SOME PROBLEMS IN THE MATHEMATICAL THEORY OF PROBABILITY

SOME PROBLEMS IN THE MATHEMATICAL THEORY OF PROBABILITY

A Thesis submitted in June 1952 for the degree of Doctor of Philosophy in the University of Oxford

by

FREDERIC GORDON FOSTER

Magdalen College

ABSTRACT

The group of problems with which we are concerned may be reduced to the consideration of a system whose state changes with time. It is assumed only that a knowledge of the past history of the system and of its present state determines a probability distribution for its state at any instant in the future. This generalizes the situation in classical mechanics, where the future development of the system is completely determined by its present state. The process in time is a stochastic process. The most usual case is where the conditional probability distribution for future states, given the present state, is unaffected by any additional knowledge of the past history of the system. The process is then a Markov process. We now make some further simplifying assumptions. 1) We regard "time" as a discrete sequence of instants. 2) We assume that the system possesses only a finite or enumerably infinite number of possible states. 3) We assume that the probability distribution for states at time $m + n$, given the state at time m, is independent of m. The process is now called a (temporally homogeneous) Markov chain. Its properties depend on those of the stochastic matrix, $[p_{ij}]$, whose elements are the one-step transition-probabilities from the i^{th} to the j^{th} state of the

system.

In Chapter 2 we treat the case of a system possessing only a finite number of possible states. The material included is classical but an attempt is made to present a unified account, and there is occasionally some novelty in the methods of proof.

In Chapter 3 we proceed to the study of Markov chains with an enumerable infinity of states. There are now new possibilities for the asymptotic behaviour of the system. Let $[p_{ij}^{(n)}]$ be the n^{th} power of the infinite stochastic matrix $[p_{ij}]$. As $n \to \infty$ the $(C, 1)$-limit of $p_{ij}^{(n)}$, π_{ij}, always exists. In the finite case, for each fixed i, $\{\pi_{ij}\}$ $(j = 0, 1, 2, \ldots)$ is a probability distribution. In the infinite case, the possibility exists that $\sum_{j=0}^{\infty} \pi_{ij} < 1$. That is, $\{\pi_{ij}\}$ may no longer be an "honest" probability distribution. (This corresponds to the possibility that the state-label may, with a positive probability, become infinitely large.) When $\sum_{j=0}^{\infty} \pi_{ij} = 1$ for every i, we say that the system is <u>non-dissipative</u>. We are chiefly concerned with finding conditions for non-dissipation. Several theorems are proved to this end, and they find their more important applications in the next two chapters. The principal theorems are quoted below.

A stable distribution is a probability distribution $\{x_j\}$ $(j = 0, 1, 2, \ldots)$ which is such that

$$\sum_{i=0}^{\infty} x_i p_{ij} = x_j , \qquad j = 0, 1, 2, \ldots .$$

A Markov chain is irreducible when it is possible for the system to pass from any state to any other state in a finite number of steps.

THEOREM 3.5.5 An irreducible chain is non-dissipative if and only if there exists a stable distribution.

THEOREM 3.5.6 An irreducible chain is non-dissipative if and only if for some integer i_0 there exists a non-negative solution of the system of "inequations"

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y_i - 1, \qquad i > i_0 ,$$

such that

$$\sum_{j=0}^{\infty} p_{ij} y_j < \infty , \qquad i \leq i_0 .$$

A sharper form of the necessary condition for non-dissipation is also true and is proved in Theorem 3.5.7.

A chain is defined to be simply reducible when the system of states consists of a finite irreducible closed set C and a set D of states from each of which C

is accessible. An irreducible chain becomes simply reducible when it is so modified that one of its states is made "absorbing", i.e. such that when the system attains this state it remains there.

THEOREM 3.6.1. A simply reducible chain is non-dissipative if and only if, for every initial state, the system is almost certain eventually to attain the closed set of states.

Some criteria for non-dissipation are connected with the possible solutions of the system of equations:

$$\sum_{j=0}^{\infty} p_{ij}\, y_j = y_i , \qquad i = 0, 1, 2, \ldots , \quad (3.6.1)$$

and, more generally, of the system of "inequations":

$$\sum_{j=0}^{\infty} p_{ij}\, y_j \leqslant y_i , \qquad i = 0, 1, 2, \ldots . \quad (3.6.2)$$

Three types of solution are distinguished:

(i) y_i constant for all i. This is the trivial solution: it always exists.

(ii) y_i bounded but not constant.

(iii) $y_i \to \infty$ as $i \to \infty$, $|y_i| < \infty$ for all i. We call this a properly divergent solution.

THEOREM 3.6.4. A simply reducible chain is dissipative if and only if there exists a bounded non-constant solution of the equations (3.6.1).

THEOREM 3.6.5. A simply reducible chain is dissipative if and only if there exists a bounded solution $\{y_i\}$ of the inequations (3.6.2), with $y_j < y_i$ for some $j \in D$ and $i \in C$.

THEOREM 3.6.6. If a properly divergent solution of the inequations (3.6.2) exists, then the chain is non-dissipative; and further it possesses a finite closed set of states.

Under some weak additional assumptions, the converse is also true, and this is proved in Theorem 3.6.7.

THEOREM 3.6.8. Any chain is non-dissipative if for some integer i_0 and some constant m, $0 < m < 1$, there exists a properly divergent solution of the system of inequations:

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq m y_i , \qquad i > i_0 ,$$

with

$$\sum_{j=0}^{\infty} p_{ij} y_j < \infty , \qquad i \leq i_0 .$$

In Chapter 4 the general theory is applied to three particular types of stochastic processes of frequent occurrence, namely: Renewal processes, Branching processes and Queueing processes.

In the Renewal process we consider a large population of elements, each of which has a finite life-span and is replaced on failure by an element of zero age. After the lapse of a sufficiently long time the distribution of ages in the population settles down to statistical equilibrium, and it is shown how this distribution may be simply calculated.

In the Branching process, an individual gives rise to a certain number of new individuals, according to a given probability distribution. Each of these in turn gives rise to further individuals, and so on. We are concerned with the random growth and decay of successive generations. The problem of the extinction of families carrying a given surname, first suggested by Francis Galton and since treated by several authors, is considered. A new proof is given of the theorem on extinction, and it is shown how it may be extended in certain respects.

In the Queueing process we consider a queue of "customers" waiting to be served at a "counter". The type of situation treated has many different applications, e.g. to trunking in telephone engineering and to the "stacking" of aircraft. New proofs are given of two theorems, due to D.G.Kendall, on the asymptotic behaviour of queue-sizes. To illustrate the theory graphs are given depicting artificial realizations of two queueing processes.

Chapter 5 is again devoted to the study of a particular Markov chain. We here consider an n-dimensional symmetrical random walk. A particle moves on the points of an n-dimensional lattice. The simplest case is where at each step the particle has equal probabilities of moving to any one of the 2n nearest neighbours of its present position (the Pólya random walk). We also study a much more general problem (generalized random walk) in which at each step the particle has a probability of moving to any point of the lattice. Let the lattice-points be labelled by vectors

$$x \equiv (x_1, x_2, \ldots, x_n)$$

with integer valued components, and suppose that the particle in position x has probability f(y) of moving

in one step into position $x + y$. Further suppose that the probability distribution $\{f(x)\}$ is symmetrical (about the origin):

$$f(x) = f(-x) \quad \text{for all } x.$$

Define $\|x\| \equiv \sum_{i=1}^{n} |x_i|$. The following theorems are proved.

THEOREM 5.2.1. In one dimension if

$$\sum_{y=m}^{\infty} f(y) = o(m^{-1}), \qquad \text{as } m \to \infty,$$

the particle will almost certainly eventually return to its initial position.

THEOREM 5.2.2. In two dimensions if

$$\sum_{\|y\| \geq m} f(y) = o(m^{-2}), \qquad \text{as } m \to \infty,$$

the particle will almost certainly eventually return to its initial position.

THEOREM 5.2.3. In three or more dimensions, provided only that $\{f(y)\}$ is at least a three-dimensional distribution, there is a positive probability that the particle will never return to its initial position.

These results generalize a well known theorem, due to

Pólya, on the simple random walk.

Denote by $p(x)$ the probability that the particle, initially at x, eventually reaches the origin 0. It is shown that $p(x)$ satisfies the system of equations

$$\Delta_f p(x) \equiv \sum_y f(y)\{p(x) - p(x+y)\} = 0, \quad x \neq 0,$$

A study is made of the solutions of

$$\Delta_f g(x) = 0,$$

and also of the more general inequations

$$\Delta_f g(x) \geq 0,$$

wherein $g(x)$ is the unknown function. A solution of the equations is a generalization of a "discrete harmonic function". A <u>properly divergent</u> solution, $g(x)$, is one such that $g(x) \to \infty$ as $\|x\| \to \infty$, while $|g(x)| < \infty$ when $\|x\| < \infty$. The set of all points accessible from the origin 0 in a finite number of steps is denoted by Ω. The following are the principal theorems proved.

THEOREM 5.3.1. <u>There does not exist a bounded non-constant solution of</u> $\Delta_f g(x) = 0, \quad x \in \Omega$.

THEOREM 5.3.4. <u>Let the lattice dimension be n = 1 or 2. When n = 1 suppose that</u> $\sum_{|y| \geq m} f(y) = o(m^{-1})$ <u>and when</u>

$n = 2$ suppose that $\sum_{\|y\| \geq m} f(y) = o(m^{-2})$. Then (i) there does not exist a bounded non-constant solution of

$$\Delta_f g(x) = 0, \quad x \in \Omega - \{o\} ;$$

and (ii) there does exist a properly divergent solution of $\Delta_f g(x) \geq 0, \quad x \in \Omega - \{o\}$.

THEOREM 5.3.5. Let the lattice dimension be $n \geq 3$, and suppose that $\{f(y)\}$ is at least a three-dimensional distribution. Then (i) there exists a bounded non-constant solution of $\Delta_f g(x) = 0, \quad x \in \Omega - \{o\}$; and (ii) there does not exist a properly divergent solution of $\Delta_f g(x) \geq 0, \quad x \in \Omega - \{o\}$, (nor a fortiori of $\Delta_f g(x) = 0, \quad x \in \Omega - \{o\}$).

These theorems generalize and extend known results on discrete harmonic functions.

Some of the material of this thesis is published in the following papers:

Proc.Cambridge Phil.Soc. 47 (1951), 77-85.

J. Roy. Statist. Soc.(B) 13 (1951), 174-8.

Proc.Cambridge Phil.Soc. 48 (1952),

Annals of Math.Stat. 23 (1952),

ACKNOWLEDGMENTS

I am very much indebted to Mr.D.G.Kendall for original suggestions, and for guidance and encouragement throughout the preparation of this thesis; and to Professor H.R.Pitt for first introducing me to the theory of Markov chains, and for assistance with the earlier part of the work.

PREFACE

The group of problems with which we are concerned may be reduced to the consideration of a system whose state changes with time, its states at future times being determined only to the extent of known probability distributions. This generalizes the situation in classical mechanics, where the future development of the system is completely determined by its present state. The process in time is a stochastic process, and we are concerned with the particular case known as a Markov chain.

We first treat the case of a system possessing only a finite number of distinct states. The material included here is classical, but an attempt has been made to present a unified account, and there is occasionally some novelty in the methods of proof. It seems of interest to use the short and elementary methods of matrix theory, available in the finite case, and to note both where these methods carry over to systems with an enumerable infinity of states, and where and how they break down.

The treatment of systems with an enumerable infinity of states is based to a certain extent on the ideas in W.Feller's recently published book. It should be noted however that no appeal is made to the new Tauberian theorem on power series on which Feller's work rests. With the aid of weaker methods, we derive most of the general theory and extend it in certain respects. The principal new theorems in Chapter 3 are Theorems 5.6, 6.4, 6.5, 6.6, 6.7 and 6.8.

In Chapter 4 new proofs, based on this theory, are given of several known theorems relating to particular types of Markov chain, and it is shown how one of these theorems may be extended in certain respects.

The material in Chapter 5 is new. Some of the results appear to hold mathematical interest in themselves (apart from their probability interpretation) in connexion with a generalization of the concept of a discrete harmonic function.

Many examples are given to illustrate the theory. Not all of these are trivial. Attention is drawn in particular to Examples 3.1, 5.1, 5.2, 6.3 and 6.4 of Chapter 3.

A reference such as 'Theorem 3.6.7' is to Chapter 3, Section 6, Theorem 7. 'Section 3.6' refers to Chapter 3, Section 6. When the reference is to the same Chapter, the first figure is omitted.

CONTENTS

CHAPTER 1 INTRODUCTION 1

1. Probability and Measure Theory 1
2. Conditional probabilities 3
3. Statistical independence 3
4. Finite product spaces 4
5. Infinite product spaces 5
6. Random variables 10
7. Stochastic processes 11
8. Markov chains 11
9. Temporally homogeneous Markov chains 14

CHAPTER 2 MARKOV CHAINS WITH A FINITE NUMBER OF STATES 16

1. Finite stochastic matrices 16
2. Asymptotic behaviour of $p_{ij}^{(n)}$ 18
3. Reducibility 20
4. Stable distributions 23
5. Recurrent and transient states 25
6. Regular Markov chains 27
7. Roots of the Characteristic equation 32
8. Periodicity 34
9. Cyclic sub-sets 38

CHAPTER 2 (CONT.)

10. Random walks with absorbing barriers: the gamblers ruin problem 42

11. Note on matrices with non-negative elements 47

CHAPTER 3 MARKOV CHAINS WITH AN ENUMERABLE INFINITY OF STATES 50

1. Dissipative and non-dissipative systems 51

2. Quasi-finite stochastic matrices 55

3. Some theory of recurrent events 56

4. Stable distributions 64

5. Irreducible systems 65

6. Simply reducible systems 75

7. Periodicity 96

8. A Markov chain derivation of discrete distributions 96

CHAPTER 4 APPLICATIONS TO RENEWAL BRANCHING AND QUEUEING PROCESSES 103

1. Renewal processes 103

2. Branching processes 105

3. Queueing processes 110

4. An artificial realization of two queueing processes 119

CHAPTER 5 RANDOM WALKS AND DISCRETE HARMONIC FUNCTIONS 127

1. Pólya's random walk problem 127

2. The generalized random walk 130

3. Solutions of the general linear difference equations and inequations 149

CHAPTER 5 (CONT.)

4. Solutions of the discrete Laplace equations 164

BIBLIOGRAPHY

CHAPTER 1

INTRODUCTION

§0. In this chapter we explain the terminology which will be used. A simple but useful form of a fundamental theorem due to Kolmogorov (1933) is stated and proved, and a Markov chain is defined.

§1. Probability and Measure theory.

The physical basis of the mathematical theory of probability is an experiment, which we imagine may be repeated an indefinite number of times under identical conditions. We postulate a set X of entities, regarded as the (mathematically) possible outcomes of the experiment. Any sub-set $A \subset X$ is an event. The set X is the sample space of the experiment and the entities are its points. We assume that a measure, p, is defined on a ring, R, of sub-sets of X, and that $X \in R$ with $p(X) = 1$. It is known (Halmos, p.54) that p has a unique extension to a measure on the σ-ring, S(R), generated by R, and that further p may be completed so as to be a measure on the extended σ-ring, $\bar{S}$, consisting of sets belonging to S augmented by sub-sets of sets of measure zero in S.

Any set of $\bar{S}$ is measurable. We refer to p as a probability measure.

p(A) is the probability of the event A. The intuitive meaning is that if the experiment were repeated a large number of times under identical conditions we should expect the frequency of the event A to be "approximately" p(A). This meaning, though very important, is not strictly relevant to the mathematical theory, and is necessarily vague.

For the most part, we shall be concerned with the particular case where the sample space X consists of an enumerable (possibly finite) set of points. Such a sample space is called discrete. Let its points be $x_0, x_1, x_2, \cdots$. We postulate that a number $p(x_i)$ is given for every x_i, and that

$$p(x_i) \geq 0 \quad , \quad \sum_{i=0}^{\infty} p(x_i) = 1 \, .$$

The probability of an event A is then the sum of the probabilities, $p(x_i)$, of its sample points:

$$p(A) \equiv \sum_{x_i \in A} p(x_i)$$

The refinements of measure theory are not here required, since every sub-set $A \subset X$ is measurable in this sense.

In the sequel it will however become apparent that non-discrete sample spaces often play a rôle in even the most elementary problems.

§2. Conditional probabilities.

Let A and B be measurable sub-sets of X, with $p(B) > 0$. The conditional probability of A given B is defined to be

$$p(A \mid B) \equiv p(A \frown B) / p(B).$$

For fixed B, $p(A|B)$ is again a probability measure on X.

§3. Statistical independence.

Two sets, A, B, are statistically independent when

$$p(A \frown B) = p(A).\ p(B).$$

This definition then implies that, when defined, $p(A|B) \equiv p(A)$.

More generally, the sets of a (finite or infinite class C are statistically independent when

$$p\left(\tilde{\bigcap_{i=1}} A_i\right) = \tilde{\prod_{i=1}} p(A_i)$$

for every finite class $\{A_i : i = 1, 2, \cdots n\}$ of distinct set of C. It is this notion, and its generalizations to special types of statistical dependence, which

gives to probability theory its own individuality as distinct from measure theory.

§4. Finite product spaces.

Let a sequence of n experiments have respectively sample spaces $X_1, X_2, \cdots X_n$. Their product space is the set,

$$X^{(n)} \equiv X_1 \times X_2 \times \cdots \times X_n$$

of all ordered n-tuples, $(x_1, x_2, \ldots x_n)$, with $x_i \in X_i$. Let R_i be a given ring of sub-sets of X_i, and containing X_i itself, and let $A_i \in R_i$, $i = 1, 2, \cdots n$. The set

$$A^{(n)} \equiv A_1 \times A_2 \times \cdots \times A_n$$

is called a rectangle. The class of all finite disjoint unions of rectangles is a ring $R^{(n)}$ containing $X^{(n)}$. Let a probability measure, p, be defined, a priori, on $R^{(n)}$.

Define

$$\bar{A}_i \equiv X_1 \times \cdots \times A_i \times \cdots \times X_n .$$

Then $A^{(n)} = \bigcap_{i=1}^{n} \bar{A}_i$. If the events $\bar{A}_i \in R^{(n)}$ are statistically independent,

$$p(A^{(n)}) = \prod_{i=1}^{n} p(\bar{A}_i) .$$

If this holds for every rectangle, $A^{(n)}$, we say that the

n experiments are statistically independent. The measure, p, is then called a product measure.

If a measure, p_i, is given on each sample space, X_i, we can clearly define a product measure on $X^{(n)}$ by

$$p(A^{(n)}) \equiv \prod_{i=1}^{n} p_i(A_i) .$$

§5. Infinite product spaces.

The infinite product space of a sequence of sample spaces $X_1, X_2, \cdots$ is the set,

$$X \equiv X_1 \times X_2 \times \cdots$$

of all sequences $(x_1, x_2, \cdots)$, with $x_i \in X_i$. For each i, let R_i be a given ring of sub-sets of X_i, containing X_i itself. When for all but a finite number of values of i, $A_i \equiv X_i$, the set

$$A \equiv A_1 \times A_2 \times \cdots ,$$

with $A_i \in R_i$, is called a rectangle. Define

$$Y^{(n)} \equiv X_{n+1} \times X_{n+2} \times \cdots .$$

Then, for some n,

$$A \equiv A^{(n)} \times Y^{(n)} .$$

The class of all finite disjoint unions of rectangles is a ring, R, containing X.

For each n, let a measure, $p^{(n)}$, be defined on $R^{(n)}$, and let the measures, $p^{(n)}$, satisfy the consistency relations,

$$p^{(n)}(A^{(n)}) = p^{(n+1)}(A^{(n)} \times X_{n+1}) , \quad n = 1, 2, \cdots .$$

It is known (Halmos, p.157) that when each $p^{(n)}$ is a product measure, a unique measure, p, is defined on R, which is such that for any rectangle, $A \equiv A^{(n)} \times Y^{(n)}$,

$$p(A) \equiv p^{(n)}(A^{(n)}) .$$

When however the $p^{(n)}$'s are not product measures, this theorem is not true in general (Halmos, p.211). Some topological condition has to be imposed on the spaces, X_i. We prove now a simple version of a theorem due to Kolmogorov (1933) which is sufficient for our purposes.

THEOREM 5.1. Let each sample space, X_i, consist of an enumerable (possibly finite) set of points, x_i. For each n, let $p^{(n)}(x_1, \cdots, x_n)$ be a probability measure on the space $X^{(n)}$, and let the consistency relations,

$$p^{(n)}(x_1, \cdots, x_n) = \sum_{n+1} p^{(n+1)}(x_1, \cdots, x_{n+1}) ,$$

be satisfied, the summation being over all points $x_{n+1} \in X_{n+1}$. Then a unique probability measure, p, is defined on the ring, R, of those sub-sets of X which are of the form

$$A^{(n)} \times Y^{(n)} ,$$

where $A^{(n)}$ is any sub-set of $X^{(n)}$; and

$$p(x_1 \times x_2 \times \cdots \times x_n \times Y^{(n)}) \equiv p^{(n)}(x_1, \cdots x_n) .$$

Proof. Let $A \in R$ and let $A \equiv A^{(n)} \times Y^{(n)}$ for some n. Define

$$p(A) \equiv \sum p^{(n)}(x_1, \cdots x_n) \equiv p^{(n)}(A^{(n)}) ,$$

the summation being over all points $(x_1, \cdots x_n) \in A^{(n)}$. The integer n is here not unique, but the consistency relations ensure the p(A) is uniquely defined. Clearly the set-function p(A) is non-negative, and p(X) = 1. To see that it is finitely additive, let $B \in R$, with $B \equiv B^{(m)} \times Y^{(m)}$, $A \cap B = 0$, and $m \leq n$. Then

$$A \cup B = C^{(n)} \times Y^{(n)} ,$$

where

$$C^{(n)} \equiv A^{(n)} \cup (B^{(m)} \times X^{(m+1)} \times \cdots \times X^{(n)}) .$$

Now

$$A^{(n)} \cap (B^{(m)} \times X^{(m+1)} \times \cdots \times X^{(n)}) = 0 ,$$

and so

$$p(A \cup B) = p^{(n)}(C^{(n)}) = p^{(n)}(A^{(n)}) + p^{(n)}(B^{(m)} \times X^{(m+1)} \times \cdots \times X^{(n)}) .$$

But, by the consistency relations,

$$p^{(n)}(B^{(m)} \times X^{(m+1)} \times \cdots \times X^{(n)}) \equiv p^{(m)}(B^{(m)}) .$$

Therefore

$$p(A \cup B) = p^{(n)}(A^{(n)}) + p^{(n)}(B^{(n)})$$

$$= p(A) + p(B)$$

The extension to any finite number of disjoint sets is immediate.

It remains only to prove that p(A) is <u>countably</u> additive. For this purpose, it is sufficient (Halmos, p.39) to prove that p(A) is continuous from above at 0. Let $\{E_i\}$, $i = 1, 2, \cdots$, be a decreasing sequence of sets belonging to R. We shall show that $\lim_{n \to \infty} p(E_n) \geq \varepsilon$ implies that $\bigcap_{i=1}^{\infty} E_i \neq 0$. There will be no loss in generality in supposing that

$$E_n \equiv E^{(n)} \times Y^{(n)} .$$

We consider first the case where each space, X_i, consists of only a finite number, N(i), of points. Since $p(E_n) \geq \varepsilon$ for all n, there exists for each n a point $x^{(n)} \in E_n \subset X$. Therefore

$$x^{(n+p)} \in E_{n+p} \subset E_n \quad \text{for all n and p.}$$

Let $x^{(n)} \equiv (x_1^{(n)}, x_2^{(n)}, \cdots)$.

Then $(x_1^{(n+p)}, \cdots x_n^{(n+p)}) \in E^{(n)}$ for all n and p. But the set $E^{(n)}$ contains only a finite number of points, and $E^{(n+1)} \subset E^{(n)} \times X_{n+1}$. Therefore, by the diagonal method, there must exist an increasing sequence of integers $\{n_\nu\}$ such that for each i

$$x_i^{(n_\nu)} \equiv x_i \quad \text{for all} \quad n_\nu .$$

Therefore $(x_1, \cdots x_n) \in E^{(n)}$ for all n.

Therefore $x \equiv (x_1, x_2, \cdots) \in E_n$ for all n.

Therefore $x \in \bigcap_1^\infty E_n$, so that $\bigcap_1^\infty E_n \neq 0$.

We turn now to the case where each space, X_i, may contain an enumerable infinity of points. Each set $E^{(n)}$ may now contain an infinity of points, and so we cannot proceed quite so directly as above.

It is possible to find a set $F^{(n)} \subset E^{(n)}$, such that $F^{(n)}$ consists of only a finite number of points and

$$p^{(n)}(E^{(n)} - F^{(n)}) \leq \varepsilon / 2^{n+1} .$$

Define $F_n \equiv F^{(n)} \times Y^{(n)}$.

Then $p (E_n - F_n) \leq \varepsilon / 2^{n+1}$.

Define $G_n \equiv \bigcap_1^n F_i$.

Then $$p(G_n) = p(E_n) - p(E_n - G_n)$$
$$\geq \varepsilon - \sum_{i=1}^{n} p(E_i - F_i)$$
$$\geq \varepsilon/2 .$$

Now $G_n \equiv G^{(n)} \times Y^{(n)}$,
where $G^{(n)} \subset F^{(n)}$.
We have therefore a decreasing sequence $\{G_n\}$ of sets which are non-empty and belong to R, and each $G^{(n)}$ consists of only a finite number of points, with $G^{(n+1)} \subset G^{(n)} \times X_{n+1}$. Therefore, by the first part of the proof, $\bigcap_1^\infty G_n \neq 0$ But each $G_n \subset E_n$. Therefore also $\bigcap_1^\infty E_n \neq 0$.

We have defined p(A) for every set $A \subset R$. The extension to a probability measure on the σ-ring S(R) is then unique (Halmos, p.54).

§6. Random variables.

A real-valued measurable point-function, x, defined on the sample space of an experiment is a <u>random variable</u>. Its range is a sub-set of the real line, and its measurability implies that a probability measure is defined on the Borel sets of the line. If we are only concerned with the values assumed by x, no further

reference to the origin sample space is required, and it may conveniently be forgotten.

In the case of a discrete sample space, measurability plays no rôle: any real-valued point-function defined on the sample space is a random variable taking an enumerable set of values. In particular, the number associated with each point by any enumeration of the sample points is a random variable.

§7. Stochastic processes.

A sequence of random variables $z_0, z_1, \cdots$ is a stochastic process if a probability measure, p, is defined on the product-space of their respective ranges. When p is a product-measure, the random variables are independent.

We proceed now to study a particular type of stochastic process which is the simplest generalization of a sequence of independent random variables.

§8. Markov chains.

Let each z_n have a range which is a discrete space, $X_n \equiv (0, 1, 2, \cdots)$. Denote by $(x_{\nu_1}, \cdots x_{\nu_n})$ the set of all points in the infinite product space,

$X \equiv X_1 \times X_2 \times \cdots$, which have the indicated components in the $\nu_1, \nu_2, \cdots \nu_{m+k}$ places. The probability measure,

$$p(x_{\nu_1}, x_{\nu_2}, \cdots x_{\nu_m})$$

is defined for every such rectangle and satisfies the consistency relations:

$$\sum_{\nu_i} p(x_{\nu_1}, \cdots x_{\nu_m}) = p(x_{\nu_1}, \cdots x_{\nu_{i-1}}, x_{\nu_{i+1}}, \cdots x_{\nu_m}).$$

The conditional probability of $(x_{\nu_{r+1}}, \cdots x_{\nu_m})$ given $(x_{\nu_1}, \cdots x_{\nu_r})$ is

$$p(x_{\nu_{r+1}}, \cdots x_{\nu_m} | x_{\nu_1}, \cdots x_{\nu_m}) \equiv \frac{p(x_{\nu_1}, \cdots x_{\nu_m})}{p(x_{\nu_1}, \cdots x_{\nu_r})},$$

defined whenever $p(x_{\nu_1}, \cdots x_{\nu_r}) \neq 0$. The stochastic process $\{z_n\}$ is now defined to be a <u>Markov chain</u> if

$$p(x_m | x_0, \cdots x_{m-1}) \equiv p(x_m | x_{m-1}), \quad m = 1, 2, \cdots,$$

whenever the L.H.S. is defined. The following are consequences of this definition

(i) $p(x_0, \cdots x_m) = p(x_0, \cdots x_r) \cdot p(x_{r+1} | x_r) \cdots p(x_n | x_{m-1})$,

<u>whenever the R.H.S. is defined</u>.

For we have, by the definition,

$$p(x_0, \cdots x_r).\, p(x_{r+1} \mid x_r) = p(x_0, \cdots x_r).\, p(x_{r+1} \mid x_0, \cdots x_r)$$

$$= p(x_0, \cdots x_{r+1}) .$$

The result follows by an induction.

(ii) $p(x_n \mid x_m) = p(x_n \mid x_a, x_b, \cdots x_m)$, $a < b < \cdots < m < n$,

whenever the R.H.S. is defined.

For from (i), by appropriate summations,

$$\frac{p(x_m, x_n)}{p(x_m)} = \frac{p(x_a, x_b, \cdots x_m, x_n)}{p(x_a, x_b, \cdots x_m)}$$

(iii) $p(x_n \mid x_\ell) = \sum_m p(x_n \mid x_m).\, p(x_m \mid x_\ell)$,

the summation excluding undefined terms.

For $p(x_n \mid x_\ell) = \sum_m \frac{p(x_\ell, x_m, x_n)}{p(x_\ell)}$

$$= \sum_m \frac{p(x_\ell, x_m).\, p(x_m, x_n)}{p(x_\ell).\, p(x_m)}$$

$$= \sum_m p(x_n \mid x_m).\, p(x_m \mid x_\ell) .$$

Property (ii) is the essential feature of a Markov chain: the probability that $z_n = j$ given $z_m = i$ is <u>independent of any further information</u> regarding z_r, $r < m$. This generalizes the concept of a sequence of independent random variables.

Now set

$$p_{ij}(m,n) \equiv p(x_n | x_m) , \qquad p_i(m) \equiv p(x_m)$$

where $x_n = j$, $x_m = i$. Then the Markov chain is determined by the <u>initial distribution</u> $\{p_i(0)\}$ and the set of <u>stochastic matrices</u>, $[p_{ij}(m, n)]$, given for all m, n, $m < n$. The fundamental relation (iii) becomes

$$p_{ij}(l, n) = \sum_{k=0}^{\infty} p_{ik}(l, m)\, p_{kj}(m, n) .$$

§9. <u>Temporally homogeneous Markov chains</u>.

The most important case arises when $p_{ij}(m, n)$ is a function of the difference $n - m$ only. Define

$$p_{ij}^{(n-m)} \equiv p_{ij}(m, n) , \qquad p_{ij} \equiv p_{ij}(m, m+1) .$$

Then the Markov chain is determined by the single matrix $[p_{ij}]$, and an initial distribution $\{p_i\}$. The fundamental relation is now:

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(m)} p_{kj}^{(n)} .$$

in which $\left[p_{ij}^{(n)}\right]$ is the $n^{\underline{th}}$ power of $\left[p_{ij}\right]$.

We shall only be concerned with this type of Markov chain. The general theory indeed has rarely been studied: some aspects were treated by Kolmogorov (1936 a).

CHAPTER 2

MARKOV CHAINS WITH A FINITE NUMBER OF STATES

§0. A short account is given of the fundamental results in the theory. The main reference is to Fréchet (1938). The subject was first treated in a series of papers (1907-18, Russian) by A.A.Markov. These papers are listed by Hostinsky (1931), where some account of Markov's work may be found.

It is of interest to note, of the theory which follows, both where it does and where it does not carry over to Markov chains with an enumerable infinity of states (to be studied in Chapter 3). The differences between the two theories are differences between finite and infinite matrix theory.

§1. Finite stochastic matrices.

Let $\{z_n\}$ $(n = 0, 1, 2, \cdots)$ be the sequence of random variables of a temporally homogeneous Markov chain. Let the possible values of each z_n be $0, 1, 2, \cdots N$. So as to conform to the usage in physical applications, we say that these values constitute the states of the given system. If $z_n = i$, we

say that the system is in the i^{th} state at <u>time</u> n. We assume that the conditional probability, p_{ij}, that $z_{n+1} = j$, given $z_n = i$, is known a priori for each i and j. The matrix

$$A \equiv [p_{ij}] \qquad (i \text{ and } j = 0, 1, \cdots N)$$

is the <u>stochastic matrix</u> of the system, its components p_{ij} being the <u>transition-probabilities</u>. The conditions on A are:

$$p_{ij} \geq 0 \text{ for all } i, j, \qquad \sum_{j=0}^{N} p_{ij} = 1 \text{ for all } i.$$

Any matrix A satisfying these conditions can serve as a stochastic matrix. We have seen in Chapter 1 that A, together with an initial distribution, completely determines the Markov chain.

Define the n^{th} power

$$[p_{ij}^{(n)}] \equiv A^n .$$

Then we have

$$p_{ij}^{(n)} \geq 0 \text{ for all } i, j, \qquad \sum_{j=0}^{N} p_{ij}^{(n)} = 1,$$

$$p_{ij}^{(m+n)} = \sum_{k=0}^{N} p_{ik}^{(m)} p_{kj}^{(n)} .$$

Also we have seen that $p_{ij}^{(n)}$ is the conditional probability that $z_n = j$ given $z_0 = i$.

§2. Asymptotic behaviour of $p_{ij}^{(n)}$.

A simple example shows that $\lim_{n\to\infty} p_{ij}^{(n)}$ does not always exist. Let a Markov chain consist of two states and have as its stochastic matrix,

$$[p_{ij}] \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} .$$

Then clearly

$$p_{00}^{(2n)} = 1 , \qquad n = 1, 2, \cdots$$

$$p_{00}^{(2n-1)} = 0 , \qquad n = 1, 2, \cdots .$$

We can however prove that the (C, 1)-limit of $p_{ij}^{(n)}$ invariably exists. The following theorem is due to Kolmogorov (1936 b, 1937). The version of the proof we give is due to Yosida and Kakutani (1939).

THEOREM 2.1 *Define* $q_{ij}^{(n)} \equiv \frac{1}{n}\sum_{r=1}^{n} p_{ij}^{(r)}$.

Then $\lim_{n\to\infty} q_{ij}^{(n)} \equiv \pi_{ij}$ *exists for all* i *and* j, *and*

$$\pi_{ij} \geq 0 , \qquad \sum_{j=0}^{N} \pi_{ij} = 1 ,$$

$$\sum_{k=0}^{N} \pi_{ik} p_{kj} = \sum_{k=0}^{N} p_{ik} \pi_{kj} = \sum_{k=0}^{N} \pi_{ik} \pi_{kj} = \pi_{ij} .$$

Proof. By the diagonal method, we can find an increasing sequence of integers $\{n_\nu\}$ such that $\lim_{\nu\to\infty} q_{ij}^{(n_\nu)}$

exists for all i, j. Define

$$\lim_{\nu \to \infty} q_{ij}^{(n_\nu)} \equiv \pi_{ij}$$

For all n,

$$\sum_{k=0}^{N} p_{ik}^{(n)} p_{kj} = \sum_{k=0}^{N} p_{ik} p_{kj}^{(n)} = p_{ij}^{(n+1)} .$$

Therefore

$$\sum_{k=0}^{N} q_{ik}^{(n)} p_{kj} = \sum_{k=0}^{N} p_{ik} q_{kj}^{(n)} = q_{ij}^{(n)} + \frac{1}{n}\{p_{ij}^{(n+1)} - p_{ij}\}$$

Let $n \to \infty$ through the sequence $\{n_\nu\}$. Then

$$\sum_{k=0}^{N} \pi_{ik} p_{kj} = \sum_{k=0}^{N} p_{ik} \pi_{kj} = \pi_{ij} .$$

Now $$\sum_{k=0}^{N} p_{ik}^{(n+1)} \pi_{kj} = \sum_{k=0}^{N} \sum_{l=0}^{N} p_{il}^{(n)} p_{lk} \pi_{kj} = \sum_{k=0}^{N} p_{ik}^{(n)} \pi_{kj} .$$

Therefore for all n,

$$\sum_{k=0}^{N} p_{ik}^{(n)} \pi_{kj} = \pi_{ij} .$$

Therefore also

$$\sum_{k=0}^{N} q_{ik}^{(n)} \pi_{kj} = \pi_{ij} ,$$

and simialrly it may be shown that for all n

$$\sum_{k=0}^{N} \pi_{ik} q_{kj}^{(n)} = \pi_{ij} .$$

Now suppose that for another increasing sequence $\{m_\nu\}$,

$$\lim_{\nu\to\infty} q_{ij}^{(m_\nu)} = \overline{\pi}_{ij} .$$

We have then

$$\sum_{k=0}^{N} \overline{\pi}_{ik}\,\pi_{kj} = \pi_{ij} = \sum_{k=0}^{N} \pi_{ik}\,\overline{\pi}_{kj} .$$

But, by interchanging the rôles of π_{ij} and $\overline{\pi}_{ij}$, we could in the same way obtain

$$\sum_{k=0}^{N} \pi_{ik}\,\overline{\pi}_{kj} = \overline{\pi}_{ij} .$$

Therefore $\pi_{ij} \equiv \overline{\pi}_{ij}$.

Clearly $\pi_{ij} \geq 0$, and since

$$\sum_{j=0}^{N} q_{ij}^{(n)} = 1 \qquad \text{for all } i \text{ and } n,$$

it follows that also

$$\sum_{j=0}^{N} \pi_{ij} = 1 .$$

This completes the proof.

§3. Reducibility.

A proper sub-set C of the set of all states is <u>closed</u> if

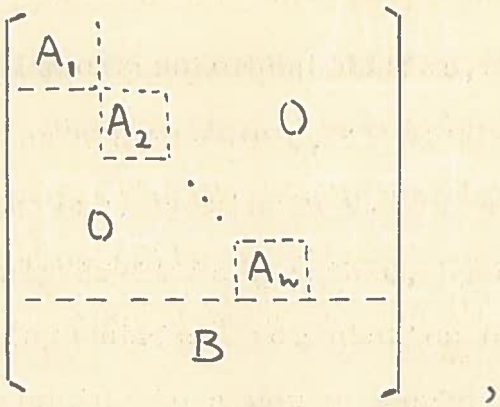
$$\sum_{j\in C} p_{ij} = 1 \qquad \text{for all } i \in C.$$

Thus one-step transitions from any state of C to any

state outside C have zero probabilities, and if ever the system attains a state of C, it will thereafter (with probability one) remain in C. A closed set exists if and only if it is possible, by a cogredient transformation (symmetrical permutation of rows and columns, corresponding to a new numbering of the states), to bring the matrix A into the form

$$\left[\begin{array}{c|c} A_1 & O \\ \hline A_2 & A_3 \end{array}\right],$$

where A_1 is a square stochastic matrix, and O is a null matrix. The matrix A is then reducible. If no such closed subset exists, it is irreducible. The same terms are applied to the corresponding Markov chain. Now the matrix A_1 may in turn be reducible. If so, a cogredient transformation will bring it into the above form. After a finite number of operations, we reach an irreducible matrix, and a corresponding irreducible closed set. A Markov chain possessing two or more irreducible closed sets is decomposable. Its matrix may be re-arranged in the form



in which each A_i is an irreducible stochastic matrix.

The states of the system may now be divided into two classes. We shall call a state final if it belongs to an irreducible closed set; if not, we shall call it intermediate. From any intermediate state there is, for some n, an n-step transition into some final state: for otherwise the state would necessarily belong to an irreducible closed set.

A single final state which itself constitutes an (irreducible) closed set is called absorbing. Clearly if state i is absorbing we must have $p_{ii} = 1$. In the sequel, the set of all states not belonging to a given set S will be denoted by S' (complement of S).

THEOREM 3.1 A Markov chain is irreducible if and only if for every pair of states i, j there exists an n such that $p_{ij}^{(n)} > 0$.

Proof. The existence of a closed sub-set C implies that $p_{ij}^{(n)} = 0$ for all n for $i \in C$, $j \in C'$.

Conversely, let $p_{ij}^{(n)} = 0$ for all n for a certain pair of states i, j. For this i, let S be the set of states j such that $p_{ij}^{(n)} > 0$ for some n. Then S is a proper sub-set of states. For $j \in S'$,

$$p_{ij}^{(n+1)} = 0 = \sum_{k=0}^{N} p_{ik}^{(n)} p_{kj}$$

$$= \sum_{k \in S} p_{ik}^{(n)} p_{kj} .$$

Therefore $p_{kj} = 0$ for $k \in S$, $j \in S'$.

Therefore $\sum_{j \in S} p_{kj} = 1$ for $k \in S$, and so S is closed.

§4. Stable distributions.

If $\{x_i\}$ $(i = 0, 1, 2, \cdots N)$ is a probability distribution such that

$$\sum_{i=0}^{N} x_i p_{ij} = x_j , \qquad j = 0, 1, \cdots N ,$$

then $\{x_i\}$ is a stable distribution of the system.

THEOREM 4.1 There exists at least one stable distribution.

Proof. This follows from Theorem 2.1 if we note that for fixed i, $\{\pi_{ij}\}$ is in fact a stable distribution.

THEOREM 4.2 Any stable distribution $\{x_i\}$ is such that $x_i = 0$ for every intermediate state i.

Proof. Let C be the set of final states, D the set of intermediate states. For all n,

$$\sum_{i \in C} x_i p_{ij}^{(n)} + \sum_{i \in D} x_i p_{ij}^{(n)} = x_j .$$

Therefore $$\sum_{i \in D} \sum_{j \in D} x_i p_{ij}^{(n)} = \sum_{j \in D} x_j .$$

But for each $i \in D$, there exists an n such that

$$\sum_{j \in D} p_{ij}^{(n)} < 1 .$$

Therefore

$$\sum_{j \in D} x_j = 0$$

COROLLARY. $\pi_{ij} = 0$ for $j \in D$ and all i.

The physical significance of a stable distribution is obtained by imagining a large number of identical systems processing simultaneously. If initially the frequency distribution of states is identical with a stable distribution, then this frequency distribution will be maintained at all future instants.

§5. Recurrent and transient states.

A state is recurrent when there is probability one that the system, initially in this state, will eventually return again to it; otherwise, the state is transient. It should be noted that the "probability of eventual return" is the probability-measure, p, of a certain set in the infinite product-space of the sequences $\{z_n\}$. The existence of this probability therefore requires an application of Theorem 1·5·1. Let $z_0 = i$. Let S_n be the set of sequences such that $z_r = i$ for some r, $1 \leq r \leq n$. Then S_n is evidently a measurable set and is monotone non-decreasing with n. Therefore the probability in question exists and is $p(\bigcup_1^\infty S_n)$. The probabilities used in the next theorem are of this type and their existence is similarly justified.

THEOREM 5.1 A state is recurrent if and only if it is final. The system initially in any intermediate state is almost certain eventually to attain the set of final states.

Proof. It is immediate that any intermediate state is transient. For, from any intermediate state there is an n-step transition of positive probability to some final state, from which there is no return. Let C be the set

of final states, and let the system be initially in state $i \in C'$. Then the probability that the set C is attained on or before the n^{th} step is

$$\sum_{j \in C} p_{ij}^{(n)}$$

This is monotone non-decreasing with n. Therefore the probability that the set C is eventually attained is

$$\lim_{n \to \infty} \sum_{j \in C} p_{ij}^{(n)} = \sum_{j \in C} \pi_{ij} = 1 ,$$

on account of Theorem 4.2 and the fact that

$$\sum_{j=0}^{N} \pi_{ij} = 1 .$$

Lastly, to show that any final state is recurrent we use a method of "matrix-modification" to which we shall return again in Section3.6. Suppose the system is initially in a final state, which, by definition, then belongs to an irreducible closed set. The system is (with probability one) confined to this set, and there is no loss in generality in restricting attention to the corresponding irreducible Markov chain. Let its states be 0, 1, 2, $\cdots$ N, and let $z_0 = 0$. Now suppose at the first transition the system moves to state i. We require the probability of the eventual return to 0. Let 0 become an absorbing state. The modified system is now

such that all the states 1, 2, ··· N are intermediate, with 0 constituting the set of final states. By the above argument the probability that 0 is eventually attained from i is one.

§6. Regular Markov chains.

A Markov chain is regular if

$$\pi_{ij} \equiv \pi_j \qquad \text{(independent of i).}$$

We may roughly describe this situation by saying that the asymptotic behaviour of the system is independent of its initial state. Such a chain is also said to satisfy the Ergodic principle.

THEOREM 6.1 A Markov chain is regular if and only if it is non-decomposable.

Proof. If the chain is non-decomposable, it consists of a single irreducible closed set C and a set D of intermediate states from each of which C is accessible. It follows that for all i and for all $j \in C$, there exists an n such that $p_{ij}^{(n)} > 0$. Now for a given $j \in C$, let

$$\pi_j \equiv \operatorname{Max}_i \pi_{ij} \equiv \pi_{i_0 j} .$$

Then

$$\pi_{i_0 j} = \sum_{k=0}^{N} p_{i_0 k}^{(n)} \pi_{kj} .$$

Therefore

$$\sum_{k=0}^{N} p_{i_0 k}^{(n)} (\pi_{kj} - \pi_{i_0 j}) = 0 .$$

Therefore $p_{ik}^{(n)} (\pi_{kj} - \pi_{ij}) = 0$ for all k.

Therefore $\pi_{kj} = \pi_{ij} = \pi_j$ when $k \in C$.

Moreover for any k, since $\pi_{kj} = 0$ for $i \in D$,

$$\pi_{kj} = \sum_{i \in C} \pi_{ki} \pi_{ij} = \pi_j \sum_{i \in C} \pi_{ki} = \pi_j .$$

Conversely, if the chain is decomposable, then by definition there exist two irreducible closed sets C_1, C_2, and for $i \in C_1$, $j \in C_2$ we have

$$\pi_{ik} = 0 \text{ but } \pi_{jk} > 0 \text{ for some } k \in C_2.$$

THEOREM 6.2 $\pi_j > 0$ <u>for</u> $j \in C$.

Proof. For any $i, j \in C$ there exists an n such that $p_{ij}^{(n)} > 0$. Now $\pi_k = 0$ for $k \in D$. Therefore

$$\sum_{i \in C} \pi_i p_{ij}^{(n)} = \pi_j$$

But $\pi_j = 0$ would imply that $\pi_i = 0$ for all $i \in C$ which is impossible since $\sum_{i \in C} \pi_i = 1$.

THEOREM 6.3 <u>A Markov chain is regular if and only if the</u> equations

$$\sum_{i=0}^{N} x_i p_{ij} = x_j , \quad j = 0, 1, \cdots N .$$

<u>have (apart from a scaler factor) a unique solution</u>.

Proof. If the solution is unique, then, since for fixed i $\{\pi_{ij}\}$ is a solution, π_{ij} must be independent of i. Conversely let $\pi_{ij} \equiv \pi_j$. Then we have

$$\sum_{i=0}^{N} x_i p_{ij}^{(n)} = x_j \quad , \quad j = 0, 1 \cdots N,$$

so that also

$$\sum_{i=0}^{N} x_i \pi_{ij} = x_j .$$

Therefore $$\pi_j \sum_{i=1}^{N} x_i = x_j ,$$

i.e. $$x_j = C \cdot \pi_j .$$

THEOREM 6.4 **A Markov chain is regular if and only if the equations**

$$\sum_{j=0}^{N} p_{ij} y_j = y_i , \quad i = 0, 1, \cdots N,$$

have a unique solution.

Proof. Since $\sum_{j=0}^{N} p_{ij} = 1$, the solution $y_i \equiv C$ (constant) always exists. If it is the only solution, then $\pi_{ij} \equiv \pi_j$ Conversely, let $\pi_{ij} \equiv \pi_j$. We have

$$\sum_{j=0}^{N} p_{ij}^{(n)} y_j = y_i .$$

Therefore $$\sum_{j=0}^{N} \pi_{ij} y_j = y_i .$$

Therefore $$y_i = \sum_{j=0}^{N} \pi_j y_j = C .$$

THEOREM 6.5 If a Markov chain is decomposable into n irreducible closed sets and an intermediate set, then there exist n orthogonal stable distributions such that any solution of

$$\sum_{j=0}^{N} x_i p_{ij} = x_j , \quad j = 0, 1, \cdots N,$$

is a linear combination of them.

Proof. Let D be the intermediate set. Then by Theorem 4.2, $x_i = 0$ for $i \in D$. Hence we need only consider the equations

$$\sum_{i \in D'} x_i p_{ij} = x_j , \quad j \in D' .$$

Now let the closed sets be $C_1, C_2, \cdots C_n$. Then $p_{ij} > 0$ only if i, j belong to the same set C_r. Therefore this set of equations splits up into the n sets

$$\sum_{i \in C_r} x_i p_{ij} = x_j , \quad j \in C_r \quad (r = 1, 2, \cdots n).$$

Each has by Theorem 6.4, a unique solution, apart from a scalar factor. Corresponding to each such solution, we obtain a solution for the whole system by placing $x_i = 0$ when $i \notin C_r$. We thus obtain n orthogonal solutions with the required properties.

EXAMPLE 6.1

Let $$[\phi_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

The chain is decomposable into the irreducible closed sets (0) and (3), and the set (1, 2) is intermediate. Any stable distribution is a linear combination of (1, 0, 0, 0) and (0, 0, 0, 1). To find the matrix $[\pi_{ij}]$ we proceed as follows. We have

$$\sum_{j=0}^{3} \phi_{ij}\, \pi_{j0} = \pi_{i0} , \qquad i = 0, 1, 2, 3 .$$

which, on substituting the given values for p_{ij}, becomes

$$\pi_{00} = \pi_{00} ,$$

$$\tfrac{1}{2}(\pi_{00} + \pi_{20}) = \pi_{10} ,$$

$$\tfrac{1}{2}(\pi_{10} + \pi_{30}) = \pi_{20} ,$$

$$\pi_{30} = \pi_{30} .$$

But we know that $\pi_{00} = 1$ and $\pi_{30} = 0$.

Therefore $$\tfrac{1}{2} + \tfrac{1}{2}\pi_{20} = \pi_{10} ,$$

$$\tfrac{1}{2}\pi_{10} = \pi_{20} .$$

Therefore $\pi_{10} = 2/3$, $\pi_{20} = 1/3$.

Similarly, we could calculate the values

$$\pi_{13} = 1/3 \quad , \quad \pi_{23} = 2/3 .$$

Since $\pi_{ij} = 0$ for $j = 1, 2$, we have finally

$$[\pi_{ij}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

§7. Roots of the Characteristic Equation.

The characteristic equation of the matrix A is

$$|A - \lambda I| = 0.$$

Its roots, λ, are the latent roots of A. Corresponding to each root, λ, there exists a row-vector x and a column-vector y such that

$$x A = \lambda x , \quad A y = \lambda y .$$

The vector x, y are latent vectors.

THEOREM 7.1 $\lambda = 1$ is always a latent root.

Proof. $\sum_{j=0}^{N} (p_{ij} - \delta_{ij}) = 0$ for each i.

THEOREM 7.2 Every latent root has modulus less than or equal to one.

Proof. Let $\{y_i\}$ be a column-latent-vector corresponding to a latent root λ :

$$\sum_{j=0}^{N} p_{ij}\, y_j \;=\; \lambda\, y_i \quad .$$

Then $$|\lambda| . |y_i| \;\leq\; \sum_{j=0}^{N} p_{ij}\, |y_j| \quad .$$

Let $$M \;\equiv\; \max_i |x_i| \quad .$$

Then for this value of i,

$$|\lambda| . M \;\leq\; M \quad .$$

And $M > 0$. Therefore $|\lambda| \leq 1$.

THEOREM 7.3 <u>Each latent root of modulus one is a root of unity</u>.

Proof. Let $\lambda = e^{i\alpha}$ be a latent root. Then for some row-vector, $\{x_j\}$,

$$\sum_{j=0}^{N} x_j\, p_{jk} \;=\; \lambda\, x_k \quad .$$

Therefore

$$\Big|\sum_{j=0}^{N} x_j\, p_{jk}\Big| \;=\; |\lambda\, x_k| \;=\; |x_k| \quad .$$

Therefore

$$\sum_{k=0}^{N} \Big|\sum_{j=0}^{N} x_j\, p_{jk}\Big| \;=\; \sum_{j=0}^{N} |x_j|$$

$$= \sum_{j=0}^{N} |x_j| \sum_{k=0}^{N} p_{jk}$$

$$= \sum_{k=0}^{N} \sum_{j=0}^{N} |x_j|\, p_{jk} \quad .$$

Therefore for each k,

$$\left| \sum_{j=0}^{N} x_j p_{jk} \right| = \sum_{j=0}^{N} | x_j p_{jk} | .$$

Therefore, whenever $p_{jk} > 0$, $x_j = \rho_j e^{i\theta}$, where θ is independent of j. Now for some k, $y_k \neq 0$. Let $y_k = \rho_k e^{i\varphi}$. Then we have

$$e^{i\alpha} \cdot \rho_k e^{i\varphi} = e^{i\theta} \sum_{j=0}^{N} \rho_j p_{jk} .$$

Therefore $\theta = \varphi + \alpha$.

Therefore if φ is an argument of some y_k, then $\varphi + \alpha$ is also an argument of some y_k. But there can only be a finite number, $r + 1$ ($\leq N+1$) of distinct arguments. Therefore, for some integer s, $\varphi + r\alpha = \varphi + 2s\pi$.

Therefore $\alpha = (s/r)\, 2\pi$,

so that $\lambda^r = 1$ as required.

EXAMPLE 7.1

The characteristic equation of $A \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

is $\lambda^3 = 1$, so that the roots are the cube roots of unity, $1, \omega, \omega^2$. This is an irreducible chain with

$$\pi_j = \tfrac{1}{3} , \quad j = 0, 1, 2 .$$

§8. Periodicity.

If the Markov chain is such that the ordinary limits

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_{ij}$$

exist for every i, j, it is a-periodic. Otherwise it is periodic.

THEOREM 8.1 A necessary and sufficient condition for a Markov chain to be aperiodic is that $\lambda = 1$ is the only latent root of modulus one.

Proof. By a classical theorem on reduction to canonical form (Ferrar, p.68), there exists a non-singular matrix T such that

$$A = T\,C\,T^{-1},$$

and C consists of the latent roots of A, repeated according to their multiplicities, along the main diagonal, with 0's and 1's along the super-diagonal and 0's elsewhere. It follows that

$$A^n = T\,C^n\,T^{-1},$$

and $A^n \longrightarrow$ limit if and only if $C^n \longrightarrow$ limit.

Let the distinct latent roots be $\lambda_1, \lambda_2, \cdots \lambda_k$. Let

$$C \equiv \text{diag}\ \{D(\lambda_1), \cdots, D(\lambda_k)\}$$

where $D(\lambda)$ is the sub matrix corresponding to λ. Then $C^n \equiv \text{diag}\ \{D^n(\lambda_1), \cdots, D^n(\lambda_k)\}$.

Clearly, when $|\lambda| < 1$, $D^n(\lambda) \to 0$ as $n \to \infty$. Now consider any root λ, with $|\lambda| = 1$, of multiplicity m. Since A^n is bounded, it follows that $D(\lambda)$ is purely diagonal of order m, so that

$$D^n(\lambda) = \begin{bmatrix} \lambda^n & & & \\ & \lambda^n & & \\ & & \ddots & \\ & & & \lambda^n \end{bmatrix}.$$

Therefore $D^n(\lambda) \to$ limit if and only if $\lambda = 1$. Since $C^n \to$ limit if and only if each $D^n(\lambda) \to$ limit, the result follows.

Remark. An application of Theorem 13 now provides an alternative proof of the result, already established in Theorem 2.1, that the (C, 1)-limit of A^n always exists. For if $\lambda \neq 1$ is a root of unity, $\frac{1}{n}\sum_{r=1}^{n} D^r(\lambda) \to 0$ as $n \to \infty$. The previous method is however of interest since we shall see that it extends to the case of an enumerable infinity of states, whereas the present method does not.

If $\lambda = 1$ has multiplicity m, the matrix $[\pi_{ij}]$ will be

$$T \cdot \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} \cdot T^{-1},$$

in which there are m 1's along the diagonal. We use this to prove the

THEOREM 8.2 A Markov chain is regular if and only if $\lambda = 1$ is a simple root.

Proof. If $\lambda = 1$ is simple, it will be seen from the above representation that each row of $[\pi_{ij}]$ must be the same. On the other hand let λ have multiplicity $m > 1$. Let $T = [t_{ij}]$ and let the rows of T^{-1} be $\rho_0, \rho_1, \cdots \rho_N$. Then the i^{th} row of $[\pi_{ij}]$ is

$$\sum_{j=0}^{m} t_{ij} \rho_j \quad .$$

The non-singularity of T and T^{-1} now imply that not all these rows are identical.

EXAMPLE 8.1

$$\text{Let} \qquad A \equiv \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} .$$

The characteristic equation is $\lambda(\lambda^3 - 1) = 0$, and so the roots are $\lambda = 0, 1, e^{2\pi i/3}, e^{4\pi i/3}$. The chain is periodic since there are roots of modulus one other than $\lambda = 1$. It is regular since $\lambda = 1$ is simple.

§9. Cyclic sub-sets.

We have seen how the set of states of a general Markov chain may be sub-divided into an intermediate set and a number of irreducible closed sets, and that the system initially in an intermediate state eventually attains one of the closed sets, thereafter remaining within this closed set. We now study in more detail its behaviour within an irreducible closed set.

The irreducible closed set forms the set of states of an irreducible chain. Let its states be $0, 1, 2, \cdots, N$, and let m be the smallest number for which $\lambda^m = 1$ whenever λ is a latent root of modulus one. We shall show that the set of all states may be sub-divided into m sub-sets $S_1, S_2, \cdots, S_m$, such that the system initially in a state of S_1 is transferred in one-step transitions (with probability one) cyclically through $S_2, S_3, \cdots, S_m$, back to S_1, and so on.

It is implicit in the proof of Theorem 8.1 that the only root of modulus one of A^m is one itself. Let the Markov chain associated with the matrix A^m possess m' irreducible closed sets. By Theorem 6.5, there exist m' orthogonal stable distributions, $u^{(i)}$ $(i = 1, 2, \cdots m)$, such that any solution of $x A^m = x$ is a linear

combination of them. Now

$$(u^{(i)}A)A^{m'} = (u^{(i)}A^{m'})A = u^{(i)}A .$$

Therefore, for some constants C_{ij},

$$u^{(i)}A = \sum_{j=1}^{m'} C_{ij}\, u^{(j)} .$$

But $u^{(i)}A^{m'} = u^{(i)}$.

Therefore $c_{ij}^{(m')} = \delta_{ij}$, where $[C_{ij}^{(m)}]$ is the m^{th} power of $[C_{ij}]$. It follows that $[C_{ij}]$ is merely a cogredient transform of the unit matrix $[\delta_{ij}]$ so that for each i, $u^{(i)}A = u^{(j)}$ for some j. Now the vector

$$x = \frac{1}{m'}\sum_{j=1}^{m'} u^{(j)}$$

is a stable distribution of A. But this is unique. Therefore, by a re-arrangement of the indices if necessary, we must have

$$u^{(1)}A = u^{(2)}, \quad u^{(2)}A = u^{(3)}, \quad \ldots, \quad u^{(m')}A = u^{(1)},$$

i.e. no sub-set of the vector, $u^{(i)}$, forms a cycle.

We can now show that $m' = m$. For we have both $u^{(i)}A^{m} = u^{(i)}$ and $u^{(i)}A^{m'} = u^{(i)}$. Moreover, $u^{(i)}A^{r} \neq u^{(i)}$ for $r < m'$. Therefore $m' \leq m$. On the other hand, let λ be any latent root of A with $|\lambda| = 1$. Then for some vector x

$$x A = \lambda x$$

Therefore $x A^m = \lambda^m x = x$.

Therefore x is a latent vector of A^m corresponding to the latent root unity. Therefore, for some constants C_1,

$$x = \sum_{i=1}^{m'} c_i u^{(i)} .$$

Therefore $x A^{m'} = \lambda^{m'} x = \sum_{i=1}^{m'} c_i u^{(i)} A^{m'} = \sum_{i=1}^{m'} c_i u^{(i)} = x$.

Therefore $\lambda^{m'} = 1$. But m is the smallest integer with this property. Therefore $m' \geq m$. Combining this with the previous inequality, we obtain $m' = m$.

Now let S_r be the set of states corresponding to the non-zero components of $u^{(r)}$, and define $S_{m+1} \equiv S_1$, $u^{(m+1)} \equiv u^{(1)}$. We have

$$\sum_{i \in S_r} u_i^{(r)} p_{ij} = u_j^{(r+1)} .$$

Therefore $$\sum_{i \in S_r} u_i^{(r)} \sum_{j \in S_{r+1}} p_{ij} = 1 .$$

But $u_i^{(r)} > 0$ for each $i \in S_r$. Therefore if $i \in S_r$, $\sum_{j \in S_{r+1}} p_{ij} = 1$. The system in S_r is thus transformed with probability one into S_{r+1}, and so on.

When m = 1, the chain is aperiodic. For then $\lambda = 1$ is the only root of modulus 1 and Theorem 8.1 applies.

THEOREM 9.1 A sufficient condition for a Markov chain to be aperiodic is that $p_{ii} > 0$ for each i.

Proof. If $p_{ii} > 0$ for each i, there cannot exist any cyclic subsets, since the system has a positive probability of remaining where it was after a one-step transition. Therefore, m = 1 and so, by the above, the system is aperiodic. A direct proof of the result can also be given.

EXAMPLE 9.1 In the chain of the previous example there are three cyclic sub-sets: (0, 1), (3), (2), and

$$(0, 1) \rightarrow (3) \rightarrow (2) \rightarrow (0, 1).$$

Thus m = 3, corresponding to the fact that there are three latent roots of modulus unity. The matrix A^3 is

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It has only $\lambda = 1$ as a latent root of modulus unity. The 3 orthogonal solutions of A^3 are

$$u^{(1)} = (\tfrac{1}{2}, \tfrac{1}{2}, 0, 0),$$

$$u^{(2)} = (0, 0, 1, 0),$$

$$u^{(3)} = (0, 0, 0, 1),$$

and the unique stable distribution of A is

$$\chi = \frac{1}{3}\left(u^{(1)} + u^{(2)} + u^{(3)}\right)$$

$$= \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{3}\right).$$

§10. Random walks with absorbing barriers: the gambler's ruin problem.

A particle moves on the points 0, 1, 2, $\cdots$ N of a line in such a way that when at point i, $0 < i < N$, it has probabilities λ, μ of jumping respectively to $i+1$ and $i-1$. The points 0 and N are such that whenever the particle reaches either one the movement ceases, i.e. the particle is "absrobed". The points 0, N are described as absorbing barriers. This is an example of a Markov chain with two absorbing states and a set $(1, 2, 3, \cdots, N-1)$ of intermediate states. It may also be given the following interpretation. A gambler, with initial capital i, wins or loses one unit in each game with probabilities λ, μ. He continues to play until either his capital is increased to N or reduced to zero. We are interested in the probability of his ruin.

The stochastic matrix of the Markov chain is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & \\ \mu & 0 & \lambda & 0 & \cdots & \\ 0 & \mu & 0 & \lambda & \cdots & \\ \vdots & \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & 0 & & 1 \end{bmatrix} ,$$

and the probability of absorption at the barrier 0 in which we are interested is π_{i0} . We therefore set down the equations

$$\sum_{j=0}^{N} p_{ij} \pi_{j0} = \pi_{i0} , \qquad i = 0, 1, \cdots N .$$

To simplify the notation, define $u_i \equiv \pi_{i0}$. In our case the equations become

$$\mu u_i + \lambda u_{i+2} = u_{i+1} , \qquad 0 \leq i \leq N-2 .$$

Therefore $\lambda (u_i - u_{i-1}) = \mu (u_{i-1} - u_{i-2}) , \quad 2 \leq i \leq N .$

Therefore $u_i - u_{i-1} = \frac{\mu}{\lambda} (u_{i-1} - u_{i-2}) .$

Therefore, by induction,

$$u_i - u_{i-1} = \left(\frac{\mu}{\lambda}\right)^{i-1} (u_1 - u_0) .$$

Therefore, again by induction,

$$u_i = u_1 + (u_1 - u_0) \sum_{r=1}^{i-1} \left(\frac{\mu}{\lambda}\right)^r \qquad (10.1)$$

Suppose first that $\mu \neq \lambda$ and define $\nu \equiv \mu/\lambda$.

Then $$u_i = u_1 \left(\frac{1-\nu^i}{1-\nu} \right) - u_0 \left(\frac{\nu-\nu^i}{1-\nu} \right). \qquad (10.2)$$

We may now determine u_1 by noting that $u_0 = 1$, $u_N = 0$. We have

$$0 = u_1 \cdot \frac{1-\nu^N}{1-\nu} - \frac{\nu-\nu^N}{1-\nu}.$$

Therefore $$u_1 = \frac{\nu-\nu^N}{1-\nu^N}.$$

Therefore, substituting in (10.2),

$$u_i = \frac{\nu^N - \nu^i}{\nu^N - 1}, \qquad 0 \leq i \leq N. \qquad (10.3)$$

When $\mu = \frac{q}{p} = \frac{1}{2}$, this of course becomes indeterminate, and we must return to the expression (10.1), which is now

$$u_i = i u_1 - i + 1.$$

Therefore $$u_N = 0 = N u_1 - N + 1,$$

so that $$u_1 = 1 - \frac{1}{N},$$

and $$u_i = 1 - \frac{i}{N}. \qquad (10.4)$$

If in (10.3) and (10.4) we allow $N \to \infty$, we shall obtain the probability of ruin in a game against an infinitely rich adversary. To justify this interpretation of the limit, we consider the corresponding random walk on the enumerable set of integers 0, 1, 2, $\cdots$ (i.e. the absorbing limit at N has passed to infinity). A

probability measure, p, is defined on the infinite product space of the sequence of random variables $\{z_n\}$, with $z_0 = i$. Consider the set in this space:

$$S_N : \left[z_n = 0 \text{ for some finite } n = n_0 ; \; z_n < N \text{ for } n < n_0 \right].$$

Then clearly

$$u_i(N) \equiv p(S_N) .$$

Now S_N is monotone non-decreasing. Therefore

$$\lim_{N \to \infty} u_i(N) = p\left(\lim_{N \to \infty} S_N\right) = p\left(\bigcup_1^\infty S_N\right),$$

and $\bigcup_1^\infty S_N$ is the set of sequences $\{z_n\}$ such that $z_n = 0$ for some n. The measure of this set is thus the probability of ruin in question. Inspection of (10.3) and (10.4) therefore gives us the result that <u>the probability of ruin in a game against an infinitely rich adversary is one if and only if</u> $\mu \geq \lambda$.

We turn now to a generalization of the random-walk problem, in which the particle in position i, $0 < i < N$, has probabilities λ_i , μ_i of jumping into positions $i + 1$, $i - 1$. 0 and N are again absorbing barriers. This has also of course a (rather artificial) gambling interpretation. Our gambler and his adversary have respectively i and $N - i$. A sequence of games is played until one of them is ruined. If after the n^{th}

game the gambler's capital is $j < N$, then in the $(n+1)^{st}$ game his chances of winning or losing are adjusted so as to be λ_j, μ_j . The matrix is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & \\ \mu_1 & 0 & \lambda_1 & 0 & 0 & \cdots & \\ 0 & \mu_2 & 0 & \lambda_2 & 0 & \cdots & \\ 0 & 0 & \mu_3 & 0 & \lambda_3 & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} ,$$

and the equations to be solved are

$$\lambda_{i-1}(u_i - u_{i-1}) = \mu_{i-1}(u_{i-1} - u_{i-2}) , \quad 2 \le i \le N,$$

with $u_0 = 1$, $u_N = 0$.

Therefore $u_i - u_{i-1} = \frac{\mu_{i-1}}{\lambda_{i-1}}(u_{i-1} - u_{i-2})$.

Therefore, by induction, $u_i - u_{i-1} = \frac{\mu_1 \mu_2 \cdots \mu_{i-1}}{\lambda_1 \lambda_2 \cdots \lambda_{i-1}}(u_1 - u_0)$.

Therefore, again by induction,

$$u_i = u_1 + (u_1 - 1)\sum_{r=1}^{i-1} v_r ,$$

where $v_r \equiv \frac{\mu_1 \mu_2 \cdots \mu_r}{\lambda_1 \lambda_2 \cdots \lambda_r}$.

Therefore

$$u_1 = \sum_{r=1}^{N-1} v_r \Big/ \left(1 + \sum_{r=1}^{N-1} v_r\right) ,$$

and* $$u_i = \sum_{r=i}^{N-1} v_r \Big/ \left(1 + \sum_{r=1}^{N-1} v_r\right) .$$

Letting $N \to \infty$, we find that the necessary and sufficient condition that $\lim_{N\to\infty} u_i = 1$ is that

$$\sum_{r=1}^{\infty} \frac{\mu_1 \mu_2 \cdots \mu_r}{\lambda_1 \lambda_2 \cdots \lambda_r} = \infty .$$

The interpretation is as before: this is the necessary and sufficient condition that the probability of ruin is one in a sequence of games against an infinitely rich adversary.

It should be noted that the case $N = \infty$ is defined on its own merits as a Markov chain with an enumerable infinity of states, and not merely as a limiting case, as it has been treated here. We shall see in the following chapter that the result of a direct treatment agrees with the above (Cf. Example 3.6.3).

§ 11. Note on matrices with non-negative elements.

Some of the results we have obtained for stochastic matrices have their counterparts in a general theory of matrices with non-negative elements which has been studied

* This result has now also been obtained by T.E.Harris (to be given in a forthcoming paper: see Bibliography).

independently of probability applications. Perron (1907) proved that a matrix, A, with positive elements, has a positive latent root, ω, which is simple and greater in absolute value than every other root, and that the elements of the adjugate matrix of $\omega I - A$ are all positive. Frobenius (1908) gave another proof of this result and extended it (1912) to the case of an irreducible matrix with non-negative elements. A simplified proof has recently been given by Wielandt (1950).

Several authors have considered the problem of finding bounds for ω. Let $A \equiv [a_{ij}]$ (i and $j = 0, 1, 2, \cdots N$). Define

$$k \equiv \min_{i,j} a_{ij} > 0,$$

$$R_i \equiv \sum_{j=0}^{N} a_{ij}, \quad R \equiv \max_i R_i, \quad r \equiv \min_i R_i,$$

$$\delta \equiv \sqrt{\max_{R_i < R_j} \left(\frac{R_i}{R_j}\right)}.$$

Lederman (1950) showed that (when $\delta < 1$)

$$r + k(\delta^{-1} - 1) < \omega < R - k(1 - \delta).$$

These inequalities have lately been improved by Ostrowski (1952), who showed that

$$r + k(\sigma^{-1} - 1) \leq \omega \leq R - k(1 - \sigma),$$

where $\sigma \equiv \sqrt{\frac{r-k}{R-k}}$.

Other references are: Collatz (1942), Brauer (1946, 1948, 1952), Perfect (1951), Taussky (1948, 1951).

CHAPTER 3

MARKOV CHAINS WITH AN ENUMERABLE INFINITY OF STATES

§0. The theory is treated as a natural extension of the work of Chapter 2. An account from a rather different point of view has recently been given by Feller (1950). We concentrate on some aspects not hitherto dealt with. The first explicit treatment of the subject seems to have been by Fortet (1935).

The Markov chain now possesses the enumerable infinity of states: 0, 1, 2, $\cdots$. We are therefore concerned with an <u>infinite stochastic matrix</u> $A \equiv [p_{ij}]$ (i and j = 0, 1, 2, $\cdots$), where components have, as before, the properties:

$$p_{ij} \geq 0 , \qquad \sum_{j=0}^{\infty} p_{ij} = 1 .$$

The definitions of a closed set, reducibility, decomposability, final and intermediate states, recurrent and transient states, stable distribution, and regularity require no alteration. Moreover Theorem 2.3.1 carries over without modification, so that <u>the system is irreducible if and only if</u> $p_{ij}^{(n)} > 0$ <u>for some</u> n <u>for every pair of states</u> i <u>and</u> j.

For the finite system the classification into final and intermediate states coincided with that into recurrent and transient states (Theorem 2.5.1). This is no longer true for the infinite system. Any intermediate state is clearly still transient. But, as we shall see, when the set of final states is infinite these may also all be transient.

A further, related, difference between the two cases which should be noted is that in the infinite system it is not necessary that <u>any</u> final states should exist. In Example 1.1 below <u>every</u> state will be seen to be intermediate.

§1. Dissipative and Non-dissipative systems.

The fundamental theorem, due to Kolmogorov (1936, 1937) on which we base our theory is an extension of Theorem 2.2.1. The version of the proof we give is again due to Yosida and Kakutani (1939).

THEOREM 1.1 <u>Define</u> $q_{ij}^{(n)} \equiv \frac{1}{n}\sum_{r=1}^{n} p_{ij}^{(r)}$.

<u>Then</u> $\lim_{n\to\infty} q_{ij}^{(n)} = \pi_{ij}$ <u>exists for all</u> i <u>and</u> j, <u>and</u>

$$\pi_{ij} \geq 0, \qquad \sum_{j=0}^{\infty} \pi_{ij} \leq 1,$$

$$\sum_{k=0}^{\infty} \pi_{ik} p_{kj} = \sum_{k=0}^{\infty} p_{ik} \pi_{kj} = \sum_{k=0}^{\infty} \pi_{ik} \pi_{kj} = \pi_{ij}. \qquad (1.1)$$

Proof. By an argument identical with that used in Theorem 2.2.1 we can show that as $n \to \infty$ through a certain sequence $\{n_\nu\}$, $q_{ij}^{(n)} \to \pi_{ij}$, and

$$\sum_{k=0}^{\infty} \pi_{ik} p_{kj} \leq \sum_{k=0}^{\infty} p_{ik} \pi_{kj} = \pi_{ij} .$$

But we must have here equalities throughout, since

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_{ik} p_{kj} = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \pi_{ik} p_{kj} = \sum_{k=0}^{\infty} \pi_{ik} .$$

Now suppose that for some other sequence $\{m_\nu\}$, $q_{ij}^{(n)} \to \overline{\pi}_{ij}$. Then, again by an argument identical with that in Theorem 2.2.1, we have that

$$\sum_{k=0}^{\infty} \overline{\pi}_{ik} \pi_{kj} \leq \sum_{k=0}^{\infty} \pi_{ik} \overline{\pi}_{kj} = \pi_{ij} .$$

But, by interchanging the rôles of π_{ij}, $\overline{\pi}_{ij}$ we obtain

$$\sum_{k=0}^{\infty} \pi_{ik} \overline{\pi}_{kj} \leq \sum_{k=0}^{\infty} \overline{\pi}_{ik} \pi_{kj} = \overline{\pi}_{ij} .$$

It follows that $\pi_{ij} \equiv \overline{\pi}_{ij}$ and

$$\sum_{k=0}^{\infty} \pi_{ik} \pi_{kj} = \pi_{ij} .$$

This completes the proof.

In contrast with the finite system, there are now three possibilities for the matrix $[\pi_{ij}]$:

1) $\pi_{ij} \equiv 0$ for all i and j.

2) $\pi_{ij} > 0$ for some i, j, but $\sum_{j=0}^{\infty} \pi_{ij} < 1$ for some i.

3) $\sum_{j=0}^{\infty} \pi_{ij} = 1$ for all i.

The following examples illustrate these possibilities.

EXAMPLE 1.1

$$[p_{ij}] \equiv \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

We have here $p_{ij}^{(n)} = 0$ whenever $j - i \neq n$. Therefore

$$\pi_{ij} = \lim_{n \to \infty} p_{ij}^{(n)} = 0 \quad \text{for all i and j.}$$

EXAMPLE 1.2

$$[p_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \alpha & 0 & 1-\alpha & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

We have here $p_{00}^{(n)} \equiv 1$, $p_{10}^{(n)} \equiv \alpha$, $p_{i0}^{(n)} \equiv 0$, $i \geq 2$,

$$p_{ij}^{(n)} \to 0 \quad \text{for } i \text{ and } j \geq 1.$$

Therefore $\sum_{j=0}^{\infty} \pi_{0j} = 1$, $\sum_{j=0}^{\infty} \pi_{ij} = \alpha$, $\sum_{j=0}^{\infty} \pi_{ij} = 0$, $i \geq 2$.

EXAMPLE 1.3

$$[p_{ij}] \equiv \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

We have here $p_{ij}^{(n)} \equiv p_{ij}$. Therefore $\pi_{ij} \equiv p_{ij}$, so that $\sum_{j=0}^{\infty} \pi_{ij} = 1$ for all i.

The above three types of system will be called respectively the completely dissipative, semi-dissipative and non-dissipative systems.* In the present chapter we shall be mainly concerned with the derivation of criteria which will enable us to recognise in practice to which category a given Markov chain belongs.

* I introduced these terms in my paper (1951)[a]. Lévy (1951) comments on them in connection with his own new classificatory system.

§2. Quasi-finite stochastic matrices.

We now consider briefly one type of infinite stochastic matrix which is of interest because of the significant reason why it has no practical applications. Suppose that as $N \to \infty$.

$$\sum_{j=0}^{N} p_{ij} \to 1 \qquad \text{uniformly in } i.$$

Then we can show that the system is non-dissipative. We have

$$\sum_{j=0}^{N} p_{ij} \geq 1-\varepsilon \quad \text{for} \quad N \geq N(\varepsilon) \quad \text{and all } i.$$

Therefore

$$\sum_{j=0}^{N} p_{ij}^{(n)} = \sum_{j=0}^{N} \sum_{k=0}^{\infty} p_{ik}^{(n-1)} p_{kj}$$

$$= \sum_{k=0}^{\infty} p_{ik}^{(n-1)} \sum_{j=0}^{N} p_{kj} \geq 1-\varepsilon.$$

Therefore also

$$\sum_{j=0}^{N} q_{ij}^{(n)} \geq 1-\varepsilon.$$

Let $n \to \infty$. We obtain

$$\sum_{j=0}^{N} \pi_{ij} \geq 1-\varepsilon \qquad \text{for all } i.$$

Therefore $$\sum_{j=0}^{\infty} \pi_{ij} \geq 1-\varepsilon \qquad \text{for all } i.$$

But since ε is arbitrarily small, we must have

$$\sum_{j=0}^{\infty} \pi_{ij} = 1 \qquad \text{for all } i.$$

In my paper (1951) such a matrix was called quasi-finite. It was also shown there that it has the properties, (i) that there are only a finite number of latent vectors corresponding to the latent value $\lambda = 1$, (ii) that the latent values, λ, with $|\lambda| = 1$ are finite in number. We shall not here re-derive these properties.

It is precisely these matrices for which the row-summations are non-uniform in i which occur in the applications.

§3. Some theory of recurrent events.

The theory of "recurrent events", due to Feller (1949, 1950), is based on a new Tauberian theorem on power series proved jointly by Erdös, Feller and Pollard (1949). We derive here with special reference to Markov chains a few elementary results which properly belong to this theory but which do not depend on the Tauberian theorem.

Define

$$f_{ij}^{(1)} \equiv p_{ij} \quad ,$$

and then, recursively,

$$f_{ij}^{(n)} \equiv p_{ij}^{(n)} - f_{ij}^{(n-1)} p_{jj}^{(1)} - \cdots - f_{ij}^{(1)} p_{jj}^{(n-1)}. \qquad (3.1)$$

Then $f_{ij}^{(n)}$ is the probability that the system, initially in the i^{th} state, attains the j^{th} state <u>for the first time</u> at the n^{th} step. (When $i = j$, $f_{jj}^{(n)}$ is the probability of a <u>first return</u> at the n^{th} step.) The probability that the j^{th} state is eventually attained is now

$$F_{ij} \equiv \sum_{n=1}^{\infty} f_{ij}^{(n)}. \qquad (3.2)$$

This interpretation of F_{ij} may be justified in the usual way in terms of the probability-measure on the sequence space of the random variables $\{z_n\}$.

When $F_{ij} = 1$, we define the <u>mean first passage time</u> from the i^{th} to the j^{th} state as

$$\mu_{ij} \equiv \sum_{n=1}^{\infty} n f_{ij}^{(n)}.$$

When $i = j$, F_{jj} is the probability of the eventual return of the system to the j^{th} state, and (when $F_{jj} = 1$) μ_{jj} is the <u>mean recurrence time</u>. In Chapter 2 we defined recurrent and transient states. It follows now that the j^{th} state is recurrent or transient according as $F_{jj} = 1$ or $F_{jj} < 1$.

Defining $p_{jj}^{(0)} \equiv 1$, we may write (3.1) in the

more symmetrical form

$$p_{ij}^{(n)} = \sum_{r=1}^{n} f_{ij}^{(r)} p_{jj}^{(n-r)} . \qquad (3.3)$$

Using (3.1), we can also easily establish by induction that

$$f_{ij}^{(n+1)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n)} . \qquad (3.4)$$

THEOREM 3.1 $F_{jj} = 1$ **if and only if** $\sum_{n=1}^{\infty} p_{jj}^{(n)}$ **diverges.**

Proof. To simplify notation, define $p_{jj}^{(n)} \equiv a_n$, $f_{jj}^{(n)} \equiv b_n$. Then from (3.3),

$$a_n = \sum_{r=1}^{n} b_r a_{n-r} .$$

Define

$$A_n = \sum_{r=0}^{n} a_r , \qquad B_n = \sum_{r=1}^{n} b_r ,$$

$$A = \lim_{n \to \infty} A_n , \qquad B = \lim_{n \to \infty} B_n .$$

We want to show that $B = 1$ if and only if $A = \infty$.

We have for $n \geq 1$,

$$A_n - 1 = \sum_{r=0}^{n-1} a_r B_{n-r} .$$

Therefore $A_n - 1 \leq B_n A_{n-1}$

Therefore $A - 1 \leq BA$

On the other hand, for any fixed $r < n$,

$$A_n - 1 \geq B_{n-r} A_r .$$

Therefore $A - 1 \geq B A_r$ for all r.

Therefore $A - 1 \geq B A$.

Combining this with the previous inequality, we obtain

$$B = \frac{A - 1}{A} ,$$

and the required result follows.

THEOREM 3.2 $F_{jj} = 1$ <u>if and only if</u> $F_{kj} = 1$ <u>for every state</u> $k \neq j$ <u>which is accessible from</u> j.

Proof. From (3.4) by summing with respect to n, we obtain

$$F_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} F_{kj} .$$

Therefore, in particular

$$F_{jj} = p_{jj} + \sum_{k \neq j} p_{jk} F_{kj} .$$

Now if $F_{kj} = 1$ for every state $k \neq j$ which is accessible from j, then in particular $F_{kj} = 1$ whenever $p_{jk} > 0$. Therefore

$$F_{jj} = \sum_{k=0}^{\infty} p_{jk} = 1 .$$

Conversely, suppose $F_{jj} = 1$. We have then

$$F_{ij} = \sum_{k=0}^{\infty} p_{ik} F_{kj} .$$

Therefore also

$$F_{ij} = \sum_{k=0}^{\infty} p_{ik}^{(n)} F_{kj} \qquad \text{for all } n.$$

Therefore

$$\sum_{k=0}^{\infty} p_{jk}^{(n)} F_{kj} = 1.$$

Therefore

$$\sum_{k=0}^{\infty} p_{jk}^{(n)} (1 - F_{kj}) = 0.$$

But $F_{kj} \leq 1$ always. Therefore

$$p_{jk}^{(n)} (1 - F_{kj}) = 0 \qquad \text{for all } n.$$

Therefore $F_{kj} = 1$ whenever $p_{jk}^{(n)} > 0$ for some n, i.e. whenever state k is accessible from j.

THEOREM 3.3 <u>For any transient state</u> j, $\pi_{ij} = 0$ <u>for all</u> i.

Proof. For any small ε, we can choose $N(\varepsilon)$ such that

$$\sum_{n=N+1}^{\infty} f_{ij}^{(n)} < \varepsilon \qquad \text{for } N > N(\varepsilon).$$

Therefore from (3.3) we have for $n > N$

$$p_{ij}^{(n)} - \varepsilon \leq \sum_{r=1}^{N} f_{ij}^{(r)} p_{jj}^{(n-r)} \leq p_{ij}^{(n)}.$$

But, by Theorem 1.1, the (C, 1)-limits, π_{ij}, of $p_{ij}^{(n)}$ exists as $n \to \infty$. Therefore

$$\pi_{ij} - \varepsilon \leq \pi_{jj} \sum_{r=1}^{N} f_{ij}^{(r)} \leq \pi_{ij}.$$

Letting $N \to \infty$ we obtain

$$\pi_{ij} = \pi_{jj} F_{ij} ,$$

and, in particular,

$$\pi_{jj} = \pi_{jj} F_{jj} .$$

Therefore $F_{jj} < 1$ implies that $\pi_{jj} = 0$ which in turn implies that $\pi_{ij} = 0$.

THEOREM 3.4 $\mu_{jj} \equiv 1 / \pi_{jj}$.

Proof. For μ_{jj} to be defined we must assume $F_{jj} = 1$.

Define $F_{jj}^{(0)} \equiv F_{jj}$

and $F_{jj}^{(N)} \equiv \sum_{n=N+1}^{\infty} f_{jj}^{(n)}$.

Then we have $\mu_{jj} = \sum_{n=0}^{\infty} F_{jj}^{(n)}$.

From (3.3),

$$p_{jj}^{(1)} = (F_{jj}^{(0)} - F_{jj}^{(1)}) p_{jj}^{(0)} .$$

Therefore

$$1 = F_{jj}^{(0)} p_{jj}^{(0)} = F_{jj}^{(0)} p_{jj}^{(1)} + F_{jj}^{(1)} p_{jj}^{(0)} ,$$

since $F_{jj}^{(0)} = 1$.

Now, by induction, we can prove that for all n,

$$\sum_{r=0}^{n} F_{jj}^{(r)} p_{jj}^{(n-r)} = 1.$$

Suppose first $\mu_{jj} < \infty$. Then we can choose $N(\varepsilon)$ such that

$$\sum_{n=N+1}^{\infty} F_{jj}^{(n)} < \varepsilon \quad \text{for} \quad N > N(\varepsilon).$$

Therefore $$1 - \varepsilon \leq \sum_{r=1}^{N} F_{jj}^{(r)} P_{jj}^{(n-r)} \leq 1.$$

Therefore, by an argument similar to that of Theorem 3.3,

$$\pi_{jj} \mu_{jj} = 1.$$

Moreover, when $\mu_{jj} = \infty$ it will be seen that

$$\pi_{jj} \mu_{jj} \leq 1.$$

Therefore when $\mu_{jj} = \infty$, $\pi_{jj} = 0$ so that in all cases

$$\mu_{jj} = 1 / \pi_{jj}.$$

There does not, in general, exist such a simple expression for μ_{ij}. We can however prove the weaker

THEOREM 3.5 *If both* $\pi_{ij} > 0$ *and* $\pi_{ji} > 0$, *then* $\mu_{ij} < \infty$.

Proof. Denote by $g_{ij}^{(n)}$ the probability that the system, initially in the i^{th} state, will leave this state at once and without re-entering it at any intermediate step, will find itself in the j^{th} state (not necessarily for the first time) at the n^{th} step. Clearly, when $i \neq j$,

$$f_{jj}^{(n+m)} \geq g_{ji}^{(m)} f_{ij}^{(n)},$$

and $\pi_{ji} > 0$ implies that for some m, $g_{ji}^{(m)} > 0$.

For such an m,

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)} \leq \sum_{n=1}^{\infty} n f_{jj}^{(n+m)} / g_{ji}^{(m)}$$

$$\leq \sum_{n=1}^{\infty} (n+m) f_{jj}^{(n+m)} / g_{ji}^{(m)}$$

$$\leq \mu_{jj} / g_{ji}^{(m)} .$$

But $\pi_{jj} \geq \pi_{ji} \pi_{ij} > 0$.

Therefore, by Theorem 3.4, $\mu_{jj} < \infty$ and so $\mu_{ij} < \infty$.

EXAMPLE 3.1

$$[p_{ij}] \equiv \begin{bmatrix} \mu_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & 0 & \lambda_1 & 0 & \cdots \\ \mu_2 & 0 & 0 & \lambda_2 & \cdots \\ \mu_3 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Let us suppose for simplicity that all the μ_i, λ_i are non-zero. We can show that the zero state is recurrent if and only if $\sum_{i=0}^{\infty} \mu_i = \infty$. We have

$$f_{00}^{(1)} = \mu_0$$

$$f_{00}^{(2)} = \lambda_0\mu_1$$

$$f_{00}^{(3)} = \lambda_0\lambda_1\mu_2 \quad , \quad \text{etc.}$$

Therefore $$\sum_{n=1}^{N} f_{00}^{(n)} = \mu_0 + \lambda_0\mu_1 + \cdots + \lambda_0\lambda_1\cdots\lambda_{N-2}\mu_{N-1}$$

$$= 1 - \lambda_0\lambda_1\cdots\lambda_{N-1} \quad .$$

Therefore $\sum_{n=1}^{\infty} f_{00}^{(n)} = 1$ if and only if

$\prod_{i=0}^{\infty} \lambda_i = 0$, i.e. if and only if $\sum_{i=0}^{\infty} \mu_i = \infty$.

§4. Stable distributions.

We saw in the case of a finite system that at least one stable distribution always existed. We have now the

THEOREM 4.1 A stable distribution exists if and only if the system is not completely dissipative.

Proof. If the system is not completely dissipative, then for some i, $0 < \sum_{j=0}^{\infty} \pi_{ij} = \lambda \leq 1$. It follows from (1.1) that $\{x_j\} = \{\frac{1}{\lambda}\pi_{ij}\}$ is a stable distribution. Conversely, suppose that $\{x_j\}$ is a stable distribution. Then

$$\sum_{i=0}^{\infty} x_i p_{ij}^{(n)} = x_j \qquad \text{for all n.}$$

Therefore $\sum_{i=0}^{\infty} x_i q_{ij}^{(n)} = x_j$ for all n

Therefore $\sum_{i=0}^{N} x_i q_{ij}^{(n)} = x_j - \sum_{i=N+1}^{\infty} x_i q_{ij}^{(n)}$

$\geq x_j - \sum_{i=N+1}^{\infty} x_i$

$> \varepsilon$ for some j and N.

Therefore $\sum_{i=0}^{N} x_i \pi_{ij} \geq \varepsilon$.

Therefore $\pi_{ij} > 0$ for some i, j.

THEOREM 4.2 Any stable distribution $\{x_j\}$ is such that $x_j = 0$ for all transient states j.

Proof. We have

$$\sum_{i=0}^{\infty} x_i \pi_{ij} = x_j$$

Now, by Theorem 3.3, $\pi_{ij} = 0$ for all transient states j. The result follows.

§5. Irreducible systems.

THEOREM 5.1 For an irreducible chain, either $\pi_{ij} > 0$ for all i, j, or $\pi_{ij} = 0$ for all i, j.

Proof. Suppose $\pi_{st} = 0$. For any i, j we can find integers M, N such that

$$p_{si}^{(M)} > 0 , \qquad p_{jt}^{(N)} > 0 .$$

Now we have

$$p_{st}^{(m+M+N)} \geq p_{si}^{(M)} p_{ij}^{(m)} p_{jt}^{(N)}$$

Therefore

$$0 = \pi_{st} \geq p_{si}^{(M)} \pi_{ij} p_{jt}^{(N)} .$$

Therefore $\pi_{ij} = 0$ for all i, j.

THEOREM 5.2 <u>If a stable distribution, $\{x_j\}$, of an irreducible chain exists, then it is unique and $x_j > 0$ for all j.</u>

Proof. We have

$$\sum_{i=0}^{\infty} x_i \pi_{ij} = x_j , \qquad j = 0, 1 \cdots ,$$

and, by Theorem 5.1, $\pi_{ij} > 0$ for all i, j. Therefore $x_j = 0$ would imply that $x_i = 0$ for all i. Therefore $x_i > 0$ for all i. Now suppose $\{\overline{x}_j\}$ is a second stable distribution. We can choose λ such that $x_0 - \lambda \overline{x}_0 = 0$. Define $u_i = x_i - \lambda \overline{x}_i$. Clearly

$$\sum_{i=0}^{\infty} u_i p_{ij} = u_j , \qquad j = 0, 1, \cdots .$$

Therefore

$$\sum_{i=0}^{\infty} u_i p_{ij}^{(n)} = u_j$$

Since $\sum |u_i| < \infty$, we also have

$$\sum_{i=0}^{\infty} u_i \pi_{ij} = u_j$$

Therefore $$\sum_{i=0}^{\infty} |u_i| \pi_{ij} \geq |u_j| .$$

But we must have equalities throughout since

$$\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |u_i| \pi_{ij} = \sum_{i=0}^{\infty} |u_i| \sum_{j=0}^{\infty} \pi_{ij} \leq \sum_{i=0}^{\infty} |u_i| .$$

Now $|u_0| = 0$. Therefore $|u_i| = 0$ for all i, so that $x_i \equiv \bar{x}_i$.

As an immediate corollary we have

THEOREM 5.3 An irreducible chain is regular and either $\pi_{ij} \equiv \pi_j > 0$ for all j or $\pi_j = 0$ for all j.

THEOREM 5.4 An irreducible chain is either completely dissipative or non-dissipative.

Proof. By (1.1),

$$\sum_{k=0}^{\infty} \pi_{ik} \pi_{kj} = \pi_{ij}$$

Therefore, since $\pi_{ij} \equiv \pi_j$,

$$\pi_j \sum_{k=0}^{\infty} \pi_k = \pi_j .$$

Therefore

$$\Big(\sum_{j=0}^{\infty} \pi_j \Big)^2 = \sum_{j=0}^{\infty} \pi_j$$

Therefore either $\sum_{j=0}^{\infty} \pi_j = 1$ or $\sum_{j=0}^{\infty} \pi_j = 0$

As an immediate corollary of this and Theorem 4.1, we have

THEOREM 5.5 An irreducible chain is non-dissipative if and only if there exists a stable distribution.

EXAMPLE 5.1 Consider again the matrix of Example 3.1. The equations for the stable distribution $\{x_j\}$ are

$$x_0 = \mu_0 x_0 + \mu_1 x_1 + \cdots$$
$$x_1 = \lambda_0 x_0$$
$$x_2 = \lambda_1 x_1 , \quad \text{etc.}$$

Therefore

$$x_1 = \lambda_0 x_0$$
$$x_2 = \lambda_0 \lambda_1 x_0$$
$$x_3 = \lambda_0 \lambda_1 \lambda_2 x_0 , \quad \text{etc.}$$

Therefore a stable distribution exists and so the system is non-dissipative if and only if

$$\sum_{n=0}^{\infty} \lambda_0 \lambda_1 \cdots \lambda_n < \infty .$$

This should be compared with the weaker condition of Example 3.1,

$$\lambda_0 \lambda_1 \cdots \lambda_n \longrightarrow 0 ,$$

which was necessary and sufficient for recurrence.

EXAMPLE 5.2

$$[p_{ij}] \equiv \begin{bmatrix} \mu_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & 0 & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & 0 & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Let us again suppose that $\mu_i > 0$, $\lambda_i > 0$ for all i. The equations for the stable distribution $\{x_j\}$ are

$$\begin{aligned} \mu_0 x_0 + \mu_1 x_1 &= x_0 \\ \lambda_0 x_0 + \mu_2 x_2 &= x_1 \\ \lambda_1 x_1 + \mu_3 x_3 &= x_2 \quad , \text{ etc.} \end{aligned}$$

It is easily proved by induction that

$$x_n = \lambda_0 x_0 \cdot \frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} .$$

Therefore a stable distribution exists if and only if

$$\sum_{n=1}^{\infty} \frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty .$$

In particular, when $\mu_i = \mu$, $\lambda_i = \lambda$ constants for all i, this condition becomes simply $\mu > \lambda$.

EXAMPLE 5.3

$$[p_{ij}] \equiv \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

The stable distribution $\{x_j\}$ is found to be such that

$$x_j = x_0 (p_j + p_{j+1} + \cdots) .$$

Therefore it exists if and only if

$$\sum_{j=1}^{\infty} j p_j < \infty .$$

i.e. if and only if the mean jump from the zero state is finite. It will be seen from inspection that even when it is infinite, every state is recurrent.

It is not always so easy to calculate the stable distribution as it is in these examples. The following criterion for non-dissipation is then often useful.

THEOREM 5.6 <u>An irreducible chain is non-dissipative if and only if for some integer i_0 there exists a non-negative solution of the system of "inequations"</u>

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y_i - 1 , \qquad i > i_0 , \qquad (5.1)$$

such that $\sum_{j=0}^{\infty} p_{ij} y_j < \infty$ for $i \le i_0$. (5.2)

Proof. Let $\sum_{j=0}^{\infty} p_{ij} y_j \le y$ for $i \le i_0$.

Define $y_i^{(n+1)} \equiv \sum_{j=0}^{\infty} p_{ij}^{(n)} y_j$, $y_i^{(1)} \equiv y_i$.

Then

$$y_i^{(n+2)} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{ij}^{(n)} p_{jk} y_k$$

$$= \sum_{j=0}^{\infty} p_{ij}^{(n)} y_j^{(2)}$$

$$\le y \sum_{j=0}^{i_0} p_{ij}^{(n)} + \sum_{j=i_0+1}^{\infty} (y_j - 1) p_{ij}^{(n)}$$

$$= y \sum_{j=0}^{i_0} p_{ij}^{(n)} + y_i^{(n+1)} - \sum_{j=0}^{i_0} p_{ij}^{(n)} y_j - 1 + \sum_{j=0}^{i_0} p_{ij}^{(n)}$$

$$\le (y+1) \sum_{j=0}^{i_0} p_{ij}^{(n)} + y_i^{(n+1)} - 1 .$$

It follows that if $y_i^{(n+1)}$ **is finite, then so also is** $y_i^{(n+2)}$. **But** $y_i^{(2)}$ **is finite for all i. Therefore** $y_i^{(n)}$ **is finite for all i and n. We have now a recurrence relation from which we obtain**

$$y_i^{(n+2)} \leq (y+1) \sum_{r=1}^{n} \sum_{j=0}^{i_0} p_{ij}^{(r)} + y_i^{(2)} - n .$$

Therefore $$\frac{1}{n} y_i^{(n+2)} \leq (y+1) \sum_{j=0}^{i_0} q_{ij}^{(n)} + \frac{1}{n} y_i^{(2)} - 1 .$$

Letting $n \to \infty$ we obtain

$$0 \leq (y+1) \sum_{j=0}^{i_0} \pi_{ij} - 1 .$$

Therefore $$\sum_{j=0}^{i_0} \pi_{ij} \geq 1/(1+y) .$$

Then $\pi_{ij} > 0$ for some i, j. Therefore by Theorem 5.4, the chain is non-dissipative.

The necessary condition for non-dissipation may be given a sharper form. We therefore state and prove the result separately as

THEOREM 5.7 <u>Given an irreducible non-dissipative chain, choose any state i_0 and define</u>

$$y_i \equiv \mu_{i i_0} , \quad i \neq i_0 ,$$
$$y_{i_0} \equiv 0 .$$

<u>Then</u> $$\sum_{j=0}^{\infty} p_{ij} y_j = y_i - 1 , \quad i \neq i_0$$

<u>and</u> $$\sum_{j=0}^{\infty} p_{i_0 j} y_j < \infty .$$

Proof. Choose $i_0 = 0$. (The proof is the same for any state.) Since for all i, $\pi_{i0} > 0$, $\pi_{0i} > 0$ it

follows from Theorem 3.5 that $\mu_{i0} < \infty$ for all i.

We have now for all i

$$\sum_{j=0}^{\infty} p_{ij} y_j = \sum_{j=1}^{\infty} p_{ij} \sum_{n=1}^{\infty} n f_{j0}^{(n)}$$

$$= \sum_{n=1}^{\infty} n f_{i0}^{(n+1)}$$

$$= \sum_{n=1}^{\infty} (n+1) f_{i0}^{(n+1)} - \sum_{n=1}^{\infty} f_{i0}^{(n+1)}$$

$$= \sum_{n=1}^{\infty} n f_{i0}^{(n)} - \sum_{n=1}^{\infty} f_{i0}^{(n)}$$

$$= \mu_{i0} - 1 .$$

Therefore

$$\sum_{j=0}^{\infty} p_{ij} y_j \begin{cases} = y_i - 1 , & i \neq 0 \\ < \infty , & i = 0 . \end{cases}$$

This is essentially given by Feller (1950, p.335. Note however that the sentence following his equations (8.13) is inaccurate, as Theorem 5.6 above shows: there does not exist a solution of the equations when $\mu_{ii_0} = \infty$].

EXAMPLE 5.4 Consider again Example 5.3. We have

$$\sum_{j=0}^{\infty} p_{0j} j = \sum_{j=1}^{\infty} j p_j ,$$

and $$\sum_{j=0}^{\infty} p_{ij} \cdot y_j = i - 1 .$$

Therefore, when $\sum j p_j < \infty$, the solution $y_j = j$ satisfies the conditions of Theorem 5.6, and so the matrix is non-dissipative. This example shows incidentally that the condition,

$$\sum_{j=0}^{\infty} p_{ij} y_j < \infty \quad \text{for} \quad i \leq i_0 ,$$

cannot be relaxed.

It may here be noted that the condition (5.1) is equivalent to

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y_i - \lambda , \quad i > i_0 , \qquad (5.3)$$

where $\lambda > 0$. For we may clearly divide through by λ to obtain (5.1) again. The following example shows that (5.3) cannot be relaxed to read:

$$\sum_{j=0}^{\infty} p_{ij} y_j < y_i , \quad i > i_0 .$$

EXAMPLE 5.5

$$[p_{ij}] \equiv \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \cdots \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & \cdots \\ 0 & \frac{3}{5} & 0 & \frac{2}{5} & 0 & \cdots \\ 0 & 0 & \frac{4}{7} & 0 & \frac{3}{7} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

This is a particular case of Example 5.2 in which we have put for $n \geq 1$:

$$\lambda_n \equiv n/(2n+1) \quad , \quad \mu_n \equiv (n+1)/(2n+1) .$$

Therefore

$$\frac{\lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \frac{2n+1}{n+1} \cdot > \frac{3}{2} \cdot \frac{1}{n} .$$

Therefore $\sum \frac{\lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \infty$, so that, by Example 5.2, no stable distribution exists and the system is completely dissipative.

It will be observed however that

$$\sum_{j=0}^{\infty} p_{oj}\, j = \frac{1}{2} < \infty ,$$

and

$$\sum_{j=0}^{\infty} p_{ij}\, j = i - \frac{1}{2i+1} < i , \quad \text{for } i \geq 1 .$$

§6. Simply reducible systems.

One type of reducible chain occurs so often in applications that it is convenient to have a name for it. A Markov chain will be said to be simply reducible if the system of states consists of a finite irreducible closed set C and a set D of intermediate states from each of which C is accessible. It will be observed that, for each i, the expression

$$\sum_{j \in C} p_{ij}^{(n)}$$

gives the probability that the system, initially in the i^{th} state, attains the set C on or before the n^{th} step. Clearly this is monotone non-decreasing with n, so that we have

$$\sum_{j \in C} \pi_{ij} = \lim_{n \to \infty} \sum_{j \in C} p_{ij}^{(n)} ,$$

and this is the probability that the set C is eventually attained. Since each state of D is transient, we have, by Theorem 3.3, that $\pi_{ij} = 0$ for all $j \in D$ and for all i. Therefore for each i

$$\sum_{j=0}^{\infty} \pi_{ij} = \sum_{j \in C} \pi_{ij} ,$$

and we have the

THEOREM 6.1 A simply reducible chain is non-dissipative if and only if for every initial state, the system is almost certain eventually to attain the closed set of states.

A simply reducible chain can clearly never be completely dissipative.

An important application of the idea of simple reducibility is to the problem of determining whether or not the states of an irreducible system are recurrent (Cf. Bartlett, 1950). To investigate the nature of the j^{th} state, we modify the irreducible system by making

this state absorbing. The modified system is then simply reducible, the closed set consisting of the j^{th} state alone. Now in the original irreducible system, every state is accessible from the j^{th} state. Therefore, by Theorem 3.2, the j^{th} state is recurrent ($F_{jj} = 1$) if and only if there is probability one that it is attained from each state i ($F_{ij} = 1$ for all $i \neq j$). By Theorem 6.1, this will be so if and only if the modified system is non-dissipative. We state this formally as

THEOREM 6.2 Any state of an irreducible chain is recurrent if and only if the corresponding simply reducible chain, obtained by making this state absorbing, is non-dissipative.

In this connection, the following result, given by Feller (1950), may also be noted.

THEOREM 6.3 All states of an irreducible chain are of the same type: either all are recurrent or all are transient.

Proof. There exist finite integers M, M', N, N' such that for any pair of states i, j

$$p_{ij}^{(M)} > 0, \quad p_{ji}^{(M')} > 0, \quad p_{ji}^{(N)} > 0, \quad p_{ij}^{(N')} > 0.$$

But $p_{ij}^{(M)} p_{jj}^{(n)} p_{ji}^{(N)} \leq p_{ii}^{(M+N+n)}$

and $p_{ji}^{(M')} p_{ii}^{(n)} p_{ij}^{(N')} \leq p_{jj}^{(M'+N'+n)}$

Therefore $\sum_{n=1}^{\infty} p_{ii}^{(n)}$ and $\sum_{n=1}^{\infty} p_{jj}^{(n)}$ converge or diverge together. The result now follows from Theorem 3.1.

We proceed now to study some criteria for non-dissipation connected with the possible solutions of the system of equations:

$$\sum_{j=0}^{\infty} p_{ij} y_j = y_i , \quad i = 0, 1, \cdots , \qquad (6.1)$$

and, more generally, of the system of "inequations":

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y_i , \quad i = 0, 1, \cdots .$$

Three types of solution will be distinguished:

(i) y_i constant for all i. This is the <u>trivial</u> solution: it always exists.

(ii) y_i bounded but not constant.

(iii) $y_i \to \infty$ as $i \to \infty$, $|y_i| < \infty$ for all i. We shall call this a <u>properly divergent</u> solution.

Since in all cases $\{k + y_i\}$ is also a solution for any non-negative constant k, there is no loss in generality

in assuming in all three cases (as we shall) that

$$0 \leq y_i < \infty \quad \text{for all } i .$$

THEOREM 6.4 <u>A simply reducible chain is dissipative if and only if there exists a bounded non-constant solution of the equations</u> (6.1).

Proof. The necessity is obvious. For we have

$$\sum_{k=0}^{\infty} p_{ik}\,\pi_{kj} = \pi_{ij} , \qquad i = 0, 1, \cdots .$$

Define

$$y_i \equiv \sum_{j \in C} \pi_{ij} .$$

Now, by Theorem 6.1, the system can only be dissipative if $y_i < 1$ for some i. But $y_i \equiv 1$ for $i \in C$. Therefore $\{y_i\}$ is the desired solution.

Conversely, suppose the system (6.1) has a bounded non-constant solution $\{y_i\}$. By Theorem 2.6.4, since C is a finite irreducible closed set, y_i = constant for $i \in C$. Therefore we may suppose

$$y_i \equiv 1 \quad \text{for } i \in C,$$

$$\text{with} \quad 0 \leq y_i \leq 2 \quad \text{for all } i.$$

Now we have for all n,

$$\sum_{j=0}^{\infty} p_{ij}^{(n)}\, y_j = y_i .$$

Therefore

$$\sum_{j=0}^{\infty} \pi_{ij}\, y_j \leq y_i .$$

Therefore $$\sum_{j=0}^{\infty} \pi_{ij} = \sum_{j \in C} \pi_{ij} \leq y_i .$$

Since $\{y_i\}$ is non-constant, we must have for some i, either $y_i < 1$ or $y_i > 1$. In the former case we have the desired result that $\sum \pi_{ij} < 1$. In the latter case, by consideration of the new solution $\{2 - y_i\}$ we reach the same conclusion.

The following is a slight variant.

THEOREM 6.5 <u>A simply reducible chain is dissipative if and only if there exists a bounded solution</u> $\{y_i\}$ <u>of the inequations (6.2), with</u> $y_j < y_i$ <u>for some</u> $j \in D$, $i \in C$.

Proof. Only the sufficiency part of the previous proof requires alteration. We note first that it is still the case that $y_i =$ constant for $i \in C$. For we have

$$\sum_{j \in C} p_{ij}^{(n)} y_j \leq y_i , \qquad i \in C$$

Therefore

$$\sum_{j \in C} \pi_{ij} y_j \leq y_i , \qquad i \in C$$

But for $i, j \in C$, $\pi_{ij} \equiv \pi_j > 0$, $\sum_{j \in C} \pi_j = 1$.

Therefore

$$\sum_{j \in C} \pi_j y_j \leq y_i , \qquad i \in C .$$

This is only possible if $y_i =$ constant for $i \in C$. We may now assume $y_i \equiv 1$ for $i \in C$, and we have $y_i < 1$ for some i. Therefore, for this i,

$$\sum_{j=0}^{\infty} \pi_{ij}\, y_j \;<\; 1 .$$

Therefore

$$\sum_{j=0}^{\infty} \pi_{ij} \;=\; \sum_{j \in C} \pi_{ij} \;<\; 1 .$$

EXAMPLE 6.1

$$[p_{ij}] \;\equiv\; \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \mu_1 & 0 & \lambda_1 & 0 & 0 & \cdots \\ \mu_2 & 0 & 0 & \lambda_2 & 0 & \cdots \\ \mu_3 & 0 & 0 & 0 & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} .$$

This is the simply reducible system corresponding to the irreducible system of Example 3.1 in which the zero state has become absorbing. We showed there that the condition,

$$\sum_{i=1}^{\infty} \mu_i \;=\; \infty ,$$

was necessary and sufficient for the zero state to be recurrent. We may re-derive this result using Theorems 6.2 and 6.4. We find easily, by induction, that the solution of the equations (6.1) is

$$y_{n+1} \;=\; y_0 \;+\; (y_1 - y_0) \,/\, \lambda_1 \lambda_2 \cdots \lambda_n , \qquad n = 1, 2 \cdots ,$$

and y_0, y_1 are arbitrary. It follows that no bounded non-constant solution exists if and only if

$\lambda_1 \lambda_2 \cdots \lambda_n \to 0$, i.e. if and only if $\sum \mu_i = \infty$.

The following theorem applies to any Markov chain, but finds its most important application in connection with simply reducible systems.

THEOREM 6.6 <u>If a properly divergent solution of the inequations (6.2) exists, then the chain is non-dissipative; and further it possesses a finite closed set of states.</u>

Proof. Since $y_i \to \infty$ as $i \to \infty$, there exists a minimum value of y_i which we may take to be zero. Moreover $y_i = 0$ for only a finite number of values of i. Let C be this set of i. Then we have for $i \in C$

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq 0 .$$

Therefore $p_{ij} = 0$ for $i \in C$, $j \notin C$.

Therefore $\sum_{j \in C} p_{ij} = 1$ for $i \in C$,

and so C is closed.

Now we have

$$\sum_{j=0}^{\infty} p_{ij}^{(n)} y_j \leq y_i \quad \text{for all } i .$$

Let $u_N \equiv \text{Min} [y_{N+1}, y_{N+2}, \cdots]$.

Then $u_N > 0$ for all sufficiently large N, and as $N \to \infty$. We have for all i and n,

$$u_N \sum_{j=N+1}^{\infty} p_{ij}^{(n)} \leq y_i .$$

Therefore $$\sum_{j=0}^{N} p_{ij}^{(n)} \geq 1 - y_i / u_N .$$

Therefore $$\sum_{j=0}^{N} \pi_{ij} \geq 1 - y_i / u_N .$$

Let $N \to \infty$. We obtain $\sum_{j=0}^{\infty} \pi_{ij} \geq 1$. But the sum on the left cannot exceed unity. Therefore $\sum_{j=0}^{\infty} \pi_{ij} = 1$ for all i.

This theorem was proved for the particular properly divergent sequence, $y_i \equiv i$ in my paper (1951)[a]. The above extension is due to D.G.Kendall (1951)[a]. We now proceed to establish a converse for the particular case of a simply reducible chain. In addition, we shall make the following two weak assumptions:

<u>Each pair of states, i and j D are interaccessible, i.e.</u> $p_{ij}^{(n)} > 0$ <u>for some n.</u> (6.3)

<u>There exists a number</u> δ, $0 \leq \delta < 1$, <u>such that if</u> $t_{i(\delta)}$ <u>denotes the number of positive integers</u> t <u>for which</u> $\sum_{j \in C} p_{ij}^{(t)} \leq \delta$, <u>then</u> $t_i \to \infty$ <u>as</u> $i \to \infty$. (6.4)

These could be weakened in various ways, but will almost

invariably be satisfied in any practical application. (6.4) has a simple intuitive meaning. The sum $\sum_{j \in C} p_{ij}^{(t)}$ is the probability that the system, initially in the i^{th} state, is absorbed into the closed set C within t steps, and is non-decreasing with t. $t_i(\delta)$ is the largest number of steps which may be taken before the probability of such absorption exceeds δ. We may think of $t_i(\delta)$ as a measure of the "remoteness" of the i^{th} state from C, and (6.4) asserts that the remoteness of the i^{th} state tends to infinity with i. When the system is non-dissipative, $\sum_{j \in C} p_{ij}^{(t)} \to 1$ as $t \to \infty$, so that t_i is finite for each i.

THEOREM 6.7 <u>If a simply reducible chain is non-dissipative and satisfies (6.3) and (6.4), then there exists a properly divergent solution of the inequations (6.2).</u>

Proof. We shall suppose, for simplicity, that C consists of a single absorbing state, which we may take to be zero state, so that $p_{00} = 1$. The extension to the more general case is immediate. We have, by Theorem 6.1, $F_{i0} = 1$ for all i.

Consider first the simple special case when the finiteness of μ_{i0} is assumed. Defining $y_i \equiv \mu_{i0}$

$i \neq 0$, $y_0 \equiv 0$, we could show (Cf. the more general proof below) that $\{y_i\}$ is a properly divergent solution. However, the finiteness of μ_{i0} excludes most of the interesting cases, and we therefore adopt a more general procedure.

To simplify notation, define $f_i^{(n)} \equiv f_{i0}^{(n)}$.

The series, $$\sum_{n=1}^{\infty} f_i^{(n)} = 1 , \quad i \neq 0 ,$$

is non-terminating. For suppose all terms after the n^{th} were zero. The system, initially in the i^{th} state, is then almost certain to attain the zero state in n steps. Therefore, from (6.4), the system must almost certainly be in the zero state at the $(n-1)^{st}$ step. This leads to a contradiction if $i \neq 0$. Now define

$$\lambda^{(r)} \equiv 1 / \sqrt{\left\{ \sum_{s=r}^{\infty} f_1^{(s)} \right\}} .$$

Then $\lambda^{(r)}$ is finite and positive for each r. By a theorem of Dini (Cf. Knopp, p.293), we have

$$\sum_{r=1}^{\infty} \lambda^{(r)} f_1^{(r)} < \infty .$$

Define

$$y_0 \equiv 0 ,$$

$$y_i \equiv \sum_{r=1}^{\infty} \lambda^{(r)} f_i^{(r)} , \quad i \neq 0 .$$

We prove that $\{y_i\}$ is a properly divergent solution of (6.2).

Clearly $y_i \geq 0$ always. We show that y_i is finite for all i. By (6.3), for some integer t, $p_{ij}^{(t)} > 0$. From (3.4), since $p_{00} = 1$, we obtain by a simple induction that

$$f_i^{(t+r)} = \sum_{k=1}^{\infty} p_{ik}^{(t)} f_k^{(r)}$$

$$\geq p_{ij}^{(t)} f_j^{(r)} .$$

Therefore $$y_j \leq \sum_{r=1}^{\infty} \lambda^{(r)} f_i^{(t+r)} / p_{ij}^{(t)} .$$

But $\lambda^{(r)}$ is monotone non-decreasing with r. Therefore

$$y_j \leq \sum_{r=1}^{\infty} \lambda^{(t+r)} f_i^{(t+r)} / p_{ij}^{(t)}$$

$$\leq y_i / p_{ij}^{(t)} < \infty .$$

To prove that $y_i \to \infty$ as $i \to \infty$, we make use of (6.4). The integer t_i is zero for only a finite number of values of i. We need therefore only consider the behaviour of y_i when i is such that $t_i \geq 1$. For such values of i,

$$y_i \geq \sum_{r=t_i+1}^{\infty} \lambda^{(r)} f_i^{(r)}$$

$$\geq \lambda^{(t_i)} \left(1 - p_{i0}^{(t_i)} \right)$$

$$\geq \lambda^{(t_i)} (1 - \delta) .$$

But $\lambda^{(n)} \to \infty$ with n and $t_i \to \infty$ with i. Therefore $y_i \to \infty$ with i.

Finally, we prove that $\{y_i\}$ satisfies the inequations (6.2). For $i \neq 0$, since $y_0 = 0$,

$$\sum_{j=0}^{\infty} p_{ij} y_j = \sum_{j=1}^{\infty} p_{ij} \sum_{r=1}^{\infty} \lambda^{(r)} f_j^{(r)} = \sum_{r=1}^{\infty} \lambda^{(r)} f_i^{(r+1)}$$

$$\leq \sum_{r=1}^{\infty} \lambda^{(r+1)} f_i^{(r+1)} \leq y_i .$$

Moreover,

$$\sum_{j=0}^{\infty} p_{0j} y_j = p_{00} y_0 = y_0 .$$

This completes the proof. To extend it to the more general case of a finite closed set C, we define $y_i \equiv 0$ for $i \in C$, and only a few verbal alterations are required.

EXAMPLE 6.2

$$[p_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & \cdots \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

This is a particular case of Example 6.1. It is of interest in showing that (6.4) is not a <u>necessary</u> condition for the existence of a properly divergent solution of (6.2). For (6.4) is here clearly <u>not</u> satisfied. On the other hand, $\{i\}$ $(i = 0, 1, 2 \cdots)$ is a properly divergent solution of (6.2) and (6.3) is also clearly satisfied.

EXAMPLE 6.3

$$[p_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \mu_1 & 0 & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & 0 & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & 0 & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

This is the simply reducible system corresponding to the irreducible system of Example 5.2, in which the zero state

has become absorbing. (6.3) and (6.4) are both satisfied. Therefore, by Theorems 6.6 and 6.7, the system is non-dissipative if and only if there exists a properly divergent solution of the inequations (6.2). These are:

$$\mu_{i+1} y_i + \lambda_{i+1} y_{i+2} \leq y_{i+1}, \quad i = 0, 1, \cdots .$$

They may be solved. We have

$$\lambda_{i+1} (y_{i+2} - y_{i+1}) \leq \mu_{i+1} (y_{i+1} - y_i) .$$

Therefore

$$y_{i+2} - y_{i+1} \leq \frac{\mu_{i+1}}{\lambda_{i+1}} (y_{i+1} - y_i)$$

$$\leq \frac{\mu_1 \mu_2 \cdots \mu_{i+1}}{\lambda_1 \lambda_2 \cdots \lambda_{i+1}} (y_1 - y_0) .$$

Therefore $$y_{i+2} \leq y_1 + \sum_{n=1}^{i+1} \nu_n (y_1 - y_0),$$

wherein $$\nu_n \equiv \frac{\mu_1 \mu_2 \cdots \mu_n}{\lambda_1 \lambda_2 \cdots \lambda_n} .$$

The existence of a properly divergent solution implies that $\sum \nu_n$ diverges. Conversely, if $\sum \nu_n$ diverges, a solution for the corresponding system of <u>equations</u>, which may be found as above, is properly divergent (if we take $y_0 = 0$, $y_1 = 1$).

We have therefore the result that the system is

non-dissipative if and only if

$$\sum_{n=1}^{\infty} \frac{\mu_1 \mu_2 \cdots \mu_n}{\lambda_1 \lambda_2 \cdots \lambda_n} = \infty .$$

By Theorem 6.2, this is the necessary and sufficient condition for the states of the irreducible matrix of Example 5.2 to be recurrent. It is instructive to compare this condition with the condition we obtained in Example 5.2 that the irreducible matrix should be non-dissipative. When $\mu_i = \mu$, $\lambda_i = \lambda$ the above condition is satisfied if and only if $\mu \geq \lambda$ (In Example 5.2, we required the stronger condition, $\mu > \lambda$).

The above has a random walk interpretation (Cf. Section 2.10). A particle moves on the integers, $i = 0, 1, 2, \cdots$, in such a way that when it is at $i \neq 0$ it has probabilities μ_i, λ_i of stepping respectively to $i - 1$, $i + 1$, with $\lambda_i + \mu_i = 1$. When at $i = 0$, it has probabilities μ_0, λ_0 of either remaining where it is or stepping to $i = 1$. We have above the necessary and sufficient condition that all positions are recurrent. In particular, the symmetric random walk (having $\lambda_i = \mu_i = \frac{1}{2}$) is such that all positions are recurrent. Since however the condition of Example 5.2 is not satisfied, the system is dissipative, so that the mean

recurrence times are infinite. We can also have recurrence even when there is a "drift" away from the zero position. Consider, for example, $\lambda_i \equiv (i+1)/(2i+1)$, $\mu_i \equiv i/(2i+1)$. The divergence of $\sum n^{-1}$ shows at once that the positions are recurrent (with of course infinite mean recurrence times).

EXAMPLE 6.4 We consider now a generalised <u>symmetric</u> random walk on the integers

$$\cdots, -2, -1, 0, 1, 2, \cdots$$

When it is in position i, suppose the particle has a probability p_j of moving in one step to position i + j. Further suppose that

$$\phi_j \equiv \phi_{-j} > 0 \qquad \text{for all } j.$$

To see whether the positions are recurrent, we consider the corresponding simply reducible system in which we make the zero state absorbing. The modified system has the matrix $[p_{ij}]$ with

$$\phi_{00} = 1, \qquad \phi_{ij} \equiv \phi_{i-j}, \quad i \neq 0.$$

This matrix is two-way enumerably infinite, the states being numbered from $-\infty$ to $+\infty$. This however raises no special difficulty, and the system is non-dissipative if there exists a solution of

$$\sum_{j=-\infty}^{\infty} p_{i-j}\, y_j \leq y_i , \quad i \neq 0 ,$$

which is properly divergent in the sense that $y_j \to \infty$ as $|j| \to \infty$, with $|y_j| < \infty$ for all j.

It is possible to prove indeed that for certain distributions, $\{p_j\}$, the system is dissipative, so that no properly divergent solution exists. We shall not now consider the general problem (Cf. Chapter 5), but shall examine only one particular case in which we put

$$p_j \equiv e^{-|j|}/\lambda ,$$

λ being a normalizing constant. We can show that there exists a constant μ such that

$$y_j \equiv |j| + \mu , \quad j \neq 0 , \qquad y_0 \equiv 0 ,$$

is a properly divergent solution of the equations. We have for any $i \neq 0$,

$$\sum_{j=-\infty}^{\infty} p_{i-j}\, y_j = \mu (1 - p_i) + |i| + 2 \sum_{j=|i|}^{\infty} (j - |i|) p_j$$

$$= |i| + \mu - \frac{\mu}{\lambda} e^{-|i|} + \frac{2}{\lambda} e^{-|i|} \sum_{j=1}^{\infty} j e^{-j} .$$

We now obtain the required result if we choose

$$\mu = 2 \sum_{j=1}^{\infty} j e^{-j}$$

The following is a variant of Theorem 6.6.

THEOREM 6.8 Any Markov chain is non-dissipative if for some integer i_0 and some constant m, $0 < m < 1$, a properly divergent solution, $\{y_j\}$, exists of the system of inequations

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq m y_i , \qquad i > i_0 ,$$

with

$$\sum_{j=0}^{\infty} p_{ij} y_j < \infty , \qquad i \leq i_0 .$$

Proof. We may suppose

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y \qquad \text{for } i \leq i_0 .$$

Therefore for all i,

$$\sum_{j=0}^{\infty} p_{ij} y_j \leq y + m y_i .$$

Define $$\rho_i^{(n)} \equiv \sum_{j=0}^{\infty} p_{ij}^{(n)} y_j .$$

Then $$\rho_i^{(n+1)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} \sum_{j=0}^{\infty} p_{kj} y_j$$

$$\leq \sum_{k=0}^{\infty} p_{ik}^{(n)} (y + m y_k)$$

$$= y + m \rho_i^{(n)}$$

It follows by induction that

$$p_i^{(n)} \leq y \sum_{r=0}^{n-1} m^r + m^n \cdot y_i$$

$$< y_i + \frac{y}{1-m} = k_i \text{ , say,}$$

where k_i is independent of n.

Define $u_N \equiv \min [y_{N+1}, y_{N+2}, \cdots]$.

Then for large enough N, $u_N > 0$ and $u_N \to \infty$ as $N \to \infty$. We have now for any i,

$$u_N \sum_{j=N+1}^{\infty} p_{ij}^{(n)} \leq k_i .$$

Therefore

$$\sum_{j=0}^{N} p_{ij}^{(n)} \geq 1 - k_i / u_N .$$

Therefore

$$\sum_{j=0}^{N} \pi_{ij} \geq 1 - k_i / u_N .$$

Letting $N \to \infty$, we obtain

$$\sum_{j=0}^{\infty} \pi_{ij} \geq 1 .$$

But the sum on the left cannot exceed unity, so the result follows.

EXAMPLE 6.5

$$[p_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \lambda_0 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \cdots \\ \lambda & 0 & 0 & \mu & 0 & \cdots \\ \lambda & 0 & 0 & 0 & \mu & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

Suppose that $\sum_{j=1}^{\infty} j\,\lambda_j < \infty$.

We have then $\sum_{j=0}^{\infty} p_{ij}\, j < \infty$ for all i.

Now $$\sum_{j=0}^{\infty} p_{ij}\, j = \mu\,(i+1) \qquad (i \geq 2)$$

$$= \frac{i+1}{i} \cdot \mu \cdot i .$$

Now, given $\mu < 1$ there exists an integer i_0 and a constant m, $0 < m < 1$, such that

$$\frac{i+1}{i}\,\mu \leq m \quad \text{for} \quad i > i_0.$$

We have then

$$\sum_{j=0}^{\infty} p_{ij}\, j < \infty \quad \text{for} \quad i \leq i_0 ,$$

$$\sum_{j=0}^{\infty} p_{ij}\, j \leq m\, i \quad \text{for} \quad i > i_0 ,$$

and so the system is non-dissipative. It follows that,

assuming $\lambda_0 \neq 0$, the zero state is almost certainly eventually attained from any state.

§7. Periodicity.

The methods of Sections 2.8 and 2.9 do not carry over to chains with an enumerable infinity of states. However the result of Section 2.9 remains true: *the set of states of any irreducible chain may be divided into a finite number, m (≥ 1) of sub sets, $S_1, S_2, \cdots, S_m$, such that the system in a state of S_r is almost certainly transferred in a one-step transition into S_{r+1} (with $S_{m+1} \equiv S_1$).* When $m = 1$, *the chain is aperiodic.* A proof, based on the Erdös-Feller-Pollard theorem (1949), is given by Feller (1950). It will not here be reproduced.

It will now be seen that Theorem 2.9.1 carries over without modification, and so we are provided with a useful criterion of aperiodicity.

§8. A Markov chain derivation of discrete distributions.

We have seen that, when the chain is irreducible, either $\pi_{ij} \equiv 0$ or there exists a unique stable distribution $\{x_j\}$ with $\pi_j \equiv x_j > 0$ for all j. In this way to every irreducible stochastic matrix A there

corresponds a vector, $\{x_j\}$, which is either null or represents a probability distribution. Conversely, to any probability distribution, $\{x_j\}$, (with $x_j > 0$ for all j), it can be shown that there corresponds an irreducible stochastic matrix A (not necessarily unique). We give now a simple method of constructing such a matrix A.

Let $\{a_i\}$ $(i = 0, 1, 2, \cdots)$ be a sequence of positive numbers and define $A_n \equiv \sum_{i=0}^{n} a_i$ $(n = 0, 1, 2, \cdots)$. Now define

$$A \equiv \begin{bmatrix} a_0/A_1 & a_1/A_1 & 0 & 0 & \cdots \\ a_0/A_2 & a_1/A_2 & a_2/A_2 & 0 & \cdots \\ a_0/A_3 & a_1/A_3 & a_2/A_3 & a_3/A_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Then A is clearly an irreducible stochastic matrix. Since its diagonal elements are all positive, it is also aperiodic. Now consider the equations

$$x A = x .$$

The solution is easily obtained. We have

$$x_0/a_0 = x_1/a_1 = x_0/A_1 + x_1/A_2 + \cdots ,$$

$$x_2/a_2 = x_1/A_2 + x_2/A_3 + \cdots ,$$

$$x_3/a_3 = x_2/A_3 + x_3/A_4 + \cdots ,$$

etc.

Therefore $$x_1 = \frac{a_1}{a_0} x_0 ,$$

and $$x_{n+1} = a_{n+1}\left(\frac{x_n}{a_n} - \frac{x_{n-1}}{A_n}\right), \quad n \geq 1.$$

Therefore, by induction,

$$x_n = x_0 \cdot \frac{a_1 a_2 \cdots a_n}{A_0 A_1 \cdots A_{n-1}}, \quad n \geq 1.$$

The solution is therefore uniquely determined, apart from a common factor. For it to represent a probability distribution, we must have

$$x_0\left(1 + \sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}}\right) = 1.$$

It follows that the matrix

$$\Pi \equiv \lim_{n \to \infty} A^n$$

is non-null if and only if

$$\sum_{n=1}^{\infty} \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} < \infty .$$

When $\prod$ is non null, each row will consist of the probability distribution $\{x_j\}$.

Now suppose the distribution $\{x_j\}$ is given. To calculate the sequence $\{a_j\}$ with the above property, we have only to put

$$\frac{a_n}{A_{n-1}} = \frac{x_n}{x_{n-1}} , \qquad n = 1, 2, \cdots .$$

Then we find, as required, that

$$x_n = x_0 \cdot \frac{a_1 \cdots a_n}{A_0 \cdots A_{n-1}} .$$

We have now

$$A_n - A_{n-1} = a_n = A_{n-1} \cdot \frac{x_n}{x_{n-1}} .$$

Therefore

$$A_n = A_{n-1} \left(1 + \frac{x_n}{x_{n-1}} \right) ,$$

and so, by iteration,

$$A_n = a_0 \prod_{i=1}^{n} \left(1 + \frac{x_i}{x_{i-1}} \right) .$$

Therefore, putting (as we may, since a common factor is unimportant) $a_0 = 1$, we have

$$a_n = \frac{x_n}{x_{n-1}} \prod_{i=1}^{n-1} \left(1 + \frac{x_i}{x_{i-1}} \right) , \qquad n \geq 1 .$$

We may give this procedure the following interpretation. Consider a particle performing a random walk on the integers 0, 1, 2, $\cdots$, in such a manner that when in position n it has probabilities in the ratios,

$$a_0 : a_1 : \cdots : a_{n+1} ,$$

of jumping in a one step-transition into one of the positions 0, 1, 2, $\cdots$, n + 1. In other words, the particle can move either one step forward or back to any previous position. The distribution $\{x_j\}$ may then be interpreted as giving the asymptotic probabilities for its position after a large number of steps. We have seen how the sequence $\{a_j\}$ may be calculated to give any required asymptotic distribution $\{x_j\}$ (with $x_j > 0$ for all j).

In the following two examples we consider cases where $\{x_j\}$ is a recognised distribution and the sequence $\{a_j\}$ has a particularly simple form.

EXAMPLE 8.1 <u>The Poisson distribution.</u>

Let $x_n \equiv e^{-\lambda}\lambda^n/n!$, $\lambda > 0$, $n = 0, 1, \cdots$.

We find that

$$a_n = \lambda(\lambda+1)\cdots(\lambda+n-1)/n! , \quad n = 1, 2, \cdots ,$$

with $a_0 = 1$.

The n^{th} row of A is thus a truncated Negative-Binomial distribution having n + 1 terms. In particular, when $\lambda = 1$, A takes the very simple form

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \cdots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

and the rule governing the motion of the particle is that when it is in the n^{th} position it has equal probabilities of jumping into any one of the positions 0, 1, 2, $\cdots$ n + 1.

EXAMPLE 8.2 The Negative-Binomial distribution.

Let $$x_n \equiv (1-\beta)^{\lambda}\beta^n \lambda(\lambda+1)\cdots(\lambda+n-1)/n!, \quad n = 1, 2, \cdots,$$

$$x_0 \equiv (1-\beta)^{\lambda},$$

where $0 < \beta < 1$ and $\lambda > 0$.

We find that

$$a_n = \frac{\lambda+n-1}{n!}(1+\lambda\beta)(2+\overline{\lambda+1}\,\beta)\cdots(n-1+\overline{\lambda+n-2}\,\beta), \quad n = 1, 2, \cdots,$$

with

$$a_0 = 1.$$

In particular, when $\lambda = 1$,

$$a_m = (1+\beta)^{m-1}\beta ,$$

$$a_0 = 1 ,$$

and each row of A is a truncated Modified-Geometric distribution.

CHAPTER 4

APPLICATIONS TO RENEWAL, BRANCHING AND QUEUEING PROCESSES

§1. Renewal processes.

We consider an element such as an electric light bulb, whose life-span is finite and is a random variable assuming discrete values (seconds, days, years). As soon as the element fails it is replaced by a new element of the same kind. Let $\{a_n\}$ $(n = 1, 2, \dots)$ be the probability distribution for its life-span. Define

$$A_n \equiv \sum_{i=n+1}^{\infty} a_i , \quad \mu_n \equiv a_{n+1}/A_n , \quad \lambda_n \equiv A_{n+1}/A_n .$$

Then μ_n is the probability that an element of age n fails at once (and is replaced by an element of zero age), and λ_n is the probability that it lasts at least to age $n+1$. We have now a Markov chain whose states are the possible ages of an element: $0, 1, 2, \dots$. The stochastic matrix is one previously considered, namely:

$$\begin{bmatrix} \mu_0 & \lambda_0 & 0 & 0 & \cdot\cdot \\ \mu_1 & 0 & \lambda_1 & 0 & \cdot\cdot \\ \mu_2 & 0 & 0 & \lambda_2 & \cdot\cdot \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

If we consider a whole population of elements of like kind, each of which is replaced at once on failure, we might enquire as to the statistical equilibrium distribution of ages. If this exists, it will be given by the stable distribution of the Markov chain. We have seen (Example 3.5.1) that this is $\{x_n\}$, where

$$x_0 = 1/\left(1 + \sum_{n=0}^{\infty} \lambda_0\lambda_1 \cdots \lambda_n\right)$$

and $$x_n = x_0 . \lambda_0\lambda_1 \cdots \lambda_n , \qquad n \geq 1 .$$

Now $$\lambda_0\lambda_1 \cdots \lambda_n = \frac{A_1}{A_0} . \frac{A_2}{A_1} \cdots \frac{A_{n+1}}{A_n} = A_{n+1} .$$

Therefore $$1 + \sum_{n=0}^{\infty} \lambda_0\lambda_1 \cdots \lambda_n = \sum_{n=0}^{\infty} A_n$$

$$= \sum_{n=1}^{\infty} n a_n$$

$$= \mu ,$$

where μ is the mean life-span. We have now $x_n = A_n/\mu$, and a stable distribution exists if and only if the mean life-span, μ, is finite.

There exists a large literature on Renewal Theory. Reference may be made to Fréchet (1949) for a bibliography.

§2. Branching processes.

A population of individuals gives rise to a new population, the probability that one individual gives rise to 0, 1, 2, ... new individuals being $a_0, a_1, a_2, \ldots$. These probabilities are the same for all individuals and are independent. The new population forms a first generation which in turn produces a second generation, and so on. Only the individuals of the n^{th} generation are supposed capable of producing the $(n+1)^{st}$ generation. We shall assume $a_0 > 0$.

The process described has an application to chain re-actions in nuclear fission. The individuals are the neutrons, which are subject to chance hits by other particles. On fission each neutron gives birth to a number, m, of direct descendants. The more homely, but similar, problem of the random growth and extinction of families carrying a given surname appears to have been the first branching process to be studied. The problem was suggested by Francis Galton in 1873 and treated in the following year by a probability method by the Rev.H.W. Watson (1874). Further references and historical remarks may be found in Kendall (1949). A complete solution to Galton's problem appears to have been first given by Steffensen (1930, 1933). Define

$$F(x) \equiv a_0 + a_1 x + a_2 x^2 + \cdots .$$

Steffensen showed that there is always a positive probability of the ultimate extinction of the population (consisting initially of one individual) and that it is given by the smallest root of the equation,

$$F(x) = x ,$$

in the interval $a_0 \leq x \leq 1$. Ultimate extinction is almost certain if and only if the mean number of individuals produced by one individual,

$$\rho \equiv \sum_{j=1}^{\infty} j\, a_j$$

is less than or equal to unity. A recent (independent) proof was given by Good (1949). Similar proofs may be found in Bartlett (1949) and Feller (1950, p.226). We give here a new proof of the result and show how it may be extended in certain respects.

The population sizes constitute the states of a Markov chain, p_{ij} denoting the probability that a population of size i reproduces one of size j. It is easily seen that

$$p_{ij} \equiv \text{coefficient of } x^j \text{ in } [F(x)]^i ,$$

and in particular, $p_{00} = 1$. Since $p_{i0} = a_0^i > 0$, for all i ; it follows that the matrix $[p_{ij}]$ is simply

reducible, the closed set consisting of the zero state alone. We have now

$$\rho = \sum_{j=1}^{\infty} j a_j = \left[\frac{d}{dx} F(x) \right]_{x=1} .$$

Suppose first that $\rho \leq 1$. Then for all i,

$$\sum_{j=0}^{\infty} j p_{ij} = \left[\frac{d}{dx} \{ F(x) \}^i \right]_{x=1}$$

$$= i.\rho \leq i .$$

Therefore, by Theorem 3.6.6, the system is non-dissipative. Since it is also simply reducible, it follows, by Theorem 3.6.1, that the system, initially in any state, will almost certainly eventually attain the zero state. In other words, the eventual extinction of the population is almost certain.

Now suppose that $\rho > 1$. We have

$$F(0) = a_0 > 0 , \quad F(1) = 1 , \quad F'(1) = \rho > 1 .$$

Moreover F(x) and F'(x) are both monotone non-decreasing as x increases from 0 to 1. Therefore there exists a unique root, ξ, $a_0 \leq \xi < 1$, of the

equation

$$F(x) = x .$$

Therefore

$$\sum_{j=0}^{\infty} p_{ij} \xi^j = [F(\xi)]^i = \xi^i , \quad i = 0, 1, \cdots .$$

By Theorem 6.6.4, the system is dissipative; we obtain however slightly more information by proceeding directly. We have

$$\sum_{j=0}^{\infty} p_{ij}^{(n)} \xi^j = \xi^i , \quad i = 0, 1, \cdots .$$

The series $\{\xi^j\}$ is absolutely convergent. Therefore

$$\sum_{j=0}^{\infty} \pi_{ij} \xi^j = \xi^i , \quad i = 0, 1, \cdots .$$

But because the system is simply reducible, $\pi_{ij} = 0$ for $i, j \neq 0$. Therefore

$$\pi_{i0} = \xi^i ,$$

and we have in particular that the probability of the eventual extinction of a population of initially one individual is ξ.

Let us drop now the assumption that the individuals reproduce independently. We consider, that is, a matrix, $[p_{ij}]$, which is subject only to the condition that it is simply reducible with $p_{00} = 1$. The individuals of a generation of i individuals are now supposed to interact with each other, so that the resulting probability

distribution, $\{p_{ij}\}$ $(j = 0, 1, 2, \ldots)$, depends on i in an arbitrary way and bears no simple relation to $\{p_{1j}\}$. Define

$$\rho_i \equiv \sum_{j=1}^{\infty} j\, p_{ij} ,$$

the expected number of individuals produced by a population of i individuals. It will be seen that the condition,

$$\rho_i \leq i \qquad \text{for all } i ,$$

is still <u>sufficient</u> for the almost certain ultimate extinction of the population.

In certain situations, however, it would be more realistic to suppose that the <u>reproduction rate</u>

$$r_i \equiv \rho_i / i , \qquad r_0 \equiv 0 ,$$

is greater than one for small values of i, but that it falls off rapidly for large i. Let us suppose only that

$$r_i < \infty \qquad \text{for} \quad i \leq i_0 ,$$

but that

$$r_i \leq m < 1 \qquad \text{for} \quad i > i_0 .$$

Then it follows that the ultimate extinction of the population (of initially any number of individuals) is

almost certain. For we have

$$\sum_{j=0}^{\infty} j p_{ij} < \infty \quad \text{for} \quad i \leq i_0 ,$$

$$\sum_{j=0}^{\infty} j p_{ij} \leq m i \quad \text{for} \quad i > i_0 .$$

The result is now immediate on application of Theorems 3.6.1 and 3.6.8.

§3. Queueing processes.

We consider a queue of customers waiting to be served at a counter. "Customer" and "Counter" are here used in a technical sense, and the problems we discuss have many applications, e.g. to trunking in telephone engineering and to the "stacking" of aircraft.

We assume that customers arrive in the queue "at random", i.e. that the number of arrivals in time t is a Poisson variable of expectation t/a, say. The arrival-time, u, that is, the time which elapses between two consecutive arrivals, then has the negative-exponential distribution

$$e^{-u/a} . du/a \qquad (0 < u < \infty),$$

and successive u-variables will be statistically independent. The mean value of u is a. The service-time,

v, is the time which elapses while a particular customer is being served. It is natural to assume that successive service-times are statistically independent of one another and of the input. Let v have the distribution

$$dB(v) \qquad (0 < v < \infty).$$

Two special cases of importance are (i) where v has a negative-exponential distribution, and (ii) where v is a constant. Let b in all cases be the mean-value of v, which we shall assume to be finite.

The stochastic process we have described, evolving in continuous time, is in general "non-Markovian", in the sense that if we regard the queue-size as the state of the system, then a knowledge of previous states will in general modify the conditional probability distribution of future states given the present state. Kendall (1951[6]) has pointed out that, <u>provided the input is Poissonian</u>, the sequence of queue-sizes immediately after the departure of each customer forms a realization of a Markov chain. Define

$$k_r \equiv \frac{1}{r!}\int_0^\infty e^{-v/a}\left(\frac{v}{a}\right)^r dB(v), \qquad r = 0, 1, \cdots .$$

Then k_r is the probability of precisely r arrivals in a single service time. It is easy to prove that

$$\sum_{r=1}^{\infty} r k_r = b/a .$$

We define $\rho \equiv b/a$ as the <u>traffic intensity</u>. Denote by p_{ij} the probability that a queue of size i, left behind by a departing customer, becomes of size j after the next departure. Then the Markov chain which is thus associated with the queueing process has the stochastic matrix:

$$[p_{ij}] \equiv \begin{bmatrix} k_0 & k_1 & k_2 & k_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & k_0 & k_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Since $k_r > 0$ for all r, this chain is irreducible. Let us consider its behaviour after the lapse of a long time (or, equivalently, after the passage of a large number of customers), when we should expect the distribution of queue-sizes to have attained statistical equilibrium. Kendall has proved the following two theorems.

(i) The system is non-dissipative (so that a stable distribution exists) if and only if $\rho < 1$.

(ii) Each state is recurrent if and only if

$$\rho \leq 1 .$$

We now show how both these results may be obtained as consequences of the general theorems of Chapter 3.

Let us consider (i). To prove that $\rho < 1$ is sufficient for non-dissipation, we have only to note that

$$\sum_{j=0}^{\infty} p_{0j} j = \sum_{j=1}^{\infty} j k_j = \rho < \infty ,$$

and

$$\sum_{j=0}^{\infty} p_{ij} j = \sum_{j=0}^{\infty} k_j (i+j-1) = i-(1-\rho) , \quad i \neq 0.$$

Therefore the sequence, $\{y_j\}$, with $y_j \equiv j/(1-\rho)$ satisfies the conditions of Theorem 3.5.6, and the desired result follows. To prove that $\rho < 1$ is also necessary for non-dissipation, we first calculate the mean first-passage times, μ_{i0}, from the i^{th} to the zero state. On the assumption of non-dissipation, these are all finite by Theorem 3.3.5 and Theorem 3.5.1. Now, from the structure of the matrix, we see that to every sequence of states by which a queue of size one becomes of size zero there corresponds a sequence of equal probability by which a queue of size j becomes one of size j - 1.

Therefore $\mu_{10} \equiv \mu_{j,j-1}$. Moreover from state j, the zero state can be reached only via state j - 1. Therefore $\mu_{j0} = \mu_{j,j-1} + \mu_{j-1,0}$. It follows by an induction that

$$\mu_{j0} = j \cdot \mu_{10} , \qquad j \geq 1 .$$

Define

$$y_0 \equiv 0$$
$$y_j \equiv \mu_{j0} , \qquad j \geq 1 .$$

Then, by Theorem 3.5.7,

$$\sum_{j=0}^{\infty} p_{1j} y_j = y_1 - 1 .$$

Therefore

$$y_1 \sum_{j=1}^{\infty} j k_j = y_1 - 1 .$$

Therefore

$$\rho = 1 - 1/y_1 < 1 ,$$

since $0 < k_0 \leq y_1 < \infty$.

We turn now to the proof of (ii). By Theorem 3.6.3 it will be sufficient to consider the recurrence of the zero state, and for this purpose we consider the simply reducible system obtained by making the zero state absorbing. The matrix now is

$$[p_{ij}] \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & k_0 & k_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

and, by Theorem 3.6.2, we have only to show that $\rho \leq 1$ is necessary and sufficient for non-dissipation. Now if $\rho \leq 1$, we have

$$\sum_{j=0}^{\infty} j p_{ij} \leq i \qquad \text{for all } i.$$

Therefore, by Theorem 3.6.6, the system is non-dissipative. Conversely, suppose that $\rho > 1$. Following Kendall, define

$$K(x) \equiv \sum_{n=0}^{\infty} k_n x^n .$$

We then have, by an argument identical with that used in Section 2, that there exists a unique root, ξ, $k_0 \leq \xi < 1$, of the equation

$$K(x) = x .$$

Therefore

$$k_0 \xi^j + k_1 \xi^{j+1} + \cdots = \xi^{j+1} .$$

We have thus

$$\sum_{j=0}^{\infty} p_{0j} \xi^j = \xi^0 ,$$

$$\sum_{j=0}^{\infty} p_{ij} \xi^j = \xi^i , \qquad i \geq 1 .$$

Therefore $$\sum_{j=0}^{\infty} p_{ij}^{(n)} \xi^j = \xi^i , \qquad i = 0, 1, \cdots .$$

Since $\sum \xi^i$ is absolutely convergent, we have also

$$\sum_{j=0}^{\infty} \pi_{ij} \xi^j = \xi^i .$$

But, since the system is simply reducible with $p_{00} = 1$, $\pi_{ij} = 0$ for $i, j \neq 0$. Therefore

$$\pi_{i0} = \xi^i .$$

It follows that the zero state is not recurrent, and we have in addition the result that the probability of a queue of i eventually becomes zero is ξ^i. This may be compared with our previous result for branching processes in Section 2.

It is to be emphasized that we have been discussing a Markov chain <u>associated with</u> the queueing process. The latter, being of Non-Markovian type, is

evolving in continuous time. Intuition suggests that when $\rho < 1$ this also settles down to a régime of statistical equilibrium, possessing what would naturally be termed a stable distribution of queue-sizes. This stable distribution will not be the same as that for the Markov chain. This will be appreciated if it is observed that associated with the queueing process there is also a Markov chain whose states are queue-sizes immediately before a departure. If the original Markov chain has the stable distribution $\{\pi_j\}$ $(j = 0, 1, 2, \ldots)$, that for the second Markov chain will be

$$0, \ \pi_0, \ \pi_1, \ \cdots \ .$$

A simple heuristic argument now shows that the stable distribution (if this exists) for the queueing process itself will be

$$\tfrac{1}{2}\pi_0, \ \ \tfrac{1}{2}(\pi_0 + \pi_1), \ \ \tfrac{1}{2}(\pi_1 + \pi_2), \ \cdots \ .$$

The following example illustrates this remark.

EXAMPLE 3.1 We consider the special case where both the arrival-times and the service-times have negative exponential distributions, with means of respectively a and b, and $a > b$. On account of the characteristic property of the negative-exponential distribution, the queueing process itself is now "Markovian"

(Cf. Kendall, 1951[b]). We have in the matrix of the associated Markov chain

$$k_r = \left(\frac{\rho}{1+\rho}\right)^r \cdot \frac{1}{1+\rho}, \quad r = 0, 1, \cdots.$$

It may now be verified that

$$\pi_r = (1-\rho)\rho^r, \quad r = 0, 1, \cdots,$$

is the unique stable distribution.

Now let us consider the queueing process itself. The fact that it is "Markovian" makes it easy to calculate the stable distribution as follows. Confining our attention to those epochs at which jumps take place, we see that at any time the queue-size $i(\geq 1)$ has probabilities

$$\mu = \frac{a}{a+b}, \qquad \lambda = \frac{b}{a+b}$$

of becoming at the next jump respectively $i-1$, $i+1$; the zero queue-size having probability one of becoming of size one at the next jump. It is clear that the stable distribution for the queueing process will be identical with that for the simple Markov chain describing the system at the sequence of epochs which marks all the departures and arrivals. This chain has the stochastic matrix

$$[p_{ij}] \equiv \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ \mu & 0 & \lambda & 0 & 0 & \cdots \\ 0 & \mu & 0 & \lambda & 0 & \cdots \\ 0 & 0 & \mu & 0 & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

with $\lambda/\mu \equiv b/a \equiv \rho < 1$. We have considered a matrix of this type in Example 3.5.2. Its stable distribution is found to be

$$\tfrac{1}{2}(1-\rho),\ \tfrac{1}{2}(1-\rho)(1+\rho),\ \tfrac{1}{2}(1-\rho)(1+\rho)\rho,\ \cdots .$$

§4. An artificial realization of two queueing processes.

A realization of two queueing processes, systems A and B, was artificially constructed by a method similar to that used by Kendall (1950) for his "birth-and-death" process. System "A" was given a constant service-time of one unit. The service-times for "B" were a sample drawn from a negative-exponential distribution with a mean value of one unit. The same sample of arrival-times, from a negative-exponential distribution with a mean value of $^3/_2$, was fed into both systems. Thus in both A and B

the traffic intensity was $\rho = 2/3$, so that the systems were non-dissipative.

A sample of 105 service-times, v, for system B was constructed by using the transformation

$$x = e^{-v}$$

so that the variables x must be uniformly distributed in (0, 1), and could be read off (to four, or when necessary for uniform accuracy, to eight decimal places) from a table of " random numbers". The conversion to the variables v was then carried out by reading backwards from the table of the negative-exponential function of the National Bureau of Standards.

In the same manner, a sample of 105 arrival-time intervals, u, was constructed. The transformation required here was

$$x = e^{-\frac{2}{3}\cdot u} .$$

The actual working for system A was set out as indicated in Table 1. The sequence of arrival-time intervals, u, were recorded in the first column. The servic-times in the fourth column were constantly one unit for system A. The figures in the remaining columns

may now be calculated on the basis of this data. The arrival-times in the second column are calculated cumulatively from the first column. Thus the first customer arrives at time t = 0. He has not to wait for service, but is served immediately during one unit of time and departs at time t = 1·00, leaving behind a zero queue, since the second customer does not arrive till later, at time t = 2·19. This customer also has zero waiting time and departs at time t = 3·19, leaving behind a queue of one. Each customer thus departs at time:

$$t = 1 + \max(\text{previous arrival time}, \text{previous departure}).$$

To complete the sixth column we take the arrival time of a customer, then count the number of previous departure times which are <u>greater</u> than this number. To complete the seventh column, we take the departure-time of a customer, then count the number of succeeding arrival-time entries which are <u>less</u> than this number.

Table 1

u	Arr t	Wait	Ser	Dep t	Finds Q	Leaves Q
0	0	0.00	1	1.00	0	0
2.19	2.19	0.00	1	3.19	0	1
0.42	2.61	0.58	1	4.19	1	2
0.79	3.40	0.79	1	5.19	1	5
0.27	3.67	1.52	1	6.19	2	
0.77	4.44	1.75	1	7.19	2	
0.41	4.85	2.34	1	8.19	3	
0.05	4.90	3.29	1	9.19	4	
0.21	5.11	4.08	1	10.19	5	
2.20	7.31	2.88	1	11.19	3	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
0.45	142.44	2.21	1	145.65	3	2
0.12	142.56	3.09	1	146.65	4	1
2.37	144.93	1.72	1	147.65	2	1
2.49	147.42	0.23	1	148.65	1	0
1.74	149.16	0	1	150.16	0	-

The working for system B was set out in a similar way as indicated in Table 2.

Table 2

Arr t	Wait	Ser	Dep t	Finds Q	Leaves Q
0	0	1.38	1.38	0	0
2.19	0	6.90	9.09	0	
2.61	6.48	5.07	14.16	1	
3.40	10.76	1.26	15.42	2	
3.67	11.75	0.12	15.54	3	
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
142.44	1.87	0.01	144.32	1	1
142.56	1.76	2.33	146.65	2	1
144.93	1.72	0.37	147.02	1	0
147.42	0	1.10	148.52	0	0
149.16	0	1.37	150.53	0	-

Column one is the same sequence of arrival-times t as used for system A. The service-times in column three are now variable. The entries are made as before.

The resulting graphs of queue-sizes against all the recorded values of t are shown for both systems in

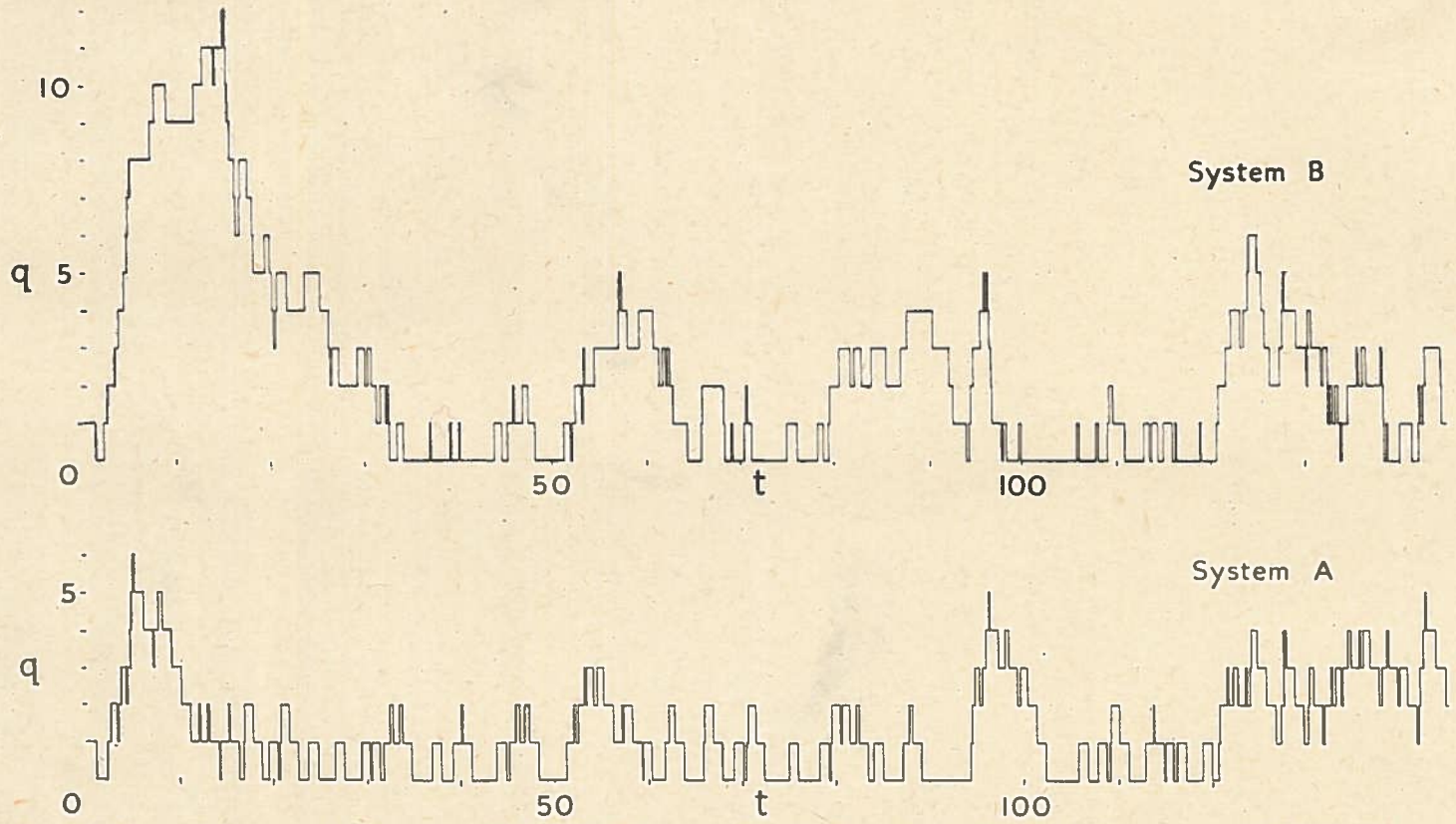


FIG. 1.—*The graphs compare the fluctuations with time (t) in the queue-sizes (q) of a realization of two congestion processes, A, B, having identical epochs of arrival, the time-intervals between consecutive arrivals being from a negative-exponential distribution with a mean value of ½ units; system A has a constant service-time of one unit, system B has service-times from a negative-exponential distribution with a mean value of one unit.*

the Figure 1. They illustrate the greater fluctuations to be expected when there is a variable service-time. It may be noted that system B is Markovian, while A is of the stationary Non-Markovian type. These graphs give the complete picture of the development of the processes in continuous time.

The difference in the efficiency of the two systems is brought out even more clearly in Figure 2, where the waiting times of the first 100 customers are compared. Theoretically, the mean waiting times for the two systems are in the ratio 1:2. On the other hand, the fraction of the total time ($1 - \rho = \frac{1}{3}$) during which the counter is empty should be the same for both systems. A measurement on Figure 1 shows this to be approximately so.

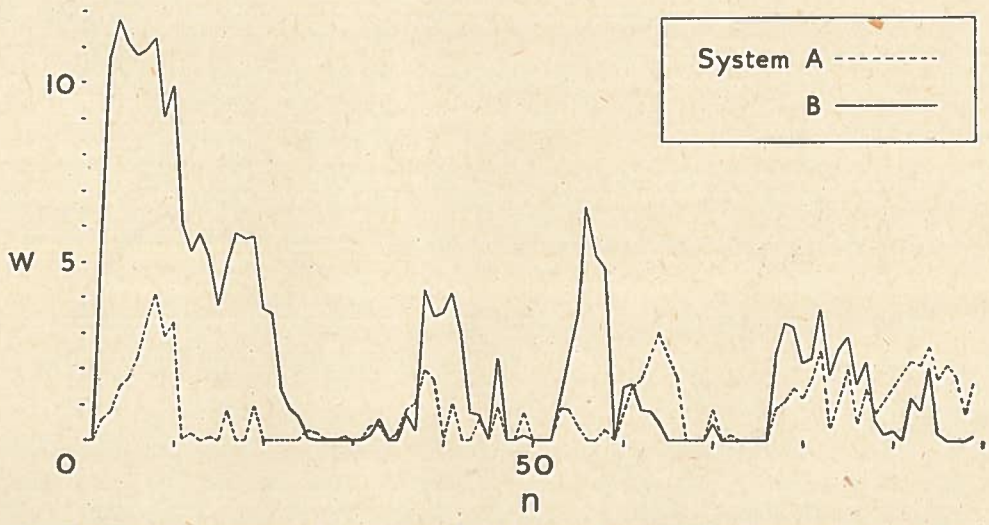


FIG. 2.—*The graphs compare the sequence of waiting-times (w) of 100 customers (n = 0 to 99), having identical arrival-times, for the two systems, A, B.*

The greater spread of queue-sizes to be expected in system B is illustrated in Table 3.

Table 3

Queue size	System A: Frequencies Observed	System A: Frequencies Expected	System A: Cumulated Expected Frequencies (%)	System B: Cumulated Expected Frequencies (%)	System B: Frequencies Expected	System B: Frequencies Observed
0	34	33·3	33·3	33·3	33·3	31
1	30	31·6	64·9	55·6	22·2	17
2	20	18·3	83·2	70·4	14·8	21
3	11	9·0	92·1	80·2	9·9	10
4	4	4·2	96·3	86·8	6·6	8
5	1	2·0	98·3	91·2	4·4	3
6		0·9	99·2	94·1	2·9	2
7		0·4	99·6	96·1	2·0	2
8		0·2	99·8	97·4	1·3	1
9		0·1	99·9	98·3	0·9	2
10				98·8	0·6	2
11				99·2	0·4	1
Mean queue size	1·2	1·3			2·0	2·2

The table shows the frequencies of different queue-sizes for systems A, B, observed after epochs of departures, and compares these with their expected values. The two central columns compare the cumulated expected frequencies (expressed as percentages) for the two systems.

CHAPTER 5

RANDOM WALKS AND DISCRETE HARMONIC FUNCTIONS

§0. This Chapter is again devoted to the study of a particular type of Markov chain, namely the random walk in n dimensions. Some related problems were previously touched upon in Section 2.10 and Examples 3.6.3 and 3.6.4. Our primary purpose is now to describe and prove a generalization of a classical theorem due to Pólya (1921). (An account of Pólya's theorem may be found in Feller's book, 1950.) The results we obtain are then used to establish certain existence theorems regarding possible solutions of the linear difference equations which arise in connection with this random walk problem. These equations are of some mathematical interest in themselves, apart from their probability interpretation, and the results generalize some theorems on discrete harmonic functions due to Heilbronn (1949).

§1. Pólya's random walk problem.

We consider a particle which moves on the points of an n-dimensional lattice. Each move carries it to one of the 2n nearest neighbours of its former position, and each move has the same probability, $1/2n$. Pólya's

theorem (1921) states that

<u>for n = 1 or 2, the particle will almost certainly eventually return to its initial position. In three or more dimensions there is a positive probability of non-return.</u>

A closely related difference between the two- and three-dimensional cases occurs in electrical networks. Consider a very large (infinite) network of wires of uniform resistance in the form of a two- or three- dimensional lattice, and suppose a steady current to enter at one intersection point of the lattice. Then in the two- dimensional case the potential difference between this point and any other selected point of the lattice will tend to infinity as the selected point tends to infinity. In the three-dimensional case on the other hand it will remain bounded. The analogy with the probability problem may be brought out as follows. Let the lattice in both cases have unit mesh, and let us take the two-dimensional case as typical. Choosing an origin of coordinates, let each point be labelled by a vector (x_1, x_2) with integer valued components. In the potential problem denote by $p(x_1, x_2)$ the potential at (x_1, x_2), the origin (0, 0) being arbitrarily given the potential $p(0, 0) = 1$. In the probability problem denote by

$p(x_1, x_2)$ the probability that the origin is eventually attained from the point (x_1, x_2), and let us define $p(0, 0)$ to be equal to one. Then in both interpretations $p(x_1, x_2)$ will satisfy the system of linear difference equations:

$$4\,p(x_1, x_2) = p(x_1+1, x_2) + p(x_1-1, x_2) + p(x_1, x_2+1) + p(x_1, x_2-1),$$
$$(x_1, x_2) \neq (0, 0),$$

with $p(0, 0) = 1$.

This system of equations is the discrete analogue of Laplace's equation:

$$\Delta f(x_1, x_2) \equiv \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} = 0,$$

for continuous variables x_1, x_2. For this reason any solution of the difference equations is said to be "discrete harmonic". As is well known the solutions of Laplace's equation (with a corresponding boundary condition) display in the two- and three-dimensional cases the same difference in asymptotic behaviour as that noted above in the case of the electrical network problem. Thus in two dimensions the potential at distance r from a mass point m is m (log 1/r), while in three dimensions

it is m/r.

The foregoing suggests that the probability problem might be investigated by means of a "discrete potential" theory. We shall in fact carry out such an investigation for a generalization of the random walk problem by means of a corresponding generalization of the concept of a discrete harmonic function.

§ 2. The generalized random walk.

We have a generalization of the Pólya random walk as follows. We consider as before an infinite n-dimensional lattice, but instead of restricting the particle to moves to nearest neighbours we suppose that in one step it may reach any point of the lattice, and according to a quite general probability distribution. This distribution will however always be supposed symmetrical in the sense that the probability of a move from a point A to a point B will be the same as that from B to A. It will be shown that, under very general conditions, Pólya's theorem, quoted above, still holds good for this much more general random walk. It is possible, by means of counter-examples, to show that Pólya's theorem is not in fact universally true for any symmetrical distribution, and that the conditions under which our theorems

are proved below cannot be substantially weakened. These counter-examples will be produced in a forthcoming paper by I.J.Good and the author.

We first introduce some notation. Let us label each point of the lattice by a vector

$$x \equiv (x_1, x_2, \cdots x_n),$$

where each x_i is a positive or negative integer or zero, in such a way that the lattice has unit mesh. Then we shall have an origin of coordinates

$$0 \equiv (0, 0, \cdots 0)$$

and it is convenient to define, for $1 \leq i \leq n$,

$$u^i \equiv -u^{-i} \equiv (0, 0, \cdots, 1, \cdots 0),$$

where the unit is in the i^{th} place. Further define

$$\|x\| \equiv \sum_{r=1}^{n} |x_r|$$

a measure of the distance of the point x from 0.

Let $\{f(x)\}$ denote a probability distribution over the lattice points x:

$$f(x) \geq 0, \quad \sum_x f(x) = 1.$$

As we shall in the sequel always assume this distribution symmetrical, we make once and for all the assumption:

$$f(x) = f(-x) \quad \text{for all } x.$$

We now identify f(y) with the probability that the particle, initially at any point x, moves in one step to the point x + y. We next introduce the probability, p(x), that the particle, initially at x, eventually reaches the origin 0. Underlying the assumption of the existence of this probability there is an infinite-product space whose points are the sequences

$$x^0, x', \cdots ,$$

where x^i represents the position of the particle at the i^{th} step (with $x^0 \equiv x$). In this space the distribution $\{f(y)\}$ defines a measure μ, and the probability p(x) is the measure of the set of paths

$$S : [x^i = 0 \text{ for some } i] .$$

This set S is evidently measurable, which ensures the existence of the probability in question:

$$p(x) \equiv \mu(S) ,$$

and clearly

$$p(0) = 1 .$$

For the proofs of the theorems below it is now convenient to introduce a modified random walk in which all points x with $\|x\| \geq m$ (m a positive integer) are supposed to be "absorbing". We then consider the

modified probability, $p_m(x)$, of reaching the origin from x. This is evidently given by the measure of the set

$$S_m : \left[x^i = 0 \text{ for some } i = i_0 ; \ \|x^i\| < m \text{ for } i < i_0 \right].$$

Thus $$\phi_m(x) \equiv \mu(S_m),$$

with $$\phi_m(0) = 1$$

and $$\phi_m(x) = 0 \quad \text{for} \quad \|x\| \geq m.$$

Since S_m is monotone non-decreasing with m, we have

$$\phi(x) = \mu(S) = \mu\left(\lim_{m\to\infty} S_m\right) = \lim_{m\to\infty} \mu(S_m) = \lim_{m\to\infty} \phi_m(x).$$

The probabilities $p_m(x)$ satisfy the relation

$$\phi_m(x) = \sum_y f(y)\phi_m(x+y), \quad 0 < \|x\| < m,$$

with the boundary conditions noted above, namely:

$$\phi_m(0) = 1$$

$$\phi_m(x) = 0, \quad \|x\| \geq m.$$

Defining therefore

$$\Delta_f \phi_m(x) \equiv \sum_y f(y)\{\phi_m(x) - \phi_m(x+y)\},$$

we have that

$$\Delta_f \phi_m(x) = 0, \quad 0 < \|x\| < m.$$

with $$\Delta_f \phi_m(0) = 1 - \sum_y f(y)\phi_m(y).$$

It will be observed that $\Delta_f \phi_m(0)$ may be interpreted as the probability that the particle <u>never</u> returns to the origin. We now derive an alternative expression for this probability which plays an important rôle in the sequel and corresponds to the Dirichlet Norm in Potential theory. Let

$$H(\phi_m) \equiv \frac{1}{2} \sum_y f(y) \sum_x \{\phi_m(x) - \phi_m(x+y)\}^2 .$$

We show that

(i) $H(\phi_m) \equiv \Delta_f \phi_m(0)$,

(ii) $H(\phi_m) \leq H(g_m)$ <u>where $g_m(x)$ is any lattice function satisfying the same boundary conditions as $p_m(x)$.</u>

To establish these properties we note that $H(p_m)$ may be regarded as a quadratic form in the finite number of variables, $p_m(x)$, $\|x\| < m$. Hence we may write

$$H(\phi_m) \equiv \frac{1}{2} \sum_{\|x\| < m} \frac{\partial H}{\partial \phi_m(x)} \cdot \phi_m(x) .$$

But, taking into account the symmetry of $\{f(y)\}$, we have for each x

$$\frac{\partial H}{\partial p_m(x)} = \sum_y f(y)\{p_m(x) - p_m(x+y)\} - \sum_y f(y)\{p_m(x-y) - p_m(x)\}$$

$$= 2\,\Delta_f p_m(x).$$

It follows that $H(p_m) \equiv \Delta_f p_m(0)$,

and $$\frac{\partial H}{\partial p_m(x)} = 0, \qquad 0 < \|x\| < m.$$

Now consider any $H(g_m)$ in which

$$g_m(x) = 0, \qquad \|x\| \geq m.$$

$H(g_m)$ is a positive-definite form in the variables $g_m(x)$, $\|x\| < m$. For, whenever $H(g_m) = 0$ we must have

$$g_m(x) = g_m(x+y) \quad \text{for all } x,$$

and for y such that $f(y) \neq 0$. Therefore

$$g_m(x) = g_m(x + ry)$$

for all integers r. But for some integer r, $\|x + ry\| \geq m$, so that $g_m(x + ry) = 0$. Therefore $g_m(x) = 0$ for all x. A standard argument now shows that $H(g_m)$ has a minimum, subject to the constraint

$$g_m(0) = 1.$$

The values of $g_m(x)$ which give this minimum will be among the solutions of

$$\Delta_f g_m(x) = 0, \qquad 0 < \|x\| < m .$$

But, with the given boundary conditions, this set of equations has the probabilities $p_m(x)$ as its unique solution. For suppose there were two solutions, $p_m(x)$ and $q_m(x)$, say. Then

$$g_m(x) \equiv p_m(x) - q_m(x)$$

would also be a solution with the boundary values

$$g_m(0) = 0, \qquad g_m(x) = 0, \quad \|x\| \geq m .$$

We can also suppose that $g_m(x) \geq 0$ for some x in the range $0 < \|x\| < m$. Since this set of points is finite, $g_m(x)$ will then attain its maximum in this set at some point u. It then follows that

$$g_m(u+y) = g_m(u)$$

whenever $f(y) \neq 0$. Therefore also (if $0 < \|u+y\| < m$)

$$g_m(u+2y) = g_m(u) .$$

But for some r, $\|u + ry\| \geq m$. Hence $g_m(u) = 0$. Therefore $p_m(x) \leq q_m(x)$, and a repetition of the same argument with the rôles of p_m and q_m interchanged gives $q_m(x) \leq p_m(x)$. Therefore $p_m(x) \equiv q_m(x)$. Property (ii) above now follows.

It is now clear that

$$\lim_{m\to\infty} H(\phi_m) = \lim_{m\to\infty} \Delta_f \phi_m(0) = 1 - \sum_y f(y)\phi(y)$$

is the probability of non-return to the origin in the unmodified random walk. We proceed to establish the following theorem, which shows that Pólya's one-dimensional result remains true for a very wide class of symmetrical random walks.

THEOREM 2.1 <u>In the one-dimensional random walk, provided that</u>

$$\sum_{y=m}^{\infty} f(y) = o(m^{-1}) \quad \text{as} \quad m \to \infty$$

<u>the particle will almost certainly eventually return to its initial position.</u>

Proof. We wish to show that

$$H(\phi_m) \to 0 \quad \text{as} \quad m \to \infty.$$

By the minimum property (ii) of $H(p_m)$ it will be sufficient to show that

$$H(g_m) \to 0 \quad \text{as} \quad m \to \infty,$$

where $g_m(x)$ is any function satisfying the same boundary conditions as $p_m(x)$, namely

$$g_m(0) = 1, \qquad g_m(x) = 0, \quad |x| \geq m.$$

We define

$$g_m(x) \equiv 1 - \frac{|x|}{m} , \quad |x| < m$$

$$g_m(x) \equiv 0 , \quad |x| \geq m .$$

For any fixed integer y with $|y| = r < m$, we have

$$\sum_x \{g_m(x) - g_m(x+y)\}^2 \leq \frac{1}{m^2} \sum_{x=-m+1}^{m-r-1} \{|x| - |x+r|\}^2 + 2 \sum_{x=m-r}^{m} \left(1 - \frac{x}{m}\right)^2$$

$$\leq r^2(2m-r-1)/m^2 - r(r+1)(2r+1)/3m^2$$

$$\leq K\left(r^2/m - r^3/m^2\right) ,$$

where K is an absolute constant. (K will not necessarily denote the same constant each time it occurs.) For <u>any</u> integer y,

$$\sum_x \{g_m(x) - g_m(x+y)\}^2 < 4m .$$

Therefore

$$H(\phi_m) \leq K\left\{m^{-1} \sum_{r=1}^{m-1} r^2 f(r) + m^{-2} \sum_{r=1}^{m-1} r^3 f(r) + m \sum_{r=m}^{\infty} f(r)\right\} .$$

Writing $F(n) \equiv \sum_{r=n}^{\infty} f(r)$ we have

$$\sum_{r=1}^{m-1} r^2 f(r) = \sum_{r=1}^{m-1} (2r-1) F(r) - (m-1)^2 F(m) .$$

By hypothesis, $mF(m) \to 0$ as $m \to \infty$.

Therefore

$$m^{-1} \sum_{r=1}^{m-1} r\, F(r) \to 0 \quad \text{as} \quad m \to \infty,$$

and so

$$m^{-1} \sum_{r=1}^{m-1} r^2 f(r) \to 0 \quad \text{as} \quad m \to \infty.$$

Moreover

$$m^{-2} \sum_{r=1}^{m-1} r^3 f(r) \leq m^{-1} \sum_{r=1}^{m-1} r^2 f(r).$$

Therefore, altogether, $H(p_m) \to 0$ as $m \to \infty$.

We turn now to the corresponding generalization of Pólya's two-dimensional result.

THEOREM 2.2 <u>In the two-dimensional random walk, provided that</u>

$$\sum_{\|y\| \geq m} f(y) = o(m^{-2}) \quad \text{as} \quad m \to \infty,$$

<u>the particle will almost certainly eventually return to its initial position.</u>

Proof. This is along the same lines. We define

$$g_m(x) \equiv 1 - \frac{\log(1 + \|x\|)}{\log(1+m)}, \quad \|x\| < m,$$

$$g_m(x) \equiv 0, \quad \|x\| \geq m.$$

Choosing now a fixed vector y, $\|y\| = r < m$ we perform the summation,

$$G_m(y) \equiv \sum_x \{ g_m(x) - g_m(x+y) \}^2 ,$$

as follows. We note first that $g_m(x)$ is constant over $\|x\| = t$ $(t = 0, 1, \cdots m-1)$. Thus we may write

$$g_m(x) \equiv g_m[t] \quad \text{when} \quad \|x\| = t .$$

Now in the summation, to each $g_m(x)$ there will correspond two partners, $g_m(x+y)$, $g_m(x-y)$, and at least one of the points, $x \pm y$, is not nearer the origin than X (the distance of x being defined as $\|x\|$). Choosing always a partner which is father out, and both when both are farther out, we perform the summation round each of the squares, $\|x\| = t$ in turn. This ensures that all terms in the summation are accounted for. Now $g_m[v] < g_m[u]$ for any integers $v > u$. Therefore

$$G_m(y) = \sum_{t=0}^{2m} \sum_{\|x\|=t} \{ g_m(x) - g_m(x+y) \}^2$$

$$\leq 2 \{ g_m[0] - g_m[r] \}^2 + 8 \sum_{t=1}^{m} t \{ g_m[t] - g_m[t+r] \}^2$$

$$\leq 2 \left\{ \frac{\log(1+r)}{\log(1+m)} \right\}^2 + 8 \sum_{t=1}^{m-r-1} t \left\{ \frac{\log(1+t)}{\log(1+m)} - \frac{\log(1+t+r)}{\log(1+m)} \right\}^2$$

$$+ \sum_{t=m-r}^{m} t \left\{ 1 - \frac{\log(1+t)}{\log(1+m)} \right\}^2$$

$$\leqslant \frac{8}{(\log m)^2}\left[\sum_{t=0}^{m-r-1}(1+t)\left\{\log\left(1+\frac{r}{1+t}\right)\right\}^2 + \sum_{t=m-r}^{m} t\left\{\log\left(\frac{1+t}{1+m}\right)\right\}^2\right]$$

$$\leqslant \frac{8}{(\log m)^2}\left[\sum_{t=0}^{m-r-1}\frac{r^2}{1+t} + \sum_{t=m-r}^{m} t\left(\frac{m-t}{1+t}\right)^2\right]$$

$$\leqslant K r^2/\log m .$$

To obtain the last line but one, we used the simple inequalities,

$$\log(1+x) < x \qquad (x > 0),$$

$$-\log(1-x) < \frac{x}{1-x} \qquad (0 < x < 1).$$

Now for all y, we have

$$G_m(y) < 10\, m^2 .$$

This follows from the fact that there are less than $5m^2$ lattice points in the domain $\|x\| \leqslant m$. We have therefore altogether

$$H(g_m) \leqslant \frac{K}{\log m}\sum_{r=1}^{m-1} r^2 \sum_{\|y\|=r} f(y) + K m^2 \sum_{r=m}^{\infty}\sum_{\|y\|=r} f(y).$$

Define $h(r) \equiv \sum_{\|y\|=r} f(y)$.

By hypothesis, $m^2 \sum_{r=m}^{\infty} h(r) \to 0$ as $m \to \infty$.

Therefore, defining $F(m) \equiv \sum_{r=m}^{\infty} h(r)$,

we have $m^2 F(m) \to 0$ as $m \to \infty$.

Now $$\sum_{r=1}^{m-1} r^2 \sum_{\|y\|=r} f(y) = \sum_{r=1}^{m-1} r^2 h(r) = \sum_{r=1}^{m-1} (2r-1)F(r) - (m-1)^2 F(m).$$

An application of the Cauchy-Toeplitz limit-theorem (Cf. Knopp, p.72) now shows that

$$\frac{1}{\log m} \sum_{r=1}^{m} \frac{1}{r} \{ r^2 F(r) \} \to 0 \quad \text{as} \quad m \to \infty.$$

It follows at once that $H(p_m) \to 0$ as $m \to \infty$.

We may note here that the condition under which Theorem 2.1 was proved would be satisfied if the first absolute moment, $\sum |y| f(y)$ were finite. Correspondingly, the finiteness of $\sum \|y\|^2 f(y)$ implies the condition of Theorem 2.2. It is possible to replace the $o(m^{-1})$ and $o(m^{-2})$ of the theorems by the weaker $O(m^{-1})$ and $O(m^{-2})$ (Cf. a forthcoming paper by I.J. Good and the author). The problem of finding weaker sufficient conditions and of finding necessary conditions

for almost certainty of return will not be further investigated here.

In three or more dimensions we can give a "best possible" result.

THEOREM 2.3 In the three- or higher-dimensional random walk, provided that $\{f(y)\}$ is at least a three-dimensional symmetrical distribution, there is a positive probability that the particle will never return to its initial position.

Remarks. In saying that the distribution $\{f(y)\}$ is of dimensionality greater than or equal to three, we mean that there exist three linearly independent vectors, say y^1, y^2, y^3, such that

$$f(y^1) > 0 \quad , \quad f(y^2) > 0 \quad , \quad f(y^3) > 0 \, .$$

If this were not the case the problem would be a one- or two-dimensional one, and the preceeding theorems would apply. No other condition is imposed upon $\{f(y)\}$ (apart from its symmetry).

Proof. It will be shown that the problem may be reduced to that of the three-dimensional Pólya random walk described in Section 1. We therefore first note that a

result of Heilbronn (1949) suggests an interesting new proof of the three-dimensional part of Pólya's theorem. The following essentially reproduces Heilbronn's method.

In the case of Pólya's three-dimensional random walk the expression, $H(p_m)$, becomes

$$h(p_m) \equiv \frac{1}{12} \sum_{i=-3}^{3}{}' \sum_{x} \{p_m(x) - p_m(x+u^i)\}^2 ,$$

the summation, $\sum'$, excluding $i = 0$. Then $h(p_m)$ is the probability of non-return to the origin, and we have to show that $h(p_m)$ has a positive lower bound as $m \to \infty$. Let C be a cube of unit volume with edges parallel to the coordinate axes and vertices at lattice points. Define for each point, (ξ_1, ξ_2, ξ_3), of space within or on the boundary of C the function,

$$u_m(\xi_1,\xi_2,\xi_3) \equiv \alpha_{000} + \alpha_{100}\xi_1 + \alpha_{010}\xi_2 + \alpha_{001}\xi_3 + \alpha_{110}\xi_1\xi_2 + \alpha_{011}\xi_2\xi_3 + \alpha_{101}\xi_1\xi_3 + \alpha_{111}\xi_1\xi_2\xi_3 ,$$

where the eight constants are so chosen that $u_m(\xi_1, \xi_2, \xi_3)$ is equal to $p_m(x)$ at each of the eight vertices of C. A function, $u_m(\xi_1, \xi_2, \xi_3)$, may in this way be uniquely defined everywhere.

For each cube C of the lattice

$$\frac{1}{4}\sum_{C}\{\phi_m(x)-\phi_m(x')\}^2 \geq \iiint_C \left\{\left(\frac{\partial u_m}{\partial \xi_1}\right)^2+\left(\frac{\partial u_m}{\partial \xi_2}\right)^2+\left(\frac{\partial u_m}{\partial \xi_3}\right)^2\right\}^2 d\xi_1\, d\xi_2\, d\xi_3 ,$$

where the summation is over all pairs of nearest neighbour vertices of C. It will be sufficient to establish this for the particular cube C which has 0, u^1, u^2, u^3 for four of its vertices: the proof for any other cube is similar. We have then

$$\frac{\partial u_m}{\partial \xi_1} = u_m(1,\xi_2,\xi_3) - u_m(0,\xi_2,\xi_3) .$$

Therefore

$$\iiint_C \left(\frac{\partial u_m}{\partial \xi_1}\right)^2 d\xi_1\, d\xi_2 d\xi_3 = \int_0^1\int \left(\frac{\partial u_m}{\partial \xi_1}\right)^2 d\xi_2\, d\xi_3$$

$$\geq \int_0^1 d\xi_3 \cdot \frac{1}{2}\left[\left\{u_m(11\xi_3)-u_m(01\xi_3)\right\}^2+\left\{u_m(10\xi_3)-u_m(00\xi_3)\right\}^2\right]$$

$$\geq \frac{1}{4}\left[\left\{u_m(111)-u_m(011)\right\}^2+\left\{u_m(110)-u_m(010)\right\}^2 + \cdots + \cdots\right] .$$

These inequalities are consequences of the fact that $\frac{\partial u_m}{\partial \xi_1}$ is linear both in ξ_2 and in ξ_3. We may obtain two similar inequalities in connexion with $\frac{\partial u_m}{\partial \xi_2}$, $\frac{\partial u_m}{\partial \xi_3}$. Combining all three, we obtain the desired result.*

* Heilbronn states (without proof) the weaker inequality without the factor $\frac{1}{4}$.

If we now sum over all cubes of the lattice, we obtain

$$\sum \{ \phi_m(x) - \phi_m(x') \}^2 \geq \iiint \{ (\frac{\partial u_m}{\partial \xi_1})^2 + (\frac{\partial u_m}{\partial \xi_2})^2 + (\frac{\partial u_m}{\partial \xi_3})^2 \} \, d\xi_1 \, d\xi_2 \, d\xi_3$$

when the summation is over all pairs of nearest neighbour lattice points, and the integration is over the whole of space. The factor, $\frac{1}{2}$, has disappeared since each edge of a cube belongs to 4 different cubes. Now we have on the left-hand side,

$$\sum \{ \phi_m(x) - \phi_m(x') \}^2 = \frac{1}{2} \sum_{i=-3}^{3}{}' \sum_x \{ \phi_m(x) - \phi_m(x + u^i) \}^2$$

$$= 6 h(\phi_m) .$$

Let D be the domain, $1 \leq \xi_1^2 + \xi_2^2 + \xi_3^2 \leq 4m^2$. Then we have, a fortiori,

$$6 h(\phi_m) \geq \iiint_D \{ (\frac{\partial u_m}{\partial \xi_1})^2 + (\frac{\partial u_m}{\partial \xi_2})^2 + (\frac{\partial u_m}{\partial \xi_3})^2 \} \, d\xi_1 \, d\xi_2 \, d\xi_3 .$$

Consider a cube, C, which has the origin, O, for one of its vertices. Let λ be the lower bound of $u_m(\xi_1, \xi_2, \xi_3)$ inside and on the surface of C. Since u_m is a trilinear function of the coordinates inside C, λ is attained on one of the faces of C. Since u_m is

a bilinear function on each surface, λ is attained on an edge. Finally, since u_m is linear along each edge, λ is attained at one of the vertices of C. It follows that

$$u_m(\xi_1, \xi_2, \xi_3) \geq \lambda > 0 \quad \text{for } \xi_1^2 + \xi_2^2 + \xi_3^2 = 1.$$

We also have clearly,

$$u_m(\xi_1, \xi_2, \xi_3) \equiv 0 \quad \text{for } \xi_1^2 + \xi_2^2 + \xi_3^2 \geq 4m^2.$$

Now $u_m(\xi_1, \xi_2, \xi_3)$ may be shown to be continuous everywhere, and to have continuous bounded partial first derivations everywhere except on the faces of the cubes. We may therefore apply the Cauchy-Schwarz inequality (Cf. Heilbronn) to obtain

$$\iiint_D \left\{ \left(\frac{\partial u_m}{\partial \xi_1}\right)^2 + \left(\frac{\partial u_m}{\partial \xi_2}\right)^2 + \left(\frac{\partial u_m}{\partial \xi_3}\right)^2 \right\} d\xi_1 d\xi_2 d\xi_3 \geq 4\pi\lambda^2 \cdot \frac{2m}{2m-1}.$$

Therefore

$$h(p_m) \geq \tfrac{2}{3}\pi\lambda^2 \quad \text{for all } m.$$

It follows, a fortiori, that

$$h(g_m) \geq \tfrac{2}{3}\pi\lambda^2$$

where $g_m(x)$ is any lattice-function satisfying the same boundary conditions as $p_m(x)$.

It is interesting to note here that, since $h(p_m) \equiv 1 - p_m(u^1)$, we have for λ the inequality,

$$\tfrac{2}{3}\pi\lambda^2 + \lambda \leq 1 .$$

This gives an upper bound for λ of just less than $\frac{1}{2}$. A numerical value of 0.35 (approx) for the probability of eventual return to the origin has been calculated by McCrea and Whipple (1940).

We turn now to the general symmetrical random walk in $n \geq 3$ dimensions. Let y^1, y^2, y^3 be three linearly independent vectors such that

$$f(y^i) \geq c > 0 , \qquad i = 1, 2, 3 .$$

Then we have

$$H(\phi_m) \geq \tfrac{1}{2} c \sum_{i=1}^{3} \left[\sum_x \left\{ p_m(x) - p_m(x+y^i) \right\}^2 + \sum_x \left\{ p_m(x) - p_m(x-y^i) \right\} \right] .$$

But the three vectors, y^i, positioned at the origin, define the fundamental cube for a three-dimensional sub-lattice within the given n-dimensional lattice. Thus we may re-label the points of this sub-lattice in such a way that the origin is

$$O \equiv (0, 0, 0) ,$$

the point y^i becomes u^i and $-y^i$ becomes u^{-i}. We now regard the sub-lattice as a cubic lattice of unit mesh and any point originally labelled x (n-dimensional) is re-labelled z (3-dimensional). Choose M corresponding to a given m as the smallest integer such that

$$p_m(z) = 0 \quad \text{for all } \|z\| \geq M,$$

and let us write

$$p_m(z) \equiv g_M(z).$$

Then

$$H(p_m) \geq \tfrac{1}{2} c \sum_{i=-3}^{3}{}' \sum_{z} \{ g_M(z) - g_M(z+u^i) \}^2,$$

and $g_M(z)$ is such that

$$g_M(0) = 1, \quad g_M(z) = 0, \quad \text{for } \|z\| \geq M.$$

Therefore

$$H(p_m) \geq 6c \cdot h(g_M) \geq 4\pi c \lambda^2,$$

by the above result for Pólya's random walk, and the proof is completed.

§3. Solutions of the general linear difference equations and inequations.

It will be seen from Section 2 that in the unmodified random walk the probability, p(x), that the

particle, initially at x, eventually reaches the origin satisfies the system of equations

$$\Delta_f p(x) \equiv \sum_y f(y)\{p(x) - p(x+y)\} = 0, \quad x \neq 0.$$

We proceed now to study the solutions of

$$\Delta_f g(x) = 0,$$

and also of the more general "inequations" ¤

$$\Delta_f g(x) \geq 0.$$

Here g(x) is the unknown function and $\{f(x)\}$ is a symmetrical probability distribution, both f(x) and g(x) having as their arguments n-dimensional vectors with integer valued components. (Where necessary, to make the expression $\Delta_f g(x)$ meaningful, we shall assume that the summation over the lattice points is in some given order.) Three types of solution will be

¤ It should be noted that the expression which we denote by $\Delta g(x)$ would more usually be called $-\Delta g(x)$. A solution of the inequations is thus a generalization of a superharmonic function. The change in sign was motivated by the probability meaning of $\Delta_f p(0)$.

distinguished (Cf. Section 3.6):

(i) $g(x) \equiv$ constant. This is the trivial solution: it always exists.

(ii) $g(x)$ bounded for all x but not constant.

(iii) $g(x) \to \infty$ as $\|x\| \to \infty$, $|g(x)| < \infty$ for all x.

We shall call this a <u>properly divergent</u> solution.

It should be noted that we shall not exclude the possibility that $f(y) = 0$ for some y; there may then exist points such that there is zero probability that the particle will ever visit them. In such cases we have the following situation. Let $f^{(r)}(y)$ denote the probability that the particle, initially at the origin, is at the point y at the r^{th} step. Then $f^{(r)}(y)$ is defined recursively by

$$f^{(1)}(y) \equiv f(y)$$

$$f^{(r+1)}(y) \equiv \sum_x f(x) f^{(r)}(x+y), \quad r \geq 1.$$

Now for any point x^i let Ω_i denote the set of points, y, such that

$$f^{(r)}(y - x^i) > 0 \quad \text{for some } r.$$

Then $x^i \in \Omega_i$ and any point $y \in \Omega_i$ generates in this way precisely the set Ω_i. The lattice points may thus be decomposed into a number of mutually disjoint sets, Ω_i, $(i = 1, 2, \ldots)$, such that any point of the lattice belongs to one of them. (When the dimensionality of $\{f(y)\}$ is the same as that of the lattice the number of such sets will be finite.)

It will now be seen that the system of equations (and inequations) may be divided up into a number of independent systems such that all the arguments of $g(x)$ occurring in the i^{th} system belong to Ω_i. These systems are equivalent, and having fixed on an origin of coordinates 0, we may restrict consideration to the one set, Ω, generated by 0.

In the Markov chain terminology (Cf. Section 3.0), when Ω is the set of all lattice points the random walk is irreducible. In other cases it is decomposable. Any pair of points x, y belonging to Ω will be interaccessible:

$$f^{(r)}(y-x) \equiv f^{(r)}(x-y) > 0 \quad \text{for some } r.$$

and in particular the origin will be accessible from any point of Ω. It will be seen that Ω is an infinite Abelian group under the operation of vector addition.

The following three theorems apply to lattices of general dimension n. In each case we consider functions g(x) which satisfy the system of equations, or inequations, for <u>all</u> $x \in \Omega$. In the remaining theorems we shall <u>except</u> the point x = 0.

THEOREM 3.1 <u>The trivial solution is the only bounded solution of</u>

$$\Delta_f g(x) = 0, \qquad x \in \Omega.$$

Proof. Suppose that $|g(x)| \leq M$ for all $x \in \Omega$. Choose any pair of points $a, b \in \Omega$. We show that g(a) = g(b). Let b = a + u. Then for some r,

$$f^{(r)}(u) = \lambda > 0.$$

Define $$h(x) \equiv g(x+u) - g(x).$$

Then h(x) also satisfies

$$\Delta_f h(x) = 0, \qquad x \in \Omega.$$

Suppose if possible that h(x) $\neq$ 0 for some $x \in \Omega$. Then we may assume without loss of generality that there exists a constant m > 0 such that

$$|h(x)| \leq m \leq 2M \quad \text{for all } x \in \Omega,$$

and that for any given $\varepsilon > 0$ there exists an $x \in \Omega$ such that

$$h(x) > m - \varepsilon .$$

Now we have

$$h(x) = \sum_y f^{(r)}(y)\, h(x+y) , \qquad x \in \Omega .$$

Therefore

$$h(x) \leq (1-\lambda)m + \lambda h(x+u) , \qquad x \in \Omega .$$

We have now a recurrence relation from which we obtain

$$h(x) \leq (1-\lambda^n)m + \lambda^n h(x+nu) , \qquad x \in \Omega .$$

Therefore

$$\lambda^{-n}\{h(x) - m\} + m \leq h(x+nu) .$$

But

$$\sum_{n=0}^{t-1} h(x+nu) = g(x+tu) - g(x) \leq 2M .$$

Therefore

$$tm - t\lambda^{-t}\{m - h(x)\} \leq 2M .$$

But this gives a contradiction, since we can choose t such that $tm > 2M + \delta$ and then choose x such that $m - h(x) < \delta t^{-1}\lambda^t$. It follows that m = 0, and thus that g(a) = g(b).

THEOREM 3.2 Any non-constant solution of

$$\Delta_f g(x) = 0, \quad x \in \Omega,$$

which has either an upper or a lower bound cannot attain this bound.

Proof. Suppose g(x) to have an upper bound M, and that for some point a, g(a) = M.

We have

$$\sum_y f^{(r)}(y)\, g(x+y) = g(x) \quad \text{for all } r \text{ and for } x \in \Omega.$$

But for any point $z \in \Omega$, there exists an integer r such that

$$f^{(r)}(z-a) > 0.$$

Now $$\sum_y f^{(r)}(y) = 1.$$

Therefore $g(z) < M$ would imply that for some r

$$\sum_y f^{(r)}(y)\, g(a+y) < g(a).$$

This is a contradiction. Therefore g(x) ≡ constant. In the same way, if g(x) attained a lower bound m at some point a, the assumption that g(x) > m for some x would imply that for some r

$$\sum_y f^{(r)}(y)\, g(a+y) > g(a).$$

THEOREM 3.3 Any non-constant solution of

$$\Delta_f g(x) \geq 0 , \qquad x \in \Omega ,$$

cannot attain a lower bound.

Proof. If g(x) had a lower bound m with g(a) = m, we should find as above that for some r

$$\sum_y f^{(r)}(y)\, g(a+y) > g(a) .$$

COROLLARY. There does not exist a properly divergent solution of $\Delta_f g(x) \geq 0 , \quad x \in \Omega .$

For a properly divergent solution g(x) attains a lower bound.

We prove next two lemmas which connect up the solutions of the difference equations and inequations with the probability results of Section 2.

LEMMA 1. In the random walk on Ω let p(x) denote the probability that the particle, initially at $x \in \Omega$, eventually reaches the origin 0. Then p(x) < 1 for some x if and only if there exists a bounded non-constant solution of

$$\Delta_f g(x) = 0 , \qquad x \in \Omega - \{0\} .$$

Proof. By a change of notation this lemma could be reduced to a special case of Theorem 3.6.4. It is however just as quick to give a direct proof.

The necessity of the condition is obvious, since we have seen that $p(x)$ is a solution of

$$\Delta_f g(x) = 0, \qquad x \in \Omega - \{0\},$$

with $p(0) = 1$.

To prove the sufficiency, we note first that for any constants c, d, with $d \neq 0$, $c + dg(x)$ is also a solution. Therefore it is possible to arrange that

$$0 \leq g(x) \leq 2 \quad \text{for all } x,$$

with $$g(0) = 1.$$

Let $p^{(r)}(x)$ denote the probability that the particle reaches 0 from x on or before the r^{th} step. Then $p^{(r)}(x)$ is defined recursively by

$$p^{(1)}(x) \equiv f(x),$$

$$p^{(r+1)}(x) \equiv \sum_y f(y)\, p^{(r)}(x+y), \qquad x \in \Omega - \{0\},$$

with $$p^{(r)}(0) \equiv 1.$$

We see that $p^{(r)}(x)$ is monotone non-decreasing with r and

$$p^{(r)}(x) \to p(x) \quad \text{as} \quad r \to \infty .$$

But $$g(x) = \sum_y f(y)\, g(x+y) , \quad x \in \Omega - \{0\} .$$

Therefore

$$g(x) \geq f(x) = p^{(1)}(x) .$$

Therefore, by induction,

$$g(x) \geq p^{(r)}(x) \quad \text{for all } r.$$

Therefore

$$p(x) \leq g(x) \quad \text{for all } x .$$

But since g(x) is by hypothesis non-constant with g(0) = 1, we must have for some x either g(x) < 1 or g(x) > 1. In the former case we reach the desired conclusion that for this x, p(x) < 1. In the latter case, by consideration of the new solution 2 - g(x), we reach the same conclusion.

LEMMA 2. <u>In the random walk on Ω the particle will almost certainly eventually return to the origin if and only if there exists a properly divergent solution of</u>

$$\Delta_f g(x) \geq 0 , \quad x \in \Omega - \{0\} .$$

Proof. It will be more convenient this time to show, by

changing the notation, that the sufficiency of the condition follows from Theorem 3.6.6, and the necessity from Theorem 3.6.7.

The enumerable set of points of Ω is the set of states of an irreducible Markov chain with a stochastic matrix which we obtain by numbering the states in the following way. We take in turn the points x with $\|x\| = t$ $(t = 0, 1, 2, \ldots)$, and number those belonging to Ω in some lexicographical order, $i = 0, 1, 2, \ldots$. Thus to each $x \in \Omega$ there corresponds an integer i $(x \leftrightarrow i)$ and

$$0 \leftrightarrow 0, \quad i < \infty \text{ when } \|x\| < \infty, \quad i \to \infty \text{ as } \|x\| \to \infty.$$

The stochastic matrix is now $[p_{ij}]$ (i and j = 0, 1, 2 ...) where

$$p_{ij} \equiv f(y-x), \quad x \leftrightarrow i, \quad y \leftrightarrow j.$$

Define

$$g(x) \equiv a_i, \quad x \leftrightarrow i.$$

Let us suppose that a properly divergent solution exists of the inequations,

$$\Delta_f g(x) \geq 0, \quad x \in \Omega - \{0\}.$$

These now become

$$\sum_{j=0}^{\infty} p_{ij} a_j \leq a_i , \quad i \neq 0 ,$$

and

$$a_i \to \infty \quad \text{as} \quad i \to \infty ,$$

with $|a_i| < \infty$ for all i.

We consider the corresponding simply reducible system in which we make the zero state absorbing. Since now $p_{00} = 1$, we have

$$\sum_{j=0}^{\infty} p_{ij} a_j \leq a_i \quad \text{for all } i .$$

The required result that the zero state is recurrent if a properly divergent solution exists now follows an application of Theorems 3.6.2 and 3.6.6.

To prove the necessity of the given condition, before we can apply Theorem 3.6.7, we have to examine the conditions (3.6.3) and (3.6.4), as applied to the simply reducible system. With one trivial exception, it will be seen that (3.6.3) is always satisfied. The exception is the one-dimensional Pólya random walk. Here the origin is an impassable barrier between the positive and negative integers. Since this system is however thereby separated into two systems, for each of which (3.6.3) is true, there is no real difficulty. We prove

now (3.6.4), with the closed set C consisting of the zero state alone. Choose any δ, $0 < \delta < 1$, and let n_i be the integer such that

$$n_i^2 \leq \|x\| < (n_i+1)^2, \qquad x \Leftrightarrow i .$$

Now let ε_i be the smallest number for which

$$\sum_{\|y\| \leq n_i} f(y) \geq 1 - \varepsilon_i .$$

For $t < n_i$, we have

$$p_{io}^{(t)} \leq \text{Prob}\left[\|y^1\| + \|y^2\| + \cdots + \|y^t\| > t n_i\right] \leq 1-(1-\varepsilon_i)^t,$$

where $\|y^j\|$ denotes the "step-length" of the j^{th} step. Now let r_i be the number of positive integers t for which

$$1 - (1-\varepsilon_i)^t \leq \delta .$$

Define $t_i(\delta) \equiv \min(r_i, n_i)$. Then, when $t_i(\delta) \geq 1$ we have

$$p_{io}^{(t)} \leq \delta \quad \text{for} \quad t \leq t_i(\delta) .$$

Now let $i \to \infty$. Then $\|x\| \to \infty$, and so $n_i \to \infty$. But as $n_i \to \infty$, $\varepsilon_i \to 0$, so that $r_i \to \infty$. Hence $t_i \to \infty$, and condition (3.6.4) is established. The required result, that the zero state is recurrent only if a properly divergent solution of the inequations exists, now follows an application of Theorems 3.6.2

and 3.6.7.

THEOREM 3.4 Let the lattice dimension be n = 1 or 2. When n = 1 suppose that

$$\sum_{y=m}^{\infty} f(y) = o(m^{-1})$$

and when n = 2 suppose that

$$\sum_{\|y\| \geq m} f(y) = o(m^{-2}) .$$

Then (i) there does not exist a bounded non-constant solution of

$$\Delta_f g(x) = 0, \quad x \in \Omega - \{0\},$$

and (ii) there does exist a properly divergent solution of

$$\Delta_f g(x) \geq 0, \quad x \in \Omega - \{0\}.$$

Proof. Let p(x) denote the probability of attaining the origin from $x \in \Omega$, with $\phi(0) = 1$. By Theorems 2.1 and 2.2, we know there is probability one that the particle, initially at the origin, eventually returns to the origin. Therefore

$$\sum_x f(x)\phi(x) = 1 .$$

It follows that also

$$\sum_x f^{(r)}(x)\phi(x) = 1 , \text{ for all } r.$$

But for any $x \in \Omega$, $f^{(r)}(x) > 0$ for some r. Therefore we must have $p(x) \equiv 1$ for all $x \in \Omega$. The non-existence of the bounded non-constant solution now follows from Lemma 1, and the existence of the properly divergent solution from Lemma 2.

THEOREM 3.5 Let the lattice dimension be $n \geq 3$, and suppose $\{f(y)\}$ to be at least a three-dimensional distribution. Then (i) there exists a bounded non-constant solution of

$$\Delta_f g(x) = 0 , \qquad x \in \Omega - \{0\} ,$$

and (ii) there does not exist a properly divergent solution of

$$\Delta_f g(x) \geq 0 , \qquad x \in \Omega - \{0\} .$$

Remark. It follows, a fortiori, that there does not exist a properly divergent solution of

$$\Delta_f g(x) = 0 , \qquad x \in \Omega - \{0\} .$$

Proof. By Theorem 2.3, return to the origin is not certain. It follows that $p(x) < 1$ for some $x \in \Omega$. The existence of the bounded non-constant solution then follows from Lemma 1, and the non-existence of the

properly divergent solution from Lemma 2.

THEOREM 3.6 There always exists a bounded non-constant solution of

$$\Delta_f g(x) \geq 0, \quad x \in \Omega - \{0\}$$

Proof. We have merely to define

$$g(x) \equiv \lambda > 0, \quad x \in \Omega - \{0\}$$

$$g(0) = 0.$$

One question remains open, namely whether there exists for n = 1 or 2, a properly divergent solution of the equations

$$\Delta_f g(x) = 0, \quad x \in \Omega - \{0\}.$$

It will be seen in the next section that this exists at least in the case of the simpler equations which are associated with the Pólya random walk.

§4. Solutions of the discrete Laplace equations.

The results in this section are derived independently of the preceeding theorems. We here study the particular case of $\Delta_f g(x) = 0$ which is the

discrete analogue of Laplace's equation. Define

$$\Delta g(x) \equiv \frac{1}{2n} \sum_{i=-n}^{n}{}' \{ g(x) - g(x+u^i) \},$$

and consider

$$\Delta g(x) = 0, \quad x \neq 0.$$

This is the system of equations associated with Pólya's random walk problem, and a solution is said to be discrete harmonic (Cf. Section 1). Explicit solutions will be obtained.

As before let $f^{(t)}(x)$ denote the probability that the particle, initially at 0, is at point x at the t^{th} step. Then $f^{(t)}(x)$ may be calculated explicitly by a consideration of the number of possible ways of going from 0 to x in t steps to nearest neighbours (not excluding arrivals at x before the t^{th} step. Alternatively,

$$f^{(t)}(x) = (2\pi)^{-n} n^{-t} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} (\cos\theta_1 + \cdots + \cos\theta_n)^t \cos x_1\theta_1 \cdots \cos x_n\theta_n \cdot d\theta_1 \cdots d\theta_n,$$

where n is the lattice dimensions (Cf. Feller, 1950, p.306, Ex.17). Now let n = 3. (The situation is similar when n > 3.) Define

$$g(x) \equiv \sum_{t=0}^{\infty} f^{(t)}(x).$$

We show first that g(x) is bounded. It will be seen from the formula that

$$f^{(t)}(x) \leq f^{(t)}(0) \quad \text{when } \|x\| \text{ is even},$$

$$f^{(t+1)}(x) \leq f^{(t)}(0) \quad \text{when } \|x\| \text{ is odd}.$$

Therefore $g(x) \leq g(o)$ for all x. We have therefore only to show that g(o) is finite. Now

$$\left| \tfrac{1}{3} (\cos\theta_1 + \cos\theta_2 + \cos\theta_3) \right|$$

attains its maximum value of unity only at $\theta_1 = \theta_2 = \theta_3 = 0$ and $\pm\pi$. Therefore the series for g(o) converges or diverges with

$$\sum_{t=0}^{\infty} \iiint_0^{\pi/2} \left\{ \tfrac{1}{3} (\cos\theta_1 + \cos\theta_2 + \cos\theta_3) \right\}^t d\theta_1\, d\theta_2\, d\theta_3$$

But the integrands are now all positive. Therefore this series is equal to

$$\iiint_0^{\pi/2} \frac{d\theta_1\, d\theta_2\, d\theta_3}{1 - \tfrac{1}{3} (\cos\theta_1 + \cos\theta_2 + \cos\theta_3)} ,$$

obtained by summing a geometric series. Near $\theta_1 = \theta_2 = \theta_3 = 0$ this integral behaves like

$$\iiint_0^{\delta} \frac{d\theta_1\, d\theta_2\, d\theta_3}{\theta_1^2 + \theta_2^2 + \theta_3^2} ,$$

which is finite. Therefore g(o) is finite.

Next it is almost obvious that g(x) is non-constant. For we have $f^{(t)}(x) = 0$ for $t \leq T$, where T is the greatest integer less than $\|x\|$. Therefore, for such an x,

$$g(x) \leq \sum_{T-1}^{\infty} f^{(t)}(o) < \sum_{0}^{\infty} f^{(t)}(o).$$

Finally, g(x) is discrete harmonic for all $x \neq 0$. We have

$$f^{(t)}(x) = \frac{1}{2n} \sum_{i=-n}^{n}{}' f^{(t-1)}(x + u^i), \quad t \geq 1.$$

Now $f^{(o)}(x) = 0$ for $x \neq 0$. Therefore

$$2n\, g(x) = \sum_{i=-n}^{n}{}' \sum_{t=1}^{\infty} f^{(t-1)}(x + u^i)$$

$$= \sum_{i=-n}^{n}{}' \sum_{t=0}^{\infty} f^{(t)}(x + u^i)$$

$$= \sum_{i=-n}^{n}{}' g(x + u^i),$$

as required.

We turn now to the one- and two-dimensional cases. When n = 1 we see at once that

$$g(x) \equiv |x|$$

is a properly divergent solution of $\Delta g(x) = 0$, $x \neq 0$. Now let n = 2. An argument similar to that given above shows that the series

$$\sum_{t=0}^{\infty} f^{(t)}(x)$$

is now divergent. We therefore define instead

$$g(x) \equiv \sum_{t=0}^{\infty} \{ f^{(t)}(0) - f^{(t)}(x) \} .$$

We prove first that g(x) is finite for each x. Again using an argument similar to one above, we find that this series convereges or diverges with

$$\iint_{0}^{\delta} \frac{1 - \cos x_1\theta_1 \cos x_2\theta_2}{1 - \frac{1}{2}(\cos\theta_1 + \cos\theta_2)} \, d\theta_1 \, d\theta_2 ,$$

δ being a small positive number. But this integral will be seen to be finite.

We next prove that g(x) is properly divergent. We note again that $f^{(t)}(x) = 0$ for $t \leq T$, where T is the greatest integer less than $\|x\|$. Therefore

$$g(x) = \sum_{t=0}^{T} f^{(t)}(0) + \sum_{t=T+1}^{\infty} \{ f^{(t)}(0) - f^{(t)}(x) \} .$$

But the second term tends to zero as $T \to \infty$, and the series

$$\sum_{t=0}^{\infty} f^{(t)}(0)$$

is divergent. Therefore $g(x) \to \infty$ as $\|x\| \to \infty$.

Finally, $g(x)$ is discrete harmonic for all $x \neq 0$. We have

$$g(x) = f^{(0)}(0) + \sum_{t=1}^{\infty} \{ f^{(t)}(0) - f^{(t)}(x) \}$$

$$= f^{(0)}(0) + \sum_{t=1}^{\infty} \{ f^{(t)}(0) - \tfrac{1}{4} \sum_{i=-2}^{2}{}' f^{(t-1)}(x+u^i) \}$$

$$= \tfrac{1}{4} \sum_{i=-2}^{2}{}' \left[f^{(0)}(0) + \sum_{t=1}^{\infty} \{ f^{(t)}(0) - f^{(t-1)}(x+u^i) \} \right]$$

$$= \tfrac{1}{4} \sum_{i=-2}^{2}{}' \sum_{t=0}^{\infty} \{ f^{(t)}(0) - f^{(t)}(x+u^i) \}$$

$$= \tfrac{1}{4} \sum_{i=-2}^{2}{}' g(x+u^i) .$$

We have now incidentally established the

THEOREM 4.1 For $n = 1$ or 2, there exists a properly divergent solution of

$$\Delta g(x) = 0, \quad x \neq 0.$$

It may be noted that for $n = 1$ or 2, the solution of the electrical network problem of Section 1 is $1 + Cg(x)$; for $n = 3$, it is $1 + C\{g(0) - g(x)\}$. These solutions have been given in another form by McCrea and Whipple (1940), and (for $n = 2$) by Van der Pol and Bremmer (1950, p.371).

BIBLIOGRAPHY

BARTLETT, M.S.

1949. Some evolutionary stochastic processes. J. Roy. Statist. Soc. (B) 11, 211-29.

1950. Recurrence times. Nature 165, 727-8.

BRAUER, A.

1946. Limits for the characteristic roots of a matrix. Duke Math. J. 13, 387-95.

1947. ibid. 14, 21-6.

1948. ibid. 15, 871-7.

1952. ibid. 19, 75-91.

COLLATZ, L.

1942. Einschliessungsatz für die characteristischen Zahlen von Matrizen. Math. Z. 48, 221-6.

ERDÖS, P., FELLER, W. and POLLARD, H.

1949. A property of power series with positive coefficients. Bull. Amer. Math. Soc. 55, 201-4.

FELLER, W.

1949. Fluctuation theory of recurrent events. Trans. Amer. Math. Soc. 67, 98-119.

1950. An Introduction to Probability Theory and its Applications. I. Wiley, New York.

FERRAR, W.L.

1951. Finite Matrices. Oxford.

FORTET, R.

1935. Sur des probabilités en chaîne. C. R. Acad. Sci. Paris 201, 184-6.

FOSTER, F.G.

1951a. Markoff chains with an enumerable infinity of states and a class of cascade processes. Proc. Cambridge Phil. Soc. 47, 77-85.

1951b. J. Roy. Statist. Soc. (B) 13, 174-8.

1952a. On Markov chains with an enumerable infinity of states. Proc. Cambridge Phil. Soc. 48,

1952b. A Markov chain derivation of discrete distributions. Ann. Math. Stat. 23,

FRÉCHET, M.

1938. Recherches théoriques modernes sur le calcul des probabilités, vol. 2 (théorie des événements en chaine). Paris.

1949. Statistical Self-renewing Aggregates. Cairo.

FROBENIUS, G.

1908. Über Matrizen aus positiven Elementen. S.-B. Preuss. Akad. Wiss., 471-6.

1909. ibid., 514-8.

1912. Über Matrizen aus nicht negativen Elementen. ibid., 456-77.

GALTON, F.

1889. Natural Inheritance. London. (Appendix F is devoted to the problem of the extinction of surnames.)

GOOD, I.J.

1949. The number of individuals in a cascade process. Proc. Cambridge Phil. Soc. 45, 360-3.

HALMOS, P.R.

1950. Measure Theory. Van Nostrand, New York.

HARRIS, T.E.

1952. First passage and recurrence distributions. Trans. Amer. Math. Soc. 70,

HEILBRONN, H.A.

1949. On discrete harmonic functions. Proc. Cambridge Phil. Soc. 45, 194-206.

HOSTINSKÝ, M.B.

1931. Méthodes générales du calcul des probabilités. Mémor. Sci. Math. 52.

KENDALL, D.G.

1949. Stochastic processes and population growth. J.Roy. Statist. Soc. (B) 11, 230-64.

, 1950. An artificial realization of a simple birth-and-death process. ibid. 12, 116-9.

1951a. On non-dissipative Markoff chains with an enumerable infinity of states. Proc. Cambridge Phil. Soc. 47, 633-4.

1951b. Some problems in the theory of queues. J. Roy. Statist. Soc. (B) 13, 151-73, 184-5.

KNOPP, K.

1928. Theory and Application of Infinite Series. Blackie, London.

KOLMOGOROV, A.N.

1933. Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin.

1936a. Zur Theorie der Markoffschen Ketten. Math. Ann. 112, 155-60.

1936b. Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. Rec. Math. N.S. 1, 607-10.

1937. Bull. Univ. État. Moscou (A) 1, 1-15.

LEDERMANN, W.

1950. Bounds for the greatest latent roots of a positive matrix. J. London Math. Soc. 25, 265-8.

LÉVY, P.

1951. Systèmes markoviens et stationnaires; cas dénombrable. Ann. Sci. Écol. Norm. Sup. 68, 327-81.

McCREA, W.H.

1936. A problem on random paths. Math. Gaz. 20, 311-7.

McCREA, W.H. and WHIPPLE, F.J.W.

1940. Random paths in two and three dimensions. Proc. Roy. Soc. Edinburgh 60, 281-98.

OSTROWSKI, A.

1952. Bounds for the greatest latent root of a positive matrix. J. London Math. Soc. 27, 253-6.

PERFECT, H.

1951. On matrices with positive elements. Quart. J. Math. Oxford (2) 2, 286-90.

PERRON, O.

1907a. Jacobischer Kettenbruchalgorithmus. Math. Ann. 64, 1-76.

1907b. ibid., 248-63.

PÓLYA, G.

1921. Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz. Math. Ann. 84, 149-60.

STEFFENSEN, J.F.

1930. Om Sandsynligheden for at Afkommet uddør. Mat. Tidsskr. B, 19-33.

1933. Deux problèmes du calcul des probabilités. Ann. Inst. H. Poincaré 3, 319-44.

TAUSSKY, O.

1948. Bounds for characteristic roots of matrices. Duke Math. J. 15, 1043-4.

1951. Bounds for characteristic roots of matrices II. J. Research Nat. Bur. Standards 46, 124-5.

VAN DER POL, B. and BREMMER, H.

1950. Operational Calculus. Cambridge.

WATSON, H.W. and GALTON, F.

1874. On the probability of the extinction of families. J. Anthrop. Inst. 4, 138-44.

WIELANDT, H.

1950. Unzerlegbare, nicht negative Matrizen. Math. Z. 52, 642-8.

YOSIDA, K. and KAKUTANI, S.

1939. Markoff process with an enumerable infinite number of possible states. Jap. J. Math. 16, 47-55.