Advanced Nonparametric Inference

M. Stat. 2nd Year 2024–25 Indian Statistical Institute

Semester Examination

28 April 2025 Total Marks: 50 Time: 3 hours

Instructions

- This is an open notes exam. You can bring your handwritten notes.
- Answer as much as you can. The maximum you can score is 50.
- Calculator or any other electronic device is not allowed during the exam.
- You can use any result obtained in the class. Clearly state the results that you use.
- Clearly state all the assumptions that you make. Failure to do so will lead to deduction of marks.
- Keep your answers to the point. Any unnecessary or irrelevant writing will be penalized.

Question 1. Total marks: 28

Let $X_1, \ldots, X_n$ be i.i.d. observations from the $\mathsf{Uniform}(0, 1)$ distribution. Consider the problem of density estimation based on the observations $X_1, \ldots, X_n$. Let $\widehat{f}_{\text{hist}}$ be the histogram density estimator with bins $B_k = \left(\left(k - \frac{1}{2}\right)h, \left(k + \frac{1}{2}\right)h\right]$, $k = 0, 1, 2, \ldots$. Assume that the bin-width h is of the form $h = 1/m$ for some $m \in \mathbb{N}$.

(a) Find the pointwise bias and variance of $\widehat{f}_{\text{hist}}$. [2+2]

(b) Compute the mean integrated squared error (MISE) of $\widehat{f}_{\text{hist}}$. Obtain the optimal bin-width h by minimizing the MISE. Find the value of the optimal MISE. [2+2+2]

Now, consider the kernel density estimator $\widehat{f}_{\text{kde}}$ with bandwidth $h = 1/m$ and the triangular kernel $K(t) = (1 - |t|)\mathbf{1}_{\{-1 \le t \le 1\}}$.

(c) Find the pointwise bias and variance of $\widehat{f}_{\text{kde}}$. [3+5]

(d) Compute the MISE of $\widehat{f}_{\text{kde}}$. Obtain the optimal bandwidth h by minimizing the MISE. Find the value of the optimal MISE. [4+2+2]

(e) Which of these two density estimators do you find more appropriate for this problem? Justify your answer. [2]

Question 2. Total marks: 30

Let $X_1, \ldots, X_n$ be i.i.d. $\mathsf{Bernoulli}(p)$ random variables. Suppose that our parameter of interest is $\theta = p^2$. Consider the statistic $T_n = \bar{X}_n^2$ as an estimator for θ.

(a) Derive $\mathbb{E}(T_n)$ and $\mathrm{Var}(T_n)$. [1+3]

(b) Find the jackknife bias estimator b_{jack} and the bias-corrected jackknife estimator T_{jack} for the problem. Is jackknife able to reduce bias in this example? Justify. [2+2+1]

(c) Calculate the mean squared errors (MSEs) of the two estimators T_n and T_{jack}. Which one has lesser MSE? [4+1]

(d) Find the jackknife variance estimator v_{jack}. Is it a consistent estimator of $\text{Var}(T_n)$? Justify. [3+4]

(e) Find the naive bootstrap variance estimator v_{boot}. Is it consistent for $\text{Var}(T_n)$? Justify. [3+4]

(f) Among v_{jack} and v_{boot}, which one would you prefer and why? [2]