CS215 Assignment 3
Question 2

Atishay Jain - 210050026
Gohil Megh Hiteshkumar - 210050055

November 1, 2022

1 Introduction

For each $N \in \{5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4\}$, we repeated the experiment for $M \geq 100$ times. The experiment includes generating a data sample $x_1, x_2, ...x_N$ of N points from a Uniform Distribution on $[0, 1]$ and Transform this data x to generate a transformed data sample y, where for each datum the transformed data is defined as -

$$y = -\frac{1}{\lambda} log(x)$$

where λ is a parameter, and $\lambda = 5$ is used for above data generation.
Now, the transformed data y has some distribution with parameter λ. We find its analytical form and given this distribution, we have to estimate λ using different estimation methods. Finally, we calculated Relative errors between true value of λ and its estimates $\hat{\lambda}$. Relative Error calculated as -

$$Error = \frac{|\hat{\lambda} - \lambda_{true}|}{\lambda_{true}}$$

2 Analytical Form of Transformed Data

We have transformation fucntion as -

$$y = g(x) = -\frac{1}{\lambda} log(x)$$

Since, $g(x)$ is a monotonically decreasing function for $x > 0$, the PDF of y can be derived by applying Transformation of Random Variable formula as follows -

$$x = g^{-1}(y) = e^{-\lambda y}$$

$$P_Y(y) = P_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1} y \right| \tag{1}$$

As x is a Uniform Distribution in $[0, 1]$,

$$P_X(x) = \begin{cases} 1 & , 0 \leq x \leq 1 \\ 0 & , otherwise \end{cases}$$

Putting $P_X(x)$ in eq.(1) and solving, we get -

$$P_Y(y) = \begin{cases} \lambda e^{-\lambda y} & , y \geq 0 \\ 0 & , otherwise \end{cases}$$

3 Maximum Likelihood Estimate of λ

For ML Estimate of λ, given data $y_1, y_2, ...y_N$ from distribution $P_Y(y)$, we need to find $\hat{\lambda}^{ML}$ which maximizes the Likelihood function. Solving it for $y \geq 0$ (as $P_Y(y)$ is 0 for $y < 0$) -

$$\begin{aligned} L(\lambda; y_1, ..., y_N) &= P_Y(y_1, y_2, ..., y_N|\lambda) \\ &= \prod_{i=1}^{N} P_Y(y_i|\lambda) \\ &= \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i} \end{aligned}$$

For simplifying calculation, we are taking *log* of Likelihood function and maximizing it wrt. λ. As *log* is a monotonically increasing function, maximizing *log* of Likelihood is equivalent to maximizing Likelihood function wrt. λ

$$\log(L(\mu; x_i, ..., x_N)) = N\log(\lambda) - \lambda \sum_{i=1}^{N} y_i$$

Differentiating above expression wrt. λ and equating it is to zero -

$$\frac{d}{d\lambda}\left(N\log(\lambda) - \lambda \sum_{i=1}^{N} y_i\right) = 0$$

$$\frac{N}{\lambda} - \sum_{i=1}^{N} y_i = 0$$

$$\therefore \hat{\lambda}^{ML} = \frac{N}{\sum_{i=1}^{N} y_i}$$

4 Maximum Posterior Mean Estimate of λ

- Joint Likelihood function is same as $P_Y(y_1, y_2, ..., y_N|\lambda)$ as calculated for MLE above -

$$P_Y(y_1, y_2, ..., y_N|\lambda) = \begin{cases} \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i} & , y_i \geq 0 \ \forall i \\ 0 & , otherwise \end{cases}$$

- Using a Gamma Prior for λ with parameters α and β -

$$P(\lambda|\alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

- So, Posterior Function is -

$$\begin{aligned} P(\lambda|y_1, y_2, ..., y_N) &= \frac{P_Y((y_1, y_2, ..., y_N|\lambda)P(\lambda)}{P(y_1, y_2, ..., y_N)} \\ &= \frac{P_Y((y_1, y_2, ..., y_N|\lambda)P(\lambda)}{\int_\lambda P_Y((y_1, y_2, ..., y_N|\lambda)P(\lambda) \cdot d\lambda} \\ &= \frac{\lambda^N e^{-\lambda \sum_{i=1}^{N} y_i} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_\lambda \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i} \lambda^{\alpha-1} e^{-\beta\lambda} \cdot d\lambda} \end{aligned}$$

Therefore, the Posterior Mean is -

$$E_{P(\lambda|y_1,y_2,...,y_N)}[\lambda] = \int_\lambda \lambda P(\lambda|y_1,y_2,...,y_N)\cdot d\lambda$$
$$= \frac{\int_0^\infty \lambda^{N+\alpha} e^{-\lambda(\beta+\sum_{i=1}^N y_i)}\cdot d\lambda}{\int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda(\beta+\sum_{i=1}^N y_i)}\cdot d\lambda}$$

Using the formula for Gamma function of $\int_0^\infty x^{\alpha-1}e^{-\beta x}\cdot dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$,

$$E_{P(\lambda|y_1,y_2,...,y_N)}[\lambda] = \frac{\Gamma(N+\alpha+1)/(\beta+\sum_{i=1}^N y_i)^{N+\alpha+1}}{\Gamma(N+\alpha)/(\beta+\sum_{i=1}^N y_i)^{N+\alpha}}$$

Using expression of Gamma function $\Gamma(x) = (x-1)!$ and solving further, we get the formula for **Posterior Mean** as -

$$\hat{\lambda}^{PosteriorMean} = E_{P(\lambda|y_1,y_2,...,y_N)}[\lambda] = \frac{N+\alpha}{(\sum_{i=1}^N y_i)+\beta}$$

5 Boxplot of Relative Errors

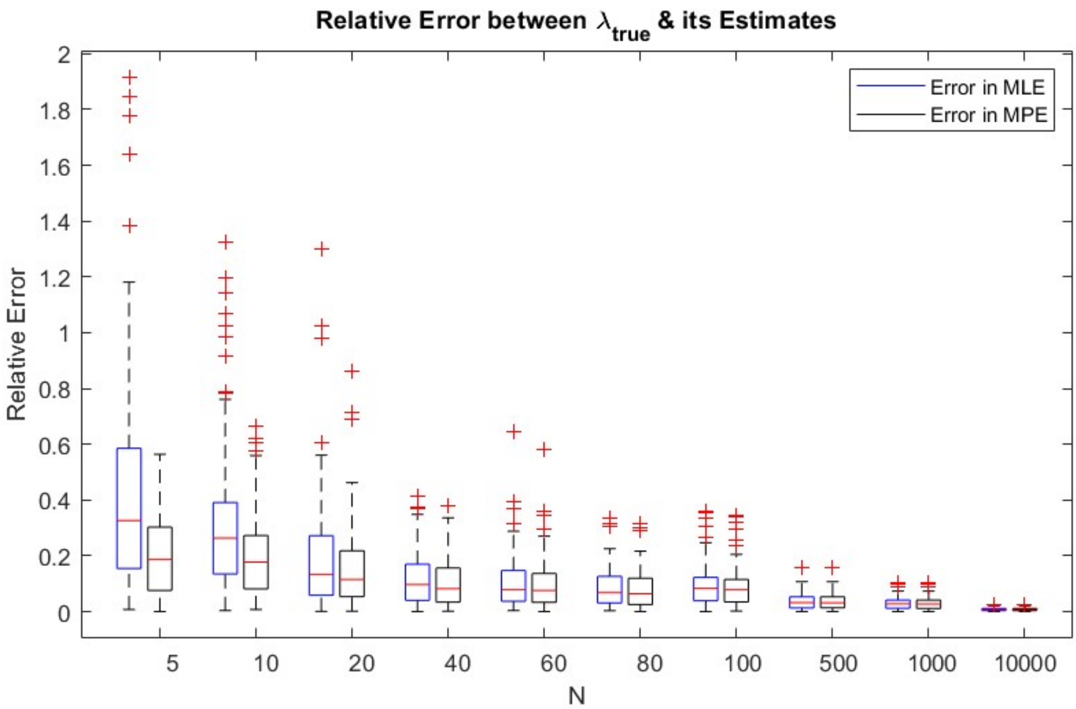

Figure 1: Boxplot of Relative errors in both the estimates for each N

The above Boxplot shows the Relative errors between true value and estimated values of the parameter λ. The Blue ones are the Boxplot for Maximum Likelihood Estimation and the Black ones are for Maximum Posterior Estimation of λ. As seen from the plot, the error in values estimated using Posterior Mean are relatively lesser.

6 Interpretation of Plot

6.1 What happens when N increases?

- We can see that the spread distribution of error decreases as the data size i.e. N increases.
- The median of distribution of error decreases as the data size increases.
- We can observe that the differences in the spread of errors for both the estimates are more at small values of N
- But as N increases, all the spread are almost nearly equal and are very small, i.e. both estimates tend to true mean at larger values of N.

6.2 Preferable Estimate

- As we can see from the box-plot graph at large values of N, the box-plot are nearly same for both estimates.
- So it means that both of the estimates give almost same estimates at large values of N. So any one of them is preferable at large values of N.
- But at lower values of N, we can see that the spread of error follows the trend -

$$\text{Maximum Likelihood estimate} > \text{Maximum Posterior Mean estimate.}$$

- Hence considering all values of N, Preferable estimate is the **Maximum Posterior Estimate** for Posterior Mean with Prior taken as Gamma function.