Multivariate normal cont'd; missing data and imputation

Dr. Olanrewaju Michael Akande

Feb 26, 2020

OUTLINE

- Multivariate normal/Gaussian model
  - Recap
  - Reading example cont'd
  - Answering questions
  - Jeffreys' prior
- Missing data and imputation
  - Missing data mechanisms
  - Multivariate normal/Gaussian model
  - Example

Reading example cont'd

Reading example

- Y_{i1}: pre-instructional score for student i and Y_{i2}: post-instructional score for student i.
- Model:
  - $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})^T \sim \mathcal{N}_2(\boldsymbol{\theta}, \Sigma)$,
  - $\pi(\boldsymbol{\theta}) = \mathcal{N}_2(\boldsymbol{\mu}_0, \Lambda_0)$, and
  - $\pi(\Sigma) = \mathcal{IW}_2(\nu_0, \boldsymbol{S}_0)$.
- Then,

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

where

$$\begin{aligned}
\Lambda_n &= \left[\Lambda_0^{-1} + n\Sigma^{-1}\right]^{-1} \\
\boldsymbol{\mu}_n &= \Lambda_n \left[\Lambda_0^{-1}\boldsymbol{\mu}_0 + n\Sigma^{-1}\bar{\boldsymbol{y}}\right] \\
\boldsymbol{\mu}_0 &= \begin{pmatrix} 50 \\ 50 \end{pmatrix}; \quad \Lambda_0 = \begin{pmatrix} 156 & 78 \\ 78 & 156 \end{pmatrix}.
\end{aligned}$$

Reading example: posterior computation

- and

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

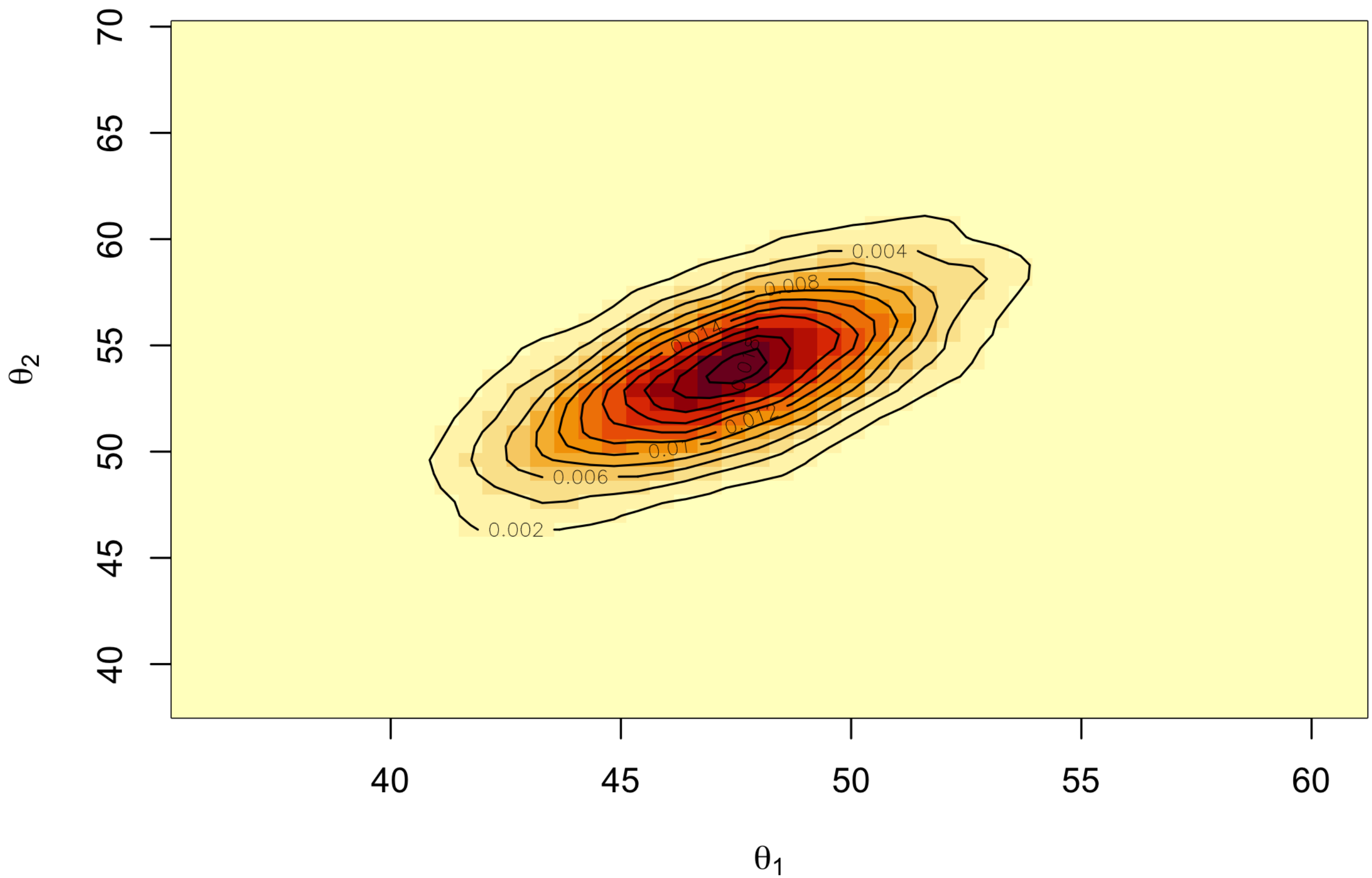
or using the notation in the book, $\mathcal{IW}_2(\nu_n, \boldsymbol{S}_n^{-1})$, where

$$\begin{aligned}
\nu_n &= \nu_0 + n \\
\\
\boldsymbol{S}_n &= [\boldsymbol{S}_0 + \boldsymbol{S}_\theta] \\
&= \left[\boldsymbol{S}_0 + \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\theta})(\boldsymbol{y}_i - \boldsymbol{\theta})^T\right].
\end{aligned}$$

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

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

Posterior distribution of the mean



Answering questions of interest

- Questions of interest:
  - Do students improve in reading comprehension on average?
- Need to compute $\Pr[\theta_2 > \theta_1 | \boldsymbol{Y}]$. 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 $\Pr[\theta_2 - \theta_1 | \boldsymbol{Y}]$. 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 $\Pr[\sigma_{12} > 0|\boldsymbol{Y}]$. 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 or 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 on the board.**

Jeffreys' prior

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

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

and

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

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

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

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

Missing data and imputation

Missing data

- Missing data/nonresponse is fairly common in real data. For example,
  - Failure to respond to survey question
  - Subject misses some clinic visits out of all possible
  - Only subset of subjects asked certain questions
- Recall that our posterior computation usually depends on the data through $\mathcal{L}(Y;\theta)$, which cannot be computed when some of the y_i values are missing.
- The most common software packages often throw away all subjects with incomplete data (can lead to bias and precision loss).
- Some individuals impute missing values with a mean or some other fixed value (ignores uncertainty).
- As you will see, imputing missing data is actually quite natural in the Bayesian context.

Missing data mechanisms

- Data are said to be missing completely at random (MCAR) if the reason for missingness does not depend on the values of the observed data or missing data.

- For example, suppose

  - you handed out a double-sided survey questionnaire of 20 questions to a sample of participants;

  - questions 1-15 were on the first page but questions 16-20 were at the back; and

  - some of the participants did not respond to questions 16-20.

- Then, the values for questions 16-20 for those people who did not respond would be MCAR if they simply did not realize the pages were double-sided; they had no reason to ignore those questions.

- **This is rarely plausible in practice!**

Missing data mechanisms

- Data are said to be missing at random (MAR) if, conditional on the values of the observed data, the reason for missingness does not depend on the missing data.
- Using our previous example, suppose
  - questions 1-15 include demographic information such as age and education;
  - questions 16-20 include income related questions; and
  - once again, some participants did not respond to questions 16-20.
- Then, the values for questions 16-20 for those people who did not respond would be MAR if younger people are more likely not to respond to those income related questions than old people, where age is observed for all participants.
- **This is the most commonly assumed mechanism in practice!**

Missing data mechanisms

- Data are said to be missing not at random (MNAR or NMAR) if the reason for missingness depends on the actual values of the missing (unobserved) data.
- Continuing with our previous example, suppose again that
  - questions 1-15 include demographic information such as age and education;
  - questions 16-20 include income related questions; and
  - once again, some of the participants did not respond to questions 16-20.
- Then, the values for questions 16-20 for those people who did not respond would be MNAR if people who earn more money are less likely to respond to those income related questions than old people.
- **This is usually the case in real data, but analysis can be complex!**

Mathematical formulation

- Consider the multivariate data scenario with $\boldsymbol{Y}_i = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^T$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^T$, for $i = 1, \ldots, n$.
- For now, we will assume the multivariate normal model as the sampling model, so that each $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^T \sim \mathcal{N}_p(\boldsymbol{\theta}, \Sigma)$.
- Suppose now that $\boldsymbol{Y}$ contains missing values.
- We can separate $\boldsymbol{Y}$ into the observed and missing parts, that is, $\boldsymbol{Y} = (\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis})$.
- Then for each individual, $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i,obs}, \boldsymbol{Y}_{i,mis})$.

Mathematical Formulation

- Let
  - j index variables (where i already indexes individuals),
  - $r_{ij} = 1$ when y_{ij} is missing,
  - $r_{ij} = 0$ when y_{ij} is observed.
- Here, r_{ij} is known as the missingness indicator of variable j for person i.
- Also, let
  - $\boldsymbol{R}_i = (r_{i1}, \ldots, r_{ip})^T$ be the vector of missing indicators for person i.
  - $\boldsymbol{R} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n)$ be the matrix of missing indicators for everyone.
  - $\boldsymbol{\psi}$ be the set of parameters associated with $\boldsymbol{R}$.
- Assume $\boldsymbol{\psi}$ and $(\boldsymbol{\theta}, \Sigma)$ are distinct.

Mathematical Formulation

- MCAR:

$$p(\boldsymbol{R}|\boldsymbol{Y}, \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) = p(\boldsymbol{R}|\boldsymbol{\Psi})$$

- MAR:

$$p(\boldsymbol{R}|\boldsymbol{Y}, \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) = p(\boldsymbol{R}|\boldsymbol{Y}_{obs}, \boldsymbol{\Psi})$$

- MNAR:

$$p(\boldsymbol{R}|\boldsymbol{Y}, \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) = p(\boldsymbol{R}|\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\Psi})$$

Implications for likelihood function

- Each type of mechanism has a different implication on the likelihood of the observed data $\boldsymbol{Y}_{obs}$, and the missing data indicator $\boldsymbol{R}$.

- Without missingness in $\boldsymbol{Y}$, the likelihood of the observed data is

$$\mathcal{L}(\boldsymbol{Y}_{obs}; \boldsymbol{\theta}, \Sigma) \propto p(\boldsymbol{Y}_{obs} | \boldsymbol{\theta}, \Sigma)$$

- With missingness in $\boldsymbol{Y}$, the likelihood of the observed data is instead

$$\begin{aligned} L(\boldsymbol{Y}_{obs}, \boldsymbol{R}; \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) &\propto p(\boldsymbol{Y}_{obs}, \boldsymbol{R} | \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) \\ &= \int p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \end{aligned}$$

- Since we do not actually observe $\boldsymbol{Y}_{mis}$, we would like to be able to integrate it out so we don't have to deal with it.

- That is, we would like to infer $(\boldsymbol{\theta}, \Sigma)$ (and sometimes, $\boldsymbol{\psi}$) using only the observed data.

Likelihood function: MCAR

- For MCAR, we have:

$$\begin{aligned}
L(\boldsymbol{Y}_{obs}, \boldsymbol{R}; \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) &\propto p(\boldsymbol{Y}_{obs}, \boldsymbol{R} | \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) \\
&= \int p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= \int p(\boldsymbol{R} | \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= p(\boldsymbol{R} | \boldsymbol{\psi}) \cdot \int p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= p(\boldsymbol{R} | \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs} | \boldsymbol{\theta}, \Sigma).
\end{aligned}$$

- For inference on $(\boldsymbol{\theta}, \Sigma)$, we can simply focus on $p(\boldsymbol{Y}_{obs} | \boldsymbol{\theta}, \Sigma)$ in the likelihood function, since $(\boldsymbol{R} | \boldsymbol{\psi})$ does not include any $\boldsymbol{Y}$.

Likelihood function: MAR

- For MAR, we have:

$$
\begin{aligned}
L(\boldsymbol{Y}_{obs}, \boldsymbol{R}; \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) &\propto p(\boldsymbol{Y}_{obs}, \boldsymbol{R} | \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) \\
&= \int p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= \int p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{\psi}) \cdot \int p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis} \\
&= p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs} | \boldsymbol{\theta}, \Sigma).
\end{aligned}
$$

- For inference on $(\boldsymbol{\theta}, \Sigma)$, we can once again focus on $p(\boldsymbol{Y}_{obs} | \boldsymbol{\theta}, \Sigma)$ in the likelihood function, although there can be some bias if we do not account for $p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{\theta})$, since it contains observed data.

- Also, if we want to infer the missingness mechanism through ψ, we would need to deal with $p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{X}, \boldsymbol{\theta})$ anyway.

Likelihood function: MNAR

- For MNAR, we have:

$$
\begin{aligned}
L(\boldsymbol{Y}_{obs}, \boldsymbol{R}; \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) &\propto p(\boldsymbol{Y}_{obs}, \boldsymbol{R} | \boldsymbol{\theta}, \Sigma, \boldsymbol{\psi}) \\
&= \int p(\boldsymbol{R} | \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}, \boldsymbol{\psi}) \cdot p(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis} | \boldsymbol{\theta}, \Sigma) \mathrm{d}\boldsymbol{Y}_{mis}.
\end{aligned}
$$

- The likelihood under MNAR cannot simplify any further.
- In this case, we cannot ignore the missing data when making inferences about $(\boldsymbol{\theta}, \Sigma)$.
- We must include the model for $\boldsymbol{R}$ and also infer the missing data $\boldsymbol{Y}_{mis}$.

How to tell in practice?

- So how can we tell the type of mechanism we are dealing with?
- In general, we don't know!!!
- Rare that data are MCAR (unless planned beforehand); more likely that data are MNAR.
- **Compromise**: assume data are MAR if we include enough variables in model for the missing data indicator $\boldsymbol{R}$.
- Whenever we talk about missing data in this course, we will do so in the context of MCAR and MAR.

Bayesian inference with missing data

- As we have seen, for MCAR and MAR, we can focus on $p(\boldsymbol{Y}_{obs}|\boldsymbol{\theta}, \Sigma)$ in the likelihood function, when inferring $(\boldsymbol{\theta}, \Sigma)$.
- While this is great, for posterior sampling under most models (especially multivariate models), we actually do need all the $\boldsymbol{Y}$'s to update the parameters.
- In addition, we may actually want to learn about the missing values, in addition to inferring $(\boldsymbol{\theta}, \Sigma)$.
- By thinking of the missing data as **another set of parameters**, we can sample them from the "posterior predictive" distribution of the missing data conditional on the observed data and parameters:

$$p(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\theta}, \Sigma) \propto \prod_{i=1}^{n} p(\boldsymbol{Y}_{i,mis}|\boldsymbol{Y}_{i,obs}, \boldsymbol{\theta}, \Sigma).$$

- In the case of the multivariate model, each $p(\boldsymbol{Y}_{i,mis}|\boldsymbol{Y}_{i,obs}, \boldsymbol{\theta}, \Sigma)$ is just a normal distribution, and we can leverage results on conditional distributions for normal models.

Gibbs sampler with missing data

At iteration $s+1$, do the following

1. Sample $\boldsymbol{\theta}^{(s+1)}$ from its multivariate normal full conditional

$$p(\boldsymbol{\theta}^{(s+1)}|\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}^{(s)}, \Sigma^{(s)}).$$

2. Sample $\Sigma^{(s+1)}$ from its inverse-Wishart full conditional

$$p(\Sigma^{(s+1)}|\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}^{(s)}, \boldsymbol{\theta}^{(s+1)}).$$

3. For each $i = 1, \ldots, n$, with at least one zero value in the missingness indicator vector $\boldsymbol{R}_i$, sample $\boldsymbol{Y}_{i,mis}^{(s+1)}$ from the full conditional

$$p(\boldsymbol{Y}_{i,mis}^{(s+1)}|\boldsymbol{Y}_{i,obs}, \boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)}),$$

   which is also multivariate normal, with its form derived by original sampling model but with the updated parameters, that is, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^T = (\boldsymbol{Y}_{i,obs}, \boldsymbol{Y}_{i,mis})^T \sim \mathcal{N}_p(\boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)})$.

Gibbs sampler with missing data

- Rewrite $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i,mis}, \boldsymbol{Y}_{i,obs})^T \sim \mathcal{N}_p(\boldsymbol{\theta}^{(s+1)}, \Sigma^{(s+1)})$ as

$$\boldsymbol{Y}_i = \begin{pmatrix} \boldsymbol{Y}_{i,mis} \\ \boldsymbol{Y}_{i,obs} \end{pmatrix} \sim \mathcal{N}_p \left[\begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right],$$

  so that we can take advantage of the conditional normal results.

- That is, we have

$$\boldsymbol{Y}_{i,mis} | \boldsymbol{Y}_{i,obs} = \boldsymbol{y}_{i,obs} \sim \mathcal{N}\left(\boldsymbol{\theta}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y}_{i,obs} - \boldsymbol{\theta}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right).$$

  as the multivariate normal distribution (or univariate normal distribution if $\boldsymbol{Y}_i$ only has one missing entry) we need in step 3 of the Gibbs sampler.

- This sampling technique actually encodes MAR since the imputations for $\boldsymbol{Y}_{mis}$ depend on the $\boldsymbol{Y}_{obs}$.

- Now let's revisit the reading comprehension example again. We will add missing values to the original data and refit the model.

Reading example with missing data

```
Y <- as.matrix(dget("http://www2.stat.duke.edu/~pdh10/FCBS/Inline/Y.reading"))

#Add 20% missing data; MCAR
set.seed(1234)
Y_WithMiss <- Y #So we can keep the full data
Miss_frac <- 0.20
R <- matrix(rbinom(nrow(Y)*ncol(Y),1,Miss_frac),nrow(Y),ncol(Y))
Y_WithMiss[R==1]<-NA
Y_WithMiss[1:12,]
```

```
##       pretest posttest
##  [1,]      59       77
##  [2,]      43       39
##  [3,]      34       46
##  [4,]      32       NA
##  [5,]      NA       38
##  [6,]      38       NA
##  [7,]      55       NA
##  [8,]      67       86
##  [9,]      64       77
## [10,]      45       60
## [11,]      49       50
## [12,]      72       59
```

```
colMeans(is.na(Y_WithMiss))
```

```
##   pretest  posttest
## 0.1363636 0.2272727
```

Reading example with missing data

```r
#ACTUAL ANALYSIS STARTS HERE!!!
#Data dimensions
n <- nrow(Y); p <- ncol(Y)

#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)

#Define missing data indicators
##we already know R. This is to write a more general code for when we don't
R <- 1*(is.na(Y_WithMiss))
R[1:12,]
```

```
##       pretest posttest
##  [1,]       0        0
##  [2,]       0        0
##  [3,]       0        0
##  [4,]       0        1
##  [5,]       1        0
##  [6,]       0        1
##  [7,]       0        1
##  [8,]       0        0
##  [9,]       0        0
## [10,]       0        0
## [11,]       0        0
## [12,]       0        0
```

Reading example with missing data

```
#Initial values for Gibbs sampler
Y_Full <- Y_WithMiss #So we can keep the data with missing values as is
for (j in 1:p) {
Y_Full[is.na(Y_Full[,j]),j] <- mean(Y_Full[,j],na.rm=TRUE) #start with mean imputation
}

Sigma <- S_0 # can't really rely on cov(Y) because we don't have full Y

#Set null objects to save samples
THETA_WithMiss <- NULL
SIGMA_WithMiss <- NULL
Y_MISS <- NULL

#first set number of iterations and burn-in, then set seed
n_iter <- 10000; burn_in <- 0.3*n_iter
```

Gibbs sampler with missing data

```
#library(mvtnorm) for multivariate normal
#library(MCMCpack) for inverse-Wishart

Lambda_0_inv <- solve(Lambda_0) #move outside sampler since it does not change

for (s in 1:(n_iter+burn_in)){
  ##first we must recalculate ybar inside the loop now since it changes every iteration
  ybar <- apply(Y_Full,2,mean)

  ##update theta
  Sigma_inv <- solve(Sigma) #invert once
  Lambda_n <- solve(Lambda_0_inv + n*Sigma_inv)
  mu_n <- Lambda_n %*% (Lambda_0_inv%*%mu_0 + n*Sigma_inv%*%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)
```

Gibbs sampler with missing data

```r
##update missing data using updated draws of theta and Sigma
  for(i in 1:n) {
    if(sum(R[i,]>0)){
       obs_index <- R[i,]==0
       mis_index <- R[i,]==1
       Sigma_22_obs_inv <- solve(Sigma[obs_index,obs_index]) #invert just once
       Sigma_12_Sigma_22_obs_inv <- Sigma[mis_index,obs_index]%*%Sigma_22_obs_inv

       Sigma_cond_mis <- Sigma[mis_index,mis_index] -
         Sigma_12_Sigma_22_obs_inv%*%Sigma[obs_index,mis_index]

       mu_cond_mis <- theta[mis_index] +
         Sigma_12_Sigma_22_obs_inv%*%(t(Y_Full[i,obs_index])-theta[obs_index])

      Y_Full[i,mis_index] <- rmvnorm(1,mu_cond_mis,Sigma_cond_mis)
      }
    }

  #save results only past burn-in
  if(s > burn_in){
  THETA_WithMiss <- rbind(THETA_WithMiss,theta)
  SIGMA_WithMiss <- rbind(SIGMA_WithMiss,c(Sigma))
  Y_MISS <- rbind(Y_MISS, Y_Full[R==1] )
  }
}

colnames(THETA_WithMiss) <- c("theta_1","theta_2")
colnames(SIGMA_WithMiss) <- c("sigma_11","sigma_12","sigma_21","sigma_22") #symmetry in sig
```

Diagnostics

```r
#library(coda)
THETA_WithMiss.mcmc <- mcmc(THETA_WithMiss,start=1); summary(THETA_WithMiss.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 45.70 3.085  0.03085        0.03346
## theta_2 54.09 3.560  0.03560        0.04055
##
## 2. Quantiles for each variable:
##
##          2.5%   25%   50%   75% 97.5%
## theta_1 39.58 43.67 45.69 47.77 51.76
## theta_2 47.05 51.76 54.11 56.44 61.26
```

Diagnostics

```
SIGMA_WithMiss.mcmc <- mcmc(SIGMA_WithMiss,start=1); summary(SIGMA_WithMiss.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_11 204.5 64.00   0.6400         0.6075
## sigma_12 154.9 61.62   0.6162         0.6602
## sigma_21 154.9 61.62   0.6162         0.6602
## sigma_22 262.0 83.48   0.8348         0.8757
## 
## 2. Quantiles for each variable:
## 
##           2.5%   25%   50%   75% 97.5%
## sigma_11 114.9 159.7 193.1 235.8 360.5
## sigma_12  65.7 111.8 144.3 186.2 306.0
## sigma_21  65.7 111.8 144.3 186.2 306.0
## sigma_22 145.8 203.2 246.3 304.1 463.9
```

Compare to inference from full data

With missing data:

```
apply(THETA_WithMiss,2,summary)
```

```
##          theta_1  theta_2
## Min.    33.33374 40.96378
## 1st Qu. 43.66655 51.76391
## Median  45.68843 54.10700
## Mean    45.70132 54.09332
## 3rd Qu. 47.77084 56.43931
## Max.    59.35106 71.24531
```

Based on true data:

```
apply(THETA,2,summary)
```

```
##          theta_1  theta_2
## Min.    35.50314 37.80999
## 1st Qu. 45.35465 51.53327
## Median  47.36177 53.68602
## Mean    47.29978 53.68529
## 3rd Qu. 49.22875 55.82192
## Max.    60.94924 69.92354
```

Very similar for the most part.

Compare to inference from full data

With missing data:

```
apply(SIGMA_WithMiss,2,summary)
```

```
##           sigma_11   sigma_12   sigma_21 sigma_22
## Min.      80.74256   7.013531   7.013531 100.0315
## 1st Qu. 159.74990 111.820018 111.820018 203.2094
## Median  193.05114 144.321157 144.321157 246.2548
## Mean    204.50545 154.872072 154.872072 261.9525
## 3rd Qu. 235.78476 186.237581 186.237581 304.1320
## Max.    683.92189 613.032587 613.032587 840.7878
```

Based on true data:

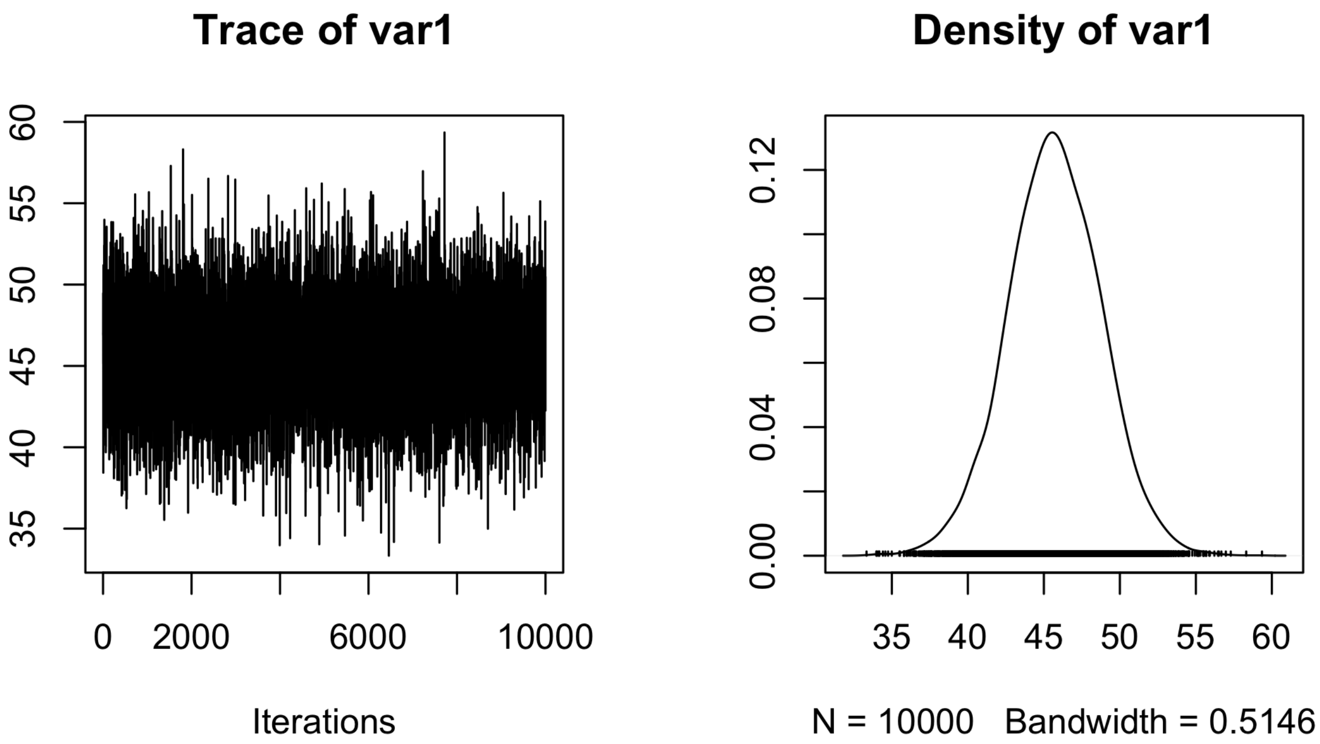
```
apply(SIGMA,2,summary)
```

```
##           sigma_11  sigma_12  sigma_21  sigma_22
## Min.      79.44258  11.41663  11.41663  93.65776
## 1st Qu. 158.21469 113.23258 113.23258 203.21138
## Median  190.77854 144.74881 144.74881 244.56334
## Mean    202.34721 155.33355 155.33355 260.07072
## 3rd Qu. 234.77319 186.50429 186.50429 300.90761
## Max.    671.16538 613.88088 613.88088 947.39333
```

Also very similar. A bit more uncertainty in dimension of Y_{i2} because we have more missing data there.

Diagnostics: trace plots

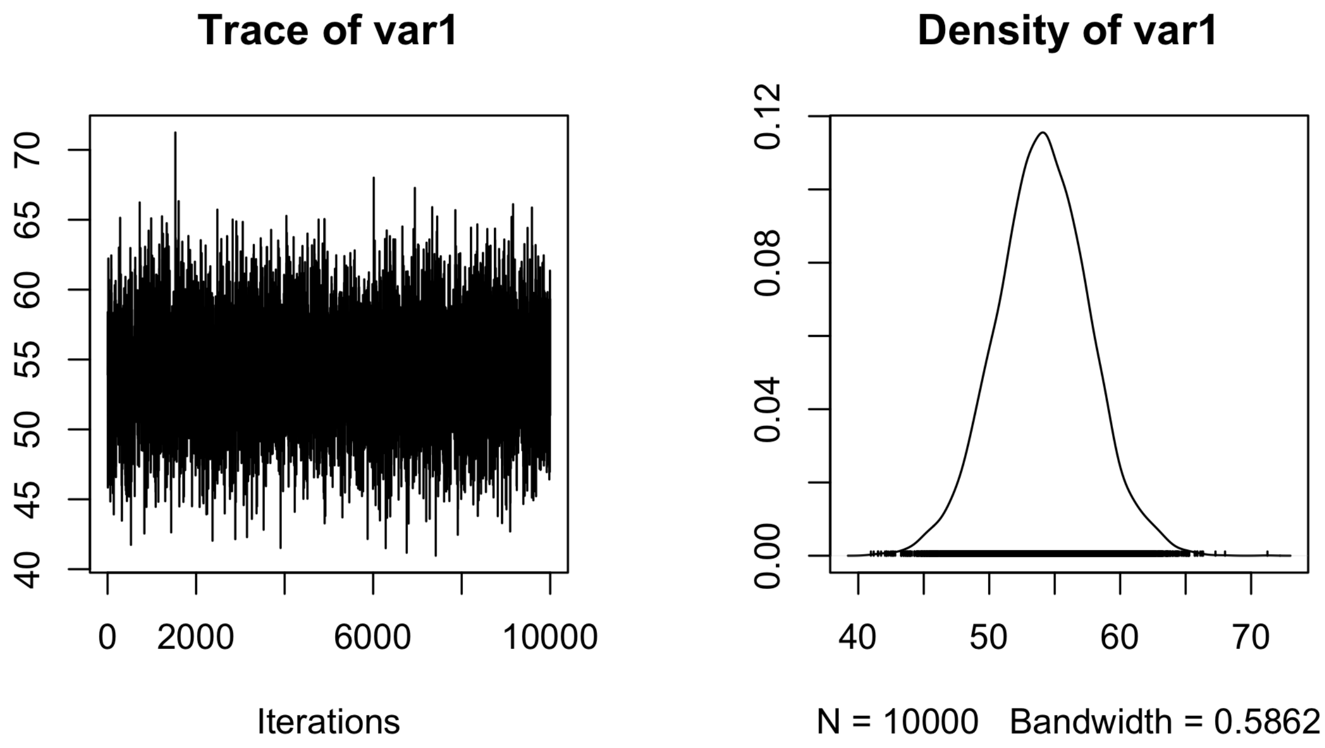
```
plot(THETA_WithMiss.mcmc[,"theta_1"])
```



Looks good!

Diagnostics: trace plots

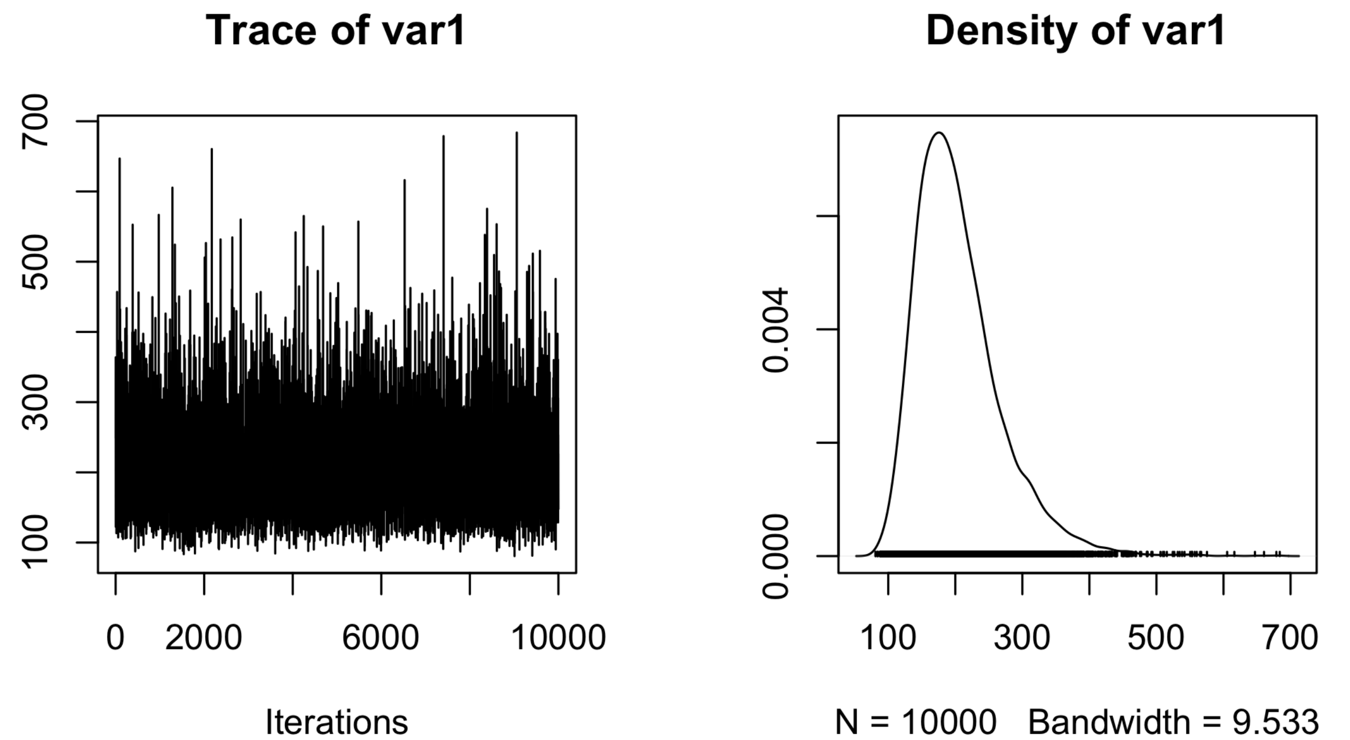
```
plot(THETA_WithMiss.mcmc[,"theta_2"])
```



Looks good!

Diagnostics: trace plots

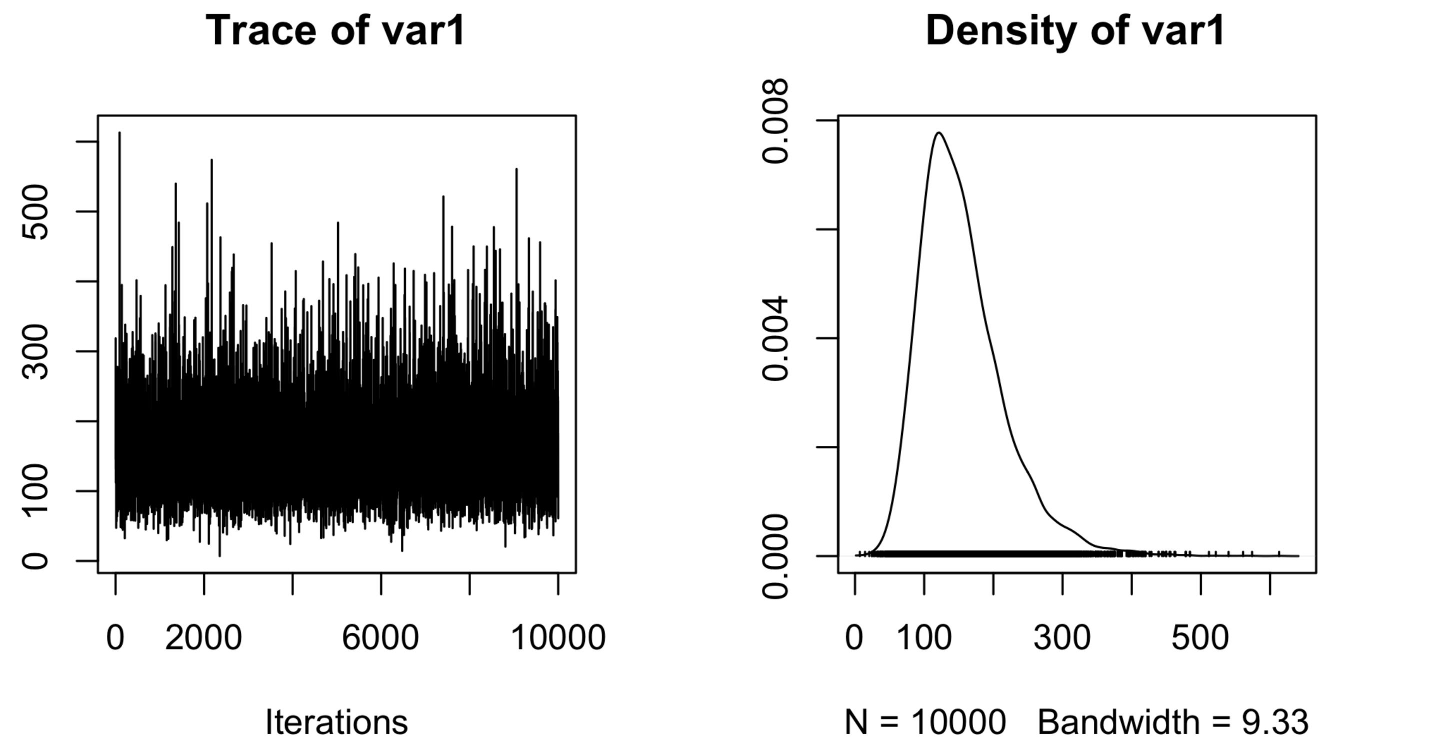
```
plot(SIGMA_WithMiss.mcmc[,"sigma_11"])
```



Looks good!

DIAGNOSTICS: TRACE PLOTS

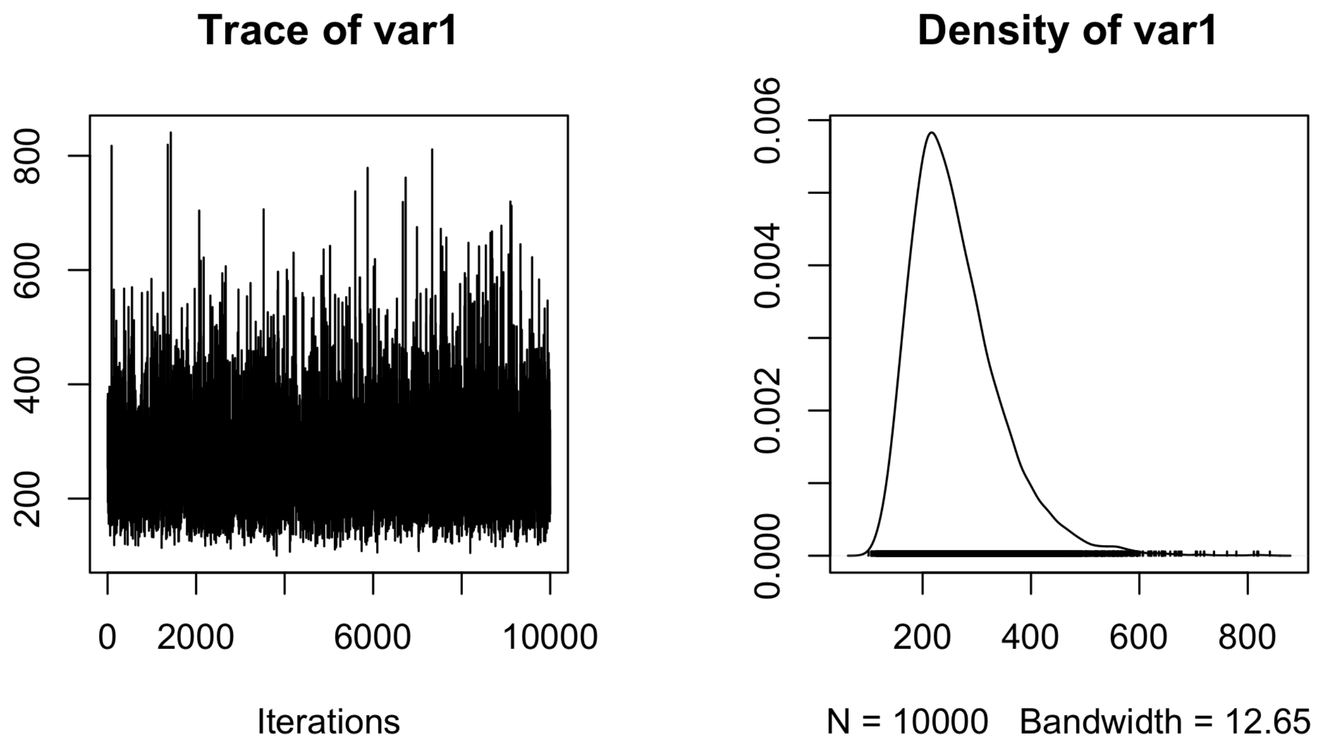
```
plot(SIGMA_WithMiss.mcmc[,"sigma_12"])
```



Looks good!

Diagnostics: trace plots

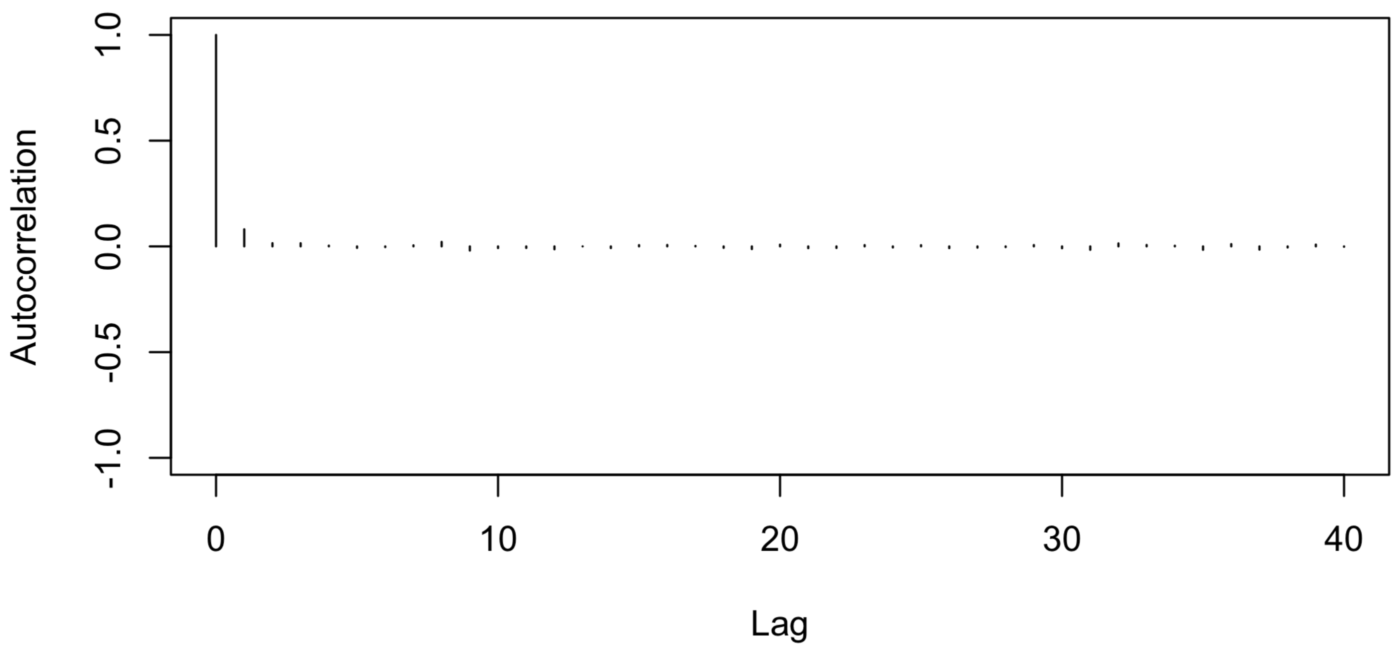
```
plot(SIGMA_WithMiss.mcmc[,"sigma_22"])
```



Looks good!

Diagnostics: autocorrelation

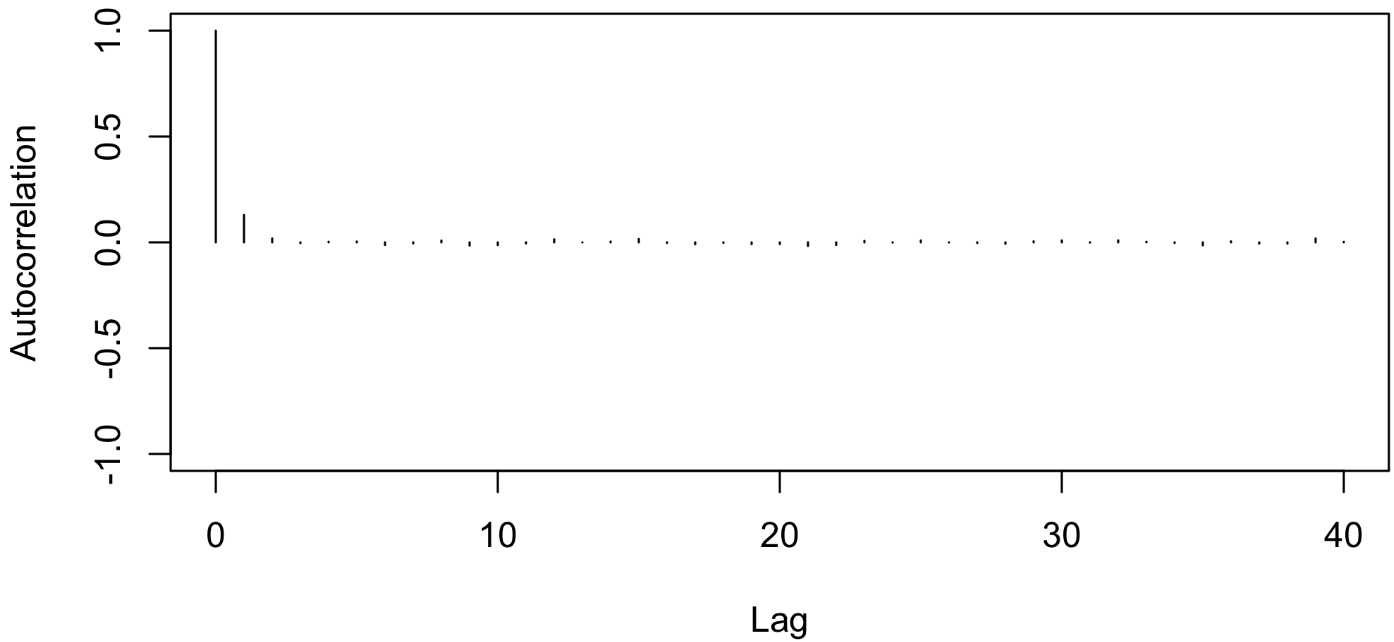
```
autocorr.plot(THETA_WithMiss.mcmc[,"theta_1"])
```



Looks good!

Diagnostics: autocorrelation

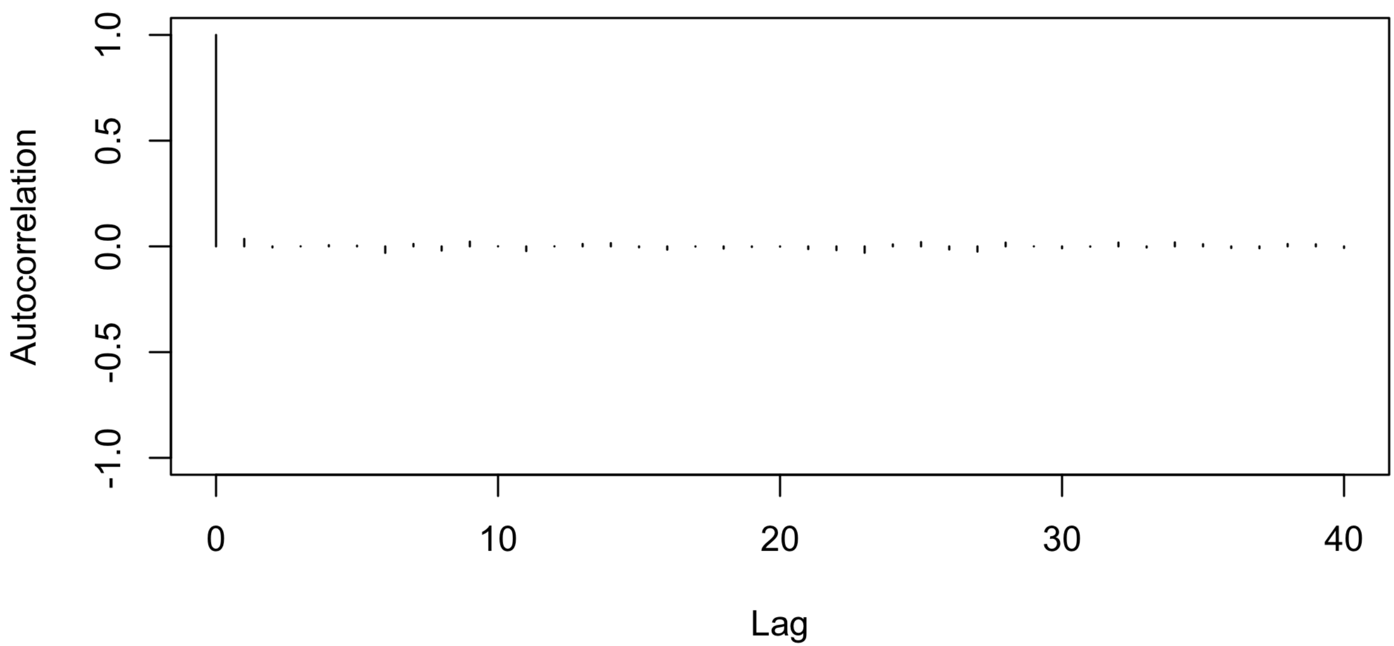
```
autocorr.plot(THETA_WithMiss.mcmc[,"theta_2"])
```



Looks good!

Diagnostics: autocorrelation

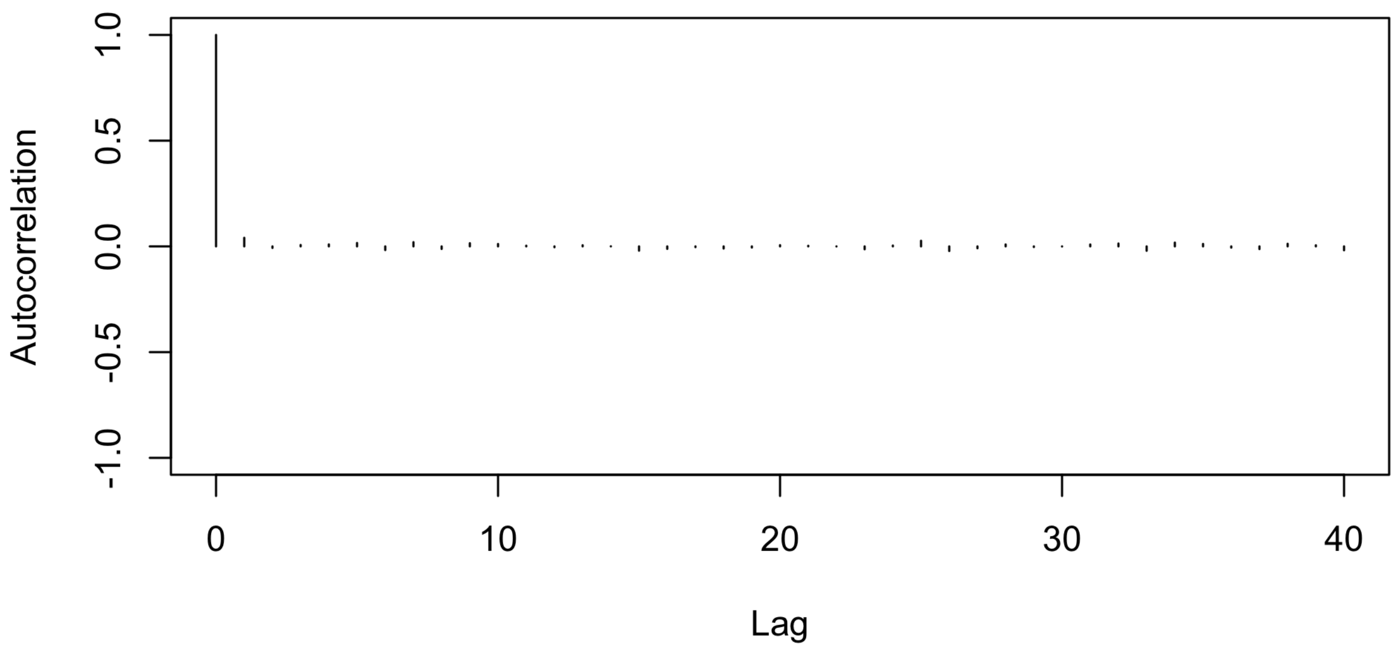
```
autocorr.plot(SIGMA_WithMiss.mcmc[,"sigma_11"])
```



Looks good!

Diagnostics: autocorrelation

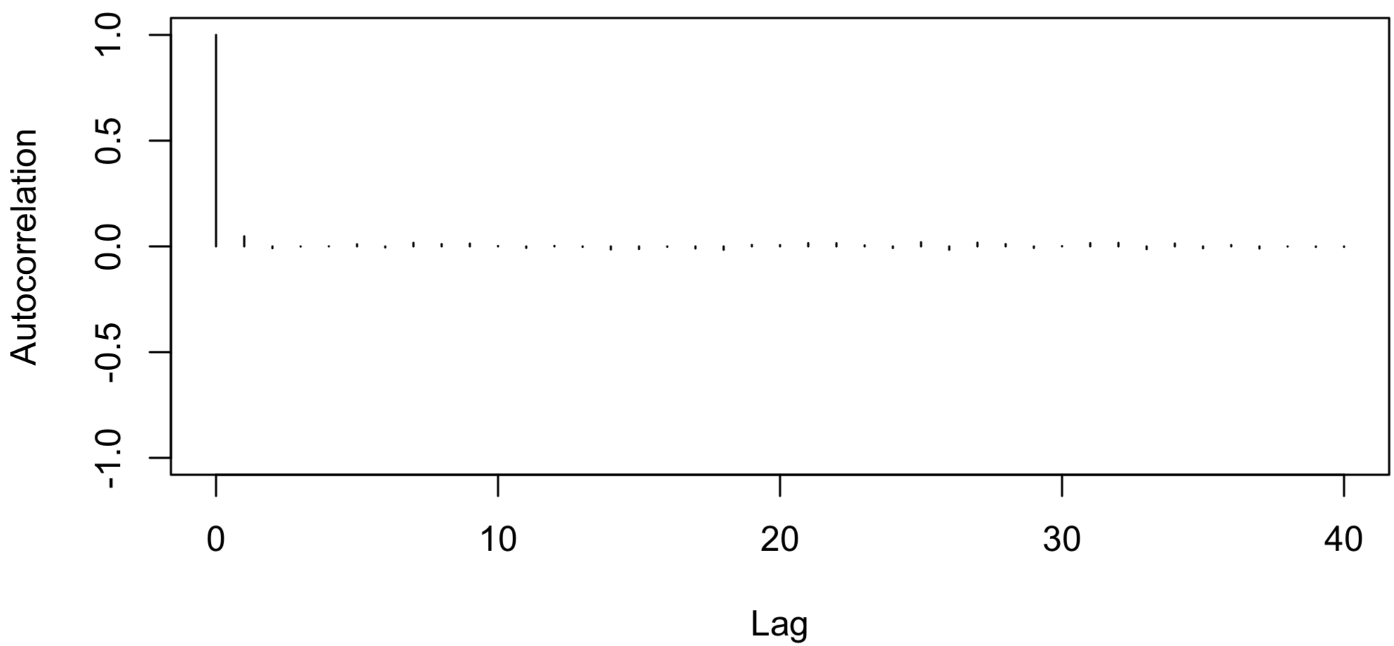
```
autocorr.plot(SIGMA_WithMiss.mcmc[,"sigma_12"])
```



Looks good!

Diagnostics: autocorrelation

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



Looks good!

Posterior distribution of the mean