Introduction to Bayesian Data Analysis

Problem Set 6 - Solutions

(1) (a) $\theta_B = \theta \times \gamma$ and $p(\theta_B|\theta,\gamma) = p(\theta) \times p(\gamma)$ because θ and γ are independent, and $p(\theta_B|\theta,\gamma) \neq p(\theta_B|\gamma)$. Now $\theta_A = \theta$, and so $p(\theta_B|\theta,\gamma) = p(\theta_B|\theta_A,\gamma) = p(\theta_A) \times p(\gamma) \neq p(\theta_B|\gamma)$, hence θ_B and θ_A are not independent under this prior distribution.

This prior is justified if we believe there exists an underlying rate θ for the population and it is modified by γ for those men that do not have a Bachelor's degree.

(b) $p(\mathbf{y}_A|\theta) = \prod_{i=1}^{n_A} \frac{e^{-\theta}\theta^{y_{A,i}}}{y_{A,i}!}$; $p(\mathbf{y}_B|\theta,\gamma) = \prod_{i=1}^{n_B} \frac{e^{-\theta\gamma}(\theta\gamma)^{y_{B,i}}}{y_{B,i}!}$

$p(\theta) = \frac{b_\theta^{a_\theta}}{\Gamma(a_\theta)}\theta^{a_\theta-1}e^{-b_\theta\theta}$; $p(\gamma) = \frac{b_\gamma^{a_\gamma}}{\Gamma(a_\gamma)}\gamma^{a_\gamma-1}e^{-b_\gamma\gamma}$

Note that $p(\theta,\gamma|\mathbf{y}_A,\mathbf{y}_B) = \frac{p(\theta,\gamma,\mathbf{y}_A,\mathbf{y}_B)}{p(\mathbf{y}_A,\mathbf{y}_B)} \propto p(\mathbf{y}_A|\theta)p(\mathbf{y}_B|\theta,\gamma)p(\theta,\gamma)$;

so

$$\begin{aligned}
p(\theta|\mathbf{y}_A,\mathbf{y}_B,\gamma) &\propto p(\mathbf{y}_A|\theta)p(\mathbf{y}_B|\theta,\gamma)p(\theta) \\
&\propto \left[\prod_{i=1}^{n_A} e^{-\theta}\theta^{y_{A,i}}\right]\left[\prod_{i=1}^{n_B} e^{-\theta\gamma}(\theta\gamma)^{y_{B,i}}\right]\theta^{a_\theta-1}e^{-b_\theta\theta} \\
&\propto e^{-n_A\theta}\theta^{\sum_{i=1}^{n_A} y_{A,i}}e^{-n_B\theta\gamma}(\theta\gamma)^{\sum_{i=1}^{n_B} y_{B,i}}\theta^{a_\theta-1}e^{-b_\theta\theta} \\
&\propto e^{-(n_A+n_B\gamma+b_\theta)\theta}\theta^{n_A\bar{y}_A+n_B\bar{y}_B+a_\theta-1}
\end{aligned}$$

Therefore, $\theta|\mathbf{y}_A,\mathbf{y}_B,\gamma \sim \text{gamma}(n_A\bar{y}_A + n_B\bar{y}_B + a_\theta, n_A + n_B\gamma + b_\theta)$

(c)

$$\begin{aligned}
p(\gamma|\mathbf{y}_A, \mathbf{y}_B, \theta) &\propto p(\mathbf{y}_B|\theta, \gamma)p(\gamma) \\
&\propto \left[\prod_{i=1}^{n_B} e^{-\theta\gamma}(\theta\gamma)^{y_{B,i}}\right] \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\propto e^{-(n_B\theta + b_\gamma)\gamma} \gamma^{n_B \bar{y}_B + a_\gamma - 1}
\end{aligned}$$

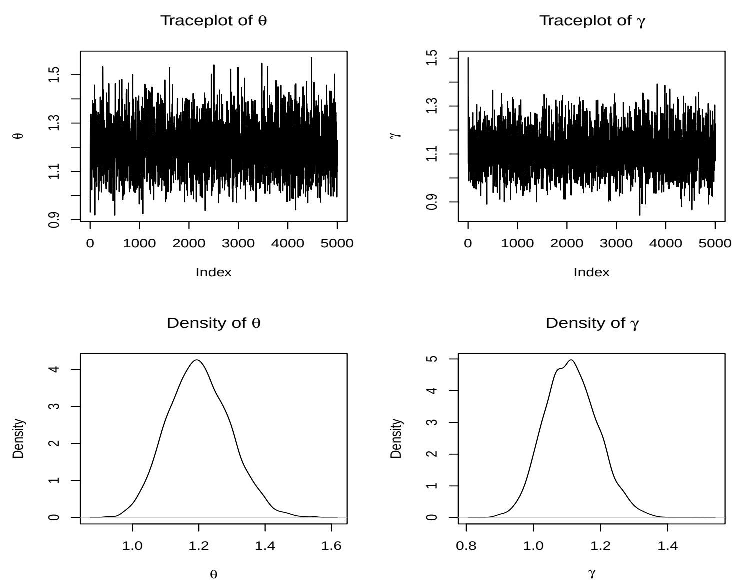
Therefore, $\gamma|\mathbf{y}_A, \mathbf{y}_B, \theta \sim \text{gamma}(n_B\bar{y}_B + a_\gamma, n_B\theta + b_\gamma)$

(d)
```
a_theta<-2
b_theta<-1
a_gamma<-c(8,16,32,64,128)

yA <- read.table("menchild30bach.txt", header = F)
yA<-yA[,1]
nA<-length(yA)
ybarA<-mean(yA)
yB <- read.table("menchild30nobach.txt", header = F)
yB<-yB[,1]
nB<-length(yB)
ybarB<-mean(yB)

#Gibbs sampling
S<-5000
PHI<-matrix(nrow=S,ncol=2)
PHI[1,]<-phi<-c( ybarA ,ybarB/ybarA)

mean.B.M.A<-NULL
for (i in 1:5){
b_gamma<-a_gamma[i]
for(s in 2:S) {

# generate a new theta value from its full conditional
phi[1]<-rgamma(1,nA*ybarA+nB * ybarB+a_theta,nA+nB*phi[2]+b_theta)

# generate a new gamma value from its full conditional
phi[2]<- rgamma(1, nB*ybarB+a_gamma[i], nB*phi[1]+b_gamma)
PHI[s,]<-phi         }
```

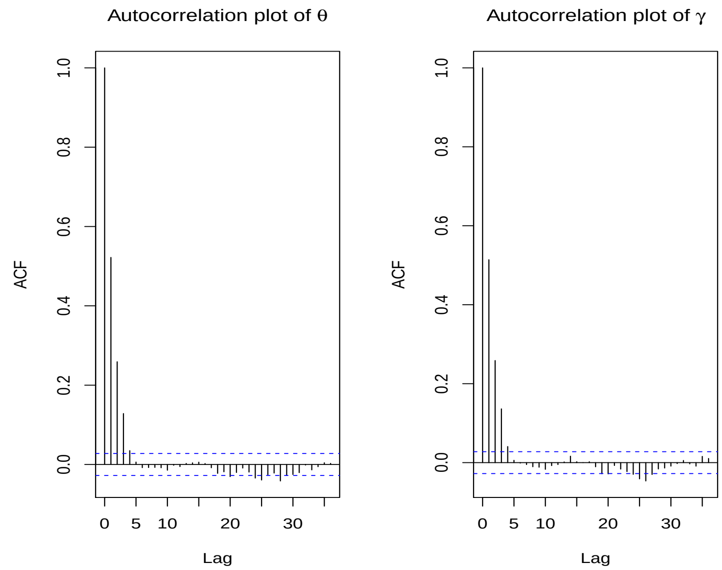
```
mean.B.M.A<-c(mean.B.M.A,mean(PHI[,2]*PHI[,1]-PHI[,1]))
}
> mean.B.M.A
[1] 0.3780609 0.3355366 0.2714495 0.2016046 0.1324830
```

As the hyper-parameters in the prior for gamma increase, the mean difference of $(\theta_B - \theta_A|\mathbf{y}_A, \mathbf{y}_B)$ gets smaller.

```
> effectiveSize(PHI[,1])
   var1
1760.03
> effectiveSize(PHI[,2])
    var1
1741.798
```



The effective sample sizes are large, the traceplots show the sequence of parameter draws randomly moves around the parameter space, and the autocorrelation plots do not show high autocorrelations to be concerned about.

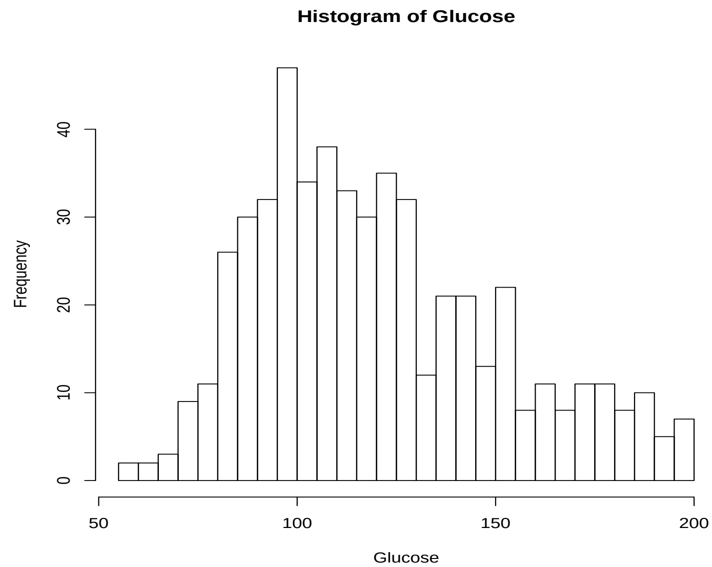


(2) (a) The empirical distribution appears slightly skewed to the right.

(b) $f(X_1, ..., X_n, p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2 | \mathbf{y}) \propto$
$f(\mathbf{y}|X_1, ..., X_n, p, \theta_1, \theta_2, \sigma_1^2, \sigma_2^2) f(X_1, ..., X_n|p) f(p) f(\theta_1) f(\theta_2) f(\sigma_1^2) f(\sigma_2^2)$.
Let $\phi = (\theta_1, \theta_2, \sigma_1^2, \sigma_2^2)$. Let $X = (X_1, ..., X_n)$
The Gibbs sampler cycle is:

1. For $i = 1, ..., n$ update $X_i \sim f(y_i|\phi, y_i) f(X_i|p)$ (conditional independence)
2. Update $p \sim f(p|X)$
3. For $k = 1, 2$, update $\theta_k \sim f(\theta_k|X, p, \theta_{-k}, \sigma_1^2, \sigma_2^2)$
4. For $k = 1, 2$, update $\sigma_k^2 \sim f(\sigma_k^2|X, p, \theta_1, \theta_2, \sigma_{-k}^2)$

The joint posterior distribution is



$$
\begin{aligned}
f(X,\phi|\mathbf{y}) &\propto \left(\prod_{i=1}^{n} f(y_i|\phi,x_i)f(x_i|p)\right) f(p)f(\phi) \\
&\propto \left(\prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2_{x_i}}} exp\left(-\frac{1}{2\sigma^2_{x_i}}(y_i-\theta_{x_i})^2\right) p^{2-x_i}(1-p)^{x_i-1}\right) \\
&\quad p^{\alpha-1}(1-p)^{\beta-1} exp(-\frac{1}{2\tau_0^2}(\theta_1-\mu_0)^2) exp(-\frac{1}{2\tau_0^2}(\theta_2-\mu_0)^2) \\
&\quad (\sigma_1^2)^{-(\nu_0/2+1)} exp(-\frac{\nu_0\sigma_0^2/2}{\sigma_1^2})(\sigma_2^2)^{-(\nu_0/2+1)} exp(-\frac{\nu_0\sigma_0^2/2}{\sigma_2^2})
\end{aligned}
$$

The conditional distribution on the class labels:

$$
\begin{aligned}
f(x_i|y_i,p,\phi) &= \frac{f(y_i|x_i,\theta,p)}{\sum_{k=1}^{2} f(y_i|x_i=k,\theta,p)} \\
&\propto \frac{\frac{1}{\sqrt{\sigma^2_{x_i}}} exp\left(-\frac{1}{2\sigma^2_{x_i}}(y_i-\theta_{x_i})^2\right) p^{2-x_i}(1-p)^{x_i-1}}{\sum_{k=1}^{2} \frac{1}{\sqrt{\sigma^2_k}} exp\left(-\frac{1}{2\sigma^2_k}(y_i-\theta_k)^2\right) p^{2-k}(1-p)^{k-1}}
\end{aligned}
$$

The conditional distribution on p:

$p| \sim Beta(\alpha + 2n - \sum_i x_i, \beta + \sum_i x_i - n)$ The conditional distribution on the group means $(\theta_k, (k = 1, 2))$:

Let n_k be the current size of class k; $n_k = \#\{i : x_i = k\}$

and so $\theta_k | X, \mathbf{y}, p, \theta_{-k}, \sigma_1^2, \sigma_2^2) \sim N(\mu_k, \sigma_k^2)$ where $\mu_k = \frac{\mu_0/\tau_0^2 + n_k \bar{y}_k/\sigma_k^2}{1/\tau_0^2 + n_k/\sigma_k^2}$

where $\bar{y}_k = \frac{1}{n_k} \sum_{i:x_i=k} y_i$

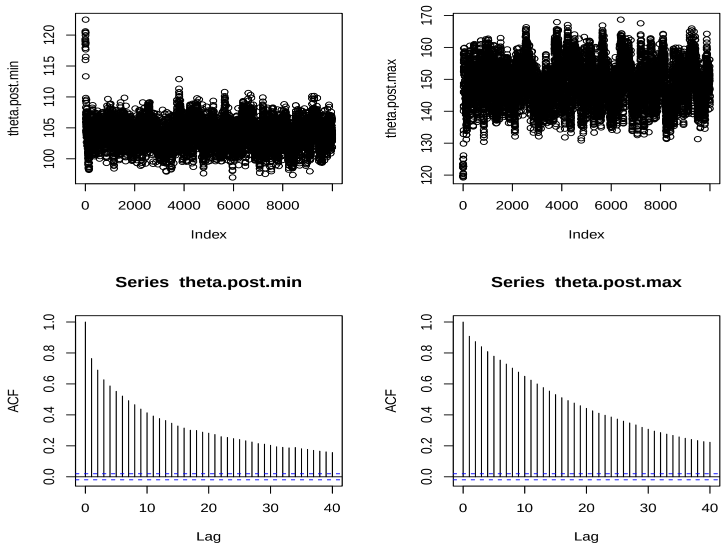
and $\sigma_k^2 | X, \mathbf{y}, p, \sigma_{-k}^2, \theta_1, \theta_2) \sim \mathrm{Inv-gamma}(\nu_k/2, \nu_k \sigma_k^2(\theta_k)/2)$

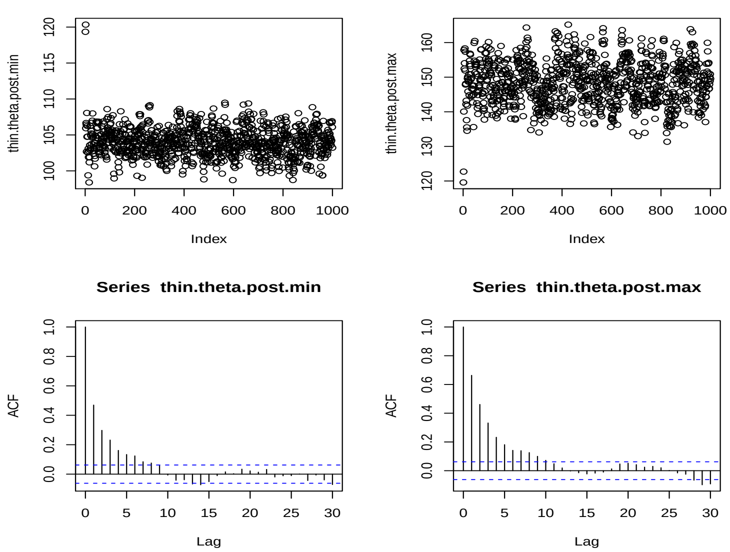
where $\nu_k = \nu_0 + n_k$, $\sigma^2(\theta_k) = \frac{1}{\nu_k}[\nu_o \sigma_0^2 + n_k s_k^2(\theta_k)]$ and $s_k^2(\theta_k) = \sum_{i:x_i=k} (y_i - \theta_k)^2 / n_k$

(c)
```
alpha<-1
beta<-1

mu0<-120
tau20<-200
sigma20<-1000
nu0<-10
mcmc.mix<-function(y, niter=10000, theta1=mean(y),theta2=mean(y),
                   sigma12=var(y),sigma22=var(y)){
           n=length(y); K=2; theta=c(theta1,theta2)
                 sigma2=c(sigma12,sigma22); p=0.5; x=NULL; nk=NULL
          p.mcmc=NULL; theta.mcmc=NULL; sigma2.mcmc=NULL; x.mcmc=NULL
for (nit in 1:niter) {
for (i in 1:n) {
p.z<-exp(-(y[i]-theta)^2/2/sigma2)/sqrt(sigma2)*c(p,1-p)
x[i]<-sample(1:2,1,prob=p.z/sum(p.z))}
#print(p.z/sum(p.z))
x.mcmc=rbind(x.mcmc,x)
for (k in 1:K){
nk[k]=sum(x==k)
muk<-(mu0/tau20+nk[k]*mean(y[x==k])/sigma2[k])/(1/tau20+nk[k]/sigma2[k])
t2n<-1/(1/tau20+nk[k]/sigma2[k])
theta[k]<-rnorm(1,muk,sqrt(t2n))
         nuk<-nu0+sum(x==k)
         sk2<-(nu0*sigma20+(nk[k]-1)*var(y[x==k])+nk[k]*
                      (mean(y[x==k])-theta[k])^2)/nuk
         sigma2[k]<-rinvgamma(1,nuk/2,nuk*sk2/2)
```

```
}

theta.mcmc<-rbind(theta.mcmc,theta)
sigma2.mcmc<-rbind(sigma2.mcmc,sigma2)
p<-rbeta(1,alpha+2*n-sum(x),beta+sum(x)-n)
p.mcmc<-rbind(p.mcmc,p)

}
return(list(x=x.mcmc, theta=theta.mcmc,sigma2=sigma2.mcmc,p=p.mcmc))
}

obj<-mcmc.mix(y, niter=10000, theta1=mean(y),theta2=mean(y),
                          sigma12=var(y),sigma22=var(y))
> theta.post<-obj$theta
> thin.theta.post.min<-theta.post.min[seq(1,length(theta.post.min),10)]
> thin.theta.post.max<-theta.post.max[seq(1,length(theta.post.max),10)]
> effectiveSize(theta.post.min[9001:10000])
    var1
144.9595
> effectiveSize(theta.post.max[9001:10000])
    var1
45.40116
> effectiveSize(thin.theta.post.min)
    var1
421.3194
> effectiveSize(thin.theta.post.max)
    var1
291.4248
```

Traceplots of $\theta_{(1)}^{(s)}$ and $\theta_{(2)}^{(s)}$ indicated immediate convergence. However, the autocorrelation plots show a high degree of autocorrelation. If we thin each sequence by taking every tenth draw, the second set of plots show a lower degree of autocorrelation and increased effective sample sizes.


theta.post.min
Index
theta.post.max
Index
Series theta.post.min
ACF
Lag
Series theta.post.max
ACF
Lag


thin.theta.post.min
Index
thin.theta.post.max
Index
Series thin.theta.post.min
ACF
Lag
Series thin.theta.post.max
ACF
Lag

(d) The histogram of posterior draws of $\tilde{Y}_s$ is quite similiar to the empirical data density except that the frequency of higher blood glucose levels is underestimated. Posterior predictive checks on tail-area percentiles show that the empirical tail area probabilities are adequately captured by the two-component mixture model. Hence, we conclude that the two-component mixture model is reasonable for inferential purposes.

```
p.post<-obj$p
sigma.post<-obj$sigma2
x.post<-NULL
for (i in 1:10000){
x.post<-c(x.post,sample(c(1,2),1,prob=c(p.post[i],1-p.post[i])))
}

y.post<-NULL
for (i in 1:10000){
y.post<-c(y.post,rnorm(1,theta.post[i,x.post[i]],
                sqrt(sigma.post[i,x.post[i]])))
}

#Posterior predictive checks
> result<-NULL
>
> for (i in 1:10000){
+  x<-sample(c(1,2),length(y),prob=c(p.post[i],1-p.post[i]),replace=TRUE)
+  y.s<-rnorm(length(y),theta.post[i,x],sqrt(sigma.post[i,x]))
+  result<-cbind(result,c(min(y.s),quantile(y.s,probs=c(0.75,0.9,0.95,0.975)
       max(y.s)))
+ }
>
```

```
> mean(result[1,]>=min(y))
```

```
[1] 0.2049
> mean(result[2,]>=quantile(y,0.75))
[1] 0.3813
> mean(result[3,]>=quantile(y,0.90))
[1] 0.1458
> mean(result[4,]>=quantile(y,0.95))
[1] 0.318
> mean(result[5,]>=quantile(y,0.975))
[1] 0.5239
> 1-mean(result[6,]>=max(y))
[1] 0.0043
> mean(result[7,]>=mean(y))
[1] 0.4651
```