MCMC and Dependence in the Bayesian Linear Model

Multivariate Normal example: Suppose that

$$\mathbf{X} \sim Normal_d(\mathbf{0}, \Sigma)$$

where $d = 20$ and

$$\Sigma = \sigma_s^2 \mathbf{I}_d + \sigma_c^2 \mathbf{1}_d \mathbf{1}_d^\top$$

where $\mathbf{I}_d$ is the $d \times d$ identity matrix, and $\mathbf{1}_d$ is the $d \times 1$ vector of 1s.

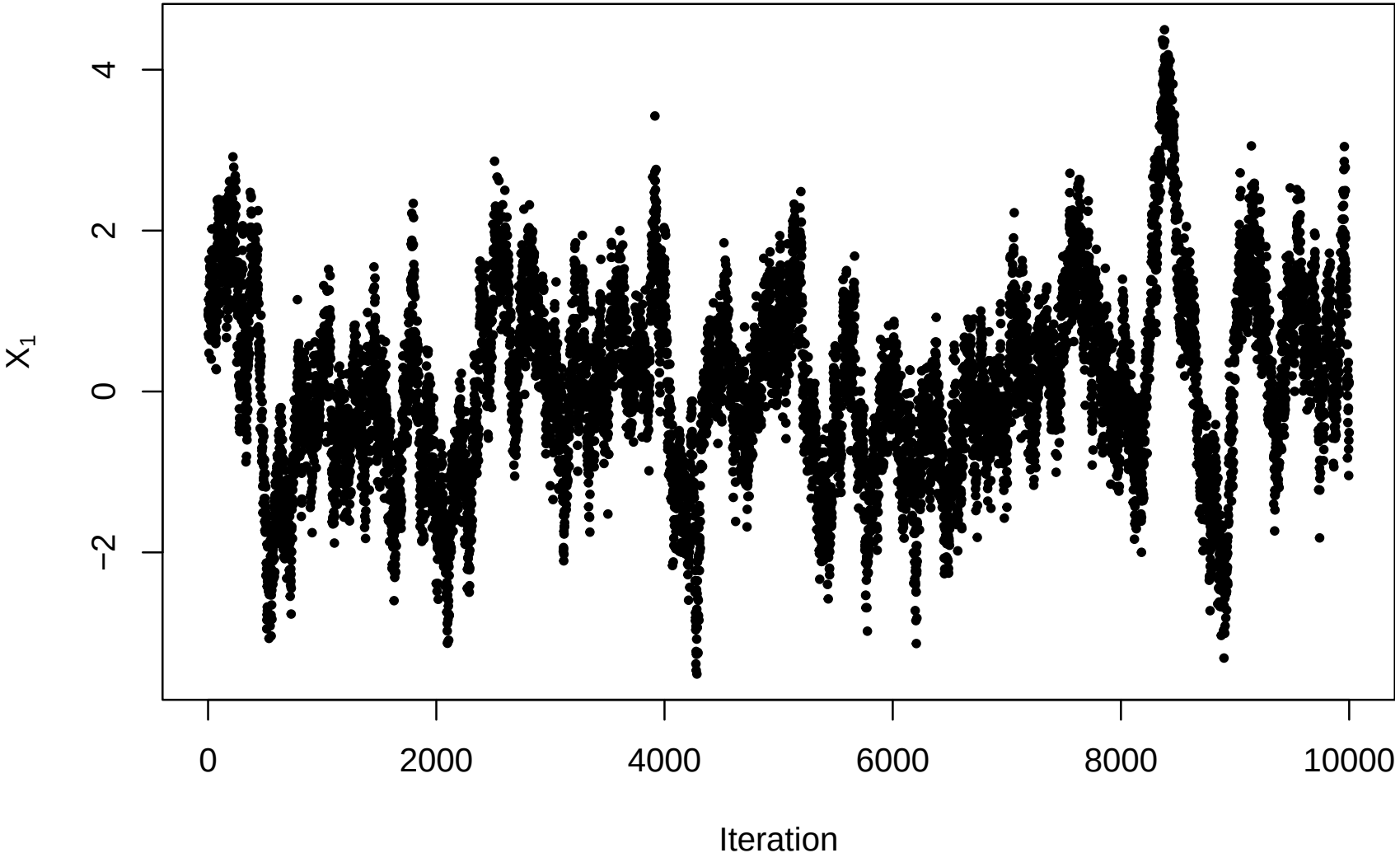
We have

$$\text{Corr}[X_j, X_k] = \frac{\sigma_c^2}{\sigma_s^2 + \sigma_c^2} \qquad j \neq k$$

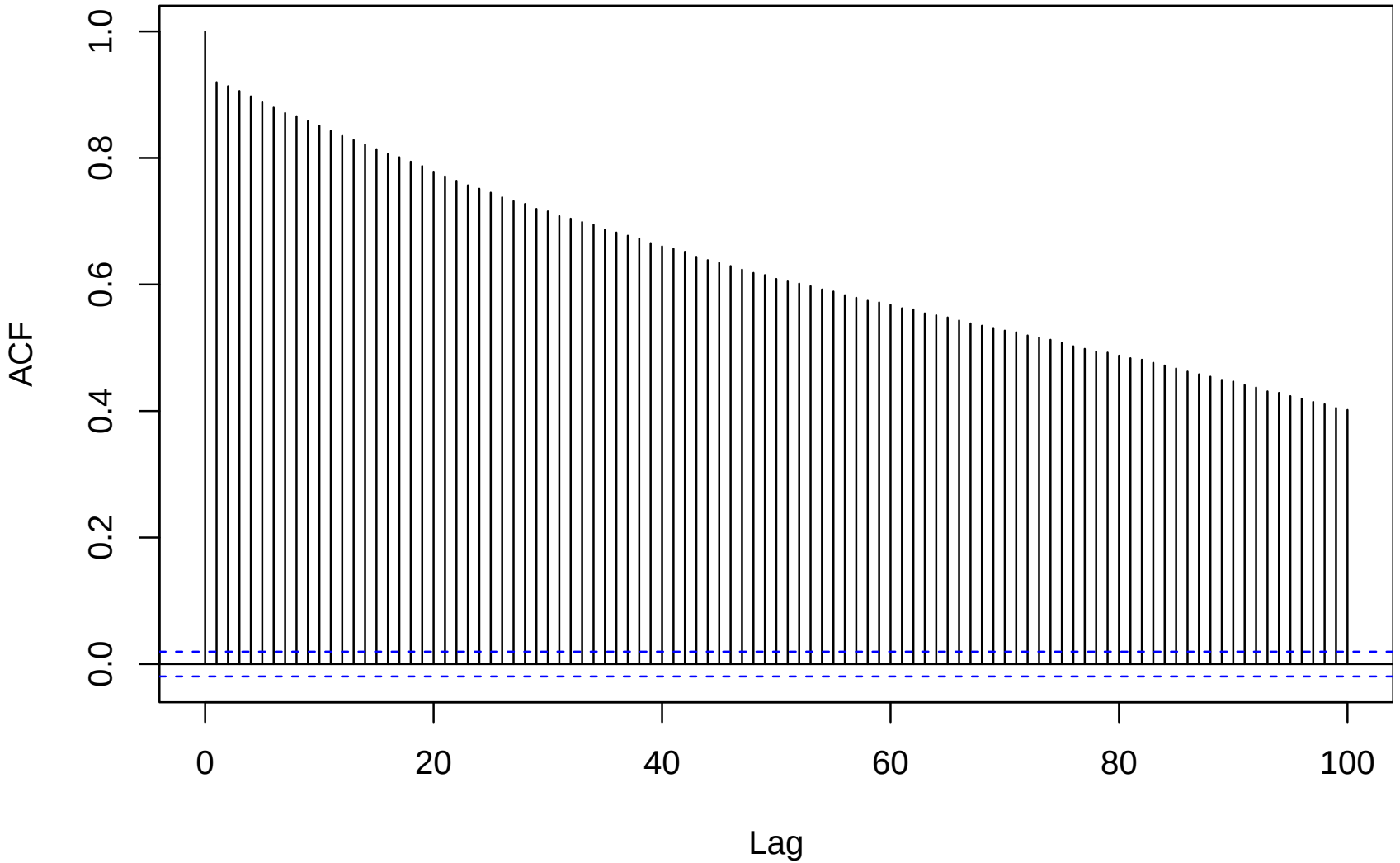
If $\sigma_s^2 = 0.1$ and $\sigma_c^2 = 1$, we have that $\text{Corr}[X_j, X_k] = 0.9090909$. We run the Gibbs sampler to collect a sample of 10000 points, taking every draw from the sampler, after a burn in of 100 iterations.

MCMC and Dependence in the Bayesian Linear Model

We examine the output for X_1 (as all variables will behave similarly).



MCMC and Dependence in the Bayesian Linear Model



There is clear evidence of significant autocorrelation. The effective sample sizes confirm that the efficiency of the sampler is quite poor.

```
(ESS<-effectiveSize(Xsamp))

+     var1     var2     var3     var4     var5     var6     var7     var8
+ 48.86716 52.39946 53.79902 50.00617 49.96306 52.16385 51.70826 51.27096
+     var9    var10    var11    var12    var13    var14    var15    var16
+ 48.42472 53.29529 53.12614 49.47482 50.13396 50.44028 52.21867 50.55783
+    var17    var18    var19    var20
+ 50.11824 51.17478 49.78654 51.30757
```

The efficiency is approximately 0.49 per cent of a procedure that takes independent draws from the joint distribution.

We can alter the procedure to incorporate blocks of parameters. Suppose we choose to block the 20 parameters into four blocks of five,

- Block 1: variables 1 to 5
- Block 2: variables 6 to 10
- Block 3: variables 11 to 15
- Block 4: variables 16 to 20

and use the Gibbs sampler on the blocks.

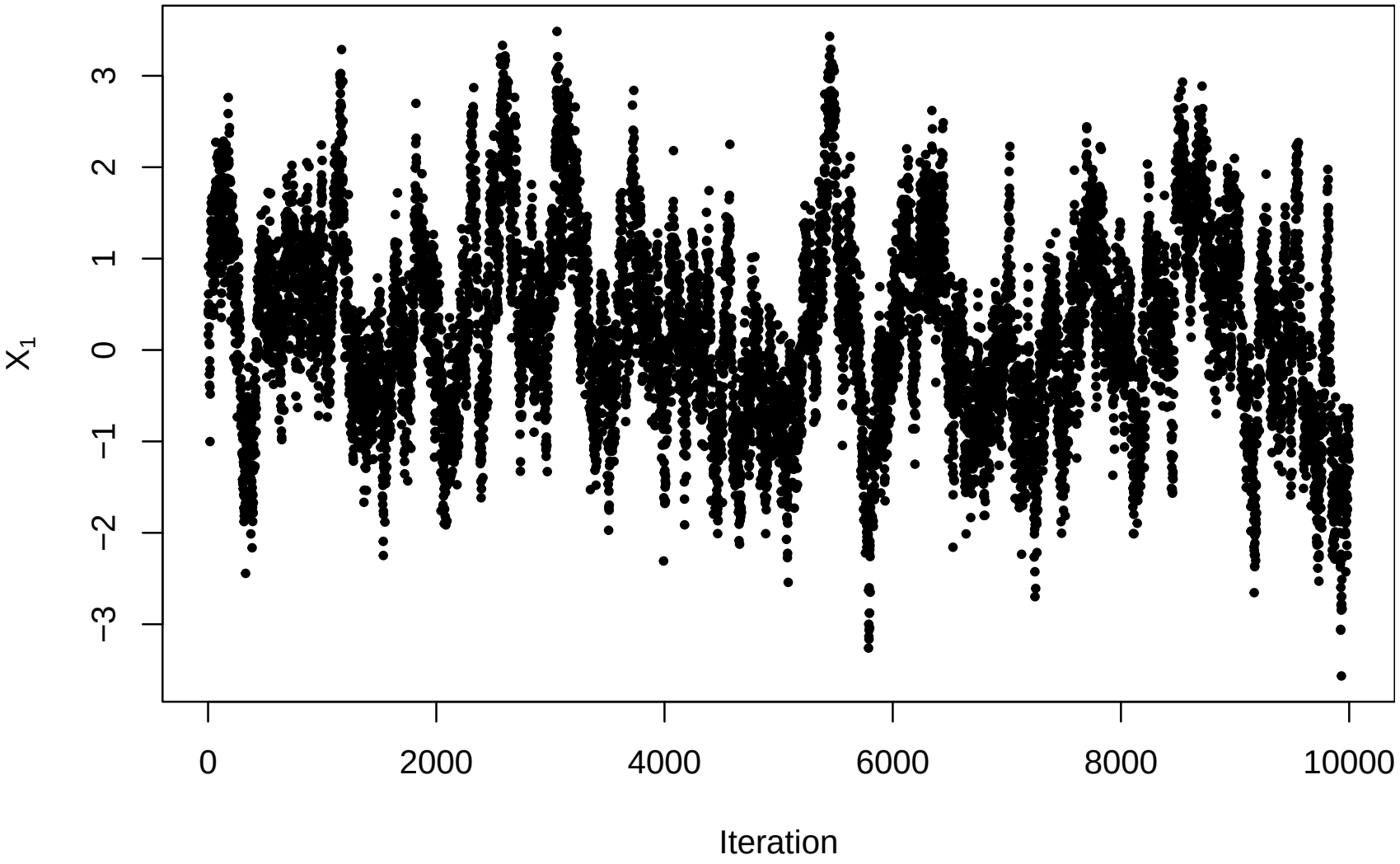
All the (block) full conditionals are also Normal.

The effective sample sizes confirm that the efficiency has improved slightly.

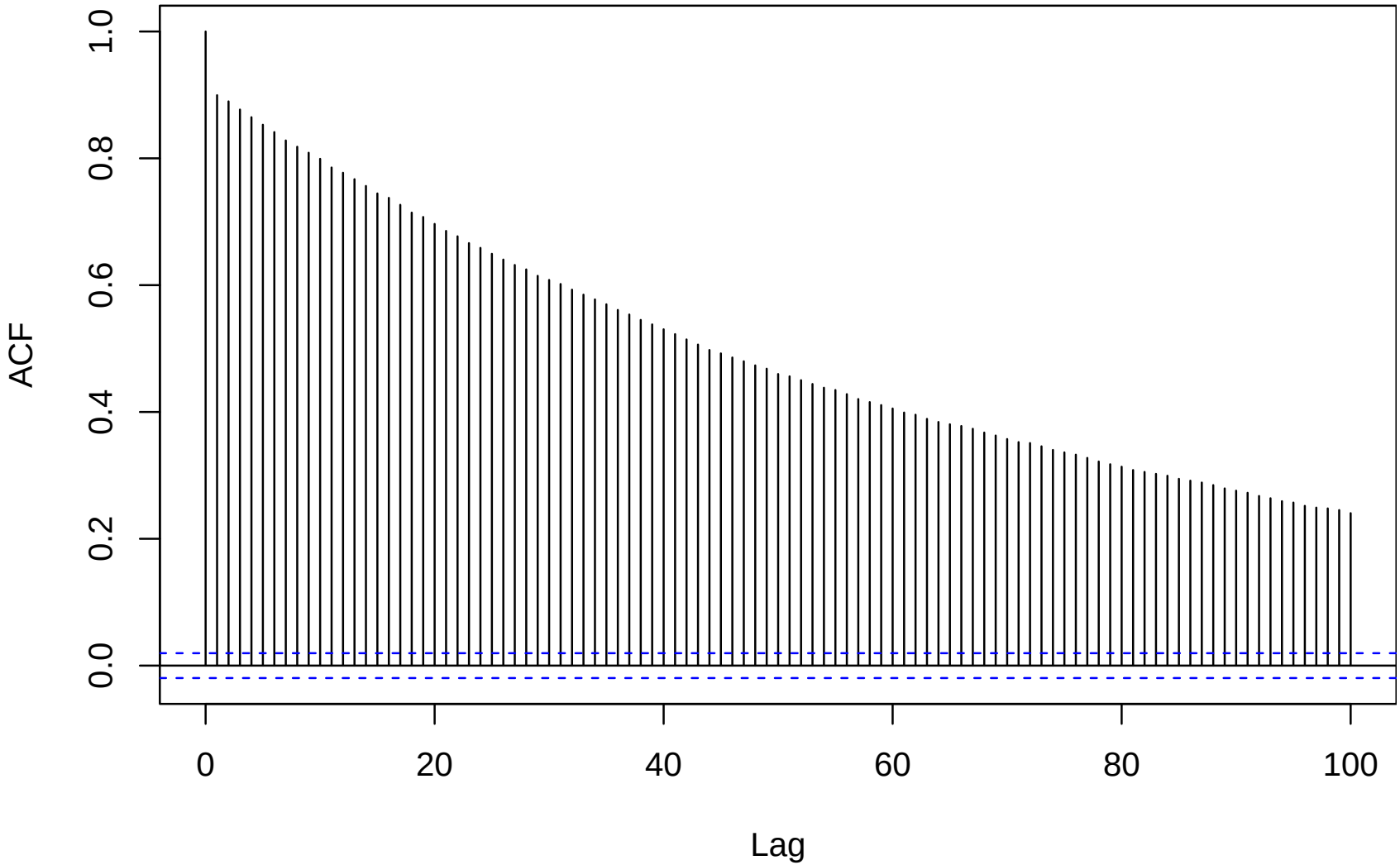
```
+     var1     var2     var3     var4     var5     var6     var7     var8
+ 74.73385 71.64560 75.53826 81.21618 73.69573 77.17450 73.05129 72.90464
+     var9    var10    var11    var12    var13    var14    var15    var16
+ 81.51397 74.52383 70.54674 74.00416 69.73784 79.88789 77.08057 74.21768
+    var17    var18    var19    var20
+ 73.55481 78.81785 70.55299 75.90545
```

We again examine the output for X_1.

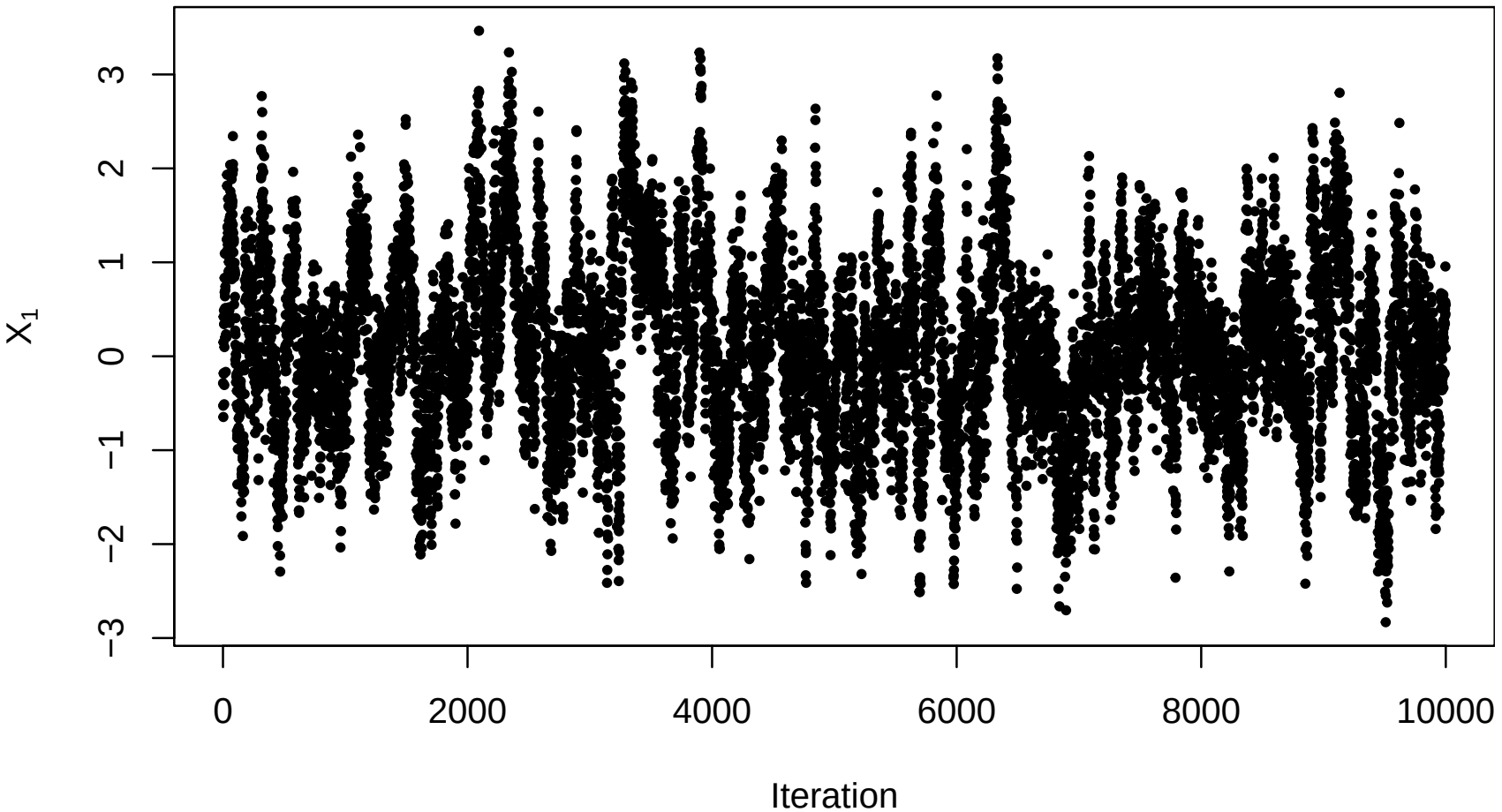
MCMC and Dependence in the Bayesian Linear Model



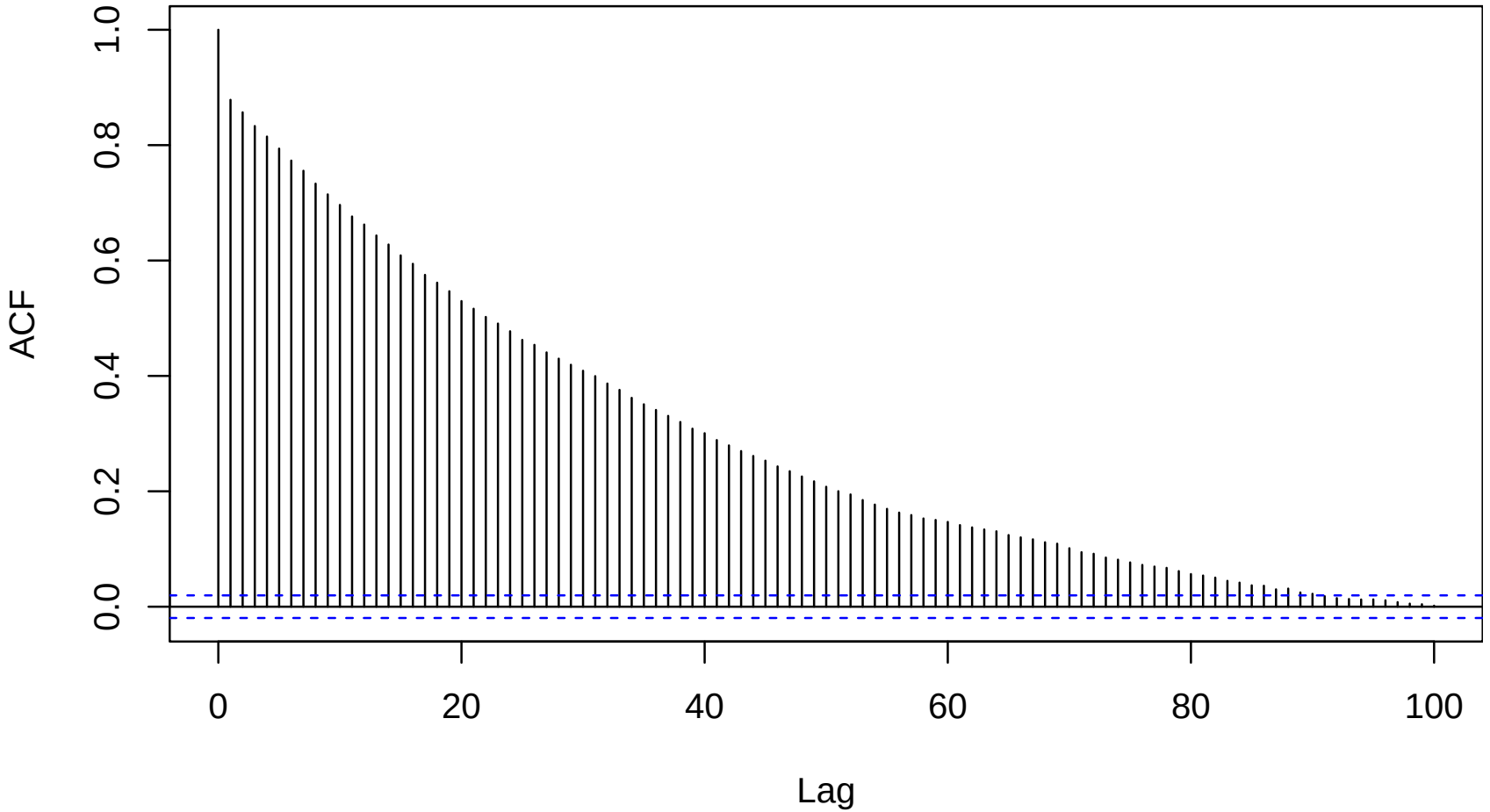
MCMC and Dependence in the Bayesian Linear Model



We now repeat using two blocks of 10 variables.



MCMC and Dependence in the Bayesian Linear Model



There is again evidence of autocorrelation, but it is further reduced, and the efficiency has improved further.

```
(ESS<-effectiveSize(Xsamp3))

+     var1     var2     var3     var4     var5     var6     var7     var8
+ 142.3526 158.1403 145.1260 147.8015 152.6840 146.1852 152.8961 146.3396
+     var9    var10    var11    var12    var13    var14    var15    var16
+ 151.8224 151.3570 144.2619 150.8587 148.2810 154.4746 148.6937 156.1808
+    var17    var18    var19    var20
+ 153.5165 152.5049 152.2063 152.7954
```

The conclusion is that updating parameters in blocks can improve sampler performance.

Bayesian ANOVA example: The two-way layout (ANOVA) model is an example of the linear model with two factor predictors X_1 (taking J levels) and X_2 (taking K levels).

		X_2			
		1	2	3	4
X_1	1				
	2				
	3				
	4				
	5				
	6				

Suppose that for $i = 1, \ldots, n$ we write

$$Y_i = \sum_{j=1}^{J} \beta_{1j} \mathbb{1}_j(x_{i1}) + \sum_{k=1}^{K} \beta_{2k} \mathbb{1}_k(x_{i2}) + \epsilon_i$$

where $\epsilon_i \sim Normal(0, \sigma^2)$ for $i = 1, \ldots, n$ are independent. That is,

$$\mathbb{E}[Y_i | X_{i1} = j, X_{i2} = k] = \beta_{1j} + \beta_{2k}.$$

This is a linear model in the $J + K$ parameters

$$\beta_{11}, \ldots, \beta_{1J}, \beta_{21}, \ldots, \beta_{2K}.$$

However, in this form, the model is not well-specified, as there is parameter non-identifiability – we may change parameters by additive constants in such a way that the overall conditional mean for Y_i is unchanged.

Although this is not insurmountable in a Bayesian analysis (as the prior specification can be used to induce identifiability) it is more common to force 'likelihood' identifiability by amending the specification.

For example,

$$Y_i = \sum_{j=1}^{J} \beta_{1j} \mathbb{1}_j(x_{i1}) + \sum_{k=2}^{K} \beta_{2k} \mathbb{1}_k(x_{i2}) + \epsilon_i$$

that is, we pick a baseline cross-categorization, here $j = 1$, and remove one parameter for the effect of X_2 for that group. The resulting model has $1 + (J-1) + (K-1) = J + K - 1$ parameters. We have

$$\mathbb{E}[Y_i | x_{i1} = j, x_{i2} = k] = \begin{cases} \beta_{1j} & j \geq 1, k = 1 \\ \beta_{1j} + \beta_{2k} & j \geq 1, k > 1 \end{cases}$$

For simplicity, we consider a *balanced* design, where there are r observations of each cross-category.

Modelled means:

		X_2			
		1	2	3	4
X_1	1	β_{11}	$\beta_{11} + \beta_{22}$	$\beta_{11} + \beta_{23}$	$\beta_{11} + \beta_{24}$
	2	β_{12}	$\beta_{12} + \beta_{22}$	$\beta_{12} + \beta_{23}$	$\beta_{12} + \beta_{24}$
	3	β_{13}	$\beta_{13} + \beta_{22}$	$\beta_{13} + \beta_{23}$	$\beta_{13} + \beta_{24}$
	4	β_{14}	$\beta_{14} + \beta_{22}$	$\beta_{14} + \beta_{23}$	$\beta_{14} + \beta_{24}$
	5	β_{15}	$\beta_{15} + \beta_{22}$	$\beta_{15} + \beta_{23}$	$\beta_{15} + \beta_{24}$
	6	β_{16}	$\beta_{16} + \beta_{22}$	$\beta_{16} + \beta_{23}$	$\beta_{16} + \beta_{24}$

$$\{\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16}, \beta_{22}, \beta_{23}, \beta_{24}\}$$

Example: We take $J = 6, K = 4, r = 5$: total of $n = 120$ observations.

```
+                  Y X1 X2
+   [1,]  0.1630229  1  1
+   [2,]  0.5467725  2  1
+   [3,]  0.1367214  3  1
+   [4,] -3.1211497  4  1
+   [5,]  3.3878850  5  1
+   [6,]  0.4526632  6  1
+   [7,]  6.0772107  1  2
+   [8,] -0.2481566  2  2
+   [9,]  2.0898470  3  2
+  [10,] -9.5359364  4  2
+  [11,] -4.1870555  5  2
+  [12,] -4.4757723  6  2
+  [13,] -1.6083612  1  3
+  [14,]  1.8752678  2  3
+  [15,] -1.4504321  3  3
```

We can construct the design matrix $\mathbf{X}$ for the linear model

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

where $\mathbf{X}$ is now $n \times (J + K - 1) \equiv 120 \times 9$, and

$$\beta = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16}, \beta_{22}, \beta_{23}, \beta_{24})^\top$$

is 9×1.

For a Bayesian analysis of this model, as before, the likelihood is

$$\mathcal{L}_n(\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta)\right\}.$$

A conjugate prior in this setting can be factorized

$$\pi_0(\beta, \sigma^2) = \pi_0(\sigma^2)\pi_0(\beta|\sigma^2)$$

where, for $d = 9$,

$$\pi_0(\sigma^2) \equiv \mathit{InvGamma}(a_0/2, b_0/2) \qquad \pi_0(\beta|\sigma^2) \equiv \mathit{Normal}_d(\mathbf{m}_0, \sigma^2\mathbf{M}_0)$$

where $a_0, b_0, \mathbf{m}_0$ and $\mathbf{M}_0$ are fixed constant *hyperparameters*.

$$\pi_0(\sigma^2) = \frac{(b_0/2)^{a_0/2}}{\Gamma(a_0/2)} \left(\frac{1}{\sigma^2}\right)^{a_0/2+1} \exp\left\{-\frac{b_0}{2\sigma^2}\right\}$$

$$\pi_0(\beta|\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{d/2} \frac{1}{|\mathbf{M}_0|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta - \mathbf{m}_0)^\top \mathbf{M}_0^{-1}(\beta - \mathbf{m}_0)\right\}$$

Under the conjugate prior, if

$$\mathbf{m}_n = (\mathbf{X}^\top\mathbf{X} + \mathbf{M}_0^{-1})^{-1}(\mathbf{X}^\top\mathbf{y} + \mathbf{M}_0^{-1}\mathbf{m}_0) \qquad \mathbf{M}_n = (\mathbf{X}^\top\mathbf{X} + \mathbf{M}_0^{-1})^{-1}$$

and

$$c_n = \mathbf{y}^\top\mathbf{y} + \mathbf{m}_0^\top\mathbf{M}_0^{-1}\mathbf{m}_0 - \mathbf{m}_n^\top\mathbf{M}_n^{-1}\mathbf{m}_n$$

we have marginally that

$$\pi_n(\sigma^2) \equiv InvGamma(a_n/2, b_n/2)$$

where $a_n = n + a_0, b_n = c_n + b_0$, and conditionally

$$\pi_n(\beta|\sigma^2) \equiv Normal(\mathbf{m}_n, \sigma^2\mathbf{M}_n).$$

Marginally, we have

$$\pi_n(\beta) = \frac{\Gamma((a_n + d)/2)}{\Gamma(a_n/2) a_n^{d/2} \pi^{d/2}} \frac{1}{|\Sigma_n|^{1/2}} \frac{1}{\left\{1 + a_n^{-1}(\beta - \mathbf{m}_n)^\top \Sigma_n^{-1}(\beta - \mathbf{m}_n)\right\}^{\frac{a_n + d}{2}}}$$

which is the Student-t distribution as in the `R` package `mvnfast`, where $a_n \Sigma_n = b_n M_n$.

For the analysis, we choose $a_0 = b_0 = 2$, and $\mathbf{m}_0 = \mathbf{0}_9$, with $\mathbf{M}_0$ being diagonal with diagonal elements set to be 2500 – this represents a very diffuse prior.

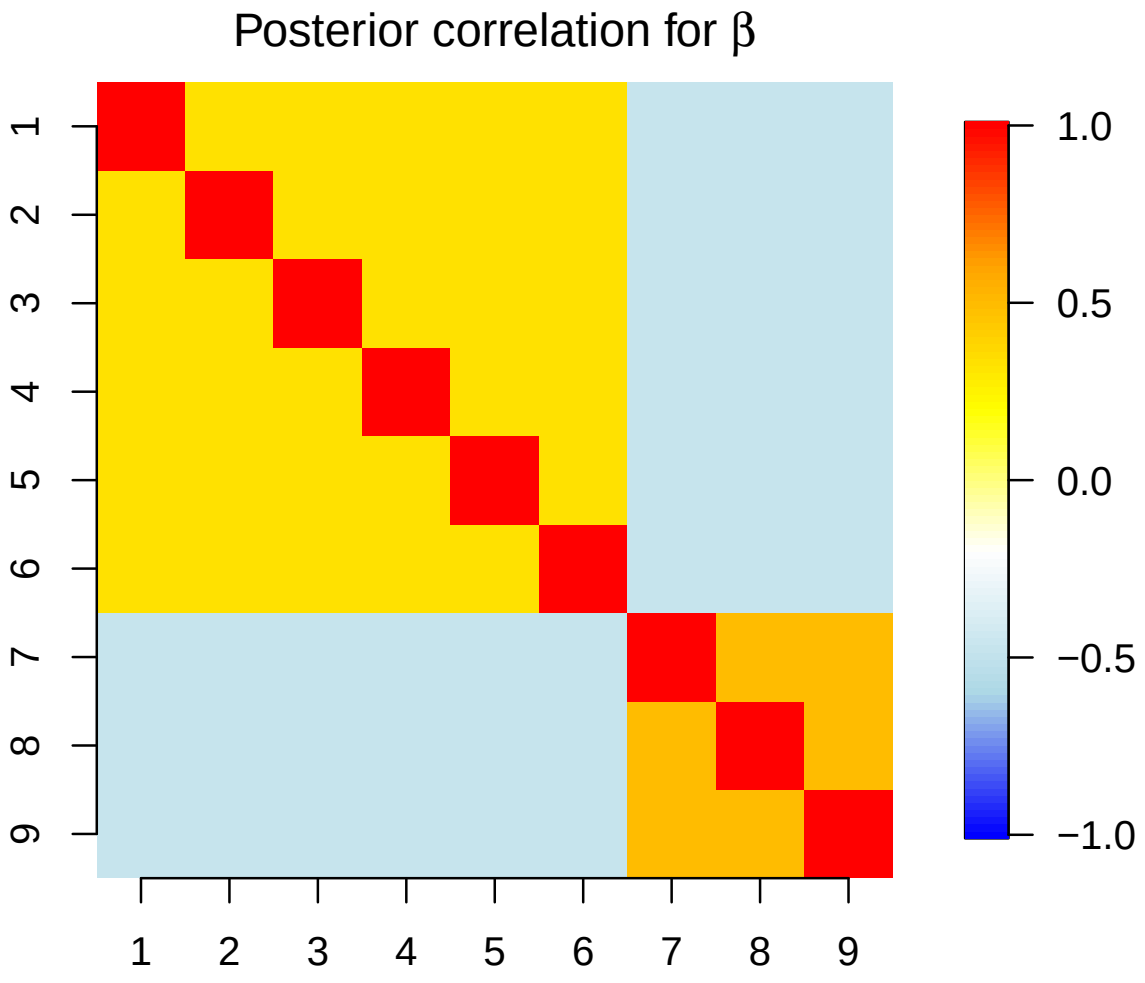
The posterior covariance matrix is

$$\left(\frac{a_n}{a_n - 2}\right)\Sigma_n,$$

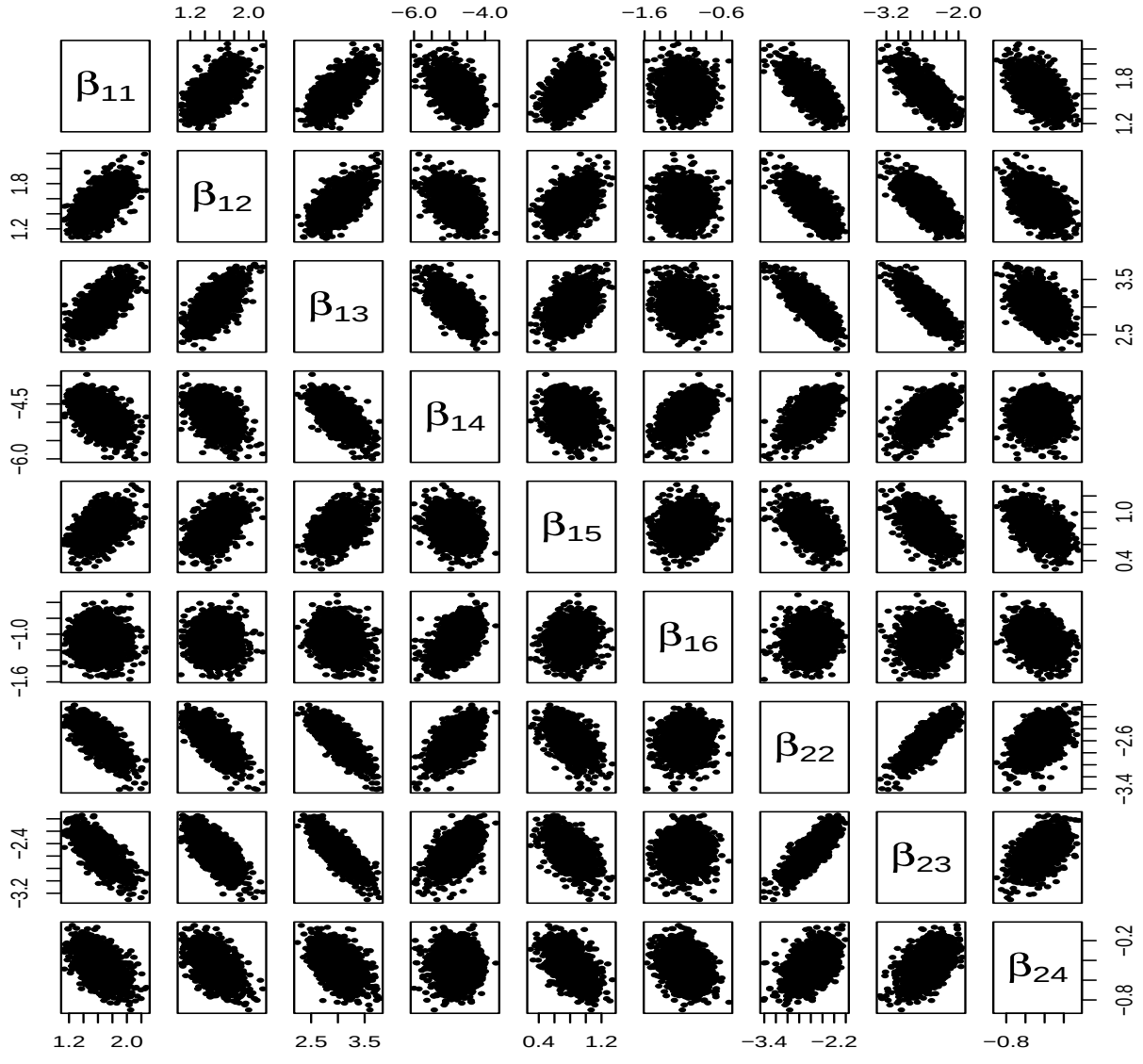
and from this we can compute the posterior correlation matrix.

```
+           [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
+   [1,]  1.0000  0.3333  0.3333  0.3333  0.3333  0.3333 -0.4714 -0.4714 -0.4714
+   [2,]  0.3333  1.0000  0.3333  0.3333  0.3333  0.3333 -0.4714 -0.4714 -0.4714
+   [3,]  0.3333  0.3333  1.0000  0.3333  0.3333  0.3333 -0.4714 -0.4714 -0.4714
+   [4,]  0.3333  0.3333  0.3333  1.0000  0.3333  0.3333 -0.4714 -0.4714 -0.4714
+   [5,]  0.3333  0.3333  0.3333  0.3333  1.0000  0.3333 -0.4714 -0.4714 -0.4714
+   [6,]  0.3333  0.3333  0.3333  0.3333  0.3333  1.0000 -0.4714 -0.4714 -0.4714
+   [7,] -0.4714 -0.4714 -0.4714 -0.4714 -0.4714 -0.4714  1.0000  0.5000  0.5000
+   [8,] -0.4714 -0.4714 -0.4714 -0.4714 -0.4714 -0.4714  0.5000  1.0000  0.5000
+   [9,] -0.4714 -0.4714 -0.4714 -0.4714 -0.4714 -0.4714  0.5000  0.5000  1.0000
```

MCMC and Dependence in the Bayesian Linear Model



MCMC and Dependence in the Bayesian Linear Model



We may construct a linear reparameterization of the model and write in the original specification $\theta = \mathbf{A}\beta$ for any $d \times d$ non-singular matrix $\mathbf{A}$, so that $\beta = \mathbf{A}^{-1}\theta$ and hence

$$\mathbf{Y} = \mathbf{X}\mathbf{A}^{-1}\theta + \varepsilon = \mathbf{X}_A\theta + \varepsilon$$

say. The induced prior on θ is

$$\pi_0(\theta) \equiv Normal_9(\mathbf{A}\mathbf{m}_0, \mathbf{A}\mathbf{M}_0\mathbf{A}^\top)$$

and we can readily repeat the analysis.

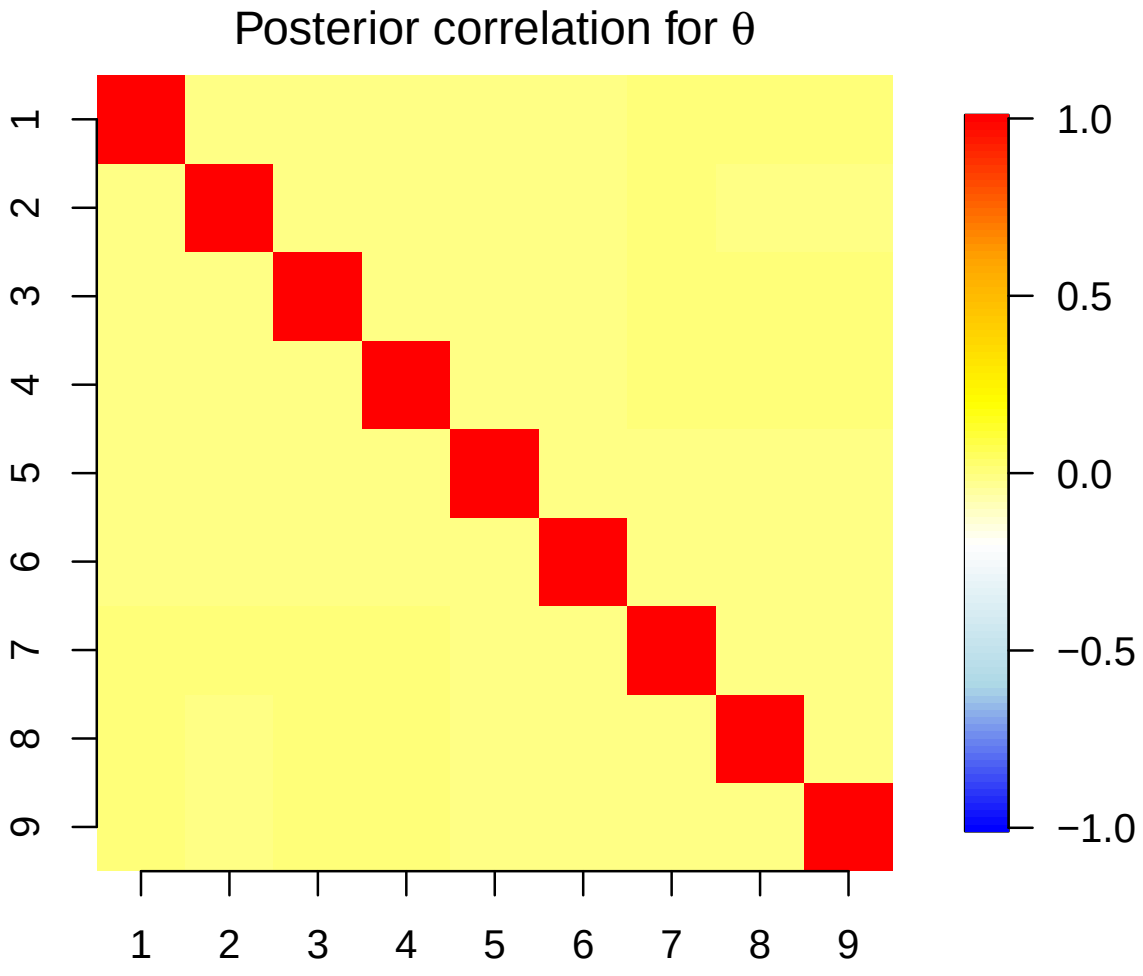
We choose $\mathbf{A}$ inspired by the posterior covariance computed above via the Cholesky decomposition: the posterior covariance for β is determined by

$$\mathbf{M}_n = (\mathbf{X}^\top \mathbf{X} + \mathbf{M}_0^{-1})^{-1}$$

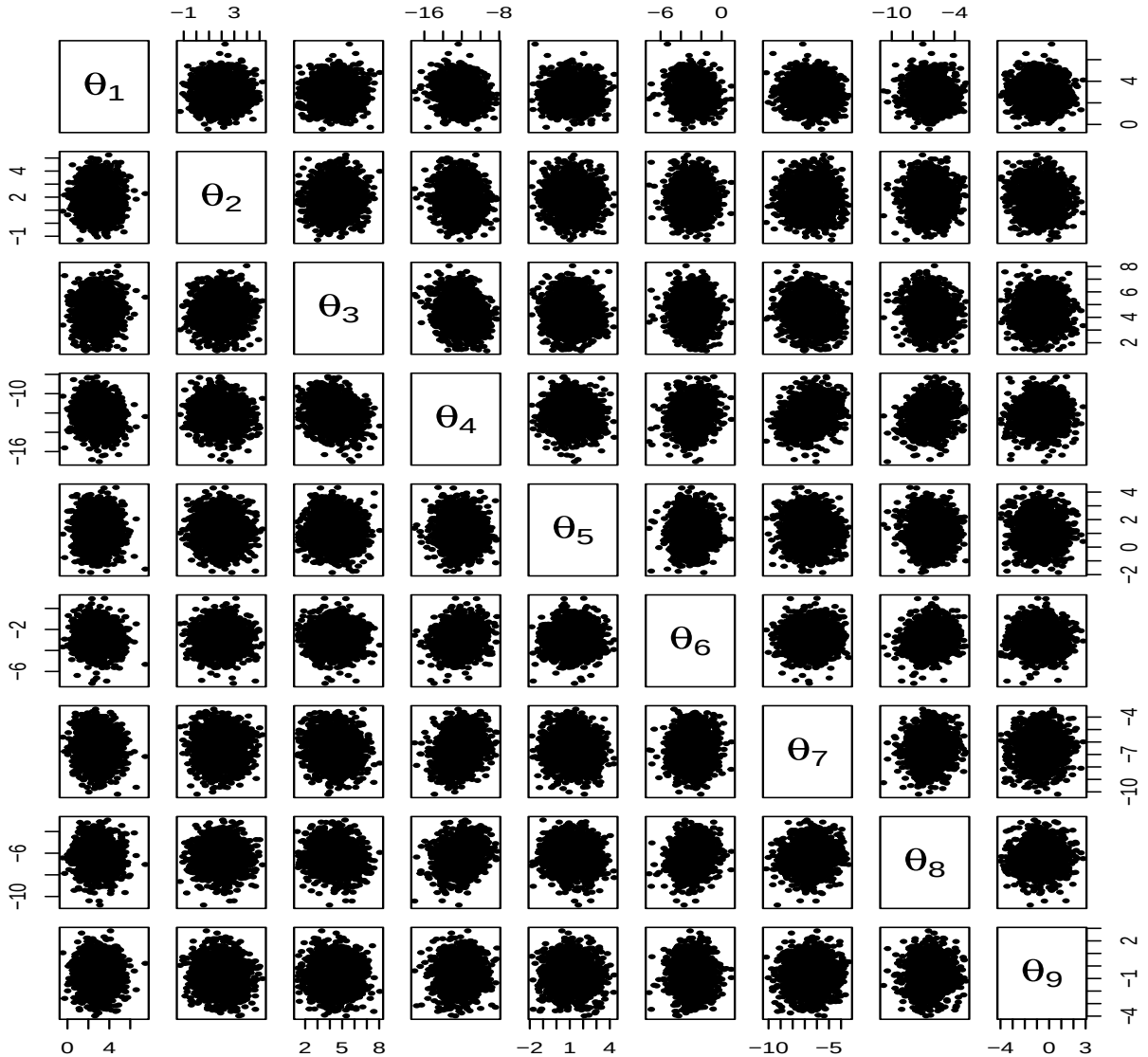
so therefore the posterior covariance for θ is determined by $\mathbf{A}\mathbf{M}_n\mathbf{A}^\top$; if we write $\mathbf{M}_n = \mathbf{C}_n\mathbf{C}_n^\top$ which is always possible as $\mathbf{M}_n$ is positive definite and symmetric, we have that if $\mathbf{A} = \mathbf{C}_n^{-1}$, then

$$\mathbf{A}\mathbf{M}_n\mathbf{A}^\top = \mathbf{C}_n^{-1}\mathbf{C}_n\mathbf{C}_n^\top\mathbf{C}_n^{-\top} = \mathbf{I}_d.$$

Note that the `chol` function in `R` returns $\mathbf{C}_n^\top$. The components of θ are uncorrelated in the new posterior.


Posterior correlation for θ
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9
1.0
0.5
0.0
−0.5
−1.0

MCMC and Dependence in the Bayesian Linear Model



Contrast parameterization: It is common to re-write the original model via the contrast parameterization, where

$$Y_i = \beta_0 + \sum_{j=2}^{J} \delta_j \mathbb{1}_j(x_{i1}) + \sum_{k=2}^{K} \gamma_k \mathbb{1}_k(x_{i2}) + \epsilon_i$$

that is, with β_0 representing the baseline mean, and then the δ and γ parameter representing contrasts from baseline

$$\mathbb{E}[Y_i | x_{i1} = j, x_{i2} = k] = \begin{cases} \beta_0 & j = k = 1 \\ \beta_0 + \gamma_k & j = 1, k > 1 \\ \beta_0 + \delta_j & j > 1, k = 1 \\ \beta_0 + \delta_j + \gamma_k & j > 1, k > 1 \end{cases}$$

Modelled means:

		X_2			
		1	2	3	4
X_1	1	β_0	$\beta_0 + \gamma_2$	$\beta_0 + \gamma_3$	$\beta_0 + \gamma_4$
	2	$\beta_0 + \delta_2$	$\beta_0 + \delta_2 + \gamma_2$	$\beta_0 + \delta_2 + \gamma_3$	$\beta_0 + \delta_2 + \gamma_4$
	3	$\beta_0 + \delta_3$	$\beta_0 + \delta_3 + \gamma_2$	$\beta_0 + \delta_3 + \gamma_3$	$\beta_0 + \delta_3 + \gamma_4$
	4	$\beta_0 + \delta_4$	$\beta_0 + \delta_4 + \gamma_2$	$\beta_0 + \delta_4 + \gamma_3$	$\beta_0 + \delta_4 + \gamma_4$
	5	$\beta_0 + \delta_5$	$\beta_0 + \delta_5 + \gamma_2$	$\beta_0 + \delta_5 + \gamma_3$	$\beta_0 + \delta_5 + \gamma_4$
	6	$\beta_0 + \delta_6$	$\beta_0 + \delta_6 + \gamma_2$	$\beta_0 + \delta_6 + \gamma_3$	$\beta_0 + \delta_6 + \gamma_4$

$$\{\beta_0, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \gamma_2, \gamma_3, \gamma_4\}$$

The new parameterization (denoted ϑ say) is a linear transformation of the original parameterization, via some $d \times d$ matrix matrix $\mathbf{L}$ say, $\vartheta = \mathbf{L}\beta$. The induced prior in the new parameterization is

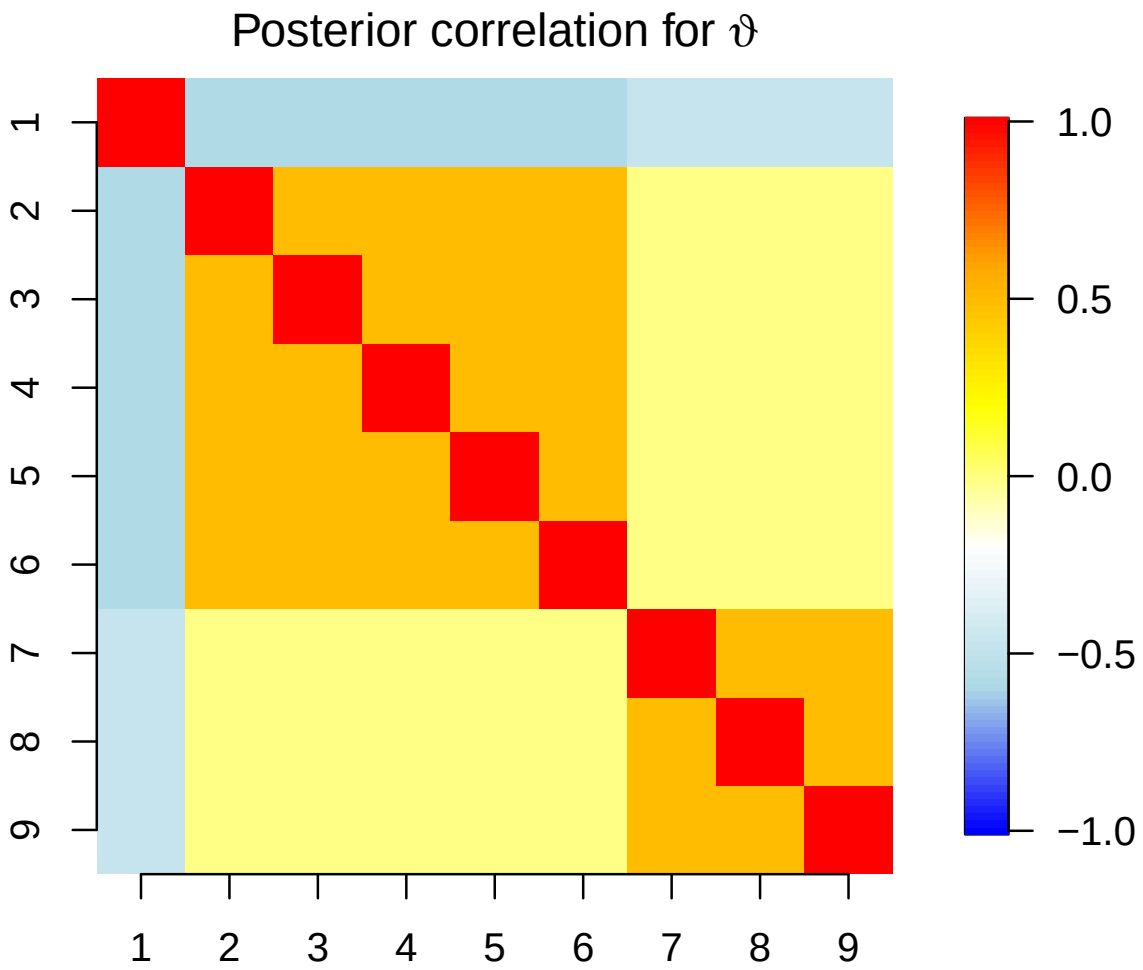
$$Normal_d(\mathbf{L}\mathbf{m}_0, \mathbf{L}\mathbf{M}_0\mathbf{L}^\top)$$

and prior independence for the components β does not imply prior independence for the components ϑ.

```
+ [1] "Induced prior covariance"
+       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
+  [1,] 2500 2500 2500 2500 2500 2500    0    0    0
+  [2,] 2500 5000 2500 2500 2500 2500    0    0    0
+  [3,] 2500 2500 5000 2500 2500 2500    0    0    0
+  [4,] 2500 2500 2500 5000 2500 2500    0    0    0
+  [5,] 2500 2500 2500 2500 5000 2500    0    0    0
+  [6,] 2500 2500 2500 2500 2500 5000    0    0    0
+  [7,]    0    0    0    0    0    0 2500    0    0
+  [8,]    0    0    0    0    0    0    0 2500    0
+  [9,]    0    0    0    0    0    0    0    0 2500
```

```
+ [1] "Posterior correlation"
+          [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
+  [1,]  1.0000 -0.5772 -0.5772 -0.5772 -0.5772 -0.5772 -0.4714 -0.4714 -0.4714
+  [2,] -0.5772  1.0000  0.4999  0.4999  0.4999  0.4999 -0.0001 -0.0001 -0.0001
+  [3,] -0.5772  0.4999  1.0000  0.4999  0.4999  0.4999 -0.0001 -0.0001 -0.0001
+  [4,] -0.5772  0.4999  0.4999  1.0000  0.4999  0.4999 -0.0001 -0.0001 -0.0001
+  [5,] -0.5772  0.4999  0.4999  0.4999  1.0000  0.4999 -0.0001 -0.0001 -0.0001
+  [6,] -0.5772  0.4999  0.4999  0.4999  0.4999  1.0000 -0.0001 -0.0001 -0.0001
+  [7,] -0.4714 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001  1.0000  0.5000  0.5000
+  [8,] -0.4714 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001  0.5000  1.0000  0.5000
+  [9,] -0.4714 -0.0001 -0.0001 -0.0001 -0.0001 -0.0001  0.5000  0.5000  1.0000
```

The conclusion is that parameterization affects the posterior dependence, and therefore has an impact on the MCMC strategy deployed.


Posterior correlation for ϑ
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9
1.0
0.5
0.0
−0.5
−1.0

MCMC and Dependence in the Bayesian Linear Model

