Bayesian Estimation

Assignment 3

Ankan Sarkar (210050013)
Soham Joshi (210051004)

Contents

Introduction

This is the report of Assignment-3 of the course CS215[Awa22] offered in the autumn semester of '22 in IIT Bombay, by Prof. Suyash Awate. In this report, we will cover the solutions, along with empirical observations and how well they align with the existing theory. We have coded this assignment using MATLAB. The entire code can be accessed in this repository[22] under the folder "code", the graphs and results are given under the folder "results" and this report can be found under the folder "report". Sections $1, \cdots, 3$ of this report correspond to questions $1, \cdots, 3$ of the problem statement. So without further ado, let's start exploring.

1 ML vs MAP estimates

In this section, we shall compare the accuracy of maximum likelihood estimates, and the maximum-a-posteriori estimates given uniform and Gaussian priors given by

1. Gaussian prior with $\mu_{prior} = 10.5$ and $\sigma_{prior} = 1$
2. Uniform prior over $[9.5, 11.5]$

A Gaussian distribution is generated with mean $\mu_{true} = 10$ and standard deviation $\sigma_{true} = 4$. We consider the problem of using the data to get an estimate $\hat{\mu}$ of this Gaussian mean, assuming it is unknown, when the standard deviation σ_{true} is known.
Now, we consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$. For each sample size N, we repeat the following experiment $M \geq 100$ times; generate the data, get the maximum likelihood estimate $\hat{\mu}^{ML}$, get the maximum-a-posteriori estimates $\hat{\mu}^{MAP1}$ and $\hat{\mu}^{MAP2}$, and measure the relative error $\frac{\|\hat{\mu}-\mu_{true}\|}{\mu_{true}}$ for all the three estimates.

Now, we plot the grouped box plot and obtain the following results:

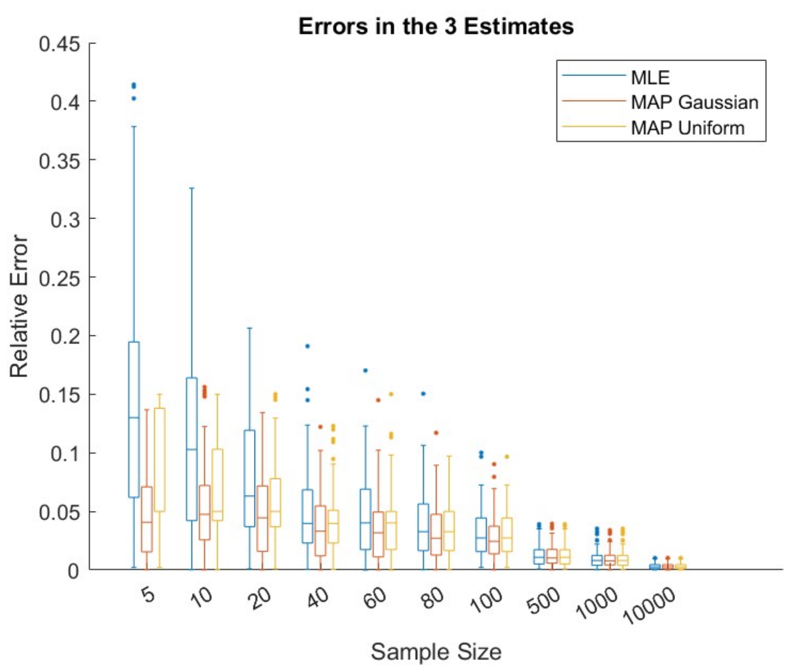

Figure 1: Comparing Errors via Box plots

1. As the value of N increases, we observe that errors in all three cases converge to 0.

2. We observe that MAP Estimate with Gaussian Prior gives lower errors for most values of N, however, all three Estimates differ only marginally for large values of N. Hence MAP Estimate with Gaussian Prior is the preferred estimator.

2 ML vs Posterior Mean

In this section, we generate a data sample of N points from the uniform distribution on $[0,1]$. Now, we transform the resulting data x to generate a transformed data sample where each datum is given by

$$Y := \frac{-1}{\lambda} log(X) \text{ where } \lambda = 5 \tag{1}$$

Now, let us compute the PDF of y denoted as $G(y)$ analytically. Let the PDF of x be given as:

$$P(x) = \begin{cases} 1, 0 < x < 1 \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

Now, we have

$$Y = g(X) = \frac{-1}{\lambda} log(X) \tag{3}$$

$$\Rightarrow X = g^{-1}(Y) = e^{-\lambda Y} \tag{4}$$

Also,

$$G(Y) = P(g^{-1}(Y)) \| \frac{dg^{-1}(Y)}{dY} \| \tag{5}$$

Since, $g^{-1}(Y)$ lies in $(0,1)$, so $P(g^{-1}(Y))$ is always 1

$$\Rightarrow G(Y) = \| \frac{de^{-\lambda Y}}{dY} \| \tag{6}$$

$$\Rightarrow G(Y) = \lambda e^{-\lambda Y} \tag{7}$$

Now, we derive the expression for posterior mean, denoted as $H(\lambda)$

$$gamma(\lambda; \alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \tag{8}$$

$$\Rightarrow H(\lambda) = \frac{(\prod_{i=1}^{N} G(y_i|\lambda))G(\lambda)}{P(y_1, \cdots, y_N)} \tag{9}$$

$$= \frac{\lambda^n e^{-\lambda \sum_{i=1}^{N} y_i} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}}{\int_0^\infty \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}} \tag{10}$$

$$= \frac{\lambda^{N+\alpha-1} e^{-\lambda(\beta + \sum_{i=1}^{N} y_i)}}{\int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda(\beta + \sum_{i=1}^{N} y_i)}} \tag{11}$$

$$= \frac{(\beta + \sum_{i=1}^{N} y_i)^{N+\alpha} \lambda^{N+\alpha-1} e^{-\lambda(\beta + \sum_{i=1}^{N} y_i)}}{\Gamma(N+\alpha)} \tag{12}$$

We observe that the resulting distribution is another gamma distribution, with $\alpha^{'} = N+\alpha$, $\beta^{'} = \beta + \sum_{i=1}^{N} y_i$. Using the fact that the mean of the gamma distribution is $\frac{\alpha}{\beta}$ we get :

$$mean = \frac{\alpha^{'}}{\beta^{'}} \tag{13}$$

$$\boxed{mean = \frac{N+\alpha}{\sum_{i=1}^{N} y_i + \beta}} \tag{14}$$

Moreover, we will use the fact that the maximum likelihood estimate is given as $\frac{N}{\sum_{i=1}^{N} y_i}$, putting values of X generated from a uniformly random distribution on $[0,1]$ for different values of N, and performing this experiment $M = 100$ times, we get the following grouped box plot:

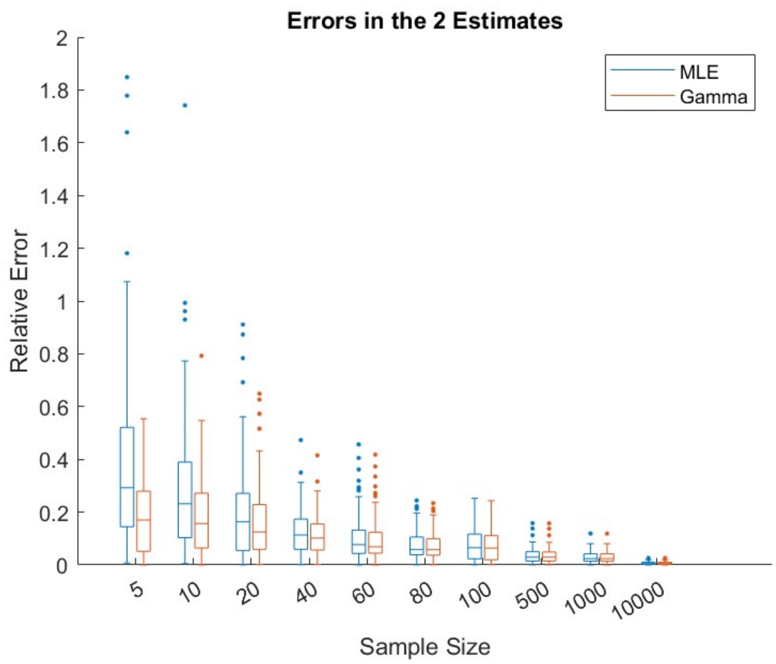

Figure 2: Comparing Errors via Box plots

1. As the value of N increases, we observe that errors in both cases converge to 0
2. We observe that Posterior mean estimator gives us significantly lower errors for small values of N, where as for large values of N, it is marginally better. Hence, Posterior mean estimator is the preferred estimator here.

3 Pareto distribution prior

In this section, we shall examine the MAP, ML and Posterior Mean estimates for the Pareto distribution prior. Let's consider a random variable X with a uniform distribution over $[0, \theta]$, where the parameter θ is unknown. Let $P(\theta; \alpha, \theta_m)$ denote the Pareto distribution given as :

$$P(\theta) = \begin{cases} c(\frac{\theta_m}{\theta})^{\alpha}, \theta \geq \theta_m \\ 0, \text{ otherwise} \end{cases} \tag{15}$$

First, let us compute the constant c. Using the fact that θ is a continuous random variable, $\alpha > 1$, and integrating the PDF, we get :

$$\int_{\theta_m}^{\infty} c(\frac{\theta_m}{\theta})^{\alpha} d\theta = 1 \Rightarrow \boxed{c = \frac{\alpha - 1}{\theta_m}} \tag{16}$$

3.1 ML Estimate

Now, let us compute the maximum likelihood estimate. Since X is uniform over $[0, \theta]$, we get the likelihood function as :

$$L(\theta) = \prod_{i=1}^{N} P(X_i; \theta) \tag{17}$$

$$\Rightarrow L(\theta) = \begin{cases} (\frac{1}{\theta})^N, \forall_i 0 \le X_i \le \theta \\ 0, \text{ otherwise} \end{cases} \tag{18}$$

This likelihood function is maximised when θ is minimised. Hence,

$$\boxed{\hat{\theta}^{ML} = \max X_i} \tag{19}$$

3.2 MAP Estimate

Let the posterior PDF be given by $G(\theta)$, is non-zero only for case 1 of posterior distribution and Likelihood function (i.e. $\forall_i 0 \le X_i \le \theta$ and $\theta \ge \theta_m$). Let $K = \max X_i, \theta_m$ and using the formula for the posterior distribution, we get

$$G(\theta) = \frac{(\frac{1}{\theta})^N * c(\frac{\theta_m}{\theta})^\alpha}{\int_K^\infty (\frac{1}{\theta})^N * c(\frac{\theta_m}{\theta})^\alpha d\theta} \tag{20}$$

Cancelling out constant terms and using 16 to evaluate the denominator, we get

$$G(\theta) = \frac{N + \alpha - 1}{K} (\frac{K}{\theta})^{\alpha + N} \tag{21}$$

That is, including all cases :

$$G(\theta) = \begin{cases} \frac{N+\alpha-1}{K}(\frac{K}{\theta})^{\alpha+N}, \text{ if } \forall_i 0 \le X_i \le \theta, \theta \ge \theta_m \\ 0, \text{ otherwise} \end{cases} \tag{22}$$

Since, $G(\theta)$ is monotonically decreasing, and θ_m, X_i must be upper bounded by θ we get :

$$\boxed{\hat{\theta}^{MAP} = \max X_i, \theta_m} \tag{23}$$

We know that maximum likelihood is a **consistent estimator** for a uniform distribution. Hence,

$$\forall \epsilon > 0, \exists_N \| \max X_i - \theta_{true} \| < \epsilon, \forall i > N \tag{24}$$

Let $\theta_m < \theta_{true}$. Use the fact that $X_i \le \theta_{true}$, then take $\epsilon = \theta_{true} - \theta_m$, giving us that $\exists N$ such that $\theta_{true} - \max X_i < \epsilon, \forall i > N$, hence, for any such i, $X_i > \theta_m$. Hence, we get the following results as $N \to \infty$:

$$\hat{\theta}^{MAP} = \begin{cases} \theta_m, \text{ if } \theta_m \ge \theta_{true} \\ \theta_{true}, \text{ otherwise} \end{cases} \tag{25}$$

Hence, $\hat{\theta}^{MAP} \to \hat{\theta}^{ML}$ if and only if $\boxed{\theta_m < \theta_{true}}$.

This behaviour is not very desirable since for some choices of the prior, i.e. the parameter θ_m the MAP Estimate does NOT actually converge to the ML Estimate.

3.3 Posterior Mean Estimate

Now, $G(\theta)$ has PDF given as

$$G(\theta) = \begin{cases} \frac{N+\alpha-1}{K}(\frac{K}{\theta})^{\alpha+N}, & \text{if } \forall_i 0 \leq X_i \leq \theta, \theta \geq \theta_m \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

Now, let us estimate the mean of this distribution. (using $\alpha > 1, N \geq 1$)

$$\hat{\theta}^{\text{PosteriorMean}} = \int_K^\infty \theta \frac{N+\alpha-1}{K}(\frac{K}{\theta})^{\alpha+N} d\theta \tag{27}$$

$$= \int_K^\infty (N+\alpha-1)(\frac{K}{\theta})^{\alpha+N-1} d\theta \tag{28}$$

$$= \boxed{\frac{N+\alpha-1}{N+\alpha-2} \max X_i, \theta_m} \tag{29}$$

As $N \to \infty$, we get $\frac{N+\alpha-1}{N+\alpha-2} \to 1$. Hence, similar to result 25,as $N \to \infty$ we get :

$$\hat{\theta}^{\text{PosteriorMean}} = \begin{cases} \theta_m, & \text{if } \theta_m \geq \theta_{true} \\ \theta_{true}, & \text{otherwise} \end{cases}$$

Hence, $\hat{\theta}^{\text{PosteriorMean}} \to \hat{\theta}^{\text{ML}}$ if and only if $\boxed{\theta_m < \theta_{true}}$.

Again, this behaviour is not very desirable since for some choices of the prior, i.e. the parameter θ_m the Posterior Mean Estimate does NOT actually converge to the ML Estimate.

Bibliography

[Awa22] Suyash Awate. *Data Analysis and Interpretation*. University Lecture. 2022.

[22] *Github repository*. 2022. URL: https://github.com/Ihsoj-Mahos/CS215-Assignment3.