- выборка - набор НЗ одинаково распред. случ. велич.
- в статистике нет абсолютной точности

Основная задача статистики

Природа случ. явлений $\underset{\text{Стат}}{\overset{\text{ТВ}}{\rightleftarrows}}$ характеристики

Пр. 1 есть N чел., на лекцию пришли M чел.

ТВ — выб. n чел. P(среди случ. n есть ровно m на лекции) - ?

статистика - неизвестно M, выбрали n; присутствует из них m → чему равно M?

Пр. 2. Пусть $\xi \sim N(1,4)$ → ТВ: мат. ожид. $E X = 1$, $P(\xi \in [0,1])$ - ?

стат: $\xi \sim N(a, \sigma^2)$, $\underbrace{a, \sigma}_{\text{найти}}$ - неизв., но есть набор $X_1, \dots, X_n$ НЗ реализ. $\xi$.

**Основная задача статистики**

по НЗ реализациям эксперимента высказать суждение о природе этого эксперимента

## ① Статистики и оценки

пр.: люди тратят какое-то кол-во денег на покупки в интернете в неделю. 1 неделя — 1 выборка

def • Выборка - набор независимых (совм. в т.ч. параметры) одинаково распределённых случайных величин

not.: $X_1, \dots, X_n$

$X_i : \Omega \to \mathbb{R}$

def • Реализация выборки — в.р $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, где $x_i = X_i(\omega)$ → эл. исход

def • Статистика — функция от выборки

def • Пусть $\exists$ семейство распределений $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$.

$\theta$ - параметр семейства $\mathcal{P}$;

$\Theta$ - мн.во параметров

напр., $\mathcal{P} = \{N(a, \sigma^2) \mid a \in \mathbb{R}, \sigma \in \mathbb{R}_+\}$ → здесь $\theta = (a, \sigma)$, $\Theta = \mathbb{R} \times \mathbb{R}_+$

Статистика $\hat{\theta}(X)$ наз. оценкой, если её мн.во знач. вложено в $\Theta$.

число $-x_i^2$ не явл. оценкой для $\sigma$, т.к. $\sigma \in \mathbb{R}_+$

Примеры статистик.

1) $\overline{g(X)} \overset{def}{=} \frac{1}{n} \sum_{i=1}^{n} g(X_i)$ - выборочная х-ка ф-ии g

$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ - выборочное среднее

$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ – выборочный момент k-го порядка

2) $S^2 \overset{def}{=} \overline{X^2} - \overline{X}^2$ – выборочная дисперсия

$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

3) Вариационный ряд: Отсорт. выборку по возраст.:

$X_{(1)}, \dots, X_{(n)}$; $X_{(k)}$ наз. k-й порядковой статистикой

Примеры реализац. выб.

$x = (2, 1, 5) \to \overline{x} = \frac{2+1+5}{3} = \frac{8}{3}$; $S^2 = 10 - \left(\frac{8}{3}\right)^2 = \frac{26}{9}$

вариц. ряд – 1, 2, 5

$x = (0, 1, -1) \to \overline{x} = 0$, $S^2 = 2/3$, вар. ряд: $-1, 0, 1$

$X_{(3)} = \max(X_1, X_2, X_3) \to X_{(3)}(\omega_1) = \max(2, 1, 5) = 5$

$X_{(3)}(\omega_2) = \max(0, -1, 1) = 1$

Упр. $E\overline{X}, D\overline{X}$ – ?

выборка – одинак. нар. $\to E$

$E\overline{X} = \frac{1}{n}\sum_{i=1}^{n} EX_i = \frac{1}{n} n EX_1 = EX_1$

$D\overline{X} = D\frac{1}{n}\sum_{i=1}^{n} X_i \underset{X_i - \text{НЗ}}{=} \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2} n DX_1 = \frac{1}{n} DX_1$ } при усред E сохр, D ↓ в n р.

② Свойства оценок

Обозначения: $P_\theta$, $E_\theta$, $D_\theta$; $P_\theta$ - п.н., $d_\theta$ } св-ва в предп. истинности $\theta$

def. Оценка $\hat{\theta}(x)$ наз. несмещённой оценкой пар-ра $\theta$, если $\forall \theta \in \Theta \to E_\theta \hat{\theta} = \theta$

└ предп. истинность – выборка

def$_1$. Пусть $X_1, \dots, X_n$ – выборка $\to$ оценка $\hat{\theta}_n(X_1, \dots, X_n)$ наз. состоятельной, если $\forall \theta \in \Theta \to$ сход. по вер-ти $\hat{\theta}_n \xrightarrow{P_\theta} \theta$, т.е.

$\forall \theta \in \Theta\ \forall \varepsilon > 0 \to P_\theta(\|\hat{\theta}_n - \theta\| > \varepsilon) \xrightarrow[n\to\infty]{} 0$

def$_2$. Пусть $X_1, \dots, X_n$ – выборка $\to$ оценка $\hat{\theta}_n(X_1, \dots, X_n)$ наз. сильно состоят., если $\forall \theta \in \Theta$: $\hat{\theta}_n \xrightarrow{P_\theta\text{-п.н.}} \theta$, т.е. $\forall \theta \in \Theta \to P(\lim_{n\to\infty} \hat{\theta}_n = \theta) = 1$

def$_3$. Пусть $X_1, \dots, X_n$ – выборка $\to$ оценка $\hat{\theta}_n(X_1, \dots, X_n)$ наз. асимптотически (а.н.о.) нормальной, если $\forall \theta \in \Theta$: $\underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}_{\text{распр. этого сход. сюда}} \xrightarrow{d_\theta} N(0, \Sigma(\theta))$, где $\Sigma(\theta)$ – асимп. матр. ковариаций разм. $\dim\Theta \times \dim\Theta$.

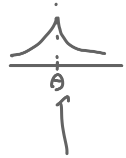
если $\dim\Theta = 1$, то $\sigma^2(\theta)$ - ас. дисп.

Смысл св-в

1) отклонение $\to 0$

ЗБЧ. $\{\xi_n\}$ – выборка: $E|\xi_i|<+\infty \xrightarrow{п.н.} E\xi_1$

ЦПТ. $\{\xi_n\}$ – выборка: $D\xi_i<+\infty$:
$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i - E\xi_1\right)\xrightarrow{d} N(0, D\xi_1)$$



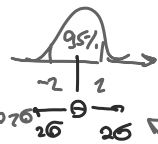

Пр. $X_1,\dots,X_n$ – выборка из распред. Лапласа плотностью $p_\theta(x)=\frac{1}{2}e^{-|x-\theta|}$

$\to E_\theta X_1=\theta,\ D_\theta X_1=2$; ЗБЧ: $\overline{X}\xrightarrow{P_\theta\text{-п.н.}}\theta\ (\forall\theta\in\Theta)\to$

$\to \hat\theta=\overline{X}$ – несмещ., сост. и сильно сост.

ЦПТ $\to \sqrt{n}(\overline{X}-\theta)\xrightarrow{d} N(0,2) \to \hat\theta=\overline{X}$ – а.н.о. $\theta$ с ас. дисп. $\sigma^2(\theta)=2$

приближ.: $\overline{X}\sim N(\theta, 2/n)$




; $\overline{X}\in(\theta\pm 2\sqrt{2/n})$ ⇔ $\theta\in(\overline{X}\pm 2\sqrt{2/n})$

$2\sigma$

реализация с.в. лежит в интервал. от $-2\sigma$ до $2\sigma$ с вер. 95%

$d=\dim\Theta$

применим.

доверительный интервал

Утв. 1) $\hat\theta$ – (сильно) сост. оценка $\theta$ и $h:\mathbb{R}^d\to\mathbb{R}^k$ непр. на $\Theta$, то $h(\hat\theta)$ – (сильно) сост. оценка $h(\theta)$

2) сильн. сост. $\Rightarrow$ состоят.
ас. норм. $\Rightarrow$ состоят.

3) Дельта-метод

ас. матр. ков.

$\exists\ \hat\theta$ – ано $\theta$ с $\Sigma(\theta)$, и $h(\theta)\in C^1$ на $\Theta$ $\Rightarrow$

$\Rightarrow h(\hat\theta)$ – а.н.о. $h(\theta)$ с а.м.к. $D(\theta)\Sigma(\theta)D(\theta)^T$, где $D(\theta)$ – матр. производных:
$$D(\theta)=\left(\frac{\partial h_k(\theta)}{\partial\theta_j}\right)_{kj}$$

## ③ Доверительные интервалы.

def. Пара статистик $(T_1(X), T_2(X))$ наз. доверит. интерв. для $\theta$ с ур-нем доверия $\alpha$, если $\forall\theta\in\Theta \hookrightarrow P_\theta(\theta\in(T_1(x),T_2(x)))\geq\alpha$.

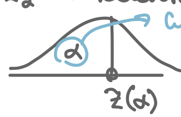
Д.и. наз. асимптотич., если $\liminf P(\theta\in(T_1(x),T_2(x)))\geq\alpha$.

∘ $\overline{X}-2\sqrt{2/n}<\theta<\overline{X}+2\sqrt{2/n}$ интервал накрывает $\theta$ (фикс.) с вер-ю 95%

$n=200$, $\overline{X}=1$ $\Rightarrow$ $0{,}8<\theta<1$ – реализация Д.и.

∘ Квантиль.

$z_\alpha$ – $\alpha$-квантиль $N(0,1)$

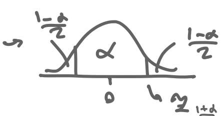



слева от отметки $z(\alpha)$

Метод построения ас. Д.и.

1. Взять а.н.о. $\hat\theta_n$ – ано $\theta$ с ас. дисп. $\sigma^2(\theta)$
2. $\sigma(\theta)$ непр. $\Rightarrow$ $\hat\sigma=\sigma(\hat\theta)$ – сост. оценка $\theta$
3. $\Rightarrow$ $\sqrt{n}\,\frac{\hat\theta-\theta}{\hat\sigma}\xrightarrow{d_\theta} N(0,1)$

$\Rightarrow$




$z_{\frac{1+\alpha}{2}}$ – слева – вер. масса $\frac{1+\alpha}{2}$

$\Rightarrow$ интервал: $(\hat\theta\pm z_{\frac{1+\alpha}{2}}\hat\sigma/\sqrt{n})$

Задача 1. $X_1,\dots,X_n\sim Exp(\theta)$. Найд. разные оценки $\theta$

$\Rightarrow E_\theta X_1=1/\theta$, $D_\theta X_1=1/\theta^2$. ЗБЧ $\Rightarrow$ $\overline{X}\to 1/\theta$ $\Rightarrow$ $\overline{X}$ – (сильно) сост. оц. $1/\theta$.

применим непр. при $x>0$ ф-ю $h(x)=1/x$ $\Rightarrow$ $1/\overline{X}=h(\overline{X})$ – (сильно) сост. оценка $h(1/\theta)=\theta$

ЦПТ $\Rightarrow$ $\sqrt{n}(\overline{X}-1/\theta)\xrightarrow{d_\theta} N(0, 1/\theta^2)$

↪ по def а.н.о. $\bar{X}$ – ас. норм. оценка $1/\theta$ с ас. дисп. $1/\theta^2$

- по дельта-методу с ф-ей $h(x)=1/x$ $\Rightarrow$ $h(\bar{X})=1/\bar{X}$ – а.н.о. $h(1/\theta)=\theta$,
  с ас. дисп. $\frac{1}{\theta^2}\cdot\left[\left(\frac{1}{x}\right)'\right]^2\Big|_{x=1/\theta} = \theta^2$
- интервал: $\left(\frac{1}{\bar{X}} \pm z_{\frac{1+\alpha}{2}}\frac{1}{\bar{X}^2\sqrt{n}}\right)$