NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

SORS 5104

FACULTY OF APPLIED SCIENCES

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

SORS 5104: COMPUTATIONAL STATISTICS

MSc. BIG DATA: PART I

DECEMBER 2024 EXAMINATION

Time : 3 hours

Answer **ANY FOUR** questions .
Answers for each question should start on a fresh page.

A1. (a) Let $ps(X)$ be a set of random numbers, where

$$ps(X) = \{0.44, 0.81, 0.14, 0.05, 0.93\}.$$

Test whether the random numbers in $ps(X)$ are uniformly distributed using the Kolmogorov-Smirnov test. [12]

(b) Determine whether the following 40 random numbers are independent of each other using the Runs Up and Runs Down Test.

$$0.41, 0.68, 0.89, 0.94, 0.74, 0.91, 0.55, 0.62, 0.36, 0.27$$
$$0.19, 0.72, 0.75, 0.08, 0.54, 0.02, 0.01, 0.36, 0.16, 0.28$$
$$0.18, 0.01, 0.95, 0.69, 0.18, 0.47, 0.23, 0.32, 0.82, 0.53$$
$$0.31, 0.42, 0.73, 0.04, 0.83, 0.45, 0.13, 0.57, 0.63, 0.29$$

[13]

A2. (a) Discuss any five properties of pseudo-random number generators. [10]

(b) Given that

$$f(x) = \frac{4}{1+x^2}$$

and

$$Z_0 = 8193,$$

i) Set Z_0 as the seed and generate five pseudo-random numbers using the von-Neumann Algorithm. [5]

ii) Use the von-Neumann Algorithm output to compute the Monte Carlo integral of $f(x)$. [10]

A3. (a) The profit obtained after producing λ units of product X and θ units of product Y is approximately modelled as,

$$f(\lambda, \theta) = 8\lambda + 10\theta - 0.001(\lambda^2 + \lambda\theta + \theta^2) - 10000.$$

Find the product level that produces the maximum profit. [10]

(b) Starting at point X=(1,1), use the Gradient Search Method to solve the unconstrained maximisation problem,

$$f(X) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2.$$

[15]

A4. (a) A sample, **s** = {Moyo, Ncube, Nyoni, Sibanda}, was bootstrapped.

i) Find the probability that all four people appear in the bootstrap sample. [5]

ii) Compile a python code for calculating the probability that N people will appear in a bootstrap sample if the sample size for **s** is adjusted to N. [8]

(b) A university has N employees where N is a large number. Each employee has one of three possible job classifications (Technician, Administrator and Lecturer). The employee job classifications change independently according to a Markov chain with transition probabilities

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}.$$

What percentage of employees will be in each job classification on the long run? [12]

A5. (a) Given that a=5, c=1, m=8 and X_0=3, generate the first 10 pseudo random numbers using the Linear Congruential Generator. [13]

(b) Consider θ, a Biased Maximum Likelihood Estimator for the sample variance given a Normal distribution, where

$$\begin{aligned}\theta &= \frac{\sum_1^n (x_i - \bar{x})^2}{n} \\ &= \frac{\sum_1^n x_i^2}{n} - \bar{x}^2.\end{aligned}$$

Show that the Jackknife Method can be used to convert the Biased Maximum Likelihood Estimator (θ) to an Unbiased Maximum Likelihood Estimator (θ_{JK}), where

$$\theta_{JK} = \frac{\sum_i^n (x_i^2 - n\bar{x}^2)}{n-1}.$$

[12]

END OF QUESTION PAPER