Indian Institute of Technology Guwahati
Statistical Inference (MA682)
Project I

Full Marks: 30 Last Date of Submission: **March 26, 2024**

Instructions

- Attempt **all** questions.
- Notations are standard and same as used during the lectures.
- **Write a report in details**. The report should contain necessary calculations, tables, graphs, seed, interpretations and conclusions. There should be **one report file** for all questions. Convert it into PDF file and rename as **Project1Group⟨Your group number⟩.pdf**. Submit it through an email.
- **Submit codes** (either .R file or .py files). There should be **separate files for each question**. Rename the file for Question x as **Project1Group⟨Your group number⟩Question$x$.R** or **Project1Group⟨Your group number⟩Question$x$.py** and upload all the files. The code files should be well commented for easy readability. The code file for each question should be standalone file so that all necessary results (numerical, graphical, etc.) can be obtained by running the code file only once.
- **One of the group members must send all files** (one report in PDF format and three code files) to `aganguly@iitg.ac.in`. This email must be copied to all other group members. The subject of the email must be **[MA682] Project I Submission (Group ⟨ Your Group Number ⟩)**.
- There will be **no credits** if either appropriate works are not shown in the report and/or error free codes are not submitted.
- You can use inbuilt function or routine for generating random number from uniform distribution. **Do not use** inbuilt functions/routines for generating random numbers from any other distributions.

Problems

1. Consider the probability density function

$$f(x) = \begin{cases} (2x+1)\, e^{-\left(x^2+x\right)} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) (3 points) Device an algorithm using inverse transformation technique to generate random number from the above probability density function. Write each step clearly in the report. Implement it in R or python code.

   (b) (3 points) Obtain an estimate of the expectation corresponding to the above probability density function using Monte Carlo technique. Mention each step in the report including the value of seed and the number of random numbers used in this calculation.

(c) (2 points) Generate 10,000 random numbers using the same seed as used in the previous part. Draw the histogram for the generated random numbers. On the same graph, draw (superimpose) the probability density function. Comment on the graph.

(d) (2 points) Draw empirical distribution function based on the random number generated in the part (c). Superimpose the cumulative distribution function corresponding to the above probability density function on the same graph. Comment on the graph.

2. Assume that the data-set given in the CSV file DataP1Q2.csv is a realization of a random sample of size 53 from the following cumulative distribution function

$$F(x) = \begin{cases} \left(1 - e^{-\frac{x}{\theta}}\right)^2 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0, \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) (5 points) Find maximum likelihood estimate of the parameter θ based on the data given in the CSV file.

(b) (5 points) Use Monte Carlo technique to estimate the bias and mean square error of the maximum likelihood estimator for $\theta = 0.5,\ 1,\ 2$. Clearly mention each step (including any specific values needed) in the report.

3. (10 points) A manufacturing company of an electric product claims that the average lifetime of their product is 15 years with standard deviation 13 years. To verify their claim, a random sample of size 20 is drawn from the population of all products. The CSV file DataP1Q3.csv contains corresponding lifetimes in years. It seems plausible to assume that the data are realization of a random sample from a population with following probability density function:

$$f(x;\, \mu,\, \sigma^2) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\ln x - \mu\right)^2\right) & \text{if } x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are unknown parameters. Construct a problem of hypothesis testing and verify the claim of the company based on the data. Clearly mention each step including null and alternative hypothesis, method used, derivation of test function (in implementable form).