CS6190: Probabilistic Modeling Homework 1
Exponential Families, Conjugate Priors

Gopal Menon

08 February, 2018

Written Part

1. **Expectation of sufficient statistics:** Consider a random variable X from a continuous exponential family with natural parameter $\eta = (\eta_1, \ldots, \eta_n)$. Recall this means the pdf is of the form:

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

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

Hint: Start with the identity $\int p(x)dx = 1$, and take the derivative with respect to η.

$$\begin{aligned}
\int p(x)dx &= 1 \\
\Rightarrow \int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx &= 1 \\
\nabla \left(\int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx\right) &= 0 \\
\int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) . \left(T(x) - \nabla A(\eta)\right) dx &= 0 \\
\int T(x) h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx &= \nabla A(\eta) \int h(x) \exp\left(\eta \cdot T(x) - A(\eta)\right) dx \\
\int T(x) p(x) dx &= \nabla A(\eta) \int p(x) dx \\
E\left[T(x)\right] &= \nabla A(\eta)
\end{aligned}$$

(b) Verify this formula works for the Gaussian distribution with unknown mean, μ, and known variance, σ^2.

Hint: Start by thinking about what the natural parameter η and the function $A(\eta)$ are, then verify that the expectation of the Gaussian is the same as $\nabla A(\eta)$.

The pdf for a Gaussian distribution with unknown mean, μ, and known variance, σ^2 is given by:

$$
\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\ln\left(\frac{1}{\sigma}\right)\right] \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\ln(\sigma) - \frac{(x-\mu)^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\ln(\sigma) - \frac{x^2}{2\sigma^2} + \frac{2x\mu}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)\right] \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[\begin{bmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{bmatrix} \cdot \begin{bmatrix} x \\ x^2 \end{bmatrix} - \left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)\right]
\end{aligned}
$$

Based on the form of the exponential family pdf, we can see that:

$$
\begin{aligned}
A(\eta) &= \ln(\sigma) + \frac{\mu^2}{2\sigma^2}, \\
\eta &= \begin{bmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{bmatrix}, \text{ and} \\
T(x) &= \begin{bmatrix} x \\ x^2 \end{bmatrix}
\end{aligned}
$$

Based on the equation for $E[T(x)]$ above:

$$
\begin{aligned}
E[T(x)] &= \nabla A(\eta) \\
\Rightarrow E\left[\begin{bmatrix} x \\ x^2 \end{bmatrix}\right] &= \begin{bmatrix} \frac{\partial A}{\partial \eta_1} \\ \frac{\partial A}{\partial \eta_2} \end{bmatrix}, \text{ where } \eta_1 = \frac{\mu}{\sigma^2}, \text{ and } \eta_2 = \frac{-1}{2\sigma^2} \\
\begin{bmatrix} E[x] \\ E\left[x^2\right] \end{bmatrix} &= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln(\sigma) + \frac{\mu^2}{2\sigma^2}\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\sigma^2}{2}\frac{\mu^2}{\sigma^4}\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln\left(\sqrt{\frac{-1}{2\eta_2}}\right) - \mu^2\eta_2\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial\left(\ln(\sigma) + \frac{\sigma^2}{2}\eta_1^2\right)}{\partial \eta_1} \\ \frac{\partial\left(\ln\left(\frac{i}{\sqrt{2}\sqrt{\eta_2}}\right) - \mu^2\eta_2\right)}{\partial \eta_2} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma^2}{2}.2\eta_1 \\ \frac{\sqrt{2}\sqrt{\eta_2}}{i}.\frac{i}{\sqrt{2}}.\frac{-1}{2}\eta_2^{\frac{-3}{2}} - \mu^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma^2}{2}.2\frac{\mu}{\sigma^2} \\ \frac{-1}{2\eta_2} - \mu^2 \end{bmatrix} \\
&= \begin{bmatrix} \mu \\ \sigma^2 - \mu^2 \end{bmatrix} \\
\Rightarrow E[x] &= \mu, \text{ and} \\
E\left[x^2\right] &= \sigma^2 - \mu^2, \text{ which is not the case since } \sigma^2 = E\left[x^2\right] - \mu^2
\end{aligned}
$$

2. **Noninformative priors for the Poisson distribution:** Let $X \sim Pois(\lambda)$. Recall that the pmf of the Poisson is

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

(a) Rewrite the above pmf in exponential family form. What is the natural parameter? What is the sufficient statistic?

$$\begin{aligned} P(X = k; \lambda) &= \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \frac{1}{k!} \exp\left(\ln \lambda^k e^{-\lambda}\right) \\ &= \frac{1}{k!} \exp\left(k \ln \lambda - \lambda\right) \\ \eta &= \ln \lambda, \text{ natural parameter} \\ h(k) &= k, \text{ sufficient statistic} \end{aligned}$$

(b) Give at least two different options for noninformative priors for $p(\lambda)$.

i. **Uniform Prior:** In this case the prior is

$$p(\lambda) = c, \text{ for } 0 \leq x \leq \infty, \text{, where c is a constant}$$

ii. **Jeffrey's Prior:** In this case the prior is

$$
\begin{aligned}
\mathcal{I}(\lambda) &= E\left[\left(\frac{\partial}{\partial\lambda}\ln p(x|\lambda)\right)^2\right] \\
&= -E\left[\frac{\partial^2}{\partial\lambda^2}\ln p(x|\lambda)\right] \\
&= -E\left[\frac{\partial^2}{\partial\lambda^2}\ln\frac{\lambda^x e^{-\lambda}}{x!}\right] \\
&= -E\left[\frac{\partial^2}{\partial\lambda^2}\left(x\ln\lambda - \lambda - \ln x!\right)\right] \\
&= E\left[\frac{\partial}{\partial\lambda}\left(\frac{x}{\lambda} - 1\right)\right] \\
&= -E\left[\frac{-x}{\lambda^2}\right] \\
&= E\left[\frac{x}{\lambda^2}\right] \\
&= \frac{1}{\lambda^2}E\left[x\right] \\
&= \frac{1}{\lambda^2}\sum_{i=1}^{\infty} xp(x) \\
&= \frac{1}{\lambda^2}\sum_{i=1}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{e^{-\lambda}}{\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} \\
&= \frac{e^{-\lambda}}{\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots\right) \\
&= \frac{e^{-\lambda}}{\lambda}e^{\lambda} \\
&= \frac{1}{\lambda} \\
p(\lambda) &= \sqrt{\mathcal{I}(\lambda)} \\
&= \sqrt{\frac{1}{\lambda}}
\end{aligned}
$$

(c) What are the resulting posteriors for your two options? Are they proper (i.e., can they be normalized)?

i. **Uniform Prior:** The posterior is

$$
\begin{aligned}
p(\lambda|x) &\propto p(x|\lambda)p(\lambda) \\
&= \prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}.c, \text{based on the assumtion of i.i.d. samples} \\
&\propto \prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}
\end{aligned}
$$

Even though the prior cannot be normalized, since it will not add up to 1, the posterior shown above, where the constant c has been dropped, can be normalized. The reason is that the posterior is the product of distributions that can themselves be normalized.

ii. **Jeffrey's Prior:** The posterior is

$$\begin{aligned} p(\lambda|x) &\propto p(x|\lambda)p(\lambda) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \cdot \sqrt{\frac{1}{\lambda}}, \text{based on the assumtion of i.i.d. samples} \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i - \frac{1}{2}} e^{-\lambda}}{x_i!} \end{aligned}$$

This can be normalized since it is the product of the likelihood, which is a product of Poisson distributions, and so can be normalized, and the prior, which can also be normalized.

3. **Non-conjugate priors:** Let X_i be from a Gaussian with known variance σ^2 and mean μ with uniform prior, i.e.,

$$\begin{aligned} \mu &\sim Unif(a, b) \\ X_i &\sim N(\mu, \sigma^2) \end{aligned}$$

What is the posterior pdf, $p(\mu|x_1, \ldots, x_n; \sigma^2, a, b)$?

Hint: There will be an integral that you won't be able to analytically solve (just leave it in integral form).

$$\begin{aligned} p(\mu|x_1, \ldots, x_n; \sigma^2, a, b) &= p(x_1, \ldots, x_n|\mu; \sigma^2)p(\mu; a, b) \\ &= \int_{-\infty}^{\infty} p(x|\mu; \sigma^2)p(\mu; a, b)dx \\ &= \int_{a}^{b} p(x|\mu; \sigma^2)\frac{1}{b-a}dx \\ &= \frac{1}{b-a}\frac{1}{\sqrt{2\pi}\sigma}\int_{a}^{b} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \end{aligned}$$

R Coding Part

4a

```r
# Density function for Inverse Gamma: conjugate prior for Gaussian sigma^2
density_inverse_gamma <- function(x_coordinates, alpha, beta) {

  return (beta^alpha / gamma(alpha) * x_coordinates^(-alpha - 1) * exp(-beta / x_coordinates))

}

# Density function for Normal-Inverse-Gamma with parameters mu0, lambda, alpha, beta
density_normal_inverse_gamma <-
  function(distribution_of_mean, distribution_of_variance, mu0 = 0, lambda = 1, alpha = 8, beta = 16)
{
  return (dnorm(distribution_of_mean, mean = mu0, sd = sqrt(distribution_of_variance / lambda)) *
            density_inverse_gamma(distribution_of_variance, alpha, beta))
}

# Read data for hippocampus volume and patient details
hippocampus_volume <- read.csv(file="oasis_hippocampus.csv", header=TRUE, sep=",")
```

```r
patient_details <- read.csv(file="oasis_cross-sectional.csv", header=TRUE, sep=",")

# Join both frames on identifer
patient_hippocampus_volume <- merge(hippocampus_volume, patient_details, by="ID")

# Select right hippocampal volume for control group
control_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR==0.0, select=RightHippoVol)

# Select right hippocampal volume for mild dementia group
mild_dementia_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR!=0.0 & CDR!=2.0 & !is.null(CDR),
         select=RightHippoVol)

# Select right hippocampal volume for dementia group
dementia_group_rt_hippo_vol <-
  subset(patient_hippocampus_volume, CDR==2.0, select=RightHippoVol)

library(fitdistrplus)
```

```
## Loading required package: MASS

## Loading required package: survival
```

```r
library(MASS)
library(survival)

# Model each group as a normal variable with its own mean and variance
control_group_rt_hippo_vol_normal_model <-
  fitdistr(control_group_rt_hippo_vol[[1]],"normal")
mild_dementia_group_rt_hippo_vol_normal_model <-
  fitdistr(mild_dementia_group_rt_hippo_vol[[1]],"normal")
dementia_group_rt_hippo_vol_normal_model <-
  fitdistr(dementia_group_rt_hippo_vol[[1]],"normal")
```

The joint posterior $p(\mu_j, \sigma_j^2|y_{ij})$ computation is given by:

$$
\begin{aligned}
p(\mu_j, \sigma_j^2|y_{ij}) &\propto p(y|\mu,\sigma^2)p(\mu,\sigma^2), \text{ where } p(\mu,\sigma^2) \sim N-IG(\mu_0, n_0, \alpha, \beta) \\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(y_i-\mu_n)^2}{2\sigma^2}\right]\right)\frac{\sqrt{n_0}}{\sigma\sqrt{2\pi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1} exp\left[-\frac{2\beta+n_0(\mu-\mu_0)^2}{2\sigma^2}\right] \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{n_0^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\frac{n}{2}-\frac{1}{2}-\alpha-1} exp\left[\frac{\sum_{i=1}^{n}(y_i-\mu)^2-n_0(\mu-\mu_0)^2-2\beta}{2\sigma^2}\right] \\
p(\sigma_j^2|y_{ij}) &= \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{n_0^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-\frac{n}{2}-\frac{1}{2}-\alpha-1}\int_{-\infty}^{\infty} exp\left[\frac{\sum_{i=1}^{n}(y_i-\mu)^2-n_0(\mu-\mu_0)^2-2\beta}{2\sigma^2}\right] d\mu
\end{aligned}
$$

Since the integral of a normal distribution with a non-zero mean will be the same as the integral of a normal distribution with a zero mean (the area under the curve stays the same), this can be written as:

$$
\begin{aligned}
p(\sigma_j^2|y_{ij}) &= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{n_0^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\frac{n}{2}-\frac{1}{2}-\alpha-1} \int_{-\infty}^{\infty} exp\left[\frac{-(n_0+n)\mu^2 - 2\beta}{2\sigma^2}\right] d\mu \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{n_0^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\frac{n}{2}-\alpha-\frac{3}{2}} \frac{\sqrt{2\pi}\sigma}{\sqrt{n_0+n}} exp\left[\frac{-2\beta}{2\sigma^2}\right] \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{n_0^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\frac{n}{2}-\alpha-1} \frac{\sqrt{2\pi}}{\sqrt{n_0+n}} exp\left[\frac{-\beta}{\sigma^2}\right]
\end{aligned}
$$

The above expression for the marginal posterior is of the Inverse Gamma distribution form with parameters. The definite integral was done online [1].

$$
\begin{aligned}
-\alpha_n - 1 &= -\frac{n}{2} - \alpha - 1 \\
\alpha_n &= \alpha + \frac{n}{2}
\end{aligned}
$$

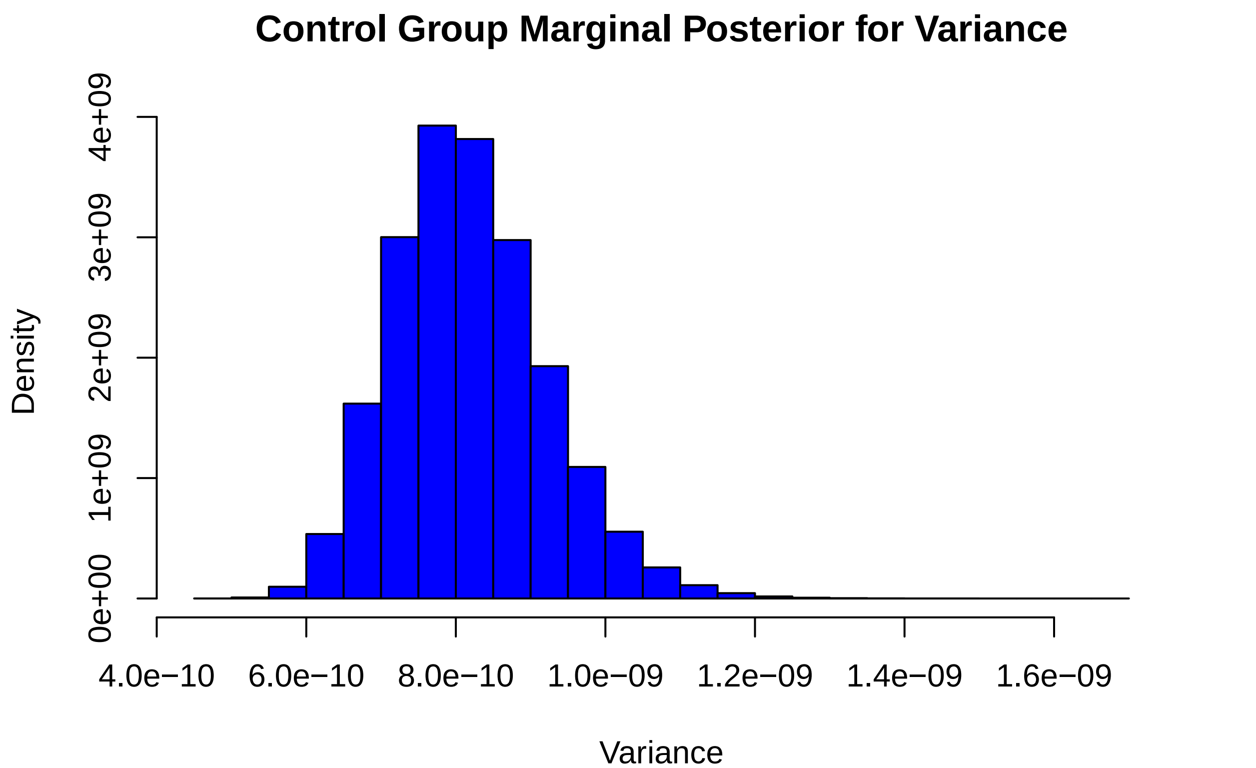
and β_n remaining unchanged from that of the N-IG distribution. The value of β_n, as per the class notes is:

$$
\beta_n = \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{2}\frac{n_0 n}{n_0+n}(\mu_0 - \bar{x})^2
$$

The following is based on Wikipedia formulas [5].

```r
# Parameters for the N-IG distribution
PRIOR_PSEUDO_SAMPLE_MEAN_MU = 0
PRIOR_PSEUDO_OBSERVATION_COUNT_NU = 10^(-6)
PRIOR_PSEUDO_VARIANCE_OBS_CNT_ALPHA = 10^(-6)
PRIOR_PSEUDO_SUM_SQUARE_DEV_BETA = 10^(-6)

# Statistics from observations
control_grp_observations_count = length(control_group_rt_hippo_vol[[1]])
control_grp_sample_mean = control_group_rt_hippo_vol_normal_model$estimate[[1]]
control_grp_variance = control_group_rt_hippo_vol_normal_model$sd[[1]]^2

# Compute the parameters for the N-IG posterior distribution
control_grp_posterior_mu = (PRIOR_PSEUDO_OBSERVATION_COUNT_NU * PRIOR_PSEUDO_SAMPLE_MEAN_MU +
                              control_grp_observations_count * control_grp_sample_mean) /
  (PRIOR_PSEUDO_OBSERVATION_COUNT_NU + control_grp_observations_count)

control_grp_posterior_nu = PRIOR_PSEUDO_OBSERVATION_COUNT_NU + control_grp_observations_count

control_grp_posterior_alpha = PRIOR_PSEUDO_VARIANCE_OBS_CNT_ALPHA + 0.5 * control_grp_observations_coun

control_grp_posterior_beta = PRIOR_PSEUDO_SUM_SQUARE_DEV_BETA +
  0.5 * sum((control_group_rt_hippo_vol[[1]] - control_grp_sample_mean)^2) +
  (control_grp_observations_count * PRIOR_PSEUDO_OBSERVATION_COUNT_NU /
  (PRIOR_PSEUDO_OBSERVATION_COUNT_NU + control_grp_observations_count)) *
  0.5 * (control_grp_sample_mean - PRIOR_PSEUDO_SAMPLE_MEAN_MU)^2

# Sample from the inverse gamma marginal posterior for variance
NUMBER_OF_SAMPLES = 10^6
samples_from_ctrl_grp_variance_marginal_posterior =
  1/rgamma(n=NUMBER_OF_SAMPLES, shape=control_grp_posterior_alpha, scale=control_grp_posterior_beta)
```

```r
# Plot a histogram of the samples from the marginal posterior for variance
hist(samples_from_ctrl_grp_variance_marginal_posterior, freq=FALSE,
     main= "Control Group Marginal Posterior for Variance", xlab="Variance", ylab="Density", col="blue")
```




```r
library(MCMCpack)
```

```
## Loading required package: coda

## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2018 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park

## ##
## ## Support provided by the U.S. National Science Foundation

## ## (Grants SES-0350646 and SES-0350613)
## ##
```

```r
control_grp_posterior_sigma =
  1/rgamma(n=1000, shape=control_grp_posterior_alpha, scale=control_grp_posterior_beta)
control_grp_posterior_density =
  dinvgamma(control_grp_posterior_sigma,
            shape=control_grp_posterior_alpha, scale=control_grp_posterior_beta)
```

I tried plotting the density function over the histogram (see code after histogram), but got all zero values. Also I did not plot the sample variance over the histogram as the value was very different from that shown on the histogram. The control group variance and the variance values on the histogram are different in orders of magnitude.

I skipped the plotting of the remaining two groups as the values for the control group were wrong.

4b

The marginal posterior $p(\mu_j|y_{ij})$ is given by:

$$p(\mu_j|y_{ij}) \propto \int_{-\infty}^{\infty} p(y|\mu,\sigma^2)p(\mu,\sigma^2)d\sigma^2, \text{ where } p(\mu,\sigma^2) \sim N-IG(\mu_0,n_0,\alpha,\beta)$$

If we integrate out the variance from the Normal Inverse-Gamma prior, we should be left with the likelihood times the normal distribution of the prior for μ. This will be a Normal distribution since the likelihood is the product of Normally distributed variables and so is the prior for μ. I tried doing this analytically, but did not get very far as the integral did not reduce to a recognizable form.

As given in chapter 2 of Bayesian Data Analysis[2], the parameters for marginal posterior will be:

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \text{ and } \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

where subscripts 0 and n are for the prior and posterior respectively, n is the number of samples, $\bar{y}$ is the sample mean, and τ is the precision. However this did not help as τ_0 and μ_0 are unknown.

4c

Given that the random variable $d_{12} = \mu_1 - \mu_2$ is the difference of two normally distributed random variables, d_{12} will have a Normal distribution with $\mu_{d_{12}} = \mu_{\mu_1} - \mu_{\mu_2}$ and $\sigma_{12}^2 = \sigma_{\mu_1} + \sigma_{\mu_2}$ [3]. Since I could not work out the previous question on the posterior marginal for μ_j, I am not able to proceed further on this part.

5

The joint posterior $p(\mu_j, \sigma_j^2|y_{ij})$ computation for the Jeffrey's prior $p(\mu_j, \sigma_j) \propto \sigma^{-2}$ is given by:

$$\begin{aligned}
p(\mu_j,\sigma_j|y_{ij}) &\propto p(y|\mu,\sigma)p(\mu_0,\sigma_0), \text{ where } p(\mu_0,\sigma_0) \propto \sigma^{-2} \\
&\propto \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma^2}}exp\left[-\frac{(y_i-\mu_n)^2}{2\sigma^2}\right]\right)\sigma^{-2} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}(\sigma^2)^{\frac{n}{2}-1}exp\left[\frac{-\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right] \\
p(\sigma_j|y_{ij}) &\propto \frac{1}{(2\pi)^{\frac{n}{2}}}(\sigma^2)^{-\frac{n}{2}-1}\int_{-\infty}^{\infty} exp\left[\frac{-\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right]d\mu \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}(\sigma^2)^{-\frac{n}{2}-1}\int_{-\infty}^{\infty} exp\left[\frac{-n\mu^2}{2\sigma^2}\right]d\mu, \text{ for the same reason as given in 4a} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}(\sigma^2)^{-\frac{n}{2}-1}\frac{\sqrt{\pi}\sqrt{2\sigma^2}}{\sqrt{n}}, \text{ using [1]} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{\sqrt{2\pi}}{\sqrt{n}}(\sigma^2)^{-\frac{n}{2}-1+\frac{1}{2}} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{\sqrt{2\pi}}{\sqrt{n}}(\sigma^2)^{-\frac{n}{2}-\frac{1}{2}} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{\sqrt{2\pi}}{\sqrt{n}}\sigma^{-n-1}
\end{aligned}$$

As per this paper [4] on non-informative priors, the posteriors are given by:

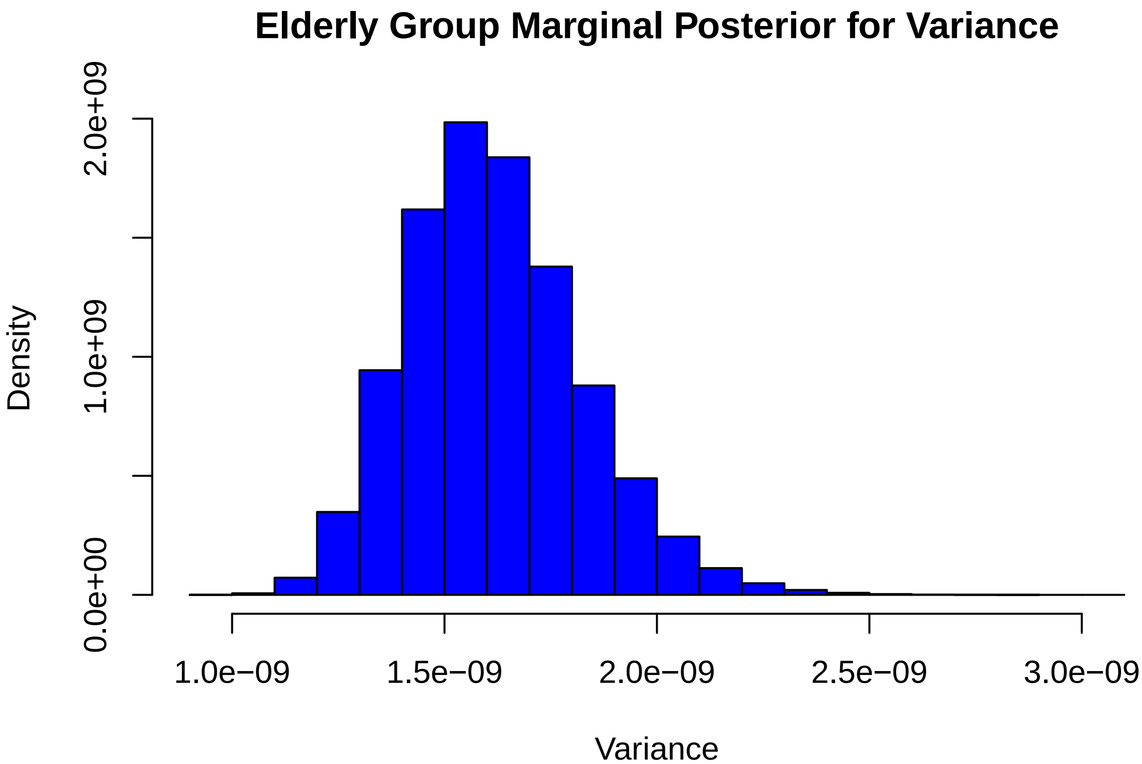
$$p(\mu|D) \sim T\left(n-1, \bar{x}, \frac{S^2}{n(n-1)}\right), \text{ and}$$
$$p(\sigma^2|D) \sim IG\left(\frac{(n-1)}{2}, \frac{2}{S^2}\right), \text{ where}$$
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \text{ and}$$
$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

```
# Marginal posterior for variance using Jeffrey's prior
ctrl_grp_sum_of_squared_deviations = sum((control_group_rt_hippo_vol[[1]] - control_grp_sample_mean)^2)
ctrl_grp_variance_jeffr_posterior_samples = 1/rgamma(n=NUMBER_OF_SAMPLES,
                                                     shape=(control_grp_observations_count-1)/2,
                                                     scale=2/ctrl_grp_sum_of_squared_deviations)
```

I got an error while plotting the histogram for the variance posterior using the Jeffrey's prior. So I skipped the rest of the analysis.

6

```
# Prior parameter values for N-IG distribution
ELDERLY_PRIOR_SAMPLE_MEAN = 2133
ELDERLY_PRIOR_SAMPLE_COUNT = 127
ELDERLY_PRIOR_ALPHA = ELDERLY_PRIOR_SAMPLE_COUNT/2
ELDERLY_PRIOR_STD_DEV = 279
ELDERLY_PRIOR_BETA = ELDERLY_PRIOR_SAMPLE_COUNT * ELDERLY_PRIOR_STD_DEV^2

# Sample from the inverse gamma marginal posterior for variance
samples_from_elderly_grp_variance_marginal_posterior =
  1/rgamma(n=NUMBER_OF_SAMPLES, shape=ELDERLY_PRIOR_ALPHA, scale=ELDERLY_PRIOR_BETA)

# Plot a histogram of the samples from the marginal posterior for variance
hist(samples_from_elderly_grp_variance_marginal_posterior, freq=FALSE,
     main= "Elderly Group Marginal Posterior for Variance", xlab="Variance", ylab="Density", col="blue")
```



The variance values seem to be incorrect as they are very small.

My Results for the R Coding Part I started the coding part with mathematical analysis as I did not know the formulas. That exercise took a long time and was not successful. After the end of the class on Thursday February 8, I got some pointers from fellow students on the Wikipedia page on conjugate priors that had details on how to arrive at posterior model parameters. I also got information on another page relating to the Jeffrey's prior. I attempted to use the information to arrive at model parameters, but I was not getting accurate results - the variance values were too small to be accurate. I was not left with much time to get this to work.

Although the assignment is structured so as to provide a better understanding of evaluating model parameters, since I did not get any results, I do not have a good understanding of the process of doing the prediction.

References

[1] "Integral Calculator." *Integral Calculator With Steps!*, www.integral-calculator.com/.

[2] "2.5 Normal Distribution with Known Variance." *Bayesian Data Analysis*, by Andrew Gelman, CRC Press/Taylor & Francis, 2013.

[3] "Normal Difference Distribution." *Normal Difference Distribution – from Wolfram Mathworld*, mathworld.wolfram.com/NormalDifferenceDistribution.html.

[4] "A Catalog of Noninformative Priors.", http://www.stats.org.uk/priors/noninformative/YangBerger1998.pdf.

[5] "Conjugate Prior." *Wikipedia*, Wikimedia Foundation, 15 Jan. 2018.