Задача мат. статистики

$$\begin{tikzcd} \text{Природа случ. явл.} \arrow[r, bend right, "\text{терВер}"'] & \text{характеристика} \arrow[l, bend right, "\text{мат статистика}"'] \end{tikzcd}$$

<u>Ситуация 1</u> Из N студентов на лекции ходят M

$\boxed{\text{ТВ}}$ P(среди n человек m придет на лекцию) = ?

N, M - известны

$\boxed{\text{МС}}$ <u>M неизв</u>. Среди n случ. человек на лек m

Оценим <u>M</u>

<u>Ситуация 2</u>

$\xi \in N(a, \sigma^2)$

$\boxed{\text{ТВ}}$ $E\xi$, $P(\xi \in [0,1])$, $F_\xi(x)$, ... -?

$\boxed{\text{МС}}$ $X_1, \dots, X_n$ — независимые реализации ξ.

Оценить a, σ^2

„Формальная" модель

$\exists X_1, \dots, X_n$ — численные хар-ки n-кратного повторения некоторого сл. явл. — независимые случ. величины ξ

Задача. По $x_1, \dots, x_n$ высказать нек. суждение о распред. P (пример убрать шум $\sim (0, \sigma^2)$ усреднением)

Опр.
Случ. вектор $X = (X_1, \dots, X_n)$ наз-ся выборкой размера n

Замечание
X — вектор независимых одинак. распред. величин.

! $(1, 2, \dots, 10)$ — не выборка, а её реализация

Задача точечного оценивания

$\exists$ $X = (X_1, \dots, X_n)$ — выборка из неизв. распр.
$P \in \mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$

Задача: найти θ или $\tau(\theta)$

$(X_i \in \mathcal{X})$

Опр.: Статистикой $S: \mathcal{X}^n \to E$ — ф-ия от выборки (**замеч.**: S — случ. в.)

Опр.: оценка $\hat{\theta}$ — статистика при $E = \Theta$

Замеч: $S(X)$ — случ. вел.
$S(x)$ — реализация статистики

Примеры статистик:

① Выборочные хар-ки:

$$\overline{g(X)} := \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

частн. случ. $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ — выборочное среднее

$\overline{X^k}$ — выборочный k-тый м-т

② Ф-ии от выб. х-к

$$S^2 = \overline{X^2} - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$ — выборочная дисп.

③ Порядковые статистики

Упоряд. X по возр $\left(X_{(1)}, X_{(2)}, \dots, X_{(n)}\right)$

$X_{(1)}$ — min, $X_{(n)}$ — max

$X_{(k)}$ — k-ая порядковая статистика

Пример: $U[0,\theta] \leftarrow \underbrace{X_{(n)}}_{\theta \text{ оценим этим}}$

З1 $X = (X_1, \dots, X_n)$ — выборка

- $E\overline{X} = E\frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{лин}}{=\!=} \frac{1}{n}\sum_{i=1}^{n} EX_i = EX_i$ ← любого т.к. н.о.р.св
- $D\overline{X} = D\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n^2} D\sum_{i=1}^{n} X_i \underset{\text{незав}}{=\!=} \frac{DX_i}{n}$ ← любого

Примеры семейств распред.

① Физический эксперимент

$x_i = \underset{\text{конст.}}{a} + \underset{\text{шум}}{\varepsilon}$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

$x_i \sim \mathcal{N}(a, \sigma^2)$

$\mathcal{P} = \{$ м-во таких штук $\}$

Оценить a : $\hat{\theta} = \overline{X}$

② Логистическая регрессия

$\{(x_i, y_i)\}_{i=1}^{n}$

дано свыше   случ. велич.

$x \in \mathbb{R}^d \longmapsto P_\theta(y = 1 \mid X) = \sigma(x^T \theta)$

$\mathcal{P} = \{ \mathrm{Bern}(\sigma(x^T\theta)) \mid \theta \in \mathbb{R}^d \}$

$\hat{\theta} : L(X, y, \theta) \longrightarrow \min_\theta$

③ Гауссовская регрессия

$y = X\theta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma \underbrace{I_d}_{\text{Ед. м-ца}})$

отклик

признаки (даются свыше)   параметры   случ (шум)

$y \sim \mathcal{N}(X\theta, \sigma^2)$

$\mathcal{P} = \{ \mathcal{N}(X\theta, \sigma^2 I_d), \ \theta \in \mathbb{R}^d \}$

$\hat{\theta} = (X^T X)^{-1} X^T y$ — МНК оценка

$\hat{\theta} = (X^T X)^{-1} X^T y + \lambda I_d$

с регуляризацией (Ridge, Lasso)

взвешенная

Виды сх-тей

$\xi_n \to \xi$ п.н., P, d

Законы

ЗБЧ, УЗБЧ, ЦПТ.

$\mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$

Для каждого P_θ будем писать E_θ, D_θ

$X = (X_1, \dots, X_n) \in P \in \mathcal{P}$

Опр. Оценка $\hat{\theta}(X)$ наз-ся несмещенной оценкой $\tau(\theta)$, если

$\forall \theta \in \Theta \hookrightarrow E_\theta \tilde{\theta} = \tau(\theta)$

$\boxed{B_1}$ $\overline{X}$: $E_\theta \overline{X} = E_\theta X_1$

$\downarrow$ оценивает

$E_\theta X_1$

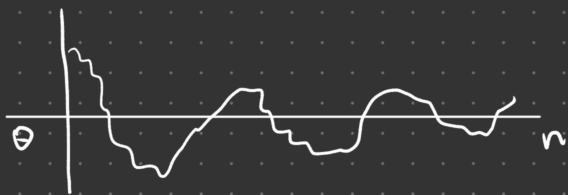
$\Rightarrow$ несмещ.

$\boxed{B_2}$ $\hat{\theta} = X_1$: $E_\theta \hat{\theta} = E_\theta X_1$

$\Rightarrow$ $\hat{\theta}$ несмещ оц. $E_\theta X_1$

Типо ЗБЧ

<u>смысл:</u> в среднем то, что нужно.



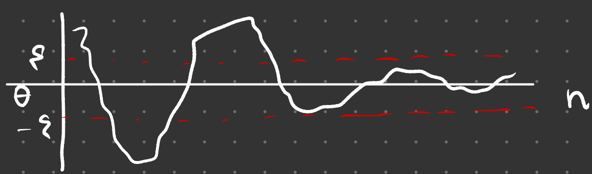
Асимптотические св-ва оценок:

<u>опр.</u> <u>состоятельность</u>

для п.н.

Оценка $\hat{\theta}(X)$ наз-ся (сильно) состоятельной оценкой $\tau(\theta)$, если $\hat{\theta}(X) \xrightarrow[(\text{п.н.})]{P} \tau(\theta)$ $\forall \theta \in \Theta$

<u>смысл:</u> при больших n не будет сильного отклонения



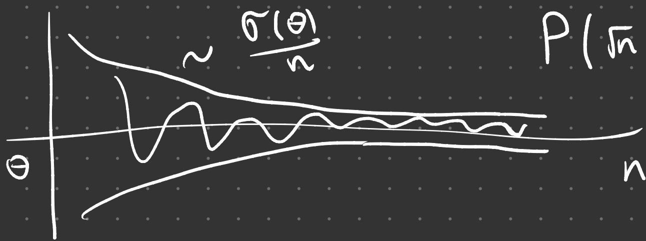
Типо УЗБЧ

Опр. оценка $\hat{\theta}(X)$ наз-ся асимп-ки нормальной оценкой ~~оценкой~~ θ с матр ковариаций $\Sigma(\theta)$, если

$$\sqrt{n}\,(\hat{\theta}-\theta) \xrightarrow{d_\theta} \mathcal{N}(a, \Sigma(\theta))$$

1D : $\Sigma(\theta) = \sigma^2(\theta)$ — асимптотич. дисп.

при больш. n $\quad \sqrt{n}\,\frac{\hat{\theta}-\theta}{\sigma(\theta)} \approx \mathcal{N}(0,1)$

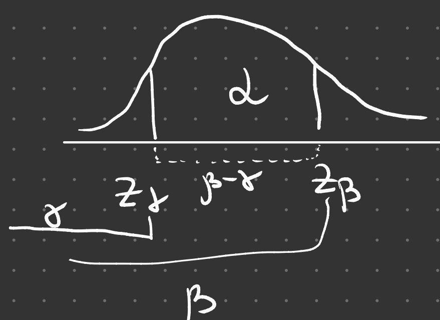


$$P\left(\sqrt{n}\,\frac{|\hat{\theta}-\theta|}{\sigma(\theta)} < 2\right) \approx 0{,}95$$

Типо ЦПТ

Квантили $\quad \alpha \in (0,1)$, то квантиль u_α :

$$u_\alpha := \min\{u : F(u) \geq \alpha\}$$



Для $\mathcal{N}(0,1)$

$$P\left(z_{\frac{1-\alpha}{2}} < \xi < z_{\frac{1+\alpha}{2}}\right)$$

$$\beta = \frac{1+\alpha}{2}$$

$$\gamma = \frac{1-\alpha}{2}$$

$$\sqrt{n}\,\frac{\hat{\theta}-\theta}{\sigma(\theta)} \approx \mathcal{N}(0,1)$$

$$P\left(\underset{\substack{\| \\ -z_{\frac{1+\alpha}{2}}}}{z_{\frac{1-\alpha}{2}}} < \sqrt{n}\,\frac{\hat{\theta}-\theta}{\sigma(\theta)} < z_{\frac{1+\alpha}{2}}\right) = \alpha$$

$$\Downarrow$$

$$P\left(\sqrt{n}\,\frac{|\hat{\theta}-\theta|}{\sigma(\theta)} < z_{\frac{1+\alpha}{2}}\right) = \alpha$$

<u>Пример:</u> $X_1, \ldots, X_n \sim Laplace(\theta)$

с $p_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$

нужно – оценить θ (а.н.о)

– АДИ θ

1) $\hat{\theta} = \overline{X}$ – несмещ, сильно сост. по УЗБЧ

$\sqrt{n}\left(\frac{\hat{\theta}-\theta}{\sigma(\theta)}\right) \xrightarrow{d} \mathcal{N}(0,1)$ по ЦПТ $\Rightarrow$ а.н.о

$\sigma^2(\theta) = \underset{\substack{\| \\ 2}}{D_a X_1}$ из ЦПТ

АДИ (асимпт. довер. интервал)

а.н.о. (ас. нормальная оценка)

АДИ ур-ня дов. α

$$\theta \in \left(\bar{X} \pm \frac{z_{\frac{1+\alpha}{2}}}{\sqrt{n}} \sqrt{\Sigma} \right)$$

Утв. сильная сост. $\Rightarrow$ сост | Отличается от сх-ти случ. величин.

А.Н.О $\nearrow$

Th (наследование св-в)

1) Пусть $\hat{\theta}$ — (сильно) сост. оценка θ, $\tau(\theta)$ непрер. на Θ. Тогда $\tau(\hat{\theta})$ — (сильно) состоятельная оценка $\tau(\theta)$

2) $\hat{\theta}$ — А.Н.О с $\Sigma(\theta)$, $\tau(\theta)$ — непр. диф. на Θ

Тогда $\tau(\hat{\theta})$ — а.н.о. $\tau(\theta)$ с $\Sigma'(\theta)$

$$\Sigma'(\theta) = D(\theta)\,\Sigma(\theta)\,D(\theta), \quad D(\theta) = \frac{\partial \tau}{\partial x}\bigg|_{x=\theta}$$

Пример:

$X_1, \dots, X_n \sim \exp(\theta)$

$E_\theta X_1 = \frac{1}{\theta}$

$\hat{\theta} = \bar{X}$ А.Н. оценка $\frac{1}{\theta}$ — несмещ., А.Н.О., сост

Th

$\Rightarrow \quad \tau(x) = \frac{1}{x}$ – непр. диф. на $\{x>0\}$

$$\mathcal{D}(\theta) = -\frac{1}{x^2}\Big|_{x=\frac{1}{\theta}} = -\theta^2$$

$$\Sigma'(\theta) = (-\theta^2)^2 \frac{1}{\theta^2} = \theta^2$$

$\tau(\hat{\theta}) = \frac{1}{\overline{x}}$ а.н.о. $\frac{1}{1/\theta} = \theta$ с $\sigma^2(\theta) = \theta^2$

$$\Rightarrow \quad \theta \in \left(\frac{1}{\overline{X}} \pm \frac{z_{\frac{1+\alpha}{2}}}{\sqrt{n}}\theta\right)$$

интервал зависит от θ, что делать?