

MORE ON THE PROBABILITY THEORY AND MARKOV CHAINS

Doç.Dr. Özcan BAYTEKİN^{1*}

Abstract

In this paper, it will be emphasized the importance of choosing mathematical models in order to study some observational phenomena. By comparing two methods, for solving a probability problem, important properties of the entries of Markov Chains matrices (Stochastic Matrices) will be explained. On the other hand, when we solve a problem by classical probability methods, some computational difficulties arise, but then when we solve the same problem by a completely different method, we see that the same difficulties arise in same other interesting place. This has been another secondary aim of this article.

Key words: Markov Chains, stochastic matrices, entry of a matrix, combinatory analysis, matrix multiplication, probability.

OLASILIK TEORİSİ VE MARKOV ZİNCİRLERİ ÜZERİNE YENİ YAKLAŞIM

Özet

Bu makalede, bazı gözlemsel olayları incelerken seçilecek matematiksel modelin uygunluğunun önemi üzerinde durulacaktır. Bir olasılık problemini çözerken, iki uygun metod karşılaştırılmış ve ilave katkı olarak da Markov matrisleri elemanlarının (Stokastik Matrisler) bazı özellikleri de ortaya konulmuştur. Bir problemi klasik olasılık metotları ile çözerken ortaya çıkan hesaplama zorluklarının, yine aynı problemi tamamen farklı bir metod ile çözerken aynı zorlukların başka bir şekilde ve yerde karşımıza çıkmasının gösterilmesi bu makalenin ikinci bir amacını teşkil etmektedir.

¹ Marmara University, Faculty of Economics and Administrative Sciences, Dept. of Business Administration, Kadıköy – İstanbul – TURKEY

* Fakültemiz öğretim üyesi Doç. Dr. Özcan Baytekin makalesi dergimiz hakem süresinde iken vefat etmiştir. Değerli hocamızı bir kez daha saygıyla anıyoruz.

Anahtar kelimeler: Markov zincirleri, stokastik matrisler, bir matris elemanı, kombinasyon analizi, matris çarpımı, olasılık.

1. Introduction

To explain the purpose of this paper, we shall make a brief summary of Markov Chains.

1.1. Probability Vectors, Stochastic Matrices

A vector $u = (u_1, u_2, \dots, u_n)$ is called a probability vector if components are nonnegative and their sum is 1.

A square matrix $P = (P_{ij})$ is called a stochastic matrix if each of its rows is a probability vector, i.e. if each entry of P is nonnegative and the sum of the entries in each row is 1.

1.2. Regular Stochastic Matrices

A stochastic matrix P is said to be regular if all the entries of some power P^m are positive. For this reason let us consider the stochastic matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 2 \end{pmatrix}$, is regular since.

$$A^2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 4 & 4 \end{pmatrix} \text{ is positive in every entry.}$$

1.3. Fixed Points and Regular Stochastic Matrices

The fundamental property of regular stochastic matrices is contained in the following theorem.

Theorem 1.3.1: Let P be a regular stochastic matrix .Then

(i) P has a unique fixed probability vector t , and the components of t are all positive.

(ii)The sequence P, P^2, P^3, \dots of powers of P approaches the matrix T whose rows are each fixed point t ,

(iii)If P is any probability vector, then the Sequence of vectors $P^P, P^{P^2}, P^{P^3}, \dots$ approaches the fixed point t ,

1.4. Markov Chains

We now consider a sequence of trials whose outcomes say, $X_1, X_2 \dots$ satisfy the following two properties:

(i) Each outcome belongs to a finite set of outcomes (a_1, a_2, \dots, a_m) called the state space of the system ; if the outcome on the n^{th} step.

(ii)The outcome of any trial depends at most upon the outcome of the immediately preceding trial and not upon any other previous outcome; with each pair of states (a_i, a_j) there is given the probability P_{ij} that a_j occurs immediately after a_i occurs.

Such a stochastic process is called a (finite) Markov Chain. The numbers P_{ij} called the transition probabilities can be arranged in a matrix

$$P = \begin{pmatrix} P_{11} & P_{12} & \dots & P_{1m} \\ P_{21} & P_{22} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & \dots & P_{mm} \end{pmatrix}$$

called the transition matrix.

Thus with each state a_i ; there corresponds the i^{th} row $(P_{i1}, P_{i2}, \dots, P_{im})$ of the transition matrix P ; if the system is in state a_i , then this row vector represents the probabilities of all the possible outcomes of the next trial and so it is a probability vector. Accordingly;

Theorem 1.4.1. the transition matrix P of a Markov Chain is a stochastic matrix.

1.5. Higher Transition Probabilities

The entry P_{ij} in the transition matrix P of a Markov Chain is the probability that the system changes from the state a_i to the state a_j in one step $a_i \rightarrow a_j$. Question: What is the probability denoted by $P_{ij}^{(n)}$ that system changes from the state a_i to the state a_j in exactly n steps

$$a_i \rightarrow a_{k_1} \rightarrow a_{k_2} \rightarrow \dots \rightarrow a_{k_{n-1}} \rightarrow a_j$$

The next theorem answers this question; here the $P_{ij}^{(n)}$ are arranged in a matrix $P^{(n)}$ called the n -step matrix.

Theorem 1.5.1. Let P be the transition matrix of a Markov Chain process. Then the n -step transition matrix is equal to the n^{th} power of P ; that is, $P^{(n)} = P^n$.

Now suppose that , at the same arbitrary time ,the probability that the system in state a_i is P_i ; we denote these probabilities by the probability vector $P=(P_1, P_2, \dots, P_n)$

which is called the probability distribution of the system at that time. In particular, we shall let

$$P^{(P)} = (P_1^{(0)}, P_2^{(0)}, \dots, P_m^{(0)})$$

Denote the initial probability distribution, i.e. the distribution when the process begins, and we shall let

$$P^{(n)} = (P_1^{(n)}, P_2^{(n)}, \dots, P_m^{(n)})$$

Denote the n^{th} step probability distribution, i.e. the distribution after the first n steps.

The following theorem applies:

Theorem 1.5.2. Let P be the transition matrix of a Markov Chain process. If $P = (P_{ij})$ is the probability distribution of the system at some arbitrary time, then P^P is the probability distribution of the system one later and P^{P^n} is the probability distribution of the system n steps later. In particular

$$P^{(1)} = P^{(0)P}, P^{(2)} = P^{(1)P}, P^{(3)} = P^{(2)P}, \dots, P^{(n)} = P^{(0)P^n}$$

2. Fundamental Concepts

In this paper, we will present two solutions for one problem. One of these two methods will use fundamental principles of probability and the other will use Markov Chains. Using these two methods, we will try to find out (namely to clarify) the difference between the two following probabilities:

Probability P_A : The probability obtained after n steps. This probability is explained by Theorem 1.5.2. and this probability is determined by Markov Chains. This is a cumulated probability but in literature this property is not explained clearly. One of the aims of this paper is to explain and verify this property.

Probability P_B : Probability obtained at the n^{th} step. We will determine P_B , by using fundamental principles of probability. Clearly, these two probabilities are different from each other. All we have explained so far will be explain by a numerical example

2.1. A numerical example

In order to explain our idea we consider the following case:

Two players bet \$1 for each of the successive tosses of a coin. Each has a bank of \$6. What is the probability that one player, say Jones, wins all the money on the tenth toss of the coin?

In the first method, we will develop a mathematical model, without use of Markov Chains (but basic principles of probability).

2.2. First Method of Solution

The game which is being played, can be represented by a series of W's and L's, W representing by a win for player Jones (and hence an increase of \$1 for his bank) and L representing a loss (and hence a decrease of \$1 for his bank). The tosses are independent

and,

$$P(W) = P(L) = \frac{1}{2}$$

For this problem, consider that the coin has been tossed ten times. There are 2^{10} possible sequences of W's and L's, each with probability $\left(\frac{1}{2}\right)^{10}$. It is necessary to enumerate n_a , the number of sequences which result in player Jones winning the game at exactly tenth toss.

In order for the event of interest to occur, player Jones must have \$11 at trial 9 and must win on trial ten. That is, after 9 trials (during which \$9 will change hands) Jones must win seven times and loose only twice in order that the total gain will be

$$\$7 - \$2 = \$5$$

And he will be left with a total of \$11 after 9 trials.

If the two losses were placed randomly among 9 trials which can be done in

$$\binom{9}{2} = 36 \text{ ways, it is possible that the game might end before the appointed 10}$$

trials. These arrangements must be eliminated from the 36 possibilities. Notice that player Jones can only win on an even numbered trial, since he must win \$6. That is, the difference between the number of wins and losses is 6, so that the sum of the numbers of wins and losses must be even number.

The number of arrangements of 2 L's and 7 W's for which Jones would win on trial 6 are shown below:

W W W W W W W L L
W W W W W W L L W
W W W W W W L W L

The number of arrangements for which Jones would win on trial 8 are found using the same argument as for the event "win on trial 10" above. They are:

L W W W W W W W L
 W L W W W W W W L
 W W L W W W W W L

 W W W L W W W W L
 W W W W L W W W L
 W W W W W L W W L

These 9 arrangements are eliminated from consideration, so that,

$$n_a = \binom{9}{2} - 9 = 27$$

and the probability of interest is $27 \left(\frac{1}{2}\right)^{10} = \frac{27}{1024}$

As the reader can notice, there are difficult points that need attention in this first method of solutions. This difficult part does not exist in the second method of solutions.

2.3. Second Method Using Markov Chains

The transition matrix for the considered problem can be constructed as follows:

$$a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8 \quad a_9 \quad a_{10} \quad a_{11} \quad a_{12}$$

$$\begin{matrix}
 & a_0 \\
 a_1 & \left[\begin{array}{cccccccccccc}
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{array} \right]
 \end{matrix}$$

Figure I. Transition Matrix

The probability distribution vector P_0 is

$P^{(0)} = (0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$ since each player begins with \$6

We seek $P^{(10)}$, the probability that the system is in state a_i after 10 steps. Now let us compute the tenth step probability distribution $P^{(10)}$:

$$P^{(1)} = P^{(0)} P = (0 \ 0 \ 0 \ 0 \ 0 \ \frac{1}{2} \ 0 \ \frac{1}{2} \ 0 \ 0 \ 0 \ 0 \ 0)$$

$$P^{(2)} = P^{(1)} P = (0 \ 0 \ 0 \ 0 \ \frac{1}{4} \ 0 \ \frac{1}{2} \ 0 \ \frac{1}{4} \ 0 \ 0 \ 0 \ 0)$$

$$P^{(3)} = P^{(2)} P = \begin{pmatrix} 0 & 0 & 0 & 1/8 & 0 & 3/8 & 0 & 3/8 & 0 & 1/8 & 0 & 0 & 0 \end{pmatrix}$$

$$P^{(4)} = P^{(3)} P = \begin{pmatrix} 0 & 0 & 1/16 & 0 & 1/4 & 0 & 3/8 & 0 & 1/4 & 0 & 1/16 & 0 & 0 \end{pmatrix}$$

$$P^{(5)} = P^{(4)} P = \begin{pmatrix} 0 & 1/32 & 0 & 5/32 & 0 & 10/32 & 0 & 10/32 & 0 & 5/32 & 0 & 1/32 & 0 \end{pmatrix}$$

$$P^{(6)} = P^{(5)} P = \begin{pmatrix} 1/64 & 0 & 6/64 & 0 & 15/64 & 0 & 20/64 & 0 & 15/64 & 0 & 6/64 & 0 & 1/64 \end{pmatrix}$$

$$P^{(7)} = P^{(6)} P = \begin{pmatrix} 2/128 & 6/128 & 0 & 21/128 & 0 & 35/128 & 0 & 35/128 & 0 & 21/128 & 0 & 6/128 & 2/128 \end{pmatrix}$$

$$P^{(8)} = P^{(7)} P = \begin{pmatrix} 10/256 & 0 & 27/256 & 0 & 56/256 & 0 & 70/256 & 0 & 56/256 & 0 & 27/256 & 0 & 10/256 \end{pmatrix}$$

$$P^{(9)} = P^{(8)} P = \begin{pmatrix} 20/512 & 27/512 & 0 & 83/512 & 0 & 126/512 & 0 & 125/512 & 0 & 83/512 & 0 & 27/512 & 20/512 \end{pmatrix}$$

$$P^{(10)} = P^{(9)} P = \begin{pmatrix} 67/1024 & 0 & 110/1024 & 0 & 209/1024 & 0 & 252/1024 & 0 & 209/1024 & 0 & 110/1024 & 0 & 67/1024 \end{pmatrix}$$

Now suppose that at some arbitrary time the probability that the system is in state a_i is P_i , we denote these probabilities by the probability vector $P = (P_1, P_2, \dots, P_n)$ which is called the probability distribution of the system at that time. In particular we shall let:

$$P^{(0)} = (P_1^{(0)}, P_2^{(0)}, \dots, P_m^{(0)})$$

denote the initial probability distribution, i.e. the distribution when the process begins, and we shall let:

$$P^{(n)} = (P_1^{(n)}, P_2^{(n)}, \dots, P_m^{(n)})$$

denote the n^{th} step probability distribution i.e. , distribution after the first n steps.

Now let us consider another matrix M whose rows are constituted by the probability distribution vectors

$$P^{(0)}, P^{(1)}, P^{(2)}, \dots, P^{(n)}$$

$$a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8 \quad a_9 \quad a_{10} \quad a_{11} \quad a_{12}$$

$$M = \begin{matrix} P_0 \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \\ P_8 \\ P_9 \\ P_{10} \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cancel{1/2} & 0 & \cancel{1/2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cancel{1/4} & 0 & \cancel{1/2} & 0 & \cancel{1/4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cancel{1/8} & 0 & \cancel{3/8} & 0 & \cancel{3/8} & 0 & \cancel{1/8} & 0 & 0 & 0 \\ 0 & 0 & \cancel{1/16} & 0 & \cancel{1/4} & 0 & \cancel{3/8} & 0 & \cancel{1/4} & 0 & \cancel{1/16} & 0 & 0 \\ 0 & \cancel{1/32} & 0 & \cancel{5/32} & 0 & \cancel{10/32} & 0 & \cancel{10/32} & 0 & \cancel{5/32} & 0 & \cancel{1/32} & 0 \\ \cancel{1/64} & 0 & \cancel{6/64} & 0 & \cancel{15/64} & 0 & \cancel{20/64} & 0 & \cancel{15/64} & 0 & \cancel{6/64} & 0 & \cancel{1/64} \\ \cancel{2/128} & \cancel{6/128} & 0 & \cancel{21/128} & 0 & \cancel{35/128} & 0 & \cancel{35/128} & 0 & \cancel{2/128} & 0 & \cancel{6/128} & \cancel{2/128} \\ \cancel{10/256} & 0 & \cancel{27/256} & 0 & \cancel{56/256} & 0 & \cancel{70/256} & 0 & \cancel{56/256} & 0 & \cancel{27/256} & 0 & \cancel{10/256} \\ \cancel{20/512} & \cancel{27/512} & 0 & \cancel{83/512} & 0 & \cancel{126/512} & 0 & \cancel{126/512} & 0 & \cancel{83/512} & 0 & \cancel{27/512} & \cancel{20/512} \\ \cancel{67/1024} & 0 & \cancel{110/1024} & 0 & \cancel{209/1024} & 0 & \cancel{252/1024} & 0 & \cancel{209/1024} & 0 & \cancel{110/1024} & 0 & \cancel{67/1024} \end{bmatrix}$$

Figure II. Probability Distribution Matrix M at Each Step

Now let us consider again $P^{(10)}$:

$$P^{(10)} = \begin{pmatrix} \frac{67}{1024} & 0 & \frac{110}{1024} & 0 & \frac{209}{1024} & 0 & \frac{252}{1024} & 0 & \frac{110}{1024} & 0 & \frac{67}{1024} \end{pmatrix}$$

Now what is the probability of getting \$12 after 10 trials. According to the above vector the answer is

$$P_{12}^{(10)} = \frac{67}{1024}$$

But in section 2.1., the probability of getting \$12 at the 10th trial was $\frac{27}{1024}$.

The difference between these two answers comes from the fact that $P_{12}^{(10)}$ contains also probabilities of the previous steps. One of the aims of this paper is to verify this property.

If we consider the two probability distributions $P^{(n)}$ and $P^{(n-1)}$ namely

$$P^{(n-1)} = \left(P_1^{(n-1)} P_2^{(n-1)} \dots P_m^{(n-1)} \right) \text{ and}$$

$$P^{(n)} = \left(P_1^{(n)} P_2^{(n)} \dots P_m^{(n)} \right) \text{ then}$$

The probability of the event a_m and a_0 at the n^{th} step is :

$$P(a_m) = P^{(a_0)} = P_m^{(n)} - P_m^{(n-1)}$$

Let us consider now the example (2.1.)

The probability that the player Jones gains \$6 and finally, he gets \$12 is, using the Formula (A):

$$P = a_{12}^{(10)} - a_{12}^{(9)} = \frac{67}{1024} - \frac{20}{512} = \frac{27}{1024}$$

and this is the same result obtained in section 2.1., by using the fundamental principles of probability.

2.4. Finals Remarks and Conclusions

In this paper, we presented two solutions for one problem. One of these two methods used fundamental principles of probability and the other Markov Chains. Using these two methods, we tried to precise if there is a difference between the two following probabilities

(i) Probability P_A : Probability obtained after n steps (obtained by Markov Chains)

(ii) Probability P_B : Probability obtained at the n^{th} step (obtained by fundamental principles of probability)

Obviously there is a difference between P_A and P_B , because the first one, namely P_A is a cumulated probability, that is, if we consider again example 2.1. and the solution obtained by Markov Chain method, probability distribution vector $P^{(10)}$ was as follows:

$$P^{(10)} = \left(\frac{67}{1024} \ 0 \ \frac{110}{1024} \ 0 \ \frac{209}{1024} \ 0 \ \frac{252}{1024} \ 0 \ \frac{209}{1024} \ 0 \ \frac{110}{1024} \ 0 \ \frac{67}{1024} \right)$$

Here the $P_{12}^{(10)} = \frac{67}{1024}$ is the probability of gaining \$12 after 10 days. This is the

sum of probabilities of obtaining \$12, at sixth trial, at the seventh trial, at the eighth trial, at the ninth trial and at the tenth trial. The reason of this lies behind the structure of the transition. Matrix shown in Fig. I. In this matrix the first entry of row 1 is the others are zeros, and the last entry of row 12 is a 1, and the others are zeros. In this case $P_0^{(n)}$ denotes the probability that the man reaches the state a_0 on or before the n^{th} step. Similarly $P_{12}^{(n)}$ denotes the probability that reaches the state a_{12} on or before the n^{th} step.

The formula (A) is only valid for a_0 and a_n and not for any other states because of the structure of the transition matrix.

We added the matrix M shown in Fig. II. for the interested readers. If we determine the probability distribution when n is going to infinity, this probability distribution will take the following form:

$$P = \left(\frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \frac{1}{2} \right)$$

3. References

KREYSZIG, Erwin, **Advanced Engineering Mathematics**, John Wiley & Sons, 1998

NAHMIAS, Steven, **Production and Operation Analysis**, McGraw-Hill/*Irwin*, 2001

The Staff of Research & Education Association, **The Linear Algebra Problem Solver**,
New Jersey 1994

WAGNER, Harwey M., **Principles of Operations Research**, Prentice – Hall, Inc.,
Englewood Cliffs, New Jersey, 1997

WINSTON ,Wayne L., **Operations Research**, Learning EMEA, 1996