Chapter 8 - Bootstrap

Q_1: Computer experiment

find solution in Jupyter notebook

Q_2: Computer experiment

find solution in Jupyter notebook

Q_3: Computer experiment

find solution in Jupyter notebook

Q_4: $X_1, \dots, X_n$ distinct observations, no ties

show ther are $\binom{2n-1}{n}$ distinct bootstrap samples

hint: imagine putting n balls into n buckets.

$$\underbrace{\bullet}_{X_1} \mid \underbrace{\ \ }_{X_2} \mid \underbrace{\bullet\bullet}_{X_3} \mid \underbrace{\ \ }_{X_4} \mid \underbrace{\bullet\bullet}_{X_5} \quad \begin{cases} n \text{ dots} \\ n-1 \text{ bars} \end{cases}$$

$$\Rightarrow \text{number at distinct bootstrap samples} = \frac{(n+n-1)!}{n!\,(n-1)!} = \binom{2n-1}{n}$$

Q_5: $X_1, \dots, X_n$ distinct observations, $X_1^*, \dots, X_n^*$ bootstrap sample

find $E(\bar{X}_n^* \mid X_1, \dots, X_n)$, $V(\bar{X}_n^* \mid X_1, \dots, X_n)$, $E(\bar{X}_n^*)$, $V(\bar{X}_n^*)$

$$E(\bar{X}^* \mid x_1, \dots, x_n) = E(X_1^* \mid x_1, \dots, x_n) = E(X) = \bar{X} = \mu \quad ①$$

$$E(\bar{X}^*) = E(X_1^*) = E(X) = \mu$$

$$V(\bar{X}^* \mid X_i) = \frac{1}{n^2} \sum V(X_i^* \mid X_i) = \frac{1}{n} V(X_1^* \mid X_i) = \frac{1}{n^2} \sum (X_i - \mu)^2$$

$$\Rightarrow \quad V(\bar{X}^* \mid X_i) = \frac{n-1}{n^2} S_n^2 \quad ②$$

$V(\bar{X}^*) = ?$ let $y = \bar{X}^*$, $X = X_1, \dots, X_n$

we know that $V(y) = EV(y \mid X) + VE(y \mid X)$

$$VE(y \mid X) \overset{①}{=} V(\bar{X}) = \frac{\sigma^2}{n}$$

$$EV(y \mid X) \overset{②}{=} \frac{n-1}{n^2} E(S_n^2) = \frac{n-1}{n^2} \sigma^2$$

$$\Rightarrow V(\bar{X}^*) = \frac{\sigma^2}{n} + \frac{n-1}{n^2} \sigma^2$$

$$= \left(\frac{2n-1}{n^2}\right) \sigma^2 = \frac{2n-1}{n} V(\bar{X})$$

Q6: Computer experiment

find solution in Jupyter notebook

Q7: $X_1, \dots, X_n \sim Unif(0, \theta)$ $\hat{\theta} = X_{max} = max\{X_1, \dots, X_n\}$

generate dataset size 50, $\theta = 1$

a) dist. $\hat{\theta} = ?$ compare true dist. vs. bootstrap histogram

dist. can be found in Q2 of chapter 6

computer simulation in Jupyter notebook.

b) prove bootstrap does poorly in this case

show $P(\hat{\theta} = \hat{\theta}) = 0$ & $P(\hat{\theta}^* = \hat{\theta}) \approx 0.632$

hint: show $P(\hat{\theta}^* = \hat{\theta}) = 1 - (1 - \frac{1}{n})^n$ then take limit

$P(\hat{\theta} = c) = 0 \ \forall c$, since X is continous dist.

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

$A = \{X_1, \dots, X_n\}$, $\hat{\theta} = max\{A\} = X_j$

resample from $A \to A^* = \{X \in A\}$

if $X_j \in A^* \Rightarrow \hat{\theta}^* = \hat{\theta}$ how to calc $P(X_j \in A)$?

note: in each draw, there is $\frac{1}{n}$ chance to choose X_j

$\Rightarrow P(X_j \notin A^*) = (1 - \frac{1}{n})^n$

$\Rightarrow P(X_j \in A^*) = 1 - (1 - \frac{1}{n})^n$

$$\Rightarrow \lim_{n\to\infty} P(\hat{\theta}^* = \hat{\theta}) = \lim_{n\to\infty} 1-\left(1-\frac{1}{n}\right)^n = 1-e^{-1} \approx 0.632$$

Q_8: $T_n = \bar{X}_n^2$ $\qquad \mu = E(X_1)$

$$\alpha_k = \int |x-\mu|^k \, dF(x) \qquad \hat{\alpha}_k = n^{-1}\sum_{i=1}^{n} |X_i - \bar{X}_n|^k$$

Show $$v_{boot} = \frac{4\bar{X}_n^2 \hat{\alpha}_2}{n} + \frac{4\bar{X}_n \hat{\alpha}_3}{n^2} + \frac{\hat{\alpha}_4}{n^3}$$

$$v_{boot} = V(\bar{X}_n^2) = E(\bar{X}_n^4) - E(\bar{X}_n^2)^2 \quad (1)$$

$$\bar{X}_n = \mu + \frac{1}{n}\sum(X_i-\mu) = \mu + S_n \quad , \begin{cases} S_n = \frac{1}{n}\sum(X_i-\mu) \\ E(S_n)=0 \end{cases}$$

$$E(\bar{X}_n^2) = E(\mu^2 + S_n^2 + 2\mu S_n) = \mu^2 + E(S_n^2)$$

$$\Rightarrow E(\bar{X}_n^2)^2 = \mu^4 + E(S_n^2)^2 + 2\mu^2 E(S_n^2)$$

$$E(S_n^2) = n^{-2} E\left(\sum (X_i-\mu)^2 + \cancel{\sum_i \sum_{j\neq i} (X_i-\mu)(X_j-\mu)} \right)$$

assume X_i indep $\Rightarrow E[(X_i-\mu)(X_j-\mu)] = E(X_i-\mu)E(X_j-\mu) = 0$

$$\Rightarrow E(S_n^2) = \frac{\hat{\alpha}_2}{n}$$

$$\Rightarrow E(\bar{X}_n^2)^2 = \mu^4 + \frac{\hat{\alpha}_2^2}{n^2} + 2\mu^2 \frac{\hat{\alpha}_2}{n} \quad (2)$$

$$E(\bar{X}_n^4) = E(\mu^4 + S_n^4 + 4S_n^3\mu + 4S_n\mu^3 + 6S_n^2\mu^2)$$

$$= \mu^4 + E(S_n^4) + 4\mu E(S_n^3) + 4\mu^3 \cancel{E(S)}^{0} + 6\mu^2 E(S_n^2)$$

$E(S_n^2) = n^{-1}\hat{\alpha}_2$ from prev. section

$$E(S_n^3) = n^{-3} E\left(\sum (X_i-\mu)^3 + \sum\sum \cancel{(X_i-\mu)}^{0}(X_j-\mu)^2\right) = n^{-2}\hat{\alpha}_3$$

$$E(S_n^4) = n^{-4} E\left(\sum (X_i-\mu)^4 + \sum_i \sum_{j\neq i} (X_i-\mu)^2(X_j-\mu)^2 + \sum\sum \cancel{(X_i-\mu)}^{0}(X_j-\mu)^3\right)$$

$$= n^{-4}(n\hat{\alpha}_4 + n(n-1)\hat{\alpha}_2^2) = n^{-3}\hat{\alpha}_4 + (n-1)n^{-3}\hat{\alpha}_2^2$$

$$\Rightarrow E(\bar{X}_n^4) = \mu^4 + n^{-3}\hat{\alpha}_4 + (n-1)n^{-3}\hat{\alpha}_2^2 + 4\mu n^{-2}\hat{\alpha}_3 + 6\mu^2 n^{-1}\hat{\alpha}_2$$

$\Rightarrow$ Using 1 & 2

$$V(\bar{X}_n^2) = \cancel{\mu^4} + n^{-3}\hat{\alpha}_4 + \cancel{(n-1)n^{-3}\hat{\alpha}_2^2} + 4\mu n^{-2}\hat{\alpha}_3 + 6\mu^2 n^{-1}\hat{\alpha}_2$$

$$- \cancel{\mu^4} - \cancel{n^{-2}\hat{\alpha}_2^2} - 2\mu^2 n^{-1}\hat{\alpha}_2$$

approximate $(n-1)n^{-3} \approx n^{-2}$

$$\Rightarrow V(\bar{X}_n^2) = 4\mu^2 n^{-1}\hat{\alpha}_2 + 4\mu n^{-2}\hat{\alpha}_3 + n^{-3}\hat{\alpha}_4$$