AM INTRODUCTION TO MARKOV CHAIN ANALYSIS

Lyndhurst Collins

ISSN 0306-6142

ISBN 0 902246 43 7

Printed in Great Britain by Headley Brothers Ltd The Invicta Press Ashford Kent and London

CATMOG (Concepts and Techniques in Modern Geography)

1 An introduction to Markov chain analysis - L. Collins
2 Distance decay in spatial interactions - P.J. Taylor
3 Understanding canonical correlation analysis - D. Clark
4 Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5 An introduction to trend surface analysis - D. Unwin
6 Classification in geography - R.J. Johnston
7 An introduction to factor analytical techniques - J.B. Goddard & A. Kirby
8 Principal components analysis - S. Daultrey
9 Causal inferences from dichotomous variables - N. Davidson
10 Introduction to the use of logit models in geography - N. Wrigley
11 Linear programming: elementary geographical applications of the transportation problem - A. Hay
12 An introduction to quadrat analysis - R.W. Thomas
13 An introduction to time-geography - N.J. Thrift
14 An introduction to graph theoretical methods in geography - K.J. Tinkler
15 Linear regression in geography - R. Ferguson
16 Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17 Sampling methods for geographical research - C. Dixon & B. Leach
18 Questionnaires and interviews in geographical research - C. Dixon & B. Leach
19 Analysis of frequency distributions - V. Gardiner & G. Gardiner
20 Analysis of covariance and comparison of regression lines - J. Silk
21 An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22 Transfer function modelling: relationship between time series variables - Pong-wai Lai *(in preparation)*
23 Stochastic processes in one-dimensional series: an introduction - K.S. Richards *(in preparation)*
24 Linear programming: the Simplex method with geographical applications - James E. Killen *(in preparation)*

Price each: 75p or $1.50 except Catmogs 16 & 19 which are £1.25 ($2.50)

The series is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers. For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London SW7. The series is published by GEO ABSTRACTS, University of East Anglia, Norwich NR4 7TJ, to whom all other enquiries should be addressed.

AN INTRODUCTION TO MARKOV CHAIN ANALYSIS

by

Lyndhurst Collins
(University of Edinburgh)

CONTENTS

Acknowledgement

We thank W.F. Lever for permission to reproduce tables of transition matrices from The intra-urban movement of manufacturing: a Markov approach, Transactions I.B.G., 1972.

I. INTRODUCTION

(i) Geographic attraction of Markov models

Markov chain models are particularly useful to geographers concerned with problems of movement, both in terms of movement from one location to another and in terms of movement from one "state" to another. "State", in this context, may refer to the size class of a town, to income classes, to type of agricultural productivity, to land use, or to some other variable. Markov chain models are neat and elegant conceptual devices for describing and analysing the nature of changes generated by the movement of such variables; in some cases Markov models may be used also to forecast future changes. Markov chain models, therefore, are valuable both in studies of migration in which the aim may be to assess the predominant direction or the rate of change and in studies of growth or development of say an urban system in which the aim may be to determine which sort of cities are tending to increase in size and which cities are tending to decline. These preliminary and basic comments concerning Markov chain models are best exemplified with reference to a specific example.

(ii) The concept of probability

In the real world we may attach, conceptually, a probability to the occurrence of every event. There is a probability, however remote, that an aeroplane will crash or an ocean liner will sink this year; there is a probability that a red car will win the next grand prix; there is a probability that at least two towns in Britain will double their size during the next decade, and there is a probability that during the next five years five food manufacturing factories will relocate from London to Birmingham. We do not know in most cases the value of these true underlying fixed probabilities and for most courses of action we must assume or estimate the probabilities. Thus in terms of aeroplanes or shipping, insurance companies will charge a rate which, based on their judgement of experience of past trends, will more than cover the cost incurred by probable air crashes or sea disasters. In either case the companies need to know not which particular plane or ship will be involved but only the "size" or probable value. However, if an insurance company's judgement is grossly wrong the company will either go bankrupt because it has not attached a sufficiently high probability to the occurrence of a possible disaster or the company will charge rates of such a high value that its clients will seek insurance coverage elsewhere. It is in the interest of all insurance companies, therefore, to "estimate" the probabilities of disaster as accurately as possible. Likewise, so that they too can adopt the correct course of action, planners and administrators concerned with the present and future spatial organization of the landscape are interested in the probabilities of changes in town size or in the affects of changes of human migration.

(iii) A simple Markov model

Information relating to the observed probabilities of past trends, say over the last ten years, can be organized into a matrix which is the basic framework of a Markov model. To use Harvey's (1967) example, let us assume that between 1950 and 1960 the probabilities of movement between London city, its suburbs, and the surrounding country are described by the following matrix. (Table 1).

Table 1

	London	Suburbs	Country
London	0.6	0.3	0.1
Suburbs	0.2	0.5	0.3
Country	0.4	0.1	0.5

The three locations in this matrix form the "states" of the model, and each element represents the value of the probability of moving from one state to any other state; in this context, therefore, we refer to the probabilities as transition probabilities which in this example are assumed. We assume there ore, t at etween 950 and 1960 of all the people who were living in London in 1950, 60 percent or 0.6 were still in London in 1960, 30 percent (0.3) moved from London to the suburbs and 10 percent moved from London to the country. Thus, each row of the matrix, unlike the columns, sums to 100 percent or 1.0. During the same period, of those people living in the suburbs in 1950, 0.2 of them had moved into London by 1960, and of those living in the country in 1950, 0.1 of them had moved into the suburbs. Thus, the matrix of transition probabilities or transition matrix describes the probability of movement from one state to any other state during a specified or discrete time interval (10 years). With further reference to Harvey's example, let us assume that between 1950 and 1960 there was no change in the total number of the population in the three states. We are concerned, therefore, only with the redistribution of a constant population. Let us further assume that in 1950 of the total population (10 million) in the three states 50 percent were in London, 30 percent were in the suburbs, and 20 percent were in the country. This initial state of the system can be expressed in the form of a probability vector:

$$p^{(0)} = (p_1^{(0)}\ p_2^{(0)}\ p_3^{(0)}) = (0.5,\ 0.3,\ 0.2)$$

The initial state vector $p(^0)$, therefore, refers to the state of the system in 1950; $p(^1)$ would refer to the state of the system in 1960, $p(^2)$ to 1970 and so on. Similarly, for notational convenience we can refer to the complete transition matrix as P. Using theorems of matrix algebra we can obtain $p(^1)$ by multiplying the initial state vector $p(^0)$ by P so that

$$p^{(0)}\,\mathbf{P} = p^{(1)} \qquad (1)$$

In our example

$$(0.5,\ 0.3,\ 0.2) \times \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}$$

$$= \begin{array}{rcl} 0.5 \times 0.6 &=& .30 \\ 0.3 \times 0.2 &=& .06 \\ 0.2 \times 0.4 &=& .08 \\ \hline && 0.44 \end{array} \qquad \begin{array}{rcl} 0.5 \times 0.3 &=& .15 \\ 0.3 \times 0.5 &=& .15 \\ 0.2 \times 0.1 &=& .02 \\ \hline && 0.32 \end{array} \qquad \begin{array}{rcl} 0.5 \times 0.1 &=& .05 \\ 0.3 \times 0.3 &=& .09 \\ 0.2 \times 0.5 &=& .10 \\ \hline && 0.24 \end{array}$$

so that $p^{(1)} = (0.44,\ 0.32,\ 0.24)$.

Thus, in 1960 of the 10 million people in the three state system 44 percent will be in London (50 percent in 1950), 32 percent will be in the suburbs, and 24 percent will be in the country. If we assume that the transition probabilities for the 1950-1960 period will remain constant for the 1960-1970

period, and if we assume also that the total population will remain the same then we can determine the distribution of the population in 1970 (p(4 0 by multiplying the new state of the system in 1960 (p(1)) by P. Thus

$$p^{(2)} = p^{(1)}\,\mathbf{P}$$

$$(0.44,\ 0.32,\ 0.24) \times \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.1 & 0.5 \end{pmatrix} = (0.424,\ 0.316,\ 0.260)$$

By 1970, therefore, the population in London will have fallen still further to 42.4 percent of the total, whereas the proportion in the country will have increased to 26 percent. Note that between 1950 and 1960 the suburban proportion increased from 30 per cent to 32 per cent but in the next decade declined to 31.6 per cent. By applying the same vector-matrix multiplication procedure we can determine the expected distribution of the population for the three state system in 1980, 1990 and so on. In general then

$$p^{(n)} = p^{(n-1)}\mathbf{P}. \qquad (2)$$

Alternatively, instead of multiplying the initial transition matrix by successive new states of the system to derive the next proportional distribution, the same result can be obtained by multiplying each successive power of the initial transition matrix by the initial state vector.

Thus

$$\begin{aligned} p^{(1)} &= p^{(0)}\mathbf{P} \\ p^{(2)} &= p^{(0)}\mathbf{P}^2 \\ p^{(3)} &= p^{(0)}\mathbf{P}^3 \\ p^{(n)} &= p^{(0)}\mathbf{P}^n \end{aligned} \qquad (3)$$

To compute P^2 we follow the same procedure as outlined for the vector-matrix multiplication. Each row of the first matrix is regarded as a vector and is multiplied with each column of the second matrix. The values of each row-column multiplication operation are summed to give the respective elements of the new matrix - P2.

$$\mathbf{P}^2 = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.1 & 0.5 \end{pmatrix} \times \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.1 & 0.5 \end{pmatrix}$$

First Row of First Matrix

$$\begin{array}{rcl} 0.6 \times 0.6 &=& .36 \\ 0.3 \times 0.2 &=& .06 \\ 0.1 \times 0.4 &=& .04 \\ \hline && .46 \end{array} \qquad \begin{array}{rcl} 0.6 \times 0.3 &=& .18 \\ 0.3 \times 0.5 &=& .15 \\ 0.1 \times 0.1 &=& .01 \\ \hline && .34 \end{array} \qquad \begin{array}{rcl} 0.6 \times 0.1 &=& 0.07 \\ 0.3 \times 0.3 &=& 0.09 \\ 0.1 \times 0.5 &=& 0.05 \\ \hline && .20 \end{array}$$

Second Row

$$\begin{array}{rcl} 0.2 \times 0.6 &=& \ldots \\ 0.5 \times 0.2 &=& \ldots \\ 0.3 \times 0.4 &=& \ldots \\ \hline && 0.34 \end{array} \qquad \begin{array}{rcl} 0.2 \times \ldots &=& \ldots \\ 0.5 \times \ldots &=& \ldots \\ 0.3 \times \ldots &=& \ldots \\ \hline && \end{array}$$

Third Row ... x ... = ...
... x ... = ...
... x ... = ...

Thus $$P^2 = \begin{pmatrix} 0.46 & 0.34 & 0.20 \\ 0.34 & 0.34 & 0.32 \\ 0.46 & 0.22 & 0.32 \end{pmatrix}$$

and $$P^3 = P \times P^2$$
$$P^4 = P \times P^3 \text{ or } P^2 \times P^2$$

Although initially the second procedure may seem more complicated, it is the one most commonly used in geographic computer programs (see Marble, 1967). The main reason for its more widespread use is that further descriptive measures (see later sections) can be derived from successive powers of the transition matrix whereas, by contrast, the first procedure reveals only the state of the system at the end of each time interval. In essence, these are the basic principles of Markov chain analysis. This simple example has shown how, in a Markov model, information relating to the probability of events occurring probability outcomes can be organised into a concise format for descriptive, analytic and predictive purposes. It must be emphasized, however, that in our example we have assumed a constant population, we have assumed a set of transition probabilities, we have assumed that these probabilities have remained constant or stationary, and we have assumed that the information can in fact be approximated by a Markov model. The aims of the following sections are fourfold: to define more precisely the concepts of Markov chain analysis so that it will be possible to determine to what extent one is justified in using the technique for geographical analysis, to elaborate many of the descriptive qualities of Markov models, to discuss the implications of the underlying assumptions, and lastly to outline particular applications of Markov models to geographical problems.

(iv) Markov chain models as particular cases of stochastic process models

A Markov chain is one particular type of stochastic process. In general terms we can define a stochastic model as one which provides only the probability or likelihood associated with a set of possible future outcomes. Thus, whereas in a deterministic law state X is always followed by state Y, in a stochastic law state X is followed by state Y with probability p, and by state Z with probability q = 1 - p. In the same way that all models may be classified into say deterministic or stochastic models, stochastic models may be classified on the basis of being "discrete" or "continuous", depending on whether or not the time variable is treated in these terms, or depending on whether or not they possess the Markov property. Markov process models possess this property and can be regarded as generalizations of Markov chains; in a Markov process model a transition from one state to another can take place at any point in time, i.e., time is continuous, but in a Markov chain the state varies only at discrete time intervals. A Markov chain can be defined as:

> ... a stochastic process in which the future development depends only on the present state, but not on the past history of the process or the manner in which the present state was reached (Feller, 1968, p. 444).

This degree of dependency is the Markov property. With reference back to the example in section I(iii) it is seen that in a Markov chain a system of states changes, according to some probability law, with time t in such a manner that the system changing from a given state S_i at time t_{0+1} depends only on the state S_i at time t_0 and is independent of the states of the system prior to t_0. Accordingly, in the earlier example knowledge of the state of the system in 1950 and of the probability of moving from one state to any other was sufficient to estimate the state of the system in each successive ten year interval. No knowledge of the system prior to 1950 was necessary. If the state of the system at time t_{0+1} is dependent only on the state of the system at time t_0 plus some independent random component, the process is referred to as a first-order Markov' chain. In a second-order Markov chain the state of the system at time t_2 would depend on the states of the system at both time t_0 and t_1. In this way, provided sufficient data are available, the "dependence" of a Markov chain can be extended to any length though in geography the tendency has been to adopt first-order Markov chains. It should be noted however, that even in a first-order Markov chain, although at time t_1, the state S_1 depends on S_0 and need not refer to S_{0-1}, S_{0-2}, ..., S_{0-n}, proportions of S_{0-1}, S_{0-2}, ..., S_{0-n}, are included in S_0 which represents the sum of past history. To express this degree of dependency of a given state upon previous states, Feller (1968 p. 329) provides the term "memory". With reference to this concept of memory it is possible to outline the position occupied by Markov chain models within the broad conceptual spectrum that includes classical deterministic models at one extreme and at the other purely random models. In a purely random model, for example a Poisson process, the state of the system at any instant or point in time or space is wholly independent of its state at any other instant or point and is completely specified by the underlying fixed probabilities. A purely random model, therefore, has a marked lack of memory. On the other hand, a classical deterministic model, in which the state of a system at t0 depends upon all previous states, has a long memory. The first-order Markov chain model, therefore, although it embodies the sum of past history contained in state S_0, occupies a position of partial dependence. These general comments on the relationship between Markov models and other models serve as a framework for a discussion of the more formal properties of regular finite Markov chains.

II. PROPERTIES OF REGULAR FINITE MARKOV CHAINS

(i) Transition probabilities

The transition probabilities, the p_{ij}'s which give the probability that the process will move from state S_i to state S_j are given for every pair of states The "set" of probabilities or "outcome functions" describes the process as it moves through any finite number of steps. For computational reasons and notational simplicity the transition probabilities can be generalised in the form of a transition matrix **P** (see over page)

The elements of **P** denote the probability of moving from state S_i to S_j in the next step. Since the elements of this matrix must be non-negative and the sum of the elements in any row is 1, each row is called a probability vector and the matrix **P** is a stochastic matrix. If for some power of the matrix **P** there are only positive entries the transition matrix is described as regular.

		S_1	S_2	S_3	. . .	S_n
	S_1	p_{11}	p_{12}	p_{13}		p_{1n}
	S_2	p_{21}	p_{22}	p_{23}		p_{2n}
P =	.					.
	.					.
	.					.
	S_n	p_{n1}	p_{n2}	p_{n3}	. . .	p_{nn}

where
$$\sum_{j=1}^{n} p_{ij} = 1 \qquad (4)$$
and
$$p_{ij} \geq 0 \quad \text{for all } i \text{ and } j$$

(ii) Markov theorems

For regular Markov chains two important theorems relating to the equilibrium properties are provided by Kemeny and Snell (1967, pp. 69-98).

Theorem I If P is a transition matrix for a regular Markov chain then:

(1) the powers of P approach a matrix **A**

(2) each row of **A** is the same probability vector a

(3) -the elements of a are all positive

Theorem II If **P** is a transition matrix for a regular Markov chain and **A** and α are as in Theorem I, then the unique vector α is the unique probability vector such that α **P** = α . The matrix **A** is defined as the limiting matrix.
These theorems are best exemplified by specitic reference to a hypothetical example. Consider a constant sample of merchant ships distributed throughout the three oceans: the Atlantic, the Pacific and the Indian. Assume that at time t0, 20 per cent of the total number of ships in the three oceans are located in the Indian Ocean, 20 per cent are in the Pacific Ocean and, 60 per cent are in the Atlantic Ocean. Thus the initial state of the system can be represented by the initial distribution vector - $p^{(0)}$ - which is expressed as

$$p^{(0)} - (20, 20, 60) = (.2, .2, .6)$$

Assume also that the probability of a ship's moving from its trading activities in one state (ocean) to any other state during a specified time period (one year interval) is described by the following transition matrix

		S_1 Indian	S_2 Pacific	S_3 Atlantic
	S_1 (Indian)	.6	.2	.2
P =	S_2 (Pacific)	.3	.4	.3
	S_3 (Atlantic)	.2	.2	.6

(iii) The limiting matrix

The matrix P shows that the probability of a ship's remaining in the Indian ocean during a given time interval is .6, whereas the probability of a ship's moving from the Indian ocean to the Pacific is .2 and so on. Given this initial transition matrix it is possible to compute the transition probabilities after 1, 2, 3, ..., n, stages by calculating the relevant power of the matrix.

After two stages the matrix takes the form:

		S_1	S_2	S_3
	S_1	.46	.24	.30
$\mathbf{P}^2$ =	S_2	.36	.28	.36
	S_3	.30	.24	.46

and after four stages:

		S_1	S_2	S_3
	S_1	.3880	.2496	.3624
$\mathbf{P}^4$ =	S_2	.3744	.2512	.3744
	S_3	.3624	.2496	.3880

and by the eighth stage the transition matrix takes the form

		S_1	S_2	S_3
	S_1	.37533	.25000	.37467
$\mathbf{P}^8$ =	S_2	.37500	.25000	.37500
	S_3	.37467	.25000	.37533

A comparison of the above four matrices indicates a rapid convergence towards some average state of the system. This average state is represented by the limiting or **A** matrix:

		S_1	S_2	S_3
	S_1	.3750	.2500	.3750
A =	S_2	.3750	.2500	.3750
	S_3	.3750	.2500	.3750

and the probability vector α = (.3750, .2500, .3750) holds the system in equilibrium.

(iv) The concept of equilibrium

In the present context, the notion of equilibrium can be defined as that distribution for which the average number of ships entering a particular ocean in a given time interval equals the average number of ships leaving it. This concept of equilibrium is analogous in some ways to the comparatively stable pattern of national population pyramids. The age structure of a population, for example, under constant socio-economic conditions tends to maintain a state of equilibrium, but the shape of these equilibrium distributions may vary among countries. If, however, there is some major disruption such as war, famine, or technological change (e.g., birth control, vaccine) the age distribution may be altered significantly. In pre-World War 1 Europe the population pyramids of Britain, France, Germany and Italy had similar configurations. But after the conflict which had a particularly high toll among the male 20-45 age groups, especially in Germany, the patterns were significantly disrupted. Sixty years later, despite another World War the respective population pyramids are being restored to their former patterns. This is not to say that the particular patterns will remain absolutely the same in the long term because socio-economic conditions are changing continuously. In Markov chanin analysis, therefore. for modelling purposes the equilibrium distribution is of interest not as a forecast of the future state of the system (distribution of ships in three oceans) but as a projection of what would be if the observed pattern of movement continued unhampered. In our example the limiting probability a_j (an element of the matrix A) of being in state S_j is independent of the starting state and represents the fraction of the time that the process can be expected to be in state S_j during a large number of transitions and after a large number of steps from $p^{(0)}$. This arises from the law of large numbers for regular Markov chains. In terms of our example, after a large number of time periods 37.5 per cent of the ships will be in the Indian ocean, 25 per cent will be in the Pacific and 37.5 per cent will be in the Atlantic.

(v) The Fundamental matrix

The limiting matrix is but one important property of a regular Markov chain. Other descriptive properties can be computed from the Z or fundamental matrix. The details of the matrix algebra associated with these computations are outlined by Kemeny and Snell (1967) and in order to allow comparison the notation used below, unless otherwise indicated, is the same. In matrix notation

$$\mathbf{Z} = (\mathbf{I} - (\mathbf{P} - \mathbf{A}))^{-1} \quad (5)$$

where
- $\mathbf{I}$ is an identity matrix
- $\mathbf{P}$ is a regular matrix
- $\mathbf{A}$ is the limiting matrix of $\mathbf{P}$.

In using the ship and ocean example above the P and A matrices are given so that

$(\mathbf{I} - \mathbf{P} + \mathbf{A}) =$

	S_1	S_2	S_3
S_1	.775	.050	.175
S_2	.075	.850	.075
S_3	.175	.050	.775

Inversion of this matrix provides

$\mathbf{Z} =$

	S_1	S_2	S_3
S_1	1.36458	-1.06250	-0.30208
S_2	-0.09375	1.18750	-0.09375
S_3	-0.30208	-0.06250	1.36459

Whereas the fundamental matrix Z has several properties in common with a transition matrix it does not necessarily have non-negative entries. The Z matrix simply describes how the system approaches equilibrium from a given initial distribution: from any initial state, the expected percentage of time the system will spend in state j approaches a_j as the number of time periods n becomes large; however, starting from a given state i, this expected percentage differs from a_j by approximately $(Z_{ij} - a_j)/n$. The important quality of the **Z** matrix however, is that it can be used to provide a measure of the time it takes on average to move from one state to another. The distribution describing this random variable is called the first passage time distribution, and the expected value is called mean first passage time.

(vi) The matrix of mean first passage times

The matrix of mean first passage times is denoted by M and the entries, the m_{ij}'s give the expected time to move from S_i to S_j for the first time. For a regular Markov chain the mean first passage time matrix is given by

$$\mathbf{M} = (\mathbf{I} - \mathbf{Z} + \mathbf{E}\mathbf{Z}_{dg})\,\mathbf{D} \quad (6)$$

where
- I is an identity matrix
- Z is the fundamental matrix
- E is a matrix with all entries 1
- $\mathbf{Z}_{dg}$ results from Z by setting off - diagonal entries equal to 0
- $\mathbf{D}$ is the diagonal matrix with j-th entry $1/a_j$.

For our example

M =

	S_1	S_2	S_3
S_1	2.66667	5.00002	4.44447
S_2	3.88890	4.00002	3.88891
S_3	4 44446	5.00002	2.66668

Since the entries $m_{ij} = M_i / \bar{f}_{ij}$ represent the mean number of time periods - in our case one year intervals - to arrive at any given state, the average ship would take 5.0 years to relocate from the Indian ocean to the Pacific, and 3.9 years to move from the Pacific to the Atlantic.

(vii) Matrix of standard deviations

In the same way that the significance of means can be determined by standard deviations the variance of the mean first passage times provides similar useful descriptive information. The variance of M is given by

$$Var_i [\bar{f}_j] = M_i [\bar{f}_j^2] - M_i [\bar{f}_j]^2 \quad (7)$$

Kemeny and Snell (1967 p. 82) denote $M_i [\bar{f}_j^2]$ by which represents the matrix of second moments of number of steps required to reach S_j. The matrix satisfies the equation

$$\mathbf{W} = \mathbf{M}(2\mathbf{Z}_{dg}\mathbf{D} - \mathbf{I}) + 2(\mathbf{ZM} - \mathbf{E}(\mathbf{ZM})_{dg}) \quad (8)$$

where $(\mathbf{ZM})_{dg}$ results from the product of the fundamental and mean first passage matrices by setting off - diagonal entries equal to 0. The entries of the matrix $M_i [f_j]^2$ are the squares of the first moments for the **M** matrix. The hadamard product of **M** is best denoted by $\mathbf{M}_{sq}$.
A simple matrix operation gives:

$$Var_i [\bar{f}_j] = \mathbf{V} = \mathbf{W} - \mathbf{M}_{sq} \quad (9)$$

so that

		S_1	S_2	S_3
	S_1	9.62973	20.00024	14.07428
V =	S_2	13.08653	18.00024	13.08661
	S_3	14.07420	20.00027	9.62978

Since, in the present example, the standard deviations of the first passage times - the v_{ij}'s - are of a greater magnitude than the means - the m_{ij}'s - the means can not be taken as typical values.

(viii) Central Limit theorem for Markov chains

The most important descriptive measure is derived from the Central Limit theorem for Markov chains. The computation of this property requires the elements of the limiting covariance matrix. The entries, the c_{ij}'s - of the **C** matrix are given by:

$$c_{ij} = a_i Z_{ij} + a_j Z_{ji} - a_i d_{ij} - a_i a_j \quad (10)$$

where a_{ij} are the elements of the limiting matrix A
Z_{ij} the entries of the fundamental matrix
d_{ij} the entries of a diagonal matrix with diagonal elements d_{ii} equal to the reciprocal of the diagonal elements of the limiting matrix **A**.

In the illustrative example the limiting covariance matrix is:

		S_1	S_2	S_3
	S_1	0.50781	-0.14062	-0.36718
C =	S_2	-0.14062	0.28125	-0.14062
	S_3	-0.36718	-0.14062	0.50781

The limiting variances, the diagonal entries of C, are denoted by:

$$\beta = [b_j] = [c_{jj}]$$

and in the example the vector

$$\beta = (0.50781, 0.28125, 0.50781)$$

These quantities appear in the Central Limit theorem defined in the following manner by Kemeny and Snell (1967, p. R91 For any regular Markov Chain, let $y^{(n)}_j$ be the number of times in state S_j in the first n steps and let $\alpha = a_j$ and $\beta = b_j$ respectively be the fixed vector and the vector of limiting variances. Then if $b_j \neq 0$ for any numbers $r < s$

$$Pr_k \left[r < \frac{y^{(n)}_j - na_j}{\sqrt{nb_j}} < s \right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_r^s e^{-x^2/2} dx \quad (11)$$

as $n \to \infty$, for any choice of starting state k.

Evaluation of the integral would provide aproximate values of 0.681, when r = -1 and s = 1, 0.954 when r = -2 and s = 2, and 0.997 when r = -3 and s = 3.

Using the alpha and beta values the Central Limit theorem gives for the Pacific ocean (S_2) in the illustrative example:

$$\frac{y^{(n)}_2 - .25n}{\sqrt{.28125\ n}} \quad (12)$$

which would for large n have approximately a normal distribution. The high limiting variances, however, suggest the low predictive value for the long term of this illustrative model. Equation 12, for example, gives that after approximately 100 time intervals (years) the percentage of ships in the Pacific would, with probability 0.68, not deviate from 25 per cent by more than

$$\sqrt{100 \times .2812} = 5.3\%$$

All the properties discussed in this section refer only to regular finite Markov chains which are to be distinguished from absorbing Markov chains. In an absorbing Markov model there is at least one state which once entered cannot be left. Thus in modelling the age structure of a human life cycle, "death" would be an absorbing state. The corresponding properties of absorbing Markov models are not discussed here because in the geographic literature the use of absorbing chains has been extremely limited. One noteworthy exception is that of Marble (1964). The formal properties of absorbing Markov chains are elaborated by Kemeny and Snell (1967, Ch. 3) and a program to compute most of these properties is contained in Marble (1967).

III A PROCEDURE FOR CONSTRUCTING A MARKOV MODEL

(i) The Markov property

In the example outlined in Section II it is assumed throughout that the movement of ships in the three oceans does typify a first-order Markov process. But the valid application of a Markov model is contingent on the identification of the specific-order property. Given adequate data this property can be readily determined by a substantial body of statistical theory. It should be noted, however, that any transition matrix suggests a Markovian model but the partial dependence or Markovity property of a Markov chain renders it unsuitable for the analysis of an independent series of events. A transition matrix depicting such a series is often described as zero-order. Before determining the specific-order of a stochastic matrix, therefor is essential to test the validity of the Markov property assumption. For this the Maximum Likelihood Ratio Criterion, which may be extended to determine the specific-order of the process, is an appropriate test statistic. The design of these tests involving asymptotic distribution theory and the closely related Chi-square tests of the form used in contingency tables is elaborated in two studies of Markov methods by Anderson and Goodman (1957) and Kullback, Kupperman and Ku (1962). Basic to all these tests is the actual number of observations for all cells represented in a "Tally Matrix". Assume that the initial transition probability matrix of the three ocean example above was derived from the observations tabulated in Table 2.

Table 2: Tally Matrix

	S_1	S_2	S_3	Marginal Totals
S_1	120	40	40	200
S_2	60	80	60	200
S_3	120	120	360	600
	300	240	460	1000

By using Anderson and Goodman's (1957) maximum likelihood ratio criterion test for the Markov property it is possible to test the null hypothesis - that the movement of ships from one ocean to another is statistically independent as against the alternative that the observations exhibit partial dependence.

(ii) Maximum likelihood ratio criterion test for the Markov property

In general this tests the null hypothesis that a stationary transition matrix is of "zero"- order, that is pij = pj for all i, against the alternative of a first-order chain.

The ratio criterion is:

$$\lambda = \prod_{i,j} (\hat{p}_j / \hat{p}_{ij})^{f_{ij}} \qquad (13)$$

where the marginal probability

$$\hat{p}_j = \sum_i f_{ij} / \sum_i \sum_j f_{ij} = f_{.j} / f_{..}$$

and

$$\hat{p}_{ij} = f_{ij} / \sum_j f_{ij} = f_{ij} / f_{i.}$$

and f_{ij} is the number of observations in each cell. The required statistic is $-2 \log_e \lambda$ which under the null hypothesis has an asymptotic Chi-square distribution with $(n-1)^2$ degrees of freedom. Equation (13) may be written as:

$$-2 \log_e \lambda = 2 \sum_{i=j}^{n} \sum_{j=i}^{n} f_{ij} \log_e \frac{f_{ij}\, f_{..}}{f_{i.}\, f_{.j}} \qquad (14)$$

The $\hat{p}_j$ values in (13) are obtained by summing the columns in each state and by converting to proportions; p_1 in the example is given by:

$$300 / 1000 = .3$$

and the procedure is repeated for all p_j. Since p_{11} is 0.6 the required statistic for S_{11} is:

$$120 \log_e (.6 / .3) = 83.17 \qquad (15)$$

For each matrix element the computations are repeated and summed algebraically; the N log P matrix is shown in Table 3.

Table 3: N log P Matrix

	S_1	S_2	S_3
S_1	83.177	-7.292	-33.316
S_2	0.000	40.866	-25.646
S_3	-48.655	-21.875	95.652

The sum of the elements is doubled and compared with the Chi-square distribution for the chosen level of significance and $(n-1)^2$ degrees of freedom. Since, in the hypothetical example $-2 \log_e \lambda = 165.8$ which is greatly in excess of the tabled value of Chi-square for four degrees of freedom the hypothesis of an independent trials process is rejected.

(iii) The first-order Markov property

Alone, the dependence property is insufficient justification for adopting a specific-order Markov model. The assumption of most studies that any transition matrix typifies a first-order Markov process, is akin to that of assuming normality for the application of standard statistical procedures. Such assumptions usually arise from inadequate data. Anderson and Goodman's test for a first-order Markov chain cannot be applied to a simple two dimensional tally matrix. Their test requires observations on individual movements through at least two time intervals; such observations are distinct from the aggregate observations derived by a comparison of the states at two dates. Assume, for example, that the tally matrix presented above is for the 1941-1951 period. This indicates that during the interval, 40 ships moved from S_1 to S_2 and 120 ships moved from S_3 to S_2, and so on. But to determine that the data typify a first-order process it is necessary to observe the individual movements of these ships during the next time interval (1951-19611. In this way we can attach a probability to a ship's moving to S_3 in the next interval given that it has already moved from S_1 to S_2. The notion is best illustrated by a conditional probability tree (Fig. 1). Formally, this shows the probability of moving to the j-th state in the k+1 "realization" given that movement has already occurred from the i-th state to a j-th state in the k-th realization.

The probability tree shows that of the 40 ships which moved from S_1 to S_2 between 1941 and 1951, 28 remained in the same ocean during the 1951-61 period but two returned to S_1 and ten moved to S_3. Such data are best presented in a three way or cubic matrix which takes the general form shown in Fig. 2. For the hypothetical example the three facets or leaves of the cubic matrix are presented also in Fig. 2.

(iv) Maximum likelihood ratio criterion test for a first-order Markov chain

Given such data it is possible to test the null hypothesis that the chain is first-order against the alternative that it is second order. The null hypothesis is that $p_{1jk} = p_{2jk} = \dots = p_{njk} = p_{jk}$, for j, k = 1, ... n. The likelihood ratio criterion for testing this hypothesis is:

$$\lambda = \prod_{i,j,k=1}^{n} (\hat{p}_{jk} / \hat{p}_{ijk})^{f_{ijk}}$$

where

$$\hat{p}_{jk} = \sum_i f_{ijk} / \sum_i \sum_k f_{ijk} = f_{.jk} / f_{.j.}$$

and

$$\hat{p}_{ijk} = f_{ijk} / \sum_k f_{ijk} / f_{ij.}$$

Under the null hypothesis, $-2 \log \lambda$ is asymtotically χ^2 with $n(n-1)^2$ degrees of freedom.

This test was applied by Collins (1972) to a 14x14 cubic matrix of manufacturing establishment size categories (size measured in terms of employment). Data were available for each year for the 1961-1965 period and the resulting values are shown in Table 4. For all realizations, the values of $-2 \log_e \chi$

Fig 1 PROBABILITY TREE FOR A FIRST – ORDER MARKOV CHAIN

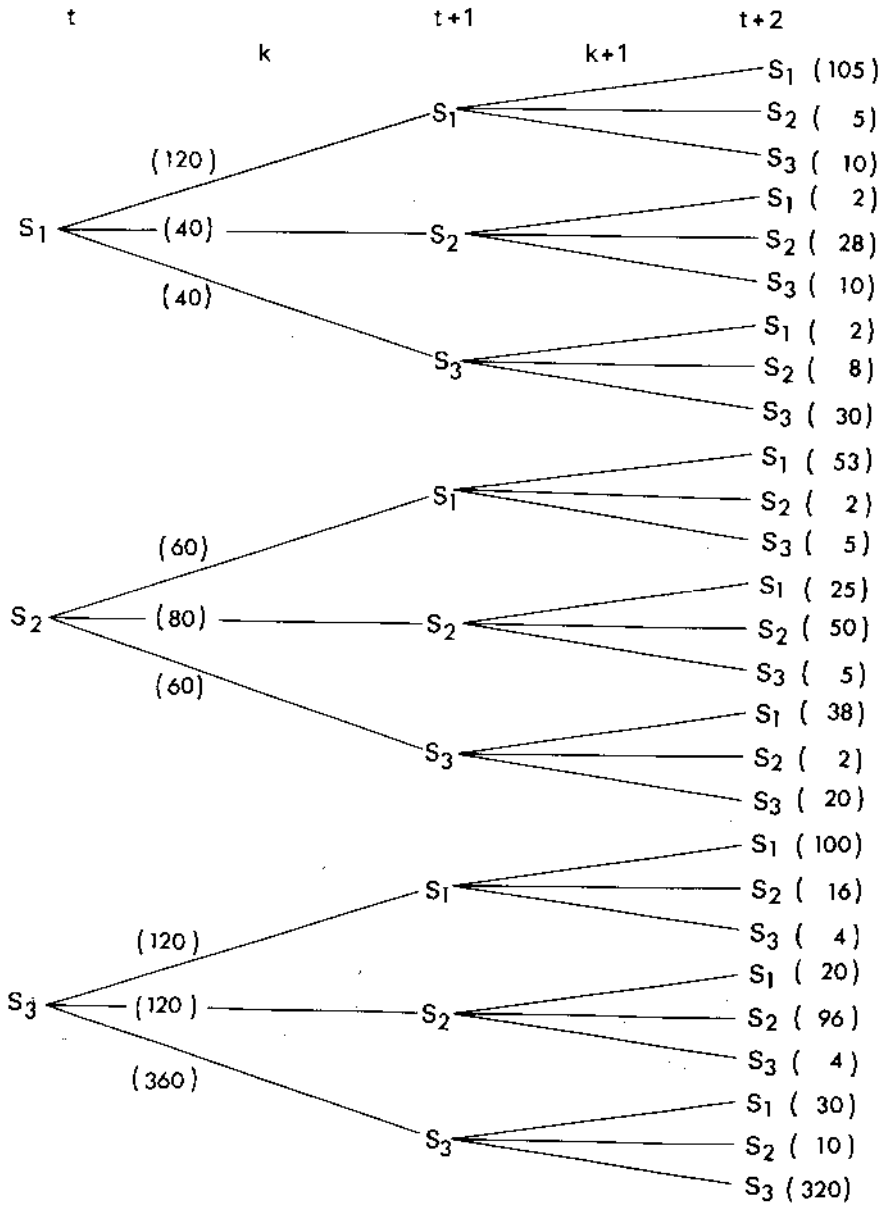
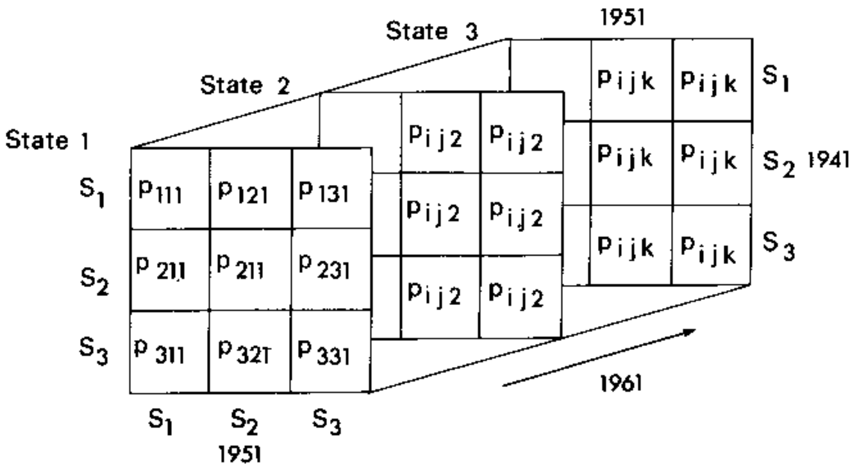


Fig 2 CUBIC MATRIX FOR A FIRST—ORDER MARKOV CHAIN



State 1 : 1961

	S1	S2	S3
S1	105	2	2
S2	53	25	38
S3	100	20	30

Rows: 1941; columns: 1951

State 1 : 1961

S1	5	28	8
S2	2	50	2
S3	16	96	10

Rows: 1941; columns: 1951

State 1 : 1961

S1	10	10	30
S2	5	5	20
S3	4	4	320

Rows: 1941; columns: 1951

are less than the appropriate degrees of freedom. The null hypothesis - that the chain is first-order against the alternative that it is second-order is not rejected and the change a plant's employment structure is considered to typify a first-order Markov process.

Table 4: Test of first-order property for 14 x 14 matrices of establishment size categories

Realization	$-2 \log_e \lambda$	D.F. $n(n-1)^2$
1961-63	697.62	2,366
1962-64	631.21	2,366
1963-65	607.78	2,366
1961-65	610.16	2,366

(v) The concept of stationarity

Fundamental Markov theory requires, in addition to the first-order property, that the parameters be stationary. This implies that the estimated transition probabilities are fixed or constant throughout the predictive period, and is, therefore, a restricting assumption of Markov theory. It is often possible, however, to estimate a series of transition matrices or a set of realizations which typify recent trends. the constancy of these trends can be determined by statistical tests. The test statistics discussed here have been developed using the notions of information theory (Kullback, Kupperman, and Ku, 1962). They are particular cases of Minimum Discrimination Information Statistics (m.d.i.s.) and are labelled as:

(a) "i - homogeneity" which is the two-way independence component
(b) "conditional homogeneity" or Markov property component
(c) "(i.j) - homogeneity" which is the two-way by one-way independence component.

(vi) Statistical test for homogeneity

Under the hypotheses of homogeneity (similarity), the statistics are distributed as central Chi-square variables with appropriate degrees of freedom. The information table for the general case of s sub-tables of n rows is given in Table 5. (see over page)

The table entries are f_{kij}, indicating the value in the k-th sub-table, i-th row and j-th column. The subscript dots indicate summation over the ranges of indexes replaced, and $m=f_{...}$.

A good illustrative example is provided in a report prepared for the Canada Land Inventory by the Assessment, Determination and Evaluation Team of Spartan Air Services Ltd. (1967). The example concerns the problem of regionalization on the basis of similarity in land use changes.
For convenience land use is classified as:
(a) Pasture (b) Woodland (c) Unproductive land (d) Cultivated areas.

Assume that random samplings from arbitrarily defined parcels of land yield the following transition matrices for counties A (Table 6), and B (Table 7).

Table 5: Information Table for Statistical Tests of Homogeneity

Component due to	Information	Degrees of Freedom	
(i)	$2\sum_{k=1}^{s}\sum_{i=1}^{n} f_{ki.} \log_e \frac{m f_{ki.}}{f_{k..} f_{.i.}}$	$(s-1)(n-1)$	(16)
(j/i) conditional homogeneity	$2\sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{n} f_{kij} \log_e \frac{f_{kij}}{\frac{f_{ki.}\, f_{.ik}}{f_{.i.}}}$	$n(s-1)(n-1)$	(17)
(i,j) homogeneity	$2\sum_{k=1}^{s}\sum_{i=1}^{n}\sum_{j=1}^{n} f_{kij} \log_e \frac{m f_{kij}}{f_{k..} f_{.ij}}$	$(s-1)(n^2-1)$	(18)

Table 6

	P	C	U	W
P	272	72	26	0
C	31	143	42	4
U	0	24	53	3
W	0	1	7	22

Table 7

	P	C	U	W
P	195	47	18	0
C	24	102	29	5
U	1	16	36	2
W	0	0	5	15

Is land use change in the two counties proceeding in a similar manner? Are the changes proceeding at the same rate? In other words are the two transition matrices homogeneous? The computations are organized as follows:

Table 8: Entries f_{kij}

K	i \ j	P	C	U	W	Total
County A	P	272	72	26	0	370
	C	31	143	42	4	220
	U	0	24	53	3	80
	W	0	1	7	22	30
County B	P	195	47	18	0	260
	C	24	102	29	5	160
	U	1	16	36	2	55
	W	0	0	5	15	20
Total		523	405	216	51	1195

Table 9: Entries f_{ki}

k \ i	P	C	U	W	Total
County A	370	220	80	30	700
County B	260	160	55	20	495
Total	630	380	135	50	1195

Table 10: Entries $f_{.ij}$

i \ j	P	C	U	W
P	467	119	44	0
C	55	245	71	9
U	1	40	89	5
W	0	1	12	37
Total	523	405	216	51

These data provide the following values:

$$\sum_{k=1}^{2}\sum_{i=1}^{4}\sum_{k=1}^{4} 2f_{kij} \log_e f_{kij} = 10803.9 \quad (19)$$

$$\sum_{k=1}^{2}\sum_{i=1}^{4} 2f_{ki.} \log_e f_{ki.} = 12730.3 \quad (20)$$

$$\sum_{i=1}^{4}\sum_{j=1}^{4} 2f_{.ij} \log_e f_{.ij} = 12429.1 \quad (21)$$

$$\sum_{k=1}^{2} 2f_{k..} \log_e f_{k..} = 15314.0 \quad (22)$$

$$\sum_{i=1}^{4} 2f_{.j.} \log_e f_{.j.} = 14351.7 \quad (23)$$

$$2f_{...} \log_e f_{...} = 16935.3 \quad (24)$$

Using these values the three information components are derived in the following way:

j - homogeneity = (27)+(31)-(29)-(30) = 0.1; D.F. = (s-1)(n-1) = 3

conditional homogeneity = (26)+(30)-(27)-(28) = 3.8; D.F. = n(s-1)(n-1) = 12

(j,k) - homogeneity = (26)+(31)-(29)-(28) = 3.9; D.F. = $(s-1)(n^2-1)$ = 15

Table 11: Analysis of information

Component due to	Information	D.F.
j - homogeneity	0.1	3
conditional homogeneity	3.8	12
(j,k) homogeneity	3.9	15

Since none of the "information values" is significant at the 0.1 level similarity between the two counties in terms of total land use and rate of change is indicated.

Assuming therefore, that we have two realizations for the three ocean example - 1941-51, and 1951-61 - we can test the hypothesis that the two realizations come from the same but unspecified matrix of transition probabilities. In our case the set comprises S(i.e. 2) realizations of a first-order Markov chain with n(i.e. 3) states. For this the null hypothesis is that the probability of moving from state i to state j in the k-th realization, that is $p_k(S_j/S_i)$ is the same for all k(k=1,2,...,S)for every possible pairing of i and j where i=1, 2,...,n, and j=1,2,...,n. It must be emphasized that by demonstrating that the differences between the series of realizations are small enough to be attributed to random or chance fluctuations we show only that past trends have been constant; but in so doing we give substance, in some cases, to the assumption that future short term trends will continue at a constant pace.

IV DERIVATION OF TRANSITION PROBABILITIES

The transition probabilities, the p_{ij}'s, form the heart of any Markov chain model; their derivation is of the utmost importance. In the illustrative example used so far the transition probabilities have been assumed. Such probabilities, however, cannot be used in models of real-world situations. In a mathematical model the probability would be derived theoretically from a probability distribution but for most models the parameters are usually derived empirically. Two main approaches have been used: conceptual and statistical estimation.

(i) Conceptual

The general problem is a lack of suitable disaggregated data showing individual

temporal or spatial movements of selected economic variables. Krenz (1964), in a study of temporal changes of farm size in North Dakota, adopted a conceptual approach by postulating rules of behaviour for the micro units (farms). Basic data were obtained from the Quinquennial Census of Agriculture which only enumerates the total number of farms in each of several size categories (size based on acreage); no information concerning the movement of individual farms from one size group to another was available. Krenz postulated that: if possible, farm operators would always expand their acreage; the farms most likely to expand are initially larger than average; increases in farm size are most likely to result from gradual increases in acreage; and that a farm is more likely to go out of business than to decrease its acreage. Krenz adopted a six state system (state refers to farm size) and used the assumptions to postulate three rules of behaviour. First, farms in the largest category S_6, remain in this category. Second, increases in number of farms in any state S_i move from the next smaller state S_{i-1}. Finally, any decrease in the number of farms in any state, other than from the second rule, results in a movement to S_0 which represents a cover-all state for going out of business. An absorbing Markov model is thereby applied. Clearly, this approach in highly dependent on an intimate knowledge and on the conceptualization of the processes involved.

(ii) Statistical estimation from aggregate data

Where a detailed knowledge of the underlying processes is absent the parameters must be statistically estimated from either aggregate occurrence data of the type used by Krenz or from observations of individual movements between states. One alternative to estimating transition probabilities from aggregate or total occurrence data is the use of linear or quadratic programming procedures. A linear programming solution, for example, was used by Scott (1965) in a study of the diffusion of the Negro population in the south-side of Chicago between 1930 and 1960. Aggregate data for six contiguous regions were available for the years 1930, 1940, 1950 and 1960. Scott defines a time period as "... an increase in the Negro population in the study area of approximately 37,000 people". By assuming a constant population growth in all six regions Scott generated five sets of artificial time periods providing him with nine distribution vectors. As did Krenz, Scott hypothesized rules of behaviour for the assumed outward migration; transitions in any one time period were only allowed into adjacent regions.

(iii) Statistical estimation from individual observations

Given observations on the individual movements of economic variables, Anderson and Goodman provide the maximum likelihood techninue for estimating transition probabilities. Maximum likelihood estimates of the p_{ij}'s are derived by dividing the number of times micro units move from S_i to S_j by the total number of occurrences of S_i; the total number of occurrences and individual movements are obtained froth empirical observation. Thus:

$$\hat{P} = \left[\hat{p}_{ij} \right] = \left[\frac{n_{ij}}{\sum_{j=1}^{n} n_{ij}} \right] \geq 0$$

where nij is the number of movements of the sample elements from state Si to Sj. Thus, in Table 2 the transition probabilities are derived by dividing each pi• or element of the matrix, by the appropriate marginal totals. One distinct advantage of being able to use the maximum likelihood criterion technique for estimating the transition probabilities is that the original tally matrices can be subjected to statistical tests for determining the specific order of the chain (See Section III)

In summary, it is emphasized that considerable caution must be exercised in the initial classification of states which must be meaningful for the particular process under study. The results of any Markov chain analysis for example, can be interpreted only in the framework of the adopted classification scheme.

V SOME GEOGRAPHICAL IMPLICATIONS OF MARKOV ASSUMPTIONS

No matter how sophisticated are the mathematical techniques employed in the formulation of a predictive model, intelligent generalizations must depend, in the final analysis, on the acceptance of certain restricting assumptions. The four main Markovian assumptions, already mentioned, are further discussed in this section to emphasize their interdependence.

(i) The system of states

The first assumption is one of definition. In discontinuous Markov processes it must be assumed that the system is typified by distinctive states and that transitions occur at discrete time intervals. The actual classification of states adopted depends in part on the processes analysed. Industrial activity, for example, is not distributed in a continuum across the landscape; rather it is concentrated into selected nodes separated by conspicuously non-industrialized areas. A discontinuous system of spatial states may, therefore, be an appropriate framework for a Markov chain model of industrial activity. As with any formal regionalization of the landscape, such a system of spatial states calls for subjectivity in both the selection of states and their delimitation. The same problem arises in a dynamic analysis of industrial structure or of urban systems. States, in these cases, would refer to establishment or city size categories. In such analyses, however, the system of states would of necessity be continuous within the limits of the actual range of the size criteria examined. As in the spatial states, the upper and lower bounds of the structural frequencies must be subjectively determined and the size of the state will affect the estimated transition probabilities. The degrees to which the probabilities and predictions are affected can be determined only by an iterative procedure with a change of limits for each iteration. Generally, the smaller the states the greater is the tendency for "noise" elements to appear in the off-diagonals of the matrix, whereas the larger the states the more pronounced is the main diagonal. It follows that the size of the states and the method of classification may, therefore, determine the underlying Markov property.

(ii) First-order assumption

Methods for verifying the first-order property have been discussed in Section III. Several studies of industrial activity have shown that the new location of a plant will be dependent upon its existing location though not

necessarily on previous locations. It is reasonable to hypothesize then that the pattern of industrial location for Si, at time t, is a function of the industrial location pattern at time t-1 plus some component of change which may be defined by a set of probabilities (Harvey, 1967). The implications of the first-order assumption for a structural analysis of manufacturing activity, as opposed to a strictly spatial analysis, are translatable as the evolution of manufacturing establishments as a size dependent stochastic process. This implies that the underlying determinants of change in the size distribution of plants during one period may be represented by a probability of plant movement from one size category to another; in this we are assuming that such movement is independent of activity in previous time periods, and that economic factors such as entrepreneurship, corporate structure and position, potential technological innovation and scale economies are all correlated with size (Adelman, 1958).

(iii) Assumption of stationarity

The third assumption - that of stationarity - is also dependent in part on the method of classification. Moreover, by definition, this assumption is dependent on the first-order property which is assumed in statistical tests of stationarity; similarly statistical tests for the first-order property assume the constancy of the parameters. The assumption of stationarity requires that the parameters remain constant throughout the predictive period and is therefore a severe constraint; a constraint which lends support to the notion that Markov chains should be used primarily for descriptive rather than predictive purposes (Brown, 1970). Acceptance of the assumption for long-term prediction of industrial activity, for example, may not always be justifiable since technological change could have a significant impact upon existing trends. On the other hand, there is no evidence to suggest that any technological innovation, so far, has profoundly affected the spatial distribution of manufacturing establishments during a short-term period. The two factors most likely to affect the spatial rearrangement of manufacturing activity are direct government subsidies and the construction of new superhighways and airports; but such factors have only a long-term effect. Technological changes may well influence the size structure of industry to a greater degree but again there is little evidence to suggest that even this has occurred for aggregate numbers in the short-term.

(iv) Assumption of uniform probabilities

The fourth assumption concerns the notion of uniform probabilities. The desirability of assigning spatial and structural probabilities to broad groups of plants depends upon careful grouping which again relates to the method of classification and size or content of states. Bourne (1969), for example, was faced with the problem of reducing a 100 x 100 land use matrix into a more manageable 10 x 10 matrix. For industrial location analysis a plant employing 200 employees locating in a country town is unlikely to satisfy the notion of uniform probabilities but as part of a large industrial complex it might well do so. Similarly, in terms of state content a factory manufacturing pencils is more likely to relocate than a plant manufacturing locomotives; plants in urban areas are more likely to relocate than plants in rural locations or vice versa. There is evidence to suggest that branch plants have a higher propensity to relocate than other types of plants of comparative size and conceivably other sub-groups within the broader aggregate categories could be distinguished as having a higher/lower propensity to migrate. In a similar vein sociological studies of occupational mobility have found a marked negative correlation between length of time in any one state and the tendency to move out of that state (Blumen, Kogan and McCarthy, 1955). Such observations encourage the adoption of a Mover-Stayer dichotomy in which one transition matrix represents occupational transients and a second matrix represents those who possess occupational stability. Formally, this may be stated as:

$$S(t+1) = S(t) \times (I-M = S(t) \times Nix\ P) \quad (33)$$

where M is a diagonal matrix of the probabilities that each person in state i is subjected to the mover process, and I is the identity matrix (Brown, 1970). It is unlikely, however, that this mover-stayer dichotomy is applicable to industrial establishments; there is no theoretical or empirical evidence to suggest any correlation between the length of time a plant remains in a location and the likelihood of its relocating. The classification scheme adopted, then, can influence the additional assumptions of the Markov property, stationarity and uniform probabilities.

VI GEOGRAPHICAL APPLICATIONS OF MARKOV MODELS

(i) The existing literature

Geographical applications of Markov models have been reviewed by Brown (1970), and by Collins, Drewett and Ferguson (1974). Brown's review focuses on movement research in human geography and distinguishes between the use of the Markov chain as a descriptive device and as a model of geographic process. In their review Collins, Drewett and Ferguson appraise the use of Markov chains in both human and physical geography but they distinguish between those applications which have been essentially aspatial and those which have adopted a system of spatial states. Brown (1963) was the first to introduce a Markov frame-work to geography in a study of the innovation of diffusion and has since been followed, - among others, by Marble (1964), Scott (1965), Clark (1965), Rogers (1966), Rogers (1968), Harris (1968), Bourne (1969), Compton (1969), Drewett (1969), Collins (1972), Lever (1972 and 1973), Morley and Thornes (1972) and Thornes (1973). A more comprehensive list is referenced in both the reviews mentioned above. In accordance with these reviews the following comments concern the way in which the various applications have devised aspatial and spatial models and also how some models have been used mainly for descriptive purposes whereas others have focused on the predictive capabilities. The section terminates with brief comments relating to some possible modifications and refinements of the basic Markov model.

(ii) Aspatial systems of states

Marble's (1964) application is unusual in that it is one of the few in geography to adopt an absorbing chain. Marble analysed trip structures in Metropolitan Chicago with data taken from the Chicago Area Transportation Study. He calibrated a six state "trip purpose" transition matrix in which the states were designated as home. (the absorbing state), shopping, school, social-recreation, and others. Each state, therefore, is discrete and clearly identifiable. The states, however, are not space dependent so that both the model and its output are aspatial. Another example of an aspatial model is that of Clark (1965) who analysed the change in average contract rents of housing in central city housing tracts. The census tracts were divided into ten classes, each class being a ten dollar (rent) interval; the intervals ranged from 0-10

to 90 and over. In this it is presumed that it is possible for the average contract rent of any one tract to move in any one time period from any one size class to any other. The important point to note here is that the division of these interval classes is an arbitrary arithmetic progression (see also Lever, 1973). Clark points out, for example, that a five rather than a ten dollar interval would possibly decrease the probability of remaining in the starting state and would increase the probability of moving up one or more states. Studies involving the use of size classes in related disciplines such as economics, have in some cases adopted theoretical guidelines for selecting the size of the class intervals. The lognormal distribution which approximates the configuration of many economic variables such as income, wealth, size of firms and cities has long been seen to be the end product of a simple Markov process. Accordingly several economists have used a geometric progression with a fixed ratio to define the size limits of the respective states. In Champernowne's (1953) study a 10910 scale was used, whereas, Hart and Prais (1956) used log2. Lognormal frameworks have been used in geographic contexts by Collins (1973) for the analysis of manufacturing establishments and by Murota (1967) in a study of reservoir sedimentation by landslips of different sizes. Other examples of aspatial models are found in Bourne (1969) and Drewett (1969).

(iii) Spatial systems of states

The problems involved in developing a system of spatial states are much greater and solutions have ranged from the simple to the sophisticated though it is not yet clear whether in the latter case the results are more meaningful. In Collins' (1972) study of industrial activity in Ontario a subjective solution was employed. Within the literature there is ample evidence to suggest that manufacturing establishments relocate once, twice or more times during their existence and several investigators have recognised the tendency of plants to relocate over a considerable range of distances so that spatial states of origin and destination typified by well defined geographic areas are clearly discernible. In Collins' study information concerning the relocations of establishments in Ontario was available for over 500 municipalities but it would not have been computationally feasible to use a transition matrix of this size. Moreover the number of zero entries would have been too large to exhibit any meaningful pattern or trend. It was necessary, therefore, to group the municipalities into spatial units or states. This was done subjectively according to the industrial attractiveness or similarities of the respective municipalities. The six designated states were 1) the city of Toronto; 2) the suburbs of Toronto; 3) the four metropolitan cities of Hamilton, Windsor, London and Ottawa 4) the suburbs of the cities in 3; 5) all other towns with over 10,000 people; 6) the rest of Ontario. Note, however, that the states in this system are not contiguous nor are the components within each state contiguous with one another. Brown and Longbrake (1970) also developed a system of non-contiguous spatial states in their study of intra-urban migration flows in Cedar Rapids, Iowa. Their five state system, however, was determined objectively by using principal components analysis with varimax rotation. This was used to group forty-eight socio-economic variables according to similarities in their variance over 64 origin-destination zones.

One of the few examples of a contiguous system of spatial states is that devised by Lever (1972) in a study of the intra-urban movement of manufacturing in Glasgow. The dominant trend in that city, has been one of suburbanization of manufacturing activity away from the old traditional manufacturing areas centred around the upper Clyde and city centre to the new industrial estates in the suburban periphery. Lever superimposed a kilometre square grid over a map of Glasgow and computed an index for each square on the basis of the proportions of stayers, movers and new firms in each square. The squares were then grouped subject to the constraint that differences within zones were minimized and inter zone differences were maximized. The resulting system of states comprised four contiguous concentric zones.

In each of these three approaches the dominant objective has been to devise a system of states in which the transition probabilities are as uniform or homogeneous as possible. Also the particular configuration of each system of states has been influenced by the prevailing geographic process. Nevertheless, in their review Collins, Drewett and Ferguson (1974) conclude that in general the attempts to calibrate a system of spatial states have been less satisfactory than those dealing with aspatial systems. This inadequacy is seen to stem largely from the lack of spatial theory to provide guidelines as exist for the analysis of the size distribution of firms and cities where the log-normal distribution can be used as a conceptual framework. The over-riding difficulty is that the size, number and configuration of the spatial states will have a strong influence on the outcome of the analyses. Indeed as noted in section V the classification adopted will affect the Markov property, the first-order property, the degree of stationarity and homogeneity of the matrices.

(iv) Markov models as descriptive tools

The most commonly used descriptive Markov property is the equilibrium distribution. In his study of interregional migration flows Rogers (1968, p.92) views the equilibrium vector as a "... kind of speedometer which describes the ultimate consequences of the current movement pattern if it remains unchanged". He computed equilibrium vectors for both white and negro population flows among California's metropolitan areas and by comparing the vectors he derived mobility indices for the two populations. Rogers also used mean first passage time matrices as a measure of "migrant distance". In Rogers study the actual values of these migrant distances proved to be meaningless but when considered in relative terms they suggest interesting conclusions concerning the spatial and aspatial contiguities among California's major S.M.S.A.'s (Rogers, 1968, pp. 96-97). His matrices show, for example, that the non-white migrant distance between the San Francisco-Oakland and San Jose S.M.S.A.'s is thirteen times the distance between the San Francisco and the Los Angeles S.M.S.A.'s and that for selected age groups the migrant distance from San Jose to Sacremento is three times that of the reverse distance. Such matrices are useful also for describing temporal changes. Rogers shows that for almost all the regions sampled the migrant distances declined noticeably between the 1935-40 period and the 1955-60 period - a reflection of increased geographical mobility. Beshers and Laumann (1967) used mean first passage times in a similar context for measuring "social distance", and Collins (1972) used the matrices as a measure of migrant distance for manufacturing establishments. In the latter study the matrix of standard deviations was used to indicate the general unreliability of the mean first passage time values.

(v) Markov models as predictive mechanisms

In general, following the initial example in Section II the Markov chain has been considered as a model of closed systems. If it is desired touse a Markov

model to predict or forecast, at least in a geographical context, it is usually necessary to convert or transform the model. In many geographical studies, for example, it is desirable to model not only the movement which may be observed in a closed system and in which total numbers remain constant through successive time intervals, but also the birth and death of selected variables, such as people, establishments and cities. Temporal observations of these variables are usually typified by a fluctuating population.

There are several alternative procedures for converting a closed Markov model into one which can accommodate the birth and death process as well as movement out of the observed system. Only two alternatives are mentioned here. Rogers (1966) uses separate birth and death vectors which are used as operators in conjunction with the matrix of transition probabilities, a method which has subsequently been adopted by Lindsay and Barr (1972). The second alternative was devised by Adelman (1958) and has been used since by both Lever (1972) and Collins (1972). Lever's original tally matrix estimated from a sample population in a four zone system for 1959-1969 appears as:

	Zone 1	Zone 2	Zone 3	Zone 4
Zone 1	118	13	4	14
Zone 2	6	33	8	6
Zone 3	1	1	68	5
Zone 4	2	0	3	43

(Source: Lever, 1972, p.30)
and as a transition probability matrix

P =	Zone 1	Zone 2	Zone 3	Zone 4
Zone 1	0.79	0.09	0.03	0.09
Zone 2	0.11	0.63	0.15	0.11
Zone 3	0.01	0.01	0.91	0.07
Zone 4	0.04	0.00	0.06	0.90

(Source: Lever, 1972, p.30)
The sample population of births and deaths are arranged alongside the original tally matrix in the following manner

	Z_1	Z_2	Z_3	Z_4	X
Z_1	118	13	4	14	63
Z_2	6	33	8	6	20
Z_3	1	1	68	5	24
Z_4	2	0	3	43	17
X	17	24	17	36	-

(Source: Lever, 1972, p.32)
The bottom row X represents the 1969 locations of births and in-movers into Glasgow, and the right hand column X shows the 1959 locations of deaths and out-movers from Glasgow. To complete this new matrix the problem is one of filling the bottom right hand square which may be viewed as a reservoir acting as a source of potential entrants into the system and as a pool for

liquidated firms. The difficulty was first overcome by Adelman (1958) claiming that any large number will suffice and that the size of the number will not affect the final prediction of the model. To support this claim she provides a mathematical proof. Lever endorses this with empirical proof by comparing the results of two matrices. In the first matrix he assigns a value of 906 to the reservoir so that the elements of row X total 1000. Lever's transition matrix with birth and death vectors thereby takes the form

		Z_1	Z_2	Z_3	Z_4	X
P =	Z_1	0.56	0.06	0.02	0.07	0.29
	Z_2	0.08	0.41	0.11	0.08	0.27
	Z_3	0.01	0.01	0.69	0.05	0.24
	Z_4	0.03	0.00	0.05	0.66	0.26
	X	0.02	0.02	0.02	0.04	0.90

(Source: Lever, 1972, p.32)

The original numbers of sample firms in the four zones in 1959 were 212, 73, 99 and 65 respectively, and in X 1000. For the whole system the initial probability vector $p^{(0)}$ is, therefore,

$$(0.147, 0.151, 0.068, 0.043, 0.692)$$

To derive $p^{(1)}$ - the distribution in 1969 - the initial probability vector is multiplied by the transition matrix **P** so that

$$p^{(1)} = p^{(0)} \mathbf{P} = (0.100, 0.047, 0.072, 0.073, 0.708)$$

This means that of the 447 actual firms in existence in 1959 and the 1000 potential entrants in the 1959-69 period, there is a 0.100 probability that a firm will be in Zone 1 in 1969, a 0.47 probability that it will be in Zone 2 and so on. Put another way, of the 1447 actual and potential firms 10 percent of them would be in Zone 1 in 1969, 7.2 per cent would be in Zone 3 and so on. Lever continues to power the original transition matrix until the system reaches equilibrium at $p^{(8)}$. The intermediate probability distributions are listed in Table 12.

Table 12: Successive values of $p^{(n)}$

	Zone 1	Zone 2	Zone 3	Zone 4	X
$p^{(0)}$	0.147	0.050	0.068	0.043	0.692
$p^{(1)}$	0.100	0.047	0.072	0.073	0.708
$p^{(2)}$	0.077	0.043	0.075	0.090	0.715
$p^{(3)}$	0.064	0.040	0.078	0.100	0.718
$p^{(4)}$	0.057	0.038	0.078	0.105	0.721
$p^{(5)}$	0.054	0.037	0.079	0.109	0.721
$p^{(6)}$	0.051	0.035	0.080	0.112	0.722
$p^{(7)}$	0.050	0.034	0.080	0.113	0.723
$p^{(8)}$	0.049	0.034	0.080	0.115	0.722

(Source: Lever 1972, p.33)

Adelman shows that at this stage it is possible to disregard the proportion of firms in X since the concern is only with those plants in existence. Thus by summing the proportions in Z_1, Z_2, Z_3 and Z_4 for each $p^{(n)}$ Lever computes the percentage of firms expected in each time period. For t_1 (1969) the sum is 0.292 and the respective percentages in each Zone are 34, 16, 25 and 25. The series of percentages tabulated by Lever are listed in Table 13.

Table 13: Predicted distribution of firms (percentage)
(X = 1000)

	Zone 1	Zone 2	Zone 3	Zone 4
t_o (1959)	48	16	22	14
t_1 (1969)	34	16	25	25
t_2 (1979)	27	15	26	32
t_3	23	14	28	35
t_4	20	14	28	38
t_5	19	14	28	39
t_6	18	13	29	40
t_7	18	12	29	41
t_8 (2039)	18	12	29	41

(Source: Lever, 1972, p. 33)

Lever is able to conclude, therefore, that if current trends continue then by 2039 Zone 1 will account for 18 percent of all firms in Glasgow, whereas in 1959 it accounted for almost half (48%). Conversely Zone 4 will increase its share from 14 percent in 1959 to 41 percent in 2039. Lever then proceeds to verify Adelman's mathematical proof by demonstrating that when the reservoir size is doubled to 2000 the predicted proportions (Table 14) are almost the same as those derived from a reservoir of 1000.

It must be noted at this stage that both Adelman's proof and Lever's empirical verification refer only to the predicted proportions and carry no implications concerning the total number of establishments in Glasgow (Lever, 1972, p.35). By changing the size of the reservoir the predicted total number of plants may be influenced quite considerably even though the proportions in each state will remain the same. Thus, although the Markov predictions may indicate that the proportion of plants in Zone 1 will be halved relative to all plants, the actual number of plants in that Zone may increase. The size of the reservoir, therefore, must be treated as an empirically derived parameter and its selection must be undertaken with caution if predicted numbers are required.

(vi) Modifications and refinements

This sub-section is deliberately terse and contains only brief reference to some modifications of the basic Markov model. At this stage it is necessary to refer to the individual studies for more detailed treatment.

In 1967 Hamilton noted that Markov chains "... are still in their infancy..." and in terms of subsequent geographical applications there is no real evidence

Table 14: Predicted distribution of firms (percentages)
(X = 2000)

	Zone 1	Zone 2	Zone 3	Zone 4
t_o (1959)	48	16	22	14
t_1	34	16	25	25
t_2	27	15	26	32
t_3	23	14	27*	36*
t_4	20	13*	28	39
t_5	19	13*	28	40*
t_6	19*	12*	29	40
t_7	18	12	29	41
t_8	18	12	29	41

Note: *value differs from Table 13

(Source: Lever, 1972, p.34).

to suggest that the technique has been developed much beyond a youthful stage. Much work has been done and much progress has been made but, in general, conceptual developments have kept well ahead of actual applications. The main difficulty of operationalizing the various conceptual developments has been lack of suitable data.

Mention has been made already of Scott's early attempt to derive Markov transition probabilities by using a linear-programming procedure. In a more recent study which compares the viability of the Markov framework with Monte Carlo simulation Lindsay and Barr (1972) follow Brown's (1970, p.399) suggestion of using a gravity interaction model to estimate the transition probabilities.

One limitation of the standard transition probability matrix, is that it can be used only to provide estimates for time intervals equal to the time interval covered by the initial matrix. Hence, a probability matrix for the 1950-1960 period (Bourne 1969) can be used only for providing decennial estimates, while a 1961-1965 matrix (Collins, 1972) provides only quinquennial estimates. To provide annual forecasts from such a matrix Collins computed an average annual matrix through the fractional dissaggregation of the longer period matrix. The relevant root of the matrix can be computed either by Newtonian approximation or by the binomial theorem (Waugh and Abel, 1967).

A common characteristic of many aspatial matrices, especially those in which the transition probabilities relate to change in town size or establishment size, is the existence of discontinuities in the probability distribution across the rows. Such discontinuities represent an incomplete estimation of the underlying fixed probabilities. If it is observed during a given time interval that manufacturing establishments can increase their size (expressed in terms of say employment or sales) from category 3 to categories 4, 5, and 7 then it is reasonable to expect, that, given enough observations, establishments could also increase their size to category 6 - though not necessarily

to 8, 9 or higher. Theoretically, therefore, there is an underlying fixed probability of establishments increasing their size from class 3 to 6. To estimate this "missing" probability it is possible to fit a smoothing surface to the transition matrix (Collins, 1974). The surface will fill in the discontinuities and will modify the remaining probabilities accordingly. Such a procedure is comparable to the fitting of conventional trend surfaces to observed data points (Tobler, 1967). Yet another way of modifying estimated transition probabilities in the standard matrix is by incorporating the notion of entropy into the Markov framework (Fano, 1969; Drewett and Bell, 1974).

More complex Markov models have been suggested by Olsson and Gale (1968). They recognise that traditional Markov models employ ordinary matrix structures which limit the number of analysed variables to one. If Markov models are to be applied to spatial behaviour present knowledge is too scant to suggest which of the many possible variables should be employed. Olsson and Gale claim that the answer to this problem comes in a logical extension of the matrix form such that the ordinary two-dimensional matrix is expanded into three dimensions or, in the more general case, into a matrix of q-dimensions.This notion of q-way matrices or multidimensional Markov frameworks remains to be explored. Another brief suggestion which needs to be investigated is the combining of Markov chain theory with input-output analysis as a technique for constructing better models of inter-regional migration (Hamilton, 1967, p.412).

The modifications discussed so far refer only to Markov models using discrete-space and discrete-time. Other distinctive Markov processes are classified in Collins, Drewett and Ferguson (1974) as 1) discrete-space and continuous-time; 2) continuous-space and discrete-space; and 3) continuous-space and continuous-time. In dealing with any form of continuous Markov process the mathematics become increasingly difficult and sometimes intractable. It is not surprising, therefore, that there are only a few such applications of geographical relevance (Harris, 1966; Drewett, 1969; Drewett and Bell, 1974). Finally all the comments made so far refer only to the concept of first-order Markov chains. Some processes need to be modelled by second-order or higher-order chains. Applications of this nature have been limited mainly to earth sciences such as climatology (Pattison, 1965) and geomorphology (Thornes, 1973).

It should be clear, therefore, that much more research is required in all branches of geography before it is possible to make firm statements concerning the viability of Markov theory as an explanatory and predictive framework in geographical analysis. The exploratory stage is not yet over.

BIBLIOGRAPHY

A. Basic texts for mathematical treatment

1. KEMENY, J.G. and SNELL, J.L. (1967), *Finite Markov chains.* Princeton, New Jersey. D. Van Nostrand Co.
2. FELLER, W. (1969), *An introduction to probability theory and its applications.* Vol. 1. Third Edition. New York: John Wiley & Sons.

B. Statistical tests relating to Markov theory

3. ANDERSON, T.W. and GOODMAN, L.A. (1957), Statistical inference about Markov chains. *Annals, Mathematical Statistics,* 28, 89-109.
4. KULLBACK, S., KUPPERMAN, M., and KU, H.H. (1962), Tests for contingency tables and Markov chains, Technometrics, 4, 572-608.

C. Computer programs for Markov chain analysis

5. MARBLE, D. (1967), *Some computer programs for geographic research.* Special Publication No. 1, Department of Geography, Northwestern University.

D. Markov applications and examples

6. BOURNE, L.S. (1969), A spatial allocation land use conversion model of urban growth. *Journal, Regional Science,* 9, 261-272.
7. CLARK, W.A.V. (1965), Markov chain analysis in geography: an application to the movement of rental housing areas. *Annals, Association of American Geographers,* 55, 351-359.
8. COLLINS, L. (1972), *Industrial migration in Ontario: forecasting aspects of industrial activity through Markov chain analysis.* Ottawa: Statistics Canada.
9. COMPTON, P.A. (1969), Internal migration and population change in Hungary between 1959 and 1965. *Transactions, Institute of British Geographers,* 47, 111-130.
10. HARVEY, D. (1967), Models of the evolution of spatial patterns in human geography. in: *Models in Geography,* (eds) R.J. Chorley and P. Haggett, London: Methuen Ch. 14.
11. LEVER, W.F. (1972), The intra-urban movement of manufacturing: a Markov approach. *Transactions, Institute of British Geographers,* 56, 21-38.
12. LEVER, W.F. (1973), A Markov approach to the optimal size of cities in England and Wales. *Urban Studies,* 10, 353-365.
13. MARBLE, D.F. (1964), A simple Markovian model of trip structures in a Metropolitan region. *Papers, Regional Science Association, Western Section,* 1, 150-156.
14. ROGERS, A. (1966), Matrix methods of population analysis. *Journal, American Institute of Planners,32,* 40-44.

E. Further Reading

15. ADELMAN, I.G. (1958), A stochastic analysis of the size distribution of firms. *Journal, American Statistical Association,53,* 893-904.
16. BESHERS, J.M. and LAUMANN, E.O. (1967), Social distance: a network approach. *American Sociological Review,* 32, 225-236.
17. BLUMEN, I., KOGAN, M. and McCARTHY, P.J. (1955), *The industrial mobility of labor as a probability process.* Ithaca, New York: Cornell University Press.
18. BROWN, L.A. (1963), The diffusion of innovation; a Markov chain-type approach, Discussion Paper No. 3. Department of Geography, Northwestern University.
19. BROWN, L.A. (1970), On the use of Markov chains in movement research. *Economic Geography,* 46 (Supplement), 393-403.
20. BROWN, L.A. and LONGBRAKE, D.B. (1970), Migration flows in intra-urban space: place utility considerations. *Annals, Association American Geographers,* 60, 368-384.
21. CHAMPERNOWNE, D.G. (1953), A model of income distribution. *Economic Journal,* 63, 318-351.
22. COLLINS, L. (1973), Industrial size distributions and stochastic processes. in: *Progress in Geography,* 5, (eds) C. Board et a/. London: Edward Arnold, 121-165.
23. COLLINS, L. (1974), Estimating Markov transition probabilities from micro-unit data. *Journal, Royal Statistical Society,* Series C, 23, 355-371.
24. COLLINS, L., DREWETT, R. and FERGUSON, R. (1974), Markov models in geography. *The Statistician,* 23, 179-209.
25. DREWETT, J.R. (1969), A stochastic model of the land conversion process. *Regional Studies,* 3, 269-280.
26. DREWETT, J.R. and BELL, E.J. (1974), Stochastic models of urban analysis. *Progress in Planning,* (forthcoming).
27. HAMILTON, F.E.I. (1967), Models of industrial location. in: *Models in Geography,* (eds) R.C. Chorley and Peter Haggett, London: Methuen Ch. 10.
28. HARRIS, C.C. (1968), A stochastic process model of residential development. *Journal, Regional Science,* 8, 29-39.
29. HART, P.E. and PRAIS, S.J. (1956), The analysis of business concentration: a statistical approach. Journal, *Royal Statistical Society,* 119, 150-181.
30. KRENZ, R.D. (1964), Projections of farm numbers for North Dakota with Markov chains. *Agricultural Economics Research,* 16, 77-83.
31. LINDSAY, I. and BARR, B.M. (1972), Two stochastic approaches to migration: comparison of Monte Carlo simulation and Markov chain models. *Geografiska Annaler,* 54B, 56-67.

32. MORLEY, C.D. and THORNES, J.B. (1972), A Markov decision model for network flows. *Geographical Analysis,* 4, 180-193.

33. PATTISON, A. (1965), Synthesis of hourly rainfall data. *Water Resources Research,* 1, 489-498.

34. OLSSON, G. and GALE, S. (1968), Spatial theory and human behaviour. *Papers, The Regional Science Association, 21*, 229-242.

35. ROGERS, A. (1968), *Matrix analysis of interregional population growth and distribution.* Berkeley: University of California Press.

36. SCOTT, A.J. (1965), A procedure for the estimation of Markov transition probabilities, Discussion Paper No. 8, Regional Science Research Institute, University of Pennsylvania.

37. THORNES, J.B. (1973), Markov chains and slope series: the scale problem. *Geographical Analysis,* 5, 322-328.

38. TOBLER, W.C. (1967), Of maps and matrices. *Journal, Regional Science,* 7, (Supplement), 275-280.

39. WAUGH, F.U. and ABEL, M.E. (1967), On fractional powers of a matrix. *Journal, American Statistical Association,* 62, 1018-1021.