

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2024 – 25

Resampling Techniques: MStat-I

Date: 30 April 2025

Maximum Marks: 100

Duration: 3 Hours

You can use any result discussed in Class making appropriate reference
and mentioning the relevant assumptions.

Calculators are allowed, but no sharing between students.

This paper carries 108 marks. Attempt all question. The maximum you can score is 100.

1. For an i.i.d. sample data of size n , (a) show that the Jackknife estimator of variance $v_{jack}(T_n)$, for a statistic T_n , will remain unchanged if we order the sample values. (b) Is this true for bootstrap? [10 + 8 = 18]

2. Consider D1: Data set 1 ($n = 10$): {3.09, 0.26, 0, 2.16, 1, 3, 1.22, 4.57, 2, 1}.
For D1, we want to estimate the distribution. Compare the performance of delete-2 Jackknife, Grouped Jackknife (delete -1) and Balanced Bootstrap. (State all assumptions carefully). [18]

3. Let $x_{(1)} < x_{(2)} < x_{(3)} < x_{(4)} < x_{(5)} < x_{(6)} < x_{(7)}$ be an ordered sample of size $n = 7$. Let x^* be a bootstrap sample, and $s(x^*)$ be the corresponding bootstrap replication of the median. Show that

(a) $s(x^*)$ equals one of the original data values $x_{(i)}$, $i = 1, 2, \dots, 7$.

(b) $s(x^*)$ equals $x_{(i)}$ with probability

$$p(i) = \sum_{j=0}^3 \left\{ Bi \left(j; n, \frac{i-1}{n} \right) - Bi \left(j; n, \frac{i}{n} \right) \right\},$$

where $Bi(j; n, p)$ is the binomial probability $\binom{n}{j} p^j (1-p)^{n-j}$. [6 + 12 = 18]

4. Consider the problem of estimating the third central moment of a distribution. Based on i.i.d. samples, suppose we start with the estimator T_n which is the third sample moment.

Find its bias, the jackknife bias corrected estimator and the bias of this estimator.

[5 + 8 + 5 = 18]

5. Give one example of a situation where the Bootstrap accelerated bias corrected percentile strategy will work worse than the simple bootstrap percentile strategy for creating confidence sets. (You may use D1.) [12]

6. Let $X_1, X_2, \dots, X_n$ be independent and identically distributed as Uniform $(0, \theta)$, and hence, $\hat{\theta}_{mle} = X_{(n)}$. Consider two different Bootstrap approaches as follows:

(a) Parametric Bootstrap: Sample $X_i \sim \text{Uniform}(0, \hat{\theta}_{mle})$, and we generate 10,000 samples, for each sample we obtain $\hat{\theta}_{mle} = X_{(n)}^*$, the corresponding bootstrap estimate and its empirical distribution.

(b) Nonparametric Bootstrap: We take a bootstrap resample (with replacement) from the original sample and obtain the similar estimate and its distribution.

We claim that the Nonparametric Bootstrap fails to provide valid inference on θ . Do you agree? Explain. [12]

7. Suppose we have a sample of n data points from some unknown distribution, and we denote them as $X_1, X_2, \dots, X_n$. A problem of interest is to know the shape of the (unknown) probability distribution which is the data generating process (DGP) for this sample. Suppose, we want to test if the distribution is unimodal or there are multiple modes in it. A Bootstrap-based approach for this problem was proposed in Silverman (1981, 1983) and was discussed in class.

Based on this approach, develop a Resampling-based Statistical test for testing the following hypothesis: H_0 : There are exactly two modes in the original distribution vs. H_1 : The number of modes > 2 . Write each step clearly. [12]