CS-215-Assignment 3

Akshat Goyal (210050009) and Anshul Verma (210050015)

October 2022

Problem 1

1 ML-Estimate

We know that the likelihood function

$$
\begin{aligned}
L(x_1, x_2....x_N|\mu) &\propto \prod_{i=1}^{N} e^{\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&\propto e^{\sum_{i=1}^{N} \frac{(x_i-\mu)^2}{2\sigma^2}} \\
&\propto e^{\frac{\sum_{i=1}^{N}(x_i)^2+N\mu^2-2\sum_{i=1}^{N} x_i\mu}{2\sigma^2}} \\
&\propto e^{\frac{(\mu-\frac{\sum_{i=1}^{N}(x_i)}{N})^2}{\frac{2\sigma^2}{N}}} \\
&= G(\frac{\sum_{i=1}^{N}(x_i)}{N}, \frac{\sigma^2}{N})
\end{aligned}
$$

We can see that this is a gaussian in μ with mean as the sample mean. The gaussian takes its maximum value at the mean so the ML-Estimate is the sample average.

$$\hat{\mu}_{ML} = \frac{\sum_{i=1}^{N} X_i}{N}$$

2 Maximum A Posteriori Estimate 1

Here we take our prior to be a gaussian with

$$\mu_{prior} = 10.5 \text{ and } \sigma_{prior} = 1$$

$$Prior = P(\mu) = G(\mu_{prior}, \sigma^2_{prior})$$

We know likelihood

$$L(x_1, x_2....x_N|\mu) = G\left(\frac{\sum_{i=1}^{N}(x_i)}{N}, \frac{\sigma^2}{N}\right)$$

So the posterior

$$Posterior = P(\mu|x_1, x_2....x_N) \propto L(x_1, x_2....x_N|\mu)P(\mu)$$

Now the posterior pdf is the product of two Gaussians which is itself a Gaussian

$$P(\mu|x_1, x_2....x_N) = G(\frac{\bar{x}\sigma_{prior}^2 + \mu_0\frac{\sigma^2}{N}}{\sigma_{prior}^2 + \frac{\sigma^2}{N}}, \frac{\sigma_{prior}^2\frac{\sigma^2}{N}}{\frac{\sigma^2}{N} + \sigma_{prior}^2})$$

So we get the MAP estimated μ to be-

$$\hat{\mu}_{MAP1} = \frac{\bar{x}\sigma_{prior}^2 + \mu_0\frac{\sigma^2}{N}}{\sigma_{prior}^2 + \frac{\sigma^2}{N}}$$

Where $\bar{x}$ is the sample mean and σ is the given variance.

3 Maximum A Posteriori Estimate 2

When we take our prior to be a uniform distribution in range [9.5,11.5].

$$Prior = P(\mu) = \begin{cases} 0 & \text{if} \quad \mu < 9.5 \\ 1/2 & \text{if} \quad 9.5 \le \mu \le 11.5 \\ 0 & \text{if} \quad \mu > 11.5 \end{cases}$$

We know likelihood

$$L(x_1, x_2....x_N|\mu) = G(\frac{\sum_{i=1}^{N}(x_i)}{N}, \frac{\sigma^2}{N})$$

So the posterior

$$Posterior = P(\mu|x_1, x_2....x_N) \propto L(x_1, x_2....x_N|\mu)P(\mu)$$

$$P(\mu|x_1, x_2....x_N) = \begin{cases} 0 & \text{if} \quad \mu < 9.5 \\ G(\frac{\sum_{i=1}^{N}(x_i)}{N}, \frac{\sigma^2}{N}) & \text{if} \quad 9.5 \le \mu \le 11.5 \\ 0 & \text{if} \quad \mu > 11.5 \end{cases}$$

So the Posterior is a Gaussian limited to [9.5,11.5].
Let $\bar{x} = \frac{\sum_{i=1}^{N} X_i}{N}$, now for the MAP estimate, we have three cases.

1. If $\bar{x} < 9.5$ then the mode will occur at 9.5.
2. If $\bar{x} > 11.5$ then the mode will occur at 11.5.
3. Else mode will be $\bar{x}$.

So we get the MAP estimated μ to be

$$\hat{\mu}_{MAP2} = \begin{cases} 9.5 & \text{if} \quad \bar{x} < 9.5 \\ \bar{x} & \text{if} \quad 9.5 \leq \bar{x} \leq 11.5 \\ 11.5 & \text{if} \quad \bar{x} > 11.5 \end{cases}$$

4 Boxplots

For each value of N, we repeated the experiment M=100 times. Each time we generated data of samples of sizes N and then got the three estimates $\hat{\mu}_{ML}$ and $\hat{\mu}_{MAP1}$ and $\hat{\mu}_{MAP2}$. Then we took these data and plotted boxplots for all of them using the boxplotGroup() function in MATLAB.

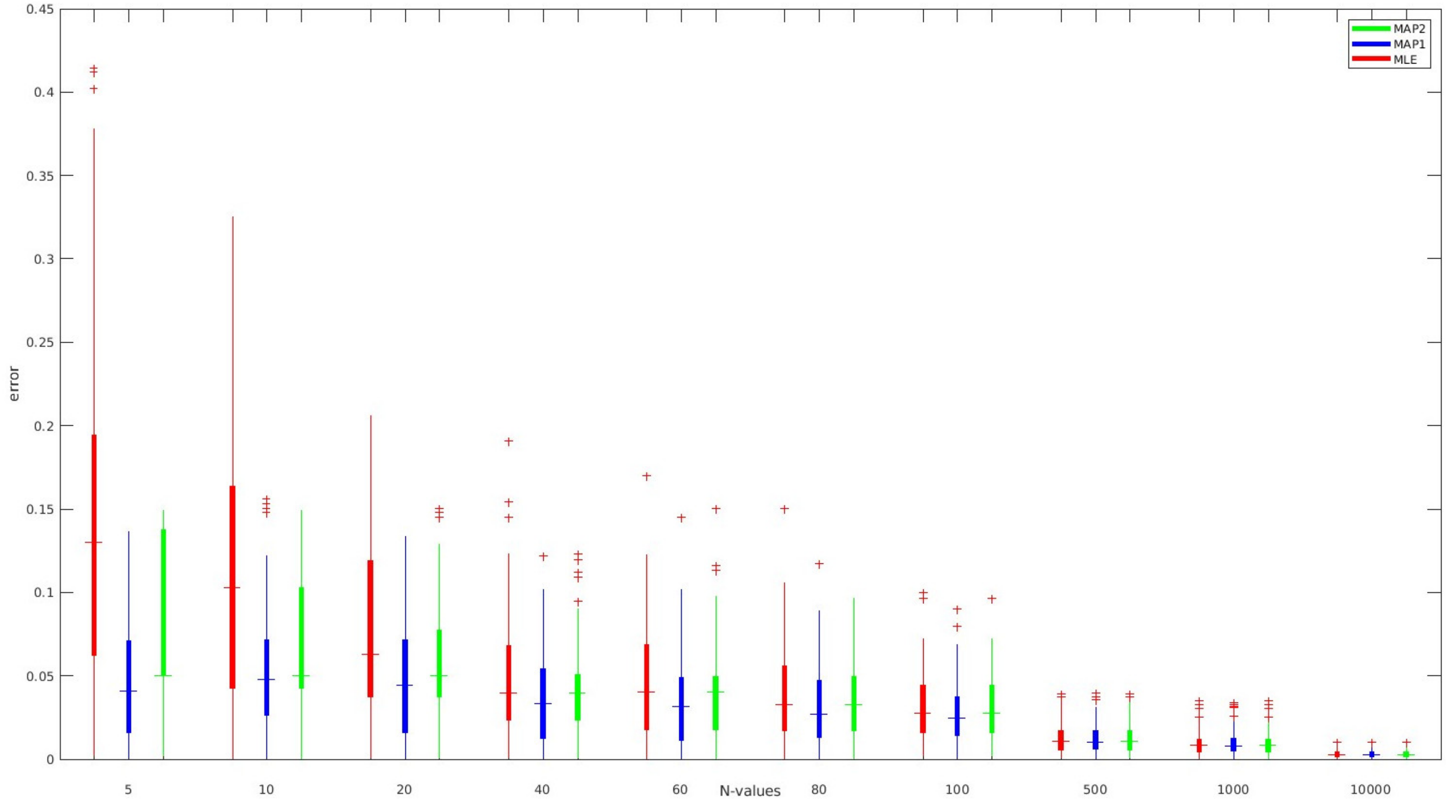

Figure 1: $\frac{|\mu_{true}-\mu_{estimate}|}{\mu_{true}}$

5 Interpretation

5.1 Increasing N

As we increase N the error decreases. We can see that the boxplots come closer to 0 and their width also decreases for each of the three graphs. This is because every time we increase N we get a better estimate of the mean as we have more information. We can also say that since the variance of the estimator is inversely proportional to N the spread of the data decreases.

5.2 Comparison

For smaller values of N (small sample size), we can see that MAP estimate 1 outperforms MAP estimate 2 which outperforms the ML estimate. But for larger values of N, we can see that the boxplots of all the estimates appear similar so all the estimators perform equally. So the MAP estimates tend to the ML estimate as N becomes very large. So in general we would prefer the MAP estimator with a gaussian prior which gives good estimates for all sample sizes.