CS 215 Assignment-3

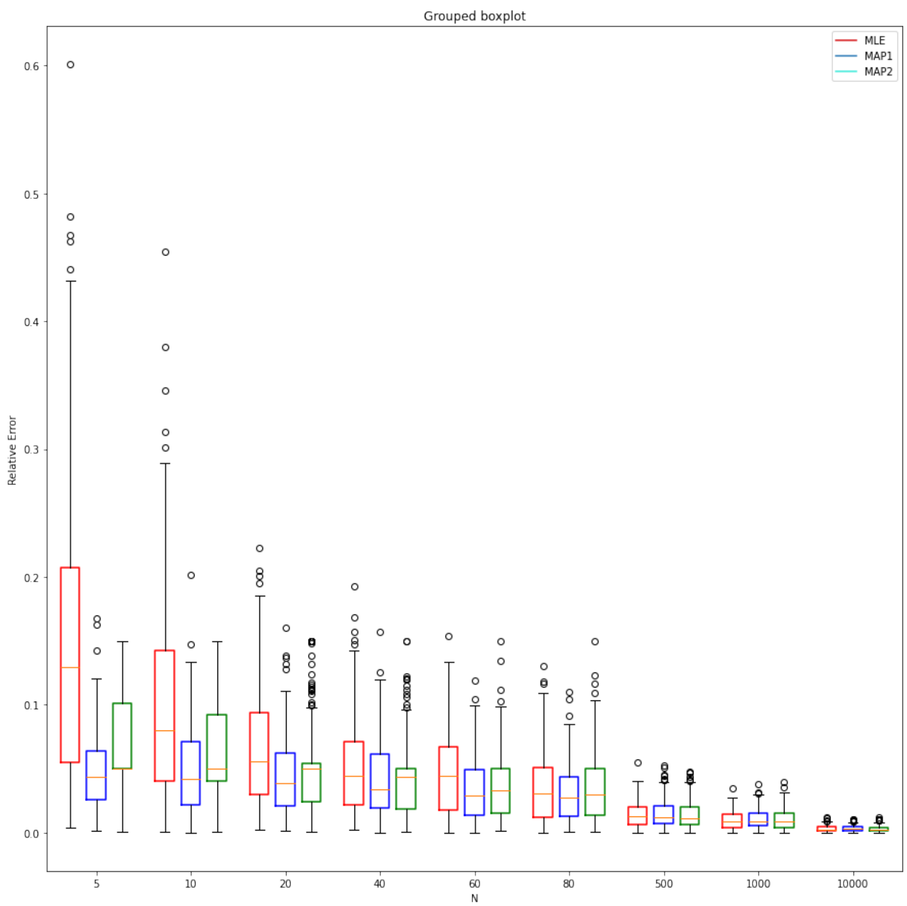
Aryan Mathe & Yash Ruhatiya

October 2022

Contents

1 Estimation of Mean from given Standard Gaussian

1.1 Boxplot of relative errors for different N



MAP1 $\rightarrow$ Gaussian prior on mean

MAP2 $\rightarrow$ Uniform prior on mean

1.2 Observations

The following observations can be made from the above plots.

- As N increases the error in the estimate gradually decreases with N. Also, MAP estimate and ML estimate both tends to **true mean**, i.e. μ_{true}
- From the boxplot, it can be observed that the error produced in the case of posterior being generated from the gaussian prior and gaussian likelihood function will be preferred because of fewer relative errors.

2 Posterior Mean estimation for a given transformation

2.1 Formula for Posterior mean

$$\text{Transformation:} \quad y = -\frac{1}{\lambda} log(x)$$

Now, we have to find the probability distribution for the transformed random variable.

$$q(y) = p(g^{-1}(y)) \cdot |\frac{d}{dy}(g^{-1}(y))|$$

where $q(y)$ is the probability distribution for transformed random variable y.

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

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

$$q(y) = \lambda e^{-\lambda y}$$

The maximum likelihood estimator for the following distribution can be calculated as follows,

$$L(y_1, y_2, ..., y_n|\lambda) = \prod_{i=1}^{N} q(y_i)$$

$$L(y_1, y_2, ..., y_n|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{N} y_i}$$

Now,

$$log(L) = Nlog(\lambda) - \lambda \sum_{i=1}^{N} y_i$$

Using $\frac{dL}{d\lambda} = 0$, to get the **ML estimate**

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

$$\hat{\lambda}_{MLE} = \frac{N}{\sum_{i=1}^{N} y_i}$$

Now **MAP** estimate can be found using the given prior distribution $P(\lambda|y_1, y_2, ..., y_n)$ and likelihood function.

$$P(\lambda|y_1, y_2, ..., y_n) = \frac{P(y_1, y_2, ..., y_n|\lambda) \cdot q(\lambda)}{\int_{\lambda} P(y_1, y_2, ..., y_n, \lambda)}$$

$$P(\lambda|y_1,y_2,...,y_n) = \frac{k\lambda^{N+\alpha-1}e^{-\lambda(\beta+\sum y_i)}}{\int_\lambda k(\prod_{i=1}^N \lambda e^{-\lambda y_i})\lambda^{\alpha-1}e^{-\beta\lambda}d\lambda} \quad ,k = \frac{\beta^\alpha}{\tau(\alpha)}$$

$$I(N+\alpha-1) = \int_0^\infty \lambda^{N+\alpha-1}e^{-\lambda(\beta+\sum y_i)}d\lambda$$

Now using integration by parts we get the following recurrence relation,

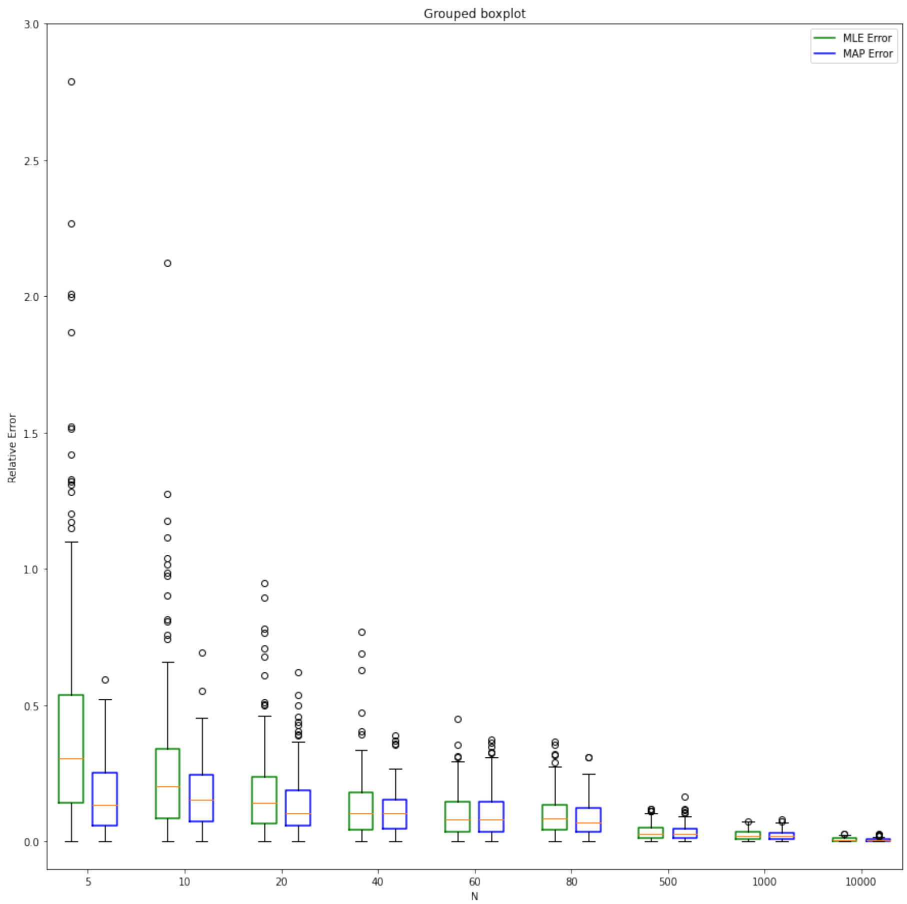
$$\frac{I(N+\alpha-1)}{I(N+\alpha-2)} = \frac{N+\alpha-1}{\beta+\sum y_i}$$

Now the posterior distribution can be written as,

$$P(\lambda|y_1,y_2,...,y_n) = \frac{\lambda^{N+\alpha-1}e^{-\lambda(\beta+\sum y_i)}}{I(N+\alpha-1)}$$

$$\begin{aligned}
\hat{\lambda}^{PosteriorMean} &= \int_0^\infty \lambda P(\lambda|y_1,y_2,...,y_n)d\lambda \\
&= \int_0^\infty \frac{\lambda^{N+\alpha}e^{-\lambda(\beta+\sum y_i)}d\lambda}{I(N+\alpha-1)} \\
&= \frac{I(N+\alpha)}{I(N+\alpha-1)} \\
\hat{\lambda}^{PosteriorMean} &= \frac{N+\alpha}{\beta+\sum y_i}
\end{aligned}$$

2.2 Plot for Relative Errors



2.3 Observations

The following interpretations can be made from the boxplots shown above

- Since the relative error between true mean and the estimates continuously decreases, it can be observed that as the $N \to \infty$, both estimates will converge to the true value of the parameter.
- Also, it can be seen that the relative errors in the case of MAP estimate is less than ML estimate for most values of N. Hence MAP estimate will be the desired estimate.

3 Parameter Estimation using MAP and ML estimator

3.1 ML Estimate and MAP Estimate

Maximum Likelihood Estimate

To find the **ML** estimate we will first calculate the likelihood function using uniform probability distribution over the range $[0, \theta]$

$$\begin{aligned} P(x_1, x_2, ..., x_n|\theta) &= \prod_{i=1}^{n} P(x_i|\theta) \\ &= (\frac{1}{\theta})^n \end{aligned}$$

Now,

$$\begin{gathered} L(X|\theta) = P(x_1, x_2, ..., x_n|\theta) \\ L(X|\theta) = (\frac{1}{\theta})^n \\ \log(L) = -n\log(\theta) \\ \frac{dL}{d\theta} = -\frac{nL}{\theta} \end{gathered}$$

which is always negative, hence to maximize the value of Likelihood function, θ should be estimated as the $\max_i(x_i)$.

$$\hat{\theta}_{MLE} = max(x_i|1 \leq i \leq n)$$

Maximum a Posteriori Estimate

The prior distribution i.e, the **Pareto** distribution is given as

$$P(\theta) = \begin{cases} 0 & \theta < \theta_m \\ (\frac{\theta_m}{\theta})^\alpha & \theta \geq \theta_m \end{cases}$$

Now ,

$$P(\theta|x_1, x_2, ..., x_n) = \frac{P(x_1, x_2, ..., x_n|\theta)P(\theta)}{P(x_1, x_2, ..., x_n)}$$

It can be observed from the given prior distribution that the MAP estimate has to be greater than θ_m, else the prior would be zero, and hence the posterior would be zero, which is not suitable as we wish to maximise the posterior distribution.

Also, it can be seen that θ should be greater than each of the $x_i^{\prime}s$, else the posterior would be zero, and hence would not give us the estimate.

So we must enforce that, $\forall i, x_i \leq \theta$ and $\theta \geq \theta_m$ which is equivalent to saying that ,

$$\theta \geq max(x_1, x_2, ..., x_n, \theta_m)$$

$$\text{Let , } X_m = max(x_1, x_2, ..., x_n)$$

So , we have,

$$\theta \geq max(X_m, \theta_m)$$

So now proceeding with this constraint on theta we have,

$$P_{Posterior}(\theta|x_1, x_2, ..., x_n) = \frac{(\prod_{i=1}^{n} P(x_i|\theta))P(\theta)}{\int_{max(X_m,\theta_m)}^{\infty}(\prod_{i=1}^{n} P(x_i|\theta))P(\theta)d\theta}$$

$$= \frac{(\frac{1}{\theta})^n(\frac{\theta_m}{\theta})^\alpha}{\int_{max(X_m,\theta_m)}^{\infty}(\frac{1}{\theta})^n(\frac{\theta_m}{\theta})^\alpha d\theta}$$

$$P_{Posterior}(\theta|x_1, x_2, ..., x_n) = \frac{\frac{\theta_m^\alpha}{\theta^{n+\alpha}}}{\int_{max(X_m,\theta_m)}^{\infty}\frac{\theta_m^\alpha}{\theta^{n+\alpha}}d\theta}$$

So , now to find the MAP estimate for θ, we need to find the θ that maximises the posterior probability distribution, i.e

$$\hat{\theta}_{MAP} = argmax_\theta \ P_{posterior}(\theta|x_1, x_2, ..., x_n)$$

Now,

$$P(X|\theta) = P(\theta|x_1, x_2, ..., x_n)$$

$$log(P) = -(n + \alpha)log(\theta) + constant$$

$$\frac{dP}{d\theta} = -\frac{(n + \alpha)P}{\theta}$$

which is always negative, hence to maximize the value, we need to minimize θ. So, the MAP estimate would be the least possible value of θ that satisfies all the equations above, i.e

$$\hat{\theta}_{MAP} = max(X_m, \theta_m)$$

3.2 Relation between ML and MAP

Let us first get the condition for $\hat{\theta}_{ML}$ and $\hat{\theta}_{MAP}$ being equal,

$$\hat{\theta}_{ML} = \hat{\theta}_{MAP}$$

$$X_m = max(X_m, \theta_m)$$

$$i.e \ X_m \geq \theta_m$$

So, now let us check the probability of this happening, and evaluate that at infinity. To do this we can see that the only case where $X_m < \theta_m$ would be when all of the $x_i's$ generated are less than θ_m and, if $X_m \geq \theta_m$ then,

$$\begin{aligned}
P(X_m \geq \theta_m) &= 1 - P(X_m < \theta_m) \\
&= 1 - \prod_{i=1}^{n} P(\theta_m > x_i) \\
&= 1 - (\frac{\theta_m}{\theta})^n \\
P(\hat{\theta}_{ML} = \hat{\theta}_{MAP}) &= 1 - (\frac{\theta_m}{\theta})^n
\end{aligned}$$

Here, we had assumed that $\theta = X_m \& X_m > \theta_m$, hence $\frac{\theta_m}{\theta} < 1$, so

$$P(\hat{\theta}_{ML} = \hat{\theta}_{MAP}) \to 1 \ as \ n \to \infty$$

So , **Yes**, $\hat{\theta}_{MAP}$ tends to $\hat{\theta}_{ML}$ as the sample size tends to infinity. This is also desirable as for large datasets the prior holds little significance in determining the posterior distribution, and hence the MAP estimate is majorly dependent on the data generated and hence should tend to the ML estimate.

3.3 Mean Estimate of Posterior Distribution

$$\hat{\theta}^{PosteriorMean} = \int_{max(X_m,\theta_m)}^{\infty} \theta P(\theta|x_1, x_2, ..., x_n) d\theta + \int_{0}^{max(X_m,\theta_m)} \theta * 0 d\theta$$

Substituting the value of posterior distribution for $\theta > max(X_m, \theta_m)$ above , we get

$$\hat{\theta}^{PosteriorMean} = \int_{max(X_m,\theta_m)}^{\infty} \frac{\theta \frac{\theta_m^\alpha}{\theta^{n+\alpha}} d\theta}{\int_{max(X_m,\theta_m)}^{\infty} \frac{\theta_m^\alpha}{\theta^{n+\alpha}} d\theta}$$

Now, since the denominator in the integrand is independent of θ, we can write,

$$\hat{\theta}^{PosteriorMean} = \frac{\int_{max(X_m,\theta_m)}^{\infty} \frac{\theta_m^\alpha}{\theta^{n+\alpha-1}} d\theta}{\int_{max(X_m,\theta_m)}^{\infty} \frac{\theta_m^\alpha}{\theta^{n+\alpha}} d\theta}$$

$$\hat{\theta}^{PosteriorMean} = \frac{\int_{max(X_m,\theta_m)}^{\infty} \theta^{-n-\alpha+1} d\theta}{\int_{max(X_m,\theta_m)}^{\infty} \theta^{-n-\alpha} d\theta}$$

Now, we have been given that $\alpha > 1$, also $n \geq 1$ as some data has to be generated for finding the posterior, so $n+\alpha > 2$, so we can directly use the standard integral result here that, $\int_{\theta}^{\infty} x^{-\alpha} dx = \frac{\theta^{-\alpha+1}}{-\alpha+1}$, for some constant $\alpha > 1$, giving us

$$\hat{\theta}^{PosteriorMean} = \frac{n+\alpha-1}{n+\alpha-2} max(X_m, \theta_m)$$

3.4 Relation between Posterior Mean and ML

As calculated above, we have,

$$\hat{\theta}^{PosteriorMean} = \frac{n+\alpha-1}{n+\alpha-2} max(X_m, \theta_m)$$

Now as $n \to \infty$, the fraction $\frac{n+\alpha-1}{n+\alpha-2}$ tends to 1 and hence

$$\hat{\theta}^{PosteriorMean} \to max(X_m, \theta_m)$$

that is,

$$\hat{\theta}^{PosteriorMean} \to \hat{\theta}_{MAP}$$

Now as proved in one of the previous sections,

$$\hat{\theta}_{MAP} \to \hat{\theta}_{ML} \;\; as \; n \to \infty$$

Hence,

$$\hat{\theta}^{PosteriorMean} \to \hat{\theta}_{ML} \;\; as \; n \to \infty$$

So,**Yes**, the Posterior mean estimate tends to the ML estimate as the sample size tends to infinity. This is also desirable, as for large datasets, all consistent estimators should approach the true value of the parameter and hence their estimates should approach each other.