Problems:

1. you are fired
2. Strange external course
2.1. spss??????
2.2. coursework without data ?????

---

wiki.cs.hse.ru →

grades:

Vol grades = 0.3 · coursew + 0.7 · exam.

Fall grade = 0.4 · HA + 0.6 · Fall exam

Final grade = 0.5 · Fall grade + 0.2 · HA + 0.3 · Final exam

Bootstrap, Jackknife, CV.

Stand. Simple problem

$X_1, \ldots\ldots, X_n$ n is big

$X_i \sim iid$

$\mu = E(X_i)$

a) point estimator

$\hat{\mu} = \bar{X}$

b) CI 95% for μ:

$$CI = [\hat{\mu} - 1.96\, se(\hat{\mu})\ ;\ \hat{\mu} + 1.96\, se(\hat{\mu})]$$

$$se(\hat{\mu}) = se(\bar{X}) = \sqrt{\widehat{Var}(\bar{X})} = \sqrt{\frac{1}{n} \cdot \widehat{Var}(X_1)} =$$

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

$$Var(\bar{X}) = Var\left(\frac{X_1 + \ldots + X_n}{n}\right) =$$

$$= \frac{1}{n^2} \cdot n \cdot Var(X_1) = \frac{1}{n} \cdot Var(X_1)$$

real-life

$a = \text{Med}(X_1)$

ratio $\quad b = \frac{\mu}{\sigma}$

point estimator: $\hat{b} = \frac{\hat{\mu}}{\hat{\sigma}} = \frac{\bar{X}}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}}$

$se(X_i)$

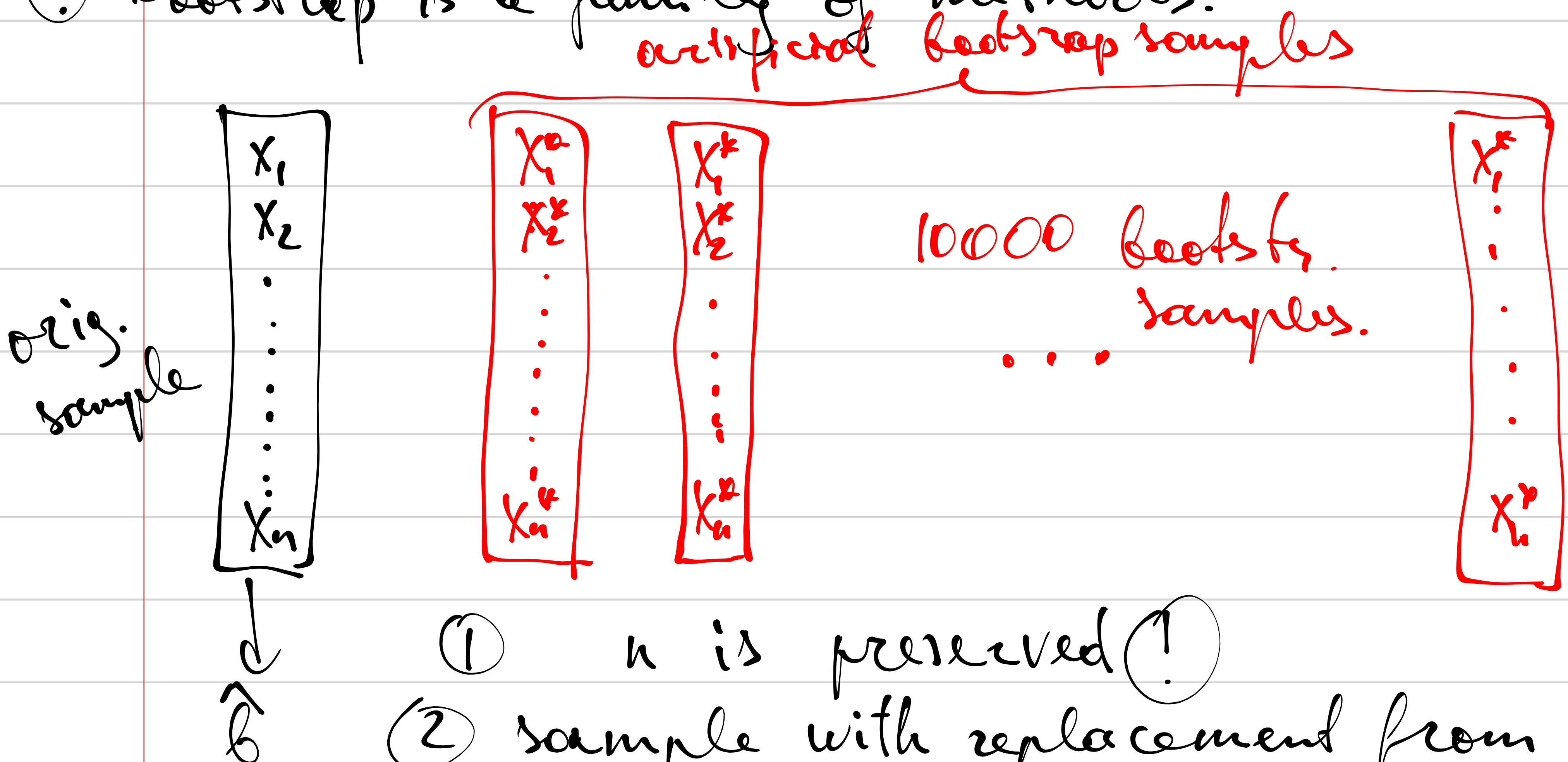
$Var(X_j)$

$Var(\bar{X}) = \frac{1}{n} \cdot Var(X_1)$

interval estimation $\quad \boxed{\text{CI for } b\text{?}}$

$\boxed{\text{Sol 1.}}$ Naive Bootstrap.

(!) Bootstrap is a family of methods.

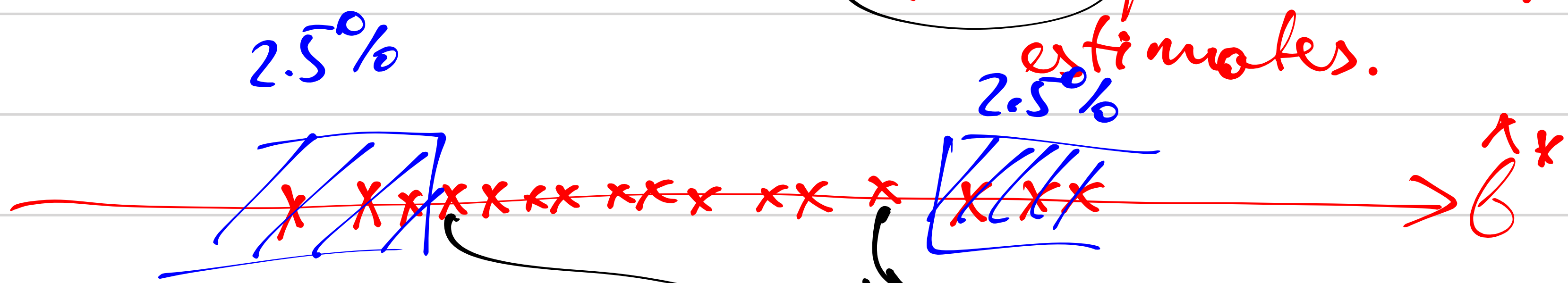


(1) n is preserved (!)

(2) sample with replacement from orig. sample.

$X_1=2$	Bootstr.-sample №1	Boot-sample №2	Boot-sample №10000
$X_2=0.5$	$X_1^*=7$	$X_1^*=-1$	X_1^*
$X_3=7$	$X_2^*=-1$	$X_2^*=-1$	⋮
$X_4=-1$	$X_3^*=7$	$X_3^*=-1$	⋮
	$X_4^*=2$	$X_4^*=-1$	X_4^*
	$\downarrow$	$\downarrow$	$\downarrow$
	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$	$\hat{\theta}_{10000}^*$

10000 point bootstr. estimates.



95% naive bootstrap CI: [,]

$$\left[q_{0.025}(\hat{\theta}^*) \,;\, q_{0.975}(\hat{\theta}^*)\right]$$

q - sample quantile

Adv:

1. no need for $se(\hat{\theta})$ formula!
2. CI for any θ for which we have consistent est. $\hat{\theta}$
3. more robust

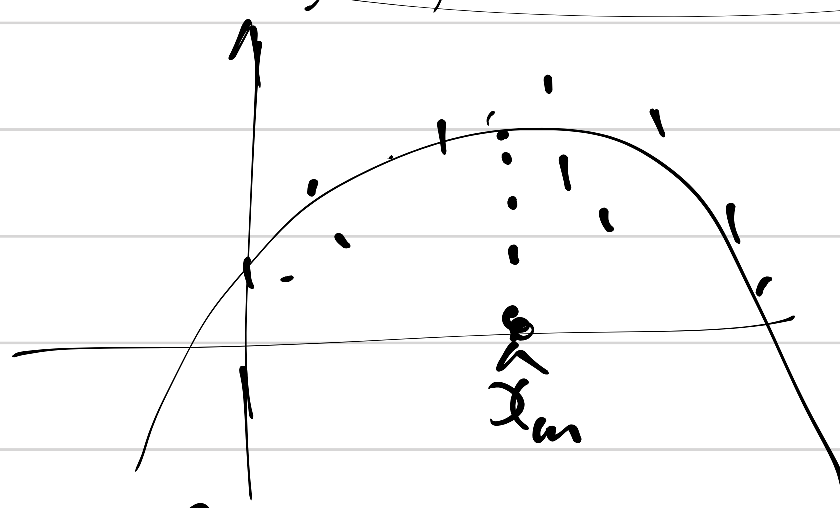
Dis

1. takes time
2. Asymptotic method $n \to \infty$

if n is small real coverage prob. $\neq$ nominal coverage probability

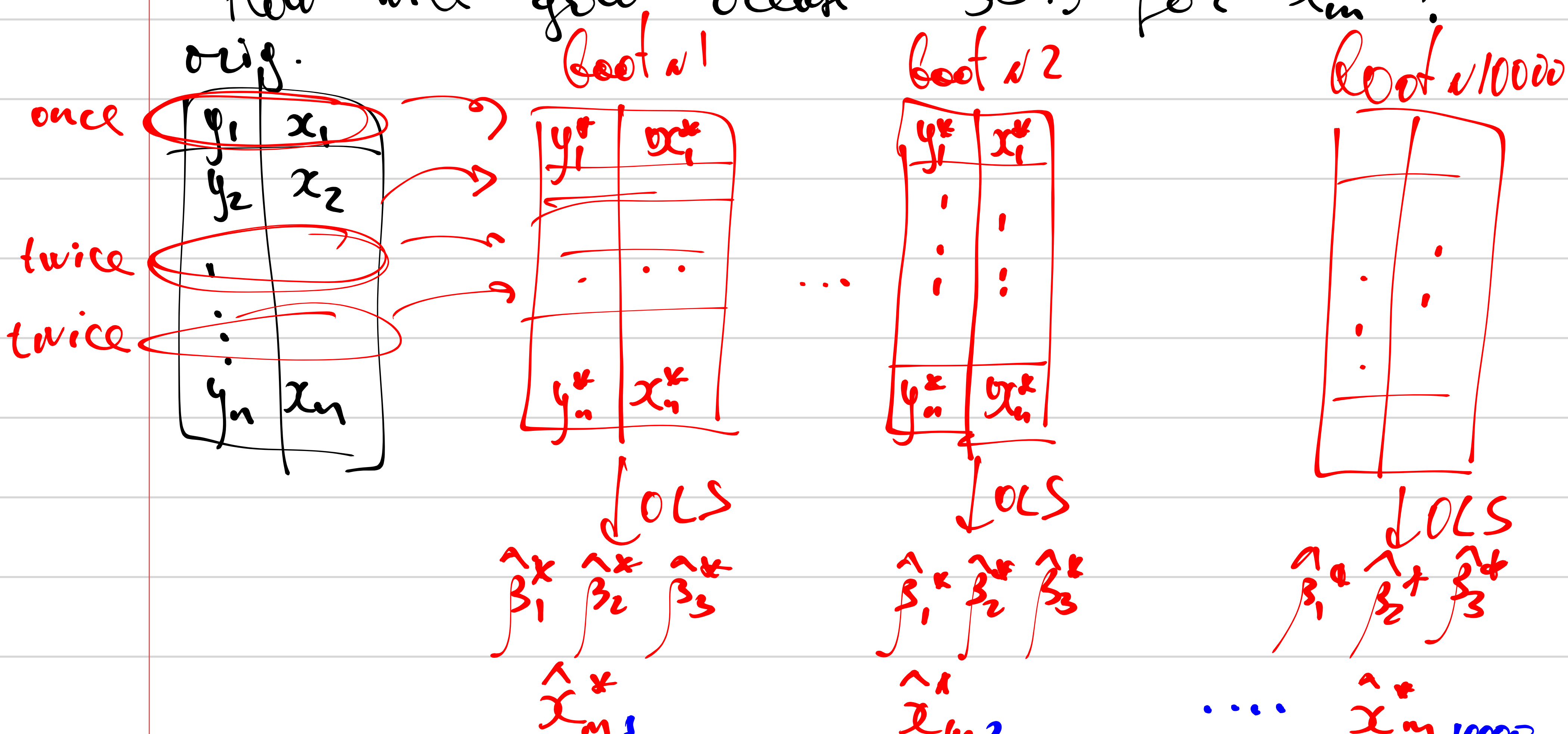
Question?

OLS: $\hat{y}_i = \hat\beta_1 + \hat\beta_2 \cdot x_i + \hat\beta_3 \cdot x_i^2$



$$\hat{x}_m = \boxed{\frac{-b}{2a}} = \frac{-\hat\beta_2}{2\hat\beta_3}$$

How will you create 95% for x_m?



t-stat bootstrap

higher speed of convergence than naive bootstrap.

$n = 100$

real cover. prob. 89% ← naive bootstrap nominal cover. prob 95%

93% ← t-stat bootstr with nominal cover. prob 95%

t-stat. bootstrap

Ideal sit.: $\boxed{CLT}$:

If $X_1, X_2, \dots$ are iid with $\mu = E(X_i)$, $\sigma^2 = Var(X_i)$ then $\frac{\bar{X} - \mu}{se(\bar{X})} \xrightarrow[n\to\infty]{distr} N(0;1)$

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

quantiles

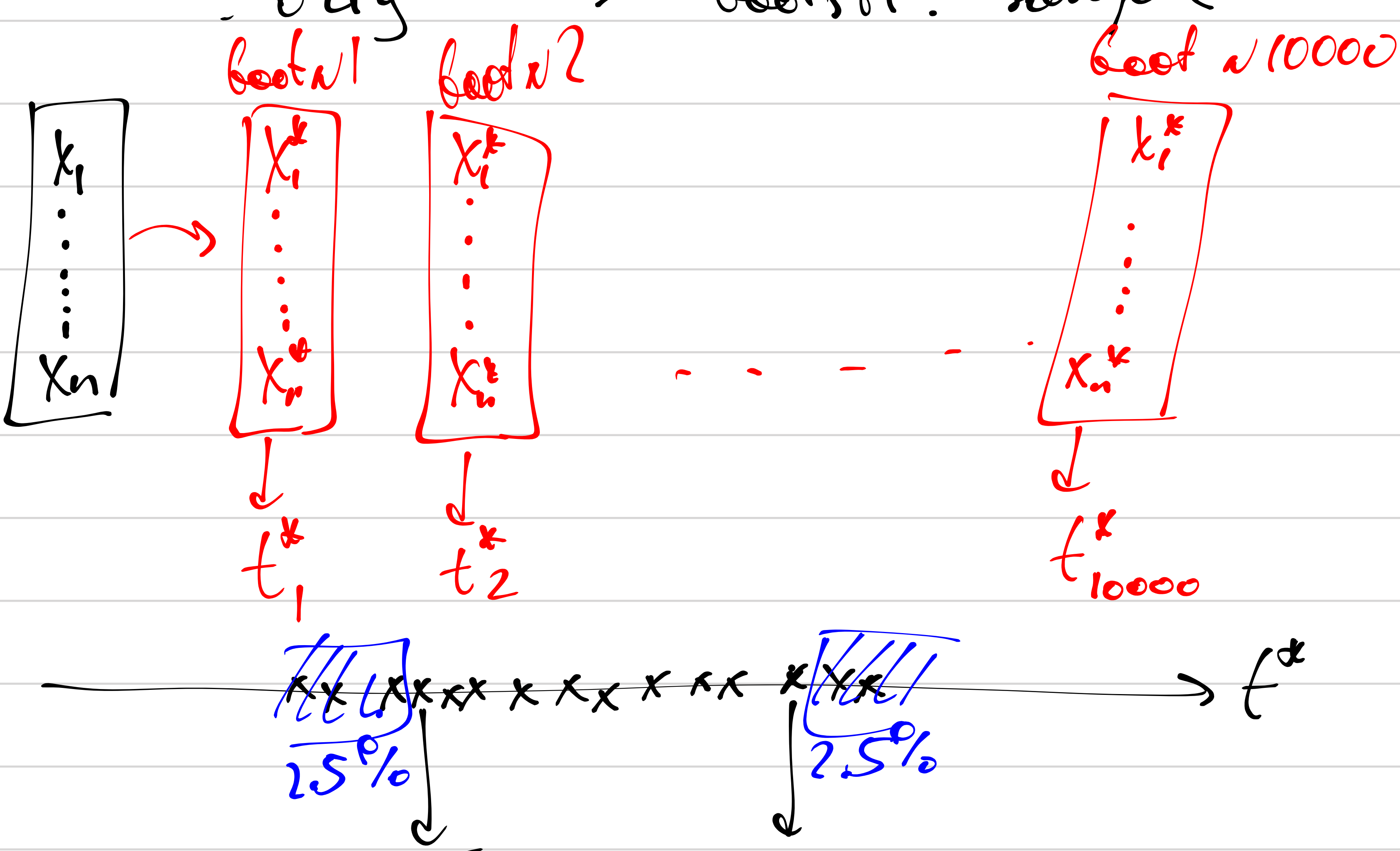
$$[\bar{X} - 1.96\cdot se(\bar{X}) ; \bar{X} + 1.96\cdot se(\bar{X})]$$

t-stat bootstrap: we use right resampled quantiles:

$$[\bar{X} - ?\cdot se(\bar{X}) ; \bar{X} + ?\cdot se(\bar{X})]$$

$$t = \frac{\bar{X} - \mu}{se(\bar{X})} \longrightarrow t^* = \frac{\bar{X}^* - \bar{X}}{se(\bar{X}^*)}$$

orig → bootstr. sample



CI for μ: $[\bar{X} - 1.8\, se(\bar{X}) ; \bar{X} + 2.5\, se(\bar{X})]$

CLT:

$$-1{,}96 \le \frac{\bar{X} - \mu}{se(\bar{X})} \le 1{,}96$$

t-stat boot:

$$-2.5 \le \frac{\bar{X} - \mu}{se(\bar{X})} \le 1.8$$

! mult by (-1)!

$$\mu \in [\bar{X} - 1{,}8\, se(\bar{X})\,;\ \bar{X} + 2.5\, se(\bar{X})]$$

$$\mu \in [\bar{X} - q_{0.975}(t^*) \cdot se(\bar{X});$$

$$\bar{X} - q_{0.025}(t^*) \cdot se(\bar{X})]$$

!

! Tim Hesterberg

What teachers should know about bootstrap?