C. 1

STOCHASTIC MODELS OF CHANGES IN POPULATION DISTRIBUTION AMONG CATEGORIES

by

YIGAL GERCHAK

B.A., Tel-Aviv University, 1970

M.Sc., Tel-Aviv University, 1975

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Faculty of Commerce and Business Administration)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 1980

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Faculty ~~Department~~ of Commerce and Business Administration

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date Oct. 23, 1979

Supervisor: Dr. Shelby L. Brumelle

ABSTRACT

There are very many processes in the natural and social sciences which can be represented as a set of flows of objects or people between categories of some kind. The Markov chain model has been used in the study of many of them. The basic form of the Markov chain model is, however, rarely adequate to describe social, occupational and geographical mobility processes. We shall therefore discuss a number of generalizations designed to introduce greater realism.

In Chapter I we formulate and investigate a general model which results from relaxing the assumptions of sojourn-time's memorylessness and independence of origin and destination states, and of population homogeneity. The model (a mixture of semi-Markov processes) is then used in two ways. First, it provides a framework in which various special cases (which correspond to models which were used by social scientists) can be analytically compared. We pay particular attention to comparisons of rate of mobility in related versions of various models and to compatability of popular parametric forms with observed mobility patterns. Second, any result obtained for the general model can be specialized for the various cases and subcases.

In Chapter II we formulate a system-model allowing interaction among individuals (components), which has been motivated by Conlisk. We define processes on this model and analyze their properties. A major effort is then devoted to establishing that when the population size becomes large, this rather complex stochastic model can be approximated by a single deterministic recursion due to Conlisk (1976). Nevertheless, we draw attention to certain aspects (particularly steady-state behavior) in which the approximation may fail.

In Chapter III we address ourselves to the issue of measurement of (what we refer to as) social inheritance in intergenerational mobility processes. We distinguish between various aspects and concepts of social inheritance and outline the implications that certain "social values" may have on constructing a measure (or index). In the mathematical discussion which follows certain mechanisms for generating "families" of measures are indicated, and the properties of some particular combinations are investigated.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I wish to thank my thesis supervisor, Professor Shelby L. Brumelle, for his invaluable help in this research. His knowledge and rigorousness as well as patience and friendliness were a real asset to me, and I learned a lot from him.

I am also grateful to my committee members, Professors K. R. MacCrimmon, I. Vertinsky and J. Walsh for their help and encouragement.

This is an opportunity to thank the following people for helping me through the Ph.D. program. To Professors M. Avriel and W. T. Ziemba for "bringing" me to U.B.C.; to Ms. B. Wynne-Edwards and Professor K. R. MacCrimmon for helping me in obtaining financial assistance; and to the U.B.C. awards-office for granting it; to Professor L. G. Mitten for enabling me to get teaching experience and to Ms. Colleen Colclough and Ms. Suzan Altimas for excellent typing of this thesis.

I would also like to thank my fellow doctoral students, and especially E. Choo, J. Kallberg and R. Solanki for their constant help and friendliness.

Finally, I owe a special debt of thanks to my family, without whom all this would have been impossible.

INTRODUCTION

Since the early fifties, researchers in various disciplines have attempted to devise models to describe the dynamics of social systems. The first models described changes in voting behavior (Anderson 1954), social mobility (Prais 1955) and occupational mobility (Blumen et al. 1955). These models were stochastic in nature, and used Markov chains as their main tool. Markov chains were used later to model other types of social systems,(*) such as geographical mobility [for a recent survey of models used in various mobility processes see Stewman (1976)], diffusion of innovations, educational systems, and flow of people among different states of health [see Bartholomew (1973) for a general survey]. They were also used to model buying behavior (brand switching; e.g. Massy et al. 1970), internal labour markets (e.g. Vroom and MacCrimmon 1968) and health delivery systems (e.g. Meredith 1973).(**)

In the area of intergenerational social mobility, in addition to the interest in devising descriptive models, there was an accompanying interest in <u>measuring mobility</u> of certain societies. The first attempt to devise an index (measure) of mobility was made by Prais (1955a), and was followed, among others, by Matras (1960), Yasudo (1964), Goodman (1969), Bartholomew (1973, pp. 23-7), Boudon (1973), Pullum (1975) and Sommers and Conlisk (1979). Recently Shorrocks (1978) approached this issue more systematically, setting

(*) Similar work was done in the biosciences for nonhuman populations (e.g. Thompson and Vertinsky 1975).

(**) The term "social systems" to be used here should be broadly interpreted to include all the above systems.

some desired properties of such indices as axioms and making inferences from them. The above mentioned measures were suggested for a Markov chain model, and are actually functions defined on probability transition matrices ("mobility tables").

More recently, however, researchers found it necessary to generalize and extend the basic model. In order to understand what motivated this trend, one should realize that, in general, a new model of a certain phenomenon usually emerges as a result of one of three causes.

The first, and most natural to those who view models as devices for fitting data, is poor fit (or inadequate predictions) of the existing model. In the area of intragenerational mobility, researchers observed an empirical regularity in the form of "clustering on the main diagonal", which standard Markov chain models were not able to account for. This prompted some researchers to relax the assumption of "population homogeneity" (implicit in early models), resulting in the "Mover-Stayer Model" and its subsequent extensions (Blumen et al., 1955, McFarland 1970, Spilerman 1972a, 1972b).(*)

Another rationale for extending existing models occurs when some behavioral patterns, which were unknown before, are discovered. Typical examples, in the area of geographical migration processes, include the realization that duration in one's current location reduces the chances of leaving it ("Axiom of Cumulative Inertia"; McGinnis 1968, Huff and Clark 1978), the demonstration that the destination one moves to is not independent of the duration in one's current location (Ginsberg 1978a) and

(*) Similar extensions were proposed also in buying-behavior literature (Jones 1973, Givon and Horsky 1978).

observation of decline of mobility rate with age (Mayer 1972). An attempt to accomodate such effects gave rise to the "Cornell Mobility Model" (McGinnis 1968, Henry et al. 1971) and, more generally, to a semi-Markov model (Ginsberg 1971).(*) A somewhat related avenue of research involves hypothesizing some logical patterns and using a model to test them. An example is the work by Wise (1975) on the effects of academic achievement on graduates' careers.

In order to understand the third class of rationales for extending models, it is important to realize that "...not all mathematical models are intended to fit empirical data; not infrequently mathematical models are developed to work out implications of postulates..." (McFarland 1974, p. 883). In fact, this latter rationale for modelling becomes central when we move from the field of Sociology (where the above citation appeared) to the policy sciences, where one frequently wants to assess future impacts of policies which might be implemented.

With particular attention to the rationale for modelling discussed in the above paragraph, and in order to come up with a comprehensive model which captures more features of real systems, Conlisk (1976) argued in favor of dropping the assumption of individuals "moving" independently of each other, which was implicit in all above mentioned models. Conlisk formulated a model which allows for interactions among individuals, and Smallwood and Conlisk (1978) made an interesting use of some of its versions to model adaptive behavior of poorly informed consumers.

(*) This has been made possible by the development of the theory of semi-Markov processes in the early sixties and by using certain notions from reliability theory which were developed about the same time.

While Markov and semi-Markov models could make direct use of available stochastic process theory, once population heterogeneity or interaction between individuals is introduced, the resulting models have no immediate counterparts in classical theory, so some specific analysis is required. A particular problem is that "Traditional Markov chain theory pertains to a single object moving from state to state; but in applications to social mobility and other social systems, one considers an entire population, each person moving probabilistically from state to state" (McFarland 1970, p. 463). As long as the processes which individuals in the population follow were assumed to be (stochastically) the same and independent of each other, the law of large numbers justifies approximating the proportion of the population (assumed large) in a given state by an individual's probability of being in that state. Without these assumptions, however, one should be very careful when formulating models and attempting to infer population-level (macroscopic) quantities from individual-level (microscopic) ones and vice versa.

This work addresses shortcomings of existing literature on three levels:

1. By a careful formulation and analysis of models. Since relaxing the assumption that the individuals behave independently of each other seems to be the most rewarding in terms of the spectrum of systems it may enable us to model, and since it seems to be the most difficult step towards a comprehensive system-model, we devote a major effort to such a model.
2. By providing a logical hierarchy for various models (with

particular attention to those with heterogeneous populations) within the theory of stochastic processes.

3. By testing analytically whether some models indeed have the desired properties, and whether they always alter predictions in a manner which is consistent with the way they were motivated.

4. By differentiating among various aspects of social mobility, singling out, for purposes of measurement, a pure social-inheritance aspect, and approaching it systematically on both substantive and mathematical levels.

In Chapter I we formulate and investigate a general model which results from relaxing the assumptions of sojourn-time's memorylessness and independence of origin and destination states, and of population homogeneity. The model (a mixture of semi-Markov processes) is then used in two ways. First, it provides a framework in which various special cases (which correspond to models which were used by social scientists) can be analytically compared. We pay particular attention to comparisons of rate of mobility in related versions of various models and to compatability of popular parametric forms with observed mobility patterns. Second, any result obtained for the general model can be specialized for the various cases and subcases.

In Chapter II we formulate a system-model allowing interaction among individuals (components), which has been motivated by Conlisk. We define processes on this model and analyze their properties. A major effort is then devoted to establishing that when the population size becomes large, this rather complex stochastic model can be approximated by a single deterministic recursion due to Conlisk (1976). Nevertheless, we draw attention to certain aspects (particularly steady-state behavior) in which the approximation may fail.

In Chapter III we return to the issue of measurement of (what we refer to as) social inheritance in intergenerational mobility processes. We distinguish between various aspects and concepts of social inheritance and outline the implications that certain "social values" may have on constructing a measure (or index). In the mathematical discussion which follows, certain mechanisms for generating "families" of measures are indicated, and the properties of some particular combinations are investigated.

Although the models in Chapters I and II were motivated on substantive grounds, some of the observations and results there may be viewed as context-free and, hopefully, perhaps of an independent probabilistic interest. Chapter III, on the other hand, is geared, essentially, to intergenerational social mobility issues only.

Despite the common sources and rationales of the various models and problems addressed by this work (which were outlined in this introduction), the technical aspects of the various chapters (and sections) vary significantly. This is one of the reasons why the three chapters are self-contained and with little cross-references. The main mathematical "tools" used in this work are:

- Chapter I: Markov and semi-Markov processes (Sections 1.1 - 1.3; Ross 1970, Cinlar 1975); stochastic dominance (subsection 1.4.1; Brumelle and Vickson 1975); spectral representations (subsection 1.4.2; Cinlar 1975, Appendix); reliability theory (particularly in section 1.5; Barlow and Proschan 1975).

- Chapter II: Products of finite non-negative square matrices (subsection 2.1.4; Hajnal 1958); mathematical probability (throughout: Breiman 1968); weak convergence of probability measures (section 2.2 and, particularly 2.3; Billingsley 1968).
- Chapter III: Linear Transformations (subsection 3.2.1; Halmos 1958); Orders (subsection 3.2.2; Krantz et al. 1971); metrics and norms (subsection 3.2.3; Royden 1963); ergodic coefficient, rate of convergence of Markov chains (subsection 3.2.4; Isaacson and Madsen 1976); information theory (subsection 3.2.4; Khinchin 1957); rate of convergence notions (subsection 2.3.5; Ortega and Rheinboldt 1970).

Due to this rather "local" use of various concepts, we have chosen to define concepts and quote results only when we need them. Nevertheless, we attempted to make the work virtually self-contained.

CHAPTER I

ANALYTICAL COMPARISON OF MOBILITY MODELS IN A HETEROGENEOUS SEMI-MARKOV CONTEXT

When Markov chain models failed to describe certain aspects of human mobility, and when empirical evidence showed that their assumptions were not compatible with human behavior in certain types of social systems, researchers suggested two major directions of extending them. One was the Mover-Stayer Model and its subsequent extensions (Blumen, Kogan and McCarthy 1955, McFarland 1970, Spilerman 1972b), which introduced population heterogeneity. The other direction was semi-Markov models (McGinnis 1968, Ginsberg 1971), which allowed the distribution of time between moves to depend on the origin and destination states. A semi-Markov model has also been used to model movement of personnel through a hierarchical organization (Grinold and Marshall 1977, Section 4.4).

Since researchers were typically interested in explaining particular phenomena and modeling specific systems, they usually formulated rather specialized models[(*)] (e.g. "cumulative inertia" of duration length - a particular form of a semi-Markov process). Even when they related models to each other, the basis of comparison was their relative success in fitting a given set of data. Very little was done towards arranging the various models in some logical order, establishing relations among them, and comparing their predictions analytically.[(**)]

(*) Some exceptions are the works of Ginsberg (1971), Singer and Spilerman (1974, 1976) and Schinnar and Stewman (1978).

(**) One exception, in the area of personnel prediction models, was a theoretical comparison of a cross-sectional (Markov) model and a longitudinal (cohort) model by Marshall (1973) [see also Grinold and Marshall (1977, Section 4.5)].

In this chapter, following Singer and Spilerman (1974), we formulate a general model (a mixture of semi-Markov processes) which includes all of the models mentioned above as special cases. We use this model in two ways. First, it provides a framework in which the various special cases can be compared. Second, any result obtained for the general model will hold for the various special cases (possibly assuming some special forms). This is done in Sections 1 through 3.

As was pointed out in the literature, (e.g. Singer and Spilerman 1974) there are two equivalent ways of interpreting mixtures of stochastic processes. The population may be considered to consist of subpopulations which follow distinct processes.(*) Alternatively, each individual may be considered to "draw" the process that he will follow from some probability distribution over processes (or parameters). The wording of our formulations will follow the latter interpretation.

One property to which we pay particular attention is the rate of mobility (Section 4). Many of the above extensions of the simple Markov model were motivated by the empirically observed fact that the simple model overestimated some measures of mobility. We show that particular extensions of continuous-time Markov chains in the direction of "cumulative inertia" duration times (McGinnis 1968) and in the direction of Mover-Stayer models (Spilerman 1972b) result in stochastically longer durations and in stochastically fewer transitions in any time interval (subsection 4.1).

Another comparison concerns the "extended mover-stayer model with

(*) See also Lazarsfeld and Henry (1968).

rate heterogeneity" (Spilerman 1972b). Bartholomew (1973, pp. 48-49) wondered whether an attempt to model such a system as a discrete-time Markov chain always underestimates the proportion of individuals who remain in their initial state. A counterexample shows that this is not always true, but we give some sufficient conditions for its validity (subsection 4.2).

We also check whether some specific parametric forms of the models, which were suggested in the literature, have properties which March and March (1977, p. 380) consider desirable for models of career patterns (Section 5).

1.1 A General Model and Some of Its Properties

Suppose that $\mathcal{K}$ is a finite set of categories, e.g. K regions $\mathcal{K} = \{1,2, \ldots, K\}$. Let $X(t)$ be the category of a given individual at time t. The stochastic process $\{X(t):t\geq 0\}$ is defined as follows.

(i) The given individual chooses a parameter Z from a set of parameters $\mathcal{A}$(*), according to a probability measure μ. i.e., for any event $A \subseteq \mathcal{A}$, $\mu(A) = \Pr[Z \in A]$. If $\mathcal{A} = R_+^K$ we shall denote the corresponding distribution function by G.

(ii) Given $Z = z$ we assume that $\{X_z(t):t\geq 0\}$ is a semi-Markov process.(**) In order to characterize the semi-Markov

(*) $\mathcal{A}$ can be interpreted as the population. Individuals can be identified with $z \in \mathcal{A}$.

(**) Note that the unconditional process $\{X(t):t\geq 0\}$ will in general _not_ be semi-Markov.

process (Cinlar 1975) one needs to specify the two quantities ${}_zB(t)$ and ${}_zQ(t)$ defined next.

a) ${}_zB(t)$ is the distribution of initial conditions at time 0. More precisely, it is a matrix-valued function such that ${}_zB_{ij}(t)$ is the joint (conditional of Z = z) distribution of initial category i, the category j to which the first transition is made and the time until this transition. Let $\nu_z(i)$ be the (marginal) probability of X(0) = i (given Z = z) associated with this joint distribution; i.e., $\nu_z(i) = [\Pr X(0)=i|Z=z] = \sum_{j=1}^{K} {}_zB_{ij}(\infty)$. For every i, let Γ_i be the conditional probability measure of Z given X(0) = i; i.e., $\Gamma_i(A) = \Pr[Z\varepsilon A|X(0)=i]$. It is obtained from $\nu_z(i)$ and μ via Bayes' Formula.

b) ${}_zQ_{ij}(t)$ is the conditional probability, given that Z = z and that transition into category i has been made at time s, that the next transition will be into state j and will occur before time s + t. Since this probability does not depend on s, the process $\{X_z(t):t\geq 0\}$ is time-homogeneous.

For every semi-Markov process $\{X_z(t):t\geq 0\}$ define:

${}_zP_{ij} = {}_zQ_{ij}(\infty)$ = probability that the state which will be occupied after i is j.

${}_zF_{ij}(t) = {}_zQ_{ij}(t)/{}_zP_{ij}$ = probability that, given that the process occupies state i at time s and later moves to state j, this transition will take place before time s + t. (If ${}_zP_{ij} = 0$ then ${}_zF_{ij}(t)$ is arbitrary.)

$${}_z\mu_i = \int_0^\infty t\,d\Big(\sum_{j=1}^{K} {}_zQ_{ij}(t)\Big) = \text{mean time between transitions in state } i.$$

${}_zN(t)$ = number of transitions in $(0,t]$.

${}_zP_{ij}(t) = \Pr[X_z(t) = j \mid X_z(0) = i]$.

In general, any function defined on the stochastic processes $\{X_z(t):t\geq 0\}$, will be indexed by z. The same function defined on $\{X(t):t\geq 0\}$ will be denoted by the same symbol, but without the index. For example,

$$P_{ij}(t) = \Pr[X(t) = j \mid X(0) = i] = \int_A \Pr[X(t) = j \mid X(0) = i,\ Z = z]\ \Gamma_i(dz).$$

We will, in this case, use the matrix notation $P(t) = E({}_ZP(t))$ for the above integrals.

The behavior of the process $\{X(t):t\geq 0\}$ can be deduced from the behavior of the processes $\{X_z(t):t\geq 0\}$. Of particular interest are results about the limiting behavior of the process $\{X(t):t\geq 0\}$, e.g. about $\lim_{t\to\infty} P(t)$, which is equal to $\lim_{t\to\infty} E({}_ZP(t))$. Since each ${}_zP_{ij}(t)$ is bounded (between zero and one), it follows from Lebesgue's Dominated Convergence Theorem that $\lim_{t\to\infty} E_Z P(t) = E \lim_{t\to\infty} {}_ZP(t)$. However, since the processes $\{X_z(t):t\geq 0\}$ are semi-Markov, Theorem 5.16 in Ross (1970, p. 104) show that if ${}_zP$ is irreducible and aperiodic with steady state probabilities ${}_z\pi_i$, then

$$\lim_{t\to\infty} {}_zP_{kj}(t) = \frac{{}_z\pi_j \cdot {}_z\mu_j}{\sum_{i=1}^{K} {}_z\pi_i \cdot {}_z\mu_i} \quad \text{for every } k\ . \qquad (1.1)$$

Hence

$$\lim_{t\to\infty} P_{kj}(t) = E \lim_{t\to\infty} {}_Z P_{kj}(t) = E \frac{{}_Z\pi_j \cdot {}_Z\mu_j}{\sum_{i=1}^{K} {}_Z\pi_i \cdot {}_Z\mu_i} \quad \text{for every } k. \tag{1.2}$$

Consider a special case in which for every $Z = z$, ${}_z\mu_i = {}_z\mu$ for every i. Then the above equation reduces to $\lim_{t\to\infty} P_{kj}(t) = E_Z\pi_j = E \lim_{m\to\infty} ({}_ZP^m)_{kj}$ for every k. (The limiting behavior of the special case ${}_z\mu = 1$ for every z and μ was analyzed by Morrison et al. 1971.) If, in addition, ${}_z\pi = \pi$ for every z, then

$$\lim_{t\to\infty} P_{kj}(t) = \pi_j \ . \tag{1.3}$$

Some of the specialized models discussed in the next subsection have these properties.

1.2 Special Cases

In this section we identify some important special cases of the general model which has been used to model mobility. We assume throughout this section that a transition has just occured at time 0, so that ${}_zB_{ij}(t) = \nu_z(t) \cdot {}_zQ_{ij}(t)$.

The models are divided into two main classes. Those in class A correspond to homogeneous populations and those in class B correspond to heterogeneous populations.

A. In the models numbered 1 through 4 below, we shall assume that ${}_zQ$ and that ν_z do not depend on z, so that ${}_zQ = Q$ and $\nu_z = \nu$ for each $z \in A$. The process $\{X(t):t\geq 0\}$ then reduces to a semi-

Markov process(*) (Ginsberg 1971, 1978a, 1978b, Grinold and Marshall 1977, Section 4.4).

1) $F_{ij}(t) = \begin{cases} 0 & t < 1 \\ 1 & t \geq 1 \end{cases}$ for every i,j.

This is the case where X(t) is a discrete-time Markov chain - the "classical" model (see, for example, Blumen et al. 1955).

2) $F_{ij}(t) = 1 - e^{-\lambda_i t}$, $t \geq 0$ for every i,j.

This is the case of a continuous-time Markov chain (see, for example, Coleman 1964, Tuma, Hannan and Groenveld 1979). A special subcase is (2s): $\lambda_i = \bar{\lambda}$ for every i. $\{N(t):t\geq 0\}$ then becomes a Poisson process.

3) The F_{ij}'s are "decreasing failure rate" (DFR) distributions; i.e. $\bar{F}_{ij}(x + t)/\bar{F}_{ij}(t)$ is increasing in $t \geq 0$ for each $x \geq 0$. This is equivalent to the "Axiom of Cumulative Inertia".

   a) $F_{ij}(t)$ is arithmetic, i.e. $F_{ij}(t^+) \neq F_{ij}(t^-)$ only if t is an integer.

   This is the "Cornell Mobility Model" (McGinnis 1968, Henry et al. 1971).

   b) $F_{ij}(t) = F(t)$ for every i and j (which reduces N(t) to a renewal process), and F(t) is a mixture of exponential distributions; i.e. $F(t) = E(1 - e^{-Yt})$, $t \geq 0$, where the random variable Y has distribution function L.

(*) Bartholomew (1973, p. 54) argues that "The long-run behavior of such a system [Class A] will depend only on the transition matrix [of the embedded Markov chain]...". As can be seen from equation (1.1), this is only true if $\mu_i = \mu$ for every i.

For a proof that a mixture of exponential distributions is a DFR distributuion see Barlow and Proschan (1975, p. 103, Theorem 4.7(a)). This point was also mentioned by Ginsberg (1971, pp. 253-4).

4) $F_{ij}(t) = F_i(t)$ for every i,j, where the F_i's are arithmetic and $F_i(\infty) < 1$ (i.e. the F_i's are defective distributions). This case corresponds to Mayer's "Absorbing State Model" (Stewman 1976, pp. 218-9).

B. In the following cases ${}_zQ(t)$ and $\nu_z(i)$ do depend on z, corresponding to models with heterogeneous populations.

5) a) ${}_zF_{ij}(t) = F(t) = \begin{cases} 0 & t < 1 \\ 1 & t \geq 1 \end{cases}$ for every z, i, j.

The processes $\{X_z(t):t\geq 0\}$ are discrete-time Markov chains. The process $\{X(t):t\geq 0\}$ corresponds to an Extended Mover-Stayer Model that postulates population heterogeneity with respect to transition probabilities (McFarland 1970, Morrison et al. 1971, Spilerman 1972a, Bartholomew 1973, pp. 34-7, Singer and Spilerman 1974, pp. 375-5, Example 2).(*) A "promotion" model due to Wise (1975) is, essentially, a special subcase with ${}_zP_{ii} = p_z$ and ${}_zP_{i,i+1} = 1 - p_z$ for every i.

b) A continuous-time version of 5a: ${}_zF_{ij}(t) = F(t) = 1-e^{-\bar{\lambda}t}$, $t \geq 0$ for every z, i, j. For this model we get:

(*) For a related model see Lazarsfeld and Henry (1968, Section 9.3).

$$ {}_zP(t) = \sum_{r=0}^{\infty} \frac{e^{-\bar{\lambda}t}(\bar{\lambda}t)^r}{r!} {}_zP^r = e^{\bar{\lambda}t({}_zP-I)} , $$

and thus

$$ P(t) = Ee^{\bar{\lambda}t({}_ZP-I)} \qquad (1.4) $$

6) ${}_zP = P$ for every z, $A = R_+$ and ${}_zF_{ij}(t) = 1 - e^{-zt}$, $t \geq 0$ for every i, j.

This case corresponds to an Extended Mover-Stayer Model that postulates population heterogeneity with respect to transition rate. Such models were considered by Spilerman (1972b), Bartholomew (1973, pp. 46-54), and Singer and Spilerman (1974, pp. 375-9, Example 3), although they also assumed that $\nu_z(i) = \nu(i)$ for every z.

For this model we have

$$ {}_zP(t) = \sum_{r=0}^{\infty} \frac{e^{-zt}(zt)^r}{r!} P^r = e^{zt(P-I)} , $$

and thus

$$ P(t) = Ee^{Zt(P-I)} . \qquad (1.5) $$

Note that this model satisfies the assumptions under which (1.3) was obtained, so its limiting behavior coincides with that of its embedded Markov chain. This was also proved specifically for Model 6 by Spilerman (1972b, Appendix A). [For a related result see Bartholomew (1973, p. 52)].

After addressing mover-stayer models like Model 6, Bartholomew (1973, p. 54) concludes by saying "... the general semi-Markov model [i.e. Class A] ... includes them all as special cases". This classification was repeated by Singer and Spilerman (1974, p. 377) and Stewman (1976, Table 3), and it seems that they identified

Model 6 with Model 3b (with corresponding parameters). These models are, however, distinct. In Model 6, Z is chosen at time 0 and its realization is adhered to throughout, while under Model 3b, Y is rechosen after each transition. Under Model 6 "... the length of stay between two successive moves ... are dependent because people with high propensities to move are likely to have two short invervals and people with low propensities to move are likely to have long ones" (Ginsberg 1971, p. 254).(*) The time intervals between transitions are independent of each other in Model 3b, as well as in any semi-Markov process for which $F_{ij} = F$ for every i, j.(**)

7) $\mathcal{A} = R^K_+$, ${}_zP = P$ for each $z\varepsilon\mathcal{A}$, and ${}_zF_{ij}(t) = 1 - e^{-z_i t}$ for each $t \geq 0$ and $i, j\varepsilon\mathcal{K}$ where $z = (z_1, z_2,\ldots,z_K)$. This is a generalization of Model 6 (Singer and Spilerman 1974, pp. 380-5, Example 4) in which K parameters $(z_1,z_2,\ldots,z_K)$ are chosen at time 0; z_i is the transition rate in category i.

For each $z\varepsilon\mathcal{A}$, let $\tilde{z}$ be the diagonal matrix

$$\tilde{z} = \begin{pmatrix} z_1 & & 0 \\ & z_2 & \\ 0 & & \ddots \\ & & & z_K \end{pmatrix}.$$

(*) To put it in other words, the realization of the first duration time T_1 provides some information about Z, which in turn influence the posterior distribution of T_2. More on this in subsection 1.5.2

(**) Though not for the general Class A as claimed by Ginsberg (1971, 1978a, 1978b).

Then ${}_zP(t) = e^{t\tilde{z}(P-I)}$, and $\{X_z(t): t \geq 0\}$ is a Markov process with generator ${}_zA = \tilde{z}(P-I)$. Thus

$$P(t) = Ee^{t\tilde{Z}(P-I)} . \tag{1.6}$$

1.3 A Comparison of Population Heterogeneity Induced through Transition Rates and through Probability Transition Functions

In this section we show that Models 6 and 7, which introduce population heterogeneity through the transition rate are, subject to a bound on the transition rate, a special case of Model 5b, which has constant transition rate $\bar{\lambda}$, but heterogeneous probability transition matrices.

Consider Model 7, and suppose there exists a number $\bar{\lambda}$ such that $Z_i \leq \bar{\lambda}$ with probability 1 for every i.(*) For each z, define

$${}_z\bar{P} = I + \frac{1}{\bar{\lambda}} \cdot {}_zA,$$

where ${}_zA$ is the generator defined in Model 7. Then by Theorem 8.4.31 in Cinlar (1975), a version of Model 5b with probability transition matrices $\{{}_z\bar{P}: z \in A\}$ and rate $\bar{\lambda}$ will have exactly the same transition function ${}_zP(t)$ as does the version of Model 7 with which we started.

So we conclude that extensions of the basic Models 1 and 2 in the

---

(*) If the distribution G of Z does not satisfy this condition, we can pick some (large) λ and truncate G to obtain

$$G_\lambda(z) = \begin{cases} G(z)/G(\lambda,\ldots,\lambda) & \text{if } \; 0 \leq z_i \leq \lambda \quad \text{for every i,} \\ 1 & \text{otherwise.} \end{cases}$$

Since the distributions of times between transitions will still be mixtures of exponentials, they will retain the DFR property. Hence the condition is not too restrictive.

direction of population heterogeneity through the transition rate are, essentially, special cases of the extension that postulates population heterogeneity through transition probabilities. Population heterogeneity is expressed through one parameter in Model 6 and K parameters in Model 7, which are the "rate-difference" cases, while it is expressed through K^2 parameters in Model 5, which postulates "transition probability differences".(*)

1.4 Rate of Mobility in the Various Models

The rate aspect of (intragenerational) mobility which was not accounted for properly by Markov chain models was the tendency to remain in the same category. In fact, the other models were introduced specifically in order to decrease the "rate of mobility" predicted by Markov chains. In order to be able to assess their success in achieving this goal, we have to focus on quantities which are related to the "rate of mobility" in the system. Three quantities come to mind:

a) durations (inter-transition times);

b) number of transitions in a given time inverval; and

c) diagonal elements of P(t) matrices.

(*) Yet another kind of population heterogeneity (through the order of the Markov chain) was postulated in the brand-choice literature (Jones 1973, Givon and Horsky 1978). Consumers are classified into three categories: those of the first type choose a brand independently of their previous choice (a "zero order" Markov chain): the choice of the consumers of the second type is affected only by their most recent choice (a "first order" Markov chain), while the choices of the rest are affected by all their previous choices (linear learning - an "infinite order" Markov chain).

Since the first quantity is most meaningful in models which evolve in continuous-time, we use it to compare Model 2s to Models 3b and 6. Models 3b and 6 may be viewed as two alternative methods of introducing an additional stochastic component to Model 2s. A comparison between these models in terms of the second quantity can then be deduced using some monotonicity properties.

Historically, however, the direct motivation for introducing the more complex models was that Model 1 underestimated the third quantity. It is thus of interest to check whether ignoring population heterogeneity (i.e. using Model 1 when the system is actually Class B type) results in systematic underprediction of diagonal elements.

1.4.1 Inter-Transition Times and Number of Transitions Per Unit Time

In this section we compare Models 2s, 3b, and 6. Assume that the three models have the same probability transition matrix P. Recall that in Model 2s there is a constant transition rate $\bar{\lambda}$; in Model 3b the transition rate Y is chosen anew at each transition from a distribution L; and in Model 6 the transition rate Z is chosen initially from a distribution G. Assume for the purposes of comparing the three models that $G \equiv L$ (we will use G to denote the common distribution, even in the context of Model 3b), and that $EZ = \bar{\lambda}$ (= EY). Let $T^{(3b)}$, $T^{(6)}$ and $T^{(2s)}$ be random inter-transitions times for these models. Then

$$\Pr(t^{(3b)} > t) = \Pr(T^{(6)} > t) = Ee^{-Zt}, \quad t \geq 0, \tag{1.7}$$

while

$$\Pr(T^{(2s)} > t) = e^{-\bar{\lambda}t} = e^{-tEZ}, \quad t \geq 0. \tag{1.8}$$

Since e^{-xt} is a convex function in x, we have by Jensen's inequality that

$$Ee^{-Zt} \geq e^{-tEZ^{(*)}} . \qquad (1.9)$$

Consequently

$$Pr(T^{(3b)} > t) = Pr(T^{(6)} > t) \geq Pr(T^{(2s)} > t), \; t \geq 0. \qquad (1.10)$$

When two random variables X and Y have the same distribution we write $X \overset{st}{=} Y$ and when $Pr(X \geq x) \geq Pr(Y > x)$ for every x we write $X \overset{st}{\geq} Y$ (see, for example, Brumelle and Vickson 1975). Using this notation, (1.10) can be rewritten as $T^{(3b)} \overset{st}{=} T^{(6)} \overset{st}{\geq} T^{(2s)}$.

Consider now the number of transitions in the time interval (0,t] predicted by these three models.(**) Let $N_{2s}(t)$ be the number of transitions in (0,t] for Model 2s. Since $\{N_{2s}(t):t \geq 0\}$ is a Poisson process, its renewal function $m_{2s}(t) \equiv EN_{2s}(t)$ equals $\bar{\lambda}t$. $\{{}_zN_6(t):t \geq 0\}$ is also a Poisson process, so ${}_zm_6(t) = zt$; unconditioning we get $m_6(t) = \bar{\lambda}t$. Hence Models 2s and 6 (with corresponding parameters) pre-

---

(*) F of Model 3b is DFR with mean

$$ET^{(3b)} = E(E(T^{(3b)}|Y)) = E(\tfrac{1}{Y}) = E(\tfrac{1}{Z}) .$$

Hence, in combination with the upper-bound on the survival distribution provided by Barlow and Proschan (1975, p. 116, Theorem 6.10), we get

$$e^{-tEZ} \leq Pr(T^{(3b)} > t) \leq \begin{cases} e^{-t/E(1/Z)} & \text{for } t \leq E(1/Z) \\ E(1/Z)/et & \text{for } t \geq E(1/Z) . \end{cases}$$

(**) If $P_{ii} \neq 0$ for some i, some of the transitions will not involve a real category change. But since P is the same for all three models, the proportion of transition of this type will be the same in all of them, so we can use the total number of transitions as means of comparison.

For any semi-Markov process, there exists some semi-Markov process with $P_{ii} = 0$ for every i which has the same distribution of sample paths. Moreover, a Markov process retains its Markovian property under such transformation.

dict the same expected number of transitions.

Let us now compare $N_{2s}(t)$ with $N_{3b}(t)$. By definition

$$N_{2s}(t) = \sup \{n: \sum_{i=1}^{n} T_i^{(2s)} \leq t\}$$

and

$$N_{3b}(t) = \sup \{n: \sum_{i=1}^{n} T_i^{(3b)} \leq t\}.$$

The N(t)'s are thus decreasing functions of the corresponding T_i's. Since by (1.10) $T_i^{(3b)} \overset{st}{\geq} T_i^{(2s)}$, it follows from the independence of the T_i's that

$$N_{3b}(t) \overset{st}{\leq} N_{2s}(t), \quad t \geq 0. \tag{1.11}$$

Hence

$$m_{3b}(t) \leq m_6(t) = m_{2s}(t), \quad t \geq 0.$$

So we conclude that, compared to Model 2s, Model 3b reduces the rate of mobility as measured by both durations and expected number of transitions per unit time, while Model 6 does so with respect to durations only.

1.4.2 Diagonal Elements and Eigenvalues

It has been already shown by a counterexample (Bartholomew 1973, p. 37) that Model 1 does not necessarily underpredict the diagonal elements of the P(t)'s of a process that actually evolves according to Model 5a.(*) For a process that evolves according to (the more speci-

(*) Bartholomew provides a sufficient condition (reversibility of the processes $\{X_z(t):t\geq 0\}$, for each z) under which underprediction will occur.

alized) Model 6, previous examples (Spilerman 1972b, Bartholomew 1973, pp. 48-9) did exhibit underprediction of diagonal elements by Model 1, but no general proof is available.

We shall prove that if the eigenvalues of P are real, the sum of the diagonal elements generated by Model 6 will indeed be larger than the one generated by Model 1. In other words, under this hypothesis, the expected total number of individuals who at any future time are in the state they started from is larger in Model 6 than would have been predicted by a Markov chain model. We shall restrict ourselves to the case in which $\nu_z(i) = \nu(i)$ for every z.

Theorem 1.1

In Model 6, if all the eigenvalues of P are real, then

$$\text{trace } P(k) \geq \text{trace } P(1)^k \quad k = 1,2,3, \ldots \tag{1.12}$$

Proof:

First, note that

$$\text{trace } P(k) = \text{trace } E[e^{Z(P-I)}]^k \text{ and that}$$
$$\text{trace } P(1)^k = \text{trace } [Ee^{Z(P-I)}]^k .$$

If the eigenvalues of P are real, so are those of P-I. Denote the eigenvalues of P-I by $\lambda_1, \ldots, \lambda_K$. The eigenvalues of $e^{zk(P-I)}$ are then (Cinlar 1975, Appendix) $e^{zk\lambda_1}, \ldots, e^{zk\lambda_K}$, and it can be shown(*) that

(*) Suppose that for every x, ${}_x\lambda$ is an eigenvalue of the generator ${}_xA$, i.e. there exists a vector V such that $({}_x\lambda)V = V({}_xA)$. Then taking expectations of both sides (with respect to the distribution of X) we get $(E_X\lambda)V = VE({}_XA)$. Hence $E({}_X\lambda)$ is an eigenvalue of $E({}_XA)$.

those of $Ee^{Zk(P-I)}$ are $Ee^{ZK\lambda_1}, \ldots, Ee^{Zk\lambda_K}$. Since a matrix's trace equals the sum of its eigenvalues, we have

$$\text{trace } P(k) = \sum_{j=1}^{K} E(e^{Z\lambda_j})^k$$

and

$$\text{trace } P(1)^k = \sum_{j=1}^{K} (Ee^{Z\lambda_j})^k.$$

Let $g(z) = (e^{z\lambda_j})^k$. Then for $k > 1$, g is a convex function of z. Hence by Jensen's inequality

$$Eg(Z) \geq gE(Z);$$

i.e.

$$E(e^{Z\lambda_j})^k \geq (Ee^{Z\lambda_j})^k \qquad j = 1, \ldots, K. \tag{1.13}$$

Summing over j completes the proof.

Remarks:

1) Since 0 is always an eigenvalue of (P-I), and since the sum of the eigenvalues, being equal to the trace, is always real, the other eigenvalue in the two-categories case is also real. Hency for two-categories systems the assertion of the Theorem always holds.

2) Inequalities (1.13) are stronger than the assertion of the Theorem. They imply that each eigenvalue of a k-step transition matrix is larger than the corresponding one of $P(1)^k$.

We now provide an example for which the relation (1.12) does not hold. This example, of course, has complex eigenvalues and shows the necessity of some restriction on P such as the assumption of real eigenvalues in the Theorem.

Example

Let

$$P = \begin{bmatrix} .5 & .0 & .5 \\ .3 & .7 & .0 \\ .0 & .4 & .6 \end{bmatrix}$$

and let

$$Z = \begin{Bmatrix} 1 \text{ with probability } .9 \\ 9 \text{ with probability } .1 \end{Bmatrix} .$$

We now wish to calculate the traces of the matrices P(k) of Model 6 and of the corresponding Model 1. Since we shall do so by summing eigenvalues, it should be noted that complex eigenvalues appear in conjugate pairs. Let $\lambda = a + bi$ be an eigenvalue of (P-I), and let $\bar{\lambda}$ be its conjugate. Now,

$$\begin{aligned} e^{\lambda kz} + e^{\bar{\lambda} kz} &= e^{kz(a+bi)} + e^{kz(a-bi)} \\ &= e^{akz}(\cos bkz + i\sin bkz) \\ &+ e^{akz}[\cos(-bkz) + i\sin(bkz)] \\ &= 2e^{akz}\cos bkz. \end{aligned}$$

Also,

$$\begin{aligned} (Ee^{\lambda Z})^k + (Ee^{\bar{\lambda} Z})^k &= 2\mathrm{Re}(Ee^{\lambda Z})^k \\ &= 2\mathrm{Re}\{E[e^{aZ}(\cos bZ + i\sin bZ)]\}^k \end{aligned}$$

In our example the eigenvalue of (P-I) are 0, $-.6 + .3317i$ and $-.6 - .3317i$. Hence

$$\text{trace } P(k) = 1 + 2E(e^{-.6kZ}\cos .3317kZ)$$

and

$$\text{trace } P(1)^k = 1 + 2\text{Re}\{E[e^{-.6Z}(\cos .3317Z + i\sin .3317Z)]\}^k.$$

Using double-precision, we obtained the following values:

k	trace P(k)	trace $P(1)^k$
1	1.933120	1.933120
2	1.427165	1.383567
3	1.161984	1.130630
4	1.039448	1.028467
5	0.992150	0.994745
6	0.979972	0.988163
7	0.981579	0.990234
8	0.986916	0.993771
9	0.991969	0.996566
10	0.995607	0.998313

We see that there are some cases (e.g. k=6) where trace $P(k) <$ trace $P(1)^k$. This, of course, implies that at least one diagonal element of $P(6)$ is smaller than the corresponding one in $P(1)^6$. So some restrictions on P have to be imposed if (1.12) is to hold.(*)

1.5 Compatibility of Parametric Families of Distributions with Observed Properties of Career Patterns

In this section we mention several observations made by March and March (1977) about career patterns. We then check whether some specific parametric forms of the models, which were suggested in the literature, are compatible with these properties.

(*) It is also evident that it is not sufficient (as conjectured by Blumen et al. 1955) that the diagonal elements of P will be large.

1.5.1 Decreasing Rate of Inertia Accumulation

Recent mobility studies indicated that, in addition to the "cumulative-inertia" phonomenon, the distributions of time between transitions exhibit some additional properties. Tuma (1976) and March and March (1977) argue that "...although selection, retention, adaptation and depletion processes characteristically result in changes in average durability, the rate of change declines over the duration of the match" (March and March 1977, p. 380). Translated to our vocabulary, this means that although inertia is being accumulated, the rate of accumulation is decreasing. Assuming (for simplicity) that F is twice differentiable, we can express it as:

$$\frac{\partial}{\partial t}\left[\frac{\bar{F}(t+x)}{\bar{F}(t)}\right] \geq 0 \qquad \text{for every } x \tag{DFR}$$

$$\frac{\partial^2}{\partial t^2}\left[\frac{\bar{F}(t+x)}{\bar{F}(t)}\right] < 0 \qquad \text{for every } x, \tag{1.14}$$

where $\bar{F} = 1 - F$.

Denote by T a random time between successive transitions, and let $\mu = ET$. Then it is known (Barlow and Proschan 1975) that DFR implies a property called "New Worse than Used in Expectation", which can be stated as

$$E(T - t \mid T > t) > \mu \; .$$

A popular DFR distribution in the mobility literature (Silcock 1954, Spilerman 1972b) is a mixture of exponential distributions with gamma mixing distribution. However, Morrison (1978) proved that for such distribution

$$E(T - t \mid T > t) = at + b \qquad a > 0.$$

Consequently,

$$\frac{\partial}{\partial t} E(T - t|T > t) = a > 0$$

$$\frac{\partial^2}{\partial t^2} E(T - t|T > t) = 0 \quad \text{for every } t.$$

Hence this distribution does not satisfy (1.14).

The Weibul distribution

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}} \quad t \geq 0 \quad \lambda > 0$$

with $0 < \alpha \leq 1$, and the Gompertz distribution

$$F(t) = 1 - e^{-\lambda(e^{\beta t}-1)/\beta} \quad t > 0$$

with $\beta < 0$, have been also suggested for modelling mobility (Ginsberg 1978a). These distributions are DFR, but in both cases as $t \to \infty$ the "hazard rate" approaches zero (Ginsberg 1971, p. 253, Barlow and Proschan 1975, p. 73). So (1.14) will not be satisfied here either.

A DFR distribution which does satisfy (1.14) is a gamma, which has density

$$f(t) = \frac{\lambda(\lambda t)^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \qquad t \geq 0, \lambda > 0,$$

if $0 < \alpha \leq 1$ (Ginsberg 1971, p. 253, Barlow and Proschan 1975, pp. 73-5).

1.5.2 Heterogeneity of Subpopulations Selected by Their Durations

Referring to career patterns March and March (1977, p. 380) claim that selection and retention rules commonly used by employers "... reduce the variance [and thus] reduce over the duration of the matches the heterogeneity of populations". In our notation, this statement becomes

$\mathrm{Var}\ (Z|T > t)$ is decreasing in t(*), where T denotes a sojourn time.

Not every joint distribution of Z and T will have this property, however. Suppose that a large majority of the population have short sojourn times with low variability, making the overall variance of Z small. Now, if a realization of sojourn time turns out to be long, the individual is likely to belong to the less mobile minority, which may have a high variability among its members.

In the particular case of Model 6

$$\Pr(T > t) = Ee^{-Zt};$$

so we obtain that

$$\mathrm{Var}(Z|T > t) = \frac{EZ^2e^{-Zt}}{Ee^{-Zt}} - \left(\frac{EZe^{-Zt}}{Ee^{-Zt}}\right)^2.$$

Hence

$$\frac{\partial}{\partial t}\mathrm{Var}(Z|T > t) = \frac{1}{(Ee^{-ZT})^3}\ [3Ee^{-ZT}\ EZe^{-Zt}\ EZ^2e^{-Zt}$$

$$- (Ee^{-Zt})^2\ EZ^3e^{-Zt} - 2(EZe^{-Zt})^3].$$

(*) A related property is

$$\mathrm{Var}\ (Z|T_1 = t) \leq \mathrm{Var}\ Z \text{ for every } t \geq 0,$$

where T_1 is the sojourn time in the initial category. These kinds of questions, comparing properties of posterior distributions to those of prior ones, are common in Bayesian analysis, and the answers depend on the joint distribution of Z and T.

Upon replacing variance by <u>entropy</u> as a measure of dispersion, on the other hand, similar properties become valid for <u>any</u> two random variables (cf. Khinchin 1957, pp. 2-9).

For the desired property to hold, the distribution G of Z has to be such that

$$3Ee^{-Zt}\ EZe^{-Zt}\ EZ^2e^{-Zt} - (Ee^{-Zt})^2\ EZ^3e^{-Zt} - 2(EZe^{-Zt})^3 < 0.$$

Consider, for example, the distribution

$$Z = \begin{cases} X & \text{with probability } \beta \\ M & \text{with probability } 1-\beta \end{cases},$$

where X has an exponential distribution with rate 1 and M is a constant. Here we get

$$EZ^m e^{-Zt} = \frac{(1-\beta)\ M^m(t+1)^{m+1}\ e^{-Mt} + \beta}{(t+1)^4} \qquad \text{for } m \geq 1,$$

and hence

$$\frac{\partial}{\partial t}\text{Var}(Z|T>t) = \frac{e^{-2Mt}\beta(1-\beta)(t+1)[M^3(t+1)^3-3M^2(t+1)^2+3M(t+1)-1][(1-\beta)(t+1)-\beta e^{Mt}]}{(t+1)^6(Ee^{-Zt})^3}.$$

Thus

$$\frac{\partial}{\partial t}\text{Var}(Z|T>t) < 0$$

if and only if

$$\beta > \frac{t+1}{t+1+e^{Mt}}.$$

So for a given M, the sign of the derivative may depend on t.

However, the commonly used mixture of exponentials with gamma mixing distribution has the desired property. Let the density of Z be

$$g_Z(z) = z^{\alpha-1}\ e^{-z/\Gamma(\alpha)} \qquad z \geq 0\ ,\ \alpha > 0$$

Then

$$EZ^m e^{-Zt} = \Gamma(\alpha+m)/\Gamma(\alpha)(t+1)^{\alpha+m},$$

so

$$\mathrm{Var}(Z|T > t) = \frac{\Gamma(\alpha)\Gamma(\alpha+2) - [\Gamma(\alpha+1)]^2}{[\Gamma(\alpha)]^2\ (t+1)^2} = \frac{\alpha}{(t+1)^2},$$

which is clearly decreasing in t.

CHAPTER II

A STOCHASTIC MODEL ALLOWING INTERACTION AMONG INDIVIDUALS

Many systems, arising in a variety of contexts, consist of a number of individuals moving among various categories or states. Typically, the individuals in such systems do not move independently of each other, but instead interact in some way. Several authors have pointed out the need for explicitly modeling such interactions. Social mobility (Matras 1967) and promotion chances in an organization (White 1970) are affected by the existence of opportunities, which are created (among other reasons) by other people's movements. Popularity affects political affiliation (Holley and Liggett 1975),(*) consumers' brand switching (Smallwood 1975, Smallwood and Conlisk 1979), and modal choice (Krishnan and Beckman 1979). The effect of crowding on internal migration was modeled by Cordey-Hayes and Gleave (1974). Models of epidemics and diffusion of rumours (e.g. Bartholomew 1973, Ch. 9, 10) incorporate effects of human contacts.

Conlisk (1976) introduced a generalization of Markov chains in which an individual's next category depends on his current category and on the distribution of the population among the categories. Models of this type, which combine "push" flows with "pull" flows (Bartholomew 1973, p. 26) were also suggested by Matras (1967) and Smallwood (1975) and are used by Smallwood and Conlisk (1979). This type of model appears to be appropriate for systems such as those mentioned in the first paragraph.

(*) This model originated from statistical mechanics (see also Spitzer 1970).

Conlisk's model assumes that the probability Q_{ij} of moving from category i to category j is a function of how the other individuals are distributed among the categories. Thus instead of having one probability transition matrix, this model has a probability transition matrix Q(y) for each vector y (called a profile), whose components are fractions of the population in each category. However, Conlisk does not analyze the above model, which we call the <u>finite population model</u>. In fact, he does not define it unambigously. We thus start section 1 of this chapter by carefully defining a finite population model. In structuring our model, we have attempted to be consistent with the motivation and examples which are so nicely developed in Conlisk (1976) and Smallwood and Conlisk (1979). We then investigate some of the model's properties and computational aspects.

Conlisk argues that for a large population the profiles at time t and time t+1 (denoted by row vectors y_t and y_{t+1}, respectively) should be related by

$$y_{t+1} = y_t \, Q(y_t) \ . \tag{2.1}$$ (*)

In this model, which we call the <u>infinite population model</u>, the process $y_0, y_1, y_2, \ldots$ is deterministic once y_0 is specified (in the finite population model the profiles will be truly stochastic). Conlisk (1976, p. 158) states that "[y_t] is stochastic [in the finite population model] and the equation [(2.1)] must be viewed as approximate; but, for a large population the approximation error is negligible". The infinite popula-

(*) Matras (1967) also suggests this relation. If the functional Q happens to be one-to-one, (2.1) becomes a special case of a demographic model due to Cohen (1976).

tion model (2.1) is also used in Smallwood and Conlisk (1978) with a similar comment about the approximation. However, neither paper substantiates the claim that the approximation error is negligible. Now, since the infinite population model has obvious computational advantages over the corresponding finite population one, the later sections of this chapter are devoted to examining the validity of (2.1) as an approximation of the finite population model.

In Section 2, we investigate the degree to which the infinite population model approximates profiles in the finite population model. Loosely speaking, a real valued function (with some continuity assumptions) defined on the finite population profile process converges to the same function defined on the infinite population process. Over a finite horizon, the continuity assumptions on the functions are not very restrictive. However, over an infinite horizon they are bothersome. In particular, one must be very careful about making inferences about the equilibrium or steady state behavior of the finite population model from the corresponding behavior of the infinite population model. For example, the lack of a globally stable fixed point for the map $y \to yQ(y)$ in the infinite population model does not imply that the finite population profile process lacks asymptotic stationarity (for definitions, see subsection 2.2.2). Put differently, in general

$$\lim_{N\to\infty} \lim_{t\to\infty} \Pr(Y_t^N \varepsilon B) \neq \lim_{t\to\infty} \lim_{N\to\infty} \Pr(Y_t^N \varepsilon B),$$

where Y_t^N is the profile at time t in a model with population size N, and B is some set. However, we do show that if the map $y \to yQ(y)$ is globally stable then the infinite population model does approximate the finite population model even over infinite time horizon.

Section 2.3 develops the mathematical theory needed for the discussion in Section 2.2. The basic result is that if the probability transition functions of a sequence of Markov processes converge weakly and uniformly to some probability transition function p, then the sequence of Markov processes converges weakly to a Markov process with probability transition function p. Although the theory developed in Section 2.3 is used in Section 2.2, it is placed later so as not to interupt the discussion of the model.

2.1 The Finite Population Model and its Properties

2.1.1 Formulation

Let $K = \{1,2,\ldots,K\}$ be a set of categories (e.g. social or occupational classes, geographical regions, brands, etc.). Let $S = \{y=(y_1,y_2\ldots,y_K): \sum_{i=1}^{K} y_i = 1 \text{ and } y_j \geq 0 \ (j=1,2,\ldots,K)\}$. An element of $y\varepsilon S$ is called a population profile and y_j is the fraction of the population in category j. Let N be the population size. For a particular population size, only certain profiles can occur and these are included in S_N, which is the subset of S consisting of profiles y such that each component of Ny is integer.(*)

We will now define a stochastic process $\{X_t^i: t=0,1,2,\ldots\}$ for each

(*) The number of profiles in S_N is given by $\binom{N+K-1}{N}$.

of the N individuals, which are named $i=1,2,\ldots,N$. The random variable X_t^i is the category of individual i at time period t. The profile at time t is defined to be the random vector Y_t^N whose k-th component is

$$Y_t^N(k) := \frac{1}{N}\sum_{i=1}^{N} I_{[X_t^i=k]}{}^{(*)} \tag{2.2}$$

for each category k.

The first assumption is that the future of the individuals depends only on their current categories and not on how they arrive at their categories. More formally, we assume that the stochastic process $\{(X_t^1,X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is a vector valued Markov chain; that is

$$\begin{aligned} &\Pr[X_{t+1}^1=k_1,\ X_{t+1}^2=k_2,\ldots,X_{t+1}^N=k_N \mid (X_n^1,X_n^2,\ldots,X_n^N),\ n=0,1,2,\ldots,t] \\ &= \Pr[X_{t+1}^1=k_1,\ X_{t+1}^2=k_2,\ldots,X_{t+1}^N=k_N \mid X_t^1,X_t^2,\ldots,X_t^N] \end{aligned} \tag{2.3}$$

for each $t=0,1,2,\ldots$ and categories $k_i \varepsilon \mathcal{K}$.(**)

The second assumption is that each individual's decision as to his next category depends only on his current state and the current profile, and is taken independently of the other individuals' decisions. This assumption means that the model does not explicitly allow for leadership of influential individuals, since any individual's decision does not depend on the particular category of any other particular individual, but only on the distribution of the other individuals among the various

---

(*) I_A is the indicator function of the event A; i.e. $I_A = 1$ if A occurs, and $I_A = 0$ if A does not occur.

(**) For a general discussion of vector-valued Markov processes see Moyal (1962).

categories.(*) More formally, define $Q_{jk}(y)$ to be the probability that an individual moves from category j to category k if the current profile is y. That is, for each i=1,2,...,N and $y \in S_N$,

$$Q_{jk}(y) := \Pr[X^i_{t+1} = k \mid X^i_t = j,\ Y_t = y] \text{ for } t=0,1,2,... \tag{2.4}$$

Then the assumption can be formulated as

$$\Pr[X^1_{t+1}=k_1,\ X^2_{t+1}=k_2,...,X^N_{t+1}=k_N \mid X^1_t=j_1,\ X^2_t=j_2,...,X^N_t=j_N]$$

$$= \prod_{i=1}^{N} Q_{j_i k_i}(y) \quad \text{for } t=0,1,2,\text{...and for any categories } j_i \text{ and } k_i,$$

where $y = Y_t$ is defined in terms of X_t by (2.1).(**)

So far the formulation of the model. A continuous time version of the special case K=2 was studied by Holland and Leinhardt (1977) in the context of social networks.

2.1.2 Example (Smallwood and Conlisk 1979)

Consider a product whose quality can be tested only by using it, and not by simple inspection or prior information. The product lasts one period (e.g. automobile insurance policy), and each of the N consumers buys one unit of the product each period. There are K equally-priced brands, the quality of which is defined solely in terms of their "break-

(*) However, the presence of leaders can be accommodated in this model by creating an extra set of categories for each leader.

(**) The special case, in which Q is a constant function of y (i.e. no interaction among individual), corresponds to the "classical" model in which the individuals follow independent and identically distributed Markov chains.

down" probabilities $b_1,\ldots,b_K$. Let us assume that for the product of interest breakdown merely means moderately unsatisfactory product performance. If the product does not breakdown, the consumer repurchases the same brand. If it does breakdown, he chooses his next period brand randomly among all brands, with probabilities proportional to the current market shares $[Y_t^N(i)]^\sigma$ $i=1,\ldots,K$, where σ is a non-negative parameter which may be interpreted as the degree of confidence in market popularity implicit in consumer behavior. Hence we have a special case of the finite population model in which Q has the following form

$$Q_{ij}(y) = \begin{cases} 1 - b_i + b_i\left(y_i^\sigma \Big/ \sum_{k=1}^{K} y_k^\sigma\right) & j = i \\ b_i\left(y_j^\sigma \Big/ \sum_{k=1}^{K} y_K^\sigma\right) & j \neq i \end{cases}$$

2.1.3 <u>Properties</u>

I. The distribution of the stochastic process $\{(X_t^1, X_t^2,\ldots,X_t^N): t=0,1,2,\ldots\}$ is defined recursively by (2.3) and (2.5) once the distribution of $(X_0^1, X_0^2,\ldots,X_0^N)$ is specified. The parameters of this time homogeneous, vector valued Markov chain are the population size N, the matrix valued function $Q(\cdot)$, and the distribution of $(X_0^1, X_0^2,\ldots,X_0^N)$. The profile process $\{Y_t^N: t=0,1,2,\ldots\}$ is defined by (2.2) and is also a time homogeneous vector valued Markov chain.

It follows from (2.5) that if Q is continuous in y (an assumption which we shall make throughout) a small change in Q will not cause a large change in the population-level transition probabilities. More specifically, (2.5) implies that if

$$\sup_{\substack{i,j\varepsilon K \\ y\varepsilon S}} | Q_{ij}(y) - Q'_{ij}(y) | < \delta$$

then

$$\sup_{\substack{j_1,\ldots,j_N\varepsilon K \\ k_1,\ldots,k_N\varepsilon K}} \Big| \Pr[X_{t+1}^{1}=k_1,\ldots,X_{t+1}^{N}=k_N \,|\, X_t^1=j_1,\ldots,X_t^N=j_N] - \Pr[X'^{1}_{t+1}=k_1,\ldots,X'^{N}_{t+1}=k_N \,|\, X'^{1}_t=j_1,\ldots,X'^{N}_t=j_N] \Big| \leq \delta^N.$$

Since, in particular, steady-state probabilities vary continuously with transition probabilities,(*) the above implies that the model's equilibrium will be "stable" under small perturbations of Q.

Due to the model's symmetry in individuals, the following statements are equivalent for given N and Q(•):

a. $\{(X_t^1, X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is irreducible and aperiodic.

b. $\{Y_t^N :t=0,1,2,\ldots\}$ is irreducible and aperiodic.

c. There exists t such that $\Pr(X_t^i=k|X_0^i=j,Y_0^N=y) > 0$ for every $y\varepsilon S_N$ and categories j and k.

These three statements will be used interchangeably in the next subsection, where sufficient conditions for the above will be given.

(*) The steady-state probabilities are a solution of a system of linear equations, which is known to be continuous in the coefficients.

II. One would like to be able to interpret $Y_t^N(k)$ as the probability that an individual is in category k; i.e.,

$$Y_t^N(k) = \Pr(X_t^i = k | Y_t^N). \tag{2.6}$$

This is clearly valid if individual i is chosen at random (with the individuals equally likely to be chosen) at time t. Equivalently, one can number the individuals at time 0 at random (with each of the N! permutations equally likely) after they have been assigned to their initial categories. Then (2.6) holds for each i = 1,2,...,N.

Relation (2.6) has the following intuitive interpretation. In deciding, say, on our political affiliation at time t, our choice will be stochastically the same (given the population profile Y_t^N) if we choose party k with probability $Y_t^N(k)$ or if we select a person at random and switch to the party which he currently supports.

Taking the expectation of each side of (2.6) gives

$$EY_t^N(k) = \Pr(X_t^i=k). \tag{2.7}$$

We shall make use of this equality later on.

III. Although $Q(\cdot)$ has been defined on all of S, some of its values will have no impact on the behavior of the finite population model regardless of population size.(*) In particular, if for some $k \in K$ we let $A_k = \{y \in S : y(k)=0\}$ then for any $y \in A_k$ the values of $Q_{k\cdot}(y)$ will have no effect on the model's behavior. Hence any condition imposed on such values is not at all restrictive. This fact should be kept in mind when

(*) They will have no impact on the infinite population model either.

considering the conditions imposed on $Q(\cdot)$ in the next subsection.

2.1.4 Conditions for Irreducibility and Aperiodicity

Without some restrictions, the properties of the above mentioned processes can be significantly dependent on the population size N. We shall thus look for conditions under which the process $\{(X_t^1, X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is irreducible and aperiodic for every N.

We shall write $A > 0$ (where A and 0 are matrices of the same order, the latter of which consists of zeroes only) if a_{ij} is (strictly) positive for every i and j.

The basic result is the following:

Theorem 2.1

If there exists an integer n such that for any $y_m \varepsilon S$ $m = 0,1,\ldots,n$ $\prod_{m=0}^{n} Q(y_m) > 0$, then $\{(X_t^1, X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is irreducible and aperiodic for any population size N.

Proof:

The condition

$$\prod_{m=0}^{n} Q(y_m) > 0 \text{ for any } y_m \varepsilon S \quad m=0,1,\ldots,n$$

is equivalent to

$$\sum_{i_1=1}^{K} \cdots \sum_{i_n=1}^{K} \prod_{m=0}^{n} Q_{i_m, i_{m+1}}(y_m) > 0 \quad \text{for any } y_m \varepsilon S \quad m = 0,1,\ldots,n \text{ for}$$

any $i_0, i_{t+k} \varepsilon K$,

which implies that for any $y_m \varepsilon S$ $m=0,1,\ldots,n$ and any $i_0, i_{t+1} \varepsilon K$ there exist $i_1,\ldots,i_{n-1} \varepsilon K$ such that

$$\prod_{m=0}^{n} Q_{i_m, i_{m+1}}(y_m) > 0. \tag{2.8}$$

Now, since $\{(X_t^1, X_t^2,\ldots,X_t^N): t=0,1,2,\ldots\}$ is a Markov chain, for any categories $i_0{}^1,\ldots,i_0{}^N$ and $i_{t+1}{}^1,\ldots,i_{t+1}{}^N$

$$\Pr(X_{n+1}{}^1 = i_{n+1}{}^1,\ldots,X_{n+1}{}^N = i_{n+1}{}^N | X_0^1 = i_0^1,\ldots,X_0^N = i_0{}^N)$$

$$= \sum_{1\le i_1^1,\ldots,i_1^N \le K} \cdots\cdots\cdots\cdots \sum_{1\le i_n^1,\ldots,i_n^N \le K} \prod_{m=0}^{n} \Pr(X_{m+1}{}^1 = i_{m+1}{}^1,\ldots,X_{m+1}{}^N = i_{m+1}{}^N | X_m^1 = i_m^1,\ldots,X_m^N = i_m^N) .$$

By the model's assumptions the product in the last expression equals

$\prod_{m=0}^{n} \prod_{\ell=1}^{N} Q_{i_m^\ell, i_{m+1}^\ell} [y(X_m^1,\ldots,X_m^N)]$, which in turn equals

$\prod_{\ell=1}^{N} \prod_{m=0}^{n} Q_{i_m^\ell, i_{m+1}^\ell} [y(X_m^1,\ldots,X_m^N)]$. But by (2.8), for each ℓ we can choose

$i_1^\ell,\ldots,i_n^\ell$ such that $\prod_{m=0}^{n} Q_{i_m^\ell, i_{m+1}^\ell} [y(X_m^1,\ldots,X_m^N)] > 0$.

For such $\{i_1^\ell,\ldots,i_n^\ell\}_{\ell=1}^N$

$$\prod_{\ell=1}^{N} \prod_{m=0}^{n} Q_{i_m^\ell, i_{m+1}^\ell} [y(X_m^1,\ldots,X_m^N)] > 0 .$$

Hence $\quad \Pr(X_{n+1}^1 = i_{n+1}{}^1,\ldots,X_{n+1}^N = i_{n+1}{}^N | X_0^1 = i_0^1,\ldots,X_0^N = i_0^N)$

is strictly postive for any categories $i_0^1,\ldots,i_0^N$ and $i_{n+1}^1,\ldots,i_{n+1}^N$. The Markov chain $\{(X_t^1, X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is thus irreducible and aperiodic for every N. ■

Remark: In the non-interactive case $[Q(y) \equiv Q$ for every $y \in S]$ the Theorem reduces to stating that if the individual-level Markov chain $\{X_t^i:t=0,1,2,\ldots\}$ is irreducible and aperiodic then so is the population level chain $\{(X_t^1,\ldots,X_t^N):t=0,1,2,\ldots\}$, which is what one would expect intuitively.

What we would like now is to find some verifiable conditions on $Q(\cdot)$ under which the Theorem's hypothesis will be satisfied.

Define the pattern of Q to be a matrix P such that $P_{ij} = \inf_{y\in S} Q_{ij}(y)$. Thus $P_{ij} > 0$ if and only if $Q_{ij}(y) > 0$ for every $y \in S$. We say that P is regular if $P^n > 0$ for some n. The theory of products of finite non-negative square matrices (Hajnal 1958) implies that if Q has a regular pattern then there exists n such that

$$\prod_{m=0}^{n} Q(y_m) > 0 \text{ for any } y_m \in S,\ m=0,1,\ldots,n.$$

Thus if Q has a regular pattern then the process $\{(X_t^1,X_t^2,\ldots,X_t^N):t=0,1,2,\ldots\}$ is irreducible and aperiodic for any population size N.

The following special case of the Theorem is particularly useful.

Corollary 2.1

If $Q(y) > 0$ for every $y \in S$ then for any N and t $\Pr(Y_{t+1}^N=\hat{y}\,|\,Y_t^N=y) > 0$ for every $y,\hat{y} \in S_N$.

2.1.5 Some Computational Aspects

The finite population model can be simulated in a straightforward manner: at each period t the random variables X_t^i, $i=1,\ldots,N$ will be randomly generated according to the probability mass function $Q_{j\cdot}(y)$ where $j=X_{t-1}^i$ and $y=Y_{t-1}^N$. When the population size becomes very large, however, the above procedure becomes rather costly. Instead, we can make use of the following observation.

Given $Y_t^N=y$ the random variable NY_{t+1}^N can be written as a sum of K independent multinomialy distributed random variables M_k^N, $k=1,2,\ldots,K$ with parameters $(Ny(k);\ Q_{k1}(y),\ldots,Q_{kK}(y))$. Hence the variables NY_{t+1}^N can be generated by generating the multinomials M_k^N, $k=1,2,\ldots,K$.

Now, there are two ways of generating a multinomial. First, by combining all the categories (and parameters) but one we can generate one of its components by generating a binomial variable. Then, updating the population size and the parameters, single out another category from the remaining ones and continue in that fashion (Bishop, Feinberg and Holland 1975, Section 13.4). Second, there are routines which generate multinomials directly.

The computational saving is due to the well known fact that, even for moderately large populations, binomials can be adequately approximated by normal distributions. Moreover, multinomial distributions can be directly approximated by the multivariate normal distribution (Bishop et al. 1975, pp. 469-70).

Hence the following computational scheme seems reasonable. If $Ny(k)$ is small, generate M_k^N by assigning each individual separately (via $Q_{k\cdot}(y)$). If it is large, generate M_k^N by a multivariate-normal approximation. Thus the computational effort required for the simulation is proportional to the number of categories and, essentially, does not depend on the population size.

2.2 Approximation of the Profile Process

Recall from the introduction that Conlisk (1976) introduced the deterministic recursion $y_{t+1} = y_t Q(y_t)$ relating the profile vectors at times t and t+1, and suggested it as a model for the evolution of population profiles for large populations.(*)(**) He also assumed that each component of Q(y), say $Q_{ij}(y)$, is a continuous function of y. We, too, make this assumption for both the finite and infinite population models.

This section investigates the degree to which Conlisk's infinite population model approximates the finite population model. Theorem 2.2 implies that the approximation is "good" over one period, and Theorem 2.3 implies that the approximation is "good" over any finite horizon. Theorem 2.4 implies that approximation of steady state behavior is "good" if the infinite population model has a globally stable fixed point.

Denote the one step probability transition function of the finite-population profile process by

$$p^N(y,B) := \Pr[Y^N_{t+1} \varepsilon B | Y^N_t = y] \text{ for } y \varepsilon S_N \text{ and } B \varepsilon \mathcal{B}(S),$$

(*) It should be noted that in the "classical" non-interactive model (constant Q) this type of recursion is frequently used to approximate the stochastic profile process.

(**) A continous time version of the recursion was considered in Conlisk (1978b).

where $\mathcal{B}(S)$ is the Borel σ-field of S. Its t-step probability transition function is similarly denoted by

$$p_t^N(y,B) := \Pr[Y_{t+s}^N \varepsilon B | Y_s^N = y],$$ and its initial probability measure by

$$\mu^N(B) := \Pr[Y_0^N \varepsilon B] \text{ for } B\varepsilon\mathcal{B}(S).$$

The recursion (1.1) determines the infinite population process $\{Y_t : t=0,1,2,\ldots\}$ once Y_0 is specified. Let Y_0 have probability measure μ defined on $\mathcal{B}(S)$; i.e. $\mu(B) = \Pr[Y_0 \varepsilon B]$, $B\varepsilon\mathcal{B}(S)$. Note that conditioned on $Y_0 = y$, the profile process is deterministic with $Y_{t+1} = Y_t Q(Y_t)$.

It is convenient to introduce some additional notation for the infinite population model. Let $F(y) := F^{(1)}(y) := yQ(y)$ and let $F^{(t+1)}(y) := F(F^{(t)}(y))$ for $y\varepsilon S$. Let $p(y,B) := I_{[F(y)\varepsilon B]}$ and $p_t(y,B) := I_{[F^{(t)}(y)\varepsilon B]}$ for $y\varepsilon S$, $B\varepsilon\mathcal{B}(S)$. Note that p is the (deterministic) probability transition function of the Markov chain $Y_0, Y_1, Y_2, \ldots$. Since Conlisk's model allows profiles to take any value in S, the comparison between our model and Conlisk's is facilitated by extending p^N to a probability transition function on all of S. Define the extension $\hat{p}^N$ by $\hat{p}^N(y,B) := p^N(\hat{y},B)$ where $\hat{y}$ is the point in S_N nearest to y. Any fixed rule can be used to break ties; e.g. if several points in S_N are equally close to y, then choose the lexicographically smallest. This extension is only for technical convenience. Note that it does not alter the definition of the finite population model and $\hat{p}^N(y,B) = p^N(y,B)$ for all $y\varepsilon S_N$. For typographical convenience, the "^" will be suppressed and $p^N(y,B)$ will always refer to the extended probability transition function.

The notion of convergence used in the paper is summarized here. Suppose V is any metric space (e.g. any of the spaces S^t, $t=1,2,\ldots,\infty$) and Z^N and Z are random elements taking values in V with respective probability measures μ^N and μ. If $Ef(Z^N) \to Ef(Z)$ as $N \to \infty$ for each real bounded continuous f on V, then we say that Z^N converges in distribution to Z ($Z^N \xrightarrow{\mathcal{D}} Z$) or that μ^N converges weakly to μ ($\mu^N \xrightarrow{w} \mu$) (Billingsley 1968). Recall that p^N and p are, respectively, probability transition functions for the Markov chains $\{Y_t^N, t=0,1,2,\ldots\}$ and $\{Y_t : t=0,1,2,\ldots\}$ representing the finite and infinite population profile processes. We say that $p^N \xrightarrow{w} p$ uniformly on S if

$$\int f(y,x)p^N(y,dx) \longrightarrow \int f(y,x)p(y,dx)$$

uniformly in y for each real function f continuous on S^2.

2.2.1 Approximation over Finite Time Horizon

Theorem 2.2 $\quad p^N \xrightarrow{w} p$ uniformly on S.

Proof: During some transition, say the t-th, let Z_{ij} be the number of individuals moving from category i to category j. Given that $Y_t^N = y = (y_1,y_2,\ldots,y_K) \varepsilon S_N$, we have that $Z_{1j},\ldots,Z_{Kj}$ are independent for each j, and that Z_{ij} has a binomial distribution characterized by

$$EZ_{ij} = Ny_iQ_{ij}(y), \text{ and}$$

$$\text{Var } Z_{ij} = Ny_iQ_{ij}(y)(1-Q_{ij}(y)) \leqslant Ny_i/4.$$

Since $Y_{t+1}^N(j) = \frac{1}{N} \sum_{i=1}^{K} Z_{ij}$, it follows that

$$E[Y^N_{t+1}(j)|Y^N_t = y] = yQ._j(y)$$

and (2.9)

$$\mathrm{Var}[Y^N_{t+1}(j)|Y^N_t = y] \leqslant 1/4N \quad \text{for } j=1,2,\ldots,K, \text{ (*)}$$

where $Q._j(y)$ is the j-th column of $Q(y)$.

For $y \varepsilon \mathcal{S}$, define $|y| = \sup_i |y_i|$. The function $y \to |y|$ is a norm on $\mathcal{S}$ and any function continuous with respect to Euclidean norm (e.g. $F(y)=yQ(y)$) will be continuous with respect to this norm. Since F is continuous and $\mathcal{S}$ is compact, F is uniformly continuous. Thus there exists a function $\delta(\in)$ such that $|F(y) - F(\hat{y})| < \in/2$ whenever $|y-\hat{y}| < \delta(\in)$. Note that for any $y\varepsilon\mathcal{S}$, the distance $|y-\hat{y}|$ between y and its nearest neighbor $\hat{y}$ in $\mathcal{S}_N$ cannot exceed 1/N. Given $\in > 0$, choose N_0 large enough so that $1/N_0 < \delta(\in)$. Then for each $N \geqslant N_0$ and $y\varepsilon\mathcal{S}$ with nearest neighbor $\hat{y}\varepsilon\mathcal{S}_N$, we have $|y-\hat{y}| < \delta(\in)$, and it follows that

$$\Pr[|Y^N_{t+1} - F(y)| > \in \; |Y^N_t = y]$$

$$= \Pr[|Y^N_{t+1} - F(y)| > \in \; |Y^N_t = \hat{y}]$$

$$\leqslant \Pr[|Y^N_{t+1} - F(\hat{y})| + |F(\hat{y}) - F(y)| > \in |Y^N_t = \hat{y}]$$

$$\leqslant \Pr[|Y^N_{t+1} - F(\hat{y})| > \in/2 \; |Y^N_t = \hat{y}].$$

(*) The argument so far is similar to Smallwood and Conlisk (1979, Footnote 11).

By (2.9), Chebychev's inequality, and the inequality $\Pr(\bigcup_{i=1}^{K} A_i) \leqslant \sum_{i=1}^{K} \Pr(A_i)$,

$$\Pr[|Y_{t+1}^N - F(\hat{y})| > \epsilon/2 | \; Y_t^N = \hat{y}] \leqslant K/\epsilon^2 N. \tag{2.10}$$

Consequently, for each $y \varepsilon S$

$$\Pr[|Y_{t+1}^N - F(y)| > \epsilon | Y_t = y] \to 0 \text{ uniformly in } y \text{ as } N \to \infty.$$

Let $f \varepsilon C(S)$ and $y \varepsilon S$ be arbitrary, but fixed throughout the remainder of the proof. Since f is continuous and S is compact, f is uniformly continuous; hence, given $\epsilon > 0$ there exists $\delta(\epsilon)$ such that $x \varepsilon A(\epsilon) := [x \varepsilon S: \; | \; f(x) - f(yQ(y))| < \epsilon]$ whenever $|x - yQ(y)| < \delta(\epsilon)$. So (2.10) can be rewritten as

$$\Pr[Y_{t+1}^N \notin A(\epsilon) | Y_t^N = y] \leqslant \frac{K}{\delta(\epsilon)^2 N} .$$

In order to show that $\int_S f(x) p^N(y,dx) \to f(yQ(y)) = \int_S f(x)p(y,dx)$, we write $\int_S$ as $\int_{A(\epsilon)} + \int_{A(\epsilon)^c}$.

First note that

$$\left| \int_{A(\epsilon)} f(x)p^N(y,dx) - \int_{A(\epsilon)} f(x)p(y,dx) \right| = \left| \int_{A(\epsilon)} f(x)p^N(y,dx) - f(yQ(y)) \right|$$

$$\leqslant \epsilon \; p^N(A \; (\epsilon),y) \leqslant \epsilon .$$

Next,

$$\left| \int_{A(\in)^c} f(x)p^N(y,dx) - \int_{A(\in)^c} f(x)p(y,dx) \right| = \left| \int_{A(\in)^c} f(x)p^N(y,dx) \right|$$

$$\leqslant (\sup f) \cdot p^N(A(\in)^c,y) \quad \frac{K}{\delta(\in)^2 N} \cdot \sup f,$$

where supf: = sup(f(y):y$\varepsilon\mathcal{S}$) is finite since f is continuous and $\mathcal{S}$ compact. Consequently, $p^N(y,\cdot)\xrightarrow{w}p(y,\cdot)$ and the convergence is uniform in y.

■

Since $Q(\cdot)$ is continuous, $\int f(y,v)p(y,dv) = f(y,yQ(y))\varepsilon C(\mathcal{S})$ whenever $f\varepsilon C(\mathcal{S}^2)$. This observation, together with the fact that $p^N \xrightarrow{w} p$ uniformly on $\mathcal{S}$ allow us to apply Theorems 2.5 and 2.6 from section 2.3. To illustrate the usefulness of these theorems in the present context, some simple consequences are given in the next theorem.

Theorem 2.3

(i) If $Y_0^N \xrightarrow{\mathcal{D}} y$ and if f is a bounded, measurable, real function on S^∞ which is continuous (with respect to the product topology) at $(y,F(y),F^{(2)}(y),\ldots)$, then $Ef(Y_0^N,Y_1^N,Y_2^N,\ldots)\to f(y,F(y),F^{(2)}(y),\ldots)$ as $N\to\infty$.

(ii) $p_t^N \xrightarrow{w} p_t$ uniformly on $\mathcal{S}$ for each t.

(iii) If $Y_0^N \xrightarrow{\mathcal{D}} y$ (i.e. y is the random variable which only takes the value y$\varepsilon\mathcal{S}$), then for each $\in>0$ and each t,

$$\Pr\left[\max_{n=0,1,2,\ldots,t} |Y_n^N - F^{(n)}(y)| < \in\right] \to 1 \text{ as } N\to\infty .$$

Remark: If $Y_0^N \xrightarrow{\mathcal{D}} y$, then it follows from (2.7) and the above Theorem that for large N, the individual-level $\Pr(X_t^i = k)$ approximately equals the k-th coordinate of $F^{(t)}(y)$.(*)

Proof

(i) This is just a restatement of Theorem 2.6 in the present context.

(ii) Theorem 2.5 implies that for $f \varepsilon C(\mathcal{S}^\infty)$

$$E[f(Y_0^N, Y_1^N, \ldots, Y_t^N) \mid Y_0^N = y] \to f(y, F(y), F^{(2)}(y), \ldots, F^{(t)}(y)).$$

uniformly in y. Let π_t be the map $(y_0, y_1, y_2, \ldots) \to y_t$ on $\mathcal{S}^\infty$. Then for each $g \varepsilon C(\mathcal{S})$, $g(\pi_t(\cdot)) \varepsilon C(\mathcal{S}^\infty)$, and so

$$E[g(Y_t^N) \mid Y_0^N = y] \to g(F^{(t)}(y)) \text{ uniformly in y.}$$

But this is equivalent to $p_t^N \xrightarrow{w} p_t$ uniformly on $\mathcal{S}$.

(iii) Theorem 2.5 and $Y_0^N \xrightarrow{\mathcal{D}} y$ imply that

$$(Y_0^N, Y_1^N, \ldots Y_t^N) \xrightarrow{\mathcal{D}} (y, F(y), F^{(2)}(y), \ldots, F^{(t)}(y)).$$

Since the latter quantity is a constant, convergence in distribution is equivalent to convergence in probability (Billingsley 1968, pages 24, 25, and Theorem 5.1 Corollary 2), which proves (iii).

(*) In an early paper (Gerchak 1978), this property was proved by first showing that the variables X_t^i and X_t^j are asymptotically (when $N \to \infty$) independent. A result similar to part (iii) of Theorem 2.3 then followed.

Note that part (i) of Theorem 2.3 does not include steady state results, because of the continuity requirement. E.g. if we define $f_B(y_0,y_1,y_2,\ldots):=\lim_{t\to\infty} \frac{1}{t+1} \sum_{i=0}^{t} I_{[y_i \varepsilon B]}$ as the fraction of time the trajectory of profile vectors $(y_0,y_1,y_2,\ldots)$ is in the set B, then f_B is not continuous on S^∞.

2.2.2 Approximation of Steady State

Before we investigate the relation between the steady state behavior of the finite and infinite population models some relevant concepts from the theory of Markov chains (e.g. Breiman 1968) are summarized and the special meaning they obtain in the (deterministic) infinite population model is indicated.

Steady state behavior of Markov chains is characterized by the equivalence classes of communicating states and their corresponding stationary measures. A measure π of a Markov chain with probability transition function q is stationary if $\pi(B) = \int q(v,B)\pi(dv)$ for each Borel set B. For the Markov chain $\{Y_0,Y_1,Y_2,\ldots\}$, π is thus stationary if $\pi(B) = \int I_{[F(y)\varepsilon B]}\pi(dy)$ for each $B\varepsilon\mathcal{B}(S)$. The Markov chains $\{Y_0^N,Y_1^N,\ldots\}$ are finite state Markov chains. For each $y\varepsilon S$, $\pi(y,\cdot)$ defined by

$$\pi(y,B) = E[\lim_{t\to\infty} \frac{1}{t+1} \sum_{n=0}^{t} I_{[Y_n^N \varepsilon B]} | Y_0^N=y]$$

is a stationary probability measure. There is a unique stationary measure if and only if the chain is irreducible. If it is also aperiodic, then π defined by $\pi(B) = \lim_{t\to\infty} Pr[Y_t^N \varepsilon B]$ exists for any distribution of Y_0^N

and is the unique stationary measure, and the chain is said to be <u>asymptotically stationary</u>. The Markov chain $\{Y_0, Y_1, Y_2, \ldots\}$ has the deterministic probability transition function $p(y,B) = I_{[F(y)\varepsilon B]}$ and, given $Y_0 = y$, $Y_t = F^{(t)}(y)$. Again, $\pi(y,\cdot)$ defined by $\pi(y,B) = \lim_{t\to\infty} \frac{1}{t+1} \sum_{n=0}^{t} I_{[F^{(n)}(y)\varepsilon B]}$ exists for each $y\varepsilon \mathcal{S}$ and $B\varepsilon \mathcal{B}(\mathcal{S})$ and is a stationary probability measure. Define ϵ_y to be the measure on $\mathcal{B}(\mathcal{S})$ such that

$$\epsilon_y(B) = \begin{cases} 1 \text{ if } y\varepsilon B \\ 0 \text{ otherwise.} \end{cases}$$

Then y is a fixed point of F if and only if ϵ_y is a stationary measure for $\{Y_0, Y_1, Y_2, \ldots\}$. A fixed point of F exists by Brouwer's fixed point theorem, since $\mathcal{S}$ is convex and compact and F is continuous. The function F is said to be <u>globally stable</u> if there exists a $y^*\varepsilon\mathcal{S}$ such that $\lim_{t\to\infty} F^{(t)}(y) = y^*$ for each $y\varepsilon\mathcal{S}$. It is straight forward to check that F has a globally stable fixed point if and only if there exists an unique stationary measure μ of $\{Y_0, Y_1, Y_2, \ldots\}$ and in this case $\mu = \epsilon_{y^*}$.

<u>Theorem 2.4</u> Suppose Y_0^N has probability measure μ^N and Y_0 probability measure μ.

(i) If μ^N is a stationary measure for $\{Y_0^N, Y_1^N, Y_2^N, \ldots\}$ and $\mu^N \xrightarrow{w} \mu$, the μ is a stationary measure for $\{Y_0, F(Y_0), F^{(2)}(Y_0), \ldots\}$.

(ii) If μ is the unique stationary measure of $\{Y_0, F(Y_0), F^{(2)}(Y_0), \ldots\}$ (or equivalently, $\mu = \epsilon_{y^*}$ where y^* is a globally stable fixed point of F), and if for each N, μ^N is a stationary measure of $\{Y_0^N, Y_1^N, \ldots\}$, then $\mu^N \xrightarrow{w} \mu = \epsilon_{y^*}$. In this case,

$$\Pr[\max_{n=0,1,2,\ldots t} |Y_n^N - y^*| < \epsilon] \to 1 \text{ as } N\to\infty, \text{ and}$$

$$Ef(Y_0^N, Y_1^N, \ldots) \to f(y^*, y^*, y^*, \ldots) \text{ as } N\to\infty$$

for each bounded, measurable, real function f on S^∞ continuous at $(y^*, y^*, y^*, \ldots)$.

Remark: If F is globally stable with fixed point y^*, it follows from (2.7) and the above Theorem that the individual-level $\Pr(X_t^i=k)$ is approximately equal to the k-th coordinate of y^* for large N and large t.

Proof

(i) By hypothesis, $\mu^N(B) = \int p^N(y,B)\mu^N(dy)$. Also by hypothesis, $\mu^N \to \mu$, and by Theorem 3.1, $p^N \xrightarrow{w} p$ uniformly on S. Hence by Lemma 2.2 [interpret $q^N(y,\cdot) = \mu^N(\cdot)$ and $q(y,\cdot) = \mu(\cdot)$], μ satisfies $\mu(B) = \int p(y,B)\mu(dy)$ and is thus stationary.

(ii) Since S is compact, $\{\mu^N\}$ is relatively compact and hence any subsequence of μ^N contains a further subsequence which converges weakly (Billingsley 1968, Prokorov's Theorem). Suppose the subsequence $\mu^{N'}$ converges weakly to Π. Then since each $\mu^{N'}$ is a stationary measure, it follows from (i) that Π is a stationary measure for $\{Y_0, F(Y_0), F^{(2)}(Y_0), \ldots\}$. By hypothesis, μ is the unique stationary measure; hence $\Pi = \mu$. Thus each subsequence of $\{\mu^N\}$ contains a further subsequence with converges to μ. By Billingsley (1968, Theorem 2.3), this implies that $\mu^N \xrightarrow{w} \mu$. The last statement holds by Theorem 2.3, since $\mu = \epsilon_{y^*}$ and $y^* = F(y^*)$.

In light of the previous Theorem, it is tempting to conjecture that for a sufficiently large population, if the initial profile is y, then the profile after a long enough period of time t (i.e. "at steady state") should be near $F^{(t)}(y)$ with high probability. More precisely, this conjecture is that $\lim_{t\to\infty} Pr[|Y_t^N - F^{(t)}(y)|<\in| Y_0^N = y]$ should be close to 1 for N sufficiently large. Of course, Theorem 2.4 shows that the conjecture is true if F is globally stable.

The following example shows that the above conjecture is not true in general. In the example, the finite population models are asymptotically stationary for each N, so that $\lim_{t\to\infty} Pr[Y_t^N \varepsilon B|Y_0^N = y] = \mu^N(B)$ exists for each $B\varepsilon\mathcal{B}(S)$ and does not depend on y. However, $\lim_{t\to\infty} F^{(t)}(y)$ does depend on y since F has three fixed points in the example. Consequently, the conjecture cannot be true in general.

For the example, suppose that there are two categories (K=2), and that Q is defined by

$$Q_{11}(y) = \begin{cases} 1/4 + (7/8)y_1 & 0 \leqslant y_1 \leqslant 5/7 \\ 7/8 & 5/7 \leqslant y_1 \leqslant 1, \end{cases}$$

and $Q_{22}(y) = \frac{255}{256} - \frac{y_1}{2}$ where $Q_{12}(y) = 1 - Q_{11}(y)$, $Q_{21}(y) = 1 - Q_{22}(y)$, and $y = (y_1, 1-y_1)$. The function $F(y) = yQ(y)$ has three fixed points y^*, y^{**}, and y^{***}, where $y_1^* = 0.015751$, $y_1^{**} = 0.661332$, and $y_1^{***} = 0.752568$. The fixed points y^* and y^{***} are stable; y^{**} is unstable. Since $Q(y) > 0$ for each y, it follows that $Pr[Y_{t+1}^N = \hat{y}|Y_t^N = y] > 0$ for each y, $\hat{y}\varepsilon S_N$ and for each N. Hence the processes $\{Y_t^N: t = 0,1,2,\ldots\}$

are asymptotically stationary for each N, although F has several fixed points.

Neither is uniqueness of fixed point sufficient. Conlisk (1976, Appendix I, Case 3) provides a two-categories example for which the behavior of the infinite population model is characterized by a unique but unstable fixed point and a stable limit cycle (see graph on p. 178 there). However for that example, again, $Q(y) > 0$ for each y, so the finite population processes are asymptotically stationary. Hence the conjecture fails even here.

2.3 Weak Convergence of Sequences of Markov Chains

In this section we show that if the probability transition functions $p^N(\cdot,y) \xrightarrow{w} p(\cdot,y)$ uniformly in y and if the initial distributions $\mu^N \xrightarrow{w} \mu$, then the Markov chains corresponding to p^N and μ^N converge weakly to the Markov chain corresponding to p and μ. The Markov chains considered in this section are general and are not restricted to the profile processes defined in the previous sections.

The key result is Lemma 2.2, which demonstrates that uniform weak convergence is preserved when probability transition functions are composed. Theorem 2.5 then applies this result to show that the Markov chains, considered over a finite number of periods, converge weakly. A standard result from weak convergence then allows us to conclude that the entire processes converge weakly.

Let $\mathcal{R}$ be the real line and $\mathcal{R}^k$ the space of k-dimensional real vectors. For any metric space V, let $\mathcal{B}(V)$ be the Borel sets of V; i.e.

$\mathcal{B}(V)$ is the smallest σ-field containing all of the open sets in V. Let $C(V)$ be the set of real-valued, continuous, bounded functions on V. If U is also a metric space, define the map $(u,V)\to p(u,V)$ for $u\in U$ and $V\in\mathcal{B}(V)$ to be a probability transition function from U to V if for each $u\in U$, $p(u,\cdot)$ is a probability measure on $\mathcal{B}(V)$, and if for each $V\in\mathcal{B}(V)$, $p(\cdot,V)$ is a measurable function on U.

If ν and ν_N $(N=0,1,2,\ldots)$ are probability measures defined on $\mathcal{B}(V)$, then $\nu_N \xrightarrow{w} \nu$ if and only if

$$\int f(x)\nu_N(dx) \to \int f(x)\nu(dx) \text{ for each } f\in C(V).^{*} \quad \text{(Billingsley 1968).}$$

If p and p_N are probability transition functions from U to V then $p_N \xrightarrow{w} p$ uniformly on U if and only if

$$\int f(u,v)p_N(u,dv) \to \int f(u,v)p(u,dv) \text{ uniformly on } U \text{ as } N\to\infty \text{ for each}$$

$f\in C(U\text{x}V)$.

The next lemma is used in the proof of Theorem 2.5 and follows immediately from the definition of continuity.

Lemma 2.1 If $\int_V f(u,v)p(u,dv)\in C(U)$ whenever $f\in C(U\text{x}V)$ then $\int_V f(s,u,v)p(u,dv)\in C(S\text{x}U)$ whenever $f\in C(S\text{x}U\text{x}V)$.

Lemma 2.2 Suppose that S, U, and V are separable metric spaces, that q and $q_N(N=0,1,2,\ldots)$ are probability transition functions from S to U, and that p and $p_N(N=0,1,2,\ldots)$ are probability transition functions from U to V.

---

* We sometimes omit the domain of integration when it is the whole space.

If $q_N \xrightarrow{w} q$ uniformly on $\mathcal{S}$, $p_N \xrightarrow{w} p$ uniformly on $\mathcal{U}$, and $\int f(u,v)p(u,dv) \varepsilon C(\mathcal{U})$ whenever $f \varepsilon C(\mathcal{U} x \mathcal{V})$, then $\nu_N \to \nu$ uniformly on $\mathcal{S}$, where ν and ν_N (N=0,1,2,...) are probability transition functions from $\mathcal{S}$ to $\mathcal{U}x\mathcal{V}$ and are defined by

$$\nu_N(s,UxV) := \int_U p_N(u,V)q_N(s,du)$$

and

$$\nu(s,UxV) := \int_U p(u,V)q(s,du) \ .$$

Note that ν_N and ν are actually probability transition functions from S to $\mathcal{U}x\mathcal{V}$. Since $\mathcal{U}$ and $\mathcal{V}$ are separable, any measure defined on $\{UxV: U\varepsilon\mathcal{B}(\mathcal{U}), V\varepsilon\mathcal{B}(\mathcal{V})\}$ has an unique extension to $\mathcal{B}(\mathcal{U}x\mathcal{V})$ (Billingsley 1968, page 225). By Theorem III.2.1 (Neveu 1965) $\nu(\cdot,W)$ and $\nu_N(\cdot,W)$ are measurable for each $W\varepsilon\mathcal{U}x\mathcal{V}$.

Proof

We need to show that

$$\int_{\mathcal{U}x\mathcal{V}} f(s,w)\nu_N(s,dw) \to \int_{\mathcal{U}x\mathcal{V}} f(s,w)\nu(s,dw) \text{ uniformly in } s\varepsilon\mathcal{S}$$

for each $f\varepsilon C(\mathcal{S}x\mathcal{U}x\mathcal{V})$. By Fubini's Theorem and the definitions of ν_N and ν given in our Theorem, this may be rewritten as

$$\int_{\mathcal{U}}\int_{\mathcal{V}} f(s,u,v)p_N(u,dv)q_N(s,du) \to \int_{\mathcal{U}}\int_{\mathcal{V}} f(s,u,v)p(u,dv)q(s,du)$$

uniformly in $s\varepsilon\mathcal{S}$ for each f $C(\mathcal{S}x\mathcal{U}x\mathcal{V})$.

Applying the triangle inequality,

$$\left| \int_{U} \int_{V} f(s,u,v)\ p_N(u,dv)q_N(s,du) - \int_{U} \int_{V} f(s,u,v)p(u,dv)q(s,du) \right|$$

$$\leqslant \left| \int_{U} \int_{V} f(s,u,v)p_N(u,dv)q_N(s,du) - \int_{U} \int_{V} f(s,u,v)p(u,dv)q_N(s,du) \right|$$

$$+ \left| \int_{U} \int_{V} f(s,u,v)p(u,dv)q_N(s,du) - \int_{U} \int_{V} f(s,u,v)p(u,dv)q(s,du) \right| .$$

Suppose we are given $\in > 0$. Since, by hypothesis for $s\varepsilon S$ we have

$$\int_{V} f(s,u,v)p_N(u,dv) \longrightarrow \int_{V} f(s,u,v)p(u,dv) \text{ uniformly on } U,$$

it follows that the first term on the right hand side of the above inequality is less than $\in/2$ for N sufficiently large. Also by hypothesis, $\int_{V} f(s,u,v)p(u,dv)$ is continuous and bounded on $S\mathrm{x}U$ and $q_N \xrightarrow{w} q$ uniformly on S; hence the second term is less than $\in/2$ for N sufficiently large. Thus the desired limit has been established. ■

Assume that S is a Polish space, i.e., a complete, separable, metric space (e.g. $S \subseteq \mathbb{R}^K$). Let $\{Y_0, Y_1, Y_2, \ldots\}$ be a time homogeneous Markov chain with state space S, with probability transition function p from S to S, and with the initial conditions given by the probability measure $\mu(B) = \Pr[Y_0 \varepsilon B]$, $B \varepsilon \mathcal{B}(S)$.

Given that $Y_0 = y$, denote the joint distribution of $Y_1, Y_2, \ldots Y_t$ by $\nu_t(y,B) := \Pr[(Y_1,Y_2,Y_3,\ldots,Y_t)\varepsilon B | Y_0 = y]$ for $y\varepsilon S$, $B\varepsilon\mathcal{B}(S^t)$. Using Theorem V.1.1 from Neveu (1965), ν_t is a probability transition function from S to S^t. The functions ν_t are also characterized by the recursive relation

$$\nu_{t+1}(y,B_1 x B) = \int_{B_1} \nu_t(x,B)p(y,dx) \qquad (2.11)$$

for $B_1 \subseteq S$ and $B \subseteq S^t$.

Unconditioning on Y_0, define the joint distribution function

$$\hat{\nu}_t(B) := \Pr[(Y_0,Y_1,\ldots,Y_t)\varepsilon B] \text{ for } B\varepsilon \mathcal{B}(S^{t+1}).$$

The measures $\hat{\nu}_t$ are characterized by the relation

$$\hat{\nu}_t(B_1 x B) = \int_{B_1} \nu_t(y,B)\mu(dy) \quad \text{for } B_1 \subseteq S \text{ and } B\subseteq S^t. \qquad (2.12)$$

Now, in addition, consider a family of Markov chains $\{Y_0^N, Y_1^N, Y_2^N, \ldots\}$ for N = 1,2,..., each defined on the same state space S. For each process, define p^N, μ^N, ν_t^N, and $\hat{\nu}_t^N$ as above. (Since the probability space actually changes as we change N, we should also use $\Pr^N$. However, the particular probability space will be clear from the context and this superscript will be omitted.) Of course, (2.11) and (2.12) apply with the superscripted measures.

<u>Theorem 2.5</u> If $p^N \xrightarrow{w} p$ uniformly on S and if $\int f(u,v)p(u,dv)\varepsilon C(S)$ whenever $f\varepsilon C(S^2)$ and $u\varepsilon S$, then for each t, $\nu_t^N \xrightarrow{w} \nu_t$ uniformly on S as $N\to\infty$. If, in addition, $\mu^N \xrightarrow{w} \mu$, then for each t $\hat{\nu}_t^N \xrightarrow{w} \hat{\nu}_t$ as $N\to\infty$; or equivalently

$$(Y_0^N, Y_1^N, Y_2^N, \ldots, Y_t^N) \xrightarrow{\mathcal{D}} (Y_0, Y_1, Y_2, \ldots, Y_t).$$

Proof

We will prove by induction that

(i) $\int f(u,v)\ \nu_t(u,dv) \varepsilon\ C(S)$ whenver $f \varepsilon C(S^{t+1})$, and

(ii) $\nu_t^N \to \nu_t$ uniformly on S as $N\to\infty$.

Since $\nu_1 = p$, the above statements are true by hypothesis when $t = 1$. Suppose statements (i) and (ii) hold for some t.

$$\text{Then} \int_{S^{t+1}} f(y,x)\nu_{t+1}(y,dx) = \int_{x\varepsilon S} \int_{\hat{x}\varepsilon S^t} f(y,x,\hat{x})\nu_t(x,d\hat{x})\ p(y,dx)$$

follows from (2.11). The inner integral belongs to $C(S^2)$ by Lemma 2.1 and induction hypothesis (i). Consequently, the outer integral belongs to $C(S)$. Now apply Lemma 2.2 to show that ν_{t+1}^N, characterized by

$\nu_{t+1}^N(y,B_1xB) = \int_{B_1} \nu_t^N(x,B)\ p^N(y,dx)$, converges weakly and uniformly on S to ν_{t+1} characterized by (2.11). This completes the proof of (i) and (ii).

The second conclusion of the Theorem follows by applying Lemma 2.2 to the characterization of $\hat{\nu}_t^N$ and $\hat{\nu}_t$ given by (2.12). The hypotheses of Lemma 2.2 are satisfied by (i), (ii), and the assumption that $\mu^N \xrightarrow{w} \mu$ uniformly on S.

■

Theorem 2.6 Assume that the hypotheses of Theorem 2.5 hold, including $\mu^N \xrightarrow{w} \mu$. Then $(Y_0^N,Y_1^N,Y_2^N,\ldots) \xrightarrow{\mathcal{D}} (Y_0,Y_1,Y_2,\ldots)$. And for any bounded, measurable, real valued function f defined on S^∞ with discontinuity set D_f such that $\hat{\nu}(D_f) = 0$,

$$Ef(Y_0^N,Y_1^N,Y_2^N,\ldots) \to Ef(Y_0,Y_1,Y_2,\ldots). \tag{2.13}$$

Proof The Theorem follows from Theorem 2.5 and the discussion at the beginning of Section 5 and page 19 in Billingsley. (Although Billingsley assumes $S \subseteq \mathcal{R}$, only completeness is used in his argument.) If $f \varepsilon C(S^{\infty})$, then (2.13) follows from the definition of "$\xrightarrow{\mathcal{D}}$".

Theorem 5.2 in Billingsley shows that the continuity of f can be relaxed to the stated conditions.

■

2.4 Concluding Remarks

Although a major part of this chapter was concerned with the approximation of the finite population model by the infinite population one, the possibility of using the finite population model directly should not be ignored. As we pointed out, it can be simulated without difficulty. Also, fairly general and easily checked conditions for asymptotic stationarity were given, whereas comparable results for global stability (of the infinite population model) have been obtained (Conlisk 1976, 1978a, Smallwood and Conlisk 1979) only for special cases.

Nevertheless, for certain parametric families of the infinite population model it is sometimes feasible (Smallwood and Conlisk 1979) to discover the dependence of the model's equilibrium on some underlying parameters. This is certainly an advantage of the infinite population model which, coupled with its computational simplicity, justifies the effort we have put into proving its validity.

CHAPTER III
THE MEASUREMENT OF SOCIAL INHERITANCE

A major aim of social scientists, who engaged in modelling intergenerational mobility processes, was often either to study the international differences in the rates of mobility, or to analyze mobility trends over time. One aspect, of particular interest to sociologists, was the dependence of a person's social (occupational; income) class on that of his father. This gave rise to what is frequently referred to as "measurement of mobility" and led to the formal problem of building an acceptable index of mobility. For historical account and survey see Boudon (1973).

The term "rate of mobility" is, however, quite ambiguous and may be given entirely different interpretations even within the context of a single phenomenon like, say, occupational mobility. A common interpretation is that of the net redistribution of the working force by functional categories - industries and occupations (sometimes referred to as "structural mobility"). This aspect of occupational mobility is of particular interest in the context of development economics, since economic growth is known to be accompanied by such net redistribution (see Smelser and Lipset 1966, especially the contribution by Duncan).

Social scientists with primary interest in issues of equity, on the other hand, focus more on patterns of gross mobility. They are interested in assessing the deviation of a given society from some social ideal, such as the one in which son's class does not depend on that of previous generations in his family line. Shorrocks (1978) refers to this aspect as the "predictability" of society - the extent to which future positions are dictated by the current place in the distribution.(*) (Sometimes referred to also as "circulation" mobility).

Since the predictability (deviation from social ideal) aspect of social mobility seems most interesting and challenging as far as measurement is concerned, we shall focus our attention on it. We make it explicit by referring to the problem specifically as measurement of "social inheritance",(**),(***) rather than "measurement of mobility". Since high degree of social inheritance corresponds to what was previously labelled as low

(*) Still another aspect of mobility, of particular importance in migration and other types of intragenerational mobility, is the "physical" rate at which individuals change locations (e.g. number of residence changes per unit time; see Long (1970) and Section 1.4). Although "...as more movement is observed it would be normal to expect the class occupied in the future to become less dependent on the present position. In general, therefore, they [the rate of movement and the non-predictability of society] should be in harmony" (Shorrocks 1978, p. 1016), these two aspects of mobility are not perfectly correlated, and should not be confused. [See also Sommers and Conlisk (1979, pp. 254-255)].

(**) Pullum (1975) used the term "occupational inheritance". Another possible term is "measurement of equality of social opportunity".

(***) This will also be consistent with an argument made by Duncan (1966) according to which available mobility tables provide only "inheritance", and not "mobility", data.

degree of mobility, our ordering will reverse the more common one.(*),(**)

An underlying assumption in the following discussion is that the processes of interest are time-homogeneous, which seems to have some empirical support as far as intergenerational mobility is concerned (Bartholomew 1973). However, even if the process is not really time-homogeneous, it might be interesting to ask what are the long term implications of the current trend. The degree of social inheritance in two societies can be compared on the basis of data pertaining to similar dates, perhaps combining "generation-specific" indices to an overall one.(***)

In Section 1 we discuss, in non-mathematical terms, the properties one would like such measures to have (they are related to the ones mentioned by Shorrocks 1978). Various concepts through which social inheritance is manifested are then discussed.

The second section then formulates a (Markovian) model as a mapping over the unit simplex. It then states desirable properties and inheritance concepts mathematically and discusses their implications. Various concept-dependent ways to measure "non-constancy" of the operator are suggested, and it is shown that many known measures, as well as new ones, can be obtained in this way. Some special cases are analyzed in detail. A method of introducing "period consistency" (Shorrocks 1978) is then discussed.

(*) Though it will be consistent with Goodman's (1969) notion of "status persistence" and with Sommers and Conlisk's (1979) "Immobility".

(**) For an interesting discussion of measures of (static) income inequality see Kolm (1976).

(***) See also Shorrocks (1978, p. 1021).

3.1 Preliminary Discussion

As we shall soon see, even zeroing in on this single aspect of mobility does not provide sufficient guidance for constructing measures. But already at this stage we can lay down some general requirements from such a measure. For the moment, they will be stated in non-mathematical form; in Section 2, we shall restate them mathematically. An underlying assumption is, however, that social mobility over generations is Markovian. The "states" may, however, be taken to be either social classes or distributions over classes ("profiles") so the set-up is general enough to include Markovian models of previous chapters.

3.1.1 Basic Requirements

If inequities due to differences in origin entirely disappear within few generations, the degree of social inheritance is as low as it can be.(*) All other types of societies exhibit some degree of inheritance. The slower the "rate" at which inequities due to origin disappear, the higher should the society rank on a social inheritance scale. At the extreme, we find societies where any person's class is always identical to that of his father - a perfect caste system. Those should be assigned the highest value of the measure.

Since one may wish to compare social inheritance in societies for which the length of time interval for which data are available might be different (say, one generation vs. two), measures should be required to

(*) Within this class, we may rank societies according to the number of generations it takes.

be "Period Consistent" (Shorrocks 1978). This is to say that a society whose degree of social inheritance is ranked higher than another's on the basis of one generational inheritance data, should be so ranked on the basis of data corresponding to any number of generations.

3.1.2 Concepts of Social Inheritance

The aspect of social mobility (inheritance) which we wish to measure has many facets which are not necessarily perfectly correlated with each other. We shall now discuss some facets of social inheritance and their implications for purposes of measurement.

As we said, a measure of social inheritance should indicate the rate at which inequities, due to differences in origin classes, disappear over generations. Now, what if for a certain society, regardless of how many generations pass, some "basic" inequities due to origin still prevail (while others, perhaps, disappear)? Loosely speaking, in such cases one could lump social classes together and end up with a perfect caste system over the lumped classes. Hence we feel that there is a justification to assign all such processes (and not only those which correspond to perfect caste systems over the original classes) the highest value of the measure.

The above point gives rise, however, to a general fundamental issue, which has (surprisingly) achieved little attention in the mobility measurement literature. Given any concept of social inheritance, a measure can focus on a facet which exhibits "highest" social inheritance, or it can "average" all the facets (circumstances). For example, social critics frequently point out that the rate of reduction in some specific inequities in a given society (say, a gap in access to high education between certain groups) is too slow, and disregard (possible) fast reductions in equities simultaneously taking place in another part (or aspect) of society. Specifically, the above distinction is particularly relevant in the treatment of:

a. "Destination" classes (or profiles).

   The chances of joining certain social classes may depend on origin class more than chances to join other classes do. The measure will then be an "average" weighted over all destination classes, and the weights will "express" the user's social priorities. At one extreme, all weight will be assigned to the most origin-dependent destination, and on the other all destinations will be weighted equally.

b. "Origin" classes (or profiles).

   Using "differences" between distributions over destination classes for each pair of origin classes as a basic tool, the measure will be a weighted average over all these pairs. At one extreme, all weight will be assigned to the pair of origins which generates the largest difference. At the other extreme, all pairs will be weighted equally.

   Even if inequities due to difference in origin virtually disappear after a sufficient number of generations, this may take longer for the offsprings of some origin-classes (or, more generally, for some initial distributions) than for others. Again, one can focus on either the "slowest" origin-effect to disappear, or compute an average over all origins.

One manifestation of the rate at which inequities due to difference in origin-class disappear are the differences between the distribution over classes of individuals whose fathers belonged to different classes (i.e. the amount of "social scrambling" that takes place from one generation to the next). The relation of such two successive generations' distributions to each other after many generations had passed may serve as a basis for a measure of social inheritance.

We shall now turn to a mathematical discussion.

3.2 Mathematical Discussion

3.2.1 Definitions

Let $\mathcal{K} = \{1,2,...,K\}$ be a set of categories (social classes or distributions over classes). Let P_{i*} be a conditional probability mass function over $\mathcal{K}$, so that P_{ij} is, say, the probability that a son of class-i-father belongs to class j. Let $\mathcal{P}$ be the set of probability mass functions over $\mathcal{K}$, i.e. if $p\varepsilon\mathcal{P}$ then $p=(p_1,...,p_K)$ such that p_i = probability that a given individual belongs to class i (or that the profile is of type i).

$\mathcal{P}$ can be modelled by a K-1 dimensional simplex S, which is a convex combination of K linearly independent points $\{V_1,...V_K\} \subseteq R^K$. The points $\{V_1,...,V_K\}$ are arbitrary points in R^K except for the restriction that they be linearly independent. The vertices $V_1,\ldots,V_K$ correspond to the probability mass functions (or profiles) $(1,0,\ldots,0)$,..., $(0,\ldots,0,1)$ respectively. The probability distribution $p=(p_1,\ldots p_K)$ corresponds to the point $\sum_{i=1}^{K} p_i V_i$.

The conditional probability mass functions $(P_{i*}:i=1,2,\ldots,K)$ can be characterized as a map T from S to itself, with the property that $T(\sum_{i=1}^{K} p_i V_i) = \sum_{i=1}^{K} p_i T(V_i)$ for each $\sum_{i=1}^{K} p_i V_i \varepsilon S$.(*) There is a unique extension of T from S to a linear map on R^K. Let $T^{(n+1)}(X) = T(T^{(n)}(X))$, where $T^{(1)}(X) = T(X)$. If $\{V_1,\ldots,V_K\}$ happens to be used as the basis of R^K, then the matrix representation of T is P whose i-th row is the vector P_{i*}, and $Tp = pP$.(**) If a different basis $\{b_1,\ldots,b_K\}$ in R^K is chosen, then there exists an invertible matrix A such that if $x=\sum p_i V_i$ then $x = \sum (pA^{-1})_i b_i$, and the matrix representation of T is APA^{-1} (see, for example, Halmos 1958).

T will be considered a "constant" map if there exists p_T^* such that $Tp = p_T^*$ for every $p \varepsilon S$. It is possible that although T itself is not constant, there exists n such that $T^{(n)}p = p_T^*$ for every $p \varepsilon S$ (Brosh and Gerchak 1978). T is an "identity" map (denoted by I) if $TV_i = V_i$ for every $i \varepsilon K$.(***)

(*) We shall sometimes abreviate T(X) by TX.

(**) This representation is the common one in the literature since it describes the Markovian nature of the process.

(***) For any $p \varepsilon S \ \exists\ \alpha_1,\ldots,\alpha_K$ such that $p = \sum_{i=1}^{K} \alpha_i V_i$. Now, $I\, p = I(\sum_{i=1}^{K} \alpha_i V_i) = \sum_{i=1}^{K} \alpha_i I V_i = \sum_{i=1}^{K} \alpha_i V_i = p$. Hence the above is enough to characterize an identity map over all S.

3.2.2 Mathematical Statement of Desirable Properties

Let us fix $\{V_i: i=1,\ldots,K\}$, use it as a basis, and consider the aspect of social inheritance. What we actually want is to impose a simple order on maps of the kind defined above. A simple order is a relation which is connected, transitive and antisymmetric (for details, see Krantz, Luce, Suppes and Tversky 1971). Now, the number of K-state rational-valued probability transitions matrices is countable and this set is dense in the set of all K-state probability transition matrices. Hence Theorem 2 in Krantz et al. (1971, p. 40) implies that for any simple order over those transition matrices there exists an isomorphism into the real line. Hence we shall restrict ourselves to real-valued functions (measures). Moreover, since in the previous section we have identified processes for which we wish the measure to attain its maximum and minimum, we can restrict ourselves to any closed interval, and in particular to the common choice [0,1].

Denote functions which map the T's to [0,1] by M. Some of the general properties required above can now be stated as follows.

a. $M(T) = 0$ if and only if there exists an integer n such that $T^{(n)}$ is a constant map. The "if" part implies Shorrocks' "Perfect Mobility" condition ["if T is constant $M(T) = 0$"]; his "strong" version, however, assigns the extreme value only if n=1.

b) $M(T) = 1$ if and only if $\lim_{n\to\infty} T^{(n)}p$ is not constant in p. This implies Shorrocks' "Immobility" condition ($M(I)=1$) since the identity map has this property. His "strong" version, however, assigns the extreme value to the measure only if T is an identity map.

Thus the remaining task is to rank non-constant irreducible-aperiodic Markov chains by the degree of social inheritance they exhibit.

The "Period Consistency" condition can be stated as follows: if $M(T) \geq M(T')$ then $M(T^n) \geq M(T'^n)$ for every n.(*)

Shorrocks, who assumed that individuals follow independent and identically distributed Markov chains (i.e. the categories in his model were the usual social classes) introduced the following partial order over the probability transition matrices, which he called "Monotonicity": $P \prec P'$ if for every $j \neq i$ $P_{ij} \leq P'_{ij}$ and for some $j \neq i$ $P_{ij} < P'_{ij}$. However, since constant maps were previously assigned the lowest value of the measure, then in order to achieve consistency between the two some further restriction of the partial order is necessary. Indeed, Shorrocks restricted it to P's which have a "quasi-maximal diagonal": there exists positive numbers $\mu_1,\ldots,\mu_K$ such that $\mu_i P_{ii} \geq \mu_k P_{ik}$ for every i and k.

The above partial order will be too strong if one wishes to focus on the extremes of society. In particular, if what we wish to express in the measure is the "largest" inequity in society, "improvements" in other parts (facets) of society should not affect the measure. This may be achieved by weakening the notion of Monotonicity to "Weak Monotonicity": $P \precsim P'$ if for every $j \neq i$ $P_{ij} \leq P'_{ij}$.

Although it will be nice if the total orders we shall come up with will be consistent with the above partial order (and some will indeed

(*) Suppose that $M(T) \geq M(T') \Rightarrow M(T^n) \geq M(T'^n)$ for every T,T' and $n \geq 1$, and assume that for some T,T' and n^* $M(T^{n^*}) > M(T'^{n^*})$. Suppose now that for these T and T' $M(T) < M(T')$. Then, by the first assumption, $M(T^n) \leq M(T'^n)$ for every n, and in particular for n^*. Contradiction. Hence "Period Consistency", the way it was defined, implies the seemingly stronger property "if $M(T^k) \geq M(T'^k)$ for some integer k then $M(T^n) \geq M(T'^n)$ for every integer n".

be), it is not that crucial given our approach as it has been with Shorrocks'. We shall essentially arrive at measures in a constructive manner, using various facets of social inheritance as bases, while Shorrocks had only "Monotonicity" to help him evaluate the suitability of arbitrary functions as measures.

As pointed out by Sommers and Conlisk (1979, p. 254) "...immobility might be thought of as the slowness with which the state probabilities of a Markov chain 'escape' the effects of initial conditions on route to their equilibrium values. The slower this convergence, the more strongly the parent's state influences the life chances of the child, grandchild, and so on. That is, the 'mathematical' problem of measuring the slowness of a Markov chain's convergence to equilibrium is closely akin to the immobility measurement problem". It will be nice, then, if our measures will be related to this rate of convergence.

3.2.3 Measuring the Non-Constancy of the Operator

Measuring second-generation inequities due to differences in origin naturally involves measuring the non-constancy of the operator T. If we "allow" the process to start from any distribution,(*) it amounts to evaluating differences among all $\{T(p): p \in S\}$. If we restrict our interest to vertex-origins only, we evaluate differences among only $\{T(V_i): i=1,\ldots,K\}$. If $\lim_{n\to\infty} T^{(n)}(p) = p_T^*$ for every $p \in S$, we may choose to measure the distances between $\{T(p): p \in S \ [p=V_i: i=1,\ldots,K]\}$ and p_T^*.

(*) Say, varying a random mechanism for choosing an individual.

We now wish to operationalize the notion of differences/distances between distributions. Let p and q be any probability mass functions over K, and let d(p,q) be a "distance-function". Such a real-valued function is referred to as <u>metric</u> (Royden 1962) if for every p,q and r on K:

i. $d(p,q) \geq 0$;

ii. $d(p,q) = 0$ if and only if $p=q$;

iii. $d(p,q) = d(q,p)$; and

iv. $d(p,q) \leq d(p,r) + d(r,q)$.

If condition (ii) is relaxed to read $d(p,p)=0$, the function d is called a pseudometric.

A most common way to construct metrics (pseudometrics) is by defining a <u>norm</u> over differences of distributions. A norm is a non-negative real-valued function $\|.\|$ such that

1. $\|x\| = 0$ if any only if $x = o$;
2. $\|x+y\| \leq \|x\| + \|y\|$; and
3. $\|\alpha x\| = |\alpha| \|x\|$.

If condition (1) is relaxed to read $\|0\| = 0$ the function is called a pseudonorm.

A natural step in the direction of constructing measures of non-constancy is thus to select a function g and a norm $\|.\|$, and consider expressions like $\|g(Tp)-g(Tq)\|$.

The most obvious function to consider is just the identity function, i.e. to focus on $\|Tp - Tq\|$. The differences $T(p) - T(q)$ are K-dimensional, so a natural family of norms to consider with such g is

$L_a = (\sum_{i=1}^{K} |x_i|^a)^{1/a}$. For a=1 we obtain the summation ("averaging") norm $\sum_{i=1}^{K} |x_i|$ (to be denoted by $\| \|_1$). For $a = \infty$ we get the supremum norm $\sup_i |x_i|$ (to be denoted by $\| \|_\infty$).

Another natural direction to proceed is to try and come up with a function g which by itself maps from the boundary of the K-dimensional unit simplex to the real line (all L_a-type norms then reduce to the absolute-value norm). A well-known such function which turned out to be useful in many contexts, is the entropy of the distribution (e.g. Khinchin 1957). For any probability mass function p it is defined as $H_p = \sum_i p_i \log {}^1/_{p_i}$. The pseudometric(*) $|H(Tp)-H(Tq)|$ is then an "alternative" to the pseudometric(**) $\| T(p) - T(q)\|$ as a "basis" for a measure of non-constancy.

Let us now try to combine these pseudometrics with the notions of social inheritance discussed in subsection 3.1.2 and in the beginning of this subsection. A measure of non-constancy will be a choice of combination of:

a. nature of origins (vertices vs. distributions);

b. distance generated by pairs of origins vs. distance to steady state;

c. pseudometric over destinations (see above);

d. norm over origins.

(*) $H_p = H_q$ does not imply that p=q (consider, for example, any two permutations).

(**) T(p) = T(q) does not imply that p=q.

We shall now specify and discuss some of the more interesting combinations.

3.2.4 Some Special Cases

A. For every pair of origin vertices, calculate their images under T and their difference using the L_1-norm. Take the measure to be the supremum of the above over all such pairs. We obtain the function $\sup_{i,k}\|TV_i - TV_k\|_1$.

$\sup_{i,k}\|TV_i - TV_k\|_1$ equals (twice) the "delta coefficient" of T, which equals one minus the "ergodic coefficient" of T - a useful tool in the analysis of Markov chains (e.g. Isaacson and Madsen 1976). The ergodic coefficient has been actually suggested, in an entirely different context, as a measure of the "scrambling power of a matrix ... the degree to which it approaches a matrix with identical rows which scrambles all traces of the past" (Hajnal 1958, p. 236). Since this notion seems close to that of constancy, we shall now "operationalize" the delta coefficient as a measure of social inheritance, and investigate its properties. This will be done using $\{V_1,\ldots,V_K\}$ as a basis - the common representation. In this setting the delta-coefficient becomes (Isaacson and Madsen 1976)

$$\delta(P) = 1 - \min_{i,k} \sum_{j=1}^{K} P_{ij} \wedge P_{kj},$$

where $P_{ij} \wedge P_{kj} = \inf(P_{ij}, P_{kj})$.

A delta-coefficient-based ordinal which we shall discuss as a measure is the following:

$M(P) < M(P')$ if and only if $\delta(P^n) < \delta(P'^n)$, where $n = \inf\{m: \delta(P^m) \neq \delta(P'^m)\}$; if $n = \infty$ then $M(P) = M(P')$.(*)

We shall now state (and, for some non-standard assertions, prove) some of the properties of the above measure.

a. $0 \leq \delta(P) \leq 1$ for every probability transition matrix P [Shorrocks (1978) referred to this property as "Normalization"].

b. $\exists m$ such that $\delta(P^m) < 1$ if and only if $\exists k$ such that $P^k > 0$ (i.e. all elements of P^k are strictly positive for some k). Thus $\delta(P^n) = 1$ for every n if and only if P is reducible.

Viewed in another way, $\delta(P) < 1$ if and only if, for any two origin classes, there exists at least one destination class whose occupants might have come from either one of them (i.e. if and only if some social "scrambling" takes place) - an attractive feature of this measure. Note that the above statement "looked" at the processes backwards in time, which is rather natural in

---

(*) It is known that if $\inf\{m:\delta(P^m) < 1\} < \infty$ then $\inf\{m:\delta(P^m) < 1\} \leq [{(K-1)^2}/2 + 1]$, where [ ] denote the "integer part of". Hence, in order to determine this infimum, only a relatively small number of powers of P has to be checked.

the context of "future's independence of past".

c. $\delta(P^m) = 0$ if and only if all rows of P^m are equal, i.e. if and only if the Markov chain converges in a finite number of periods.

d. Let us now restrict the partial order of "weak monotonicity" $\preceq$ to a class P such that $P \varepsilon P$ only if for every i $P_{ii} \geq P_{ji}$ for every j.(*) Given any $P \varepsilon P$ and i^*, $j^* \varepsilon K$ define P^ε as follows:

$$P_{ij}^{\varepsilon} = \begin{cases} P_{ij} & \text{for } i \neq i^* \\ P_{i^*j} & \text{for } i = i^* \; j \neq i^* \; j \neq j^* \\ P_{i^*i^*} - \varepsilon & \text{for } i = i^* \; j = i^* \\ P_{i^*j^*} + \varepsilon & \text{for } i = i^* \; j = j^* \end{cases},$$

where $\varepsilon \geq 0$ is any number such that $P_{i^*i^*} - \varepsilon \geq P_{ki^*}$ for every $k \neq i^*$, and that $P_{i^*j^*} + \varepsilon \leq P_{j^*j^*}$ (so that $P^\varepsilon \varepsilon P$).

$$\text{Now, } 1 - \delta(P^\varepsilon) = \min_{i,k} \sum_{j=1}^{K} P_{ij}^{\varepsilon} \wedge P_{kj}^{\varepsilon}$$

$$= \{\min_{\substack{i \neq i^* \\ k \neq i^*}} \sum_{j=1}^{K} P_{ij} \wedge P_{kj}\} \wedge \{\min_{k} \sum_{j=1}^{K} P_{i^*j}^{\varepsilon} \wedge P_{kj}\} .$$

But

$$\min_{k} \sum_{j=1}^{K} P_{i^*j}^{\varepsilon} \wedge P_{kj} = \min_{k}\{ \sum_{\substack{j \neq i^* \\ j \neq j^*}} P_{i^*j} \wedge P_{kj} + (P_{i^*i^*} - \varepsilon) \wedge P_{ki^*} + (P_{i^*j^*} + \varepsilon) \wedge P_{kj^*}\}$$

(*) This condition is somewhat more restrictive than Shorrocks' "quasi-maximal diagonal" (for which it is sufficient). We chose to use it here since it is more convenient to work with, as well as having an intuitive appeal. Also, as pointed out by Shorrocks, it is "easy to confirm by inspection and holds for a large number of transition matrices [reported in the empirical literature]".

$$= \min_k \{ \sum_{j \neq j^*} P_{i^*j} \wedge P_{kj} + (P_{i^*j^*} + \varepsilon) \wedge P_{kj^*} \}$$

$$\geq \min_k \sum_{j=1}^{K} P_{i^*j} \wedge P_{kj} \quad .$$

Hence $1 - \delta(P^{\varepsilon}) \geq 1 - \delta(P)$,

i.e. $\delta(P^{\varepsilon}) \leq \delta(P)$.

So we proved that if for every i $P_{ij} \geq P'_{ij}$ for every $j \neq i$, then $\delta(P) \leq \delta(P')$. This is the "weak" version of "Monotonicity", which is natural since the delta coefficient focuses on the pair of origins which generates the largest difference in next-generation distributions.

e. Suppose that $\lim_{n\to\infty} P^n = P^*$ (where P^* has equal rows), and let $\|A\| = \sup_i \sum_j |a_{ij}|$ for any matrix A. Then $\|P^n - P^*\| = \|I \cdot P^n - P^* \cdot P^n\|$ $= \|(I-P^*)P^n\|$. Hence by Lemmas V.2.3 and V.2.4 in Isaacson and Madsen (1976) $\|P^n - P^*\| \leq \|I-P^*\| \delta(P^n) \leq \|I-P^*\| [\delta(P)]^n$. We see then that the geometric rate of convergence of P^n to P^* can be expressed in terms of the delta coefficient of P, as well as in terms of the second largest eigenvalue of P (to be discussed next).

f. Let $\hat{T}$ be the operator which maps distributions over profiles into distributions over profiles. Define the profile-level delta coefficient $\delta_{\text{prof.}}(\hat{T}) = \max_{p.p'} \|\hat{T}p - \hat{T}p'\|_1$. If the individual-level processes are independent and identically distributed then $\hat{T}p = \frac{1}{N} \sum_{n=1}^{N} TV_n$, where V_n is the "location vector" of the n-th individual. As before, on individual level $\delta_{\text{ind.}}(T) = \max_{i,k} \|TV_i - TV_k\|_1$, and denote the pair of origin classes for which this maximum is attained by (i^*,k^*).

Since we can choose p to be a profile corresponding to $V_n = i^*$ $n=1,\ldots,N$ [In this case $\hat{T}p = TV_{i^*}$] and p' to be the one corresponding to $V'_n = k^*$ $n=1,\ldots,N$, it follows that $\delta_{pop}(\hat{T}) \geq \delta_{ind}(T)$. Now, by the triangle inequality $\| \sum_{n=1}^{N} T(V_n - V'_n) \| \leq \sum_{n=1}^{N} \| TV_n - TV'_n \|$. Hence

$$\delta_{pop}(\hat{T}) = \max_{p,p'} \left\| \frac{1}{N} \sum_{n=1}^{N} T(V_n - V'_n) \right\|$$

$$\leq \frac{1}{N} \max_{\substack{(V_1,\ldots,V_N) \\ (V'_1,\ldots,V'_N)}} \sum_{n=1}^{N} \| TV_n - TV'_n \| .$$

Taking the maxima sequentially (individual by individual: first over (V_1,V'_1) etc.), we get

$$\frac{1}{N} \max_{\substack{(V_1,\ldots,V_N) \\ (V'_1,\ldots,V'_N)}} \sum_{n=1}^{N} \| TV_n - TV'_n \| = \frac{1}{N} N \| TV_{i^*} - TV_{k^*} \| = \delta_{ind}(T).$$

Hence $\delta_{pop}(\hat{T}) = \delta_{ind}(T)$.

Conclusion: In the i.i.d. case the delta coefficient assigns equal values to the profile-level operator as to the individual-level one.

B. Consider distributions over origins and the L_∞ distance of their images under T to the steady-state distribution. Consider then the norm of the map $(T(\ ) - p^*_T)$ (Royden 1963, p. 160), namely $\sup_p \dfrac{\| T(p) - p^*_T \|_\infty}{\| p - p^*_T \|_\infty}$.

This function is nothing but $|\lambda_2|$, where λ_2 is the second-largest (in norm) eigenvalue of T. If b_2 is the eigenvector of T corresponding to λ_2, then the supremum defined above is attained for $\hat{p} = p^*_T + \bar{\alpha}\, b_2$, where $\bar{\alpha} = \max\{\alpha : p^* + \alpha b_2 \in S\}$, i.e. $\hat{p}$ represents the "worst" direction (in terms of rate of convergence to steady state).

$|\lambda_2|$ has been suggested as a measure of (what we refer to as) social inheritance by Theil (1972, Ch. 5) and Shorrocks (1978), who also discussed some of its properties. A strong case in favor of this measure(*) was made recently by Sommers and Conlisk (1979). Among other things, they show its intimate relation to parent-child status correlation measure, and to the process' rate of regression to the mean. If, instead of considering the worst direction, we would average by summing over all eigenvector directions $b_1, \ldots, b_K$ we would get the sum of the eigenvalues,(**) which (in the usual setting) is nothing but the trace of P.(***) It has been discussed as (basis for) a measure of (what we refer to as) social inheritance by Shorrocks.

(*) Sommers and Conlisk also suggested to use the second largest eigenvalue of the matrix $P_s = \frac{1}{2}(P + \Pi^{-1} P^t \Pi)$, where $\Pi = \mathrm{diag}(\bar{\pi})$.

(**) If, instead of summing, we would have multiplied the eigenvalues by each other, we would have obtained the determinant of the transition matrix. It was discussed as a measure by Bartholomew (1973) and Shorrocks (1978). This function is, however, heavily influenced by the aspect of society with least inheritance, and as such is not very interesting. Perhaps a more rewarding function of this nature will be $1 - \prod_{i=2}^{K} (1-\lambda_i)$.

(***) When K=2 it can be easily shown that $|\alpha(P)| = |1 - \text{trace } P| = |\lambda_2|$. Thus in this special case the above mentioned measures coincide.

C. Consider now the pseudometric obtained by taking the difference in entropies of two distributions. In particular, consider $|H_{T(V_i)} - H_{p_T^*}|$ weighted by the steady-state probabilities $p^*_{T_i}$ i=1,...,K. Explicitly, we are referring to the measure

$$\left|-\sum_{i=1}^{K} p_i^* \sum_{j=1}^{K} P_{ij} \log P_{ij} + \sum_{j=1}^{K} p_j^* \log p_j^*\right|.$$

The above function can be shown to be the logarithm of the expression $\prod_{i,j} \left\{\frac{p_j^*}{P_{ij}}\right\}^{P_{ij} p_i^*}$. This last expression, is, in turn, nothing but the likelihood-ratio for testing "H_o: for every j $P_{ij} = p_j^*$ for every i" (Hoel 1954, Anderson and Goodman 1957).(*) Since this H_o corresponds to a constant map, any statistic used for testing it can be a basis for a measure of non-constancy.

The expression $-\sum_i \sum_j \pi_i P_{ij} \log P_{ij}$ is usually referred to (e.g. Khinchin 1957) as the entropy of a Markov chain with probability transition matrix P and steady-state probabilities π. $-\sum_{j=1}^{K} \pi_j \log \pi_j$ is then the entropy of the corresponding constant map. In order to prove "strong perfect mobility", we have to show that this pseudometric indeed attains its minimum (zero) when and only when P is the constant map. We shall now prove that this is indeed the case.

(*) For general relations between information measures (e.g. entropy) and likelihood ratios, see Kullback (1959).

Lemma 3.1

The function $H = -\sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i P_{ij} \log P_{ij}$ attains its maximum over P, subject to the constraints

$$P_{ij} \geq 0 \qquad \text{for every i and j;}$$

$$\sum_{j=1}^{K} P_{ij} - 1 = 0 \qquad \text{for every i; and}$$

$$\sum_{i=1}^{K} \pi_i P_{ij} - \pi_j = 0 \quad \text{for every j,}$$

at $P_{ij} = \pi_j$ for every i and j. This maximum is unique if $\pi_j > 0$ for every j.

Proof:(*)

The function $H = -\sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i P_{ij} \log P_{ij}$ is concave and if $\pi_i > 0$ for every i (which is what we assume) it is strictly concave $\left(\frac{\partial^2 H}{\partial P_{ij}^2} = \frac{-\pi_i}{P_{ij}} < 0\right)$. Since we maximize it subject to linear constraints, the Kuhn-Tucker conditions (Zangwill 1969) are necessary and sufficient for maximum. Let the sets of corresponding multipliers be $\{\alpha_{ij}\}$, $\{\beta_i\}$ and $\{\gamma_j\}$. The Kuhn-Tucker conditions for this case are

$$\pi_i(\log P_{ij} + 1) = \alpha_{ij} + \beta_i + \gamma_j \pi_i \text{ for every i and j, and}$$

$\alpha_{ij} P_{ij} = 0$ for every i and j. For the particular choice, $P_{ij} = \pi_j$ for every i and j, they reduce to $\pi_i \log \pi_j + \pi_i = \beta_i + \pi_i \gamma_j$ for every i and j. Since this system has a solution ($\beta_i = \pi_i$ for every i,

(*) This proof may be viewed as a natural generalization of the one commonly used (e.g. Theil 1972) in the case of (static) distributions. I am indebted to E. Choo and J. Kallberg for the idea.

$\gamma_j = \log \pi_j$ for every j), our choice corresponds to a (unique) maximum. ■

The idea of using entropy as a "measure of distance" in the social mobility context is not new; it has been used by Theil (1972, Ch. 5). However, the notion of distance that he used, $\sum_i q_i \log \frac{q_i}{p_i}$, is not a pseudo-metric, since it does not satisfy the triangle inequality. Our notion of the "entropy excess" of a Markov map over its corresponding constant map thus seems more natural. Nevertheless, Theil made a particular use of his notion of distance to obtain a measure of mobility which we wish to generalize in the last subsection.

3.2.5 <u>From Non-Constancy to Social Inheritance - Introducing Period-Consistency</u>

As we already mentioned in the previous section, a natural way of constructing measures of social inheritance is obtained by observing the rate at which the measures of non-constancy go to zero, i.e. the rate of convergence to zero of the sequence $\{f(T^n)\}$.

Now, there are two common families of indicators of assymptotic rate of convergence (see Ortega and Rheinboldt 1970). For any measure of non-constancy f(T), those will be the "quotient convergence factors"

$$Q_m(f,T) = \limsup_{n \to \infty} \frac{f(T^n)}{[f(T^{n-1})]^m} \qquad m \in [1,\infty),$$

and the "root-convergence factors"

$$R_m(f,T) = \begin{cases} \limsup\limits_{n \to \infty} \; [f(T^n)]^{1/n} \quad (*) & \text{if } m = 1 \\ \\ \limsup\limits_{n \to \infty} \; [f(T^n)]^{1/m^n} & \text{if } m > 1 \,. \end{cases}$$

What is encouraging is that both families of measures are period-consistent for <u>every</u> non-constancy measure f:

<u>Lemma 3.2</u>

Both $Q_m(f,T)$ and $R_m(f,T)$ are period consistent for every non-constancy measure f.

Proof:

We can assume without loss of generality, that m = 1. Now, consider maps T and T' such that $Q_1(f,T) \geq Q_1(f,T')$ and $R_1(f,T) \geq R_1(f,T')$.

$$Q_1(f,T^k) = \limsup_{n \to \infty} \frac{f[(T^k)^n]}{f[(T^k)^{n-1}]} = \limsup_{n \to \infty} \frac{f(T^{kn})}{f(T^{kn-k})}$$

$$= \limsup_{n \to \infty} \frac{f(T^{kn})}{f(T^{kn-1})} \cdot \frac{f(T^{kn-1})}{f(T^{kn-2})} \cdots \frac{f(T^{kn-k+1})}{f(T^{kn-k})}$$

$$= \prod_{i=0}^{k-1} \limsup_{n \to \infty} \frac{f(T^{kn-i})}{f(T^{kn-i-1})}$$

$$\geq \prod_{i=0}^{k-1} \limsup_{n \to \infty} \frac{f(T'^{kn-i})}{f(T'^{kn-i-1})} = Q_1(f,T'^k).$$

Also,

$$R_1(f,T^k) = \limsup_{n \to \infty} \{f[(T^k)^n]\}^{1/n}$$

$$= \limsup_{n \to \infty} \{[f(T^{kn})]^{1/kn}\}^k$$

(*) This particular convergence factor was used by Sommers and Conlisk (1979) to construct a status-correlation measure.

$$= \{\limsup_{n \to \infty} [f(T^{kn})]^{1/kn}\}^k$$

$$\geq \{\limsup_{n \to \infty} [f(T'^{kn})]^{1/kn}\}^k = R_1(f, T'^k) .$$

■

One conclusion from this Lemma is that any function M(T) which can be written in the form $Q_m(f,T)$ or $R_m(f,t)$ for some f (although this may not be the natural way of defining or calculating it) is period-consistent. In particular, Theil (1972, Ch. 5) showed that the second largest eigenvalue can be obtained as $Q_1(f,T)$.(*) Hence $|\lambda_2|$ is period consistent (Shorrocks 1978).

Let M be a measure of social inheritance and let T and T' be maps such that $M(T) > M(T') > 0$, $M(T'^k) > 0$ and $M(T^k) = 0$ for some k (i.e. T^k is a constant map while T'^k is not). Such measure M is not period consistent, and both delta-coefficient and entropy suffer from this deficiency.(**) The entropy-based measure, however, exhibits a property, related to period-consistency, which is not without some appeal, and will be described below.

(*) Theil defined I_t to be $\sum_i \pi_i \log \frac{\pi i}{Y_t(i)}$, approximated it by the quadratic approximation $\frac{1}{2} \sum_i \frac{(Y_t(i)-\pi_i)^2}{\pi_i}$, and showed that $\lim_{t\to\infty} {}^{I_t}/_{I_{t-1}} = \lambda_2^2$. Shorrocks also obtained $|\lambda_2|$ by a limiting procedure of this nature, using the concept of "assymptotic half life".

(**) This particular problem, related to finite-convergent Markov chains, can be eliminated by some appropriate domain restriction, but it is doubtful whether complete period consistency will be achieved.

Let $P_{ik_1k_2...k_r}$ be the (joint) probability that a process which starts in class i first goes to k_1, then to k_2 etc., defined over r generations. Denote the entropy of this distribution by $H_i^{(r)}$, and let $H^{(r)} = \sum_{i=1}^{K} \pi_i H_i^{(r)}$, where $H^{(1)} = H$ is the entropy of the Markov chain as defined previously. Then it can be shown (Khinchin 1957) that $H^{(r+s)} = H^{(r)} + H^{(s)}$ for every r and s.

3.3 Conclusion

In this chapter we investigated the problem of measuring social inheritance. Although due to the multiplicity of issues involved no clear-cut procedures emerge, we believe that the systematic scheme of generating measures presented can assist social scientists in this task.

Generally, once the social inheritance aspect of social mobility has been singled out as the object to be measured, it is of central importance to make one's "social priorities" within this object explicit. Operationalizing these concepts is then a major step towards constructing a measure of non-constancy, though a certain amount of flexibility still remains in the actual choice, which will be made by trading-off mathematical properties and convenience.

Measures of social inheritance can then be constructed by observing the rate at which the measures of non-constancy, for increasingly long periods, converge to zero. This will also ensure period-consistency, even if the non-constancy measure did not have it by itself.

BIBLIOGRAPHY

Anderson T.W. (1954) "Probability Models for Analyzing Time Changes in Attitudes", in Mathematical Thinking in the Social Sciences, P.F. Lazarsfeld (ed.), The Free Press of Glencoe, New York.

__________, and L.A. Goodman (1957) "Statistical Inference About Markov Chains" Ann. Math. Stat., 28, pp. 89-110.

Barlow R.E. and F. Proschan (1975) Statistical Theory of Reliability and Life Testing, Holt, Rinehart and Winston.

Bartholomew D.J. (1973) Stochastic Models for Social Processes (2nd edition), John Wiley & Sons.

Billingsley P. (1968) Convergence of Probability Measures, Wiley, New York.

Bishop Y.M.M., S.E. Feinberg, and P.W. Holland (1975) Discrete Multivariate Analysis: Theory and Practice, The M.I.T. Press.

Blumen, I., M. Kogan and P.J. McCarthy (1955) The Industrial Mobility of Labour as a Probability Process, Cornell University Press, Ithaca, New York.

Boudon R. (1973) Mathematical Structures of Social Mobility, American Elsevier Co., New York.

Breiman L. (1968) Probability, Addison-Wesley.

Brosh I., and Y. Gerchak (1978) "Markov Chains with Finite Convergence Time", Stochastic Processes and Their Applications, 7, pp. 247-253.

Brumelle S.L. and R.G. Vickson (1975) "A Unified Approach to Stochastic Dominance" in Stochastic Models in Finance, W.T. Ziemba and R.G. Vickson, Eds. Academic Press.

Cinlar E. (1975) Introduction to Stochastic Processes, Prentice-Hall.

Cohen J.E. (1976) "Ergodicity of the Age Structure in Populations with Markovian Vital Rates, I: Countable States", J. Amer. Stat. Assoc., 71, pp. 325-339.

Coleman J.S. (1964) Introduction to Mathematical Sociology, The Free Press of Glencoe and Collier-Macmillan, London.

Conlisk J. (1976) "Interactive Markov Chains", J. Math Soc. 4, pp. 157-85.

__________, (1978a) "A Stability Theorem for an Interactive Markov Chain", J. Math. Soc., 6, pp. 163-8.

__________, (1978b) "The Law of Supply and Demand as a Law of Markov Chains", Discussion Paper 78-14, Dept. of Economics, Univ. of Cal., San Diego.

Cordey-Hayes M. and D. Gleave (July 1974) "Dynamic Models of the Interaction between Migration and the Differential Growth of Cities", IIASA Report RR-74-9.

Duncan O.D. (1966) "Methodological Issues in the Analysis of Social Mobility", in Smelser and Lipset (1966).

Gerchak Y. (Oct. 1978) "Markovian Models Allowing Interaction Among Individuals", Working Paper No. 611, Faculty of Commerce and Business Administration, Univ. of B.C.

Ginsberg R.B. (1971) "Semi-Markov Processes and Mobility" J. Math Sociol., 1, pp. 233-62.

__________, (1978a) "The Relationship between Timing of Moves and Choice of Destination in Stochastic Models of Migration", Envir. Plann. A, 10, pp. 667-79.

__________, (1978b) "Testing Stochastic Models of Timing in Mobility Histories: Comparison of Information Derived from Different Observation Plans", Fels Discussion Paper #134, School of Public and Urban Policy, University of Pennsylvania.

Givon M. and D. Horsky (1978) "Market Share Models as Approximators of Aggregated Heterogeneous Brand Choice Behavior", Man. Sci., 24, pp. 1404-16.

Goodman L.A. (1969) "On the Measurement of Social Mobility: An Index of Status Persistence", Amer. Sociol. Rev., 34, pp. 831-850.

Grinold R.C. and K.T. Marshall (1977) Manpower Planning Models, North Holland.

Hajnal J. (1958) "Weak Ergodicity in Nonhomogeneous Markov Chains", Proc. Cambr. Phil. Soc., 54, pp. 233-246.

Halmos P.R. (1958) Finite Dimensional Vector Spaces, 2nd Ed., D. Van Nost and Company.

Henry N.W., R. McGinnis and H.W. Tegtmeyer (1971) "A Finite Model of Mobility", J. Math. Sociol., 1, pp. 107-16.

Hoel P.G. (1954) "A Test for Markoff Chains", Biometrica, 41, pp. 430-433.

Holland P.W. and S. Leinhardt (1977) "A Dynamic Model for Social Networks", J. Math. Sociol., 5, pp. 5-20.

Holley R.A. and T.M. Liggett (1975) "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model", Ann. Prob., 3, pp. 643-63.

Huff J.O. and W.A.V. Clark (1978) "Comulative Stress and Comulative Inertia: A Behavioral Model of the Decision to Move", Envir. Plann., A10, pp. 1101-19.

Isaacson D.L. and R.W. Madsen (1976) Markov Chains: Theory and Applications, Wiley.

Jones J.M. (1973) "A Composite Heterogeneous Model for Brand Choice Behavior", Man. Sci., 19, pp. 499-509.

Khinchin A.I. (1957) Mathematical Foundations of Information Theory, Dover Publications.

Kolm S.C. (1976) "Unequal Inequalities. I", J. Econ. Theo., 12, pp. 416-42.

Krantz, D.H., R.D. Luce, P. Suppes and A. Tversky (1971) Foundations of Measurement, Academic Press.

Krishnan K.S., and M.J. Beckman (1979) "Dynamic Disaggregate Choice Models, with an Application in Transportation", Decision Sciences, 10, pp. 218-231.

Kullback S. (1959) Information Theory and Statistics, Wiley.

Lazarsfeld P.F., and N.W. Henry (1968) Latent Structure Analysis, Houghton Mifflin Company, Boston.

Long L.H. (1970) "On Measuring Geographic Mobility, J. Amer. Stat. Ass., 65, pp. 1195-1203.

McFarland D. (1970) "Inter-Generational Social Mobility as a Markov Process: Including a Time-Stationary Markovian Model that Explains Observed Declines in Mobility Rates over Time", Amer. Sociol. Rev., 35, pp. 463-476.

__________, (1974) "On the Relationship between Mathematical Models and Empirical Data, and the Relationship between Duration of Stay and Subsequent Mobility", Amer. Sociol. Rev., 39, pp. 883-884.

McGinnis R. (1968) "A Stochastic Model of Social Mobility", Amer. Sociol. Rev., 33, pp. 712-721.

March J.C. and J.G. March (1977) "Almost Random Careers: The Wisconsin School Superintendency 1940-1972", Adm. Sci. Quar., 22, pp. 377-409.

Marshall K.T. (1973) "A Comparison of Two Personnel Prediction Models", Oper. Res., 21, pp. 810-822.

Massy W.F., D.B. Montgomery and D.G. Morrison (1970) Stochastic Models of Buying Behavior, The M.I.T. Press, Cambridge, Mass.

Matras J. (1960) "Differential Fertility, Inter-Generational Occupational Mobility and Change in the Occupational Distribution: Some Elementary Interrelationships", Population Studies, 15, pp. 187-97.

__________, (1967) "Social Mobility and Social Structure: Some Insights from the Linear Model", Amer. Soc. Rev., 32, pp. 608-14.

Mayer T.F. (1972) "Models of Intragenerational Mobility", in Sociological Theories in Progress, Vol. 2, J. Berger, M. Zeldtich Jr. and B. Anderson (eds.), New York: Houghton Miffin.

Meredith J. (1973) "A Markovian Analysis of a Geriatric Ward", Man. Sci., 19, pp. 604-12.

Morrison D.G. (1978) "On Linearly Increasing Mean Residual Lifetimes", J. Appl. Prob., 15, pp. 617-620.

__________, W.F. Massy, and F.N. Silverman (1971) "The Effect of Non-Homogeneous Populations on Markov Steady-State Probabilities", J. Amer. Statist. Ass., 66, pp. 268-274.

Moyal J.E. (1962) "The General Theory of Stochastic Population Processes", Acta Math., 108, pp. 1-31.

Neveu J. (1965) Mathematical Foundations of the Calculus of Probability, Holden-Day.

Ortega J.M. and W.C. Rheinboldt (1970) Iterative Solution of Nonlinear Equations in Several Variables, Academic Press.

Prais S.J. (1955a) "Measuring Social Mobility", J.R. Statist. Soc., A118, pp. 56-66.

__________, (1955b) "The Formal Theory of Social Mobility", Population Studies, 9, pp. 72-81.

Pullum T.W. (1975) Measuring Occupational Inheritance, Elsevier Amsterdam, New York.

Ross S.M. (1970) Applied Probability Models with Optimization Applications, Holden-Day.

Royden H.L. (1963) Real Analysis, The Macmillan Company.

Schinnar A.P. and S. Stewman (1978) "A Class of Markov Models of Social Mobility with Duration Memory Patterns", J. Math. Sociol., 6, pp. 61-86.

Shorrocks A.F. (1978) "The Measurment of Mobility", Econometrica, 46, pp. 1013-24.

Silcock H. (1954) "The Phenomenon of Labour Turnover," J. R. Statist. Soc., A117, pp. 429-40.

Singer B. and S. Spilerman (1974) "Social Mobility Models for Heterogeneous Populations" in H.L. Costner (ed.) Sociological Methodology 1973-4, San Francisco: Jossey-Bass.

__________, and __________, (1976) "Some Methodological Issues in the Analysis of Longitudinal Survey," Annals of Economic and Social Measurement, 5, pp. 447-74.

Smallwood, D.E. (1975) "Competition, Regulation and Product Quality in the Automobile Insurance Industry" in Promoting Competition in Regulated Markets, Phillips A. (ed.) The Brookings Institution.

__________, and J. Conlisk (1979) "Product Quality in Markets where Consumers are Imperfectly Informed," Quar. J. Econ., 93, pp. 1-23.

Smelser N.J. and S.M. Lipset (1966) Social Structure and Mobility in Economic Development, Chicago, Illinois: Aldine.

Sommers P.M. and J. Conlisk (1979) "Eigenvalue Immobility Measures for Markov Chains," J. Math. Sociol., 6, pp. 253-276.

Spilerman, S. (1972a) "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain," Amer. Sociol. Rev., 37, pp. 277-294.

__________, (1972b) "Extensions of the Mover-Stayer Model," Amer. J. Sociol., 78, pp. 599-626.

Spitzer F. (1970) "Interaction of Markov Processes," Adv. Math., 5, pp. 246-290.

Stewman S. (1976) "Markov Models of Occupational Mobility: Theoretical Development and Empirical Support. Part I: Careers," J. Math. Sociol., 4, pp. 201-245.

Theil H. (1972) Statistical Decomposition Analysis, North-Holland, Amsterdam.

Thompson W.A., and I. Vertinsky (1975) "Application of Markov Chains to Analysis of a Simulation of Birds' Foraging," J. Theor. Biol., 53, pp. 285-307.

Tuma N.B. (1976) "Rewards, Resources, and the Rate of Mobility: A Non-stationary Multivariate Stochastic Model," Amer. Sociol. Rev., 41, pp. 338-360.

__________, M. Hannan and L.P. Groenveld (1979) "Dynamic Analysis of Event Histories," Amer. J. Sociol., 84, pp. 820-854.

Vroom V.H. and K.R. MacCrimmon (1968) "Towards a Stochastic Model of Managerial Careers", Admin. Sci. Quart., 13, pp. 26-46.

White H.C. (1970) Chains of Opportunity, Harvard University Press, Cambridge, Mass.

Wise D.A. (1975) "Personal Attributes, Job Performance, and Probability of Promotion", Econometrica, 43, pp. 913-931.

Yasudo S. (1964) "A Methodological Inquiry into Social Mobility", Amer. Sociol. Rev., 29, pp. 16-23.

Zangwill W.I. (1969) Non-Linear Programming - A Unified Approach, Prentice-Hall.