In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

3/17/65
b

COMPATIBILITY OF MARKOV CHAIN DEVELOPMENTS

A THESIS

Presented to

The Faculty of the Graduate Division

by

Selwyn Lester Flournoy, Jr.

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

January, 1966

COMPATIBILITY OF MARKOV CHAIN DEVELOPMENTS

Approved:

____
Chairman

____

____

____

Date approved by Chairman: 1/13/66

ACKNOWLEDGMENTS

The author wishes to express his gratitude to his thesis advisor, Dr. John D. Neff, whose teaching and helpful suggestions made the writing of this paper possible.

The author also wishes to thank the other members of his Reading Committee, Doctors Eric R. Immel and H. M. Wadsworth, Jr., for their help.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

Chapter

I. INTRODUCTION 1

II. PREREQUISITES. 3

Notation
Basic Definitions
Definitions: Matrices
Definitions: Sets
Definitions: States
Definitions: Chains
Preliminary Theorems

III. RELATIONSHIPS BETWEEN CLASSIFICATIONS OF STATES. 12

IV. RELATIONSHIPS BETWEEN CLASSIFICATIONS OF SETS. 20

V. RELATIONSHIPS BETWEEN CLASSIFICATIONS OF CHAINS. 25

VI. SUMMARY OF RELATIONSHIPS 38

BIBLIOGRAPHY 40

CHAPTER I

INTRODUCTION

As in many other aspects of mathematics, the theory of finite Markov chains can be developed by several methods. It is the purpose of this paper to relate three such developments of the theory of finite Markov chains with discrete parameters.

F. R. Gantmacher, in his book *Applications of the Theory of Markov Chains*, studies Markov chains by using matrix theory. On the other hand, K. L. Chung, in his book *Markov Chains with Stationary Transition Probabilities*, approaches the study of Markov chains without any use of matrices; and the development of Markov chain theory by J. G. Kemeny and J. L. Snell in their book *Finite Markov Chains* depends only partially on matrix theory. The author of this paper has attempted to relate these particular approaches by interrelating the definitions of types of states, sets, and chains given in these three books.

In this thesis, Chapter II is devoted to listing basic definitions and theorems, the knowledge of which will be essential to a reader of this paper. Chapters III, IV, and V contain statements and proofs of the relationships between different classifications of states, sets, and chains, respectively. For quick reference, a summary of these relationships is listed in Chapter VI.

To indicate from which of the three books a theorem or classification is taken, the author has used the letters G, K, or C following the theorem or classification; G indicating Gantmacher's book, K indicating Kemeny's book, and C indicating Chung's book.

By relating the various definitions used in different developments of Markov chain theory, the author hopes that this paper will serve as a handy reference to anyone interested in the study of Markov chains.

CHAPTER II

PREREQUISITES

Notation

$s_i, s_j, s_k, \ldots$: states $i,j,k=0,1,2,\ldots$

p_{ij}: transition probability from state s_i to s_j.

P: transition matrix for a Markov chain.

$p_{ij}^{(n)}$: probability of being in state s_j n steps after being in state s_i. This is the ij^{th} entry of the matrix P^n. $n=0,1,2,\ldots$

$i \rightarrow j$: state s_i leads to state s_j; i.e., there exists an integer $m>0$ such that $p_{ij}^{(m)}>0$.

$i \leftrightarrow j$: state s_i leads to state s_j and state s_j leads to state s_i; also read "states s_i and s_j communicate."

d_i: the period of state s_i; i.e., the greatest common divisor of the set of positive n such that $p_{ij}^{(n)}>0$.

P^{∞}: limiting transition matrix (if it exists). $P^{\infty} = \lim_{n \to \infty} P^n$ if this limit exists.

$A \sim B$: A is B and B is A.

f_{ij}^{*}: probability of being in state s_j at least once, beginning in state s_i.

Basic Definitions

1. Equivalence class: An equivalence class of states is either a set of two or more mutually communicating states or

a set consisting of a single state.

2. Period of equivalence class: An equivalence class is said to have period d if every state in the class has period d.

3. Cyclic class: An equivalence class with period d is composed of d cyclic classes determined as follows: two states s_i and s_j belong to the same cyclic class if and only if $p_{ij}^{(kd)}>0$ for some positive integer k.

Definitions: Matrices

Definition 1M: *reducible matrix:* The n x n matrix A is reducible if there is a permutation of the indices which reduces it to the form

$$\bar{A} = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

where B and D are square matrices.

Definition 2M: *irreducible matrix:* A square matrix which is not reducible is irreducible.

Definition 3M: *primitive matrix:* Consider a non-negative, irreducible matrix with dominant characteristic root r. If there is only one characteristic root of modulus r, the matrix is called primitive.

Definition 4M: *imprimitive matrix:* Consider a non-negative, irreducible matrix with dominant characteristic root r. If there is more than one characteristic root of modulus r, the matrix is called imprimitive.

Definitions: Sets

Definition 1C: *closed set:* A set of states, A, is closed if and only if $\sum_{s_j \varepsilon A} p_{ij} = 1$ for all $s_i \varepsilon A$.

Definition 2C: *minimal set:* A minimal set is a closed set which contains no proper subset that is closed.

Definition 1K: *ergodic set:* The ergodic sets are the minimal elements of the partial ordering of equivalence classes. (Once a process enters an ergodic set, it can never leave it.)

Definition 2K: *transient set:* Any element in the partial ordering of equivalence classes which is not an ergodic set is a transient set. (Once a process leaves a transient set, it can never return.)

Definition 3K: *regular set:* An equivalence class with only one cyclic class is a regular set. (An equivalence class with period one is a regular set.)

Definition 4K: *cyclic set:* An equivalence class with more than one cyclic class is a cyclic set. (An equivalence class with period two or more is a cyclic set.)

Definition 5K: *open set:* A set S of states is open if from every state in S, it is possible to go to a state which is not in S.

Definitions: States

Definition 3C: *essential state*(C): A state s_i is essential (C) if and only if $i \to j$ implies $j \to i$ for every state s_j.

Definition 4C: *inessential state*(C): A state s_i is inessential (C) if and only if there exists a state s_j such that $i \to j$, but not $j \to i$.

Definition 5C: *recurrent state:* A state s_i is recurrent if $f_{ii}^* = 1$.

Definition 6C: *nonrecurrent state:* A state s_i is nonrecurrent if $f_{ii}^* < 1$.

Definition 7C: *positive state:* A recurrent state s_i is positive if $\lim_{n \to \infty} p_{ii}^{(nd_i)} > 0$.

Definition 8C: *null state:* A recurrent state s_i is null if $\lim_{n \to \infty} p_{ii}^{(nd_i)} = 0$.

Definition 6K: *ergodic state:* A state s_i is ergodic if it is a member of an ergodic set.

Definition 7K: *transient state:* A state s_i is transient if it is a member of a transient set.

Definition 8K: *absorbing state:* The state contained in an ergodic set containing only one state is an absorbing state.

Definition 1G: *essential state*(G): Consider the normal form of the transition matrix P for a proper Markov chain with n states (see theorem 4G). Partition the states $s_1, \ldots, s_n$ of the chain into groups $\Sigma_1, \Sigma_2, \ldots, \Sigma_s$, the number of states in group Σ_i being equal to the number of rows in the block Q_i, $i=1,2,\ldots,s$. The essential states (G) are the states of the various groups $\Sigma_1, \ldots, \Sigma_g$.

Definition 2G: *inessential state* (G): Under the same considerations as in Definition 1G, the states of the groups $\Sigma_{g+1}, \ldots, \Sigma_s$ are called inessential states (G).

<u>Definitions: Chains</u>

Definition 9C: *indecomposable chain:* A chain is indecomposable if the set of all states of the chain does not contain two disjoint closed sets.

Definition 10C: *irreducible chain*(C): A chain is irreducible if the set of all states of the chain forms a single class (with period one).

Definition 9K: *ergodic chain:* A chain consisting of a single ergodic set is an ergodic chain.

Definition 10K: *regular chain*(K): A regular chain (K) is an ergodic chain with the ergodic set regular.

Definition 11K: *cyclic chain:* A cyclic chain is an ergodic chain with the ergodic set cyclic.

Definition 12K: *absorbing chain:* A chain with transient sets and with all ergodic sets unit sets is an absorbing chain.

Definition 13K: *chain type I:* A chain with no transient sets and with more than one ergodic set is called a chain type I.

Definition 14K: *chain type II:* A chain with transient sets and with all ergodic sets regular, but not all unit sets, is called a chain type II.

Definition 15K: *chain type III:* A chain with transient sets and with all ergodic sets cyclic is called a chain type III.

Definition 16K: *chain type IV:* A chain with transient sets and with both cyclic and regular ergodic sets is called a chain type IV.

Definition 3G: *proper chain:* A chain with corresponding transition matrix having no characteristic roots ($\neq 1$) of modulus 1 is a proper chain.

Definition 4G: *regular chain*(G): A proper chain with corresponding transition matrix having 1 as a simple characteristic root is a regular chain (G).

Definition 5G: *irreducible chain*(G): A chain with an irreducible transition matrix is called irreducible (G).

Definition 6G: *reducible chain:* A chain with a reducible transition matrix is called reducible.

Definition 7G: *aperiodic chain:* A chain with a primitive transition matrix is called aperiodic.

Definition 8G: *periodic chain:* A chain with an imprimitive transition matrix is called periodic.

Preliminary Theorems

The following theorems will be stated without proof, as their proofs are given on the pages of the books indicated. These theorems will be used throughout this paper.

Theorem 1C. An essential class is minimal closed. The union of a finite number of inessential classes is not closed. [p. 14(C)].

Theorem 2C. An inessential state is nonrecurrent. [p. 18(C)].

Theorem 3C. If s_i is recurrent and $i \to j$, then s_j is recurrent. [p. 19(C)].

Theorem 4C. The state s_i is recurrent or nonrecurrent according as the series $\sum_{n=o}^{\infty} P_{ii}^{(n)}$ diverges or converges. [p. 22(C)].

Theorem 1G. A necessary and sufficient condition that a homogeneous Markov chain possess limiting transition probabilities between every pair of states is that the chain be proper. [p. 112(G)].

Theorem 2G. A necessary and sufficient condition that the limiting transition probabilities of a proper homogeneous Markov chain be independent of the initial state is that the chain be regular. [p. 112(G)].

Theorem 3G. A necessary and sufficient condition that the limiting transition probabilities of a homogeneous Markov chain all be positive is that the chain be aperiodic. [p. 112(G)].

<u>Theorem 4G</u>. Let P be a stochastic transition matrix for a proper Markov chain. The normal form of P is

$$P = \begin{bmatrix} Q_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & Q_g & 0 & \cdots & 0 \\ U_{g+1,1} & \cdots & U_{g+1,g} & Q_{g+1} & & \vdots \\ \vdots & & \vdots & & \ddots & \vdots \\ U_{s,1} & \cdots & U_{s,s} & \cdots & U_{s,s-1} & Q_s \end{bmatrix}$$

where each of $Q_1,\ldots,Q_g$ is a primitive stochastic matrix, and where each of the irreducible matrices $Q_{g+1},\ldots,Q_s$ has dominant characteristic root less than 1. [p. 109(G)].

<u>Theorem 1K</u>. If P is the transition matrix for a Markov chain, then the canonical form of P is:

$$P = \begin{array}{c} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_k \end{array} \left(\begin{array}{c|c|c|c|c} P_1 & & & & \\ \hline R_2 & P_2 & & & 0 \\ \hline \multicolumn{2}{c|}{R_3} & P_3 & & \\ & \vdots & & \ddots & \\ & R_k & & & P_k \end{array} \right)$$

where $u_1,u_2,\ldots,u_k$ are equivalence classes with the minimal sets enumerated first, the sets one level above the minimal sets enumerated next, etc.; P_i, $i=1,\ldots,k$, are transition matrices within an equivalence class; the region 0 consists entirely of zeroes; and R_i, i=1,...,k, will be entirely zero if P_i is an ergodic set, but will have non-zero elements otherwise. [p. 36(k)].

Theorem 2K. A set S of states is open if and only if no ergodic set is a subset of S. [p. 59(K)].

Theorem 3K. A transition matrix is regular if and only if for some N, P^N has no zero entries. [p. 69(K)].

Theorem 4K. If P is a regular transition matrix, then

a. The powers P^n approach a probability matrix P^∞.

b. Each row of P^∞ is the same probability vector
$\alpha = \{a_1, a_2, \ldots, a_n\}$.

c. The components of α are positive.

d. For any probability vector π, $\pi \cdot P^n$ approaches the vector α as n approaches ∞ . [p. 70, 71(K)].

Theorem 5K. If P is the transition matrix for a cyclic chain, then P^n cannot converge. [p. 99(K)].

Theorem 6K. 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. [p. 43(K)].

CHAPTER III

RELATIONSHIPS BETWEEN CLASSIFICATIONS OF STATES

Theorem 2.1. An essential state (C) $\sim$ ergodic state (K).

Proof.

(i) Let state s_i be an essential state (C). Then, by definition, $i \to j$ implies that $j \to i$ for every state s_j.

Suppose s_i is not an ergodic state (K); i.e, that s_i is a transient state (K).

From Theorem 6K, it can be seen that a Markov process always moves toward the ergodic sets; and once a process leaves a transient set, it can never return. Thus, there exists a state s_j in an ergodic set such that $i \to j$, but not $j \to i$.

But this implies that s_i is an inessential state (C), giving a contradiction.

Therefore, s_i is an ergodic state (K).

(ii) Now, let s_i be an ergodic state (K). Consider state $s_j \not\sim s_i$.

If s_j is in the same ergodic set as s_i, then $i \to j$ implies that $j \to i$, by property of belonging to the same ergodic set.

If s_j is not in the same ergodic set as s_i, then $i \not\to j$, since once a process enters an ergodic set, it can never leave it.

Thus, $i \to j$ implies $j \to i$ for every state s_j, which shows that s_i is an essential state (C).

Therefore, by (i) and (ii), an essential state (C) $\sim$ ergodic state (K).

Theorem 2.2. An essential state (G) $\sim$ ergodic state (K).

Proof.

(i) Let state s_i be an essential state (G).

Then, by definition, s_i belongs to one of the groups $\Sigma_1, \Sigma_2, \ldots, \Sigma_g$ mentioned in Definition 1G; say, s_i belongs to Σ_j.

By comparing the normal form of a transition matrix P given in Theorem 4G with the canonical form of the matrix given in Theorem 1K, one can easily see that the normal form of P is precisely the canonical form of P where the groups $\Sigma_1, \ldots, \Sigma_g$ associated with $Q_1, \ldots, Q_g$ of the normal form are the ergodic sets of P.

Then, since s_i belongs to Σ_j, which is the group of states associated with Q_j, s_i is an ergodic state (K); i.e., s_i belongs to the ergodic set Σ_j.

(ii) Let state s_i be an ergodic state (K).

Then s_i belongs to an ergodic set, and this ergodic set will be one of the $\Sigma_1, \ldots, \Sigma_g$ mentioned in (i).

Thus, s_i belongs to one of the $\Sigma_1, \ldots, \Sigma_g$, which implies that s_i is an essential state (G).

Therefore, by (i) and (ii), an essential state (G) $\sim$ ergodic state (K).

Corollary 2.3. An essential state (G) $\sim$ essential state (C) $\sim$ ergodic state (K).

Proof. This is a direct consequence of Theorems 2.1 and 2.2.

Theorem 2.4. An inessential state (C) $\sim$ transient state (K).

Proof.

(i) Let s_i be an inessential state (C).

Then s_i is not an essential state (C), which implies that s_i is not an ergodic state (K) by Theorem 2.1.

Thus, s_i is a transient state (K), since every state is either transient (K) or ergodic (K).

(ii) Let s_i be a transient state (K).

Then s_i is not an ergodic state (K), which implies that s_i is not an essential state (C) by Theorem 2.1.

Thus, s_i is an inessential state (C), since every state is either essential (C) or inessential (C).

Therefore, by (i) and (ii), an inessential state (C) $\sim$ transient state (K).

<u>Theorem 2.5</u>. An inessential state (G) is an inessential state (C).

Proof. Let s_i be an inessential state (G).

Then s_i is not an essential state (G), which implies that s_i is not an essential state (C) by Corollary 2.3.

Thus, s_i is an inessential state (C), since every state is either essential (C) or inessential (C).

Note: An inessential state (C) is not necessarily an inessential state (G), since an inessential state (G) is only defined for a proper Markov chain; i.e., it is <u>not</u> true that every state is either essential (G) or inessential (G).

<u>Corollary 2.6</u>. An inessential state (G) is a transient state (K).

Proof. This is a direct consequence of Corollary 2.3 and Theorem 2.5.

Theorem 2.7. A transient state (K) is a nonrecurrent state (C).

Proof. Let s_i be a transient state (K).

Then s_i is an inessential state (C) by Theorem 2.4.

But this implies, by use of Theorem 2C, that s_i is a nonrecurrent state (C).

Theorem 2.8. An ergodic state (K) $\sim$ recurrent state (C).

Proof.

(i) Let s_i be a recurrent state (C).

Then s_i is not a nonrecurrent state (C), and hence not a transient state (K) by Theorem 2.7.

Thus, s_i is an ergodic state since every state is either transient (K) or ergodic (K).

Therefore, a recurrent state (C) is an ergodic state (K).

(ii) Let s_i be an ergodic state (K).

Then s_i belongs to an ergodic set (K).

Let P be the transition matrix for the chain consisting of this set.

Suppose that the ergodic set is cyclic (K).

Then, by Theorem 5K, the sequence $P_{ii}^{(0)}, P_{ii}^{(1)}, \ldots, P_{ii}^{(n)}, \ldots$ diverges, which implies that $\sum_{n=0}^{\infty} P_{ii}^{(n)}$ diverges (otherwise, $\lim_{n \to \infty} P_{ii}^{(n)} = 0$, contradicting the convergence of the sequence $P_{ii}^{(n)}$).

Thus, if s_i is a cyclic state (K), $\sum_{n=0}^{\infty} P_{ii}^{(n)}$ diverges.

Suppose that the ergodic set is regular (K).

Then, by Theorem 4K, the powers P^n approach a probability matrix P^∞, each row of P^∞ being the same probability vector $\alpha = \{a_1, a_2, \ldots, a_m\}$, and each $a_j, j=1,\ldots,m$, being positive; i.e., $\lim_{n \to \infty} P_{ii}^{(n)} = K > 0$ where K is one of the $a_j, j=1,\ldots,m$.

Thus, $\sum_{n=0}^{\infty} P_{ii}^{(n)}$ diverges by the same reasoning as in the first part of proof.

Therefore, we have shown that if s_i is an ergodic state (K), $\sum_{n=0}^{\infty} P_{ii}^{(n)}$ diverges.

But, by Theorem 4C, $\sum_{n=0}^{\infty} P_{ii}^{(n)}$ diverging implies that s_i is a recurrent state (C).

Thus, an ergodic state (K) is a recurrent state (C).

Therefore, by (i) and (ii), an ergodic state (K) $\sim$ recurrent state (C).

Corollary 2.9. A recurrent state (C) $\sim$ essential state (C) $\sim$ essential state (G) $\sim$ ergodic state (K).

Proof. This is a direct consequence of Corollary 2.3 and Theorem 2.8.

Theorem 2.10. A nonrecurrent state (C) $\sim$ transient state (K).

Proof. Let s_i be a nonrecurrent state (C).

Then s_i is not a recurrent state (C), which implies that s_i is not an ergodic state (K) by Theorem 2.8.

Thus, s_i is a transient state (K), since every state is either ergodic (K) or transient (K).

Therefore, a nonrecurrent state (C) is a transient state (K).

But, by Theorem 2.7, a transient state (K) is a nonrecurrent state (C).

Thus, a transient state (K) ∿ nonrecurrent state (C).

Corollary 2.11: A non-recurrent state (C) ∿ inessential state (C) ∿ transient state (K).

Proof. This is a direct consequence of Theorems 2.4 and 2.10.

Corollary 2.12. An absorbing state (K) is an essential state (C), and hence is an essential state (G), and hence is a recurrent state (C).

Proof. Let s_i be an absorbing state (K).

Then, by definition, s_i is an ergodic state.

The three conclusions of the Corollary follow directly from Corollary 2.9.

Example 1. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

State s_1 is ergodic (K) (since it belongs to the ergodic set $\{s_1, s_2\}$), and hence s_1 is essential (C), essential (G), and recurrent (C); however s_1 is not absorbing (K).

This example shows, for instance, that an essential state (C) is not necessarily an absorbing state (K).

Theorem 2.13. An ergodic state (K) with period one is a positive state (C).

Proof. Let s_i be an ergodic state (K) with period d_i=1.

Then s_i belongs to a regular set (K).

Let P be the transition matrix for the chain consisting of this set, and let this set contain m states.

Then, by Theorem 4K, the powers P^n approach a probability matrix P^∞, each row of P^∞ being the same probability vector $\alpha = \{a_1, a_2, \ldots, a_m\}$, and each $a_j, j=1,\ldots,m$, being positive; i.e., $\lim_{n \to \infty} P_{ii}^{(n)} = K > 0$ where K is one of the $a_j, j=1,\ldots,m$.

Since $d_i = 1$, $\lim_{n \to \infty} P_{ii}^{(nd_i)} = \lim_{n \to \infty} P_{ii}^{(n)} = K > 0$.

s_i is recurrent (C) since it is ergodic (K).

Therefore, by Definition 7C, s_i is a positive state (C).

Thus, an ergodic state (K) with period 1 is a positive state (C).

Example 2. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

The period of state s_1 is two; i.e., $d_1 = 2$.

But $\lim_{n \to \infty} P_{11}^{(nd_1)} = \lim_{n \to \infty} P_{11}^{(2n)} = 1/2$ as can be seen from the following:

$$P^2 = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

from which it is obvious that

$$\lim_{n \to \infty} P^{2n} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

which implies that $\lim_{n \to \infty} P_{11}^{(2n)} = 1/2$.

State s_1 is certainly recurrent (C) since it is ergodic (K) (it is a member of the ergodic set $\{s_1,s_2,s_3\}$).

Thus, s_1 is a positive state (C).

This example shows that a positive state (C) is not necessarily an ergodic state (K) with period one.

Theorem 2.14. A positive state (C) is an ergodic state (K), and hence is an essential state (G).

Proof. A positive state (C) is necessarily a recurrent state (C) by definition, and thus is an ergodic state (K) and an essential state (G) by Corollary 2.9.

CHAPTER IV

RELATIONSHIPS BETWEEN CLASSIFICATIONS OF SETS

Theorem 3.1. An ergodic set (K) $\sim$ minimal set (C).

Proof.

(i) Let D be an ergodic set (K).

Then every member of D is an ergodic state, and for any two states $s_i, s_j \in D$, $i \to j$ and $j \to i$.

But since an ergodic state (K) is an essential state (C), every state in D is also an essential state (C); and since $i \to j$ and $j \to i$ for any two states s_i, $s_j \in D$, D is a minimal class by definition.

By Theorem 1C, an essential class (C) is minimal (C), implying that D is a minimal set (C).

Therefore, an ergodic set (K) is a minimal set (C).

(ii) Let D be a minimal set (C).

Then, by definition, D is closed (i.e., $\sum_{s_j \in D} P_{ij} = 1$ for all $s_i \in D$), and D contains no proper subset that is closed.

(a) First, we will show that D cannot be a transient set or a finite number of transient sets.

By Theorem 1C, the union of a finite number of inessential classes (C) is not closed.

An inessential class is a transient set, for an inessential class is composed of inessential states, and any two of these states s_i, s_j are such that $i \to j$ and $j \to i$, implying that an inessential class is composed of transient states; and any two of these states s_i, s_j are such that $i \rightarrow j$ and $j \rightarrow i$, implying that an inessential class is a transient set.

Thus, Theorem 1C says that the union of a finite number of transient sets (K) is not closed.

Hence, D cannot be a transient set or a union of a finite number of transient sets, since D is closed.

(b) Next, we will show that D cannot strictly contain an ergodic set.

By part (i) of this proof, an ergodic set (K) is minimal (C) and hence is closed (C). Since D contains no proper subset that is closed, D cannot strictly contain an ergodic set.

(c) Finally, we will show that D cannot contain only part of a transient set or part of an ergodic set. Suppose D contains part of a set E, where E is either transient or ergodic. Then there exists a $p_{kl} > 0$ for a state $s_k \in E \cap D$ and a state $s_l \in E-D$.

Then $\sum_{s_j \in D} p_{kj} < 1$ since $p_{kl} > 0$, which implies that

$\sum_{s_j \in D} p_{ij} \neq 1$ for at least one $s_i \in D$; i.e., that D is <u>not</u> closed. However, this gives a contradiction since D is closed. Thus, D does not contain only part of a transient set or part of an ergodic set. Therefore, from (a), (b), and (c), we see that D must be an ergodic set (K).

Thus, a minimal set (C) is an ergodic set (K).

Therefore, by (i) and (ii), an ergodic set (K) $\sim$ minimal set (C).

<u>Corollary 3.2</u>. An ergodic set (K) is a closed set (C).

Proof. Let D be an ergodic set (K).

Then D is minimal (C), which implies that D is closed (C), since a minimal set (C) is a closed set (C).

Therefore, an ergodic set (K) is a closed set (C).

<u>Example 3</u>. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array}$$

It is easily verified that the set $\{s_1,s_2\}$ is closed (C), but is not ergodic (K).

This example shows that a closed set (C) is not necessarily an ergodic set (K).

<u>Theorem 3.3</u>. A closed set (C) is not an open set (K).

Proof. Let S be a closed set (C).

Let $\bar{S}$ be composed of all those states not contained in S.

Then for state $s_k \varepsilon \bar{S}$, $p_{ik} = 0$ for any state $s_i \varepsilon S$, since by definition, $\sum_{s_j \varepsilon S} p_{ij}=1$ for all $s_i \varepsilon S$.

Thus, it is not possible to go from any state in S to a state in $\bar{S}$, and hence, by definition, S is not an open set (K).

Therefore, a closed set (C) is not an open set (K).

Theorem 3.4. An open set (K) is not a closed set (C).

Proof. Let S be an open set (K).

Suppose S is also a closed set (C).

Then, by Theorem 3.3, S is not an open set (K), which contradicts our hypothesis.

Therefore, S is not a closed set (C).

Thus, an open set (K) is not a closed set (C).

Example 4. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}$$

Consider the set $S = \{s_1, s_2\}$.

S is not an open set (K), since, for instance, it is impossible to go from state s_1 in S to state s_3 not in S.

S is not a closed set (C), since, for instance, $p_{21} + p_{22} = 1/3 + 1/3 = 2/3 < 1$.

This example shows that a set which is not open (K) is not necessarily closed (C); and that a set which is not closed (C) is not necessarily open (K).

Theorem 3.5. A transient set (K) is not a closed set (C).

Proof. Let S be a transient set (K).

Then S is clearly an open set (K), since a Markov process tends to the ergodic sets (K) (see Theorem 6K), implying that it is possible to go from a state in S to a state not in S.

Therefore, by Theorem 3.4, S is not a closed set (C).

Hence, a transient set (K) is not a closed set (C).

CHAPTER V

RELATIONSHIPS BETWEEN CLASSIFICATIONS OF CHAINS

Theorem 4.1. A regular chain (K) is a regular chain (G).

Proof. Let P be the transition matrix for a regular chain (K). Then, by Theorem 4K, the limiting matrix P^{∞} is independent of the initial state.

But, by Theorem 2G, this is a sufficient condition to guarantee that P is the transition matrix for a regular chain (G).

Thus, a regular chain (K) is a regular chain (G).

Example 5. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \end{array}$$

In this case,

$$P^{\infty} = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \end{array}$$

The characteristic roots of P are 1, 1/2.

Hence, P represents a regular chain (G), since 1 is a simple characteristic root of P; however, P does not represent a regular chain (K), since the chain represented by P has a transient set $\{s_2\}$.

This example shows that a regular chain (G) is not necessarily a regular chain (K).

Corollary 4.2. A regular chain (K) is a proper chain (G).

Proof. This follows immediately from Theorem 4.1, and the fact that a regular chain (G) is a proper chain (G).

Theorem 4.3. A regular chain (K) $\sim$ aperiodic chain (G).

Proof.

(i) Let P be the transition matrix for a regular chain (K).

Then, by Theorem 3K, for some positive interger N, P^N has no zero entries, which implies that P^∞ is positive (has all positive entries).

It follows, from Theorem 3G, that P is the transition matrix for an aperiodic chain (G).

Thus, a regular chain (K) is an aperiodic chain (G).

(ii) Let P be the transition matrix for an aperiodic chain (G).

Then P^∞ is positive, by Theorem 3G, implying that P^N is positive for some positive integer N.

It follows, from Theorem 3K, that P is the transition matrix for a regular chain (K).

Thus, an aperiodic chain (G) is a regular chain (K).

Therefore, by (i) and (ii), a regular chain (K) $\sim$ aperiodic chain (G).

<u>Theorem 4.4</u>. Any chain containing a transient set (K) is reducible (G).

Proof. Let P be the transition matrix for a chain with n states, m of which are transient ($m > 0$).

Certainly $m < n$, since every chain must contain at least one ergodic state.

Without loss of generality, assume that $s_1,\ldots,s_{n-m}$ are the ergodic states and $s_{n-m+1},\ldots,s_n$ are the transient states.

Then there exists a permutation of the indices such that P becomes

$$\bar{P} = \begin{array}{c|c|c} & s_1 \quad s_2 \ldots s_{n-m} & s_{n-m+1} \ldots s_n \\ \hline s_1, s_2, \ldots, s_{n-m} & B & A \\ \hline s_{n-m+1}, s_{n-m+2}, \ldots, s_n & C & D \end{array}$$

Clearly, [B] is an $n-m \times n-m$ square matrix, and [D] is an $m \times m$ square matrix.

Also, the matrix [A] = [0] since, in any Markov chain, an ergodic state cannot lead to a transient state.

Thus, by a permutation of the indices, P can be put in the form $\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$, where B and D are square matrices. This implies that P is reducible (G), by definition.

Therefore, any chain containing a transient set (K) is reducible (G).

Corollary 4.5. Absorbing chains (K) and chains of types II, III, and IV (K) are all reducible (G).

Proof. Since absorbing chains (K) and chains of types II, III, and IV (K) all contain at least one transient set (K), the result is immediate from Theorem 4.4.

Theorem 4.6. A chain with more than one ergodic set (K) is reducible (G).

Proof. If the chain contains any transient sets (K), then it is reducible (G) by Theorem 4.4.

Let P be the transition matrix for a chain with n states, with no transient sets, but with more than one ergodic set.

Without loss of generality, assume that $s_1, s_2, \ldots s_m$ represent the states in one of the ergodic sets, and that $s_{m+1}, \ldots, s_n$ represent the states in the remaining ergodic sets.

Then, there exists a permutation of the indices of P such that P becomes

$$\bar{P} = \begin{array}{c} s_1 \\ \vdots \\ s_m \\ s_{m+1} \\ \vdots \\ s_n \end{array} \overset{\begin{array}{cc} s_1 \ldots\ldots s_m & s_{m+1} \ldots\ldots s_n \end{array}}{\left[\begin{array}{c|c} B & A \\ \hline C & D \end{array}\right]}$$

Clearly, [B] is an m x m square matrix, and [D] is an n-m x n-m square matrix.

Also, the matrix [A] = [0] since an ergodic set has the property that once a process enters it, the process can never leave it; i.e., a state in an ergodic set cannot lead to a state in another ergodic set.

Thus, by a permutation of the indices, P can be put in the form $\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$ where B and D are square matrices.

This implies that P is reducible (G), by definition.

Therefore, any chain containing more than one ergodic set (K) is reducible (G).

<u>Theorem 4.7.</u> An ergodic chain (K) $\sim$ irreducible chain (G).

Proof.

(i) Let P be the transition matrix for an ergodic chain with n states.

Suppose the ergodic chain is reducible (G).

Then, by a permutation of the indices of P, P becomes $\bar{P} = \begin{pmatrix} B & 0 \\ C & D \end{pmatrix}$ where B and D are square matrices.

Let [B] be an m x m matrix, m < n; then [D] is an n-m x n-m matrix.

Then, by a renumeration of the states if necessary,

$$\bar{P} = \begin{array}{c} s_1 \\ \vdots \\ s_m \\ s_{m+1} \\ \vdots \\ s_n \end{array} \overset{\begin{array}{cc} s_1 \ldots s_m & s_{m+1} \ldots s_n \end{array}}{\left(\begin{array}{c|c} B & 0 \\ \hline C & D \end{array} \right)}$$

But then it can be seen that $\bar{P}$ represents a chain for which it is impossible for one of the states $s_1, \ldots, s_m$ to lead to any of the states $s_{m+1}, \ldots, s_n$; however, this gives a contradiction since an ergodic chain consists of a single ergodic set, implying that every state in the chain can lead to every other state in the chain.

Thus, P is the transition matrix for an irreducible chain (G), and hence, an ergodic chain (K) is an irreducible chain (G).

(ii) Let P be the transition matrix for an irreducible chain (G).

P is the transition matrix for one or more of an ergodic chain (K), an absorbing chain (K), a chain type I (K), a chain type II (K), a chain type III (K), and a chain type IV (K), since this listing categorizes every chain.

But P is not the transition matrix for an absorbing chain (K), or a chain of type I, II, III, or IV (K), since all of these chains are reducible (G) by Corollary 4.5 and Theorem 4.6.

Thus, P must be the transition matrix for an ergodic chain (K).

Therefore, by (i) and (ii), an ergodic chain (K) $\sim$ irreducible chain (G).

Theorem 4.8. A cyclic chain (K) $\sim$ periodic chain (G).

Proof.

(i) A cyclic chain (K) is an ergodic chain (K), and hence is an irreducible chain (G) by Theorem 4.7.

An irreducible chain (G) is either periodic (G) or aperiodic (G), since the transition matrix for an irreducible chain is either primitive or imprimitive, and an imprimitive matrix represents a periodic chain (G) and a primitive matrix represents an aperiodic chain (G) (see Definitions 3M, 4M, 7G, and 8G).

Thus, a cyclic chain (K) is either periodic (G) or aperiodic (G).

But a cyclic chain (K) cannot be aperiodic (G), since, by Theorem 4.3, an aperiodic chain (G) is regular (K).

Hence, a cyclic chain (K) is a periodic chain (G).

(ii) A periodic chain (G) is irreducible (G), since a periodic chain (G) is defined in terms of an imprimitive matrix which is only defined for an irreducible chain (G).

Thus, a periodic chain (G) is an ergodic chain (K) by Theorem 4.7, and hence is either a regular chain (K) or a cyclic chain (K).

But a periodic chain (G) cannot be regular (K), since, by Theorem 4.3, a regular chain (K) is aperiodic (G).

Hence, a periodic chain (G) is a cyclic chain (K).

Therefore, by (i) and (ii), a cyclic chain (K) $\sim$ periodic chain (G).

Theorem 4.9. A cyclic chain (K) is not a proper chain (G), and a proper chain (G) is not a cyclic chain (K).

Proof.

Let P be the transition matrix for a cyclic chain (K).

Then P^{∞} does not exist, since, by Theorem 5K, P^{n} does not converge.

Thus, the chain is not proper (G) by Theorem 1G.

Hence, a cyclic chain (K) is not proper (G).

If a proper chain (G) could be a cyclic chain (K), we would contradict what we just proved—that every cyclic chain (K) is not proper (G).

Hence, a proper chain (G) is not a cyclic chain (K).

We have shown in Theorems 4.1 and 4.9 that a regular chain (K) is a regular chain (G), and that a cyclic chain (K) is not a proper chain (G) and hence not a regular chain (G).

We will now give examples to show that an absorbing chain (K) may be regular (G) or not regular (G).

Example 6. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

The characteristic roots of P are 1, $\frac{1 + \sqrt{3}}{4}$, and $\frac{1 - \sqrt{3}}{4}$.

Hence P represents a regular chain (G), since 1 is a simple characteristic root of P.

P also represents an absorbing chain (K), since the chain has a transient set $\{s_2,s_3\}$, and its only ergodic set $\{s_1\}$ is a unit set.

This is an example of an absorbing chain (K) which is a regular chain (G).

Example 7. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \end{array}$$

The characteristic roots of P are 1, 1, and 1/2.

Hence P does not represent a regular chain (G), since 1 is not a simple characteristic root of P.

But P does represent an absorbing chain (K), since the chain has a transient set $\{s_3\}$, and its only ergodic sets, $\{s_2\}$ and $\{s_1\}$, are unit sets.

This is an example of an absorbing chain (K) which is not a regular chain (G).

Theorem 4.10. A chain containing only one ergodic set (K) ∿ an indecomposable chain (C).

Proof.

(i) First, we will prove that a chain containing only one ergodic set (K) is an indecomposable chain (C).

Suppose there exists a chain containing only one ergodic set (K) which is not indecomposable (C). Let this chain consist of n states s_1, $s_2, \ldots s_n$.

Then the set of n states contains at least two disjoint closed sets, say (by a renumeration of the indices, if necessary) $A = \{s_1, s_2, \ldots, s_m\}$ and $B = \{s_{m+1}, s_{m+2}, \ldots, s_q\}$, $q \leq n$.

Then, by definition of a closed set, once the process enters into set A, it can never leave it.

Hence, set A must contain the only ergodic set, since every Markov process must eventually lead to an ergodic set.

But, by a similar argument, set B must contain the only ergodic set, which is impossible since sets A and B are disjoint.

Thus, a contradiction arises implying that a chain containing only one ergodic set (K) is an indecomposable chain (C).

(ii) Next, we will show that an indecomposable chain (C) is a chain containing only one ergodic set (K) by proving that a chain with more than one ergodic set (K) is not indecomposable (C).

By Corollary 3.2, an ergodic set (K) is a closed set (C).

Any two ergodic sets in the same chain are disjoint, since once a process enters an ergodic set, it can never leave it.

Hence, any chain with at least two ergodic sets (K) contains at least two disjoint closed sets.

Thus, by definition, any chain with more than one ergodic set (K) is not indecomposable.

Therefore, by (i) and (ii), a chain containing only one ergodic set (K) $\sim$ an indecomposable chain (C).

Corollary 4.11. An ergodic chain (K) is an indecomposable chain (C).

Proof. This follows from the fact that an ergodic chain contains only one ergodic set (K), and by use of Theorem 4.10.

Theorem 4.12. An irreducible chain (G) is an indecomposable chain (C).

Proof. An irreducible chain (G) is an ergodic chain (K) by Theorum 4.7; and an ergodic chain (K) is an indecomposable chain (C) by Corollary 4.11.

Hence, it follows directly that an irreducible chain (G) is an indecomposable chain (C).

Corollary 4.13. An aperiodic chain (G) is an indecomposable chain (C). A periodic chain (G) is an indecomposable chain (C).

Proof. This follows directly from Theorem 4.12, since both an aperiodic chain (G) and a periodic chain (G) are irreducible chains (G).

Example 8. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \end{array}$$

P represents an indecomposable chain (C), since the set $\{s_1,s_2\}$ does not contain two disjoint closed sets.

However, P does not represent an ergodic chain (K), since $\{s_2\}$ is a transient set. (An ergodic chain contains no transient sets.)

This is an example of an indecomposable chain (C) which is not an ergodic chain (K), and hence not an irreducible chain (G), and hence neither an aperiodic chain (G) nor a periodic chain (G).

Theorem 4.14. An irreducible chain (C) $\sim$ regular chain (K).

Proof. By definition, an irreducible chain (C) consists of only one class (set), each state in this set having period one.

This set must be ergodic (K), since every chain must contain an ergodic set.

Furthermore, this ergodic set is regular (K), since every state in the set has period one (i.e, this set contains only one cyclic class).

Hence, an irreducible chain (C) consists of only one regular set (K).

But this is precisely the definition of a regular chain (K).

Therefore, an irreducible chain (C) $\sim$ regular chain (K), since both have the same definition.

Corollary 4.15. An irreducible chain (C) $\sim$ aperiodic chain (G).

Proof. This follows directly from Theorem 4.14, since, as previously proved, an aperiodic chain (G) $\sim$ regular chain (K).

Theorem 4.16. An irreducible chain (C) is an irreducible chain (G).

Proof. An aperiodic chain (G) is an irreducible chain (G), since an aperiodic chain is defined in terms of a primitive matrix which is only defined for an irreducible chain (G).

Hence, by Corollary 4.15, an irreducible chain (C) is an irreducible chain (G).

Example 9. Consider the chain represented by the following transition matrix:

$$P = \begin{array}{c} \\ s_1 \\ s_2 \end{array} \begin{array}{c} \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

P represents an irreducible chain (G), since the matrix P is an irreducible matrix.

P does not represent an irreducible chain (C), since the set of all states $\{s_1,s_2\}$ has period equal to two; i.e., the states s_1 and s_2 each have period two.

This example shows that an irreducible chain (G) is not necessarily an irreducible chain (C).

CHAPTER VI

SUMMARY OF RELATIONSHIPS

A. States

1. Essential state (G) ∿ essential state (C) ∿ ergodic state (K) ∿ recurrent state (C).
2. Inessential state (G) is an inessential state (C) ∿ transient state (K) ∿ nonrecurrent state (C).
3. Absorbing state (K) is an essential state (C) ∿ essential state (G) ∿ recurrent state (C).
4. Ergodic state (K) with period one is a positive state (C).
5. Positive state (C) is an ergodic state (K) ∿ essential state (G).

B. Sets

1. Ergodic set (K) ∿ minimal set (C).
2. Ergodic set (K) is a closed set (C).
3. Closed set (C) is not an open set (K).
4. Open set (K) is not a closed set (C).
5. Transient set (K) is not a closed set (C).

C. Chains

1. Regular chain (K) ∿ aperiodic chain (G).
2. Regular chain (K) is a regular chain (G).
3. Regular chain (K) is a proper chain (G).
4. Ergodic chain (K) ∿ irreducible chain (G).

5. Cyclic chain (K) $\sim$ periodic chain (G).
6. Cyclic chain (K) is not a proper chain (G).
7. Proper chain (G) is not a cyclic chain (K).
8. Any chain containing a transient set (K) is a reducible chain (G).
9. Absorbing chain (K) and chains of Types II, III, and IV (K) are all reducible chains (G).
10. Chain with more than one ergodic set (K) is a reducible chain (G).
11. Chain with only one ergodic set (K) $\sim$ indecomposable chain (C).
12. Ergodic chain (K) is an indecomposable chain (C).
13. Irreducible chain (G) is an indecomposable chain (C).
14. Aperiodic chain (G) is an indecomposable chain (C).
15. Periodic chain (G) is an indecomposable chain (C).
16. Regular chain (K) $\sim$ irreducible chain (C).
17. Aperiodic chain (G) $\sim$ irreducible chain (C).
18. Irreducible chain (C) is an irreducible chain (G).

BIBLIOGRAPHY

1. Chung, K. L., *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag Ohg. Berlin·Gottingen·Heidelberg, 1960.

2. Dartmouth College Writing Group, *Modern Mathematical Methods and Models*, volume 2, Cushing-Malloy, Inc., Ann Arbor, Michigan, 1958.

3. Gantmacher, F. R., *Applications of the Theory of Matrices*, Interscience Publishers Inc., New York, 1959.

4. Gantmacher, F. R., *The Theory of Matrices*, volume 2, Chelsea Publishing Company, New York, 1959.

5. Kemeny, J. G. and Snell, J. L., *Finite Markov Chains*, D. Van Nostrand Co., Inc., Princeton, New Jersey, 1960.

6. Parzen, Emanuel, *Stochastic Processes*, Holden-Day, Inc., San Francisco, California, 1962.