Kyle Knudson

11/27/17

Stats 417

**Project # 3**

**Part 1: Problem description**

The goal of this project was to utilize R to simulate a Markov chain consisting of the states of Sunny, Cloudy and Rainy. The Probability of a sunny day followed by two cloudy days was determined from this simulation, along with the case that given today was rainy, the probability the sun will shine the day after tomorrow, was calculated. These probabilities were extrapolated by utilizing the transition matrix that is provided in exercise 9.3.1.

**Part 2: Discussion of results:**

**Part a:**

For this part of the problem, the theoretical value of a sunny day followed by two cloudy days was 0.100. When I ran the simulation for these states, on various different values for n (the population of the sample size), as N grew larger the experiment value of a sunny day followed by two cloudy days grew closer and closer to the theoretical value for this simulation. The experimental values for this simulation were obtained across several runs were as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| N - value | 100 | 1000 | 10000 | 100000 |
| Probability | 0.08333333 | 0.082 | 0.10349 | 0.09960 |

These values were very close to the theoretical value and therefore indicate a successful simulation. The last value that was produced with an N-value of 100000 yielded a probability of 0.09960 which is extremely close to the theoretical value of 0.1. This low error result demonstrates that the larger the sample sized used will push the probability towards the theoretical value. This simulation could be improved by favoring the early states, as it would provide a more accurate picture of what might occur in reality.

**Part b:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| N-Value | 100 | 1000 | 10000 | 100000 |
| Probability | 0.3181818 | 0.3871 | 0.34128 | 0.36393 |

For this part of the problem, the theoretical value of the probability of the case that given today is rainy, that the day after tomorrow will be sunny is 0.36. When running my simulations for this case, at low values of N, the probability was inaccurate and inconsistent. However, as the value for N that was used increased, the probability that was determined from the simulation approached the theoretical value of 0.36. The last simulation run with an N-value of 100000 had the closest value of 0.36393 which yields a very small error.

**Part 3**: **Code**

P = matrix(c(0.7, 0.3, 0.2, 0.2, 0.5, 0.6, 0.1, 0.2, 0.2), 3)

print(P)

x = c("S", "C", "R") ## The weather state space

n = 1000000 ## We plan to generate 10000 states one for each of the 10000 consecutive days.

states = character(n+100) ## Intentionally generate 100 more and the first 100 (called burn-in's) will be abandoned in order to generate a quality sequence.

states[1] = "C" ## Set the initial state to any of the states, say "C" here. If the steady-state distribution is available, we can generate the initial state according to this distribution.

for (i in 2:(n+100)){

if (states[i-1] == "S") {cond.prob = P[1,]}

else if (states[i-1] == "C") {cond.prob = P[2,]}

else {cond.prob = P[3,]}

states[i]=sample(x, 1, prob = cond.prob )

}

print(states[1:100])

states = states[-(1:100)] ## Abandon the bur-in's

head(states)

tail(states)

states[1:200]

#part a

collapsed <- paste(states, collapse="")

length(gregexpr("S", collapsed)[[1]])

length(gregexpr("SCC", collapsed)[[1]])

prob =(length(gregexpr("SCC", collapsed)[[1]])) / length(gregexpr("S", collapsed)[[1]])

prob

#part b

collapsed <- paste(states, collapse="")

length(gregexpr("R", collapsed)[[1]])

length(gregexpr("RRS", collapsed)[[1]])

length(gregexpr("RSS", collapsed)[[1]])

length(gregexpr("RCS", collapsed)[[1]])

prob2 =((length(gregexpr("RRS", collapsed)[[1]])) + length(gregexpr("RSS", collapsed)[[1]]) + length(gregexpr("RCS", collapsed)[[1]]) ) / length(gregexpr("R", collapsed)[[1]])

prob2