Total: 75

Duration: 3 hours

QUESTION 1 **[9]**

1.1 Provide a short definition of the bootstrap (2)

1.2 Explain the difference between the concepts of the "ideal bootstrap estimator" and the "approximate bootstrap estimator". (2)

1.3 Consider using the EDF, F_n, as an estimator for F. The EDF is defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x), \text{ where } I(\cdot) \text{ is the indicator function.}$$

1.3.1 Explain how one would go about drawing bootstrap samples from the EDF. (1)

1.3.2 If it is known that the EDF is an unbiased estimator for $F(x)$, show that the variance of the EDF, F_n is equal to $\frac{1}{n}F(x)\big(1-F(x)\big)$. (2)

1.4 Prove that the ideal Bootstrap estimate of the expected value of the statistic $\hat{\theta}_n = \bar{X}_n$ is given by $\bar{X}_n$.

(2)

QUESTION 2 **[10]**

Let $X_1, X_2, \dots, X_n$ be a random sample from an unknown distribution F with parameter $\theta = t(F)$. Let $\hat{\theta}_n = t(F_n)$ be the plug-in estimate of θ.

2.1 Suppose one wants to use the bootstrap to estimate the following quantity

$$\delta = Var\left(\hat{\theta}^3 - E\left(\hat{\theta}\right)^3\right)^{\theta}$$

By only making use of the plug-in principle, state the ideal bootstrap estimate of δ (using "bootstrap" notation). Call this $\hat{\delta}$ (2)

2.2 By only making use of the plug-in principle, *state* the ideal bootstrap estimate of Var($\hat{\delta}$) (using "bootstrap" notation). (3)

2.3 Provide an algorithm that can be used to approximate the estimate given in 2.2. (5)

QUESTION 3 **[8]**

3.1 A sample of data $X_1, X_2, \dots, X_n$ is collected from some population. Suppose that the population has the property that $[E(X)]^2 = Var(X) = 1$. Describe, in detail, the procedure you would use to obtain bootstrap samples $X_1^*, X_2^*, \dots, X_n^*$. (Only state the method used to obtain the bootstrap samples). (2)

3.2 A sample of data $Y_1, Y_2, \dots, Y_n$ is collected from a different population. Suppose that it is known that this population is normally distributed. Describe, in detail, the procedure you would use to obtain bootstrap samples $Y_1^*, Y_2^*, \dots, Y_n^*$. (Only state the method used to obtain the bootstrap samples). (2)

3.3 Consider the smoothed Bootstrap sample $Y_i^* = X_i^* + hZ_i$, $i = 1,2,\dots,n$ where X_i^* are i.i.d. from the EDF, F_n, and Z_i are i.i.d. from some symmetric distribution function $K(x)$ and $Var(Z_i) = 1$ and $E(Z_i) = 0$. Moreover, assume that $\{X_i^*\}$ and $\{Z_i\}$ are independent. Prove that creating the sample $Y_i^* = X_i^* + hZ_i$ is the same as sampling from the smoothed kernel distribution function (4)

$$F_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

QUESTION 4 **[10]**

Assume that G denotes the distribution function of the statistic $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$, where $X_1, X_2, \dots, X_n$ is a random sample from an unknown distribution F with parameter θ.

4.1 The $100(1-\alpha)\%$ bootstrap-t interval is given by

$$\left[\hat{\theta}_n - \hat{H}^{-1}\left(1 - \frac{\alpha}{2}\right)\hat{\sigma}_n \,;\, \hat{\theta}_n - \hat{H}^{-1}\left(\frac{\alpha}{2}\right)\hat{\sigma}_n\right],$$

where $\hat{H}(x)$ is the plug-in estimate of $H(x) = P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \le x\right)$, and $\hat{\sigma}_n^2$ is the bootstrap estimator for $Var(\hat{\theta}_n)$. Provide an algorithm that you would use to approximate this bootstrap interval. (5)

4.2 Suppose we are interested in a $100(1-\alpha)\%$ confidence **lower bound** for θ, i.e., the interval $[\,C\,;\infty)$ such that θ is contained in this interval with probability $1-\alpha$. Derive an expression for C by making use of the same procedures used to obtain the "Hybrid" percentile bootstrap confidence interval for θ. (5)

QUESTION 5 **[10]**

Suppose we have a random sample $X_1, X_2, \dots, X_n$ from an unknown distribution F and we want to estimate the 3rd "raw" moment $\mu_3 = E(X^3)$ with the statistic $\hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$.

5.1 Provide the ideal bootstrap estimate of the bias of this statistic (only provide the "bootstrap" expression for this estimator; do not provide an algorithm). (1)

5.2 It is possible that $\hat{\mu}_3$ is a biased estimator for μ_3. Explain how the estimator might be improved through the use of "bias correction". Motivate your answer. (2)

5.3 Provide a basic algorithm to approximate the bootstrap estimate of bias of $\hat{\mu}_3$. (4)

5.4 How can this *algorithm* be improved? Define all quantities used in this improved algorithm carefully using the scenario provided. (3)

QUESTION 6 **[13]**

Let $X_1, X_2, \dots, X_n$ be a random sample from an unknown distribution F with finite mean $\mu_1 = E(X)$, finite second "raw" moment $\mu_2 = E(X^2)$, finite variance σ^2 and median η.

6.1 Consider testing the hypothesis

$$H_0\colon \sigma^2 = 100 \ \text{ vs } \ H_A\colon \sigma^2 > 100,$$

where σ_0^2 is a constant. Explain in detail how you will approximate the bootstrap p-value to test this hypothesis using an algorithm. State the decision rule you would use to reject (or not reject) H_0. (6)

6.2 Consider testing the hypothesis

$$H_0: \eta = \eta_0 \text{ vs } H_A: \eta < \eta_0,$$

where η_0 is a constant. Describe how you will transform $X_1, X_2, \dots, X_n$ to construct a sample (a "bootstrap population") that satisfies the null hypothesis. (2)

6.3 Consider testing the hypothesis

$$H_0: \mu_3 = 0 \text{ vs } H_A: \mu_3 > 0,$$

where $\mu_3 = E(X^3)$ is the 3rd "raw" moment. Instead of transforming the data in order to "mimic" the null hypothesis, one can change the probability associated with each data value in order to mimic the null hypothesis. That is, we can find the modified EDF (denoted F_{Tilt}) that assigns probability p_i to X_i, $i = 1,2, \dots, n$.

Find an expression, by solving an appropriate minimization problem, for the probabilities p_i , $i = 1, \dots, n$. (*Provide the expression for* $p_i, i = 1, \dots, n$ *that can be obtained as far as possible without having to resort to numerical optimisation procedures*). (5)

QUESTION 7 **[12]**

Assume that we have the following statistical relationship between the i.i.d. random variable $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$ with unknown distribution F (*dependent* on unknown parameters $\beta_0, \beta_1, \beta_2$ and β_3), the fixed dependent variable $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ and the i.i.d. random errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$, with unknown distribution G (*independent* of the parameters $\beta_0, \beta_1, \beta_2$ and β_3 and $E(\varepsilon_i) = 0, i = 1,2, \dots, n$):

$$Y_i = (\beta_0 + \beta_1 X_i + \beta_2 \varepsilon_i)^{\beta_3}$$

Given sample data $Y_1, \dots, Y_n$ we can estimate β_0, β_1, β_2 and β_3 with $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

7.1 Provide a bootstrap algorithm that could be used to approximate the bootstrap estimate of standard error of $\hat{\beta}_3$. (4)

7.2 How will your answer in 9.1 change if there are heteroskedasticity present in the model? Define all expressions used in your explanation carefully (state the desired distributional properties associated with the random variables methods that are introduced when applying this method). (3)

7.3 How will your method of resampling change in Question 9.1 if we did not explicitly assume that the model $Y_i = (\beta_0 + \beta_1 X_i + \beta_2 \varepsilon_i)^{\beta_3}$ generated the data? (2)

Suppose now we are interested in using the bootstrap to test the hypothesis

$$H_0: \beta_3 = 1 \quad vs. \quad H_A: \beta_3 > 1$$

using some test statistic, say, $T_n = T_n(\boldsymbol{Y}, \boldsymbol{X})$.

7.4 Provide the bootstrap algorithm you would use to test this hypothesis (specifically explain how you would go about drawing your bootstrap samples for this test – to answer this question, you may simply indicate how the algorithm given in 9.1 changes). (3)

QUESTION 8 **[3]**

Let $X_1, X_2, \dots, X_n$ be a sample drawn from an unknown population. Assume that the sample values are **unique.** Denote by $X_1^*, X_2^*, \dots, X_n^*$ a bootstrap sample drawn with replacement from the original sample.

Calculate $P^*(X_k^* \leq X_{k+1}^*)$ for some $k = 1, \dots, n-1$. (3)

TOTAL: 75