CS 215 Assignment 3 Report

Diwan Anuj Jitendra - 170070005 Soumya Chatterjee - 170070010

October 21, 2018

Problem 4

Let the data $D = \{x_1, x_2, \ldots, x_n\}, n \geq 1$.(If $n = 0$, there is no data!) Since the samples are drawn from a uniform distribution on $(0, \theta)$, we assume each sample $x_i \geq 0$ since $Pr(x_i < 0) = 0$ i.e. the event of any x_i being negative is *almost surely* impossible.

- Likelihood:

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

  i.e

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

- Prior:

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

Thus the joint likelihood is:

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

0.1 MLE and MAP

Seeing the joint likelihood above, we have:

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

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 $\left(\frac{1}{\theta}\right)^n > 0 \ \forall \ \theta \in \mathbb{R}$, we consider the first case where we must have $\forall i, x_i \in (0, \theta)$. This is equivalent to $\theta \geq \max(x_1, x_2, \ldots, x_n)$, 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 $\left(\frac{1}{\theta}\right)^n$ i.e minimizing θ subject to $\theta \geq \max(x_1, x_2, \ldots, x_n)$ means

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

Now, the posterior distribution:

$$\begin{aligned} P(\theta|\{x_i\}) &= \frac{P(\{x_i\}|\theta)P(\theta)}{P(\{x_i\})} \\ &\sim P(\{x_i\}|\theta)P(\theta) \\ &\sim \left(\frac{1}{\theta}\right)^n \prod_{i=1}^{n} \mathrm{I}(x_i \in (0,\theta)) \times \frac{(\theta_m)^\alpha}{\theta^\alpha} \mathrm{I}(\theta \geq \theta_m) \\ &\sim \frac{(\theta_m)^\alpha}{\theta^{n+\alpha}} \left(\prod_{i=1}^{n} \mathrm{I}(x_i \in (0,\theta))\right) \mathrm{I}(\theta \geq \theta_m) \end{aligned}$$

Thus,

$$P(\theta|\{x_i\}) \sim \begin{cases} \frac{(\theta_m)^\alpha}{\theta^{n+\alpha}} & \text{if } \forall i, x_i \in (0,\theta) \text{ and } \theta \geq \theta_m \\ 0 & \text{otherwise.} \end{cases}$$

Thus, similarly, we must enforce $\forall i, x_i \in (0,\theta)$ and $\theta \geq \theta_m$. This is equivalent to $\theta \geq \max(x_1, x_2, \ldots, x_n, \theta_m)$. 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:

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

0.2 MAP vs MLE asymptotically

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

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

. i.e

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

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

$$\Pr(\max(x_1, x_2, \ldots, x_n, \theta_m) \neq \max(x_1, x_2, \ldots, x_n)) \to 0$$

i.e

$$\Pr(\max(x_1, x_2, \ldots, x_n, \theta_m) = \max(x_1, x_2, \ldots, x_n)) \to 1$$

i.e.

$$(\max(x_1, x_2, \ldots, x_n, \theta_m) \to \max(x_1, x_2, \ldots, x_n)) \text{ a.s.}$$

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.

0.3 Mean of the posterior distribution

As seen in the previous proof,

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

Thus,

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

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,

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

The integrand here is non-zero only when $\forall i, x_i \in (0,\theta)$ and $\theta \geq \theta_m$ i.e when

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

.Thus,

$$\begin{aligned}
&= K \times {\theta_m}^\alpha \int_{\max(x_1,x_2,\ldots,x_n,\theta_m)}^{\infty} \frac{1}{\theta^{n+\alpha}} \\
&= K \times {\theta_m}^\alpha \times \frac{1}{1-(n+\alpha)} \times (\theta^{1-(n+\alpha)})\Big|_{\max(x_1,x_2,\ldots,x_n,\theta_m)}^{\infty}
\end{aligned}$$

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

$$\begin{aligned}
&= K \times {\theta_m}^\alpha \times \frac{-\max(x_1, x_2, \ldots, x_n, \theta_m)^{1-(n+\alpha)}}{1-(n+\alpha)} \\
&= K \times {\theta_m}^\alpha \times \frac{\max(x_1, x_2, \ldots, x_n, \theta_m)^{1-(n+\alpha)}}{(n+\alpha)-1}
\end{aligned}$$

Thus,

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

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

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

The integrand here is non-zero only when $\forall i, x_i \in (0,\theta)$ and $\theta \geq \theta_m$ i.e when

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

.Thus,

$$\begin{aligned}
&= K \times {\theta_m}^\alpha \int_{\max(x_1,x_2,\ldots,x_n,\theta_m)}^{\infty} \frac{1}{\theta^{n+\alpha-1}} \\
&= K \times {\theta_m}^\alpha \times \frac{1}{2-(n+\alpha)} \times (\theta^{2-(n+\alpha)})\Big|_{\max(x_1,x_2,\ldots,x_n,\theta_m)}^{\infty}
\end{aligned}$$

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

$$\begin{aligned}
&= K \times {\theta_m}^{\alpha} \times \frac{-\max(x_1,x_2,\ldots,x_n,\theta_m)^{2-(n+\alpha)}}{2-(n+\alpha)} \\
&= \frac{(n+\alpha)-1}{{\theta_m}^{\alpha}(\max(x_1,x_2,\ldots,x_n,\theta_m)^{1-(n+\alpha)})} \times {\theta_m}^{\alpha} \times \frac{\max(x_1,x_2,\ldots,x_n,\theta_m)^{2-(n+\alpha)}}{(n+\alpha)-2} \\
&= \frac{(n+\alpha)-1}{(n+\alpha)-2} \times \frac{\max(x_1,x_2,\ldots,x_n,\theta_m)^{2-(n+\alpha)}}{\max(x_1,x_2,\ldots,x_n,\theta_m)^{1-(n+\alpha)}} \times \frac{{\theta_m}^{\alpha}}{{\theta_m}^{\alpha}} \\
&= \frac{(n+\alpha)-1}{(n+\alpha)-2}\max(x_1,x_2,\ldots,x_n,\theta_m)
\end{aligned}$$

Thus,

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

0.4 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

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

asymptotically almost surely tends pointwise to the function $\max(x_1,x_2,\ldots,x_n)$ (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.