UiO : Matematisk institutt

Det matematisk-naturvitenskapelige fakultet

STK-4051/9051 Computational Statistics Spring 2020
Slides Exercise 11

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

Exersice 7.6

$$p(\theta,\lambda_1,\lambda_2,\alpha) \propto p(\alpha)p(\theta)\cdot p(\lambda_1|\alpha)\cdot p(\lambda_2|\alpha)\times L(\theta,\lambda_1,\lambda_2)$$

$$= \frac{10^{10}\alpha^{10-1}}{\Gamma(10)}e^{-10\alpha}\frac{1}{111}\frac{\alpha^3\lambda_1^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_1}\frac{\alpha^3\lambda_2^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_2}\times\prod_{j=1}^{\theta}\frac{\lambda_1^{x_j}e^{-\lambda_1}}{x_j!}\prod_{j=\theta+1}^{112}\frac{\lambda_2^{x_j}e^{-\lambda_2}}{x_j!}$$

$$p(\lambda_1,\lambda_2|\cdots)\propto\frac{\alpha^3\lambda_1^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_1}\frac{\alpha^3\lambda_2^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_2}\times$$

$$\prod_{j=1}^{\theta}\frac{\lambda_1^{x_j}e^{-\lambda_1}}{x_j!}\prod_{j=\theta+1}^{112}\frac{\lambda_2^{x_j}e^{-\lambda_2}}{x_j!}$$

$$\propto\lambda_1^{2+\sum_{j=1}^{\theta}x_j}e^{-(\alpha+\theta)\lambda_1}\lambda_2^{2+\sum_{j=\theta+1}^{112}x_j}e^{-(\alpha+\theta)\lambda_2}$$

$$\propto\text{Gamma}(\lambda_1;3+\sum_{j=1}^{\theta}x_j,\alpha+\theta)\text{Gamma}(\lambda_2;3+\sum_{j=\theta+1}^{112}x_j,\alpha+112-\theta)$$

Exersice 7.6

$$p(\theta, \lambda_1, \lambda_2, \alpha) \propto p(\alpha)p(\theta) \cdot p(\lambda_1|\alpha) \cdot p(\lambda_2|\alpha) \times L(\theta, \lambda_1, \lambda_2)$$

$$\frac{10^{10}\alpha^{10-1}}{\Gamma(10)}e^{-10\alpha}\frac{1}{111}\;\frac{\alpha^3\lambda_1^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_1}\frac{\alpha^3\lambda_2^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_2}\times\prod_{j=1}^{\theta}\frac{\lambda_1^{x_j}e^{-\lambda_1}}{x_j!}\prod_{j=\theta+1}^{112}\frac{\lambda_2^{x_j}e^{-\lambda_2}}{x_j!}$$

$$\begin{aligned} p(\alpha|\cdots) &\propto \frac{10^{10}\alpha^{10-1}}{\Gamma(10)}e^{-10\alpha}\frac{\alpha^3\lambda_1^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_1}\frac{\alpha^3\lambda_2^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_2} \\ &\propto \alpha^{15}e^{-(10+\lambda_1+\lambda_2)\alpha} \\ &\propto \mathrm{Gamma}(\alpha; 16, 10+\lambda_1+\lambda_2) \end{aligned}$$

Exersice 7.6

$$p(\theta,\lambda_1,\lambda_2,\alpha) \propto p(\alpha)p(\theta)\cdot p(\lambda_1|\alpha)\cdot p(\lambda_2|\alpha) \times L(\theta,\lambda_1,\lambda_2)$$

$$\frac{10^{10}\alpha^{10-1}}{\Gamma(10)}e^{-10\alpha}\frac{1}{111}\ \frac{\alpha^3\lambda_1^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_1}\frac{\alpha^3\lambda_2^{3-1}}{\Gamma(3)}e^{-\alpha\lambda_2}\times\prod_{j=1}^{\theta}\frac{\lambda_1^{x_j}e^{-\lambda_1}}{x_j!}\prod_{j=\theta+1}^{112}\frac{\lambda_2^{x_j}e^{-\lambda_2}}{x_j!}$$

$$p(\theta|\ldots) \propto \frac{1}{111}\prod_{j=1}^{\theta}\frac{\lambda_1^{x_j}e^{-\lambda_1}}{x_j!}\prod_{j=\theta+1}^{112}\frac{\lambda_2^{x_j}e^{-\lambda_2}}{x_j!}$$

$$\propto \lambda_1^{\sum_{j=1}^{\theta}x_j}e^{-\theta\lambda_1}\lambda_2^{\sum_{j=\theta+1}^{112}x_j}e^{-(111-\theta)\lambda_2}$$

Not a known distribution
But the partial sum of x_j
is a factor we will need

```
#Calculating sum_xi for all values of theta
xsum = rep(NA,n-1)
for(i in 1:(n-1))
  xsum[i] = sum(coal$disasters[1:i])
Xsum = sum(coal$disasters)
```

7.6 b

```
for(i in 1:N)
{
  #Sample theta
  logp = xsum*log(lambda1)-lambda1*seq(1,n-1) + (Xsum-xsum)*log(lambda2)-lambda2*seq(n-1,1)
  p = exp(logp)/sum(exp(logp))
  theta = sample(1:(n-1),1,prob=p)
  #Sample lambda
  lambda1 = rgamma(1,shape=3+xsum[theta],rate=alpha+theta)
  lambda2 = rgamma(1,shape=3+Xsum-xsum[theta],rate=alpha+n-theta)
  #Sample alpha
  alpha = rgamma(1,shape=16,rate=10+lambda1+lambda2)
  thetaM = c(thetaM,theta)
  lambdaM1 = c(lambdaM1,lambda1)
  lambdaM2 = c(lambdaM2,lambda2)
  alphaM = c(alphaM,alpha)
}
```

$$\propto \lambda_1^{\sum_{j=1}^{\theta} x_j} e^{-\theta\lambda_1} \lambda_2^{\sum_{j=\theta+1}^{112} x_j} e^{-(111-\theta)\lambda_2}$$

$$\text{Gamma}(\lambda_1; 3 + \sum_{j=1}^{\theta} x_j, \alpha + \theta)$$

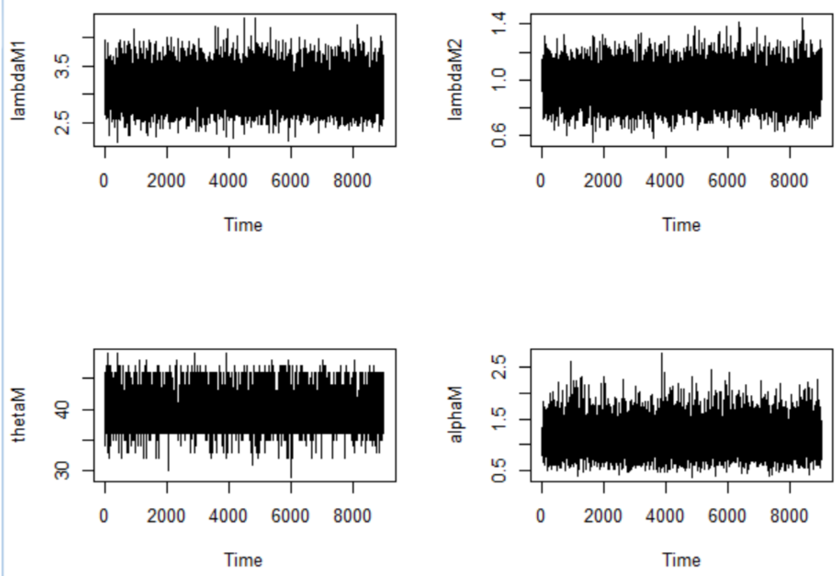
$$\text{Gamma}(\lambda_2; 3 + \sum_{j=\theta+1}^{112} x_j, \alpha + 112 - \theta)$$

$$\propto \text{Gamma}(\alpha; 16, 10 + \lambda_1 + \lambda_2)$$

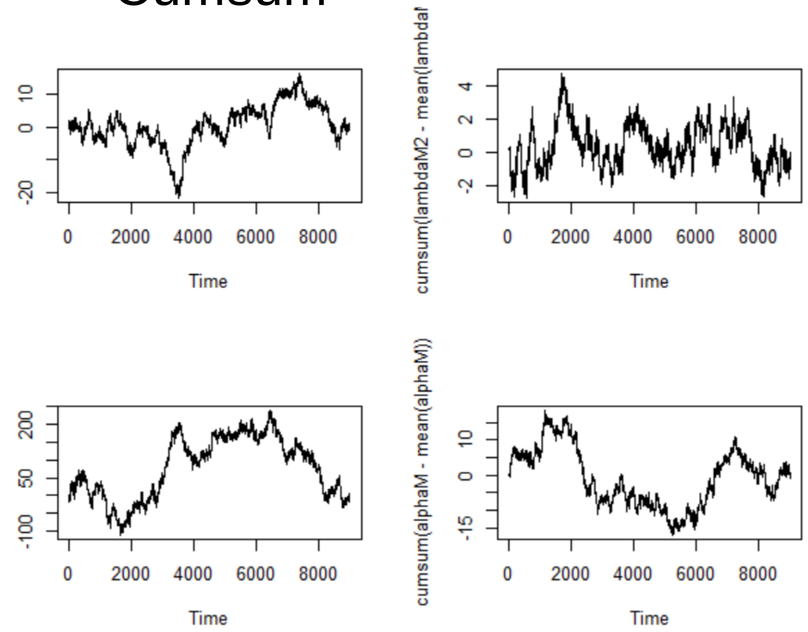
Diagnostics

Have removed the 1000 first

Trace plot

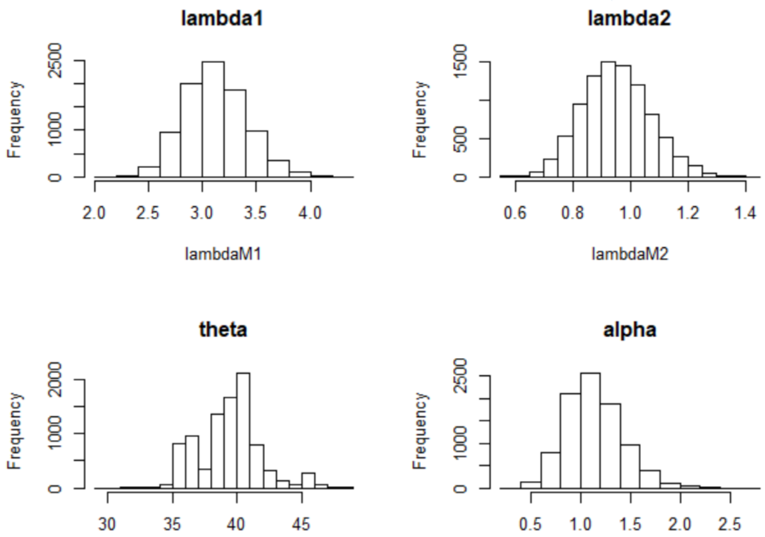


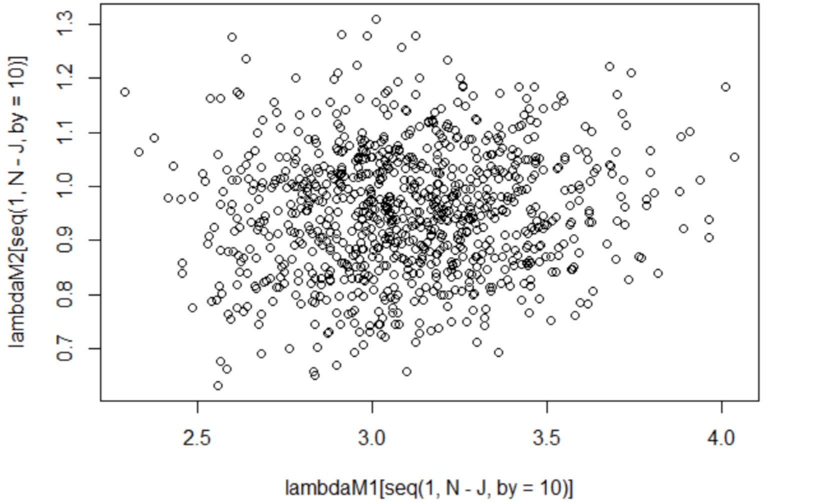
Cumsum

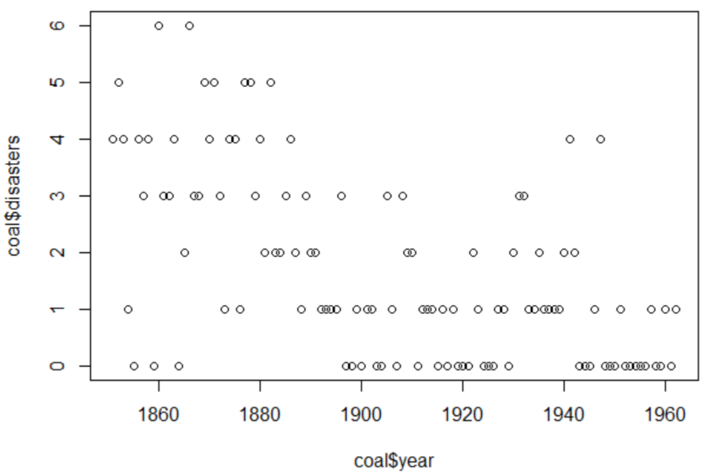


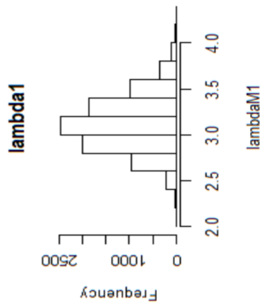
Results

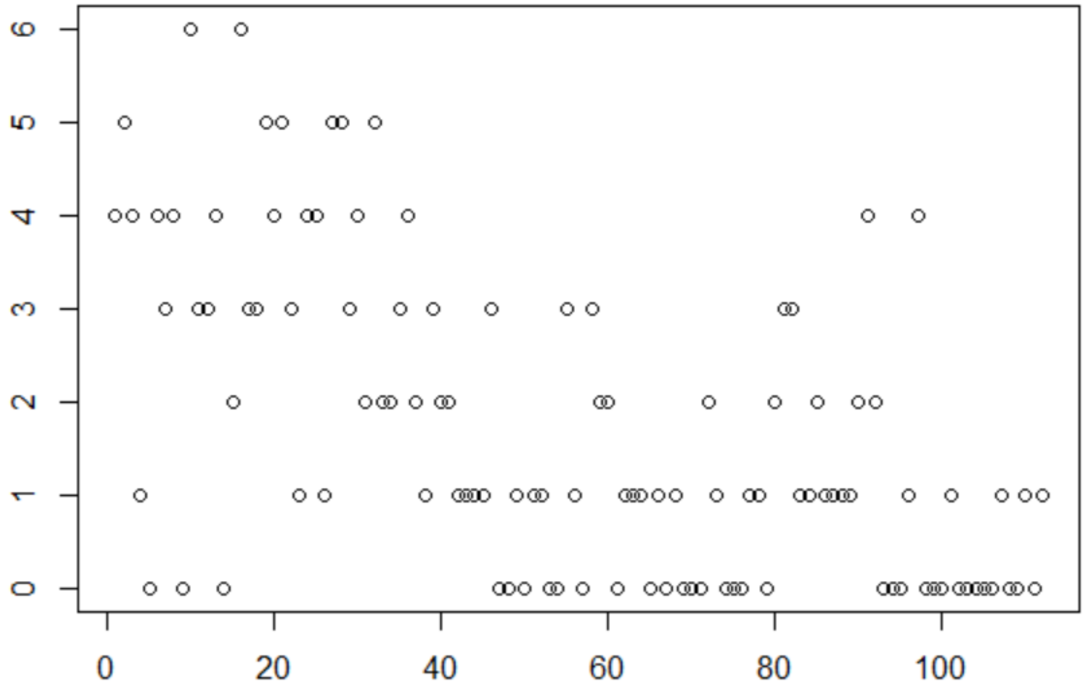
```
par(mfrow=c(1,1))
plot(lambdaM1[seq(1,N-J,by=10)],lambdaM2[seq(1,N-J,by=10)])
```

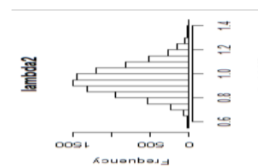


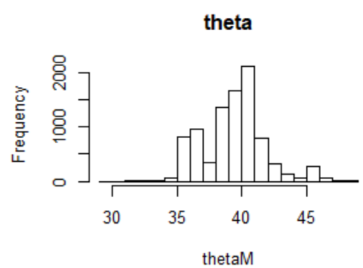




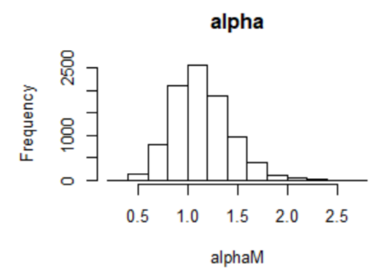

lambda1
lambdaM1
Frequency
2.0
2.5
3.0
3.5
4.0
0
1000
2500


0
1
2
3
4
5
6
0
20
40
60
80
100


lambda2
Frequency


theta
Frequency
0
1000
2000
30
35
40
45
thetaM

850


alpha
Frequency
0
1000
2500
0.5
1.0
1.5
2.0
2.5
alphaM

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

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

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

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

This gives

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

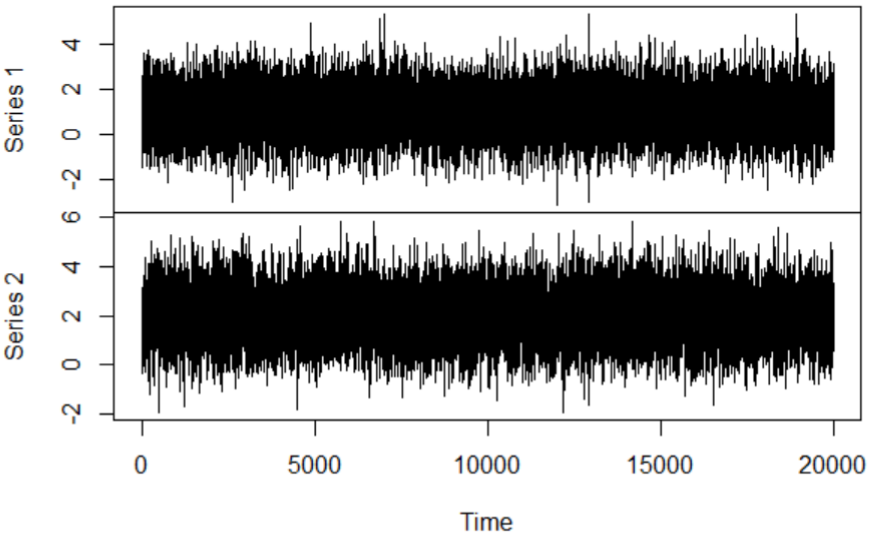
Gibbs sampler

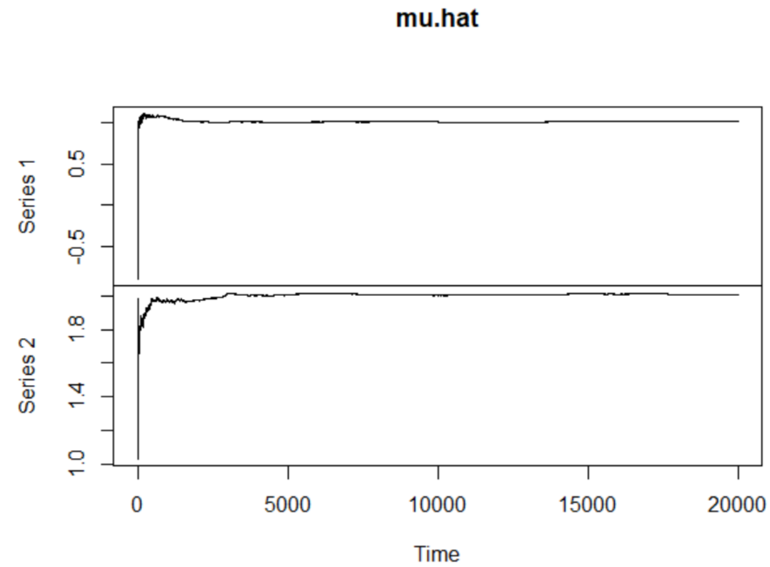
$$
\begin{aligned}
E[x_1|x_2] =& 1 + a(x_2 - 2) = 1 - 2a + ax_2 \\
\text{Var}[x_1|x_2] =& 1 - a^2 \\
E[x_2|x_1] =& 2 + a(x_1 - 1) = 2 - a + ax_1 \\
\text{Var}[x_1|x_2] =& 1 - a^2
\end{aligned}
$$

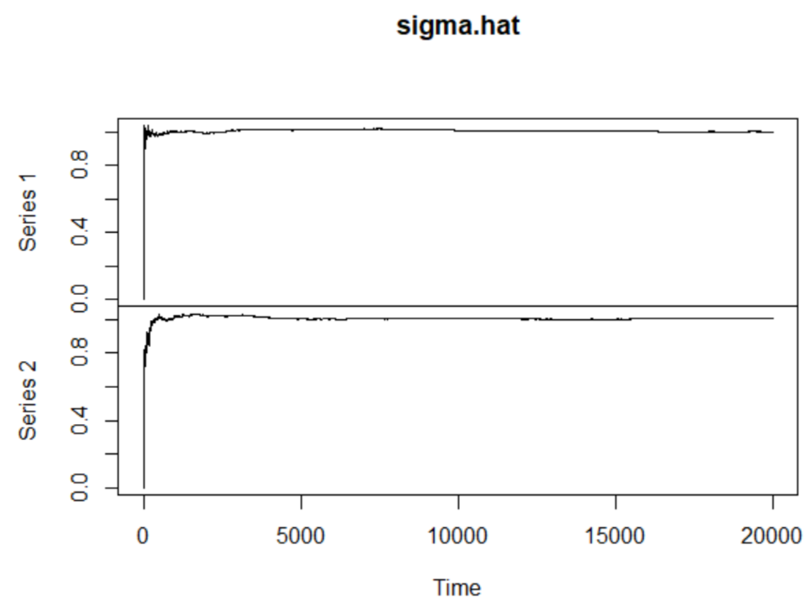
```
#Initialization
X[1,] = rnorm(2)

#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

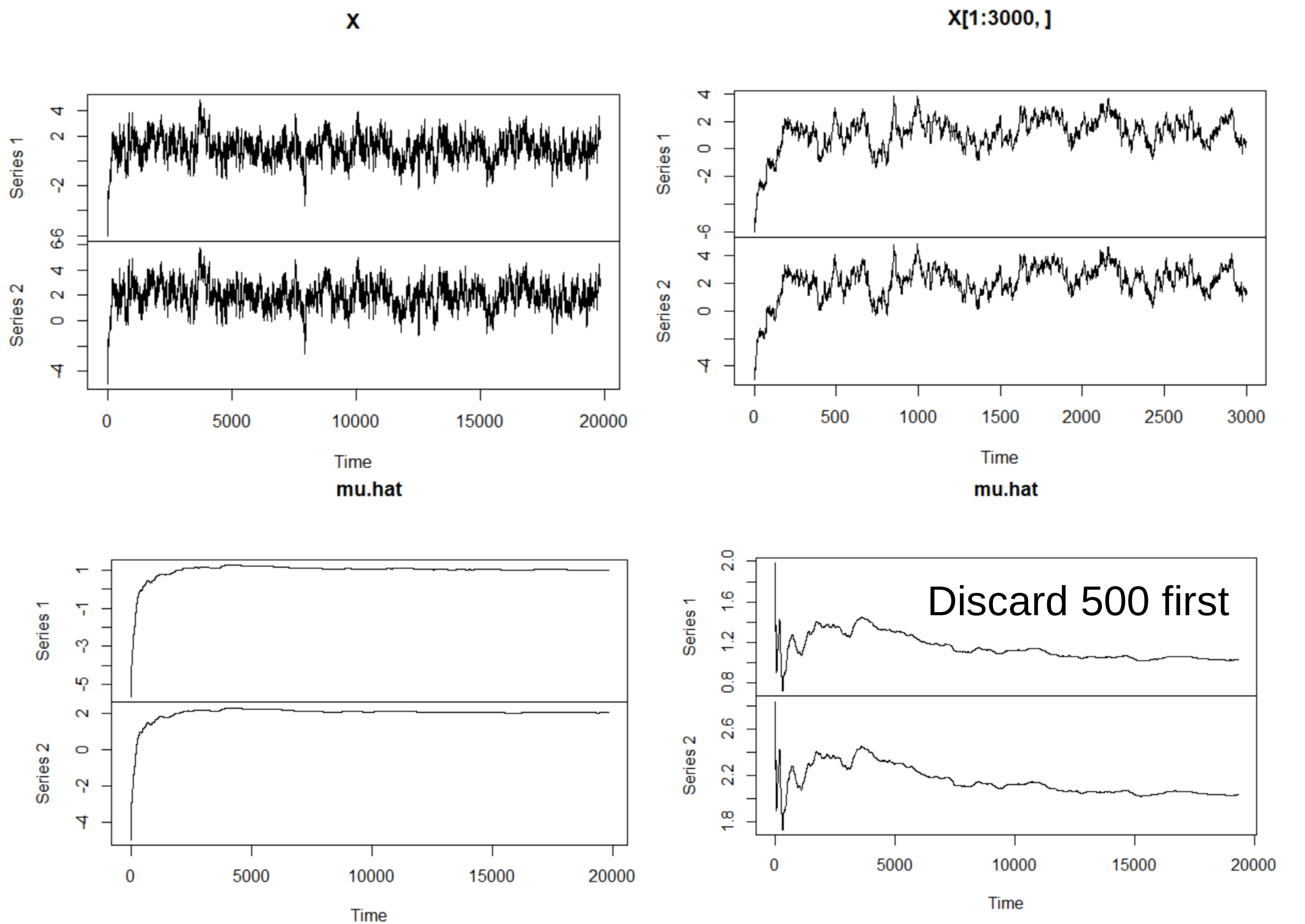


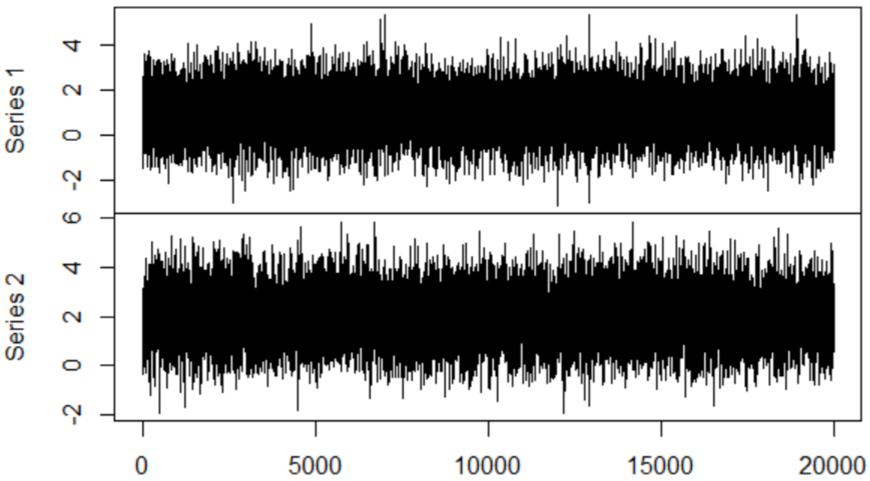




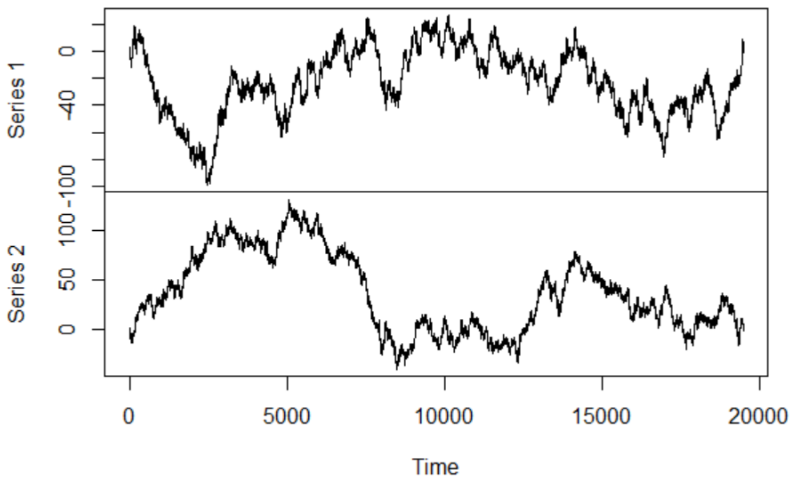
Instant convergence!

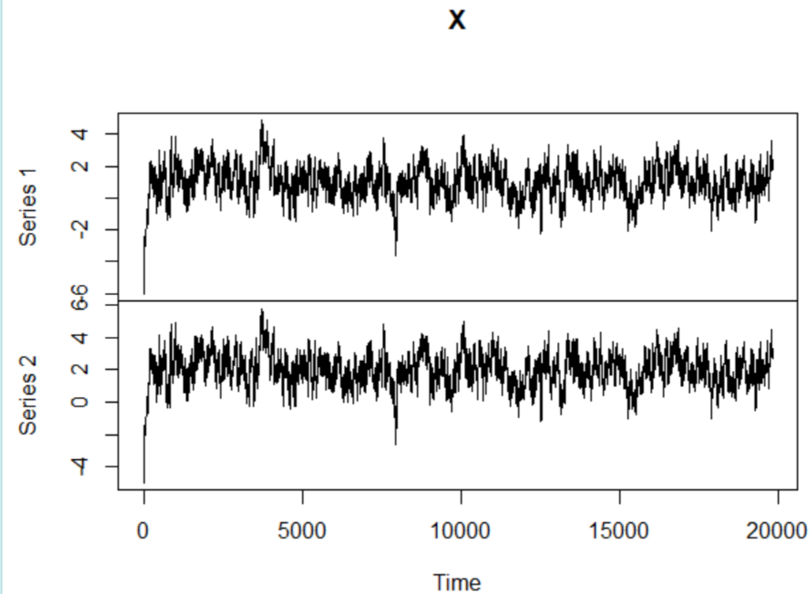
Results a=0.99





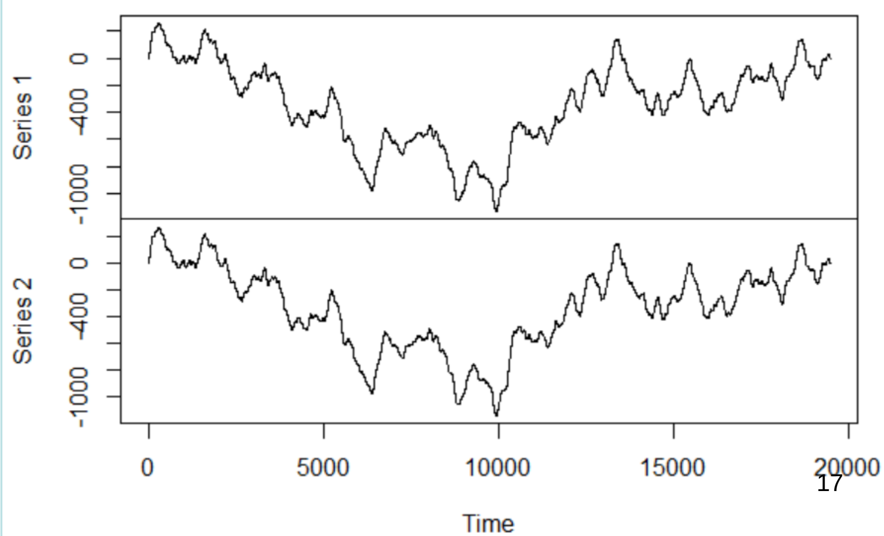
a=0.00



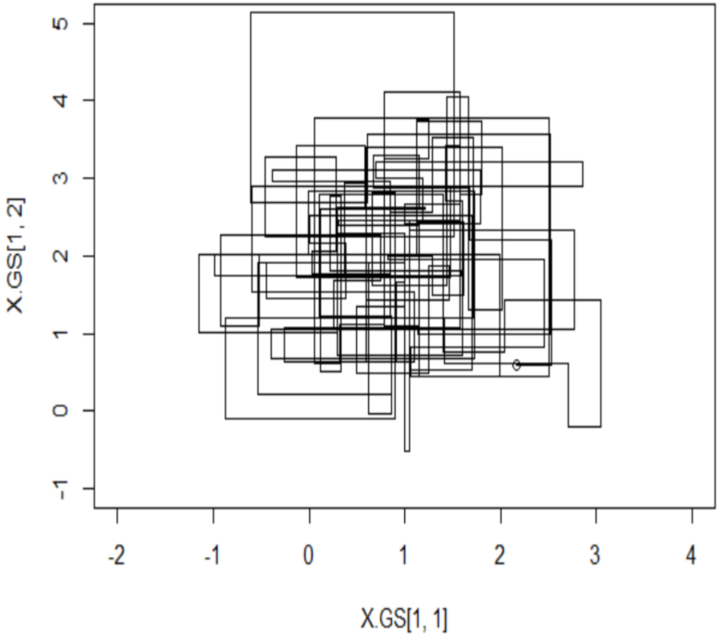


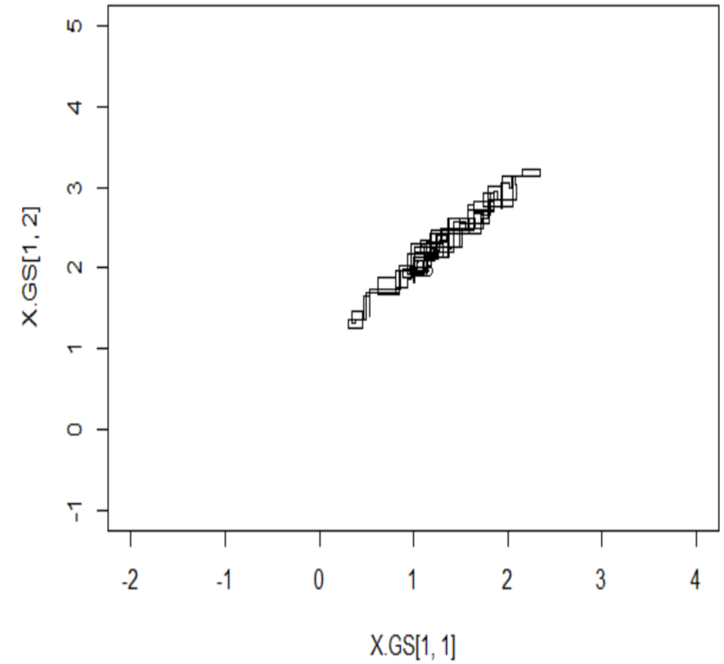
cumsumDiag

a=0.99



100 Samples in a row

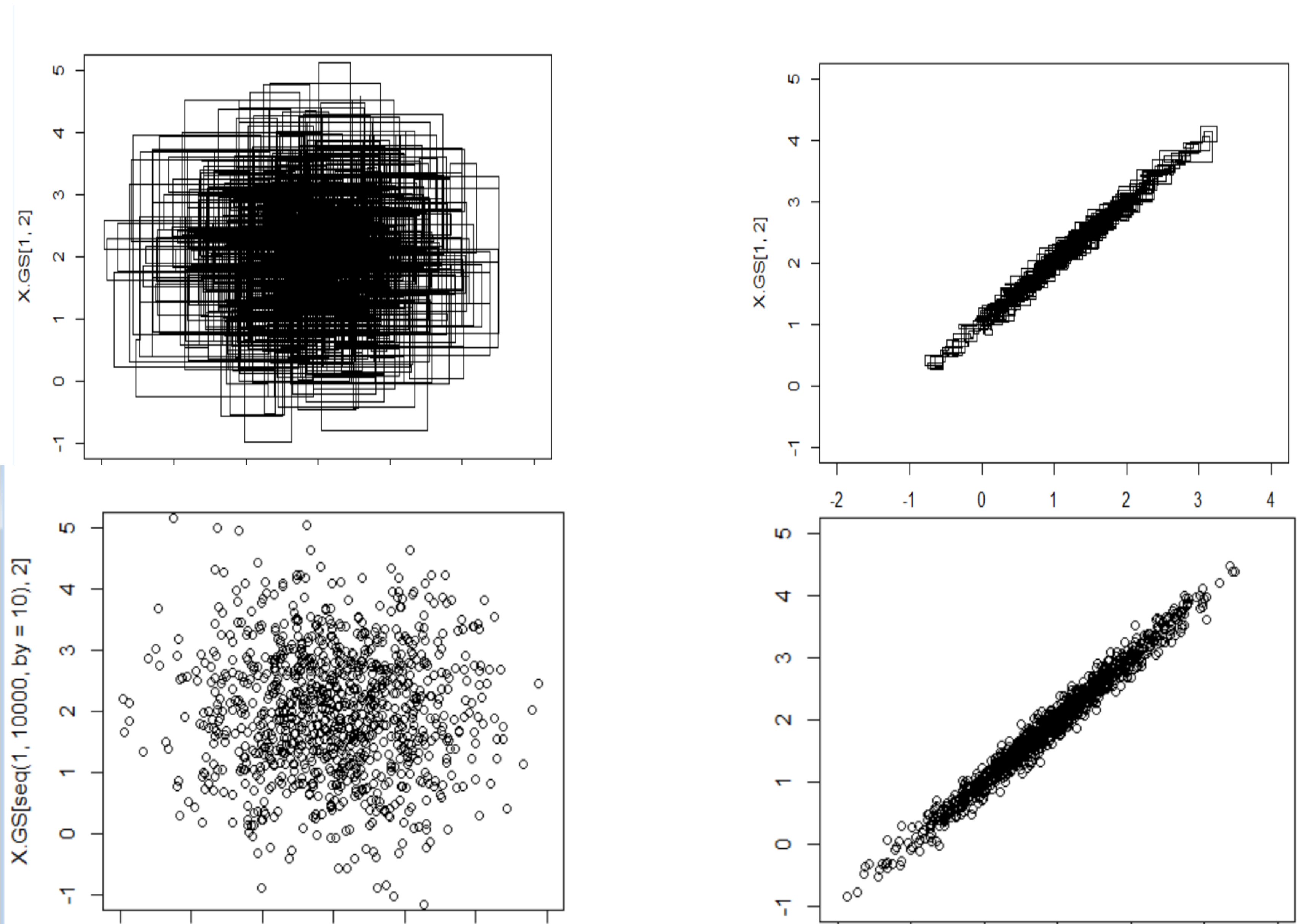




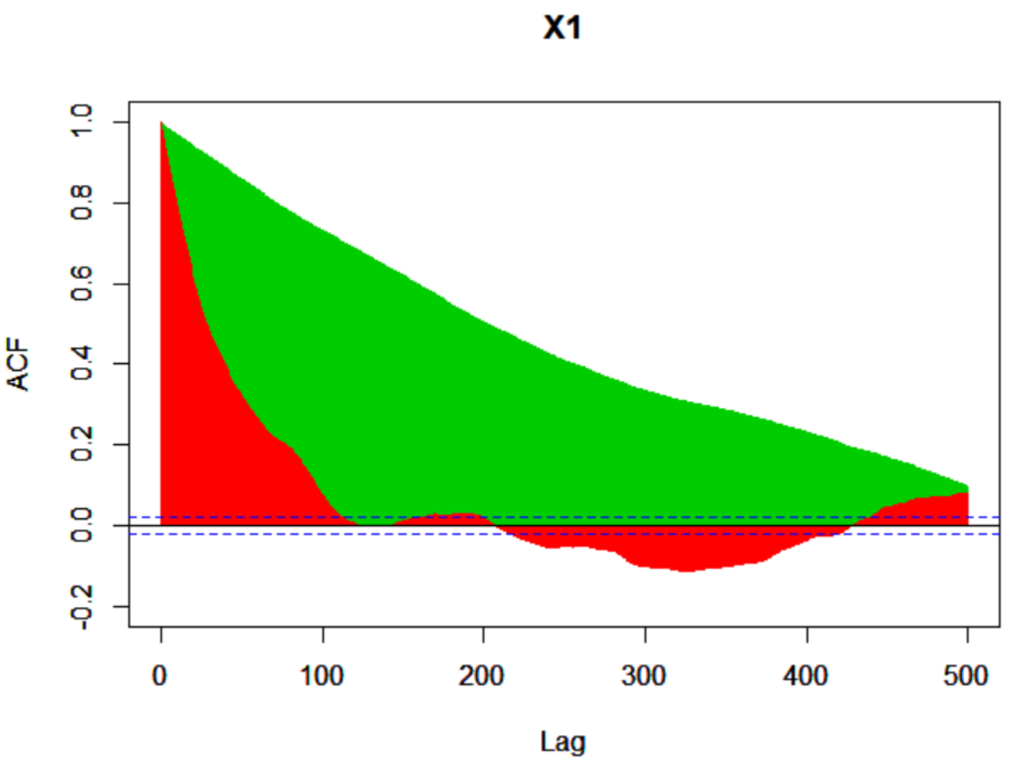
Fore update of individual coordinates

```
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
Complex code

Ex 36

- $\varepsilon_t \sim \text{Unif}([-a, a])$
- $I_t = \begin{cases} 1 \text{ if } U_t \leq \frac{h(X_t + \varepsilon_t)}{h(X_t)} \\ 0 \qquad \text{otherwise} \end{cases}$
- $X_{t+1} = X_t + I_t \cdot \varepsilon_t$

Ex 37 a

- $\boldsymbol{\varepsilon}_t \sim \text{Unif}([-a, a] \times [-b, b])$
- $I_t = \begin{cases} 1 \text{ if } U_t \leq \frac{h(X_t + \varepsilon_t)}{h(X_t)} \\ 0 \qquad \text{otherwise} \end{cases}$
- $\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + I_t \cdot \boldsymbol{\varepsilon}_t$

Ex 37 b

$\varepsilon_{1,t} \sim \text{Unif}([-a,a])$
$\varepsilon_{2,t} \sim \text{Unif}([-b,b])$

$U_{1,t} \sim \text{Unif}([0,1])$
$U_{2,t} \sim \text{Unif}([0,1])$

$$I_{1,t} = \begin{cases} 1 \text{ if } U_{1,t} \leq \dfrac{h\left(\boldsymbol{X}_t + \begin{bmatrix} \varepsilon_{1,t} \\ 0 \end{bmatrix}\right)}{h(\boldsymbol{X}_t)} \\ 0 \quad \text{otherwise} \end{cases}$$

$$X_{t+0.5} = X_t + I_{1,t} \cdot \begin{bmatrix} \varepsilon_{1,t} \\ 0 \end{bmatrix}$$

$$I_{1,t} = \begin{cases} 1 \text{ if } U_{2,t} \leq \dfrac{h\left(\boldsymbol{X}_{t+0.5} + \begin{bmatrix} 0 \\ \varepsilon_{2,t} \end{bmatrix}\right)}{h(\boldsymbol{X}_{t+0.5})} \\ 0 \quad \text{otherwise} \end{cases}$$

$$X_{t+1} = X_{t+0.5} + I_{2,t} \cdot \begin{bmatrix} 0 \\ \varepsilon_{2,t} \end{bmatrix}$$