HW2 Solutions
36-724: Applied Bayesian and Computational Statistics

February 13, 2006

Problem 1

We are given that y is the number of heads in n spins of a coin, where the probability of heads is θ. Thus, $y \sim$ Binomial(n, θ).

a Prior Predictive Distribution

Let $p(\theta) = 1$ for $0 \leq \theta \leq 1$.

$$\begin{aligned} P(Y = k) &= \int P(Y = k|\theta)P(\theta)d\theta \\ &= \int_0^1 P(Y = k|\theta)d\theta) \\ &= \int_0^1 \binom{n}{k}\theta^k(1-\theta)^{n-k}d\theta \end{aligned}$$

Note that this is the kernal for a beta distribution, with parameters $\alpha = k + 1, \beta = n - k + 1$.

$$\begin{aligned} P(Y = k) &= \int_0^1 \binom{n}{k}\theta^k(1-\theta)^{n-k}d\theta \\ &= \binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \end{aligned}$$

Because k and n are both integers, this reduces to $\frac{1}{n+1}$.

b Posterior Means

We know that the beta and the binomial are conjugate priors, so we can get the mean using that fact. We can also show it using the above fact, the likelihood and the prior. The posterior of θ given y is beta(α+y, $n - y + \beta$). Thus the mean of the posterior is $\frac{\alpha+y}{\alpha+\beta+n}$

I will show that this is between $\frac{y}{n}$ and $\frac{\alpha}{\alpha+\beta}$ when $\frac{\alpha}{\alpha+\beta} \leq \frac{y}{n}$. The other case can be shown similarly.

$$\begin{aligned} \frac{\alpha+y}{\alpha+\beta+n} - \frac{\alpha}{\alpha+\beta} &> 0 \Leftrightarrow \\ y\alpha + y\beta - \alpha n &> \Leftrightarrow \\ \frac{y}{n} &> \frac{\alpha}{\alpha+\beta} \end{aligned}$$

Since the last equation is true, and it is equivalent to the first equation, we know that $\frac{\alpha+y}{\alpha+\beta+n} > \frac{\alpha}{\alpha+\beta}$

Similarly

$$\begin{aligned} \frac{y}{n} - \frac{\alpha+y}{\alpha+\beta+n} &> 0 \Leftrightarrow \\ y\alpha + y\beta - n\alpha &> 0 \Leftrightarrow \\ \frac{y}{n} &> \frac{\alpha}{\alpha+\beta} \end{aligned}$$

Again, because the last equation is true, and it is equivalent to the first equation, we know that $\frac{y}{n} > \frac{\alpha+y}{\alpha+\beta+n}$

c Posterior Variance

Here, the posterior distribution is proportional to $\binom{n}{y}\theta^y(1-\theta)^{n-y}$. Thus, the posterior is a beta distribution with parameters $\alpha = y+1$ and $\beta = n-y+1$. So, the variance is $\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$. We want to show that this is less than $\frac{1}{12}$, the variance of a uniform$(0,1)$.

Well, $\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} = \frac{1}{n+3}(\frac{y+1}{n+2})(1-\frac{y+1}{n+2})$. We know that $\frac{1}{n+3} < \frac{1}{3}$. We also know that for all $0 \leq a \leq 1$, the quantity $a(1-a)$ is minimized at $\frac{1}{4}$ when $a = 0.5$. Thus $\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} \leq \frac{1}{4}\frac{1}{3} \leq \frac{1}{12}$

d Example ¿From part b), we know that the posterior distribution is a beta with parameters $\alpha_1 = \alpha + y$ and $\beta_1 = n - y + \beta$ (where α and β are the parameters on the prior). We know that the prior variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ and the posterior variance is $\frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)^2(\alpha_1+\beta_1+1)}$. Let $\alpha = 1, \beta = 10, y = 1, n = 1$, then $\alpha_1 = 2, \beta_1 = 10$. The prior variance is $\frac{10*1}{11^2*12} = 0.006$ and the posterior variance is $\frac{2*10}{12^2*13} = 0.01$

The reason this is happening, is because the prior variance is small, so prior to the data, it is believed that θ will be near the mean – in this case $\frac{1}{11}$. Thus, the probability that $y = 1$ is 0.091. However, with one data point, $y = 1$, so that conflicts with the prior causing posterior confusion!

Problem 2

a Normalized Posterior

We know that the posterior is the likelihood times the prior. I will assume that the data points are independent.

$$p(\theta|y) \propto \prod_{i=1}^{5} \frac{1}{1+(y_i-\theta)^2}$$

A graph of the normalized posterior is in Figure 1. The R-Code used to produce the graph of the normalized posterior is below.

```
> y=c(-2, -1, 0, 1.5, 2.5)
> grid.size=1000
> theta=seq(0, 1, length=grid.size)
> post=dcauchy(y[1], theta, 1)
```

```
> for (i in 2:5){
+ post=post*dcauchy(y[i], theta, 1)}
> post=post/sum(post)*grid.size
> plot(theta, post, type='l')
```

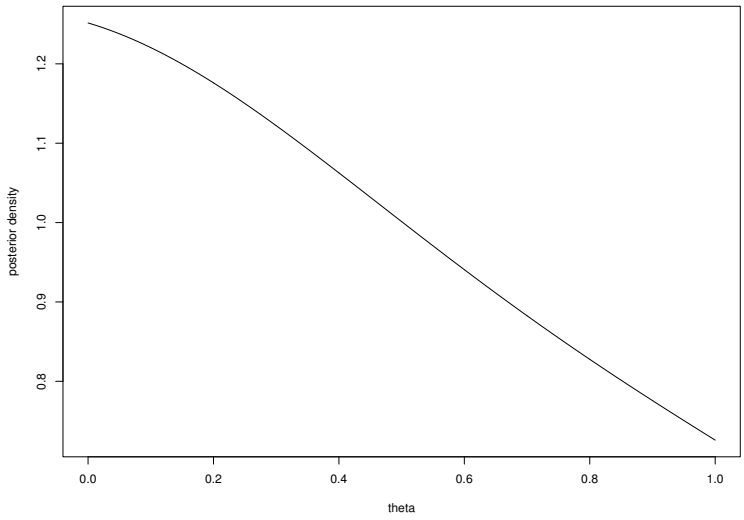

Figure 1: Posterior of theta for problem 2a

b Sample from the posterior density

To sample from the posterior density, generate a sample from the grid points, and then associate them back to the θ values. That will generate a sample from θ. See Figure 2.

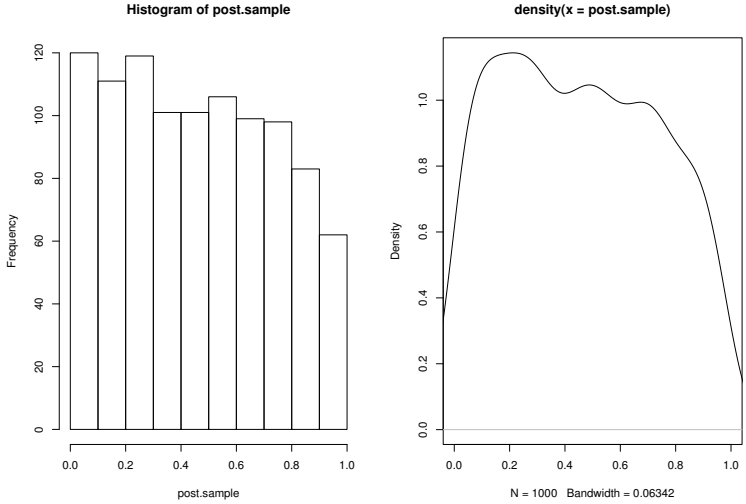

Figure 2: Histogram and Density for Problem 2b

```
> par(mfrow=c(1,2))
```

```
> hist(post.sample, xlim=c(0, 1))
> plot(density(post.sample), type="l", xlim=c(0, 1))
```

c Posterior predictive

To sample from the posterior predictive, first sample 1000 $\theta's$ from the posterior distribution, then use those $\theta's$ to generate y values. See Figure 3.

```
> index2=sample(1:grid.size, grid.size, replace=T, prob=post)
> post.sample2=theta[index]
> predsamp=rcauchy(1000, post.sample2, 1)
> par(mfrow=c(1, 2))
> hist(predsamp)
> plot(density(predsamp), type='l')
```

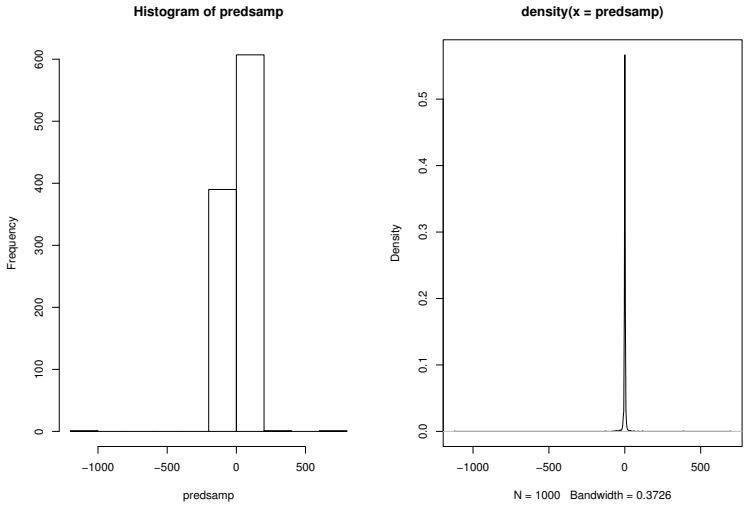

Figure 3: Histogram and Density for Problem 2c

d Rejection Sampling Method

I will use the simplification described on page 7 of the notes from January 17th. The code is below, and the graphs are in Figure 4. The results are very similar to part b. This is good since they are both ways to sample from the posterior.

```
> negloglike=function(theta){
+ return(-sum(log(dcauchy(y, theta, 1))))}
> mle=nlm(negloglike, 0.5)$estimate
> post.sample=rep(NA, 1000)
> for(i in 1:1000){
+
+   repeat{
+     theta=runif(1, 0, 1)
+     U=runif(1, 0, 1)
```

```
+      if (U <= prod(dcauchy(y, theta, 1))/prod(dcauchy(y, mle, 1))){
+        break
+      }
+     }
+    post.sample[i]=theta
+ }
> hist(post.sample)
> plot(density(post.sample), xlim=c(0,1))
```

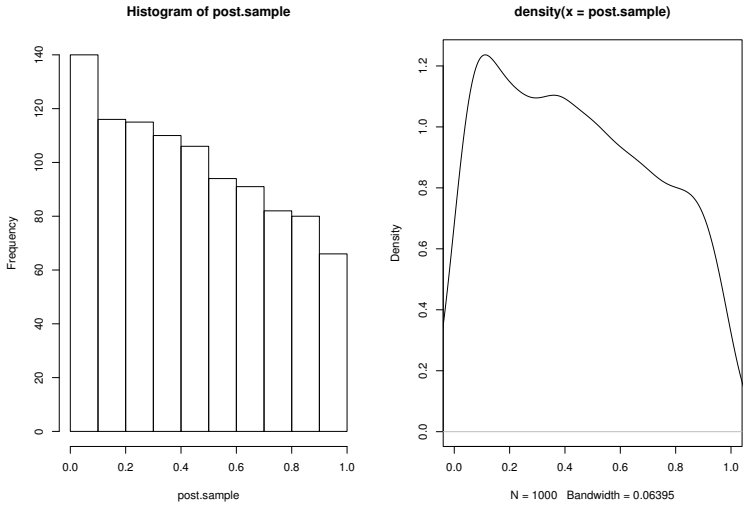

Figure 4: Histogram and Density for Problem 2d

e Different Data Set Lengths

The code in part a) was redone for the different lengths of data. The graph is in Figure 5.

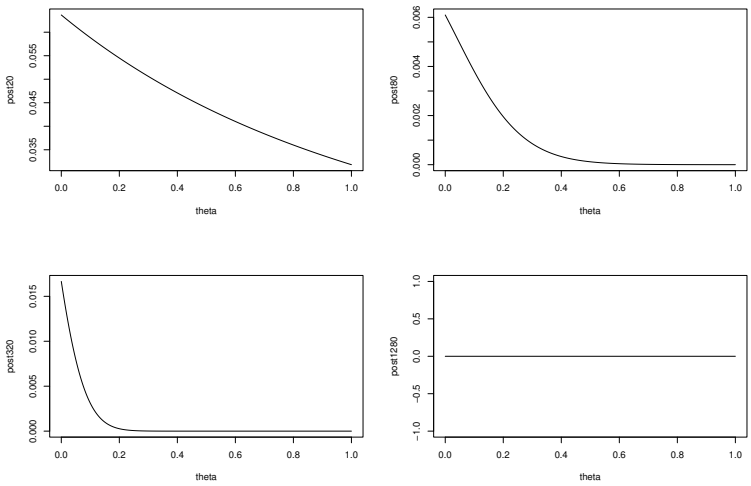

Figure 5: Posterior Samples for Problem 2e

As the sample size grows, the posterior density becomes more and more concentrated around zero. The problem with the length 1280 is due to rounding errors when calculating the likelihood ie. the likelihood is smaller than the smallest number representable by a floating point number in the computer. This problem can be avoiding by taking logs, which we do not do here.

Problem 3

a Prior Density

We are given that $p(\sigma) \propto \frac{1}{\sigma}$. To get $p(\sigma^2)$ use a transformation. Let $X = \sigma$ and $Y = \sigma^2 = g(X)$. Then we know that $f_Y(y) = f_X(g^{-1}(y))|\frac{\partial}{\partial y}g^{-1}(y)|$. Here $g(x) = x^2$ and $g^{-1}(y) = \sqrt{y}$. Thus, $f_Y(y) = f_X(\sqrt{y})|\frac{1}{2}y^{-\frac{1}{2}} \propto \frac{1}{\sigma^2}$

b HPD's

I will assume that the variable ν represents the data. I want to find both $p(\sigma|\nu)$ and $p(\sigma^2|\nu)$. We are given that $\frac{n\nu}{\sigma^2}|\sigma^2 \sim \chi^2_n$.

$$p(\frac{n\nu}{\sigma^2}|\sigma^2) \quad = \quad \frac{2^{-n/2}}{\Gamma(n/2)}(\frac{n\nu}{\sigma^2})^{n/2-1}e^{\frac{n\nu}{2\sigma^2}}.$$

Let ν represent the data. I first find the posterior of the data using a transformation of variables. Realize that $\theta = \frac{n\nu}{\sigma^2}$ and $\nu = \frac{\theta\sigma^2}{n}$. Thus, $g(\theta) = \theta\sigma^2/n$ and $g^{-1}(\theta) = \frac{n\nu}{\sigma^2}$.

$$\begin{aligned} f_\nu(\nu) &= f_\theta(g^{-1}(\theta))|\frac{\partial}{\partial\theta}g^{-1}(\theta)| \\ &\propto (\frac{n\nu}{\sigma^2})^{(\frac{n}{2}-1)}e^{\frac{n\nu}{2\sigma^2}}|\frac{n}{\sigma^2}| \end{aligned}$$

Note that this is a kernal of a Gamma($\alpha = \frac{n}{2}, \beta = \frac{n}{2\sigma^2}$). Putting this together with the prior, we get the posterior of σ^2 given ν.

$$p(\sigma^2|\nu) \propto (\sigma^2)^{(\frac{-n}{2}-1)}e^{\frac{-n\nu}{2(\sigma^2)}}$$

Using another transformation of variables, we can computer $p(\sigma|\nu)$.

$$p(\sigma|\nu) \quad \propto \quad \sigma^{-n-1}e^{\frac{n\nu}{2\sigma^2}}$$

We know that a characteristic of the HPD's is that the density of the function at the endpoints of the interval match each other. Let (a_2, b_2) be an HPD interval for the posterior of σ^2. Put $a_1 = \sqrt{a_2}, b_1 = \sqrt{b_2}$ and assume it is an HPD. We will show that $a_1 = b_1$, and thus, for this case, *no* HPD interval for σ^2 is obtained by squaring a HPD interval for σ. In particular, the 95% HPD interval for σ^2 is not the square of the 95% HPD interval for σ when using the given priors.

$$a_2^{\frac{-n}{2}-1}e^{\frac{n\nu}{2a_2}} = b_2^{\frac{-n}{2}-1}e^{\frac{n\nu}{2b_2}} \text{ and } \sqrt{a_2}^{-n-1}e^{\frac{n\nu}{2a_2}} = \sqrt{b_2}^{-n-1}e^{\frac{n\nu}{2b_2}}$$

Note here that there are two equations and two unknowns. Take the log of the equations and solve to get that $\frac{1}{2}log(a_2) = \frac{1}{2}log(b_2)$. Thus, under the assumption that t $a_1 = \sqrt{a_2}, b_1 = \sqrt{b_2}$, we see that $a_1 = b_1$, or, the interval is null.

Note that here we also could have used a simulation. I used the grid method to sample 10,000 elements from the two distributions, with $n = 10$ and $\nu = 0.001$. I then found the confidence intervals.

```
> sigma=seq(0.01, 10, length=1000)
> post.sig2
function(sig, nu, n){
return(sig^(-n/2-1)*exp(n*nu/(2*sig)))}
> post.sig
function(sig, nu, n){
return(sig^(-n-1)*exp(-n*nu/(2*sig^2)))}
> sample2=sample(sigma, 10000, replace=T, prob=post.sig2(sigma, nu, n))
> sample=sample(sigma, 10000, replace=T, prob=post.sig(sigma, nu, n))
> n
[1] 10
> nu
[1] 0.001
> sort(sample)[c(25, 975)]
[1] 0.02 0.03
> sort(sample2)[c(25, 975)]
[1] 0.01 0.01
```

Here we see that the CI from sample (i.e. the posterior of σ) is not the square root of the CI of sample2 (i.e. the posterior of σ^2).

Problem 4

a Prior for θ

We are given that $\lambda = \log(\frac{\theta}{1-\theta})$ and that $p(\lambda) \propto 1$. We need to do a transformation of variable to get the distribution of θ.

$$\begin{aligned} p_\theta(\theta) &\propto p_\lambda(\log(\frac{\theta}{1-\theta})) * \frac{\partial}{\partial\theta}\log(\frac{\theta}{1-\theta}) \\ &= 1 * (\frac{1-\theta}{\theta}) * \frac{(1-\theta)*1-\theta*(-1)}{(1-\theta)^2} \\ &= \theta^{-1}(1-\theta)^{-1} \end{aligned}$$

Note that this is a kernel of a beta distribution as both parameters reach zero in the limit. Note that this is not a proper distribution. Technically α and β need to be larger than zero for this distribution to be proper.

b Improper Posterior

First compute the posterior.

$$\begin{aligned} p(\theta|y) &\propto \theta^y(1-\theta)^{n-y}\theta^{-1}(1-\theta)^{-1} \\ &= \theta^{y-1}(1-\theta)^{n-y-1} \end{aligned}$$

This is a beta distribution with parameters $\alpha = y - 1$ and $\beta = n - y - 1$. For the distribution to be proper, both α and β need to be positive. This does not happen if $y = 0$ or if $y = n$ since the function x^a is not integrable on the interval $(0, c)$ if $a < -1$.

Problem 5

a Jeffreys' Prior

Recall Jeffreys' prior is $\sqrt{I(\theta)}$ and $I(\theta) = -E(\frac{\partial^2}{\partial^2\theta} log(p(y|\theta)))$.

$$\begin{aligned} I(\theta) &= -E(\frac{\partial^2}{\partial^2\theta}\log(\frac{e^{-\theta}\theta^x}{x!})) \\ &= -E(\frac{\partial^2}{\partial^2\theta} - \theta + x\log(\theta) - \log(x!)) \\ &= -E(-\frac{x}{\theta^2}) \\ &= \theta^{-1} \end{aligned}$$

Thus, Jeffrey's prior is $\theta^{-\frac{1}{2}}$, which is the limit of an Gamma with $\alpha = 0.5$ as $\beta \to 0$.

b Posterior

The posterior is proportional to the prior times the likelihood.

$$\begin{aligned} p(\theta|y) &\propto \theta^{-\frac{1}{2}} e^{-n\theta}\theta^{\Sigma x} \\ &= e^{-n\theta}\theta^{\Sigma x - \frac{1}{2}} \end{aligned}$$

Note that this is a Gamma with parameters $\alpha = \Sigma x + 0.5$ and $\beta = n$.

c Jeffreys as limiting conjugate prior.

The conjugate of a Poisson is a Gamma. The kernel of a Gamma is $\theta^{\alpha-1}e^{\beta\theta}$. We can see that if $\alpha = 0.5$ and $\beta \to 0$, then it is the Jeffrey's prior. Another way to see this is that the variance of the gamma is $\frac{\alpha}{\beta}$. As β approaches zero, the variance increases, providing less and less information.

Problem 6

a Fatal Accidents – Poisson(θ)

I will set a prior for θ to be Gamma, as it is the conjugate prior. I will allow the paramters to be $\alpha = 0.5$, $\beta = 0.01$ as it is near the non-informative Jeffrey's prior shown in the last problem. Thus, the posterior distribution can be calculated.

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &= e^{-n\theta}\theta^{\sum y_i}\theta^{-\frac{1}{2}}e^{-0.01\theta} \\ &= e^{-\theta(n+0.01)}\theta^{\sum y_i - \frac{1}{2}} \end{aligned}$$

Note that this is a Gamma($\alpha = \sum y_i + \frac{1}{2}, \beta = n + 0.01$).

I get the 95% predictive interval using simulation. I found the interval to be [14, 35].

```
> theta=rgamma(1000, sum(y)+0.5, length(y)+0.01)
> ypred=rpois(1000, theta)
> sort(ypred)[c(25, 975)]
[1] 14 35
```

Problem 7

a Rejection Sample from Model II

It is not possible to use rejection sampling from the posterior sampling from Model II using the prior as the proposal distribution. This is because the prior distribution is improper.

b Rejection Sampling for Model II

To pick a proper prior, there are a number of options. I will pick uniform distributions as they are non-informative on μ and $\log(\sigma^2)$. Note that since μ is a measurement, it cannot be less than zero. I will pick an upper limit of 50, which is about 4 times larger than the largest data point (arbitrary). I choose the lower limit of $\log(\sigma^2)$ to be close to zero (i.e. log(0.01)), and I let the upper limit be log(100) (arbitrary). I will use the following priors.

$$\begin{aligned} \mu &\sim \text{Uniform}(0, 50) \\ \log(\sigma^2) &\sim \text{Uniform}(\log(0.01), \log(100) \end{aligned}$$

Using these priors, the analytic expression of the posterior is as follows, where $\Phi(y|\mu, \sigma^2)$ is the cdf of the normal distribution with mean μ and variance σ^2 evaluated at the point y.

$$p(\mu, \sigma^2|y) \propto p(\mu, \sigma^2) \prod_{i=1}^{5} (\Phi(y_i + 0.5|\mu, \sigma^2) - \Phi(y_i - 0.5|\mu, \sigma^2))$$

A graph of the posterior densities of μ and σ^2 are in Figure 6. I used the following code to obtain the samples.

```
> y=c(10, 10, 12, 11, 9)
> neglike=function(musigma2){
+ return(-prod(pnorm(y+0.5, musigma2[1], sqrt(musigma2[2]))
   -pnorm(y-0.5, musigma2[1], sqrt(musigma2[2]))))}
> mle=nlm(neglike, c(mean(y), var(y)))$estimate
> mle
[1] 10.4004165  0.9536996
> mu.postsamp=rep(NA, 1000)
> sigma2.postsamp=rep(NA, 1000)
> for (i in 1:1000){
+   repeat{
+ mu = runif(1, 0, 50)
+ sig=runif(1, log(0.01), log(100))
+ sig2=exp(sig)
```

```
+ U=runif(1)
+ if (U <= -neglike(c(mu, sig2))/-neglike(mle)) break
+ }
+ mu.postsamp[i]=mu
+ sigma2.postsamp[i]=sig2
+ }
> mean(mu.postsamp)
[1] 10.39239
> mean(sigma2.postsamp)
[1] 2.325363
> var(mu.postsamp)
[1] 0.5306943
> var(sigma2.postsamp)
[1] 10.82086
```

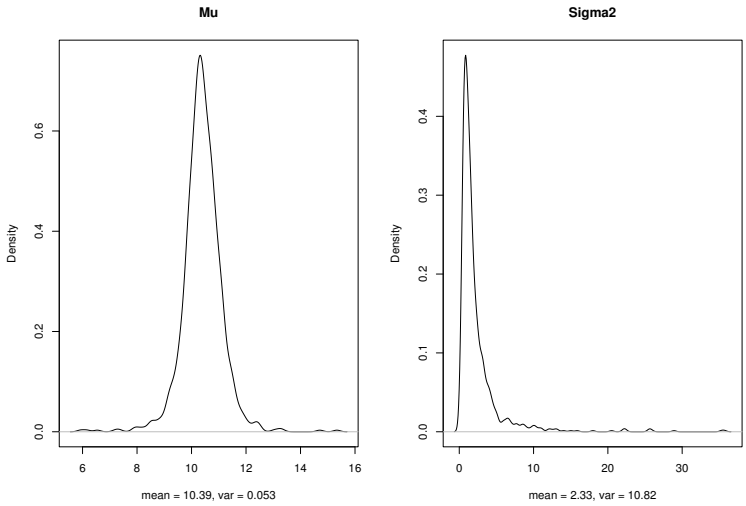

Figure 6: Posterior of mu and sigma2 for 1b

Comparing this to what was done in class, (recall that the mean and stdev for μ was (10.372, 0.716) and the mean and stdev for σ^2 was (2.76 and 4.18). These are very close to my values for μ of (10.39, 0.728) and for σ^2 of (2.32, 3.28). the values of μ are closer than the values of σ^2, however overall they are still pretty close (the confidence intervals have significant overlap).

c Data Augmentation

(a) Iterated Expected Values

We know using iterated expected values that $E_{(\mu,\sigma^2)}(E(I_{Z\leq t})|y,\mu,\sigma^2) = E_{\mu,\sigma^2}(Pr(Z \leq t|y,\mu,\sigma^2)) = Pr(Z \leq t|y)$. Another way of writing this is below.

$$p(Z \leq t|y) \quad = \quad \int\int p(Z \leq t|y,\mu,\sigma^2)p(\mu,\sigma^2|y)d\mu d\sigma^2$$

¿From this we can get $p(Z|y)$.

$$\begin{aligned} p(Z|y) &= \frac{\partial}{\partial z} F_Z(z) \\ &= \int\int p(z|y,\mu,\sigma^2) p(\mu,\sigma^2|y) d\mu d\sigma^2 \end{aligned}$$

(b) Summarize Posterior of $(z_1 - z_2)^2$

I used the following code to generate 1000 draws of $z_1, ..., z_5$.

```
> zvals=matrix(NA, nrow=5, ncol=1000)
> for (i in 1:5){
+ zvals[i,]=rnorm(1000, mu.postsamp, sqrt(sigma2.postsamp))}
> z1minusz2=(zvals[1,]-zvals[2,])^2
> mean(z1minusz2)
[1] 4.626152
> sqrt(var(z1minusz2))
[1] 15.38466
```

Figure 7 contains a histogram of the values of $(z_1 - z_2)^2$. As we can see, the standard deviation is quite large, and the distribution has a large right tail.

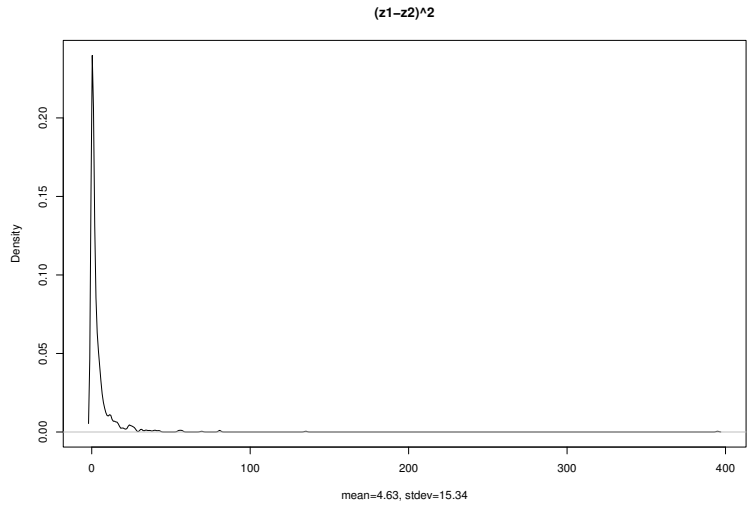

Figure 7: Posterior of $(z1 - z1)^2$ for 1cii

d Posterior Distributions of Model I and Model II

I will plot the posterior density from model I using contour. I used a small grid due to time considerations. The posteriors are in Figures 8.

```
> z1=matrix(NA, 100, 100)
> like2=function(mu, sigma){
```

```
+ return(prod(dnorm(y, mu, sigma)))}
> dens1=function(mu, sigma){
+ return(1/sigma^2*like2(mu, sigma))}
> for (i in 1:100){
+ for(j in 1:100){
+   z[i,j]=dens1(mu.grid[i], sigma.grid[j])
+ }
+ }
```

I will plot the posterior density from model II using contour. I used a small grid due to time considerations. The posteriors are in Figures 8.

```
> z=matrix(NA, 100, 100)
> mu.grid=seq(0, 50, length=100)
> sigma.grid=seq(0.01, 100, length=100)
> dens=function(mu, sigma){
+ return(1/50*1/99.99*-1*neglike(c(mu, sigma^2)))}
> for (i in 1:100){
+ for(j in 1:100){
+   z[i,j]=dens(mu.grid[i], sigma.grid[j])
+ }}
> contour(mu.grid, sigma.grid, z, nlevels=10, xlim=c(5, 15), ylim=c(0, 4), mair
```

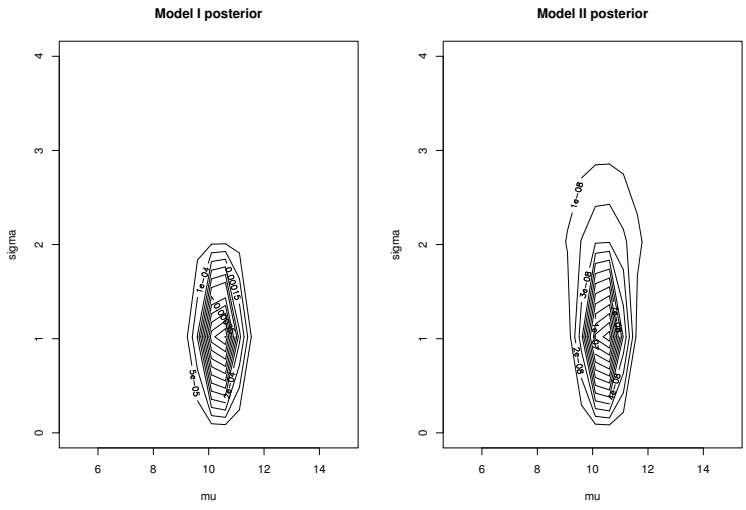

Figure 8: Posterior for Model II for 1d

We can see that the density for model 2 has a larger range for σ and looks similar for μ.

Problem 8: Gelman pg. 99 #10

We want to show that $(s_1^2/s_2^2)/(\sigma_1^2/\sigma_2^2) \sim F(n_1 - 1, n_2 - 1)$ under the conditions listed in the book. Equivalently, we may show that the posterior for $(n_i - 1)s_i^2/\sigma_i^2 \sim \chi^2_{n_i-1}$ for each i and that the posterior

for σ_1^2 and σ_2^2 are independent. This follows from the definition of the F distribution as the distribution of the $(X/n_y)/(Y/n_x)$ where X and Y are independent χ^2 random variables with degrees of freedom n_x, n_y respectively.

Let $\vec{y_i} = (y_{i,1}, ..., y_{i,n_i})$ ie. the vector of observations belonging to the i^{th} group. In case there is any confusion, we explicitly assume $\vec{y_1}|(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ is independent of $\vec{y_1}|(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$.

$$\begin{aligned}
p(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2|\vec{y_1}, \vec{y_2}) &\propto \\
p(\vec{y_1}, \vec{y_2}|\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)p(\mu_1, \sigma_1^2)p(\mu_2, \sigma_2^2) &\propto \text{(By Bayes rules and independence of priors)} \\
p(\vec{y_1}|\mu_1, \sigma_1^2)p(\mu_1, \sigma_1^2)p(\vec{y_2}|\mu_2, \sigma_2^2)p(\mu_2, \sigma_2^2) &\propto \begin{pmatrix}\text{By our independence assumption and} \\ \text{since } \vec{y_i}\text{'s distribution only depends on} \\ \mu_i, \sigma_i\end{pmatrix}
\end{aligned}$$

Since the posterior density factorizes into a function of σ_1^2 times a function of σ_2^2, the random variables σ_1^2 and σ_2^2 are independent in the posterior distribution. From the notes, we know the posterior distribution of σ_i^2 is $Inv-_{\chi^2}(n_i - 1, s_i^2)$ for each i, or equivalently, $(n_i - 1)s_i^2/\sigma_i^2 \sim \chi^2_{n_i-1}$, which is what we needed to show.

It is worth noting that the independence result in this problem generalizes easily. Given $\vec{y_1}$ and $\vec{y_2}$ whose distributions depend only on parameters $\vec{\theta_1}$ and $\vec{\theta_2}$ respectively and given independent priors for the parameters, if $\vec{y_1}$ and $\vec{y_2}$ are *conditionally* independent given the parameters, then $\vec{\theta_1}$ and $\vec{\theta_2}$ are independent in the posterior distribution (or conditionally independent given the data). Note that it is *not* true that the elements within $\vec{y_1}$ are (unconditionally) independent. They are dependent through the parameter $\vec{\theta_1}$. They are, however, conditionally independent given $\vec{\theta_1}$. The notion of observations being conditionally independent given the parameters is an important notion in the subjectivist Bayesian "philosophy." It moves away from the frequentist notion of pure iid random variables and is tied with an alternative concept called exchangeability which we won't go into further here.