Экспоненциальный класс распределений

3.6

Опр. Семейство $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ принадлежит **эксп кл распр.**, если $p_\theta(x) = \dfrac{g(x)}{h(\theta)} e^{a(\theta)^T u(x)}$, где

$$g(x) > 0, \quad u(x) \text{ — борелевские}$$

$$h(\theta) = \int_{\mathcal{X}} g(x)\, e^{a(\theta)^T u(x)}\, dx$$

Если $a(\theta) = \theta$, то говорят о **естественной параметризации**

Коши, U $\notin$ эксп. кл. распр.

Exp, N, Γ $\in$ эксп. кл. распр.

Пример $\mathcal{P} = \{ N(a, \sigma) \mid a \in \mathbb{R},\ \sigma > 0 \}$ $\in$ эксп. кл.?

$$\blacktriangle \quad p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2} + \frac{ax}{\sigma^2} - \frac{a^2}{2\sigma^2}} =$$

$$\left[u(x) = \begin{pmatrix} -x^2 \\ x \end{pmatrix} \qquad \theta = \begin{pmatrix} 1/2\sigma^2 \\ a/\sigma^2 \end{pmatrix} \right]$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{a^2}{2\sigma^2}} e^{\frac{-x^2}{2\sigma^2} + \frac{ax}{\sigma^2}} \qquad \left[\frac{a^2}{2\sigma^2} = \frac{\theta_2^2}{4\theta_1} \right]$$

$$p_\theta(x) = \sqrt{\frac{\theta_1}{\pi}}\, e^{\frac{\theta_2^2}{4\theta_1}} e^{\theta^T u(x)}$$

$$g(x) = 1$$

Найдём достаточные статистики

$$p_\theta(x_1 \ldots x_n) = h(\theta)^{-n} \prod_{i=1}^{n} g(x_i)\, e^{a(\theta)^T \sum_{i=1}^{n} u(x_i)}$$

Тогда по кр. факторизации:

$$S(x) = \sum_{i=1}^{n} u(x_i) \text{ — достаточная статистика}$$

хорошая статистика, т.к. надо хранить только 1 число

Критерий $\quad p_\theta(x) = \psi(S(x), \theta)\, h(x)$

$$L_x(\theta)' = \left(\sum \ln \psi(S(x), \theta) + \ln h(x)\right)' = f(S(X))$$

т.е. не надо хранить выборку для ОМП, хватит $S(X) \in \mathbb{R}$

Теорема (б/д) $\mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$, т.ч. $p_\theta(x)$ – непр. дифф. по x

Носитель $p_\theta(x)$ не зависит от θ.

$S(X)$ – достаточная статистика фикс. размерности

т.е. не зависит от $|X|$

То $\mathcal{P} \in$ экспоненциальному классу

(при этом $\dim S(X) \geq \dim a(\theta)$)

Следствие: если плотность „достаточно хорошая", то только семейства из экспоненциального класса допускают сжатия данных с помощью дост. статистик

Примеры 1) Cauchy(θ) не лежат в эксп. классе;

выполнены усл. теоремы $\Rightarrow$ нет дост. ст. фикс. размера

2) $U[0, \theta]$ носитель зависит от θ;

не лежит в эксп. классе

есть дост. ст. $S(X) = X_{(n)}$

Далее предполагаем след. условие на ρ

1) Естественная параметризация

2) $g(x)$, $u(x)$ непр

3) условие равн. сх. инт. по параметру

$$\forall s \quad \forall j \leq k \quad \exists \varphi(x) \quad \forall \theta \qquad \left| g(x)\, u_s^j(x)\, e^{\theta^T u(x)} \right| \leq \varphi(x)$$

степень дифф.

$$\int_{\mathcal{X}} \varphi(x)\, dx \text{ — сходится}$$

$\Downarrow$

[легко посчитать $E\, u(x)$, $D\, u(x)$]

Следствия: 1) $h(\theta)$ непр. дифф k раз

2) $p_\theta(x)$ непр дифф по θ k раз

3) можно менять местам $\frac{\partial}{\partial\theta}$ и $\int$

Утв. $E_\theta\, u(X_1) = \nabla \ln h(\theta) = \left(\frac{\partial \ln h(\theta)}{\partial \theta_j}\right)_j$ ← док-во

$0 \leq D_\theta\, u(X_1) = \nabla\nabla \ln h(\theta) = \left(\frac{\partial^2 \ln h(\theta)}{\partial\theta_j\, \partial\theta_k}\right)_{jk}$ ← ДЗ

▲ $\frac{\partial h(\theta)}{\partial \theta_j} = \frac{\partial}{\partial\theta_j}\int g(x)\, e^{\theta^T u(x)}\, dx = \int g(x)\, u_j(x)\, e^{\theta^T u(x)}\, dx = \left[\cdot h(\theta)/h(\theta)\right] =$

$$= h(\theta) \int \frac{g(x)}{h(\theta)} u_j(x)\, e^{\theta^T u(x)}\, dx = h(\theta)\, E_\theta u_j(x_1)$$

$$\Rightarrow E_\theta u_j(x_1) = \frac{1}{h(\theta)} \cdot \frac{\partial h(\theta)}{\partial \theta_j} = \frac{\partial \ln h(\theta)}{\partial \theta_j}$$

Утв. Θ — выпуклое, то ОМП $\exists!$

▲ $$L_X(\theta) = \underbrace{-n \underbrace{\ln h(\theta)}_{\text{выпуклая}}}_{\text{вогнутая}} + \underbrace{\sum_{i=1}^{n} \ln g(x_i)}_{\text{не зав. от } \theta} + \underbrace{\theta^T \sum_{i=1}^{n} u(X_i)}_{\text{и выпуклая и вогнутая}} \text{ — вогнутая}$$

$\Rightarrow$ Задача оптимизации вогнутого функционала

$\Rightarrow \cap \Rightarrow$ ОМП $\exists!$

ОМП

$$\left[\begin{aligned} &L_X(\theta)' = -n E_\theta X_1 + \sum_{i=1}^{n} u(x_i) = 0 \\ &E_\theta X_1 = \overline{u(x_i)} \end{aligned}\right]$$

есть обман

$N(\theta, \theta^2)$

θ — одномерная

но дост. ст. разм 2

$\Downarrow$

проблемы

Утв. Если Θ — выпуклое, открытое, то выполнено L1 – L9

▲ L5 – L7 выполнены из следствия и предположения

L8 $$i(\theta) = E_\theta \left(\frac{\partial \ln p_\theta(X_1)}{\partial \theta} \right)^2$$

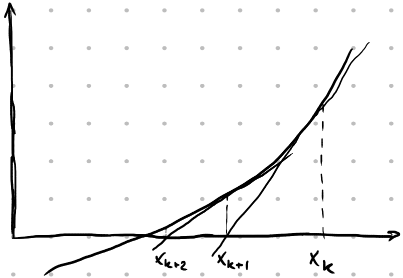
$$\frac{\partial \ln p_\theta(x)}{\partial \theta} = \frac{-1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta} + u(x)$$ [дальше как-то руками помахали]

L9 $$\frac{\partial^3 \ln p_\theta(x)}{\partial \theta^3} = \frac{\partial^2}{\partial \theta^2} \left(-\frac{1}{h(\theta)} \frac{\partial h(\theta)}{\partial \theta} + u(x) \right)$$ — не зависит от x

Следствие: выполнено св-во ОМП

Приближённый поиск ОМП

Метод Ньютона для решения $f(x) = 0$



x_0 — старт

Ур-е касательной в x_k:

$$y = f(x_k) + f'(x_k)(x - x_k) = 0$$

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

Утв. Квадратичная скорость сходимости.

Пусть $X = (X_1 \dots X_n)$ — выборка из $P_\theta \in \mathcal{P}$

$p_\theta(x)$ — плотность

$\Theta \subset \mathbb{R}^d$

θ^* — ОМП

$$\frac{\partial L_X(\theta)}{\partial \theta} = 0$$ — ур-е правдоподобия

Пусть $\hat{\theta}_0$ — начальное приближение

$$\hat{\theta}_{k+1} = \hat{\theta}_k - L''_X(\hat{\theta}_k)^{-1} \, L'_X(\hat{\theta}_k)$$

Теорема (б/д) [L1 – L9] (часть усл. ↑)

Пусть $\hat{\theta}_0$ — ас. норм. оц. θ

наим ас. дисп. в асимпт. подходе

Тогда $\hat{\theta}_1$ — ас. норм. оц. θ с ас. дисп. $i(\theta)^{-1}$

одношаговая оценка

ОМП

$\hat{\theta}_1$ — ас. экв. ОМП : $\sqrt{n}\,(\hat{\theta}_1 - \theta^*) \xrightarrow[n\to\infty]{P_\theta} 0$

▲ Идея для $d = 1$

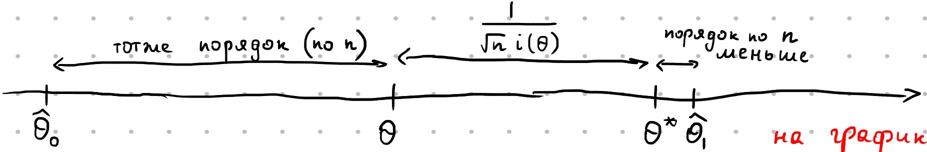
(б/д) $\hat{\theta}_1 - \theta^* = (\hat{\theta}_0 - \theta^*)\,\varepsilon_n(\theta)$, где $\varepsilon_n(\theta) \xrightarrow{P_\theta} 0$

Ас. экв. $\sqrt{n}\,(\hat{\theta}_1 - \theta^*) = \sqrt{n}\,(\hat{\theta}_0 - \theta^*)\,\varepsilon_n(\theta) = \left[\pm\,\theta\,\varepsilon_n(\theta)\sqrt{n}\right] =$

$$= \underbrace{\underbrace{\sqrt{n}\,(\hat{\theta}_0 - \theta)}_{\substack{\downarrow d \\ N(\cdot,\cdot)}} \underbrace{\varepsilon_n(\theta)}_{\substack{\downarrow P_\theta \\ 0}}}_{\substack{\downarrow \\ 0}} - \underbrace{\underbrace{\sqrt{n}\,(\theta^* - \theta)}_{\substack{\downarrow d \\ N(\cdot,\cdot)}} \underbrace{\varepsilon_n(\theta)}_{\substack{\downarrow P_\theta \\ 0}}}_{\substack{\downarrow \\ 0}} \xrightarrow{d} 0 \Rightarrow \xrightarrow{P_\theta} 0$$

Ас. норм.: $\sqrt{n}\,(\hat{\theta}_1 - \theta) = [\pm \theta^* \sqrt{n}] = \underbrace{\sqrt{n}\,(\hat{\theta}_1 - \theta^*)}_{\substack{\downarrow P_\theta \\ 0}} + \underbrace{\sqrt{n}\,(\theta^* - \theta)}_{\substack{\downarrow d \\ N\left(0, \frac{1}{i(\theta)}\right)}} \xrightarrow{d} N\left(0, \frac{1}{i(\theta)}\right)$

Смысл: порядок отклонения θ^* от θ > порядка отклонения $\hat{\theta}_1$ от θ^*



на графике $\theta^* \approx \hat{\theta}_1$

Замечание $L''_x(\hat{\theta}_k)$ можно заменить на $E_\theta L''_x(\hat{\theta}_k) = -n \cdot i(\theta)$

док-во TODO

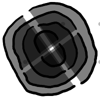
$$\hat{\theta}_{k+1} = \hat{\theta}_k + (n \cdot i(\theta))^{-1} L'_x(\hat{\theta}_k)$$

Пример Cauchy(θ) $\quad i(\theta) = 1/2$ [D3]

$\hat{\mu}$ – а.н.о. с а.д. $\frac{\pi^2}{4} \approx 2{,}47$

ОМП на 23% эффективнее, чем выборочная медиана

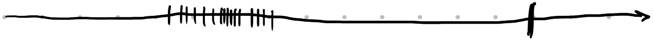
Одношаговая: $\hat{\theta}_1 = \dfrac{\sum\limits_{i=1}^{n} \dfrac{x_1 - \hat{\mu}}{1 + (x_i - \hat{\mu})^2}}{\sum\limits_{i=1}^{n} \dfrac{1 - (x_i - \hat{\mu})^2}{(1 + (x_i - \hat{\mu})^2)^2}}$



Робастность и симметричные распределения

Пусть $X_1 \ldots X_n \sim N(\theta, \sigma^2)$

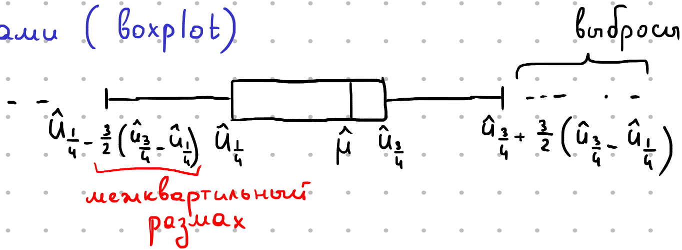
$\hat{\theta} = \bar{X}$ – куча хороших свойств



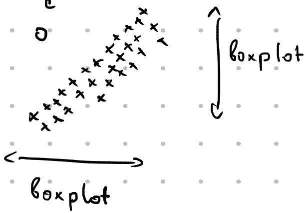
$\Rightarrow \bar{X}$ сместится

Если в данных есть выбросы, то св-ва теряются

Ящик с усами (boxplot)



на boxplot-ах не видно



Робастные оценки — оценки, допускающие отклонения от заданной модели.

Опр Пусть оценка имеет вид $\hat{\theta} = f(\underbrace{X_{(1)} \ldots X_{(n)}}_{\text{вариационный ряд}})$

k_n^* := наим. число k при котором

1) $x_1 \ldots x_{k+1} \to -\infty$, $x_{k+2} \ldots x_n$ – фикс $\Rightarrow f(x_1 \ldots x_n) \to -\infty$

или

2) $x_1 \ldots x_{n-k-1}$ – фикс., $x_{n-k} \ldots x_n \to +\infty$ $\Rightarrow f(x_1 \ldots x_n) \to +\infty$

Асимптотической толерантностью оценки $\hat{\theta}$ наз-ся

$$\tau_\theta = \lim_{n\to\infty} \frac{k_n^*}{n},$$ если он существует

$\shortparallel$

[какую долю плохих значений можно принять и не сломаться]

Пример

- $\hat{\theta}_1 = \overline{X}$ $\qquad \tau_{\overline{X}} = 0$
- $\hat{\mu}$ $\qquad K_n^* = \left\lceil \frac{n}{2} \right\rceil \Rightarrow \tau_{\hat{\mu}} = \frac{1}{2}$

Смысл: наибольшая доля выбросов, которую способна выдержать оценка без смещения в ∞

Будем искать оценки:

1) дост. эфф. (асим. эфф-ть) на $\mathcal{P}$

2) робастные (которые допускают отклонение от распр из $\mathcal{P}$)

Далее рассмотрим класс $\mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$, где

- P_0 имеет плотность $p_0(x)$ – чёт., непр, носитель $(-c; c)$
  $0 < c \leq +\infty$

- $\theta \in \Theta = \mathbb{R}$ – параметр сдвига $p_\theta(x) = p_0(x - \theta)$

① Усечённое среднее $\alpha \in (0; \frac{1}{2})$

$k = \lceil \alpha n \rceil$ по-моему тут $\lfloor \alpha n \rfloor$ надо

Ус. ср. порядка α наз-ся $\overline{X}_\alpha = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} X_{(i)}$

$\alpha = 0 \quad \overline{X}_\alpha = \overline{X}$

$\alpha = \frac{1}{2} \quad \overline{X}_\alpha \approx \hat{\mu}$

Теорема $X_1 \ldots X_n$ — выборка из $P_\theta \in \mathcal{P}$

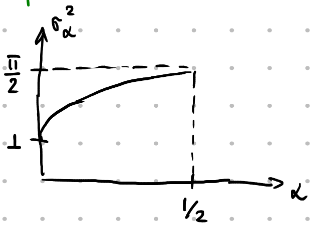
Тогда $\sqrt{n}\,(\overline{X}_\alpha - \theta) \xrightarrow{d_\theta} N(0, \sigma_\alpha^2)$

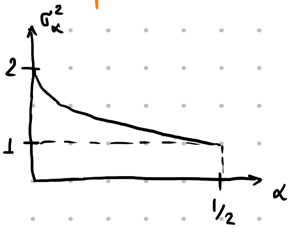
$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left(\int_0^{u_{1-\alpha}} x^2 p_\theta(x)\,dx + \alpha u_{1-\alpha}^2\right)$$

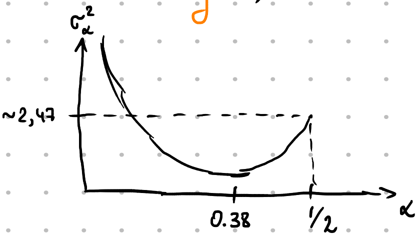
p_θ — плотность P_θ

$u_{1-\alpha}$ — $(1-\alpha)$ квантиль P_θ

Примеры $N(\theta, 1)$ Laplace (θ) Cauchy (θ)







Пример $N(\theta, 1)$

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

$$= \frac{1}{\sigma_\alpha^2}$$

$\alpha = \frac{1}{2}$, то защита от выбросов 12,5 %

потеря эф-ти 6 %

Теорема Пусть $D_\theta X_1 < +\infty$

Тогда $ARE_{\overline{X}_\alpha, \overline{X}} \geqslant (1-2\alpha)^2$

▲ $\frac{1}{2} D_\theta X_1 = \frac{1}{2}\int_{-\infty}^{\infty} x^2 p_0(x)\,dx = \int_0^{+\infty} x^2 p_0(x)\,dx = \int_0^{u_{1-\alpha}} x^2 p_0(x)\,dx + \int_{u_{1-\alpha}}^{+\infty} x^2 p_0(x)\,dx \geqslant$

т.к. D не зав. от сдвига

по т. выше

$$\geqslant \int_0^{u_{1-\alpha}} x^2 p_0(x) + u_{1-\alpha}^2 \alpha = \frac{(1-2\alpha)^2}{2}\sigma_\alpha^2$$

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

② Медиана средних Уолша

$$Y_{ij} = \frac{X_i + X_j}{2} \quad 1 \le i \le j \le n \quad - \text{ средние Уолша}$$

Мед. ср. Уолша $W :=$ медиана Y_{ij}

Теорема (б/д) $X_1, \dots, X_n$ – выборка из $P_\theta \in \mathcal{P}$

Тогда $\sqrt{n}\,(W - \theta) \xrightarrow{d_\theta} N(0, \sigma_W^2)$

где $$\sigma_W^2 = \frac{1}{12\left(\int\limits_{\mathbb{R}} p_0^2(x)\,dx\right)^2}$$

Утв. Для $P_\theta \in \mathcal{P}$ $ARE_{w,\bar{x}} \geq \frac{108}{125} \approx 0.864$

Утв. $\tau_w \approx 0.293$, т.е. оценка устойчива при $\approx 29\%$ выбросов