WST312 Stochastic Processes 2018
Practical 6

Simulating a Markov Process and Estimating Parameters

We use Monte Carlo simulation. This is done via the following procedure:
Step 1: Generate an observation from a uniform(0,1) distribution, say u.
Step 2: Generate an observation x from the Markov chain, i.e. $x = F^{-1}(u)$ where $F(x)$ is the distribution function for the Markov chain.

Consider a time-homogeneous Markov chain $\{X_t : t = 0,1,2,...\}$ with state space $S = \{1,2,...m\}$. Note that the transition matrix P gives the probabilities for the process and so specifies the distribution function (row i (corresponding to state i) of P gives the conditional distribution of X_t given that $X_{t-1} = i$). If the process is in state i at time $t-1$, the simulated value u needs to specify one of the states (i.e. the state j for time t) so we divide the interval [0,1] into m sub-intervals (one for each state) of lengths proportional to the probabilities given by row i of P. The interval the observed value u falls into specifies the state j.

Consider the following one-step transition matrix with state space $S = \{1,2,3\}$

$$P = \begin{bmatrix} 0.45 & 0.05 & 0.5 \\ 0.21 & 0.24 & 0.55 \\ 0.33 & 0.31 & 0.36 \end{bmatrix}.$$

a) Write code to generate a random initial state for the process if $q_1 = 0.1$, $q_2 = 0.55$ and $q_3 = 0.35$. (Use a seed of 1.)
b) Using the initial state obtained in question (a) write code to generate 15 observations (excluding the initial state generated in (a) for this Markov chain. Use a seed of 1 and print the values obtained.
c) Write code to find the values of n_{ij} (the number of transitions from a state i to a state j) and n_i (the number of transitions from state i to some state (note the last generated state doesn't go anywhere), and thus the maximum likelihood estimates $\hat{p}_{ij}$ for the p_{ij}'s. Note that $\sum_i n_i$ should equal 15 and not 16. Compare the estimates to the original matrix P and comment.
d) Run your code again but produce 100 observations for the process instead of 15. How do your estimates differ?
e) Run your code again but produce 1000 observations for the process instead of 100. How do your estimates differ?

For all code mentioned above the question should be done using SAS proc iml as well as in R