Markov Chain and its applications

Angela (Xinye) Yang

Abstract

This paper will explore concepts of the Markov Chain and demonstrate its applications in probability prediction area and financial trend analysis. The historical background and the properties of the Markov's chain are analyzed.

1. Introduction of Markov Chain

Markov Chain is a statistic model developed by a Russian Mathematician Andrei A. Markov (1856–1922). It decibels a sequence of possible events that the probability of each event is the dependent of the previous event (Gagniuc 2017).

1.1 Background

In 1887, Markov's teacher outlined a proof of the central limit theorem which in 1995 Markov succeeded in proving the general result using Chebyshev's method. During this period of time, Markov extended the law of large numbers and central limit theorem to certain sequences of dependent random variables forming special classes. This theorem is known for Markov Chain. "He was also very interested in poetry and the first application he found of Markov chains was in fact in a linguistic analysis of Pushkin's work Eugene Onegin" (A. Andersson, Introduktion 2014).

1.2 Limitations and Purposes

This paper will not explore very deep theory regarding Markov's Chain; instead, the variety of applications of the theorem are explored, especially in the area of finance and population predictions.

2. Theorems

2.1 Introduction to the Model

Suppose there are four types of drinks provided at Hoover Dining Hall, DePauw University every meal: Tea, Coffee, Lemonade and Milk.

Assume that 20% of those who drink tea drink coffee next time, 20% drink lemonade, and 30% drink milk. From those who drink coffee, 10% drink tea, 25% drink lemonade, and 25% drink milk next time. From those who drink milk, 30% drinks tea, 10% drink coffee, and 30%drink lemonade. Those who drink lemonade, 20% drink milk, 25% drink coffee, and 30% drink tea. We call this situation a system. A student eats at Hoover-Hall and picks a drink from one of these four drinks, each of them called a state. In our example, the system has four states. In this situation, it is worth to find out the extent to which the drinks company should supply the drinks on a weekly basis.

In a **discrete-time Markov system**, suppose there is a physical or mathematical system that has possible states and at any one time, the system is in one and only one of its states. And suppose that at a given observation period, said period, the probability of the system being in a particular state depends on its status at the period, such a system is called **Markov Chain or Markov process.**

In the drinks example above, there are four states for the system. Define to be the probability of the system to be in state after it was in state j (at any observation). The matrix is called the **Transition matrix** of the Markov Chain.

Definition: the transition matrix P of a Markov chain is said to be regular if for some power of P all the components are positive. The chain is then called a regular Markov chain.

$$A = \begin{bmatrix} .25 & .20 & .25 & .30 \\ .20 & .30 & .25 & .30 \\ .25 & .20 & .40 & .10 \\ .30 & .30 & .10 & .30 \end{bmatrix}$$

The first column represents the state of drinking lemonade (L), the second column represents the state of drinking tea (T), the third column represents the state of drinking coffee (C), and the fourth column represents the state of drinking milk (M).

Similarly, the rows respectively represent drinking lemonade, drinking tea, drinking coffee and drinking milk.

$$A = \begin{array}{c} \begin{array}{cccc} L & T & C & M \end{array} \\ \begin{bmatrix} .25 & .20 & .25 & .30 \\ .20 & .30 & .25 & .30 \\ .25 & .20 & .40 & .10 \\ .30 & .30 & .10 & .30 \end{bmatrix} \end{array}$$

“If the Markov chain has possible states, the matrix will be a matrix, such that entry is the probability of transitioning from state to state. In addition, the transition matrix must be a stochastic matrix, a matrix whose entries in each row must add up to exactly 1" (Soni). Since the matrix is highly related to probability theory, each row represents its own probability distribution. Thus, the summation of each row and column should add up to one.

Now comes the question that is variable to know the answer: how many drinks the supplier company should provide for the next 1 week, next 2 week or next ith week?

Thus, we define the state vector. For a Markov Chain, which has k states, the state vector for an observation period, is a column vector defined b

$$x^{(n)} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

x_i = probability that the system is in the i^{th} state at the time of observation. Note that the sum of the entries of the state vector has to be one. Any column vector x,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$$

The probability vector is the vector that all the elements add up to 1. In this drinks example, we have,

$$x^{(1)} = Ax^{(0)} = \begin{bmatrix} .25 \\ .20 \\ .25 \\ .30 \end{bmatrix}$$

At the end of the 2nd week, the state vector is

$$x^{(2)} = Ax^{(1)} = \begin{bmatrix} .25 & .20 & .25 & .30 \\ .20 & .30 & .25 & .30 \\ .25 & .20 & .40 & .10 \\ .30 & .30 & .10 & .30 \end{bmatrix} \begin{bmatrix} .25 \\ .20 \\ .25 \\ .30 \end{bmatrix} = \begin{bmatrix} .2550 \\ .2625 \\ .2325 \\ .2500 \end{bmatrix}$$

Similarly, we can derive the third, forth$\cdots$ith week by applying the equation:

$$x^{(2)} = Ax^{(1)} = A\left(Ax^{(0)}\right) = A^2x^0$$

$$x^{(n)} = Ax^{(n-1)} = A\left(Ax^{(n-1)}\right) = A^n x^{n-1}$$

From this equation, we can find out the 5th, 30th, ⋯, infinity.

$$x^{(5)} = A^5 x^{(0)} = \begin{bmatrix} .2495 \\ .2634 \\ .2339 \\ .2532 \end{bmatrix}$$

2.1 Discrete-time Markov chain (DTMC) verses Continuous-time Markov chain (CTMC)

In a discrete model, N(t) changes as discrete instants of time. The stochastic processes are valid only at the value of the integer of time, n=1,2, 3, ⋯

A finite Markov chain is often described by a matrix [P], of the chain have M states, then [P] is an M by M matrix with elements P_{ij}.

In this graphic illustration, there are six states in the Markov chain:

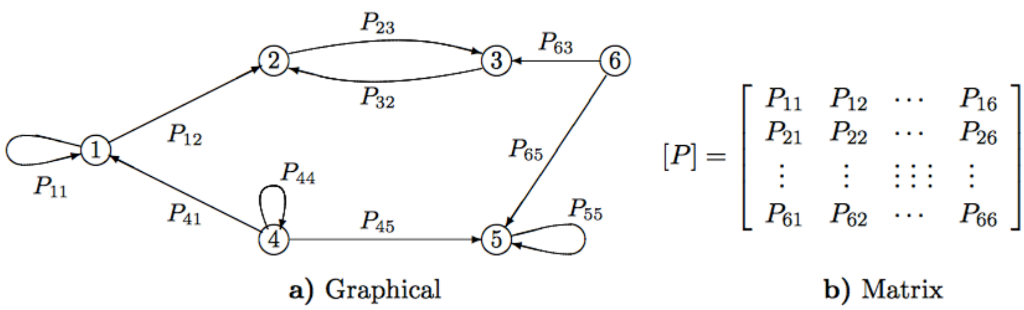


Figure from -R. G. Gallager, “Finite-state Markov chains” p.162

The Markov chain can be time-homogeneous with a finite state space, which refers to the transition matrix P is the same after each step. Thus, the k^{th} step after can be computed as P to the power of k, namely $[P]^k$.

Additionally, the Markov chain can be stationary as the final predictions are independent of the initial state. We can see the results from the population predictions example later.

Markov chain helps people predict future behavior by looking at the subject’s previous preference. However, it is impossible to give an absolute answer while the trend can be analyzed through the Continuous-time Markov chain. This paper will not focus on the continuous model.

2.2 Theorem for predictions

Considered a regular Markov having initial vector x_0 and transition matrix P, then

1. $x_{0,}x_{1,}x_{3,} \cdots \rightarrow x, where\ x\ satisfies\ Px = x.$
   Thus x is an eigenvector of P corresponding to $\lambda = 1$.
2. $P^1, P^2, P^3 \cdots \rightarrow$
   $Q, where\ Q\ is\ a\ stochstic\ matrix. The\ columns\ of\ Q\ are\ all\ identical, each\ being\ an$ eigenvector of P corresponding to $\lambda = 1$.

3. Applications

Research indicates that there are a variety of applications of Markov chain in the area of natural sciences and social sciences, weather predictions, economics and finances etc. Some of the topics are chosen and being analyzed in this paper.

3.1 Population Predictions

The population of U.S. metropolitan and nonmetropolitan areas in 2012 is described by the vector $x_0 = \begin{bmatrix} 255 \\ 52 \end{bmatrix} \begin{matrix} metro \\ nonmetro \end{matrix}$ (Initial populations)

$$\mathrm{P} = \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix} \begin{matrix} metro \\ nonmeteo \end{matrix} \quad \text{(from metro to non-metro)}$$

The populations in the following years are given by a Markov chain with the transition matrix P. Determine the long-term predictions for metro and non-metro populations, assuming no change in their total population.

Solutions:

∵ Observe all the elements of P are positive

∴ The chain is regular and the results of the preceding theorem can be applied to given the long-term trends.

The theorem tells us that P will have an eigenvalue of 1 and that the steady-state vector x is a corresponding eigenvector.

$$\mathrm{Px} = \mathrm{x}$$

$$(\mathrm{P} - I_2)\mathrm{x} = 0$$

$$\begin{bmatrix} 0.99 - 1 & 0.02 \\ 0.01 & 0.98 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

This leads to the systems of equations

$$-0.01x_1 + 0.02x_2 = 0$$
$$0.01x_1 - 0.02x_2 = 0$$

giving $x_1 = 2x_2$. The solutions to this system of equations are $x_2 = r, x_1 = 2r$, where r is a scalar. Thus the eigenvector of P corresponding to $\lambda = 1$ are non − zero vectors of the form

$$r\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The steady-state vector x will be a vector of this form. Assume that there is no total annual population change over the years. Therefore, the sums of the elements of x and x_0 are equal,

$$2r + r = 255 + 52$$
$$3r = 307$$
$$r = 102.33$$

The steady-state vector is thus

$$x = \begin{bmatrix} 204.67 \\ 102.33 \end{bmatrix}$$

This implies the following long-term predictions:

U.S. city populations→204.67 million

U.S. suburban populations→102.33 million

The above theorem gives further information about long-term population trends. Each column of the matrix Q is an eigenvector corresponding to the eigenvalue 1. Let

$$Q = \begin{bmatrix} 2s & 2s \\ s & s \end{bmatrix}$$

Since Q is a stochastic matrix the sum of the elements in each column is 1. Thus,

$$s + 2s = 1$$
$$s = 0.33$$

We get (exhibiting the elements to two decimal places for each reading)

$$P1 = \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix}$$

$$P_2 = P_1 P_1$$

$$= \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix}\begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix}$$

$$= \begin{bmatrix} 0.99\times0.99 + 0.02\times0.01 & 0.99\times0.02 + 0.02\times0.98 \\ 0.01\times0.99 + 0.98\times0.01 & 0.01\times0.02 + 0.98\times0.98 \end{bmatrix}$$

$$= \begin{bmatrix} 0.99 & 0.04 \\ 0.02 & 0.96 \end{bmatrix}$$

$$\begin{array}{cccc} P_1 & P_2 & & Q \\ P_1 = \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix} & P_2 = \begin{bmatrix} 0.99 & 0.04 \\ 0.02 & 0.96 \end{bmatrix} & \cdots\cdots & \rightarrow \begin{bmatrix} 0.67 & 0.67 \\ 0.33 & 0.33 \end{bmatrix} \end{array}$$

We focus on (1,2) element in each matrix and the interpretation is similar to every element in the matrix, we get

$$0.02, 0.04, \cdots, 0.67$$

These are the possibilities of moving from the suburb to city in 1 year, 2 years and so on. From the data, we can see the probability slowly increases, approaching the predicted value of 0.67. Therefore, Q is thus the **long-term transition matrix** of this model. It gives the long-term possibilities of living in metropolitan and nonmetropolitan areas in 2012.

(from)

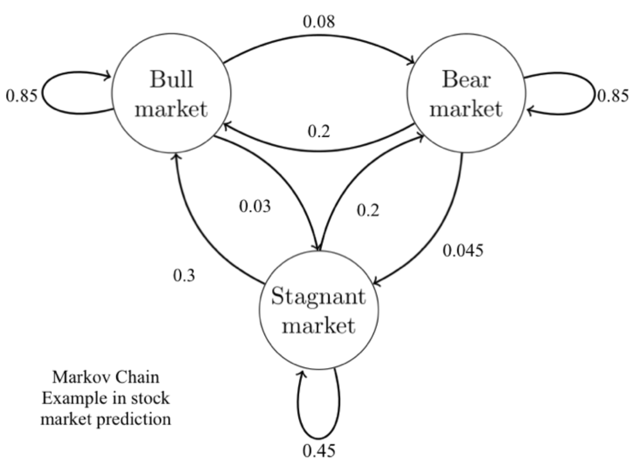
$$\begin{array}{cc} & \begin{array}{cc} \text{city} & \text{suburb} \end{array} \\ Q = & \begin{bmatrix} 0.67 & 0.67 \\ 0.33 & 0.33 \end{bmatrix} \end{array}$$

Long-term probabilities

As we recall the properties of the Markov chain, the Long-term probabilities are independent of the initial state:

The long-term probability of living in metropolitan is about 67% while living in non-metropolitan area is about 33%. Thus, we can come to the conclusion that these possibilities are independent of the initial location.

3.2 Stock Market Predictions



According to the figure, a bull week is followed by another bull week 85% of the time, a bear week 8% of the time, and a stagnant week the other 2 % of the time. It is the same logic with the other two.

The graph above is revised from the finance model [change the numbers] https://en.wikipedia.org/wiki/Markov_chain#/media/File:Finance_Markov_chain_example_state_space.svg

To From	Bull	Bear	Stagnant
Bull	0.85	0.02	0.03
Bear	0.2	0.85	0.045
Stagnant	0.3	0.2	0.45

While the table shows the probability clearly, we define the matrix P:

$$P = \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix}$$

P is a 3 by 3 matrix which contains the information about the probabilities for Bill, Bear and Stagnant markets. Three situations can happen:

I. This week is a Bullish market, we denote the stochastic row vector X $[1 \quad 0 \quad 0]$;
II. This week is a Bearish market, we denote the stochastic row vector X $[0 \quad 1 \quad 0]$;
III. This week is a Stagnant market, we denote the stochastic row vector X $[0 \quad 0 \quad 1]$.

Suppose in this case it is a bull market, during the 1st week we have,

$$X \times P = [1 \quad 0 \quad 0] \times \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix} = [0.85 \quad 0.2 \quad 0.3]$$

The 2nd week from now, we have $X \times P^2$, denoted,

$$X \times P^2 = [1 \quad 0 \quad 0] \times \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix}^2$$

$$= [1 \quad 0 \quad 0] \times \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix} \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix}$$

$$= [0.85 \quad 0.2 \quad 0.3] \times \begin{bmatrix} 0.85 & 0.02 & 0.03 \\ 0.2 & 0.85 & 0.045 \\ 0.3 & 0.2 & 0.45 \end{bmatrix}$$

$$= [0.85 \times 0.85 + 0.04 + 0.09 \quad 0.02 \times 0.85 + 0.17 + 0.06 \quad 0.0255 + 0.09 + 1.35]$$

$$= [0.85 \quad 0.247 \quad 1.47]$$

For the 53rd week from the first week, we can derive the general formula that

$$\boldsymbol{x^{n+53} = x^{(n)} P^{53}}$$

By using the transition matrix, we can imply the long-term stock market trend (n=52 implies that approximately one year later) as it is a finite series.

4. Summary

Markov chain is a useful tool for prediction which its extended applications benefit and facilitate several areas including physics, chemistry, biology, statistics, genetics, games and social sciences etc. The central theorem (discrete Markov chain) implies that it is possible to predict future trends by analyzing the object's previous behaviors. Some of the variations like Time-homogeneous Markov chains (or stationary Markov chains) and Markov chain with memory both provide different dimensions to the whole picture.

As a stochastic process, Markov chain has the properties of reducibility, periodically, periodicity and Steady-state analysis & limiting distributions. Although this paper focuses on the discrete part, the theory of continuous-time Markov processes is worth to discuss as well. Introduced by Andrei Kolmogorov in his paper in 1931, "he studied a particular set of Markov processes known as diffusion processes, where he derived a set of differential equations describing the processes" (Kendall and Batchelor). A lot of mathematicians continuously contributed to the Markov chains' foundation.

Multiple states are highly possible in the modification process of the Markov chain. This property renders the theorem to be applicable to different areas from physics to finance. Constructing tables and if possible diagrams facilitate calculations. Moreover, the nonhomogeneous calculations are usually more complex and reliable than the homogenous calculations due to the fact that the transition matrix P being the same after each step (Myers, Wallin, and Wikström).

Markov chain proves that people can predict future events, ranging from the weather to the stock market trend by applying the stochastic calculations correctly and analyzing empirical data from past years.

5. References

1. B. Sericola, *Markov Chains: Theory, Algorithms and Applications* . London: ISTE Ltd and John Wiley & Sons Inc, 2013.
2. Everitt, B.S. (2002) *The Cambridge Dictionary of Statistics*. CUP. ISBN 0-521-81099-X
3. Encyclopaedia Britannica "Andrey Andreyevich Markov" https://www.britannica.com/biography/Andrey-Andreyevich-Markov
4. Gagniuc, Paul A. (2017). Markov Chains: From Theory to Implementation and Experimentation. USA, NJ: John Wiley & Sons. pp. 1–235. ISBN 978-1-119-38755-8.
5. Graph at the stock market predictions is revised from the finance model [I change the numbers] https://en.wikipedia.org/wiki/Markov_chain#/media/File:Finance_Markov_chain_example_state_space.svg
6. Kendall, D. G.; Batchelor, G. K.; Bingham, N. H.; Hayman, W. K.; Hyland, J. M. E.; Lorentz, G. G.; Moffatt, H. K.; Parry, W.; Razborov, A. A.; Robinson, C. A.; Whittle, P. (1990). "Andrei Nikolaevich Kolmogorov (1903–1987)". *Bulletin of the London Mathematical Society*. **22** (1): . doi:10.1112/blms/22.1.31. ISSN 0024-6093.
7. "Markov chain | Definition of Markov chain in US English by Oxford Dictionaries". Oxford Dictionaries | English. Retrieved 2017-12-14.
8. Myers, Wallin and Wikström. "An introduction to Markov chains and their applications within finance" http://www.math.chalmers.se/Stat/Grundutb/CTH/mve220/1617/redingprojects16-17/IntroMarkovChainsandApplications.pdf
9. R. G. Gallager, "Finite-state markov chains" in Stochastic processes: theory for applications. United Kingdom: Cambridge university press, 2013, 160-213. [Online] Avalible: http://www.rle.mit.edu/rgallager/documents/6.262-4vaw.pdf p.162
10. Valeriy Skorokhod (5 December 2005). *Basic Principles and Applications of Probability Theory*. Springer Science & Business Media. p. 146. ISBN 978-3-540-26312-8. Archived from the original on 23 March 2017.
11. Soni, Devin. "Introduction to Markov Chains: What are Markov chains, when to use them, and how they work" Mar 5 *Towards Data Science* https://towardsdatascience.com/introduction-to-markov-chains-50da3645a50d
12. Williams, Gareth. Linear Algebra with applications (8^{th} edition) p145-50; 197-22