STAT_205_HW2

HONGLEI REN

1/20/2020

Problem 1

1.1 Propose a conjugate Normal Prior $\mu \sim N(\mu_0, \tau_0)$

According to the percentile table of standard normal distribution, we know that:

$$z = \frac{x - \mu_0}{\sqrt{\tau_0}} = \frac{42 - 40}{\sqrt{\tau_0}} = \frac{2}{\sqrt{\tau_0}} = 1.96$$

, and we got $\tau_0 = (\frac{2}{1.96})^2$. Therefore, the Normal prior is $\mu \sim N(40, (\frac{2}{1.96})^2)$.

1.2 The Posterior Expectation

Suppose $\tau_0 = \frac{\sigma^2}{m} = \frac{10^2}{m} = (\frac{2}{1.96})^2$, therefore, $m = 96.04$.

Then, we have $Y|\mu, \sigma^2 \sim Normal(\mu, 10^2), \mu \sim N(40, \frac{10^2}{96.04})$. The posterior is:

$$\mu|Y \sim N(w\overline{Y} + (1 - w)\mu_0, \frac{\sigma^2}{n + m})$$

, where $w = \frac{n}{n+m}$, $\overline{Y} = 45.283$ is the sample mean.

Therefore, the posterior mean is 40.586, which is very close to the prior mean although the sample mean 45.283 is much higher, and this is caused by a strong prior.

In our case, the relative contribution of prior is 0.89 and data for posterior expectation is 0.11.

1.3 The Posterior Distribution

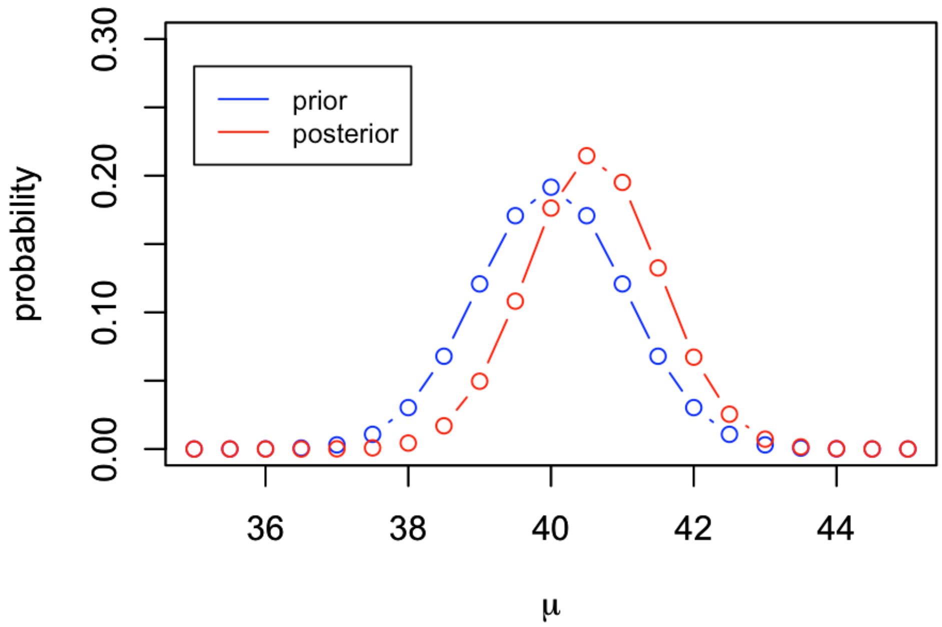
From the figure below, we can see that posterior distribution mu shifts towards right of the prior, which means the data suggest average snowfall observed are higher than our prior knowledge.

```
#Data
snowfall = c(38.6, 42.4, 57.5, 40.5, 51.7, 67.1, 33.4, 60.9, 64.1, 40.1, 40.7, 6.4)
mean_snowfall = mean(snowfall)

#Params
n = length(snowfall)
m = 96.04
mu0 = 40
sigma = 10
w = n / (n + m)
mu = seq(35, 45, 0.5)
```

```
# Calculate prior and posterior
prior_mu = dnorm(mu, mean = mu0, sd= (sigma^2) / m)
posterior_mean = w * mean_snowfall + (1 - w) * mu0
posterior_sd = (sigma^2) / (n + m)
posterior_mu = dnorm(mu, mean = posterior_mean, sd= posterior_sd)

# Plotting
plot(mu, prior_mu/sum(prior_mu),col="blue", type="b", xlab = expression(mu), ylab =
"probability", ylim=range(c(0,0.3)))
par(new=TRUE)
plot(mu, posterior_mu/sum(posterior_mu), col="red", type="b", xlab = expression(mu)
, ylab = "probability", ylim=range(c(0,0.3)))
legend(35, 0.28, legend=c("prior", "posterior"),
       col=c("blue", "red"), lty=1:1, cex=0.8)
```



1.4 Quantiles of Posterior Distribution

```
S = 100000 #Sample size
mu_star = rnorm(S, posterior_mean, posterior_sd)
pct_10 = quantile(mu_star, 0.1)
pct_10
```

```
##      10%
## 39.39421
```

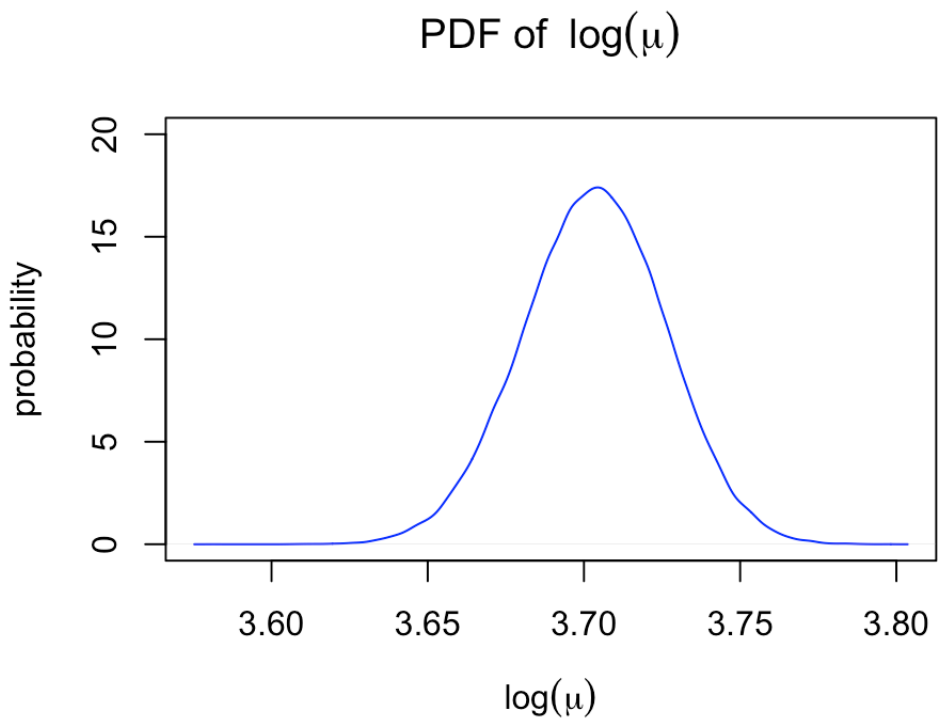
```
pct_90 = quantile(mu_star, 0.9)
pct_90
```

```
##      90%
## 41.77048
```

The 10% quantiles of the posterior distribution is 39.3942072 and the and 90% percentile is 41.7704777.

1.5 Posterior mean of log(μ)

```
S = 100000 #Sample size
mu_star = rnorm(S, posterior_mean, posterior_sd)
log_mu = log(mu_star)
mean_log_mu = mean(log_mu)
# Plotting
plot(density(log_mu),col="blue", type="l", xlab = expression(log(mu)), ylab = "prob
ability", ylim=range(c(0, 20.0)), main=expression("PDF of " ~ log(mu)))
legend(3.5, 0.28, legend=c("log(mu)"),
       col=c("blue"), lty=1:1, cex=0.8)
```




```
mean_log_mu
```

```
## [1] 3.703151
```

I first draw 10000 samples from the posterior distribution of μ, and then perform a **log** transformation. Finally, I calculate the mean from the log transformed samples of μ, and the posterior mean of $log(\mu)$ =3.703151.

Problem 2

2.1 Derive posterior

The number of ASE for N patient follows $Y \sim Poisson(N\lambda)$ and $\lambda \sim Gamma(a, b)$.

The likelihood function is :

$$f(Y|\lambda) \propto e^{-N\lambda}\lambda^{Y}$$

, where $N = 50, Y = 12 + 6 * 2 + 2 * 10 = 44$.

The prior function is:

$$\pi(\lambda) \propto e^{-b\lambda}\lambda^{a}$$

Therefore, the posterior is:

$$p(\lambda|Y) \propto e^{-N\lambda}\lambda^{Y}e^{-b\lambda}\lambda^{a} \propto e^{-B\lambda} * \lambda^{A}$$

, where $A = a + Y, B = b + N$.

$$p(\lambda|Y) = Gamma(A, B)$$

.

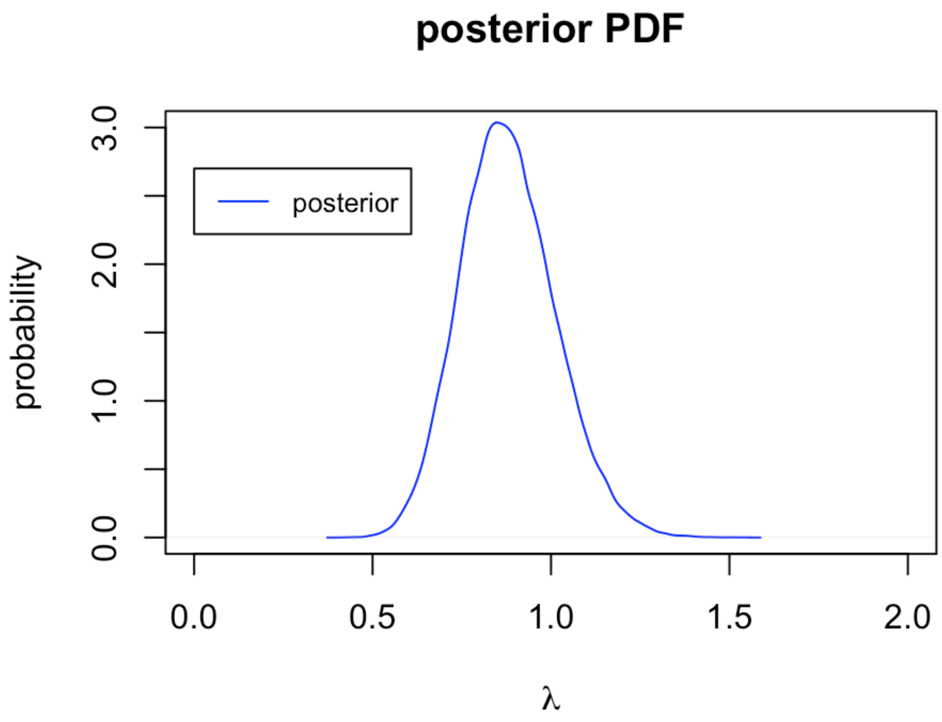
2.2 Posterior distribution

Given $a = b = 0.01$, the posterior is

$$p(\lambda|Y) = Gamma(44.01, 50.01)$$

.

```
N = 50
Y = 44
a = b = 0.01
A = Y + a
B = N + b
S = 100000
posterior_lambda = rgamma(S, shape = A, rate=B)
posterior_mean = mean(posterior_lambda)
# Plotting
plot(density(posterior_lambda), col="blue", type="l", xlab = expression(lambda), yl
ab = "probability", main = "posterior PDF", xlim = range(c(0,2)), ylim=range(c(0,3)
))
legend(0, 2.7, legend=c( "posterior"),
       col=c("blue"), lty=1:1, cex=0.8)
```

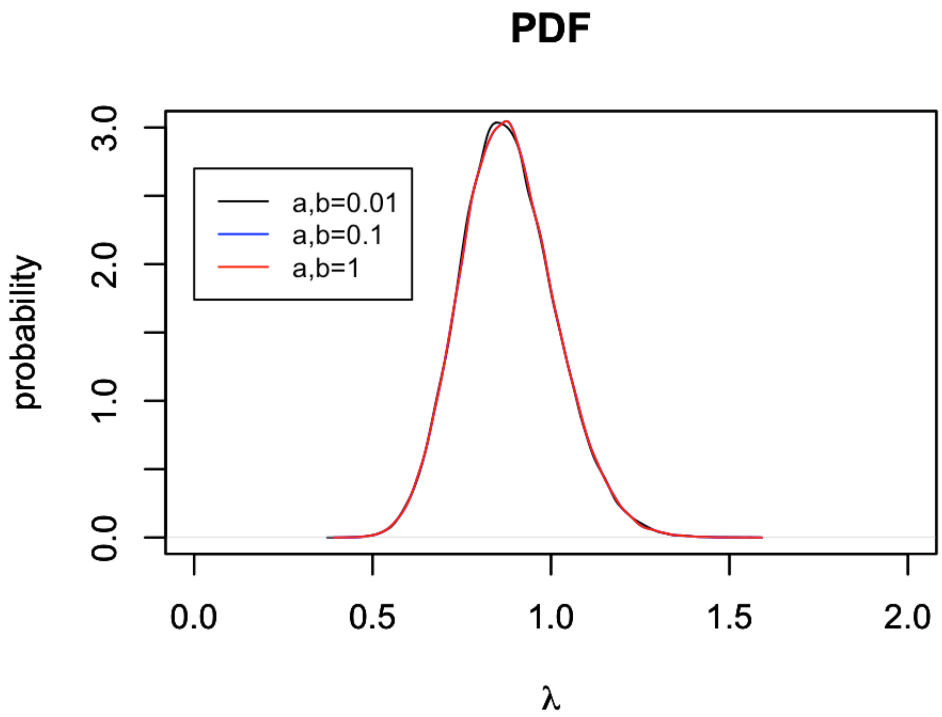


Posterior Mean is 0.880264, and the 95% credible interval is (0.6393088, 1.1586223).

2.3 Sensitivity Analysis

The distribution with differnt priors are given below, we can hardly see the difference in their posterior distribution, which means our posterior is pretty robust to the prior parameters.

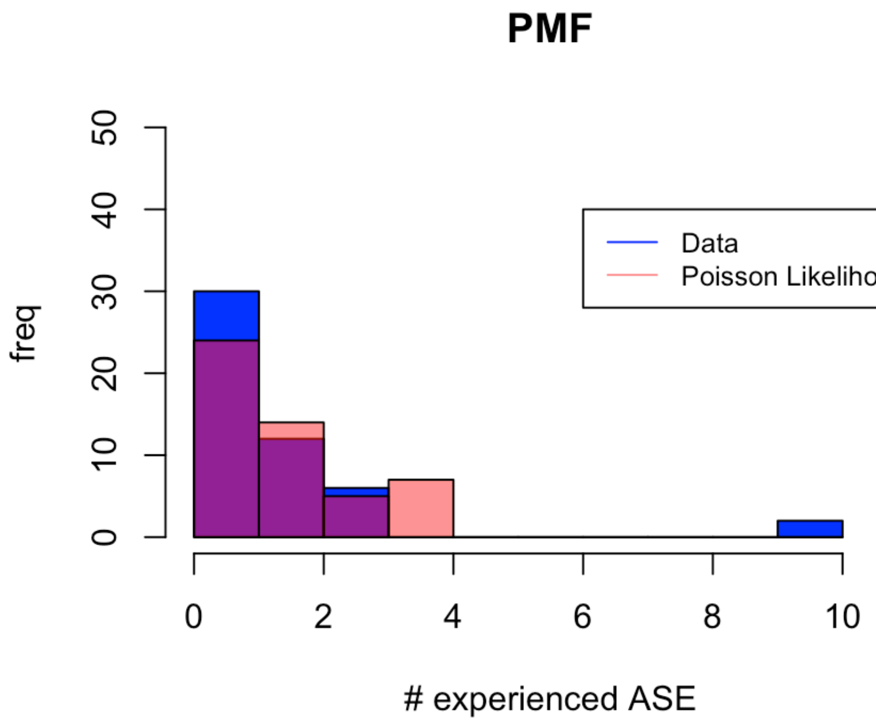
```
posterior_lambda_2 = rgamma(S, shape = 44.1, rate=50.1)
posterior_lambda_3 = rgamma(S, shape = 45, rate=51)
# Plotting
plot(density(posterior_lambda), col="black", type="l", xlab = expression(lambda), y
lab = "probability", main = "PDF", xlim = range(c(0,2)), ylim=range(c(0,3)))
par(new=T)
plot(density(posterior_lambda_2),col="blue", type="l", xlim = range(c(0,2)),ylim=ra
nge(c(0,3)), xlab = expression(lambda), ylab = "probability", main = "PDF")
par(new=T)
plot(density(posterior_lambda_2), col="red", type="l", xlab = expression(lambda), y
lab = "probability", main = "PDF", xlim = range(c(0,2)), ylim=range(c(0,3)))
legend(0, 2.7, legend=c("a,b=0.01", "a,b=0.1", "a,b=1"),
       col=c("black", "blue", "red"), lty=1:1, cex=0.8)
```



2.4 Compare Fitted Poisson with data

According to the fitted results below, we can see the fitting is fairly good, but the Poisson likelihood can not capture two extreme data points at # experienced $ASE = 10$, and the frequency at lower ASE value in likelihood are higher than data.

```
data = c(rep(0, 30), rep(1, 12), rep(2, 6), rep(10, 2))
result = rpois(N, lambda = posterior_mean)
hist(data, col="blue", xlim = range(c(0, 11)),ylim=range(c(0, N)),breaks = 0:10,rig
ht=F, xlab = "# experienced ASE", ylab = "freq", main = "PMF")
hist(result, xlim = range(c(0, 11)),ylim=range(c(0, N)),breaks = 0:10,right=F, xlab
= "# experienced ASE", ylab = "freq", main = "PMF", col=rgb(1,0,0,0.5), add=T)
legend(6, 40, legend=c("Data", "Poisson Likelihood"),
       col=c("blue", rgb(1,0,0,0.5)), lty=1:1, cex=0.8)
```



2.5 Probabilty that new medication has higher side effect rate than the previous one

```
p1 = sum(posterior_lambda > 1, na.rm = T) / S
p2 = sum(posterior_lambda_2 > 1, na.rm = T) / S
p3 = sum(posterior_lambda_3 > 1, na.rm = T) / S
```

The probabilty that new medication has higher side effect rate than the previous one (p) is :

For $a = b = 0.01$, p = 0.18065

For $a = b = 0.1$, p = 0.18179

For $a = b = 1$, p = 0.18317

It can be seen, the probability is not sensitive to prior params.

Problem 3 Wildfires

Because $E(X) = \frac{a}{b}, Var(X) = \frac{a}{b^2}$, therefore, by setting our goal $E(X) = \frac{a}{b} = 75, Var(X) = \frac{a}{b^2} = 100.$ We get: $a = 56.25, b = 0.75$

Therefore, the prior can be set to $Gamma(a = 56.25, b = 0.75)$.