Markov chain sampling

Week3-ex1, problem statement

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

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 $\phi \to 0$ and when $\phi \to 1$. Assume that σ is fixed in these cases.
2. Fill in the below table

$\mathrm{Var}[\theta^{(i)}]$	ϕ	σ^2	$\mathrm{Corr}[\theta^{(i)}, \theta^{(i+1)}]$
1		1	
1	0.5		
1		0.2	
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 ϕ?
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.

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.