Видно? Слышно?

① Вопросы по дом?

Bootstrap.   Бутстрэп
(е/а)

Проблема!
→ Часто есть асимпт. теория для $\hat{\theta}$ или CI

если [.....] то

при $n \to \infty$
$$P\left(\theta \in [\hat{a}_L ; \hat{a}_R] \right) \longrightarrow 0.95$$

→ требуется слишком большое n.

Проблема 2
→ нет теории (есть, но плох)

при $X_i \sim$ нез. один. распр (непр. распр) $E(X_i^?) < \infty$

$\mu = E(X_1)$

$$\mu \in [\bar{X} - 1.96 \cdot se(\bar{X}) ; \bar{X} + 1.96 \cdot se(\bar{X})]$$

$\nu = Med(X_i)$

$$\nu \in [s.Med(x) - ? ; s.Med(x) + ?]$$

Бутстрэп — семейство методов, осн-х
на оценке ф-ции распр-ия интере-
сующей нас статистики.

наивный, бутстрэп t-стат, бут в бут, BCA-бут.
$\longrightarrow x_{boot}(\hat{\theta})$, CI для θ

Наивный бутстрэп. (naive / percentile)

Задача. $X_1, \ldots, X_n \sim$ нз. один. распр. (непр).

$$Med(X_i) = m \qquad P(X_i > m) = P(X_i < m) = \frac{1}{2}$$

тот. оц: $\left(\hat{m}\right) = $ выб. медиана $= SMed(X_1, \ldots, X_n)$:
$$\uparrow$$
$$\text{sample}$$
$\rightarrow$ упоряд. X_i : $X_{(1)} \leq X_{(2)} \leq \ldots X_{(n)}$
$\rightarrow$ берем тот $X_{(i)}$, что посередине
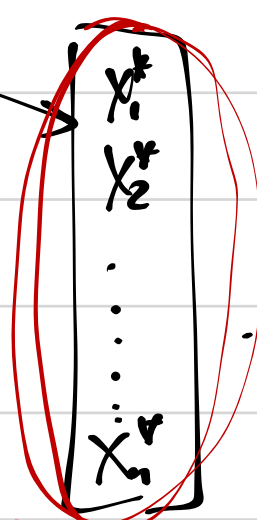
? 95% CI (asy) для m ?

Идея бутстрэпа:
исх. выборка.

* интерпретируем по исх-ой
выборке случ- ую
выборку того же
размера с повторениями.
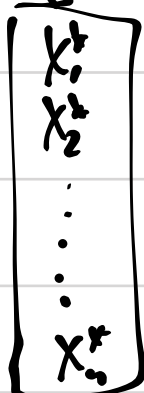
интуиц:
$X_1 \ldots X_n$ - нзов
$X_1^* \ldots X_n^*$ зав
совпадет с $n \uparrow$

$F_{\hat{m}} \overset{approx}{\sim} F_{\hat{m}^*}$

исх-ая   бут-выборка   бут-выборка
выборка   №1   №10000

$\hat{m}$ $\hat{m}^*$ .... $\hat{m}^*$ бут-стрэп-нные
                                                  выбор-ые медианы.

выб. медиана

2.5% $\hat{m}^*$ $\hat{m}^*$ $\hat{m}^*$ IR
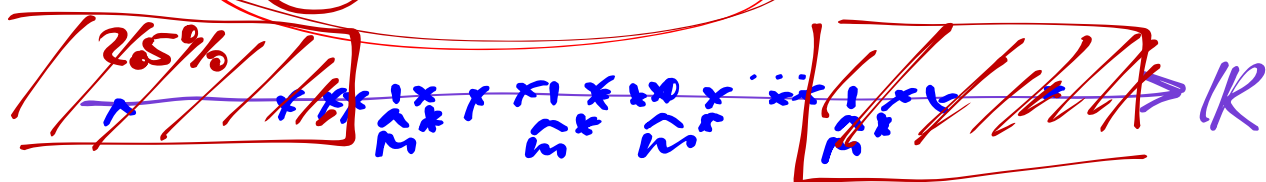
наивн. boot - CI : $[q_L(\hat{m}^*); q_R(\hat{m}^*)]$

q_L - левый квантиль

q_R - правый квантиль

⊕ ничего кроме формул для самой оценки не требуется

⊖ время боот велико.

исп-ие №2

получить $se_{naive}(\hat{m})$

n - чи сло наблюд.

кол-во дуб-боот-к $n_{boot} =$
$= 10000$

исх. выб.

$\hat{m}$

нет формулы $se(\hat{m})$

бутстрэп выборки.

$\hat{m}^*_1$ $\hat{m}^*_K$

$se_{boot}(\hat{m}) = \sqrt{\dfrac{\sum\limits_{j=1}^{n_{boot}}(\hat{m}^*_j - \overline{\hat{m}^*})^2}{n_{boot} - 1}}$
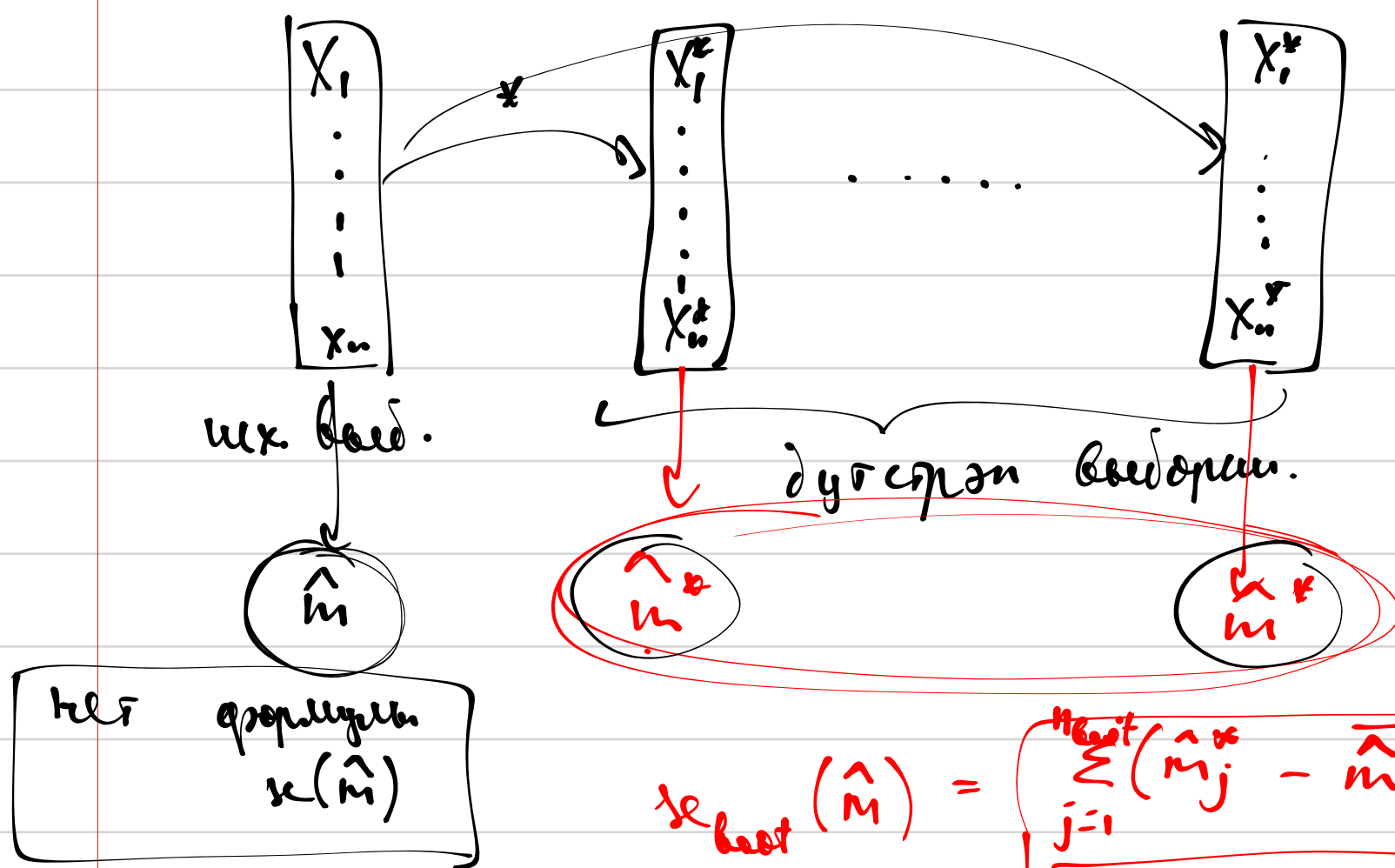
еще один вар-т

CI : $[\hat{m} - 1.96 \cdot se_{boot}(\hat{m}); \hat{m} + 1.96 \cdot se_{boot}(\hat{m})]$

предупреждения: правила

① n велико

② n_boot велико   = 10000

③ нельзя работать
    с гран. стат-ми.

Пример.

$$X_i \sim \text{незав.}, \quad U[0;\theta] \qquad \begin{matrix} X_1 \\ \vdots \\ X_n \end{matrix}$$

а) прид-те оценку (сост-ую) для θ.

$$\hat{\theta} = \max\{X_1, \ldots, X_n\}.$$

б) убедитесь, что наивн. бутстрэп строит
некорр-ый дов. ый интервал.
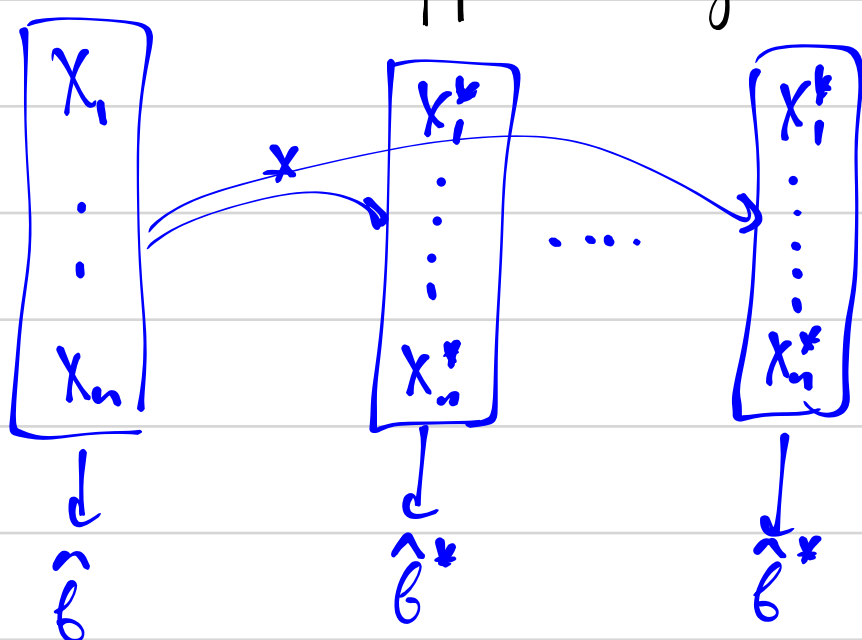
$n = 10^6$

$n_{boot} = 10^5$

как правило
в X_i^* даже
не все X_i
попадут
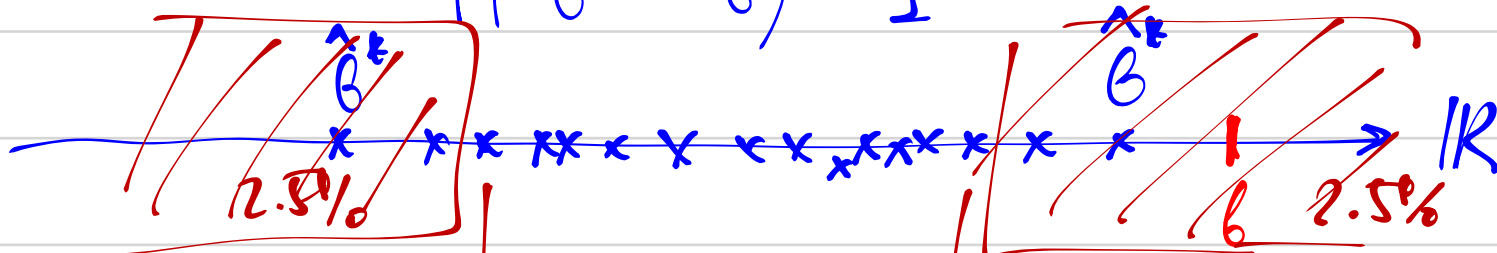
$$P(\hat{\theta} < \theta) \overset{?}{=} 1$$

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

$$P(\hat{\theta}^* < \theta) = 1$$

$$P\left([q_L(\hat{\theta}^*); q_R(\hat{\theta}^*)] \ni \theta\right) = 0 \not\to 0.95$$

$n, n_{boot} \to \infty$

Бутстрэп t-статистики.

(наивн. бут) (теория про кси) → будем хлерировать бутстрэп-ные статистики!

[⊕ более высокая скорость сх-сти]

Задача.

$$E(X_i^2) < \infty$$

$$X_1, X_2, \dots, X_n \sim \text{незав один распр}$$

$$\mu = E(X_i) \qquad CI?$$

а) [ML + t-тест.]   (95% CI)

$$\hat{\mu} = \frac{X_1 + \dots + X_n}{n} = \bar{X}$$

(CI:)   Ц.П.Т. $\left(\dfrac{\bar{X} - \mu}{se(\bar{X})}\right) \xrightarrow[n \to \infty]{dist} N(0;1)$

$$se(\bar{X}) = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n(n-1)}}$$

при $n \to \infty$ $-1.96 \leq \dfrac{\bar{X} - \mu}{se(\bar{X})} \leq 1.96$

$$\mu \in \left[\bar{X} - 1.96\, se(\bar{X}); \ \bar{X} + 1.96\, se(\bar{X})\right]$$

б) бутстрэп. t-стат-ки.

Эти значения не из фактич-го распр-ия $\dfrac{\bar{X} - \mu}{se(\bar{X})}$, а из предельного при $n \to \infty$

давайте вместо предельных крит-х значений возьмём их оценки.

$$\begin{array}{ccc}
\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} & \xrightarrow{\hspace{1cm} x \hspace{1cm}} \begin{bmatrix} X_1^* \\ \vdots \\ X_n^* \end{bmatrix} \quad \cdots\cdots & \begin{bmatrix} X_1^* \\ \vdots \\ X_n^* \end{bmatrix}
\end{array}$$

среди $\left\{\begin{array}{l} CI \ (\mu \text{ не знаем}) \\ \\ \hat{\mu} = \overline{X} \quad !! \\ \\ x(\overline{X}) \quad !! \end{array}\right.$

μ — не знаем

$t = \dfrac{\overline{X} - \mu}{x(\overline{X})}$

$\hat{\mu}^* = \overline{X}^* \qquad \xleftarrow{\text{разн}} \qquad \hat{\mu}^*_\bullet$

$x(\overline{X}^*) \qquad \xleftarrow{\text{разн}} \qquad x(X^{\bullet})$

$\mu^* = \overline{X} \ \textcolor{red}{[!]} \quad \xleftarrow{\text{один}} \quad \mu^* = \overline{X}$

$t^* = \dfrac{\overline{X}^* - \overline{X}}{x(\overline{X}^*)} \quad \xleftarrow{\text{разн}} \quad t^\bullet$

$$x(\overline{X}^\bullet) = \sqrt{\dfrac{\Sigma(X_i^* - \overline{X^*})^2}{n\cdot(n-1)}}$$

$t^k \qquad t^k \quad t^k \qquad\qquad t^k$

$\times \ \times \ \times\!\times_k\!\times \ \times\times \ \times\!\times\!\times \ \times\!\times\!\times\!\times\times \quad \longrightarrow \ \mathbb{R}$

классич CI.:

$\boxed{-1.96} \leq \dfrac{\overline{X} - \mu}{x(\overline{X})} \leq \boxed{+1.96}$

$\downarrow$ решаем нер-во

$\mu \in \left[\overline{X} - 1.96\cdot x(\overline{X}) ; \ \overline{X} + 1.96\, x(\overline{X}) \right]$

бутстрэп t- статистики.

$\boxed{\textcolor{red}{\varphi_L(t^*)} \leq \dfrac{\overline{X} - \mu}{x(\overline{X})} \leq \textcolor{red}{\varphi_R(t^*)}}$

$\downarrow$ решаем нер.

$\mu \in \left[\overline{X} - \varphi_R(t^*)\cdot x(\overline{X}) ; \ \overline{X} - \varphi_L(t^*)\cdot x(\overline{X}) \right]$

??? иллюзия ошибки

всё прав. !

как правило $\varphi_L(t^*) < 0.$