200050030 - 200050138

CS215: Data Analysis & Interpretation
Assignment: Bayesian Estimation

Report for Problem 3

Given,

Random variable $X \sim U([0,\theta])$ where U represents uniform distribution with unknown parameter θ.

We consider random variable Θ and (for parameter θ) Pareto distribution prior on θ,

$$P_{\Theta}(\theta) \propto (\theta_m/\theta)^{\alpha} \quad \text{for } \theta \geq \theta_m$$

$$P_{\Theta}(\theta) = 0 \quad \text{otherwise}$$

where $\theta_m > 0$

$\alpha > 1$.

(i) Find maximum-likelihood estimate $\hat{\theta}^{ML}$ and the maximum-a-posteriori estimate $\hat{\theta}^{MAP}$:

Consider, n-samples drawn from $X|\Theta$

$x_1, x_2, \ldots, x_n$ not necessarily in that order.

$$P(X=x_i) = \frac{1}{\theta} \cdot 1(0 \leq x_i \leq \theta)$$

where $1(0 \leq x_i \leq \theta)$ takes value 1 if $0 \leq x_i \leq \theta$ or 0 otherwise.

We now define likelihood function for θ as

$$L(\theta: x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} \cdot \prod_{i=1}^{n} 1(x_i \le \theta)$$

(Considering samples drawn are greater than 0)

The above expression can be simplified when we consider

$$c_1 = \max(x_1, x_2, \dots, x_n).$$

and

$\prod_{i=1}^{n} 1(x_i \le \theta)$ can be simplified

into $1(c_1 \le \theta)$

since the former expression would take value 0 if atleast one of $x_i > \theta$ (which would be greater than other x_j's with $x_j \le \theta$)

Hence, Likelihood function is given as

$$L(\theta: x_1, x_2, \dots, x_n) = \frac{1}{\theta^n} \cdot 1(c_1 \le \theta)$$

~~and log-likelihood function $\ell(\theta: x_1, \dots, x_n) = \log(L(\theta: x_1, \dots, x_n))$ is given as~~

From

$$L(\theta: x_1, x_2, \ldots, x_n) = \frac{1}{\theta^n} \cdot 1(c_1 \leq \theta)$$

$L(\theta) = 0$ if $\theta < c_1 = \max(x_1, \ldots, x_n)$.

$L(\theta) = \frac{1}{\theta^n}$ if $\theta \geq$ ~~most~~ $c_1 = \max(x_1, \ldots, x_n)$

in order to maximise $L(\theta)$ we need to minimize θ to a value $\hat{\theta}^{ML}$ such that $\hat{\theta}^{ML}$ also satisfies $\theta \geq c_1$

Hence, we get $\hat{\theta}^{ML} = c_1 = \max(x_1, \ldots, x_n)$ as the maximum likelihood estimation for θ.

For the maximum-a-posteriori estimate, we have,

prior, $P_\theta(\theta) \propto \left(\frac{\theta_m}{\theta}\right)^\alpha$

$P(x|\theta) \sim U([0, \theta])$

posterior, $P(\theta|x) \propto P(x|\theta) \cdot P(\theta)$ (Bayes's Rule)

For posteriori estimation, we have.

$$P(\theta | x_1, x_2, \ldots x_n) \propto P(x_1, x_2, \ldots x_n | \theta) \cdot P(\theta)$$

$$P(\theta | x_1, x_2, \ldots, x_n) \propto \frac{1}{\theta^n} \cdot 1(c_1 \leq \theta) \cdot \left(\frac{\theta_m}{\theta}\right)^{\alpha} \cdot 1(\theta_m \leq \theta).$$

$$P(\theta | x_1, x_2, \ldots, x_n) \propto \frac{\theta_m^{\alpha}}{\theta^{n+\alpha}} \cdot 1(\theta \geq \max(c_1, \theta_m))$$

$$\propto \frac{1}{\theta_{m_2}^n} \cdot \left(\frac{\theta_{m_2}}{\theta}\right)^{n+\alpha} \cdot 1(\theta \geq \theta_{m_2})$$

where $\theta_{m_2} = \max(c_1, \theta_m)$.

Here, we get

$P(\theta | x_1, x_2, \ldots, x_n)$ as having

a Pareto distribution with

parameter

$n+\alpha$, $\theta_{m_2} = \max(c_1, \theta_m)$

Here, again, if $\theta < \theta_{m_2}$

$P(\theta | x_1, \ldots x_n) = 0$

if $\theta \geq \theta_{m_2}$ $P(\theta | x_1, \ldots, x_n) \propto \frac{1}{(\theta)^{n+\alpha}}$ which decreases for $\theta \geq \theta_{m_2}$.

Hence, to find ~~maximum~~ posterior estimate we minimise θ to satisfy $\theta \geq \theta_{m_2}$, ($\because$ inverse proportionality)

Hence, we get $\hat{\theta}^{MAP} = \theta_{m_2} = \max(c_1, \theta_m)$.

Hence,

$$\hat{\theta}^{ML} = c_1 = \max(x_1, x_2, \ldots, x_n)$$

$$\hat{\theta}^{MAP} = \max(c_1, \theta_m).$$

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

As the sample size tends to infinity,

consider the quantity

$$\Delta\theta = \hat{\theta}^{MAP} - \hat{\theta}^{ML}$$

$$= \max(c_1, \theta_m) - c_1$$

if $c_1 \geq \theta_m$

$\Delta\theta = 0$.

else, $c_1 \leq \theta_m$, we have

$\Delta\theta = \theta_m - c_1$

But, considering $\theta \geq \theta_m$ and some positive $\epsilon = \theta - \theta_m > 0$ and

$\lim_{n \to \infty} (x_1, x_2, ----, x_n)$ samples drawn

for $\theta - \epsilon$, there might be corresponding samples with probability $\frac{1}{\theta}$ for the n samples ($\because n \to \infty$)

and hence, $c_1 = \max(x_1, x_2, ---, x_n) \geq \theta - \epsilon$

$\Rightarrow c_1 \geq \theta - \epsilon = \theta_m$

which contradicts the fact that:

but $c_1 \leq \theta_m$ (for the case considered)

$\because c_1 \leq \theta_m$ and $c_1 \geq \theta_m$

$c_1 - \theta_m \to 0$ as $n \to \infty$

$\Rightarrow \Delta\theta \to 0.$ as $n \to \infty$

Hence, $\hat{\theta}^{MAP} - \hat{\theta}^{ML} \to 0$ as $n \to \infty$

$\Rightarrow \hat{\theta}^{MAP} \to \hat{\theta}^{ML}$ as $n \to \infty$.

This is desirable since we get a definite/precise estimation for the maximum likelihood of θ with or without the information of the posterior obtained from the prior as sample size tends to infinity.

(iii) Find estimator of the mean of the posterior distribution $\hat{\theta}$ Posterior Mean

We have the posterior distribution

$$P(\theta|x_1, x_2, \ldots, x_n) \sim \text{Pareto}(n+\alpha, \theta_{m_2})$$

for pareto distribution Pareto(α, θ_m),

we find mean as follows

$$P(\theta) = k\left(\frac{\theta_m}{\theta}\right)^{\alpha} \cdot I(\theta \geq \theta_m).$$

$$\theta_{mean} = \frac{\int_{\theta_m}^{\infty} k\theta \left(\frac{\theta_m}{\theta}\right)^{\alpha} d\theta}{\int_{\theta_m}^{\infty} k\left(\frac{\theta_m}{\theta}\right)^{\alpha} d\theta}$$

$$= \frac{\frac{1}{2-\alpha} \cdot \frac{1}{\theta_m^{\alpha-2}}}{\frac{1}{1-\alpha} \cdot \frac{1}{\theta_m^{\alpha-1}}}$$

$\Rightarrow \theta_{mean} = \theta_m \cdot \frac{\alpha-1}{\alpha-2}$

Hence, for the posterior distribution,

$$\hat{\theta}_{mean \; P(\theta|x_1, x_2, \ldots, x_n)} = \theta_{m_2} \cdot \frac{n+\hat{\alpha}-1}{n+\hat{\alpha}-2}$$

where we estimat $\hat{\alpha}$ for the posterior distribution.

for some n_2 draws from posterior distribution.

we define likelihood function for α

as $L(\alpha : \theta_1, \theta_2, \ldots, \theta_{n_2}) \propto \prod_{i=1}^{n_2} \left(\frac{\theta_{m_2}}{\theta_i}\right)^{n+\alpha}$

log-likelihood,

$$\log [L(\alpha : \theta_1, \ldots, \theta_{n_2})] = C_3 + n_2(n+\alpha) \log \theta_{m_2} - (n+\alpha) \sum_{i=1}^{n_2} \log \theta_i$$

and $\frac{\partial l(\alpha : \theta_1, \ldots, \theta_{n_2})}{\partial \alpha} = \frac{\partial C_3}{\partial \alpha} + n_2 \log \theta_{m_2} - \sum \log \theta_i$

We get, $c_3 = \log(\alpha^{n_2}) = n_2 \log \alpha$

and hence,

$$\frac{n_2}{\alpha} + n_2 \log \theta_{m_2} - \sum_{i=1}^{n_2} \ln(\theta_i') = 0$$

$$\boxed{\hat{\alpha} = \frac{n_2}{\sum_{i=1}^{n_2} \ln(\theta_i') - n_2 \log \theta_{m_2}}}$$

and hence the $\hat{\theta}$ posterior mean is given

as $\hat{\theta}^{\text{Posterior Mean}} = \frac{n+\hat{\alpha}-1}{n+\hat{\alpha}-2} \cdot \theta_{m_2}$

where $\hat{\alpha}$ is found over n_2 sample

$$\hat{\alpha} = \frac{n_2}{\sum_{i=1}^{n_2} \ln(\theta_i) - n_2 \log \theta_{m_2}}$$

(iv) Does $\hat{\theta}$ Posterior Mean tend to $\hat{\theta}^{ML}$ as the sample size tends to infinity? Is this desirable or not?

~~As~~

As sample size $n \to \infty$,

$$\lim_{n\to\infty} \hat{\theta}^{\text{Posterior Mean}} = \frac{n+\hat{\alpha}-1}{n+\hat{\alpha}-2} \cdot \theta_{m_2} = \theta_{m_2} = \hat{\theta}^{MAP} \underset{\text{as } n\to\infty}{\to} \hat{\theta}^{ML}$$

Hence, as sample size, $n \to \infty$, $\hat{\theta}^{\text{Posterior Mean}}$

tends to $\hat{\theta}^{ML}$.

Since Posterior Mean gives us information about the distribution of ~~the~~ θ based on prior, it is important to know how various estimators differ from one another to come to a conclusion on the precision of estimation, since $\hat{\theta}^{\text{Posterior Mean}}$ tends to $\hat{\theta}^{ML}$ as sample size tends to infinity, we understand that the estimation can be precise and it is also desirable for prediction of data.