Homework 8

Samuel Eure

11/6/2018

Problem 8.3

8.3a

Importing the Data

```
library(MASS)
library(coda)
Y = list();
theta = 1:8; #initial theta for the Gibbs Sampler
N = 1:8;
yBar = 1:8;
for(i in 1:8){
  name     = paste("school", toString(i), sep = "");
  fileName = paste(name ,'.dat', sep = "");
  groupI   = read.table(fileName, header = F, col.names = c(name));
  Y[[i]]   = groupI;
  theta[i] = mean(groupI[,1]);
  yBar[i]  = mean(groupI[,1]);
  N[i]     = length(groupI[,1]);
}
names(Y) = 1:8;
```

Functions

```
getSumSquares <- function(Y, theta){
  total = 0;
  for(j in 1:length(theta)){
    y.j = Y[[j]]; theta.j = theta[j];
    diff.j = sum((y.j - theta.j)^2);
    total = total + diff.j;
  }
  return(total);
}
```

Obtaining the Samples

```
mu0 <- 7; g2.0 <- 5; t2.0 <- 10; nu0 <- 2; s2.0 <- 15;

M = 8;
Samples = 10000
#Creating the data structures
THETA   = matrix(0, nrow = Samples, ncol = M);
```

```r
SIGMA2 = matrix(0, nrow = Samples, ncol = 1);
MU     = matrix(0, nrow = Samples, ncol = 1);
TAU2   = matrix(0, nrow = Samples, ncol = 1);

#Set Initial Parameters
tau2 = t2.0; tau2.post.a = (nu0+M)/2;
sigma2 = s2.0; sigma2.post.a = (nu0 + sum(N))/2;

for(s in 1:Samples){
  #mu
  mu.var = 1/(M/tau2+1/g2.0);
  mu = rnorm(1, mu.var*(M*mean(theta)/tau2 + mu0/g2.0) , sqrt(mu.var));
  #tau
  tau2 = 1/rgamma(1, tau2.post.a, (nu0*t2.0+sum((theta-mu)^2))/2);
  #theta
  for(j in 1:M){
    theta.j.var  = 1/(N[j]/sigma2 + 1/tau2);
    theta.j.mean = (N[j]*yBar[j]/sigma2 + mu/tau2)*theta.j.var;
    theta[j]     = rnorm(1, theta.j.mean, sqrt(theta.j.var ));
  }
  #sigma
  sigma.2.post.b =  ( nu0*s2.0 + getSumSquares(Y, theta))/2;
  sigma2 = 1/rgamma(1, sigma2.post.a, sigma.2.post.b);
  THETA[s,] = theta; SIGMA2[s,1] = sigma2; MU[s,1] = mu; TAU2[s,1] = tau2;
}
```

Effective Sample Sizes

```r
print("Effective Sample Sizes for Thetas")
```

```
## [1] "Effective Sample Sizes for Thetas"
```

```r
effectiveSize(THETA)
```

```
##      var1      var2      var3      var4      var5      var6      var7 
##  9271.384 10000.000  9288.051  9406.174  8527.373 10274.512  9439.380 
##      var8 
##  9582.132
```

```r
print("For sigma.squared")
```

```
## [1] "For sigma.squared"
```

```r
effectiveSize(SIGMA2)
```

```
##     var1 
## 9274.501
```

```r
print("For tau.squared")
```

```
## [1] "For tau.squared"
```

```r
effectiveSize(TAU2)
```

```
##     var1 
## 6772.759
```

```r
print("For mu")
```

```
## [1] "For mu"
```

```r
effectiveSize(MU)
```

```
##     var1
## 8206.523
```

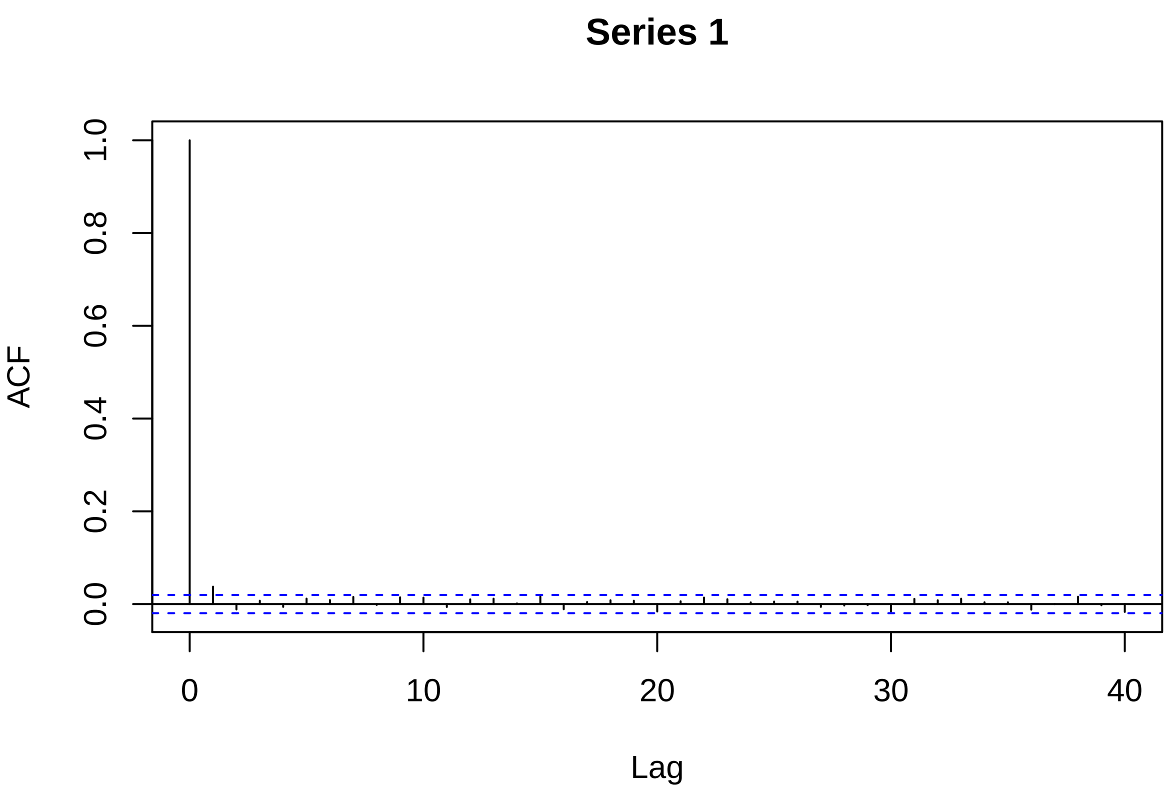
Convergence Analysis

Autocorrelation

```r
print("For sigma.squared")
```

```
## [1] "For sigma.squared"
```

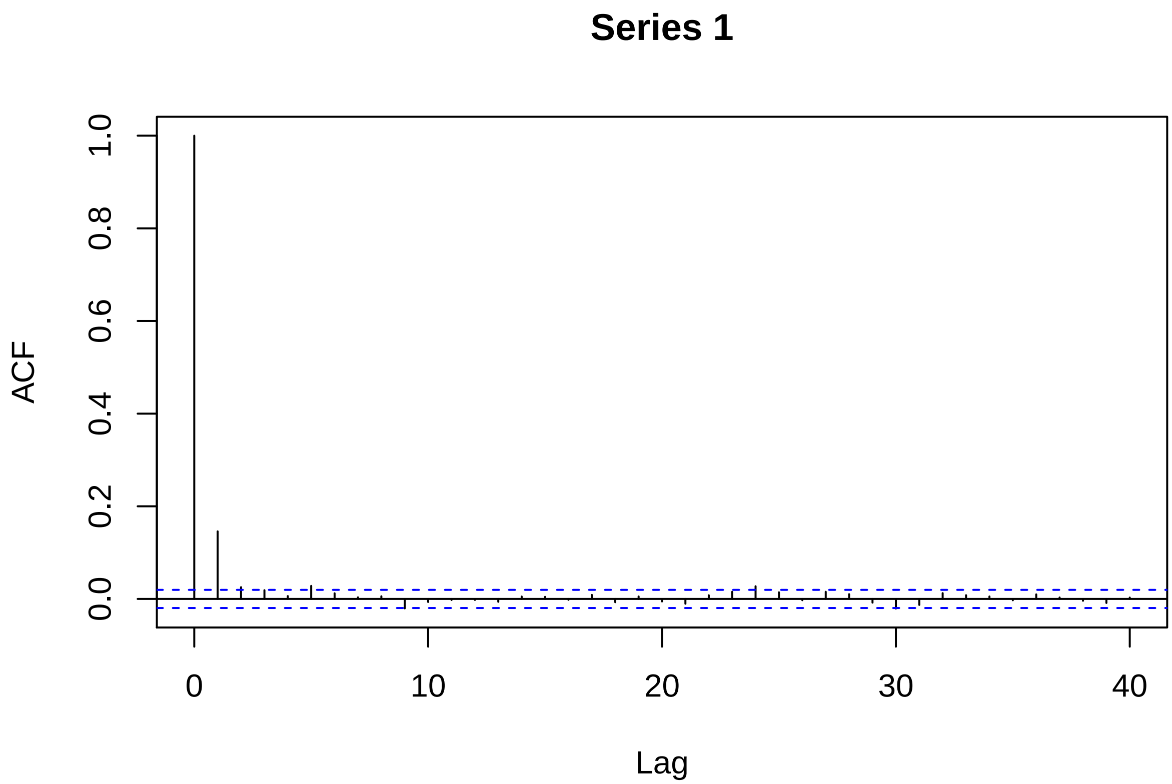
```r
acf(SIGMA2)
```




```r
print("For tau.squared")
```

```
## [1] "For tau.squared"
```

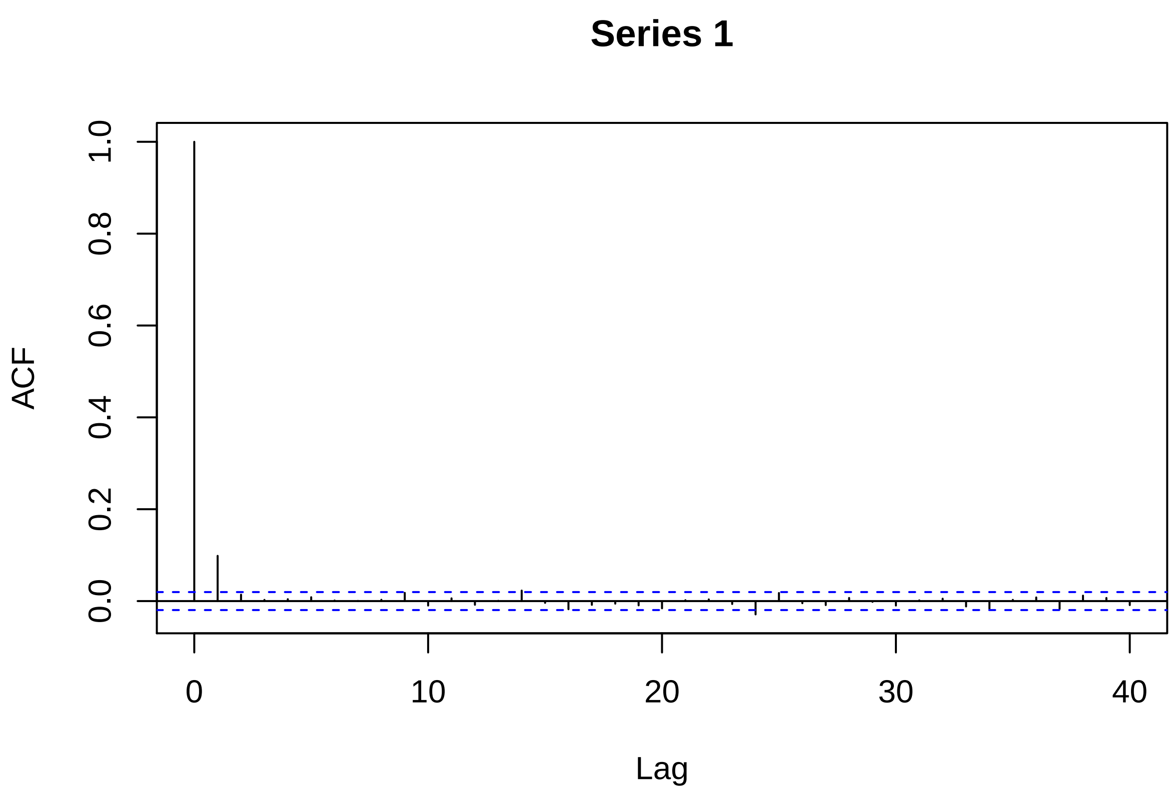
```r
acf(TAU2)
```




```r
print("For mu")
```

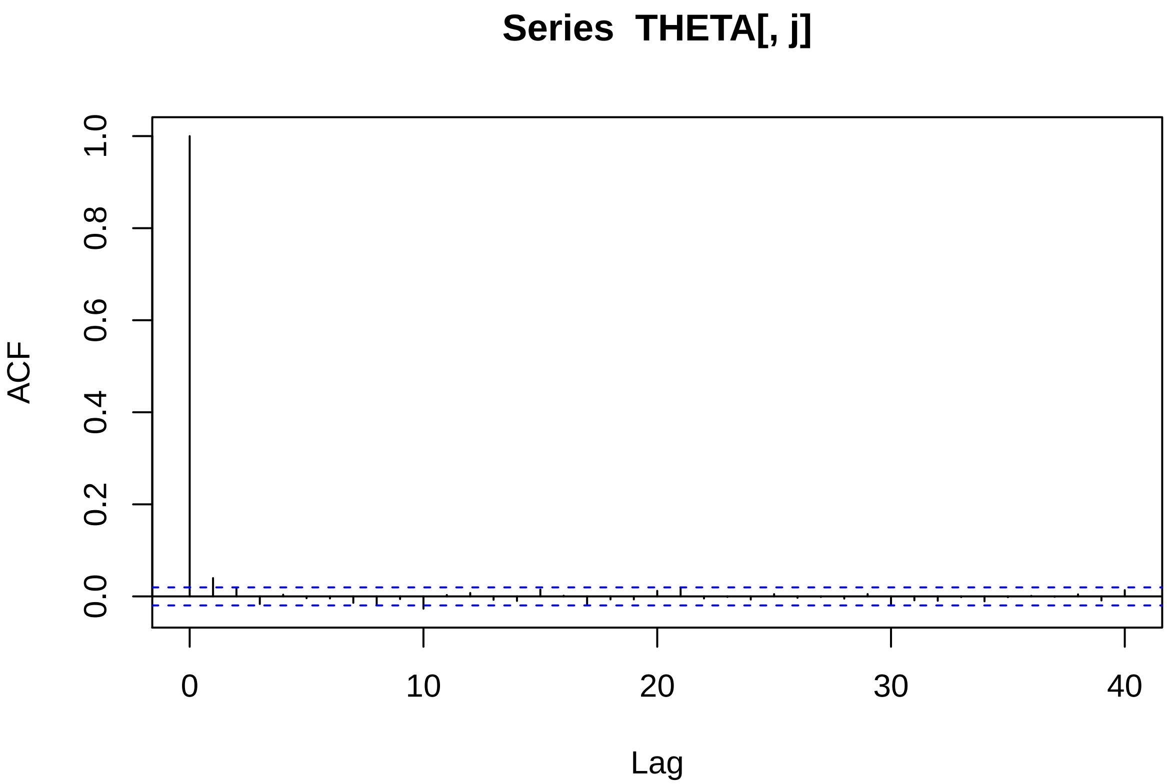
```
## [1] "For mu"
```

```r
acf(MU)
```

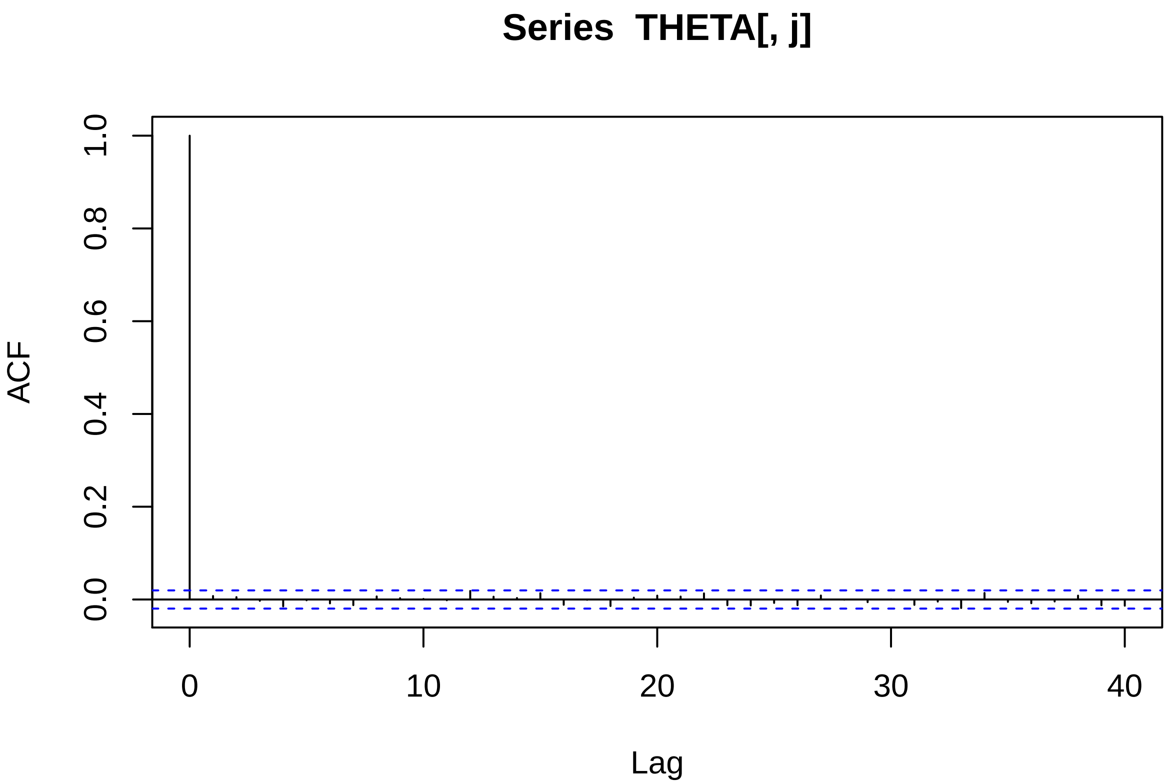


```r
for(j in 1:M){
  print(paste('For theta', toString(j), sep = ""))
  acf(THETA[,j])
}
```

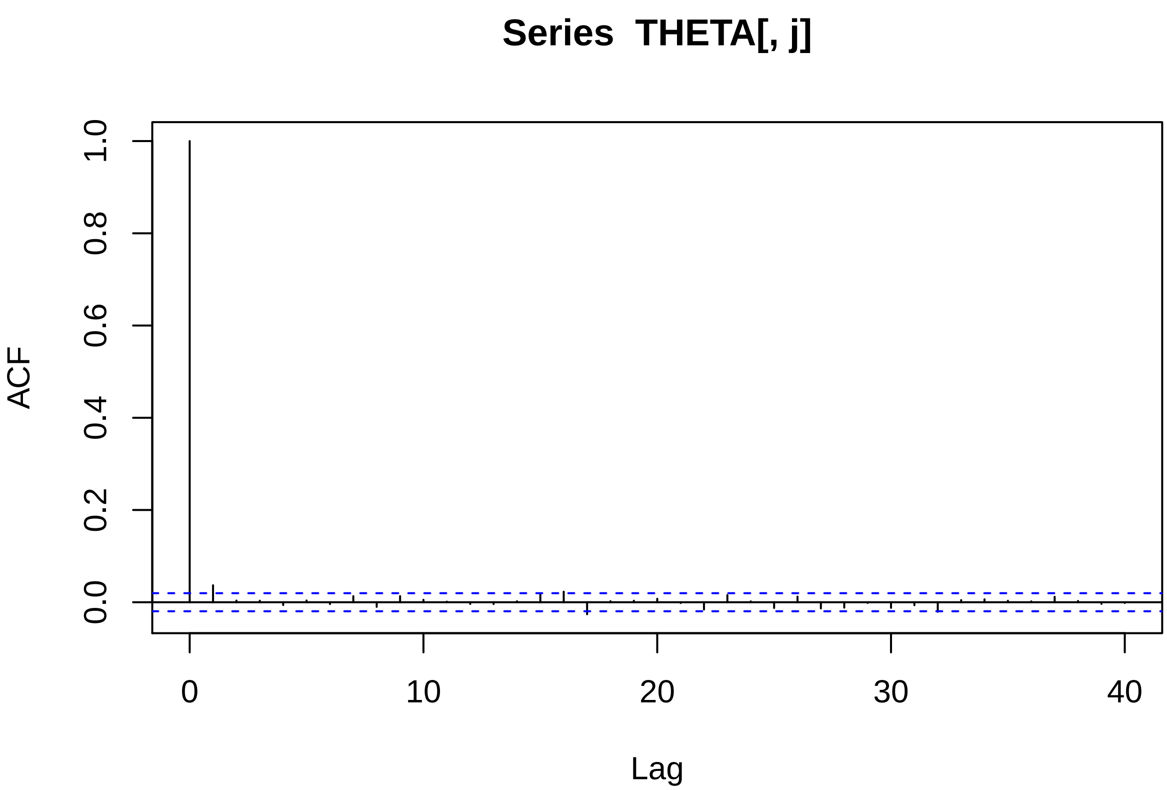
```
## [1] "For theta1"
```



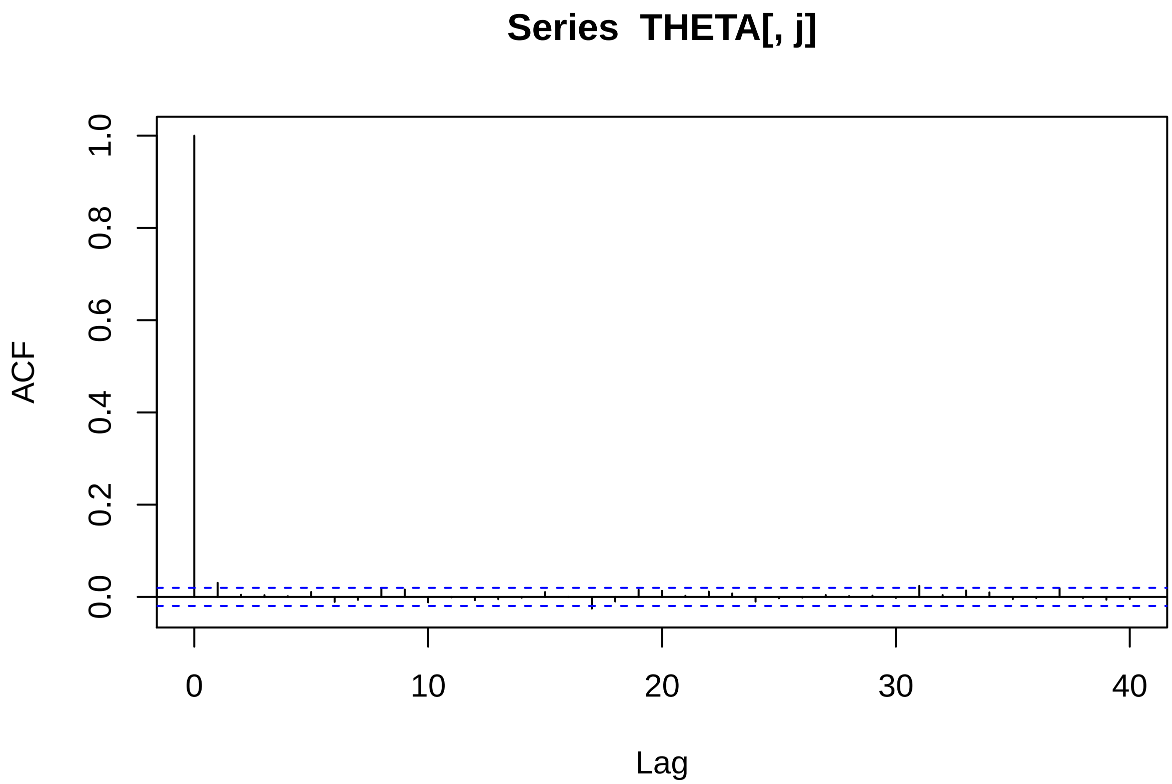

```
## [1] "For theta2"
```



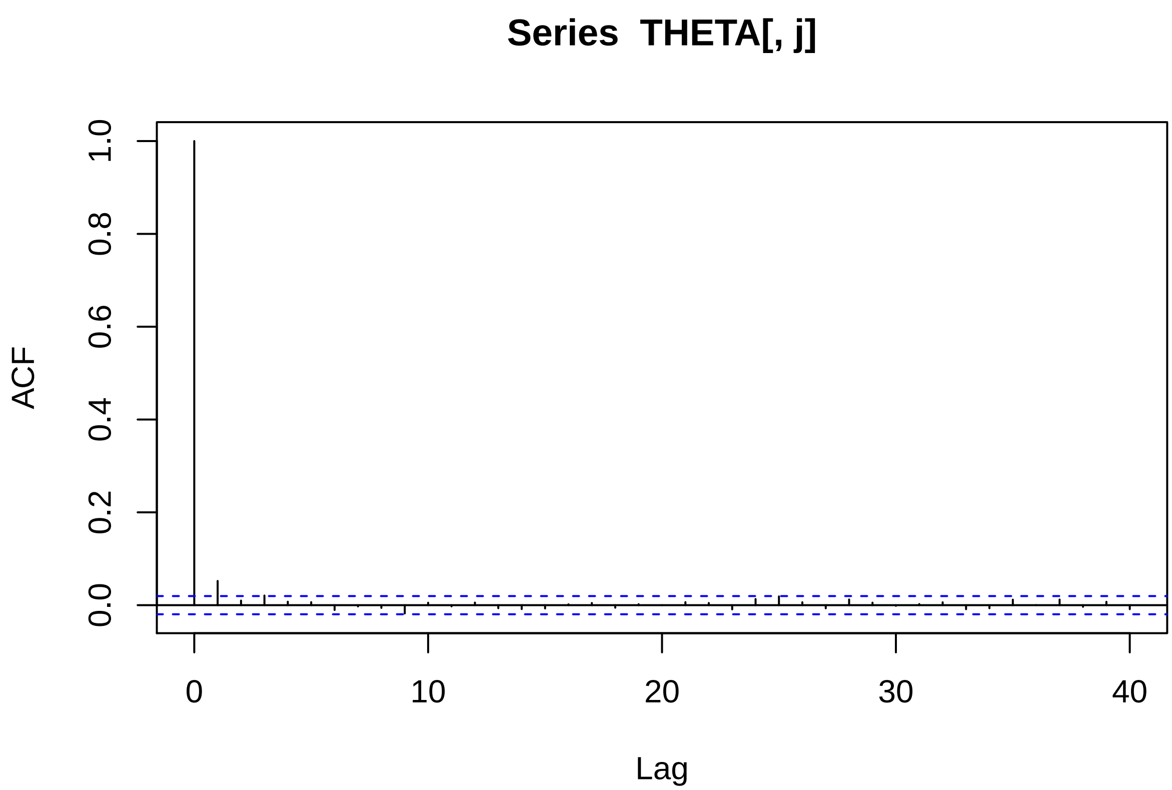

```
## [1] "For theta3"
```



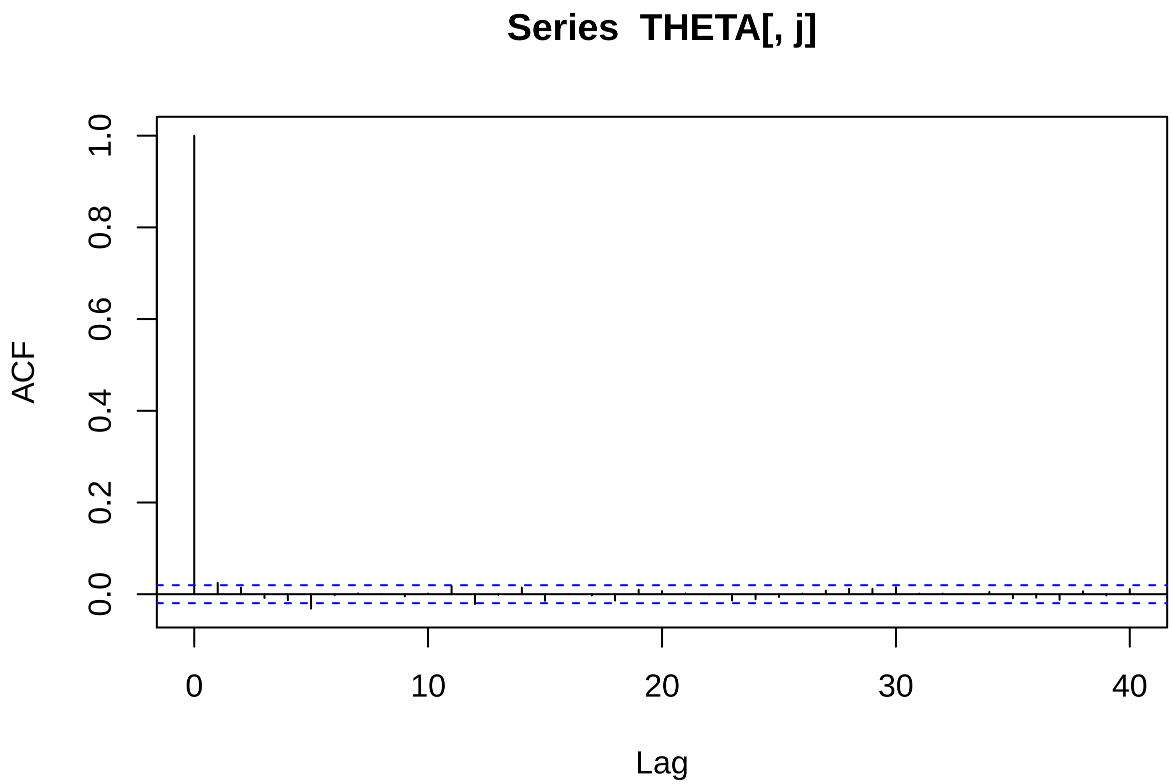

```
## [1] "For theta4"
```



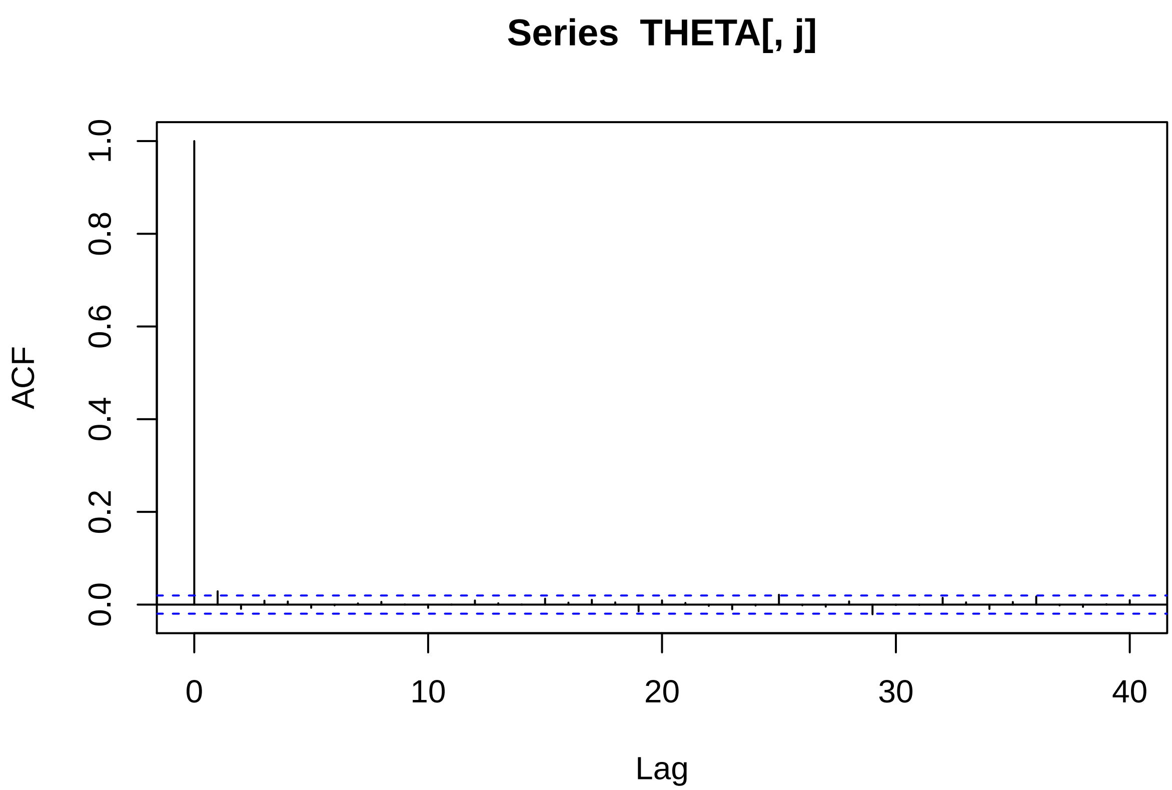

```
## [1] "For theta5"
```



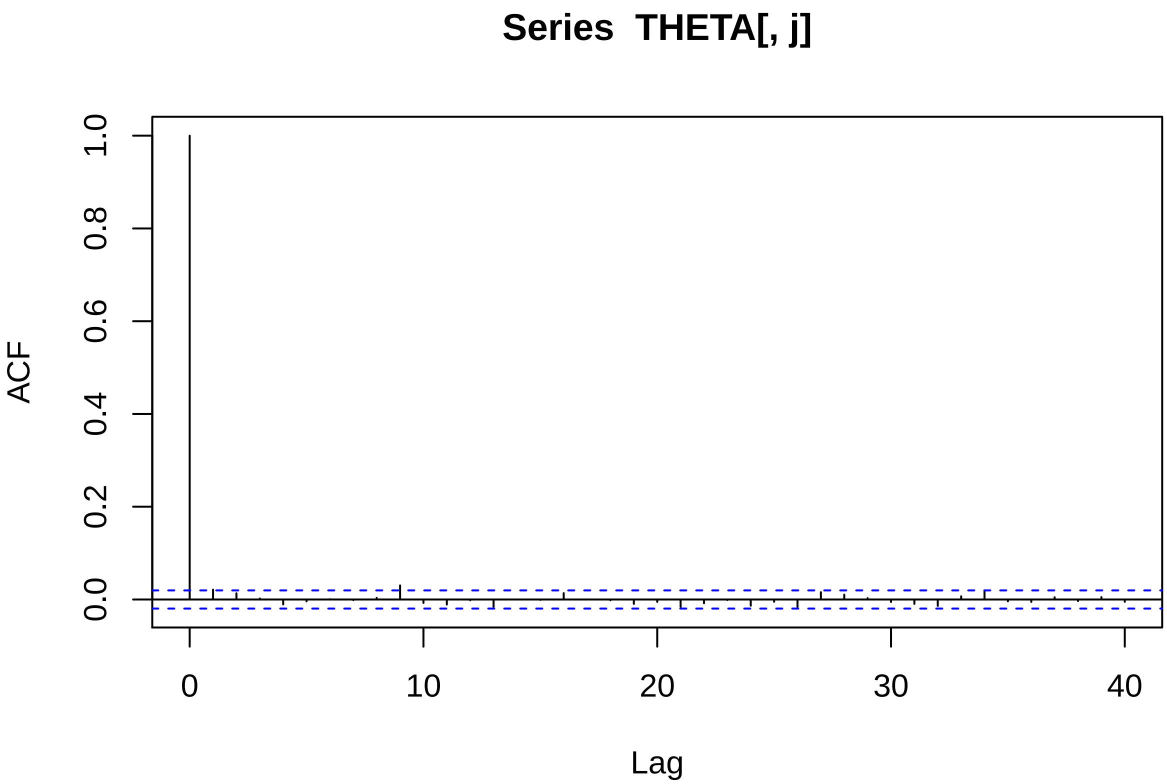

```
## [1] "For theta6"
```




```
## [1] "For theta7"
```




```
## [1] "For theta8"
```



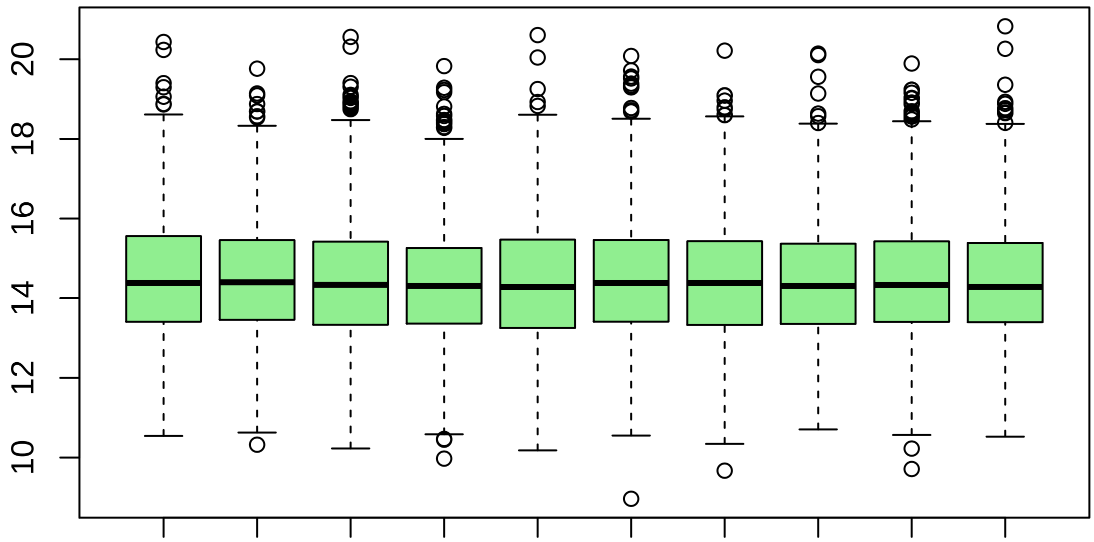
Autocorrelation assessment

As evidenced by the fact that the autocorrelation scores of each variable, the samples I have obtained see to be have a very weak correlation. This makes sence since my effective sample sizes are very close to the actual number of samples I took.

Mixing

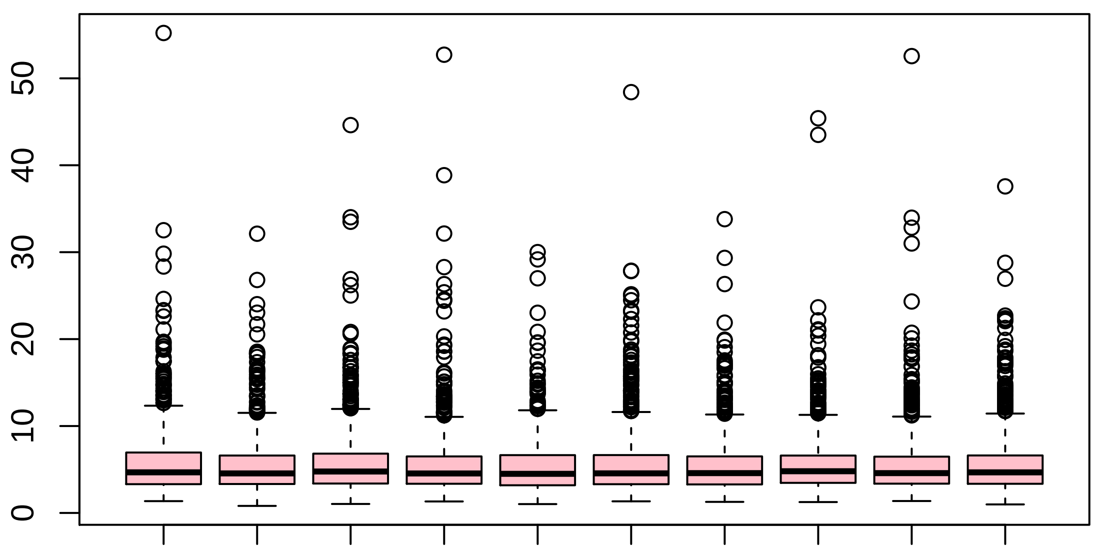
```
boxplot(SIGMA2[1:1000],SIGMA2[1001:2000],SIGMA2[2001:3000], SIGMA2[3001:4000],
        SIGMA2[4001:5000], SIGMA2[5001:6000], SIGMA2[6001:7000],SIGMA2[7001:8000],
        SIGMA2[8001:9000], SIGMA2[9001:10000], main = 'Stationarity Plots for sigma^2',
        col = 'lightgreen')
```

Stationarity Plots for sigma^2

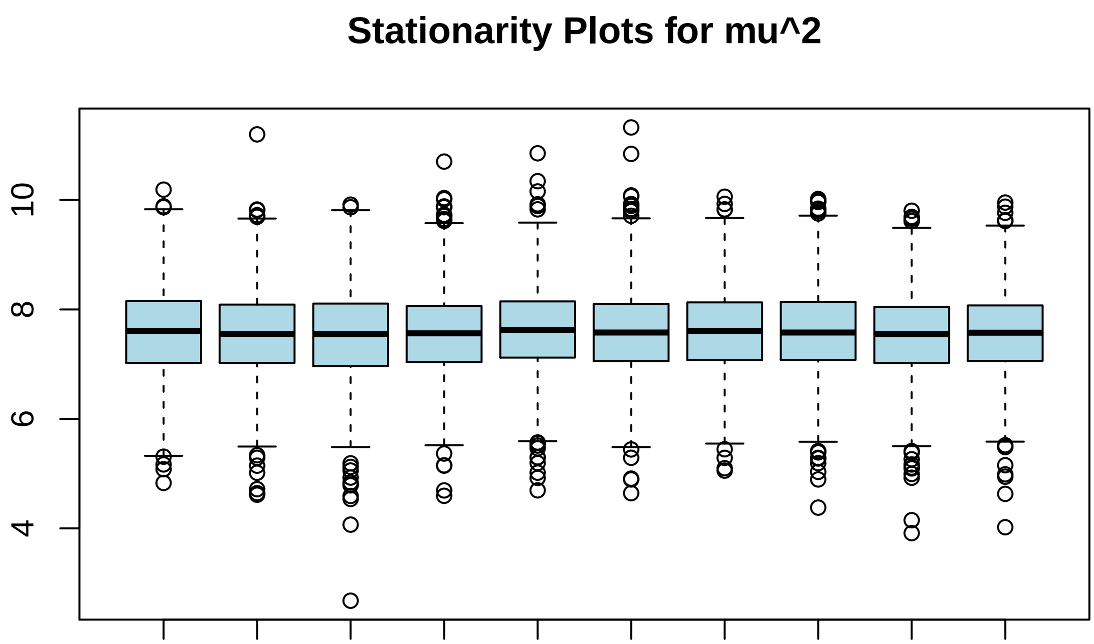



```
boxplot(TAU2[1:1000],TAU2[1001:2000],TAU2[2001:3000], TAU2[3001:4000],
        TAU2[4001:5000], TAU2[5001:6000], TAU2[6001:7000],TAU2[7001:8000],
        TAU2[8001:9000], TAU2[9001:10000], main = 'Stationarity Plots for tau^2',
        col = 'pink')
```

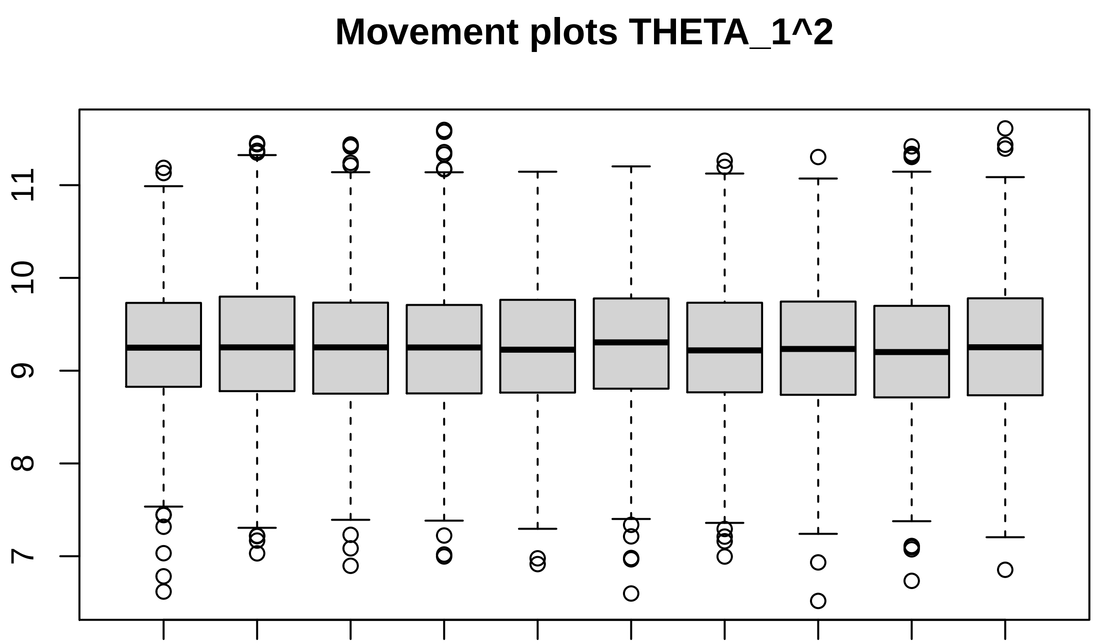
Stationarity Plots for tau^2



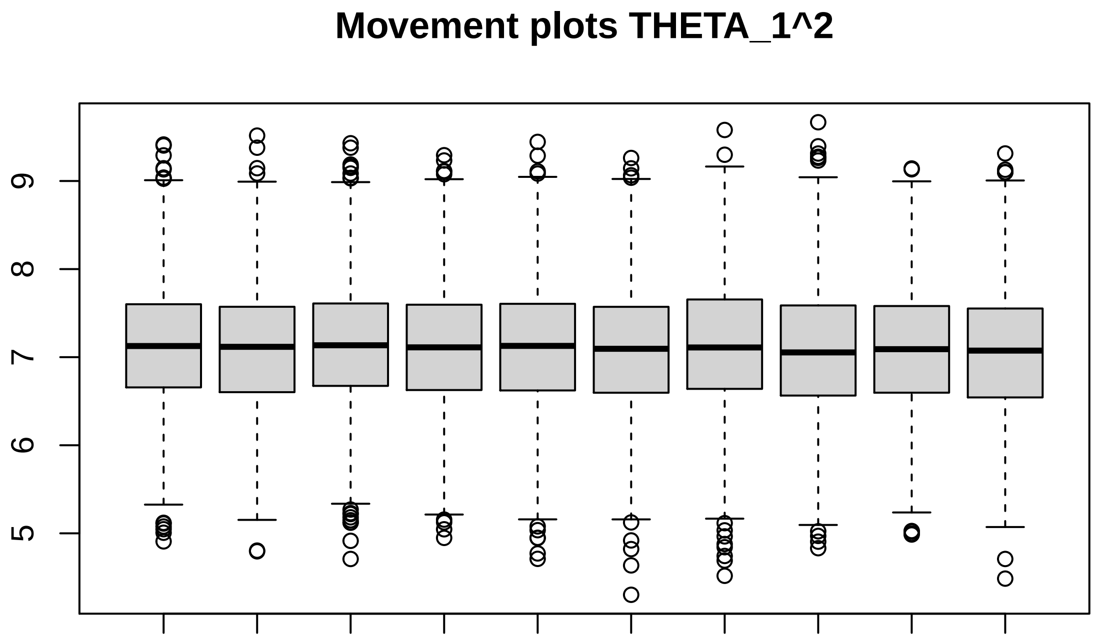

```
boxplot(MU[1:1000],MU[1001:2000],MU[2001:3000], MU[3001:4000],
        MU[4001:5000], MU[5001:6000], MU[6001:7000],MU[7001:8000],
        MU[8001:9000], MU[9001:10000], main = 'Stationarity Plots for mu^2',
        col = 'lightblue')
```



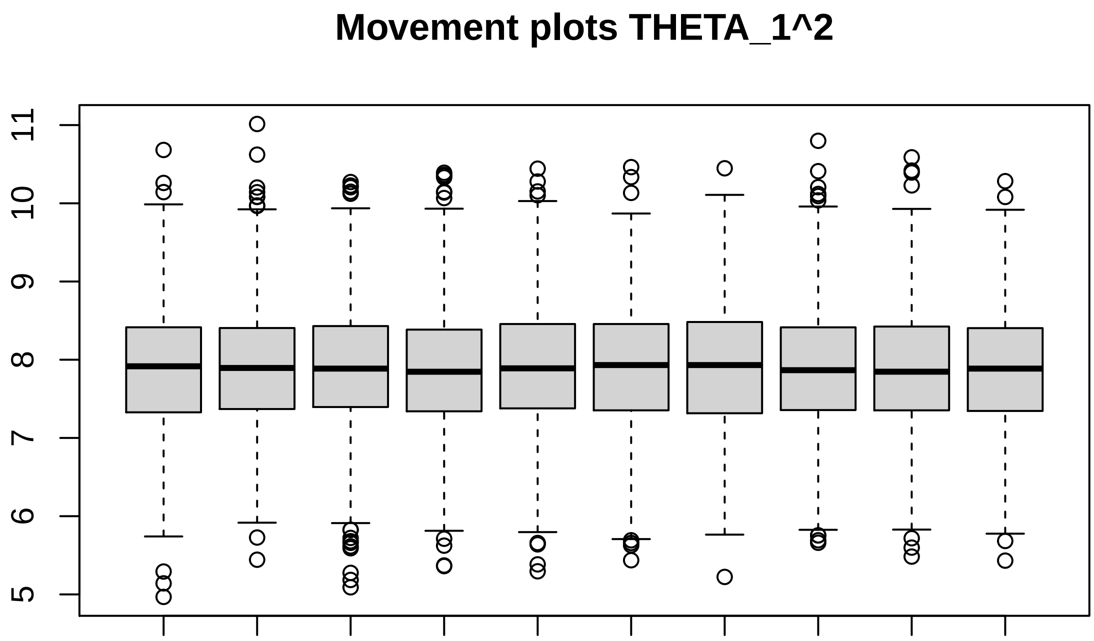

```
boxplot(THETA[1:1000,1],THETA[1001:2000,1],THETA[2001:3000,1], THETA[3001:4000,1],
    THETA[4001:5000,1], THETA[5001:6000,1], THETA[6001:7000,1],THETA[7001:8000,1],
    THETA[8001:9000,1], THETA[9001:10000,1], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```



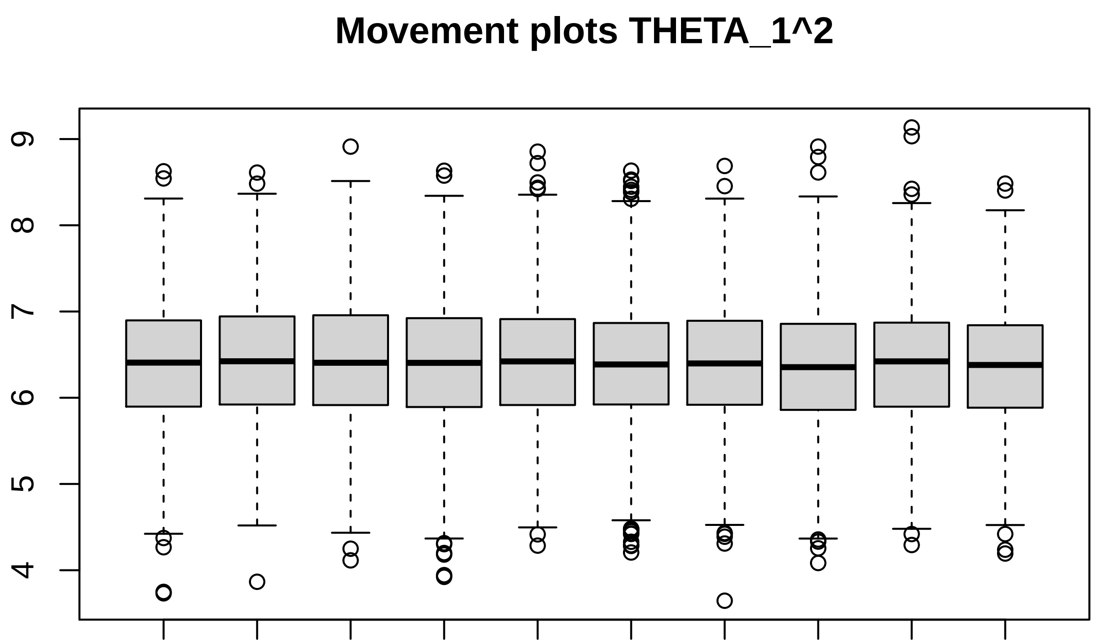

```
boxplot(THETA[1:1000,2],THETA[1001:2000,2],THETA[2001:3000,2], THETA[3001:4000,2],
    THETA[4001:5000,2], THETA[5001:6000,2], THETA[6001:7000,2],THETA[7001:8000,2],
    THETA[8001:9000,2], THETA[9001:10000,2], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```



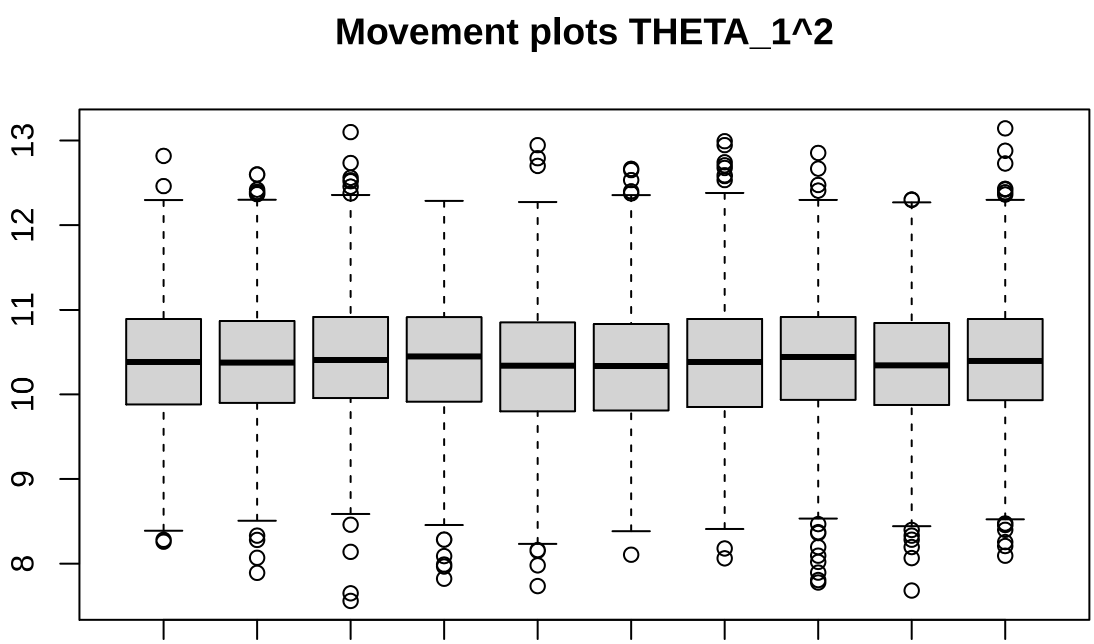

```r
boxplot(THETA[1:1000,3],THETA[1001:2000,3],THETA[2001:3000,3], THETA[3001:4000,3],
    THETA[4001:5000,3], THETA[5001:6000,3], THETA[6001:7000,3],THETA[7001:8000,3],
    THETA[8001:9000,3], THETA[9001:10000,3], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```



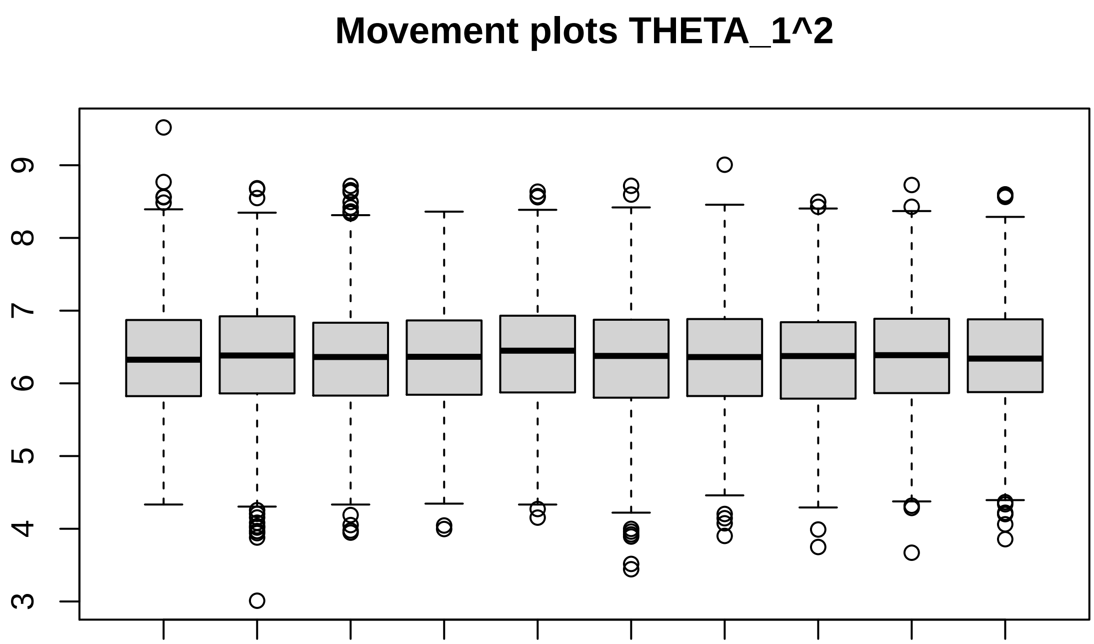

```r
boxplot(THETA[1:1000,4],THETA[1001:2000,4],THETA[2001:3000,4], THETA[3001:4000,4],
    THETA[4001:5000,4], THETA[5001:6000,4], THETA[6001:7000,4],THETA[7001:8000,4],
    THETA[8001:9000,4], THETA[9001:10000,4], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```



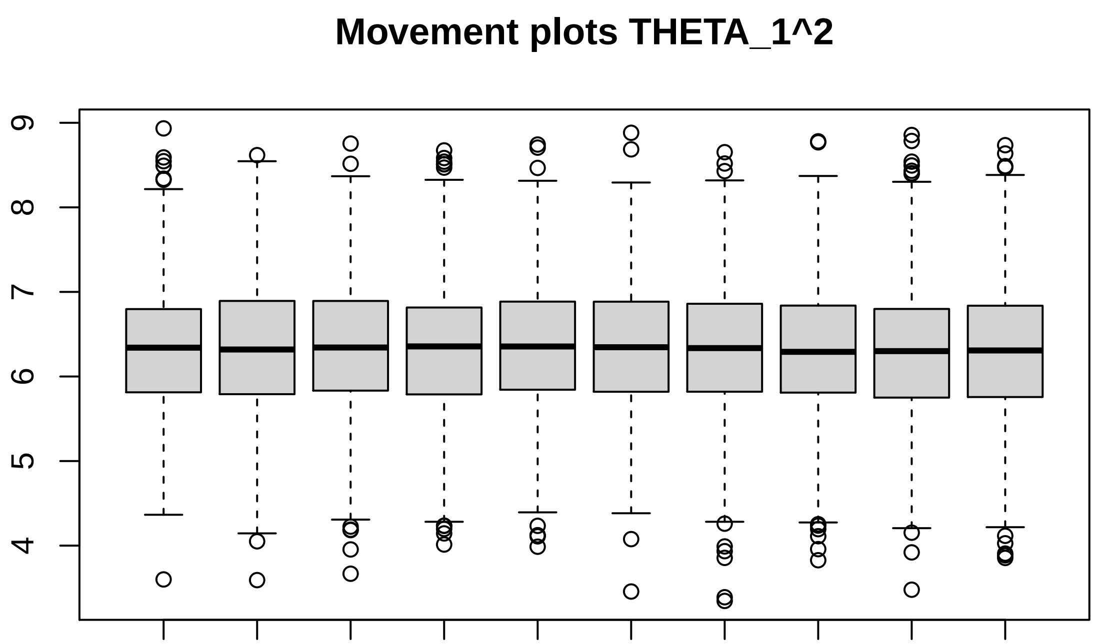

```
boxplot(THETA[1:1000,5],THETA[1001:2000,5],THETA[2001:3000,5], THETA[3001:4000,5],
    THETA[4001:5000,5], THETA[5001:6000,5], THETA[6001:7000,5],THETA[7001:8000,5],
    THETA[8001:9000,5], THETA[9001:10000,5], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```



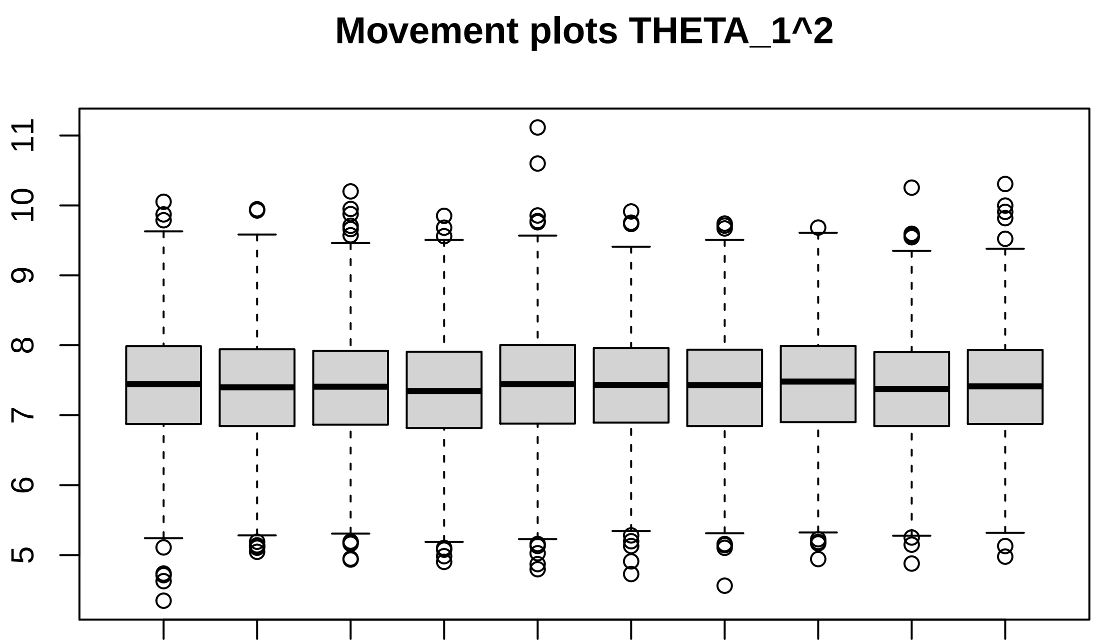

```
boxplot(THETA[1:1000,6],THETA[1001:2000,6],THETA[2001:3000,6], THETA[3001:4000,6],
    THETA[4001:5000,6], THETA[5001:6000,6], THETA[6001:7000,6],THETA[7001:8000,6],
    THETA[8001:9000,6], THETA[9001:10000,6], main = paste('Movement plots THETA_',toString(1), '^2', se
    col = 'lightgrey')
```




```r
boxplot(THETA[1:1000,7],THETA[1001:2000,7],THETA[2001:3000,7], THETA[3001:4000,7],
    THETA[4001:5000,7], THETA[5001:6000,7], THETA[6001:7000,7],THETA[7001:8000,7],
    THETA[8001:9000,7], THETA[9001:10000,7], main = paste('Movement plots THETA_',toString(1), '^2', sep
    col = 'lightgrey')
```




```r
boxplot(THETA[1:1000,8],THETA[1001:2000,8],THETA[2001:3000,8], THETA[3001:4000,8],
    THETA[4001:5000,8], THETA[5001:6000,8], THETA[6001:7000,8],THETA[7001:8000,8],
    THETA[8001:9000,8], THETA[9001:10000,8], main = paste('Movement plots THETA_',toString(1), '^2', sep
    col = 'lightgrey')
```



Convergence assessment

As shown in the box plots above, all of my parameters seem to have reached stationarity since the distribution doesn't seem to be changin over time, as evidenced by the similarities between the box plots created from samples over different time intervals.

8.3b

For σ^2

```
mean(SIGMA2[,1]);
```

```
## [1] 14.4556
```

```
quantile(SIGMA2[,1], c(.025, .975));
```

```
##     2.5%    97.5%
## 11.70546 17.80369
```

For τ^2

```
mean(TAU2[,1]);
```

```
## [1] 5.535315
```

```
quantile(TAU2[,1], c(.025, .975))
```

```
##     2.5%    97.5%
##  1.92748 14.50586
```

For μ

```
mean(MU);
```

```
## [1] 7.569696
```

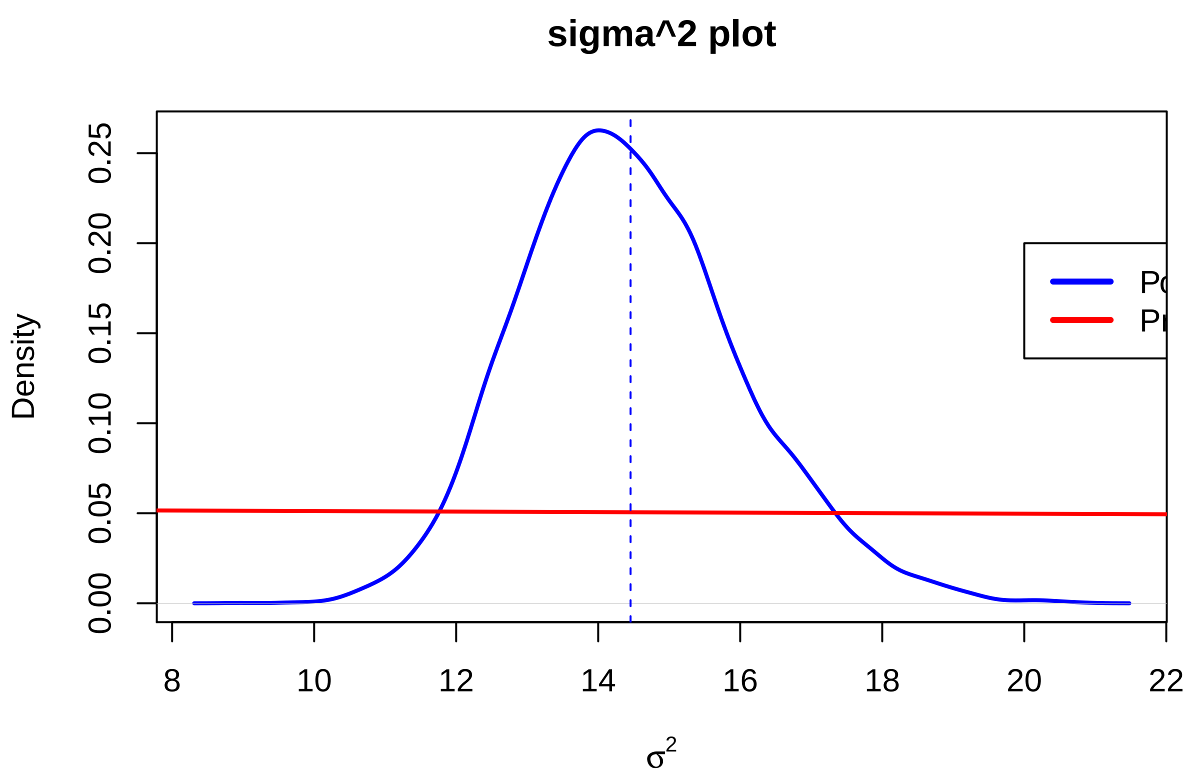
```
quantile(MU[,1], c(.025, .975))
```

```
##     2.5%    97.5%
## 5.946877 9.142112
```

Densities

For σ^2

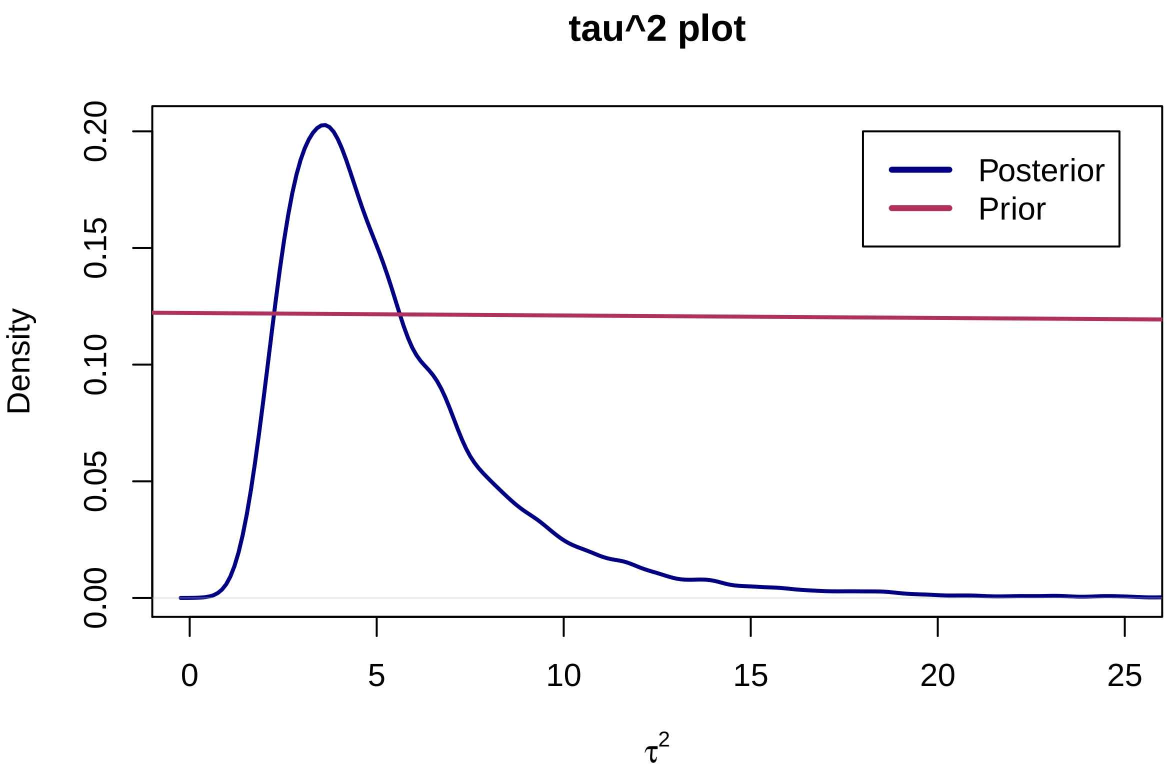
```
priorSigma2 = 1/rgamma(10000,nu0/2,nu0*s2.0/2);
plot(density(SIGMA2), col = 'blue', lwd =2,
     main = 'sigma^2 plot', xlab = expression(sigma^2))
lines(density(priorSigma2),
      col = 'red', lwd = 2)
legend(x = 20, y = .2, legend = c('Posterior', 'Prior'),
       col = c('blue','red'), lty=1, lwd = 3)
abline(v = mean(SIGMA2), col = 'blue', lty = 2)
```



Given that our prior was quite diffuse, we learned quite a bit about the distribution of σ^2 from our data. Firstly, we learned that it's distribution has a mean which is slighly lower than that of our prior (about 14.5 as opposed to 15) and we have a better understanding of the variance of σ^2 – given the posterior distribution, it seems likely that the true value of σ^2 lies between 11 and 18 (as evidenced by the convidence interval above).

For τ^2

```r
priorTau2 = 1/rgamma(10000, nu0/2, nu0*t2.0/2);
plot(density(TAU2), col = 'navy', lwd =2, main = 'tau^2 plot',
     xlim = c(0,25), xlab = expression(tau^2))
lines(density(priorTau2), col = 'maroon', lwd = 2)
legend(x = 18, y = .2, legend = c('Posterior', 'Prior'),
       col = c('navy','maroon'), lty=1, lwd = 3)
```

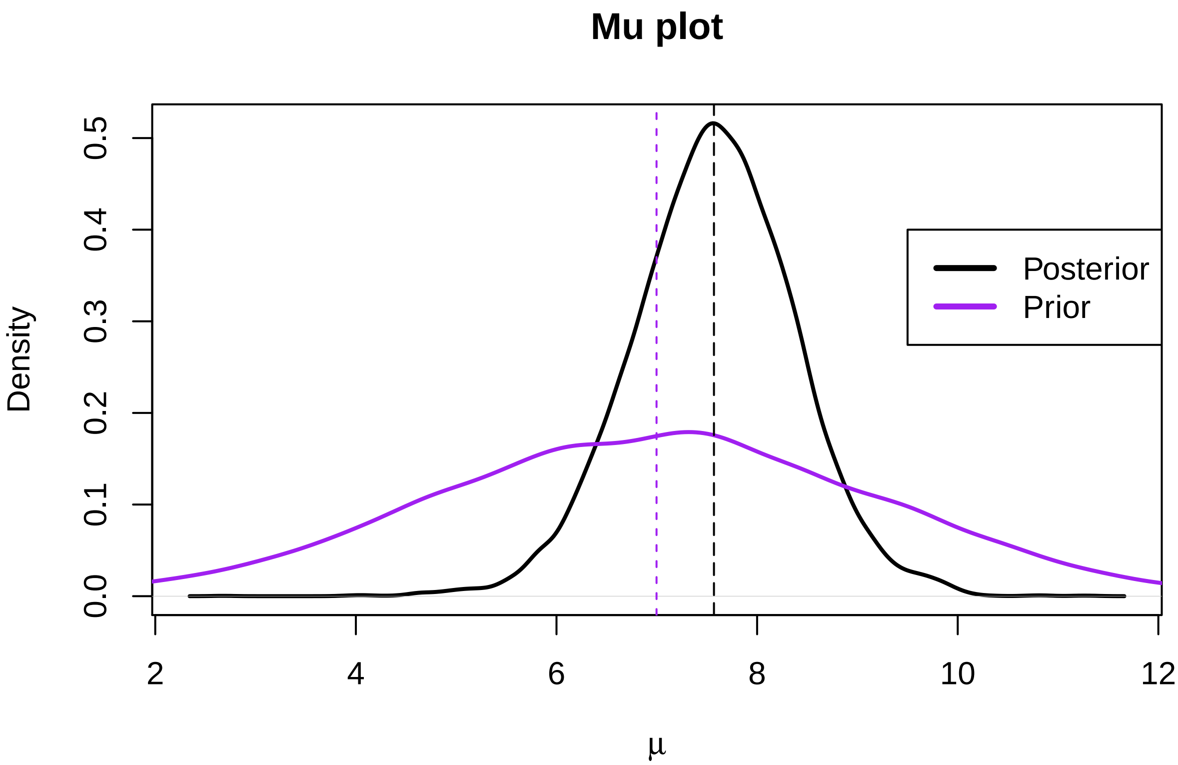


The information we learned about τ^2 is similar to what we learned about σ^2, however the mean of τ^2. Furthermore, since the prior distribution of τ^2 was "flat", our posterior distribution allows us to see the spread of the values τ^2 is likely to equal.

For mu

```r
priorDist = rnorm(10000,mu0,sqrt(g2.0))
plot(density(MU), col = 'black', lwd =2,
     main = 'Mu plot', xlab = expression(mu))
lines(density(priorDist), col = 'purple', lwd = 2)
legend(x = 9.5, y = .4, legend = c('Posterior', 'Prior'),col = c('black','purple'), lty=1, lwd = 3)
```

```r
abline(v = mean(priorDist), lty = 2, col = 'purple')
abline(v = mean(MU), lty = 5, col = 'black')
```

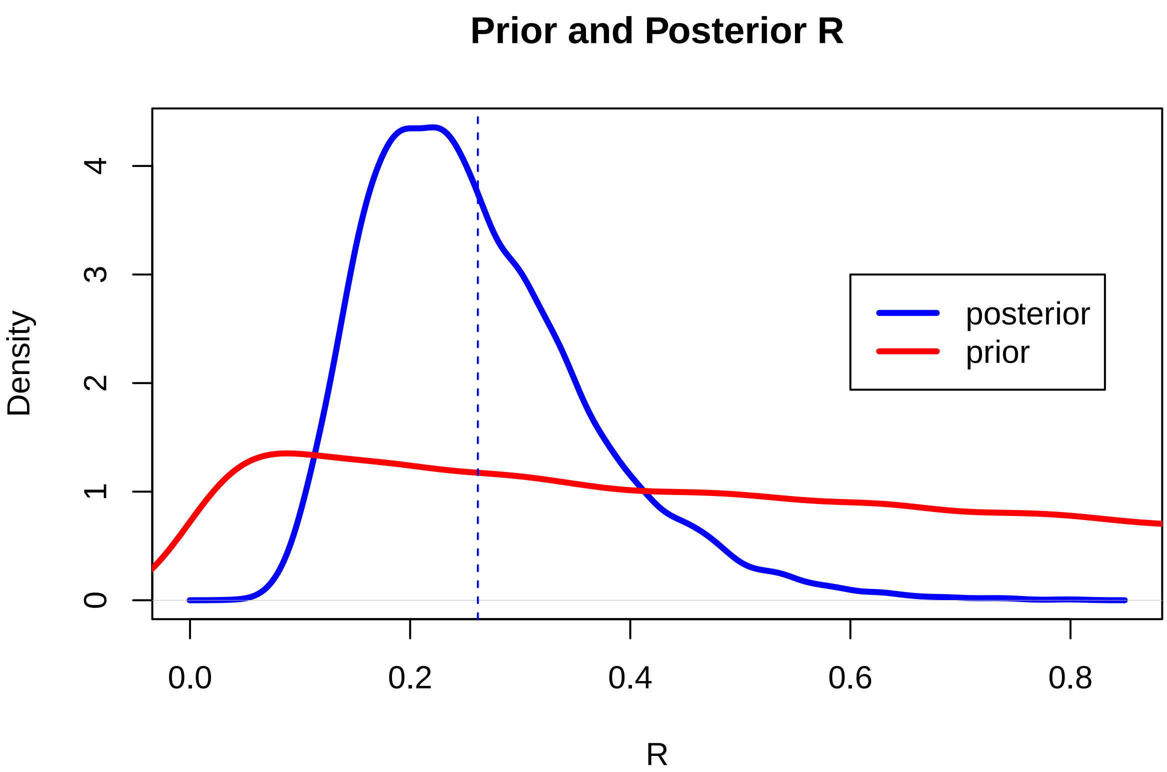


The posterior distribution of μ suggests that the mean of μ is larger than previously expected (this is evidenced by the dashed lines which indicate the mean of the respective distributions) and the between group variance is smaller than previously expected (depicted by the "sharper" peak of the posterior).

8.3c

Let $R = \frac{\tau^2}{\tau^2+\sigma^2}$

```r
R = TAU2/(SIGMA2 + TAU2);
priorR = priorTau2/(priorSigma2+priorTau2);
plot(density(R), col = 'blue', lwd = 3,
     main = "Prior and Posterior R", xlab = "R")
lines(density(priorR), col = 'red', lwd = 3)
legend(x = .6, y = 3, legend = c('posterior', 'prior'),
       col = c('blue','red'),lty = 1, lwd = 3)
abline(v = mean(R), lty = 2, col = 'blue')
```



Our prior belief leads us to believe that between-school variation could be quite large, quite small, and everywhere in between (it is essentially flat). However, the posterior distribution of R is centered roughly within the interval $(.25, .3)$, suggesting that τ^2 is nonzero and is about half as large as σ^2, implying between-group variation.

8.3d

$P(\theta_7 < \theta_6)$

```
mean(THETA[,7]< THETA[,6])
```

```
## [1] 0.5124
```

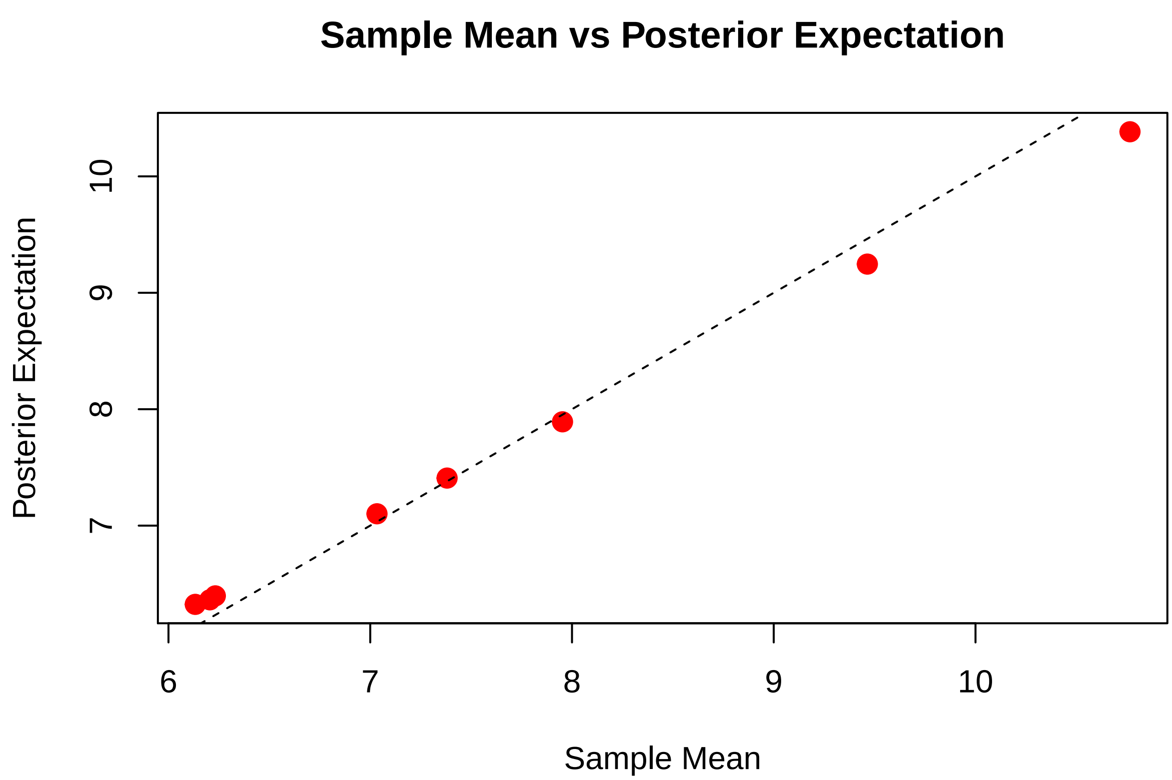
$P(\min(\theta_1, ..., \theta_8) = \theta_7)$

```
mean(apply(THETA,1, min) == THETA[,7])
```

```
## [1] 0.3166
```

8.3e

```
postE = apply(THETA,2,mean)
plot(yBar,postE, col = 'red', pch = 20, cex = 2,
     main = "Sample Mean vs Posterior Expectation", ylab = "Posterior Expectation",
     xlab = 'Sample Mean')
abline(a = 0, b = 1, lty = 2)
```



It appears that as our sample mean becomes larger, the sample mean tends to be larger than the posterior expectation of the test scores, while samples with lower observed means are lower than their corresponding posterior expectations.

Cumulative Average value of Y and posterior μ

```
posteriorMu = mean(MU[,1])
posteriorMu
```

```
## [1] 7.569696
```

```
overallY = N%*%yBar/sum(N)
overallY
```

```
##          [,1]
## [1,] 7.691278
```

The posterior expectation of μ is slightly lower than the mean of all my observations. This is due to the fact that my prior $\mu_0 = 7$, which seems to have pulled my posterior expecttion down a bit.