TRANSIENT BEHAVIOUR IN FINITE ABSORBING MARKOV CHAINS

BY

E.SENETA B.Sc.(Hons.)
MATHEMATICS DEPARTMENT
UNIVERSITY OF ADELAIDE

Submitted for the Degree of M.Sc.
December, 1964

CONTENTS

Summary

Several possible distributions for describing quasi-stationary behaviour within the transient states of a finite Markov absorbing chain are proposed and studied. The study is motivated by some work of Bartlett, and possible population model implications.

The submatrix corresponding to the transient states is found to be of fundamental importance, the most relevent results obtained being expressible in terms of the eigen-vectors corresponding to its spectral radius, when this matrix is indecomposable. Whereas most attention is devoted to the primitive and cyclic alternatives of this case, the theory is extended to the decomposable set of states by some results of Mandl, and developed analogously for the finite case with continuous parameter.

During the evolution of the subject matter, it becomes clear that, in effect, a generalization of ordinary Markov chains is being considered, this giving rise to the idea of quasi-chains, defined on the transient states. The most interesting result in relation to this, is the existence and properties of a reverse Markov chain, defined also only on the transient states.

The last part of the thesis is devoted to examining the concept of quasi-stationary distributions for absorbing chains with a denumerably infinite state space, in a few special cases. In the process a theorem of Yaglom, concerning simple discrete branching processes, is generalized. Finally, an appendix gives some notes on generating functions in our context, together with some remarks on the related problems for diffusion processes.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and to the best of the author's knowledge and belief the thesis contains no material previously published or written by another person except when due reference is made in the text of the thesis.

Signed..

(E. Seneta)

INTRODUCTION

It is both theoretically and practically interesting to study the behaviour within the transient states of a a finite absorbing Markov chain. Theoretically, such a study is profitable because the problem has received virtually no attention in the literature, apart from a few isolated instances(Bartlett, [2] ; Ewens, [IO] ; Yaglom, [33].)where the matter arises only incidentally. From a practical point of view, such a study is significant in relation to genetic and stochastic population models, for such models are often just finite absorbing Markov chains (e.g. Ewens, [II] ; Watterson, [32].). In such chains the probability of absorption in a finite time is unity, but "It may still happen that the time to extinction is so long that it is still of more relevence to consider the effectively ultimate distribution, called a 'quasi-stationary' distribution,..."(Bartlett, [2]). It is with such quasi-stationary distributions and their variants that we concern ourselves in this thesis.

Finite chains, with discrete time parameter are given the most extensive treatment, and the results are developed analogously for the case of continuous time The methods used are not applicable when the number of states is denumerably infinite. However Chapter IO, in which a few special cases are considered, is included for completeness.

The general method used is to set up quantities which are thought to provide suitable descriptions of the behaviour, in some sense, as time goes on. Thus we often consider limits as the time parameter increases beyond all bounds. After investigating existence and aptness, we compare such quantities. Of considerable interest also, is the comparison of the various expressions with corresponding expressions occuring in the ordinary theory of

Markov chains, leading to a somewhat parallel evolution to it. Where relevent, results of other authors (e.g. Mandl, [22]) are appropriate, they are adapted to our needs,with due citation.

The material of Chapters I, 2, and 8 is not new apart from the lemmas in I.3. The content of Chapters 3, 4, 5, 6, and 9 is essentially new, except where otherwise stated. For the most part, Chapter 7 contains new matter, apart from 7.2, 7.3, 7.4 , while Chapter IO contains nothing original apart from the generalization of Yaglom's theorem.

The most significant result of the investigation is that it is in fact possible to develop a reasonable, unified theory of quasi-stationary behaviour. Moreover, it is shown that such a theoryis, in fact, more general than the ordinary theory of finite Markov chains, in that certain results are obtained for sub-stochastic matrices, which contain those for stochastic matrices as particular cases. Hence it may be justifiable to speak of finite Markov chains as a particular case of more general Markov 'quasi-chains'. However as regards these statements, this thesis can be regarded as only a beginning.

I. BASIC DEFINITIONS AND RESULTS.

I.I Non-negative Matrices.

The theory of non-negative matrices has been applied to finite Markov chains in several books: see Frechet [I4], Gantmacher [I5], and Romanovskii [29]. Good descriptions of the theory, in English, may be found in Debreu and Herstein [8], and Gantmacher; we merely state what we require.

Definition I.I We will call a matrix $\mathbf{A}$ with real elements non-negative(positive) if all its elements are non-negative(positive), and write it $A \geq 0$ $(A > 0)$

Definition I.2 The square matrix $A = [a_{ik}]_1^n$ is called decomposable if thre is a permutation of indices (i.e. simultaneous interchange of rows and columns) which reduces it to the form

$$\tilde{A} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

where A_{11}, A_{22}, are square. Otherwise the matrix $\mathbf{A}$ is indecomposable.

Definition I.3 The non-negative indecomposable matrix $\mathbf{A}$ is called primitive if there is a power of $\mathbf{A}$ which is positive i.e. $A^p > 0$ for some $p \geq 1$. Otherwise it is imprimitive(or cyclic).

THEOREM I.I An indecomposable non-negative matrix $\mathbf{A}$

a) always has a positive eigenvalue r, which is a simple root of the characteristic equation;

b) has the moduli of the other eigenvalues at most r;

c) has corresponding to r positive left and right eigenvectors; and

d) if $\mathbf{A}$ has precisely h eigenvalues $r, \lambda_2, \ldots \lambda_h$ of modulus equal to r, then these numbers are all different from each other and are roots of the equation

$$\lambda^h - r^h = 0$$

THEOREM I.2 An indecomposable non-negative matrix **A** is primitive iff **r** exceeds the modulus of any other eigenvalue. Hence if $h>1$, such **A** is cyclic.

THEOREM I.3 If **A** is cyclic, it may be reduced to the following cyclic form by a permutation of indices:

$$A = \begin{bmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ 0 & 0 & A_{23} & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & A_{h-1\,h} \\ A_{h1} & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where the zero matrices along the diagonal are square. Thus **h** is is called the period of the cyclic matrix.

THEOREM I.4 (Romanovskii, [29] pp. I5-I6) For an indecomposable non-negative matrix $A = [a_{ik}]_1^n$

$$\min_i \sum_{k=1}^{n} a_{ik} \le r \le \max_i \sum_{k=1}^{n} a_{ik}$$

with one equality holding iff both hold.

Definition I.4 We shall refer to **r** as the spectral radius of the non-negative indecomposable matrix **A**.

Definition I.5 A substochastic matrix is a non-negative square matrix whose row sums do not exceed unity. (A particular case is the stochastic matrix, whose row sums are all unity.)

I.2 Ergodic Chains and Absorbing Chains.

We follow the terminology of Kemeny and Snell [I9], but the few remarks below serve to redefine the various types of chains in terms of non-negative stochastic matrices. This is convenient for comparison purposes, later. However, for the properties and general theory of finite Markov chains, the reader is referred to [19], the approach there being a matrix one also, although the authors are careful to avoid spectral theory, and follow the more common classification of chains in terms of communication relations.

a) A regular chain is one with a primitive transition matrix.

b) A <u>cyclic</u> chain is one with a cyclic transition matrix.

c) An <u>ergodic</u> chain is one which is either regular or cyclic.

We recall that an <u>absorbing</u> chain is one whose transition matrix can be written in the form

$$(1.1) \qquad P = \begin{bmatrix} 1 & \underset{\sim}{0}' \\ \underset{\sim}{p_0} & Q \end{bmatrix} \qquad \underset{\sim}{p_0} \neq \underset{\sim}{0}$$

where Q is sxs and $\underset{\sim}{p_0}$, $\underset{\sim}{0}$ are both sxl; where the totality of statesfor this, as any other finite chain in this thesis, shall be denoted by $S, = \{0,1,2,....s\}$. The set of transient states, $\{1,2,.....s\}$ in the absorbing chain, shall be denoted by T (in general we shall consider only one <u>absorbing</u> state, 0;). Thus Q is the matrix of transitions within the transient states, and has at least one row sum less than unity(since $\underset{\sim}{p_0} \neq \underset{\sim}{0}$).

It is found convenient to define certain random variables associated with an absorbing chain, which simplify what is to follow. If the process starts in state $i \epsilon T$ at time 0, let $Y_{ij}^{(n)}=1$ or 0 according as the process is or is not in state j at time n. Therefore $Y_{ij}^{(0)} = \delta_{ij}$. Further, let

$$X_{ij} = \sum_{n=0}^{\infty} Y_{ij}^{(n)} \quad , \quad X_i = \sum_{j \epsilon T} X_{ij}; \; i, j \epsilon T.$$

It follows that

a) X_{ij} is the total number of visits to state j before absorption (or alternately, the time spent in state j before absorption) having started from i;

b) X_i is the time to absorption, starting from i.

Moreover, when the process starts with an initial probability distribution whose component over the transient set T is $\underset{\sim}{\pi}$. $(0 < \underset{\sim}{\pi}' \underset{\sim}{e} \leq 1)$, it is convenient to define random variables $X_{\underset{\sim}{\pi} j}$, $X_{\underset{\sim}{\pi}}$ to have analogous meanings. Thus

$$X_{\underset{\sim}{f}_i,j} = X_{ij} \quad , \qquad X_{\underset{\sim}{f}_i} = X_i \ .$$

where – we denote by $\underset{\sim}{f}_i$ the vector with one as the i th coordinate the others being zero, and by $\underset{\sim}{e}$ the unit vector, so that $\underset{\sim}{e}' = [1,1,\cdots\cdots 1]$.

Finally, we note that from theorems I.I, I.2 and I.4 we have for an absorbing chain, with matrix Q primitive, that

$$(1.2) \qquad Q^m = \rho^m \underset{\sim}{w}\, \underset{\sim}{v}' + O(n^k |\rho_2|^m) .$$

Here $\rho < 1$ is the spectral radius of Q, $\underset{\sim}{w}$ and $\underset{\sim}{v}'$ are the corresponding positive left and right eigenvectors such that $\underset{\sim}{v}'\underset{\sim}{w} = 1$, and $|\rho_2| < \rho$ where ρ_2 is the eigenvalue whose modulus is closest to ρ. The number k+1 is the maximum multiplicity taken over those eigenvalues whose modulus is $|\rho_2|$; hence it is w.l.o.g. the multiplicity of ρ_2. The vectors $\underset{\sim}{w}$ and $\underset{\sim}{v}'$ are also assumed to be normalized in such a way that $\underset{\sim}{v}'\underset{\sim}{e} = 1$, where $\underset{\sim}{e}$ is the unit vector. It is important to note that (1.2) is true for any primitive matrix Q, apart from the bound on ρ.

I.3 Two Basic Lemmas.

Since it is appropriate to use generating functions several times in the sections which deal with the discrete time parameter, we utilize a fairly well known approach to counting problems in such theory (e.g. Bhat, [5]; Good, [16]) to prove the first of the lemmas. The second gives an expression for the derivative of a matrix power which enters into our generating functions.

LEMMA I.I Let Q be the substochastic matrix which corresponds to the set of transient states T in an absorbing Markov chain. Let $D_j(\omega)$ be the diagonal matrix of the same dimension whose j th diagonal element is ω, the others being ones. Then, putting

$$\underset{\sim}{\pi}' D_j(\omega) = \underset{\sim}{\pi}_j(\omega)$$

$$Q\, D_j(\omega) = Q_j(\omega)$$

we have

(1.3) $\sum_{x=0}^{n} x\, P[X_{\underset{\sim}{\pi},j} = x,\ X_{\underset{\sim}{\pi}} = n] = \frac{d}{d\omega}[\underset{\sim}{\pi}_j'(\omega)\, Q_j^{n-1}(\omega)[I-Q]\underset{\sim}{e}]_{\omega=1}$

Proof: $E[\omega^{X_{\underset{\sim}{\pi},j}}] = \sum_{i\in T} \pi_i\, \omega^{\delta_{ij}} \sum_{i_1, i_2, \ldots, i_{n-1}} p_{i i_1} \omega^{\delta_{i_1 j}} p_{i_1 i_2} \omega^{\delta_{i_2 j}} \ldots p_{i_{n-2} i_{n-1}} \omega^{\delta_{i_{n-1} j}} \underset{\sim}{p}_{i_{n-1} 0}$

where $X_{\underset{\sim}{\pi}} = n$ is fixed.

$$= \underset{\sim}{\pi}' D_j(\omega)[Q D_j(\omega)]^{n-1}[I-Q]\underset{\sim}{e}$$

$$= \underset{\sim}{\pi}_j(\omega)\, Q_j^{n-1}(\omega)[I-Q]\underset{\sim}{e}$$

Hence (1.3), by differentiation and evaluation at .

LEMMA I.2 When Q is primitive

(1.4) $\frac{d}{d\omega}[Q_j^{n-1}(\omega)]_{\omega=1} = (n-1)\rho^{n-2} v_j w_j \underset{\sim}{w}\, \underset{\sim}{v}' + O(\rho^n)$

where v_j, w_j are the j th elements of the eigenvectors $\underset{\sim}{v}'$ and $\underset{\sim}{w}$ respectively (see (1.2)).

Proof:

$$\frac{d}{d\omega}[Q_j^{n-1}(\omega)]_{\omega=1} = \sum_{m=0}^{n-2} Q^m \frac{d}{d\omega} Q_j(\omega)\, Q^{n-2-m}$$

$$= \sum_{m=0}^{n-2} Q^m \underset{\sim}{q}_j \underset{\sim}{\delta}_j'\, Q^{n-2-m}$$

where $\underset{\sim}{q}_j$ is the j th column of Q. From (1.1)

$$= \sum_{m=0}^{n-2}\left\{\rho^{n-2}\underset{\sim}{w}\,\underset{\sim}{v}'\underset{\sim}{q}_j\underset{\sim}{\delta}_j'\underset{\sim}{w}\,\underset{\sim}{v}'\right\} + O\left\{\rho^{n-2}\sum_{m=0}^{n-2} m^k\left(\frac{|\rho_2|}{\rho}\right)^m\right\}$$

$$= (n-1)\rho^{n-2} w_j v_j \underset{\sim}{w}\,\underset{\sim}{v}' + O(\rho^n)$$

since $\underset{\sim}{v}'\underset{\sim}{q}_j = \rho v_j$ ($\underset{\sim}{v}'P = \rho\underset{\sim}{v}'$) and $\sum_{m=0}^{n-2} m^k \left(\frac{|\rho_2|}{\rho}\right)^m$ is bounded, being convergent. Hence (1.4).

2. MOTIVATION AND ANALYSIS.

2.I Bartlett's Approach.

While dealing with stochastic population models of various kinds, Bartlett ([2], p.21) introduces the idea of a quasi-stationary distribution during the discussion of one which is a Markov chain, with transition matrix (adapted to our notation and context)

$$(2.1) \qquad P = \left[\begin{array}{c|cc} 1-\varepsilon & \varepsilon & \underset{\sim}{0}' \\ \hline \underset{\sim}{p}_0 & & Q \end{array}\right]$$

Here the dimensions of the various components are as in (1.1) except that 0 is now (s-2)x1. The states 0,1,2,..,s actually represent numbers of individuals, and the transition probabilities are given by

$$\begin{aligned} p_{i,i+1} &> 0 \quad , \quad 1 \le i \le s-1 \\ p_{i,i-1} &> 0 \quad , \quad 1 \le i \le s \\ p_{i,i} &> 0 \quad , \quad 1 \le i \le s. \end{aligned}$$

Thus when $\varepsilon > 0$, it is clear that the chain is regular, whereas if $\varepsilon = 0$, the state 0 is absorbing (an "extinction state) but the matrix Q is primitive. Both these situations can, of course, be described in terms of random walks or a general birth-and-death process.

Now, when $\varepsilon = 0$, extinction occurs in a finite time with probability 1. To obtain an effectively ultimate distribution, when the time to absorption may be long, he considers a chain which is in an obvious sense "close" to the absorbing chain, by taking $\varepsilon > 0$, but small. Since the chain is now regular, there exists a limiting distribution independent of any initial distribution, for which Bartlett derives an expression, this last depending on ε.

At this stage an assumption is made to the effect that, for certain states, the limiting probability is considerably greater than for the zero state, this leading to an approximation in which all the limits of the non-zero states are independent of $\varepsilon > 0$, which cancels in the approximation. This last appears to say that

$$\mu_m(\varepsilon) \sim \frac{\mu_m(\varepsilon)}{1 - \mu_0(\varepsilon)}, \quad m = 1, 2, \ldots, s \tag{2.2}$$

where $\mu_n(\varepsilon)$ is the limiting probability of being in state n, and is intuitively obvious under Bartlett's assumption. It will be shown in 2.3 that the right hand side is independent of $\varepsilon > 0$; Bartlett reasons that because of this the right hand side provides a description for the corresponding absorbing chain, i.e. when $\varepsilon = 0$. This viewpoint will also be discussed in 2.3.

2.2 Ewens' "Pseudo-transient" Distribution.

Ewens [10] is concerned (in a genetic context) with the formal solution of the stationary equation of a diffusion process, when both boundaries are "exit" (i.e. absorbing) inwhich case no non-trivial stationary distribution can exist. With suitable interpretation, this pseudo-stationary distribution is shown to be the proportion of time spent in any specified range before absorption. We recognize this as being the diffusion analogue of the type of distribution for which we search.

The parallel discussion for the discrete case is also carried out in [10] and is again concerned with matrix (2.1), where $\varepsilon = 0$. It is clear that the proportion of time spent in any state j before absorption is

$$b_j(\underset{\sim}{\pi}) = \frac{\sum_{i \in T} \pi_i E[X_{ij}]}{\sum_{i \in T} \pi_i E[X_i]} = \frac{E[X_{\underset{\sim}{\pi}, j}]}{E[X_{\underset{\sim}{\pi}}]}, \quad j \in T \tag{2.3}$$

Ewens calls this distribution a pseudo-distribution since it obeys the formal requirements for a probability distribution (i.e. always positive, and of total mass unity), although it is not the distribution of any variate in the absorbing process. He gives it the name 'pseudo-transient' because he thinks that "...it describes in a sense the behaviour of the process before absorption" (Ewens, [10]). The sequence of Ewens' papers on this topic is [10], [11], and [12].

2.3 Discussion.

We first generalize and then discuss Bartlett's approach to the problem. Let us take the chain with matrix (2.4) where $1 > \varepsilon > 0$ and Q merely indecomposable:

$$P(\varepsilon, \underset{\sim}{\alpha}) = \begin{bmatrix} 1-\varepsilon & \varepsilon\underset{\sim}{\alpha}' \\ \underset{\sim}{p}_0 & Q \end{bmatrix} \qquad \underset{\sim}{p}_0 \neq \underset{\sim}{0} \tag{2.4}$$

where $\underset{\sim}{\alpha}'\underset{\sim}{e} = 1$. Such a matrix represents a regular Markov chain in which there is a probability of escape out of state 0 equal to ε. Whereas Bartlett took $\underset{\sim}{\alpha}' = \underset{\sim}{\delta}_1' = [1, 0, \ldots 0]$, here $\underset{\sim}{\alpha}$ is a general probability vector whose elements give the conditional probabilities of escape into states 1, 2, ..., s.

Hence there exists a unique stationary (and limiting) distribution, denoted by

$$[\mu_0(\varepsilon, \underset{\sim}{\alpha}), \underset{\sim}{\mu}'(\varepsilon, \underset{\sim}{\alpha})],$$

whence it follows that

$$[\mu_0(\varepsilon, \underset{\sim}{\alpha}), \underset{\sim}{\mu}'(\varepsilon, \underset{\sim}{\alpha})]\begin{bmatrix} 1-\varepsilon & \varepsilon\underset{\sim}{\alpha}' \\ \underset{\sim}{p}_0 & Q \end{bmatrix} = [\mu_0(\varepsilon, \underset{\sim}{\alpha}), \underset{\sim}{\mu}'(\varepsilon, \underset{\sim}{\alpha})]$$

and therefore the stationary distribution conditional on the process being in T, is

$$\underset{\sim}{\alpha}'(\underset{\sim}{\alpha}) = \frac{\underset{\sim}{\mu}'(\varepsilon, \underset{\sim}{\alpha})}{1 - \mu_0(\varepsilon, \underset{\sim}{\alpha})} = \frac{\underset{\sim}{\alpha}'[I-Q]^{-1}}{\underset{\sim}{\alpha}'[I-Q]^{-1}\underset{\sim}{e}} \tag{2.5}$$

and therefore independent of $\varepsilon > 0$. This is a generalization of (2.2). (Note that $[I-Q]$ is non-singular ([19], p.22))

It is, however, quite clear that the absorbing chain with matrix $P(\varepsilon, \underset{\sim}{\alpha})$ where $\varepsilon = 0$, and the related chain where $\varepsilon > 0$ are only superficially related. Intuitively, there is a vast difference between a finite regular and an absorbing chain, no matter how small is i.e. there is a discontinuity at $\varepsilon = 0$ as regards the physical system being described. More concretely, we refer to a paper of Sinkhorn [31] who is concerned with a certain iterative procedure for positive stochastic matrices, and discusses, by means of examples, the effect of replacing the zero entries of non-negative matrices by "small" functions $p_{ij}(\varepsilon) > 0$. He comments that "Even the apparently natural artifice of replacing zero entries by 'small' functions ....and (subsequently) letting $\varepsilon \to 0$ leads to difficulties." and also "It may....be a poor policy to use a strictly positive approximation...., unless there is a very good reason for a particular selection.".

Another considerable objection to the use of $\underset{\sim}{a}(\underset{\sim}{\alpha})$ in describing quasi-stationary behaviour, is that it is a far too general function of $\underset{\sim}{\alpha}$. It can be made into almost any probability distribution $\underset{\sim}{z}$ over T by a suitable choice of $\underset{\sim}{\alpha}$, for

$$\underset{\sim}{a}'(\underset{\sim}{\alpha}) = \underset{\sim}{z} \qquad \text{iff} \qquad \underset{\sim}{\alpha}' = \frac{\underset{\sim}{z}'[I-Q]}{\underset{\sim}{z}'[I-Q]\underset{\sim}{e}}$$

providing the denominator is not zero. It is also interesting that the distributions which are excluded by the denominator being zero, are just those forms which if regarded as an initial distribution, allow no immediate absorption. This last, however, seems to be just that kind of property which would be pleasing in a quasi-stationary distribution. Bartlett's quasi-stationary distribution was obtained by taking $\underset{\sim}{\alpha}' = [1,0,\ldots 0]$ because his state 1 corresponds to a population of size 1.

From now on, for the above reasons, we shall concern ourselves only with absorbing chains, although we shall have occasion to refer to the distribution $\mathbf{a}(\alpha)$ obtained from the related regular chain, again. The generalization of Ewens' suggestion is given in 3.1, where we first take up the extension of possible modes of description of behaviour within the transient states, basing our study on the ideas put forward in this chapter.

3. TWO QUASI-STATIONARY DISTRIBUTIONS: DISCRETE CASE

3.I The Ratio of Means.

The distribution (2.3) seems a reasonable one, even when the substochastic matrix **Q** corresponding to T is arbitrary, so long as it is not stochastic, of course. The matrix of means ([19], p.46) can be written

$$M = [E[X_{ij}]]$$
$$= [I - Q]^{-1}$$

for any such **Q**, and hence (2.3) becomes

$$(3.1) \qquad \underset{\sim}{b}'(\underset{\sim}{\pi}) = \frac{\underset{\sim}{\pi}'[I-Q]^{-1}}{\underset{\sim}{\pi}'[I-Q]^{-1}\underset{\sim}{e}}$$

Possibly the most interesting characteristic of this ratio of means distribution is that

$$\underset{\sim}{b}'(\underset{\sim}{\pi}) = \underset{\sim}{a}'(\underset{\sim}{\pi})$$

providing an interesting link between the two. The reason for this is as follows.

In the chain governed by $P(\varepsilon, \underset{\sim}{\pi})$, the limiting proportion of time spent in j is $\mu_j(\varepsilon, \underset{\sim}{\pi})$, the limiting distribution. Now, we can think of each passage to the state zero in this chain as a complete realization of the correponding absorbing chain with initial probability $\underset{\sim}{\alpha} = \underset{\sim}{\pi}$. Thus the limiting proportion of time spent in state $j \neq 0$ in the regular chain gives the ratio of the expected time spent in $j \in T$ to the expected absorption time in the absorbing one. A more general application of this "return" process is described by Kemeny and Snell [19], p.117. It was also used by Ewens, and is the discrete analogue of the elementary return process of diffusion theory.

The great disadvantage of this distribution is that no simple expression for it can be obtained, even when $\boldsymbol{Q}$ is primitive. Secondly, it is not independent of $\underset{\sim}{\pi}$, whereas it is reasonable to hope that a 'good' quasi-stationary distribution would be. Both these facts are more readily seen by realizing that

$$E[X_{ij}] = \sum_{n=0}^{\infty} p_{ij}^{(n)} \qquad , \ i, j \in T$$

and considering, for instance, the spectral resolution (1.2) for primitive $\boldsymbol{Q}$. We refer also to the example in 3.3.

3.2 <u>The Mean Ratio</u>.

In view of 3.1, we are led to consider something similar, namely the expected fraction of the time spent in T, as a possible candidate for the same purpose. Thus we define, for $\sum_{i \in T} \pi_i = 1$,

$$(3.2) \qquad c_j(\underset{\sim}{\pi}) = E\left[\frac{X_{\underset{\sim}{\pi}j}}{X_{\underset{\sim}{\pi}}}\right] \qquad , \ j \in T$$

as the mean ratio distribution. Now

$$E\left[\frac{X_{\underset{\sim}{\pi},j}}{X_{\underset{\sim}{\pi}}}\right] = \underset{X_\pi}{E}\left[\frac{1}{X_{\underset{\sim}{\pi}}} \underset{X_{\underset{\sim}{\pi},j}/X_{\underset{\sim}{\pi}}}{E}[X_{\underset{\sim}{\pi},j} / X_{\underset{\sim}{\pi}}]\right]$$
$$= \sum_{n=1}^{\infty} n^{-1} \sum_{x=0}^{\infty} P[X_{\underset{\sim}{\pi},j} = x, X_{\underset{\sim}{\pi}} = n]$$

Using LEMMA I.I, (1.3)

$$(3.3) \qquad c_j(\underset{\sim}{\pi}) = \frac{d}{d\omega}\left[\underset{\sim}{\pi}_j'(\omega) \sum_{n=1}^{\infty} n^{-1} Q_j^{n-1}(\omega)[I-Q]\underset{\sim}{e}\right]_{\omega=1}$$

Once again, this expression for $c_j(\underset{\sim}{\pi})$ does not reduce to an easily manageable form, nor is it independent of $\underset{\sim}{\pi}$ (see next section).

This unfortunate dependence upon $\underset{\sim}{\pi}$ in both $\underset{\sim}{b}(\underset{\sim}{\pi})$ and $\underset{\sim}{c}(\underset{\sim}{\pi})$ can be explained by recalling that any realization of the absorbing chain remains in T only a finite time, and this is not long enough for the dependence on the initial distribution to'wear off'.

3.3 An Example.

The example we consider is the substochastic analogue of the independent trials process. Let

$$Q = \underset{\sim}{\alpha}\underset{\sim}{\beta}' \quad ,$$

where $\alpha_i, \beta_i > 0$, i=1,2,...s; $\underset{\sim}{\beta}'\underset{\sim}{e} = 1$, $\underset{\sim}{\alpha}'\underset{\sim}{\beta} = \gamma < 1$. Thus Q is primitive, with spectral radius and all other eigenvalues zero. The corresponding eigenvectors are Then:-

a) Ratio of Means.

$$b_j(\underset{\sim}{\pi}) = \frac{\sum_{i \in T} \pi_i \sum_{m=0}^{\infty} p_{ij}^{(m)}}{\sum_{i \in T} \pi_i \sum_{m=0}^{\infty} (1 - p_{i0}^{(m)})}$$

$$= \frac{(1-\gamma)\pi_j + \beta_j \sum_{i \in T} \pi_i \alpha_i}{(1-\gamma) + \sum_{i \in T} \pi_i \alpha_i} \, ,$$

which even in this very simple case depends on $\underset{\sim}{\pi}$;

b) Mean Ratio.(from (3.3))

$$c_j(\underset{\sim}{\pi}) = \frac{d}{d\omega}[\underset{\sim}{\pi}_j'(\omega)[I-Q]\underset{\sim}{e}]_{\omega=1} + \sum_{m=2}^{\infty} m^{-1} \frac{d}{d\omega}[\underset{\sim}{\pi}_j'(\omega) Q_j^{m-1}(\omega)[I-Q]\underset{\sim}{e}]_{\omega=1}$$

$$= \pi_j \underset{\sim}{\delta}_j'[I-Q]\underset{\sim}{e} + \sum_{m=2}^{\infty} m^{-1} \pi_j \underset{\sim}{\delta}_j' Q^{m-1}[I-Q]\underset{\sim}{e} + \sum_{m=2}^{\infty} m^{-1} \underset{\sim}{\pi}' \frac{d}{d\omega}[Q_j^{m-1}(\omega)[I-Q]\underset{\sim}{e}]_{\omega=1} .$$

Applying LEMMA I.2, (I.4) to the last summand

$$c_j(\underset{\sim}{\pi}) = \pi_j \underset{\sim}{\delta}_j'[I-\underset{\sim}{\alpha}\underset{\sim}{\beta}']\underset{\sim}{e} + \pi_j \underset{\sim}{\delta}_j' \underset{\sim}{\alpha}\underset{\sim}{\beta}'[I-\underset{\sim}{\alpha}\underset{\sim}{\beta}']\underset{\sim}{e} \sum_{m=2}^{\infty} \frac{\gamma^{m-2}}{m} + \sum_{m=2}^{\infty} \frac{(m-1)}{m} \gamma^{m-2} \underset{\sim}{\pi}'\underset{\sim}{\alpha}\underset{\sim}{\beta}'[I-\underset{\sim}{\alpha}\underset{\sim}{\beta}']\underset{\sim}{e}\, u_j v_j ,$$

$$= \pi_j[1-\alpha_j] + \gamma^{-1}\alpha_j\beta_j \underset{\sim}{\pi}'\underset{\sim}{\alpha} - \gamma^{-2}[1-\gamma]\alpha_j[\log(1-\gamma)+\gamma][\pi_j - \gamma^{-1}\beta_j \underset{\sim}{\pi}'\underset{\sim}{\alpha}]$$

which is also dependent on $\underset{\sim}{\pi}$.

At this stage it is useful to introduce some further notation. Let $X_{ij}^{*} = X_{ij} - \delta_{ij}$, $X_i^{*} = X_i - 1$. It follows that these random variables disregard the initial occupation (of state i). Then the modified quantity

$$(3.4) \qquad \frac{\sum_{i\in T} \pi_i E[X_{ij}^{*}]}{\sum_{i\in T} \pi_i E[X_i^{*}]} = \frac{\sum_{i\in T} \pi_i \sum_{m=1}^{\infty} p_{ij}^{(m)}}{\sum_{i\in T} \pi_i \sum_{m=1}^{\infty} (1 - p_{i0}^{(m)})}$$

$$= \beta_j \quad (= \nu_j),$$

which, in this special case, is independent of $\underset{\sim}{\pi}$. We cannot however consider a modified mean ratio, since $X_{\underset{\sim}{\pi}}^{*}$ can assume the value zero, even if $\sum_{i\in T} \pi_i = 1$.

4. THE LIMITING CONDITIONAL MEAN RATIO: DISCRETE CASE

On account of the remarks at the conclusion of 3.2, it is plausible to consider those quasi-stationary distributions which can be derived, roughly speaking, by considering only those realizations of an absorbing chain in which the time to absorption is long. In a regular chain, the stationary probability of any state is the limiting proportion of the time spent in that state. One analogue of this for an absorbing chain is $c_j(\underset{\sim}{\pi})$. Another, but this time conforming to the above criterion, is

$$E\left[\frac{X_{\underset{\sim}{\pi},j}}{X_{\underset{\sim}{\pi}}} \Big/ X_{\underset{\sim}{\pi}} = n\right],$$

where n is large; it is studied in this chapter only for the case when Q is primitive, as are also its variants.

4.1 Derivation.

For primitive Q corresponding to T in an absorbing Markov chain

$$\lim_{n\to\infty} E\left[\frac{X_{\underset{\sim}{\pi},j}}{X_{\underset{\sim}{\pi}}} \Big/ X_{\underset{\sim}{\pi}}=n\right] = w_j v_j \quad. \tag{4.1}$$

We have in fact that

$$E\left[\frac{X_{\underset{\sim}{\pi},j}}{X_{\underset{\sim}{\pi}}} \Big/ X_{\underset{\sim}{\pi}}=n\right] = \frac{1}{n} \; \frac{\sum_{x=0}^{n} x \, P[X_{\underset{\sim}{\pi},j}=x, X_{\underset{\sim}{\pi}}=n]}{P[X_{\underset{\sim}{\pi}}=n]} .$$

It is well known that

$$P[X_{\underset{\sim}{\pi}}=n] = \underset{\sim}{\pi}' Q^{n-1}[I-Q]\underset{\sim}{e}$$

from e.g. Bartlett ([1], p.68); and the numerator is given by (1.3). Hence it follows that the above is

$$\frac{\frac{d}{d\omega}[\underset{\sim}{\pi}_j'(\omega)\, Q_j^{n-1}(\omega)\,[I-Q]\underset{\sim}{e}]_{\omega=1}}{n\,\underset{\sim}{\pi}'\, Q^{n-1}\,[I-Q]\,\underset{\sim}{e}} .$$

Utilizing (1.2) and (1.4)

$$= \frac{\pi_j \underset{\sim}{\delta}_j' Q^{n-1}[I-Q]\underset{\sim}{e} + \underset{\sim}{\pi}'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e}.\{(n-1)\rho^{n-1} v_j \omega_j\} + O(\rho^n)}{n\rho^{n-1}\underset{\sim}{\pi}'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e} + O(n(n-1)^k|\rho_2|^{n-1})}$$

$$= \frac{\pi_j \rho^{n-1}\underset{\sim}{\delta}_j'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e} + (n-1)\rho^{n-1}\omega_j v_j \underset{\sim}{\pi}'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e} + O(\rho^n)}{n\rho^{n-1}\underset{\sim}{\pi}'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e} + O(n(n-1)^k|\rho_2|^{n-1})}$$

and since $\underset{\sim}{\pi}'\underset{\sim}{\omega}\underset{\sim}{v}'[I-Q]\underset{\sim}{e}$ is positive for primitive Q, we finally obtain

$$\text{(4.2)} \qquad E\left[\frac{X_{\underset{\sim}{\pi},j}}{X_{\underset{\sim}{\pi}}} \Big/ X_{\underset{\sim}{\pi}} = n\right] = \omega_j v_j + O(\tfrac{1}{n})$$

whence (4.1).

Therefore there is no dependence on $\underset{\sim}{\pi}$ in the limit. This expression also has an interesting resemblance to certain expressions for a regular chain; this will be taken up later.

4.2 A Related Quantity.

Examining (4.2), it may seem physically more relevent to study a modified distribution, which has the physical interpretation

"limiting proportion of time n spent in state j T, given that the time to absorption exceeds n"

for j=1,2,...s; viz.

$$\lim_{n\to\infty} E\left[\frac{X_{ij}^{(n)}}{n} \Big/ X_i > n\right]$$

if the initial state is i, where

$$X_{ij}^{(n)} = \sum_{m=0}^{n} Y_{ij}^{(m)}$$

is the number of visits to state j upto time n. The limit may be found by a method similar to that used to derive (4.2); however there is an alternative which we apply. This, conversely, could have been used to derive (4.2). We have

$$E\left[\frac{X_{ij}^{(n)}}{n} \Big/ X_i > n\right] = \frac{1}{n} E\left[\sum_{m=0}^{n} Y_{ij}^{(m)} \Big/ X_i > n\right]$$

$$= \frac{1}{n} \sum_{m=0}^{n} E\left[Y_{ij}^{(m)} \Big/ X_i > n\right]$$

$$= \frac{1}{n}\left\{\delta_{ij} + \sum_{m=1}^{n} \frac{p_{ij}^{(m)}(1 - p_{jo}^{(n-m)})}{1 - p_{io}^{(n)}}\right\}$$

$$= \frac{1}{n}\left\{\delta_{ij} + \sum_{m=1}^{n} \frac{p_{ij}^{(m)} \sum_{s \in T} p_{js}^{(n-m)}}{\sum_{s \in T} p_{is}^{(n)}}\right\}$$

$$= \frac{1}{n}\left\{\delta_{ij} + \frac{\left[\sum_{s \in T} n\rho^n \omega_i v_j \omega_j v_s\right] + O(\rho^n)}{\left[\sum_{s \in T} \rho^n \omega_i v_s\right] + O(n^k |\rho_2|^n)}\right\}$$

from (1.2); i.e.

$$(4.3) \qquad E\left[\frac{X_{ij}^{(n)}}{n} \Big/ X_i > n\right] = \omega_j v_j + O\left(\frac{1}{n}\right),$$

which gives the same limit as before. A similar procedure gives the same result for arbitrary distribution $\underset{\sim}{\pi}$ over T.

There are considerable similarities in the two methods, although results (4.2) and (4.3) are not intuitively obvious as equivalent. The 'reason' for the similarity, is the close unity between either and the spectral decomposition (see also Appendix). Whereas it would be easy to proceed in 4.1 by realizing that

$$E\left[Y_{ij}^{(m)} / X_i = n\right] = \begin{cases} \dfrac{p_{ij}^{(m)} \sum_{k \in T} p_{jk}^{(n-m-1)} p_{ko}}{\sum_{k \in T} p_{ik}^{(n-1)} p_{ko}}, & 0 < m < n \\ \delta_{ij}, & m = 0 \\ 0, & m \geq n, \end{cases}$$

the p.g.f. approach to derive (4.3), on the contrary, is cumbersome.

4.3 Comparison with the Stochastic Case.

It is clear that the discussion in 4.2 is valid for the corresponding stochastic case i.e. for the regular Markov chain (noting that $P[X > n] = 1$), where now $\rho = 1$ $\underset{\sim}{w} = \underset{\sim}{e}$, and is the unique stationary distribution. Thus we have been discussing a generalization of this case. It is of interest to note that the quasi-stationary probability $w_j v_j$ depends on both the left and right eigenvectors corresponding to the spectral radius, although for the above particular case, it simplifies to v_j. An interesting problem is therefore posed: is there a quasi-stationary probability that depends only upon the left eigenvector ? In fact there is, and hence again an interesting comparison with the regular chain (see Chapter 5).

4.4 Extensions.

There are several extensions of 4.1 and 4.2 in the sense that the same result is obtained in the limit, of expressions closely related to those of these sections. The most important of these is one which gives the quantity $w_j v_j$ an interpretation as a limiting proba bility, extending the analogy with the regular chain; it is suggested by the derivation of (4.3). Consider for $n > m$

(4.4) $\quad P[$ in state j at time m/not absorbed at time n $]$,

$$= P[\, i_m = j \;/\; i_n \neq 0 \,]$$

$$= \frac{\sum_{i \in T} \pi_i \, p_{ij}^{(m)} \left(1 - p_{j0}^{(n-m)}\right)}{\sum_{i \in T} \pi_i \left(1 - p_{i0}^{(n)}\right)}$$

$$= \frac{\sum_{k \in T} \sum_{i \in T} \pi_i \, p_{ij}^{(m)} \, p_{jk}^{(n-m)}}{\sum_{k \in T} \sum_{i \in T} \pi_i \, p_{ik}^{(n)}}$$

$$= w_j v_j + O_m\left(m^k \left(\frac{|\rho_2|}{\rho}\right)^m\right) + O_n\left((n-m)^k \left(\frac{|\rho_2|}{\rho}\right)^{n-m}\right)$$

from (1.2).

This converges to $w_j v_j$ as $n\to\infty$. The result is once more valid for a regular chain.

Of particular importance is the case when n=m in (4.4), when the final expression derived is not valid; this forms a special topic and is deferred to the next chapter.

5. THE STATIONARY CONDITIONAL DISTRIBUTION: DISCRETE CASE

The stationary conditional distribution, which we are about to develop, is again derived by effectively considering only those realizations of the absorbing chain in which the time to absorption is long. Moreover it has a limiting interpretation, and is much like the stationary-limiting distribution for a regular chain.

5.I Definition and Existence.

Let us consider the probability distribution over all (s+1) states at time n, when Q is indecomposable:

$$[\ \pi_0(n)\ ,\ \underset{\sim}{\pi}'(n)\ ]'$$

and denote by $\underset{\sim}{d}(n)$ the conditional distribution restricted to the transient states i.e.

$$\underset{\sim}{d}(n) = \frac{\underset{\sim}{\pi}(n)}{1 - \pi_0(n)} .$$

Seeking a quasi-stationary distribution, we shall call d a stationary conditional distribution over T if

(5.1) $$\underset{\sim}{d}(n+1) = \underset{\sim}{d}(n) = \underset{\sim}{d}$$

Since

$$[\ \pi_0(n),\ \underset{\sim}{\pi}'(n)\ ]\ \underline{P} = [\ \pi_0(n+1),\ \underset{\sim}{\pi}'(n+1)\ ],$$

it follows that

$$\underset{\sim}{\pi}'(n)\ Q = \underset{\sim}{\pi}'(n+1) .$$

Hence

$$\underset{\sim}{d}'(n)\ Q = \rho(n)\ \underset{\sim}{d}'(n+1)$$

where $\rho(n)$ is a function of n. If $\underset{\sim}{d}$ exists,

$$\underset{\sim}{d}'Q = \rho(n)\underset{\sim}{d}' \quad .$$

Since Q is indecomposable and $\underset{\sim}{d} \geq 0$, it follows from THEOREM I.I and the orthogonality of left and right eigen-vectors belonging to different eigenvalues, that

$$\rho(n) = \rho \quad , \quad \underset{\sim}{d} = \underset{\sim}{v}$$

and indeed $\underset{\sim}{d} = \underset{\sim}{v}$ satisfies the required condition (5.1).

5.2 Interpretation as a Limit.

To interpret as a limiting distribution, when Q is indecomposable we treat the two distinct cases, when Q is primitive and when it is cyclic, seperately. The latter is treated in Chapter 6, which is devoted entirely to it.

If the process starts in state $i \in T$ with probability π_i , the probability it has not been absorbed by time n is

$$(5.2) \qquad \sum_{i \in T} \pi_i (1 - p_{i0}^{(n)}) = \underset{\sim}{\pi}' Q^n \underset{\sim}{e} ,$$

and the probability of being in state $j \in T$ at time n, given that the process is still in T is

$$(5.3) \qquad \frac{\sum_{i \in T} \pi_i p_{ij}^{(n)}}{\sum_{i \in T} \pi_i (1 - p_{i0}^{(n)})} \equiv \frac{\underset{\sim}{\pi}' Q^n \underset{\sim}{\delta}_j}{\underset{\sim}{\pi}' Q^n \underset{\sim}{e}} .$$

Notice that this is just (4.4) when n=m. From (1.2)

$$(5.4) \qquad \frac{\sum_{i \in T} \pi_i p_{ij}^{(n)}}{\sum_{i \in T} \pi_i (1 - p_{i0}^{(n)})} = v_j + O\left(n^k \left(\frac{|\rho_2|}{\rho}\right)^n\right) .$$

Therefore the limit as n in (5.3) is , which is independent of (c.f. Chapter 4). Arather cumbersome proof of (5.4) is given in Theorem 1 of Mandl [22], although the context is somewhat different.

Worth mentioning also is an interesting phenomenon noticed by Mandl [22], which involves the stationary and limiting nature of $\underset{\sim}{v}$. The result is of interest particularly in the absorbing chain context i.e. $\rho<1$, although as usual, it is also true for the regular chain in the same sense as before. We have that

$$P[X_{\underset{\sim}{\pi}} > n+m \,/\, X_{\underset{\sim}{\pi}} > n] = \frac{\underset{\sim}{\pi}' Q^{n+m} \underset{\sim}{e}}{\underset{\sim}{\pi}' Q^{n} \underset{\sim}{e}}$$

$$\begin{cases} = \rho^{m} & \text{if } \underset{\sim}{\pi} = \underset{\sim}{v} \\ \to \rho^{m} & \forall \pi,\ 0 < \underset{\sim}{\pi}'\underset{\sim}{e} \leq 1. \end{cases}$$

Thus

$$P[X_{\underset{\sim}{\pi}} = n+m \,/\, X_{\underset{\sim}{\pi}} > n] \overset{=}{\to} \rho^{m-1}(1-\rho),$$

which is the frequently occuring geometric distribution, with parameter ρ. (This shows that the assertion of Kemeny and Snell [19] that eigenvalues have no direct probability interpretation in Markov chain theory is not entirely true.)

5.3 Rates of Convergence.

From a physical point of view, a question of the utmost importance concerning limiting quasi-stationary distributions, is that of the rate at which they are approached. In particular, we may say that such a distributionis of most relevence, when the rate of convergence to it exceeds the rate of convergence to zero of the probability of being within the transient set at time n. This statement is equivalent to saying it is of most relevence, when the time to absorption in most realizations is 'long' (c.f. Bartlett, [2] pp. 22-25). Therefore we briefly turn to this, at least for the case when the matrix **Q** is primitive, the case of fundamental importance.

The probability of still being in T at time n is

$$(5.5) \qquad \sum_{i \in T} \pi_i (1 - p_{i0}^{(n)}) = O(\rho^n)$$

from (1.2), where , the spectral radius of **Q** is less than unity. Now, recalling (5.4),

$$\frac{\sum_{i \in T} \pi_i p_{ij}^{(n)}}{\sum_{i \in T} \pi_i (1 - p_{i0}^{(n)})} = v_j + O\left(n^k \left(\frac{|\rho_2|}{\rho}\right)^n\right).$$

It follows that if $|\rho_2|$ is small compared to ρ, the quantity v_j is most satisfactory, in the sense above. THEOREM I.4 gives that

$$(5.6) \qquad \max_{i \in T} (1 - p_{i0}) \geq \rho \geq \min_{i \in T} (1 - p_{i0}),$$

with either equality true iff both are; thus, roughly speaking, we require that $\min_{i \in T} (1 - p_{i0})$ be close to unity. The same remarks apply to the probability $w_j v_j$, from (4.4) and the subsequent calculations.

Bartlett [3], comparing the limiting conditional distribution v with his conditional limiting distribution, (2.2), comments that the former "...might be defined even if this distribution had small 'absolute' probability content,..." whereas, mentioning (2.2), "...the whole idea of my quasi-stationary distribution was that it had large 'absolute' content, at least over some time interval." He effectively points out, however, that if $|\rho_2|$ is small compared to ρ, the distribution $\underset{\sim}{v}$ is indeed relevent. Moreover, as has been mentioned previously, it seems more satisfactory than (2.2) for several reasons.. However, there is an interesting 'relationship' between the two which will be discussed in the next section.

5.4 Comparison with a Preceding Distribution.

We compare $\underset{\sim}{v}$, which has been seen to be both a limiting conditional and conditional stationary distribution, for primitive Q, with $\underset{\sim}{a}(\underset{\sim}{\alpha})$ (2.3; (2.5)) which was seen to be a conditional distribution, in a certain sense. Using the same notation as before, for $\varepsilon > 0$

$$a_j(\underset{\sim}{\alpha}) = \frac{\mu_j(\varepsilon,\underset{\sim}{\alpha})}{1-\mu_0(\varepsilon,\underset{\sim}{\alpha})} = \frac{\lim\limits_{n\to\infty} p_{ij}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}{\lim\limits_{n\to\infty}(1-p_{i0}^{(n)}(\varepsilon,\underset{\sim}{\alpha}))}$$

$$= \lim_{n\to\infty} \frac{p_{ij}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}{1-p_{i0}^{(n)}(\varepsilon,\underset{\sim}{\alpha})},$$

and since $a_j(\underset{\sim}{\alpha})$ is independent of $\varepsilon > 0$,

$$= \lim_{\varepsilon\to 0+}\lim_{n\to\infty} \frac{p_{ij}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}{1-p_{i0}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}.$$

On the other hand,

$$v_j = \lim_{n\to\infty}\lim_{\varepsilon\to 0+} \frac{p_{ij}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}{1-p_{i0}^{(n)}(\varepsilon,\underset{\sim}{\alpha})}.$$

Thus the two differ because the limit operations do not commute.

Another difference emerges from the comparison of $\underset{\sim}{v}$ with $\underset{\sim}{a}(\underset{\sim}{\pi})$ $(= \underset{\sim}{b}(\underset{\sim}{\pi}))$, for

$$a_j(\underset{\sim}{\pi}) = b_j(\underset{\sim}{\pi}) = \frac{\sum_{i\in T}\pi_i E[X_{ij}]}{\sum_{i\in T}\pi_i E[X_i]}$$

$$= \lim_{m\to\infty}\frac{\sum_{i\in T}\pi_i\sum_{n=0}^{m} p_{ij}^{(n)}}{\sum_{i\in T}\pi_i\sum_{n=0}^{m}(1-p_{i0}^{(n)})}$$

$$= \lim_{m\to\infty}\frac{\sum_{n=0}^{m}\sum_{i\in T}\pi_i p_{ij}^{(n)}}{\sum_{n=0}^{m}\sum_{i\in T}\pi_i(1-p_{i0}^{(n)})},$$

which bears a certain resemblence to

$$v_j = \lim_{n\to\infty}\frac{\sum_{i\in T}\pi_i p_{ij}^{(n)}}{\sum_{i\in T}\pi_i(1-p_{i0}^{(n)})}.$$

Before proceeding to the cyclic case, it is, once more, worth noting that all the preceding discussion is valid when Q is a stochastic matrix i.e. the transition matrix of a regular chain, since we have nowhere used the fact that $\rho < 1$. In this special case the probability of non-absorption after n steps:

$$\sum_{i \in T} \pi_i \left(1 - p_{i0}^{(n)}\right),$$

is just unity, for all n.

6. THE CYCLIC CASE

The purpose of this chapter is to consider the various limiting distributions when the matrix $\mathbf{Q}$ corresponding to the transient set T, is cyclic. As with regular and cyclic chains, the present case can be discussed by utilizing the material for primitive $\mathbf{Q}$ from previous chapters.

6.I Preliminaries.

For simplicity, the period of $\mathbf{Q}$ is taken as 3, in which case the matrix can be written in the form

$$Q = \begin{bmatrix} 0 & Q_{12} & 0 \\ 0 & 0 & Q_{23} \\ Q_{31} & 0 & 0 \end{bmatrix},$$

where the diagonal matrices are square. Let us denote the corresponding cyclic subsets of states T by T_1, T_2, and T_3. Then

$$Q^2 = \begin{bmatrix} 0 & 0 & Q_{12}Q_{23} \\ Q_{23}Q_{31} & 0 & 0 \\ 0 & Q_{31}Q_{12} & 0 \end{bmatrix}, \qquad Q^3 = \begin{bmatrix} Q_{12}Q_{23}Q_{31} & 0 & 0 \\ 0 & Q_{23}Q_{31}Q_{12} & 0 \\ 0 & 0 & Q_{31}Q_{12}Q_{23} \end{bmatrix}$$

where the square matrices $[Q_{i\,i+1} \times Q_{i+1\,i+2} \times Q_{i+2\,i+3}]$, i = 1,2,3 (subscripts to be reduced mod.3) are primitive, each with the same spectral radius, ρ^3, and corresponding to the subsets T_i, i = 1,2,3 respectively. Writing left and right eigenvectors corresponding to $Q_{i\,i+1}\,Q_{i+1\,i+2}\,Q_{i+2\,i+3}$ as $\underset{\sim}{v}^{(i)}$ and $\underset{\sim}{w}^{(i)}$ respectively (these being normalized as usual, so that $\sum_{j \in T_i} w_j^{(i)} v_j^{(i)} = 1$, $\sum_{j \in T_i} v_j^{(i)} = 1$ for every i), the following relations must hold:

$$\underset{\sim}{v}^{(1)\prime} Q_{12}\, Q_{23}\, Q_{31} = \rho^3\, \underset{\sim}{v}^{(1)\prime} \tag{6.1a}$$

$$\underset{\sim}{v}^{(2)\prime} Q_{23}\, Q_{31}\, Q_{12} = \rho^3\, \underset{\sim}{v}^{(2)\prime} \tag{6.1b}$$

$$\underset{\sim}{v}^{(3)\prime} Q_{31}\, Q_{12}\, Q_{23} = \rho^3\, \underset{\sim}{v}^{(3)\prime} \tag{6.1c}$$

Multiplying (6.1a) by Q_{12} and (6.1b) by $Q_{12} \cdot Q_{23}$, and invoking the uniqueness of $\underset{\sim}{v}^{(i)\prime}$, we have

(6.2a) $$\underset{\sim}{v}^{(2)\prime} = \frac{\underset{\sim}{v}^{(1)\prime} Q_{12}}{\underset{\sim}{v}^{(1)\prime} Q_{12} \underset{\sim}{e}}$$

(6.2b) $$\underset{\sim}{v}^{(3)\prime} = \frac{\underset{\sim}{v}^{(1)\prime} Q_{12} Q_{23}}{\underset{\sim}{v}^{(1)\prime} Q_{12} Q_{23} \underset{\sim}{e}}$$

We are now in a position to proceed.

6.2 The Limiting Conditional Mean Ratio.

We shall restrict ourselves to considering

$$E\left[\frac{X_{ij}^{(n)}}{n} \Big/ X_i > n\right] = \frac{1}{n} \frac{\sum_{s=0}^{n} p_{ij}^{(s)}(1 - p_{jo}^{(n-s)})}{1 - p_{io}^{(n)}}$$

which is cumbersome enough, as always happens in the cyclic case. For convenience, it is considered in several stages.

a) i, j $\in T_i$, i = 1,2,3.

n = 3k , k a positive integer

$$E\left[\frac{X_{ij}^{(n)}}{n} \Big/ X_i > n\right] = \frac{1}{n}\left\{\frac{\sum_{s=1}^{n} p_{ij}^{(s)}(1 - p_{jo}^{(n-s)})}{1 - p_{io}^{(n)}} + \delta_{ij}\right\}$$

$$= \frac{1}{3k}\left\{\frac{\sum_{s=1}^{3k} p_{ij}^{(s)}(1 - p_{jo}^{(3k-s)})}{1 - p_{io}^{(3k)}} + \delta_{ij}\right\}$$

$$= \frac{1}{3k}\left\{\frac{\sum_{h=1}^{k} p_{ij}^{(3h)}(1 - p_{jo}^{3(k-h)})}{1 - p_{io}^{(3k)}} + \delta_{ij}\right\}$$

since $p_{ij}^{(s)} = 0$ if s ≠ 3h, h a positive integer. Since the elements $p_{ij}^{(3h)}$, i, j $\in$ T are the elements of the h th power of the primitive matrix $Q_{i\,i+1} Q_{i+1\,i+2} Q_{i+2\,i+3}$, it follows from the theory of 4.1 that this expression becomes

$$\frac{1}{3} w_j^{(i)} v_j^{(i)} + O\left(\frac{1}{n}\right).$$

b) i $\in T_i$, j $\in T_{i+1}$, i = 1,2,3.

n = 3k , k a positive integer.

$$E\left[\frac{X_{ij}^{(m)}}{m} / X_i > m\right] = \frac{1}{3k}\left\{\frac{\sum_{s=1}^{3k} p_{ij}^{(s)}(1-p_{jo}^{(3k-s)})}{1-p_{io}^{(3k)}} + \delta_{ij}\right\}$$

and since $p_{ij}^{(s)} = 0$ if $s \neq 3h - 2$, h a positive integer, $p_{jl}^{3(k-h)+2} = 0$ if $l \neq T$, the above becomes

$$= \frac{1}{3k}\left\{\frac{\sum_{h=1}^{k} p_{ij}^{(3h-2)} \sum_{l \in T_i} p_{jl}^{3(k-h)+2}}{\sum_{l \in T_i} p_{il}^{(3k)}} + \delta_{ij}\right\}$$

$$= \frac{1}{3k}\left\{\frac{\sum_{h=1}^{k} \sum_{f \in T_i} p_{if}^{3(h-1)} p_{fj} \sum_{l \in T_i} \sum_{d \in T_{i+1}} p_{jd}^{3(k-h)} p_{dl}^{(2)}}{\sum_{l \in T_i} p_{il}^{(3k)}} + \delta_{ij}\right\}.$$

The calculation now becomes unpleasant, and so is omitted. We mention only that with the aid of (1.2) and (6.1a)-'.. (6.2b) the last expression can be written as

$$\frac{1}{3} w_j^{(i+1)} v_j^{(i+1)} + O\left(\frac{1}{m}\right),$$

as can be expected.

c) $i \in T_i$, $j \in T_{i+2}$, $i = 1,2,3$.
 $n = 3k$, k a positive integer.

We state the result without derivation:

$$E\left[\frac{X_{ij}^{(m)}}{m} / X_i > m\right] = w_j^{(i+2)} v_j^{(i+2)} + O\left(\frac{1}{m}\right)$$

d) $i \in T_i$, $j \in T_i$, $i = 1,2,3$.
 $n = 3k + 1$, k a positive integer.

Using obvious relations,

$$\frac{1}{(3k+1)}\left\{\frac{\sum_{s=1}^{3k+1} p_{ij}^{(s)}(1-p_{jo}^{(3k+1-s)})}{1-p_{io}^{(3k+1)}} + \delta_{ij}\right\}$$

$$= \frac{1}{(3k+1)}\left\{\frac{\sum_{h=1}^{k} p_{ij}^{(3h)} \sum_{l \in T_{i+1}} p_{jl}^{(3(k-h)+1)}}{\sum_{l \in T_{i+1}} p_{il}^{(3k+1)}} + \delta_{ij}\right\}$$

$$= \frac{1}{(3k+1)}\left\{\frac{\sum_{h=1}^{k} p_{ij}^{(3h)} \sum_{l \in T_{i+1}} \sum_{d \in T_i} p_{jd}^{(3(k-h))} p_{dl}}{\sum_{l \in T_{i+1}} \sum_{d \in T_i} p_{id}^{(3k)} p_{dl}} + \delta_{ij}\right\}$$

$$= \frac{1}{3} w_j^{(i)} v_j^{(i)} + O\left(\frac{1}{m}\right),$$

by (1.2). All further calculations are omitted. The results are summarized by the statement

$$(6.3)\qquad \lim_{n\to\infty} \frac{1}{n} E[X_{ij}^{(n)} / X_i > n]$$

$$= \frac{1}{3} w_j^{(i)} v_j^{(i)} \qquad \text{if } i, j \in T_i$$

$$= \frac{1}{3} w_j^{(i+1)} v_j^{(i+1)} \qquad \text{if } i \in T_i,\ j \in T_{i+1}$$

$$= \frac{1}{3} w_j^{(i+2)} v_j^{(i+2)} \qquad \text{if } i \in T_i,\ j \in T_{i+2}.$$

For any given $i \in T$, this can be expressed in matrix form viz.

$$(6.4)\ \lim_{n\to\infty}\left[\frac{1}{n} E[X_{ij}^{(n)}/X_i > n]\right] = \frac{1}{3}\underset{\sim}{e}\,[w_1^{(1)}v_1^{(1)} \ldots w_{t_1}^{(1)}v_{t_1}^{(1)}, w_1^{(2)}v_1^{(2)} \ldots w_{t_2}^{(2)}v_{t_2}^{(2)} w_1^{(3)}v_1^{(3)} \ldots w_{t_3}^{(3)}v_{t_3}^{(3)}]$$

where t_i is the number of states in T_i, $t_1 + t_2 + t_3 = s$. The reader will notice the similarity of the above to the case of the cyclic chain matrix, of which it is clearly a generalization, since we have nowhere used that $\rho < 1$. This will be considered more fully in 6.4, in the general comparison.

Note particularly that the result is independent of the initial state i; it would have been inconvenient to have considered an initial distribution as was done previously, though it is not difficult to see that in this more general case the result is independent of .

6.3 The Limiting Conditional Distribution.

For simplicity we define and investigate the quantity

$$(6.5)\qquad s_{ij}^{(n)} = \frac{p_{ij}^{(n)}}{\sum_{j\in T} p_{ij}^{(n)}} = \frac{\underset{\sim}{\delta}_i' Q^n \underset{\sim}{\delta}_j}{\underset{\sim}{\delta}_i' Q^n \underset{\sim}{e}} \qquad i, j \in T$$

Since the maximal modulus eigenvalues of Q are given by

$$\rho_m = \rho\, e^{\frac{2\pi i m}{3}} \qquad m = 0, 1, 2.$$

from THEOREM I.I, it is clear that

$$\lim_{m\to\infty} s_{ij}^{(m)}$$

does not exist. However, from 6.I

$$\lim_{m\to\infty} s_{ij}^{(3m)} \geq 0 ,$$

being positive iff i and j belong to the same cyclic subset T . In fact then

$$(6.6a) \qquad \lim_{m\to\infty} s_{ij}^{(3m)} = v_j^{(i)} , \qquad i,j \in T_i .$$

To get a complete description of the behaviour, we must investigate the other non-trivial cases, viz.

a) $\lim\limits_{m\to\infty} s_{ij}^{(3m+1)} , \qquad i \in T_i, j \in T_{i+1}$

b) $\lim\limits_{m\to\infty} s_{ij}^{(3m+2)} , \qquad i \in T_i , j \in T_{i+2} ,$

the other possibilities having zero value for all n.

a)

$$\lim_{m\to\infty} s_{ij}^{(3m+1)} = \lim_{m\to\infty} \frac{p_{ij}^{(3m+1)}}{\sum_{j\in T_{i+1}} p_{ij}^{(3m+1)}}$$

$$= \lim_{m\to\infty} \frac{\sum_{l\in T_i} p_{il}^{(3m)} p_{lj}}{\sum_{j\in T_{i+1}} \sum_{l\in T_i} p_{il}^{(3m)} p_{lj}}$$

$$= \lim_{m\to\infty} \frac{w_i^{(i)} \sum_{l\in T_i} v_l^{(i)} p_{lj}}{w_i^{(i)} \sum_{j\in T_{i+1}} \sum_{l\in T_i} v_l^{(i)} p_{lj}}$$

(6.6b)

$$\text{i.e.} \lim_{m\to\infty} s_{ij}^{(3m+1)} = v_j^{(i+1)} ,$$

from 6.I.

b) $$\lim_{m\to\infty} s_{ij}^{(3m+2)} = v_j^{(i+2)}$$

in a similar manner.

Once more, the results can be most conveniently expressed in matrix form, corresponding to Q^{3m}, Q^{3m+1}, Q^{3m+2},

(6.7a) $$\lim_{m\to\infty} [s_{ij}^{(3m)}] = \begin{bmatrix} e_1 v^{(1)'} & 0 & 0 \\ 0 & \frac{e_2 v^{(1)'} Q_{12}}{v^{(1)'} Q_{12} e_2} & 0 \\ 0 & 0 & \frac{e_3 v^{(1)'} Q_{12} Q_{23}}{v^{(1)'} Q_{12} Q_{23} e_3} \end{bmatrix}$$

(6.7b) $$\lim_{m\to\infty} [s_{ij}^{(3m+1)}] = \begin{bmatrix} 0 & \frac{e_1 v^{(1)'} Q_{12}}{v^{(1)'} Q_{12} e_2} & 0 \\ 0 & 0 & \frac{e_2 v^{(1)'} Q_{12} Q_{23}}{v^{(1)'} Q_{12} Q_{23} e_3} \\ e_3 v^{(1)'} & 0 & 0 \end{bmatrix}$$

(6.7c) $$\lim_{m\to\infty} [s_{ij}^{(3m+2)}] = \begin{bmatrix} 0 & 0 & \frac{e_1 v^{(1)'} Q_{12} Q_{23}}{v^{(1)'} Q_{12} Q_{23} e_3} \\ e_2 v^{(1)'} & 0 & 0 \\ 0 & \frac{e_3 v^{(1)'} Q_{12}}{v^{(1)'} Q_{12} e_2} & 0 \end{bmatrix}$$

where $v^{(1)'} Q_{12}\ Q_{23}\ Q_{31} = \rho^3\ v^{(1)'}$ and $v^{(1)'} e = 1$. (Here e_i is the unit vector with t_i elements.) Again note the similarity to the cyclic case, except that here the limit matrices are obtained as the limits of elements only, whereas in the cyclic case the corresponding matrices may also be obtained as limits of the powers P^{3m}, P^{3m+1}, P^{3m+2}, where P is the cyclic transition matrix. ($P \equiv Q$ in the above, for the special case when Q is stochastic)

6.4 Contrast with and Similarity to Cyclic Chains.

Since the matter is of considerable importance, we stress, once more, that allthe results in 6.1–6.3 have been obtained for an absorbing chain with cyclic transient T; however since the fact that $\rho < 1$ has nowhere been used, everything is valid when Q is a stochastic matrix i.e. the transition matrix of a cyclic chain (see 4.3 and 5.4). This is in accordance with our aim of getting a more general theory, of which the corresponding Markov chain is a particular case.

Therefore, making the appropri... ...anges in (6.3)-(6.7c) i.e. putting $w^{(i)} = e_i$ and noting that $Q_{i\,i+1} e_{i+1} = e_i$ in which case all the denominators of (6.7a)-(6.7c) become unity, familiar results for the cyclic chain emerge. For example, (6.3) yields

$$\lim_{m\to\infty} \frac{1}{m} \sum_{m=0}^{m} p_{ij}^{(m)} = v_j \qquad i,j \in T$$

where v_j is the j th element of $\frac{1}{3}[v^{(1)'}, v^{(2)'}, v^{(3)'}]$, which is, in the stochastic case, the unique positive left eigenvector corresponding to ρ. This is just an expression of the Césaro summability of $\{p_{ij}^{(m)}\}$, for a cyclic Markov chain (see e.g. [19], p.101). Similarly (6.7a)-(6.7c) yield the well known limiting results for the n-step transition probabilities in the cyclic case.

The fact that this treatment is a 'considerable' generalization, emerges more clearly by noting that, when Q is not stochastic,

$$\frac{1}{3}[\underset{\sim}{v}^{(1)'}, \underset{\sim}{v}^{(2)'}, \underset{\sim}{v}^{(3)'}]$$

is not the unique positive left eigenvector, corresponding to $\rho \neq 1$. For suppose we denote this last by $\underset{\sim}{h}' = [\underset{\sim}{h}_1' \ \underset{\sim}{h}_2' \ \underset{\sim}{h}_3']$, such that $\underset{\sim}{h}'\underset{\sim}{e} = 1$. Then

$$\underset{\sim}{h}_3' Q_{31} = \rho \underset{\sim}{h}_1' \tag{6.8a}$$

$$\underset{\sim}{h}_1' Q_{12} = \rho \underset{\sim}{h}_2' \tag{6.8b}$$

$$\underset{\sim}{h}_2' Q_{23} = \rho \underset{\sim}{h}_3' \tag{6.8c}$$

Moreover, $\underset{\sim}{h}'$ is a left eigenvector of Q^3, corresponding to ρ^3 i.e. from (6.1a)-(6.1c) we have that

$$\underset{\sim}{h}_1' = \delta_1 \underset{\sim}{v}^{(1)'}, \quad \underset{\sim}{h}_2' = \delta_2 \underset{\sim}{v}^{(2)'}, \quad \underset{\sim}{h}_3' = \delta_3 \underset{\sim}{v}^{(3)'}.$$

Therefore we must determine $\delta_1, \delta_2, \delta_3 \geq 0$, such that $\delta_1 + \delta_2 + \delta_3 = 1$. From (6.8a) and (6.8c), it follows that

$$\delta_1 \underset{\sim}{v}^{(1)'} Q_{12} \underset{\sim}{e}_2 = \rho \delta_2, \quad \delta_3 \underset{\sim}{v}^{(3)'} Q_{31} \underset{\sim}{e}_1 = \rho \delta_1,$$

whence

$$(6.9) \qquad \mathbf{h}' = \left[\theta_1 \mathbf{v}^{(1)'},\ \theta_1 \frac{\mathbf{v}^{(1)'} Q_{12} \mathbf{e}_2 \mathbf{v}^{(2)'}}{\rho},\ \frac{\theta_1 \rho \mathbf{v}^{(3)'}}{\mathbf{v}^{(3)'} Q_{31} \mathbf{e}_1} \right],$$

where θ_1 is determined from the normalizing condition

$$\theta_1 \left(1 + \frac{\mathbf{v}^{(1)'} Q_{12} \mathbf{e}_2}{\rho} + \frac{\rho}{\mathbf{v}^{(3)'} Q_{31} \mathbf{e}_1} \right) = 1.$$

If Q is stochastic, the normalizing condition becomes $3\theta_1 = 1$ i.e. $\theta_1 = \frac{1}{3}$ and

$$\mathbf{h}' = \frac{1}{3} \left[\mathbf{v}^{(1)'},\ \mathbf{v}^{(2)'},\ \mathbf{v}^{(3)'} \right]$$

as required.

Finally, it is natural to enquire in this more general case about the Césaro limit

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n} s_{ij}^{(m)}, \qquad i, j \in T.$$

This exists and is positive for each i, j $\in$ T –and is independent of i –by the same argument as e.g. that used by Kemeny and Snell,[19] p.101. The result in matrix form is

$$(6.10) \qquad \lim_{n \to \infty} \left[\frac{1}{n} \sum_{m=0}^{n} s_{ij}^{(m)} \right] = \frac{1}{3} \begin{bmatrix} \mathbf{e}_1 \mathbf{v}^{(1)'} & \dfrac{\mathbf{e}_1 \mathbf{v}^{(1)'} Q_{12}}{\mathbf{v}^{(1)'} Q_{12} \mathbf{e}_2} & \dfrac{\mathbf{e}_1 \mathbf{v}^{(1)'} Q_{12} Q_{23}}{\mathbf{v}^{(1)'} Q_{12} Q_{23} \mathbf{e}_3} \\ \mathbf{e}_2 \mathbf{v}^{(1)'} & \dfrac{\mathbf{e}_2 \mathbf{v}^{(1)'} Q_{12}}{\mathbf{v}^{(1)'} Q_{12} \mathbf{e}_2} & \dfrac{\mathbf{e}_2 \mathbf{v}^{(1)'} Q_{12} Q_{23}}{\mathbf{v}^{(1)'} Q_{12} Q_{23} \mathbf{e}_3} \\ \mathbf{e}_3 \mathbf{v}^{(1)'} & \dfrac{\mathbf{e}_3 \mathbf{v}^{(1)'} Q_{12}}{\mathbf{v}^{(1)'} Q_{12} \mathbf{e}_2} & \dfrac{\mathbf{e}_3 \mathbf{v}^{(1)'} Q_{12} Q_{23}}{\mathbf{v}^{(1)'} Q_{12} Q_{23} \mathbf{e}_3} \end{bmatrix}$$

as expected.

NOTE:- If $\mathbf{w}$ is the unique right eigenvector corresponding to Q, in a manner analogous to the above one may write it as

$$\mathbf{w} = \left[\theta_1^* \mathbf{w}^{(1)'},\ \theta_2^* \mathbf{w}^{(2)'},\ \theta_3^* \mathbf{w}^{(3)'} \right]'$$

but the normalizing conditions are awkward again. We require that $\underset{\sim}{v}'\underset{\sim}{w} = 1$, which implies

$$\theta_2^* = \frac{\rho\,\theta_1^*}{\underset{\sim}{v}^{(1)'} Q_{12}\, \underset{\sim}{w}^{(2)}} \qquad \theta_3^* = \frac{\underset{\sim}{v}^{(3)'} Q_{31}\, \underset{\sim}{w}^{(1)}\,\theta_1^*}{\rho}$$

$$\theta_1\,\theta_1^* \left[1 + \frac{\underset{\sim}{v}^{(1)'} Q_{12}\, \underset{\sim}{e}_2}{\underset{\sim}{v}^{(1)'} Q_{12}\, \underset{\sim}{w}^{(2)}} + \frac{\underset{\sim}{v}^{(3)'} Q_{31}\, \underset{\sim}{w}^{(1)}}{\underset{\sim}{v}^{(3)'} Q_{31}\, \underset{\sim}{e}_1} \right] = 1 .$$

Thus, except when Q is stochastic (then $\theta_1\theta_1^* = \frac{1}{3}$),

$$[\, w_1 v_1,\ w_2 v_2,\ w_3 v_3,\ \ldots\ \ldots,\ w_s v_s \,]$$

$$\neq \frac{1}{3}\left[w_1^{(1)} v_1^{(1)}, \ldots, w_{t_1}^{(1)} v_{t_1}^{(1)},\ w_1^{(2)} v_1^{(2)}, \ldots\ w_{t_2}^{(2)} v_{t_2}^{(2)},\ w_1^{(2)} v_1^{(2)}, \ldots, w_{t_3}^{(3)} v_{t_3}^{(3)} \right].$$

7. TRANSITION AND ABSOLUTE QUASI-PROBABILITIES

7.I The Ergodic Analogy.

It is possible to derive many results common to regular and cyclic Markov chains without distinguishing between them; thus the two cases may be treated under the common name of ergodic chains. The same is true for our generalized analogues to these two cases, which have their fundamental interest in the absorbing chain sense.

Reviewing Chapters 4, 5, and 6, which deal with an indecomposable Q, we see that the quantities

$$s_{ij}^{(n)} = \frac{p_{ij}^{(n)}}{\sum_{j \in T} p_{ij}^{(n)}} \qquad i, j \in T$$

behave in a manner similar to transition probabilities of the corresponding ergodic chains. Moreover, quasi-stationary distributions defined on T exist, involving the left eigenvector of the matrix Q. Not only are the $s_{ij}^{(n)}$ Césaro summable as expected, but there is another generalization of the Césaro summability, viz.

$$\lim_{n \to \infty} \frac{1}{n} E[X_{ij}^{(n)} / X_i > n],$$

which also behaves analogously to the case of ergodic chains. Thus we have some justification in calling the quantities $s_{ij}^{(n)}$ n-step transition quasi-probabilities; the quantities

$$\frac{\sum_{i \in T} \pi_i p_{ij}^{(n)}}{\sum_{i \in T} \pi_i \sum_{j \in T} p_{ij}^{(n)}}$$

the absolute quasi-probabilities at time n; and consider the whole 'process' as being a quasi-chain (ergodic), which is not necessarily a Markov chain.

Possibly the most interesting consequence of taking this point of view, is that a <u>reverse chain</u> can be defined on the states of T for an ergodic quasi-chain, the former being in fact a true Markov chain, and being significant in the quasi-stationary context. The procedure is analogous to that for the ordinary ergodic chain (see Feller, [13] p.373).

7.2 <u>The Reverse Process.</u>

A reverse process can be defined on the transient set T of an absorbing Markov chain, the transition probability from j to $k \in T$ being

$$P[\, i_m = k \;/\; i_{m+1} = j \,] = \frac{\pi_k^{(m)} p_{kj}}{\pi_j^{(m+1)}}, \qquad j, k \in T. \tag{7.1}$$

This reverse process is a Markov <u>chain</u> iff this quantity is constant with respect to m for all k, $j \in T$. The necessary and sufficient condition for this (for indecomposable $\mathbf{Q}$) is

$$\frac{\pi_k^{(m)}}{\pi_j^{(m+1)}} = \frac{v_k}{\rho\, v_j}, \qquad j, k \in T. \tag{7.2}$$

The sufficiency is obvious. Necessity for primitive $\mathbf{Q}$ follows, since for all allowable distributions i.e. with some mass in T,

$$\frac{\pi_k^{(m)}}{\pi_j^{(m+1)}} \to \frac{v_k}{\rho\, v_j}, \qquad j, k \in T.$$

For cyclic Q, necessity does not appear to be straightforward.

Let g be the period of $\mathbf{Q}$. Then the maximal modulus eigenvalues are all different and given by

$$\rho_n = \rho\, e^{\frac{2\pi i n}{g}}, \qquad n = 0, 1, \dots, g-1,$$

as before. Hence

$$\rho > |\rho_{g+1}| \geq |\rho_{g+2}| \geq \dots \dots \geq |\rho_{s-g}|.$$

The spectral decomposition is therefore

$$Q^m = \rho^m \underset{\sim}{w}\,\underset{\sim}{v}' + \rho^m \Big[\sum_{s=1}^{g-1} \underset{\sim}{w_s}\,\underset{\sim}{v_s}' e^{\frac{2\pi i s m}{g}} \Big] + O(m^\phi |\rho_{g+1}|^m),$$

where $\phi+1$ is the multiplicity w.l.o.g. of ρ_{g+1}. It is clear that no element of the matrix

$$(7.3) \qquad \Big[\sum_{s=1}^{g-1} \underset{\sim}{w_s}\,\underset{\sim}{v_s}' e^{\frac{2\pi i s m}{g}} \Big]$$

can be constant with respect to m, for no element can be zero for all m; if, for instance, the j, k th element were, then

$$\frac{p_{jk}^{(m)}}{\rho^m}$$

would have a limit $w_j v_k > 0$ as $m \to \infty$, – impossible as $\frac{Q}{\rho}$ is cyclic. Therefore, each element of (7.3) is a function of m, bounded but having no limit as $m \to \infty$. We can therefore write that for suitable $\underset{\sim}{\pi}$ and sufficiently large m

$$\frac{\pi_k^{(m)}}{\pi_j^{(m+1)}} = \frac{\sum_{\alpha\in T} \pi_\alpha w_\alpha v_k + \psi_k(m)}{\rho \sum_{\alpha\in T} \pi_\alpha w_\alpha v_j + \rho e^{\frac{2\pi i}{g}} \psi_j(m)} + O(m^\phi |\rho_{g+1}|^m)$$

where the function $\psi_j(m)$ is obvious from (7.3), for all j, and depends on $\underset{\sim}{\pi}$. Now, if the left hand side is constant for all m, for any j, $k \in T$, then putting j=k

$$\frac{\pi_j^{(m)}}{\pi_j^{(m+1)}} = \frac{\sum_{\alpha\in T} \pi_\alpha w_\alpha v_j + \psi_j(m)}{\rho \sum_{\alpha\in T} \pi_\alpha w_\alpha v_j + \rho e^{\frac{2\pi i}{g}} \psi_j(m)} = c_j,$$

where $c_j > 0$. Therefore

$$\psi_j(m) \Big[1 - \rho e^{\frac{2\pi i}{g}} \Big] = \sum_{\alpha\in T} \pi_\alpha w_\alpha [c_j \rho v_j - v_j].$$

Hence $\psi_j(m)$ is constant, and since $\psi_j(m+1) = e^{\frac{2\pi i}{g}} \psi_j(m)$, $\psi_j(m) = 0$

for all $j \in T$. Therefore

$$\frac{\pi_k^{(m)}}{\pi_j^{(m+1)}} = \frac{v_k}{\rho v_j} .$$

It is, moreover, clear that the above, (7.2), holds for the quasi-ergodic case if the forward process starts in state $j \in T$ with probability $\lambda v_j, 0 < \lambda \le 1$, $j \in T$. In the case of primitive Q (quasi-regular case), it also holds if a sufficiently long time has elapsed for $\frac{\pi_k^{(m)}}{\pi_j^{(m+1)}}$ to converge to $\frac{v_k}{\rho v_j}$.

Therefore, when the reverse process is a Markov chain, its transition matrix, defined on the states of T is

$$P^* = [\, p_{ij}^* \,]$$

where $$p_{ij}^* = \frac{v_j p_{ji}}{\rho v_i} \qquad i, j \quad T$$

i.e. $$P^* = \frac{1}{\rho} V^{-1} Q' V$$

where $V = \mathrm{diag}[v_1, v_2, \ldots\ldots, v_s]$. The stationary probability vector of this ergodic Markov chain is easily seen to be

$$\underset{\sim}{v}^{*\prime} = [w_1 v_1, w_2 v_2, \ldots\ldots\ldots, w_s v_s].$$

This is just the quasi-stationary(vector)distribution discussed in Chapter 4, at least when Q is primitive (see NOTE, 6.4 also). This is therefore a pleasing link with the material of this chapter. In the present case $\rho^* = 1$, and $\rho_2^* = \frac{\rho_2}{\rho}$ with multiplicity k+1.

Finally, for primitive Q (and hence P^*), an intuitive description is enlightening. Go forward in time sufficently far for $\pi_j^{(m)}/\pi_i^{(m+1)}$ to be replaced by $v_j/\rho v_i$. Given that the forward processis still in T, come backward for a length of time m along the reverse process. The probability of being in state j is then

$$w_j v_j + O(m^k (\frac{|\rho_2|}{\rho})^m).$$

This interpretation of $w_j v_j$ is qualitatively similar to (4.4).

7.3 The Chapman-Kolmogorov Equation: An Example.

In view of the foregoing, the question that naturally arises is, when do the n-step probabilities $s_{ij}^{(n)}$, i, j $\in$ T actually behave as the transition probabilities of an ordinary Markov chain i.e. when do they satisfy the Chapman-Kolmogorov equation?

This is not true in general, so that we are interested in necessary and sufficient conditions on the $s_{ij}^{(n)}$ to make them satisfy

$$\sum_{j \in T} p_{ij}^{(m)} p_{jk}^{(n)} = p_{ik}^{(m+n)}, \quad m, n = 0, 1, 2, \ldots \quad i, j \in T \tag{7.4}$$

A sufficient condition is that the rows of Q have equal sums $\lambda \leq 1$, say. In the case of interest $\lambda < 1$, the transient states of T are said to be lumpable (Kemeny and Snell, [19]). Then

$$\rho = \lambda \quad , \quad \underset{\sim}{w} = \underset{\sim}{e}$$

and the quasi-stationary probabilities v_j and $w_j v_j$ are equal. An example of such a matrix is the matrix Q of 3.3, in which case it is particularly interesting that

$$s_{ij}^{(n)} = \frac{p_{ij}^{(n)}}{1 - p_{i0}^{(n)}} = v_j = \beta_j$$

for all n.

7.4 A More General Context.

Looking back at the material upto the present, let us notice that we began by considering the important question of pseudo-stationarity in absorbing Markov chains, this giving rise to the allied concept of what has been described as quasi-chain behaviour. All our theory to date has therefore been developed in the context of the transient states of an absorbing Markov chain. It is clear, however, that all our quantities can be defined on and treated by the use of only the substochastic matrix $\mathbf{Q}$ (and of course the initial probability). Thus, in a broader sense, we are not at all concerned with the explicit fact that the 'parent' chain is absorbing, but only with a set of states T, transitions between which are governed by a substochastic matrix $\mathbf{Q}$, the set T being not necessarily exhaustive (if it is , this becomes the case when $\mathbf{Q}$ is stochastic). Notice that this viewpoint is closely related to Bartlett's objection, discussed in 5.3.

Thus, Mandl [22] considers any homogeneous Markov chain with an initial probability distribution; in particular he takes any subset T of states and discusses the asymptotic behaviour of the conditional probability

$$(7.5) \qquad P[i_n = i \;/\; T^n]$$

that the system will be in state $i \in T$ at time n, under the condition that it does not leave the class T.

The relation to our ideas is easy to see if we denote by $\mathbf{Q}$ the matrix of transition probabilities p_{ij} between the states of T only. Then if we define by $p_{ij}^{(n)}$, i, $j \in T$ the elements of the matrix $\mathbf{Q}$, then (7.5) becomes

$$(7.6) \qquad \frac{\sum_{i \in T} \pi_i \, p_{ij}^{(n)}}{\sum_{i \in T} \pi_i \sum_{j \in T} p_{ij}^{(n)}}$$

with which quantity we have dealt, at least for indecomposable Q, being nothing more than (5.3). Notice that Mandl's treatment also allows Q to be stochastic, i.e. T may include all the states of his chain. Before proceeding, it is worth noticing that (7.5) is quite different from Chung's concept of 'taboo' probabilities, [7] .

The way in which Mandl's paper is valuable as far as this thesis is concerned, is that he also treats the limiting behaviour of (7.5), and so (7.6), for the case when Q is decomposable. We shall return to this later.

7.5 Note on a Paper of Breny.

It seems appropriate, as a conclusion to this chapter, to mention an interesting paper of Breny [6] , primarily because we can arrive at an interesting consequence of some of the theory we have developed. A second reason is that the method put forward leads to the construction of several absorbing chains with one absorbing state from one with several absorbing states. Since our theory is only relevent to the transient set T, his ideas could be further developed as a means of 'transforming' the matrix Q. How this comes about will be explained below; however we will not pursue the topic beyond this section, although it is related, as will be shown, to the nature of the problems under investigation.

Breny's more general absorbing chain has transition matrix

$$(7.7) \qquad P = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}$$

where P is (s+p)x(s+p), I is the identity matrix, pxp, and Q corresponding to the set T is sxs as usual; R is the matrix of absorption transitions and is sxp, being the generalization of p_0 previously. (We adopt Breny's argument to our needs.) Giving as his reason the fact that the study of the passageof a homogeneous chain through its indecomposable groups, has as its chief tool the theory of absorbing chains, he sets out to derive for the chain governed by P the probabilities

$$P\,[\,i_{n+1} = j \mid i_n = i,\ \omega = k\,] \tag{7.8}$$

i.e 'transition probabilities', given that the state ω into which the system is absorbed, is k, for i, j $\in$ T or = k. It is understood that ω, j can be reached from all i $\in$ T. Subsequently, he proves that the derived process, defined on T and k is Markovian, the transition probabilities (7.8) also being independent of the time parameter n. Thus for every absorbing state k, a new Markov absorbing chain can be defined, having transition matrix

$$P_{|k} = \begin{bmatrix} 1 & \underset{\sim}{0}' \\ R_{|k} & Q_{|k} \end{bmatrix}$$

where

$$Q_{|k} = D_k^{-1} Q\, D_k, \quad s\times s; \quad R_{|k} = D_k^{-1} R\, \underset{\sim}{e}_k, \quad s\times 1;$$

$$D_{|k} = \mathbf{diag}\,(\, \underset{\sim}{f}_1'[I - Q]^{-1} R \underset{\sim}{f}_k, \ldots\ldots, \underset{\sim}{f}_s'[I - Q]^{-1} R \underset{\sim}{f}_k)$$

If, for simplicity, we suppose Q primitive, we obtain an interesting connection between it and $Q_{|k}$, as regards our theory, for since both have exactly the same set of eigenvalues, the rates of convergence to the quasi-stationary distributions are the same. Moreover, since the eigenvectors are transformed according to

$$\underset{\sim}{v} \xrightarrow{D_k} D_k \underset{\sim}{v}, \qquad w \xrightarrow{D_k} D_k^{-1} w$$

we see that although the quasi-stationary distribution $\underset{\sim}{v}$

changes, the other viz. $[w_1v_1, \ldots\ldots\ldots, w_sv_s]$ does not.

We have seen in 7.2 already, that a modified similarity transformation of Q has an important consequence, the reverse chain. In this section, another similarity transformation has also led to interesting properties.

8. THE DECOMPOSABLE TRANSIENT SET OF STATES

This chapter consists of a brief survey of Mandl's paper, [22] , in which he shows the existence of and derives expressions for

$$\lim_{n\to\infty} \frac{\sum_{i\in T} \pi_i \, p_{ij}^{(n)}}{\sum_{i\in T} \pi_i \sum_{j\in T} p_{ij}^{(n)}}$$

when the set of states T (and so the matrix Q) is decomposable.

8.I The Decomposable Transition Submatrix.

The Markov chain which has a decomposable transition matrix P, is one with a finite number of closed indecomposable sets of states, the stochastic submatrices corresponding to these being indecomposable. There may also be a set of transient states T, which the system leaves in a finite time entering the closed sets. The limiting behaviour of the n-step transition probabilities of such a chain (apart from the probabilities of transition into the absorbing sets) can be studied by using the results for an ergodic chain, which has indecomposable matrix. No difficulty is presented by the elements of Q^n , since $Q^n \to 0$ as $n\to\infty$. The analogy to the above in the present case is the study of the above <u>conditional</u> probabilities for a set of transient states T which is decomposable. It is clear that the eigenvalue structure of a decomposable Q is comparitively more complex than for a decomposable P, and in fact the former requires quite involved treatment, 22 , as compared to the latter. If we rewrite the matrix Q in the same form as we would its stochastic analogue by a suitable relabelling of the states in T, we have

$$Q = \begin{bmatrix} Q_{11} & & & \\ & Q_{22} & 0 & \\ & & \ddots & \\ 0 & & Q_{pp} & \\ \cdots & R & \cdots & \vdots\; Q_{p+1\,p+1} \end{bmatrix}$$

where Q_{11} , ..., $Q_{p+1\,p+1}$ are square. It will be seen that we cannot in general neglect the effect of the matrix $Q_{p+1\,p+1}$, which may itself be decomposable. Moreover, Q_{11} , Q_{22} , ...$Q_{p+1\,p+1}$ may each have any spectral radius not exceeding unity, which is one cause of the relatively greater difficulty in treatment, than in the stochastic analogue.

8.2 A Summary of Mandl's Results.

Any set of states whose submatrix of transition probabilities is primitive is defined as regular, and the largest positive eigenvalue (spectral radius) of the submatrix is called its characteristic number. Before proceeding, Mandl restricts the complete arbitrariness of the set T, by assuming that every state of T communicates with itself within T. This enables every state of T to fall into one of the subsets T of communicating states into which the set T can be divided. Secondly, these subsets are assumed regular. They can then be partially ordered in the usual way, i.e. define a relation '$\leq$ ' on the class of subsets of T, to mean that $T_k \leq T_j$ if it is possible to pass from T_k to T_j . This is the well known ordering of antecedent to consequent, and may be found e.g. in Chung [7] and Romanovskii [29] . A last assumption states that at time zero the system will, with positive probability, be in the minimal subsets T i.e. those T for which there exists no $T_k \leq T_j$, where $T_k \neq T_j$. In the paper, this is referred to as 'Condition A', but is in effect trivial, as will be seen in a moment.

The theory is developed by first additionally assuming that T is split into sets T_1 , T_2 , ..., T_l such that from T_j the system can pass only into T_{j+1}. It is then proved that

$$\lim_{n \to \infty} \frac{\sum_{i \in T_1} \pi_i\, p_{ij}^{(n)}}{\rho^n\, n^{h-1}}$$

exists, and is positive, where $\sum_{i \in T_1} \pi_i = 1$; $j \in T_1$,

$$\rho = \max_{j=1,2,\ldots,l} \rho_j \quad ,$$

ρ_j being the characteristic number of T_j , and h is the number of times ρ occurs among the ρ_j . An expression is derived for this limit.

When the additional assumption is removed, it is proved that for every $j \in T$

$$\lim_{n \to \infty} \frac{\sum_{i \in T} \pi_i \, p_{ij}^{(n)}}{\sum_{i \in T} \pi_i \, (1 - p_{io}^{(n)})}$$

exists, and the necessary and sufficient condition for it to be positive is given. Its value in general depends on the initial distribution, on the arrangement of the subsets T_k within T, and on their characteristic numbers.

A result of particular interest to us, Mandl's Theorem 5, states that if there is a unique h-member sequence of subsets, each with characteristic number where h is the largest positive number (integer) for which such a sequence can exist i.e.

$$T_{s_1} \leq T_{s_2} \leq \quad . \quad . \quad . \quad . \quad . \quad \leq T_{s_h} \quad , \quad T_{s_i} \neq T_{s_j}$$

then the limit is independent of the initial distribution. An example is given in the next section.

8.3 Comment and Criticism.

The reason why Mandl assumes that every state within T communicates with itself, is to avoid increased complexity in the proofs of his theorems e.g. Theorem 2. If there exists a state of T for which the above condition does not hold, then it can be made into a group by itself, and fitted into the partial ordering, as usual, [7]. Since its transition submatrix is the zero element, this complicates matters as far as the spectral radius is concerned. It is therefore more convenient not to consider sets T with such 'nuisance' subsets, which involve modifications in the proof, nor really fit in with the assumption that all the T are regular.

Condition A, stated here as in Mandl's paper, is ambiguous. What is intended, as emerges from the paper, is that <u>each</u> of the minimal subsets must have a positive probability of being occupied initially. If not, then the minimal subsets having zero initial probability can be disregarded as, clearly, they they do not affect the behaviour in other subsets, and have zero limiting probabilities themselves. That this condition is really trivial follows from the fact that if there is an arbitrary initial probability distribution over T, then if T_j is a least subset having initial positive probability, (in the sense that for all k such that $T_k \leq T_j$ $k \neq j$, there is zero initial probability), then all $T_k \leq T_j$, $k \neq j$, have zero limiting conditional probability for every state of T_k, and can be discarded as they have no effect on the behaviour within other subsets. Hence there is really no loss of generality when Condition A is applied.

8.4 <u>A Numerical Example</u>.

Mandl [22] gives an interesting example at the conclusion of his paper, but the following numerical one is helpful in demonstrating the validity of his Theorem 5 (see 8.2). Consider

$$Q = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1/2 & 0 & 0 \\ \frac{1}{6} & 2/3 & 0 \\ \frac{1}{6} & \frac{1}{12} & 2/3 \end{bmatrix} \end{array} .$$

Here the regular subsets are in fact the states 1, 2, and 3 themselves, and there is a unique two member sequence of maximal length h = 2 and characteristic number 2/3. Utilizing the partial fractions approach (or equivalently

Perron's formula, [29])

$$[I - zQ]^{-1} = \begin{bmatrix} (1-\tfrac{1}{2}z)^{-1} & 0 & 0 \\ \tfrac{z}{6}(1-\tfrac{2}{3}z)^{-1}(1-\tfrac{1}{2}z)^{-1} & (1-\tfrac{2}{3}z)^{-1} & 0 \\ \tfrac{z}{6}(1-\tfrac{7}{12}z)(1-\tfrac{2}{3}z)^{-2}(1-\tfrac{1}{2}z)^{-1} & \tfrac{z}{12}(1-\tfrac{2}{3}z)^{-2} & (1-\tfrac{2}{3}z)^{-1} \end{bmatrix}$$

whence

$$Q^{n} = \begin{bmatrix} (\tfrac{1}{2})^n & 0 & 0 \\ (\tfrac{2}{3})^n - (\tfrac{1}{2})^n & (2/3)^n & 0 \\ -(\tfrac{1}{2})^{n+1} + \tfrac{1}{8}(n+1)(\tfrac{2}{3})^n + \tfrac{3}{8}(\tfrac{2}{3})^n & \tfrac{n}{12}(2/3)^{n-1} & (\tfrac{2}{3})^n \end{bmatrix}$$

Thus if the initial distribution over T is $\underset{\sim}{\pi} = [\pi_1, \pi_2, \pi_3]'$ $0 < \underset{\sim}{\pi}'\underset{\sim}{e} \leq 1$, then considering

$$\lim_{n \to \infty} \frac{\sum_{i \in T} \pi_i \, p_{ij}^{(n)}}{\sum_{i \in T} \pi_i \, (1 - p_{i0}^{(n)})}$$

<u>j = 1</u> $$\lim_{n \to \infty} \frac{\pi_1(\tfrac{1}{2})^n + \pi_2((\tfrac{2}{3})^n - (\tfrac{1}{2})^n) + \pi_3(-(\tfrac{1}{2})^{n+1} + \tfrac{1}{8}(n+1)(\tfrac{2}{3})^n + \tfrac{3}{8}(2/3)^n)}{\pi_1(\tfrac{1}{2})^n + \pi_2(2(\tfrac{2}{3})^n - (\tfrac{1}{2})^n) + \pi_3(-(\tfrac{1}{2})^{n+1} + \tfrac{1}{8}(n+1)(\tfrac{2}{3})^n + \tfrac{3}{8}(\tfrac{2}{3})^n + \tfrac{n}{12}(\tfrac{2}{3})^{n-1} + (\tfrac{2}{3})^n)};$$

if $\pi_3 > 0$

$$\lim_{n \to \infty} \frac{\pi_3 \left(\tfrac{1}{8}(n+1)(\tfrac{2}{3})^n \right)}{\pi_3 \left(\tfrac{1}{8}(n+1)(\tfrac{2}{3})^n + \tfrac{n}{12}(\tfrac{2}{3})^{n-1} \right)} = \frac{1}{2}$$

if $\pi_3 = 0$, $\pi_2 > 0$, then the limit is $\frac{1}{2}$.

if $\pi_3 = 0$, $\pi_2 = 0$, $\pi_1 > 0$, the limit is 1.

<u>j = 2</u>

if $\pi_3 > 0$, the limit is $\frac{1}{2}$.

if $\pi_3 = 0$, $\pi_2 > 0$, the limit is $\frac{1}{2}$.

if $\pi_3 = 0$, $\pi_2 = 0$, $\pi_1 > 0$, the limit is zero.

<u>j = 3</u>

In all cases the limit is zero.

Condition A requires $\pi_3 > 0$, and notice e.g. for j = 1 that π_3 cancelsout; and the answer is independent of $\underset{\sim}{\pi}$. If $\pi_3 = 0$, T_3 can be disregarded, and if $\pi_2 > 0$ the independence of $\underset{\sim}{\pi}$ is obtained for the same reason, on T_1 and T_2. If $\pi_2 = \pi_3 = 0$, the same reasoning applies, giving the obvious.

9. THE FINITE CONTINUOUS PARAMETER CASE

This chapter summarizes the results for continuous time chains arising out of investigation analogous to the discrete time case. It is therefore much shorter than the preceding material, and also because cyclicity does not arise with the continuous parameter.

9.I Prerequisites.

For the basic theory of finite continuous parameter chains we refer to Doob [9] . Assume, as usual, that if the (stationary) matrix transition function is $P(t)$, that

$$\lim_{t\to 0} p_{ij}(t) = \delta_{ij} .$$

This implies that $p_{ij}(t)$ has a derivative $p_{ij}'(t)$ for all $t \geq 0$, whence we can put

$$q_i = \lim_{t\to 0} \frac{1 - p_{ii}(t)}{t} = - p_{ii}'(0)$$

$$q_{ij} = \lim_{t\to 0} \frac{p_{ij}(t)}{t} = p_{ij}'(0), \quad i \neq j .$$

The matrix $[q_{ij}]$, where q_{ii} is $-q_i$, is denoted by R. Then we have

$$q_{ij} \geq 0, \quad q_i \geq 0, \quad \sum_{j\neq i} q_{ij} = q_i$$

$$P(t) = e^{Rt} . \tag{9.1}$$

The simple absorbing chain in which we are interested has transition function which can be written

$$P(t) = \begin{bmatrix} 1 & \underset{\sim}{0}' \\ \underset{\sim}{p}_0(t) & Q(t) \end{bmatrix} \qquad (\underset{\sim}{p}_0(t) \neq \underset{\sim}{0} , t > 0) \tag{9.2}$$

where the matrix is (s+1)x(s+1) and the other dimensions are as before. However, it is as usual of more relevence to consider the matrix R of infinitesimal transition probabilities (probability densities)

$$(9.3)\qquad R = \begin{bmatrix} 0 & \underset{\sim}{0}' \\ \underset{\sim}{q}_0 & C \end{bmatrix}, \qquad \underset{\sim}{q}_0 \neq \underset{\sim}{0}$$

with dimensions analogous to (9.2). Notice that since $\underset{\sim}{q}_0 \neq \underset{\sim}{0}$, and $\underset{\sim}{q}_0 \geq \underset{\sim}{0}$ then at least one of the row sums of C must be less than zero.

The set T of transient states 1,2,...,s shall be called regular if every state of T can be reached from every other state. This determines an indecomposable (though not non-negative !) matrix C, and is the case of fundamental importance. Then the matrix

$$(9.4)\qquad Q(t) = e^{Ct}$$

is an indecomposable substochastic matrix for any t 0; since we are dealing with an absorbing chain, it is, as before, strictly substochastic. Following Mandl [23] , Theorem 1, there corresponds to C a chracteristic root ρ , having maximal real part. This root is real,simple, and less than zero. Corresponding to it are positive left and right eigenvectors $\underset{\sim}{v}', \underset{\sim}{w}$, and we can write

$$(9.5)\qquad Q(t) = e^{Ct} = e^{t\rho}\underset{\sim}{w}\,\underset{\sim}{v}' + o(e^{t\rho'})$$

where $\underset{\sim}{v}'\underset{\sim}{w}, \underset{\sim}{v}'\underset{\sim}{e} = 1$, and $\rho' < \rho$. This expression is the analogue of (1.2).

Before proceeding, we note that we can no longer use the expressions

'time spent in j T before absorption'

and

'number of visits to j T before absorption'

interchangeably, the former being the relevent one in our is the ratio of means distribution, which is dependent on $\underset{\sim}{\pi}$. This expression is valid when T is regular.

9.3 The Limiting Conditional Mean Ratio.

We will only derive the expression analogous to (4.3) i.e. an expression for

$$\lim_{t\to\infty} E\left[\frac{X_{ij}(t)}{t} / X_i > t\right] , \quad i,j \in T$$

where $X_{ij}(t)$ is the time spent in state $j \in T$ in time t, having started from $i \in T$. Referring to 4.2, it is not difficult to see that

$$E\left[\frac{X_{ij}(t)}{t} / X_i > t\right] = \frac{1}{t}\,\frac{\int_0^t p_{ij}(x)(1-p_{jo}(t-x))\,dx}{1-p_{io}(t)}$$

$$= \frac{\int_0^t e^{\rho t}(w_i v_j w_j \sum_{k\in T} v_k)\,dx + O(e^{\rho' t})}{t\, e^{\rho t} w_i \sum_{k\in T} v_k + o(t e^{t\rho'})}$$

$$= w_j v_j + O\left(\frac{1}{t}\right),$$

for regular T, using (9.5). Hence

$$\lim_{t\to\infty} E\left[\frac{X_{ij}(t)}{t} / X_i > t\right] = w_j v_j \tag{9.8}$$

as is expected. The expressions related to (9.8) hold true also, and can be derived similarly e.g. notice that

$$E\left[\frac{X_{ij}}{X_i} / X_i = t\right] = \frac{1}{t}\,\frac{\int_0^t p_{ij}(x)\sum_{k\in T} p_{jk}(t-x)\,q_{ko}\,dx}{\sum_{k\in T} p_{ik}(t)\,q_{ko}},$$

and that $\sum_{k\in T} v_k q_{ko} > 0$, from (9.3).

It is not practicable to use a generating function, e.g. Laplace transform, approach here.

9.4 The Stationary Conditional Distribution.

Let us again consider the conditional probability distribution, restricted to the transient states, where the distribution over all (s+1) states at time t is

$[\pi_0(t), \underset{\sim}{\pi}'(t)]$. This is

$$(9.9) \qquad d(t) = \frac{\underset{\sim}{\pi}(t)}{1-\pi_0(t)},$$

and call it a quasi-stationary distribution if

$$d(t) = d, \qquad t \geq 0.$$

Then since

$$\underset{\sim}{\pi}'(t_1)\, Q(t_2) = \underset{\sim}{\pi}'(t_1 + t_2), \qquad t_1, t_2 \geq 0$$

it follows

$$d(t_1)\, Q(t_2) = \rho(t_1, t_2)\, \underset{\sim}{d}'(t_1 + t_2).$$

If d exists it satisfies

$$\underset{\sim}{d}'\, Q(t_2) = \rho(t_1, t_2)\, \underset{\sim}{d}'$$

true for all t_1, $t_2 \geq 0$. Now, since T is regular, for fixed $t_2 > 0$ $Q(t_2)$ is indecomposable, with spectral radius $e^{\rho t_2}$ and corresponding eigenvectors $\underset{\sim}{w}$, $\underset{\sim}{v}'$ which are positive etc. Hence it follows

$$e^{\rho t_2} = \rho(t_1, t_2), \quad \underset{\sim}{d} = \underset{\sim}{v}.$$

Conversely, if $\underset{\sim}{v} = d(0)$, then $d(t) = \underset{\sim}{v}$, $t > 0$.

The interpretation as a limiting conditional distribution follows from

$$(9.10) \qquad \frac{\sum_{i \in T} \pi_i p_{ij}(t)}{\sum_{i \in T} \pi_i (1 - p_{i0}(t))} = \frac{\underset{\sim}{\pi}' Q(t) \underset{\sim}{e}_j}{\underset{\sim}{\pi}' Q(t) \underset{\sim}{e}} = v_j + o(e^{-t(\rho-\rho')}),$$

from (9.5). The rate of convergence is exponential as in

(5.4). The bounds on ρ, analogous to (5.6), are

$$\min_{i\in T}\sum_{j\in T} q_{ij} \le \rho \le \max_{i\in T}\sum_{j\in T} q_{ij} ,$$

with either equality holding iff both do.

9.5 Mandl's Results.

Mandl's paper [23] for the continuous parameter, in particular for T which is not regular, follows the pattern of [22], the results being analogous.

The corresponding subsets T_j of mutually communicating states arise from a communication relation defined in terms of the elements $q_{ij} \in C$, viz. $j \in T$ is a consequent of $i \in T$ if there exists a sequence $i_1, i_2, \ldots, i_s$ of states of T such that $q_{ii_1}\, q_{i_1 i_2} \cdots q_{i_r i_{r+1}} \cdots q_{i_s j}$ are all positive. If i is also a consequent of j then i and j are said to communicate. No assumption is made about all states being able to communicate with themselves. Instead, every state is said to communicate with itself. The usual partial ordering follows as in the discrete parameter case. The new feature is that among the submatrices $Q_{jj} \in C$ corresponding to sets T_j, there may now occur one-dimensional zeros, the others being indecomposable.

Condition A is as before; however there appears to be an additional assumption that the initial probability distribution is non-zero only in T: this seems unnecessary.

The proofs follow a similar pattern to the discrete case, with the matrices $Q_{ij} \in C$, $i \neq j$ playing the role of transition step matrices between the subsets T_i and T_j $i \neq j$, and convolution integrals playing a prominent role.

NOTE. In conclusion to this chapter, we remark that 9.3-9.5, and 9.I (making appropriate adjustments) hold also for regular sets T such that $Q(t) = e^{Ct}$ is stochastic i.e. for regular chains as in the discrete case as regards 9.3 and 9.4, and for decomposable chains, as regards 9.5.

IO. A DENUMERABLE INFINITY OF STATES

For completeness, we present here the few known facts gleaned from the literature for particular special cases, that are relevent to our study in the case of absorbing chains with a countably infinite number of states.

IO.I A Survey.

It is apparent that virtually all the preceding matter is dependent upon the spectral resolution of a finite matrix. In the present case the matrices are infinite, and the same technique is not generally applicable.

Sarymsakov [30], §§ 22-24 attempts to deal with the discrete parameter case by a passage from finite to infinite matrices with the reason that this will afford a computational procedure. His method is to consider a sequence of matrices

$$\{\Phi_m\} \qquad m = 1, 2, \ldots$$

where the matrix Φ_m is composed of the first m rows and columns of the infinite matrix $\mathbf{P}$, taken to correspond to a single essential aperiodic class of states (we use Chung's terminology, [7]). A further assumption is made, to the effect that after a finite m all Φ_m are indecomposable, and then several theorems about 'eigenvalues' and 'cofactors' are proved. However, the theory is far from complete, and even when $\mathbf{Q}$ is of the same form as the above mentioned $\mathbf{P}$, does not enable us to obtain results for e.g. the limiting conditional distribution.

Reuter and Ledermann [28] have developed some spectral theory, for the special case of birth-and-death processes (continuous parameter) along lines which somewhat resemble Sarymsakov's.

Karlin and McGregor [17] have obtained some results for this special case also, and for the discrete (random walk) analogue also, [18]. Their results are applied in IO.3 and IO.4.

Yaglom's paper [33] is important as regards the limiting conditional distribution. A generalization of his basic theorem is given in IO.2.

In the infinite discrete case, we again consider a chain with one absorbing state and a set T of non-essential and therefore non-recurrent states which, for simplicity may again be called transient, [13]. The non-essential states fall into non-essential classes, which may or may not be periodic. Chung [7], p.55 has pointed out that, in general, the limit of an individual ratio

$$\frac{p_{ij}^{(n)}}{p_{kl}^{(n)}}$$

does not exist as $n \to \infty$ for a recurrent aperiodic class, whereas in the corresponding finite case it does. Whether or not this is also true for i, j, k, l, $\in$ T with T a single non-essential aperiodic class, does not appear to be known. A knowledge of what happens in this case, would clearly be helpful in considering the limit of

$$s_{ij}^{(n)} = \frac{p_{ij}^{(n)}}{\sum_{j \in T} p_{ij}^{(n)}} .$$

The remaining point of interest which must be mentioned, is that it only seems relevent to consider the countable case if absorption is certain i.e.

$$\sum_{n=0}^{\infty} \sum_{j \in T} p_{ij}^{(n)} p_{j0} = 1 \quad , \text{all } i \in T$$

whereas, in general, the above sum $\leq$ 1.

IO.2 The Simple Discrete Branching Process.

A simple discrete branching process defines an absorbing Markov chain with a countable number of states, 0, 1, 2,...... representing the number of individuals in the population. (For a sketch of branching processes, and extensive bibliography, see Bharucha-Reid, [4] .) Let the determining probability distribution, which gives the probability that the number of individuals **x** produced in any one generation by any one individual, be **p(x)**, the reproduction taking place independently of other individuals; and the corresponding generating function be

$$F(s) = \sum_{x=0}^{\infty} p(x) s^x \qquad |s| \leq 1$$

Then

$$p_{ij} = \text{coefficient of } s^j \text{ in } [F(s)]^i$$

To have an absorbing chain, we must allow the possibility of the population dying out i.e. we must have $0 < p(0) < 1$, which means that absorption can take place in one step from any state i = 1, 2, 3,......, these being transient.

It is a consequence of the fundamental theorem of branching processes, that if

$$\lim_{s \to 1-} F'(s) = m \leq 1$$

then the probability of absorption (extinction) is unity, for any initial probability distribution $\underset{\sim}{\pi}$, although $\underset{\sim}{\pi}$ is usually taken as $\underset{\sim}{\rho_i}$, since in general we are concerned with starting from a fixed number of individuals, i. Thus, when $m \leq 1$, we have an analogy to the finite case. Under the assumptions that $m < 1$ and $F''(1) < \infty$, Yaglom proves that

$$\lim_{n \to \infty} \frac{p_{ij}^{(n)}}{\sum_{j=1}^{\infty} p_{ij}^{(n)}} = v_j \quad , \left(\sum_{j=1}^{\infty} v_j = 1 \right).$$

He therefore proves the existence of the limiting conditional distribution for a particular $\underset{\sim}{\pi}$. We extend this result, following the lines of his theorem, to obtain the existence of the limit for a large class of $\underset{\sim}{\pi}$, this limit being independent of $\underset{\sim}{\pi}$.

We need another result before proceeding. Let us denote the probability generating function of the number of individuals in the n th generation, starting with one initially, by $F_n(s)$. It is well known that (e.g. [4] p. 19)

$$F_n(s) = F[F_{n-1}(s)], \tag{10.1}$$

and that if the original number of individuals is i, then the corresponding generating function is $[F_n(s)]^i$. Therefore

$$p_{ij}^{(n)} = \text{coefficient of } s^j \text{ in } [F_n(s)]^i. \tag{10.2}$$

THEOREM: Generalized theorem of Yaglom.

If $m<1$, $F''(1-)<\infty$, $0<\sum_{i\in T}\pi_i\leq 1$, and the initial distribution has finite first moment i.e. $\sum_{i=1}^{\infty} i\pi_i<\infty$ then

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{\infty}\pi_i p_{ij}^{(n)}}{\sum_{i=1}^{\infty}\pi_i(1-p_{i0}^{(n)})} = v_j \quad , \quad j\in T$$

where $\sum_{j=1}^{\infty} v_j = 1$, and v_j is independent of $\underset{\sim}{\pi}$.

Proof:

Let

$$G_n(s) = \frac{\sum_{j=1}^{\infty}\pi_i p_{ij}^{(n)} s^j}{\sum_{i=1}^{\infty}\pi_i(1-p_{i0}^{(n)})} \qquad |s|\leq 1$$

$$= \frac{\sum_{i=1}^{\infty}\pi_i(F_n^i(s)-F_n^i(0))}{\sum_{i=1}^{\infty}\pi_i(1-F_n^i(0))}$$

from (10.2), i.e.

(10.3) $$G_m(s) = 1 + \frac{\sum_{i=1}^{\infty} \pi_i \, (F_m^{i}(s) - 1)}{\sum_{i=1}^{\infty} \pi_i \, (1 - F_m^{i}(0))}.$$

We now use an extension of a classical result due to Königs, by Kneser [21] (c.f. Yaglom, [33]). This is concerned with functional iteration in the form of (10.1), and considers a function $f(s)$ defined in a neighbourhood of $s=c$, where c is a fixpoint of the function i.e. $f(c)=c$ and

(10.4) $$|(f(s)-c) - a(s-c)| \leq M|s-c|^{\delta}$$

in the neighbourhood of c, with a, $M>0$, $\delta>1$ constant and $0<|a|<1$. Then, for s sufficiently close to c,

(10.5) $$\xi(s) = \lim_{n\to\infty} [a^{-n}(f_n(s)-c)]$$

is a solution of Schröder's functional equation

$$\xi(f(s)) = a\,\xi(s)$$

and the conditions $\xi(c) = 0$, $\xi'(c) = 1$.

In our case, we know the function F(s) is analytic for $-1 < s < 1$, with $F(1-) = 1$. Considering the left neighbourhood of the point $c = 1$ we need only verify that condition (10.4) is satisfied to be able to apply Kneser's result. In fact from Taylor's theorem with remainder

$$F(s) = F(1-) + F'(1-)(s-1) + O((s-1)^2)$$

(since $F''(1-) < \infty$) for $-1 < s \leq 1$; and $0<m=F'(1-)<1$ as required, m corresponding a in (10.4). Hence, from (10.5)

$$(10.6) \qquad \frac{F_n(s) - 1}{m^n} = L_n(s)$$

has the property that

$$\lim_{n \to \infty} L_n(s) = H(s)$$

which satisfies, in a left neighbourhood of 1, the equation

$$H(F(s)) = m\, H(s)$$

and
$$H(1-) = 0 \;, \quad H'(1-) = 1 \,.$$

Notice that $H(0) < 0$, since $H(s)$ is non-decreasing for $0 \leq s \leq 1$, from (10.6), and is strictly decreasing at $s = 1-$ where it is zero.

Now from (10.6)

$$F_n^{\,i}(s) = (1 + m^n L_n(s))^i$$

$$= 1 + i m^n L_n(s) + \binom{i}{2}(m^n L_n(s))^2 \ldots (m^n L_n(s))^i,$$

since i is a positive integer. Hence

$$\frac{F_n^{\,i}(s) - 1}{i\, m^n} = L_n(s) + O(m^n)$$

$$(10.7) \quad \lim_{n \to \infty} \frac{F_n^{\,i}(s) - 1}{i\, m^n} = H(s)$$

$$(10.8) \therefore \lim_{n \to \infty} \frac{1 - F_n^{\,i}(0)}{i\, m^n} = -H(0) > 0$$

The series

$$\sum_{i=1}^{\infty} \pi_i \left(\frac{F_n^{\,i}(s) - 1}{m^n} \right)$$

is uniformly convergent with respect to n: from (10.7) and (10.8) for $0 \leq s \leq 1$, we have for $K > 0$ and n sufficiently large

$$\left| \frac{F_m^{\,i}(s) - 1}{m^n} \right| \leq i\,[-H(0) + K]$$

and since $\sum_{i=1}^{\infty} i\pi_i$ is finite by assumption, the uniform convergence follows from Weierstrass' M-test.

Therefore, from (10.3), for $0 \leq s \leq 1$

$$\lim_{m \to \infty} G_m(s) = 1 + \frac{\sum_{i=1}^{\infty} i\,\pi_i\,H(s)}{-\sum_{i=1}^{\infty} i\,\pi_i\,H(0)}$$

$$= 1 - \frac{H(s)}{H(0)} .$$

The series $G_n(s)$ is uniformly convergent with respect to n for $0 \leq s < 1$, hence the assertion of the theorem.

It is convenient to express the conditions on H(s) in terms of conditions on G(s), where

$$1 - \frac{H(s)}{H(0)} = G(s) .$$

Denoting -H(0) by B, we have from (10.6) that

$$G(F(s)) = 1 + m\,(G(s) - 1)$$

i.e. $$G(F(s)) = (1 - m) + m\,G(s)$$

with $$G(1-) = 1 + \frac{H(1-)}{B} = 1 \; ; \; G'(1-) = \frac{H'(1-)}{B} = B^{-1}.$$

10.3 Random Walks.

Karlin and McGregor [18] have considered the existence of the limit of

$$\frac{p_{ij}^{(m)}}{p_{kl}^{(m)}} \qquad i, j, k, l \in T$$

in that class of discrete Markov chains termed semi-finite random walks. The transition matrix **P** of such processes is a Jacobi matrix i.e. $p_{ij} = 0$ if $|i - j| > 1$, where the set of states is 0, 1, 2,..... We are concerned with the particular random walk with an absorbing state (or barrier) at 0, and therefore assume

$$\begin{array}{lll} p_{00} = 1, & p_{i,i-1} = q_i > 0 & i \geq 1 \\ p_{i,i+1} = p_i > 0, & p_{i,i} = r_i \geq 0 & i \geq 1 \\ q_i + p_i + r_i = 1 & & i \geq 1. \end{array}$$

(This is in keeping with our previous labelling of states although -1 is taken as the absorbing state in [18].) Put

$$\tau_1 = 1, \qquad \tau_n = \frac{p_1 p_2 \ldots p_{n-1}}{q_2 q_3 \ldots q_n}, \; n \geq 2.$$

We then have the following important results of Karlin and McGregor:

i) If $r_i = 0$ for all $i \in T$ (i.e. $T = \{1,2,3,\ldots\}$ forms a periodic non-essential class, period 2) then

$$\lim_{n\to\infty} \frac{p_{ij}^{(2n)}}{p_{kl}^{(2n)}} = \frac{\tau_j Q_i(\alpha) Q_j(\alpha)}{\tau_l Q_k(\alpha) Q_l(\alpha)} \qquad \text{if} \begin{Bmatrix} i-j \\ k-l \end{Bmatrix} \text{even}$$

$$\lim_{n\to\infty} \frac{p_{ij}^{(2n+1)}}{p_{kl}^{(2n+1)}} = \frac{\tau_j Q_i(\alpha) Q_j(\alpha)}{\tau_l Q_k(\alpha) Q_l(\alpha)} \qquad \text{if} \begin{Bmatrix} i-j \\ k-l \end{Bmatrix} \text{odd}$$

where $\infty > Q_i(\alpha) > 0$ (see [18]); i, j, k, l, $\in$ T.

ii) If all $r_i \geq \delta > 0$, $i \in T$ then

$$\lim_{n\to\infty} \frac{p_{ij}^{(n)}}{p_{kl}^{(n)}} = \frac{\tau_j Q_i(\alpha) Q_j(\alpha)}{\tau_l Q_k(\alpha) Q_l(\alpha)}$$

for i, j, k, l, $\in$ T, and is therefore finite and positive.

Only the latter result will be considered here, this being a case when T is a non-essential aperiodic class and corresponds to a finite set with primitive submatrix (except that absorption into 0 may no longer be certain). Let us now examine the limiting conditional probability

$$\lim_{m\to\infty} \frac{\sum_{i=1}^{\infty} \pi_i p_{ij}^{(m)}}{\sum_{i=1}^{\infty} \pi_i (1-p_{i0}^{(m)})} .$$

First take $\underset{\sim}{\pi} = \underset{\sim}{f}^i$: then the above becomes

$$\begin{aligned}
\lim_{m\to\infty} s_{ij}^{(m)} &= \lim_{m\to\infty} \frac{p_{ij}^{(m)}}{1-p_{i0}^{(m)}} \qquad i,j \in T \\
&= \lim_{m\to\infty} \frac{1}{\sum_{k=1}^{\infty} \frac{p_{ik}^{(m)}}{p_{ij}^{(m)}}} \\
&= \frac{1}{\sum_{k=1}^{\infty} \frac{\tau_k Q_k(\alpha)}{\tau_j Q_j(\alpha)}} \\
&= \frac{\tau_j Q_j(\alpha)}{\sum_{k=1}^{\infty} \tau_k Q_k(\alpha)}
\end{aligned}$$

providing it is permissible to interchange the limiting and summation procedures. Notice that if so, the limit is independent of i. In the general case

$$\begin{aligned}
\lim_{m\to\infty} \frac{\sum_{i=1}^{\infty} \pi_i p_{ij}^{(m)}}{\sum_{i=1}^{\infty} \pi_i (1-p_{i0}^{(m)})} &= \lim_{m\to\infty} \frac{\sum_{i=1}^{\infty} \pi_i p_{ij}^{(m)}}{\sum_{i=1}^{\infty} \pi_i \sum_{k=1}^{\infty} p_{ik}^{(m)}} \\
&= \frac{(\sum_{i=1}^{\infty} \pi_i Q_i(\alpha))\, \tau_j Q_j(\alpha)}{(\sum_{k=1}^{\infty} \tau_k Q_k(\alpha))(\sum_{i=1}^{\infty} \pi_i Q_i(\alpha))} \\
&= \frac{\tau_j Q_j(\alpha)}{\sum_{k=1}^{\infty} \tau_k Q_k(\alpha)}
\end{aligned}$$

independent of $\underset{\sim}{\pi}$, provided all the operations carried out to this point are valid (e.g. $\sum_{i=1}^{\infty} \pi_i Q_i(\alpha) < \infty$).

The validity of such operations has not been investigated, but it seems clear that some restriction on $\underset{\sim}{\pi}$ is necessary (c.f. 10.2). Moreover, it is conjectured that for these operations to be valid, absorption with probability 1 is necessary.

Finally, let us remark that everything discussed here holds true, from [18], when

$$q_i + p_i + r_i \leq 1 \quad , \quad i \geq 1$$

in which case the matrix Q corresponding to states 1, 2,. ... is a Jacobi matrix, although P is not. This inequality therefore gives a slight extension of the usual random walk theory.

Thus we have at least some indication of the existence of a quasi-stationary distribution independent of π, in this particular case at least.

IO.4 The Birth-and-death Process.

In continuous time, the process analogous to a random walk (IO.3, q.v.) is the birth-and-death process, which has also been investigated by Karlin and McGregor 17 . The results of their two papers [17] and [18] are analogous, apart from the fact that periodicity cannot occur when the parameter is continuous.

The birth-and-death process of interest, is the one with states 0, 1, 2,......with 0 absorbing, as before, is specified by the following scheme:

$$\left.\begin{aligned} p_{i,i-1}(t) &= q_i t + o(t) \\ p_{i\,i+1}(t) &= p_i t + o(t) \\ p_{ii}(t) &= 1 - (p_i + q_i)t + o(t) \end{aligned}\right\}$$

as $t \to 0$, where p_i , $q_i > 0$. Then for τ_n, $n \geq 1$ defined as in IO.3, we have the result corresponding to ii) of that section:

$$\lim_{t\to\infty} \frac{p_{ij}(t)}{p_{kl}(t)} = \frac{\tau_j\, Q_i(\alpha)\, Q_j(\alpha)}{\tau_l\, Q_k(\alpha)\, Q_l(\alpha)}, \quad i, j, k, l \in T.$$

(The reader is referred to [17] for the remaining symbols.) The remaining discussion is analogous to IO.3.

APPENDIX

A.I Some Remarks on Generating Functions for Finite Discrete Markov Chains.

According to Romanovskiĭ [29], pp 35-36, generating functions for finite discrete chains were used by Markov himself to prove his limit theorems; subsequently, much use has been made of them as regards frequency counts (e.g. Bhat, [5] : Good, [16] ; Neuts, [27].), and other purposes.

It is noticeable that whenever this method has been used in this thesis i.e. 1.3, 3.3, and 4.1, the eigen-values and eigenvectors of the transient state transition submatrix Q, have figured prominently. A good understanding of this relationship may be obtained by studying Romanovskiĭ's chapter on characteristic functions; we shall only briefly indicate the basis of the connection. The notation is that of Chapter 1 of this thesis.

Suppose we consider a more general case than previously i.e. the problem of obtaining the joint p.g.f. of $X_{\underset{\sim}{\pi},1}$, $X_{\underset{\sim}{\pi},2}$, ..., $X_{\underset{\sim}{\pi},s}$, with time to absorption fixed viz. $\sum_{i\in T} X_{\underset{\sim}{\pi},i} = X_{\underset{\sim}{\pi}} = n$. It then follows, by a method analogous to that of 1.3, that this is

$$g_n(\omega_1, \omega_2, \ldots, \omega_s) = \underset{\sim}{\pi}'(\omega_1, \omega_2, \ldots, \omega_s) Q^{n-1}(\omega_1, \ldots, \omega_s)[I-Q]\underset{\sim}{e}$$

where

$$Q(\omega_1, \omega_2, \ldots, \omega_s) = \begin{bmatrix} p_{11}\omega_1 & p_{12}\omega_2 & \cdots & p_{1s}\omega_s \\ p_{21}\omega_1 & p_{22}\omega_2 & \cdots & p_{2s}\omega_s \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ p_{s1}\omega_1 & p_{s2}\omega_2 & \cdots & p_{ss}\omega_s \end{bmatrix}, \quad \underset{\sim}{\pi}(\omega_1, \ldots, \omega_s) = \begin{bmatrix} \pi_1\omega_1 \\ \pi_2\omega_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \pi_s\omega_s \end{bmatrix}$$

or, as in 1.3

$$g_n(\omega_1, \omega_2, \ldots, \omega_s) = \underset{\sim}{\pi}'D\,[QD]^{n-1}[I-Q]\underset{\sim}{e}$$

where $D = \mathrm{diag}\,[\omega_1, \omega_2, \ldots, \omega_s]$.

According to the Cayley-Hamilton theorem QD satisfies the equation

$$|\lambda I - QD| = 0. \tag{A.1}$$

Now, it is certainly true that $g_n = g_n(\omega_1, \omega_2, \ldots \omega_s)$ satisfies the difference equation

$$|EI - QD|g_n = 0, \tag{A.2}$$

from (A.1), where E is the shift operator in the finite difference calculus (see also Bartlett, [1] p.32). The characteristic equation of (A.2) is (A.1) and its characteristic values are the eigenvalues of QD (depending on $\omega_1, \ldots\ldots, \omega_s$, but reducing to those of Q when the parameters are equated to unity).

It is now clear that the p.g.f. approach is tantamount to a spectral resolution procedure, explicitly or implicitly. Extensive discussion of the behaviour of the generating function in the vicinity of $(\omega_1, \ldots, \omega_s) = (1, 1, \ldots, 1)$ is given by Romanovskiĭ [29], although he discusses characteristic functions rather than probability generating ones.

A.2 The Problem for Diffusion Processes.

Apart from his papers [22] and [23], Mandl has been concerned with the analogous problem for diffusion processes bounded on one side by a reflecting, absorbing or elastic barrier, the results of his investigations being contained in [24], [25] and [26] (c.f. Ewens, [10], [11], [12]). We give his statement for the case of absorbing and reflecting barriers, from [25]:

"...to find the conditions for the covergence to a limit distribution of the probability distribution of the position of the particle at time t, under the condition that the particle was not absorbed before the time t."

However, it is not within the scope of this thesis to study diffusion processes, therefore we do not pursue it further, but hope to take up the study in the future.

Conclusion.

In this thesis, we have developed some ideas relevent to the description of transient behaviour in absorbing Markov chains. In particular, this has been done for discrete finite chains, and extended to finite continuous chains. The material lent itself easily to the concept of a generalized finite Markov chain, and several analogous results to the ordinary Markov chain theory were obtained.

There is obviously a great deal yet to be done, particularly when the number of states is denumerably infinite. Clearly, more sophisticated tools need to be used than are to be found in the present thesis. Moreover, a vast new field for these ideas is contained in the theory of diffusion processes, in which several papers have already appeared. In particular, the interpretation of the formal solution of the stationary equation, when no stationary distribution can exist, has caused considerable controversy.

It is hoped that this thesis may be a small beginning to the study of the ideas outlined in it.

Acknowledgement.

The author wishes to acknowledge the help, guidance and patience of his supervisor, Dr.J.N. Darroch, who originated the idea of studying quasi-stationary distributions, and suggested several of the possibilities contained here, in particular the limiting conditional probability.

Bibliography

[I] Bartlett, M.S.(I955)An Introduction to Stochastic Processes. Cambridge University Press, Cambridge.

[2] ———— (I960)Stochastic Population Models. Methuen, London.

[3] ———— (I963)Private Communication(to Dr. J.N.Darroch).

[4] Bharucha-Reid, A.T.(I960)Elements of the Theory of Markov Processes and their Applications. McGraw-Hill, New York.

[5] Bhat, B.R. (I96I)Some Properties of regular Markov chains. Ann.Math.Statist., 32, 59-7I.

[6] Breny, H. (I962)Cheminements conditionnels de chaînes de Markov absorbantes. Ann.Soc. Sci.Bruxelles, 76, 8I-87.

[7] Chung, K.L. (I960)Markov Chains with Stationary Transition Probabilities. Springer, Berlin.

[8] Debreu, G. and Herstein, I.N.(I953)Non-negative square matrices. Econometrica, 2I, 597-607.

[9] Doob, J.L. (I953)Stochastic Processes. Wiley, New York.

[IO] Ewens, W.J. (I963)The diffusion equation and a pseudo-distribution in genetics. J.Roy.Statist. Soc. Ser.B., 25, 405-4I2.

[II] ———— (I963)The mean time to absorption in a process genetic type. J.Aust.Math.Soc., 3, 375-383.

[I2] ———— (I964)The pseudo-transient distribution and its uses in genetics. J.Appl.Prob., I, I4I-I56.

[I3] Feller, W. (I957)Introduction to Probability Theory and its Applications, Vol.I. Wiley, New York.

[I4] Fréchet, M. (I938) Recherches modernes sur la théorie des probabilités. Traité du Calcul des Probabilités (ed. E. Borel), Vol. I, No. 3. Gauthier-Villars, Paris.

[I5] Gantmacher, F.R. (I959) Applications of the Theory of Matrices. Interscience, New York.

[I6] Good, I.J. (I96I) The frequency count of a Markov chain and transition to continuous time. Ann. Math. Statist., 32, 4I-48.

[I7] Karlin, S. and McGregor, J.L. (I957) The classification of birth and death processes. Trans. Amer. Math. Soc., 86, 366-400.

[I8] ——— (I959) Random Walks. Illinois J. Math., 3, 66-8I.

[I9] Kemeny, J.G. and Snell, J.L. (I960) Finite Markov Chains. Van Nostrand, Princeton, N.J.

[20] ——— (I96I) Finite continuous time Markov chains. Teor. Veroyatnost. i Primenen, 6, II0-II5.

[2I] Kneser, H. (I950) Reele analytische Lösungen der Gleichung $\phi\{\phi(x)\}=e^x$ und verwandte Funktionalgleichungen. J. Reine Angew. Math., I87, 56-67.

[22] Mandl, P. (I959) On the asymptotic behaviour of the probabilities in groups of states of a homogeneous Markov chain. (in Russian.) Časopis Pěst. Mat., 84, I40-I49.

[23] ——— (I960) On the asymptotic behaviour of the probabilities within groups of states of a homogeneous Markov process. (in Russian.) Časopis Pěst. Mat., 85, 448-456.

[24] ——— (I960) Les propriétés limites des répartitions de probabilité des processus de Markov bornés. Transactions of the Second Prague Conference on Information Theory. Prague.

[25] Mandl, P. (1961) Spectral theory of semigroups connected with diffusion processes and its application. Czechoslovak Math.J., 11, 226-236.

[26] ——— (1962) On the distribution of time until one boundary is reached in the process of Uhlenbeck. (in Czech.) Aplikace Mat., 7, 141-148.

[27] Neuts, M.F. (1964) General transition probabilities for finite Markov chains. Proc.Camb. Phil.Soc., 60, 83-91.

[28] Reuter, G.E.H. and Ledermann, W. (1954) Spectral theory ~~for~~ for differential equations of simple birth and death processes. Phil.Trans.Roy.Soc. London, 246, 321-369.

[29] Romanovskiĭ, V.I. (1949) Discrete Markov Chains. (in Russian.) Gosudarstv.Izdat.Tehn.-Teor.Lit., Moscow.

[30] Sarymsakov, T.A. (1954) Elements of the Theory of Markov Processes. (in Russian.) Gosudarstv.Izdat. Tehn.-Teor.Lit., Moscow.

[31] Sinkhorn, R. (1964) A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann.Math.Statist., 35, 876-879.

[32] Watterson, G.A. (1961) Markov chains with absorbing states: a genetic example. Ann.Math.Statist., 32, 716-729.

[33] Yaglom, A.M. (1947) Certain limit theorems of the theory of branching stochastic processes. (in Russian.) Doklady Akad.Nauk S.S.S.R., n.s. 56, 795-798.