Assignment_1

Yash Gangrade

February 8, 2018

Written Part

Question 1

a)

Solution: We have,

$$p(x) = h(x)exp(\eta \cdot T(x) - A(\eta))$$

Now, we will use the general identity here:

$$\int p(x)dx = 1$$

$$\int h(x)exp(\eta \cdot T(x) - A(\eta))dx = 1$$

$$\int h(x)exp(\eta \cdot T(x)) = exp(A(\eta))$$

Now, we will do the partial derivative of the above function w.r.t η, so we have

$$\frac{\partial \int h(x)exp(\eta \cdot T(x))dx}{\partial \eta} = \frac{\partial exp(A(\eta))}{\partial \eta}$$

$$\int h(x)\frac{\partial exp(\eta \cdot T(x))}{\partial \eta}dx = \frac{\partial exp(A(\eta))}{\partial \eta}$$

$$\int h(x)T(x)exp(\eta \cdot T(x))dx = exp(A(\eta))\frac{\partial A(\eta)}{\partial \eta}$$

$$\frac{\int h(x)T(x)exp(\eta \cdot T(x))dx}{exp(A(\eta))} = \frac{\partial A(\eta)}{\partial \eta}$$

$$\int h(x)T(x)exp(\eta \cdot T(x) - A(\eta))dx = \frac{\partial A(\eta)}{\partial \eta}$$

$$\int T(x)h(x)exp(\eta \cdot T(x) - A(\eta))dx = \frac{\partial A(\eta)}{\partial \eta}$$

From initial equation and the knowledge of expectations i.e. $E[g(x)|y] = \int g(x)p(x|y)dx$. So, here we have,

$$\int T(x)h(x)exp(\eta \cdot T(x) - A(\eta))dx = \frac{\partial A(\eta)}{\partial \eta}$$

$$\int T(x)p(x|\eta)dx = \frac{\partial A(\eta)}{\partial \eta}$$

$$E[T(x)|\eta] = \nabla A(\eta) = \left(\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, \ldots, \frac{\partial A}{\partial \eta_d}\right)$$

Hence proved.

b)

Solution: Now, we need to verify the above result. Here, we are given a Gaussian with known variance but unknown mean so we have,

$$X|\mu \sim N(\mu, \sigma^2)$$

So, we have the conditional probability distribution and we need to write it in form of an exponential family first.

$$p(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

We have after expanding,

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2\sigma^2}(x^2 + \mu^2 - 2x\mu)\right)$$

$$p(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2\sigma^2}\right) exp\left(-\frac{\mu^2}{2\sigma^2} + \frac{x\mu}{\sigma^2}\right)$$

Comparing the above equation with the standard exponential family function $p(x) = h(x)exp(\eta \cdot T(x) - A(\eta))$, we have,

$$\eta = \mu$$
$$h(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2\sigma^2}\right)$$
$$T(x) = \frac{x}{\sigma^2}$$
$$A(\mu) = \frac{\mu^2}{2\sigma^2}$$

So, now we have,

$$E[T(x|\eta)] = \frac{E[x]}{\sigma^2} = \frac{\mu}{\sigma^2} \tag{1}$$

And, we have,

$$\frac{\partial A}{\partial \mu} = \frac{2\mu}{2\sigma^2} = \frac{\mu}{\sigma^2} \tag{2}$$

So, from the above two equations, we have verified the result in the first part.

Question 2

a)

Solution:

We have, $X \sim Pois(\lambda)$ and the pmf as:

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Now, we need to get the exponential family form so, we do,

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$
$$P(X = x; \lambda) = \frac{e^{x \ln \lambda} e^{-\lambda}}{x!}$$
$$P(X = x; \lambda) = \frac{exp(x \ln \lambda - \lambda)}{x!}$$

Comparing with the general form of exponential family form we have,

$$\eta(\lambda) = \ln \lambda \qquad \textit{Natural Parameter}$$
$$h(x) = \frac{1}{x!}$$
$$T(x) = x \qquad \textit{Sufficient Statistic}$$
$$A(\lambda) = \lambda$$

b) and c)

Solution: Two non-informative priors for this distribution are: \ **Jeffrey's Prior** We have the pmf of the gaussian as,

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Now, we will find the fischer's information i.e. $I(\lambda)$ for the jeffrey's prior as follows:

$$I(\lambda) = E\left[-\frac{\partial^2 \ln p(x|\lambda)}{\partial \lambda^2}\right]$$
$$I(\lambda) = E\left[-\frac{\partial^2}{\partial \lambda^2}(x \ln \lambda - \lambda - \ln x!)\right]$$
$$I(\lambda) = E\left[-\frac{\partial}{\partial \lambda}\left(\frac{x}{\lambda} - 1\right)\right]$$
$$I(\lambda) = E\left[\frac{x}{\lambda^2}\right]$$

We know that for a poisson, $E[x] = \lambda$

$$I(\lambda) = \frac{1}{\lambda}$$

Now, we know that jeffrey's prior is just dependent on the fischer's information and it is calculated as follows:

$$P(\lambda) = \sqrt{I(\lambda)}$$
$$P(\lambda) = \sqrt{\frac{1}{\lambda}} \qquad \forall x \in (0, \infty)$$

3

Following the definition from one of the blogs online (http://lesswrong.com/lw/6uk/against/_improper/_priors/), this prior is an improper prior. Now, we need to calculate the posterior distribution for this prior. The calculations are as follows:

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{p(\lambda)p(X|\lambda)}{\sum p(\lambda)p(X|\lambda)}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{\frac{1}{\sqrt{\lambda}}\prod_{i=1}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}}{\sum_\lambda \prod_{i=1}\frac{1}{\sqrt{\lambda}}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{\frac{\lambda^{\sum x_i - \frac{1}{2}} \cdot e^{-n\lambda}}{x_1!x_2!....x_n!}}{\sum_\lambda \frac{\lambda^{\sum x_i - \frac{1}{2}} \cdot e^{-n\lambda}}{x_1!x_2!....x_n!}}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{\lambda^{\sum x_i + \frac{1}{2} - 1} \cdot e^{-n\lambda}}{\sum_\lambda \lambda^{\sum x_i + \frac{1}{2} - 1} \cdot e^{-n\lambda}}$$

The formula above looks similar to the pdf of a Gamma Distribution \ (https://en.wikipedia.org/wiki/Gamma/_distribution), thus we can define proportionality to it with the factors:

$$p(\lambda|x_1, x_2, ..., x_n) \propto Gamma\left(\frac{1}{2} + \sum x_i, n\right)$$

So one interesting thing to observe here is that even though our prior was improper, the posterior is a proper distribution.

Uniform Prior

We have the pmf of the gaussian as,

$$P(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Consider a uniform prior as it is the max entropy non-informative prior as the prior. Therefore, we have,

Following the procedure similar to jeffrey's prior, we will do the calculation as;

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{p(\lambda)p(X|\lambda)}{\sum p(\lambda)p(X|\lambda)}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{1 \cdot \prod_{i=1}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}}{\sum_\lambda \prod_{i=1}\frac{1}{\sqrt{\lambda}}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{\frac{\lambda^{\sum x_i} \cdot e^{-n\lambda}}{x_1!x_2!....x_n!}}{\sum_\lambda \frac{\lambda^{\sum x_i} \cdot e^{-n\lambda}}{x_1!x_2!....x_n!}}$$

$$p(\lambda|x_1, x_2, ..., x_n) = \frac{\lambda^{\sum x_i + 1 - 1} \cdot e^{-n\lambda}}{\sum_\lambda \lambda^{\sum x_i + 1 - 1} \cdot e^{-n\lambda}}$$

The formula above looks similar to the pdf of a Gamma Distribution \ (https://en.wikipedia.org/wiki/Gamma/_distribution), thus we can define proportionality to it with the factors:

$$p(\lambda|x_1, x_2, ..., x_n) \propto Gamma\left(1 + \sum x_i, n\right)$$

4

So one interesting thing to observe here is that even though our prior was improper, the posterior is a proper distribution.

Question 3

Solution: Given, we have that the X_i is from a Gaussian distribution with a known variance σ^2 and an unknown μ with a uniform prior. Essentially, we have,

$$\mu \sim Unif(a, b)$$

$$p(\mu) = \frac{1}{b-a} \qquad \forall \mu \in [a, b]$$

Similarly, we have,

$$X_i \sim N(\mu, \sigma^2)$$

$$p(x_i|\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

For all the n variables

$$p(x_1, x_2, ..., x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$p(x_1, x_2, ..., x_n|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right)$$

Now, we need to find the posterior pdf, we have,

$$p(\mu|x_1, x_2, ..., x_n; \sigma^2, a, b) = \frac{p(x_1, x_2, ..., x_n|\mu; \sigma^2, a, b) \cdot p(\mu)}{\int_a^b p(x_1, x_2, ..., x_n|\mu; \sigma^2, a, b) \cdot p(\mu) d\mu}$$

$$p(\mu|x_1, x_2, ..., x_n; \sigma^2, a, b) = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right) \cdot \frac{1}{b-a}}{\int_a^b \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right) \cdot \frac{1}{b-a} d\mu}$$

$$p(\mu|x_1, x_2, ..., x_n; \sigma^2, a, b) = \frac{exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right)}{\int_a^b exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right) d\mu}$$

Thus, we can see that the posterior pdf is also a Gaussian Distribution.

R Coding

Question 4

a)

Solution: For the given distributions we have the joint posterior density function as:

$$p(\mu_j, \sigma_j^2 | y_{i,j}) \sim N - IG(\mu_n, n_n, \alpha_n, \beta_n)$$

Here, for each j in range from 1 to 3, we have the number of elements as $n_1, ... n_j$, we have for a particular (i, j) pair,

$$
\begin{aligned}
\bar{y} &= \frac{\sum_{i=1}^{n} y_i}{n} \\
\mu_n &= \frac{\mu_0 n_0 + n\bar{y}}{n_0 + n} \\
n_n &= n_0 + n \\
\alpha_n &= \alpha_0 + \frac{n}{2} \\
\beta_n &= \beta_0 + \frac{\sum(y_i - \bar{y})^2}{2} + \frac{nn_0}{n_0 + n}\frac{(y_i - \mu_0)^2}{2}
\end{aligned}
$$

And, the marginal posterior is:

$$p(\sigma^2 | y_{i,j}) \sim IG(\alpha_n, \beta_n)$$

Now, using the information, we have sampled from the distribution and then plot a histogram for the posterior with pdf.

```r
# We have the given prior information as (mu_j, sigma_j^2) ~ N-IG(mu_0, n_0, alpha_0, beta_0)
mu_0 = 0
n_0 = 0.000001
alpha_0 = 0.000001
beta_0 = 0.000001


# Reading and pre-processing the data

f1 = read.csv("oasis_cross-sectional.csv")
f2 = read.csv("oasis_hippocampus.csv")
f = merge(f1,f2, by = 'ID')

f$group[f$CDR >= 1.0] = "Dementia"
f$group[f$CDR == 0.5] = "Mild"
f$group[f$CDR == 0.0] = "Control"
f$group = factor(f$group, levels = c("Dementia", "Mild", "Control"))

# Right Hippocampal volume for each of the three groups
dementia_y3 = f$RightHippoVol[f$group == "Dementia"]
mild_y2 = f$RightHippoVol[f$group == "Mild"]
control_y1 = f$RightHippoVol[f$group == "Control"]

#Removing NA entries from the three arrays
control_y1 = control_y1[!is.na(control_y1)]
```

```r
mild_y2 = mild_y2[!is.na(mild_y2)]
dementia_y3 = dementia_y3[!is.na(dementia_y3)]

# We can calculate sample mean, variances, and lengths of the data for each of the three groups
mu_y3 = mean(dementia_y3, na.rm = TRUE)
mu_y2 = mean(mild_y2, na.rm = TRUE)
mu_y1 = mean(control_y1, na.rm = TRUE)

var_y3 = mean((dementia_y3 - mu_y3)^2, na.rm= TRUE)
var_y2 = mean((mild_y2 - mu_y2)^2, na.rm= TRUE)
var_y1 = mean((control_y1 - mu_y1)^2, na.rm = TRUE)

n_y3 = length(dementia_y3)
n_y2 = length(mild_y2)
n_y1 = length(control_y1)

#Function definitions

# function to get n random samples from a inverse gamma distribution
rinvgamma <- function(num, alpha, beta){
  inv = rgamma(n = num, shape = alpha, rate = beta)
  inv = 1/inv
  return(inv)
}
# Inverse Gamma PDF
dinvgamma <- function(x, alpha, beta){
  a = exp(alpha*log(beta) - log(gamma(alpha))) * x^(-alpha - 1) * exp(-beta / x)
  return(a)
}

# Parameters for the posterior N-IG as discussed in class
mun_y1 = (n_0 * mu_0 + n_y1*mu_y1)/(n_0 + n_y1)
nn_y1 = n_0 + n_y1
alphan_y1 = alpha_0 + n_y1/2
betan_y1 = beta_0 + (var_y1 * n_y1 /2) + ((n_y1*n_0 / (n_0 + n_y1))*((mu_y1 - mu_0)^2/2))

mun_y2 = ((n_0 * mu_0) + n_y2*mu_y2)/(n_0 + n_y2)
nn_y2 = n_0 + n_y2
alphan_y2 = alpha_0 + n_y2/2
betan_y2 = beta_0 + (var_y2 * n_y2 /2) + ((n_y2*n_0 / (n_0 + n_y2))*((mu_y2 - mu_0)^2/2))

mun_y3 = ((n_0 * mu_0) + n_y3*mu_y3)/(n_0 + n_y3)
nn_y3 = n_0 + n_y3
alphan_y3 = alpha_0 + n_y3/2
betan_y3 = beta_0 + (var_y3 * n_y3 /2) + ((n_y3*n_0 / (n_0 + n_y3))*((mu_y3 - mu_0)^2/2))


#Part a of Question 4
n = 1000000
#x = seq(10^5, 7*10^5,10)
# We need to plot for only the control group
sig = rinvgamma(n, alphan_y1, betan_y1)
p_sig = dinvgamma(sig, alphan_y1, betan_y1)
```
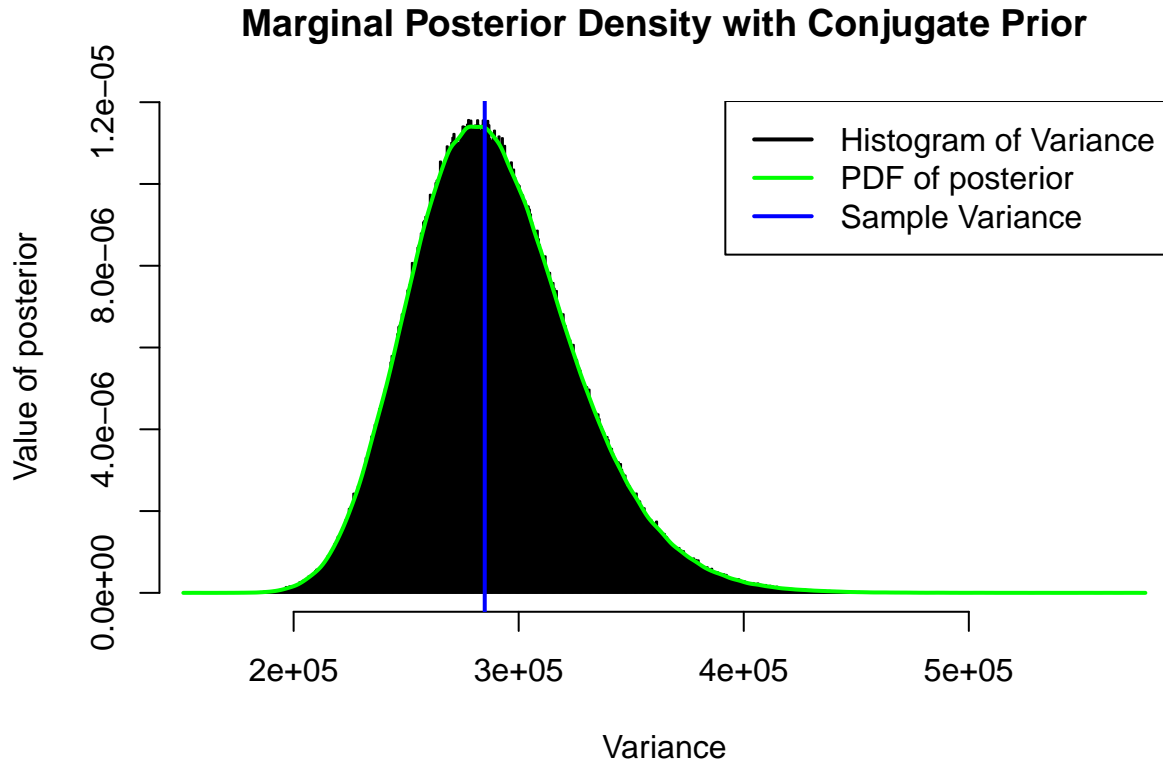
```
h = hist(sig,main = "Marginal Posterior Density with Conjugate Prior", xlab = "Variance", ylab = "Value
lines(density(sig), col = "green", lwd=2)
abline(v= var_y1, col = "blue", lwd = 2)
legend('topright', c("Histogram of Variance", "PDF of posterior", "Sample Variance"), col = c("black",
```

Marginal Posterior Density with Conjugate Prior

b)

Solution: Here, we want to find the marginal posterior distrubution $p(\mu_j|y_{ij})$. Next, we want to plot this for each of the groups as well. Finally, we will draw the sample means of each group as a vertical line. After looking for the marginal distribution of N-IG from Wikipedia, we found that it can be written as a t-distribution with a certain shift and scaling parameter. We will be using dt function in R as the student-t disctribution. Essentially, we have

$$p(\mu_j|y_{ij}) \sim t_{2\alpha_n}(\mu_j \cdot scale + shift)$$

Here,

$$Scale = \sqrt{\frac{\beta_n}{\alpha_n n_n}} \qquad Shift = \mu_n$$

It can also be written as

$$T_{2\alpha}(\mu_p, \frac{\beta_p}{\alpha_p n_p})$$

The code is as follows:

```
n = 1000000

#For the control dataset, the shift, scale and degree of freedom parameters are
scale1 = sqrt(alphan_y1 * nn_y1/betan_y1)
shift1 = mun_y1
```

```r
df = 2*alphan_y1

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)
# Pdf of y from dt
pdf_y = dt(y, df, ncp = shift1)
pdf_y = pdf_y/scale1
plot(y,pdf_y, col ="green", xlim = c(2000, 6000), ylim=c(0,0.12),lwd=3, main = "Marginal Posterior over
abline(v=mu_y1, col = "blue", lty=5)

# For the Mild Dataset, we follow the same process
shift2 = mun_y2
scale2 = sqrt(alphan_y2 * nn_y2/betan_y2)
df = 2*alphan_y2

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)

pdf_y = dt(y, df, ncp = shift2)
pdf_y = pdf_y/scale2
lines(y, pdf_y, col="yellow", lwd=3)
abline(v=mu_y2, col = "red", lty=5)

# For the dementia Dataset, we follow the same process again
shift3 = mun_y3
scale3 = sqrt(alphan_y3 * nn_y3/betan_y3)
df = 2*alphan_y3

#Now, we will use the in-built t-distribution random number generator
y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)

pdf_y = dt(y, df, ncp = shift3)
pdf_y = pdf_y/scale3
lines(y, pdf_y, col="black", lwd=3)
abline(v=mu_y3, col = "grey", lty=5)
legend('topright', c("Control", "Sample mean - control", "Mild", "Sample mean - mild","Dementia", "Sampl
```
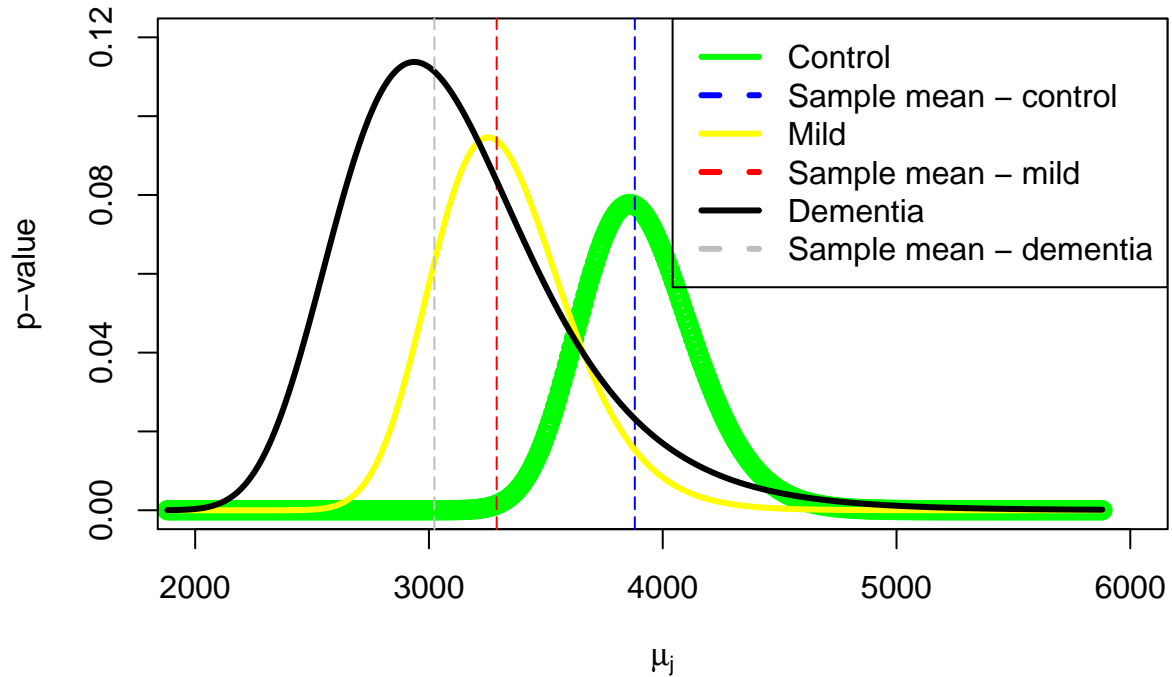
Marginal Posterior over mean using conjugate prior

c)

Solution: Here, we have the random variable $d_{12} = \mu_1 - \mu_2$. Now, we need to find the conditional density function i.e. $p(d_{12}|\sigma_1^2, \sigma_2^2, y_{ij})$ So, we know that the difference between two gaussians is also a gaussian with a different mean and variance, the process is as follows:

$$p(d_{12}|\sigma_1^2, \sigma_2^2, y_{ij}) \sim N(\mu_1, \sigma_1^2) - N(\mu_2, \sigma_2^2)$$

$$p(d_{12}|\sigma_1^2, \sigma_2^2, y_{ij}) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

So, now we will code it for these values, the process is as follows:

```
n = 1000000
#Calculate the parameters for the new distribution

#Datesets 1 and 2
new_mean = mun_y1 - mun_y2
var1 = rinvgamma(n, alphan_y1, betan_y1)
var2 = rinvgamma(n, alphan_y2, betan_y2)
new_sigma = sqrt((var1/nn_y1)+(var2/nn_y2))
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req1 = sum(sampled_values < 0)/length(sampled_values)
#hist(sampled_values)
#Datasets 1 and 3
new_mean = mun_y1 - mun_y3
```

```r
var1 = rinvgamma(n, alphan_y1, betan_y1)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var1/nn_y1)+ (var3/nn_y3))

sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req2 = sum(sampled_values < 0)/length(sampled_values)

#Datasets 2 and 3
new_mean = mun_y2 - mun_y3
var2 = rinvgamma(n, alphan_y2, betan_y2)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var2/nn_y2)+ (var3/nn_y3))
#new_sigma = sqrt(var2+ var3)
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req3 = sum(sampled_values < 0)/length(sampled_values)
#hist(sampled_values)
#print the values out
p_req1
```

```
## [1] 0
```

```r
p_req2
```

```
## [1] 0
```

```r
p_req3
```

```
## [1] 0.031965
```

So, we have the results written above as the output from the R code. The values are $p(d_{12} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$, $p(d_{13} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$, and $p(d_{23} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$. We can rewrite these probabilities i.e $p(d_{12} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$ as $p(\mu_1 < \mu_2|\sigma_1^2, \sigma_2^2, y_{ij})$. Similarly, $p(d_{13} < 0|\sigma_1^2, \sigma_3^2, y_{ij})$ as $p(\mu_1 < \mu_3|\sigma_1^2, \sigma_3^2, y_{ij})$ and $p(d_{23} < 0|\sigma_2^2, \sigma_3^2, y_{ij})$ as $p(\mu_2 < \mu_3|\sigma_2^2, \sigma_3^2, y_{ij})$. For the d_{12} and d_{13} case, we are getting infinitesimal near to zero probability, this can be verified from the plots in 4b where we can see that the overlapped region is very small for the (Control, Mild) and (Control, Dementia) pairs. But for the (Mild, Dementia) pair, we have much more signifcant overlap and thus the probability is higher.

d)

Solution: Here, we just need to perform t-test on the three combinations. The tests are written as follows:

```r
t1 = t.test(control_y1, mild_y2, alternative='greater')
t2 = t.test(control_y1, dementia_y3, alternative = 'greater')
t3 = t.test(mild_y2, dementia_y3, alternative = 'greater')
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  control_y1 and mild_y2
## t = 7.2205, df = 131.41, p-value = 1.868e-11
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  455.3136      Inf
## sample estimates:
## mean of x mean of y
##  3880.606  3289.735
```

t2

```
##
##  Welch Two Sample t-test
##
## data:  control_y1 and dementia_y3
## t = 6.4866, df = 31.197, p-value = 1.504e-07
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  633.216      Inf
## sample estimates:
## mean of x mean of y
##  3880.606  3023.360
```

t3

```
##
##  Welch Two Sample t-test
##
## data:  mild_y2 and dementia_y3
## t = 1.8925, df = 39.069, p-value = 0.03293
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  29.23099      Inf
## sample estimates:
## mean of x mean of y
##  3289.735  3023.360
```

The results of the t-test are attached above. In terms of statistics, the probability of finding the observed, or more extreme results when the null hypothesis (H_0) for a study in question is true (ref. statsdirect.com). In the context of our problem, p-value can be considered as the probability that the difference of two means is negative. It is the also considered as the frequentist or likelihood approach to estimate the difference. Now, as far as the result goes, the probability obtained in c part matches with the p-values obtained through t tests. In other words, it is a frequentist approach verification of our bayesian analysis.

Question 5 - Jeffrey's Prior

(a)

We have the joint posterior density function for Jeffrey Priors as (ref. Baysian Data Analysis by Gelman pp.65)

$$p(\mu_j, \sigma_j^2 | y_{i,j}) = \sigma_j^{-n-2} exp\left(-\frac{(n-1)s^2 + n(\bar{y} - \mu_j)^2}{2\sigma_j^2}\right)$$

For the formula above, we have,

$$\text{Sample Mean } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\text{Sample Variance } s^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Also, we have the marginal posterior (saw online) as:

$$p(\sigma_j^2|y_{i,j}) = IG\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

We have to follow all the steps from 4th question again, the (a) part of the above question is coded as follows. We will reuse some of the variables like sample mean, variance from our previous calculations.

```
# We will calculate the new alphas and betas here
alphan_y1 = (n_y1 - 1)/2
betan_y1 = var_y1*(n_y1-1)/2

alphan_y2 = (n_y2-1)/2
betan_y2 = var_y2*(n_y2-1)/2

alphan_y3 = (n_y3 - 1)/2
betan_y3 = var_y3*(n_y3-1)/2

#Part a of Question 4
n = 1000000
#x = seq(10^5, 7*10^5,10)
# We need to plot for only the control group
sig = rinvgamma(n, alphan_y1, betan_y1)
p_sig = dinvgamma(sig, alphan_y1, betan_y1)

h = hist(sig,main = "Marginal Posterior Density function with Conjugate Prior", xlab = "Variance", ylab
lines(density(sig), col = "green", lwd=2)
abline(v= var_y1, col = "blue", lwd = 2)
legend('topright', c("Histogram of Variance", "PDF of posterior", "Sample Variance"), col = c("black",
```

Marginal Posterior Density function with Conjugate Prior
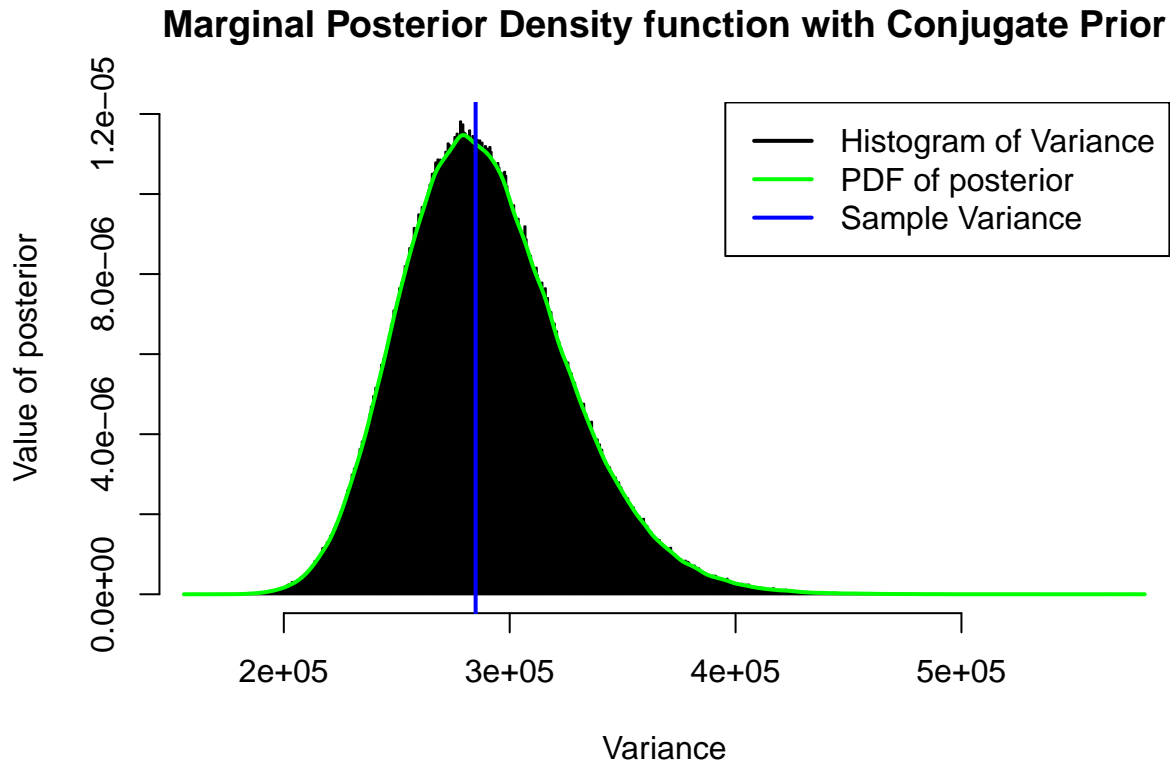
From the plot of marginal posterior density function above, we can see that it looks similar to the plot in question 4a of conjugate prior.

b)

Now, repeating the b part from the question 4. Again from the same reference as before, the marginal posterior for the jeffrey prior are:

$$p(\mu_j|y_{i,j}) \sim t_{n-1}(scale * \mu_j + shift)$$

Here, the scale and shift are

$$Scale = \frac{s}{\sqrt{n-1}} \qquad Shift = \bar{y}$$

It can also be written as $T_{n-1}(\bar{y}, \frac{s^2}{n})$. Code for b part is as follows:

```
n = 1000000

#For the control dataset, the shift, scale and degree of freedom parameters are
scale1 = sqrt((nn_y1-1)/var_y1)
shift1 = mu_y1
df = nn_y1 - 1

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)
# Pdf of y from dt
pdf_y = dt(y, df, ncp = shift1)
pdf_y = pdf_y/scale1
plot(y,pdf_y, col ="green", xlim = c(2000, 6000), ylim=c(0,0.12),lwd=3, main = "Marginal Posterior over
abline(v=mu_y1, col = "blue", lty=5)

# For the Mild Dataset, we follow the same process
```

```
scale2 = sqrt((nn_y2-1)/var_y2)
shift2 = mu_y2
df = nn_y2 - 1

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)

pdf_y = dt(y, df, ncp = shift2)
pdf_y = pdf_y/scale2
lines(y, pdf_y, col="yellow", lwd=3)
abline(v=mu_y2, col = "red", lty=5)

# For the dementia Dataset, we follow the same process again
scale3 = sqrt((nn_y3-1)/var_y3)
shift3 = mu_y3
df = nn_y3 - 1
#Now, we will use the in-built t-distribution random number generator
y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)

pdf_y = dt(y, df, ncp = shift3)
pdf_y = pdf_y/scale3
lines(y, pdf_y, col="black", lwd=3)
abline(v=mu_y3, col = "grey", lty=5)
legend('topright', c("Control", "Sample mean - control", "Mild", "Sample mean - mild","Dementia", "Sampl
```
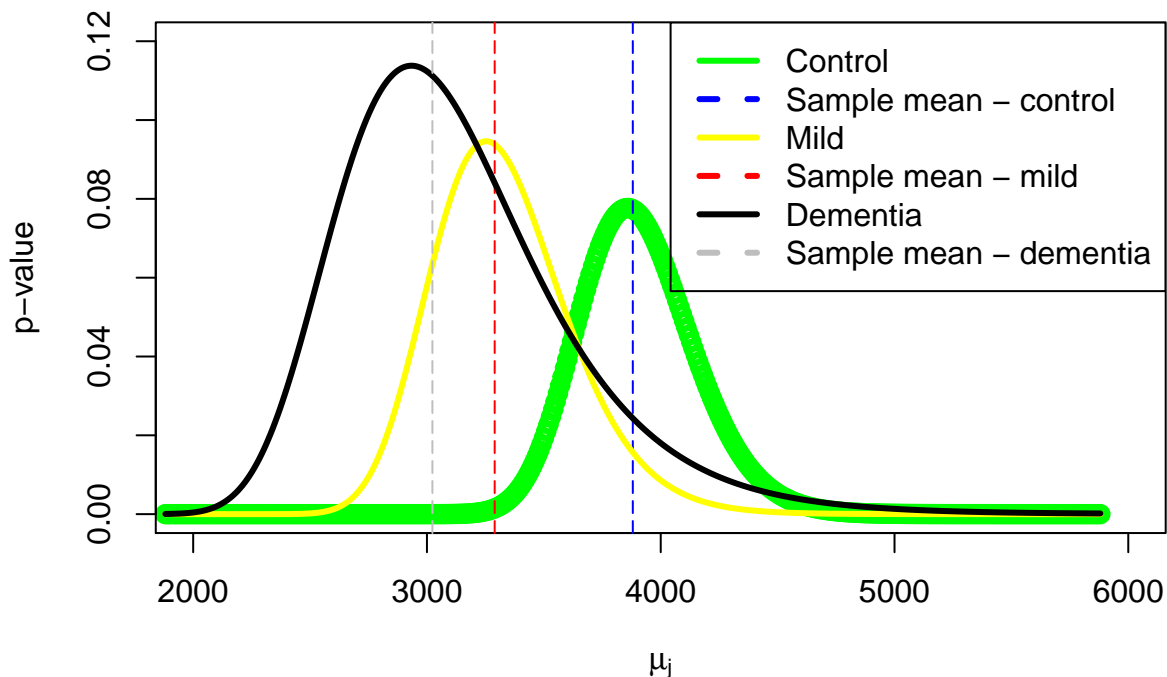
Marginal Posterior over mean using conjugate prior

c)

Solution: Here, we have the random variable $d_{12} = \mu_1 - \mu_2$. Now, we need to find the conditional density function i.e. $p(d_{12}, \sigma_1^2, \sigma_2^2, y_{ij})$ So, we know that the difference between two gaussians is also a gaussian with

15

a different mean and variance, the process is as follows:

$$p(d_{12}, \sigma_1^2, \sigma_2^2, y_{ij}) \sim N(\mu_1, \sigma_1^2) - N(\mu_2, \sigma_2^2)$$

$$p(d_{12}, \sigma_1^2, \sigma_2^2, y_{ij}) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

The code to calculate is as follows:

```
n = 1000000
#Calculate the parameters for the new distribution

#Datesets 1 and 2
new_mean = mun_y1 - mun_y2
var1 = rinvgamma(n, alphan_y1, betan_y1)
var2 = rinvgamma(n, alphan_y2, betan_y2)
new_sigma = sqrt((var1/nn_y1)+(var2/nn_y2))
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req1 = sum(sampled_values < 0)/length(sampled_values)

#Datasets 1 and 3
new_mean = mun_y1 - mun_y3

var1 = rinvgamma(n, alphan_y1, betan_y1)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var1/nn_y1)+ (var3/nn_y3))

sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req2 = sum(sampled_values < 0)/length(sampled_values)

#Datasets 2 and 3
new_mean = mun_y2 - mun_y3
var2 = rinvgamma(n, alphan_y2, betan_y2)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var2/nn_y2)+ (var3/nn_y3))
#new_sigma = sqrt(var2+ var3)
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req3 = sum(sampled_values < 0)/length(sampled_values)
p_req1
```

```
## [1] 0
```

```
p_req2
```

```
## [1] 1e-06
```

```
p_req3
```

```
## [1] 0.032061
```

So, we have the results written above as the output from the R code. The values are $p(d_{12} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$, $p(d_{13} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$, and $p(d_{23} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$. We can rewrite these probabilities i.e $p(d_{12} < 0|\sigma_1^2, \sigma_2^2, y_{ij})$ as $p(\mu_1 < \mu_2|\sigma_1^2, \sigma_2^2, y_{ij})$. Similarly, $p(d_{13} < 0|\sigma_1^2, \sigma_3^2, y_{ij})$ as $p(\mu_1 < \mu_3|\sigma_1^2, \sigma_3^2, y_{ij})$ and $p(d_{23} < 0|\sigma_2^2, \sigma_3^2, y_{ij})$ as $p(\mu_2 < \mu_3|\sigma_2^2, \sigma_3^2, y_{ij})$. For the d_{12} and d_{13} case, we are getting infinitesimal near to zero probability, this can be verified from the plots in 4b where we can see that the overlapped region is very small for the (Control, Mild) and (Control, Dementia) pairs. But for the (Mild, Dementia) pair, we have much more signifcant overlap and thus the probability is higher.

d) Relationship between Jeffrey's prior and the conjugate prior

So, from the 4th ques, we have the conjugate prior as,

$$p(\mu_j, \sigma_j^2) \sim N - IG(\mu_0, n_0, \alpha_0, \beta_0)$$

And, we have the jeffrey prior as

$$p(\mu_j, \sigma_j^2) \sim\propto \frac{1}{\sigma^2}$$

Also, as discussed in class, we know that the Jeffrey prior is the limiting case of the conjugate prior i.e. we can say, the pdf of conjugate would look something like,

$$p(\mu_j, \sigma_j^2) \sim N - IG(0, 0, 0, 0)$$

This is essentially the pdf of Jefferey prior.

Now, from the plots above, we can safely infer that the posteriors obtained when using Jeffrey's prior are almost identical to the posteriors obtained using the conjugate prior. The reason is very simple, as we discussed above, in the limiting case, conjugate prior is similar to the jeffrey's prior and it becomes flatter (non-informativeness increases). So, here the parameters of the conjugate Normal-Inverse Gamma prior are very small (in the order of 10^{-6}), so we are seeing very closely similar plots and results for the Jeffrey and Conjugate priors.

Question 6

Most of this question is similar to the procedure in 4th question except here we know the priors for the distribution in form of values. As discussed in the class, the initial parameters for this N-IG will be,

$$n_0 = 127$$
$$\mu_0 = 2133$$
$$\sigma_0 = 279$$
$$\alpha_0 = \frac{n_0}{2} = 63.5$$
$$\beta_0 = \frac{n_0\sigma_0^2}{2} = 4942903.5$$

The code is as follows:

```
#Most of the variables like sample mean, sample variance etc. can be reused but since the initial param
n_0 = 127
mu_0 = 2133
sigma_0 = 279
alpha_0 = 63.5
```

```
beta_0 = n_0 * sigma_0^2/2

# Parameters for the posterior N-IG as discussed in class
mun_y1 = (n_0 * mu_0 + n_y1*mu_y1)/(n_0 + n_y1)
nn_y1 = n_0 + n_y1
alphan_y1 = alpha_0 + n_y1/2
betan_y1 = beta_0 + (var_y1 * n_y1 /2) + ((n_y1*n_0 / (n_0 + n_y1))*((mu_y1 - mu_0)^2/2))

mun_y2 = ((n_0 * mu_0) + n_y2*mu_y2)/(n_0 + n_y2)
nn_y2 = n_0 + n_y2
alphan_y2 = alpha_0 + n_y2/2
betan_y2 = beta_0 + (var_y2 * n_y2 /2) + ((n_y2*n_0 / (n_0 + n_y2))*((mu_y2 - mu_0)^2/2))

mun_y3 = ((n_0 * mu_0) + n_y3*mu_y3)/(n_0 + n_y3)
nn_y3 = n_0 + n_y3
alphan_y3 = alpha_0 + n_y3/2
betan_y3 = beta_0 + (var_y3 * n_y3 /2) + ((n_y3*n_0 / (n_0 + n_y3))*((mu_y3 - mu_0)^2/2))
```

a)

Solution: The code to plot the histogram, pdf and sample variance is written below.

```
#Part a of Question 4 repeated
n = 1000000
# We need to plot for only the control group
sig = rinvgamma(n, alphan_y1, betan_y1)
p_sig = dinvgamma(sig, alphan_y1, betan_y1)

h = hist(sig,main = "Marginal Posterior Density" ,xlab = "Variance", ylab = "Value of posterior", xlim=
lines(density(sig), col = "green", lwd=2)
abline(v= var_y1, col = "blue", lwd = 2)
legend('topright', c("Histogram of Variance", "PDF of posterior", "Sample Variance"), col = c("black",
```

Marginal Posterior Density
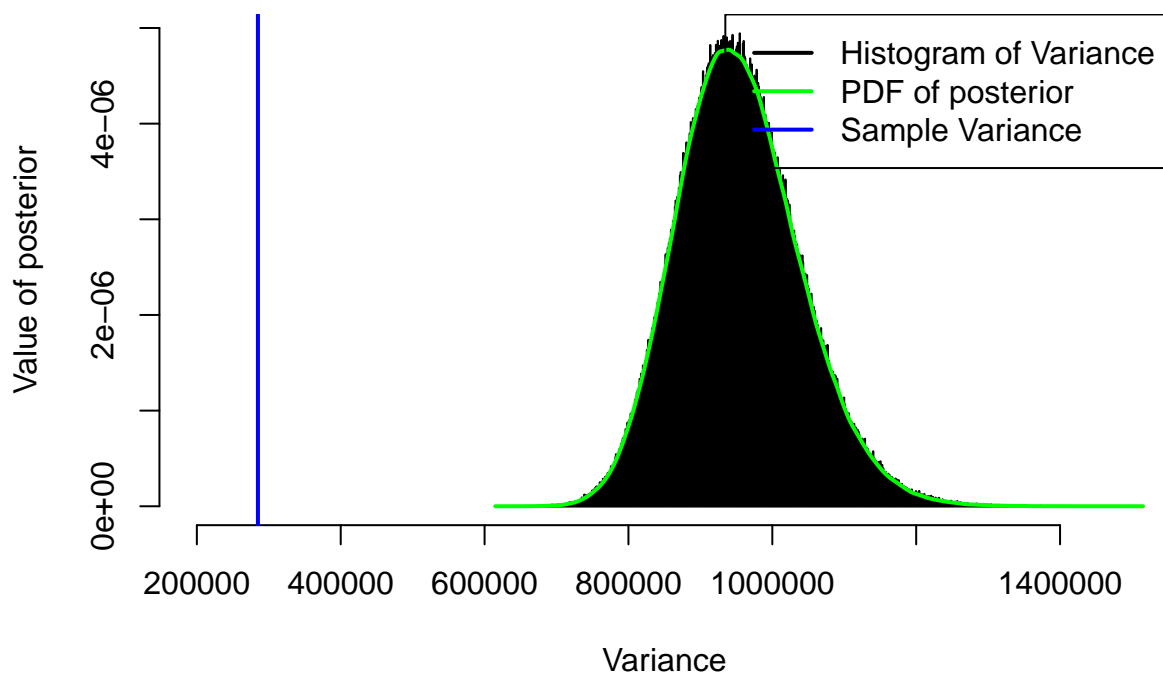
From the above plot, we can conclude that the sample of the σ_j^2 parameter distribution is shifted a lot as compared to the sample variance.

b)

Solution: Similar to the b part of question 4, the code is as follows:

```
### Part b of question 4 repeated
n = 1000000

#For the control dataset, the shift, scale and degree of freedom parameters are
scale1 = sqrt(alphan_y1 * nn_y1/betan_y1)
shift1 = mun_y1
df = 2*alphan_y1

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)
# Pdf of y from dt
pdf_y = dt(y, df, ncp = shift1)
pdf_y = pdf_y/scale1
plot(y,pdf_y, col ="green", xlim = c(2000, 5500), ylim=c(0,0.20),lwd=3, main = "Marginal Posterior over
abline(v=mu_y1, col = "blue", lty=5)

# For the Mild Dataset, we follow the same process
shift2 = mun_y2
scale2 = sqrt(alphan_y2 * nn_y2/betan_y2)
df = 2*alphan_y2

y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)
```
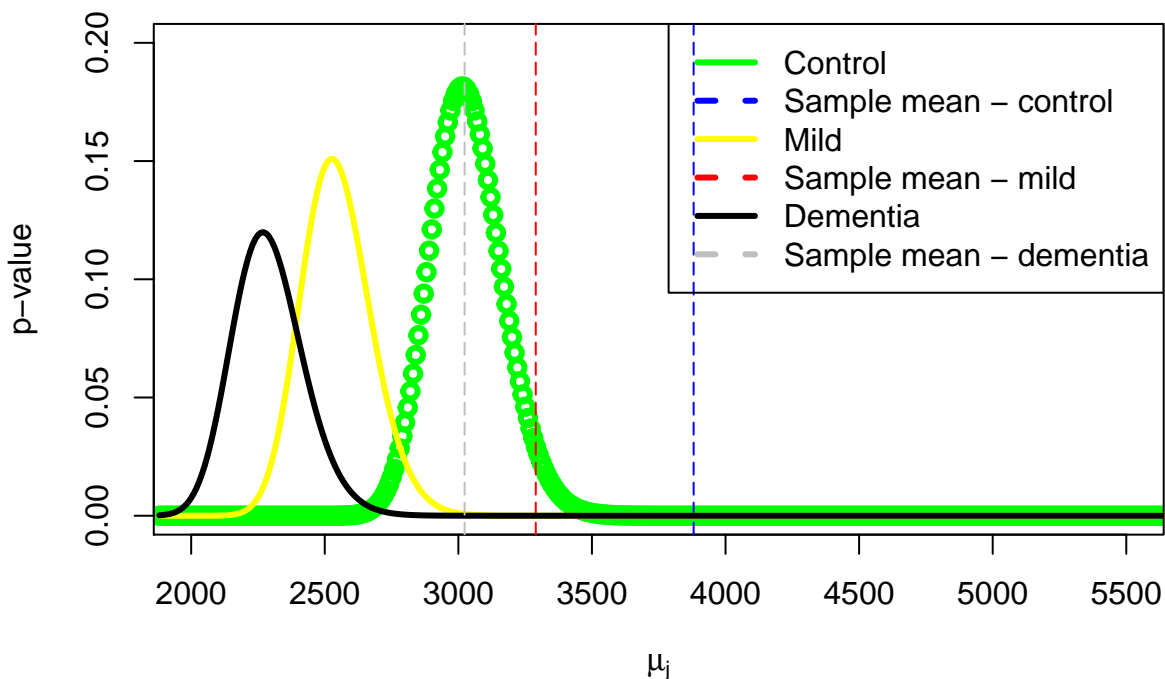
```
pdf_y = dt(y, df, ncp = shift2)
pdf_y = pdf_y/scale2
lines(y, pdf_y, col="yellow", lwd=3)
abline(v=mu_y2, col = "red", lty=5)

# For the dementia Dataset, we follow the same process again
shift3 = mun_y3
scale3 = sqrt(alphan_y3 * nn_y3/betan_y3)
df = 2*alphan_y3

#Now, we will use the in-built t-distribution random number generator
y = seq(mu_y1 - 2000, mu_y1 + 2000, 10)

pdf_y = dt(y, df, ncp = shift3)
pdf_y = pdf_y/scale3
lines(y, pdf_y, col="black", lwd=3)
abline(v=mu_y3, col = "grey", lty=5)
legend('topright', c("Control", "Sample mean - control", "Mild", "Sample mean - mild","Dementia", "Sampl
```

Marginal Posterior over mean using conjugate prior

From the above plot it is clear that the sample of the mean values extracted in the graph are very different from the sample means for all the three groups.

c)

Solutions: Similar to question 4c, the process and the code is as follows:

```
n = 1000000
#Calculate the parameters for the new distribution
```

```r
#Datesets 1 and 2
new_mean = mun_y1 - mun_y2
var1 = rinvgamma(n, alphan_y1, betan_y1)
var2 = rinvgamma(n, alphan_y2, betan_y2)
new_sigma = sqrt((var1/nn_y1)+(var2/nn_y2))
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req1 = sum(sampled_values < 0)/length(sampled_values)
#Datasets 1 and 3
new_mean = mun_y1 - mun_y3

var1 = rinvgamma(n, alphan_y1, betan_y1)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var1/nn_y1)+ (var3/nn_y3))

sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req2 = sum(sampled_values < 0)/length(sampled_values)

#Datasets 2 and 3
new_mean = mun_y2 - mun_y3
var2 = rinvgamma(n, alphan_y2, betan_y2)
var3 = rinvgamma(n, alphan_y3, betan_y3)
new_sigma = sqrt((var2/nn_y2)+ (var3/nn_y3))
sampled_values = rnorm(n, new_mean, new_sigma)

#Now we will calculate the mentioned probability
p_req3 = sum(sampled_values < 0)/length(sampled_values)
#hist(sampled_values)
#print the values out
p_req1
```

```
## [1] 0
```

```r
p_req2
```

```
## [1] 0
```

```r
p_req3
```

```
## [1] 2.9e-05
```

This question is different from ques 4 in terms of the initial parameters. Here, the parameters are nowhere near zero in contrast to the question 4. This change affects the results obtained. From the results in the above three parts, we can infer that this is not a good prior to use as the posterior means are very different from the sample means and the same goes for the variance. Essentially, this can be a sign of biasness of the prior which in theory should be an un-informative prior like Jeffrey's prior.