

# Markov chain sampling

## Week3-ex1, solution

The purpose of this exercise is to study the properties of Markov chains and how they can be used to produce samples for Monte Carlo estimation.

Consider a Markov chain defined as follows:

- set  $\theta^{(0)} = C$ , where  $C$  is some constant number.
- for  $i = 1, \dots$  sample  $\theta^{(i)} \sim N(\phi\theta^{(i-1)}, \sigma^2)$  where  $\phi \in [0, 1)$  is a parameter controlling the autocorrelation between samples.

It can be shown that as  $i \rightarrow \infty$  (that is, at the limit of long chain) the *marginal* distribution of each  $\theta^{(i)}$  is a Gaussian with mean zero and variance  $\text{Var}[\theta^{(i)}] = \frac{\sigma^2}{1-\phi^2}$  and that the correlation between  $\theta^{(i)}$  and  $\theta^{(i+t)}$  is  $\text{Corr}([\theta^{(i)}, \theta^{(i+t)}]) = \phi^t$

1. What are the variance of  $\theta^{(i)}$  and the correlation between  $\theta^{(i)}$  and  $\theta^{(i+t)}$  at the limit of large  $i$  when  $\phi \rightarrow 0$  and when  $\phi \rightarrow 1$ . Assume that  $\sigma$  is fixed in these cases.
2. Fill in the below table

Var[ $\theta^{(i)}$ ]	$\phi$	$\sigma^2$	Corr[ $\theta^{(i)}, \theta^{(i+1)}$ ]
1	0.5	1	
1		0.2	
1			
1	0.1		

3. Implement the above Markov chain with R and use it to sample random realizations of  $\theta^{(i)}$  where  $i = 1, \dots, 100$  with the parameter values given in the above table. As an initial value use  $C = 10$ . Plot the sample chain and based on the visual inspection, what can you say about the convergence and mixing properties of the chain with the different choices of  $\phi$ ?
4. Choose the parameter combination where  $\sigma^2 = 0.2$  from the above table. Run three Markov chains with initial values  $C_1 = 10$ ,  $C_2 = -10$  and  $C_3 = 5$ . Find a burn-in value at which the chains have converged according to the PSRF ( $\hat{R}$ ) statistics. This is implemented in function `Rhat` in RStan. Note,  $m = 100$  samples might not be enough here.

Note! This is a Markov chain that is constructed very differently from how Stan constructs the Markov chains to sample from the posterior distributions. However, the properties related to autocorrelation and initial value are analogous.

## Solution:

### Markov chain sampling

The purpose of this exercise is to study the properties of Markov chains and how they can be used to produce samples for Monte Carlo estimation.

Consider a Markov chain defined as follows:

- set  $\theta^{(0)} = C$ , where  $C$  is some constant number.
- for  $i = 1, \dots$  sample  $\theta^{(i)} \sim N(\phi\theta^{(i-1)}, \sigma^2)$  where  $\phi \in [0, 1]$  is a parameter controlling the autocorrelation between samples.

Note! This is a Markov chain that is constructed very differently from how Stan constructs the Markov chains to sample from the posterior distributions. However, the properties related to autocorrelation and initial value are analogous.

#### Result for task 1: the limits

$\text{Var}[\theta^{(i)}] = \frac{\sigma^2}{1-\phi^2} \rightarrow \frac{\sigma^2}{1-0} = \sigma^2$  when  $\phi \rightarrow 0$  and  $\text{Var}[\theta^{(i)}] = \frac{\sigma^2}{1-\phi^2} \rightarrow \frac{\sigma^2}{1-1} = \infty$  when  $\phi \rightarrow 1$ .

$\text{Corr}([\theta^{(i)}, \theta^{(i+t)}]) = \phi^t \rightarrow 0^t = 0$  as  $\phi \rightarrow 0$  and  $\text{Corr}([\theta^{(i)}, \theta^{(i+t)}]) = \phi^t \rightarrow 1^t = 1$  as  $\phi \rightarrow 1$ .

#### Result for task 2: the table

Given the marginal variance for  $\theta^{(i)}$  we can solve for  $\phi$  and  $\sigma^2$  when the other is given

$$\phi = \sqrt{1 - \sigma^2 / \text{Var}(\theta^{(i)})}$$

$$\sigma^2 = \text{Var}(\theta^{(i)})(1 - \phi^2)$$

```
varTheta = 1
sigma2.1 = 1
sigma2.3 = 0.2
phi.2 = 0.5
phi.4 = 0.1
phi.1 = sqrt(1-sigma2.1/varTheta)
phi.3 = sqrt(1-sigma2.3/varTheta)
sigma2.2 = varTheta*(1-phi.2^2)
sigma2.4 = varTheta*(1-phi.4^2)
table.entries = matrix(nrow=4, ncol=4, data=c(
  varTheta, phi.1, sigma2.1, phi.1,
  varTheta, phi.2, sigma2.2, phi.2,
  varTheta, phi.3, sigma2.3, phi.3,
  varTheta, phi.4, sigma2.4, phi.4
))
table.entries <- t(table.entries) # take transpose since matrix fills in the elements in columnwise
colnames(table.entries) <- c("var(theta)", "phi", "sigma2", "corr")
print(table.entries)
```

##	var(theta)	phi	sigma2	corr
## [1,]	1	0.0000000	1.00	0.0000000
## [2,]	1	0.5000000	0.75	0.5000000
## [3,]	1	0.8944272	0.20	0.8944272
## [4,]	1	0.1000000	0.99	0.1000000

#### Result for task 3

Let's Implement the above Markov chain with R and use it to sample random realizations of  $\theta^{(i)}$  where  $i = 1, \dots, 100$  with the parameter values given in the above table. As an initial value use  $C = 10$ . Plot

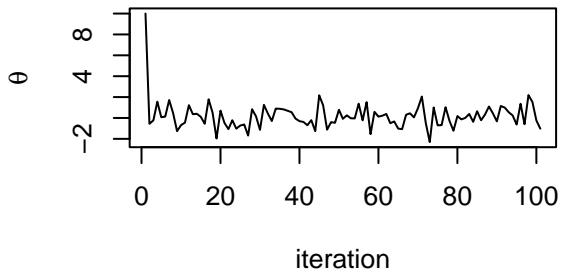
the sample chain and based on the visual inspection, what can you say about the convergence and mixing properties of the chain with the different choices of  $\phi$ ?

```
set.seed(123)
# let's first define a function to conduct the sampling
MarkovChain <- function(phi,sigma2,initial,m){
  theta = vector(length=m)
  theta[1] = initial
  for (i1 in seq(1,m,1)){
    theta[i1+1] = phi*theta[i1] + rnorm(1,0,sqrt(sigma2))
  }
  return(theta)
}

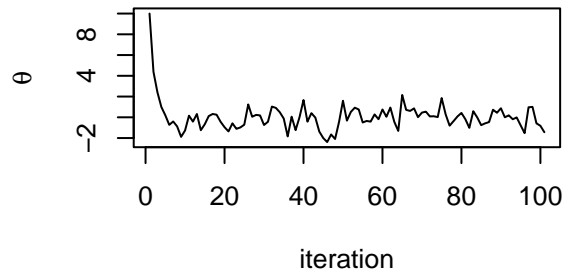
# Then we can sample from the Markov chain with the alternative phi and sigma values
m = 100
initial = 10
par(mfrow=c(2,2)) # Open figure for plotting the examples
for (i1 in c(1,2,3,4)){
  theta = MarkovChain(table.entries[i1,"phi"],table.entries[i1,"sigma2"],initial,m)

  plot(theta, type="l", xlab="iteration", ylab=expression(theta),
        main=sprintf("sigma2=%.1f, phi=%.2f",
                      table.entries[i1,"sigma2"],table.entries[i1,"phi"]))
}
```

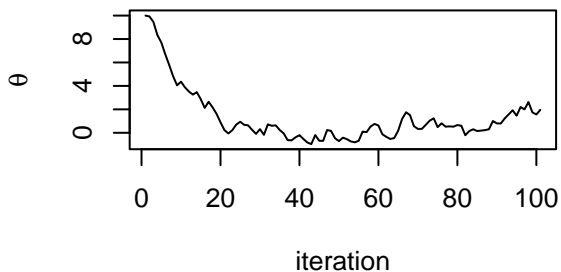
**sigma2=1.0, phi=0.00**



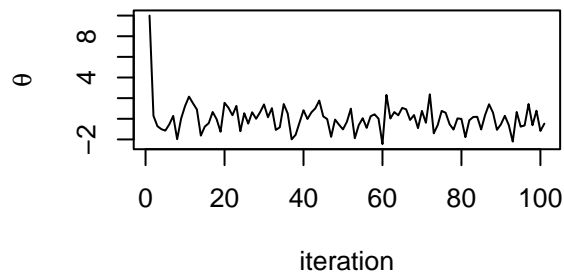
**sigma2=0.8, phi=0.50**



**sigma2=0.2, phi=0.89**



**sigma2=1.0, phi=0.10**



#### Result for task 4

Choose the parameter combination where  $\sigma^2 = 0.2$  from the above table. Run three Markov chains with initial values  $C_1 = 10$ ,  $C_2 = -10$  and  $C_3 = 5$ . Find a burn-in value at which the chains have converged

according to the PSRF ( $\hat{R}$ ) statistics. This is implemented in function `Rhat` in `RStan`. Note,  $m = 100$  samples might not be enough here.

```
# Define the sample size
m = 500

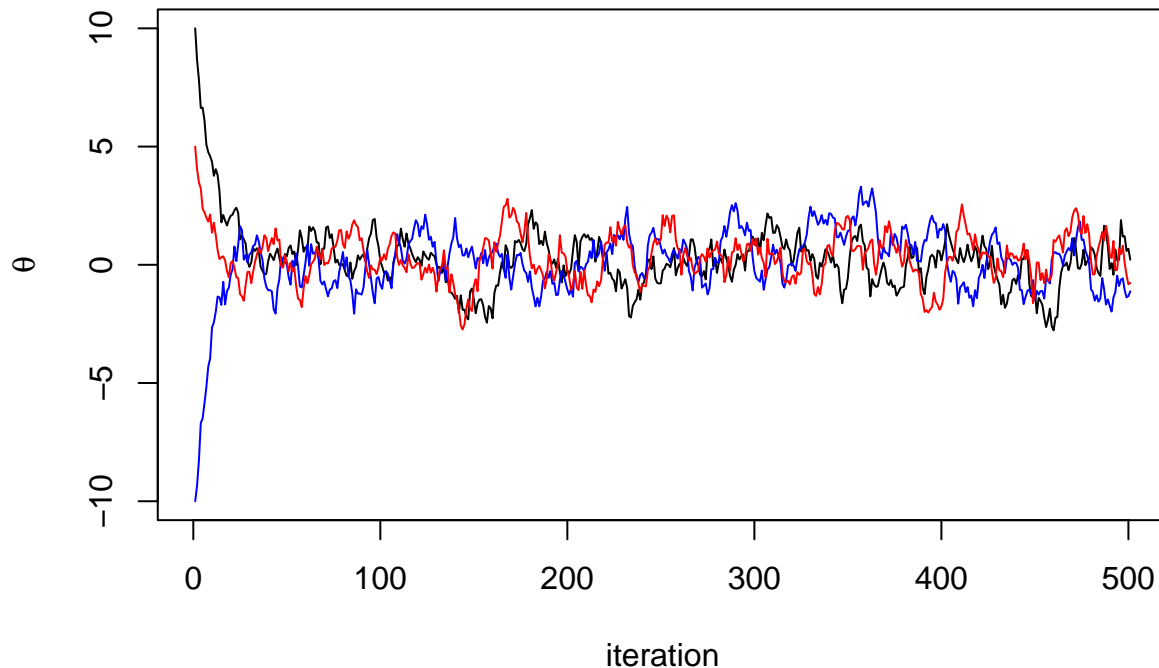
# let's sample from three independent Markov chains as instructed in the exercise
set.seed(1)
theta1 = MarkovChain(table.entries[3,"phi"],table.entries[3,"sigma2"],10,m)
set.seed(2)
theta2 = MarkovChain(table.entries[3,"phi"],table.entries[3,"sigma2"],-10,m)
set.seed(3)
theta3 = MarkovChain(table.entries[3,"phi"],table.entries[3,"sigma2"],5,m)

# Let's examine the chains visually
plot(theta1, type="l", xlab="iteration", ylab=expression(theta), main=sprintf("sigma2=%.1f, phi=%.2f", ,
lines(theta2, col="blue")
lines(theta3, col="red")

# It seems the chains have converged to the same stationary distribution
# after 50 steps. Hence, let's try PSRF
library(rstan)           # First, load the necessary tool - that is Rhat function

## Loading required package: StanHeaders
## Loading required package: ggplot2
## rstan (Version 2.21.2, GitRev: 2e1f913d3ca3)
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
```

**sigma2=0.2, phi=0.89**



```
# ?Rhat
THETA = rbind(theta1,theta2,theta3) # Put the chains into list where each row is one Markov chain
Rhat(t(THETA[,50:m])) # note we need transpose since Rhat assumes the chains are in columns

## [1] 1.014315

# if Rhat value is greater than 1.05 try to remove more samples from the beginning. For example,
Rhat(t(THETA[,100:m])) # note we need transpose since Rhat assumes the chains are in columns

## [1] 1.013971

# Or directly through cbind
THETA = cbind(theta1,theta2,theta3) # Put the chains into list where each column is one Markov chain
Rhat(THETA[50:m,])

## [1] 1.014315

# if Rhat value is greater than 1.05 try to remove more samples from the beginning. For example,
Rhat(THETA[100:m,])

## [1] 1.013971

# However, note that you may need to increase the sample size in order to get
# reliable estimate for PSRF. Note also that the Rhat values change each time
# you rerun the Markov chains. Hence, even if with one realization the PSRF
# looked fine it might not be so in another even with same number of samples.
```

## Grading:

**Total 10 points:** 2 points for correct answer for step 1. 3 points for correct answer to step 2. 3 points for correct answer for step 3. 2 points for correct answer for step 4. **Note,** You should not penalize from wrong parameter values in step 3 and 4 if the table was filled wrong in step 2.