AUSTRALIAN NATIONAL UNIVERSITY
RESEARCH SCHOOL OF FINANCE ACTUARIAL STUDIES, AND APPLIED STATISTICS

INTRODUCTION TO BAYESIAN DATA ANALYSIS (STAT3016/4116/7016)
SEMESTER 2 2017

ASSIGNMENT 3 - SOLUTIONS

(see `assign3.R` for R code)

Problem 1 [STAT7016 ONLY]

$$U_i|\alpha,\mu,\tau^2,\nu,y \sim \text{InvGamma}\left(\frac{\nu+1}{2}, \frac{\nu\tau^2+((y_i-\mu)/\alpha)^2}{2}\right)$$

$$\mu|\alpha,\tau^2,U,\nu,y \sim N\left(\frac{\sum_{i=1}^n \frac{1}{\alpha^2 U_i} y_i}{\sum_{i=1}^n \frac{1}{\alpha^2 U_i}}, \frac{1}{\sum_{i=1}^n \frac{1}{\alpha^2 U_i}}\right)$$

$$\tau^2|\alpha,\mu,Y,\nu,y \sim \text{Gamma}\left(\frac{n\nu}{2}, \frac{\nu}{2}\sum_{i=1}^n \frac{1}{U_i}\right)$$

$$\alpha^2|\mu,\tau^2,U,\nu,y \sim \text{InvGamma}\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^n \frac{(y_i-\mu)^2}{U_i}\right)$$

Problem 2

(a)
```
p2.data<-read.table("interexp.dat",header=T)
p2.data
attach(p2.data)
thetaA<-mean(yA, na.rm=T)
thetaB<-mean(yB, na.rm=T)
varA<-var(yA,na.rm=T)
varB<-var(yB,na.rm=T)
rho<-cor(yA,yB,use="complete.obs")
```

The empirical estimates are

$\hat{\theta}_A$	$\hat{\theta}_B$	$\hat{\sigma}_A^2$	$\hat{\sigma}_B^2$	ρ
24.20	24.81	4.09	4.69	0.62

Table 1: Empirical estimates - interrupted stimulus experiment

Note: the missing data method described in the question to obtain the above empirical estimates is known as 'casewise' deletion, where cases that do not contain missing data for the variables(s) selected for analysis are included in the analysis.

(b)
```
> yB.hat<-yB
> yB.hat[is.na(yB)==TRUE]<-thetaB+(yA[is.na(yB)==TRUE]-thetaA)*rho*sqrt(varB/varA)
> yA.hat<-yA
> yA.hat[is.na(yA)==TRUE]<-thetaA+(yB[is.na(yA)==TRUE]-thetaB)*rho*sqrt(varA/varB)
>
> t.test(yA.hat,yB.hat,paired=TRUE)
t = -3.2807, df = 57, p-value = 0.00177
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9850730 -0.2383347
sample estimates:
mean of the differences
             -0.6117038
```

The p-value of the paired t-test is 0.00177, indicating a significant difference in mean reaction time between stimulus A and B. Specifically, stimulus A produces a faster mean reaction time.

(c)
```
#Jeffreys prior
p<-2

#Gibbs sampler for imputation
#prior parameters
n<-dim(p2.data)[1]
p<-dim(p2.data)[2]

Sigma<-diag(1,2,2)

#Starting values
Sigma<-S0
Y.full<-p2.data
O<-1*(!is.na(p2.data))

for (j in 1:p){
 Y.full[is.na(Y.full[,j]),j]<-mean(Y.full[,j],na.rm=TRUE)
}

#Gibbs sampler
THETA<-SIGMA<-Y.MISS<-NULL
set.seed(1)
for (s in 1:1000){
 ##update theta
  mun<-apply(Y.full,2,mean)
  Ln<-Sigma/n
  theta<-rmvnorm(1,mun,Ln)
  ##update Sigma
  Sn<-(t(Y.full)-c(theta))%*%t(t(Y.full)-c(theta))
  Sigma<-solve(rwish(solve(Sn),n))
  ##update missing data
  for (i in 27:n){
   b<-(O[i,]==0)
   a<-(O[i,]==1)
   iSa<-solve(Sigma[a,a])
   beta.j<-Sigma[b,a]%*%iSa
   Sigma.j<-Sigma[b,b]-Sigma[b,a]%*%iSa%*%Sigma[a,b]
   theta.j<-theta[b]+beta.j%*%(t(Y.full[i,a])-theta[a])
   Y.full[i,b]<-rmvnorm(1,theta.j,Sigma.j)
  }
```

```
   ##save results
   THETA<-rbind(THETA,theta)
   SIGMA<-rbind(SIGMA,c(Sigma))
   Y.miss<-rbind(Y.MISS,Y.full[O==0])
 }
> mean(THETA[,1]-THETA[,2])
[1] -0.623123
>  quantile(THETA[,1]-THETA[,2],c(0.025,0.975))
      2.5%       97.5%
-1.27869140  0.04522384
>  mean(THETA[,1]>THETA[,2])
[1] 0.035
```

A similar estimate is obtained for the mean difference $\theta_A - \theta_B$ but the 95% posterior confidence interval contains zero. The posterior probability that $\theta_A > \theta_B$ is 0.035. We conclude that based on the Bayesian model, there is only weak evidence of a significant difference in mean reaction times between stimuli.

Discussion: The paired t-test in part (b) fails to take into account the uncertainty in the parameter estimates due to incomplete data. Therefore, the variance of the mean difference is underestimated, and hence, we over state the significance of the difference in mean reaction times between stimulus A and stimulus B. The 95% confidence interval based on the paired t-test is deemed permissive (that is, anti-conservative).

(d) The estimate $\hat{y}_{i,B} = \hat{\theta}_B$ assumes there is no relationship between reaction times from stimulus A and reaction times from stimulus B. However, since we know that each subject is given both stimulus A and stimulus B, we know the reaction times are correlated. Hence, $\hat{y}_{i,B} = \hat{\theta}_B + (y_{i,A} - \hat{\theta}_A)\hat{\rho}\sqrt{\hat{\sigma}_B^2/\hat{\sigma}_A^2}$ is a better estimate. But to properly account for the uncertainty in our parameter estimates due to the presence of missing values, we should implement the Gibbs sampler in part (c) to iteratively impute the missing values before performing our statistical inference to test for a significant stimulus effect.

Problem 3

(a)
```
p3.data<-read.table("azdiabetes.dat",header=T)
p3.data
attach(p3.data)

#Extract data for two groups
Lyes<-diabetes=="Yes"
Lno<-diabetes=="No"
diab_yes<-p3.data[Lyes,]
diab_no<-p3.data[Lno,]


##Yes-diabetes
Y<-diab_yes[,1:7]
n<-dim(Y)[1]
ybar<-apply(Y,2,mean)
Sigma<-cov(Y)
THETA_yes<-SIGMA_yes<-NULL

##Prior parameters
mu0<-ybar
L0<-S0<-Sigma
nu0<-9

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(solve(Sn),nu0+n))

##save results
THETA_yes<-rbind(THETA_yes,theta)
SIGMA_yes<-rbind(SIGMA_yes,c(Sigma))
}
```

```
##No-diabetes
Y<-diab_no[,1:7]
n<-dim(Y)[1]
ybar<-apply(Y,2,mean)
Sigma<-cov(Y)
THETA_no<-SIGMA_no<-NULL

##Prior parameters
mu0<-ybar
L0<-S0<-Sigma
nu0<-9

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(solve(Sn),nu0+n))

##save results
THETA_no<-rbind(THETA_no,theta)
SIGMA_no<-rbind(SIGMA_no,c(Sigma))
}
```

```
> pp<-NULL
> for (i in 1:7){
+  pp<-c(pp,mean(THETA_yes[,i]>THETA_no[,i]))
+
+ }
>  names(pp)<-names(Y)
> pp
npreg   glu    bp  skin   bmi   ped   age
    1     1     1     1     1     1     1
```

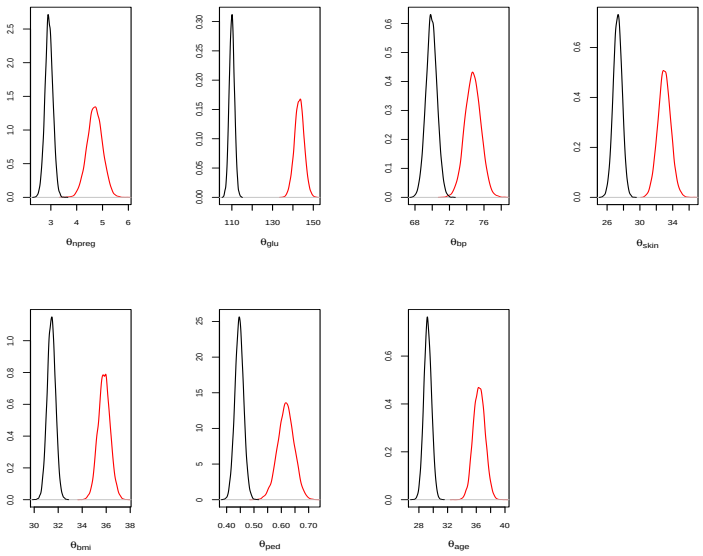

Figure 1: Comparison of marginal posterior distributions of $\theta_{d,j}$ and $\theta_{n,j}$ (diabetes==yes in red)

From Figure 1, we see that none of the marginal posterior distributions for the seven variables overlap between the diabetes group and the non-diabetes group. Specifically, the marginal posterior densities for the diabetes group lie to the right of the densities for the non-diabetes group and show greater spread, indicating that the average number of pregnancies, glucose level, blood pressure, skin fold thickness, body mass index, diabetes pedigree and age are higher for diabetic women than non-diabetic women. These results are confirmed by the calculation of the posterior probabilities $Pr(\theta_{d,j} > \theta_{n,j}|\mathbf{Y})$, which are 1 for all j.

(b)
```
Sigma_yes_mean<-apply(SIGMA_yes,2,mean)
Sigma_no_mean<-apply(SIGMA_no,2,mean)

pdf("Fig2.pdf")
plot(Sigma_yes_mean,Sigma_no_mean,xlab=expression(Sigma[yes]),
ylab=expression(Sigma[no]))
abline(0,1)
dev.off()
```

The posterior means for variances and covariances are roughly similar, except the posterior mean estimate of glucose variance is much larger for the diabetes group.

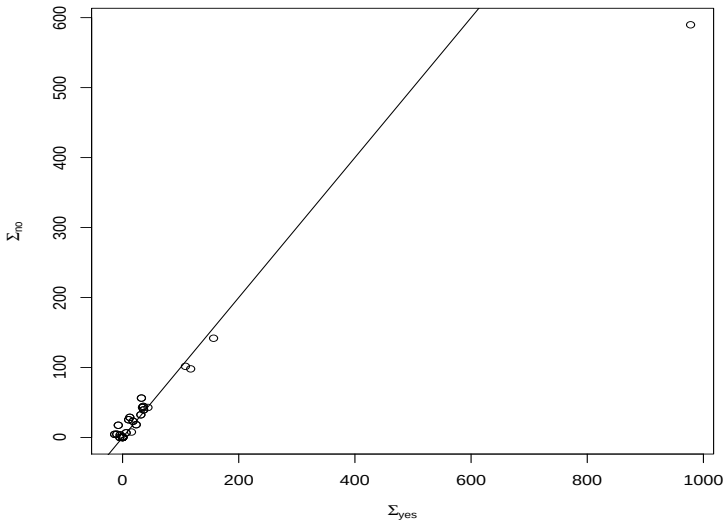

Figure 2: Comparison of posterior means of Σ_d and Σ_n (45 degreee line in black)

Problem 4

(a) The full conditional of β depends only on $\mathbf{z}$

$$
\begin{aligned}
p(\beta|\mathbf{y},\mathbf{x},\mathbf{z},c) &\propto p(\beta)p(\mathbf{z}|\beta,\mathbf{x}) \\
&\propto \exp\left\{-\frac{1}{2\tau_\beta^2}\beta^2\right\}\exp\left\{-\frac{1}{2\tau_\beta^2}\sum_{i=1}^n(z_i-x_i\beta)^2\right\} \\
&= \exp\left\{-\frac{1}{2\tau_\beta^2}\beta^2\right\}\exp\left\{-\frac{1}{2\tau_\beta^2}\sum_{i=1}^n(z_i^2-2\beta x_i z_i+x_i^2\beta)\right\} \\
&= \exp\left\{-\frac{1}{2}\left(\beta^2\left(\frac{1}{\tau_\beta^2}+\sum_{i=1}^n x_i^2\right)-2\beta\sum_{i=1}^n x_i z_i+\sum_{i=1}^n z_i^2\right)\right\} \\
&= \exp\left\{-\frac{A}{2}\left(\beta^2-\frac{2\beta\sum_{i=1}^n x_i z_i}{A}+\frac{\sum_{i=1}^n z_i^2}{A}\right)\right\} \\
&\propto \exp\left\{-\frac{A}{2}\left(\beta-\frac{\sum_{i=1}^n x_i z_i}{A}\right)^2\right\}
\end{aligned}
$$

(where $A=\left(\frac{1}{\tau_\beta^2}+\sum_{i=1}^n x_i^2\right)$. From the last line we see that $\beta|\mathbf{y},\mathbf{x},\mathbf{z},c) \sim \mathrm{N}\left(\frac{\sum_{i=1}^n x_i z_i}{A},\frac{1}{A}\right)$

(b)

- Given $\mathbf{z}$ and $\mathbf{y}$, we know that c must be between $a_k = \{\max(z_i) : y_i = 0\}$ and $b_k = \{\min(z_i) : y_i = 1\}$. Therefore the full conditional distribution of c is $N(0, \tau_c^2)$ but constrained on the interval $[a_k, b_k)$.
- $$z_i|\mathbf{y}, \mathbf{x}, \mathbf{z}_{-i}, \beta, c \sim \begin{cases} \mathrm{N}(\beta x_i, 1) \times I_{\{z_i > c\}} & \text{if } y_i = 1 \\ \mathrm{N}(\beta x_i, 1) \times I_{\{z_i \leq c\}} & \text{if } y_i = 0 \end{cases}$$

(c)
```
X<-p4.data[,1]
y<-p4.data[,2]
ranks<-match(y,sort(unique(y)))
uranks<-sort(unique(ranks))
n<-length(X)
p<-1

#starting values
beta<-rep(0,p)
z<-qnorm(rank(y,ties.method="random")/(n+1))
g<-rep(NA,1)
K<-length(uranks)
mu_c<-0
S<-10000

#prior parameters
tau2_beta<-tau2_c<-16

A<-1/tau2_beta+sum(X^2)
#matrices to store posterior draws in
BETA<-matrix(NA,S,p)
Z<-matrix(NA,S,n)
G<-matrix(NA,S,length(g))

#Gibbs sampler
for (s in 1:S){
#update g using its full conditional
    for (k in 1:K-1){
     a<-max(z[y==0])
     b<-min(z[y==1])
     u<-runif(1,pnorm((a-mu_c)/sqrt(tau2_c)),pnorm((b-mu_c)/sqrt(tau2_c)))
     g[k]<-mu_c+sqrt(tau2_c)*qnorm(u)
    }
```

```
    #update beta
    beta<-rnorm(1,mean=sum(X)/A,sqrt(1/A))

    #update z
    ez<-X%*%t(beta)
    a<-c(-Inf,g)[match(y,0:K)]
    b<-c(g,Inf)[match(y,0:K)]
    u<-runif(n,pnorm(a-ez),pnorm(b-ez))
    z<-ez+qnorm(u)

    #Store values
    BETA[s,]<-beta
    Z[s,]<-z
    G[s,]<-g

#Thinning by taking every 10th value in the chains
BETA<-BETA[10*(1:(S/10)),]
Z<-Z[10*(1:(S/10)),]
G<-G[10*(1:(S/10)),]

#Effective Sizes
>library(coda)
> apply(Z,2,effectiveSize)
 [1] 814.5956 1000.0000 1000.0000  906.8203 1000.0000 1000.0000
   1000.0000 1000.0000 1000.0000  911.9785 1000.0000 1000.0000
   820.3050  889.8450  896.7642  783.2733 1000.0000  945.5848 1000.0000
   947.7887 1266.9027 851.0850 1000.0000 1000.0000 1000.0000
> effectiveSize(BETA)
    var1
1237.202
> effectiveSize(G)
    var1
463.5543
}
```

The effective sizes are all reasonably large, except for the cut off point 'c' (labelled 'G' in the R code). We could run the Gibbs sampler longer such that all effective sample sizes are greater than 1000.

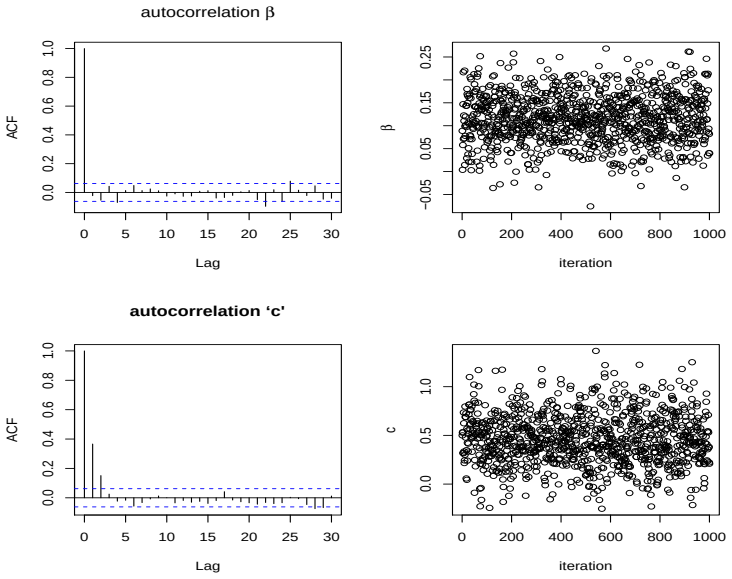

Figure 3: Autocorrelation plot for β and c

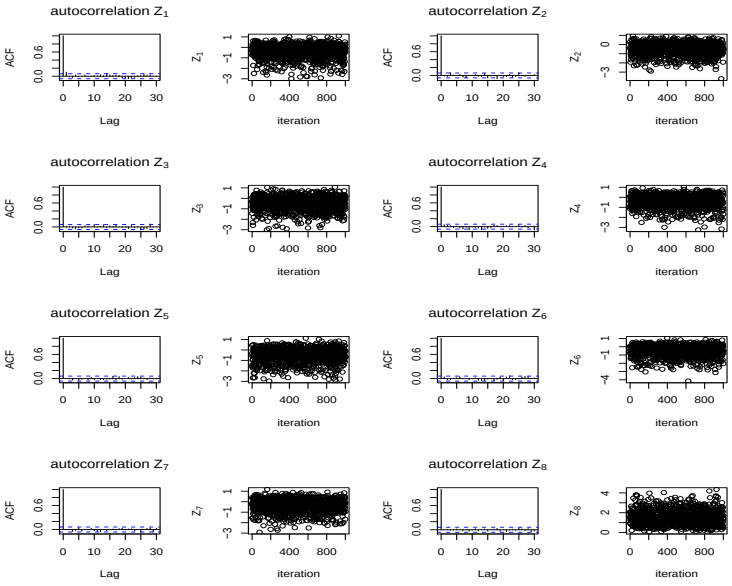

Figure 4: Autocorrelation plot for Z_1 ....Z_8

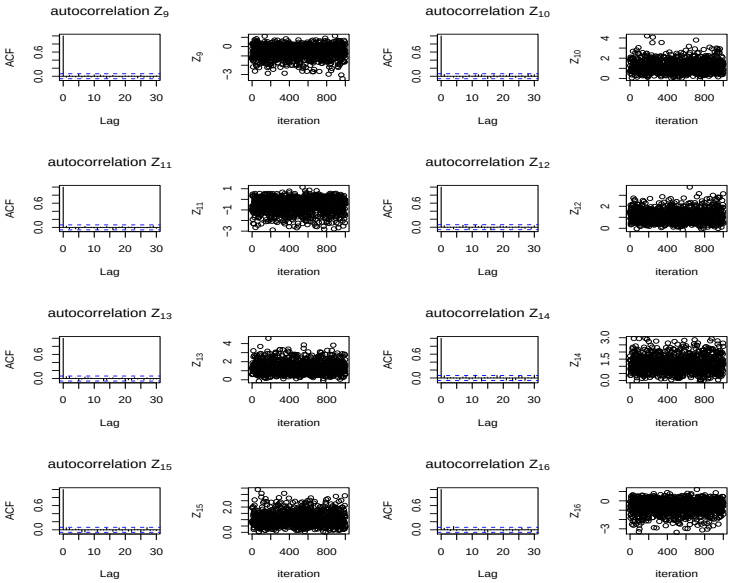

Figure 5: Autocorrelation plot for Z_9 ....Z_{16}

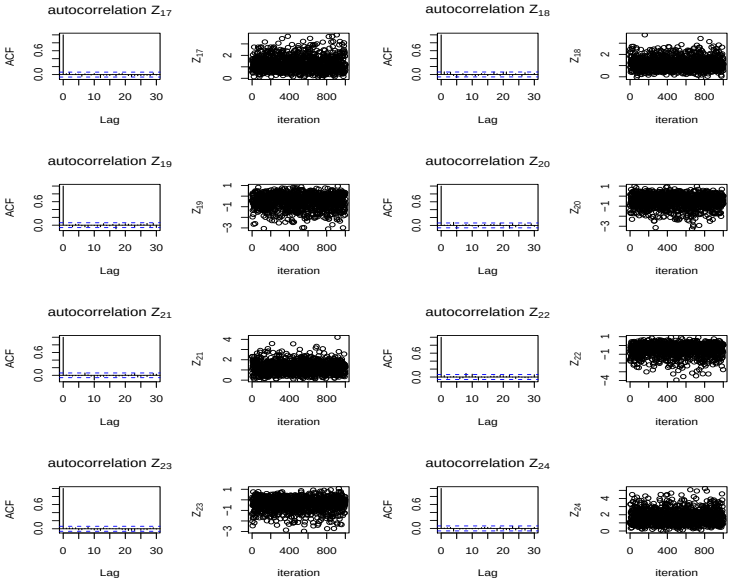

Figure 6: Autocorrelation plot for Z_{17} ....Z_{24}

The autocorrelation plots for β and $\mathbf{Z}$ show no correlation to be concerned about. The autocorrelation plot for c shows non negligible correation at lags 1 and 2 but the correlation quickly diminishes for lags thereafter. We could run the sequence for longer and discard a burn-in period or thin the sequence more to remove this autocorrelation.

We see from the traceplots that the speed of mixing is good for all parameters. That is, the chain moves around the respective parameter spaces quickly and does not get stuck in a particular region.

(d)
```
 > mean(BETA>0)
[1] 0.988
> quantile(BETA,c(0.025,0.975))
      2.5%      97.5%
0.01447776 0.21462669
```

A 95% posterior confidence interval for β is (0.014, 0.215) and $Pr(\beta > 0|\mathbf{y}, \mathbf{x}) = 0.988$, which is a strong indication that the relationship between the latent variable z and the covariate x is positive.

Problem 5

(a) $p(\theta|y) \propto \frac{1}{\sigma}\prod_{i=1}^{n}\phi\left(\frac{y_i^\lambda-1}{\lambda},\mu,\sigma\right)y_i^{\lambda-1}$ where $\phi(x,\mu,\sigma)$ is the normal density with mean $=\mu$ and standard deviation $=\sigma$ evaluated at x.

(b) The R function is:

```
logpost<-function(theta){
  sigma<-theta[3]
  mu<-theta[2]
  lambda<-theta[1]
  logf<- -log(sigma)+sum(log(dnorm(((y^lambda-1)/lambda),mu,sigma)))
                +sum((lambda-1)*log(y))
  return(logf)
}
```

(c) Using the `laplace` function, we obtain the following posterior mode estimate:

```
> fit<-laplace(logpost,start)
> fit$mode
[1] lambda=0.1025895; mu=3.0913611; sigma=1.4290467
```

(d) The full conditional distributions we need to draw from are:

$\underline{\mu}$:

$$\mu|\sigma,\lambda,y \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^\lambda-1}{\lambda}-\mu\right)^2\right) = \exp\left(-\frac{n}{2\sigma^2}\left(\mu-\sum_{i=1}^{n}\frac{y_i^\lambda-1}{n\lambda}\right)^2\right)$$

and so

$$\mu|\sigma,\lambda,y \sim N\left(\frac{1}{n\lambda}\sum_{i=1}^{n}(y_i^\lambda-1),\sigma^2/n\right)$$

$\underline{\sigma}$:

$$\begin{aligned}\sigma|\mu,\lambda,y &\propto (\sigma)^{-1}\sigma^{-n}exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^\lambda-1}{\lambda}-\mu\right)^2\right)\\ &= \sigma^{-n-1}exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^\lambda-1}{\lambda}-\mu\right)^2\right)\\ &\propto (\sigma^2)^{-1/2}(\sigma^2)^{\frac{-n-1}{2}}exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^\lambda-1}{\lambda}-\mu\right)^2\right)\\ &= (\sigma^2)^{-n/2-1}exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^\lambda-1}{\lambda}-\mu\right)^2\right)\end{aligned}$$

applying the change of variable from σ to σ^2 in line 3. We recognise the last line to be the probability density of an $InverseGamma(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i^{\lambda}-1}{\lambda}-\mu\right)^2$ distribution.

Setting up the Gibbs sampler in R, we will draw iteratively from the two conditional distributions derived above for the ten values of λ in the sequence (0.05,0.10,0.15,....,0.85,0.95). The figures below plot the 95% posterior intervals for μ and σ^2. The results do not appear to be sensitive to the value of λ. The 95% confidence interval for μ is approximately (2.3,3.0) and the 95% confidence interval for σ^2 is approximately (0.90, 2.0).

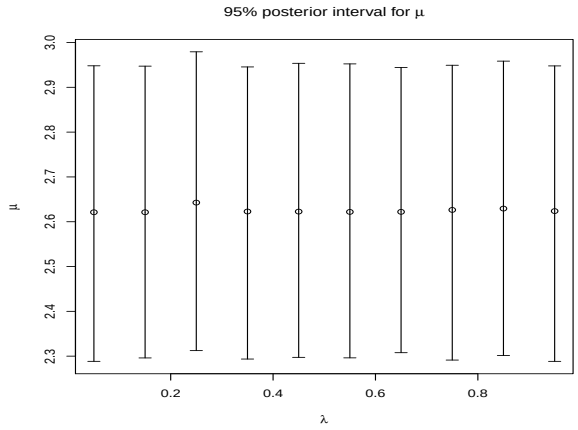

Figure 7: 95% posterior intervals for μ

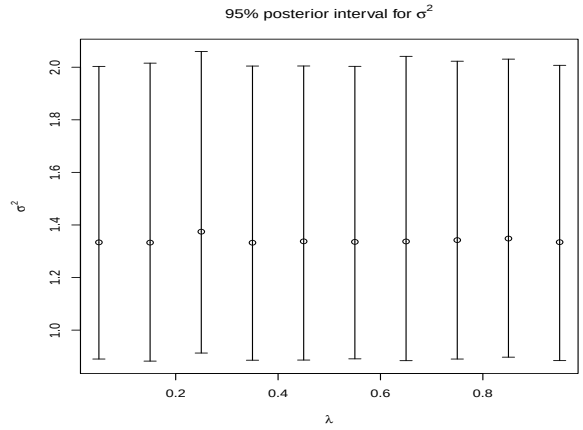

Figure 8: 95% posterior intervals for $\sigma 2$