STA 601: Lab 10

Akbota Anuarbek

(1) Model

n is number of days in the sample.

Hidden variable $z_i = \begin{cases} 1 & \text{if ith observation is from a weekday} \\ 0 & \text{if ith observation is from a weekend} \end{cases}$

$z_i \sim Bernoulli(\theta)$, where θ is the probability that given observation is from a weekday.

$\theta \sim Beta(5, 2)$

$L(z|\theta) = \theta^{\sum_{i=1}^n z_i}(1-\theta)^{n-\sum_{i=1}^n z_i}$

Weekdays

$f(y_i|z_i = 1) = \frac{1}{y_i\sigma_1\sqrt{2\pi}}e^{-\frac{(\ln y_i - \mu_1)^2}{2\sigma_1^2}}$

Weekends

$f(y_i|z_i = 0) = \frac{1}{y_i\sigma_2\sqrt{2\pi}}e^{-\frac{(\ln y_i - \mu_2)^2}{2\sigma_2^2}}$

$L(y|z, \mu_1, \mu_2, \sigma_1, \sigma_2) = \prod_{i=1}^n \left(\frac{1}{y_i\sigma_1\sqrt{2\pi}}e^{-\frac{(\ln y_i - \mu_1)^2}{2\sigma_1^2}}\right)^{z_i}\left(\frac{1}{y_i\sigma_2\sqrt{2\pi}}e^{-\frac{(\ln y_i - \mu_2)^2}{2\sigma_2^2}}\right)^{(1-z_i)}$

$\pi(\mu_1, \mu_2) \propto N(\tilde{\mu}_1, \tau_1)N(\tilde{\mu}_2, \tau_2)\mathbf{1}(\mu_1 > \mu_2)$

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

(2) Derive full conditionals and run Gibbs in R

$\underline{\pi(\mu_1, \mu_2|y, z, \sigma_1^2, \sigma_2^2)} = N(\mu_1; M_1, V_1)N(\mu_2; M_2, V_2)\mathbf{1}(\mu_1 > \mu_2)$, where

$V_1 = (\tau_1^{-1} + \sigma_1^{-2}\sum_{i=1}^n z_i)^{-1}$

$M_1 = V_1(\tilde{\mu}_1\tau_1^{-1} + \sigma_1^{-2}\sum_{i=1}^n z_i \ln y_i)$

$V_2 = (\tau_2^{-1} + \sigma_2^{-2} \sum_{i=1}^n (1 - z_i))^{-1}$

$M_2 = V_2(\tilde{\mu}_2 \tau_2^{-1} + \sigma_2^{-2} \sum_{i=1}^n (1 - z_i) \ln y_i)$

$\underline{\pi(\sigma_1^{-2}|\mu_1, y, z)} = Gamma\left(a + \frac{1}{2}\sum_{i=1}^n z_i, b + \frac{1}{2}\sum_{i=1}^n (\ln y_i - \mu_1)^2 z_i\right)$

$\underline{\pi(\sigma_2^{-2}|\mu_2, y, z)} = Gamma\left(c + \frac{1}{2}\sum_{i=1}^n (1 - z_i), d + \frac{1}{2}\sum_{i=1}^n (\ln y_i - \mu_2)^2 (1 - z_i)\right)$

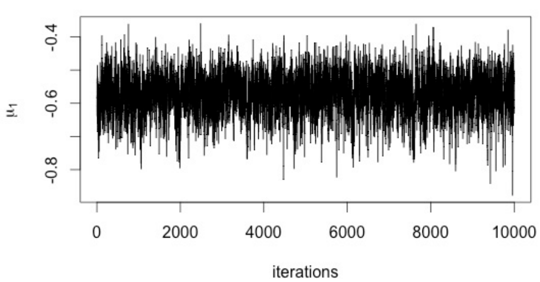
$\underline{\pi(\theta|z)} = Beta(5 + \sum_{i=1}^n z_i, 2 + n - \sum_{i=1}^n z_i)$

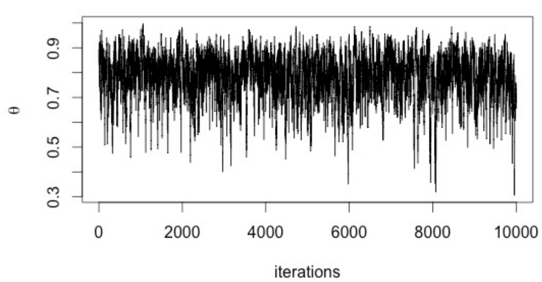
$\underline{\pi(z_i|y, \theta, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)} = Bernoulli(\tilde{\theta})$, where

$$\tilde{\theta} = \frac{\frac{\theta}{y_i \sigma_1 \sqrt{2\pi}} e^{-\frac{(\ln y_i - \mu_1)^2}{2\sigma_1^2}}}{\frac{\theta}{y_i \sigma_1 \sqrt{2\pi}} e^{-\frac{(\ln y_i - \mu_1)^2}{2\sigma_1^2}} + \frac{1-\theta}{y_i \sigma_2 \sqrt{2\pi}} e^{-\frac{(\ln y_i - \mu_2)^2}{2\sigma_2^2}}}$$

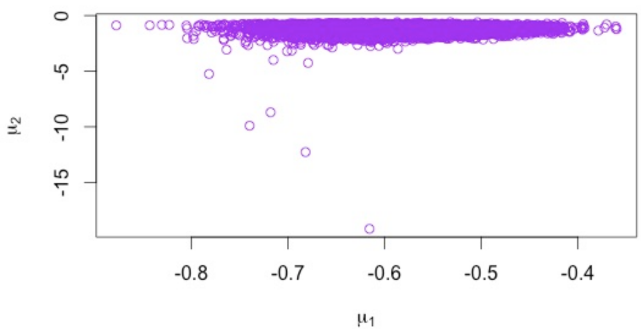
Set $a = b = c = d = 1$, $\tilde{\mu}_1 = \tilde{\mu}_2 = 0$, $\tau_1 = \tau_2 = 100$

Run Gibbs in R, R code is provided in the end of the pdf. Burn-in first 2000 observations. Look at trace plots, to see that sample has converged.





(3)



(4)

θ is the probability that given day is a weekday, thus:

Point estimate: $\frac{\sum_{t=1}^{N} \theta^{(t)} * n}{N} = 78.34 \approx 78$

95% CI: (56.82, 93.35)

Prob-ty that technicians is coming less often on weekend: $P(\theta \geq 5/7) \approx \frac{\sum_{t=1}^{N} \mathbf{1}(\theta^{(t)} \geq 5/7)}{N} = 0.7976$

* *Calculations are given in the end as R code*

(5)

	μ_1	μ_2	σ_1^2	σ_2^2
Posterior point estimates	**-0.572**	**-1.311**	**0.147**	**0.457**
95% CIs	**(-0.703 -0.460)**	**(-1.892 -0.817)**	**(0.090 0.239)**	**(0.185 1.049)**
Last lab's post.p.est's	-0.469	-1.385	0.110	0.160
Last lab's 95% CIs	(-0.546, -0.392)	(-1.532, -1.237)	(0.080, 0.152)	(0.093, 0.267)

The posterior inferences are quite similar in both lab's. Posterior point estimates for σ_2^2 is the only one for which posterior point estimates are significantly different. Moreover, confidence intervals are mostly different in two labs, even though some point estimates are close to each other.

```
library(truncnorm)

setwd('/Users/akbota/Documents/STA601/Labs/lab10')
```

```
data<- read.table("data.txt", stringsAsFactors=F)

y<-as.numeric(data[2:101,1])

N<-12000
a=1
b=1
c=1
d=1
m1=0
m2=0
tau1=100
tau2=100
n=length(y)
ss1<-c(3)
ss2<-c(3)
mu1<-c(2)
mu2<-c(1)
th<-c(0.5)
z<-c()

#Gibbs

for(t in 2:N){

  for(i in 1:n){
   term1=th[t-1]*(y[i]*sqrt(ss1[t-1]*2*pi))^(-1)*exp(-((log(y[i])-mu1[t-1])^2)/(2*ss1[t-1]))
   term2=(1-th[t-1])*(y[i]*sqrt(ss2[t-1]*2*pi))^(-1)*exp(-((log(y[i])-mu2[t-1])^2)/(2*ss2[t-1]))

    z[i]<-rbinom(1, 1, prob=term1/(term1+term2))
  }

  V1=(1/tau1+sum(z)/ss1[t-1])^(-1)
  M1=V1*(m1/tau1+sum(z*log(y))/ss1[t-1])
  V2=(1/tau2+sum(1-z)/ss2[t-1])^(-1)
  M2=V2*(m2/tau2+sum((1-z)*log(y))/ss2[t-1])

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

  f<-rgamma(1, shape=a+0.5*sum(z), rate=b+0.5*sum(z*(log(y)-mu1[t])^2))
  ss1[t]=1/f

  f<-rgamma(1, shape=c+0.5*sum(1-z), rate=d+0.5*sum((1-z)*(log(y)-mu2[t])^2))
```

```
  ss2[t]=1/f

  th[t]=rbeta(1, shape1 = 5+sum(z), shape2 = 2+n-sum(z))

  if(t%%100==0){
    print(t)
  }
}

plot(mu1[2001:N], type='l', xlab='iterations', ylab=expression(mu[1]))
plot(th[2001:N], type='l', xlab='iterations', ylab=expression(theta))

#part3
plot(mu1[2001:N],mu2[2001:N],xlab=expression(mu[1]),ylab=expression(mu[2]),col="purple")

#part4
mean(th[2001:N]*100)
quantile(th[2001:N]*100, c(0.025, 0.975))
mean(th[2001:N]>=5/7)

#part5
mean(mu1[2001:N])
quantile(mu1[2001:N], c(0.025, 0.975))
mean(mu2[2001:N])
quantile(mu2[2001:N], c(0.025, 0.975))
mean(ss1[2001:N])
quantile(ss1[2001:N], c(0.025, 0.975))
mean(ss2[2001:N])
quantile(ss2[2001:N], c(0.025, 0.975))
```