Лекция 6 (от 7.10)

2.8. Приближенный поиск ОМП

Метод Ньютона: Пусть $f : \mathbb{R} \to \mathbb{R}$ — функция. Нужно решить уравнение $f(x) = 0$.

x_0 — начальное приближение
Формула касательной в точке $x_k : y = f(x_k) + f'(x_k)(x - x_k)$. Получим соотношение

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Пусть $X = (X_1, \ldots, X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta \mid \theta \in \Theta\}, \Theta \subset \mathbb{R}^d$. Пусть $\theta^{\backslash *}$ — ОМП. Хотим приблизить оценку θ^*.

Уравнение правдоподобия: $\dfrac{\partial l_X(\theta)}{\partial \theta} = 0$. Применим метод Ньютона для функции $l'_X(\theta)$.
$\widehat{\theta}_0$ — начальное приближение. Шаг метода:

$$\widehat{\theta}_{k+1} = \widehat{\theta}_k - \underbrace{(l''_X(\widehat{\theta}_k))^{-1}}_{\text{матрица}} \cdot \underbrace{l'_X(\widehat{\theta}_k)}_{\text{вектор}}.$$

> **Теорема:** В условиях регулярности $L1 - L9$, если $\widehat{\theta}_0$ — а.н.о, то
>
> 1. $\widehat{\theta}_1$ — а.н.о с асимт. дисперсией $(i(\theta))^{-1}$.
> 2. $\widehat{\theta}_1$ асимптотически эквивалентна ОМП θ^*, т.е
>
> $$\sqrt{n}(\widehat{\theta}_1 - \theta^*) \xrightarrow{P_\theta} 0.$$

Доказательство: (для $d = 1$, идея)
Утв. (б/д): $\widehat{\theta}_1 - \theta^* = (\widehat{\theta}_0 - \theta^*)\varepsilon_n(\theta)$, где $\varepsilon_n(\theta) \xrightarrow{P_\theta} 0$.

(2). $\sqrt{n}(\widehat{\theta}_1 - \theta^*) = \sqrt{n}(\widehat{\theta}_0 - \theta^*)\varepsilon_n(\theta) =$

$$= \underbrace{\underbrace{\sqrt{n}(\widehat{\theta}_0 - \theta)}_{\xrightarrow{d_\theta} \mathcal{N}(0, \ldots)} \underbrace{\varepsilon_n(\theta)}_{\xrightarrow{d_\theta} 0}} + \underbrace{\underbrace{\sqrt{n}(\theta - \theta^*)}_{\xrightarrow{d_\theta} \mathcal{N}(0, \ldots)} \underbrace{\varepsilon_n(\theta)}_{\xrightarrow{d_\theta} 0}}.$$

По лемме Слуцкого первое слагаемое $\xrightarrow{d_\theta} 0$, второе слагаемое $\xrightarrow{d_\theta} 0$. Применяя еще раз лемму Слуцкого для их суммы, получим $\sqrt{n}(\widehat{\theta}_1 - \theta^*) \xrightarrow{d_\theta (\iff P_\theta, \text{т.к const})} 0$.

(1). $\sqrt{n}(\widehat{\theta}_1 - \theta) = \underbrace{\sqrt{n}(\widehat{\theta}_1 - \theta^*)}_{\xrightarrow{P_\theta} 0 (\text{из } (2))} - \underbrace{\sqrt{n}(\widehat{\theta}_0 - \theta)}_{\xrightarrow{d_\theta} \mathcal{N}(0, \frac{1}{i(\theta)}) \text{ (ОМП)}}$. По лемме Слуцкого

$$\sqrt{n}(\widehat{\theta}_1 - \theta) \xrightarrow{d_\theta} \mathcal{N}\left(0, \frac{1}{i(\theta)}\right) \qquad \square.$$

Замечание: Утверждение теоремы не изменится, если заменить $l''_X(\theta)$ на $E_\theta l''_X(\theta) = -ni(\theta)$, т.е.

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{i(\theta)^{-1}}{n} l'_X(\theta).$$

Оценка $\widehat{\theta}_1$ называется *одношаговой оценкой*.

Смысл:
Отклонение $\widehat{\theta}_1$ от θ^* на порядок меньше, чем отклонение θ^* от θ. Значит отклонение $\widehat{\theta}_1$ от θ тоже имеет порядок $\sqrt{\frac{1/i(\theta)}{n}}$.

Пример (γ-котики):
$\widehat{\mu}$ — а.н.о. с асимпт. дисперсией $\pi^2/4 \approx 2.47$. При этом $i(\theta) = 1/2$, т.е наименьшая возможная асимпт. дисперсия равна 2. Запишем одношаговую оценку:

$$\widehat{\theta}_1 = \widehat{\mu} + \frac{\sum\limits_{i=1}^{n} \frac{X_i - \widehat{\mu}}{1 + (X_i - \widehat{\mu})^2}}{\sum\limits_{i=1}^{n} \frac{1 - (X_i - \widehat{\mu})^2}{(1 + (X_i - \widehat{\mu})^2)^2}}.$$

$\widehat{\theta}_1$ — наиболее асимптотически эффективная оценка.

2.9. Робастность и симметричные распределения

Пусть $X = (X_1, \ldots, X_n)$ — выборка из $\mathcal{N}(\theta, \sigma^2)$, σ известна.
Оценка $\widehat{\theta} = \overline{X}$ обладает всеми хорошими свойствами (сильная состоятельность, асимптотическая нормальность, ОМП и т. д.). Однако если в данных есть выбросы, то все свойства теряются.
Для того, чтобы визуализировать выбросы в данных, можно использовать *ящик с усами (box plot)*.

Будем рассматривать только одномерный случай.

> **Определение:** *Робастная оценка* — оценка, допускающая отклонение от заданной модели.

> **Определение:** Пусть оценка имеет вид $\widehat{\theta} = f(X_{(1)}, \ldots, X_{(n)})$.
> Пусть k_n^* — наименьшее число k, т. ч. выполнено одно из условий:
>
> 1. Если $x_1, \ldots, x_{k+1} \to -\infty$, а $x_{k+2}, \ldots, x_n$ фиксированы,то $f(x_1, \ldots, x_n) \to -\infty$.
> 2. Если $x_{n-k}, \ldots, x_n \to +\infty$, а $x_1, \ldots, x_{n-k+1}$ фиксированы, то $f(x_1, \ldots, x_n) \to +\infty$.
>
> Тогда число $\tau_{\widehat{\theta}} = \lim\limits_{n \to \infty} \dfrac{k_n^*}{n}$ называется *асимптотической толерантностью оценки* $\widehat{\theta}$.

Смысл: $\tau(\theta)$ — наибольшая доля выбросов, которые способна выдержать оценка, не смещаясь на $\pm\infty$.

Примеры:

- $\overline{X} : k_n^* = 0, \tau_{\overline{X}} = 0$
- $\widehat{\mu} : k_n^* = \lceil n/2 \rceil - 1, \tau_{\widehat{\mu}} = 1/2.$

Далее будем рассматривать класс распределений $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, т. ч.

- P_0 имеет плотность $p_0(x)$ — симметричная, непрерывная, носитель плотности имеет вид $(-c, c),\ 0 < c \leqslant +\infty$.
- θ — параметр сдвига, т. е. $p_\theta(x) = p_0(x-\theta $.

Будем искать оценки, которые:

1. Достаточно эффективные в классе $\mathcal{P}$ (в асимптотическом подходе).
2. Робастные — допускают отклонение от $\mathcal{P}$.

1. Усеченное среднее

> **Определение:** Пусть $\alpha \in (0, 1/2),\ k = \lceil \alpha n \rceil$. Тогда *усеченным средним по выборке* $X_1, \ldots, X_n$ называется оценка
>
> $$\overline{X}_\alpha = \frac{1}{n-2k}(X_{(k-1)} + \cdots + X_{(n-k)}).$$

- $\alpha = 0$: $\overline{X}_\alpha = \overline{X}$
- $\alpha = 1/2 : \overline{X}_\alpha = \widehat{\mu}.$

Асимптотическая толерантность: $\tau_{\overline{X}_\alpha} = \alpha$.

> **Теорема** *(б/д)*: Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения $P \in \mathcal{P}$. Тогда
>
> $$\sqrt{n}(\overline{X}_\alpha - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma_\alpha^2),\ \text{где}$$
>
> $$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left(\int\limits_0^\theta x^2 p_0(x)dx + \alpha u_{1-\alpha}^2\right),$$
>
> $u_{1-\alpha}$ — $(1-\alpha)$-квантиль распределения P_0.

Пример: для $\mathcal{N}(0,1)$

α	0	$1/20$	$1/8$	$1/4$	$3/8$	$1/2$
$\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}}$	1	0.99	0.94	0.84	0.74	0.64

При $\alpha = 1/8$ достигается защита от 12.5% загрязнения выборки, но эффективность теряется на 6%.

Утв: Если $D_\theta X_1 < +\infty$, то $\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}} \geqslant (1-2\alpha)^2$.

$\triangle \quad \overline{X}_\alpha$ — а.н.о θ с асимпт. дисперсией σ_α^2.
Из ЦПТ: $\overline{X}$ — а.н.о θ с асимпт. дисперсией $D_\theta X_1$. Так как дисперсия не зависит от сдвига, посчитаем дисперсию при $\theta = 0$:

$$\frac{1}{2} D_\theta X_1 = \frac{1}{2}\int\limits_{\mathbb{R}} x^2 p_0(x)dx = \int\limits_0^{+\infty} x^2 p_0(x)dx =$$

$$= \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \int\limits_{u_{1-\alpha}}^{+\infty} x^2 p_0(x)dx \geqslant$$

$$\geqslant \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + u_{1-\alpha}^2 \underbrace{\int\limits_{u_{1-\alpha}}^{+\infty} p_0(x)dx}_{=\alpha} = \int\limits_0^{u_{1-\alpha}} x^2 p_0(x)dx + \alpha u_{1-\alpha}^2 = \frac{\sigma_\alpha^2 (1-2\alpha)^2}{2}.$$

Отсюда $\mathrm{ARE}_{\overline{X}_\alpha, \overline{X}} = \dfrac{D_\theta X_1}{\sigma_\alpha^2} \geqslant (1-2\alpha)^2 \qquad \square.$

α	0	$1/20$	$1/8$	$1/4$	$3/8$	$1/2$
$(1-2\alpha)^2$	1	0.81	0.56	0.25	0.06	0

При $\alpha = 1/8$ возможна потеря эффективности до 44%.

2. Медиана средних Уолша

> **Определение:** $Y_{ij} = \dfrac{X_i + X_j}{2}$ — *среднее Уолша.*
>
> $W = \mathrm{med}\{Y_{ij},\ 1 \leqslant i \leqslant j \leqslant n\}$ — *медиана средних Уолша.*

> **Теорема:** Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $P \in \mathcal{P}$. Тогда
>
> $$\sqrt{n}(W - \theta) \xrightarrow{d_\theta} \mathcal{N}(0, \sigma^2), \text{ где}$$
>
> $$\sigma^2 = \frac{1}{12\left(\int\limits_{\mathbb{R}} p_0^2(x)dx\right)^2}.$$

Пример: $\mathcal{N}(0,1) : \mathrm{ARE}_{W,\overline{X}} \approx 0.955$ (потеря эффективности на 4.5%).

Утверждение: Для $P_\theta \in \mathcal{P}$ $\mathrm{ARE}_{W,\overline{X}} \geqslant \frac{108}{125} = 0.864$ (в худшем случае теряем 14% эффективности). Равенство достигается при

$$p_0(x) = \frac{3\sqrt{5}}{100}(5 - x^2) I\{|x| < \sqrt{5}\}.$$

Утверждение: $\tau_W \approx 0.293$ (доказательство см. в ДЗ).

Глава 3. Сложные оценки параметров

3.1. Доверительные интервалы

> **Определение:** Пусть $X = (X_1, \dots, X_n)$ — выборка из неизвестного распределения $P \in \{P_\theta \mid \theta \in \Theta\}$.
>
> - Если $\Theta \subset \mathbb{R}$, то пара статистик $(T_1(X), T_2(X))$ называется *доверительным интервалом для θ уровня доверия α*, если
>
> $$\forall \theta \in \Theta \quad P_\theta(T_1(X) \leqslant \theta \leqslant T_2(X)) \geqslant \alpha.$$

> - Если $\Theta \subset \mathbb{R}^d$, то статистика $S(X) \subset \Theta$ называется *доверительной областью для θ уровня доверия α*, если
>
> $$\forall \theta \in \Theta \quad P_\theta(\theta \in S(X)) \geqslant \alpha.$$
>
> - Если равенство точное, то интервал называтся *точным*.

Замечание:

1. Если $X = (X_1, \ldots, X_n)$ — выборка, то утверждение $P_\theta(T_1(X) \leqslant \theta \leqslant T_2(X)) = \alpha$ имеет смысл ($(T_1(X), T_2(X))$ — доверительный интервал).
2. Если $x = (x_1, \ldots, x_n)$ — реализация выборки, то утверждение $P_\theta(T_1(x) \leqslant \theta \leqslant T_2(x)) = \alpha$ некорректно.

$(T_1(x), T_2(x))$ — *реализация доверительного интервала*.

Первая магическая константа статистики: $\alpha = 0.95(\text{ она же } 0.05)$.