ISyE 6420: Homework 4

Laura Schultz

10/7/2019

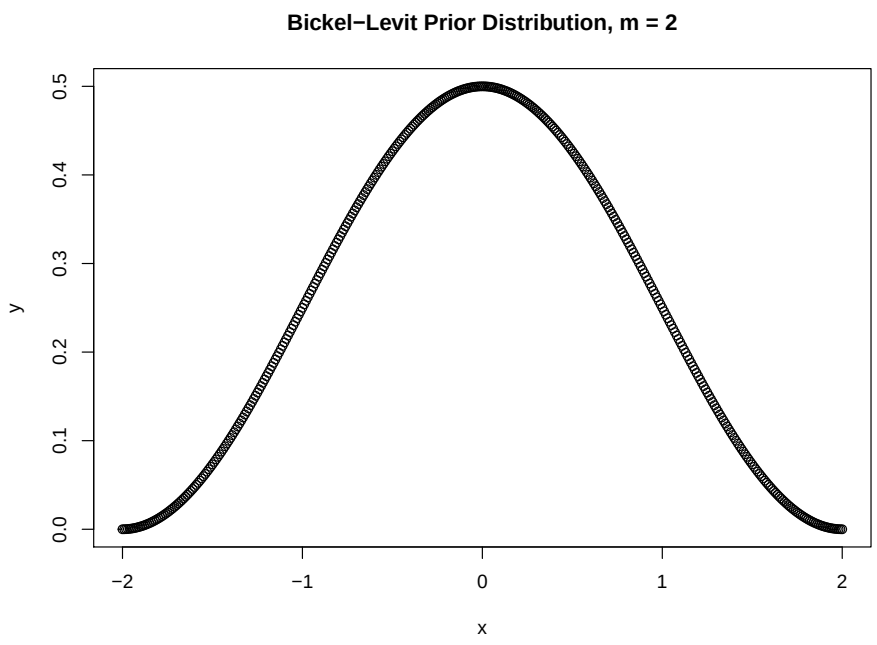
Problem 1: Metropolis: The Bounded Normal Mean.

Suppose that we have information that the normal mean θ is bounded between $-m$ and m, for some known number m. In this case, it is natural to elicit a prior on θ with the support on the interval $[-m, m]$. A prior with interesting theoretical properties supported on $[-m, m]$ is the Bickel-Levit prior,

$$\pi(\theta) = \frac{1}{m}\cos^2\left(\frac{\pi\theta}{2m}\right), \ -m \leq \theta \leq m.$$

Here, we will assume that $m = 2$. Out of curiosity, I plotted $f(x) = \frac{1}{2}\cos^2\left(\frac{\pi x}{4}\right)$ for $-2 \leq x \leq 2$.

```
x <- seq(from=-2, to=2, by=0.01)
y = 0.5 * cos(pi*x/4)^2
plot(x,y, main="Bickel-Levit Prior Distribution, m = 2")
```



As I expected, this function has the appearance of a truncated normal distribution that is centered at 0 and bounded between -2 and 2.

Suppose that we observe a sample $[-2, -3, 4, -7, 0, 4]$ from a normal distribution

$$f(y|\theta) \propto \sqrt{\tau}\exp\left\{-\frac{\tau}{2}(y-\theta)^2\right\},$$

with a known precision $\tau = \frac{1}{\sigma^2} = 1/4$ (shape $\sigma^2 = \frac{1}{\tau} = 4$) and unknown mean θ.

Using the given data and $\tau = \frac{1}{4}$, we can calculate the likelihood as

$$\mathcal{L}(\theta|y_1, \dots y_6) = \prod_{i=1}^{6} f(y_i|\theta) \propto \left(\frac{1}{2}\right)^6 \exp\left\{-\frac{\frac{1}{4}}{2}\sum_{i=1}^{6}(y_i-\theta)^2\right\} =$$

$$\frac{1}{64}\exp\left\{-\frac{1}{8}\sum_{i=1}^{6}(y_i-\theta)^2\right\}.$$

Let $Y \sim N(\theta, 4)$ with Bickel-Levit prior

$$\pi(\theta) = \frac{1}{2}\cos^2\left(\frac{\pi\theta}{4}\right), \ -2 \le \theta \le 2,$$

such that the posterior is

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta) \propto \frac{1}{64}\exp\left\{-\frac{1}{8}\sum_{i=1}^{6}(y_i-\theta)^2\right\} \times \frac{1}{2}\cos^2\left(\frac{\pi\theta}{4}\right) \propto \frac{\cos^2\left(\frac{\pi\theta}{4}\right)}{128\exp\left\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta)^2\right\}}.$$

This result does not suggest an explicit posterior, nor is it feasible to integrate to find the marginal distribution. Hence, I will utilize a Metropolis algorithm to sample from the posterior distribution of θ, and then use my simulation results to calculate the Bayes estimator $\hat{\theta}_B$ and generate a 95% equitailed credible set.

Let the target density function be

$$g(\theta) = \pi(\theta|y) \propto \frac{\cos^2\left(\frac{\pi\theta}{4}\right)}{128\exp\left\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta)^2\right\}},$$

which is the posterior that I derived above. The independence candidate density will be $\mathcal{U}(-2, 2)$, such that

$$q(\theta'|\theta_n) = q(\theta_n|\theta') = \frac{1}{2-(-2)} = \frac{1}{4}, \text{ for } -2 \le y \le 2 \text{ (otherwise 0)}.$$

Then, I can define the acceptance probability as

$$\rho = \min\left\{1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)}\right\} = \min\left\{1, \frac{\pi(\theta')}{\pi(\theta)}\right\} = \min\left\{1, \frac{\cos^2\left(\frac{\pi\theta'}{4}\right)}{128\exp\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta')^2\}} \times \frac{128\exp\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta)^2\}}{\cos^2\left(\frac{\pi\theta}{4}\right)}\right\}$$

$$= \min\left\{1, \frac{\cos^2\left(\frac{\pi\theta'}{4}\right)\exp\left\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta)^2\right\}}{\cos^2\left(\frac{\pi\theta}{4}\right)\exp\left\{\frac{1}{8}\cdot\sum_{i=1}^{6}(y_i-\theta')^2\right\}}\right\}.$$

(a) Simulate 10,000 observations from the posterior, after discarding the first 500 observations (burn-in), and plot the histogram of the posterior.

I used R to implement the following Metropolis algorithm for generating values from the posterior:

- Step 1: Start with (arbitrary) $\theta_0 = 0$
- Step 2: At stage i, generate proposal θ' from $\mathcal{U}(-2, 2)$.

- Step 3: Let

$$\theta_{i+1} = \begin{cases} \theta', & \text{w.p. } \rho \\ \theta_i, & \text{w.p. } 1-\rho \end{cases}$$

  Generate $U \sim \mathcal{U}(0,1)$, and accept the proposed θ' as θ_{i+1} if $U \leq \rho$. Otherwise, retain θ_i as θ_{i+1}.

- Step 4: Go back to step 2 and repeat until 10,500 observations have been simulated.

```
# Clear the environment
rm(list=ls())

#Load tictoc package for timing
require(tictoc)
```

```
## Loading required package: tictoc
```

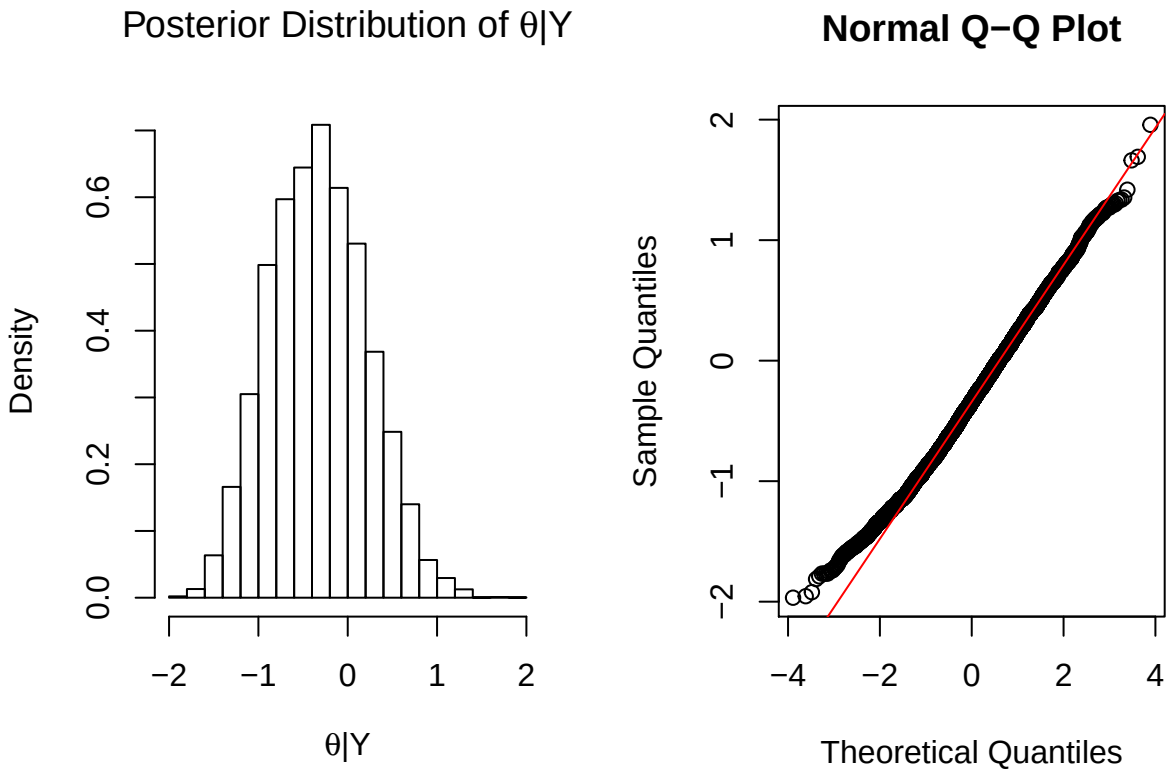
```
# Observed data
y = c(-2,-3,4,-7,0,4)

# Set seed for reproducible results
set.seed(66)

# Number of trials and burn-in
m = 10000
burnin = 500
m = m + burnin

# Initialize theta
theta = 0  #I started in the middle of the U(-2,2) distribution.

# Initialize vector to store simulated thetas
thetas = theta

tic() # Start timer
# Generate proposal
 for (i in 1:m){
   theta_prop = runif(1,min=-2,max=2)
   SS_theta = sum((y - theta)^2) # Used to calculate the numerator of ratio r
   SS_theta_prime = sum((y - theta_prop)^2) # Used to calculate the denominator of r
   r= (cos(pi*theta_prop/4)^2)*exp(1/8*SS_theta)/
     (cos(pi*theta/4)^2*exp(1/8*SS_theta_prime))

   rho = min(1,r)
   if (runif(1) <= rho){
     theta = theta_prop
   }

   thetas = c(thetas, theta)
 }

toc() # See how long the simulation took
```

```
## 0.474 sec elapsed
```

```r
# Discard trials from burn-in
thetas = thetas[(501:m)]
length(thetas)  # Check that I have kept 10,000 simulated thetas
```

```
## [1] 10000
```

```r
# Plot relative frequency histogram and normal Q-Q plot of simulated thetas
par(mfrow= c(1,2))
hist(thetas, freq = FALSE, right=FALSE,
    main=expression(paste("Posterior Distribution of ",
    theta,"|Y")), xlab=expression(paste(theta,"|Y")))
qqnorm(thetas, main="Normal Q-Q Plot")
qqline(thetas, col = 2)
```




```r
# Look at summary for simulated posterior distribution
summary(thetas)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.9669 -0.7252 -0.3397 -0.3297  0.0417  1.9571
```

My simulated posterior distribution of $\theta|Y$ is nearly bell-shaped (but very slightly right-skewed) and centered at a mean of -0.3297. The thetas in this distribution ranged from -1.9669 to 1.9571, which is aligned with my expectation that the posterior distribution would be bounded on $[-2, 2]$. The slight concave-up curvature visible in the normal Q-Q plot supports my impression that the simulated posterior distribution isn't quite normal.

(b) Find Bayes estimator of θ and 95% credible set based on the simulated observations.

```
# Calculate Bayes estimator (and variance)
mean(thetas)
```

```
## [1] -0.3296525
```

```
var(thetas)
```

```
## [1] 0.296071
```

```
# Find 95% credible set
quantile(thetas, c(0.025,0.975))
```

```
##       2.5%      97.5%
## -1.3178813  0.7508018
```

The Bayesian estimator is $\hat{\theta}_B = -0.3297$, which is the mean of my simulated posterior distribution, and the 95% equitailed credible set is $\theta \in [-1.3179, 0.7508]$.

Problem 2: Gibbs Sampler and High/Low Protein Diet in Rats.

Armitage and Berry (1994) described a study in which 19 female rats were randomly assigned to consume either a high-protein ($n_1 = 12$) or a low-protein ($n_2 = 7$) diet and then had their weight gain recorded after 28 to 84 days on the assigned diet. The following table shows the observed weight gains.

High Protein	Low Protein
134	70
146	118
104	101
119	85
124	107
161	132
107	94
83	
113	
129	
97	
123	

I will use the Gibbs sampler to conduct a Bayesian test of H_0 : Rats fed a high-protein gain more weight than those fed a low-protein diet ($\theta_1 > \theta_2$) vs. H_1 : Rats fed a low-protein diet gain as much or more weight than those fed a high-protein diet ($\theta_2 \geq \theta_1$).

Assume that the weight gains for rats fed the high-protein diet, y_{1i}, come from the normal distribution $\mathcal{N}(\theta_1, \frac{1}{\tau_1})$ with density

$$f(y_{1i}|\theta_1,\tau_1) \propto \tau_1^{1/2} \exp\left\{ -\frac{\tau_1}{2}(y_{1i}-\theta_1)^2 \right\},\ i=1,\ldots,12.$$

Likewise, assume that the weight gains for rats fed the low-protein diet, $y_{2,i}$, come from the normal distribution $\mathcal{N}(\theta_2, \frac{1}{\tau_2})$ with density

$$f(y_{2i}|\theta_2,\tau_2) \propto \tau_2^{1/2} \exp\left\{ -\frac{\tau_2}{2}(y_{2i}-\theta_2)^2 \right\},\ i=1,\ldots,7.$$

Assume that θ_1 and θ_2 have normal priors, $\mathcal{N}(\theta_{10}, \frac{1}{\tau_{10}})$ and $\mathcal{N}(\theta_{20}, \frac{1}{\tau_{20}})$, respectively. Let the prior means be $\theta_{10} = \theta_{20} = 110$ and the prior precisions be $\tau_{10} = \tau_{20} = \frac{1}{100}$.

Assume that τ_1 and τ_2 have gamma priors, $\mathcal{G}a(a_1, b_1)$ and $\mathcal{G}a(a_2, b_2)$, with shapes $a_1 = a_2 = 0.01$ and rates $b_1 = b_2 = 4$.

Putting it all together, the weight gain data, y_{ij} for $i = 1,\ldots,n_j$ and $j = 1, 2$ are assumed to be independently normally distributed within each of the $J = 2$ groups with means θ_j and variances $\frac{1}{\tau_j}$, where $n_1 = 12$ and $n_2 = 7$.

The likelihood for θ_1 is

$$\mathcal{L}(\theta_1|y_{11},\ldots y_{112}) = \prod_{i=1}^{12} f(y_{1i}|\theta_1,\tau_1) \propto \tau_1^6 \exp\left\{ -\frac{\tau_1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2 \right\},$$

and the joint distribution for group 1 (rats fed a high-protein diet) is

$$f(y_{1i},\theta_1,\tau_1) \propto \tau_1^6 \exp\left\{ -\frac{\tau_1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2 \right\} \pi(\theta_1)\pi(\tau_1).$$

From Brani's Gibbs Sampler handout, I can deduce that the full conditional for θ_1 is

$$\pi(\theta_1|\theta_2,\tau_1,\tau_2,\mathbf{y_1}) \propto \exp\left\{ -\frac{1}{2}(\tau_{10}+n_1\tau_1)\left(\theta_1 - \frac{\tau_1\sum_{i=1}^{n_1} y_{1i} + \tau_{10}\theta_{10}}{\tau_{10}+n_1\tau_1}\right)^2 \right\},$$

which is a kernel of the normal $\mathcal{N}\left(\frac{\tau_1\sum y_{1i}+\tau_{10}\theta_{10}}{\tau_{10}+n_1\tau_1}, \frac{1}{\tau_{10}+n_1\tau_1}\right)$ distribution. Given that $\theta_{10} = 110$ and $\tau_{10} = 0.01$ are fixed hyperparameters and $n_1 = 12$, the full conditional for θ_1 becomes

$$\theta_1|\theta_2,\tau_1,\tau_2,\mathbf{y_1} \sim \mathcal{N}\left(\frac{\tau_1\sum y_{1i} + 1.1}{0.01+12\tau_1}, \frac{1}{0.01+12\tau_1}\right).$$

Similarly, I deduce that the full conditional for θ_2 is

$$\pi(\theta_2|\theta_1,\tau_1,\tau_2,\mathbf{y_1}) \propto \exp\left\{ -\frac{1}{2}(\tau_{20}+n_2\tau_2)\left(\theta_2 - \frac{\tau_2\sum_{i=1}^{n_2} y_{2i} + \tau_{20}\theta_{20}}{\tau_{20}+n_2\tau_2}\right)^2 \right\}.$$

Given hyperparameters $\theta_{20} = 110$ and $\tau_{20} = 0.01$ and $n_2 = 7$, the full conditional becomes

$$\pi(\theta_2|\theta_1,\tau_1,\tau_2,\mathbf{y_2}) \propto \exp\left\{ -\frac{1}{2}(0.01+7\tau_2)\left(\theta_2 - \frac{\tau_2\sum_{i=1}^{7} y_{2i} + 1.1}{0.01+7\tau_2}\right)^2 \right\}.$$

Hence, the full conditional for θ_2 is

$$\theta_2|\theta_1,\tau_1,\tau_2,\mathbf{y_2} \sim \mathcal{N}\left(\frac{\tau_2\sum y_{2i}+1.1}{0.01+7\tau_2}, \frac{1}{0.01+7\tau_2}\right).$$

Referring again to the Gibbs sampler handout, I have deduced that the full conditional for τ_1 is

$$\pi(\tau_1|\theta_1,\theta_2,\tau_2,\mathbf{y_1}) \propto \tau_1^6 \exp\left\{-\frac{\tau_1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2\right\}\tau_1^{a_1-1}\exp\{-b_1\tau_1\} =$$

$$\tau_1^{6+a_1-1}\exp\left\{-\tau_1\left[b_1+\frac{1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2\right]\right\},$$

which is a kernel of the gamma $\mathcal{G}a(a_1+6,\, b_1+\frac{1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2)$ distribution. Given the hyperparameters $a_1=0.01$ and $b_1=4$, the kernel for the full conditional of τ_1 yields

$$\tau_1|\theta_1,\theta_2,\tau_2,\mathbf{y_1} \sim \mathcal{G}a(6.01,\, 4+\frac{1}{2}\sum_{i=1}^{12}(y_{1i}-\theta_1)^2).$$

Similarly, I have deduced that the full conditional for τ_2 is

$$\pi(\tau_2|\theta_1,\theta_2,\tau_1,\mathbf{y_2}) \propto \tau_2^{7/2} \exp\left\{-\frac{\tau_2}{2}\sum_{i=1}^{7}(y_{2i}-\theta_2)^2\right\}\tau_2^{a_2-1}\exp\{-b_2\tau_2\} =$$

$$\tau_2^{7/2+a_2-1}\exp\left\{-\tau_2\left[b_2+\frac{1}{2}\sum_{i=1}^{7}(y_{2i}-\theta_2)^2\right]\right\}.$$

Given the hyperparameters $a_2=0.01$ and $b_2=4$, the full conditional of τ_2 can be expressed as

$$\tau_2|\theta_1,\theta_2,\tau_1,\mathbf{y_2} \sim \mathcal{G}a(3.51,\, 4+\frac{1}{2}\sum_{i=1}^{7}(y_{2i}-\theta_2)^2).$$

(a) Construct a Gibbs sampler that will sample θ_1, τ_1, θ_2, and τ_2 from their posteriors.

```
# Clear the environment
rm(list=ls())

# Data for rats fed high-protein diet (Group 1)
y_1 <- c(134, 146, 104, 119, 124, 161, 107, 83, 113, 129, 97, 123)
sum_y1 <- sum(y_1)
n_1 <- length(y_1)
sum_y1/n_1 # Sample mean for group 1
```

```
## [1] 120
```

```
# Data for rats fed low-protein diet (Group 2)
y_2 <- c(70,118, 101, 85, 107, 132, 94)
sum_y2 <- sum(y_2)
n_2 <- length(y_2)
sum_y2/n_2 # Sample mean for group 2
```

```
## [1] 101
```

```r
# Note: I  hardcoded the hyperparameters into the full-conditional pdfs
# theta_10 = theta_20 = 110 and tau_10 = tau_20 = 0.01
# a_1 = a_2 = 0.01 and b_1 = b_2 = 4.
#I also hardcoded the sample sizes, n_1 = 12 and n_2 = 7

set.seed(36)

# Number of trials and burn-in
NN = 10000
burnin=500
NN = NN + burnin

# Initialize vectors
thetas_1 <- c()
taus_1 <- c()
thetas_2 <- c()
taus_2 <- c()
thetas_diff <- c()

# Initial values
theta_1 <- 110
theta_2 <- 110
tau_1 <- 0.1
tau_2 <- 0.1

# Gibbs sampler
for (i in 1:NN){
   new_theta1 <- rnorm(1, (tau_1 * sum_y1 + 1.1)/(0.01 + 12*tau_1),
                       sqrt(1/(0.01 + 12 * tau_1)))
   new_theta2 <- rnorm(1, (tau_2 * sum_y2 + 1.1)/(0.01 + 7*tau_2),
                       sqrt(1/(0.01 + 7 * tau_2)))
   new_tau1 <- rgamma(1, shape = 6.01, rate = 4 + 0.5 * sum((y_1 - theta_1)^2))
   new_tau2 <- rgamma(1, shape = 3.51, rate = 4 + 0.5 * sum((y_2 - theta_2)^2))

   thetas_diff <- c(thetas_diff, new_theta1 - new_theta2)
   thetas_1 <- c(thetas_1, new_theta1)
   thetas_2 <- c(thetas_2, new_theta2)
   taus_1 <- c(taus_1, new_tau1)
   taus_2 <- c(taus_2, new_tau2)

   theta_1 <- new_theta1
   tau_1 <- new_tau1
   theta_2 <- new_theta2
   tau_2 <- new_tau2

}

# Exclude the burn-in trials
thetas_1 <- thetas_1[501:NN]
thetas_2 <- thetas_2[501:NN]
taus_1 <- taus_1[501:NN]
taus_2 <- taus_2[501:NN]
```

```r
thetas_diff <- thetas_diff[501:NN]

# Bayes estimator for theta_1
mean(thetas_1)
```

```
## [1] 116.9891
```

```r
# Bayes estimator for theta_2
mean(thetas_2)
```

```
## [1] 104.6635
```

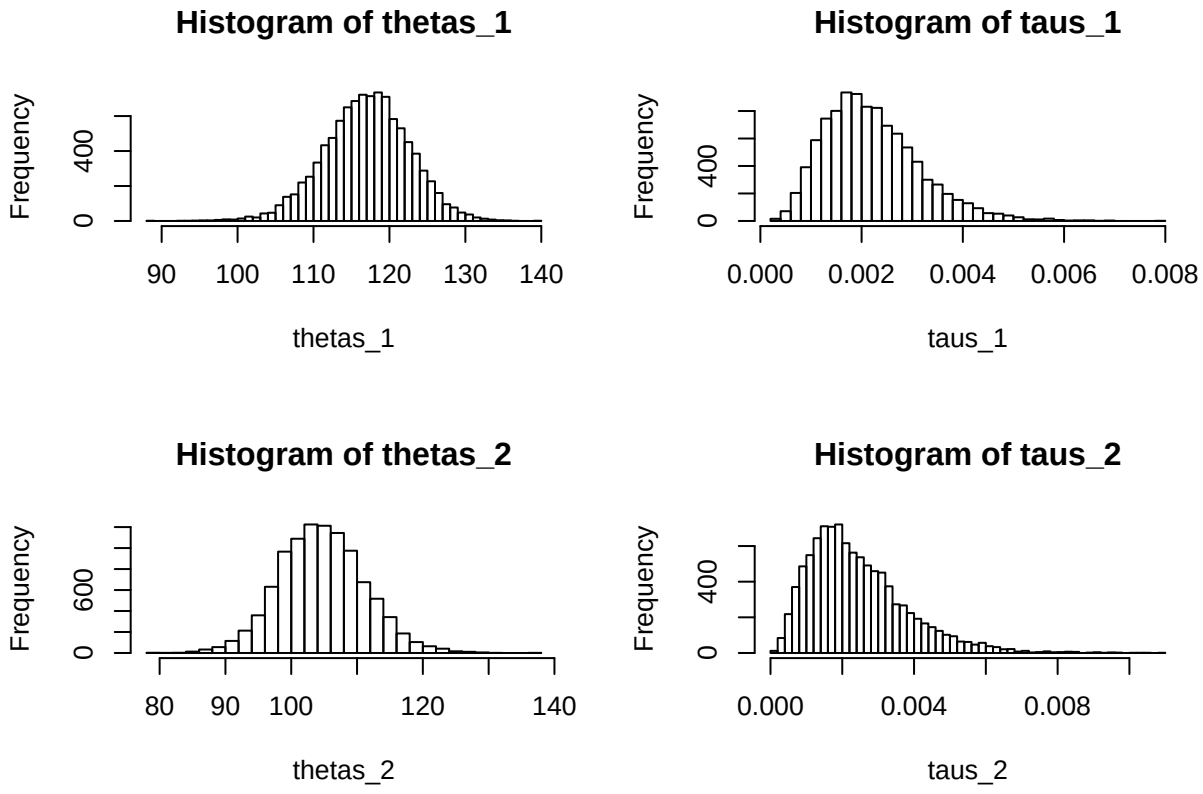
```r
# Bayes estimator for tau_1
mean(taus_1)
```

```
## [1] 0.002203912
```

```r
# Bayes estimator for tau_2
mean(taus_2)
```

```
## [1] 0.002428625
```

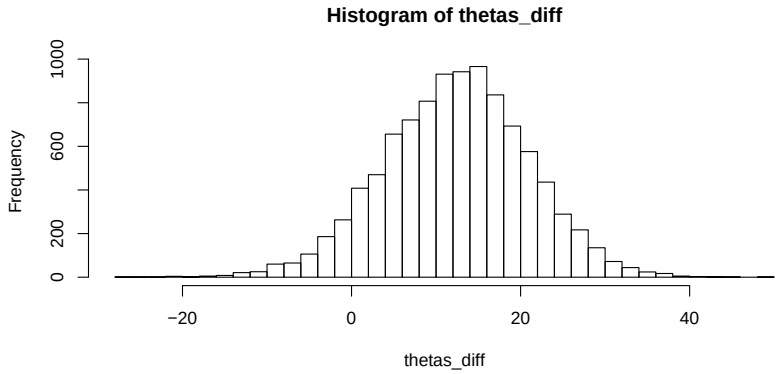
```r
#Plot the simulated posteriors
par(mfrow= c(2,2))
hist(thetas_1, 40)
hist(taus_1, 40)
hist(thetas_2, 40)
hist(taus_2, 40)
```



My Gibbs sampler yielded the following Bayesian estimators: $\hat{\theta}_{1B} = 116.99$ (slightly smaller than the MLE estimator $\hat{\theta}_{1MLE} = \bar{x}_1 = 120$), $\hat{\theta}_{2B} = 104.66$ (slightly larger than the MLE estimator $\hat{\theta}_{2MLE} = \bar{x}_2 = 101$), $\hat{\tau}_{1B} = 0.0022$, and $\hat{\tau}_{2B} = 0.0024$.

(b) Find the sample differences $\theta_1 - \theta_2$. The proportion of positive differences approximates the posterior probability of the hypothesis $H_0 : \theta_1 > \theta_2$. What is this proportion if the number of simulations is 10,000 with burn-in of 500?

```
# Plot the simulated theta_1 - theta_2 differences
hist(thetas_diff, 40)
```




```
# Bayes estimator for theta_1 - theta_2
mean(thetas_diff)
```

```
## [1] 12.32561
```

```
# Confirm that the 500 burn-in trials were removed
length(thetas_diff)
```

```
## [1] 10000
```

```
# Find proportion of differences that are positive
pos_count = sum(thetas_diff > 0)
pos_prop = pos_count / length(thetas_diff)
pos_prop
```

```
## [1] 0.9252
```

The posterior probability of $H_0 : \theta_1 > \theta_2$ is approximately 0.9252, which suggests that rats fed a high-protein gain more weight, on average, than do rats fed a low-protein diet (since $p_0 = 0.9252$ is much greater than $p_1 = 1 - 0.9252 = 0.0748$). Given that we assumed equal priors, the Bayes factor B_{01} in support of H_0 can be calculated as

$$B_{01} = \frac{p_0/p_1}{\pi_0/\pi_1} = \frac{p_0}{p_1} = \frac{0.9252}{0.0748} = 12.369.$$

Calibrating the Bayes factor, I find that $\log_{10}(12.369) = 1.092$, which indicates that there is strong evidence in support of $H_0 : \theta_1 > \theta_2$.

```r
# Calculate Bayes factor
p_0 = 0.9252
p_1 = 1 - p_0
B_01 = p_0/p_1
B_01
```

```
## [1] 12.36898
```

```r
# Calibrate Bayes factor
log10(B_01)
```

```
## [1] 1.092334
```

(c) Using sample quantiles, find the 95% equitailed credible set for $\theta_1 - \theta_2$. Does this set contain 0?

```r
# Find 95% credible set for theta_1 - theta_2
quantile(thetas_diff, c(0.025,0.975))
```

```
##     2.5%    97.5%
## -4.80479 28.63448
```

The Bayes estimator $(\widehat{\theta_1} - \widehat{\theta_2})_B = 12.326$ (a positive value) coupled with the posterior probability $p_0 = 0.9252$ provided seemingly strong support for $H_0 : \theta_1 > \theta_2$. However, the 95% equitailed credible set of $\theta_1 - \theta_2 \in [-4.805, 28.634]$ includes zero, which suggests that there is still a possibility that $\theta_1 \leq \theta_2$ (and that the high-protein diet actually is not better than the low-protein diet). If I were to interpret this outcome from the perspective of a frequentist statistician, I would point out that the hypothesis test was one-tailed, while the 95% credible set is inherently two-tailed. (A frequentist would construct a one-sided confidence interval to be consistent with the one-sided alternative hypothesis, but confidence intervals and credible sets are conceptually different.) Given my understanding of Bayesian statistics, I would think that the only way we would not have seen zero contained within the 95% credible set would be if the posterior probability for H_0 was closer to 1 (i.e., at least 0.95). At 0.9252, p_0 was still pretty close to 1, so it makes sense that only a small portion of the credible set lies below 0.

Out of curiosity, I ran a classical two-sample t-test to see how the results compared.

```r
# Two-sample t-test
t.test(y_1, y_2, alternative = "greater", paired = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  y_1 and y_2
## t = 1.9107, df = 13.082, p-value = 0.0391
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.398247      Inf
## sample estimates:
## mean of x mean of y
##       120       101
```

The frequentist hypothesis test also supports the conclusion that rats who are fed a high-protein diet gain significantly more weight on average than those fed a low-protein diet ($t_{13} = 1.9107, P = 0.0391,$ right-tailed). Also, the one-sided 95% confidence interval of $1.398 < \mu_1 - \mu_2$ is fully positive, as I expected.

Laura Schultz

ISyE 6420, Fall 2019

Built with R version 3.5.1