

STA 360/602L: MODULE 4.4

MULTIVARIATE NORMAL MODEL IV

DR. OLANREWAJU MICHAEL AKANDE

READING EXAMPLE: POSTERIOR COMPUTATION

- Recall that we have

$$\pi(\boldsymbol{\theta}|\Sigma, \mathbf{Y}) = \mathcal{N}_2(\boldsymbol{\mu}_n, \Lambda_n)$$

where

$$\Lambda_n = [\Lambda_0^{-1} + n\Sigma^{-1}]^{-1}$$

$$\boldsymbol{\mu}_n = \Lambda_n [\Lambda_0^{-1}\boldsymbol{\mu}_0 + n\Sigma^{-1}\bar{\mathbf{y}}],$$

- For our reading example,

$$\boldsymbol{\mu}_0 = (\mu_{0(1)}, \mu_{0(2)})^T = (50, 50)^T$$

$$\Lambda_0 = \begin{pmatrix} 156 & 78 \\ 78 & 156 \end{pmatrix}$$

READING EXAMPLE: POSTERIOR COMPUTATION

- We also have

$$\pi(\Sigma|\boldsymbol{\theta}, \mathbf{Y}) = \mathcal{IW}_2(\nu_n, \mathbf{S}_n)$$

or using the notation in the book, $\mathcal{IW}_2(\nu_n, \mathbf{S}_n^{-1})$, where

$$\nu_n = \nu_0 + n$$

$$\begin{aligned}\mathbf{S}_n &= [\mathbf{S}_0 + \mathbf{S}_\theta] \\ &= \left[\mathbf{S}_0 + \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^T \right].\end{aligned}$$

- Again, for our reading example,

$$\nu_0 = p + 2 = 4$$

$$\Sigma_0 = \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix}$$

POSTERIOR COMPUTATION

```
#Data summaries
n <- nrow(Y)
ybar <- apply(Y,2,mean)

#Hyperparameters for the priors
mu_0 <- c(50,50)
Lambda_0 <- matrix(c(156,78,78,156),nrow=2,ncol=2)
nu_0 <- 4
S_0 <- matrix(c(625,312.5,312.5,625),nrow=2,ncol=2)

#Initial values for Gibbs sampler
#No need to set initial value for theta, we can simply sample it first
Sigma <- cov(Y)

#Set null matrices to save samples
THETA <- SIGMA <- NULL
```

Next, the code for the Gibbs sampler.

POSTERIOR COMPUTATION

```
#Now, to the Gibbs sampler
#library(mvtnorm) for multivariate normal
#library(MCMCpack) for inverse-Wishart

#first set number of iterations and burn-in, then set seed
n_iter <- 10000; 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
```

Note that the text also has a function to sample from the Wishart distribution.



DIAGNOSTICS

```
#library(coda)  
THETA.mcmc <- mcmc(THETA,start=1); summary(THETA.mcmc)
```

```
##  
## Iterations = 1:10000  
## Thinning interval = 1  
## Number of chains = 1  
## Sample size per chain = 10000  
##  
## 1. Empirical mean and standard deviation for each variable,  
##    plus standard error of the mean:  
##  
##           Mean      SD Naive SE Time-series SE  
## theta_1 47.30 2.956 0.02956      0.02956  
## theta_2 53.69 3.290 0.03290      0.03290  
##  
## 2. Quantiles for each variable:  
##  
##           2.5%   25%   50%   75% 97.5%  
## theta_1 41.55 45.35 47.36 49.23 53.08  
## theta_2 47.08 51.53 53.69 55.82 60.13
```

```
effectiveSize(THETA.mcmc)
```

```
## theta_1 theta_2  
## 10000 10000
```



DIAGNOSTICS

```
SIGMA.mcmc <- mcmc(SIGMA,start=1); summary(SIGMA.mcmc)
```

```
##
## Iterations = 1:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 10000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##           Mean      SD Naive SE Time-series SE
## sigma_11 202.3 63.39   0.6339      0.6511
## sigma_12 155.3 60.92   0.6092      0.6244
## sigma_21 155.3 60.92   0.6092      0.6244
## sigma_22 260.1 81.96   0.8196      0.8352
##
## 2. Quantiles for each variable:
##
##           2.5%   25%   50%   75% 97.5%
## sigma_11 113.50 158.2 190.8 234.8 357.3
## sigma_12  67.27 113.2 144.7 186.5 305.4
## sigma_21  67.27 113.2 144.7 186.5 305.4
## sigma_22 145.84 203.2 244.6 300.9 461.0
```

```
effectiveSize(SIGMA.mcmc)
```

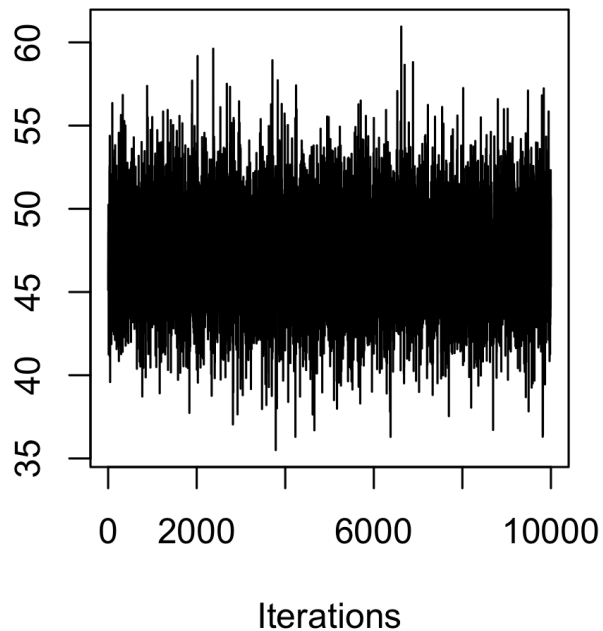
```
## sigma_11 sigma_12 sigma_21 sigma_22
## 9478.710 9517.989 9517.989 9629.352
```



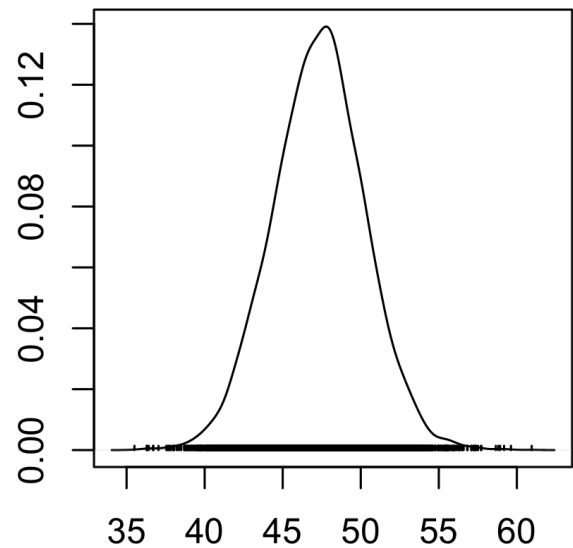
DIAGNOSTICS: TRACE PLOTS

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

Trace of var1



Density of var1



N = 10000 Bandwidth = 0.4857

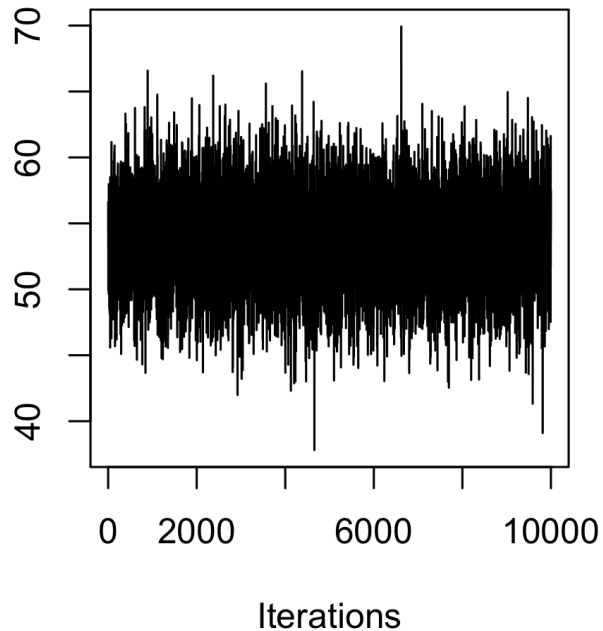
Looks good!



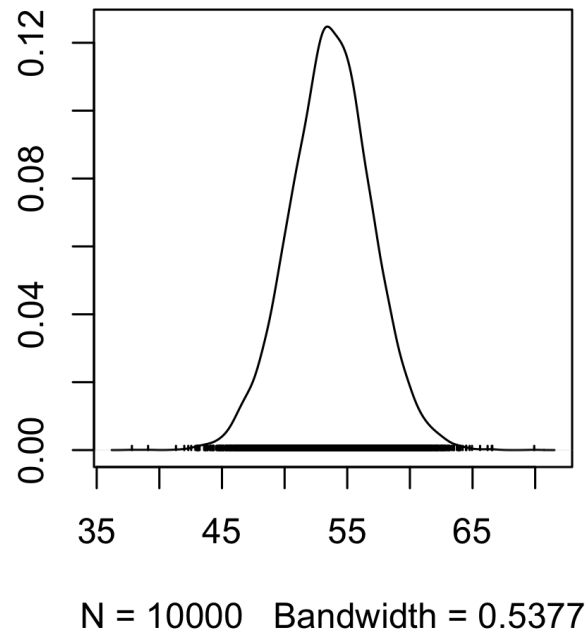
DIAGNOSTICS: TRACE PLOTS

```
plot(THETA.mcmc[, "theta_2"])
```

Trace of var1



Density of var1



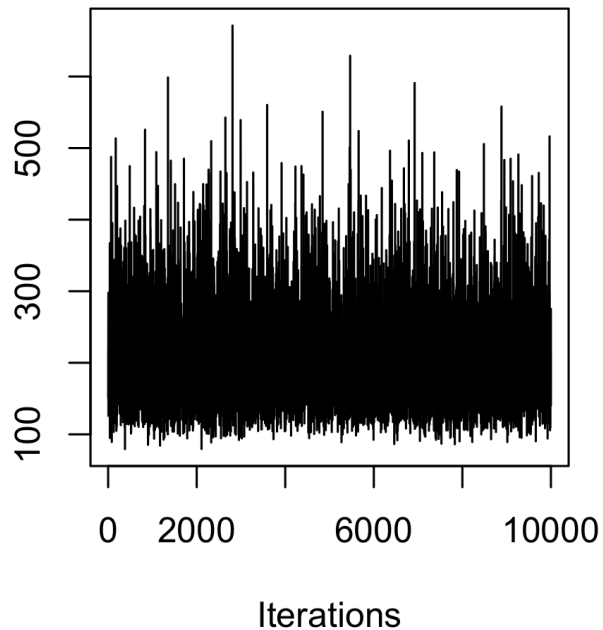
Looks good!



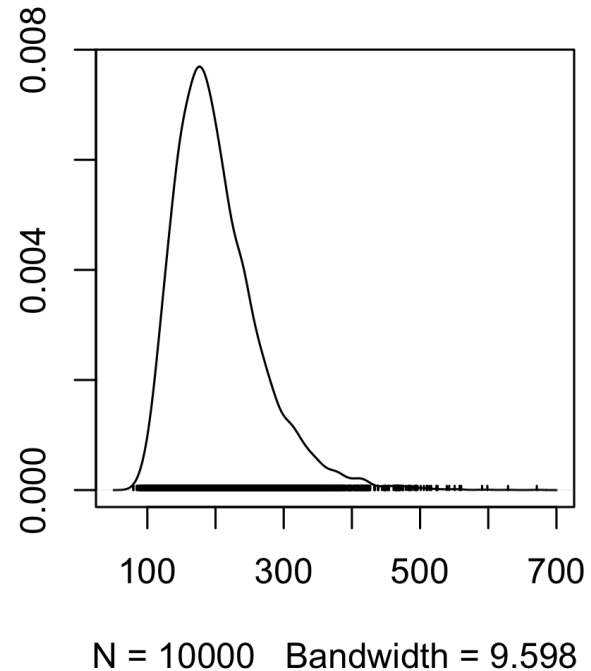
DIAGNOSTICS: TRACE PLOTS

```
plot(SIGMA.mcmc[, "sigma_11"])
```

Trace of var1



Density of var1



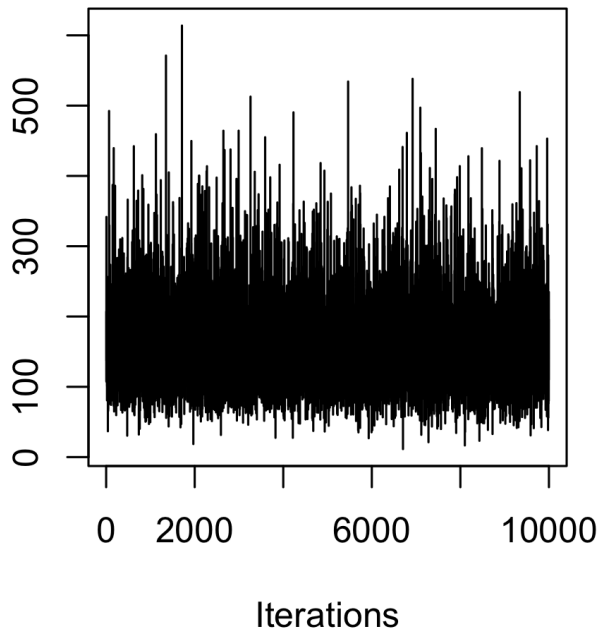
Looks good!



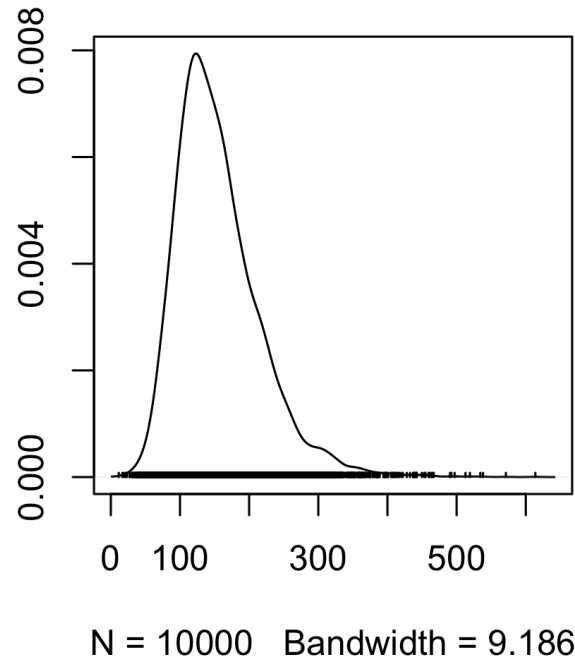
DIAGNOSTICS: TRACE PLOTS

```
plot(SIGMA.mcmc[, "sigma_12"])
```

Trace of var1



Density of var1



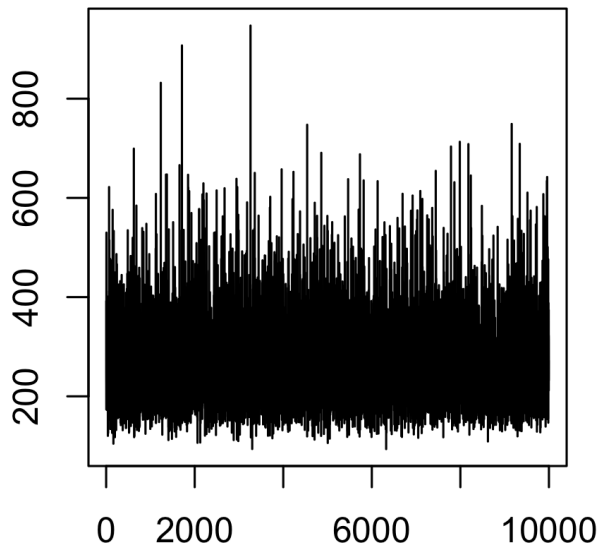
Looks good!



DIAGNOSTICS: TRACE PLOTS

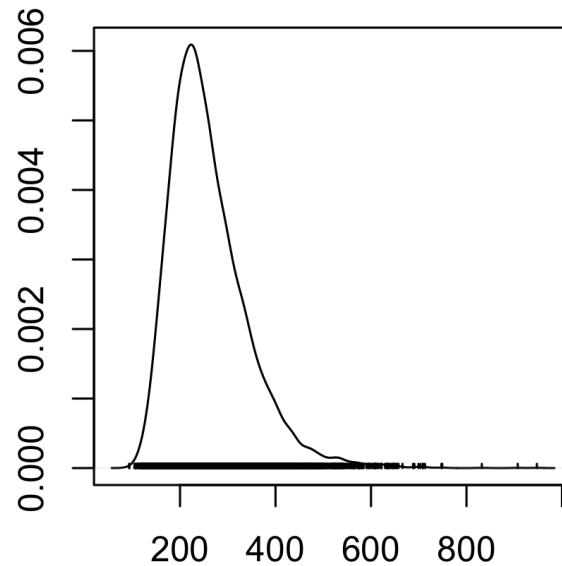
```
plot(SIGMA.mcmc[, "sigma_22"])
```

Trace of var1



Iterations

Density of var1



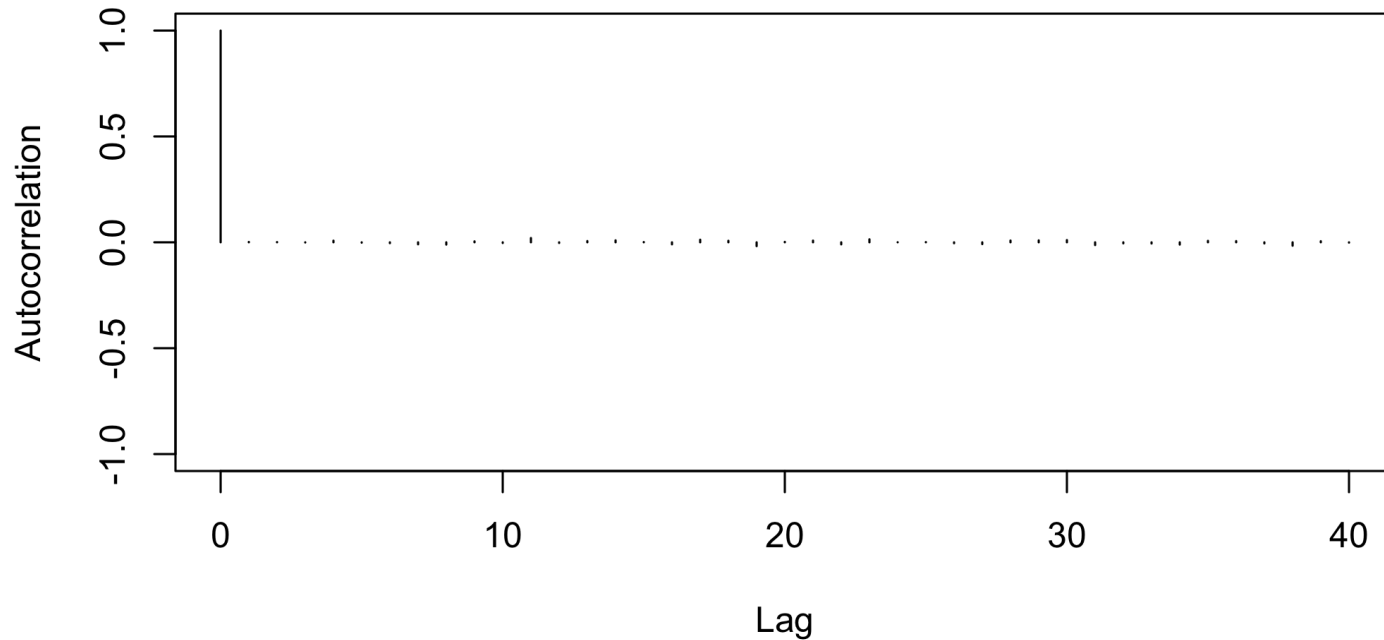
N = 10000 Bandwidth = 12.25

Looks good!



DIAGNOSTICS: AUTOCORRELATION

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

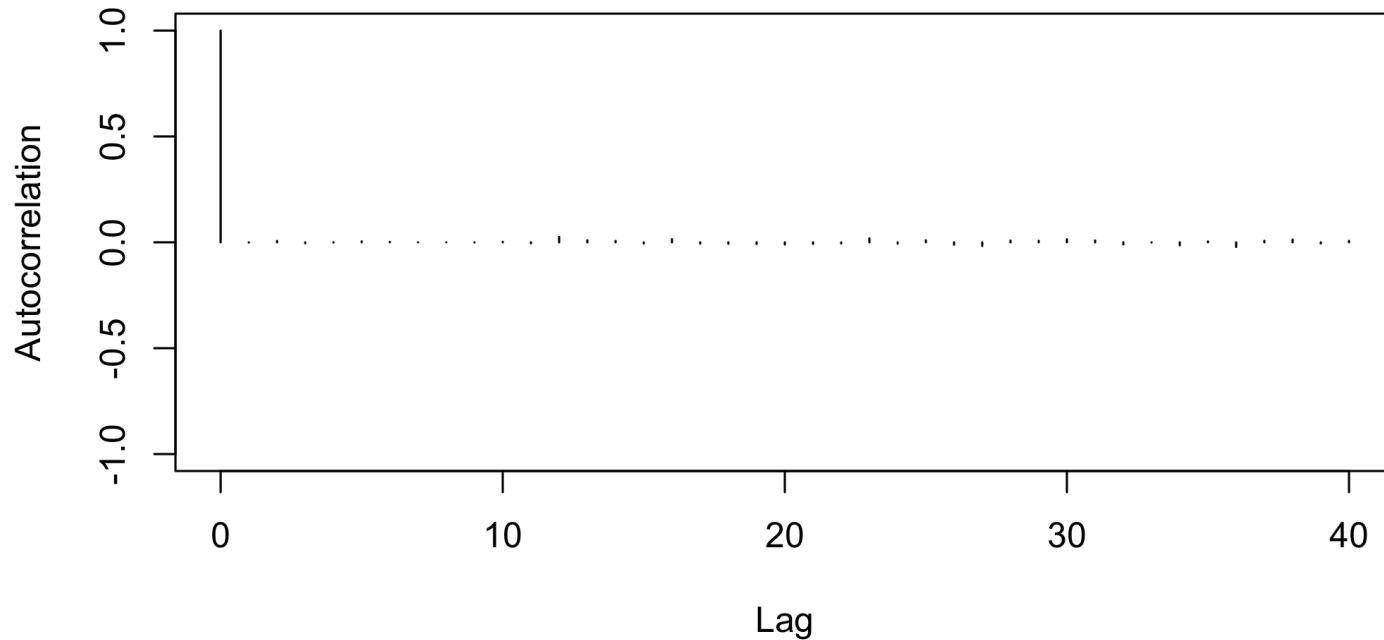


Looks good!



DIAGNOSTICS: AUTOCORRELATION

```
autocorr.plot(THETA.mcmc[, "theta_2"])
```

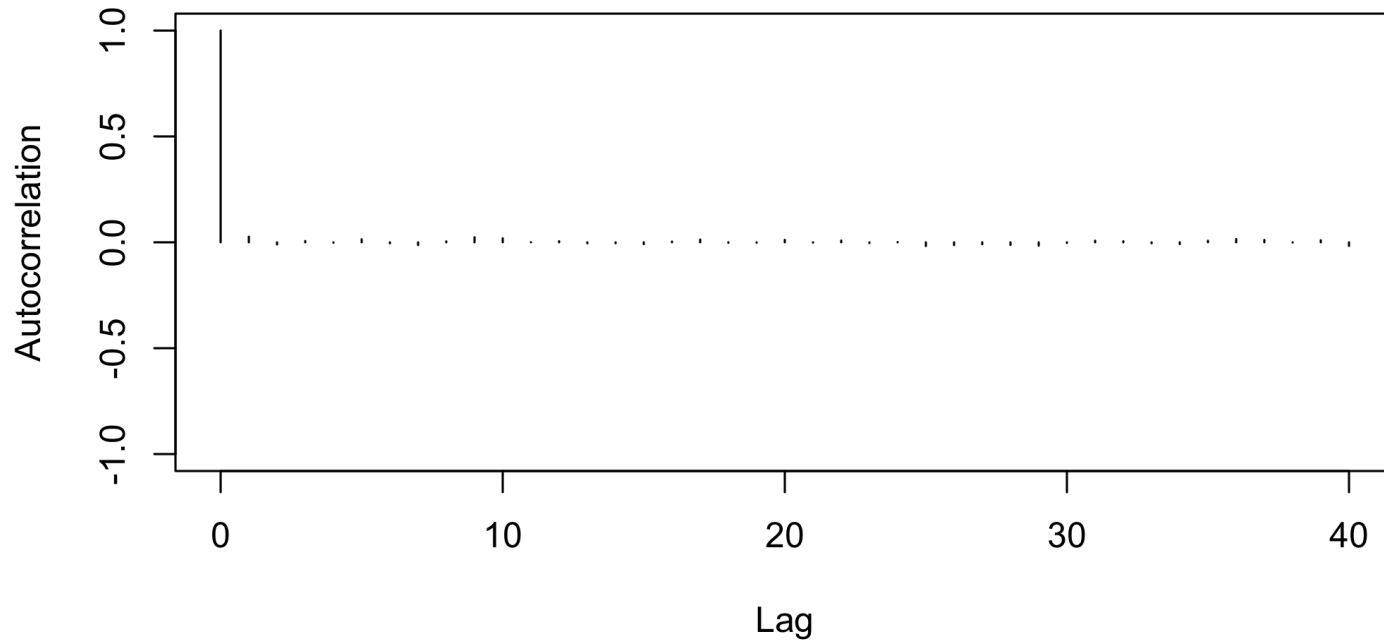


Looks good!



DIAGNOSTICS: AUTOCORRELATION

```
autocorr.plot(SIGMA.mcmc[, "sigma_11"])
```

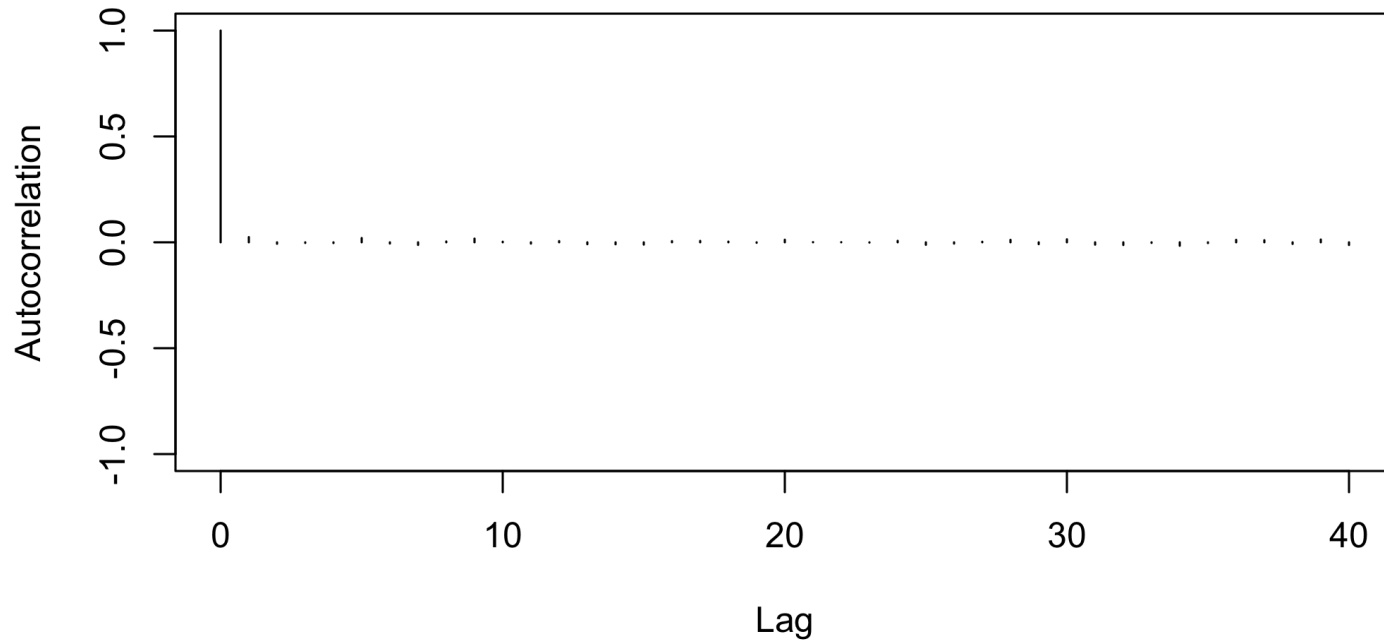


Looks good!



DIAGNOSTICS: AUTOCORRELATION

```
autocorr.plot(SIGMA.mcmc[, "sigma_12"])
```

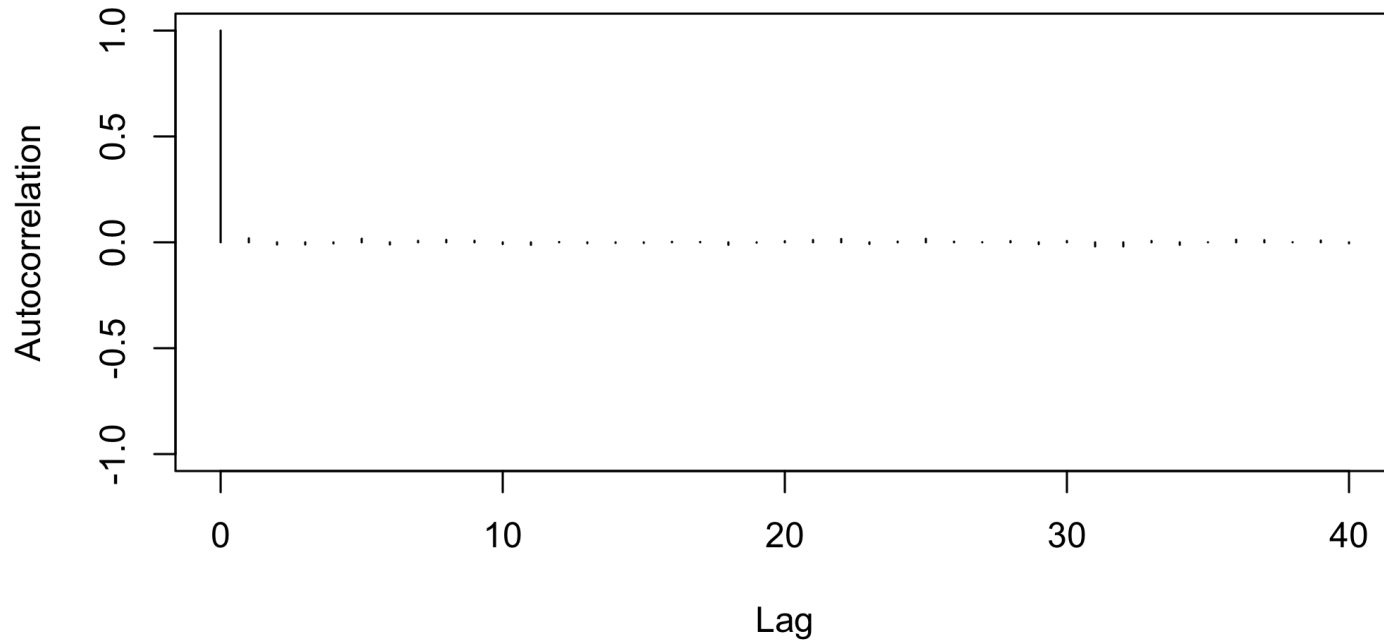


Looks good!



DIAGNOSTICS: AUTOCORRELATION

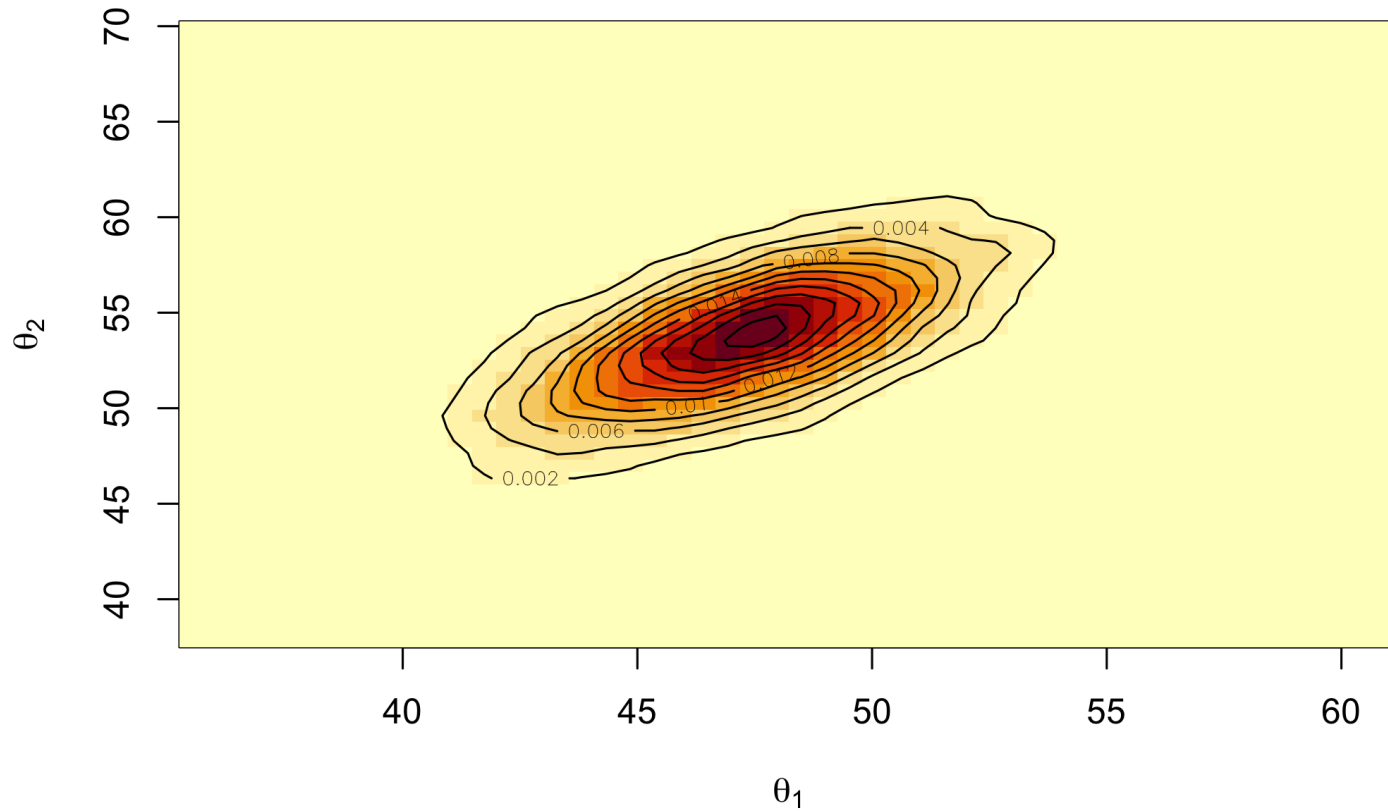
```
autocorr.plot(SIGMA.mcmc[, "sigma_22"])
```



Looks good!



POSTERIOR DISTRIBUTION OF THE MEAN



ANSWERING QUESTIONS OF INTEREST

- Questions of interest:
 - Do students improve in reading comprehension on average?
- Need to compute $\Pr[\theta_2 > \theta_1 | \mathbf{Y}]$. In R,

```
mean(THETA[,2]>THETA[,1])
```

```
## [1] 0.992
```

- That is, posterior probability > 0.99 and indicates strong evidence that test scores are higher in the second administration.

ANSWERING QUESTIONS OF INTEREST

- Questions of interest:
 - If so, by how much?
- Need posterior summaries of $\pi[\theta_2 - \theta_1 | \mathbf{Y}]$. In R,

```
mean(THETA[,2] - THETA[,1])
```

```
## [1] 6.385515
```

```
quantile(THETA[,2] - THETA[,1], prob=c(0.025, 0.5, 0.975))
```

```
##      2.5%      50%      97.5%  
## 1.233154 6.385597 11.551304
```

- Mean (and median) improvement is ≈ 6.39 points with 95% credible interval (1.23, 11.55).

ANSWERING QUESTIONS OF INTEREST

- Questions of interest:
 - How correlated (positively) are the post-test and pre-test scores?
- We can compute $\Pr[\sigma_{12} > 0 | \mathbf{Y}]$. In R,

```
mean(SIGMA[,2]>0)
```

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

- Posterior probability that the covariance between them is positive is basically 1.

ANSWERING QUESTIONS OF INTEREST

- Questions of interest:
 - How correlated (positively) are the post-test and pre-test scores?
- We can also look at the distribution of ρ instead. In R,

```
CORR <- SIGMA[,2]/(sqrt(SIGMA[,1])*sqrt(SIGMA[,4]))  
quantile(CORR,prob=c(0.025, 0.5, 0.975))
```

```
##      2.5%      50%      97.5%  
## 0.4046817 0.6850218 0.8458880
```

- Median correlation between the 2 scores is 0.69 with a 95% quantile-based credible interval of (0.40, 0.85)
- Because density is skewed, we may prefer the 95% HPD interval, which is (0.45, 0.88).

```
#library(hdrcde)  
hdr(CORR,prob=95)$hdr
```

```
##      [,1]      [,2]  
## 95% 0.4468522 0.8761174
```

JEFFREYS' PRIOR

- Clearly, there's a lot of work to be done in specifying the hyperparameters (two of which are $p \times p$ matrices).
- What if we want to specify the priors so that we put in as little information as possible?
- We already know how to do that somewhat with Jeffreys' priors.
- For the multivariate normal model, turns out that the Jeffreys' rule for generating a prior distribution on $(\boldsymbol{\theta}, \Sigma)$ gives

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

- Can we derive the full conditionals under this prior?
- **To be done during discussion session.**

JEFFREYS' PRIOR

- We will leverage previous work. For the likelihood we have both

$$p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma) \propto \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}^T (n\Sigma^{-1}) \boldsymbol{\theta} + \boldsymbol{\theta}^T (n\Sigma^{-1} \bar{\mathbf{y}}) \right\}$$

and

$$p(\mathbf{Y}|\boldsymbol{\theta}, \Sigma) \propto |\Sigma|^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \text{tr} [\mathbf{S}_\theta \Sigma^{-1}] \right\},$$

where $\mathbf{S}_\theta = \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})^T$.

- Also, we can rewrite any $\mathcal{N}_p(\boldsymbol{\mu}_0, \Lambda_0)$ as

$$p(\boldsymbol{\theta}) \propto \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}^T \Lambda_0^{-1} \boldsymbol{\theta} + \boldsymbol{\theta}^T \Lambda_0^{-1} \boldsymbol{\mu}_0 \right\}.$$

- Finally, $\Sigma \sim \mathcal{IW}_p(\nu_0, \mathbf{S}_0)$,

$$\Rightarrow p(\Sigma) \propto |\Sigma|^{\frac{-(\nu_0+p+1)}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{S}_0 \Sigma^{-1}) \right\}.$$

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!