Homework 6

Samuel Eure

10/23/2018

Problem 5.3

Given $Y_1, ..., Y_n \sim i.i.d.$**Normal**$(\theta, \sigma^2)$**,**

$(\theta \mid \sigma^2) \sim$ **Normal**$(\mu_0, \sigma^2/\kappa_0)$**,**

Let $\tilde{\sigma}^2 = 1/\sigma^2$

and $\tilde{\sigma}^2 \sim$ **Gamma**$(\nu_0/2, \nu_0\sigma_0^2/2)$

Derive $p(\theta \mid y_1, ..., y_n)$ **and** $p(1/\sigma^2 \mid y_1, ..., y_n)$

Where $\tau_0^2 = \sigma^2/\kappa_0$

$$P(\theta \mid Y_{1:n}) \propto \int P(\theta, \tilde{\sigma}^2 \mid Y_{1:n}) d\tilde{\sigma}^2 \propto \int P(Y_{1:n} \mid \theta, \tilde{\sigma}^2) P(\theta \mid \tilde{\sigma}^2) P(\tilde{\sigma}^2) d\tilde{\sigma}^2$$

$$\propto \int (\tilde{\sigma}^2)^{n/2} \Big[\Pi_i^n \exp(-\frac{\tilde{\sigma}^2}{2}(y_i - \theta)^2\Big] (\tilde{\sigma}^2)^{1/2} \exp(-\frac{\kappa_0 \tilde{\sigma}^2}{2}(\theta - \mu_0)^2)(\tilde{\sigma}^2)^{(\nu_0/2-1)} \mathrm{e}(-\frac{\nu_0\sigma_0^2}{2}\tilde{\sigma}^2) d\tilde{\sigma}^2$$

$$\propto \int (\tilde{\sigma}^2)^{\frac{n+\nu_0+1}{2}-1} \exp(-\frac{\tilde{\sigma}^2}{2}[\sum_i^n (y_i - \theta)^2 + k_0(\theta - \mu_0)^2 + \nu_0\sigma_0^2]) d\tilde{\sigma}^2$$

Which looks like the distribution of a Gamma($\frac{n+\nu_0+1}{2}, \frac{\sum_i^n (y_i-\theta)^2 + k_0(\theta-\mu_0)^2 + \nu_0\sigma_0^2}{2}$**) for the precision, so it must integrate to**

$$P(\theta \mid Y_{1:n}) \propto \frac{\Gamma(\frac{n+\nu_0+1}{2})}{(\frac{\sum_i^n (y_i-\theta)^2 + k_0(\theta-\mu_0)^2 + \nu_0\sigma_0^2}{2})^{(n+\nu_0+1)/2}}$$

$$P(\theta \mid Y_{1:n}) \propto \big(\sum_i^n (y_i - \theta)^2 + k_0(\theta - \mu_0)^2 + \nu_0\sigma_0^2\big)^{-(n+\nu_0+1)/2}$$

To find the marginal posterior distribution of $\tilde{\sigma}^2 \mid y_1, ..., y_n$**. Since our sampling scheme is given as** $(Y_1, ..., Y_n) \sim$ **Normal**(θ, σ^2)**, the posterior can be found by**

$$p(\tilde{\sigma}^2 \mid y_1, ..., y_n) \propto p(y_1, ..., y_n \mid \tilde{\sigma}^2) p(\tilde{\sigma}^2) \propto p(\tilde{\sigma}^2) \int p(y_1, ..., y_n \mid \theta, \tilde{\sigma}^2) p(\theta \mid \tilde{\sigma}^2) d\theta$$

$$\propto (\tilde{\sigma}^2)^{\frac{\nu_0}{2}-1} e^{\left(\frac{-\nu_0\sigma_0^2}{2}\tilde{\sigma}^2\right)} \left(\tilde{\sigma}^2\right)^{n/2} \int exp\Big[-\frac{\tilde{\sigma}^2}{2}\sum_i^n (y_i - \theta)^2\Big] \Big((\tilde{\sigma}^2)^{1/2} e^{-\frac{\kappa_0\tilde{\sigma}^2}{2}(\theta-\mu_0)^2}\Big) d\theta$$

$$\propto (\tilde{\sigma}^2)^{\frac{\nu_0+n}{2}-1} e^{\left(\frac{-\nu_0\sigma_0^2}{2}\tilde{\sigma}^2\right)} \int \frac{1}{(\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}\left(\kappa_0(\theta-\mu_0)^2 + \sum_i^n (y_i-\theta)^2\right)\right] d\theta$$

Focusing only on the part in the exponential in the integrand, since

$$\kappa_0(\theta-\mu_0)^2 + \sum_i^n (y_i-\theta)^2 = \kappa_0((\theta-\bar{y})+(\bar{y}-\mu_0))^2 + \sum_i^n ((y_i-\bar{y})+(\bar{y}-\theta))^2$$

Where

$$\kappa_0((\theta-\bar{y})+(\bar{y}-\mu_0))^2 = \kappa_0(\bar{y}-\theta)^2 + \kappa_0(\bar{y}-\mu_0)^2 + 2\kappa_0(\theta-\bar{y})(\bar{y}-\mu_0)$$

and

$$\sum_i^n \big((y_i-\bar{y})+(\bar{y}-\theta)\big)^2 = \sum_i^n (y_i-\bar{y})^2 + n(\bar{y}-\theta)^2$$

with

$$\sum_i^n \big(y_i-\bar{y}\big)\big(\bar{y}-\theta\big) = n\bar{y}\bar{y} - n\bar{y}\bar{y} - n\theta\bar{y} + n\theta\bar{y} = 0$$

Thus,

$$\propto (\tilde{\sigma}^2)^{\frac{\nu_0+n}{2}-1} e^{\left(\frac{-\nu_0\sigma_0^2+\frac{\kappa_0 n}{\kappa_0+n}(\bar{y}-\mu_0)^2+(n-1)s_n^2}{2}\tilde{\sigma}^2\right)} \int \frac{1}{(\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}\left((\bar{y}-\theta)^2 + 2\kappa_0(\theta-\bar{y})(\bar{y}-\mu_0)\right)\right] d\theta$$

Where the integrand has the shape of a normal distribution, which will integrate to a constant, leaving the remainder to have the form of a gamma distribution of

$$\tilde{\sigma}^2 \mid Y_{1:n} \sim Gamma(\frac{\nu_0+n}{2}, \frac{\nu_0\sigma_0^2+(n-1)s_n^2+\frac{\kappa_0 n}{\nu_0+n}(\bar{y}-\mu_0)^2}{2})$$

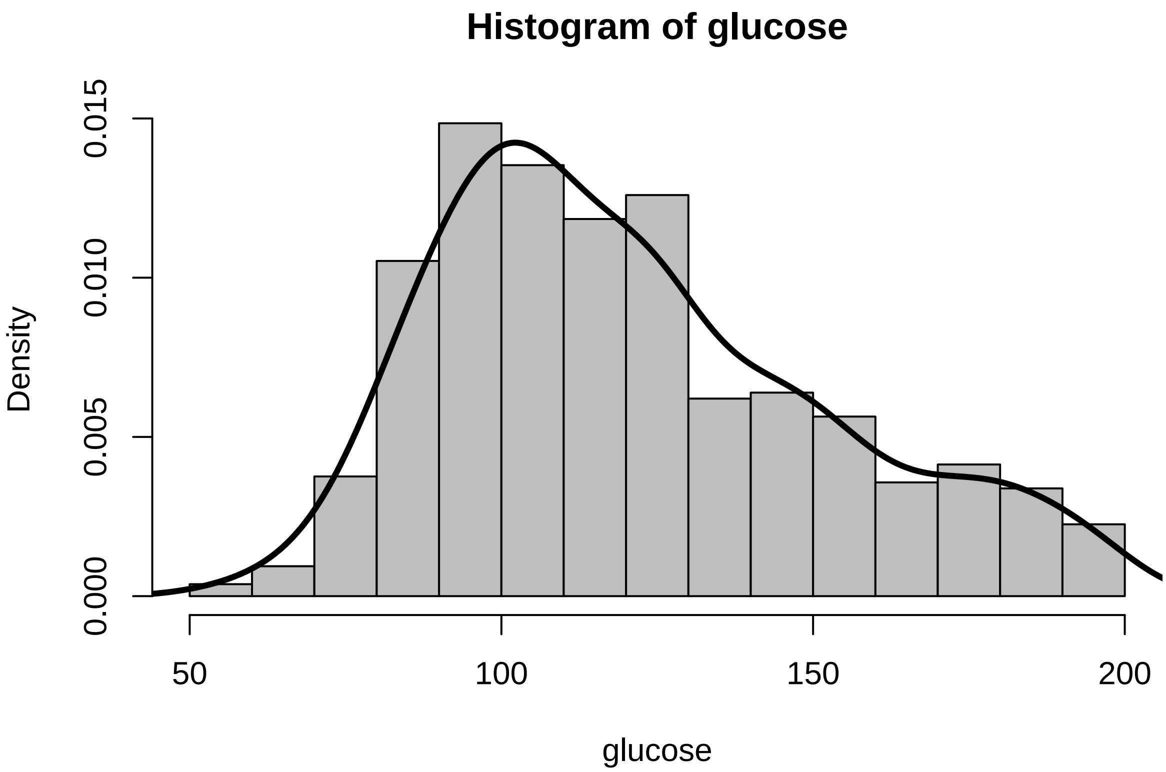
Problem 6.2

6.2a

The Glucose data is plotted below.

```r
set.seed(1)
glucose = read.table("glucose.dat", header = FALSE)
names(glucose) = c('plasma')
glucose = glucose$plasma
```

```r
hist(glucose, col = 'grey', freq=FALSE)
lines(density(glucose), lwd = 3)
```



As seen in the plots above, the distribution of the glucose readings are not symetrical, and seem to have a smooth left size (where lower level glucose measurements are located) and a "bumpy" right side as noted byt the two small local maxima at around 140 and 175. Additionally, there seems to be quite a shart cut off at around 200, which would not be seen with a normal distribution.

6.2b

First, the joint posterior distribution is

$$P(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n, p \mid y_1, ..., y_n) \propto P(y_1, ..., y_n, p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n)$$

$$\propto P(y_1, ..., y_n \mid p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n) P(X_1, ..., X_n \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, p) P(\theta_1, \theta_2, \sigma_1^2, \sigma_2^2 \mid p) P(p)$$

$$\propto P(y_1, ..., y_n \mid \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n) P(X_1, ..., X_n \mid p) P(\theta_1) P(\theta_2) P(\sigma_1^2) P(\sigma_2^2) P(p)$$

Since $\theta_1, \theta_2, \sigma_1^2, and \sigma_2^2$ are all independent and X_i is conditional on p, and $Y_1, ..., Y_n$ are conditionally independent given $X_1, ..., X_n$. Thus, the full conditional of p is

$$P(p \mid y_1, ..., y_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n) \propto P(y_1, ..., y_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n \mid p) P(p)$$

$$\propto P(X_1, ..., X_n \mid p) P(p)$$

Since X_i is drawn as a Bernoulli(p) iid random variable if a "success" is defined as $X_i = 1$, if $I()$ is the indicator function, then

$$p\big((X_1, ..., X_n) \mid p\big) = p^{\sum_i^n (I(X_i=1))}(1-p)^{n-\sum_i^n (I(X_i=1))}$$

And thus

$$P(p \mid y_1, ..., y_n, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2, X_1, ..., X_n) = p^{\sum_i^n (I(X_i=1))}(1-p)^{n-\sum_i^n (I(X_i=1))}p^{a-1}(1-p)^{b-1}$$

$$= p^{\sum_i^n (I(X_i=1))+a-1}(1-p)^{n-\sum_i^n (I(X_i=1))+b-1}$$

Which is $Beta(\sum_i^n(I(X_i = 1)) + a, n - \sum_i^n(I(X_i = 1)) + b)$**. Further, the full condition for X_i is given as**

$$P(X_i \mid X_{(1:n,-i)}, p, Y_{1:n}, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) \propto P(X_i)P(Y_i, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2 \mid X_i)$$

$$\propto \big(p * \text{dnorm}(y_i \mid \theta_1, \sigma_1)\big)^{p^{I(X_i=1)}} \big((1-p) * \text{dnorm}(y_i \mid \theta_2, \sigma_2)\big)^{p^{I(X_i=2)}}$$

Which looks like a Bernoulli($\frac{A}{A+B}$) with

$$A = p * \text{dnorm}(y_i \mid \theta_1, \sigma_1)$$

$$B = (1-p) * \text{dnorm}(y_i \mid \theta_2, \sigma_2)$$

Next,I shall consider the cases of θ_1

$$P(\theta_1 \mid ...) \propto P(Y_{1:n}, X_{1:n}, \mid \theta_1, \sigma_1^2)P(\theta_1)$$

$$\propto \Pi_i^n \Big(\exp(-\frac{1}{\sigma_1^2}(y_i - \theta_0)^2)\Big)^{I(X_i=1)} P(\theta_1)$$

Which we have already seen to have the form of

$$(\theta_1 \mid ...) \sim \text{Normal}\Big(\frac{\mu_0/\tau_0^2 + \sum_i^n (y_i)I(x_i = 1)/\sigma_1^2}{1/\tau_0^2 + \sum_i^n I(x_i = 1)/\sigma_1^2}, (1/\tau_0^2 + \sum_i^n I(x_i = 1)/\sigma_1^2)^{-1}\Big)$$

Which, by symetry

$$(\theta_2 \mid ...) \sim \text{Normal}\Big(\frac{\mu_0/\tau_0^2 + \sum_i^n (y_i)I(x_i = 2)/\sigma_2^2}{1/\tau_0^2 + \sum_i^n I(x_i = 2)/\sigma_2^2}, (1/\tau_0^2 + \sum_i^n I(x_i = 2)/\sigma_2^2)^{-1}\Big)$$

And finally, since the full conditional of θ_1 is just he posterior distribution of θ_1 given the data points of $X_i = 1$, based on how this full conditional was derived, the full condition of σ_1^2 should just be the posterior distribution of σ_1^2 based on the values of $X_i = 1$. Thus,

With

$$(\sigma_k^2 \mid ...) \sim \text{Inv-Gamma}(\alpha, \beta)$$

$$\alpha = \frac{v_0 + \sum_i^n I(x_i = k)}{2}$$

$$\beta = \frac{\left[v_0\sigma_0^2 + \sum_i^n [I(x_i = k)(y_i - \frac{1}{\sum I(X_i=k)} \sum_i^n (y_i * I(x_i = k)))^2] + \frac{\nu_0 * \sum I(X_i=k)}{\nu_0 + \sum I(X_i=k)} (\frac{\sum_i^n (y_i * I(x_i=k))}{\sum_i^n (I(x_i=k))} - \mu_0)^2\right]}{2}$$

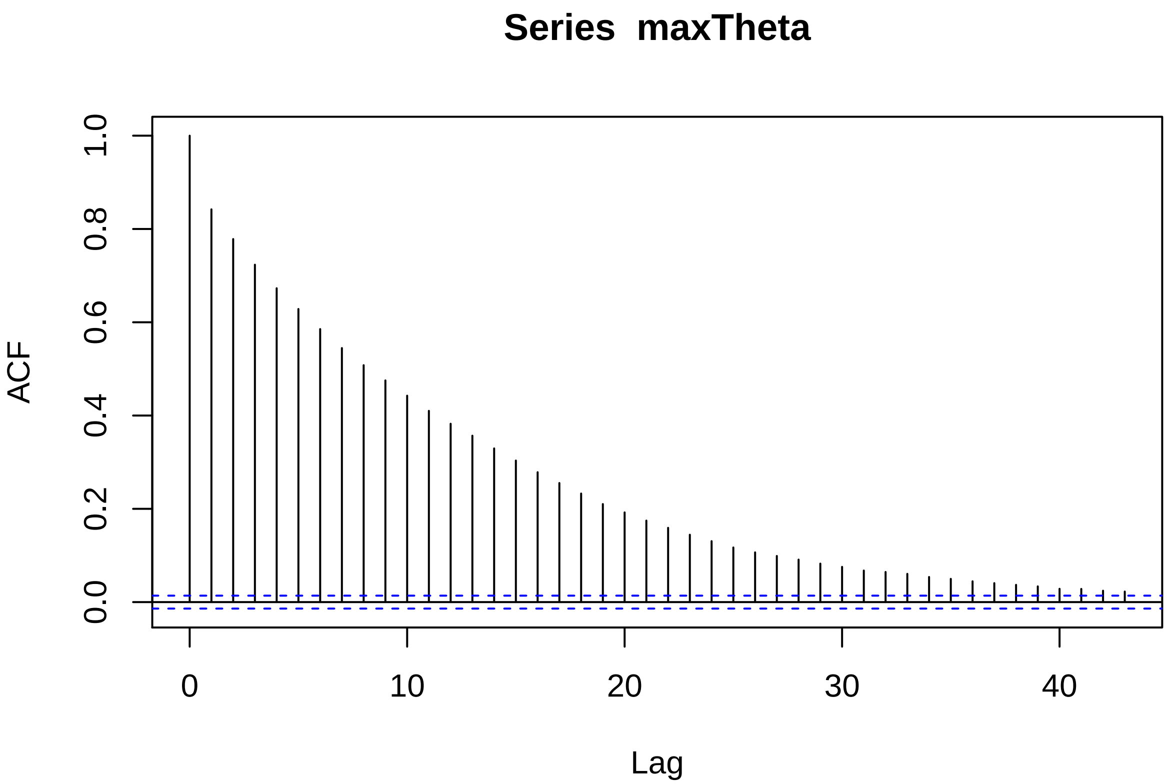
6.2c

```
library(coda)
a =1; b=1; mu0 = 120; tau0.squared = 200; sigma0.squared = 1000; nu0 = 10;
Size = 20000;
n = length(glucose)
partDYSamples = numeric(Size-1)

X.values = matrix(nrow =Size, ncol = n)
theta1 = numeric(Size); theta2 = numeric(Size); sigma1.squared = numeric(Size);
sigma2.squared = numeric(Size); p.values = numeric(Size);
maxTheta = numeric(Size-1)
minTheta = numeric(Size-1)
X.values[1,] = sample(x = c(1,2), n, prob = c(.5,.5), replace = TRUE)

for(i in 2:Size){
  if(i%%5000 ==0){
    print(i)
  }
  #p
  x1.bool = X.values[i-1,]==1
  x1.size = sum(x1.bool)

  x2.bool = X.values[i-1,]==2
  x2.size = sum(x2.bool)
  p.values[i] = rbeta(1, a + x1.size, b+x2.size)

  y1.values = glucose[x1.bool]
  y1.sum    = sum(y1.values)
  y2.values = glucose[x2.bool]
  y2.sum    = sum(y2.values)

  alpha1 = (nu0+x1.size)/2
  beta1  = (nu0*sigma0.squared + sum((y1.values - mean(y1.values))^2) + (nu0*x1.size)*(mean(y1.values)-

  sigma1.squared[i] = 1/rgamma(1, alpha1, beta1)
```

```
  alpha2 = (nu0+x2.size)/2
  beta2  = (nu0*sigma0.squared + sum((y2.values - mean(y2.values))^2) + (nu0*x2.size)*(mean(y2.values)-m
  sigma2.squared[i] = 1/rgamma(1, alpha2, beta2)


  newVar1 = 1/(1/tau0.squared+x1.size/sigma1.squared[i])
  newMu1 = (mu0/tau0.squared+x1.size*mean(y1.values)/sigma1.squared[i])/(1/newVar1)
  theta1[i] = rnorm(1,newMu1 , sqrt(newVar1))

  newVar2 = 1/(1/tau0.squared+x2.size/sigma2.squared[i])
  newMu2 = (mu0/tau0.squared+x2.size*mean(y2.values)/sigma2.squared[i])/(1/newVar2)
  theta2[i] = rnorm(1,newMu2 , sqrt(newVar2))

  maxTheta[i-1] = max(theta2[i], theta1[i])
  minTheta[i-1] = min(theta2[i], theta1[i])


  partDX = sample(x = c(1:2), 1, prob= c(p.values[i], 1-p.values[i]))
  partDTheta = theta1[i]^(partDX==1)*theta2[i]^(partDX==2)
  partDVar = sigma1.squared[i]^(partDX==1)*sigma2.squared[i]^(partDX==2)
  partDYSamples[i-1] = rnorm(1, partDTheta, sqrt(partDVar))


  for(k in 1:n){
    A = p.values[i]*dnorm(glucose[k], theta1[i], sqrt(sigma1.squared[i]))
    B = (1-p.values[i])*dnorm(glucose[k], theta2[i], sqrt(sigma2.squared[i]))
    localP = A/(A+B)
    X.values[i, k] = sample(x = c(1:2), 1, prob = c(localP, 1-localP), replace = TRUE)
  }
}
```

```
## [1] 5000
## [1] 10000
## [1] 15000
## [1] 20000
```

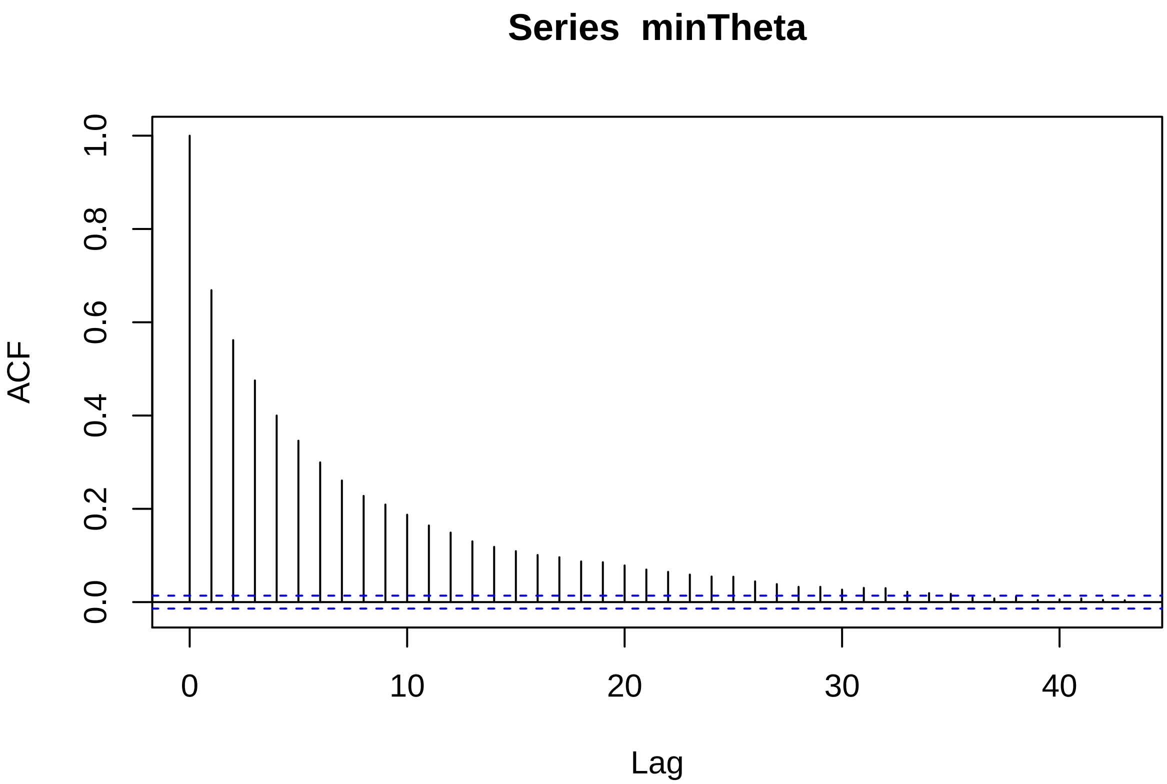
```
#For the Max Theta
acf(maxTheta)
```




```
effectiveSize(maxTheta)
```

```
##      var1
## 797.6106
```

```
#For the Min Theta
acf(minTheta)
```

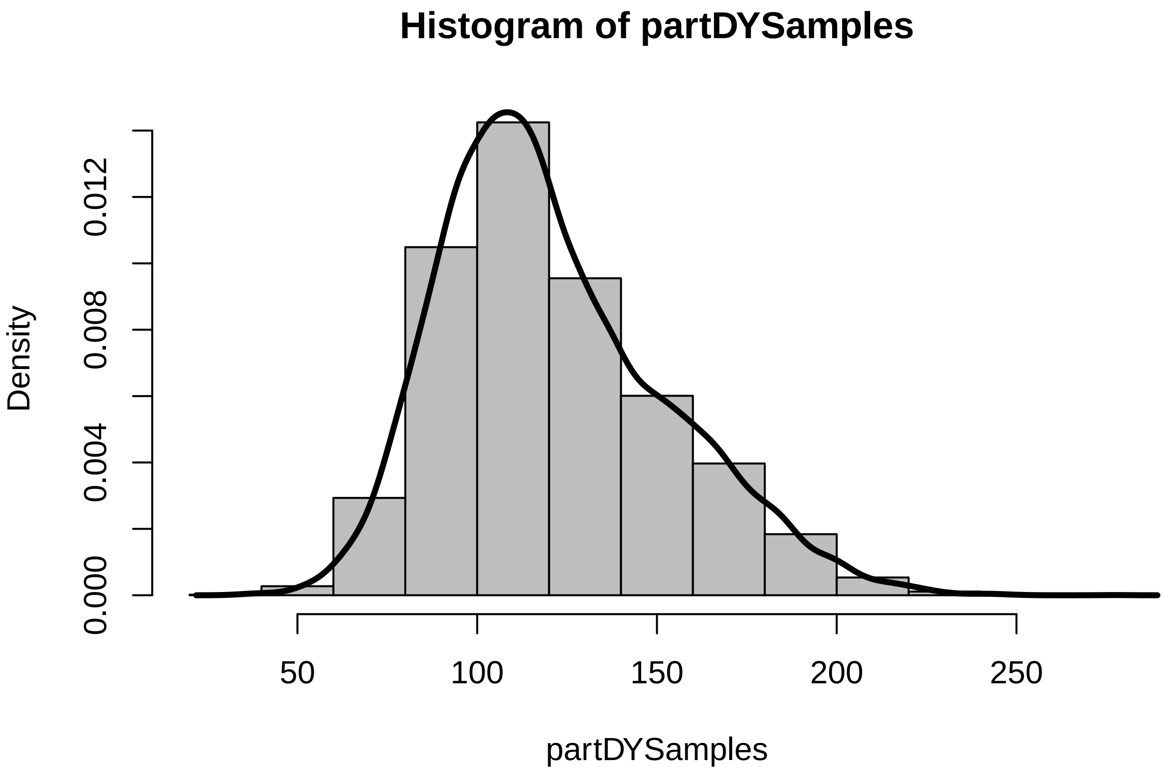



```
effectiveSize(minTheta)
```

```
##     var1
## 1833.873
```

6.2d

```
hist(partDYSamples, col = 'grey', freq=FALSE)
lines(density(partDYSamples), lwd = 3)
```



Compared with the density plot in part a, this histogram seems to possess the general shape of the observed data, however it still lacks key features such as the "humps" in the data at around 140 and 175. Additionally, the histogram generated from the Gibbs sampler seems to have a smoother right tail than that of the observed data, which as a sharp cut off at around 200. This being said, our model would have been nicer (in comparison to the observed sample) had we made a correction to put more weight between 150 and 200 more uniformly. Our current model fails to capture this feature of the data. Perhaps a three component model that doesn't allow values larger than 200 would be more appropriate given the shape of the observed data, as a three component model may allow us to simulate the "humps" in the observed distribution, and the cut off of 200 would allow us to replicate the seemingly natural threshold of plamsa levels observed in the data.

Problem 6.3

6.3a

$$P(\beta \mid y_{1:n}, x_{1:n}, z_{1:n}, c) \propto P(\beta, y_{1:n}, x_{1:n}, z_{1:n}, c) \propto P(z_{1:n} \mid x_{1:n}, z_{1:n}, \beta, c)P(\beta)$$

$$\propto \Pi_i^n \left(e^{-\frac{1}{2}(z_i-\beta x_i)^2}\right) e^{-\beta^2\tau_\beta^{-2}/2} \propto e^{-\frac{\sum_i^n (z_i-\beta x_i)^2+\beta^2\tau_\beta^{-2}}{2}}$$

$$\propto e^{-\sum_i^n x_i^2+\tau_\beta^{-2} \frac{\beta^2-2\beta\frac{\sum_i^n z_i x_i}{\sum_i^n x_i^2+\tau_\beta^{-2}}+\frac{\sum_i^n z_i^2}{\sum_i^n x_i^2+\tau_\beta^{-2}}}{2}}$$

Which smells like a normal distribution with mean $\frac{\sum_i^n z_i x_i}{\sum_i^n x_i^2 + \tau_\beta^{-2}}$ and variance $\sum_i^n x_i^2 + \tau_\beta^{-2}$.

$$(\beta \mid c, z_{1:n}, y_{1:n}, x_{1:n}) \sim \text{Normal}\left(\frac{\sum_i^n z_i x_i}{\sum_i^n x_i^2 + \tau_\beta^{-2}}, \sum_i^n x_i^2 + \tau_\beta^{-2}\right)$$

6.3b

$$P(c \mid y_{1:n}, x_{1:n}, z_{1:n}, \beta) \propto P(y_{1:n} \mid z_{1:n}, c, x_{1:n}, \beta)P(c)$$

since $y_i = 1$ iff $z_i > c$ and $y_i = 0$ iff $z_i < c$, given $z_{1:n}$ and c, P(y_i) is deterministic and either 1 or 0. Thus,

$$\propto \Pi_i^n \left[I(z_i > c)^{I(y_i=1)} I(z_i \leq c)^{I(y_i=0)} \right] e^{-c^2/2\tau_c^2}$$

Thus, the full conditional of c has a normal distribution, however, as implied above, the probability of c is zero when that value of c results in $I(z_i \leq c)^{I(y_i=1)}$ or $I(z_i > c)^{I(y_i=0)}$ for some $\{z_i, y_i\}$. Thus, it must be that

$$c \in [max(z_i \mid y_i = 0), min(z_i \mid y_i = 1)]$$

since values of c outsite of this interval automatically result in a c of probabillity zero.

As for values of z_i

$$(z_i \mid x_i, \beta) \sim \text{Normal}(x_i\beta, 1)$$

Since ϵ_i is a standard uniform random variable. Thus,

$$P(z_j \mid c, z_{1:n,-j} y_{1:n}, x_{1:n}, \beta) \propto e^{-(z_i - x_i\beta)^2/2} I(z_i \leq c)^{I(y_i=0)} I(z_i > c)^{I(y_i=1)}$$

Thus, z_i has a truncated normal distirubtion with a probability of zero whenever $\{z_i \leq c$ and $y_i = 1\}$ or $\{z_i > c$ and $y_i = 0\}$

6.3c

```r
library(MCMCglmm)
```

```
## Loading required package: Matrix
## Loading required package: ape
```

```r
divorce = read.table('divorce.dat', header = F)
divorce = data.frame(divorce)
names(divorce) = c('age', "divorce")
tau.b.squared = 16; tau.c.squared = 16;
Size = 13000;
n = length(divorce$age)
z.values = matrix(nrow = Size, ncol = n)
```

```r
c.values = numeric(Size)
b.values = numeric(Size)
b.values[1] = 0
c.values[1] = 0
for(i in 2:Size){
  for(k in 1:n){
    divorced = divorce$divorce[k]
    ageDiff  = divorce$age[k]
    if(divorced){
      z.values[i,k] = rtnorm(1, b.values[i-1]*ageDiff, 1, lower = c.values[i-1])
    }
    else{
      z.values[i,k] = rtnorm(1, b.values[i-1]*ageDiff, 1, upper = c.values[i-1])
    }
  }
  y0.bool = divorce$divorce == 0
  y1.bool = divorce$divorce == 1
  z0 = z.values[i,][y0.bool]
  z1 = z.values[i,][y1.bool]
  cMin = min(z.values[i,])
  cMax = max(z.values[i,])
  c.values[i] = rtnorm(1, 0, sqrt(tau.c.squared), lower = cMin, upper = cMax)
  bVar = (sum(divorce$age*divorce$age)+1/tau.b.squared)
  bMean = sum(z.values[i,]*divorce$age)/bVar
  b.values[i] = rnorm(1, bMean, sqrt(bVar))
}
print("effective size Beta")
```

```
## [1] "effective size Beta"
```

```r
effectiveSize(mcmc(b.values))
```

```
##     var1
## 1115.022
```

```r
print("effective size Z values")
```

```
## [1] "effective size Z values"
```

```r
effectiveSize(mcmc(z.values[2:Size,1]))
```

```
##     var1
## 4273.571
```

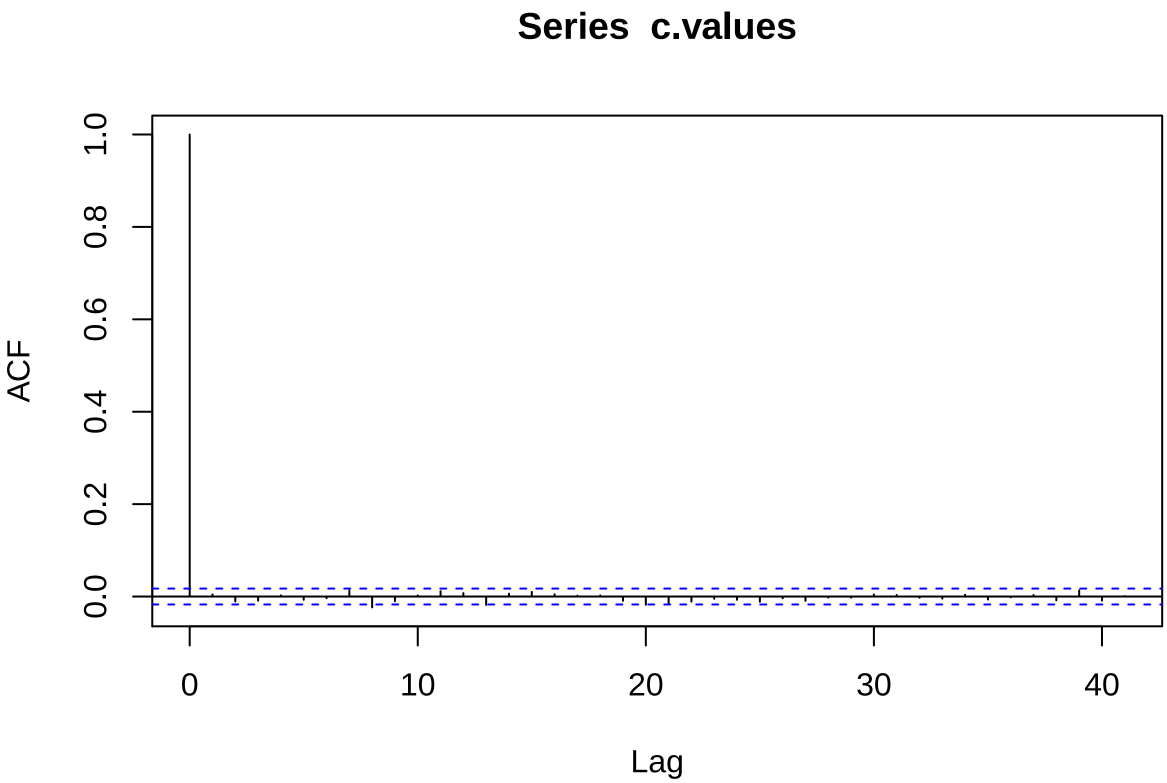
```r
print("effective size c values")
```

```
## [1] "effective size c values"
```

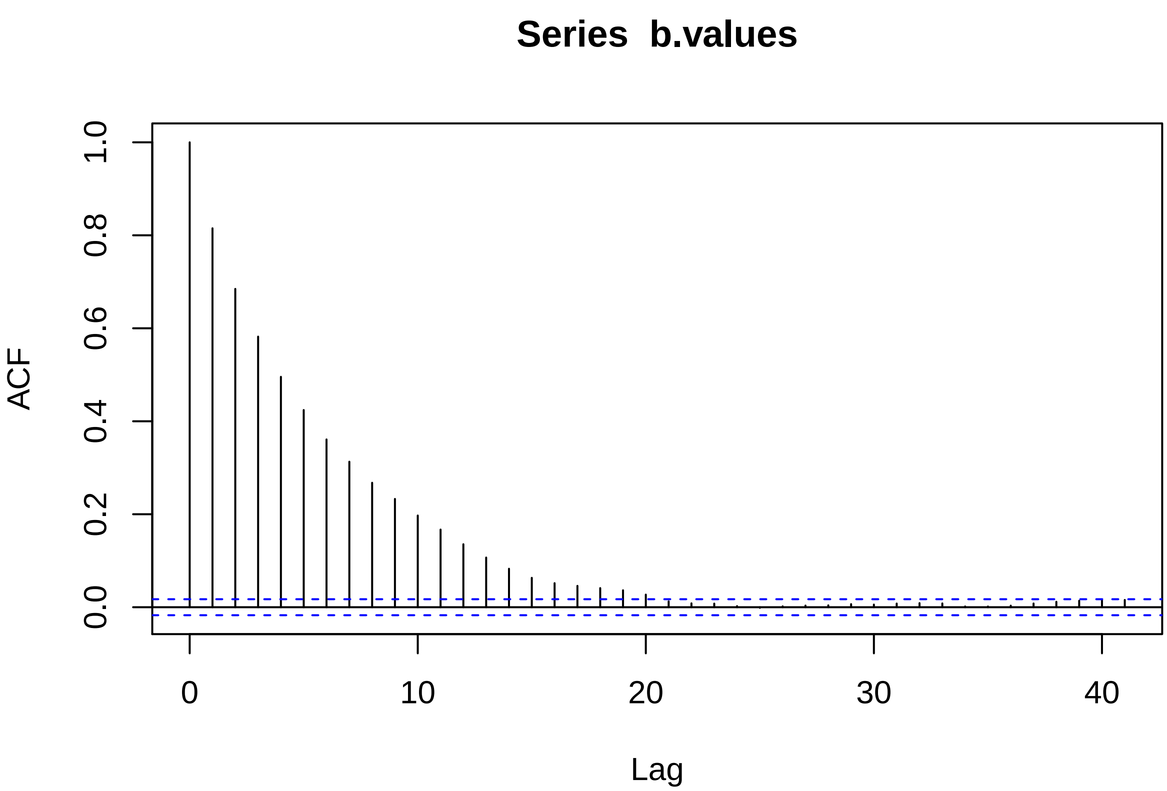
```r
effectiveSize(mcmc(c.values))
```

```
##  var1
## 13000
```

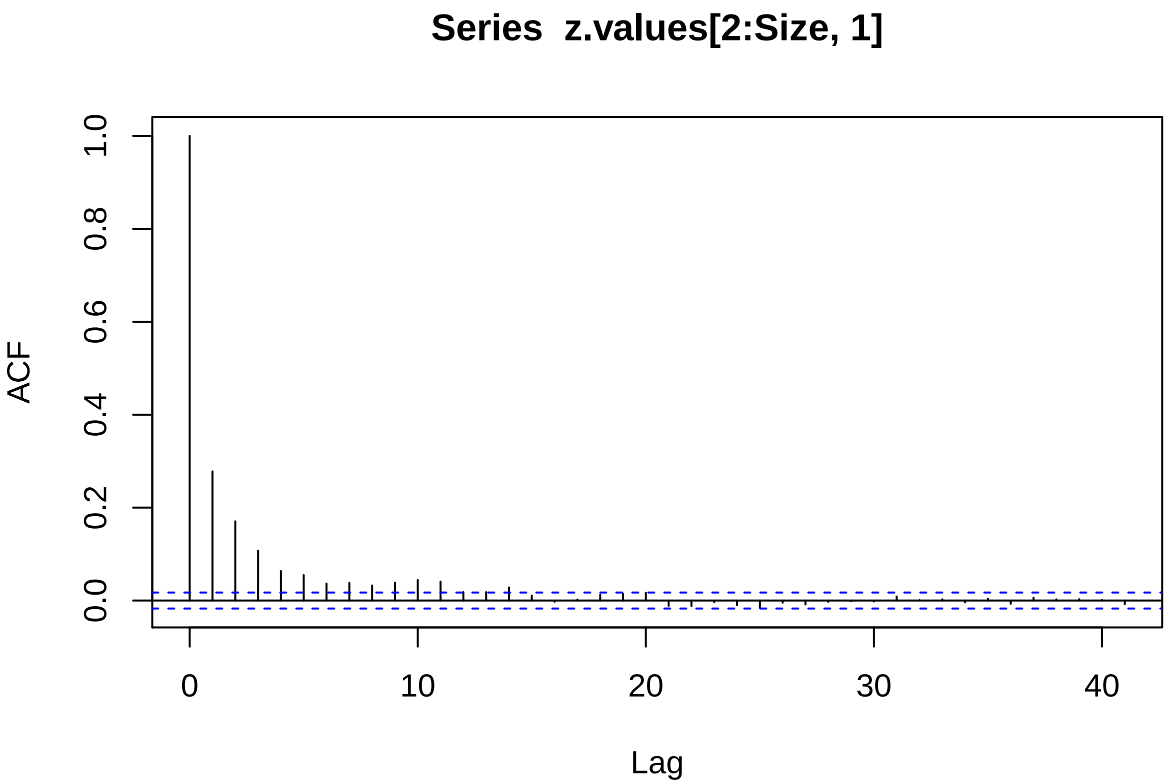
```r
acf(c.values)
```




```r
acf(b.values)
```



```r
acf(z.values[2:Size,1])
```



It appears that my autocorrelation values are quite low for my c and b values but slightly autocorrelated for z, indicating that the Markov chain mixes more slowly for the z values than for the other two values. This being said,

Moreover, it appears that all three values acheive stationarity since their autocorrelations aproach negligable values.

3.3d

```r
#I know the function for this, so why not an HPD.
HPDinterval(mcmc(b.values), .95)
```

```
##          lower    upper
## var1 -29.45388 93.58568
## attr(,"Probability")
## [1] 0.95
```

Since samples from the joint distribution can be thought of as samples from the marginal distribution, $P(\beta > 0 \mid y_{1:n}, x_{1:n})$ is approximately

```r
sum(b.values>0)/length(b.values)
```

```
## [1] 0.7925385
```