MTH210a (2023): Quiz 4

Name ..Dootika Vats.......... Roll. No. ..Solutions..........

Let $(X_i, Y_i) \overset{\text{iid}}{\sim} F$, where F is some unknown bivariate distribution with mean vector $(\theta, \theta)^T$ and finite marginal variances. In order to estimate θ, you choose the estimator:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i + Y_i}{2}\,.$$

1. (~~7~~ 10 marks) Write the steps of bootstrapping to obtain a confidence interval around $\hat{\theta}$. Explain all steps carefully.

Bonus, 2. (3 marks) Propose an alternative way (without using bootstrap) to obtain a confidence interval around $\hat{\theta}$?

① $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n) \overset{iid}{\sim} F$ with mean $\begin{pmatrix}\theta\\ \theta\end{pmatrix}$.

(−5) if parametric bootstrap done correctly.

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i + Y_i}{2}$$

In order to obtain C.I around $\hat{\theta}$, we will implement non-parametric bootstrap.

For $b = 1, \dots, B$. we draw a sample of size n from $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ with replacement. (−2) if details not given about sampling method.

Bootstrap Sample 1: $(X_{11}^*, Y_{11}^*), (X_{12}^*, Y_{12}^*), \dots, (X_{1n}^*, Y_{1n}^*)$: set $\hat{\theta}_1^* = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{1i}^* + Y_{1i}^*}{2}$

Bootstrap sample 2: $(X_{21}^*, Y_{21}^*), (X_{22}^*, Y_{22}^*), \dots, (X_{2n}^*, Y_{2n}^*)$: set $\hat{\theta}_2^* = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{2i}^* + Y_{2i}^*}{2}$

⋮

Bootstrap Sample B: $(X_{B1}^*, Y_{B1}^*), (X_{B2}^*, Y_{B2}^*), \dots, (X_{Bn}^*, Y_{Bn}^*)$: set $\hat{\theta}_B^* = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{Bi}^* + Y_{Bi}^*}{2}$.

The $100(1-\alpha)$ confidence interval is then

$$\left(\hat{\theta}^*_{(\lfloor \frac{\alpha}{2} B \rfloor)},\ \hat{\theta}^*_{(\lfloor (1-\frac{\alpha}{2}) B \rfloor)}\right).$$

(−2) if orderstatistic is missing $\theta_{(\cdot)}$

(−3) if CIs are missing.

② Note that

$$\text{Var}(\hat{\theta}) = \frac{1}{n}\text{Var}\left(\frac{X_1 + Y_1}{2}\right) = \frac{1}{4n}\left[\text{Var}(X_1) + \text{Var}(Y_1) + 2\,\text{Cov}(X_1, Y_1)\right]$$

then by CLT $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \tau^2)$ where $\tau^2 = \dfrac{\text{Var}(X_1) + \text{Var}(Y_1) + 2\,\text{Cov}(X_1, Y_1)}{4}$.

Let $\hat{\tau}^2 = \dfrac{S_X^2 + S_Y^2 + \cancel{2S} \; 2S_{XY}}{4}$ where

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}).$$

Then $\hat{\theta} \pm z_{\frac{\alpha}{2}}\sqrt{\dfrac{\hat{\tau}^2}{n}}$ is a potentian alternative for a confidence interval..

Points given only if completely correct.