Chapter 7 - Estimating CDF

Q_1: prove Theorem 7.3

$$E(\hat{F}_n(x)) = F(x)$$

$$\hat{F}_n = \frac{1}{n}\sum I(X_i \le x) \Rightarrow E(\hat{F}_n) = \frac{1}{n}\sum E(I(X_i \le x))$$

$$E(I(X_i \le x)) = P(X_i \le x) = F(x)$$

$$\Rightarrow E(\hat{F}_n) = F(x)$$

$$V(\hat{F}_n(x)) = \frac{1}{n} F(x)(1-F(x))$$

$$V(\hat{F}_n) = \frac{1}{n^2}\sum V(I(X_i \le x))$$

$$V(I(X_i \le x)) = E(I(X_i \le x)^2) - E(I(X_i \le x))^2 = F(x) - F(x)^2$$

$$\Rightarrow V(\hat{F}_n) = \frac{1}{n} F(x)(1-F(x))$$

$$MSE = \frac{1}{n} F(x)(1-F(x)) \to 0$$

$$MSE = \text{bias}(\hat{F})^2 + V(\hat{F}) = \frac{1}{n} F(x)(1-F(x))$$

$$\Rightarrow \lim_{n\to\infty} MSE = 0$$

$$\hat{F}_n \xrightarrow{P} F(x)$$

$$MSE \to 0 \Rightarrow \hat{F}_n \xrightarrow{qm} F(x) \Rightarrow \hat{F}_n \xrightarrow{P} F(x)$$

Q_2: $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ $\quad Y_1, \dots, Y_n \sim \text{Bernoulli}(q)$

find estimator & se for p & $p-q$

find 90% conf. interval for p & $p-q$

Estimator for p: $\hat{p} = \overline{X}_n$, $\hat{se}^2(\hat{p}) = \frac{\sigma^2}{n} = \frac{1}{n}\hat{p}(1-\hat{p})$

Estimator for $p-q$: $\hat{p} - \hat{q} = \overline{X}_n - \overline{Y}_n$, $\hat{se}^2(\hat{p}-\hat{q}) = \frac{1}{n}\left[\hat{p}(1-\hat{p}) + \hat{q}(1-\hat{q})\right]$

90% conf. interval : $\hat{p} \pm Z_{\frac{0.1}{2}} \hat{se}$

$Z_{0.05} = 1.65 \Rightarrow$ 90% C.I. $\begin{cases} \overline{X}_n \pm 1.65\, \hat{se}(\hat{p}) \\ \overline{X}_n - \overline{Y}_n \pm 1.65\, \hat{se}(\hat{p}-\hat{q}) \end{cases}$

Q_3: computer experiment

find solution in Jupyter notebook.

Q_4: $X_1, \dots, X_n \sim F$ & empirical dist. $\hat{F}$

use CLT to find limiting dist. of $\hat{F}$

$\hat{F}_n = \frac{1}{n}\sum I(X_i \leq x)$, assume x fixed, $Z_i = I(X_i \leq x)$

Z_i are IID $\Rightarrow$ Using CLT we have:

$\hat{F}_n = \overline{Z} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{n}\right)$ where $\begin{cases} \mu = E(Z_i) = F(x) \\ \sigma^2 = \frac{1}{n} F(x)(1-F(x)) \end{cases}$

Q_5: $Cov(\hat{F}_n(x), \hat{F}_n(y)) = ?$

$$Cov(\hat{F}(x), \hat{F}(y)) = E(\hat{F}_x \hat{F}_y) - E(\hat{F}_x)E(\hat{F}_y)$$

$$= E(\hat{F}_x \hat{F}_y) - F_x F_y \quad ①$$

$$E(\hat{F}_x \hat{F}_y) = \frac{1}{n^2} E\left[\sum_i I(X_i \leq x) \sum_j I(X_j \leq y)\right]$$

$$= \frac{1}{n^2} E\left[\sum_{i=j} I(X_i \leq x) I(X_j \leq y) + \sum_{i \neq j} I(X_i \leq x) I(X_j \leq y)\right]$$

$$= \frac{1}{n^2}\left[n \underbrace{E(I(X_1 \leq x) I(X_1 \leq y))}_{F(min(x,y))} + n(n-1) \underbrace{E(I(X_1 \leq x) I(X_2 \leq y))}_{F(x)F(y)}\right]$$

$$\Rightarrow E(\hat{F}_x \hat{F}_y) = \frac{1}{n} F(min(x,y)) + \frac{n-1}{n} F(x)F(y) \quad ②$$

$$①, ② \Rightarrow Cov(\hat{F}_x, \hat{F}_y) = \frac{1}{n} F(min(x,y)) - \frac{1}{n} F_x F_y$$

Q_6: $X_1, \dots, X_n \sim F$, empirical $\hat{F}_n$

let $a < b$, $\theta = T(F) = F(b) - F(a)$

$$\hat{\theta} = T(\hat{F}) = \hat{F}_n(b) - \hat{F}_n(a)$$

$\hat{se}(\hat{\theta}) = ?$ $\quad$ $1-\alpha$ conf. interval for θ?

$$V(\hat{\theta}) = V(\hat{F}(b)) + V(\hat{F}(a)) - 2\,Cov(\hat{F}(b), \hat{F}(a))$$

from prev. question: $Cov(\hat{F}(b), \hat{F}(a)) = \frac{F(a)}{n}(1 - F(b))$

$$\Rightarrow V(\hat{\theta}) = \frac{1}{n}\left[F(b)(1-F(b)) + F(a)(1-F(a)) - 2F(a)(1-F(b)) \right]$$

$$= \frac{1}{n}\left(F(b)-F(a)\right)\left(1-F(b)+F(a)\right)$$

$$\Rightarrow \hat{se}(\hat{\theta}) = \frac{1}{n}\left(\hat{F}(b)-\hat{F}(a)\right)\left(1-\hat{F}(b)+\hat{F}(a)\right)$$

$1-\alpha$ confidence interval: $\hat{\theta} \pm Z_{\alpha/2}\,\hat{se}(\hat{\theta})$

Q7: Computer experiment

find solutions in Jupyter notebook.

Q8: Computer experiment

find solutions in Jupyter notebook.

Q9: $X_1, \dots, X_{100}$, $P_1 = 0.9$

$Y_1, \dots, Y_{100}$, $P_2 = 0.85$

$\theta = P_1 - P_2 \Rightarrow \hat{\theta} = ?\quad \hat{se}(\hat{\theta}) = ?\quad$ 80%, 95% CI

Same as question 2: $\hat{\theta} = \hat{P}_1 - \hat{P}_2 = 0.9 - 0.85 = 0.05$

$1-\alpha$ CI: $\hat{\theta} \pm Z_{\alpha/2}\,\hat{se}$

$$\hat{se} = \sqrt{\frac{0.9 \times 0.1 + 0.85 \times 0.05}{100}} = 0.047$$

$$\left.\begin{array}{l} 80\%\ C.I. \Rightarrow Z_{\alpha/2} = 1.28 \\ 95\%\ C.I. \Rightarrow Z_{\alpha/2} = 1.96 \end{array}\right\} \quad C.I. = 0.05 \pm Z_{\alpha/2} \times 0.047$$

Q_{10}: Computer experiment

find solutions in Jupyter notebook.