

Homework 5

Bingying Liu

February 25, 2020

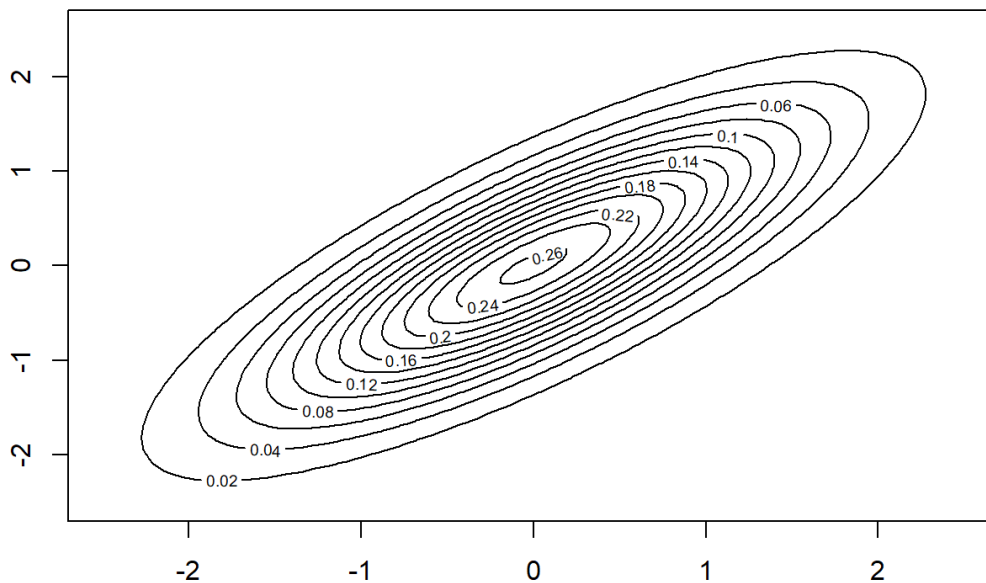
Question 1

Part(a): Estimate the MLE of (θ, Σ) . Make two contour plots, one using the true values of (θ, Σ) , and the other using the MLEs. Comment on the differences, if any.

```
## Simulate data
n = 100
set.seed(1234)
theta = c(0,0)
sigma = matrix(c(1,0.8,0.8,1),nrow=2,ncol=2)
Y = rmvnorm(n, mean = theta, sigma= sigma)

## true value
x.points <- seq(-2.5,2.5,length.out=100)
y.points <- x.points
z <- matrix(0,nrow=100,ncol=100)
for (i in 1:100) {
  for (j in 1:100) {
    z[i,j] <- dmvnorm(c(x.points[i],y.points[j]),
                      mean=theta,sigma=sigma)
  }
}
contour(x.points,y.points,z, main= expression(paste('Contour plot using true value of ',theta, ' and ',Sigma)))
```

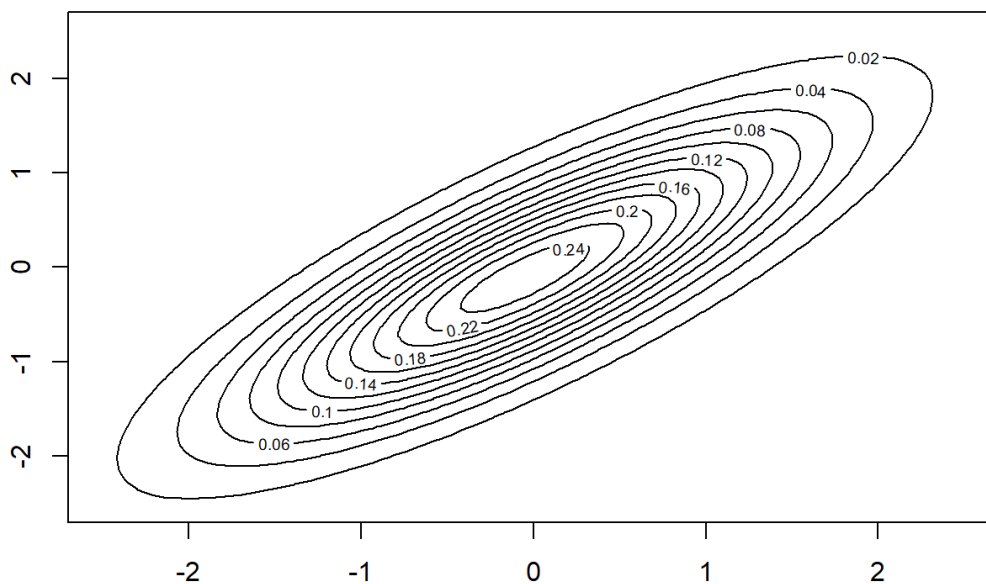
Contour plot using true value of θ and Σ



```
## MLE
theta_MLE = apply(Y,2,mean)
sigma_MLE = t(Y-theta_MLE)%*(Y-theta_MLE)/n

x.points <- seq(-2.5,2.5,length.out=100)
y.points <- x.points
z <- matrix(0,nrow=100,ncol=100)
for (i in 1:100) {
  for (j in 1:100) {
    z[i,j] <- dmvnorm(c(x.points[i],y.points[j]),
                      mean=theta_MLE,sigma=sigma_MLE)
  }
}
contour(x.points,y.points,z, main= expression(paste('Contour plot using MLE of ',theta, ' and ',Sigma)))
```

Contour plot using MLE of θ and Σ



There is not much difference between true value of (θ, Σ) and MLE of (θ, Σ) since sample size is big enough for MLE to approximate the true values.

Part(b): Assuming independent normal & inverse-Wishart priors for θ and Σ , that is, $\pi(\theta, \Sigma) = \pi(\theta)\pi(\Sigma)$, run Gibbs sampler (hyperparameters up to you but you must justify your choices) to generate posterior samples for (θ, Σ) .

Since we don't have prior knowledge about θ , I set μ_o to be $(0, 0)$ and set λ_0 to identity matrix multiplied by 1000 (which is a diffuse prior setting and we don't assume any correlation between the two variables). And since we also don't have prior guess for Σ , we set ν_0 to be $p + 2 = 4$ and S_0 to be identity matrix.

```

gibbs_normal = function(mu_0, Lambda_0, nu_0, S_0, n_iter, Y){
  n = nrow(Y)
  ybar = apply(Y,2,mean)

  # initial values for Gibbs sampler
  Sigma <- cov(Y)
  # set null matrices to save samples
  THETA <- SIGMA <- NULL

  # first set number of iterations and burn-in, then set seed
  burn_in <- 0.3*n_iter
  set.seed(1234)

  for (s in 1:(n_iter+burn_in)){
    # update theta using its full conditional
    Lambda_n <- solve(solve(Lambda_0) + n*solve(Sigma))
    mu_n <- Lambda_n %*% (solve(Lambda_0)%*%mu_0 + n*solve(Sigma)%*%ybar)
    theta <- rmvnorm(1,mu_n,Lambda_n)
    # update Sigma
    S_theta <- (t(Y)-c(theta))%*%t(t(Y)-c(theta))
    S_n <- S_0 + S_theta
    nu_n <- nu_0 + n
    Sigma <- riwish(nu_n, S_n)
    # save results only past burn-in
    if(s > burn_in){
      THETA <- rbind(THETA,theta)
      SIGMA <- rbind(SIGMA,c(Sigma))
    }
  }
  colnames(THETA) <- c("theta_1","theta_2")
  colnames(SIGMA) <- c("sigma_11","sigma_12","sigma_21","sigma_22") #symmetry in sigma

  return(list(theta = THETA, sigma = SIGMA))
}

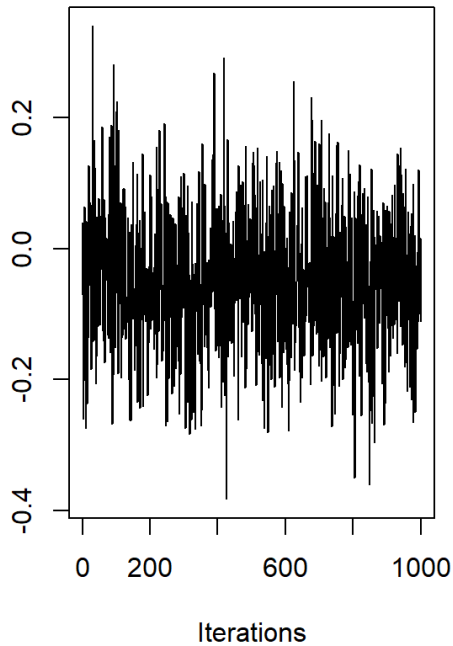
# Initialize the hyperparameters
ybar = apply(Y,2,mean)
mu_0 = c(0,0)
Lambda_0 = matrix(c(1,0.5,0.5,1),nrow=2, ncol=2)
nu_0 = 4
S_0 = matrix(c(1,0,0,1),nrow=2, ncol=2)

# Performs Gibbs Sampling
normal_mcmc = gibbs_normal(mu_0, Lambda_0, nu_0, S_0, n_iter=1000, Y=Y)

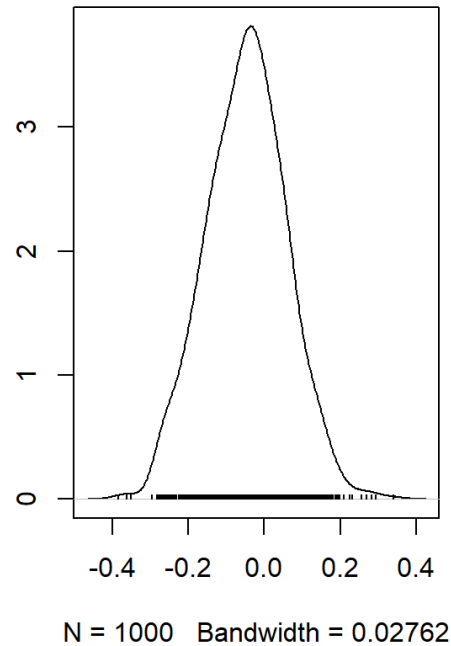
# Diagnostics check (autocorrelation plot seems reasonable)
THETA.mcmc <- mcmc(normal_mcmc$theta,start=1);
plot(THETA.mcmc[, "theta_1"])

```

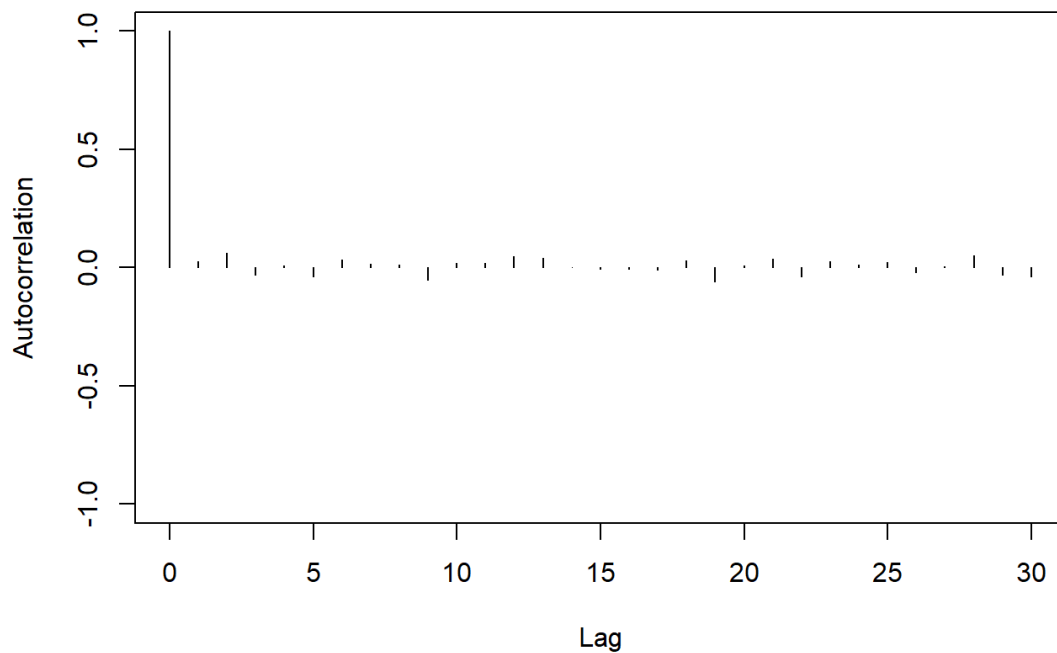
Trace of var1



Density of var1



```
autocorr.plot(THETA.mcmc[, "theta_1"])
```



Part(c): Compare Bayes estimates of (θ, Σ) to MLE and truth. Comment on the similarities and differences. How much influence do you think your prior had on the Bayes estimates in particular?

```
# THETA
apply(normal_mcmc$theta, 2, mean); # Bayes estimate
```

```
##      theta_1      theta_2
## -0.04776418 -0.09744347
```

```
theta_MLE; # MLE
```

```
## [1] -0.05010387 -0.10488054
```

```
theta # truth
```

```
## [1] 0 0
```

```
# SIGMA
apply(normal_mcmc$sigma, 2, mean) # Bayes estimate
```

```
## sigma_11 sigma_12 sigma_21 sigma_22
## 1.0982339 0.8893127 0.8893127 1.0896247
```

```
sigma_MLE # MLE
```

```
##           [,1]      [,2]
## [1,] 1.0978773 0.8933898
## [2,] 0.8933898 1.0792508
```

```
sigma # truth
```

```
##           [,1] [,2]
## [1,] 1.0 0.8
## [2,] 0.8 1.0
```

Since MLE is very close the true value of (θ, Σ) and also since we're using a diffuse/non-informative prior, prior is drown out by data. In this case, I calculated the expectation of Bayes estimate of θ and Σ . Both of the expectations are very close to MLE values, which shows that it's data that dominates the expected value of Bayes estimates. My prior had little influence on Bayes estimate.

Part(d): Given y_{i2} values for 50 new (test) subjects, describe in as much detail as possible, how you would use the Gibbs sampler to predict new y_{i1} values, given the y_{i2} values, from the “conditional posterior predictive distribution” of $(y_{i1}|y_{i2})$.

For each iteration s:

- We use Gibbs sampler to sample one θ and one Σ from posterior distribution $p(\theta, \Sigma|y_{1:100})$.
- We derive $p(y_{i1}^{101:150}|y_{i2}^{101:150}, \theta, \Sigma)$ by putting the sampled θ and Σ into conditional distribution $Y_1|(Y_2 = y_2) \sim N(\theta_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \theta_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Repeat the process for n_iter.

Question 2: Hoff 7.4

a) Before you look at the data, use your own knowledge to formulate a semiconjugate prior distribution for $\theta = (\theta_h, \theta_w)^T$ and Σ , where θ_h , θ_w are mean husband and wife ages, and Σ is the covariance matrix.

Considering the life expectancy and marriage age, I think most ages will fall between $(25, 85)$. The average age of women might be higher than men, but I don't have scientific proof to support it. It seems intuitive that there are less married couples at age 25 and 85 than there are married couples around 55 ($\frac{25+85}{2}$). So the bell curve centers around 55 with variance $(\frac{55-25}{2})^2 = 225$ such that approximately 95% of my prior falls within range $(25, 85)$. I also believe the age of married couples are close to each other, with a

correlation if 0.8, which results in the covariance to be $225 * 0.8 = 180$. Thus the prior covariance matrix for θ is $\Lambda_0 = \begin{bmatrix} 225 & 180 \\ 180 & 225 \end{bmatrix}$.

For variance Σ , the same logic of the ranges of ages still applies here. I'll set $S_0 = \Sigma_0$, but make Σ only loosely centered around Σ_0 by setting $\nu_0 = p + 2 = 4$.

b) Generate a prior predictive dataset of size $n = 100$, by sampling (θ, Σ) from your prior distribution and then simulating $Y_1, \dots, Y_n \sim \text{i.i.d multivariate normal } (\theta, \Sigma)$. Generate several such datasets, make bivariate scatterplots for each dataset, and make sure they roughly represent your prior beliefs about what such a dataset would actually look like.

```
n = 100
m = 15 # number of datasets

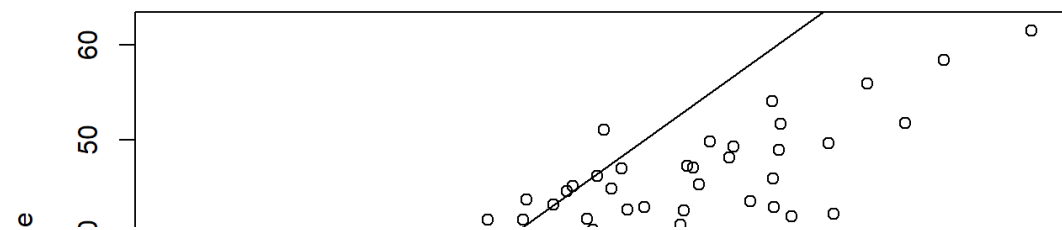
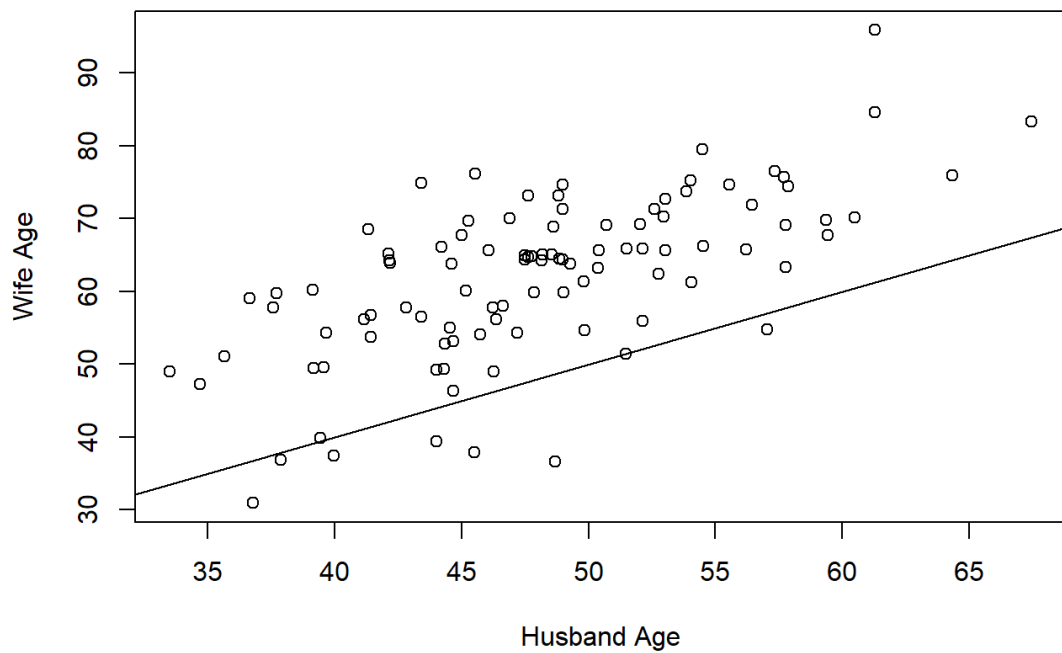
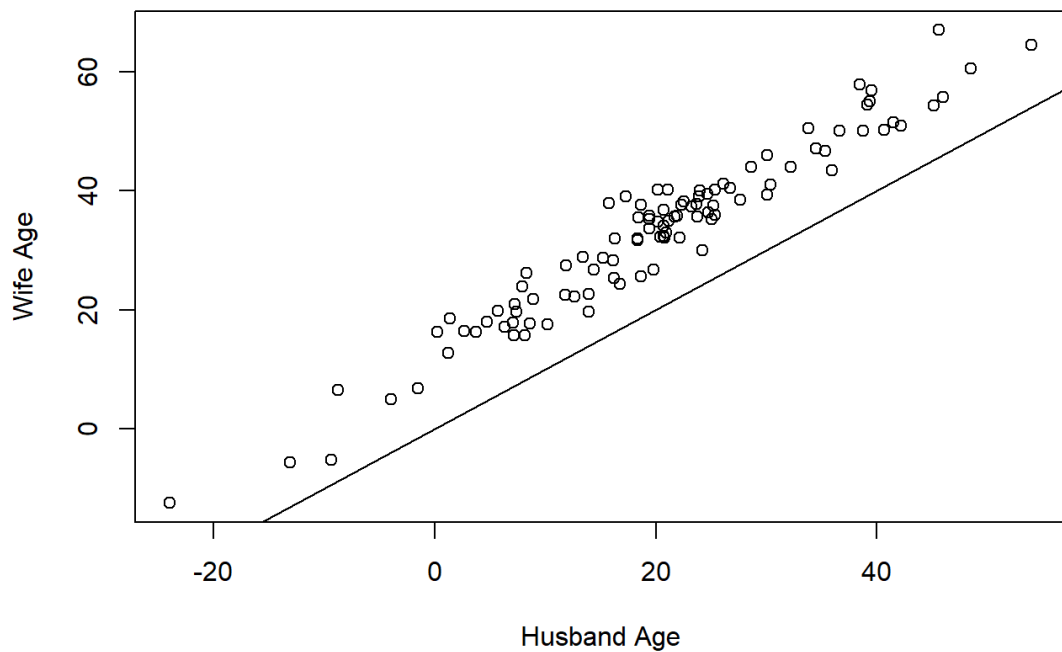
mu_0 = c(55,55)
Lambda_0 = matrix(c(225,180,180,225),nrow=2, ncol=2)
nu_0 = 4
S_0 = Lambda_0

# Set null matrices to save samples
THETA <- SIGMA <- Y.prior <- NULL

for (i in 1:m){
  # sample
  theta = rmvnorm(1, mu_0, Lambda_0)
  Sigma = riwish(nu_0, S_0)
  Y = rmvnorm(n, theta, Sigma)

  # generate
  THETA <- rbind(THETA,theta)
  SIGMA <- rbind(SIGMA,c(Sigma))
  Y.prior <- rbind(Y.prior,c(Y))
}

# Plotting, visual check
ids = sample(1:m, 3)
for (id in ids){
  plot(Y.prior[id, 1:100], Y.prior[id, 101:200], xlab = 'Husband Age', ylab = 'Wife Age')
  abline(a=0, b=1)
}
```





The generated prior predictive datasets do represent my prior beliefs on the real dataset.

c) Using your prior distribution and the 100 values in the dataset, obtain an MCMC approximation to $p(\theta, \Sigma | y_1, \dots, y_{100})$. Plot the joint posterior distribution of θ_h and θ_w , and also the marginal posterior density of the correlation between Y_h and Y_w , the ages of a husband and wife. Obtain 95% posterior confidence intervals for θ_h and θ_w and the correlation coefficient.

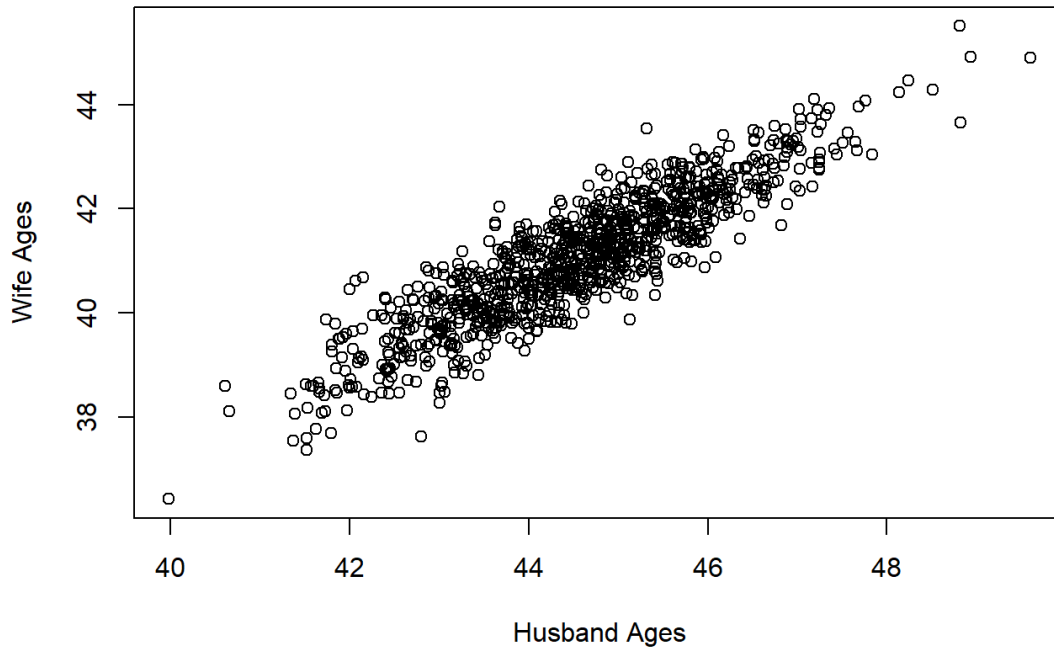
```
# Read in data
age = read.table('agehw.dat', header = TRUE)

# Hyperparameters for the priors
mu_0 = c(55,55)
Lambda_0 = matrix(c(225,180,180,225),nrow=2, ncol=2)
nu_0 = 4
S_0 = Lambda_0

# Gibbs Sampler
normal_age_mcmc = gibbs_normal(mu_0, Lambda_0, nu_0, S_0, n_iter=1000, Y=age)

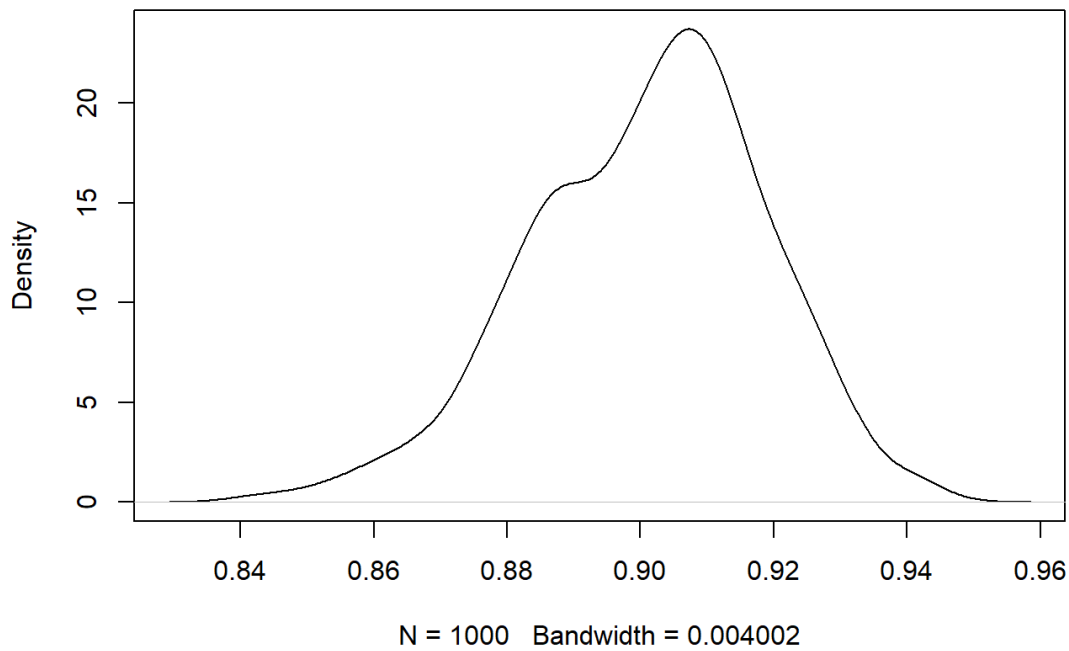
# Plots
plot(normal_age_mcmc$theta["theta_1"], normal_age_mcmc$theta["theta_2"], xlab="Husband Ages", ylab="Wife Ages", main = "Joint posterior distribution of THETA")
```


Joint posterior distribution of THETA



```
CORR = normal_age_mcmc$sigma[,2]/sqrt((normal_age_mcmc$sigma[, 1]* normal_age_mcmc$sigma[, 4]))  
plot(density(CORR), main = "Posterior Density of Correlation Between Yh and Yw")
```

Posterior Density of Correlation Between Yh and Yw



```
# 95% posterior confidence intervals for THETA and correlation coefficient  
quantile(normal_age_mcmc$theta[, "theta_1"], probs = c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 41.85173 47.22440
```

```
quantile(normal_age_mcmc$theta[, "theta_2"], probs = c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 38.50918 43.41570
```

```
quantile(CORR, probs = c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 0.8634539 0.9318348
```

d) Obtain 95% posterior confidence intervals for θ_h and θ_w and the correlation coefficient using the following prior distributions

i. Jeffrey's prior

```
# Initialize
n = nrow(age)
ybar = apply(age, 2, mean)
Sigma = cov(age)
THETA <- SIGMA <- NULL

# Gibbs Sampler
n_iter <- 1000; burn_in <- 0.3*n_iter
set.seed(1234)

for (s in 1:(n_iter+burn_in)){
  # update theta using its full conditional
  theta <- rmvnorm(1, ybar, Sigma/n)
  # update Sigma
  S_theta <- (t(age)-c(ybar))%*%t(t(age)-c(ybar))
  Sigma <- riwish(n+1, S_theta)
  # save results only past burn-in
  if(s > burn_in){
    THETA <- rbind(THETA, theta)
    SIGMA <- rbind(SIGMA, c(Sigma))
  }
}
colnames(THETA) <- c("theta_1", "theta_2")
colnames(SIGMA) <- c("sigma_11", "sigma_12", "sigma_21", "sigma_22") #symmetry in sigma

# Correlation coefficient
CORR = SIGMA[,2]/sqrt((SIGMA[, 1]* SIGMA[, 4]))

# 95% posterior confidence interval
quantile(THETA[, "theta_1"], probs = c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 41.73296 47.14183
```

```
quantile(THETA[, "theta_2"], probs = c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 38.37914 43.34355
```

```
quantile(CORR, probs = c(0.025,0.975))
```

```
##      2.5%      97.5%  
## 0.8640805 0.9327544
```

iii. Diffuse prior with μ_0

```
# Hyperparameters for prior  
mu_0 = c(0,0)  
Lambda_0 = 10^5 * diag(2)  
nu_0 = 3  
S_0 = 10^3 * diag(2)  
  
# Gibbs sampler  
normal_age_mcmc2 = gibbs_normal(mu_0, Lambda_0, nu_0, S_0, n_iter=1000, Y=age)  
  
# Correlation coefficient  
CORR = normal_age_mcmc2$sigma[,2]/sqrt((normal_age_mcmc2$sigma[, 1]* normal_age_mcmc2$sigma[, 4]))  
  
# 95% posterior confidence interval  
quantile(normal_age_mcmc2$theta[, "theta_1"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%  
## 41.63187 47.18515
```

```
quantile(normal_age_mcmc2$theta[, "theta_2"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%  
## 38.37408 43.38290
```

```
quantile(CORR, probs = c(0.025,0.975)) # the correlation is even higher
```

```
##      2.5%      97.5%  
## 0.7973876 0.8982505
```

e) Compare the confidence intervals from d) to those obtained in c). Discuss whether or not you think that your prior info is helpful in estimating θ and Σ , or if you think one of the alternatives in d) is preferable. What about if the sample size were much smaller, say $n=25$?

I think my prior is helpful in estimating θ and Σ , but the choice of different priors doesn't matter much in this case. Sample size of n is 100, which is pretty large compared to the amount of information in any of those priors. The prior sample size $\nu_0 \leq 4$ for any of the priors above, which is much less informative than the data itself. Thus prior effects are drawn and the results above are pretty much similar.

If we have a smaller sample size, this may be different. The priors would be more informative compared to the data. This might have a greater effect on the posterior. I randomly choose 25 samples from the original data and test on my prior and the diffuse prior.

```
age25 = age[sample(1:100, 25),]

## my prior
mu_0 = c(55,55)
Lambda_0 = matrix(c(225,180,180,225),nrow=2, ncol=2)
nu_0 = 4
S_0 = Lambda_0

normal_age25_mcmc = gibbs_normal(mu_0, Lambda_0, nu_0, S_0, n_iter=1000, Y=age25)

# post analysis
CORR = normal_age25_mcmc$sigma[,2]/sqrt((normal_age25_mcmc$sigma[, 1]* normal_age25_mcmc$sigma[, 4]))
quantile(normal_age25_mcmc$theta[, "theta_1"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 35.44038 45.62840
```

```
quantile(normal_age25_mcmc$theta[, "theta_2"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 32.20460 42.01807
```

```
quantile(CORR, probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 0.8339850 0.9588367
```

```
## diffuse prior
mu_0 = c(0,0)
Lambda_0 = 10^5 * diag(2)
nu_0 = 3
S_0 = 10^3 * diag(2)

normal_age25_mcmc = gibbs_normal(mu_0, Lambda_0, nu_0, S_0, n_iter=1000, Y=age25)

# post analysis
CORR = normal_age25_mcmc$sigma[,2]/sqrt((normal_age25_mcmc$sigma[, 1]* normal_age25_mcmc$sigma[, 4]))
quantile(normal_age25_mcmc$theta[, "theta_1"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 34.16441 45.95725
```

```
quantile(normal_age25_mcmc$theta[, "theta_2"], probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 30.83500 41.88588
```

```
quantile(CORR, probs = c(0.025,0.975))
```

```
##      2.5%      97.5%
## 0.5076869 0.8662577
```

In this case, the effect of prior on correlation is greater when sample size is smaller, especially when using the diffuse prior. Since the diffuse prior assumes that the age of husband has zero correlation to the age of wife, which is equivalent of the two being independent. Thus, the posterior correlation is being dragged towards 0 with a greater effect when sample size is small (n=25).