Лекция 3

> **Теорема (дельта-метод):** Пусть $\hat{\theta}_n$ — асимптотически нормальная оценка $\theta \in \Theta \subset \mathbb{R}^d$ 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:

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

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

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

□

Пример: $X_1, \dots X_n \sim \mathrm{Exp}(\theta)$, $\theta > 0$. 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:

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

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

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

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

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.
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 $E(h(\xi_n)) \to$ $Ef(h(\xi)) \Rightarrow h(\xi_n) \xrightarrow{d} h(\xi)$ 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 $F_{\xi+c}(x)$.

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

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

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

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

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

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 $\lim\limits_{n\to\infty} F_{\xi_n+\eta_n}(t) = F_{\xi+c}(t) \Rightarrow \xi_n + \eta_n \xrightarrow{d} \xi + c$.

$\square$

2.4. Методы нахождения оценок

(1) Метод моментов

Идея: приравняем друг к другу теоретические и выборочные моменты.

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

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

Решение этой системы называется оценкой θ по методу моментов.

Обобщенный метод моментов

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

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

Пример: $X_1, \ldots, X_n \sim Exp(\theta)$. 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: $E_\theta I(X_1 > 1) = \overline{I(X > 1)}$. $E_\theta I\{X_1 > 1\} = \int_1^{+\infty} \theta e^{-\theta x} dx = e^{-\theta}$, $\overline{I\{X > 1\}} = \dfrac{1}{n}\sum_{i=1}^{n} I\{X_i > 1\}$. 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 $(e^{-\theta} - e^{-2\theta}) \cdot ((-\ln x)')^2\Big|_{e^{-\theta}} = (e^{-\theta} - e^{-2\theta}) \cdot \dfrac{1}{x^2}\Big|_{e^{-\theta}} = e^{\theta} - 1 = \sigma_2^2(\theta)$.

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 $g_1(x), \ldots, g_d(x)$ — борелевские функции такие, что $|E_\theta g_i(x_i)| < +\infty$.

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

> **Утверждение:**
>
> 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. $\square$

(2) Метод максимального правдоподобия

Пусть $X = (X_1, \ldots X_n)$ — 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:** $L_X(\theta) = p_\theta(X) = \prod_{i=1}^{n} p_\theta(X_i)$ — функция правдоподобия (как функция от θ).

> **Определение:** $l_X(\theta) = L_X(\theta)$ — логарифмическая функция правдоподобия.

Замечание: При фиксированном θ функция правдоподобия равна плотности выборки, в которую в качестве аргумента подставлена сама выборка.

Смысл: "вероятность" выборки в зависимости от значения параметра. Степень доверия к конкретному значению параметра. Интересует только относительное значение.

Пример: пусть x_1 — наблюдение.

Рисунок

Видимо θ_2 более правдоподобно, чем θ_1 и θ_3.

> **Определение:** $\hat{\theta} = \arg\max\limits_{\theta\in\Theta} L_X(\theta)$ называется оценкой максимального правдоподобия.

> **Утверждение:** 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:** Пусть $\forall n,\ \forall x_1, \ldots, x_n$ 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 коварияций $i(\theta)^{-1}$, 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: $X_1, \ldots, X_n \sim \mathrm{Exp}(\theta)$. 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 $i(\theta)^{-1} = \theta^2$.