

## Gate ST 24.2023

Rusna Shaikh  
EE22BTECH11048

Q.24 Let  $X_1, X_2, \dots, X_n$  be a random sample of size  $n$  from a population having uniform distribution over the interval  $(\frac{1}{3}, \theta)$ , where  $\theta > \frac{1}{3}$  is an unknown parameter. If  $Y = \max \{X_1, X_2, \dots, X_n\}$ , then which one of the following statements is true?

1.  $(\frac{n+1}{n})(Y - \frac{1}{3}) + \frac{1}{3}$  is an unbiased estimator of  $\theta$
2.  $(\frac{n}{n+1})(Y - \frac{1}{3}) + \frac{1}{3}$  is an unbiased estimator of  $\theta$
3.  $(\frac{n+1}{n})(Y + \frac{1}{3}) - \frac{1}{3}$  is an unbiased estimator of  $\theta$
4.  $Y$  is an unbiased estimator of  $\theta$

**Solution:**

Any one of the above expressions is an unbiased estimator of  $\theta$  if

$$E(\text{estimator}) = \theta$$

We will first calculate expected value of  $Y$

The PDF of  $X_i$  can be expressed as

$$p_{X_i}(x) = \frac{1}{\theta - \frac{1}{3}}, \quad \frac{1}{3} < \theta < 3 \quad (1)$$

$$F_{X_i}(y) = \int_{\frac{1}{3}}^y p_{X_i}(x) dx \quad (2)$$

$$= \int_{\frac{1}{3}}^y \frac{1}{\theta - \frac{1}{3}} dx \quad (3)$$

$$= \frac{y - \frac{1}{3}}{\theta - \frac{1}{3}} \quad (4)$$

$$F_X(y) = \Pr(X \leq y) \quad (5)$$

$$= \Pr(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \quad (6)$$

$$= \Pr(X_1 \leq y) \Pr(X_2 \leq y) \dots \Pr(X_n \leq y) \quad (7)$$

$$= \left( \frac{y - \frac{1}{3}}{\theta - \frac{1}{3}} \right)^n \quad (8)$$

$$p_X(y) = \frac{d}{dy} F_X(y) \quad (9)$$

$$= \frac{n}{\left(\theta - \frac{1}{3}\right)^n} \left(y - \frac{1}{3}\right)^n \quad (10)$$

$$E(Y) = \int_{\frac{1}{3}}^{\theta} y p_X(y) dy \quad (11)$$

$$= \int_{\frac{1}{3}}^{\theta} y \frac{n}{\left(\theta - \frac{1}{3}\right)^n} \left(y - \frac{1}{3}\right)^n dy \quad (12)$$

$$= \frac{n}{\left(\theta - \frac{1}{3}\right)^n} \int_{\frac{1}{3}}^{\theta} y \left(y - \frac{1}{3}\right)^{n-1} dy \quad (13)$$

$$\text{Let } y - \frac{1}{3} = t \quad (14)$$

$$\implies y = t + \frac{1}{3} \implies dy = dt \quad (15)$$

$$\text{Therefore,} \quad (16)$$

$$E(Y) = \frac{n}{\left(\theta - \frac{1}{3}\right)^n} \int_0^{\theta - \frac{1}{3}} \left(t^n + \frac{t^{n-1}}{3}\right) dt \quad (17)$$

$$= \frac{n}{\left(\theta - \frac{1}{3}\right)^n} \left( \frac{\left(\theta - \frac{1}{3}\right)^{n+1}}{n+1} + \frac{\left(\theta - \frac{1}{3}\right)^n}{3n} \right) \quad (18)$$

$$= \frac{n}{n+1} \left(\theta - \frac{1}{3}\right) + \frac{1}{3} \quad (19)$$

$$= \frac{3n\theta + 1}{3(n+1)} \neq \theta \quad (20)$$

Therefore, fourth option is incorrect.

1.

$$E\left(\left(\frac{n+1}{n}\right) \left(Y - \frac{1}{3}\right) + \frac{1}{3}\right) = \frac{n+1}{n} E(Y) - \frac{n+1}{3n} + \frac{1}{3} \quad (21)$$

$$= \frac{3n\theta + 1 - (n+1) + n}{3n} \quad (22)$$

$$= \theta \quad (23)$$

It is the unbiased estimator of  $\theta$ .

2. Similarly,

$$E\left(\left(\frac{n}{n+1}\right)\left(Y - \frac{1}{3}\right) + \frac{1}{3}\right) = \frac{n}{n+1}E(Y) - \frac{n}{3(n+1)} + \frac{1}{3} \quad (24)$$

$$= \frac{n(3n\theta + 1) - n(n+1) + (n+1)^2}{3(n+1)^2} \quad (25)$$

$$= \frac{3n^2 + n}{3(n+1)^2} \neq \theta \quad (26)$$

It is not an unbiased estimator of  $\theta$ .

3. In the same way,

$$E\left(\left(\frac{n+1}{n}\right)\left(Y + \frac{1}{3}\right) - \frac{1}{3}\right) = \frac{n+1}{n}E(Y) + \frac{n+1}{3n} - \frac{1}{3} \quad (27)$$

$$= \frac{3n\theta + 1 + n + 1 - n}{3n} \quad (28)$$

$$= \frac{3n\theta + 2}{3n} \neq \theta \quad (29)$$

It is not an unbiased estimator of  $\theta$ .

Hence, the first option is the correct option.

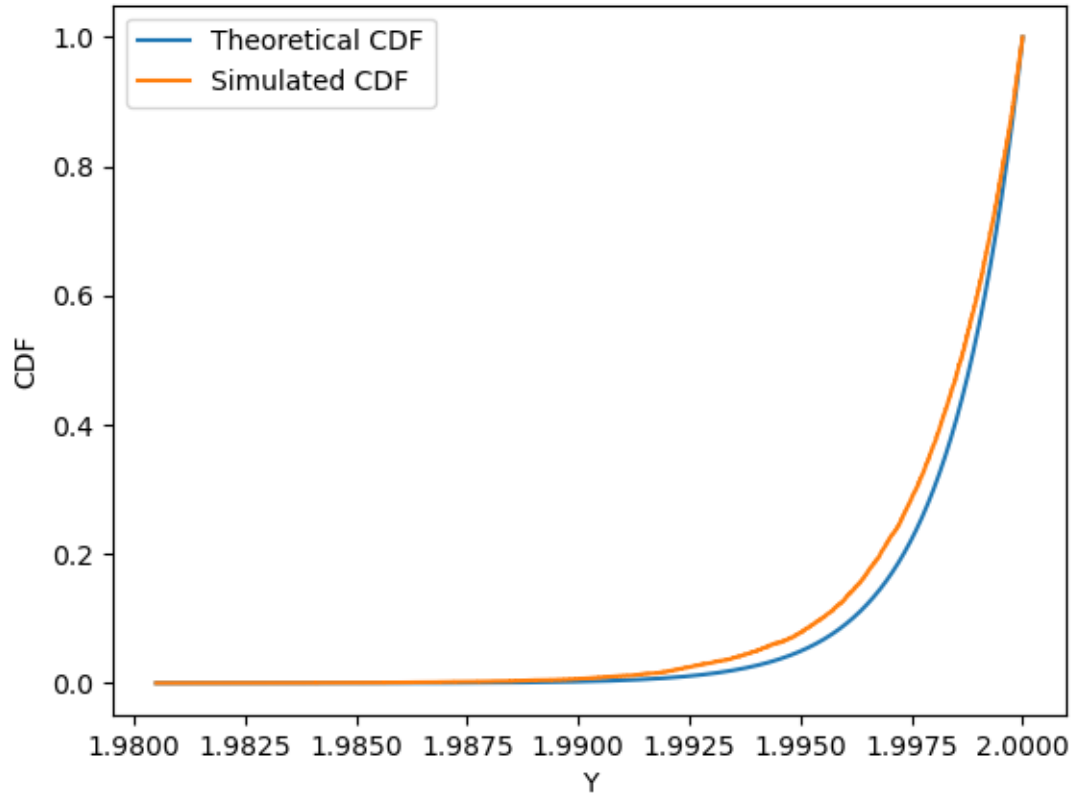


Figure 1: Theoretical and Simulated CDFs

Steps for the Simulation using C:

1. Include Header Files: Start the C program by including necessary header files, particularly '`<stdio.h>`' to perform standard input and output functions.
2. Declare Variables: Declare variables for sample size, number of simulations and  $\theta$ .
3. Open the file: Use '`fopen`' to open a file for writing the simulated data. The file is named '`estimations.txt`'.
4. Simulate data: Use loop to perform multiple simulations. In each iteration, generate random samples, calculate the estimator, and store the estimations in an array.

5. Write data to file: Inside the loop, use ‘fprintf’ to write each estimation to the opened file.
6. Close the file: After all simulations are complete and data is written, close the file using ‘fclose’ to save the data.

Python code for Plotting and Analysis:

Steps:

1. Include Required Libraries: In Python code, include libraries like NumPy for data manipulation and Matplotlib for plotting.
2. Load data: Use NumPy to load the data from the file ‘estimations.txt’ into a python array or NumPy array.
3. Define theoretical CDF: Define the function that calculates the theoretical CDF.
4. Sort the data: Sort the loaded data in ascending order.
5. Calculate the empirical CDF: Calculate the empirical CDF for the sorted data by dividing the rank of each value by the total number of data points.
6. Plot the CDFs: Use Matplotlib to create a plot that compares the theoretical and simulated CDFs.