Question 1

Avadhoot Jadhav - 210050027
Hrishikesh Jedhe Deshmukh - 210050073

October 2022

Let x_1, x_2, ... x_N be the observed values of some random variable (say X) whose mean we want to estimate.

$$\begin{aligned}
\text{Likelihood} = \mathbb{P}(X|\mu) &= G(x_1, \mu, \sigma^2_{\text{true}}) \cdot G(x_2, \mu, \sigma^2_{\text{true}}) ......G(x_N, \mu, \sigma^2_{\text{true}}) \\
&= \Pi_{i=1}^N G(x_i, \mu, \sigma^2_{\text{true}}) \\
&= \Pi_{i=1}^N G(\mu, x_i, \sigma^2_{\text{true}}) \\
&= exp\left(-0.5\left[N\mu^2 - 2\mu\sum x_i\right]/\sigma^2_{\text{true}} + \text{ terms independent of } \mu\right) \\
&= exp\left(-0.5\left[\mu^2 - 2\mu\bar{x}\right]/(\sigma^2_{\text{true}}/N) + \text{ terms independent of } \mu\right) \\
&\propto G(\mu, \bar{x}, \sigma^2_{\text{true}}/N)
\end{aligned}$$

To find $\hat{\mu}^{ML}$, we need to find μ such that likelihood is maximized. As value of a gaussian in maximum at it's mean, so -

$$\boxed{\hat{\mu}^{ML} = \bar{x} = \frac{\sum x_i}{N} = \text{ Sample Mean}}$$

- **Maximum-a-posteriori for Gaussian Prior ($\hat{\mu}^{MAP1}$)**

$$\text{Prior} = \mathbb{P}(\mu) = G(\mu, \mu_{\text{prior}}, \sigma^2_{\text{prior}})$$
$$\text{Evidence} = \mathbb{P}(X) = \text{Constant w.r.t } \mu$$

We know the following about product of 2 Gaussians :-

$$G(x, \mu_1, \sigma_1) \cdot G(x, \mu_2, \sigma_2) \propto G\left(x, \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)$$

$$\begin{aligned}
\text{Posterior} = \mathbb{P}(\mu|X) &= \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \\
&\propto G(\mu, \bar{x}, \sigma^2_{\text{true}}/N) \cdot G(\mu, \mu_{\text{prior}}, \sigma^2_{\text{prior}})
\end{aligned}$$

$$\boxed{\therefore \hat{\mu}^{MAP1} = \frac{\bar{x}\sigma^2_{\text{prior}} + \mu_{\text{prior}}\sigma^2_{\text{true}}/N}{\sigma^2_{\text{prior}} + \sigma^2_{\text{true}}/N}}$$

- **Maximum-a-posteriori for Uniform Prior ($\hat{\mu}^{MAP2}$)**

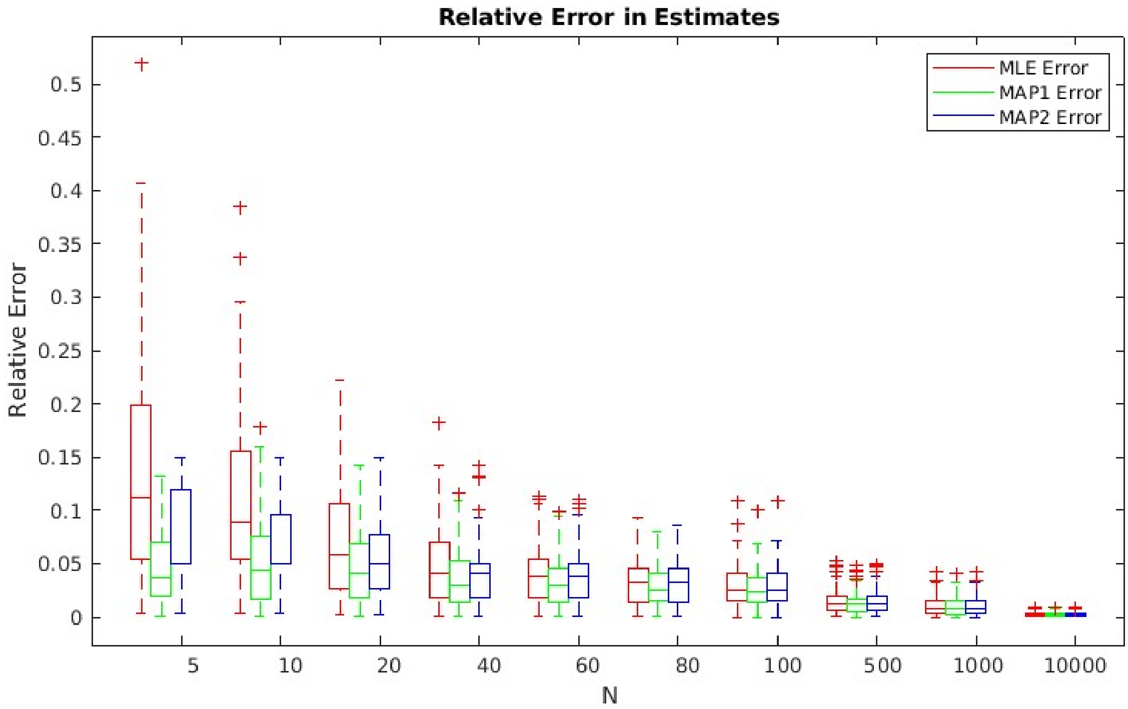
$$Prior = \mathbb{P}(\mu) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq \mu \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Evidence} = \mathbb{P}(X) = \text{Constant w.r.t } \mu$$

$$\text{Posterior} = \mathbb{P}(\mu|X) = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$
$$= \begin{cases} \text{constant} \times G(\mu, \bar{x}, \sigma^2_{\text{true}}/N) & \text{if } a \leq \mu \leq b \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\mu}^{MAP2} = \begin{cases} \bar{x} & \text{if } a \leq \bar{x} \leq b \\ a & \text{if } \bar{x} < a \\ b & \text{if } b < \bar{x} \end{cases}$$

- **Interpretation of graph**



For small N, maximum-likelihood estimate has large errors compared to maximum-a-posteriori estimate. This is due to the fact that, for small data sizes, there can be a large variation of the data from the true mean, leading to bad estimate given by MLE. But in this case if we use MAP estimate, we get better results with less errors, because we have priors that are pretty close to what true mean is and therefore these priors are nullifying a part of variation in data.

For infinite N, all 3 of the estimators tend to sample mean which in turn tends to μ_{true}.

As value of N increases, error in all 3 of the estimates decreases. There might be some outliers, but most of the errors are close to zero for large values of N. This is due to the fact that all 3 of the estimates converge to true mean as $N \to \infty$, so error in them decreases.

The best estimator among the 3 is MAP1 i.e. Gaussian prior. Average error for this estimator is small compared to other estimators. The next best estimator is MAP2 i.e. Uniform prior. This estimator is not much different from MAP1, but has a higher average error. The least reliable estimator in this case the the MLE, which has very high errors for small data sizes.

So we would prefer to use MAP1 estimator which is with a Gaussian Prior.