CS215 Assignment 3
Question 1

Atishay Jain - 210050026
Gohil Megh Hiteshkumar - 210050055

November 1, 2022

1 Relative Errors in Estimates

For each $N \in \{5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4\}$, we repeated the experiment of generating data sample of N points $(x_1, x_2, x_3, ..., x_N)$ for $M \geq 100$ times from a Gaussian distribution $(G(\mu, \sigma^2))$. For this data we assumed that σ is known and μ is unknown. So, we have to find an estimate $\hat{\mu}$ of mean of the data using different estimation methods as shown below. And, finally we plotted Boxplot of relative error of estimate $\hat{\mu}$ w.r.t. true Mean μ_{true}, where relative error is calculated as -

$$Error = \frac{|\hat{\mu} - \mu_{true}|}{\mu_{true}}$$

1.1 Maximum Likelihood Estimate

For ML Estimate of Mean of a Gaussian $G(\mu, \sigma^2)$ given data $x_1, x_2, ...x_N$, we need to find μ^{ML} which maximizes the Likelihood function -

$$L(\mu; x_i, ..., x_N) = \prod_{i=1}^{N} P(x_i; \mu) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

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

$$log(L(\mu; x_i, ..., x_N)) = -\sum_{i=1}^{N} (\frac{(x_i - \mu)^2}{2\sigma^2} + log(\sigma\sqrt{2\pi}))$$

Differentiating above expression wrt. μ and setting it is to zero -

$$\frac{d}{d\mu}\left(\sum_{i=1}^{N} -\frac{(x_i - \mu)^2}{2\sigma^2} - log(\sigma\sqrt{2\pi})\right) = 0$$

$$\sum_{i=1}^{N} \frac{2(x_i - \mu)}{2\sigma^2} = 0 \implies \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\therefore \hat{\mu}^{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

1.2 Maximum A-Posteriori Estimate

For finding maximum a-posteriori estimate, we have to find a $\hat{\mu}^{MAP}$ that maximizes the Posterior function. The Posterier function is given as -

$$\begin{aligned} P(\mu|x_1, ..., x_N) &= \frac{P_{likelihood}(x_1, ..., x_N|\mu)P_{prior}(\mu)}{\int_\mu P(x_1, ..., x_N, \mu)d\mu} \\ &= \frac{\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} P_{prior}(\mu)}{\int_\mu P(x_1, ..., x_N, \mu)d\mu} \end{aligned}$$

Here the denominator is for normalisation and it is independent of μ, so maximizing the numerator will be equivalent to maximizing whole expression. For this, we will differentiate Posterior function's numerator wrt. μ and put it equal to 0. We have two cases for Prior function -

1.2.1 Using a Gaussian Prior Function

We have a Gaussian prior function with $\mu_{prior} = 10.5$ and $\sigma_{prior} = 1$ -

$$P_{prior}(\mu) = \frac{1}{\sigma_{prior}\sqrt{2\pi}} e^{-\frac{(\mu-\mu_{prior})^2}{2\sigma^2_{prior}}}$$

We put above Prior function in Posterior expression. So the numerator of the posterior looks like -

$$\left(\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \frac{1}{\sigma_{prior}\sqrt{2\pi}} e^{-\frac{(\mu-\mu_{prior})^2}{2\sigma^2_{prior}}}$$

Taking *log* of above expression and then differentiating,

$$\frac{d}{d\mu}\left[-\left(\sum_{i=1}^{N} log(\sigma\sqrt{2\pi}) + \frac{(x_i-\mu)^2}{2\sigma^2}\right) - log(\sigma_{prior}\sqrt{2\pi}) - \frac{(\mu-\mu_{prior})^2}{2\sigma^2_{prior}}\right] = 0$$

$$\sum_{i=1}^{N} \frac{(x_i-\mu)}{\sigma^2} + \frac{(\mu-\mu_{prior})}{\sigma^2_{prior}} = 0$$

On rearranging the above expression and solving further, we get,

$$\hat{\mu}^{MAP1} = \frac{(\sum_{i=1}^{N} x_i)\sigma^2_{prior} + \mu_{prior}\sigma^2}{N\sigma^2_{prior} + \sigma^2}$$

1.2.2 Using a Uniform Prior Function

We have a uniform prior function as -

$$P_{prior}(\mu) = \begin{cases} \frac{1}{2} & , 9.5 \le \mu \le 11.5 \\ 0 & , otherwise \end{cases}$$

We will put this Prior function in Posterior and maximize it wrt. μ. As the Prior PDF is 0 for all μ except $\mu \in [9.5, 11.5]$, we will consider the Posterior function in $\mu \in [9.5, 11.5]$ and look at three cases for values of μ obtained -

- If μ obtained is less than 9.5, then the Posterior will be maximized at $\mu = 9.5$ as it is zero for all $\mu < 9.5$
- If μ obtained $\in [9.5, 11.5]$, then we have Posterior as non-zero function and we will differentiate it and set it to zero. Numerator of posterior is -

$$\left(\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right) \frac{1}{2}$$

  Taking log of above function and maximizing it, we get,

$$\frac{d}{d\mu}\left(-log(2) + \sum_{i=1}^{N} -\frac{(x_i-\mu)^2}{2\sigma^2} - log(\sigma\sqrt{2\pi})\right) = 0$$

$$\sum_{i=1}^{N} \frac{2(x_i-\mu)}{2\sigma^2} = 0 \implies \mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- If μ obtained is greater than 11.5, then the Posterior will be maximized at $\mu = 11.5$ as it is zero for all $\mu > 11.5$

Therefore, we have,

$$\hat{\mu}^{MAP2} = \begin{cases} 9.5 & , \mu \leq 9.5 \\ \frac{1}{N}\sum_{i=1}^{N} x_i & , 9.5 \leq \mu \leq 11.5 \\ 11.5 & , \mu \geq 11.5 \end{cases}$$

where, $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$, i.e. the sample mean of data generated.

1.3 Boxplot of Relative Errors

We plotted the below Boxplot which shows the Relative errors for each N for each Estimate among MLE, MAP1 and MAP2 -

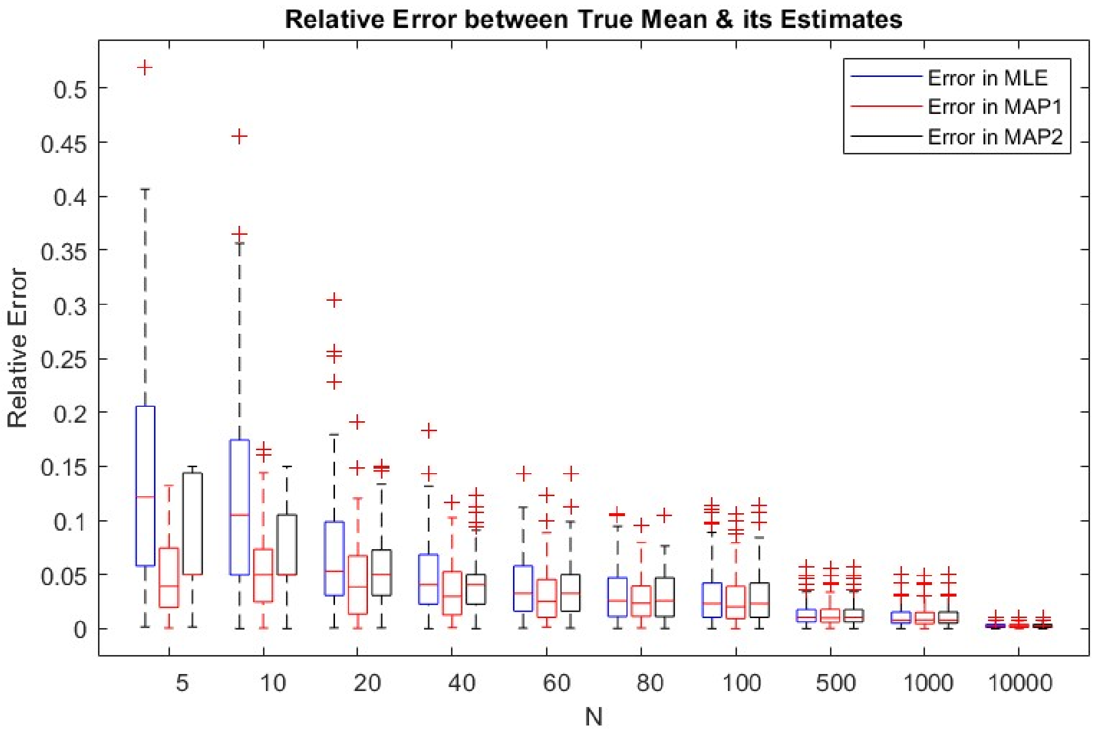

Figure 1: Boxplot of Relative errors in all the 3 estimates for each N

2 Interpretation of Plot

2.1 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 three 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. all of the three estimates tends to true mean at larger values of N.

2.2 Preferable Estimate

- As we can see from the box-plot graph at large values of N, the box-plot are nearly same for all three estimates.
- So it means that all three 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 -

$$\text{ML estimate} > \text{MAP2 estimate} > \text{MAP1 estimate.}$$

- Hence considering all the values of N, the Preferable estimate is **MAP1**, that is the **Maximum A-Posteriori Estimate** with Prior taken as a **Gaussian** Distribution.