UiO : Matematisk institutt

Det matematisk-naturvitenskapelige fakultet

STK-4051/9051 Computational Statistics Spring 2024
Slides Exercise 10

Instructor: Odd Kolbjørnsen, oddkol@math.uio.no

Exercise 41 (Combinations of Markov chains)

Assume you have two Markov chains described by the transition densities $P_1(y|x)$ and $P_2(y|x)$, both having a target distribution $\pi(x)$ as stationary distribution, that is

$$\pi(y) = \int_x \pi(x)P_j(y|x)dx.j = 1, 2.$$

Define now a new Markov chain by the transition density

$$P(y|x) = \alpha P_1(y|x) + (1-\alpha)P_2(y|x)$$

were $\alpha \in [0, 1]$.

(a). Show that this new transition density also have $\pi(y)$ as stationary distribution

(b). Discuss the implication of this result with respect to constructing Markov chain Monte Carlo metohds.

Exercise 41

(a). Show that this new transition density also have $\pi(y)$ as stationary distribution

$$P(y|x) = \alpha P_1(y|x) + (1-\alpha)P_2(y|x)$$

were $\alpha \in [0,1]$.

$$\int_x \pi(x)P(y|x)dx = \int_x \pi(x)[\alpha P_1(y|x) + (1-\alpha)P_2(y|x)]dx$$

$$= \alpha \int_x \pi(x)P_1(y|x)dx + (1-\alpha)\int_x \pi(x)P_2(y|x)dx$$

$$= \alpha\pi(y) + (1-\alpha)\pi(y) = \pi(y)$$

Exersice 38

Assume again we are interested in simulating from the bivariate Gaussian distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix},$$

that is the same setting as Exercise 35.

(a). Find the conditional distributions for x_1 given x_2 and x_2 given x_1.

(b). Implement the Gibbs sampler based on the condional distributions.

(c). Run the algorithm 1000 iterations for $a = 0$. Use simulations from the standard Gaussian distribution as starting points.

Estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from your simulations. Use traceplots to see how many of the first samples you should discard.

Make a plot of your simulations in the two-dimensional space, drawing a line between each iteration.

Comment on the results.

(d). Now repeat the previous point with $a = 0.99$.

Make similar estimates and plots and comment on the results.

Conditional distributions Ex 38

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix},$$

$$\begin{aligned} E[\boldsymbol{x}_1|\boldsymbol{x}_2] =& \boldsymbol{\mu}_1 + \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}[\boldsymbol{x}_2 - \boldsymbol{\mu}_2] \\ \mathrm{Var}[\boldsymbol{x}_1|\boldsymbol{x}_2] =& \boldsymbol{\Sigma}_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}. \end{aligned}$$

This gives

$$\begin{aligned} E[x_1|x_2] =& 1 + a(x_2 - 2) = 1 - 2a + ax_2 \\ \mathrm{Var}[x_1|x_2] =& 1 - a^2 \\ E[x_2|x_1] =& 2 + a(x_1 - 1) = 2 - a + ax_1 \\ \mathrm{Var}[x_1|x_2] =& 1 - a^2 \end{aligned}$$

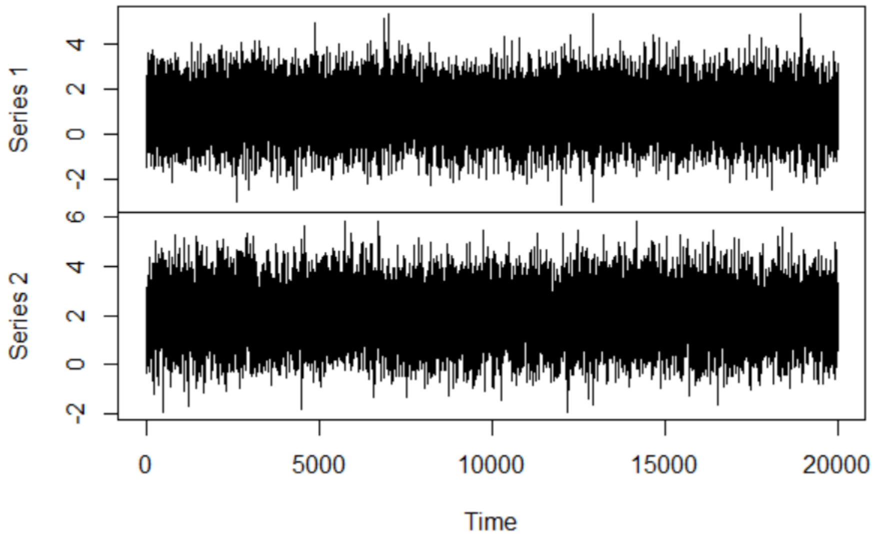
Gibbs sampler

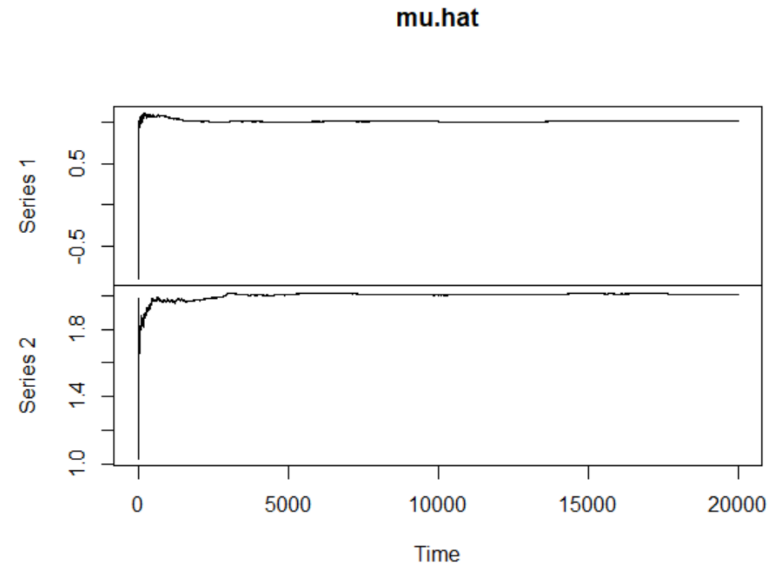
$$E[x_1|x_2] = 1 + a(x_2 - 2) = 1 - 2a + ax_2$$
$$\text{Var}[x_1|x_2] = 1 - a^2$$
$$E[x_2|x_1] = 2 + a(x_1 - 1) = 2 - a + ax_1$$
$$\text{Var}[x_1|x_2] = 1 - a^2$$

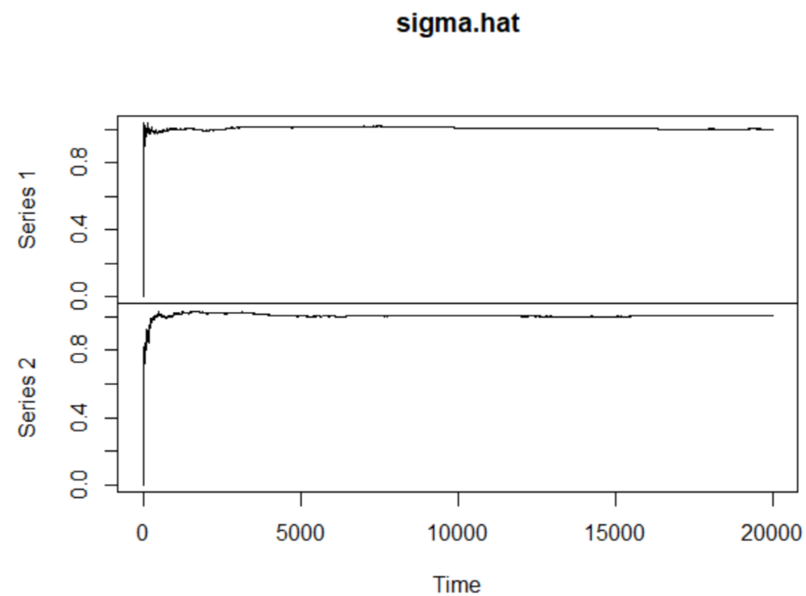
```
#Initialization
X[1,] = rnorm(2)

#Gibbs sampler
for(i in 2:N)
{
  X[i,1] = rnorm(1,1-2*a+a*X[i-1,2],sqrt(1-a^2))
  X[i,2] = rnorm(1,2-a+a*X[i,1],sqrt(1-a^2))
}
```

Results a=0

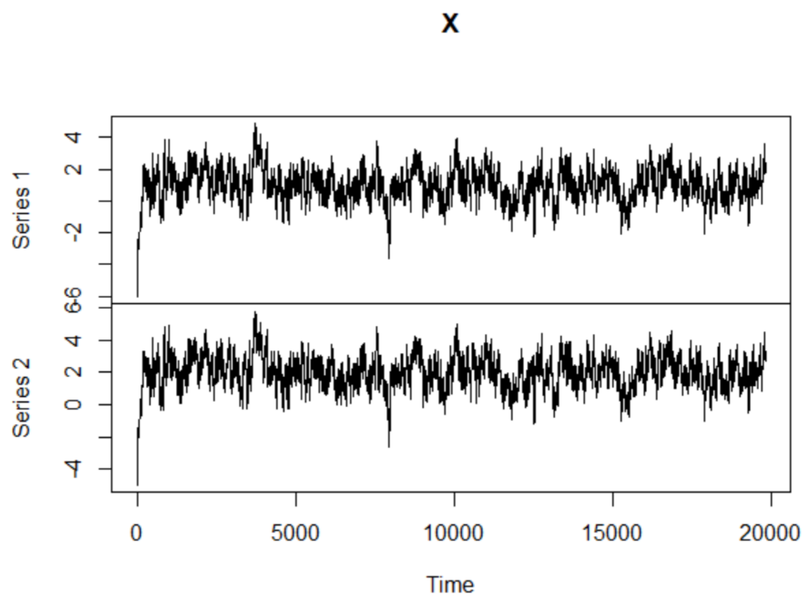


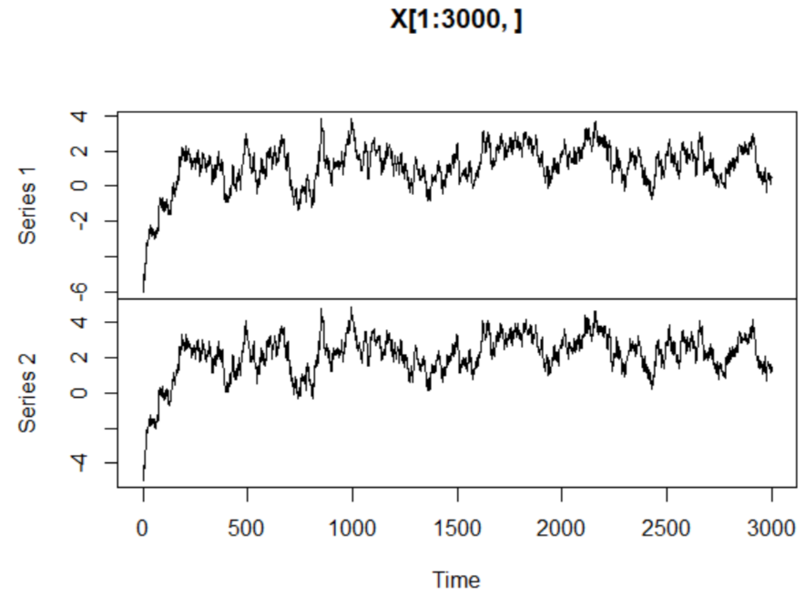


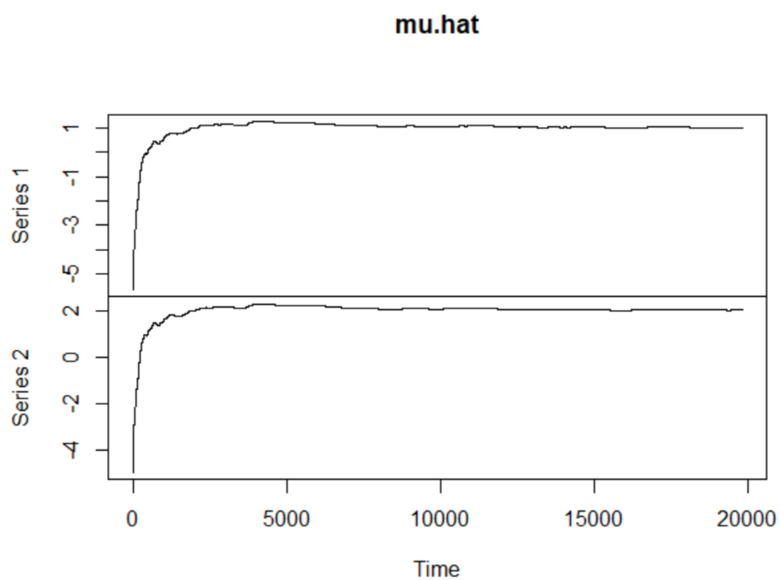


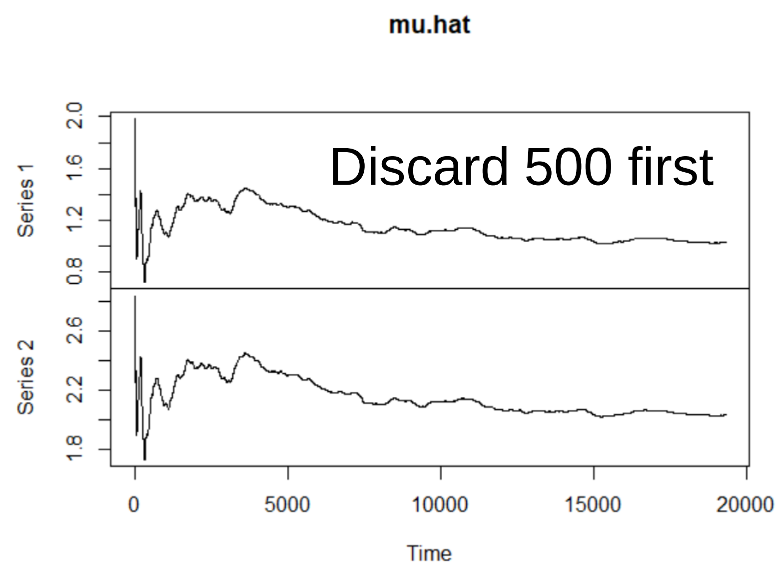
Instant convergence!

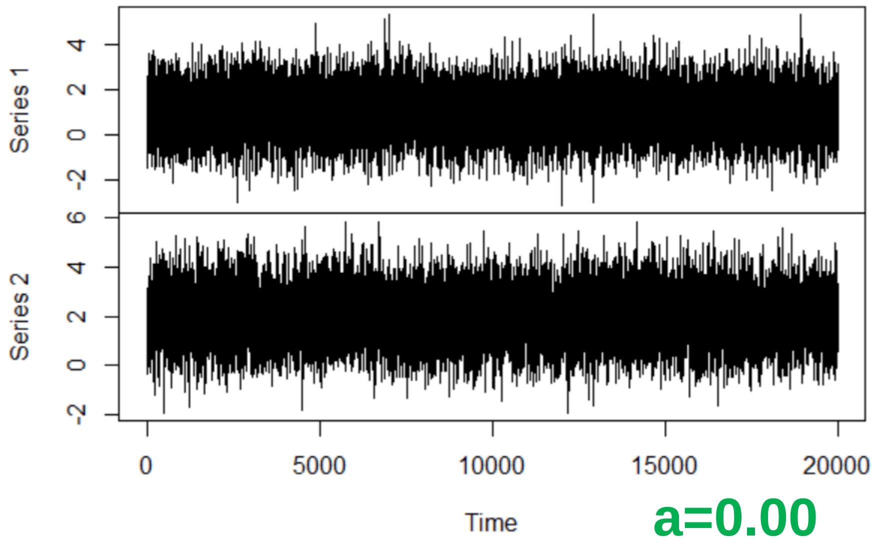
Results a=0.99



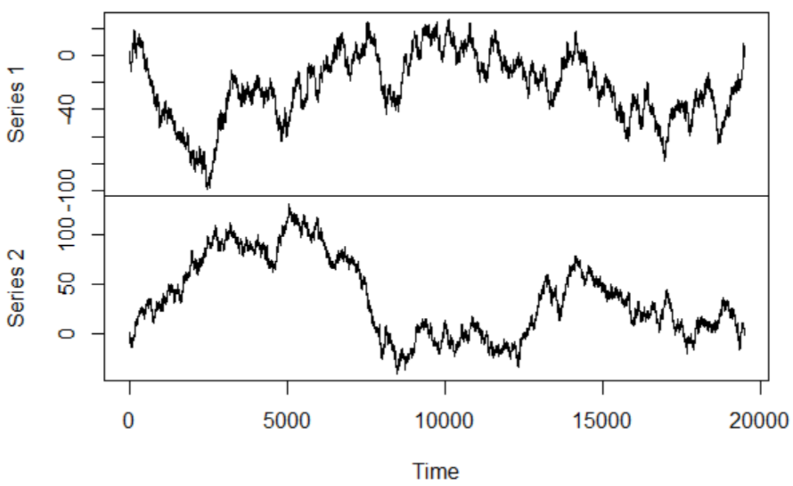


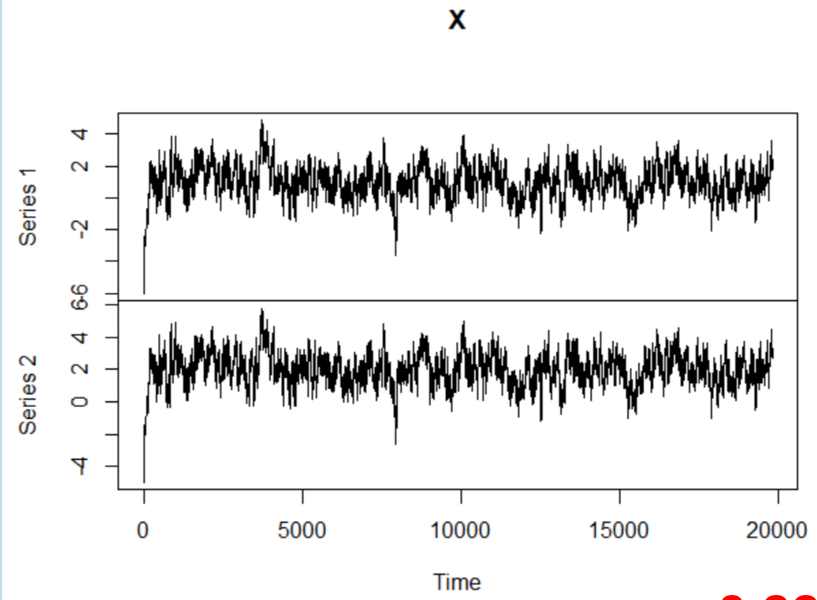






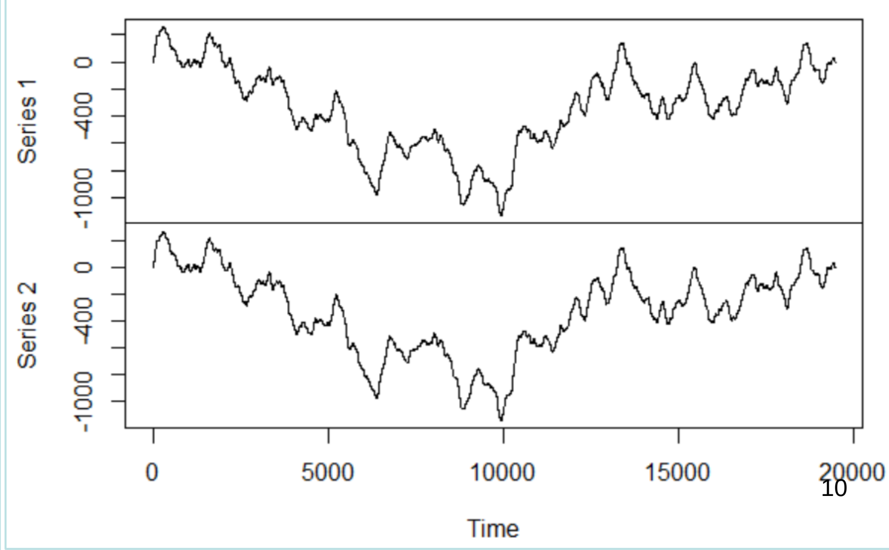
a=0.00



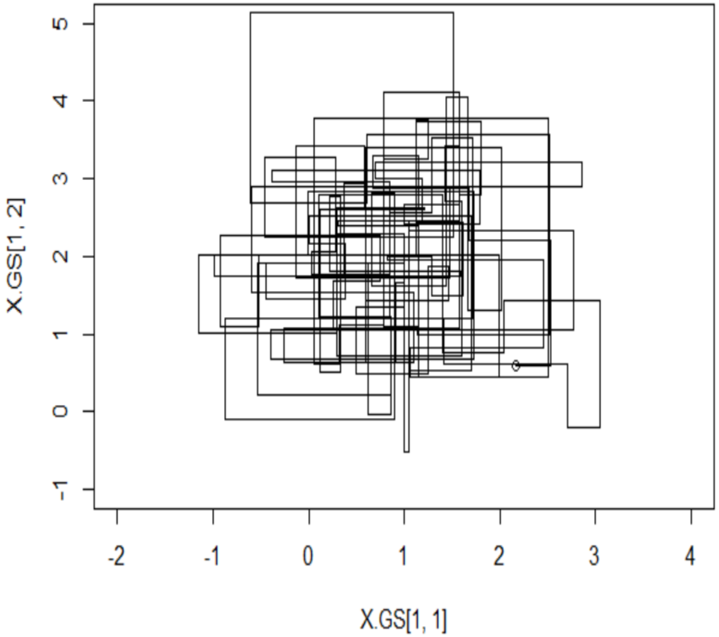


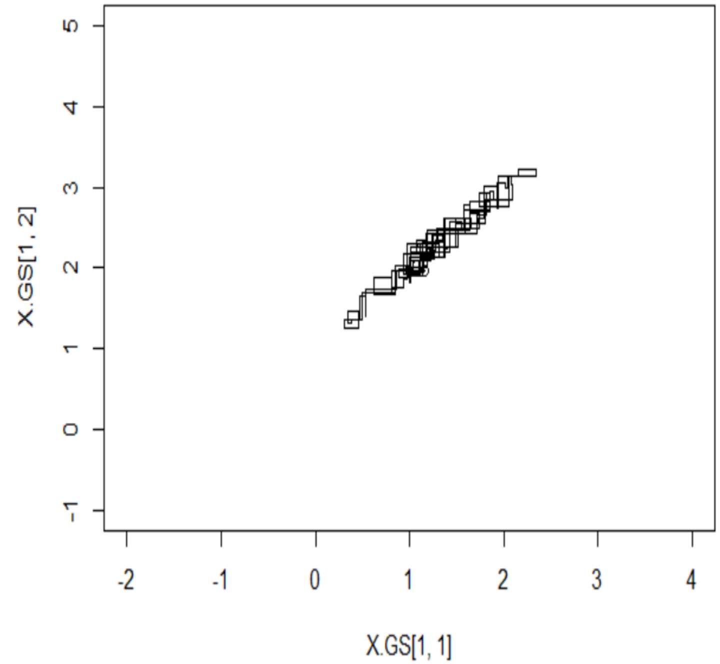
cumsumDiag

a=0.99



100 Samples in a row

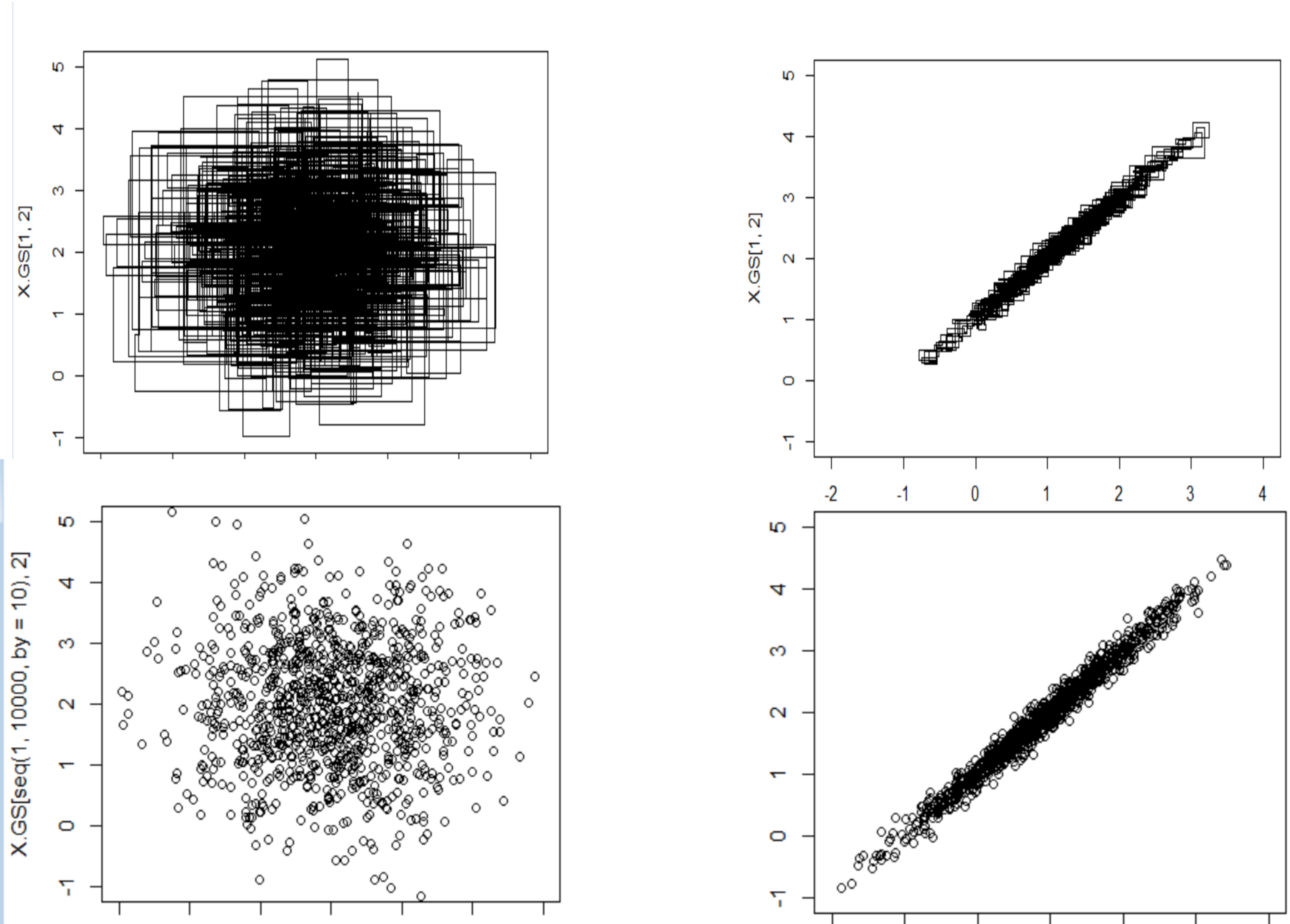




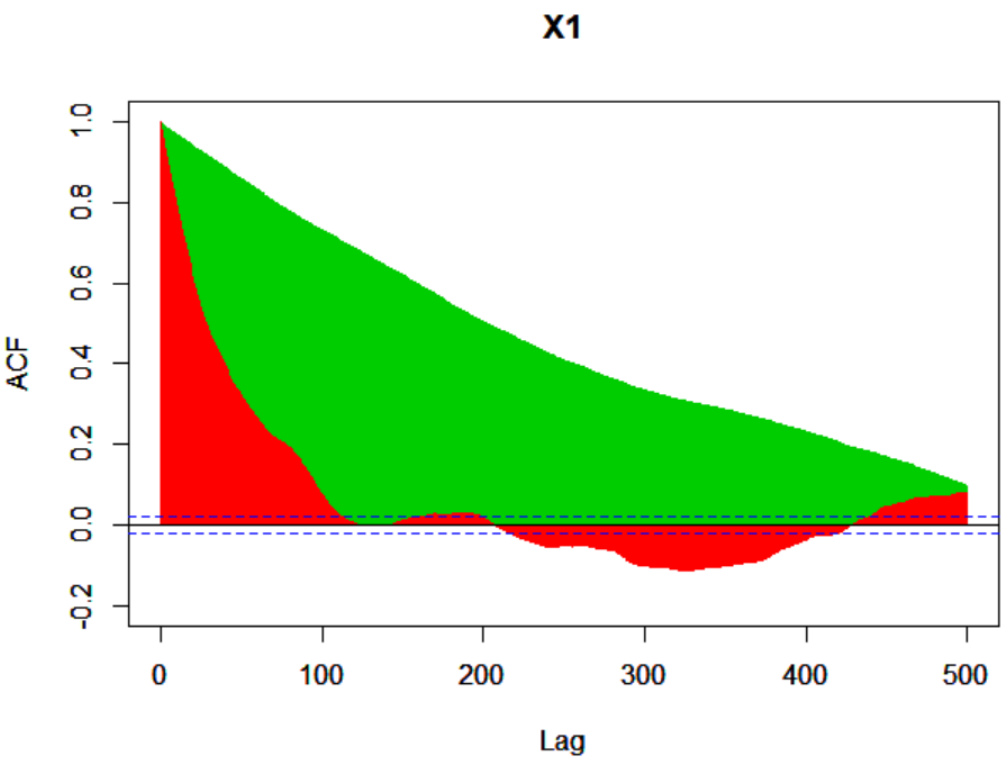
Fore update of individual coordinates

```
plot(X.GS[1,1],X.GS[1,2],xlim=c(-2,4),ylim=c(-1,5))
for(i in 1:1000){
  lines(X.GS[i:(i+1),1],X.GS[c(i,i),2])
  lines(X.GS[c(i+1,i+1),1],X.GS[i:(i+1),2])
 }
```

1000 Samples in a row



Compare ex 35, M-H vs Gibbs



In general Compare also
* runtime per iteration
* Complexity of code (=probability of error)

Ex 39

Exercise 39 (The Gibbs sampler)
The Gibbs sampler applies to vectors of random variables. We shall in this exercise consider random pairs (X, Y). The pedagogical idea is the same as in Exercises 36 and 37. By trying out the algorithm on a simple example where it is not needed, insight is gained into how the scheme operates and into the factors influencing its efficiency. The algorithm is so simple that it is self-explanatory.

Algorithm
Select X (initialization)
Repeat
Sample Y from its conditional distribution given X.
Sample X from its conditional distribution given Y.

It can under general conditions be proved that a simulation of (X, Y) appears in the limit as the loop is continued on and on. We shall below actually prove this result in the simple example considered.

Let (X, Y) be bivariate normal, with means $E(X) = E(Y) = 0$, variances $\text{var}(X) = \text{var}(Y) = 1$ and correlation $\text{corr}(X, Y) = \rho$. The conditional distribution of Y given $X = x$ is then normal with mean ρx and variance $1 - \rho^2$, and the conditional distribution of X given $Y = y$ is defined by symmetry. Let $\{Z_n\}$ and $\{V_n\}$ be sequences of independent normal variables $(0, 1)$. Also assume independence between sequences.

39 a

Let (X, Y) be bivariate normal, with means $E(X) = E(Y) = 0$, variances $\mathrm{var}(X) = \mathrm{var}(Y) = 1$ and correlation $\mathrm{corr}(X, Y) = \rho$. The conditional distribution of Y given $X = x$ is then normal with mean ρx and variance $1 - \rho^2$, and the conditional distribution of X given $Y = y$ is defined by symmetry. Let $\{Z_n\}$ and $\{V_n\}$ be sequences of independent normal variables $(0, 1)$. Also assume independence between sequences.

$$X|Y \sim N(\rho Y, 1 - \rho^2)$$
$$Y|X \sim N(\rho X, 1 - \rho^2)$$

or

$$X|Y = \rho Y + \sqrt{1 - \rho^2} Z$$
$$Y|X = \rho X + \sqrt{1 - \rho^2} V$$

which, when using the Gibbs sampler gives the recursion

$$Y_n = \rho X_n + \sqrt{1 - \rho^2} Z_n,$$
$$X_n = \rho Y_{n-1} + \sqrt{1 - \rho^2} V_n.$$

were all the $\{Z_n\}$ and $\{V_n\}$ are independent variables

39b

It will be proved that as $n \to \infty$, (X_n, Y_n) converges to a sample of (X, Y) for any starting point $X_0 = \mu_0$. We shall also study the rate of convergence. Consider $\{X_n\}$ first.

(b). Show that $X_n = \rho^2 X_{n-1} + \varepsilon_n$, where $\varepsilon_n = \sqrt{1-\rho^2}(\rho Z_{n-1} + V_n)$.

Note that ε_n is normal with mean 0 and variance $\sigma_\varepsilon^2 = 1 - \rho^4$. Stochastic processes of the form $X_n = aX_{n-1} + \varepsilon_n$ is known as autoregressive of order one (or AR(1) for short). They are known to converge in distribution to a limit if $|a| < 1$. Take this result for granted. Explain why it applies here.

(b). From the equation above we have

$$\begin{aligned} X_n =& \rho Y_{n-1} + \sqrt{1-\rho^2} V_n \\ =& \rho[\rho X_{n-1} + \sqrt{1-\rho^2} Z_{n-1}] + \sqrt{1-\rho^2} V_n \\ =& \rho^2 X_{n-1} + \sqrt{1-\rho^2}[\rho Z_{n-1} + V_n] \\ =& \rho^2 X_{n-1} + \varepsilon_n \end{aligned}$$

were $E[\varepsilon_n] = 0$ and

$$\text{Var}[\varepsilon_n] = (1-\rho^2)[\rho^2 + 1] = 1 - \rho^4 \equiv \sigma_\varepsilon^2$$

Since the factor in front of X_{n-1} in the reccursion has an absolute value less than 1, (here: $\rho^2 < 1$), the general results about AR(1) processes applies.

39c

(c). Why is $E(X_n) = \rho^2 E(X_{n-1})$? Use this to establish that $E(X_n) = \rho^{2n}\mu_0$.

(c). We have

$$E[X_n] = E[\rho^2 X_{n-1} + \varepsilon_n] = \rho^2 E[X_{n-1}]$$

which recursively gives $E[X_n] = \rho^{2n}\mu_0$ were $\mu_0 = E[X_0]$.

39d

(d). Show that $\text{var}(X_n) = \rho^4 \text{var}(X_{n-1}) + \sigma_\varepsilon^2$. Since $\text{var} X_0 = 0$, this yields

$$\text{var}(X_n) = \frac{\sigma_\varepsilon^2}{1-\rho^4}(1-\rho^{4n}).$$

Prove it.

(d). We have

$$\begin{aligned}\text{Var}[X_n] =& \text{Var}[\rho^2 X_{n-1} + \varepsilon_n] \\ =& \rho^4 \text{Var}[X_{n-1}] + \sigma_\varepsilon^2\end{aligned}$$

Assume now the statement about the variance is true for n. Then

$$\begin{aligned}\text{Var}[X_{n+1}] =& \rho^4 \text{Var}[X_n] + \sigma_\varepsilon^2 \\ =& \rho^4 \tfrac{\sigma_\varepsilon^2}{1-\rho^4}(1-\rho^{4n}) + \sigma_\varepsilon^2 \\ =& \sigma_\varepsilon^2 \tfrac{\rho^4(1-\rho^{4n})+1-\rho^4}{1-\rho^4} \\ =& \sigma_\varepsilon^2 \tfrac{1-\rho^{4(n+1)}}{1-\rho^4} = 1-\rho^{4(n+1)}\end{aligned}$$

39e

(e). What is the limit for $E(X_n)$ and $\mathrm{var}(X_n)$ when $n \to \infty$? Insert for σ_ε^2.

$$E[X_n] = \rho^{2n}\mu_0 \qquad \mathrm{Var}[X_{n+1}] = \sigma_\varepsilon^2 \frac{1-\rho^{4(n+1)}}{1-\rho^4}$$

(e). When $n \to \infty$ we have

$$E[X_n] \to 0$$
$$\mathrm{Var}[X_n] \to \frac{\sigma_\varepsilon^2}{1-\rho^4} = 1$$

39f

(f). The analysis for $\{Y_n\}$ is similar. Carry it out, i.e. repeat b)-e).

(f). We have

$$
\begin{aligned}
Y_n =& \rho X_n + \sqrt{1-\rho^2} Z_n \\
=& \rho[\rho Y_{n-1} + \sqrt{1-\rho^2} V_n] + \sqrt{1-\rho^2} Z_n \\
=& \rho^2 Y_{n-1} + \sqrt{1-\rho^2}[\rho V_n + Z_n]
\end{aligned}
$$

39g & h

(g). Show that $E(X_nY_n) = \rho E(X_n^2)$ and use this to calculate a formula for this expection.

(g). We have

$$\begin{aligned} E[X_nY_n] &= E[X_n(\rho X_n + \sqrt{1-\rho^2}Z_n)] \\ &= \rho E[X_n^2] \\ &= \rho[1 - \rho^{4n} + \rho^{4n}\mu_0^2] \\ &= \rho + \rho^{4n+1}(\mu_0^2 - 1) \end{aligned}$$

(h). What happens to $E(X_nY_n)$ when $n \to \infty$?

(h). We see that $E[X_nY_n] \to \rho$

39i

(i). Summarize your findings. What is the limit distribution of (X_n, Y_n)? Discuss the convergence speed. What is its dependence on ρ?

(i). We then see that the limit distribution for $X_n, Y_n)$ indeed is the target distribution.

We see that the convergence speed is geometric in ρ^2 for the mean and geometric in ρ^4 for the variances and the correlations.

Ex 42

Exercise 42 (Transformations)
Consider the sampling from the same bivariate normal distribution as in the exercises 40 and 38. We then know that (X, Y) can be represented as

$$X = Z \qquad Y = \rho Z + \sqrt{1-\rho^2}V,$$

where Z and V are independent random variables, both normal $(0, 1)$. Explain why this is so.

(a). What is the convergence rate if you apply the Gibbs sampler to (Z, V) and obtain simulations of (X, Y) by using the transformation after the Gibbs sampler has converged?

Solution to exercise 42. (a). Assume (X, Y) is the current sample, and we draw new ones by

$$X^* = Z^* \qquad Y^* = \rho Z^* + \sqrt{1-\rho^2}V^*.$$

Note that in this case (X^*, Y^*) do not depend on (X, Y) at all! This means that we get immidiate convergence in this case.

Ex 42 b

Exercise 42 (Transformations)
Consider the sampling from the same bivariate normal distribution as in the exercises 40 and 38. We then know that (X, Y) can be represented as

$$X = Z \qquad Y = \rho Z + \sqrt{1-\rho^2}V,$$

where Z and V are independent random variables, both normal $(0,1)$. Explain why this is so.

(a). What is the convergence rate if you apply the Gibbs sampler to (Z, V) and obtain simulations of (X, Y) by using the transformation after the Gibbs sampler has converged?

(b). Discuss possible general lessons for designing simulation algorithms based on the Gibbs sampler or the Metropolis scheme.

Discussion 42 b

(b). If there are components of the full random vector that has a simple marginal distribution, this should be utilized and can influence the convergence rate considerable. In most cases it might not be possible to obtain such marginal distributions. However, in some cases with three sets of variables, we may have that

$$p(x,y|z) = p(x|z)p(y|x,z)$$

where $p(x|z)$ is available. An ordinary Gibbs sampler would be to simulate

$$\begin{aligned} x &\sim p(x|y,z) \\ y &\sim p(y|x,z) \\ z &\sim p(z|x,y) \end{aligned}$$

which would be to jump between three variables or blocks (if x, y or z are vectors). With $p(x|z)$ is available, we can simulate through

$$\begin{aligned} (x,y) &\sim p(x,y|z) \\ z &\sim p(z|x,y) \end{aligned}$$

which only contains two blocks and will typically have much faster convergence. We can diretly simulate from $p(x,y|z) = p(x|z)p(y|x,z)$ due to the assumptions made.