UNIVERSITY OF JOHANNESBURG

FACULTY OF SCIENCE

DEPARTMENT OF STATISTICS	
MODULE	**STA8X03** Nonparametric Methods
CAMPUS	**APK**
EXAM	**NOVEMBER 2019**

DATE: 11 November 2019	**SESSION: 8:30 - 11:30**
ASSESSOR	**Dr SJM Steeb**
EXTERNAL MODERATOR	**Prof J Allison**
DURATION: 3 HOURS	**MARKS: 91**

NUMBER OF PAGES: 3
INSTRUCTIONS: Answer all the questions.

QUESTION 1 [15]

Consider a sample $X_1, \ldots, X_n$ from a population $x_1, \ldots, x_N$ with mean $\mu = \sum_{j=1}^N x_j$ and variance $\sigma^2 = \sum_{j=1}^N (x_j - \mu)^2/N$. The sample mean $\bar{X} = \sum_{i=1}^n X_i$ is used to estimate the population mean μ.

1.1 Find the standard error of $\bar{X}$ for samples with and without replacement. (8)

1.2 Find and explain the relationship between the standard errors found in 1.1. (4)

1.3 How do you expect the bootstrap estmate of standard error of $\bar{X}$ to compare with the standard error of $\bar{X}$? (3)

QUESTION 2 [22]

Suppose that a random sample $X_1, X_2, \ldots, X_n$, is drawn from a population with cumulative distribution function F.

2.1 Use indicators to show that the empirical distribution function $\hat{F}$ is the plug-in estimate of F. (3)

2.2 Is information lost when, instead of $X_1, X_2, \ldots, X_n$, the reduced representation $\hat{F}$ is used? Motivate your answer by considering the two independent samples problem. (8)

2.3 Show that

$$E(\hat{F}(y)) = F(y) \qquad \text{and} \qquad \mathrm{var}(\hat{F}(y)) = \frac{1}{n} F(y)(1 - F(y)) \,.$$

(5)

2.4 Is $\hat{F}$ a consistent estimator of $F(y)$? Motivate your answer. (2)

2.5 Discuss the limiting distribution of $\hat{F}$. (4)

QUESTION 3 [9]

Consider a random sample $X_1, X_2, \ldots, X_n$ from the unknown distribution of a real valued random variable with mean μ and variance σ^2.

3.1 Find the plug-in estimate of σ^2. (7)

3.2 What property of the plug-in principle is illustrated when considering the standard error of $\hat{\sigma}^2$? (2)

QUESTION 4 [10]

Consider the linear regression model

$$Y_i = \boldsymbol{c}_i \boldsymbol{\beta} + \varepsilon_i \quad \text{for} \quad i = 1, 2, \ldots, n$$

with $\boldsymbol{c}_i = (c_{i1}, c_{i2}, \ldots, c_{ip})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^t$. Suppose that the regression parameters β_j are estimated and that the bootstrap standard error estimates have to be calculated.

Discuss the application of the bootstrap algorithm using the residuals $\hat{\varepsilon}_i = y_i - \boldsymbol{c}_i \hat{\boldsymbol{\beta}}\,, i = 1, 2, \ldots, n$, to find the ideal bootstrap standard error estimates $\hat{se}_\infty(\hat{\beta}_j)$ for $j = 1, 2, \ldots, p$.

QUESTION 5 [15]

Let θ be the parameter of interest and $\hat{\theta}$ be a biased estimator of θ, with

$$\mathrm{bias}_F = \mathrm{bias}_F(\hat{\theta} - \theta) = E_F(\hat{\theta}) - \theta \,.$$

Discuss how the bootstrap method and the jackknife method, respectively, are applied to estimate bias_F. Where possible, comment on the relationships between the estimators.

QUESTION 6 [8]

A random sample $\boldsymbol{x} = (X_1, X_2, \ldots, X_n)$ is observed from a probability distribution of real numbers F and it is desired to estimate the variance $\theta = \mathrm{var}_F(X)$. The plug-in estimate is $\hat{\theta} = \sum_{i=1}^n (X_i - \bar{X})^2/n$. Show that the jackknife bias-corrected estimate is the usual unbiased estimate of variance, that is

$$\bar{\theta} = \hat{\theta} - \hat{\mathrm{bias}}_{\text{jack}} = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)\,.$$

QUESTION 7 [12]

Suppose that $\hat{\theta}$ is a linear statistic such that

$$\hat{\theta} = \mu + \frac{1}{n}\sum_{i=1}^{n} \alpha_i,$$

where $\alpha_i = \alpha(X_i)$.

7.1 Show that $se_{\hat{F}}(\hat{\theta}) = \sqrt{\frac{n-1}{n}}\hat{se}_{\text{jack}}$. (9)

7.2 Discuss the result given in 7.1. (3)