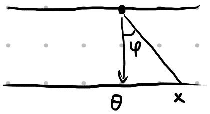
$$\left(X_{(1)} ; \frac{X_{(1)}}{1-\sqrt[n]{\alpha}} \right) \qquad \left(X_{(n)} ; \frac{X_{(n)}}{\sqrt[n]{1-\alpha}} \right)$$

$\downarrow$ возрастает при $n \to \infty$ $\qquad$ $\downarrow$ убывает при $n \to \infty$

т.е. интервал $(0; +\infty)$

бесполезный

5) Ас. дов. инт. для Коши



$\varphi \sim U\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$

$\operatorname{tg} \varphi = X_1 - \theta$

$\varphi = \operatorname{atan}(X_1 - \theta)$

Считаем точный инт. методом центр. ф-ций

Он будет хуже, чем ас. дов. инт. через медиану, т.к. он использует только X_1, а ас. инт. через медиану использует всю выборку.

6) Частая ошибка:

Теорема [L1 – L5] $\Rightarrow$ сост.

L2 – невып. $\not\Rightarrow$ сост.

7) bias – var:

$$MSE_{\hat{\theta}}(\theta) = (E_\theta \hat{\theta} - \theta)^2 + D_\theta \hat{\theta} + \underbrace{D f(x\varepsilon)}_{noise}$$

Мотивация

Как бороться с выбросами?

ас. эф. оц. для θ $\quad \sigma^2 = 2$

$\hat{\mu}$, $\sigma^2 = \pi^2/4 \approx 2{,}47$ – хуже

Решать ОМП не хотим, т.к. сложно

$\Rightarrow$ робастные оценки + метод Ньютона

$|S(X)| = \text{const}$

$\Uparrow$

• Достаточные статистики: (+ экспон. семейство распр.)

для оценки θ надо $(X_1 \ldots X_n)$ – выборка

она большая $\Rightarrow$ не хотим хранить $\Rightarrow$ $s = S(X)$ дост. ст.

3.5

Достаточные статистики

Опр $S(X)$ — достаточная статистика для сем-ва $\{P_\theta \mid \theta \in \Theta\}$, если
важно

$\forall B \in \mathcal{B}(\mathcal{X}) \quad P_\theta(X \in B \mid S(X))$ — не зависит от θ

Теорема (критерий факторизации Неймана–Фишера)

Пусть $\mathcal{P}$ — дом. семейство с пл-ю $p_\theta(x)$

$(X_1 \ldots X_n) \sim P \in \mathcal{P}$

Тогда $S(X)$ — дост. ст. $\iff p_\theta(x) = h(x) \cdot \psi(S(x), \theta)$

Задача 1 $X_1 \ldots X_n \sim N(a, \sigma^2)$ $\theta = (a, \sigma^2)$

Найти дост. статистику.

▲ $$p_\theta(x_1 \ldots x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum \frac{(x_i - a)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{a\sum x_i}{\sigma^2} - \frac{a^2}{2\sigma^2}}$$

$$h(x) = 1$$

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

$$\Rightarrow S = \begin{pmatrix} \sum x_i^2 \\ \sum x_i \end{pmatrix}$$

$$S_{new} = S_{old} + \begin{pmatrix} x_{new}^2 \\ x_{new} \end{pmatrix}$$

Задача 2 $X_1 \ldots X_n \sim$ Exp со сдвигом θ

Найти дост. статистику.

▲ $p_\theta(x) = e^{-(x-\theta)} I\{x \geq \theta\}$

$$p_\theta(x_1 \ldots x_n) = e^{-\sum x_i} e^{n\theta} I\{X_{(1)} \geq \theta\}$$

$$h(x) = e^{-\sum x_i}$$

$$\psi(S(\theta), \theta) = e^{n\theta} I\{X_{(1)} \geq \theta\}$$

$$S(\theta) = X_{(1)}$$

$$S_{new} = \min(S_{old}, x_{new})$$

Робастные оценки

[вообще мы работаем с θ - парам. сдвига для симм. распределений]

Идея: выкинуть выбросы

Опр $\hat{\theta}(X_{(1)} \dots X_{(n)})$ — оценка $[X_{(k_n^*)} \leq \hat{\theta} \leq X_{(n-k_n^*)}]$

Пусть $k_n^* :=$ наим k, при котором

$$\left.\begin{array}{l} x_1 \dots x_{k+1} \to -\infty \text{ при фикс } x_{k+2} \dots x_n \\ \text{либо} \\ x_{n-k} \dots x_n \to +\infty \text{ при фикс } x_1 \dots x_{n-k-1} \end{array}\right. \quad \hat{\theta} \to +\infty$$

Тогда асимптотической толерантностью наз-ся

$$\tau_\theta = \lim_{n \to +\infty} \frac{k_n^*}{n}$$

Пример

$$\hat{\theta}(x_{(1)} \ldots x_{(n)}) = \frac{x_2 + \ldots + x_{n-2}}{n-3}$$

$$k^* = 1$$

$$\tau_\theta = \lim_{n\to\infty} \frac{1}{n} = 0$$

Пример

$$\hat{\theta} = \frac{\sum\limits_{i=\ln(n)+1}^{n-\ln(n)} x_{(i)}}{n - 2\ln n}$$

$$k^* = \ln n \quad (\text{идi } \pm 1)$$

$$\tau_{\hat{\theta}} = \lim_{n\to\infty} \frac{\ln n}{n} = 0$$

Опр. Усечённое среднее :=

$$\alpha \in (0; 1/2)$$

$$k = \lceil \alpha n \rceil$$

$$:= \overline{X}_\alpha = \frac{\sum_{i=k+1}^{n-k} X_{(i)}}{n-2k}$$

Опр. Медиана ср. Уолша := $W = \operatorname{med}\left\{ \frac{x_i + x_j}{2},\ 1 \le i \le j \le n \right\}$

Теорема Пусть 1) P — сим распр с пл-ю $p(x)$

2) Носитель пл-ти $p(x)$ — инт $[-c; c]$ $0 \le c \le +\infty$

3) θ — пар-р сдвига, т.е. $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$

4) $X_1 \ldots X_n \sim P_\theta \in \mathcal{P}$

Тогда $\sqrt{n}\,(\overline{X}_\alpha - \theta) \to N(0, \sigma_\alpha^2)$ $\sigma_\alpha^2 = \dfrac{2}{(1-2\alpha)^2}\left(\int\limits_0^{u_{1-\alpha}} x^2 p(x)\, dx + \alpha u_{1-\alpha}^2\right)$

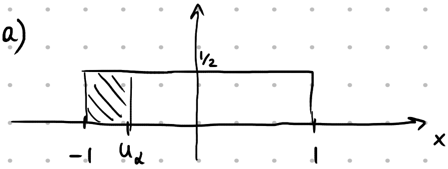
$u_{1-\alpha}$ — $1-\alpha$ кв. для P_0

$\sqrt{n}\,(W - \theta) \to N(0, \sigma_W^2)$ $\sigma_W^2 = \dfrac{1}{12\left(\int\limits_{\mathbb{R}} p^2(x)\, dx\right)^2}$

Задача 3 $U[-1+\theta; 1+\theta]$

Найти ас. дисп. $\sigma^2_{\overline{X}_\alpha}$ и σ^2_W

▲ a)

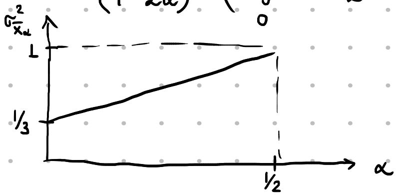


$$\frac{1}{2}(u_\alpha + 1) = \alpha$$

$$u_\alpha = 2\alpha - 1$$

$$u_{1-\alpha} = 1 - 2\alpha$$

$$\sigma^2_{\overline{X}_\alpha} = \frac{2}{(1-2\alpha)^2}\left(\int_0^{1-2\alpha} x^2 \frac{1}{2}\,dx + \alpha(1-2\alpha)^2\right) = \frac{1}{3}(1-2\alpha) + 2\alpha = \frac{4\alpha+1}{3}$$



б) $\sigma_w^2 = \dfrac{1}{12\left(\int\limits_{-1}^{1} \frac{1}{4}\,dx\right)^2} = \dfrac{1}{3}$

$\tau_{\overline{x}} = 0$

$\Rightarrow$ w более робастая, чем $\overline{x}$

$\tau_w = 0.298$

Метод Ньютона

ОМП $= \arg\max L_x(\theta)$ — задача оптимизации

Алгоритм:

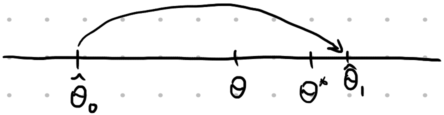
1) $\hat{\theta}_0$ — начальное приближение
2) $\hat{\theta}_{k+1} = \hat{\theta}_k - \left(L''_x(\hat{\theta}_k) \right)^{-1} L'_x(\hat{\theta}_k)$

θ^* — ОМП для θ

Теорема [L1 – L9] $\hat{\theta}_0$ — а.н.о. θ

Тогда $\hat{\theta}_1$ — а.н.о. θ с а.д. $i^{-1}(\theta)$

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



Задача 4 $X_1 \ldots X_n \sim Exp(\theta)$

Куда придём за 1 шаг, если $\hat{\theta}_0 =$ ОМП

▲ $\theta^* = \frac{1}{\overline{X}}$

$L_x(\theta) = n \ln \theta - \theta \sum x_i$

$L_x'(\theta) = \frac{n}{\theta} - \sum x_i$

$L_x''(\theta) = -\frac{n}{\theta^2}$

$$\hat{\theta}_1 = \hat{\theta}_0 + \frac{\theta^2}{n}\left(\frac{n}{\theta} - \sum x_i\right) = \hat{\theta}_0 + \hat{\theta}_0 - \frac{\hat{\theta}_0^2}{n} \sum x_i = 2\hat{\theta}_0 - \hat{\theta}_0^2 \overline{X} =$$

$$= \frac{2}{\overline{X}} - \frac{1}{\overline{X}} = \frac{1}{\overline{X}}$$