STAT 753: Stochastic Models and Simulations

Jaleesa Houle
University of Nevada, Reno - Spring 2024

Homework 4

Theory

1. Find the probability that starting from 2, we hit 1 earlier than 3, in the Markov chain with transition matrix

$$P = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

.

Solution

The probability that 1 will be hit earlier than 3 if starting from 2 can be found by using the system of equations

$$\begin{aligned} p_1 &= 1 \\ p_3 &= 0 \\ p_2 &= \frac{1}{3}p_1 + \frac{1}{3}p_2 + \frac{1}{3}p_3 \end{aligned} \tag{1}$$

Solving this set of equations for p_2 gives a value of 1/2, so the probability of hitting 1 before 3 when starting from 2 with the given transition matrix is 50%.

For problems 2-3, consider the Markov chain with transition matrix

$$\text{A} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0.2 & 0.3 \\ 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

2. Find the recurrent and transient states.

Solution

The recurrent states in this Markov chain are x_1 and x_3, and the transient states are x_2 and x_4. This is obvious by inspection of A, but can also be shown by finding the stationary distribution.

$$\pi = \pi A$$

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0.2 & 0.3 \\ 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{bmatrix} = \begin{bmatrix} 0.5\pi_1 + 0.25\pi_2 + \pi_3 + 0.25\pi_4 \\ 0.25\pi_2 + 0.25\pi_4 \\ 0.5\pi_1 + 0.2\pi_2 + 0.25\pi_4 \\ 0.3\pi_2 + 0.25\pi_4 \end{bmatrix}^T$$

$$\begin{Bmatrix} \pi_1 = 0.5\pi_1 + 0.25\pi_2 + \pi_3 + 0.25\pi_4 \\ \pi_2 = 0.25\pi_2 + 0.25\pi_4 \\ \pi_3 = 0.5\pi_1 + 0.2\pi_2 + 0.25\pi_4 \\ \pi_4 = 0.3\pi_2 + 0.25\pi_4 \end{Bmatrix}$$

$$\begin{Bmatrix} \pi_1 = 2/3 \\ \pi_2 = 0 \\ \pi_3 = 1/3 \\ \pi_4 = 0 \end{Bmatrix}$$

It can be seen that the transient states x_2 and x_4 have a distribution of 0 as $N => \infty$, since once they leave those states they cannot return. The Markov chain with only recurrent states x_1 and x_3 is irreducible and aperiodic, therefore it is ergodic and will converge to the stationary distribution.

3. Find the average time spent in state 2 if we start from 4.

Solution

The average time spent in state 2 if starting from state 4 can be found by using the relationship $M = P_T M + I$, where M is a square matrix of m_{ij} entries, each corresponding to the mean time spent at state j if starting from state i, P_T is the reduced Markov chain transition matrix that contains only probabilities of transient states, and I is the identity matrix of size nxn. where n is the number of transient states. Rearranging this formula, we get $M = (I - P_T)^{-1}$. Solving for the transient states x_2 and x_4 in matrix A, we find

$$\begin{aligned}
\begin{bmatrix} m_{22} & m_{24} \\ m_{42} & m_{44} \end{bmatrix} &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} p_{2\to2} & p_{2\to4} \\ p_{4\to2} & p_{4\to4} \end{bmatrix} \right)^{-1} \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.25 & 0.3 \\ 0.25 & 0.25 \end{bmatrix} \right)^{-1} \\
&= \left(\begin{bmatrix} 0.75 & -0.3 \\ -0.25 & 0.75 \end{bmatrix} \right)^{-1} \\
&\approx \begin{bmatrix} 1.5385 & 0.6154 \\ 0.5128 & 1.5385 \end{bmatrix}
\end{aligned} \tag{3}$$

Thus, the average time spent in state 2 if starting from state 4 is $m_{42} = 0.5128$.

Coding

1. Simulate the Markov chain with transition matrix P 1000 times, starting from 2, until it hits 1 or 3. Find the empirical probability that it hits 1 before 3. Compare with the answer from Theory 1.

Solution

```python
import numpy as np

def MC_sim_1(P,N,x0): #function for 1 simulation
    sim = []
    for i in range(1,N):
        random_x = np.random.choice(np.arange(1,4), p = P[x0-1])
        if random_x != 2:
            sim.append(random_x)
            break
    return sim

#givens
P = np.array([[0.25,0.25,0.5],[(1/3), (1/3), (1/3)],[.5, .5, 0]])
x0 = 2
N = 1000

np.random.seed(1)
sim_list = np.empty((N, 0)).tolist()
for i in range(0,N):
    sim_list[i] = MC_sim_1(P,N,x0) #run simulation N times

unique_states, counts = np.unique(sim_list, return_counts=True)

print('Probability of hitting 1 or 3 first:', counts/N)

>> Probability of hitting 1 or 3 first: [0.491 0.509]
```

The theoretical probability of hitting 1 before 3 was found in question 1 to be 1/2. From 1000 simulations, the empirical probability was found to be 0.491. The percent error between the empirical and theoretical solutions is approximately 1.8%, demonstrating that running 1000 simulations gives a fairly accurate result.

2. Simulate the Markov chain with transition matrix A 1000 times, starting from 4, until it leaves transient states forever. Find the empirical average of number of steps spent at state 2. Compare with the answer from Theory 3.

Solution

```python
def MC_sim_2(P,N,x0, recurrent_states):
    sim = []
    sim.append(x0)
    for i in range(1,N):
        random_x = np.random.choice(np.arange(1,len(P)+1), p = P[sim[i-1]-1])
        if random_x not in recurrent_states:
            sim.append(random_x)
        else:
            break
    return sim

A = np.array([[0.5,0,0.5,0],[.25, .25, 0.2, 0.3],[1,0,0,0],[.25,.25,.25,.25]])
x0= 4
N=1000
recurrent_states = [1,3]

np.random.seed(1)
sim_list = np.empty((N, 0)).tolist()
for i in range(0,N):
    sim_list[i] = MC_sim_2(A,N,x0,recurrent_states)

count_2= []
for i,numbers in enumerate(sim_list):
    count_2.append(numbers.count(2))

print('Mean steps spent at state 2:', np.mean(count_2))

>> Mean steps spent at state 2: 0.487
```

The theoretical average time spent in state 2 if starting from state 4 was found in question 3 to be 0.5128. From 1000 simulations, the empirical average was found to be 0.487. The percent error between the empirical and theoretical solutions is approximately 5.03%, demonstrating that running 1000 simulations gives some slight variability in the average number of steps, but is still close to the theoretical solution.