For a Jeffrey's prior

$$\pi(\mu, \Sigma) \propto |\Sigma|^{-\frac{p+2}{2}}$$

, and a multivariate-normal likelihood, its conjugate posterior is

$$\pi(\Sigma|Y) \sim \mathrm{Inv}-\mathrm{Wishart}(n, S^{-1})$$

$$\pi(\mu|\Sigma, Y) \sim N(\bar{y}, \frac{\Sigma}{n})$$

1. Using only the complete data.

Set a prior of $(\mu1, \mu2) = (25,25), \Sigma = (\begin{matrix} 1 & 0.5 \\ 0.5 & 1 \end{matrix})$, using Monte Carlo simulation, we can get a posterior mean of ***-0.6230237,*** with 95% credible interval ***(-1.0288563, -0.2202139)***

2. Using mean to substitute the missing data.

Set a prior of $(\mu1, \mu2) = (25,25), \Sigma = (\begin{matrix} 1 & 0.5 \\ 0.5 & 1 \end{matrix})$, using Monte Carlo simulation, we can get a posterior mean of ***-0.6055888,*** with 95% credible interval***(-0.8830074 -0.3316835)***

3. Gibbs sampler approximating the missing data.

Set a prior of $(\mu1, \mu2) = (25,25), \Sigma = (\begin{matrix} 1 & 0.5 \\ 0.5 & 1 \end{matrix})$, using Gibbs sampling, we can get a posterior mean of ***-0.6066454,*** with 95% credible interval***(-0.8707112 -0.3422293)***

Compare the 3 posteriors, we can see that using the mean and Gibbs samplers give closer result than throwing away missing data. In fact, when we do the first method, we are throwing away many useful information and thus our posterior belief is "wider", compared to the other 2 method.

When we compare the last 2 methods, we know that setting all the missing data to the mean, is probably causing bias to the result as we ignored all the uncertainty.

Additionally, to check the convergence of μ and Σ , we look at their trace plots after burn-in. From the plots we see they converge well and the simulation is good.

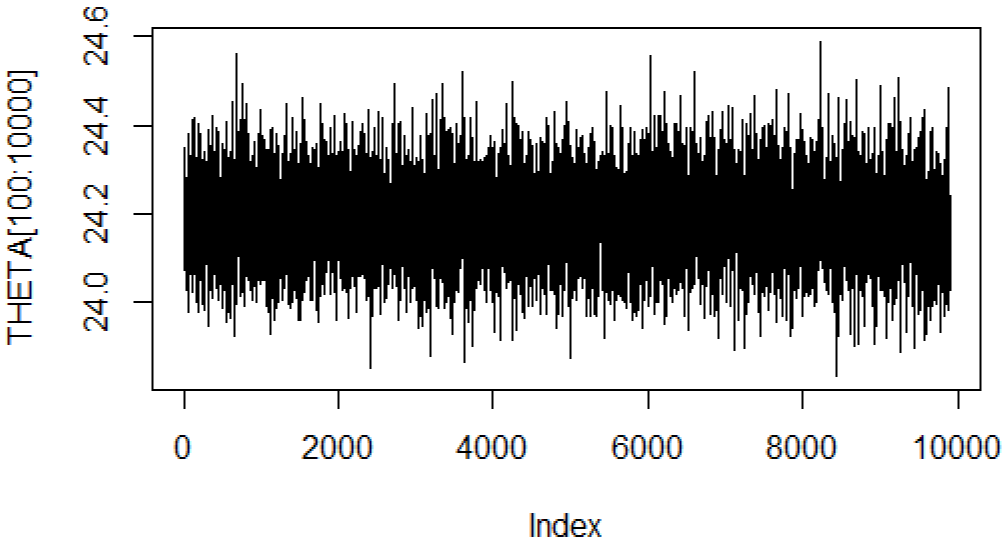

Figure 1: Trace plot of μ

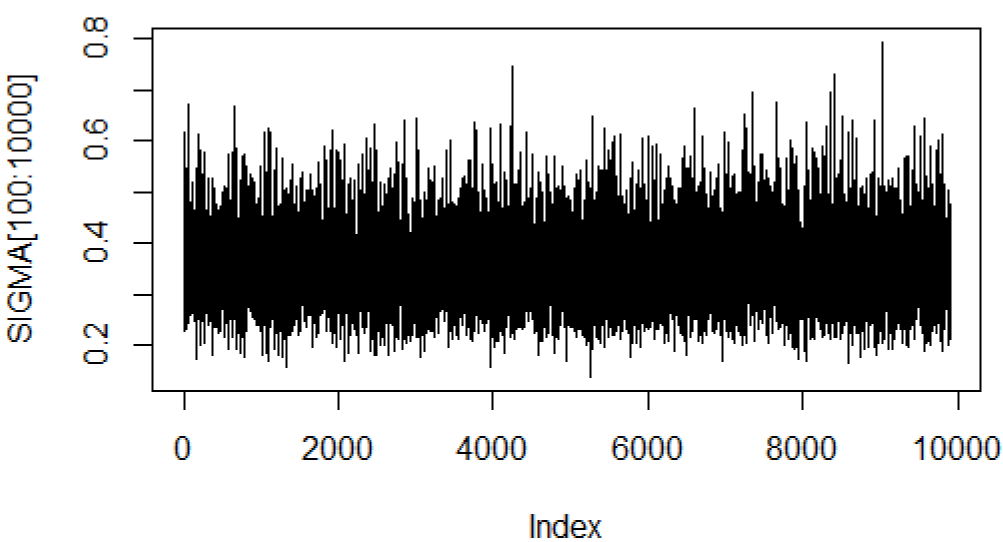

Figure 2: Trace plot of Σ

CODE

####1######

```
### Code from Hoff's website
### sample from the Wishart distribution
rwish<-function(n,nu0,S0)
{
  sS0 <- chol(S0)
  S<-array( dim=c( dim(S0),n ) )
  for(i in 1:n)
  {
    Z <- matrix(rnorm(nu0 * dim(S0)[1]), nu0, dim(S0)[1]) %*% sS0
    S[,,i]<- t(Z)%*%Z
  }
  S[,,1:n]
}
#########
library(mvtnorm)
Y <-read.table(header = TRUE,"http://www.stat.washington.edu/hoff/Book/Data/hwdata/interexp.dat")
data <- Y[1:26,]
n<−dim(data)[1] ; p<−dim(data)[2]
mu0 <- c(25,25)
Sigma0 <- matrix(1,0.5,0.5,1)

THETA<-SIGMA<-NULL
for (t in 1:10000)
{
  y.bar <- colMeans(data,1)
  S <- (t(data)-c(y.bar))%*%t(t(data)-c(y.bar))
  Sigma <- rwish(1,n,solve(S))
  Theta <- rmvnorm(1,y.bar,1/n*Sigma)

  THETA<-rbind(THETA,Theta)
  SIGMA<-rbind(SIGMA,c(Sigma))
}

result <- THETA[,1]-THETA[,2]
print(mean(result))
quantile(result,c(0.025,0.975))
```

######2########

```
n<−dim(Y)[1] ; p<−dim(Y)[2]
mu0 <- c(25,25)
Sigma0 <- matrix(1,0.5,0.5,1)
```

```
THETA<-SIGMA<-NULL
Y.full <−Y
O <−1*(!is.na(Y))
for(j in 1:p)
{
  Y.full[is.na(Y.full[,j]),j]<−mean(Y.full[,j],na.rm=TRUE)
}

for (t in 1:10000)
{
  y.bar <- colMeans(Y.full,1)
  S <- (t(Y.full)-c(y.bar))%*%t(t(Y.full)-c(y.bar))
  Sigma <- rwish(1,n,solve(S))
  Theta <- rmvnorm(1,y.bar,1/n*Sigma)

  THETA<-rbind(THETA,Theta)
  SIGMA<-rbind(SIGMA,c(Sigma))
}

result <- THETA[,1]-THETA[,2]
print(mean(result))
quantile(result,c(0.025,0.975))
```

######### 3 #########

```
library(mvtnorm)
Y <- read.table(header = TRUE,"http://www.stat.washington.edu/hoff/Book/Data/hwdata/interexp.dat")

### pri r parameters
n<−dim(Y)[1] ; p<−dim(Y) [2]
mu0<−c(25,25)
sd0<−(mu0/2)
L0<−matrix(.1,p,p) ; diag(L0)<−1 ; L0<− L0*outer(sd0,sd0)
nu0<−p+2 ; S0<−L0

###
### s t a r t i n g va lue s
Sigma<−S0
Y.full <−Y
O <−1*(!is.na(Y))
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 1:10000)
{
 ###update Sigma
 ybar<−apply (Y.full,2,mean)
 S <- (t(Y.full)-c(ybar))%*%t(t(Y.full)-c(ybar))
 Sigma <- rwish(1,n,solve(S))
 ###
 ###update the theta

 theta <- rmvnorm(1,ybar,1/n*Sigma)


 ###update missing data
 for (i in 27:n)
{
   b <− which( O[ i ,]==0 )
   a <− which( 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 r e s u l t s
 THETA<−rbind(THETA, theta) ; SIGMA<−rbind(SIGMA, c(Sigma))
 Y.MISS<−rbind(Y.MISS,Y.full[O==0])
 ###
}
plot(THETA[100:10000],type='l')
plot(SIGMA[100:10000],type='l')
result <- THETA[,1]-THETA[,2]
print(mean(result))
quantile(result,c(0.025,0.975))
```