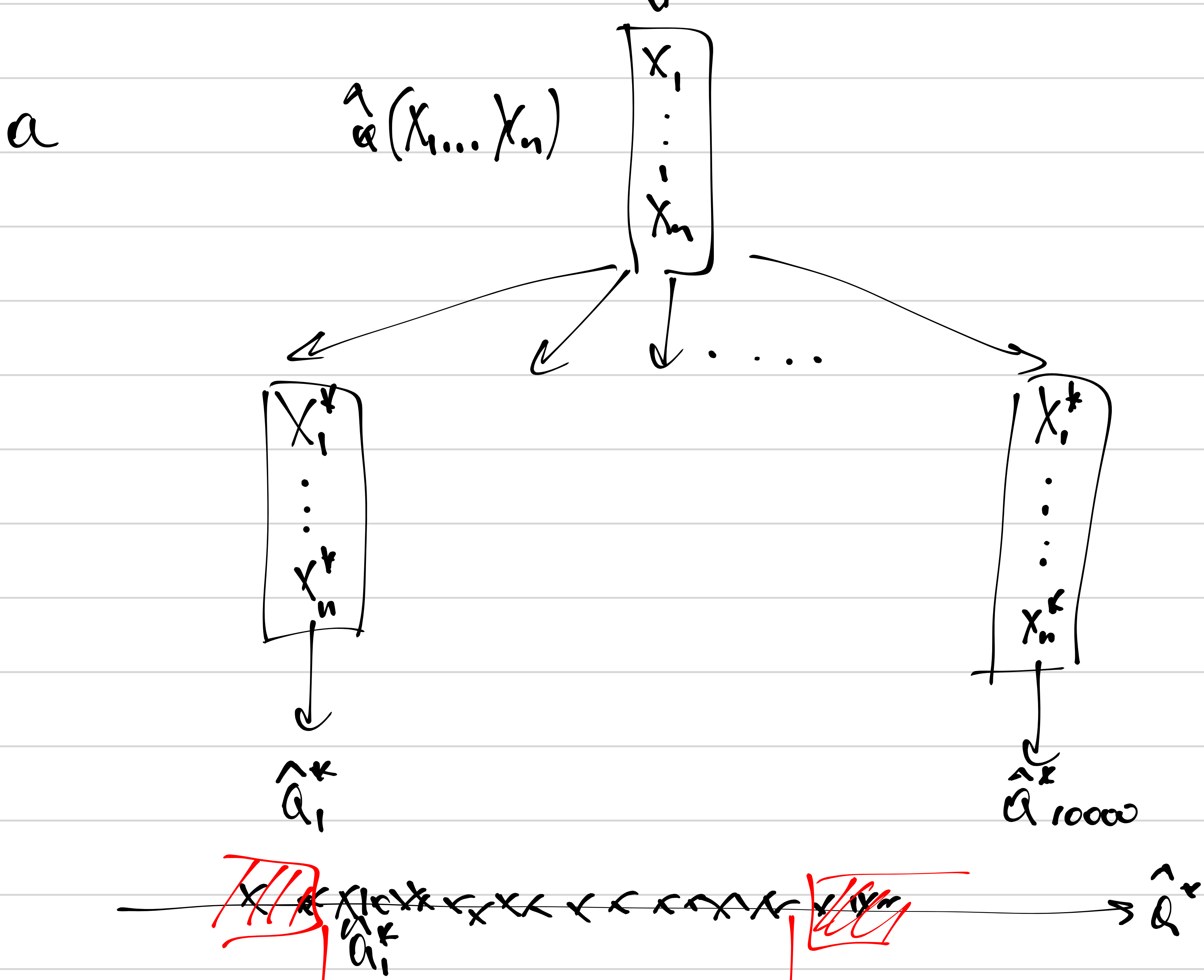
Hi :)

Bootstrap + Jackknife

naive bootstrap

t-stat bootstrap



$[\, q_L(\hat{a}^*) \; ; \; q_R(\hat{a}^*) \,]$

real cov-ge prob $\neq$ nominal cov-ge prob.

$n \to \infty$ real cov-ge prob $\to$ nom. cov-ge prob-ty

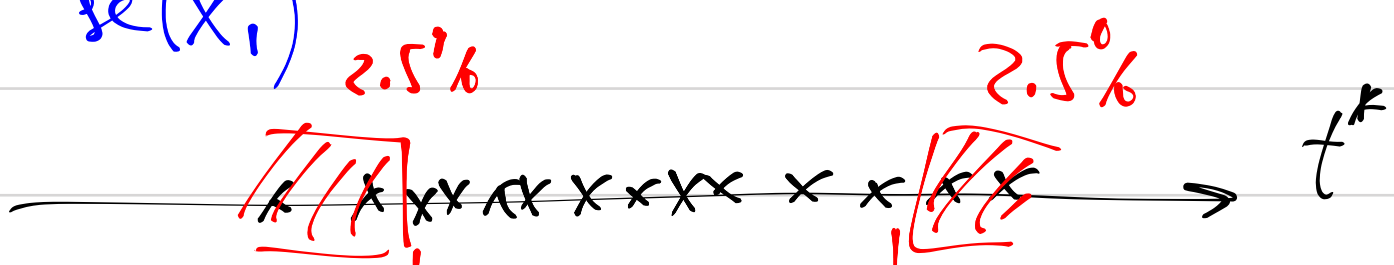
$$\begin{pmatrix} X_1^* \\ \vdots \\ X_n^* \end{pmatrix} \quad \cdots \cdots \quad \begin{pmatrix} X_1^* \\ \vdots \\ X_n^* \end{pmatrix}$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

$$\hat a_1^* \qquad\qquad\qquad \hat a_{10000}^*$$

$$t_1^* = \frac{\hat a_1^* - \hat a}{se(\hat a_1^*)} \qquad\qquad t_{10000}^* = \dots$$

$$t_1^* = \frac{\bar X_1^* - \bar X}{se(\bar X_1^*)}$$



unicorn: -1.96 $\quad$ 1.96 $\quad$ ideal s.t

real

real: q_L $\quad$ q_R

$$[\hat a - q_R \cdot se(\hat a) ; \hat a - q_L \, se(\hat a)].$$

t-stat bootstrap has lower difference $|\alpha_{real} - \alpha_{nom}|$ $\oplus$

we need $se(\hat a)$ $\ominus$

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

$$se(\hat p) = \sqrt{\frac{\hat p \cdot (1 - \hat p)}{n}}$$

$$se(\hat a) \qquad ?$$

Jackknife

$$\hat{a} = \hat{a}(X_1, X_2 \ldots \ldots X_n)$$

$$\hat{a}_{-3} = \hat{a}(X_1, X_2, \cancel{X_3}, X_4, X_5 \ldots X_n)$$

$$\hat{a}_{-i} = \hat{a}(X_1, \ldots X_{i-1}, \cancel{X_i}, X_{i+1}, X_{i+2} \ldots X_n)$$

$$\hat{a}_{-1}, \quad \hat{a}_{-2} \quad \ldots \quad \hat{a}_{-n}$$

$$\bar{\hat{a}}_{\bullet} = \frac{\sum_{i=1}^{n} \hat{a}_{-i}}{n}$$

$$\hat{\sigma}^2_{JK}(\hat{a}) = \frac{n-1}{n} \cdot \sum (\hat{a}_{-i} - \bar{\hat{a}}_{\bullet})^2$$

$$se_{JK}(\hat{a}) = \sqrt{\frac{n-1}{n} \cdot \sum (\hat{a}_{-i} - \bar{\hat{a}}_{\bullet})^2}$$

Theorem If $\hat{a} = \bar{X}$ then $se_{JK}(\hat{a})$ coincides with well-known

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

check!

$$\frac{n-1}{n} \sum (\hat{a}_{-i} - \bar{\hat{a}}_{\bullet})^2 \vee \frac{1}{n}\frac{1}{n-1} \sum (X_i - \bar{X})^2$$

$$(n-1)^2 \sum (\hat{a}_{-i} - \bar{\hat{a}}_{\bullet})^2 \vee \sum (X_i - \bar{X})^2$$

$$\bar{\hat{a}}_{\bullet} = \frac{\sum \bar{X}_{-i}}{n} = \frac{X_1 \cdot \frac{(n-1)}{n-1} + X_2 \cdot \frac{(n-1)}{(n-1)} + \ldots X_n \frac{(n-1)}{(n-1)}}{n} =$$

$$= \frac{\sum X_i}{n} = \bar{X}$$

$$\frac{\frac{X_2 + X_3 + \ldots + X_n}{n-1} + \frac{X_1 + X_3 + \ldots + X_n}{n-1} + \ldots}{n}$$

$$(n-1)^2 \cdot \sum (\hat{a}_{-i} - \bar{X})^2 \vee \sum (X_i - \bar{X})^2$$

LHS: $\hat{a}_{-1} - \bar{X} = \frac{X_2 + \dots + X_n}{n-1} - \frac{X_1 + \dots + X_n}{n}$

$$= -X_1 \cdot \frac{1}{n} + X_2 \cdot \left(\frac{1}{n-1} - \frac{1}{n}\right) + \dots X_n \cdot \left(\frac{1}{n-1} - \frac{1}{n}\right)$$

$$\frac{1}{n-1} - \frac{1}{n} = \frac{1}{(n-1)n}$$

RHS: $X_1 - \bar{X} = X_1 - \frac{X_1 + \dots + X_n}{n} =$

$$= X_1 \cdot \left(1 - \frac{1}{n}\right) + X_2 \left(-\frac{1}{n}\right) + \dots =$$

term RHS $= X_1 \frac{n-1}{n} + X_2 \cdot \left(-\frac{1}{n}\right) + \dots X_n \left(-\frac{1}{n}\right)$

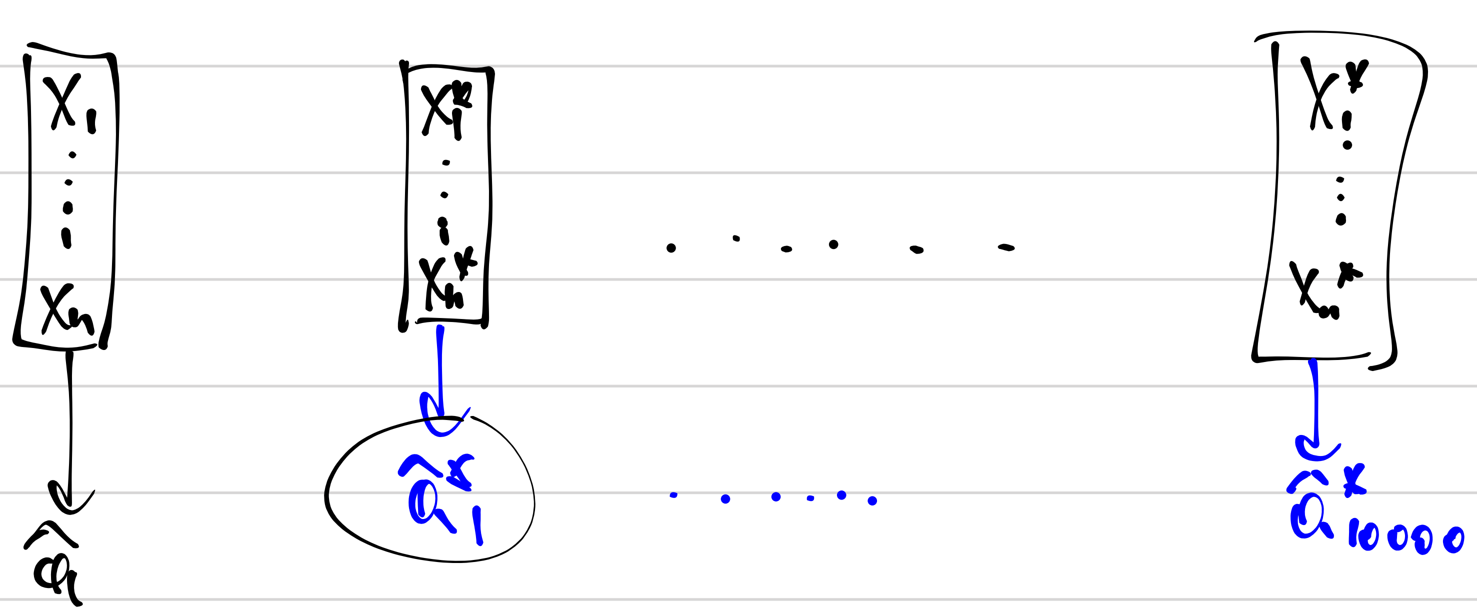
term LHS $= X_1 \cdot \left(-\frac{1}{n}\right) + X_2 \cdot \left(\frac{1}{(n-1)n}\right) + \dots X_n \left(\frac{1}{(n-1)n}\right)$

happy: if $\hat{a} = \bar{X}$ then

$$\hat{\sigma}^2_{JK}(\hat{a}) = se^2_{class}(\bar{X})$$

$$\hat{\sigma}^2_{JK} = \frac{n-1}{n} \cdot \sum_{i=1}^{n} \left(\hat{a}_{-i} - \bar{\hat{a}}_{\cdot}\right)^2$$ ■

Bootstrap estimator of variance.



case 1 ~~$se(\hat{a})$~~

case 2: you have $se(\hat{a})$ but under too restr. assump.

$$se^2_{boot}(\hat{a}) = \frac{\sum_{i=1}^{10000} \left(\hat{a}_i^* - \bar{\hat{a}}^*\right)^2}{10000 - 1}$$

The price we pay for

(+) faster converg. rate of $\alpha_{nom} - \alpha_{real}$

(+) no knowledge of $se(\hat{a})$.

(−) time! memory

---

t-stat boot will have 2 levels now!

$$X_1, \dots, X_n \rightarrow \hat{a}$$

$$se_{boot}(\hat{a}) = \frac{\sum (\hat{a}_i^* - \overline{\hat{a}^*})^2}{n_{1boot} - 1}$$

Level 1 boot

$X_1^*, \dots, X_n^*$ $\quad$ $X_1^*, \dots, X_n^*$ $\quad \dots \quad$ $X_1^*, \dots, X_n^*$

$n_{1boot} = 10000$

$\hat{a}_1^*$ $\quad$ $\hat{a}_2^*$ $\quad$ $\hat{a}_{10000}^*$

$$t_1^* = \frac{\hat{a}_1^* - \hat{a}}{se(\hat{a}_1^*)} \qquad t_2^* = \frac{\hat{a}_2^* - \hat{a}}{se(\hat{a}_2^*)}$$

no formula for se!

Level 2 boot

$X_1^{**}, \dots, X_n^{**}$ $\quad \dots \quad$ $X_1^{**}, \dots, X_n^{**}$ $\quad$ $X_1^{**}, \dots, X_n^{**}$ $\quad \dots \quad$ $X_1^{**}, \dots, X_n^{**}$ $\quad$ $X_1^{**}, \dots, X_n^{**}$ $\quad \dots \quad$ $X_1^{**}, \dots, X_n^{**}$

$n_{2boot} = 1000$

$\hat{a}_1^{**}$ $\quad$ $\hat{a}_{1000}^{**}$

t^*

$q_L(t^*)$ $\quad$ $q_R(t^*)$

$se(\hat{a}_1^*)$ no theory!

use $$se_{boot}(\hat{a}_1^*) = \frac{\sum (\hat{a}_i^{**} - \overline{\hat{a}^{**}})^2}{n_{2boot} - 1}$$

CI for a

$$\left[\hat{a} - q_R(t^*) \cdot se_{boot}(\hat{a});\ \hat{a} - q_L(t^*) \cdot se_{boot}(\hat{a})\right]$$

$p, \hat{p}$ real life

request nomin cov. prob $= 0.95$

class. method with se

real cov prob $= 0.87$

real. cov. naive Bootstrap $= 0.88$

real cov t-stat Bootstrap $= 0.92$