Бутстрэп

Эфрон 1979

n большое - работает

kurtosis

Med $[Med_L \quad Med_R]$

$H_0: Med = 5$

$H_A: \neq$

Природа:

$$\underset{\text{неизв. р.-е.}}{F_\theta(x)} \longrightarrow \overset{\text{выборка}}{x_1, x_2, \dots, x_n} \longrightarrow \hat\theta$$

Бутстрап:

$$\underset{\text{эмп. р.-е}}{\hat F_n(x)} \longrightarrow \overset{\text{бутстрап - выборка}}{x_1^*, \dots, x_n^*} \longrightarrow \hat\theta^*$$

$B < n^n$

Природа	Наш аналог
$F_\theta(x)$	$\hat F_n(x)$
θ	$\hat\theta$
$\hat\theta$	$\hat\theta^*$
$\hat\theta - \theta$	$\hat\theta^* - \hat\theta$

Перцентильный Д.И.:

$x_1, \dots, x_n$

выборки с повторением

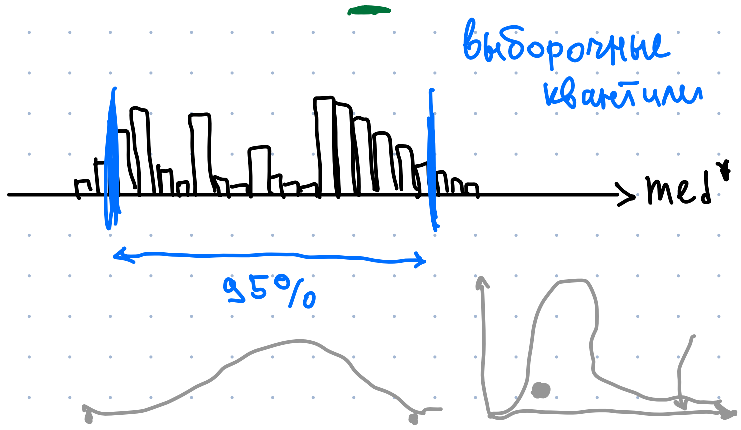
$$\begin{array}{cccc} x_1^* & x_1^* & \dots & x_1^* \\ x_2^* & x_2^* & \dots & x_2^* \\ \vdots & \vdots & & \vdots \\ x_n^* & x_n^* & \dots & x_n^* \\ \downarrow & \downarrow & & \downarrow \\ med_1^* & med_2^* & \dots & med_B^* \end{array}$$

сэмплируем: $x_1^*, \dots, x_n^*$

считаем: $\hat\theta^*$

распределение: $\hat\theta_1^*, \dots, \hat\theta_B^*$

интервал: $[\hat\theta^*_{\frac{\alpha}{2}}; \hat\theta^*_{1-\frac{\alpha}{2}}]$



Если р.-е. скошенное, тогда такой д.и. будет плохим (он съедет и $\gamma < 0.95$)

Упражнение

1, 3 — выборка

1, 1	→	$\bar{x}^* = 1$
1, 3	→	$\bar{x}^* = 2$
3, 1	→	$\bar{x}^* = 2$
3, 3	→	$\bar{x}^* = 3$

$\bar{x}^*$	1	2	3
$\mathbb{P}(\dots)$	1/4	1/2	1/4

2^2

n^n

Обратный перцентильный д.и.

сэмплируем: $x_1^*, \dots, x_n^*$

считаем: $\hat{q}^* = \hat{\theta}^* - \hat{\theta}$ центрировали

распределение: $\hat{q}_1^*, \dots, \hat{q}_B^*$

интервал: $[\hat{\theta} - \hat{q}^*_{1-\frac{\alpha}{2}}; \hat{\theta} - \hat{q}^*_{\frac{\alpha}{2}}]$

$$q_L \le \hat{\theta} - \theta \le q_R$$
$$-q_R \le \theta - \hat{\theta} \le -q_L$$
$$\hat{\theta} - q_R \le \theta \le \hat{\theta} - q_L$$

Всё ещё съезжает, но уже с другой стороны

Это уже будет работать норм

t-перцентильный д.и.

сэмплируем: $x_1^*, \dots, x_n^*$

считаем: $t^* = \dfrac{\hat{\theta}^* - \hat{\theta}}{se(\hat{\theta}^*)}$

$se(x_1^*, \dots, x_n^*)$ откуда взять?

распределение: $t_1^*, \dots, t_B^*$

$se(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$

интервал: $[\hat{\theta} - t^*_{1-\frac{\alpha}{2}} \cdot se(\hat{\theta}); \hat{\theta} - t^*_{\frac{\alpha}{2}} \cdot se(\hat{\theta})]$

Inception — Бутстрэп в бутстрэпе

„начало"

$x_1, \dots, x_n$ $\quad\hat\theta$

$x_1^*, \dots, x_n^*$ $\qquad$ $x_1^*, \dots, x_n^*$ $\quad\dots\quad$ $x_1^*, \dots, x_n^*$

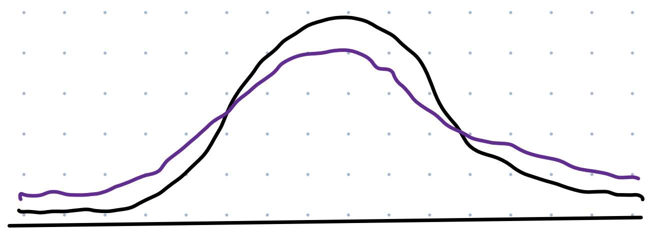
$\hat\theta_1^*$ $\qquad$ $\underbrace{\hat\theta_2^*}_{se(\hat\theta_2^*)}$ $\quad\dots\quad$ $\underbrace{\hat\theta_B^*}_{se(\hat\theta_B^*)}$

$$t_1^* = \frac{\hat\theta_1^* - \hat\theta}{se(\hat\theta_1^*)}$$

$x_1^{**}, \dots, x_n^{**}$ $\quad$ $x_1^{**}, \dots, x_n^{**}$ $\quad\dots\quad$ $x_1^{**}, \dots, x_n^{**}$

$\hat\theta_{11}^{**}$ $\quad$ $\hat\theta_{12}^{**} \dots \hat\theta_{1B}^{**}$

$se(\hat\theta_1^*)$



Проверка гипотез:

$H_0: \ \theta = \theta_0$

$H_A: \ \theta > \theta_0$

$x_1, \dots, x_n$

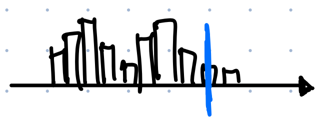
одна выборка

$$t = \frac{\hat\theta - \theta_0}{se(\hat\theta)}$$

генерим при верности H_0

θ_0

$$t^* = \frac{\hat\theta^* - \hat\theta}{se(\hat\theta^*)}$$



t_{obs} vs $t^*_{1-\alpha}$

А если две выборки?

$x_1, \dots, x_n$ $\bar{x}$ $\hat{\sigma}_x^2$ $H_0: \mu_x = \mu_y$

$y_1, \dots, y_m$ $\bar{y}$ $\hat{\sigma}_y^2$ $H_A: \mu_x > \mu_y$

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{\sigma}_x^2}{n} + \frac{\hat{\sigma}_y^2}{m}}}$$

$$z^* = \frac{\bar{x}^* - \bar{y}^*}{\sqrt{\frac{\hat{\sigma}_x^{*2}}{n} + \frac{\hat{\sigma}_y^{*2}}{m}}}$$

H_0 верна

$$\bar{h} = \frac{x_1 + \dots x_n + y_1 + \dots + y_m}{n+m}$$

$$\left.\begin{aligned} x_i' &= x_i - \bar{x} + \bar{h} \\ y_i' &= y_i - \bar{y} + \bar{h} \end{aligned}\right]$$ Отсюда бутстрапирую

$$z_{obs} \text{ vs } z^*_{1-\frac{\alpha}{2}}$$