CS215 Assignment3 Problem 2

Parth Pujari and Anish Kulkarni

October 2022

1 Question

Use the Matlab function rand() to generate a data sample of N points from the uniform distribution on $[0, 1]$. Transform the resulting data x to generate a transformed data sample where each datum $y = \left(\frac{1}{\lambda}\right) log(x)$ with $\lambda = 5$. The transformed data y will have some distribution with parameter λ; what is its analytical form ? Use a Gamma prior on the parameter , where the Gamma distribution has parameters $\alpha = 5.5$ and $\beta = 1$.
Consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 103, 104$. For each sample size N, repeat the following experiment $M \geq 100$ times: generate the data, get the maximum likelihood estimate $\hat{\lambda}^{ML}$, get the Bayesian estimate as the posterior mean $\hat{\lambda}^{\text{PosteriorMean}}$, and measure the relative errors $|\hat{\lambda} - \lambda_{true}|/\lambda_{true}$ for both the estimates.

2 Formula for posterior mean

Solution-
Note that $y = g(x) = \left(\frac{-1}{\lambda}\right) log(x)$ has pdf given by

$$f(y) = -P(g^{-1}(y)) \times \frac{d}{dy}g^{-1}(y) = -1 \times \frac{d}{dy}e^{-\lambda y} = \lambda e^{-\lambda y}$$

Hence the Likelihood function is

$$\prod_{i=1}^{N} \lambda e^{-\lambda Y_i} = \lambda^N e^{-N\lambda\overline{Y}}$$

The Gamma distribution with $\alpha = 5.5$ and $\beta = 1$ is $C\lambda^{4.5}e^{-\lambda}$. So Posterior distribution is

$$C\lambda^N e^{-N\lambda\overline{Y}}\lambda^{4.5}e^{-\lambda} = C\lambda^{N+4.5}e^{-\lambda(1+N\overline{Y})}$$

This is a Gamma distribution with $\alpha = N + 5.5$ and $\beta = 1 + N\overline{Y}$ and hence has mean

$$\frac{\alpha}{\beta} = \frac{N + 5.5}{1 + N\overline{Y}}$$

3 Plotting and Results

For each sample size N, 100 trials were calculated and their mean $\bar{x}$ was used to find the estimate as above. The error was calculated as

$$|Est - \lambda_{true}|/\lambda_{true}$$

The box plots for each estimate (maximum likelihoodm and Posterior mean in that order) for each sample size are shown below.

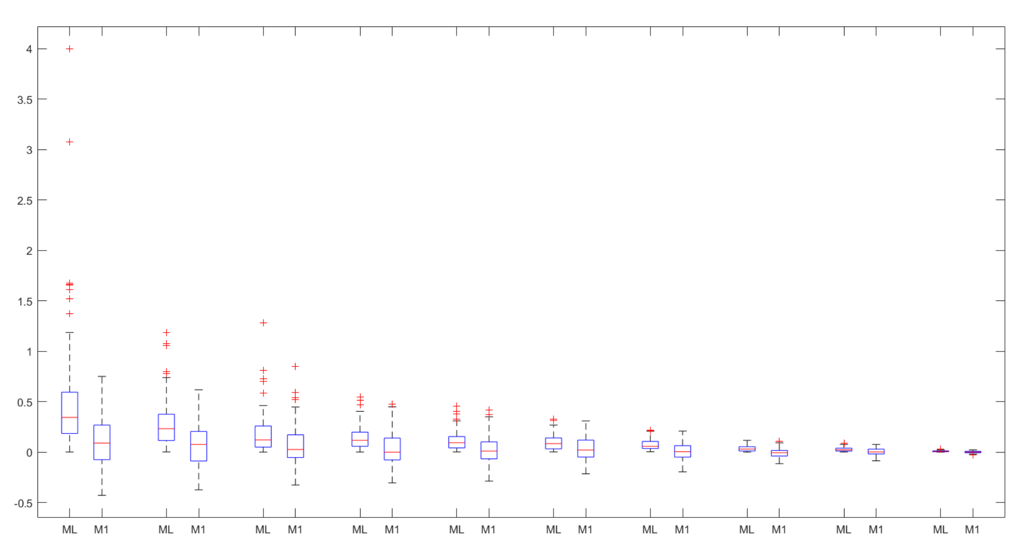

Figure 1: Boxplots for Estimators

4 Interpretation

As the sample size N increases, owing to the law of large numbers, the estimate errors decrease. The posterior mean is closer to the parameter however it has a larger variance.
Owing to better accuracy the posterior mean is the better estimator