



Australian
National
University

THESES SIS/LIBRARY
R.G. MENZIES LIBRARY BUILDING NO:2
THE AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200 AUSTRALIA

TELEPHONE: +61 2 6125 4631
FACSIMILE: +61 2 6125 4063
EMAIL: library.theses@anu.edu.au

USE OF THESES

This copy is supplied for purposes
of private study and research only.
Passages from the thesis may not be
copied or closely paraphrased without the
written consent of the author.

SOME PROBLEMS
IN THE THEORY AND APPLICATIONS OF
MARKOV CHAINS

by
Joseph Gani

A thesis submitted to the
Australian National University for
the degree of Doctor of Philosophy
in the Department of
Statistics

Canberra

1955

CONTENTS

	Page
PREFACE	viii
SUMMARY	xi

Part I

PROBLEMS IN THE THEORY OF
MARKOV CHAINS

Chapter 1

STATISTICAL PROBLEMS IN THE THEORY OF
MARKOV CHAINS

1. Introduction	2
2. Statistical problems in the theory of stochastic processes	3
3. Tests of statistical hypotheses based on a Markov chain	8
4. Sequential estimation as a case of estimation in a Markov chain	12

Chapter 2

SOME PROPERTIES OF SIMPLE
MARKOV CHAINS

I. Basic Definitions

- | | |
|------------------------------------|----|
| 1. Transition probabilities | 15 |
| 2. Higher transition probabilities | 17 |

II. Regularity and Positive Regularity

- | | |
|---|----|
| 3. Stationary probabilities | 19 |
| 4. Necessary and sufficient conditions for
regularity and positive regularity | 22 |
| 5. Higher transition probabilities and the
latent roots of a stochastic matrix | 28 |
| 6. Evaluation of the stationary probabilities | 33 |

III. Transition Frequencies and their

Moment Generating Function

- | | |
|---|----|
| 7. Transition frequencies | 35 |
| 8. Moment generating function of the transition
frequencies | 42 |
| 9. Latent roots of the matrix $\underline{R}' = \{p_{ij} \exp t_{ij}\}$
for a positively regular chain | 44 |

Chapter 3

OPTIMUM PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF A PARAMETER IN A MARKOV CHAIN

- | | |
|--|----|
| 1. The likelihood equation for the realisation
of a chain | 48 |
|--|----|

	iv
2. Optimum properties of the estimator	50
3. Consistency of the estimator	51
4. Uniqueness of the consistent estimator	54
5. Asymptotic normality of the estimator	56

Chapter 4

SUFFICIENCY CONDITIONS FOR A MARKOV CHAIN WITH A FIXED INITIAL STATE

1. Introduction	62
2. The form of the likelihood function admitting a sufficient estimator, for discrete parent distributions with constant variate intervals	63
3. Form of the probabilities $p_i(\theta)$ for a multinomial distribution admitting a sufficient estimator T of θ	67
4. Form of the transition probabilities $p_{ij}(\theta)$ for a simple positively regular Markov chain admitting a sufficient estimator T of θ	71

Chapter 5

SUFFICIENCY CONDITIONS FOR A MARKOV CHAIN WITH AN UNRESTRICTED INITIAL STATE

I. Introduction

1. The general problem 83
2. The Markov chain with two states 86

II. The Markov Chain with s States

3. The general form of the stationary probabilities 89
4. Stochastic matrices with stationary probabilities $P_i = \alpha_i e^{k_i \lambda_1(\theta) + \lambda_2(\theta)}$ 93
5. Stochastic matrices with constant stationary probabilities 98

III. The Markov Chain with Three States

6. Stochastic matrices with stationary probabilities $P_i = \alpha_i e^{k_i \lambda_1(\theta) + \lambda_2(\theta)}$ 102
7. Stochastic matrices with constant stationary probabilities 107

Part II

APPLICATIONS OF MARKOV CHAINS TO PROBLEMS OF STORAGE

Chapter 6

PROBLEMS OF STORAGE

I. Some Theory of Storage

	vi
1. Introduction	111
2. Pitt's problems of provisioning	115
3. Moran's problems in the theory of dams	118
4. Provisioning and dam problems as general storage problems	120
 <u>II. Interpretation of Pitt's Results</u> <u>for Dam Problems</u>	
5. Problems of the infinite dam	126
6. Stationary distribution of $Z(t)$ for release rule (6.5.2)	128
7. Stationary distribution of $Z(t)$ for release rule (6.5.3)	131
 <u>III. Application of Moran's Methods to</u> <u>Provisioning Problems</u>	
8. Provisioning with a finite positive stock	133
9. Provisioning with replacements at fixed times without time-lag	135
10. Provisioning with replacements at fixed times $t = kMa^{-1} + T$ ($k = 1, 2, \dots$), with time-lag $T = Ma^{-1}$	138
11. Provisioning with orders at fixed times $t = kMa^{-1}$, and replacements at fixed times $t = (k+1)Ma^{-1}$ ($k = 1, 2, \dots$) with time-lag Ma^{-1}	150

Chapter 7

AN EXACT SOLUTION OF THE STORAGE PROBLEM WITH
A POISSON INPUT

1. Introduction	156
2. The discrete model	158
3. The continuous analogue	165

Chapter 8

THE SOLUTION OF A STORAGE PROBLEM
BY MONTE CARLO METHODS

1. Introduction	171
2. Estimators of the stationary probabilities and their variances	173
3. Application of Monte Carlo methods to a storage problem	176
4. Extension to the continuous case	182
REFERENCES	184

PREFACE

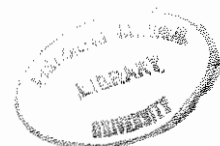
I

This thesis, written during my two-year term between January 1954 and January 1956 as a Research Student of the Australian National University, consists of some work carried out during 1954 and early 1955. Sections of Chapter 2, and all of Chapters 3, 4, 6, 7, 8, are either published or in process of publication, and are available in a slightly different form in *Biometrika* (1955, 42), and the *Australian Journal of Applied Science* (1955, 6); Chapter 5 is also being submitted for publication.

II

I feel it is safe to claim the greater part of the thesis as original work; but, in the circumstances, it would perhaps be more appropriate to specify the original parts of each chapter in some detail.

Chapter 1 is a review of recent contributions to the field considered; to this no great originality can



be ascribed. Chapter 2 consists of a connected account of those properties of Markov chains, most of them well known, which are required in the remaining parts of the thesis. My own contributions to this chapter are:

- 1) the somewhat new presentation of the proof in § 2.5, p. 30-33, that the latent root 1 of the stochastic matrix for a regular chain is simple;
- 2) the new theorem in § 2.7, p. 39-42, derived from a theorem of Fréchet, that $\lim_{n \rightarrow \infty} \sigma_{ij}^2 n^{-1} \neq 0$;
- 3) the entirely new theorem in § 2.9, p. 44-47, for the latent roots of the matrix $\underline{R}' = \{p_{ij} \exp t_{ij}\}$.

The work in Chapters 3, 4, 5, 6, 7, though based in part on suggestions of Professor P.A.P. Moran, is fully my own. Naturally, some results due to other authors are used, or briefly summarised, but these are always clearly acknowledged.

Finally, Chapter 8 was prepared in collaboration with Professor Moran; each of us can, I think, claim an equal share of the work. I began by working out the discrete dam problem by Monte Carlo methods; Professor Moran redrafted the entire chapter, eliminated several errors and added various improvements. The final part § 8.4 on the continuous dam problem is entirely his,

and has only been included for the sake of completeness.

III

I should like to take this opportunity of thanking Professor A.L. Blakers and the University of Western Australia for granting me leave of absence to allow me to continue my post-graduate studies; also, the Australian National University without whose financial support and excellent library facilities, the research could not have been conducted.

It is to Professor Moran, however, that I feel the greatest debt of gratitude. He suggested several of the problems which are considered in this thesis, discussed my difficulties with me, and criticised my work at all stages in its development. His warm interest, particularly at times when progress was slow, proved the greatest stimulus of all. I shall remain permanently indebted to him for convincing me that research could be a deeply satisfying activity.

Department of Statistics,
Australian National University,
Canberra, A.C.T.,
July 1955.


J. Gani

SUMMARY

SOME PROBLEMS
IN THE THEORY AND APPLICATIONS OF
MARKOV CHAINSPart I
PROBLEMS IN THE THEORY OF
MARKOV CHAINS

The first part of this thesis is concerned with some theory of estimation of an unknown parameter θ defining the transition probabilities $p_{ij}(\theta)$ of a positively regular Markov chain.

Chapter 1 begins with a review of existing work on statistical problems in the theory of Markov chains and other stochastic processes. A brief account is given of some contributions to the problem of estimation of an unknown parameter in certain stochastic processes. It is pointed out that no systematic study of problems in statistical inference exists in the theory of Markov chains, although some work has been done on frequency goodness of fit tests of statistical hypotheses based

on chain models. Sequential estimation may also be regarded as a case of estimation in a Markov chain with absorbing states, but this, however, will not be fitted into a comprehensive theory, since our own work is concerned only with estimation theory in positively regular chains.

In Chapter 2, stochastic matrices $\underline{p}$ and $\underline{p}^n$ of transition and higher transition probabilities for Markov chains are defined; in particular, regular and positively regular chains are considered, and the necessary and sufficient conditions satisfied by their stochastic matrices are obtained. The theorem is proved that the stochastic matrix $\underline{p}$ of a regular chain has a latent root 1 which is simple. Finally, transition frequencies are defined, and some of their properties discussed; their moment generating function is derived, and a theorem concerning the latent roots of a matrix $\underline{R}' = \{p_{ij} \exp t_{ij}\}$ in this function is proved.

Chapter 3 consists of the proofs of the usual theorems for the optimum properties of the maximum likelihood estimator of an unknown parameter θ defining the transition probabilities $p_{ij}(\theta)$ of a simple positively regular Markov chain. The likelihood equation is shown to have a consistent solution T ,

and this consistent estimator of θ is proved to be unique and asymptotically normally distributed.

In Chapter 4, we proceed to establish the form of the transition probabilities $p_{ij}(\theta)$ which admit a sufficient estimator of θ when the initial state of a realisation of the chain is fixed. To do this, the form of the likelihood function admitting a sufficient estimator when the parent distribution is discrete is first derived; this is used to obtain the form of the probabilities $p_i(\theta)$ for a multinomial distribution admitting a sufficient estimator of θ , and the result is finally generalised for the transition probabilities $p_{ij}(\theta)$ of the positively regular Markov chain with s states. It is proved that for conditional sufficiency, the transition probabilities in any row i of the stochastic matrix are of the form

$$p_{ij}(\theta) = \alpha_{ij} e^{K_{ij} \Lambda_1(\theta) + \lambda_2(\theta)} \quad (j = 1, \dots, s),$$

with $\exp - \lambda_2(\theta) = \sum_j \alpha_{ij} \exp K_{ij} \Lambda_1(\theta)$, and the number of distinct K_{ij} is less than or equal to the number of states s . Some examples for Markov chains with two and three states are given as illustrations.

Chapter 5 is concerned with conditions for the

sufficiency of the parameter θ defining the non-zero transition probabilities $p_{ij}(\theta)$ of an initially stationary, positively regular Markov chain, when the initial state in a realisation of the chain is unrestricted. The general problem presents serious difficulties; we examine the more limited problem of obtaining from among the matrices $\{p_{ij}\}$ with non-zero elements, which admit a sufficient estimator of θ when the initial state is fixed, those particular ones which do so equally when the initial state is unrestricted. First the Markov chain with two states is considered; then some general results, which are possibly not exhaustive, are obtained for the Markov chain with s states. For this chain, two forms of stationary probabilities such that a sufficient estimator of θ can be obtained for an unrestricted initial state are examined: the first where all

$$P_i(\theta) = \alpha_i e^{K_i \lambda_1(\theta) + \lambda_2(\theta)} \quad (i = 1, \dots, s)$$

are similar to the $p_{ij}(\theta)$, and the second where all the stationary probabilities P_i are constants. In the first case, it is proved that if the K_i are all distinct, the states of the chain must be independent; if not, the states may, but need not, be independent. In the second case, two simple solutions are given: the

doubly stochastic matrix, and another derived from it. The chapter ends with a review of the case of the Markov chain with three states, for which these results prove to be exhaustive.

Part II

APPLICATIONS OF MARKOV CHAINS TO PROBLEMS OF STORAGE

The second part of this thesis examines problems of storage, for some of which, solutions both numerical and exact are obtained.

Chapter 6 begins with a short review of work done on inventory, provisioning and dam storage problems; a more detailed outline is given of two problems in the theory of provisioning with a discrete stock considered by Pitt (1946), and of Moran's work (1954,1955) in the theory of finite dams. It is pointed out that these provide two different methods of attack, each appropriate to certain conditions, on problems in the probability theory of a general storage function $S(t)$ defined at time t by

$$S(t) = I(t) - D(t) - F(t),$$

where $I(t)$, $D(t)$, $F(t)$ are respectively an input, output, and overflow function. The storage function is identified with the stock deficit in provisioning theory, or the dam content in dam theory, so that any problem and its solution in the one theory has an exact analogue in the other. Pitt's results of the theory of provisioning are used in two analogous cases of the infinite discrete dam. These are followed by the application of Moran's methods of the theory of dams in some analogous problems of provisioning with a discrete finite stock; exact solutions are obtained for the discrete and continuous cases of a particular problem in which ordering and replacement times coincide.

In Chapter 7, an exact solution for a storage problem with a Poisson input is obtained. First a discrete model is constructed; in this, a finite discrete storage function $S(t)$, fed by a discrete input function of Poisson type, has a small prescribed discrete output at fixed time intervals $n \delta t$ ($n=0,1,\dots$) when $S(t) \neq 0$, a zero output if $S(t) = 0$, and an overflow function such that $S(t)$ never exceeds a certain maximum value. The set of linear equations arising from this model is solved. By allowing $\delta t \rightarrow 0$,

we obtain the solution for the continuous analogue, where a finite continuous storage function $S(t)$, fed by a discrete input function of Poisson type has a continuous output with a steady rate when $S(t) \neq 0$, a zero output if $S(t) = 0$, and an overflow function not permitting $S(t)$ to exceed a given maximum value.

In Chapter 8, Monte Carlo methods of estimating the elements of an eigen-vector satisfying certain equations in the theory of Markov chains are considered; the eigen-vector, suitably scaled, is the vector of stationary probabilities associated with the stochastic matrix of a positively regular chain. The variances of the estimators of these stationary probabilities are given, and a way of estimating them discussed. This is applied to a particular problem of storage, a case of Moran's dam equations; some numerical results are obtained leading to an estimate of the number of trials required to reach a prescribed accuracy for the elements of the vector. The conclusion is reached that, in general, a numerical evaluation of the eigen-vector is to be preferred to an evaluation by Monte Carlo methods; the Monte Carlo method may, however, offer advantages when applied directly to the continuous problem rather than to Moran's matrix equations, which are an approximation to it.

Part I

PROBLEMS IN THE THEORY OF
MARKOV CHAINS

Chapter 1

STATISTICAL PROBLEMS IN THE THEORY OF MARKOV CHAINS

1. Introduction

Since Markov's original work (1912) initiating the systematic study of those sequences of dependent events now bearing his name, there have been numerous contributions to the probability theory of Markov chains. The most important of these are due, for finite chains to Doeblin, Feller, Fortet, Fréchet, Hostinsky, Mihoc, Onicescu, Romanovsky, and for denumerable infinite chains to Doob, Foster, Kolmogorov, Yosida and Kakutani. These, and other works, are listed in the bibliographies given by Hostinsky (1931) in his memoir, and Fréchet (1952) in his comprehensive treatise where the essentials of the theory of finite simple Markov chains, including many of his own results are collected. A brief account of those parts of the

theory which are specifically required in connection with the problems we are to consider, is given in Chapter 2 of this thesis.

It is not, however, until recently that statistical problems in the theory of Markov chains, such as the testing of hypotheses based on Markov chain models, or the estimation of an unknown parameter defining a chain, have been broached. In this aspect, Markov chain theory appears to have lagged behind the study of statistical problems of the same type in the field of other stochastic processes.

2. Statistical problems in the theory of stochastic processes

An early attack on the problem of estimating an unknown parameter of a discrete stochastic process was made by Wald (1948). In his paper, Wald drew attention to the fact that although the asymptotic properties of maximum likelihood estimators had been studied for independent observations, the case of stochastically dependent observations had not until then been considered. Assuming certain general restrictions on the joint probability distribution of the observations, Wald proves that if θ is the parameter



to be estimated, then

- a) the maximum likelihood equation has a root which is a consistent estimator of θ ;
- b) any consistent root of this equation is asymptotically efficient at least in the wide sense, if not necessarily in the strict sense, that is with a limiting normal distribution.

The discrete stochastic process which he considers, includes the Markov chain, so that in some ways, the results of our Chapter 3 will already have been obtained more generally; it remains of interest, however, to consider in detail some aspects of estimation theory specific to Markov chains, such as for example the asymptotic efficiency in the strict sense of any consistent root of the maximum likelihood equation for a parameter θ defining a positively regular Markov chain.

In the case of time-dependent stochastic processes, the problem of estimation was first raised by Kendall in his paper on "Stochastic processes and population growth" (1949) in which he considers the estimator of a parameter in an evolutive process, the Furry simple birth process. Here, a system is determined at time t by a non-negative integer $N(t)$; if at time t the

probabilities that this be $n-1$, and n are respectively $p_{n-1}(t)$ and $p_n(t)$, then the probability that it be n at time $t + dt$ is given for $n > 1$ by

$$p_n(t+dt) = p_{n-1}(t)(n-1)\lambda dt + p_n(t)(1-n\lambda dt) + o(dt),$$

and for $n = 1$ by

$$p_1(t+dt) = p_1(t)(1-\lambda dt) + o(dt).$$

Fixing a time-period T , and taking all realisations of the process lasting this time, Kendall obtains as the estimator $\hat{\lambda}$ of the parameter λ , for a known initial value $N(0)$,

$$\hat{\lambda} = \{N(T) - N(0)\} / \int_0^T N(t) dt,$$

with sampling variance

$$v(\hat{\lambda}) = \lambda^2 / \{N(0)(e^{\lambda T} - 1)\}.$$

Moran (1951) considering the same problem, obtains for the set of realisations in which the epoch $N(T) - N(0)$ of the system is fixed, but T may vary, the same value for the estimator $\hat{\lambda}$, with a different variance

$$v(\hat{\lambda}) = \lambda^2 / \{N(T) - N(0)\}.$$

This is to be expected, since the variances refer to different hypothetical populations of processes. It is shown, however, that under certain conditions the two variances converge in probability to the same value. The paper continues with an extension of this method to the estimation of the sum of two parameters $(\lambda + \mu)$ in the simplest birth and death process, where the state of the system is again defined at time t by a non-negative integer $N(t)$. In this case, if at time t the probabilities that this be $n-1$, n , and $n+1$ are respectively $p_{n-1}(t)$, $p_n(t)$, and $p_{n+1}(t)$, then the probability that at time $t + dt$ it be n is given for $n \geq 1$ by

$$p_n(t+dt) = p_{n-1}(t)(n-1)\lambda dt + p_{n+1}(t)(n+1)\mu dt + p_n(t)\{1 - n(\lambda + \mu)dt\} + o(dt),$$

and for $n=0$ by

$$p_0(t+dt) = p_1(t)\mu dt + p_0(t)\{1 - \lambda dt\} + o(dt).$$

The problem is considered again by Moran (1953), and $\lambda(\lambda + \mu)^{-1}$ estimated by a method which is in fact one of sequential analysis, and for which we can give a Markov chain model. Writing $p = \lambda(\lambda + \mu)^{-1}$, and $q = \mu(\lambda + \mu)^{-1}$, the process is factorised so that it can be represented by a random walk with an infinite

ERRATA

P.7 : In line 1, and also in the matrix considered, the states E_{N+s-1} , E_{N+s} should read E_{N-1} , E_N respectively.

P.7 : Line 7 should read

$$y = -(s-1), \dots, 0, \dots, N.$$

P.7 : The second part of equation (1.2.1) should read

$$\hat{p}_2 = (N+y)/2N.$$

This estimation problem, since it is effectively that of estimating a parameter in a regular Markov chain with an infinite number of states, one of which is an absorbing state, is of interest to us. It has not yet, however, been possible to extend our theory of estimation beyond that for positively regular chains, so that Moran's contribution has yet to be reconsidered in a projected extension of the work which follows.

Further details on the theory of testing statistical hypotheses, and of estimation in continuous parameter stochastic processes can be found in Grenander's fundamental paper (1950), the first work containing a systematic treatment of problems in statistical inference for these processes.

3. Tests of statistical hypotheses based on a Markov chain

The earliest mention of general statistical problems connected with Markov chains appears in a paper by Romanovsky (1938). In it, an attempt is made to test the hypothesis that a set of events $E_1, \dots, E_s$, forms a simple Markov chain; this is done by testing for the goodness of fit of the transition frequencies n_{ij} from events E_i to E_j , $i, j = 1, 2, \dots, s$, the value of χ^2

being given in the form

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - m_{ij})^2}{m_{ij}},$$

where $m_{1j} = \xi(n_{1j})$ is either known or easily estimated. A similar test is also given to determine whether a chain is simple or multiply dependent. The results quoted are, however, incorrect; the values of χ^2 used by Romanovsky are suitable only for a set of independent events, so that his contribution consists of little more than the statement of problems to which he had intended a solution.

A second reference to a similar test, which is in fact Romanovsky's test of the hypothesis that the events $E_0, E_1, \dots, E_9$ associated with the digits 0, 1, ..., 9 form a Markov chain, is to be found in Kendall and Babington Smith's work (1938, 1939) on the randomness of sampling numbers. In their "serial test" that no digit tends to be followed by any other digit, they suggest a goodness of fit test based on the criterion

$$\psi^2 = \sum_{ij} \frac{(n_{ij} - m)^2}{m},$$

where $m = \sum_{ij} n_{1j}/100$, and ψ^2 is stated to be

asymptotically distributed as χ^2 on 90 degrees of freedom owing to the existence of 10 linear constraints on the n_{ij} . Again this result is incorrect for the reason already stated.

The correct form of χ^2 in the goodness of fit test for simple Markov chains was obtained by Bartlett (1951) as

$$\chi^2 = (\underline{n} - \underline{m})' \underline{V}^{-1} (\underline{n} - \underline{m}) ,$$

where $\underline{n}$, $\underline{m}$ are the column vectors of elements n_{ij} and $m_{ij} = E(n_{ij})$, and where $\underline{V}$ is the variance covariance matrix of the n_{ij} . His work contains a detailed discussion of χ^2 as the asymptotic distribution of the likelihood criterion λ , and of the correct value for its degrees of freedom in the case where the events form a positively regular Markov chain, not only simple but also multiply dependent. The fact that Kendall and Bahington Smith's assertion that χ^2 has a χ^2 distribution asymptotically as $\sum_{i,j} n_{ij} \rightarrow \infty$, is an error, is implicit in Bartlett's work; in spite of this, Bartlett mistakenly agrees with them in one brief paragraph of his paper.

It was finally Good (1953), in his paper on tests of randomness for sampling numbers, who cleared

up the misconceptions on the distribution of χ^2 , and proceeded to prove some general theorems for the equivalent criterion

$$\chi^2 = \sum_{i_1, \dots, i_\nu} \frac{(n_{i_1, \dots, i_\nu} - m)^2}{m},$$

in the case of ν -dependent chains to test the hypothesis of randomness in sequences of random numbers for which $m = \mathcal{E}(n_{i_1, \dots, i_\nu})$.

Basing his work on Bartlett's and Good's results, Patankar (1954) discusses the application of these to the particular cases where the processes are of the Poisson and normal Markov form. In both Bartlett's and Patankar's papers, some mention is made of the problem of estimation. Bartlett obtains maximum likelihood estimators of transition frequencies, and hints at several results which we elaborate in some detail in Chapters 2 and 3 of this thesis. Patankar mentions that work is in progress on the estimation of parameters by a modified minimum χ^2 test, and quotes some of his results in anticipation. It appears nevertheless, that problems arising in the estimation of a parameter defining a simple Markov chain, with which we are concerned, have not as yet been the subject of any systematic study.

4. Sequential estimation as a case of estimation in a Markov chain

The final remark in the previous section is perhaps not entirely true if it is remembered that, as in the case of Moran's work summarised in § 1.2, it is possible to consider sequential sampling as an example of a random walk with absorbing barriers (Feller, 1950; 313). The problem of sequential estimation is then a particular case of the estimation of a parameter defining a simple Markov chain with absorbing states.

Methods of sequential estimation have been studied in some detail, and are reviewed in Anscombe's paper (1953); among others referred to, Girshick, Mosteller and Savage's contribution (1946) to the estimation of the parameter p for a binomial population is of particular interest.

In their paper, samples drawn from a binomial distribution with parameter p are discussed, and an estimator $\hat{p}$ defined as follows. Sampling is considered as a random process, with the event whose probability of occurrence is p at the i -th trial represented by a jump from the point $\alpha_i = (x, y)$ to $\alpha_{i+1} = (x, y+1)$; the non-occurrence of the event whose probability is

$q = 1 - p$, is represented by a jump from $\alpha_i = (x, y)$ to $\alpha_{i+1} = (x+1, y)$. The sampling is allowed to proceed until a point α is reached on the boundary of a certain prescribed region R ; the estimator $\hat{p}$ of p is then defined as

$$(1.4.1) \quad \hat{p}(\alpha) = k^*(\alpha) / k(\alpha),$$

where $k^*(\alpha)$ and $k(\alpha)$ are respectively the number of paths in R from the points $(0, 1)$ and $(0, 0)$ to the boundary point α . It is proved that under certain conditions for the region R , $\hat{p}$ is the unique unbiased estimator of p , and that in some cases it is also sufficient.

Following this work, Wolfowitz (1947) has also shown that the estimator (1.4.1) is consistent, while we have seen in § 1.2 that Moran (1953) using an analogous though slightly different representation of the random walk, showed that the maximum likelihood estimators $\hat{p}_1$ and $\hat{p}_2$ of equation (1.2.1), which are in some cases close to the value of $\hat{p}(\alpha)$, are sufficient and asymptotically normally distributed under certain restrictive conditions.

These results of sequential estimation can be considered as scattered solutions of problems in the

estimation of a parameter defining certain types of Markov chains with absorbing states. They are decidedly of relevance to the general theory of estimation of parameters defining Markov chains, but we shall be unable to fit them into a comprehensive theory, since we have developed only that for positively regular chains. It is hoped that when this theory is extended to include other regular chains, among them those with absorbing states which are at the basis of sequential analysis, sequential estimation will appear as an extension of that estimation theory of parameters defining positively regular chains which we now proceed to examine.

Chapter 2

SOME PROPERTIES OF SIMPLE
MARKOV CHAINSI. Basic Definitions1. Transition probabilities

Suppose that a system consisting of a finite set of s states $E_1, \dots, E_s$, is such that in a realisation of $n+1$ trials, the outcome is a sequence S of states

$$E_1, E_j, \dots, E_k, E_\ell,$$

where the outcome of a particular trial is some state E_r in the set, and this depends on the outcome of the trial preceding it. To every pair of consecutive states (E_i, E_j) say, let there correspond the conditional probability $\Pr(E_j/E_i) = p_{ij} \geq 0$, which we shall call the transition probability from the state E_i to the state E_j of the system; further, let the

probabilities of the states at the initial trial be defined by

$$\Pr(E_i) = a_i \quad (i=1,2,\dots,s),$$

where the $a_i \geq 0$ are known as the initial probabilities; then the probability of the sequence S is clearly

$$(2.1.1) \quad \Pr(S) = a_i p_{ij} \dots p_{kl}.$$

Certain conditions are necessarily satisfied by the initial and the transition probabilities: these are

$$\sum_{i=1}^s a_i = 1,$$

and

$$(2.1.2) \quad \sum_{j=1}^s p_{ij} = 1 \quad (i=1,2,\dots,s).$$

A sequence of trials such as S is known as a simple Markov chain with a finite number of states. We shall not be concerned with the somewhat more complicated multiply dependent chains for which, given the set of s events $E_1, \dots, E_s$, the transition probabilities are defined by

$$\Pr(E_{i_v} / E_{i_1} \dots E_{i_{v-1}}) = p_{i_1 \dots i_v} \quad (2 \leq v \leq s),$$

for all values of $i_1, i_2, \dots, i_v = 1, 2, \dots, s$.

It is convenient for the simple chain, that the

s^2 elements $p_{ij} \geq 0$ be arranged in a matrix of transition probabilities $\underline{p}$ given by

$$(2.1.3) \quad \underline{p} = \begin{pmatrix} p_{11} & \cdots & p_{1s} \\ \cdots & \cdots & \cdots \\ p_{s1} & \cdots & p_{ss} \end{pmatrix};$$

from (2.1.2), it is clear that the sum of the elements in each row is unity. A matrix of non-negative elements satisfying this condition is known as a stochastic matrix, and together with the initial distribution $\{a_i\}$ completely defines a simple Markov chain. If in $\underline{p}$ the columns as well as the rows each have a unit sum so that

$$(2.1.4) \quad \sum_{j=1}^s p_{ij} = 1 = \sum_{i=1}^s p_{ij} \quad (i, j = 1, 2, \dots, s),$$

then the matrix is known as doubly stochastic.

2. Higher transition probabilities

It is frequently necessary to obtain the probability of transition from a state E_i to a state E_j in exactly n trials; this can occur in several different ways, and we denote the probability that the system be in state E_j exactly n trials after starting from the

state E_i by $p_{ij}^{(n)}$, a higher transition probability.

It is easy to see that

$$p_{ij}^{(1)} = p_{ij},$$

and by taking into account all possible ways in which E_j can be reached from E_i in 2 trials, that

$$p_{ij}^{(2)} = \sum_{r=1}^s p_{ir} p_{rj}.$$

It follows quite simply that

$$(2.2.1) \quad p_{ij}^{(n+1)} = \sum_{r=1}^s p_{ir} p_{rj}^{(n)},$$

or more generally that

$$p_{ij}^{(n+m)} = \sum_{r=1}^s p_{ir}^{(n)} p_{rj}^{(m)}.$$

Since, starting from the state E_i , the system must necessarily reach one of the states $E_1, \dots, E_s$ in n trials, it follows that

$$\sum_{j=1}^s p_{ij}^{(n)} = 1,$$

and the higher transition probabilities $p_{ij}^{(n)}$ are clearly the (i,j) -th elements of the n -th power of the stochastic matrix p , so that

$$p^n = \begin{pmatrix} p_{11}^{(n)} & \dots & p_{1s}^{(n)} \\ \dots & \dots & \dots \\ p_{s1}^{(n)} & \dots & p_{ss}^{(n)} \end{pmatrix}.$$

II. Regularity and Positive Regularity

3. Stationary probabilities

It is important to investigate the behaviour of the higher transition probabilities for increasing values of n ; following Fréchet (1952; 25-26), we give a useful inequality for these.

Suppose that in the matrix p^n , the smallest and largest elements in a particular column j consisting of $p_{1j}^{(n)}, \dots, p_{sj}^{(n)}$, are respectively written $b_j^{(n)}$ and $B_j^{(n)}$. Then, from (2.2.1), since

$$b_j^{(n+1)} \geq \sum_r p_{ir} b_j^{(n)}, \quad B_j^{(n+1)} \leq \sum_r p_{ir} B_j^{(n)},$$

it follows that for all $j = 1, \dots, s$,

$$b_j^{(n)} \leq b_j^{(n+1)} \leq B_j^{(n+1)} \leq B_j^{(n)}.$$

This means that, with the column j of the stochastic matrices $p, p^2, \dots, p^n, \dots$ there are associated two monotonic and bounded infinite sequences with elements satisfying the following inequalities

$$(2.3.1) \quad \begin{cases} b_j^{(1)} \leq b_j^{(2)} \leq \dots \leq b_j^{(n)} \leq \dots, \\ B_j^{(1)} \geq B_j^{(2)} \geq \dots \geq B_j^{(n)} \geq \dots; \end{cases}$$

these must converge to limits b_j and B_j respectively, where $b_j \leq B_j$.

One possible case of particular interest is that for which, as $n \rightarrow \infty$, the limits b_j and B_j are identically equal to some P_j , so that for all $i=1, \dots, s$, we have

$$(2.3.2) \quad \lim_{n \rightarrow \infty} p_{ij}^{(n)} = P_j,$$

some probability independent of the suffix i . This means that no matter from what state E_i one may start, the probability of eventually reaching the state E_j is the same. The result (2.3.2) may be written in matrix form as

$$\lim_{n \rightarrow \infty} \underline{p}^n = \underline{1} \underline{P}',$$

where $\underline{P}'$ is the row vector of probabilities P_j , and $\underline{1}$ is the column vector of unit elements. It is clear from (2.2.1) and (2.3.2) that

$$\underline{1} \underline{P}' = \lim_{n \rightarrow \infty} \underline{p}^{n+1} = \lim_{n \rightarrow \infty} \underline{p}^n \underline{p} = \underline{1} \underline{P}' \underline{p},$$

so that the column vector $\underline{P}$ of probabilities P_j satisfying the condition $\sum_{j=1}^s P_j = 1$, is a solution of the matrix equation

$$(2.3.3) \quad \underline{P} = \underline{p}' \underline{P}.$$

This solution is unique; for suppose that another column vector $\underline{X}$ with elements X_i satisfying the condition $\sum_{i=1}^s X_i = 1$ is also a solution of the equation (2.3.3) $\underline{X} = \underline{p}' \underline{X}$. Then iterating n times, and taking the limit as $n \rightarrow \infty$ we obtain that

$$\underline{X} = (\underline{p}')^n \underline{X} \rightarrow (\underline{P} \underline{1}') \underline{X};$$

and since $\underline{1}' \underline{X} = \sum_i X_i = 1$, it follows that $\underline{X}$ is identical with $\underline{P}$.

A chain of this kind, which in its final state is stationary and independent of initial conditions, is known as regular, and the P_j as its stationary probabilities. Since all $p_{ij} \geq 0$, it must follow that all $P_j \geq 0$, and as $\sum_j P_j = 1$, at least one stationary probability is non-zero. The particular case of the regular chain for which all the stationary probabilities P_j are non-zero is known as positively regular.

There is some confusion in the use of terms to describe what we have called a "regular" chain; the term "stationary" is quite unequivocal and can be interchanged with "regular", but we have avoided "ergodic" which is occasionally used in a restricted sense for "regular", and also more widely for what we have described as a "positively regular" chain. We shall

use, throughout this thesis, Fréchet's terms "regular" and "positively regular", which we find both clear and adequate.

4. Necessary and sufficient conditions for regularity and positive regularity

The necessary and sufficient conditions that a chain be regular is that there exist for some sufficiently large value n_0 of n , a matrix $\underline{p}^{n_0}$ among the infinite sequence of matrices $\underline{p}, \underline{p}^2, \dots, \underline{p}^n, \dots$ for which at least one column has non-zero elements. If the chain is to be positively regular, then there must exist a matrix $\underline{p}^{n_0}$ for which all elements are non-zero. This can be interpreted as meaning that a positively regular chain is one for which there exists a number n_0 , sufficiently large to permit any state E_j to be reached in n_0 trials starting from any initial state E_i . The regular chain is one for which there exists a sufficiently large number n_0 of trials to permit at least one state E_j to be reached, starting from any initial state E_i . The proofs given below will be found in Fréchet (1952; 26 et seq.), that for the sufficient condition being a variation on one originally due to Markov.

It is easily seen that if the chain with stochastic matrix p is regular, that is such that

$$\lim_{n \rightarrow \infty} b_j^{(n)} = \lim_{n \rightarrow \infty} B_j^{(n)} = P_j \quad (j = 1, 2, \dots, s),$$

where at least one P_j is non-zero, then the matrix p^{n_0} , for some sufficiently large value n_0 of n , must have at least one column of non-zero elements. If this is so, the same column will naturally be non-zero in the matrices p^n for all values of n greater than n_0 , since by (2.3.1), the smallest element in the column will satisfy the condition

$$b_j^{(n)} \geq b_j^{(n_0)} > 0 \quad (n > n_0).$$

Similarly, it is clear that the necessary condition for a chain to be positively regular is that the matrix p^{n_0} , for some sufficiently large value n_0 of n , consist entirely of non-zero elements. Then for all values of n greater than n_0 , the matrices p^n will also consist of non-zero elements.

We now show that these conditions are sufficient. Consider the sequence of stochastic matrices starting with the matrix p^{n_0} : p^{n_0} , p^{2n_0} , ..., p^{vn_0} , $p^{(v+1) \cdot n_0}$, ..., where v is any positive integer; the value n_0 of n will later be taken as that for which at least one, or

all, columns of the stochastic matrix consist of positive elements. If $p_{ij}^{(\overline{v+1}, n_0)}$, $p_{kj}^{(\overline{v+1}, n_0)}$ are any two elements in the j -th column of $p_{\overline{v+1}, n_0}$, their difference is

$$(2.4.1) \quad p_{ij}^{(\overline{v+1}, n_0)} - p_{kj}^{(\overline{v+1}, n_0)} = \sum_r (p_{ir}^{(n_0)} - p_{kr}^{(n_0)}) p_{rj}^{(v, n_0)}.$$

Now the elements $p_{ir}^{(n_0)}$, $p_{kr}^{(n_0)}$ ($r=1, 2, \dots, s$), in the i -th and k -th rows of p_{n_0} respectively, will be such that for some particular values r' of r , the differences

$$u_{r'} = p_{ir'}^{(n_0)} - p_{kr'}^{(n_0)} \geq 0,$$

and the remaining differences for the other values r'' of r will be

$$u_{r''} = p_{ir''}^{(n_0)} - p_{kr''}^{(n_0)} < 0.$$

Since

$$\sum_{r'} |u_{r'}| - \sum_{r''} |u_{r''}| = \sum_r u_r = \sum_r (p_{ir}^{(n_0)} - p_{kr}^{(n_0)}) = 0,$$

we may write

$$\sum_{r'} |u_{r'}| = \sum_{r''} |u_{r''}| = \Theta,$$

where it is clear that Θ depends only on i and k , and satisfies the inequalities

$$0 \leq \Theta \leq 1.$$

In addition, θ also satisfies the inequalities

$$(2.4.2) \quad \theta = \sum_{r'} p_{ir'}^{(n_0)} - p_{kr'}^{(n_0)} \leq 1 - \sum_{r'} p_{kr'}^{(n_0)} \leq 1 - \sum_{r'} b_{r'}^{(n_0)},$$

and similarly,

$$(2.4.3) \quad \theta \leq 1 - \sum_{r''} b_{r''}^{(n_0)}.$$

Another way of writing (2.4.1) is

$$p_{ij}^{(\overline{v+1}, n_0)} - p_{kj}^{(\overline{v+1}, n_0)} = \sum_{r'} |u_{r'}| p_{rj}^{(vn_0)} - \sum_{r''} |u_{r''}| p_{r''j}^{(vn_0)},$$

and we now consider upper bounds for this difference, depending on whether $\theta > 0$, or $\theta = 0$. If $\theta > 0$, then

$$(2.4.4) \quad p_{ij}^{(\overline{v+1}, n_0)} - p_{kj}^{(\overline{v+1}, n_0)} = \theta \left\{ \frac{\sum_{r'} |u_{r'}| p_{rj}^{(vn_0)}}{\sum_{r'} |u_{r'}|} - \frac{\sum_{r''} |u_{r''}| p_{r''j}^{(vn_0)}}{\sum_{r''} |u_{r''}|} \right\} \leq \theta \{ B_j^{(vn_0)} - b_j^{(vn_0)} \};$$

for the two fractions are effectively weighted means of some of the $p_{rj}^{(vn_0)}$, and must therefore lie between their largest and smallest values. If $\theta = 0$, it is easily seen that the differences u_r must be zero for all values of r , and therefore also every

$$p_{ij}^{(\overline{v+1}, n_0)} - p_{kj}^{(\overline{v+1}, n_0)} = 0 \quad (j=1, 2, \dots, s).$$

It follows that the inequality (2.4.4) holds for all possible values of θ in the range $0 \leq \theta \leq 1$.

Now let the condition be assumed that for the value n_0 of n , at least one column of p^{n_0} , say the j -th,

consist of non-zero elements, of which $b_j^{(n_0)}$ will be the smallest. This j -th column may be one of the set r' or alternatively one of the set r'' ; in the first case, we have

$$\sum_{r'} b_{r'}^{(n_0)} \geq b_j^{(n_0)} = \varepsilon ,$$

and in the second

$$\sum_{r''} b_{r''}^{(n_0)} \geq b_j^{(n_0)} = \varepsilon ,$$

so that in either case, from (2.4.2) and (2.4.3), the value of θ is given by

$$0 \leq \theta \leq 1 - \varepsilon < 1 ,$$

where ε is clearly independent of the initial row values i and k . We may therefore write from (2.4.4), that

$$p_{ij}^{(\sqrt{v+1} \cdot n_0)} - p_{kj}^{(\sqrt{v+1} \cdot n_0)} \leq (1 - \varepsilon) \{ B_j^{(v n_0)} - b_j^{(v n_0)} \} ,$$

or since the right-hand side of this is independent of the row values i, k , that

$$B_j^{(\sqrt{v+1} \cdot n_0)} - b_j^{(\sqrt{v+1} \cdot n_0)} \leq (1 - \varepsilon) \{ B_j^{(v n_0)} - b_j^{(v n_0)} \} .$$

It follows that

$$(2.4.5) \quad B_j^{(\sqrt{v+1} \cdot n_0)} - b_j^{(\sqrt{v+1} \cdot n_0)} \leq (1 - \varepsilon)^v \{ B_j^{(n_0)} - b_j^{(n_0)} \} ,$$

and therefore, as $n \rightarrow \infty$, that the limits B_j and b_j of $B_j^{(n)}$ and $b_j^{(n)}$ are identically equal to some P_j , so that the chain is regular. If, instead of the condition that for some value n_0 of n , at least one column of p^{n_0} consist of non-zero elements, we have the condition that all columns of p^{n_0} consist of non-zero elements, the same argument applies for all values of the columns j , and the chain is then proved to be positively regular.

Some idea of the convergence of the higher transition probability $p_{ij}^{(n)}$ to its limit P_j in cases of regularity or positive regularity can be obtained as follows. Let n be a positive integer such that $vn_0 < n < (v+1)n_0$; since P_j must lie between $B_j^{(vn_0)}$ and $b_j^{(vn_0)}$, we have that

$$|p_{ij}^{(n)} - P_j| \leq B_j^{(n)} - b_j^{(n)} \leq B_j^{(vn_0)} - b_j^{(vn_0)},$$

so that from (2.4.5), we obtain

$$(2.4.6) \quad |p_{ij}^{(n)} - P_j| \leq (1-\varepsilon) \{B_j^{(n_0)} - b_j^{(n_0)}\} < q_{n_0}^{n-1} \{q_{n_0}^{-1} (1-\varepsilon)^{-1} (B_j^{(n_0)} - b_j^{(n_0)})\},$$

where $q_{n_0} = (1-\varepsilon)^{1/n_0} < 1$. We see that the term $|p_{ij}^{(n)} - P_j|$ converges to zero at least as fast as the terms of a certain convergent geometric progression with the ratio q_{n_0} already defined. Fréchet (1952; 29-30)

considers some improvements of this ratio which give geometric progressions converging faster than the one above.

5. Higher transition probabilities and the latent roots of a stochastic matrix

Sylvester's theorem (Frazer, Duncan, Collar, 1947; 83-85) enables us to write for $\underline{p}^n$, the n -th power of the stochastic matrix $\underline{p}$, the expression

$$(2.5.1) \quad \underline{p}^n = \sum_{r=1}^s \mu_r^n \underline{Z}_0(\mu_r) + \underline{R},$$

where the μ_r are roots of the characteristic equation $D(\mu) = |\underline{p} - \mu \underline{I}| = 0$, some of which may be multiple, $\underline{Z}_0(\mu_r)$ are finite matrices, and $\underline{R}$ is a remainder matrix involving the n -th powers of those roots μ_r which are multiple, and polynomials in n of degree at most one less than their multiplicity. It is clear, then, that the evaluation of higher transition probabilities $p_{ij}^{(n)}$, which are elements of the matrix $\underline{p}^n$, and also of the stationary probabilities for regular chains, is closely connected with the latent roots of the stochastic matrix $\underline{p}$.

A simple result which applies to all stochastic matrices is the following: if μ_r ($r = 1, 2, \dots, s$) are

the latent roots of the stochastic matrix $\underline{p}$, then they all have moduli $|\mu_r| \leq 1$, and at least one of them is unity. For if $D(\mu) = |\underline{p} - \mu \underline{I}| = 0$, then there exists at least one non-trivial vector solution $\underline{P}(\mu)$ of the matrix equation

$$(2.5.2) \quad (\underline{p} - \mu \underline{I}) \underline{P}(\mu) = 0,$$

where $\underline{P}(\mu)$ may have real or complex elements. Let the modulus of the largest element $P_1(\mu)$ in $\underline{P}(\mu)$ be M ; then from (2.5.2), since $\sum_j p_{1j} P_j = \mu P_1$, it follows that

$$|\mu| M \leq \sum_j p_{1j} |P_j| \leq M,$$

and $|\mu| \leq 1$. It is easily seen that at least one of the roots of the characteristic equation is unity; for consider the determinant $D(\mu)$, adding its columns we obtain the equation for the latent roots as

$$D(\mu) = \begin{vmatrix} 1-\mu & p_{12} & \cdots & p_{1s} \\ 1-\mu & p_{22}-\mu & \cdots & p_{2s} \\ \cdots & \cdots & \cdots & \cdots \\ 1-\mu & \cdots & \cdots & p_{ss}-\mu \end{vmatrix} = 0,$$

where it is obvious that $\mu = 1$ is a root.

A more important result, which we shall deduce, and which follows from (2.5.1) is that for a regular



chain, the characteristic equation $D(\mu) = 0$ of the stochastic matrix $\underline{p}$ has a root $\mu_1 = 1$ which is simple, and the remaining roots μ_r ($r = 2, \dots, s$) with moduli $|\mu_r| < 1$. In order to prove this, we shall obtain from (2.5.1) an expression for the (i, j) -th element $p_{ij}^{(n)}$ of the matrix $\underline{p}^n$, and consider the consequences when its limit as $n \rightarrow \infty$ is finite.

A simple way of obtaining in detail the expression for $p_{ij}^{(n)}$, that is, of evaluating the elements of the remainder matrix $\underline{R}$ of (2.5.1), is to consider for some suitable μ the expansion in powers of $\underline{p}$ of the matrix

$$(\mu \underline{I} - \underline{p})^{-1} = \mu^{-1} (\underline{I} + \mu^{-1} \underline{p} + \mu^{-2} \underline{p}^2 + \dots)$$

This is convergent for $\mu > 1$. It is clear that the (k, i) -th element of this matrix is

$$(2.5.3) \{(\mu \underline{I} - \underline{p})^{-1}\}_{ki} = \mu^{-1} (\delta_{ki} + \mu^{-1} p_{ki} + \mu^{-2} p_{ki}^{(2)} + \dots),$$

so that if $\{(\mu \underline{I} - \underline{p})^{-1}\}_{ki}$ can be expanded in some other way in descending powers of μ , it is possible to obtain an expression for $p_{ki}^{(n)}$ by equating the coefficients of $\mu^{-(n+1)}$ in the two expansions.

Consider now the matrix

$$\mu \underline{I} - \underline{p}' = \begin{pmatrix} \mu - p_{11} & -p_{12} & \dots & -p_{1s} \\ \dots & \dots & \dots & \dots \\ -p_{ts} & \dots & \dots & \mu - p_{ss} \end{pmatrix};$$

the cofactors $C_{ki}(\mu)$ of each of the elements $\mu \delta_{ik} - p_{ik}$ of the matrix are of degree no greater than $s-1$ in μ , whereas the degree of $-D(\mu) = |\mu \underline{I} - \underline{p}|$ is always one greater than that of any C_{ki} , that is no greater than s . It follows that

$$\{(\mu \underline{I} - \underline{p})^{-1}\}_{ki} = C_{ki}(\mu) |\mu \underline{I} - \underline{p}|^{-1} = C_{ki}(\mu) / (\mu - \mu_1) \cdots (\mu - \mu_s),$$

where $\mu_1, \dots, \mu_s$, are the roots of the characteristic equation $-D(\mu) = |\mu \underline{I} - \underline{p}| = 0$, some of which may be multiple.

Let $\mu_1, \mu_2, \dots, \mu_{j-1}$ represent simple roots, and $\mu_j, \dots, \mu_t$ roots of multiplicity $m_j, \dots, m_t$ respectively; then by the method of partial fractions, we obtain that

$$\{(\mu \underline{I} - \underline{p})^{-1}\}_{ki} = \sum_{r=1}^{j-1} \frac{A_r^{(ki)}}{(\mu - \mu_r)} + \sum_{r=j}^t \frac{B_r^{(ki)}(\mu)}{(\mu - \mu_r)^{m_r}},$$

where the $A_r^{(ki)}$ are constants, and the $B_r^{(ki)}(\mu)$ polynomials in μ of degree no greater than $m_r - 1$.

Since $\mu > 1$, and we know that for any stochastic matrix $\underline{p}$, $|\mu_r| \leq 1$ for all r , we may expand this expression as follows:

$$\begin{aligned} \{(\mu \underline{I} - \underline{p})^{-1}\}_{ki} &= \mu^{-1} \sum_{r=1}^{j-1} A_r^{(ki)} \{1 + \mu^{-1} \mu_r + \mu^{-2} \mu_r^2 + \cdots\} \\ &\quad + \sum_{r=j}^t B_r^{(ki)} \mu^{-m_r} \{1 + m_r \mu^{-1} \mu_r + \cdots + \binom{m_r+n-1}{n} \mu^{-n} \mu_r^n + \cdots\}. \end{aligned}$$

Now for all values of n greater or equal to $m_{r^*} - 1$, where m_{r^*} is the largest value among $m_j, \dots, m_t$, we have that since the polynomials $B_r^{(ki)}(\mu)$ are of the form

$$B_r^{(ki)}(\mu) = b_{r0}^{(ki)} + b_{r1}^{(ki)}\mu + \dots + b_{r, m_r-1}^{(ki)}\mu^{m_r-1},$$

the coefficient of $\mu^{-(n+1)}$ in the element $\{(\mu I - p)^{-1}\}_{ki}$ is equal to

$$\sum_{r=1}^{j-1} A_r^{(ki)} \mu_r^n + \sum_{r=j}^t \left\{ b_{r, m_r-1}^{(ki)} \mu_r^n \binom{m_r+n-1}{n} + \dots + b_{r0}^{(ki)} \mu_r^n \binom{n-m_r+1}{n-m_r+1} \right\};$$

for simplicity, this can be written as

$$\sum_{r=1}^{j-1} A_r^{(ki)} \mu_r^n + \sum_{r=j}^t \beta_r^{(ki)}(n) \mu_r^n,$$

where the $\beta_r^{(ki)}(n)$ are polynomials in n of degree not greater than $m_r - 1$.

From (2.5.3), we see that this is the transition probability $p_{ki}^{(n)}$ for values of n greater than a certain value; this is in any case all that is required, since we are to consider the limit of $p_{ki}^{(n)}$ as $n \rightarrow \infty$. Writing, for sufficiently large n , the result

$$(2.5.4) \quad p_{ki}^{(n)} = \sum_{r=1}^{j-1} A_r^{(ki)} \mu_r^n + \sum_{r=j}^t \beta_r^{(ki)}(n) \mu_r^n,$$

we see that if the chain is regular, so that $p_{ki}^{(n)} \rightarrow P_i$ then the unit root cannot be multiple, for if $\mu_1 = 1$

say, were multiple, then

$$p_{ki}^{(n)} \rightarrow \beta_i^{(ki)} (n) \rightarrow \infty,$$

which contradicts the assumption of regularity. It follows that there is one simple root $\mu_1 = 1$ of the characteristic equation $D(\mu) = 0$ for the stochastic matrix $\underline{p}$ of a regular chain, the remaining roots having moduli $|\mu_r| < 1$ ($r = 2, 3, \dots, s$).

6. Evaluation of the stationary probabilities

We shall later require an interesting method of evaluating the stationary probabilities, which was first derived by Mihoc (1934); an account of this is given in Fréchet (1952; 114). If in the matrix

$$(2.6.1) \quad \underline{p}^i - \underline{I} = \begin{pmatrix} p_{11}^{-1} & p_{21} & \dots & p_{s1} \\ \dots & \dots & \dots & \dots \\ p_{1s} & \dots & p_{ss}^{-1} \end{pmatrix},$$

the cofactors of the i -th diagonal elements are denoted by Δ_i , it is shown that the stationary probabilities P_i associated with the stochastic matrix $\underline{p}$ are

$$(2.6.2) \quad P_i = \Delta_i \left(\sum_{i=1}^s \Delta_i \right)^{-1} \quad (i = 1, 2, \dots, s).$$

Let us write

$$D(\mu) = |\underline{p}' - \mu \underline{I}| = \begin{vmatrix} p_{11} - \mu & p_{21} & \cdots & p_{s1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1k} & \cdots & p_{sk} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1s} & \cdots & p_{ss} - \mu \end{vmatrix},$$

then, summing the columns, and replacing the k -th row in the determinant by this sum, we obtain that

$$D(\mu) = \begin{vmatrix} p_{11} - \mu & \cdots & p_{s1} \\ \vdots & \ddots & \vdots \\ 1 - \mu & \cdots & 1 - \mu \\ \vdots & \ddots & \vdots \\ p_{1s} & \cdots & p_{ss} - \mu \end{vmatrix} = (1 - \mu) \sum_{i=1}^s D_{ki},$$

where the D_{ki} are cofactors of the i -th elements $p_{ik} - \mu \delta_{ik}$ in the k -th row of the determinant $D(\mu)$. It follows that on differentiating $D(\mu)$ with respect to μ , we obtain

$$D'(\mu) = - \sum_i D_{ki} + (1 - \mu) \sum_i D'_{ki},$$

and if we put $\mu = 1$, this is

$$D'(1) = - \left\{ \sum_i D_{ki} \right\}_{\mu=1} = - \sum_i \Delta_{ki},$$

where the Δ_{ki} are cofactors of the i -th elements $p_{ik} - \delta_{ik}$ in the k -th row of the matrix $\underline{p}' - \underline{I}$ of (2.6.1). We see that this sum of cofactors is

independent of the row k , and that further, since $\mu = 1$ is a simple root of $D(\mu)=0$ for a regular chain, then this sum is necessarily non-zero.

Now if we consider any row k of the matrix (2.6.1) it is clear that since the stationary probabilities P_i are given by $(\underline{p}' - \underline{I}) \underline{P} = 0$, then

$$\frac{P_1}{\Delta_{k1}} = \frac{P_2}{\Delta_{k2}} = \dots = \frac{P_s}{\Delta_{ks}} = \frac{1}{\sum_i \Delta_{ki}},$$

so that for any value of $k=1,2,\dots,s$, we have

$$P_i = \Delta_{ki} \left(\sum_k \Delta_{ki} \right)^{-1}.$$

Since $\sum_i \Delta_{ki}$ has the constant value $-D'(1)$ for all k , the values of the cofactors Δ_{ki} are independent of their row value k , and are equal. The stationary probabilities P_i can therefore be written as in (2.6.2), a form frequently convenient to determine.

III. Transition Frequencies and their Moment Generating Function

7. Transition frequencies

Suppose that a realisation in $n+1$ trials for a

finite chain with s states results in the sequence S :

$$E_1, E_j, \dots, E_k, E_\ell.$$

We have seen in (2.1.1) that the probability of S is

$$\Pr(S) = a_1 p_{1j} \dots p_{k\ell},$$

where the a_1 are initial, and the p_{ij} transition probabilities. It is possible, however, if the particular transition probability p_{ij} is repeated n_{ij} times in $\Pr(S)$, to group together the s^2 distinct values of the p_{ij} so as to obtain

$$(2.7.1) \quad \Pr(S) = a_1 \prod_{i,j=1}^s p_{ij}^{n_{ij}}.$$

The n_{ij} , which are the number of transitions from state E_i to state E_j in the realisation S , are known as transition frequencies, and clearly satisfy the equation $\sum_{j=1}^s n_{ij} = n$.

It is important to note that the n_{ij} are not linearly independent; the number of transitions from the state E_1 to the states $E_1, \dots, E_s$, will except for a possible end effect, be equal to the number of transitions from the states $E_1, \dots, E_s$ to the state E_1 , so that

$$\sum_{j=1}^s n_{1j} = \sum_{r=1}^s n_{r1},$$

where the sign $\doteq$ indicates equality or a possible difference of 1 between the sums. If n increases, however, we may accept the equations

$$(2.7.2) \quad \sum_{j=1}^s n_{ij} = \sum_{r=1}^s n_{ri} = n_i \quad (i=1,2,\dots,s),$$

where

$$\sum_{i=1}^s n_i = \sum_{i,j=1}^s n_{ij} = n.$$

The transition frequencies n_{ij} are variates associated with the given Markov chain; it is frequently useful to regard them as the sum of n distinct variates $X_{ij}^{(r)}$, such that

$$n_{ij} = \sum_{r=1}^n X_{ij}^{(r)},$$

where $X_{ij}^{(r)}$ takes the value 1 or 0 at the r -th transition between the $(r-1)$ -th and r -th trials, depending on whether this transition is or is not from the state E_i to the state E_j . This method is described in Fréchet (1952; 73), and will be used to find the expectation and variance of the transition frequencies n_{ij} in the case of the regular chain.

For simplicity, we shall assume that the chain is initially stationary so that the initial probabilities a_i will be equal to the stationary probabilities P_i .

It is clear then, that

$$\begin{aligned}\xi(x_{ij}^{(r)}) &= P_i p_{ij} \\ \psi(x_{ij}^{(r)}) &= \xi\{(x_{ij}^{(r)})^2\} - \{\xi(x_{ij}^{(r)})\}^2 = P_i p_{ij} - P_i^2 p_{ij}^2, \\ &\quad (r = 1, 2, \dots, n),\end{aligned}$$

and

$$\begin{aligned}\phi(x_{ij}^{(r)}, x_{ij}^{(t)}) &= \xi(x_{ij}^{(r)} x_{ij}^{(t)}) - \xi(x_{ij}^{(r)}) \xi(x_{ij}^{(t)}) = P_i p_{ij}^{(t-r-1)} p_{ji} - P_i^2 p_{ij}^2 \\ &\quad (r, t = 1, 2, \dots, n; t > r),\end{aligned}$$

where, for $t-r-1 \geq 1$, $p_{ji}^{(t-r-1)}$ is the probability that a transition from the state E_j to the state E_i occur in $t-r-1$ steps, and is the (j, i) -th element of the matrix $\underline{p}^{t-r-1}$, but for $t-r-1 = 0$, $p_{ji}^{(0)} = \delta_{ji}$. It follows that the expectation and variance of the transition frequency n_{ij} are given by

$$(2.7.3) \quad \xi(n_{ij}) = \sum_{r=1}^{\infty} \xi(x_{ij}^{(r)}) = n P_i p_{ij},$$

and

$$(2.7.4) \quad \begin{cases} \sigma_{ij}^2 = \psi(n_{ij}) = \psi\left(\sum_{r=1}^{\infty} x_{ij}^{(r)}\right) = \sum_{r=1}^{\infty} \psi(x_{ij}^{(r)}) + 2 \sum_{t>r} \phi(x_{ij}^{(r)}, x_{ij}^{(t)}) \\ \quad = n(P_i p_{ij} - P_i^2 p_{ij}^2) + 2 P_i p_{ij}^2 \sum_{t>r} (p_{ji}^{(t-r-1)} - P_i), \end{cases}$$

where there are $\frac{1}{2}n(n-1)$ terms of the form $(p_{ji}^{(t-r-1)} - P_i)$.

We shall later require the limit of $\sigma_{ij}^2 n^{-1}$; it can be fairly easily seen that this exists and is

independent of n . For since the chain is assumed to be regular, we have from equation (2.4.6) that the terms $|p_{ji}^{(t-r-1)} - P_i|$ converge to zero at least as fast as the terms of a certain convergent geometric progression. If, therefore, we write the variance (2.7.4) in the somewhat different form

$$\sigma_{ij}^2 = n(P_i p_{ij} - P_i^2 p_{ij}^2) + 2 P_i p_{ij}^2 \sum_{k=0}^{n-2} (n-k-1)(p_{ji}^{(k)} - P_i),$$

we see that

$$\sigma_{ij}^2 n^{-1} = \{P_i p_{ij} - P_i^2 p_{ij}^2 + 2 P_i p_{ij}^2 \sum_{k=0}^{n-2} (p_{ji}^{(k)} - P_i)\} - 2 P_i p_{ij}^2 \sum_{k=0}^{n-2} n^{-1}(k+1)(p_{ji}^{(k)} - P_i),$$

which will clearly converge to

$$(2.7.5) \quad \lim_{n \rightarrow \infty} \sigma_{ij}^2 n^{-1} = P_i p_{ij} - P_i^2 p_{ij}^2 + 2 P_i p_{ij}^2 s_{ji} = A,$$

where A is some value independent of n , and

$$s_{ji} = \sum_{k=0}^{\infty} (p_{ji}^{(k)} - P_i).$$

We now prove that for positively regular chains, the limit A must be non-zero. To do this, we use a theorem given in Fréchet (1952; 86-88) which applies to the frequencies n_i of (2.7.2) giving the number of times in a realisation that the system is in state E_i . The variances $\sigma_i^2 = \mathcal{V}(n_i)$ of these can be shown, by a method similar to that used for the n_{ij} above, to be such that $\lim_{n \rightarrow \infty} \sigma_i^2 n^{-1}$ is some value, finite and independent of n . The theorem proves that for a positively regular

chain of the type we consider, $\lim_{n \rightarrow \infty} \sigma_1^2 n^{-1}$ cannot be zero.

In order to apply this result to our transition frequencies n_{ij} , we re-define our system of states in the following fashion, assuming first that the p_{ij} are all non-zero. We define a system of s^2 states E_{ij} ($i, j = 1, 2, \dots, s$), in which the system will be in state E_{ij} when there is a transition from state E_i to E_j of the original system; the new stochastic matrix for this system will be

$$\underline{\gamma} = \begin{pmatrix} \begin{bmatrix} p_{11} \dots p_{1s} & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & p_{21} \dots p_{2s} & \dots & 0 \dots 0 \\ \dots & \dots & \dots & \dots \\ 0 \dots 0 & \dots & \dots & p_{s1} \dots p_{ss} \end{bmatrix} \\ \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix} \\ \dots \\ \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix} \end{pmatrix}.$$

If the original stochastic matrix is written as

$$\underline{p} = (\underline{\alpha}_1 \dots \underline{\alpha}_s) = \begin{pmatrix} \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_s \end{pmatrix}$$

where $\underline{\alpha}_i$, $\underline{\beta}_i$ are respectively column and row vectors, then $\underline{\gamma}$ can be written as

$$\underline{\gamma} = \begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \beta_s \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

It is clear that n_{ij} will now indicate the number of times in a realisation of the chain that the system is in the state E_{ij} ; in other words, the n_{ij} in the new system are the analogues of the n_i in the original system. All that remains to be proved is that if the original system is positively regular, so also will the new system; this is intuitively obvious, and can be easily shown by powering the matrix $\underline{\gamma}$ as follows. If we write for the n -th power of the matrix $\underline{p}$

$$\underline{p}^n = (\alpha_1^{(n)} \dots \alpha_s^{(n)}),$$

it is seen directly that on multiplying the matrix by itself, we obtain

$$\underline{\gamma}^2 = \begin{pmatrix} p_{11} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \end{bmatrix} & \dots & p_{1s} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \end{bmatrix} & p_{21} \begin{bmatrix} \alpha_2 \\ \vdots \\ \alpha_2 \end{bmatrix} & \dots & p_{2s} \begin{bmatrix} \alpha_2 \\ \vdots \\ \alpha_2 \end{bmatrix} & \dots & p_{s1} \begin{bmatrix} \alpha_s \\ \vdots \\ \alpha_s \end{bmatrix} & \dots & p_{ss} \begin{bmatrix} \alpha_s \\ \vdots \\ \alpha_s \end{bmatrix} \end{pmatrix},$$

and similarly, since for $n \rightarrow \infty$, $\lim_{n \rightarrow \infty} \alpha_l^{(n)} = P_l \underline{1}$, then it follows that

$$\underline{\gamma}^{n+1} = \left(p_{11} \begin{bmatrix} \alpha_1^{(n)} \\ \vdots \\ \alpha_1^{(n)} \end{bmatrix} \cdot \cdot \cdot p_{ss} \begin{bmatrix} \alpha_s^{(n)} \\ \vdots \\ \alpha_s^{(n)} \end{bmatrix} \right) \rightarrow \begin{pmatrix} p_{11} P_1 & p_{ss} P_s \\ \vdots & \vdots \\ p_{11} P_1 & p_{ss} P_s \end{pmatrix}.$$

We see therefore by Fréchet's theorem that $\lim_{n \rightarrow \infty} \sigma_{1j}^2 n^{-1} \neq 0$. If a certain p_{1j} is zero, then in our new system the state E_{1j} must be eliminated, since no transition into or from this state is possible. It can be verified, however, that the results above will hold equally well in such a case.

8. Moment generating function of the transition frequencies

In obtaining the asymptotic distribution of the transition frequencies n_{1j} ($i, j = 1, 2, \dots, s$), we shall later require their moment generating function

$$\begin{aligned} M(\underline{t}) &= M(t_{11}, \dots, t_{1s}; t_{21}, \dots, t_{2s}; \dots; t_{s1}, \dots, t_{ss}) \\ &= \mathcal{E}(\exp \sum_{i,j=1}^s t_{ij} n_{ij}), \end{aligned}$$

where the matrix $\underline{t}$ represents the s^2 variables t_{ij} ($i, j = 1, 2, \dots, s$).

Following Montroll (1947) and Bartlett (1951),

we have that the function $\exp \sum_{i,j=1}^s t_{ij} n_{ij}$ can be written in the form of the n products

$$\prod_{r=1}^{\infty} \exp \sum_{i,j=1}^s t_{ij} X_{ij}^{(r)} = \prod_{r=1}^{\infty} g^{(r)},$$

and its expectation can be evaluated in n steps so that, for example

$$\xi \left(\prod_{r=1}^{\infty} g^{(r)} \right) = \xi_{n-1} \left\{ \prod_{r=1}^{n-1} g^{(r)} \xi_{n/n-1}(g^{(n)}) \right\}.$$

Here, ξ_{n-1} gives the expectation over all the variables up to the $(n-1)$ -th transition, and $\xi_{n/n-1}$ indicates the conditional expectation over the variables at the n -th transition for a prescribed initial state and given values of the variables at all transitions up to the $(n-1)$ th.

This has the same structure as the evaluation of a final probability distribution after n transitions, with elements $\{p_{ij} \exp t_{ij}\}$ in place of the elements p_{ij} . Hence, if we denote by $\underline{R}$ a matrix with the transpose

$$(2.8.1) \quad \underline{R}' = \begin{pmatrix} p_{11} \exp t_{11} & \dots & p_{1s} \exp t_{1s} \\ \dots & \dots & \dots \\ p_{s1} \exp t_{s1} & \dots & p_{ss} \exp t_{ss} \end{pmatrix},$$

we have that

$$M(\underline{t}) = \xi \left(\exp \sum_{i,j} t_{ij} n_{ij} \right) = \underline{1}' \underline{R}^n \underline{a},$$

where $\underline{1}$ ' is the row vector of unit elements, and $\underline{a}$ the column vector of initial probabilities a_i . If the chain is initially stationary, the vector $\underline{a}$ is equal to $\underline{P}$ with elements P_i , the stationary probabilities, and the moment generating function is given by

$$(2.8.2) \quad M(\underline{t}) = \underline{1}' \underline{R}^n \underline{P}.$$

9. Latent roots of the matrix $\underline{R}' = \{p_{ij} \exp t_{ij}\}$ for a positively regular chain

The latent roots $\mu_r(\underline{t})$ of the matrix $\underline{R}'$ will be given by the determinantal equation

$$(2.9.1) \quad | \underline{R} - \mu \underline{I} | = 0,$$

and will clearly be continuous in the t_{ij} . For $\underline{t} = 0$, this becomes the characteristic equation for the stochastic matrix $\underline{p}$,

$$(2.9.2) \quad | \underline{p} - \mu \underline{I} | = 0,$$

with roots $\mu_1(0), \dots, \mu_s(0)$, not necessarily all distinct; we may, without loss of generality, assume these to be arranged so that their moduli are in descending order of magnitude. For a regular chain,

we have seen in § 2.5 that these roots are such that

$$\mu_1(0) = 1, \quad |\mu_r(0)| < 1 \quad (r = 2, 3, \dots, s),$$

and it follows from the continuity of the roots, that for $\underline{t}$ in the neighbourhood of $\underline{t} = 0$, we have

$$|\mu_r(\underline{t})| < |\mu_1(\underline{t})| \quad (r = 2, 3, \dots, s).$$

We prove that for a positively regular chain, for some $\underline{t}$ in the neighbourhood of $\underline{t} = 0$, the latent root $\mu_1(\underline{t})$ is not identically equal to 1. For suppose that $\mu_1(\underline{t}) \equiv 1$: then for $\underline{t}$ such that $t_{11} \neq 0$, and $t_{1j} = 0$ for all other values of i, j , the equation (2.9.1) would give

$$(2.9.3) \quad \begin{vmatrix} p_{11} \exp t_{11} - 1 & p_{12} & \dots & p_{1s} \\ p_{21} & p_{22} - 1 & \dots & p_{2s} \\ \dots & \dots & \dots & \dots \\ p_{s1} & \dots & \dots & p_{ss} - 1 \end{vmatrix} = 0.$$

On expansion, this could be written

$$(p_{11} \exp t_{11} - 1) C_{11} + p_{12} C_{12} + \dots + p_{1s} C_{1s} = 0,$$

where the C_{1j} are cofactors of the elements in the first row and j -th column. For t_{11} small, this is

$$p_{11} t_{11} C_{11} + (p_{11} - 1) C_{11} + p_{12} C_{12} + \dots + p_{1s} C_{1s} = 0.$$

Now if $t_{11} = 0$ also, so that $\underline{t} = 0$, equation (2.9.3) would give

$$(p_{11} - 1)C_{11} + p_{12}C_{12} + \dots + p_{1s}C_{1s} = 0;$$

we see therefore that $\mu_1(\underline{t}) \equiv 1$ only if $p_{11}C_{11}t_{11} = 0$, so that $p_{11} = 0$, or $C_{11} = 0$, or both are zero. Now C_{11} cannot be zero, for since $\mu_1(0) = 1$ is a simple root, then

$$\left\{ \frac{d}{d\mu} (|\underline{p} - \mu \underline{I}|) \right\}_{\mu=1} \neq 0,$$

or on expansion

$$-\{C_{11} + C_{22} + \dots + C_{ss}\} \neq 0,$$

where the C_{ii} are cofactors of the elements in the leading diagonal of (2.9.3) when $t_{11} = 0$. At least one of these C_{ii} is non-zero, and we may without loss of generality assume that C_{11} is such a non-zero cofactor. If, in addition, p_{11} is non-zero, then $\mu_1(\underline{t}) \neq 1$ for at least the case when $t_{11} \neq 0$ and $t_{1j} = 0$ for all other values of i, j .

It is possible, however, that p_{11} be zero; then in a positively regular chain, at least two of the remaining probabilities $p_{12}, \dots, p_{1s}$ will be non-zero. For it is known that, in a positively regular chain,

$0 \leq p_{ij} \leq 1$, but no $p_{ii} = 1$ ($i, j = 1, \dots, s$); if, as is possible, a particular $p_{ij} = 1$ for $i \neq j$, then the states E_i, E_j can be contracted, and we may therefore accept that $0 \leq p_{ij} < 1$ ($i, j = 1, \dots, s$). It follows therefore that in any row of the stochastic matrix, there must be at least two non-zero elements.

From equation (2.9.2), we see that in this case

$$-C_{11} + p_{12}C_{12} + \dots + p_{1s}C_{1s} = 0,$$

so that it is clear that at least one of the cofactors C_{1r} ($r = 2, \dots, s$) which multiplies a non-zero probability p_{1r} is non-zero. Let $\underline{t}$ now be such that $t_{1r} \neq 0$, and the remaining t_{1j} are all zero; then in exactly the same way as before, for a small value of t_{1r} , we can show that $\mu_1(\underline{t}) \equiv 1$ only if $p_{1r}C_{1r}t_{1r} = 0$, which is impossible. So that even if $p_{11} = 0$, there still exists some value of $\underline{t}$ for which $\mu_1(\underline{t}) \neq 1$. It follows therefore in general, that for a positively regular chain, for some $\underline{t}$ in the neighbourhood of $\underline{t} = 0$, the latent root $\mu_1(\underline{t})$ is not identically equal to 1.

Chapter 3

OPTIMUM PROPERTIES OF THE
MAXIMUM LIKELIHOOD ESTIMATOR OF A PARAMETER IN A
MARKOV CHAIN

1. The likelihood equation for the realisation
of a chain

Consider the positively regular Markov chain with s possible states $E_1, \dots, E_s$, for which the transition probabilities p_{ij} are all functions of an unknown parameter θ , so that

$$p_{ij} = p_{ij}(\theta) \quad (i, j = 1, 2, \dots, s).$$

If a realisation of the chain results in the observed sequence S of $n+1$ states in the following order,

$$E_1, E_j, \dots, E_k, E_\ell,$$

then we have seen in (2.7.1) that the probability of this sequence is

$$\Pr(S) = a_1 \prod_{i,j=1}^s p_{ij}^{n_{ij}},$$

where the initial probabilities a_1 are assumed to be known functions of the parameter θ , and the n_{ij} are transition frequencies.

The likelihood function will be given by

$$(3.1.1) \quad L = \ln a_i + \sum_{i,j=1}^s n_{ij} \ln p_{ij},$$

and providing the $a_i(\theta)$ and $p_{ij}(\theta)$ are differentiable with respect to θ , the derivative of the likelihood is

$$(3.1.2) \quad \frac{dL}{d\theta} = \frac{1}{a_i} \frac{da_i}{d\theta} + \sum_{i,j=1}^s \frac{n_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta}.$$

As n increases, the second parts of (3.1.1) and (3.1.2) become dominant, and the likelihood function and its derivative will be

$$(3.1.3) \quad L \sim \sum_{i,j} n_{ij} \ln p_{ij},$$

and

$$(3.1.4) \quad \frac{dL}{d\theta} \sim \sum_{i,j} \frac{n_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta},$$

respectively. The maximum likelihood estimator T of will in this case be given by

$$(3.1.5) \quad \left\{ \frac{dL}{d\theta} \right\}_T \sim \left\{ \sum_{i,j} \frac{n_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta} \right\}_T = 0,$$

where the n_{ij} are not linearly independent, and satisfy equations (2.7.2).

2. Optimum properties of the estimator

Following closely the proofs given by Cramér (1946) and Rao (1952) for continuous distributions, we shall deduce that for a positively regular chain, the maximum likelihood estimator T of θ obtained as a solution of (3.1.5) has the following optimum properties:

- a) the likelihood equation (3.1.5) has a solution T which is consistent: that is, convergent in probability as $n \rightarrow \infty$ to the true value θ_0 of θ ;
- b) any consistent solution of (3.1.5) is such that, as $n \rightarrow \infty$, the probability that the likelihood tend to a maximum converges to 1: the consistent solution is therefore unique;
- c) the consistent solution of (3.1.5) is asymptotically normally distributed about θ_0 : it is fully efficient.

In order to establish these properties, we require the conditions that the derivatives $\frac{dp_{ij}}{d\theta}$, $\frac{d^2 p_{ij}}{d\theta^2}$, $\frac{d^3 p_{ij}}{d\theta^3}$, exist and are continuous for every θ in a range R including the true value θ_0 . In equation (3.1.1), no assumption was made about the chain being initially

stationary, so that the initial probabilities a_i are not necessarily equal to the stationary probabilities P_i ; for a positively regular chain, these are non-zero, and it is clear that since they are given by the matrix equation (2.3.3), $\underline{P} = \underline{p}' \underline{P}$, together with the condition $\sum_i P_i = 1$, the P_i are also functions of θ . If therefore, the process is initially stationary, we can write $a_i = P_i$ in equations (3.1.1) and (3.1.2). In all that follows, we shall for simplicity consider an initially stationary chain; however, the results obtained will also hold asymptotically for a chain which is not initially stationary.

3. Consistency of the estimator

If the estimator T obtained as a solution of (3.1.5) is to be consistent, it is necessary that if θ_0 is the real value of θ , then as $n \rightarrow \infty$,

$$(3.3.1) \quad P_r \{ |T - \theta_0| < \delta \} \rightarrow 1,$$

where δ is any small positive number.

Consider now the expansion

$$(3.3.2) \quad \frac{dL}{d\theta} = \left\{ \frac{dL}{d\theta} \right\}_{\theta_0} + (\theta - \theta_0) \left\{ \frac{d^2L}{d\theta^2} \right\}_{\theta_0} + \frac{1}{2} (\theta - \theta_0)^2 \left\{ \frac{d^3L}{d\theta^3} \right\}_{\theta_0},$$

where θ_1 lies in (θ, θ_0) . For simplicity, we write $n^{-1}\{\frac{dL}{d\theta}\}_{\theta_0}$, $n^{-1}\{\frac{d^2L}{d\theta^2}\}_{\theta_0}$, and $n^{-1}\{\frac{d^3L}{d\theta^3}\}_{\theta_1}$ as B_0 , B_1 and B_2 respectively, where these are obtained from (3.1.4) as

$$(3.3.3) \quad \begin{cases} B_0 = \left\{ \sum_{i,j} \frac{n_{ij} n^{-1}}{p_{ij}} \frac{dp_{ij}}{d\theta} \right\}_{\theta_0}, \\ B_1 = \left\{ \sum_{i,j} n_{ij} n^{-1} \left[\frac{1}{p_{ij}} \frac{d^2 p_{ij}}{d\theta^2} - \frac{1}{p_{ij}^2} \left(\frac{dp_{ij}}{d\theta} \right)^2 \right] \right\}_{\theta_0}, \\ B_2 = \left\{ \sum_{i,j} n_{ij} n^{-1} \left[\frac{1}{p_{ij}} \frac{d^3 p_{ij}}{d\theta^3} - \frac{3}{p_{ij}^2} \frac{dp_{ij}}{d\theta} \frac{d^2 p_{ij}}{d\theta^2} + \frac{2}{p_{ij}^3} \left(\frac{dp_{ij}}{d\theta} \right)^3 \right] \right\}_{\theta_1}. \end{cases}$$

It is clear that as $n \rightarrow \infty$, the expectations of the $n_{ij} n^{-1}$ will be given from (2.7.3) for all $i, j = 1, 2, \dots, s$, by

$$E(n_{ij} n^{-1}) = P_i p_{ij}.$$

Further, since on differentiation with respect to θ , the equations $\sum_{j=1}^s p_{1j} = 1$ ($i = 1, 2, \dots, s$) give

$$\sum_j \frac{dp_{ij}}{d\theta} = \sum_j \frac{d^2 p_{ij}}{d\theta^2} = \sum_j \frac{d^3 p_{ij}}{d\theta^3} = 0,$$

we obtain for the expectations of B_0 , B_1 , B_2 , the results

$$(3.3.4) \quad \begin{cases} E(B_0) = \left\{ \sum_i P_i \sum_j \frac{dp_{ij}}{d\theta} \right\}_{\theta_0} = 0, \\ E(B_1) = - \left\{ \sum_i P_i \sum_j \frac{1}{p_{ij}} \left(\frac{dp_{ij}}{d\theta} \right)^2 \right\}_{\theta_0} = -i(\theta_0), \\ E(B_2) = \left\{ \sum_i P_i \sum_j \left[\frac{2}{p_{ij}^3} \left(\frac{dp_{ij}}{d\theta} \right)^3 - \frac{3}{p_{ij}^2} \frac{dp_{ij}}{d\theta} \frac{d^2 p_{ij}}{d\theta^2} \right] \right\}_{\theta_1} = K(\theta_1), \end{cases}$$

where $i(\theta)$ is obviously finite and positive. We now prove that the variates $n_{ij}n^{-1}$ converge in probability to their expectations; from this it will follow that although the $n_{ij}n^{-1}$ are not independent, B_0, B_1, B_2 , which are linear functions of these variates also jointly converge in probability to their expectations.

We have seen in § 2.7 that for a positively regular chain, as $n \rightarrow \infty$, $\sigma_{ij}^2 n^{-1} = \psi(n_{ij})n^{-1}$ tends to a finite non-zero limit A . We conclude from Tchebyshev's theorem that the variates $n_{ij}n^{-1}$ converge in probability to their expectations, since for ε a small positive number,

$$\Pr \{ |n_{ij}n^{-1} - p_i p_j| \geq \varepsilon \} \leq \sigma_{ij}^2 (n\varepsilon)^{-2} \leq A n^{-1} \varepsilon^{-2}.$$

We may now express the joint convergence in probability of B_0, B_1, B_2 to their expectations in the form

$$\Pr \{ |B_0| < \delta^2, B_1 < -\frac{1}{2} i(\theta_0), B_2 < 2K(\theta_1) \} > 1 - \varepsilon,$$

where δ, ε , are any two small positive numbers. Let F denote the set of points for which this inequality holds; then for every point in F , we have that

$$B_0 + \frac{1}{2} B_2 \delta^2 < \{ 1 + |K(\theta_1)| \} \delta^2,$$

and

$$B_1 \delta < -\frac{1}{2} i(\theta_0) \delta.$$

If $\theta = \theta_0 \pm \delta$ are two such points in F , then from (3.3.2), we have for these points that

$$\pi^{-1} \frac{dL}{d\theta} = B_0 \pm \delta B_1 + \frac{1}{2} \delta^2 B_2 .$$

Provided that δ is so defined that

$$\delta < \frac{1}{2} i(\theta_0) \{1 + |K(\theta_1)|\}^{-1} ,$$

then, for $\theta = \theta_0 + \delta$,

$$\pi^{-1} \frac{dL}{d\theta} < \{1 + |K(\theta_1)|\} \left[\delta - \frac{1}{2} i(\theta_0) \{1 + |K(\theta_1)|\}^{-1} \right] \delta < 0 ,$$

whereas for $\theta = \theta_0 - \delta$,

$$\pi^{-1} \frac{dL}{d\theta} > \{1 + |K(\theta_1)|\} \left[\delta + \frac{1}{2} i(\theta_0) \{1 + |K(\theta_1)|\}^{-1} \right] \delta > 0 .$$

We see from (3.1.4) that $\frac{dL}{d\theta}$ is a continuous function of θ in R , and it follows therefore that the likelihood equation (3.1.5) has a root in $\theta_0 \pm \delta$ with probability tending to 1. This establishes the consistency of the estimator T of θ .

4. Uniqueness of the consistent estimator

This second property of the consistent solution of (3.1.5) is due to Huzurbazar (1948) who obtained it for the case of continuous distributions. For the

positively regular chain, the proof is identical in all its essential points with the former, and will only be briefly outlined for the sake of completeness.

The consistent estimator of θ , T will satisfy equation (3.3.1). Now, from the mean value theorem, we may write

$$n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_T - n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_{\theta_0} = n^{-1} \left\{ \frac{d^3 L}{d\theta^3} \right\}_{\theta_1} (T - \theta_0),$$

so that as $n \rightarrow \infty$,

$$P_r \left[\left| n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_T - n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_{\theta_0} \right| < |K(\theta_1)|\delta \right] \rightarrow 1,$$

or $n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_T$ converges stochastically to $-i(\theta_0)$, a finite negative value. For an arbitrarily small positive number ε' , such that $-i(\theta_0) + \varepsilon' < 0$, we may write

$$P_r \left[n^{-1} \left\{ \frac{d^2 L}{d\theta^2} \right\}_T < -i(\theta_0) + \varepsilon' \right] \rightarrow 1,$$

so that

$$P_r \left[\left\{ \frac{d^2 L}{d\theta^2} \right\}_T < 0 \right] \rightarrow 1.$$

This proves that any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tending to 1 as $n \rightarrow \infty$.

To show that a consistent solution of the

likelihood equation is unique, we assume that there are two consistent solutions T_1 and T_2 of the equation. From Rolle's theorem, it will follow that $\frac{d^2L}{d\theta^2} = 0$ has at least one solution T_3 lying between T_1 and T_2 , and since these converge stochastically to θ_0 , so also will T_3 . From this it would follow that as $n \rightarrow \infty$,

$$Pr \left[\left\{ \frac{d^2L}{d\theta^2} \right\}_{T_3} < 0 \right] \rightarrow 1,$$

which contradicts the equation $\left\{ \frac{d^2L}{d\theta^2} \right\}_{T_3} = 0$. The consistent solution of likelihood equation is therefore unique.

5. Asymptotic normality of the estimator

Let the consistent solution of the likelihood equation be T ; then from equation (3.3.2), we have that

$$\pi^{-1} \left\{ \frac{dL}{d\theta} \right\}_T = \pi^{-1} \left\{ \frac{dL}{d\theta} \right\}_{\theta_0} + (T - \theta_0) \pi^{-1} \left\{ \frac{d^2L}{d\theta^2} \right\}_{\theta_0} + \frac{1}{2} (T - \theta_0)^2 \pi^{-1} \left\{ \frac{d^3L}{d\theta^3} \right\}_{\theta_1} = 0,$$

or more simply

$$B_0 + (T - \theta_0) B_1 + \frac{1}{2} (T - \theta_0)^2 B_2 = 0,$$

so that

$$(T - \theta_0) = -B_0 \left\{ B_1 + \frac{1}{2} (T - \theta_0) B_2 \right\}^{-1}.$$

We have seen that the expectation of B_0 is zero; its variance

$$(3.5.1) \quad \mathcal{V}(B_0) = \mathcal{V}\left(\left\{\sum_{i,j} n_{ij} \kappa^{-1} \left(\frac{1}{p_{ij}} \frac{dp_{ij}}{d\theta}\right)\right\}_T\right) = \kappa^{-2} \mathcal{E}\left(\left\{\sum_{i,j} \frac{n_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta}\right\}_T\right)^2$$

can be evaluated in the following manner. As in § 2.7, we treat each n_{ij} as the sum of the variates $X_{ij}^{(r)}$ which take the values 1 or 0 at each of the n transitions ($r = 1, \dots, n$), depending on whether these transitions are or are not from the state E_1 to the state E_j .

We then have

$$\begin{aligned} \kappa^{-2} \mathcal{E}\left(\sum_{i,j} \frac{n_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta}\right)^2 &= \kappa^{-2} \mathcal{E}\left(\sum_{i,j,r} \frac{X_{ij}^{(r)}}{p_{ij}} \frac{dp_{ij}}{d\theta}\right)^2 \\ &= \kappa^{-2} \sum_r \mathcal{E}\left(\sum_{i,j} \frac{X_{ij}^{(r)}}{p_{ij}} \frac{dp_{ij}}{d\theta}\right)^2 + 2 \kappa^{-2} \sum_{r < t} \mathcal{E}\left(\sum_{i,j} \frac{X_{ij}^{(r)}}{p_{ij}} \frac{dp_{ij}}{d\theta} \sum_{\ell,m} \frac{X_{\ell m}^{(t)}}{p_{\ell m}} \frac{dp_{\ell m}}{d\theta}\right) \\ &= \kappa^{-2} \sum_r \mathcal{E}\left(\sum_{i,j} \left[\frac{X_{ij}^{(r)}}{p_{ij}}\right]^2 \left[\frac{dp_{ij}}{d\theta}\right]^2\right) + \kappa^{-2} \sum'_{i,j} \sum'_{\ell,m} \frac{X_{ij}^{(r)}}{p_{ij}} \frac{X_{\ell m}^{(r)}}{p_{\ell m}} \frac{dp_{ij}}{d\theta} \frac{dp_{\ell m}}{d\theta} \\ &\quad + 2 \kappa^{-2} \sum_{r < t} \sum_{i,j} \sum_{\ell,m} \mathcal{E}\left(\frac{X_{ij}^{(r)}}{p_{ij}} \frac{X_{\ell m}^{(t)}}{p_{\ell m}} \frac{dp_{ij}}{d\theta} \frac{dp_{\ell m}}{d\theta}\right), \end{aligned}$$

where $\sum_r, \sum_{i,j}, \sum_{\ell,m}$ indicate summation over all possible values of $r = 1, 2, \dots, n$; i, j and $\ell, m = 1, 2, \dots, s$, but where $\sum'_{i,j}, \sum'_{\ell,m}$ indicate summation over those values of i, j and $\ell, m = 1, 2, \dots, s$, for which at least one of the conditions $i \neq j, \ell \neq m$ holds. Since for these values of i, j and ℓ, m we have that $X_{ij}^{(r)} X_{\ell m}^{(r)} = 0$, and since further $\sum_m \frac{dp_{\ell m}}{d\theta} = 0$ for all ℓ , then this is equal to

$$\begin{aligned} n^{-2} E \left(\sum_{ij} \frac{\pi_{ij}}{p_{ij}} \frac{dp_{ij}}{d\theta} \right)^2 &= n^{-2} \sum_{ij} E \left(\sum_{r=1}^n \frac{X_{ij}^{(r)}}{p_{ij}} \left[\frac{dp_{ij}}{d\theta} \right]^2 \right) + 2n^{-2} \sum_{r \neq s} \sum_{ij} \sum_{\ell, m} P_i P_{j\ell}^{(r-s-1)} \frac{p_{ij} p_{\ell m}}{p_{ij} p_{\ell m}} \frac{dp_{ij}}{d\theta} \frac{dp_{\ell m}}{d\theta} \\ &= n^{-1} \sum_{ij} \frac{P_i}{p_{ij}} \left[\frac{dp_{ij}}{d\theta} \right]^2. \end{aligned}$$

It follows that $V(B_0) = n^{-1} E(B_1) = i(\theta_0)n^{-1}$, so that

$$(T - \theta_0) \{i(\theta_0)n\}^{\frac{1}{2}} = \frac{\frac{B_0}{\{i(\theta_0)n\}^{\frac{1}{2}}}}{\left\{ \frac{B_1}{-i(\theta_0)} + \frac{(T - \theta_0)B_2}{-2i(\theta_0)} \right\}}$$

has zero expectation and unit variance, for the denominator converges to 1 in probability, and the numerator has zero expectation and unit variance. If we can further show that B_0 is asymptotically normally distributed, then T will be normally distributed with expectation θ_0 , and variance $\{i(\theta_0)n\}^{-1}$.

Now from (3.3.3), we see that B_0 is a linear function of the n_{ij} , so that it is sufficient to prove that as $n \rightarrow \infty$, these are jointly normally distributed. This has been proved for the positively regular chain by Bartlett (1951), who in his work also implied some of the properties of maximum likelihood estimators developed in detail in this chapter. Two differences in the proof which is briefly outlined below, however, are that it is no longer necessary for positively regular chains to postulate that



- a) $\ln \mu_1(\underline{t}) \neq 0$, since it was proved in § 2.9 that for some $\underline{t}$ in the neighbourhood of $\underline{t} = 0$, the latent root $\mu_1(\underline{t})$ is not identically equal to 1;
- b) the quantities $\lim_{n \rightarrow \infty} \sigma_{ij}^2 n^{-1}$ are non-zero, since in § 2.7 it was shown that these limits must not only be finite, but also necessarily non-zero.

We have defined the moment generating function of the n_{ij} in § 2.8 as

$$M(\underline{t}) = E(\exp \sum_{i,j} t_{ij} n_{ij}) = E(\exp \sum_{i,j} t_{ij} \sum_r x_{ij}^{(r)}),$$

and seen that if $\underline{R}$ denotes the square matrix with elements $\{\underline{R}\}_{j1} = p_{ij} \exp t_{ij}$, this is given by equation (2.8.2) as

$$M(\underline{t}) = \underline{1}' \underline{R} \underline{P},$$

where $\underline{1}'$ denotes the row vector of unit elements, and $\underline{P}$ the column vector of the stationary probabilities P_i .

If $\underline{R}$ has the latent roots $\mu_1(\underline{t}), \dots, \mu_s(\underline{t})$, we have seen in § 2.9 that for small $\underline{t}$, $\mu_1(\underline{t})$ is the simple dominant root such that $\mu_1(0) = 1$. Following Frazer, Duncan and Collar (1947; 134-136), we have that Sylvester's theorem gives for $n \rightarrow \infty$,

$$\underline{R}^n \sim \{\mu_1(\underline{t})\}^n \underline{Z}_0(\mu_1),$$

where the matrix $Z_0(\mu_1)$ is finite. It follows that

$$M(\underline{t}) \sim \{\mu_1(\underline{t})\}^{-1} Z_0(\mu_1) \underline{P},$$

so that $\ln M(\underline{t}) \sim n \ln \mu_1(\underline{t})$, a non-zero value, since for a positively regular chain $\mu_1(\underline{t}) \neq 1$.

$M(\underline{t})$ is thus asymptotically equivalent to the moment generating function of a sum of n identical independent vector variates. Since for a positively regular chain, the n_{ij} have variances of order n , so that $\sigma_{ij}^2 n^{-1}$ tends to a finite and non-zero value, they will tend to simultaneous normality as $n \rightarrow \infty$ by the central limit theorem for $n = \sum_{ij} n_{ij}$ variates. It follows that B_0 will be asymptotically normally distributed.

Finally, a general convergence theorem due to Cramér (1946) enables us to conclude that $(T - \theta_0)\{i(\theta_0)n\}^{\frac{1}{2}}$ is asymptotically normally distributed with zero expectation and unit variance; this is that if a variate ξ_n with distribution function $f_n(x)$ tending to $f(x)$, and a variate η_n converging in probability to a positive constant c as $n \rightarrow \infty$ exist, then the distribution function of $\xi_n \eta_n^{-1}$ tends to $f(cx)$.

Lastly, since $\{n i(\theta_0)\}^{-1} = -\{\mathcal{E}(\{\frac{d^2 \mu}{d\theta^2}\}_\theta)\}^{-1}$ is the minimum possible variance, we see clearly that the

consistent solution for the maximum likelihood equation (3.1.5) is fully efficient.

Chapter 4

SUFFICIENCY CONDITIONS FOR A MARKOV CHAIN WITH A FIXED INITIAL STATE

1. Introduction

In the theory of maximum likelihood estimation, it is usual to prove a general theorem linking sufficient estimators with the solution of the likelihood equation; Cramér (1946; 500), under certain conditions, proves it for continuous distributions, and it applies equally with minor modifications to the case of the Markov chain. The theorem is that:

if a sufficient estimator T of the parameter θ defining the transition probabilities $p_{ij}(\theta)$ ($i, j = 1, \dots, s$) of a Markov chain exists, any solution of the likelihood equation will be a function of T .

In obtaining the maximum likelihood estimator of θ , it may therefore be of interest to determine the form of the transition probabilities which admit a sufficient

statistic.

Koopman (1936), and Pitman (1936) have given the general form of continuous distributions admitting a sufficient statistic, but although it has been obvious that a similar form exists for discrete distributions, no account of the proof for this appears to be available. We briefly outline the proof, use it to obtain the form of the probabilities $p_1(\theta), \dots, p_k(\theta)$ in the particular case of the multinomial distribution, and finally generalise our results in the case of the simple Markov chain which is positively regular.

2. The form of the likelihood function admitting a sufficient estimator, for discrete parent distributions with constant variate intervals

We consider those discrete distributions $p(x, \theta)$ for which the variate x is defined at equal intervals. We can then, without loss of generality, assume the interval to be unity so that x takes consecutive integral values in any specific range, finite or infinite; this includes the standard discrete distributions for which x is a frequency, as well as others for which x may take positive or negative integral values.

Let $x_1, \dots, x_n$ be n independent observations of a variate x with the discrete distribution $p(x, \theta)$, where x takes a set of consecutive integral values in a given range. The probability of this sample is

$$\varphi(x_1, \dots, x_n; \theta) = p(x_1, \theta) \dots p(x_n, \theta).$$

If this is to admit a sufficient estimator T of θ , then the following factorisability condition must hold:

$$\varphi(x_1, \dots, x_n; \theta) = f(x_1, \dots, x_n) F(T, \theta),$$

where $f(x_1, \dots, x_n)$, a conditional distribution function given T , is independent of θ ; the likelihood function $L = \ln \varphi$ is given by

$$(4.2.1) \quad L = \ln f(x_1, \dots, x_n) + \ln F(T, \theta).$$

Assuming that the functions $p(x_1, \theta)$, $F(T, \theta)$ are differentiable with respect to θ , we obtain the equation

$$(4.2.2) \quad \sum_i \frac{\partial}{\partial \theta} \{ \ln p(x_i, \theta) \} = \frac{\partial}{\partial \theta} \{ \ln F(T, \theta) \} = G(T, \theta).$$

Since this holds for all values of θ within the range allowed, a particular value θ_0 lying in this range, will give when substituted in this equation

$$(4.2.3) \quad u = \sum_i u(x_i) = g(T)$$

which connects T with the statistic

$$u = \sum_i u(x_i) = \sum_i \left[\frac{\partial}{\partial \theta} \{ \ln p(x_i, \theta) \} \right]_{\theta_0}.$$

Now since sufficiency is a property which holds for all values of the sample $x_1, \dots, x_n$, we may allow x_i , a particular discrete sample value, to increase by 1, while the remaining $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$, remain unchanged. We can then difference equation (4.2.3) with respect to x_i to obtain

$$\Delta_{x_i} u = \Delta_{x_i} u(x_i) = \Delta_{x_i} g(T).$$

Similarly for equation (4.2.2)

$$\Delta_{x_i} \frac{\partial}{\partial \theta} \{ \ln p(x_i, \theta) \} = \Delta_{x_i} G(T, \theta),$$

so that for all values of i , we have the equation

$$(4.2.4) \quad \frac{\Delta_{x_i} \frac{\partial}{\partial \theta} \{ \ln p(x_i, \theta) \}}{\Delta_{x_i} u(x_i)} = \frac{\Delta_{x_i} G(T, \theta)}{\Delta_{x_i} g(T)}.$$

Now if we assume the condition that $G(T, \theta)$ and $g(T)$ be differentiable with respect to T , we have that

$$\begin{aligned} \Delta_{x_i} G(T, \theta) &= G(T(x_1, \dots, x_{i-1}, x_i+1, x_{i+1}, \dots, x_n), \theta) - G(T(x_1, \dots, x_i, \dots, x_n), \theta) \\ &= G(T_1, \theta) - G(T_0, \theta) \\ &= (T_1 - T_0) \left\{ \frac{\partial G}{\partial T} \right\}_{T_2} \\ &= (\Delta_{x_i} T) \left\{ \frac{\partial G}{\partial T} \right\}_{T_2}, \end{aligned}$$

where T_2 is some function $T_2(x_1, \dots, x_{i-1}, x_i, \dots, x_n; \theta)$ of the sample values and of θ such that $T_0 \leq T_2 \leq T_1$.

Further, since the estimator T is a symmetrical function of the x_i , the function T_2 will be the same in the case of every differencing with respect to each x_i ($i = 1, \dots, n$). Similarly,

$$\Delta_{x_i} g(T) = (\Delta_{x_i} T) \left\{ \frac{dg}{dT} \right\}_{T_2(x_1, \dots, x_i, \dots, x_n; \theta_0)}$$

where $T_0 \leq T_2(x_1, \dots, x_i, \dots, x_n; \theta_0) \leq T_1$. It follows that equation (4.2.4) can be written for all values of i as

$$\frac{\Delta_{x_i} \frac{\partial}{\partial \theta} \{ \ln p(x_i, \theta) \}}{\Delta_{x_i} u(x_i)} = \frac{\left\{ \frac{\partial G}{\partial T} \right\}_{T_2(x_1, \dots, x_n; \theta)}}{\left\{ \frac{dg}{dT} \right\}_{T_2(x_1, \dots, x_n; \theta_0)}},$$

which is equal to a fixed value, not depending on i .

Hence

$$\frac{\Delta_{x_1} \frac{\partial}{\partial \theta} \{ \ln p(x_1, \theta) \}}{\Delta_{x_1} u(x_1)} = \frac{\Delta_{x_2} \frac{\partial}{\partial \theta} \{ \ln p(x_2, \theta) \}}{\Delta_{x_2} u(x_2)} = \dots = \frac{\Delta_{x_n} \frac{\partial}{\partial \theta} \{ \ln p(x_n, \theta) \}}{\Delta_{x_n} u(x_n)},$$

and this can only be so if these are all equal to a function of θ only, $K_1(\theta)$. It follows from (4.2.4) that

$$\Delta_{x_i} G(T, \theta) = K_1(\theta) \Delta_{x_i} g(T),$$

so that

$$G(T, \theta) = K_1(\theta) g(T) + K_2(\theta).$$

On integrating this with respect to θ , we obtain

$$\ln F(T, \theta) = \Lambda_1(\theta) g(T) + \Lambda_2(\theta) + h(T),$$

so that the likelihood (4.2.1) is of the form

$$(4.2.5) \quad L = \ln f(x_1, \dots, x_n) + \Lambda_1(\theta) g(T) + \Lambda_2(\theta),$$

and the probability of the sample is of the form

$$\varphi(x_1, \dots, x_n; \theta) = f(x_1, \dots, x_n) e^{\Lambda_1(\theta) g(T) + \Lambda_2(\theta)}.$$

3. Form of the probabilities $p_i(\theta)$ for a multinomial distribution admitting a sufficient estimator T of θ .

Consider the k mutually exclusive events $E_1, \dots, E_k$ with non-zero probabilities $p_1(\theta), \dots, p_k(\theta)$ such that $\sum_{i=1}^k p_i(\theta) = 1$. In the event of a total of n independent trials in which E_1 occurs x_1 times, ..., and E_k occurs x_k times, the probability of the sample of frequencies $x_1, \dots, x_k$ is

$$\varphi(x_1, \dots, x_k; \theta) = p_1^{x_1}(\theta) \dots p_k^{x_k}(\theta).$$

For this to admit a sufficient statistic, the likelihood $L = \ln \varphi$ must be of the form (4.2.5) so that

$$L = \sum_{i=1}^k x_i \ln p_i(\theta) = \ln f(x_1, \dots, x_k) + \Lambda_1(\theta)g(T) + \Lambda_2(\theta).$$

We note in this case, that $\Lambda_2(\theta)$ involves the number of trials $n = \sum_i x_i$, so that it can be written

$$\Lambda_2(\theta) = \sum_{i=1}^k x_i \cdot \lambda_2(\theta),$$

where $\lambda_2(\theta)$ is a function of θ only. Also from the form of the likelihood, we see that the first term can only be of the form

$$\ln f(x_1, \dots, x_k) = \sum_{i=1}^k A_i x_i.$$

We can therefore write the likelihood function admitting a sufficient statistic T of θ as

$$(4.3.1) \quad L = \sum_i x_i \ln p_i(\theta) = \sum_i A_i x_i + \Lambda_1(\theta)g(T) + \lambda_2(\theta) \sum_i x_i,$$

which holds for all values of n , and of the frequency values $x_1, \dots, x_n$. If, in a total of $n+1$ independent trials, the frequency of the event E_1 was x_1+1 , while the frequencies of the remaining events E_r ($r \neq 1$) was unchanged, then on differencing equation (4.3.1) with respect to x_1 , we obtain

$$\ln p_i(\theta) = A_i + \Lambda_1(\theta) \Delta_{x_i} g(T) + \lambda_2(\theta),$$

or

$$\{\Lambda_1(\theta)\}^{-1} \{ \ln p_i(\theta) - A_i - \lambda_2(\theta) \} = \Delta_{x_i} g(T).$$

That is, a function of θ only, equals a function of the sample values x_1 , which is possible only if both equal a constant K_1 .

It follows that

$$\begin{aligned} p_i(\theta) &= e^{K_i \Lambda_1(\theta) + \lambda_2(\theta) + A_i} \\ (4.3.2) \quad &= \alpha_i e^{K_i \Lambda_1(\theta) + \lambda_2(\theta)} \quad (i = 1, \dots, k), \end{aligned}$$

where $A_1 = \ln \alpha_1$, is the form of the probability allowing a sufficient estimator T of θ . A condition to which the probabilities $p_i(\theta)$ are subject, is

$$\sum_{i=1}^k p_i(\theta) = 1, \text{ so that in our case, from (4.3.2),}$$

$$\sum_i \alpha_i e^{K_i \Lambda_1(\theta) + \lambda_2(\theta)} = 1,$$

or

$$\sum_i \alpha_i e^{K_i \Lambda_1(\theta)} = e^{-\lambda_2(\theta)}.$$

An alternative way of writing the probabilities $p_i(\theta)$ is therefore

$$\begin{aligned} p_i(\theta) &= \alpha_i e^{K_i \Lambda_1(\theta)} \left\{ \sum_i \alpha_i e^{K_i \Lambda_1(\theta)} \right\}^{-1} \\ (4.3.3) \quad &= \alpha_i \lambda_1^{K_i}(\theta) \left\{ \sum_i \alpha_i \lambda_1^{K_i}(\theta) \right\}^{-1}, \end{aligned}$$

where $\lambda_1(\theta) = \exp \Lambda_1(\theta)$. We can now tell immediately in any practical case whether it is possible to obtain

a sufficient statistic or not, by an examination of the form of the multinomial probabilities. Three simple illustrations from trinomial distributions are given as examples.

Example (4.3.a) The trinomial distribution with probabilities

$$p_1 = \theta(\theta + \theta^2 + 2\theta^3)^{-1}, \quad p_2 = \theta^2(\theta + \theta^2 + 2\theta^3)^{-1}, \quad p_3 = 2\theta^3(\theta + \theta^2 + 2\theta^3)^{-1},$$

of the form (4.3.3) with α_i equal to 1, 1, 2 respectively, and K_i equal to 1, 2, 3, where $\lambda_1(\theta) = \theta$, will give an estimator of θ which is sufficient.

Example (4.3.b) Probabilities of the type

$$p_1 = \theta, \quad p_2 = 2\theta, \quad p_3 = 1 - 3\theta,$$

which can be written in the form (4.3.2)

$$p_i(\theta) = \alpha_i e^{K_i \ln \theta (1-3\theta)^{-1} + \ln (1-3\theta)},$$

where α_i equal 1, 2, 1, and K_i equal 1, 1, 0, respectively will also give an estimator of θ which is sufficient.

Example (4.3.c) However, probabilities of the form

$$p_1 = \frac{1}{4}(2+\theta), \quad p_2 = \frac{1}{2}(1-\theta), \quad p_3 = \frac{1}{4}\theta,$$

which occur in genetics, cannot be written in the forms

(4.3.2) or (4.3.3), and can be verified to admit no sufficient estimator T of θ .

4. Form of the transition probabilities $p_{ij}(\theta)$
for a simple positively regular Markov chain
admitting a sufficient estimator T of θ

Consider the realisation of the simple positively regular Markov chain with s possible states $E_1, \dots, E_s$, which results in the observed sequence S of $n + 1$ states in the following order

$$E_1, E_j, \dots, E_k, E_\ell.$$

Assuming the initial state E_1 to be fixed, the probability distribution of the sequence S is given by

$$\varphi(S) = \prod_{i,j=1}^s p_{ij}^{n_{ij}},$$

and the likelihood can be written as

$$L = \sum_{i,j} n_{ij} \ln p_{ij}.$$

It is important to note that we are specifically considering the case of sufficiency for a distribution conditional on the given initial state E_1 . In the general case where any state E_1 may occur initially,

the likelihood is

$$L = \sum_{i=1}^s x_i \ln a_i + \sum_{i,j} n_{ij} \ln p_{ij},$$

where the variates x_i are 0 or 1 according to whether the state E_i is or is not the initial state. In addition to the usual conditions for the n_{ij} , the x_i are subject to the condition that $\sum_{i=1}^s x_i = 1$. This more difficult case will not, for the moment, be considered.

A positively regular chain (§ 2.9; 47) is such that for any row i of the stochastic matrix, and for all values of j , $0 \leq p_{ij}(\theta) < 1$, so that, in particular, for $i = j$ no state E_i is an absorption state; some, though not all $p_{ij}(\theta)$ in a row may equal zero, subject to the usual condition that $\sum_{j=1}^s p_{ij}(\theta) = 1$ ($i = 1, \dots, s$).

Now if the chain admits a sufficient estimator T of θ , we see that the likelihood may be written, much in the same way as for the multinomial case,

$$(4.4.1) \quad L = \sum_{i,j} n_{ij} \ln p_{ij} = \sum_{i,j} A_{ij} n_{ij} + \Lambda_1(\theta) g(T) + \Lambda_2(\theta) \sum_{i,j} n_{ij},$$

where the A_{ij} are constants, and $\sum_{i,j} n_{ij} = n$.

Before proceeding, as in the similar case of the multinomial, to increase the number of trials in the

realisation of the chain by one, so as to allow this equation to be differenced with respect to the n_{ij} in order to obtain the non-zero values of the $p_{ij}(\theta)$, we must consider some difficulties which this method presents. Since to any zero transition probability $p_{ij}(\theta)$ there will necessarily correspond a zero value of the transition frequency n_{ij} , neither will appear in the likelihood function (4.4.1). We are concerned with the n_{ij} corresponding to non-zero values of $p_{ij}(\theta)$, and differencing occurs only with respect to these; however, owing to the set of relations given in § 2.7 for the sums of transition frequencies

$$\sum_{j=1}^s n_{ij} \doteq \sum_{i=1}^s n_{ri} \quad (i = 1, 2, \dots, s),$$

where the sign $\doteq$ indicates equality or a possible difference of one between the sums, it is not always possible in increasing the number of trials in a realisation of the chain by 1, from $n+1$ to $n+2$, to increase each of the n_{ij} associated with a non-zero transition probability $p_{ij}(\theta)$ singly by 1, while leaving the values of the remaining n_{ij} unchanged.

As a simple example to illustrate this point, we consider the case of a chain with two states E_1 , and E_2 , for which a possible realisation of three trials,

starting with the fixed state E_1 results in the sequence E_1, E_1, E_2 . Assuming all the transition probabilities to have non-zero values, we see immediately that the transition frequency matrix with elements n_{ij} is

$$\underline{n} = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} ;$$

a realisation of four trials starting with E_1 will give among others, the following transition frequency matrices

$$\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

for the sequences E_1, E_1, E_1, E_2 ; E_1, E_1, E_2, E_1 ; and E_1, E_1, E_2, E_2 . Here we see that n_{11} , n_{21} , n_{22} have singly been increased by 1, while the remaining n_{ij} remain constant; it is, however, in no way possible starting with the given matrix $\underline{n}$, to increase n_{12} by 1, while leaving n_{11} , n_{21} , n_{22} unchanged.

In order to resolve this difficulty, we use the fact that for a positively regular chain, providing the number of trials $n+1$ is sufficiently large in a realisation starting with the given state E_1 , it is

possible to end with any one of the states $E_1, \dots, E_s$. In our particular case, assuming a sufficiently large number of trials, the realisation of the chain leads to a sequence S which starts with the fixed state E_i and ends with the state E_e ; a further trial may result in the sequence S' with an additional state E_r , for all r for which $p_{er} \neq 0$. The probability of this sequence is

$$\varphi(S') = p_{ij}(\theta) \cdots p_{ke}(\theta) p_{er}(\theta),$$

and the likelihood function $L' = \ln \varphi(S')$ is identical with $L = \ln \varphi(S)$ except that the transition frequency n_{er} is increased to $n_{er} + 1$. This allows us to difference the equation (4.4.1) with respect to n_{er} for all r for which $p_{er} \neq 0$, in order to obtain

$$\ln p_{er}(\theta) = A_{er} + \Lambda_1(\theta) \Delta_{n_{er}} g(T) + \lambda_2(\theta),$$

where $\Delta_{n_{er}} g(T)$ is some function of the transition frequencies n_{ij} . Now since

$$\{\Lambda_1(\theta)\}^{-1} \{\ln p_{er} - \lambda_2(\theta) - A_{er}\} = \Delta_{n_{er}} g(T),$$

that is, a function of θ equals a function of the n_{ij} , then both must equal a constant K_{er} . This allows us to write for all r for which $p_{er} \neq 0$,

$$p_{\ell r}(\theta) = \alpha_{\ell r} e^{K_{\ell r} \Lambda_1(\theta) + \lambda_2(\theta)},$$

where $A_{\ell r} = \ln \alpha_{\ell r}$. If we further admit zero values of $\alpha_{\ell r}$, we may accept this equation as the general form for both zero and non-zero values of $p_{\ell r}$ ($r=1, \dots, s$). Since, for a sufficiently large number of trials in a realisation of the chain starting with the given state E_1 , the final state E_ℓ may be such that ℓ can take any of the values $1, 2, \dots, s$, it follows that we may write as the general form of the transition probabilities $p_{ij}(\theta)$ admitting a sufficient estimator T of θ , the equation

$$(4.4.2) \quad p_{ij}(\theta) = \alpha_{ij} e^{K_{ij} \Lambda_1(\theta) + \lambda_2(\theta)}.$$

This general form of the $p_{ij}(\theta)$ has been obtained for sufficiently large values of $n = \sum_{i,j} n_{ij}$; this means that in this case, the necessary condition that a sufficient estimator of θ exist is that the $p_{ij}(\theta)$ be of form (4.4.2). It is easily verified that the sufficient condition also follows: that if the $p_{ij}(\theta)$ are of the form (4.4.2), the estimator of θ is sufficient. This, however, holds generally, irrespective of the size of n . It is of interest to note that if we restrict ourselves to the case where all the $p_{ij}(\theta)$ are non-zero,

all possible values of $n \geq 1$ are sufficiently large for the necessary condition to hold, since any final state E_e may be reached in a realisation of one or more trials starting with the given state E_1 . This means that for the positively regular chain with non-zero $p_{ij}(\theta)$, a sufficient estimator of θ exists for any number of trials $n \geq 1$, if and only if the $p_{ij}(\theta)$ are of the form (4.4.2).

The transition probabilities (4.4.2) are subject to the condition that $\sum_{j=1}^s p_{ij} = 1$, so that

$$\sum_j \alpha_{ij} e^{K_{ij} \Lambda_1(\theta)} = e^{-\lambda_i(\theta)},$$

or

$$(4.4.3) \quad \sum_j \alpha_{ij} e^{K_{ij} \Lambda_1(\theta)} = \sum_j \alpha_{rj} e^{K_{rj} \Lambda_1(\theta)} \quad (i, r = 1, \dots, s).$$

If we write $\lambda_1(\theta) = \exp \Lambda_1(\theta)$, we see that the $p_{ij}(\theta)$ can also be put in the form

$$(4.4.4) \quad p_{ij}(\theta) = \alpha_{ij} \lambda_1^{K_{ij}}(\theta) \left\{ \sum_j \alpha_{ij} \lambda_1^{K_{ij}}(\theta) \right\}^{-1}$$

where $\left\{ \sum_j \alpha_{ij} \lambda_1^{K_{ij}}(\theta) \right\}^{-1}$ has the same value for all rows $i=1, \dots, s$.

Two possible cases now exist for those exponents among $K_{11}, K_{12}, \dots, K_{1s}$ associated with the non-zero probabilities among $p_{11}, p_{12}, \dots, p_{1s}$, in the i -th row

of the transition probability matrix $\underline{p}$: the first in which all such exponents are distinct and the second in which only some such exponents are distinct. For simplicity, we assume that in the i -th row, all transition probabilities $p_{ij}(\theta)$ ($j=1, \dots, s$) are non-zero; the slightly more general case where some $p_{ij}(\theta)$ may be zero follows without difficulty.

Consider the first case, where for the particular row i , there are s exponents K_{ij} , all distinct, associated with the s non-zero transition probabilities $p_{ij}(\theta)$ ($j=1, \dots, s$); then from equation (4.4.3), it follows that for any other row r ,

$$(4.4.5) \quad \sum_j \alpha_{ij} \lambda_1^{K_{ij}} = \sum_j \alpha_{rj} \lambda_1^{K_{rj}}.$$

This is possible only if the exponents K_{ij} and K_{rj} ($j=1, \dots, s$) have the same s distinct values, though these may be arranged in different orders. This means that there is only a single set of s distinct values

$$\alpha_{i1} \lambda_1^{K_{i1}} \left\{ \sum_j \alpha_{ij} \lambda_1^{K_{ij}} \right\}^{-1}, \dots, \alpha_{is} \lambda_1^{K_{is}} \left\{ \sum_j \alpha_{ij} \lambda_1^{K_{ij}} \right\}^{-1},$$

for the transition probabilities of the matrix $\underline{p}$, and these must appear, possibly in different orders, in every row of the matrix.

Two simple examples for Markov chains with two and

three states respectively, will illustrate the previous points.

Example (4.4.a) The transition probability matrix of the form

$$\underline{p} = \begin{pmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{pmatrix}$$

will always provide a sufficient estimator of θ , for we may write $p_{1j}(\theta)$ ($j=1,2$) in the form (4.4.2) as

$$p_{1j} = \alpha_{1j} e^{K_{1j} \ln\{\theta(1-\theta)^{-1}\} + \ln(1-\theta)},$$

where the values of α_{11} , α_{12} and K_{11} , K_{12} are 1, 1, and 1, 0 respectively. The second row consists of the same transition probabilities in a different order.

Example (4.4.b) The stochastic matrix

$$\underline{p} = \begin{pmatrix} 2\theta(2\theta+\theta^2+\theta^3)^{-1} & \theta^2(2\theta+\theta^2+\theta^3)^{-1} & \theta^3(2\theta+\theta^2+\theta^3)^{-1} \\ \theta^3(2\theta+\theta^2+\theta^3)^{-1} & 2\theta(2\theta+\theta^2+\theta^3)^{-1} & \theta^2(2\theta+\theta^2+\theta^3)^{-1} \\ \theta^2(2\theta+\theta^2+\theta^3)^{-1} & \theta^3(2\theta+\theta^2+\theta^3)^{-1} & 2\theta(2\theta+\theta^2+\theta^3)^{-1} \end{pmatrix}$$

also provides a sufficient estimator of θ , for in (4.4.4), we have $\lambda_1(\theta) = \theta$, and the values of α_{1j} and K_{1j} are 2, 1, 1 and 1, 2, 3 respectively. The second and third rows consist of the same values for the transition probabilities, but in different orders.

following $s-k+1$ forms

$$(4.4.7) \quad \lambda_1^k \left\{ \sum_j \alpha_{ij} \lambda_1^{k_{ij}} \right\}^{-1}, \lambda_1^{k_{i,k+1}} \left\{ \sum_j \alpha_{ij} \lambda_1^{k_{ij}} \right\}^{-1}, \dots, \lambda_1^{k_{is}} \left\{ \sum_j \alpha_{ij} \lambda_1^{k_{ij}} \right\}^{-1},$$

appearing in various arrangements. The s coefficients associated with these $s-k+1$ distinct forms for the transition probabilities, may also differ from row to row, but are subject to the condition which follows from (4.4.5) that for all rows ($r=1, \dots, s$), the sums of the coefficients for each distinct form in (4.4.7) must equal the fixed values

$$(4.4.8) \quad \sum_{j=1}^k \alpha_{ij}, \alpha_{i,k+1}, \dots, \alpha_{is},$$

respectively.

Two simple examples of this case for Markov chains with three states follow.

Example (4.4.c)

$$\underline{p} = \begin{pmatrix} 2\theta(3\theta+\theta^3)^{-1} & \theta(3\theta+\theta^3)^{-1} & \theta^3(3\theta+\theta^3)^{-1} \\ \frac{1}{3}\theta^3(3\theta+\theta^3)^{-1} & \frac{2}{3}\theta^3(3\theta+\theta^3)^{-1} & 3\theta(3\theta+\theta^3)^{-1} \\ \frac{11}{4}\theta(3\theta+\theta^3)^{-1} & \theta^3(3\theta+\theta^3)^{-1} & \frac{1}{4}\theta(3\theta+\theta^3)^{-1} \end{pmatrix}$$

where in equation (4.4.6), $\lambda_1(\theta) = \theta$, and the two distinct forms for the transition probabilities appearing in each row are

$$\theta(3\theta+\theta^3)^{-1}, \quad \theta^3(3\theta+\theta^3)^{-1}.$$

The values of the coefficients for the form $\theta(3\theta + \theta^3)^{-1}$ vary from row to row, subject to the condition (4.4.8) that their sum is always equal to 3, while those for the form $\theta^3(3\theta + \theta^3)^{-1}$ have a sum always equal to 1.

Example (4.4.d)

$$\underline{p} = \begin{pmatrix} \frac{1}{2}\theta & 1-\theta & \frac{1}{2}\theta \\ \frac{1}{4}\theta & \frac{3}{4}\theta & 1-\theta \\ \frac{1}{2}(1-\theta) & \frac{1}{2}(1-\theta) & \theta \end{pmatrix}$$

where in equation (4.4.6), $\lambda_1(\theta) = \theta(1-\theta)^{-1}$, and the two distinct forms for the transition probabilities appearing in each row are

$$\{\theta(1-\theta)^{-1}\}^{K_{ij}} (1-\theta) \quad (K_{ij} = 0, 1).$$

Again the values of the coefficients for θ vary from row to row, subject to the condition (4.4.8) that their sum is always equal to 1, while those for $(1-\theta)$ also have a sum equal to 1.

Chapter 5

SUFFICIENCY CONDITIONS FOR A MARKOV CHAIN
WITH AN UNRESTRICTED INITIAL STATEI. Introduction1. The general problem

In the last chapter on positively regular Markov chains with s possible states $E_1, \dots, E_s$, we have considered the form of stochastic matrices $\underline{p}$ with elements $p_{ij}(\theta)$ admitting a sufficient estimator of θ when the initial state in a realisation of the chain is fixed.

For a realisation of this chain resulting in the sequence S of $n + 1$ states in the following order

$$(5.1.1) \quad E_1, E_j, \dots, E_k, E_\ell,$$

it was seen that the probability of this sequence was

$$\varphi(S) = \prod_{i,j=1}^s p_{ij}^{n_{ij}}(\theta),$$

where the n_{ij} are transition frequencies from state E_i to state E_j , such that $\sum_{j=1}^s n_{ij} = n$, and its likelihood was clearly

$$(5.1.2) \quad L = \sum_{i,j=1}^s n_{ij} \ln p_{ij}(\theta).$$

We restrict ourselves throughout this chapter to stochastic matrices $\underline{p}$ for which all the transition probabilities $p_{ij}(\theta)$ are non-zero. For these, it has been proved that, given an initial state E_i , the necessary and sufficient condition for all values of $n \geq 1$, that a sufficient estimator of θ exist is that the transition probabilities in any row i of the stochastic matrix $\underline{p}$ be of the form

$$(5.1.3) \quad p_{ij}(\theta) = \alpha_{ij} e^{K_{ij}\Lambda_1(\theta) + \lambda_2(\theta)} \quad (j=1, \dots, s),$$

where the constants $\alpha_{ij} > 0$, and the number of distinct K_{ij} in the row is less than or equal to s . We have seen that since $\sum_{j=1}^s p_{ij} = 1$, the functions $\Lambda_1(\theta)$ and $\lambda_2(\theta)$ are related by the equation

$$(5.1.4) \quad e^{-\lambda_2(\theta)} = \sum_{j=1}^s \alpha_{ij} e^{K_{ij}\Lambda_1(\theta)}.$$

If the values of the K_{ij} in a row i of the matrix $\underline{p}$ are all distinct, the transition probabilities in the remaining rows of the matrix have the same s distinct

values as in row i , possibly in different orders; if, however, the K_{ij} in a row i have only $r < s$ distinct values, these determine, apart from coefficients, the r distinct forms of the transition probabilities which appear in some arrangement in each row of the matrix $\underline{p}$, subject to the condition that the sums of the coefficients for each distinct form are equal in every row.

Suppose now that a realisation of the chain occurs for which the initial state is not fixed, and let this result in a particular case in the sequence S' of the same form as (5.1.1). For a chain known to be initially stationary, the probabilities P_i ($i=1, \dots, s$) that the realisation begin with the state E_i are given by the matrix equation

$$(5.1.5) \quad \underline{p}' \underline{P} = \underline{P}$$

together with the condition $\sum_{i=1}^s P_i = 1$, where $\underline{P}$ is the column vector of elements P_i . It is clear that these stationary probabilities P_i will in general be functions of θ depending on the form of the $p_{ij}(\theta)$. The likelihood of the sequence S' , for which the initial state E_i is unrestricted, will be given by

$$(5.1.6) \quad L = \sum_{i=1}^s x_i \ln P_i(\theta) + \sum_{i,j=1}^s n_{ij} \ln p_{ij}(\theta),$$

where the x_i are 1 or 0 according to whether E_i is or is not the initial state, and are subject to the condition $\sum_{i=1}^s x_i = 1$.

The study of the general form of the transition probabilities $p_{ij}(\theta)$, such that these and the associated stationary probabilities $P_i(\theta)$ obtained as a solution of (5.1.5) will admit a sufficient estimator of θ in the case where the initial state of a realisation is not fixed, presents serious difficulties. We may, however, more easily consider the restricted problem of obtaining from among the stochastic matrices $\underline{p}$ with non-zero transition probabilities $p_{ij}(\theta)$ already admitting a sufficient estimator of θ when the initial state of a realisation is fixed, that subclass which will also admit a sufficient estimator of θ when the initial state is unrestricted.

2. The Markov chain with two states

For the positively regular Markov chain with the two states E_1, E_2 , the stochastic matrix $\underline{p}$ admitting a sufficient estimator of θ for a fixed initial state can have only two possible forms; these are reducible to the types

$$(a) \quad \underline{p} = \begin{pmatrix} p & q \\ p & q \end{pmatrix},$$

where $p = p(\theta)$ is some function of θ , and $q = 1 - p$, and for which the states E_1, E_2 are independent, or alternatively

$$(b) \quad \underline{p} = \begin{pmatrix} q & p \\ p & q \end{pmatrix},$$

where the matrix $\underline{p}$ is doubly stochastic. In both these forms, the transition probabilities $p_{ij}(\theta)$ satisfy the equation (5.1.3), since for any row i , these can be written as

$$p_{ij}(\theta) = e^{\{K_{ij} \ln p q^{-1} + \ln q\}} \quad (K_{ij} = 0, 1),$$

and these two values for the two distinct values of the K_{ij} , appear in each row of the matrix (a) or (b).

For the form (a), the stationary probabilities are clearly $P_1 = p$, $P_2 = q$, which are identical with the values p_{i1}, p_{i2} ($i = 1, 2$) of the matrix $\underline{p}$. This is the binomial case, which is known to admit a sufficient estimator of p , and so also of θ , for both fixed and unrestricted initial states.

For the form (b), in which the matrix $\underline{p}$ is doubly stochastic, it is known (Feller, 1950; 327) and easily

seen that the stationary probabilities P_1, P_2 , which are given by (5.1.5)

$$(p' - \underline{I}) \underline{P} = \begin{pmatrix} -p & p \\ p & -p \end{pmatrix} \underline{P} = 0 ,$$

together with the condition $P_1 + P_2 = 1$, are each equal to $\frac{1}{2}$. The likelihood (5.1.6) for a realisation of the chain in $n+1$ trials for which the initial state is unrestricted, can be written

$$L = (x_1 + x_2) \ln \frac{1}{2} + (n_{12} + n_{21}) \ln p + (n_{11} + n_{22}) \ln q ,$$

where x_i ($i=1,2$) is 1 or 0 according to whether E_i does or does not occur at the first trial, and $x_1 + x_2 = 1$, and where the n_{ij} ($i,j=1,2$) are the transition frequencies from state E_i to state E_j such that $n_{11} + n_{12} + n_{21} + n_{22} = n$. This will clearly provide a sufficient estimator of p , and so also of θ , which in fact has the same value as that obtained in the case where the initial state is fixed.

We have, therefore, that for the two-state case, the transition probabilities $p_{ij}(\theta)$ of the form (5.1.3) admitting a sufficient estimator of θ when the initial state is fixed, will also admit a sufficient estimator of θ when the initial state is unrestricted.

This simple case, though in itself trivial, is suggestive of similar but more general results in the case of the Markov chain with more than two states.

II. The Markov Chain with s States

3. The general form of the stationary probabilities

In the general case where a positively regular chain consists of the s states $E_1, \dots, E_s$, we consider the stochastic matrix

$$(5.3.1) \quad \underline{P} = \begin{pmatrix} p_{11}(\theta) & \dots & p_{1s}(\theta) \\ \dots & \dots & \dots \\ p_{s1}(\theta) & \dots & p_{ss}(\theta) \end{pmatrix}$$

with non-zero elements $p_{ij}(\theta)$. Since a sufficient estimator of θ exists in the case where the initial state of a realisation is fixed, these transition probabilities will satisfy the condition (5.1.3); for the moment, no further assumptions will be made about the number of distinct K_{ij} in row i of (5.3.1).

We consider now the form of the stationary probabilities $P_i(\theta)$ for the chain with the stochastic



matrix (5.3.1); these will be given by the equation (5.1.5), which when written in full is

$$(5.3.2) \quad (\underline{p}' - \underline{I}) \underline{P} = \begin{pmatrix} p_{11}-1 & \cdots & p_{s1} \\ \cdots & \cdots & \cdots \\ p_{1s} & \cdots & p_{ss}-1 \end{pmatrix} \underline{P} = \begin{pmatrix} -(p_{12}+\cdots+p_{1s}) & \cdots & p_{s1} \\ \cdots & \cdots & \cdots \\ p_{1s} & \cdots & -(p_{s1}+\cdots+p_{s,s-1}) \end{pmatrix} \underline{P} = 0$$

In addition, the stationary probabilities P_i satisfy the condition $\sum_i P_i = 1$.

Following exactly the account given in § 2.6, 33-35, earlier in this thesis, we have that the stationary probabilities P_i can be written as

$$(5.3.3) \quad P_i = \Delta_i \left\{ \sum_{i=1}^s \Delta_i \right\}^{-1} \quad (i=1, \dots, s),$$

where the Δ_i are cofactors of the i -th diagonal elements of the matrix $\underline{p}' - \underline{I}$ of (5.3.2).

To examine the general form of these stationary probabilities $P_i(\theta)$, we may first consider $P_1(\theta)$; the cofactor Δ_1 is

$$\Delta_1 = \begin{vmatrix} -(p_{21}+p_{23}+\cdots+p_{2s}) & p_{32} & \cdots & p_{s2} \\ p_{23} & -(p_{31}+\cdots+p_{3s}) & \cdots & p_{s3} \\ \cdots & \cdots & \cdots & \cdots \\ p_{2s} & \cdots & \cdots & -(p_{s1}+\cdots+p_{s,s-1}) \end{vmatrix}$$

which is a sum of terms, each of the form $p_{2j_2} p_{3j_3} \cdots p_{sj_s}$,

and all of which are positive or negative with $(-1)^{s-1}$. For, on expanding the determinant, it is clear that all off-diagonal terms other than the product of the diagonal elements $(-1)^{s-1}(p_{21}+p_{23}+\dots+p_{2s}) \dots (p_{s1}+\dots+p_{s,s-1})$ will, except possibly for their signs, consist of products contained in this diagonal term. Now, if the sign of an off-diagonal term is different from that of its counterpart in the diagonal, the two will cancel; if their signs are identical, the off-diagonal term will reinforce its counterpart in the diagonal. In both cases, the cofactor Δ_1 will be a sum of terms

$$\Delta_1 = \sum_{j_2, \dots, j_s} (-1)^{s-1} \alpha_{j_2, \dots, j_s} p_{2j_2} \dots p_{sj_s} \quad (j_2, \dots, j_s = 1, \dots, s),$$

where $\alpha_{j_2, \dots, j_s}$ may be zero, or some integer equal to or greater than 1.

If the $p_{ij}(\theta)$ are of the form (5.1.3), this is an expression of the form

$$\Delta_1 = (-1)^{s-1} \sum_t \beta_t^{(1)} e^{\{k_t^{(1)} \Lambda_1(\theta) + (s-1) \lambda_2(\theta)\}},$$

where all the coefficients $\beta_t^{(1)}$ are positive, and in which, depending on the ordering of the p_{ij} in the rows of the stochastic matrix $\underline{p}$, the terms are possibly not all distinct. Since the cofactors Δ_i ($i=1, 2, \dots, s$)

are similar, the stationary probabilities $P_i(\theta)$ will all be of the form

$$(5.3.4) \quad P_i(\theta) = \frac{\sum_t \beta_t^{(i)} e^{\{k_t^{(i)} \Lambda_1(\theta) + (s-1) \Lambda_2(\theta)\}}}{\sum_{i=1}^s \sum_t \beta_t^{(i)} e^{\{k_t^{(i)} \Lambda_1(\theta) + (s-1) \Lambda_2(\theta)\}}} = \frac{\sum_t \beta_t^{(i)} e^{k_t^{(i)} \Lambda_1(\theta)}}{\sum_{i=1}^s \sum_t \beta_t^{(i)} e^{k_t^{(i)} \Lambda_1(\theta)}},$$

and will in general be ratios of two irreducible polynomials in $\exp \Lambda_1(\theta)$, each a sum of positive terms only, and such that the numerator is of degree less than or equal to that of the denominator.

It is obvious that, in general, these values of $P_i(\theta)$ will not result in a likelihood function (5.1.6) admitting a sufficient estimator of θ . It may occur, however, that a reduction of the $P_i(\theta)$ can be carried out, and that of all possible reduced forms, we obtain the only two which, for an unrestricted initial state, will admit a sufficient estimator of θ when substituted in the likelihood function (5.1.6). These stationary probabilities are

$$(5.3.5) \quad P_i(\theta) = \alpha_i e^{K_i \Lambda_1(\theta) + \Lambda_2(\theta)},$$

which are identical in form to the transition probabilities (5.1.3), or

$$(5.3.6) \quad P_i(\theta) = \alpha_i,$$

which are constants. We shall consider only the two cases where the stationary probabilities are all of the form (5.3.5), or all of the form (5.3.6); these may well be the only two permissible cases, so that there exists no stochastic matrix resulting in a set of mixed stationary probabilities, some of form (5.3.5) and others of form (5.3.6). It has not as yet, however, been possible to prove or disprove this in the general case of the chain with s states.

4. Stochastic matrices with stationary

probabilities $P_i = \alpha_i e^{K_i \Lambda_1(\theta) + \lambda_2(\theta)}$

When the stationary probabilities are all of the form (5.3.5), it is possible for the s exponents K_i to be all distinct, or alternatively for only a number $r < s$ of them to be distinct. We prove that when the exponents K_i ($i=1, \dots, s$) are all distinct, the stochastic matrix $\underline{p}$ necessarily corresponds to the case where the states $E_1, \dots, E_s$ of the chain are independent. For, in this case, we see from (5.1.4), that the equations $\sum_{i=1}^s P_i = \sum_{j=1}^s p_{ij} = 1$, will result in

$$(5.4.1) \quad e^{-\lambda_2} = \sum_{i=1}^s \alpha_i e^{K_i \Lambda_1} = \sum_{j=1}^s \alpha_{ij} e^{K_{ij} \Lambda_1},$$

so that in addition to the fact that the s distinct values for the K_i are identical to the s distinct values for the K_{ij} of the transition probabilities $p_{ij}(\theta)$, we have that the coefficients α_i must also equal the α_{ij} . It follows that the rows of the stochastic matrix consist of the s distinct values of the P_i in some order; this order is shown to be unique so that

$$(5.4.2) \quad \underline{p} = \begin{pmatrix} P_1 & P_2 & \dots & P_s \\ P_1 & P_2 & \dots & P_s \\ \vdots & \vdots & \ddots & \vdots \\ P_1 & P_2 & \dots & P_s \end{pmatrix},$$

for which the states are independent.

Suppose that $\underline{p}$ is the stochastic matrix (5.3.1), where it is understood that each row consists of the values $P_1, \dots, P_s$ in some order. For some suitable value of θ , let the stationary probabilities P_i be arranged in an ascending order of magnitude; we may without loss of generality, assume that for $\theta = \theta_0$, $P_1 < P_2 < \dots < P_s$. Then from (5.1.5), we have that

$$(5.4.3) \quad \sum_{i=1}^s p_{ij} P_i = P_j \quad (j=1, \dots, s),$$

so that, considering the equation for $j=1$, we have

$$p_{11}P_1 + p_{21}P_2 + \dots + p_{s1}P_s = P_1.$$

Now, since for all i, j , we know that $p_{ij} \geq P_1$, then

$$p_{11}P_1 + p_{21}P_2 + \dots + p_{s1}P_s \geq P_1(P_1 + \dots + P_s) \geq P_1,$$

the required equality holding only if $p_{11} = P_1$ for all i . By a repetition of the same method, it can be shown that any i -th row of the transpose matrix $\underline{p}'$ must consist of elements P_1 only, so that $\underline{p}$ must be of the form (5.4.2).

The equations (5.4.3) can also be written, if the values $p_{ij} = \alpha_{ij} e^{K_{ij} \Lambda_1} \left\{ \sum_i \alpha_i e^{K_i \Lambda_1} \right\}^{-1}$, and $P_1 = \alpha_1 e^{K_1 \Lambda_1} \left\{ \sum_i \alpha_i e^{K_i \Lambda_1} \right\}^{-1}$ are substituted in it, as

$$(5.4.4) \quad \sum_{i=1}^s \alpha_i e^{K_i \Lambda_1} \{ \alpha_{ij} e^{(K_{ij} - K_j) \Lambda_1} - \alpha_j \} = 0 \quad (j = 1, \dots, s).$$

We have proved that the only possible solution of these equations is given by $K_{ij} = K_1$, and $\alpha_{ij} = \alpha_1$, a result not directly obvious from a consideration of (5.4.4), and which will prove to be useful in what follows.

We now consider the somewhat different case when the exponents K_i of the stationary probabilities P_i are not all distinct. Suppose there are $r < s$ distinct values for the K_i , then from equation (5.4.1) we see that the set of r distinct K_i is exactly the same as the set of distinct K_{ij} for the transition probabilities p_{ij} . Let us assume that $P_1, \dots, P_u$ have the same

exponent $K_1 = \dots = K_u = K^{(1)}$; . . . ; and $P_v, \dots, P_s$ have the same exponent $K_v = \dots = K_s = K^{(r)}$. In this case the equations (5.4.4) for $j=1, \dots, s$, may be written as

$$(5.4.5) \quad e^{K^{(1)}\lambda_1} \left\{ \sum_{i=1}^u \alpha_i (\alpha_{ij} e^{(K_{ij}-K_j)\lambda_1} - \alpha_j) \right\} + \dots + e^{K^{(r)}\lambda_1} \left\{ \sum_{i=v}^s \alpha_i (\alpha_{ij} e^{(K_{ij}-K_j)\lambda_1} - \alpha_j) \right\} = 0,$$

which, by analogy with (5.4.4) can be seen to have at least one solution given by $K_{ij} = K_j$ ($i=1, \dots, s$). It has not, however, been possible to prove the uniqueness of this solution for the general case of the chain with s states, although this may well be true. For the solution $K_{ij} = K_j$, we require that

$$(5.4.6) \quad \sum_{i=1}^u \alpha_i (\alpha_{ij} - \alpha_j) = \dots = \sum_{i=v}^s \alpha_i (\alpha_{ij} - \alpha_j) = 0;$$

this is certainly satisfied if all $\alpha_{ij} = \alpha_j$ ($i=1, \dots, s$), which is again the case of the independent states, but it may equally well be satisfied by other positive values of the α_{ij} .

To illustrate this clearly in a simple case, we consider the Markov chain with four states, for which the stationary probabilities have two distinct exponents only, and are written as

$$P_1 = \alpha p, \quad P_2 = (1-\alpha)p, \quad P_3 = \beta q = \beta(1-p), \\ P_4 = (1-\beta)q \quad (0 < \alpha, \beta < 1).$$

It is clear that the matrix

$$\underline{p} = \begin{pmatrix} \alpha p & (1-\alpha)p & \beta q & (1-\beta)q \\ \alpha p & (1-\alpha)p & \beta q & (1-\beta)q \\ \alpha p & (1-\alpha)p & \beta q & (1-\beta)q \\ \alpha p & (1-\alpha)p & \beta q & (1-\beta)q \end{pmatrix}$$

which corresponds to the case of independent states, will give the required probabilities P_1 . However, assuming the more general matrix which results in these stationary probabilities to be of the form

$$(5.4.7) \quad \underline{p} = \begin{pmatrix} ap & (1-a)p & eq & (1-e)q \\ bp & (1-b)p & fq & (1-f)q \\ cp & (1-c)p & gq & (1-g)q \\ dp & (1-d)p & hq & (1-h)q \end{pmatrix}$$

$$(0 < a, b, c, d, e, f, g, h < 1),$$

we find that (5.4.3) result in four distinct equations

$$\begin{aligned} a\alpha + b(1-\alpha) - c\beta - d(1-\beta) &= 0, \\ c\beta + d(1-\beta) &= \alpha, \\ g\beta + h(1-\beta) - e\alpha - f(1-\alpha) &= 0, \\ e\alpha + f(1-\alpha) &= \beta, \end{aligned}$$

for which the following solutions are obtained

$$\begin{aligned} a &= \alpha^{-1} \{ \alpha + b(\alpha - 1) \} \\ c &= \beta^{-1} \{ \alpha + d(\beta - 1) \} \\ e &= \alpha^{-1} \{ \beta + f(\alpha - 1) \} \\ g &= \beta^{-1} \{ \beta + h(\beta - 1) \}. \end{aligned}$$

It follows that for any positive values of

$$b < \alpha(1-\alpha)^{-1}, \quad h < \beta(1-\beta)^{-1},$$

$$(\alpha-\beta)(1-\beta)^{-1} < d < \alpha(1-\beta)^{-1}, \quad (\beta-\alpha)(1-\alpha)^{-1} < f < \beta(1-\alpha)^{-1},$$

which ensure that the corresponding values of a, c, e, g , lie between zero and 1, the stochastic matrix (5.4.7) will have the required stationary probabilities.

We see therefore that in the case where the exponents of the stationary probabilities $P_i(0)$ are not all distinct, the stochastic matrix p corresponding to these may be not only of the type with independent states, but also of the type with dependent states such that the transition probabilities in each column have the same exponent, and are subject to certain conditions for their coefficients.

5. Stochastic matrices with constant stationary probabilities

Suppose now that all the stationary probabilities P_i are constants. For this to be so, it is necessary that (5.3.4) result in constants for all values of i ; a systematic study of these cases is difficult, but two simple solutions are known.

The first is the doubly stochastic matrix $\underline{p}$ such that $\sum_{j=1}^s p_{ij} = \sum_{i=1}^s p_{ij} = 1$ ($i, j = 1, \dots, s$). Among these are the circulant matrix

$$(5.5.1) \quad \underline{p} = \begin{pmatrix} p_1 & p_2 & \dots & p_s \\ p_s & p_1 & \dots & p_{s-1} \\ \dots & \dots & \dots & \dots \\ p_2 & p_3 & \dots & p_1 \end{pmatrix}$$

in which the transition probabilities $p_1, \dots, p_s$ clearly satisfy the conditions for a doubly stochastic matrix, and the $s!$ matrices with the rows of (5.5.1) rearranged in all possible orders. In general, these are by no means the only possible doubly stochastic matrices; this is clear if we consider as an example the two matrices for the Markov chain with four states,

$$\begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ p_4 & p_1 & p_2 & p_3 \\ p_3 & p_4 & p_1 & p_2 \\ p_2 & p_3 & p_4 & p_1 \end{pmatrix} \quad \begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ p_2 & p_1 & p_4 & p_3 \\ p_4 & p_3 & p_1 & p_2 \\ p_3 & p_4 & p_2 & p_1 \end{pmatrix}$$

where the first is the circulant matrix (5.5.1), but the second, also doubly stochastic, is not one of the $s!$ matrices obtained by a rearrangement of the rows of the first. There is no restriction on the exponents

of the p_i ; these may all be distinct, or such that there are only $r < s$ distinct values for them. It is clear that the stationary probabilities $P_i = s^{-1}$ ($i=1, \dots, s$) will satisfy the equations $\underline{p}'\underline{P} = \underline{P}$, and since the chain is regular, these are known (§ 2.3, 21) to be unique.

A particular case of interest is that in which the doubly stochastic matrix consists of transition probabilities with only two distinct exponents: we may for simplicity write these, apart from coefficients, as p and $q = 1 - p$. One of the possible doubly stochastic matrices would then be

$$(5.5.2) \quad \underline{p} = \begin{pmatrix} q & \alpha_{12}p & \dots & \alpha_{1s}p \\ \alpha_{21}p & q & \dots & \alpha_{2s}p \\ \dots & \dots & \dots & \dots \\ \alpha_{s1}p & \alpha_{s2}p & \dots & q \end{pmatrix}$$

where all $\alpha_{ij} > 0$, and $\sum_{\substack{j=1 \\ i \neq j}}^s \alpha_{ij} = \sum_{\substack{i=1 \\ i \neq j}}^s \alpha_{ij} = 1$.

This leads us to a second type of stochastic matrix for which the stationary probabilities are constants: in this case, $\underline{p}$ is not doubly stochastic, but may easily be derived as a generalisation of (5.5.2). Consider a stochastic matrix of form similar to (5.5.2),

which may be written simply as $\underline{p} = \{\delta_{ij} + \alpha_{ij}p_j\}$, and where for $i \neq j$, $\alpha_{ij} > 0$, and $\alpha_{ii} = -\sum_{j=1, j \neq i}^s \alpha_{ij} = -1$ ($i=1, \dots, s$). Writing the matrix $\underline{p}' - \underline{I}$ of (5.3.2) as

$$(5.5.3) \quad \begin{pmatrix} -1 & \alpha_{12} & \cdot & \cdot & \cdot & \alpha_{1s} \\ \alpha_{21} & -1 & \cdot & \cdot & \cdot & \alpha_{2s} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha_{s1} & \alpha_{s2} & \cdot & \cdot & \cdot & -1 \end{pmatrix} \underline{p}$$

it is directly obvious that the cofactors Δ_i will all be of the form $\Delta_i = C_i p^{s-1}$, where the C_i are themselves cofactors of the i -th diagonal terms in the matrix of constants in (5.5.3); the stationary probabilities will have the constant values

$$(5.5.4) \quad P_i = C_i \left\{ \sum_{i=1}^s C_i \right\}^{-1} \quad (i=1, \dots, s).$$

III. The Markov Chain with Three States

It has been stated that the various forms of the stochastic matrices given in §§5.4-5.5 are possibly not exhaustive in the general case of s states. We discuss as an example the case of the chain with three states, for which the matrices discussed in §§5.4-5.5

are the only ones which give stationary probabilities of the forms (5.3.5) or (5.3.6).

6. Stochastic matrices with stationary

probabilities $P_i = \alpha_i e^{K_i \lambda_1 + \lambda_2}$

In the case where the stationary probabilities are all of the form (5.3.5) with the exponents K_i all distinct, there is only a single possibility for the stochastic matrix. For we have seen in § 5.4 that if

$$P_1 = \alpha_1 e^{K_1 \lambda_1 + \lambda_2}, P_2 = \alpha_2 e^{K_2 \lambda_1 + \lambda_2}, P_3 = \alpha_3 e^{K_3 \lambda_1 + \lambda_2} \quad (K_1 \neq K_2 \neq K_3),$$

there is a unique stochastic matrix for three independent states

$$(5.6.1) \quad P = \begin{pmatrix} P_1 & P_2 & P_3 \\ P_1 & P_2 & P_3 \\ P_1 & P_2 & P_3 \end{pmatrix}$$

which will have these stationary probabilities.

It is conceivable, however, if the stationary probabilities have only two distinct exponents K_1, K_2 , that they take two forms; the first

$$(5.6.2) \quad P_1 = \alpha_1 e^{K_1 \lambda_1 + \lambda_2}, P_2 = \alpha_2 e^{K_2 \lambda_1 + \lambda_2}, P_3 = \alpha_3 e^{K_2 \lambda_1 + \lambda_2},$$

for which all the stationary probabilities are of the form (5.3.5), and the second

$$P_1 = \alpha_1 e^{K_1 \lambda_1 + \lambda_2}, P_2 = \alpha_2 e^{K_2 \lambda_1 + \lambda_2}, P_3 = \alpha_3,$$

for which two of the probabilities are of the form (5.3.5), and the third is of the form (5.3.6). In the particular case of the chain with three states, we prove that this second possibility, with mixed forms, is not permissible.

For this case, the equations $\sum_i P_i = \sum_j p_{ij} = 1$ will result in

$$\alpha_1 e^{K_1 \lambda_1 + \lambda_2} + \alpha_2 e^{K_2 \lambda_1 + \lambda_2} + \alpha_3 = \sum_j \alpha_{ij} e^{K_{ij} \lambda_1 + \lambda_2} \quad (i = 1, 2, 3),$$

from which we see that there are only two distinct values for the exponents K_{ij} of the transition probabilities. The equations (5.4.3) give for P_1 ,

$$(5.6.3) \quad \alpha_{11} \alpha_1 e^{K_{11} \lambda_1} + \alpha_{21} \alpha_2 e^{(K_{21} + K_2 - K_1) \lambda_1} + \{ \alpha_{31} \alpha_3 e^{(K_{31} - K_1) \lambda_1} - \alpha_1 \} (\alpha_1 e^{K_1 \lambda_1} + \alpha_2 e^{K_2 \lambda_1} + \alpha_3) = 0.$$

Consider the exponent K_{31} in the third term; we know that it can have either the value K_1 or K_2 . If $K_{31} = K_1$, and K_{11} , K_{21} are either of K_1 or K_2 , then (5.6.3) becomes

$$\alpha_{11} \alpha_1 e^{K_{11} \lambda_1} + \alpha_{21} \alpha_2 e^{(K_{21} + K_2 - K_1) \lambda_1} + \{ \alpha_{31} \alpha_3 - \alpha_1 \} (\alpha_1 e^{K_1 \lambda_1} + \alpha_2 e^{K_2 \lambda_1} + \alpha_3) = 0,$$

so that $(\alpha_{31}\alpha_3 - \alpha_1)\alpha_3 = 0$. This is true if either $\alpha_3 = 0$, which is impossible for a positively regular chain, or $(\alpha_{31}\alpha_3 - \alpha_1) = 0$ so that the sum of the two remaining positive terms is zero, which is also impossible. If $K_{31} = K_2$, then we obtain for (5.6.3)

$$\begin{aligned} & \alpha_{11}\alpha_1 e^{K_1\lambda_1} + \alpha_{21}\alpha_2 e^{(K_2 + K_2 - K_1)\lambda_1} + \alpha_1(\alpha_{31}\alpha_3 - \alpha_2)e^{K_2\lambda_1} \\ & + \alpha_{31}\alpha_3\alpha_2 e^{(2K_2 - K_1)\lambda_1} + \alpha_{31}\alpha_3^2 e^{(K_2 - K_1)\lambda_1} - \alpha_1^2 e^{K_1\lambda_1} - \alpha_1\alpha_3 = 0, \end{aligned}$$

where again, an examination of the terms shows that the hypothesis is not tenable. It follows therefore, that at least for the Markov chain with three states, the mixed case is impossible: we are left with (5.6.2) as the only remaining possibility.

In §5.4 it was shown that for a case such as (5.6.2), where the number $r=2$ of distinct exponents for the stationary probabilities is less than the number $s=3$ of states, there are at least two solutions for the stochastic matrix. The first was of the form (5.6.1) where the states are independent, and the second of the form

$$(5.6.4) \quad \underline{p} = \begin{pmatrix} \alpha_{11} e^{K_1\lambda_1 + \lambda_2} & \alpha_{12} e^{K_2\lambda_1 + \lambda_2} & \alpha_{13} e^{K_2\lambda_1 + \lambda_2} \\ \alpha_{21} e^{K_1\lambda_1 + \lambda_2} & \alpha_{22} e^{K_2\lambda_1 + \lambda_2} & \alpha_{23} e^{K_2\lambda_1 + \lambda_2} \\ \alpha_{31} e^{K_1\lambda_1 + \lambda_2} & \alpha_{32} e^{K_2\lambda_1 + \lambda_2} & \alpha_{33} e^{K_2\lambda_1 + \lambda_2} \end{pmatrix},$$

where the transition probabilities in each column have the same exponents, and the coefficients satisfy some conditions such as (5.4.6). It was not possible, in the general case of s states, to say whether these were exhaustive; for the case of three states, however, we prove that no other solutions exist.

Consider once again the equation for P_1 in (5.4.3),

$$\alpha_{11}\alpha_1 e^{K_{11}A_1} + \alpha_{21}\alpha_2 e^{(K_{21}+K_2-K_1)A_1} + \alpha_{31}\alpha_3 e^{(K_{31}+K_2-K_1)A_1} = \alpha_1 \{ \alpha_1 e^{K_1 A_1} + (\alpha_2 + \alpha_3) e^{K_2 A_1} \};$$

for it to hold, we must have two of the exponents in the left hand side of the equation equal, either $K_{11} = K_{21} + K_2 - K_1$ or $K_{11} = K_{31} + K_2 - K_1$ or $K_{21} + K_2 - K_1 = K_{31} + K_2 - K_1$. The first is not possible, since it results in $K_{11} - K_{21} = K_2 - K_1$ which is true only if $K_{11} = K_2$, $K_{21} = K_1$, and the last exponent must therefore be $K_{31} + K_2 - K_1 = K_1$ where $K_{31} = 2K_1 - K_2$ which is not in general possible. The second will give rise to a similar result, and is therefore also not possible. The third will give $K_{21} = K_{31}$ which can only be equal to K_1 , and the third exponent K_{11} can also only equal K_1 . This means that the exponents of the transition probabilities in the first column are all equal to K_1 ;

we see as a result of this in the equation, that

$$\alpha_{11} = \alpha_1, \quad \alpha_{21}\alpha_2 + \alpha_{31}\alpha_3 = \alpha_1(\alpha_2 + \alpha_3).$$

Now the coefficients α_{21} , α_{31} can only be less than or equal to α_1 , so that

$$\alpha_{21}\alpha_2 + \alpha_{31}\alpha_3 \leq \alpha_1(\alpha_2 + \alpha_3);$$

the required equality will hold only if $\alpha_{21} = \alpha_{31} = \alpha_1$, and it follows therefore that (5.6.4) is the only possible type of stochastic matrix with the stationary probabilities (5.6.2).

Writing the equations for P_2 and P_3 in (5.4.3) as

$$\begin{aligned} \alpha_{12}\alpha_1 e^{K_1\Lambda_1} + \alpha_{22}\alpha_2 e^{K_2\Lambda_1} + \alpha_{32}\alpha_3 e^{K_2\Lambda_1} &= \alpha_2 \{ \alpha_1 e^{K_1\Lambda_1} + (\alpha_2 + \alpha_3) e^{K_2\Lambda_1} \}, \\ \alpha_{13}\alpha_1 e^{K_1\Lambda_1} + \alpha_{23}\alpha_2 e^{K_2\Lambda_1} + \alpha_{33}\alpha_3 e^{K_2\Lambda_1} &= \alpha_3 \{ \alpha_1 e^{K_1\Lambda_1} + (\alpha_2 + \alpha_3) e^{K_2\Lambda_1} \}, \end{aligned}$$

it is clear that the coefficients α_{12} , α_{13} are given by $\alpha_{12} = \alpha_2$, $\alpha_{13} = \alpha_3$, so that the most general matrix with the stationary probabilities (5.6.2) is

$$\underline{P} = \begin{pmatrix} \alpha_1 e^{K_1\Lambda_1 + \Lambda_2} & \alpha_2 e^{K_2\Lambda_1 + \Lambda_2} & \alpha_3 e^{K_2\Lambda_1 + \Lambda_2} \\ \alpha_1 e^{K_1\Lambda_1 + \Lambda_2} & \alpha_{22} e^{K_2\Lambda_1 + \Lambda_2} & \alpha_{23} e^{K_2\Lambda_1 + \Lambda_2} \\ \alpha_1 e^{K_1\Lambda_1 + \Lambda_2} & \alpha_{32} e^{K_2\Lambda_1 + \Lambda_2} & \alpha_{33} e^{K_2\Lambda_1 + \Lambda_2} \end{pmatrix},$$

where the unknown coefficients satisfy the conditions

$$\alpha_{22}\alpha_2 + \alpha_{32}\alpha_3 = \alpha_2(\alpha_2 + \alpha_3),$$

$$\alpha_{13}\alpha_2 + \alpha_{33}\alpha_3 = \alpha_3(\alpha_2 + \alpha_3).$$

We may for simplicity write the probabilities (5.6.2) in the somewhat different form

$$P_1 = q = 1 - p, \quad P_2 = \alpha p, \quad P_3 = (1 - \alpha)p, \quad (0 < \alpha < 1),$$

and the most general stochastic matrix associated with these as

$$\underline{p} = \begin{pmatrix} q & \alpha p & (1 - \alpha)p \\ q & \lambda p & (1 - \lambda)p \\ q & \mu p & (1 - \mu)p \end{pmatrix}, \quad (0 < \lambda, \mu < 1),$$

where it is easily found that the coefficients λ, μ , are related by the equation

$$\mu = \alpha(1 - \alpha)^{-1}(1 - \lambda).$$

We may have any values $0 < \alpha < 1$ whatsoever, but $\alpha^{-1}(2\alpha - 1) < \lambda < 1$ to ensure that $\mu < 1$.

7. Stochastic matrices with constant stationary probabilities

For this case, two possibilities are given in § 5.5. The first is any matrix $\underline{p}$ which is doubly stochastic; however, in the case of the chain with three states, the most general doubly stochastic matrices prove to be the circulant matrix of form (5.5.1)

$$(5.7.1) \quad \underline{p} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_3 & p_1 & p_2 \\ p_2 & p_3 & p_1 \end{pmatrix},$$

and any of its five other rearrangements

$$\begin{pmatrix} p_1 & p_2 & p_3 \\ p_2 & p_3 & p_1 \\ p_3 & p_1 & p_2 \end{pmatrix}, \begin{pmatrix} p_3 & p_1 & p_2 \\ p_1 & p_2 & p_3 \\ p_2 & p_3 & p_1 \end{pmatrix}, \begin{pmatrix} p_3 & p_1 & p_2 \\ p_2 & p_3 & p_1 \\ p_1 & p_2 & p_3 \end{pmatrix}, \begin{pmatrix} p_2 & p_3 & p_1 \\ p_1 & p_2 & p_3 \\ p_3 & p_1 & p_2 \end{pmatrix}, \begin{pmatrix} p_2 & p_3 & p_1 \\ p_3 & p_1 & p_2 \\ p_1 & p_2 & p_3 \end{pmatrix},$$

for all of which the stationary probabilities are $P_1 = P_2 = P_3 = 1/3$. The exponents of the transition probabilities may be all distinct, or alternatively only two of them may be distinct, in which case the matrix (5.7.1) may be written as

$$\underline{p} = \begin{pmatrix} q & \alpha p & (1-\alpha)p \\ (1-\alpha)p & q & \alpha p \\ \alpha p & (1-\alpha)p & q \end{pmatrix}, \quad (0 < \alpha < 1).$$

The second possibility is the matrix

$$\underline{p} = \begin{pmatrix} q & \alpha p & (1-\alpha)p \\ (1-\beta)p & q & \beta p \\ \gamma p & (1-\gamma)p & q \end{pmatrix}, \quad (0 < \alpha, \beta, \gamma < 1),$$

with the constant probabilities

$$P_1 = \frac{1 - \beta(1-\gamma)}{\{3 - (\alpha + \beta + \gamma) + \alpha\beta + \alpha\gamma + \beta\gamma\}}, \quad P_2 = \frac{1 - \gamma(1-\alpha)}{\{3 - (\alpha + \beta + \gamma) + \alpha\beta + \alpha\gamma + \beta\gamma\}}, \quad P_3 = \frac{1 - \alpha(1-\beta)}{\{3 - (\alpha + \beta + \gamma) + \alpha\beta + \alpha\gamma + \beta\gamma\}}$$

It is found by a process of enumeration that these are the only possible cases.

Part II

APPLICATIONS OF MARKOV CHAINS TO PROBLEMS OF STORAGE

Chapter 6

PROBLEMS OF STORAGE

I. Some Theory of Storage1. Introduction

Problems in the theory of storage can perhaps be briefly described in words which Dvoretzky, Kiefer and Wolfowitz (1952a) have used of the inventory problem: they consist essentially of "stochastic processes of a fairly general nature, for which, at various instants of time, one may exert some degree of control over the distribution functions in question in order to achieve a desired result."

Owing to the way in which developments in the theory have arisen, the literature in the subject presents these problems in various guises. In one set, usually known as inventory problems, and which is of interest

to economists, it is required for some specific conditions to find optimal stock and replacement policies. These give rise to results not directly related to the kind with which we are concerned; Arrow, Harris and Marschak (1951), Dvoretzky, Kiefer and Wolfowitz (1952a, 1952b, 1953) and Rosenblatt (1954) have obtained for a variety of cases in which the consumption is a random variable with a known (or unknown) probability distribution, some optimal stock and replacement policies.

Other problems of storage appear either as the provisioning problem, where the distribution of the quantity of components in a stock is to be determined for given consumption and replacement policies, or alternatively as problems in the theory of dams, where for known input and release rules, the distribution of the water content of a dam is to be found. It will be shown that these two problems are in effect analogous, and can be adequately described in terms of a general storage function, fed by an input function, and depleted always by an output function, but also, when the storage function is finite, by an overflow function to ensure that this never rises above a prescribed maximum.

Elementary probability methods are used in the study of storage problems of the provisioning type by Pitt (1946), and Hammersley (1955). Pitt (1946) obtains the distribution of a stock when the consumption is known to be of Poisson type, with orders of unit size, and there are two possible replacement policies depending on this consumption. Hammersley (1955) considers a store which is completely restocked at each replacement, and for which the policy of replacements is such that time-intervals between these are independent and their distribution is known. The consumption is again of Poisson type, but the sizes of orders may vary, these being independent of all other orders and of the times of ordering, and having a known distribution. Both the distribution of the consumption in the time between two replacements, and the probability that the store does not become empty in this interval are found.

A second method of solving some storage problems in the theory of dams depends on the construction of a Markov chain model. This use of Markov chain methods is foreshadowed in Kendall's (1951) paper on the theory of queues. Here some results are obtained for a single-counter queue fed by an input of Poisson type,

and with a service time distribution of some general form. By considering the epochs at which individual customers depart from the queue counter as a sequence of regeneration points, the stochastic process describing the queue size, which is not in general Markovian, can be adequately represented by an enumerable Markov chain. For this chain, Kendall derives some general theorems. This queue problem can be thought of as a special case of a discrete storage problem with continuous time, where (a) the positive storage function (queue size) can be infinite, (b) the discrete input of single units is of Poisson type, and (c) a unit output will always follow a unit input after a random time-lag (service time), this having some distribution of a general form when the storage function is not zero. Then, under some specific conditions, several properties of the storage function can be obtained.

Moran (1954), starting with some practical considerations on the uses of a storage system, evolves a finite Markov chain model for his probability theory of dams. For a finite discrete dam content (storage function), defined at fixed and equally spaced time-intervals, the distribution of the random input between

time-intervals is assumed known, and the output, a fixed quantity, is released at the end of these time-intervals. Equations are framed for the distributions of the dam content at consecutive release times, from which the stationary distribution of the dam content can always be obtained numerically. In a subsequent paper, Moran (1955), these are used to derive an exact solution for the stationary distribution of the dam content in a discrete case where the input has a distribution of geometric type; this is then used to obtain the solution of the continuous analogue where the input has a negative exponential distribution.

Our own contributions consist, in the main, of applications of Markov chain methods to the solution of some storage problems stemming from Pitt's and Moran's work. We proceed therefore to summarize their relevant results in some detail.

2. Pitt's problems of provisioning

In his work on the theory of provisioning, Pitt (1946) is concerned with the derivation of stationary probability distributions for a discrete stock function $s(t)$. This stock function represents the number of

components in stock at a time t , and is allowed to take integral values ranging between K and $-\infty$, where $K = s(0)$ is the initial stock, and negative values indicate that components may be borrowed if they are not in stock. The function is defined for all values of time t (t continuous), by the equation

$$(6.2.1) \quad s(t) = K + r(t) - c(t),$$

where $c(t)$ is the consumption function, the number of components consumed up to time t , a random increasing step function taking positive integral values, and $r(t)$ is the replacement function, the number of components delivered and added to the stock up to time t , a function depending on $c(t)$ and always smaller than or equal to it.

The probability distribution associated with the consumption function $c(t)$ is the Poisson with parameter a , so that the probabilities that in a small interval of time δt there be one and no components consumed respectively are

$$(6.2.2) \quad \begin{cases} \Pr\{c(t+\delta t) - c(t) = 1\} = a\delta t, \\ \Pr\{c(t+\delta t) - c(t) = 0\} = 1 - a\delta t, \end{cases}$$

and increases in $c(t)$ in non-overlapping intervals are independent. Two types of replacement functions are considered: type 1

$$(6.2.3) \quad r_1(t) = [c(t-T)M^{-1}]M,$$

where we define $[x]$ as the integral part of x ; here, orders for a constant number of components M are sent out at irregular time-intervals $t - T$, when $c(t - T)$ is an integral multiple of M , and are delivered and added to the stock at time t , after a positive time-lag T ; and type 2

$$(6.2.4) \quad r_2(t) = c\left([(t-T)aM^{-1}]Ma^{-1}\right);$$

here, orders for an irregular number of components equal to the consumption in the previous time-interval Ma^{-1} are sent out at regular times kMa^{-1} ($k=1,2,\dots$), and are delivered and added to stock at times $kMa^{-1} + T$ after a positive time-lag T .

For both these types of replacement function, the stationary probability distributions of $s(t)$ are derived by arguments which will be followed exactly in §§6.6-6.7. Using these probability distributions, Pitt concludes with a comparison of the means of both the positive and negative values of the stock function

$s(t)$ when the replacement functions are of types 1 and 2.

3. Moran's problems in the theory of dams

The problem discussed by Moran (1954), that of obtaining the stationary probability distribution of the amount of water in a dam, appears at first sight to be of a different nature. A finite dam of capacity K units, whose content at times $t = 0, 1, 2, \dots$, after water has been released according to a prescribed rule, is $Z(t)$, where $Z(t)$ will be found to lie in the range $(0, K - M)$, is subject to the following conditions:

(1) a discrete input $X(t)$ which flows into the dam during the interval of time $(t, t+1)$, where the series $\{X(t)\}$ is serially independent, and the probability that the input be i units ($i = 0, 1, 2, \dots$) is p_i ;

(2) an overflow rule such that in any time-interval $(t, t+1)$ there is no overflow from the dam if $Z(t) + X(t) \leq K$, and there is an overflow $Z(t) + X(t) - K$ if $Z(t) + X(t) > K$;

(3) a release rule such that at time $t+1$, M units are released from the dam if its total content $Z(t) + X(t) \geq M$, and the total content $Z(t) + X(t)$ is released if this is less than M .

probabilities P_i can always be evaluated numerically for any known matrix of coefficients p' , and these are in fact computed for a special case. An extension of this discrete theory also allows equations for the continuous stationary distribution of $U(t) = Z(t) + X(t)$ to be written when $Z(t)$, $X(t)$ are continuous and $U(t)$ lies in the range $(0, \infty)$.

4. Provisioning and dam problems as general storage problems

We now proceed to identify these problems of the stock function $s(t)$ of provisioning, and of the dam content $Z(t)$ in dam theory as particular problems of a general storage function $S(t)$. This may be discrete or continuous, and defined for continuous time t or at fixed times such as $t = 0, 1, 2, \dots$, according to the conditions of the problem, by the equation

$$(6.4.1) \quad S(t) = I(t) - D(t) - F(t),$$

where $I(t)$, $D(t)$, $F(t)$ are positive increasing functions, discrete or continuous. $I(t)$ is a random input function which feeds the storage function, $D(t)$ is an output function which depletes the storage function and may

depend on $I(t)$ or be otherwise defined, and $F(t)$ is an overflow function which operates to restrict the range of $S(t)$ below a prescribed maximum value.

To show that the problems of Pitt's stock function $s(t)$ of provisioning can be regarded as problems of a particular storage function $S(t)$, we consider the previously defined stock function (6.2.1)

$$s(t) = K + r(t) - c(t),$$

which lies in the range $-\infty < s(t) \leq K$. We note that this equation holds at any time t (t continuous), and that $c(t)$, the random discrete consumption function, and $r(t)$, the discrete replacement function ($r(t) \leq c(t)$) are defined as in § 6.2. We can here equally well consider the stock deficit,

$$(6.4.2) \quad K - s(t) = c(t) - r(t),$$

which ranges from zero when the stock is at its maximum K , to ∞ as the stock decreases through 0 to $-\infty$. Naturally this is possible only when the stock has a maximum value K which it cannot exceed. Comparing this with a storage function

$$(6.4.3) \quad S(t) = I(t) - D(t),$$

for which $F(t) = 0$ for all t , it is clear that a discrete storage function $S(t)$ defined in the range $(0, \infty)$ for all time t (t continuous), is identifiable with the deficit $K - s(t)$. Similarly, $I(t)$, a discrete random input function, is identifiable with the discrete random consumption function $c(t)$, and $D(t)$, a discrete output function ($D(t) \leq I(t)$), with a discrete replacement function $r(t)$. $F(t)$ will be zero at all times, since the range of the stock deficit is $(0, \infty)$, and no overflow is required to restrict this range below any prescribed value.

In the specific problems considered by Pitt, where the consumption function $c(t)$ was of Poisson type with parameter a , and the replacement functions of the types previously defined in (6.2.3), (6.2.4) as $r_1(t)$ and $r_2(t)$, the general storage equation would be such that

$$S(t) = K - s(t), \quad I(t) = c(t)$$

$$D(t) = r_1(t) \text{ or } r_2(t), \quad F(t) = 0.$$

However, generally, for any stock function $s(t)$, whether continuous or discrete, defined for continuous time t or at discrete intervals of time, providing only that the maximum stock K is finite, it is always



possible to frame the general storage function $S(t)$ for the stock deficit as in equation (6.4.1). The finite maximum stock K results in the storage function $S(t)$ having a range with lower bound 0; its upper bound may be infinite, as in the case just considered, or finite as we shall see in §§6.8-6.11. $I(t)$, $D(t)$, and $F(t)$ will be appropriately chosen to fit the conditions of the problem.

Similarly, we proceed to show that Moran's discrete dam content $Z(t)$ with range $(0, K - M)$ can be regarded as a particular storage function. In this case, however, the function $Z(t)$ is defined only at the times $t = 0, 1, 2, \dots$, and we shall also define the storage function at these times. In our storage equation (6.4.1), we may identify $S(t)$ with $Z(t) - Z(0)$, the increase in dam content after a time t ; if without loss in generality, we assume that $Z(0) = 0$, that is, the dam is empty at time $t = 0$, we have that $S(t)$ the storage function is equal to the dam content $Z(t)$. The random discrete input function $I(t)$ would be the sum of the discrete random inputs $X(0), X(1), \dots, X(t-1)$ in the time-intervals $(0, 1), (1, 2), \dots, (t-1, t)$, so that

$$I(t) = \sum_{\tau=1}^t X(\tau-1).$$

The discrete output function $D(t)$ would be the sum of discrete outputs released at times $\tau = 1, 2, \dots, t$; let these be written $d(1), d(2), \dots, d(t)$, then we have that

$$d(\tau) = M \quad \text{if} \quad Z(\tau - 1) + X(\tau - 1) \geq M,$$

or $d(\tau) = Z(\tau - 1) + X(\tau - 1)$ if this is less than M ,

so that the output function is

$$D(t) = \sum_{\tau=1}^t d(\tau).$$

Finally, the overflow function $F(t)$ would be the sum of the discrete overflows in the time intervals $(0,1), (1,2), \dots, (t-1,t)$; let these be written $f(0), f(1), \dots, f(t-1)$, then we have that

$$f(\tau - 1) = 0 \quad \text{if} \quad Z(\tau - 1) + X(\tau - 1) \leq K$$

or $f(\tau - 1) = Z(\tau - 1) + X(\tau - 1) - K$ if this is greater than zero, so that the overflow function is

$$F(t) = \sum_{\tau=1}^t f(\tau - 1).$$

The equation for $Z(t)$ at times $t = 0, 1, 2, \dots$ could be written

$$(6.4.4) \quad Z(t) = \sum_{\tau=1}^t X(\tau - 1) - \sum_{\tau=1}^t d(\tau) - \sum_{\tau=1}^t f(\tau - 1),$$

a particular case of the storage function $S(t)$. For a continuous dam content, a similar equation could also be written; more generally, for a $Z(t)$ continuous or discrete, defined at any time t (t continuous) or at discrete times, with a finite or infinite range, it is possible to frame a general storage equation.

In neither the case of problems of provisioning nor of dam problems will the storage equation itself necessarily provide a method of solution enabling the stationary probabilities for the function to be found. But as we are always able to reduce provisioning or dam problems to a general storage problem, this means that a provisioning problem can be interpreted equally well as a dam problem, and vice versa. In other words, any problem and its solution in the one theory will have an exact analogue in the other; equally, any method of attack on a problem in the one theory may be useful in analogous problems of the other.

We proceed to interpret Pitt's results in provisioning theory as solutions in two analogous cases of the infinite dam. We shall also apply Moran's methods in the theory of dams to some problems in the theory of provisioning.

II. Interpretation of Pitt's Results for Dam Problems

5. Problems of the infinite dam

Pitt's equation for the stock deficit

$$K - s(t) = c(t) - r(t) \quad (r(t) \leq c(t)),$$

was seen to be an example of the storage equation

$$S(t) = I(t) - D(t),$$

where at any time t (t continuous), $S(t)$ is discrete and lies in the range $(0, \infty)$, $I(t)$ is a random discrete input function, and $D(t)$ is a discrete output function depending on $I(t)$ ($D(t) \leq I(t)$). Equally, a storage equation of this form can represent the discrete dam content $Z(t)$, ($S(t) = Z(t)$) at any time t , where the dam is infinite, $0 \leq Z(t) < \infty$. In this case, interpreting the specific conditions of Pitt's problems of provisioning for an infinite dam, we have that the discrete input function $I(t)$ will be of Poisson type with parameter λ such that, in a small interval of time Δt the probabilities that one and no units of water

flow into the dam are respectively

$$(6.5.1) \quad \begin{cases} P_r \{ I(t+\delta t) - I(t) = 1 \} = a \delta t, \\ P_r \{ I(t+\delta t) - I(t) = 0 \} = 1 - a \delta t, \end{cases}$$

and increases in $I(t)$ in non-overlapping intervals are independent. The release rules corresponding to Pitt's two replacement functions are such that the resulting output will be of type 1:

$$(6.5.2) \quad D_1(t) = [I(t-T)M^{-1}]M;$$

here, releases of M units are made at irregular time-intervals t , after a time-lag T from the times $t - T$ when $I(t - T)$ is an integral multiple of M . In practice, this might mean that at every time when the input increased by M units, a decision would be taken to release M units from the dam, but this decision would be carried out after a time-lag T ; alternatively, the output may be of type 2:

$$(6.5.3) \quad D_2(t) = I([(t-T)aM^{-1}]Ma^{-1});$$

here, releases of an irregular number of units equal to the input in the previous time-interval $((k-1)Ma^{-1}, kMa^{-1})$ are made at regular times $kMa^{-1} + T$, $(k=1,2,\dots)$ with time-lag T after the interval. In practice, this

might mean that at the regular times kMa^{-1} , a decision would be taken to release the input in the previous interval Ma^{-1} , but this decision would be carried out after a time-lag T .

We proceed to follow Pitt exactly in deriving the stationary distributions of the dam content $Z(t)$ when the release rules of types (6.5.2) and (6.5.3) are functioning, and in comparing the mean contents of the dam for these cases.

6. Stationary distribution of $Z(t)$ for release rule (6.5.2)

In this case, where the output function given by (6.5.2) holds, the storage equation for the dam content can be written

$$Z(t) = I(t) - [I(t - T)M^{-1}] M \quad (0 \leq Z(t) < \infty).$$

Let the input function at times t , and $t - T$ be such that $I(t - T) = i$, and $I(t) - I(t - T) = j$, where i and j can take the values $0, 1, 2, \dots$. Then, writing

$$P_n(t) = \Pr \{Z(t) = n\} \quad (n = 0, 1, 2, \dots),$$

for the probability that the dam content be n units

at time t , we have that

$$P_n(t) = \Pr \{i+j - [iM^{-1}]M = n\}.$$

Now suppose that $kM \leq i \leq (k+1)M - 1$ ($k=0,1,\dots$), then it follows that $[iM^{-1}] = k$, $i = kM + v$, and $j = n - v$, where v can take the values $0,1,\dots,M-1$. From the independence condition for increases in $I(t)$ in non-overlapping intervals, it follows that

$$\begin{aligned} (6.6.1) \quad P_n(t) &= \sum_{i,j} \Pr \{I(t-T) = i\} \cdot \Pr \{I(t) - I(t-T) = j\} \\ &= \sum_{v=0}^{M-1} \sum_{k=0}^{\infty} \left(e^{-a(t-T)} \{a(t-T)\}^{kM+v} / (kM+v)! \right) \cdot g_{n-v}, \end{aligned}$$

where we write

$$g_r = \begin{cases} e^{-aT} (aT)^r / r! & (r \geq 0), \\ 0 & (r < 0). \end{cases}$$

The stationary distribution is obtained when $t \rightarrow \infty$; we shall write it $P_n = \lim_{t \rightarrow \infty} P_n(t)$. It follows that

$$(6.6.2) \quad P_n = \sum_{v=0}^{M-1} g_{n-v} \cdot \lim_{t \rightarrow \infty} \sum_{k=0}^{\infty} e^{-a(t-T)} \{a(t-T)\}^{kM+v} / (kM+v)!$$

Now consider the series $\sum_{k=0}^{\infty} e^{-y} y^{kM+v} / (kM+v)!$, where $y = a(t-T)$; if $y = v + (N+1)M$, we have that since $v \leq M-1$, then for $k=0$,

$$(6.6.3) \quad M^{-1} \{1 + y + \dots + \frac{y^v}{v!}\} < \frac{y^v}{v!} < M^{-1} \left\{ \frac{y^v}{v!} + \dots + \frac{y^{v+M-1}}{(v+M-1)!} \right\}.$$

For $k = 1, 2, \dots, N$, the following inequalities hold

$$(6.6.4) \quad M^{-1} \left\{ \frac{y^{v+(k-1)M+1}}{(v+(k-1)M+1)!} + \dots + \frac{y^{v+kM}}{(v+kM)!} \right\} \\ < \frac{y^{v+kM}}{(v+kM)!} < M^{-1} \left\{ \frac{y^{v+kM}}{(v+kM)!} + \dots + \frac{y^{v+(k+1)M-1}}{(v+(k+1)M-1)!} \right\}$$

but for $k = N+1$, only one side of the inequality is valid

$$(6.6.5) \quad M^{-1} \left\{ \frac{y^{v+kM+1}}{(v+kM+1)!} + \dots + \frac{y^{v+(N+1)M}}{(v+(N+1)M)!} \right\} < \frac{y^{v+(N+1)M}}{(v+(N+1)M)!}$$

Further, for all values of k greater than $N+1$ ($k = N+2, \dots$), the inequalities are reversed

$$(6.6.6) \quad M^{-1} \left\{ \frac{y^{v+kM+1}}{(v+kM+1)!} + \dots + \frac{y^{v+(k+1)M}}{(v+(k+1)M)!} \right\} \\ < \frac{y^{v+kM}}{(v+kM)!} < M^{-1} \left\{ \frac{y^{v+(k-1)M}}{(v+(k-1)M)!} + \dots + \frac{y^{v+kM-1}}{(v+kM-1)!} \right\}$$

Summing the inequalities (6.6.3) to (6.6.6) for all values of k , and multiplying the result by e^{-y} , we obtain

$$M^{-1} \left\{ 1 - \sum_{i=1}^M e^{-y} \frac{y^{v+(N+1)M+i}}{(v+(N+1)M+i)!} \right\} < \sum_{k=0}^{\infty} e^{-y} \frac{y^{v+kM}}{(v+kM)!} < M^{-1} \left\{ 1 + e^{-y} \frac{y^{v+(N+1)M}}{(v+(N+1)M)!} \right\}.$$

Now let $N \rightarrow \infty$; y will also tend to infinity, and we have that

$$M^{-1} \leq \lim_{y \rightarrow \infty} \sum_{k=0}^{\infty} e^{-y} \frac{y^{v+kM}}{(v+kM)!} \leq M^{-1},$$

so that the equation (6.6.2) for P_n can be written

$$(6.6.7) \quad P_n = M^{-1} \sum_{v=0}^{M-1} g_{n-v}.$$

The mean content of the dam is given by

$$\begin{aligned} (6.6.8) \quad \xi(Z) &= M^{-1} \sum_{n=0}^{\infty} n \sum_{v=0}^{M-1} g_{n-v} \\ &= M^{-1} \left\{ \sum_{n=0}^{\infty} (n + (n+1) + \dots + (n+M-1)) g_n \right\} \\ &= aT + \frac{1}{2} (M-1). \end{aligned}$$

7. Stationary distribution of $Z(t)$ for release rule (6.5.3)

In this case, the output function is (6.5.3), and releases are made at regular times $kMa^{-1} + T$ ($k = 0, 1, \dots$). Consider the time-interval $kMa^{-1} + T \leq t < (k+1)Ma^{-1} + T$ between two releases, where T may have any value ($T \leq Ma^{-1}$); at any time in this interval, the content $Z(t)$ of the dam has the value

$$\begin{aligned} Z(t) &= I(t) - I\left(\lceil (t - T)aM^{-1} \rceil Ma^{-1}\right) \\ &= I(t) - I(kMa^{-1}). \end{aligned}$$

Now if we write for the probability that the dam content be n units at time t ,

$$P_n(t) = \Pr \{Z(t) = n\} \quad (n = 0, 1, \dots),$$

then we have that

$$(6.7.1) \quad P_n(t) = \Pr \{ I(t) - I(kMa^{-1}) = n \} = e^{-a(t-kMa^{-1})} \frac{(at-kM)^n}{n!}$$

This is periodic in Ma^{-1} , so that in order to obtain the stationary distribution $\{P_n\}$, we write

$$\begin{aligned} P_n &= a M^{-1} \int_{t=kMa^{-1}+T}^{(k+1)Ma^{-1}+T} P_n(t) dt \\ &= M^{-1} \int_0^M e^{-(u+aT)} \frac{(u+aT)^n}{n!} du, \end{aligned}$$

where $u = a(t-kMa^{-1}-T)$. On expanding the binomial, we have

$$\begin{aligned} (6.7.2) \quad P_n &= \sum_{v=0}^{\infty} e^{-aT} \frac{(aT)^{n-v}}{(n-v)!} M^{-1} \int_0^M e^{-u} \frac{u^v}{v!} du \\ &= \sum_{v=0}^{\infty} g_{n-v} Q_v, \end{aligned}$$

where $Q_v = M^{-1} \int_0^M (e^{-u} u^v / v!) du$.

The mean of the dam content in this case is given by

$$\begin{aligned} \xi(Z) &= \sum_{n=0}^{\infty} n \sum_{v=0}^{\infty} g_{n-v} Q_v \\ &= \sum_{v=0}^{\infty} Q_v \sum_{n=0}^{\infty} (n+v) g_n \\ &= \sum_{v=0}^{\infty} Q_v (aT+v). \end{aligned}$$

It is easy to see that

$$\sum_{v=0}^{\infty} Q_v = M^{-1} \int_0^M \sum_{v=0}^{\infty} e^{-u} \frac{u^v}{v!} du = 1;$$

$$\sum_{v=0}^{\infty} v Q_v = M^{-1} \int_0^M \sum_{v=0}^{\infty} v e^{-\frac{u}{M}} \frac{u^v}{v!} du = \frac{1}{2} M,$$

whence the mean is obtained as

$$(6.7.3) \quad \xi(Z) = aT + \frac{1}{2} M.$$

Comparing (6.6.8) with (6.7.3), we note that the second release rule gives a larger mean value of the dam than the first.

III. Application of Moran's Methods to Provisioning Problems

8. Provisioning with a finite positive stock

In dealing with provisioning, our notation will be more suggestive of the conditions of the problem if we write the general storage equation in the form

$$(6.8.1) \quad S(t) = K - s(t) = c(t) - r(t) - F(t),$$

where $S(t)$ is the stock deficit, $s(t)$ the stock with maximum value K , and $c(t)$, $r(t)$, and $F(t)$, consumption, replacement and overflow functions respectively.

Pitt considered cases where, since components not

in stock could be borrowed, the deficit could become infinite, so that for replacement functions always smaller than or equal to consumption functions ($r(t) \leq c(t)$), no overflow function was necessary, and $F(t)$ was equal to zero. We shall discuss the case, in practice equally realistic, where the stock is finite and positive ($0 \leq s(t) \leq K$), so that the deficit is also finite and positive ($0 \leq S(t) \leq K$). This is the case where no borrowing is allowed, and consumption is lost between times when the stock becomes zero (or the deficit K) until there is a replacement which raises the stock above zero (or decreases the deficit below K). This will require, in cases where $r(t)$ is smaller than $c(t)$, that an overflow function operate so as to restrict the deficit to a value no greater than K , $S(t) \leq K$. In these conditions, it is difficult to discuss $S(t)$ for all time t (t continuous), though it remains comparatively simple to consider it at specific time-intervals. This was the case for the dam content $Z(t)$, defined at times $t = 0, 1, 2, \dots$, for which Moran framed a set of equations which gave its stationary probability distribution; we proceed to apply Moran's methods to three particular problems of provisioning with finite positive stock.

9. Provisioning with replacements at fixed times without time-lag

This problem was designed to provide an example of an exact analogue in provisioning theory of the dam problem considered by Moran; in practice, however, it gives rise to a perfectly possible situation. Suppose that at fixed times $t = 1, 2, \dots$, a consignment of M components arrives at a store for a stock replacement; if at these times there is a deficit equal to or greater than M , the replacement consists of all M components, but if the deficit is smaller than M , only enough components are added to the stock from the consignment to reduce the deficit to zero. There is no time-lag between the evaluation of the deficit at times $t - 0$, and the deliveries of components for replacements at times t . This might well be the case for a truck containing a consignment of M components which, in its regular deliveries at a store, discharges for replacements only as many of the components as may be required. This replacement policy can be expressed by the replacement function $r(t)$, a step-function with jumps at times $t = 1, 2, \dots$, such that

$$r(t) - r(t - 1) = M \quad \text{if } S(t - 0) \geq M,$$

or

$$r(t) - r(t - 1) = S(t - 0) \quad \text{if } S(t - 0) < M.$$

Let the discrete consumption function $c(t)$ for the store be defined in any interval of time θ by the distribution $\{p_i(\theta)\}$ ($i = 0, 1, 2, \dots$) such that

$$(6.9.1) \quad \begin{cases} (1) & \Pr \{c(t+\theta) - c(t) = i\} = p_i(\theta), \\ (2) & \text{Increases in } c(t) \text{ in non-overlapping} \\ & \text{time-intervals are independent.} \end{cases}$$

The consumption will be met by components from stock only so long as the stock is positive (or the deficit less than K), but when the stock is zero (or the deficit K), no further orders can be met until the replacement arrives. This condition defines the overflow function $F(t)$, for in the interval of time $(t - 1, t)$ between replacements, the overflow required to restrict the deficit to a value K will be such that

$$F(t) - F(t - 1) = 0, \quad \text{if } S(t - 1) + c(t) - c(t - 1) < K,$$

$$F(t) - F(t - 1) = S(t - 1) + c(t) - c(t - 1) - K, \quad \text{if}$$

this is greater than zero. The storage equation (6.8.1) can now be written for times $t = 0, 1, 2, \dots$, since the consumption, replacement and overflow functions are all defined; this, however, presents no great advantages

as it does not lead to a method of solving the problem of finding the stationary probability distribution for the deficit $S(t)$.

If for $t = 0, 1, 2, \dots$, we write the probabilities that the deficit be i components at times t and $t+1$ respectively as

$$P_i = \Pr\{S(t) = i\}, \quad P_i^{(1)} = \Pr\{S(t+1) = i\},$$

where i takes the values $0, 1, \dots, K-M$, and if we write $p_i = p_i^{(1)}$ and $q_j = \sum_{i=0}^{\infty} p_i$, then for $K-M \gg M$, the equations relating the probabilities are identical to (6.3.1), those obtained by Moran in his dam problem. In matrix form, we write $\underline{P}^{(1)} = \underline{p}' \underline{P}$, where $\underline{P}$, $\underline{P}^{(1)}$ are column vectors with elements P_i , $P_i^{(1)}$ respectively, and $\underline{p}'$ is the matrix of coefficients. Exactly as in Moran's problem, the stationary distribution $P_0, P_1, \dots, P_{K-M}$ is obtained by solving the equations $\underline{P} = \underline{p}' \underline{P}$ together with $\sum_{i=0}^{K-M} P_i = 1$; these $\{P_i\}$ cannot always be obtained in an explicit form, but they can always be evaluated numerically for any given values of the $\{p_i\}$.

10. Provisioning with replacements at fixed times $t = kMa^{-1} + T$ ($k = 1, 2, \dots$), with time-lag $T < Ma^{-1}$

(a) General equations for the stationary distributions

This problem stems from Pitt's second problem of provisioning under replacement policy of type 2, with replacement function (6.2.4). The differences arising are due to the restriction of the stock $s(t)$ to the range $0 \leq s(t) \leq K$, or the deficit $S(t)$ to the range $0 \leq S(t) \leq K$, and to the consideration of the deficit at specific time-intervals only. The consumption function $c(t)$ is defined by (6.9.1) as in the previous problem, and the replacement function (6.2.4) is

$$r_2(t) = c\left(\left[(t - T)aM^{-1}\right]Ma^{-1}\right),$$

where an irregular number of components equal to the consumption in the previous time-interval Ma^{-1} are ordered at regular times kMa^{-1} ($k = 1, 2, \dots$), and are delivered at times $kMa^{-1} + T$, after a positive time-lag T , where $T < Ma^{-1}$. As the deficit is restricted to values no greater than K , consumption is lost in the intervals after the deficit reaches the value K until a replacement arrives to reduce the deficit to a value

less than K . This defines the overflow function in such a way that, in the intervals kMa^{-1} , $kMa^{-1}+T$, between orders and deliveries, the overflow is

$$F(kMa^{-1}+T) - F(kMa^{-1}) = 0$$

(6.10.1) if $S(kMa^{-1}) + c(kMa^{-1}+T) - c(kMa^{-1}) \leq K$, or

$$F(kMa^{-1}+T) - F(kMa^{-1}) =$$

$$S(kMa^{-1}) + c(kMa^{-1}+T) - c(kMa^{-1}) - K \text{ if this}$$

is greater than zero.

Similarly, in the intervals $kMa^{-1}+T$, $(k+1)Ma^{-1}$, between deliveries and new orders, the overflow is

$$F((k+1)Ma^{-1}) - F(kMa^{-1}+T) = 0$$

if $S(kMa^{-1}+T) + c((k+1)Ma^{-1}) - c(kMa^{-1}+T) \leq K$, or

$$F((k+1)Ma^{-1}) - F(kMa^{-1}+T) =$$

$$S(kMa^{-1}+T) + c((k+1)Ma^{-1}) - c(kMa^{-1}+T) - K \text{ if this}$$

is greater than zero.

The storage equation (6.8.1) is now fully defined at times kMa^{-1} , $kMa^{-1}+T$, but as in the previous problem, this does not lead to a solution of the stationary distribution of $S(t)$, the deficit. We can, however, obtain sets of equations relating the probabilities

$$P_0, P_1, \dots, P_K; \quad P_0^{(a)}, P_1^{(a)}, \dots, P_K^{(a)}; \quad P_0^{(a)}, P_1^{(a)}, \dots, P_K^{(a)};$$

or in matrix form, $\underline{P}^{(2)} = \underline{p}^{(1)'} \underline{P}^{(1)}$:

$$(6.10.5) \quad \begin{pmatrix} P_0^{(2)} \\ P_1^{(2)} \\ P_2^{(2)} \\ \vdots \\ P_K^{(2)} \end{pmatrix} = \begin{pmatrix} p_0^{(1)} & 0 & 0 & \dots & \dots & 0 \\ p_1^{(1)} & p_0^{(1)} & 0 & \dots & \dots & 0 \\ p_2^{(1)} & p_1^{(1)} & p_0^{(1)} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ q_K^{(1)} & q_{K-1}^{(1)} & q_{K-2}^{(1)} & \dots & \dots & q_0^{(1)} \end{pmatrix} \begin{pmatrix} P_0^{(1)} \\ P_1^{(1)} \\ P_2^{(1)} \\ \vdots \\ P_K^{(1)} \end{pmatrix}$$

The relations between $\underline{P}^{(2)}$ and $\underline{P}$, the probabilities of $S(t)$ at consecutive ordering times kMa^{-1} , $(k+1)Ma^{-1}$, are given by

$$(6.10.6) \quad \underline{P}^{(2)} = \underline{p}^{(1)'} \underline{p}' \underline{P},$$

where the matrix $\underline{p}^{(1)'} \underline{p}'$ is square. For ordering times, the stationary probability distribution

$P_0, P_1, \dots, P_K$ can be found by putting $P_i^{(2)} = P_i$ for $i = 0, 1, \dots, K$, and solving the set of equations

$\underline{P} = \underline{p}^{(1)'} \underline{p}' \underline{P}$ together with $\sum_{i=0}^K P_i = 1$. This solution may not be obtained easily in an explicit form,

but it can always be computed numerically for any

given distribution $\{p_i(\theta)\}$, ($i = 0, 1, \dots$). We note,

in passing, that this stationary distribution for

ordering times will be easier to compute if we first

obtain the stationary probability distribution

$P_0^{(1)}, P_1^{(1)}, \dots, P_K^{(1)}$, for delivery times. Writing

$P_i^{(3)} = \Pr \{S((k+1)Ma^{-1} + T) = i\}$ for the probability that $S(t)$ have a value i at delivery time $(k+1)Ma^{-1} + T$, we have that the relations between $\underline{P}^{(3)}$ and $\underline{P}^{(1)}$, the probabilities of $S(t)$ at consecutive delivery times $kMa^{-1} + T$, $(k+1)Ma^{-1} + T$, is given by

$$(6.10.7) \quad \underline{P}^{(3)} = \underline{p}' \underline{p}^{(1)'} \underline{P}^{(1)},$$

where the matrix $\underline{p}' \underline{p}^{(1)'}$ is triangular. The stationary probability distribution for delivery times, $P_0^{(1)}$, $P_1^{(1)}$, ..., $P_K^{(1)}$, can be found by putting $P_i^{(3)} = P_i^{(1)}$ for $i = 0, 1, \dots, K$, and solving the equations $\underline{P}^{(1)} = \underline{p}' \underline{p}^{(1)'} \underline{P}^{(1)}$ together with $\sum_{i=0}^K P_i^{(1)} = 1$. For numerical computation of $\underline{P}^{(1)}$, the work is considerably lightened by the fact that $\underline{p}' \underline{p}^{(1)'}$ is triangular; once the stationary probability distribution for delivery times is obtained, the simple relation $\underline{P} = \underline{p}^{(1)'} \underline{P}^{(1)}$ will enable the stationary distribution for ordering times, P_0 , P_1 , ..., P_K to be found.

(b) A numerical example

As an illustration of the computations necessary to evaluate the stationary distributions $\{P_i\}$ and $\{P_i^{(1)}\}$ for ordering and delivery times, we construct an

example in which the consumption function $c(t)$ has, associated with it, a Poisson distribution such that

$$p_i(\theta) = e^{-a\theta} (a\theta)^i / i! \quad (i = 0, 1, \dots).$$

We choose the maximum deficit $K = 9$, the time-lag between order and delivery $T = \frac{1}{2}$, and the time-interval between orders $Ma^{-1} = 1$, while the mean consumption per unit time is $a = 4$. Then, since $T = Ma^{-1} - T = \frac{1}{2}$, we have that the p_i and $p_i^{(1)}$ defined previously are equal,

$$p_i = p_i^{(1)} = e^{-2} 2^i / i! \quad (i = 0, 1, \dots).$$

From Molina's tables for 'Poisson's Exponential Binomial Limit', we obtain

$p_0 = p_0^{(1)} = 0.135335$	$q_0 = q_0^{(1)} = 1.000000,$
$p_1 = p_1^{(1)} = 0.270671$	$q_1 = q_1^{(1)} = 0.864665,$
$p_2 = p_2^{(1)} = 0.270671$	$q_2 = q_2^{(1)} = 0.593994,$
$p_3 = p_3^{(1)} = 0.180447$	$q_3 = q_3^{(1)} = 0.323324,$
(6.10.8) $p_4 = p_4^{(1)} = 0.090224$	$q_4 = q_4^{(1)} = 0.142877,$
$p_5 = p_5^{(1)} = 0.036089$	$q_5 = q_5^{(1)} = 0.052653,$
$p_6 = p_6^{(1)} = 0.012030$	$q_6 = q_6^{(1)} = 0.016564,$
$p_7 = p_7^{(1)} = 0.003437$	$q_7 = q_7^{(1)} = 0.004534,$
$p_8 = p_8^{(1)} = 0.000859$	$q_8 = q_8^{(1)} = 0.001097$
	$q_9 = q_9^{(1)} = 0.000237$

The equations giving the stationary distribution $\{P_i^{(1)}\}$ for delivery times are $\underline{P}^{(1)} = \underline{p}' \underline{p}^{(1)} \underline{P}$, together with $\sum_{i=0}^K P_i^{(1)} = 1$; these, in the case of the $\{p_i\}$ and $\{q_i\}$ given in (6.10.8), result in eleven equations, not all independent, for ten unknowns:

$$\begin{aligned}
 & -0.864460P_0^{(1)} + 0.136284P_1^{(1)} + 0.139256P_2^{(1)} + 0.149657P_3^{(1)} \\
 & + 0.180862P_4^{(1)} + 0.258876P_5^{(1)} + 0.414902P_6^{(1)} + 0.645941P_7^{(1)} \\
 & + 0.882981P_8^{(1)} + P_9^{(1)} = 0, \\
 & 0.271117P_0^{(1)} - 0.727584P_1^{(1)} + 0.276590P_2^{(1)} + 0.287625P_3^{(1)} \\
 & + 0.310012P_4^{(1)} + 0.339183P_5^{(1)} + 0.343934P_6^{(1)} + 0.270671P_7^{(1)} \\
 & + 0.117019P_8^{(1)} = 0, \\
 & 0.271486P_0^{(1)} + 0.273334P_1^{(1)} - 0.722144P_2^{(1)} + 0.285591P_3^{(1)} \\
 & + 0.290341P_4^{(1)} + 0.270671P_5^{(1)} + 0.197408P_6^{(1)} + 0.080388P_7^{(1)} = 0, \\
 & 0.181348P_0^{(1)} + 0.182615P_1^{(1)} + 0.183837P_2^{(1)} - 0.819553P_3^{(1)} \\
 & + 0.160777P_4^{(1)} + 0.111934P_5^{(1)} + 0.043756P_6^{(1)} = 0, \\
 & 0.090630P_0^{(1)} + 0.090224P_1^{(1)} + 0.086834P_2^{(1)} + 0.075304P_3^{(1)} \\
 & - 0.949117P_4^{(1)} + 0.019336P_5^{(1)} = 0, \\
 & 0.035683P_0^{(1)} + 0.033922P_1^{(1)} + 0.028904P_2^{(1)} + 0.019136P_3^{(1)} \\
 & + 0.007126P_4^{(1)} - P_5^{(1)} = 0, \\
 & 0.011129P_0^{(1)} + 0.009368P_1^{(1)} + 0.006111P_2^{(1)} + 0.002242P_3^{(1)} \\
 & - P_6^{(1)} = 0, \\
 & 0.002623P_0^{(1)} + 0.001692P_1^{(1)} + 0.000614P_2^{(1)} \\
 & - P_7^{(1)} = 0,
 \end{aligned}$$

$$\begin{array}{rcccc}
0.000413P_0^{(1)} + 0.000148P_1^{(1)} & & & & \\
& -P_8^{(1)} & & & = 0, \\
0.000032P_0^{(1)} & & & & \\
& & -P_9^{(1)} & & = 0, \\
P_0^{(1)} & +P_1^{(1)} & +P_2^{(1)} & +P_3^{(1)} & \\
+P_4^{(1)} & +P_5^{(1)} & +P_6^{(1)} & +P_7^{(1)} & \\
+P_8^{(1)} & +P_9^{(1)} & & & = 1.
\end{array}$$

Solving these equations by straightforward elimination, which is greatly simplified by the fact that $\underline{p}'\underline{p}^{(1)'} is triangular, we obtain the following values for $P_i^{(1)}$:$

$$\begin{array}{ll}
P_0^{(1)} = 0.148685 & P_5^{(1)} = 0.026833, \\
P_1^{(1)} = 0.281325 & P_6^{(1)} = 0.006382, \\
(6.10.9) \quad P_2^{(1)} = 0.277080 & P_7^{(1)} = 0.001036, \\
P_3^{(1)} = 0.177621 & P_8^{(1)} = 0.000103, \\
P_4^{(1)} = 0.080930 & P_9^{(1)} = 0.000005.
\end{array}$$

From these we can compute the stationary distribution $\{P_i\}$ for ordering times, using the relations between $\underline{P}$ and $\underline{p}^{(1)}$ which are given in the matrix form $\underline{P} = \underline{p}^{(1)'}\underline{p}^{(1)}$. This is easier than solving directly the equations $\underline{P} = \underline{p}^{(1)'}\underline{p}'\underline{P}$ and $\sum_{i=0}^{\infty} P_i = 1$, where the matrix $\underline{p}^{(1)'}\underline{p}'$ is square. We obtain for $\{P_i\}$ the values:

$$\begin{array}{ll}
 P_0 = 0.020122 & P_5 = 0.154360, \\
 P_1 = 0.078318 & P_6 = 0.099024, \\
 P_2 = 0.153890 & P_7 = 0.053655, \\
 P_3 = 0.202012 & P_8 = 0.025004, \\
 P_4 = 0.198206 & P_9 = 0.015410.
 \end{array}$$

(c) Comparison with Pitt's results for an infinite deficit

It is interesting to compare the stationary distribution of the deficit $S(t)$ in the present case, where it is finite and restricted to the range $0 \leq S(t) \leq K$, with that obtained by Pitt for an infinite deficit with the same types of consumption and replacement functions. Pitt's problem for the infinite deficit, when the consumption is of Poisson type, and the replacement function is the $r_2(t)$ defined in (6.2.4), is the exact analogue of the problem of the infinite dam considered in §6.7. With a slight change in the notation of (6.7.2) to avoid confusion, we write the stationary distribution of the deficit $S(t)$ when the range is $0 \leq S(t) < \infty$, as $\{\pi_i\}$ ($i = 0, 1, \dots$), where

$$(6.10.10) \quad \pi_i = \sum_{v=0}^i p_{i-v} Q_v.$$

Here, the $p_{i-v} = e^{-aT} (aT)^{i-v} / (i-v)!$, and the Q_v are given by

$$\begin{aligned} Q_v &= M^{-1} \int_0^M (e^{-u} u^v / v!) du \\ &= M^{-1} \sum_{i=v+1}^{\infty} e^{-M} M^i / i! . \end{aligned}$$

We proceed to obtain a similar stationary distribution, also averaged over the time-interval Ma^{-1} , for the present case of the finite deficit, and we shall write it $\{\pi_i^{(1)}\}$ ($i = 0, 1, \dots, K$). To do this we consider the time-interval θ , where $\theta \leq Ma^{-1}$ such that $kMa^{-1} + T + \theta$ is any time between two deliveries at times $kMa^{-1} + T$, and $(k+1)Ma^{-1} + T$. Let the probabilities that there be consumption i in this interval be

$$p_i(\theta) = e^{-a\theta} (a\theta)^i / i! ,$$

then if $\{P_i(kMa^{-1} + T + \theta)\}$ represent the distribution of the deficit $S(t)$ at the time $kMa^{-1} + T + \theta$, this will be given in matrix form by

$$\begin{pmatrix} P_0(kMa^{-1} + T + \theta) \\ P_1(kMa^{-1} + T + \theta) \\ \dots \\ P_{K-1}(kMa^{-1} + T + \theta) \\ P_K(kMa^{-1} + T + \theta) \end{pmatrix} = \begin{pmatrix} p_0(\theta) & 0 & \dots & \dots & 0 \\ p_1(\theta) & p_0(\theta) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ p_{K-1}(\theta) & p_{K-2}(\theta) & \dots & \dots & p_0(\theta) & 0 \\ q_K(\theta) & q_{K-1}(\theta) & \dots & q_1(\theta) & q_0(\theta) \end{pmatrix} \begin{pmatrix} P_0^{(1)} \\ P_1^{(1)} \\ \dots \\ P_{K-1}^{(1)} \\ P_K^{(1)} \end{pmatrix}$$

where $q_j(\theta) = \sum_{i=j}^{\infty} p_i(\theta)$, and $P_0^{(1)}, P_1^{(1)}, \dots, P_K^{(1)}$, are the stationary probabilities of the deficit at delivery times. We can write these in the forms

$$P_i(kMa^{-1} + T + \theta) = \sum_{v=0}^i P_{i-v}^{(1)} p_v(\theta) \quad (i = 0, 1, \dots, K-1),$$

$$P_K(kMa^{-1} + T + \theta) = \sum_{v=0}^K P_{K-v}^{(1)} q_v(\theta).$$

The stationary distribution $\{\pi_i^{(1)}\}$ will then be obtained from the $P_i(kMa^{-1} + T + \theta)$ by averaging them over the interval of time Ma^{-1} , so that

$$\begin{aligned} \pi_i^{(1)} &= a M^{-1} \int_0^{Ma^{-1}} \sum_{v=0}^i P_{i-v}^{(1)} p_v(\theta) d\theta \\ &= \sum_{v=0}^i P_{i-v}^{(1)} Q_v \quad (i = 0, 1, \dots, K-1), \end{aligned}$$

$$\begin{aligned} (6.10.11) \quad \pi_K^{(1)} &= a M^{-1} \int_0^{Ma^{-1}} \sum_{v=0}^K P_{K-v}^{(1)} q_v(\theta) d\theta \\ &= \sum_{v=0}^K P_{K-v}^{(1)} \sum_{j=v}^{\infty} Q_j, \end{aligned}$$

where there is some similarity of form with (6.10.10).

In the case of our numerical example in §6.10(b), where $K = 9$, $T = \frac{1}{2}$, $M = a = 4$, the values of the p_{i-v} in (6.10.10) are given by equation (6.10.8) for values of $i = 0, 1, \dots, 8$; the values of the Q_v in (6.10.10) and (6.10.11) are given for $v = 0, 1, \dots, 8$ by

$$Q_0 = 0.254421 \quad Q_3 = 0.141633 \quad Q_6 = 0.027669,$$

$$\begin{aligned}
 Q_1 &= 0.227106 & Q_4 &= 0.092791 & Q_7 &= 0.027669, \\
 Q_2 &= 0.190474 & Q_5 &= 0.053718 & Q_8 &= 0.005341;
 \end{aligned}$$

and the values of the $\{P_i^{(1)}\}$ are given by (6.10.9).

Using these, we compute

$$\begin{aligned}
 \pi_0 &= 0.033214 & \pi_0^{(1)} &= 0.036490, \\
 \pi_1 &= 0.097164 & \pi_1^{(1)} &= 0.102810, \\
 \pi_2 &= 0.153677 & \pi_2^{(1)} &= 0.160212, \\
 \pi_3 &= 0.176480 & \pi_3^{(1)} &= 0.181162, \\
 \pi_4 &= 0.165573 & \pi_4^{(1)} &= 0.166619, \\
 \pi_5 &= 0.134440 & \pi_5^{(1)} &= 0.132132, \\
 \pi_6 &= 0.097291 & \pi_6^{(1)} &= 0.093169, \\
 \pi_7 &= 0.063731 & \pi_7^{(1)} &= 0.059328, \\
 \pi_8 &= 0.038132 & \pi_8^{(1)} &= 0.034385, \\
 \sum_{i=9}^{\infty} \pi_i &= 0.040298 & \pi_9^{(1)} &= 0.033693;
 \end{aligned}$$

there would appear to be a close similarity in the values of the $\{\pi_i\}$ and $\{\pi_i^{(1)}\}$, so that we could use the more easily computed values of π_i as approximations for the $\pi_i^{(1)}$; however, it is not possible without further work to state within what bounds this approximation would be valid.

11. Provisioning with orders at fixed times

$t = kMa^{-1}$, and replacements at fixed times

$t = (k+1)Ma^{-1}$ ($k=1,2,\dots$) with time-lag Ma^{-1}

- (a) General equations, and an exact solution for the case of consumption with a discrete geometric distribution

This is a special case of the previous problem §6.10.(a), when the times for delivery of replacements and for the ordering coincide; that is, a new order for replacements is put in every time a delivery occurs, at time-intervals Ma^{-1} . The consumption function is of the same form (6.9.1), the replacement function is $r_2(t)$ of (6.2.4), the overflow function is defined by (6.10.1), where the time-lag T is Ma^{-1} ; in these conditions, with the same notation as before, the equations (6.10.3) relate the probability distributions $\{P_i\}$ and $\{P_i^{(1)}\}$ for $S(t)$ at consecutive ordering and replacement times kMa^{-1} , and $(k+1)Ma^{-1}$ ($k=1,2,\dots$). To obtain the stationary probability distribution $P_0, P_1, \dots, P_K$, for these times, we write $P_i^{(1)} = P_i$ ($i=0,1,\dots,K$), and solve the equations $\underline{P} = \underline{p}'\underline{P}$ together with $\sum_{i=0}^K P_i = 1$.

$$\begin{aligned}
 (6.11.2) \quad P_i &= \left\{ p_i \left(1 - \sum_{n=K-i+1}^K P_n \right) + q_{i+1} p_{K-i} \sum_{n=0}^{i-1} P_n \right\} (1 - q_{i+1} q_{K-i})^{-1} \\
 P_{K-i} &= p_{K-i} \sum_{n=0}^{i-1} P_n + q_{K-i} P_i \quad (i = 1, \dots, K-1).
 \end{aligned}$$

A distribution $\{p_i\}$ which gives a simple explicit solution of these equations is the geometric; it is important to note, however, that this does not satisfy the independence condition for non-overlapping intervals of time postulated in (6.9.1). For the geometric distribution

$$(6.11.3) \quad p_i = AB^i \quad (i = 0, 1, \dots),$$

where $A + B = 1$, we find from the recurrence relations (6.11.2) that the stationary distribution of the deficit, $\{P_i\}$ is given by

$$(6.11.4) \quad P_i = AB^i (1 - B^{K+1})^{-1} \quad (i = 0, 1, \dots, K).$$

- (b) An exact solution for the case of consumption with a continuous exponential distribution

Suppose that instead of dealing with integral units of the deficit, consumption and replacement, we deal

with units of magnitude Δ ; Δ will tend to zero, and K and i will tend to infinity in such a way that the maximum deficit $K\Delta$ tends to the value k , and a deficit, consumption or replacement $i\Delta$ tends to the value s , ($i = 0, 1, \dots, K$). The distribution of the consumption then becomes continuous; for if we put

$$A = 1 - e^{-\mu\Delta}$$

$$B = e^{-\mu\Delta},$$

and let $\Delta \rightarrow 0$ in (6.11.3), so that $i\Delta \rightarrow s$, we have that the distribution of the consumption in the interval of time Ma^{-1} is

$$\begin{aligned} p(s)ds &= \lim_{\Delta \rightarrow 0} \Delta^{-1} p_i ds \\ (6.11.5) \quad &= \lim_{\Delta \rightarrow 0} \Delta^{-1} e^{-\mu i \Delta} (1 - e^{-\mu \Delta}) ds \\ &= \mu e^{-\mu s} ds, \end{aligned}$$

which is continuous, and of the exponential form. The solutions (6.11.4) of equations (6.11.1) will lead by the same limiting process to an exact solution of the following provisioning problem.

A deficit $S(t)$ is defined continuously in the range $0 \leq S(t) \leq k$, at the fixed times nMa^{-1} ($n = 0, 1, \dots$), and is subject to the following conditions:

(1) a consumption function $c(t)$ such that in a time-interval Ma^{-1} the distribution of the consumption is continuous and of the exponential type

$$\begin{aligned} p(s)ds &= \Pr\{s < c((n+1)Ma^{-1}) - c(nMa^{-1}) < s+ds\} \\ &= \mu e^{-\mu s} ds; \end{aligned}$$

(2) a replacement function such that

$$r_2((n+1)Ma^{-1}) = c(nMa^{-1}),$$

where irregular quantities lying between the values $(s, s+ds)$ equal to the consumption in the previous time-interval Ma^{-1} are ordered at times nMa^{-1} and delivered and added to the stock at times $(n+1)Ma^{-1}$;

(3) an overflow function such that in any time-interval Ma^{-1} between ordering and delivery times nMa^{-1} , $(n+1)Ma^{-1}$, the consumption is lost at any time after the deficit rises to the value k , and is otherwise met from stock.

For small Δ , the equations (6.11.1) hold, with the proviso that $\{p_i\}$ represent the probabilities that the consumption in an interval Ma^{-1} is $i\Delta$, and $\{P_i\}$ the stationary probabilities that the deficit have a value $i\Delta$, so that the solutions (6.11.4) to them also hold. In the limit as Δ tends to zero in the manner defined above, the continuous stationary distribution

of the deficit $S(t)$ is given by

$$\begin{aligned}
 (6.11.6) \quad f(s)ds &= \lim_{\Delta \rightarrow 0} \Delta^{-1} p_1 ds \\
 &= \mu e^{-\mu s} (1 - e^{-\mu k})^{-1} ds \quad (0 \leq s \leq k),
 \end{aligned}$$

a truncated exponential distribution.

This solution can also be obtained by taking the limit as Δ tends to zero for equations (6.11.1); this gives an integral equation for the continuous stationary probability distribution of the deficit,

$$\begin{aligned}
 (6.11.7) \quad f(s)ds &= p(s)ds \int_0^{k-s} f(u)du \\
 &\quad + f(k-s)ds \int_s^\infty p(u)du.
 \end{aligned}$$

It is easily verified that (6.11.6) will satisfy this equation.

Chapter 7

AN EXACT SOLUTION OF THE STORAGE PROBLEM WITH
A POISSON INPUT1. Introduction

In attempting to find algebraic solutions for the stationary probability distribution $\{P_i\}$ of the dam content $Z(t)$ in Moran's problem (§6.3), where $\underline{P} = \underline{p}'\underline{P}$, $\underline{P}$ being the column vector of the P_i , it is found that no simple solution in an explicit form exists for an input function of the Poisson type with parameter a , such that in the interval of time $t, t+1$, the probabilities of an input $X(t)$ being i units is

$$p_i = e^{-a} a^i / i! \quad (i = 0, 1, \dots).$$

However, if in this problem the conditions for the input of Poisson type and for the overflow are left unchanged, but the release rule is altered so that

instead of the release rule in §6.3 (M units at times $t+1$ if the total content of the dam $Z(t)+X(t)$ is greater or equal to M , or the total content of the dam $Z(t)+X(t)$ if this is less than M), a new release rule is prescribed so that there is an output with a steady flow when the dam contains water, and no output when it is empty, then $Z(t)$ is continuous, and it is possible to obtain solutions in an explicit form for its stationary distribution.

Our method of approach will be to consider Moran's set of equations (6.3.1) for an infinitesimal interval of time δt , instead of a unit interval of time; the input will remain of Poisson type, and be a multiple of a definite discrete unit which will not tend to zero, while a small discrete release will be made at the end of each infinitesimal interval of time under a rule of the same type as that in §6.3. The equations can then be solved explicitly, and when $\delta t \rightarrow 0$, these solutions will provide the continuous stationary probability distribution for the dam content $Z(t)$, which is then itself continuous, when the release rule prescribed gives an output with a steady flow when the dam contains water, or no output when the dam is empty.

2. The discrete model

Before proceeding further, it is preferable, since the theory of dams was identified in §6.4 with the general storage theory, to view the problem as one of the general storage function. It then becomes possible to interpret the problem as one of the theory of provisioning as well; here we would have a consumption of Poisson type, a replacement rule such that there is continuous delivery at a steady rate when the stock is not at its maximum value, and zero when it is, and an overflow function such that the consumption is lost when there is no stock to meet it. In practice, this could conceivably be the case of a stock of grain in a silo, where the replacement is a steady flow of grain which is stopped when the silo is full, and the consumption of Poisson type is lost when the silo is empty, and met from the stock otherwise.

We frame the conditions of the problem for the storage function as follows: a discrete storage function $S(t)$ is defined in the range $0 \leq S(t) \leq (K-1)\Delta$, where $K = bH + U$ (b, H, U integers) at the fixed times $t = 0, \delta t, 2\delta t, \dots$, and is subject to the conditions that



(1) it is fed by a discrete input function $I(t)$ of Poisson type with parameter a , such that in a small interval of time δt there may be inputs of H or of no units respectively with probabilities

$$\begin{aligned} p &= \Pr\{I(t+\delta t) - I(t) = H\Delta\} = 1 - e^{-a\delta t}, \\ q &= \Pr\{I(t+\delta t) - I(t) = 0\} = e^{-a\delta t}; \end{aligned}$$

(2) the output function is such that at fixed times $(n+1)\delta t$ ($n = 0, 1, \dots$), the discrete output consists of a quantity $\Delta = C\delta t$ units, where C is a constant, provided that the storage function at a time just previous to $(n+1)\delta t$ is greater than Δ ,

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) \geq \Delta,$$

but is zero otherwise;

(3) the overflow function is such that in the interval $n\delta t$, $(n+1)\delta t$ there is no overflow if

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) \leq K\Delta,$$

but there is an overflow

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) - K\Delta,$$

if this is greater than zero.

Writing P_i , $P_i^{(1)}$ for the probabilities that $S(t)$ take the values $i \Delta$ at times $n\delta t$, $(n+1)\delta t$ respectively, where $i = 0, 1, \dots, K-1$, we have that the relations between these probabilities are

$$\begin{aligned} P_0^{(1)} &= qP_0 + qP_1 \\ P_1^{(1)} &= \dots + qP_2 \end{aligned}$$

.....

$$P_{H-2}^{(1)} = \dots + qP_{H-1}$$

$$\begin{aligned} P_{H-1}^{(1)} &= pP_0 + \dots + qP_H \\ P_H^{(1)} &= \dots + pP_1 + \dots + qP_{H+1} \end{aligned}$$

.....

$$P_{2H-2}^{(1)} = \dots + pP_{H-1} + \dots + qP_{2H-1}$$

.....

(7.2.1)

$$P_{jH-1}^{(1)} = \dots + pP_{(j-1)H} + \dots + qP_{jH}$$

.....

$$P_{(j+1)H-2}^{(1)} = \dots + pP_{jH-1} + \dots + qP_{(j+1)H-1}$$

.....

$$P_{bH-1}^{(1)} = \dots + pP_{(b-1)H} + \dots + qP_{bH}$$

.....

$$P_{K-1}^{(1)} = P_{bH+U-1}^{(1)} = p(P_{(b-1)H+U} + \dots + P_{bH+U-1}),$$



or in matrix form, $\underline{P}^{(1)} = \underline{p}'\underline{P}$, where $\underline{P}$, $\underline{P}^{(1)}$ are column vectors with elements P_i , $P_i^{(1)}$ respectively, and $\underline{p}'$ is the matrix of coefficients, $\underline{p}$ being a stochastic matrix.

Formally, this is Moran's set of equations (6.3.1) in which we have set $p_0 = q$, $p_H = p$, and all other $p_i = 0$ for i not equal to 0 or H . For the stationary distribution of $S(t)$, we put $P_i^{(1)} = P_i$, and solve the equations $\underline{P} = \underline{p}'\underline{P}$, together with $\sum_{i=0}^{K-1} P_i = 1$. We see clearly that the solution of these equations will fall into the $(b+1)$ distinct classes:

- (0)th class: $P_1, P_2, \dots, P_{H-1}$;
- (1)th class: $P_H, P_{H+1}, \dots, P_{2H-1}$;
-
- (j)th class: $P_{jH}, P_{jH+1}, \dots, P_{(j+1)H-1}$;
-
- (b)th class: $P_{bH}, P_{bH+1}, \dots, P_{K-1}$.

The solution for the (0)th class is immediately obvious from (7.2.1),

$$(7.2.2) \quad P_R = pP_0 q^{-R} \quad (R = 1, 2, \dots, H-1);$$

for the (j)th class, we obtain from our equations (7.2.1) the equation

$$(7.2.3) \quad P_{jH+R} = q^{-(R+1)} \left\{ P_{jH-1} - p \sum_{N=0}^R q^N P_{(j-1)H+N} \right\} \\ (j = 1, \dots, b; R = 0, \dots, H-1),$$

relating the probability P_{jH+R} in the (j) th class with the probabilities $P_{(j-1)H+N}$, P_{jH-1} of the $(j-1)$ th class, so that starting with the solutions (7.2.2) of the (0) th class, we can obtain solutions progressively for one class after another.

Consider the formula

$$(7.2.4) \quad P_{jH+R} = p P_0 q^{-(jH+R)} \left[\sum_{v=0}^j (-1)^v p^{v-1} q^{v(H-1)} \right. \\ \left. \times \left\{ \binom{(j-v)H+R+v-1}{v-1} + p \binom{(j-v)H+R+v-1}{v} \right\} \right]$$

where $R = 1, 2, \dots, H-1$ for $j = 0$, and $R = 0, 1, \dots, H-1$ for $j = 1, 2, \dots, b$, and where it is understood that $\binom{n}{r} = 0$ if $r < 0$, or if $n < r$. For $j = 0$, this gives the solution (7.2.2) for the (0) th class. We prove that formula (7.2.4) holds for all values of j by induction; assume that it holds for the $(j-1)$ th class, then substituting for $P_{(j-1)H+N}$ and P_{jH-1} given by (7.2.4) in the equation (7.2.3), we obtain for the value of the probability P_{jH+R}

$$\begin{aligned}
 P_{jH+R} = & q^{-(R+1)} \left[p P_0 q^{-(jH-1)} \left(\sum_{v=0}^{j-1} (-1)^v p^{v-1} q^{v(H-1)} \left\{ \binom{(j-v)H+v-2}{v-1} + p \binom{(j-v)H+v-2}{v} \right\} \right) \right. \\
 & \left. - p^2 P_0 q^{-(j-1)H} \sum_{N=0}^R \left(\sum_{v=0}^{j-1} (-1)^v p^{v-1} q^{v(H-1)} \left\{ \binom{(j-1-v)H+N+v-1}{v-1} + p \binom{(j-1-v)H+N+v-1}{v} \right\} \right) \right]
 \end{aligned}$$

(7.2.5)

$$\begin{aligned}
 P_{jH+R} = & p P_0 q^{-(jH+R)} \left[1 + \sum_{v=1}^{j-1} (-1)^v p^{v-1} q^{v(H-1)} \left\{ \binom{(j-v)H+v-2}{v-1} + \sum_{N=0}^R \binom{(j-v)H+N+v-2}{v-2} \right\} \right. \\
 & \left. + p \left\{ \binom{(j-v)H+v-2}{v} + \sum_{N=0}^R \binom{(j-v)H+N+v-2}{v-1} \right\} \right] \\
 & + (-1)^j p^{j-1} q^{j(H-1)} \left\{ \sum_{N=0}^R \binom{N+j-2}{j-2} + p \sum_{N=0}^R \binom{N+j-2}{j-1} \right\}
 \end{aligned}$$

Now in (7.2.5), consider

$$\left\{ \binom{(j-v)H+v-2}{v-1} + \sum_{N=0}^R \binom{(j-v)H+N+v-2}{v-2} \right\} ;$$

since

$$\sum_{t=0}^{\infty} \binom{t}{w} = \sum_{t=w}^{\infty} \binom{t}{w} = \binom{\infty+1}{w+1},$$

then

$$\left\{ \binom{(j-v)H+v-2}{v-1} + \sum_{N=0}^R \binom{(j-v)H+N+v-2}{v-2} \right\} =$$

$$\begin{aligned}
&= \left\{ \binom{(j-v)H+v-2}{v-1} + \sum_{N=-(j-v)H}^R \binom{(j-v)H+N+v-2}{v-2} - \sum_{N=-(j-v)H}^{-1} \binom{(j-v)H+N+v-2}{v-2} \right\} \\
&= \binom{(j-v)H+R+v-1}{v-1}.
\end{aligned}$$

In exactly the same way, we reduce the other terms of (7.2.5), so that

$$\left\{ \binom{(j-v)H+v-2}{v} + \sum_{N=0}^R \binom{(j-v)H+N+v-2}{v-1} \right\} = \binom{(j-v)H+R+v-1}{v},$$

and

$$\sum_{N=0}^R \binom{N+j-2}{j-2} + p \sum_{N=0}^R \binom{N+j-2}{j-1} = \binom{R+j-1}{j-1} + p \binom{R+j-1}{j}.$$

We thus obtain for P_{jH+R} in (7.2.5) the following expression

$$\begin{aligned}
(7.2.6) \quad P_{jH+R} &= p P_0 q^{-(jH+R)} \\
&\times \left[1 + \sum_{v=1}^j (-1)^v p^{v-1} q^{v(H-1)} \left\{ \binom{(j-v)H+R+v-1}{v-1} + p \binom{(j-v)H+R+v-1}{v} \right\} \right]
\end{aligned}$$

which is equivalent to (7.2.4). Now, since this has been shown to hold for the (0)th class, then it will hold for the (1)th class, and so on up to class (b); this proves that (7.2.4) or its slightly different form (7.2.6) is the solution required for the stationary

distribution of $S(t)$ in the discrete case considered.

We write out in full the formulae for a few classes so as to give an idea of their structure:

$$P_R = p P_0 q^{-R} \quad (R = 1, 2, \dots, H-1);$$

$$P_{H+R} = p P_0 q^{-(H+R)} \left\{ 1 - q^{H-1} (1 + p(R)) \right\},$$

$$P_{2H+R} = p P_0 q^{-(2H+R)} \left\{ 1 - q^{H-1} (1 + p(R)) + p q^{2(H-1)} \left(\binom{R+1}{1} + p \binom{R+1}{2} \right) \right\},$$

$$P_{3H+R} = p P_0 q^{-(3H+R)} \left\{ 1 - q^{H-1} (1 + p(R)) + p q^{2(H-1)} \left(\binom{H+R+1}{1} + p \binom{H+R+1}{2} \right) - p^2 q^{3(H-1)} \left(\binom{R+2}{2} + p \binom{R+2}{3} \right) \right\} \quad (R = 0, 1, \dots, H-1).$$

For the complete solution in any particular case, the value of P_0 is also required; this can be found from the equation $\sum_{i=0}^{K-1} P_i = 1$.

3. The continuous analogue

We now proceed to the continuous case by letting $\delta t \rightarrow 0$, and consequently $\Delta = C \delta t \rightarrow 0$, where the constant C can now be interpreted as the rate of output per unit time. We allow Δ to tend to zero, and the previously defined integers K , H , and R to tend to infinity in such a way that

(1) the maximum value of the storage function $K\Delta \rightarrow k$,

or $(bH+U)\Delta \rightarrow bh+u$,

(2) the input $H\Delta \rightarrow h$,

(3) the values of the storage function

$(jH+R)\Delta \rightarrow jh+r$, which we shall also write as

$(j = 0, 1, \dots, b; 0 \leq r < h)$.

As a result of this limiting process, the conditions of the problem which we have just considered for the storage function are changed; the function $S(t)$ has now a continuous range of values, $0 \leq S(t) < k$, where $k = bh+u$, and is defined for all time t (t continuous) in such a way that

(1) it is fed by a discrete input of Poisson type with parameter a , so that in a small interval of time δt there may be inputs of h or of no units with probabilities

$$\begin{aligned} p &= \Pr\{I(t+\delta t) - I(t) = h\} = 1 - e^{-a\delta t} = 1 - e^{-a\Delta/C} \\ &= 1 - e^{-\mu\Delta} \\ q &= \Pr\{I(t+\delta t) - I(t) = 0\} = e^{-a\delta t} = e^{-a\Delta/C} \\ &= e^{-\mu\Delta} \end{aligned}$$

where $\mu = a/C$;

(2) the output is such that it has a constant rate C per unit time, provided the storage function is not

zero, but is zero if the storage function is zero;

(3) the overflow function is such that in any interval of time δt there is no overflow if

$$S(t) + I(t + \delta t) - I(t) \leq k$$

but there is an overflow

$$S(t) + I(t + \delta t) - I(t) > k$$

if this is greater than zero.

We write the continuous stationary probability density for $S(t)$ as $f(s)$, so that

$$f(s)ds = \Pr\{s < S(t) < s + ds\}.$$

Now for $s = \lim_{\Delta \rightarrow 0} (jH + R)\Delta \rightarrow jh + r$, we have that

$$f(s)ds = \lim_{\Delta \rightarrow 0} P_{jH+R} \Delta^{-1} ds;$$

this enables us to obtain from (7.2.2) and (7.2.6) the equations for $f(s)$ in various ranges of magnitude h . For $j = 0$ ($0 < r < h$) we have

$$\begin{aligned} (7.3.1) \quad f(r) &= \lim_{\Delta \rightarrow 0} \Delta^{-1} P_0 q^{-R} = \lim_{\Delta \rightarrow 0} P_0 \Delta^{-1} (1 - e^{-\mu\Delta}) e^{\mu\Delta R} \\ &= \mu e^{\mu r} P_0. \end{aligned}$$

For $j = 1, 2, \dots, b$ and $0 \leq r < h$, we have

$$\begin{aligned}
 f(jh+r) &= \lim_{\Delta \rightarrow 0} \Delta^{-1} p P_0 q^{-(jH+R)} \left[1 + \sum_{v=1}^j (-1)^v p^{v-1} q^{v(H-1)} \right. \\
 &\quad \left. \times \left\{ \binom{(j-v)H+R+v-1}{v-1} + p \binom{(j-v)H+R+v-1}{v} \right\} \right] \\
 (7.3.2) \quad &= \lim_{\Delta \rightarrow 0} \Delta^{-1} (1 - e^{-\mu \Delta}) e^{\mu \Delta (jH+R)} P_0 \\
 &\quad \times \left[1 + \sum_{v=1}^j (-1)^v e^{-\mu v \Delta (H-1)} \left\{ (\mu \Delta)^{v-1} \binom{(j-v)H+R+v-1}{v-1} + (\mu \Delta)^v \binom{(j-v)H+R+v-1}{v} \right\} \right] \\
 &= \mu e^{\mu(jh+r)} P_0 \left[1 + \sum_{v=1}^j (-1)^v e^{-\mu v h} \left\{ \mu^{v-1} \frac{((j-v)h+r)^{v-1}}{(v-1)!} + \mu^v \frac{((j-v)h+r)^v}{v!} \right\} \right]
 \end{aligned}$$

In addition to these probability densities, which are continuous in the intervals in which they are defined, there is also a concentration of probability P_0 at $s = 0$. To complete the solution, we require the value of P_0 ; this will be given in any particular case where the value of $k = bh + u$ is known by

$$(7.3.3) \quad P_0 + \sum_{j=0}^{b-1} \int_{r=0}^h f(jh+r) dr + \int_{r=0}^u f(bh+r) dr = 1.$$

We write out in full some of the formulae for the $f(jh+r)$, in order to indicate their structure:

$$\begin{aligned}
 f(r) &= \mu e^{\mu r} P_0 \quad (0 < r < h), \\
 f(h+r) &= \mu e^{\mu(h+r)} P_0 \{ 1 - e^{-\mu h} (1 + \mu r) \}, \\
 (7.3.4) \quad f(2h+r) &= \mu e^{\mu(2h+r)} P_0 \left\{ 1 - e^{-\mu h} (1 + \mu(h+r)) + \mu e^{-2\mu h} \left(r + \mu \frac{r^2}{2!} \right) \right\}, \\
 f(3h+r) &= \mu e^{\mu(3h+r)} P_0 \left\{ 1 - e^{-\mu h} (1 + \mu(2h+r)) + \mu e^{-2\mu h} \left((h+r) + \mu \frac{(h+r)^2}{2!} \right) \right. \\
 &\quad \left. - \mu^2 e^{-3\mu h} \left(\frac{r^2}{2!} + \mu \frac{r^3}{3!} \right) \right\} \\
 &\quad (0 \leq r < h).
 \end{aligned}$$

It is interesting to note that these formulae for $f(jh+r)$ ($j = 1, 2, \dots, b$) can also be obtained from the limit of equation (7.2.3), when $\Delta \rightarrow 0$,

$$\begin{aligned}
 (7.3.5) \quad f(jh+r) &= \lim_{\Delta \rightarrow 0} P_{jh+r} \Delta^{-1} \\
 &= \lim_{\Delta \rightarrow 0} q^{-(R+1)} \left\{ P_{jh-1} \Delta^{-1} - \Delta^{-1} \sum_{N=0}^R q^N (P_{(j-1)h+N} \Delta^{-1}) \Delta \right\}.
 \end{aligned}$$

Now for $j = 1$, we have

$$\begin{aligned}
 (7.3.6) \quad f(h+r) &= \lim_{\Delta \rightarrow 0} P_{h+r} \Delta^{-1} \\
 &= \lim_{\Delta \rightarrow 0} q^{-(R+1)} \left\{ P_{h-1} \Delta^{-1} - \Delta^{-1} \sum_{N=0}^R q^N (P_N \Delta^{-1}) \Delta - \Delta^{-1} P_0 \right\} \\
 &= e^{\mu r} f(h-0) - \mu e^{\mu r} \left\{ \int_0^r e^{-\mu n} f(n) dn + P_0 \right\},
 \end{aligned}$$

since $P_N \Delta^{-1} \rightarrow f(n)$, where $N\Delta \rightarrow n$ for all $N \neq 0$, but there is a concentration of probability P_0 at $s = 0$. For $j = 2, 3, \dots, b$ the equation (7.3.5) gives

$$(7.3.7) \quad f(jh+r) = e^{\mu r} f(jh-0) - \mu e^{\mu r} \int_0^r e^{-\mu n} f((j-1)h+n) dn.$$

The formulae (7.3.4) can be obtained by putting the formula (7.3.1) for $f(r)$ in (7.3.6) and integrating it to find $f(h+r)$, then repeating this process for $f(jh+r)$ by using (7.3.7). More generally, it can be proved by induction that our solution (7.3.1) and

(7.3.2) satisfy the integral equations (7.3.6) and (7.3.7).

It is found that (7.3.3) does not lend itself to the computation of P_0 as well as an alternative equation for the integral of the probabilities, which we now obtain. To do so, we write the probability density $f(s)$ for the various intervals in the following manner:

$$\begin{aligned} f(s) &= \mu e^{\mu s} P_0 \quad (0 < s < h) \\ f(s) &= \mu e^{\mu s} P_0 \left(1 + \sum_{v=1}^j (-1)^v e^{-\mu v h} \left\{ \mu^{v-1} \frac{(s-vh)^{v-1}}{(v-1)!} + \mu^v \frac{(s-vh)^v}{v!} \right\} \right) \\ &\quad (jh \leq s < (j+1)h; j = 1, 2, \dots, b). \end{aligned}$$

The equation (7.3.3) can then be written as

$$\begin{aligned} P_0 \left[1 + \int_0^{bh+u} \mu e^{\mu s} ds \right. \\ \left. + \sum_{j=1}^b (-1)^j \int_{s=jh}^{bh+u} \mu e^{\mu(s-jh)} \left\{ \mu^{j-1} \frac{(s-jh)^{j-1}}{(j-1)!} + \mu^j \frac{(s-jh)^j}{j!} \right\} ds \right] = 1, \end{aligned}$$

which shortens considerably the work involved in the computation of P_0 .

Chapter 8

THE SOLUTION OF A STORAGE PROBLEM

BY MONTE CARLO METHODS*

1. Introduction

A problem which is of interest in Markov chain theory, and has been solved in some cases for the sets of equations arising in Moran's theory of dams and in the other problems of storage considered in §§6-7, is that of finding the eigen-vector associated with the root $\mu = 1$ of the stochastic matrix for a positively regular chain. This eigen-vector, when suitably scaled, so that its elements have a sum unity, gives the set of stationary probabilities for the stochastic matrix. Explicit solutions for these may be found in some particular cases, as we have seen, but

*This chapter was written in collaboration with Professor P.A.P. Moran, who is entirely responsible for §8.4

more usually, only a numerical solution is possible. An alternative to the numerical solution of the set of linear equations involved, is the evaluation of these stationary probabilities by Monte Carlo methods.

Starting with an idea of Von Neumann and Ulam, Forsythe and Leibler (1950), Wasow (1952) and Edmundson (1953) have discussed an allied problem, that of matrix inversion by Monte Carlo methods. Curtiss (1953), more comprehensively, has considered the use of Monte Carlo methods for the solution of a class of problems in which a function u is to be estimated which satisfies the equation

$$u = L(u) + c,$$

where c and L are known; in particular, a detailed account of the application of this method to the problem of inverting matrices is given. However, the specific problem of evaluating the eigen-vector associated with the root $\mu = 1$ of the stochastic matrix for a positively regular chain, that is of obtaining the stationary probabilities for the stochastic matrix, although certainly subsumed in the general problem, is not discussed. We proceed to consider it in some detail.

2. Estimators of the stationary probabilities and their variances

Consider a Markov chain with s positive states $E_1, \dots, E_s$, and with a matrix of transition probabilities $\underline{p} = \{p_{ij}\}$. We restrict ourselves to positively regular cases, that is as we have seen in §2.4, those for which there exists an integer n_0 such that all the elements of $\underline{p}^{n_0}$ are positive. This is equivalent to saying that there exists an integer n_0 such that it is always possible for the system to change from any state E_i to any other state E_j in n_0 steps with non-zero probability. This will always be the case in the storage problems we are considering, and implies (§2.5) that the eigen-value 1 of $\underline{p}$ is simple and that all the other eigen-values have moduli less than 1.

Now consider a realisation of the chain in n trials, where the states $E_1, \dots, E_s$ appear $n_1, \dots, n_s$ times respectively, and $\sum_i n_i = n$. Then we can use

$$(8.2.1) \quad \hat{P}_i = n_i n^{-1}$$

as estimators of the elements P_i of the eigen-vector $\underline{P}$ of stationary probabilities which satisfy the equation $\underline{P} = \underline{p}'\underline{P}$. It is known that these estimators are

consistent and asymptotically normally distributed as n tends to infinity (§ 3.5). They would be unbiased if the realised chain had been started up by choosing the initial state with probability P_i , but this is just what we cannot do in an artificial experiment, as the P_i are the very quantities we wish to find. The P_i are, however, asymptotically unbiased, and we could get rid of most of the bias in practice by allowing the process to run for a little while before starting the sequence on which the estimates are based. From the point of view of theory, we may therefore assume that the process is already stationary.

Writing $X_i^{(r)} = 1$ or 0 according to whether the chain is in the state E_i at the r -th trial or not, we have

$$(8.2.2) \quad \bar{X}_i(n) = \bar{X}_i(n^{-1}) = \bar{X}_i(n^{-1}) = P_i.$$

The variance of this estimator has been obtained by Patankar (1954) and Fréchet (1952; 76-79). We have

$$\begin{aligned} V(\hat{P}_i) &= n^{-2} V(\bar{X}_i) \\ &= n^{-2} \left\{ \sum_{r=1}^n V(X_i^{(r)}) + 2 \sum_{r < s} \text{Cov}(X_i^{(r)}, X_i^{(s)}) \right\}. \end{aligned}$$

$$\begin{aligned} \text{Now } V(X_i^{(r)}) &= \bar{X}_i(X_i^{(r)})^2 - \{\bar{X}_i(X_i^{(r)})\}^2 \\ &= P_i - P_i^2, \end{aligned}$$

and

$$\begin{aligned} E(X_i^{(r)} X_i^{(t)}) &= E(X_i^{(r)} X_i^{(t)}) - E(X_i^{(r)}) E(X_i^{(t)}) \\ &= P_i p_{ii}^{(t-r-1)} - P_i^2, \end{aligned}$$

where $p_{ii}^{(t-r-1)}$ is the (i,i) -th element of the matrix p^{t-r-1} . Thus

$$\begin{aligned} (8.2.3) \quad \mathcal{V}(\hat{P}_i) &= n^{-2} \left\{ n(P_i - P_i^2) + 2 \sum_{t=r}^{n-1} (P_i p_{ii}^{(t-r-1)} - P_i^2) \right\} \\ &= n^{-2} \left\{ n(P_i - P_i^2) + 2 P_i \sum_{v=1}^{n-1} (n-v)(p_{ii}^{(v)} - P_i) \right\}. \end{aligned}$$

This is exactly analogous to the formula for the variance of the mean of a stationary stochastic process in terms of the serial correlations. In the positively regular case, $p_{ii}^{(v)}$ tends to P_i as v increases, at least as fast as the terms of a certain convergent geometric progression. For n large, $n \mathcal{V}(\hat{P}_i)$ tends asymptotically to

$$(8.2.4) \quad \lim_{n \rightarrow \infty} n \mathcal{V}(\hat{P}_i) = P_i - P_i^2 + 2 P_i \sum_{v=1}^{\infty} (p_{ii}^{(v)} - P_i),$$

which is independent of n .

Even if we insert in this formula the estimated values $\hat{P}_i$, we would still have to find the value of the infinite series. It is possible, however, to give an upper bound for the ratio of a majorising geometric series, which is based on simple calculations on the

p_{ij} (see §2.4). However, since the P_i must themselves be estimated, this does not seem worthwhile. A better method, which is used in the following example, is to split the realisation of the process up into a number of sub-sections, and estimate the variance of the whole from the variance of the observed values in the sub-sections. This suffers from the defect that successive sub-sections are not independent, but this is unlikely to bias the result seriously in most cases, and can be avoided if necessary by taking a gap between the sub-sections.

3. Application of Monte Carlo methods to a storage problem

A simple problem of storage, briefly summarised in §6.3, is that in the theory of dams considered by Moran (1954). Here, a set of equations is constructed for the case where, between times t and $t+1$ ($t=0,1,\dots$) an input of $0,1,2,\dots$, units enters a dam with probabilities $p_0, p_1, p_2,\dots$, respectively; the capacity of the dam is K units, and a release of M units or the total content of the dam is made at times $t+1$, according to whether the dam content is greater

than or equal to M units, or less than M units. The capacity K of the dam is a maximum, and any surplus overflows. For this case, we have seen that the system of equations (6.3.1) relates the probabilities $P_i^{(1)}$ at times $t+1$, and P_i at times t , that there be i units ($i = 0, 1, \dots, K-M$) in the dam, provided that $K - M \geq M$; these equations are written in the simpler matrix form $\underline{P}^{(1)} = \underline{p}' \underline{P}$, where $\underline{P}$, $\underline{P}^{(1)}$ are column vectors with elements P_i , $P_i^{(1)}$ respectively, and $\underline{p}$ is a stochastic matrix whose elements are known. The stationary distribution $\{P_i\}$ that there be i units in the dam is given by $\underline{P} = \underline{p}' \underline{P}$, together with $\sum_{i=0}^{K-M} P_i = 1$.

Moran has solved, among others, the particular case where the values of the p_i , which appear in the elements of the stochastic matrix $\underline{p}$, are given by

$$p_0 = \int_0^{\frac{1}{2}} f(x) dx; \quad p_i = \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} f(x) dx \quad (i > 0),$$

and $f(x) = \{\Gamma(3/2)\}^{-1} e^{-x} x^{\frac{1}{2}}$; the values of the capacity of the dam and of the release are $K = 9$, and $M = 3$ respectively. Solving the dam equation in this case, where the stochastic matrix is

$$\underline{p} = \begin{pmatrix} 0.6790 & 0.1085 & 0.0737 & 0.0490 & 0.0322 & 0.0208 & 0.0368 \\ 0.5164 & 0.1626 & 0.1085 & 0.0737 & 0.0440 & 0.0322 & 0.0576 \\ 0.3176 & 0.1988 & 0.1626 & 0.1085 & 0.0737 & 0.0490 & 0.0898 \\ 0.0812 & 0.2364 & 0.1988 & 0.1626 & 0.1085 & 0.0737 & 0.1388 \\ 0 & 0.0812 & 0.2364 & 0.1988 & 0.1626 & 0.1085 & 0.2125 \\ 0 & 0 & 0.0812 & 0.2364 & 0.1988 & 0.1626 & 0.3210 \\ 0 & 0 & 0 & 0.0812 & 0.2364 & 0.1988 & 0.4836 \end{pmatrix}$$

by straightforward elimination, the stationary probabilities are found to be

$$\begin{aligned} P_0 &= 0.3029 & P_2 &= 0.1060 & P_4 &= 0.1121 \\ (8.3.1) \quad P_1 &= 0.1060 & P_3 &= 0.1087 & P_5 &= 0.0846 \\ & & P_6 &= 0.1798 \end{aligned}$$

The Monte Carlo method for the solution of this dam equation consists in the construction, by means of a table of random numbers, of a sequence of states $E_i^{(t)}$ ($i = 0, 1, \dots, 6$), where $E_i^{(t)}$ indicates that the dam has a quantity of i units in it at time t . We shall use Fisher and Yates's tables of random numbers, in groups of four. The partitioning of these random numbers to fit the probability distributions forming the rows of the stochastic matrix $\underline{p}$ is achieved by associating with each cell in the row of the matrix shown below,

a range of numbers obtained from the progressive cumulative totals of the probabilities in the row.

	E_0	E_1	E_2	E_3	E_4	E_5	E_6
E_0	0.6790 (0000- 6790)	0.1085 (6791- 7875)	0.0737 (7876- 8612)	0.0490 (8613- 9102)	0.0322 (9103- 9424)	0.0208 (9425- 9632)	0.0368 (9633- 9999)
E_1	0.5164 (0000- 5164)	0.1626 (5165- 6790)	0.1085 (6791- 7875)	0.0737 (7876- 8612)	0.0490 (8613- 9102)	0.0322 (9103- 9424)	0.0576 (9425- 9999)
E_2	0.3176 (0000- 3176)	0.1988 (3177- 5164)	0.1626 (5165- 6790)	0.1085 (6791- 7875)	0.0737 (7876- 8612)	0.0490 (8613- 9102)	0.0898 (9103- 9999)
E_3	0.0812 (0000- 0812)	0.2364 (0813- 3176)	0.1988 (3177- 5164)	0.1626 (5165- 6790)	0.1085 (6791- 7875)	0.0737 (7876- 8612)	0.1388 (8613- 9999)
E_4	0	0.0812 (0000- 0812)	0.2364 (0813- 3176)	0.1988 (3177- 5164)	0.1626 (5165- 6790)	0.1085 (6791- 7875)	0.2125 (7876- 9999)
E_5	0	0	0.0812 (0000- 0812)	0.2364 (0813- 3176)	0.1988 (3177- 5164)	0.1626 (5165- 6790)	0.3210 (6791- 9999)
E_6	0	0	0	0.0812 (0000- 0812)	0.2364 (0813- 3176)	0.1988 (3177- 5164)	0.4836 (5165- 9999)

The construction of the sequence of states $E_i^{(t)}$ is obtained by starting with the arbitrary state $E_0^{(0)}$, referring to the row 0, reading off the first random number in the tables (0347) and finding the cell in row 0 to which this belongs. This is in the column 0, so



that the state to which a transition has been made is $E_0^{(1)}$. Repeating this procedure, we have that the first few transitions are

Random No.: 0347 4373 8636 9647 3661 4698 6371

State $E_1^{(t)}$: $0 \rightarrow 0 \rightarrow 0 \rightarrow 3 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow 4$

Continuing this artificial method of constructing a sequence for the process in question, we find estimates for the stationary probabilities as follows:

	P_0	P_1	P_2	P_3	P_4	P_5	P_6
(1) After 100 trials	0.29	0.14	0.18	0.11	0.11	0.09	0.08
(2) After 500 trials	0.308	0.102	0.120	0.150	0.068	0.092	0.160
(3) After 1000 trials	0.276	0.109	0.110	0.133	0.096	0.092	0.184
(4) Exact values	0.303	0.106	0.106	0.109	0.112	0.085	0.180
(5) Difference (3)-(4)	-0.027	0.003	0.004	0.024	-0.016	0.007	0.004
(6) Estimated S.E. of (3)	0.023	0.013	0.016	0.015	0.012	0.011	0.022

where the exact values (4) are those in equations (8.3.1) found by the explicit solution, and the estimated standard errors of the estimates from the realised series are found by splitting the series of 1000 trials into ten

series of 100 each. We might have used 20 series of 50 trials each, but it would be unwise to go too far in splitting the series up, for this would ultimately introduce biases, and upset the effective independence of the sub-series. It will be seen that in no case does the deviation exceed twice the estimated standard error.

These estimates of standard errors enable us to estimate how long the Monte Carlo process should continue to obtain any prescribed accuracy. Thus to reduce all the above standard errors to 0.01, we should need about $(2.3)^2 10^3 = 5290$ trials, whilst to reduce them to 0.001, we should need 529000 trials. Thus on the whole, it would appear better to try to find the eigen-vectors numerically; there are, however, some practical cases in which Monte Carlo methods may be useful. In the case of the dam, the linear equations are only an approximation to a continuous process, and if the release rule in the dam is at all complicated, it may be difficult to get an adequate approximation without increasing the number of equations to an unmanageable extent. Thus in practice, it may be better to apply the Monte Carlo method direct to the continuous problem; this is in fact what is done by Barnes(1953).

4. Extension to the continuous case

Consider the continuous dam problem to which the linear equations (6.3.1) are an approximation. The variate is the amount of water X_t in the dam at the time t just after the water has been released, t in this case being a discrete variable taking integral values. The probability distribution of X_t is partly discrete and partly continuous, for there are non-zero probabilities $p_t(0)$ and $p_t(K-M)$ that it takes the values 0 and $K - M$, where K is the capacity of the dam, and M the amount released when the dam is full. In addition, there is a continuous probability distribution $f_t(x)$, on the interval $(0, K-M)$ such that the probability that X_t lies in $(x, x+dx)$ is $f_t(x)dx$. Then

$$p_t(0) + \int_0^{K-M} f_t(x)dx + p_t(K-M) = 1.$$

The probability distribution at time $t+1$ is obtained from this and the equations

$$\begin{aligned} p_{t+1}(0) &= p_t(0)P_{11} + \int_0^{K-M} f_t(y)P_{12}(y)dy + p_t(K-M)P_{13}, \\ f_{t+1}(x) &= p_t(0)P_{21}(x) + \int_0^{K-M} f_t(y)P_{22}(x,y)dy + p_t(K-M)P_{23}(x), \\ p_{t+1}(K-M) &= p_t(0)P_{31} + \int_0^{K-M} f_t(y)P_{32}(y)dy + p_t(K-M)P_{33}, \end{aligned}$$

where the P 's are certain constants and functions which form an array analogous to the matrix of transition probabilities in a Markov chain. These P 's are defined as soon as we know the probability distribution of the input, and the rule of release. Doob shows, (1953; 190) that under a very wide set of conditions, a Markov process of this kind has a stationary solution in which $p_{t+1}(0) = p_t(0)$, $f_{t+1}(x) = f_t(x)$, $p_{t+1}(K) = p_t(K)$, and that starting from any initial point, the probabilities at time $t+n$ tend to those of this stationary state as n increases.

The analytical solution of these equations is very difficult, and is known in only one particular case (Moran 1955). A solution by Monte Carlo methods is easy, however complicated the release rule. The result will appear in the form of the frequencies with which the values 0 and $K - M$ occur, together with a set of values $x_1, \dots, x_n$ in the open interval $(0, K-M)$. Once again, the only practicable method of estimating the accuracy of the solution is to divide the whole record into sub-series, and estimate the variance from these, the variance for the continuous part of the distribution being, say, that of the frequency of the values of x_1 occurring in any specified interval.

REFERENCES

1. Anscombe F.J.
(1953) Sequential estimation. J.R.Statist.Soc., B, 15, 1-21.
2. Arrow K.J., Harris T., Marschak J.
(1951) Optimal inventory policy. Econometrica 19, 250-272.
3. Bartlett M.S.
(1951) The frequency goodness of fit test for probability chains. Proc.Camb.Phil.Soc., 47, 86-95.
4. Barnes F.B.
(1953) Storage required for a city water supply. J. Instn.Engnrs.Aust., 26, 198.
5. Cramér H.
(1946) Mathematical Methods of Statistics. Princeton.
6. Curtiss J.H.
(1953) Monte Carlo methods for the iteration of linear operators. J.Math.Phys., 32, 209-232.
7. Doob J.L.
(1953) Stochastic Processes. Wiley, New York.

8. Dvoretzky A., Kiefer J., Wolfowitz J.
- (1952a) The inventory problem. I. Case of known distributions of demand. *Econometrica*, 20, 187-222.
 - (1952b) The inventory problem. II. Case of unknown distributions of demand. *Econometrica*, 20, 450-466.
 - (1953) On the optimal character of the (s, S) policy in inventory theory. *Econometrica*, 21, 586-596.
9. Edmundson H.P.
- (1953) Monte Carlo matrix inversion and recurrent events. *Math. Tables and Other Aids to Computation*, 7, 18-21.
10. Feller W.
- (1950) An Introduction to Probability Theory and its Applications. Wiley, New York.
11. Forsythe G.E., Leibler R.A.
- (1950) Matrix inversion by a Monte Carlo method. *Math. Tables and Other Aids to Computation*, 4, 127-129.
12. Frazer R.A., Duncan W.J., Collar A.R.
- (1947) Elementary Matrices. Macmillan.
13. Fréchet M.
- (1952) Recherches Théoriques Modernes sur le Calcul des Probabilités, Second Livre. Gauthier-Villars, Paris.
14. Ganí J.
- (1955a) Some problems in the theory of provisioning and of dams. *Biometrika*, 42.
 - (1955b) Some theorems and sufficiency conditions

for the maximum likelihood estimator of an unknown parameter in a simple Markov chain. Biometrika, 42 (to appear).

15. Gani J., Moran P.A.P.

(1955) The solution of dam equations by Monte Carlo methods. Aust.J.Applied Sci., 6, (to appear).

16. Girshick M.A., Mosteller F., Savage L.J.

(1947) Unbiased estimates for certain binomial sampling problems with applications. Ann.Math.Statist., 17, 13-23.

17. Good I.J.

(1953) The serial test for sampling numbers and other tests of randomness. Proc.Camb. Phil.Soc., 49, 276-284.

18. Grenander U.

(1950) Stochastic processes and statistical inference. Arkiv for Matematik, Bd. 1 nr. 17, 195-277.

19. Hammersley J.M.

(1955) Storage Problems. Math.Annalen, Bd. 128, 475-478.

20. Hostinsky M.B.

(1931) Méthodes Générales du Calcul des Probabilités. Mémorial des Sciences Mathématiques. Gauthier-Villars, Paris.

21. Huzurbazar V.S.

(1948) The likelihood equation, consistency and maxima of the likelihood function. Ann. Eug. London, 14, 185-200.

22. Kendall D.G.

- (1949) Stochastic processes and population growth. J.R.Statist.Soc., B, 11, 230-264.
- (1951) Some problems in the theory of queues. J.R.Statist.Soc., B, 13, 151-185.

23. Kendall M.G., Babington Smith B.

- (1938) Randomness and random sampling numbers. J.R.Statist.Soc., 101, 147-166.
- (1939) Second paper on random sampling numbers. J.R.Statist.Soc., Suppl. 6, 51-61.

24. Koopman B.O.

- (1936) On distributions admitting a sufficient statistic. Trans.Amer.Math.Soc., 39, 399-409.

25. Markov A.A.

- (1912) Wahrscheinlichkeitsrechnung. Leipzig-Berlin.

26. Mihoc G.

- (1934) On the general properties of stochastic variables in a chain. Unpublished thesis in Rumanian.

27. Montroll E.W.

- (1947) On the theory of Markoff chains. Ann. Math.Statist., 18, 18-36.

29. Moran P.A.P.

- (1951) Estimation methods for evolutive processes. J.R.Statist.Soc., B, 13, 141-146.
- (1953) The estimation of the parameters of a birth and death process. J.R.Statist.Soc., B, 15, 241-245.

- (1954) A probability theory of dams and storage systems. Aust.J.Applied Sci., 5, 116-124.
- (1955) A probability theory of dams and storage systems: Modifications of the release rules. Aust.J.Applied Sci., 6, 117-130.
30. Patankar V.N.
- (1954) The goodness of fit of frequency distributions obtained from stochastic processes. Biometrika, 41, 450-462.
31. Pitman E.J.G.
- (1936) Sufficient statistics and intrinsic accuracy. Proc.Camb.Phil.Soc., 32, 567-579.
32. Pitt H.R.
- (1946) A theorem on random functions with applications to a theory of provisioning. J.London Math.Soc., 21, 16-22.
33. Rao C.R.
- (1952) Advanced Statistical Methods in Biometric Research. Wiley, New York.
34. Romanovsky V.
- (1938) Quelques problèmes nouveaux de la théorie des chaînes de Markoff. Actualités Scientifiques et Industrielles, 737, 43-66.
35. Rosenblatt M.
- (1954) An inventory problem. Econometrica, 22, 244-247.
36. Wald A.
- (1948) Asymptotic properties of the maximum likelihood estimate of an unknown parameter



of a discrete stochastic process. Ann.
Math.Statist., 19, 40-46.

37. Wasow W.R.

(1952) A note on the inversion of matrices by
random numbers. Math.Tables and Other
Aids to Computation, 6, 78-81.

38. Wolfowitz J.

(1947) Consistency of sequential binomial
estimates. Ann.Math.Statist., 18, 131-135.