"MARKOV CHAINS"

by

Virginia M. Gray

Ball State University

for

I.D. 499

Advisor: Dr. John Beekman

TABLE OF CONTENTS

INTRODUCTION

The Markov process is a method of analyzing the current movement of some variable in an effort to predict the future movement of this same variable. An extensive amount of work has been done by various mathematicians concerning such a process. Out of this work has come a system of classification for the states of such a process and numerous theorems which have greatly increased its effectiveness.

Of particular interest in the following discussion is the Markov chain. The chain falls into the category of a Markov process or stochastic process. The author of this paper has delved into several of the works of most capable mathematicians in order to obtain information about this Markov chain and its originator, A. A. Markov. The paper here offered gives only a brief outline of the Markov chain. Also included here is a brief biography of Mr. Markov himself.

Chapter I

The modern theory of Markov chains has its origins in the studies of the Russian mathematician, Andrei Andreevich Markov. Mr. Markov was born in 1856. The exact location of his birth is not known, other than it was in Russia. Very little information is available about his early years or his parents or family.

In regard to Markov's education, it is known that he spent the academic grades at the University of Saint Petersburg. He was appointed ordinary professor by the University in 1886. A few years later, in 1896, he became an ordinary member of the Academy of Sciences. The majority of Markov's years at the University were devoted to research.

The earlier part of his work was devoted to number theory and analysis. He made notable contributions to continued fractions, the limits of integrals, the approximation theory, and the convergence of series. After about 1900, Markov concentrated on probability theory. Using fairly general assumptions, he proved the central limit theorem, which states the asymptotically normal distribution of the sum of a large number of independent random variables. This distribution is often called the Gaussian distribution.

In 1906-1907, Mr. Markov turned to the study of mutually dependent variables. In this research, he introduced the notion of chained events, and worked with the concept of chain dependence. These chains have become known as "Markov chains" and will be the topic of the following paper. The initial research was done by Mr. Markov in an attempt to describe mathematically the physical phenomenon known as Brownian motion.[1] In working in this area, Markov did the basic pioneering work on the class of Markov processes used in mathematics today.

Of Markov's later life, it is known that he married and had a son. His son, named Andrey Andreyevich Markov Jr., was born in 1903. Following in his father's footsteps, the younger Markov also became a mathematician. He has contributed research concerning topology, topological algebra and the theory of dynamic systems. He also proved the impossibility of solving some problems of the associative system theory and also problems of integral matrices.[2]

Both father and son have made numerous contributions to the field of mathematics. The elder Markov contributed the class

---

[1]In 1828, an English botanist, R. Brown noticed that fine particles suspended in liquid move chaotically, incessantly changing their direction of motion. It was hypothesized that various questions about this motion could be answered by using the tools and theory of mathematics.

[2]Encyclopedia of Russia and the Soviet Union, ed. by Michael T. Florishy (McGraw-Hill, 1961), p. 332.

of Markov processes and the younger his research in topology. The elder Markov died in 1922, having extended several classical results concerning independent events to certain types of chains. This work is one of the starting points of the modern theory of stochastic processes.

Chapter II

In considering the outcome of any experiment, we must look at the various trials involved. Frequently we encounter an outcome to one experiment which is completely independent of all previous outcomes. Another possibility would be an experiment in which a certain trial is dependent upon all of the previous trials. The third possibility and that with which we are concerned, is one where the n^{th} trial is dependent upon only the $n-1^{st}$ trial and is completely independent of the first $n-2$ trials. Cases of this type are classified as "Markov chains."

Let us define a "Markov chain." Consider an experiment such that its sequence of trials have outcomes, $E_1, E_2, \ldots, E_a$. This set of outcomes has the property that the probability of the n^{th} trial of the experiment resulting in outcome E_k is dependent on the outcome of, at most, the $n-1^{st}$ trial, E_j. Thus to each pair (E_j, E_k) there corresponds a conditional probability p_{jk} such that given E_j has occurred, the next trial will result in E_k. "Instead of saying 'the n^{th} trial results in E_k,' we shall say that at time n the system is in state E_k."[3] This sequence of trials with outcomes $E_1, E_2, \ldots, E_a$ is called a

---

[3] Willian Feller, An Introduction to Probability Theory and Its Applications (New York: John Wiley & Sons, 1957), I, p. 340.

Markov chain (with constant transition probabilities) if the probabilities of sample spaces are defined by:

$$P(E_{j_0}, E_{j_1}, E_{j_2}, \ldots, E_{j_n}) = (a_{j_0})(P_{j_0 j_1})(P_{j_1 j_2})\ldots(P_{j_{n-1} j_n})$$

in terms of an initial probability distribution (a_k) for the states at time 0, and fixed conditional probabilities p_{jk} of E_k, given that E_j has occurred at the preceding trial. It is more convenient to call these chains simply "Markov chains" and this terminology shall be used from this point. The words "with constant transition probabilities" are used in standard terminology due to the fact that in more general Markov chains, the transition probabilities may be dependent upon which trial is involved. In this definition the term "state" is used. This term, borrowed from the science of physics, is simply the condition of the process at a given instance. Let us consider an example (1) where the "state" of the accident record of an individual auto driver might be defined according to the cumulative total number of accidents appearing in that record. If the record shows k accidents, then let the record be in the state designated as "S_k". A "state" need not be defined in arithmetic terms, however. In a given instance, the exact fashion in which the states are defined will determine the specific characteristics of the process with which associated and may determine whether or not the process is or is not a Markov chain.[4]

---

[4]Kenneth L. McIntosh, "An Introduction to Finite Markov Chains," Casualty Actuarial Society paper.

The conditional probability p_{jk} is called the probability of the transition $E_j \rightarrow E_k$, that is, from state E_j to state E_k. These transition probabilities p_{jk} may be arranged in a matrix of transitional probabilities.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1a} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2a} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3a} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ p_{a1} & p_{a2} & p_{a3} & \cdots & p_{aa} \end{bmatrix}$$

A Markov chain has several properties which can be determined by examining this matrix of transition probabilities. P is clearly a square matrix. Also P consists of non-negative elements, since $p_{jk} \geq 0$ for all j and k. Thirdly, the row sums are equal to unity, since $\sum_{k=1}^{a} p_{jk} = 1$ for all j. Thus P is a stochastic, or Markov, matrix since it meets the above mentioned conditions. Any stochastic matrix can serve as a matrix of transition probabilities. Such a matrix together with our initial distribution (a_k) completely defines a Markov chain with states $E_1, E_2, \ldots, E_a$.

We might consider example (2) to see how this matrix of transition probabilities is determined. A man is playing two slot machines. The first machine pays off with probability $\frac{1}{2}$, the second with probability $\frac{1}{4}$. If he loses, he plays the same machine again; if he wins, he switches to the other machine. The transition matrix is

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

We see that this matrix meets the properties mentioned above and thus is a stochastic matrix.[5]

In the above case, we have considered only a one-step transition. We can also have a two or more step transition. This indicates the number of paths possible of various number steps to go from one machine to another. Let us denote the first slot-machine as machine A and the second as machine B. Then a two-step transition from A to B might be A to A to B or A to B to B. Likewise three, four, or more-step transitions can be determined.

We can also determine the sum of the probabilities of all such possible paths. The general symbolism for this probability is $P_{jk}^{(n)}$. This indicates the probability that the transition goes from a state E_j to the state E_k in n steps. Using this idea, we can determine matrices of transition probabilities for any number n of steps.

The next major consideration in our discussion of Markov chains is that of classification of states and chains. This procedure is accomplished by considering well defined properties of various chains. One of the most important aspects is whether it is possible to go from one given state to another given state.

One of the first types to be considered is the closed set. A set C of states is closed if no state outside C can be reached from any state E_j in C. Furthermore, the smallest

---

[5]John Kemeny and Laurie Snell, Finite Markov Chains (Princeton, N.J.: D. Van Nostrand Co., Inc., 1960), p. 31.

closed set containing C is called the closure of C. A single state forming a closed set is called an absorbing state. As an example (3), let us consider an experiment having a matrix of transition probabilities

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ 0 & 1 \end{pmatrix}$$

Then the set E_2 would be a closed set and would also be an absorbing state since no single state other than E_2 can be reached from E_2.

A Markov chain is said to be irreducible if there exists no closed set other than the set of all sets. This says that a chain is irreducible if and only if, every state can be reached from every other state. The Markov chain of example (1) is an irreducible chain.

Consider an arbitrary fixed state E_j, where the system is located initially. Then $f_j^{(n)}$ is the probability that the first return to E_j occurs at time n. Thus $f_j^{(1)} = p_{jj}$ and $f_j^{(2)} = p_{jj}^{(2)} - f_j^{(1)} p_{jj}$, and so on. Looking at $f_j^{(2)}$, we note that this means that the probability of a first return to E_j at time 2 is the same as the probability that its return to E_j is at time 2 minus the probability that it returned at time 1 and returned again at time 2. Using this same philosophy, the general pattern would be

$$f_j^{(n)} = p_{jj}^{(n)} - f_j^{(1)} p_{jj}^{(n-1)} - f_j^{(2)} p_{jj}^{(n-2)} - \dots - f_j^{(n-1)} p_{jj} .$$

The sum $f_j = \sum_{n=1}^{\infty} f_j^{(n)}$ is the probability that, starting from E_j, the system ever returns to E_j. The state E_j is said to be <u>persistent</u> or <u>recurrent</u> if $f_j = 1$. In this case, the mean recurrence time is $\mu_j = \sum_{n=1}^{\infty} nf_j^{(n)}$.[6] We call the state E_j <u>transient</u> if $f_j = 0$ or $f_j < 1$. A state E_j is a <u>persistent null</u> state if $f_j = 1$, but the mean recurrence time $\mu_j = \infty$.[7] We can also consider a transient set of states. This is a set in which every state can be reached from every other state. Each of these states can be left, however.

We next need to introduce the notion of the period of a state, or a periodic state. The state E_j is said to have period $t > 1$ if $p_{jj}^{(n)} = 0$ whenever n is not divisible by t and t is the smallest integer with this property. That is, a return to state E_j is impossible except, perhaps, in $t, 2t, 3t, \ldots$ steps.[8] Persistent states which are neither periodic nor null states will be called <u>ergodic</u>. An ergodic state is an element of an ergodic set, i.e. a set in which every state can be reached from every other state, but the set cannot be left once it is

---

[6]Feller, <u>Probability Theory</u>, pp. 352-353.

[7]A.T. Bharucha-Reid, <u>Elements of the Theory of Markov Processes and Their Applications</u> (New York: McGraw-Hill Co., 1960), pp. 15-17.

[8]Feller, <u>Probability Theory</u>, p. 353.

entered. As an example of this type of set, let us consider a process whose matrix P is such that the system starts in, or at least enters, at some time, any one of the states $E_1 - E_4$, inclusive. It thereafter bounces around among the four of them for as long as the process may be continued. This set is an ergodic set, a set which can be entered but once entered cannot be left.

From our classification of states, we can arrive at a classification of Markov chains. The first chain which we shall consider is an irreducible Markov chain. In such a chain all states belong to the same class and in each case have the same period. Either the states are all transient, all persistent null states or all persistent non-null states. As a result of this, it is seen that in a finite Markov chain there are no null states, and it is impossible that all states are transient.

We shall restrict our discussion to aperiodic (non-periodic) chains since the formulations for periodic chains are very involved. Then the above theorem can be restated: In an irreducible aperiodic Markov chain either (1) the states are all transient or null, or (2) all states are ergodic.[9] As a type of chain whose states are all transient, we have the absorbing chain. An absorbing chain is one which has at least one absorbing state, and this absorbing state can be reached from every state. There

---

[9]Ibid., p. 356.

are other possibilities in this classification too. In the second category where all states are ergodic, we have an ergodic chain, a chain in which it is possible to go from every state to every other state. In this case we have two possibilities. If the ergodic set is regular, i.e. one where some power of P contains no zero elements, we have a regular Markov chain. If the ergodic set is cyclic, we have a cyclic Markov chain, i.e., an ergodic chain in which each state can only be entered at certain periodic intervals.[10]

Let us now consider some of the more basic theorems concerning Markov chains. Of course there are many theorems covering various aspects of the Markov chain and its use, but it is virtually impossible to list all such theorems here.

One of the most important theorems concerned with Markov chains is the "Mean Ergodic Theorem." Before stating this theorem, we must define the term stationary distribution. The probability vector, t, which is uniquely determined for a Markov chain whose states are ergodic, is called the <u>stationary distribution</u> of the chain. Now consider an arbitrary chain where $A_{jk}^{(n)}$ is defined to be $\frac{1}{n}\sum_{v=1}^{n} p_{jk}^{(v)}$. If E_j and E_k belong to the same irreducible closed set, i.e. the same ergodic set, then

---

[10]Kemeny and Snell, <u>Finite Markov Chains</u>, pp. 36-37.

$\lim_{n \to \infty} A_{jk}^{(n)} = u_k$ where u_k is the stationary probability whenever u_k exists. Many different proofs of this theorem can be found in the resources consulted.[11] If E_j and E_k belong to different closed sets, $p_{jk}^{(n)} = 0$ for all n, so $A_{jk}^{(n)} = 0$ for all n. Also, if E_k is a transient state, then $\lim_{n \to \infty} A_{jk}^{(n)} = 0$.[12]

In our discussion of finite Markov chains, another important theorem to be considered is: In a finite Markov chain the probability of the system's staying forever in the transient states is zero. The proof of this theorem can also be found in any one of several sources.

Earlier in our discussion we deleted the case of periodic chains. We can use the various theorems found in relationship to aperiodic chains to characterize the behaviour of $p_{jk}^{(n)}$ in irreducible periodic chains. This process would be much easier to accomplish having considered the previous material. This study of periodic chains will not be attempted at this time, however.

We have dealt with a very tightly restricted group of Markov chains. At this time, let us look at the definition of the general Markov process.

---

[11] Bharucha-Reid, Elements of the Theory of Markov Processes, p. 32.

A sequence of discrete-valued random variables is a Markov process if, for every finite collection of integers $n_1 < n_2 < \cdots < n_r < n$, the joint distribution of $(X^{(n_1)}, X^{(n_2)}, \ldots, X^{(n_r)}, X^{n})$ is defined in such a way that the conditional probability of the relation $X^{(n)} = x$ on the hypothesis $X^{(n_1)} = x_1, \ldots, X^{(n_r)} = x_r$ is identical with the conditional probability of $X^{(n)} = x$ on the single hypothesis $X^{(n_r)} = x_r$. Here $x_1, \ldots, x_r$, are arbitrary numbers for which the hypothesis has a positive probability.[13] In simpler terms, this tells us that given the state x_r at time n_r, no additional data concerning states of the system at previous times can alter the conditional probability of the state x at a future time n. Thus we see that our Markov chains are general Markov processes.

---

[13]Ibid., p. 369.

Chapter III

Markov chains have found application in many areas in our world today. These areas include physics, sociology, genetics, and economics. One very important use of this Markov process in the last few years has been as a marketing aid for examining and predicting consumer behavior in terms of brand loyalty. The following example illustrates how Markov chains can be used in this area.

Assume that A, B, and C are three dairies and they supply all of the milk consumed in the town where they are located. The consumers switch from dairy to dairy because of advertising, dissatisfaction with service, and various other reasons. To simplify the mathematics, we shall assume that no old customers leave the market and no new customers enter the market during the period under consideration. All three dairies keep records concerning the number of their customers and the dairy from which they obtained each new customer. Suppose that Table 1 is the true explanation of the exchange of customers among the three dairies for one month. From it we see that a somewhat complex movement of customers involving all three dairies occurred.

Table 1

Actual exchanges of customers

Dairy	June 1 customer	Gain	Loss	July 1 customer
A	200	60	40	220
B	500	40	50	490
C	300	35	45	290

Each dairy needs details about the movement in order to do the best marketing job possible. This type of information enables the dairy to decide if it is spending too much time promoting new customers and neglecting its old customers or vice versa.

This simple analysis in terms of net gain or net loss of customers is inadequate for good management. More detailed analysis concerning the rate of gains from and losses to all competitors is needed. The Markov chain will provide this needed analysis.

Our first step is to compute transition probabilities for all three dairies. For example, from Table 1 we know that dairy B loses 50 customers this month. This is the same as saying that it has a probability of .9 of retaining customers. Table 2 gives these transition probabilities.

Table 2

Transition Probabilities for Retention of Customers

Dairy	June 1 customer	Number Lost	Number Retained	Probability of Retention
A	200	40	160	.8
B	500	50	450	.9
C	300	45	255	.85

Calculation of a complete set of these transition probabilities would require data on the flow of customers among all of the dairies. Table 3 gives this information.

Table 3

Flow of Customers

		Gains			Losses			
Dairy	June 1 customers	from A	from B	from C	to A	to B	to C	July 1 customers
A	200	0	35	25	0	20	20	220
B	500	20	0	20	35	0	15	490
C	300	20	15	0	25	20	0	290

This table allows us to note net gain or loss of each dairy plus the interrelationship between the gains and losses of customers by each of the dairies.

The next step is to convert Table 3 into a more concise form, one where the gains and losses take the form of transition probabilities as illustrated in Equation 1. The rows in this matrix show the retention of customers and the gain of customers; the columns represent the retention of customers and the loss of customers.

Matrix of transition probabilities

$$
\begin{array}{c c}
 & \begin{array}{ccc} A & B & C \end{array} \\
\begin{array}{c} A \\ B \\ C \end{array} &
\begin{pmatrix} .800 & .070 & .083 \\ .100 & .900 & .067 \\ .100 & .030 & .850 \end{pmatrix}
\end{array}
$$

The columns of the matrix are read: Column I indicates that dairy A retains .800 of its customers, loses .100 of its customers to dairy B, and loses .100 of its customers to dairy C. The other columns refer to B and C respectively. Row I indicates that dairy A retains .800 of its customers, gains .070 of B's customers and gains .083 of C's customers. Again the other rows would be read similarly.

A first-order Markov chain is used which is based upon the assumption that the probability of the choice this month depends upon the customers choice last month. We wish to predict the market shares of each dairy for future periods. Assuming the matrix of transition probabilities remains fairly stable and that the July 1 market shares are A = 22%, B = 49%, and C = 29%, we can calculate the probable share of the total market likely to be had by each dairy on August 1. The following equation shows how we reach this share

$$\begin{array}{c} \text{Transition probabilities} \\ \begin{matrix} A \\ B \\ C \end{matrix} \begin{pmatrix} .800 & .070 & .083 \\ .100 & .900 & .067 \\ .100 & .030 & .850 \end{pmatrix} \end{array} \times \begin{array}{c} \text{July 1} \\ \begin{pmatrix} .22 \\ .49 \\ .29 \end{pmatrix} \end{array} = \begin{array}{c} \text{August 1} \\ \begin{pmatrix} .234 \\ .483 \\ .283 \end{pmatrix} \end{array}$$

Thus we see that we must use matrix multiplication to accomplish this task.

The probable market share on September 1 can be calculated by squaring the matrix of transition probabilities and multiplying by the July 1 market shares:

$$\begin{pmatrix} .800 & .070 & .083 \\ .100 & .900 & .067 \\ .100 & .030 & .085 \end{pmatrix}^{2} \times \begin{pmatrix} .22 \\ .49 \\ .29 \end{pmatrix} = \begin{pmatrix} .244 \\ .478 \\ .278 \end{pmatrix}$$

Following this pattern, to obtain the market shares after three periods we would have:

$$\begin{pmatrix} .800 & .070 & .083 \\ .100 & .900 & .067 \\ .100 & .030 & .085 \end{pmatrix}^{3} \times \begin{pmatrix} .22 \\ .49 \\ .29 \end{pmatrix} = \text{October 1 shares}$$

This will work for any time period. Thus to find the share of the market obtained by each dairy after n periods, we would have:

$$\begin{pmatrix} .800 & .070 & .083 \\ .100 & .900 & .067 \\ .100 & .030 & .085 \end{pmatrix}^{n} \times \begin{pmatrix} .22 \\ .49 \\ .29 \end{pmatrix}$$

Often computers are used to perform the calculations. Thus we see that Markov chains can be used in the modern business world.

Recall that in the beginning of our example, we agreed that no new customers would join the market and no old customers would leave. We know this is seldom the case. In these instances our matrix of transition probabilities would not be stable and adjustments would have to be made in our calculations. The Markov chain would still be a very important tool in computing future profits even in this case, for the actual process would be the same.[14]

We have briefly skimmed the surface of the area of Probability dealing with Markov chains. As we see from our example, a basic knowledge of matrix multiplication and the arts of computation are necessary to make our Markov chain workable.

This topic is so extensive that many books have been written on it in great thoroughness. This paper is merely an introduction to the topic stating some of the more basic definitions and theorems. The final illustration is only one of many which could have been used to show the usefulness of Markov chains in fields other than pure mathematics. More information is available and if desired can be obtained from any of the books listed in the bibliography.

---

[14]Levin and Kirkpatrick, Quantitative Approaches to Management (McGraw-Hill, 1965), p. 330.

SELECTED BIBLIOGRAPHY

(1) Bharucha-Reid, A. T., Elements of the Theory of Markov Processes and Their Applications (New York: McGraw-Hill, 1960), pp. 1 - 160.

(2) Dynkin, E. B., Theory of Markov Processes (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961).

(3) Dynkin, E. B., Markov Processes (New York: Academic Press, 1965), Vol. I, II.

(4) Feller, William, An Introduction to Probability Theory and Its Applications (New York: John Wiley & Sons, 1957), Vol. I, pp. 338 - 379.

(5) Fisz, Morek, Probability Theory and Mathematical Statistics (New York: Wiley & Sons, 1963), pp. 250 - 270.

(6) Hadley, G. and Whitin, T.M., Analysis of Inventory Systems (Prentice Hall, 1963), pp. 127 - 135.

(7) Kemeny, John and Snell, J. Laurie, Finite Markov Chains (Princeton, N.J.: D. Van Nostrand Co., Inc., 1960), pp. 1 - 206.

(8) Levin and Kirkpatrick, Quantitative Approaches to Management (McGraw-Hill, 1965), pp. 310 - 332.

(9) Parzen, Emanuel, Stochastic Processes, (San Francisco: Holden-Day, Inc., 1962), pp. 187 - 306.

(10) "A. A. Markov," Encyclopedia Britannica, 1960, XIV, 921.

(11) "A. A. Markov," Encyclopedia of Russia and the Soviet Union, 1961, 332.

(12) McIntosh, Kenneth L., "An Introduction to Finite Chains", Casualty Actuarial Society Paper.

(13) Research report, Britannica Library Research Service.

(14) Beehler, Jerry, "Markov Chains," Master's course paper, Ball State University, Autumn, 1966.