**An Introduction to Markov Chains: Concepts and Applications**

Yining Guo

College of William and Mary

MATH400: Mathematical Connections

Professor Chi-Kwong Li

Oct. 18, 2020

**An Introduction to Markov Chains: Concepts, History, and Applications**

Stochastic processes describe the evolution of a random phenomenon with respect to time. Particularly, the term ‘stochastic’ refers to a system where each observation at a certain time has a certain probability to generate a certain outcome. In general, this probability depends on observations obtained previously. The more observations one has, the more accurate one can predict the outcome later on (Physics of Life. The Phycist's Road to Biology, 2007). Of all types of stochastic process, Markov chains are one of the most commonly used. Proposed by Russian mathematician Andrei Markov, their unique property that the outcome only depends on the latest observation and whatever before the latest has no influence enables researchers to form a well-defined and concise model without. In this paper, I aim to deliver a brief overview of Markov chains and relevant applications. I first introduce a short example of Markov chains and important mathematical concepts associated with it. Then I discuss three applications in subjects other than mathematics, ranging from finance to computer science algorithm. Instead of delving deeply into the mathematical analysis of Markov chains, I would like to depict the usefulness of analyzing real-world problems from the mathematical perspective.

**Transition Probabilities, Transition Matrices, and Probability Distributions**

To begin, we consider a mouse in a maze as shown in figure 1. The maze consists of five different rooms with channels connecting to each other. The mouse can move by one step to one of the rooms adjacent to its current room. For example, if the mouse is currently in room 1, it can move to any of room 2, 5, and 4, but cannot move to room 3. To simplify the problem, two additional assumptions are made here:

1. The mouse’s decisions are independent of any historical movements and only depend on the current room the mouse is in;
2. The mouse cannot stay in the same room forever.



Let *i* denote the current room the mouse is in, and let *j* denote other rooms ($i,jε\{1,2,3,4,5\}$). Mathematically, each *i* and *j* is defined as a state. Then, the transition probability $Pji$ is the probability for the mouse to move from room *i* to room *j*. For example, $P\_{53}=\frac{1}{3}$ is the probability going from 3 to 5: three identical channels are available in room 3 with equal probabilities. The transition probability matrix is formed by writing all transition probabilities in a matrix form, and such a matrix $P$ is shown below:

$$\left[\begin{matrix}0&^{1}/\_{3}&0&^{1}/\_{3}&^{1}/\_{4}\\^{1}/\_{3}&0&^{1}/\_{3}&0&^{1}/\_{4}\\0&^{1}/\_{3}&0&^{1}/\_{3}&^{1}/\_{4}\\^{1}/\_{3}&0&^{1}/\_{3}&0&^{1}/\_{4}\\^{1}/\_{3}&^{1}/\_{3}&^{1}/\_{3}&^{1}/\_{3}&0\end{matrix}\right]$$

To find the transition probability of moving from state *i* to state *j*, one could simply look up the entry $P\_{ji}$*.*

In the previous case, the starting state, the room the mouse currently stays, is given. Now, consider that the mouse could be in any one the five rooms. We then assign a probability distribution vector to specify the starting state. Let $q = \left[\begin{matrix}q\_{1}\\q\_{2}\\q\_{3}\\q\_{4}\\q\_{5}\end{matrix}\right]$ be a probability distribution vector, where each $q\_{i}$ representing the probability of starting from room *i.* By multiplying the vector $q$ by the matrix $P$, one can get the probabilities of arriving at each room by one move. (Math Explorer’s Club: The Mouse, the Maze and the Markov Chain, 2008). Such kind of processes is a simple case of Markov chains.

**Mathematical Expression and Steady State Theorem**

Mathematically, the Markov property holds for a sequence of random variables $X\_{0}, X\_{1}, ... $with values in a countable set $S$ if at any time *n*, the future values of variables $X\_{n+1},X\_{n\\_2}$ ,... depend on the history $X\_{0}, X\_{1} $*,…*only through the present state, $X\_{n}$ (Serfozo, 2009). It is also known as the memoryless property. In terms of probability theories, for any $s, i\_{0}, i\_{1}, ...\in S$,

$P(X\_{n}=s | X\_{0}=i\_{0},... X\_{n-1}=i\_{n-1} ) = P(X\_{n}=s | X\_{n-1}=i\_{n-1})$*.*

A discrete stochastic process with such property is a Markov chain, and a continuous process with the property is called a Markov process. Within a Markov chain, given $X\_{n-1}$ and the transition probability matrix *A*, one can compute $X\_{n}=X\_{n-1}A$*,* and so $X\_{n}=X\_{n-1}A=X\_{n-2}A^{2}= ... = X\_{0}A^{n}$*.* If $X=XA$ for some $X$ after successive iterations from any starting distributions, then $X$ is called the steady state distribution for the Markov chain. Two attributes are related to the formation of a steady state distribution of a Markov chain: irreducibility and aperiodicity. In particular, a Markov chain is irreducible if there exists a path from every state to every other state. A Markov chain is periodic if for some time *t*, there exists some state$S $that can be reached only at time $\{t, 2t, 3t, ...\}$. That is, there is a period *t* associated with the state $S$. A Markov chain is hence called aperiodic without such a *t*. The steady state theorem states that a Markov chain that is both aperiodic and irreducible has a steady state distribution.

**Applications in Stock Market Trend Predictions**

 The analyses of steady states of Markov chains and transition probabilities have been widely used in financial predictions, risk management and other issues in macroeconomic studies. Aiming at showing how the concepts are utilized, I present a simplified model of a stock market and related computing processes here.

Table 1 represents a hypothetical market with probabilities of falling into three trends: the bull, bear and stagnant market (Myers, Wallin, & Wikström). A trend of a stock market is the overall direction of all stocks’ prices. A bull market trend means a general increase in all stocks’ prices; a bear market means a general decrease; a stagnant market means that it neither grows nor shrinks. While economists usually incorporate various indicators when predicting the future trend, I consider all these probabilities are fixed and given to make it more straightforward. Each entry in the table denotes the probability of the market going from one to another. For example, the probability of going from the bull market to the bear market is 0.075, but the probability of going from the bear market to the bull market is 0.15. A state here is correspondingly defined as a time period. Suppose that a state is one week long, and the current week is set as bearish. Then, the vector representing the initial state is $S\_{0}=\left[\begin{matrix}0\\1\\0\end{matrix}\right]$.



 Table 1: A Hypothetical Stock Market

We now calculate the probabilities of a bull, bear or stagnant market from any numbers of weeks into the future. Steps of calculation are shown below:

One week from now:

$S\_{1}= \left[\begin{matrix}0.9&0.075&0.025\\0.15&0.8&0.05\\0.25&0.25&0.5\end{matrix}\right]\left[\begin{matrix}0\\1\\0\end{matrix}\right] = \left[\begin{matrix}0.15\\0.8\\0.05\end{matrix}\right]$

This means that given that the current market is bearish, then it has a probability $P\_{1} = 0.15 $ of becoming a bull market, a probability $P\_{2} = 0.8 $ of staying bearish, and a probability $P\_{3} = 0.05 $ of becoming a bull market in one week.

 Five weeks from now:

$$S\_{5}= \left[\begin{matrix}0.9&0.075&0.025\\0.15&0.8&0.05\\0.25&0.25&0.5\end{matrix}\right]^{5}\left[\begin{matrix}0\\1\\0\end{matrix}\right] = \left[\begin{matrix}0.48\\0.45\\0.07\end{matrix}\right]$$

 Fifty two weeks from now:

$$S\_{52}= \left[\begin{matrix}0.9&0.075&0.025\\0.15&0.8&0.05\\0.25&0.25&0.5\end{matrix}\right]^{52}\left[\begin{matrix}0\\1\\0\end{matrix}\right] = \left[\begin{matrix}0.63\\0.31\\0.05\end{matrix}\right]$$

 A hundred weeks from now:

$$S\_{100}= \left[\begin{matrix}0.9&0.075&0.025\\0.15&0.8&0.05\\0.25&0.25&0.5\end{matrix}\right]^{100}\left[\begin{matrix}0\\1\\0\end{matrix}\right] = \left[\begin{matrix}0.63\\0.31\\0.05\end{matrix}\right]$$

 Eventually, as the number of weeks goes to infinity, the probabilities of market trend changes will converge to a steady state. The vector $\left[\begin{matrix}0.63\\0.31\\0.05\end{matrix}\right]$ is the steady state distribution in this case. In addition, according to the steady state theorem, this steady state distribution is constant regardless of the initial state. That is, the current market trend will not impact the future trend as time goes to infinity.

**Applications in Literature Analysis**

One of the most famous examples of using markov models to mathematize literature dates back to 1913, when Andrei Markov himself lectured on his research at Royal Academy of Science in St. Petersburg (Markov, 2006). He investigated the frequency of vowels and consonants and the potential connections between the two. He manually collected and classified the first 200,000 Russian letters of Alexander Pushkin’s novel Eugene Onegin by dividing them into 200 groups and wrote each group in a square table with ten rows and ten columns. Figure 2 shows a screenshot of Markov’s manuscript.



Figure 2: A screenshot of Markov’s manuscript (Hilgers & Langville, 2006)

A letter was either a vowel or a consonant. Markov counted the number of vowels appearing in each column, and joined two numbers in pairs (the 1st and 6th, 2th and 7th, 3rd and 8th, 4th and 9th, and 5th and 10 th), and wrote down the five sums in a column. In this way, he obtained 5 numbers for each 100 letters, each representing the number of vowels in each group of 20 letters, separated by four letters in the text. Markov made 40 tables each with 5 of these columns, each table representing 500 letters of the 20,000. By adding all entries vertically and horizontally, he obtained a new column and a new row where each entry represented the numbers of vowels per 500 letters. Markov then calculated the arithmetic mean of entries in the new row and the new column of each table and got a result of 43.2, the average number of vowels in 100 characters. Consequently, the average number of consonants was 56.8. Hence the probability of any single letter being a vowel is $P = 0.43$*,* and the probability of any single letter being a consonant is  $P = 0.57$*.* Apart from these independent probabilities, four events hence occur:

1. $P\_{1}$: A vowel is followed by a vowel;
2. $P\_{2}$: A vowel is followed by a consonant;
3. $P\_{3}$: A consonant is followed by a vowel;
4. P4: A  consonant is followed by a consonant.

Theoretically, if we only consider the two independent probabilities discussed above regardless of any sequential influences, the four probabilities are calculated as: $P\_{1}=0.43×0.43 = 0.185$, $P\_{2}=0.43×0.57 = 0.245$*,* $P\_{3}=0.57×0.43 = 0.245,$and$P\_{4}=0.57×0.57 = 0.325$*.*

However, it was more possible that each letter in consecutive series was dependent on one another: a certain number of preceding consonants forced the next letter to be a vowel and vice versa (Link, 2006). Markov hence introduced the concept of dependent quantities. Borrowing an equation in his *Investigation of a notable case of dependent samples,* Markov claimed that the appearance of a vowel depended only on the attribute of the previous letter but nothing more than that. Facing a lack of suitable empirical experiment, Markov manually re-counted the number of vowels, taking preceding letters into consideration. The result showed that the probability of a vowel varied according to the preceding letter: when it was a consonant, the probability was 0.663, and when it was a vowel, the probability was about 0.128. The full transition matrix is written as below:

$$\left[\begin{matrix}0.128&0.872\\0.662&0.337\end{matrix}\right]$$

This was the first time that his mathematical assumptions regarding dependent quantities were proved experimentally. Later he revised his hypothesis and stated that the probability of a vowel depended on the previous two letters. This claim was proved in his *On a Case of Samples Connected in Multiple Chains* in 1911.

Markov’s applications of Markov models and transition probabilities to letters have influenced other disciplines like computational linguistics and artificial intelligence in the 20th and 21st centuries. For example, Khmelev and Tweedue (2001) extended the application of Markov chains similar to Markov’s, the probabilities of a subsequent letter given a letter preceded, and designed an algorithm used for authorship attribution. The occurrence of some particular sequences of letters were considered as a signature. Of all 387 cases, the algorithm identified 288 correctly.

**Applications in Web Search**

Apart from stock market predictions and literature analysis, Markov chains are frequently used in web search identifications. PageRank, an algorithm used by Google to order web pages in their search engine results, uses a special Markov chain to compute the rank of importance of web pages and determine in which order all pages should be listed in search results to users. A web page may contain links that direct to other pages or to another part within this page. The algorithm assigns a score of importance to a web page between 0 and 10 based on two criteria: the extent to which the text could match and the number of links a web page contains. Figure 3 presents a web graph with five websites, where$x$has no links to others, and $w$ has one to $x$*.*



 Figure 3: An example of a Web Graph

To understand the rationale for PageRank from the Markov chain perspective, one can regard each state of the Markov chain as one web page, and each transition probability being the probability of moving from one page to another. To generate the transition probability matrix, we further define an $n×n$ matrix *M*, where the entry $M\_{ij}$ is 1 if there is an outgoing link from page $i$ to page $j$, and $M\_{ij}$ is 0 otherwise*.*Hence, the outgoing degree of page $i$, the number of pages that can be reached out from age $i$, is the row sums calculated as $s\_{i}$*:*

$$s\_{i}=\sum\_{}^{}Mij$$

Normalize the matrix *M* by its row sums, we get a new matrix *W*, where each entry $W\_{ij}$ is:

$$W\_{ij} = \left\{\begin{array}{c}^{M\_{ij}}/\_{s\_{i}}, if s\_{i}\geq 1\\^{1}/\_{n}, if s\_{i}=0\end{array}\right.$$

This matrix simulates the behavior of a user at page $i$*.* If there exist outgoing links on page $i$, the user will pick one randomly as the next page to visit. If there are no links, the user randomly picks one according to the uniform distribution to visit. By this, the matrix *W* can be regarded as a transition probability matrix of the Markov chain. By filtering out the web pages with no texts matching, the algorithm could get a finite number of web pages and additionally the Markov chain involved could be proved irreducible and aperiodic. Therefore, there exists a unique steady state distribution, called $π$, which is used to rank all related pages: the page $i$ with the largest $π\_{i}$ will be ranked first, the second largest be ranked second, and so on (Ye, 2013).

**Conclusion**

Markov chains are one of the most fundamental in stochastic processes. First introduced by Andrei Markov in 1907, the concept of Markov chains has been extensively employed in emerging subjects, ranging from economic predictions to computational linguistics. Its unique property that the future state of a variable depends only on its current state simplifies data collection and cleaning process in empirical studies. Mathematically, a Markov chain consists of a state space, whose elements are possible values of a random variable, a transition probability matrix, and an initial state. Regarding applications in economic and financial predictions, computing the transition matrix is critical to the accuracy of the model. Additionally, the transition matrix may vary when certain factors change. In reality, a Markov model could be more complicated, including multiple parameters other than the transition matrix that depict factors involving in the prediction process. By estimating and testing the parameters, one may distinguish certain patterns and hence form the final model. In this paper, I present an example of Markov chains and related mathematical representations. Three applications of Markov chains are introduced briefly and show that Markov chains are tremendously powerful in solving real-world problems.

# **Bibliography**

Markov, A. (2006, December). An Example of Statistical Investigation of the Text Eugene Onegin Concerning the Connection of Samples in Chains. *Science in Context, 19*(4), 591-600.

Link, D. (2006). Traces of the Mouth. Andrei Andreyevich Markov's Mathematization of Writing. *History of Science*, 321-348.

Khmelev, D. V., & Tweedie, F. J. (2001). Using Markov Chains for Identification of Writers. *Literary and Linguistic Computing, 16*(4), 299-307.

Serfozo, R. (2009). *Basics of Applied Stochastic Processes.* Springer-Verlag Berlin Heidelberg.

Myers, D. S., Wallin, L., & Wikström, P. (n.d.). *An introduction to Markov chains and their applications within finance.* Retrieved from http://www.math.chalmers.se/Stat/Grundutb/CTH/mve220/1617/redingprojects16-17/IntroMarkovChainsandApplications.pdf

*Math Explorer’s Club: The Mouse, the Maze and the Markov Chain*. (2008). Retrieved from Math Explorer's Club: http://pi.math.cornell.edu/~mec/Summer2008/youssef/markov.pdf

Blomberg, C. (2007). *Physics of Life. The Phycist's Road to Biology.*

Ye, W. (2013). *On PageRank Algorithm and Markov Chain Reduction.* Retrieved from https://www.cise.ufl.edu/~wye/Pagerank.pdf

Hilgers, P. V., & Langville, A. N. (2006). The five greatest applications of Markov chains.