STA 602 - Intro to Bayesian Statistics

Lecture 13

Li Ma

Duke University

Bayesian inference under independent Normal-inverse-Wishart priors

- Recall our sampling model

$$\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n \mid \boldsymbol{\theta}, \Sigma \stackrel{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \Sigma).$$

- Indepednent prior specification:

$$p(\boldsymbol{\theta}, \Sigma) = p(\boldsymbol{\theta}) p(\Sigma)$$

with

$$\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}_0, \Lambda_0)$$

and

$$\Sigma \sim \mathrm{IW}(v_0, \boldsymbol{S}_0).$$

Find the full conditionals

- Let's derive the full conditionals so we can do Gibbs sampling
- The likelihood is

$$L(\boldsymbol{\theta}, \Sigma; \mathbf{y}_1, \ldots, \mathbf{y}_n) = p(\mathbf{y}_1, \ldots, \mathbf{y}_n | \boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-n/2} e^{-\frac{1}{2}\sum_i (\mathbf{y}_i - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta})}.$$

- The prior is given by

$$p(\boldsymbol{\theta}) \propto e^{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_0)' \Lambda_0^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_0)}$$

and

$$p(\Sigma) \propto |\Sigma|^{-\frac{\nu_0 + p + 1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0 \Sigma^{-1})}$$

The full joint probability

- The full joint probability is then

$$
\begin{aligned}
p(\mathbf{y}_1, \ldots, \mathbf{y}_n, \boldsymbol{\theta}, \Sigma) \propto & |\Sigma|^{-n/2} e^{-\frac{1}{2}\sum_i (\mathbf{y}_i - \boldsymbol{\theta})' \Sigma^{-1} (\mathbf{y}_i - \boldsymbol{\theta})} \times \\
& e^{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_0)' \Lambda_0^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_0)} \times \\
& |\Sigma|^{-\frac{\nu_0 + p + 1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0 \Sigma^{-1})}
\end{aligned}
$$

The full conditional of $\boldsymbol{\theta}$

- Viewing the joint probability as a function in $\boldsymbol{\theta}$, we have

$$p(\boldsymbol{\theta}|\mathbf{y}_1,\ldots,\mathbf{y}_n,\Sigma) \propto e^{-\frac{1}{2}\sum_i(\mathbf{y}_i-\boldsymbol{\theta})'\Sigma^{-1}(\mathbf{y}_i-\boldsymbol{\theta})-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_0)'\Lambda_0^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_0)}$$
$$\propto e^{-\frac{1}{2}\left[n\boldsymbol{\theta}'\Sigma^{-1}\boldsymbol{\theta}-2n\bar{y}'\Sigma^{-1}\boldsymbol{\theta}+\boldsymbol{\theta}'\Lambda_0^{-1}\boldsymbol{\theta}+2\boldsymbol{\mu}_0'\Lambda_0^{-1}\boldsymbol{\theta}\right]}$$
$$\propto e^{-\frac{1}{2}\left[\boldsymbol{\theta}'(n\Sigma^{-1}+\Lambda_0^{-1})\boldsymbol{\theta}-2(n\bar{y}'\Sigma^{-1}+\boldsymbol{\mu}_0'\Lambda_0^{-1})\boldsymbol{\theta}\right]}$$

- Now let Λ_n be such that

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

  and

$$\boldsymbol{\mu}_n = \Lambda_n\left(n\Sigma^{-1}\bar{y} + \Lambda_0^{-1}\boldsymbol{\mu}_0\right).$$

- Then by completion of squares we have

$$p(\boldsymbol{\theta}|\mathbf{y}_1,\ldots,\mathbf{y}_n,\Sigma) \propto e^{-\frac{1}{2}\left(\boldsymbol{\theta}'\Lambda_n^{-1}\boldsymbol{\theta}-2\boldsymbol{\mu}_n'\Lambda_n^{-1}\boldsymbol{\theta}\right)}$$
$$\propto e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu}_n)'\Lambda_n^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}_n)},$$

  which is exactly a $\mathrm{N}(\boldsymbol{\mu}_n, \Lambda_n)$.
- Compare this with univariate case — similar interpretation.

Full conditional of Σ

- Viewing the joint probability as a function in Σ, we have

$$p(\Sigma|\mathbf{y}_1,\ldots,\mathbf{y}_n,\boldsymbol{\theta}) \propto |\Sigma|^{-(n+\nu_0+p+1)/2} e^{-\frac{1}{2}\left[\sum_i (\mathbf{y}_i-\boldsymbol{\theta})'\Sigma^{-1}(\mathbf{y}_i-\boldsymbol{\theta})+\mathrm{tr}(\boldsymbol{S}_0\Sigma^{-1})\right]}.$$

- Two useful properties of $\mathrm{tr}(\cdot)$:

  (i) For two matrix $A_{k\times m}$ and $B_{m\times k}$, we can prove (by definition of trace) that

  $$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

  Note that AB is $k\times k$ and BA is $m\times m$.

  (ii) Also by definition of trace, we have for two square matrix A and B,

  $$\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B).$$

Full conditional of Σ

- Now note that $(\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\theta})$ is a scalar (i.e., a 1×1), and so it is equal to its trace

$$(\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\theta}) = \mathrm{tr}\left((\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}\right)$$

  by Property (i).
- Then by Property (ii)

$$\begin{aligned}\sum_i (\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}(\mathbf{y}_i - \boldsymbol{\theta}) &= \sum_i \mathrm{tr}\left((\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}\right) \\ &= \mathrm{tr}\left(\sum_i (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})'\Sigma^{-1}\right) \\ &= \mathrm{tr}\left(\boldsymbol{S}_{\boldsymbol{\theta}}\Sigma^{-1}\right)\end{aligned}$$

  where

$$\boldsymbol{S}_{\boldsymbol{\theta}} = \sum_i (\mathbf{y}_i - \boldsymbol{\theta})(\mathbf{y}_i - \boldsymbol{\theta})',$$

  which is the *residual sum of squared matrix* of the observations.

Full conditional of Σ

- By Property (ii) again,

$$\mathrm{tr}\left(\boldsymbol{S}_{\boldsymbol{\theta}}\Sigma^{-1}\right)+\mathrm{tr}\left(\boldsymbol{S}_0\Sigma^{-1}\right)=\mathrm{tr}\left(\left(\boldsymbol{S}_{\boldsymbol{\theta}}+\boldsymbol{S}_0\right)\Sigma^{-1}\right)$$

- Therefore, putting all pieces together, we have

$$p(\Sigma|\mathbf{y}_1,\ldots,\mathbf{y}_n,\boldsymbol{\theta})\propto|\Sigma|^{-(n+\nu_0+p+1)/2}e^{-\frac{1}{2}\mathrm{tr}\left((\boldsymbol{S}_{\boldsymbol{\theta}}+\boldsymbol{S}_0)\Sigma^{-1}\right)},$$

  which we recognize as again an inverse-Wishart$(\nu_n,\boldsymbol{S}_0+\boldsymbol{S}_{\boldsymbol{\theta}})$ distribution where $\nu_n=\nu_0+n$.
- That is

$$\Sigma|\mathbf{y}_1,\ldots,\mathbf{y}_n,\boldsymbol{\theta}\sim\mathrm{IW}(\nu_n,\boldsymbol{S}_0+\boldsymbol{S}_{\boldsymbol{\theta}}).$$

- ν_0 is the prior degrees of freedom and $\boldsymbol{S}_0$ the prior sum of squares matrix.
- The full conditional expectation of Σ is

$$\begin{aligned}\mathrm{E}(\Sigma|\mathbf{y}_1,\ldots,\mathbf{y}_n,\boldsymbol{\theta})&=\frac{\boldsymbol{S}_0+\boldsymbol{S}_{\boldsymbol{\theta}}}{\nu_n-p-1}\\&=\frac{\nu_0-p-1}{\nu_n-p-1}\cdot\underbrace{\frac{\boldsymbol{S}_0}{\nu_0-p-1}}_{\text{prior mean of }\Sigma}+\frac{n}{\nu_n-p-1}\cdot\underbrace{\frac{\boldsymbol{S}_{\boldsymbol{\theta}}}{n}}_{\text{Empirical estimate of }\Sigma}.\end{aligned}$$

Gibbs sampling

- Initialize the chain at $(\boldsymbol{\theta}^{(0)}, \Sigma^{(0)})$.
- For $t = 1, 2, \ldots,$
  - Update $\boldsymbol{\theta}$:
    - Compute

$$\left(\Lambda_n^{(t)}\right)^{-1} = n\left(\Sigma^{(t-1)}\right)^{-1} + \Lambda_0^{-1},$$
$$\boldsymbol{\mu}_n^{(t)} = \Lambda_n^{(t)}\left(n\left(\Sigma^{(t-1)}\right)^{-1}\bar{y} + \Lambda_0^{-1}\boldsymbol{\mu}_0\right)$$

    - Draw

$$\boldsymbol{\theta}^{(t)} \sim \mathrm{N}(\boldsymbol{\mu}_n^{(t)}, \Lambda_n^{(t)})$$

  - Update Σ:
    - Compute

$$\boldsymbol{S}_{\boldsymbol{\theta}^{(t)}} = \sum_i (\mathbf{y}_i - \boldsymbol{\theta}^{(t)})(\mathbf{y}_i - \boldsymbol{\theta}^{(t)})'.$$

    - Draw

$$\Sigma^{(t)} \sim \mathrm{IW}\left(\nu_n, \boldsymbol{S}_0 + \boldsymbol{S}_{\boldsymbol{\theta}^{(t)}}\right).$$

- Discard suitable burn-in steps.

Back to air pollutant example

- Suppose now we measure two pollutants (e.g., PM2.5 and SO2) concurrently 16 times on a day, so our data are bivariate

$$(104,100),(105,102),(103,101),(102,104),(105,108),(107,108),$$
$$(106,103),(104,104),(103,106),(106,107),(105,105),(102,101),$$
$$(102,100),(108,106),(105,105),(104,105)$$

Prior specification

- Suppose based on historical data, both pollutants have average level of 100. So we set

$$\boldsymbol{\mu}_0 = (100, 100)'.$$

- Suppose the historical distribution of PM2.5 has a standard deviation of about 10, and SO2 has a standard deviation of about 5, and their values tend to rise or fall together with a correlation of about 0.3. Turning this into our prior,

$$\Lambda_0 = \begin{pmatrix} 10^2 & 0.3 \times 10 \times 5 \\ 0.3 \times 10 \times 5 & 5^2 \end{pmatrix} = \begin{pmatrix} 100 & 15 \\ 15 & 25 \end{pmatrix}.$$

Prior specification on Σ

- Suppose we have very little idea about the value of Σ, so we want to choose a weak prior.
- This can be achieved by setting ν_0 to a small value, such as $\nu_0 = p+2$. (Note that the inverse-Wishart prior requires $\nu_0 > p+1$ to have a finite mean.)
- As for the prior mean $\boldsymbol{S}_0$, (e.g., $\boldsymbol{S}_0/(\nu_0 - p - 1)$), for simplicity, we use a diagonal matrix

$$\boldsymbol{S}_0 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

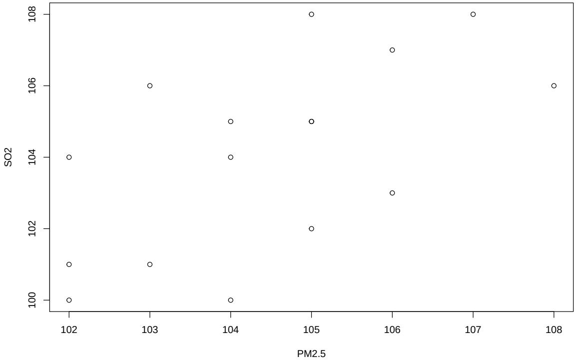
  where the diagonal is chosen to be 4, corresponding to marginal standard deviations of 2 for the device readings.
- The off-diagonals are set to 0 because we don't really know anything about how the device reading errors are correlated for the two pollutants.

Reading the data

```
library(mvtnorm) # for drawing multivariate normal
library(MCMCpack) # for drawing inverse-Wishart
library(coda) # for MCMC diagnostics
options(digits=4) # tidy up print outs

y <- matrix(c(104,100,105,102,103,101,102,104,105,108,107,108,106,103,104,104,
              103,106,106,107,105,105,102,101,102,100,108,106,105,105,104,105),
            ncol=2,byrow=TRUE)
n <- nrow(y) # sample size
p <- ncol(y) # dimensionality

plot(y[,1],y[,2],xlab="PM2.5",ylab="SO2")
```



Prior specification

```r
# prior for theta
mu.0 <- c(100,100)
Lambda.0 = matrix(c(100,15,15,25),ncol=2,byrow=TRUE)

# prior for Sigma
nu.0 <- p + 2  # a very weak prior
S0 <- (nu.0-p-1)*matrix(c(4,0,0,4),ncol=2,byrow=TRUE)

ybar <- apply(y,2,mean)
nu.n <- nu.0 + n
```

Gibbs sampling

```
niter <- 10000 # total number of iterations
nburnin <- 1000 # 1000 burn-in steps

THETA <- matrix(NA,nrow=niter,ncol=p) # matrix for storing the draws for theta
colnames(THETA) <- c("theta1","theta2")

THETA.init <- ybar # Initial values set to sample mean
THETA.curr <- THETA.init # the theta value at current iteration

SIGMA <- matrix(NA,nrow=niter,ncol=p*p) # matrix for storing the draws for Sigma
colnames(SIGMA) <- c("sigma11","sigma12","sigma21","sigma22")

SIGMA.init <- cov(y) # intial value set to sample covariance
SIGMA.curr <- SIGMA.init # the Sigma value at current iternation

### Start Gibbs sampling
for (t in 1:niter) {

  ## Update theta
  Lambda.n <- solve(n*solve(SIGMA.curr)+solve(Lambda.0))
  mu.n <- Lambda.n %*% (n*solve(SIGMA.curr,ybar)+solve(Lambda.0,mu.0))
  THETA.curr <- rmvnorm(1,mean=mu.n,sigma=Lambda.n)

  ## Update Sigma
  S.theta <- (t(y)-c(THETA.curr))%*% t(t(y)-c(THETA.curr))
  SIGMA.curr <- riwish(v=nu.n,S=S0+S.theta)

  ## Save the current iteration
  THETA[t,] <- THETA.curr
  SIGMA[t,] <- SIGMA.curr

}
```

Histogram of MCMC draws for $\boldsymbol{\theta}$

```r
ggplot(data.frame(THETA), aes(x=theta1, y=theta2) ) +
  labs(x=expression(theta_1),y=expression(theta_2)) +
  geom_bin2d(bins=70) +
  scale_fill_continuous(type = "viridis") +
  lims(x=c(101,107),y=c(101,107))
```

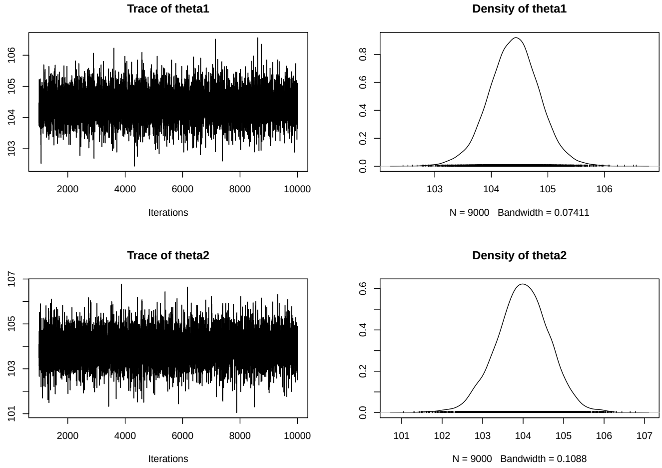
MCMC diagnostics

```
THETA.mcmc <- mcmc(THETA[-(1:nburnin),],start=nburnin+1)
summary(THETA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##        Mean    SD Naive SE Time-series SE
## theta1  104 0.448  0.00472        0.00472
## theta2  104 0.658  0.00694        0.00694
##
## 2. Quantiles for each variable:
##
##        2.5% 25% 50% 75% 97.5%
## theta1  104 104 104 105   105
## theta2  103 104 104 104   105
```

Trace plots for $\boldsymbol{\theta}$

```r
plot(THETA.mcmc)
```



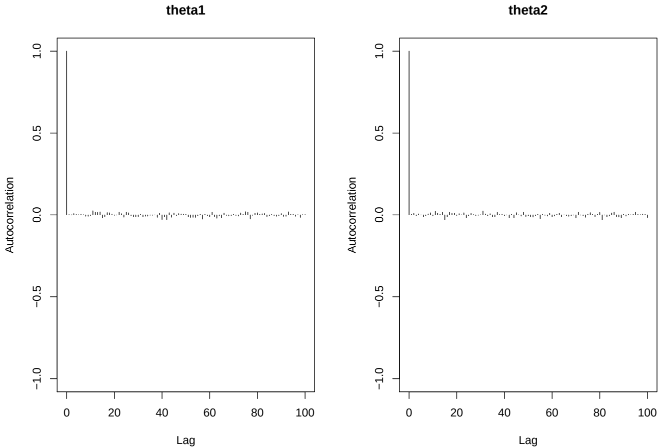
ESS for $\boldsymbol{\theta}$

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

```
## theta1 theta2
##   9000   9000
```

Autocorrelation plot for $\boldsymbol{\theta}$

```
autocorr.plot(THETA.mcmc,lag.max=100)
```



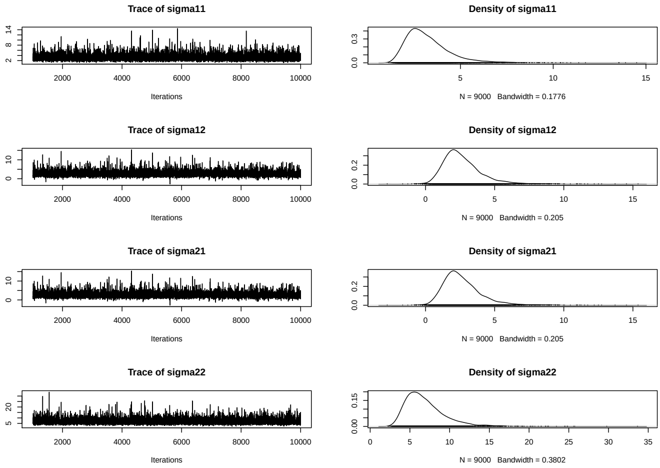
MCMC diagnostics

```r
SIGMA.mcmc <- mcmc(SIGMA[-(1:nburnin),],start=nburnin+1)
summary(SIGMA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##         Mean   SD Naive SE Time-series SE
## sigma11 3.23 1.21   0.0128         0.0128
## sigma12 2.65 1.42   0.0150         0.0157
## sigma21 2.65 1.42   0.0150         0.0157
## sigma22 6.92 2.59   0.0273         0.0291
##
## 2. Quantiles for each variable:
##
##          2.5%  25%  50%  75% 97.5%
## sigma11 1.644 2.40 2.99 3.79  6.33
## sigma12 0.635 1.71 2.39 3.31  6.07
## sigma21 0.635 1.71 2.39 3.31  6.07
## sigma22 3.553 5.13 6.38 8.10 13.34
```

Trace plots for Σ

```r
plot(SIGMA.mcmc)
```



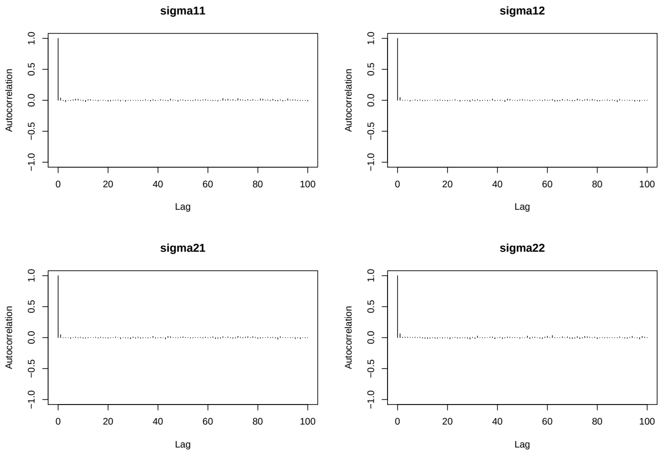
ESS for Σ

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

```
## sigma11 sigma12 sigma21 sigma22
##    8916    8200    8200    7948
```

Autocorrelation plot for Σ

```
autocorr.plot(SIGMA.mcmc,lag.max=100)
```



With a stronger prior on Σ with $v_0 = 50$

- Suppose we have a better idea about the value of Σ, so we want to choose a stronger prior.
- For example we set $v_0 = 50$.
- Now the prior mean $\boldsymbol{S}_0/(v_0 - p - 1) = \boldsymbol{S}_0/47$. Thus if we want the same prior mean as before, we set

$$\boldsymbol{S}_0 = 47 \cdot \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = \cdot \begin{pmatrix} 188 & 0 \\ 0 & 188 \end{pmatrix}$$

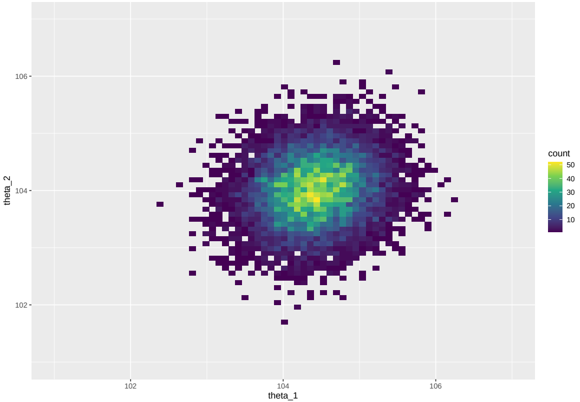
```
# prior for theta
mu.0 <- c(100,100)
Lambda.0 = matrix(c(100,15,15,25),ncol=2,byrow=TRUE)

# prior for Sigma
nu.0 <- 50  # a stronger prior on covariance now
S0 <- (nu.0-p-1) * matrix(c(4,0,0,4),ncol=2,byrow=TRUE) # maintaining the same prior mean

ybar <- apply(y,2,mean)
nu.n <- nu.0 + n
```

Histogram of MCMC draws for $\boldsymbol{\theta}$

```
ggplot(data.frame(THETA), aes(x=theta1, y=theta2) ) +
  labs(x=expression(theta_1),y=expression(theta_2)) +
  geom_bin2d(bins=70) +
  scale_fill_continuous(type = "viridis") +
  lims(x=c(101,107),y=c(101,107))
```



- How does this compare to the case with $\nu_0 = 4$?
- Why is the posterior spread of $\boldsymbol{\theta}$ smaller?

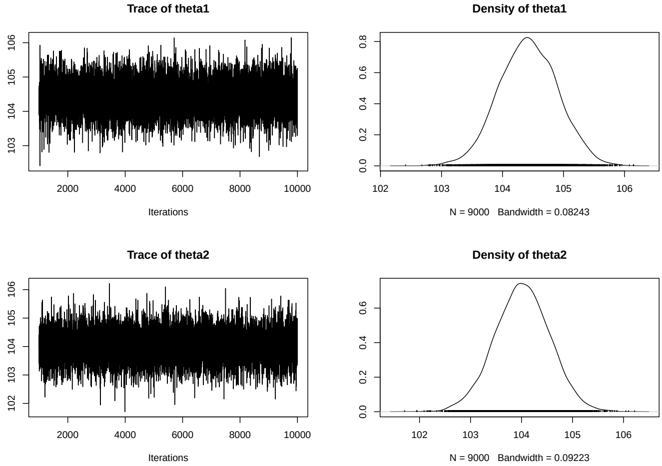
MCMC diagnostics

```r
THETA.mcmc <- mcmc(THETA[-(1:nburnin),],start=nburnin+1)
summary(THETA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##        Mean    SD Naive SE Time-series SE
## theta1  104 0.480  0.00506        0.00506
## theta2  104 0.538  0.00568        0.00568
##
## 2. Quantiles for each variable:
##
##        2.5% 25% 50% 75% 97.5%
## theta1  103 104 104 105   105
## theta2  103 104 104 104   105
```

Trace plots for $\boldsymbol{\theta}$

```r
plot(THETA.mcmc)
```



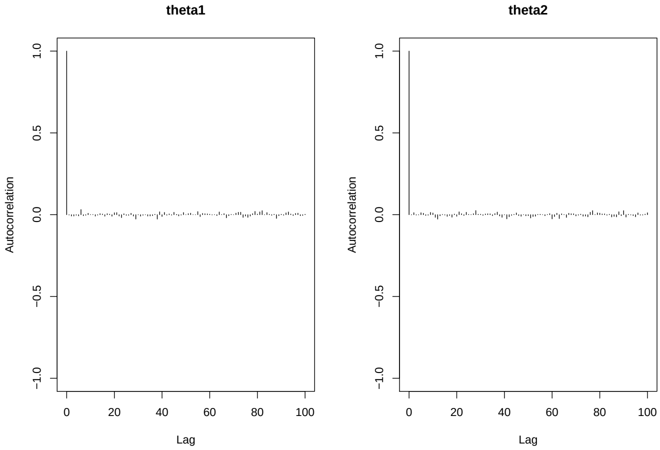
ESS for $\boldsymbol{\theta}$

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

```
## theta1 theta2
##   9000   9000
```

Autocorrelation plot for $\boldsymbol{\theta}$

```
autocorr.plot(THETA.mcmc,lag.max=100)
```



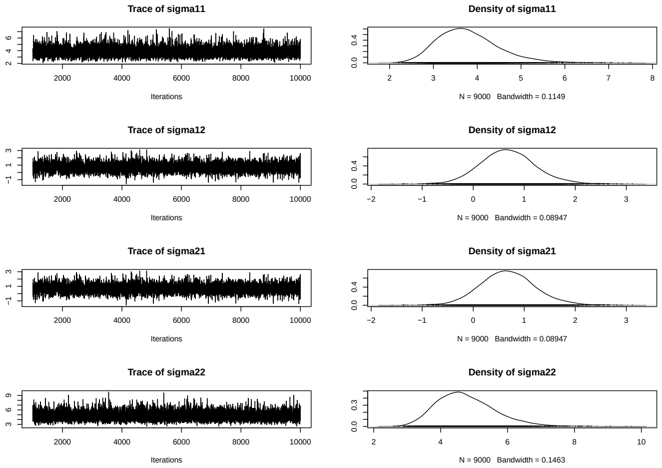
MCMC diagnostics

```r
SIGMA.mcmc <- mcmc(SIGMA[-(1:nburnin),],start=nburnin+1)
summary(SIGMA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##          Mean    SD Naive SE Time-series SE
## sigma11 3.798 0.700  0.00738        0.00738
## sigma12 0.685 0.546  0.00576        0.00576
## sigma21 0.685 0.546  0.00576        0.00576
## sigma22 4.760 0.872  0.00919        0.00919
##
## 2. Quantiles for each variable:
##
##           2.5%   25%   50%  75% 97.5%
## sigma11  2.661 3.300 3.724 4.20  5.39
## sigma12 -0.337 0.324 0.668 1.02  1.82
## sigma21 -0.337 0.324 0.668 1.02  1.82
## sigma22  3.323 4.138 4.658 5.28  6.71
```

Trace plots for Σ

```
plot(SIGMA.mcmc)
```



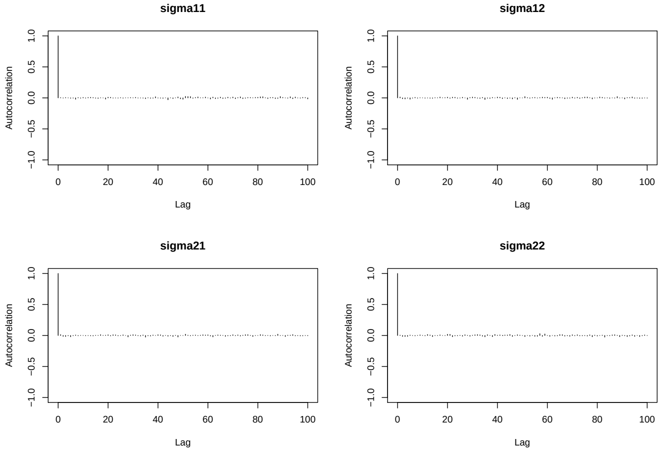
ESS for Σ

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

```
## sigma11 sigma12 sigma21 sigma22
##    9000    9000    9000    9000
```

Autocorrelation plot for Σ

```
autocorr.plot(SIGMA.mcmc,lag.max=100)
```



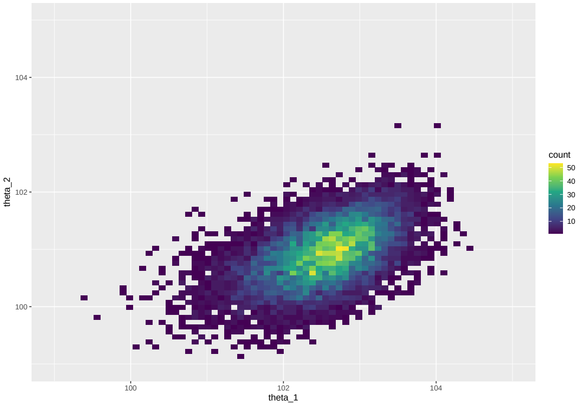
Stronger prior on $\boldsymbol{\theta}$

- Suppose the historical distribution of PM2.5 has a standard deviation of about *1* (vs 10 before), and SO2 has a standard deviation of about *0.5* vs (5 before), still with a correlation of about 0.3. Turning this into our prior,

$$\Lambda_0 = \begin{pmatrix} 1^2 & 0.3 \times 1 \times 0.5 \\ 0.3 \times 1 \times 0.5 & 0.5^2 \end{pmatrix} = \begin{pmatrix} 1 & 0.15 \\ 0.15 & 0.25 \end{pmatrix}.$$

With weak prior on Σ: $\nu_0 = p + 2$

```
ggplot(data.frame(THETA), aes(x=theta1, y=theta2) ) +
  labs(x=expression(theta_1),y=expression(theta_2)) +
  geom_bin2d(bins=70) +
  scale_fill_continuous(type = "viridis") +
  lims(x=c(99,105),y=c(99,105))
```



- Why is there a lot *more* shrinkage toward prior mean?

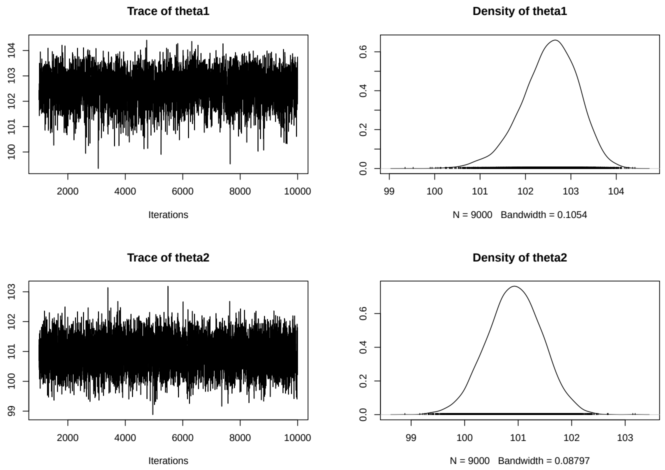
MCMC diagnostics

```r
THETA.mcmc <- mcmc(THETA[-(1:nburnin),],start=nburnin+1)
summary(THETA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##        Mean    SD Naive SE Time-series SE
## theta1  103 0.618  0.00651        0.01197
## theta2  101 0.513  0.00540        0.00707
##
## 2. Quantiles for each variable:
##
##         2.5% 25% 50% 75% 97.5%
## theta1 101.2 102 103 103   104
## theta2  99.9 101 101 101   102
```

Trace plots for $\boldsymbol{\theta}$

```
plot(THETA.mcmc)
```



ESS for $\boldsymbol{\theta}$

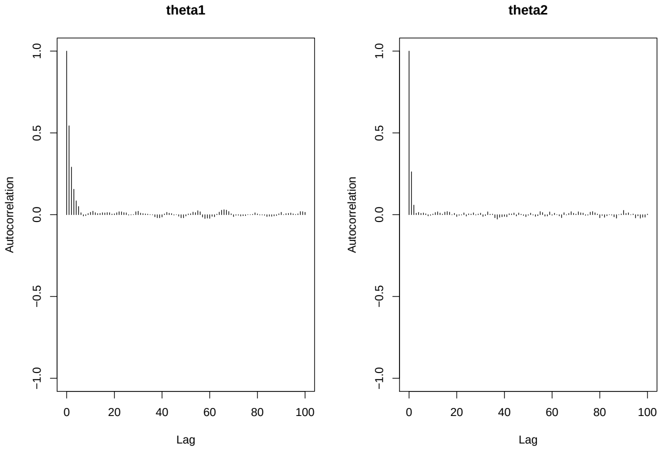
```r
effectiveSize(THETA.mcmc)
```

```
## theta1 theta2
##   2662   5259
```

- This stronger prior induced much stronger autocorrelation and smaller ESS.
- Why?
  - The prior concentrated near a value far from the empirical estimate.
  - Stronger posterior dependency between $\boldsymbol{\theta}$ and Σ.

Autocorrelation plot for $\boldsymbol{\theta}$

```r
autocorr.plot(THETA.mcmc,lag.max=100)
```



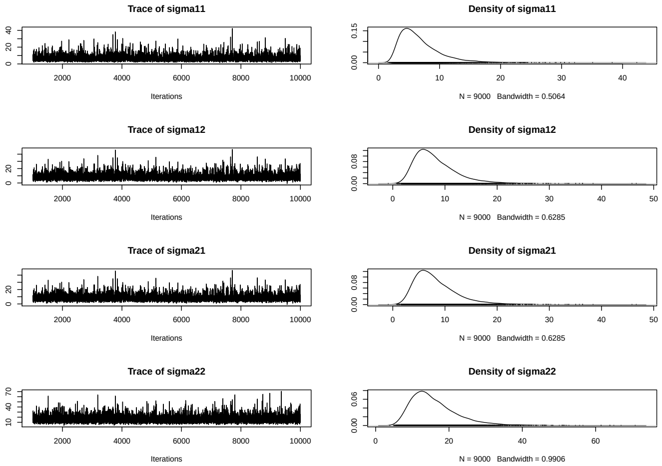
MCMC diagnostics

```
SIGMA.mcmc <- mcmc(SIGMA[-(1:nburnin),],start=nburnin+1)
summary(SIGMA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##          Mean   SD Naive SE Time-series SE
## sigma11  6.80 3.55   0.0374         0.0626
## sigma12  8.24 4.24   0.0447         0.0645
## sigma21  8.24 4.24   0.0447         0.0645
## sigma22 16.00 6.63   0.0699         0.0842
##
## 2. Quantiles for each variable:
##
##         2.5%   25%   50%  75% 97.5%
## sigma11 2.63  4.35  5.93  8.3  15.9
## sigma12 2.73  5.30  7.32 10.2  18.9
## sigma21 2.73  5.30  7.32 10.2  18.9
## sigma22 7.33 11.38 14.59 19.1  32.3
```

Trace plots for Σ

```r
plot(SIGMA.mcmc)
```



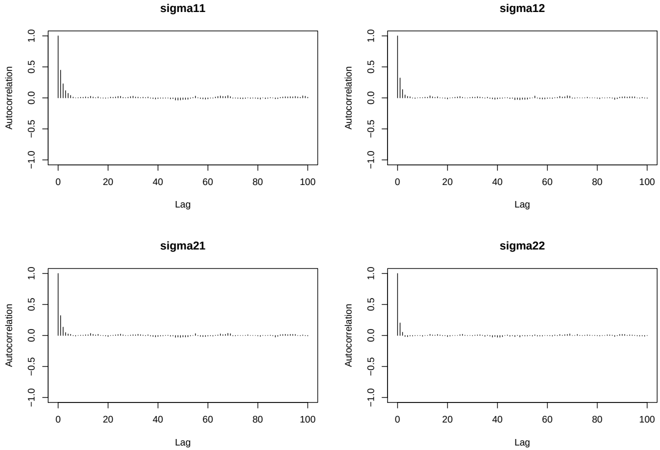
ESS for Σ

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

```
## sigma11 sigma12 sigma21 sigma22
##    3212    4310    4310    6200
```

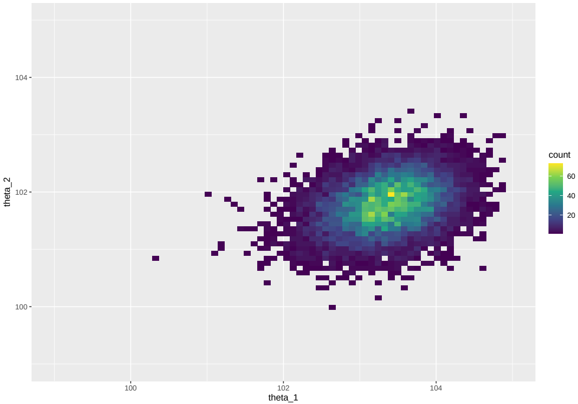
Autocorrelation plot for Σ

```
autocorr.plot(SIGMA.mcmc,lag.max=100)
```



With a strong prior on Σ: $\nu_0 = 50$

```r
ggplot(data.frame(THETA), aes(x=theta1, y=theta2) ) +
  labs(x=expression(theta_1),y=expression(theta_2)) +
  geom_bin2d(bins=70) +
  scale_fill_continuous(type = "viridis") +
  lims(x=c(99,105),y=c(99,105))
```



- Why is there a lot *less* shrinkage toward prior mean?

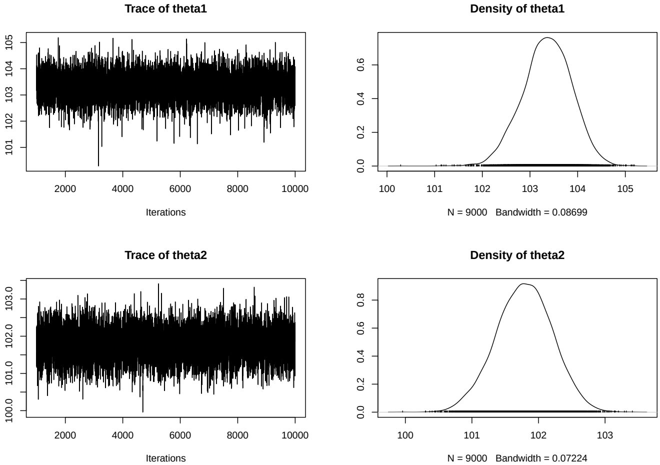
MCMC diagnostics

```r
THETA.mcmc <- mcmc(THETA[-(1:nburnin),],start=nburnin+1)
summary(THETA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##        Mean    SD Naive SE Time-series SE
## theta1  103 0.513  0.00540        0.00681
## theta2  102 0.421  0.00444        0.00537
##
## 2. Quantiles for each variable:
##
##        2.5% 25% 50% 75% 97.5%
## theta1  102 103 103 104   104
## theta2  101 102 102 102   103
```

Trace plots for $\boldsymbol{\theta}$

```r
plot(THETA.mcmc)
```



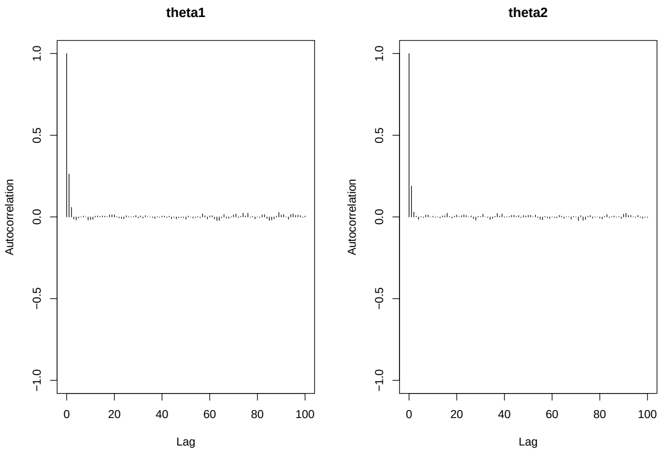
ESS for $\boldsymbol{\theta}$

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

```
## theta1 theta2
##   5673   6139
```

Autocorrelation plot for $\boldsymbol{\theta}$

```
autocorr.plot(THETA.mcmc,lag.max=100)
```



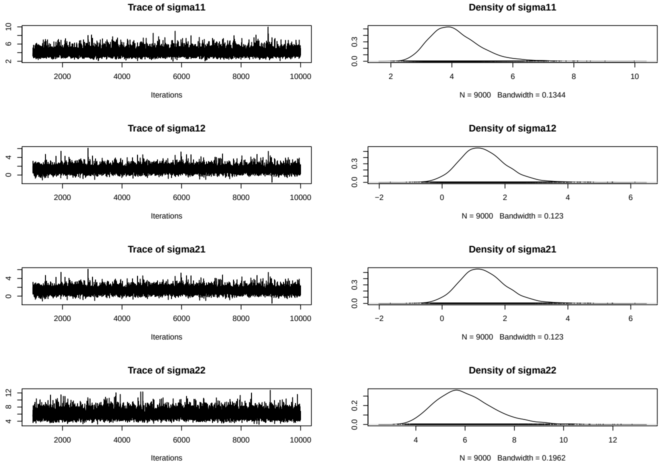
MCMC diagnostics

```
SIGMA.mcmc <- mcmc(SIGMA[-(1:nburnin),],start=nburnin+1)
summary(SIGMA.mcmc)
```

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##          Mean    SD Naive SE Time-series SE
## sigma11 4.12 0.827  0.00872         0.0101
## sigma12 1.32 0.757  0.00798         0.0101
## sigma21 1.32 0.757  0.00798         0.0101
## sigma22 6.03 1.200  0.01265         0.0148
##
## 2. Quantiles for each variable:
##
##             2.5%   25%  50%  75% 97.5%
## sigma11  2.8141 3.534 4.01 4.58  6.06
## sigma12 -0.0119 0.803 1.26 1.76  2.95
## sigma21 -0.0119 0.803 1.26 1.76  2.95
## sigma22  4.1017 5.188 5.88 6.72  8.76
```

Trace plots for Σ

```r
plot(SIGMA.mcmc)
```



ESS for Σ

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

```
## sigma11 sigma12 sigma21 sigma22
##    6675    5604    5604    6590
```

Autocorrelation plot for Σ

```
autocorr.plot(SIGMA.mcmc,lag.max=100)
```