Question 3

Avadhoot Jadhav - 210050027
Hrishikesh Jedhe Deshmukh - 210050073

October 2022

(20 points) Suppose random variable X has a uniform distribution over $[0, \theta]$, where the parameter θ is unknown. Consider a Pareto distribution prior on θ, with a scale parameter $\theta_m > 0$ and a shape parameter $\alpha > 1$, as $P(\theta) \propto (\theta_m/\theta)^\alpha$ for $\theta \geq \theta_m$ and $P(\theta) = 0$ otherwise.

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

Let x_1, x_2, ... x_N be the observed values of random variable X.

$$\text{Likelihood} = \mathbb{P}(X|\theta) = \begin{cases} \left(\frac{1}{\theta}\right)^N & \text{if } 0 \leq x_i \leq \theta \quad \forall i = 1, 2, ...N \\ 0 & \text{otherwise} \end{cases}$$

To maximize likelihood, we have to minimize θ but with an added constraint that $\theta \geq max(x_i)$

$$\boxed{\hat{\theta}^{ML} = max(x_i)}$$

$$\text{Prior} = \mathbb{P}(\theta) = \begin{cases} c\left(\dfrac{\theta_m}{\theta}\right)^\alpha & \forall \theta \geq \theta_m \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Evidence} = \mathbb{P}(X) = \text{Independent of } \theta$$

$$\begin{aligned} \text{Posterior} = \mathbb{P}(\theta|X) &= \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}} \\ &= \begin{cases} \dfrac{\left(\dfrac{1}{\theta}\right)^N \times c\left(\dfrac{\theta_m}{\theta}\right)^\alpha}{\text{Constant}} & \text{if } \theta \geq \theta_m \text{ and } \theta \geq max(x_i) \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} c'\left(\dfrac{1}{\theta}\right)^{N+\alpha} & \text{if } \theta \geq max(\theta_m, max(x_i)) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Thus to maximize the probability of posterior, we need to minimize θ but by keeping in mind that we have an added constraint of $\theta \geq max(\theta_m, max(x_i))$.

$$\boxed{\hat{\theta}^{MAP} = max(\theta_m, max(x_i)) = max(\theta_m, \hat{\theta}^{ML})}$$

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

To answer this question, let's first check the consistency of maximum-likelihood estimator of θ, i.e. $\hat{\theta}^{ML}$.

$\forall\epsilon > 0$ and $\epsilon < \theta^{TRUE}$, let's calculate the probability that the maximum-likelihood estimator gives an answer that is more than ϵ off of the θ^{TRUE}. That is we need to calculate $\mathbb{P}(\theta^{TRUE} - \hat{\theta}^{ML} \geq \epsilon)$.

$$\begin{aligned}\mathbb{P}(\theta^{TRUE} - \hat{\theta}^{ML} \geq \epsilon) &= \mathbb{P}(b - x_1 \geq \epsilon)\mathbb{P}(b - x_2 \geq \epsilon).......\mathbb{P}(b - x_N \geq \epsilon)\\ &= \mathbb{P}(x_1 \leq b - \epsilon)\mathbb{P}(x_2 \leq b - \epsilon).......\mathbb{P}(x_N \leq b - \epsilon)\\ &= \left(\frac{b-\epsilon}{b}\right)^N\end{aligned}$$

Thus we can see that, $\mathbb{P}(\theta^{TRUE} - \hat{\theta}^{ML} \geq \epsilon) \to 0$ as $N \to \infty$. So $\hat{\theta}^{ML}$ is a consistent estimator.

Now, $\hat{\theta}^{MAP} = max(\theta_m, \hat{\theta}^{ML})$ so,

$$\hat{\theta}^{MAP} = \begin{cases}\hat{\theta}^{ML} & \text{if } \hat{\theta}^{ML} > \theta_m\\ \theta_m & \text{if } \hat{\theta}^{ML} < \theta_m\end{cases}$$

Thus, $\hat{\theta}^{MAP}$ tends to $\hat{\theta}^{ML}$ if and only if $\hat{\theta}^{ML} \geq \theta_m$. So, the reliability of estimate we get from maximum-a-posteriori estimate depends strongly on the prior that we have chosen.

So, if we have chosen a prior such that $\theta_m < \theta^{TRUE}$, then we are almost sure that $\hat{\theta}^{MAP}$ will tend to θ^{TRUE} for large values of N. But, if we choose a prior such that $\theta_m > \theta^{TRUE}$, then $\hat{\theta}^{MAP}$ will almost never tend to θ^{TRUE}. This is not desirable as there is a large margin of error depending on the prior we choose. To mitigate this, we can possibly choose a prior with fairly low θ_m.

- **Find an estimator of the mean of the posterior distribution $\hat{\theta}^{PosteriorMean}$.**

$$\text{Posterior} = \mathbb{P}(\theta|X) = \begin{cases}c'\left(\frac{1}{\theta}\right)^{N+\alpha} & \text{if } \theta \geq max(\theta_m, max(x_i))\\ 0 & \text{otherwise}\end{cases}$$

$$\therefore \hat{\theta}^{PosteriorMean} = \frac{\int_{-\infty}^{+\infty}\theta\cdot\mathbb{P}(\theta|X)d\theta}{\int_{-\infty}^{+\infty}\mathbb{P}(\theta|X)d\theta}$$

Let, $max(\theta_m, max(x_i)) = \theta_o$

$$\begin{aligned}\therefore \hat{\theta}^{PosteriorMean} &= \frac{\int_{\theta_0}^{+\infty}\theta\left(\frac{1}{\theta}\right)^{N+\alpha}d\theta}{\int_{\theta_0}^{+\infty}\left(\frac{1}{\theta}\right)^{N+\alpha}d\theta}\\ &= \frac{\int_{\theta_o}^{+\infty}\theta^{-n-\alpha+1}d\theta}{\int_{\theta_o}^{+\infty}\theta^{-n-\alpha}d\theta}\\ &= \frac{\left[\frac{\theta^{-n-\alpha+2}}{-n-\alpha+2}\right]_{\theta_o}^{\infty}}{\left[\frac{\theta^{-n-\alpha+1}}{-n-\alpha+1}\right]_{\theta_o}^{\infty}}\\ &= \left(\frac{n+\alpha-1}{n+\alpha-2}\right)\theta_o\\ \hat{\theta}^{PosteriorMean} &= \left(\frac{n+\alpha-1}{n+\alpha-2}\right)max(\theta_m, max(x_i))\end{aligned}$$

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

$$\lim_{N \to \infty} \hat{\theta}^{PosteriorMean} = max(\theta_m, max(x_i)) = max(\theta_m, \hat{\theta}^{ML})$$

Thus, $\hat{\theta}^{PosteriorMean}$ tends to $\hat{\theta}^{ML}$, if we have chosen prior in such a way that $\theta^{TRUE} > \theta_m$. But if we have chosen prior such that $\theta_m > \theta^{TRUE}$, then $\hat{\theta}^{PosteriorMean}$ will almost never tend to $\hat{\theta}^{ML}$. This is not desirable, as the estimate we get strongly depends on our choice of prior and so there is a large margin for error.