

Бутстреп

- Эфрон 1979
- n большое - асимптотическая техника (не связана с ЦПТ)

• мед

Реальность:

$F_{\theta}(x)$
неизв. р.-е.

выборка
 x_1, \dots, x_n
 \downarrow
 $\hat{\theta}$
статистика

Бутстреп

$\hat{F}_n(x)$
эмп. р.-е.

бутстреп-выборка

x_1^*, \dots, x_n^*

\downarrow
 $\hat{\theta}^*$
бутстр.-стат.

Упражнение

1, 3
выборка

выборки
с повтор.

Хоту построить

бутстр.-р.-е

для \bar{x}

1, 1

$\bar{x} = 1$

1, 3

$\bar{x} = 2$

$2^2 = 4$

3, 1

$\bar{x} = 2$

3, 3

$\bar{x} = 3$

1 3 3 2
1 1 3 3

\bar{x}	1	2	3
$P(\bar{x}=1)$	$1/4$	$1/2$	$1/4$

$1/2^2$ $1/2^2$

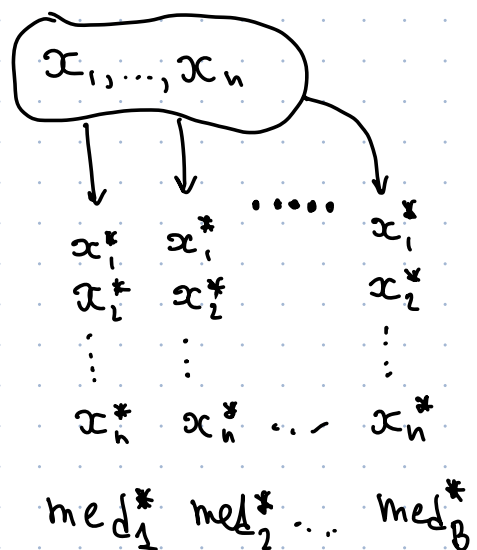
Всего
 n^n вариантов
выборок

можно было бы как с параметр. критериями
 строить табличные р.-я и делать выводы по ним,
 но это долго

⇒ Симуляции (выборки с повторениями)

Будем делать в случ. выборок размера n

① Эфров г.ч. (Терцентильный г.ч.)

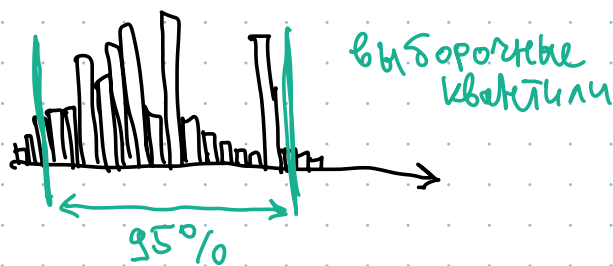


сэмплируем: x_1^*, \dots, x_n^*

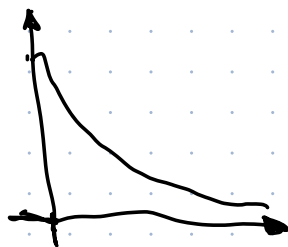
считаем: $\hat{\theta}^*$

распределение: $\hat{\theta}_1^*, \dots, \hat{\theta}_n^*$

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



Pl.hist(med)



Если р.-е. выборки
 смещенные, тогда
 Такой г.ч.

Будет усиливаться
 смещение
 (от тупо следует)

② Обратный перу. г.ч.

$$F_{\theta}(x) \quad \hat{F}(x)$$

$$\theta$$

$$\hat{\theta}$$

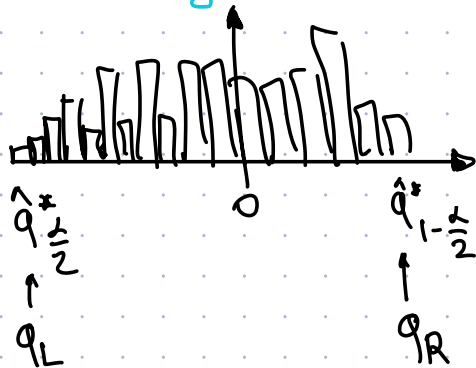
$$\hat{\theta}$$

$$\hat{\theta}^*$$

$$\hat{\theta} - \theta$$

$$\hat{\theta}^* - \hat{\theta}$$

Центрирование



сэмплируем: x_1^*, \dots, x_n^*

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

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_R

q_L

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

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

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

г.ч. и.б.
несмещен-
тривный

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

$\frac{\alpha}{2}$

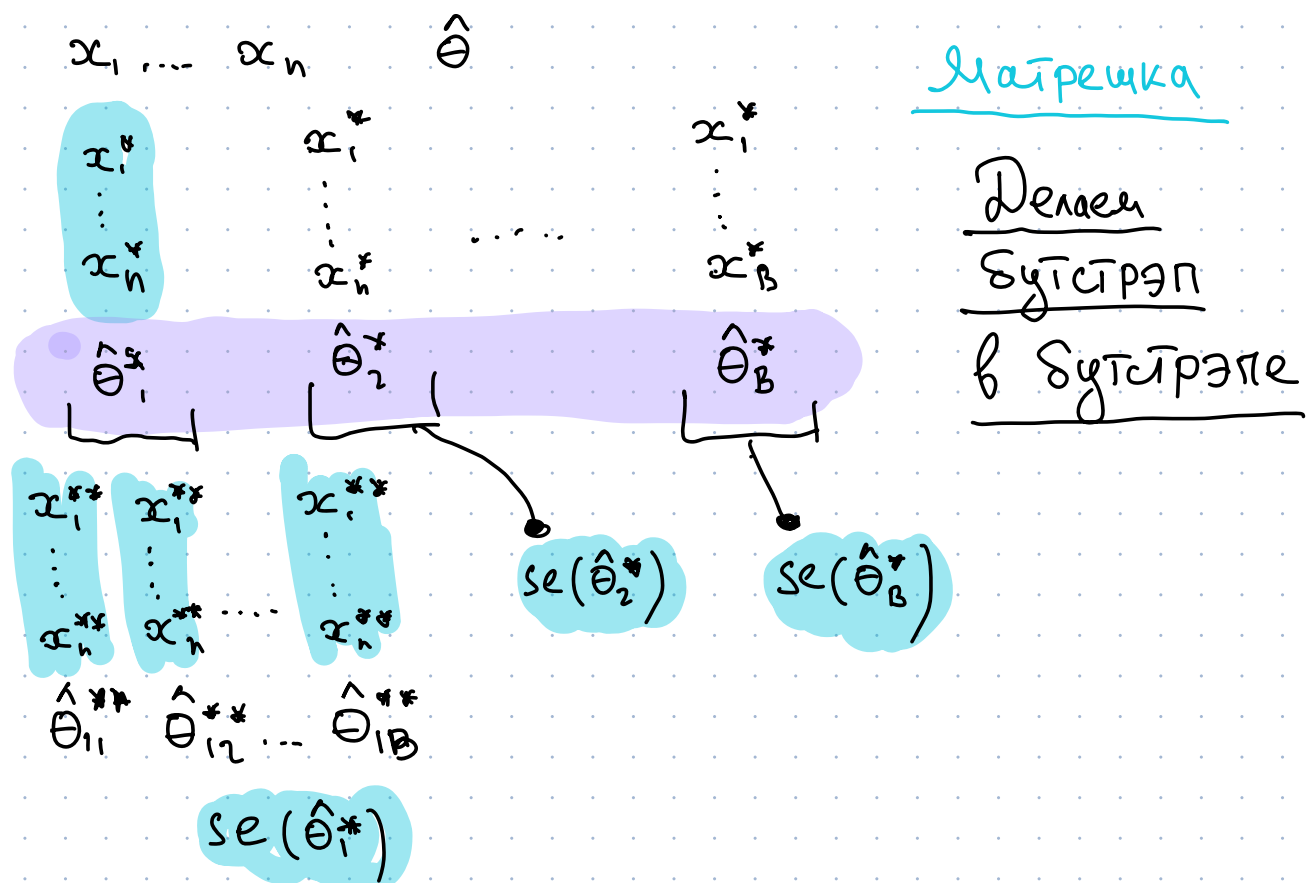
③ t-процентильный г.ч.

сэмплируем: x_1^*, \dots, x_n^*

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

распределение: t_1^*, \dots, t_B^*

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



Проверка гипотез:

$$H_0: \theta = \theta_0$$

$$H_A: \theta > \theta_0$$

"t-статистика";

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

t_{obs}

V_S

Например, гипотеза
про медиану....

Сутстрэп-анализ;

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

t_{1-2}^*

$$\begin{array}{ccc} x_1, \dots, x_n & \bar{x} & \hat{\sigma}_x^2 \\ y_1, \dots, y_m & \bar{y} & \hat{\sigma}_y^2 \end{array}$$

$$H_0: \mu_x = \mu_y$$

$$H_A: \mu_x \neq \mu_y$$

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

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

$$\begin{array}{l} 1. \quad x'_i = x_i - \bar{x} + \bar{z} \\ \quad y'_i = y_i - \bar{y} + \bar{z} \end{array} \quad \left. \begin{array}{l} \text{Отсюда убер} \\ \text{систему} \end{array} \right\}$$

$$2. \quad y_1^*, \dots, y_m^* \quad x_1^*, \dots, x_n^*$$

$$3. \quad t^* = \frac{\bar{x}^* - \bar{y}^*}{\sqrt{\frac{\hat{\sigma}_x^{*2}}{n} + \frac{\hat{\sigma}_y^{*2}}{m}}}$$

$$SE(\hat{d})$$

$$\hat{d} = \mu_x - \mu_y$$

$$4. \quad t_{obs} \quad \text{vs} \quad t_{1-\frac{\alpha}{2}}$$

Пяссковскии Бутстреп (Google 2012)

$$\lim_{n \rightarrow \infty} \text{Bin}(n, \frac{1}{n}) = \text{Poiss}(1)$$

(2) (1) (7) (4)

x_1, x_2, x_3, x_4

$$\begin{array}{l} B=1 \\ B=2 \\ B=3 \\ \dots \end{array} \left| \begin{array}{cccc} 4 & 7 & 2 & 4 \\ 2 & 2 & 1 & 2 \\ 7 & 4 & 2 & 4 \end{array} \right| \begin{array}{cccc} 1 & 0 & 1 & 2 \\ 3 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 \end{array}$$

можно тащить
набл.

я могу считать
частоты

$$Y_i \sim \text{Bin}(n; \frac{1}{n})$$

$$Y_1 + \dots + Y_n = n$$

можно звать n
и считать
сколько
набл. уже
осталось
добавить

$$Y_1, \dots, Y_n \sim \text{Mult}(n, \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$$

частоты

с которыми
поддаются

набл. в выборку

мы знаем $X_i \sim \text{Poiss}(1)$

$$N = X_1 + \dots + X_n \sim \text{Poiss}(n)$$

и бутстрап не работает

\Rightarrow можно распараллелить
генерацию

