FINITE MARKOV CHAIN THEORY AND ITS CONNECTION WITH MATRIX THEORY

A Thesis

Presented to

the Faculty of the Department of Mathematics

University of Houston

In partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

ames

Tsu-ching J. Teng

August 1969

FINITE MARKOV CHAIN THEORY AND ITS CONNECTION WITH MATRIX THEORY

An Abstract of a Thesis

Presented to

the Faculty of the Department of Mathematics

University of Houston

In partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Tsu-ching J. Teng

August 1969

ABSTRACT

To study finite Markov chains, we begin with the theory of order relations to classify states and chains. Then we define various functions on the chain and use the theory of probability and statistics to find their means and variances. Throughout the whole study, however, the connection with matrix theory is built-in since a finite Markov chain can be represented as a stochastic matrix.

Many questions concerning finite Markov chains can be answered, directly or indirectly, by investigating only two kinds of chains: absorbing Markov chains and regular Markov chains. Though these chains are different, the studies of these chains offer many striking similarities.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

§ I. Basic concepts and definitions.

In the study of an experiment which takes place in stages, we usually indicate the possible outcomes by a tree-shaped diagram. Each possible sequence of outcomes may be identified with a path through the tree. Each path consists of line segments called branches. We can assign probabilities to the branches and call them branch probabilities since it is assumed that the probability for each outcome at a given stage is known when the previous stages are known. The weight of a path is just the product of the probabilities assigned to the components of the path. For each j, we obtain a tree U_j which indicates all possible outcomes of the first j stages. The set of all paths of this tree may be considered a suitable probability space for any statement whose truth value depends on the outcome of the first j experiments.

Let U_n be the set of all paths of a tree for an n stage experiment. Let f_j be a function with domain U_n and value the outcome at the j-th stage. Then the functions $f_1, f_2, \ldots, f_n$ are called outcome functions. The set of functions $f_1, \ldots, f_n$ is called a finite stochastic process.

For two statements p and q relative to the same probability space, let $p \wedge q$ denote the statement that is true if both p and q are true. The conditional probability of p given q is denoted by $Pr[p|q]$.

Definition 1.1 A finite Markov process is a finite stochastic precess such that, for any statement P whose truth value depends only on the outcomes before the n-th,

$$Pr[f_n = S_j \mid (f_{n-1} = S_i) \wedge P] = Pr[f_n = S_j \mid f_{n-1} = S_i] .$$

For a finite Markov process, therefore, we can define the n-th step transition probability, denoted by $P_{ij}(n)$, to be

$$P_{ij}(n) = Pr[f_n = S_j \mid f_{n-1} = S_i] .$$

Definition 1.2 A finite Markov chain is a finite Markov process such that the transition probabilities $P_{ij}(n)$ do not depend on n. We denote them by P_{ij} and call any possible outcome a state.

Definition 1.3 A matrix is called nonnegative if all the entries are nonnegative real numbers. If $B = \{b_{ij}\}$ is a nonnegative m x n matrix such that

$$\sum_{j=1}^{n} b_{ij} = 1, \quad i = 1, \ldots, m.$$

then B is stochastic.

Theorem 1.1 If A and B are stochastic matrices and A•B is defined, then A•B is stochastic.

Definition 1.4 Let f be a function with domain $U = \{u_1, \ldots, u_k\}$, a probability space and range $R = \{r_1, \ldots, r_n\}$, another probability space for the same experiment. The induced measure for f is the probability measure on the set R given by $Pr[f = r_i]$, $i = 1, \ldots, n$.

§ 2. Matrix representation of a finite Markov chain.

Definition 1.5 The transition matrix for a finite Markov chain is the matrix P with entries P_{ij}, the transition probabilities.

We see immediately, by Definition 1.3, that any transition matrix is a stochastic matrix.

Definition 1.6 The initial probability vector is the row vector $\pi_o = \left\{ P_j^{(o)} \right\} = \left\{ \Pr \left[f_o = S_j \right] \right\}$ where f_o is the outcome function with value the initial position.

§ 3. Basic connection with matrix theory.

The matrix representation of a finite Markov chain is clearly justified by the following two important theorems:

Theorem 1.2 Let f_n be the outcome function at n-th stage for a finite Markov process involving r states, then

$$\Pr \left[f_n = S_v \right] = \sum_{u=1}^{r} \Pr \left[f_{n-1} = S_u \right] \cdot P_{uv}(n).$$

Proof: Let Q be the set of finite sequences of n positive integers chosen from 1 to r such that $\left\{ j, k, \cdots, u \right\} \varepsilon Q$ if and only if $S_j, S_k, \cdots, S_u, S_v$ form a path through the tree U_n. Clearly, we have

$$\Pr \left[f_n = S_v \right] = \sum_{\{j,k,\cdots,u\}\varepsilon Q} \Pr \left[f_o = S_j \wedge f_1 = S_k \wedge \cdots \wedge f_{n-1} = S_u \wedge f_n = S_v \right]$$

$$= \sum_{\{j,k,\cdots,u\}\varepsilon Q} \Pr \left[f_o = S_j \wedge f_1 = S_k \wedge \cdots \wedge f_{n-1} = S_u \right] \cdot \Pr \left[f_n = S_v \middle| f_o = S_j \wedge \cdots \wedge f_{n-1} = S_u \right]$$

Since we are dealing with Markov processes, By Definition 1.1, this is

$$\sum_{\{j,k,\cdots,u\}\varepsilon Q} \Pr \left[f_o = S_j \wedge \cdots \wedge f_{n-1} = S_u \right] \cdot P_{uv}(n).$$

By keeping u fixed and summing over the remaining indices, we obtain $\Pr \left[f_n = S_v \right] = \sum_{u=1}^{r} \Pr \left[f_{n-1} = S_u \right] P_{uv}(n)$. This completes the proof.

Theorem 1.3 Let π_n be the induced measure for the outcome function f_n for a finite Markov chain with initial probability vector π_o and transition matrix P. Then $\pi_n = \pi_o \cdot P^n$.

Proof: Let $P_j^{(n)} = Pr[f_n = S_j]$, then $\pi_n = \{P_1^{(n)}, \ldots, P_r^{(n)}\}$. By the previous theorem, we have for $n \geq 1$, $\pi_n = \pi_{n-1} \cdot P(n)$ where $P(n) = \{P_{ij}(n)\}$. Apply this result successively, we obtain $\pi_n = \pi_o \cdot P(1) \cdot P(2) \cdots P(n)$. If the Markov process is actually a Markov chain, then all the $P(n)$'s are the same and we have $\pi_n = \pi_o \cdot P^n$.

In addition to basic connection with matrix theory established by these two theorems, we need the following definitions and theorems to see further connections.

Definition 1.7 A square matrix with each entry 0 or 1 is called a permutation matrix if there is only one non-zero entry in each row and each column.

Definition 1.8 A nonnegative square matrix A is reducible if there exists a permutation matrix P such that $PAP^T = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$ where B and D are square. Otherwise, A is irreducible.

Definition 1.9 An irreducible matrix A is said to be primitive if it has a characteristic root r with the property that, if d is any characteristic root of A other than r, then $|d| < |r|$.

Theorem 1.4 A nonnegative square matrix is primitive if and only if A^p is a positive matrix for some positive integer p.

Definition 1.10 Let both $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ be r x s matrices. Then $A \geq B$ means that $a_{ij} \geq b_{ij}$ for all i and j.

CHAPTER II

EXTERNAL STRUCTURE OF THE THEORY

As basically a part of the theory of probability and statistics, finite Markov chain theory also relies on the theory of matrices and the theory of order relations.

§ 1. Theory of order relations.

Let T be a weak ordering defined on a finite set U, then T is reflexive and transitive, but not necessarily symmetric for every pair of elements in U. Thus we can obtain an equivalence relation $\widetilde{T}$ by letting $x\widetilde{T}y$ if and only if xTy and yTx. Consider the set $\widetilde{U}$ of all equivalence classes resulting from $\widetilde{T}$. For each pair of classes u and v in $\widetilde{U}$, let uT^*v hold if every element of u bears the relation T to every element of v. Then T^* is a weak ordering that is never symmetric for each pair of classes in $\widetilde{U}$. Therefore, T^* is actually a partial ordering on $\widetilde{U}$; and we have minimal and maximal elements of T^* in $\widetilde{U}$.

Let $U = \{u_1, \ldots, u_n\}$ be the set of states of a finite Markov chain. Let u_iTu_j mean that the process can go from state u_i to state u_j or that $u_i = u_j$. Then $\widetilde{T}$ partitions U into equivalemce classes where two states are in the same class if the process can go from either one of them to the other. Moreover, T^* partially orders all equivalence classes so that we can classify the states of a chain through the following definition.

<u>Definition 2.1</u> The minimal elements of the partial ordering T^* of equivalence classes obtained from $\widetilde{T}$ are called ergodic sets. The remaining elements are called transient sets. The elements of a transient set are called transient states. The elements of an ergodic

set are called ergodic states. If an ergodic set has only one state, then that state is called an absorbing state.

For every finite Markov chain, there must be at least one ergodic set since a finite partial ordering must have at least one minimal element. Certainly, it is possible for a chain to have no transient set. Therefore, we reach the following preliminary classification.

(1) Ergodic chains. These chains consist of a single ergodic set. If a chain does not have any transient set but has more than one ergodic set, it may be studied separately as several ergodic chains insomuch that there is no interaction between them.

(2) Absorbing chains. A chain all of whose non-transient states are absorbing is called an absorbing chain. As will be seen in Theorem 3.1, for such a chain the process is eventually trapped in a single (absorbing) state. In general, in any chain having transient sets, the process moves toward the ergodic sets; and it can not leave an ergodic set once entered it. Therefore, questions concerning the behavior of the chain after entering an ergodic set can be answered by considering that particular ergodic set as an ergodic chain. If we are only concerned about its behavior up to the moment that it enters an ergodic set, we may reduce the chain to an absorbing one by making all ergodic states into absorbing states since the nature of the ergodic states is entirely irrelevant to our concern.

To reach further classification, let us consider again the equivalence relation $\widetilde{T}$ which partitions the states of a chain into equivalence classes. By means of a number-theoretical result, it can be shown that a given equivalence class consists of one or more cyclic classes. The process moves cyclically from class to class. After sufficient time has elapsed, the process can be in any state of the one cyclic class to which the originating state belongs.

This result is obtained, however, by forbidding the process to leave the equivalence class in which we are concerned. Therefore, we can apply this result unconditionally to any ergodic set since the process will never leave the set once entered. Accordingly, we subdivide ergodic chains via following definitions.

Definition 2.2 A regular Markov chain is an ergodic chain containing only one cyclic class. For such a chain, its transition matrix is called a regular transition matrix.

Definition 2.3 A cyclic Markov chain is an ergodic chain having more than one cyclic classes.

We observe immediately that an ergodic chain is regular if and only if there exists nonzero entries on the main diagonal. On the other hand, the transition matrix of an ergodic chain has all zeros down the main diagonal only in case the chain is cyclic.

Regualr chains can be interpreted as a special case of cyclic chains by taking the number of cyclic classes to be 1. This special case, however, turns out to be the most important case of cyclic chains. Theoretical problems concerning cyclic chains are much easier to handle if the chain happens to be a regular one. Moreover, results obtained in this manner can be easily generalized and become applicable to any ergodic chains (techniques of generalization will not be discussed in this paper).

If we combine this result with our preliminary classification, we see that regular chains and absorbing chains should be investigated first in more detail.

§ 2. Theory of probability and statistics.

To investigate the behavior of a chain, we have to define suitable functions on the set of all states and find means and variances of these functions. Results from probability and statistics make up the bulk of the theory.

§ 3. Connections with matrix theory.

We can put the transition matrix of a chain containing transient sets into a canonical form that is much easier to deal with. The idea is to simultaneously permute the rows and columns of a transition matrix so that the ergodic states come first. In other words, there is a permutation matrix H such that

$$HPH^T = \begin{array}{cc} \overbrace{}^{r-s} & \overbrace{}^{s} \\ \end{array} \hspace{-6em} \begin{array}{c} \\ \left[\begin{array}{cc} S & O \\ R & Q \end{array}\right] \end{array} \begin{array}{l} \\ \}\, r-s \\ \}\, s \end{array}$$

where P is the transition matrix of a chain containing s transient states and r-s ergodic states. The region O must consist entirely of 0's since all states involved in the region S are ergodic, and the process never goes from an ergodic set to a transient one. By definition 1.8, P is a reducible matrix.

If we explore the reducibility of S and Q successively, P can be brought into the following form which is usually called the canonical form of a reducible matrix, [1, p. 74] .

$$\begin{bmatrix} A_{1,1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & A_{2,2} & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & A_{g,g} & 0 & \cdots & 0 \\ A_{g+1,1} & A_{g+1,2} & \cdots & A_{g+1,g} & A_{g+1,g+1} & \cdots & A_{g+1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,g} & A_{n,g+1} & \cdots & A_{n,n} \end{bmatrix}$$

where $A_{k,k}$ is irreducible for $k = 1, \ldots, n$. A_{ii} is called an isolated block if $1 \leq i \leq g$. It is interesting to note that the states involved in an isolated block must form an ergodic set, while those involved in non-isolated blocks must be transient states.

Of course, the transition matrix of an ergodic chain is irreducible. Therefore, our observation on transition matrices concerning their being stochastic and their reducibility has been fruitful enough to warrant believing there is a strong connection between finite Markov chain theory and the theory of matrices.

For example, it follows from Theorems 1.3 and 1.4 that a regular transition matrix must be primitive since a chain is regular if and only if it is possible to be in any state after a certain number of steps regardless of the starting state.

As will be seen in Theorem 3.3, the limiting vector α for a regular Markov chain is the unique probability vector such that $\alpha P = \alpha$ where P is the transition matrix of this regular chain. Since $\alpha^T = (\alpha P)^T = P^T\alpha^T$, this theorem merely states that the limiting vector is actually the transpose of the probability eigenvector of P^T corresponding to the eigenvalue 1. Naturally, the probability eigenvector is defined to be the eigenvector whose entries add up to 1. 1 is certainly an eigenvalue for P and P^T since P is stochastic, and P and P^T are similar matrices.

Another striking connection with matrix theory lies in the following theorem which is of essential importance in deriving most of our formulas.

<u>Theorem 2.1</u> For a square matrix A, if $\lim_{n\to\infty} A^n = 0$, the zero matrix, then (I-A) has an inverse and $(I-A)^{-1} = \sum_{k=0}^{\infty} A^k$.

Proof: By hypothesis, $\lim_{n\to\infty} (I-A^n) = I$, but $I-A^n = (I-A) \cdot \sum_{k=0}^{n-1} A^k$ ----(1)

Since $\det(I) = 1$, there exists a positive integer N such that $\det(I-A^N) \neq 0$. Hence, $0 \neq \det\left((I-A) \cdot \sum_{k=0}^{N-1} A^k\right) = \det(I-A) \cdot \det\left(\sum_{k=0}^{N-1} A^k\right)$ which implies that $\det(I-A) \neq 0$. Therefore I-A has an inverse.

Multiply both sides of (1) by $(I-A)^{-1}$, we have

$(I-A)^{-1} \cdot (I-A^n) = \sum_{k=0}^{n-1} A^k$, and also

$\sum_{k=0}^{\infty} A^k = \lim_{n\to\infty} (I-A)^{-1} \cdot (I-A^n) = (I-A)^{-1} \cdot \lim_{n\to\infty} (I-A^n) = (I-A)^{-1}$ which completes the proof.

CHAPTER III

INTERNAL STRUCTURE OF THE THEORY

As was pointed out in closing §1 of chapter II, there are mainly two kinds of chains to be studied: absorbing Markov chains and regular Markov chains. It is quite surprising, as will be seen, that the process of our investigation into both chains are structurally the same.

§ 1. Asking legitimate questions.

Questions that could possibly be answered differ widely from chain to chain. Asking legitimate questions constitutes a part of the theory which is just as important as are the answers to these questions.

In a regular chain, the process keeps moving through all the states no matter where it starts. Thus given any pair of states S_i and S_j, it does make sense to study the length of time to go from S_i to S_j for the first time.

Definition 3.1 For a regular Markov chain, the first passage of time f_k is a function whose value is the number of steps before entering S_k for the first time after the initial position.

We can find the mean and variance of the function f_k and put the results in matrix form. The mean first passage matrix, denoted by M, is the matrix $\{m_{ij}\} = \{M_i[f_j]\}$ where $M_i\,[f_j]$ is the mean of f_j computed at S_i.

More often than not the first passage of time function is undefined on an absorbing chain, because the transition of S_i to S_j may never be accomplished. We do know, however, that the process will eventually be trapped by an ergodic set. Therefore, it is legitimate to ask how many steps are needed for absorption in a process starting in a transient state.

This problem can be easily solved as a byproduct of another important result which furnishes us the most important information about an absorbing chain. This concerns the matrix $\left\{ M_i \left[n_j \right] \right\}$ where $M_i \left[n_j \right]$ is the mean of the function n_j evaluated at S_i. We define n_j to be the function giving the total number of times that the process is in S_j.

§ 2. Fundamental theorems.

The following two theorems, Theorem 3.1 and 3.2, are of fundamental importance in developing theories of absorbing chains and regular chains respectively.

Theorem 3.1 In any finite Markov chain, no matter where the process starts, the probability after n steps that the process is in an ergodic state tends to 1 as n tends to infinity.

Proof: If the process starts in an ergodic state, then it can never leave that ergodic set to which the initial position belongs. The theorem holds trivially in this case. Suppose the process starts in a transient state. By Definition 2.1, this state belongs to an equivalence class (resulting from $\widetilde{T}$) which is not a minimal element of the partial ordering T^*. Therefore, it must be possible to reach one of the minimal elements, i.e. ergodic sets. For each transiont state S_i, let h_i be the number of steps after which the process has a possibility to reach an ergodic set starting in S_i. Put $h = \max \left\{ h_i \,\middle|\, S_i, \text{ a transient state} \right\}$. For each transient state S_i, let p_i be the probability to reach an ergodic set in h steps starting in S_i. Put

$$p = \min \left\{ p_i \,\middle|\, S_i, \text{ a transient state} \right\}.$$

Therefore, from any transient state, the probability of entering an ergodic state in at most h steps is at least p. For each transient state S_i, let $q_i^{(n)}$ be the probability of not reaching an ergodic state in n steps. Let $d_n = \max\left\{ q_i^{(n)} \,\middle|\, S_i, \text{ a transient state} \right\}$. Then $\{d_n\}$ is clearly a monotonically decreasing sequence which is bounded below by 0. Hence, $\{d_n\}$ converges. Since $d_h = 1 - p$, we have $d_{kh} = (1-p)^k$ for each positive integer k. Thus $\lim_{k\to\infty} d_{kh} = \lim_{k\to\infty} (1-p)^k = 0$. We have found a subsequence of $\{d_n\}$ converges to 0. Therefore, $\lim_{n\to\infty} d_n = 0$ which completes the proof.

<u>Corollary</u> There are numbers $b > 0$, $0 < c < 1$ such that $P_{ij}^{(n)} \leq b \cdot c^n$ for any pair of transient states S_i and S_j and any nonnegative integer n.

Proof: Choose $c = (1-p)^{1/h}$ and $b = \frac{1}{1-p} = c^{-h}$, where p and h are as defined in the above proof. For each nonnegative integer n, $n = kh + n_1$ for some nonnegative integer k and $0 \leq n_1 < h$. Clearly,

$d_n \leq d_{kh} = (1-p)^k = c^{n-n_1} \leq c^{-h} \cdot c^n = b \cdot c^n$ since $\{d_n\}$ is non-increasing. The corollary follows by noticing that $P_{ij}^{(n)} \leq d_n$ for any pair of transient states S_i, S_j and any nonnegative integer n.

<u>Lemma</u> Let P be an r x r transition matrix having no zero entries. Let ϵ be the smallest entry in P. Let x be any r-component column vector, having maximum component M_0 and minimum component m_0, and let M_1 and m_1 be the maximum and minimum components for the vector Px. Then

$M_1 \leq M_o$, $m_1 \geq m_o$, and $M_1 - m_1 \leq (1-2\epsilon)\cdot(M_o-m_o)$.

Proof: Let x' be the vector obtained from x by replacing all components, except one m_o component, by M_o. Since P is stochastic, each component of Px' is of the form $a\cdot m_o + (1-a)\cdot M_o = M_o - a(M_o-m_o)$ where $a \geq \epsilon$. Thus each such component is less than or equal to $M_o - \epsilon(M_o-m_o)$. Since $x \leq x'$, $Px \leq Px'$ and M_1 is a component of Px, we have

$$M_1 \leq M_o - \epsilon(M_o-m_o) \quad \text{---------------}(1)$$

Apply this result to the vector $-x$, we obtain

$$-m_1 \leq -m_o - \epsilon(-m_o+M_o) \quad \text{-------------}(2)$$

Adding (1) and (2), we have

$$M_1 - m_1 \leq M_o - m_o - 2\epsilon(M_o-m_o) = (1-2\epsilon)\cdot(M_o-m_o).$$

Theorem 3.2 If P is a regular transition matrix, then

(i) $\lim_{n\to\infty} P^n = A$ where A is stochastic.

(ii) $A = \xi\alpha$ where ξ is a column vector having all components equal to 1 and α is a probability vector.

(iii) The components of α are positive.

Proof: We shall prove the theorem under two cases. Assume first that P has no zero entries. Let ϵ be the minimum entry. Let P_j be a column vector with a 1 in the j-th component and 0 in the remaining components. Let M_n and m_n be the maximum and minimum components of the vector P^nP_j. Since $P^nP_j = P\cdot P^{n-1}\cdot P_j$, we have, from the previous lemma, that

$$M_1 \geq M_2 \geq \,.\,.\,.\,.\,.\,.\,.\,.\,. \qquad \text{----------}(1)$$

$$m_1 \leq m_2 \leq \,.\,.\,.\,.\,.\,.\,.\,.\,. \qquad \text{----------}(2)$$

and $M_n - m_n \leq (1-2\epsilon)\ (M_{n-1}-m_{n-1})$ for $n \geq 1$. Since both $\{M_n\}$ and $\{m_n\}$ are bounded above and below by 1 and 0 respectively, it follows from (1) and (2) that both sequences converge. Put $d_n = M_n - m_n$, then $d_n \leq (1-2\epsilon)^n \cdot d_o = (1-2\epsilon)^n$. Since P is stochastic, $0 < \epsilon \leq 1/2$ if P has more than one entry. Thus the sequence $\left\{(1-2\epsilon)^n\right\}$ converges to zero. This makes $\{d_n\}$ also converge to 0 by the comparison test. Therefore, $\lim\limits_{n\to\infty} M_n = \lim\limits_{n\to\infty} m_n$. In other words, $\lim\limits_{n\to\infty} P^n P_j$ exists and is a column vector with all components the same for $j = 1,\ldots\ldots, r$. Let a_j be this common value, then $m_n \leq a_j \leq M_n$ for $n = 1, 2, \ldots.$ and $j = 1, \ldots\ldots, r$. In particular, $0 < m_1 \leq a_j \leq M_1 < 1$, $j = 1, \ldots\ldots, r$. This would prove (iii) if $\alpha = (a_1 \ldots a_j \ldots a_r)$, where α is the probability vector required in (ii). It turns out that we really have $\lim\limits_{n\to\infty} P^n = \xi\alpha$ since $P^n P_j$ is actually the j-th column of P^n. By Theorem 1.1, P^n is stochastic for $n = 1, 2, \cdot\cdot\cdot$, $\lim\limits_{n\to\infty} P^n$ must also be stochastic. This completes the proof for the first case. Consider next the case that P is only assumed to be regular. Let N be such that P^N has no zero entries. Applying the first part of the proof, we have

$d_{kN} \leq (1-2\epsilon')^k$ where ϵ' is the smallest entry of P^N. Therefore, the non-increasing sequence $\{d_n\}$ has a subsequence tending to zero. Thus $\{d_n\}$ tends to zero since $\{d_n\}$ is bounded below. This reduces the present case to the previous one and completes our proof.

The following theorem is a direct consequence of the above one.

<u>Theorem 3.3</u> If P is a regular transition matrix and A and α are as given in Theorem 3.2, then

(i) For any probability vector π, $\lim_{n\to\infty} \pi P^n = \alpha$.

(ii) The vector α is the unique probability vector such that $\alpha P = \alpha$.

(iii) PA = AP = A.

<u>Definition 3.2</u> The matrix A and vector α, as given by Theorem 3.2 and 3.3, are called the limiting matrix and limiting vector for the Markov chain determined by P.

These fundamental theorems determine, both theoretically and technically, the process of our investigation into both kinds of chains. In fact, they clearly reveal the general behavior pattern and trend of both kinds of chains. In an absorbing chain, the process moves toward inevitable absorption. A regular chain, on the other hand, will eventually reach a state of equilibrium--though it never can stop once the process started.

By Theorem 3.1 and its corollary, we can prove that $M_i[n_j]$ and $Var_i[n_j]$, which were defined in §1, are finite.

<u>Theorem 3.4</u> $M_i[n_j]$ is finite for any absorbing chain and any pair of transient states S_i and S_j.

Proof: Let u_j^k be a function, defined on the set of all states of an absorbing chain, that is 1 if the process is in state S_j after k steps, and is 0 otherwise. Then, $M_i[n_j] = M_i(\sum_{k=0}^{\infty} u_j^k) = \sum_{k=0}^{\infty} M_i[u_j^k]$. Clearly, $M_i[u_j^k]$ is the probability that the process is in S_j on step k starting in S_i. Hence, $M_i[n_j] = \sum_{k=0}^{\infty} P_{ij}^{(k)}$. By the corollary to Theorem 3.1, there are numbers $b > 0$ and $0 < c < 1$ such that

$P_{ij}^{(k)} \leq b \cdot c^k$ for $k = 0, 1, \cdots,$ Thus

$$\sum_{k=0}^{\infty} P_{ij}^{(k)} \leq \sum_{k=0}^{\infty} b \cdot c^k = b \sum_{k=0}^{\infty} c^k \text{ which is finite.}$$

<u>Theorem 3.5</u> $Var_i[n_j]$ is finite for any absorbing chain and any pair of transient states S_i and S_j.

Proof: Since $Var_i[n_j] = M_i[n_j^2] - M_i[n_j]^2$ and $M_i[n_j]$ is finite, it remains to be shown that $M_i[n_j^2]$ is finite.

$$M_i[n_j^2] = M_i\left[(\sum_{k=0}^{\infty} u_j^k)^2\right] = M_i\left[\sum_{k=0}^{\infty}\sum_{q=0}^{\infty} u_j^k u_j^q\right]$$

$$= \sum_{k=0}^{\infty}\sum_{q=0}^{\infty} M_i\left[u_j^k u_j^q\right].$$

Clearly, $M_i[u_j^k u_j^q]$ is the probability that the process is in state S_j both on step k and q starting in S_i. Let $m = \min\{k, q\}$, $d = |k-q|$; then $M_i[u_j^k u_j^q]$ is the probability of being in S_j after m steps, and of returning d steps later. Hence, $M_i\left[u_j^k \; u_j^q\right] = P_{ij}^{(m)} P_{jj}^{(d)}$. By the corollary to Theorem 3.1, there are numbers $b > 0$ and $0 < c < 1$ such that $P_{ij}^{(m)} \leq b \cdot c^m$ and $P_{jj}^{(k)} \leq b \cdot c^k$. Thus

$$M_i\left[n_j^2\right] = \sum_{k=0}^{\infty} \sum_{q=0}^{\infty} P_{ij}^{(m)} \; P_{jj}^{(d)}$$

$$\leq \sum_{k=0}^{\infty} \sum_{q=0}^{\infty} (b \cdot c^m)\,(b \cdot c^d)$$

$$= b^2 \sum_{k=0}^{\infty} \sum_{q=0}^{\infty} c^t \quad \text{where } t = \max\left\{k, q\right\}$$

$$= b^2 \sum_{t=0}^{\infty} (2t + 1)\, c^t \text{ which is finite.}$$

By virtue of these two theorems, the two matrices $\left\{M_i[n_j]\right\}$ and $\left\{Var_i[n_j]\right\}$ can be found. Many other results on absorbing chains can be easily obtained from the matrix $\left\{M_i[n_j]\right\}$. As an example, let b_{ij} be the probability that the process starting in S_i ends up in an absorbing state S_j. We can show that $\left\{b_{ij}\right\} = \left\{M_i[n_j]\right\} \cdot R$ where R is that submatrix of the transition matrix such that r_{ij} is an entry of R only in case S_i is transient and S_j is absorbing.

Unlike the finite feature of absorbing chains, most results on regular chains appear in limiting form. Of course, this is due to Theorem 3.2, the fundamental theorem of regular chains. The following theorem, known as the Law of Large Numbers, illustrates how the limiting process helps to give quantitative estimates concerning the behavior of a regular chain.

Theorem 3.6 Given a regular Markov chain with limiting vector $\alpha = (a_1\ a_2 \cdots a_r)$, for any initial vector π, $\lim_{n\to\infty} M_\pi \left[v_j^{(n)}\right] = a_j$, where $v_j^{(n)}$ is the probability of passing S_j in first n steps (not counting the initial position).

This theorem states, to our amazement, that in the long run we can expect about a fraction a_j of the steps to be in S_j no matter what initial probability vector we start with.

Let $\bar{y}_j^{(n)}$ be the number of times that the process is in state S_j in the first n steps including the initial position. In an attempt to get a mean of this function $\bar{y}_j^{(n)}$, we obtain a better-than-nothing estimation of $M\left[\bar{y}_j^{(n)}\right]$.

Theorem 3.7 For any regular Markov chain and any initial vector π,

$$\lim_{n\to\infty}\left(\left\{M_\pi\left[\bar{y}_j^{(n)}\right]\right\} - n\alpha\right) = \pi Z - \alpha$$

where α is the limiting vector and Z is the fundamental matrix for a regular chain which will be introduced in next section.

To see how accurate this prediction about $\bar{y}_j^{(n)}$ is, again we must appeal to limiting process to extablish the following result.

Theorem 3.8 $$\lim_{n\to\infty}\left\{\mathrm{Var}\left[\frac{\bar{y}_j^{(n)}}{\sqrt{n}}\right]\right\} = \left\{a_j\ (2z_{jj} - 1 - a_j)\right\}.$$

§3. Fundamental matrices.

Let Q be the submatrix of the transition matrix of an absorbing Markov chain such that P_{ij} is an entry of Q if and only if both S_i and S_j are transient states. By Theorem 3.1, the process will inevitably be trapped by an absorbing state. Using matrix language which is made available by Theorem 1.3, this says that Q^k tends to a zero matrix as k tends to infinity. It follows from Theorem 2.1 that I-Q has an inverse and $(I-Q)^{-1} = \sum_{k=0}^{\infty} Q^k$.

Definition 3.3 For an absorbing Markov chain, we define the fundamental matrix N to be $(I-Q)^{-1}$.

For a square matrix A, let A_{dg} denote the square matrix that agrees with A on the main diagonal but is zero elsewhere. The matrix A_{sq} is formed from A by squaring each entry.

Theorem 3.9

(i) $\left\{ M_i[n_j] \right\} = N$, $\left\{ Var_i[n_j] \right\} = N(2N_{dg}-I) - N_{sq}$ where S_i and S_j are transient states.

(ii) Let t be the function giving the number of steps (including the original position) in which the process is in a transient state. Then $\left\{ M_i[t] \right\} = \tau$, $\left\{ Var_i[t] \right\} = (2N-I)\tau - \tau_{sq}$ where $\tau = N\xi$.

This theorem, which we stated without proof, shows that most of the important quantities can be expressed in terms of the fundamental matrix N.

For regular chains, there is a corresponding fundamental matrix which is of similar importance.

Theorem 3.10 Let P be the transition matrix for a regular Markov chain. Let A be the limiting matrix. Then $Z = (I - (P - A))^{-1}$ exists and

$$Z = I + \sum_{n=1}^{\infty} (P^n - A).$$

Proof: Since $A^2 = \xi\alpha\xi\alpha = \xi\alpha = A$, $A^k = A$ for any positive integer k.

$$(P-A)^n = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} P^i A^{n-i}$$

$$= P^n + \sum_{i=0}^{n-1} \binom{n}{i} (-1)^{n-i} A \qquad \text{(by Theorem 3.3(iii))}$$

$$= P^n - A.$$

Since $\lim_{n\to\infty} (P^n - A) = 0$, $\lim_{n\to\infty} (P-A)^n = 0$. By Theorem 2.1, $I - (P-A)$ has an inverse and $(I - (P-A))^{-1} = \sum_{n=0}^{\infty} (P-A)^n = I + \sum_{n=1}^{\infty} (P^n - A)$.

This completes the proof.

<u>Definition 3.4</u> Let P be a regular transition matrix. The matrix $Z = (I - (P-A))^{-1}$ is called the fundamental matrix for the Markov chain determined by P.

<u>Theorem 3.11</u> The mean first passage matrix M, which was introduced immediately following Definition 3.1, is given by $M = (I - Z + EZ_{dg})D$ where D is the diagonal matrix with diagonal elements $d_{ii} = 1/a_i$ and E is a square matrix with all entries 1.

<u>Theorem 3.12</u> Let f_j be the first passage time function of a regular chain. Then

$$\left\{Var_i\left[f_j\right]\right\} = M\,(2Z_{dg}D - I) + 2\,(ZM - E\,(ZM)_{dg}) - M_{sq}.$$

These two theorems, together with Theorems 3.7 and 3.8, indicate how fundamental the matrix Z is in finding basic formulas for regular chains.

CHAPTER IV

APPLICATION OF PRELIMINARY RESULTS

It is very surprising to see that we can achieve a tremendous extension of our preliminary results on finite Markov chains without the help of any additional theorems. All we have to do is to define a suitable new chain on the original one and analyze the new chain with the theorems at our disposal. The following theorem is a typical example.

Theorem 4.1 Let S be a subset of transient states in an absorbing chain with transition matrix P. Let Q be the s x s submatrix of P corresponding to these states. Let the process start in S_i. Then

(i) The ij component of $N = (I-Q)^{-1}$ is the mean number of times the process is in S_j before leaving S.

(ii) The ij component of $N_2 = N(2N_{dg} - I) - N_{sq}$ is the variance of the same function.

(iii) The i-th component of $N\xi$ is the mean number of steps needed to leave S.

(iv) The i-th component of $(2N - I)N\xi - (N\xi)_{sq}$ is the variance of the same function.

Proof: We form a new process in which all states not in S are made absorbing states. Since S is a subset of transient states in the original process, from each element of S we can reach a state not in S which must be an absorbing state in the new chain. Hence, the new chain is absorbing and the elements of S must all be transient states in the new process. By applying Theorem 3.9 to the new chain, we get all four results.

BIBLIOGRAPHY

1. Gantmacher, F. R., The Theory of Matrices, Vol. 2, Chelsea, New York, 1959.

2. Kemeny, J. G. and Snell, J. L., Finite Markov Chains. Van Nostrand, Princeton, N.J. 1959.