CS215 Assignment 3
Bayesian Estimation

Josyula Venkata Aditya
210050075

Kartik Sreekumar Nair
210050083

October 22, 2022

Contents

Preface

This assignment is done in a group. Group members:

1. **Josyula Venkata Aditya - 210050075**
2. **Kartik Sreekumar Nair - 210050083**

The submission folder contains four folders(in addition to this file and the problem statement file):

code - contains 3 *MATLAB*(`.m`) files, corresponding to each of the 3 problems.

results/fig - contains 3 sub-folders; containing the respective plots of that problem. This is the main folder of plots, please perform the evaluations of plots using this folder.

results/eps - contains all the plots, but in `.eps` format, used in the report. This format may not be readable from your device, hence for evaluation purposes, please check the plots in `results/fig` folder.

report - contains the reports of the 3 problems.

Problem 1

Use the Matlab function randn() to generate a data sample of N points drawn from a Gaussian distribution with mean $\mu_{\text{true}} = 10$ and standard deviation $\sigma_{\text{true}} = 4$. Consider the problem of using the data σ_{prior} to get an estimate $\hat{\mu}$ of this Gaussian mean, assuming it is unknown, when the standard deviation σ_{true} is known. Consider using one of the two prior prior distributions on the mean:

(i) a Gaussian prior with mean $\mu_{\text{prior}} = 10.5$ and standard deviation $\sigma_{\text{prior}} = 1$

$$f_1(\mu) = \frac{1}{\sqrt{2\pi}\sigma_{\text{prior}}} e^{-\frac{(\mu-\mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}} \tag{1.1}$$

(ii) a uniform prior over $[9.5, 11.5]$.

$$f_2(\mu) = \begin{cases} \frac{1}{2} & 9.5 \leq \mu \leq 11.5 \\ 0 & \text{o.w.} \end{cases} \tag{1.2}$$

Consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$. For each sample size N, repeat the following experiment $M \geq 100$ times: generate the data, get the maximum likelihood estimate $\hat{\mu}^{\text{ML}}$, get the maximum-a-posteriori estimates $\hat{\mu}_1^{\text{MAP}}$ and $\hat{\mu}_2^{\text{MAP}}$, and measure the relative errors $\frac{|\hat{\mu}-\mu_{\text{true}}|}{\mu_{\text{true}}}$ for all three estimates.

We are given the pdf of the data given its mean μ,

$$f_X(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma_{\text{true}}} e^{-\frac{(x-\mu)^2}{2\sigma^2_{\text{true}}}} \tag{1.3}$$

For the entire data of N sample points, we get the Likelihood L:

$$L(\mu) = f(\{x_i\}|\mu) = \prod_{i=1}^{N} f_X(x_i|\mu) \tag{1.4}$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{\text{true}}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2_{\text{true}}}} \tag{1.5}$$

$$\log L(\mu) = \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma_{\text{true}}} - \frac{(x_i-\mu)^2}{2\sigma^2_{\text{true}}} \tag{1.6}$$

$$= c_0 - \frac{1}{2\sigma^2_{\text{true}}} \sum_{i=1}^{N} (x_i-\mu)^2 \tag{1.7}$$

$$= c_0 - \frac{N}{2\sigma^2_{\text{true}}} \left(\mu^2 - 2\frac{\sum x_i}{N}\mu + \frac{\sum x_i^2}{N}\right) \tag{1.8}$$

$$= c_1 - \frac{N}{2\sigma^2_{\text{true}}} (\mu - \frac{\sum x_i}{N})^2 \tag{1.9}$$

The ML estimate $\hat{\mu}^{\text{ML}}$ is

$$\begin{aligned}\hat{\mu}^{\text{ML}} &= \arg\max_{\mu} L(\mu) = \arg\max_{\mu} \log L(\mu) && (1.10)\\ &= \arg\min_{\mu} \frac{N}{2\sigma^2_{\text{true}}}(\mu - \frac{\sum x_i}{N})^2 && (1.11)\\ &= \frac{\sum x_i}{N} = \overline{x}, && (1.12)\end{aligned}$$

the sample mean.
The posterior is

$$\begin{aligned}P = f(\mu|\{x_i\}) &= \frac{f(\{x_i\}|\mu)f(\mu)}{f(\{x_i\})} && (1.13)\\ &= \frac{1}{f(\{x_i\})}L(\mu)f(\mu) && (1.14)\\ \log P &= c' + \log L(\mu) + \log f(\mu) && (1.15)\end{aligned}$$

We find the MAP estimates for both the priors

(i)

$$\begin{aligned}\hat{\mu}_1^{\text{MAP}} &= \arg\max_{\mu} \log P && (1.16)\\ &= \arg\max_{\mu} c' + \log L(\mu) + \log f_1(\mu) && (1.17)\\ &= \arg\max_{\mu} c - \frac{N}{2\sigma^2_{\text{true}}}(\mu - \overline{x})^2 - \frac{(\mu - \mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}} && (1.18)\\ &= \arg\min_{\mu} \frac{N}{2\sigma^2_{\text{true}}}(\mu - \overline{x})^2 + \frac{(\mu - \mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}} && (1.19)\end{aligned}$$

Do your magic JV

$$\frac{\partial}{\partial \mu}\left[\frac{N}{2\sigma^2_{\text{true}}}(\mu - \overline{x})^2 + \frac{(\mu - \mu_{\text{prior}})^2}{2\sigma^2_{\text{prior}}}\right]_{\mu = \hat{\mu}_1^{\text{MAP}}} = \frac{N}{\sigma^2_{\text{true}}}(\mu - \overline{x}) + \frac{(\mu - \mu_{\text{prior}})}{\sigma^2_{\text{prior}}}\Bigg|_{\mu = \hat{\mu}_1^{\text{MAP}}} = 0 \qquad (1.20)$$

Hence

$$\hat{\mu}_1^{\text{MAP}} = \frac{\sigma^2_{\text{true}}\mu_{\text{prior}} + N\sigma^2_{\text{prior}}\overline{x}}{\sigma^2_{\text{true}} + N\sigma^2_{\text{prior}}} \qquad (1.21)$$

(ii)

$$\begin{aligned}\hat{\mu}_2^{\text{MAP}} &= \arg\max_{\mu} \log P && (1.22)\\ &= \arg\max_{\mu} c' + \log L(\mu) + \log f_2(\mu) && (1.23)\\ &= \arg\max_{\mu \in [9.5, 11.5]} \log L(\mu) && (1.24)\\ &= \arg\min_{\mu \in [9.5, 11.5]} (\mu - \overline{x})^2 && (1.25)\end{aligned}$$

$$\hat{\mu}_2^{\text{MAP}} = \begin{cases} 9.5 & \overline{x} < 9.5 \\ \overline{x} & 9.5 \le \overline{x} \le 11.5 \\ 11.5 & 11.5 < \overline{x} \end{cases} \qquad (1.26)$$

- Plot a single graph that shows the relative errors for each value of N as a box plot (use the Matlab boxplot() function), for each of the three estimates.

- Interpret what you see in the graph.
  - (i) What happens to the error as N increases ?
  - (ii) Which of the three estimates will you prefer and why ?
    Please gib knowledge JV

Problem 2

Reorder everything pls
Use the Matlab function rand() to generate a data sample of N points from the uniform distribution on $[0, 1]$. Transform the resulting data x to generate a transformed data sample where each datum $y := -\frac{1}{\lambda}\log(x)$ with $\lambda = 5$. The transformed data y will have some distribution with parameter λ; what is its analytical form ? Use a Gamma prior on the parameter λ, where the Gamma distribution has parameters $\alpha = 5.5$ and $\beta = 1$.

$$f(\lambda) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda/\beta} \tag{2.1}$$

Consider various sample sizes $N = 5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4$. For each sample size N, repeat the following experiment $M \geq 100$ times: generate the data, get the maximum likelihood estimate $\hat{\lambda}^{\text{ML}}$, get the Bayesian estimate as the posterior mean $\hat{\lambda}^{\text{PosteriorMean}}$, and measure the relative errors $\frac{|\hat{\lambda}-\lambda_{\text{true}}|}{\lambda_{\text{true}}}$ for both the estimates.

$$X \sim \text{Uniform}[0, 1] \tag{2.2}$$

We take

$$Y = g(X) \tag{2.3}$$

where

$$g(x) = -\frac{1}{\lambda}\log x, \tag{2.4}$$

and its inverse

$$g^{-1}(y) = e^{-\lambda y}. \tag{2.5}$$

Using transformation of random variables,

$$f_Y(y) = f_X(g^{-1}(y)\left|\frac{dg^{-1}(y)}{dy}\right| \tag{2.6}$$

Given $\lambda > 0$

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & 0 \leq e^{-\lambda y} \leq 1 \\ 0 & \text{o.w.} \end{cases} \tag{2.7}$$

$$= \begin{cases} \lambda e^{-\lambda y} & 0 \leq y \leq \infty \\ 0 & \text{o.w.} \end{cases} \tag{2.8}$$

An exponential distribution, We look at the likelihood given a prior on λ, given N sample points

$$L(\lambda) = f(\{y_i\}|\lambda) = \prod_{i=1}^{N} f_Y(y_i|\lambda) \tag{2.9}$$

$$\log L(\lambda) = \sum_{i=1}^{N} \log f_Y(y_i|\lambda) \tag{2.10}$$

$$= \sum_{i=1}^{N} \log(\lambda) - \lambda y_i \tag{2.11}$$

$$= N\log(\lambda) - \lambda \sum_{i=1}^{N} y_i, \tag{2.12}$$

given $y_i > 0 \forall 1 \leq i \leq N$.

$$\hat{\lambda}^{\mathrm{ML}} = \arg\max_{\lambda} \log L \tag{2.13}$$

Say stuff JV

$$\frac{\partial}{\partial\lambda}\left[\log L\right]_{\lambda=\hat{\lambda}^{\mathrm{ML}}} = \left.\frac{N}{\lambda} - \sum_{i=1}^{N} y_i\right|_{\lambda=\hat{\lambda}^{\mathrm{ML}}} = 0. \tag{2.14}$$

$$\hat{\lambda}^{\mathrm{ML}} = \frac{N}{\sum_{i=1}^{N} y_i} = \frac{1}{\overline{y}} \tag{2.15}$$

- Derive a formula for the posterior mean

  The posterior mean,

$$\hat{\lambda}^{\mathrm{PosteriorMean}} = E_{f(\lambda|\{y_i\})}[\lambda] \tag{2.16}$$

$$\hat{\lambda}^{\mathrm{PosteriorMean}} = \int f(\lambda|\{y_i\})\lambda d\lambda \tag{2.17}$$

$$= \int_0^\infty \frac{f(\{y_i\}|\lambda)f(\lambda)}{f(\{y_i\})}\lambda d\lambda \tag{2.18}$$

$$= \frac{\int_0^\infty f(\{y_i\}|\lambda)f(\lambda)\lambda d\lambda}{\int_0^\infty f(\{y_i\}|\lambda)f(\lambda)d\lambda} \tag{2.19}$$

$$f(\{y_i\}|\lambda)f(\lambda) = \lambda^N e^{-\lambda N\overline{y}} \cdot \frac{\beta^{-\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda/\beta} \tag{2.20}$$

$$= k\,\lambda^{\alpha+N-1}e^{-\lambda(N\overline{y}+1/\beta)} \tag{2.21}$$

We evaluate,

$$\int_0^\infty f(\{y_i\}|\lambda)f(\lambda)\lambda d\lambda = \int_0^\infty k\,\lambda^{\alpha+N}e^{-\lambda(N\overline{y}+1/\beta)} \tag{2.22}$$

$$= \left[-k\,\frac{\lambda^{\alpha+N}e^{-(N\overline{y}+1/\beta)\lambda}}{(N\overline{y}+1/\beta)}\right]_0^\infty + \frac{\alpha+1}{(N\overline{y}+1/\beta)}\int_0^\infty k\,\lambda^{\alpha+N-1}e^{-\lambda(N\overline{y}+1/\beta)}d\lambda \tag{2.23}$$

$$= \frac{\alpha+N}{(N\overline{y}+1/\beta)}\int_0^\infty f(\{y_i\}|\lambda)f(\lambda)d\lambda \tag{2.24}$$

Therefore

$$\hat{\lambda}^{\mathrm{PosteriorMean}} = \frac{\alpha+N}{(N\overline{y}+1/\beta)} \tag{2.25}$$

- Plot a single graph that shows the relative errors for each value of N as a box plot (use the Matlab boxplot() function), for both the estimates.
- Interpret what you see in the graph.
  - (i) What happens to the error as N increases ?
  - (ii) Which of the two estimates will you prefer and why ?

Problem 3

Suppose random variable X has a uniform distribution over $[0, \theta]$, where the parameter θ is unknown.

$$f_X(x|\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{o.w.} \end{cases} \tag{3.1}$$

Consider a Pareto distribution prior on θ, with a scale parameter $\theta_m > 0$ and a shape parameter $\alpha > 1$, as

$$f(\theta) = \begin{cases} c(\theta_m/\theta)^\alpha & \theta \ge \theta_m \\ 0 & \text{o.w.} \end{cases} \tag{3.2}$$

- Find the maximum-likelihood estimate $\hat{\theta}^{\text{ML}}$ and the maximum-a-posteriori estimate $\hat{\theta}^{\text{MAP}}$.

  The likelihood, given N i.i.d sample points $\{x_i\}$

$$L(\theta) = f(\{x_i\}|\theta) = \prod_{i=1}^{N} f(x_i|\theta) \tag{3.3}$$

$$= \begin{cases} \theta^{-N} & 0 \le x_i \le \theta \;\; \forall 1 \le i \le N \\ 0 & \text{o.w.} \end{cases} \tag{3.4}$$

$$\hat{\theta}^{\text{ML}} = \arg\max_{\theta} L \tag{3.5}$$

$$= \arg\max_{\theta:\theta \ge \max_i x_i} L \tag{3.6}$$

$$= \arg\max_{\theta:\theta \ge \max_i x_i} \theta^{-N} \tag{3.7}$$

as $\theta > 0$ and θ^{-N} is strictly decreasing.

$$\therefore \hat{\theta}^{\text{ML}} = \max_i x_i \tag{3.8}$$

For MAP

The Posterior,

$$P = f(\theta|\{x_i\}) = \frac{f(\{x_i\}|\theta) f(\theta)}{f(\{x_i\})} \tag{3.9}$$

$$= k' L(\theta) f(\theta) \tag{3.10}$$

$$= \begin{cases} k\theta^{-\alpha-N} & \theta \ge x_i, \theta_m \\ 0 & \text{o.w} \end{cases} \tag{3.11}$$

$$\hat{\theta}^{\text{MAP}} = \arg\max_{\theta} P = \arg\max_{\theta:\theta \ge x_i, \theta_m} \theta^{-\alpha-N} \tag{3.12}$$

$$= \max(\theta_m, \max_i x_i) \tag{3.13}$$

- Does $\hat{\theta}^{\text{MAP}}$ tend to $\hat{\theta}^{\text{ML}}$ as the sample size tends to infinity ? Is this desirable or not ?
  $\hat{\theta}^{\text{ML}} = \hat{\theta}^{\text{MAP}}$ iff $\theta_m \leq \max_i x_i$. interpret this JV
- Find an estimator of the mean of the posterior distribution $\hat{\theta}^{\text{PosteriorMean}}$.

$$\begin{aligned}
\hat{\theta}^{\text{PosteriorMean}} &= E_{f(\theta|\{x_i\})}[\theta] && (3.14)\\
&= \int P\theta d\theta && (3.15)\\
&= \frac{\int f(\{x_i\}|\theta) f(\theta)\theta d\theta}{\int f(\{x_i\}|\theta) f(\theta) d\theta} && (3.16)\\
&= \frac{\int_{\max(\theta_m, \max_i x_i)}^{\infty} k\theta^{-\alpha-N}\theta d\theta}{\int_{\max(\theta_m, \max_i x_i)}^{\infty} k\theta^{-\alpha-N} d\theta} && (3.17)\\
&= \frac{\alpha+N-1}{\alpha+N-2}\max(\theta_m, \max_i x_i) && (3.18)
\end{aligned}$$

- Does $\hat{\theta}^{\text{PosteriorMean}}$ tend to $\hat{\theta}^{\text{ML}}$ as the sample size tends to infinity ? Is this desirable or not ?