STA 360: Lab 11

Michael Lin

April 17, 2015

1. Let w_i be the indicator for whether observation is from a weekday or weekend. In particular, $w_i = 1$ with probability p if i is weekday, and $w_i = 0$ with probability $1 - p$. In other words,

$$w_i \sim \text{Be}(p)$$

Henceforth, let ξ denote the p.d.f. Given $p \sim \text{Beta}(5, 2)$, the posterior p.d.f. of p is:

$$\begin{aligned}
\xi(p|w_{1:n}) &\propto \xi(w_{1:n}|p)\xi(p) \\
&= \Big(\prod_{i=1}^{n} p^{w_i}(1-p)^{1-w_i}\Big)\frac{\Gamma(7)}{\Gamma(5)\Gamma(2)}p^4(1-p) \\
&\propto p^{\sum w_i}(1-p)^{n-\sum w_i}p^4(1-p) \\
&= p^{(5+\sum w_i)-1}(1-p)^{(n+b-\sum w_i)-1} \\
&\propto \text{Beta}(p|5 + \sum w_i, n + 2 - \sum w_i)
\end{aligned}$$

Recall the following distributions:

$$\begin{aligned}
y_i|(w_i = 1) &\sim \text{LN}(\mu_1, \sigma_1^2) \\
y_i|(w_i = 0) &\sim \text{LN}(\mu_2, \sigma_2^2)
\end{aligned}$$

Thus by Bayes' Rule:

$$\mathbb{P}(w_i = 1|y_i, p) = \frac{\frac{1}{y_i\sqrt{2\pi\sigma_1^2}}\exp\{-(\log y_i - \mu_1)^2/2\sigma_1^2\}p}{\frac{1}{y_i\sqrt{2\pi\sigma_1^2}}\exp\{-(\log y_i - \mu_1)^2/2\sigma_1^2\}p + \frac{1}{y_i\sqrt{2\pi\sigma_2^2}}\exp\{-(\log y_i - \mu_2)^2/2\sigma_1^2\}(1-p)}$$

Define $\pi_i = \mathbb{P}(w_i = 1|y_i, p)$. For each iteration of the Gibb's sampler, we would sample:

$$w_i \sim \text{Be}(\pi_i)$$

Then we would sample $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ as we did in lab 10.

See below for derivation of full conditionals from lab 10:

From last time, we saw that given priors:

$$p(\mu) = \text{Normal}(\mu \mid \mu_0, \sigma_0^2)$$
$$p(\sigma^2) = IG(\sigma^2 \mid \alpha, \beta)$$

The posteriors are:

$$p(\mu \mid x_{1:n}) = \text{Normal}\left(\mu \,\middle|\, \frac{\mu_0\sigma^2 + \sigma_0^2 \sum \ln x_i}{\sigma^2 + n\sigma_0^2}, \frac{\sigma_0^2\sigma^2}{\sigma^2 + n\sigma_0^2}\right)$$

$$p(\sigma^2 \mid x_{1:n}) = IG\left(\sigma^2 \,\middle|\, \alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_i (\ln x_i - \mu)^2\right)$$

1) Since we want $\mu_1 > \mu_2$, we will draw from these priors

$$\mu_2 \sim No(\mu_2 \mid m_2, s_2^2)$$

$$\mu_1 = \mu_2 + \delta \quad \text{where} \quad \delta \sim No(\delta \mid m_1, s_1^2)\mathbb{1}_{\delta > 0}$$

Thus $\delta = \mu_1 - \mu_2 \sim No(\delta \mid m_1, s_1^2)\mathbb{1}_{\delta > 0}$

For σ_1 and σ_2, we have

$$\sigma_1^2 \sim IG(\sigma_1^2 \mid a_1, b_1)$$
$$\sigma_2^2 \sim IG(\sigma_2^2 \mid a_2, b_2)$$

1

$$p(\mu_2 \mid y_2) \propto p(y_2 \mid \mu_2, \sigma_2^2) \cdot p(\mu_2)\, p(\sigma_2^2)$$

$$\propto \prod_{i=1}^{n_2} \exp\left\{-\frac{(\ln(y_{2i})-\mu_2)^2}{2\sigma_2^2}\right\} \cdot \exp\left\{-\frac{(\mu_2-m_2)^2}{2S_2^2}\right\}$$

$$= \exp\left\{-\frac{\sum_{i=1}^{n_2}(\ln(y_{2i})-\mu_2)^2}{2\sigma_2^2}\right\} \cdot \exp\left\{-\frac{(\mu_2-m_2)^2}{2S_2^2}\right\}$$

$$\propto \exp\left\{-\frac{\sum_{i=1}^{n_2}(\mu_2^2 - 2\mu_2\ln(y_{2i}))}{2\sigma_2^2}\right\} \exp\left\{-\frac{\mu_2^2 - 2\mu_2 m_2}{2S_2^2}\right\}$$

$$= \exp\left\{-\frac{n_2\mu_2^2 - 2\mu_2\sum_{i=1}^{n_2}\ln(y_{2i})}{2\sigma_2^2} - \frac{\mu_2^2 - 2\mu_2 m_2}{2S_2^2}\right\}$$

$$= \exp\left\{-\left[\frac{S_2^2(n_2\mu_2^2 - 2\mu_2\sum\ln(y_{2i})) + \sigma_2^2(\mu_2^2 - 2\mu_2 m_2)}{2\sigma_2^2 S_2^2}\right]\right\}$$

$$= \exp\left\{-\frac{S_2^2 n_2\mu_2^2 - 2S_2^2\mu_2\sum_i\ln(y_{2i}) + \sigma_2^2\mu_2^2 - 2\sigma_2^2\mu_2 m_2}{2\sigma_2^2 S_2^2}\right\}$$

$$= \exp\left\{-\frac{(S_2^2 n_2 + \sigma_2^2)\mu_2^2 - 2\left[S_2^2\sum\ln(y_{2i}) + \sigma_2^2 m_2\right]\mu_2}{2\sigma_2^2 S_2^2}\right\}$$

$$= \exp\left\{-\frac{\mu_2^2 - 2\left(\frac{S_2^2\sum\ln(y_{2i}) + \sigma_2^2 m_2}{S_2^2 n_2 + \sigma_2^2}\right)\mu_2}{2\sigma_2^2 S_2^2 / (S_2^2 n_2 + \sigma_2^2)}\right\}$$

$$\propto \mathrm{No}\left(\mu_2 \,\middle|\, \frac{S_2^2\sum\ln(y_{2i}) + \sigma_2^2 m_2}{S_2^2 n_2 + \sigma_2^2}, \frac{\sigma_2^2 S_2^2}{S_2^2 n_2 + \sigma_2^2}\right)$$

$p(\sigma_2^2 \mid y_2) = \cdots$ (See last week's derivation).

$$= IG\left(\sigma_2^2 \,\middle|\, \alpha_2 + \frac{n_2}{2}, \beta_2 + \frac{1}{2}\sum_{i=1}^{n_2}(\ln y_{2i} - \mu_2)^2\right)$$

Similarly,

$$p(\sigma_1^2 \mid y_1) = IG\left(\sigma_1^2 \,\middle|\, \alpha_1 + \frac{n_1}{2}, \beta_1 + \frac{1}{2}\sum_{i=1}^{n_1}(\ln y_{1i} - \mu_1)^2\right)$$

(2)

$$p(\delta \mid y_1) \propto p(y_1 \mid \mu_1, \sigma_1^2)\, p(\delta) \cdot p(\sigma_1^2)$$

$$\propto \prod_{i=1}^{n_1} \exp\left\{-\frac{(\ln(y_{1i}) - \mu_1)^2}{2\sigma_1^2}\right\} \cdot \exp\left\{-\frac{(\delta - m_1)^2}{2s_1^2}\right\} \mathbb{1}_{\delta>0}$$

$$= \prod_{i=1}^{n_1} \exp\left\{-\frac{(\ln y_{1i} - (\delta + \mu_2))^2}{2\sigma_1^2}\right\} \exp\left\{-\frac{(\delta - m_1)^2}{2s_1^2}\right\} \mathbb{1}_{\delta>0}$$

$$\propto \exp\left\{-\frac{n(\delta+\mu_2)^2 - 2(\delta+\mu_2)\sum \ln y_{1i}}{2\sigma_1^2}\right\} \exp\left\{-\frac{\delta^2 - 2m_1\delta}{2s_1^2}\right\} \mathbb{1}_{\delta>0}$$

$$\propto \exp\left\{-\frac{n(\delta^2 + 2\delta\mu_2) - 2\delta \sum_i \ln y_{1i}}{2\sigma_1^2}\right\} \exp\left\{-\frac{\delta^2 - 2m_1\delta}{2s_1^2}\right\} \mathbb{1}_{\delta>0}$$

$$= \exp\left\{-\frac{\delta^2 + 2\delta\mu_2 - 2\delta \frac{\sum \ln y_{1i}}{n_1}}{2\sigma_1^2/n_1} - \frac{\delta^2 - 2m_1\delta}{2s_1^2}\right\} \mathbb{1}_{\delta>0}$$

$$= \exp\left\{-\frac{s_1^2\delta^2 + 2\delta\mu_2 s_1^2 - 2\delta s_1^2 \frac{\sum \ln y_{1i}}{n_1} + \left(\delta^2\frac{\sigma_1^2}{n_1} - 2m_1\delta\frac{\sigma_1^2}{n_1}\right)}{2\sigma_1^2 s_1^2/n_1}\right\} \mathbb{1}_{\delta>0}$$

$$= \exp\left\{-\frac{\left(s_1^2 + \frac{\sigma_1^2}{n_1}\right)\delta^2 + 2\left(\mu_2 s_1^2 - s_1^2\frac{\sum \ln y_{1i}}{n_1} - m_1\frac{\sigma_1^2}{n_1}\right)\delta}{2\sigma_1^2 s_1^2/n_1}\right\} \mathbb{1}_{\delta>0}$$

$$= \exp\left\{-\frac{\left(s_1^2 + \frac{\sigma_1^2}{n_1}\right)\delta^2 - 2\left(\frac{s_1^2 \sum \ln y_{1i} + m_1\sigma_1^2 - n_1\mu_2 s_1^2}{n_1}\right)\delta}{2\sigma_1^2 s_1^2/n_1}\right\} \mathbb{1}_{\delta>0}$$

$$= \exp\left\{-\frac{\delta^2 - 2\left(\frac{s_1^2 \sum_i \ln y_{1i} + m_1\sigma_1^2 - n_1\mu_2 s_1^2}{n_1\left(s_1^2 + \frac{\sigma_1^2}{n_1}\right)}\right)\delta}{2\frac{\sigma_1^2 s_1^2}{n_1\left(s_1^2 + \frac{\sigma_1^2}{n_1}\right)}}\right\} \mathbb{1}_{\delta>0}$$

$$\propto \mathrm{No}\left(\delta \,\middle|\, \frac{s_1^2 \sum \ln y_{1i} + m_1\sigma_1^2 - n_1\mu_2 s_1^2}{n_1 s_1^2 + \sigma_1^2}, \frac{\sigma_1^2 s_1^2}{n_1 s_1^2 + \sigma_1^2}\right)$$

③

$(\sigma^2)^{-1/2}$

$$p(\sigma^2 \mid X_{1:n}) = p(\sigma^2)\, p(X_{1:n} \mid \mu, \sigma^2),$$

$$\propto \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(\frac{-\beta}{\sigma^2}\right) \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(\ln x_i - \mu)^2}{2\sigma^2}\right\}$$

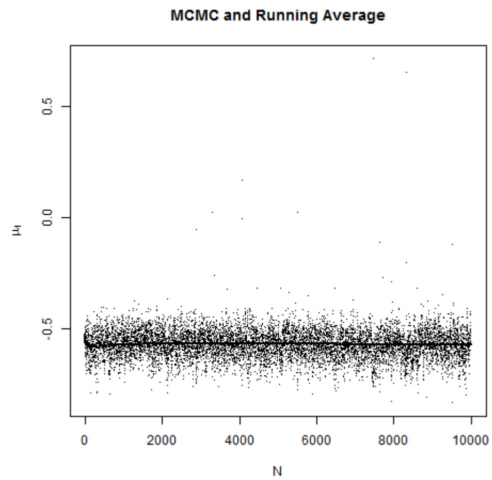
$$\propto (\sigma^2)^{-\alpha-1} \cdot (\sigma^2)^{-n/2} \exp\left\{-\frac{2\beta + \sum_{i=1}^{n} (\ln(x_i) - \mu)^2}{2\sigma^2}\right\}$$

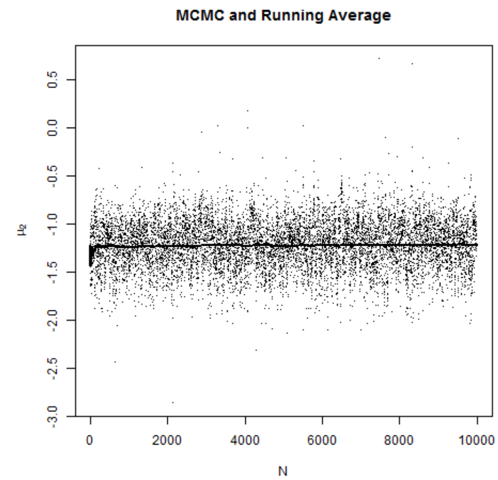
$$\propto (\sigma^2)^{-\alpha-n/2-1} \exp\left\{-\frac{2\beta + \sum_{i=1}^{n} (\ln x_i - \mu)^2}{2\sigma^2}\right\}$$

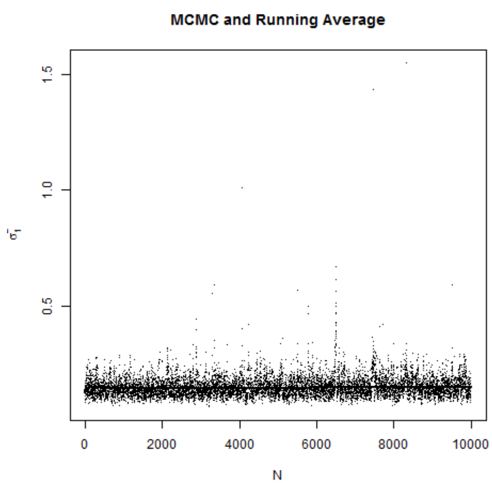
$$= (\sigma^2)^{-(\alpha+n/2)-1} \exp\left\{-\frac{\beta + \frac{1}{2}\sum_{i=1}^{n} (\ln x_i - \mu)^2}{\sigma^2}\right\}$$

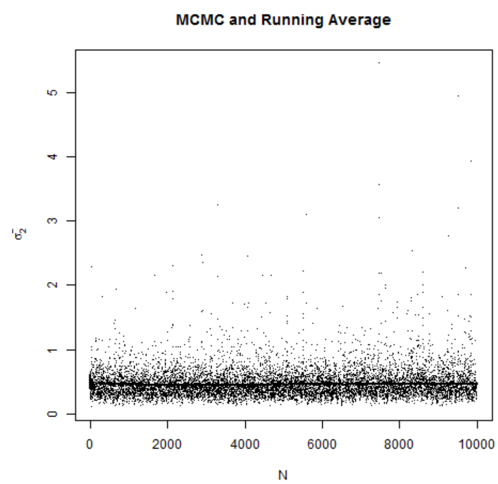
$$= IG\left(\sigma^2 \,\middle|\, \alpha + \frac{n}{2},\ \beta + \frac{1}{2}\sum_{i=1}^{n} (\ln x_i - \mu)^2\right)$$

2. The sampler has converged. See traceplots and running averages below:

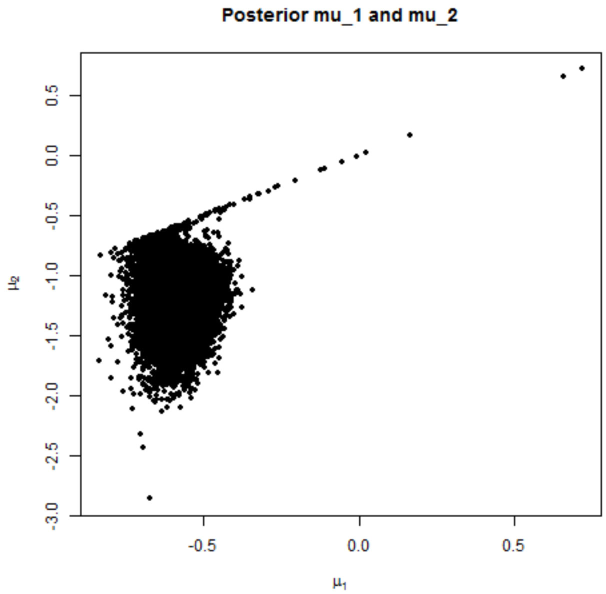








3. Below is a plot of μ_2 vs. μ_1:



4.
   - Number of days in the sample which are weekdays: 77.6965. 95% posterior credible interval is [58, 92].
   - Probability that the technician is coming in less often on weekends than on weekdays: 0.7908.

5. Below are point estimates and 95% confidence intervals:

Lab 11	Mean	Confidence Interval
μ_1	-0.5728062	[-0.7031991, -0.4558429]
μ_2	-1.218221	[-1.7240026, -0.7504473]
σ_1^2	0.1494581	[0.09059487, 0.24865874]
σ_2^2	0.46506	[0.1903877, 1.0277780]

Below are point estimates and 95% confidence intervals from lab 10:

Lab 10	Mean	Confidence Interval
μ_1	-0.4708718	[-0.5466545, -0.3949325]
μ_2	-1.377839	[-1.5240060, -1.2294750]
σ_1^2	0.1104881	[0.07978934, 0.15334020]
σ_2^2	0.1597617	[0.09405123, 0.26961285]

In general, the point estimates of parameters are comparable between results from lab 10 and lab 11, but the variance is larger in lab 11. However, the point estimates for σ_2^2 is very different between the two samplers.

See below for R code:

```
## load data ##
data = read.table("data1.txt", header = T)
library(MCMCpack)
library(truncnorm)
data$weekday = rbinom(100,1,0.5)

X = data
Y1 = subset(X, weekday==1)
row.names(Y1) = NULL
Y1 = Y1[,1]
Y2 = subset(X, weekday==0)
row.names(Y2) = NULL
Y2 = Y2[,1]
Y = X[,1]
zero.mat = matrix(0, nrow = length(Y), ncol = 4)
Y.dat = cbind(Y, zero.mat)
colnames(Y.dat) = c("Y","w","Y*w","Y*(1-w)", "pw=1")
#Y[,1] =Y | Y[,2] = w | Y[,3] = Yw | Y[,4] = Y(1-w) | Y[,5] = p(w=1)#

## set prior parameters ##
p.a = 5; p.b = 2
m1 = m2 = 0
s21 = s22 = 1
a1 = a2 = 1
b1 = b2 = 1
n = nrow(Y.dat)
B = 2000
N = 10000 + B

## set seed priors ##
mu1.temp = 0 #mean(Y1)
sig1.temp = 10 #var(Y1)
mu2.temp = 0 #mean(Y2)
sig2.temp = 10 #var(Y2)


## sample priors ##
# sample p #
wlist = list()
wvec = Y.dat[,2]
wvec.temp = rep(0, n) #new w's
pw1 = rep(0, n) #p(w=1)
pyw1 = rep(0, n) #p(y|w=1)
pyw0 = rep(0, n) #p(y|w=0)

p.temp = rbeta(1, p.a + sum(wvec), n + p.b - sum(wvec))


for(j in 1:n){
  pyw1[j] = dlnorm(Y.dat[j,1], meanlog = mu1.temp, sdlog = sqrt(sig1.temp))
  pyw0[j] = dlnorm(Y.dat[j,1], meanlog = mu2.temp, sdlog = sqrt(sig2.temp))
  pw1[j] = pyw1[j]*p.temp/(pyw1[j]*p.temp + pyw0[j]*(1-p.temp))
  wvec.temp[j] = rbinom(1, 1, pw1[j])
```

```
}



Y.dat[,5] = pw1
Y.dat[,2] = wvec.temp
Y.dat[,3] = Y.dat[,1]*Y.dat[,2]
Y.dat[,4] = Y.dat[,1]*(1-Y.dat[,2])

n1 = sum(Y.dat[,2])
n2 = n-n1

mu2.temp = rnorm(1, mean = (s22*sum(log(Y.dat[,4][Y.dat[,4]!=0]))+sig2.temp*m2)/(s22*n2 + sig2.temp),
                 sd = sqrt((sig2.temp*s22)/(s22*n2+sig2.temp)))

del.temp = rtruncnorm(1, a = 0, b = Inf, mean = (s21*sum(log(Y.dat[,3][Y.dat[,3]!=0]))+m1*sig1.temp-
                 sd = sqrt(sig1.temp*s21/(n1*s21+sig1.temp)) )

mu1.temp = mu2.temp + del.temp


DEL = rep(0, N)
MU1 = rep(0, N)
MU2 = rep(0, N)
SIG1 = rep(0, N)
SIG2 = rep(0, N)



## gibbs sampler ##

for(i in 1:N){
  wvec = Y.dat[,2]
  wvec.temp = rep(0, n) #new w's
  pw1 = rep(0, n) #p(w=1)
  pyw1 = rep(0, n) #p(y|w=1)
  pyw0 = rep(0, n) #p(y|w=0)


  p.temp = rbeta(1, p.a + sum(wvec), n + p.b - sum(wvec))


  for(j in 1:n){
    pyw1[j] = dlnorm(Y.dat[j,1], meanlog = mu1.temp, sdlog = sqrt(sig1.temp))
    pyw0[j] = dlnorm(Y.dat[j,1], meanlog = mu2.temp, sdlog = sqrt(sig2.temp))
    pw1[j] = pyw1[j]*p.temp/(pyw1[j]*p.temp + pyw0[j]*(1-p.temp))
    wvec.temp[j] = rbinom(1, 1, pw1[j])
  }

  wlist[[i]] = wvec.temp

  Y.dat[,5] = pw1
  Y.dat[,2] = wvec.temp
  Y.dat[,3] = Y.dat[,1]*Y.dat[,2]
```

```
  Y.dat[,4] = Y.dat[,1]*(1-Y.dat[,2])

  n1 = sum(Y.dat[,2]==1)
  n2 = n-n1

  mu2.temp = rnorm(1, mean = (s22*sum(log(Y.dat[,4][Y.dat[,4]!=0]))+sig2.temp*m2)/(s22*n2 + sig2.temp
                    sd = sqrt((sig2.temp*s22)/(s22*n2+sig2.temp)))

  sig2.temp = rinvgamma(1, shape = a2+n2/2, scale = b2 + 0.5*sum((log(Y.dat[,4][Y.dat[,4]!=0])-mu2.te

  sig1.temp = rinvgamma(1, shape = a1+n1/2, scale = b1 + 0.5*sum((log(Y.dat[,3][Y.dat[,3]!=0])-mu1.te




  del.temp = rtruncnorm(1, a = 0, b = Inf, mean = (s21*sum(log(Y.dat[,3][Y.dat[,3]!=0]))+m1*sig1.temp
                    sd = sqrt(sig1.temp*s21/(n1*s21+sig1.temp)) )


  mu1.temp = mu2.temp + del.temp

  DEL[i] = del.temp
  MU1[i] = mu1.temp
  MU2[i] = mu2.temp
  SIG1[i] = sig1.temp
  SIG2[i] = sig2.temp

  if(i %% 1200 ==0){
    print(i/N)
  }
}


MU1 = MU1[(B+1):N]
MU2 = MU2[(B+1):N]
SIG1 = SIG1[(B+1):N]
SIG2 = SIG2[(B+1):N]
N = N-B

# traceplots
traceplot(mcmc(MU1))
traceplot(mcmc(MU2))
traceplot(mcmc(SIG1))
traceplot(mcmc(SIG2))

# running avg
MU1.avg = cumsum(MU1)/seq(1,N)
MU2.avg = cumsum(MU2)/seq(1,N)
SIG1.avg = cumsum(SIG1)/seq(1,N)
SIG2.avg = cumsum(SIG2)/seq(1,N)

# No. 2: MCMC plot
png('mu1.png')
plot(MU1, pch = '.', xlab = "N", ylab = expression(mu[1]),
```

```
     main = "MCMC and Running Average")
lines(MU1.avg, lwd = 2)
dev.off()

png('mu2.png')
plot(MU2, pch = '.', xlab = "N", ylab = expression(mu[2]),
     main = "MCMC and Running Average")
lines(MU2.avg, lwd = 2)
dev.off()

png('sig1.png')
plot(SIG1, pch = '.', xlab = "N", ylab = expression(sigma[1]^{2}),
     main = "MCMC and Running Average")
lines(SIG1.avg, lwd = 2)
dev.off()

png('sig2.png')
plot(SIG2, pch = '.', xlab = "N", ylab = expression(sigma[2]^{2}),
     main = "MCMC and Running Average")
lines(SIG2.avg, lwd = 2)
dev.off()

# No. 3
png('post.png')
plot(MU1, MU2, pch = 20, xlab = expression(mu[1]), ylab = expression(mu[2]),
     main = "Posterior mu_1 and mu_2")
dev.off()

# No. 4
wlist.avg = rep(0, N)
wlist.sum = rep(0, N)
for(i in 1:N){
  wlist.avg[i] = mean(wlist[[i]])
  wlist.sum[i] = sum(wlist[[i]])
}

mean(wlist.sum)
quantile(wlist.sum, c(0.025, 0.975))

mean(wlist.avg > 5/7)

# summary statistics
mean(MU1)
quantile(MU1, c(0.025, 0.975))

mean(MU2)
quantile(MU2, c(0.025, 0.975))

mean(SIG1)
quantile(SIG1, c(0.025, 0.975))

mean(SIG2)
quantile(SIG2, c(0.025, 0.975))
```