INTRODUCTION TO BAYESIAN DATA ANALYSIS (STAT3016/4116/7016)
SEMESTER 2 2021

ASSIGNMENT 2 - SOLUTIONS

Problem 1 [15 marks]

(a) [3 marks] This is the R code

```
tennis<-read.csv("tennis.csv",header=TRUE)
Grp1<-tennis[1:50,]
Grp2<-tennis[51:100,]
n1<-nrow(Grp1)
n2<-nrow(Grp2)

#Group 1 (top50) multivariate analysis
Y<-Grp1
n<-n1
ybar<-apply(Y,2,mean) #sample means for each variable
Sigma<-cov(Y) #sample covariance matrix
THETA_Grp1<-SIGMA_Grp1<-NULL #objects to store posterior draws in

##Prior parameters
mu0<-ybar
L0<-S0<-Sigma
nu0<-dim(Y)[2]+2

set.seed(1)
for (s in 1:10000){
  ###update theta
  Ln<-solve(solve(L0)+n*solve(Sigma))
  mun<-Ln%*%(solve(L0)%*%mu0+n*solve(Sigma)%*%ybar)
  theta<-rmvnorm(1,mun,Ln)

  ##update Sigma
  Sn<-S0+(t(Y)-c(theta))%*%t(t(Y)-c(theta))
  Sigma<-solve(rwish(1,nu0+n,solve(Sn)))

  ##save results
  THETA_Grp1<-rbind(THETA_Grp1,theta)
  SIGMA_Grp1<-rbind(SIGMA_Grp1,c(Sigma))
```

```
}

#Group 2 (rank 51-100) multivariate analysis
Y<-Grp2
n<-n2
ybar<-apply(Y,2,mean) #sample means for each variable
Sigma<-cov(Y) #sample covariance matrix
THETA_Grp2<-SIGMA_Grp2<-NULL #objects to store posterior draws in

##Prior parameters
mu0<-ybar
L0<-S0<-Sigma
nu0<-dim(Y)[2]+2

set.seed(1)
for (s in 1:10000){
  ###update theta
  Ln<-solve(solve(L0)+n*solve(Sigma))
  mun<-Ln%*%(solve(L0)%*%mu0+n*solve(Sigma)%*%ybar)
  theta<-rmvnorm(1,mun,Ln)

  ##update Sigma
  Sn<-S0+(t(Y)-c(theta))%*%t(t(Y)-c(theta))
  Sigma<-solve(rwish(1,nu0+n,solve(Sn)))

  ##save results
  THETA_Grp2<-rbind(THETA_Grp2,theta)
  SIGMA_Grp2<-rbind(SIGMA_Grp2,c(Sigma))
}
```

(1 mark) correct prior parameter assumptions
(1 mark) correct code for conditional posterior distributions
(1 mark) correct separation of data into two groups

(b) We can estimate the posterior mean of the correlation matrix for each group using the following code.

```
p<-dim(Y)[2]
COR_Grp1 <- array( dim=c(p,p,10000) )
for(s in 1:10000)
{
  Sig<-matrix( SIGMA_Grp1[s,] ,nrow=p,ncol=p)
  COR_Grp1[,,s] <- Sig/sqrt( outer( diag(Sig),diag(Sig) ) )
}

COR_Grp2 <- array( dim=c(p,p,10000) )
for(s in 1:10000)
{
  Sig<-matrix( SIGMA_Grp2[s,] ,nrow=p,ncol=p)
  COR_Grp2[,,s] <- Sig/sqrt( outer( diag(Sig),diag(Sig) ) )
}
```

The estimates of the correlation between each pair of variables (based on the posterior mean of the correlation matrix) are shown in the tables below

	1st_Serve %	1st_Serve_Pts%	2nd_Serve_Pts%	BPts_Saved%	ServG_Won%
1st_Serve %	1.00000	-0.6952	-0.3780	-0.03915	-0.4202
1st_Serve_Pts%	-0.69520	1.0000	0.3581	0.44379	0.8229
2nd_Serve_Pts%	-0.37797	0.3581	1.0000	0.22517	0.5334
BPts_Saved%	-0.03915	0.4438	0.2252	1.00000	0.7352
ServG_Won%	-0.42019	0.8229	0.5334	0.73521	1.0000

Table 1: Estimated correlations - Top 50 players

	1st_Serve %	1st_Serve_Pts%	2nd_Serve_Pts%	BPts_Saved%	ServG_Won%
1st_Serve %	1.00000	-0.5396	0.2461	0.0385	0.08989
1st_Serve_Pts%	-0.53957	1.0000	-0.1129	0.2043	0.55397
2nd_Serve_Pts%	0.24607	-0.1129	1.0000	-0.1169	0.40528
BPts_Saved%	0.03850	0.2043	-0.1169	1.0000	0.40180
ServG_Won%	0.08989	0.5540	0.4053	0.4018	1.00000

Table 2: Estimated correlations - Top 51-100 players

Based on the results in Tables 1 and 2, we observe the following notable differences

- Strong positive correlation between First_Serve_Points% and Service_Games_Won%, and Break_Points_Saved% and Service_Games_Won% for the Top 50 players, whereas the correlation for these pairs of variables is only moderate for players ranked 51-100.
- The estimated correlations are larger in absolute value in Group 1 (Top 50 players)
- First_Serve% is negatively correlated with all other variables in Group 1, but positively correlated with all but one variable in Group 2.

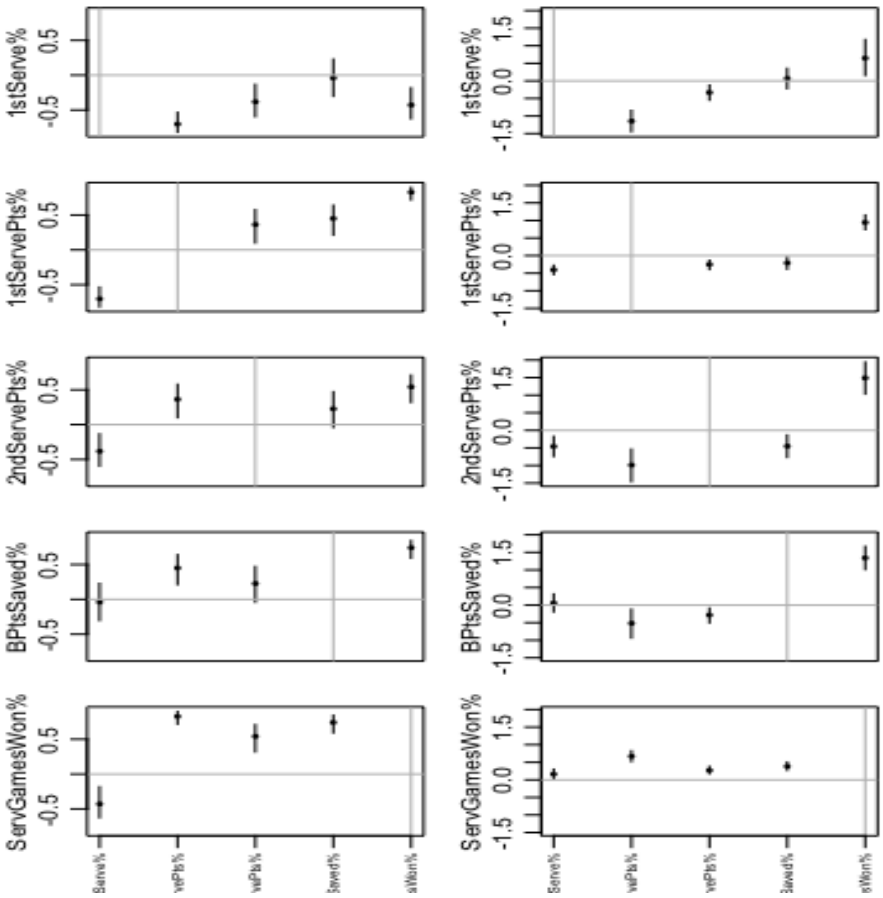

Figure 1: 95% posterior intervals for correlations - Group 1 (Top 50 players)

In the posterior interval plots for the correlations, we see that the regression adjusted correlations are much more similar between the two groups. The confidence intervals for Group 1 are narrower than the confidence intervals for Group 2. The stronger

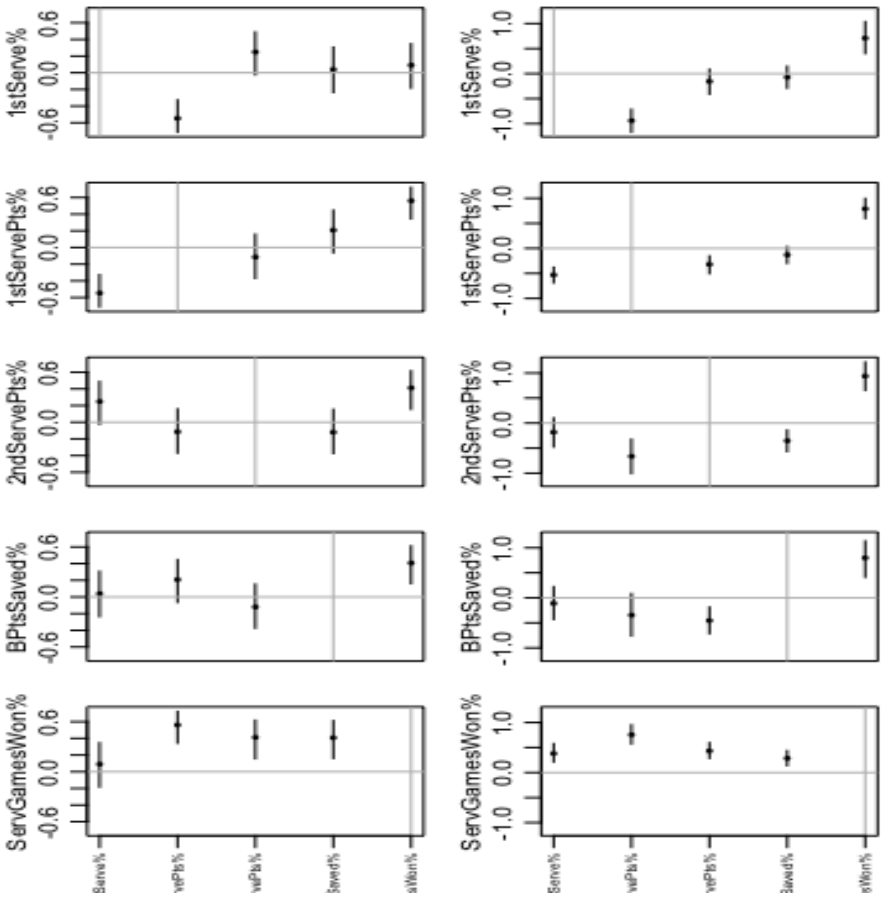

Figure 2: 95% posterior intervals for correlations - Group 2 (Top 51-100 players)

correlations in Group 1 and lower posterior variance suggest that players in the top 50 demonstrate specific match performance combinations that are associated with being a top 50 player, whereas evidence for such 'winning' combinations is not as strong for those players outside the top 50.

(1 mark) for estimate of correlations based on posterior mean
(1 mark) for plot of 95% posterior confidence intervals for the correlations
(2 marks) for at least valid comments on differences between correlations. (note practical explanation for reason behind differences not required)

(c) [4 marks] The code to produce the marginal posterior plots is

```
for (i in 1:5){
  xname<-bquote(theta[.(names(Y)[i])])
  xl<-c(min(range(THETA_Grp1[,i])[1],range(THETA_Grp2[,i])[1]),
       max(range(THETA_Grp1[,i])[2],range(THETA_Grp2[,i])[2]))
  d_Grp1<-density(THETA_Grp1[,i])
  d_Grp2<-density(THETA_Grp2[,i])
  yl<-c(min(min(d_Grp1[2]$y),min(d_Grp2[2]$y)),max(max(d_Grp1[2]$y),max(d_Grp2[2]$y)))
  plot(d_Grp1,xlab=xname,ylab="",main="",xlim=xl,ylim=yl,col="red",cex.axis=0.8)
  lines(d_Grp2)
  legend(x="topleft",c("Grp1","Grp2"),col=c("red","black"),lty=c(1,1),cex=0.5)
}
```

From Figure 3 we see that Top 50 players have a higher average percentage of points won on first serve or second serve, a higher average percentage of break points saved and a higher average percentage of service games won. The marginal posterior density plots for First_Serve% overlap each other and there does not appear to be a significant difference in average percentage of first serves in between the two groups. These observations are verified by estimation of the marginal posterior probabilities $Pr(\theta_{1,k} > \theta_{2,k}|\mathbf{Y})$ as below.

```
pp<-NULL
for (i in 1:5){
  pp<-c(pp,mean(THETA_Grp1[,i]>THETA_Grp2[,i]))
  }
names(pp)<-names(Y)
```

	1st_Serve %	1st_Serve_Pts%	2nd_Serve_Pts%	BPts_Saved%	ServG_Won%
$Pr(\theta_{1,k} > \theta_{2,k}\|\mathbf{Y})$	0.0014	1.000	1.000	1.000	1.000

Table 3: $Pr(\theta_{1,k} > \theta_{2,k}|\mathbf{Y})$

(2 marks) marginal posterior density comparison plot
(1 mark) estimation of $Pr(\theta_{1,k} > \theta_{2,k}|\mathbf{Y})$
(1 mark) Stating which variables show differnces

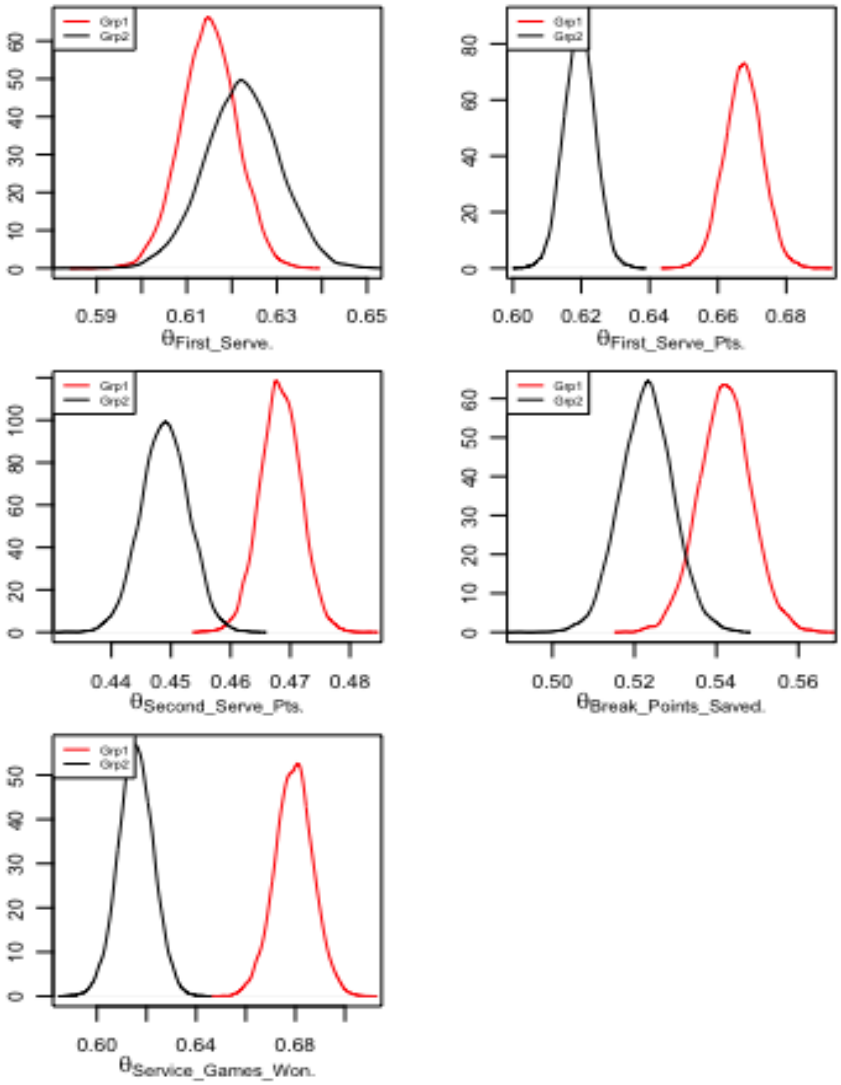

Figure 3: Marginal posterior distribution comparison for θ_j

(d) The effective sizes calculated using the code below show that all effective sizes are at least 8000 which is more than enough to justify independence for valid Monte Carlo approximations. Note some effective sizes are greater than 10000 (the total number of simulations run) which suggests there may be negative autocorrelations on odd lags.

```
> apply(THETA_Grp1, 2, function(x) effectiveSize(x))
       First_Serve.    First_Serve_Pts.   Second_Serve_Pts. Break_Points_Saved.
              10363               10000               10000               10000
 Service_Games_Won.
              10000
> apply(THETA_Grp2, 2, function(x) effectiveSize(x))
       First_Serve.    First_Serve_Pts.   Second_Serve_Pts. Break_Points_Saved.
              10357               10000               10111               10000
 Service_Games_Won.
              10000
> apply(SIGMA_Grp1, 2, function(x) effectiveSize(x))
 [1]  9563  9593  9532  9374  9647  9593  9706 10503  9338  9663  9532 10503  9890
[14]  9524 10566  9374  9338  9524  8854  9134  9647  9663 10566  9134  9448
> apply(SIGMA_Grp2, 2, function(x) effectiveSize(x))
 [1]  9655 10000 10000  9119  9614 10000 10000  9967  9381 10000 10000  9967  9908
[14]  9666 10550  9119  9381  9666  9528  8724  9614 10000 10550  8724  9468
```

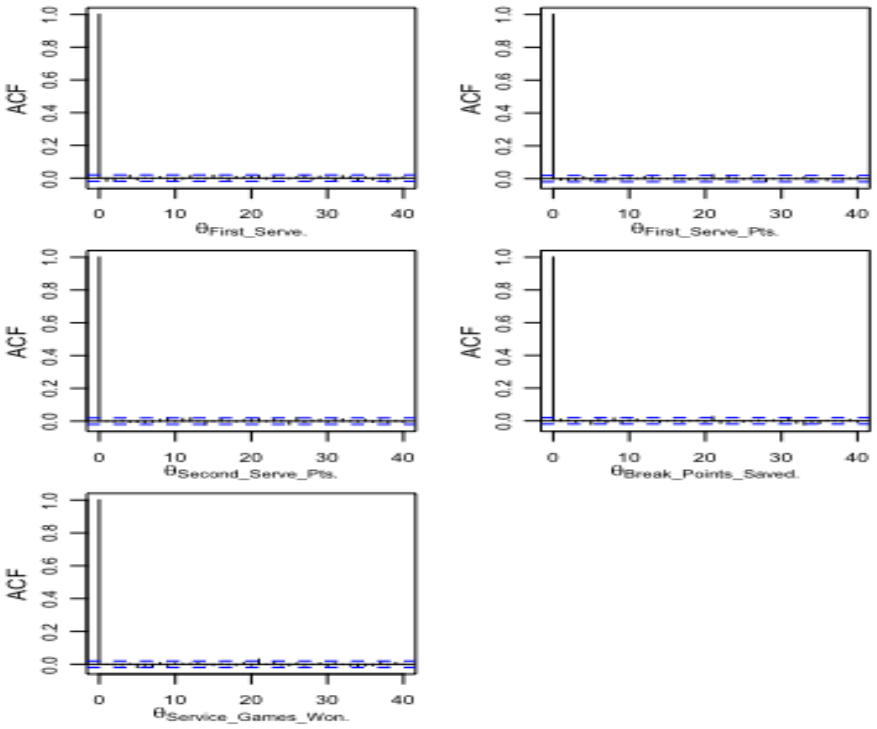

Figure 4: ACF plot θ_k - Group 1

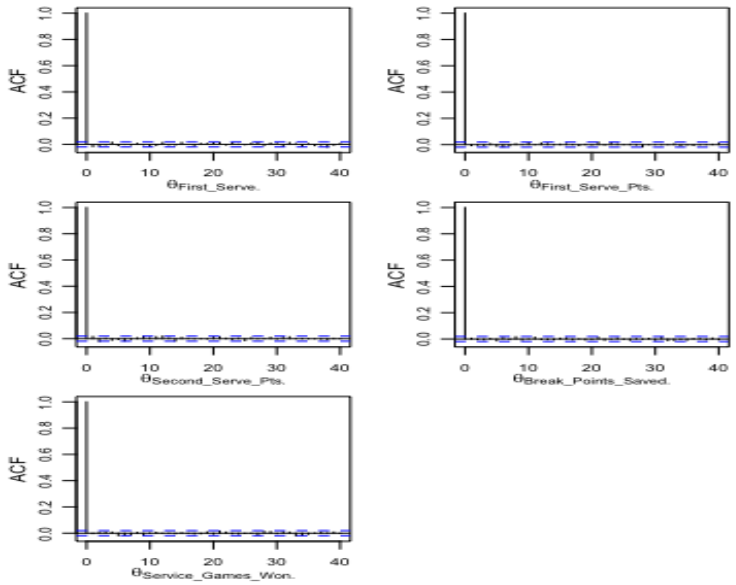

Figure 5: ACF plot θ_k - Group 2

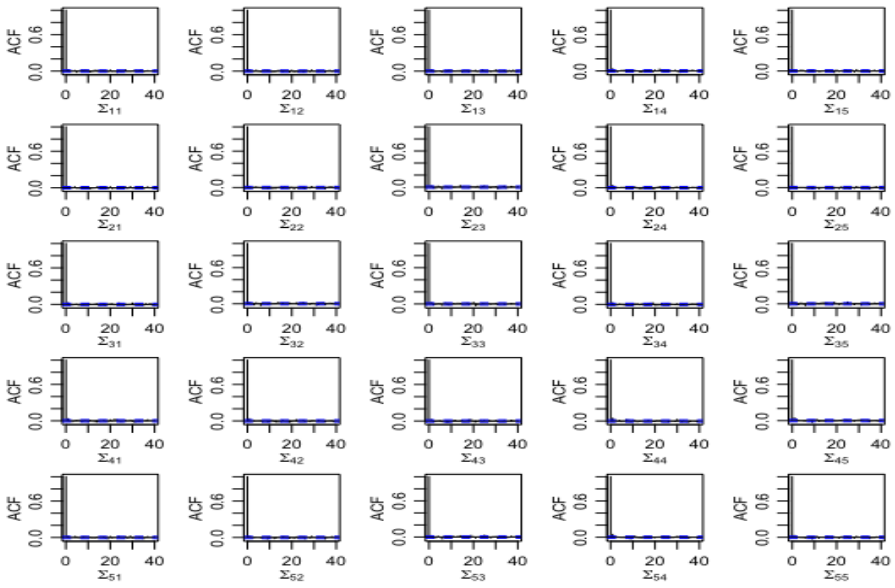

Figure 6: ACF plot Σ - Group 1

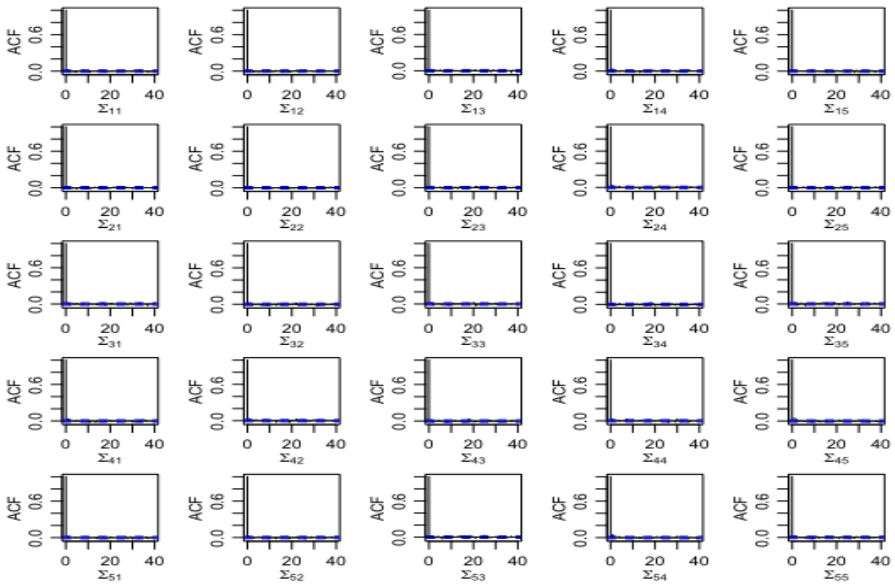

Figure 7: Stationarity plot Σ - Group 2

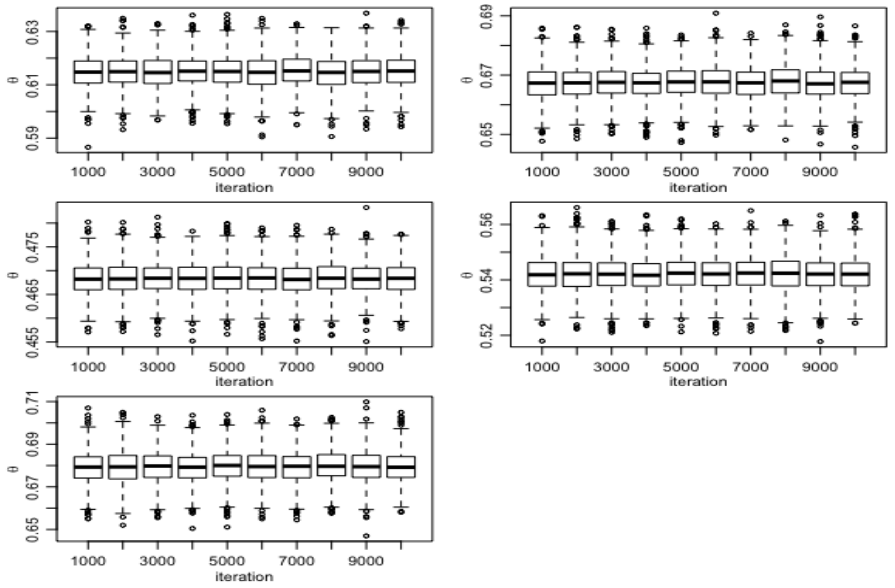

Figure 8: Stationarity plot θ_k - Group 1

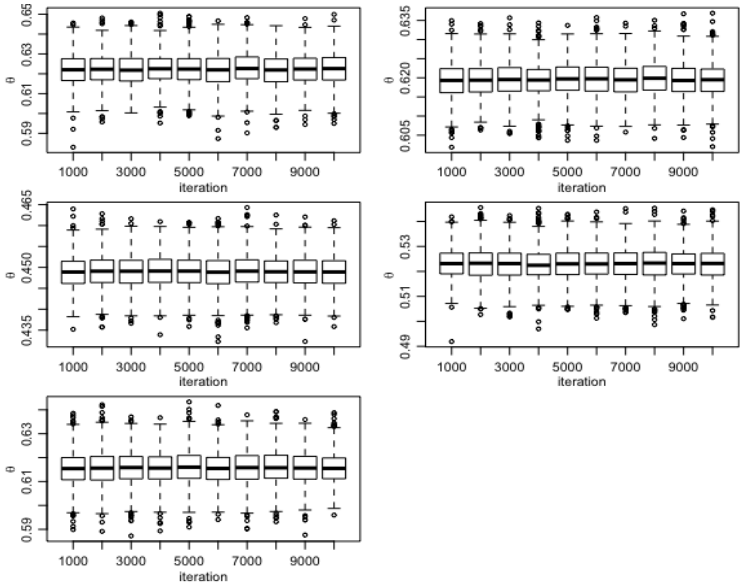

Figure 9: Stationarity plot θ_k - Group 2

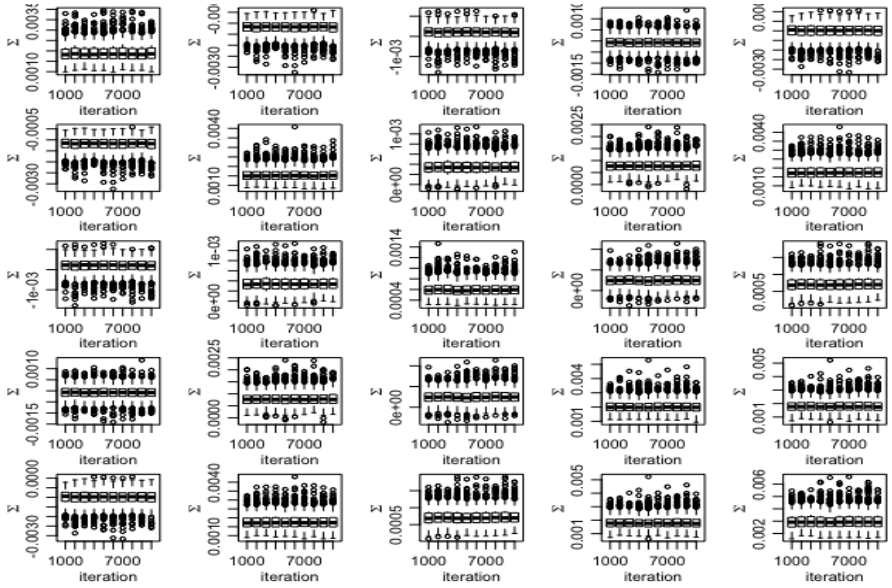

Figure 10: Stationarity plot Σ - Group 1

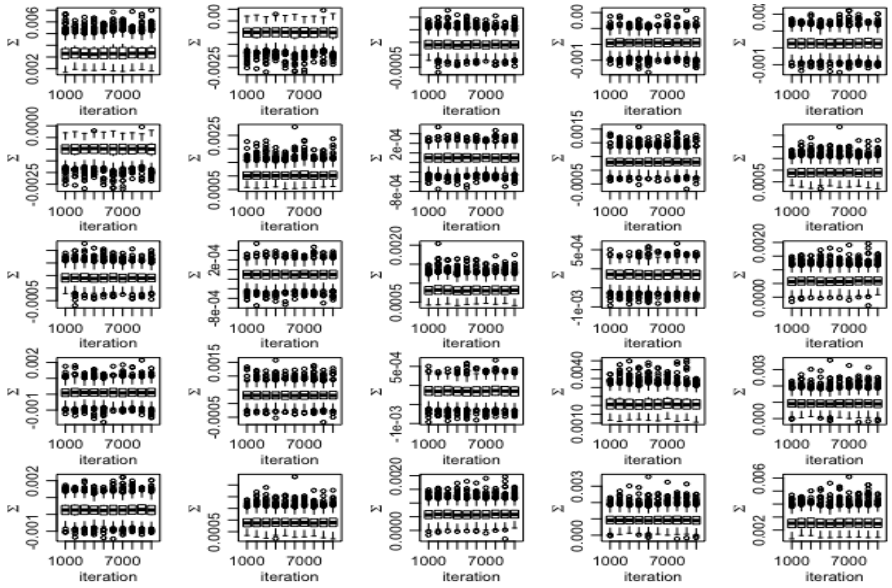

Figure 11: Stationarity plot Σ - Group 2

ACF plots for posterior draws of θ_k and $\Sigma_{i,j}$ for both groups show minimal autocorrelation. The stationarity ploys show all Markov Chains have converged.

(1 mark) ACF plots and comments
(1 mark) stationarity plots and comments
(1 mark) effective Sizes and comments
(1 mark) provide comments and plots on both θ and σ

Problem 2 [15 marks]
Using the same `tennis.csv` data set from Problem 1, your task is to run a Bayesian linear regression analysis to assess the relative importance of the variables First_Serve%, First_Serve_Pts%, Second_Serve_Pts% and Break_Points_Saved% as predictors of Service_Games_Won%. Include all two-way interaction terms in your model. In this question you do not need to split the players into Group 1 and Group 2. Assume weakly informative priors for the parameters of your model.

(a) The Bayesian model assumptions are:

Sampling model: $\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2 \sim MVN(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

Prior assumptions: $\boldsymbol{\beta}|\sigma^2, \mathbf{X} \sim MVN(\mathbf{0}, n\sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$ (weakly informative g-prior on $\boldsymbol{\beta}$ with $k = g\sigma^2$ and $g = n$).

$\sigma^2 \sim InvGamma(\nu_0/2 = 1/2, \nu_0\sigma_0^2/2 = \hat{\sigma}^2_{ols}/2)$ (unit prior on σ^2)

where $\mathbf{X}$ is a 100×11 matrix with a first column of 1's (intercept term), columns two to four are for the main effects for each of the four predictors, and the remaining columns represent the values of the six two-way interaction terms (note here, quadratic terms x_i^2 are ignored but can be included).

Let $z_j = 1$ if variable j is selected to be included in the model ($j = 1, .., 10$). Let $\mathbf{X}_z$ denotes the columns of $\mathbf{X}$ for which the corresponding model selection indicator variable $z_j = 1$. The steps of the sampling algorithm at iteration t are are:

Step 1: update model selection indicator variables $\mathbf{z} = c(z_1, z_2, z_3, z_4, ...z_{15}, z_{16})$
Step 1a: Set $\mathbf{z} = z^{(t-1)}$
Step 1a: For $j \in \{1, ..., 16\}$ in random order, replace z_j with a sample from $p(z_j|\mathbf{z}_{-j}, \mathbf{y}, \mathbf{X})$
Step 1a: Set $\mathbf{z}^{(t)} = \mathbf{z}$

Step 2: update σ^2 (residual variance of linear regression model)

Sample $\sigma^2_{(t)} \sim InvGamma((\nu_0 + n)/2, (\nu_0\sigma_0^2 + SSR_g)/2)$.
Assuming weakly informative priors we have $\nu_0 = 1$ and $\sigma_0^2 = \hat{\sigma}^2_{OLS}$ and $g = n$. ($\hat{\sigma}^2_{OLS} = 0.0005065$
$SSR_g = \mathbf{y}^T\left(\mathbf{I} - \frac{g}{g+1}\mathbf{X}_z(\mathbf{X}_z^T\mathbf{X}_z)^{-1}\mathbf{X}_z\right)\mathbf{y}$

Step 3: update $\boldsymbol{\beta}$ (vector of regression coefficients)

Sample $\boldsymbol{\beta}^{(t)} \sim MVN(\boldsymbol{\beta}_m, V_{\boldsymbol{\beta}})$.
where $\boldsymbol{\beta}_m = V_{\boldsymbol{\beta}}\mathbf{X}_z^T\mathbf{y}$ and $V_{\boldsymbol{\beta}} = \frac{g}{g+1}\sigma^2_{(t)}(\mathbf{X}_z^T\mathbf{X}_z)^{-1}$

(Note: students answers should be almost a direct copy of lecture slides)
(1 mark) for each correctly stated conditional distribution, must state the form of the conditional posterior distribution (eg inverse gamma or MVN, and provide parameters of conditional posterior distribution)
(1 mark) statement of assumed values for prior parameters, that is, ν_0, σ_0^2, $\boldsymbol{\beta}_0$, Σ_0.

(b) The R code is

```
source("regression_gprior.R")

Y<-tennis[1:100,5]
X<-tennis[1:100,1:4]
n<-length(Y)
X<-cbind(c(rep(1,n)),X)
colnames(X)<-c("Intercept","X1","X2","X3","X4")
#main effects
X1<-X[,2]
X2<-X[,3]
X3<-X[,4]
X4<-X[,5]
#two-way interactions
X1X2<-X1*X2
X1X3<-X1*X3
X1X4<-X1*X4
X2X3<-X2*X3
X2X4<-X2*X4
X3X4<-X3*X4
X1sq<-X1^2
X2sq<-X2^2
X3sq<-X3^2
X4sq<-X4^2


X<-cbind(X,X1X2,X1X3,X1X4,X2X3,X2X4,X3X4)
X<-as.matrix(X)

p<-dim(X)[2] # number of regression coefficients including intercept
S<-20000

BETA<-Z<-matrix(NA,S,p) #object to store posterior draws in
S2<-NULL
```

```
z<-rep(1,dim(X)[2] ) #starting values of z
lpy.c<-lpy.X(Y,X[,z==1,drop=FALSE]) #starting marginal log likelihood of data

for(s in 1:S)
{
  for(j in sample(2:p)) #always include intercept term
  {
    zp<-z ; zp[j]<-1-zp[j]
    lpy.p<-lpy.X(Y,X[,zp==1,drop=FALSE])
    r<- (lpy.p - lpy.c)*(-1)^(zp[j]==0)
    z[j]<-rbinom(1,1,1/(1+exp(-r)))
    if(z[j]==zp[j]) {lpy.c<-lpy.p}
  }

  beta<-z
  if(sum(z)>0){
    temp<-lm.gprior(Y,X[,z==1,drop=FALSE],S=1)
    beta[z==1]<-temp$beta
    s2<-temp$s2}

  Z[s,]<-z
  BETA[s,]<-beta
  S2<-c(S2,s2)
}
```

(c)
```
par(mfrow=c(1,1))
plot(apply(Z,2,mean,na.rm=TRUE),xlab="regressor index",ylab=expression(
  paste( "Pr(",italic(z[j] == 1),"|",italic(y),",X)",sep="")),type="h",lwd=2)
```

Figure 12 shows the estimated posterior probabilities $Pr(z_j = 1|\mathbf{y}, \mathbf{X})$. We see that only the model inclusion indicator for the 9th regression coefficient (corresponding to the variable for the interaction term between the variables First_Serve_Pts% and Second_Serve_Pts%) has posterior probability greater than 0.5. So only the joint behaviour of the variables First_Serve_Pts% and Second_Serve_Pts% is strongly predictive of Service_Games_Won%. This suggests that the combination of winning the point on either first or second serve is key to winning the service game which makes sense as winning points on serve gets the player closer to winning the game.

(1 mark) provide numerical values for or a plot of $Pr(z_j = 1|\mathbf{y}, \mathbf{X})$
(1 mark) valid comment on which variables are strongly predictive (note practical reasoning not required).

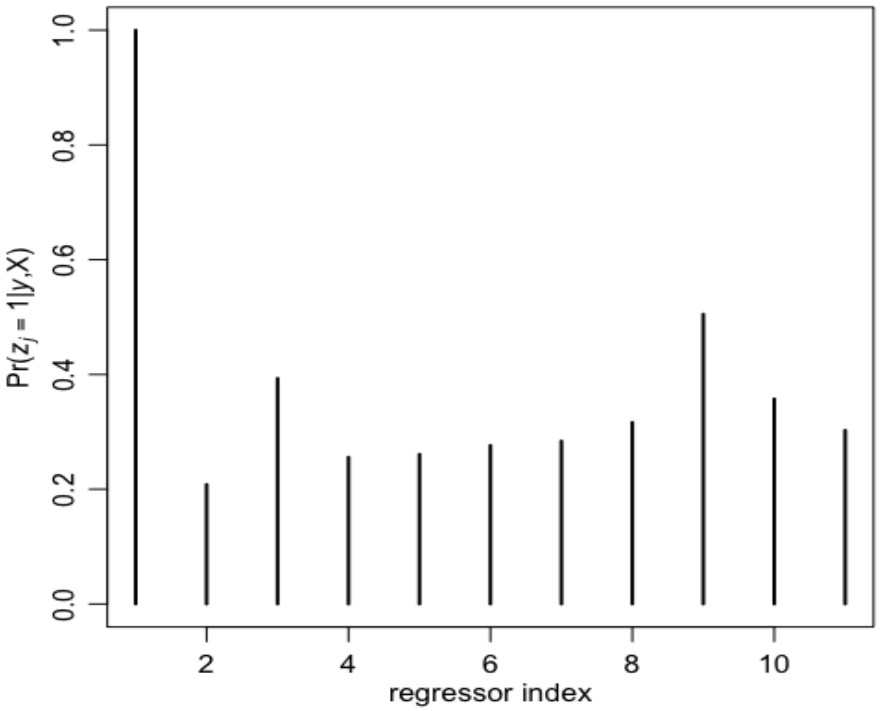

Figure 12: Posterior probabilities that each coefficient is non-zero

(d) The R code to create posterior predictive residuals and associated plots is

```
y.pred<-NULL #object to store posterior predictive values
for (s in 1:S){
  beta<-BETA[s,]
  s2<-S2[s]
  y.pred<-rbind(y.pred,c(X%*%beta)+rnorm(n,0,sqrt(s2)))

}
#Residual plot diagnostics
e.pred<-NULL
for (i in 1:S){
  e.pred<-rbind(e.pred,Y-y.pred[s,])
}

e.pred.avg<-apply(e.pred,2,mean)
par(mfrow=c(1,1))
```

```
plot(e.pred.avg,pch=19)
abline(h=0)
```

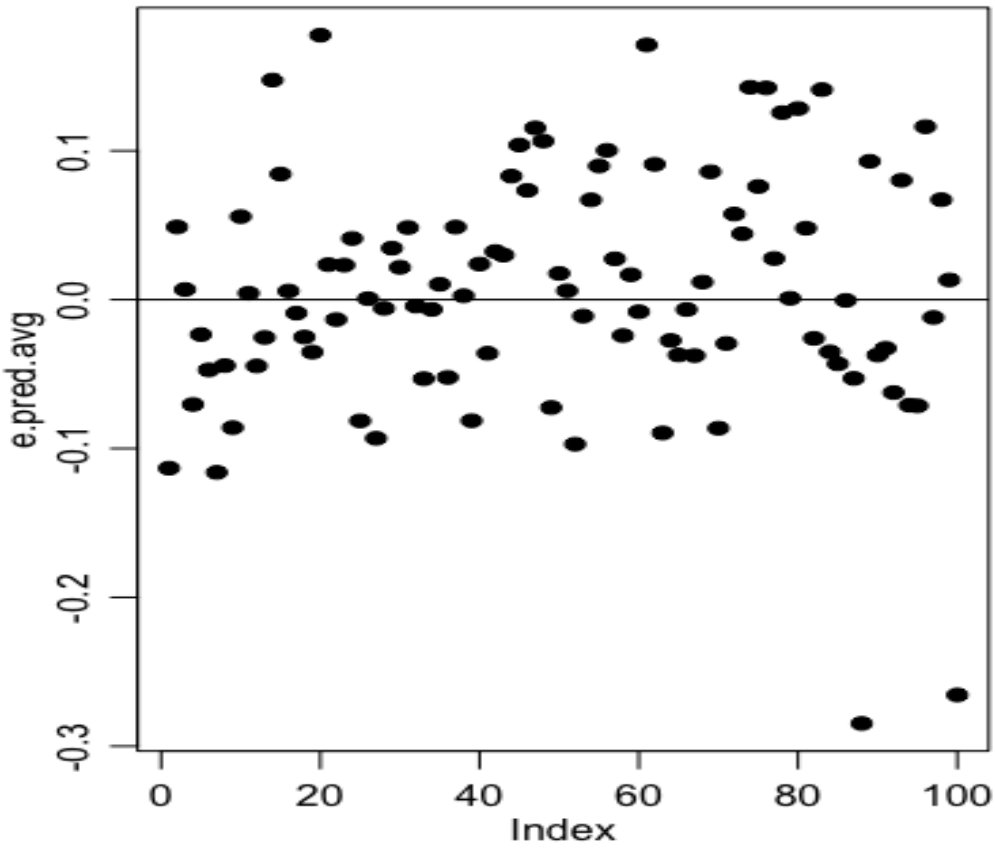

Figure 13: Posterior predictive residual plot

The residual plot based on the average posterior predictive residual value for each player is shown in Figure 13. The plot shows a random scatter of points and no issues which would indicate violation of the normal linear model assumptions.

(2 marks) for residual plot(s)
(1 mark) for valid commentary

(e) Initial ACF plots showed high autocorrelation up to lag 20 so the sequence of draws was thinned by taking every 20th iteration. The ACF plots shown in Figure 14 show no issues with high autocorrelation as all plots decay quickly to zero.

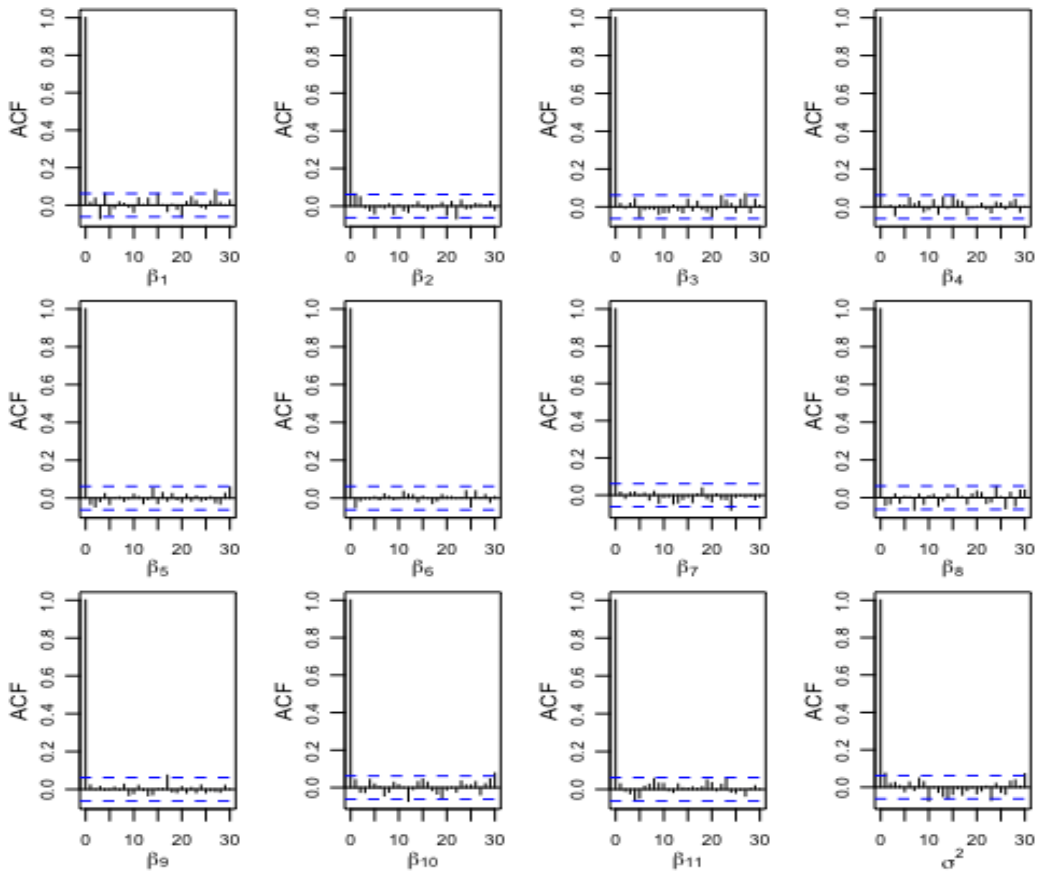

Figure 14: ACF plots of every 20th value of the Markov Chain

After thinning most effective sizes are equal to 1000 which is just enough to provide valid Monte Carlo approximations. The effective size for β_1 is below 1000, but as the variable First_Serve% was not identified as strongly predictive, this is not of great concern. The effective size for σ^2 could be improved by running the chain for more than 20000 iterations.

```
> apply(BETA[thin,],2,function(x) effectiveSize(x))
 [1] 1009.5  879.6 1001.0 1001.0 1001.0 1102.5 1001.0 1001.0 1001.0 1001.0 1001.0
```

```
> effectiveSize(S2[thin])
 var1
861.8
```

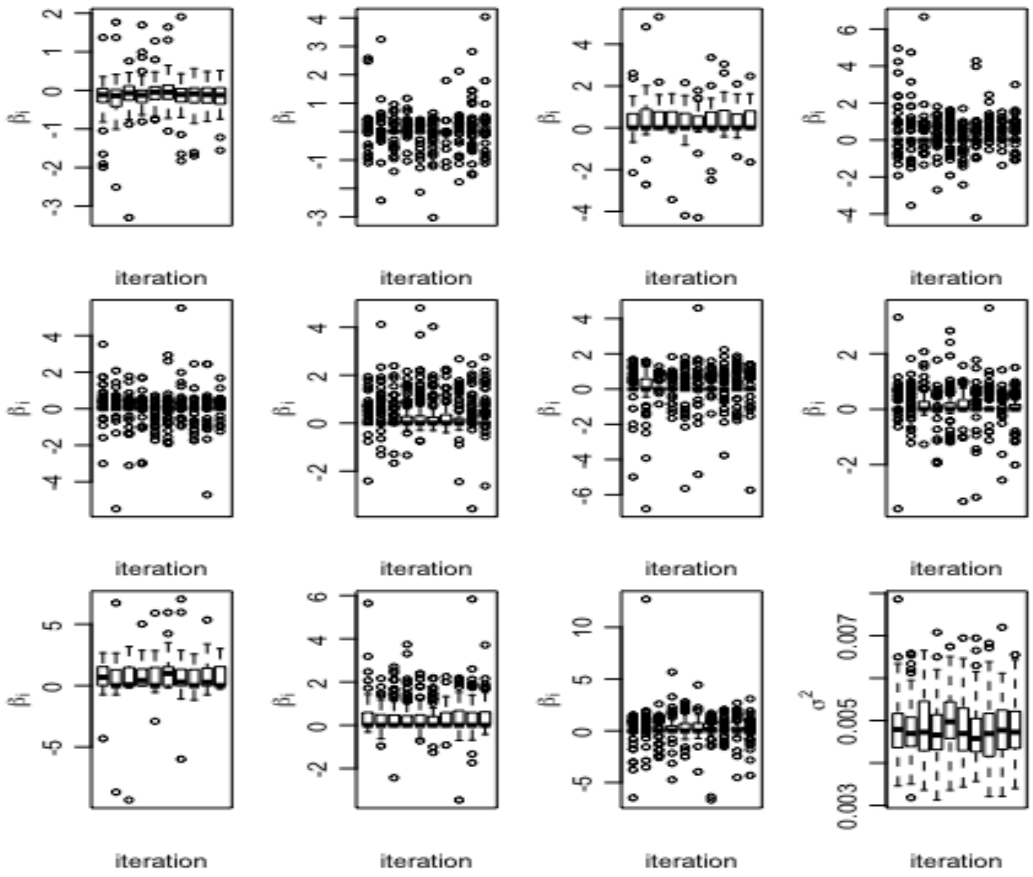

Figure 15: Stationarity Plots

The stationarity plots show the sequence of draws of each parameter have reached convergence. (Note diagnostic plots for the z_j indicator variables are less informative and not shown given the binary nature of the outcome).

(1 mark) ACF plots and comment
(1 mark) effective sizes and comment
(1 mark) stationarity plot and comment

Note students can receive full marks if recognise that σ^2 and $\boldsymbol{\beta}$ are direct samples from $p(\sigma^2, \boldsymbol{\beta}|\mathbf{y}, \mathbf{X})$ and hence stationarity convergence diagnostics are not required.

Problem 3 [15 marks]

(a) μ represents the common mean (for all training groups). We can assume a semi-conjugate prior, $\mu \sim Normal(\mu_0, \gamma_0^2)$, and set the hyperprior mean $\mu_0 = 0$ (that is, a priori, we assume no ferritin loss). For a weakly informative prior, assign a relatively large value for the hyperprior variance $\gamma_0^2 = 1000$.

(1 mark) state normal distribution
(1 mark) justification of prior mean (note other justifications plausible eg μ_0 equal to overall sample mean or average of sample mean of three groups
(1 mark) justification of prior variance (note a small value of γ_0^2 must be appropriately justified).

(b) At iteration t of the Gibbs sampler

- Step 1: For $j = 1, ..., 3$, sample θ_j from its full conditional distribution $\theta_j|\mathbf{y}_j, \mu_{(t)}, \sigma^2_{(t)}, \tau^2_{(t)}, \sim Normal\left(\frac{\bar{y}_j n_j/\sigma^2_{(t)} + \mu_{(t)}/\tau^2_{(t)}}{n_j/\sigma^2_{(t)} + 1/\tau^2_{(t)}}, \frac{1}{n_j/\sigma^2_{(t)} + 1/\tau^2_{(t)}}\right)$ where $\bar{y}_j$ and n_j are the observed sample means and sample sizes for ferritin loss in group j as provided in the summary statistics (Table 4) $\mathbf{y}_j$ denotes the data vector of ferritin loss in group j, and the subscript $_{(t)}$ refers to the parameter values at iteration t.
- Step 2: sample σ^2 from its full conditional distribution $\sigma^2|\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \theta_{1,(t+1)}, \theta_{2,(t+1)}, \theta_{3,(t+1)} \sim InverseGamma(\frac{\nu_0+n}{2}, \frac{\nu_0\sigma_0^2 + \sum_{j=1}^3 (n_j-1)s_j^2 + n_j(\bar{y}_j - \theta_{j,(t+1)})^2}{2})$ where $n = \sum_{j=1}^3 n_j = 22$ and s_j^2 are the observed sample variances of ferritin loss for each group.
- Step 3: sample μ from its full conditional distribution $\mu|\theta_{1,(t+1)}, \theta_{2,(t+1)}, \theta_{3,(t+1)}, \tau^2_{(t)} \sim Normal\left(\frac{3\bar{\theta}_{(t+1)}/\tau^2_{(t)} + \mu_0/\gamma_0^2}{3/\tau^2_{(t)} + 1/\gamma_0^2}, \frac{1}{3/\tau^2_{(t)} + 1/\gamma_0^2}\right)$ where $\bar{\theta}_{(t+1)} = \frac{1}{3}\sum_{j=1}^3 \theta_{j,(t+1)}$
- Step 4: sample τ^2 from its full conditional distribution $\tau^2|\theta_{1,(t+1)}, \theta_{2,(t+1)}, \theta_{3,(t+1)}, \mu_{(t+1)} \sim InverseGamma\left(\frac{\eta_0+3}{2}, \frac{\eta_0\tau_0^2 + \sum_{j=1}^3 (\theta_{j,(t+1)} - \mu_{(t+1)})^2}{2}\right)$

The hyperparameter values are $\mu_0 = -3.856$; $\gamma_0^2 = 1000$; $\nu_0 = 2$; $\sigma_0^2 = 1$; $\eta_0 = 2$; $\tau_0^2 = 1$.

We use empirical estimates based on the observed data as starting values for the parameters. Specifically, $\theta_{j,(t=0)} = \bar{y}_j$; $\sigma^2_{(t=0)} = \frac{1}{3}\sum_{j=1}^3 s_j^2 = 277.2$; $\mu_{(t=0)} = \bar{\theta}_{(t=0)} = -3.856$; $\tau^2_{(t=0)} = \frac{\sum_{j=1}^3 (\theta_{j,(t=0)} - \bar{\theta}_{(t=0)})^2}{3-1} = 108.8$

The Gibbs sampler was run for S=100000 iterations.

The R code to run the Gibbs sampler is

	IHE	Heat	Placebo
$\bar{y}_j$	-1.525	5.214	-15.257
s_j^2	334.5	175.5	321.4
n_j	8	7	7

Table 4: Table of summary statistics for athlete ferritin loss data

```
athlete<-read.csv("athlete.csv",header=TRUE)
athlete$y<-athlete$Post-athlete$Pre
#subset data by Group
IHE<-subset(athlete,Group=="IHE")
Heat<-subset(athlete,Group=="Heat")
Placebo<-subset(athlete,Group=="Placebo")

m<-3 #number of groups

#list object to store ferritin loss by group
Y<-list()
Y[[1]]<-IHE$y
Y[[2]]<-Heat$y
Y[[3]]<-Placebo$y

#summary statistics by group
y1<-mean(IHE$y)
y2<-mean(Heat$y)
y3<-mean(Placebo$y)

sv1<-var(IHE$y)
sv2<-var(Heat$y)
sv3<-var(Placebo$y)

n1<-length(IHE$y)
n2<-length(Heat$y)
n3<-length(Placebo$y)

ybar<-c(y1,y2,y3)
sv<-c(sv1,sv2,sv3)
n<-c(n1,n2,n3)
```

```
## weakly informative priors
nu0<-1  ; s20<-1
eta0<-1 ; t20<-1
mu0<-0 ; g20<-1000

#starting values
theta<-ybar ; sigma2<-mean(sv) ; mu<-mean(theta) ; tau2<-var(theta)

set.seed(1)
S<-100000
THETA<-matrix( nrow=S,ncol=m) #object to store posterior draws of theta_j in
MST<-matrix( nrow=S,ncol=3) #object to store posterior draws of sigma2, mu, tau

## MCMC algorithm
for(s in 1:S)
{ # sample new values of the thetas
  for(j in 1:m)
  {
    vtheta<-1/(n[j]/sigma2+1/tau2)
    etheta<-vtheta*(ybar[j]*n[j]/sigma2+mu/tau2)
    theta[j]<-rnorm(1,etheta,sqrt(vtheta)) }

  #sample new value of sigma2
  nun<-nu0+sum(n)
  ss<-nu0*s20;for(j in 1:m){ss<-ss+sum((Y[[j]]-theta[j])^2)}
  sigma2<-1/rgamma(1,nun/2,ss/2)

  #sample a new value of mu
  vmu<- 1/(m/tau2+1/g20)
  emu<- vmu*(m*mean(theta)/tau2 + mu0/g20)
  mu<-rnorm(1,emu,sqrt(vmu))

  # sample a new value of tau2
  etam<-eta0+m
  ss<- eta0*t20 + sum( (theta-mu)^2 )
  tau2<-1/rgamma(1,etam/2,ss/2)

  #store results
  THETA[s,]<-theta
  MST[s,]<-c(mu,sigma2,tau2) }
```

(1 mark) writing out steps of Gibbs sampler
(1 mark) R code
(1/2 mark) appropriate starting parameter values
(1/2 mark) correct values for hyperparameters $\sigma_0^2, \nu_0, \tau_0^2, \eta_0$.

(c) Autocorrelation plots on the full sequence of 100000 draws showed evidence of high autocorrelation up to lag 50. The sequence was thinned by taking draws from every 50th iteration. The resultant diagnostic plots are below. The ACF plots of the thinned sequence (Figure 17) show no problems with high autocorrelation and the stationarity boxplots in Figure 16 show convergence has been achieved as the distribution of the box plots for each parameter are similar.

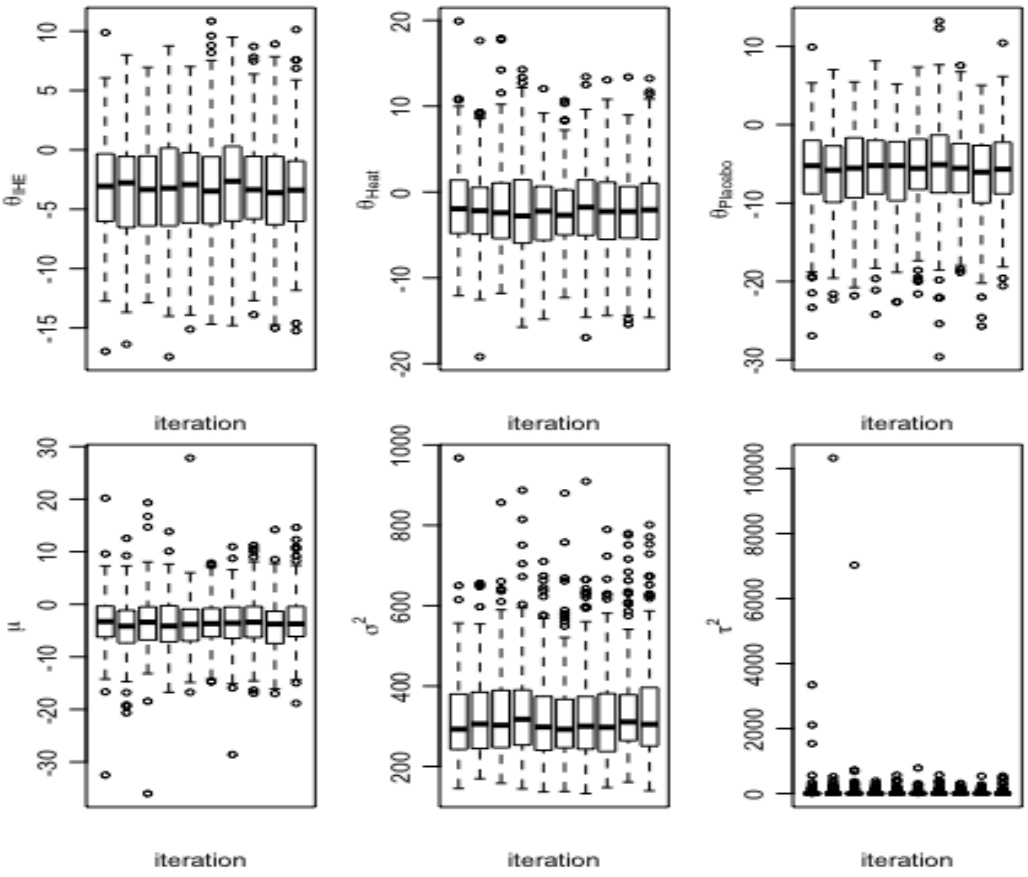

Figure 16: Stationarity Plots - ferritin loss study - thinned sequence

```
> #effective size
```

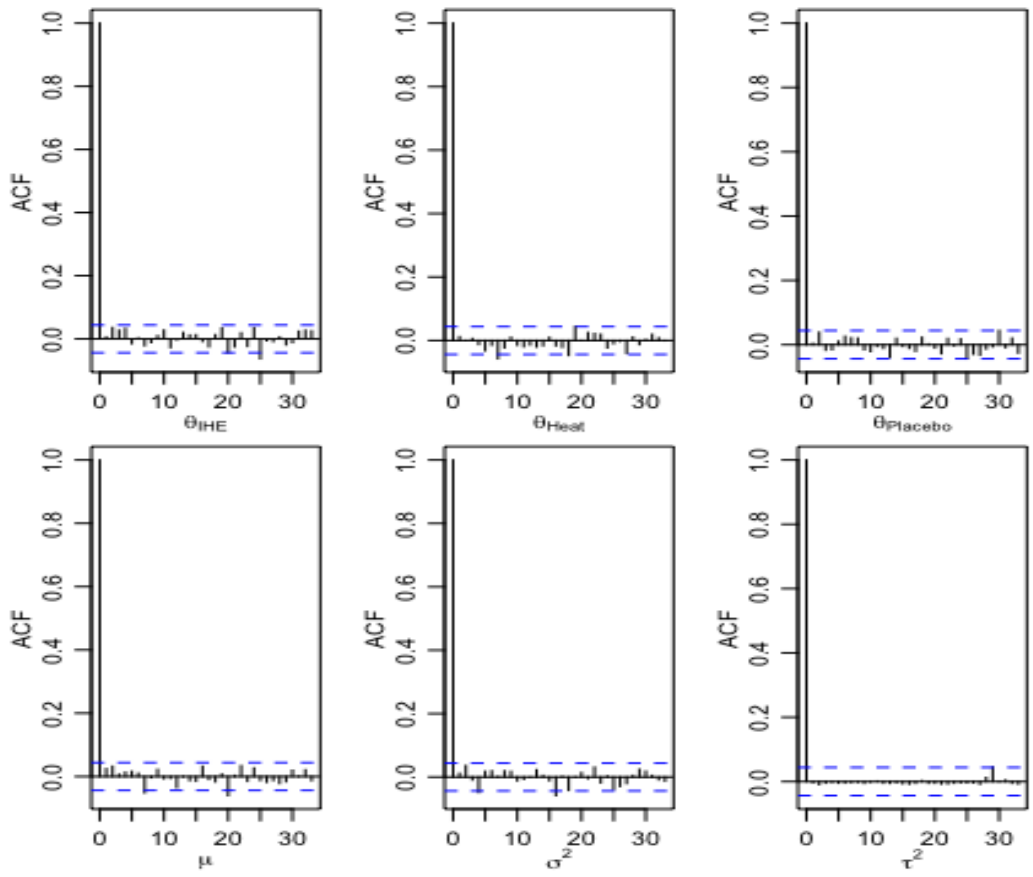

Figure 17: ACF plots - ferritin loss study - thinned sequence

```
> apply(MST[thin,],2,function(x) effectiveSize(x))
[1] 2001 2001 2001
```

The effective sizes for σ^2, μ, and τ^2 are all equal to 2001 (so > 1000) and independence among the sequence of draws for these parameters is not a problem.

(1 mark) ACF plot and valid comment
(1 mark) Stationarity plot and valid comment
(1 mark) effective sizes and valid comment

(d) For $j = 1, 2, 3$ we want to evaluate the posterior probabilities $Pr(\theta_j > max(\boldsymbol{\theta}_{-j}))$ and $Pr(\theta_j < min(\boldsymbol{\theta}_{-j}))$ where $\boldsymbol{\theta}_{-j}$ refers to the vector of mean ratings for all training groups except group$j$.

The R code to evaluate this is:

```
min.prop<-max.prop<-rep(0,m)
for (i in 1:m){
  min.ind<-apply(THETA[thin,-i],1,min)
  max.ind<-apply(THETA[thin,-i],1,max)
  min.prop[i]<-mean(THETA[thin,i]<min.ind)
  max.prop[i]<-mean(THETA[thin,i]>max.ind)
}
```

The posterior probabilities are tabulated below in Table 5:

Training Group	$Pr(\theta_j < min(\boldsymbol{\theta}_{-j}))$	$Pr(\theta_j > max(\boldsymbol{\theta}_{-j}))$	$E(\theta_j \vert y)$
IHE	0.2349	0.3243	-3.234
Heat	0.1779	0.5207	-2.060
Placebo	0.5872	0.1549	-5.827

Table 5: Posterior probabilities for ranking training programs

As the outcome variable y is defined as Post Ferritin - Pre-Ferritin, positive (or less negative) values of θ_j indicate more effectiveness at reducing iron-loss. From Table 5, we see that $Pr(\theta_j > max(\boldsymbol{\theta}_{-j}))$ is greatest for 'Heat' so we conclude that the 'Heat' training environment is most effective at reducing iron-loss. In contrast, the estimate for $Pr(\theta_j < min(\boldsymbol{\theta}_{-j}))$ is greatest for the Placebo group and so we conclude that the Placebo training environment is least effective.

(1 mark) for calculation of relevant posterior probabilities (note if only $E(\theta_j|y)$ is reported, then deduct 0.5 mark)
(1 mark) for correct interpretation of posterior probabilities calculated to identify most and least effective training programs.

(e) The 90% credible (quantile-based) intervals and interval widths for θ_j $(j = 1, 2, 3)$ are shown in Table 6 below.

Training Group	90% Interval	Interval Width
IHE	(-10.293, 3.736)	14.03
Heat	(-9.667, 6.589)	16.26
Placebo	(-15.350, 2.242)	17.59

Table 6: 90% credible intervals

The training group with the shortest interval width is 'IHE'. This group had a marginally larger sample size (8 athletes). The other two training groups had 7 athletes each respectively. So we see that there is more posterior uncertainty regarding θ_j where we have less information from the observed data because of a smaller sample size.

(1 mark) interval width calculation
(1 mark) for valid explanation

(f) A 90% posterior interval for R is (0.001, 0.185) and Figure 18 is the posterior density plot of $R = \frac{\tau^2}{\sigma^2+\tau^2}$. We can see the posterior mode of R is close to zero, indicating that there is minimal variation in average ferritin loss between training environments, and much more variation within a training environment.

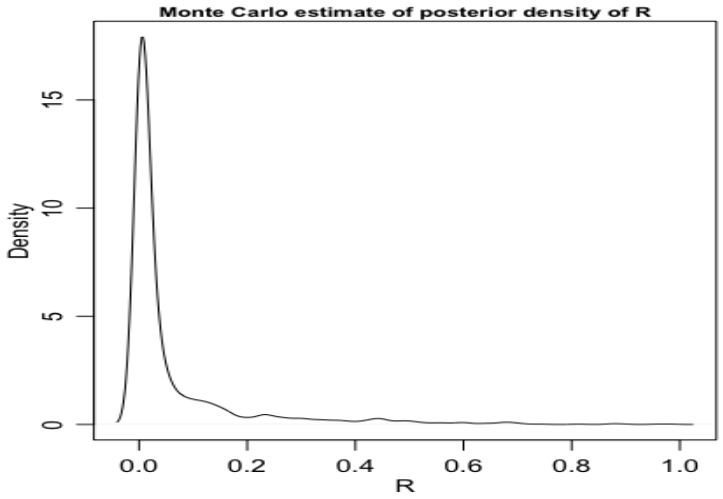

Figure 18: Posterior density plot for R

(1 mark) credible interval; (1 mark) density plot of R; (1 mark) valid interpretation

Problem 4 [STAT4116/STAT7016 ONLY] [10 marks]

(a) [1 mark]

$$y_{ijk}|z_i, \theta_j \overset{iid}{\sim} Bernoulli(\theta_j)$$

where θ_j is the probability a person in latent group j answers the question correctly. Note here we assume, the probability of answering each question correctly is the same for all questions ($k = 1, ..., 50$) within a latent group.

Note: sampling distribution must be conditional on latent group indicator z_i.
The probability of answering the question correctly needs to depend on the latent group (that is, have the subscript j) but does not need to vary by question k.

(b) [1 mark]

$$z_i|\pi \overset{iid}{\sim} Bernoulli(\pi)$$

where π is the probability a person is a visual thinker.

(c) [1 mark]

$$\theta_j|a, b \sim Beta(a, b)$$

Assume a conjugate prior for θ_j with prior parameters a and b which is common to both groups.

(d) [1 mark]

$$\pi|u, v \sim Beta(u, v)$$

Assume a conjugate prior for π with prior parameters u and v.

(e) [1 mark] $a \sim Gamma(0.001, 0.001)$ and $b \sim Gamma(0.001, 0.001)$

Note the support of the hyperprior must match the allowable range of values for a and b (specifically, $a > 0$ and $b > 0$). Setting small values for the shape and rate parameter of a Gamma distribution is equivalent to a weakly informative prior. Hyperprior distribution assumptions are not required for parameters of the Beta(u, v) prior on π as this is not a common prior shared among multiple groups (that is, the distribution for π does not vary by any group definition).

(f) [4 marks]

Define $y_i = \sum_{i=1}^{50} y_{ijk}$ (that is, the total number of correct responses from person i). So we have $y_i|z_i, \boldsymbol{\theta} \sim Binomial(50, \theta_j)$. Define $\mathbf{y} = (y_1, ..., y_n)$

Define $\theta_{j=1}$ to be the probability of answering a question correctly for a person labelled as a visual thinker, and $\theta_{j=2}$ to be the probability of answering a question correctly for a person labelled as a non-visual thinker.

At iteration t

Step 1: Update θ_j. - Gibbs step

Draw a value θ_j from its conditional posterior distribution

$$\theta_j^{(t)}|a^{(t-1)}, b^{(t-1)}, \mathbf{z}^{(t-1)}, \mathbf{y} \sim Beta(a^{(t-1)} + \sum_{i=1}^{n} y_i(\mathbb{1}_{z_i^{(t-1)}==j}), b^{(t-1)} + \sum_{i=1}^{n}(50 - y_i)(\mathbb{1}_{z_i^{(t-1)}==j}))$$

where $\mathbb{1}_{z_i^{(t-1)}==j}$ is the indicator function to indicate that the observed outcome y_i of only current latent group j members at the end of iteration (t-1) are to be included in the summation.

Step 2: Update a. - Metropolis-Hastings step

The conditional posterior distribution for a is

$$\begin{aligned} p(a^{(t)}|\theta_1^{(t)}, \theta_2^{(t)}, b^{(t-1)}) &\propto p(a) \prod_{j=1}^{2} p(\theta_j^{(t)}|a, b^{(t-1)}) \\ &\propto a^{0.001-1} \exp^{-0.001a} \times \left(\frac{\Gamma(a + b^{(t-1)})}{\Gamma(a)} \right)^2 \times \left(\prod_{j=1}^{2} \theta_j^{(t)} \right)^a \end{aligned}$$

(Note that the conditional posterior of a does not depend on the latent group indicators z_i or the binary outcomes y_i).

Step 2a:

Sample $a^* \sim Gamma(mean = a^{(t-1)}, var = \delta_a)$. The Gamma proposal density is defined on the positive real line which matches the parameter space of a. The proposal distribution is centred at the current value in the Markov Chain $a^{(t-1)}$, and δ_a is the tuning parameter for the update of a. Note, one can solve for the shape and rate parameter of the Gamma proposal distribution, such that its mean is equal to $a^{(t-1)}$ and the variance set to δ_a.

Step 2b: Compute the log of the acceptance ratio for the update of a

$$log(r) = log(p(a^*|b^{(t-1)}, \boldsymbol{\theta}^{(t)})) - log(p(a^{(t-1)}|b^{(t-1)}, \boldsymbol{\theta}^{(t)})) + log(J(a^{(t-1)}|a^*)) - log(J(a^*|a^{(t-1)}))$$

Step 2c: Sample $u \sim Unif(0,1)$. Set $a^{(t)} = a^*$ if $log(u) < log(r)$ and to $a^{(t-1)}$ otherwise.

Step 3: Update b. - Metropolis-Hastings step

The conditional posterior distribution for b is

$$\begin{aligned} p(b^{(t)}|\theta_1^{(t)}, \theta_2^{(t)}, a^{(t)}) &\propto p(b) \prod_{j=1}^{2} p(\theta_j^{(t)}|a^{(t)}, b) \\ &\propto b^{0.001-1} \exp^{-0.001b} \left(\frac{\Gamma(a^{(t)} + b)}{\Gamma(b)} \right)^2 \times \left(\prod_{j=1}^{2} (1 - \theta_j^{(t)}) \right)^b \end{aligned}$$

(Note that the conditional posterior of b does not depend on the latent group indicators z_i or the binary outcomes y_i).

Step 3a:

Sample $b^* \sim Gamma(mean = b^{(t-1)}, var = \delta_b)$. The Gamma proposal density is defined on the positive real line which matches the parameter space of b. The proposal distribution is centred at the current value in the Markov Chain $b^{(t-1)}$, and δ_b is the tuning parameter for the update of b. Note, one can solve for the shape and rate parameter of the Gamma proposal distribution, such that its mean is equal to $b^{(t-1)}$ and the variance set to δ_b.

Step 3b: Compute the log of the acceptance ratio for the update of b

$$log(r) = log(p(b^*|a^{(t)}, \boldsymbol{\theta}^{(t)})) - log(p(b^{(t-1)}|a^{(t)}, \boldsymbol{\theta}^{(t)})) + log(J(b^{(t-1)}|b^*)) - log(J(b^*|b^{(t-1)}))$$

Step 3c: Sample $u \sim Unif(0,1)$. Set $b^{(t)} = b^*$ if $log(u) < log(r)$ and to $b^{(t-1)}$ otherwise.

Step 4: Update latent group indicators z_i. For $i = 1, ..., 200$, evaluate

$$\begin{aligned} r_i = p(z_i = 1|y_i, \boldsymbol{\theta}^{(t)}, \pi^{(t-1)}) &= \frac{p(z_i = 1)p(y_i|z_i = 1, \theta_1^{(t)})}{p(z_i = 1)p(y_i|z_i = 1, \theta_1^{(t)}) + p(z_i = 0)p(y_i|z_i = 0, \theta_2^{(t)})} \\ &= \frac{\pi^{(t-1)}(\theta_1^{(t)})^{y_i}(1 - \theta_1^{(t)})^{50-y_i}}{\pi^{(t-1)}(\theta_1^{(t)})^{y_i}(1 - \theta_1^{(t)})^{50-y_i} + (1 - \pi^{(t-1)})\theta_2^{y_i}(1 - \theta_2^{(t)})^{50-y_i}} \end{aligned}$$

Then update the value of z_i by drawing a Bernoulli outcome with probability of success equal to r_i.

Step 5: Update latent group probability π. - Gibbs step

$$\begin{aligned} p(\pi|z_i^{(t)}) &\propto p(\pi) \times \prod_{i=1}^{n} p(z_i^{(t)}|\pi) \\ &\propto \pi^{u-1}(1-\pi)^{v-1} \prod_{i=1}^{n} \pi^{z_i^{(t)}}(1-\pi)^{1-z_i^{(t)}} \\ &\propto \pi^{u+\sum_{i=1}^{n} z_i^{(t)}-1}(1-\pi)^{v+n-\sum_{i=1}^{n} z_i^{(t)}-1} \end{aligned}$$

So π is updated by a random draw from the distribution $Beta(u+\sum_{i=1}^{n} z_i^{(t)}, v+n-\sum_{i=1}^{n} z_i^{(t)})$

(1 mark) correct algorithm for update of $\boldsymbol{\theta}$
(0.5 mark) correct algorithm for update of a
(0.5 mark) correct algorithm for update of b
(1 mark) correct algorithm for update of z_i's
(1 mark) correct algorithm for update of π

(g) [1 mark] The hierarchical model allows for information sharing between the two latent groups. For both groups we are interested to measure the same thing (performance on the psychometric assessment) so it makes sense to pool information from both groups. Also, the latent group indicator is unknown and for some iterations, individual i may belong to group 1 and for other iterations they may belong to group 2, so they contribute observed data information to inform the posterior distribution of both θ_1 and θ_2, and the hierarchical structure also captures this information sharing idea. Also, if by chance, most people are identified as visual thinkers or non-thinkers, then the hierarchical approach can provide more reliable parameter estimates for θ_j for the group with smaller sample size.

(1 mark) for any valid explanation but must explain what information sharing between groups means for this question.