Лекция 4

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

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

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

$$l_X(\theta) = \ln L_X(\theta) = \sum X_i \ln\theta + (n - \sum X_i)\ln(1-\theta)$$

$$\frac{\partial l_X(\theta)}{\partial\theta} = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta} = 0$$

$$(1-\theta)\sum X_i = \theta(n - \sum X_i)$$

$$\sum X_i = n\theta \Rightarrow \theta = \overline{X}.$$

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

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

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

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 $\overline{X}$: $\varphi_X(t) = Ee^{itX}$. Для Коши $\varphi_{X_1} = e^{-|t|}$ $(\theta = 0)$.

$$\varphi_{\overline{X}}(t) = Ee^{it\overline{X}} = Ee^{it\frac{1}{n}\sum X_i} = E\prod_{i=1}^{n} e^{i(t/n)X_i} = /\text{незав.}/ = \prod_{i=1}^{n} Ee^{i(t/n)X_i} = |X_i \overset{d}{=} X_1| =$$

$$= \left(Ee^{i(t/n)X_1}\right)^n = e^{-|t|} = \varphi_{X_1}(t) \Rightarrow \text{ по теореме о единственности } \overline{X} \overset{d}{=} X_1.$$

Вывод: усреднение ничего не дает.

4. Медиана - рассмотрим далее.

□

Выборочные квантили

Определение 1. 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 медиантой.

Пример 2. $Exp(1)$, $F(x) = 1 - e^{-x}$, $u_p = -\ln(1-p)$ — p-квантиль $Exp(1)$.

Определение 2. 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 3. $X = (7, 9, 15, 8, 12, 1, 8, 5, 17, 21)$. Найти выборочные квантили уровней 0.01, 0.1, 0.25 и медиану.

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

Теорема 1. *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.