HW1 R Markdown Poisson_1

Santanu Mukherjee

9/19/2021

R Markdown

Problem 1 a, b, c

Conjugate Prior, Posterior Distribution, Posterior Mean & Variance.

STA6113 SANTANU MUKHERJEE Zes254

Problem 1

The data distribution given is Poisson

$$P(Y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!} I(y \in \{0,1,2\ldots\}).$$

(a) for the Poisson data model, which is a data model within the exponential family, the conjugate families are always available, which means the prior and posterior distribution is GAMMA. So the conjugate prior is GAMMA

(b) $$P(\lambda|y) = \frac{P(y|\lambda)P(\lambda)}{P(y)}$$

$$P(\lambda|y) \propto \underbrace{P(y|\lambda)}_{\text{Likelihood function}} \underbrace{P(\lambda)}_{\text{Prior}} \to ①$$

[y is the given data and is independent of λ, so treated as constant)

As we said, $P(\lambda)$ prior is Gamma distribution.

$$\lambda \sim G(a,b), \quad P(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$\text{So } P(\lambda) \propto \lambda^{a-1} e^{-b\lambda} \to ②$$ (Omitting the terms independent of λ, treating as constant)

So using ②, substituting in ①,

$$P(\lambda|y) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i} \lambda^{a+1} e^{-b\lambda}$$ (Omitting the constants not depending on λ)

$$P(\lambda|y) \propto e^{-n\lambda} \lambda^{n\bar{y}+a-1} e^{+b\lambda}$$

$$P(\lambda|y) \propto \lambda^{n\bar{y}+a-1} e^{-(b+n)\lambda}$$

(c) Posterior mean of λ/y :- $(n\bar{y}+a)/b+n$

Posterior variance of λ/y :- $(n\bar{y}+a)/(b+n)^2$

Problem 1 d

MLE of Lambda,Relationship between MLE & Posterior Mean.

(d) MLE of λ can be found out by first finding the log likelihood of Poisson distribution and then find the MLE for λ.

Likelihood function of λ is

$$L(\lambda, y_1 \ldots y_n) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

(LE) Log likelihood function is

$$\log L(\lambda, y_1 \ldots y_n) = -n\lambda - \sum_{i=1}^{n} \ln(y_i!) + \ln(\lambda) \sum_{i=1}^{n} y_i$$

To find the MLE, we have to do first order derivative $= 0$

$$\frac{d}{d\lambda}\left(\log L(\lambda, y_1, \ldots y_n)\right) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i$$

$$0 = -n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i$$

$$\hat{\lambda} = \bar{y}.$$

From (c) we found out that Posterior Mean of λ $= \dfrac{n\bar{y} + a}{b+n}$

$$\frac{n\bar{y} + a}{b+n} = \left(\frac{n}{b+n}\right)\underbrace{\bar{y}}_{\text{MLE}} + \left(\frac{b}{b+n}\right)\underbrace{\left(\frac{a}{b}\right)}_{\text{prior mean}}$$

Posterior Mean $=$ weighted average of (MLE) $+$ weighted average of Prior Mean

Problem 1 e

Here we are generating 10 random Poisson variables. So, n = 10, lambda = 3

```
## [1] "Equal_Tailed Credible Interval: 1.6176 3.4633"
```

```
## [1] "HPD Credible Interval: 1.5653 3.3959"
```

```
## [1] "Equal Tailed Range: 1.8457"
```

```
## [1] "HPD Range: 1.8306"
```

Problem 1 f

Here is the dataset that I have used : **2018 World Cup Soccer Stats (https://fbref.com/en/comps/1/schedule/FIFA-World-Cup-Scores-and-Fixtures)**. My interest is on the average goals scored **2.64**, which is my lambda. For reference: The total matches played was **64** and total goals scored was **169**. This dataset can be used as a Poisson distribution because:

• (y goals in a match - occurs at fixed times and y = 0,1,2,..) • The occurrence of k goals in a match does not impact the probability of goals in the second match. That means events occur independently. • They are identically distributed. • 2 goals cannot happen at the same instant in the same match.

Because of the above conditions y is a **Poisson random variable.** and the distribution of y is a **Poisson distribution.**

Problem 1 g

Here lambda is 2.64, n = 64, a = 2, b = 1 (values of a and b are assumed)

```
##  [1] 5 1 1 6 3 2 1 2 1 1 2 1 3 3 3 3 4 1 1 1 2 1 3 2 2 3 7 3 3 7 4 3 3 3 4 2 0 2
## [39] 3 3 2 3 4 2 1 1 1 3 7 3 2 2 2 5 1 2 2 3 2 4 1 3 2 6
```

```
## [1] "Posterior Mean Estimate is : 2.63076923076923"
```

```
## [1] "Sum of Weighted Averages of MLE and Prior Mean is : 2.63076923076923"
```

Based on the data, we can say that the posterior mean is the weighted averages of the MLE and prior mean (the prior and posterior are both Gamma with hyper parameters a and b)

```
##  [1] 5 1 1 6 3 2 1 2 1 1 2 1 3 3 3 3 4 1 1 1 2 1 3 2 2 3 7 3 3 7 4 3 3 3 4 2 0 2
## [39] 3 3 2 3 4 2 1 1 1 3 7 3 2 2 2 5 1 2 2 3 2 4 1 3 2 6
```

```
## [1] "Equal_Tailed Credible Interval: 2.2512 3.0394"
```

```
## [1] "HPD Credible Interval: 2.2415 3.0287"
```

```
## [1] "Equal Tailed Range: 1.4218"
```

```
## [1] "HPD Range: 1.4634"
```

Problem 2 a

Posterior distribution of Population Mean.

Problem 2

Given prior distribution for $\mu \sim N(0,1)$.

It is also given that sagi $\sim N(\mu, \sigma^2)$,

(a) meaning $P(x/\mu)$, e.g. $X_i \sim N(\mu, \sigma^2)$

where $\sigma = 0.25$.

$\bar{x}$ = mean of $\{5.19, 4.72, 4.81, 4.87, 4.88\}$

$= 4.894$.

$n = 5$,

$\sigma = 0.25$, so, $\sigma^2 = 0.0625$

We need to find the distribution of $P(\mu|x)$

We know,

$$P(\mu|x) \propto P(x|\mu)\, P(\mu)$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{1}{2}\mu^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\sum x_i^2 - 2n\bar{x}\mu + n\mu^2 + \sigma^2\mu^2\right]\right\}$$

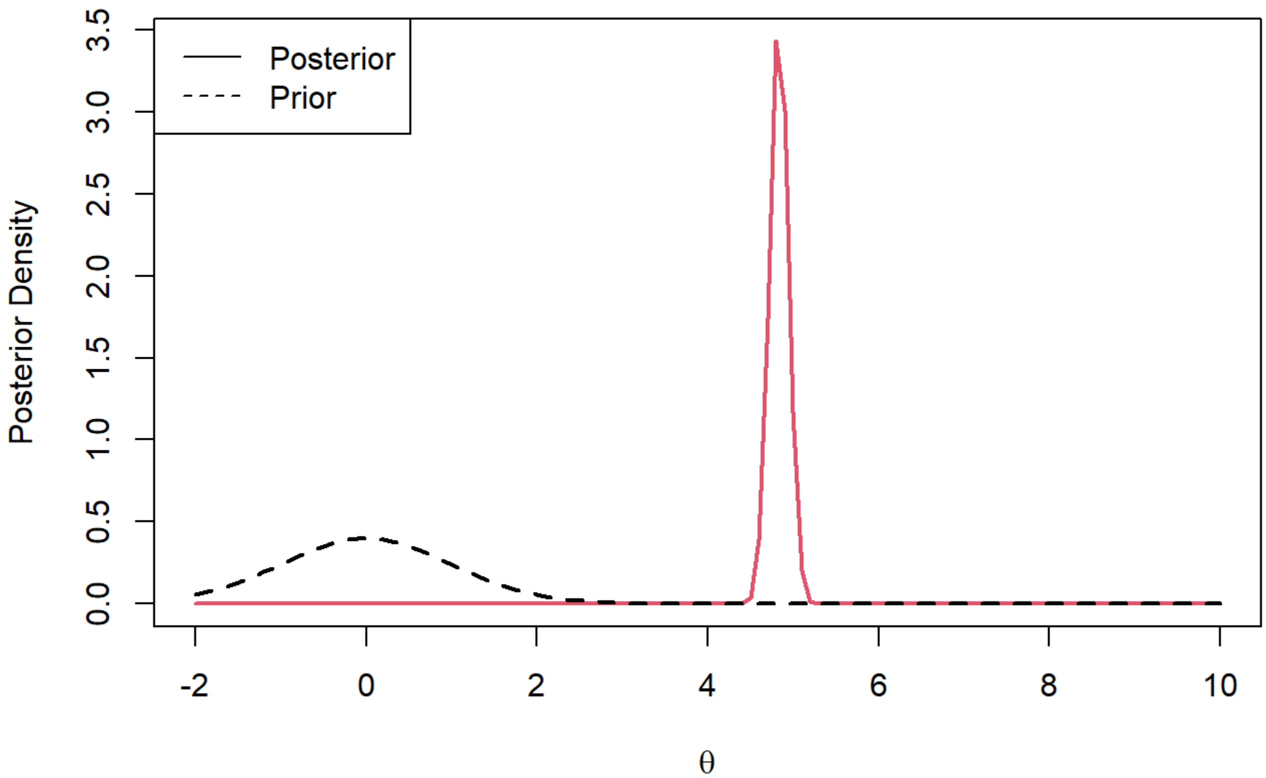
$$\propto \exp\left\{-\frac{1}{2\frac{\sigma^2}{n+\sigma^2}}\left[\mu - \frac{n\bar{x}}{n+\sigma^2}\right]^2\right\} \cancel{\exp\left\{-\frac{1}{2\sigma^2}\sum x_i^2\right\}}$$

So $P(\mu|x)$ is normally distributed with Variance $= \frac{\sigma^2}{n+\sigma^2}$

and mean $= \frac{n\bar{x}}{n+\sigma^2}$.

$$P(\mu|x) \sim N\left(\frac{n}{n+\sigma^2}\bar{x}, \frac{\sigma^2}{n+\sigma^2}\right)$$

Problem 2 b


Posterior
Prior
Posterior Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-2
0
2
4
6
8
10
θ