CS215 Assignment 3
Question 3

Atishay Jain - 210050026
Gohil Megh Hiteshkumar - 210050055

November 1, 2022

1 Maximum Likelihood Estimate $\hat{\theta}^{ML}$

The given Random Variable X has uniform distribution over $[0, \theta]$

$$P(X = x|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } x \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

To ease all the calculation, lets define a new function $f(x)$

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

Hence above PDF can be rewritten as,

$$P(X = x|\theta) = \frac{1}{\theta} f(x)$$

Now, lets calculate the Likelihood function -

$$\text{Likelihood : } P(\text{data}|\theta) = \prod_{i=1}^{N} P(X = x_i|\theta) = \left(\frac{1}{\theta}\right)^N \prod_{i=1}^{N} f(x_i)$$

where all the data is generated from uniform distribution with parameters 0 and $\theta^{true}(\theta^{true} > 0)$.
The above function can be rewritten as,

$$P(data|\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^N & \text{if } \forall i, x_i \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

Maximum value of likelihood cannot be zero since $\left(\frac{1}{\theta}\right)^N$ is a positive number. Hence, to maximize likelihood function we must consider the case where all $x_i \in [0, \theta]$.
To maximize the likelihood further we need to take smallest possible value of θ.
So given the condition $\forall i, x_i \in [0, \theta]$, $\theta = max(x_1, x_2, x_3, ..., x_N)$.

Thus, we get :-

$$\boxed{\hat{\theta}^{ML} = max(x_1, x_2, x_3, ..., x_N)}$$

2 Maximum-a-Posteriori estimate $\hat{\theta}^{MAP}$

Given,
Prior distribution with PDF as,

$$P(\theta) = \begin{cases} NF_1 * \left(\frac{\theta_m}{\theta}\right)^{\alpha} & \text{if } \theta \geq \theta_m \\ 0 & \text{otherwise} \end{cases}$$

where, NF_1 is normalizing factor
Posterior PDF will be

$$\begin{aligned} P(\theta|data) &= NF_2 * P_{uni}(data|\theta) * P_{prior}(\theta) \\ &= NF_2 * P_{prior}(\theta) * \prod_{i=1}^{N} P_{uni}(X = x_i|\theta) \\ &= NF_2 * P_{prior}(\theta) * \left(\frac{1}{\theta}\right)^{N} \prod_{i=1}^{N} f(x_i) \end{aligned}$$

Thus,

$$P(\theta|data) = \begin{cases} NF_2 * NF_{1*} \left(\frac{\theta_m}{\theta}\right)^{\alpha} * \left(\frac{1}{\theta}\right)^{N} & \text{if } \theta \geq \theta_m \text{ and } \forall i, x_i \in [0, \theta] \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Now to maximize above function, $\theta \geq max(x_1, x_2, x_3, ..., x_N, \theta_m)$
Thus with above condition

$$P(\theta|data) = NF * \left(\frac{1}{\theta}\right)^{\alpha+m} \tag{2}$$

Now to maximize above expression θ should be as small as possible and it should be greater than or equal to all x_i and θ_m
Hence, $\theta = max(x_1, x_2, x_3, ..., x_N, \theta_m)$
Therefore, we get :-

$$\boxed{\hat{\theta}^{MAP} = max(x_1, x_2, x_3, ..., x_N, \theta_m)}$$

3 $\hat{\theta}^{ML}$ and $\hat{\theta}^{MAP}$ as $N \to \infty$

Suppose, the true value of θ is θ^{true} from where the data was generated. So almost surely $\forall x_i \in [0, \theta^{true}]$.
There are two cases -

3.1 CASE 1 : $\theta_m > \theta^{true}$

Suppose the prior is chosen with scale parameter $\theta_m > \theta^{true}$ then for any size of the data $\hat{\theta}^{MAP} \neq \theta^{true}$ almost surely since $\theta_m > max(x_1, x_2, x_3, ..., x_N)$ thus, $\hat{\theta}^{MAP} = \theta_m \neq \theta^{true}$.
So even if size of data i.e. $N \to \infty$, $\hat{\theta}^{MAP}$ does not tend to θ^{true}. But $\hat{\theta}^{ML} \to \theta^{true}$ as $N \to \infty$.
It can be shown as follows, $\forall \epsilon > 0$

$$\lim_{N\to\infty} P(|\hat{\theta}^{ML} - \theta^{true}| \geq \epsilon) = \lim_{N\to\infty} P(\hat{\theta}^{ML} \leq \theta^{true} - \epsilon)$$

$\hat{\theta}^{ML}$ is $max(x_1, x_2, x_3, ..., x_N)$, which means that $\forall x_i < \theta^{true} - \epsilon$

$$\begin{aligned}
\lim_{N\to\infty} P(\hat{\theta}^{ML} \leq \theta^{true} - \epsilon) &= \lim_{N\to\infty} \prod_{i=1}^{N} P(x_i \leq \theta^{true} - \epsilon) \\
&= \lim_{N\to\infty} \left(\frac{\theta^{true} - \epsilon}{\theta^{true}}\right)^N \\
&= 0
\end{aligned}$$

So $\hat{\theta}^{ML} \to \theta^{true}$ as $N \to \infty$ but, $\hat{\theta}^{MAP}$ does not tend to θ^{true} which means $\hat{\theta}^{MAP}$ does not tend to θ^{ML} in this case.

3.2 CASE 2 : $\theta_m \leq \theta^{true}$

Suppose the prior is chosen with scale parameter $\theta_m \leq \theta^{true}$
When $\theta_m > max(x_1, x_2, x_3, ..., x_N)$ then $\hat{\theta}^{ML} \neq \hat{\theta}^{MAP}$
Otherwise, $\hat{\theta}^{ML} = \hat{\theta}^{MAP}$

$$P(X < \theta_m) = \left(\frac{\theta_m}{\theta}\right)$$

$$P(\forall i, x_i < \theta_m) = \left(\frac{\theta_m}{\theta}\right)^N$$

Now $N \to \infty \implies \left(\frac{\theta_m}{\theta}\right)^N \to 0 \implies P(\forall i, x_i < \theta_m) \to 0$
which indicates that $\exists$ almost surely at least one x_i such that $x_i \geq \theta_m$ as $N \to \infty$
which means $\hat{\theta}^{MAP} \to \hat{\theta}^{ML}$ as $N \to \infty$

3.3 Final Conclusion

So from above two cases we can see that it depends on the prior chosen if $\hat{\theta}^{MAP} \to \hat{\theta}^{ML}$ or not. But, it is desirable if $\hat{\theta}^{MAP} \to \hat{\theta}^{ML}$ at large size of data since $\hat{\theta}^{ML} \to \theta^{true}$.If it does not converge to $\hat{\theta}^{ML}$ then the MAP estimate is not consistent.

4 Mean estimator of the Posterior distribution ($\hat{\theta}^{PosteriorMean}$)

$$E_{P(\theta|data)}[\theta|data] = \int_{-\infty}^{\infty} \theta P(\theta|data) d\theta$$

Let $\beta = max(x_1, x_2, x_3, ..., x_N, \theta_m)$
From equation (1),

$$\begin{aligned}
E_{P(\theta|data)}[\theta|data] &= \int_{-\infty}^{\infty} \theta P(\theta|data) d\theta \\
&= \int_{-\infty}^{\beta} \theta * 0 d\theta + \int_{\beta}^{\infty} \theta * NF * \left(\frac{1}{\theta}\right)^{N+\alpha} d\theta \\
&= NF \int_{\beta}^{\infty} \left(\frac{1}{\theta}\right)^{N+\alpha-1} d\theta \\
&= \frac{NF}{2 - N - \alpha} \left[\frac{1}{\theta^{N+\alpha-2}}\right]_{\beta}^{\infty}
\end{aligned}$$

$$E_{P(\theta|data)}[\theta|data] = \frac{NF}{(N+\alpha-2)*\beta^{N+\alpha-2}} \tag{3}$$

Now to calculate value of NF which is normalizing factor, integrate the PDF and equate to 1,

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} P(\theta|data)d\theta \\
&= \int_{-\infty}^{\beta} 0 d\theta + \int_{\beta}^{\infty} NF * \left(\frac{1}{\theta}\right)^{N+\alpha} d\theta \\
&= NF \int_{\beta}^{\infty} \left(\frac{1}{\theta}\right)^{N+\alpha} d\theta \\
&= \frac{NF}{1-N-\alpha}\left[\frac{1}{\theta^{N+\alpha-1}}\right]_{\beta}^{\infty} \\
&= \frac{NF}{N+\alpha-1}\frac{1}{\beta^{N+\alpha-1}}
\end{aligned}$$

Thus,

$$NF = (N+\alpha-1)\beta^{N+\alpha-1}$$

Hence,

$$\boxed{\hat{\theta}^{PosteriorMean} = \frac{N+\alpha-1}{N+\alpha-2}\beta}$$

5 $\hat{\theta}^{ML}$ and $\hat{\theta}^{PosteriorMean}$ as $N \to \infty$

We have to calculate $\lim_{N\to\infty} \frac{N+\alpha-1}{N+\alpha-2}\beta$
Since $\lim_{N\to\infty} \frac{N+\alpha-1}{N+\alpha-2}$ and $\lim_{N\to\infty} \beta$ both exist, we can write limit as -

$$\lim_{N\to\infty} \frac{N+\alpha-1}{N+\alpha-2}\beta = \left(\lim_{N\to\infty} \frac{N+\alpha-1}{N+\alpha-2}\right) * \left(\lim_{N\to\infty} \beta\right)$$

So as $N \to \infty$,

$$\lim_{N\to\infty} \hat{\theta}^{PosteriorMean} = \lim_{N\to\infty} \beta$$

The above value is same as the $\hat{\theta}^{MAP}$ and we have shown earlier that it depends on the chosen prior if $\hat{\theta}^{MAP}$ tends to $\hat{\theta}^{ML}$ or not as N tends to ∞. So here as well it will depend on the prior chosen.

Yes, It is desirable if $\hat{\theta}^{PosteriorMean} \to \hat{\theta}^{ML}$, since $\hat{\theta}^{ML} \to \theta^{true}$ as seen earlier.