

# М-оценки (M-estimators)

Звонарев Никита

Слушатели: 4-й курс бакалавриата

Санкт-Петербургский Государственный Университет  
Кафедра Статистического моделирования



Санкт-Петербург – 2024 г.

*М-оценки* — статистические оценки, получаемые с помощью максимизации некоторой функции от выборки.

Зачем нужны?

- Обобщение ОМП (оценок максимального правдоподобия)
- Нелинейная регрессия
- Робастное оценивание
- Оценивание в случае цензурированной выборки
- ...

# Формальное определение

$X_1, \dots, X_n, X_i \in \mathcal{X}$  — выборка из независимых одинаково распределенных величин с распределением  $\mathcal{P} = \mathcal{P}_{\theta_0}$ ,  
 $\theta_0 \in \Theta$  — параметр распределения.

Нужна статистика  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$  — оценка  $\theta_0$ .

Введем  $m_\theta(x) : \mathcal{X} \rightarrow \overline{\mathbb{R}}$  (допускаются значения  $+\infty, -\infty$ ) — “функция-критерий” и  $M_n(\theta) = \frac{1}{n} \sum_{i=1}^n m_\theta(X_i)$ .

Тогда  $\hat{\theta}_n = \arg \max_{\theta \in \Theta} M_n(\theta)$  называется М-оценкой.

## Замечание

При  $m_\theta(x) = \log p_\theta(x)$ , где  $p_\theta(x)$  — плотность распределения  $\mathcal{P}_\theta$ , М-оценка является оценкой максимального правдоподобия.

# Оценки пси-типа (Z-оценки)

Аналогично можно рассмотреть  $\Psi_n(\theta) = \frac{1}{n} \sum_{i=1}^n \psi_\theta(x_i)$ .

Тогда  $\tilde{\theta}_n$ :  $\Psi_n(\tilde{\theta}_n) = 0$  называется Z-оценкой.

При условии, что есть дифференцируемость  $m_\theta(x)$  (+ условия достаточности для глобального экстремума):  $\psi_\theta(x) = \frac{\partial m_\theta(x)}{\partial \theta}$ .

## Замечание

*M-оценки и Z-оценки не обязательно эквивалентны. Пример: пусть  $\mathcal{P}_\theta$  — равномерное распределение на  $[0; \theta]$ . В этом случае M-оценка существует, Z-оценки нет.*

## Замечание

*На слайде представлен только случай одномерной  $\theta$ . Для  $k$ -мерной  $\theta$  Z-оценка обычно должна удовлетворять  $k$  условиям вида  $\Psi_{n,k}(\tilde{\theta}_n) = 0$ , а  $\psi_{\theta,i}(x)$  должна быть частной производной  $m_\theta(x)$  по  $i$ -й компоненте.*

# Состоятельность. Теорема Вальда

Рассмотрим метрику  $d$  на  $\Theta$ ,  $\theta_0$  — “истинное” значение параметра.  
Вопрос: когда  $d(\hat{\theta}_n, \theta_0) \rightarrow 0$  по вероятности?

- ❶ Отображение  $m_\theta$  — полунепрерывно сверху по  $\theta$ , то есть для почти всех  $x$ :  $\limsup_{\theta_n \rightarrow \theta} m_{\theta_n}(x) \leq m_\theta(x)$ .
- ❷ Для любого достаточно маленького шара  $\mathcal{U} \in \Theta$  отображение  $q_{\mathcal{U}}(x) = \sup_{\theta \in \mathcal{U}} m_\theta(x)$  измеримо и  $\int q_{\mathcal{U}}(x) d\mathcal{P} < +\infty$ .
- ❸ Множество максимумов  
 $\Theta^* = \{\theta^* \in \Theta : \int m_{\theta^*}(x) d\mathcal{P} = \sup_{\theta \in \Theta} \int m_\theta(x) d\mathcal{P}\}$ .  
“Идеальная” ситуация:  $\Theta^* = \{\theta_0\}$ .

## Теорема (Wald, 1949)

Пусть  $m_\theta$  удовлетворяет условиям 1 и 2. Тогда для любой оценки  $\hat{\theta}_n$  такой, что  $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$ , любого  $\varepsilon > 0$ , любого компакта  $\mathcal{K}$ :

$$P(d(\hat{\theta}_n, \Theta^*) \geq \varepsilon, \hat{\theta}_n \in \mathcal{K}) \rightarrow 0 \text{ при } n \rightarrow +\infty.$$

Рассмотрим  $m_\theta(x) = \log \frac{p_\theta(x)}{p_{\theta_0}(x)}$ .

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)}.$$

$M(\theta) = \int \left( \log \frac{p_\theta(x)}{p_{\theta_0}(x)} \right) p_{\theta_0}(x) dx$  — достигает максимума в точке  $\theta_0$  (см. нер-во Йенсена!).

По ЗБЧ:  $M_n(\theta) \rightarrow M(\theta)$  по вероятности.

Оставшееся условие на плотность.

## Утверждение (Идентифицируемость)

Пусть из того, что  $p_{\theta_1} = p_{\theta_2}$  почти всюду следует то, что  $\theta_1 = \theta_2$ .

Тогда  $\Theta^* = \{\theta_0\}$ .

Чтобы выполнялась состоятельность ОМП, нужны теорема Вальда + идентифицируемость.

## Утверждение

При выполнении некоторых условий регулярности:

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \rightarrow N \left( 0, I_{\theta_0}^{-1} \right) \text{ по распределению,}$$

где  $I_{\theta_0} = \mathbb{E}_{\theta_0} \left[ \frac{\partial}{\partial \theta} \log p_{\theta}(x) \cdot \frac{\partial}{\partial \theta} \log p_{\theta}(x)^T \right]$  — информационная матрица Фишера.

Пример условий: рассмотрим  $\psi_{\theta}(x) = \frac{\partial \log p_{\theta}(x)}{\partial \theta}$ ,  $\psi(\theta) = \mathbb{E}_{\theta_0} \psi_{\theta}$ .

Пусть для  $\psi_{\theta}(x)$  выполняется  $\|\psi_{\theta_1}(x) - \psi_{\theta_2}(x)\| \leq \hat{\psi}(x) \|\theta_1 - \theta_2\|$ , при этом  $\mathbb{E}_{\theta_0} \hat{\psi}^2 < +\infty$ ,  $\mathbb{E}_{\theta_0} \|\psi_{\theta_0}\|^2 < +\infty$  и  $\psi(\theta)$  дифференцируема в своём нуле  $\theta_0$  с невырожденной матрицей производных, оценка  $\hat{\theta}_n$  состоятельна и  $\frac{1}{n} \sum_{i=1}^n \psi_{\hat{\theta}_n}(X_i) = o_p(n^{-1/2})$ , то выполнена асимптотическая нормальность.

Асимптотически задаётся через т.н. Wald approximation:

$$n(\hat{\theta}_n - \theta_0)^T I_{\theta_0} (\hat{\theta}_n - \theta_0) \sim \chi_k^2$$

при  $n \rightarrow +\infty$  ( $k$  – размерность  $\theta$ ).

Проблема: при фиксированной квантили распределения  $\chi_k^2$  множество допустимых параметров это эллипсоид, а нам нужен способ понизить размерность (в том числе до 1 — т.е. до доверительного интервала).

# Метод профилей правдоподобия (Profile likelihood)

Разобьём параметр на два подмножества  $\theta = (\theta_1, \theta_2)$ ,  $\theta_1 \in \mathbb{R}^k$ .

Введем  $R(\theta_1) = \max_{\theta_2} L(\theta_1, \theta_2)$  — профиль правдоподобия, где  $L$  — функция правдоподобия. Выполнено:

$$2 \log \frac{R(\hat{\theta}_1)}{R(\theta_1)} \rightarrow_D \chi_k^2$$

при  $n \rightarrow \infty$ .

При  $k = 1$  имеем способ построить доверительный интервал для  $i$ -й координаты.

При этом следует обращать внимание на форму профиля возле точки максимума. Модель можно считать достаточно близкой к асимптотической (т.е. к нормальной), если в этом регионе профиль хорошо приближается параболой. Если этого нет, то доверять таким интервалам нельзя.

Пакет `bbmle` позволяет проверить качество приближения параболой взятием корня из значения профиля. После этого он должен выглядеть “галочкой” (как функция  $|x|$ ).