Точные доверительные интервалы в нормальной модели

Опр. Пара статистик $(T_1(X), T_2(X))$ наз-ся Д.И. уровня доверия α, если

$$P(T_1(X) \le \theta \le T_2(X)) \geq \alpha$$

Если нер-во точное, то Д.И. точный

$\exists$ $X = (X_1, \dots, X_n) \sim \mathcal{N}(a, \sigma)$

Нужно Д.И. для a

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

1) $\exists$ σ известна

$$X_1 \sim \mathcal{N}(a, \sigma^2) \Rightarrow X_1 + \dots + X_n \sim \mathcal{N}(na, n\sigma^2)$$

Стандартизируем: $\frac{\sum X_i - na}{\sqrt{n}\,\sigma} = \sqrt{n}\,\frac{\overline{X} - a}{\sigma} \sim \mathcal{N}(0,1)$

$$P\left(z_{\frac{1-\alpha}{2}} \le \sqrt{n}\,\frac{\overline{X} - a}{\sigma} \le z_{\frac{1+\alpha}{2}}\right) = \alpha$$

$$a \in \left(\overline{X} \pm \frac{z_{\frac{1+\alpha}{2}}}{\sqrt{n}}\sigma\right)$$ <u>если знаем σ</u>

2) $\exists$ <u>σ неизвестна</u>

Th 1 (б/д)

$\exists$ $X_1, \dots, X_n \sim \mathcal{N}(a, \sigma^2)$. Тогда

(1) $\overline{X}$ и s^2 независимы

(2) $\frac{n s^2}{\sigma^2} \sim \chi^2_{n-1}$

(3) $\sqrt{n-1}\,\frac{\overline{X}-a}{s} \sim T_{n-1}$

◁ Д-м только (3)

$$\sqrt{n}\,\frac{\overline{X}-a}{\sigma} \sim \mathcal{N}(0,1)$$

из (2) $\Rightarrow$ $\frac{n s^2}{\sigma^2} \sim \chi^2_{n-1}$

Из презентации: $\xi \sim \mathcal{N}(0,1)$ $\eta \sim \chi^2_k$ нез-мы $\Rightarrow$

$$\frac{\xi}{\sqrt{\frac{\eta}{k}}} = \frac{\sqrt{n}\,\frac{\overline{X}-a}{\sigma}}{\sqrt{\frac{n s^2}{\sigma^2 (n-1)}}} = \sqrt{n-1}\,\frac{\overline{X}-a}{s} \sim T_{n-1}$$

$$P\left(T_{n-1,\frac{1-\alpha}{2}} \le \sqrt{n-1}\,\frac{\overline{X}-a}{s} \le T_{n-1,\frac{1+\alpha}{2}}\right) = \alpha$$

$\|$ т.к. T симметр.

$-T_{n-1,\frac{1+\alpha}{2}}$

для норм. распред. это совпадало бы с $z_{\frac{1+\alpha}{2}}$

$$a \in \left(\overline{X} \pm \frac{T_{n-1,\frac{1+\alpha}{2}}}{\sqrt{n-1}}\, s\right)$$

Оценка максимального правдоподобия

$\exists\ X = (X_1, \dots, X_n) \in P \in \mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$

где все P_θ либо абс. непр., либо дискр.

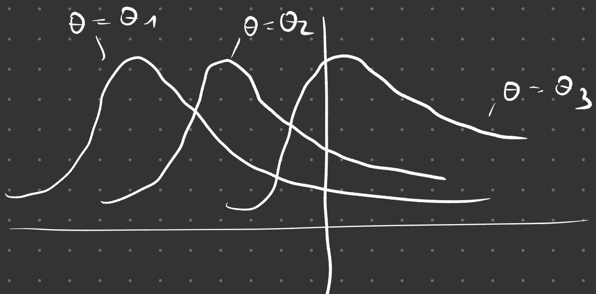
Если P_θ дискр, то $p_\theta(x) = P(X_1 = x)$

Если P_θ непр., то $p_\theta(x)$ – пл-ть

<u>Опр.</u> $L_X(\theta) = \prod_{i=1}^{n} p_\theta(X_i)$ – ф-ция правдоподобия

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

<u>Замеч.</u> Аргумент ф-ии θ (не X)



смысл „вероятность" выборки в зав-ти от θ

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

$$p_\theta(x) = \begin{cases} \theta, & \text{если } x = 1 \\ 1-\theta, & x = 0 \end{cases} = \theta^x (1-\theta)^{1-x}$$

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

$$l_X(\theta) = \sum_{i=1}^{n} \left[X_i \ln\theta + (1-X_i)\ln(1-\theta) \right]$$

$-l_x(x)$ — logLoss (из логрега)

<u>Опр.</u> $\hat{\theta}_{ОМП} = \arg\max\limits_{\theta \in \Theta} L_x(\theta)$ — ОМП (оценка макс. пл-ти)

<u>Th1</u> Пусть

1) носитель пл-ти $(\{x : p_\theta(x) > 0\})$ не зависит от θ

2) Θ — открытое м-во в $\mathbb{R}^d$

3) Пл-ть $p_\theta(x)$ „достаточно гладкая"

4) Ур-ие правд. $\frac{\partial l_x}{\partial \theta} = 0$ имеет ровно 1 решение

Тогда: $\hat{\theta}_{ОМП}$ — состоятельная и ас. нормальная оценка θ с ас. матрицей ковариаций

$$\Sigma(\theta) = i^{-1}(\theta), \text{ где } (i(\theta))_{jk} = \mathbb{E}_\theta\left(\frac{\partial l_{x_1}}{\partial \theta_j} \cdot \frac{\partial l_{x_1}}{\partial \theta_k}\right)$$

здесь $n=1$

Пример:

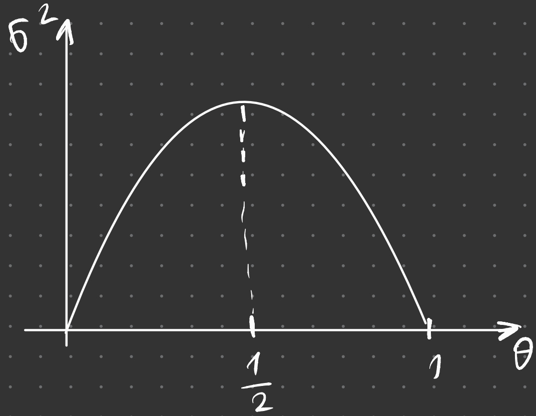
$$l_x(\theta) = \sum_{i=1}^{n} [x_i \ln\theta + (1-x_i)\ln(1-\theta)]$$

$$l_{x_1}(\theta) = x \ln\theta + (1-x_1)\ln(1-\theta)$$

$$\frac{\partial l_{x_1}}{\partial \theta} = \frac{x_1}{\theta} - \frac{1-x_1}{1-\theta} = \frac{x_1-\theta}{\theta(1-\theta)}$$

$$i(\theta) = \mathbb{E}_\theta \frac{(X_1-\theta)^2}{\theta^2(1-\theta)^2} = \frac{D_\theta X_1}{\theta^2(1-\theta)^2} = \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2} \Rightarrow$$

$$\sigma^2(\theta) = \theta(1-\theta)$$



$$\frac{\partial l_x}{\partial \theta} = \frac{\sum (x_i-\theta)}{\theta(1-\theta)} = 0$$

$$\Rightarrow \boxed{\hat{\theta}_{ОМП} = \overline{X_i}}$$

<u>Пример</u>: $X_1, \ldots, X_n \sim \text{Cauchy}(\theta)$

$$\sigma^2_{\hat{\mu}} = \frac{\pi^2}{4}$$ — для медианы

оценка медианой

$$L_x(\theta) = \prod_{i=1}^{n} \frac{1}{\pi} \frac{1}{1+(x_i-\theta)^2}$$

$$l_x = -n \ln \pi - \sum_{i=1}^{n} \ln(1+(x_i-\theta))$$

$$\frac{\partial l_x}{\partial \theta} = \sum_{i=1}^{n} \frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = 0$$

(в общем случае не решается формульно, только численно. И решение м.б. не единственно)

$$\mathbb{E}_\theta \left(\frac{\partial l_{X_1}}{\partial \theta} \right)^2 = \mathbb{E}_\theta \frac{4 (X_1 - \theta)^2}{(1 + (X_1 - \theta)^2)^2} = \frac{1}{2}$$

$$\mathbb{E} f(x) = \int f(x)\, p(x)\, dx$$

$$\stackrel{\text{Th}}{\Longrightarrow} \quad \sigma^2_{\text{ОМП}} = 2 < \sigma^2_{\hat{\mu}}$$