TWO CLASSIFICATION THEOREMS OF STATES OF DISCRETE MARKOV CHAINS

by

Gary Leslie Phillips

B.Sc., Simon Fraser University, 1967

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

November, 1970

APPROVAL

Name: Gary Leslie Phillips

Degree: Master of Science

Title of Thesis: Two Classification Theorems of States of Discrete Markov Chains

Examining Committee:

(A. H. Lachlan)

Chairman

(C. W. Kim)

Senior Supervisor

(H. Gerber)

Examining Committee

(J. J. Sember)

Examining Committee

(N. R. Reilly)

Examining Committee

Date approved: November 30, 1970

ABSTRACT

The purpose of this paper is to prove, from the point of view of ergodic theory, two classification theorems of states of discrete Markov chains.

To prove these theorems we first indicate several basic definitions and prove some elementary theorems from Markov chain theory. Following this we develop a basis in discrete ergodic theory, as developed by E. Hopf. We begin by indicating the connection between the Markov operators P and T and the related Markov chains. Next we define the conservative part C, and the dissipative part D, of the state space of a Markov chain, indicating that the sets C and D partition the state space. We finally introduce the family $\mathcal{C}$ of subsets $B \subset C$ for which $P1_B = 1_B$ on C, and prove that $\mathcal{C}$ is a σ-algebra with atoms $C_k = \{j: \Sigma_n p_{kj} = \infty\}$ where $k \in C$. We further indicate that this development differs from that of Kim [6] only by the choice of representation of the atoms of $\mathcal{C}$. This difference, however, appears to make the argument more clear.

Using this information we are able to prove the two classification theorems. First we prove that the state space of a discrete Markov chain can be partitioned into the set of all nonrecurrent (transient) states and recurrent classes. We then prove the corresponding theorem for idempotent Markov chains.

We conclude the paper by indicating how some idempotent Markov chains arise quite naturally as a result of taking various limits.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author would like to thank Dr. C. W. Kim of the Mathematics Department, Simon Fraser University for suggesting the topic and for his continuous aid and encouragement in the development of this thesis. I am also indebted to Simon Fraser University and to the National Research Council (grant # A 4844) for their financial aid which made this thesis possible. Finally, I would like to thank Miss Sue Smithe for her patient and accurate typing.

CHAPTER O - THE INTRODUCTION

The purpose of this paper is to examine some of the basic properties of Markov chains and the related stochastic matrices. Specifically we will prove, from the viewpoint of ergodic theory, two classification theorems of the states of Markov chains: the classification of the states into the set of all nonrecurrent (transient) states and recurrent classes (Feller [3], Theorem 3, p. 392), and the corresponding classification for an idempotent Markov chain (Doob [2], Theorem 2, p. 39). This argument employs a different method than that of Kim [6] and appears more transparent.

A basic knowledge of general probability theory is assumed in this paper. We begin with a review of the basic concepts of Markov chain theory following the outline of Chung [1] and Feller [3]. After defining the Markov property and indicating its importance we define the transition probabilities and transition matrices associated with Markov chains. From here we examine the properties of reducibility and recurrence of Markov chains. Then we shall use the weakly doubly stochastic matrices as defined by Révész [8] to expand the notion of a stochastic matrix.

In Chapter II we develop a discrete version of the ergodic theory due to E. Hopf and establish useful facts which are needed in Chapter III. We begin by indicating the relationship between Markov chains and Markov operators. We define the conservative part of C and dissipative part, D, of S, the state space of a discrete Markov chain. Then, we introduce the family $\mathcal{C}$ of subsets $B \subset C$ for which $P1_B = 1_B$ on C . We next show that $\mathcal{C}$ is a σ-algebra with atoms $C_k = \{j: \Sigma_n p^n_{kj} = \infty\}$ where $k \in c$.

In Chapter III we will actually prove these two classification theorems relying constantly on the information set down in Chapter II.

After noting the importance of idempotent Markov chains in the second classification theorem we devote Chapter IV to examining how these idempotent Markov chains arise. We exhibit the simplest example of an idempotent transition matrix, show how idempotent transition matrices arise quite naturally and conclude by showing how such idempotent matrices may be constructed.

CHAPTER I - DISCRETE PARAMETER MARKOV CHAINS

§1.1 The Markov Property

Following the approach of Chung [1], Feller [3], and Doob [2] we will outline the basic concepts of Markov chain theory.

We consider an abstract space Ω , called the probability space (or sample space), with the generic element ω , called the elementary event (or sample point), a Borel field $\mathcal{F}$ of subsets of Ω, called measurable sets or events, and a countably additive probability measure P defined on $\mathcal{F}$. We call the triple $(\Omega,\mathcal{F},P)$ a probability triple.

A discrete parameter stochastic process is a sequence of discrete random variables (measurable functions)

$$\{X_n: \ n \geq 0\}$$

defined with respect to some probability triple $(\Omega,\mathcal{F},P)$. If all the random variables X_n are discrete, the union of all possible values of all X_n is a denumerable set S called the (minimal) state space of the process. Each element of S is a state.

For the following we may assume without loss of generality that $S = \{1,2,3,\ldots\}$

<u>Definition 1.1</u> A <u>discrete time parameter Markov chain</u> is a sequence of discrete random variables $\{X_n: n \geq 0\}$ possessing the following property: for any integer $n \geq 2$, $0 \leq t_1 < \ldots < t_n$ and any states $i_1, i_2, \ldots i_n$ in the state space S we have

$$P\{X_{t_n}(\omega) = i_n : X_{t_1}(\omega) = i_1, \ldots, X_{t_{n-1}}(\omega) = i_{n-1}\}$$

$$= P\{X_{t_n}(\omega) = i_n : X_{t_{n-1}}(\omega) = i_{n-1}\}.$$

This property is referred to as the <u>Markov property</u>.

The Markov property is equivalent to the apparently weaker condition

$$P\{X_n(\omega) = i_n : X_0(\omega) = i_0, \ldots, X_{n-1}(\omega) = i_{n-1}\}$$

$$= P\{X_n(\omega) = i_n : X_{n-1}(\omega) = i_{n-1}\}.$$

An important consequence of the Markov property is that for any $n \geq 2$, $0 \leq t_1 < \ldots < t_n < \ldots < t_{n+m}$ and any $i_1, i_2, \ldots, i_{n+m}$ in S, we have

$$P\{X_{t_\nu}(\omega) = i_\nu,\ n \leq n+m : X_{t_\nu}(\omega) = i_\nu,\ 1 \leq \nu \leq n-1\}$$

$$= P\{X_{t_\nu}(\omega) = i_\nu,\ n \leq \nu \leq n+m : X_{t_{n-1}} = i_{n-1}\}$$

This result is obtained by induction on m. (Chung [1]).

One may therefore verbally describe the Markov property with the following statement: The probability of an event, knowing the states at several previous moments, is the same as the probability knowing only the last given state.

§1.2 Transition Probabilities and Transition Matrices

A discrete time parameter Markov chain $\{X_n\}$ is said to have stationary transition probabilities if

$$P\{X_n(\omega) = j: \ X_{n-1}(\omega) = i\} = p_{ij} \quad \text{for all} \quad n \geq 1 \quad \text{and}$$

all $i,j \in S$ for which this conditional probability is defined.

Unless otherwise specified, all Markov chains considered in this paper will be stationary (have stationary transition probabilities).

The probability p_{ij} is called the (one step) transition probability from i to j. The matrix (p_{ij}) $i,j \in S$ is called the (one step) transition matrix of the related Markov chain.

Definition 1.2 The distribution $\{p_i: \ i \in S\}$, $p_i = P\{X_0(\omega) = i\}$, where $i \in S$ is called the (initial) distribution of the Markov chain $\{X_n\}$.

The following relations are clearly satisfied by the initial and transition probabilities of a Markov chain.

$$p_i \geq 0 \quad \text{and} \quad \Sigma_i \, p_i = 1 \quad \text{for all} \quad i \in S$$

and $\quad p_{ij} \geq 0$ and $\Sigma_j p_{ij} = 1$ for all $i,j \in S$.

We now state without proof two fundamental theorems.

Theorem 1.1 (Kolmogorov) (Chung [1], Theorem 1, p. 7). Given $\{p_i\}_{1 \leq i < \infty}$ and $\{p_{ij}: \ i,j = 1,2,\ldots\}$ such that

$$p_i \geq 0 \text{ and } \Sigma_i p_i = 1 \quad \text{for all } i \in S = \{1,2,\ldots\}$$

and $\quad p_{ij} \geq 0$ and $\Sigma_j p_{ij} = 1$ for all $i \in S = \{1,2,\ldots\}$.

Then there exists a probability space (Ω,F,P) and a Markov chain $\{X_n\}$ with state space S , initial distribution $\{p_i\}$ and transition matrix (p_{ij}) .

We call $P\{X_{n+k}(\omega) = j: \ X_k(\omega) = i\} = p_{ij}^{(n)}$, for $n \geq 1,\ k \geq 0$ the n-step transition probability from i to j . The matrix $(p_{ij}^{(n)}) = (p_{ij})^n$ is similarly called the n-step transition matrix. In the future we will denote $p_{ij}^{(n)}$ by p_{ij}^n . Thus we state the other theorem.

Theorem 1.2 (Chapman - Kolmogorov Equation)

$$p_{ij}^{m+n} = \Sigma_{k \in S}\, p_{ik}^m\, p_{kj}^n \quad \text{for all } m,n \geq 0 \ .$$

§1.3 Classifications

We classify the states of a state space S with regard to their basic transition properties. We say i leads to j , $i \to j$, if there exists a positive integer m such that $p_{ij}^{m} > 0$. The states i and j are said to *communicate* if $i \to j$ and $j \to i$, then we write $i \leftrightarrow j$. It is obvious that the relation "$\leftrightarrow$" is symmetric and transitive. It is also clear that if $i \leftrightarrow j$ for some j then $i \leftrightarrow i$, however, it is not true in general that $i \leftrightarrow i$. Thus the relation "$\leftrightarrow$" partitions the state space into subsets (equivalence classes) as follows.

Definition 1.3 A *class* *C(i)* is a subset of the state space and

a) Consists of all states mutually communicating with i

or b) Consists of only i if i does not communicate with any state.

A property defined for all states is called a *class property* if its possession by one state implies its possession by all states in that class.

A state which communicates with every state it leads to is called an *essential state*. Other states are called *inessential states*. We may now state the following theorem.

Theorem 1.3 An essential state cannot lead to an inessential state.

Proof. Let i be an essential state and j an inessential state. Suppose $i \to j$. Since j is an inessential state there is some state, say k, such that $j \to k$ but j does not communicate with k. However, since $i \to j$ and $j \to k$ then $i \to k$ by transitivity.

Moreover i is an essential state so $i \leftrightarrow k$ so $k \rightarrow i$. By transitivity it follows that $k \rightarrow j$; the contradiction. □

<u>Corollary 1.1</u> The property of being essential (or inessential) is a class property. We call a class essential or inessential according as its states are essential or inessential.

<u>Proof</u>. Since all states in a class mutually communicate and no essential state can lead to an inessential state, the states of a class must all be essential or all be inessential. □

If $i \rightarrow i$, then the greatest common divisor of $\{n: p_{ii}^{n} > 0\}$ is called the <u>period</u> of i (denoted d_i). If i does not lead to itself then we do not define the period of i. On closely examining the periods of states the following theorem is suggested.

<u>Theorem 1.4</u> The property of having period equal to d is a class property.

<u>Proof</u>. Suppose i and j are in the same class, then $i \leftrightarrow j$. Thus there are positive integers m and n such that $p_{ij}^{m} > 0$ and $p_{ji}^{n} > 0$. Now if for some integer $s > 0$, $p_{ii}^{s} > 0$ then $p_{jj}^{n+s+m} \geq p_{ji}^{n} p_{ii}^{s} p_{ij}^{m} > 0$. We also have $p_{ii}^{2s} > 0$ so $p_{jj}^{n+2s+m} > 0$. Thus the period of j (d_j) divides both $(n+s+m)$ and $(n+2s+m)$ so it also divides the difference, $[(n+2s+m) - (n+s+m)] = s$.

Therefore for all s such that $p_{ii}^{s} > 0$, d_j divides s. We conclude that d_j divides d_i. Exchanging roles of i and j we see that d_i divides d_j. Hence for all i and j in the same class $d_i = d_j$ which completes the proof. □

Theorem 1.5 (Chung [1], p. 14) If $d_i > 0$ then there exists an integer N_i such that $p_{ii}^{nd_i} > 0$ for $n \geq N_i$.

Proof. By the definition of period there are finitely many positive integers n_s, $1 \leq s \leq t$ such that $p_{ii}^{n_s} > 0$ and such that d_i is their greatest common divisor. From elementary number theory we see that there exists an N such that $n \geq N$ implies the existence of positive integers c_s, $1 \leq s \leq t$ satisfying

$$nd_i = \Sigma_{s=1}^{t} c_s n_s .$$

It follows then that if $n \geq N$

$$p_{ii}^{nd_i} = p_{ii}^{\Sigma_{s=1}^{t} c_s n_s} \geq \prod_{s=1}^{t} (p^{n_s})^{c_s} > 0 .$$

Thus let $N_i = N$ and the proof is complete. □

This leads immediately to the following corollary.

Corollary 1.2 If $p_{ij}^{m} > 0$, then for sufficiently large n we have

$$p_{ij}^{nd_i+m} > 0$$

Proof. By Theorem 1.5 we have $p_{ii}^{nd_i} > 0$ for sufficiently large n. It follows that, for sufficiently large n,

$$p_{ij}^{nd_i+m} \geq p_{ii}^{nd_i} \cdot p_{ij}^{m} > 0 \qquad \square$$

If for some state i, $d_i = 1$, then i is called an aperiodic state.

§1.4 Reducibility

We say that a set A, of states, is (stochastically) closed if we have

$$\Sigma_{j \varepsilon A}\, p_{ij} = 1 \quad \text{for every} \quad i \varepsilon A$$

Clearly then for all n, $S_{j \varepsilon A}\, p_{ij}^{n} = 1$ if A is closed. A set is minimal closed if it is closed and has no proper, closed subset.

Theorem 1.6 A set of states is minimal closed if and only if it is an essential class.

Proof. Since an essential state leads only to states with which it communicates, it is obvious that an essential class is minimal closed. Thus any closed set which contains an essential state must consist of the corresponding essential class. Now let a closed set C consist of only inessential states. Let i be a state of C and

$A = \{j \in C: i \to j$ and $j \to i\}$. Then $A - \{i\}$ is nonempty, closed and a proper subset of C . Therefore C is not minimal closed. □

<u>Definition 1.4</u> A Markov chain with a state space which is minimal closed (i.e. consisting of one essential class) is called <u>irreducible</u>.

A transition matrix $P = (p_{ij})$ is irreducible if the corresponding Markov chain is irreducible.

<u>Theorem 1.7</u> Let $P = (p_{ij})$ be an irreducible transition matrix then i) For each i , $\exists\, j_i$ such that $p_{ij_i} > 0$

ii) For each j , $\exists\, i_j$ such that $p_{i_j j} > 0$

<u>Proof.</u> i) Since for each $i \in S$, $\Sigma_j\, p_{ij} = 1$ we have $p_{ij_i} > 0$ for some $j_i \in S$. Thus the assertion holds for all transition matrices

ii) Suppose the assertion fails. That is

$$p_{kj} = 0 \quad \text{for all} \quad k .$$

By the Chapman - Kolmogorov equation, we have

$$p_{kj}^{h+1} = \Sigma_i\, p_{ki}^{n}\, p_{ij} = 0 \quad \text{for all} \quad k \text{ , and all} \quad n \geq 1 .$$

Thus $p_{kj}^{n} = 0$ for all k and all n . However, if we denote the period by d , then $d \geq 1$. By Theorem 1.5 we have

$$p_{ij}^{nd} > 0 \quad \text{for sufficiently large} \quad n .$$

Therefore the supposition leads to a contradiction, completing the proof. □

We notice that conditions i) and ii) of the previous theorem are not sufficient for P being irreducible. For example

$$\text{Let } P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

P satisfies conditions i) and ii) but is clearly not irreducible.

A matrix (p_{ij}) is said to be <u>doubly stochastic</u> if $\Sigma_i\, p_{ik} = 1$ for all k and $\Sigma_j\, p_{hj} = 1$ for all h. We notice that every finite doubly stochastic matrix satisfies conditions i) and ii). However, not all finite doubly stochastic matrices are irreducible as shown by the example above.

<u>Theorem</u> 1.8 Let $P = (p_{ij})$ be an irreducible matrix. Then

i) $\forall n, \forall i, \exists j_i$ such that $p^n_{ij_i} > 0$

ii) $\forall n, \forall j, \exists i_j$ such that $p^n_{i_j j} > 0$

<u>Proof</u>. By Theorem 1.7 it is clear that i) is true for $n = 1$. Now suppose it is true for $n = h \geq 1$. Then $\forall k, \exists j_k$ such that $p^h_{kj_k} > 0$. It follows from Theorem 1.7 that $\exists l$ such that $p_{j_k l} > 0$.

Using the Chapman - Kolmogorov equation, we have

$$p_{kl}^{h+1} = \Sigma_i \, p_{ki}^{h} \cdot p_{il} \geq p_{kj_k}^{h} \cdot p_{j_k l} > 0$$

Thus (i) is proven by induction on n .

(ii) is proven similarly. □

§1.5 Recurrence

Henceforth we shall let the event $\{\omega: \; X_n(\omega) = i\}$ denote the statement that the Markov chain $\{X_n\}$ is in the state i at the time n (or at the n^{th} step). Similarly the statement "The probability that the Markov chain will be in state j for the first time at the n^{th} step, given that it starts from i " will be denoted by

$$P\{X_{m+\nu}(\omega) \neq j, \; 0 < \nu < n, \; X_{m+n}(\omega) = j: \; X_m(\omega) = i\}$$

where $m \geq 0$ and $P\{X_m(\omega) = i\} > 0$. We note that, due to the Markov property, there is no dependence on m , whenever it is defined. Following this,

$$P\{X_{m+n}(\omega) = i \quad \text{for some} \quad n > 0: \; X_m(\omega) = i\} = 1$$

if it is defined, denotes the statement "if $X_0(\omega) = i$ then $X_n(\omega) = i$ for some $n > 0$".

We write

$$f_{ij}^{n} = P\{X_\nu(\omega) \neq j, \; 0 < \nu < n; \; X_n(\omega) = j: \; X_0(\omega) = i\}$$

and $f^*_{ij} = \Sigma_{n=1}^{\infty} f^n_{ij}$.

We note that f^*_{ij} is the probability that the Markov chain $\{X_n\}$ will be in state j at least once, given that it starts from state i . Therefore we can rewrite it as

$$f^*_{ij} = P\{X_n(\omega) = j \text{ for some } n > 0: X_0(\omega) = i\}$$

$$= P\{\cup_{n=1}^{\infty}[\omega: X_n(\omega) = j]: X_0(\omega) = i\}$$

We also use the symbol g_{ij} to denote the probability that the Markov chain will be in state j infinitely often (i.o.) given that it started from state i . Then we have

$$g_{ij} = P\{X_n(\omega) = j \text{ i.o.}: X_0(\omega) = i\}$$

$$= \lim_{m\to\infty} P\{X_n(\omega) = j \text{ for some } n \geq m: X_0(\omega) = i\}$$

$$= \lim_{m\to\infty} \Sigma_k p^m_{ik} f^*_{kj} .$$

Suppose $i \to j$, then $p^n_{ij} > 0$ for some $n \geq 1$. Thus $f^*_{ij} \geq p^n_{ij} > 0$. Similarly if $f^*_{ij} > 0$ then for some $n \geq 1, p^n_{ij} > 0$ thus $i \to j$. Hence $i \to j$ if and only if $f^*_{ij} > 0$. It follows directly that $i \leftrightarrow j$ if and only if $f^*_{ij} \cdot f^*_{ji} > 0$.

We can now prove the following theorem.

<u>Theorem 1.9</u> $g_{ij} = f^*_{ij}\, g_{jj}$

Proof. Certainly

$$g_{ij} = \Sigma_{n=1}^{\infty} P\{X_\nu(\omega) \neq j,\ 0 < \nu < n\ ;\ X_n(\omega) = j;\ X_s(\omega) = j \text{ i.o. for } s > n:\ X_0(\omega) = i\}$$

$$= \Sigma_{n=1}^{\infty} P\{X_\nu(\omega) \neq j,\ 0 < \nu < n\ ;\ X_n(\omega) = j:\ X_0(\omega) = i\} \cdot P\{X_s(\omega) = j \text{ i.o. for } s > n:\ X_n(\omega) = j\}$$

$$= \Sigma_{n=1}^{\infty} f_{ij}^{n}\, g_{jj} = f_{ij}^{*}\, g_{jj} \quad \square$$

Definition 1.5 A state i is called recurrent or transient according as

$$f_{ii}^{*} = 1 \text{ or } f_{ii}^{*} < 1 .$$

This gives us the following theorem.

Theorem 1.10 $g_{ij} = f_{ij}^{*}$ or 0 according as j is recurrent or transient.

Proof. For $m \geq 1$ let

$$g_{ij}(m) = P\{X_n(\omega) = j \text{ for at least } m \text{ values of } n > 0:\ X_0(\omega) = i\}.$$

Then $$g_{ij}(m+1) = f_{ij}^{*}\, g_{jj}(m)$$

and $$g_{ij}(1) = f_{ij}^{*}$$

By induction on m we have

$$g_{ij}(m+1) = f^*_{ij} \; (f^*_{jj})^m .$$

Now taking the limit as $m \to \infty$ gives

$$g_{ij} = \begin{cases} f^*_{ij} & \text{if} \quad f^*_{jj} = 1 \\ 0 & \text{if} \quad f^*_{jj} < 1 \end{cases}$$

which completes the proof. □

We notice that for any states i and j and $n \geq 1$ we have

$$p^n_{ij} = \Sigma^n_{\nu=1} \; f^\nu_{ij} \; p^{n-\nu}_{jj} \tag{1.1}$$

by the definition of p^n_{ij} , f^ν_{ij} and $p^{n-\nu}_{jj}$.

We may now expand the notion of recurrence as we state the following theorem.

<u>Theorem 1.11</u> The following are equivalent.

i) $f^*_{ii} = 1$ (i is a recurrent state)

ii) $\Sigma^\infty_{n=1} \; p^n_{ii} = \infty$

iii) $g_{ii} = 1$

<u>Proof</u>. It is clear from Theorem 1.10 that $f^*_{ii} = 1$ if and only if $g_{ii} = 1$. So (i) is equivalent to (iii).

Suppose that $f_{ii}^{*} = 1$, then for all $\varepsilon > 0$ there is some N such that for $k \geq N$

$$\Sigma_{n=1}^{k} f_{ii}^{n} > 1-\varepsilon .$$

Thus for $k \geq N$ and using equation (1.1) we have the following.

$$\Sigma_{n=1}^{2k} p_{ii}^{n} = \Sigma_{n=1}^{2k}[\Sigma_{\nu=1}^{n} f_{ii}^{\nu} p_{ii}^{n-\nu}] = \Sigma_{n=1}^{2k}[\Sigma_{\nu=0}^{n-1} f_{ii}^{n-\nu} p_{ii}^{\nu}]$$

$$= \Sigma_{\nu=0}^{2k-1} p_{ii}^{\nu} \Sigma_{n=\nu+1}^{2k} f_{ii}^{n-\nu} = \Sigma_{\nu=0}^{2k-1} p_{ii}^{\nu} \Sigma_{n=1}^{2k-\nu} f_{ii}^{n}$$

$$\geq (\Sigma_{\nu=0}^{k} p_{ii}^{\nu})(\Sigma_{n=0}^{k} f_{ii}^{n}) > (\Sigma_{\nu=0}^{k} p_{ii}^{\nu})(1-\varepsilon) .$$

Thus for $k \geq N$ we have

$$\frac{\Sigma_{n=1}^{2k} p_{ii}^{n}}{\Sigma_{\nu=0}^{k} p_{ii}^{\nu}} > 1-\varepsilon$$

Hence
$$\lim_{k\to\infty} \frac{\Sigma_{n=1}^{2k} p_{ii}^{n}}{\Sigma_{n=0}^{k} p_{ii}^{n}} \geq 1 .$$

Now suppose $\lim_{k\to\infty} \Sigma_{n=1}^{k} p_{ii}^{n} = K < \infty$ then $\lim_{k\to\infty} \Sigma_{n=0}^{k} p_{ii}^{n} = K + p_{ii}^{0}$.

Since $p_{ii}^{0} > 0$

$$1 \leq \lim_{k\to\infty} \frac{\Sigma_{n=1}^{2k} p_{ii}^{n}}{\Sigma_{n=0}^{k} p_{ii}^{n}} = \frac{\lim_{k\to\infty} \Sigma_{n=1}^{2k} p_{ii}^{n}}{\lim_{k\to\infty} \Sigma_{n=0}^{k} p_{ii}^{n}} < 1$$

a contradiction. Therefore $\lim_{k\to\infty} \Sigma_{n=1}^{k} p_{ii}^{n} = \Sigma_{n=1}^{\infty} p_{ii}^{n} = \infty$.

On the other hand, if $\Sigma_{n=1}^{\infty} p_{ii}^{n} = \infty$ then given $K > 0$ there exists $N > 0$ such that for $m \geq N$, $\Sigma_{n=1}^{m} p_{ii}^{n} > K$. Using equation (1.1) we have

$$\Sigma_{n=1}^{m} p_{ii}^{n} = \Sigma_{n=0}^{m-1} p_{ii}^{n} \Sigma_{\nu=1}^{m-n} f_{ii}^{\nu} \leq \Sigma_{n=0}^{m-1} p_{ii}^{n} (\Sigma_{\nu=0}^{m} f_{ii}^{\nu}) .$$

Hence

$$\frac{\Sigma_{n=1}^{m} p_{ii}^{n}}{\Sigma_{n=0}^{m-1} p_{ii}^{n}} \leq \Sigma_{\nu=0}^{m} f_{ii}^{\nu} .$$

But
$$\Sigma_{n=1}^{m} p_{ii}^{n} < \Sigma_{n=0}^{m-1} p_{ii}^{n} < \Sigma_{n=1}^{m} p_{ii}^{n} + 1 .$$

Hence

$$\frac{\Sigma_{n=1}^{m} p_{ii}^{n}}{\Sigma_{n=0}^{m-1} p_{ii}^{n}} > \frac{\Sigma_{n=1}^{m} p_{ii}^{n}}{\Sigma_{n=1}^{m} p_{ii}^{n} + 1} > \frac{K}{K+1} = 1 - \frac{1}{K+1} .$$

So given $\varepsilon > 0$ find $K > 0$ such that $\frac{1}{K+1} \leq \varepsilon$, choose N as before. Then for $m \geq N$

$$1 \geq \Sigma_{\nu=0}^{m} f_{ii}^{\nu} > 1 - \frac{1}{K+1} \geq 1 - \varepsilon .$$

Thus
$$\lim_{m\to\infty} \Sigma_{\nu=0}^{m} f_{ii}^{\nu} = f_{ii}^{*} = 1 .$$

So $\Sigma_{n=1}^{\infty} p_{ii}^{n} = \infty$ is equivalent to $f_{ii}^{*} = 1$. □

We shall now redefine recurrence in the form which we shall use for the remainder of this paper.

<u>Definition 1.6</u> A state i is said to be <u>recurrent</u> or <u>transient</u> according as

$$\Sigma_{n=1}^{\infty} p_{ii}^{n} = \infty \quad \text{or} \quad \Sigma_{n=1}^{\infty} p_{ii}^{n} < \infty .$$

If i is a recurrent state and $i \leftrightarrow j$ (i.e. $j \in C(i)$) then for some $k > 0$ and $h > 0$, $p_{ij}^{k} > 0$ and $p_{ji}^{h} > 0$.

Thus $$\Sigma_{n=0}^{\infty} p_{jj}^{n} \geq \Sigma_{n=k+h}^{\infty} p_{jj}^{n} \geq \Sigma_{n=0}^{\infty} p_{ji}^{h} p_{ii}^{n} p_{ij}^{k} = \infty .$$

On the other hand, if $j \in C(i)$ and j is recurrent then clearly $i \in C(j) = C(i)$ so as above i is recurrent. Hence we have the following.

<u>Corollary 1.3</u> Recurrence is a class property.

We may also take this opportunity to point out that a recurrent state i may further be identified as either <u>positive recurrent</u> or <u>null recurrent</u> according as

$$\lim_{n\to\infty} p_{ii}^{nd_i} > 0 \quad \text{or} \quad = 0 .$$

Note that a positive recurrent state i with $d_i = 1$ (aperiodic) is called an <u>ergodic state</u>.

<u>Theorem 1.12</u> The property of a state being positive (or null) recurrent is a class property.

<u>Proof</u>. Suppose i is positive recurrent and $i \leftrightarrow j$. Let $d = d_i = d_j$. Set m and n such that

$$p_{ij}^{m} > 0 \quad \text{and} \quad p_{ji}^{n} > 0 \, .$$

Since
$$p_{jj}^{m+\nu d+n} \geq p_{ji}^{n} \, p_{ii}^{\nu d} \, p_{ij}^{m} \, ,$$

$$\lim_{\nu\to\infty} p_{jj}^{\nu d} = \lim_{\nu\to\infty} p_{jj}^{m+\nu d+n} \geq p_{ji}^{n} \, p_{ij}^{m} \lim_{\nu\to\infty} p_{ii}^{\nu d} = \infty . \quad \square$$

<u>§1.6 Stochastic (Markov) Matrices</u>

Matrix $A = (a_{ij})$ is a non-negative matrix if all entries a_{ij} are non-negative. All matrices considered throughout this paper will be non-negative.

An $n \times m$ matrix $A = (a_{ij})$ is <u>row stochastic</u> if $\Sigma_{j=1}^{m} a_{ij} = 1$ for all $i = 1,2,\ldots,n$. Similarly A is <u>column stochastic</u> if $\Sigma_{i=1}^{n} a_{ij} = 1$ for all $j = 1,2,$ [redacted] . A matrix which is both row stochastic and column stochastic is called a <u>doubly stochastic</u> matrix.

The following definition is due to Révész [8].

<u>Definition 1.7</u> An $n \times m$ matrix $A = (a_{ij})$ is said to be <u>row (column) weakly doubly stochastic</u> if for all i,j

$$\Sigma_{j=1}^{m} a_{ij} = 1 \ (\le 1)$$

and $$\Sigma_{i=1}^{n} a_{ij} \le 1 \ (= 1) .$$

Therefore an $n \times m$ matrix is doubly stochastic if it is both row weakly doubly stochastic and column weakly doubly stochastic.

The following theorems now become clear.

<u>Theorem 1.13</u> Every $n \times n$ row (column) weakly doubly stochastic matrix $A = (a_{ij})$ is doubly stochastic.

<u>Proof</u>. Suppose A is row weakly doubly stochastic, then

$$\Sigma_{1\le i,j\le n} a_{ij} = \Sigma_{i=1}^{n} (\Sigma_{j=1}^{n} a_{ij}) = \Sigma_{i=1}^{n} (1) = n .$$

Also $$\Sigma_{i=1}^{n} a_{ij} \le 1 \quad \text{for all} \quad j = 1,2,\ldots,n .$$

If for some $k = 1,2,\ldots,n$, $\Sigma_{i=1}^{n} a_{ik} < 1$,

then $\Sigma_{1\le i,j\le n} a_{ij} = \Sigma_{j=1}^{n} (\Sigma_{i=1}^{n} a_{ij}) \le \Sigma_{i=1}^{n} a_{ik} + \Sigma_{j\ne k} (1) < 1 + (n-1) = n .$

Thus no such k exists and the matrix is doubly stochastic. Exchanging the roles of i and j we can similarly prove the theorem for a column weakly doubly stochastic matrix. □

<u>Theorem 1.14</u> If $A = (a_{ij})$ is an $m \times n$ row (column) weakly doubly stochastic matrix, then $m \le n$ $(n \le m)$.

Proof. Let A be a row weakly doubly stochastic matrix. Then $\Sigma_{j=1}^{n} a_{ij} = 1$ and $\Sigma_{i=1}^{m} a_{ij} \le 1$ for all i,j.

Thus $$\Sigma_{i,j}\, a_{ij} = \Sigma_{i=1}^{m} \Sigma_{j=1}^{n} a_{ij} = \Sigma_{i=1}^{m} 1 = m .$$

Hence $$m = \Sigma_{i,j}\, a_{ij} = \Sigma_{j=1}^{n} \Sigma_{i=1}^{m} a_{ij} \le \Sigma_{j=1}^{n} 1 = n$$

so $m \le n$. Once again by interchanging the roles of i and j we can similarly prove the theorem for a column weakly doubly stochastic matrix. □

It is now obvious that if matrix $A = (a_{ij})$ is an $m \times n$ doubly stochastic matrix then $m = n$. We shall prove the following stronger result.

Theorem 1.15 If A is an $m \times n$ row (column) weakly doubly stochastic matrix, then A is doubly stochastic if and only if $m = n$.

Proof. (only if) This is an obvious consequence of Theorem 1.12 and the comment that A is doubly stochastic if and only if A is row and column weakly doubly stochastic.

(if) Suppose A is a $m \times m$ row weakly doubly stochastic matrix and is not doubly stochastic. Then $\Sigma_{i=1}^{m} a_{ij} \le 1$ and $\Sigma_{j=1}^{m} a_{ij} = 1$ for all i,j, and for some $k = 1,2,\ldots,n$,

$$\Sigma_{i=1}^{m} a_{ik} < 1 .$$

Therefore we have

$$\Sigma_{i,j}\, a_{ij} = \Sigma_{i=1}^{m} \Sigma_{j=1}^{m} a_{ij} = \Sigma_{i=1}^{m} (1) = m$$

so $m = \Sigma_{i,j} a_{ij} = \Sigma_{j=1}^{m} \Sigma_{i=1}^{m} a_{ij} \le \Sigma_{i=1}^{m} a_{ik} + \Sigma_{j\ne k}(1) < m$, a contradiction. Interchanging the roles of i and j we use the same proof for column weakly doubly stochastic matrices. □

We also note the following result.

<u>Theorem 1.16</u> Let $P = (p_{ij})$ be an $n \times n$ doubly stochastic matrix. Then P^{ν} is an $n \times n$ doubly stochastic matrix for all $\nu = 1,2,\ldots$.

<u>Proof</u>. Let (p_{ij}^{ν}) represent P^{ν} for $\nu = 1,2,\ldots$. Since P is a non-negative doubly stochastic matrix,

$$0 \le p_{ij} \le 1 \quad \text{for} \quad i,j = 1,2,\ldots,n .$$

Suppose for some $m \ge 1$

$$0 \le p_{ij}^{m} \le 1$$

then

$$p_{ij}^{m+1} = \Sigma_{k=1}^{n} p_{kj}^{m} p_{ik} \quad \text{for} \quad m = 1,2,\ldots;\ i,j = 1,2,\ldots,n .$$

Thus

$$0 \le \min_{k} (p_{kj}) \le p_{ij}^{m+1} \le \max_{k} (p_{kj}^{m}) \le 1 .$$

Hence

$$0 \le p_{ij}^{\nu} \le 1 \quad \text{for} \quad \nu = 1,2,\ldots .$$

Now suppose for some m, P^m is doubly stochastic, then since P was doubly stochastic we have

$$\Sigma_{j=1}^{n} p_{ij}^{(m+1)} = \Sigma_{j=1}^{n} (\Sigma_{k=1}^{n} p_{kj}^{m} p_{ik}) = \Sigma_{k=1}^{n} \Sigma_{j=1}^{n} p_{kj}^{m} p_{ik} = \Sigma_{k=1}^{n} 1 \cdot p_{ik} = 1$$

and

$$\Sigma_{i=1}^{n} p_{ij}^{(m+1)} = \Sigma_{i=1}^{n} (\Sigma_{k=1}^{n} p_{ik}^{m} p_{kj}) = \Sigma_{k=1}^{n} \Sigma_{j=1}^{n} p_{ik}^{m} p_{kj} = \Sigma_{k=1}^{n} 1 \cdot p_{kj} = 1 .$$

Therefore P^{ν} is doubly stochastic for all $\nu = 1,2,\ldots$. □

CHAPTER II - ELEMENTS OF ERGODIC THEORY

In this chapter we will develop a discrete version of the ergodic theory due to E. Hopf and establish useful facts which are needed in Chapter III.

§2.1 Markov Chains and Markov Operators

In this section we shall indicate the relationship between Markov chains and Markov operators.

Let S be the set of all positive integers and let $\Pi = \{\Pi(i) = \Pi_i\}$ be a strictly positive probability measure on S. Let $\{X_n : n \geq 0\}$ be a Markov chain with the state space S, the initial distribution Π, and the transition matrix (p_{ij}). This notation will remain fixed throughout this chapter.

By a <u>Markov operator</u> T on $L_1(S,\Pi)$ we mean a positive linear operator from $L_1(S,\Pi)$ into itself which preserves the integral

$$\int_S Tf \, d\Pi = \int_S f \, d\Pi \ , \quad \text{for} \quad f \in L_1(S,\Pi) \ .$$

That is,

$$\Sigma_{i\in S} Tf(i)\cdot\Pi(i) = \Sigma_{i\in S} f(i)\cdot\Pi(i) \ , \quad \text{for} \quad f \in L_1(S,\Pi) \ .$$

We write L_1 for $L_1(S,\Pi)$ and L_∞ for $L_\infty(S,\Pi)$. It is evident that L_∞ is identified with l_∞.

For each $A \subset S$, 1_A denotes the indicator function of the set A.

<u>Theorem 2.1</u> For each Markov operator T on L_1 there exists a unique positive linear operator P from L_∞ into itself such that

i) $\Sigma_i Tf(i) \cdot g(i) \cdot \Pi(i) = \Sigma_i f(i) \cdot Pg(i) \cdot \Pi(i)$ for $f \in L_1$ and

$g \in L_\infty$

ii) $P1 = 1$.

<u>Proof</u>. Fix $g \in L_\infty$. Then

$$F(f) = \Sigma_i Tf(i) \cdot g(i) \cdot \Pi(i)$$

is a linear functional on L_1. In view of the Riesz Representation Theorem there is a unique $\tilde{g}$ in L_∞ such that

$$F(f) = \Sigma_i Tf(i) \cdot g(i) \cdot \Pi(i) = \Sigma_i f(i) \cdot \tilde{g}(i) \cdot \Pi(i) \qquad (2.1)$$

for all $f \in L_1$

Thus for each $g \in L_\infty$ we have a unique $\tilde{g} \in L_\infty$. Let $P_g = \tilde{g}$. Since T is a positive linear operator it follows that P must be positive. Furthermore we can show that P is linear.

Let $f = 1_{\{k\}} \in L_1$ for some $k \in S$ and let $g = g_1 + g_2$ for some $g_1, g_2 \in L_\infty$. Then

$$Pg(k) = \frac{1}{\Pi(k)} \Sigma_i \; Tf(i) \cdot g(i) \cdot \Pi(i)$$

$$= \frac{1}{\Pi(k)} (\Sigma_i \; Tf(i) \cdot g_1(i) \cdot \Pi(i) + \Sigma_i \; Tf(i) \cdot g_2(i) \cdot \Pi(i))$$

$$= \frac{1}{\Pi(k)} (\Sigma_i \; f(i) \cdot Pg_1(i) \cdot \Pi(i) + \Sigma_i \; f(i) \cdot Pg_2(i) \cdot \Pi(i))$$

$$= \frac{1}{\Pi(k)} (Pg_1(k) \cdot \Pi(k) + Pg_2(k) \cdot \Pi(k))$$

$$= Pg_1(k) + Pg_2(k) \; .$$

Letting k run through S we have

$$Pg = P(g_1+g_2) = Pg_1 + Pg_2 \; .$$

Similarly for any real number a, let $g = ag^* \in L_\infty$. Then setting $f = 1_{\{k\}}$ for some $k \in S$ we have

$$Pg(k) = \frac{1}{\Pi(k)} \Sigma_i \; Tf(i) \cdot g(i) \cdot \Pi(i)$$

$$= \frac{a}{\Pi(k)} \Sigma_i \; Tf(i) \cdot g^*(i) \cdot \Pi(i)$$

$$= \frac{a}{\Pi(k)} \Sigma_i \; f(i) \cdot Pg^*(i) \cdot \Pi(i)$$

$$= \frac{a}{\Pi(k)} (Pg^*(k) \cdot \Pi(k))$$

$$= a \; Pg^*(k) \; .$$

Thus $Pg = Pag^* = a \; Pg^*$ on S. So P is a linear operator.

P is also unique since suppose there exists another positive linear operator P' from L_∞ into itself such that

$$\Sigma_i \, Tf(i) \cdot g(i) \cdot \Pi(i) = \Sigma_i \, f(i) \cdot P'g(i) \cdot \Pi(i)$$

for all $f \in L_1$ and $g \in L_\infty$.

Then by fixing $g \in L_\infty$ we have the linear functional F on L_1 such that

$$F(f) = \Sigma_i \, Tf(i) \cdot g(i) \cdot \Pi(i) = \Sigma_i \, f(i) \cdot P'g(i) \cdot \Pi(i)$$

for all $f \in L_1$.

Since $\tilde{g}$ in equation (2.1) is unique

$$P'g = \tilde{g} = Pg \; .$$

Hence P is unique.

Now we need only show that $P1 = 1$. Let $f = 1_{\{k\}}$ for any $k \in S$. Then

$$P1_{\{k\}} = \frac{1}{\Pi(k)} \Sigma_i \, Tf(i) \cdot 1 \cdot \Pi(i) = \frac{1}{\Pi(k)} \int_S Tf \, d\Pi \; .$$

But from the definition of a Markov operator T , for all $f \in L_1$

$$\int_S Tf \, d\Pi = \int_S f \, d\Pi \; .$$

Thus

$$P1_{\{k\}} = \frac{1}{\Pi(k)} \int_S Tf \, d\Pi = \frac{1}{\Pi(k)} \int_S f \, d\Pi$$

$$= \frac{1}{\Pi(k)} \Sigma_i \; f(i) \cdot \Pi(i) = 1 \; . \quad \square$$

<u>Definition 2.1</u> The linear operator P in Theorem 2.1 is called the <u>adjoint operator</u> of T. We call P a Markov operator on L_∞.

<u>Corollary 2.1</u> For all $k = 0,1,2,\ldots$; $f \in L_1$ and $g \in L_\infty$

$$\Sigma_{i \in S} T^k f(i) \cdot g(i) \cdot \Pi(i) = \Sigma_{i \in S} f(i) \cdot P^k g(i) \cdot \Pi(i) \; ,$$

where P is the adjoint operator of T.

<u>Proof.</u> For $k = 0$ the result is trivial. The result for $k = 1$ was proved in Theorem 2.1. Suppose the result is true for some $N \geq 0$. Then

$$\Sigma_{i \in S} T^{N+1} f(i) \cdot g(i) \cdot \Pi(i) = \Sigma_{i \in S} T(T^N f)(i) \cdot g(i) \cdot \Pi(i)$$

$$= \Sigma_{i \in S} T^N f(i) \cdot Pg(i) \cdot \Pi(i)$$

$$= \Sigma_{i \in S} f(i) \cdot P^N (Pg)(i) \cdot \Pi(i)$$

$$= \Sigma_{i \in S} f(i) \cdot P^{N+1} g(i) \cdot \Pi(i) \; .$$

Thus we have the desired result by induction on k. $\square$

<u>Theorem 2.2</u> For each Markov operator T on L_1 there is a unique stochastic matrix (p_{ij}) such that

$$Tf(j) \cdot \Pi(j) = \Sigma_i \ f(i) \cdot \Pi(i) \cdot p_{ij} \quad \text{for all} \quad f \in L_1 \ . \tag{2.2}$$

Conversely, each stochastic matrix (p_{ij}) defines a unique Markov operator T on L_1 satisfying (2.2).

Proof. For each Markov operator T and for each $j \in S$ we define $F_j(f) = \Sigma_{i\in S} \ Tf(i) \cdot 1_{\{j\}} \cdot \Pi(i) = Tf(j) \cdot \Pi(j)$ for $f \in L_1$. It is clear that F_j is a positive linear functional on L_1 . We also note that

$$|F_j(f)| = |Tf(j) \cdot \Pi(j)| \le \Sigma_i \ |f(i)| \cdot \Pi(i) = ||f||_1$$

$$\text{for} \quad f \in L_1 \ .$$

Thus $||F|| \le 1$.

From the Riesz Representation Theorem we have

$$L_1^* = L_\infty \ ,$$

and there exists a unique vector $\overline{p_j} = (p_{1j}, p_{2j}, \ldots) \in L_\infty$ such that $p_{ij} \ge 0$ and

$$F_j(f) = \Sigma_i \ f(i) \cdot \overline{p_j}(i) \cdot \Pi(i) = \Sigma_{i\in S} \ f(i) \cdot p_{ij} \cdot \Pi(i)$$

$$\text{for} \quad f \in L_1 \ .$$

In particular we have

$$T1_{\{k\}}(j) \cdot \Pi(j) = p_{kj} \cdot \Pi(k) \ .$$

Since

$$\Sigma_{i\varepsilon S}\, T1_{\{k\}} \cdot \Pi(i) = \Sigma_{i\varepsilon S}\, T1_{\{k\}} \cdot 1 \cdot \Pi(i) = \Sigma_{i\varepsilon S}\, 1_{\{k\}} \cdot P1 \cdot \Pi(i)$$

$$= \Sigma_{i\varepsilon S}\, 1_{\{k\}} \cdot \Pi(i) = \Pi(k) \ ,$$

then
$$\Sigma_j\, p_{kj} \cdot \Pi(k) = \Pi(k)$$

or
$$\Sigma_j\, p_{kj} = 1 \ .$$

Thus (p_{ij}) is a stochastic matrix satisfying (2.2).

Conversely, for each stochastic matrix (p_{ij}) and each $j \varepsilon S$,

$$F_j(f) = \Sigma_{i\varepsilon S}\, f(i) \cdot \Pi(i) \cdot p_{ij}$$

defines a positive linear functional F_j on L_1. F_j is also bounded since

$$|F_j(f)| = |\Sigma_i\, f(i) \cdot p_{ij} \cdot \Pi(i)| \le \Sigma_i\, |f(i)| \cdot p_{ij} \cdot \Pi(i) \le$$

$$\Sigma_i |f(i)| \cdot \Pi(i) = ||f||_1 \quad \text{for} \quad f \varepsilon L_1 \ .$$

Let $Tf(j) = \dfrac{F_j(f)}{\Pi(j)}$ for $f \varepsilon L_1$. Then we have

$$||Tf||_1 = \Sigma_j |Tf|(j) \cdot \Pi(j) = \Sigma_j\, |\Sigma_i\, f(i) \cdot p_{ij} \cdot \Pi(i)|$$

$$\le \Sigma_j\, \Sigma_i |f(i)| \cdot p_{ij} \cdot \Pi(i) = \Sigma_i |f(i)| \cdot \Pi(i) \cdot (\Sigma_j\, p_{ij})$$

$$= \Sigma_i |f(i)| \cdot \Pi(i) = ||f||_1 \ .$$

Thus $||Tf||_1 \le 1$. It is easy to see that T is a positive linear operator.

We also notice that for $f \in L_1$, using Fubini's Theorem

$$\Sigma_i Tf(i) \cdot \Pi(i) = \Sigma_i F_i(f) = \Sigma_i \Sigma_j f(j) \cdot \Pi(j) \cdot p_{ji} =$$

$$\Sigma_j f(i) \cdot \Pi(j) \cdot (\Sigma_i p_{ji}) = \Sigma_j f(j) \cdot \Pi(j) .$$

Hence T is a Markov operator on L_1. □

The Markov operator P on L_∞ satisfies the equation

$$\Sigma_i f(i) \cdot Ph(i) \cdot \Pi(i) = \Sigma_i Tf(i) \cdot h(i) \cdot \Pi(i)$$

$$\text{for } f \in L_1 , h \in L_\infty .$$

Letting $f = 1_{\{i\}}$, yields

$$\Sigma_{k\in S} f(k) \cdot Ph(k) \cdot \Pi(k) = Ph(i) \cdot \Pi(i) .$$

Thus by equation (2.2)

$$Tf(j) = \frac{1}{\Pi(j)} \Sigma_{k\in S} f(k) \cdot \Pi(k) \cdot p_{kj} = \frac{1}{\Pi(j)} \cdot \Pi(i) \cdot p_{ij} .$$

Therefore

$$\Sigma_{j\in S} Tf(j) \cdot h(j) \cdot \Pi(j) = \Sigma_{j\in S} \frac{1}{\Pi(j)} \cdot \Pi(i) \cdot p_{ij} \cdot h(j) \cdot \Pi(j) =$$

$$\Pi(i) \Sigma_j p_{ij} \cdot h(j) .$$

Hence we have equation

$$Ph(i) = \Sigma_j p_{ij} \cdot h(j) \quad \text{for } h \in L_\infty . \tag{2.3}$$

Which proves the following.

<u>Corollary 2.2</u> For each stationary Markov chain $\{X_n\}$ with state space S and initial distribution Π and the transition matrix (p_{ij}), there exists a unique Markov operator T on L_1 such that its adjoint P is representable by the transition matrix (p_{ij}) as in equation (2.3).

Now since each stationary Markov chain can be identified with a Markov operator T on L_1, we shall use Hopf's theory of Markov operators to solve problems of Markov chains.

We now state without proof the following fundamental theorem of Hopf (Neveu [7] Theorem V. 5.2. p. 196).

<u>Theorem 2.3</u> For every Markov operator T on L_1 there exists a unique subset C of S, for example we set

$$C = \{i: \Sigma_{k=0}^{\infty} T^k 1(i) = \infty\} \quad \text{and} \quad D = \{i: \Sigma_{k=0}^{\infty} T^k 1(i) < \infty\}$$

such that for all $f \in L_1^+$,

$$\Sigma_{k=0}^{\infty} T^k f = 0 \text{ or } +\infty \text{ on } C$$

$$\Sigma_{k=0}^{\infty} T^k f < \infty \text{ on } D \text{ (the complement of } C)$$

For each $f \in L_1^+$ we have

$$\{i: \Sigma_{k=0}^{\infty} T^k f(i) = \infty\} = C \cap \{i: \Sigma_{k=0}^{\infty} T^k f(i) > 0\} .$$

In future we may denote $\Sigma_{k=0}^{\infty} T^k$ by T_∞, and for $f \in L_1^+$, $C_f = \{i: \Sigma_{k=0}^{\infty} T^k f(i) = \infty\}$.

<u>Definition 2.2</u> The sets C and D in Theorem 2.3 are called, respectively, the conservative and the dissipative parts of S relative to T.

Recall (Definition 1.6) that a state i is said to be recurrent or transient according as

$$\Sigma_n p_{ii}^n = \infty \quad \text{or} \quad \Sigma_n p_{ii}^n < \infty .$$

We will show that the conservative part C of S consists of all recurrent states, or, equivalently the dissipative part D consists of all transient states. We note that from (2.2)

$$T^n f(j) \cdot \Pi(j) = \Sigma_i f(i) \cdot \Pi(i) \cdot p_{ij}^n \quad \text{for } f \in L_1 ,$$

and

$$\Sigma_{n=0}^{\infty} T^n f(j) \cdot \Pi(j) = \Sigma_{n=0}^{\infty} \Sigma_i f(i) \cdot \Pi(i) \cdot p_{ij}^n$$

$$= \Sigma_i f(i) \cdot \Pi(i) (\Sigma_n p_{ij}^n) \quad \text{for } f \in L_1^+ .$$

We now prove the following.

<u>Theorem 2.4</u> i) $C = \{i: \Sigma_{n=0}^{\infty} p_{ii}^n = \infty\}$

ii) $D = \{i: \Sigma_{n=0}^{\infty} p_{ii}^n < \infty\}$.

<u>Proof</u>. Recall

$$D = \{j: \; T_\infty 1(j) < \infty\} = \{j: \; \Sigma_i \Pi(i) \cdot (\Sigma_n p_{ij}^n) < \infty\} .$$

Thus for $j \in D$ we have

$$\Sigma_{n=0}^{\infty} p_{ij}^n < \infty \quad \text{for each} \quad i \in S .$$

So in particular

$$\Sigma_{n=0}^{\infty} p_{jj}^n < \infty .$$

Therefore

$$D \subset \{j: \; \Sigma_{n=0}^{\infty} p_{jj}^n < \infty\}$$

or equivalently

$$D \supset \{j: \; \Sigma_{n=0}^{\infty} p_{jj}^n = \infty\} .$$

However, for $j \in C$

$$\{i: \; T_\infty 1_{\{j\}}(i) = \infty\} = C \cap \{i: \; T_\infty 1_{\{j\}}(i) > 0\}$$

or equivalently

$$\{i: \; \Sigma_{n=0}^{\infty} p_{ji}^n = \infty\} = C \cap \{i: \; \Sigma_{n=0}^{\infty} p_{ji}^n > 0\} .$$

Since $1 = p_{jj}^0 \leq \Sigma_{n=0}^{\infty} p_{jj}^n$ we have $j \in \{i: \; \Sigma_{n=0}^{\infty} p_{ji} = \infty\}$.

Thus

$$C = \{j: \; \Sigma_{n=0}^{\infty} p_{jj}^n = \infty\} \quad \text{and} \quad D = \{j: \; \Sigma_{n=0}^{\infty} p_{jj} < \infty\} . \qquad \square$$

§2.2 Invariant Sets

In this section we shall first define subsets $C_k \subset S$

$$C_k = \{j: \ \Sigma_n \ p_{kj}^n = \infty\} \quad \text{where} \quad k \in C \ .$$

Next we shall identify a family $\mathcal{C}$ consisting of all subsets $B \subset C$ such that $P1_B = 1_B$ on C. We shall further prove that $\mathcal{C}$ is the σ-algebra generated by the partition $\{C_k\}$. Hence $\mathcal{C}$ is the family of invariant sets as identified by Hopf.

Recall that $T1_{\{k\}}(j) = \frac{\Pi(k)}{\Pi(j)} \cdot p_{kj}$. Thus for all n

$$T^n 1_{\{k\}}(j) = \frac{\Pi(k)}{\Pi(j)} \cdot p_{kj}^n \ .$$

Therefore we have

$$\{j: \ \Sigma_n \ T^n 1_{\{k\}}(j) = \infty\} = \{j: \ \Sigma_n \ p_{kj}^n = \infty\} \tag{2.4}$$

This set contains $\{k\}$ if and only if $k \in C$. In particular we define the set

$$C_k = \{j: \ \Sigma_n \ p_{kj}^n = \infty\} \quad \text{where} \quad k \in C \ .$$

<u>Lemma 2.1</u> The sets C_k and C are stochastically closed. That is

$$P1_{C_k} = 1_{C_k} \quad \text{on} \quad C_k$$

$$P1_C = 1_C \quad \text{on} \quad C \ .$$

__Proof.__ We shall first prove for fixed $i \in C_k$ we have

$$Pl_{\{i\}}(j) = p_{ji} = 0 \quad \text{for all} \quad j \in C_k \ . \tag{2.5}$$

Recall for $j \in C_k$, $T_\infty 1_{\{k\}}(j) = \infty$, but for $i \notin C_k$

$$\Sigma_{j\in C_k} T_\infty 1_{\{k\}}(j) \cdot Pl_{\{i\}}(j) \cdot \Pi(j) \le \Sigma_{j\varepsilon S} T_\infty 1_{\{k\}}(j) \cdot Pl_{\{i\}}(j) \cdot \Pi(j)$$

$$= \Sigma_{j\varepsilon S}(\Sigma_{n\ge 1} T^n 1_{\{k\}}(j)) \cdot 1_{\{i\}}(j) \cdot \Pi(j)$$

$$\le T_\infty 1_{\{k\}}(i) \cdot \Pi(i) < \infty \ .$$

Thus $Pl_{\{i\}}(j) = 0$ for all $j \in C_k$, $i \notin C_k$.

Similarly, for each $i \in D$ we have

$$Pl_{\{i\}}(j) = p_{ji} = 0 \quad \text{for all} \quad j \in C \ . \tag{2.6}$$

Since for $i \in D$ we have, as above

$$\Sigma_{j\varepsilon C} T_\infty 1(j) \cdot Pl_{\{i\}}(j) \cdot \Pi(j) \le \Sigma_{j\varepsilon S} T_\infty 1(j) \cdot Pl_{\{i\}}(j) \cdot \Pi(j)$$

$$= \Sigma_{j\varepsilon S}(\Sigma_{n=1}^{\infty} T^n 1)(j) \cdot 1_{\{i\}}(j) \cdot \Pi(j)$$

$$= \Sigma_{n=1}^{\infty} T^n 1(i) \cdot \Pi(i) < \infty \ .$$

Thus $Pl_{\{i\}}(j) = 0$ for all $j \in C$, $i \notin C$.

From equations (2.5) and (2.6) it is clear that

$$Pl_{C_k^c}(j) = \Sigma_{i\varepsilon C_k^c} p_{ji} = 0 \quad \text{for each} \quad j \in C \ .$$

and

$$P1_D(j) = 0 \quad \text{for each} \quad j \in C \ .$$

By induction on n we also have for all $n \geq 0$

$$P^n 1_{C_k^c}(j) = 0 \quad \text{for} \quad j \in C_k \tag{2.7}$$

and

$$P^n 1_D(j) = 0 \quad \text{for} \quad j \in C \ . \tag{2.8}$$

Consequently

$$1 = \Sigma_{h \in S}\, p_{jh} = \Sigma_{h \in C_k}\, p_{jh} = P1_{C_k}(j) \quad \text{for each} \quad j \in C_k$$

and

$$1 = \Sigma_{h \in S}\, p_{jh} = \Sigma_{h \in C}\, p_{jh} = P1_C(j) \quad \text{for each} \quad j \in C \ . \quad \square$$

<u>Corollary 2.3</u> If $f \in L_1^+$ is such that $f = 0$ on D then

$$T_\infty f(i) = 0 \quad \text{for} \quad i \in D \ .$$

<u>Proof</u>. By the formula

$$\Sigma_{i \in D}\, T^n f(i) \cdot \Pi(i) = \Sigma_{i \in S}\, T^n f(i) \cdot 1_D(i) \cdot \Pi(i)$$

$$= \Sigma_{i \in S}\, f(i) \cdot P^n 1_D(i) \cdot \Pi(i)$$

$$= \Sigma_{i \in C}\, f(i) \cdot P^n 1_D(i) \cdot \Pi(i)$$

and equation (2.8), we have $T^n f = 0$ on D for all n, so that

$T_\infty f(i) = 0$ on D . □

<u>Lemma 2.2</u> The following equalities hold on C .

$$Pl_{C_k} = l_{C_k} , \ Pl_C = l_C , \ Pl_{C_k^c} = l_{C_k^c} .$$

<u>Proof</u>. We see that $Pl_{C_k} \geq l_{C_k}$ on C , since $Pl_{C_k}(j) = 1 = l_{C_k}(j)$ for $j \in C_k$ and $Pl_{C_k}(j) \geq 0 = l_{C_k}(j)$ otherwise. Let $f = l_C$; then by Corollary 2.3 we have

$$\Sigma_{i\varepsilon C}(\Sigma_{n=0}^{m} T^n f(i))(Pl_{C_k}(i) - l_{C_k}(i)) \cdot \Pi(i)$$

$$= \Sigma_{i\varepsilon S}(\Sigma_{n=0}^{m} T^n f(i))(Pl_{C_k} - l_{C_k})(i) \cdot \Pi(i)$$

$$= \Sigma_{i\varepsilon S} f(i) \Sigma_{n=0}^{m} P^n (Pl_{C_k} - l_{C_k})(i) \cdot \Pi(i)$$

$$= \Sigma_{i\varepsilon C}(P^{(m+1)} l_{C_k}(i) - l_{C_k}(i)) \cdot \Pi(i)$$

$$\leq \Sigma_{i\varepsilon C} P^{(m+1)} l_{C_k}(i) \cdot \Pi(i) \leq \Pi(C) \leq 1 , \text{ for } m = 1,2,\ldots .$$

Hence we have

$$\Sigma_{i\varepsilon C} T_\infty l_C(i)(Pl_{C_k}(i) - l_{C_k}(i)) \cdot \Pi(i) \leq 1 .$$

But $T_\infty l_C = \infty$ on C so $Pl_{C_k} = l_{C_k}$ on C . We have $Pl_{C_k} = l_C$ on C from Lemma 2.2.

Now clearly $$Pl_{C_k^c} = Pl - Pl_{C_k} ,$$

so by Theorem 2.1 and the above $Pl_{C_k^c} = 1 - Pl_{C_k} = l_{C_k^c}$ on C . □

Using induction on n we arrive at the immediate corollary.

Corollary 2.4 The following equalities hold on C .

$$P^n 1_{C_k} = 1_{C_k} \ , \ P^n 1_{C_k^c} = 1_{C_k^c} \ , \ P^n 1_C = 1_C \quad \text{for} \quad n = 1,2,\ldots \ .$$

We may now take a closer look at C_k and prove the following.

Lemma 2.3 $C_k = \{j: \ \Sigma_n p_{kj}^n > 0\} = \{j: \ k \to j\}$ where $k \in C$.

Proof. Since $C_k = \{j: \ \Sigma_n p_{kj}^n = \infty\} = \{j: \ T_\infty 1_{\{k\}}(j) = \infty\}$

$$= C \cap \{j: \ T_\infty 1_{\{k\}}(j) > 0\}$$

$$= C \cap \{j: \ \Sigma_n p_{kj}^n > 0\} \ ,$$

we have
$$C_k \subset \{j: \ \Sigma_n p_{kj}^n > 0\} \ .$$

However, for $k \in C$, $p_{kj}^n = 0$ for all $j \notin C_k$ and $n = 1,2,\ldots$.
So $\{j: \ \Sigma_n p_{kj}^n > 0\} \subset C_k$. Moreover

$$\{j: \ \Sigma_n p_{kj}^n > 0\} = \{j: \ k \to j\}$$

and the proof is complete. □

Lemma 2.4 $C_k = C_j$ for each $j \in C_k$.

Proof. We shall first show that $C_j \subset C_k$. Since $j \in C_k$ we have

$k \to j$, thus for any $i \in S$, if $j \to i$ then $k \to i$.
Therefore $C_j = \{i:\ j \to i\} \subset \{i:\ k \to i\} = C_k$.

Now to show that $C_k \subset C_j$ it is sufficient, on observing the above, to show that $k \in C_j$. Suppose $k \notin C_j$, then by Corollary 2.4, $P^n 1_{C_j}(k) = 0$ for all n . Thus $\Sigma_n p^n_{kj} = 0$ which contradicts $j \in C_k$. □

We can now state several immediate consequences of the previous two lemmas.

<u>Lemma 2.5</u> (i) $\Sigma_n p^n_{ij} = \infty$ for $i,j \in C_k$.

(ii) $C_i \cap C_j = \emptyset$ or $C_i = C_j$.

(iii) $\Sigma_n p^n_{ij} = 0$ whenever $i \in C_h$, $j \in C_k$, $C_h \cap C_k = \emptyset$.

(iv) $C_k = \{j:\ k \leftrightarrow j\}$ i.e. C_k is a recurrent class.

<u>Proof</u>. (i) If $i,j \in C_k$ then by Lemma 2.4 $i,j \in C_k = C_i = C_j$, so $\Sigma_n p^n_{ij} = \infty$ and $\Sigma_n p^n_{ji} = \infty$.

(ii) Suppose $C_i \cap C_j \neq \emptyset$ then for some $k \in S$, $k \in C_i \cap C_j$. Thus $C_k = C_i = C_j$ by Lemma 2.4.

(iii) Take $i \in C_h$, $j \in C_k$ where $C_h \cap C_k = \emptyset$. Then $C_i = C_h$ and $C_j = C_k$ so we have

$$i \in C_i \subset S - C_k = S - C_j = \{x:\ \Sigma_n p^n_{jx} = 0\}$$

$$j \in C_j \subset S - C_h = S - C_i = \{x:\ \Sigma_n p^n_{ix} = 0\} .$$

Hence $\Sigma_n p_{ij}^n = 0$.

(iv) From Lemma 2.3 we have $C_k = \{j: k \rightarrow j\}$.

Now by Lemma 2.4, for all $j \in C_k$ we have $C_j = C_k = \{i: j \rightarrow i\}$.

Since $k \in C_k = C_j$ we have $j \rightarrow k$ for all $j \in C_k$.

Thus $C_k \subset \{j: j \rightarrow k\}$ so $C_k = \{j: j \rightarrow k\} \cap \{j: k \rightarrow j\} = \{j: k \leftrightarrow j\}$ □

Having sets C_i and C_j defined in Lemma 2.5(ii) we may assume without loss of generality that $\{C_k\}_k$ denotes a countable partition of C . We shall now consider the measure theoretic aspects of the family $\{C_k\}_k$.

<u>Theorem 2.5</u> Let $\mathcal{C}$ be the class of subsets $B \subset C$ such that $Pl_B = 1_B$ on C . Then $\mathcal{C}$ is the σ-algebra generated by the partition $\{C_k\}_k$.

<u>Proof</u>. The class $\mathcal{C}$ is not empty since by Lemma 2.2 it contains the partition $\{C_k\}$ and set C . The σ-algebra $\mathcal{B}$ generated by the partition $\{C_k\}$ consists of all unions of sets C_k . Thus if $B \in \mathcal{B}$ then $\bigcup_t C_{k_t} = B$ and the C_{k_t} are pairwise disjoint, so

$$Pl_B(j) = Pl_{\bigcup_t C_{k_t}}(j) = \Sigma_{t \in \bigcup_t C_{k_t}} p_{ji} = \Sigma_t \Sigma_{i \in C_{k_t}} p_{ji}$$

$$= \Sigma_t Pl_{C_{k_t}}(j) .$$

Hence if $j \in B$ then $j \in C_{k_r}$ for exactly one $r \in \{t\}$, and

$$Pl_B(j) = \Sigma_t Pl_{C_{k_t}}(j) = Pl_{C_{k_r}}(j) = 1 \text{ by Lemma 2.2}$$

On the other hand, if $j \in (C - B)$ then $j \in C$, but for all t , $j \notin C_{k_t}$ and

$$Pl_B = l_B \text{ on } C , \text{ and } \mathcal{B} \subset \mathcal{C} \text{ by Lemma 2.2}$$

Thus

$$Pl_B = l_B \text{ on } C , \text{ and } \mathcal{B} \subset \mathcal{C} .$$

It remains only to prove that $B = \bigcup_t C_{k_t}$ for all $B \in \mathcal{C}$. Since $\{C_k\}$ is a partition of C we have $B = \bigcup_k (B \cap C_k)$. So we shall prove that if $B \cap C_k \neq \emptyset$ then $C_k \subset B$. If this is not the case then for $j \in C_k - B$ and $i \in B \cap C_k$, then recalling $Pl_B = l_B$ on C we have $p^n_{ji} = 0$ for all n , or equivalently $\Sigma_n p^n_{ji} = 0$. This contradicts Lemma 2.5(i) so $C_k \subset B$. Thus $\mathcal{C} \subset \mathcal{B}$ and the proof is complete. □

Using the terminology of Hopf as used by Neveu [7] we call a set $B \in \mathcal{C}$ an <u>invariant set</u>.

In view of Lemma 2.5(iv) it is clear that C_k is a recurrent class in the Markov chain sense, and C_k is also the atom of the σ-algebra $\mathcal{C}$, containing the state k .

It was pointed out by Neveu ([7], Corollary 2, p. 200) that the atom of $\mathcal{C}$ containing a recurrent state k could be identified by

$$C \cap \{j: \Sigma_n p_{jk} = \infty\} = C \cap \{j: \Sigma_n p_{jk} > 0\} \qquad (2.9)$$

We conclude this section by proving that the set defined by equation (2.9) is the atom C_k . Thus the proofs of the classification theorems in this paper differ from those given by Kim [6] only by the choice of representation of atoms of the σ-algebra $\mathcal{C}$.

<u>Lemma 2.7</u> For each $k \in C$,

$$\{j: \Sigma_n p_{kj}^n = \infty\} = C \cap \{i: \Sigma_n p_{ik}^n = \infty\} .$$

<u>Proof</u>. Let $C_k = \{j: \Sigma_n p_{kj}^n = \infty\}$ and $B_k = C \cap \{i: \Sigma_n p_{ik}^n = \infty\}$. For each $j \in C_k$ recall $k \in C_k = C_j$ so $\Sigma_n p_{ik}^n = \infty$.

Thus $C_k \subset B_k$. On the other hand, $\{C_j\}$ is a partition of C so if $i \in B_k$ then $i \in C_j$ for some j , say C_{j_i} . Since $\Sigma_n p_{ik}^n = \infty$, it follows that $C_{j_i} = C_k$. □

Similarly it is clear that

$$C_k = \{j: \Sigma_n p_{kj}^n > 0\} = C \cap \{i: \Sigma_n p_{ik}^n > 0\} \quad \text{for} \quad k \in C .$$

Thus we have

<u>Theorem 2.6</u> For all $k \in C$

$$C_k = C \cap \{i: \Sigma_n p_{ik}^n = \infty\} = C \cap \{i: \Sigma_n p_{ik}^n > 0 \} .$$

By the following example we realize that the set $D \cap \{i: \Sigma_n p_{ik}^n = \infty\}$, where $k \in C$ is not necessarily void.

<u>Example 2.1</u> Let a,b,c be in $[0,1]$ such that $c \neq 1$, and $a + b + c = 1$.

Let $I = \{1,2,3\}$ and let (p_{ij}) for $i,j \in I$ be such that $p_{11} = p_{22} = 1$, $p_{31} = a$, $p_{32} = b$, $p_{33} = c$ and other $p_{ij} = 0$. Then $C = \{1,2\}$, $D = \{3\}$ are the conservative and dissipative parts of I. For each $n \geq 1$ we have

$$p^n_{31} = \frac{a(1-c^n)}{1-c}, \quad p^n_{32} = \frac{b(1-c^n)}{1-c}, \quad p^n_{33} = c^n,$$

$$p^n_{11} = p^n_{22} = 1, \quad p^n_{ij} = 0 \text{ otherwise.}$$

So $\Sigma_n p^n_{31} = \infty$ or 0 according as $a > 0$ or $a = 0$

$\Sigma_n p^n_{32} = \infty$ or 0 according as $b > 0$ or $b = 0$.

Hence

$$\{i: \Sigma_n p^n_{i1} = \infty\} = \begin{cases} \{1,3\} & \text{if } a > 0 \\ \{1\} & \text{if } a = 0 \end{cases}$$

and

$$\{i: \Sigma_n p^n_{i2} = \infty\} = \begin{cases} \{2,3\} & \text{if } b > 0 \\ \{2\} & \text{if } b = 0. \end{cases}$$

CHAPTER III - TWO CLASSIFICATION THEOREMS OF STATES

We will consider the Markov chain $\{X_n: n \geq 0\}$ with state space $S = \{1,2,3,...\}$, strictly positive initial distribution Π, and transition matrix (p_{ij}) as previously defined. Let T be the Markov operator corresponding to Markov chain $\{X_n\}$ in the sense of Corollary 2.2. We will prove two classification theorems of the states of Markov chains from the viewpoint of ergodic theory as developed in Chapter II.

It is worth noting that the two classification theorems involve only transition matrices and are independent of the initial distribution of the Markov chain $\{X_n\}$. Thus our restriction of a strictly positive initial distribution is not a serious restriction on the generality of the theorems.

§3.1 Classification Theorems

We shall first show that S can be divided into a set consisting of all transient states, and a family of recurrent classes. (Feller [3], Theorem 3, p. 392). We restate this theorem as

Theorem 3.1 Let $\{X_n\}$ be a discrete Markov chain.

Then the space S has a unique partition $\{D,C_1,C_2,\ldots\}$ such that D is the set of all transient states and C_i is a countable family of recurrent classes.

We then prove the corresponding theorem for idempotent discrete Markov chains. (Doob [2], Theorem 2, p. 39). An idempotent Markov chain is a Markov chain which has an idempotent transition matrix.

Before restating this theorem we notice that every substochastic matrix may easily be enlarged to a stochastic matrix by adding a top row $(1,0,0,\ldots)$ and a column whose entries are the deficiencies of the appropriate rows of the original matrix. We may now restate this second classification theorem as

Theorem 3.2 Let $\{X_n\}$ be an idempotent Markov chain. Then the state space S can be partitioned uniquely into the set D of all nonrecurrent states (or equivalently all inessential states) and a countable family of positive recurrent aperiodic classes $\{C_i\}$ such that

i) $j \in D$ if and only if $p_{ij} = 0$ for all $i \in S$

ii) if $x \in C_i$ and $y \in C_j$ then $p_{xy} = \delta_{ij}\, \nu_j(y)$

where ν_j is a probability measure with support C_j

iii) there are nonnegative numbers $\{\rho_{ti}\}$ such that

$$\Sigma_i\, \rho_{ti} = 1 \text{ and}$$

$$\rho_{tx} = \rho_{ti}\, \nu_i(x) \quad \text{for} \quad t \in D \,,\; x \in C_i \,.$$

§3.2 Proof of Theorem 3.1

In view of Corollary 2.2, we shall henceforth denote the unique Markov operator on L_1 corresponding to the Markov chain $\{X_n\}$ by T. Let C and D be the conservative and dissipative parts of S relative to T. Then $\{C_i\}$ as defined in Chapter II is a family of recurrent classes, as proven in Lemma 2.5. The dissipative part of S, set D consists of all transient states as proven in Theorem 2.4. From Lemma 2.5 we clearly see that $\{C_i\}$ partitions C and therefore the family $\{D, C_1, C_2, \ldots\}$ partitions S. This partition is unique by virtue of the uniqueness of the recurrence classes. Thus the proof of Theorem 3.1 is complete. □

§3.3 Proof of Theorem 3.2

For this section we assume that the Markov chain $\{X_n\}$ is idempotent, that is, its transition matrix (p_{ij}) is idempotent. Hence the Markov operators T and P are also idempotent. If we set $\mu = T1$, then $T\mu = \mu$ and it is easily seen that $\mu(x) > 0$ iff $T_\infty \mu(x) > 0$ iff $T_\infty \mu(x) = \infty$. So we have

$$\{x\colon\ \mu(x) > 0\} = \{x\colon\ T_\infty \mu(x) = \infty\} \in C \,.$$

Since we have $C = \{x\colon\ T_\infty 1(x) = \infty\}$, if $i \in C$ then $T_\infty \mu(i) = \Sigma_{n=1}^{\infty} T^n 1(i) \geq T_\infty 1(i) - 1 = \infty$. Thus $C = \{x\colon\ \mu(x) > 0\}$.

Proof of Theorem 3.2

(i) We notice from the remark above that $D = \{j: \mu(j) = 0\}$. Since $\mu(j) = T1(j) = \frac{1}{\Pi(j)} \Sigma_i p_{ij} \Pi_i$, then

$$j \in D \text{ if and only if } p_{ij} = 0 \text{ for each } i \in S .$$

This proves (i). In particular $P1_D(i) = 0$ for all $i \in S$, thus we have

$$1 = P1 = P1_C + P1_D = P1_C . \quad (3.1)$$

(ii) Recall $\Sigma_n p^n_{xy} = \infty$ for $x,y \in C_k$ as in Lemma 2.5, but for all n , $p^n_{xy} = p_{xy}$ so $p_{xy} > 0$ for $x,y \in C_k$. Also by Lemma 2.5 we see that $p_{xy} = 0$ for $x \in C_i$, $y \in C_j$ if $i \neq j$.

We shall now show that for a given $y \in C_i$, $p_{xy} = p_{yy}$ for all $x \in C_i$. Since $P^2 = P$, letting $g = P1_{\{y\}}$ we have

$$Pg = P(P1_{\{y\}}) = P1_{\{y\}} = g .$$

Hence, by the argument of [7, pp. 198-199] the function $g(z) = P1_{\{y\}}(z) = p_{zy}$ is $\mathcal{C}$-measurable. C_i is an atom of the σ-algebra $\mathcal{C}$, so the set $B_i = C_i \cap \{z: p_{zy} < p_{yy}\}$ is either C_i or $\emptyset$. If $B_i = C_i$ then $p_{yy} < p_{yy}$ so $B_i = \emptyset$. Similarly the set $C_i \cap \{z: p_{zy} > p_{yy}\}$ is empty. Thus given $y \in C_i$, $p_{xy} = p_{yy}$ for all $x \in C_i$. In particular $p_{xy} = p_{iy}$ for all $x,y \in C_i$.

Let us associate the mapping $\nu_i: S \to [0,1]$ with C_i , for $i \in C$, as follows.

$$\nu_i(z) = p_{iz} \quad \text{for} \quad z \in S .$$

Now ν_i is a probability measure on S with support C_i, since $\nu_i(z) = p_{iz}$ is positive or zero according as $z \in C_i$ or $z \notin C_i$ and by Lemma 2.2 we have

$$\Sigma_{z \in C_i} \nu_i(z) = \Sigma_{z \in C_i} p_{iz} = P1_{C_i}(i) = 1_{C_i}(i) = 1 .$$

This completes the proof of (ii).

(iii) For $t \in D$ and $x \in C_i$ simple calculation yields

$$p_{tx} = p^2_{tx} = \Sigma_{z \in S} p_{tz} p_{zx} = \Sigma_{z \in C_i} p_{tz} p_{zx} = P1_{C_i}(t) \nu_i(x) .$$

If we set $\rho_{ti} = P1_{C_i}(t)$, then $p_{tx} = \rho_{ti} \nu_i(x)$ and $0 \leq \rho_{ti} \leq 1$. So in view of (3.1)

$$\Sigma_i \rho_{ti} = \Sigma_i P1_{C_{(i)}}(t) = P1_C(t) = P1(t) = 1$$

which establishes (iii).

Since $p^n_{xx} = p_{xx} > 0$ for all $x \in C_i$, $n \geq 1$, each C_i is obviously a positive recurrent class of period 1. We note however from part (i) that each state in D which is transient, is an inessential state.

From part (i) it is clear that for state $x \in C_i$, if $x \to y$ then $y \in C$. Then by Lemma 2.3 $y \in C_i$. But, by Lemma 2.5 it is clear that $x \leftrightarrow y$. Thus for all i, each state of C_i is essential. Hence each state of C is essential.

Therefore D consists of all inessential states. □

CHAPTER IV - LIMITING PROPERTIES OF STOCHASTIC MATRICES

§4.1 Construction of Idempotent Stochastic Matrices

In Chapter III we noted an important result which holds for idempotent Markov chains (Theorem 3.2), which can be represented by idempotent stochastic matrices. We now show how these idempotent stochastic matrices arise quite naturally and how they can be constructed.

We shall first give the simplest example of an idempotent stochastic matrix, then list two special cases which result in idempotent stochastic matrices.

Example 4.1 The simplest example of an idempotent stochastic matrix is one of the following type.

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \cdots \alpha_n \cdots \\ \alpha_1 & \alpha_2 & \alpha_3 \cdots \alpha_n \cdots \\ \cdot & \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot & \cdot \quad\quad \cdot \end{pmatrix}$$

such that $\alpha_i \geq 0$ for all i and $\Sigma_i \alpha_i = 1$.

This matrix is clearly stochastic and it is also idempotent. Let $AA = B = (b_{ij})$. Then

$$b_{ij} = \Sigma_{k\geq 1}(\alpha_k \cdot \alpha_j) = \alpha_j(\Sigma_{k\geq 1} \alpha_k) = \alpha_j \quad \text{for all } i,j .$$

Hence $B = A$ so A is idempotent.

Theorem 4.1 Let P be an irreducible recurrent aperiodic (ergodic) Markov chain. Then

$\lim P^n = A = (a_{ij})$ exists, that is, such that for each j ,

$$a_{ij} = \alpha_j \begin{cases} > 0 \text{ with } \Sigma_j \alpha_j = 1 ; & \text{if } P \text{ is positive recurrent} \\ = 0 & \text{if } P \text{ is null recurrent.} \end{cases}$$

If P is positive, $\alpha = (\alpha_1, \alpha_2, \ldots)$ is a unique invariant probability measure for P . That is

$$\alpha_j = \Sigma_i \alpha_i p_{ij} \quad \text{for } j = 1,2,\ldots . \tag{4.1}$$

Proof. It is well known (Chung [1], Theorem 1, p. 28; Karlin [4], Theorem 6-38, p. 153) that

$$\lim_{n\to\infty} P^n = A \text{ exists, that is}$$

$$\lim_{n\to\infty} P^n_{ij} = \alpha_j > 0 \text{ for all } i,j = 1,2,\ldots .$$

Since, for each n , $1 = \Sigma_j p^n_{ij}$, Fatou's Lemma yields

$$1 \geq \Sigma_j \lim_n p^n_{ij} = \Sigma_j \alpha_j .$$

However we also have

$$p_{ij}^{n+1} = \Sigma_k \, p_{ik}^n \, p_{kj} \quad \text{for} \quad n = 1,2,\ldots \; .$$

Again using Fatou's Lemma we have

$$\alpha_j = \lim_n p_{ij}^{n+1} \geq \Sigma_k \lim_n p_{ik}^n \, p_{kj} = \Sigma_k \, \alpha_k \, p_{kj} \; , \; j = 1,2,\ldots \; .$$

It follows that

$$\alpha_j \geq \Sigma_k \, \alpha_k \, p_{kj}^n \; , \; j = 1,2,\ldots \; ; \; n = 1,2,\ldots \; .$$

However the inequality reduces to equality. Otherwise we have

$$\alpha_{j_0} > \Sigma_k \, \alpha_k \, p_{kj_0}^{n_0} \quad \text{for some} \quad j_0 \quad \text{and} \quad n_0 \; .$$

This gives us

$$\Sigma_j \, \alpha_j > \Sigma_j \, \Sigma_k \, \alpha_k \, p_{kj}^{n_0} = \Sigma_k \, \alpha_k \, \Sigma_j \, p_{kj}^{n_0} = \Sigma_k \, \alpha_k$$

which is a contradiction. Thus we have (4.1). This equation implies that

$$\alpha_j = \Sigma_i \, \alpha_i \, p_{ij}^n \quad \text{for} \quad j = 1,2,\ldots \; ; \; n = 1,2,\ldots \; .$$

and by the Lebesque Bounded Convergence Theorem,

$$\alpha_j = \Sigma_i \, \alpha_i \, \alpha_j = (\Sigma_i \, \alpha_i) \, \alpha_j \; .$$

Hence

$$\Sigma_i \, \alpha_i = 1 \; .$$

Furthermore $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots)$ is an invariant probability measure for P, that is, $\alpha P = \alpha$. In fact α is the only invariant probability measure for P. If $\beta = \beta P$, that is $\beta_j = \Sigma_i \beta_i p_{ij}$ then

$$\beta_j = \Sigma_i \beta_i p_{ij}^n \quad \text{and}$$

$$\beta_j = \Sigma_i \beta_i \lim_{n\to\infty} p_{ij}^n = \Sigma_i \beta_i \alpha_j = \alpha_j .$$

It is readily seen that the limit matrix A is an idempotent stochastic matrix and $AP = PA = A = A^2$. □

<u>Definition 4.1</u> A sequence of matrices $\{A_m\} = \{(a_{ij}^m)\}$ is said to converge to $B = (b_{ij})$ if it converges coordinatewise. That is

$$\lim_{m\to\infty} a_{ij}^m = b_{ij} \quad \text{for all } i,j .$$

<u>Theorem 4.2</u> If $P = (p_{ij})$ is an $N \times N$ stochastic matrix, then there is a stochastic matrix $Q = (q_{ij})$ such that

$$\lim_{m\to\infty} \frac{1}{m} \Sigma_{k=1}^m P^k = Q ,$$

or equivalently,

$$\lim_{m\to\infty} \frac{1}{m} \Sigma_{k=1}^m p_{ij}^k = q_{ij} , \quad i,j = 1,2,\ldots, N .$$

Moreover, $QP = PQ = Q$ and $Q^2 = Q$.

Proof. Let Q_m denote $\frac{1}{m}\Sigma_{\nu=1}^{m} P^{\nu}$ for $m = 1,2,\ldots$ and let (q_{ij}^{m}) denote Q_m.

It is obvious that

$$0 \le q_{ij}^{m} = \frac{1}{m}\Sigma_{\nu=1}^{m} p_{ij}^{\nu} \le 1 \quad \text{for} \quad i,j = 1,\ldots,N .$$

For every pair (i,j), the sequence $\{q_{ij}^{m}\}_{m=1}^{\infty} \subset [0,1]$ has a convergent subsequence $q_{ij}^{m_\nu}$ such that

$$\lim_{m_\nu \to \infty} q_{ij}^{m_\nu} = q_{ij}$$

By repeating this process N^2-times all entries converge. Thus we have a subsequence of matrices which converge. Call this subsequence $\{Q_{m_k}\}$ such that

$$\lim_{m_k \to \infty} Q_{m_k} = Q = (q_{ij}) .$$

We shall now show that $Q = (q_{ij})$ is stochastic. Recall

$$q_{ij} = \lim_{m_k \to \infty} q_{ij}^{m_k} = \lim_{m_k \to \infty} \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} p_{ij}^{\nu} .$$

So

$$\Sigma_{j=1}^{N} q_{ij} = \Sigma_{j=1}^{N} \lim_{m_k \to \infty} q_{ij}^{m_k} = \lim_{m_k \to \infty} \Sigma_{j=1}^{N} q_{ij}^{m_k}$$

$$= \lim_{m_k \to \infty} \Sigma_{j=1}^{N} \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} p_{ij}^{\nu} = \lim_{m_k \to \infty} \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} \Sigma_{j=1}^{N} p_{ij}^{\nu}$$

$$= \lim_{m_k \to \infty} 1 = 1 .$$

Thus Q is stochastic.

Furthermore $PQ = Q = QP$ since for all m_k we have

$$PQ_{m_k} = P(\frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu}) = \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu+1} = Q_{m_k} P .$$

Hence

$$PQ = \lim_{m_k \to \infty} PQ_{m_k} = \lim_{m_k \to \infty} Q_{m_k} P = QP .$$

Moreover, for all m_k

$$\begin{aligned} PQ_{m_k} &= \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu+1} \\ &= \frac{1}{m_k} (P^2 + P^3 + \ldots + P^{m_k+1}) \\ &= \frac{1}{m_k} (P + P^2 + \ldots + P^{m_k}) - \frac{1}{m_k} (P - P^{m_k+1}) \\ &= Q_{m_k} - \frac{1}{m_k} (P - P^{m_k+1}) \end{aligned}$$

But $p_{ij} \varepsilon [0,1]$, and $p_{ij}^{m_k+1} \varepsilon [0,1]$ for all m_k and $i,j = 1,2,\ldots,N$. Thus

$$\lim_{m_k \to \infty} \frac{1}{m_k} p_{ij} = 0 \quad \text{and} \quad \lim_{m_k \to \infty} \frac{1}{m_k} p_{ij}^{m_k+1} = 0 .$$

So we conclude that

$$PQ = \lim_{m_k} PQ_{m_k} = \lim_{m_k} Q_{m_k} - \lim_{m_k} \frac{1}{m_k} P + \lim_{m_k} \frac{1}{m_k} P^{m_k+1} = Q . \tag{4.2}$$

Similarly we see that for any $m = 1,2,\ldots$ we have

$$Q = Q(\frac{1}{m} \Sigma_{\nu=1}^{m} P^{\nu}) = \frac{1}{m}(\Sigma_{\nu=1}^{m} P^{\nu})Q . \tag{4.3}$$

If $\{Q_{n_h}\}$ is a subsequence of $\{Q_m\}$ such that

$$\lim_{n_h \to \infty} Q_{n_h} = Q'$$

then from (4.3) it is clear that

$$Q = QQ' = Q'Q .$$

Now interchanging the roles of Q and Q' we have

$$Q' = Q'Q = QQ' \quad \text{so} \quad Q = Q' .$$

Thus $Q^2 = Q$, and we conclude that every convergent subsequence of $\{Q_m\}$ converges to Q . Therefore $\{Q_m\}$ converges to Q which completes the proof. □

We would like to be able to generalize Theorem 4.1 to include infinite stochastic matrices. With this end in mind we shall prove the following.

Theorem 4.3 If $P = (p_{ij})$ is an infinite stochastic matrix, there is a substochastic matrix $Q = (q_{ij})$ such that

$$\lim_{n\to\infty} \frac{1}{m} \Sigma_{k=1}^{m} P^k = Q .$$

Moreover, $QP = PQ = Q = Q^2$.

Proof. We may consider the matrix P to be an infinite dimensional real vector. If $M = \{B\colon\ B$ is a real infinite matrix$\}$, it is clear that $\mathcal{M}$ is an infinite dimensional vector space. We can topologize $\mathcal{M}$ with what we will call the coordinatewise limit topology, which corresponds with the usual topology on R^{∞}. Now

$$A = \{B = (b_{ij}): \forall\ i,j = 1,2,\dots\ ;\ 0 \le b_{ij} \le 1\}$$

is compact by the Tychonoff Theorem.

It now follows by simple calculation that P^{ν} is stochastic for all $\nu = 1,2,\dots$. For each m let $Q_m = \frac{1}{m} \Sigma_{\nu=1}^{m} P^{\nu} = (q_{ij})$. Thus for $m = 1,2,\dots$ we have

$$\Sigma_{j=1}^{\infty} q_{ij}^{m} = \Sigma_{j=1}^{\infty} \frac{1}{m} \Sigma_{\nu=1}^{m} p_{ij}^{\nu} = \frac{1}{m} \Sigma_{\nu=1}^{m} \Sigma_{j=1}^{\infty} p_{ij}^{\nu} = 1 .$$

Therefore Q_m is stochastic for all m , and the sequence $\{Q_m\}_{m=1}^{\infty}$ has a convergent subsequence, call it $\{Q_{m_k}\}$. So we have

$$\lim_{m_k \to \infty} Q_{m_k} = Q \text{ for some matrix } Q .$$

We further note that

$$PQ = Q = QP \tag{4.4}$$

since for all m_k

$$PQ_{m_k} = P(\frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu}) = \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu+1} = Q_{m_k} P .$$

Thus

$$PQ = \lim_{m_k \to \infty} PQ_{m_k} = \lim_{m_k \to \infty} Q_{m_k} P = QP .$$

It is also clear that $PQ = Q$ since

$$PQ_{m_k} = \frac{1}{m_k} \Sigma_{\nu=1}^{m_k} P^{\nu+1} = \frac{1}{m_k}(\Sigma_{\nu=1}^{m_k} P^{\nu} - (P - P^{m_k+1}))$$

$$= Q_{m_k} - \frac{1}{m_k} P + \frac{1}{m_k} P^{m_k+1} .$$

It follows that

$$PQ = \lim_{m_k \to \infty} PQ_{m_k} = \lim_{m_k \to \infty} Q_{m_k} - \lim_{m_k \to \infty} (\frac{1}{m_k} (P - P^{m_k+1})) = Q .$$

We also have

$$Q = Q(\frac{1}{m} \Sigma_{\nu=1}^{m} P^{\nu}) = (\frac{1}{m} \Sigma_{\nu=1}^{m} P^{\nu}) Q . \qquad (4.5)$$

Now if $\{Q_{n_h}\}$ is a subsequence of $\{Q_m\}$ and $\lim_{n_h \to \infty} Q_{n_h} = Q'$, then by (4.2) we have

$$Q = QQ' = Q'Q .$$

Interchanging the roles of Q and Q' it follows that

$$Q = Q' .$$

So $Q^2 = Q$, and every convergent subsequence of Q_m converges to Q.

Hence $\lim_{m \to \infty} Q_m = Q$.

It remains only to prove that Q is substochastic. We have previously shown that for all m

$$Q_m = \frac{1}{m} \Sigma_{\nu=1}^{m} P = (q_{ij}^{m})$$

is stochastic, and

$$q_{ij}^m = \frac{1}{m} \Sigma_{\nu=1}^m p_{ij}^\nu \quad \text{for all } i \text{ and } j .$$

Recall that

$$q_{ij} = \lim_{m\to\infty} q_{ij}^m \quad \text{for } i,j = 1,2,\ldots .$$

Thus for each i , by Fatou's Lemma we have

$$\Sigma_{j=1}^\infty q_{ij} = \Sigma_{j=1}^\infty \lim_{m\to\infty} q_{ij}^m \le \lim_{m\to\infty} \Sigma_{j=1}^\infty q_{ij}^m = 1 .$$

Hence Q is substochastic. □

It is now straightforward to prove the following.

<u>Theorem 4.4</u> If $P = (p_{ij})$ is an infinite (finite) substochastic matrix then there is a substochastic matrix $Q = (q_{ij})$ such that

$$\lim_{m\to\infty} \frac{1}{m} \Sigma_{k=1}^m P^k = Q \quad \text{and} \quad QP = PQ = Q = Q^2 .$$

To set out necessary and sufficient conditions for the substochastic matrix Q in Theorem 4.3 being stochastic we prove the following.

<u>Theorem 4.5</u> Let $P = (p_{ij})$ be a stochastic matrix and let $Q = \lim_{m\to\infty} \frac{1}{m} \Sigma_{k=1}^m P^k$. Let $Q = (q_{ij})$ and let $q_{ij}^m = \frac{1}{m} \Sigma_{k=1}^m p_{ij}^k$. Then Q is a stochastic matrix if and only if, for each i , the series

$$\Sigma_{j=1}^\infty q_{ij}^m$$

in j converges uniformly with respect to m .

<u>Proof</u>. (If) Suppose that, for a fixed i , the series $\Sigma_{j=1}^{\infty} q_{ij}^{m}$ in j converges uniformly with respect to m . We notice that

$$\Sigma_{j=1}^{\infty} q_{ij}^{m} = \Sigma_{j=1}^{\infty} \frac{1}{m} \Sigma_{k=1}^{m} p_{ij}^{k} = \frac{1}{m} \Sigma_{k=1}^{m} \Sigma_{j=1}^{\infty} p_{ij}^{k} = \frac{1}{m} \Sigma_{k=1}^{m} 1 = 1 .$$

For $\varepsilon > 0$, let j_ε be such that

$$\Sigma_{j>j_\varepsilon} q_{ij}^{m} = 1 - \Sigma_{j\le j_\varepsilon} q_{ij}^{m} < \varepsilon \quad \text{for al} \quad m .$$

It follows that

$$\overline{\lim_{m\to\infty}} \Sigma_{j>j_\varepsilon} q_{ij}^{m} = 1 - \lim_{m\to\infty} \Sigma_{j\le j_\varepsilon} q_{ij}^{m} \le \varepsilon . \tag{4.6}$$

Since

$$\lim_{m\to\infty} \Sigma_{j\le j_\varepsilon} q_{ij}^{m} = \Sigma_{j\le j_\varepsilon} \lim_{m\to\infty} q_{ij}^{m} = \Sigma_{j\le j_\varepsilon} q_{ij} ,$$

from (4.6) we have

$$\Sigma_{j=1}^{\infty} q_{ij} = 1 .$$

Hence Q is a stochastic matrix.

(Only if) Suppose that Q is a stochastic matrix. Since $\Sigma_{j=1}^{\infty} q_{ij} = 1$, for each i and a given $\varepsilon > 0$, there is a j_1 such that

$$\Sigma_{j>j_1} q_{ij} < \varepsilon .$$

Then we have

$$\lim_{m\to\infty} \Sigma_{j>j_1} q_{ij}^m = 1 - \lim_{m\to\infty} \Sigma_{j\le j_1} q_{ij}^m = 1 - \Sigma_{j\le j_1} q_{ij} = \Sigma_{j>j_1} q_{ij} < \varepsilon .$$

It follows, for some m_1, that

$$\Sigma_{j>j_1} q_{ij}^m < \varepsilon \quad \text{for} \quad m > m_1 .$$

However, we also have

$$\Sigma_{j=1}^{\infty} q_{ij}^m = 1 \quad \text{for} \quad m \le m_1 ,$$

so there is j_2 such that

$$\Sigma_{j>j_2} q_{ij}^m < \varepsilon \quad \text{for} \quad m \le m_1 .$$

Setting $j_0 = \max(j_1, j_2)$ we have

$$1 - \Sigma_{j\le j_0} q_{ij}^m = \Sigma_{j>j_0} q_{ij}^m < \varepsilon$$

uniformly with respect to m, which completes the proof. □

§4.2 Euler-summability of Irreducible Stochastic Matrices.

Recall (Theorem 4.1) that for P, the transition matrix for an irreducible recurrent aperiodic Markov chain, the $\lim_{n\to\infty} P^n$ exists. This limit on the other hand need not exist in the case of a Markov chain with period $d > 1$. For example if $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ then $\lim_{n\to\infty} P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\lim_{n\to\infty} P^{2n+1} = P$. However, it is known (Kemeny, Snell [5], Theorem 5.1.1, pp. 99, 100) that for each irreducible positive recurrent finite Markov chain, the sequence of powers of its transition matrix, P^n, is Euler-summable to a limiting matrix $A = (a_{ij})$. That is for $0 < \lambda < 1$

$$\lim_{n\to\infty} (\lambda I + (1 - \lambda)P)^n = A$$

where I is the identity matrix.

Moreover, for each j,

$$a_{ij} = \alpha_j > 0 \ , \ i = 1,2,\ldots \quad \text{and} \quad \Sigma_j \alpha_j = 1 \ .$$

We should now like to see what analogs of the above facts hold for an irreducible infinite Markov chain. We shall therefore prove the following theorem.

Theorem 4.6 If P is a stochastic matrix for an irreducible Markov chain then the sequence $\{P^n\}$ is Euler-summable, that is

$$\lim_{n\to\infty} (\lambda I + (1 - \lambda)P)^n = A = (a_{ij})$$

exists for $0 < \lambda < 1$.

Moreover, for each j

$$a_{ij} = \alpha_j \begin{cases} > 0 \text{ with } \Sigma_j \alpha_j = 1 & \text{if } Q = (\lambda I + (1 - \lambda)P) \text{ is positive recurrent.} \\ = 0 & \text{otherwise.} \end{cases}$$

Proof. We shall begin by proving for each $0 < \lambda < 1$ that $Q = (\lambda I + (1 - \lambda)P)$ is an irreducible aperiodic Markov chain.

For $0 < \lambda < 1$, let $Q = (q_{ij}) = \lambda I + (1 - \lambda)P$, then

$$q_{ij} = \lambda \delta_{ij} + (1 - \lambda)p_{ij} \quad \text{and} \quad \Sigma_j q_{ij} = 1 .$$

where δ_{i_j} is Kronecker's delta.

Thus Q is a stochastic matrix. For each $n \geq 1$ we have

$$q^n_{ij} = \Sigma^n_{k=0} \binom{n}{k} \lambda^k (1 - \lambda)^{n-k} p^{n-k}_{ij} .$$

Since P is irreducible, that is, for each pair (i,j) , there are m and n such that $p^n_{ij} > 0$ and $p^m_{ji} > 0$, we have

$$q^n_{ij} \geq (1 - \lambda)^n p^n_{ij} > 0$$

and

$$q^m_{ji} \geq (1 - \lambda)^m p^m_{ji} > 0 .$$

Hence Q is irreducible. Furthermore it is clear that $0 < \lambda \leq q_{11}$ so Q is aperiodic.

Now we shall prove the following relation.

$$\Sigma^\infty_{n=1} q^n_{11} = \frac{\lambda}{1-\lambda} + \frac{1}{1-\lambda} \Sigma^\infty_{n=1} p^n_{11} . \tag{4.7}$$

We come upon (4.7) by the following calculations.

$$\Sigma^\infty_{n=1} q^n_{11} = \Sigma^\infty_{n=1} \Sigma^n_{k=0} \binom{n}{k} \lambda^{n-k} (1 - \lambda)^k p^k_{11} \tag{4.8}$$

$$= \Sigma^\infty_{n=1} \lambda^n + \Sigma^\infty_{k=1} (1 - \lambda)^k p^k_{11} \Sigma^\infty_{n=k} \binom{n}{k} \lambda^{n-k}$$

$$= \frac{\lambda}{1-\lambda} + \Sigma^\infty_{k=1} (1 - \lambda)^k p^k_{11} S_k$$

where $S_k = \Sigma_{n=k}^{\infty} \begin{matrix} n \\ k \end{matrix} \lambda^{n-k}$.

It remains then to prove

$$S_k = \frac{1}{(1-\lambda)^{k+1}} \quad \text{for} \quad k = 1,2,\ldots \ . \tag{4.9}$$

We prove (4.7) by induction on k . It is easy to see that $S_1 = \frac{1}{(1-\lambda)^2}$. Now suppose (4.7) holds for an arbitrary k . Then

$$(1 - \lambda)\ S_{k+1} = \Sigma_{m=0}^{\infty} \binom{k+1+m}{k+1} \lambda^m - \Sigma_{m=0}^{\infty} \binom{k+1+m}{k+1} \lambda^{m+1}$$

$$= 1 + \Sigma_{m=1}^{\infty} \binom{k+1+m}{k+1} \lambda^m - \Sigma_{m=1}^{\infty} \binom{k+m}{k+1} \lambda^m$$

$$= 1 + \Sigma_{m=1}^{\infty} \binom{k+m}{k} \lambda^m$$

$$= \Sigma_{n=k}^{\infty} \binom{n}{k} \lambda^{n-k} = S_k \ .$$

Thus $S_{k+1} = \frac{1}{(1-\lambda)^{k+2}}$ which proves (4.9).

Now (4.7) follows directly from (4.8) and (4.9). Hence we have

$$\Sigma_{n=0}^{\infty} q_{11}^n = \frac{1}{1-\lambda} \Sigma_{n=0}^{\infty} p_{11}^n \ .$$

Thus P is recurrent iff Q is recurrent.

Suppose P is irreducible recurrent, then Q is irreducible, recurrent and aperiodic. A well known theorem of Kolmogorov (Theorem 4.1) states that for an irreducible aperiodic recurrent Markov chain with transition matrix Q

$$\lim_{n\to\infty} Q^n = A = (a_{ij})$$

exists, and for each j

$$a_{ij} = \alpha_j \begin{cases} > 0 \quad \text{with } \Sigma_j \alpha_j = 1 \text{ , } & \text{if } Q \text{ is positive} \\ = 0 & \text{if } Q \text{ is null recurrent.} \end{cases}$$

Thus the theorem holds in case P is recurrent.

Now suppose P is transient; then Q is also transient. Then $\lim_{n\to\infty} Q^n = 0$ (Chung [1], Theorem 5, p. 24). This completes the proof. □

BIBLIOGRAPHY

[1] Chung, K. L. (1967). Markov Chains with Stationary Transition Probabilities. Springer-Verlag, New York.

[2] Doob, J. L. (1942). Topics in the Theory of Markoff Chains. Trans. Amer. Math. Soc. 52 37-64.

[3] Feller, W. (1967). An Introduction to Probability Theory and its Applications. Wiley, New York.

[4] Karlin, S. (1966). A First Course in Stochastic Processes. Academic Press, New York.

[5] Kemeny, J. G. and Snell, J. L. (1967). Finite Markov Chains. Van Nostrand, New Jersey.

[6] Kim, C. W. (1970). Two Classification Theorems of States of Markov Chains. J. Appl. Prob. 7 490-496.

[7] Neveu, J. (1965). Mathematical Foundations of the Calculus of Probability. Holden-Day, San Francisco.

[8] Révész, P. (1962). A Probabilistic Solution of Problem 111 of G. Birkhoff. Acta Mathematica, Budapest 187-198.