Indian Institute of Management
Ahmedabad

e-Post Graduate Diploma in Advanced Business Analytics, 2021-2022

Bayesian Analysis
Assignment

Bharath Dasari
PGDAE202100006

Bayesian Analysis Assignment

Question 1:

Q1)

x	0.415	0.409
$p(x)$	0.5	0.5

$$E(x) = \Sigma x \cdot p(x)$$
$$= 0.415 \times 0.5 + 0.409 \times 0.5$$
$$= 0.412$$

$$Var(x) = E(x^2) - [E(x)]^2$$

$$E(x^2) = \Sigma x^2 \cdot p(x)$$
$$= (0.415)^2(0.5) + (0.409)^2(0.5)$$
$$= 0.1697$$

$$Var(x) = 0.1697 - (0.412)^2$$
$$\simeq 0.000009$$

Now, matching moments

$$E(x) = 0.412 = \frac{a}{a+b}$$

$$Var(x) = 0.000009 = \frac{ab}{(a+b)^2(a+b+1)}$$

$$\frac{a}{a+b} = 0.412$$

$$a = 0.412a + 0.412b$$

$$0.588a = 0.412b$$

$$\boxed{a = 0.7b.} \rightarrow 1$$

$$\cancel{\frac{0.7b}{\frac{0.7b+b}{0.7b}}} = \cancel{0.412}$$

$$\frac{ab}{(a+b)^2(a+b+1)} = 0.000009$$

$$\frac{0.7b \cdot b}{(0.7b+b)^2(0.7b+b+1)}$$

$$\frac{0.7b^2}{(1.7b)^2(1.7b+1)} = \frac{0.7\cancel{b^2}}{2.89\cancel{b^2}(1.7b+1)}$$

$$= \frac{0.2424}{(1.7b+1)} = 0.000009$$

$$(1.7b + 1) = 26938.8$$

$$1.7b = 26937.8$$

$$\boxed{\begin{aligned} b &= 15845.8 \\ a &= 11102.8 \end{aligned}}$$

Part (ii)

$$a_1 = a + n$$

$$b_1 = b + n - s$$

$$\cancel{s = \frac{61}{150} = 0.4}$$

$$\cancel{a_1 = 11102.8 + 2}$$

$$a_1 = 11102.8 + 61$$

$$b_1 = 15845.8 + 150 - 61$$

$$a_1 = 11163.8$$

$$b_1 = 15934.8$$

Posterior : $B(a_1, b_1)$

20^{th} percentile = qbeta(0.2, a_1, b_1)

$$= 0.409$$

$$\boxed{= 40.9\%}$$

Question 2:

Q2) X → random variable

$$f(x) = \mu e^{-\mu x} \qquad x \in (0, \infty)$$
$$\mu > 0$$

μ = rate parameter

(avg rate of arrivals per unit time)

→ This follows exponential distribution with known rate.

Prior for μ:-

$\mu \sim$ Gamma distribution $G(\nu_0, \lambda_0)$

$\mu \sim G(1, 0.2)$

Data:- $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$

7 5 8 4 6

$$L(\underset{\sim}{x} \mid \mu) = \mu^{n} e^{-\sum_{i=1}^{n} \mu_i x} \qquad x \in (0, \infty)$$

$$= \mu^{5} e^{-\sum_{i=1}^{5} (\mu_i x)}$$

$$\boxed{L(\underset{\sim}{x} \mid \mu) = \mu^{n} e^{-\mu \sum_{i=1}^{n} x_i}}$$

Part (i)

Posterior for μ

$$\pi(\mu|x) \propto L(x|\mu)\,\pi(\mu) \qquad \mu>0$$

$$\pi(\mu|x) \propto \mu^5 e^{-\sum_{i=1}^{5}(\mu x_i)} \times \underbrace{\lambda_0^{\gamma_0}}_{\text{constant}} \times \mu^{\gamma_0-1} e^{-\lambda_0 \mu}$$

$$\propto \mu^{5+\gamma_0-1} \cdot e^{-\left[\lambda_0+\sum_{i=1}^{5}\mu_i x\right]} \qquad \mu>0$$

$$\pi(\mu|x) \equiv \text{Gamma}(\gamma_1, \lambda_1) \qquad n=5$$

$\gamma_1 \to n+\gamma_0$

$\lambda_1 \to \lambda_0 + \sum_{i=1}^{n} \mu_i x_i$

rate parameter

$$-\mu\left(\sum_{i=1}^{n} x_i - \lambda_0\right)$$

$$\gamma_1 = \gamma_0 + n$$

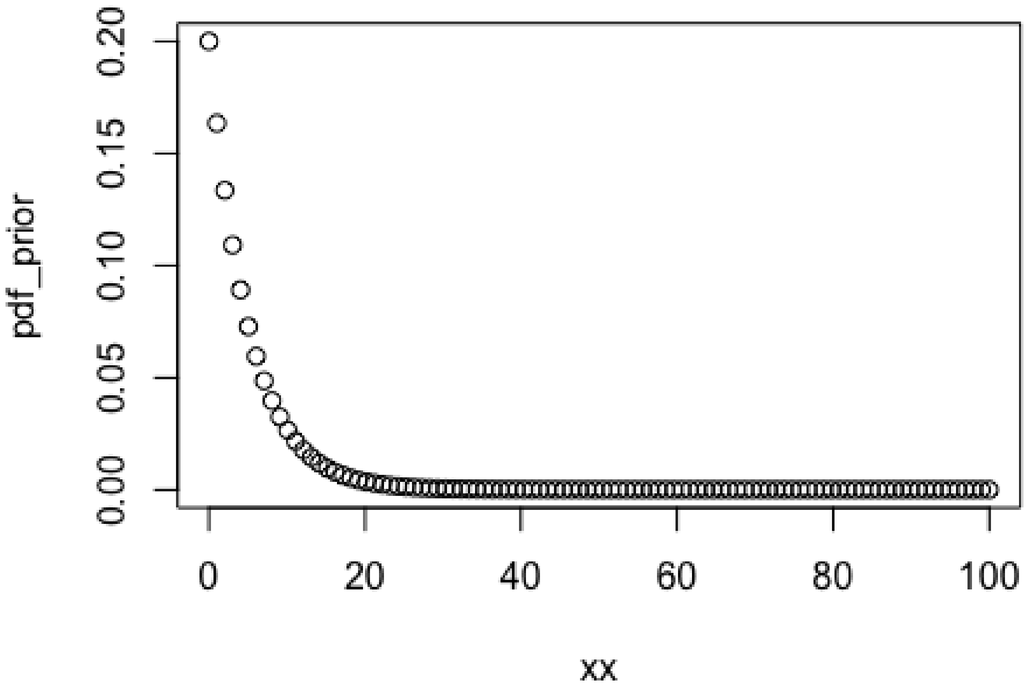
$$\lambda_1 = \lambda_0 + \sum_{i=1}^{n} x_i$$

R Code:

```
#Data
x <- c(7,5,8,4,6)
n<- length(x)

#Prior
nu0<- 1
lambda0<- 0.2

#Prior Distribution
xx<- seq(0,100,length=100)
pdf_prior<- dgamma(xx,shape=nu0,rate=lambda0)
plot(xx,pdf_prior)
```




```
#Posterior for mu
nu1<- nu0+n
nu1

## [1] 6

lambda1<- lambda0 + sum(x)
lambda1

## [1] 30.2

post_mean<- nu1/lambda1
post_mean

## [1] 0.1986755
```

Part (ii)

```r
# 95% CL
high<-qgamma(.975, shape=nu1, rate=lambda1)
low<-qgamma((1-.975),shape=nu1, rate=lambda1)
c(low,high)
```

```
## [1] 0.07291041 0.38636861
```

Part (iii)

```r
#Predictive distribution
N<-1000
rtheta<- rgamma(N, shape=nu1, rate=lambda1)
prob<- 1- pexp(q = 7, rate = rtheta)
mean(prob)
```

```
## [1] 0.2868927
```

Question 3:

(3) Sampling distribution is Poisson with sample size n,

Prior distribution $\rightarrow$ constant/uniform

Likelihood:

$$L(x|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Prior $P(\lambda) = \frac{1}{a}$ $\quad a = \text{constant}$.

Posterior:

$$P(\lambda|x) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \cdot \frac{1}{a}$$

$$\Rightarrow P(\lambda|x) \propto e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}$$

$$\sim \text{Gamma}\left(\gamma_1, \lambda_1\right)$$

$$\Rightarrow \quad \sum_{i=1}^{n} x_i \leftarrow \gamma_1$$

$$n \leftarrow \lambda_1$$

$$\text{posterior mean} = \frac{\gamma_1}{\lambda_1} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\boxed{\lambda \mid x = \text{Gamma}\left(\sum x_i , n\right)}$$

(iii) Posterior predictive distribution:-

$$P(x_{new} \mid x) = \int_0^\infty P(x_{new} \mid \lambda) . P(\lambda \mid x) . d\lambda$$

$$= \int_0^\infty \left[\frac{\lambda^{x_{new}} e^{-\lambda}}{(x_{new})!}\right] \cdot \left[\frac{(n)^{\sum x_i}}{\Gamma(\sum x_i)} \lambda^{\sum x_i - 1} e^{-n\lambda}\right] d\lambda$$

$$= \frac{(n)^{\Sigma x_i}}{\Gamma(\Sigma x_i)\Gamma(x_{new}+1)} \int_0^\infty \lambda^{x_{new}+\Sigma x_i} e^{-n\lambda} \, d\lambda$$

$$= \frac{n^{\Sigma x_i}}{\Gamma(\Sigma x_i)\Gamma(x_{new}+1)} \frac{\Gamma(x_{new}+\Sigma x_i)}{n^{x_{new}+\Sigma x_i}}$$

$$= \frac{\Gamma(x_{new}+\Sigma x_i)}{\Gamma(\Sigma x_i)\Gamma(x_{new}+1)} \left(\frac{n}{n+1}\right)^{\Sigma x_i} \left(\frac{1}{n+1}\right)^{x_{new}}$$

$\Rightarrow$ $\boxed{\text{Negative binomial with mean } \frac{\Sigma x_i}{n} \text{ and Variance } \frac{\Sigma x_i}{n^2}(n+1)}$

(ii) average = 200   10 weeks.

$\Rightarrow$ average $\times 10 = \Sigma x_i$

$\Rightarrow \Sigma x_i = 200 \times 10 = 2000$

$\Sigma x_i = 2000$

$n = 10$

$\boxed{\text{Gamma prior} \Rightarrow \text{Gamma}(2000, 10)}$

Predictive:-

N = 1000

rtheta - rgamma(N, 2000, 10)

Problem-3.R

```
#Data

x<- c(28, 15, 11, 21, 11, 17, 21, 15, 14, 18)
```

```r
n<- length(x)

#Prior is flat uniform distribution

#Posterior for mu

nu1<- sum(x)
nu1
```

```
## [1] 171
```

```r
lambda1<- n
lambda1
```

```
## [1] 10
```

```r
# 95% CL
high<-qgamma(.975, shape=nu1, rate=lambda1)
low<-qgamma((1-.975),shape=nu1, rate=lambda1)
c(low,high)
```

```
## [1] 14.63300 19.75635
```

Question 4:

Problem-4.R

```r
#prior
values<- seq(991,1010)
probability<- rep(.05,20)

#Data
x<-c(1003,995,999,1001,994,998,996,1002,1004,998,994,994,995,995,1001,997,1006,997,998
,994)
n<-length(x)
sigma<-sqrt(32)
sigma
```

```
## [1] 5.656854
```

Part(i)

```r
#Posterior distribution for mu
posterior<-c()
for(i in 1:20){
  mu<-values[i]
  likelihood<- prod(dnorm(x,mu,sigma))
  posterior<-c(posterior,likelihood*0.05)
}

posterior_prob<-posterior/sum(posterior)
result <- data.frame(Value=values, Proba=round(posterior_prob,3))
result
```

```
##    Value Proba
## 1    991 0.000
## 2    992 0.000
## 3    993 0.000
## 4    994 0.002
```

```
## 5    995 0.017
## 6    996 0.085
## 7    997 0.223
## 8    998 0.315
## 9    999 0.238
## 10  1000 0.096
## 11  1001 0.021
## 12  1002 0.002
## 13  1003 0.000
## 14  1004 0.000
## 15  1005 0.000
## 16  1006 0.000
## 17  1007 0.000
## 18  1008 0.000
## 19  1009 0.000
## 20  1010 0.000
```

```
#Posterior mean
posterior_mean<- sum(values*posterior_prob)
posterior_mean

## [1] 998.05

#Normal Prior for mu ~N(mu0, tau0^2)
mu0<-mean(values)
tau0<-sd(values)
xbar<-mean(x)
mu1<- (xbar * n/sigma^2 + mu0 * 1/tau0^2)/( n/sigma^2+ 1/tau0^2)
mu1

## [1] 998.1571

tau1<- sqrt(1/( n/sigma^2+ 1/tau0^2))
c(mean=mu1, sd=tau1)

##       mean         sd
## 998.157104   1.236954

#Credible Interval
c(low=qnorm(.025,mu1,tau1), high=qnorm((1-.025),mu1,tau1))

##       low      high
##  995.7327 1000.5815
```

Part (ii):

```
#P(mu<1000 under posterior distribution)
N<-5000
rmu<-rnorm(N, mu1, tau1)
rprob<- pnorm(1000, rmu, sigma)
mean(rprob)

## [1] 0.6249315
```

Part (iii)

$$\mu/\theta \sim N\left(\mu = 998, \frac{\sigma^2}{2}\right)$$

$$\sim N\left(\mu = 998, \frac{1}{2\theta}\right) \qquad \theta = \frac{1}{\sigma^2}$$

$$\theta \sim Gamma(\nu_0, \lambda_0)$$

$$\theta \sim Gamma(40, 10)$$

$$\pi(\theta) \propto \lambda_0^{\nu_0} \theta^{\nu_0 - 1} e^{-\lambda_0 \theta} \qquad \theta \in (0, \infty)$$

$$\nu_0 = 40 \quad \lambda_0 = 10$$

Posterior for θ

$$\pi(\theta|x) \propto L(x|\theta) \cdot \pi(\theta) \qquad \theta > 0$$

$$\pi(\theta|x) \propto \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{20} e^{-\frac{\sum_{i=1}^{20}(x_i-\mu)^2}{2\sigma^2}} \cdot \lambda_0^{\nu_0} \theta^{\nu_0-1} e^{-\lambda_0\theta} \qquad \theta > 0$$

$$\propto \frac{1}{\sqrt{2\pi}} \theta^{\frac{-20}{2}} e^{-\frac{\sum_{i=1}^{20}(x_i-\mu)^2}{2}\theta} \cdot \lambda_0^{\nu_0} \theta^{\nu_0-1} e^{-\lambda_0\theta} \qquad \theta > 0$$

$$\propto \theta^{-10+\nu_0-1} e^{-\left(\lambda_0 + \frac{\sum_{i=1}^{20}(x_i-\mu)^2}{2}\right)\theta} \qquad \theta > 0$$

$$\pi(\theta|x) \equiv Gamma(\nu_1, \lambda_1)$$

$$\equiv \frac{n}{2} + \nu_0, \quad \lambda_0 + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2} \qquad n = 20$$

```
#Posterior distribution for σ2
nu1<- n/2+40
nu1
```

```
## [1] 50
```

```
lambda1<-10+sum(((x-998)^2/2))
lambda1
```

```
## [1] 138.5
```

σ2 ~ *Gamma(50,138.5)*

```
posterior_sig<-c()
for(i in 1:20){
  mu<-values[i]
  likelihood<- prod(dgamma(x,shape = nu1,rate = lambda1))
  posterior_sig<-c(posterior_sig,likelihood*0.05)
}

posterior_prob1<-posterior_sig/sum(posterior_sig)
result1 <- data.frame(Value=values, Proba=round(posterior_prob1,3))
```