CS-215-Assignment 3

Akshat Goyal (210050009) and Anshul Verma (210050015)

October 2022

Problem 3

1 Pareto Distribution

A Pareto Distribution with scale parameter $\theta_m > 0$ and shape parameter $\alpha > 1$ is given as

$$P(\theta) = (\frac{\alpha - 1}{\theta_m})(\theta_m/\theta)^\alpha \quad \theta \geq \theta_m$$

$$P(\theta) = 0 \quad otherwise$$

1.1 Mode of the Distribution

$$\frac{\partial}{\partial\theta}P(\theta) = (-\alpha)(\frac{\alpha - 1}{\theta_m})(\theta_m)^\alpha(\theta)^{-\alpha-1}$$

.

Hence the derivative of this function is negative and so it is a strictly decreasing function. Hence the maximum value of the PDF occurs at $\theta = \theta_m$.

$$\textbf{Mode} = \theta_m$$

1.2 Mean of the Distribution

We have

$$\bar{\theta} = \int_{\theta_m}^{\infty} \theta.P(\theta)\, d\theta$$

.

$$\Rightarrow \bar{\theta} = \int_{\theta_m}^{\infty} \theta.(\frac{\alpha - 1}{\theta_m})(\theta_m/\theta)^\alpha\, d\theta$$

$$\Rightarrow \bar{\theta} = \int_{\theta_m}^{\infty} \theta^{1-\alpha}(\alpha - 1)(\theta_m)^{\alpha-1}\, d\theta$$

$$\Rightarrow \bar{\theta} = \frac{\theta^{2-\alpha}}{2-\alpha}(\alpha - 1)(\theta_m)^{\alpha-1}\Big|_{\theta_m}^{\infty}$$

$$\Rightarrow \textbf{Mean} = \bar{\theta} = (\frac{\alpha - 1}{\alpha - 2})(\theta_m)$$

2 Maximum Likelihood Estimate

We are given that X is a random variable that has a Uniform PDF i.e $U[0, \theta]$.

Suppose $x_1, x_2, \ldots x_N$ are N observations from the PDF of x such that $x_1 \leq x_2 \leq \ldots \leq x_N$ then the likelihood function is given as

$$L(\theta) = P(x_1, x_2, \ldots, x_n|\theta) = \left(\frac{1}{\theta}\right)^N$$

$$\Rightarrow \frac{\partial L(\theta)}{\partial \theta} = \frac{-N}{\theta^{N+1}}$$

Hence Likelihood function is maximum for the case when θ is minimum.

Also, we know that

$$\theta \geq max(x_1, x_2, \ldots, x_n)$$

$$\Rightarrow min(\theta) = max(x_1, x_2, \ldots, x_n) = x_n$$

$$\Rightarrow \hat{\theta}^{ML} = x_n$$

Hence the maximum likelihood estimate of the random variable X is the maximum value of the sample.

3 Maximum-a-Posteriori Estimate

We know that Posterior Distribution is given as:

$$Posterior = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

$$P\left(\theta \mid x_1, x_2, \ldots x_N\right) = \frac{P\left(x_1, x_2, \ldots, x_N \mid \theta\right) \times P(\theta)}{\int P\left(x_1, x_2, \ldots, x_N \mid \theta\right) P(\theta) d\theta}$$

$$\Rightarrow P\left(\theta \mid x_1, x_2, \ldots x_N\right) \propto \left(\frac{\theta_m}{\theta}\right)^{N+\alpha} \quad \text{where } \theta \geq max(max(x_1, x_2, \ldots, x_n), \theta_m)$$

Hence it is a conjugate PDF to the Pareto Prior and has Pareto Distribution with $\alpha' = N + \alpha$ and $\theta'_m = max(max(x_1, x_2, \ldots, x_n), \theta_m) = max(x_n, \theta_m)$

$$\Rightarrow P\left(\theta \mid x_1, x_2, \ldots x_N\right) = \left(\frac{N + \alpha - 1}{\theta'_m}\right)(\theta'_m/\theta)^{N+\alpha} \quad \theta \geq \theta'_m$$

$$P\left(\theta \mid x_1, x_2, \ldots x_N\right) = 0 \quad otherwise$$

$$\Rightarrow \hat{\theta}^{MAP} = \theta'_m$$

$$\Rightarrow \hat{\theta}^{MAP} = max(x_n, \theta_m)$$

4 Comparison of estimates for large sample size

$$\hat{\theta}^{MAP} = max(max(x1, x2, \ldots, x_n), \theta_m)$$
$$\hat{\theta}^{ML} = max(x1, x2, \ldots, x_n)$$

When Sample size $N \to \infty$,

- if $\theta_m \leq \theta_{true}$

  $$\hat{\theta}^{MAP} = max(max(x1, x2, \ldots, x_n), \theta_m) \to \theta_{true}$$

  $$\hat{\theta}^{ML} = max(x1, x2, \ldots, x_n) \to \theta_{true}$$
  $$\Rightarrow \hat{\theta}^{MAP} \to \hat{\theta}^{ML}$$

  (Since all the data lies in [0, θ_{true}] and $\theta_m \leq \theta_{true}$)
  This case is desirable since the Maximum likelihood estimate has asymptotically the lowest Mean Squared Error and hence $\hat{\theta}^{MAP}$ is a good estimator.

- if $\theta_m > \theta_{true}$

  $$\hat{\theta}^{MAP} = max(max(x1, x2, \ldots, x_n), \theta_m) = \theta_m \quad \forall N$$

  $$\hat{\theta}^{ML} = max(x1, x2, \ldots, x_n) \to \theta_{true}$$
  $$\Rightarrow \hat{\theta}^{MAP} > \hat{\theta}^{ML}$$

  (Since all the data lies in [0, θ_{true}] and $\theta_m > \theta_{true}$)
  This case is undesirable because the MAP estimate has a constant bias of ($\theta_m - \theta_{true}$), so its Mean Squared Error never tends to 0. Whereas the ML-estimate's bias tends to 0 asymptotically and its Mean Squared Error also tends to 0.

In many cases due to a bad choice of prior Model, $\hat{\theta}^{MAP}$ would not tend to $\hat{\theta}^{ML}$ and this is undesirable and makes the MAP estimator a poor estimator.
For a correctly chosen Prior Model the MAP estimate tends to the ML estimate and also may be slightly better than it.

5 Posterior Mean Estimate

Since the posterior distribution is a Pareto distribution with the parameters as:

$$\alpha' = N + \alpha$$

$$\theta'_m = max(max(x_1, x_2, \ldots, x_n), \theta_m)$$

Posterior Mean estimate is (Assuming squared error Loss function):

$$\hat{\theta}^{PosteriorMean} = E_{Posterior}[\theta] = \int_{\theta'_m}^{\infty} \theta.P(\theta|X)\, d\theta$$

$$\Rightarrow \hat{\theta}^{PosteriorMean} = \int_{\theta'_m}^{\infty} \theta.(\frac{N + \alpha - 1}{\theta'_m})(\theta'_m/\theta)^{N+\alpha}\, d\theta$$

$$\Rightarrow \hat{\theta}^{PosteriorMean} = \int_{\theta'_m}^{\infty} (N + \alpha - 1)(\theta'_m)^{N+\alpha-1}.(\theta^{-N-\alpha+1})\, d\theta$$

$$\Rightarrow \hat{\theta}^{PosteriorMean} = \frac{(N + \alpha - 1)}{(N + \alpha - 2)}(\theta'_m)$$

6 Comparison Posterior-Mean and ML estimates

As $N \to \infty$ we have $\hat{\theta}^{PosteriorMean} \to \theta'_m$ again using the previous argument,

- if $\theta_m \leq \theta_{true}$

  $$\hat{\theta}^{PosteriorMean} = \frac{(N + \alpha - 1)}{(N + \alpha - 2)}.max(max(x1, x2, \ldots, x_n), \theta_m) \to \theta_{true}$$

  $$\hat{\theta}^{ML} = max(x1, x2, \ldots, x_n) \to \theta_{true}$$

  $$\Rightarrow \hat{\theta}^{PosteriorMean} \to \hat{\theta}^{ML}$$

  (Since all the data lies in [0, θ_{true}] and $\theta_m \leq \theta_{true}$)
  This case is desirable since the Maximum likelihood estimate has asymptotically the lowest Mean Squared Error and hence $\hat{\theta}^{PosteriorMean}$ is a good estimator.

- if $\frac{(N+\alpha-1)}{(N+\alpha-2)}.\theta_m > \theta_{true}$

  $$\hat{\theta}^{PosteriorMean} = \frac{(N + \alpha - 1)}{(N + \alpha - 2)}.max(max(x1, x2, \ldots, x_n), \theta_m) = \frac{(N + \alpha - 1)}{(N + \alpha - 2)}.\theta_m \quad \forall N$$

  $$\hat{\theta}^{ML} = max(x1, x2, \ldots, x_n) \to \theta_{true}$$

  $$\Rightarrow \hat{\theta}^{PosteriorMean} > \hat{\theta}^{ML}$$

  (Since all the data lies in [0, θ_{true}] and $\theta_m > \theta_{true}$)

> This case is undesirable because the MAP estimate always has a positive bias whose minimum value is $(\theta_m - \theta_{true})$, so its Mean Squared Error never tends to 0. Whereas the ML-estimate's bias tends to 0 asymptotically and its Mean Squared Error also tends to 0.

In many cases due to a bad choice of prior Model, $\hat{\theta}^{PosteriorMean}$ would not tend to $\hat{\theta}^{ML}$ and this is undesirable and makes the Posterior Mean estimator a poor estimator.
For a correctly chosen Prior Model the Posterior Mean estimate would tend to the ML estimate and also may be slightly better than it.