Sam Chen
EE 381
013502214
12/11/20
Lab 6

PROBLEM 1. A three-state Markov Chain

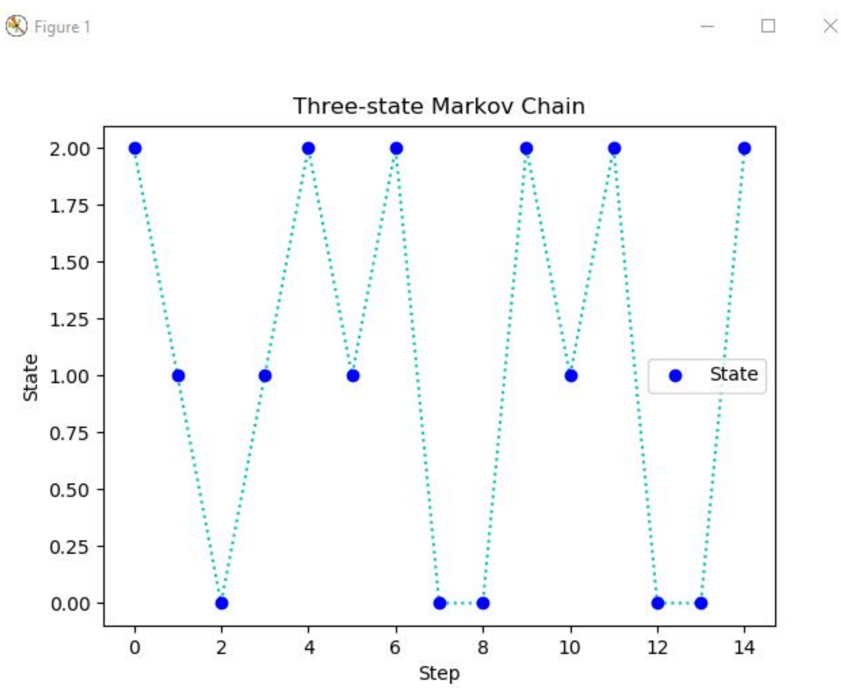
Follow the previous code as a guide, and simulate the Markov chain defined in problem 11.1 of the AMS text by Grinstead & Snell, p. 406 (reference [1]).

For this problem use the provided values of the state transition matrix and the initial probability distribution.

1. Run each experiment for $n = 15$ time steps. In order to obtain meaningful statistical data perform a total of $N = 10,000$ experiments.
2. After your experimental simulations are completed, use the State Transition Matrix approach and compare the results.

SUBMIT a report with the results and your code. You must follow the guidelines given in the syllabus regarding the structure of the report. Points will be taken off, if you do not follow the guidelines. The report should contain:

- One plot showing one single-simulation run for $n = 15$ steps, similar to Figure 0.3, but for a three-state chain. Make sure that your plot has the appropriate labels and title.
- A plot with the simulated probabilities for the states R, N, S, similar to Figure 0.5
- A plot with the probabilities obtained through the state transition matrix approach for the three states R, N, S, similar to Figure 0.6
- The code in an Appendix
- **Make sure that all the plots are properly labeled**



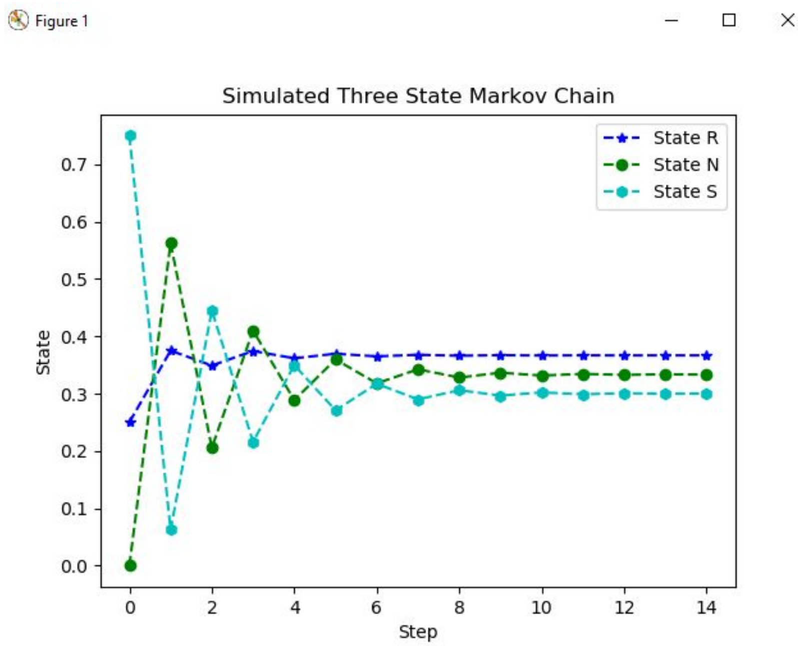


Figure 1
Simulated Three State Markov Chain
State R
State N
State S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
State
0
2
4
6
8
10
12
14
Step

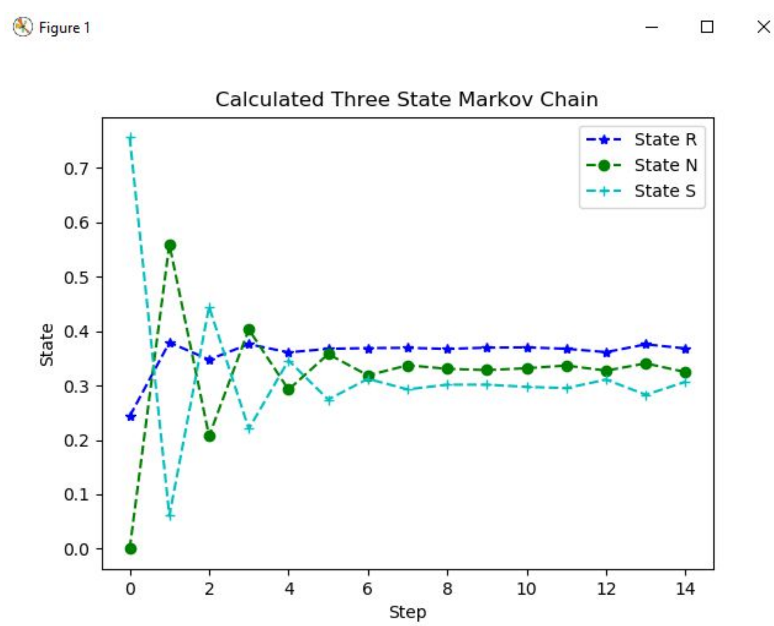


Figure 1
Calculated Three State Markov Chain
State R
State N
State S
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
State
0
2
4
6
8
10
12
14
Step

PROBLEM 2. The Google PageRank Algorithm

This problem presents an introduction to the algorithm used for ranking web pages for searching purposes. The algorithm was developed by Google's founders Page and Brin with Motwani and Winograd, and was first published in 1998 (see reference [2]). The *PageRank* algorithm allowed *Google* to rise to the top of all web search engines within a matter of months after its implementation, outperforming the established search engines of the time such as *AltaVista* and *Yahoo*. The algorithm is based on the theory of Markov chains, utilizing the information on the number of links leading to an existing webpage.

The current problem uses a simplified web in order to show how the algorithm works (see reference [3]). The simplified web consists of 5 pages only, and it is shown schematically in Figure 2.1.

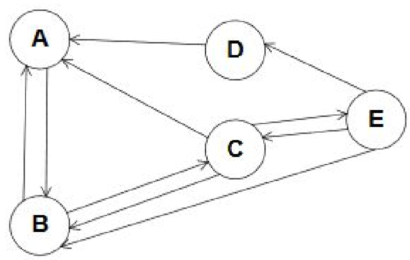


Figure 2.1: A five-page web

A Markov chain model of an *impartial surfer* visiting web pages is created as following:

- At time k the surfer is on page X, where $X \in \{A, B, C, D, E\}$.
- At time $(k+1)$ the surfer will move randomly to another page Y, which must be linked to X.
- The state S_k of the Markov chain at time k is defined as the page which the surfer is visiting at time k: $S_k \in \{A, B, C, D, E\}$.
- To create the Markov chain model, the algorithm assumes that all pages that can be reached from X have the same probability to be visited by the surfer. Thus:

$$\text{Prob}\left(S_{k+1}=Y \mid S_k=X\right)=\begin{cases}\dfrac{1}{(\text{Total number of links leaving page X})} & , \quad \text{if } Y \text{ can be reached from } X \\ 0 & , \quad \text{if } Y \text{ cannot be reached from } X\end{cases}$$

- For example: Prob $(S_{k+1}=A \mid S_k=C)=\frac{1}{3}$; Prob $(S_{k+1}=E \mid S_k=D)=0$; etc.

- Based on these probabilities the State Transition Matrix (P) of the Markov chain is constructed. The left eigenvector w of P corresponding to the eigenvalue $\lambda=1$ (which is a *fixed vector* of the Markov chain) is computed from: $wP=w$
- The *PageRank* algorithm uses the vector w to rank the pages of the five-page web in terms of visiting importance.

To complete this problem follow the steps below:

1. Create the 5×5 State Transition Matrix P with the values of the transition probabilities
2. Assume that initially all pages are equally likely to be visited, i.e. use the initial state probability distribution vector: $v_1=\begin{bmatrix}A & B & C & D & E\end{bmatrix}=\begin{bmatrix}\frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5}\end{bmatrix}$
3. Calculate the probability vectors for $n=1,2,\ldots 20$ steps, **using the State Transition matrix only.** Note that *you ARE NOT asked to run the complete simulation* of the chain. You are only asked to use the State Transition matrix approach to calculate the probability vectors for n steps.
4. Rank the pages $\{A,B,C,D,E\}$ in order of importance based on the results from the previous step
5. Create a plot of the calculated state probabilities for each of the five states $\{A,B,C,D,E\}$ vs. the number of steps for $n=1,2,\ldots 20$. This should be similar to the plot in Figure 0.6
6. Assume that the surfer always starts at his/her home page, which for the purposes of this problem is page E. Repeat steps (2)-(5) with the new state probability distribution vector: $v_2=\begin{bmatrix}A & B & C & D & E\end{bmatrix}=\begin{bmatrix}0 & 0 & 0 & 0 & 1\end{bmatrix}$

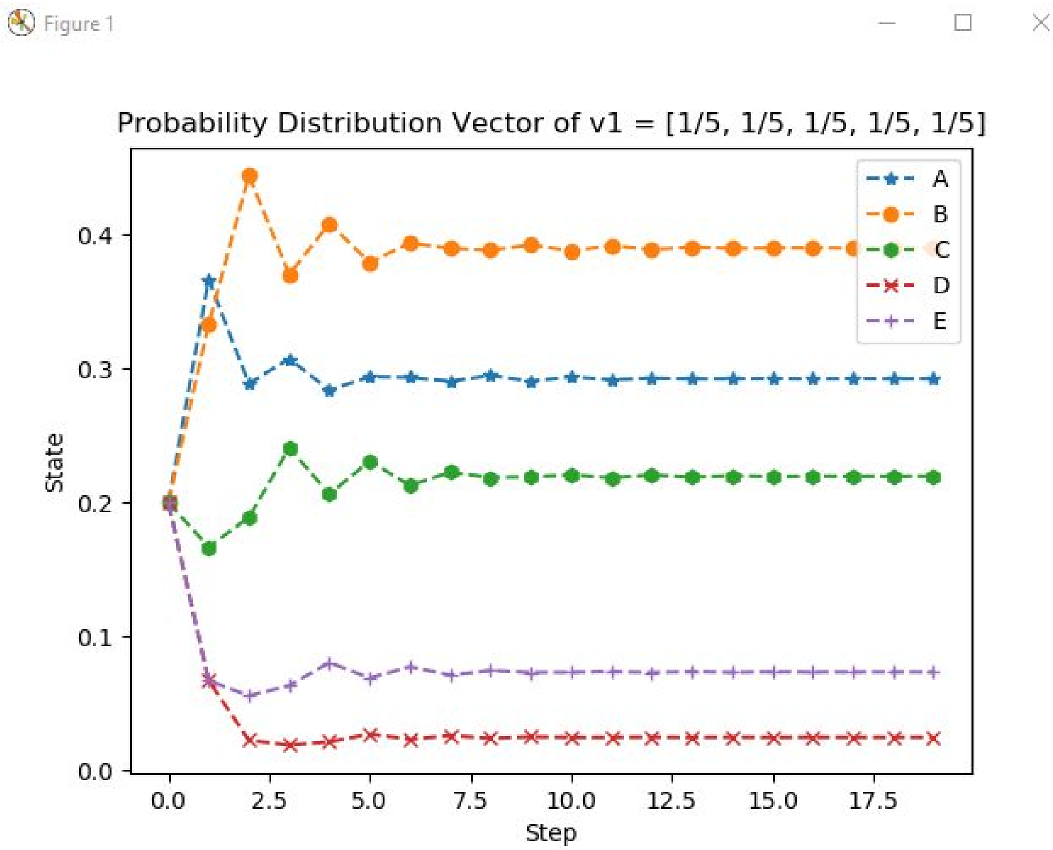


Figure 1
Probability Distribution Vector of v1 = [1/5, 1/5, 1/5, 1/5, 1/5]
A
B
C
D
E
State
0.0
0.1
0.2
0.3
0.4
Step
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

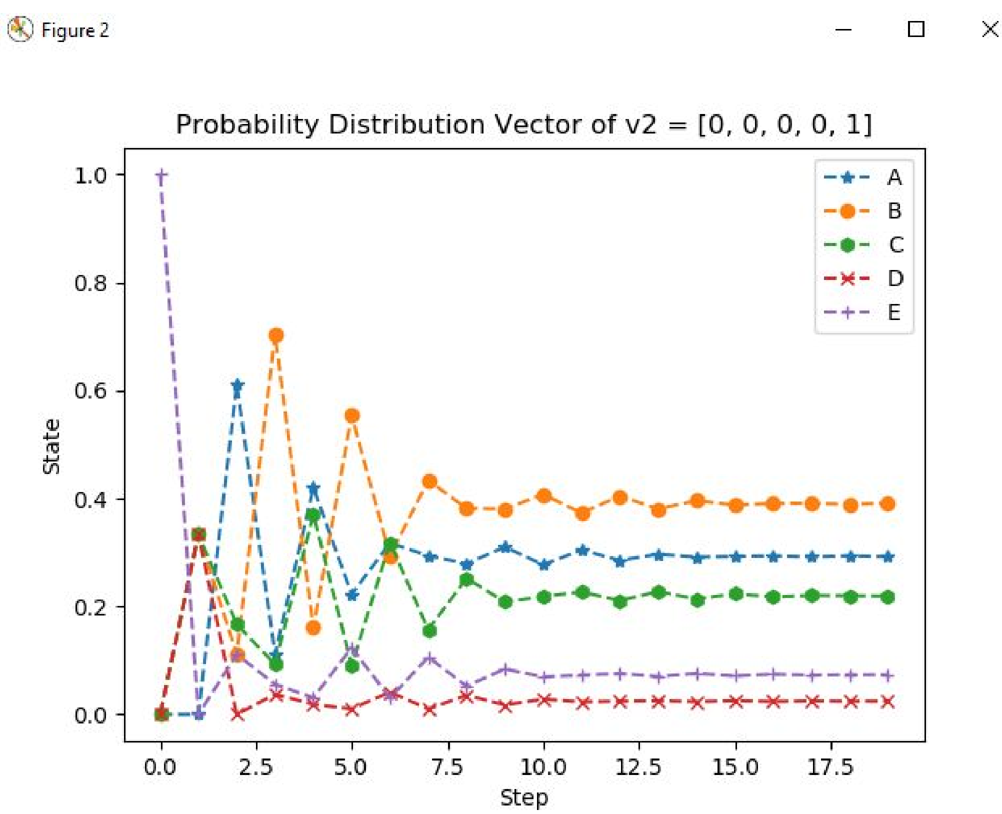


Figure 2
Probability Distribution Vector of v2 = [0, 0, 0, 0, 1]
A
B
C
D
E
State
0.0
0.2
0.4
0.6
0.8
1.0
Step
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

Initial Probability Vector1		
Rank	Page	Probability
1	#2	0.22646251829795874
2	#1	0.17702084436473658
3	#3	0.1315813336300578
4	#5	0.04500286926401215
5	#4	0.015371756967251321

Initial Probability Vector2		
Rank	Page	Probability
1	#2	0.16347717616775292
2	#1	0.1276048701376245
3	#3	0.09544943380035929
4	#5	0.03222900471230626
5	#4	0.011258800404676925

PROBLEM 3. Simulate a five-state absorbing Markov chain

Simulate the Markov chain defined given by the probabilities shown in Figure 1.1 below.

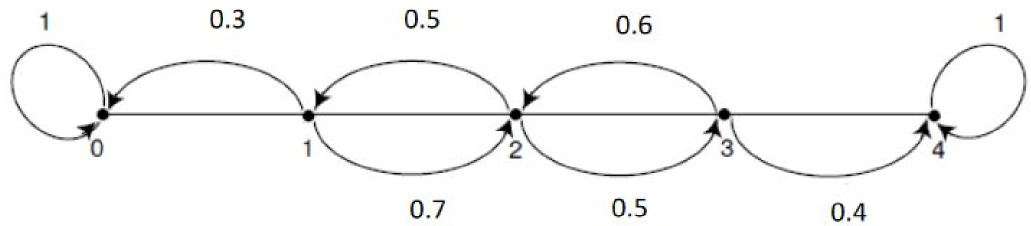
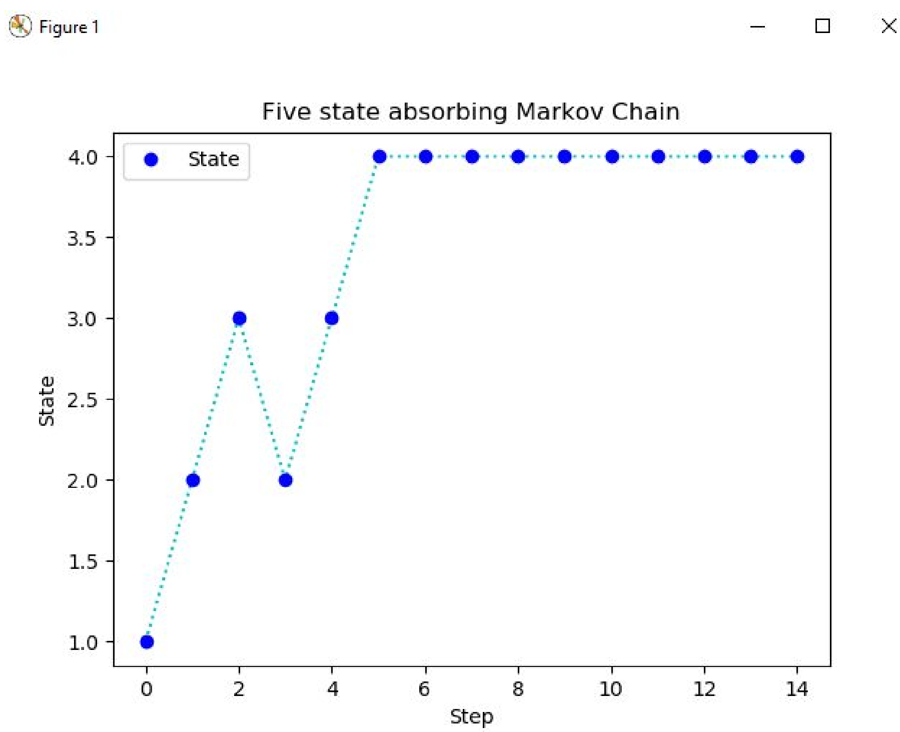


Figure 1.1

3. Define the State Transition Matrix P
4. Create a single-run simulation of the chain for $n = 15$ time steps. The chain can start at a random transient state, but it *should get absorbed by state 0*. Provide the results in a figure similar to Figure 0.2. Your figure must be properly labeled.
5. Create a single-run simulation of the chain for $n = 15$ time steps. The chain can start at a random transient state, but it *should get absorbed by state 4*. Provide the results in a figure similar to Figure 0.3. Your figure must be properly labeled.



PROBLEM 4. Compute the probability of absorption using the simulated chain

1. Run the single simulation of the chain for $N = 10{,}000$ times, *using the following probabilities for the initial state:* $[0.0 \quad 0.0 \quad 1.0 \quad 0.0 \quad 0.0]$, i.e. the initial state in always *state 2.*
2. After the 10,000 runs are completed, count how many times the chain ended at state 0 and how many times the state ended at state 4, *when the starting state is state 2.* Based on these counts compute the probabilities of absorption b_{20} and b_{24}.
3. Use the format of Table 1 below to report the probabilities of absorption on your report. Points will be taken off, if the given format is not used.

Absorption probabilities (via simulations)			
b_{20}	0.4124	b_{24}	0.5388