Introduction to Markov Chains

Ryan McPherson

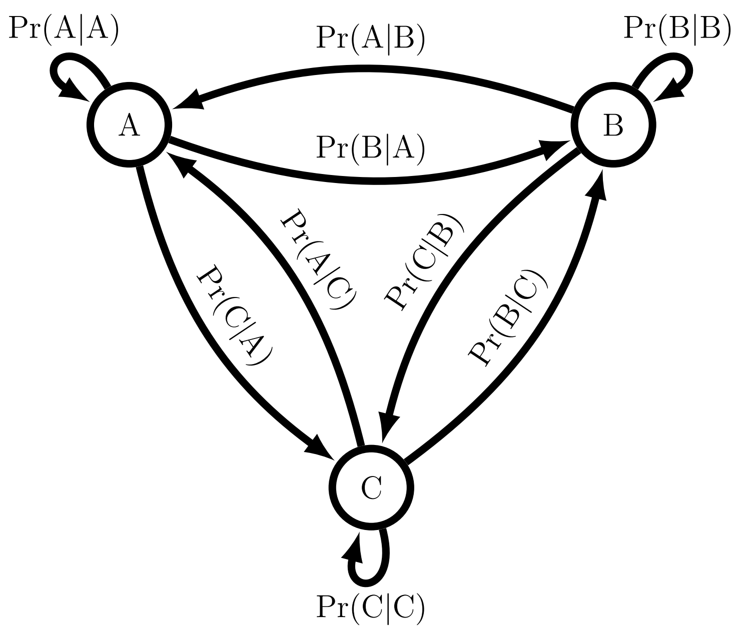
June 2021

1 Introduction

Markov Chains were first created in an attempt to analyze poetry mathematically [2]. The core idea comes from being able to repeatedly apply a transition matrix over a state space. The state space has special conditions needed for a Markov Chain to be formed. Additionally, Markov Chains open the door for many other important concepts. Through this paper I hope to give a better understanding about what makes Markov Chains interesting, how we can tie this in to our learning about eigenvectors, and shed some light on the use of Markov Chain's in Google's original page rank algorithm. A great starting point is [1].

2 State Space

A **state space** is the set of all possible states of a given system. State spaces can be both finite and infinite, but Markov Chains generally are only formed in state spaces of finite dimension. The equivalence of a Markov Chain over an infinite state space is a random walk. We can model the a state space and the transitions between states with a graph of nodes:



The current state can be represented as a vector with only an entry of 1 for index of the state, for this paper I will be using row vectors. In a state space that a Markov Chain can be created, each transition is the probability of changing states given only the current state. In the diagram this is being represented by Pr(X|Y), read as probability of X given Y. By using the probabilities for changing states we can create a transition matrix that represents the state space.

3 Stochastic Matrix

The matrix we can create is a **stochastic matrix**. A stochastic matrix is a square transition matrix where each entry is represented by $p_{i,j}$ = Pr(i|j), where i is the i element row wise and j is the j element column wise. This probability is the representation of transitioning from state i to state j. If we apply the stochastic matrix to our vector representation of the current state we will have a vector of probabilities for following states. Using this idea, our initial state vector is also a vector of probabilities, given we know what the state is the probability must be 1. A full matrix of these probabilities is as follows:

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,j} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i,1} & p_{i,2} & \cdots & p_{i,j} \end{bmatrix}$$

In a right stochastic matrix the sum of any row will be equal to 1:

$$x \in \{x \in \mathbb{Z} | 1 \leq x \leq j\}$$

$$\sum_{m=1}^{x} p_{x,m} = 1$$

In a left stochastic matrix the sum of any column will be equal to 1:

$$y \in \{y \in \mathbb{Z} | 1 \leq y \leq i\}$$

$$\sum_{n=1}^{y} p_{n,y} = 1$$

For a stochastic matrix that satisfies both of these conditions it is considered to be doubly stochastic [3].

4 Markov Chains and Steady States

A **Markov Chain** comes from the idea that we can take any state from our state space, use it as an initial state, then repeatedly apply our stochastic matrix. The repeated application of transformations or transitions is what creates our Markov Chain. With our stochastic matrix we can assemble a Markov Chain of length n, where $n \in \mathbb{N}$:

$$P^n$$

This is useful for answering questions about which state may be present after a given amount of repetitions. After an infinite amount of repetitions of our Markov Chain the vector we get a **steady state**. All stochastic matrices have a steady state that the Markov Chain will eventually arrive at.

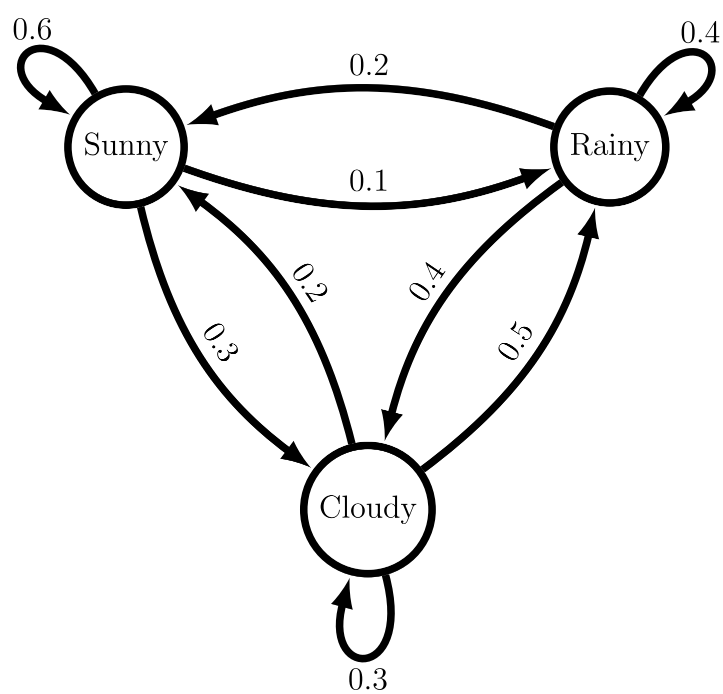
A clear connection to our class can be seen through steady states and their relation to eigenvectors. Given that the Markov Chain has a steady state, it is easy to imagine that eventually we will have a vector which is no longer changed by our stochastic matrix. If we were to write said vector, s, we would have sW$= s$. This is identical to how we describe an eigenvector for a row vector, $v\text{A} = \lambda v$, however in our case the eigenvalue, $\lambda = 1$. While other eigenvalues for a stochastic matrix may exist, we only concern ourselves with the eigenvalue of 1. Because the steady state is made of probabilities, the components added together must be 1, we can make use of the eigenvalue of 1 to ensure our sum does not change. In turn, it is quick to compute the steady state by using $\lambda = 1$, going about the processes we had learned in class.

5 Example

To give an example of a finite state space, we can use possible weather states:

$$\{Sunny, Rainy, Cloudy\}$$

We can create a diagram like the one previous:



Now using our probabilities established in the diagram, we can create a transition matrix that models our state space:

$$\begin{bmatrix} \Pr(Sunny|Sunny) & \Pr(Rainy|Sunny) & \Pr(Cloudy|Sunny) \\ \Pr(Sunny|Rainy) & \Pr(Rainy|Rainy) & \Pr(Cloudy|Rainy) \\ \Pr(Sunny|Cloudy) & \Pr(Rainy|Cloudy) & \Pr(Cloudy|Cloudy) \end{bmatrix} = W = \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

We can prove this matrix double stochastic through the following:

Proof. Given the matrix of transitions above, we can find the sum of each column, which must be 1 if our matrix is left stochastic.

$$0.6 + 0.2 + 0.2 = 1$$

$$0.1 + 0.4 + 0.5 = 1$$

$$0.3 + 0.4 + 0.3 = 1$$

Because all column's values add to 1, the matrix is left stochastic. We can sum each row to find if the matrix is right stochastic and inherently doubly stochastic.

$$0.6 + 0.1 + 0.3 = 1$$

$$0.2 + 0.4 + 0.4 = 1$$

$$0.2 + 0.5 + 0.3 = 1$$

Seeing that every row and column sum to 1, we can say that this matrix is doubly stochastic. □

Now that we know our matrix is stochastic and represents the transitions of our state space we can create a Markov Chain. Say we want to know what type of weather to expect given 3 days from today. If the starting day is a sunny day we can begin by representing in row vector form.

$$d_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Now to find predictions for tomorrow's weather, we have:

$$d_1 = d_0 W = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.6 & 0.1 & 0.3 \end{bmatrix}$$

And now if we assemble a Markov Chain to find the 3rd day we have:

$$d_3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.376 & 0.299 & 0.325 \end{bmatrix}$$

This vector allows us to see that there is a 37.6% chance of it being sunny, 29.9% chance of rain, and 32.5% chance of it being cloudy. Unfortunately, these probabilities don't give too much insight for what to pack for vacation. In a much larger state space, we may see that some states become heavily favored and others may go to zero. However, for this example, we have similar chances for each outcome.

We can also find the steady state for this state space. Knowing that the steady state occurs at $\lambda = 1$, we can create the following equation:

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 0.6-1 & 0.1 & 0.3 \\ 0.2 & 0.4-1 & 0.4 \\ 0.2 & 0.5 & 0.3-1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

$$-0.4x + 0.2y + 0.2z = 0$$

$$0.2x - 0.6y + 0.4z = 0$$

$$0.2x + 0.5y - 0.7z = 0$$

And if we solve this system of equations we will have:

$$x = 1, y = 1, z = 1$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$$

To find the steady state vector we must ensure that our entries all sum to 1. The way to do so is multiply our eigenvector by $\frac{1}{\text{sum of the values}}$. We know that the resulting vector will have the same span, and because the eigenvalue that scales our vector is 1, we dont have to worry about our values being changed. In our example we have:

$$\frac{1}{1+1+1} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

So the steady state of our state space is $\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$. And finally to prove that the vector is a steady state we can apply our stochastic matrix and show that vW = v.

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0.6 & 0.1 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.2 & 0.5 & 0.3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

We can make a connection between the steady state to our previous results. Our probabilities after only 3 applications of our stochastic matrix had already started to approach the steady state. Given infinite repetitions, one could see how the steady state would naturally emerge.

6 Google's PageRank Algorithm

The **PageRank algorithm** is patented by Google and used to determine which order to list website results for a given web search. PageRank makes use of the set of all websites on the internet, and creates transitions between them based on their links. Each page divides its value of 1 among the webpages it links to. Links to the same webpage are not counted, leaving 0's along the diagonal. Given websites W_1, W_2, W_3, W_4 we can create an example. If:

$$\begin{aligned}
&W_1 \text{ links to } W_2, W_4 \\
&W_2 \text{ links to } W_1, W_3, W_4 \\
&W_3 \text{ links to } W_1 \\
&W_4 \text{ links to none}
\end{aligned}$$

Then we create the following matrix:

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

With columns of 0's we dont have a stochastic matrix, and we cannot find the steady state. To solve this problem, we replace any column of all 0's with $\frac{1}{n}$, where n is the number of websites in our set. Performing this action results in:

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

We can clearly see that the sum of each column in our matrix S equals 1, meaning we have a left stochastic matrix. The final step in creating Google's PageRank matrix is to apply a **dampening factor**, α. The dampening factor is a scalar between 0 and 1, representing the idea that randomly navigating from website to website will have to stop at some point. The assumed dampening factor of Google's PageRank is $\alpha = 0.85$. So finally we have our matrix G given by:

$$G = \alpha S + \frac{1-\alpha}{n}\mathbf{1}$$

$\mathbf{1}$ is a square matrix of size n where every entry is 1. Following our example, our matrix G is:

$$G = \begin{bmatrix} 0 & 0.85 * \frac{1}{3} & 0.85 & 0.85 * \frac{1}{4} \\ 0.85 * \frac{1}{2} & 0 & 0 & 0.85 * \frac{1}{4} \\ 0 & \frac{1}{3} & 0 & 0.85 * \frac{1}{4} \\ 0.85 * \frac{1}{2} & 0.85 * \frac{1}{3} & 0 & 0.85 * \frac{1}{4} \end{bmatrix} + \frac{0.15}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} 0.0375 & 0.32083\ldots & 0.8875 & 0.25 \\ 0.4625 & 0.0375 & 0.0375 & 0.25 \\ 0.0375 & 0.37083\ldots & 0.0375 & 0.25 \\ 0.4625 & 0.32083\ldots & 0.0375 & 0.25 \end{bmatrix}$$

If we calculate the eigenvector at $\lambda = 1$ we get: $v \approx [1.05611, 0.781163, 0.594821, 1]$. The steady state of G is then equal to $\frac{v}{n}$=[0.264028, 0.195291, 0.148705, 0.25] We can display each of our websites with their page ranks in the order they would appear:

$$\begin{aligned} W_1 &= 0.264028 \\ W_4 &= 0.25 \\ W_2 &= 0.195291 \\ W_3 &= 0.148705 \end{aligned}$$

The matrix used by Google operates over the set of all websites in the accessible internet, meaning an incomprehensible amount of entries. Over time Google has optimized the process to keep search results appearing quickly. There have also been weights introduced for variables such as website age, and combative measures against websites that try to outsmart the algorithm. More information on the matter can be found through [4].

References

[1] DeGroot, Morris and Schervish. *Probability and Statistics Fourth Edition.* (2012). pp. 188-200.

[2] *First Links in the Markov Chain.* American Scientist. (2018, September 6). https://www.americanscientist.org/article/first-links-in-the-markov-chain.

[3] Gagniuc, Paul A. *Markov Chains, From Theory to Implementation and Experimentation.* John Wiley & Sons (2017.

[4] Margalit, Rabinoff, and Williams. *Interactive Linear Algebra UBC edition.* Section 6. Georgia Institute of Technology (2017) and University of British Columbia(2020). https://secure.math.ubc.ca/ tbjw/ila/index.html