

## Question 4

Given a  $\mathbf{X} \sim U(0, \theta)$ , where  $\theta$  has a Pareto prior Distribution, with scale parameter  $\theta_m$  and shape parameter  $\alpha$  with

$$P(\theta) = \begin{cases} \alpha \frac{\theta_m^\alpha}{\theta^{\alpha+1}} & \text{if } \theta > \theta_m \\ 0 & \text{otherwise} \end{cases}$$

### 5.1 : Part A : MLE, MAP

For MLE,

$$P(x_i|\theta) = \begin{cases} \frac{1}{\theta} & \text{if } \theta \geq x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Hence the Likelihood is given by -

$$\begin{aligned} \text{Likelihood} &= P(x_1|\theta) \cdot P(x_2|\theta) \dots P(x_n|\theta) \\ \implies \text{Posterior} &= P(\{x_i\}|\theta) \\ \implies \text{Likelihood} &= \theta^{-n} \end{aligned}$$

Now, to maximize this likelihood, theta must be smallest. But we know that theta has to be atleast as large as the largest value of the input data X.  
Hence

$$\hat{\theta}^{\text{MLE}} = \max\{\mathbf{X}_i\}$$

For the Posterior,

$$\begin{aligned} \text{Posterior} &\propto P(\theta) \cdot P(\{x_i\}|\theta) \\ \implies \text{Posterior} &\propto \alpha \frac{\theta_m^\alpha}{\theta^{\alpha+1}} \cdot \theta^{-n} \\ \implies \text{Posterior} &\propto \begin{cases} \theta^{-(n+\alpha+1)} & \text{if } \theta > \theta_m \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Now, to maximize this posterior, theta must be smallest, but greater than  $\theta_m$ , since the posterior is zero at all other points.  
But we know that theta has to be atleast as large as the largest value of the input data X.

Hence

$$\hat{\theta}^{\text{MAP}} = \max\{\theta_m, \max\{\mathbf{X}_i\}\}$$

Now, the posterior takes the form of another pareto distribution. We know that for a pareto distribution, the mode is the scale parameter.

Hence,

$$\text{Posterior} = P(\theta|\{x_i\}) = \text{Pareto}[(n + \alpha), \max\{\theta_m, \max\{X_i\}\}]$$

## 5.2 : Part B

As  $n \rightarrow \infty$ , we can obviously have a value of  $\theta_m$  that is greater than all the values in the data set. (For example, we take  $\theta_m$  as '1' greater than the max input data value)

In this case,

$$\begin{aligned}\hat{\theta}^{MAP} &= \max\{\theta_m, \max\{X_i\}\} \\ \hat{\theta}^{MAP} &\rightarrow \theta_m \\ \implies \hat{\theta}^{MAP} &\neq \max\{X_i\} \neq \hat{\theta}^{MLE}\end{aligned}$$

This is undesirable and **not a very good prior** since it is not converging to the MLE. As the sample size increases to infinity, MLE is the most efficient and accurate estimator. Since given the liberty of such large datasets, any estimator should give the best possible result, and hence should converge to MLE.

## 5.3 : Part C

For a pareto distribution with parameters  $\theta_m$  and  $\alpha$ , mean can be calculated as,

$$\begin{aligned}Mean &= \int_{\theta_m}^{\infty} \theta \cdot \alpha \frac{\theta_m^\alpha}{\theta^{\alpha+1}} d\theta \\ Mean &= \alpha \cdot \theta_m^\alpha \int_{\theta_m}^{\infty} \theta^{-\alpha+1} d\theta \\ Mean &= \frac{\alpha \theta_m^\alpha}{\alpha - 1}\end{aligned}$$

Hence for our posterior, we just have to substitute different parameters in the above obtained result.

$$Posterior = P(\theta|_{\{x_i\}}) = \text{Pareto}[(n + \alpha), \max\{\theta_m, \max\{X_i\}\}]$$

$$PosteriorMean = \frac{(\alpha + n) \cdot (\max\{\theta_m, \max\{X_i\}\})^\alpha}{\alpha + n - 1}$$

## 5.4 : Part D

Now as  $n \rightarrow \infty$ ,

$$PosteriorMean \rightarrow \max\{\theta_m, \max\{X_i\}\}$$

Giving a similar argument as given in part B, we can conclude that as the sample size increases, the posterior mean will not converge to  $\theta^{MLE}$

Again this is **not desirable** since the mean isn't converging to MLE and since MLE is the best possible estimate at large sample size, it always is desirable that the estimate converges to MLE for large data.