STA 601: Lab 9

Akbota Anuarbek

Weekdays

$f_X(x;\mu_1,\sigma_1) = \frac{1}{x\sigma_1\sqrt{2\pi}} e^{-\frac{(\ln x-\mu_1)^2}{2\sigma_1^2}}, x > 0$

Weekends

$f_Y(y;\mu_2,\sigma_2) = \frac{1}{y\sigma_2\sqrt{2\pi}} e^{-\frac{(\ln y-\mu_2)^2}{2\sigma_2^2}}, y > 0$

1. Likelihoods:

$L(x|\mu_1,\sigma_1^2) = \prod_{i=1}^{n_1} \frac{1}{x_i\sigma_1\sqrt{2\pi}} e^{-\frac{(\ln x_i-\mu_1)^2}{2\sigma_1^2}}$

$L(y|\mu_2,\sigma_2^2) = \prod_{i=1}^{n_2} \frac{1}{y_i\sigma_2\sqrt{2\pi}} e^{-\frac{(\ln y_i-\mu_2)^2}{2\sigma_2^2}}$

Priors:

$\pi(\mu_1,\mu_2) \propto N(\mu_{1.0},\tau_1)N(\mu_{2.0},\tau_2)\mathbf{1}(\mu_1 > \mu_2)$

$\sigma_1^{-2} \sim Gamma(a,b)$
$\sigma_2^{-2} \sim Gamma(c,d)$

Posteriors:

$\pi(\sigma_1^{-2}|\mu_1,x) \propto Gamma\big(a+\frac{n_1}{2}, b+0.5\sum_{i=1}^{n_1}(\ln x_i-\mu_1)^2\big)$

$\pi(\sigma_2^{-2}|\mu_2,y) \propto Gamma\big(c+\frac{n_2}{2}, d+0.5\sum_{i=1}^{n_2}(\ln y_i-\mu_2)^2\big)$

$\pi(\mu_1,\mu_2|\sigma_1^2,\sigma_2^2,x,y) \propto N(M_1,V_1)N(M_1,V_1)\mathbf{1}(\mu_1 > \mu_2)$, where

$M_1 = V_1\big(\frac{\sum_{i=1}^{n_1}\ln x_i}{\sigma_1^2} + \frac{\mu_{1.0}}{\tau_1}\big)$
$V_1 = \big(\frac{n_1}{\sigma_1^2} + \frac{1}{\tau_1}\big)^{-1}$
$M_2 = V_2\big(\frac{\sum_{i=1}^{n_2}\ln y_i}{\sigma_2^2} + \frac{\mu_{2.0}}{\tau_2}\big)$
$V_2 = \big(\frac{n_2}{\sigma_2^2} + \frac{1}{\tau_2}\big)^{-1}$

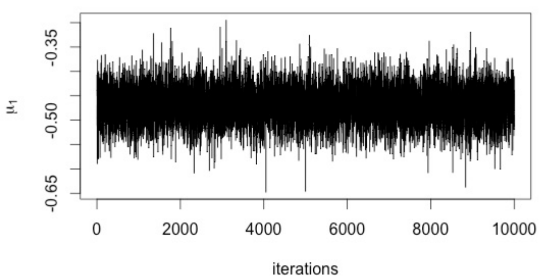
(Derivations are very straightforward and almost the same as in lab 8, thus are omitted)

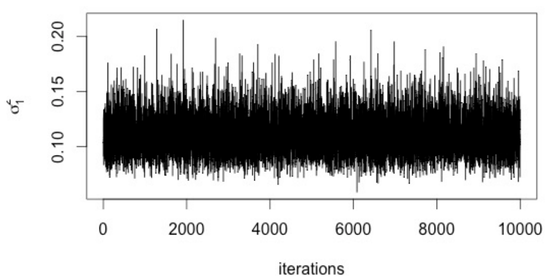
2. Use Gibbs sampling in R to obtain 10,000 post burn-in draws.

```
setwd('/Users/akbota/Documents/STA601/Labs/lab9')
data <- read.table("data.txt", stringsAsFactors=F)
data=data[-1,]

#weekdays
x=as.numeric(data[data[,2]!="Saturday" & data[,2]!="Sunday",1])

#weekends
y=as.numeric(data[data[,2]=="Saturday" | data[,2]=="Sunday",1])

#initialize variables
a=1
b=1
c=1
d=1
mu1.0=0
mu2.0=0
tau1=100
tau2=100
N=12000
n1=length(x)
n2=length(y)
sig_sq1<-c(3)
sig_sq2<-c(3)
mu1<-c(2)
mu2<-c(1)


#Gibbs
for(i in 2:N){

  V1=(n1/sig_sq1[i-1]+1/tau1)^(-1)
  M1=V1*(sum(log(x))/sig_sq1[i-1]+mu1.0/tau1)

  V2=(n2/sig_sq2[i-1]+1/tau2)^(-1)
  M2=V2*(sum(log(y))/sig_sq2[i-1]+mu2.0/tau2)

  mu1[i]<-rnorm(1,mean=M1, sd=sqrt(V1))
  mu2[i]<-rtruncnorm(1,b=mu1[i],mean=M2,sd=sqrt(V2))
```

```
  sumforf1=0
  for(j in 1:n1){
    sumforf1=(log(x[j])-mu1[i])^2+sumforf1
  }
  f1<-rgamma(1,shape=a+n1/2, rate=b+0.5*sumforf1)
  sig_sq1[i]=1/f1

  sumforf2=0
  for(j in 1:n2){
    sumforf2=(log(y[j])-mu2[i])^2+sumforf2
  }
  f2<-rgamma(1,shape=c+n2/2, rate=c+0.5*sumforf2)
  sig_sq2[i]=1/f2

  if(i%%2000==0){
    print(i)
  }
}

#Burn-in
m1=mu1[2001:N]
m2=mu2[2001:N]
ss1=sig_sq1[2001:N]
ss2=sig_sq2[2001:N]

#traceplots
plot(m1, type='l')
plot(ss1, type='l')
```

Both traceplots illustrate that the sampler has converged:





3.

	μ_1	μ_2	σ_1^2	σ_2^2
Posterior point estimates	-0.469	-1.385	0.110	0.160
95% CIs	(-0.546, -0.392)	(-1.532, -1.237)	(0.080, 0.152)	(0.093, 0.267)

```
#posterior point estimates
mean(m1)
mean(m2)
mean(ss1)
mean(ss2)

#CIs
quantile(m1, probs=c(0.025,0.975))
quantile(m2, probs=c(0.025,0.975))
quantile(ss1, probs=c(0.025,0.975))
quantile(ss2, probs=c(0.025,0.975))
```

4. Posterior probability that $\mu_1 > \mu_2$ is equal to 1.
Posterior probability that $\sigma_1^2 > \sigma_2^2$ is 0.1391.

```
#posterior probabilities
mean(m1>m2)
mean(ss1>ss2)
```

5. Posterior probability that pollution level on randomly chosen future Tuesday is higher than pollution level on randomly chosen future Saturday is 0.9369. *(Calculations and R code are given below)*

```
tues<-c()
sat<-c()
for(k in 1:length(m1)){
  sat[k]<-rlnorm(1,meanlog=exp(m2[k]+ss2[k]/2), sdlog=sqrt((exp(ss2[k])-1)*exp(2*m2[k]+ss2[k])))
  tues[k]<-rlnorm(1,meanlog=exp(m1[k]+ss1[k]/2), sdlog=sqrt((exp(ss1[k])-1)*exp(2*m1[k]+ss1[k])))
}
mean(tues > sat)
```