# Introduction to Bayesian Data Analysis

## Tutorial 10 - Solutions

(1)

$$
\begin{aligned}
Pr(\phi^{(s+1)} \in A) &= \int_x Pr(\phi^{(s+1)} \in A, \phi^{(s)} = x)\ d\phi^{(s)} \\
&= \int_x p(\phi^{(s)}) Pr(\phi^{(s+1)} \in A \big| \phi^{(s)} = x)\ d\phi^{(s)} \\
&= \int_x p(\phi^{(s)}) \int_A p(\phi^{(s+1)} | \phi^{(s)} = x) d\phi^{(s+1)})\ d\phi^{(s)} \\
&= \int_A p(\phi^{(s)}) \left[ \int_x p(\phi^{(s+1)} | \phi^{(s)} = x) d\phi^{(s+1)}) \right] d\phi^{(s)} \\
&= \int_A p(\phi^{(s)} d\phi^{(s)} \text{ (as required).}
\end{aligned}
$$

(2) (a) The R-Code to implement the MH algorithm is as follows:

```
library(MCMCpack)
library(mvtnorm)
diab.data<-read.table("azdiabetes.dat",header=TRUE)
attach(diab.data)
names(diab.data)

n<-nrow(diab.data)
y<-rep(0,n)
y[diabetes=="Yes"]<-1

X<-cbind(rep(1,n),npreg,bp,bmi,ped,age)
X<-t( (t(X)-apply(X,2,mean))/apply(X,2,sd))

mle.model<-glm(y~-1+X,family="binomial")
summary(mle.model)
```

```
p<-6

#proposal variance for beta
beta.var.prop<-summary(mle.model)$cov.unscaled
#prior parameters
pmn.beta<-rep(0,6)
psd.beta<-c(4,rep(2,5))
#starting values for beta and gamma
beta<-coef(mle.model)
acs_beta<-0
gamma<-c(1,rbinom(5,1,0.5))
#MH algorithm parameters and results matrices
S<-10000
BETA<-NULL
GAMMA<-NULL
set.seed(1)
#inverse logit function
ilogit<-function(x) exp(x)/(1+exp(x))

for(s in 1:S){ #lpy.c==currrent log-likelihood
   lpy.c<-sum(dbinom(y,1,ilogit(X[,gamma==1,drop=FALSE]%*%beta[gamma==1]),log=T))
    #UPDATE GAMMAs
      for(j in sample(2:p))
    {
      gamma_p<-gamma ; gamma_p[j]<-1-gamma_p[j]
      #lpy.p==proposal loglikelihood
      lpy.p<-sum(dbinom(y,1,ilogit(X[,gamma_p==1,drop=FALSE]%*%
           beta[gamma_p==1,drop=FALSE]),log=T))
      lhr<-(lpy.p-lpy.c)*(-1)^(gamma_p[j]==0)
      gamma[j]<-rbinom(1,1,1/(1+exp(-lhr)))
      if(gamma[j]==gamma_p[j]) {lpy.c<-lpy.p}
      }
      GAMMA<-rbind(GAMMA,gamma)
     #UPDATE BETA
      beta.p<-rmvnorm(1,beta,beta.var.prop)
      lpy_beta.p<-sum(dbinom(y,1,ilogit(X[,gamma==1,drop=FALSE]%*%
             beta.p[gamma==1]),log=T))
      lhr.beta<-lpy_beta.p-lpy.c+sum(dnorm(beta.p,pmn.beta,psd.beta,log=T))-
            sum(dnorm(beta,pmn.beta,psd.beta,log=T))
     if(log(runif(1))<lhr.beta) {beta<-beta.p; acs_beta<-acs_beta+1}
        BETA<-rbind(BETA,beta)   }
```

Traceplots for the sequences $\beta_j^{(s)}$, $\beta_j^{(s)} \times \gamma_j^{(s)}$ and $\gamma_j^{(s)}$ are given below.

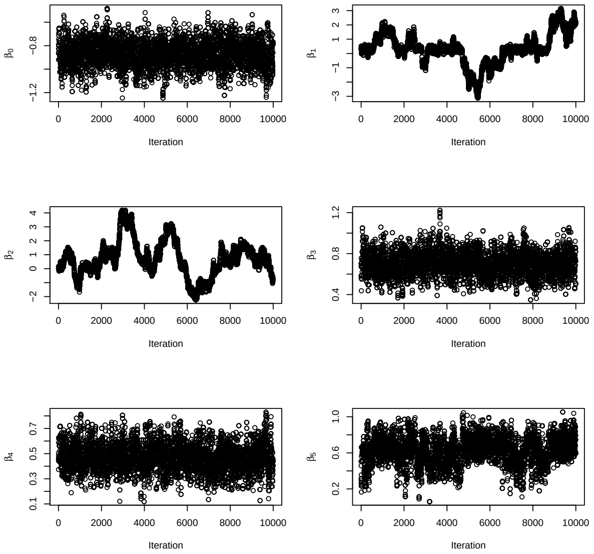

Figure 1: Traceplots $\beta_j^{(s)}$

Figure 1 shows non-stationarity of the MCMC chains for $\beta_1$ and $\beta_2$, which is reflective of the predominant draws of $\gamma_1^{(s)} = 0$ and $\gamma_2^{(s)} = 0$. Occassionally we accept draws of $\gamma_1^{(s)} = 1$ or $\gamma_2^{(s)} = 1$, as displayed in Figures 2 and 3, which results in the irregular peaks and troughs in the MCMC chains for $\beta_1$ and $\beta_2$ in Figure 1. That is, our results indicate that the number of pregnancies and blood pressure are not predictors of diabetes incidence.

The acceptance rate for $\boldsymbol{\beta}$ is 0.32 which is reasonable.

(b)

```
>  a<-apply(GAMMA,2,function(x) mean(x==1))
>  names(a)<-c("Intercept","npreg","bp","bmi","ped","age")
>  a
Intercept      npreg         bp        bmi        ped        age
   1.0000     0.4035     0.0658     1.0000     0.9999     0.9995
```

The values for $Pr(\gamma_j = 1|\mathbf{x}, \mathbf{y})$ are displayed above, and indicate that bmi, diabetes pedigree and age have strong associations with the incidence of diabetes. The number of pregnancies and blood pressure have weak associations with the incidence of diabetes, which from a medical viewpoint is surprising.

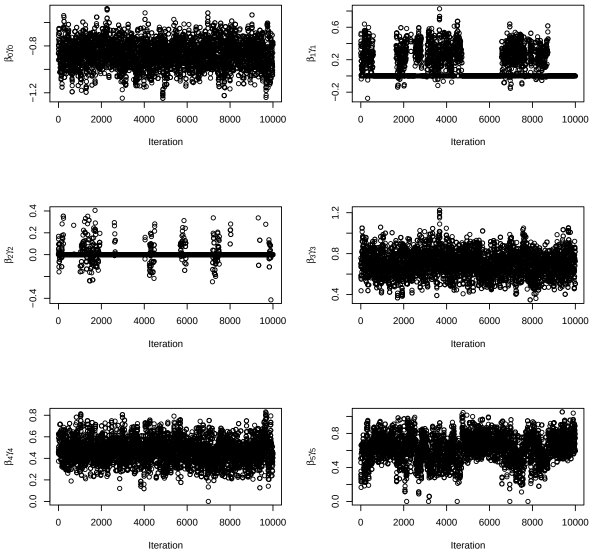

Figure 2: Traceplots $\beta_j^{(s)} \times \gamma_j^{(s)}$

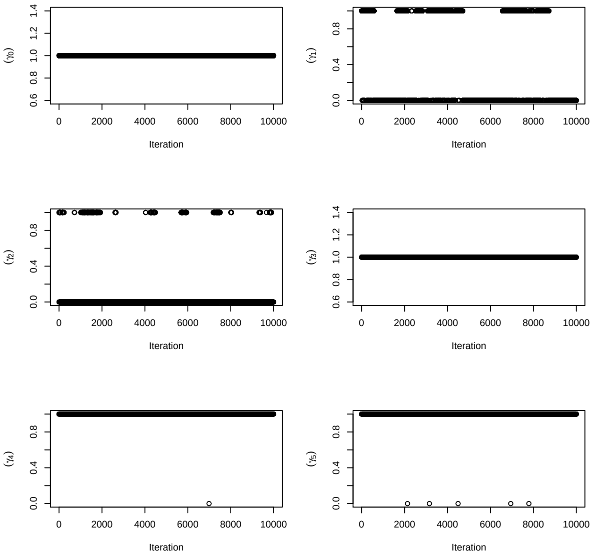

Figure 3: Traceplots $\gamma_j^{(s)}$

However, for the important predictors, the autocorrelation plots show high autocorrelations to be concerned about, so the current estimates of $Pr(\gamma_j = 1|\mathbf{x}, \mathbf{y})$ are not reliable. The sequences should be thinned and/or the chains run for a longer number of iterations and estimates for $Pr(\gamma_j = 1|\mathbf{x}, \mathbf{y})$ recalculated.

Also note that the estimates in parts (b) and (c) are bayesian model averaged estimates, because we average over all iterations where the variables are not always active in the logistic model, depending on whether $\gamma_j^{(s)} = 1$ or 0.

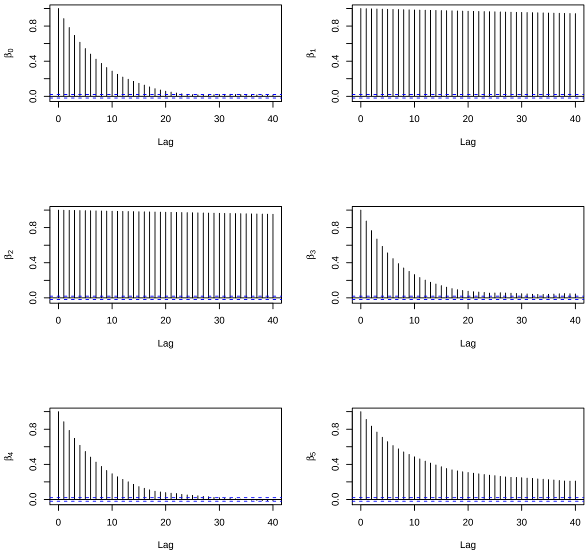

Figure 4: Autocorrelation plots for $\beta_j^{(s)}$

(c) From Figure 6, we see that the posterior density plot of $\beta_1\gamma_1$ is somewhat bimodal (with one mode around 0) which is reflective of the posterior probability $Pr(\gamma_1 = 1|\mathbf{x}, \mathbf{y}) = 0.4035$ (which is not $\approx 0$ or $\approx 1$ unlike for the other variables).

```
>  b<-apply(GAMMA*BETA,2,mean)
>  names(b)<-c("Intercept","npreg","bp","bmi","ped","age")
>  b
Intercept     npreg        bp       bmi       ped       age
  -0.8647    0.1126    0.0018    0.7004    0.4673    0.6193
```

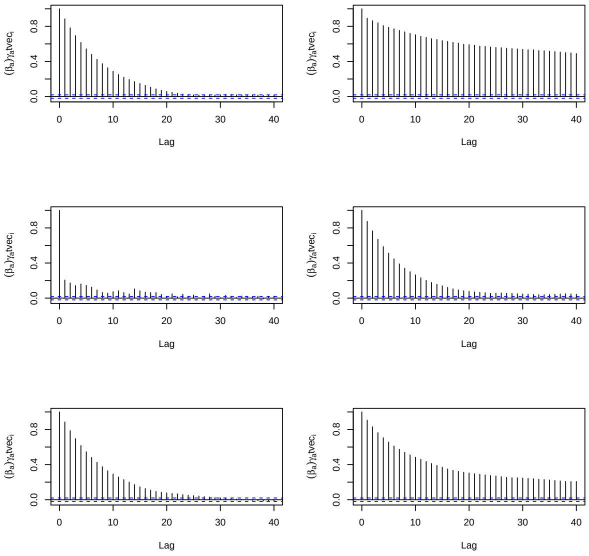

Figure 5: Autocorrelation plots for $\beta_j^{(s)} \times \gamma_j^{(s)}$

Posterior mean estimates for $\beta_j \gamma_j$ are given in Table 1.

| npreg | bp | bmi | ped | age |
|---|---|---|---|---|
| 0.1126 | 0.0018 | 0.7004 | 0.4673 | 0.6193 |

Table 1: Posterior mean estimates for $\beta_j \gamma_j$

The association of each variable with the indicidence of diabetes is better communicated on the odds scale. Posterior mean estimates for $\exp(\beta_j \gamma_j)$ are given in Table 2.

```
>  b<-apply(GAMMA*BETA,2, function(x) mean(exp(x)))
>  names(b)<-c("Intercept","npreg","bp","bmi","ped","age")
>  b
Intercept      npreg         bp        bmi        ped        age
     0.42       1.13       1.00       2.03       1.60       1.88
```

All variables (apart from blood pressure) are associated with an increase in the odds of diabetes. The estimated increase in odds of diabetes ranges from 13% (per unit increase in the number of pregnancies) to 203% (per unit increase in bmi).

| npreg | bp | bmi | ped | age |
|---|---|---|---|---|
| 1.13 | 1.00 | 2.03 | 1.60 | 1.88 |

Table 2: Posterior mean estimates for $\exp(\beta_j \gamma_j)$

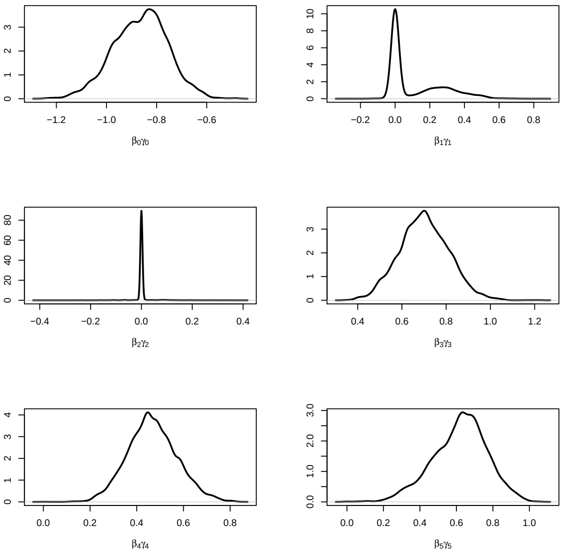

Figure 6: Posterior density plots for $\beta_j \gamma_j$