MARKOV CHAINS

by

ABDUL HAFEZ HIJJAWI
B. S. Ain Shams University, Cairo, 1958

A MASTER'S REPORT

submitted in partial fulfillment of the

reqirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

Approved by:

Gary L. Krause
Major Professor

TABLE OF CONTENTS

INTRODUCTION 1
Definition 1
Definition 2
CLASSIFICATION OF MARKOV CHAINS 3
MARKOV CHAINS 4
Transition Probabilities and Stochastic Matrices .. 5
The Chapman Kolmogrov Equation 6
Absolute Probabilities 7
DEFINITIONS 8
Decomposition of Markov Chains 9
Reducible and Irreducible Matrices 10
FINITE MARKOV CHAINS 11
REGULAR MARKOV CHAINS AND THEIR STOCHASTIC MATRICES. 12
EXAMPLES AND APPLICATIONS 15
OCCUPATION TIMES AND FIRST PASSAGE TIMES 22
VARIANCE OF FIRST PASSAGE TIME 26
RECURRENT AND NON-RECURRENT STATES 27
ERGODIC MARKOV CHAINS 31
ACKNOWLEDGEMENT 33
REFERENCES 34

INTRODUCTION

Definition

To formulate a mathematical model one can trace either one of the following approaches:

(i) The deterministic approach

(ii) The stochastic approach.

A basic role is played by the deterministic approach in the field of classical physics, where the analysis of a physical system at a given time t_2 depends functionally on its state at any earlier or later time t_1 and is independent of the state of the system before or after time t_1.

Physical phenomena which obey stochastic laws are of no less importance or less frequent occurence than those subject to deterministic laws. In the same manner, as deterministic laws, the probability that a given system will be in a given state at a given time t_2 may depend on its state at any earlier time t_1 and be independent of its state before time t_1.

Probabilists had studied dependent and independent events and their probabilities a long time before Markov. But it was he who studied a more general case of stochastic processes in 1907. A remarkable contribution from Markov which tempted probabilists to give his name to that class of stochastic processes for which the future of the process is only dependent on the present of the process and is independent of its past.

Examples of Markov processes are very numerous in physical sciences. The following example is taken from Bharucha-Reid (1960):

Consider a stochastic process $(X_n, n=0, 1, 2, \ldots)$, that is a family of random variables, defined on the space $\mathbb{X}$ of all possible values that the random variables can assume. The space $\mathbb{X}$ is called the state space of the process, and the elements x which belong to $\mathbb{X}$, the different values that X_n can assume, are called the states. In this example the state space $\mathbb{X}$ will be taken to be the set of non-negative integers x=1, 2, 3, ...

The process $(X_n, n = o, 1, 2, \ldots)$ will be said to represent a simple discrete branching process if the following conditions are satisfied:

(a) $X_0 = x_o = 1$

(b) $p(x) = P(X_1 = x)$, with $\Sigma\, p(x) = 1$

(c) The conditional distribution of $X_n = j$, is the same as the sum of j independent random variables, each having the same probability distribution as X_1. Hence if we denote by $p_{n+1}(k)$ the probability that there are k individuals in the population in the (n+1)st generation, then

$$p_{n+1}(k) = \Sigma\, p_n(j) \left[p(k)\right]^{*j}$$

in other words the conditional distribution is the j-fold convolution of p(k) with itself.

Definition

The following definition is due to Parzen (1962):

A discrete parameter stochastic process $(X(t), t = 0, 1, \ldots)$ or a continuous parameter stochastic process $(X(t), t \geq 0)$ is said to be a Markov process if, for any set on n time points $t_1 < t_2 < \ldots < t_n$

in the index set of the process, the conditional distribution of the random variable $X(t_n)$, for a given values of $X(t_1), \ldots, X(t_{n-1})$ depends only on $X(t_{n-1})$, the most recent known value; more precisely, for any real numbers $x_1, x_2, \ldots, x_n$

$$P(X(t_n) = x_n \mid X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = P(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}).$$

The set to which x_i, i=1, ..., n belongs is said to be state space and will be denoted by X and the set T, to which t_1, t_2, ..., t_n belongs is said to be the index set.

CLASSIFICATION OF MARKOV CHAINS

Markov processes can be classified according to:

(i) The index set T:

(a) T: t; t=0, 1, ..., i.e. t can take only positive integral values.

(b) T: t; t 0, i.e. T is continuous.

(ii) The state space X:

(a) X: x; x=0, 1, ..., i.e. X is discrete.

(b) X: x; x 0, i.e. X is the continuum.

Further subdivision of Markov processes may be done with respect to the number of states the process can visit, Parzen (1962):

(i) Finite: x=o, 1, ..., n.

(ii) Infinite: x=0, 1, ...

In the above classification it is noticed that the number of states is either finitely countable or infinitely countable.

Markov processes with a finite number of states were considered first by Frechet (1937).

MARKOV CHAINS

A Markov process whose state space and time space are both discrete is called a Markov chain, Kemney and Snell (1959).

Montroll (1947), classified Markov chains into simple Markov chains and multiple or complex Markov chains, where he defined a simple Markov chain as:

"By a simple Markov chain it is meant a sequence of events, each of which leads to one of ν possible results and which occur in such a manner that if the result of the kth event is a_k, the probability of the (k+1)st one yielding a result a_{k+1} is proportional to $p(a_k, a_{k+1})$".

Montroll's definition can be put in the form:

Let (X(t), t T) be a stochastic process whose state space and time space are both discrete. Such a process is said to be a simple Markov chain if:

$$P(X(t_{n+1}) = x_{n+1} | X(t_n) = x_n) = \Phi(p(x_n, x_{n+1}))$$

where $p(x_n, x_{n+1})$ is the joint probability function of the two dimensional random variable (x_n, x_{n+1}) and $\Phi(p(x_n, x_{n+1}))$ is a function of $p(x_n, x_{n+1})$.

An example of a simple Markov chain would be encountered in the analysis of a sequence of tosses of a coin with a memory. Let the probability of two successive heads or two successive tails be p, while the probability of having different faces on two successive tosses is q:

$$p(H,H) = p(T,T) = p$$

$P(H,T) = p(T,H) = q$

$p + q = 1.$

A multiple Markov chain is one for which the result of each event depends on those of its n predecessors. If each simple event leads to any one of n possible results, a multiple event could lead to any one of ν^n events.

Transition Probabilities and Stochastic Matrices

The conditional probability p_{ij} will be called the probability of a transition from state i to state j. The transition probabilities of a chain can be arranged in a matrix of the form:

$$P = \begin{Vmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{Vmatrix} .$$

P is a square matrix of size nxn, where n is the number of states the system can visit. The elements of P can take only non-negative values and each row sum is unity. P is called a Markov matrix or a stochastic matrix, Feller (1957).

A Markov chain is completely defined by a stochastic matrix P whose elements are the transition probabilities and column vector, say:

$$Q = \begin{bmatrix} q(1) \\ q(2) \\ \cdot \\ \cdot \\ \cdot \\ q(n) \end{bmatrix}$$

which gives the probability distribution for the states x = 0, 1, 2, ..., at time zero, i. e. initially, Bharucha-Reid (1960).

Transition from one state to another may occur in a period of time which is unit-time on our scale and is such a case the chain will be called a one-step chain. The transition probabilities will be denoted by $p_{ij}^{(1)}$ or p_{ij}. If transition from one state to another happens in n time-units then the chain is said to be an n-step chain. The transition probabilities, in such a case will be denoted by $p_{ij}^{(n)}$.

The Chapman-Kolmogrov Equation

Transition probabilities of an n-step Markov chain are defined recursively in the following manner, Feller (1957),

$$p_{ij}^{(1)} = p_{ij}$$

$$p_{ij}^{(n+1)} = \sum_k p_{ik}^{(n)} p_{kj} .$$

By the use of mathematical induction the following relation is seen to hold:

$$p_{ij}^{(v)} = P(X_{m+v} = j \mid X_m = i).$$

A more general form of the relation among transition probabilities can be obtained from the use of the Chapman-Kolmogrov equation:

$$p_{ij}^{(m+n)} = \Sigma p_{ik}^{(m)} p_{kj}^{(n)}.$$

The Chapman-Kolmogrov equation can be put in a matrix form, Bharucha-Reid (1960), as follows:

$$P^{m+n} = P^m P^n.$$

This functional equation, which characterizes Markov chains, is of fundamental importance in the theory of Markov chains. It is this equation which establishes the connection between Markov chains and the theory of semigroups of operators. The theory of semigroups is discussed by Hille and Phillips (1957).

One way of studying the general properties of the Markovian processes is to study the properties of the solutions of the Chapman-Kolmogrov equation. It should be realized that there exist non-Markovian stochastic processes whose transition probabilities satisfy the Chapman-Kolmogrov equation. Thus while it is true that the transition probabilities of a Markov chain satisfy the Chapman-Kolmogrov equation, it is not true that a stochastic process is Markovian if its transition probabilities satisfy the Chapman-Kolmogrov equation, Parzen (1962).

Absolute Probabilities

Let Q be the vector of initial probabilities, i. e. q(i) is

the probability of the system being at state i at time zero, i.e. initially. This implies that:

$$q(i) = P(X_0 = i).$$

The unconditional probability of finding the system at time n in state j is given by:

$$q^{(n)}(j) = \sum_i q(i)\, p_{ij}^{(n)}.$$

The unconditional probabilities are called absolute probabilities, Bharucha-Reid (1960). Hence given q(i) and the n-step transition probabilities $p_{ij}^{(n)}$, the $q^{(n)}$ can be calculated.

DEFINITIONS

The following definitions will serve as an orientation to some terminology to be used later.

Definition: A state S_j is said to be accessible from a state S_i if, for some integer $n \geq 0$, $p_{ij}^{(n)} > 0$, and this will be denoted by $S_i \dashrightarrow S_j$.

Definition: If state S_j is accessible from state S_i and state S_i is accessible from state S_j then the two states are said to be communicating and this is denoted as $S_i \dashleftarrow\dashrightarrow S_j$.

Definition: A family of communicating states is said to be a communicating class.

By the use of the Chapman-Kolmogrov equation one can show

that communicating states satisfy:

(i) If $S_i \longleftrightarrow S_j$, then $S_j \longleftrightarrow S_i$.

(ii) If $S_i \longrightarrow S_j$, then, $S_j \longrightarrow S_i$.

(iii) If $S_i \longleftrightarrow S_k$, and $S_k \longleftrightarrow S_j$, then, $S_i \longleftrightarrow S_j$.

(iv) Two communicating classes C_1 and C_2 are either identical or have no states in common.

Definition: A state which communicates with itself is said to be a return state.

Definition: A state which does not communicate withe itself is said to be a non-return state.

The above properties enable one to express a Markov chain as the union of a finite or countably infinite family of disjoint classes of states. This can be expressed in the form:

$$S = C_1 \cup C_2 \cup C_3 \cup \ldots$$

where $C_i C_j = \emptyset$, for $i \neq j$, and each set C_r is either communicating class of states, or contains exactly one non-return state.

Decomposition of Markov Chains

Let S be a set of states of a Markov chain, if no one-step transition is possible from any state into any other state outside S then S is said to be a closed set. In terms of transition probabilities a set of states is closed if:

$$p_{ij} = \begin{cases} 0, & \text{for } i,\ j \in S \\ 0, & \text{for } i \in S,\ j \in \bar{S}\ . \end{cases}$$

Definition: The smallest closed set containing S is called the closure of S.

Definition: A set which is not closed is said to be open.

Definition: A Markov chain whose space X contains two or more closed sets of states is said to be decomposable or reducible.

Definition: A chain which is not decomposable is said to be irreducible.

Reducible and Irreducible Matrices

Definition: The transition matrix of a decomposable Markov chain is said to be reducible.

Definition: The transition matrix of an irreducible Markov chain is said to be indecomposable or irreducible.

Kemney et al (1957), considered reducibility and irreducibility of Markov chains. They gave the following criterion to t test such properties of a Markov chain as:

Let M be the class of all nxn Markov matrices, where n=2 and let I be that subset of all irreducible matrices contained in M. Let A(1) be a nonempty subset of M, and for k 1 let A(k) be the set of all m M such that m can be expressed as the product of at most k elements of A(1). Let

$$A = U\, A(k)$$

then the following is true:

If, $A(2^{n^2}) \subset I$, then $A \subset I$

and hence the elements of A are irreducible matrices.

FINITE MARKOV CHAINS

So far, we have been concerned with chains whose time space is a countable set. In practice most of the chains encountered are of finite nature. Since there are only a finite number of states the system can visit.

The stochastic matrix will be written in the form:

$$P = \begin{array}{c|cccccc} & S_1 & S_2 & S_3 & \cdots & S_n \\ \hline S_1 & p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ S_2 & p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ S_n & p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{array}$$

where S_i refers to category i and classification into categories depends on the nature of the problem under consideration, i. e. S_i may be the number of bushels of apples produced or a sequence of dry-wet days, etc.

Consider a finite Markov chain with states S_i, i = 1, 2, ... n, and transition probability matrix:

$$P = \| p_{ij} \| \qquad i, j = 1, 2, 3, \ldots, n.$$

Let the probability vector at time t be given by:

$$Q(t) = \left[q_i(t)\right]$$

where Q is an nxl matrix. $q_i(t)$ is the probability that the process is in state S_i at time t. The Matrix P and the vector Q are related in the following manner:

$$Q'(t+1) = Q'(t)\, P$$

where Q' is the transpose of Q. Considering the initial state of the system:

$$Q'(t) = Q'(0)\, P^t.$$

An initial position must be chosen always as a starting point for the system.

REGULAR MARKOV CHAINS AND THEIR STOCHASTIC MATRICES

A stochastic matrix P is said to be regular if and only if, P^t has no zero entries for some positive integer t.

A Markov chain whose stochastic matrix is regular is said to be a regular chain.

Regularity of stochastic matrices implies that any state S_j, can be reached from any other state, S_i, in time t. Since a zero entry in the stochastic matrix implies that the state corresponding to that element cannot be visited. A non-zero element, $p_{ij}^{(n)}$, will indicate that there is a chance for the system to be in either state S_i or state S_j. Such a chance depends

on how large $p_{ij}^{(n)}$ is.

Tests for regularity of stochastic matrices follow a trial a and error scheme, but one may save time be considering:

$$P, P^2, P^4, \ldots.$$

The following theorem is from Kemney and Snell (1959):

If P is a regular transition matrix then:

(i) The powers P^n approach a stochastic matrix P^*.

(ii) Each row of P^* is the same probability vector, π' where:

$$\pi' = \left(\pi_1, \pi_2, \ldots, \pi_n \right) .$$

(iii) The components of π' are all positive.

(iv) For any probability vector π, πP^n approaches the vector π' as n tends to infinity.

(v) The vector π' is the unique probability vector such that:

$$P' \pi = \pi .$$

(vi) $P P^* = P^* P = P^*$.

The matrix P^* is called the limiting matrix and the vector π is called the stationary probability vector for the Markov chain determined by P and Q.

A point to be clarified concerns the convergence of the stochastic matrix. Such convergence may be fast or slow. It is left to the experimenter to explain this convergence, since it depends on the different factors and variables envolved in the process. An analytic test for that type of convergence has

not been published yet. It may draw the interest of probablists as well as algebraists in the future. Once the limiting matrix has been reached, say P^n is the limiting matrix , then whatever inference made about the system at time n, would also apply at time n+1, n+2, ...

The equation:

$$Q'(t) = Q'(0)\, P^t$$

shows that if the process started in such a way that the initial states have probability distribution Q(0), then the probability distribution for the states after time t is given by:

$$Q'(t) = Q'(0)\, P^t .$$

According to the theorem given above one has:

$$\lim_{t \to \infty} Q'(t)\, P^t = \pi'.$$

The existence of the limit is gauranteed by the theorem, and since π is dependent only on P, one can notice that Q(t) is almost independent of the initial distribution, Q(0) for sufficiently large t.

Numereous techniques may be used to obtain π or P^*. Such techniques may be numerical or analytical.

A basic quantity used to compute most of the interesting descriptive quantities for the behavior of a regular Markov chain is its fundemental matrix, given by:

$$Z = (I - P + P^*)^{-1}$$

where I is the identity matrix with the same dimensions as P.

The properties of Z may be given as follows:

(i) $P Z = Z P$.

(ii) $\pi Z = \pi$.

(iii) $Z \not{s} = \not{s}$, i. e. row sums of $Z = 1$.

(iv) $I - Z = A - PZ$.

EXAMPLE AND APPLICATION

Markov processes may be used to evaluate changes in the size distribution of firms within an industry. The general operation of this model consists of observing movement of firms between specific size categories over specific time periods and generating an equilibrium size distribution of firms which would be expected to result if the type of activity initially observed continued indefinitely. A two size-category model may be constructed as follows. Size category 1 may be specified as firms with assets of 1,000 units or over, size category 2 as firms with assets less than 1,000 units. Assume that forces associated with firm growth between time t_1 and time t_2 may be reqresented by a probability of movement from one category to another, for example $p_{11} = .6$ is the probability that firms which are in category 1 at time t_1 will remain in category 1 at time t_2, $p_{12} = .4$ is the probability that firms which are in category 1 at time t_1 will move to category 2 at time t_2, $p_{21} = .3$ and $p_{22} = .7$. These probabilities may be set in a matrix P:

$$P = \begin{Vmatrix} .6 & .4 \\ .3 & .7 \end{Vmatrix}$$

As shown above P is a regular stochastic matrix.

$$P^2 = \begin{Vmatrix} .48 & .52 \\ .39 & .61 \end{Vmatrix}$$

$$P^4 = \begin{Vmatrix} .4332 & .5668 \\ .4251 & .5749 \end{Vmatrix}$$

$$P = \begin{Vmatrix} .428571 & .571429 \\ .428571 & .571429 \end{Vmatrix}$$

and $w = (.428571 \quad .571429)$

Assuming that the initial configuration of firms is 100 in each category. This configuration at time t_1 may be expressed as a vector,

$$C_{t_1} = (100 \quad 100)$$

The expected configuration at time t_2 is obtained by matrix multiplication,

$$C_{t_2} = C_{t_1} P$$

$$= (100 \quad 100) \begin{Vmatrix} .6 & .4 \\ .3 & .7 \end{Vmatrix}$$

$$= (90 \quad 110) .$$

That is, 90 firms are in category 1 and 110 are in category 2.

The expected configuration in time t_3 would likewise be obtained by multiplying C_{t_2} times P or C_{t_1} times P^2:

$$C_{t_3} = C_{t_2} P$$

$$= (90 \quad 110) \begin{Vmatrix} .6 & .4 \\ .3 & .7 \end{Vmatrix}$$

$$= (87 \quad 113)$$

$$= C_{t_1} P^2$$

$$= (100 \quad 100) \begin{Vmatrix} .48 & .52 \\ .39 & .61 \end{Vmatrix}$$

$$= (87 \quad 113)$$

Therefore, it follows that:

$$C_t = C_{t_1} P$$

$$= (100 \quad 100) \begin{Vmatrix} .4286 & .5714 \\ .4286 & .5714 \end{Vmatrix}$$

$$= (85.71 \quad 114.29).$$

In this way the unique vector π associated with each regular matrix of transition probabilities gives a unique equilibrium configuration of firms. C_t can be obtained by:

$$C_t = 200 \ (.4286 \quad .5714)$$

$$= (85.71 \quad 114.29)$$

where 200 is the total number of firms.

A solution technique as shown by Adlman (1958), may be

outlined as follows:

$$P = \begin{Vmatrix} .6 & .4 \\ .3 & .7 \end{Vmatrix}$$

$$P' = \begin{Vmatrix} .6 & .3 \\ .4 & .7 \end{Vmatrix}$$

$$P^{*} = P' - I$$

$$= \begin{Vmatrix} -.4 & .3 \\ .4 & -.3 \end{Vmatrix}$$

Let $P^{**} = P^{*}$ with the ith row replaced by a row containing all 1's, let V be a column vector with all zero elements except the ith element which is 1, then

$$(P^{**})^{-1} V = \pi' .$$

For the above example, where i = 2,

$$P' = \begin{Vmatrix} .6 & .3 \\ .4 & .7 \end{Vmatrix}$$

$$P^{*} = \begin{Vmatrix} -.4 & .3 \\ .4 & -.3 \end{Vmatrix}$$

$$P^{**} = \begin{Vmatrix} -.4 & .3 \\ 1 & 1 \end{Vmatrix}$$

$$(P^{**})^{-1} = \begin{Vmatrix} -1/.7 & 3/.7 \\ 1/.7 & 4/.7 \end{Vmatrix}$$

$$(P^{**})^{-1} = \begin{Vmatrix} -1.42857 & .428571 \\ 1.42857 & .571429 \end{Vmatrix}$$

$$V = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$(P^{**})^{-1} V = \pi' = \begin{bmatrix} .428571 \\ .571429 \end{bmatrix}$$

Therefore: $\pi = (.428571 \quad .571429)$.

Another method to get π is illustrated in the following example:

Let

$$P = \begin{Vmatrix} .900 & .100 & .000 \\ .810 & .090 & .100 \\ .729 & .081 & .190 \end{Vmatrix} .$$

Since,

$$\Sigma p_{ij} = 1, \text{ and } p_{ij} \geq 0, \ i, j = 1, 2, 3,$$

holds for this example; P is seen to be a stochastic matrix. For t=2, one has:

$$P^t = P^2 = \begin{Vmatrix} .8910 & .0990 & .0100 \\ .8748 & .0972 & .0280 \\ .8602 & .0956 & .0442 \end{Vmatrix} .$$

It is noticed that $p_{ij} > 0$ for all i, and j, which implies that P is regular. It is noticed also that P^2 is a stochastic matrix. Consider the equation:

$$\begin{Vmatrix} .900 & .100 & .000 \\ .810 & .090 & .100 \\ .729 & .081 & .190 \end{Vmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

The above matrix equation shows that $\lambda_1 = 1$ is a characteristic root of P and the vector:

$$V_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

is a characteristic vector of P. From the theory of matrices it is known that for a constant c, cV_1 is a characteristic vector also. Solving the set of linear equations:

$$P' \pi = \pi$$

one obtains:

$$\begin{Vmatrix} .900 & .810 & .729 \\ .100 & .090 & .081 \\ .000 & .100 & .190 \end{Vmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix}.$$

We see that the stationary distribution is:

$$\pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} = \begin{bmatrix} .889 \\ .099 \\ .012 \end{bmatrix}$$

is the characteristic vector of P' corresponding to $\lambda_1 = 1$. To

examine the convergence of the stochastic matrix P^t as t increases, one could look

$$\begin{Vmatrix} .889 & .099 & .012 \\ .886 & .098 & .015 \\ .884 & .098 & .018 \end{Vmatrix}$$

and for t=5, one obtains:

$$\begin{Vmatrix} .889 & .099 & .012 \\ .889 & .098 & .015 \\ .884 & .098 & .018 \end{Vmatrix} \quad P^2 = \begin{Vmatrix} .889 & .099 & .012 \\ .889 & .099 & .012 \\ .889 & .099 & .012 \end{Vmatrix}$$

this result implies that:

$$P^5 = P^* .$$

Consider an arbitrary probability vector Q(0) given by:

$$Q(0) = \begin{Vmatrix} .66 \\ .13 \\ .21 \end{Vmatrix}$$

$$Q'(0)\, P^5 = \begin{pmatrix} .66 & .13 & .21 \end{pmatrix} \begin{Vmatrix} .889 & .099 & .012 \\ .889 & .099 & .012 \\ .889 & .099 & .012 \end{Vmatrix}$$

$$= \begin{pmatrix} .889 & .099 & .012 \end{pmatrix}$$

$$= \pi' .$$

The previous result shows that P^5 is a good approximation of P^*, and that the initial condition of the Markov chain has little effect on the distribution after 5 time-intervals. In general the value of t required to insure a good estimate of P^* cannot be predicted. One can make a good guess if he

notices the speed of convergence of the powers of P.

OCCUPATION TIMES AND FIRST PASSAGE TIMES

The states of a Markov chain may be characterized by the number of times the system visits a particular state. Two states to be considered:

(i) A state which the system can visit infinitely often.

(ii) A state which the system can visit finitely often.

Consider a Markov chain $(X_n, n=1, 2, 3, \ldots)$. For any state S_k and $n=1, 2, \ldots$, let $N_k(n)$ be the number of times that state is visited by the system in the first n transitions. $N_k(n)$ will be called the occupation time of the state S_k in the first n transitions. The total occupation time can be obtained as follows:

$$\text{Total occupation time} = N_k(\infty)$$

$$= \lim_{n\to\infty} N_k(n).$$

For a regular Markov chain one must define a function whose value depends on the number of steps before the system enters a certain state initially, after the starting position. This function has the particular state under consideration as its argument; Kemney and Snell (1959), denoted such a funtion by f_k and called it "first passage time" of the state S_k. The

mean value of the function f_k can be shown to be finite no matter which state the system was in initially. The mean first passage matrix is of great importance in the analysis of Markov chains. The mean first passage matrix is defined as:

$$M = \| m_{ij} \|$$

where,

$$m_{ij} = E_i(f_j)$$

and E_i refers to the mean of f_j with the system being in state S_i initially. Hence for an initial vector of probabilities π, the mean first passage times are the components of the vector πM.

Properties of the matrix M were discussed by Kemney and Snell (1959), and are summerized in the following:

(i) The matrix M satisfies the equation:

$$M = P (M - M_{dg}) + E$$

where,

M_{dg} = the matrix M with all the off diagonal elements equal to zero.

E = a matrix of the same dimensions as P and all its elements are unity.

(ii) Let $\pi = (\pi_1, \ldots, \pi_n)$ be the stationary probability vector of the matrix P. Then,

$$m_{ij} = 1/\pi_i \text{ , for } j=i \text{ and zero for } j \neq i.$$

the vector π, as mentioned earlier, is any row of the matrix P^*.

(iii) The equation:

$$M = P\,(M - M_{dg}) + E$$

has a unique solution.

(iv) The mean first passage matrix M is given by:

$$M = (I - Z + E\,Z_{dg})\,D$$

where Z is the fundamental matrix of P and D is the diagonal matrix whose diagonal elements are the reciprocals of the corresponding components of π and its off diagonal elements are all zeros.

(v) Let P be the transition matrix for an independent trials process. The mean first passage matrix for an independent trial process is given by:

$$M = \left[m_{ij}\right]$$

where, $$m_{ij} = 1/p_{ij}\,.$$

This follows from the fact that for a sequence of independent trials the fundamental matrix Z can be shown to be the identity matrix.

(vi) For a regular Markov chain:

$$\pi M = \left[E_{\pi_j}(f_{ij})\right]$$

$$= \pi Z_{dg} D$$

$$= z_{ij}/\pi_i$$

where, $E_{\pi}(f_j)$ is the mean of the first passage time f_j with π as the initial probability vector. η is a row vector with entries all equal to 1.

(vii) Let,

$$C = \Sigma z_{ij}$$

then,

$$C = M \pi'$$

where, π' is the transpose of π, the initial vector of probabilities.

(viii) Let π_1 and π_2 be any two initial probability vectors, then π_1 and π_2 are related as follows:

$$(E_{\pi_1}(f_j) - E_{\pi_2}(f_j)) = (\pi_1 - \pi_2)(I-Z) D$$

for proofs of these results and more details one may refer to Kemney and Snell (1959).

It was mentioned earlier that a Markov chain is completely determined by a transition matrix and an initial probability vector. With these results, one can completely determine a Markov chain by the matrix of the mean passage times, M. The numbers m_{ij}, $i \neq j$ will be used as the non-zero entries of the matrix:

$$\overline{M} = M - D .$$

This matrix has n(n-1) non-zero entries, which suffices to determine the chain completely. The matrix $\overline{M}$ is characterized by:

(i) $\bar{\bar{M}}$ has an inverse.

(ii) $\pi = (C-1)\,(\bar{\bar{M}}^{-1})'$.

(iii) $P = I + (D - E)\,\bar{\bar{M}}^{-1}$.

VARIANCE OF FIRST PASSAGE TIME

The first passage time has a variance as well as a mean. The variance of the first passage time can be determined in the usual way, namely:

$$\mathrm{Var}(f_{ij}) = E(f_{ij}^2) - (E(f_{ij}))^2 .$$

To find the variance of the first passage time, the two right hand side of the above equation need to be determined. Since $E(f_{ij})$ was determined before, so it remains to fine $E(f_{ij}^2)$. Let us consider the matrix:

$$W = \left(E(f_{ij})\right).$$

The matrix W satisfies the equation:

$$W = P(W - W_{dg}) - 2P(Z - E\,Z_{dg})D + E .$$

The values of $E(f_{ij}^2)$ are determined by:

$$W_{dg} = D(2\,Z_{dg}\,D - I)$$

the unique solution of this equation is given by:

$$W = M(2\,Z_{dg}\,D - I) + 2(Z\,M - E\,(Z M)_{dg})$$

hence,

$$\mathrm{Var}(f_{ij}) = W - M_{sq}$$

where the subscript sq is used to denote that the elements of

M_{sq} are the squares of the corresponding elements of M. As usual the variance will be denoted by M_2, i.e. the 2nd moment about the mean.

If the trials were independent then:

$$W = E\,D\,(2D - I)$$
$$= \left(\frac{1}{p_{ij}}\right)\left(\frac{2}{p_{ij}} - I\right)$$

and,

$$M_2 = E\,(D^2 - D)$$
$$= \left(\frac{1}{p_{ij}}\right)^2 - \frac{1}{p_{ij}}$$

where, D in the above expression stands for the first passage time.

RECURRENT AND NON-RECURRENT STATES

Consider a sequence of repeated trials with possible outcomes $E_1, E_2, \ldots$. Let $\acute{E}$ be an attribute of finite sequences, i. e. it is uniquely determined whether a sequence $(E_{j_1}, \ldots, E_{j_n})$ has or has not the characteristic $\acute{E}$. The attribute $\acute{E}$ defines a recurrent event if:

(i) In order that $\acute{E}$ occurs at the nth and the (n+m)th place of the sequence $(E_{j_1}, \ldots, E_{j_{n+m}})$ it is necessary and sufficient that $\acute{E}$ occurs at the last place of the subsequences $(E_{j_1}, \ldots, E_{j_n})$ and $(E_{j_{n+1}}, \ldots, E_{j_{n+m}})$.

(ii) Whenever this is the case:

$$P(E_{j_1}, \ldots, E_{j_{n+m}}) = P(E_{j_1}, \ldots, E_{j_n}) P(E_{j_{n+1}}, \ldots, E_{j_{n+m}}).$$

Let S_j be an arbitrary, but fixed, initial state of the system. Every time the system visits S_j the process recommences. Therefore the return to S_j, as defined above is a recurrent event. In terms of probabilities this is expressed as:

A state S_j is said to be recurrent if and only if,

$$Q_{jj} = 1,$$

or equivalently, $L_{jj} = 1$

where,

$$Q_{ij} = P(X_h = j \text{ infinitely often} \mid X_0 = i)$$

and,

$$L_{ij} = P(X_h = j \text{ for atleast one } h \mid X_0 = i).$$

A summary of the basic properties of recurrent events is given by Feller (1957), as follows:

(i) Recurrence time T_i of a state S_i is defined as:

Let S_i be a recurrent state with:

$$P(X_n = i) > 0$$

Given that $X_n = i$, define a random variable T_i as

$T_i = m$, if $X_{n+k} \neq i$, for $1 \leq k \leq m$, and $X_{n+m} = i$.

(ii) A state S_j is transient if, T_j 1. The necessary and sufficient condition for a state S_j to be transient is:

$$\sum_{ij} p_{ij}^{(n)} < \infty$$

and hence,

$$p_{ij}^{(n)} < \infty \quad \text{, for each i.}$$

(iii) A state S_j is a persistent nul state if $T_j = 1$, and the expected value of $T_j = \infty$. A necessary and sufficient condition for a state to be a persistent null state is:

$$\sum_{ij} p_{ij}^{(n)}$$

and,

$$\lim_{n \to \infty} p_{ij}^{(n)} = 0 \text{ , for each state } S_i .$$

(iv) A state S_j has period w 1 if,

$$p_{jj}^{(n)} = 0$$

whenever n is not divisible by w and w is the smallest integer with this property, i. e. a return to the state S_j is impossible, except, perhaps, in w, 2w, ..., steps.

(v) If S_j is persistent and aperiodic, i. e. not periodic then,

$$\lim p_{ij}^{(n)} = T_{ij} / E(T_j)$$

where, $E(T_j)$ is the mean value of T_j. In particular we have:

$$\lim p_{jj}^{(n)} = 1/E(T_j) .$$

If S_j is a null state, then, one puts:

$$E(T_j) = \infty$$

(vi) If S_j is persistent and has period w, then the equation:

$$\lim p_{jj}^{(n)} = 1/E(T_j)$$

is replaced by:

$$\lim p_{ij}^{(nw)} = w/E(T_j) .$$

Persistent states which are neither periodic nor null are called ergodic and will be discussed later.

Consider the quantity:

$$K_{ij}^{(n)} = P(X_n = j \mid X_0 = i,\ X_h \neq i,\ i = h = n)$$

in words this means:

The conditional probability of the system being in state S_j at time n given that at time zero it was in state S_i and has not been in state S_j before time n. Hence we have:

$$P(T_i = n) = K_{ii}^{(n)} .$$

Hence, the relation;

$$E(T_i) = n\,P(T_i = n)$$
$$= n\,K_{ii}^{(n)} .$$

A set of recurrent states will be called a recurrence class. Recurrence states and their classes were discussed by Parzen (1962). Recurrent states whose mean recurrence time is finite and positive are called positive. If the mean recurrence time is infinite, the state is said to be a null state.

ERGODIC MARKOV CHAINS

As mentioned before, a recurrent state which is neither null nor periodic is said to be an ergodic state.

Due to the absence of a well-agreed-upon terminology about Markov chains it is felt that quoting W. Feller(1957), in the following may save some ambiguity and orient to different terminology:

"Unfortunately, no generally accepted terminology exists, persistent states are called recurrent states which confuses by obscuring the parallism between Markov chains and recurrent events. Kolmogrov called transient states 'unessential states' but new research has shown that the main interest, both theoritical and practical, centers on transient states. The term 'ergodic', being synonymous to 'persistent', 'non-null', 'non-periodic', is rather generally accepted, but 'positive' state is one of the existing alternatives, and sometimes ergodic is equated to persistent".

An ergodic chain is simply a chain where the system can go from every state to any other state. Hence an ergodic chain is a chain which consists of one and only one comunic-

ating class. The stochastic matrix of an ergodic chain will be called an ergodic transition matrix.

ACKNOWLEDGEMENT

The author wishes to aexpress his sincere appreciation to his major professor, Dr. Gary F. Krause, who gave every helpful suggestion and criticism in the writing of this report. Appreciation is also extended to Dr. H. C. Fryer and all the faculty members in the Department of Statistics for their continuous encouragement and fruitful instruction through his graduate study. A deep feeling of indebtedness is extended to his wife and to his children for their understanding and confidence and the many spoiled weekends.

REFERENCES

Adelman, I. G.
"A Stochastic Analysis of the Size Distribution of Firms," Journal of American Statistical Association, 1958, 53:893-904

Anderson, T. W. and Goodman, Leo A.
"Statistical Inference About Markov Chains," Annals of Mathematical Statistics, March, 1957, 28:89-110.

Bharucha-Reid, A. T.
Elements of the Theory of Markov Processes and Their Applications. New York: McGraw-Hill Book Company, Inc., 1960.

Feller, William.
An Introduction to Probability Theory and its Applications. New York: John Wiley and Sons, Inc., 1957.

Foster, F. G.
"Markov chains with Enumerable Number of States and a Class of Cascade Processes." Proc. Cambridge Phil. Soc.,1952, 48:587-591

Kemney, John G. and Snell, J. Laurie.
Finite Markov Chains. New York: D. Van Nostrand Company Inc., 1959.

Kemney, John G., Snell, Laurie and Thompson, G. L.
Introduction to finite mathematics. Englewood Cliffs, N. J.: PrenticeHall Inc., 1957.

Montroll, Elliott.
"On the Theory of Markoff Chains." The Annals of Mathematical Statistics, March, 1947, 18:18-38

Parzen, Emanuel.
Stochastic Processes. San Francisco: Holden-Day, Inc., 1962.

Phillips, R. S. and Hille, E.
"Fundamental Analysis and Semi-groups". New York: American Mathematical Society, 1957.

MARKOV CHAINS

by

ABDUL H. HIJJAWI

B. SC. AIN SHAMS UNIVERSITY, CAIRO, 1958

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1964

In the applications of probability theory one can often assume that the random events or random variables under consideration are independent. There are many problems in science where the assumptions of independence are not satisfied. The study of dependent random events is fof great importance in probability theory. To abandon the assumption of independence, among random events creates serious complications in the reasoning and in the proofs.

Markov, in the investigation of dependent events, distinguished a scheme of events, which can be considered as the simplest generalization of the scheme of independent events, and that was named after him as "Markov Processes". In a Markov process, the outcome of any event in a sequence of events depends only on the outcome of the directly preceding event.

Study of the stochastic matrix of a Markov chain enables one to describe the behavior of the process, as well as classifying it such as, communicating, regular, ergodic, absorbing, etc. Powers of the stochastic matrix gives a clue about the rate of convergence of that matrix. A fast convergence implies that the process will reach equillibrium in a short period of time, while a slow convergence indicates that the process needs a long time to reach its stationary distribution. The powers of the stochstic matrix where used as a tool to predict the distribution of the process at a certain time, with respect to a prespecified starting position.

Markov chains are found to be a suitable tool to analyse

some economic data.