Report for CS215 Assignment 3

AMEEYA RANJAN SETHY - 200050006

October 30, 2021

Question 3

Let the Data D = $x_1, x_2, ..., x_n$, $n \geq 1$. Since the samples are drawn from a uniform distribution on $(0, \theta)$, we assume that each sample $x_1 \geq 0$ because the possibility of any x_i being negative is confirmed to be almost impossible.

- Likelihood:

$$P(x|\theta) = \begin{cases} (\frac{1}{\theta}) & \text{if } \mathrm{x}_i \in (0, \theta) \\ 0 & \text{otherwise.} \end{cases}$$

$$P(x|\theta) = \frac{1}{\theta}(x \in (0, \theta))$$

- Prior:

$$P(\theta) \propto \frac{\theta_m^\alpha}{\theta^\alpha} I(\theta \geq \theta_m)$$

So the joint Likelihood is:

$$P(\{x_i\}, \theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} I(x_i \in (0, \theta))$$

A MLE and MAP

By looking at the joint Likelihood above, we have:

$$P(\{x_i\}|\theta) = \begin{cases} (\frac{1}{\theta})^n & \text{if } \mathrm{x}_i \in (0, \theta) \forall i \\ 0 & \text{otherwise.} \end{cases}$$

since if even one x_i is not in this range, the indicator will become 0 and so will the joint likelihood. Since we wish to maximize the joint likelihood, and since $(\frac{1}{\theta})^n > 0 \forall \theta \in R$, we consider the first case where we must have $\forall i\ x_i \in (0, \theta)$ This is equivalent to $\theta \geq \max(x_1, x_2, ..., x_n)$, since $x_i \leq \theta\ \forall i$, thus the largest of them must also be smaller than θ, and conversely if the largest of them is smaller than θ, then so are all of them.

Thus, maximizing the joint likelihood $(\frac{1}{\theta})^n$ i.e minimizing θ subject to $\theta \geq \max(x_1, x_2, ..., x_n)$ means

$$\hat{\theta}^{MLE} = max(x_1, x_2, ..., x_n)$$

Now the Posterior Distribution:

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

$$P(\theta|\{x_i\}) \propto P(\{x_i\}|\theta)P(\theta))$$

$$P(\theta|\{x_i\}) \propto \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} I(x_i \in (0, \theta)) \times \frac{\theta_m^\alpha}{\theta^\alpha} I(\theta \geq \theta_m)$$

$$P(\theta|\{x_i\}) \propto \left(\frac{\theta_m^\alpha}{\theta^{n+\alpha}}\right) \left(\prod_{i=1}^{n} I(x_i \in (0, \theta))\right) I(\theta \geq \theta_m)$$

Thus finally we get

$$P(\theta|\{x_i\}) = \begin{cases} (\frac{1}{\theta}) & \text{if } \forall i \ x_i \in (0,\theta) \ and \ \theta \geq \theta_m \\ 0 & \text{otherwise.} \end{cases}$$

Thus, similarly, we must enforce $\forall i x_i \in (0,\theta)$ and $\theta \geq \theta_m$. This is equivalent to $\theta \geq \max(x_1, x_2, ..., x_n, \theta_m)$. And since maximizing posterior i.e maximizing $\frac{\theta_m^\alpha}{\theta^{n+\alpha}}$ is equivalent to minimizing θ since $n + \alpha > 2$, so denominator's power is positive, we again get:

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

B MAP vs MLE Asymptotically

Yes, MAP estimate tends to MLE estimate as sample size tends to infinity. This is because of the following:

$$max(x_1, x_2, ..., x_n, \theta_m) \neq max(x_1, x_2, ..., x_n) \iff \forall i \ \theta_m \geq x_i$$

i.e.

$$Pr(max(x_1, x_2, ..., x_n, \theta_m) \neq max(x_1, x_2, ..., x_n)) = \left(\frac{\theta_m}{\theta}\right)^n$$

since $\Pr(\theta > \theta_m) = 1$(all the probability mass exists after θ_m). As $n \to \infty$, the RHS will tend to 0. Thus, the

$$Pr(max(x_1, x_2, ..., x_n, \theta_m) \neq max(x_1, x_2, ..., x_n)) = \left(\frac{\theta_m}{\theta}\right)^n \to 0$$

$$Pr(max(x_1, x_2, ..., x_n, \theta_m) = max(x_1, x_2, ..., x_n)) = \left(\frac{\theta_m}{\theta}\right)^n \to 1$$

$$max(x_1, x_2, ..., x_n, \theta_m) = max(x_1, x_2, ..., x_n) always$$

i.e the MAP almost surely tends to the MLE. This is desirable because, since the ML estimate is the best estimate asymptotically(no other consistent estimator has a lower asymptotic MSE than the ML estimate), the MAP estimate is getting all the nice asymptotic properties (consistency, asymptotic normality, efficiency) of the ML estimate.

C Mean of the Posterior Distribution

As seen in the previous proof,

$$P(\theta|\{x_i\}) \propto \left(\frac{\theta_m^\alpha}{\theta^{n+\alpha}}\right)\left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) I(\theta \geq \theta_m)$$

Thus,

$$P(\theta|\{x_i\}) = K\left(\frac{\theta_m^\alpha}{\theta^{n+\alpha}}\right)\left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) I(\theta \geq \theta_m)$$

where K is a normalization constant independent of θ. Let us obtain it now. To do so, we shall integrate over all θ and since the LHS is a posterior distribution, it integrates to one. Thus,

$$1 = \int_{-\infty}^{\infty} K\left(\frac{\theta_m^\alpha}{\theta^{n+\alpha}}\right)\left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) I(\theta \geq \theta_m)$$

$$1 = K \times \theta_m^\alpha \int_{-\infty}^{\infty} \frac{1}{\theta^{n+\alpha}}\left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) I(\theta \geq \theta_m)$$

The integrand here is non-zero only when

$$\forall 1, x_i \in (0,\theta) \text{ and } \quad \text{i.e when } \theta \geq max(x_1, x_2, ..., x_n, \theta_m)$$

Thus,

$$1 = K \times \theta_m^\alpha \int_{max(x_1,x_2,....,x_n,\theta_m)}^{\infty} \frac{1}{\theta^{n+\alpha}}$$

$$1 = K \times \theta_m^\alpha \times \frac{1}{1-(n+\alpha)} \times \left(\theta^{1-n+\alpha}\right)\Big|_{max(x_1,x_2,....,x_n,\theta_m)}^{\infty}$$

Since $1-(n+\alpha) < 0$, since $n \geq 1$, $\alpha \geq 1$, it evaluates to 0 at ∞. Thus,

$$1 = K \times \theta_m^\alpha \times \frac{-max(x_1,x_2,....,x_n,\theta_m)^{1-(n+\alpha)}}{1-(n+\alpha)}$$

$$1 = K \times \theta_m^\alpha \times \frac{max(x_1,x_2,....,x_n,\theta_m)^{1-(n+\alpha)}}{(n+\alpha)-1}$$

Thus,

$$K = \frac{(n+\alpha)-1}{\theta_m^\alpha \times max(x_1,x_2,....,x_n,\theta_m)^{1-(n+\alpha)}}$$

Now, to find the mean, we find the expectation of θ over the posterior distribution.

$$E_{P(\theta|\{x_i\})}(\theta|\{x_i\}) = \int_{-\infty}^{\infty} \theta \times K \times \frac{(\theta_m)^\alpha}{\theta^{n+\alpha}} \times \left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) \times I(\theta \geq \theta_m)$$

$$E_{P(\theta|\{x_i\})}(\theta|\{x_i\}) = \int_{-\infty}^{\infty} K \times \frac{(\theta_m)^\alpha}{\theta^{n+\alpha-1}} \times \left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) \times I(\theta \geq \theta_m)$$

$$E_{P(\theta|\{x_i\})}(\theta|\{x_i\}) = K \times (\theta_m)^\alpha \int_{-\infty}^{\infty} \frac{1}{\theta^{n+\alpha-1}} \times \left(\prod_{i=1}^{n} I(x_i \in (0,\theta))\right) \times I(\theta \geq \theta_m)$$

Thus,

$$= K \times (\theta_m)^\alpha \int_{max(x_1,x_2,....,x_n,\theta_m)}^{\infty} \frac{1}{\theta^{n+\alpha-1}}$$

$$= K \times (\theta_m)^\alpha \times \frac{1}{2-(n+\alpha)} \times \left(\theta^{2-(n+\alpha)}\right)\Big|_{max(x_1,x_2,....,x_n,\theta_m)}^{\infty}$$

Since $2-(n+\alpha) < 0$, since n ≥ 1, $\alpha \geq 1$, it evaluates to 0 at ∞. Thus,

$$= K \times (\theta_m)^\alpha \times \frac{-max(x_1,x_2,...,x_n,\theta_m)^{2-(n+\alpha)}}{2-(n+\alpha)}$$

$$= \frac{(n+\alpha)-1}{\theta_m^\alpha \times max(x_1,x_2,...,x_n,\theta_m)^{1-(n+\alpha)}} \times (\theta_m)^\alpha \times \frac{max(x_1,x_2,...,x_n,\theta_m)^{2-(n+\alpha)}}{2-(n+\alpha)}$$

$$= \frac{(n+\alpha)-1}{(n+\alpha)-2} \times max(x_1,x_2,....,x_n,\theta_m)$$

Thus,

$$E_{P(\theta|\{x_i\})}(\theta|\{x_i\}) = \frac{(n+\alpha)-1}{(n+\alpha)-2} \times max(x_1,x_2,....,x_n,\theta_m)$$

D Mean of posterior and MLE asymptotically

We can clearly see that since

$$\lim_{n\to\infty} \frac{(n+\alpha)-1}{(n+\alpha)-2} = 1$$

and using the previous result, the function

$$E_{P(\theta|\{x_i\})}(\theta|\{x_i\}) = \frac{(n+\alpha)-1}{(n+\alpha)-2} \times max(x_1, x_2, ..., x_n, \theta_m)$$

asymptotically almost surely tends pointwise to the function $\max(x_1, x_2, ..., x_n)$ (one function converges to the other at every point) and thus the posterior mean tends to the ML estimate as $n \to \infty$.

This is, as explained before, desirable since the ML estimate is the best estimate asymptotically(no other consistent estimator has a lower asymptotic MSE than the ML estimate), the estimate obtained using the posterior mean is getting all the nice asymptotic properties(consistency, asymptotic normality, efficiency) of the ML estimate.